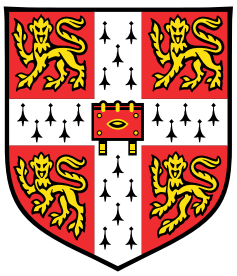
The SSIP approach to solvation



Mark David Driver

Supervisor: Prof. C. A. Hunter

Department of Chemistry
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Robinson College September 2019

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, tables and equations.

Abstract

Intermolecular interactions, including hydrogen bonding, are important for governing many processes, including solvation, protein ligand binding and crystallisation. Previously in the Hunter group, the Surface Site Interaction Point (SSIP) model of intermolecular interactions has been used to describe molecules as a set of discrete hydrogen bonding sites.

In this thesis I describe the prediction of phase properties using this molecular description, after giving an introduction to the wider area of phase property prediction.

Experimental equilibrium constants for formation of hydrogen bonded complexes was collated to allow for the reparameterisation of the relationship used to convert calculated electrostatic potentials into the hydrogen bond parameters used to describe the interaction properties of SSIPs, and a curated database of this information was created.

An overhaul of the software infrastructure, to allow greater exploitation of automation in the workflow, was undertaken. Canonicalisation of data formats for information produced during the calculation has also been undertaken to provide a foundation for work in the following chapters.

Calculation of partition coefficients for a series of molecular datasets have been used to benchmark the performance of the surface site interaction model for liquids at equilibrium (SSIMPLE) which uses the SSIP description for molecules to calculate free energies.

Functional Group Interaction Profiles (FGIP) describe the energy change of two solute SSIPs interacting as a function of solute SSIP values. These profiles provide insight into the strength of interactions within different solvents. A selection of profiles for a range of solvents are included as examples. The profiles provide a useful guide for how to tune interactions when designing systems which rely on intermolecular interactions for association.

Consideration of the solvation free energy of a single SSIP in a solvent as a function of SSIP value, has led to the development of a metric for solvent similarity. The comparison of solvation free energy curves for different solvents and solvent mixtures led to the construction of similarity dendrograms that can be used for solvent selection in experimental systems.

Expansion of the approach to be used to examine the temperature dependence of the interactions, allowed the exploration of vapour liquid equilibria.

Acknowledgements

I have been supported by a lot of people throughout this PhD that I would like to acknowledge.

I wish to start by thanking my supervisor, Prof. Chris Hunter for the opportunity to undertake my PhD and Part III project in his research group. The stimulating discussions we have had about the development of the SSIP approach and its implementation has underpinned my PhD. His professional support, guidance and involvement has been invaluable.

I want to give thanks for the warm welcome I received from the Hunter research group and the discussions in group meetings about my work. I wish to highlight the contribution of Dr. Mark Williamson for the guidance in the development of my software engineering and programming skills during the course of my PhD which have allowed me to explore the topics detailed in this PhD, in addition to proof reading this thesis. I also want to thank my companions in the computational office of the group over the last five years for their discussions throughout the course of my PhD and preceding master's project.

From outside of the research group I wish to thank Dr. Alex Thom for first introducing me to computational chemistry research, as an undergraduate summer student. In addition I wish to thank Dr. Paul Griffiths and Dr. Bill Nolan for supervision during my undergraduate degree and for directing me towards theoretical chemistry.

I also wish to thank my family for their support, and all the friends I have made in Robinson College since starting there as an undergraduate student, especially the 2011 Natural Scientists and the college staff who have been around with me since I started this journey eight years ago.

Table of contents

Chapter 1

Introduction

Understanding of the solvation properties of molecules in the liquid phase is fundamental to several industrial and biological processes.

Free energy of transfer of solutes between solvent systems is used extensively in pharmaceutical development to assess the bioavailability of possible active pharmaceutical ingredients. The association of solutes relies on the formation of favourable interactions. Modification of the solvent composition can prevent these interactions, disfavouring association and further aggregation. This reduces the complexity of supramolecular structures possible in the solution. The similarity of solvation ability of solvents is influenced by the intermolecular interactions solvents can form with solutes. Evaluation of the similarity of solvents is used to select replacements or alternatives for use in synthesis. The distribution of molecules described by the vapour-liquid equilibria (VLE) requires the assessment of the temperature dependence of interactions, to be able to accurately describe the expansion and evaporation of liquids. Selection of experimental conditions for synthesis of pharmaceutical agents at larger scales requires knowledge of VLE for the undertaking of safe and efficient reactions.

All of these processes are based on the understanding of molecular interactions. The surface site interaction point (SSIP) approach to molecular description is used in this work to describe molecular interactions, in an attempt to describe and predict these solvation effects. This is done using the Surface Site Interaction Model for the Properties of Liquids at Equilibrium (SSIMPLE).

1.1 Phase transfer energy

Phase transfer free energies are measured regularly to assess the suitability of potential drugs, as a measure of the ability of the compound to travel through the body to the target. Passage through aqueous phases (such as blood and the extracellular fluid) and cell membranes (a hydrophobic phase) is required for a pharmaceutical agent to reach the required binding site within a cell. The value of the partition coefficient between octan-1-ol and water is one of the criteria in Lipinski's 'Rule of 5' [1] for predicting suitability for potential drug candidates. A value no greater than 5 is required under these rules.

Solubility and partition coefficients can refer to the dissolution or transfer of a solute between any solvents. A partition coefficient is the ratio of concentrations of a compound in a mixture of two immiscible solvents which are in equilibrium, represented by log P, which is defined in equation (1.1). It can also be expressed in terms of the solvation energies, which is shown in equation (1.2).

$$\log P_{1,2} = \log\left(\frac{[A]_1}{[A]_2}\right) \tag{1.1}$$

$$\log P_{1,2} = \frac{\log(e)\,(\Delta G_1 - \Delta G_2)}{RT} \tag{1.2}$$

Where $\log P_{1,2}$ is the partition coefficient for transfer from solvent 1 to solvent 2 of molecule A; $[A]_1$ and $[A]_2$ are the concentrations of the unionised forms of molecule A in solvents 1 and 2 respectively; ΔG_1 and ΔG_2 are the solvation energies of molecule A in solvent 1 and 2 respectively; R is the gas constant and T is the temperature.

Equation (1.2) can be reached from equation (1.1) by relation of the chemical potential of molecule A in the phases, at the limit of infinite dilution.

Log S is used to represent solubility coefficients. This describes the partition of the compound between a solvent and the molecule in its standard state. The partition coefficient can therefore be expressed as the difference between the solubility coefficients for the molecule in the two solvents. These solubility and partition coefficients are for the neutral state of a molecule. The distribution coefficient, log D, is related to log P, which is the ratio of all forms of the molecule present in the two phases. It includes concentrations of protonated and deprotonated forms, and therefore varies with pH.

Since the partition between octan-1-ol and water is used in drug discovery, the existing methods to predict log P implicitly mean the partition between octan-1-ol and water, unless

explicit solvents are given. Likewise, solubility prediction implicitly refers to solubility of the compound in water.

1.1.1 Experimental determination of partition coefficients

The partition coefficient and solubility coefficient can be determined by experimental measurement. The classical reference method used for partition coefficients is the shake flask method, with all new methods calibrated to this method[2–4].

A flask containing the two solvents has the solute of interest added to it, and the solute is allowed to equilibrate between the two solvent phases. The concentration of solute in each phase is then measured by some analytical technique; either by quantitative spectroscopy or measuring mass of solute in extracted aliquots of known volume. This process is time consuming, leading to the need for faster and more efficient methods.

The creation of new methods or application of existing techniques to this field has been done to try and reduce the time required to collect the experimental partition coefficients.

The pH dependence of solute precipitation has been used as an alternative approach to obtain solubility coefficients [3–5]. The precipitation rate of the neutral species from a supersaturated solution is used to find the solubility after the pH of the solution is changed.

Cyclic voltammetry can also be used to determine the partition coefficients[6, 7]. This method uses the potential-pH profiles generated, along with the pK_a of the solute to be able to determine the partition coefficients.

Another approach is the use of HPLC (high performance liquid chromatography)[8–10]. This uses the retention time of the solutes in the column which is related to the partition coefficient between the two phases in the column. Calibration with compounds of known partition coefficients of the solvent of interest allows the generation of a linear relationship to relate retention time to the partition coefficient to be measured. This means the method is reliant on data from another method to be able to validate the results.

All these methods require a large amount of resources to be committed. A high throughput screening approach has been applied to some methods [2, 6], to increase speed, but still requires the synthesis of moderate quantities of each compound (on the order of 1-10mg). The prediction of these values would be of great benefit, providing a more efficient way to evaluate the potential suitability of drug candidates, in a much shorter time frame and without the need to synthesise them.

1.2 Phase transfer free energy calculation methods

Several methods have been developed to predict solubility coefficients and partition coefficients. These methods fall into broad categories which group the methods based on the fundamental approaches taken for the prediction. Reviews by Mannhold *et al.* [11], Nieto-Draghi *et al.* [12] and Skyner *et al.* [13] each provide slightly different taxonomies of the methods.

The prediction methods can be broadly categorised as either empirical functions, implicit solvent simulations or explicit solvent simulations.

1.2.1 Empirical phase transfer functions

Empirical functions for energy prediction consist of parameterised functions that use correlation to experimental measurement of molecular properties. These methods can further be classified into group contribution methods and quantitative structure property relationships (QSPR).

Group Contribution methods

Group contribution methods are based on the summation of contributions from different blocks in the chemical structures. These fall into three main categories. Fragment based, atom based and topological descriptor based methods.

Fragment based methods have molecules cut into fragments, with the application of correction factors to compensate for intramolecular interactions. Equation 1.3 shows the general formula of fragmentational methods.

$$\log P = \sum_{i=1}^{n} a_i f_i + \sum_{j=1}^{m} b_j F_j \tag{1.3}$$

The first summation considers the contribution of fragment constant f_i, with incidence a_i. The second term applies correction factor, F_j, with its frequency b_j. The definition of fragments larger than single atoms guarantees that significant electronic effects are comprised in one fragment. This advantage becomes a disadvantage if key fragments are missing from the definitions or arbitrary fragmentations are used, which can cause a calculation to fail. The methods all rely on training sets to generate the parameters used in the individual implementation of the algorithm. KLogP [14], KowWIN [15, 16], CLOGP [17], ADCLogP [18], AB/LogP [19] and also AQUAFAC [20, 21] are implementations of this method.

Atom based methods split up the molecule into individual atoms instead of fragments. This provides a simpler general form for the methods, shown in equation 1.4.

$$\log P = \sum_{k=1}^{n} a_k N_k \tag{1.4}$$

With the summation over all atom types k, where a_k, the contribution of atom type k, and N_k, the number of atoms of type k. Atom types are classified based on element as well as the chemical environment. This methodology fails to deal with long-range interactions within a molecule, and also relies on parameterisation of all the potential functional group environments. Examples of the implementation of atom-based methods are the original Ghose-Crippen approach [22] (this is used in AlogP[23], MOLCAD[24], TSAR [25] and PrologP[26]), the refined Ghose-Crippen approach [27] (which is implemented in ALOGP98), and XLOGP [28]. The XLOGP methods apply correction factors to account for longer range interactions, with the Ghose-Crippen and refined Ghose-Crippen methods are purely atom-additive based models.

Topological descriptor methods use information based on the connectivity of atoms within the molecule in the fitting of the empirical relationship. This produces relationships of similar form to the atom or fragment based methods. Longer range interactions are taken into account by considering the classification of atoms separated by multiple bonds.

MLOGP [29] was one of the first such methods developed. The two basic descriptors used are the numbers of lipophilic and hydrophilic atoms in the molecule. A series of eleven correction factors based on factors e.g. the number of unsaturated bonds and number of rings, provides a simple formula based on fitting to experimental data for 1230 compounds.

Graph molecular connectivity was used by Junghans and Pretsch to develop TLOGP [30]. This uses a vector representation of molecules to provide descriptors. Its performance depends on the similarity to the training set molecules.

E-state indices [31] cover both topological and valence states of atoms. Models developed based on these descriptors are VLOGP [32] and ALOGPS [33].

Quantitative Structure Property Relationships (QSPR)

QSPR methods are based on the assumption that molecules with similar structures will have similar properties, with changes in macroscopic properties determined by variations between the molecular structures. This method tries to quantitatively correlate an experimentally observable property to a series of molecular descriptors, which encode the features of the molecules.

Abraham's Linear Solvation Energy Relationship [34, 35] is one of the earliest QSPR methods and relies on five different descriptors. ABSOLV [36] and ChemProp [37] both contain implementations of this method. Some of the parameters used require determination by experiment, which is not always practical, so fragment based approaches were developed to overcome some of these issues.

Other examples of QSPR implementations include those in [38–41]. These methods have a limited applicability domain, depending on the training and test sets of data used, with over fitting to training data also a serious concern [42].

1.2.2 Implicit solvation methods

When undertaking electronic structure calculations of individual molecules the default view is to consider the molecule in isolation when solving the molecular Schrödinger equation. This approximation can be considered to represent the gas phase, where molecules are far apart, but this neglects effects from the surrounding solvent molecules on the electron density if you consider a molecule in a liquid. Implicit descriptions of solvent have been developed to model these effects, to remove the computational expense from having to calculate the electronic structure of solvent molecules explicitly, which would lead to a dramatic increase in computational cost due to the high scaling of electronic structure methods [43].

The implicit solvent models describe the solvent as a dielectric continuum, in which the molecule of interest is embedded into a cavity[44]. The size of the cavity is dependent on the continuum method used as well as the solvent description.

Activity coefficients are then calculated by consideration of the energy of the interaction between the molecule and solvent continuum at the cavity surface as the interactions between them are gradually turned on in an approach analogous to thermodynamic integration [44]. The solvation energies are then calculated using parameterised relationships, based on experimental data.

Examples of polarisable continuum models are the SM8 model from Truhlar *et al.*[45, 46] and the conductor-like screening model (COSMO) developed by Klamt [47]. COSMO has been widely applied to areas related to solvation and phase transfer free energies.

Conductor-like screening model (COSMO)

The conductor-like screening model (COSMO) [47–49] embeds a molecule in a cavity surrounded by a dielectric continuum with a permitivity of $\varepsilon = \infty$. The cavity surface corresponds to the solvent accessible surface, which is defined at a distance from the nucleus

of the nearest atom. The charge that is induced on the cavity surface which would screen the surroundings from the molecular charge is calculated. The surface is then partitioned into segments and a probability density function is calculated to describe the probability of finding a given surface charge density on a unit of surface. This characteristic function is a σ profile ($p^x(\sigma)$), of the charges on the surface. $p^X(\sigma)$ describes the probability of finding a mean screening charge density, σ, on a typical contact segment of the molecule X.

The sigma profile, $p^X(\sigma)$, provides a description of the molecular surface charge density, which can be used to describe intermolecular interactions. For it to be used for prediction, this must be linked to thermodynamical quantities, which was first done in the definition of COSMO for realistic solvation (COSMO-RS)[48–50].

In that work, the σ profile is linked to the chemical potential, the change in free energy with number of particles. This is done by partitioning the molecule into a series of surface segments. The mismatch energy, between two surface segments, can be computed by (1.5) (equation 2 in [48]).

$$E_{misfit}(\sigma_1, \sigma_2) = \frac{1}{2}\alpha(\sigma_1 + \sigma_2)^2 \tag{1.5}$$

Where σ_1, σ_2 are the surface charge densities, and α is given by equation (1.6) (derived in [48]).

$$\alpha = \frac{1.2\pi^{\frac{5}{2}}R_{eff}^3}{4\pi\varepsilon_0} \tag{1.6}$$

With R_{eff} the effective radius of a segment, which was chosen to be 1Å, and ε_0 is the permittivity of free space. From consideration of the partition function, the chemical potential per mole of surface segments was derived, and is shown in equation (1.7) (equation 5 in [48], derived in appendix A in [48]).

$$\mu'_S(\sigma) = -kT\ln\left[\int d\sigma' p(\sigma') exp\left\{-\left(\frac{1}{2}\alpha'(\sigma+\sigma')^2 - \mu'_S(\sigma')\right)/kT\right\}\right] \tag{1.7}$$

Where μ'_S is the chemical potential per mole of surface area, k is the Boltzmann constant, T is the temperature, α' is defined in equation (1.6).

From this calculated chemical potential, the partition coefficient was then found, by the creation of QSPR using the correlation of COSMO parameters to experimental data. This also includes information about the number of aromatic rings. The initial parameterisation

for the generation of the σ profiles made assumptions for the cavity radii for different atoms, refinement to these parameters and others was undertaken in [49].

The COSMO-segment activity coefficient (COSMO-SAC) approach by Lin *et al.* [51, 52] uses the σ profiles from COSMO, to define the activity coefficients for segments. This is used to calculate the restoring force to move from the ideal conductor to the solvent, thus using a different approach to find the solvation energy.

Refinements to the COSMO model have tried to account for the discrepancy between the idealised solvent model and the "real solvent" case, which is done with the COSMO-RS implementation [50].

Activity coefficients describe the divergence from ideality of a particle. The activity coefficients are used as the starting point in derivations for several property calculations, including phase transfer free energies [48, 50, 51, 53]. This can also be used to predict densities and molar volumes [54]

The COSMOfrag [55] implementation removes the requirement to run a QM calculation on new molecules, instead a database of σ profiles for molecules which have already been calculated. A composite σ profile for the molecule is then created through the weighted combination of fragment σ profiles. COSMOfrag is designed to reduce the computational intensity of the task, without loss in accuracy of the predicted values.

1.2.3 Explicit Solvation: Molecular Simulation

Molecular simulation methods can incorporate solvents as explicit molecules. Coarse graining of electron interactions leads to a force field using classical mechanics to describe the interactions between atoms [56, 57]. This is used to simulate large systems using either molecular dynamics (MD) or Monte Carlo (MC) approach to propagate the system to sample phase space[58].

From such simulations, the free energies of solvation can be calculated by extracting the energy components corresponding to interaction between the solvent and solute. This is primarily done by two similar techniques, thermodynamic integration (TI) [59] and free energy perturbation (FEP) [60]. In both methods the energy differences are found by consideration of a thermodynamic cycle, where the end points are linked by coupling parameter. The separation into discrete states then allows summation of the property as the states are converted.

ΔG_{ow} can be abstracted from calculations through the use of a thermodynamic cycle linking the free energies of the solute in different states, which identifies the simulations

required such that thermodynamic integration of the appropriate transitions is calculated [61]. This method has been used to predict free energies of transfer for molecules illustrated by [61–65].

The use of explicit solvent molecules in MD simulations makes the process very computationally demanding due to the size of the solvent box required. Periodic boundary conditions are used to avoid finite size effects on the system.

As the molecule of interest increases in size, the box must also increase to ensure there is no self interaction of the molecule of interest with a copy in an adjacent box. The number of solvent molecules therefore increases dramatically with system size.

Molecular simulation with implicit solvation

To reduce the computational complexity of the simulations, implicit solvent methods have been developed for MD simulations. The Poisson-Boltzmann surface area (PBSA) [66, 67] or the Generalised Born surface area (GBSA) [68] are examples of this.

1.2.4 Machine learning Approaches

Recently, machine learning has gained popularity as a tool to help in the prediction of properties of molecules, where large amounts of experimental data has previously been gathered.

Machine learning is used to infer properties and structure of the data that has been given, to create a mathematical model for the input data[69, 70]. The model can then be used to make predictions. This allows the creation of new empirical relationships without prior knowledge of which features are important in the molecules. It has been applied to phase transfer energy prediction either to augment existing approaches, or to generate new relationships. Skyner *et al.* [13] provides an overview of how the different machine learning algorithms function.

New group contribution methods have been developed using machine learning to assign coefficients and aid in the fragment assignment and characterisation during model development [71, 72]. Machine learning has also been applied to QSPR methods [38], as these methods are also based on simple mathematical models, which can be generated using machine learning frameworks.

MD simulations are a computationally expensive method to find phase transfer and solvation free energies, such that Riniker has proposed using ML with fingerprints generated from MD, MDFP (molecular dynamics fingerprints), for prediction of free energies [73].

Machine learning frameworks to predict chemical properties, including the solvation free energies of molecules, benchmarked to QM methods have also been developed [74–76].

1.2.5 Performance of predictive models

The review by Mannhold *et al.* [11] contains a comparison of 30 different log *P* models using the same molecular test set of 266 molecules. The best predictions gave an RMSD of less than 0.50 for a subset of molecules. However the training set for some of these models was contained in this subset. For the second part of the dataset, which did not overlap with the training set, the models gave poorer results with RMSD.

For a new partition coefficient prediction model to be comparable to the existing best methods from [11], the uncertainty in the predicted value should be just under 1, for log *P* (this corresponds to an error in the phase transfer energy of approximately 5 kJ mol^{-1}). If the model is to perform significantly better than existing models in [11], then an uncertainty value closer to 0.5 log units is required.

1.3 Solvent similarity

The aim to reduce the environmental impact is the core principle of green chemistry [77], leading to greater sustainability of chemical practices. Solvents are routinely used as the medium for chemical reactions as well as during purification. They therefore contribute heavily to the environmental impact of production processes of pharmaceutical agents and agrochemicals, leading to scales [78–81] designed to quantify the impact.

There are two major categories of assessment criteria. Identification of the environmental, health and safety (EHS) impact of a solvent must be considered, by application of "greenness" screening procedures [82] before use. The other approach, life cycle assessment (LCA), involves comprehensive analysis of the impact of the solvent from production, to eventual disposal after use [83]. Criteria weighting is important for the proposed solvents, with multi-criteria decision analysis (MCDA) a possible solution [78, 84].

From these approaches harmful solvents have been identified as candidates to be replaced. Solvent substitution could be undertaken if a suitable alternative with similar properties, but lower environmental impact, can be identified.

1.4 Similarity approaches

Identification of similarities between molecules requires a metric with a quantifiable scale to denote the degree of similarity. The definition of such metrics have been developed to describe molecular structure similarity. This has then allowed the exploration of prediction of properties, such as partition coefficents, that have previously been discussed in section 1.2.

1.4.1 Molecular similarity

Structural similarity between two molecules has been used as a guide for similar properties[85]. There have been multiple methods developed to describe structural similarity based on topological descriptors. The descriptors used can be divided into three classes:

- Whole molecule (1D) descriptors
- Descriptors calculated using the 2D molecule structure
- Descriptors calculated using the 3D molecule structure

Whole molecule descriptors are usually single number properties that describe bulk molecule properties for example: number of stereocentres, molecular weight, partition coefficient. A molecule could be represented by a collection of these values, normally after some standardisation process has been applied to the values.

2D descriptors are computed from a chemical structure diagram, which is encoded as a connectivity graph of all atoms in the molecule. These include topological indices and fragment substructures[31]. A molecular fingerprint can be produced from these values, which consists of a vector summarising the information. The Tanimoto [86], cosine coefficient [87] and dice coefficient [87] are examples of structural similarity metrics which use such 2D fingerprints.

3D topological descriptors are less widely used, due to the difficulty in assigning useful descriptions [87, 88]. Consideration of whether the lowest energy conformation chosen to describe the molecule is sufficient, or if a conformational search is required for descriptor generation. Alignment of the molecules to maximise overlap may also be required for some of the measures. This leads to a complex 3D picture with no clearly preferred way to describe a molecule. Shape similarity methods form the most common class of this type of descriptor [89] including the receiver operator characteristic (ROC) approach [90].

Property similarity

Molecular properties are influenced by molecular structure, so quantitative property structure relationships (QSPR) have been developed.

These produce parameterised functions which are dependent on the values of identified molecular characteristics which have shown some correlation with the property to be predicted.

The characteristics can be found from inspection of the molecular topology, or from simple *ab initio* calculations. These have been used to provide relationships for a variety of molecule properties including partition coefficients[19, 21, 34, 35, 91] and vapour-liquid equilibria [92–96].

1.4.2 Solvent similarity

Possible solvent space covers a wide range of molecules. Quantifying the relative solvation ability of different solvents requires a numerical distance between solvents in this solvent space can be described. Such a distance is also referred to as a metric.

Hildebrand [97] first proposed to use the square root of the cohesive energy density of a solvent, δ (equation 1.8), as a suitable quantity for comparison. The difference between δ values for solvents was used to assess the degree of similarity.

$$\delta = \left[\frac{\Delta H_v - RT}{V_m}\right]^{\frac{1}{2}} \tag{1.8}$$

Where ΔH_v is the heat of vapourisation, R is the gas constant and T is the temperature and V_m is the molar volume.

Hansen later proposed an extension of the Hildebrand parameter to estimate the miscibility of polar and hydrogen bonded systems [98]. This parameter for describing solvent similarity, is given in equation (1.9).

$$R_a = \sqrt{4\left(\delta_{d,2} - \delta_{d,1}\right)^2 + \left(\delta_{p,2} - \delta_{p,1}\right)^2 + \left(\delta_{h,2} - \delta_{h,1}\right)^2} \leq R_0 \tag{1.9}$$

Where R_a is the similarity, δ_d, δ_p, δ_h are the dispersion (non-polar), polar (coulombic) and hydrogen bonding cohesion solubility parameters for the solvents respectively. R_0 is the solubility radius, outside which they are not miscible. This value is not constant for all systems, so has to be experimentally determined for each solvent combination.

The Hansen parameters have been widely used in the paint and coatings industry for the discovery of appropriate solvents [99–101].

1.5 Vapour liquid equilibria (VLE)

The partitioning of molecules between the vapour and liquid phases has implications for processing and handling of substances. An equilibrium between a liquid and its vapour is a function of the temperature and pressure of the vessel it is in. For binary and higher mixtures, the proportion of each component in the two distinct phases may be different, due to different affinities. The collection of VLE data can be used to create pressure-temperature phase diagrams for species. This knowledge is exploited in distillation, to aid in extraction and purification[102].

Experimental determination of VLE information, the composition of the vapour and liquid phases of a system under given temperature and pressure conditions, is time consuming. The prediction of phase composition has therefore been explored using several different approaches.

The methods used for VLE prediction are based on generating equations of state for the thermodynamic quantities of the material being modelled [103]. Classification of the methods divides them into two broad classes: empirical approaches and simulations. There are also some semi-empirical approaches.

1.5.1 Experimental measurement

The experimental determination of the vapour and liquid compositions is undertaken through a protocol that has seen only minor changes as improvements in the equipment have allowed for more efficient sampling [104–108]. The fundamental principle is to place a sample of the material or mixture in a sealed vessel, heated by a heat bath, with a manostat to control the pressure. The system is then allowed to reach equilibrium, before small samples from the vapour and liquid phases are removed, for composition determination by spectroscopic techniques.

The process is time consuming, due to the requirement for potentially long equilibration times between measurements, and the use of high pressures and temperatures for full phase space exploration resulting in increased risks. Collection of mole fractions and partial pressures for all species in the vessel, leads to the collection of redundant information. This is due to the correlation of the thermodynamic relationships, thus a reduced set of information could be collected. The verification of the consistency of such data requires the full collection of the complete data set to quantify the errors [109]. The accurate prediction of the behaviour is therefore sought to reduce the time, cost and potential risks involved in gathering the information.

1.5.2 Equations of state

Equations of state (EoS) link thermodynamic state variables under a given set of physical conditions to describe the state of matter, relating pressure, volume and temperature for a system. The ideal gas law of Clapeyron [110] was the first equation of state developed for a state of matter, shown in equation (1.10). This was based on the relationship between pressure and volume first noted by Boyle in the 17th century[111].

$$P\bar{V} = RT \tag{1.10}$$

Where P is the pressure; $\bar{V}$ is the molar volume. For real gases this relationship does not hold as there are interactions between molecules. The development of equations of state for the description of real gases has been an area of active research since the first equation by van der Waals [112], introduced an equation based on the concept of finite volume of constituent molecules (b) and interactions between molecules (a), shown in equation (1.11).

$$\left(P + \frac{a}{\bar{V}^2}\right)(\bar{V} - b) = RT \tag{1.11}$$

Where a and b are phase composition dependent parameters. This relationship can be rearranged to produce a cubic expression in $\bar{V}$. Two solutions to the cubic expression correspond to the molar volumes in the gas or liquid phases, with the third solution being a spurious result. Further developments of such equations of state are covered in the reviews by Anderko [113], Wei and Sadus [114] and Valderrama [115].

1.5.3 Empirical methods

Empirical approaches to predict vapour-liquid behaviour rely on the creation of equations linking the properties of interest to other observable or calculable properties of a species by a mathematical relationship. Quantitative Structure Property Relationships (QSPR) use the correlation of a series of molecular descriptors to physical properties. QSPR methods have previously been discussed for phase transfer free energies in section 1.2.1. The application of this approach to the prediction of VLE behaviour has been undertaken by several groups.

The UNIFAC approach is a group contribution method which was first applied to the generation of relationships to predict VLE and liquid liquid equilibria (LLE) in 1977 [92]. Improvements to the descriptions used to describe fragments have been produced [93–96], leading to gradual improvements in the model. The applicability domain of the UNIFAC methods are restricted to regions where experimental information is already known. The

extension of the UNIFAC model to incorporate parameters for ionic liquids by Kato *et. al* [108] therefore expands the applicability domain to a classification of solvents of interest for green chemistry [116, 117].

The σ-profile descriptions of molecules generated by COSMO [47] have been used in the creation of semi-empirical relationships to predict the composition of vapour-liquid equilibria, and therefore phase diagrams. Through use of COSMO-RS, activity coefficients are calculated [50, 53], which can then be used in a QSPR function [51, 53, 108, 118–121]. The development of predictive relationships for novel systems can influence emerging fields such as biodiesel purification [122].

1.5.4 Simulations

Simulations of liquid-vapour systems can be used to predict phase properties through propagation of the components through phase space. The Monte Carlo and Molecular Dynamics techniques, previously introduced in section 1.2.3, have also been used in the prediction of VLE and liquid-liquid phase compositions. Examples of Monte Carlo simulations with application to VLE are in references [123–126], with examples of molecular dynamics calculations in references [127–130].

Coarse graining of molecular interactions to simulate species at the mesoscopic scale has been done through the use of Dissipative particle dynamics (DPD) [131, 132]. The coarse grained description used in DPD simulations consists of beads used to represent a molecule or molecule fragment. Parameterisation of the interactions between beads are used to summarise the interactions of the constituent atoms within a bead, reducing the degrees of freedom in the system. This has allowed the study of larger molecular VLE and LLE systems, such that the study of micelle or bilayer formation is possible [133, 134] with DPD.

Statistical Associating Fluid Theory (SAFT)

Statistical Associating Fluid Theory (SAFT) is a methodology to include the effects of association into a given theory to develop equations of state able to deal with VLE and LLE phenomena [135–137]. SAFT extends the thermodynamic perturbation theory developed by Wertheim [138, 139] to incorporate the treatment of mixtures [127, 135]. Overviews of the development of the SAFT approach are presented in reviews by Müller and Gubbins [136, 137]. The SAFT approach is independent of simulation framework, and used in Monte Carlo simulations [135] and molecular dynamics simulations [127, 128]. After development

of SAFT on model systems [127, 135], SAFT has been applied to a wide range of systems including alkanes and polymer mixtures [140–144].

1.5.5 Performance of predictive models

The benchmarking of performance of predictive models for VLE have not been undertaken to the same frequency or scope as in the field of phase transfer energy prediction. Anderko [113] and Valderrama [115] both make limited comparisons of the performances of a selection of a subset of methods in their reviews, which indicate that relative errors of less than 5% in concentrations indicate good performance.

1.6 Intermolecular interactions

Intermolecular interactions play an important role in solvation, the formation of crystals and molecular recognition events (including protein ligand interactions). There have been numerous attempts to try and describe these forces. This has led to the naming of different types of interactions, from our desire to partition the effects into discrete, separate phenomenological classifications. Examples of the classifications include CH−O bonds, halogen bonds, aromatic interactions, cation-π interactions, hydrophobic interactions [112, 145–148].

This leads to an obscuring of the origins of these often related forces. By consideration of the interactions more generally as arising from electrostatic interactions, Hunter [149] has been able to develop a way of describing these interactions using a unified framework for assessing the strength of these interactions.

Taft and coworkers [145, 150] originally worked on quantification of hydrogen bond strength. From experimental measurement of binding constants of various hydrogen bond acceptors with para-fluorophenol in tetrachloromethane, the pK_{HB} was developed for hydrogen bond acceptor strength. This work was then extended by the work of Abraham (summarised in [151]). Measurements of association constants for a collection of hydrogen bond donors with a complementary set of hydrogen bond acceptors was used to generate solvent dependent scales for the strength of hydrogen bond donors and acceptors. The relationship of the association constant to these scales is given in (1.12).

$$\log(K) = c_1 \alpha_2^H \beta_2^H + c_2 \tag{1.12}$$

Where K is the association constant at 298K and c_1, c_2, are solvent dependent constants, α_2^H is the hydrogen bond donor parameter for the first solute and β_2^H is the hydrogen bond acceptor parameter for the second solute.

α_2^H and β_2^H were arbitrarily scaled between 0.0 and 1.0, to provide solvent independent parameters (adjustment of c_1 and c_2 are used to provide the correct association constants in a solvent). The scale has a systematic offset in the origin due to the solvent hydrogen bond donor and acceptor properties of carbon tetrachloride, which was the solvent in which the majority of the measurements were undertaken. Functional groups which are poorer acceptors or donors are never observed to interact with the corresponding binding partner due to competition with the solvent sites which are in excess. This means no experimental parameters can be found for donors and acceptors that are less polar than carbon tetrachloride.

By consideration of the equilibrium between solvent and solute donor and acceptor sites, the above relationship was restated to include the solvent acceptor and donor interactions explicitly, given by the relationship in (1.13) by Hunter [149].

$$\Delta G^o = -RT\log(K) = -(\alpha - \alpha_s)(\beta - \beta_s) + \gamma \tag{1.13}$$

Where ΔG^o is the free energy change on formation of a 1:1 hydrogen bonded complex of the two solutes; α, β are the solute hydrogen bond donor and acceptor parameters; α_s, β_s are the solvent hydrogen bond donor and acceptor parameters respectively. γ is a constant with a value of 6 kJ mol^{-1}.

This relationship removed the need to have empirically determined solvent specific constants for prediction of the binding constants in the solvent, instead the hydrogen bond donor and acceptor parameters of the solvent are required, allowing the wider applicability to all solvents.

The process of gathering experimental information is time consuming and only gives information about the strongest interaction sites on both molecules [152–154]. Mapping of the hydrogen bonding parameters provides a description of all the molecular interactions, based on a series of discrete surface site interaction points (SSIPs) [155] for a molecule. With an understanding of how to describe intermolecular forces, we can attempt to understand and predict what would be observed in different situations.

1.7 Surface site interaction point (SSIP) approach to molecule description

The surface site interaction point (SSIP) approach describes molecules as a series of discrete surface sites. Previous work [149, 155] has shown that the calculated electrostatic potential on the 0.002 $e\ Bohr^{-3}$ isodensity surface of the molecule shows good correlation to the experimental hydrogen bond donor and acceptor parameters. This relationship was then combined with a coarse graining approach to describe the molecule as a collection of SSIPs [155].

Calculation of the molecular electrostatic potential surface (MEPS) of a molecule can be the time limiting step, and has been previously investigated [156]. The electron density isosurface used for the mapping has also been explored [157].

1.7.1 Surface Site Interaction Model for the Properties of Liquids at Equilibrium (SSIMPLE) approach

The SSIPs of a molecule can be used to describe the interactions between molecules in a liquid since they arise from surface contacts. SSIPs in a phase are either bound to another SSIP or unbound, and free to move in the phase. The SSIMPLE approach [158] calculates the speciation of these states at equilibrium for all SSIPs in a phase. From the concentrations of the free SSIPs and bound complexes at equilibrium, the free energies of molecules in the phase are calculated, allowing transfer free energies to be calculated [158].

1.7.2 Applications of the SSIP description to crystals

In crystalline states, molecules are packed in a periodic array, forming the maximum possible number of intermolecular interactions. There are many factors influencing the stability of the crystalline state, making exact energies difficult to predict [159].

Cocrystal prediction

Cocrystals contain two or more molecular components in a defined stoichiometry. They are a class of crystals of interest for the formulation of active pharmaceutical ingredients (APIs) to improve drug potency by crystallisation with new co-formers.

The pairwise interactions between SSIPs has been used to rank the relative favourability of a range of potential coformers with an API, to provide guidance for experimental exploration [160]. This has been validated in applications to a series of systems in [161–164].

Crystal solubility

Bioavailability of compounds influences the effectiveness of drugs, with water solubility key for oral bioavailability. The effects on solubility of APIs by formation of cocrystals is explored in [165]. It shows that cocrystal formulations may lead to large increases in solubility, compared to neat formulations.

Calculating SSIPs from crystal data

Crystal structures contain information on the closest contacts between molecules. The contacts represent the outcome of competition to form the strongest possible interactions. The solution based parameters of α and β were shown to correlate to the probability of observing these interactions[166–168] in the Cambridge Structure Database (CSD) [169].

1.8 Research Objectives

Building on previous work using the SSIP description of molecule interactions, in this work the applications the SSIP and SSIMPLE approach to solvation will be expanded.

The collection and validation of binding constant measurements from literature was undertaken to allow for the reparameterisation of the method used to map calculated electrostatic potentials to the experimental hydrogen bonding parameters. As part of this work in chapter 2, a detailed analysis of suitable data structures was undertaken. The design of methods and structures to display and search the data were also developed.

The robust exploration of liquid phase based properties using the SSIP approach to molecular interactions required the redesign of the computational framework, to allow for greater automation. As part of this redesign, the code was rebuilt with an underlying modularisation scheme. This process is discussed in chapter 3. Knowledge from the creation of an experimental database of information was used to inform the design of data structures to be used as inputs and outputs to the modular components of the computational work flow. This design feature has allowed addition to the infrastructure to expand the applications of the approach that have been explored in a systematic manner.

The SSIMPLE method can be used to calculate free energies of transfer for molecules as well as partition coefficients. Benchmarking of the SSIMPLE method on data sets was undertaken after the reconstruction of the software to evaluate the effectiveness of the SSIMPLE model and the SSIP descriptions used in prediction.

The interaction energy of two solutes in a solvent is a function of solvent as well as the solute hydrogen bond donor and acceptor strength. This interaction energy is plotted in a functional group interaction profile (FGIP), which is characteristic for a solvent. A method to generate these plots for any solvent is presented in chapter 5.

The solvation energy of a solute depends on the strength of its hydrogen bond donor or acceptor ability and the possible interactions with a solvent. A solute hydrogen bond donor will interact favourably with hydrogen bond acceptors and unfavourably with hydrogen bond donors in the solvent. Conversely a hydrogen bond acceptor on a solute will interact favourably with solvent hydrogen bond donors and unfavourably with solvent hydrogen bond acceptors. Therefore the difference between a solute's solvation energy in two solvents is representative of the similarity of the solvents. A new solvent similarity metric has been developed in this work based on the solvation behaviour of a single solute SSIP in the SSIMPLE framework.

Extension of the SSIMPLE approach to the study of vapour-liquid equilibria required developments to the theoretical model. Incorporation of the temperature dependence of interactions into the model, allowed calculations to be done at variable temperatures. Expressions to calculate the total concentrations of species in a phase with expansion and evaporation of the liquid were also developed in the SSIMPLE framework. Modelling of VLE as a function of temperature act as proof of concept.

Chapter 2

Data Curation

Experimental equilibrium constant information for formation of hydrogen bonding complexes has been collated, to allow the reparameterisation of the method used to map calculated electrostatic potentials to the hydrogen bonding parameters. This required the design of appropriate data structures. Storage of the data is only part of data management, with access and usage of the gathered data being equally important.

2.1 Data storage methods

Digital data can be stored in a variety of different formats, each with advantages and drawbacks. Existing databases of experimental and theoretical data use different formats. The formats include comma separated variable (csv) file [63], Structured Query Language (SQL) database [170], eXtensible Markup Language (XML) [171, 172], HDF5 [173] and JavaScript Object Notation (JSON) [174] or bespoke database structures [175, 176]. The formats were chosen for reasons including readability, query time and file size. Query time, the time taken to interrogate the database can vary widely, depending on the type of database, and scales differently with size, so must be considered [177].

For the database created in this work XML was used, with the use of the Chemical Markup Language (CML) [178] extension of XML to represent chemical structures.

2.1.1 Chosen data framework: XML

XML (eXtensible Markup Language) provides an easy to interrogate storage format. It was designed to describe data, with a focus on what the data are. Information is stored as a series of elements. An element is described by a tag, which describes the information contained

within the element. The value of an element can be a number, string or more elements (these are referred to as children). An element can also have any number of attributes associated with it. For example an element called Size could have a value of 10, and an attribute called unit, which is m^2, as in figure 2.1. The example element provides a simple example of an element with an attribute and a text value as a child.

```
<?xml version="1.0" encoding="utf-8"?>
<Size unit="m^2">10</Size>
```

Fig. 2.1: XML example element

XML [179] provides a formal syntax to store information in a machine and human readable format. XPATH [180] can be used to query an XML file, allowing for bulk retrieval of data, such as standard InChIKeys [181] of molecules. Due to its extensible nature, XML provides a very broad framework, with only a few syntactical restrictions on how information can be represented. There is no restriction on the number or type of elements and attributes in the file, so a convention needs to be defined.

This can be done by defining an XML Schema, which is an XML file which extends the ideas of the more abstract Document Type Definition (DTD)[182]. A DTD defines the expected structure of elements and attributes in a document, which can be used to validate the data structure, but the definitions provided are still fairly broad. An XML Schema Definition (XSD) is an extension of a DTD but allows greater specificity about the data stored in elements and attributes, including the type and also format for a data value, as well as namespace. For example, if a standard InChIKey is considered, which has a set format described in [181], an 'inchikey' attribute can be restricted to conform to this format using an XSD, which was not possible with a DTD. Validation tools can then be used to check for conformance against the schema.

XML specifications for chemistry fields have already been developed. CML [178, 183–185] provides a general format for representing 2D and 3D structures, as well as providing an extensible platform for computational and spectroscopic work [186]. An alternate XML based specification for thermodynamic information, ThermoML[187], has also been developed by IUPAC to standardise the data format used for storage of thermodynamic data. There is also an example of using an XML format for calculation output from NWChem [188, 189], a quantum mechanics code.

2.1.2 Alternative frameworks

There are a range of other frameworks that were also considered for storing information, but were not chosen for the reasons given below.

JavaScript Object Notation (JSON) is another format for information which is similar to XML, in general structure, but has much looser definitions on types values can take, and lacks the native rigidity provided by the definition of a schema. JSON can easily be instantiated in memory into a data structure, but traversal to find information requires assumed knowledge of the data structure to be efficient. An implementation of NWChem using this has also been developed [190]. The lack of standardised robust parsing and validation frameworks means a JSON implementation would require greater maintenance. Work is currently being undertaken to add this functionality to JSON, but it is currently still in draft form [191].

CSV files and other text file formats lack the stringent specifications of other formats and require the development of bespoke validation software for any format designed. The development and maintenance of such code would be unfeasible, hence use of existing frameworks would be preferred.

SQL and related relational databases store information as collections of tables, providing a framework for storing and manipulating big data. For output from individual calculations, this would not be suitable, due to the relatively small amount of data generated.

Hierarchical Data Format 5 (HDF5) stores information in binary files in a structured way as a series of tables that is platform independent. This is useful for applications with very large amounts of data being produced, since you can compress this information. Since it is in binary the files are not directly human readable, but visualisation is possible through programs and application program interfaces (APIs) [173], such that data manipulation for large scale systems is feasible. This has been used in astrophysics for improved efficiency in image transport [192]. Current data requirements are much smaller than this, so the HDF5 framework is not suitable.

2.2 Data collection and storage

With XML selected as the data format of choice, a schema was developed to describe how information will be stored once collected.

2.2.1 Database Schema

The full schema is shown in appendix B. Information is stored in the database on a per molecule basis, with the root element containing a list of molecule entries. Within each molecule entry, the information is partitioned into four categories:

- Structure
- Physical Properties
- Calculated Properties
- Calculation Data

Each category represents a child element of the molecule. Within structure elements information about the molecule structure (SMILES, InChIKey, CML) and names are stored. A Physical Properties element contains the experimental data collected, along with information on the experimental conditions for the measurement as well as the source of the information. A Calculated Properties element is designed to store information from calculations, and act as a curated repository for information. The Calculation Data element stores information on the calculation.

2.2.2 Data collected

Experimental data has been found by searching the literature. Data on the association constants of different molecules with a fixed host or guest has been gathered. The data [34, 35, 151–153, 193–221] used in the original parameterisation of the model was verified, and canonical SMILES were created, from which standard InChIKeys could be generated. Further data was gathered from [222–255], using a similar process. Compound names were resolved into structures using OPSIN [256]. Data from the Freesolv database [63] was also compiled, including the references for the primary sources. The information compiled is summarised in appendix D.

2.3 Data visualisation and usage

The data collection was undertaken to be used as a reference for future work, including the reparameterisation of the coarse graining.

2.3.1 Reparameterisation of SSIP footprinting approach

The information of the experimental hydrogen bonding parameters collated has been used in the reparameterisation of the mapping used in the coarse graining approach to assign surface site interaction point (SSIP) values [155]. The results of this process are discussed in 3.3.5.

2.4 Conclusions

An approach for data curation has been developed for the storage and use of experimental information. This has been used in the reparameterisation of the footprinting approach.

Chapter 3

Computational implementation and original code refactor

To be able to explore novel applications of the SSIP approach, a redesign of the workflow was first undertaken. This also allowed the optimisation of some processes by use of better scaling algorithms.

3.1 Footprinting Computational workflow overview

The proposed framework is designed to improve reproducibility and scalability of calculations undertaken in future sections. Generation of the SSIP description, as shown in figure 3.1, can be decomposed into discrete units, promoting the development of a modular code base for the workflow. Generation of a 3D structure for the molecule of interest is followed by calculation of molecular electrostatic potential surface (MEPS) of the molecule. A coarse graining algorithm then converts the MEPS data to the SSIP description. Refactoring afforded an opportunity to define the data formats required at the interfaces between these discrete steps in the workflow, such that a modular construction could be employed. This allows for easier extension, addition, development and replacement of components in the workflow. The utilisation of more computationally efficient algorithms has improved the efficiency and scalability of the footprinting process.

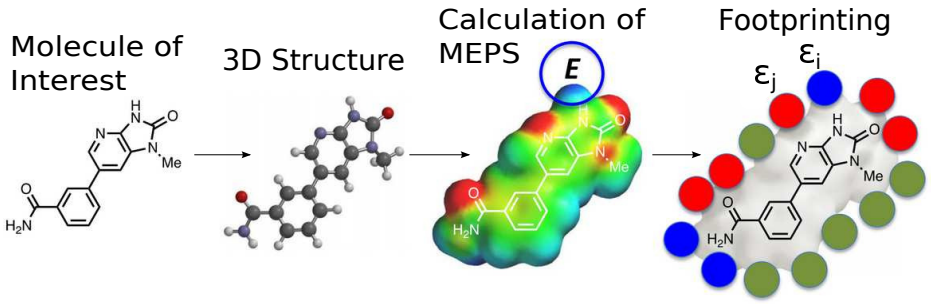


Fig. 3.1: SSIP footprinting. The process starts with a molecule of interest, for which a 3D structure is generated. The MEPS data is then calculated for the molecule, before this is coarse grained to produce a collection of SSIPs which describe the surface interaction sites.

3.2 Data formats

XML is predominately used as the file format for information transfer between different parts of the workflow, due to the benefits discussed in chapter 2. Molecule structures are expressed as CML, and the output of the SSIP footprinting process also expressed as XML. The molecular electrostatic potential surface (MEPS) data are stored in unformatted cube files.

3.2.1 SSIP Schema definition

The full schema is in appendix B. The schema is used to define the output file formats from footprinting calculations. This allows for the modularisation of the workflow, providing the ability to interchange components, such that different ideas can be explored without a full redesign of the structure being required.

The molecule CML is encapsulated along with summary information about the MEPS surfaces used in the calculation. The SSIPs are contained in a sequence of elements.

3.2.2 MEPS data files

To store the MEPS data produced during calculations, the unformatted cube file format is used. This is a raw text file that uses FORTRAN formatted information, and is outputted by QM codes to store property information, based on a specification by the Gaussian package [257].

3.3 SSIP approach to molecule description

The surface site interaction point (SSIP) approach describes intermolecular interactions via a series of discrete sites on the 0.002 e bohr^{-3} electron density isosurface [155] of a molecule. Figure 3.2 shows the modular construction of the computational framework developed to undertake this work. The process uses the aforementioned data format specifications to be used in input/output operations at the interfaces between different computational modules.

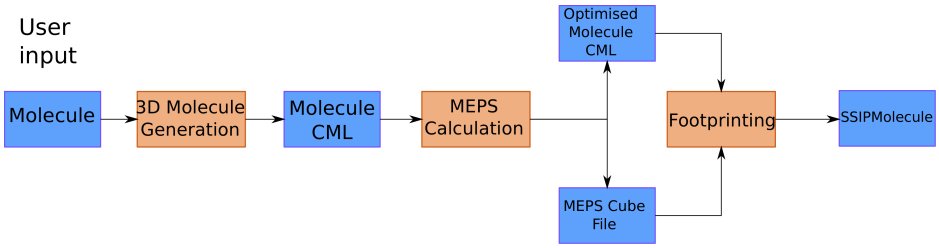


Fig. 3.2: Workflow showing the process of generating a SSIP description for a molecule from input structure to the final XML output. Blue rectangles represent input/output information, Orange rectangles represent computational processes.

3.3.1 Molecule Input

The calculation requires a 3D structure of the molecule written in a Chemical Markup Language (CML) format [178], that has namespace qualified elements and attributes. CML can be generated by conversion of molecule information stored in other 3D structure file formats, using the openbabel software [258]. The Python library cmlgenerator [259] created for this work provides a wrapper for this functionality, to be able to process molecules in batches.

3D structure generation

A SMILES [260] string is a 2D structural representation, which can be outputted from ChemDraw™. Generation of the corresponding 3D structure is required. The RDKit [261] library can perform this operation, but outputs a mol2 formatted structure, which requires further conversion using openbabel. The ETDKG2 [262] approach to conformer generation is used, with a UFF[263] force field to optimise the produced structures. The minimum energy conformer, from a search of 1000 conformers, is then selected for use in future steps.

The Python library, cmlgenerator, created for this work, acts as a wrapper with a command line interface (CLI) to be able to undertake this transformation in one step.

3.3.2 MEPS generation

The Flow Diagram in figure 3.3 summarises the process of MEPS generation. The MEPS is generated on the 0.002 e bohr^{-3} electron density isosurface [155]. NWChem [188] was used for calculations of the MEPS with a DFT method, the B3LYP functional [264–267]. B3LYP was chosen due to the reliable results produced with a low computational cost [268]. The basis set used was 6-31G* [269, 270] for all atoms, except Bromine, Selenium and Iodine, where a 6-311G** [271] basis set was used.

The MEPS calculation is a multi-step process, which was automated and the functionality was packaged into the nwchemcmlutils repository [272]. The Python application programme interface within NWChem [188, 273] was used to improve the efficiency of the workflow.

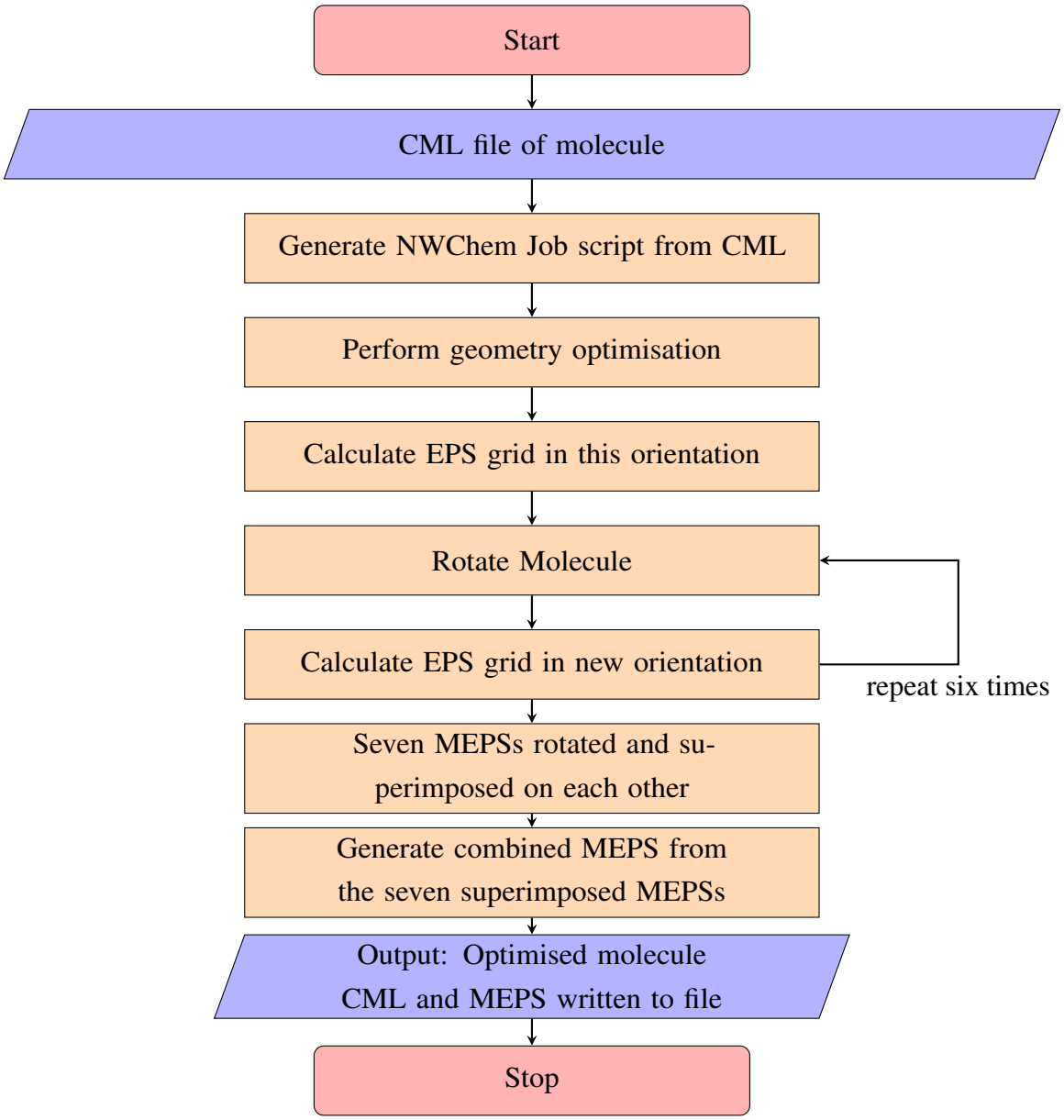


Fig. 3.3: Flow diagram for MEPS generation. Red rounded rectangles are start/stop points, blue trapeziums show input/output operations, orange rectangles are processes.

The MEPS is calculated in seven different orientations such that a coarser grid can be used to reduce computational time, without loss of accuracy, and to reduce the likelihood of discretisation errors. The final surface which is outputted is the combination of the seven MEPS grids, which provides a comparable result to the use of a much finer grid. Figure 3.4 shows the MEPS of water generated from a single orientation, showing the points lie in planes due to the grid. An uneven distribution of points on the surface is heavily reliant on

the initial geometry and the grid properties. Rotation by 45 or 120 degrees about the x, y and z axes is used to generate a further six surfaces. The resultant surface from the superposition of these seven surfaces after reorientation shown in figure 3.5, contains an even distribution of points over the entire surface, with a higher density.

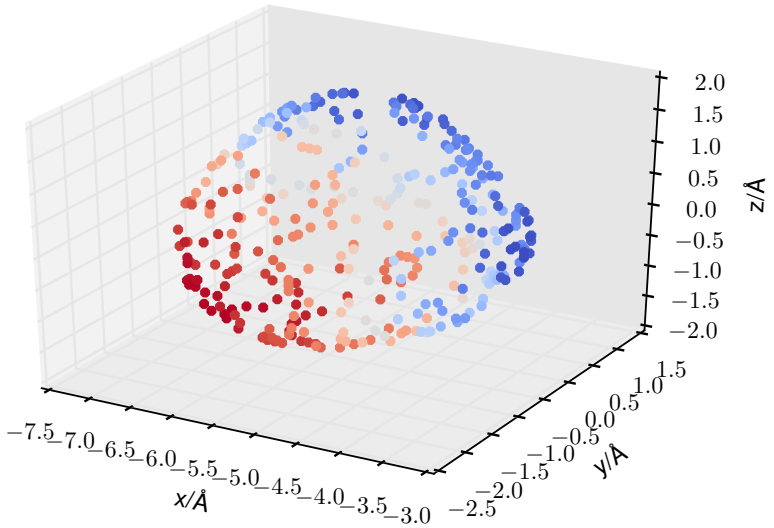


Fig. 3.4: MEPS for water in one orientation (atoms not shown). Calculation used B3LYP with a 6-31G* basis set, a grid padding of 2.0 Å and step size of 0.088 Å, on the 0.002 e bohr^{-3} electron isosurface, with a tolerance of 0.00003 e bohr^{-3} in the electron density. The colour map goes from red (negative potentials) to blue (positive potentials).

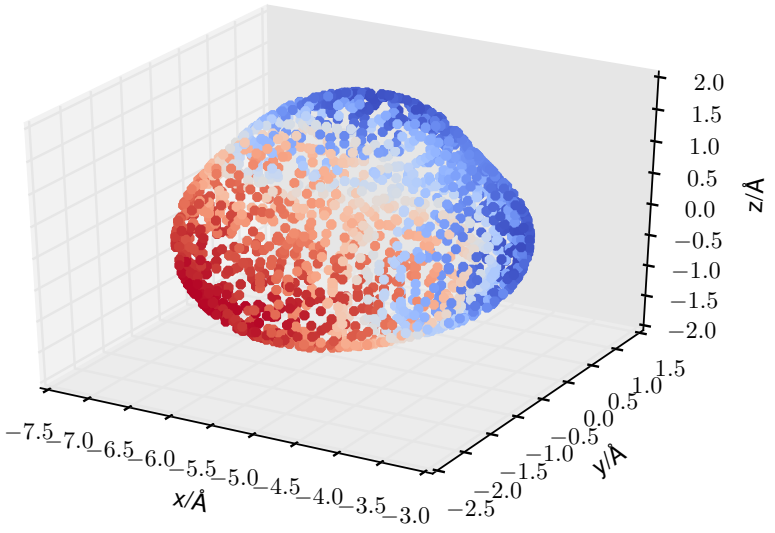


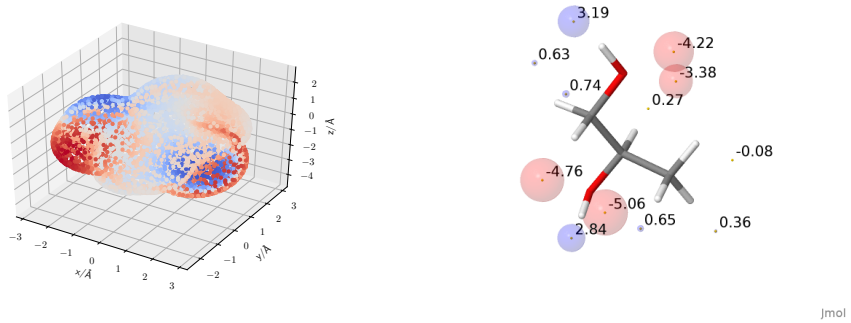
Fig. 3.5: MEPS for water after all seven orientations have been superposed (atoms not shown). Calculation used B3LYP with a 6-31G* basis set, a grid padding of 2.0 Å and step size of 0.088 Å, on the 0.002 e bohr^{-3} electron isosurface, with a tolerance of 0.00003 e bohr^{-3} in the electron density. The colour map goes from red (negative potentials) to blue (positive potentials).

Molecule Volume

The volume enclosed by an isosurface is also calculated during this stage within NWChem. This information is required for the work in chapter 7. The number of voxels with electron densities greater than the specified surface is calculated for each orientation. The mean volume enclosed is then written in the merged cube file that is outputted from the calculation.

3.3.3 Footprinting Algorithm

Once a MEPS for a molecule has been generated, it can then be footprinted, to produce the SSIP description. Figure 3.6 contains example output of the footprinting process for 1,2-propandiol.



(a) 1,2-propandiol MEPS (b) 1,2-propandiol with SSIPs

Fig. 3.6: The conversion of the MEPS (left, with no atoms shown) to the SSIPs (right, with atoms shown) for 1,2-propandiol. A colour map is used for the MEPS, going from blue (for positive MEPS points) to red (negative electrostatic potentials), with the depth of colour representing the magnitude of the MEPS. For the SSIPs blue is used for positive SSIPs and red for negative SSIPs, with size of sphere representing magnitude.

A SSIP represents a single surface segment. The partitioning of a molecule into a set of SSIPs requires knowledge of the number of SSIPs to be assigned to the molecule. This is calculated from the molecular surface area. Water is known to have four hydrogen bond interaction sites. Calero *et al.* [155] used this to define the surface area of a SSIP as a quarter the surface area of water which corresponds to $A_{SSIP} = 9.35\text{Å}^2$.

Hydrogen bond interactions occlude a segment of the surface from other interactions. This places a restriction on the proximity of SSIPs on a molecular surface, to avoid clashing. The minimum separation of SSIPs on the surface, d (in figure 3.7), influences the SSIP description produced. It was parameterised by Calero *et al.* [155], and assigned a value of 3.2Å.

The flow diagram in figure 3.7 summarises the footprinting process, described in [155].

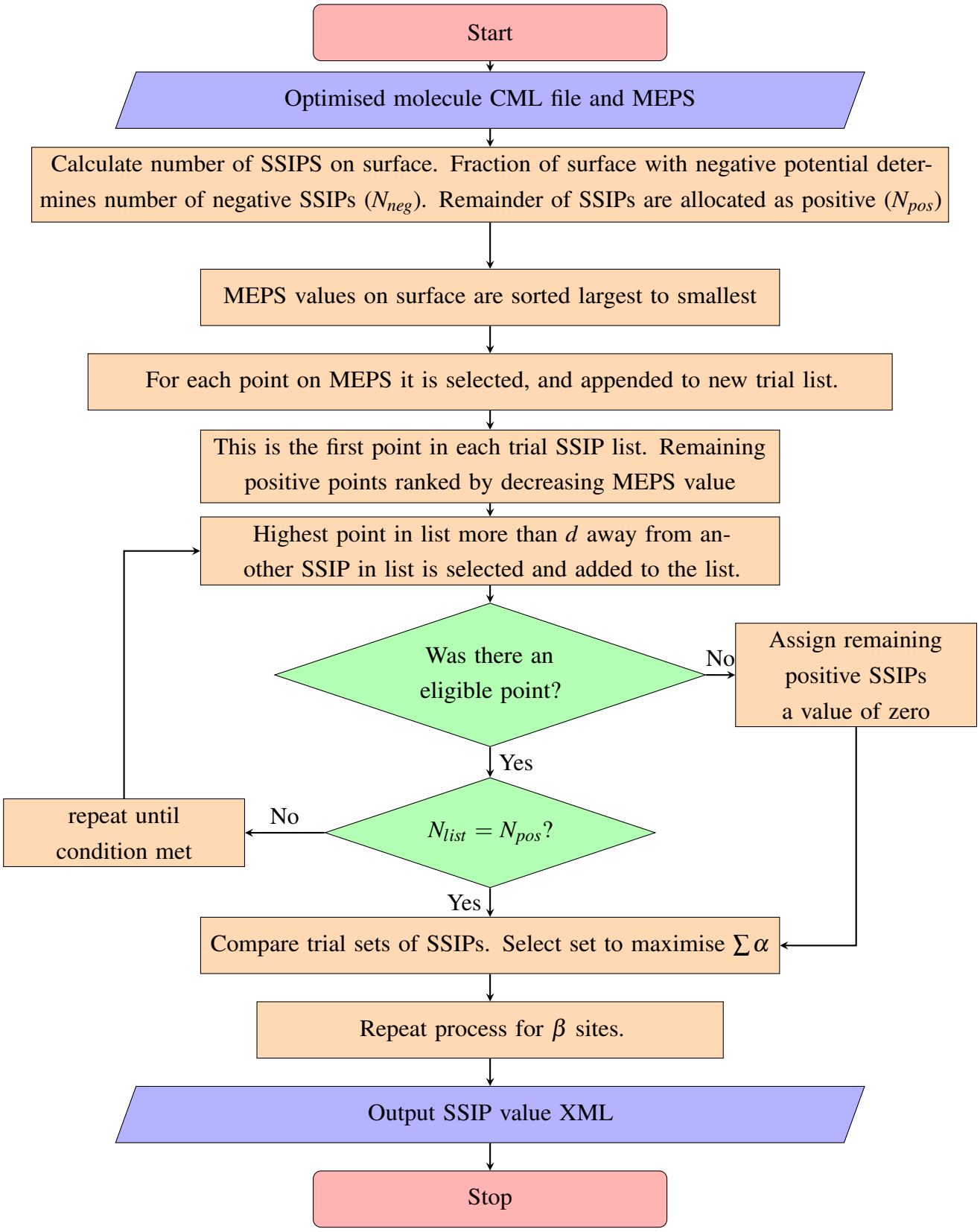


Fig. 3.7: Flow diagram summarising footprinting process for a single molecule. Red rounded rectangles are start/stop points, blue trapeziums show input/output operations, orange rectangles are processes and green diamonds are decision nodes.

Surface Area Calculation

The molecular surface area is calculated to be able to determine the required number of SSIPs to assign to a molecule in the first step of footprinting. This is done using the positional information for the MEPS points, which lie on a convex hull that encapsulates the molecule. The molecular surface area, S.A., is calculated using equation (3.1).

$$S.A. = \sum_{r_i}^{N} D_{r_i} \pi r_D^2 \tag{3.1}$$

Where r_i is the position of the ith point on the surface, D_{r_i} is the density of points in the local environment, given by equation (3.2). The local environment is defined by r_D, the density radius, and $\delta(r_i, r_j)$ (equation (3.2) and equation (3.3)).

$$D_{r_i} = \frac{1}{\sum_j^N \delta(r_i, r_j)} \tag{3.2}$$

$$\delta(i,j) = \begin{cases} 1 \text{ if } |r_i - r_j| \leq r_D \\ 0 \text{ otherwise} \end{cases} \tag{3.3}$$

The density of points in the local environment is simply the reciprocal of the sum of the number of other points within the defined search radius. Each point on the surface contributes a circular area, equal to the area occluded by the density radius, weighted by the density of points in the local environment. A value of $r_D = 0.5$ Å was used.

Efficient searching for nearest neighbours

Calculating the distance between all points on the MEPS is an $\mathcal{O}\left(N^2\right)$ operation, which could become very expensive for large molecules. To reduce the computational cost, a KDTree is used [274, 275], to reduce the scaling to $\mathcal{O}\left(NlogN\right)$ making the search for points satisfying the distance cutoffs faster. The KDTree produces a space-partitioned data structure. A KDTree data structure can be more efficiently traversed than a standard array structure, which was used previously.

The implementation used in the code was written by Teodor Nikolov.

MEPS value mapping to Hydrogen bond parameter

The mapping of molecular electrostatic potential surface (MEPS) value to hydrogen bond parameters uses second order polynomial functions, parameterised in [155]. SSIP values

for positive MEPS segments are mapped to hydrogen bond donor parameters with equation (3.4). The negative MEPS values are mapped to hydrogen bond acceptor parameters with equation (3.5).

$$\varepsilon = 1.12 \times 10^{-5} \psi_{0.002}^2 + 1.14^{-2} \psi_{0.002} \text{if} \psi_{0.002} > 0.0 \tag{3.4}$$

$$\varepsilon = c \left(8.33 \times 10^{-5} \psi_{0.002}^2 - 2.08 \times 10^{-2} \psi_{0.002} \right) \text{if} \psi_{0.002} < 0.0 \tag{3.5}$$

Where $\psi_{0.002}^2$ is the MEPS value on the 0.002 e bohr^{-3} electron density isosurface in kJmol^{-1}, and c is an empirical correction factor based on functional group. The correction factors are contained in table 3.1.

Functional Group	c
Nitrile nitrogen	0.77
Primary amine nitrogen	1.02
Secondary amine nitrogen	1.24
Tertiary amine nitrogen	1.34
Pyridine type nitrogen	1.07
Ether type oxygen	1.16
Alcohol type oxygen	0.86
Aldehyde carbonyl oxygen	0.89
Ester carbonyl oxygen	0.92
Carbonate carbonyl oxygen	0.90
Nitro oxygen	0.77
Any oxygen atom bonded to sulfur	1.00
Any oxygen atom bonded to phosphorus	1.10

Table 3.1 Empirical functional group correction factors for negative surface segments

Functional group identification

As noted in [155] the acceptor sites have a correction factor applied. The requirement of an empirical correction factor to improve the prediction of hydrogen bond acceptors produces an extra requirement of functional group identification within the code.

This is done by expressing the molecule in a graph representation, with atoms as vertices, and bonds as edges. A subgraph isomorphism approach can then be used to identify any occurrences of the functional group, if it is also described as a graph. The algorithm used is the VF2 of Cordella [276], which has been implemented in this code by Teodor Nikolov.

The functional group identification routine employed prevents the assignment of multiple correction factors. This is done by the canonical ordering of functional group subgraphs by use of an enumeration object. Once an assignment to a functional group is made, no others can be assigned. The functional groups are therefore ordered so the largest collection of atoms would be assigned first, i.e. an oxygen atom would be assigned to an ester rather than an ether.

Evaluation of trial SSIP configuration

The positional assignment uses the MEPS value, rather than the ε value of the SSIP. This is done to ensure the correct surface features are found. Once the best configuration is found, the region around each SSIP assigned is then further examined. Since a SSIP interaction involves a small contact area and not just a single point, all MEPS points within this contact area must be inspected. This led to the assignment of a contact radius, r, assigned value of 1.1Å, in work by Calero *et al.* [155]. A SSIP assigned to the surface is replaced by a more polar point that is found within this radius, to more accurately account for the polarity of the region.

Canonicalisation of output

To ensure reproducibility of footprints from the calculation, the output ordering of SSIPs when written to file must be consistent, to allow for unit testing. The SSIPs are ordered by decreasing numerical value. If there are any SSIPs with the same value, the distance of the SSIPs to the molecule (mass unweighted) centroid is then used to ensure consistent ordering.

This is required because the traversal of the MEPS surfaces generating the different trial SSIP collections is not deterministic, so the same solution may be found with different additions to the SSIP description.

3.3.4 Limitations of Footprinting

The process of generating a SSIP description for a molecule has some known issues which are discussed in the following sections.

Through space effects

Intermolecular interaction strengths are influenced by the functional groups adjacent to the interaction site. The influence of atoms in the wider environment and not just the closest atom,

on the MEPS of a molecule, is therefore desirable to be able to describe the through space effects. Overestimation of these through space effects, however can lead to the assignment of incorrect descriptions.

The presence of two adjacent negative regions produce a build up of negative charge on the surface, thus the assigned SSIPs appear much more polar than expected. Conversely the presence of an adjacent positive region and negative region in space can lead to a cancellation on the MEPS, thus making the assigned SSIPs much less polar. For rigid molecules, there is minimal conformational flexibility, so these effects are always present. Figure 3.8 contains the SSIP description for 2,6-dimethoxyphenol, which exhibits both of the features discussed.

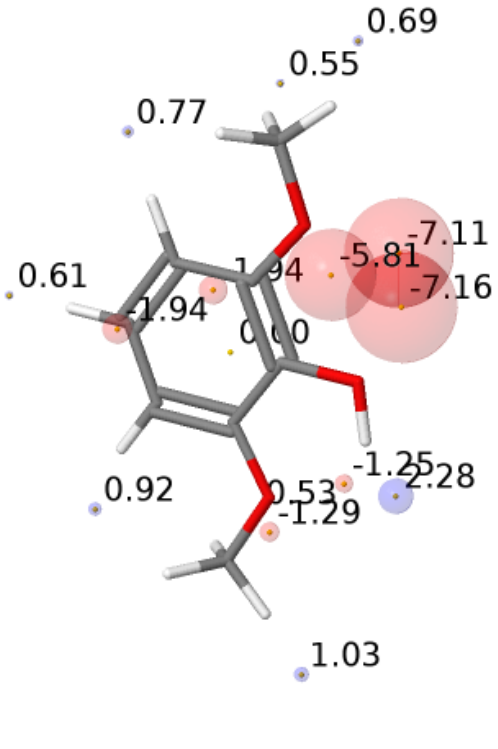


Fig. 3.8: SSIP description of 2,6-dimethoxyphenol. The values of the SSIPs are displayed with a coloured sphere indicating position with the value written nearby. The magnitude of the value is also shown by the size of the translucent sphere, coloured red for acceptors and blue for donors.

The reinforcement shown between the phenol oxygen and the top methoxy oxygen in figure 3.8 leads to the assignment of three SSIPs for the region with values approximately two units higher than expected for the oxygens if they were isolated. This therefore increases the strength of the polar interactions of the molecule made at these sites, which is not realistic.

The interaction between the phenol hydrogen and the lower methoxy group leads to the assignment of a much weaker hydrogen bond donor and two very weak hydrogen bond acceptors. This formation represents an intramolecular hydrogen bond, which would therefore remove SSIPs from interacting with the other molecules in a phase, thus exposing a more non-polar surface as shown here. The presence or lack of intramolecular hydrogen bonds in a description will influence the calculations of interactions with other molecules. Molecules with very polar groups would compete for interactions with the groups involved in the intramolecular hydrogen bond. If these interactions were sufficiently strong this could mean an open system was preferred (without the intramolecular hydrogen bond), which cannot be formed if the system with the intramolecular hydrogen bond is used to represent the molecule.

In 2,6-dimethylphenol four fewer sites were assigned to the surface than required by the total surface area. This meant four extra null SSIPs were assigned, three sites were from positive MEPS regions and one site from negative MEPS regions.

Conformational Effects

The through space interactions present on an MEPS are a function of conformation. Conformational selection processes for non-rigid molecules can therefore influence the description produced for the molecule, leading to variation in the outcome of any calculations using the description.

An example of this is 1,2-propandiol. When the alcohol groups are in a gauche conformation, as shown in figure 3.9, the interactions between the oxygen lone pairs lead to a reinforcement of the negative charge on the surface, such that there are three SSIPs assigned to oxygen lone pairs, instead of 4 with a much larger central SSIP, with a value almost double what is normally expected for an alcohol.

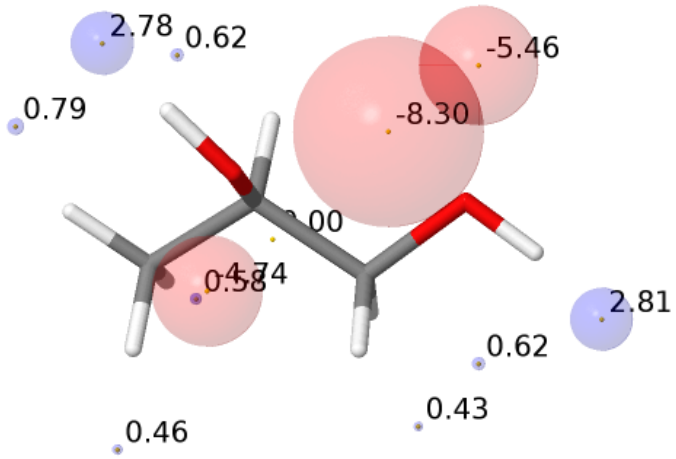


Fig. 3.9: 1,2-propandiol with the alcohol groups gauche, and the O−H hydrogen atoms pointing in opposite directions. The values of the SSIPs are displayed with a coloured sphere indicating position with the value written nearby. The magnitude of the value is also shown by the size of the translucent sphere, coloured red for acceptors and blue for donors.

Maintaining a gauche conformation, but rotating the OH bond of the secondary alcohol, leads to the formation of an intramolecular hydrogen bond. This leads to a reduction in the surface electrostatic potential in the region, as evidenced by the SSIPs shown in figure 3.10 for this conformation. There is now a much better description for the 2-hydroxy acceptor sites. However there is now a very poor donor that is not aligned with the OH σ bond, and one of the acceptor sites is missing on the other alcohol group.

It is also important to note the effect on the positive SSIP of the primary alcohol donor after the formation of the intramolecular hydrogen bond. There is a significant increase in its donor ability, due to polarisation of the oxygen. The polarisation of atoms, leading to a modification to the strength of hydrogen bonding interactions, has been noted experimentally to be significant for alcohols [277].

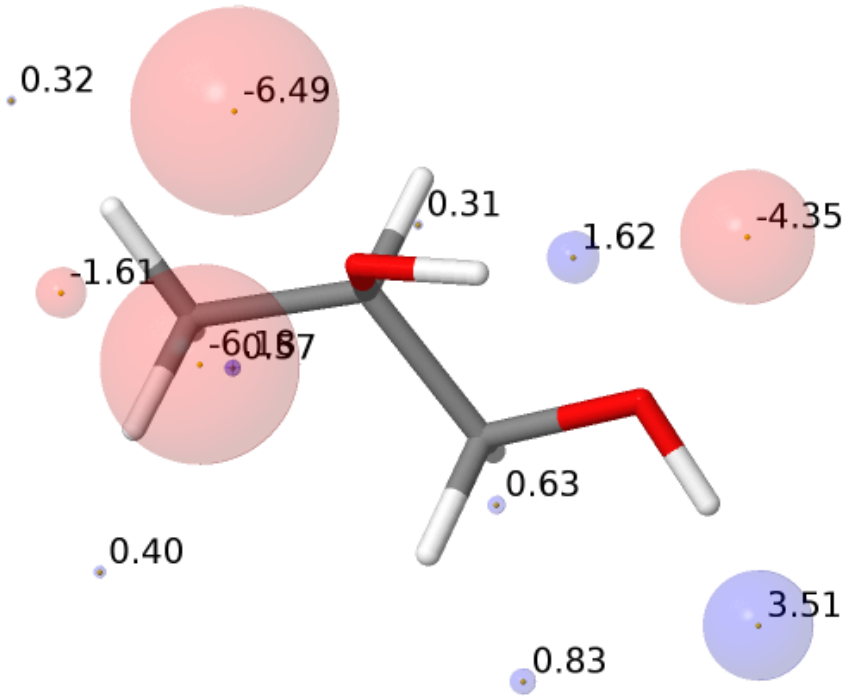


Fig. 3.10: 1,2-propandiol with the alcohol groups gauche, and the O−H hydrogen atoms pointing in the same direction. The values of the SSIPs are displayed with a coloured sphere indicating position with the value written nearby. The magnitude of the value is also shown by the size of the translucent sphere, coloured red for acceptors and blue for donors.

Positioning of the alcohol groups trans, by rotation of the C−C bond, avoids these interactions completely, as shown in figure 3.11. This provides a conformation free of intramolecular interactions.

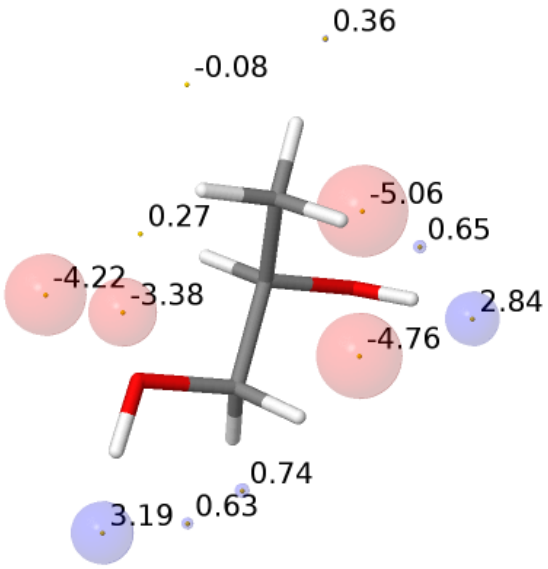


Fig. 3.11: 1,2-propandiol with alcohol groups trans. The values of the SSIPs are displayed with a coloured sphere indicating position with the value written nearby. The magnitude of the value is also shown by the size of the translucent sphere, coloured red for acceptors and blue for donors.

The conformer used in calculation of the SSIP description has a large effect on the resulting description. By using a conformer generation process if a 3D structure is not provided the reproducibility of the SSIP description should be improved. However it may not lead to the generation of the best description for a molecule.

Charged species

Experimental α and β values for a selection of charged species have been gathered [154]. The current approach functions for only neutral species. The inclusion of a charge on a molecule causes a large change in the MEPS compared to a neutral isoelectronic molecule. For anions, the entire surface is much more negative, whereas for cations the surface is more positive. Generating a SSIP description of the molecule is therefore untreatable with the current approach.

3.3.5 Reparameterisation of SSIP footprinting process

To remove the dependence on empirical correction factors based on functional group recognition, a new method has been proposed, using more computational parameters. This work was carried out by Nicola De Mitri, with the results of his investigation summarised here.

In this approach the calculation of hydrogen bond acceptor SSIPs uses information from three different MEPSs for the molecule. The hydrogen bond donor protocol remains the same. This is referred to as the tri-surface approach to footprinting, with the previously published approach (in [155], equation (3.5)), referred to as the mono-surface approach.

The selection of negative surface segments uses the electrostatic potential of an inner electron surface, the 0.0104 a.u. MEPS for the trial SSIP configuration. This is to try to identify lone pair directions.

The second modification is the mapping function of MEPS value to the SSIP ε value. The new function is in (3.6). It has a similar quadratic functional form to the original approach.

$$\varepsilon = \pi\left(\psi_{0.002}, \psi_{0.005}\right) Q\left(\psi_{0.002}\right) \tag{3.6}$$

Where $\psi_{0.002}$ is the MEP value on the 0.002 a.u. surface, $\psi_{0.005}$ is the MEP value on the 0.005 a.u. surface. $Q\left(\psi_{0.002}\right)$ is a quadratic function, shown in (3.7). The correction factor has been replaced with the function, $\pi\left(\psi_{0.002}, \psi_{0.005}\right)$, shown in (3.8).

$$Q\left(\psi_{0.002}\right) = b_a\left(\psi_{0.002} + b_b \psi_{0.002}^2\right) \tag{3.7}$$

$$\pi\left(\psi_{0.002}, \psi_{0.005}\right) = 1.00 + b_d * \left|(1 - b_f) * \left(\psi_{0.005} - \psi_{0.002}\right) + b_f * \psi_{0.002} - b_c\right| \tag{3.8}$$

Where the parameters b_a, b_b, b_c, b_d and b_f are defined in table 3.2 for MEPS data in kJmol^{-1}.

Parameter Symbol	Parameter Value
b_a	-63.9536393
b_b	-4.72726144
b_c	-0.00610595816
b_d	25.8334241
b_f	-0.00908829891

Table 3.2 Parameters for the tri-surface MEPS mapping equations (3.7) and (3.8).

The π function replaces the use of empirical functional group correction factors in the method described in [155]. The inclusion of information about another MEPS surface leads to a greater understanding about the type of acceptor, without the need to add empirical functional group correction factors.

Influence on through space effects: reinforcement

Propan-1,2-diol in the gauche hydroxyl configuration in figure 3.9 showed a large reinforcement effect when using the mono-surface approach. Using the same conformation, but now with the tri-surface methodology, provides a direct comparison for the degree to which this problem still occurs. Figure 3.12 shows a more even SSIP description, with the largest hydrogen bond acceptor SSIP less polar, so the influence of the through space effects were reduced. However still only three acceptors are present in the region between the two hydroxyl oxygens, rather than 4.

The effect of reinforcement of SSIP values due to the occurrence of adjacent negative segments on the values of SSIPs has been reduced using the tri-surface approach.

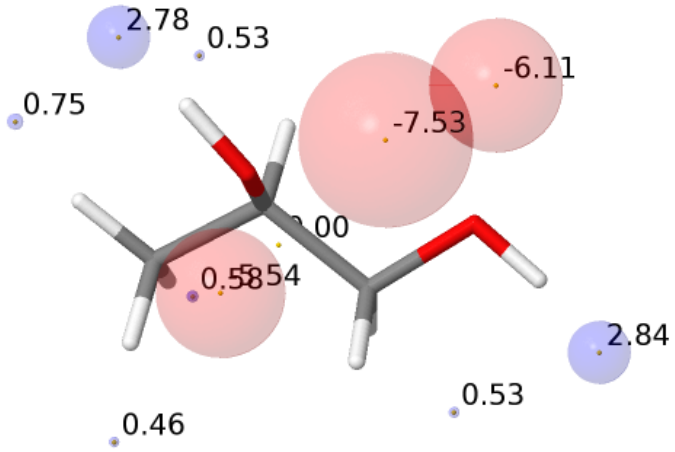


Fig. 3.12: 1,2-propandiol with the alcohol groups gauche, and the O-H hydrogen atoms pointing in opposite directions. Calculation of the SSIPs done using the tri-surface approach. The values of the SSIPs are displayed with a coloured sphere indicating position with the value written nearby. The magnitude of the value is also shown by the size of the translucent sphere, coloured red for acceptors and blue for donors.

This could be due to the distance parameter, d, used to give good separation of SSIPs over the molecular surface, which was optimised for the results from the previous mapping. The use of the inner surface for position assignment is better able to detect lone pairs, so the distance parameter could be refined.

3.4 Conclusion

The refactor described provides the core functionality required for the calculations presented in the following chapters. The current framework follows a simple workflow to generate SSIP descriptions of molecules in a reproducible and highly automated process.

Chapter 4

Phase Transfer Energy Calculation benchmarking

The calculation of phase transfer free energies have been used in pharmaceutical development to aid in drug selection for pharmacodynamic property selection, based on Lipinski's rule of 5 [1]. Multiple methods have been created to predict these energies, with reviews [11–13] categorising existing approaches into three broad classes: empirical functions, implicit solvation simulations or explicit solvation simulations.

Empirical methods consist of parameterised functions that use the correlation of molecular properties to experimental measurement. These functions either use the summation of information for individual fragments in the chemical structure in group contribution type methods [14–28, 30, 31], or quantitative structure property relationships (QSPR) where molecular descriptors are used [34–38, 40–42].

For implicit solvation methods the solvent medium is treated as a dielectric continuum into which the molecule of interest is embedded in a cavity within an electronic structure calculation framework [44]. Parameterised relationships are then used to convert the activity coefficients generated from such calculations to free energies [45–48, 51, 52].

Full atomic simulations are required for explicit solvation models, using molecular dynamics (MD) or Monte Carlo (MC) frameworks to propagate a system in phase space. Calculation of free energies requires summation of the free energy components of interactions between solute and solvent [61–68].

These approaches are all described in more detail in section 1.2.

The Surface Site interaction model for the properties of Liquids at Equilibrium (SSIMPLE) [158] will be used to calculate free energies of transfer and partition coefficients for a series of molecules in this chapter.

4.1 Surface Site Interaction Point (SSIP) approach to interactions

To describe thermodynamic properties of liquids, all intermolecular interactions must be considered. The molecular surface can be described by a collection of surface site interaction points (SSIPs). This representation can be used to describe the properties of the molecule surface, by the assignment of interaction parameters. The parameter assigned to a SSIP, i, is ε_i, which describes polar interactions between surface segments.

The process of value and position assignment for SSIPs is undertaken by footprinting of molecular electrostatic potential surface (MEPS) data. Two footprinting approaches are used in this work. The first is the mono surface approach, using the MEPS at the 0.002 e bohr^{-3} electron density isosurface described in chapter 3.1 (based on work in [155]). The second is the tri-surface approach detailed in 3.3.5, which requires the MEPS at the 0.0020 e bohr^{-3}, 0.0050 e bohr^{-3} and 0.0104 e bohr^{-3} electron density isosurfaces.

4.2 SSIMPLE

In the SSIMPLE approach, the free energy of transfer for a molecule is found from considering the concentrations of free and bound SSIPs. Each molecule in a phase has a set of SSIPs that describe intermolecular interactions, with figure 4.1 showing a cartoon representation of the interactions present at a snapshot in time. The equilibrium population of the free and bound species must be calculated.

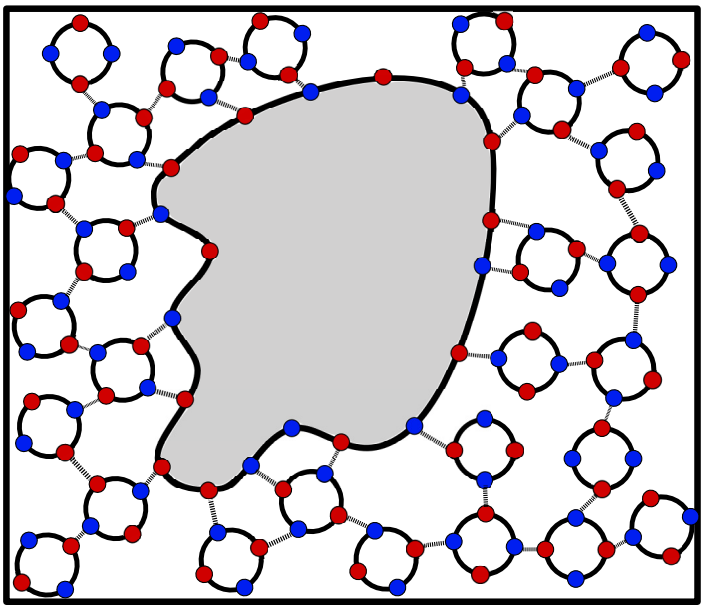


Fig. 4.1: Cartoon representation for a molecule (grey) in solution surrounded by solvent molecules (clear). Solvent and solute interactions shown by dashed lines.

Pairwise interactions are considered between all SSIPs in a phase. The association constant for formation of a bound contact between two SSIPs, i and j, is in equation (4.1) (this is equation 5 in [158]).

$$K_{ij} = \frac{1}{2} e^{-\frac{\varepsilon_i \varepsilon_j + E_{vdW}}{RT}} \tag{4.1}$$

Where $E_{vdW} = -5.6$ kJ mol^{-1}; ε_i, ε_j are the values of the SSIPs; R is the gas constant and T is the temperature. E_{vdW} is the energy of interaction from van der Waals interactions between two SSIPs. From Hunter's study of van der Waals interactions in non polar liquids [278], a contribution of 0.3 kJ mol^{-1} Å^{-2} to the interaction energy was found for noble gases. SSIPs have a fixed surface area of 9.35 Å^2 ([155], discussed in chapter 3). The energy of van der Waals interactions between SSIPs is therefore constant. The polar interactions are directional, whereas the van der Waals interactions are non-directional. If the polar interactions are misaligned, only van der Waals interactions are possible between SSIPs. For repulsive interactions, i.e. $\varepsilon_i \varepsilon_j > 0$, it is assumed that such a state can always be reached, so $\varepsilon_i \varepsilon_j$ is set to zero.

4.2.1 Standard state of a SSIP

K_{ij} is a dimensionless quantity, so depends on the standard state used as a reference, normally 1 M for the solution phase. Definition of a concentration scale, that can be used to treat a system with interactions between solvent and solute SSIPs was required. A standard state where SSIP concentrations are expressed as the fraction of total volume occupied by the SSIPs was defined [158].

The maximum density of SSIPs in a phase, c_{max}, is a function of packing density, and the volume of a SSIP. c_{max} can be defined by equation (4.2), assuming the packing density corresponds to that of the hypothetical zero point solid of cylindrical particles (90%) [278].

$$c_{max} = \frac{0.9}{N_a V_{SSIP}} \tag{4.2}$$

Where V_{SSIP} is the volume of a SSIP, and N_a is Avogadro's constant. The volume of a SSIP was estimated using the volume of water enclosed within the 0.002 e bohr^{-3} isodensity surface. V_{SSIP} was defined to be 5 Å^3 [155], which is the zero point void volume. Under these conditions c_{max} corresponds to a value of 300 M. Future discussions use concentrations normalised by c_{max} unless otherwise stated.

4.2.2 Phase speciation

From consideration of the ensemble of SSIPs the free and bound concentrations are calculated. The concentration of the bound species formed between i and j, $[ij]$ is given by (4.3).

$$[ij] = K_{ij} \left[i_{free}\right] \left[j_{free}\right] \tag{4.3}$$

$\left[i_{free}\right]$, $\left[j_{free}\right]$ are the unbound concentrations of the ith and jth SSIP respectively. The total concentration of the ith species is given by (4.4), which is the sum of the free concentration and all bound species.

$$[i] = \left[i_{free}\right] + \sum_{j=1}^{N} ([ij] + [ji]) \tag{4.4}$$

Phase speciation calculation

The phase speciation calculation is done using a modified form of the concentration of generalised species (COGS) algorithm [279, 280], which is detailed below.

The total concentration of the ith SSIP, in (4.4), is equal to the concentration of the molecular species in the phase. Since the total concentration is known at the start of the

speciation calculation, it provides a condition that the free concentration must satisfy. The free concentration can then be found numerically by an iterative process. An initial guess that $\left[i_{free}\right] = [i]$ is used to calculate the concentrations of all of the $i \cdot j$ species and then (4.5) is used to iteratively converge on the free concentration, where n is the cycle number. This gives a smooth convergence on the phase speciation, which is reached when the concentration difference between iterations is less than a tolerance value. For phase speciation calculations a tolerance of 10^{-10} in SSIP normalised concentration units was used.

$$\left[i_{free}\right]_{n+1} = \frac{\left[i_{free}\right]_n}{\sqrt{\frac{[i]_n}{[i]}}} \tag{4.5}$$

This process allows the treatment of multi-component phases, where a solute and solvent are present. The solvent is also not restricted to a single species, but can be a complex mixture, as long as the total concentrations of each molecular component are defined *a priori*.

4.3 Solvation energies

Solvation of a SSIP considers the change in energy of transfer from the gas phase to the solution phase with the solvent. The gas phase state of the SSIP used considered a free state where there are no interactions. This means the binding free energy of solvation for a SSIP is given by equation (4.6).

$$\Delta G_{b,i} = RTln\left(\frac{\left[i_{free}\right]}{[i]}\right) \tag{4.6}$$

Where $[i]$ is the total concentration of the SSIP, and $\left[i_{free}\right]$ is the free concentration.

Confinement of SSIPs to a condensed phase results in the binding energy overestimating the probability of interaction. This confinement energy can be calculated by using $K_{ij} = 1$ for all SSIP interactions [158]. The concentrations of SSIPs in such a phase are therefore given by equation (4.7).

$$[i] = \left[i_{free}\right] + 2\left[i_{free}\right]^2 \tag{4.7}$$

Rearrangement of equation (4.7), yields the probability of a SSIP being free, P_f, if there is no interaction energy between any SSIPs in the phase, in equation (4.8).

$$P_f = \frac{\left[i_{free}\right]}{[i]} = \frac{\sqrt{1+8\theta}-1}{4\theta} \tag{4.8}$$

Where θ is the fractional occupancy of the phase, equal to $[i]$.

The confinement energy, given in equation (4.9), is the free energy associated with restricting a SSIP to a condensed phase.

$$\Delta G_{c,i} = -RTln\left(P_f\right) = -RTln\left(\frac{\sqrt{1+8\theta}-1}{4\theta}\right) \tag{4.9}$$

Note that ΔG_c for a SSIP, i, is a constant for the phase, as it only depends on the fractional occupancy of the phase and the temperature. For a single SSIP, i, the solvation free energy is therefore the sum of the binding and confinement energies, given in (4.10).

$$\Delta G_{S,i} = \Delta G_{b,i} + \Delta G_{c,i} \tag{4.10}$$

4.3.1 Molecular Solvation energy

A molecule is described by a collection of SSIPs, so the solvation energy of the molecule is a function of the solvation energies of all SSIPs in the molecule. Since the interactions of each SSIP in a molecule are assumed to be completely independent of all others, then the total solvation energy of a molecule is the sum of the SSIP solvation energies for all SSIPs in the molecule. This is shown in (4.11), where N is the total number of SSIPs in the molecule.

$$\Delta G_S = \sum_{i=1}^{N} \Delta G_{S,i} \tag{4.11}$$

4.3.2 Phase transfer free energies

Calculation of the free energy of transfer between two solvents is simply the difference between the solvation free energies of the molecule in the two solvents, in (4.12).

$$\Delta G_{1/2} = \Delta G_{S,2} - \Delta G_{S,1} \tag{4.12}$$

Where the transfer is from solvent 1 to solvent 2.

The free energies calculated using the SSIMPLE approach correspond to the free energy for a molefraction standard state. Thus to get the free energy change for a 1 M standard state the concentrations of the solvents must be included as shown in (4.13).

$$\Delta G^o_{1/2} = \Delta G_{1/2} - RTln\left(\frac{[2]}{[1]}\right) \tag{4.13}$$

Where $[1]$ and $[2]$ are the concentrations of solvent 1 and 2 respectively.

Partition Coefficients

The partition coefficient is defined in (4.14) for the partition between solvents 1 and 2 of a molecule M.

$$\log P_{1/2} = \log\left(\frac{[M]_1}{[M]_2}\right) = -\frac{\Delta G^o_{1/2}\log(e)}{RT} \tag{4.14}$$

Where $[M]_1$ and $[M]_2$ are the concentrations relative to the 1 M standard state of the molecule in solvent 1 and 2 respectively.

4.4 Calculation of free energy of transfer and partition coefficients

Four different data sets of free energy or partition coefficient information have been used to evaluate the performance of the model. Two different approaches to footprint the molecules to obtain SSIP descriptions were used. The mono-surface footprinting approach described in 3.3.3 and the tri-surface described in 3.3.5.

The first test set are molecules from the original SSIMPLE paper [158]. The second set are molecules from a private communication from Reynolds [281]. The third set are molecules from the set of chemically diverse molecules from [282]. The fourth set are the eleven molecules that form the SAMPL6 challenge, [283].

Solvent Descriptions

The solvent SSIP descriptions used are the same for all data sets and footprinting approaches, for consistency. The full descriptions used are in appendix E.1.

Solute Description

The solute SSIP descriptions were generated using the footprinting processes in chapter 3. Both footprinting approaches, the mono-surface approach from [155] (detailed in section 3.3.3) and the new tri-surface method detailed in section 3.3.5 were used to generate descriptions.

A concentration of 0.001 M was used for all solutes for the SSIMPLE calculation.

4.4.1 Dataset 1- SSIMPLE paper molecules

Calculation of the free energies of transfer for the molecules in the dataset from the SSIMPLE paper [158] has been undertaken. In that work, a manually assigned SSIP description for the molecules was used. The results from using the mono-surface description are shown in figure 4.2. The results from use of the tri-surface footprinting approach are shown in figure 4.3. The tabulated values of the plotted data are in appendix E.

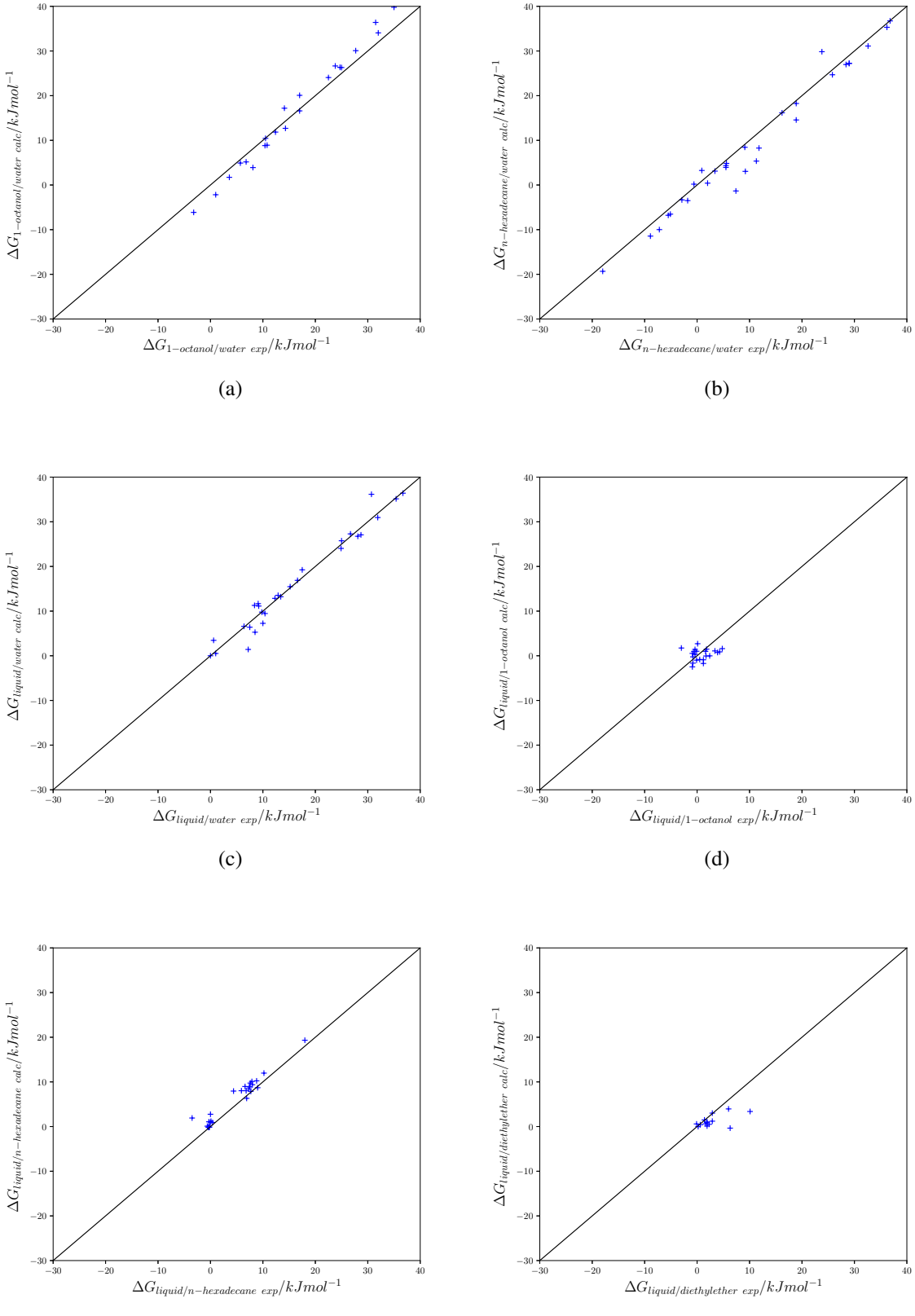
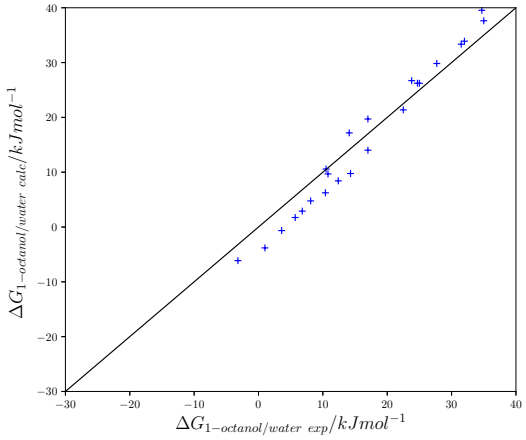
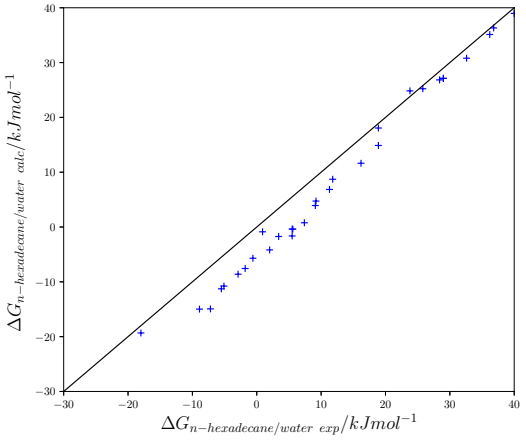


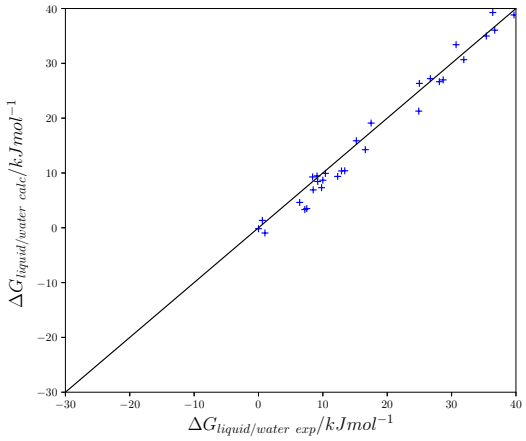
Fig. 4.2: Calculated against experimental free energies of transfers as blue crosses using the mono-surface footprint. y=x plotted as a black line for reference on all plots.



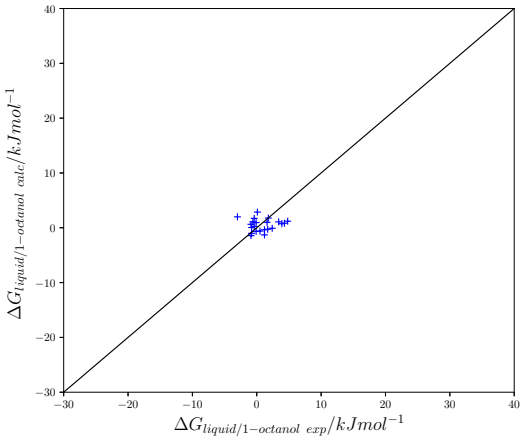
(a)



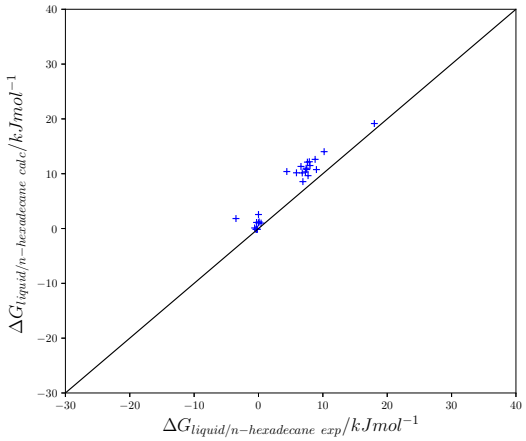
(b)



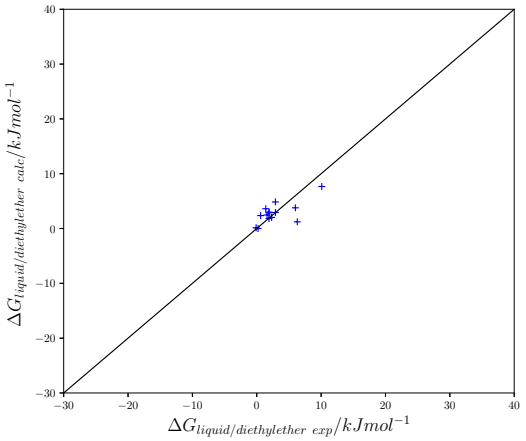
(c)



(d)



(e)



(f)

Fig. 4.3: Calculated against experimental free energies of transfers as blue crosses using the tri-surface footprint. y=x plotted as a black line for reference on all plots.

Performance of the different footprinting approaches on free energy prediction are summarised in table 4.1.

Solvent System	Ref [158]	Mono surface	Tri-surface
1-octanol/water	2.7	2.8	3.1
n-hexadecane/water	2.7	2.9	4.4
pure liquid/water	1.5	2.2	2.2
pure liquid/1-octanol	2.4	2.2	2.2
pure liquid/n-hexadecane	1.2	1.8	3.1
pure liquid/diethyl ether	2.6	2.8	2.0

Table 4.1 RMSE (kJ mol^{-1}) of the transfer free energy predictions for the different footprinting methods

The RMSE values in table 4.1 show the errors with each footprinting approach are similar. The manually assigned footprints provide the lowest error for four transfer energies; the mono-surface and the tri-surface assigned footprints perform best for one data set each. Performance is generally marginally worse for the tri-surface footprint than the mono-surface footprint. This could possibly be attributed to the slightly more negative hydrogen bond acceptors that are generally assigned to a molecule (see appendix F for examples).

4.4.2 Dataset two- Reynolds communication

A substantial collection of different transfer data collated by Reynolds [281]. Figure 4.4 contains the results from the mono-surface footprints and figure 4.5 contains the results from the tri-surface footprints. The RMSE data for the two approaches for the different transfer energies are in table 4.2. The mean RMSE of the free energy of transfer datasets is 0.37 kJ mol^{-1} better for the mono-surface description (3.48 kJ mol^{-1}) than the tri-surface description (3.85 kJ mol^{-1}).

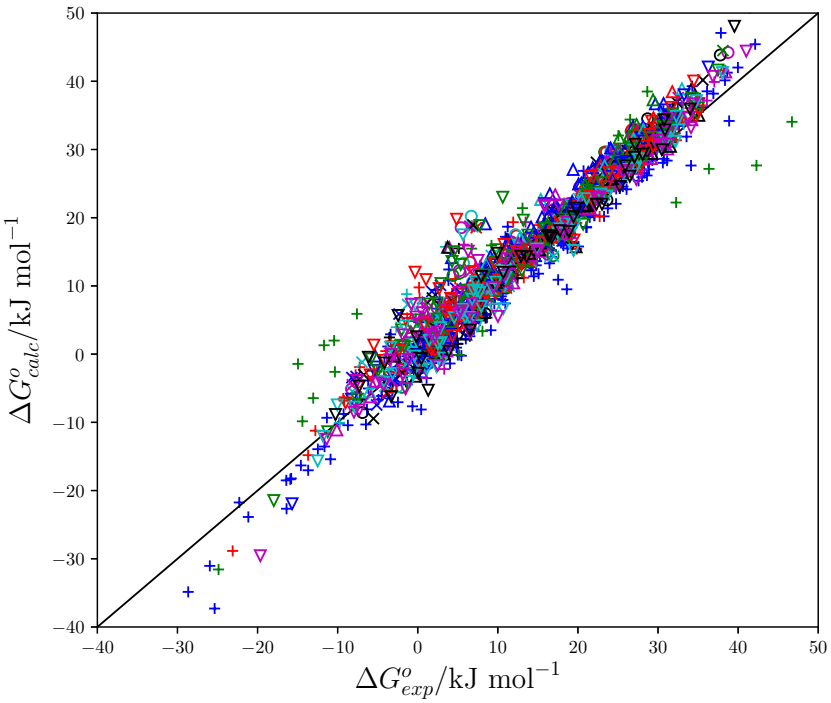


Fig. 4.4: Calculated free energies of transfer using the mono-surface footprint against experimental free energies of transfer for Reynolds' dataset[281]: $\Delta G_{n-hexadecane/water}$ (blue plus), $\Delta G_{n-perfluorohexane/water}$ (green plus), $\Delta G_{carbontetrachloride/water}$ (red plus), $\Delta G_{acetone/water}$ (cyan plus), $\Delta G_{tetrahydrofuran/water}$ (magenta plus), $\Delta G_{diethylether/water}$ (black plus), $\Delta G_{3-methyl-1-butanol/water}$ (blue circle), $\Delta G_{2-methyl-2-propanol/water}$ (green circle), $\Delta G_{2-methyl-1-propanol/water}$ (red circle), $\Delta G_{2-butanol/water}$ (cyan circle), $\Delta G_{2-propanol/water}$ (magenta circle), $\Delta G_{1-decanol/water}$ (black circle), $\Delta G_{1-octanol/water}$ (blue cross), $\Delta G_{1-hexanol/water}$ (green cross), $\Delta G_{1-pentanol/water}$ (red cross), $\Delta G_{1-butanol/water}$ (cyan cross), $\Delta G_{1-propanol/water}$ (magenta cross), $\Delta G_{ethanol/water}$ (black cross) $\Delta G_{methanol/water}$ (blue upward triangle) $\Delta G_{acetonitrile/water}$ (green upward triangle) $\Delta G_{propionitrile/water}$ (red upward triangle) $\Delta G_{n-butyronitrile/water}$ (cyan upward triangle) $\Delta G_{2-butanone/water}$ (magenta upward triangle) $\Delta G_{cyclohexanone/water}$ (black upward triangle) $\Delta G_{dichloromethane/water}$ (blue downward triangle) $\Delta G_{chloroform/water}$ (green downward triangle) $\Delta G_{1,2-dichloroethane/water}$ (red downward triangle) $\Delta G_{chlorobenzene/water}$ (cyan downward triangle) $\Delta G_{toluene/water}$ (magenta downward triangle) $\Delta G_{dibutylether/water}$ (black downward triangle) y=x plotted as a black line for reference.

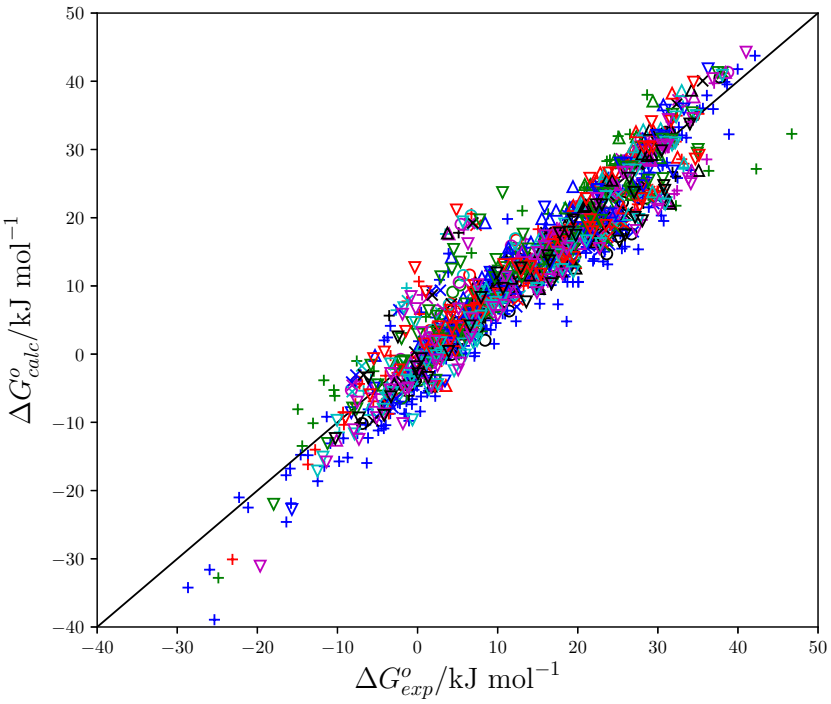


Fig. 4.5: Calculated free energies of transfer using the tri-surface footprint against experimental free energies of transfer for Reynolds' dataset[281]: $\Delta G_{n-hexadecane/water}$ (blue plus), $\Delta G_{n-perfluorohexane/water}$ (green plus), $\Delta G_{carbontetrachloride/water}$ (red plus), $\Delta G_{acetone/water}$ (cyan plus), $\Delta G_{tetrahydrofuran/water}$ (magenta plus), $\Delta G_{diethylether/water}$ (black plus), $\Delta G_{3-methyl-1-butanol/water}$ (blue circle), $\Delta G_{2-methyl-2-propanol/water}$ (green circle), $\Delta G_{2-methyl-1-propanol/water}$ (red circle), $\Delta G_{2-butanol/water}$ (cyan circle), $\Delta G_{2-propanol/water}$ (magenta circle), $\Delta G_{1-decanol/water}$ (black circle), $\Delta G_{1-octanol/water}$ (blue cross), $\Delta G_{1-hexanol/water}$ (green cross), $\Delta G_{1-pentanol/water}$ (red cross), $\Delta G_{1-butanol/water}$ (cyan cross), $\Delta G_{1-propanol/water}$ (magenta cross), $\Delta G_{ethanol/water}$ (black cross) $\Delta G_{methanol/water}$ (blue upward triangle) $\Delta G_{acetonitrile/water}$ (green upward triangle) $\Delta G_{propionitrile/water}$ (red upward triangle) $\Delta G_{n-butyronitrile/water}$ (cyan upward triangle) $\Delta G_{2-butanone/water}$ (magenta upward triangle) $\Delta G_{cyclohexanone/water}$ (black upward triangle) $\Delta G_{dichloromethane/water}$ (blue downward triangle) $\Delta G_{chloroform/water}$ (green downward triangle) $\Delta G_{1,2-dichloroethane/water}$ (red downward triangle) $\Delta G_{chlorobenzene/water}$ (cyan downward triangle) $\Delta G_{toluene/water}$ (magenta downward triangle) $\Delta G_{dibutylether/water}$ (black downward triangle) y=x plotted as a black line for reference.

Solvent System	Mono surface footprint	Tri-surface footprint
$\Delta G_{n-hexadecane/water}$	3.37	5.17
$\Delta G_{n-perfluorohexane/water}$	7.94	6.57
$\Delta G_{carbontetrachloride/water}$	3.29	3.44
$\Delta G_{acetone/water}$	3.00	3.86
$\Delta G_{tetrahydrofuran/water}$	2.78	4.28
$\Delta G_{diethylether/water}$	4.40	4.97
$\Delta G_{3-methyl-1-butanol/water}$	2.61	3.40
$\Delta G_{2-methyl-2-propanol/water}$	2.92	2.78
$\Delta G_{2-methyl-1-propanol/water}$	2.99	2.78
$\Delta G_{2-butanol/water}$	4.05	3.82
$\Delta G_{2-propanol/water}$	3.78	2.96
$\Delta G_{1-decanol/water}$	3.72	4.18
$\Delta G_{1-octanol/water}$	2.92	3.92
$\Delta G_{1-hexanol/water}$	3.39	3.81
$\Delta G_{1-pentanol/water}$	2.54	3.58
$\Delta G_{1-butanol/water}$	3.02	3.28
$\Delta G_{1-propanol/water}$	3.24	3.20
$\Delta G_{ethanol/water}$	3.33	3.27
$\Delta G_{methanol/water}$	4.26	3.20
$\Delta G_{acetonitrile/water}$	3.72	2.99
$\Delta G_{propionitrile/water}$	3.23	3.75
$\Delta G_{n-butyronitrile/water}$	3.24	3.17
$\Delta G_{2-butanone/water}$	2.86	3.68
$\Delta G_{cyclohexanone/water}$	2.89	4.63
$\Delta G_{dichloromethane/water}$	3.02	3.35
$\Delta G_{chloroform/water}$	3.83	3.90
$\Delta G_{1,2-dichloroethane/water}$	3.97	4.15
$\Delta G_{chlorobenzene/water}$	2.75	4.53
$\Delta G_{toluene/water}$	3.34	4.96
$\Delta G_{dibutylether/water}$	4.00	4.05
Mean	3.48	3.85

Table 4.2 RMSE (kJ mol^{-1}) of the transfer free energy predictions for the different footprinting methods for Reynolds communication[281].

4.4.3 Dataset three: Martel dataset

The set of chemically diverse molecules collated by Martel and coworkers [282] is the third dataset. Wet 1-octanol was used in these calculations (0.271 molefraction water content in a 1-octanol solution).

Figures 4.6 and 4.7 show the results for the two different approaches, with the tabulated values in appendix E.

Comparison of the errors for these methods shows that the tri-surface approach is significantly better than the mono-surface approach, by 0.66 log units, however this is for a reduced dataset. Both footprinting methods have greater errors than the majority of methods benchmarked in [11], where the RMSEs of the best methods were under 0.5 log units.

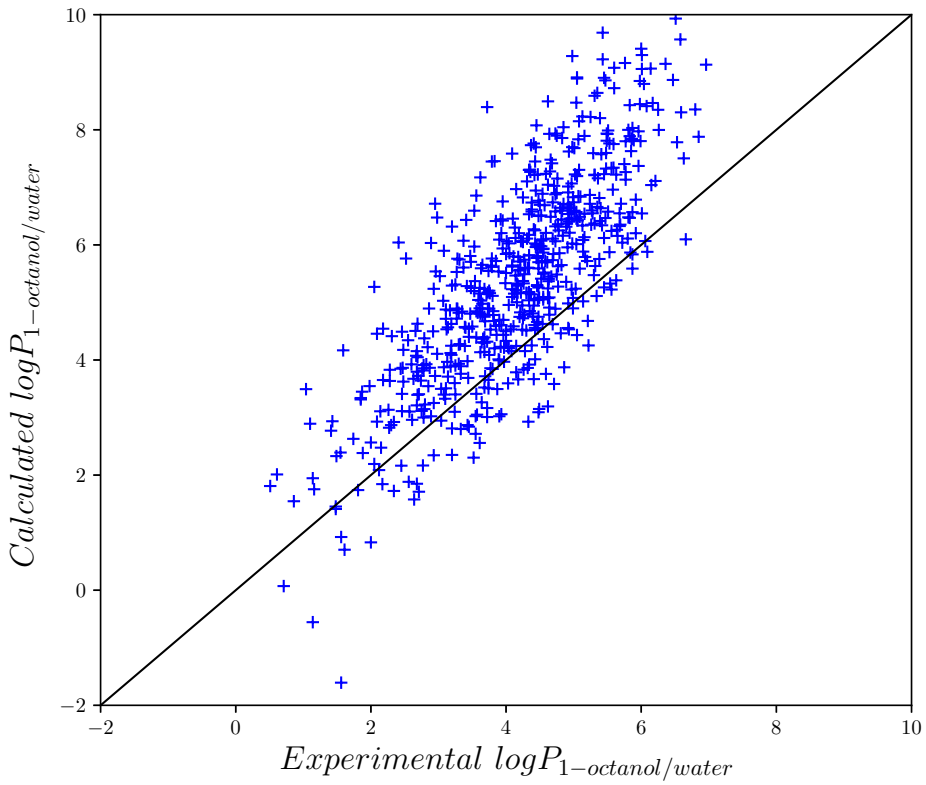


Fig. 4.6: Log $P_{1-octanol/water}$ calculated with mono-surface footprint against experimental log $P_{1-octanol/water}$ (blue crosses), y=x (black line) shown for reference. RMSE of 1.70.

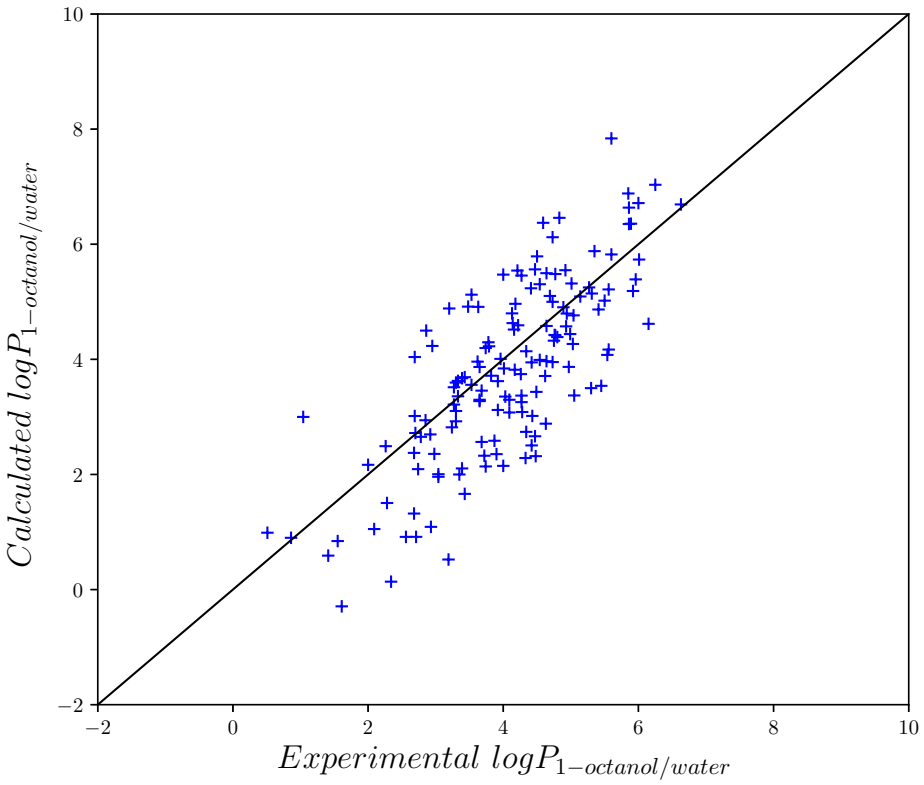


Fig. 4.7: Log $P_{1-octanol/water}$ calculated with tri-surface footprint against experimental log $P_{1-octanol/water}$ (blue crosses), y=x (black line) shown for reference. RMSE of 1.04.

4.4.4 Dataset 4- SAMPL6 molecules

The SAMPL6 logP challenge contained 11 complex molecules [283]. Wet 1-octanol was used in these calculations (0.271 molefraction water content in a 1-octanol solution), to represent the experimental conditions. Figures 4.8 and 4.9 show the results for the two different approaches, with the tabulated values in appendix E.

The tri-surface approach has performed significantly better on this set of molecules than the mono-surface approach, with an RMSE nearly half that of the mono-surface approach.

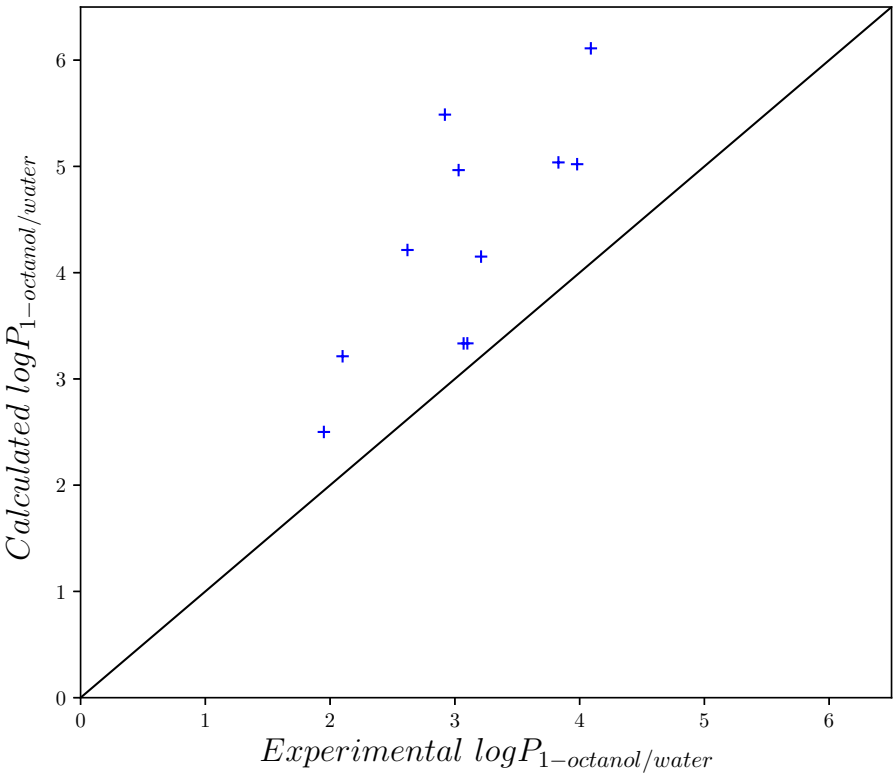


Fig. 4.8: Log $P_{1-octanol/water}$ calculated with mono-surface footprint against experimental log $P_{1-octanol/water}$ (blue crosses), y=x (black line) shown for reference. RMSE of 1.42.

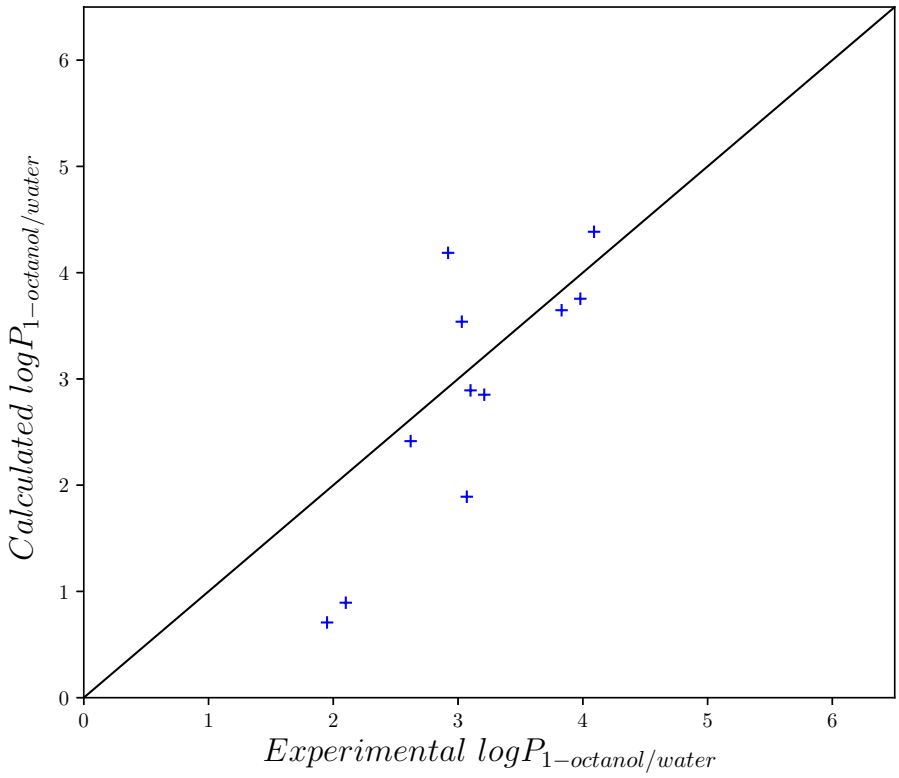


Fig. 4.9: Log $P_{1-octanol/water}$ calculated with tri-surface footprint against experimental log $P_{1-octanol/water}$ (blue crosses), y=x (black line) shown for reference. RMSE of 0.78.

Figure 4.10 shows the mono-surface and tri-surface SSIP descriptions for two of the molecules in the SAMPL6 dataset. The hydrogen bond acceptor sites for both molecules show significant differences between the two approaches. The tri-surface SSIP description assigned for the hydrogen bond acceptor surface, is more negative than the mono-surface descriptions assigned for the hydrogen bond acceptor surface. This is also seen for the other molecules in the SAMPL6 set. This reduces impact of through space effects on the description for the occluded aromatic nitrogen atoms.

This more polar description leads to stronger interactions of the solute with solvent hydrogen bond donor sites, such that there are more favourable interactions with water, for these large aromatic molecules.

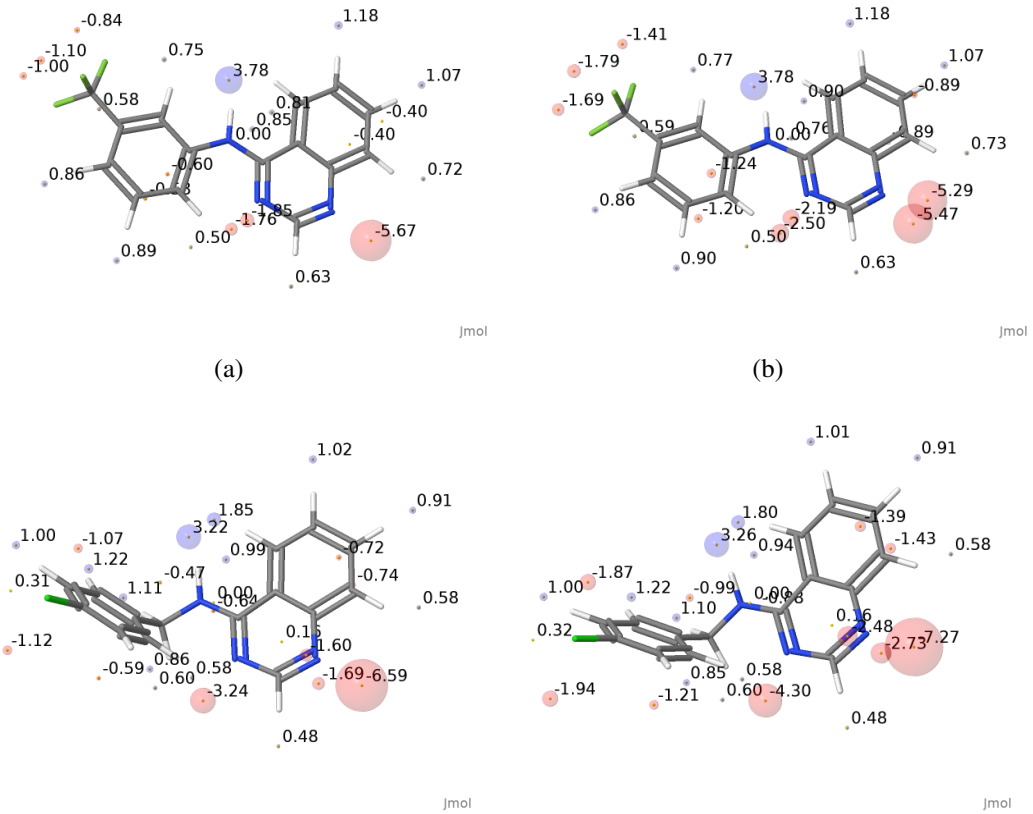


Fig. 4.10: SAMPL6 molecule SM02 SSIP description with mono-surface approach (a) and tri-surface approach (b). SAMPL6 molecule SM04 SSIP description with mono-surface approach (c) and tri-surface approach (d). The molecules are diplayed by a liquorice representation of aatoms and bonds. SSIP value is written in text near to yellow sphere indicating position. Translucent spheres for hydrogen bond donors (blue) and hydrogen bond acceptors (red) indicate the SSIP magnitude.

4.5 Conclusions

The performance of the SSIMPLE approach, using two different approaches to footprinting has been benchmarked on a series of datasets. The performance of the two footprinting approaches are similar for all data sets presented in this work, with the tri-surface approach giving lower RMSEs on average over all the datasets. The mono-surface approach provided better predictions for the molecules that compose the SSIMPLE [158] and the Reynolds

[281] datasets, whereas the Martel [282] and SAMPL6 [283] datasets received more accurate energy predictions using the tri-surface description.

Chapter 5

Functional Group Interaction Profiles

Understanding of intermolecular interactions can be divided into two parts, classification and quantification. The classification of intermolecular interactions is based on the functional groups involved, such as CH−O bonds, halogen bonds, aromatic interactions, cation-π interactions, hydrophobic interactions [112, 145–148]. The quantification of intermolecular interactions requires measurement or calculation of the energy change upon association of two interacting functional groups.

The association of two solutes in solution can be described by the equilibrium is shown in figure 5.1. On the left, a solute donor is interacting with a solvent acceptor, and a solute acceptor with a solute donor. On the right the solute donor and acceptor are now interacting, and the solvent donor and acceptor are also interacting. The position of this equilibrium is dependent on the relative strength of these four interactions.

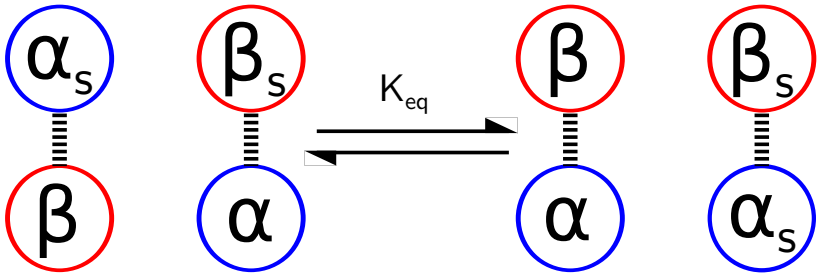


Fig. 5.1: Association of two solutes, with equilibrium constant K_{eq}. In the free state the solute hydrogen bond acceptor (red) and a solute hydrogen bond donor (blue) are initially interacting with a solvent hydrogen bond acceptor (red) and solvent hydrogen bond donor (blue) respectively.

Experimental quantification of the strength of the observed interactions requires the measurement of this equilibrium constant. The development of predictive scales for the strength

of intermolecular interactions uses large collections of experimental data to parameterise such scales [145, 150, 151, 195].

Hunter [149] used the experimental data on intermolecular association constants to formulate the predictive equation (5.1).

$$\Delta G^o = -RT\,log\,(K) = -(\alpha - \alpha_s)(\beta - \beta_s) + \gamma \tag{5.1}$$

Where ΔG^o is the free energy change for formation of the 1:1 complex between two solutes, R is the gas constant, T is the temperature, α, β are the solute hydrogen bond donor and acceptor values, α_s, β_s are the solvent hydrogen bond donor and acceptor parameters respectively. γ is a constant with a value of 6 kJ mol^{-1}.

A Functional Group Interaction Profile (FGIP) for a solvent is a plot of the free energy change of interaction for all possible solute combinations in a solvent.

5.1 $\Delta\Delta G_{H\ bond}$

The functional group interaction profile (FGIP) is designed to give insight into the effects on binding between solutes as a function of solute hydrogen bond acceptor and donor strength.

By consideration of the relationship in equation (5.1), the change in hydrogen bond interaction energy, is the energy difference between the four hydrogen bonding interactions in figure 5.1 given by equation (5.2).

$$\Delta\Delta G_{H\ bond} = -(\alpha - \alpha_s)(\beta - \beta_s) \tag{5.2}$$

5.1.1 Interpretation of $\Delta\Delta G_{H\ bond}$

Interpretation of the expression for $\Delta\Delta G_{H\ bond}$ presented in (5.2), is simplified by considering four different conditions. These conditions are represented by the generic FGIP quadrant diagram depicted in figure 5.2, which shows which interaction dominates under the different regimes.

Under the conditions $\alpha > \alpha_S$ and $\beta > \beta_S$, the solute-solute interactions dominate and $\Delta\Delta G_{H\ bond}$ will be negative, therefore solute-solute association is favoured. This is due to stronger interactions between the solute molecules than with the solvent. This is satisfied in the upper right quadrant of figure 5.2.

When the solvent is a better hydrogen bond donor and acceptor than the solutes the conditions $\alpha < \alpha_S$ and $\beta < \beta_S$ are both satisfied. Solvent-solvent interactions are more favourable

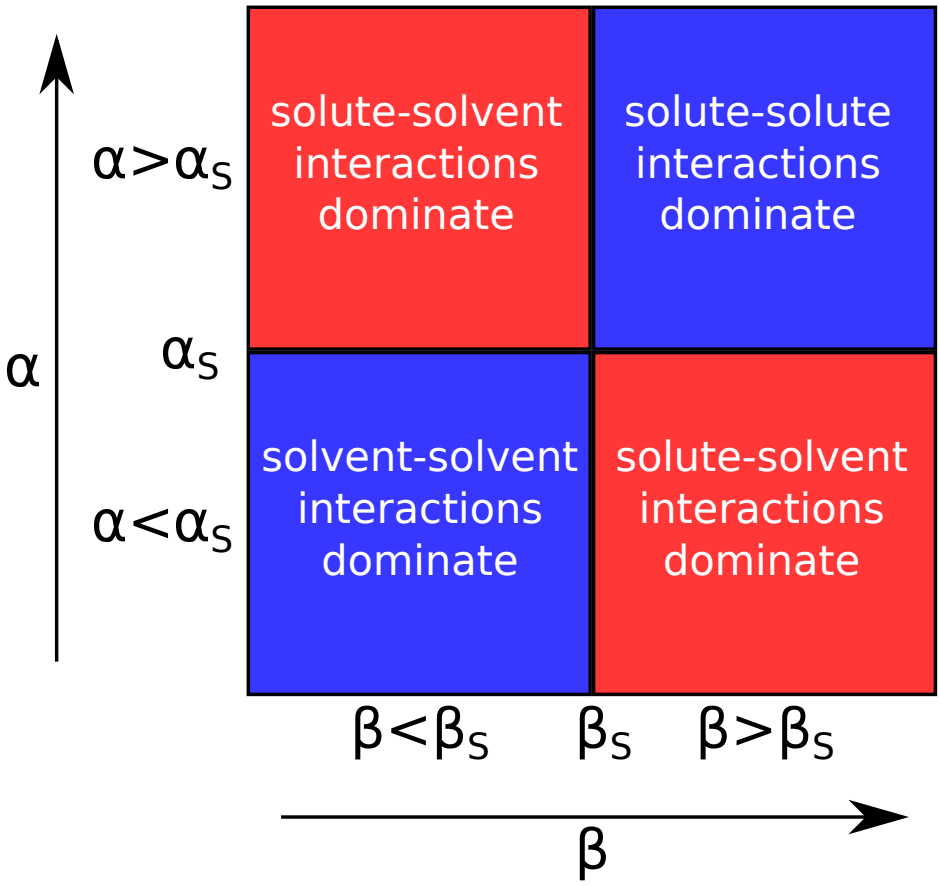


Fig. 5.2: Generic FGIP for the interaction of a hydrogen bond donor (α) with a hydrogen bond acceptor (β) in solvent S, Blue regions have favourable solute-solute interactions (negative $\Delta\Delta G_{H\ bond}$ values), and red regions have unfavourable solute-solute interactions (positive $\Delta\Delta G_{H\ bond}$ values).

than solvent-solute interactions, so solute-solute interactions are preferred. $\Delta\Delta G_{H\ bond}$ will be negative, due to the solvophobic effect. This is satisfied in the bottom left quadrant of figure 5.2.

The remaining two possibilities are: the solvent is a better hydrogen bond donor than the solute ($\alpha < \alpha_S$), but a worse hydrogen bond acceptor ($\beta > \beta_S$); the solvent is a better hydrogen bond acceptor than the solute ($\beta < \beta_S$), but a worse hydrogen bond donor ($\alpha > \alpha_S$. The greater strength of the solvent-solute interactions dominates, making solute-solute interactions unfavourable. This is satisfied in the upper left and bottom right quadrants of figure 5.2.

The boundaries between these regions, where the solute and solvent are of similar polarity, are of interest and are defined by $\alpha = \alpha_S$ and $\beta = \beta_S$.

5.1.2 Calculation of ΔG_{int}

$\Delta G_{H\ bond}$ in equation (5.2) can only be calculated if the solvent has a single type of hydrogen bond acceptor and a single type of hydrogen bond acceptor, so is unsuitable to study complex solvents. The calculated free energy of interaction ΔG_{int} represents the $\Delta\Delta G_{H\ bond}$ calculated with the surface site interaction model for the properties of liquids at equilibrium (SSIMPLE) approach [158]. This process allows for the treatment of solvents which have a collection of different hydrogen bond donor and acceptor sites, rather than a single hydrogen bond donor and acceptor in equation (5.2), so more complex systems can be evaluated.

In the SSIMPLE approach a molecule is described by a collection of surface site interaction points (SSIPs). A SSIP is assigned an interaction parameter, ε_i, which is equivalent to the hydrogen bond donor parameter (α) for positive sites or the hydrogen bond acceptor parameter ($-\beta$) for negative sites [149]. Assignment of these values is done by coarse graining of the *ab initio* calculated Molecular Electrostatic Potential Surface (MEPS) of the molecule in the gas phase [155].

To describe a liquid, SSIP interactions are treated in a pairwise manner, such that the association constant for binding between the ith and jth SSIP, K_{ij}, is given in equation (5.3).

$$K_{ij} = \frac{1}{2} e^{-\frac{\varepsilon_i \varepsilon_j + E_{vdW}}{RT}} \tag{5.3}$$

where $E_{vdW} = -5.6$ kJ mol^{-1}, ε_i, ε_j are the values of the SSIPs. E_{vdW} is the energy of van der Waals interactions between two SSIPs. The polar interaction term, $\varepsilon_i \varepsilon_j$, is set to zero for repulsive interactions (i.e. $\varepsilon_i \varepsilon_j > 0$). It is assumed that a state where the directional polar

interactions are misaligned such that only the non-directional van der Waals interactions are possible between SSIPs can be found.

Treatment of interactions between solvent and solute SSIPs requires the definition of a standard state to ensure K_{ij} is dimensionless. The standard state used is the theoretical density of SSIPs in the zero point solid, $c_{max} = 300$ M. The populations of free and bound species are then computed for the equilibrium state, as described by Hunter in [158] (detailed in chapter 4).

Calculation of the free energy difference between the interactions made in the bound state and the interactions made with the solvent in the free state, requires a new treatment of the bound state. ΔG_{int}, in equation (5.4), allows calculation of the equivalent $\Delta G_{H\ bond}$ based on the solvation energies of the two solutes in the free state and the free energy of the bound state.

$$\Delta G_{int}(\varepsilon_1, \varepsilon_2) = \Delta G_{bound}(\varepsilon_1, \varepsilon_2) - \Delta G_S(\varepsilon_1) - \Delta G_S(\varepsilon_2) \tag{5.4}$$

Where ε_1, ε_2 are the SSIP values of solute 1 and 2 respectively; $\Delta G_{bound}(\varepsilon_1, \varepsilon_2)$ is the free energy of the bound state defined below; $\Delta G_S(\varepsilon_1)$, $\Delta G_S(\varepsilon_2)$ are the solvation free energies of solute 1 and solute 2 respectively.

Solute Solvation free energy

The free energy of solvation for a SSIP, $\Delta G_{S,i}$, is given in equation (5.5), where $\Delta G_{b,i}$, $\Delta G_{c,i}$ are the binding energy and confinement energy of the SSIP respectively.

$$\Delta G_{S,i} = \Delta G_{b,i} + \Delta G_{c,i} \tag{5.5}$$

The binding energy for a SSIP, $\Delta G_{b,i}$, is given by equation (5.6), which arises due to the interactions with the solvent SSIPs in the condensed phase.

$$\Delta G_{b,i} = RT\,ln\left(\frac{\left[i_{free}\right]}{[i]}\right) \tag{5.6}$$

Where $\left[i_{free}\right]$ is the concentration of the free SSIP which is not bound to a solvent SSIP and $[i]$ is the total concentration of the SSIP in the phase.

SSIPs are confined to the condensed phase in order to calculate the speciation of the phase, but there is still probability of interaction between SSIPs in a condensed phase. The

confinement energy is the free energy change per SSIP for SSIPs confined to the condensed phase, $\Delta G_{c,i}$, shown in equation (5.7), from [158].

$$\Delta G_{c,i} = -RTln\left(\frac{\sqrt{1+8\theta}-1}{4\theta}\right) \tag{5.7}$$

Where θ is the fractional occupancy of the phase, and is the total concentration of SSIPs in the phase expressed relative to c_{max}.

To reduce the computational overhead of FGIP generation, the fitting of polynomial functions to the solvation free energies are used in calculation of $\Delta G_{b,i}$ (described in appendix C.2).

Free energy of the bound state

In order to use the solvation energies calculated with SSMIPLE, the free energy of the bound state must be defined relative to the same reference state. Therefore the probability that one SSIP does not interact with the other SSIP in a phase, which describes the bound state is required. For the interaction of two solute SSIPs, consider the bound state to be a phase where only the two SSIPs of interest are present, they can only interact with each other and the total SSIP concentration is the same as in the solution state. The total concentrations of each SSIP are given by equations (5.8) and (5.9).

$$[1] = \left[1_f\right] + 2K_{12}\left[1_f\right]\left[2_f\right] + 2K_{11}\left[1_f\right]^2 \tag{5.8}$$

$$[2] = \left[2_f\right] + 2K_{12}\left[1_f\right]\left[2_f\right] + 2K_{22}\left[2_f\right]^2 \tag{5.9}$$

Where $\left[1_f\right]$ and $\left[2_f\right]$ are the free concentrations of the two solute SSIPs in the bound state; $[1]$ and $[2]$ are the total concentrations of the two solute SSIPs in the bound state; K_{12} is the association constant for the interaction between the two solute SSIPs, given by (5.3), and the factor of 2 is a statistical factor as complexes $1\cdot 2$ and $2\cdot 1$ are equivalent. K_{11} and K_{22} correspond to the self interaction of the SSIPs. All concentrations are expressed relative to the SSIP standard state.

In the bound state, the total concentrations of each SSIP are the same. The constants K_{11} and K_{22} are equivalent to K_{VdW}, since it is assumed that the dipoles of repulsive interactions are misaligned, thus giving a purely non-polar interaction. This means that the free concentrations for both SSIPs are also the same.

For very strong polar interactions, $K_{12} >> K_{VdW}$, the concentrations in self interacting complexes are negligible. For non-polar SSIPs the self interactions are equally likely for any other non-polar SSIP, so are important for the complete description of the bound state of non-polar species.

Rearrangement of equations (5.8) and (5.9) to find the non interacting concentrations is therefore possible. The probability a SSIP is free in the bound state, P_f, is given in equation (5.10).

$$P_{f,1} = P_{f,2} = \frac{[1_f]}{[1]} = \frac{\sqrt{1+4\left(K_{12}+K_{VdW}\right)\theta}-1}{2\left(K_{12}+K_{VdW}\right)\theta} \tag{5.10}$$

The confinement of the SSIPs to a condensed phase leads to overestimation of the probability of interaction in the bound state. This can be accounted for by the inclusion of the confinement energy of SSIPs in the bound phase. This leads to the free energy of the bound state, ΔG_{bound}, shown in equation (5.11).

$$\begin{aligned}\Delta G_{bound}\left(\varepsilon_1,\varepsilon_2\right) &= RTln\left(P_{f,1}\right)+RTln\left(P_{f,2}\right)+\Delta G_{c,1}+\Delta G_{c,2} \\ &= 2RTln\left(\frac{\sqrt{1+4\left(K_{12}+K_{VdW}\right)\theta}-1}{2\left(K_{12}+K_{VdW}\right)\theta}\right)+2\Delta G_{c,i}\end{aligned} \tag{5.11}$$

Substitution into equation (5.4) leads to equation (5.12).

$$\begin{aligned}\Delta G_{int}\left(\varepsilon_1,\varepsilon_2\right) &= \Delta G_{bound}\left(\varepsilon_1,\varepsilon_2\right)-\Delta G_S(\varepsilon_1)-\Delta G_S(\varepsilon_2) \\ &= 2RTln\left(\frac{\sqrt{1+4\left(K_{12}+K_{VdW}\right)\theta}-1}{2\left(K_{12}+K_{VdW}\right)\theta}\right)+2\Delta G_{c,i} \\ &\quad -\Delta G_{b,i}(\varepsilon_1)-\Delta G_{c,i}-\Delta G_{b,i}(\varepsilon_2)-\Delta G_{c,i} \\ &= 2RTln\left(\frac{\sqrt{1+4\left(K_{12}+K_{VdW}\right)\theta}-1}{2\left(K_{12}+K_{VdW}\right)\theta}\right)-\Delta G_{b,i}(\varepsilon_1)-\Delta G_{b,i}(\varepsilon_2)\end{aligned} \tag{5.12}$$

5.2 FGIP Plots

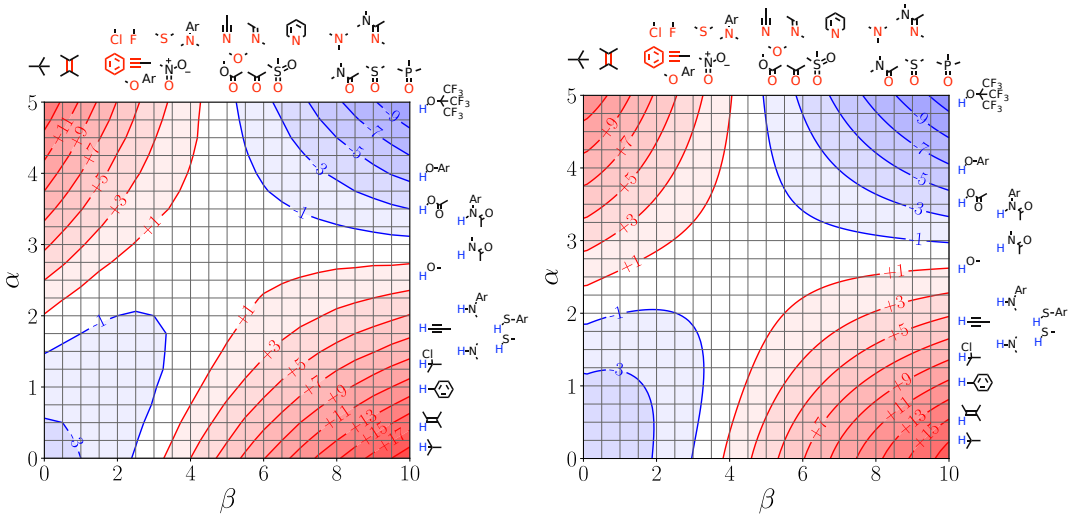
The calculation of ΔG_{int} for visualisation of FGIPs can be done for any solvent system. All the analysis presented here is for solvents at 298K.

5.2.1 FGIPs for single molecule solvents

The calculation of FGIPs for pure solvent systems was undertaken for a set of 261 solvents. Information on solvent concentrations used is in appendix G.1, with the full set of FGIPs generated in appendix H.

To show that calculation of ΔG_{int} is appropriate for generation of FGIPs, figure 5.3 compares the published FGIP for water (left) [149] calculated using $\Delta G_{H\ bond}$ with the corresponding FGIP calculated using ΔG_{int} (right). The agreement is excellent. In particular, the value of ΔG_{int} is zero when $\alpha = \alpha_S$ and $\beta = \beta_S$.

Water is a moderately good donor ($\alpha = 2.8$) and also moderately good acceptor ($\beta = 4.5$). This leads to a significant region of favourable solute-solute interactions for non-polar solutes, due to the strong solvent-solvent interactions, which are preferred to the weaker solvent-solute interactions. It looks similar to the quadrant diagram (figure 5.2), since water has values of α_S and β_S, which lie in the centre of both ranges.



(a) $\Delta G_{H\ bond}$ (b) ΔG_{int}

Fig. 5.3: FGIPs for two solutes in water at 298K. The left shows $\Delta G_{H\ bond}$ (kJ mol^{-1}) calculated using equation (5.2), and the right shows ΔG_{int} (kJ mol^{-1}) calculated using equation (5.4). The solute-solute interactions are favourable when negative (blue), and unfavourable when positive (red).

Figure 5.4 shows the FGIP for dimethyl sulfoxide (DMSO). DSMO is a very good hydrogen bond acceptor ($\beta = 8.5$) and a poor hydrogen bond donor ($\alpha = 1.4$). Comparison to the quadrant diagram in figure 5.2, shows that the upper left quadrant covers most of the

plot. This means DMSO is a very useful solvent for dissolving many molecules, due to the extensive range where solvent-solute interactions dominate.

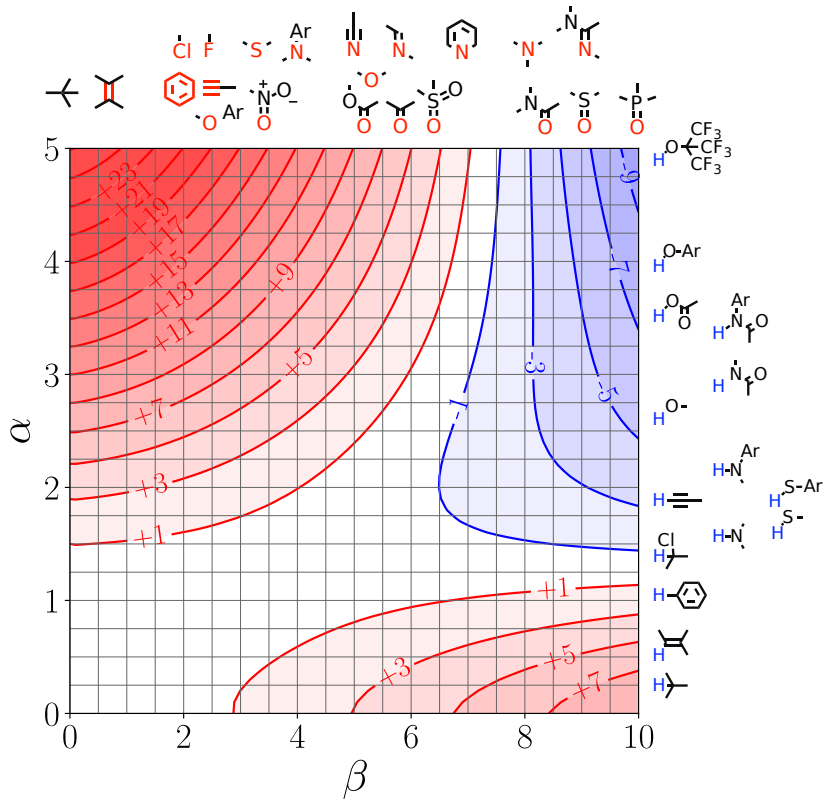


Fig. 5.4: FGIP showing ΔG_{int} (kJ mol^{-1}) for two solutes in dimethyl sulfoxide (DMSO) at 298K. The solute-solute interactions are favourable when negative (blue), and unfavourable when positive (red).

Toluene is a much less polar solvent, which is shown by the FGIP in figure 5.5. Toluene is a weak acceptor ($\beta = 2.0$), as well as weak hydrogen bond donor ($\alpha = 0.6$). Comparison to the quadrant diagram, figure 5.2, shows that the upper right quadrant would encompass most of the diagram. Solute-solute interactions are very likely to be stronger than interactions with the toluene, so intermolecular interactions between solutes are highly favourable, and dominate the FGIP.

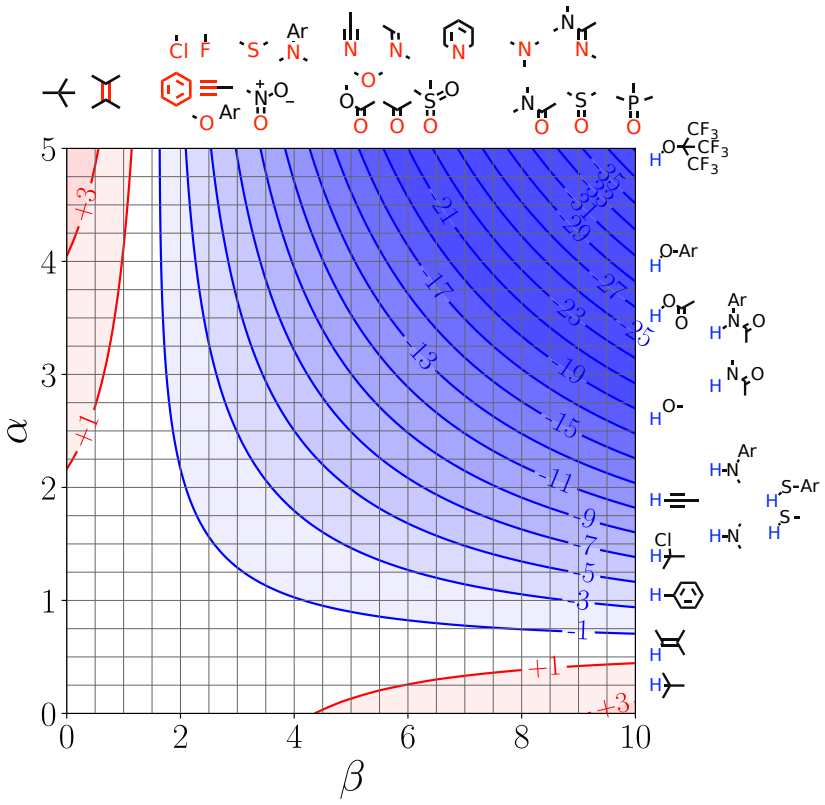


Fig. 5.5: FGIP showing ΔG_{int} (kJ mol^{-1}) for two solutes in toluene at 298K. The solute-solute interactions are favourable when negative (blue), and unfavourable when positive (red).

5.2.2 FGIPs for solvent mixtures

The expression for $\Delta\Delta G_{H\ bond}$ in equation (5.2), is unsuitable for the study of solvent mixtures as it assumes only a single type of solvent hydrogen bond donor and a single type of solvent hydrogen bond acceptor are present. This is untrue for solvent mixtures, where at least two different solvent hydrogen bond donors and at least two different hydrogen bond acceptors could be present. This leads to ambiguity as to which solvent parameter should be used in the relationship, with the stoichiometries of the two solvent components adding additional complexity.

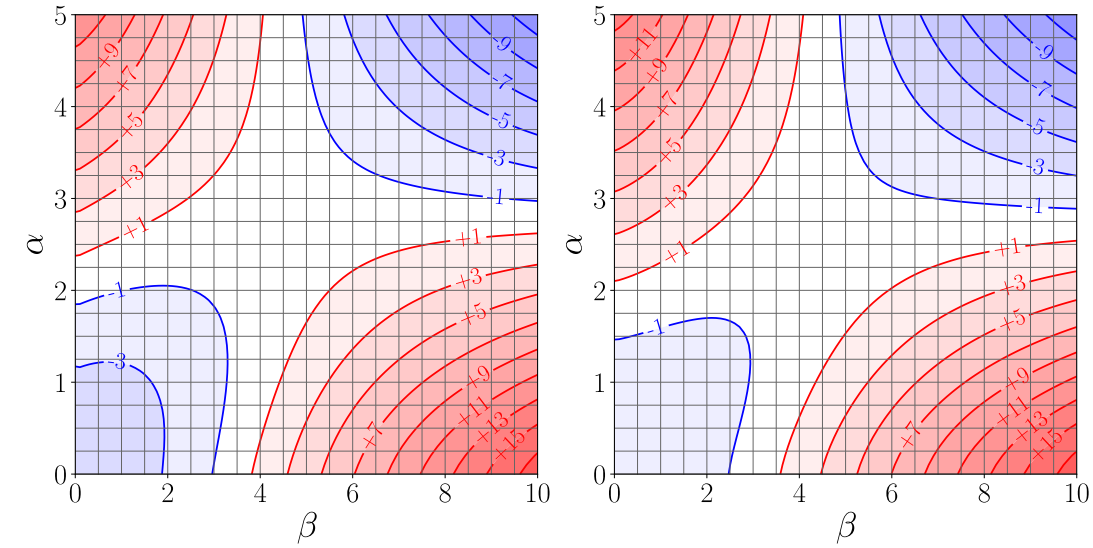
Treatment of interactions using the SSIMPLE approach, with equation (5.4) developed in this work removes any ambiguity, as all the interactions are considered during calculation, providing easy evaluation of interaction energies in solvent mixtures.

Two binary solvent mixtures have been investigated, water-ethanol and tetrahydrofuran-chloroform. These solvent mixtures were chosen as they are completely miscible across the full solvent composition range.

The FGIPs for the solvent mixtures were calculated at 5% volume intervals of the solvents. A subset are shown in the following sections, to highlight key features, with the FGIPs for all 5% volume fraction mixture in appendix H for reference.

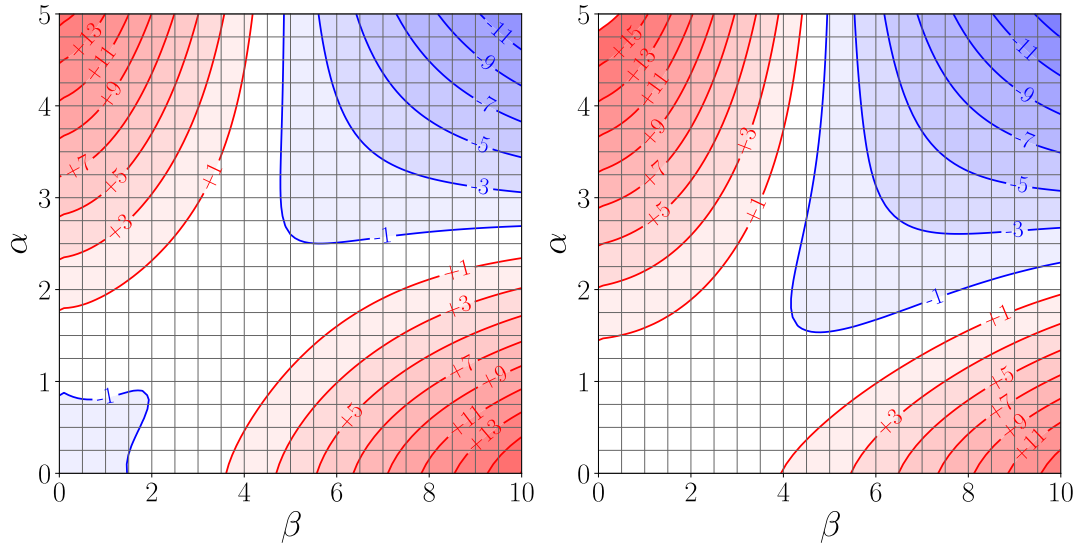
Water-Ethanol mixtures

Water and ethanol are both polar protic solvents. Using the experimental relationship in equation (5.2), if the strongest solvent donor and acceptor were picked to describe the solvent, then the FGIP for ethanol would be equivalent to that of water. However the alkyl chain has a significant effect on the solvation properties of non-polar solutes. This increases the solvation of non-polar solutes, so the region where solvent-solvent interactions dominate is much less pronounced as compared with water since the solvophobic interactions are weaker. Figure 5.6 highlights the effect of ethanol addition to water on increasing the strength of solvent interactions with non-polar solutes.



(a) Water 100% Ethanol 0%

(b) Water 70% Ethanol 30%



(c) Water 30% Ethanol 70%

(d) Water 0% Ethanol 100%

Fig. 5.6: FGIPs showing ΔG_{int} (kJ mol^{-1}) for two solutes in ethanol-water mixtures at 298K. The solute-solute interactions are favourable when negative (blue), and unfavourable when positive (red).

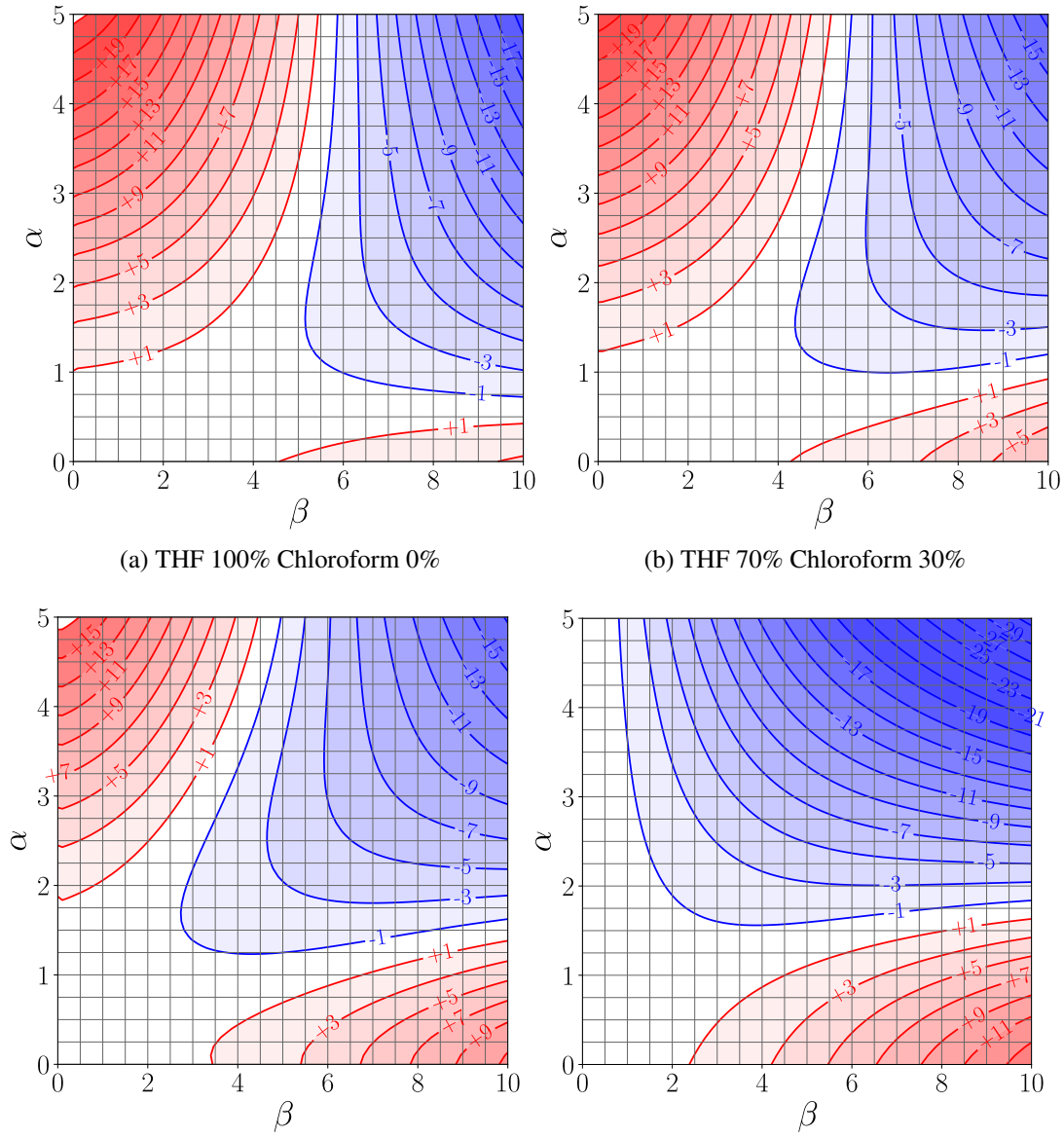
As the proportion of ethanol increases, the bottom left solvophobic region, starts to disappear. The alkyl chain on ethanol contains non-polar SSIPs which interact favourably with the non-polar solute SSIPs.

Tetrahydrofuran-Chloroform mixtures

Tetrahydrofuran (THF) and chloroform are two solvents commonly used during synthesis or in purification. Both solvents are moderately polar solvents, with THF being a moderate acceptor, and chloroform a moderate donor. Therefore THF preferentially interacts more strongly with donors, and chloroform preferentially interacts more strongly with acceptors.

Figure 5.7 shows the FGIPs for a selection of THF-chloroform mixtures. The FGIPs for the pure solvents appear to have just two regions, instead of the four of the quadrant diagram due to the difference in donor and acceptor strengths. THF contains a moderate hydrogen bond acceptor, but poor hydrogen bond donors, so the upper two quadrants extend down to cover the majority of the FGIP, such that favourable solute-solute interactions dominate the right hand side of the FGIP. Chloroform on the other hand has a good hydrogen bond donor and poor hydrogen bond acceptors, so the two right quadrants extend to cover most of the plot and in this case favourable solute-solute interactions dominate the top half of the FGIP.

As the solvents are mixed, competition with the solute SSIPs increases. This leads to a decrease in the solute interaction strength for the mixtures as stronger interactions with the solute hydrogen bond donors and acceptors from the solvent are now both possible. The region of favourable solute-solute interactions becomes restricted to the upper right region of the FGIPs in the solvent mixtures, with the regions between $0 < \varepsilon < 2$ and $0 > \varepsilon > -4$ showing large changes in interaction energy as solvent composition changes. Favourable interactions between the solvent and solute now dominate these low polarity regions.



(a) THF 100% Chloroform 0%

(b) THF 70% Chloroform 30%

(c) THF 30% Chloroform 70%

(d) THF 0% Chloroform 100%

Fig. 5.7: FGIPs showing ΔG_{int} (kJ mol^{-1}) for two solutes in chloroform-THF mixtures at 298K. The solute-solute interactions are favourable when negative (blue), and unfavourable when positive (red).

5.3 Origin of the constant in equation (5.1)

ΔG^o is the experimentally measured free energy change for formation of a 1:1 complex between two solutes.

This is defined using the association constant, given by equation (5.13).

$$\Delta G^o = -RTln(K) = -RTln\left(\frac{[A \cdot D]}{[A][D]}\right) \tag{5.13}$$

Where $[A \cdot D]$ is concentration of the complex; $[D]$ is concentration of uncomplexed hydrogen bond donor; $[A]$ is the concentration of uncomplexed solute hydrogen bond acceptor. Concentrations are expressed relative to a standard state of 1M in equation (5.13). This can also be expressed by considering the change in hydrogen bonding interactions, shown in equation (5.14), with $\Delta\Delta G_{H\ bond}$ defined in equation (5.2).

$$\Delta G^o = \Delta\Delta G_{H\ bond} + \gamma \tag{5.14}$$

Where γ is 6 kJ mol^{-1}.

5.3.1 Calculation of ΔG^o

It is also possible to calculate ΔG^o using the SSIMPLE approach. Starting from equation (5.13), relabelling the acceptor as solute 1 and donor as solute 2, yields equation (5.15).

$$\Delta G^o = -RTln\left(\frac{[1 \cdot 2]}{[1][2]}\right) \tag{5.15}$$

The concentrations of all three species in (5.15) can be calculated from the speciation in SSIMPLE, thus allowing the calculation of the observed free energy change from a SSIMPLE calculation.

$$[1 \cdot 2] = 2K_{1\cdot2}[1_f][2_f] \tag{5.16}$$

$[1 \cdot 2]$ is the concentration of the complex defined relative to the SSIP standard state (c_{max}), defined in equation (5.16); $[1_f]$, $[2_f]$ are the unbound concentrations of solute 1 and 2 respectively, i.e. the concentrations of SSIPs that do not interact with each other or the solvent; $K_{1\cdot2}$ is defined in (4.1).

The concentrations of solute 1 not in the complex, $[1]$, is given by equation (5.17), where $[1_S]$ is the concentration of species 1 bound to solvent, $K_{S,1}$ is the overall equilibrium constant defining the interaction of all solvent SSIPs with species 1. A similar expression for $[2]$, the concentration of solute 2 not in the complex, is in equation (5.18), where $[2_S]$ is the

concentration of species 2 bound to solvent, $K_{S,2}$ the equilibrium constant defining the interaction of all solvent SSIPs with species 2.

$$[1] = [1_f] + [1_S] = [1_f] + K_{S,1}[1_f] = (1 + K_{S,1})[1_f] \tag{5.17}$$

$$[2] = [2_f] + [2_S] = [2_f] + K_{S,2}[2_f] = (1 + K_{S,2})[2_f] \tag{5.18}$$

Where the equilibrium constant for solvation of the ith species, $K_{S,i}$, is defined in equation (5.19).

$$K_{S,i} = \frac{[i_S]}{[i_f]} \tag{5.19}$$

Substitution of equations (5.16), (5.17) and (5.18) into equation (5.15) and the conversion of concentrations from the SSIP standard state to the 1M standard state gives equation (5.20).

$$\Delta G^o = -RTln\left(\frac{2K_{1\cdot2}[1_f][2_f]}{c_{max}\left((1 + K_{S,1})[1_f]\right)\left((1 + K_{S,2})[2_f]\right)}\right) \tag{5.20}$$

Separation of the expression yields equation (5.21).

$$\Delta G^o = -RTln\left(2K_{1\cdot2}\right) + RTln\left(c_{max}\right) + RTln\left(1 + K_{S,1}\right) + RTln\left(1 + K_{S,1}\right) \tag{5.21}$$

By noting that the third and fourth terms in equation (5.21) correspond to the binding energies of solvation (see appendix C.1 for full derivation), the free energy for formation of the complex can be written as equation (5.22).

$$\Delta G^o = -RTln\left(2K_{1\cdot2}\right) + RTln\left(c_{max}\right) - \Delta G_{S,1} - \Delta G_{S,2} + 2\Delta G_c \tag{5.22}$$

Substitution of $K_{1\cdot2}$ (equation (4.1)) into equation (5.22) yields equation (5.23).

$$\Delta G^o = \varepsilon_i \varepsilon_j + E_{vdW} + RTln\left(c_{max}\right) - \Delta G_{S,1} - \Delta G_{S,2} + 2\Delta G_c \tag{5.23}$$

5.4 Derivation of γ

The difference between ΔG^o and $\Delta\Delta G_{Hbonds}$, as shown in (5.13) is γ, which experimentally was found to be 6 kJ mol^{-1} as described in [149]. The expressions derived for ΔG^o and ΔG_{int} provide a method for calculating γ, shown in equation (5.24).

$$\begin{aligned}\gamma &= \Delta G^o - \Delta G_{int} \\ &= \varepsilon_i\varepsilon_j + E_{vdW} + RTln\left(c_{max}\right) - \Delta G_{S,1} - \Delta G_{S,2} + 2\Delta G_c \\ &\quad - 2RTln\left(\frac{\sqrt{1+4\left(K_{12}+K_{VdW}\right)\theta}-1}{2\left(K_{12}+K_{VdW}\right)\theta}\right) + \Delta G_{b,i}(\varepsilon_1) + \Delta G_{b,i}(\varepsilon_2) \\ &= \varepsilon_i\varepsilon_j + E_{vdW} + RTln\left(c_{max}\right) - 2RTln\left(\frac{\sqrt{1+4\left(K_{12}+K_{VdW}\right)\theta}-1}{2\left(K_{12}+K_{VdW}\right)\theta}\right)\end{aligned} \tag{5.24}$$

From equation (5.24), note that γ is not a constant, but actually has a dependence on the solute SSIP values, as $K_{1\cdot2}$, ε_i and ε_j appear in the equation.

5.4.1 Tight binding limit

In the limit of tight binding, when $K_{1\cdot2}\theta >> 1$ ($K_{1\cdot2} >> K_{VdW}$ also holds), the probability that a SSIP is not interacting in the bound state (equation (5.10)) simplifies, such that γ becomes equation (5.25).

$$\begin{aligned}\gamma_{tight} &= \varepsilon_i\varepsilon_j + E_{vdW} + RTln\left(c_{max}\right) + RTln\left(K_{1\cdot2}\theta\right) \\ &= -RTln\left(2K_{1\cdot2}\right) + RTln\left(c_{max}\right) + RTln\left(K_{1\cdot2}\theta\right) \\ &= RTln\left(\frac{\theta c_{max}}{2}\right)\end{aligned} \tag{5.25}$$

This expression shows that the constant arises from the solvent-solvent interactions. The concentration inside the bracket is the concentration of solvent-solvent SSIP interactions if the SSIPs were fully bound. This suggests that the origin of the γ term in the experimental equation is related to the concentration of solvent.

In (5.13) the solvent is not considered and the standard state is 1 M, but the concentration of the solvent is much higher (10 M for carbon tetrachloride, the solvent used to determine the α/β scales). This must be accounted for when the free energy change of the hydrogen bonding interactions in figure 5.1 is considered. In other words the solvent concentration

should be included in the expression for equilibrium constant. The equilibrium constant for the interaction of the two solutes in solution in equation (5.26).

$$K_{eq} = \frac{[A \cdot D]\,[S \cdot S]}{[A \cdot S]\,[D \cdot S]} \tag{5.26}$$

Where $[S \cdot S]$ is the concentration of solvent. Thus an extra term in our expression for the free energy change from experimental measurement which is related to the concentration of solvent-solvent interactions is found.

Equation (5.25) provides an expression that is now independent of the SSIP values of interest. The relationship in equation (5.25) provides a way to calculate a value of γ for solvents. The plot in figure 5.8 shows the distribution of γ values for the set of 261 pure solvents (details in appendix G). The mean value is 10.54 kJ mol^{-1} with a standard deviation of 0.27 kJ mol^{-1}.

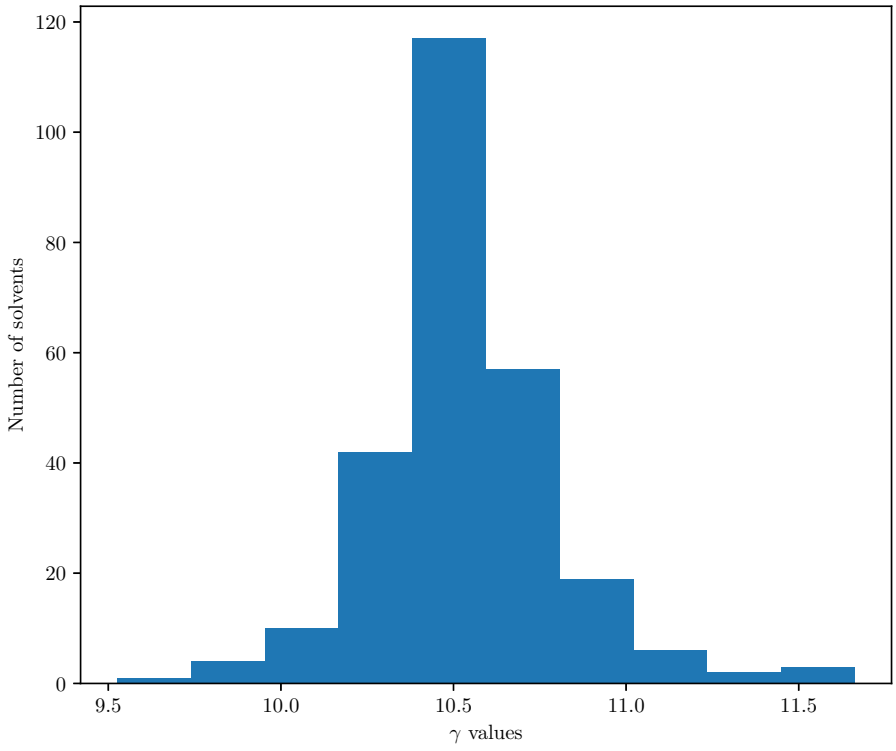


Fig. 5.8: Histogram of the distribution of γ for the set of solvents listed in appendix G, calculated using equation (5.25).

5.4.2 Comparison of γ to experimental value

The calculated value of γ is 10.5 kJ mol^{-1}, which is significantly more than the experimental value. The disparity can be traced to the creation of the hydrogen bond donor and acceptor scales in [149]. As discussed in chapter 1, the scales are derived from experimental association constant measurements [145, 149–151] in carbon tetrachloride.

The experimental parameters for carbon tetrachloride are $\alpha_S = 1.4$, $\beta_S = 0.6$. This differs to the calculated description which has less polar SSIPs with the largest $\alpha = 1.1$ and the largest $\beta = 0.15$, (plot of all SSIPs in appendix G). The less polar SSIP description of the solvent is offset by the more positive the constant, γ.

Comparison of the values of ΔG^o calculated using equation (5.1) against ΔG^o calculated using equation (5.23) for carbon tetrachloride is shown in figure 5.9. The most significant deviation occurs for the most polar solutes, with errors of less than 2 kJ mol^{-1} for the strongest interactions.

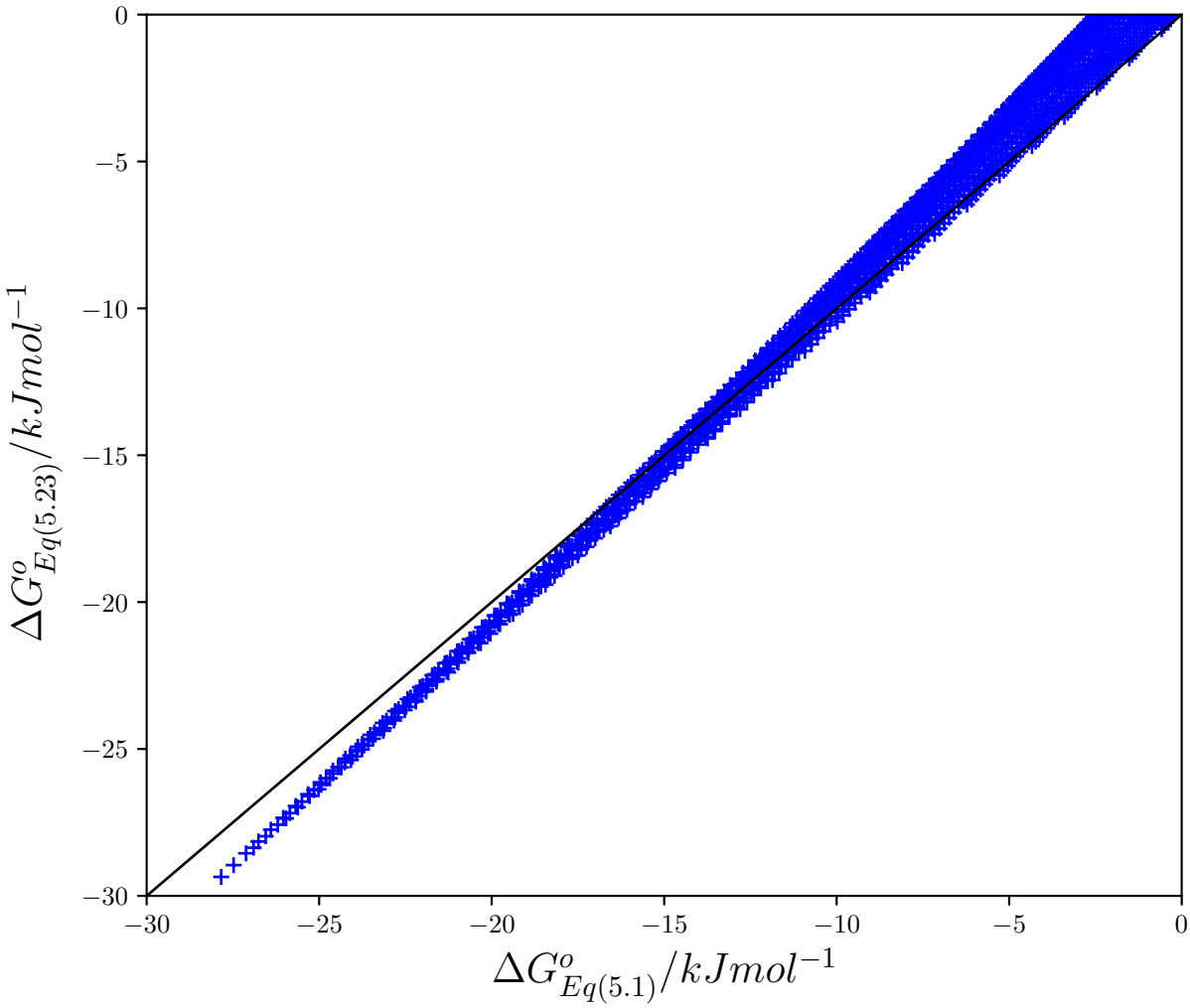


Fig. 5.9: Comparison of ΔG^o calculated using equation (5.1) against ΔG^o (equation (5.23)), plotted with blue crosses. Black line is y=x.

5.5 Conclusions

Generation of functional group interaction profiles for any solvent composition is now possible with the equations derived in this work. Examples of nearly 300 systems at 298K are included for reference in appendix H, including a selection of binary solvent mixtures which previous approaches were unable to describe. The origin of the difference between

$\Delta\Delta G_{H\ bond}$ and ΔG^o, γ (defined in equation (5.24)), was shown to correspond to the treatment of the solvent concentration in the experimental definition of equilibrium constants. The calculated value of γ was shown to be constant.

Chapter 6

Solvent Similarity

Solvent-solute interactions play an important role in dissolution and solvation of compounds. Solvents are used in several industries, as a medium for reactions to synthesise compounds such as in the pharmaceutical and agrochemical industries, or for the application of surface coatings and paints. Finding the solvent with the correct properties to solvate the required components is therefore important. Multiple solvents may possess similar efficacy at solvation, so secondary factors such as cost and environmental effects also play a role in selection. With the increasing desire to reduce environmental impact and improve efficiency in processes, the concept of 'green' chemistry [77] is becoming more prominent, with scales to assess the environmental impact of solvents developed [78–81]. This has led to the use of alternative solvents with lower environmental impact in pharmaceutical companies [79, 284].

To find suitable solvent replacements in industry, solvents with similar properties must be identified. The prediction of solvents with similar properties requires the definition of a metric with which to compare them. Previously Hansen [98] has generated a similarity metric for solvents, based on work originally undertaken by Hildebrand [97].

A new solvent similarity metric to use for solvent comparisons is presented here. The difference between solvation profiles of a single solute surface site interaction point (SSIP) [149, 155], calculated using the Surface Site Interaction Model for Liquids at Equilibrium (SSIMPLE) [158] is proposed as a suitable metric.

6.1 Surface Site Interaction Model for Liquids at Equilibrium (SSIMPLE)

A collection of Surface Site Interaction Points (SSIPs) is used to describe the interactions that a molecule makes with the environment, e.g. the solvent. Each SSIP is assigned an interaction parameter, ε_i, which is equivalent to the experimentally measured hydrogen bond donor parameter (α) for positive sites or the hydrogen bond acceptor parameter ($-\beta$) for negative sites [149]. The assignment of these values is done by footprinting of the *ab initio* calculated Molecular Electrostatic Potential Surface (MEPS) of the molecule [155].

To describe interactions between SSIPs in the liquid phase, all possible SSIP interactions are treated in a pairwise manner, with the association constant for interaction between the ith and jth SSIPs in a phase given by (6.1).

$$K_{ij} = \frac{1}{2} e^{-\frac{\varepsilon_i \varepsilon_j + E_{vdW}}{RT}} \tag{6.1}$$

Where $E_{vdW} = -5.6$ kJ mol^{-1}, ε_i, ε_j are the values of the SSIPs, R is the gas constant and T is the temperature.

E_{vdW} is the energy attributed to van der Waals interactions between two SSIP fragments and comes from [278]. The polar interaction term, $\varepsilon_i \varepsilon_j$, is set to zero for repulsive interactions (i.e. $\varepsilon_i \varepsilon_j > 0$). It is assumed that a state where the directional polar interactions are misaligned such that only the non-directional van der Waals interactions are possible between SSIPs can be found. Since K_{ij} is dimensionless, the concentrations of SSIPs must be defined relative to a standard state. The standard state used is the theoretical density of SSIPs in the zero point solid, $c_{max} = 300$ M. The populations of free and bound species are then computed for the equilibrium state, as described by Hunter in [158] (detailed in chapter 4).

The solvation free energy of a SSIP, is defined as the change in energy for transfer from the gas phase to the solution phase. The gas phase is assumed to be sufficiently dilute that there are no interactions with the molecule, i.e. all its SSIPs are completely free.

For a single SSIP, i, the solvation free energy is the sum of the binding and confinement energies, shown in equation (6.2).

$$\Delta G_{S,i} = \Delta G_{b,i} + \Delta G_{c,i} \tag{6.2}$$

The binding energy for a SSIP, $\Delta G_{b,i}$, is given by equation (6.3), which arises due to the interactions with solvent SSIPs in the condensed phase.

$$\Delta G_{b,i} = RTln\left(f^i\right) \tag{6.3}$$

$$f^i = \frac{\left[i_{free}\right]}{[i]} \tag{6.4}$$

Where f^i, is the fraction free (equation 6.4) of the SSIP, $\left[i_{free}\right]$ is the concentration of free i (calculated using the SSIMPLE approach, [158]) and $[i]$ is the total concentration of i in the phase.

In order to calculate the speciation of the phase, the SSIPs are confined to the condensed phase. The confinement energy, shown in equation (6.5), is the free energy change for SSIPs being confined to the condensed phase [158].

$$\Delta G_{c,i} = -RTln\left(\frac{\sqrt{1+8\theta}-1}{4\theta}\right) \tag{6.5}$$

Where θ is the fractional occupancy of the phase, and is the total concentration of SSIPs in the phase expressed relative to the standard state of the zero point solid concentration, c_{max}. Note that $\Delta G_{c,i}$ for a SSIP is a constant for the phase, as it only depends on the fractional occupancy of the phase and the temperature.

6.2 Similarity Quantification using SSIMPLE

Figure 6.1 shows plots of solvation energies as a function of all possible SSIP values for a selection of solvents. These solvation profiles are smooth so a polynomial function can be fitted to be able to regenerate the curves on the fly (detailed in appendix C.2), increasing the speed of future analysis.

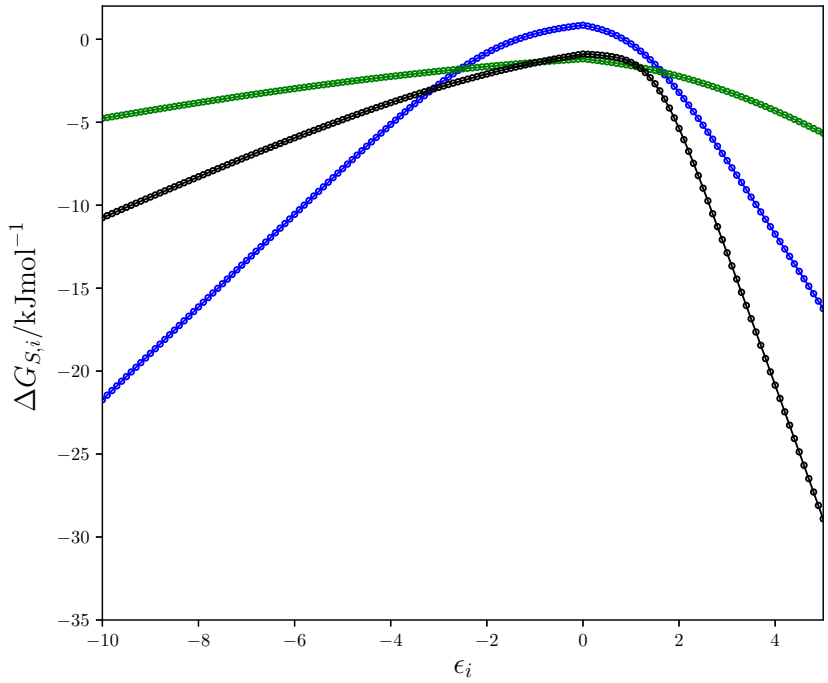


Fig. 6.1: Solvation profiles for water (blue), DMSO (black) and toluene (green). The calculated solvation free energy for a single SSIP ($\Delta G_{S,i}$ kJ mol^{-1}) plotted as a function of SSIP value, ε_i, as circles, with polynomial fit used during analysis plotted as a solid line (described in appendix C).

The solvation profiles in figure 6.1 are dramatically different for each of the solvents plotted. Toluene is a non-polar solvent, so has poor interactions with hydrogen bond donors and acceptors and has a relatively shallow curve.

Dimethyl sulfoxide is a polar aprotic solvent. It has a strong hydrogen bond acceptor, therefore it solvates donors very strongly, with a large ΔG_S for positive SSIPs, but is only a weak donor so solvates acceptors poorly; with a smaller negative value of ΔG_S for negative SSIPs. This produces a more assymetric curve.

Water is a moderate donor and acceptor, so has a sharp inverted 'u' shaped curve. It is classed as a polar protic solvent. The interaction of water is very weak with non-polar hydrogen bond acceptor and donor sites, leading to hydrophobic interactions with non-polar molecules. Preferential interaction between the water hydrogen bond donor and acceptor groups means that for neutral SSIPs the solvation energy is unfavourable; $\Delta G_{S,i}$ is positive for water around $\varepsilon_i = 0.0$, but is negative for the other solvents displayed here.

The difference between the shapes of the solvation profiles match the qualitative description of a solvent as non-polar, polar aprotic or polar protic.

6.3 Similarity Metric

Comparison of the solvation profiles requires the definition of a metric which describes the distance between the curves. The metric would quantify the extent of the difference between two solvents for all possible solute SSIP values.

Hausdorff and Fréchet distances [285, 286] are two possible distance metrics previously used in assessing the similarity of curve paths. The Hausdorff distance equation (6.6) is the largest minimum distance required to maintain a connection between the two curves along the entire curve lengths. The Fréchet distance, in equation (6.7) is the same as the Hausdorff distance, except that the traversal of the curves is unidirectional, i.e. you must increase the values of ε if you have started from the most negative end; you cannot retraverse a section of the profile.

$$d_{\mathrm{H}}(\Delta G_{S_1}, \Delta G_{S_2}) = \max\left\{ \sup_{\varepsilon_1 \in E} \inf_{\varepsilon_2 \in E} d(\Delta G_{S_1}(\varepsilon_1), \Delta G_{S_2}(\varepsilon_2)), \sup_{\varepsilon_2 \in E} \inf_{\varepsilon_1 \in E} d(\Delta G_{S_1}(\varepsilon_1), \Delta G_{S_2}(\varepsilon_2)) \right\} \tag{6.6}$$

$$F(\Delta G_{S_1}, \Delta G_{S_2}) = \inf_{\varepsilon_1, \varepsilon_2} \max_{t \in [0,1]} \left\{ d\Big(\Delta G_{S_1}(\varepsilon_1(t)), \Delta G_{S_2}(\varepsilon_2(t))\Big) \right\} \tag{6.7}$$

Where ΔG_{S_1}, ΔG_{S_2} are the solvation profiles for solvents 1 and 2 respectively; d is a distance function; ε_1, ε_2 are solute SSIP values in solvent system 1 and solvent system 2 respectively; E is the set of all permitted ε values; t is a scaling parameter. The value of t is such that the values of ε_1 and ε_2 are always incremented in the same direction.

These distances will be heavily dependent on the extrema of the curves, so will not capture the small but important differences in the non polar region well. To overcome the issue of extrema biasing, a domain averaged root mean square standard deviation between solvation profiles is proposed as the similarity metric.

Root mean square standard deviation (RMSD)

The root mean standard deviation (RMSD) between the two curves is given by equation (6.8).

$$RMSD\left(S_1,S_2\right)=\sqrt{\frac{\sum_{i=1}^{N}\left(\Delta G_{S_1}\left(\varepsilon_i\right)-\Delta G_{S_2}\left(\varepsilon_i\right)\right)^2}{N}} \tag{6.8}$$

Where S_1, S_2 are the solvents, ΔG_{S_2}, ΔG_{S_2} are the solvation profiles for a single SSIP for solvent one and solvent two respectively, and ε_i is the solute SSIP value. The summation is over all ε_i in the region of interest.

If the summation was carried out over the full curve then the RMSD would be dominated by the solvation energy differences for the extreme ends of the ε_i scale. Instead, to avoid this issue, the curves are partitioned and each domain is given an equal weighting after normalisation. The distinct domains are given by equation (6.9). This is required to produce a metric that is not biased by the interaction of a solvent with very strong donor or acceptors, which are relatively uncommon compared to the occurrence of non-polar hydrocarbon groups in chemical structures.

$$Domains=\begin{cases}-10.0\leq\varepsilon_i<-5.0\\-5.0\leq\varepsilon_i<-2.0\\-2.0\leq\varepsilon_i<0.0\\0.0\leq\varepsilon_i<1.0\\1.0\leq\varepsilon_i<3.0\\3.0\leq\varepsilon_i<5.0\end{cases} \tag{6.9}$$

The normalisation factor for the jth domain, η_j, used is defined in (6.10). This is the maximum RMSD between any two solvents in the set of all solvents, S, for the domain. All RMSDs for that domain are scaled by this value.

$$\eta_j=\max_{k,l\in S}\left\{RMSD\left(S_{k,j},S_{l,j}\right)\right\} \tag{6.10}$$

This results in scaled values between 0 and 1 for each domain. Where 0 means the domains are identical, and a value of 1 means the domains are the most dissimilar. The mean

of the scaled values of all domains was then used as the distance between solvents to describe similarity, in equation (6.11).

$$Similarity(S_k, S_l) = \frac{1}{D}\sum_{j}^{D} \frac{RMSD\left(S_{k,j}, S_{l,j}\right)}{\eta_j} \tag{6.11}$$

Where j is the domain, D is the total number of domains. The relative distance produced from scaling the solvent values produces a set dependent metric, such that exact distances are not comparable between different solvent sets, providing a qualitative guide to similarity within a collection of solvent profiles.

A quantitative measure can be created by defining normalisation factors, η_j based on a standard reference set. The set of 261 pure solvent molecules listed in appendix G were chosen to act as this reference, with the normalisation factors in table 6.1.

Domain	η_j/kJ mol^{-1}
$-10.0 \leq \varepsilon_i < -5.0$	29.573840367308566
$-5.0 \leq \varepsilon_i < -2.0$	9.4352526310819851
$-2.0 \leq \varepsilon_i < 0.0$	2.1109705595898145
$0.0 \leq \varepsilon_i < 1.0$	1.981619630257929
$1.0 \leq \varepsilon_i < 3.0$	11.244516687700679
$3.0 \leq \varepsilon_i < 5.0$	30.555524859353685

Table 6.1 Normalisation factors for SSIP domains.

6.4 Similarity Visualisation

With the distance metric defined between all curves visualisation of the distances is required to be able to interpret the similarity. For a group of N solvents a similarity matrix could be created, of size $N \times N$, which could then be plotted with a colour scale used to represent the degree of similarity. This could be useful for small N ($N \ \mathscr{O}(10)$), but becomes unintelligible for larger N. 261 pure solvent molecules have been considered (detailed in appendix G), thus this approach to visualisation is not suitable, so clustering of the data is required.

6.4.1 Clustering

Clustering of individual datapoints into larger collections provides a way to structure the similarity distance data stored in the similarity matrix to readily display it. The process of

aggregating individual solvent entries into clusters of like solvents indicates the degree of similarity by spatial positioning within the diagram.

Three types of clustering algorithm have been considered, which are:

- K-means[287, 288]
- Density-based spatial clustering of applications with noise (DBSCAN)[289]
- Hierarchical [290]

K-means clustering algorithms[287, 288] create K clusters of points, where the number of clusters, K, must be specificied a priori. An iterative process is required to assign datapoints to clusters, after initial guesses for each cluster centroids. The datapoints are assigned to one of these clusters based on the voronoi diagram of the space. A new centroid is calculated and then this process is repeated until the cluster centroids remain static.

The DBSCAN[289] clustering approach produces a collection of clusters without need to specify the number of clusters. Instead a distance cutoff, δ, and the minimum number of points (*minPts*) to create a dense region are required. Core points of a cluster have *minPts* within a radius δ. The edge of a cluster is defined by points within δ of fewer core points than *minPts*. Again, an iterative approach is required to complete assignment to the different clusters, noting that the assignment of cluster edges may be ambiguous if there are multiple points from different clusters equidistant, but the total number of points is less than *minPts*.

In Hierarchical clustering [290] the approach does not yield a total number of distinct clusters as in the previous two approaches. During a hierarchical clustering, the two most similar nodes (datapoints) are replaced by a new node representing an aggregate of the two previous nodes. The distance to this new node is calculated, and a new distance matrix is computed. This aggregation is repeated until only one node is present, when there is only one entry in the resulting distance matrix.

The sequential merging of entries into higher order nodes gives a dendrogram structure which can be displayed. This made the hierarchical approach the most suited to the task of visualisation, so to understand the degree of similarity between individual solvents instead of the formation of clusters. The degree of similarity can be determined by the distance to the closest shared node between two solvent, purely from visual inspection.

There are multiple strategies for the creation of new nodes in hierarchical clustering to represent the combination of the two most similar nodes from the previous step. These strategies are:

- Single linkage

- Complete linkage
- Unweighted pair group method with arithmetic mean (UPGMA)
- Weighted pair group method with arithmetic mean (WPGMA)
- Ward's method [291]

The single linkage method uses the minimum distance between any leaf node in the aggregated node to another node, for the updated distances. The complete linkage method uses the maximum distance between any leaf node in the aggregated node to another node, for the updated distances. UPGMA uses the mean distance of all leaf nodes to the other node as the distance to the aggregated node. For WPGMA the mean of the distances from each node cluster being combined is used to represent the distances to the aggregated node. This produces an ultrametric tree, therefore leaf nodes are equidistant to the root node. In Ward's method the clustering criteria used minimises the total in-cluster variance after merging to determine the new position.

Of these methods the UPGMA approach was evaluated to be the most logical choice, as the average similarity of the clusters is of most interest. Since the relative distances between solvents after clustering is of interest, it would be lost if WPGMA is used, so that is unsuitable. The single and complete linkage methods lead to the extreme values in the clusters dominating, which is also not suitable for the requirements. The variance within a cluster is not important as a qualitative view of similarity for visualisation is required, thus Ward clustering can be discounted.

6.5 Solvent comparison

Solvation profiles were then compared using the defined distance metric in equation (6.11). The smaller the normalised distance, the more similar two solvents are.

6.5.1 Pure Solvents

The solvation profiles for 261 pure solvents at 298K were used in the following analysis. Details of all the solvents are in appendix G, with the polynomials used to describe the solvation profiles in appendix G.

The complete dendrogram is presented in figure 6.2. This provides an overview of the similarity for all solvents in the set. Examination of this information requires truncation of

the set at distinct branch nodes to visualise the grouping of solvents within smaller regions of solvent space. The numbers on figure 6.2 correspond to the branch nodes used to generate regional dendrograms of subsets of solvents.

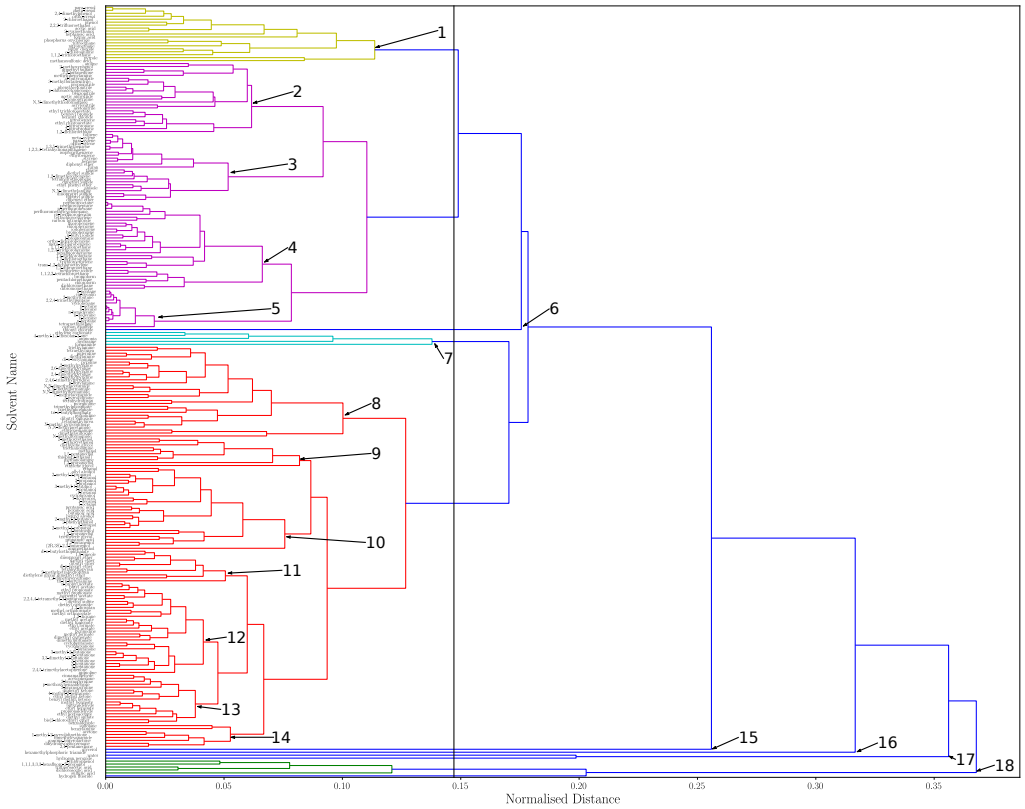


Fig. 6.2: Dendrogram showing normalised distances between 261 pure solvents at 298K using UPGMA clustering algorithm. Numbers correspond to nodes used for partitions into smaller subsets. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

A partitioning scheme with a threshold of two fifths of the maximum distance between nodes was used as a cutoff to colour different solvent clusters in all of the figures. For the complete set of solvents this identifies four distinct clusters with this threshold level.

Distribution of distances

The distribution of unique solvent distances is shown in figure 6.3, peaks between 0.1 and 0.2 normalised distance units, with a long tail. The maximum theoretical normalised distance for two solvents is 1.0, when a pair of solvents have the greatest dissimilarity in all six

domains. The observed maximum distance is smaller than this, with a value of 0.80, for hydrogen fluoride and hexamethylphosphoramide (HMPA). Hydrogen flouride contains a very good hydrogen bond donor, but poor hydrogen bond acceptors, so solvates hydrogen bond acceptors well, and hydrogen bond donors poorly. Whereas HMPA has the best hydrogen bond acceptor of any solvent, but poor hydrogen bond donors, so exhibits the opposite solvation behaviour to hydrogen fluoride.

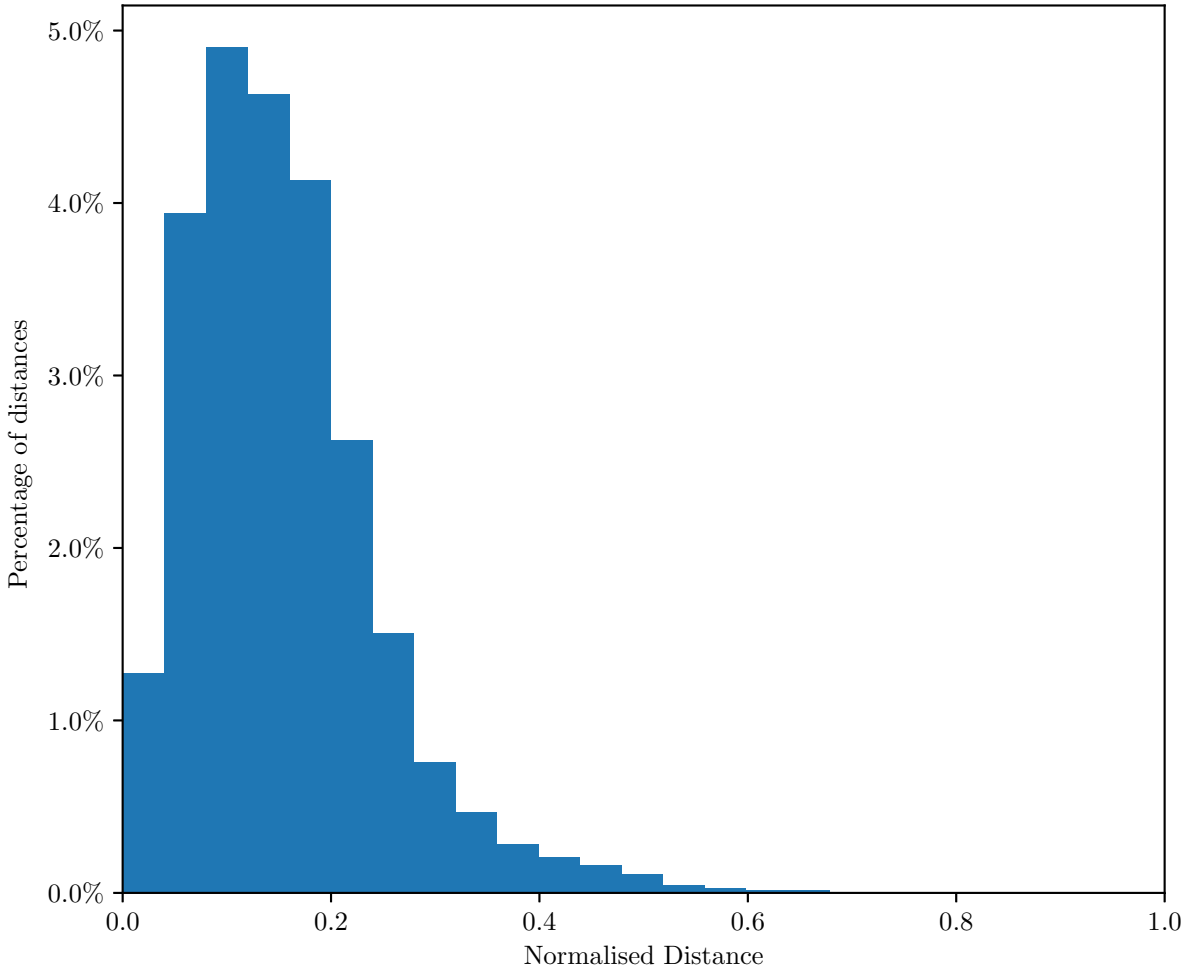


Fig. 6.3: Distribution of distances between solvents, as percentage of unique solvent distances.

Regional dendrograms

The regional dendrograms are plotted below to show the solvents to the left of the nodes labelled in figure 6.2. The normalised distances for these regional plots are those in table 6.1.

Region 1 (figure 6.4) is a collection of molecules with good hydrogen bond donors and moderate hydrogen bond acceptors. It is mainly phenols ($\alpha \approx 4$), acids and alcohols, with electron withdrawing substituents.

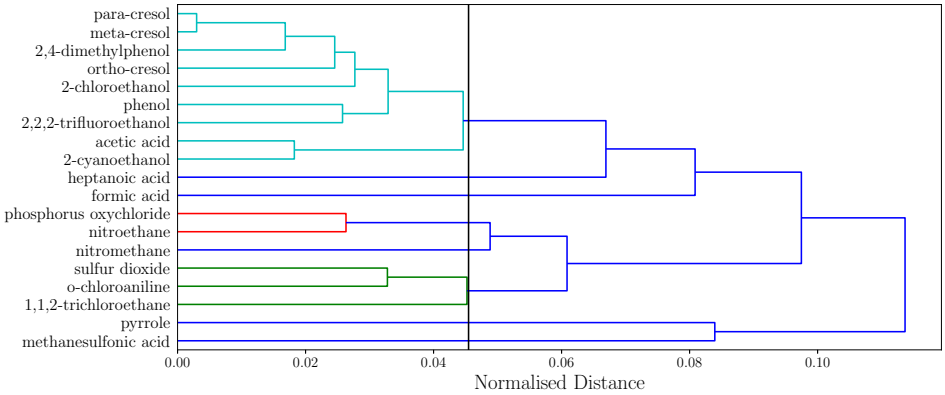


Fig. 6.4: Dendrogram for solvents to the left of node 1 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

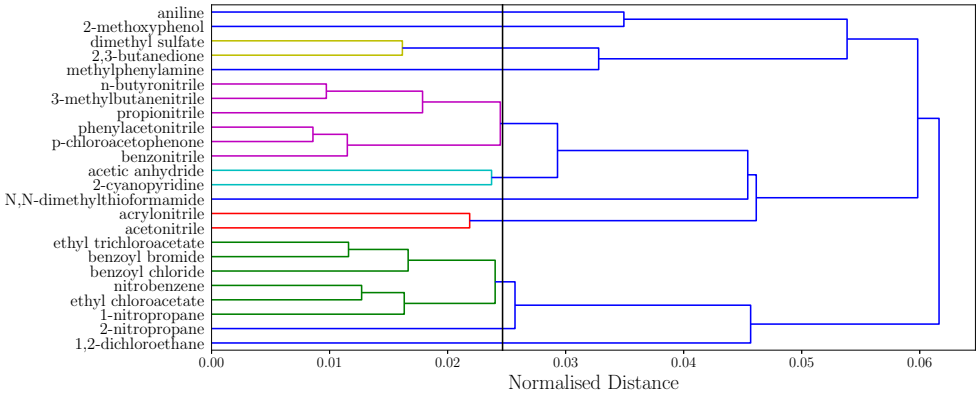


Fig. 6.5: Dendrogram for solvents to the left of node 2 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Regions 2-5 contain the most non-polar molecules. Nitriles form the dominant classification of molecules in region 2 (figure 6.5). These molecules are slightly more polar than those in regions 3-5, and contain the best hydrogen bond acceptors of any of the molecules in the super cluster formed by regions 2-5.

Aromatic compounds are found in region 3 (figure 6.6), with aromatic ethers and dialkyl sulfides. The molecules have poor hydrogen bond donors and hydrogen bond acceptors, so van der Waals interactions will be the dominant mode of interaction with solutes.

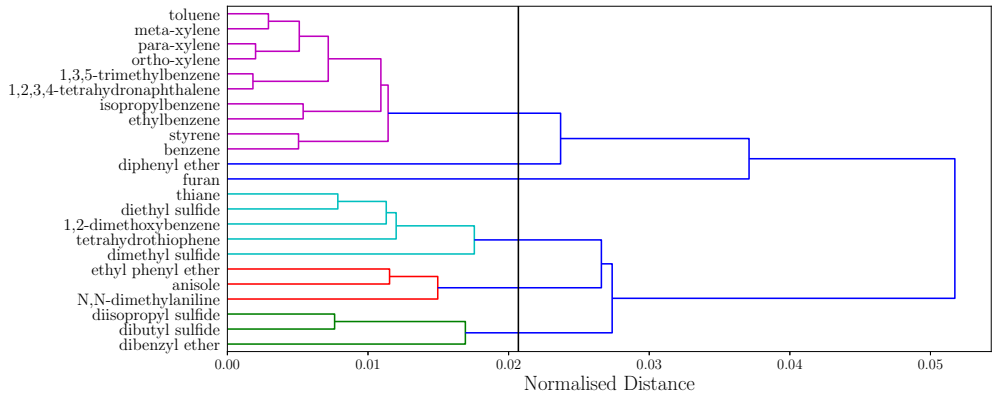


Fig. 6.6: Dendrogram for solvents to the left of node 3 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

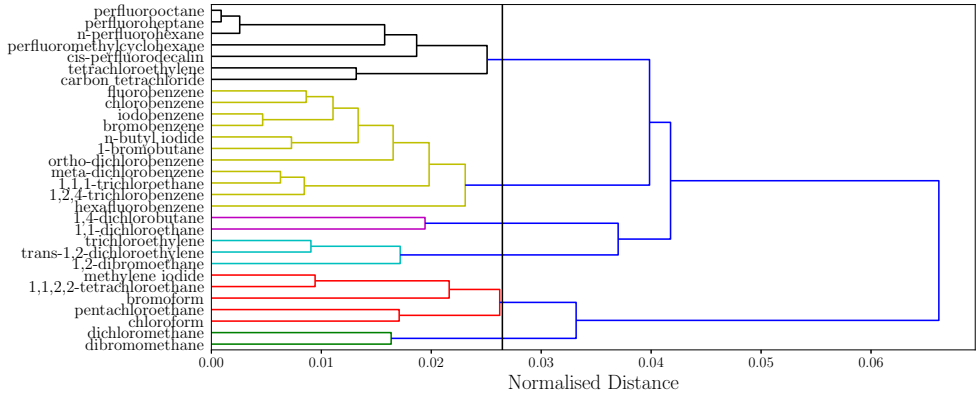


Fig. 6.7: Dendrogram for solvents to the left of node 4 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Region 4 (figure 6.7) contains halogenated molecules (haloalkanes, haloarenes and perfluorocarbons), which have poor hydrogen bond donors and also poor hydrogen bond acceptors.

Region 5 (figure 6.8) contains the least polar set of molecules, the alkanes plus tetramethylsilane and carbon disulfide. Carbon disulfide is an interesting molecule as its MEPS surface is exclusively positive, so there are no negative SSIPs. These molecules possess the weakest hydrogen bond donor groups and a minimal number of hydrogen bond acceptor sites. This results in van der Waals interactions dominating the interactions of these solvent molecules, hence giving poor solubility of polar groups.

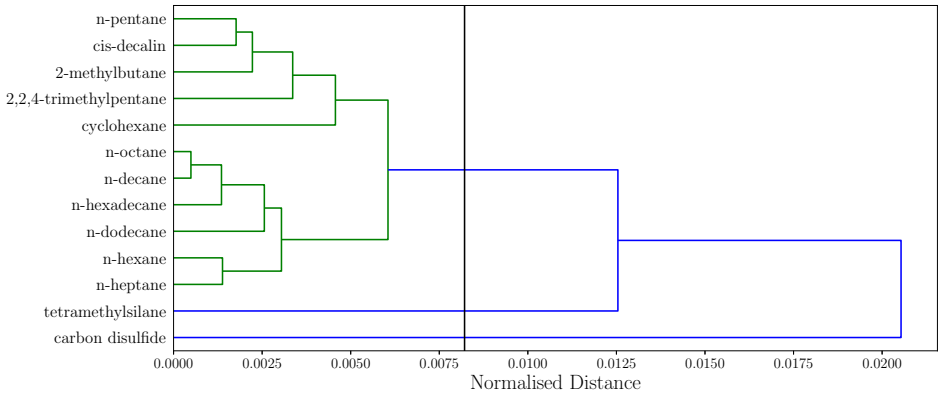


Fig. 6.8: Dendrogram for solvents to the left of node 5 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

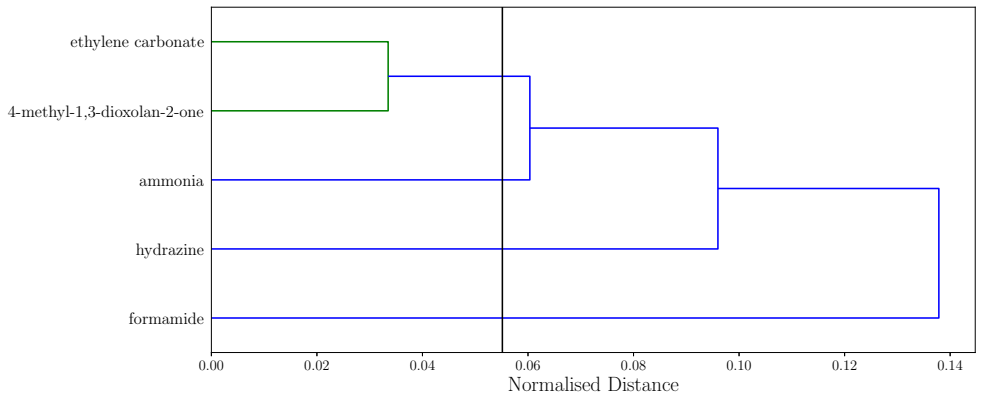


Fig. 6.9: Dendrogram for solvents to the left of node 7 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Region 6 contains only thionyl chloride, which contains moderate hydrogen bond acceptors and poor hydrogen bond donors.

Region 7 (figure 6.9) contains ethylene carbonate, 4-methyl-1,3-dioxolan-2-one, ammonia, hydrazine and formamide. These molecules contain strong hydrogen bond acceptors and weak hydrogen bond donors.

Region 8 (figure 6.10) contains molecules with strong hydrogen bond acceptors, based on pyridines ($\beta \approx -7.5$), amines ($\beta \approx -8.0$) and amides ($\beta \approx -8.3$), so therefore there is good solvation of hydrogen bond donors. Solvation of hydrogen bond acceptors is weak since the molecules contain poor hydrogen bond donors.

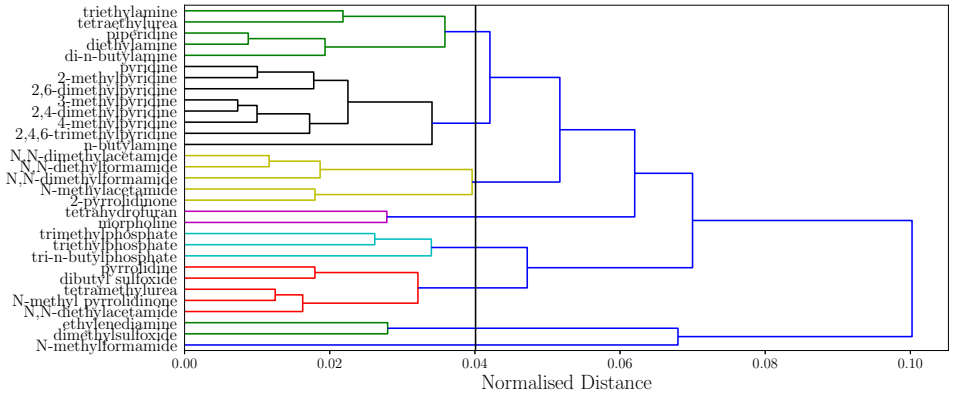


Fig. 6.10: Dendrogram for solvents to the left of node 8 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

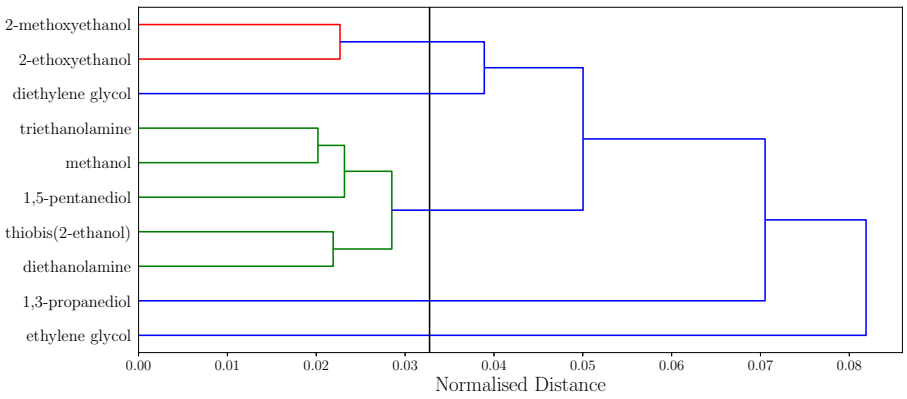


Fig. 6.11: Dendrogram for solvents to the left of node 9 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Region 9, in figure 6.11, contains substituted alcohols, which like region 10 contains molecules with at least one moderate hydrogen bond donor, and multiple moderate hydrogen bond acceptors. This allows them to solvate good hydrogen bond donors and acceptors well, due to the high proportion of polar SSIPs. Methanol is also contained in this group, instead of region 10, due to the smaller non-polar region compared to the other unsubstituted alcohols which have longer alkyl chains.

Region 10 (figure 6.12) contains carboxylic acids and alcohols (including diols). These molecules have at least one good hydrogen bond donor, plus multiple moderate hydrogen

bond acceptors. This allows for favourable interactions with solute hydrogen bond donors and acceptors of intermediate to strong ability. The small less polar regions on the molecules give more limited interactions with non-polar solute SSIPs.

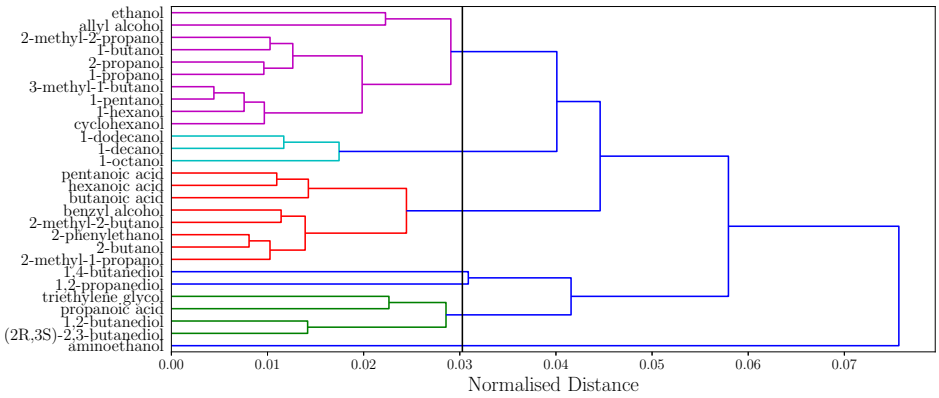


Fig. 6.12: Dendrogram for solvents to the left of node 10 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

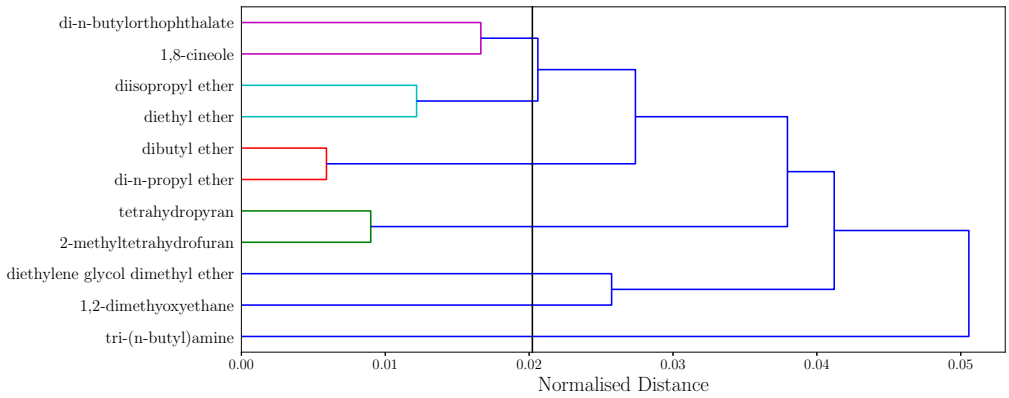


Fig. 6.13: Dendrogram for solvents to the left of node 11 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Region 11, in figure 6.13 is a collection of ethers, which have a moderate hydrogen bond acceptor ($\beta \approx 5$), but are poor hydrogen bond donors.

Region 12 (figure 6.14) contains ketones and esters.

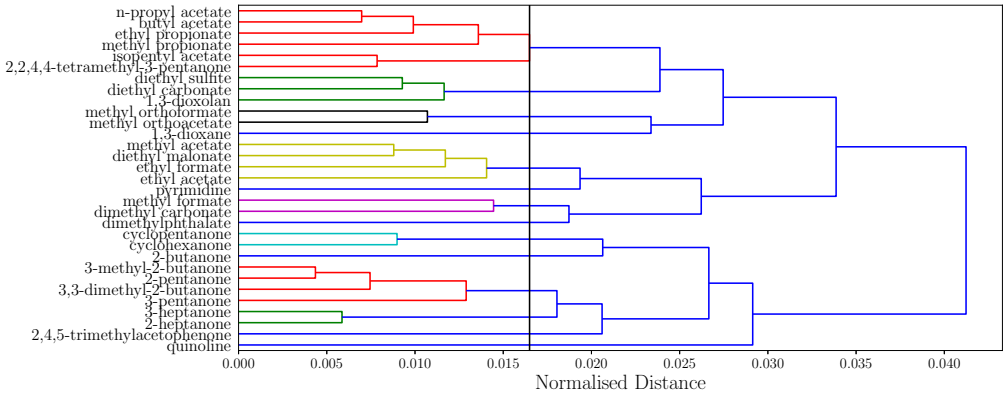


Fig. 6.14: Dendrogram for solvents to the left of node 12 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

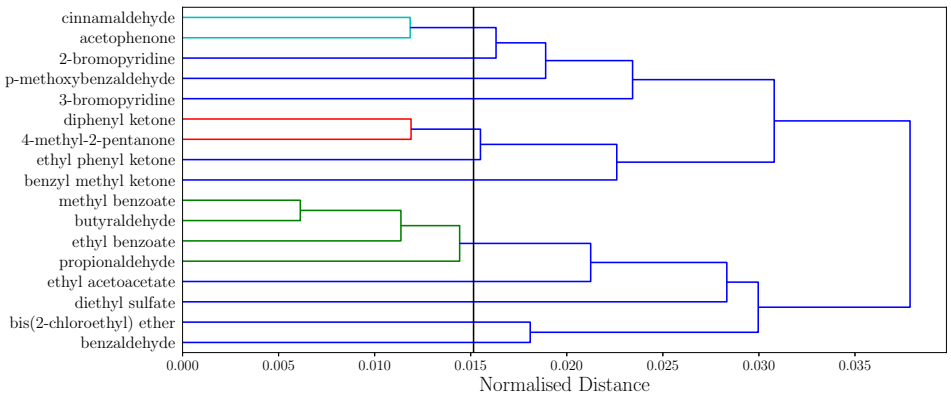


Fig. 6.15: Dendrogram for solvents to the left of node 13 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Region 13 (figure 6.15) also contains molecules with good hydrogen bond acceptors, but which lack a good donor. They are esters, aldehydes or substituted pyridines, so they can solvate hydrogen bond donors and non-polar solutes well, but not hydrogen bond acceptors, due to lack of strong positive SSIPs.

Region 14 (figure 6.16) contains an assortment of molecules with good hydrogen bond acceptors, but poor hydrogen bond donors. Regions 11-14 form a cluster of polar aprotic solvents which contain predominantly oxygen centred hydrogen bond acceptors (a few nitrogen acceptors occur in region 13).

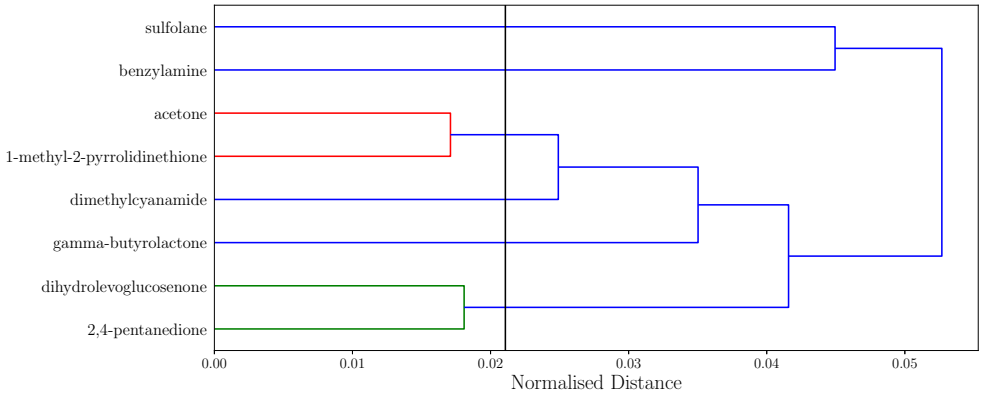


Fig. 6.16: Dendrogram for solvents to the left of node 14 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

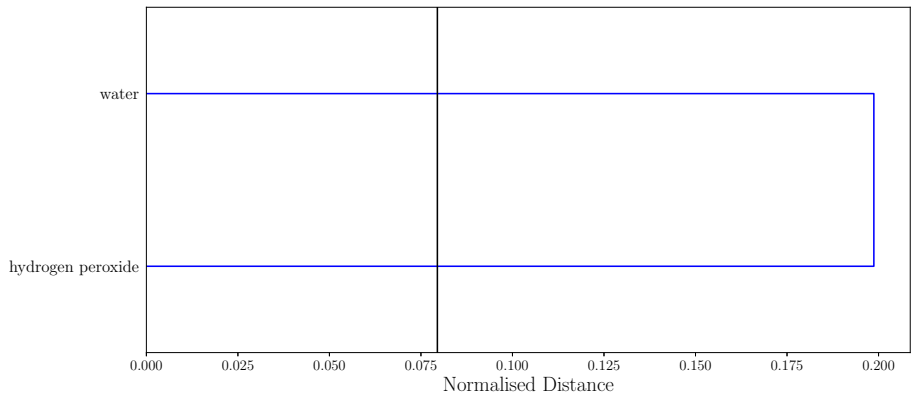


Fig. 6.17: Dendrogram for solvents to the left of node 17 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Region 15 and region 16 both contain only one molecule each: glycerol is in region 15 and HMPA (Hexamethylphosphoric triamide) is in region 16. Glycerol contains both moderate hydrogen bond donors and moderate hydrogen bond acceptors, so can solvate both hydrogen bond acceptors and hydrogen bond donors reasonably well. It also has a small non-polar region, so can interact favourably with non-polar solutes as well. HMPA has the strongest hydrogen bond acceptor, but contains only poor hydrogen bond acceptors, so solvates hydrogen bond donors very well, but strong hydrogen bond acceptors are very poorly solvated.

Region 17 in figure 6.17 contains hydrogen peroxide and water, which can solvate strong hydrogen bond donors and acceptors equally well, since both have a moderate hydrogen bond donor and acceptor groups, but are unable to solvate non-polar sites well due to the unfavourable interactions with the polar SSIPs.

Region 18 (figure 6.18) contains mainly carboxylic acids, the most polar hydrogen bond donor species, which also contain moderately good hydrogen bond acceptors. The molecules have only a small fraction of low polarity surface, so solvate non-polar molecules poorly.

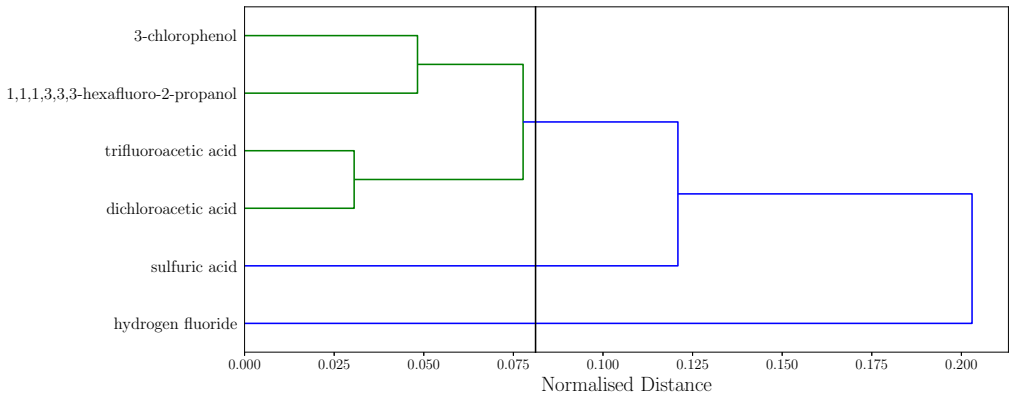


Fig. 6.18: Dendrogram for solvents to the left of node 18 in figure 6.2. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

6.5.2 Solvent Mixtures

By combining different solvents together as a mixture, a new solvent system with different properties is created. These mixtures possess properties that are a composite of the constituent parts. The similarity of the mixtures to the component solvents should exhibit a smooth change in similarity as the mixture composition is varied. This is due to the change in composition of solvent SSIPs with which a solute can interact. Competition of solute SSIP with solvent hydrogen bond acceptors and donors will vary with solvent composition. Comparison of the similarity of the solvent mixtures to the pure solvents provides a method to quantitatively measure the change in behaviour on mixing the solvents.

For these calculations the volume change of mixing was assumed to be negligible.

Water-Ethanol Mixtures

Water and ethanol appear in regions 17 and 10 of figure 6.2 respectively. Both molecules contain an hydroxyl donor and oxygen lone pair acceptors, with ethanol also containing a small alkyl chain that will solvate non-polar SSIPs. Figure 6.19 shows the dendrogram for water-ethanol mixtures in isolation. This dendrogram exhibits a homogeneous distribution of solvent mixtures.

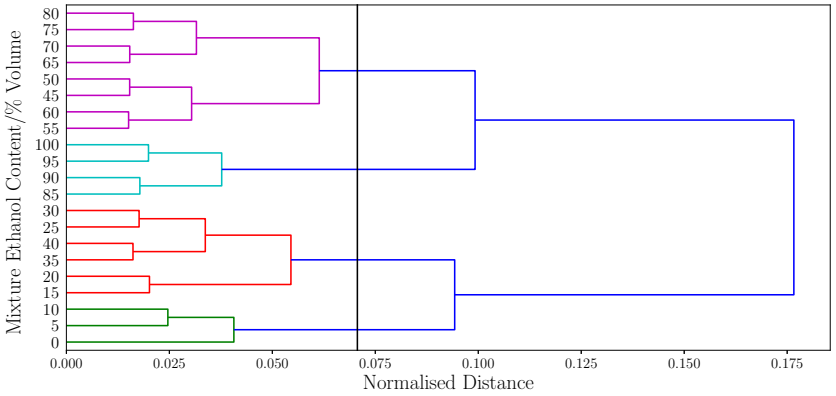


Fig. 6.19: Dendrogram showing similarity of Water-Ethanol solvent mixtures with 5% volume fraction composition differences at 298K using UPGMA clustering algorithm. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

The even distribution of the mixtures during clustering aligns with the proportion of ethanol in the mixture; the first node formed by any solvent mixture is to another solvent mixture with a composition difference of less than 10%. This behaviour is expected, as the solvent SSIP concentrations are gradually being changed, so solvation ability of the mixture should vary smoothly with the composition.

Comparison of these solvent mixtures to the set of 261 pure solvents (in appendix G) can be used to show the degree of similarity to pure solvents. The insertion points for the solvent mixtures are in table 6.2.

The mixtures fall into three major categories: $\geq$75% ethanol, 30-70% ethanol and $\leq$25% ethanol mixtures.

The mixtures with 80-95% ethanol are inserted into region 10 of the pure solvent dendrogram, with a high degree of similarity to 1,4-butanediol. The mixture with 75% ethanol is also inserted into region 10 of the pure solvent dendrogram, with a high degree of similarity

Mixture ethanol content/ % volume	Most similar pure solvent	Region	Distance to pure solvent	Distance to water	Distance to ethanol
0	water	17	0.000	0.000	0.360
5	water	17	0.028	0.028	0.332
10	water	17	0.053	0.053	0.308
15	water	17	0.075	0.075	0.286
20	water	17	0.095	0.095	0.266
25	water	17	0.114	0.114	0.247
30	formamide	7	0.112	0.131	0.230
35	formamide	7	0.098	0.148	0.213
40	formamide	7	0.084	0.164	0.197
45	formamide	7	0.073	0.180	0.181
50	formamide	7	0.063	0.195	0.166
55	formamide	7	0.056	0.210	0.151
60	formamide	7	0.053	0.225	0.136
65	formamide	7	0.056	0.240	0.121
70	formamide	7	0.064	0.256	0.105
75	1,2-propanediol	10	0.051	0.271	0.090
80	1,4-butanediol	10	0.038	0.287	0.073
85	1,4-butanediol	10	0.022	0.304	0.057
90	1,4-butanediol	10	0.011	0.322	0.039
95	1,4-butanediol	10	0.019	0.341	0.020
100	ethanol	10	0.000	0.360	0.000

Table 6.2 Water-ethanol mixture summary of most similar solvents.

to 1,2-propanediol. This is due to the increase in proportion of water in the mixture, which increases the proportion of polar hydrogen bond donors and acceptors in the solvent, with 1,2-propanediol having a larger proportion of polar hydrogen bond donors and acceptors than 1,4-butanediol.

The mixtures with 30-70% ethanol are most similar to formamide and are inserted into region 7 of the pure solvent dendrogram. The mixtures with up to 25% ethanol more closely resemble water, and are inserted into region 17.

Tetrahydrofuran-Chloroform Mixtures

Tetrahydrofuran is a moderate hydrogen bond acceptor, appearing in region 8 of figure 6.2, whereas chloroform is a moderate hydrogen bond donor, appearing in region 4 of figure 6.2.

A homogeneous distribution of solvent mixtures upon clustering are seen in figure 6.20, when only the mixtures are compared. This behaviour is expected as the composition changes lead to a gradual change in solvent SSIP composition.

The similarity of the mixtures when compared to other solvents shows a much greater dispersion than water ethanol mixtures, summarised in table 6.3.

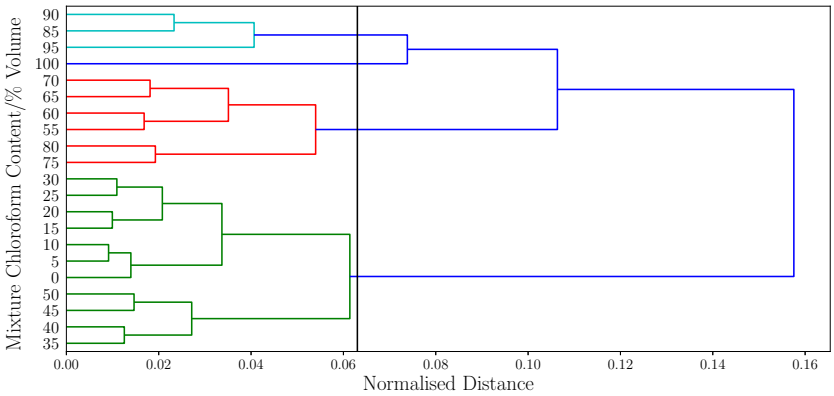


Fig. 6.20: Dendrogram showing similarity of THF-Chloroform solvent mixtures with 5% volume fraction composition differences at 298K using UPGMA clustering algorithm. Solvent clusters with a normalised distance between nodes of less than two fifths of the maximum distance, which is denoted by the vertical black line was used as a threshold to colour different clusters.

Up to 5 % volume chloroform mixtures are very similar to pure THF and after this point the mixtures transition to behave more like morpholine. Morpholine contains a moderate hydrogen bond donor group from the secondary amine, but also moderate hydrogen bond acceptors. These solvents are both in region 8 of the pure solvent dendrogram.

Once 25% volume of chloroform has been reached, the mixtures start to more closely match cyclohexanone, then quinoline at 40% volume chloroform and 3,3-dimethyl-2-butanone at 45% volume chloroform. These solvents are all in region 12 of the pure solvent dendrogram, which all have moderate hydrogen bond acceptors but fairly poor hydrogen bond donors.

50-70% volume chloroform mixtures are most similar to pure solvents in region 13 of the pure solvent dendrogram, which have moderate hydrogen bond acceptors but weak hydrogen bond donors.

As the concentration of chloroform increases (75-90% volume chloroform), the mixture solvation properties start to resemble pure solvents in region 2 (nitriles and anilines), which have weak hydrogen bond acceptors, and weak to moderate hydrogen bond donors. At 95% volume chloroform the mixture is most similar to 1,1,2-trichloroethane (region 1).

Mixture chloroform content/ % volume	Most similar pure solvent	Region	Distance to pure solvent	Distance to tetrahydro-furan	Distance to chloroform
0	tetrahydrofuran	8	0.000	0.000	0.314
5	tetrahydrofuran	8	0.010	0.010	0.306
10	morpholine	8	0.016	0.018	0.297
15	morpholine	8	0.019	0.027	0.289
20	morpholine	8	0.024	0.037	0.280
25	cyclohexanone	12	0.023	0.047	0.270
30	cyclohexanone	12	0.018	0.058	0.259
35	cyclohexanone	12	0.022	0.069	0.248
40	quinoline	12	0.024	0.082	0.237
45	3,3-dimethyl-2-butanone	12	0.024	0.095	0.224
50	benzyl methyl ke-tone	13	0.027	0.109	0.211
55	2-bromopyridine	13	0.025	0.125	0.197
60	2-bromopyridine	13	0.022	0.141	0.182
65	3-bromopyridine	13	0.030	0.159	0.167
70	3-bromopyridine	13	0.036	0.176	0.151
75	benzonitrile	2	0.037	0.194	0.134
80	phenylacetonitrile	2	0.039	0.211	0.115
85	phenylacetonitrile	2	0.046	0.228	0.095
90	aniline	2	0.048	0.246	0.073
95	1,1,2-trichloroethane	1	0.046	0.268	0.053
100	chloroform	4	0.000	0.314	0.000

Table 6.3 Tetrahydrofuran-Chloroform mixture summary of most similar solvents.

6.5.3 Replacement solvent- Dihydrolevoglucosenone

Dihydrolevoglucosenone ((1R)-7,8-Dioxabicyclo[3.2.1]octan-2-one, commercially known as Cyrene) has been suggested as a 'green' replacement for dimethyl-formamide (DMF) and N-methyl pyrrolidine (NMP) due to having similar COSMO σ-profiles [292]. Cyrene production is from cellulose, which comes from plants[293] so the feedstock required can be supplied from a sustainable source of plant biomass. At the end of its lifecycle, the solvent can be incinerated without the release of NO_x or SO_x, due to the lack of N and S hetero-atoms.

Cyrene has been shown to be a useable solvent for Sonogashira cross-coupling, Cacchi-type annulation [294] and urea synthesis [295]. Sensitivity to basic conditions were highlighted by Wilson *et al.* [294], potentially limiting which reactions it could be used as an alternative solvent for.

Since Cyrene is being used in the synthetic community as a replacement solvent it was included in the dendrogram of pure solvents (figure 6.2) and appears in Region 14 (figure 6.16).

Sherwood *et al.* [292] used the COSMO σ profile description to suggest Cyrene as a replacement for DMF and NMP. DMF and NMP both appear in region 8 (figure 6.10) of the pure solvent similarity dendrogram with the SSIMPLE solvation energy metric presented in this work.

Comparison of the solvation profiles directly using figure 6.21 shows that the solvation of solute hydrogen bond donors is much poorer in Cyrene than NMP or DMF. The solvation ability of hydrogen bond acceptors has a much smaller difference, since all have very similar hydrogen bond donors.

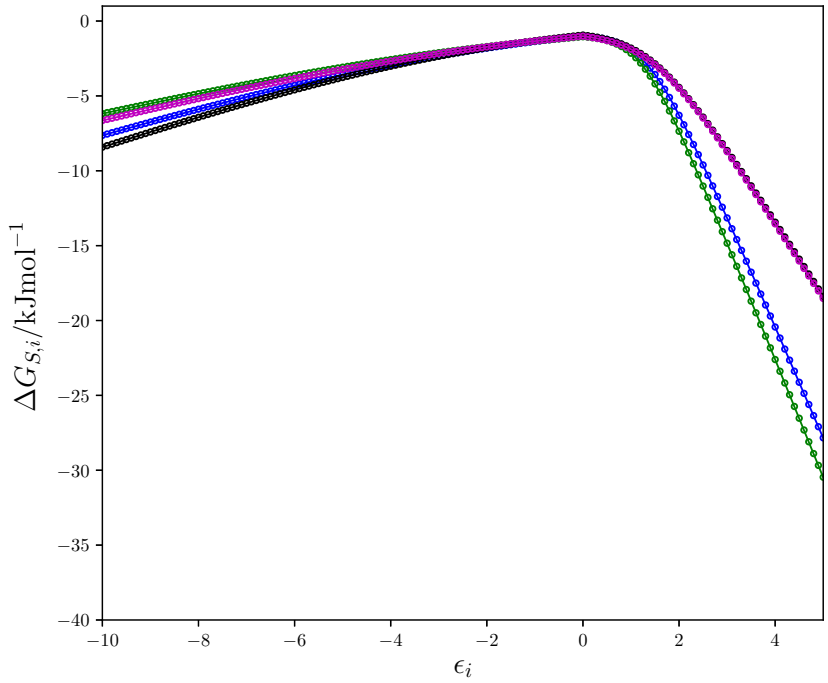


Fig. 6.21: Solvation profiles, $\Delta G_{S,i}$ (kJ mol^{-1}) against SSIP value, ε_i, for dimethyl formamide (DMF) in blue, N-methyl pyrrolidine (NMP) in green, Dihydrolevoglucosenone in black, ethyl formate in magenta.

This indicates a much greater disparity between the solvent abilities than previously described, indicating more suitable alternatives exist. Cyrene, most closely matches 2,4-pentanedione and the most common solvent that is relatively close in similarity is ethyl formate, which is illustrated in figure 6.21. The distances to cyrene for the 25 most similar solvents are displayed in table 6.4, along with the values for DMF and NMP.

Solvent	Distance to Cyrene
2,4-pentanedione	0.018
methyl formate	0.020
ethyl formate	0.024
p-methoxybenzaldehyde	0.025
1-methyl-2-pyrrolidinethione	0.027
pyrimidine	0.027
dimethylcyanamide	0.029
ethyl acetoacetate	0.029
dimethyl carbonate	0.031
cinnamaldehyde	0.031
methyl acetate	0.032
acetophenone	0.033
dimethylphthalate	0.034
2-bromopyridine	0.035
acetone	0.036
diethyl malonate	0.037
benzyl methyl ketone	0.038
4-methyl-2-pentanone	0.040
diethyl sulfite	0.041
ethyl acetate	0.041
propionaldehyde	0.041
diethyl carbonate	0.041
ethyl phenyl ketone	0.042
2-propanol	0.043
2,3-butanedione	0.043
dimethyl formamide	0.102
N-methyl pyrrolidine	0.139

Table 6.4 Distances to the 25 closest solvents to Cyrene

6.6 Conclusion

The development of a new metric for solvent similarity presented here could be a useful alternative to Hansen's solubility parameters [98]. The outcome of the clustering using this metric on the set of pure solvents aligns with the concepts of solvent type- non-polar, polar protic, polar aprotic.

Binary mixtures, such as those of THF-chloroform and water-ethanol presented here show gradual variation in solvation profile when compared to other mixtures of the same series. The behaviour of the mixtures relative to other pure solvents can be radically different to that of the components, as seen with the chloroform-tetrahydrofuran mixtures, for example a THF-chloroform 65% volume chloroform mixture most closely matches 3-bromopyridine, a much more polar solvent than either components of the mixture.

The evaluation of emerging solvents, such as Cyrene, with this metric can be used to direct experimental feasibility studies to find which conventional solvents a new green solvent might replace.

Chapter 7

Vapour-Liquid Equilibria

The study of vapour-liquid equilibria requires extensive experimental set ups [104–108]. The prediction of such properties is therefore desirable. The creation of equations of state to link the pressure, volume and temperature dependence of fluids have been in development for nearly two centuries. Clapeyron first proposed the ideal gas law [110], which was since modified by van der Waals [112] for non-ideal fluids. Further developments of more complex equations of state are covered in reviews [113–115]. Quantitative Structure Property Relationships (QSPR) have also been developed for VLE systems [93–96, 121]. Molecular simulation frameworks, such as Monte Carlo (MC) [123–126], Molecular Dynamics (MD) [127–130] and Dissipative Particle Dynamics (DPD) [131–134] have also been used to predict phase compositions.

In this work the surface site interaction model for liquids at equilibrium (SSIMPLE) [158] will be extended to consider the temperature dependence of interactions. This uses the surface site interaction point approach (SSIP) approach to molecule description [155].

7.1 Surface Site Interaction Model for Liquids at Equilibrium (SSIMPLE)

The interactions a molecule makes with the environment, e.g. the solvent, are described by a collection of surface site interaction points (SSIPs). An interaction parameter, ε_i, is assigned to each SSIP, which is equivalent to the experimentally measured hydrogen bond donor parameter (α) for positive sites or the hydrogen bond acceptor parameter ($-\beta$) for negative sites [149]. Coarse graining of the *ab initio* calculated molecular electrostatic potential surface (MEPS) of the molecule in the gas phase is used to assign these parameters [155].

All possible pairwise interactions between SSIPs in a phase are considered, with the association constant for interaction of two SSIPs, i and j, in equation (7.1).

$$K_{ij} = \frac{1}{2} e^{-\frac{\varepsilon_i \varepsilon_j + E_{vdW}}{RT}} \tag{7.1}$$

Where $E_{vdW} = -5.6$ kJ mol^{-1}; ε_i, ε_j are the values of the SSIPs; R is the gas constant and T is the temperature. E_{vdW} is the energy of interaction from van der Waals interactions between two SSIPs. The contribution to the change in free energy due to polar interactions between two SSIPs is given by $\varepsilon_i \varepsilon_j$.

The ensemble of SSIPs contained within a phase are considered to interact in a pairwise manner, such that the concentration of the bound species formed between i and j, $[i \cdot j]$ is given by (7.2).

$$[i \cdot j] = K_{ij} \left[i_{free}\right] \left[j_{free}\right] \tag{7.2}$$

Where $\left[i_{free}\right]$, $\left[j_{free}\right]$ are the concentrations of i and j that are free (not bound to any other SSIP). The speciation of a phase is calculated using the concentrations of generalized species (COGS) algorithm [279, 280]. This is detailed in chapter 4.2 and [158].

The current formulation of the SSIMPLE approach requires the total concentration of all species to be known and uses SSIP values parameterised at 298K only. Temperature-dependent changes in the structure of a liquid, which influence molecular concentrations, are governed by two distinct processes: evaporation and expansion.

Development of treatment for the temperature dependence of interactions between species and the expansion and evaporation of species in a phase are presented in this work which allow the study of vapour-liquid equilibria.

7.2 Temperature dependence of interactions

The interaction energy between two SSIPs has two components: van der Waals interactions, described by E_{vdW} and polar interactions, described by $\varepsilon_i \varepsilon_j$. Figure 7.1 depicts the formation of a hydrogen bonded bound state for two water molecules in solution. Molecular rotation within the bound state can cause the loss of directional polar interactions, whilst maintaining van der Waals interactions, the non-hydrogen bonded bound state in figure 7.1. Molecular translation can break all interactions, leading to the free state.

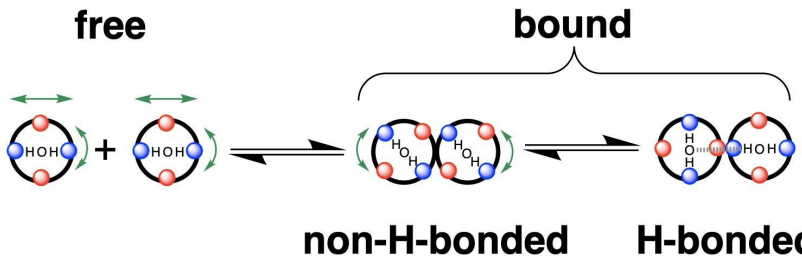


Fig. 7.1: Depiction of the three possible states of molecules in a liquid (water shown with SSIPs on surface). Free molecules have no interactions, and have no restriction to motion. Directional interactions, such as hydrogen bonding between bound molecules can be broken if there is sufficient rotational energy in the molecule whilst remaining in a bound state. The non-directional van der Waals contacts remain, maintaining formation of a non-hydrogen bonded bound complex.

7.2.1 Van der Waals interactions: E_{vdW}

Van der Waals interactions are non-directional surface interactions, that depend on surface contact area [278].

The enthalpy change of evaporation, ΔH^*, for non-polar molecules such as noble gases and alkanes can be used to explore the temperature dependence of non-polar interactions [278]. Figure 7.2 shows that the variation of ΔH^* for the non-polar molecules methane and argon is ± 0.2kJ mol^{-1}, for the entire liquid temperature range. ΔH^* is an equilibrium measurement, only obtainable at temperatures at which the liquid exists, so it cannot be measured at the same temperature for all the compounds and the data in figure 7.2 is plotted relative to the melting temperature. Figure 7.2 also shows the data for water, which is a polar molecule: $-\Delta H^*$ is not only much larger it is also strongly temperature dependent. Hydrogen bonding contributes to the large attractive interaction energies in water, but the extent of hydrogen bond formation decreases with temperature whereas the non-polar, van der Waals interactions are temperature-independent. This observation is consistent with figure 7.1 and can be used to formulate the temperature dependence of the two energy terms that contribute to K_{ij}. The van der Waals term, E_{vdW}, is temperature independent because these interactions are not affected by temperature-induced rotational motion, but the contribution due to the polar interaction term $\varepsilon_i \varepsilon_j$ depends on the relative populations of the non-hydrogen bonded and hydrogen bonded states and so is sensitive to temperature.

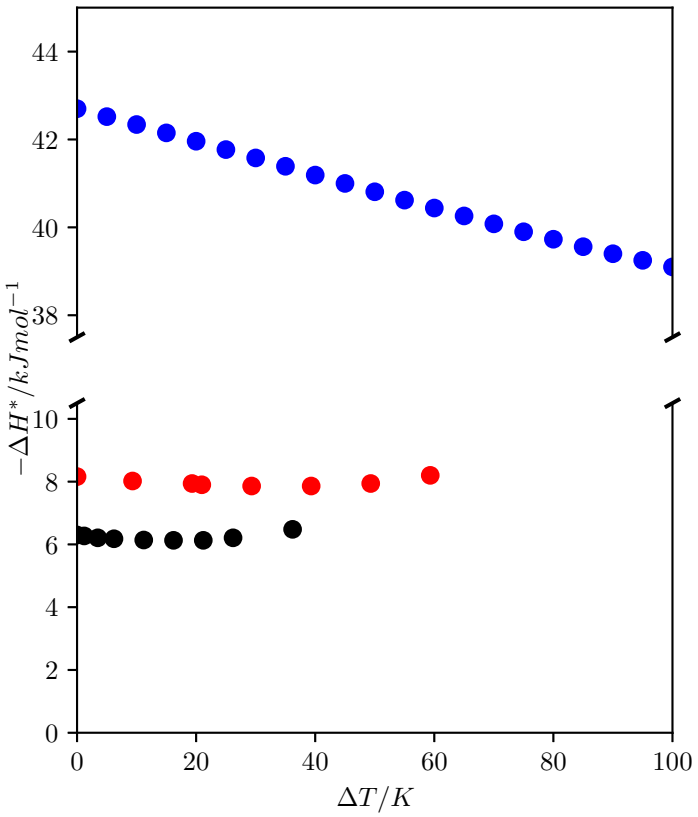


Fig. 7.2: Temperature dependence of ΔH^* for argon (black), methane (red) and water (blue). ΔT is the increase in temperature relative to the melting temperature.

7.2.2 Polar interactions: $\varepsilon_i\varepsilon_j$

The interaction parameters of the SSIPs, ε_i, were calibrated using free energies of complexation experimentally measured at 298K [149].

It is assumed that a polar interaction can populate either a hydrogen bonded state or a non-hydrogen bonded state, while two SSIPs remain in contact due to van der Waals interactions. The population of the state where the polar interaction is made is assumed to be determined by a Boltzmann distribution and this in turn determines the size of the contribution that the polar interaction makes to K_{ij} for the interaction of two SSIPs. Equation (7.3) defines the Boltzmann weighted polar interaction term at any temperature, T. Where

E_0 is the polar interaction energy of two SSIPs at zero Kelvin. The value of E_0 can be determined for any pairwise SSIP interaction, because the value of $\varepsilon_i\varepsilon_j$ at 298 K is known.

$$\left(\varepsilon_i\varepsilon_j\right)_T = \begin{cases} \frac{E_0}{1+e^{\frac{E_0}{RT}}} & \text{if } \varepsilon_{i,T}\varepsilon_{j,T} < 0 \\ 0 & \text{otherwise} \end{cases} \tag{7.3}$$

A directional Hydrogen bonded state of a repulsive polar interaction will not be significantly populated so it is assumed that unfavourable polar interactions are negligible.

7.2.3 Calculation of Experimental association constants

Association constant measurements were undertaken by Gramstad and coworkers over a range of temperatures for a series of different hydrogen bond donors and hydrogen bond acceptors [222–255]. From this large collection of data, solutes with known experimental values of hydrogen bond parameters were selected for this analysis. The ratios of the association constants of these solutes measured in carbon tetrachloride at 293 K and 323 K (the greatest temperature range studied) were used to validate the treatment of the temperature dependence of the SSIP interactions in the SSIMPLE approach (using ΔG^o in chapter 5). The calculated and experimental results are compared in figure 7.3.

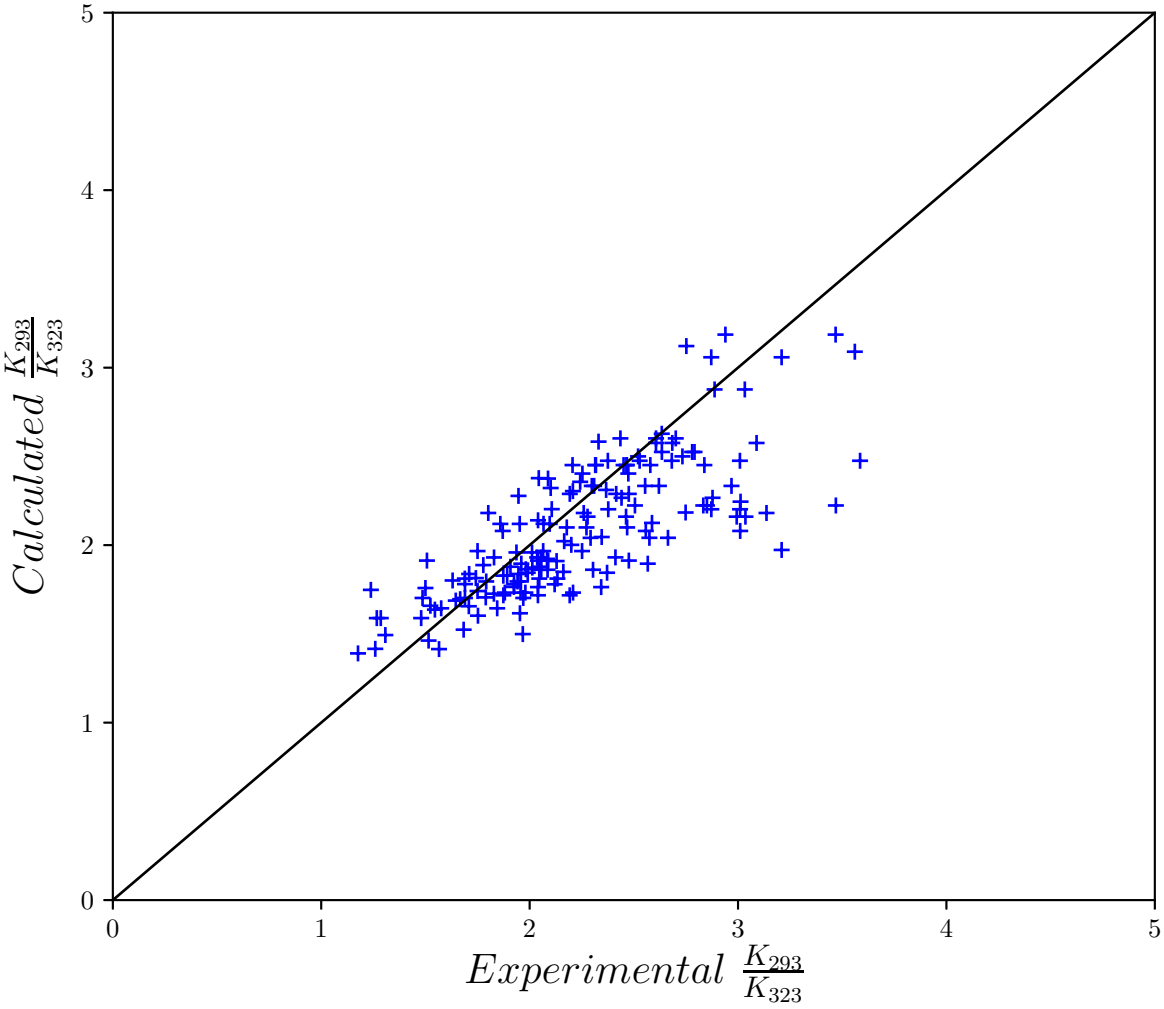


Fig. 7.3: Comparison of the ratio of association constants at two different temperatures calculated with the SSIMPLE approach with experimental measurements for 169 1:1 hydrogen bonded complexes in carbon tetrachloride.

7.3 Structural properties of the liquid

Understanding structural parameters that describe the nature of the liquid state are important for the derivation of a molecular interpretation of experimental data based on bulk physical measurements, before thermodymanic parameters are considered [97, 296–298]. Detailed

structural information for liquids is difficult to obtain compared to solids and solutes. Local information about the arrangement of neighbouring molecules is provided by X-ray scattering; the probability of finding atoms at specific separations defines a radial distribution function, which has been used to test quality of atomistic simulations of liquids [299–302]. Spectroscopy provides information on the formation of specific interactions in the liquid: for example, the extent of formation of OH···O in water [303].

The concentration of a liquid is strongly temperature dependent. As the temperature increases, the concentration decreases, with a corresponding increase in vapour pressure. The total concentration of molecules in the vapour phase and the liquid phase obeys the law of rectilinear diameter for non-polar molecules [304, 305], in equation (7.4).

$$c_T = [liquid]_T + [vapour]_T = c_0(1 - \frac{RT}{E_{exp}}) \tag{7.4}$$

Where $[liquid]_T$ is the liquid phase concentration; $[vapour]_T$ is the vapour phase concentration; E_{exp} is the expansion energy; c_0 is the hypothetical zero point concentration of the liquid and is related to molecular volume by equation (7.5) [278].

$$c_0 = \frac{r}{N_A(V_{vdW} + v)} \tag{7.5}$$

Where N_A is Avogadro's constant; $r = 0.9$, is the packing coefficient in the zero point solid for rod like molecules [278]; V_{vdW} is the molecular volume, defined by the 0.002 e bohr^{-3} electron density isosurface; $v = 5.0$Å ^{3} is the zero point void volume, from [278].

The zero point concentration, c_0, corresponds to the most concentrated state, which in practice is never reached because the liquid freezes. The lowest possible liquid concentration occurs at the critical point, when vapour and liquid concentrations are equal. Equation (7.6) defines the value of c_T for a molecule at the critical point, c_c [278].

$$c_c = \frac{r}{2N_A V_{vdW}} \tag{7.6}$$

7.4 Expansion energy

Thermal expansion results in the incorporation of additional void volume, such that the total concentration of the liquid and gas phases, c_T, will decrease. The expansion energy which governs this process, E_{exp}, is the average energy required to break interactions between

molecules. The expansion energy can be decomposed into contributions from van der Waals interactions and polar interactions, in the same way as K_{ij}.

7.4.1 Van der Waals expansion energy: E_{exp}^{vdW}

Non-polar molecules obey the law of rectilinear diameter (equation (7.4)) and can be used to examine the relationship between the expansion energy and the total van der Waals interaction energy available to a molecule in the condensed phase. At the critical temperature, T_c , the liquid and vapour concentrations are equal, such that $c_T = c_c$. Substitution of equations (7.6) and (7.5) into equation (7.4) provides a relationship between the expansion energy and critical temperature in equation (7.7).

$$E_{exp} = \frac{2RT_c}{1 - \frac{v}{V_{vdW}}} \tag{7.7}$$

Figure 7.4 shows the relationship between the values of E_{exp} calculated using experimental values of T_c and calculated values of V_{vdW} against the calculated number of SSIPs for 190 alkanes (details in appendix I). Although there is some variation, equation (7.8), plotted in black in figure 7.4, provides a reasonable description of the data.

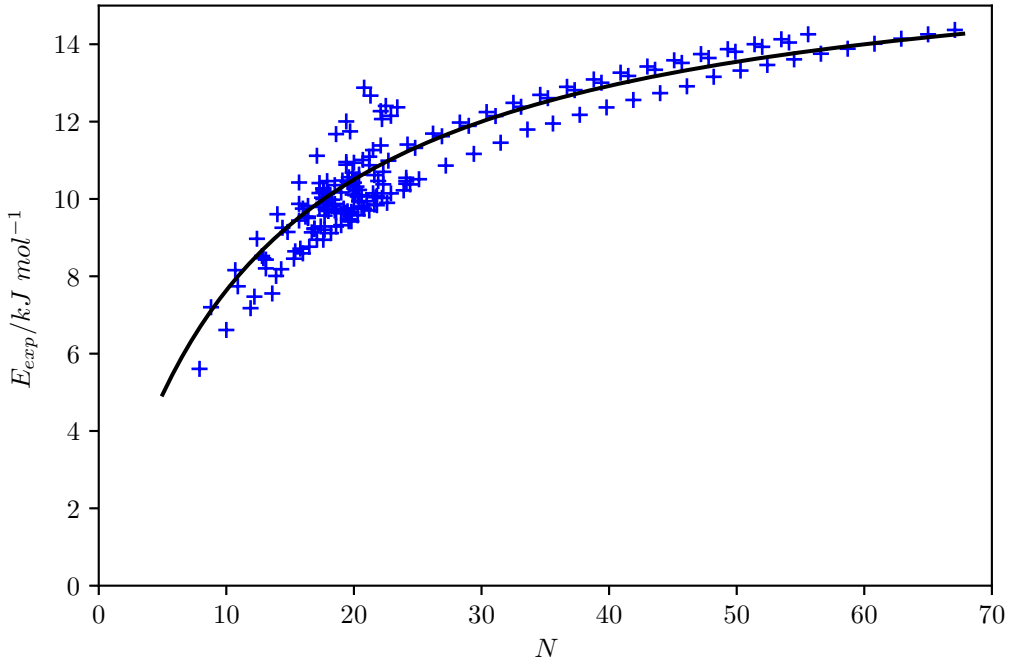


Fig. 7.4: Plot of E_{exp} against N, the number of SSIPs (blue crosses). Black line is E_{exp}^{vdW}, equation (7.8).

The contribution to the expansion energy from van der Waals interactions is in equation (7.8), where N is the number of SSIPs in the molecule. The total van der Waals interaction energy available if all SSIPs were paired is $0.5E_{vdW}N$; the energy barrier to expansion is approximately half of the total van der Waals interaction energy, but the contribution of these interactions to thermal expansion is reduced by an additional factor of $(1+\frac{N}{12})$. This reduction indicates that the cooperativity of the interactions between molecules decreases with molecular size.

$$E_{exp}^{vdW} = \frac{E_{vdW}N}{4(1+\frac{N}{12})} \tag{7.8}$$

7.4.2 Determination of the polar expansion energy contribution

In order to determine the contribution that polar interactions make to the expansion energy, the experimental temperature dependence of the concentration of water is examined. The apparent expansion energy, E_{exp}^{app}, can be computed with equation (7.9) (a rearrangement of

equation (7.4)) using the experimental values of $[liquid]_T$ and $[vapour]_T$ to calculate c_T and equation (7.5).

$$E_{exp}^{app} = \frac{RT}{1 - \frac{c_T}{c_0}} \tag{7.9}$$

Figure 7.5 shows that E_{exp}^{app} changes significantly with temperature for water; i.e. water does not obey the law for rectilinear diameter. The value of E_{exp}^{app} increases dramatically at low temperatures, where hydrogen bonded states are more highly populated. The more hydrogen bonding interactions are made, the greater the energy barrier to pulling molecules apart in the expansion process.

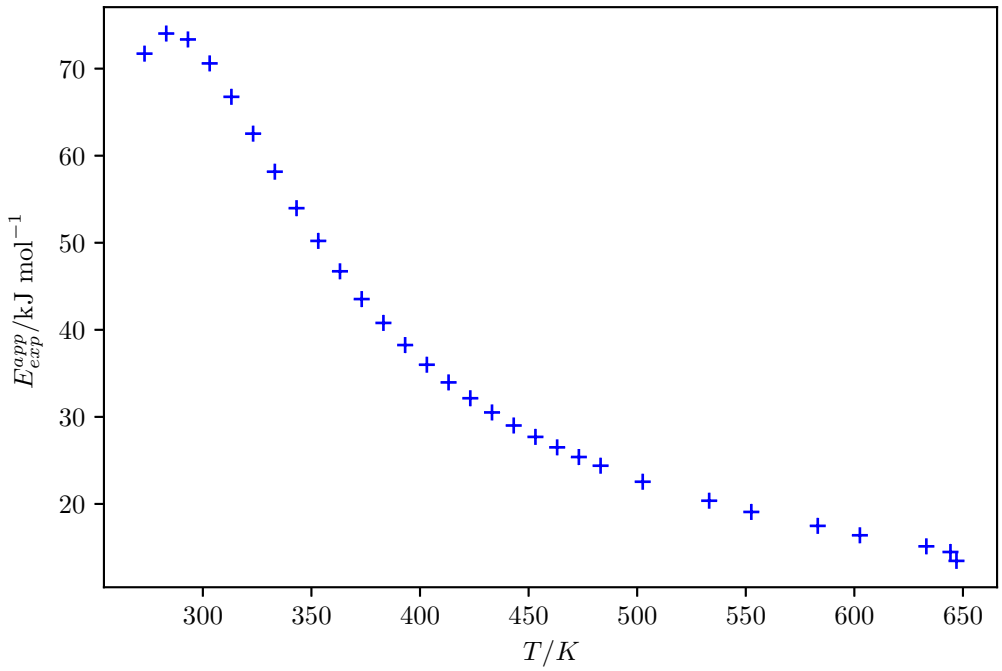


Fig. 7.5: E_{exp}^{app} against temperature for water

It is possible to calculate the extent of hydrogen bond formation in water as a function of temperature using SSIMPLE. However, the equilibrium constant for pairwise interactions between SSIPs (equation (7.1)) involves contributions from both van der Waals interactions and polar interactions. In order to determine the population of the hydrogen bonded bound state in figure 7.1, the population of the bound state calculated using SSIMPLE must be compared with the population of the non-hydrogen bounded bound state which can be

calculated using just the van der Waals term in equation (7.1) (i.e. $\varepsilon_i\varepsilon_j = 0$). For SSIPs that only make van der Waals interactions the probability that a H-bonded or a non-H-bonded state is populated would be equal because the energy difference is zero.

The extent to which polar interactions perturb these populations can therefore be calculated using equation (7.10). Where ϕ_b is the fraction of the bound state that is hydrogen bonded; ψ_f, equation (7.11), is the fraction of free SSIPs from a SSIMPLE calculation that includes polar interactions; ψ_f^{vdW} is the fraction of free SSIPs from a SSIMPLE calculation when only van der Waals interactions are made, equation (7.12).

$$\phi_b = \begin{cases} 1 - \frac{1}{2} * \frac{\psi_f}{\psi_f^{vdW}} \text{ if } \psi_f < \psi_f^{vdW} \\ \frac{1}{2} * \frac{1-\psi_f}{1-\psi_f^{vdW}} \text{ if } \psi_f > \psi_f^{vdW} \\ \frac{1}{2} \; otherwise \end{cases} \tag{7.10}$$

$$\psi_f = \frac{\sum_{SSIPs}[i_f]}{\sum_{SSIPs}[i]} \tag{7.11}$$

$$\psi_f^{VdW} = \frac{\sum_{SSIPs}[i_f]_{VdW}}{\sum_{SSIPs}[i]} \tag{7.12}$$

Where $[i]$ is the total concentration of the SSIP; $[i_f]$ is the concentration of i that is free; $[i_f]_{VdW}$ is the concentration of i that is free if only van der Waals interactions occur. The total concentration of a SSIP, $[i]$, in a phase where there are only van der Waals interactions between species is equation (7.13). K_{vdW} is the association constant in equation (7.14) (the case when $\varepsilon_i\varepsilon_j = 0$ in equation (7.1)). ψ_f^{VdW} can be written as equation (7.15), by rearrangement of equation (7.13), and noting that the probability is the same for any SSIP. θ is the fractional occupancy of the phase.

$$[i] = [i_f] + 2K_{vdW}[i_f]^2 \tag{7.13}$$

$$K_{vdW} = \frac{1}{2}e^{-\frac{E_{vdW}}{RT}} \tag{7.14}$$

$$\psi_f^{VdW} = \frac{\sqrt{1+8K_{vdW}\theta} - 1}{4K_{vdW}\theta} \tag{7.15}$$

It is possible to use the temperature dependent formulation of K_{ij} in equation (7.1) in conjunction with the experimental concentrations of the liquid and vapour phases of water to calculate the speciation in both phases as a function of temperature using SSIMPLE.

The overall fraction of bound states that are hydrogen bonded for the vapour-liquid system at temperature T is ϕ_b, in equation (7.16).

$$\phi_b = \frac{[liquid]_T \phi_b^l + [vapour]_T \phi_b^v}{[liquid]_T + [vapour]_T} \tag{7.16}$$

Figure 7.6 shows that the value of E_{exp}^{app} calculated from the experimental data for water is closely related to the fraction of the bound state that is hydrogen bonded, ϕ_b, calculated using SSIMPLE. The line of best fit through the data obtained above 313 K has a correlation coefficient (R^2) of 0.9996. The four lower temperature datapoints in figure 7.6 deviate from this line, presumably because there is a change in the structure of the liquid as it approaches the freezing point due to the geometric constraints imposed by highly H-bonded networks i.e. there is a decrease in density associated with formation of ice-like structures.

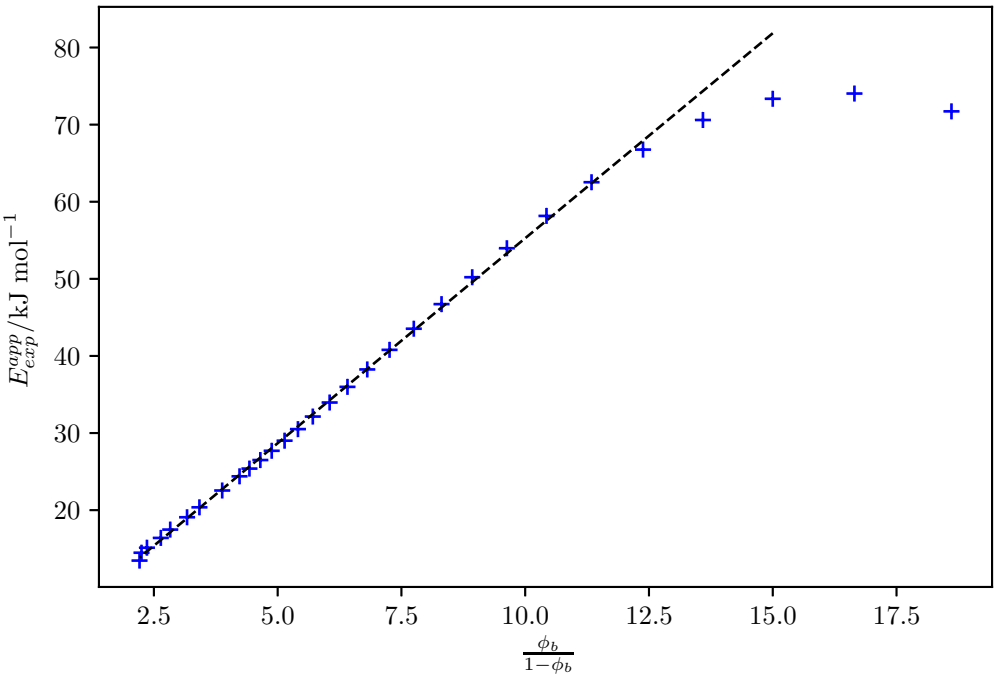


Fig. 7.6: E_{exp}^{app} against $\frac{\phi_b}{1-\phi_b}$ for water (blue crosses). Black dotted line is value of E_{exp}^{app} from a linear fit without the four lowest temperature points.

The line of best fit in figure 7.6 has a slope of 5.3 kJ mol^{-1} and an intercept of 2.1 kJ mol^{-1} . These values can be compared with the value of E_{exp}^{vdW} calculated using equation for

a molecule with four SSIPs that makes only van der Waals interactions (4.2 kJ mol^{-1}) and the total polar interaction energy for water at the zero point (twice E_0 for two water-water hydrogen bonds, which is -25.6 kJ mol^{-1}). Some of the values of E_{exp}^{app} are two to three times larger than these energies. In order to obtain an expression for E_{exp}^{app} that is consistent for both polar and non-polar molecules, it is required that if polar interactions were to make no contribution to the expansion energy of water, then ϕ_b would be 0.5 and E_{exp}^{app} would be 4.2 kJ mol^{-1}. Noting that the intercept in figure 7.6 is exactly half of 4.2 kJ mol^{-1}, these conditions are satisfied by equation (7.17).

$$\begin{aligned} E_{exp}^{app} &= \frac{1}{2}E_{exp}^{vdW} + \frac{1}{2}(E_{exp}^{vdW} + E_{exp}^{polar})\frac{\phi_b}{1-\phi_b} \\ &= \frac{E_{exp}^{vdW} + \phi_b E_{exp}^{polar}}{2(1-\phi_b)} \end{aligned} \tag{7.17}$$

Where E_{exp}^{polar} is the contribution of polar interactions to E_{exp}^{app}.

The value of E_{exp}^{polar} can therefore be determined from the slope of the line of best fit in figure 7.6, as 6.4 kJ mol^{-1}, which is half of the value of E_0 (12.8 kJ mol^{-1}) for water at the zero point. These results suggest that a suitable formulation of the expansion energy involves the sum of the van der Waals term given by equation (7.8) and the average polar interaction energy at the zero point per SSIP, in equation (7.18).

$$E_{exp}^{polar} = -\frac{\sum_{interactions} E_0}{N} \tag{7.18}$$

Equation (7.17) shows that contribution of this polar interaction term is weighted by the fraction of bound states that are hydrogen bonded, i.e the fraction of polar interactions that are actually made at any given temperature determines how much polar interaction is available to oppose expansion. The denominator in equation (7.17) can be rationalised by reformulating equation (7.4) that was used for non-polar molecules as equation (7.19) for polar molecules.

$$c_T = c_0\left(1 - \frac{2(1-\phi_b)RT}{E_{exp}}\right) \tag{7.19}$$

Where E_{exp} is in equation (7.20).

$$E_{exp} = E_{exp}^{vdW} + \phi_b E_{exp}^{polar} \tag{7.20}$$

In equation (7.19) the $2(1-\phi_b)$ term is a factor that scales the input of thermal energy that drives expansion. For molecules that make only van der Waals interactions $2(1-\phi_b)$

is equal to one, so equation 7.19 becomes identical to equation (7.4). For molecules that make polar interactions, the $2(1-\phi_b)$ term is less than one. When polar interactions are present, some of the thermal energy goes into rotational motion of the molecules that breaks directional interactions like hydrogen bonding and the rest goes into translational motion that leads to expansion. In effect, the polar interactions must be broken before expansion can take place, so the amount of thermal energy that goes into expansion is proportional to the population of non-hydrogen bonded bound states, illustrated by figure 7.1.

Expression for E_{exp}^{polar}

From equation (7.18), E_{exp}^{polar} is the negative of the mean zero point polar interaction energy per SSIP. In order to calculate E_{exp}^{polar}, we therefore require a method to determine which SSIPs interact at the zero point. The hierarchical SSIP pairing strategy described previously for cocrystal prediction [160] is used.

Interactions in the zero point liquid state are assumed to maximise the total polar interaction energy. All hydrogen bonding interactions are assumed to be independent and free to find the best partner in this approach, such that there are no steric constraints on contacts, no packing effects, and no cooperativity between sites that are close in space on the surface of the molecule. This means that the best hydrogen bond donor pairs with the best available hydrogen bond acceptor, and the second best hydrogen bond donor with the next best hydrogen bond acceptor etc. Any remaining SSIPs of the same sign that have not been paired are assumed to not form polar interactions, and therefore do not contribute.

The expansion energy from polar interactions for a molecule, E_{exp}^{polar}, is obtained by summation over all interactions as shown in equation (7.21).

$$E_{exp}^{polar} = -\frac{1}{N}\sum_{i,j}\left(\varepsilon_i\varepsilon_j\right)_{0K} \tag{7.21}$$

Where $\varepsilon_i > 0$, ordered such that $\varepsilon_i > \varepsilon_{i+1}$; $\varepsilon_j < 0$ ordered such that $\varepsilon_j < \varepsilon_{j+1}$. Repulsive interactions are assumed not to contribute to the total energy.

7.4.3 Calculation of liquid concentration

Equation (7.19), can be used to calculate the total concentration of molecules in the liquid and vapour phase through an iterative approach.

For most organic solvents at room temperature, the vapour pressure is low, and we can assume that $[liquid]_T \approx c_T$ and $[vapour]_T \approx 0$. Thus SSIMPLE can be used to calculate the concentration of the liquid phase as explained below.

An initial concentration equal to the critical concentration, c_c (equation (7.6)), was used for each SSIP. The SSIMPLE approach was used to calculate the speciation and the concentration of each SSIP was then be reevaluated, $[i]_{n,T}$, using equation (7.19).

The total concentration used for the next cycle, $[i]_{n+1}$, is shown in equation (7.22), gradually concentrates the liquid phase until the concentration converges. Convergence in concentration is reached when $[i]_{n,T} - [i]_{n-1,T}| < \delta$ where the tolerance is $\delta = 10^{-8}$.

$$[i]_{n+1,T} = \frac{[i]_{n,calc} + [i]_{n,T}}{2} \tag{7.22}$$

7.4.4 Concentration of pure liquids at 298K

Figure 7.7 shows the calculated concentrations for a set of 261 pure solvents, detailed in appendix G using the SSIMPLE approach. The major outliers, shown on Figure 7.7 are hydrogen flouride (HF), hydrogen peroxide (H_2O_2), hydrazine (N_2H_4) and ammonia (NH_3). These outliers are all small molecules where the footprinting does not perform well and the SSIP values are not very reliable.

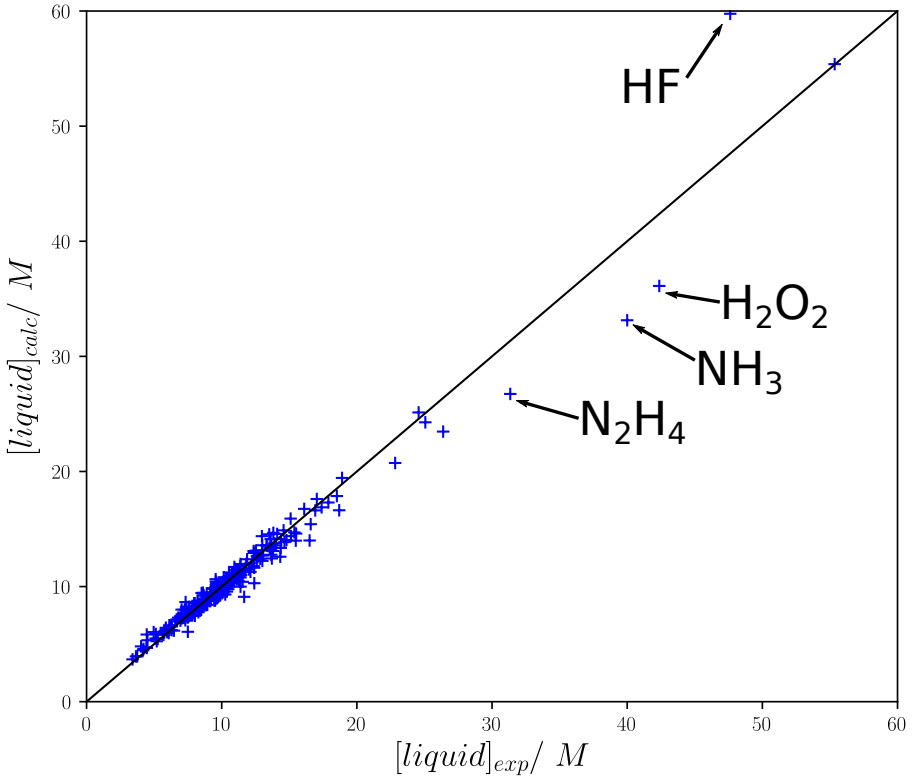


Fig. 7.7: Calculated concentration against experimental concentration for set of 261 solvents, with RMSE of 1.17M. The identities of the major outliers are indicated.

7.5 Vapour phase concentration

For higher temperatures, the approximation, $c_T \approx [liquid]_T$, does not hold. The vapour phase can be treated in exactly the same way as the liquid phase using the SSIMPLE approach, with the total concentration of SSIP i in the vapour phase, $[i]_{vapour}$ given by equation (7.23).

$$[i]_{vapour} = \frac{c_T}{2} e^{\frac{\Delta G_{S,liquid} - \Delta G_{S,vapour}}{RT}} \tag{7.23}$$

$$[i]_{liquid} = c_T - [i]_{vapour} \tag{7.24}$$

Where $\Delta G_{S,liquid}$ and $\Delta G_{S,vapour}$ are the total solvation energies for the molecule in the liquid phase and the vapour phase; $[i]_{liquid}$, equation (7.24) is the concentration of the SSIP in the liquid. The factor of two is required because at the saturated vapour pressure, the

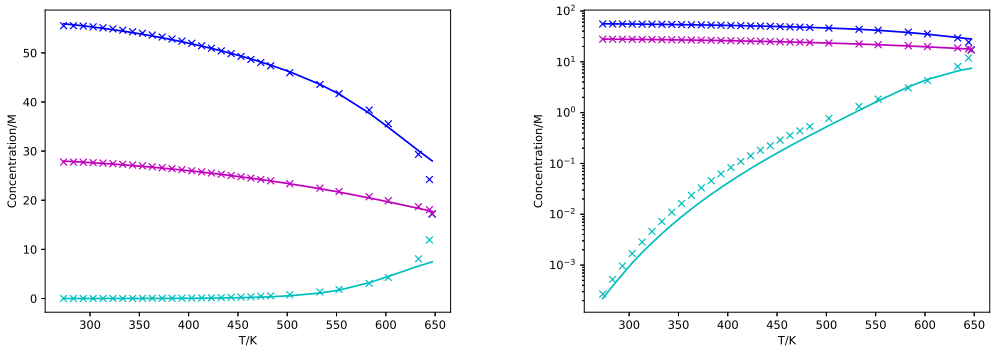
volume of the vapour plus liquid phases is double the volume of the liquid phase for a fixed number of molecules. The molecules that are not bound to the liquid phase are assumed to explore the full volume of both phases, so only half of them are in the vapour phase. This formulation satisfies the criterion that when $\Delta G_{S,liquid} = \Delta G_{S,vapour}$, the concentration in the liquid and vapour are both $\frac{1}{2}c_c$.

The initial concentration of the vapour is set equal to the tolerance, δ and the initial concentration of the liquid phase is $c_c - \delta$. The liquid is then concentrated using the procedure described above and $\Delta G_{S,liquid}$ is evaluated. Using $\Delta G_{S,liquid}$ in equation (7.23), the vapour phase concentration is calculated. SSIMPLE is then used to calculate the gas phase speciation, such that $\Delta G_{S,vapour}$ can be reevaluated, and a new vapour phase concentration calculated using equation (7.23). This new vapour phase concentration is used in the next iteration, for recalculation of the liquid concentration. This process starts by finding the maximum possible liquid concentration at a given temperature and then gradually dilutes it by transferring molecules to the gas phase until equilibrium is reached.

Iteration with equations (7.23) and (7.24) are used to reevaluate the concentrations for both phases until the amount partitioned to the vapour phase converges; i.e. $[i]_{n+1,vapour} - [i]_{n,vapour} < \delta$.

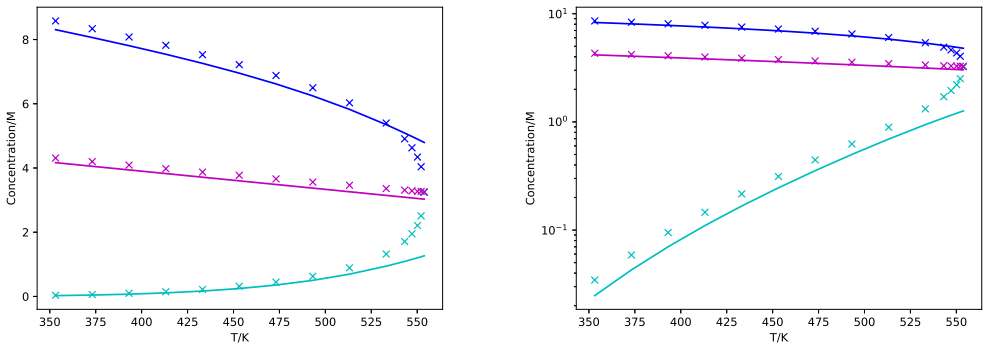
7.5.1 Calculation of vapour liquid equilibria

The equations above were used for the calculation of VLE for water in figure 7.8 and cyclohexane in figure 7.9. The prediction of the mean concentration of the two phases, $\frac{1}{2}c_T$, shows good agreement with the experimental data. At high temperatures, close to the critical point, the partitioning of species between the liquid and vapour phases becomes much poorer, with too little being deposited into the gas. This arises from the assumption made in the SSIMPLE model that molecular structures do not affect interactions: i.e. each SSIP interaction of a molecule is completely independent of any other interaction, such that the desired number of interactions can be formed. This assumptions holds well for condensed phases, but close the the critical point, where there is a large percentage of void space incorporated in the liquid, this assumption is no longer valid. This is because there are not enough nearby molecules to form an extended network of interactions required for the assumption to hold. For example water near the critical point will consist mainly of dimers due to the large amount of void space in the liquid phase, rather than existing in an extended network of hydrogen bonds, which is found in the liquid at lower temperatures.



(a) (b)

Fig. 7.8: Water concentration (linear concentration scale left, logarithmic concentration scale right) against temperature. Calculated concentrations as solid lines, experimental concentrations as crosses. $[liquid]_T$ are blue, $[vapour]_T$ are cyan, $\frac{1}{2}c_T$ are magenta.



(a) (b)

Fig. 7.9: Cyclohexane (linear concentration scale left, logarithmic concentration scale right) against temperature. Calculated concentrations as solid lines, experimental concentrations as crosses. $[liquid]_T$ are blue, $[vapour]_T$ are cyan, $\frac{1}{2}c_T$ are magenta.

7.6 Conclusions

The SSIMPLE calculation framework has been extended to be able to consider vapour liquid systems. This approach increases to the applicability of the SSIMPLE approach to solvent systems where the concentrations of solvents are not known and also under variable temperature conditions.

Chapter 8

Conclusions

The work presented in this thesis can be divided into two major areas:

- Redesign and rebuilding of the SSIP (Surface Site interaction Point) footprinting and SSIMPLE (Surface Site Interaction Model for Properties of Liquids at Equilibrium) software.
- Applications of the SSIMPLE approach.

8.1 Redesign of computational processes

Redesign of the computational processes encompasses the work on data curation, with the definition of information rich file formats for the portability of information. The rebuilding of the infrastructure to run SSIP footprinting and SSIMPLE property calculations were completed after a redesign of the computational workflow. The extension of SSIMPLE to the applications discussed in this thesis was made possible after the overhaul of the computational processes.

8.1.1 Data Curation

Definition of data formats for the storage of experimental and calculation data provides a defined framework for the interfacing of future developments. Collection and verification of experimental data into a well defined data format for storage was used in the reparameterisation of the SSIP footprinting approach. This data was also used in the evaluation of the temperature dependence of interactions using the SSIMPLE approach.

8.1.2 Computational implementation and original code refactor

The definition of clearly defined interfaces and data formats for the transfer of information provides a modular framework that was expanded to develop the applications described in the later half of this thesis. Adherence to software engineering best practices led to the creation of software modules with reproducible output along with documented application programme interfaces.

8.2 Applications of SSIMPLE

The redesigned framework was used to enable the following applications:

- benchmark performance of the SSIMPLE approach in phase transfer free energy prediction
- generate Functional Group Interaction Profiles (FGIPs)
- design a new solvent similarity metric
- extension of SSIMPLE to be able to calculate vapour-liquid equilibria (VLE)

8.2.1 Phase transfer free energy calculation benchmarking

Phase transfer free energy calculations using SSIMPLE were evaluated using two different methods for SSIP footprint generation on multiple data sets, detailed in chapter 4, showing the large domain of applicability for the approach. Both SSIP footprinting approaches, detailed in chapter 3, showed comparable accuracy in prediction of free energies of transfer and log P. Comparison of the results to existing methods show performance is worse than the leading prediction methods for log P, by around 0.5 log units.

8.2.2 Functional Group Interaction Profiles (FGIPs)

The process to generate Functional Group Interaction Profiles (FGIPs) for any solvent composition is described in chapter 5. FGIPs display the free energy of interaction between a solute hydrogen bond acceptor and solute hydrogen bond donor in a solvent as a function of solute hydrogen bond acceptor and solute hydrogen bond donor strengths. Example FGIPs for 261 single component solvent systems and 40 binary solvent mixtures at 298K are included in appendix H.

8.2.3 Solvent Similarity

A new similarity metric was developed for the assessment of solvation ability of solvents. The solvation profile, the solvation free energy of a single solute SSIP as a function of solute hydrogen bond donor/acceptor ability, of solvents are compared using a domain weighted RMSD. This approach allows a quantitative comparison of solvation ability of different solvents, including solvent mixtures.

8.2.4 Vapour-Liquid Equilibria

The law of rectilinear diameter was used to extend the SSIMPLE approach to include the expansion and evaporation processes of liquids and introduce a gas phase into calculations. The temperature dependence of interactions was also examined to extend treatment of systems with SSIMPLE to any temperature. Previously only systems at 298K, for which experimental data was available for hydrogen bond donor and hydrogen bond acceptor strengths, could be studied. With these developments the SSIMPLE approach can now be used to calculate the concentrations of species at any temperature, including partitioning of solvent into the gas phase. The calculated water and cyclohexane VLE showed agreement to experimental data within 5% error.

Bibliography

[1] C. A. Lipinski et al. “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings”. In: *Advanced Drug Delivery Reviews* 23.1-3 (Jan. 1997), pp. 3–25. DOI: 10.1016/S0169-409X(96)00423-1.

[2] E. Baka et al. “Study of equilibrium solubility measurement by saturation shake-flask method using hydrochlorothiazide as model compound”. In: *Journal of Pharmaceutical and Biomedical Analysis* 46.2 (Jan. 2008), pp. 335–341. DOI: 10.1016/j.jpba.2007.10.030.

[3] A. Avdeef and C. M. Berger. “pH-Metric Solubility. 2: Correlation Between the Acid-Base Titration and formulations for use in early animal bioavailability and toxicity studies. Later in development, solubility takes on a broader”. In: *Pharmaceutical Research* 17.1 (2000). DOI: https://doi.org/10.1023/A:1007526826979.

[4] A. Avdeef and C. M. Berger. “pH-metric solubility.: 3. Dissolution titration template method for solubility determination”. In: *European Journal of Pharmaceutical Sciences* 14.4 (2001), pp. 281–291. DOI: 10.1016/S0928-0987(01)00190-7.

[5] M. Stuart and K. Box. “Chasing Equilibrium: Measuring the Intrinsic Solubility of Weak Acids and Bases”. In: *Analytical Chemistry* 77.4 (Feb. 2005), pp. 983–990. DOI: 10.1021/ac048767n.

[6] S. M. Ulmeanu et al. “Water-oil partition profiling of ionized drug molecules using cyclic voltammetry and a 96-well microfilter plate system”. In: *Pharmaceutical Research* 20.8 (2003), pp. 1317–1322. DOI: 10.1023/A:1025025804196.

[7] A. Avdeef. “Octanol-Water Partitioning”. In: *Absorption and Drug Development*. Hoboken, NJ, USA: John Wiley & Sons, Inc., May 2012, pp. 174–219. DOI: 10.1002/9781118286067.ch4.

[8] K. Valkó. “Application of high-performance liquid chromatography based measurements of lipophilicity to model biological distribution”. In: *Journal of Chromatography A* 1037.1-2 (May 2004), pp. 299–310. DOI: 10.1016/j.chroma.2003.10.084.

[9] S. Han et al. “Determination of n-octanol/water partition coefficient for DDT-related compounds by RP-HPLC with a novel dual-point retention time correction”. In: *Chemosphere* 83.2 (Mar. 2011), pp. 131–136. DOI: 10.1016/j.chemosphere.2011.01.013.

[10] S. Han et al. “Determination of n-octanol/water partition coefficients of weak ionizable solutes by RP-HPLC with neutral model compounds”. In: *Talanta* 97 (Aug. 2012), pp. 355–361. DOI: 10.1016/j.talanta.2012.04.045.

[11] R. Mannhold et al. *Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds*. 2009. DOI: 10.1002/jps.21494.

[12] C. Nieto-Draghi et al. "A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes". In: *Chemical Reviews* 115.24 (Dec. 2015), pp. 13093–13164. DOI: 10.1021/acs.chemrev.5b00215.

[13] R. E. Skyner et al. "A review of methods for the calculation of solution free energies and the modelling of systems in solution". In: *Physical Chemistry Chemical Physics* 17.9 (2015), pp. 6174–6191. DOI: 10.1039/C5CP00288E.

[14] G. Klopman et al. "Computer Automated log P Calculations Based on an Extended Group Contribution Approach". In: *Journal of Chemical Information and Modeling* 34.4 (July 1994), pp. 752–781. DOI: 10.1021/ci00020a009.

[15] W. M. Meylan and P. H. Howard. "Atom/fragment contribution method for estimating octanol–water partition coefficients". In: *Journal of Pharmaceutical Sciences* 84.1 (Jan. 1995), pp. 83–92. DOI: 10.1002/jps.2600840120.

[16] W. M. Meylan and P. H. Howard. "Atom/fragment contribution method for estimating octanol-water partition coefficients". In: *Journal of Pharmaceutical Sciences* 84.1 (1995), pp. 83–92. DOI: 10.1002/jps.2600840120.

[17] A. Leo et al. "Calculation of hydrophobic constant (log P) from .pi. and f constants". In: *Journal of Medicinal Chemistry* 18.9 (Sept. 1975), pp. 865–868. DOI: 10.1021/jm00243a001.

[18] A. A. Petrauskas and E. A. Kolovanov. "ACD / Log P method description". In: *Most* 19.1 (2000), pp. 99–116.

[19] P. Japertas et al. "Fragmental Methods in the Design of New Compounds. Applications of The Advanced Algorithm Builder". In: *Quantitative Structure-Activity Relationships* 21.1 (2002), pp. 23–37. DOI: 10.1002/1521-3838(200205)21:1<23::AID-QSAR23>3.0.CO;2-E.

[20] P. Myrdal et al. "AQUAFAC 1: Aqueous functional group activity coefficients; application to hydrocarbons". In: *Chemosphere* 24.8 (1992), pp. 1047–1061. DOI: 10.1016/0045-6535(92)90196-X.

[21] S. Pinsuwan et al. "AQUAFAC 5: AQUeous Functional group Activity Coefficients; Application to alcohols and acids". In: *Chemosphere* 35.11 (1997), pp. 2503–2513. DOI: 10.1016/S0045-6535(97)00318-4.

[22] A. K. Ghose and G. M. Crippen. "Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity". In: *Journal of Computational Chemistry* 7.4 (Aug. 1986), pp. 565–577. DOI: 10.1002/jcc.540070419.

[23] Daylight Chemical Information Systems Inc. *PCModels*. URL: https://www.daylight.com/products/pcmodels.html (visited on 05/18/2019).

[24] M. Waldherr-Teschner et al. "MOLCAD — Computer Aided Visualization and Manipulation of Models in Molecular Science". In: *Advances in Scientific Visualization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 58–67. DOI: 10.1007/978-3-642-77334-1_5.

[25] Oxford Molecular Group PLC. *TSAR 3.2*. URL: https://www.drugdiscoveryonline.com/doc/structure-activity-software-0001 (visited on 05/18/2019).

[26] CompuDrug Ltd. *PROLOGP*. URL: https://www.compudrug.com/prologp (visited on 05/18/2019).

[27] S. Wildman and G. Crippen. "Prediction of Physicochemical Parameters by Atomic Contributions". In: *Journal of Chemical Information and Modeling* 39 (1999), pp. 868–873. DOI: 10.1021/ci990307l.

[28] R. Wang et al. "A New Atom-Additive Method for Calculating Partition Coefficients". In: *Journal of Chemical Information and Computer Sciences* 2338.96 (1997), pp. 615–621. DOI: 10.1021/ci960169p.

[29] I. Moriguchi et al. "Simple method of calculating octanol/water partition coefficient". In: *Chem. Pharm. Bull.* 40 (1992), pp. 127–130.

[30] M. Junghans and E. Pretsch. "Estimation of partition coefficients of organic compounds: local database modeling with uniform-length structure descriptors". In: *Fresenius' Journal of Analytical Chemistry* 359.1 (1997), pp. 88–92. DOI: 10.1007/s002160050541.

[31] L. H. Hall and L. B. Kier. "Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information". In: *Journal of Chemical Information and Modeling* 35.6 (1995), pp. 1039–1045. DOI: 10.1021/ci00028a014.

[32] V. K. Gombar. "Reliable assessment of logP of compounds of pharmaceutical relevance." In: *SAR and QSAR in environmental research* 10.4 (1999), pp. 371–380. DOI: 10.1080/10629369908039105.

[33] J. Huuskonen et al. "Neural network modeling for estimation of partition coefficient based on atom-type electrotopological state indices". In: *Journal of chemical information and computer sciences* 40.4 (2000), pp. 947–55. DOI: 10.1021/ci9904261.

[34] M. H. Abraham. "Hydrogen bonding. 31. Construction of a scale of solute effective or summation hydrogen-bond basicity". In: *Journal of Physical Organic Chemistry* 6.12 (Dec. 1993), pp. 660–684. DOI: 10.1002/poc.610061204.

[35] M. H. Abraham et al. "Hydrogen bonding. 39. The partition of solutes between water and various alcohols". In: *Journal of Physical Organic Chemistry* 7.12 (Dec. 1994), pp. 712–716. DOI: 10.1002/poc.610071209.

[36] *ACD/Absolv, accessed June 2016*. URL: http://www.acdlabs.com/products/percepta/predictors/absolv/details.php.

[37] *UFZ Department of Ecological Chemistry 2015. ChemProp 6.3*. URL: http://www.ufz.de/ecochem/chemprop.

[38] I. V. Tetko et al. "Prediction of n -Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices". In: *Journal of Chemical Information and Computer Sciences* 41.5 (Sept. 2001), pp. 1407–1421. DOI: 10.1021/ci010368v.

[39] D. Eros et al. "Reliability of logP Predictions Based on Calculated Molecular Descriptors: A Critical Review". In: *Current Medicinal Chemistry* 9.20 (Oct. 2002), pp. 1819–1829. DOI: 10.2174/0929867023369042.

[40] R. Liu and D. Zhou. "Using molecular fingerprint as descriptors in the QSPR study of lipophilicity". In: *Journal of Chemical Information and Modeling* 48.3 (2008), pp. 542–549. DOI: 10.1021/ci700372s.

[41] H.-F. Chen. "In silico log P prediction for a large data set with support vector machines, radial basis neural networks and multiple linear regression." In: *Chemical biology & drug design* 74.2 (2009), pp. 142–7. DOI: 10.1111/j.1747-0285.2009.00840.x.

[42] I. V. Tetko et al. "Critical assessment of QSAR models of environmental toxicity against tetrahymena pyriformis: Focusing on applicability domain and overfitting by variable selection". In: *Journal of Chemical Information and Modeling* 48.9 (2008), pp. 1733–1746. DOI: 10.1021/ci800151m.

[43] R. A. Friesner. "Ab initio quantum chemistry: Methodology and applications". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.19 (2005), pp. 6648–6653. DOI: 10.1073/pnas.0408036102.

[44] J. Tomasi et al. "Quantum Mechanical Continuum Solvation Models". In: *Chemical Reviews* 105.8 (Aug. 2005), pp. 2999–3094. DOI: 10.1021/cr9904009.

[45] A. V. Marenich et al. "Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges". In: *Journal of Chemical Theory and Computation* 3 (2007), pp. 2011–2033. DOI: 10.1021/ct7001418.

[46] C. J. Cramer and D. G. Truhlar. "ChemInform Abstract: A Universal Approach to Solvation Modeling". In: *ChemInform* 39.35 (2008). DOI: 10.1002/chin.200835275.

[47] A. Klamt and G. Schüürmann. "COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient". In: *J. Chem. Soc., Perkin Trans. 2* 5 (1993), pp. 799–805. DOI: 10.1039/P29930000799.

[48] A. Klamt. "Conductor-like Screening Model for Real Solvents : A New Approach to the Quantitative Calculation of Solvation Phenomena". In: (1995), pp. 2224–2235. DOI: 10.1021/j100007a062.

[49] A. Klamt et al. "Refinement and Parametrization of COSMO-RS". In: *The Journal of Physical Chemistry A* 102.26 (1998), pp. 5074–5085. DOI: 10.1021/jp980017s.

[50] A. Klamt and F. Eckert. "COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids". In: *Fluid Phase Equilibria* 172.1 (July 2000), pp. 43–72. DOI: 10.1016/S0378-3812(00)00357-5.

[51] S.-T. Lin and S. I. Sandler. "A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model". In: (2002), pp. 899–913. DOI: 10.1021/ie001047w.

[52] S.-T. Lin and S. I. Sandler. "A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model." In: *Industrial & Engineering Chemistry Research* 43.5 (Mar. 2004), pp. 1322–1322. DOI: 10.1021/ie0308689.

[53] A. Klamt and F. Eckert. "Prediction of vapor liquid equilibria using COSMOtherm". In: 217.March 2003 (2004), pp. 53–57. DOI: 10.1016/j.fluid.2003.08.018.

[54] J. Palomar et al. "Density and Molar Volume Predictions Using COSMO-RS for Ionic Liquids. An Approach to Solvent Design". In: *Industrial & Engineering Chemistry Research* 46.18 (Aug. 2007), pp. 6041–6048. DOI: 10.1021/ie070445x.

[55] M. Hornig and A. Klamt. “COSMOfrag: A novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry”. In: *Journal of Chemical Information and Modeling* 45 (2005), pp. 1169–1177. DOI: 10.1021/ci0501948.

[56] C. F. Kettering et al. “A Representation of the Dynamic Properties of Molecules by Mechanical Models”. In: *Phys. Rev.* 36 (3 Aug. 1930), pp. 531–543. DOI: 10.1103/PhysRev.36.531.

[57] F. H. Westheimer and Martin W. Shookhoff. “The Electrostatic Influence of Substituents on Reactions Rates. I”. In: *Journal of the American Chemical Society* 62.2 (1940), pp. 269–275. DOI: 10.1021/ja01859a009.

[58] D. L. Beveridge and F. M. DiCapua. “Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems”. In: *Annual Review of Biophysics and Biophysical Chemistry* 18.1 (June 1989), pp. 431–492. DOI: 10.1146/annurev.bb.18.060189.002243.

[59] J. G. Kirkwood. “Statistical Mechanics of Fluid Mixtures”. In: *The Journal of Chemical Physics* 3.5 (1935), p. 300. DOI: 10.1063/1.1749657.

[60] R. W. Zwanzig. “HighTemperature Equation of State by a Perturbation Method. I. Nonpolar Gases”. In: *The Journal of Chemical Physics* 22.8 (Aug. 1954), pp. 1420–1426. DOI: 10.1063/1.1740409.

[61] N. M. Garrido et al. “1-Octanol/Water Partition Coefficients of n -Alkanes from Molecular Simulations of Absolute Solvation Free Energies”. In: *Journal of Chemical Theory and Computation* 5.9 (Sept. 2009), pp. 2436–2446. DOI: 10.1021/ct900214y.

[62] N. M. Garrido et al. “Prediction of the n -hexane/water and 1-octanol/water partition coefficients for environmentally relevant compounds using molecular simulation”. In: *AIChE Journal* 58.6 (June 2012), pp. 1929–1938. DOI: 10.1002/aic.12718. arXiv: 1402.6991.

[63] D. L. Mobley and J. P. Guthrie. “FreeSolv: A database of experimental and calculated hydration free energies, with input files”. In: *Journal of Computer-Aided Molecular Design* 28.7 (2014), pp. 711–720. DOI: 10.1007/s10822-014-9747-x.

[64] O. Beckstein et al. “Prediction of hydration free energies for the SAMPL4 diverse set of compounds using molecular dynamics simulations with the OPLS-AA force field”. In: *Journal of Computer-Aided Molecular Design* 28.3 (Mar. 2014), pp. 265–276. DOI: 10.1007/s10822-014-9727-1.

[65] G. Duarte Ramos Matos et al. “Infinite Dilution Activity Coefficients as Constraints for Force Field Parametrization and Method Development”. In: *Journal of Chemical Theory and Computation* 15.5 (May 2019), pp. 3066–3074. DOI: 10.1021/acs.jctc.8b01029.

[66] P. A. Kollman et al. “Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models”. In: *Accounts of Chemical Research* 33.12 (Dec. 2000), pp. 889–897. DOI: 10.1021/ar000033j.

[67] B. R. Miller et al. “MMPBSA.py : An Efficient Program for End-State Free Energy Calculations”. In: *Journal of Chemical Theory and Computation* 8.9 (Sept. 2012), pp. 3314–3321. DOI: 10.1021/ct300418h.

[68] S. Genheden and U. Ryde. "How to obtain statistically converged MM/GBSA results". In: *Journal of Computational Chemistry* 31.16 (2009), NA–NA. DOI: 10.1002/jcc.21366.

[69] J. R. Koza et al. "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming". In: *Artificial Intelligence in Design '96*. Dordrecht: Springer Netherlands, 1996, pp. 151–170. DOI: 10.1007/978-94-009-0279-4_9.

[70] C. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006, pp. XX, 738.

[71] A. Lusci et al. "Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules". In: *Journal of Chemical Information and Modeling* 53.7 (2013), pp. 1563–1575. DOI: 10.1021/ci400187y.

[72] N. S. H. Narayana Moorthy et al. "Classification study of solvation free energies of organic molecules using machine learning techniques". In: *RSC Advances* 4.106 (2014), pp. 61624–61630. DOI: 10.1039/c4ra07961b.

[73] S. Riniker. "Molecular Dynamics Fingerprints (MDFP): Machine Learning from MD Data to Predict Free-Energy Differences". In: *Journal of Chemical Information and Modeling* 57.4 (2017), pp. 726–741. DOI: 10.1021/acs.jcim.6b00778.

[74] R. Zubatyuk et al. "Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network". In: *Science Advances* 5.8 (Aug. 2019), eaav6490. DOI: 10.1126/sciadv.aav6490.

[75] J. S. Smith et al. "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning". In: *Nature Communications* 10.1 (2019), pp. 1–8. DOI: 10.1038/s41467-019-10827-4.

[76] C. A. Bauer et al. "Machine learning models for hydrogen bond donor and acceptor strengths using large and diverse training data generated by first-principles interaction free energies". In: *Journal of Cheminformatics* 11.1 (2019), pp. 1–16. DOI: 10.1186/s13321-019-0381-4.

[77] P.T. Anastas and J.C. Warner. *Green Chemistry: Theory and Practice*. Oxford University Press, 2000. URL: https://books.google.co.uk/books?id=%5C_iMORRU42isC.

[78] M. Tobiszewski et al. "Environmental risk-based ranking of solvents using the combination of a multimedia model and multi-criteria decision analysis". In: *Green Chemistry* 19.4 (2017), pp. 1034–1042. DOI: 10.1039/C6GC03424A.

[79] R. K. Henderson et al. "Expanding GSK's solvent selection guide – embedding sustainability into solvent selection starting at medicinal chemistry". In: *Green Chemistry* 13.4 (2011), p. 854. DOI: 10.1039/c0gc00918k.

[80] K. Alfonsi et al. "Green chemistry tools to influence a medicinal chemistry and research chemistry based organisation". In: *Green Chem.* 10.1 (2008), pp. 31–36. DOI: 10.1039/B711717E.

[81] C. Jiménez-González et al. "Using the right green yardstick: Why process mass intensity is used in the pharmaceutical industry to drive more sustainable processes". In: *Organic Process Research and Development* 15.4 (2011), pp. 912–917. DOI: 10.1021/op200097d.

[82] S. Hellweg et al. “Environmental assessment of chemicals: Methods and application to a case study of organic solvents”. In: *Green Chemistry* 6.8 (2004), pp. 418–427. DOI: 10.1039/b402807b.

[83] A. Amelio et al. “Guidelines based on life cycle assessment for solvent selection during the process design and evaluation of treatment alternatives”. In: *Green Chemistry* 16.6 (2014), pp. 3045–3063. DOI: 10.1039/c3gc42513d.

[84] J. F. Martínez-Gallegos et al. “Dihydroxyacetone crystallization: Process, environmental, health and safety criteria application for solvent selection”. In: *Chemical Engineering Science* 134 (2015), pp. 36–43. DOI: 10.1016/j.ces.2015.04.047.

[85] A. Bender and R. C. Glen. “Molecular Similarity : A Key Technique in Molecular similarity : a key technique in molecular informatics”. In: February (2016), pp. 3204–3218. DOI: 10.1039/B409813G.

[86] P. Willett et al. “Chemical similarity searching”. In: *Journal of Chemical Information anddoi: 10.0.3.253/ci800151m Computer Sciences* 38.6 (1998), pp. 983–996. DOI: 10.1021/ci9800211.

[87] P. Willett. “The calculation of molecular structural similarity: Principles and practice”. In: *Molecular Informatics* 33.6-7 (2014), pp. 403–413. DOI: 10.1002/minf.201400024.

[88] J. D. Maccuish and N. E. Maccuish. “Chemoinformatics applications of cluster analysis”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 4.1 (2014), pp. 34–48. DOI: 10.1002/wcms.1152.

[89] P. W. Finn and G. M. Morris. “Shape-based similarity searching in chemical databases”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3.3 (2013), pp. 226–241. DOI: 10.1002/wcms.1128.

[90] T. S. Rush et al. “A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein-Protein Interaction”. In: *Journal of Medicinal Chemistry* 48.5 (Mar. 2005), pp. 1489–1495. DOI: 10.1021/jm040163o.

[91] G. Klopman. “MULTICASE 1. A Hierarchical Computer Automated Structure Evaluation Program”. In: *Quantitative Structure-Activity Relationships* 11.2 (1992), pp. 176–184. DOI: 10.1002/qsar.19920110208.

[92] A. Fredenslund et al. *Vapor-Liquid Equilibria Using Unifac A Group-Contribution Method*. 1st. Elsevier, 1977, p. 392.

[93] S. Skjold-Jorgensen et al. “Vapor-Liquid Equilibria by UNIFAC Group Contribution. Revision and Extension”. In: *Industrial and Engineering Chemistry Process Design and Development* 18.4 (1979), pp. 714–722. DOI: 10.1021/i260072a024.

[94] T. Holderbaum and J. Gmehling. “PSRK: A Group Contribution Equation of State Based on UNIFAC”. In: *Fluid Phase Equilibria* 70.2-3 (1991), pp. 251–265. DOI: 10.1016/0378-3812(91)85038-V.

[95] J. Gmehling et al. “A Modified UNIFAC Model. 2. Present Parameter Matrix and Results for Different Thermodynamic Properties”. In: *Industrial and Engineering Chemistry Research* 32.1 (1993), pp. 178–193. DOI: 10.1021/ie00013a024.

[96] R. Wittig et al. “Vapor-liquid equilibria by UNIFAC group contribution. 6. Revision and extension”. In: *Industrial and Engineering Chemistry Research* 42.1 (2003), pp. 183–188. DOI: 10.1021/ie020506l.

[97] J. H. Hildebrand and R. L. Scott. *Solubility of non-electrolytes*. Reinhold Pub, 1936.

[98] C. M. Hansen. *Hansen Solubility Parameters: A User's Handbook, Second Edition*. 2nd ed. CRC press, 2007. URL: https://www.crcpress.com/Hansen-Solubility-Parameters-A-Users-Handbook-Second-Edition/Hansen/p/book/9780849372483.

[99] M. Belmares et al. "Hildebrand and Hansen solubility parameters from Molecular Dynamics with applications to electronic nose polymer sensors". In: *Journal of Computational Chemistry* 25.15 (Nov. 2004), pp. 1814–1826. DOI: 10.1002/jcc.20098.

[100] T. Lindvig et al. "A Flory–Huggins model based on the Hansen solubility parameters". In: *Fluid Phase Equilibria* 203.1-2 (Dec. 2002), pp. 247–260. DOI: 10.1016/S0378-3812(02)00184-X.

[101] Y. Zhao et al. "Solvent similarity analysis-from qualitative to quantitative". In: *Chemical Engineering Science* 184 (2018), pp. 149–160. DOI: 10.1016/j.ces.2018.03.034.

[102] R. H. Perry. *Perry's chemical engineers' handbook.* eng. 7th ed. / prepared by a staff of specialists under the editorial direction of late editor Robert H. Perry ; editor, Don W. Green ; associate editor, James O. Maloney. New York [N.Y.] ; London: McGraw-Hill, 1997.

[103] D. A. McQuarrie and J. D. Simon. *Physical Chemistry: A Molecular Approach.* University Science Books., 1997. Chap. 24.

[104] B. Kolbe and J. Gmehling. "Thermodynamic properties of ethanol + water. I. Vapour-liquid equilibria measurements from 90 to 150°C by the static method". In: *Fluid Phase Equilibria* 23.2-3 (1985), pp. 213–226. DOI: 10.1016/0378-3812(85)90007-X.

[105] M. M. Abbott. "Low-pressure phase equilibria: Measurement of VLE". In: *Fluid Phase Equilibria* 29.C (1986), pp. 193–207. DOI: 10.1016/0378-3812(86)85021-X.

[106] T. Laursen and S. I. Andersen. "VLE and VLLE measurements of dimethyl ether containing systems". In: *Journal of Chemical and Engineering Data* 48.5 (2003), pp. 1085–1087. DOI: 10.1021/je025599s.

[107] S. Nebig et al. "Measurement of vapor-liquid equilibria (VLE) and excess enthalpies (HE) of binary systems with 1-alkyl-3-methylimidazolium bis(trifluoromethyl sulfonyl)imide and prediction of these properties and γ_∞ using modified UNIFAC (Dortmund)". In: *Fluid Phase Equilibria* 258.2 (2007), pp. 168–178. DOI: 10.1016/j.fluid.2007.06.001.

[108] R. Kato and J. Gmehling. "Systems with ionic liquids: Measurement of VLE and $\gamma\infty$ data and prediction of their thermodynamic behavior using original UNIFAC, mod. UNIFAC(Do) and COSMO-RS(Ol)". In: *Journal of Chemical Thermodynamics* 37.6 (2005), pp. 603–619. DOI: 10.1016/j.jct.2005.04.010.

[109] H. C. Van Ness et al. "Vapor-Liquid equilibrium: Part I. An appraisal of data reduction methods". In: *AIChE Journal* 19.2 (Mar. 1973), pp. 238–244. DOI: 10.1002/aic.690190206.

[110] B. Paul É. Clapeyron. "Mémoire sur la puissance motrice de la chaleur". In: *Journal de l'École Polytechnique* XIV (1834), pp. 153–190.

[111] J. B. West. "Robert Boyle's landmark book of 1660 with the first experiments on rarified air". In: *Journal of Applied Physiology* 98.1 (Jan. 2005), pp. 31–39. DOI: 10.1152/japplphysiol.00759.2004.

[112] J. D. van der Waals. “Over de continuiteit van den gas- en vloeistoftoestand”. PhD thesis. University of Leiden, 1873.

[113] A. Anderko. “Equation-of-state methods for the modelling of phase equilibria”. In: *Fluid Phase Equilibria* 61.1-2 (Jan. 1990), pp. 145–225. DOI: 10.1016/0378-3812(90)90011-B.

[114] Y. S. Wei and R. J. Sadus. “Equations of state for the calculation of fluid-phase equilibria”. In: *AIChE Journal* 46.1 (Jan. 2000), pp. 169–196. DOI: 10.1002/aic.690460119.

[115] J. O. Valderrama. “The state of the cubic equations of state”. In: *Industrial and Engineering Chemistry Research* 42.8 (2003), pp. 1603–1618. DOI: 10.1021/ie020447b.

[116] Robin D. Rogers and Kenneth R. Seddon, eds. *Ionic Liquids*. Vol. 818. ACS Symposium Series. Washington, DC: American Chemical Society, July 2002. DOI: 10.1021/bk-2002-0818.

[117] Peter Wasserscheid and Thomas Welton, eds. *Ionic Liquids in Synthesis*. Wiley, Oct. 2007. DOI: 10.1002/9783527621194.

[118] F. Eckert and A. Klamt. “Fast Solvent Screening via Quantum Chemistry : COSMO-RS Approach”. In: 48.2 (2002), pp. 369–385.

[119] R. P. Gerber and R. De P. Soares. “Prediction of Infinite-Dilution Activity Coefficients Using UNIFAC and COSMO-SAC Variants”. In: *Industrial & Engineering Chemistry Research* 49.16 (Aug. 2010), pp. 7488–7496. DOI: 10.1021/ie901947m.

[120] R. Xiong et al. “An Improvement to COSMO-SAC for Predicting Thermodynamic Properties”. In: *Industrial & Engineering Chemistry Research* 53.19 (May 2014), pp. 8265–8278. DOI: 10.1021/ie404410v.

[121] N. D. Austin et al. “A COSMO-based approach to computer-aided mixture design”. In: *Chemical Engineering Science* 159 (Feb. 2017), pp. 93–105. DOI: 10.1016/j.ces.2016.05.025.

[122] H. Ding et al. “Isobaric vapor-liquid equilibrium for binary system of methyl caprate + methyl laurate at 2, 4 and 6 kPa”. In: *The Journal of Chemical Thermodynamics* 116 (Jan. 2018), pp. 241–247. DOI: 10.1016/j.jct.2017.09.020.

[123] J. A. Barker and R. O. Watts. “Structure of water; A Monte Carlo calculation”. In: *Chemical Physics Letters* 3.3 (1969), pp. 144–145. DOI: 10.1016/0009-2614(69)80119-3.

[124] D. A. Kofke. “Gibbs-duhem integration: A new method for direct evaluation of phase coexistence by molecular simulation”. In: *Molecular Physics* 78.6 (1993), pp. 1331–1336. DOI: 10.1080/00268979300100881.

[125] G. C. A. M. Mooij et al. “Direct simulation of phase equilibria of chain molecules”. In: *Journal of Physics: Condensed Matter* 4.16 (Apr. 1992), pp. L255–L259. DOI: 10.1088/0953-8984/4/16/001.

[126] M. G. Martin and J. I. Siepmann. “Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes”. In: *The Journal of Physical Chemistry B* 103.21 (2002), pp. 4508–4517. DOI: 10.1021/jp984742e.

[127] W.G. Chapman et al. “SAFT: Equation-of-state solution model for associating fluids”. In: *Fluid Phase Equilibria* 52 (Dec. 1989), pp. 31–38. DOI: 10.1016/0378-3812(89)80308-5.

[128] W. G. Chapman et al. “New Reference Equation of State for Associating Liquids”. In: *Industrial and Engineering Chemistry Research* 29.8 (1990), pp. 1709–1721. DOI: 10.1021/ie00104a021.

[129] B. H. Morrow and J. A. Harrison. “Vapor-Liquid Equilibrium Simulations of Hydrocarbons Using Molecular Dynamics with Long-Range Lennard-Jones Interactions”. In: *Energy and Fuels* 33.2 (2019), pp. 848–858. DOI: 10.1021/acs.energyfuels.8b03700.

[130] M. Veit et al. “Equation of State of Fluid Methane from First Principles with Machine Learning Potentials”. In: *Journal of Chemical Theory and Computation* 15.4 (2019), pp. 2574–2586. DOI: 10.1021/acs.jctc.8b01242.

[131] R. D. Groot and P. B. Warren. “Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation”. In: *The Journal of Chemical Physics* 107.11 (Sept. 1997), pp. 4423–4435. DOI: 10.1063/1.474784.

[132] I. Pagonabarraga and D. Frenkel. “Dissipative particle dynamics for interacting systems”. In: *The Journal of Chemical Physics* 115.11 (Sept. 2001), pp. 5015–5026. DOI: 10.1063/1.1396848.

[133] J. C. Shillcock and R. Lipowsky. “Equilibrium structure and lateral stress distribution of amphiphilic bilayers from dissipative particle dynamics simulations”. In: *The Journal of Chemical Physics* 117.10 (Sept. 2002), pp. 5048–5061. DOI: 10.1063/1.1498463.

[134] S. Yamamoto et al. “Dissipative particle dynamics study of spontaneous vesicle formation of amphiphilic molecules”. In: *The Journal of Chemical Physics* 116.13 (Apr. 2002), pp. 5842–5849. DOI: 10.1063/1.1456031.

[135] W. G. Chapman et al. “Theory and simulation of associating liquid mixtures”. In: *Fluid Phase Equilibria* 29.C (1986), pp. 337–346. DOI: 10.1016/0378-3812(86)85033-6.

[136] E. A. Müller and K. E. Gubbins. “Molecular-based equations of state for associating fluids: A review of SAFT and related approaches”. In: *Industrial and Engineering Chemistry Research* 40.10 (2001), pp. 2193–2211. DOI: 10.1021/ie000773w.

[137] K. E. Gubbins. “Perturbation theories of the thermodynamics of polar and associating liquids: A historical perspective”. In: *Fluid Phase Equilibria* 416 (2016), pp. 3–17. DOI: 10.1016/j.fluid.2015.12.043.

[138] M. S. Wertheim. “Fluids with highly directional attractive forces. II. Thermodynamic perturbation theory and integral equations”. In: *Journal of Statistical Physics* 35.1-2 (Apr. 1984), pp. 35–47. DOI: 10.1007/BF01017363.

[139] M. S. Wertheim. “Fluids with highly directional attractive forces. I. Statistical thermodynamics”. In: *Journal of Statistical Physics* 35.1-2 (Apr. 1984), pp. 19–34. DOI: 10.1007/BF01017362.

[140] S. Rahman et al. "SAFT-γ Force Field for the Simulation of Molecular Fluids. 5. Hetero-Group Coarse-Grained Models of Linear Alkanes and the Importance of Intramolecular Interactions". In: *Journal of Physical Chemistry B* 122.39 (2018), pp. 9161–9177. DOI: 10.1021/acs.jpcb.8b04095.

[141] V. Papaioannou et al. "Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments". In: *Journal of Chemical Physics* 140.5 (2014). DOI: 10.1063/1.4851455.

[142] J. Jover et al. "Fluid-fluid coexistence in an athermal colloid-polymer mixture: Thermodynamic perturbation theory and continuum molecular-dynamics simulation". In: *Molecular Physics* 113.17-18 (2015), pp. 2608–2628. DOI: 10.1080/00268976.2015.1047425.

[143] J. Gross and G. Sadowski. "Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules". In: *Industrial and Engineering Chemistry Research* 40.4 (2001), pp. 1244–1260. DOI: 10.1021/ie0003887.

[144] J. Gross and G. Sadowski. "Application of the perturbed-chain SAFT equation of state to associating systems". In: *Industrial and Engineering Chemistry Research* 41.22 (2002), pp. 5510–5515. DOI: 10.1021/ie010954d.

[145] R. W. Taft et al. "Studies of Hydrogen-Bonded Complex Formation with p-Fluorophenol. V. Linear Free Energy Relationships with OH Reference Acids". In: *Journal of the American Chemical Society* 91.17 (1969), pp. 4801–4808. DOI: 10.1021/ja01045a038.

[146] A. R. Fersht. *Enzyme Structure and Mechanism*. W.H. Freeman, 1985.

[147] C. A. Hunter and J. K.M. Sanders. "The Nature of π-π Interactions". In: *Journal of the American Chemical Society* 112.14 (1990), pp. 5525–5534. DOI: 10.1021/ja00170a016.

[148] H.-J. Schneider. "Mechanisms of Molecular Recognition : Investigations of Organic Host–Guest Complexes". In: *Angewandte Chemie International Edition in English* 30.11 (Nov. 1991), pp. 1417–1436. DOI: 10.1002/anie.199114171.

[149] C. A. Hunter. "Quantifying intermolecular interactions: Guidelines for the molecular recognition toolbox". In: *Angewandte Chemie - International Edition* 43 (2004), pp. 5310–5324. DOI: 10.1002/anie.200301739.

[150] D. Gurka and R. W. Taft. "Studies of Hydrogen-Bonded Complex Formation with p-Fluorophenol. IV. The Fluorine Nuclear Magnetic Resonance Method". In: *Journal of the American Chemical Society* 91.17 (1969), pp. 4794–4801. DOI: 10.1021/ja01045a037.

[151] M. H. Abraham and J. A. Platts. "Hydrogen Bond Structural Group Constants". In: *The Journal of Organic Chemistry* 66.10 (May 2001), pp. 3484–3491. DOI: 10.1021/jo001765s.

[152] J. Mitsky et al. "Hydrogen-bonded complex formation with 5-fluoroindole. Applications of the pKHB scale". In: *Journal of the American Chemical Society* 94.10 (May 1972), pp. 3442–3445. DOI: 10.1021/ja00765a030.

[153] M. H. Abraham et al. "Hydrogen bonding. Part 1.—Equilibrium constants and enthalpies of complexation for monomeric carboxylic acids with N-methylpyrrolidinone in 1,1,1-trichloroethane". In: *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* 82.11 (1986), p. 3501. DOI: 10.1039/f19868203501.

[154] S. J. Pike et al. "H-Bond Acceptor Parameters for Anions". In: *Journal of the American Chemical Society* 139.19 (May 2017), pp. 6700–6706. DOI: 10.1021/jacs.7b02008.

[155] C. S. Calero et al. "Footprinting molecular electrostatic potential surfaces for calculation of solvation energies." In: *Physical chemistry chemical physics : PCCP* 15 (2013), pp. 18262–73. DOI: 10.1039/c3cp53158a.

[156] A. Oliver et al. "A surface site interaction point methodology for macromolecules and huge molecular databases". In: *Journal of Computational Chemistry* 38.7 (2017), pp. 419–426. DOI: 10.1002/jcc.24695.

[157] A. Oliver et al. "An improved methodology to compute surface site interaction points using high density molecular electrostatic potential surfaces". In: *Journal of Computational Chemistry* 39.28 (2018), pp. 2371–2377. DOI: 10.1002/jcc.25574.

[158] C. A. Hunter. "A surface site interaction model for the properties of liquids at equilibrium". In: *Chemical Science* 4 (2013), pp. 1687–1700. DOI: 10.1039/c3sc22124e.

[159] A. R. Oganov. "Crystal structure prediction: reflections on present status and challenges". In: *Faraday Discussions* 211 (2018), pp. 643–660. DOI: 10.1039/c8fd90033g.

[160] D. Musumeci et al. "Virtual cocrystal screening". In: *Chemical Science* 2 (2011), p. 883. DOI: 10.1039/c0sc00555j.

[161] T. Grecu et al. "Validation of a Computational Cocrystal Prediction Tool: Comparison of Virtual and Experimental Cocrystal Screening Results". In: *Crystal Growth & Design* 14.1 (Jan. 2014), pp. 165–171. DOI: 10.1021/cg401339v.

[162] T. Grecu et al. "Virtual Screening Identi fi es New Cocrystals of Nalidixic Acid". In: (2014). DOI: 10.1021/cg401889h.

[163] T. Grecu et al. "Cocrystals of spironolactone and griseofulvin based on an in silico screening method". In: *CrystEngComm* 19.26 (2017), pp. 3592–3599. DOI: 10.1039/C7CE00891K.

[164] R. Barbas et al. "Combined Virtual/Experimental Multicomponent Solid Forms Screening of Sildenafil: New Salts , Cocrystals, and Hybrid Salt-Cocrystals". In: (2018). DOI: 10.1021/acs.cgd.8b01413.

[165] C. A. Hunter and R. Prohens. "Solid form and solubility". In: *CrystEngComm* 19.1 (2017), pp. 23–26. DOI: 10.1039/C6CE02094A.

[166] J. McKenzie et al. "H-bond competition experiments in solution and the solid state". In: *CrystEngComm* 18.3 (2016), pp. 394–397. DOI: 10.1039/C5CE02223A.

[167] J. McKenzie et al. "Competitor analysis of functional group H-bond donor and acceptor properties using the Cambridge Structural Database". In: *Physical Chemistry Chemical Physics* 20.39 (2018), pp. 25324–25334. DOI: 10.1039/C8CP05470C.

[168] J. McKenzie et al. “Correction: Competitor analysis of functional group H-bond donor and acceptor properties using the Cambridge Structural Database”. In: *Physical Chemistry Chemical Physics* 21.11 (2019), pp. 6296–6296. DOI: 10.1039/C9CP90062D.

[169] C. R. Groom et al. “The Cambridge Structural Database.” In: *Acta crystallographica Section B, Structural science, crystal engineering and materials* 72.Pt 2 (2016), pp. 171–9. DOI: 10.1107/S2052520616003954.

[170] A. P. Bento et al. “The ChEMBL bioactivity database: an update”. In: *Nucleic Acids Research* 42.D1 (Jan. 2014), pp. D1083–D1090. DOI: 10.1093/nar/gkt1031.

[171] J. Goodman. “How Well Are We Using XML in Chemistry?” In: *Chemistry International – Newsmagazine for IUPAC* 24.4 (Jan. 2002), pp. 7–8. DOI: 10.1515/ci.2002.24.4.7.

[172] J. Westbrook et al. “PDBML: the representation of archival macromolecular structure data in XML”. In: *Bioinformatics* 21.7 (Apr. 2005), pp. 988–992. DOI: 10.1093/bioinformatics/bti082.

[173] The HDF Group. *Hierarchical data format version 5*. 2000. URL: %5Curl%7Bhttp://www.hdfgroup.org/HDF5%7D (visited on 06/23/2019).

[174] Ecma International. *The JSON Data Interchange Syntax*. 2017. URL: %5Curl%7Bhttp://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf%7D (visited on 06/23/2019).

[175] H. M. Berman. “The Protein Data Bank”. In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242. DOI: 10.1093/nar/28.1.235.

[176] H. Berman et al. “Announcing the worldwide Protein Data Bank”. In: *Nature Structural Biology* 10.12 (Dec. 2003), pp. 980–980. DOI: 10.1038/nsb1203-980.

[177] A. Pavlo et al. “A comparison of approaches to large-scale data analysis”. In: *Proceedings of the 35th SIGMOD international conference on Management of data - SIGMOD ’09*. Vol. 12. 2. New York, New York, USA: ACM Press, 2009, p. 165. DOI: 10.1145/1559845.1559865.

[178] P. Murray-Rust and H. S. Rzepa. “Chemical Markup, XML, and the Worldwide Web. 1. Basic Principles”. In: *Journal of Chemical Information and Computer Sciences* 39.6 (Nov. 1999), pp. 928–942. DOI: 10.1021/ci990052b.

[179] *RFC 3076, accessed 30-September-2019*. 2001. URL: https://www.ietf.org/rfc/rfc3076.txt.

[180] *RFC 5261, accessed 30-September-2019*. URL: http://www.rfc-base.org/rfc-5261.html.

[181] Stein S. et al. *IUPAC International Chemical Identifier (InChI) InChI version 1, software version 1.04 Technical Manual*. 2011.

[182] T. Bray et al. *"Extensible Markup Language (XML) 1.0 (2nd ed)", W3C REC-xml*. accessed 30-September-2019. Oct. 2000. URL: http://www.w3.org/TR/REC-xml.

[183] P. Murray-Rust and H. S. Rzepa. “Chemical Markup, XML and the World-Wide Web. 2. Information Objects and the CMLDOM”. In: *Journal of Chemical Information and Computer Sciences* 41.5 (Sept. 2001), pp. 1113–1123. DOI: 10.1021/ci000404a.

[184] P. Murray-Rust and H. S. Rzepa. “Chemical Markup, XML, and the World Wide Web. 4. CML Schema”. In: *Journal of Chemical Information and Computer Sciences* 43.3 (May 2003), pp. 757–772. DOI: 10.1021/ci0256541.

[185] P. Murray-Rust et al. “Chemical Markup, XML, and the World Wide Web. 5. Applications of Chemical Metadata in RSS Aggregators”. In: *Journal of Chemical Information and Computer Sciences* 44.2 (Mar. 2004), pp. 462–469. DOI: 10.1021/ci034244p.

[186] W. Phadungsukanan et al. “The semantics of Chemical Markup Language (CML) for computational chemistry : CompChem”. In: *Journal of Cheminformatics* 4.1 (2012), p. 15. DOI: 10.1186/1758-2946-4-15.

[187] M. Frenkel et al. “Extension of ThermoML: The IUPAC standard for thermodynamic data communications (IUPAC Recommendations 2011)”. In: *Pure and Applied Chemistry* 83.10 (Sept. 2011), pp. 1937–1969. DOI: 10.1351/PAC-REC-11-05-01.

[188] M. Valiev et al. “NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations”. In: *Computer Physics Communications* 181.9 (2010), pp. 1477–1489. DOI: 10.1016/j.cpc.2010.04.018.

[189] W. A. de Jong et al. “From data to analysis: linking NWChem and Avogadro with the syntax and semantics of Chemical Markup Language”. In: *Journal of Cheminformatics* 5.1 (2013), p. 25. DOI: 10.1186/1758-2946-5-25.

[190] M. D. Hanwell et al. “Open chemistry: RESTful web APIs, JSON, NWChem and the modern web application”. In: *Journal of Cheminformatics* 9.1 (2017), pp. 1–10. DOI: 10.1186/s13321-017-0241-z.

[191] H. Andrews and B. Hutton. *JSON Schema: A Media Type for Describing JSON Documents*. URL: %5Curl%7Bhttps://json-schema.org/work-in-progress/WIP-jsonschema-core.html%7D (visited on 06/23/2019).

[192] D. C. Price et al. “HDFITS: Porting the FITS data model to HDF5”. In: *Astronomy and Computing* 12 (2015), pp. 212–220. DOI: 10.1016/j.ascom.2015.05.001.

[193] J. L. M. Abboud et al. “Studies on amphiprotic compounds. 4. Application of the .alpha.H2 hydrogen-bonding acidity scale to complexation between pyridine N-oxide and monomeric hydrogen-bond donors in cyclohexane”. In: *The Journal of Organic Chemistry* 55.7 (Mar. 1990). Ed. by Intergovernmental Panel on Climate Change, pp. 2230–2232. DOI: 10.1021/jo00294a046. arXiv: arXiv:1011.1669v3.

[194] M. H. Abraham et al. “Hydrogen bonding. Part 3.—Enthalpies of transfer from 1,1,1-trichloroethane to tetrachloromethane of phenols, N-methylpyrrolidinone (NMP) and phenol–NMP complexes”. In: *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* 84.3 (June 1988), p. 865. DOI: 10.1039/f19888400865.

[195] M. H. Abraham et al. “Hydrogen bonding. Part 9. Solute proton donor and proton acceptor scales for use in drug design”. In: *Journal of the Chemical Society, Perkin Transactions 2* 25.10 (1989), p. 1355. DOI: 10.1039/p29890001355.

[196] M. H. Abraham et al. “Hydrogen bonding. Part 7. A scale of solute hydrogen-bond acidity based on log K values for complexation in tetrachloromethane”. In: *Journal of the Chemical Society, Perkin Transactions 2* 6 (1989), p. 699. DOI: 10.1039/p29890000699.

[197] M. H. Abraham et al. “Hydrogen-bonding 8. Possible equivalence of solute and solvent scales of hydrogen-bond basicity of non-associated compounds”. In: *Journal of Physical Organic Chemistry* 2.7 (Sept. 1989), pp. 540–552. DOI: 10.1002/poc.610020706.

[198] M. H. Abraham et al. “Hydrogen bonding. Part 10. A scale of solute hydrogen-bond basicity using log K values for complexation in tetrachloromethane”. In: *Journal of the Chemical Society, Perkin Transactions 2* 75.4 (1990), p. 521. DOI: 10.1039/p29900000521.

[199] M. H. Abraham et al. “Analysis of hydrogen-bond complexation constants in 1,1,1-trichloroethane: the α2Hβ2H relationship”. In: *Journal of the Chemical Society, Perkin Transactions 2* 1 (1998), pp. 187–192. DOI: 10.1039/a702326j.

[200] M. H. Abraham and M. Rosés. “Hydrogen bonding. 38. Effect of solute structure and mobile phase composition on reversed-phase high-performance liquid chromatographic capacity factors”. In: *Journal of Physical Organic Chemistry* 7.12 (Dec. 1994), pp. 672–684. DOI: 10.1002/poc.610071205.

[201] M. H. Abraham et al. “Hydrogen-bonding. Part 5. A thermodynamically-based scale of solutehydrogen-bond acidity”. In: *Tetrahedron Letters* 29.13 (Jan. 1988), pp. 1587–1590. DOI: 10.1016/S0040-4039(00)80360-3.

[202] M. Berthelot et al. “The Hydrogen-Bond Basicity pKHB Scale of Peroxides and Ethers”. In: *European Journal of Organic Chemistry* 1998.5 (May 1998), pp. 925–931. DOI: 10.1002/(SICI)1099-0690(199805)1998:5<925::AID-EJOC925>3.0.CO;2-F.

[203] M. Berthelot et al. “Hydrogen-bond basicity pKHB scale of six-membered aromatic N-heterocycles”. In: *Journal of the Chemical Society, Perkin Transactions 2* 2 (1998), pp. 283–290. DOI: 10.1039/a706696a.

[204] F. Besseau et al. “Hydrogen-bond basicity of esters, lactones and carbonates”. In: *Journal of the Chemical Society, Perkin Transactions 2* 3 (1994), p. 485. DOI: 10.1039/p29940000485.

[205] F. Besseau et al. “A Theoretical Evaluation of the pKHB and ΔHHB Hydrogen-Bond Scales of Nitrogen Bases”. In: *Chemistry - A European Journal* 14.34 (Oct. 2008), pp. 10656–10669. DOI: 10.1002/chem.200800977.

[206] F. Besseau et al. “Hydrogen-bond basicity pKHB scale of aldehydes and ketones”. In: *Journal of the Chemical Society, Perkin Transactions 2* 1 (1998), pp. 101–108. DOI: 10.1039/a704427e.

[207] A. Chardin et al. “Hydrogen-bond basicity of the sulfonyl group. The case of strongly basic sulfonamidates RSO2NNMe3”. In: *Journal of the Chemical Society, Perkin Transactions 2* 1.6 (1996), p. 1047. DOI: 10.1039/p29960001047.

[208] P. Goralski et al. “The hydrogen bonding of alcohols, cholesterol and phenols with cyanate and azide ions”. In: *Journal of the Chemical Society, Perkin Transactions 2* 11 (1994), p. 2337. DOI: 10.1039/p29940002337.

[209] J. Graton et al. “Hydrogen-bond basicity pKHB scale of aliphatic primary amines”. In: *Journal of the Chemical Society, Perkin Transactions 2* 5 (1999), p. 997. DOI: 10.1039/a809265f.

[210] L. Joris et al. “Effects of polar aprotic solvents on linear free-energy relations in hydrogen-bonded complex formation”. In: *Journal of the American Chemical Society* 94.10 (May 1972), pp. 3438–3442. DOI: 10.1021/ja00765a029.

[211] C. Laurence and M. Berthelot. “Observations on the strength of hydrogen bonding”. In: *Perspectives in Drug Discovery and Design* 18.1 (2000), pp. 39–60. DOI: 10.1023/A:1008743229409.

[212] C. Laurence et al. “Hydrogen-bond basicity of thioamides and thioureas”. In: *Journal of the Chemical Society, Perkin Transactions 2* 11 (1995), p. 2075. DOI: 10.1039/p29950002075.

[213] C. Laurence et al. “Hydrogen-bond basicity of nitro-compounds”. In: *Journal of the Chemical Society, Perkin Transactions 2* 3 (1994), p. 491. DOI: 10.1039/p29940000491.

[214] C. Laurence and R. Queignec. “A thermodynamic hydrogen-bond acidity scale for monosubstituted acetylenes”. In: *Journal of the Chemical Society, Perkin Transactions 2* 1.11 (1992). Ed. by Intergovernmental Panel on Climate Change, p. 1915. DOI: 10.1039/p29920001915. arXiv: arXiv:1011.1669v3.

[215] J.-Y. Le Questel et al. “Can semi-empirical calculations yield reasonable estimates of hydrogen-bonding basicity? The case of nitriles”. In: *Journal of the Chemical Society, Perkin Transactions 2* 12 (1997), pp. 2711–2718. DOI: 10.1039/a703743k.

[216] J.-Y. Le Questel et al. “Hydrogen-bond basicity of secondary and tertiary amides, carbamates, ureas and lactams”. In: *Journal of the Chemical Society, Perkin Transactions 2* 12 (1992), p. 2091. DOI: 10.1039/p29920002091.

[217] J. Marco et al. “Hydrogen Bonding of Neutral Species in the Gas Phase: The Missing Link”. In: *Journal of the American Chemical Society* 116.19 (Sept. 1994), pp. 8841–8842. DOI: 10.1021/ja00098a067.

[218] C. Ouvrard et al. “The first basicity scale of fluoro-, chloro-, bromo- and iodoalkanes: some cross-comparisons with simple alkyl derivatives of other elements”. In: *Journal of the Chemical Society, Perkin Transactions 2* 7 (1999), pp. 1357–1362. DOI: 10.1039/a901867k.

[219] C. Takayama et al. “Quantitative separation of electronic and steric substituent effects in reactions between aliphatic amines and electron acceptors”. In: *The Journal of Organic Chemistry* 44.16 (Aug. 1979), pp. 2871–2879. DOI: 10.1021/jo01330a011.

[220] M. H. Abraham et al. “Hydrogen-bonding. Part 11. A quantitative evaluation of the hydrogen-bond acidity of imides as solutes”. In: *The Journal of Organic Chemistry* 55.7 (Mar. 1990), pp. 2227–2229. DOI: 10.1021/jo00294a045.

[221] D. H. Williams and M. S. Westwell. “Aspects of weak interactions”. In: *Chemical Society Reviews* 27.1 (1998), p. 57. DOI: 10.1039/a827057z.

[222] Å. Austerheim et al. “Studies of Hydrogen Bonding. Part XXXII. The Structure of Hydrogen-bonded Complexes between p-Fluorophenol and Various Nitriles.” In: *Acta Chemica Scandinavica* 39b (1985), pp. 583–587. DOI: 10.3891/acta.chem.scand.39b-0583.

[223] U. Blindheim and T. Gramstad. “Studies of hydrogen bonding—XX. Hydrogen bonding ability of phosphoryl compounds containing N-P and S-P bonds”. In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 25.6 (June 1969), pp. 1105–1113. DOI: 10.1016/0584-8539(69)80109-1.

[224] U. Blindheim and T. Gramstad. “Studies of hydrogen bonding—XV. The influence of a S-P bond in phosphoryl compounds on the P=O stretching frequency and their hydrogen bonding ability”. In: *Spectrochimica Acta* 21.6 (June 1965), pp. 1073–1079. DOI: 10.1016/0371-1951(65)80186-2.

[225] A. J. Dale and T. Gramstad. “Studies of hydrogen bonding—XXII”. In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 28.4 (Apr. 1972), pp. 639–650. DOI: 10.1016/0584-8539(72)80032-1.

[226] Jan Ekrene and Thor Gramstad. “Studies of hydrogen bonding—XXV”. In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 28.12 (Dec. 1972), pp. 2465–2469. DOI: 10.1016/0584-8539(72)80224-1.

[227] N. Futsaeter and T. Gramstad. “Studies of hydrogen bonding—XXXI. NMR studies of solute—solute—solvent interaction”. In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 36.12 (Jan. 1980), pp. 1083–1088. DOI: 10.1016/0584-8539(80)80097-3.

[228] T. Gramstad and K. Tjessem. “Studies of hydrogen bonding Part XXX”. In: *Journal of Molecular Structure* 48.1 (Jan. 1978), pp. 147–151. DOI: 10.1016/0022-2860(78)87235-4.

[229] T. Gramstad. “Studies of hydrogen bonding—XIX. Hydrogen bonding ability of carbethoxymethylenetriphenylphosphorane”. In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 26.3 (Mar. 1970), p. 741. DOI: 10.1016/0584-8539(70)80118-0.

[230] T. Gramstad and W.J. Fuglevik. “Studies of hydrogen bonding—XIV Hydrogen bonding of indole to phosphoryl compounds”. In: *Spectrochimica Acta* 21.3 (Mar. 1965), pp. 503–509. DOI: 10.1016/0371-1951(65)80142-4.

[231] T. Gramstad and H.J. Storesund. “Studies of hydrogen bonding—XXI. A comparison between hydrogen bonding ability and infrared intensity of the proton accepting group”. In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 26.2 (Mar. 1970), pp. 426–428. DOI: 10.1016/0584-8539(70)80090-3.

[232] T. Gramstad and G. van Binst. “Studies of hydrogen bonding—XVI the complexing of pentafluorophenol with triphenylphosphine oxide”. In: *Spectrochimica Acta* 22.10 (Oct. 1966), pp. 1681–1696. DOI: 10.1016/0371-1951(66)80213-8.

[233] T. Gramstad. “Studies of hydrogen bonding—XII. Correlation of the complexing ability of organophosphorous compounds with the Taft δ^* constant”. In: *Spectrochimica Acta* 20.4 (Apr. 1964), pp. 729–731. DOI: 10.1016/0371-1951(64)80068-0.

[234] T. Gramstad. “Studies of hydrogen bonding—Part XI”. In: *Spectrochimica Acta* 19.10 (Oct. 1963), p. 1698. DOI: 10.1016/0371-1951(63)80168-X.

[235] T. Gramstad. “Studies of hydrogen bonding—part VII”. In: *Spectrochimica Acta* 19.2 (Feb. 1963), pp. 497–508. DOI: 10.1016/0371-1951(63)80060-0.

[236] T. Gramstad. “Studies of hydrogen bonding—VIII. Hydrogen-bond association between phenol and sulphoxides and nitroso compounds”. In: *Spectrochimica Acta* 19.5 (May 1963), pp. 829–834. DOI: 10.1016/0371-1951(63)80154-X.

[237] T. Gramstad. "Studies ot hydrogen bonding—part IX". In: *Spectrochimica Acta* 19.8 (Aug. 1963), pp. 1363–1369. DOI: 10.1016/0371-1951(63)80246-5.

[238] T. Gramstad. "Studies of hydrogen bonding—X. The hydrogen bonding ability of ω-dimethylaminoethyl-diethylphosphate". In: *Spectrochimica Acta* 19.8 (Aug. 1963), pp. 1391–1392. DOI: 10.1016/0371-1951(63)80250-7.

[239] T. Gramstad and E. D. Becker. "NMR studies of hydrogen bonding of phenol with several proton acceptors". In: *Journal of Molecular Structure* 5.4 (Apr. 1970), pp. 253–261. DOI: 10.1016/0022-2860(70)80028-X.

[240] T. Gramstad et al. "Studies of Hydrogen Bonding. Part XXXIV. Dipole Moments of Carbonyl Compounds and of their Hydrogen-Bonded Complexes with Phenol." In: *Acta Chemica Scandinavica* 47 (1993), pp. 68–71. DOI: 10.3891/acta.chem.scand.47-0068.

[241] T. Gramstad and W. J. Fuglevik. "Studies of hydrogen bonding—XIII. Hydrogen bonding ability of lactones and lactams". In: *Spectrochimica Acta* 21.2 (Feb. 1965), pp. 343–344. DOI: 10.1016/0371-1951(65)80015-7.

[242] T. Gramstad et al. "Studies of Hydrogen Bonding. Part VI. A Comparison between the Ability of the =P=O and the =P=S Groups to Form Addition Compounds with Iodine and to Take Part in Hydrogen Bonding." In: *Acta Chemica Scandinavica* 16 (1962), pp. 2368–2372. DOI: 10.3891/acta.chem.scand.16-2368.

[243] T. Gramstad et al. "Studies of Hydrogen Bonding. Part V. The Intermolecular Hydrogen Bond Association of Amides with Phenol and Pentachlorophenol." In: *Acta Chemica Scandinavica* 16 (1962), pp. 1369–1377. DOI: 10.3891/acta.chem.scand.16-1369.

[244] T. Gramstad et al. "Studies of Hydrogen Bonding. Part XXXVI. Dipole Moments of Pyridines, Quinolines and Acridine, and of their Hydrogen-Bonded Complexes with Phenol." In: *Acta Chemica Scandinavica* 47 (1993), pp. 985–989. DOI: 10.3891/acta.chem.scand.47-0985.

[245] T. Gramstad et al. "Studies of Hydrogen Bonding. Part XXXIII. Dipole Moments of Phosphoryl Compounds and their Hydrogen-Bonded Complexes with Phenol." In: *Acta Chemica Scandinavica* 46 (1992), pp. 1087–1091. DOI: 10.3891/acta.chem.scand.46-1087.

[246] T. Gramstad and Ø. Mundheim. "Studies of hydrogen bonding—XXIII". In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 28.7 (July 1972), pp. 1405–1413. DOI: 10.1016/0584-8539(72)80109-0.

[247] T. Gramstad et al. "Studies of Hydrogen Bonding. Part II. Intermolecular Hydrogen Bond Association between Organophosphorus Compounds and Methanol, a-Naphthol and Pentachlorophenol, respectively." In: *Acta Chemica Scandinavica* 15 (1961), pp. 1337–1346. DOI: 10.3891/acta.chem.scand.15-1337.

[248] T. Gramstad and T. Olsen. "Studies of hydrogen bonding—XXVI. Proton NMR studies of hydrogen bonding between thiophenol and various phosphoryl compounds. The Hiquchi plot. Solvent effect". In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 30.11 (Nov. 1974), pp. 2121–2131. DOI: 10.1016/0584-8539(74)80062-0.

[249] T. Gramstad and O. R. Simonsen. "Studies of hydrogen bonding—XXVII. NMR studies of hydrogen bonding and solvent effect". In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 32.4 (Jan. 1976), pp. 723–730. DOI: 10.1016/0584-8539(76)80140-7.

[250] T. Gramstad et al. "Studies of Hydrogen Bonding. Part III. Intermolecular Hydrogen Bond Association between Nitrogen Compounds and Methanol, Phenol, alpha-Naphthol and Pentachlorophenol." In: *Acta Chemica Scandinavica* 16 (1962), pp. 807–819. DOI: 10.3891/acta.chem.scand.16-0807.

[251] T. Gramstad et al. "Studies of Hydrogen Bonding. Part IV. A Comparison between the Ability of Organophosphorus Compounds to Form Addition Compounds with Iodine and their Ability to Take Part in Hydrogen Bonding." In: *Acta Chemica Scandinavica* 16 (1962), pp. 999–1014. DOI: 10.3891/acta.chem.scand.16-0999.

[252] T. Gramstad et al. "Studies of Hydrogen Bonding. Part XXXV. Proton NMR Spectral Studies of Hydrogen Bonding between Phenol and Various Proton Acceptors. Concentration Effects." In: *Acta Chemica Scandinavica* 47 (1993), pp. 605–609. DOI: 10.3891/acta.chem.scand.47-0605.

[253] T. Gramstad and K. Tjessem. "Studies of hydrogen bonding. Part XXIX." In: *Journal of Molecular Structure* 41.2 (Nov. 1977), pp. 231–242. DOI: 10.1016/0022-2860(77)80065-3.

[254] T. Gramstad et al. "Studies of Hydrogen Bonding. Part XXVIII. Hydrogen Bond Association of Phenol with 5,5-Dimethyl-2-oxo-1,3,2-dioxaphosphorinanes and Diethylphosphonates." In: *Acta Chemica Scandinavica* 31b (1977), pp. 345–353. DOI: 10.3891/acta.chem.scand.31b-0345.

[255] T. Gramstad and O. Vikane. "Studies of hydrogen bonding—XXIV". In: *Spectrochimica Acta Part A: Molecular Spectroscopy* 28.11 (Nov. 1972), pp. 2131–2139. DOI: 10.1016/0584-8539(72)80187-9.

[256] D. M. Lowe et al. "Chemical Name to Structure: OPSIN, an Open Source Solution". In: *Journal of Chemical Information and Modeling* 51.3 (Mar. 2011), pp. 739–753. DOI: 10.1021/ci100384d.

[257] M. J. Frisch et al. *Gaussian09 Revision D.01*. Gaussian Inc. Wallingford CT 2009.

[258] N. M. O'Boyle et al. "Open Babel: An Open chemical toolbox". In: *Journal of Cheminformatics* 3.1 (2011), p. 33. DOI: 10.1186/1758-2946-3-33.

[259] M.D. Driver. *cmlgenerator*. https://bitbucket.org/mdd31/cmlgenerator/src/master/. 2019.

[260] D. Weininger. "SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules". In: *Journal of Chemical Information and Computer Sciences* 28.1 (1988), pp. 31–36. DOI: 10.1021/ci00057a005.

[261] *RDKit: Open-source cheminformatics*. http://www.rdkit.org. [Online; accessed 3-March-2015].

[262] S. Riniker and G. A. Landrum. "Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation". In: *Journal of Chemical Information and Modeling* 55.12 (2015). PMID: 26575315, pp. 2562–2574. DOI: 10.1021/acs.jcim.5b00654.

[263] A. K. Rappe et al. “UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations”. In: *Journal of the American Chemical Society* 114.25 (1992), pp. 10024–10035. DOI: 10.1021/ja00051a040. eprint: https://doi.org/10.1021/ja00051a040. URL: https://doi.org/10.1021/ja00051a040.

[264] A. D. Becke. “Density-functional thermochemistry.III. The role of exact exchange”. In: *The Journal of Chemical Physics* 98.7 (1993), p. 5648. DOI: 10.1063/1.464913.

[265] C. Lee et al. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Physical Review B* 37.2 (Jan. 1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785.

[266] S. H. Vosko et al. “Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis”. In: *Canadian Journal of Physics* 58.8 (1980), pp. 1200–1211. DOI: 10.1139/p80-159.

[267] P. J. S. Devlin et al. “Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields”. In: *The journal of Physical Chemistry* 98.45 (1994), pp. 11623–11627. DOI: 10.1021/j100046a014.

[268] L. Simón and J. M. Goodman. “How reliable are DFT transition structures? Comparison of GGA, hybrid-meta-GGA and meta-GGA functionals”. In: *Org. Biomol. Chem.* 9.3 (2011), pp. 689–700. DOI: 10.1039/C0OB00477D.

[269] W. J. Hehre et al. “Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules”. In: *The Journal of Chemical Physics* 56.5 (1972), pp. 2257–2261. DOI: 10.1063/1.1677527.

[270] V. A. Rassolov. “6-31G* basis set for third-row atoms”. In: *Journal of Computational Chemistry* 22.9 (2001), pp. 976–984. DOI: 10.1002/jcc.1058.

[271] R. Krishnan et al. “Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions”. In: *The Journal of Chemical Physics* 72.1 (1980), p. 650. DOI: 10.1063/1.438955.

[272] M.D. Driver. *nwchemcmlutils*. https://bitbucket.org/mdd31/nwchemcmlutils/src/master/. 2019.

[273] M. Williamson. *NWChem fork based on 6.8 release with extra methods used in this research.* URL: https://github.com/mjw99/nwchem.

[274] I. Wald and V. Havran. “On building fast kd-Trees for Ray Tracing, and on doing that in O(N log N)”. In: *2006 IEEE Symposium on Interactive Ray Tracing*. IEEE, Sept. 2006, pp. 61–69. DOI: 10.1109/RT.2006.280216.

[275] “Orthogonal Range Searching”. In: *Computational Geometry: Algorithms and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 95–120. DOI: 10.1007/978-3-540-77974-2_5.

[276] L. P. Cordella et al. “A (sub)graph isomorphism algorithm for matching large graphs”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.10 (Oct. 2004), pp. 1367–1372. DOI: 10.1109/TPAMI.2004.75.

[277] S. Henkel et al. “Polarisation effects on the solvation properties of alcohols”. In: *Chemical Science* 9.1 (2018), pp. 88–99. DOI: 10.1039/C7SC04890D.

[278] C. A. Hunter. "van der Waals interactions in non-polar liquids". In: *Chem. Sci.* 4 (2013), pp. 834–848. DOI: 10.1039/c2sc21666c.

[279] D.D. Perrin and I.G. Sayce. "Computer calculation of equilibrium concentrations in mixtures of metal ions and complexing species". In: *Talanta* 14.7 (July 1967), pp. 833–842. DOI: 10.1016/0039-9140(67)80105-X.

[280] G. W. Quinn and D. M. Taylort. "NSPEC: a chemical speciation program for personal computers". In: *The Analyst* 117.3 (1992), p. 689. DOI: 10.1039/an9921700689.

[281] D. Reynolds. *private communication*. 2019.

[282] S. Martel et al. "Large, chemically diverse dataset of log P measurements for benchmarking studies". In: *European Journal of Pharmaceutical Sciences* 48.1-2 (2013), pp. 21–29. DOI: 10.1016/j.ejps.2012.10.019.

[283] M. Isik et al. *SAMPL6 LogP challenge repository*. 2019. URL: https://github.com/MobleyLab/SAMPL6/tree/master/physical_properties/logP.

[284] D. Prat et al. "Sanofi's Solvent Selection Guide: A Step Toward More Sustainable Processes". In: *Organic Process Research and Development* 17.12 (2013), pp. 1517–1525. DOI: 10.1021/op4002565.

[285] R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*. Vol. 317. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. DOI: 10.1007/978-3-642-02431-3.

[286] B. Aronov et al. "Fréchet Distance for Curves, Revisited". In: 2006, pp. 52–63. DOI: 10.1007/11841036_8.

[287] J. MacQueen. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1.Statistics (Aug. 1967), pp. 281–297. URL: https://projecteuclid.org/euclid.bsmsp/1200512992.

[288] S. P. Lloyd. "Least squares quantization in {PCM}. Special issue on quantization". In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 129–137.

[289] E. Schubert et al. "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN". In: *ACM Transactions on Database Systems* 42.3 (2017), pp. 1–21. DOI: 10.1145/3068335.

[290] L. Rokach and O. Maimon. *Clustering Methods*. New York: Springer-Verlag. DOI: 10.1007/0-387-25465-X_15.

[291] J. H. Ward Jr. "Hierarchical Grouping to Optimize an Objective Function". In: *Journal of the American Statistical Association* 58.301 (1963), pp. 236–244. DOI: 10.1080/01621459.1963.10500845.

[292] J. Sherwood et al. "Dihydrolevoglucosenone (Cyrene) as a bio-based alternative for dipolar aprotic solvents". In: *Chem. Commun.* 50.68 (2014), pp. 9650–9652. DOI: 10.1039/C4CC04133J.

[293] F. Cao et al. "Dehydration of cellulose to levoglucosenone using polar aprotic solvents". In: *Energy & Environmental Science* 8.6 (2015), pp. 1808–1815. DOI: 10.1039/C5EE00353A.

[294] K. L. Wilson et al. “Scope and limitations of a DMF bio-alternative within Sonogashira cross-coupling and Cacchi-type annulation”. In: *Beilstein Journal of Organic Chemistry* 12 (Sept. 2016), pp. 2005–2011. DOI: 10.3762/bjoc.12.187.

[295] L. Mistry et al. “Synthesis of ureas in the bio-alternative solvent Cyrene”. In: *Green Chemistry* 19.9 (2017), pp. 2123–2128. DOI: 10.1039/C7GC00908A.

[296] J. H. Hildebrand. “The liquid state”. In: *Proc. Phys. Soc* 56 (Nov. 1944), p. 221. DOI: 10.1080/00107519408222130.

[297] D. Henderson. “The Theory of Liquids and Dense Gases”. In: *Annual Review of Physical Chemistry* 15.1 (Oct. 1964), pp. 31–62. DOI: 10.1146/annurev.pc.15.100164.000335.

[298] W. A. P. Luck. “A Model of Hydrogen-Bonded Liquids”. In: *Angewandte Chemie International Edition in English* 19.1 (Jan. 1980), pp. 28–41. DOI: 10.1002/anie.198000281.

[299] J. A. Barker and D. Henderson. “What is "liquid"? Understanding the states of matter”. In: *Reviews of Modern Physics* 48.4 (Oct. 1976), pp. 587–671. DOI: 10.1103/RevModPhys.48.587.

[300] J.-W. Handgraaf et al. “Density-functional theory-based molecular simulation study of liquid methanol”. In: *The Journal of Chemical Physics* 121.20 (Nov. 2004), pp. 10111–10119. DOI: 10.1063/1.1809595.

[301] H.-S. Lee and M. E. Tuckerman. “Structure of liquid water at ambient temperature from ab initio molecular dynamics performed in the complete basis set limit”. In: *The Journal of Chemical Physics* 125.15 (Oct. 2006), p. 154507. DOI: 10.1063/1.2354158.

[302] Z. Ma et al. “Ab initio molecular dynamics study of water at constant pressure using converged basis sets and empirical dispersion corrections”. In: *The Journal of Chemical Physics* 137.4 (July 2012), p. 044506. DOI: 10.1063/1.4736712.

[303] M. C.R. Symons et al. “Spectroscopic studies of water - Aprotic-solvent interactions in the water-rich region”. In: *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* 76 (1980), pp. 256–265. DOI: 10.1039/F19807600256.

[304] L. Cailletet and E. C. Mathias. In: *C. R. Acad. Sci.* 102 (1886), p. 1202.

[305] L. Cailletet and E. C. Mathias. In: *C. R. Acad. Sci.* 104 (1887), p. 1563.

[306] M.D. Driver. *HunterDatabaseSchema*. https://bitbucket.org/mdd31/hunterdatabaseschema/src/master/. 2019.

[307] M.D. Driver. *databaseCreation*. https://bitbucket.org/mdd31/databasecreation/src/master/. 2019.

[308] M.D. Driver. *phasexmlparser*. https://bitbucket.org/mdd31/phasexmlparser/src/master/. 2019.

[309] M. W. Williamson et al. *SSIP*. https://bitbucket.org/mjw99/ssip/src/master/. 2019.

[310] M.D. Driver. *phasexmlcreator*. https://bitbucket.org/mdd31/phasexmlcreator/src/master/. 2019.

[311] M.D. Driver. *singlessiplist*. https://bitbucket.org/mdd31/singlessiplist/src/master/. 2019.

[312] M.D. Driver. *resultsanalysis*. https://bitbucket.org/mdd31/resultsanalysis/src/master/. 2019.

[313] M.D. Driver. *ssiplotting*. https://bitbucket.org/mdd31/ssipplotting/src/master/. 2019.

[314] M.D. Driver. *solventmapcreator*. https://bitbucket.org/mdd31/solventmapcreator/src/master/. 2019.

[315] M.D. Driver. *fgipsolventmaker*. https://bitbucket.org/mdd31/fgipsolventmaker/src/master/. 2019.

[316] M.D. Driver. *ssimplepapercalcs*. https://bitbucket.org/mdd31/ssimplepapercalcs/src/master/. 2019.

[317] M.D. Driver. *Reynolds data Calculations*. https://bitbucket.org/mdd31/phasetdataset2/src/master/. 2019.

[318] M.D. Driver. *Martel data calculations*. https://bitbucket.org/mdd31/logpdataset1/src/master/. 2019.

[319] M.D. Driver. *SAMPL6 calculations*. https://bitbucket.org/mdd31/ sampl6mddattempt/src/master/. 2019.

[320] M.D. Driver. *solventssipplotting*. https://bitbucket.org/mdd31/solventssipplotting/src/master/. 2019.

[321] M.D. Driver. *puresolventmaps*. https://bitbucket.org/mdd31/puresolventmaps/src/master/. 2019.

[322] M.D. Driver. *thfchloroformsolventsystem*. https://bitbucket.org/mdd31/thfchloroformsolventsystem/src/master/. 2019.

[323] M.D. Driver. *waterethanolsolventsystem*. https://bitbucket.org/mdd31/waterethanolsolventsystem/src/master/. 2019.

[324] M.D. Driver. *gammacalculation*. https://bitbucket.org/mdd31/gammacalculation/src/master/. 2019.

[325] M.D. Driver. *cyrenewithpuresolventsimilarity*. https://bitbucket.org/mdd31/cyrenewithpuresolventsim src/master/. 2019.

[326] M.D. Driver. *waterethanolmixsimilarity*. https://bitbucket.org/mdd31/waterethanolmixsimilarity/src/master/. 2019.

[327] M.D. Driver. *thfchloroformmixsimilarity*. https://bitbucket.org/mdd31/thfchloroformmixsimilarity/src/master/. 2019.

[328] M.D. Driver. *VLE Experimental parameterisations*. https://bitbucket.org/mdd31/expparameterisations/src/master/. 2019.

[329] M.D. Driver. *Gramstad Association constant calculations*. https://bitbucket.org/mdd31/gramstadcalculations/src/master/. 2019.

[330] M.D. Driver. *Phase concentration calculations*. https://bitbucket.org/mdd31/phaseconcentrations/src/master/. 2019.

[331] Daylight Chemical Information Systems. *SMARTS Theory Manual*. 2019. URL: http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.

[332] Y. Marcus. *The Properties of Solvents*. Wiley, 1998.

[333] Sigma Aldrich. *Cyrene product entry*. 2019. URL: https://www.sigmaaldrich.com/catalog/product/sial/807796?lang=en%5C&region=GB (visited on 09/30/2019).

[334] American Petroleum Institute. *API Technical data book*. American Petroleum Institute.

Appendix A

Repositories

The code written to generate the work presented in this thesis are detailed below.

A.1 Data storage and handling

Data collation and handling has been an integral part of this research. Therefore the processing is done in an automated fashion with validation of data files.

A.1.1 Schema and XLST

The XML Schema and XLST designed are included in the HunterDatabaseSchema repository [306]. The Schema are deployed on the Hunter group website, to allow for remote validation. The XLST are currently on the private part of the website but may be migrated to the public part upon publication and further development.

A.1.2 Databasecreation

The experimental information gathered (detailed in 2.2 was compiled into separate csv files which were then merged with bibliographic information upon the instantiation of the database in the DatabaseCreation repository [307].

A.1.3 Data file parsing

The file parsing utility for the XML based file formats using Python was consolidated into a single repository, containing validation and parsing routines which would output

dictionaries of values to be passed onto the next stage in the pipeline. This is contained in the phasexmlparser repository [308].

A.2 Calculations and input generation

The process of running jobs was summarised in figure 3.2, with the repositories required to run the calculations listed in the following sections.

A.2.1 Structure generation

The CMLGenerator repository [259] is used to generate 3D structures. It has a command line interface for easier operation, to generate 3D structures in the defined CML format from given input (including batch submission of SMILES strings).

A.2.2 MEPS generation

The MEPS for molecules is generated using NWChem [188], version 6.8.1 or later contains the required modifications developed in the fork by Mark Williamson [273]. The automated generation of NWChem input scripts and job submission scripts, as well as the manipulation of the cube files during runtime was done using NWChemCMLUtils [272].

A.2.3 SSIP footprinting

Conversion of the MEPS data to an SSIP description is done with the SSIP code developed [309]. A command line interface (CLI) can be used to calculate the SSIP description in an easy to use fashion. For the calculation of multiple SSIP molecular footprints in a single step, the nwchemcmlutils CLI can be used [272]. This can be used for either footprinting approach (mono- or tri- surface).

A.2.4 Phase transfer calculations with SSIMPLE

To convert the XML to the appropriate input format for phase transfer calculations, in addition to the inclusion of concentration and mole fraction information, the phasexmlcreator repository [310]. This includes a CLI for the most basic required inputs of solute XML or single component solvent XML generation. Generation of multi-component solvents or

phases requires small Python scripts to be written. SSIP descriptions for dummy molecules can be created using singlessiplist [311].

The SSIMPLE calculation is undertaken using the phasetransfer sub-module of the SSIP code [309].

A.3 Data analysis

Analysis post calculation can be partitioned into general analysis, common between data sets, and specific analysis.

A.3.1 Shared functionality

The parser repositories (detailed in A.1.3) are used for standard inputs. For analysis the resultsanalysis repository [312] contains the core functionality and data classes used as an internal representation of the information.

SSIP description validation

The ssipplotting [313] repository contains methods to generate the summary plots of the footprints, shown in F. This provides a graphical summary used to evaluate the effectiveness of assigned values where the values are known. SSIP values for a single molecule are plotted on the same scale.

FGIP and solvent similarity

The solventmapcreator repository [314] contains the methods for the creation of the FGIPs and also the solvent similarity calculations and cluster analysis.

A.3.2 Phase transfer calculations

The solvent descriptions used were generated using the fgipsolventcreator repository [315]. The following repositories were used to generate the plots in the phasetransfer chapter [316–319].

A.3.3 Solvent SSIP description validation

The SSIP descriptions of the solvent molecules were validated [320] using the functionality of ssipplotting [313]. The results are displayed in Appendix F.

A.3.4 FGIP

The FGIPs presented in this work were generated using [321] for pure solvents; [322] for THF-chloroform mixtures; [323] for water-ethanol mixtures. The exploration of γ in equation (5.24) was done in [324].

A.3.5 Solvent Similarity

The similarity analysis was undertaken using [325] for pure solvents; [326] for water-ethanol mixture analysis; [327] for THF-chloroform analysis.

A.3.6 VLE

The plotting for Figures 7.2, 7.4, 7.5 and 7.6 was undertaken using [328]. Calculation of association constants and subsequent analysis was done with [329]. Calcualtion of VLE and liquid data and subsequent analysis is contained within [330].

Appendix B

Schema

A copy of the document format specifications for the XML based descriptions are below.

B.1 SSIP Schema

Listing B.1 Schema for the SSIP data.

```
<?xml version="1.0" encoding="utf-8"?>
<!-- Copyright 2017 Mark Driver Licensed under the Apache License, Version
        2.0 (the "License"); you may not use this file except in compliance with
        the License. You may obtain a copy of the License at http://www.apache.org/licenses/LICENSE-2.0
        Unless required by applicable law or agreed to in writing, software distributed
        under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES
        OR CONDITIONS OF ANY KIND, either express or implied. See the License for
        the specific language governing permissions and limitations under the License. -->
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
        targetNamespace="http://www-hunter.ch.cam.ac.uk/SSIP"
        elementFormDefault="qualified" attributeFormDefault="qualified"
        xmlns:tns="http://www-hunter.ch.cam.ac.uk/SSIP" xmlns:cml="http://www.xml-cml.org/schema"
        xmlns:h="http://www.w3.org/1999/xhtml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://www-hunter.ch.cam.ac.uk/SSIP␣http://www-hunter.ch.cam.ac.uk/schema/SSIP.xsd"
        id="SSIPschema">
        <xsd:import namespace="http://www.xml-cml.org/schema"
                schemaLocation="http://www-hunter.ch.cam.ac.uk/cmlschema.xsd" />
        <xsd:simpleType name="versionNumber">
                <xsd:annotation>
                        <xsd:documentation>
                                <h:div class="summary">simpleType to represent version number.
                                </h:div>
                                <h:div class="description">
                                        <h:p>
                                                The version number is series of 3 numbers separated by a
                                                ".",
```

```
                                    so we will match using the regular expression.
                                    "([0−9])∗\.([0−9])∗\.([0−9])∗−?(\w)∗"
                                </h:p>
                            </h:div>
                        </xsd:documentation>
                    </xsd:annotation>
                    <xsd:restriction base="xsd:string">
                        <xsd:pattern value="([0−9])∗\.([0−9])∗\.([0−9])∗−?(\w)∗" />
                    </xsd:restriction>
                </xsd:simpleType>
                <xsd:attributeGroup name="ssipNearestNeighbours">
                    <xsd:annotation>
                        <xsd:documentation>
                            <h:div class="summary">The nearest atom and non-H atom to the SSIP.
                            </h:div>
                            <h:div class="description">
                                The attributes store the nearest atom and nearest
                                non-H atom to the
                                SSIP, for use in analysis.
                            </h:div>
                        </xsd:documentation>
                    </xsd:annotation>
                    <xsd:attribute name="nearestAtomID" type="cml:idType">
                        <xsd:annotation>
                            <xsd:documentation>
                                <h:div class="summary">The atom ID for the atom nearest to the SSIP.
                                </h:div>
                                <h:div class="description">
                                    The ID matches that in an atomArray for the
                                    molecule corresponding to
                                    the SSIP.
                                </h:div>
                            </xsd:documentation>
                        </xsd:annotation>
                    </xsd:attribute>
                    <xsd:attribute name="nearestNonHAtomID" type="cml:idType">
                        <xsd:annotation>
                            <xsd:documentation>
                                <h:div class="summary"> The atom ID for the non-H atom nearest to the
                                    SSIP
                                </h:div>
                                <h:div class="description">
                                    The ID matches that in an atomArray for the
                                    molecule corresponding to
                                    the SSIP.
                                </h:div>
                            </xsd:documentation>
                        </xsd:annotation>
                    </xsd:attribute>
                </xsd:attributeGroup>
                <xsd:attributeGroup name="value">
```

```
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Attribute group for value attribute
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute name="value" type="xsd:double" use="required">
                <xsd:annotation>
                    <xsd:documentation>
                        <h:div class="summary">Value attribute for the value of the SSIP.
                        </h:div>
                    </xsd:documentation>
                </xsd:annotation>
            </xsd:attribute>
        </xsd:attributeGroup>
        <xsd:attribute name="ssipSoftwareVersion" type="tns:versionNumber">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Attribute to store version number of the SSIP
                        software.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:attribute>
        <xsd:attribute name="parameterVersion" type="tns:versionNumber">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Attribute to store version number of the
                        parameterisation.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:attribute>
        <xsd:attributeGroup name="softwareVersions">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains attributes for the SSIP software
                        version and parameter version.
                    </h:div>
                    <h:div class="description">
                        The software versions for the SSIP software and
                        also the parameters
                        are used to keep track of when the SSIPList was
                        created.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute ref="tns:ssipSoftwareVersion" use="required"></xsd:attribute>
            <xsd:attribute ref="tns:parameterVersion" use="required"></xsd:attribute>
        </xsd:attributeGroup>
        <xsd:attributeGroup name="cartesianCoords3D">
```

```
            <xsd:annotation>
                  <xsd:documentation>
                        <h:div class="summary">Grouping of the 3D Cartesian coordinates x3, y3,
                              z3 used for SSIP location.
                        </h:div>
                        <h:div class="description">
                              <h:p>
                                    The Cartesian coordinates for the SSIP, are expressed as 3
                                    attributes, collected here.
                                    They use the simpleTypes defined in the
                                    cml schema. The coordinates
                                    are x3, y3, z3.
                              </h:p>
                        </h:div>
                  </xsd:documentation>
            </xsd:annotation>
            <xsd:attributeGroup ref="cml:x3" />
            <xsd:attributeGroup ref="cml:y3" />
            <xsd:attributeGroup ref="cml:z3" />
      </xsd:attributeGroup>
      <xsd:attribute name="unit" type="xsd:string">
            <xsd:annotation>
                  <xsd:documentation>
                        <h:div class="summary"> Unit attribute for surface information pieces.
                        </h:div>
                  </xsd:documentation>
            </xsd:annotation>
      </xsd:attribute>
      <xsd:element name="TotalSurfaceArea">
            <xsd:annotation>
                  <xsd:documentation>
                        <h:div class="summary">Total Surface area of MEPS.
                        </h:div>
                  </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                  <xsd:simpleContent>
                        <xsd:extension base="xsd:double">
                              <xsd:attribute ref="tns:unit" />
                        </xsd:extension>
                  </xsd:simpleContent>
            </xsd:complexType>
      </xsd:element>
      <xsd:element name="NegativeSurfaceArea">
            <xsd:annotation>
                  <xsd:documentation>
                        <h:div class="summary">Negative Surface area of MEPS.
                        </h:div>
                  </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
```

```
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attribute ref="tns:unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="PositiveSurfaceArea">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Positive Surface area of MEPS.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attribute ref="tns:unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="ElectronDensityIsosurface">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains information on the Electron Density
                        Isosurface.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attribute ref="tns:unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="NumberOFMEPSPoints">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This is the number of points
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:integer">
                        <xsd:attribute ref="tns:unit" />
                    </xsd:extension>
                </xsd:simpleContent>
```

```
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="VdWVolume">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This is the volume encompassed by the surface.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attribute ref="tns:unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="ElectrostaticPotentialMax">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary"> This is the maximum value of the electrostatic
                        potential.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attribute ref="tns:unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="ElectrostaticPotentialMin">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary"> This is the minimum value of the electrostatic
                        potential.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attribute ref="tns:unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="Surface">
            <xsd:annotation>
```

```
                <xsd:documentation>
                    <h:div class="summary">This includes details on the Surface.
                    </h:div>
                    <h:div class="description">
                        The details included are: The total surface area,
                        negative surface
                        area, positive surface area and number of ESP
                        points on the surface
                        used to generate the SSIP representation.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:TotalSurfaceArea" />
                    <xsd:element ref="tns:NegativeSurfaceArea" />
                    <xsd:element ref="tns:PositiveSurfaceArea" />
                    <xsd:element ref="tns:ElectronDensityIsosurface" />
                    <xsd:element ref="tns:NumberOFMEPSPoints" />
                    <xsd:element ref="tns:VdWVolume" minOccurs="0"
                        maxOccurs="1" />
                    <xsd:element ref="tns:ElectrostaticPotentialMax" minOccurs="0" />
                    <xsd:element ref="tns:ElectrostaticPotentialMin" minOccurs="0" />
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="SurfaceInformation">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This includes details on the Surface.
                    </h:div>
                    <h:div class="description">
                        The details included are: The total surface area,
                        negative surface
                        area, positive surface area and number of ESP
                        points on the surface
                        used to generate the SSIP representation.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:choice>
                <xsd:sequence>
                    <xsd:element ref="tns:Surface" maxOccurs="unbounded"/>
                </xsd:sequence>
                <xsd:sequence>
                    <xsd:element ref="tns:TotalSurfaceArea" />
                    <xsd:element ref="tns:NegativeSurfaceArea" />
                    <xsd:element ref="tns:PositiveSurfaceArea" />
                    <xsd:element ref="tns:ElectronDensityIsosurface" />
                    <xsd:element ref="tns:NumberOFMEPSPoints" />
```

```
                    <xsd:element ref="tns:VdWVolume" minOccurs="0"
                            maxOccurs="1" />
                    <xsd:element ref="tns:ElectrostaticPotentialMax" minOccurs="0" />
                    <xsd:element ref="tns:ElectrostaticPotentialMin" minOccurs="0" />
                </xsd:sequence>
                </xsd:choice>

        </xsd:complexType>
</xsd:element>

<xsd:complexType name="ssipType">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary"> Complex type for SSIP element.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:attributeGroup ref="tns:value"></xsd:attributeGroup>
        <xsd:attributeGroup ref="tns:cartesianCoords3D" />
        <xsd:attributeGroup ref="tns:ssipNearestNeighbours" />
</xsd:complexType>

<xsd:element name="SSIP" type="tns:ssipType">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This element contains the SSIP value.
                        </h:div>
                        <h:div class="description">
                                <h:p>
                                        The value of the SSIP is stored in the element.
                                        The element is
                                        extended by attribute Groups for the Cartesian
                                        coordinates for the
                                        location of the SSIP on the isodensity
                                        surface, and by its nearest
                                        atoms.
                                </h:p>
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:element>
<xsd:element name="SSIPs">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains a list of SSIPs.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:SSIP" minOccurs="1" maxOccurs="unbounded" />
```

```
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="SSIPMolecule">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains the CML representation of the
                        molecule, the surface information and the SSIPs.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="cml:molecule" maxOccurs="1" minOccurs="1"></xsd:element>
                    <xsd:element ref="tns:SurfaceInformation" maxOccurs="1"
                        minOccurs="1" />
                    <xsd:element ref="tns:SSIPs" minOccurs="1" maxOccurs="1"></xsd:element>
                </xsd:sequence>
                <xsd:attributeGroup ref="tns:softwareVersions" />
            </xsd:complexType>
        </xsd:element>
</xsd:schema>
```

B.2 Phase Schema

Listing B.2 Schema for the Phase data.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Copyright 2017 Mark Driver

Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
 -->
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
        targetNamespace="http://www-hunter.ch.cam.ac.uk/PhaseSchema"
            xmlns:tns="http://www-hunter.ch.cam.ac.uk/PhaseSchema"
        elementFormDefault="qualified" attributeFormDefault="qualified" xmlns:cml="http://www.xml-cml.org/schema"
        xmlns:h="http://www.w3.org/1999/xhtml" xmlns:ssip="http://www-hunter.ch.cam.ac.uk/SSIP"
        id="Phaseschema">

        <xsd:import namespace="http://www.xml-cml.org/schema"
```

```
        schemaLocation="http://www-hunter.ch.cam.ac.uk/schema/cmlschema.xsd" />

<xsd:import namespace="http://www-hunter.ch.cam.ac.uk/SSIP"
        schemaLocation="http://www-hunter.ch.cam.ac.uk/schema/SSIP.xsd" />

<xsd:simpleType name="inchikeyType">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">simpleType specifying format of standard
                                InChIKey.
                        </h:div>
                        <h:div class="description">
                                The pattern matches standard InChIKey format.
                                matches description from IUPAC site:
                                http://www.iupac.org/home/publications/e-resources/inchi/r102-summary.html
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:restriction base="xsd:string">
                <xsd:pattern value="[A-Z]{14}\-[A-Z]{10}\-[A-Z]" />
        </xsd:restriction>
</xsd:simpleType>

<xsd:simpleType name="unitTypes">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Simple type for the different possible units.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:restriction base="xsd:string">
                <xsd:enumeration value="MOLAR"></xsd:enumeration>
                <xsd:enumeration value="SSIPConcentrationNormalised"></xsd:enumeration>
                <xsd:enumeration value="KELVIN"></xsd:enumeration>
                <xsd:enumeration value="CELSIUS"></xsd:enumeration>
        </xsd:restriction>
</xsd:simpleType>

<xsd:attribute name="ssipID" type="xsd:integer">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Attribute Group for the first SSIP ID. This is
                                just an integer, and is related to the canonical ordering of the
                                SSIPs in the phase.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:attribute>

<xsd:attribute name="units" type="tns:unitTypes">
        <xsd:annotation>
```

```
                <xsd:documentation>
                    <h:div class="summary">Attribute for units.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:attribute>

        <xsd:attributeGroup name="singleSSIPID">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Group for when only one SSIPID is required for an
                        element.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute ref="tns:ssipID" use="required"></xsd:attribute>
        </xsd:attributeGroup>

        <xsd:attribute name="stdInChIKey" type="tns:inchikeyType">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This is for the standard InChIKey of a molecule.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:attribute>

        <xsd:attribute name="moleculeID" type="xsd:string">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Attribute for moleculeID.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:attribute>

        <xsd:attributeGroup name="moleculeInformation">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div>Attribute group for information about a Molecule in a
                        solvent. This includes the InChIKey of the molecule and a name.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute ref="tns:stdInChIKey">
            </xsd:attribute>
            <xsd:attribute ref="tns:moleculeID" />
        </xsd:attributeGroup>

        <xsd:attribute name="moleFraction">
            <xsd:annotation>
```

```
            <xsd:documentation>
            <h:div class="summary">This is the fraction by volume for a molecule in a solvent.
            </h:div>
            </xsd:documentation>
            </xsd:annotation>
    </xsd:attribute>

    <xsd:attributeGroup name="solventInformation">
            <xsd:annotation>
                    <xsd:documentation>
                            <h:div class="summary">Attribute group for information about a solvent
                                    system. This includes attributes for solvent names, solvent ID and
                                    number of molecules.
                            </h:div>
                    </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute name="solventName" type="xsd:string" use="required"></xsd:attribute>
            <xsd:attribute name="solventID" type="xsd:string" use="required"></xsd:attribute>
    </xsd:attributeGroup>

    <xsd:attributeGroup name="energyPhaseInformation">
            <xsd:annotation>
                    <xsd:documentation>
                            <h:div class="summary"> This contains the attributes for referencing the
                                    phases between which the molecule has been transferred for the
                                    energy contained in the element.
                            </h:div>
                    </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute name="fromSolventID" type="xsd:string"
                    use="required"></xsd:attribute>
            <xsd:attribute name="toSolventID" type="xsd:string" use="required"></xsd:attribute>
            <xsd:attribute name="valueType" type="xsd:string" use="required"></xsd:attribute>
    </xsd:attributeGroup>

    <xsd:complexType name="Concentration">
            <xsd:annotation>
                    <xsd:documentation>
                            <h:div class="summary">Element containing information about a molecule&#39;s
                                    concentration.
                            </h:div>
                    </xsd:documentation>
            </xsd:annotation>
            <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                            <xsd:attribute ref="tns:units" />
                    </xsd:extension>
            </xsd:simpleContent>
    </xsd:complexType>

    <xsd:element name="TotalConcentration" type="tns:Concentration">
```

```
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Element contains total concentration for SSIP.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
    </xsd:element>

    <xsd:element name="FreeConcentration" type="tns:Concentration">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Element contains the FreeConcentraiton of the
                    SSIP.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
    </xsd:element>

    <xsd:element name="BoundConcentration">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Element contains the concentration of the SSIP bound
                    to the SSIP with the given ID.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:complexContent>
                <xsd:extension base="tns:Concentration">
                    <xsd:attributeGroup ref="tns:singleSSIPID" />
                </xsd:extension>
            </xsd:complexContent>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="BoundConcentrationSum" type="tns:Concentration">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Element containing total bound concentration.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
    </xsd:element>

    <xsd:element name="SSIP">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary"> SSIP redefined to include concentration.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
```

```
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:TotalConcentration" />
                <xsd:element ref="tns:FreeConcentration" minOccurs="0" />
                <xsd:element ref="tns:BoundConcentrationSum" minOccurs="0" maxOccurs="1" />
                <xsd:element ref="tns:BoundConcentration" minOccurs="0" maxOccurs="unbounded" />
            </xsd:sequence>
            <xsd:attributeGroup ref="ssip:cartesianCoords3D" />
            <xsd:attributeGroup ref="ssip:ssipNearestNeighbours" />
            <xsd:attributeGroup ref="ssip:value" />
            <xsd:attribute ref="tns:units" />
            <xsd:attribute ref="tns:moleculeID" use="required"/>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="SSIPs">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary"> SSIP container.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:SSIP" minOccurs="1" maxOccurs="unbounded" />
            </xsd:sequence>
            <xsd:attribute ref="tns:units" />
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="Molecule">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">This contains the SSIPList and information about
                    the molecule concentration.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="cml:molecule" minOccurs="1"
                    maxOccurs="1" />
                <xsd:element ref="ssip:SurfaceInformation" />
                <xsd:element ref="tns:SSIPs" />
            </xsd:sequence>
            <xsd:attributeGroup ref="tns:moleculeInformation" />
            <xsd:attribute ref="ssip:ssipSoftwareVersion" />
            <xsd:attribute ref="tns:moleFraction" use="optional" />
        </xsd:complexType>
    </xsd:element>
```

```
<xsd:element name="Molecules">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Molecules container.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:Molecule" minOccurs="1" maxOccurs="unbounded" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="Solvent">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Element contains information for the solvent.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:Molecules"/>
                </xsd:sequence>
                <xsd:attributeGroup ref="tns:solventInformation"/>
                <xsd:attribute ref="tns:units"/>
        </xsd:complexType>
</xsd:element>

<xsd:element name="Solute">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Element contains information for the solvent.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:Molecule" minOccurs="1" maxOccurs="1"></xsd:element>
                </xsd:sequence>
                <xsd:attribute name="soluteID" use="required" />
        </xsd:complexType>
</xsd:element>

<xsd:element name="Solvents">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Container for solvent elements.
                        </h:div>
                </xsd:documentation>
```

```
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:Solvent" minOccurs="1" maxOccurs="unbounded" />
            </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="SolventList">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary"> Container for solvents.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:Solvents" />
            </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="Solutes">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Container for Solutes.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:Solute" minOccurs="1" maxOccurs="unbounded" />
            </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="SoluteList">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Solutes container.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
             <xsd:sequence>
                <xsd:element ref="tns:Solutes" />
             </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="Temperature">
```

```
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Element containing information about the
                    temperature of the phase.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:simpleContent>
                <xsd:extension base="xsd:double">
                    <xsd:attribute ref="tns:units"></xsd:attribute>
                </xsd:extension>
            </xsd:simpleContent>
        </xsd:complexType>
</xsd:element>

<xsd:element name="Phase">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Element contains information about the Phase.

                    The
                    information about the SSIPs is contained in Molecule Child
                    elements, including total concentrations. The free and bound
                    concentrations is then stored in later elements using the standard
                    canonicalisation. The solventID is needed when it is within a
                    PhaseCollection.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:Molecules" />
                <xsd:element ref="tns:Temperature" />
            </xsd:sequence>
            <xsd:attribute ref="tns:units" />
            <xsd:attribute name="solventID" use="optional" />
            <xsd:attribute name="phaseType" use="optional" />
        </xsd:complexType>
</xsd:element>

<xsd:element name="Phases">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">Container for phases.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:Phase" minOccurs="1" maxOccurs="unbounded"></xsd:element>
```

```
                </xsd:sequence>
            </xsd:complexType>
</xsd:element>

<xsd:element name="PhaseSystem">
<xsd:annotation>
<xsd:documentation>
<h:div class="summary">This is represents a phase system, where a condensed and gas phase are in
    equilibrium.
</h:div>
</xsd:documentation>
</xsd:annotation>
<xsd:complexType>
<xsd:sequence>
<xsd:element ref="tns:Phases" minOccurs="1" maxOccurs="1"/>
</xsd:sequence>
<xsd:attribute name="mixtureID" type="xsd:string"></xsd:attribute>
</xsd:complexType>
</xsd:element>

<xsd:element name="PhaseSystems">
<xsd:annotation>
<xsd:documentation>
<h:div class="summary">This is is a collection of phase systems, where a condensed and gas phase
    are in equilibrium.
</h:div>
</xsd:documentation>
</xsd:annotation>
<xsd:complexType>
<xsd:sequence>
<xsd:element ref="tns:PhaseSystem" minOccurs="1" maxOccurs="unbounded"/>
</xsd:sequence>
</xsd:complexType>
</xsd:element>

<xsd:element name="Mixture">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary"> This contains a number of Phase Collections,
                                each with a different solute.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:Phases"/>
                </xsd:sequence>
                <xsd:attribute name="soluteID" use="optional" />
                <xsd:attribute name="mixtureID" use="optional" />
        </xsd:complexType>
</xsd:element>
```

```
<xsd:element name="Mixtures">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Collection of mixtures.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:Mixture" minOccurs="1" maxOccurs="unbounded" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="MixtureCollection">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="sumamry"> Collection of Mixtures. Represents MixtureContainer.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:Mixtures" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="PhaseSystemMixture">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary"> This contains a number of Phase Collections,
                                each with a different solute.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:PhaseSystems"/>
                </xsd:sequence>
                <xsd:attribute name="soluteID" use="optional" />
        </xsd:complexType>
</xsd:element>

<xsd:element name="PhaseSystemMixtures">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary"> This contains a number of Phase Collections,
                                each with a different solute.
                        </h:div>
```

```
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:PhaseSystemMixture"/>
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="PhaseSystemMixtureContainer">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary"> This contains a number of Phase Collections,
                        each with a different solute.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:PhaseSystemMixtures"/>
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="TotalEnergy">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This element contains the value for the total
                        energy from all SSIP contributions.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="EnergyContribution">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary"> This is the contribution to the total energy from
                        an individual SSIP.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
```

```
                    <xsd:attributeGroup ref="tns:singleSSIPID"></xsd:attributeGroup>
                </xsd:extension>
            </xsd:simpleContent>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="EnergyContributions">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary"> Collection of Energy contribution elements.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:EnergyContribution" minOccurs="1" maxOccurs="unbounded" />
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="EnergyValue">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary"> This is the element structure for any energy
                    Element.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="tns:TotalEnergy" />
                <xsd:element ref="tns:EnergyContributions" minOccurs="1" maxOccurs="1" />
            </xsd:sequence>
            <xsd:attribute ref="tns:moleculeID"></xsd:attribute>
            <xsd:attributeGroup ref="tns:energyPhaseInformation"></xsd:attributeGroup>
        </xsd:complexType>
    </xsd:element>

    <xsd:element name="FreeEnergy" substitutionGroup="tns:EnergyValue">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary"> This element contains a free energy of transfer
                    between the phases specified, for the molecule specified.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
    </xsd:element>

    <xsd:element name="BindingEnergy" substitutionGroup="tns:EnergyValue">
        <xsd:annotation>
            <xsd:documentation>
```

```
                    <h:div class="summary"> This element contains a binding energy of
                        transfer between the phases specified, for the molecule specified.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:element>

        <xsd:element name="ConfinementEnergy" substitutionGroup="tns:EnergyValue">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary"> This element contains a confinement energy of
                        transfer between the phases specified, for the molecule specified.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:element>

        <xsd:element name="PartitionCoefficient" substitutionGroup="tns:EnergyValue">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This element contains partition coefficient for
                     the molecule specified between the phases specified.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:element>

        <xsd:element name="FreeEnergyCollection">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This acts as a container for all FreeEnergy
                        Elements present.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:FreeEnergy" minOccurs="1"
                        maxOccurs="unbounded"></xsd:element>
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="BindingEnergyCollection">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This acts as a container for all BindingEnergy
                        Elements present.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
```

```
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:BindingEnergy" minOccurs="1"
                        maxOccurs="unbounded"></xsd:element>
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="ConfinementEnergyCollection">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This acts as a container for all
                        ConfinementEnergy Elements present.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:ConfinementEnergy" minOccurs="1"
                        maxOccurs="unbounded"></xsd:element>
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="PartitionCoefficientCollection">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This acts as a container for all
                    PartitionCoefficient Elements present.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:PartitionCoefficient" minOccurs="1"
                        maxOccurs="unbounded" />
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="EnergyValues">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This element acts as a wrapper for all the energy
                        values outputted by the phase transfer code, acting as the root
                        element.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:all minOccurs="0">
```

```
                    <xsd:element ref="tns:BindingEnergyCollection" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="tns:ConfinementEnergyCollection" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="tns:FreeEnergyCollection" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="tns:PartitionCoefficientCollection" minOccurs="0" maxOccurs="1" />
                </xsd:all>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="AssociationEnergyValue">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Element contains Association energy values.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
            <xsd:attribute name="moleculeID1"/>
            <xsd:attribute name="moleculeID2"/>
            <xsd:attribute name="solventID"/>
            <xsd:attribute name="value"/>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="AssociationEnergyValueCollection">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Element contains Collection of Association energy
                        values.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:AssociationEnergyValue" minOccurs="1" maxOccurs="unbounded"/>
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="FreeConcentrationFraction">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Element contains information on free concentration
                        fraction.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:attribute ref="tns:moleculeID"/>
                <xsd:attribute name="concentrationFraction" type="xsd:double"/>
            </xsd:complexType>
        </xsd:element>
```

```
<xsd:element name="BoundConcentrationFraction">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Element contains information on bound concentration
                            fraction.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:attribute ref="tns:moleculeID"/>
                <xsd:attribute name="boundToMoleculeID" type="xsd:string"/>
                <xsd:attribute name="concentrationFraction" type="xsd:double"/>
        </xsd:complexType>
</xsd:element>

<xsd:element name="MoleculeConcentrationFraction">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Element contains information on molecule
                            concentration fractions.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:FreeConcentrationFraction" minOccurs="1" maxOccurs="1"/>
                        <xsd:element ref="tns:BoundConcentrationFraction" minOccurs="1"
                            maxOccurs="unbounded"/>
                </xsd:sequence>
                <xsd:attribute ref="tns:moleculeID"/>
        </xsd:complexType>
</xsd:element>

<xsd:element name="PhaseConcentrationFraction">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Element contains information on phase concentration
                            fractions.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="tns:MoleculeConcentrationFraction" minOccurs="1"
                            maxOccurs="unbounded"/>
                        <xsd:element ref="tns:Temperature"/>
                </xsd:sequence>
                <xsd:attribute name="solventID"/>
                <xsd:attribute name="phaseType"/>
        </xsd:complexType>
```

```
        </xsd:element>

        <xsd:element name="PhaseConcentrationFractionCollection">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Element contains information on phase concentration
                        fractions.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="tns:PhaseConcentrationFraction" minOccurs="1"
                        maxOccurs="unbounded"/>
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

</xsd:schema>
```

B.3 Database Schema

Listing B.3 Schema for the database

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Copyright 2017 Mark Driver

Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
 -->
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
        targetNamespace="http://www-hunter.ch.cam.ac.uk/HunterDatabase"
            xmlns="http://www-hunter.ch.cam.ac.uk/HunterDatabase"
        xmlns:h="http://www.w3.org/1999/xhtml" xmlns:cml="http://www.xml-cml.org/schema"
        xmlns:convention="http://www.xml-cml.org/convention/"
            xmlns:compchem="http://www.xml-cml.org/dictionary/compchem/"
        xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:ssip="http://www-hunter.ch.cam.ac.uk/SSIP"
        elementFormDefault="qualified" attributeFormDefault="qualified">
        <xsd:import namespace="http://www.xml-cml.org/schema"
                schemaLocation="http://www-hunter.ch.cam.ac.uk/schema/cmlschema.xsd" />
        <xsd:import namespace="http://www.xml-cml.org/dictionary/compchem/" />
```

```
<xsd:import namespace="http://www-hunter.ch.cam.ac.uk/SSIP"
        schemaLocation="SSIP.xsd" />

<xsd:simpleType name="nameType">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Enumeration of name type.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:restriction base="xsd:string">
                <xsd:enumeration value="IUPAC" />
                <xsd:enumeration value="CAS" />
                <xsd:enumeration value="trivial" />
                <xsd:enumeration value="systematic" />
                <xsd:enumeration value="generic" />
                <xsd:enumeration value="abbreviation" />
        </xsd:restriction>
</xsd:simpleType>

<xsd:simpleType name="inchikeyType">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">simpleType specifying format of standard
                                InChIKey.
                        </h:div>
                        <h:div class="description">
                                The pattern matches standard InChIKey format.
                                matches description from IUPAC site:
                                http://www.iupac.org/home/publications/e-resources/inchi/r102-summary.html
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:restriction base="xsd:string">
                <xsd:pattern value="[A-Z]{14}\-[A-Z]{10}\-[A-Z]" />
        </xsd:restriction>
</xsd:simpleType>

<xsd:attributeGroup name="journalData">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains the attributes for Journal Volume
                                number and page number for the article in question.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:attribute name="volume" type="xsd:string">
                <xsd:annotation>
                        <xsd:documentation>
                                <h:div class="summary">This is the volume number for the Journal.
                                </h:div>
```

```
                    </xsd:documentation>
                </xsd:annotation>
            </xsd:attribute>
            <xsd:attribute name="pageNumber" type="xsd:string">
                <xsd:annotation>
                    <xsd:documentation>
                        <h:div class="summary">This is the page numbers for the journal.
                        </h:div>
                    </xsd:documentation>
                </xsd:annotation>
            </xsd:attribute>
            <xsd:attribute name="year" type="xsd:gYear">
                <xsd:annotation>
                    <xsd:documentation>
                        <h:div class="summary">This is the year of publication for the
                            journal.
                        </h:div>
                    </xsd:documentation>
                </xsd:annotation>
            </xsd:attribute>
            <xsd:attribute name="numberOfPages" type="xsd:integer">
                <xsd:annotation>
                    <xsd:documentation>
                        <h:div class="summary">This is the number of pages for the article.
                        </h:div>
                    </xsd:documentation>
                </xsd:annotation>
            </xsd:attribute>
    </xsd:attributeGroup>

    <xsd:attributeGroup name="propertyName">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This attribute group contains the attribute for a
                        property name.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute name="name" type="xsd:string" />
    </xsd:attributeGroup>

    <xsd:attributeGroup name="jobType">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This attribute group specifies the Job type for a
                        calculation.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:attribute name="jobType" type="xsd:string" />
    </xsd:attributeGroup>
```

```
<xsd:element name="Name">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This is a name for a molecule.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:simpleContent>
                        <xsd:extension base="xsd:string">
                                <xsd:attribute name="type" type="nameType" />
                        </xsd:extension>
                </xsd:simpleContent>
        </xsd:complexType>
</xsd:element>
<xsd:element name="Names">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This is a collection of all the Names of a
                                molecule.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="Name" minOccurs="1" maxOccurs="unbounded" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="CanonicalSmiles" type="xsd:string">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains the canonicalised Smiles of the
                                molecule.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:element>

<xsd:element name="StdInChIKey" type="inchikeyType">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Contains Standard InChIKey
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:element>

<xsd:element name="MolecularStructure3D">
```

```
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains an atom array and bond array.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="cml:molecule" />
                </xsd:sequence>
                <xsd:attribute name="structureType" type="xsd:string" />
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="Value">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains the value of a data entry
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attribute name="unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>
        <xsd:element name="DOI" type="xsd:string">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This is the DOI for a paper.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
        </xsd:element>
        <xsd:element name="JournalTitle">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This is the name of the Journal were the paper
                        was published
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:string">
                        <xsd:attributeGroup ref="journalData" />
                    </xsd:extension>
                </xsd:simpleContent>
```

```
        </xsd:complexType>
    </xsd:element>
    <xsd:element name="ArticleName" type="xsd:string">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">The Name of the article from which the data was
                    retrieved.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
    </xsd:element>
    <xsd:element name="Author" type="xsd:string">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">This is an Author for the paper where the data
                    was retrieved from.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
    </xsd:element>
    <xsd:element name="AuthorList">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">This is the list of Authors on a paper.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
            <xsd:sequence>
                <xsd:element ref="Author" minOccurs="1" maxOccurs="unbounded" />
            </xsd:sequence>
        </xsd:complexType>
    </xsd:element>
    <xsd:element name="Link" type="xsd:string">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="summary">This contains the Link/URL for a source- normally
                    only used if no DOI is presentfor a Primary source, but a URL is
                    provided.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
    </xsd:element>
    <xsd:element name="PrimarySource">
        <xsd:annotation>
            <xsd:documentation>
                <h:div class="smmary">This contains the Primary source of the data, if
                    the data was not first published in the source.
                </h:div>
            </xsd:documentation>
        </xsd:annotation>
```

```
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="DOI" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="JournalTitle" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="ArticleName" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="AuthorList" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="Link" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="Notes" minOccurs="0" maxOccurs="1" />
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="Source">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains the information on the source of
                        the data.
                    </h:div>
                    <h:div class="description">
                        This includes the DOI, JournalTitle and
                        ArticleName.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="DOI" />
                    <xsd:element ref="JournalTitle" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="ArticleName" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="AuthorList" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="Link" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="PrimarySource" minOccurs="0"
                        maxOccurs="1" />
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="Uncertainty">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This is the uncertainty in a value
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:float">
                        <xsd:attribute name="unit" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
```

```
</xsd:element>
<xsd:element name="Notes" type="xsd:string">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains additional notes about a data
                                entry.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:element>
<xsd:element name="ExperimentalMethod" type="xsd:string">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains the method used to obtain the
                                experimental value.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:element>
<xsd:element name="Temperature">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains the temperature at which the
                                experiment was done.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:simpleContent>
                        <xsd:extension base="xsd:string">
                                <xsd:attribute name="unit" />
                        </xsd:extension>
                </xsd:simpleContent>
        </xsd:complexType>
</xsd:element>
<xsd:element name="Solvent" type="xsd:string">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains the solvent used in the experiment.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:element>
<xsd:element name="Host" type="xsd:string">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains the Host used in the experiment.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
</xsd:element>
```

```
<xsd:element name="Conditions">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains all the experimental conditions for
                                the measurement.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="Temperature" minOccurs="0" maxOccurs="1" />
                        <xsd:element ref="Host" minOccurs="0" maxOccurs="1" />
                        <xsd:element ref="Solvent" minOccurs="0" maxOccurs="1" />
                        <xsd:element ref="Notes" minOccurs="0" maxOccurs="1" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>
<xsd:element name="Method">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This is the method by which the data value was
                                obtained. It contains the ExperiementalMethod and a list of
                                Conditions.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="ExperimentalMethod" minOccurs="0"
                                maxOccurs="1" />
                        <xsd:element ref="Conditions" minOccurs="0" maxOccurs="1" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>


<xsd:element name="Property">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">Property element is used to store information
                                about a Property for a molecule. The property is either a
                                     numerical
                                value with an uncertainty, a SSIP or SSIPList.
                        </h:div>
                        <h:div class="description">
                                The Property element contains one of the
                                following: Value and uncertainty, SSIP or SSIPList.
                                The Source
                                element must also be present, with Method and Notes elements
                                optional. The name of the property is stored as an attribute.
                        </h:div>
```

```
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:sequence>
                        <xsd:element ref="Value" />
                        <xsd:element ref="Uncertainty" />
                    </xsd:sequence>
                    <xsd:element ref="Method" minOccurs="0" maxOccurs="1" />
                    <xsd:element ref="Source" minOccurs="1" maxOccurs="1" />
                    <xsd:element ref="Notes" minOccurs="0" maxOccurs="1" />
                </xsd:sequence>
                <xsd:attributeGroup ref="propertyName" />
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="PropertyMeanValue">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary"> This is the mean value of a property.
                    </h:div>
                    <h:div class="description">
                        This is used for properties where multiple
                        measurements have been found
                        in the literature, thus the mean value
                        of all occurances will be
                        used instead of only one.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:simpleContent>
                    <xsd:extension base="xsd:double">
                        <xsd:attributeGroup ref="cml:units" />
                    </xsd:extension>
                </xsd:simpleContent>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="PropertyPreferedValue">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains the preferred value to use if
                        multiple version of the property are stored.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="Property" />
                </xsd:sequence>
```

```
                <xsd:attributeGroup ref="propertyName" />
        </xsd:complexType>
</xsd:element>

<xsd:element name="PropertyList">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains a list of property Elements. It can
                                also contain a PropertyMeanValue, if the data is compatible.
                        </h:div>
                        <h:div class="description">
                                This groups all values of a property which have
                                been found, curated and
                                added to the database. If the data is found
                                from multiple sources/experimental runs, then the mean can be
                                calculated. This is
                                then used for parameterisation/comparison to
                                calculated values.
                                This is stored in a PropertyMeanValue element.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="PropertyMeanValue" minOccurs="0"
                                maxOccurs="1" />
                        <xsd:element ref="PropertyPreferedValue" minOccurs="0"
                                maxOccurs="1" />
                        <xsd:element ref="Property" minOccurs="1" maxOccurs="unbounded" />
                </xsd:sequence>
                <xsd:attributeGroup ref="propertyName" />
        </xsd:complexType>
</xsd:element>

<xsd:element name="InputFile">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains the input file used for a
                                calculation
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <!-- I need to better understand the compchem convention and how to use
                        it. -->
        </xsd:complexType>
</xsd:element>

<xsd:element name="Input">
        <xsd:annotation>
                <xsd:documentation>
```

```
                    <h:div class="summary">This contains an input file and also a
                        MolecularStructure3D.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="InputFile" />
                    <xsd:element ref="MolecularStructure3D" />
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="OutputFile">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains the output file for a calculation.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <!-- Need to look at compchem convention. OR this contains reference to
                    output file. -->
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="OutputData">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">Output data from the calcuilation- extra info not
                        in Outputfile.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <!-- Need to look at compchem convention -->
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="Output">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains an output file and
                        MolecularStructure3D.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="OutputFile" />
                    <xsd:element ref="MolecularStructure3D" />
```

```
                    <xsd:element ref="OutputData" minOccurs="0" />
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="Calculation">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains information on a calculation
                        carried out on the molecule. It either conforms to the compchem
                        convention for a calculation using NWChem, or our internal
                        convention for our calculations.
                    </h:div>
                    <h:div class="description">
                        The choice between either using compchem for
                        NWChem calculations or
                        the internal convention for footprinting
                        calculations. Internal Convention needs work.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                    <xsd:sequence>
                        <xsd:element ref="Input" />
                        <xsd:element ref="Output" />
                        <xsd:element ref="cml:module" />
                    </xsd:sequence>
                <xsd:attributeGroup ref="jobType" />
            </xsd:complexType>
        </xsd:element>


        <xsd:element name="Structure">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This contains all structural information.
                    </h:div>
                    <h:div class="description">
                        This contains Names, CanonicalSmiles, StdInChIKey
                        and
                        MolecularStructure3D elements.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="Names" maxOccurs="1" />
                    <xsd:element ref="CanonicalSmiles" minOccurs="1"
                        maxOccurs="1" />
                    <xsd:element ref="StdInChIKey" maxOccurs="1" />
                    <xsd:element ref="MolecularStructure3D" minOccurs="0"
```

```
                                maxOccurs="1" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>

<xsd:element name="ExperimentalProperties">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains all Experimental data gathered.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="Property" minOccurs="0" maxOccurs="unbounded" />
                        <xsd:element ref="PropertyList" minOccurs="0" maxOccurs="unbounded" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>
<xsd:element name="CalculatedProperties">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains all the calculated Property data.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
                <xsd:sequence>
                        <xsd:element ref="Property" minOccurs="0" maxOccurs="unbounded" />
                        <xsd:element ref="PropertyList" minOccurs="0" maxOccurs="unbounded" />
                </xsd:sequence>
        </xsd:complexType>
</xsd:element>
<xsd:element name="CalculationData">
        <xsd:annotation>
                <xsd:documentation>
                        <h:div class="summary">This contains Data for calculations run on the
                                molecule.
                        </h:div>
                        <h:div class="description">
                                This contains inputs and outputs for Calculations
                                carried out on the
                                molecule. Each Calculation contains an Input and
                                Output (if Job has
                                finished). The results of Calculations are used
                                to update the
                                Calculated Properties.
                        </h:div>
                </xsd:documentation>
        </xsd:annotation>
        <xsd:complexType>
```

```
                <xsd:sequence>
                    <xsd:element ref="Calculation" minOccurs="0" maxOccurs="unbounded" />
                </xsd:sequence>
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="Molecule">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary"> This contains all the information about a
                        Molecule.
                    </h:div>
                    <h:div class="description">
                        This contains Structure, PhysicalProperties,
                        CalculatedProperties and
                        CalculationData children, and a StdInChIKey
                        attribute. The
                        StdInChIKey is used to distinguish the Molecules from
                        each other, so they should be unique.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="Structure" minOccurs="1" maxOccurs="1" />
                    <xsd:element ref="ExperimentalProperties" minOccurs="1"
                        maxOccurs="1" />
                    <xsd:element ref="CalculatedProperties" minOccurs="0"
                        maxOccurs="1" />
                    <xsd:element ref="CalculationData" minOccurs="0"
                        maxOccurs="1" />
                </xsd:sequence>
                <xsd:attribute name="inchikey" type="inchikeyType"
                    use="required" />
                <xsd:attribute name="stuartFileName" type="xsd:string" />
                <xsd:attribute name="freesolvFileName" type="xsd:string" />
            </xsd:complexType>
        </xsd:element>

        <xsd:element name="MoleculeList">
            <xsd:annotation>
                <xsd:documentation>
                    <h:div class="summary">This is a container for all Molecules, and is the
                        root element of the database.
                    </h:div>
                </xsd:documentation>
            </xsd:annotation>
            <xsd:complexType>
                <xsd:sequence>
                    <xsd:element ref="Molecule" minOccurs="1" maxOccurs="unbounded" />
                </xsd:sequence>
```

```
                </xsd:complexType>
            </xsd:element>
</xsd:schema>
```

Appendix C

Supplementary Derivations

This Appendix contains an alternate approach to derive the free energy of solvation to that discussed in chapters 4, 5, 6, 7. The fitting of polynomial functions to describe the solvation free energies used in chapters 5 and 6 is also described here.

C.1 Free energy of binding alternate derivation

The SSIMPLE [158] approach to solvation considered transfer free energies to derive the free energies of transfer. This approach is taken in this appendix.

A two phase liquid system, composed of phase 1 and phase 2 with a solute, m, in both phases is considered. For a single SSIP, i, in m, at equilibrium, it will have the same chemical potential in both phases, shown in (C.1).

$$[i_{free}]_1 = [i_{free}]_2 \tag{C.1}$$

Where $\left[i_{free}\right]_1$, $\left[i_{free}\right]_2$ is the concentration of the ith solute SSIP that is free in phase 1 and 2 respectively. The free energy difference of binding between the two phases for m is given by $\Delta G_{b,1\rightarrow 2}$ in (C.2), as discussed in [158]. This is the sum of the free energy changes to transfer all SSIPs in the molecule from phase 1 to phase 2. Note that the sign in (C.2) differs to that published in equation 15 in [158], which is incorrect.

$$\Delta G_{b,1/2} = -\sum_{i}^{N} RT ln\left(\frac{[i]_2}{[i]_1}\right) \tag{C.2}$$

Where $[i]_1$, $[i]_2$ the concentration of the ith solute SSIP in phase 1 and 2 respectively, N is the total number of SSIPs in m. The fraction free of the SSIP, which is in equation (C.3) can be used to rewrite equation (C.2) as equation (C.4).

$$f^i = \frac{\left[i_{free}\right]}{[i]} \tag{C.3}$$

$$\Delta G_{b,1/2} = \sum_i^N RTln\left(\frac{f_2^i}{f_1^i}\right) \tag{C.4}$$

Alternatively equation (C.4) can also be rewritten using the solubility constant for the ith species, $K_{S,i}$, which is defined in equation (C.5).

$$K_S^i = \frac{\left[i_{bound}\right]}{\left[i_{free}\right]} \tag{C.5}$$

$$\Delta G_{b,1/2} = \sum_i^N RTln\left(\frac{1+K_{S1}^i}{1+K_{S2}^i}\right) \tag{C.6}$$

Where $[i_{bound}]$ is the concentration bound to solvent; $\left[i_{free}\right]$ is the concentration not bound; K_{S1}^i, K_{S2}^i are the solvation constants for the ith SSIP in phase 1 and 2 respectively.

Confinement of SSIPs to a condensed phase results in the binding energy over estimating the probability of interaction. Consideration of a condensed phase where $K_{ij} = 1.0$ is required, as in [158]. The concentrations of SSIPs in such a phase are given by equation (C.7).

$$[i] = \left[i_{free}\right] + 2\left[i_{free}\right]^2 \tag{C.7}$$

Rearrangement of equation (C.7), yields the probability of a SSIP being free, P_f, if there is no interaction energy between any SSIPs in the phase, in equation (C.8).

$$P_f = \frac{\left[i_{free}\right]}{[i]} = \frac{\sqrt{1+8\theta}-1}{4\theta} \tag{C.8}$$

Where θ is the fractional occupancy of the phase, relative to the zero point solid state, equal to $\frac{[i]}{c_{max}}$.

The confinement energy, given in equation (C.9), is the free energy associated with restricting a SSIP to be in a condensed phase.

$$\Delta G_{c,i} = -RTln\left(P_f\right) = -RTln\left(\frac{\sqrt{1+8\theta}-1}{4\theta}\right) \tag{C.9}$$

The confinement energy for a molecule requires summation over all SSIPs; which is $N\Delta G_{c,i}$, since $\Delta G_{c,i}$ is independent of SSIP value.

The free energy of transfer, $\Delta G_{1\to 2}$ is shown in equation (C.10); with $\Delta G^1_{c,i}$,$\Delta G^2_{c,i}$ the confinement energies per SSIP in phases 1 and 2.

$$\Delta G_{1/2} = \Delta G_{b,1/2} + N(\Delta G^2_{c,i} - \Delta G^1_{c,i}) \tag{C.10}$$

C.2 SSIP Solvation energy curves: Polynomial description

$\Delta G_b(\varepsilon_i)$ and $\Delta G_S(\varepsilon_i)$ are smooth functions of SSIP value, so we can fit a polynomial function to it, to allow for easy interpolation between a selection of calculated values. This is used to increase speed of computation and reduce memory requirements for the analysis used to generate FGIPs in chapter 5 and solvent similarity in chapter 6.

The solvation free energy does show different behaviour for donor and acceptor SSIPs, so the fit has two regimes as shown in (C.11).

$$\Delta G_S(\varepsilon_i) = \begin{cases} f_+(\varepsilon_i)\ if\ \varepsilon_i > 0 \\ f_-(\varepsilon_i)\, otherwise \end{cases} \tag{C.11}$$

Where f_+ and f_- are the polynomial functions for the different positive and negative regimes generated using regression analysis. The split fit is required due to the large variation in interaction strengths of hydrogen bond donors and acceptors with a solvent. Figure C.1 shows the $\Delta G_{S,i}$ for the simulation and the values from an eighth order polynomial fit for N,N-dimethylformamide, which also shows two distinct gradients for the different solute regions.

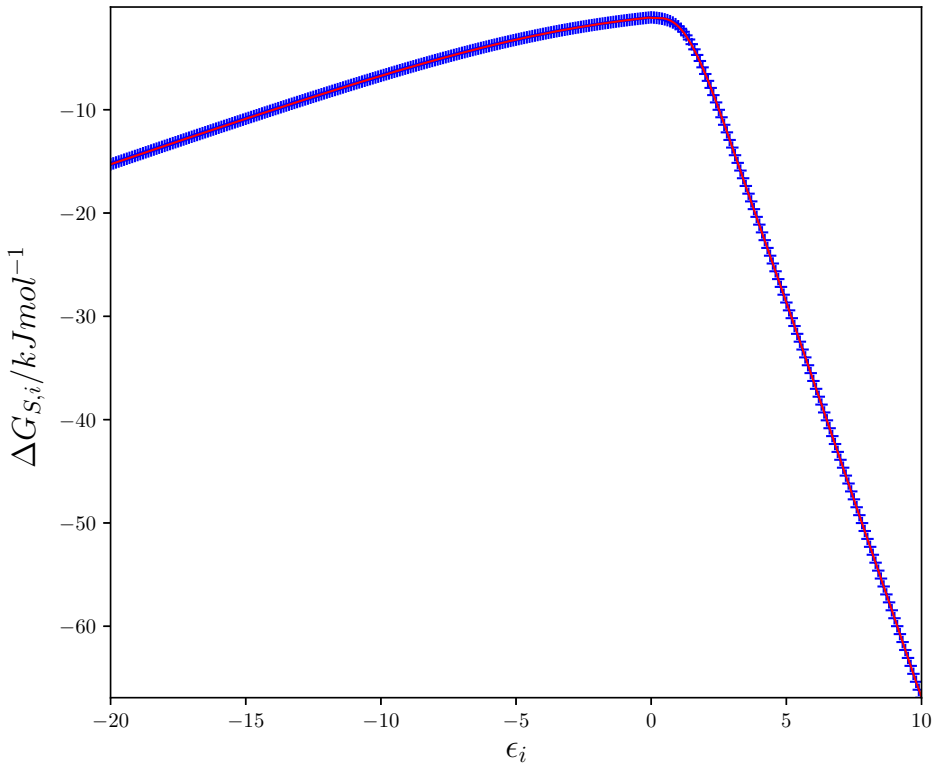
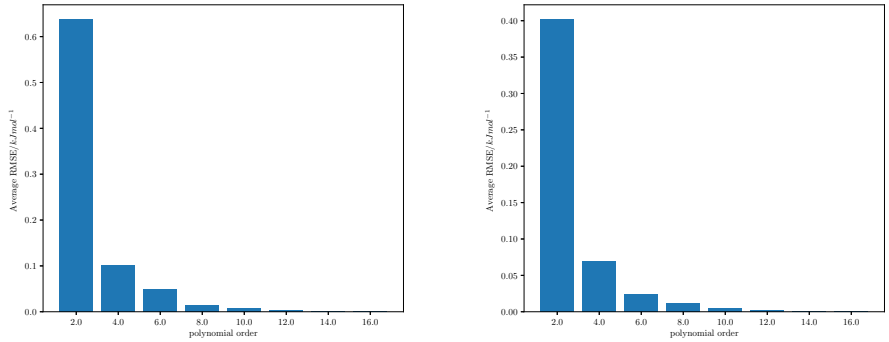


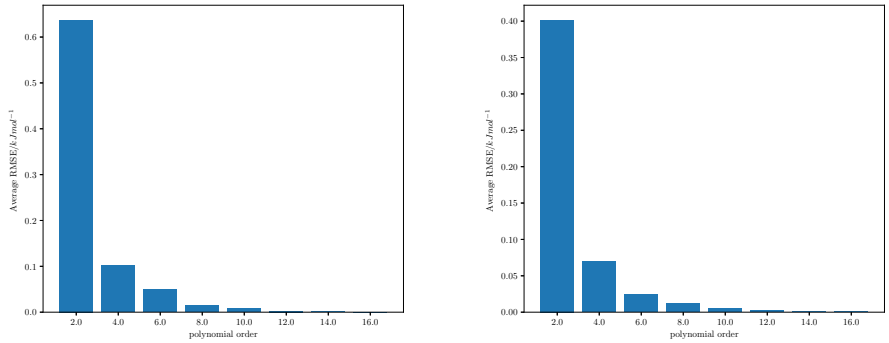
Fig. C.1: Plot of solvation energy for a solute SSIP in N,N-dimethylformamide at 298K. Simulation values as blue crosses, with red line representing the values from an eighth order polynomial fit as in equation (C.11).

An eighth order fit was used for the polynomial as this had a mean RMSE of under 0.05 kJ mol^{-1} for both positive and negative ε regions for the set of 261 solvent molecules in the pure solvent data set, listed in G.1. Figure C.2 is for the solvation profiles ($\Delta G_{S,i}$) and Figure C.3 is for the binding energy. The error in the polynomial fits are significantly lower than the errors in the solvation free energies when an eighth order fit is used, without overfitting of the data.



(a) RMSE of positive ε region (b) RMSE of negative ε region

Fig. C.2: RMSE for the polynomial fits to the positive and negative ε regions as a function of polynomial order for the solvation profile.



(a) RMSE of positive ε region (b) RMSE of negative ε region

Fig. C.3: RMSE for the polynomial fits to the positive and negative ε regions as a function of polynomial order for the binding energy, $\Delta G_{b,i}$.

The polynomials coefficients used in the generation of functional group interaction profiles (FGIPs), in chapter 5, as well as the similarity profile analysis, in chapter 6, use the polynomials in appendix G.

Appendix D

Database

A copy of the database can be found on the CD accompanying the thesis, which was created using [307]. 6441 property values were collected for 1470 molecules. Of these property values: 1001 are β values, for 770 unique molecules; 364 are α values for 275 unique molecules.

Appendix E

Partition Coefficient Data

E.1 Solvent descriptions

The solvent descriptions used for calculation of the partition coefficients and phase transfer free energies were the same used for the creation of functional group interaction profiles and solvent similarity, details are in appendix G. The descriptions were calculated with using the original parameterisation in [155]. The description of wet octanol used was a mixed 1-octanol-water solvent, with 0.271 molefraction of water, based on information from the SAMPL6 challenge [283].

E.1.1 Solvent Concentration

The concentrations of the solvent molecules are in G.1.

E.2 Partition data

Summary of data for partition plots. Experimental data is from the cited sources.

E.2.1 Dataset 1

This information is located on the attached CD.

E.2.2 Dataset 2- Reynolds

This information is located on the attached CD.

E.2.3 Dataset 3- Martel

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
ZHTYOISLJWAJCL-AWEZNQCLSA-N	4.17	5.29
NEVMHUGOCDUYCY-JTQLQIEISA-N	2.79	3.15
OGYZPWONPFHXSP-UHFFFAOYSA-N	3.96	4.44
HBWITNNIJDLPLS-UHFFFAOYSA-N	5.30	5.47
GCUGOKRWLBHMLQ-IBGZPJMESA-N	3.53	5.97
WQKMACGRTRICNN-UHFFFAOYSA-N	4.35	4.85
MYXZSSULOWIELP-ZFWWWQNUSA-N	6.21	7.11
ROQYSLYSLZPOBT-UHFFFAOYSA-N	3.14	3.71
OAJNZFCPJVBYHB-UHFFFAOYSA-N	1.81	1.74
XQLJKVUBWFWJHH-CYBMUJFWSA-N	2.34	2.92
GMCWXHUWHFUMQH-UHFFFAOYSA-N	2.98	4.11
GZJYYTDPEQCUOZ-OAHLLOKOSA-N	4.50	6.70
GNHSFPGRILAESF-UHFFFAOYSA-N	3.33	3.76
KKIVPOXYVUDHPE-UHFFFAOYSA-N	4.12	5.15
CZHBUODUVDGWLO-SNVBAGLBSA-N	3.96	4.70
ILDWRDGUEBWGLQ-UHFFFAOYSA-N	3.69	4.67
GKOUEWHYLFOVFG-UHFFFAOYSA-N	4.00	5.51
BMADQAUGPJFBPU-GFCCVEGCSA-N	3.24	3.10
PEGDWZSJRLDGOV-UHFFFAOYSA-N	1.14	-0.56
RCFRVQFKBHAWED-UHFFFAOYSA-N	5.38	5.76
CGTCXQOYTCJPIZ-UHFFFAOYSA-N	2.68	4.15
VCLVBIPTRZNCGF-YUMQZZPRSA-N	1.49	2.33
DVGGJTWGPZXYGY-UHFFFAOYSA-N	5.05	4.43
CSFNXBMHFQEMBP-FMIVXFBMSA-N	4.28	4.15
GOOSRBXROIWFPI-CHWSQXEVSA-N	3.30	4.82
PBSZHNXXFIYDBU-UHFFFAOYSA-N	4.41	4.94
HANDMCWHWMILMT-UHFFFAOYSA-N	4.14	6.12
HRQYWOOZOXTWOR-UHFFFAOYSA-N	2.46	3.41
SXJBCGWDRVUTRR-YLJYHZDGSA-N	5.48	7.33

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
CBWMOBFPWMOJKZ-QYVJPZLMSA-N	0.51	1.81
MHKHJIJXMVHRAJ-UHFFFAOYSA-N	2.61	2.96
NKCVVZUBNFITEG-ZWNOBZJWSA-N	3.68	4.30
YEXRTWOVQNJHLF-UHFFFAOYSA-N	1.56	-1.61
VCPGACZNGUNXMK-KBXCAEBGSA-N	2.46	4.49
FOCILOSWFFTFNL-UHFFFAOYSA-N	3.65	4.78
RJNURKQUNFBADL-CCEZHUSRSA-N	4.69	5.73
QEUFQFSKUSAQQE-UHFFFAOYSA-N	2.82	3.93
XTIUPBZFKHSICA-UHFFFAOYSA-N	3.90	4.60
FVLNTXOYQXQZHP-UHFFFAOYSA-N	3.92	6.44
QYMABJIOSDFFRG-DHZHZOJOSA-N	4.45	6.00
UAHAAFGEZOBXTP-AWEZNQCLSA-N	3.58	4.82
XCJIPKCPRKVCIN-KRWDZBQOSA-N	5.02	7.22
OLZITQXFEGHLND-NXEZZACHSA-N	2.44	3.85
ZGUBQUIWHJBVMV-GFCCVEGCSA-N	3.83	4.20
YEWVKBUFRDKCAG-UHFFFAOYSA-N	2.79	4.38
GSWSWFNVEVSOJC-SECBINFHSA-N	1.41	2.77
BQOCMHOWHOMIMJ-UHFFFAOYSA-N	2.34	1.72
CFGMGRLIFAVVMY-UHFFFAOYSA-N	2.92	4.50
VTRHVCVHBFETCP-UHFFFAOYSA-N	4.04	4.96
WUODFZJMRMJWIS-XUWVNRHRSA-N	5.87	6.57
NFCPRRWCTNLGSN-UHFFFAOYSA-N	2.67	4.07
YUTFQTAITWWGFH-UHFFFAOYSA-N	2.15	2.48
CISOKWVGMLHGBN-UHFFFAOYSA-N	3.04	2.95
WYCHMHPXGUPIBI-UONOGXRCSA-N	2.66	4.50
LKEZQXUSGJDRKD-FOCLMDBBSA-N	1.10	2.89
ZBNZMSUGDPAPSB-GIJQJNRQSA-N	4.81	4.46
TVWLCJVDJFYZSD-UHFFFAOYSA-N	3.45	2.84
RERURSHSFQDGJT-UHFFFAOYSA-N	3.61	2.56
ROQJDMYDPFGLEK-UHFFFAOYSA-N	4.97	5.22
FHVVYVTWZBXZBN-QWHCGFSZSA-N	2.85	3.55

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
DXFNVIVIEHJYQT-UHFFFAOYSA-N	2.70	3.72
XZERHOWNUXQBNF-VXNVDRBHSA-N	1.16	1.75
COIUQMDQIBGHQH-UHFFFAOYSA-N	3.33	3.89
ZORGBFBWSHUFAR-ZDUSSCGKSA-N	5.45	7.23
WHHFXQYCQIZJON-UHFFFAOYSA-N	4.33	5.16
MCOMUPQTXUHLTL-RDDDGLTNSA-N	0.71	0.07
IWZGIPVTSGGNME-UHFFFAOYSA-N	2.71	3.41
KNQRFJRNEDUILA-UHFFFAOYSA-N	3.24	3.60
AAVGUZZUJOHLGC-UHFFFAOYSA-N	5.86	7.84
DHKMJAHFQFYQLM-WCQYABFASA-N	4.76	7.15
ORKOCUGYWLTPIP-KGLIPLIRSA-N	4.64	4.79
CJBOKCHSMBRXAK-UHFFFAOYSA-N	3.20	5.29
ZYUUTRPXZHRTBC-UHFFFAOYSA-N	2.60	3.92
YFKZCWAATHXKMA-UHFFFAOYSA-N	3.19	4.09
BHVSCKTUCMZDIV-UHFFFAOYSA-N	4.89	5.36
XCLOVAUIMQGTOK-UHFFFAOYSA-N	4.52	6.07
YDYPMLDCUCWBKB-UHFFFAOYSA-N	4.69	4.62
ZVGRHWZDOBWBPQ-FQEVSTJZSA-N	5.77	7.80
IOYBVXGEPKKEBH-QHCPKHFHSA-N	4.38	6.74
AEXUGBKKZSVAPK-MGPUTAFESA-N	4.73	6.06
IKNZKDSAYXMKKQ-ZIAGYGMSSA-N	4.56	4.36
ZTAMDRVTGLTOKN-ZCXUNETKSA-N	6.96	9.13
CHNTVHWKBXYWAT-UHFFFAOYSA-N	6.15	7.04
LWEWLNJRHDOSHP-UHFFFAOYSA-N	5.92	6.20
HORBZFFNOVIHPI-YLJYHZDGSA-N	4.25	6.05
ALPGTOOTEBXLDZ-UHFFFAOYSA-N	5.01	6.59
AYGQLBHSQFMOEI-UHFFFAOYSA-N	5.86	7.84
UYZFELGGXDXOHW-RRFJBIMHSA-N	5.05	7.19
HCRFJPAQRKAART-JSGCOSHPSA-N	5.54	5.33
MUNBZSILFSBXQL-FSJBWODESA-N	5.16	7.89
QVDDCLSXTLUEQH-OUKQBFOZSA-N	5.53	5.87

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
DDRSZUUAUJASSI-RDJZCZTQSA-N	5.92	6.79
KZLDWBSCYBHPAW-HDICACEKSA-N	5.00	6.74
YWSZRAHHUHWDDC-XOBRGWDASA-N	3.07	5.03
LKSYGZBTKUVKQG-CYYJNZCTSA-N	6.51	9.93
INVKBFDNVBLXPS-UHFFFAOYSA-N	6.09	5.88
MLHPBLBIRWJQOT-UHFFFAOYSA-N	2.76	3.85
KKMUTMLAXBFINA-UHFFFAOYSA-N	2.89	6.03
JNIUASRGXBINOO-UHFFFAOYSA-N	4.63	5.41
BGZJVJOLKCRHGM-AYBZRNKSSA-N	3.65	4.98
CAYAHNKHBWVYHU-UHFFFAOYSA-N	4.62	8.49
AHPUQHAFKQTLEB-LQUAGZSJSA-N	6.00	9.41
SDMBWFDEPMYMGI-UHFFFAOYSA-N	3.93	3.06
LMSKFFORTXTXLR-UHFFFAOYSA-N	4.92	4.98
SFTXUGPQTNDPEG-ZWKOTPCHSA-N	4.48	5.70
VTKYHTUHVWMKSC-FQEVSTJZSA-N	5.04	6.01
NSQKTSFQLVEUHM-UHFFFAOYSA-N	4.08	4.16
UTLKAWYFHJSKQV-UHFFFAOYSA-N	4.41	3.67
YUCJLWRFTZJYIK-UHFFFAOYSA-N	5.75	7.40
INEZMZGOJOCSIS-QHCPKHFHSA-N	4.99	4.90
FAKKPEHAXNWGNO-UHFFFAOYSA-N	5.27	6.60
NRCNTMPERVVZPZ-GOTSBHOMSA-N	4.93	7.62
JXSDKBMQPWWPFP-VSGBNLITSA-N	6.65	10.58
QTZMVOQSYGIBAX-UHFFFAOYSA-N	3.69	3.72
JVTLAKZLKZFVEB-UHFFFAOYSA-N	4.83	10.29
WCJDQPQLNDQEAF-UHFFFAOYSA-N	4.55	5.05
TXKWKZXGRPPXJL-UHFFFAOYSA-N	3.55	2.72
MRPVKLSIDSIUDA-CYBMUJFWSA-N	5.04	6.57
URPZRDIJQKPJFZ-UHFFFAOYSA-N	4.48	5.26
HKVRLDHZKYJDFA-UHFFFAOYSA-N	3.95	5.94
ONGGCQFIGCVCSN-ZDUSSCGKSA-N	6.53	7.78
JJPYDFGDSYHKLR-UHFFFAOYSA-N	4.22	6.06

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
JZTMANRDBQZJEA-PBBNMVCDSA-N	5.77	7.13
ZQAHKDUZHISWBE-QHHAFSJGSA-N	3.72	4.32
FZNTXTCEOPRPML-UHFFFAOYSA-N	5.13	8.23
JGCYRGXRKXORRJ-UHFFFAOYSA-N	4.85	6.71
SVFKDRUYIIUIFE-UHFFFAOYSA-N	5.64	6.26
NRNCEYNQMPDMIW-JVWAILMASA-N	5.58	6.96
CWKKCWFXWWPOSH-UHFFFAOYSA-N	1.55	2.39
MDMSJQVAZRPBTH-FQEVSTJZSA-N	3.89	5.54
KQAVKBHOJFRDIK-UHFFFAOYSA-N	6.08	8.41
OUCUJHWBDDGFED-UHFFFAOYSA-N	4.16	5.33
QTGCYMOYQIUQDI-UHFFFAOYSA-N	4.14	5.47
IOQGCFQJRDDVBG-LRKLCTCQSA-N	5.07	6.85
JNEFIQXNVVUKLD-UHFFFAOYSA-N	2.77	2.17
WWJUWUYPLQLOEI-RGJNTOLQSA-N	1.48	1.45
SMXADHJFXXLOLN-UHFFFAOYSA-N	4.62	3.19
FGQPHLLMLBCMCK-UHFFFAOYSA-N	5.72	6.04
FGVMSGZDOANVQA-UHFFFAOYSA-N	5.35	6.66
CFWIEFRWFQHAQW-HXUWFJFHSA-N	4.18	5.79
YAMKVBLBJAAXFT-VWLOTQADSA-N	4.25	6.06
ZAKIKYBNFKWDDQ-XWRIVVANSA-N	5.29	7.59
GCFRRTILCGTGBQ-YLJYHZDGSA-N	5.50	7.93
UMPQEFOTUDZBJG-UHFFFAOYSA-N	5.47	6.47
KQMIMACSJDBGQK-ONGXEEELSA-N	4.47	4.55
CFQXQCSYUSXLDL-UHFFFAOYSA-N	2.09	2.93
POLMJDIEAXQIPX-BDTNDASRSA-N	5.98	8.45
DLRPHOSBFYTCAM-UHFFFAOYSA-N	6.85	7.88
WKDBMZQECSVNDS-UHFFFAOYSA-N	1.88	2.38
UGLGPUOORWARAH-ZCFIWIBFSA-N	2.75	3.12
DUOXRUNUARSQEB-UHFFFAOYSA-N	3.53	6.59
WBICRFXNOLHJLW-UHFFFAOYSA-N	4.92	7.01
UCEQJOWFMHDKKL-UHFFFAOYSA-N	5.47	8.86

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
NGYKAUYSNYJQEW-WKULSOCRSA-N	6.14	9.06
XPEXCQBNKAUOFM-UHFFFAOYSA-N	5.51	6.58
KGPXNTXVIMCKQE-UHFFFAOYSA-N	3.11	3.51
XZHJOCVCQOPOIY-UHFFFAOYSA-N	3.19	4.63
UYLIFRMAFKBLGA-UHFFFAOYSA-N	5.87	8.03
CAKSREDMTKQPNR-UHFFFAOYSA-N	6.72	10.66
ZHUATNWTFFMRCK-UHFFFAOYSA-N	3.68	4.39
YEOUKSXVZKMTQY-GOSISDBHSA-N	3.47	4.80
KXCHBLHOXKENCH-KBPBESRZSA-N	3.93	4.61
NQERVHWPAYNKBW-UHFFFAOYSA-N	2.30	2.87
KRVBIJNWBWCZRY-UHFFFAOYSA-N	4.31	6.11
RQVCHKZPRRZJCK-INIZCTEOSA-N	6.04	8.80
NTBDKTGBUJTAHX-UHFFFAOYSA-N	6.06	6.07
WOASUTLIZIBWDW-UHFFFAOYSA-N	5.56	6.24
HTPWJGVWDZMPLH-XDHOZWIPSA-N	4.25	4.72
ITWFHOWJWBBFJT-BQYQJAHWSA-N	4.67	5.45
YVTGQQYMCFXZOP-UHFFFAOYSA-N	6.58	9.57
CVMWEBSWRVKUDR-UHFFFAOYSA-N	4.42	5.06
XOMCFOVVVRBCRX-JOCHJYFZSA-N	4.55	5.87
XXVSGJDBVQTJAT-DTORHVGOSA-N	4.43	6.03
VYPPVOVFFZDOLO-UHFFFAOYSA-N	4.70	5.84
UNLXBPZUCFUUFH-UHFFFAOYSA-N	5.26	6.35
ZWQWUXJOSIEPAF-HMAPJEAMSA-N	4.04	4.20
HVPBYYHGVNFDHG-UHFFFAOYSA-N	2.45	2.16
MWSQBNLQCJIDNT-ZBFHGGJFSA-N	4.50	5.98
PNAWEPHPWYNHLG-KZULUSFZSA-N	4.55	5.94
FVXIXCWTNABXHN-GJZGRUSLSA-N	2.86	4.89
GKGGWGQAWVRXJJ-UHFFFAOYSA-N	5.14	5.88
IMEMJHICQKFFMG-UHFFFAOYSA-N	4.82	6.44
MIKZEONNIJGZLT-UHFFFAOYSA-N	5.09	5.59
KJOVSBNTQFQTIO-UHFFFAOYSA-N	3.62	4.14

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
ZNCVVZOMUWLFFN-ZANVPECISA-N	5.49	7.93
FWSVZRJCNNFMSR-UHFFFAOYSA-N	4.00	5.84
CJMDFPWXRRTBSV-UHFFFAOYSA-N	2.71	1.71
DERFFWPLVQTHEZ-YYHQMBLXSA-N	2.28	3.83
WPLKQUJNEKRTMS-UHFFFAOYSA-N	5.22	4.25
LBQCUXUTTHIVQP-UHFFFAOYSA-N	2.78	3.91
NSHXKNKAVFRJHR-QLENYVDCSA-N	5.87	5.83
PSFUCQBYNSURMM-PTDBIEIZSA-N	3.77	3.86
NZTIPCHEAMRWCN-OAQYLSRUSA-N	3.08	5.90
NNPHTJJEJWDYAK-VWLOTQADSA-N	4.21	4.83
YRDOMFQFSOICJQ-PEZBUJJGSA-N	4.62	5.08
OMJNHHOTHQLHNZ-GDLZYMKVSA-N	4.65	7.28
DBUWPYDSLYIBCA-JXALSKIBSA-N	6.25	8.35
NFHLXMZEEQHTEB-UHFFFAOYSA-N	5.16	6.54
ARCDRMYSNWSNEQ-CYFREDJKSA-N	6.01	9.30
XDMQFOFJMPRXRG-UHFFFAOYSA-N	5.68	6.33
IHEAMWFNSHEPHK-CCLHPLFOSA-N	4.44	7.70
CXQKIQPKNPSGDE-UHFFFAOYSA-N	5.45	5.26
SVTRDJBJMCOYEM-UHFFFAOYSA-N	4.68	7.42
KLPLWWRXBYLZHN-UHFFFAOYSA-N	2.93	3.25
BDRHHTPONBYHDD-GJZGRUSLSA-N	3.68	4.14
SFLDAQSATWBWJP-UHFFFAOYSA-N	5.34	5.11
MXFUIUPCZDUIPC-HXUWFJFHSA-N	4.18	5.56
HKKAJYQKDOSXKD-CAOOACKPSA-N	4.03	4.23
HDDFAGCFLQKITD-UHFFFAOYSA-N	4.54	4.63
WZVDLTJRQXYDGS-HXUWFJFHSA-N	5.25	8.22
DIRBFPKLPYJGBA-UHFFFAOYSA-N	3.87	3.50
KWKMYBXXAONWEO-UHFFFAOYSA-N	6.17	8.46
RHRZLSZWJLEFGE-BTYIYWSLSA-N	4.53	5.70
JCNNVRSNLKLUNU-JOCHJYFZSA-N	4.15	6.97
AQCBGMZTLIAKJI-JOCHJYFZSA-N	4.97	6.46

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
BSGDPNHSNFVIQI-UHFFFAOYSA-N	3.19	4.25
GRVALHCXOIYOPV-UHFFFAOYSA-N	5.59	7.34
ZPEIYUMLGKDPBB-UBFVSLLYSA-N	5.83	7.89
VSDNJYCAZXKHAJ-LLVKDONJSA-N	5.76	9.16
RWXUTVGACDKSDP-QPPBQGQZSA-N	5.83	10.12
SZBIPEGMKJHKTD-QHCPKHFHSA-N	4.98	9.28
WNOOCLQDMXSNLZ-VIFPVBQESA-N	4.07	4.53
HRORZUQIQZUQMI-RBUKOAKNSA-N	4.52	5.41
CACMYLJNIDZGKK-PGMHBOJBSA-N	4.64	7.93
INXOYAADWKOYAB-LLVKDONJSA-N	2.05	2.19
DSDPNAASNKSGJJ-UHFFFAOYSA-N	2.68	3.21
CHWBXRZVOJPPPR-UHFFFAOYSA-N	3.13	3.97
QDGKZHNFOUFZDW-XBVQOTNRSA-N	3.85	6.41
VWOZKEQMWIZNMT-HNNXBMFYSA-N	3.75	4.89
DIVVAQYRKSZEEW-UHFFFAOYSA-N	3.92	4.23
DDHOGDYFLHNTAA-UHFFFAOYSA-N	4.86	3.87
NFVMGWKOORQLNW-RYUDHWBXSA-N	4.25	5.37
RZSANSNOPKTWBK-UHFFFAOYSA-N	4.57	5.52
IRNBMLQJZCQHSW-UHFFFAOYSA-N	3.50	4.58
GLPFKDSABFADSV-BQYQJAHWSA-N	5.31	6.37
LLDZOOUNGYCZSH-QPJJXVBHSA-N	2.46	3.11
ZLXGVYXVUKDLNG-OAHLLOKOSA-N	3.48	5.71
CEEKAUKKCDPUKD-YOEHRIQHSA-N	4.94	6.21
YPWVYUCBFXPYRX-ONEGZZNKSA-N	1.59	4.17
OBFAAMQTFVUZJO-UHFFFAOYSA-N	5.83	8.43
FGVRQOGQXMNHGR-MRVPVSSYSA-N	4.09	5.09
ZTGKTXRVNVAAEZ-SNVBAGLBSA-N	5.83	6.47
UFJPMIQEOMDPAL-VKJFTORMSA-N	4.38	4.25
WTJAOSUCZCHUHA-UHFFFAOYSA-N	5.63	5.38
LTSCOYSDXSZJGC-OAHLLOKOSA-N	4.55	5.93
BFUGHTCDORQPRS-UHFFFAOYSA-N	3.30	3.93

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
KRJCCTZGOZLVHM-LLVKDONJSA-N	5.06	6.42
QRZFZAWIYOQACB-UHFFFAOYSA-N	4.49	3.15
GEHZJNMZLQKLDL-UHFFFAOYSA-N	2.74	3.64
XROGAVVMKSBCRD-SFHVURJKSA-N	4.55	5.50
KZOLLAACMDISDU-UHFFFAOYSA-N	5.43	9.22
UBWASVDNTIJASA-HSALFYBXSA-N	3.37	5.61
OIMISOGUXWNOTR-ONOMSOESSA-N	5.81	6.48
PGHGRJRFLRIDKV-UHFFFAOYSA-N	4.31	7.10
OTNJMHOYCFLNRI-UHFFFAOYSA-N	3.75	5.20
PCFAVTJTLIYZHT-INIZCTEOSA-N	3.95	6.42
YXVLATGCUIPYTB-UHFFFAOYSA-N	5.04	8.47
NBBJIPDAOKGSNA-IBGZPJMESA-N	3.99	5.34
SERNBTUTHRLYCV-GJZGRUSLSA-N	4.19	6.20
YOVXPJSGNSFMKE-UHFFFAOYSA-N	2.95	3.72
YNJPORWECCVABQ-ZWKOTPCHSA-N	4.48	4.75
JMJCHRSOBZHAHJ-OAHLLOKOSA-N	3.14	4.57
LMYGHJZQDHRZFC-UHFFFAOYSA-N	4.00	4.50
KTGUPAFUVGHOQS-UHFFFAOYSA-N	1.74	2.63
KHSORWOFDJUUMT-UHFFFAOYSA-N	5.70	6.71
BCSQUJYENGABAT-UHFFFAOYSA-N	2.93	2.34
FABBYUIQYBYMEC-PMACEKPBSA-N	5.49	7.81
DFWQFBNINNFHPH-UHFFFAOYSA-N	4.59	7.23
ZXPRSEABWUIJCK-HNNXBMFYSA-N	3.90	6.09
XAIDHRILHSJWHP-LJQANCHMSA-N	5.10	6.84
LCCDJOCVVMFOHD-UHFFFAOYSA-N	3.20	2.35
IHUUTTHUGLJFAF-UHFFFAOYSA-N	2.64	1.58
VUNZVDNKOMJTCK-UHFFFAOYSA-N	3.25	3.40
WHVPISQBOXRCHR-DFEHQXHXSA-N	5.24	6.87
AQSUNOZBYRWGBO-HXUWFJFHSA-N	4.48	3.09
KLAKGOHDTGZVKU-UHFFFAOYSA-N	4.34	5.18
KNTBAPSEJMDMLH-PFBJBMPXSA-N	6.47	8.87

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
STUBXTBZJGURQS-UHFFFAOYSA-N	3.02	5.46
TVUHCGKTUZXUJY-MRCUWXFGSA-N	6.36	9.15
DNKNARHPWPLDBY-UHFFFAOYSA-N	3.30	4.27
ROUZTSOXZWQINV-OAHLLOKOSA-N	4.58	5.72
MZDCJAPBIUAZNA-ROUUACIJSA-N	3.74	3.65
SEPSLMVKQVHQNT-UHFFFAOYSA-N	5.96	7.37
VZLDLRDTGAFUKP-QMMMGPOBSA-N	4.79	5.86
CPIDUHGJETUEKJ-FCHUYYIVSA-N	5.56	5.22
XCJXQCUJXDUNDN-XNVXBWHWSA-N	5.69	6.35
NVWFNGRWBYLOQL-HAHDFKILSA-N	4.66	7.48
AJNYEMLGBXCAJD-UHFFFAOYSA-N	3.04	3.50
CLBOJFVOEQYSIF-HWKANZROSA-N	1.61	0.70
XOJDPDUKYKQGLY-NXZHAISVSA-N	4.59	6.21
AZQFHIOYKBSTLM-UHFFFAOYSA-N	4.64	4.91
LCMAQUWOPVHYJP-MHZLTWQESA-N	5.45	6.12
YCINBPWVFJKMKA-GFCCVEGCSA-N	3.68	5.23
BYLPAFIYZWIQCN-QFIPXVFZSA-N	5.32	5.63
MJZVQRMHVMIJDZ-QHCPKHFHSA-N	5.43	9.69
IMRZTJQUCNZSCF-UHFFFAOYSA-N	4.25	6.82
XXLXNUKFOGCGGI-UHFFFAOYSA-N	4.43	4.56
MNEJNVRRUICAQU-UHFFFAOYSA-N	3.72	8.39
YZGUMMAITZYKOD-UHFFFAOYSA-N	2.89	3.02
YOCFCLLRCGSJIU-UHFFFAOYSA-N	4.61	4.23
FOZZNDRTRNDJBC-UHFFFAOYSA-N	4.16	5.00
QHIZTQZAHBUIEK-LFIBNONCSA-N	6.59	8.30
OPUIFMOEKSUETI-GASCZTMLSA-N	4.91	5.96
HCJFDHHBUCXINJ-CYBMUJFWSA-N	3.43	3.98
MFEWOASOCGWVJY-CQSZACIVSA-N	5.05	8.89
FESKORARFOGMKM-UHFFFAOYSA-N	5.60	8.72
DRLNFFKVXNSCFI-FYWRMAATSA-N	4.92	6.54
MDZQHXDFKBVUGZ-UHFFFAOYSA-N	4.84	6.16

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
JFSCYNSDKYRXPM-MHZLTWQESA-N	5.39	6.42
HIMPVFWXMRWJNG-UHFFFAOYSA-N	4.61	5.70
XIJNSAGCFRMZDO-UHFFFAOYSA-N	4.28	5.60
ZDNPASVKQISNOI-CQSZACIVSA-N	4.48	6.77
BSLPDPNWYWWKDM-RYUDHWBXSA-N	2.69	4.05
ZTVVEQXQBOATSI-UHFFFAOYSA-N	3.86	4.44
BRPXBFGWZBBNCG-ACRUOGEOSA-N	4.00	4.33
LABFHIJXDBVOEM-GOSISDBHSA-N	4.75	6.90
UWEABYAGFQFZFW-LBPRGKRZSA-N	3.10	3.33
ANSZZPJMTBLPCY-UHFFFAOYSA-N	0.86	1.55
UQIRGIDFUWCHDC-MDWZMJQESA-N	4.94	6.45
ZDPSBNRWMXTNCI-OAQYLSRUSA-N	5.31	8.59
WDCOJJCXZXFZJZ-KRWDZBQOSA-N	4.18	3.66
FDEMLXAFJTWNQD-UHFFFAOYSA-N	4.13	6.30
HBTOYNLTBSWHKA-UHFFFAOYSA-N	4.73	6.58
PWRSQLHEWLIIQP-UHFFFAOYSA-N	2.31	4.41
QRPPVVILVWRKMT-CVDCTZTESA-N	5.28	6.55
ZOQPZXUCYJGNGZ-UHFFFAOYSA-N	5.99	7.80
KVXNSYHNCCGHAQ-VQCBNXJZSA-N	6.26	8.00
WVDXMBBVTMQIHU-UHFFFAOYSA-N	3.71	5.15
PQENZYXOUZMDFI-UHFFFAOYSA-N	4.34	6.17
CZZXDIHZMJQBGQ-AIYVTKCESA-N	3.39	4.24
UKOXYLFGSFPIKY-UHFFFAOYSA-N	3.47	5.81
FYSQOPNXZUHHBZ-UHFFFAOYSA-N	5.03	5.24
AHEBMWPPIOTQOC-LJQANCHMSA-N	5.56	7.34
IZXBODQUOMRHIG-HQQGHWSLSA-N	6.66	6.10
NOPWOZZLZKHIRI-KLHWPWHYSA-N	3.11	4.72
NBWRFBZIRGPFSC-UHFFFAOYSA-N	2.93	5.24
ITTAHRUGUAGBGK-DQRAZIAOSA-N	5.14	5.02
NALFHGJNUXROMT-GHVJWSGMSA-N	5.84	5.87
DMKCZUXCQWLBOI-UHFFFAOYSA-N	3.27	4.89

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
HYXGIGVZSVPNGR-DEOSSOPVSA-N	5.07	7.83
GKPVOQLJHWLOIE-UHFFFAOYSA-N	5.85	7.63
DUDAZNFMQDQCTJ-FQEVSTJZSA-N	5.41	7.57
AAQGPMOPIQKGEL-UHFFFAOYSA-N	3.74	5.55
CPXSZGKFGNBILI-UHFFFAOYSA-N	4.99	5.08
MPRBOWAGTINOPY-UHFFFAOYSA-N	2.63	3.67
QGWMNAWELIVFTG-UHFFFAOYSA-N	2.78	3.00
ZTPCZESYHLPHTL-CQSZACIVSA-N	3.75	4.09
UKTLLYCWSFLLTQ-QGZVFWFLSA-N	3.73	4.89
YJXKSJWDDFJERR-VCHYOVAHSA-N	3.08	3.40
DCYUXNHCVBQGHH-ZDUSSCGKSA-N	4.27	5.49
KZTABLWIZXQULZ-SNVBAGLBSA-N	3.33	2.80
LGNNLEKUCHDMHW-UHFFFAOYSA-N	3.27	4.72
JUJFXCLBPZBDMK-HZPDHXFCSA-N	3.42	5.28
MNTNYGIDPGQTOC-OAHLLOKOSA-N	4.36	7.27
ROUIFMYBELJXEB-HNNXBMFYSA-N	3.59	5.04
HMPCZKBDXOYXEK-HXUWFJFHSA-N	4.42	7.21
ZCMGFFOIGKAQSD-FIFLTTCUSA-N	3.97	4.21
YIJYNRMUUBIRAZ-GHMZBOCLSA-N	2.80	3.23
BZUWNEFCJVOJHV-VIFPVBQESA-N	3.74	3.51
ADCZQRGTZNBXSY-UHFFFAOYSA-N	3.27	5.53
WBEXFXGSYBGDKO-SNVBAGLBSA-N	3.12	4.53
XJGNDNFOGCBVCX-CYBMUJFWSA-N	4.83	6.59
NPMKARLPTSLAKF-PHIMTYICSA-N	4.11	5.80
NMYZUPRJTGVYOA-UHFFFAOYSA-N	5.81	8.00
GQPQLHBCZMCHEZ-QMMMGPOBSA-N	6.63	7.50
ZXICNMNZPJRBPN-QMMMGPOBSA-N	3.08	3.87
ZPXKNIORQOXHPT-JKSUJKDBSA-N	4.30	4.51
WMSUKVNVAIOYTI-UHFFFAOYSA-N	4.43	7.45
NEYKKGVGPSFWOW-UHFFFAOYSA-N	1.85	3.31
RVTSNQDYGDXALU-UHFFFAOYSA-N	3.53	5.20

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
WXRPRGUHSRCLLM-UHFFFAOYSA-N	3.54	5.37
WTCNAYPUXMJTKU-HXUWFJFHSA-N	3.89	4.01
UUMRHEMBQPWFEG-NSHDSACASA-N	3.73	5.17
QSNPETXBWYINSB-UHFFFAOYSA-N	3.85	5.72
YQPMKLSMDZIKIV-ZDUSSCGKSA-N	4.52	5.98
KYIIQLZELBWXGZ-UHFFFAOYSA-N	4.92	4.56
NAPXWOKKOKBBFI-UHFFFAOYSA-N	3.90	3.02
LWEJOHOGFKEJOR-MRXNPFEDSA-N	5.16	6.13
ZPDMNDVAWLUIAG-UHFFFAOYSA-N	5.77	7.26
WFLQLAWLHFSZCD-UHFFFAOYSA-N	3.77	4.53
DVIOKQKRRBJPEK-UHFFFAOYSA-N	1.04	3.49
FEALGFZIOWUCBX-GOSISDBHSA-N	3.62	4.80
GWHMRPRMYRJABK-UHFFFAOYSA-N	4.43	4.52
HIMGRAHFJJIWAE-WAPOTWQKSA-N	3.80	5.11
OPUXZJDYQMKOMJ-UHFFFAOYSA-N	4.66	6.49
XREOKRVOJWYEGN-UHFFFAOYSA-N	3.45	4.54
HOMZNWKHRFOTDB-UHFFFAOYSA-N	4.62	7.05
GFTIFHCOENKKEV-FNCQTZNRSA-N	4.03	3.59
LRWUGAPSLMMXPA-QGZVFWFLSA-N	4.26	5.33
FZAWHHFNCLCWFV-FQEVSTJZSA-N	4.93	6.65
FSKUVTXSZTZMPW-JFMUQQRKSA-N	2.95	6.72
WUFVSQQGAYLGOP-MZYLBHOOSA-N	5.05	8.91
ZJQUDWDDZQAYJT-UHFFFAOYSA-N	3.20	2.81
CXEQEWXYGGKYDI-SFHVURJKSA-N	4.09	5.62
QCVIRJQIFSJLLW-ZEQRLZLVSA-N	4.37	7.73
QDBJFRLQQSVCQR-JYRVWZFOSA-N	2.14	3.11
OXGKYQKRPIKUSQ-UHFFFAOYSA-N	5.08	8.15
CIRWZGGESKUJHD-UHFFFAOYSA-N	4.73	4.89
HBUMYZBRQSSBIF-OAHLLOKOSA-N	4.49	5.08
XVNFPKKAPYTPDI-UHFFFAOYSA-N	4.57	6.69
HTYDUJRXOVWOKK-UHFFFAOYSA-N	3.44	3.69

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
KJGFUVLCCMTQTR-JMIUGGIZSA-N	4.93	6.40
NZRFSPNKJXIAHT-HXUWFJFHSA-N	5.63	7.22
TVBDIDDDJKLNIM-UHFFFAOYSA-N	4.24	4.53
IPHRCEPGBRGPPH-QHCPKHFHSA-N	4.34	7.30
USJBIIUVAKUIEG-UHFFFAOYSA-N	5.38	6.81
OOGLOENQRAUVGN-KRWDZBQOSA-N	3.85	4.59
WFSNJEDEIOZDHH-UHFFFAOYSA-N	3.80	3.84
XAPARILERDDUIA-RNCFNFMXSA-N	2.28	3.64
PPFVMXCGNLPTNO-UHFFFAOYSA-N	3.72	3.17
AZBUKWAUCAIIJN-UHFFFAOYSA-N	4.91	6.64
WFLKEQDJTAPERD-ZDUSSCGKSA-N	2.68	1.85
PPDOYIPVQSJRJV-UHFFFAOYSA-N	4.34	5.60
PEEXVCWBKQZBGA-UHFFFAOYSA-N	5.98	8.85
HCZZWVMQPZFMRX-UHFFFAOYSA-N	4.47	5.04
XJESGUOVKGXOEP-IAGOWNOFSA-N	5.51	7.99
CXNPTWLUNSEKLS-LLVKDONJSA-N	3.48	5.25
KKFZHSHXFMQMAT-UHFFFAOYSA-N	5.13	6.34
JWYSXYSJUHOIBO-INIZCTEOSA-N	4.44	5.96
KUKLHSJZZPQRCM-FQEVSTJZSA-N	4.21	5.57
XAAOHIZGIKSWHV-ZBXLFGCISA-N	2.18	4.54
ZKRRXUIMZUVGSN-UHFFFAOYSA-N	2.69	3.39
MJODVAOBVHVUME-WGDLNXRISA-N	3.79	5.62
ORRIBHZQKOAYQB-TYZQSYOASA-N	4.82	5.65
GZLCETKUTVIGQQ-UHFFFAOYSA-N	3.72	3.01
ZVDVVMSTHWWLHG-XZOQPEGZSA-N	5.42	6.34
HVRJMCIXBUWZCB-SFHVURJKSA-N	4.44	6.94
ULHCSQJPZWUAIA-MRXNPFEDSA-N	2.48	4.18
KEGLPYHWYBPPKK-UHFFFAOYSA-N	3.25	5.31
NIUCKOXIWCRUEH-UHFFFAOYSA-N	2.87	3.86
IWYAMGDQQATQPW-DWXRJYCRSA-N	3.79	7.45
RADPPQIJXOCSGQ-UHFFFAOYSA-N	4.46	4.66

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
BLCWNIGSFBOPQU-UHFFFAOYSA-N	2.56	1.88
DAOOKTDKSRSHHO-UPCLLVRISA-N	4.62	5.24
RBFMSOZXUUNGHX-UHFFFAOYSA-N	2.68	3.80
KAGAULBNAFMDFT-SECBINFHSA-N	1.43	2.94
NVEDNAPMXDKXSK-VIFPVBQESA-N	1.85	3.34
VGFIKAGLGYIDNO-SJORKVTESA-N	2.96	5.54
IWIZLHUNPZWOLU-UHFFFAOYSA-N	4.05	5.86
IRLHGKUNRKPBTD-LEWJYISDSA-N	3.95	6.75
PCPZUPPJCAFQRB-SDQBBNPISA-N	5.80	6.11
LNLWVMUKDKIWLS-CXUHLZMHSA-N	3.18	3.89
CUTVCQQLTOCFRM-UHFFFAOYSA-N	5.96	7.97
TUFRLCLSGMQEAK-UHFFFAOYSA-N	4.15	4.24
OZCAWKDKOKHTKN-UHFFFAOYSA-N	4.11	4.59
SOOBMJKCWOLEOI-UHFFFAOYSA-N	4.24	4.95
JJMIZFIRKGFFBT-BYPYZUCNSA-N	2.84	4.22
OOAOSOAEXSJSJL-MSOLQXFVSA-N	3.31	5.74
XHRGGNHVBWBJKI-SFHVURJKSA-N	3.97	3.97
MLQMNOGMLXHIJV-BRNYJPRKSA-N	4.71	3.58
XZZQAAPHFLQVLN-UHFFFAOYSA-N	4.93	4.53
BWONCUGSZKCCIK-DOTOQJQBSA-N	4.80	6.64
WOUPSKHTIXFKFF-GOSISDBHSA-N	4.90	6.37
KMNVRCWPZINLKV-IUXPMGMMSA-N	2.52	5.76
TXMIDRACNKSZPW-PXRHPAFUSA-N	5.45	8.90
YJYSTHAIRHSORF-UHFFFAOYSA-N	2.05	5.27
PTCSWPFIQHUFOE-JTNHKYCSSA-N	4.34	6.53
WXPHQFQCOBBPIO-VIFPVBQESA-N	3.56	3.41
GRYLSQBVFCYQBZ-UHFFFAOYSA-N	4.75	5.48
YDRBFHSWIJHFLS-UHFFFAOYSA-N	2.47	3.86
CHBFEQWFFHDSSP-UHFFFAOYSA-N	5.89	7.82
GYALPDCCJCQAEN-HSZRJFAPSA-N	5.60	9.08
XSWIQGFJJNCSFI-KRWDZBQOSA-N	4.92	6.70

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
CWECLZCKMKKXJQ-QNGOZBTKSA-N	4.41	7.76
OVDBHQYUKOYVOQ-UHFFFAOYSA-N	2.55	4.34
SEWJZKJXAAKTFG-CYBMUJFWSA-N	1.98	3.55
DLTNARWOWSMBRQ-UHFFFAOYSA-N	5.35	8.64
JYEJKRCNEMYMEF-VYYCAZPPSA-N	5.68	6.55
VVPWYMMLGFSQBL-HNONPPJISA-N	3.25	3.90
UTXUGRIYKFTWSE-DLBZAZTESA-N	3.56	4.39
DJPBASBZLCQIJC-VQTJNVASSA-N	4.34	3.92
PXKDIPHYHWDCDS-GASCZTMLSA-N	4.39	4.81
IJAMGEZYGPUHRK-UHFFFAOYSA-N	4.98	7.70
AKISEPIDMROWEK-UHFFFAOYSA-N	3.68	3.52
GELXZTXDAVDBAG-UHFFFAOYSA-N	3.43	2.87
RZKIOQDMVCVPMS-UHFFFAOYSA-N	5.31	7.21
HDYIZCUZYIIPKF-UHFFFAOYSA-N	5.17	5.79
YNFKFTQWCBXILI-UHFFFAOYSA-N	4.46	5.29
MBMVNKPLKUBNTH-JTQLQIEISA-N	4.97	7.07
QCPMXJGMPWCZJI-UHFFFAOYSA-N	5.59	7.75
KVDXSPZBMCWJLL-YFKPBYRVSA-N	0.61	2.01
KEJDGVUGSFBSTF-UHFFFAOYSA-N	1.56	0.92
ILEYOVVZWSNYBD-VXGBXAGGSA-N	4.38	6.58
QELXXCRQIHCPII-TZFALNRWSA-N	4.96	6.64
JGPOKGKCKBYCSO-UHFFFAOYSA-N	3.63	5.03
QNPSPPUXZSNJQQ-WUXMJOGZSA-N	5.36	7.16
ZRBNCQSSHOZQIA-CYBMUJFWSA-N	3.61	4.90
VKWSYMTWSQLAEF-UHFFFAOYSA-N	4.59	3.76
SVKPQZITSXAYOU-MNEOSPDGSA-N	5.40	6.64
FPFZEYWOKLQUAU-OAQYLSRUSA-N	4.54	6.89
JZVBYNZPQLWWNZ-OAHLLOKOSA-N	1.14	1.95
TYUGYIMCRDPMPJ-UHFFFAOYSA-N	1.48	1.41
BTEKQMROJFGITO-UHFFFAOYSA-N	4.47	7.22
UCKKDRLPGJIQSZ-ZVHZXABRSA-N	6.80	8.35

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
LXVOVLFIXAASDZ-UHFFFAOYSA-N	4.10	6.46
CFTRIZQAAJPXQE-UHFFFAOYSA-N	4.63	5.37
PNIFKGARWDRNGN-INIZCTEOSA-N	6.01	6.55
YONZWQXETNXNCR-UHFFFAOYSA-N	4.55	5.09
NCOQXRXMRLIQEY-UHFFFAOYSA-N	4.34	5.82
DPNXCAIGIMLFBE-ZDUSSCGKSA-N	3.90	3.99
RTMFAKDIZRPSAS-UHFFFAOYSA-N	4.99	6.83
YVSWDPCIYMZTPY-IBGZPJMESA-N	2.41	6.04
BUDUREKJMNEQPJ-UHFFFAOYSA-N	5.04	6.46
NQIRDAFLQJGUDK-UHFFFAOYSA-N	5.87	5.59
UGDRPQKXROPZSI-ZZXKWVIFSA-N	3.64	3.26
SNDKOCUTXFLZEF-UHFFFAOYSA-N	4.15	3.79
YXVRGBUHSOBVHY-UHFFFAOYSA-N	3.90	4.70
CPHOVWSNRJFISA-KSSFIOAISA-N	3.65	5.65
UIVWACWKCZLFJN-OAHLLOKOSA-N	3.56	6.86
FOZKGUULHURPKE-UHFFFAOYSA-N	4.33	2.93
LTDPUFMRVCFSNR-UHFFFAOYSA-N	5.31	7.22
ABBBQZTWZGGQMV-GFCCVEGCSA-N	3.35	4.84
KPBQUJHQPKXRGM-UHFFFAOYSA-N	4.89	6.32
MCXVVSQVUOBOQP-UHFFFAOYSA-N	2.00	0.83
SBHZPCZOCQUGMB-TTYLXTBDSA-N	4.95	6.63
UILOOYAFYSPHBY-SFQUDFHCSA-N	3.83	7.45
RWWKXMFLKNKNMZ-UHFFFAOYSA-N	3.37	6.07
PZUFTAHZHKKUQH-UHFFFAOYSA-N	4.69	6.74
HNNNVJFOFHGYTP-UHFFFAOYSA-N	4.43	6.25
ZFNLCJAOJJZMEP-HCGXMYGOSA-N	4.98	6.02
UKAXOFBQGPETQX-UHFFFAOYSA-N	2.17	1.84
XPJZKWRGHILEER-UHFFFAOYSA-N	4.09	5.23
WIJNKOGONVYBGY-CMDGGOBGSA-N	5.65	6.16
VJOLDPDJEOPQEJ-FPYGCLRLSA-N	5.08	7.30
IVHCENOEVQKPAJ-UHFFFAOYSA-N	5.39	8.20

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
KKUXNMSOKNFPNO-HNNXBMFYSA-N	3.28	4.89
YYDDJSAOOYOSCE-UHFFFAOYSA-N	5.21	4.68
JENCNZDUWJGIHF-UHFFFAOYSA-N	4.32	3.86
JTKRRRFBPXEIRJ-JTHWEQIXSA-N	3.17	4.86
KOVQRPKIMYZUMM-XZOQPEGZSA-N	6.01	9.06
SNFVYUDGJHCYHC-QZTJIDSGSA-N	4.38	6.61
IWGLMZNQDGNFCR-CMDGGOBGSA-N	5.20	5.72
LLMZJFAVXNCYMB-AWEZNQCLSA-N	3.81	4.78
AFHZHIYYMVWJKX-DXRVJIQQSA-N	4.27	5.74
UORRYFRRFINENK-UICJJRBLSA-N	5.53	7.76
IKKKFHFINZEPJS-UHFFFAOYSA-N	4.20	4.75
UVLKINHWLRXFFS-OAHLLOKOSA-N	3.32	4.28
YIKWDEDBUNSBOG-RXMQYKEDSA-N	2.12	2.09
JEUWOFGICFQRDC-SFHVURJKSA-N	4.09	7.59
UAQFYHCNUOSISM-UHFFFAOYSA-N	4.15	5.21
RRNKOOVIXANIQE-UHFFFAOYSA-N	4.71	7.03
MZXIQSBGQNWMBD-UHFFFAOYSA-N	4.28	5.59
JOIPGSCLRUCANG-KRWDZBQOSA-N	4.43	4.01
PJIOEFNPYZVAIO-UHFFFAOYSA-N	3.52	2.30
BJSAANIPBRXNLX-UHFFFAOYSA-N	2.26	3.13
QRMPGBJUZLXCKJ-UHFFFAOYSA-N	3.20	6.32
HEAYNZXNTGFSBT-UHFFFAOYSA-N	3.63	5.59
TVEBSRJZJSFOEQ-UHFFFAOYSA-N	3.55	3.02
BKNOZYOZRQYDJS-UHFFFAOYSA-N	3.78	5.64
KYPSRTCGYAUKRX-CSKARUKUSA-N	4.37	5.14
XULJJMCFNXRIKG-UHFFFAOYSA-N	3.57	3.05
SBCRDKRHZZAZNS-KRWDZBQOSA-N	4.04	6.03
FARNZWAPBICULX-JTQLQIEISA-N	2.18	3.65
KFDGDMRKBPBWHK-UHFFFAOYSA-N	2.27	2.82
OQNHJVQPZHWXJF-GOSISDBHSA-N	3.81	4.85
CUACHGSCSLHFGZ-UHFFFAOYSA-N	4.83	7.86

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
LBAFIEZOVLWYFZ-UHFFFAOYSA-N	4.92	6.91
LNOZQWIYHJQZCH-UHFFFAOYSA-N	3.41	6.43
MMNUXUOHYOFNBF-UHFFFAOYSA-N	4.73	6.32
IKPPQLDVGJTNCX-SECBINFHSA-N	4.52	6.51
YYELOBYDLMZBBQ-MDZDMXLPSA-N	4.30	5.17
PIEVAXRKLVXHQE-NSCUHMNNSA-N	4.40	4.75
UXNNRIWFHIWDPT-UHFFFAOYSA-N	3.60	5.16
GZQNBRHKYZXTRC-UHFFFAOYSA-N	4.21	5.71
SPTYXSCPYMRLLO-LLVKDONJSA-N	3.93	6.20
PWFDGTODQUGYHA-UHFFFAOYSA-N	4.78	6.25
SMNYOFAKPWSNOT-NTCAYCPXSA-N	4.34	5.04
LJPYZDKKFNSJEP-UHFFFAOYSA-N	3.18	4.12
NEEOGUOWXUGJOW-CYBMUJFWSA-N	4.09	5.64
OOFQHRCDCIKPPZ-UHFFFAOYSA-N	3.45	4.40
ZOOSCKVBKUXJNL-UHFFFAOYSA-N	2.57	3.10
SGDNURDNAGNIQA-UHFFFAOYSA-N	3.66	4.37
WNZSTHZQQPLYJE-UHFFFAOYSA-N	2.09	4.46
KBCMVDJECURDCQ-UHFFFAOYSA-N	4.48	5.59
HGTCZHGLSORINZ-UHFFFAOYSA-N	4.85	8.04
VGAHNRVZEUKAQI-HNNXBMFYSA-N	4.73	7.93
FNZSEIHPROENQZ-GQCTYLIASA-N	4.34	5.82
RNSUJULKSRLDKB-KBPBESRZSA-N	3.68	4.75
RFTUAICFFKVGAW-UHFFFAOYSA-N	3.56	4.85
SVHGOMBMZTWUKI-UHFFFAOYSA-N	4.35	7.31
KUOUTKXTVFKIMJ-UHFFFAOYSA-N	3.64	4.31
PWZYOWWGTQCTJT-CYBMUJFWSA-N	3.62	7.17
APBXPFPORVVLAM-UHFFFAOYSA-N	4.77	6.36
DBOVDDBYSUDFAU-UHFFFAOYSA-N	4.01	6.06
LZJUFSSTQXVBNZ-CQSZACIVSA-N	3.40	4.56
RFDSRVLKUXXOFR-UHFFFAOYSA-N	4.47	5.81
ZTFXCVWELNWUNG-UHFFFAOYSA-N	3.28	5.76

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
BOMSQFAAAZIQHP-LLVKDONJSA-N	4.77	5.63
FKVFUEKNZQBMRE-UHFFFAOYSA-N	4.64	5.09
CXGQEIVOZAKTFV-UHFFFAOYSA-N	2.00	2.57
JIUOPTXXGIXPMF-UHFFFAOYSA-N	4.17	5.31
DQPOYJGJZXULIA-NSHDSACASA-N	2.78	3.82
XBBNCCSGZQZHKB-UHFFFAOYSA-N	4.45	5.77
UNMNBRHGMDYLBT-UHFFFAOYSA-N	4.41	6.27
CONZAJPFHQVMJF-UHFFFAOYSA-N	3.39	4.57
HVFUALGOZLHNBG-JSGCOSHPSA-N	3.11	4.88
QAWNGMRTIHOQQI-GHMZBOCLSA-N	2.48	3.62
HTXNKCYBFXFBIZ-UHFFFAOYSA-N	5.14	6.08
DFOLGYLYIGLVRR-UHFFFAOYSA-N	4.26	3.83
HBCXUIGLMSNEFH-UHFFFAOYSA-N	3.82	4.68
VWKZGTMKCMFQLR-YVLHZVERSA-N	4.32	5.60
TUBOJVNASSNRCV-UHFFFAOYSA-N	5.48	7.60
DPMLUPHERVULNR-UHFFFAOYSA-N	3.79	5.76
MNKBPFLCLCXRLW-UHFFFAOYSA-N	5.09	6.49
CRPDGTKHEIKNSK-UHFFFAOYSA-N	4.17	5.16
LFAKWYLPUUIGIF-VXGBXAGGSA-N	4.55	6.10
GIXSHWSOHUJCQH-UHFFFAOYSA-N	4.69	6.11
HDUKGWHJQNLVKC-UHFFFAOYSA-N	2.69	4.63
QMJRCZVQSJOCCY-UHFFFAOYSA-N	4.77	5.28
WBYDPPMHWFGHIZ-UHFFFAOYSA-N	2.98	6.47
FURUFXYYTPGZFM-MHZLTWQESA-N	3.76	5.16
MJTPTDCFFIFMAM-UHFFFAOYSA-N	4.48	6.50
TXLPUBSTTXHQRR-UHFFFAOYSA-N	4.45	8.08
CCJUYZBXKBQBRM-UHFFFAOYSA-N	4.22	5.17
BCFNIXUPNOZTSK-UHFFFAOYSA-N	4.42	4.61
SZRSNULTEJROPV-SOFGYWHQSA-N	4.46	5.30
IUGXEQSYTBOJHE-YADARESESA-N	5.24	6.71
DEUFEDRUPOJLGX-UHFFFAOYSA-N	4.27	4.78

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
MDRZOKNYGRIKBA-VSGBNLITSA-N	5.01	7.68
FBNICSGIIJKYMF-JHGXHUTESA-N	4.75	7.90
WWSOROONDXPRIT-UHFFFAOYSA-N	4.73	6.88
VSPUXIYWADHSOK-UHFFFAOYSA-N	3.61	5.25
JLSSWDFCYXSLQX-UHFFFAOYSA-N	1.87	3.45

Table E.1 Data for mono-surface plot.

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
HBWITNNIJDLPLS-UHFFFAOYSA-N	5.30	3.50
GCUGOKRWLBHMLQ-IBGZPJMESA-N	3.53	3.56
GMCWXHUWHFUMQH-UHFFFAOYSA-N	2.98	2.36
GZJYYTDPEQCUOZ-OAHLLOKOSA-N	4.50	5.79
GNHSFPGRILAESF-UHFFFAOYSA-N	3.33	3.36
CZHBUODUVDGWLO-SNVBAGLBSA-N	3.96	4.01
BMADQAUGPJFBPU-GFCCVEGCSA-N	3.24	2.82
CGTCXQOYTCJPIZ-UHFFFAOYSA-N	2.68	2.37
DVGGJTWGPZXYGY-UHFFFAOYSA-N	5.05	3.37
CSFNXBMHFQEMBP-FMIVXFBMSA-N	4.28	3.09
GOOSRBXROIWFPI-CHWSQXEVSA-N	3.30	2.92
HANDMCWHWMILMT-UHFFFAOYSA-N	4.14	4.63
CBWMOBFPWMOJKZ-QYVJPZLMSA-N	0.51	0.99
FOCILOSWFFTFNL-UHFFFAOYSA-N	3.65	3.87
FVLNTXOYQXQZHP-UHFFFAOYSA-N	3.92	3.62
GSWSWFNVEVSOJC-SECBINFHSA-N	1.41	0.59
BQOCMHOWHOMIMJ-UHFFFAOYSA-N	2.34	0.14
CFGMGRLIFAVVMY-UHFFFAOYSA-N	2.92	2.70
CISOKWVGMLHGBN-UHFFFAOYSA-N	3.04	1.96
FHVVYVTWZBXZBN-QWHCGFSZSA-N	2.85	2.94
DXFNVIVIEHJYQT-UHFFFAOYSA-N	2.70	2.72

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
COIUQMDQIBGHQH-UHFFFAOYSA-N	3.33	3.63
AAVGUZZUJOHLGC-UHFFFAOYSA-N	5.86	6.64
DHKMJAHFQFYQLM-WCQYABFASA-N	4.76	4.42
CJBOKCHSMBRXAK-UHFFFAOYSA-N	3.20	4.88
BHVSCKTUCMZDIV-UHFFFAOYSA-N	4.89	4.91
AEXUGBKKZSVAPK-MGPUTAFESA-N	4.73	5.00
CHNTVHWKBXYWAT-UHFFFAOYSA-N	6.15	4.62
ALPGTOOTEBXLDZ-UHFFFAOYSA-N	5.01	5.32
AYGQLBHSQFMOEI-UHFFFAOYSA-N	5.86	6.35
HCRFJPAQRKAART-JSGCOSHPSA-N	5.54	4.08
DDRSZUUAUJASSI-RDJZCZTQSA-N	5.92	5.19
BGZJVJOLKCRHGM-AYBZRNKSSA-N	3.65	3.27
AHPUQHAFKQTLEB-LQUAGZSJSA-N	6.00	6.71
FAKKPEHAXNWGNO-UHFFFAOYSA-N	5.27	5.25
CWKKCWFXWWPOSH-UHFFFAOYSA-N	1.55	0.84
FGVMSGZDOANVQA-UHFFFAOYSA-N	5.35	5.88
CFWIEFRWFQHAQW-HXUWFJFHSA-N	4.18	4.96
GCFRRTILCGTGBQ-YLJYHZDGSA-N	5.50	5.02
CFQXQCSYUSXLDL-UHFFFAOYSA-N	2.09	1.05
DUOXRUNUARSQEB-UHFFFAOYSA-N	3.53	5.12
CVMWEBSWRVKUDR-UHFFFAOYSA-N	4.42	3.95
FVXIXCWTNABXHN-GJZGRUSLSA-N	2.86	4.50
GKGGWGQAWVRXJJ-UHFFFAOYSA-N	5.14	5.09
FWSVZRJCNNFMSR-UHFFFAOYSA-N	4.00	5.47
CJMDFPWXRRTBSV-UHFFFAOYSA-N	2.71	0.92
DERFFWPLVQTHEZ-YYHQMBLXSA-N	2.28	1.50
DBUWPYDSLYIBCA-JXALSKIBSA-N	6.25	7.03
ARCDRMYSNWSNEQ-CYFREDJKSA-N	6.01	5.73
CXQKIQPKNPSGDE-UHFFFAOYSA-N	5.45	3.54
BDRHHTPONBYHDD-GJZGRUSLSA-N	3.68	3.46
HDDFAGCFLQKITD-UHFFFAOYSA-N	4.54	3.99

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
DIRBFPKLPYJGBA-UHFFFAOYSA-N	3.87	2.59
AQCBGMZTLIAKJI-JOCHJYFZSA-N	4.97	3.87
BSGDPNHSNFVIQI-UHFFFAOYSA-N	3.19	0.52
CACMYLJNIDZGKK-PGMHBOJBSA-N	4.64	5.50
DSDPNAASNKSGJJ-UHFFFAOYSA-N	2.68	1.32
DIVVAQYRKSZEEW-UHFFFAOYSA-N	3.92	3.12
GLPFKDSABFADSV-BQYQJAHWSA-N	5.31	5.14
CEEKAUKKCDPUKD-YOEHRIQHSA-N	4.94	4.80
FGVRQOGQXMNHGR-MRVPVSSYSA-N	4.09	3.30
BFUGHTCDORQPRS-UHFFFAOYSA-N	3.30	3.60
GEHZJNMZLQKLDL-UHFFFAOYSA-N	2.74	2.09
BCSQUJYENGABAT-UHFFFAOYSA-N	2.93	1.09
DFWQFBNINNFHPH-UHFFFAOYSA-N	4.59	6.37
AQSUNOZBYRWGBO-HXUWFJFHSA-N	4.48	2.32
DNKNARHPWPLDBY-UHFFFAOYSA-N	3.30	3.10
CPIDUHGJETUEKJ-FCHUYYIVSA-N	5.56	4.17
AJNYEMLGBXCAJD-UHFFFAOYSA-N	3.04	2.00
CLBOJFVOEQYSIF-HWKANZROSA-N	1.61	-0.29
AZQFHIOYKBSTLM-UHFFFAOYSA-N	4.64	3.97
FOZZNDRTRNDJBC-UHFFFAOYSA-N	4.16	4.52
HCJFDHHBUCXINJ-CYBMUJFWSA-N	3.43	3.69
FESKORARFOGMKM-UHFFFAOYSA-N	5.60	5.82
DRLNFFKVXNSCFI-FYWRMAATSA-N	4.92	5.55
BSLPDPNWYWWKDM-RYUDHWBXSA-N	2.69	4.04
BRPXBFGWZBBNCG-ACRUOGEOSA-N	4.00	2.15
ANSZZPJMTBLPCY-UHFFFAOYSA-N	0.86	0.90
FDEMLXAFJTWNQD-UHFFFAOYSA-N	4.13	4.80
HBTOYNLTBSWHKA-UHFFFAOYSA-N	4.73	6.12
CZZXDIHZMJQBGQ-AIYVTKCESA-N	3.39	3.67
FYSQOPNXZUHHBZ-UHFFFAOYSA-N	5.03	4.27
AHEBMWPPIOTQOC-LJQANCHMSA-N	5.56	5.21

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
DMKCZUXCQWLBOI-UHFFFAOYSA-N	3.27	3.21
GKPVOQLJHWLOIE-UHFFFAOYSA-N	5.85	6.88
DUDAZNFMQDQCTJ-FQEVSTJZSA-N	5.41	4.87
AAQGPMOPIQKGEL-UHFFFAOYSA-N	3.74	4.20
CPXSZGKFGNBILI-UHFFFAOYSA-N	4.99	4.44
DCYUXNHCVBQGHH-ZDUSSCGKSA-N	4.27	3.37
BZUWNEFCJVOJHV-VIFPVBQESA-N	3.74	2.14
ADCZQRGTZNBXSY-UHFFFAOYSA-N	3.27	3.52
GQPQLHBCZMCHEZ-QMMMGPOBSA-N	6.63	6.69
DVIOKQKRRBJPEK-UHFFFAOYSA-N	1.04	3.00
FEALGFZIOWUCBX-GOSISDBHSA-N	3.62	3.96
GWHMRPRMYRJABK-UHFFFAOYSA-N	4.43	3.02
GFTIFHCOENKKEV-FNCQTZNRSA-N	4.03	3.35
FZAWHHFNCLCWFV-FQEVSTJZSA-N	4.93	4.57
FSKUVTXSZTZMPW-JFMUQQRKSA-N	2.95	4.23
CXEQEWXYGGKYDI-SFHVURJKSA-N	4.09	3.08
CIRWZGGESKUJHD-UHFFFAOYSA-N	4.73	3.95
HBUMYZBRQSSBIF-OAHLLOKOSA-N	4.49	3.43
HCZZWVMQPZFMRX-UHFFFAOYSA-N	4.47	2.66
CXNPTWLUNSEKLS-LLVKDONJSA-N	3.48	4.92
GZLCETKUTVIGQQ-UHFFFAOYSA-N	3.72	2.33
BLCWNIGSFBOPQU-UHFFFAOYSA-N	2.56	0.92
DAOOKTDKSRSHHO-UPCLLVRISA-N	4.62	3.71
CUTVCQQLTOCFRM-UHFFFAOYSA-N	5.96	5.39
BWONCUGSZKCCIK-DOTOQJQBSA-N	4.80	4.39
GRYLSQBVFCYQBZ-UHFFFAOYSA-N	4.75	4.32
CHBFEQWFFHDSSP-UHFFFAOYSA-N	5.89	6.36
GYALPDCCJCQAEN-HSZRJFAPSA-N	5.60	7.84
CWECLZCKMKKXJQ-QNGOZBTKSA-N	4.41	5.23
DJPBASBZLCQIJC-VQTJNVASSA-N	4.34	2.74
AKISEPIDMROWEK-UHFFFAOYSA-N	3.68	2.57

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
GELXZTXDAVDBAG-UHFFFAOYSA-N	3.43	1.66
FPFZEYWOKLQUAU-OAQYLSRUSA-N	4.54	5.31
BTEKQMROJFGITO-UHFFFAOYSA-N	4.47	5.56
CFTRIZQAAJPXQE-UHFFFAOYSA-N	4.63	2.88
DPNXCAIGIMLFBE-ZDUSSCGKSA-N	3.90	2.35
BUDUREKJMNEQPJ-UHFFFAOYSA-N	5.04	4.76
CPHOVWSNRJFISA-KSSFIOAISA-N	3.65	3.30
FOZKGUULHURPKE-UHFFFAOYSA-N	4.33	2.29
ABBBQZTWZGGQMV-GFCCVEGCSA-N	3.35	2.00
AFHZHIYYMVWJKX-DXRVJIQQSA-N	4.27	5.46
BJSAANIPBRXNLX-UHFFFAOYSA-N	2.26	2.49
HEAYNZXNTGFSBT-UHFFFAOYSA-N	3.63	4.91
BKNOZYOZRQYDJS-UHFFFAOYSA-N	3.78	4.30
CUACHGSCSLHFGZ-UHFFFAOYSA-N	4.83	6.46
GZQNBRHKYZXTRC-UHFFFAOYSA-N	4.21	5.54
FNZSEIHPROENQZ-GQCTYLIASA-N	4.34	4.14
APBXPFPORVVLAM-UHFFFAOYSA-N	4.77	5.48
DBOVDDBYSUDFAU-UHFFFAOYSA-N	4.01	3.84
BOMSQFAAAZIQHP-LLVKDONJSA-N	4.77	4.42
FKVFUEKNZQBMRE-UHFFFAOYSA-N	4.64	4.58
CXGQEIVOZAKTFV-UHFFFAOYSA-N	2.00	2.17
DQPOYJGJZXULIA-NSHDSACASA-N	2.78	2.65
CONZAJPFHQVMJF-UHFFFAOYSA-N	3.39	2.10
DFOLGYLYIGLVRR-UHFFFAOYSA-N	4.26	3.74
HBCXUIGLMSNEFH-UHFFFAOYSA-N	3.82	3.72
DPMLUPHERVULNR-UHFFFAOYSA-N	3.79	4.23
CRPDGTKHEIKNSK-UHFFFAOYSA-N	4.17	3.82
GIXSHWSOHUJCQH-UHFFFAOYSA-N	4.69	5.10
HDUKGWHJQNLVKC-UHFFFAOYSA-N	2.69	3.02
CCJUYZBXKBQBRM-UHFFFAOYSA-N	4.22	4.59
BCFNIXUPNOZTSK-UHFFFAOYSA-N	4.42	2.50

Continued on next page

InChIKey	Experimental $logP_{1-octanol/water}$	Calculated $logP_{1-octanol/water}$
DEUFEDRUPOJLGX-UHFFFAOYSA-N	4.27	3.26

Table E.2 Data for tri-surface plot.

E.2.4 Dataset 4- SAMPL6

Molecule ID	Experimental $logP_{1-octanol/water}$	Mono-surface $logP_{1-octanol/water}$	Tri-Surface $logP_{1-octanol/water}$
SM02	4.09	6.11	4.38
SM04	3.98	5.02	3.75
SM07	3.21	4.15	2.85
SM08	3.10	3.33	2.89
SM09	3.03	4.97	3.54
SM11	2.10	3.21	0.89
SM12	3.83	5.04	3.65
SM13	2.92	5.49	4.19
SM14	1.95	2.50	0.71
SM15	3.07	3.33	1.89
SM16	2.62	4.21	2.41

Appendix F

SSIP description validation

F.1 Displaying SSIP Summary information

The SSIP XML provides a machine readable format, with limited direct human readability. To be able to assess the suitability of the descriptions to be used to generate the FGIPs and also to be used in development of the solvent similarity metric, the descriptions were visually inspected.

F.2 Solvent SSIP summary plots

The solvent SSIP information was plotted on a single scale for each molecule. The 3D information for the descriptions is unimportant for solvation free energies, thus a one dimensional representation can be used, so multiple molecules can be compared. If an experimental value for the largest α and/or β was collected in chapter 2, then this was plotted as a red cross. Black circles represent descriptions previously generated through a combination of the previous software and manual adjustment for use in [158]. Mean functional group values were also included as horizontal black lines and are described in the next section.

F.2.1 Functional Group mean values

The mean value of the hydrogen bond donor and acceptor values were also included on the plot, as horizontal black lines. To add the values to the plots, subgraph pattern matching using RDKit [261] with the SMILES arbitrary target specification (SMARTS) [331] as input for structures.

The mean values (from [149]), and the SMARTS strings are in table F.1.

Functional Group	SSIP Value	SMARTS
Phenol	-3.0	[c][OX2][H]
Nitro alkane	-3.8	N(=O)(=O)C
Primary aniline	-4.2	[c]([NX3]([H])[H])
Secondary aniline	-4.5	[c]([NX3]([H])[C,c])
Carboxylic acid	-4.9	[CX3](=O)O[H]
Nitrile	-5.0	N#C
Alcohol	-5.0	[CX4](O[H])
Ester	-5.4	[CX3](=O)O[C,c]
Alkyl ether	-5.4	[CX4]O[CX4]
Ketone	-5.8	[#6][CX3](=O)[#6]
Primary amine	-7.9	[NX3]([H])([H])[CX4]
Secondary amine	-8.2	[NX3]([H])([CX4])[CX4]
Secondary amide	-8.3	[CX3](=O)N([H])[#6]
Pyridine	-7.4	[nX2]1ccccc1
Imidazole	-8.6	C1=C[NX2]=C[N]1([H])
Sulfoxide	-8.3	[#6][S](=[OX1])(=[OX1])[#6]
Primary amine	1.4	[NX3]([H])([H])[CX4]
Secondary amine	1.4	[NX3]([H])([CX4])[CX4]
Primary aniline	2.4	[c]([NX3]([H])[H])
Secondary aniline	2.1	[c]([NX3]([H])[C,c])
Alcohol	2.7	[CX4](O[H])
Secondary amide	3.0	[CX3](=O)N([H])[#6]
Pyrrole	3.0	C1=CC=C[N]1([H])
Indole	3.2	[nX3]([H])1c2ccccc(cc1)2
Carboxylic acid	3.6	[CX3](=O)O[H]
Imidazole	3.2	C1=C[NX2]=C[N]1([H])
Phenol	3.6	[c][OX2][H]

Table F.1 Table containing mean value by functional group and the SMARTS pattern used to detect it in a molecule.

Molecule name	SSIP α value modifications	SSIP β value modifications
water	Two values to 2.8 from 2.9, 2.9	Two values to -4.5 from -6.2, -6.3
ammonia		One value to -6.8 from -9.5
1,2-dimethoxybenzene		Two values to -3.6 from -7.0, -7.9
glycerol		Six values to -5.3 from 0.0, 0.0, -4.9, -5.1, -8.4, -8.4
bis(chloroethyl)ether		One value to -5.3 from -1.7

Table F.2 Summary of Solvent description modifications

F.2.2 Solvent description modifications.

The following solvent SSIP descriptions were modified for calculation of the FGIPs and also solvent similarity. This was done because the largest α or β SSIPs showed significant deviation from the experimental value for either the molecule, or the functional group mean value.

F.2.3 Solvent Summary Graphics

The plots showing the unmodified SSIP description produced from the mono-surface and tri-surface footprinting approaches are in the following sections for the different solvents.

Mono surface SSIP description plots

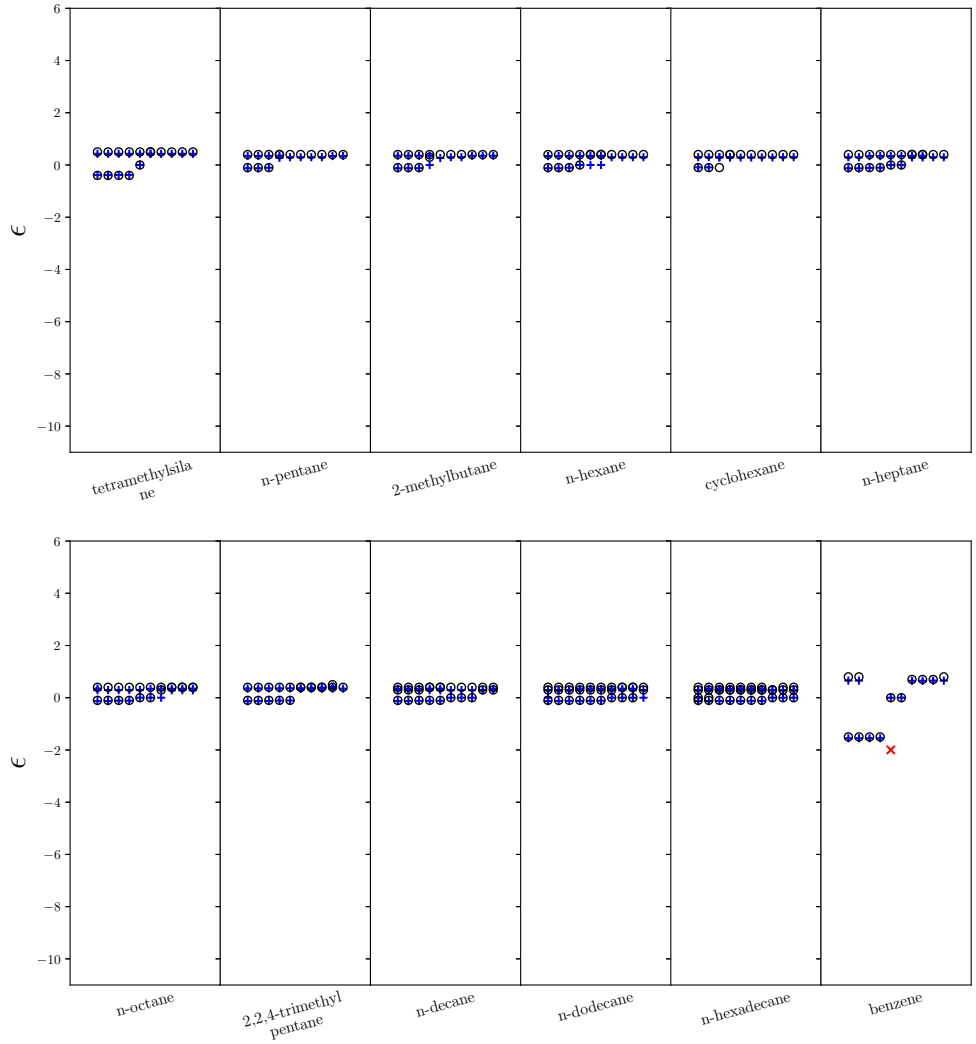


Fig. F.1: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are tetramethylsilane, n-pentane, 2-methylbutane, n-hexane, cyclohexane, n-heptane, n-octane, 2,2,4-trimethylpentane, n-decane, n-dodecane, n-hexadecane, benzene. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

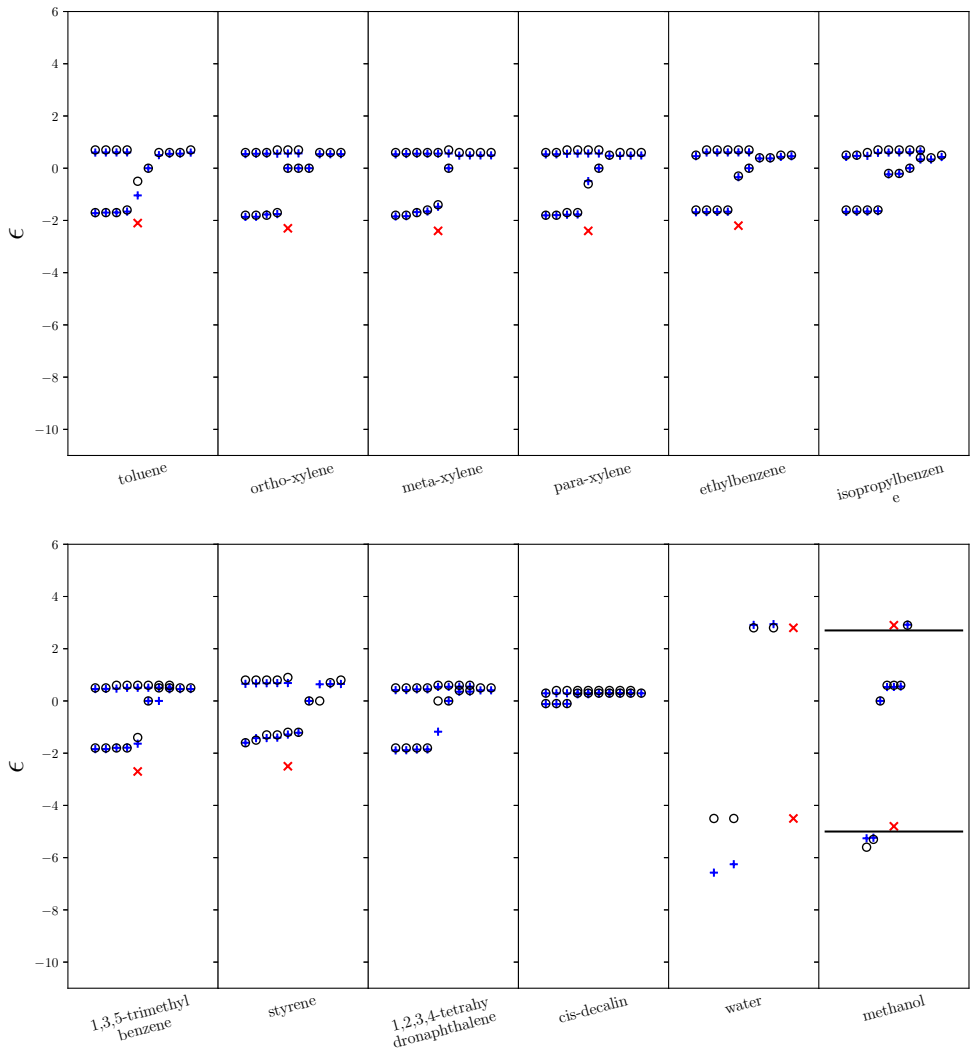


Fig. F.2: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are toluene, ortho-xylene, meta-xylene, para-xylene, ethylbenzene, isopropylbenzene, 1,3,5-trimethylbenzene, styrene, 1,2,3,4-tetrahydronaphthalene, cis-decalin, water, methanol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

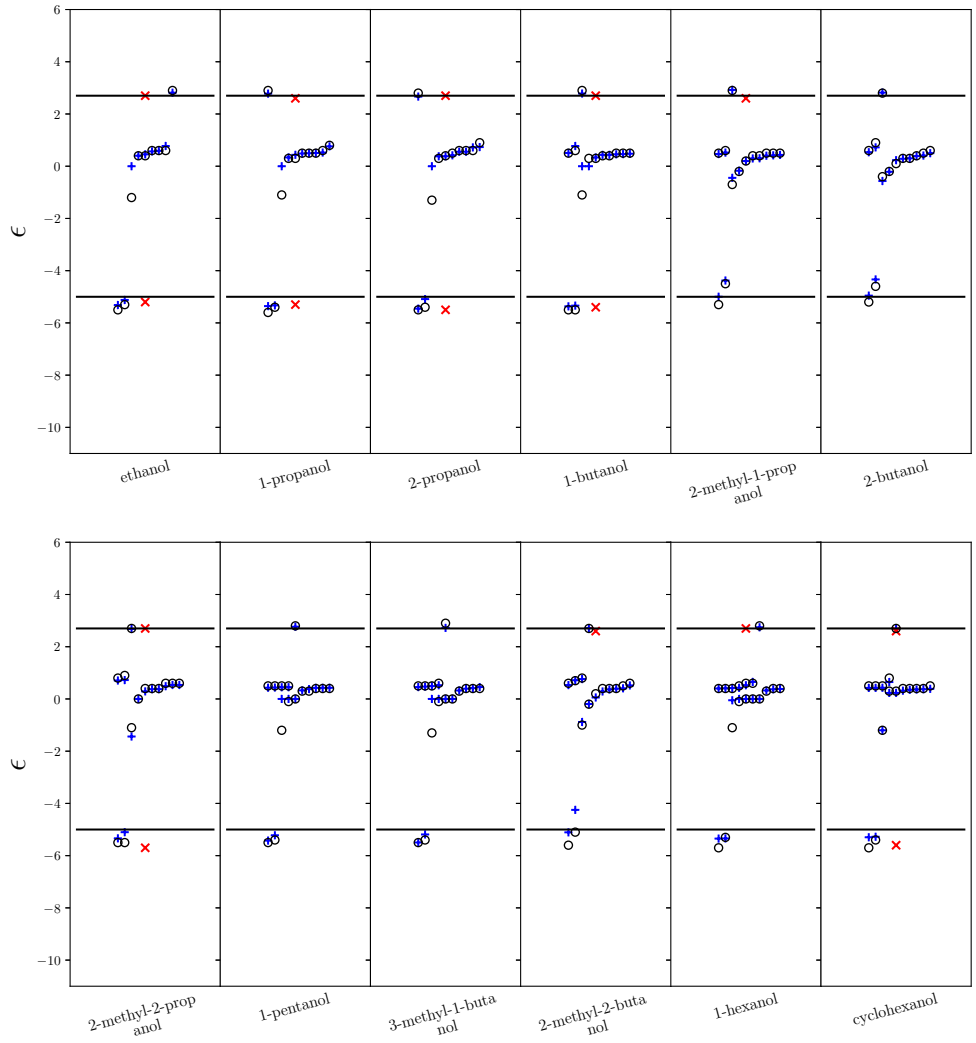


Fig. F.3: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are ethanol, 1-propanol, 2-propanol, 1-butanol, 2-methyl-1-propanol, 2-butanol, 2-methyl-2-propanol, 1-pentanol, 3-methyl-1-butanol, 2-methyl-2-butanol, 1-hexanol, cyclohexanol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

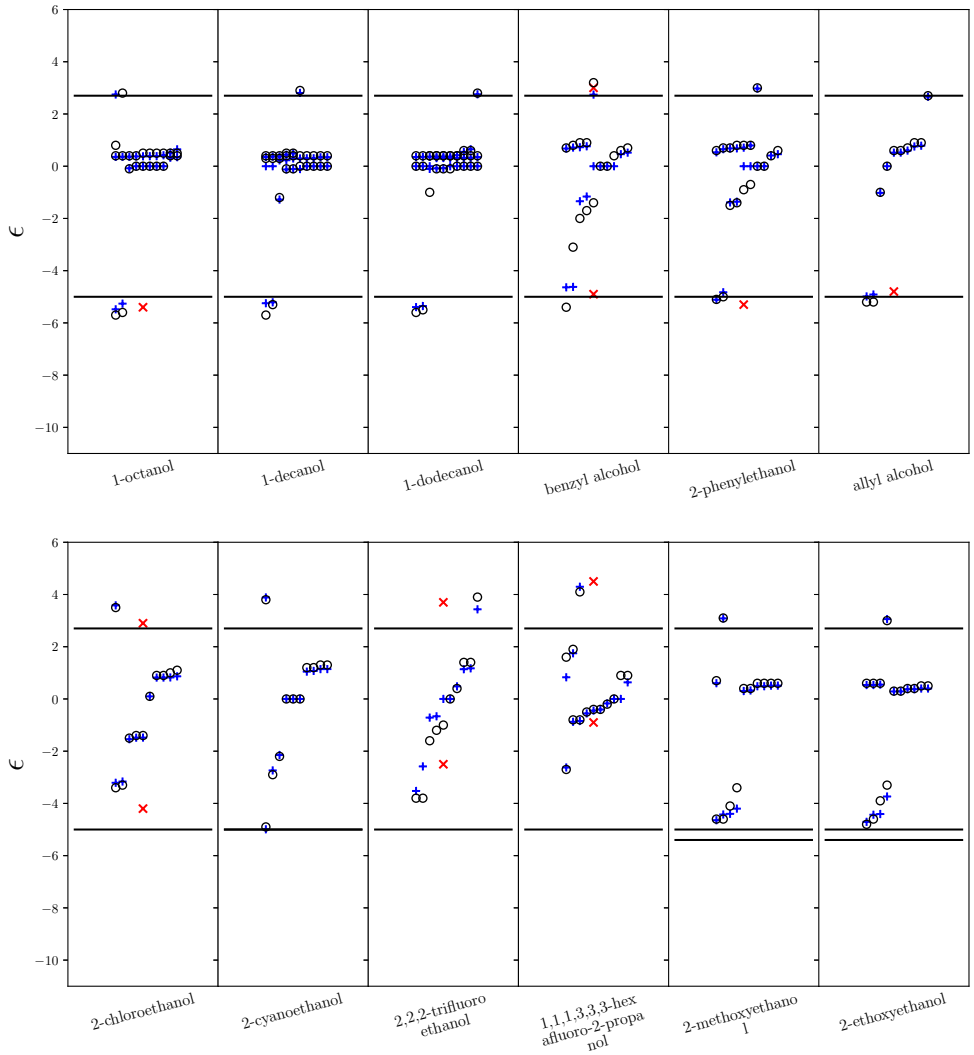


Fig. F.4: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are 1-octanol, 1-decanol, 1-dodecanol, benzyl alcohol, 2-phenylethanol, allyl alcohol, 2-chloroethanol, 2-cyanoethanol, 2,2,2-trifluoroethanol, 1,1,1,3,3,3-hexafluoro-2-propanol, 2-methoxyethanol, 2-ethoxyethanol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

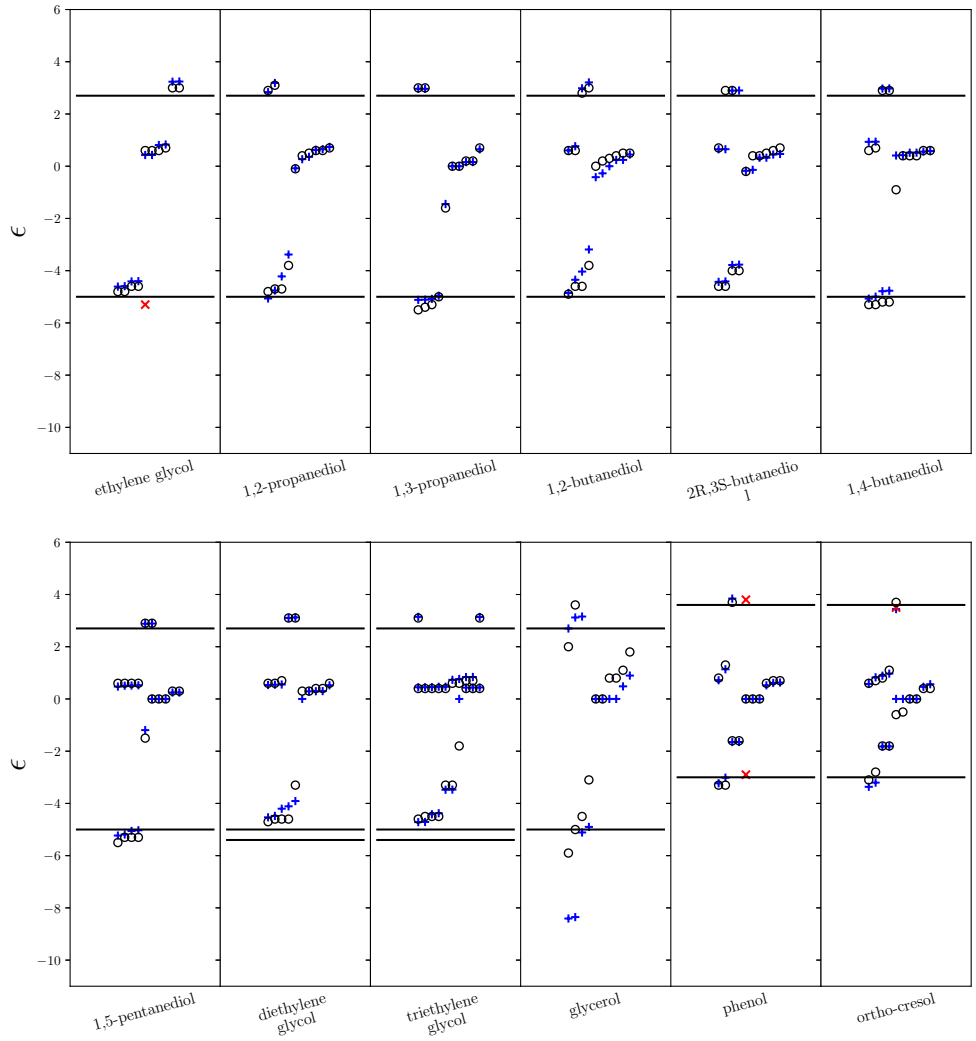


Fig. F.5: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are ethylene glycol, 1,2-propanediol, 1,3-propanediol, 1,2-butanediol, 2R,3S-butanediol, 1,4-butanediol, 1,5-pentanediol, diethylene glycol, triethylene glycol, glycerol, phenol, ortho-cresol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

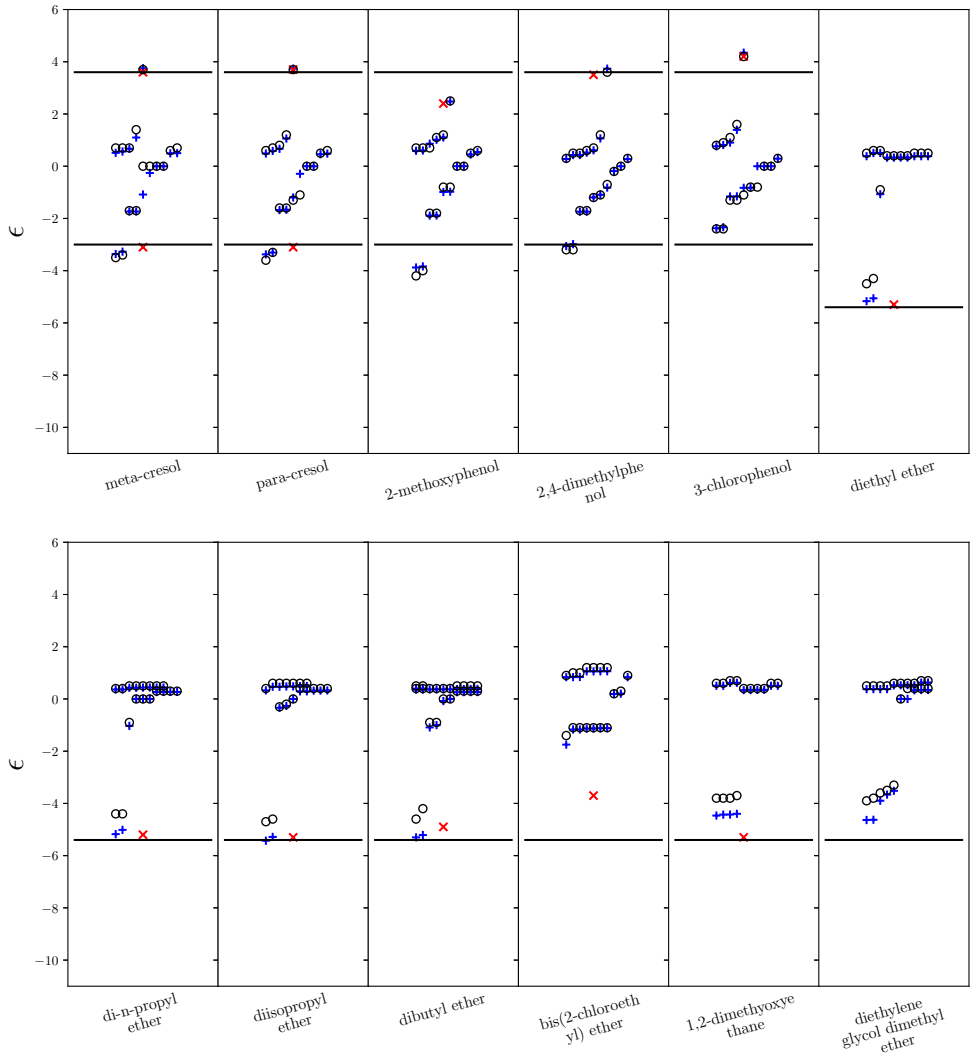


Fig. F.6: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are meta-cresol, para-cresol, 2-methoxyphenol, 2,4-dimethylphenol, 3-chlorophenol, diethyl ether, di-n-propyl ether, diisopropyl ether, dibutyl ether, bis(2-chloroethyl) ether, 1,2-dimethyoxyethane, diethylene glycol dimethyl ether. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

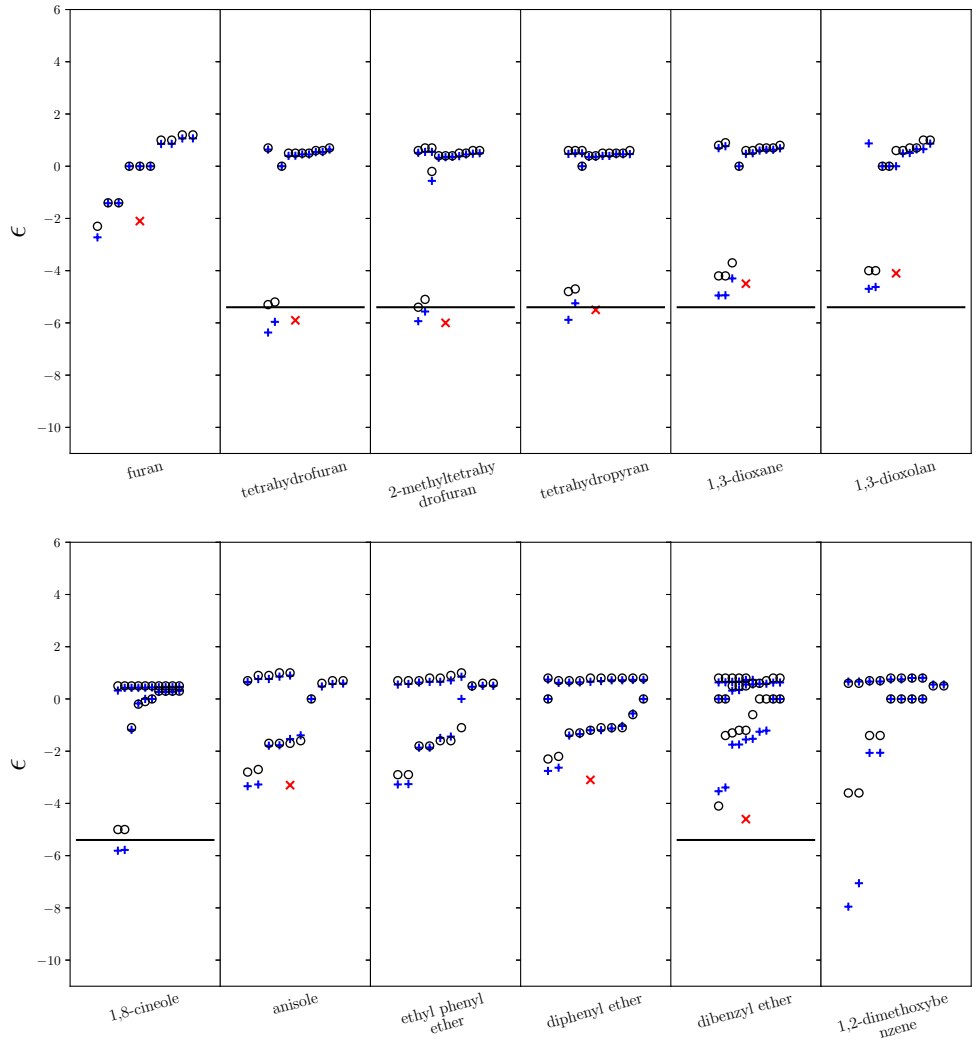


Fig. F.7: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are furan, tetrahydrofuran, 2-methyltetrahydrofuran, tetrahydropyran, 1,3-dioxane, 1,3-dioxolan, 1,8-cineole, anisole, ethyl phenyl ether, diphenyl ether, dibenzyl ether, 1,2-dimethoxybenzene. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

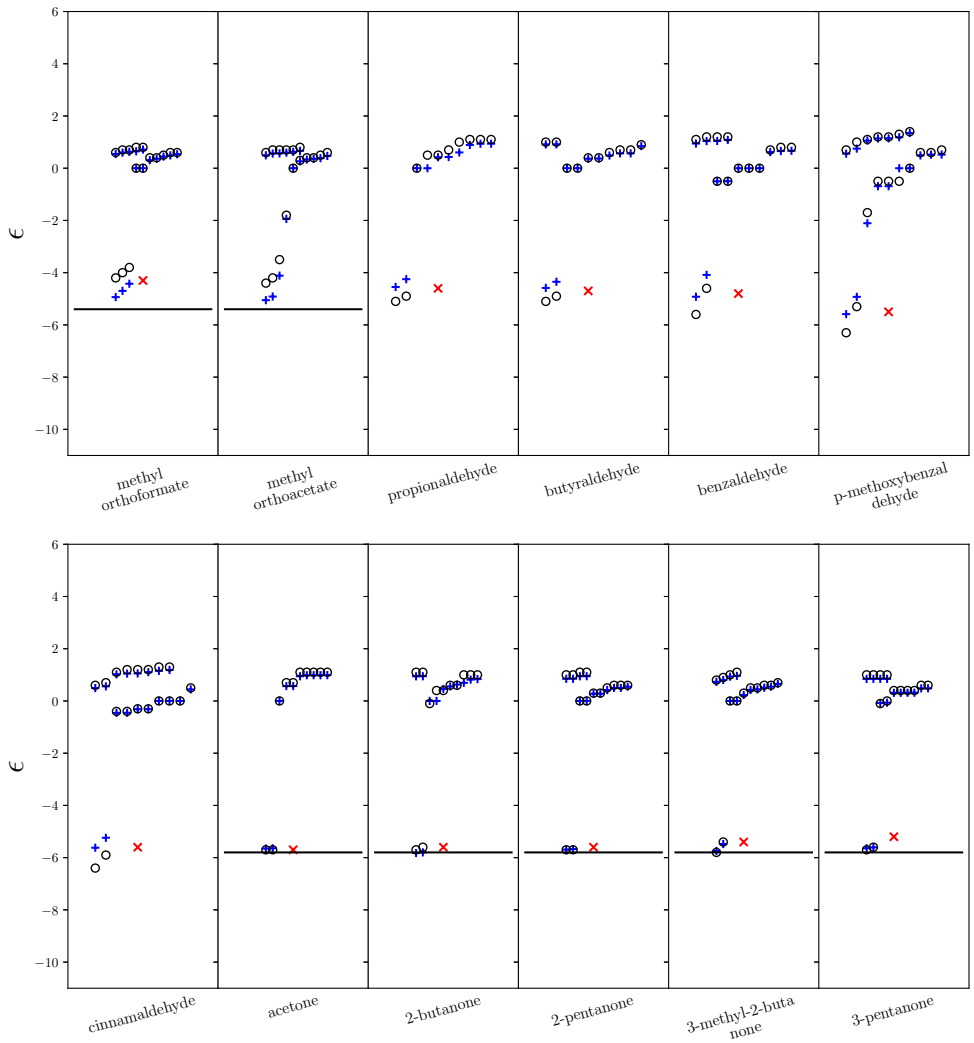


Fig. F.8: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are methyl orthoformate, methyl orthoacetate, propionaldehyde, butyraldehyde, benzaldehyde, p-methoxybenzaldehyde, cinnamaldehyde, acetone, 2-butanone, 2-pentanone, 3-methyl-2-butanone, 3-pentanone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

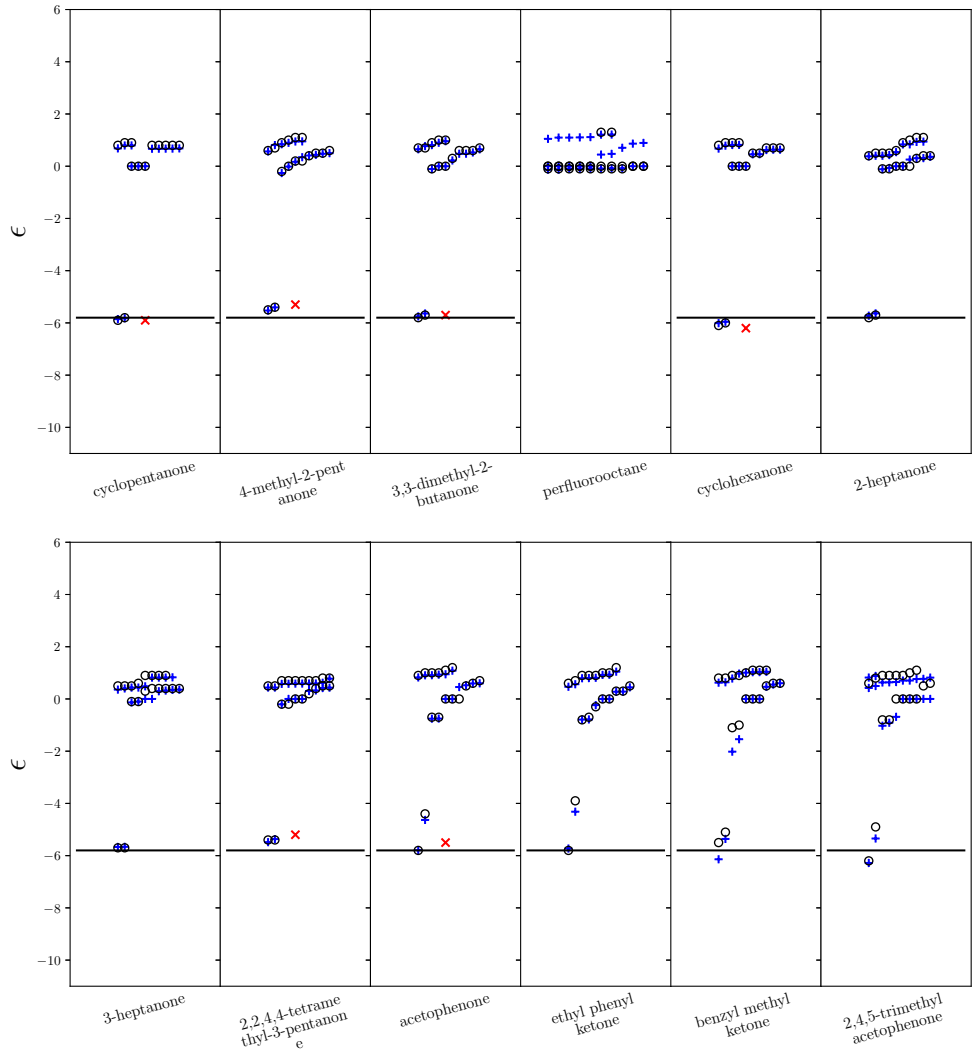


Fig. F.9: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are cyclopentanone, 4-methyl-2-pentanone, 3,3-dimethyl-2-butanone, perfluorooctane, cyclohexanone, 2-heptanone, 3-heptanone, 2,2,4,4-tetramethyl-3-pentanone, acetophenone, ethyl phenyl ketone, benzyl methyl ketone, 2,4,5-trimethylacetophenone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

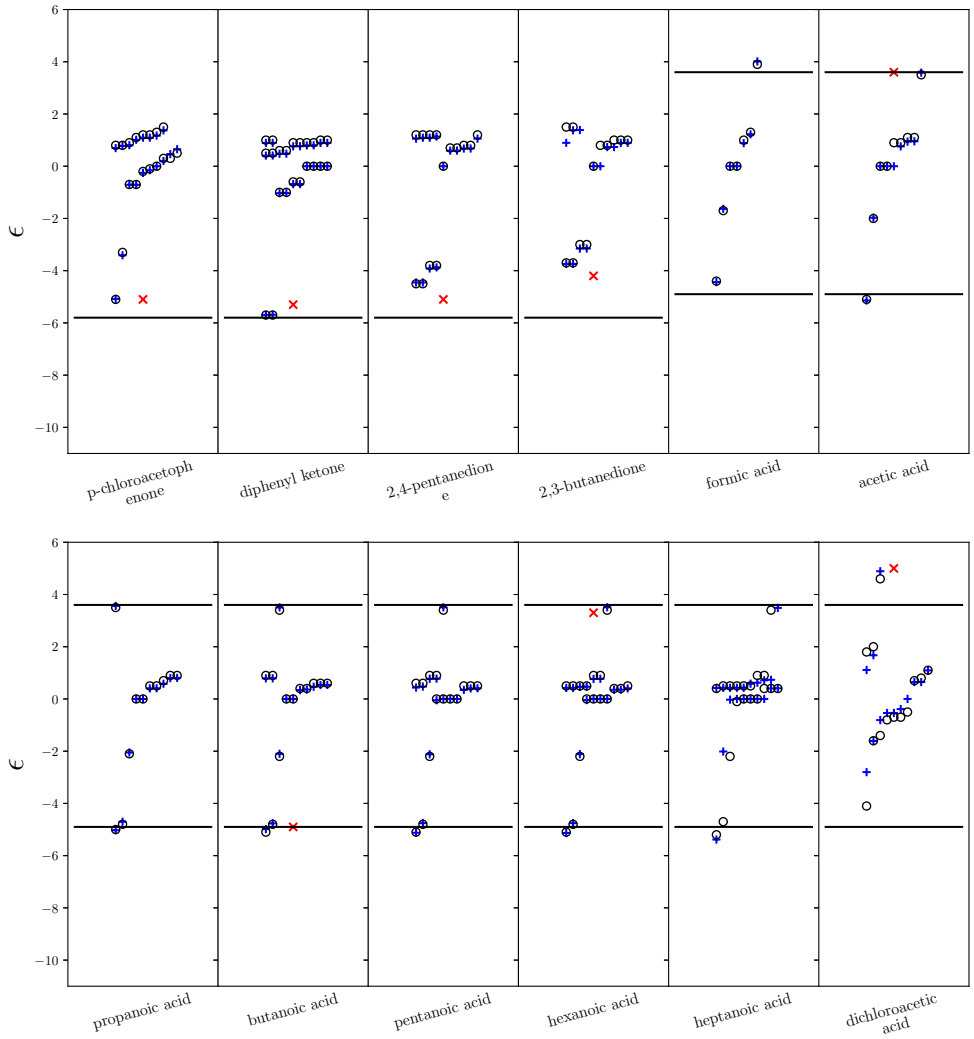


Fig. F.10: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are p-chloroacetophenone, diphenyl ketone, 2,4-pentanedione, 2,3-butanedione, formic acid, acetic acid, propanoic acid, butanoic acid, pentanoic acid, hexanoic acid, heptanoic acid, dichloroacetic acid. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

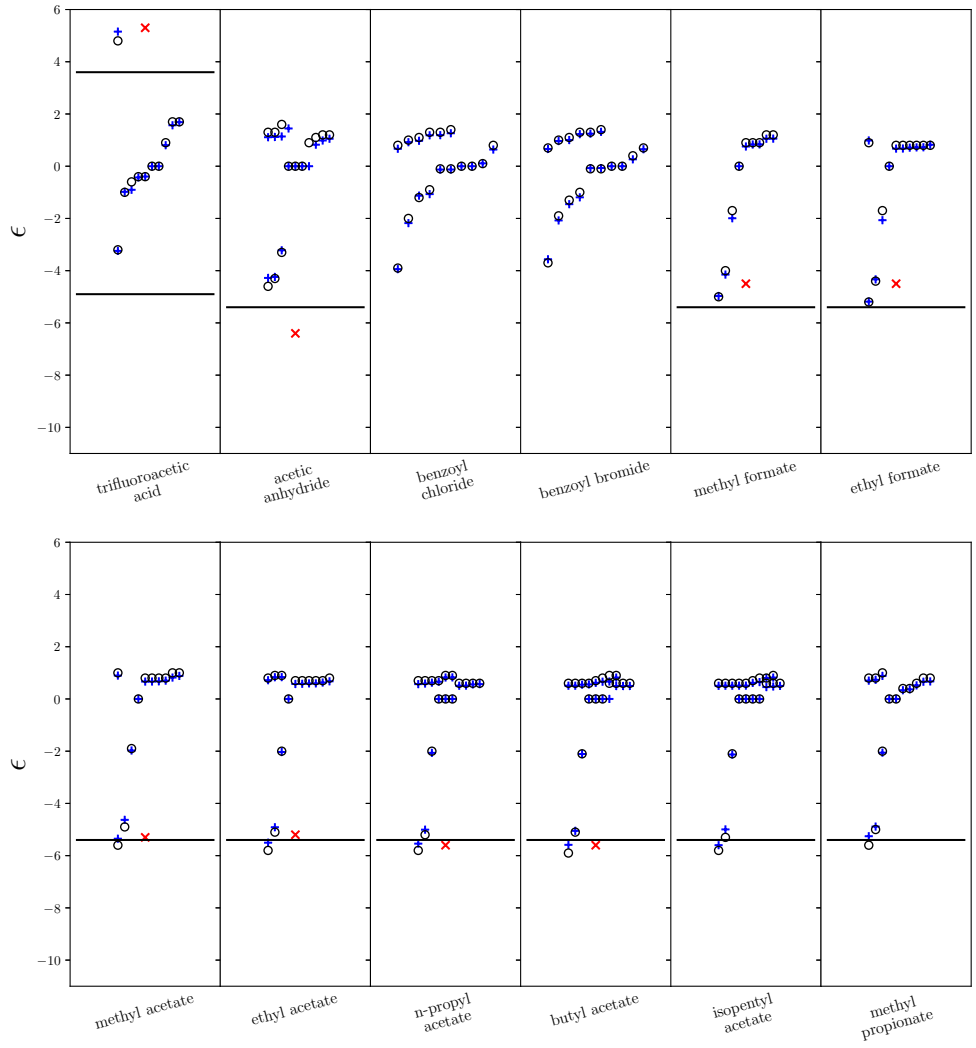


Fig. F.11: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are trifluoroacetic acid, acetic anhydride, benzoyl chloride, benzoyl bromide, methyl formate, ethyl formate, methyl acetate, ethyl acetate, n-propyl acetate, butyl acetate, isopentyl acetate, methyl propionate. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

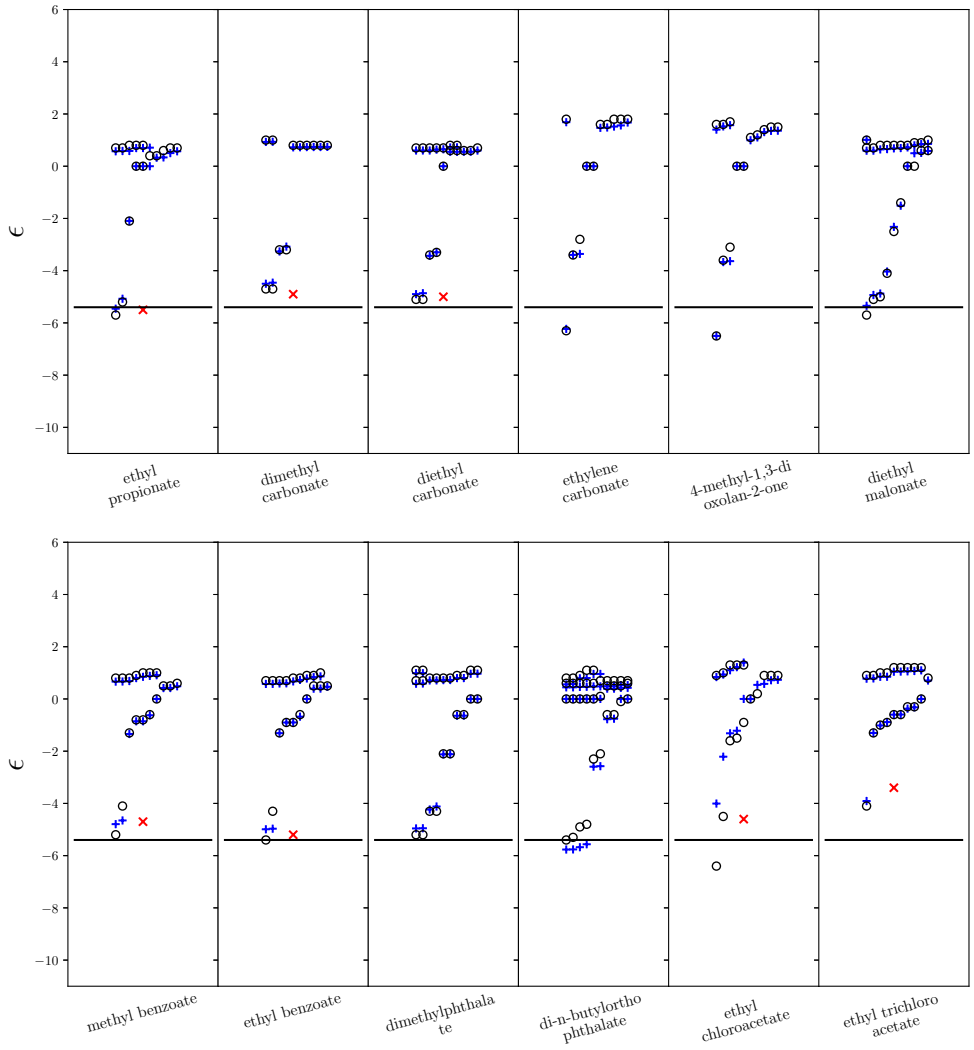


Fig. F.12: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are ethyl propionate, dimethyl carbonate, diethyl carbonate, ethylene carbonate, 4-methyl-1,3-dioxolan-2-one, diethyl malonate, methyl benzoate, ethyl benzoate, dimethylphthalate, di-n-butylorthophthalate, ethyl chloroacetate, ethyl trichloroacetate. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

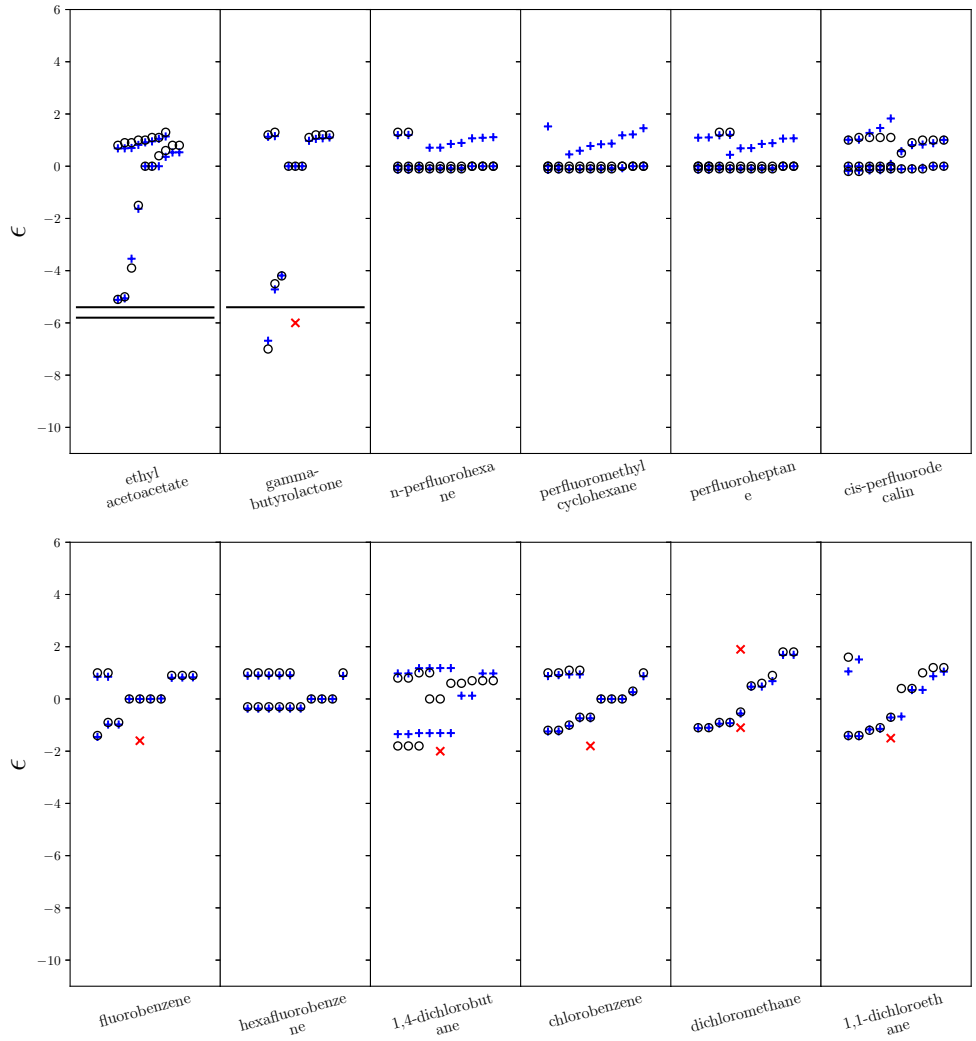


Fig. F.13: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are ethyl acetoacetate, gamma-butyrolactone, n-perfluorohexane, perfluoromethylcyclohexane, perfluoroheptane, cis-perfluorodecalin, fluorobenzene, hexafluorobenzene, 1,4-dichlorobutane, chlorobenzene, dichloromethane, 1,1-dichloroethane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

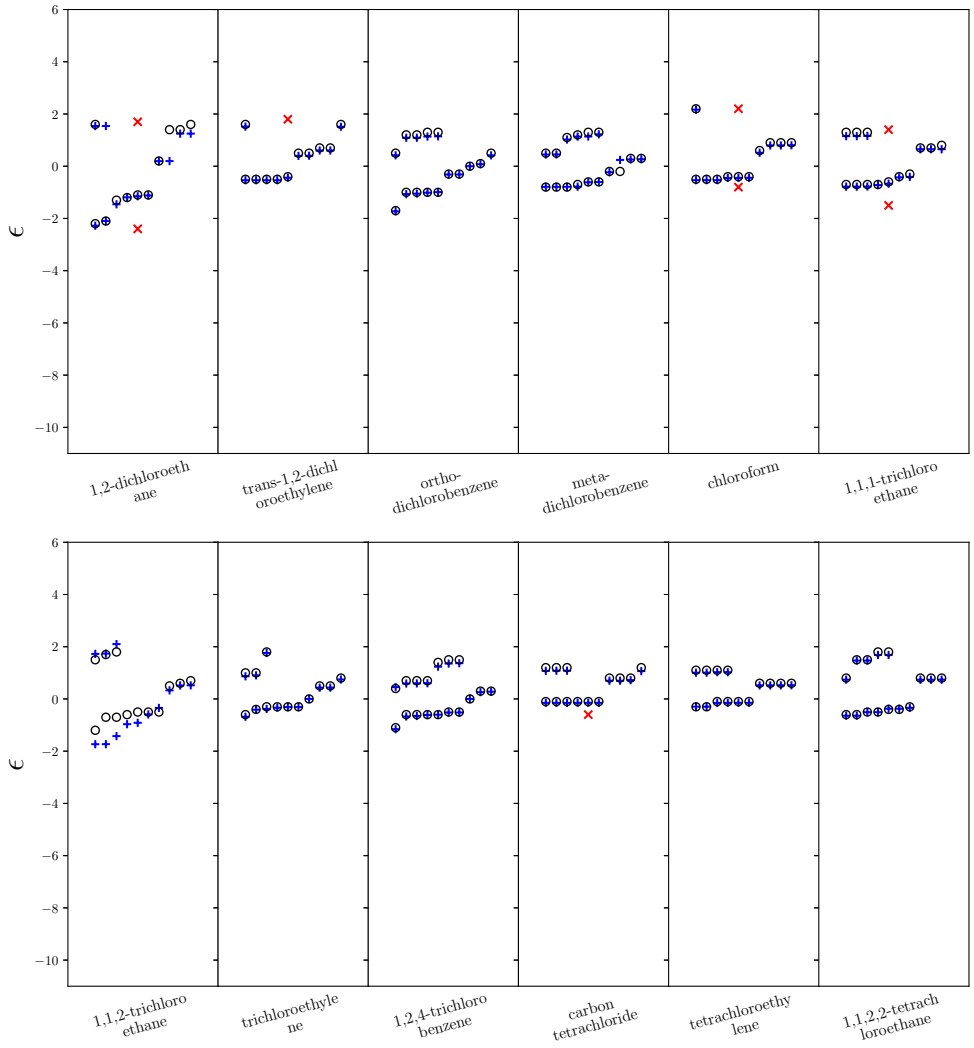


Fig. F.14: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are 1,2-dichloroethane, trans-1,2-dichloroethylene, ortho-dichlorobenzene, meta-dichlorobenzene, chloroform, 1,1,1-trichloroethane, 1,1,2-trichloroethane, trichloroethylene, 1,2,4-trichlorobenzene, carbon tetrachloride, tetrachloroethylene, 1,1,2,2-tetrachloroethane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

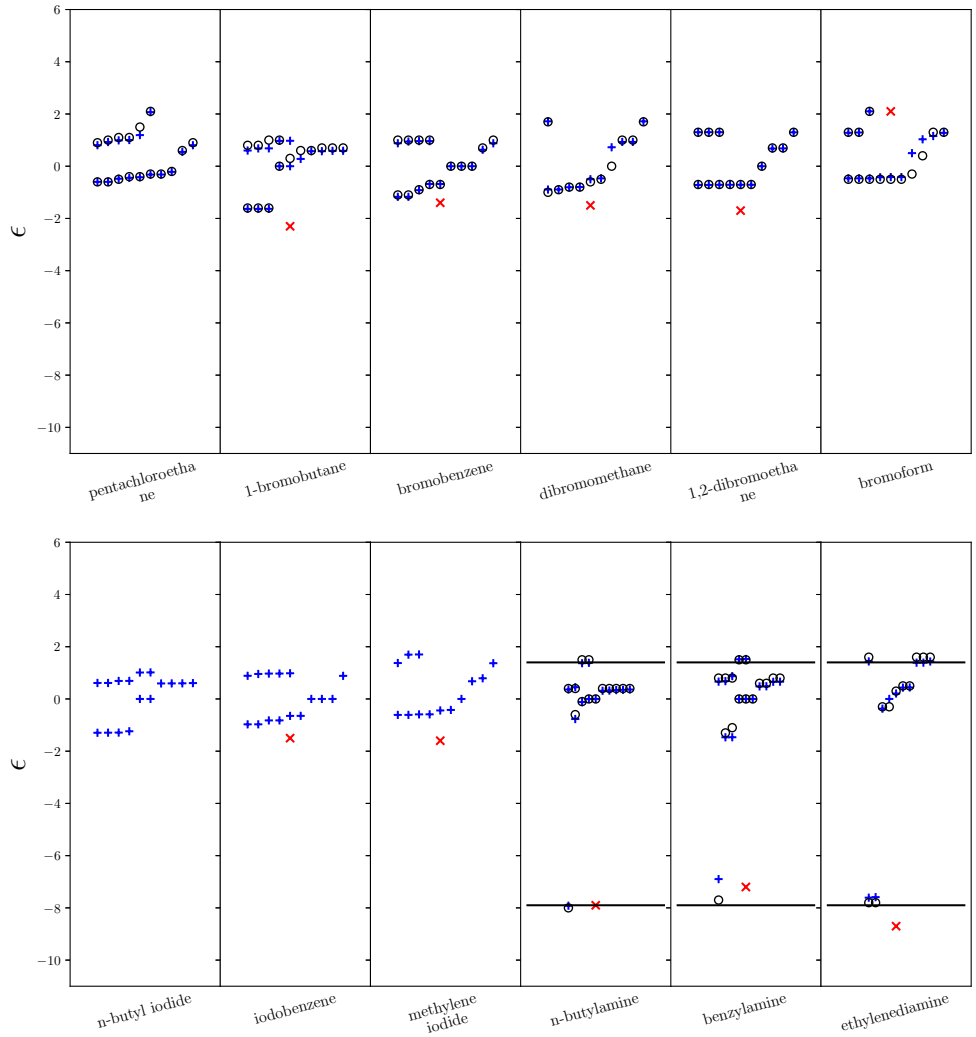


Fig. F.15: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are pentachloroethane, 1-bromobutane, bromobenzene, dibromomethane, 1,2-dibromoethane, bromoform, n-butyl iodide, iodobenzene, methylene iodide, n-butylamine, benzylamine, ethylenediamine. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

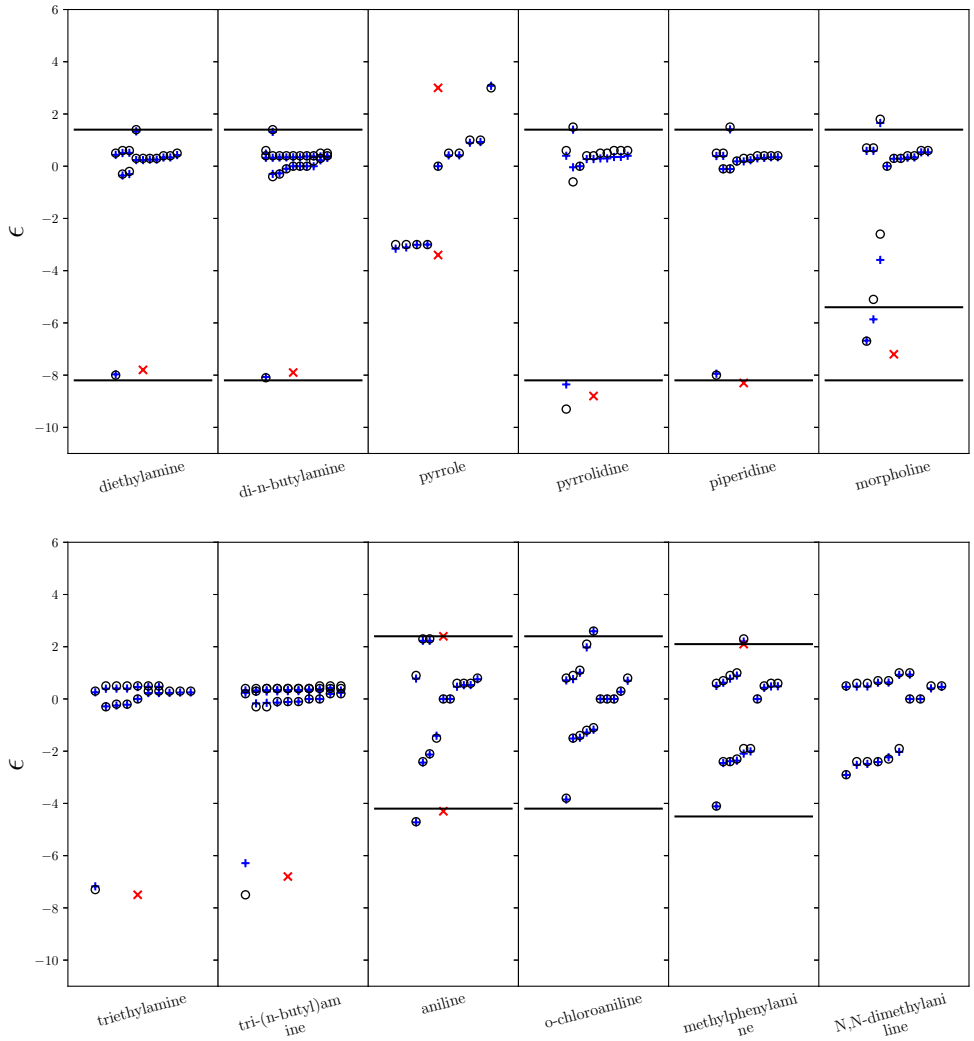


Fig. F.16: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are diethylamine, di-n-butylamine, pyrrole, pyrrolidine, piperidine, morpholine, triethylamine, tri-(n-butyl)amine, aniline, o-chloroaniline, methylphenylamine, N,N-dimethylaniline. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

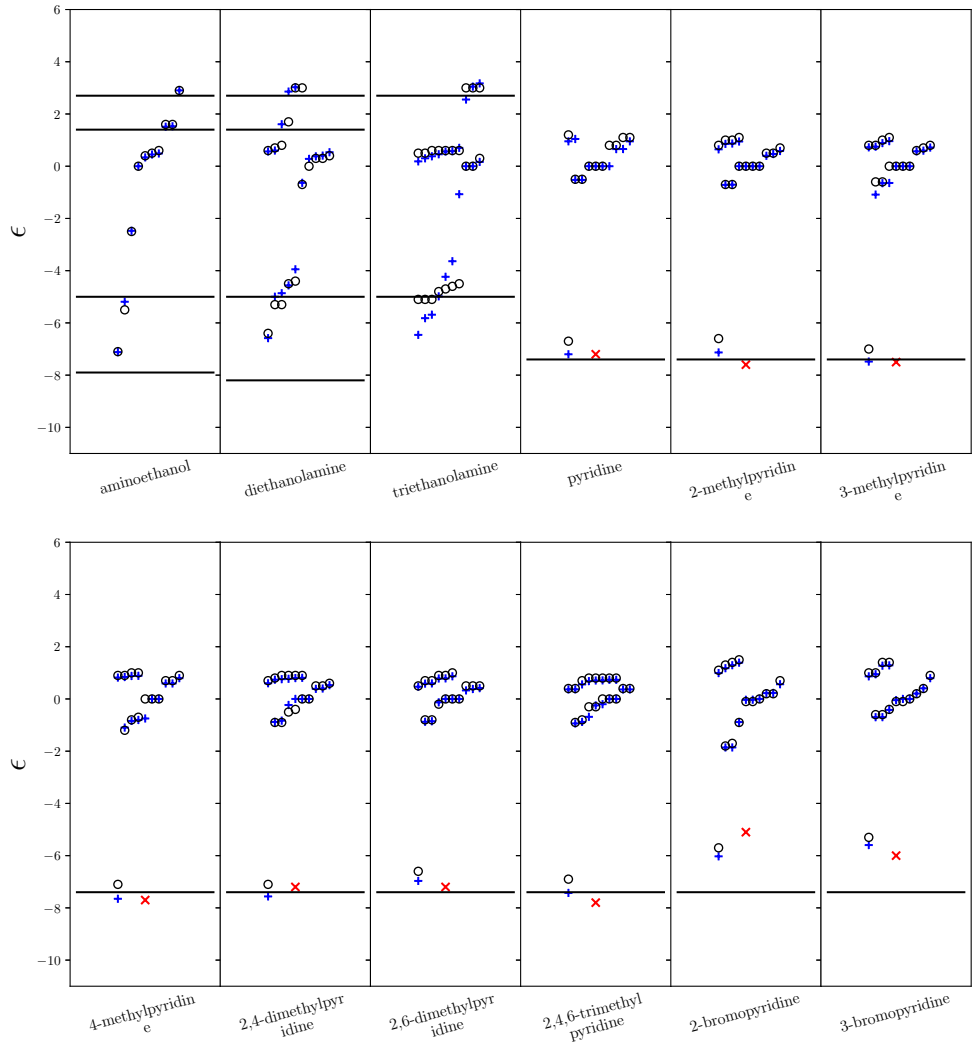


Fig. F.17: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are aminoethanol, diethanolamine, triethanolamine, pyridine, 2-methylpyridine, 3-methylpyridine, 4-methylpyridine, 2,4-dimethylpyridine, 2,6-dimethylpyridine, 2,4,6-trimethylpyridine, 2-bromopyridine, 3-bromopyridine. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

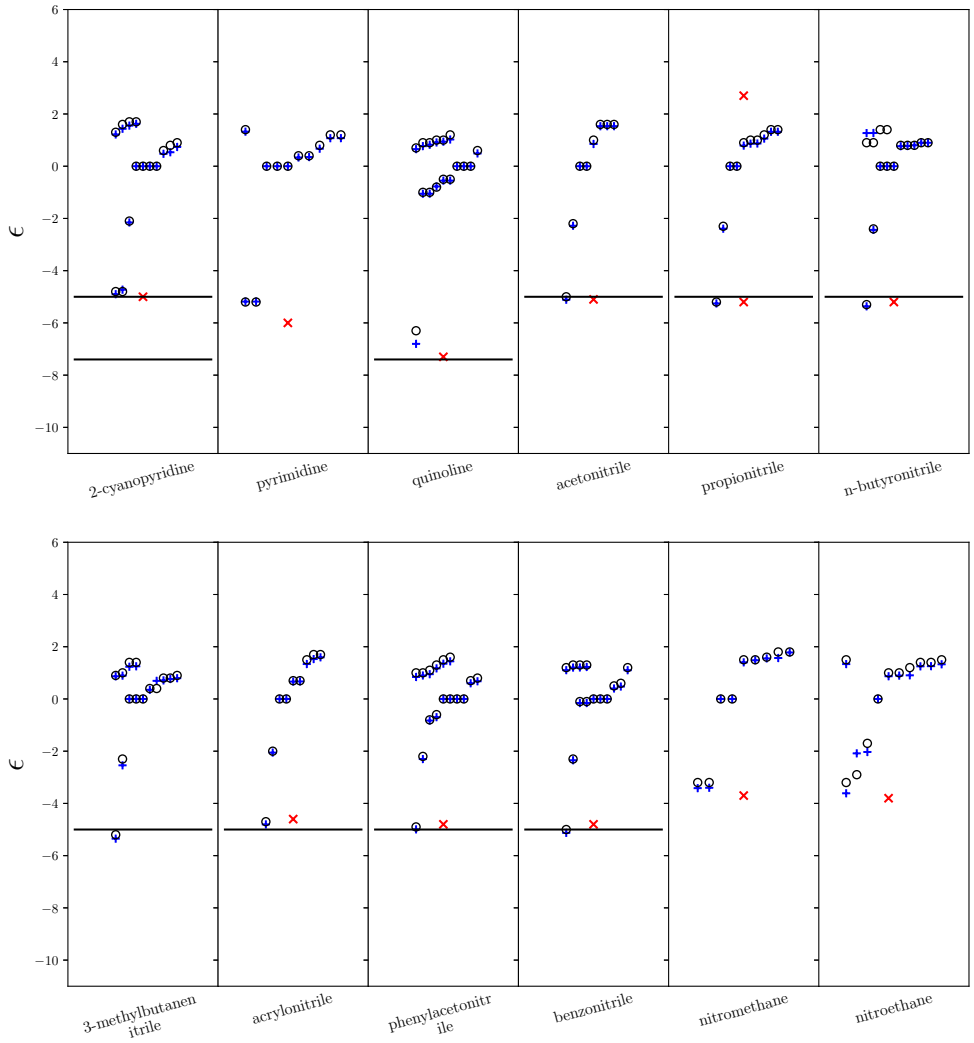


Fig. F.18: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are 2-cyanopyridine, pyrimidine, quinoline, acetonitrile, propionitrile, n-butyronitrile, 3-methylbutanenitrile, acrylonitrile, phenylacetonitrile, benzonitrile, nitromethane, nitroethane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

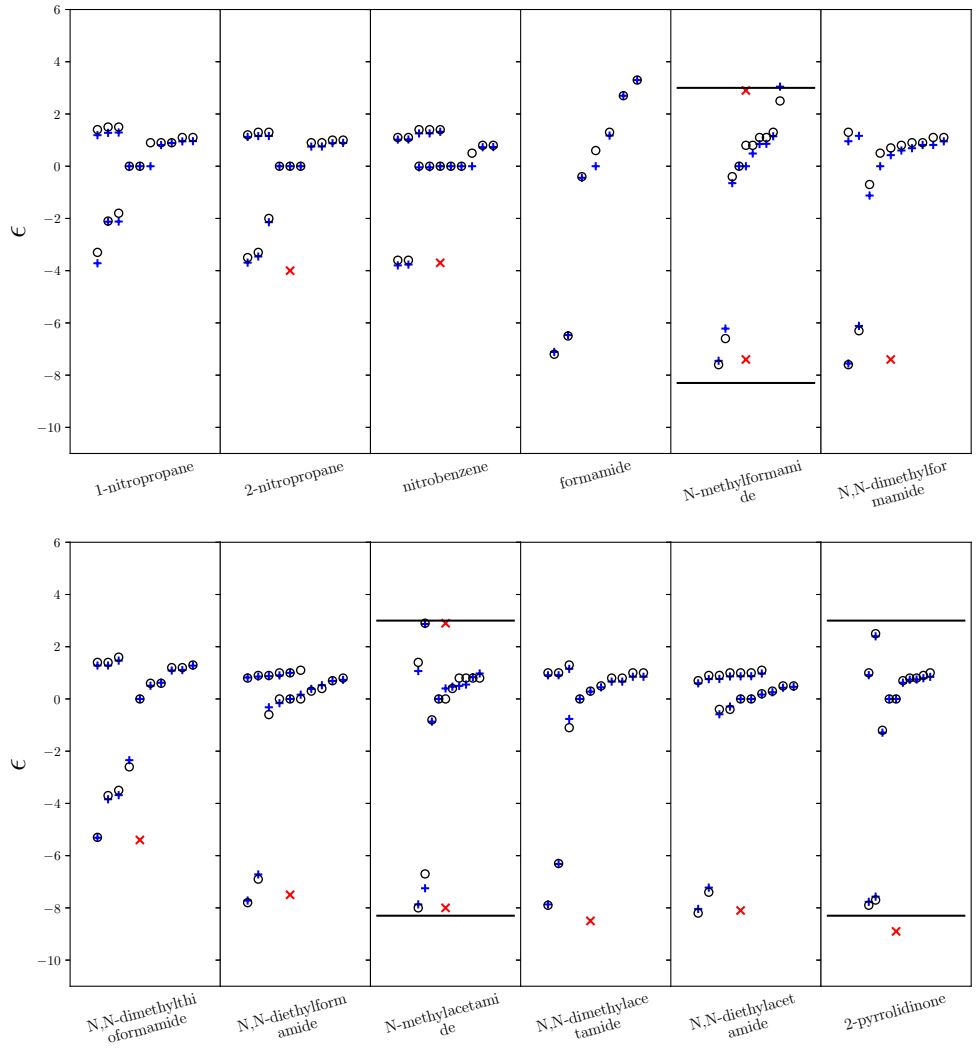


Fig. F.19: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are 1-nitropropane, 2-nitropropane, nitrobenzene, formamide, N-methylformamide, N,N-dimethylformamide, N,N-dimethylthioformamide, N,N-diethylformamide, N-methylacetamide, N,N-dimethylacetamide, N,N-diethylacetamide, 2-pyrrolidinone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

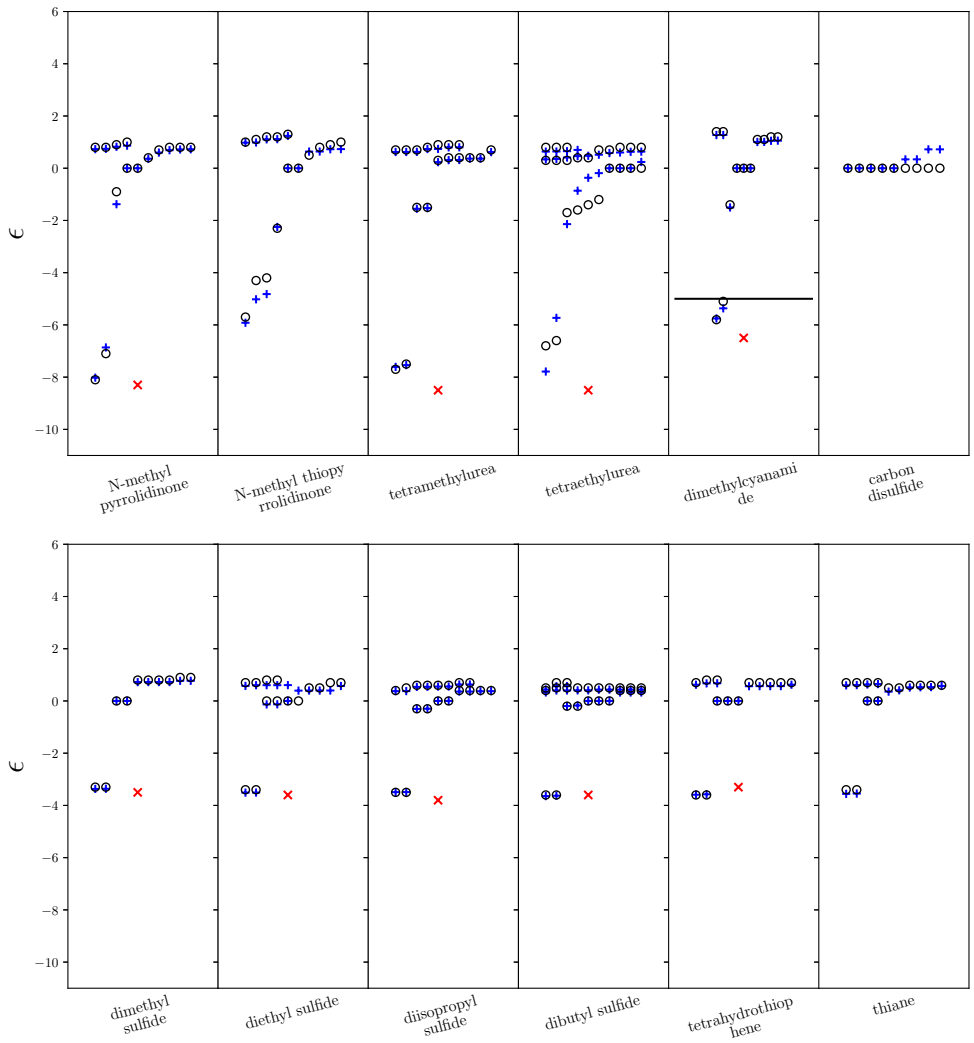


Fig. F.20: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are N-methyl pyrrolidinone, N-methyl thiopyrrolidinone, tetramethylurea, tetraethylurea, dimethylcyanamide, carbon disulfide, dimethyl sulfide, diethyl sulfide, diisopropyl sulfide, dibutyl sulfide, tetrahydrothiophene, thiane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

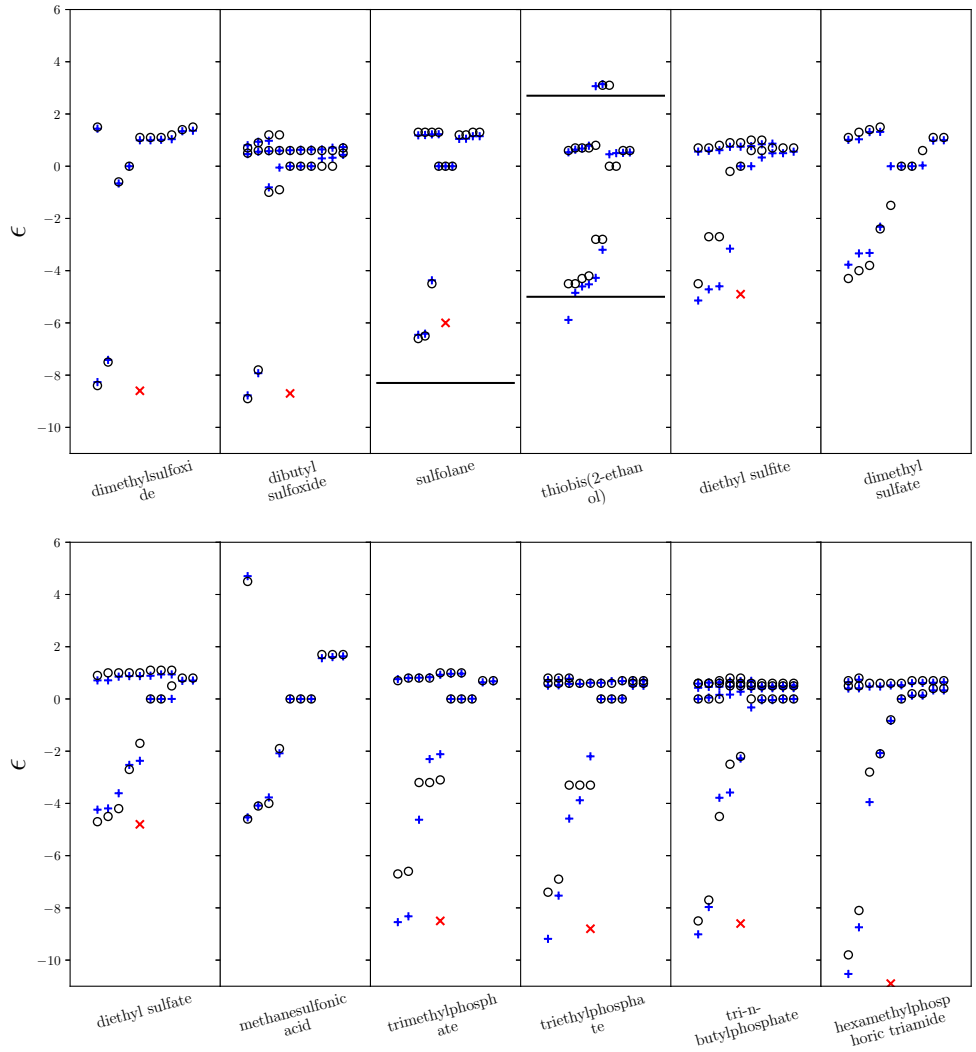


Fig. F.21: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are dimethylsulfoxide, dibutyl sulfoxide, sulfolane, thiobis(2-ethanol), diethyl sulfite, dimethyl sulfate, diethyl sulfate, methanesulfonic acid, trimethylphosphate, triethylphosphate, tri-n-butylphosphate, hexamethylphosphoric triamide. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

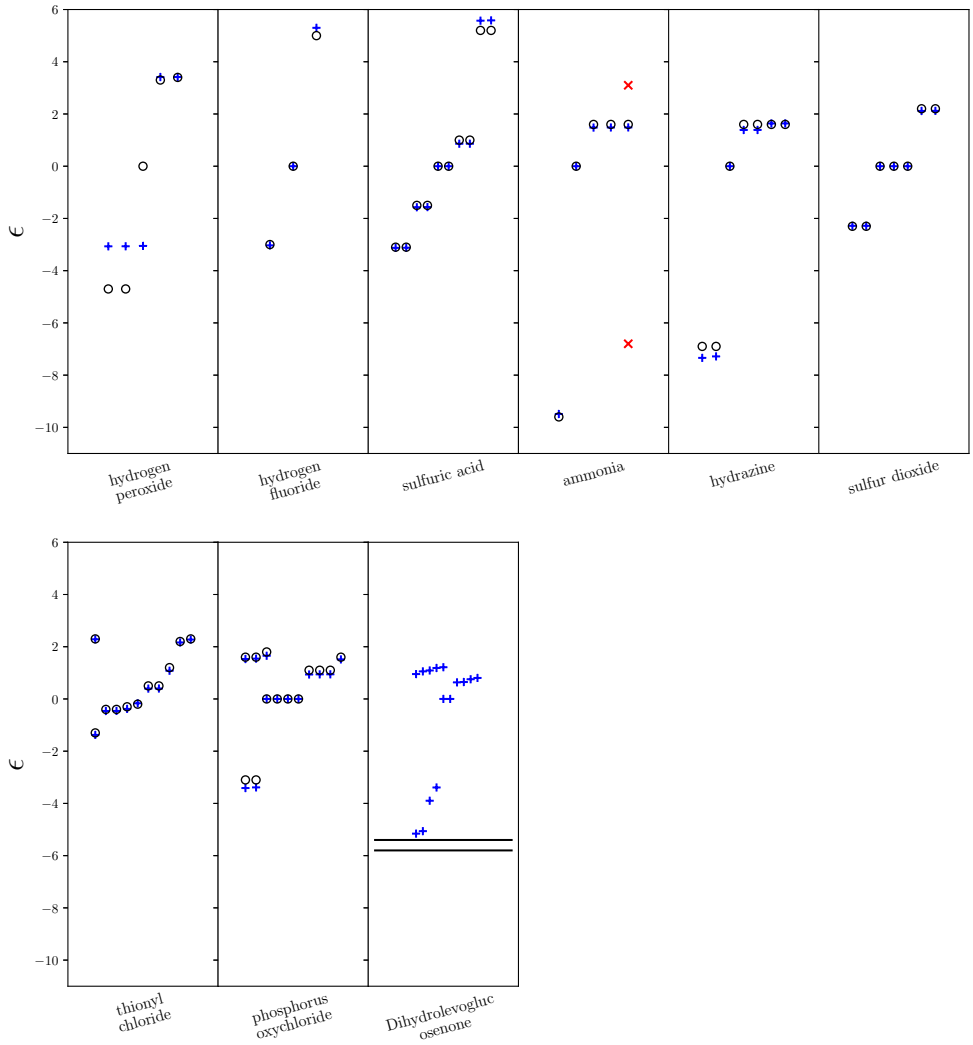


Fig. F.22: SSIP description generated by mono-surface footprinting approach. Molecules (top left to bottom right) are hydrogen peroxide, hydrogen fluoride, sulfuric acid, ammonia, hydrazine, sulfur dioxide, thionyl chloride, phosphorus oxychloride, Dihydrolevoglucosenone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

Tri-surface SSIP description plots

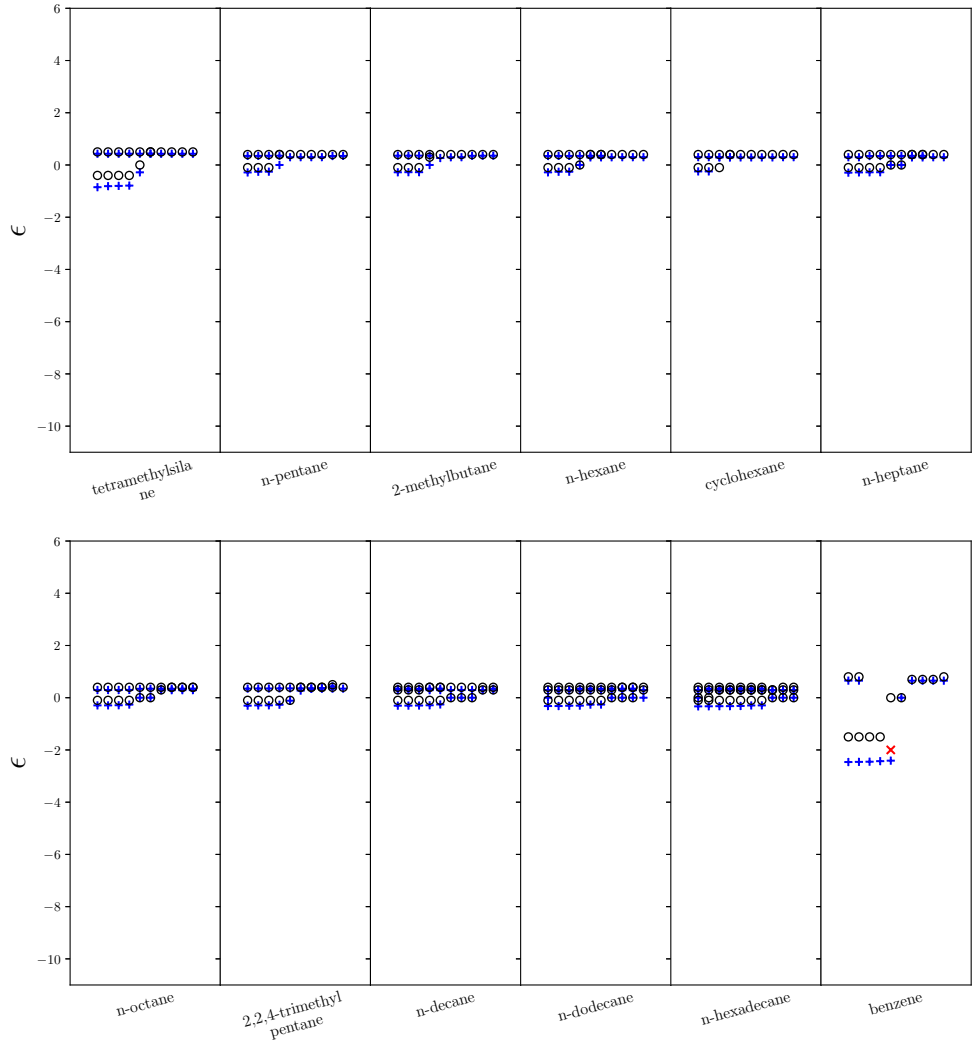


Fig. F.23: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are tetramethylsilane, n-pentane, 2-methylbutane, n-hexane, cyclohexane, n-heptane, n-octane, 2,2,4-trimethylpentane, n-decane, n-dodecane, n-hexadecane, benzene. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

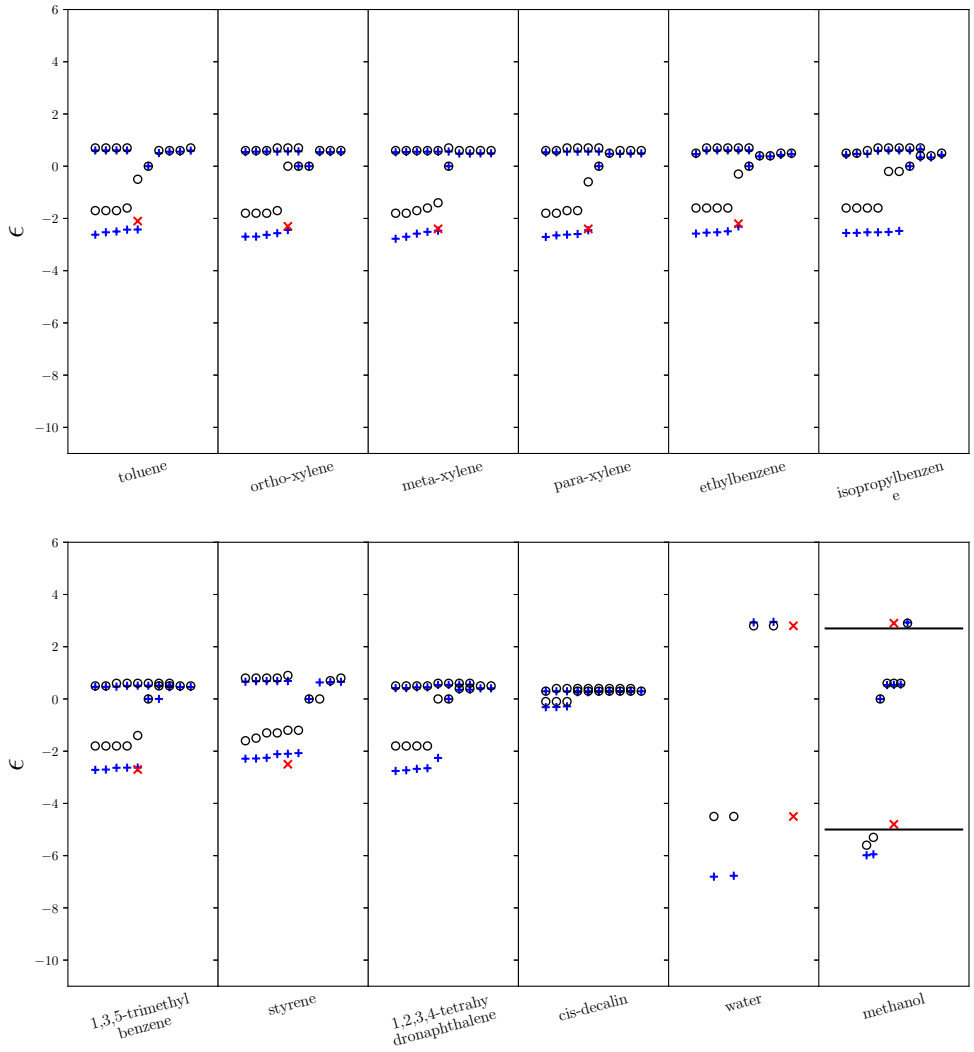


Fig. F.24: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are toluene, ortho-xylene, meta-xylene, para-xylene, ethylbenzene, isopropylbenzene, 1,3,5-trimethylbenzene, styrene, 1,2,3,4-tetrahydronaphthalene, cis-decalin, water, methanol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

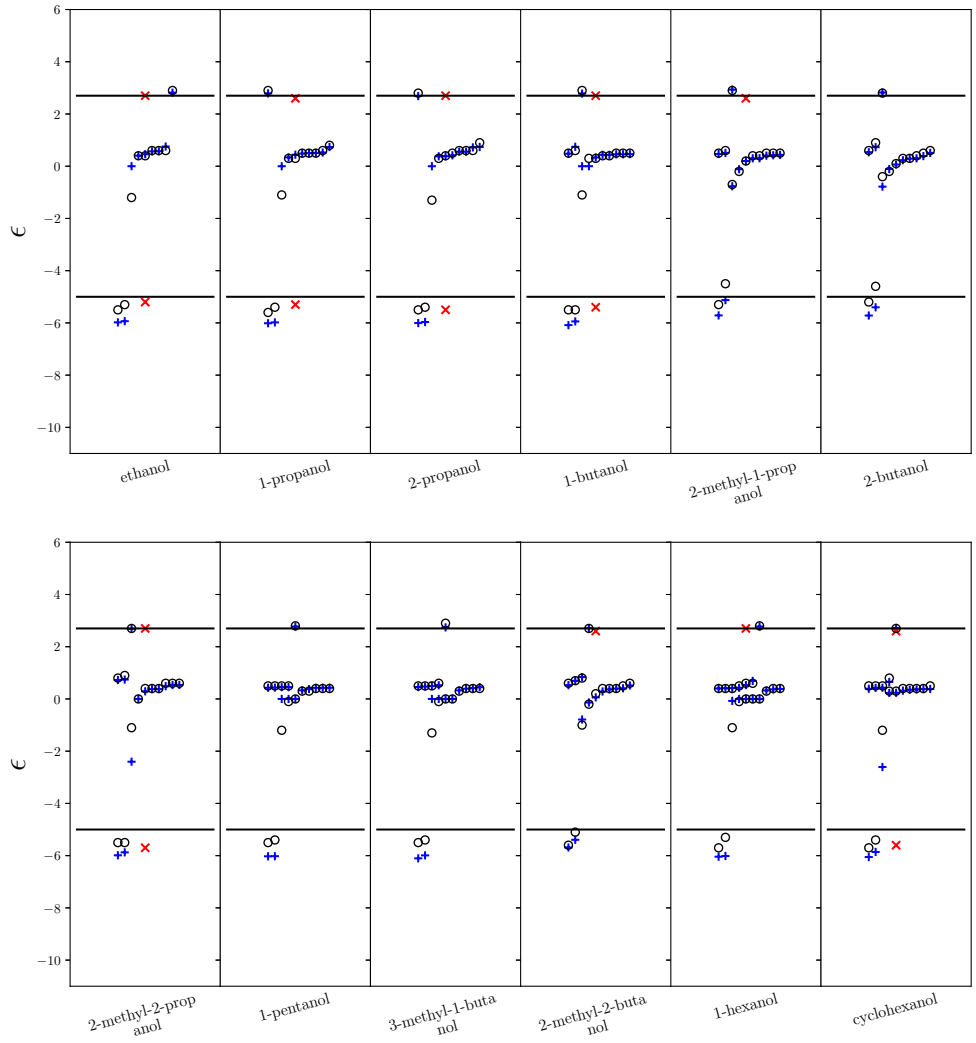


Fig. F.25: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are ethanol, 1-propanol, 2-propanol, 1-butanol, 2-methyl-1-propanol, 2-butanol, 2-methyl-2-propanol, 1-pentanol, 3-methyl-1-butanol, 2-methyl-2-butanol, 1-hexanol, cyclohexanol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

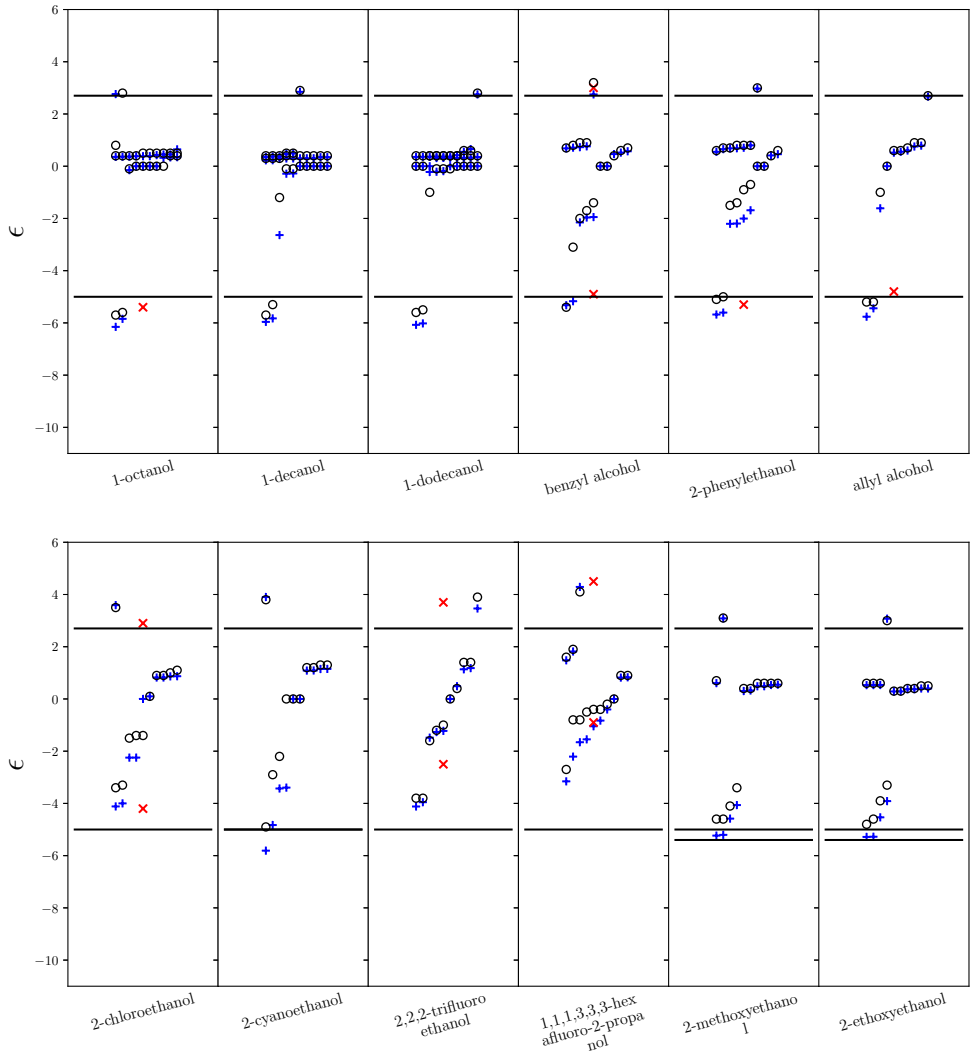


Fig. F.26: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are 1-octanol, 1-decanol, 1-dodecanol, benzyl alcohol, 2-phenylethanol, allyl alcohol, 2-chloroethanol, 2-cyanoethanol, 2,2,2-trifluoroethanol, 1,1,1,3,3,3-hexafluoro-2-propanol, 2-methoxyethanol, 2-ethoxyethanol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

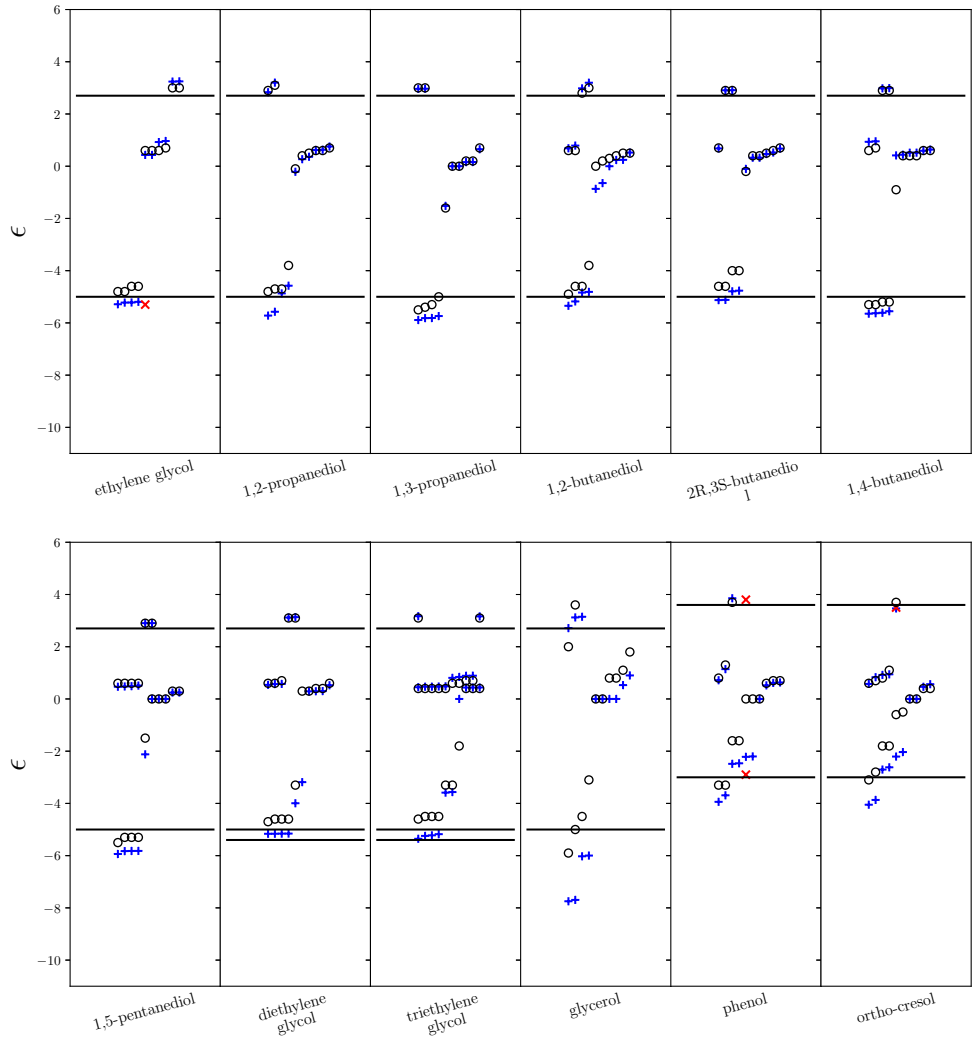


Fig. F.27: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are ethylene glycol, 1,2-propanediol, 1,3-propanediol, 1,2-butanediol, 2R,3S-butanediol, 1,4-butanediol, 1,5-pentanediol, diethylene glycol, triethylene glycol, glycerol, phenol, ortho-cresol. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

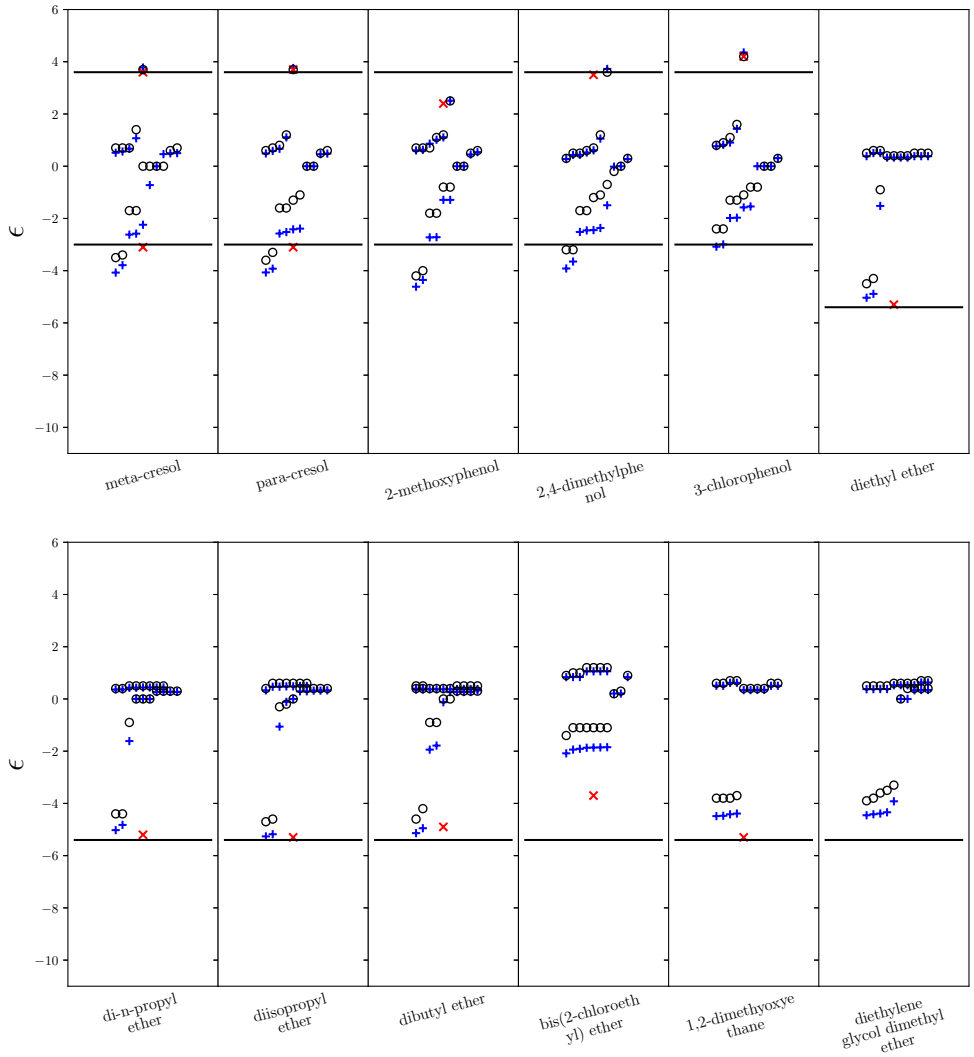


Fig. F.28: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are meta-cresol, para-cresol, 2-methoxyphenol, 2,4-dimethylphenol, 3-chlorophenol, diethyl ether, di-n-propyl ether, diisopropyl ether, dibutyl ether, bis(2-chloroethyl) ether, 1,2-dimethyoxyethane, diethylene glycol dimethyl ether. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

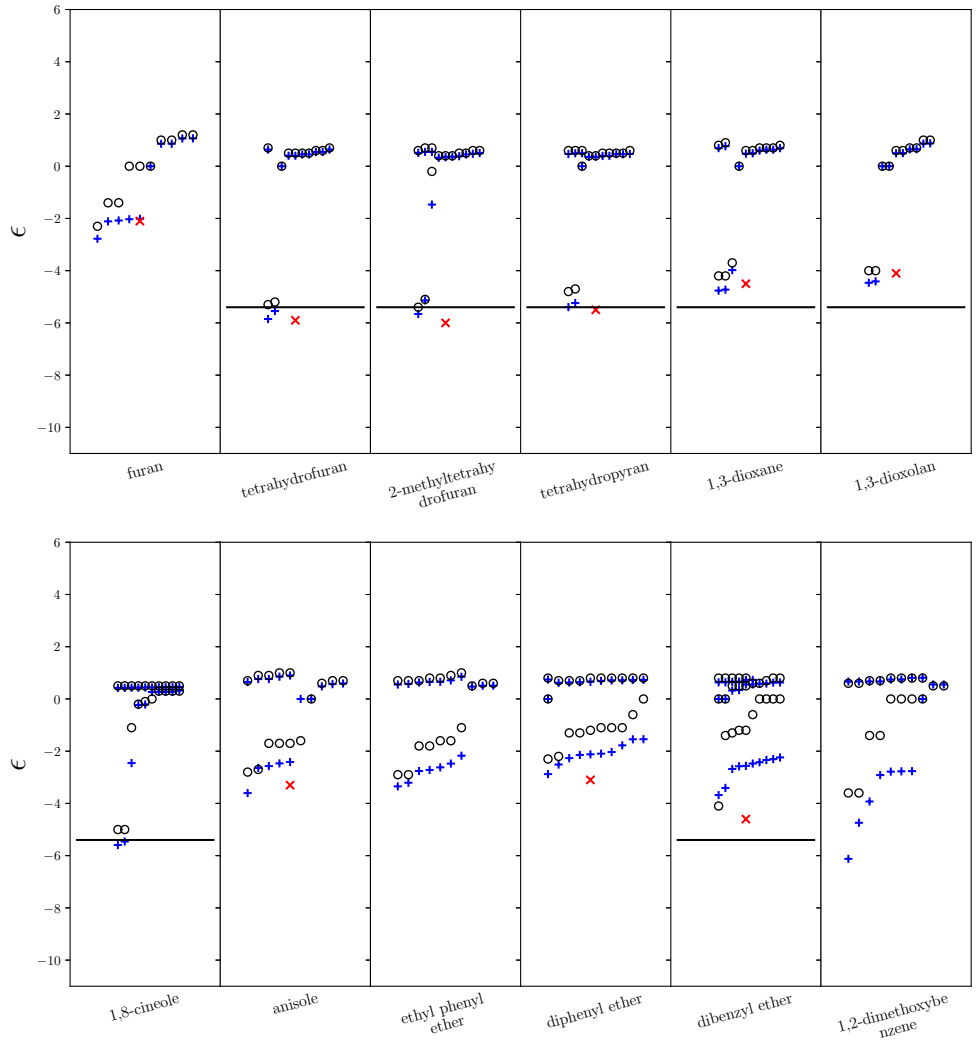


Fig. F.29: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are furan, tetrahydrofuran, 2-methyltetrahydrofuran, tetrahydropyran, 1,3-dioxane, 1,3-dioxolan, 1,8-cineole, anisole, ethyl phenyl ether, diphenyl ether, dibenzyl ether, 1,2-dimethoxybenzene. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

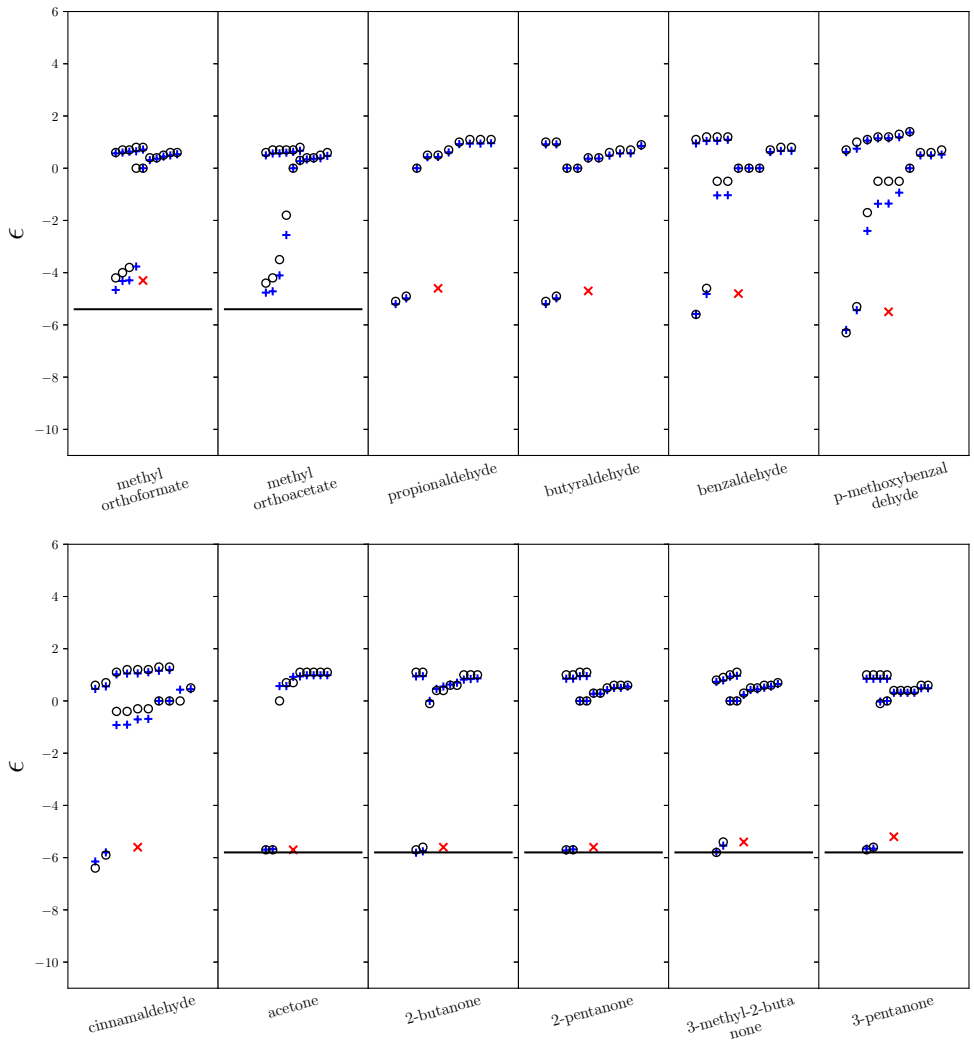


Fig. F.30: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are methyl orthoformate, methyl orthoacetate, propionaldehyde, butyraldehyde, benzaldehyde, p-methoxybenzaldehyde, cinnamaldehyde, acetone, 2-butanone, 2-pentanone, 3-methyl-2-butanone, 3-pentanone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

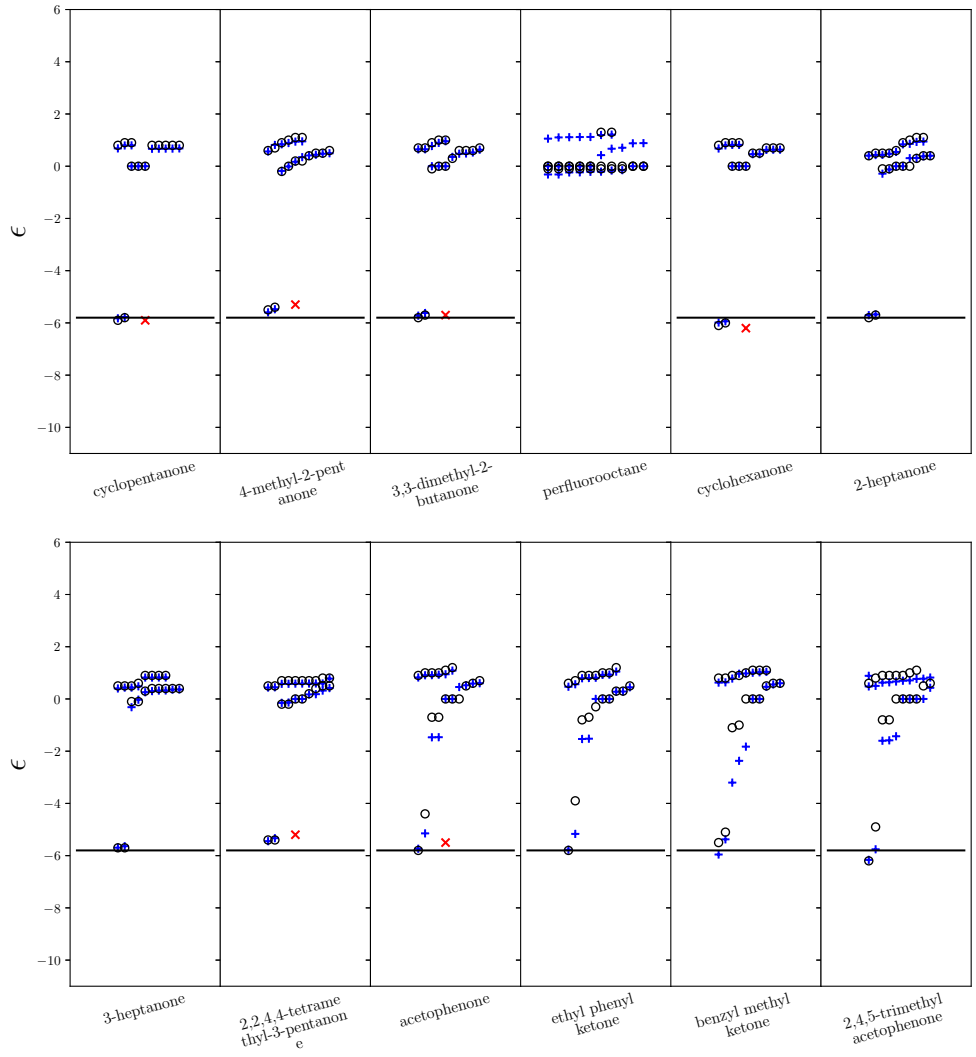


Fig. F.31: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are cyclopentanone, 4-methyl-2-pentanone, 3,3-dimethyl-2-butanone, perfluorooctane, cyclohexanone, 2-heptanone, 3-heptanone, 2,2,4,4-tetramethyl-3-pentanone, acetophenone, ethyl phenyl ketone, benzyl methyl ketone, 2,4,5-trimethylacetophenone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

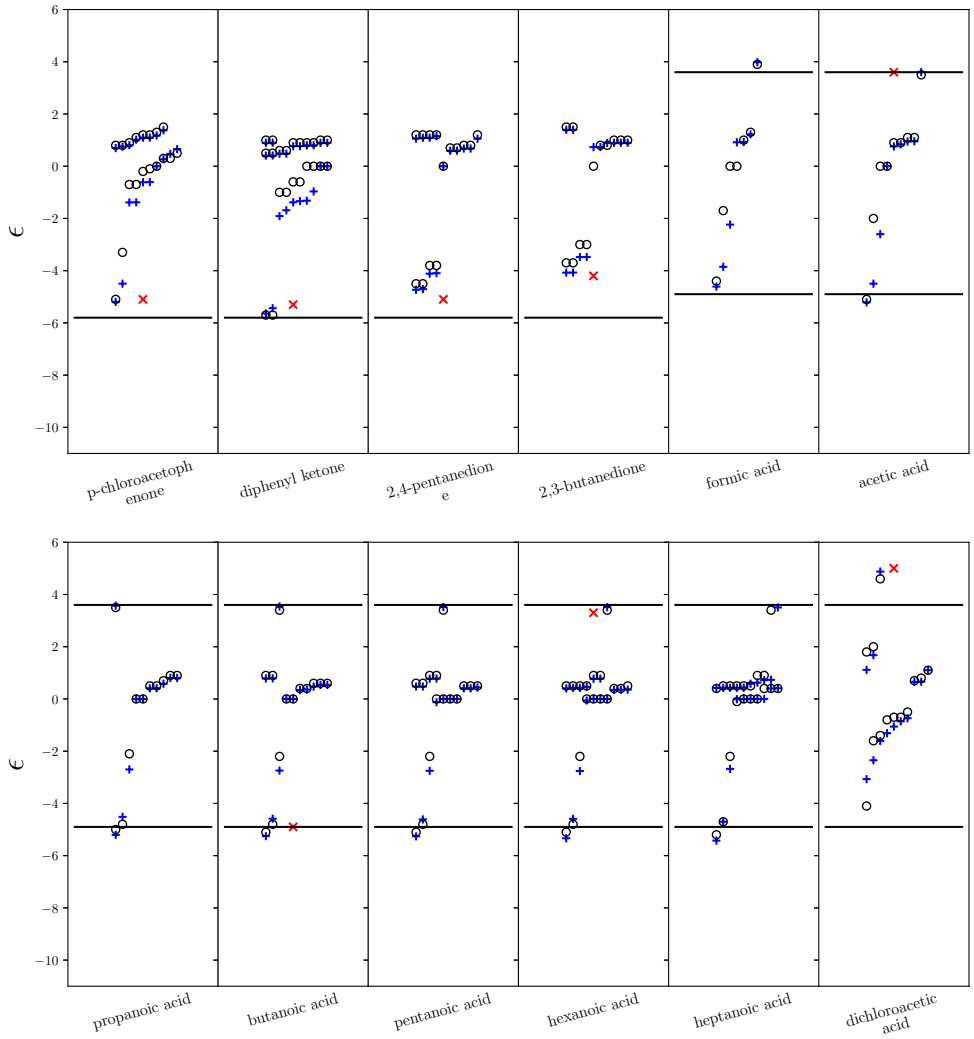


Fig. F.32: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are p-chloroacetophenone, diphenyl ketone, 2,4-pentanedione, 2,3-butanedione, formic acid, acetic acid, propanoic acid, butanoic acid, pentanoic acid, hexanoic acid, heptanoic acid, dichloroacetic acid. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

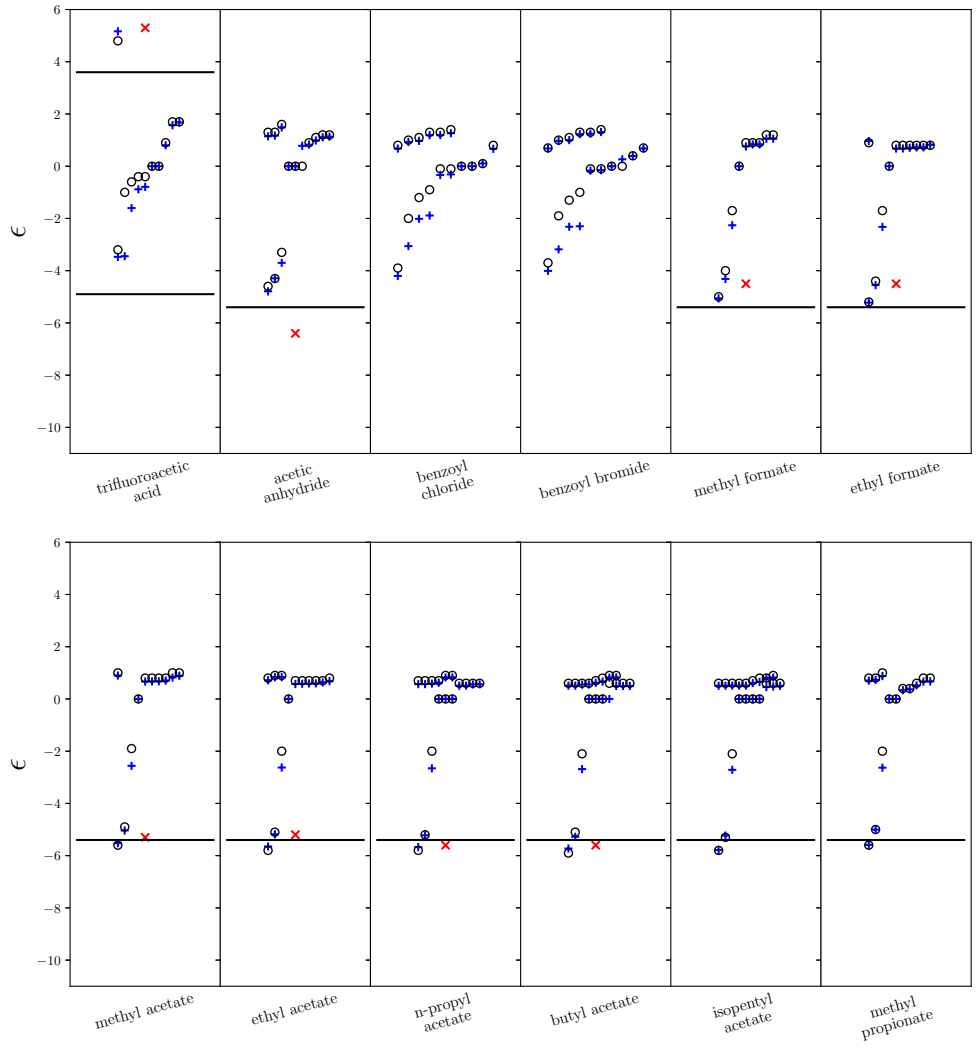


Fig. F.33: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are trifluoroacetic acid, acetic anhydride, benzoyl chloride, benzoyl bromide, methyl formate, ethyl formate, methyl acetate, ethyl acetate, n-propyl acetate, butyl acetate, isopentyl acetate, methyl propionate. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

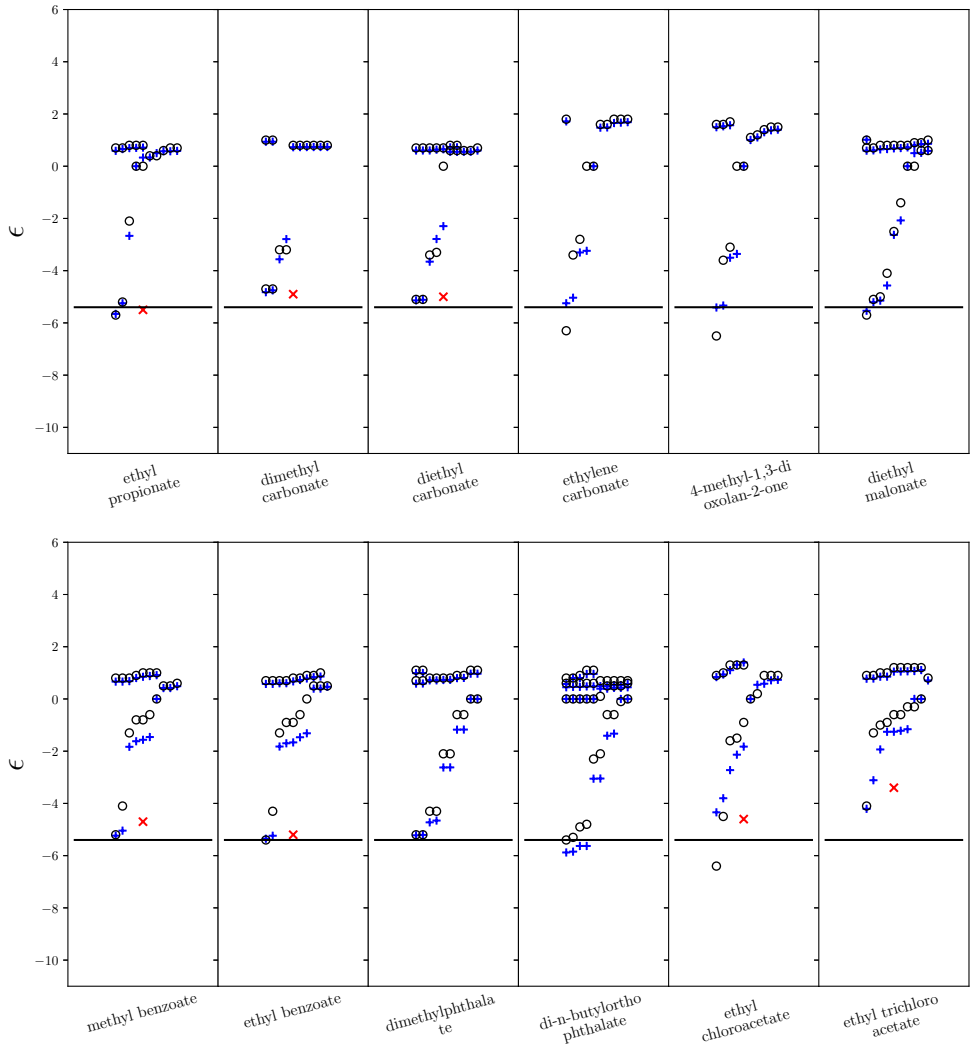


Fig. F.34: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are ethyl propionate, dimethyl carbonate, diethyl carbonate, ethylene carbonate, 4-methyl-1,3-dioxolan-2-one, diethyl malonate, methyl benzoate, ethyl benzoate, dimethylphthalate, di-n-butylorthophthalate, ethyl chloroacetate, ethyl trichloroacetate. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

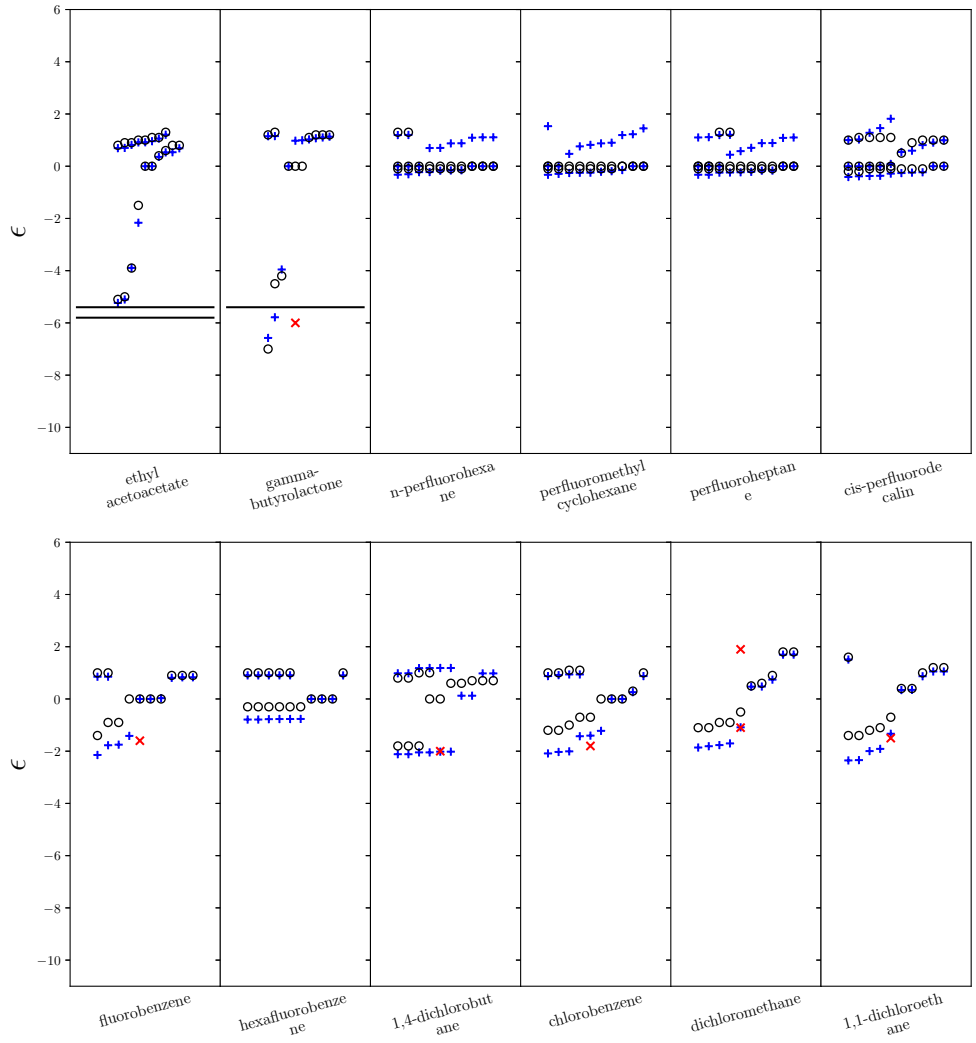


Fig. F.35: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are ethyl acetoacetate, gamma-butyrolactone, n-perfluorohexane, perfluoromethylcyclohexane, perfluoroheptane, cis-perfluorodecalin, fluorobenzene, hexafluorobenzene, 1,4-dichlorobutane, chlorobenzene, dichloromethane, 1,1-dichloroethane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

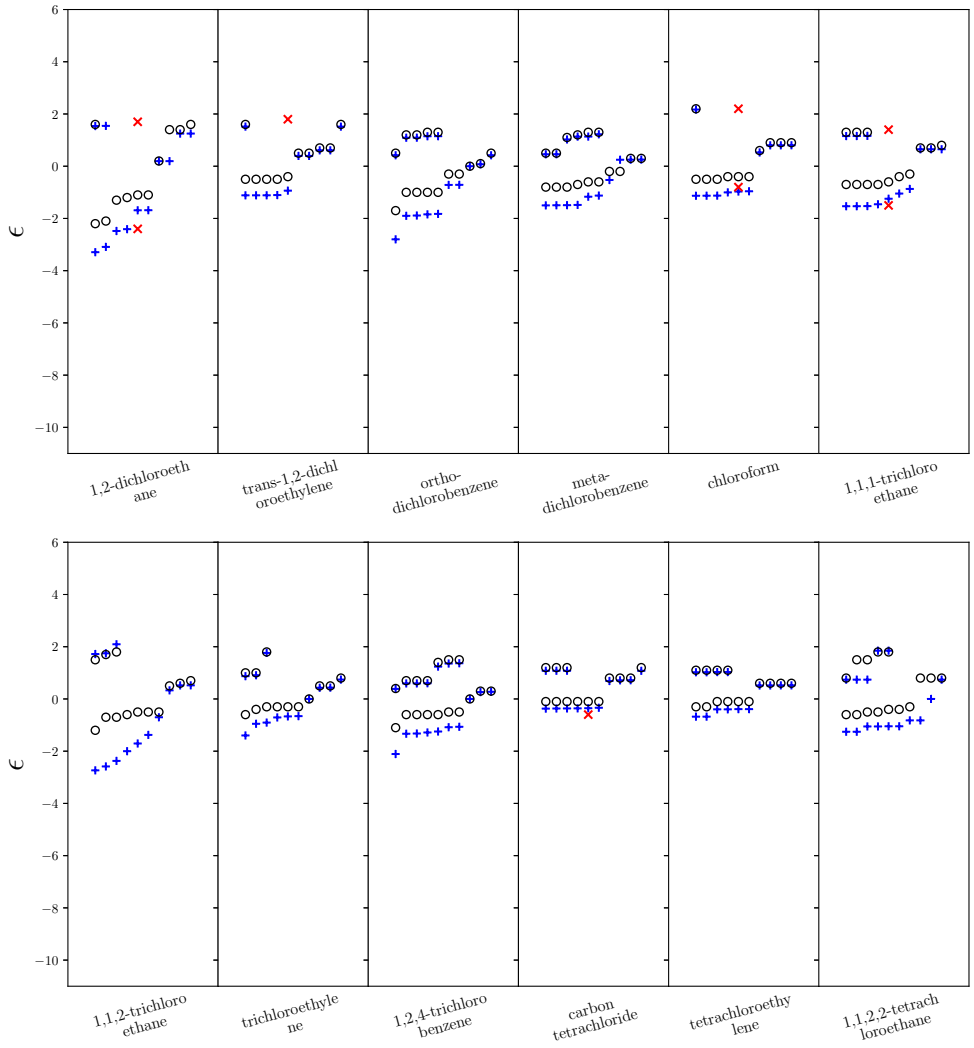


Fig. F.36: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are 1,2-dichloroethane, trans-1,2-dichloroethylene, ortho-dichlorobenzene, meta-dichlorobenzene, chloroform, 1,1,1-trichloroethane, 1,1,2-trichloroethane, trichloroethylene, 1,2,4-trichlorobenzene, carbon tetrachloride, tetrachloroethylene, 1,1,2,2-tetrachloroethane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

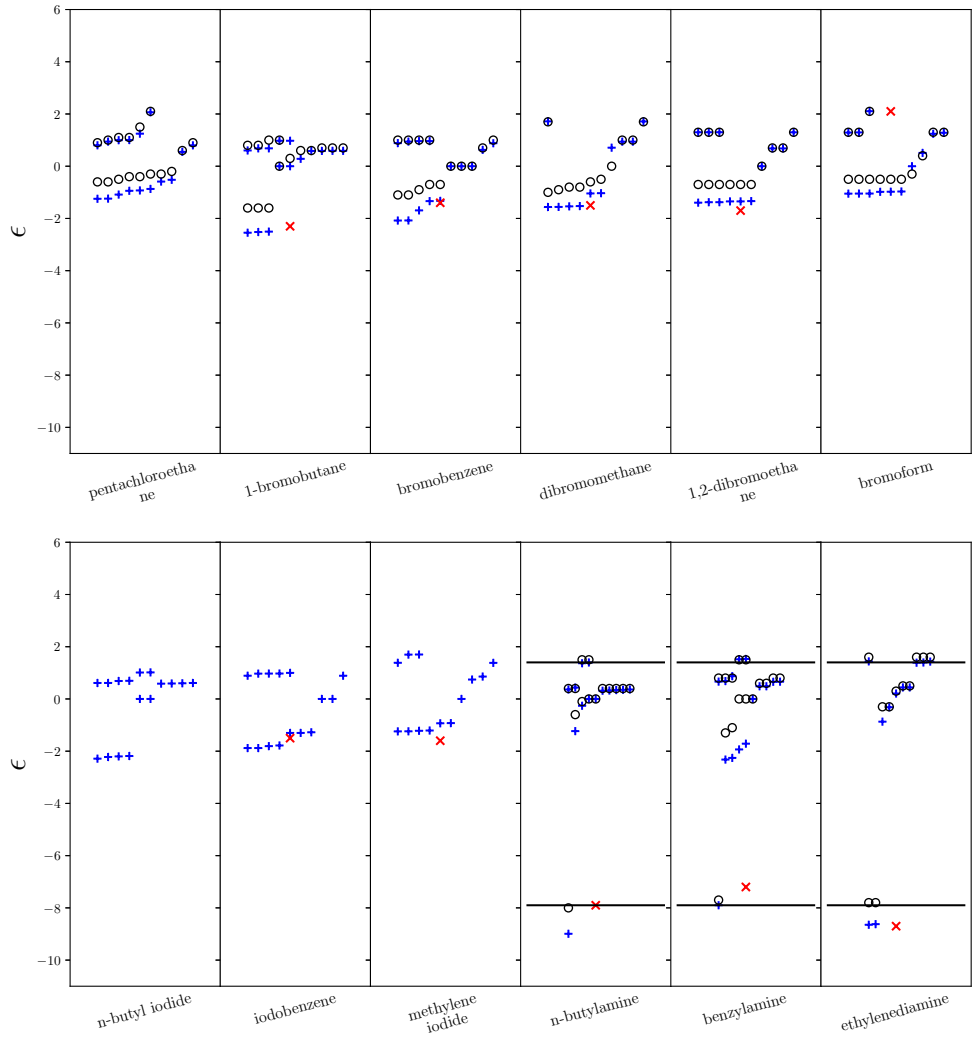


Fig. F.37: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are pentachloroethane, 1-bromobutane, bromobenzene, dibromomethane, 1,2-dibromoethane, bromoform, n-butyl iodide, iodobenzene, methylene iodide, n-butylamine, benzylamine, ethylenediamine. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

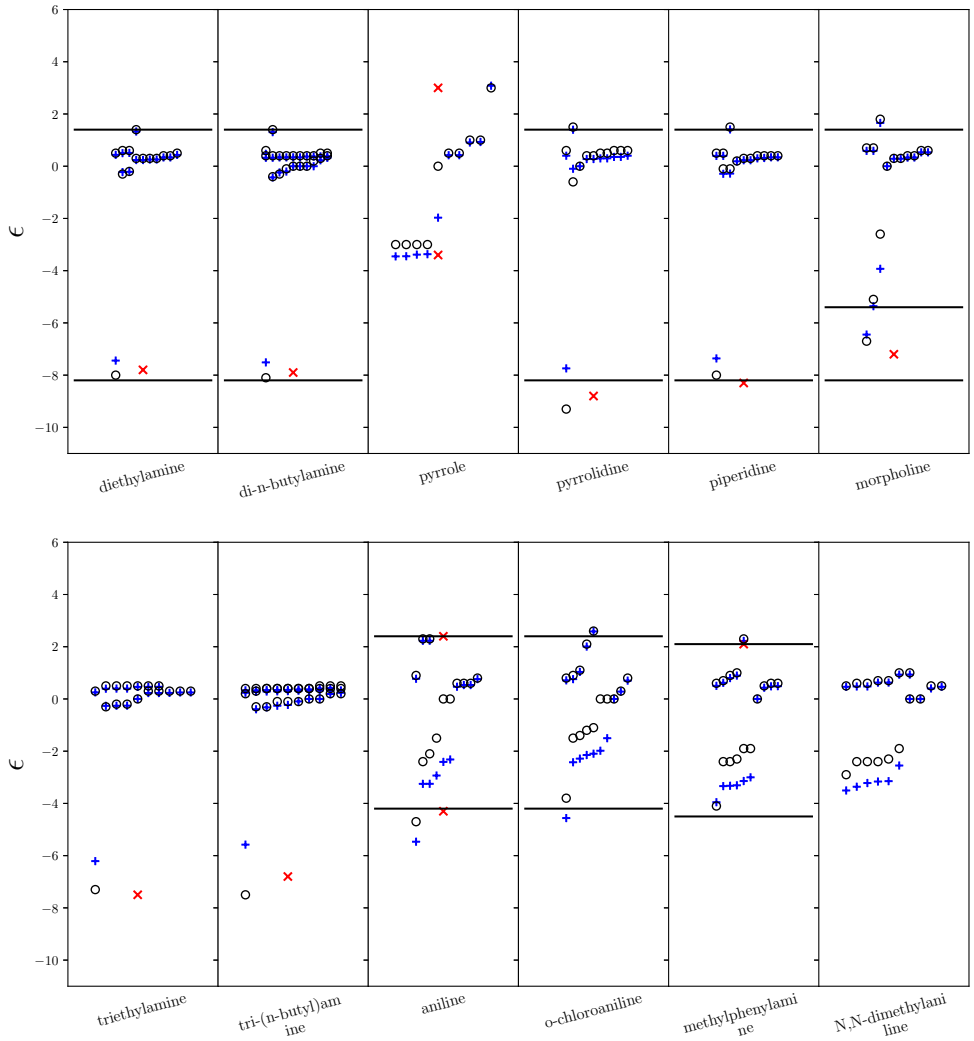


Fig. F.38: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are diethylamine, di-n-butylamine, pyrrole, pyrrolidine, piperidine, morpholine, triethylamine, tri-(n-butyl)amine, aniline, o-chloroaniline, methylphenylamine, N,N-dimethylaniline. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

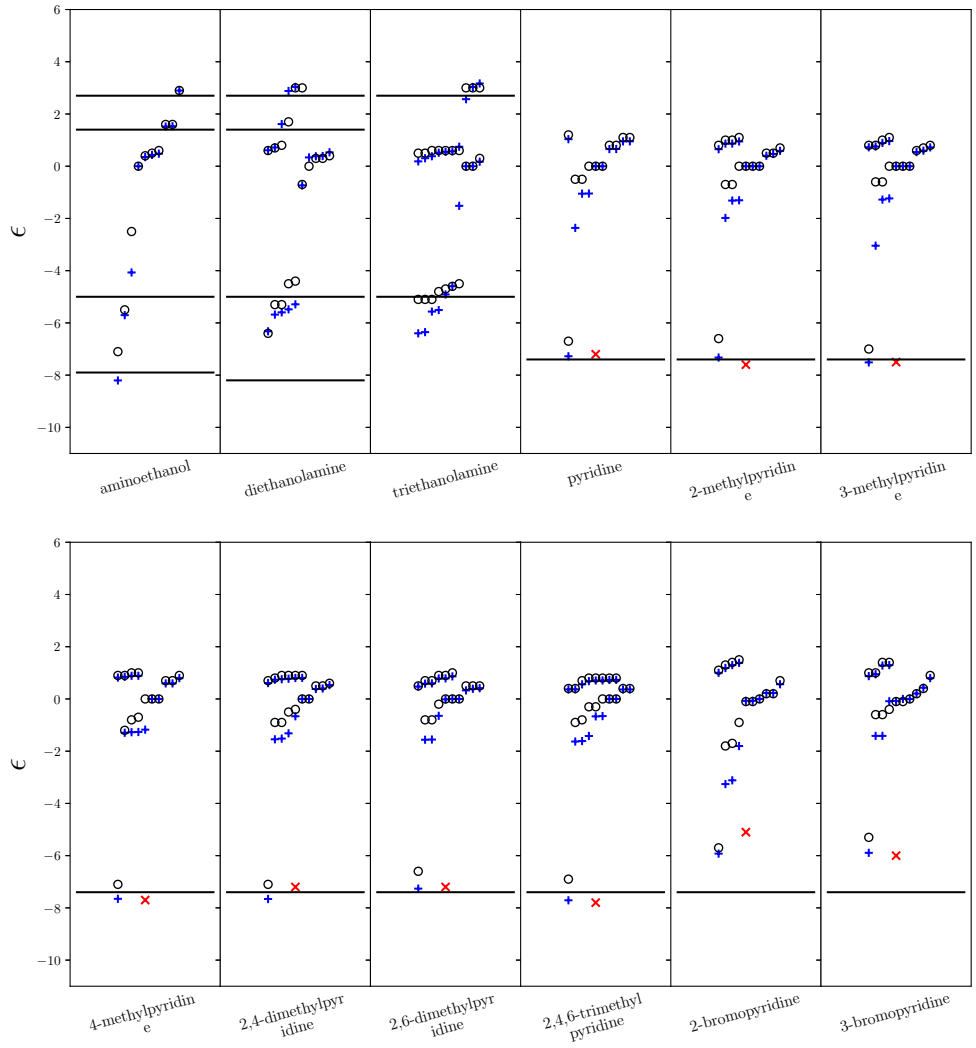


Fig. F.39: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are aminoethanol, diethanolamine, triethanolamine, pyridine, 2-methylpyridine, 3-methylpyridine, 4-methylpyridine, 2,4-dimethylpyridine, 2,6-dimethylpyridine, 2,4,6-trimethylpyridine, 2-bromopyridine, 3-bromopyridine. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

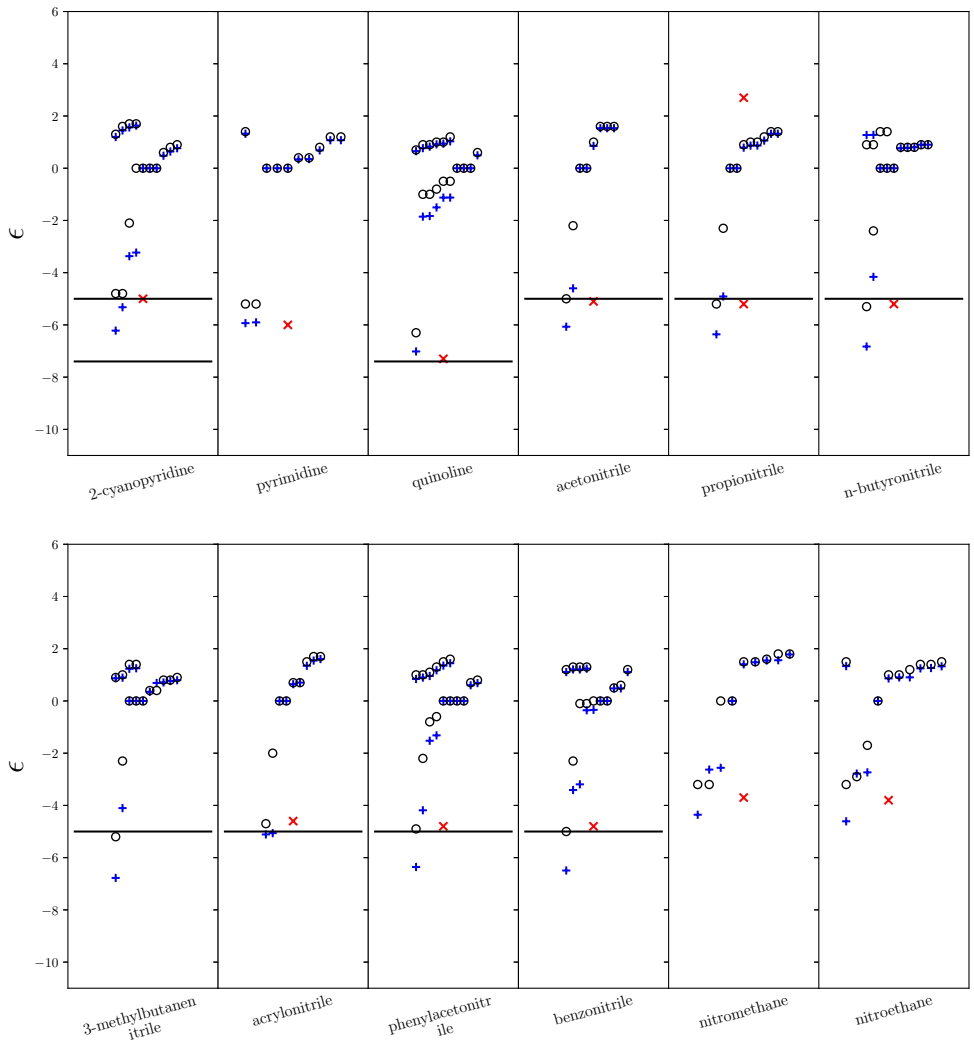


Fig. F.40: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are 2-cyanopyridine, pyrimidine, quinoline, acetonitrile, propionitrile, n-butyronitrile, 3-methylbutanenitrile, acrylonitrile, phenylacetonitrile, benzonitrile, nitromethane, nitroethane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

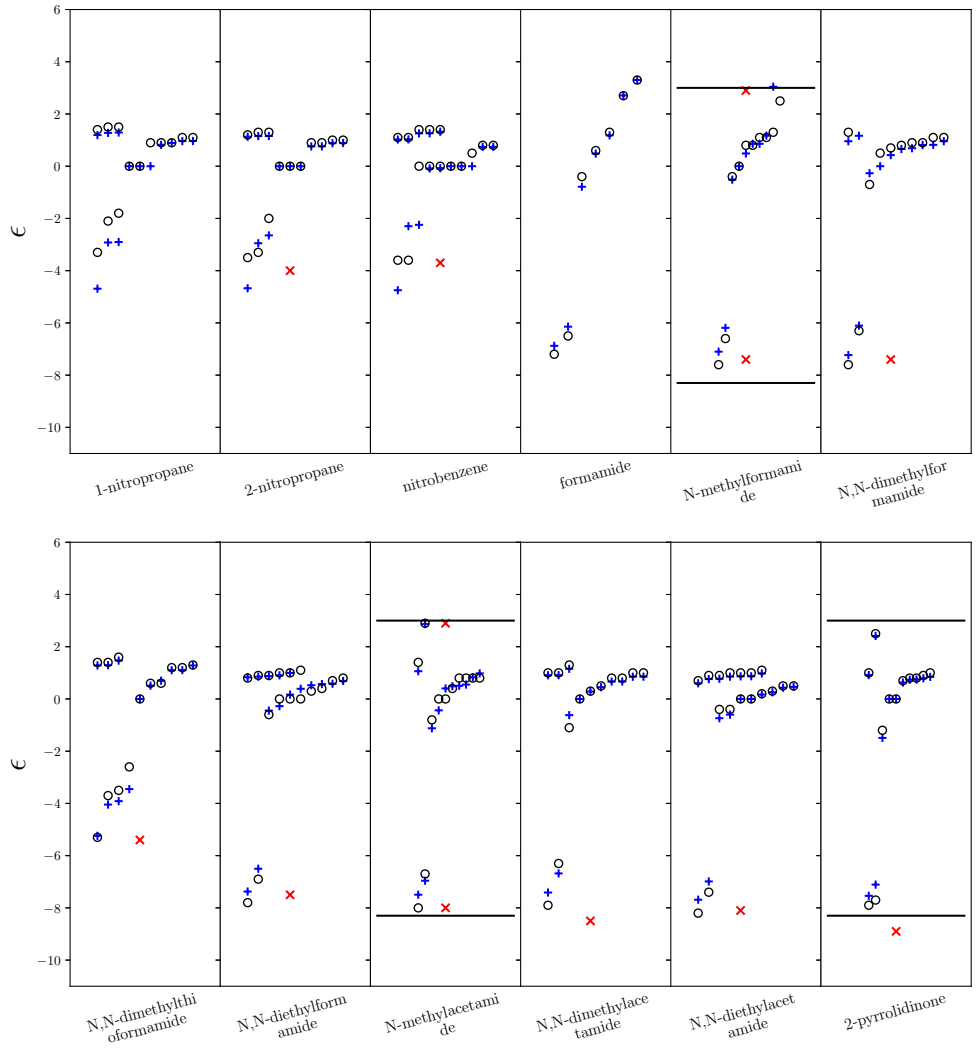


Fig. F.41: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are 1-nitropropane, 2-nitropropane, nitrobenzene, formamide, N-methylformamide, N,N-dimethylformamide, N,N-dimethylthioformamide, N,N-diethylformamide, N-methylacetamide, N,N-dimethylacetamide, N,N-diethylacetamide, 2-pyrrolidinone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

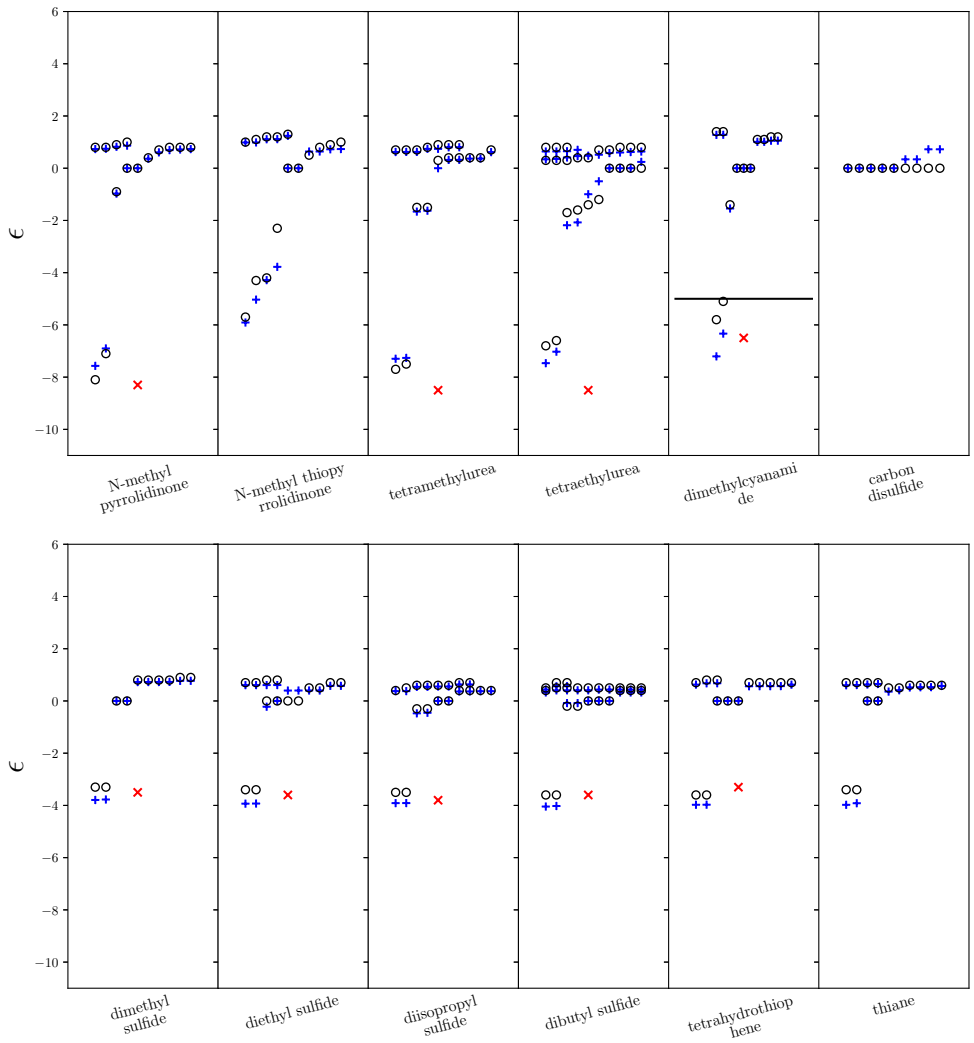


Fig. F.42: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are N-methyl pyrrolidinone, N-methyl thiopyrrolidinone, tetramethylurea, tetraethylurea, dimethylcyanamide, carbon disulfide, dimethyl sulfide, diethyl sulfide, diisopropyl sulfide, dibutyl sulfide, tetrahydrothiophene, thiane. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

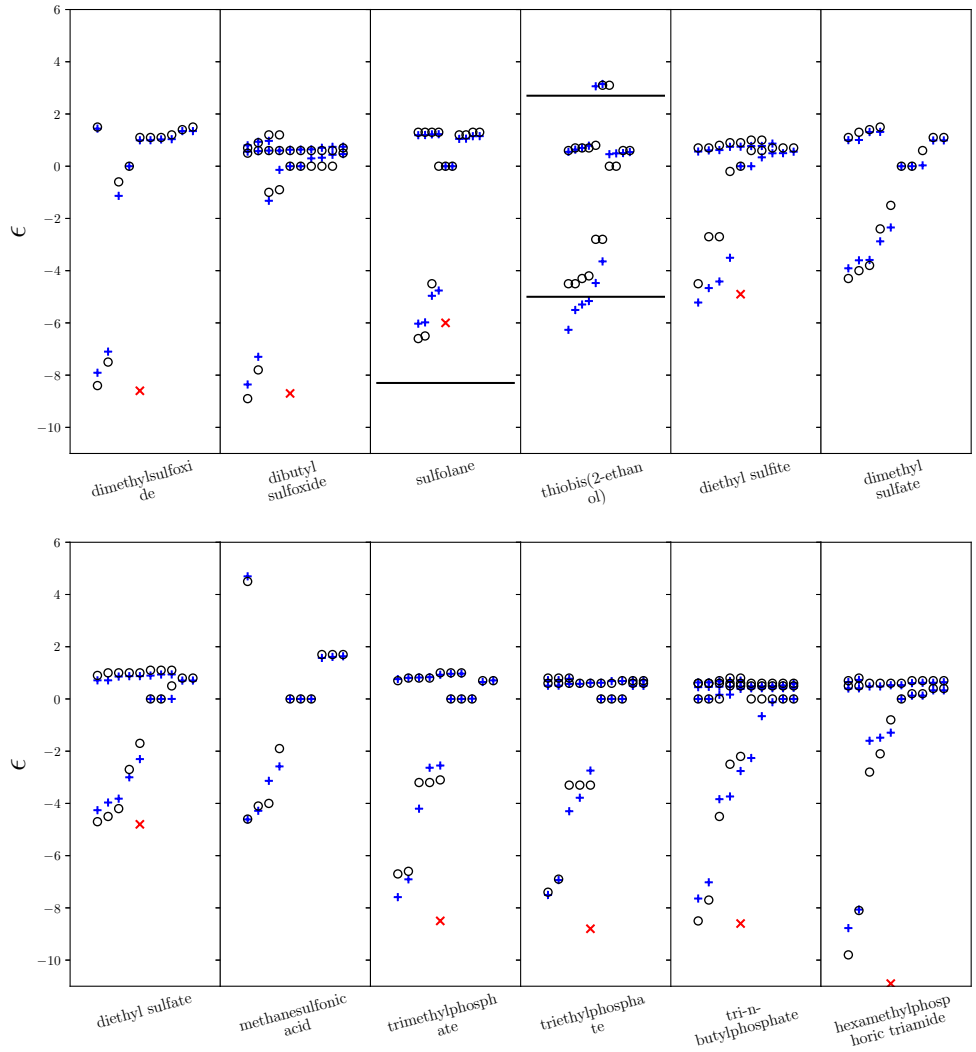


Fig. F.43: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are dimethylsulfoxide, dibutyl sulfoxide, sulfolane, thiobis(2-ethanol), diethyl sulfite, dimethyl sulfate, diethyl sulfate, methanesulfonic acid, trimethylphosphate, triethylphosphate, tri-n-butylphosphate, hexamethylphosphoric triamide. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

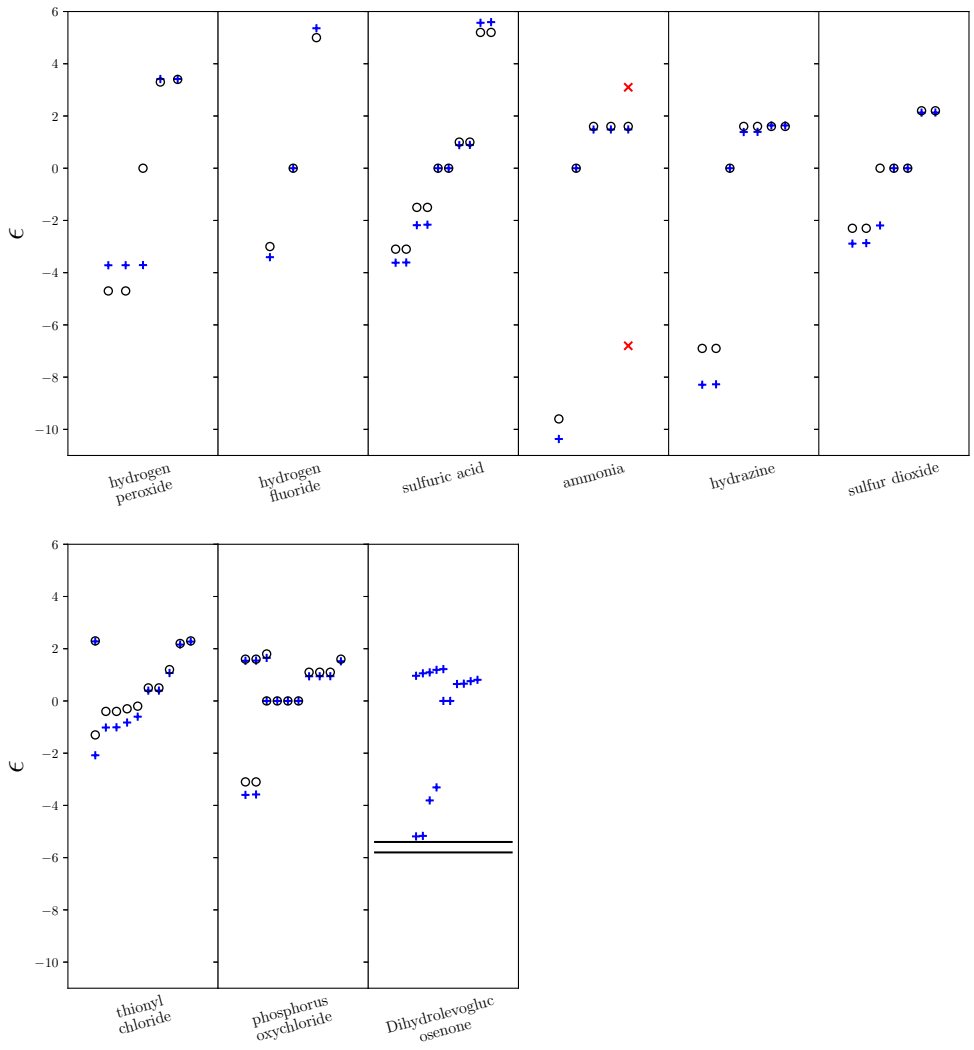


Fig. F.44: SSIP description generated by tri-surface footprinting approach. Molecules (top left to bottom right) are hydrogen peroxide, hydrogen fluoride, sulfuric acid, ammonia, hydrazine, sulfur dioxide, thionyl chloride, phosphorus oxychloride, Dihydrolevoglucosenone. Blue pluses are the calculated values, black circles are the values used in [158], black line is the mean functional group value, red crosses are for experimental values of the molecule, in D.

Appendix G

Solvent information

G.1 Pure Solvent information

G.1.1 Solvent name and concentration

Table containing information for the concentrations of the single component (pure) solvents used in table G.1. Concentrations are from [332, 333]. The ID in table G.1 was used to order the pure solvent FGIPs.

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
tetramethylsilane	1	7.33	CZDYPVPMEAXLPK-UHFFFAOYSA-N	16	1.03
n-pentane	2	8.61	OFBQJSOFQDEBGM-UHFFFAOYSA-N	14	1.05
2-methylbutane	3	8.51	QWTDNUCVQCZILF-UHFFFAOYSA-N	14	1.04
n-hexane	4	7.60	VLKZOEOYAKHREP-UHFFFAOYSA-N	16	1.05
cyclohexane	5	9.20	XDTMQSROBMDMFD-UHFFFAOYSA-N	14	1.09
n-heptane	6	6.78	IMNFDUFMRHMDMM-UHFFFAOYSA-N	18	1.06

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
n-octane	7	6.12	TVMXDCGIABBOFY-UHFFFAOYSA-N	21	1.09
2,2,4-trimethylpentane	8	6.24	NHTMVDHEPJAVLT-UHFFFAOYSA-N	19	1.04
n-decane	9	5.11	DIOQZVSQGTUSAI-UHFFFAOYSA-N	25	1.08
n-dodecane	10	4.43	SNRUBQQJIBEYMU-UHFFFAOYSA-N	29	1.09
n-hexadecane	11	3.40	DCAYPVUWAIABOU-UHFFFAOYSA-N	37	1.08
benzene	12	11.12	UHOVQNZJYSORNB-UHFFFAOYSA-N	12	1.11
toluene	13	9.35	YXFVVABEGXRONW-UHFFFAOYSA-N	14	1.10
ortho-xylene	14	8.26	CTQNGGLPUBDAKN-UHFFFAOYSA-N	16	1.11
meta-xylene	15	8.11	IVSZLXZYQVIEFR-UHFFFAOYSA-N	16	1.10
para-xylene	16	8.08	URLKBWYHVLBVBO-UHFFFAOYSA-N	16	1.09
ethylbenzene	17	8.13	YNQLUTRBYVCPMQ-UHFFFAOYSA-N	16	1.10
isopropylbenzene	18	7.14	RWGFKTVRMDUZSP-UHFFFAOYSA-N	18	1.09
1,3,5-trimethylbenzene	19	7.17	AUHZEENZYGFFBQ-UHFFFAOYSA-N	18	1.09
styrene	20	8.66	PPBRXRYQALVLMV-UHFFFAOYSA-N	15	1.10
1,2,3,4-tetrahydronaphthalene	21	7.30	CXWXQJXEFPUFDZ-UHFFFAOYSA-N	18	1.10
cis-decalin	22	6.46	NNBZCPXTIHJBJL-AOOOYVTPSA-N	19	1.06

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
water	23	55.35	XLYOFNOQVPJJNP-UHFFFAOYSA-N	4	1.48
methanol	24	24.57	OKKJLVBELUTLKV-UHFFFAOYSA-N	7	1.29
ethanol	25	17.03	LFQSCWFLJHTTHZ-UHFFFAOYSA-N	9	1.21
1-propanol	26	13.32	BDERNNFJNOPAEC-UHFFFAOYSA-N	11	1.18
2-propanol	27	13.00	KFZMGEQAYNKOFK-UHFFFAOYSA-N	11	1.16
1-butanol	28	10.87	LRHPLDYGYMQRHN-UHFFFAOYSA-N	13	1.15
2-methyl-1-propanol	29	10.76	ZXEKIIBDNHEJCQ-UHFFFAOYSA-N	13	1.15
2-butanol	30	10.83	BTANRVKWQNVYAZ-SCSAIBSYSA-N	13	1.15
2-butanol	30	10.83	BTANRVKWQNVYAZ-SCSAIBSYSA-N	13	1.15
2-butanol	30	10.83	BTANRVKWQNVYAZ-BYPYZUCNSA-N	13	1.15
2-butanol	30	10.83	BTANRVKWQNVYAZ-BYPYZUCNSA-N	13	1.15
2-methyl-2-propanol	31	10.54	DKGAVHZHDRPRBM-UHFFFAOYSA-N	13	1.13
1-pentanol	32	9.22	AMQJEAYHLZJPGS-UHFFFAOYSA-N	15	1.14
3-methyl-1-butanol	33	9.16	PHTQWCKDNZKARW-UHFFFAOYSA-N	15	1.13
2-methyl-2-butanol	34	9.13	MSXVEPNJUHWQHW-UHFFFAOYSA-N	14	1.09
1-hexanol	35	7.99	ZSIAUFGUXNUGDI-UHFFFAOYSA-N	17	1.13

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
cyclohexanol	36	9.68	HPXRVTGHNJAIIH-UHFFFAOYSA-N	15	1.17
1-octanol	37	6.31	KBPLFHHGFOOTCA-UHFFFAOYSA-N	21	1.11
1-decanol	38	5.24	MWKFXSUHUHTGQN-UHFFFAOYSA-N	26	1.13
1-dodecanol	39	4.46	LQZZUXJYWNFBMV-UHFFFAOYSA-N	30	1.12
benzyl alcohol	40	9.63	WVDDGKGOMKODPV-UHFFFAOYSA-N	15	1.17
2-phenylethanol	41	8.38	WRMNZCZEMHIOCP-UHFFFAOYSA-N	17	1.16
allyl alcohol	42	14.58	XXROGKLTLUQVRX-UHFFFAOYSA-N	10	1.17
2-chloroethanol	43	15.48	SZIFAVKTNFCBPC-UHFFFAOYSA-N	11	1.28
2-cyanoethanol	44	14.64	WSGYTJNNHPZFKR-UHFFFAOYSA-N	11	1.24
2,2,2-trifluoroethanol	45	13.82	RHQDFWAXVIIEBN-UHFFFAOYSA-N	10	1.14
1,1,1,3,3,3-hexafluoro-2-propanol	46	9.56	BYEAHWXPCBROCE-UHFFFAOYSA-N	13	1.07
2-methoxyethanol	47	12.61	XNWFRZJHXBZDAG-UHFFFAOYSA-N	12	1.20
2-ethoxyethanol	48	10.27	ZNQVEEAIQZEUHB-UHFFFAOYSA-N	14	1.16
ethylene glycol	49	17.89	LYCAIKOWRPUZTN-UHFFFAOYSA-N	10	1.32
1,2-propanediol	50	13.57	DNIAPMSPPWPWGF-GSVOUGTGSA-N	12	1.25
1,3-propanediol	51	13.79	YPFDHNVEDLHUCE-UHFFFAOYSA-N	12	1.26

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
1,2-butanediol	52	11.09	BMRWNKZVCUKKSR-SCSAIBSYSA-N	14	1.22
(2R,3S)-2,3-butanediol	53	10.95	OWBTYPJTUOEWEK-ZXZARUISSA-N	14	1.21
1,4-butanediol	54	11.24	WERYXYBDKMZEQL-UHFFFAOYSA-N	14	1.23
1,5-pentanediol	55	9.47	ALQSHHUCVQOPAS-UHFFFAOYSA-N	16	1.20
diethylene glycol	56	10.52	MTHSVFCYNBDYFN-UHFFFAOYSA-N	15	1.23
triethylene glycol	57	7.46	ZIBGPFATKBEMQZ-UHFFFAOYSA-N	21	1.22
glycerol	58	13.66	PEDCQBHIVMGVHV-UHFFFAOYSA-N	13	1.31
phenol	59	11.39	ISWSIDIOOBJBQZ-UHFFFAOYSA-N	13	1.18
ortho-cresol	60	9.62	QWVGKYWNOKOFNN-UHFFFAOYSA-N	15	1.17
meta-cresol	61	9.53	RLSSMJSEOOYNOY-UHFFFAOYSA-N	15	1.16
para-cresol	62	9.44	IWDCLRJOBJJRNH-UHFFFAOYSA-N	15	1.15
2-methoxyphenol	63	9.09	LHGVFZTZFXWLCP-UHFFFAOYSA-N	16	1.17
2,4-dimethylphenol	64	8.32	KUFFULVDNCHOFZ-UHFFFAOYSA-N	17	1.15
3-chlorophenol	65	9.87	HORNXRXVQWOLPJ-UHFFFAOYSA-N	15	1.18
diethyl ether	66	9.55	RTZKZFJDLAIYFH-UHFFFAOYSA-N	13	1.07
di-n-propyl ether	67	7.27	POLCUAVZOMRGSN-UHFFFAOYSA-N	18	1.10

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
diisopropyl ether	68	7.03	ZAFNJMIOTHYJRJ-UHFFFAOYSA-N	17	1.04
dibutyl ether	69	5.87	DURPTKYDGMDSBL-UHFFFAOYSA-N	22	1.09
bis(2-chloroethyl) ether	70	8.48	ZNSMNVMLTJELDZ-UHFFFAOYSA-N	17	1.17
1,2-dimethyoxyethane	71	9.57	XTHFKEDIFFGKHM-UHFFFAOYSA-N	14	1.12
diethylene glycol dimethyl ether	72	7.00	SBZXBUIDTXKZTM-UHFFFAOYSA-N	20	1.15
furan	73	13.68	YLQBMQCUIZJEEH-UHFFFAOYSA-N	10	1.13
tetrahydrofuran	74	12.25	WYURNTSHIVDZCO-UHFFFAOYSA-N	11	1.12
2-methyltetrahydrofuran	75	9.91	JWUJQDFVADABEY-RXMQYKEDSA-N	13	1.09
tetrahydropyran	76	10.18	DHXVGJBLRPWPCS-UHFFFAOYSA-N	13	1.11
1,3-dioxane	77	11.67	VDFVNEFVBPFDSB-UHFFFAOYSA-N	12	1.15
1,3-dioxolan	78	14.37	WNXJIVFYUVYPPR-UHFFFAOYSA-N	11	1.23
1,8-cineole	79	5.96	WEEGYLXZBRQIMU-WAAGHKOSSA-N	20	1.04
anisole	80	9.15	RDOXTESZEPMUJZ-UHFFFAOYSA-N	15	1.13
ethyl phenyl ether	81	7.87	DLRJIFUOBPOJNS-UHFFFAOYSA-N	17	1.12
diphenyl ether	82	6.29	USIUVYZYUHIAEV-UHFFFAOYSA-N	21	1.11
dibenzyl ether	83	5.19	MHDVGSVTJDSBDK-UHFFFAOYSA-N	26	1.12

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
1,2-dimethoxybenzene	84	7.83	ABDKAPXRBAPSQN-UHFFFAOYSA-N	18	1.15
methyl orthoformate	85	8.89	PYOKUURKVVELLB-UHFFFAOYSA-N	15	1.11
methyl orthoacetate	86	7.37	HDPNBNXLBDFELL-UHFFFAOYSA-N	16	1.03
propionaldehyde	87	13.62	NBBJYMSMWIIQGU-UHFFFAOYSA-N	10	1.13
butyraldehyde	88	11.05	ZTQSAGDEMFDKMZ-UHFFFAOYSA-N	12	1.11
benzaldehyde	89	9.84	HUMNYLRZRPPJDN-UHFFFAOYSA-N	14	1.13
p-methoxybenzaldehyde	90	8.23	ZRSNZINYAWTAHE-UHFFFAOYSA-N	17	1.14
cinnamaldehyde	91	7.95	KJPRLNWUNMBNBZ-QPJJXVBHSA-N	18	1.16
acetone	92	13.51	CSCPPACGZOOCGX-UHFFFAOYSA-N	10	1.12
2-butanone	93	10.93	ZWEHNKRNPOVVGH-UHFFFAOYSA-N	12	1.10
2-pentanone	94	9.30	XNLICIUVMPYHGG-UHFFFAOYSA-N	14	1.10
3-methyl-2-butanone	95	9.35	SYBYTAAJFKOIEJ-UHFFFAOYSA-N	14	1.10
3-pentanone	96	9.40	FDPIMTJIUBPUKL-UHFFFAOYSA-N	14	1.10
cyclopentanone	97	11.24	BGTOWKSIORTVQH-UHFFFAOYSA-N	13	1.17
4-methyl-2-pentanone	98	7.96	NTIZESTWPVYFNL-UHFFFAOYSA-N	16	1.08
3,3-dimethyl-2-butanone	99	8.00	PJGSXYOJTGTZAV-UHFFFAOYSA-N	15	1.05

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
perfluorooctane	100	4.03	YVBBRRALBYAZBM-UHFFFAOYSA-N	27	0.98
cyclohexanone	101	9.60	JHIVVAPYMSGYDF-UHFFFAOYSA-N	14	1.12
2-heptanone	102	7.11	CATSNJVOTSVZJV-UHFFFAOYSA-N	19	1.12
3-heptanone	103	7.11	NGAZZOYFWWSOGK-UHFFFAOYSA-N	19	1.12
2,2,4,4-tetramethyl-3-pentanone	104	5.77	UIQGEWJEWJMQSL-UHFFFAOYSA-N	20	1.02
acetophenone	105	8.52	KWOLFJPFCHCOCG-UHFFFAOYSA-N	16	1.13
ethyl phenyl ketone	106	7.53	KRIOVPPHQSLHCZ-UHFFFAOYSA-N	18	1.12
benzyl methyl ketone	107	7.57	QCCDLTOVEPVEJK-UHFFFAOYSA-N	18	1.13
2,4,5-trimethylacetophenone	108	7.50	GENBEGZNCBFHSU-UHFFFAOYSA-N	22	1.26
p-chloroacetophenone	109	7.71	BUZYGTVTZYSBCU-UHFFFAOYSA-N	18	1.14
diphenyl ketone	110	6.08	RWCCWEUUXYIKHB-UHFFFAOYSA-N	22	1.12
2,4-pentanedione	111	9.71	YRKCREAYFQTBPV-UHFFFAOYSA-N	14	1.13
2,3-butanedione	112	11.39	QSJXEFYPDANLFS-UHFFFAOYSA-N	13	1.18
formic acid	113	26.38	BDAGIHXWWSANSR-UHFFFAOYSA-N	7	1.34
acetic acid	114	17.39	QTBSBXVTEAMEQO-UHFFFAOYSA-N	9	1.22
propanoic acid	115	13.33	XBDQKXXYIPTUBI-UHFFFAOYSA-N	11	1.18

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
butanoic acid	116	10.82	FERIUCNNQQJTOY-UHFFFAOYSA-N	13	1.15
pentanoic acid	117	9.15	NQPDZGIKBAWPEJ-UHFFFAOYSA-N	15	1.13
hexanoic acid	118	7.95	FUZZWVXGSFPDMH-UHFFFAOYSA-N	17	1.12
heptanoic acid	119	7.06	MNWFXJYAOYHMED-UHFFFAOYSA-N	20	1.15
dichloroacetic acid	120	12.12	JXTHNDFMNIQAHM-UHFFFAOYSA-N	13	1.23
trifluoroacetic acid	121	12.97	DTQVDTLACAAQTR-UHFFFAOYSA-N	11	1.16
acetic anhydride	122	10.54	WFDIJRYMOXRFFG-UHFFFAOYSA-N	14	1.18
benzoyl chloride	123	8.62	PASDCCFISLVPSO-UHFFFAOYSA-N	16	1.14
benzoyl bromide	124	8.48	AQIHMSVIAGNIDM-UHFFFAOYSA-N	16	1.12
methyl formate	125	16.10	TZIHFWKZFHZASV-UHFFFAOYSA-N	9	1.17
ethyl formate	126	12.36	WBJINCZRORDGAQ-UHFFFAOYSA-N	11	1.13
methyl acetate	127	12.53	KXKVLQRXCPHEJC-UHFFFAOYSA-N	11	1.14
ethyl acetate	128	10.15	XEKOWRVHYACXOJ-UHFFFAOYSA-N	13	1.11
n-propyl acetate	129	8.65	YKYONYBAUNKHLG-UHFFFAOYSA-N	16	1.14
butyl acetate	130	7.55	DKPFZGUDAPQIHT-UHFFFAOYSA-N	18	1.13
isopentyl acetate	131	6.66	MLFHJEHSLIIPHL-UHFFFAOYSA-N	19	1.08

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
methyl propionate	132	10.31	RJUFJBKOKNCXHH-UHFFFAOYSA-N	13	1.12
ethyl propionate	133	8.66	FKRCODPIKNYEAC-UHFFFAOYSA-N	16	1.14
dimethyl carbonate	134	11.88	IEJIGPNLZYLLBP-UHFFFAOYSA-N	12	1.16
diethyl carbonate	135	8.21	OIFBSDVPJOWBCH-UHFFFAOYSA-N	17	1.14
ethylene carbonate	136	15.12	KMTRUDSVKNLOMY-UHFFFAOYSA-N	11	1.26
4-methyl-1,3-dioxolan-2-one	137	11.75	RUOJZAUFBMNUDX-GSVOUGTGSA-N	13	1.20
diethyl malonate	138	6.56	IYXGSMUGOJNHAZ-UHFFFAOYSA-N	21	1.13
methyl benzoate	139	7.96	QPJVMBTYPHYUOC-UHFFFAOYSA-N	17	1.12
ethyl benzoate	140	6.94	MTZQAGJQAFMTAQ-UHFFFAOYSA-N	19	1.11
dimethylphthalate	141	6.13	NIQCNGHVCWTJSM-UHFFFAOYSA-N	22	1.12
di-n-butylorthophthalate	142	3.75	DOIRQSBPFJWKBE-UHFFFAOYSA-N	36	1.12
ethyl chloroacetate	143	10.26	VEUUMBGHMNQHGO-UHFFFAOYSA-N	15	1.21
ethyl trichloroacetate	144	7.23	SJMLNDPIJZBEKY-UHFFFAOYSA-N	19	1.13
ethyl acetoacetate	145	7.85	XYIBRDXRRQCHLP-UHFFFAOYSA-N	18	1.15
gamma-butyrolactone	146	13.07	YEJRWHAVMIAJKC-UHFFFAOYSA-N	12	1.22
n-perfluorohexane	147	4.97	ZJIJAJXFLBMLCK-UHFFFAOYSA-N	22	0.99

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
perfluoromethylcyclohexane	148	4.45	QIROQPWSJUXOJC-UHFFFAOYSA-N	21	0.89
perfluoroheptane	149	4.51	LGUZHRODIJCVOC-UHFFFAOYSA-N	24	0.98
cis-perfluorodecalin	150	4.21	UWEYRJFJVCLAGH-XIXRPRMCSA-N	25	0.96
fluorobenzene	151	10.60	PYLWMHQQBFSUBP-UHFFFAOYSA-N	12	1.08
hexafluorobenzene	152	8.67	ZQBFAOFFOQMSGJ-UHFFFAOYSA-N	15	1.10
1,4-dichlorobutane	153	9.51	KJDRSWPQXHESDQ-UHFFFAOYSA-N	16	1.20
chlorobenzene	154	9.79	MVPPADPHJFYWMZ-UHFFFAOYSA-N	14	1.13
dichloromethane	155	15.50	YMWUJEATGCHHMB-UHFFFAOYSA-N	10	1.22
1,1-dichloroethane	156	11.81	SCYULBFZEHDVBN-UHFFFAOYSA-N	12	1.15
1,2-dichloroethane	157	12.59	WSLDOOZREJYCGB-UHFFFAOYSA-N	12	1.20
trans-1,2-dichloroethylene	158	12.85	KFUSEUYYWQURPO-OWOJBTEDSA-N	11	1.15
ortho-dichlorobenzene	159	8.84	RFFLAFLAYFXFSW-UHFFFAOYSA-N	15	1.11
meta-dichlorobenzene	160	8.73	ZPQOPVIELGIULI-UHFFFAOYSA-N	16	1.14
chloroform	161	12.40	HEDRZPFGACZZDS-UHFFFAOYSA-N	11	1.13
1,1,1-trichloroethane	162	9.97	UOCLXMDMGBRAIB-UHFFFAOYSA-N	13	1.09
1,1,2-trichloroethane	163	10.73	UBOXGVDOUJQMTN-UHFFFAOYSA-N	13	1.14

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
trichloroethylene	164	11.12	XSTXAVWGXDQKEL-UHFFFAOYSA-N	13	1.17
1,2,4-trichlorobenzene	165	8.02	PBKONEOXTCPAFI-UHFFFAOYSA-N	17	1.13
carbon tetrachloride	166	10.30	VZGDMQKNWNREIO-UHFFFAOYSA-N	13	1.12
tetrachloroethylene	167	9.74	CYTYCFOTNPOANT-UHFFFAOYSA-N	14	1.13
1,1,2,2-tetrachloroethane	168	9.45	QPFMBZIOSGYJDE-UHFFFAOYSA-N	15	1.15
pentachloroethane	169	8.27	BNIXVQGCZULYKV-UHFFFAOYSA-N	16	1.11
1-bromobutane	170	9.26	MPPPKRYCTPRNTB-UHFFFAOYSA-N	15	1.14
bromobenzene	171	9.48	QARVLSVVCXYDNA-UHFFFAOYSA-N	14	1.11
dibromomethane	172	14.33	FJBFPHVGVWTDIP-UHFFFAOYSA-N	11	1.23
1,2-dibromoethane	173	11.55	PAAZPARNPHGIKF-UHFFFAOYSA-N	13	1.19
bromoform	174	11.39	DIKBFYAXUHHXCS-UHFFFAOYSA-N	13	1.18
n-butyl iodide	175	8.73	KMGBZBJJOKUPIA-UHFFFAOYSA-N	16	1.14
iodobenzene	176	8.94	SNHMUERNLJLMHN-UHFFFAOYSA-N	15	1.12
methylene iodide	177	12.41	NZZFYRREKKOMAT-UHFFFAOYSA-N	13	1.24
n-butylamine	178	10.07	HQABUPZFAYXKJW-UHFFFAOYSA-N	14	1.15
benzylamine	179	9.16	WGQKYBSKWIADBV-UHFFFAOYSA-N	15	1.13

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
ethylenediamine	180	14.79	PIICEJLVQHRZGT-UHFFFAOYSA-N	11	1.25
diethylamine	181	9.60	HPNMFZURTQLUMO-UHFFFAOYSA-N	14	1.12
di-n-butylamine	182	5.86	JQVDAXLFBXTEQA-UHFFFAOYSA-N	22	1.09
pyrrole	183	14.39	KAESVJOAVNADME-UHFFFAOYSA-N	10	1.16
pyrrolidine	184	12.01	RWRDLPDLKQPQOW-UHFFFAOYSA-N	12	1.16
piperidine	185	10.06	NQRYJNQNLNOLGT-UHFFFAOYSA-N	13	1.10
morpholine	186	11.43	YNAVUWVOSKDBBP-UHFFFAOYSA-N	13	1.19
triethylamine	187	7.15	ZMANZCXQSJIPKH-UHFFFAOYSA-N	17	1.05
tri-(n-butyl)amine	188	4.18	IMFACGCPASFAPR-UHFFFAOYSA-N	29	1.05
aniline	189	10.93	PAYRUJLWNCNPSJ-UHFFFAOYSA-N	13	1.16
o-chloroaniline	190	9.47	AKCRQHGQIJBRMN-UHFFFAOYSA-N	15	1.16
methylphenylamine	191	9.17	AFBPFSWMIHJQDM-UHFFFAOYSA-N	15	1.13
N,N-dimethylaniline	192	7.87	JLTDJTHDQAWBAV-UHFFFAOYSA-N	17	1.12
aminoethanol	193	16.58	HZAXFHJVJLSVMW-UHFFFAOYSA-N	10	1.26
diethanolamine	194	10.42	ZBCBWPMODOFKDW-UHFFFAOYSA-N	15	1.22
triethanolamine	195	7.51	GSEJCLTVZPLZKY-UHFFFAOYSA-N	20	1.19

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
pyridine	196	12.36	JUJWROOIHBZHMG-UHFFFAOYSA-N	12	1.18
2-methylpyridine	197	10.09	BSKHPKMHTQYZBB-UHFFFAOYSA-N	14	1.15
3-methylpyridine	198	10.24	ITQTTZVARXURQS-UHFFFAOYSA-N	14	1.16
4-methylpyridine	199	10.20	FKNQCJSGGFJEIZ-UHFFFAOYSA-N	14	1.16
2,4-dimethylpyridine	200	8.66	JYYNAJVZFGKDEQ-UHFFFAOYSA-N	16	1.14
2,6-dimethylpyridine	201	8.57	OISVCGZHLKNMSJ-UHFFFAOYSA-N	16	1.13
2,4,6-trimethylpyridine	202	7.52	BWZVCCNYKMEVEX-UHFFFAOYSA-N	18	1.12
2-bromopyridine	203	10.48	IMRWILPUOVGIMU-UHFFFAOYSA-N	14	1.18
3-bromopyridine	204	10.38	NYPYPOZNGOXYSU-UHFFFAOYSA-N	14	1.17
2-cyanopyridine	205	10.39	FFNVQNRYTPFDDP-UHFFFAOYSA-N	14	1.17
pyrimidine	206	12.69	CZPWVGJYEJSRLH-UHFFFAOYSA-N	11	1.14
quinoline	207	8.44	SMWDFEZZVXVKRB-UHFFFAOYSA-N	16	1.12
acetonitrile	208	18.90	WEVYAHXRMPXWCK-UHFFFAOYSA-N	8	1.20
propionitrile	209	14.10	FVSKHRXBFJPNKK-UHFFFAOYSA-N	10	1.15
n-butyronitrile	210	11.38	KVNRLNFWIYMESJ-UHFFFAOYSA-N	12	1.13
3-methylbutanenitrile	211	9.56	QHDRKFYEGYYIIK-UHFFFAOYSA-N	14	1.12

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
acrylonitrile	212	15.10	NLHHRLWOUZZQLW-UHFFFAOYSA-N	9	1.13
phenylacetonitrile	213	8.65	SUSQOBVLVYHIEX-UHFFFAOYSA-N	16	1.14
benzonitrile	214	9.70	JFDZBHWFFUWGJE-UHFFFAOYSA-N	14	1.13
nitromethane	215	18.52	LYGJENNIWJXYER-UHFFFAOYSA-N	9	1.27
nitroethane	216	13.91	MCSAJNNLRCFZED-UHFFFAOYSA-N	11	1.21
1-nitropropane	217	11.17	JSZOAYXJRCEYSX-UHFFFAOYSA-N	13	1.17
2-nitropropane	218	11.04	FGLBSLMDCBOPQK-UHFFFAOYSA-N	13	1.16
nitrobenzene	219	9.74	LQNUZADURLCDLV-UHFFFAOYSA-N	15	1.17
formamide	220	25.06	ZHNUHDYFZUAESO-UHFFFAOYSA-N	7	1.30
N-methylformamide	221	16.92	ATHHXGZTWNVVOU-UHFFFAOYSA-N	10	1.28
N,N-dimethylformamide	222	12.90	ZMXDDKWLCZADIW-UHFFFAOYSA-N	12	1.21
N,N-dimethylthioformamide	223	11.74	SKECXRFZFFAANN-UHFFFAOYSA-N	13	1.20
N,N-diethylformamide	224	8.98	SUAKHGWARZSWIH-UHFFFAOYSA-N	15	1.12
N-methylacetamide	225	12.99	OHLUUHNLEMFGTQ-UHFFFAOYSA-N	12	1.22
N,N-dimethylacetamide	226	10.75	FXHOOIRPVKKKFG-UHFFFAOYSA-N	13	1.14
N,N-diethylacetamide	227	7.86	AJFDBNQQDYLMJN-UHFFFAOYSA-N	17	1.11

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
2-pyrrolidinone	228	13.01	HNJBEVLQSNELDL-UHFFFAOYSA-N	12	1.22
N-methyl pyrrolidinone	229	10.37	SECXISVLQFMRJM-UHFFFAOYSA-N	14	1.17
1-methyl-2-pyrrolidinethione	230	11.66	OQILOJRSIWGQSM-UHFFFAOYSA-N	15	1.30
tetramethylurea	231	8.31	AVQQQNCBBIEMEU-UHFFFAOYSA-N	17	1.15
tetraethylurea	232	5.26	UWHSPZZUAYSGTB-UHFFFAOYSA-N	24	1.08
dimethylcyanamide	233	12.38	OAGOUCJGXNLJNL-UHFFFAOYSA-N	12	1.19
carbon disulfide	234	16.50	QGJOPFRUJISHPQ-UHFFFAOYSA-N	9	1.19
dimethyl sulfide	235	13.55	QMMFVYPAHWMCMS-UHFFFAOYSA-N	10	1.12
diethyl sulfide	236	9.22	LJSQFQKUNVCTIA-UHFFFAOYSA-N	15	1.14
diisopropyl sulfide	237	6.88	XYWDPYKBIRQXQS-UHFFFAOYSA-N	18	1.07
dibutyl sulfide	238	5.73	HTIRHQRTDBPHNZ-UHFFFAOYSA-N	23	1.11
tetrahydrothiophene	239	11.27	RAOIDOHSFRTOEL-UHFFFAOYSA-N	13	1.18
thiane	240	9.63	YPWFISCTZQNZAU-UHFFFAOYSA-N	14	1.12
dimethylsulfoxide	241	14.03	IAZDPXIOMUYVGZ-UHFFFAOYSA-N	11	1.21
dibutyl sulfoxide	242	5.13	LOWMYOWHQMKBTM-UHFFFAOYSA-N	23	1.03
sulfolane	243	10.50	HXJUTPCZVOIRIF-UHFFFAOYSA-N	14	1.18

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
thiobis(2-ethanol)	244	9.67	YODZTKMDCQEPHD-UHFFFAOYSA-N	16	1.21
diethyl sulfite	245	7.84	NVJBFARDFTXOTO-UHFFFAOYSA-N	18	1.15
dimethyl sulfate	246	10.57	VAYGXNSJCAHWJZ-UHFFFAOYSA-N	14	1.18
diethyl sulfate	247	7.64	DENRZWYUOJLTMF-UHFFFAOYSA-N	18	1.13
methanesulfonic acid	248	15.37	AFVFQIVMOAPDHO-UHFFFAOYSA-N	11	1.28
trimethylphosphate	249	8.67	WVLBCYQITXONBZ-UHFFFAOYSA-N	17	1.18
triethylphosphate	250	5.87	DQWPFSLDHJDLRL-UHFFFAOYSA-N	23	1.12
tri-n-butylphosphate	251	3.65	STCOOQWBFONSKY-UHFFFAOYSA-N	36	1.10
hexamethylphosphoric triamide	252	5.69	GNOIPBMMFNIUFM-UHFFFAOYSA-N	22	1.07
hydrogen peroxide	253	42.37	MHAJPDPJQMAIIY-UHFFFAOYSA-N	5	1.44
hydrogen fluoride	254	47.62	KRHYYFGTRYWZRS-UHFFFAOYSA-N	3	1.16
sulfuric acid	255	18.69	QAOWNCQODCNURD-UHFFFAOYSA-N	10	1.35
ammonia	256	40.00	QGZKDVFQNNGYKY-UHFFFAOYSA-N	5	1.40
hydrazine	257	31.35	OAKJQQAXSVQMHS-UHFFFAOYSA-N	7	1.47
sulfur dioxide	258	22.83	RAHZWNYVWXNFOC-UHFFFAOYSA-N	7	1.24
thionyl chloride	259	13.71	FYSNRJHAOHDILO-UHFFFAOYSA-N	11	1.20

Continued on next page

Solvent	ID	[liquid]/ M	InChIKey	N_{SSIP}	$\Delta G_{c,i}$/ kJ mol^{-1}
phosphorus oxychloride	260	10.87	XHXFXVLFKHQFAL-UHFFFAOYSA-N	13	1.15
dihydrolevoglucosenone	261	9.76	WHIRALQRTSITMI-UJURSFKZSA-N	15	1.18

Table G.1 Solvent information

G.2 Polynomial Coefficient information

Polynomial coefficients for the solvation profiles ($\Delta G_{S,i}(\varepsilon_i)$) are contained in the following sections. Table G.2 contains the information for the pure solvents. The coefficients for water ethanol mixtures are in table G.3, and the coefficients for chloroform tetrahydrofuran mixtures in table G.4.

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
(2R,3S)-2,3-butanediol	positive	-0.890201	-6.497101e-01	4.300415e-01	-9.204610e-01	3.380664e-01	-6.178162e-02	6.246097e-03	-3.339554e-04	7.384658e-06
(2R,3S)-2,3-butanediol	negative	-0.762634	8.643129e-01	6.406149e-01	2.417495e-01	3.175360e-02	2.110497e-03	7.333515e-05	1.177609e-06	5.280983e-09
1,1,1,3,3,3-hexafluoro-2-propanol	positive	-1.101783	-1.525279e-01	-1.554708e-01	8.129254e-02	-3.145803e-02	3.993038e-03	-1.263393e-04	-9.480639e-06	5.583470e-07
1,1,1,3,3,3-hexafluoro-2-propanol	negative	-1.222259	-8.236826e-02	5.638164e-02	3.267681e-01	7.436811e-02	7.862788e-03	4.452899e-04	1.305879e-05	1.559838e-07
1,1,1-trichloroethane	positive	-1.194875	-2.141508e-01	-1.893664e-02	-1.413340e-03	1.591105e-04	-1.250561e-05	1.288864e-06	-7.896803e-08	1.812658e-09
1,1,1-trichloroethane	negative	-1.193311	2.573454e-01	-2.620951e-02	6.455442e-03	6.124237e-04	5.073296e-06	-1.687381e-06	-8.404279e-08	-1.258913e-09
1,1,2,2-tetrachloroethane	positive	-1.222001	-1.385446e-01	-1.024372e-02	-3.335449e-05	-1.933680e-04	4.830368e-05	-5.913206e-06	3.740997e-07	-9.566664e-09
1,1,2,2-tetrachloroethane	negative	-1.217480	4.082954e-01	-3.192065e-02	2.716476e-02	4.942599e-03	4.149763e-04	1.917130e-05	4.710287e-07	4.822575e-09
1,1,2-trichloroethane	positive	-1.111634	-3.413493e-01	-7.097159e-02	-1.203293e-03	-3.355404e-03	9.510310e-04	-1.046873e-04	5.445297e-06	-1.119244e-07
1,1,2-trichloroethane	negative	-1.095659	3.824018e-01	5.881694e-02	7.749885e-02	1.356770e-02	1.179391e-03	5.695486e-05	1.461364e-06	1.557468e-08
1,1-dichloroethane	positive	-1.167188	-3.285231e-01	-5.563326e-02	-3.441650e-03	-6.645260e-04	2.980055e-04	-3.419368e-05	1.739662e-06	-3.421443e-08
1,1-dichloroethane	negative	-1.163313	2.715932e-01	-2.149865e-02	1.065449e-02	8.795448e-04	-4.453206e-06	-3.494978e-06	-1.608593e-07	-2.368143e-09
1,2,3,4-tetrahydronaphthalene	positive	-1.200397	-2.697206e-01	-9.249236e-02	1.252853e-04	-1.013312e-02	3.146856e-03	-3.950635e-04	2.347683e-05	-5.486562e-07
1,2,3,4-tetrahydronaphthalene	negative	-1.199543	1.872945e-01	-9.461591e-03	3.999001e-04	4.829767e-05	2.506149e-06	9.777836e-08	2.642631e-09	3.281642e-11
1,2,4-trichlorobenzene	positive	-1.226141	-1.687392e-01	-1.754670e-02	-2.033351e-03	1.858162e-04	-3.911042e-05	4.861510e-06	-2.594745e-07	5.111581e-09
1,2,4-trichlorobenzene	negative	-1.222474	2.566924e-01	-1.976128e-02	8.532983e-03	3.279971e-04	-6.703982e-05	-7.121413e-06	-2.667506e-07	-3.610351e-09

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
1,2-butanediol	positive	-0.899801	-7.116919e-01	5.056025e-01	-9.521099e-01	3.398289e-01	-6.098363e-02	6.079910e-03	-3.213235e-04	7.034611e-06
1,2-butanediol	negative	-0.737606	1.009706e+00	8.319744e-01	3.085418e-01	4.176239e-02	2.921697e-03	1.103190e-04	2.070689e-06	1.417222e-08
1,2-dibromoethane	positive	-1.221197	-2.050333e-01	-1.910846e-02	-1.801006e-03	3.004630e-04	-3.958559e-05	4.434252e-06	-2.742616e-07	6.739283e-09
1,2-dibromoethane	negative	-1.217877	3.298418e-01	-2.910078e-02	1.621918e-02	2.703716e-03	2.024788e-04	8.241187e-06	1.762773e-07	1.548821e-09
1,2-dichloroethane	positive	-1.086780	-4.563994e-01	-1.047747e-01	-1.375322e-02	-2.047900e-03	1.383131e-03	-2.031858e-04	1.296764e-05	-3.165268e-07
1,2-dichloroethane	negative	-1.079245	3.141796e-01	-4.850773e-03	3.084488e-02	4.985292e-03	3.838321e-04	1.624788e-05	3.633719e-07	3.356799e-09
1,2-dimethoxybenzene	positive	-1.142760	-5.147572e-01	2.934284e-01	-4.078396e-01	9.635362e-02	-9.747940e-03	3.303550e-04	1.171936e-05	-8.119088e-07
1,2-dimethoxybenzene	negative	-1.161421	2.361978e-01	-2.048120e-02	1.183917e-03	1.201629e-04	-2.577773e-07	-3.821803e-07	-1.632188e-08	-2.229165e-10
1,2-dimethyoxyethane	positive	-1.070258	-5.280988e-01	-3.392391e-01	-6.221322e-01	2.920257e-01	-6.015154e-02	6.602993e-03	-3.762205e-04	8.761635e-06
1,2-dimethyoxyethane	negative	-1.055962	2.075163e-01	-1.046939e-02	4.321737e-04	5.204461e-05	2.590511e-06	9.421122e-08	2.429991e-09	2.975495e-11
1,2-propanediol	positive	-0.785969	-6.330375e-01	2.614069e-01	-9.509676e-01	3.714685e-01	-7.072029e-02	7.390093e-03	-4.063470e-04	9.205268e-06
1,2-propanediol	negative	-0.641951	8.842620e-01	6.073547e-01	2.115452e-01	2.403233e-02	1.214316e-03	1.921975e-05	-4.821662e-07	-1.519363e-08
1,3,5-trimethylbenzene	positive	-1.194737	-2.787201e-01	-8.979033e-02	-4.647433e-03	-7.724531e-03	2.672832e-03	-3.479194e-04	2.109508e-05	-4.998647e-07
1,3,5-trimethylbenzene	negative	-1.194159	1.809265e-01	-1.014978e-02	4.053059e-04	4.572453e-05	1.775745e-06	4.781445e-08	1.158761e-09	1.587023e-11
1,3-dioxane	positive	-1.037390	-4.829166e-01	1.628682e-02	-9.782526e-01	4.189199e-01	-8.396431e-02	9.099593e-03	-5.146063e-04	1.192481e-05
1,3-dioxane	negative	-1.026221	2.619076e-01	-1.830239e-02	7.627057e-04	1.164965e-04	4.569398e-06	2.981314e-08	-2.412470e-09	-4.780393e-11
1,3-dioxolan	positive	-1.131308	-6.166724e-01	6.839796e-01	-1.307242e+00	4.962761e-01	-9.375130e-02	9.760541e-03	-5.353703e-04	1.210468e-05
1,3-dioxolan	negative	-1.143233	2.281282e-01	-2.052025e-02	1.115286e-03	4.933203e-05	-8.158327e-06	-7.852874e-07	-2.659573e-08	-3.291020e-10
1,3-propanediol	positive	-0.897807	-5.019565e-01	-9.413541e-01	-4.080824e-01	2.560258e-01	-5.771015e-02	6.660158e-03	-3.925000e-04	9.371531e-06
1,3-propanediol	negative	-0.742587	5.175538e-01	2.825034e-01	3.262181e-02	-1.390408e-02	-2.816542e-03	-2.115544e-04	-7.304052e-06	-9.705172e-08
1,4-butanediol	positive	-0.782036	-5.870980e-01	7.026171e-01	-1.429155e+00	5.540235e-01	-1.063297e-01	1.121334e-02	-6.216786e-04	1.418391e-05
1,4-butanediol	negative	-0.695868	6.776853e-01	2.561607e-01	6.342384e-02	-2.617091e-03	-1.318562e-03	-1.141832e-04	-4.172389e-06	-5.710865e-08

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
1,4-dichlorobutane	positive	-1.161084	-2.863278e-01	-6.176222e-02	-3.319870e-03	-1.583327e-03	5.883469e-04	-7.028148e-05	3.836723e-06	-8.191659e-08
1,4-dichlorobutane	negative	-1.159320	3.526121e-01	-3.838388e-02	8.431533e-03	1.582524e-03	1.227001e-04	5.126330e-06	1.129858e-07	1.032108e-09
1,5-pentanediol	positive	-0.957860	-4.935713e-01	1.383674e-01	-1.166561e+00	4.955148e-01	-9.940832e-02	1.079746e-02	-6.120406e-04	1.421265e-05
1,5-pentanediol	negative	-0.852034	4.510116e-01	1.697979e-01	1.674567e-04	-1.760865e-02	-3.005873e-03	-2.143167e-04	-7.207140e-06	-9.420351e-08
1,8-cineole	positive	-1.152710	-6.537940e-01	1.313699e+00	-2.158599e+00	8.377638e-01	-1.626982e-01	1.737137e-02	-9.739173e-04	2.243828e-05
1,8-cineole	negative	-1.158970	1.648618e-01	-6.894482e-03	3.427407e-04	4.258169e-05	2.795689e-06	1.325842e-07	3.751459e-09	4.505801e-11
1-bromobutane	positive	-1.216247	-1.673573e-01	-7.344112e-02	1.145686e-02	-1.015224e-02	2.229956e-03	-2.199310e-04	1.048303e-05	-1.965498e-07
1-bromobutane	negative	-1.214722	2.720319e-01	-2.231445e-02	1.212152e-03	1.032782e-04	-1.510887e-06	-4.112972e-07	-1.631639e-08	-2.163230e-10
1-butanol	positive	-1.015001	-9.296050e-01	1.932317e+00	-2.250632e+00	7.992896e-01	-1.466044e-01	1.497994e-02	-8.101510e-04	1.810671e-05
1-butanol	negative	-1.022195	3.070429e-01	-5.132932e-02	-7.027418e-02	-2.697715e-02	-3.634727e-03	-2.355745e-04	-7.505902e-06	-9.465560e-08
1-decanol	positive	-1.101115	-1.220303e+00	2.369069e+00	-2.166891e+00	6.821188e-01	-1.129973e-01	1.051422e-02	-5.206054e-04	1.069688e-05
1-decanol	negative	-1.134442	2.371375e-01	-4.170727e-02	-8.540030e-02	-3.260362e-02	-4.392719e-03	-2.853697e-04	-9.119467e-06	-1.153477e-07
1-dodecanol	positive	-1.093167	-1.349119e+00	2.754538e+00	-2.396388e+00	7.418081e-01	-1.212520e-01	1.113358e-02	-5.436205e-04	1.100387e-05
1-dodecanol	negative	-1.171724	1.332946e-01	-1.561124e-01	-1.224645e-01	-3.767910e-02	-4.745450e-03	-2.979316e-04	-9.317888e-06	-1.160684e-07
1-hexanol	positive	-1.062455	-1.055548e+00	2.212318e+00	-2.345769e+00	8.064904e-01	-1.445221e-01	1.448487e-02	-7.703410e-04	1.696256e-05
1-hexanol	negative	-1.086920	2.336173e-01	-7.300032e-02	-8.427325e-02	-3.018854e-02	-3.990629e-03	-2.564848e-04	-8.134834e-06	-1.022955e-07
1-methyl-2-pyrrolidinethione	positive	-0.962651	-7.152360e-01	9.415577e-01	-1.654688e+00	6.276881e-01	-1.193967e-01	1.253090e-02	-6.925746e-04	1.576585e-05
1-methyl-2-pyrrolidinethione	negative	-0.974943	3.473443e-01	-3.442563e-02	4.498374e-03	7.472084e-04	4.788482e-05	1.561381e-06	2.404800e-08	1.149134e-10
1-nitropropane	positive	-1.091515	-3.819063e-01	9.501791e-02	-1.549414e-01	-7.431771e-03	1.088151e-02	-1.874732e-03	1.340822e-04	-3.586933e-06
1-nitropropane	negative	-1.103096	3.528543e-01	-3.872601e-02	8.872934e-03	1.619227e-03	1.244000e-04	5.174671e-06	1.137245e-07	1.036250e-09
1-octanol	positive	-1.070858	-1.202864e+00	2.523880e+00	-2.441728e+00	8.073359e-01	-1.402995e-01	1.368568e-02	-7.100381e-04	1.528042e-05
1-octanol	negative	-1.120822	1.939369e-01	-1.128419e-01	-1.019346e-01	-3.353193e-02	-4.318569e-03	-2.741080e-04	-8.628911e-06	-1.079593e-07
1-pentanol	positive	-1.051993	-9.690513e-01	2.011067e+00	-2.262460e+00	7.935028e-01	-1.442408e-01	1.462978e-02	-7.861897e-04	1.747314e-05
1-pentanol	negative	-1.060662	2.750798e-01	-5.170626e-02	-7.783727e-02	-2.936065e-02	-3.938718e-03	-2.549977e-04	-8.124673e-06	-1.024989e-07
1-propanol	positive	-0.980816	-8.096442e-01	1.626934e+00	-2.112101e+00	7.721592e-01	-1.442968e-01	1.496020e-02	-8.189513e-04	1.849502e-05
1-propanol	negative	-0.973477	3.415066e-01	-3.929623e-02	-6.205056e-02	-2.497812e-02	-3.406014e-03	-2.218706e-04	-7.088196e-06	-8.953213e-08

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
2,2,2-trifluoroethanol	positive	-0.997364	-3.793563e-01	2.041970e-02	-8.300620e-02	-2.623974e-02	1.323710e-02	-2.019365e-03	1.374169e-04	-3.573987e-06
2,2,2-trifluoroethanol	negative	-0.908180	8.557748e-01	7.544323e-01	3.942971e-01	6.867456e-02	6.247265e-03	3.177204e-04	8.566070e-06	9.552493e-08
2,2,4,4-tetramethyl-3-pentanone	positive	-1.094507	-9.245292e-01	1.934993e+00	-2.293014e+00	8.208154e-01	-1.514826e-01	1.555865e-02	-8.452087e-04	1.896393e-05
2,2,4,4-tetramethyl-3-pentanone	negative	-1.138221	2.179984e-01	-1.352736e-02	5.388750e-04	5.064939e-05	7.479288e-07	-4.731274e-08	-2.000519e-09	-2.252734e-11
2,2,4-trimethylpentane	positive	-1.238415	-1.765605e-02	-1.167380e-03	-5.701911e-05	1.354018e-04	-3.916296e-05	5.085638e-06	-3.197527e-07	7.921273e-09
2,2,4-trimethylpentane	negative	-1.237311	1.615061e-01	-5.477737e-03	3.219743e-04	4.808390e-05	3.647622e-06	1.785892e-07	4.941948e-09	5.756941e-11
2,3-butanedione	positive	-0.957952	-6.097101e-01	1.828170e-01	-5.656391e-01	2.001658e-01	-3.489969e-02	3.372913e-03	-1.730490e-04	3.686304e-06
2,3-butanedione	negative	-0.966964	3.172693e-01	-3.272390e-02	7.716039e-03	9.919664e-04	4.837025e-05	8.081493e-07	-1.157786e-08	-4.128978e-10
2,4,5-trimethylacetophenone	positive	-1.144352	-1.141974e+00	2.415895e+00	-2.728511e+00	9.724221e-01	-1.797152e-01	1.850920e-02	-1.008435e-03	2.268839e-05
2,4,5-trimethylacetophenone	negative	-1.198872	2.401702e-01	-2.125604e-02	1.303201e-03	1.187240e-04	-1.828797e-06	-5.071655e-07	-2.044638e-08	-2.743000e-10
2,4,6-trimethylpyridine	positive	-1.170222	-1.030532e+00	3.059166e+00	-3.924782e+00	1.502709e+00	-2.933149e-01	3.157664e-02	-1.785037e-03	4.143657e-05
2,4,6-trimethylpyridine	negative	-1.186389	1.888496e-01	-1.533604e-02	5.931543e-04	-1.520721e-05	-8.469822e-06	-5.597768e-07	-1.561118e-08	-1.664109e-10
2,4-dimethylphenol	positive	-1.106467	-4.111088e-01	-1.078808e-01	-1.515767e-02	-2.395171e-02	8.673753e-03	-1.195550e-03	7.657203e-05	-1.906185e-06
2,4-dimethylphenol	negative	-0.916312	1.288790e+00	1.319604e+00	5.877605e-01	9.906402e-02	8.875450e-03	4.468373e-04	1.195010e-05	1.323225e-07
2,4-dimethylpyridine	positive	-1.161837	-1.027337e+00	3.223180e+00	-4.105006e+00	1.573199e+00	-3.075244e-01	3.315225e-02	-1.876385e-03	4.360267e-05
2,4-dimethylpyridine	negative	-1.176914	2.217588e-01	-1.940308e-02	1.010516e-03	4.354258e-05	-7.793260e-06	-7.341941e-07	-2.457842e-08	-3.013660e-10
2,4-pentanedione	positive	-0.896922	-6.117613e-01	3.558251e-01	-9.368303e-01	3.566248e-01	-6.676166e-02	6.880517e-03	-3.738729e-04	8.383334e-06
2,4-pentanedione	negative	-0.905053	3.489523e-01	-3.388729e-02	4.401611e-03	7.309368e-04	4.604990e-05	1.457296e-06	2.113850e-08	8.246211e-11
2,6-dimethylpyridine	positive	-1.152276	-1.127381e+00	3.057466e+00	-3.683604e+00	1.379267e+00	-2.651796e-01	2.821752e-02	-1.580227e-03	3.639759e-05
2,6-dimethylpyridine	negative	-1.188508	1.981138e-01	-1.704868e-02	6.312015e-04	-4.356394e-05	-1.263049e-05	-8.008690e-07	-2.226273e-08	-2.390999e-10

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
2-bromopyridine	positive	-1.006249	-1.603988e+00	2.920445e+00	-2.613682e+00	8.242924e-01	-1.372013e-01	1.283572e-02	-6.390849e-04	1.320402e-05
2-bromopyridine	negative	-1.109329	2.531242e-01	-1.684152e-02	1.181144e-02	1.091236e-03	1.143585e-05	-2.917324e-06	-1.512080e-07	-2.317996e-09
2-butanol	positive	-1.023069	-9.092226e-01	1.486403e+00	-1.573215e+00	5.107924e-01	-8.633447e-02	8.181674e-03	-4.126577e-04	8.642896e-06
2-butanol	negative	-0.968545	6.321707e-01	3.254758e-01	8.454003e-02	5.500181e-04	-1.059838e-03	-1.023800e-04	-3.889664e-06	-5.434072e-08
2-butanone	positive	-1.040049	-5.384300e-01	1.162723e+00	-2.113137e+00	8.352447e-01	-1.639063e-01	1.763089e-02	-9.941886e-04	2.301270e-05
2-butanone	negative	-1.037670	3.011714e-01	-2.557917e-02	1.750138e-03	2.555240e-04	1.089778e-05	8.814789e-08	-6.177450e-09	-1.343003e-10
2-chloroethanol	positive	-0.979144	-5.667617e-01	-4.638912e-02	-1.322207e-01	2.185475e-02	2.398092e-04	-3.616235e-04	3.362895e-05	-9.987784e-07
2-chloroethanol	negative	-0.823545	1.170181e+00	1.036574e+00	4.702944e-01	7.814181e-02	6.879647e-03	3.404570e-04	8.959395e-06	9.773569e-08
2-cyanoethanol	positive	-0.953428	-4.409367e-01	1.197539e-01	-8.953514e-02	-8.180090e-02	3.222595e-02	-4.687957e-03	3.139713e-04	-8.118260e-06
2-cyanoethanol	negative	-0.755274	1.393956e+00	1.085067e+00	4.346615e-01	6.480709e-02	5.113928e-03	2.259525e-04	5.280969e-06	5.082859e-08
2-cyanopyridine	positive	-0.974113	-9.317893e-01	1.490175e+00	-1.525766e+00	4.813834e-01	-7.903134e-02	7.272935e-03	-3.561084e-04	7.237892e-06
2-cyanopyridine	negative	-1.022165	3.306715e-01	-8.293573e-03	2.558905e-02	3.734890e-03	2.524957e-04	9.021430e-06	1.589794e-07	1.004407e-09
2-ethoxyethanol	positive	-0.993060	-5.699371e-01	2.857063e-02	-8.181783e-01	3.403087e-01	-6.676237e-02	7.114118e-03	-3.968539e-04	9.093007e-06
2-ethoxyethanol	negative	-0.915065	4.051147e-01	2.492669e-02	-6.766834e-02	-3.131917e-02	-4.424432e-03	-2.940026e-04	-9.527288e-06	-1.216912e-07
2-heptanone	positive	-1.129077	-8.638787e-01	1.919537e+00	-2.436343e+00	8.969749e-01	-1.690290e-01	1.765968e-02	-9.732873e-04	2.211031e-05
2-heptanone	negative	-1.162559	2.204575e-01	-1.845211e-02	4.368498e-04	-1.325111e-04	-2.189813e-05	-1.233217e-06	-3.197388e-08	-3.242086e-10
2-methoxyethanol	positive	-0.935775	-5.378416e-01	-4.268989e-01	-5.611551e-01	2.729046e-01	-5.687227e-02	6.282805e-03	-3.594853e-04	8.397659e-06
2-methoxyethanol	negative	-0.853460	3.714764e-01	-1.860996e-02	-8.425663e-02	-3.425119e-02	-4.699678e-03	-3.083040e-04	-9.917020e-06	-1.260482e-07
2-methoxyphenol	positive	-1.042328	-6.155412e-01	3.088071e-01	-4.042506e-01	8.660035e-02	-6.684404e-03	-8.793516e-05	3.898859e-05	-1.507417e-06
2-methoxyphenol	negative	-1.017887	4.304984e-01	1.233483e-01	4.767953e-02	8.004371e-04	-5.235895e-04	-5.198240e-05	-1.974307e-06	-2.742566e-08
2-methyl-1-propanol	positive	-1.030840	-8.999038e-01	1.496157e+00	-1.621734e+00	5.352580e-01	-9.186713e-02	8.835886e-03	-4.521307e-04	9.604039e-06
2-methyl-1-propanol	negative	-0.952837	7.001680e-01	3.992854e-01	1.012856e-01	1.548118e-03	-1.115271e-03	-1.118432e-04	-4.306712e-06	-6.062771e-08
2-methyl-2-butanol	positive	-0.999257	-9.691971e-01	1.612612e+00	-1.594563e+00	4.987961e-01	-8.151030e-02	7.471339e-03	-3.642945e-04	7.369757e-06
2-methyl-2-butanol	negative	-0.979631	5.254512e-01	2.094362e-01	4.689134e-02	-4.605178e-03	-1.426048e-03	-1.162072e-04	-4.141665e-06	-5.589561e-08
2-methyl-2-propanol	positive	-0.981841	-9.770600e-01	1.745432e+00	-1.998963e+00	6.938246e-01	-1.246458e-01	1.250373e-02	-6.652757e-04	1.465433e-05
2-methyl-2-propanol	negative	-0.993531	3.227294e-01	-2.647775e-02	-5.014892e-02	-2.168523e-02	-3.004415e-03	-1.968514e-04	-6.304904e-06	-7.973295e-08
2-methylbutane	positive	-1.240942	-1.476979e-02	-1.023890e-03	-4.395708e-05	1.184994e-04	-3.493671e-05	4.593347e-06	-2.915146e-07	7.275377e-09
2-methylbutane	negative	-1.239829	1.454402e-01	-4.637962e-03	2.991283e-04	4.213525e-05	3.221511e-06	1.588783e-07	4.383144e-09	5.062381e-11

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
2-methylpyridine	positive	-1.154410	-1.018908e+00	2.958227e+00	-3.733247e+00	1.418552e+00	-2.753512e-01	2.951252e-02	-1.662364e-03	3.847356e-05
2-methylpyridine	negative	-1.177446	2.044468e-01	-2.113132e-02	1.007337e-03	-8.010979e-05	-2.516455e-05	-1.764325e-06	-5.402035e-08	-6.355713e-10
2-methyltetrahydrofuran	positive	-1.139577	-3.284804e-01	3.176746e-01	-1.581029e+00	6.764351e-01	-1.374843e-01	1.511430e-02	-8.656369e-04	2.027793e-05
2-methyltetrahydrofuran	negative	-1.112227	2.150048e-01	-9.436872e-03	3.822902e-04	5.941447e-05	3.799807e-06	1.610402e-07	4.175446e-09	4.801350e-11
2-nitropropane	positive	-1.066415	-5.362569e-01	3.728754e-01	-4.996416e-01	1.319371e-01	-1.661632e-02	1.051533e-03	-2.782027e-05	7.701682e-08
2-nitropropane	negative	-1.086239	3.183391e-01	-3.329218e-02	6.260959e-03	1.015172e-03	6.631937e-05	2.222012e-06	3.562746e-08	1.870850e-10
2-pentanone	positive	-1.079194	-6.559118e-01	1.405161e+00	-2.181007e+00	8.375815e-01	-1.617021e-01	1.719231e-02	-9.607137e-04	2.207486e-05
2-pentanone	negative	-1.090311	2.639266e-01	-2.179769e-02	1.119745e-03	1.733771e-05	-1.269512e-05	-1.044544e-06	-3.367506e-08	-4.059868e-10
2-phenylethanol	positive	-1.026310	-1.072865e+00	1.782354e+00	-1.699389e+00	5.253396e-01	-8.483241e-02	7.674972e-03	-3.689067e-04	7.347272e-06
2-phenylethanol	negative	-1.006910	5.192136e-01	1.697375e-01	1.293727e-02	-1.384265e-02	-2.542134e-03	-1.853024e-04	-6.295323e-06	-8.275295e-08
2-propanol	positive	-0.967863	-8.541049e-01	1.703331e+00	-2.106460e+00	7.578214e-01	-1.400667e-01	1.439497e-02	-7.822464e-04	1.755513e-05
2-propanol	negative	-0.972442	3.468848e-01	-2.961777e-02	-4.670123e-02	-2.029810e-02	-2.814283e-03	-1.843003e-04	-5.897902e-06	-7.451805e-08
2-pyrrolidinone	positive	-0.987363	-5.143472e-01	2.139026e+00	-3.544878e+00	1.428933e+00	-2.867431e-01	3.145891e-02	-1.803962e-03	4.235388e-05
2-pyrrolidinone	negative	-0.996889	8.363129e-02	-2.345523e-01	-7.850689e-02	-1.461857e-02	-1.280509e-03	-5.735287e-05	-1.265126e-06	-1.067032e-08
3,3-dimethyl-2-butanone	positive	-1.063622	-7.345745e-01	1.606922e+00	-2.290990e+00	8.660261e-01	-1.657657e-01	1.751881e-02	-9.744166e-04	2.230517e-05
3,3-dimethyl-2-butanone	negative	-1.082733	2.550705e-01	-2.167314e-02	1.213231e-03	7.655464e-05	-5.821501e-06	-6.865162e-07	-2.451752e-08	-3.119236e-10
3-bromopyridine	positive	-1.064741	-1.490789e+00	2.900577e+00	-2.518854e+00	7.828293e-01	-1.285830e-01	1.186896e-02	-5.827464e-04	1.186471e-05
3-bromopyridine	negative	-1.170928	2.534228e-01	-2.481352e-02	6.874995e-03	5.000537e-04	-1.662458e-05	-3.199340e-06	-1.328801e-07	-1.872091e-09
3-chlorophenol	positive	-1.111107	-2.777550e-01	-1.256709e-01	3.562941e-02	-2.562869e-02	5.946439e-03	-6.501771e-04	3.511530e-05	-7.592558e-07
3-chlorophenol	negative	-1.135119	4.814550e-01	6.914161e-01	5.233399e-01	1.043136e-01	1.042904e-02	5.713908e-04	1.637991e-05	1.923633e-07
3-heptanone	positive	-1.139016	-8.212552e-01	1.800408e+00	-2.368955e+00	8.791609e-01	-1.664329e-01	1.744641e-02	-9.640842e-04	2.194955e-05
3-heptanone	negative	-1.167710	2.163200e-01	-1.588902e-02	5.543139e-04	-3.696889e-05	-9.384812e-06	-5.282115e-07	-1.287893e-08	-1.191955e-10
3-methyl-1-butanol	positive	-1.053234	-9.604103e-01	1.998770e+00	-2.272030e+00	8.001807e-01	-1.459291e-01	1.484210e-02	-7.995343e-04	1.780785e-05
3-methyl-1-butanol	negative	-1.066424	2.421138e-01	-6.771375e-02	-8.019779e-02	-2.909788e-02	-3.859426e-03	-2.483771e-04	-7.882407e-06	-9.915039e-08

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
3-methyl-2-butanone	positive	-1.078267	-6.838775e-01	1.439279e+00	-2.166557e+00	8.254694e-01	-1.585720e-01	1.679657e-02	-9.357929e-04	2.144906e-05
3-methyl-2-butanone	negative	-1.092927	2.720050e-01	-2.156914e-02	1.107776e-03	7.205361e-05	-4.729518e-06	-5.708516e-07	-2.023274e-08	-2.548958e-10
3-methylbutanenitrile	positive	-1.034590	-1.200513e+00	2.163838e+00	-1.871514e+00	5.471009e-01	-8.344364e-02	7.079470e-03	-3.157204e-04	5.750242e-06
3-methylbutanenitrile	negative	-1.122696	3.321319e-01	-3.253393e-02	4.140397e-03	5.573427e-04	2.645320e-05	4.046310e-07	-7.526504e-09	-2.359071e-10
3-methylpyridine	positive	-1.139442	-1.063637e+00	3.136253e+00	-3.987303e+00	1.524210e+00	-2.973024e-01	3.199253e-02	-1.808024e-03	4.196101e-05
3-methylpyridine	negative	-1.157376	2.254920e-01	-2.254687e-02	1.620877e-03	8.925787e-05	-9.833865e-06	-1.083773e-06	-3.888234e-08	-5.008290e-10
3-pentanone	positive	-1.099605	-6.015719e-01	1.189729e+00	-2.031716e+00	7.909466e-01	-1.536795e-01	1.640665e-02	-9.195861e-04	2.118029e-05
3-pentanone	negative	-1.104651	2.422862e-01	-1.908788e-02	7.936843e-04	-2.697414e-05	-1.371452e-05	-9.593013e-07	-2.847044e-08	-3.231451e-10
4-methyl-1,3-dioxolan-2-one	positive	-0.796122	-1.370842e+00	2.473309e+00	-2.134372e+00	6.174558e-01	-9.332031e-02	7.855071e-03	-3.477306e-04	6.285176e-06
4-methyl-1,3-dioxolan-2-one	negative	-0.899483	5.060501e-01	-8.406628e-02	7.073692e-03	2.410429e-03	2.493096e-04	1.322871e-05	3.630058e-07	4.085229e-09
4-methyl-2-pentanone	positive	-1.065203	-9.031389e-01	1.887158e+00	-2.293542e+00	8.280320e-01	-1.537335e-01	1.586642e-02	-8.654581e-04	1.948694e-05
4-methyl-2-pentanone	negative	-1.105207	2.795487e-01	-2.251930e-02	1.209285e-03	6.466472e-05	-7.792340e-06	-8.078170e-07	-2.800327e-08	-3.514340e-10
4-methylpyridine	positive	-1.113174	-1.164421e+00	3.292704e+00	-4.127523e+00	1.574576e+00	-3.069144e-01	3.301670e-02	-1.865600e-03	4.329297e-05
4-methylpyridine	negative	-1.133833	2.341870e-01	-2.354486e-02	2.053388e-03	1.772430e-04	-3.665480e-06	-8.764673e-07	-3.580697e-08	-4.897893e-10
N,N-diethylacetamide	positive	-1.162744	3.876024e-02	9.749989e-01	-2.893522e+00	1.244727e+00	-2.573894e-01	2.877658e-02	-1.672567e-03	3.967642e-05
N,N-diethylacetamide	negative	-1.082366	2.467775e-01	-2.242817e-02	1.395823e-03	5.017602e-05	-1.267346e-05	-1.184135e-06	-4.050791e-08	-5.089947e-10
N,N-diethylformamide	positive	-1.110526	-2.107473e-01	1.320457e+00	-2.920134e+00	1.214672e+00	-2.470204e-01	2.731899e-02	-1.575340e-03	3.714222e-05
N,N-diethylformamide	negative	-1.058265	2.816857e-01	-2.513720e-02	1.963419e-03	2.218124e-04	3.534726e-06	-4.228862e-07	-2.229887e-08	-3.314444e-10
N,N-dimethylacetamide	positive	-1.053424	-3.293968e-01	1.463040e+00	-2.974736e+00	1.218761e+00	-2.458390e-01	2.704621e-02	-1.553900e-03	3.653727e-05
N,N-dimethylacetamide	negative	-1.010376	3.189667e-01	-2.792788e-02	2.343837e-03	3.142104e-04	1.257207e-05	2.980706e-08	-1.068045e-08	-2.093312e-10
N,N-dimethylaniline	positive	-1.126464	-5.269993e-01	-6.411607e-02	-1.528968e-01	4.461520e-02	-5.876325e-03	4.004190e-04	-1.291959e-05	1.310248e-07
N,N-dimethylaniline	negative	-1.130514	1.871544e-01	-1.771213e-02	4.764554e-04	-1.126826e-04	-1.851489e-05	-1.015298e-06	-2.558882e-08	-2.518874e-10

Continued on next page

Molecule Name	ε Region	Coefficient Order 0th	1th	2th	3th	4th	5th	6th	7th	8th
N,N-dimethylformamide	positive	-1.026662	-4.213129e-01	1.372187e+00	-2.758362e+00	1.122486e+00	-2.252456e-01	2.467778e-02	-1.413012e-03	3.313077e-05
N,N-dimethylformamide	negative	-0.994567	3.392553e-01	-3.026358e-02	2.563176e-03	4.098099e-04	2.291191e-05	5.775372e-07	4.026072e-09	-4.882096e-11
N,N-dimethylthioformamide	positive	-0.841709	-7.843264e-01	8.072388e-01	-1.102725e+00	3.609474e-01	-6.082947e-02	5.756769e-03	-2.906837e-04	6.107376e-06
N,N-dimethylthioformamide	negative	-0.870364	4.089936e-01	-4.491115e-02	1.095658e-02	2.156113e-03	1.784980e-04	8.023909e-06	1.915047e-07	1.906952e-09
N-methyl pyrrolidinone	positive	-1.138885	2.028837e-01	2.212266e-01	-2.329194e+00	1.050974e+00	-2.211728e-01	2.496666e-02	-1.460306e-03	3.479916e-05
N-methyl pyrrolidinone	negative	-1.045894	2.806242e-01	-2.269705e-02	1.321724e-03	1.979780e-04	8.032772e-06	3.887296e-08	-5.526114e-09	-1.101804e-10
N-methylacetamide	positive	-0.971041	-7.558858e-01	2.685987e+00	-3.815552e+00	1.495038e+00	-2.957687e-01	3.214861e-02	-1.830965e-03	4.275930e-05
N-methylacetamide	negative	-1.057572	-1.627625e-01	-4.918774e-01	-1.834077e-01	-3.454645e-02	-3.203095e-03	-1.572083e-04	-3.940175e-06	-3.975831e-08
N-methylformamide	positive	-0.917036	-7.499153e-01	2.108621e+00	-3.085488e+00	1.195129e+00	-2.339764e-01	2.521856e-02	-1.426474e-03	3.312496e-05
N-methylformamide	negative	-1.004859	-2.158321e-01	-5.544517e-01	-2.306867e-01	-4.896145e-02	-5.109253e-03	-2.828990e-04	-8.036514e-06	-9.251761e-08
acetic acid	positive	-0.975667	-6.005733e-01	5.269518e-01	-1.746808e-01	-1.296217e-01	5.357360e-02	-8.018689e-03	5.494448e-04	-1.447887e-05
acetic acid	negative	-0.848413	1.213769e+00	1.040118e+00	4.495912e-01	7.203073e-02	6.132129e-03	2.939296e-04	7.502015e-06	7.946764e-08
acetic anhydride	positive	-1.004030	-6.866539e-01	8.208229e-01	-1.015971e+00	3.260489e-01	-5.371791e-02	4.950376e-03	-2.427690e-04	4.945800e-06
acetic anhydride	negative	-1.033900	3.324099e-01	-3.558950e-02	1.091420e-02	1.879519e-03	1.413498e-04	5.749921e-06	1.227338e-07	1.076221e-09
acetone	positive	-0.946746	-6.209410e-01	1.286206e+00	-2.099822e+00	8.124743e-01	-1.574188e-01	1.677596e-02	-9.390676e-04	2.160703e-05
acetone	negative	-0.954252	3.799352e-01	-3.496309e-02	2.884898e-03	6.258445e-04	4.690239e-05	1.861574e-06	3.891459e-08	3.372833e-10
acetonitrile	positive	-0.890007	-1.028930e+00	1.591440e+00	-1.429501e+00	4.027594e-01	-5.817411e-02	4.596457e-03	-1.863540e-04	2.963100e-06
acetonitrile	negative	-0.957463	4.006980e-01	-5.103684e-02	2.417179e-02	5.037586e-03	4.598732e-04	2.277645e-05	5.954107e-07	6.451008e-09
acetophenone	positive	-1.035587	-1.103635e+00	2.098050e+00	-2.246237e+00	7.664113e-01	-1.364567e-01	1.360548e-02	-7.204309e-04	1.580409e-05
acetophenone	negative	-1.095734	2.987196e-01	-2.771455e-02	2.642062e-03	3.445964e-04	1.270915e-05	-6.249711e-08	-1.504596e-08	-2.731288e-10
acrylonitrile	positive	-0.951298	-9.088521e-01	1.338533e+00	-1.152813e+00	2.969979e-01	-3.747810e-02	2.371061e-03	-6.108352e-05	7.284002e-08
acrylonitrile	negative	-1.010254	3.911575e-01	-3.113318e-02	2.338150e-02	4.155574e-03	3.390932e-04	1.521205e-05	3.628905e-07	3.607852e-09

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
allyl alcohol	positive	-0.923042	-8.471008e-01	1.289169e+00	-1.671861e+00	5.938274e-01	-1.079852e-01	1.092842e-02	-5.857198e-04	1.298472e-05
allyl alcohol	negative	-0.903702	4.442004e-01	9.199905e-02	9.969019e-03	-9.490998e-03	-1.754741e-03	-1.274816e-04	-4.308530e-06	-5.633420e-08
aminoethanol	positive	-0.732261	-1.381674e+00	2.538456e+00	-2.554444e+00	8.379649e-01	-1.446776e-01	1.406316e-02	-7.286221e-04	1.568037e-05
aminoethanol	negative	-0.800914	3.233472e-01	-3.420746e-02	-3.945470e-03	-7.084336e-03	-1.133818e-03	-7.690779e-05	-2.476148e-06	-3.116189e-08
ammonia	positive	-1.092967	2.439675e+00	-5.694131e+00	5.016397e+00	-2.056090e+00	4.251579e-01	-4.726208e-02	2.710046e-03	-6.309725e-05
ammonia	negative	-0.911922	1.082704e+00	-7.590632e-01	-1.568647e-01	-1.941159e-02	-1.473717e-03	-6.718946e-05	-1.687094e-06	-1.792038e-08
aniline	positive	-0.983026	-5.891364e-01	3.434857e-01	-2.902169e-01	-6.416851e-03	1.807930e-02	-3.251646e-03	2.381144e-04	-6.478800e-06
aniline	negative	-0.972816	4.882414e-01	1.865492e-01	1.122971e-01	1.707675e-02	1.322987e-03	5.697119e-05	1.296679e-06	1.216124e-08
anisole	positive	-1.099428	-5.632276e-01	8.434079e-03	-1.925848e-01	4.054378e-02	-2.643298e-03	-1.199677e-04	2.315738e-05	-8.164432e-07
anisole	negative	-1.107089	2.233003e-01	-2.099917e-02	1.258410e-03	6.165841e-05	-8.757133e-06	-8.865579e-07	-3.072260e-08	-3.863969e-10
benzaldehyde	positive	-1.063402	-8.802906e-01	1.356040e+00	-1.484513e+00	4.842231e-01	-8.205588e-02	7.794973e-03	-3.941336e-04	8.276339e-06
benzaldehyde	negative	-1.109776	2.591776e-01	-2.717476e-02	4.177566e-03	4.276151e-04	6.031280e-06	-9.327595e-07	-4.899555e-08	-7.394734e-10
benzene	positive	-1.189915	-2.975202e-01	-7.854975e-02	-4.775659e-03	-3.519288e-03	1.315465e-03	-1.690335e-04	9.986133e-06	-2.302440e-07
benzene	negative	-1.189496	1.985797e-01	-1.735657e-02	8.341845e-04	4.654585e-05	-5.258833e-06	-5.368625e-07	-1.816879e-08	-2.226062e-10
benzonitrile	positive	-1.048517	-1.097759e+00	1.840836e+00	-1.536684e+00	4.182550e-01	-5.795094e-02	4.309807e-03	-1.583963e-04	2.091936e-06
benzonitrile	negative	-1.127831	2.947858e-01	-2.584035e-02	1.147186e-02	1.648366e-03	9.958969e-05	2.902069e-06	3.215559e-08	-4.757467e-11
benzoyl bromide	positive	-1.140607	-2.762925e-01	-1.132796e-01	6.273632e-02	-8.628298e-02	2.524515e-02	-3.301780e-03	2.081593e-04	-5.164174e-06
benzoyl bromide	negative	-1.142758	2.823423e-01	-2.552310e-02	9.769452e-03	1.192997e-03	5.231034e-05	3.772888e-07	-3.710464e-08	-8.229108e-10
benzoyl chloride	positive	-1.135102	-4.112152e-01	1.393608e-01	-1.109153e-01	-5.135222e-02	2.331377e-02	-3.518297e-03	2.398418e-04	-6.268021e-06
benzoyl chloride	negative	-1.151465	2.658659e-01	-2.591277e-02	8.140501e-03	9.152816e-04	3.002579e-05	-5.942907e-07	-5.931235e-08	-1.031597e-09
benzyl alcohol	positive	-1.036152	-9.222333e-01	1.339480e+00	-1.362757e+00	4.206265e-01	-6.731232e-02	6.022895e-03	-2.860547e-04	5.625470e-06
benzyl alcohol	negative	-1.006913	5.162709e-01	2.231354e-01	5.542374e-02	-2.926053e-03	-1.259071e-03	-1.071265e-04	-3.883826e-06	-5.290022e-08
benzyl methyl ketone	positive	-1.012960	-1.160353e+00	2.262328e+00	-2.505430e+00	8.762661e-01	-1.592704e-01	1.616562e-02	-8.695243e-04	1.934327e-05
benzyl methyl ketone	negative	-1.071060	2.959356e-01	-2.816856e-02	2.965240e-03	3.599461e-04	1.101198e-05	-2.557430e-07	-2.225667e-08	-3.696853e-10
benzylamine	positive	-0.957606	-2.209556e+00	4.746851e+00	-4.115330e+00	1.345991e+00	-2.342428e-01	2.294078e-02	-1.195283e-03	2.582312e-05
benzylamine	negative	-1.115044	2.981421e-01	-3.182591e-02	3.725970e-03	-5.791067e-04	-1.476949e-04	-1.090602e-05	-3.581109e-07	-4.507166e-09
bis(2-chloroethyl) ether	positive	-1.027982	-1.284074e+00	1.835173e+00	-1.461526e+00	3.742673e-01	-4.713444e-02	2.958614e-03	-7.364565e-05	-3.183870e-08

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
bis(2-chloroethyl) ether	negative	-1.114571	2.888501e-01	-2.890324e-02	5.017551e-03	6.827925e-04	3.243093e-05	4.384971e-07	-1.264316e-08	-3.465295e-10
bromobenzene	positive	-1.211649	-1.950952e-01	-2.971593e-02	-2.934090e-03	5.235983e-07	-2.277310e-06	6.115580e-06	-6.357222e-07	1.929776e-08
bromobenzene	negative	-1.210721	2.304548e-01	-2.518351e-02	3.807067e-03	3.582655e-04	6.959800e-08	-1.212658e-06	-5.587748e-08	-8.090430e-10
bromoform	positive	-1.235369	-1.270848e-01	-7.795914e-03	4.777123e-05	-1.665210e-04	4.415585e-05	-5.610851e-06	3.604316e-07	-9.277175e-09
bromoform	negative	-1.230539	4.441234e-01	-3.998380e-02	2.890628e-02	4.877835e-03	3.853956e-04	1.697268e-05	4.012182e-07	3.977461e-09
butanoic acid	positive	-0.976912	-9.148617e-01	1.403207e+00	-1.551652e+00	5.096408e-01	-8.682323e-02	8.287808e-03	-4.210345e-04	8.882872e-06
butanoic acid	negative	-0.837766	9.531500e-01	5.442380e-01	1.231030e-01	-9.759377e-04	-1.848570e-03	-1.721241e-04	-6.490239e-06	-9.059340e-08
butyl acetate	positive	-1.118116	-8.556103e-01	1.500635e+00	-1.970699e+00	7.139616e-01	-1.322924e-01	1.361568e-02	-7.406998e-04	1.663804e-05
butyl acetate	negative	-1.150851	2.184744e-01	-1.557473e-02	6.432648e-04	5.498406e-05	-4.805272e-07	-1.658881e-07	-6.128762e-09	-7.479294e-11
butyraldehyde	positive	-1.081857	-7.123522e-01	9.922994e-01	-1.353706e+00	4.759671e-01	-8.539811e-02	8.531322e-03	-4.518733e-04	9.911654e-06
butyraldehyde	negative	-1.110285	2.554868e-01	-2.125699e-02	1.082561e-03	3.492455e-05	-9.485480e-06	-8.338755e-07	-2.729232e-08	-3.305717e-10
carbon disulfide	positive	-1.307329	-1.331750e-14	1.535566e-14	-8.375219e-15	2.519390e-15	-4.455997e-16	4.617044e-17	-2.589313e-18	6.062553e-20
carbon disulfide	negative	-1.306941	1.497615e-01	-1.405268e-02	3.532408e-04	-9.846491e-05	-1.476241e-05	-7.493433e-07	-1.720725e-08	-1.504067e-10
carbon tetrachloride	positive	-1.261980	-3.877713e-02	-5.338951e-03	1.845617e-03	-3.481963e-04	2.569704e-05	5.555989e-07	-1.775134e-07	6.831965e-09
carbon tetrachloride	negative	-1.260284	3.119126e-01	-3.204632e-02	5.649235e-03	8.838734e-04	5.406136e-05	1.616274e-06	2.007596e-08	2.295846e-11
chlorobenzene	positive	-1.220774	-2.122768e-01	-3.561007e-02	-2.909212e-03	-2.999222e-04	7.370811e-05	-8.888351e-07	-3.596108e-07	1.568975e-08
chlorobenzene	negative	-1.219861	2.103213e-01	-2.369259e-02	3.164763e-03	1.627164e-04	-2.073565e-05	-2.320733e-06	-8.584762e-08	-1.138251e-09
chloroform	positive	-1.220126	-1.662607e-01	-9.942860e-03	-4.672309e-04	-2.678255e-06	1.475957e-05	-2.411077e-06	1.740120e-07	-4.822481e-09
chloroform	negative	-1.187073	4.507678e-01	1.422460e-01	9.442700e-02	1.417198e-02	1.072633e-03	4.489395e-05	9.877984e-07	8.891901e-09
cinnamaldehyde	positive	-1.049978	-1.188297e+00	2.374332e+00	-2.412035e+00	8.151099e-01	-1.441945e-01	1.429285e-02	-7.526208e-04	1.642305e-05
cinnamaldehyde	negative	-1.119384	2.714404e-01	-2.651356e-02	7.031756e-03	8.085601e-04	2.707704e-05	-4.691457e-07	-4.994430e-08	-8.723444e-10
cis-decalin	positive	-1.250649	-1.040285e-02	-5.911493e-04	-1.045394e-04	1.008522e-04	-2.630826e-05	3.257800e-06	-1.986837e-07	4.812402e-09
cis-decalin	negative	-1.249260	1.509201e-01	-3.231850e-03	2.682489e-04	4.100234e-05	3.223094e-06	1.542416e-07	4.067209e-09	4.500708e-11
cis-perfluorodecalin	positive	-1.198594	-2.176718e-02	-2.745649e-03	9.409617e-04	-2.088799e-04	2.473520e-05	-1.494943e-06	3.511522e-08	8.785907e-11
cis-perfluorodecalin	negative	-1.191385	2.839656e-01	-1.220737e-02	1.822717e-02	1.726858e-03	3.470725e-05	-3.046931e-06	-1.823644e-07	-2.916539e-09
cyclohexane	positive	-1.265174	-7.534225e-03	-2.617612e-04	-9.611194e-05	5.185752e-05	-1.027065e-05	1.013843e-06	-5.024906e-08	9.998582e-10
cyclohexane	negative	-1.263755	1.572376e-01	-3.257131e-03	2.778808e-04	4.370735e-05	3.478430e-06	1.674658e-07	4.436342e-09	4.931251e-11

Continued on next page

Molecule Name	ε Region	Coefficient Order 0th	1th	2th	3th	4th	5th	6th	7th	8th
cyclohexanol	positive	-1.045304	-1.008202e+00	1.914852e+00	-2.136170e+00	7.407897e-01	-1.332562e-01	1.338928e-02	-7.135150e-04	1.573936e-05
cyclohexanol	negative	-1.062352	2.767972e-01	-6.118314e-02	-7.460781e-02	-2.746725e-02	-3.655062e-03	-2.354432e-04	-7.473317e-06	-9.399565e-08
cyclohexanone	positive	-1.107701	-4.739699e-01	1.104379e+00	-2.164411e+00	8.683263e-01	-1.719499e-01	1.861978e-02	-1.055466e-03	2.453549e-05
cyclohexanone	negative	-1.097513	2.598276e-01	-2.030397e-02	1.032522e-03	1.255622e-04	2.575703e-06	-1.616493e-07	-9.089073e-09	-1.330960e-10
cyclopentanone	positive	-1.109450	-4.758185e-01	1.001566e+00	-2.035862e+00	8.172843e-01	-1.616524e-01	1.748089e-02	-9.896602e-04	2.298030e-05
cyclopentanone	negative	-1.100517	2.739290e-01	-2.220193e-02	1.218987e-03	1.995855e-04	9.214041e-06	1.275723e-07	-2.697551e-09	-7.573462e-11
di-n-butylamine	positive	-1.204755	-9.289133e-01	3.555454e+00	-4.612066e+00	1.789102e+00	-3.530602e-01	3.835336e-02	-2.184447e-03	5.102805e-05
di-n-butylamine	negative	-1.206633	1.414951e-01	-2.054314e-02	-2.056509e-03	-2.533343e-04	4.537314e-05	6.735651e-06	2.983533e-07	4.488261e-09
di-n-butylorthophthalate	positive	-1.127885	-7.344878e-01	1.200648e+00	-1.960801e+00	7.499961e-01	-1.438886e-01	1.521030e-02	-8.457537e-04	1.935182e-05
di-n-butylorthophthalate	negative	-1.140677	1.839395e-01	-1.492034e-02	5.728321e-04	-3.196237e-05	-6.310619e-06	-2.726810e-07	-4.456747e-09	-1.665685e-11
di-n-propyl ether	positive	-1.170545	-7.487239e-01	1.205907e+00	-1.751811e+00	6.471714e-01	-1.211503e-01	1.255745e-02	-6.869081e-04	1.550023e-05
di-n-propyl ether	negative	-1.193634	1.520120e-01	-6.143164e-03	3.265838e-04	3.848856e-05	2.615882e-06	1.284146e-07	3.668895e-09	4.388811e-11
dibenzyl ether	positive	-1.168831	-4.093266e-01	-6.220807e-03	-8.147222e-02	-2.357836e-02	1.234023e-02	-1.891521e-03	1.287451e-04	-3.345058e-06
dibenzyl ether	negative	-1.177149	1.931783e-01	-1.562828e-02	6.571086e-04	1.999906e-05	-5.490710e-06	-4.521547e-07	-1.394646e-08	-1.598459e-10
dibromomethane	positive	-1.204122	-2.497659e-01	-2.410580e-02	-1.863594e-03	2.330110e-04	-1.846143e-05	2.088461e-06	-1.488322e-07	3.994112e-09
dibromomethane	negative	-1.196433	3.778413e-01	-9.612559e-03	3.551916e-02	5.941489e-03	4.740414e-04	2.091696e-05	4.909639e-07	4.796634e-09
dibutyl ether	positive	-1.159555	-8.881057e-01	1.581952e+00	-1.987550e+00	7.135710e-01	-1.314250e-01	1.346018e-02	-7.291627e-04	1.631855e-05
dibutyl ether	negative	-1.195882	1.571710e-01	-5.213379e-03	3.072989e-04	4.358237e-05	3.256140e-06	1.585292e-07	4.360731e-09	5.042880e-11
dibutyl sulfide	positive	-1.195378	-4.756672e-01	5.123250e-01	-4.575915e-01	7.792091e-02	-1.138822e-03	-9.939841e-04	1.036985e-04	-3.256942e-06
dibutyl sulfide	negative	-1.225394	1.805517e-01	-8.020861e-03	3.741074e-04	4.811518e-05	3.068341e-06	1.398659e-07	3.895465e-09	4.685184e-11
dibutyl sulfoxide	positive	-1.190005	2.476613e-02	1.660652e+00	-3.677740e+00	1.553128e+00	-3.197085e-01	3.570075e-02	-2.074783e-03	4.923194e-05
dibutyl sulfoxide	negative	-1.097893	2.572014e-01	-1.984941e-02	9.535848e-04	7.300264e-05	-2.015851e-06	-3.505367e-07	-1.298916e-08	-1.652968e-10
dichloroacetic acid	positive	-1.093494	-1.697568e-01	-1.737701e-01	9.234743e-02	-3.999437e-02	6.488061e-03	-4.674096e-04	1.293005e-05	-1.594566e-08
dichloroacetic acid	negative	-1.243968	-2.468293e-02	1.468593e-01	4.095155e-01	9.221966e-02	9.747854e-03	5.528989e-04	1.624421e-05	1.943702e-07
dichloromethane	positive	-1.189386	-2.882673e-01	-3.833106e-02	-2.729660e-03	9.270732e-05	5.071457e-05	-5.318730e-06	1.816119e-07	-1.313649e-09
dichloromethane	negative	-1.176839	3.650568e-01	2.278269e-02	4.488576e-02	7.046246e-03	5.379154e-04	2.262496e-05	5.020085e-07	4.587114e-09

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
diethanolamine	positive	-0.757092	-5.554292e-01	-9.210169e-03	-9.690592e-01	4.046202e-01	-8.081081e-02	8.799095e-03	-5.011126e-04	1.169425e-05
diethanolamine	negative	-0.680328	4.372788e-01	1.142938e-01	5.881110e-03	-1.272094e-02	-2.241203e-03	-1.605613e-04	-5.394171e-06	-7.032271e-08
diethyl carbonate	positive	-1.055694	-6.852655e-01	5.999001e-01	-1.143515e+00	4.200298e-01	-7.729282e-02	7.880337e-03	-4.249036e-04	9.470571e-06
diethyl carbonate	negative	-1.071072	2.757722e-01	-1.766985e-02	6.181134e-04	1.225541e-04	6.722831e-06	1.801981e-07	2.197433e-09	6.778439e-12
diethyl ether	positive	-1.129426	-5.867176e-01	5.928289e-01	-1.430427e+00	5.683280e-01	-1.106344e-01	1.178859e-02	-6.589310e-04	1.513495e-05
diethyl ether	negative	-1.129687	1.852348e-01	-6.794973e-03	3.513602e-04	5.208413e-05	3.814780e-06	1.823031e-07	5.002368e-09	5.827489e-11
diethyl malonate	positive	-1.001798	-7.240751e-01	6.779848e-01	-1.307908e+00	4.915549e-01	-9.220045e-02	9.548694e-03	-5.216149e-04	1.175496e-05
diethyl malonate	negative	-1.015218	2.883752e-01	-2.281383e-02	1.219968e-03	1.740846e-04	6.890261e-06	2.691844e-08	-4.945838e-09	-9.634369e-11
diethyl sulfate	positive	-1.026366	-6.706218e-01	4.322043e-01	-7.514381e-01	2.499028e-01	-4.195147e-02	3.930571e-03	-1.960674e-04	4.067167e-06
diethyl sulfate	negative	-1.045133	2.763100e-01	-2.692733e-02	2.731084e-03	3.914227e-04	1.770740e-05	1.903441e-07	-8.590203e-09	-2.059536e-10
diethyl sulfide	positive	-1.175237	-5.215699e-01	4.944057e-01	-5.562948e-01	1.409147e-01	-1.666890e-02	9.193935e-04	-1.429057e-05	-3.488334e-07
diethyl sulfide	negative	-1.200890	2.160486e-01	-1.288420e-02	4.977308e-04	5.745433e-05	1.706581e-06	4.713001e-09	-6.426419e-10	-8.352049e-12
diethyl sulfite	positive	-1.059992	-6.468280e-01	6.185778e-01	-1.262521e+00	4.789585e-01	-9.033456e-02	9.393414e-03	-5.147929e-04	1.163241e-05
diethyl sulfite	negative	-1.070955	2.649034e-01	-1.991938e-02	9.595166e-04	1.106875e-04	1.867804e-06	-1.656863e-07	-8.525767e-09	-1.211843e-10
diethylamine	positive	-1.200219	-5.366098e-01	2.607542e+00	-4.016538e+00	1.608952e+00	-3.225842e-01	3.540326e-02	-2.031490e-03	4.772995e-05
diethylamine	negative	-1.168759	1.798610e-01	-3.347980e-02	-6.812002e-03	-1.502131e-03	-1.104747e-04	-2.976541e-06	-1.144652e-10	8.615666e-10
diethylene glycol	positive	-0.884778	-5.740656e-01	-1.533613e-01	-6.692637e-01	2.921061e-01	-5.831714e-02	6.274934e-03	-3.523312e-04	8.112067e-06
diethylene glycol	negative	-0.707477	9.143510e-01	6.287372e-01	1.917443e-01	1.684103e-02	2.641650e-04	-4.236781e-05	-2.459060e-06	-4.037309e-08
diethylene glycol dimethyl ether	positive	-1.127449	-6.142188e-01	1.275496e-01	-8.068991e-01	3.210459e-01	-6.144364e-02	6.433309e-03	-3.540053e-04	8.021571e-06
diethylene glycol dimethyl ether	negative	-1.129541	1.839603e-01	-9.825982e-03	4.332357e-04	4.081143e-05	1.720452e-06	6.578825e-08	2.021846e-09	2.809793e-11
dihydrolevoglucosenone	positive	-0.922182	-6.845493e-01	6.914088e-01	-1.270180e+00	4.700545e-01	-8.726686e-02	8.969964e-03	-4.871333e-04	1.092559e-05
dihydrolevoglucosenone	negative	-0.936907	3.414829e-01	-3.455061e-02	5.207085e-03	8.598450e-04	5.587672e-05	1.861453e-06	2.974488e-08	1.565866e-10
diisopropyl ether	positive	-1.141392	-6.696884e-01	1.162635e+00	-1.876978e+00	7.166974e-01	-1.373310e-01	1.450093e-02	-8.054800e-04	1.841297e-05
diisopropyl ether	negative	-1.154167	1.703626e-01	-7.391766e-03	3.637859e-04	4.336648e-05	2.767452e-06	1.308596e-07	3.739349e-09	4.541101e-11
diisopropyl sulfide	positive	-1.164154	-4.735286e-01	4.314330e-01	-4.382414e-01	8.634944e-02	-4.913856e-03	-4.190595e-04	6.394891e-05	-2.202846e-06
diisopropyl sulfide	negative	-1.189443	1.922374e-01	-1.085077e-02	4.543487e-04	4.217050e-05	1.305484e-06	2.880560e-08	8.252923e-10	1.391437e-11

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
dimethyl carbonate	positive	-0.916953	-6.611631e-01	1.133815e-01	-7.279292e-01	2.829781e-01	-5.314295e-02	5.480797e-03	-2.979298e-04	6.683370e-06
dimethyl carbonate	negative	-0.921456	3.334930e-01	-2.803446e-02	1.635915e-03	3.517306e-04	2.400994e-05	8.432609e-07	1.506534e-08	1.053738e-10
dimethyl sulfate	positive	-1.024729	-5.868370e-01	2.427458e-01	-5.135882e-01	1.640814e-01	-2.611119e-02	2.308199e-03	-1.083159e-04	2.108426e-06
dimethyl sulfate	negative	-1.036240	2.814025e-01	-2.610628e-02	1.223609e-02	1.797934e-03	1.143337e-04	3.675114e-06	5.299426e-08	1.812449e-10
dimethyl sulfide	positive	-1.102596	-5.273926e-01	3.699272e-01	-5.413417e-01	1.583472e-01	-2.297399e-02	1.817702e-03	-7.421731e-05	1.204088e-06
dimethyl sulfide	negative	-1.121694	2.749600e-01	-2.365304e-02	1.523237e-03	2.487826e-04	1.196477e-05	1.895516e-07	-2.668434e-09	-8.948264e-11
dimethylcyanamide	positive	-0.939983	-9.883430e-01	1.890840e+00	-2.268517e+00	8.117570e-01	-1.496598e-01	1.535609e-02	-8.334633e-04	1.868596e-05
dimethylcyanamide	negative	-0.981981	3.328570e-01	-3.674269e-02	1.062623e-02	1.930459e-03	1.503281e-04	6.345959e-06	1.416193e-07	1.311836e-09
dimethylphthalate	positive	-0.996607	-7.879669e-01	6.734243e-01	-1.145176e+00	4.107464e-01	-7.429723e-02	7.464649e-03	-3.973074e-04	8.753794e-06
dimethylphthalate	negative	-1.016631	2.577678e-01	-2.558324e-02	2.476071e-03	2.634916e-04	3.527779e-06	-5.655087e-07	-2.893867e-08	-4.284844e-10
dimethylsulfoxide	positive	-0.841684	-1.421172e+00	4.303744e+00	-4.867811e+00	1.816664e+00	-3.505053e-01	3.747169e-02	-2.108042e-03	4.875466e-05
dimethylsulfoxide	negative	-0.890509	4.681503e-01	-6.051261e-02	6.078221e-03	1.671249e-03	1.580460e-04	7.864337e-06	2.050695e-07	2.212832e-09
diphenyl ether	positive	-1.167208	-3.456227e-01	-1.422638e-01	3.292020e-02	-3.198821e-02	8.414076e-03	-1.010709e-03	5.923477e-05	-1.379568e-06
diphenyl ether	negative	-1.166215	1.956812e-01	-1.858138e-02	9.751170e-04	3.184531e-05	-9.289753e-06	-8.252295e-07	-2.727227e-08	-3.329349e-10
diphenyl ketone	positive	-1.075753	-1.181271e+00	2.353403e+00	-2.536486e+00	8.838086e-01	-1.602569e-01	1.623146e-02	-8.713758e-04	1.934987e-05
diphenyl ketone	negative	-1.137715	2.312834e-01	-2.248454e-02	1.598034e-03	7.301872e-05	-1.215582e-05	-1.220690e-06	-4.273019e-08	-5.437291e-10
ethanol	positive	-0.928881	-6.529969e-01	1.117066e+00	-1.797165e+00	6.827629e-01	-1.302703e-01	1.370624e-02	-7.590500e-04	1.730739e-05
ethanol	negative	-0.887748	4.264682e-01	4.726724e-02	-2.542575e-02	-1.845987e-02	-2.799427e-03	-1.907144e-04	-6.248878e-06	-8.025020e-08
ethyl acetate	positive	-1.028282	-7.681216e-01	1.112543e+00	-1.710938e+00	6.357115e-01	-1.193290e-01	1.239312e-02	-6.790292e-04	1.534425e-05
ethyl acetate	negative	-1.048914	2.904939e-01	-2.133067e-02	9.558287e-04	1.687813e-04	8.626755e-06	1.837425e-07	4.583940e-10	-2.744094e-11
ethyl acetoacetate	positive	-1.039025	-8.614806e-01	1.269186e+00	-1.578464e+00	5.439924e-01	-9.652216e-02	9.566741e-03	-5.035278e-04	1.098536e-05
ethyl acetoacetate	negative	-1.076533	2.838938e-01	-2.694446e-02	2.457161e-03	2.338310e-04	2.170803e-07	-7.310754e-07	-3.308122e-08	-4.707312e-10
ethyl benzoate	positive	-1.073470	-1.013164e+00	1.596361e+00	-1.735271e+00	5.783385e-01	-1.001426e-01	9.711600e-03	-5.008411e-04	1.071831e-05
ethyl benzoate	negative	-1.125498	2.410702e-01	-1.901414e-02	9.152977e-04	7.062883e-05	-2.449088e-06	-3.806346e-07	-1.388100e-08	-1.753520e-10
ethyl chloroacetate	positive	-1.122298	-4.338978e-01	1.606660e-01	-1.433931e-01	-3.795615e-02	2.059671e-02	-3.220385e-03	2.228885e-04	-5.874580e-06
ethyl chloroacetate	negative	-1.140232	3.403737e-01	-3.403231e-02	5.826819e-03	7.770173e-04	3.836046e-05	6.718088e-07	-7.485411e-09	-2.983879e-10
ethyl formate	positive	-0.989986	-7.606978e-01	8.166011e-01	-1.340678e+00	4.885063e-01	-8.983844e-02	9.163149e-03	-4.942734e-04	1.101947e-05
ethyl formate	negative	-1.010271	2.999657e-01	-2.585749e-02	1.712713e-03	2.882100e-04	1.555091e-05	3.558132e-07	1.273021e-09	-5.116056e-11

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
ethyl phenyl ether	positive	-1.122177	-5.022706e-01	1.109463e-02	-1.707757e-01	2.897006e-02	-9.246625e-05	-4.099488e-04	3.996975e-05	-1.210973e-06
ethyl phenyl ether	negative	-1.129773	2.191490e-01	-1.708823e-02	7.579715e-04	5.241600e-05	-2.533788e-06	-3.259735e-07	-1.130396e-08	-1.379863e-10
ethyl phenyl ketone	positive	-1.072284	-1.130086e+00	2.088671e+00	-2.166567e+00	7.239430e-01	-1.265873e-01	1.241939e-02	-6.480885e-04	1.402876e-05
ethyl phenyl ketone	negative	-1.136988	2.463506e-01	-2.343317e-02	1.625127e-03	3.759797e-05	-1.725572e-05	-1.527496e-06	-5.163263e-08	-6.463312e-10
ethyl propionate	positive	-1.120045	-7.629037e-01	1.179984e+00	-1.776306e+00	6.605000e-01	-1.241670e-01	1.291361e-02	-7.084135e-04	1.602521e-05
ethyl propionate	negative	-1.141378	2.160622e-01	-1.518324e-02	6.269856e-04	4.867116e-05	-1.246736e-06	-2.033506e-07	-7.007376e-09	-8.310997e-11
ethyl trichloroacetate	positive	-1.152009	-2.686945e-01	-1.546785e-01	2.132747e-01	-1.828872e-01	4.987147e-02	-6.411679e-03	4.031787e-04	-1.002572e-05
ethyl trichloroacetate	negative	-1.160545	3.024899e-01	-3.144415e-02	5.339570e-03	8.393453e-04	5.126778e-05	1.524555e-06	1.864461e-08	1.615086e-11
ethylbenzene	positive	-1.199384	-2.438849e-01	-8.221834e-02	4.706478e-03	-8.906524e-03	2.405182e-03	-2.777697e-04	1.544264e-05	-3.407344e-07
ethylbenzene	negative	-1.198417	1.974898e-01	-1.242067e-02	4.944824e-04	4.340785e-05	3.471384e-07	-5.408840e-08	-1.901737e-09	-1.921359e-11
ethylene carbonate	positive	-0.725872	-1.164653e+00	1.870403e+00	-1.595483e+00	4.229405e-01	-5.645399e-02	3.979872e-03	-1.334538e-04	1.414023e-06
ethylene carbonate	negative	-0.815968	4.949575e-01	-1.206365e-01	8.901208e-03	3.665625e-03	4.071107e-04	2.268846e-05	6.463185e-07	7.492507e-09
ethylene glycol	positive	-0.575526	-5.144307e-01	-4.835076e-01	-5.556859e-01	2.768312e-01	-5.824684e-02	6.473367e-03	-3.719944e-04	8.718879e-06
ethylene glycol	negative	-0.389614	1.006167e+00	6.676750e-01	2.265044e-01	2.504941e-02	1.181196e-03	1.182101e-05	-8.198926e-07	-2.034614e-08
ethylenediamine	positive	-0.896011	-1.169582e+00	3.699716e+00	-4.417470e+00	1.670003e+00	-3.241976e-01	3.479090e-02	-1.962499e-03	4.548254e-05
ethylenediamine	negative	-0.921676	3.655967e-01	-1.921220e-02	2.229500e-02	3.609858e-03	2.699149e-04	1.099368e-05	2.349226e-07	2.055731e-09
fluorobenzene	positive	-1.212862	-1.615090e-01	-3.392162e-02	-3.353012e-03	-4.279935e-04	-8.294712e-06	1.982070e-05	-2.018627e-06	6.151351e-08
fluorobenzene	negative	-1.212124	2.106944e-01	-2.321577e-02	2.446376e-03	1.642739e-04	-1.094350e-05	-1.464731e-06	-5.576016e-08	-7.447860e-10
formamide	positive	-0.364552	-2.410346e+00	4.997939e+00	-3.969872e+00	1.220738e+00	-2.005812e-01	1.856231e-02	-9.137634e-04	1.864203e-05
formamide	negative	-0.488836	6.367041e-01	3.948967e-01	2.140707e-01	3.426956e-02	2.861282e-03	1.342333e-04	3.356022e-06	3.487063e-08
formic acid	positive	-0.955063	-2.215387e-01	-2.437649e-01	3.253938e-01	-2.462612e-01	6.527068e-02	-8.303364e-03	5.196575e-04	-1.289210e-05
formic acid	negative	-0.876824	9.646329e-01	9.326781e-01	5.176631e-01	9.444525e-02	8.932492e-03	4.696005e-04	1.302763e-05	1.489263e-07
furan	positive	-1.164100	-2.922032e-01	-1.268954e-01	2.057692e-02	-3.224557e-02	9.282088e-03	-1.172094e-03	7.126358e-05	-1.709880e-06
furan	negative	-1.162691	2.314644e-01	-2.508634e-02	5.794114e-03	5.965018e-04	1.084555e-05	-1.135914e-06	-6.414455e-08	-9.946501e-10
gamma-butyrolactone	positive	-0.904792	-7.274744e-01	1.176102e+00	-1.969073e+00	7.398432e-01	-1.402126e-01	1.470700e-02	-8.137255e-04	1.855910e-05
gamma-butyrolactone	negative	-0.916422	3.216833e-01	-3.539274e-02	9.318102e-03	1.684076e-03	1.282204e-04	5.265921e-06	1.138902e-07	1.017680e-09
glycerol	positive	-0.588453	-5.830061e-01	-2.164204e+00	4.074295e-01	1.068188e-03	-1.325379e-02	2.223242e-03	-1.552442e-04	4.095955e-06
glycerol	negative	-0.454556	3.577487e-01	1.033620e-01	-3.358946e-02	-2.511643e-02	-3.823802e-03	-2.617649e-04	-8.620328e-06	-1.112385e-07

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
heptanoic acid	positive	-1.129264	-4.160789e-01	1.624252e-01	3.344660e-01	-3.732224e-01	1.087639e-01	-1.459041e-02	9.482665e-04	-2.422855e-05
heptanoic acid	negative	-0.945392	1.346259e+00	1.119348e+00	4.124197e-01	5.809756e-02	4.310463e-03	1.768353e-04	3.762683e-06	3.198079e-08
hexafluorobenzene	positive	-1.243311	-8.959422e-02	-6.273896e-03	1.093127e-03	-5.004731e-04	1.074659e-04	-1.256794e-05	7.645971e-07	-1.890995e-08
hexafluorobenzene	negative	-1.242590	2.224225e-01	-2.467523e-02	3.807710e-03	3.610061e-04	2.808923e-07	-1.207163e-06	-5.586577e-08	-8.101897e-10
hexamethylphosphoric triamide	positive	-1.323988	1.879068e+00	-2.787784e+00	-1.361696e+00	9.061574e-01	-2.152670e-01	2.592662e-02	-1.581826e-03	3.885784e-05
hexamethylphosphoric triamide	negative	-1.065320	2.083170e-01	-1.296873e-02	3.751806e-04	1.536084e-05	-2.300100e-06	-1.895216e-07	-5.598652e-09	-6.137355e-11
hexanoic acid	positive	-1.029724	-1.008700e+00	1.713668e+00	-1.671648e+00	5.215512e-01	-8.497963e-02	7.763548e-03	-3.771929e-04	7.601839e-06
hexanoic acid	negative	-0.908101	8.914197e-01	4.939126e-01	9.245096e-02	-7.884323e-03	-2.611172e-03	-2.169836e-04	-7.844029e-06	-1.071080e-07
hydrazine	positive	-0.574505	-1.377468e+00	4.008172e+00	-4.421845e+00	1.629684e+00	-3.112684e-01	3.299806e-02	-1.843278e-03	4.237591e-05
hydrazine	negative	-0.638507	5.462736e-01	-1.202215e-01	3.591542e-03	2.534136e-03	2.959372e-04	1.678126e-05	4.820455e-07	5.614915e-09
hydrogen fluoride	positive	-0.434075	4.562071e-02	-3.803266e-01	3.374770e-01	-1.795348e-01	4.101726e-02	-4.742655e-03	2.761243e-04	-6.460086e-06
hydrogen fluoride	negative	-0.875767	-2.098639e+00	-2.121163e+00	-2.793134e-01	-1.437268e-02	3.873878e-04	8.061455e-05	3.475677e-06	5.095788e-08
hydrogen peroxide	positive	-0.068345	-1.266862e+00	-5.581361e-01	3.917690e-02	2.015919e-02	-6.120790e-03	7.578750e-04	-4.559435e-05	1.093481e-06
hydrogen peroxide	negative	-0.105224	1.361692e-01	3.112663e-01	3.333778e-01	6.890535e-02	6.972562e-03	3.841087e-04	1.104339e-05	1.299196e-07
iodobenzene	positive	-1.215297	-1.959628e-01	-2.371654e-02	-2.283365e-03	1.992247e-04	-2.462606e-05	4.430317e-06	-3.448636e-07	9.401153e-09
iodobenzene	negative	-1.214113	2.331187e-01	-2.540824e-02	5.172498e-03	5.740135e-04	1.476000e-05	-7.128885e-07	-4.818309e-08	-7.773522e-10
isopentyl acetate	positive	-1.097834	-9.155021e-01	1.664746e+00	-2.037302e+00	7.241809e-01	-1.325258e-01	1.350680e-02	-7.287645e-04	1.625417e-05
isopentyl acetate	negative	-1.138469	2.189406e-01	-1.556222e-02	6.437046e-04	4.541145e-05	-1.174785e-06	-1.837008e-07	-6.211299e-09	-7.237247e-11
isopropylbenzene	positive	-1.206740	-2.140002e-01	-8.012677e-02	9.847253e-03	-1.021359e-02	2.460162e-03	-2.649808e-04	1.390183e-05	-2.907578e-07
isopropylbenzene	negative	-1.205497	1.882423e-01	-1.200117e-02	4.859126e-04	3.439148e-05	-3.263708e-07	-7.452622e-08	-2.125341e-09	-1.890388e-11
meta-cresol	positive	-1.066100	-4.363768e-01	7.801872e-03	-1.164830e-01	-2.594060e-03	7.301792e-03	-1.285630e-03	9.222385e-05	-2.463481e-06
meta-cresol	negative	-0.861363	1.377276e+00	1.329883e+00	5.702222e-01	9.362025e-02	8.197717e-03	4.042068e-04	1.060462e-05	1.153568e-07
meta-dichlorobenzene	positive	-1.232767	-1.755781e-01	-1.866219e-02	-9.493530e-04	-5.613048e-05	2.001723e-05	-1.517647e-06	6.159991e-08	-1.294225e-09
meta-dichlorobenzene	negative	-1.230580	2.451110e-01	-2.324895e-02	5.669636e-03	2.275100e-04	-4.435307e-05	-4.679596e-06	-1.733727e-07	-2.322036e-09
meta-xylene	positive	-1.181036	-3.004814e-01	-8.797094e-02	-6.722618e-03	-5.414048e-03	2.051606e-03	-2.719695e-04	1.657718e-05	-3.933176e-07
meta-xylene	negative	-1.180518	2.028485e-01	-1.243433e-02	4.695654e-04	5.602131e-05	1.629602e-06	-7.985833e-11	-8.090106e-10	-1.058405e-11

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
methanesulfonic acid	positive	-0.736613	-7.049307e-01	4.746359e-01	-8.014696e-01	2.679604e-01	-4.529424e-02	4.272048e-03	-2.144366e-04	4.474648e-06
methanesulfonic acid	negative	-0.459162	1.564298e+00	1.172699e+00	4.107636e-01	5.029984e-02	2.955711e-03	7.811240e-05	3.773372e-07	-1.302114e-08
methanol	positive	-0.880521	-3.322054e-01	-9.625540e-02	-1.113183e+00	4.984544e-01	-1.025204e-01	1.131925e-02	-6.494213e-04	1.522394e-05
methanol	negative	-0.783086	4.339824e-01	8.018690e-02	-2.754335e-02	-2.137320e-02	-3.271316e-03	-2.241720e-04	-7.379704e-06	-9.514762e-08
methyl acetate	positive	-0.990400	-7.375345e-01	8.307139e-01	-1.457293e+00	5.459602e-01	-1.025452e-01	1.064095e-02	-5.823671e-04	1.314550e-05
methyl acetate	negative	-1.006029	3.007967e-01	-2.576931e-02	1.693091e-03	2.952655e-04	1.635907e-05	4.005673e-07	2.539317e-09	-3.671568e-11
methyl benzoate	positive	-1.069385	-9.451184e-01	1.316253e+00	-1.459625e+00	4.775611e-01	-8.094198e-02	7.683466e-03	-3.881038e-04	8.141188e-06
methyl benzoate	negative	-1.115480	2.424574e-01	-2.168855e-02	1.324431e-03	8.582711e-05	-6.295495e-06	-7.616851e-07	-2.749668e-08	-3.524432e-10
methyl formate	positive	-0.927701	-7.319457e-01	4.677307e-01	-1.055959e+00	3.946789e-01	-7.322337e-02	7.503728e-03	-4.060812e-04	9.076867e-06
methyl formate	negative	-0.939344	3.043316e-01	-3.114216e-02	4.368768e-03	7.015402e-04	4.259785e-05	1.250762e-06	1.488538e-08	6.700833e-12
methyl orthoacetate	positive	-1.066820	-6.453582e-01	4.607156e-01	-1.180006e+00	4.590275e-01	-8.770017e-02	9.201651e-03	-5.078056e-04	1.153995e-05
methyl orthoacetate	negative	-1.072535	2.001047e-01	-1.176029e-02	4.770529e-04	4.043882e-05	6.038933e-07	-2.138585e-08	-7.039011e-10	-3.901699e-12
methyl orthoformate	positive	-1.102422	-5.540768e-01	4.048067e-01	-1.178411e+00	4.686715e-01	-9.078879e-02	9.622540e-03	-5.352500e-04	1.224201e-05
methyl orthoformate	negative	-1.103789	2.176060e-01	-1.372523e-02	5.442458e-04	5.203710e-05	5.928739e-07	-6.468856e-08	-2.622594e-09	-3.051230e-11
methyl propionate	positive	-1.074423	-7.072472e-01	8.379708e-01	-1.494888e+00	5.654666e-01	-1.069264e-01	1.115346e-02	-6.130128e-04	1.388667e-05
methyl propionate	negative	-1.088305	2.268725e-01	-1.844149e-02	8.660906e-04	3.555527e-05	-6.242811e-06	-5.704750e-07	-1.860708e-08	-2.229013e-10
methylene iodide	positive	-1.249582	-1.591452e-01	-1.147435e-02	-8.397279e-04	8.025199e-05	-3.551864e-06	-8.374754e-08	2.727642e-08	-1.155025e-09
methylene iodide	negative	-1.245753	3.935789e-01	-3.557702e-02	3.004724e-02	5.743532e-03	5.043770e-04	2.429562e-05	6.206144e-07	6.589014e-09
methylphenylamine	positive	-1.029213	-6.146947e-01	-3.677326e-02	-2.051831e-01	4.486743e-02	-3.560879e-03	1.165848e-06	1.485829e-05	-5.921148e-07
methylphenylamine	negative	-1.013995	2.906515e-01	1.255444e-02	-3.091975e-04	-6.843370e-03	-1.152867e-03	-8.053216e-05	-2.652228e-06	-3.400489e-08
morpholine	positive	-1.065013	-3.233597e-01	1.826002e-01	-1.593483e+00	6.895045e-01	-1.406989e-01	1.550838e-02	-8.900862e-04	2.088830e-05
morpholine	negative	-1.045191	1.436644e-01	-8.763352e-02	-2.858994e-02	-6.068615e-03	-5.652389e-04	-2.653602e-05	-6.193127e-07	-5.687683e-09
n-butyl iodide	positive	-1.218090	-1.793740e-01	-4.385908e-02	-1.353748e-03	-1.826447e-03	3.776148e-04	-2.576326e-05	5.062859e-07	7.079736e-09
n-butyl iodide	negative	-1.217291	2.752571e-01	-2.390009e-02	1.445329e-03	1.387454e-04	4.646063e-07	-3.685466e-07	-1.644892e-08	-2.293725e-10
n-butylamine	positive	-1.086393	-1.655196e+00	4.810264e+00	-5.155719e+00	1.898327e+00	-3.633082e-01	3.860883e-02	-2.161706e-03	4.979989e-05
n-butylamine	negative	-1.158012	2.007675e-01	-3.565863e-02	-5.794664e-03	-2.773085e-03	-3.726619e-04	-2.299041e-05	-6.912690e-07	-8.240573e-09
n-butyronitrile	positive	-1.010939	-1.196344e+00	2.144540e+00	-1.916072e+00	5.757447e-01	-9.056417e-02	7.967480e-03	-3.712036e-04	7.135328e-06
n-butyronitrile	negative	-1.094825	3.420905e-01	-3.566642e-02	5.970110e-03	9.605474e-04	6.281347e-05	2.134005e-06	3.535612e-08	2.026509e-10

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
n-decane	positive	-1.261883	-1.385071e-02	-9.313237e-04	-5.432894e-05	1.126886e-04	-3.250376e-05	4.226799e-06	-2.662652e-07	6.608045e-09
n-decane	negative	-1.260765	1.302323e-01	-3.815572e-03	2.729618e-04	3.704626e-05	2.859785e-06	1.407869e-07	3.838023e-09	4.366978e-11
n-dodecane	positive	-1.263659	-1.442047e-02	-9.848600e-04	-4.751254e-05	1.151768e-04	-3.362691e-05	4.397141e-06	-2.779805e-07	6.916747e-09
n-dodecane	negative	-1.262632	1.183338e-01	-3.700224e-03	2.630838e-04	3.341730e-05	2.568176e-06	1.278383e-07	3.511573e-09	4.008183e-11
n-heptane	positive	-1.248360	-1.475804e-02	-9.168941e-04	-8.464476e-05	1.222167e-04	-3.371722e-05	4.282788e-06	-2.654427e-07	6.506498e-09
n-heptane	negative	-1.247294	1.308505e-01	-4.120954e-03	2.831038e-04	3.795689e-05	2.920193e-06	1.451721e-07	4.007357e-09	4.611370e-11
n-hexadecane	positive	-1.257667	-1.343249e-02	-1.008666e-03	2.992371e-06	9.532460e-05	-2.996586e-05	4.042742e-06	-2.605498e-07	6.573138e-09
n-hexadecane	negative	-1.256501	1.301472e-01	-3.569590e-03	2.630774e-04	3.599597e-05	2.771214e-06	1.347324e-07	3.622555e-09	4.070947e-11
n-hexane	positive	-1.247730	-1.256684e-02	-8.148470e-04	-5.981254e-05	1.018082e-04	-2.868234e-05	3.682379e-06	-2.299093e-07	5.667254e-09
n-hexane	negative	-1.246739	1.250694e-01	-4.251915e-03	2.843878e-04	3.636050e-05	2.776796e-06	1.397910e-07	3.909035e-09	4.540583e-11
n-octane	positive	-1.263824	-1.303774e-02	-8.517058e-04	-6.147921e-05	1.068538e-04	-3.023026e-05	3.891512e-06	-2.434380e-07	6.009537e-09
n-octane	negative	-1.262736	1.302822e-01	-3.980394e-03	2.781372e-04	3.738224e-05	2.879615e-06	1.425530e-07	3.913090e-09	4.479633e-11
n-pentane	positive	-1.244461	-1.417847e-02	-8.654703e-04	-9.512188e-05	1.226424e-04	-3.343243e-05	4.227167e-06	-2.613291e-07	6.394694e-09
n-pentane	negative	-1.243218	1.550941e-01	-4.255604e-03	2.965100e-04	4.492928e-05	3.522544e-06	1.722568e-07	4.677875e-09	5.328534e-11
n-perfluorohexane	positive	-1.212019	-1.727357e-02	-9.747638e-04	-1.506005e-05	7.694863e-05	-2.144152e-05	2.681471e-06	-1.640965e-07	3.994455e-09
n-perfluorohexane	negative	-1.209537	2.468225e-01	-2.497919e-02	7.770474e-03	8.683302e-04	2.758524e-05	-6.416273e-07	-5.889121e-08	-1.011126e-09
n-propyl acetate	positive	-1.097363	-8.227055e-01	1.362403e+00	-1.875052e+00	6.846227e-01	-1.273547e-01	1.314305e-02	-7.165171e-04	1.612368e-05
n-propyl acetate	negative	-1.125868	2.417006e-01	-1.818952e-02	8.203961e-04	8.514944e-05	5.857631e-07	-1.852366e-07	-8.129636e-09	-1.091068e-10
nitrobenzene	positive	-1.117432	-5.732236e-01	7.213110e-01	-6.841618e-01	1.646596e-01	-1.795554e-02	7.886223e-04	4.944326e-06	-1.014834e-06
nitrobenzene	negative	-1.151388	3.053961e-01	-2.940485e-02	1.133688e-02	1.715646e-03	1.118218e-04	3.754049e-06	5.972460e-08	2.987623e-10
nitroethane	positive	-1.026150	-3.804724e-01	4.269414e-02	-1.286896e-01	-8.091570e-03	9.662730e-03	-1.634340e-03	1.157640e-04	-3.076074e-06
nitroethane	negative	-1.036031	4.547145e-01	-5.664858e-02	6.614125e-03	1.691185e-03	1.551953e-04	7.554083e-06	1.935485e-07	2.058472e-09
nitromethane	positive	-0.945861	-4.395990e-01	3.950244e-01	-4.119649e-01	7.988067e-02	-4.161076e-03	-4.574210e-04	6.419048e-05	-2.172072e-06
nitromethane	negative	-0.974708	5.167804e-01	-1.249680e-01	6.794627e-03	3.292921e-03	3.739307e-04	2.103843e-05	6.025984e-07	7.011115e-09
o-chloroaniline	positive	-1.095673	-2.146554e-01	-3.279741e-01	2.788627e-01	-1.757399e-01	4.340471e-02	-5.271407e-03	3.183299e-04	-7.670349e-06
o-chloroaniline	negative	-1.048256	5.161452e-01	2.310005e-01	1.396727e-01	2.202766e-02	1.788891e-03	8.138609e-05	1.971752e-06	1.984852e-08
ortho-cresol	positive	-1.074933	-3.938452e-01	5.671563e-02	-1.567331e-01	9.719158e-03	5.385538e-03	-1.124200e-03	8.525506e-05	-2.344082e-06
ortho-cresol	negative	-0.910151	1.158160e+00	1.013238e+00	4.347273e-01	6.933722e-02	5.879751e-03	2.808007e-04	7.141744e-06	7.539338e-08

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
ortho-dichlorobenzene	positive	-1.194064	-2.525996e-01	-4.780262e-02	-2.034751e-03	-1.399326e-03	2.493695e-04	-1.085863e-05	-2.637691e-07	2.197884e-08
ortho-dichlorobenzene	negative	-1.191590	2.204322e-01	-2.163471e-02	6.765231e-03	4.038749e-04	-3.156150e-05	-4.205596e-06	-1.649827e-07	-2.271590e-09
ortho-xylene	positive	-1.202962	-2.501994e-01	-9.133702e-02	1.706618e-03	-1.090924e-02	3.305303e-03	-4.115952e-04	2.435211e-05	-5.674088e-07
ortho-xylene	negative	-1.202285	1.903919e-01	-1.307440e-02	5.104806e-04	4.440758e-05	-3.446390e-07	-1.182630e-07	-4.115324e-09	-4.713614e-11
p-chloroacetophenone	positive	-1.058284	-9.769928e-01	1.499755e+00	-1.245132e+00	3.177618e-01	-3.981373e-02	2.490640e-03	-6.207721e-05	-1.242471e-08
p-chloroacetophenone	negative	-1.127488	3.202646e-01	-3.155881e-02	6.151782e-03	7.960378e-04	3.704571e-05	4.664641e-07	-1.611638e-08	-4.224140e-10
p-methoxybenzaldehyde	positive	-0.993576	-1.146541e+00	1.996120e+00	-2.021751e+00	6.624172e-01	-1.135844e-01	1.093045e-02	-5.597777e-04	1.190006e-05
p-methoxybenzaldehyde	negative	-1.058772	3.103022e-01	-2.956512e-02	6.654922e-03	7.574025e-04	2.442721e-05	-5.025233e-07	-4.845904e-08	-8.333954e-10
para-cresol	positive	-1.066021	-4.416642e-01	1.099754e-02	-1.171418e-01	-3.652492e-03	7.698399e-03	-1.343054e-03	9.607383e-05	-2.563203e-06
para-cresol	negative	-0.858866	1.381644e+00	1.334695e+00	5.690902e-01	9.313114e-02	8.132334e-03	3.999667e-04	1.046849e-05	1.136211e-07
para-xylene	positive	-1.189879	-2.645885e-01	-8.932043e-02	2.427598e-03	-9.955986e-03	2.933042e-03	-3.582781e-04	2.086118e-05	-4.793875e-07
para-xylene	negative	-1.189094	2.013898e-01	-1.228366e-02	4.655396e-04	5.479058e-05	1.591382e-06	1.356031e-09	-7.058655e-10	-8.947886e-12
pentachloroethane	positive	-1.218970	-1.284891e-01	-1.022916e-02	1.166758e-03	-5.701988e-04	1.191863e-04	-1.357791e-05	8.114165e-07	-1.980509e-08
pentachloroethane	negative	-1.205156	3.814899e-01	1.125381e-02	3.140289e-02	3.125115e-03	7.948311e-05	-4.399045e-06	-3.009193e-07	-5.005823e-09
pentanoic acid	positive	-1.017759	-9.621483e-01	1.567783e+00	-1.651485e+00	5.356288e-01	-9.044841e-02	8.565298e-03	-4.317703e-04	9.039453e-06
pentanoic acid	negative	-0.884737	8.922008e-01	5.088830e-01	9.953020e-02	-6.518948e-03	-2.474371e-03	-2.094561e-04	-7.627820e-06	-1.045701e-07
perfluoroheptane	positive	-1.208520	-1.832552e-02	-9.401020e-04	-8.775288e-05	1.071132e-04	-2.723805e-05	3.260172e-06	-1.932724e-07	4.582315e-09
perfluoroheptane	negative	-1.206209	2.610699e-01	-2.635795e-02	7.315442e-03	8.594333e-04	3.164170e-05	-2.501429e-07	-4.452367e-08	-8.176282e-10
perfluoromethyl-cyclohexane	positive	-1.162218	-2.050490e-02	-1.076080e-03	-9.291692e-05	1.235702e-04	-3.209161e-05	3.894740e-06	-2.331889e-07	5.567548e-09
perfluoromethyl-cyclohexane	negative	-1.156677	2.646818e-01	-1.527264e-02	1.488046e-02	1.549405e-03	4.533618e-05	-1.541165e-06	-1.222391e-07	-2.073805e-09
perfluorooctane	positive	-1.211180	-1.599063e-02	-8.108330e-04	-7.498639e-05	8.927569e-05	-2.246695e-05	2.670075e-06	-1.575526e-07	3.724818e-09
perfluorooctane	negative	-1.208571	2.567087e-01	-2.511037e-02	8.102079e-03	9.375152e-04	3.371473e-05	-3.526603e-07	-5.175642e-08	-9.383298e-10

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
phenol	positive	-1.050613	-3.943601e-01	-4.345966e-03	-1.127690e-01	1.780657e-03	5.630300e-03	-1.043526e-03	7.601439e-05	-2.044134e-06
phenol	negative	-0.887403	1.252752e+00	1.244587e+00	5.777139e-01	9.909845e-02	8.994252e-03	4.577596e-04	1.235926e-05	1.380229e-07
phenylacetonitrile	positive	-1.060001	-9.842373e-01	1.448481e+00	-1.112254e+00	2.487480e-01	-2.362311e-02	5.228776e-04	5.899528e-05	-3.001951e-06
phenylacetonitrile	negative	-1.132088	3.050456e-01	-2.893715e-02	1.006474e-02	1.289264e-03	6.431205e-05	1.122559e-06	-1.421476e-08	-5.440576e-10
phosphorus oxychloride	positive	-1.069196	-3.501457e-01	2.729722e-01	-2.421178e-01	6.272693e-03	1.117692e-02	-2.163992e-03	1.621651e-04	-4.459359e-06
phosphorus oxychloride	negative	-1.088793	4.276723e-01	-5.602866e-02	1.971040e-02	4.148476e-03	3.756131e-04	1.844350e-05	4.785493e-07	5.152241e-09
piperidine	positive	-1.201797	-3.814479e-01	2.306415e+00	-3.840326e+00	1.558134e+00	-3.143304e-01	3.463379e-02	-1.993005e-03	4.692863e-05
piperidine	negative	-1.163290	1.256808e-01	-5.991825e-02	-1.874677e-02	-4.062219e-03	-3.686156e-04	-1.647376e-05	-3.576127e-07	-2.946827e-09
propanoic acid	positive	-0.944659	-8.545867e-01	1.142825e+00	-1.454787e+00	4.997986e-01	-8.817255e-02	8.685193e-03	-4.543829e-04	9.857305e-06
propanoic acid	negative	-0.781888	9.945068e-01	5.959433e-01	1.425557e-01	2.112669e-03	-1.586364e-03	-1.596720e-04	-6.177788e-06	-8.736267e-08
propionaldehyde	positive	-1.058981	-6.491396e-01	7.070385e-01	-1.199892e+00	4.393788e-01	-8.083831e-02	8.236609e-03	-4.436313e-04	9.874912e-06
propionaldehyde	negative	-1.076723	2.539620e-01	-2.389850e-02	1.785179e-03	9.433778e-05	-1.159230e-05	-1.245208e-06	-4.444860e-08	-5.717427e-10
propionitrile	positive	-0.971017	-1.118543e+00	1.910137e+00	-1.750048e+00	5.252653e-01	-8.227563e-02	7.200994e-03	-3.335705e-04	6.371191e-06
propionitrile	negative	-1.046034	3.847024e-01	-4.231691e-02	7.547445e-03	1.444017e-03	1.134268e-04	4.833159e-06	1.092825e-07	1.029737e-09
pyridine	positive	-1.142033	-1.012114e+00	3.015439e+00	-3.810027e+00	1.450402e+00	-2.819805e-01	3.026311e-02	-1.706532e-03	3.953270e-05
pyridine	negative	-1.163000	2.114445e-01	-2.444976e-02	2.928828e-03	9.541925e-05	-2.708462e-05	-2.621529e-06	-9.315422e-08	-1.210845e-09
pyrimidine	positive	-1.028754	-6.594956e-01	1.119696e+00	-1.781229e+00	6.745341e-01	-1.284017e-01	1.348460e-02	-7.456310e-04	1.697947e-05
pyrimidine	negative	-1.042513	2.507960e-01	-2.498707e-02	5.397300e-03	1.824233e-04	-4.674111e-05	-4.708746e-06	-1.719945e-07	-2.286469e-09
pyrrole	positive	-0.904733	-6.649413e-01	-1.129687e-01	-2.968594e-01	1.162258e-01	-2.098714e-02	2.071462e-03	-1.080555e-04	2.336071e-06
pyrrole	negative	-0.759885	9.288828e-01	7.057711e-01	2.791028e-01	3.906776e-02	2.841171e-03	1.133545e-04	2.322809e-06	1.868959e-08
pyrrolidine	positive	-1.254375	2.412446e-02	1.716297e+00	-3.700597e+00	1.561006e+00	-3.211376e-01	3.584237e-02	-2.082134e-03	4.938889e-05
pyrrolidine	negative	-1.172113	1.481863e-01	-5.275195e-02	-1.538912e-02	-3.342220e-03	-2.954141e-04	-1.258794e-05	-2.526508e-07	-1.802850e-09
quinoline	positive	-1.063293	-1.627068e+00	3.688409e+00	-3.789140e+00	1.342421e+00	-2.486504e-01	2.570059e-02	-1.405238e-03	3.171811e-05
quinoline	negative	-1.148617	2.058809e-01	-2.359541e-02	2.128753e-03	1.178580e-05	-2.853630e-05	-2.431474e-06	-8.235088e-08	-1.040021e-09
styrene	positive	-1.168415	-3.241274e-01	-7.149878e-02	-5.076476e-03	-1.527381e-03	7.086475e-04	-9.278383e-05	5.441156e-06	-1.237551e-07
styrene	negative	-1.168016	2.150650e-01	-1.842078e-02	9.472320e-04	8.158186e-05	-2.343546e-06	-4.203003e-07	-1.588231e-08	-2.053667e-10

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
sulfolane	positive	-0.864932	-9.985330e-01	2.314000e+00	-2.811392e+00	1.025938e+00	-1.927044e-01	2.010938e-02	-1.107988e-03	2.517378e-05
sulfolane	negative	-0.905912	4.043703e-01	-5.203194e-02	8.189524e-03	1.949117e-03	1.759585e-04	8.469018e-06	2.148905e-07	2.265493e-09
sulfur dioxide	positive	-1.050420	-3.241477e-01	-6.411350e-02	-5.967235e-02	1.218686e-03	2.973482e-03	-5.464092e-04	3.925723e-05	-1.041342e-06
sulfur dioxide	negative	-1.059361	2.841481e-01	-3.180295e-02	8.513194e-02	1.847605e-02	1.840781e-03	9.895576e-05	2.777684e-06	3.198501e-08
sulfuric acid	positive	-0.806814	-3.162963e-01	-1.643550e-01	4.628932e-02	-4.327584e-02	1.135411e-02	-1.371358e-03	8.098549e-05	-1.901295e-06
sulfuric acid	negative	-1.140228	-8.441434e-01	-2.260254e-01	4.021054e-01	1.015318e-01	1.126075e-02	6.576830e-04	1.972362e-05	2.396853e-07
tetrachloroethylene	positive	-1.265246	-4.933602e-02	-6.017960e-03	2.261312e-03	-6.421346e-04	1.022174e-04	-9.415684e-06	4.693270e-07	-9.826763e-09
tetrachloroethylene	negative	-1.264034	2.766013e-01	-2.720559e-02	3.338681e-03	3.132256e-04	-5.160856e-07	-1.085564e-06	-4.900120e-08	-7.015330e-10
tetraethylurea	positive	-1.152079	-7.257795e-01	2.145393e+00	-3.305226e+00	1.295222e+00	-2.551379e-01	2.761827e-02	-1.567599e-03	3.650815e-05
tetraethylurea	negative	-1.143592	1.987939e-01	-1.357668e-02	5.411274e-04	3.140123e-05	-1.811803e-06	-1.857370e-07	-5.618865e-09	-6.052000e-11
tetrahydrofuran	positive	-1.138680	7.411591e-02	-5.839523e-01	-1.213435e+00	6.024727e-01	-1.299644e-01	1.480481e-02	-8.691612e-04	2.074313e-05
tetrahydrofuran	negative	-1.073398	2.334199e-01	-1.243114e-02	4.555848e-04	6.755023e-05	3.439403e-06	1.069819e-07	2.183149e-09	2.262727e-11
tetrahydropyran	positive	-1.156291	-3.633272e-01	4.198692e-01	-1.589446e+00	6.673391e-01	-1.343872e-01	1.468445e-02	-8.372644e-04	1.954478e-05
tetrahydropyran	negative	-1.135436	2.116226e-01	-8.625014e-03	3.547241e-04	6.080412e-05	4.214465e-06	1.873008e-07	4.920483e-09	5.632119e-11
tetrahydrothiophene	positive	-1.167220	-5.606969e-01	5.507575e-01	-6.700491e-01	1.926391e-01	-2.769238e-02	2.166409e-03	-8.689358e-05	1.367285e-06
tetrahydrothiophene	negative	-1.193194	2.369677e-01	-1.677691e-02	6.909825e-04	9.429201e-05	2.706574e-06	-4.724007e-08	-4.026573e-09	-6.146934e-11
tetramethylsilane	positive	-1.225832	-5.677245e-02	-4.234084e-03	4.351291e-04	-2.466649e-04	5.358466e-05	-6.342535e-06	3.895435e-07	-9.694918e-09
tetramethylsilane	negative	-1.225238	1.781149e-01	-8.012929e-03	3.542535e-04	5.023740e-05	3.177720e-06	1.405160e-07	3.831639e-09	4.567172e-11
tetramethylurea	positive	-1.162390	1.124485e-01	3.258999e-01	-2.373863e+00	1.065478e+00	-2.237074e-01	2.520874e-02	-1.472352e-03	3.504427e-05
tetramethylurea	negative	-1.074158	2.661570e-01	-1.718086e-02	7.151529e-04	7.082209e-05	2.257286e-07	-1.590005e-07	-6.497953e-09	-8.345030e-11
thiane	positive	-1.154776	-5.374659e-01	5.418095e-01	-6.187414e-01	1.661410e-01	-2.170689e-02	1.464814e-03	-4.502039e-05	3.589128e-07
thiane	negative	-1.181461	2.417068e-01	-1.411261e-02	5.077092e-04	7.783287e-05	3.485494e-06	7.430399e-08	6.818353e-10	1.408666e-12
thiobis(2-ethanol)	positive	-0.774729	-5.675370e-01	-3.600620e-01	-6.650358e-01	3.069092e-01	-6.322363e-02	6.965520e-03	-3.983851e-04	9.307409e-06
thiobis(2-ethanol)	negative	-0.679025	4.887397e-01	8.047651e-02	-3.558088e-02	-2.443836e-02	-3.690201e-03	-2.518825e-04	-8.281001e-06	-1.067313e-07
thionyl chloride	positive	-1.196896	-1.457694e-01	-1.740513e-02	-4.157825e-03	7.822414e-04	-2.748970e-04	3.747659e-05	-2.106553e-06	4.248102e-08
thionyl chloride	negative	-1.212708	4.155181e-01	-1.509065e-01	4.381497e-02	1.200930e-02	1.283268e-03	7.156304e-05	2.057370e-06	2.410645e-08
toluene	positive	-1.177786	-3.206211e-01	-8.500471e-02	-6.514334e-03	-3.900884e-03	1.551857e-03	-2.055472e-04	1.243243e-05	-2.923964e-07
toluene	negative	-1.177366	2.021789e-01	-1.457605e-02	5.884702e-04	5.304883e-05	-6.025115e-07	-1.708759e-07	-6.234634e-09	-7.582273e-11

Continued on next page

		Coefficient Order								
Molecule Name	ε Region	0th	1th	2th	3th	4th	5th	6th	7th	8th
trans-1,2-dichloroethylene	positive	-1.238128	-1.460258e-01	-9.831067e-03	-7.299529e-04	7.049860e-05	-2.304694e-06	-2.891648e-07	4.206150e-08	-1.536739e-09
trans-1,2-dichloroethylene	negative	-1.229822	3.218240e-01	-9.362466e-04	2.365649e-02	2.922118e-03	1.485955e-04	2.626288e-06	-3.611476e-08	-1.369085e-09
tri-(n-butyl)amine	positive	-1.120097	-1.577209e+00	3.700478e+00	-3.554637e+00	1.217473e+00	-2.196364e-01	2.220064e-02	-1.190522e-03	2.641501e-05
tri-(n-butyl)amine	negative	-1.222476	1.383496e-01	-4.139321e-03	2.656429e-04	3.524566e-05	2.634486e-06	1.271692e-07	3.426492e-09	3.865181e-11
tri-n-butylphosphate	positive	-1.246060	-1.541996e-02	1.212441e+00	-3.323946e+00	1.421967e+00	-2.935713e-01	3.280666e-02	-1.906751e-03	4.523847e-05
tri-n-butylphosphate	negative	-1.155083	1.937902e-01	-1.146366e-02	4.207604e-04	2.859326e-05	-1.961943e-07	-5.209675e-08	-1.394532e-09	-1.101487e-11
trichloroethylene	positive	-1.261510	-1.158278e-01	-9.386629e-03	1.293881e-03	-6.370444e-04	1.338099e-04	-1.548641e-05	9.374645e-07	-2.309828e-08
trichloroethylene	negative	-1.251590	2.885649e-01	1.237805e-04	1.751333e-02	6.214669e-04	-1.405913e-04	-1.502310e-05	-5.714384e-07	-7.853346e-09
triethanolamine	positive	-0.861866	-5.886552e-01	2.163389e-01	-1.226416e+00	5.029007e-01	-9.919250e-02	1.066781e-02	-6.007949e-04	1.388811e-05
triethanolamine	negative	-0.769299	4.158711e-01	1.146885e-01	-2.220269e-02	-2.182723e-02	-3.413311e-03	-2.358407e-04	-7.801522e-06	-1.009321e-07
triethylamine	positive	-1.231359	-3.802267e-01	1.666402e+00	-3.065110e+00	1.248356e+00	-2.512368e-01	2.759115e-02	-1.582701e-03	3.716152e-05
triethylamine	negative	-1.195859	1.440614e-01	-5.544946e-03	2.880718e-04	2.844419e-05	1.843024e-06	9.037126e-08	2.546438e-09	2.963325e-11
triethylene glycol	positive	-0.918207	-6.397418e-01	4.264327e-01	-9.895689e-01	3.706336e-01	-6.880861e-02	7.053106e-03	-3.817234e-04	8.532284e-06
triethylene glycol	negative	-0.800466	7.694606e-01	3.772224e-01	9.164553e-02	-8.517597e-04	-1.399950e-03	-1.293507e-04	-4.850443e-06	-6.739136e-08
triethylphosphate	positive	-1.131740	3.307742e-02	5.722574e-01	-2.801361e+00	1.226702e+00	-2.546562e-01	2.851033e-02	-1.658312e-03	3.935937e-05
triethylphosphate	negative	-1.039532	2.556713e-01	-1.838064e-02	7.975741e-04	1.128416e-04	3.718048e-06	-2.691281e-08	-4.124171e-09	-6.791830e-11
trifluoroacetic acid	positive	-1.062027	-1.894633e-02	-3.868571e-01	2.617650e-01	-9.456997e-02	1.407892e-02	-9.102512e-04	1.721765e-05	3.260146e-07
trifluoroacetic acid	negative	-1.240088	-2.185352e-01	2.407572e-02	3.997350e-01	9.321194e-02	9.993510e-03	5.715851e-04	1.688954e-05	2.029516e-07
trimethylphosphate	positive	-1.044617	7.056863e-02	2.106452e-01	-2.475752e+00	1.119010e+00	-2.353031e-01	2.652894e-02	-1.549848e-03	3.689521e-05
trimethylphosphate	negative	-0.947540	2.572889e-01	-2.577707e-02	2.701834e-03	3.182245e-04	8.141346e-06	-3.682166e-07	-2.461563e-08	-3.900022e-10
water	positive	0.838237	-5.190700e-01	-7.714317e-02	-8.138053e-01	3.524312e-01	-7.057004e-02	7.625482e-03	-4.298967e-04	9.933368e-06
water	negative	0.775179	1.258973e-02	-3.723605e-01	4.109847e-02	1.733447e-02	2.097013e-03	1.256949e-04	3.799720e-06	4.623490e-08

Table G.2 Polynomial coefficients for pure solvents

Ethanol content/ % volume	ε Region	Coefficient Order 0th	1th	2th	3th	4th	5th	6th	7th	8th
0	positive	0.838237	-0.519070	-0.077143	-0.813805	0.352431	-0.070570	0.007625	-4.298967e-04	9.933368e-06
0	negative	0.775179	0.012590	-0.372360	0.041098	0.017334	0.002097	0.000126	3.799720e-06	4.623490e-08
5	positive	0.654950	-0.532589	0.079069	-0.910200	0.379728	-0.074879	0.008015	-4.488384e-04	1.031725e-05
5	negative	0.598293	0.077947	-0.273272	0.071380	0.021855	0.002478	0.000144	4.281484e-06	5.147507e-08
10	positive	0.499760	-0.547455	0.205787	-0.986888	0.401001	-0.078166	0.008306	-4.626007e-04	1.058870e-05
10	negative	0.450298	0.139727	-0.182671	0.098447	0.025811	0.002804	0.000160	4.679731e-06	5.573467e-08
15	positive	0.364385	-0.562506	0.312642	-1.050794	0.418464	-0.080819	0.008536	-4.732653e-04	1.079385e-05
15	negative	0.322818	0.199623	-0.098298	0.123002	0.029310	0.003086	0.000173	5.007328e-06	5.915841e-08
20	positive	0.243752	-0.577280	0.405301	-1.105932	0.433387	-0.083061	0.008728	-4.820070e-04	1.095863e-05
20	negative	0.210712	0.258451	-0.018834	0.145451	0.032414	0.003329	0.000184	5.271156e-06	6.182262e-08
25	positive	0.134535	-0.591620	0.487429	-1.154881	0.446584	-0.085034	0.008895	-4.895783e-04	1.109968e-05
25	negative	0.110599	0.316623	0.056515	0.166022	0.035155	0.003534	0.000193	5.474194e-06	6.375881e-08
30	positive	0.034427	-0.605494	0.561536	-1.199422	0.458621	-0.086836	0.009048	-4.965090e-04	1.122878e-05
30	negative	0.020123	0.374318	0.128189	0.184818	0.037543	0.003703	0.000199	5.616475e-06	6.496435e-08
35	positive	-0.058241	-0.618918	0.629419	-1.240867	0.469920	-0.088542	0.009195	-5.032066e-04	1.135508e-05
35	negative	-0.062430	0.431541	0.196357	0.201842	0.039573	0.003835	0.000204	5.695502e-06	6.540719e-08
40	positive	-0.144718	-0.631916	0.692412	-1.280240	0.480823	-0.090214	0.009341	-5.100141e-04	1.148636e-05
40	negative	-0.138361	0.488142	0.260949	0.217004	0.041219	0.003926	0.000206	5.706328e-06	6.502671e-08
45	positive	-0.225968	-0.644511	0.751530	-1.318394	0.491625	-0.091906	0.009492	-5.172487e-04	1.162992e-05
45	negative	-0.208697	0.543796	0.321644	0.230114	0.042442	0.003973	0.000206	5.641384e-06	6.373174e-08
50	positive	-0.302749	-0.656703	0.807556	-1.356083	0.502604	-0.093672	0.009655	-5.252321e-04	1.179324e-05
50	negative	-0.274289	0.597958	0.377838	0.240871	0.043180	0.003969	0.000203	5.490090e-06	6.139618e-08
55	positive	-0.375677	-0.668459	0.861086	-1.394014	0.514043	-0.095568	0.009834	-5.343188e-04	1.198468e-05
55	negative	-0.335883	0.649788	0.428567	0.248831	0.043346	0.003905	0.000196	5.238236e-06	5.785208e-08
60	positive	-0.445258	-0.679688	0.912536	-1.432882	0.526245	-0.097658	0.010038	-5.449243e-04	1.221413e-05
60	negative	-0.394187	0.698028	0.472393	0.253371	0.042821	0.003769	0.000186	4.867135e-06	5.288032e-08
65	positive	-0.511926	-0.690202	0.962106	-1.473380	0.539548	-0.100016	0.010274	-5.575566e-04	1.249383e-05

Continued on next page

Ethanol content/ % volume	ε Region	Coefficient Order 0th	1th	2th	3th	4th	5th	6th	7th	8th
65	negative	-0.449936	0.740819	0.507232	0.253623	0.041445	0.003546	0.000170	4.352533e-06	4.619910e-08
70	positive	-0.576063	-0.699649	1.009678	-1.516175	0.554328	-0.102728	0.010554	-5.728451e-04	1.283915e-05
70	negative	-0.503970	0.775423	0.530098	0.248396	0.039001	0.003214	0.000149	3.663412e-06	3.745199e-08
75	positive	-0.638024	-0.707399	1.054600	-1.561817	0.570984	-0.105896	0.010889	-5.915583e-04	1.326912e-05
75	negative	-0.557348	0.797821	0.536757	0.236063	0.035204	0.002748	0.000120	2.761015e-06	2.620038e-08
80	positive	-0.698166	-0.712361	1.095313	-1.610515	0.589882	-0.109624	0.011294	-6.145811e-04	1.380618e-05
80	negative	-0.611492	0.802157	0.521274	0.214447	0.029684	0.002112	0.000082	1.599148e-06	1.193395e-08
85	positive	-0.756870	-0.712724	1.128735	-1.661706	0.611215	-0.113999	0.011781	-6.427944e-04	1.447356e-05
85	negative	-0.668386	0.780044	0.475550	0.180730	0.021985	0.001268	0.000033	1.282700e-07	-5.869163e-09
90	positive	-0.814571	-0.705662	1.149455	-1.713323	0.634731	-0.119033	0.012356	-6.767644e-04	1.528817e-05
90	negative	-0.730802	0.719954	0.389114	0.131531	0.011603	0.000175	-0.000030	-1.691163e-06	-2.756494e-08
95	positive	-0.871752	-0.687243	1.149123	-1.760901	0.659338	-0.124587	0.013011	-7.161628e-04	1.624665e-05
95	negative	-0.802440	0.607301	0.249897	0.063452	-0.001879	-0.001190	-0.000105	-3.855694e-06	-5.292259e-08
100	positive	-0.928881	-0.652997	1.117066	-1.797165	0.682763	-0.130270	0.013706	-7.590500e-04	1.730739e-05
100	negative	-0.887748	0.426468	0.047267	-0.025426	-0.018460	-0.002799	-0.000191	-6.248878e-06	-8.025020e-08

Table G.3 Coefficients for water-ethanol mixture solvation profiles.

Chloroform content/ % volume	ε Region	Coefficient Order 0th	1th	2th	3th	4th	5th	6th	7th	8th
0	positive	-1.138680	0.074116	-0.583952	-1.213435	0.602473	-0.129964	1.480481e-02	-8.691612e-04	2.074313e-05
0	negative	-1.073398	0.233420	-0.012431	0.000456	0.000068	0.000003	1.069819e-07	2.183149e-09	2.262727e-11
5	positive	-1.135735	-0.008263	-0.296479	-1.395806	0.658249	-0.139526	1.574416e-02	-9.186379e-04	2.182741e-05
5	negative	-1.040285	0.404232	0.163705	0.072443	0.014343	0.001492	8.092628e-05	2.182120e-06	2.312460e-08
10	positive	-1.131570	-0.103908	0.013727	-1.588558	0.716242	-0.149323	1.669406e-02	-9.680627e-04	2.289835e-05
10	negative	-1.054326	0.373315	0.152180	0.076513	0.017074	0.002018	1.248410e-04	3.860084e-06	4.732257e-08
15	positive	-1.126023	-0.214286	0.347950	-1.791451	0.776104	-0.159256	1.764081e-02	-1.016533e-03	2.393253e-05
15	negative	-1.079849	0.278797	0.064583	0.044874	0.011659	0.001568	1.063241e-04	3.522819e-06	4.564648e-08
20	positive	-1.118917	-0.340898	0.707081	-2.003695	0.837239	-0.169167	1.856405e-02	-1.062741e-03	2.489665e-05
20	negative	-1.103485	0.186783	-0.027996	0.008046	0.004523	0.000862	6.914949e-05	2.525688e-06	3.490824e-08
25	positive	-1.110060	-0.485168	1.091258	-2.223631	0.898670	-0.178816	1.943388e-02	-1.104807e-03	2.574331e-05
25	negative	-1.122706	0.111961	-0.107963	-0.025706	-0.002463	0.000123	2.750504e-05	1.326936e-06	2.100657e-08
30	positive	-1.099267	-0.648226	1.499270	-2.448221	0.958854	-0.187844	2.020655e-02	-1.140023e-03	2.640517e-05
30	negative	-1.137138	0.058008	-0.169957	-0.053515	-0.008599	-0.000561	-1.276263e-05	1.211658e-07	6.507742e-09
35	positive	-1.086379	-0.830517	1.927570	-2.672273	1.015393	-0.195715	2.081830e-02	-1.164501e-03	2.678648e-05
35	negative	-1.146933	0.025710	-0.212093	-0.074140	-0.013552	-0.001145	-4.868446e-05	-9.907522e-07	-7.236947e-09
40	positive	-1.071330	-1.031105	2.368677	-2.887246	1.064613	-0.201636	2.117670e-02	-1.172686e-03	2.675189e-05
40	negative	-1.152391	0.015011	-0.233587	-0.086871	-0.017121	-0.001604	-7.837625e-05	-1.944385e-06	-1.935999e-08
45	positive	-1.054253	-1.246474	2.808686	-3.079518	1.100985	-0.204448	2.114956e-02	-1.156745e-03	2.611252e-05
45	negative	-1.153914	0.025434	-0.234163	-0.091210	-0.019151	-0.001915	-1.003717e-04	-2.689645e-06	-2.918801e-08
50	positive	-1.035662	-1.468640	3.223688	-3.228135	1.116422	-0.202510	2.055323e-02	-1.105959e-03	2.461306e-05
50	negative	-1.152081	0.055833	-0.214203	-0.086876	-0.019524	-0.002061	-1.134262e-04	-3.183437e-06	-3.613755e-08
55	positive	-1.016741	-1.682544	3.575531	-3.302572	1.099718	-0.193617	1.914700e-02	-1.006539e-03	2.192987e-05
55	negative	-1.147788	0.103741	-0.175393	-0.074038	-0.018194	-0.002031	-1.166337e-04	-3.391995e-06	-3.973164e-08
60	positive	-0.999662	-1.863501	3.808986	-3.262231	1.036751	-0.175100	1.664839e-02	-8.427161e-04	1.770106e-05
60	negative	-1.142375	0.164563	-0.121626	-0.053703	-0.015264	-0.001829	-1.098284e-04	-3.302189e-06	-3.972984e-08
65	positive	-0.987756	-1.976816	3.854584	-3.060492	0.912404	-0.144267	1.278678e-02	-6.000822e-04	1.160997e-05

Continued on next page

Chloroform content/ % volume	ε Region	Coefficient Order 0th	1th	2th	3th	4th	5th	6th	7th	8th
65	negative	-1.137583	0.231279	-0.059545	-0.028068	-0.011064	-0.001480	-9.412279e-05	-2.937966e-06	-3.633065e-08
70	positive	-0.985236	-1.981977	3.640284	-2.654337	0.714243	-0.099151	7.388729e-03	-2.706336e-04	3.507033e-06
70	negative	-1.135178	0.295581	0.002322	-0.000368	-0.006154	-0.001035	-7.212123e-05	-2.368672e-06	-3.029114e-08
75	positive	-0.996514	-1.839841	3.105922	-2.013736	0.435625	-0.039097	4.395134e-04	1.437534e-04	-6.509407e-06
75	negative	-1.136357	0.350427	0.056105	0.026012	-0.001176	-0.000559	-4.731553e-05	-1.693684e-06	-2.276607e-08
80	positive	-1.025869	-1.515277	2.205410	-1.120888	0.075089	0.035385	-7.932777e-03	6.322191e-04	-1.811334e-05
80	negative	-1.141392	0.392213	0.097102	0.048637	0.003353	-0.000108	-2.290759e-05	-1.008407e-06	-1.490737e-08
85	positive	-1.078338	-0.968429	0.888593	0.042319	-0.367295	0.123153	-1.749720e-02	1.176290e-03	-3.076435e-05
85	negative	-1.149791	0.420873	0.124324	0.066477	0.007158	0.000287	-9.681064e-07	-3.779286e-07	-7.533212e-09
90	positive	-1.162540	-0.132734	-0.929945	1.505584	-0.889252	0.221460	-2.773559e-02	1.735440e-03	-4.329051e-05
90	negative	-1.160726	0.438447	0.139272	0.079604	0.010179	0.000613	1.762581e-05	1.670322e-07	-1.057960e-09
95	positive	-1.296888	1.110901	-3.349879	3.233449	-1.435572	0.312028	-3.594985e-02	2.120559e-03	-5.055678e-05
95	negative	-1.173376	0.447598	0.144418	0.088657	0.012480	0.000872	3.280372e-05	6.201256e-07	4.401588e-09
100	positive	-1.220126	-0.166261	-0.009943	-0.000467	-0.000003	0.000015	-2.411077e-06	1.740120e-07	-4.822481e-09
100	negative	-1.187073	0.450768	0.142246	0.094427	0.014172	0.001073	4.489395e-05	9.877984e-07	8.891901e-09

Table G.4 Coefficients for chloroform-tetrahydrofuran mixture solvation profiles.

Appendix H

FGIP plots for solvents

Plots of functional group interaction profiles for solvents studied are in the following sections. The solute-solute interactions are favourable when negative (blue), and unfavourable when positive (red).

H.1 Pure solvent FGIPs

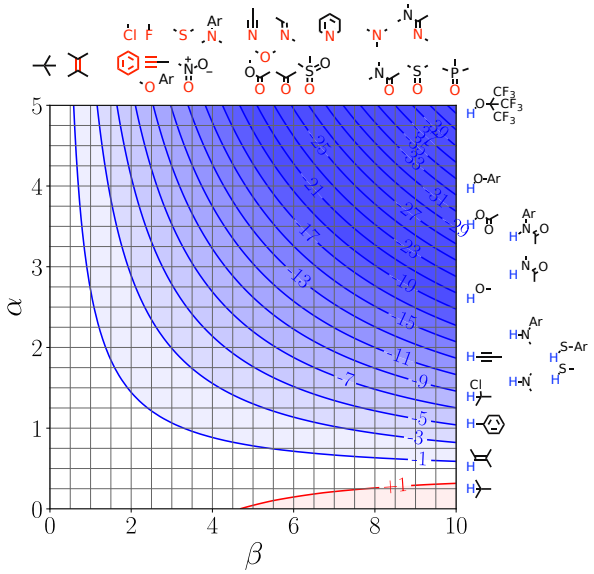


Fig. H.1: FGIP for tetramethylsilane at 298K.

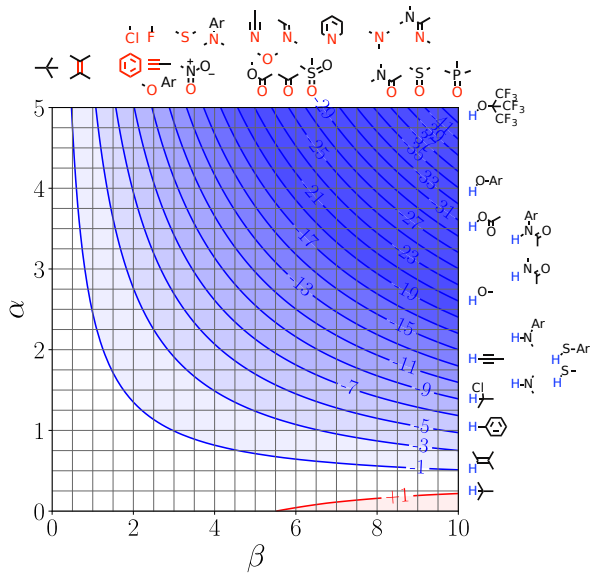


Fig. H.2: FGIP for n-pentane at 298K.

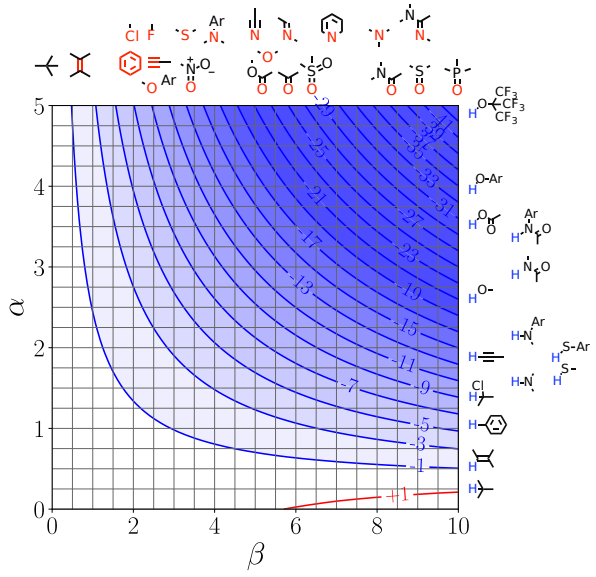


Fig. H.3: FGIP for 2-methylbutane at 298K.

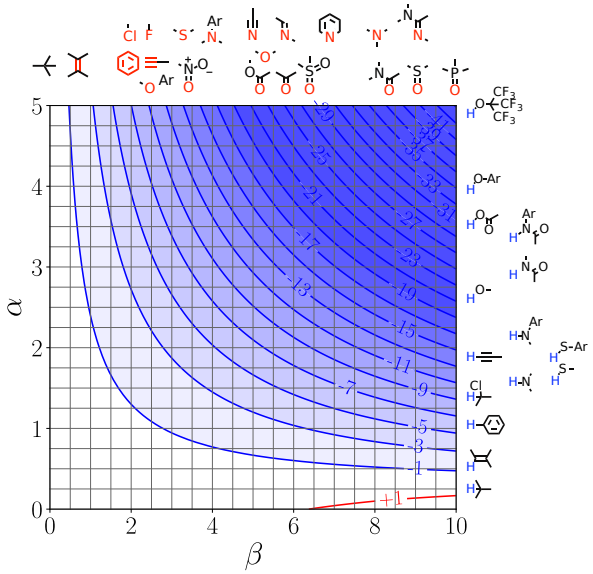


Fig. H.4: FGIP for n-hexane at 298K.

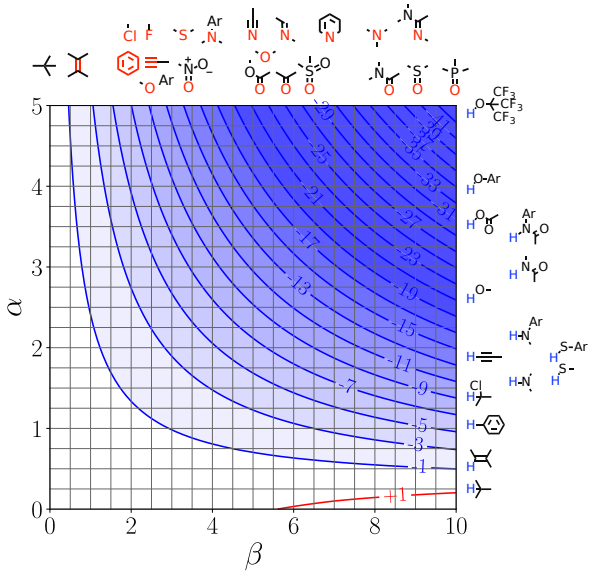


Fig. H.5: FGIP for cyclohexane at 298K.

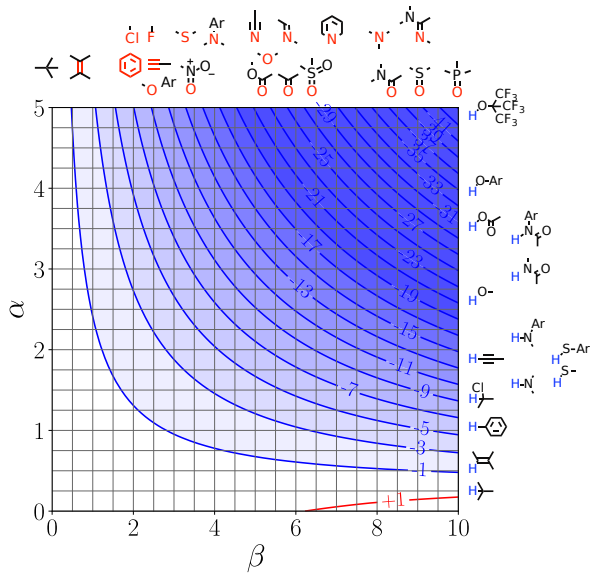


Fig. H.6: FGIP for n-heptane at 298K.

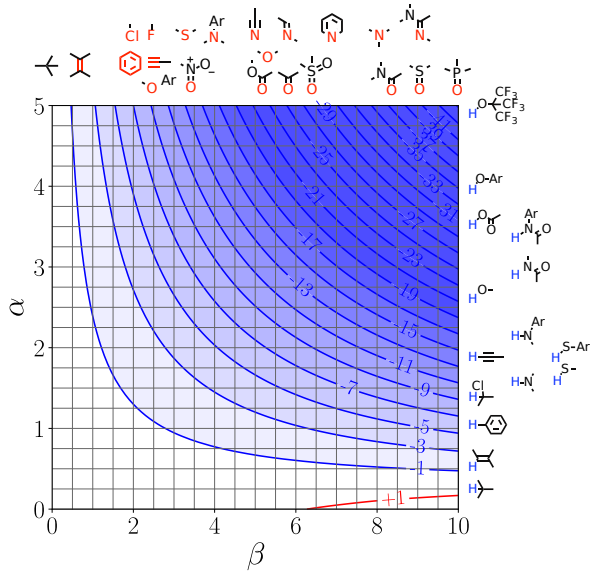


Fig. H.7: FGIP for n-octane at 298K.

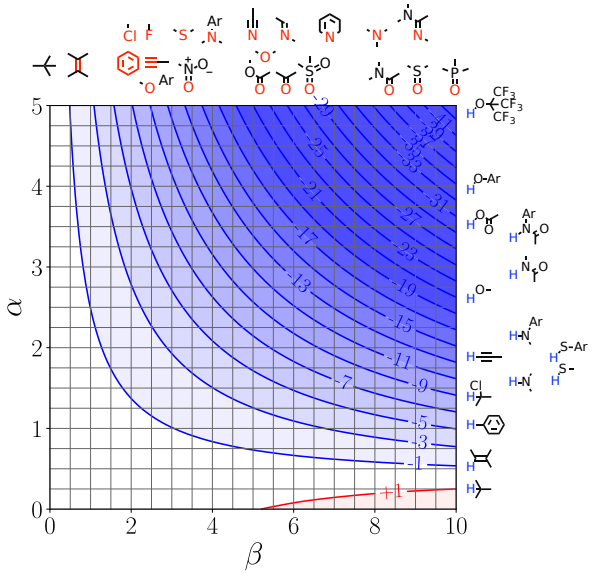


Fig. H.8: FGIP for 2,2,4-trimethylpentane at 298K.

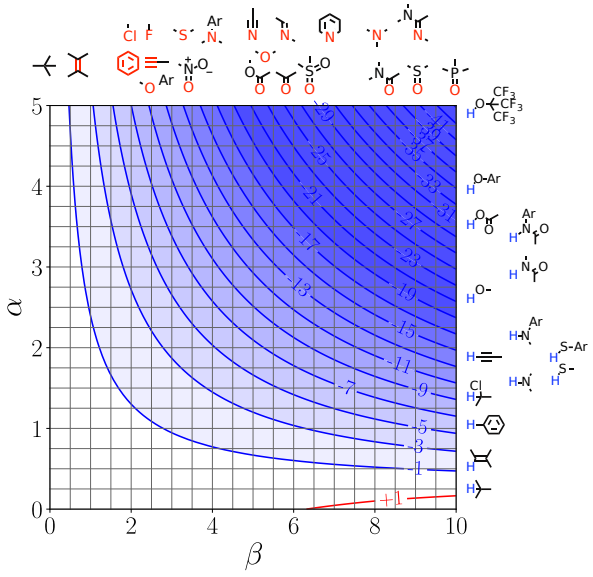


Fig. H.9: FGIP for n-decane at 298K.

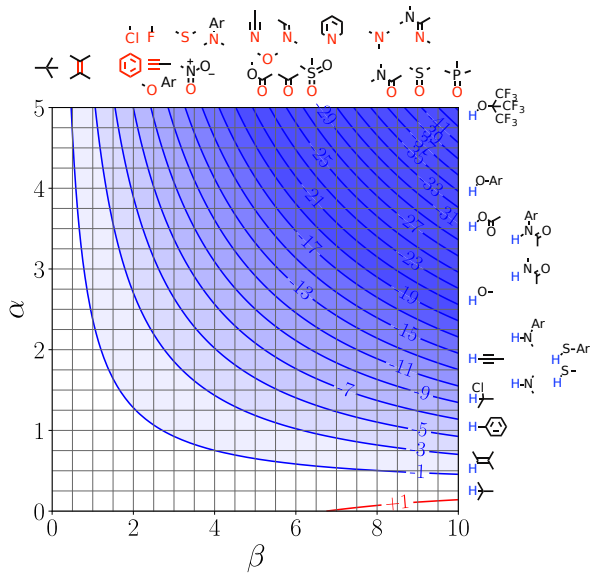


Fig. H.10: FGIP for n-dodecane at 298K.

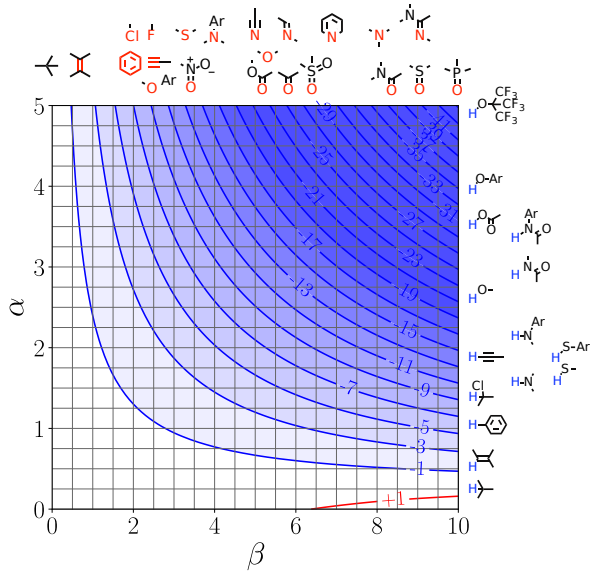


Fig. H.11: FGIP for n-hexadecane at 298K.

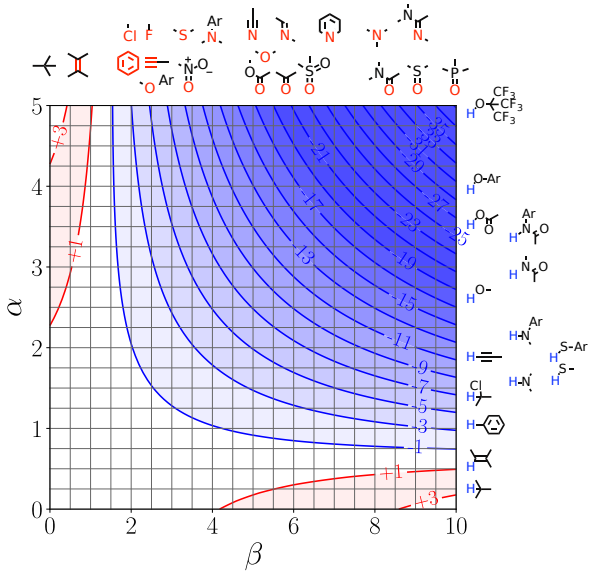


Fig. H.12: FGIP for benzene at 298K.

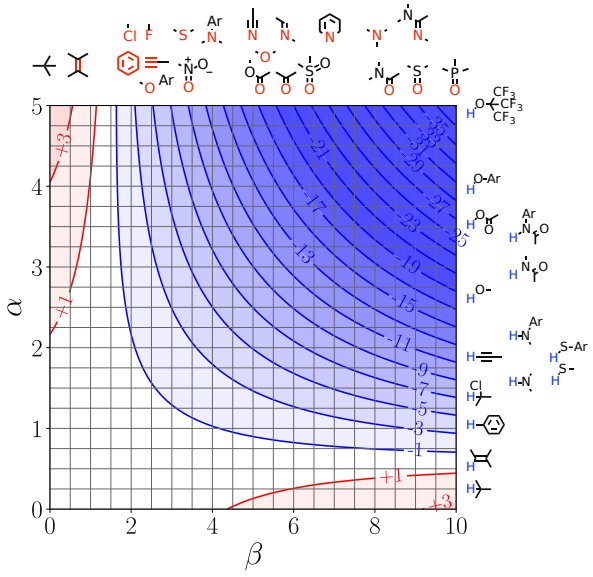


Fig. H.13: FGIP for toluene at 298K.

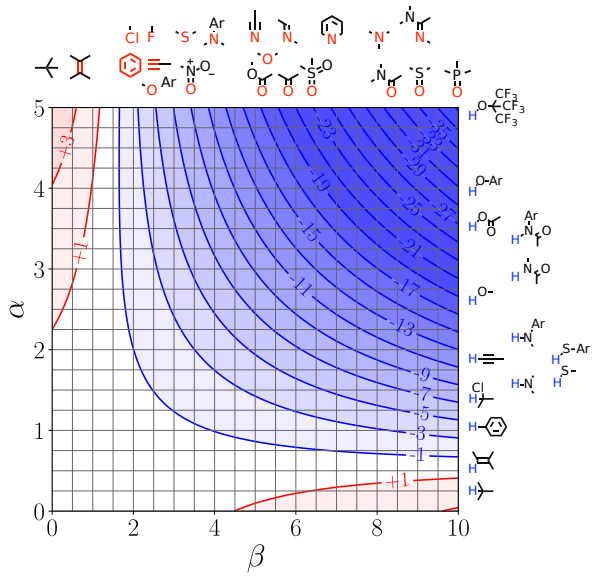


Fig. H.14: FGIP for ortho-xylene at 298K.

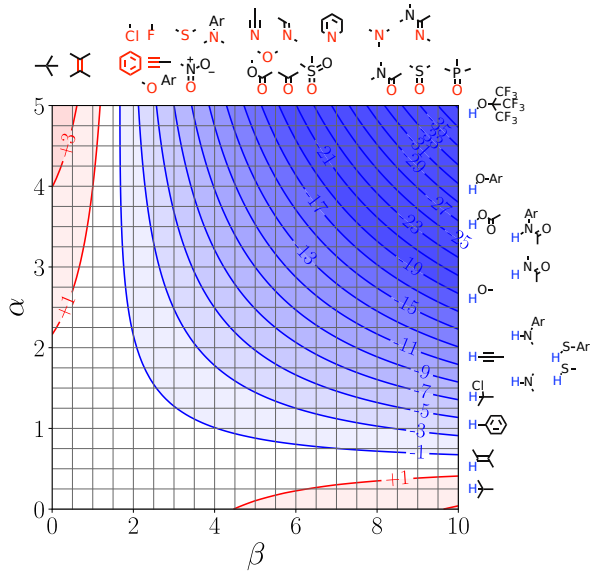


Fig. H.15: FGIP for meta-xylene at 298K.

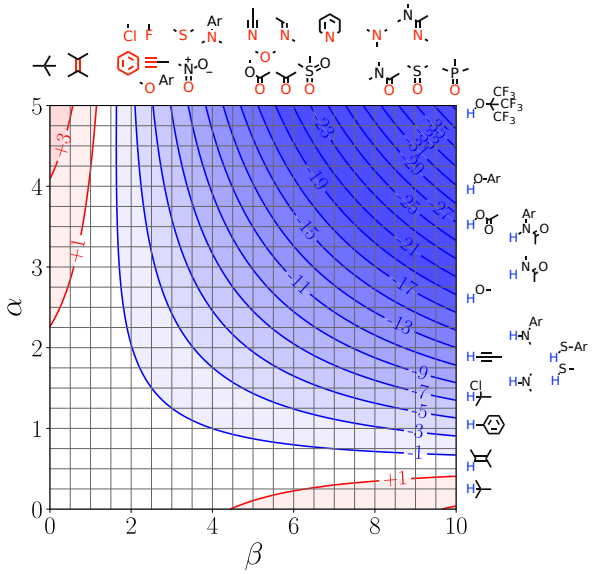


Fig. H.16: FGIP for para-xylene at 298K.

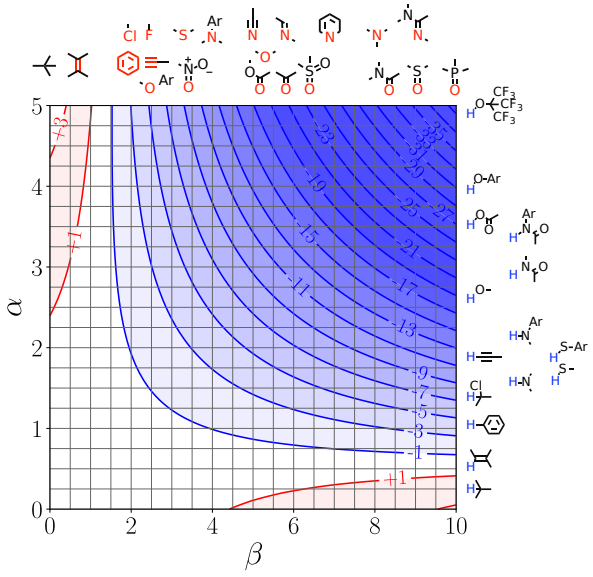


Fig. H.17: FGIP for ethylbenzene at 298K.

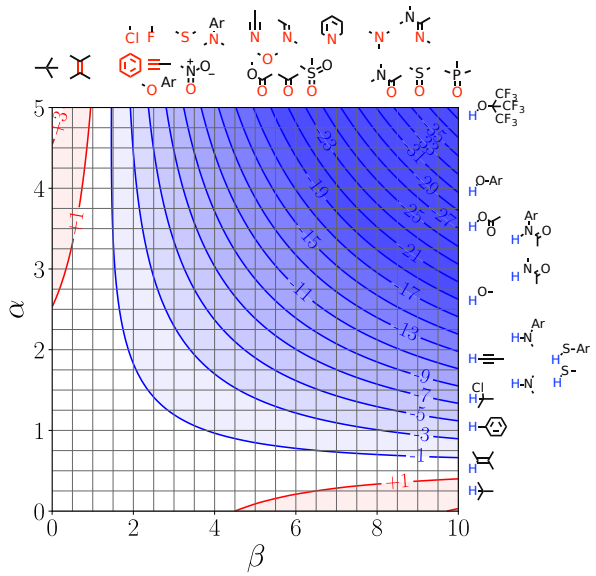


Fig. H.18: FGIP for isopropylbenzene at 298K.

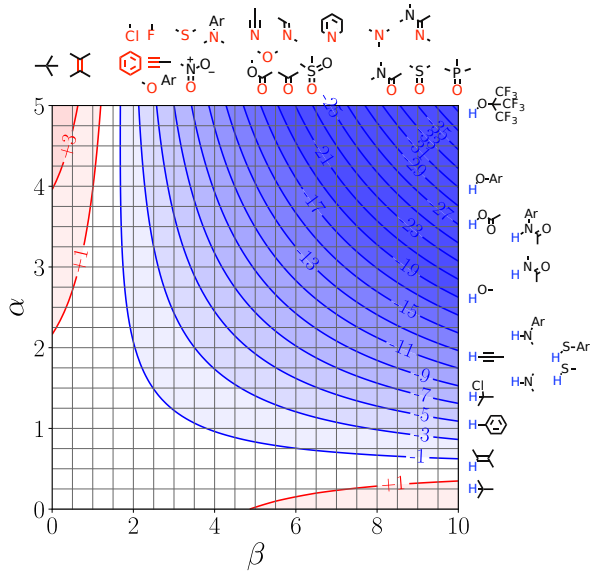


Fig. H.19: FGIP for 1,3,5-trimethylbenzene at 298K.

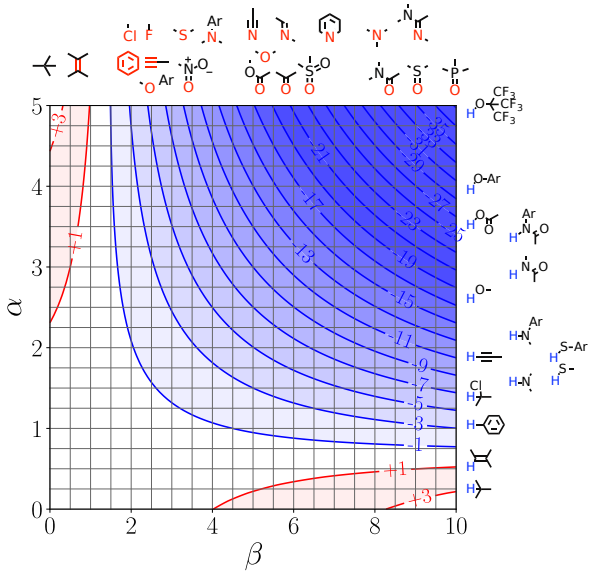


Fig. H.20: FGIP for styrene at 298K.

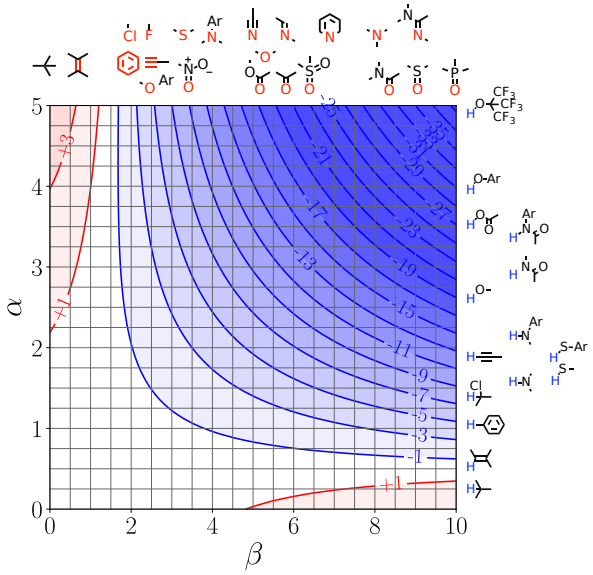


Fig. H.21: FGIP for 1,2,3,4-tetrahydronaphthalene at 298K.

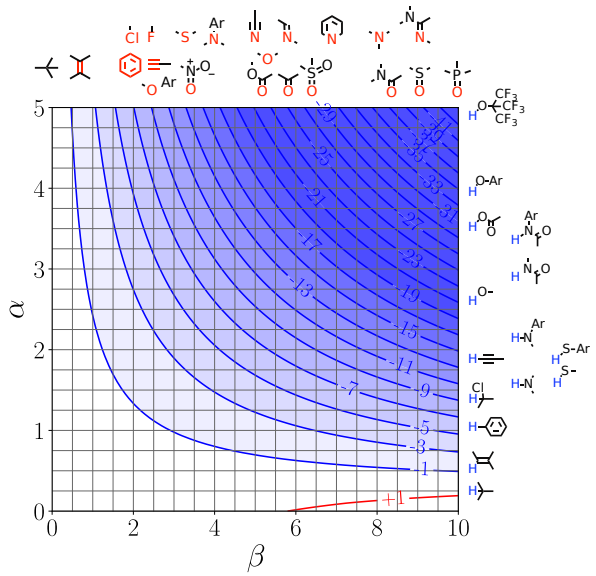


Fig. H.22: FGIP for cis-decalin at 298K.

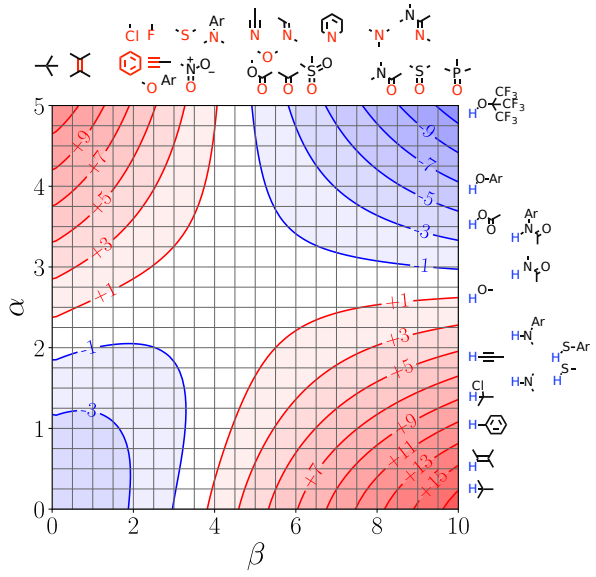


Fig. H.23: FGIP for water at 298K.

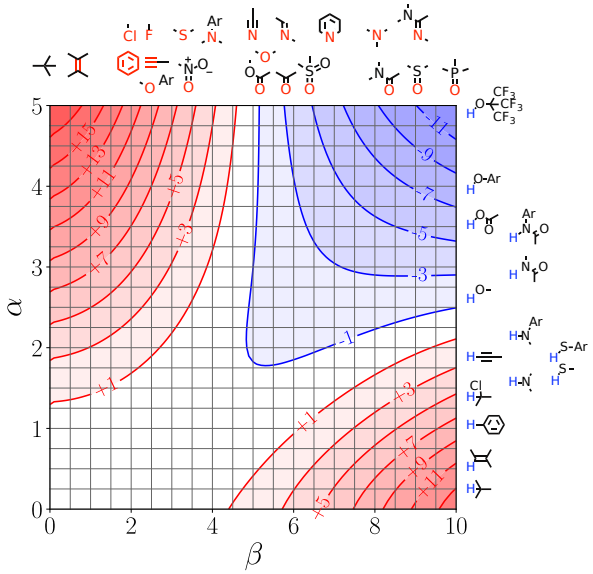


Fig. H.24: FGIP for methanol at 298K.

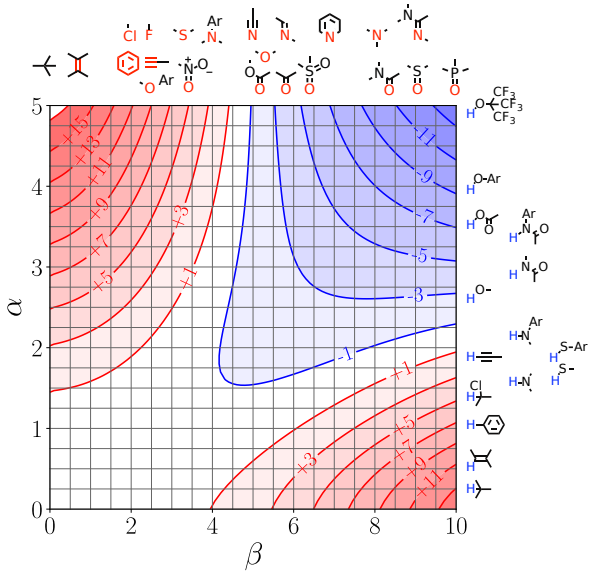


Fig. H.25: FGIP for ethanol at 298K.

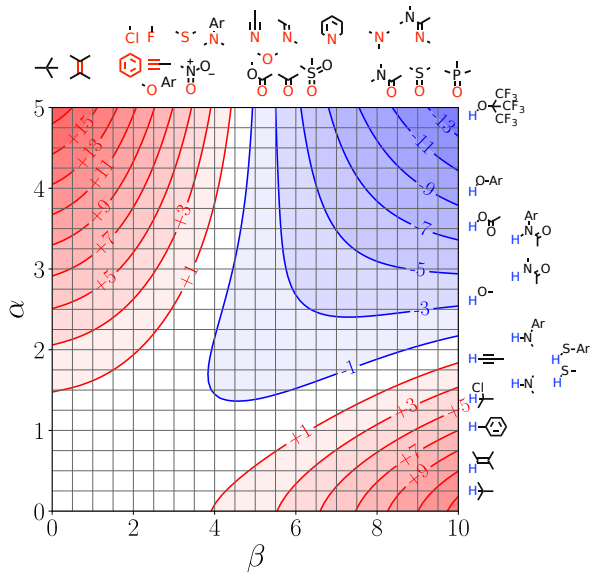


Fig. H.26: FGIP for 1-propanol at 298K.

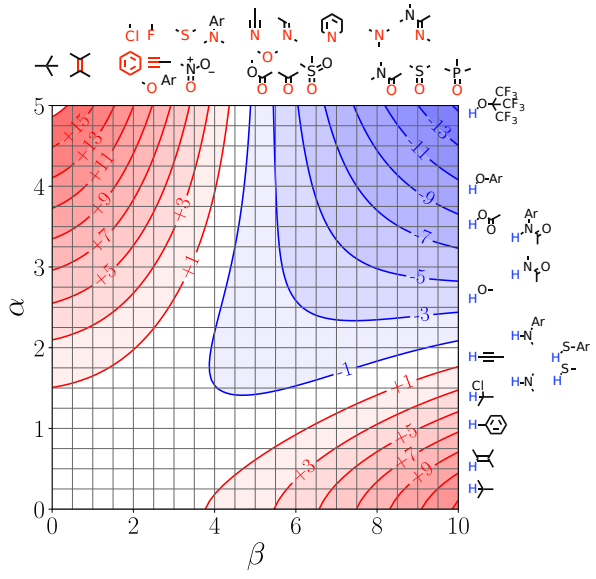


Fig. H.27: FGIP for 2-propanol at 298K.

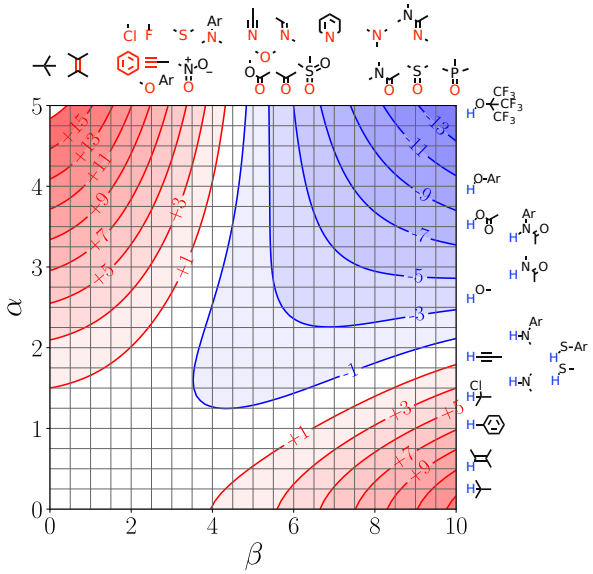


Fig. H.28: FGIP for 1-butanol at 298K.

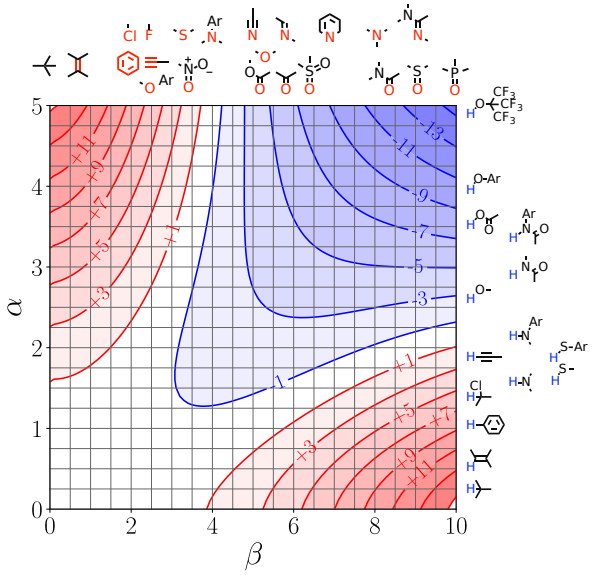


Fig. H.29: FGIP for 2-methyl-1-propanol at 298K.

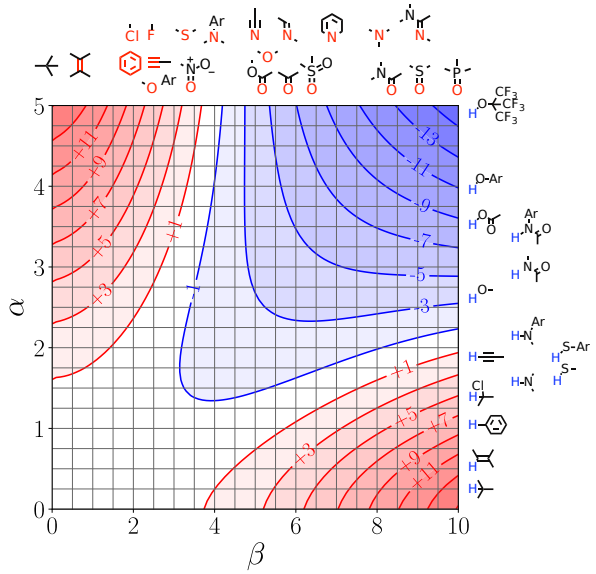


Fig. H.30: FGIP for 2-butanol at 298K.

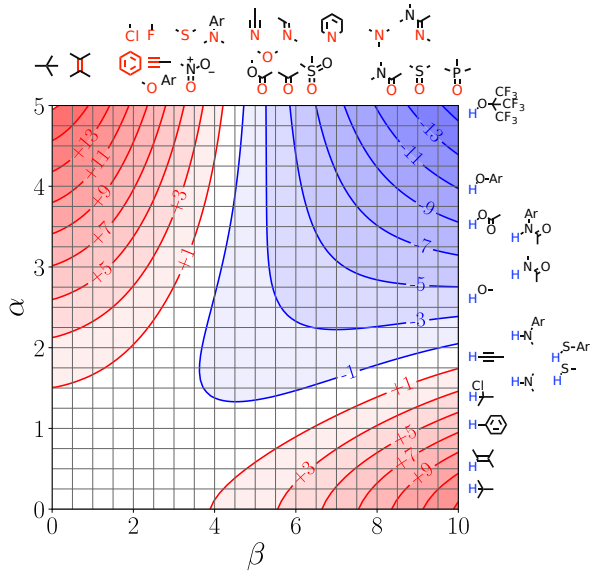


Fig. H.31: FGIP for 2-methyl-2-propanol at 298K.

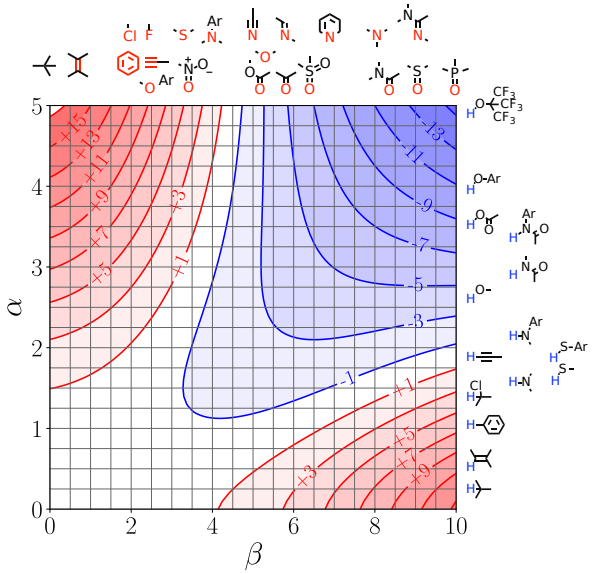


Fig. H.32: FGIP for 1-pentanol at 298K.

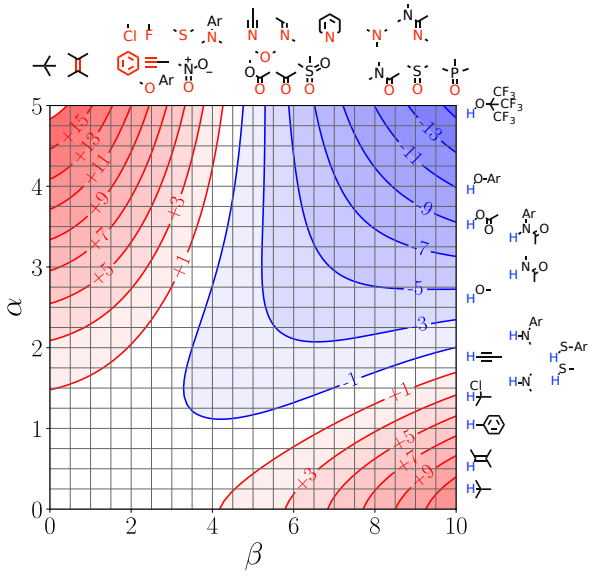


Fig. H.33: FGIP for 3-methyl-1-butanol at 298K.

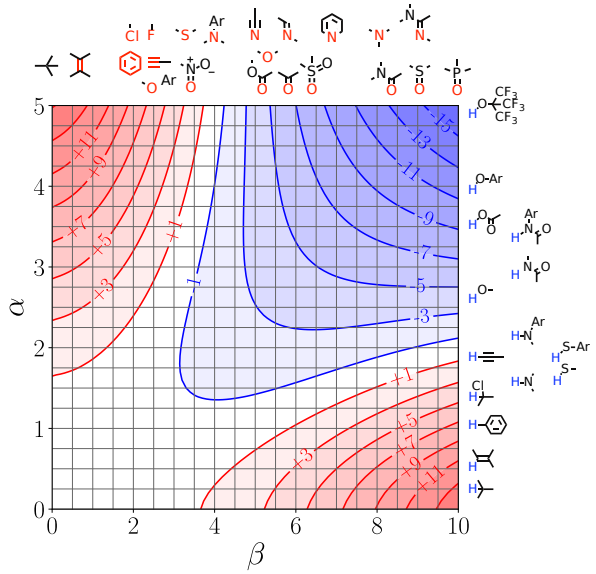


Fig. H.34: FGIP for 2-methyl-2-butanol at 298K.

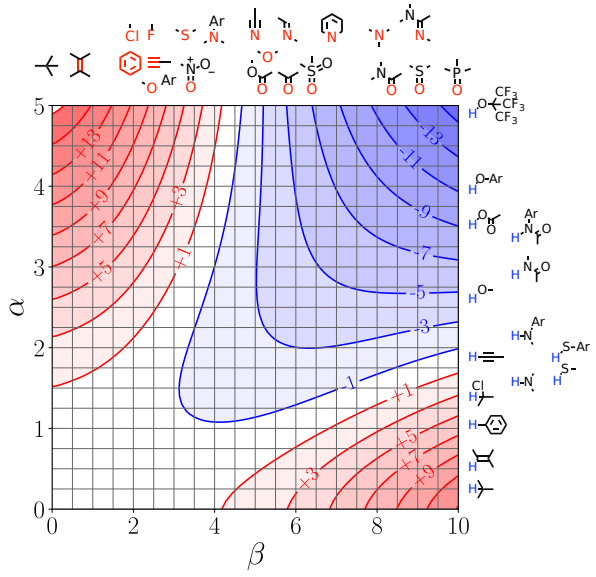


Fig. H.35: FGIP for 1-hexanol at 298K.

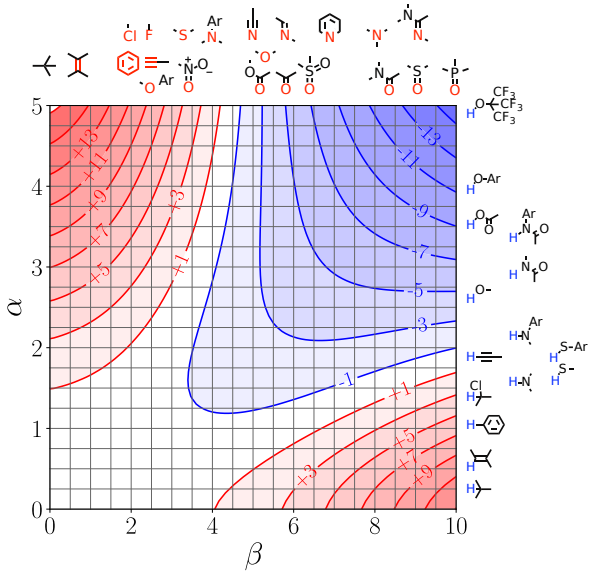


Fig. H.36: FGIP for cyclohexanol at 298K.

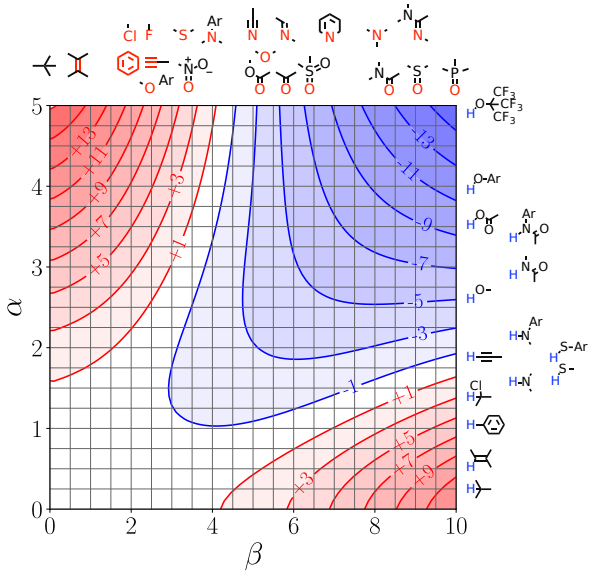


Fig. H.37: FGIP for 1-octanol at 298K.

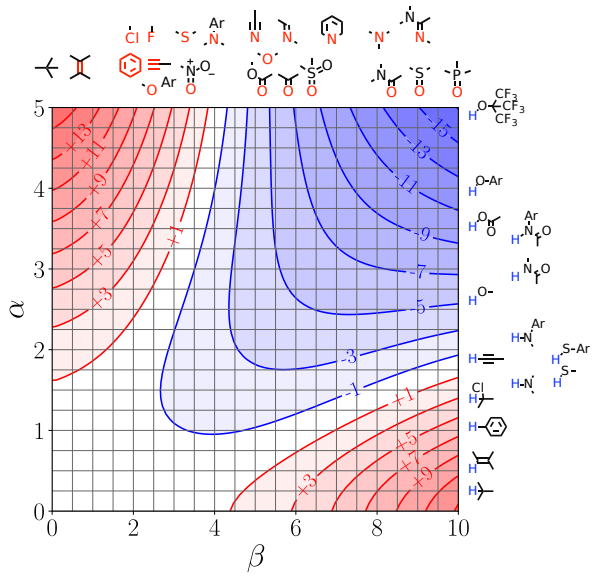


Fig. H.38: FGIP for 1-decanol at 298K.

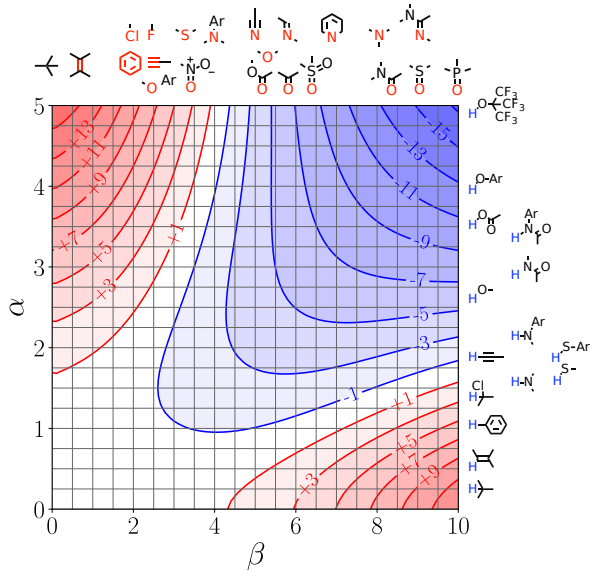


Fig. H.39: FGIP for 1-dodecanol at 298K.

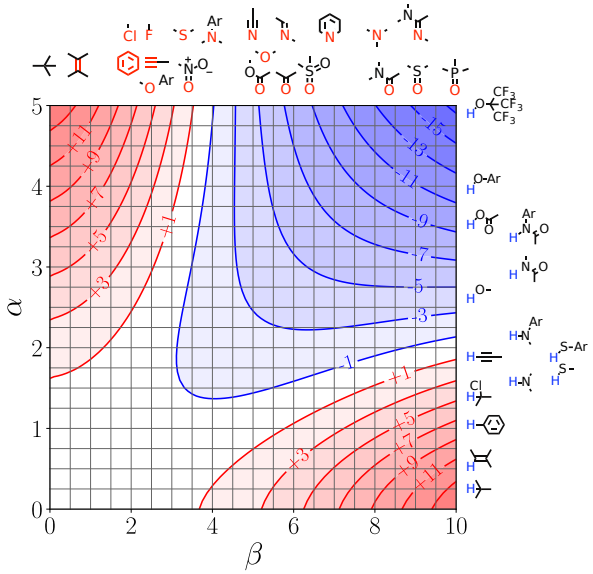


Fig. H.40: FGIP for benzyl alcohol at 298K.

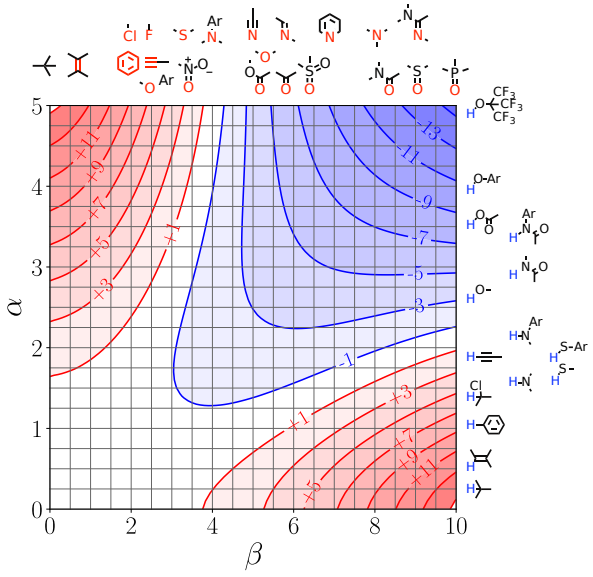


Fig. H.41: FGIP for 2-phenylethanol at 298K.

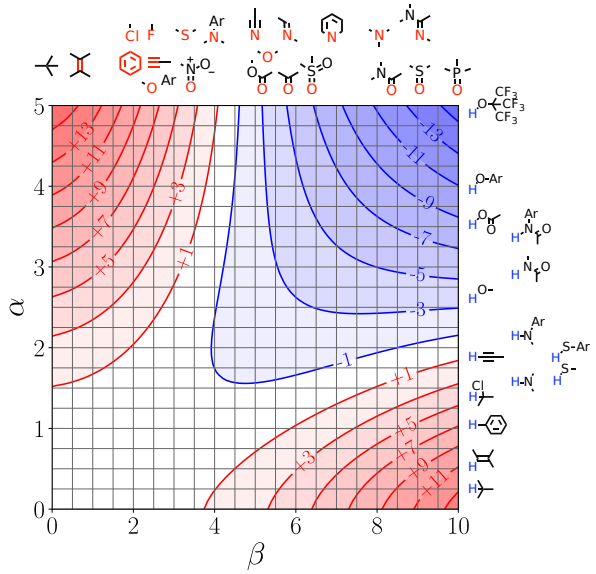


Fig. H.42: FGIP for allyl alcohol at 298K.

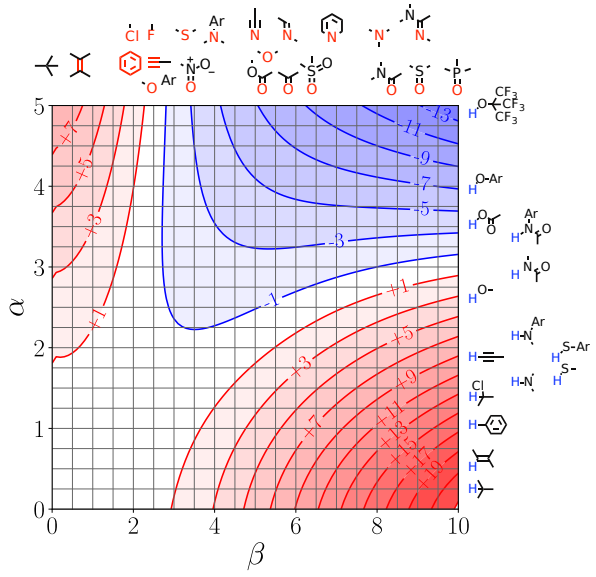


Fig. H.43: FGIP for 2-chloroethanol at 298K.

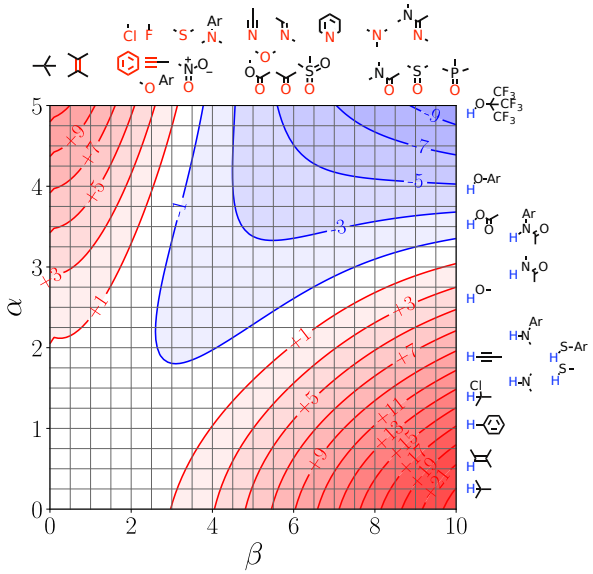


Fig. H.44: FGIP for 2-cyanoethanol at 298K.

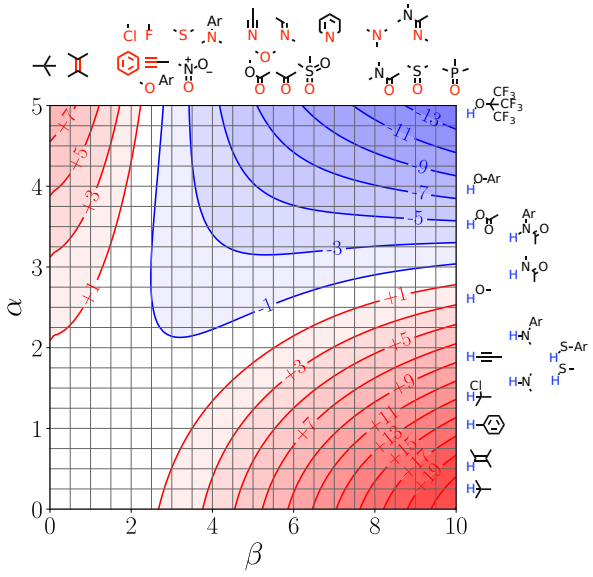


Fig. H.45: FGIP for 2,2,2-trifluoroethanol at 298K.

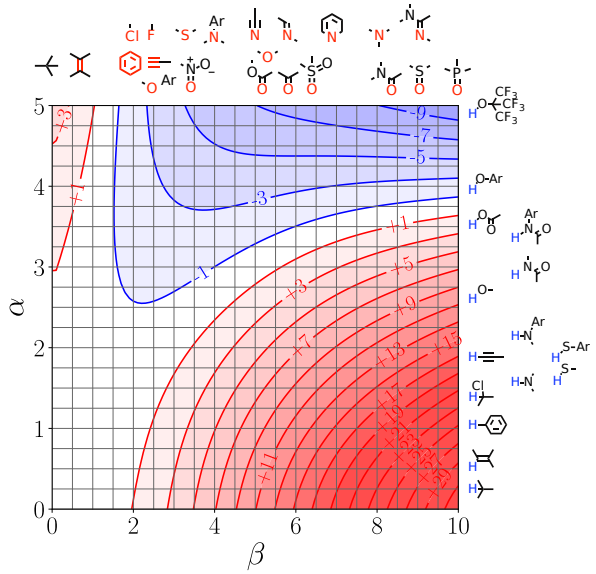


Fig. H.46: FGIP for 1,1,1,3,3,3-hexafluoro-2-propanol at 298K.

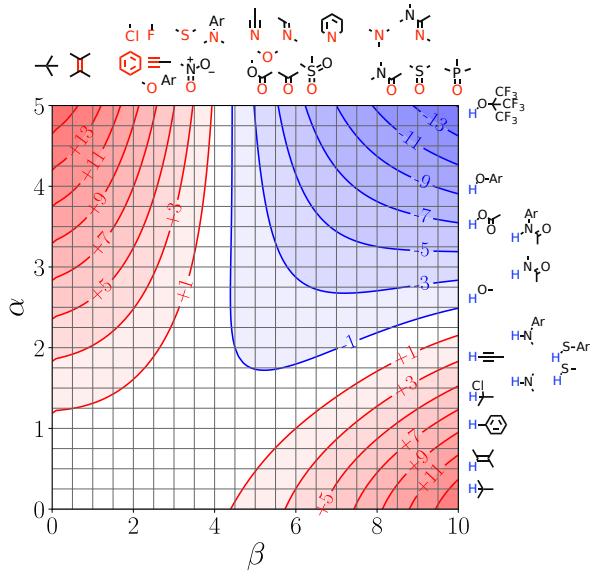


Fig. H.47: FGIP for 2-methoxyethanol at 298K.

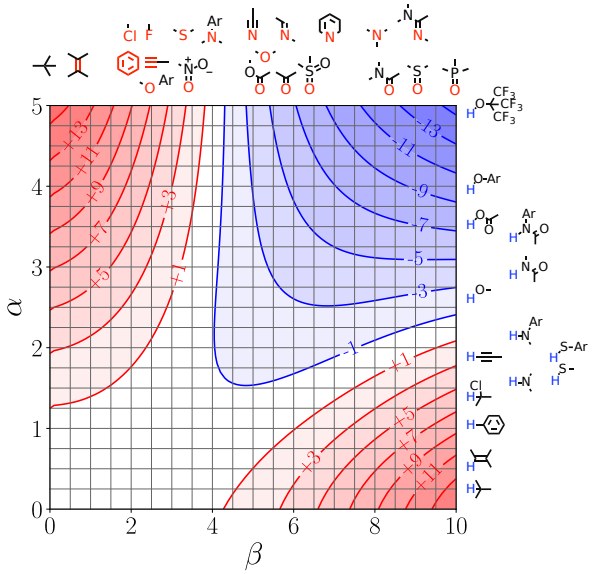


Fig. H.48: FGIP for 2-ethoxyethanol at 298K.

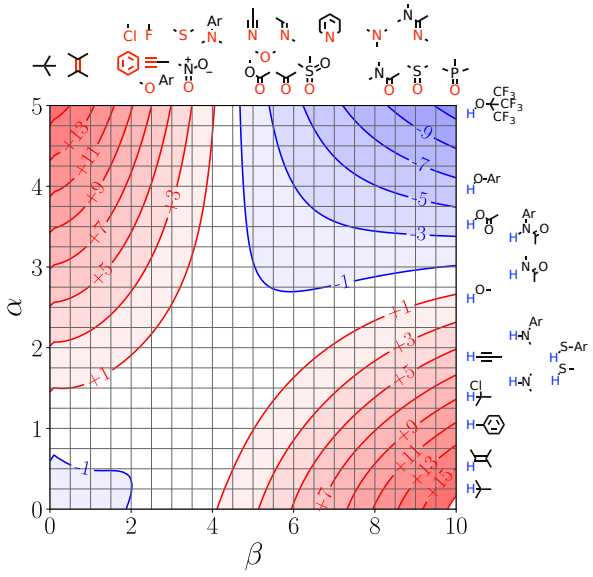


Fig. H.49: FGIP for ethylene glycol at 298K.

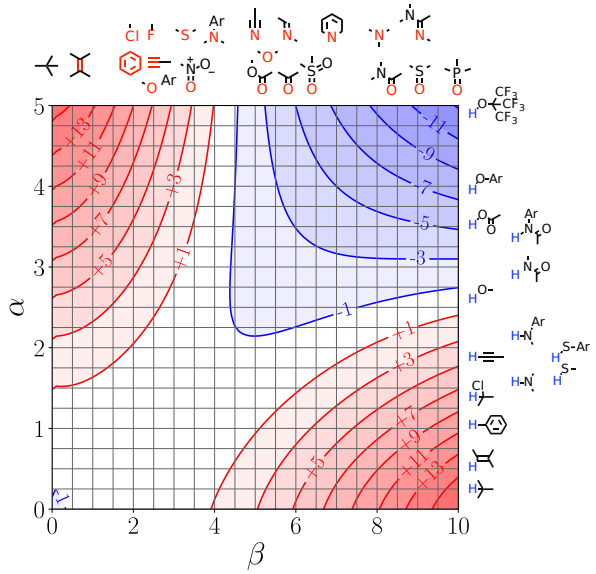


Fig. H.50: FGIP for 1,2-propanediol at 298K.

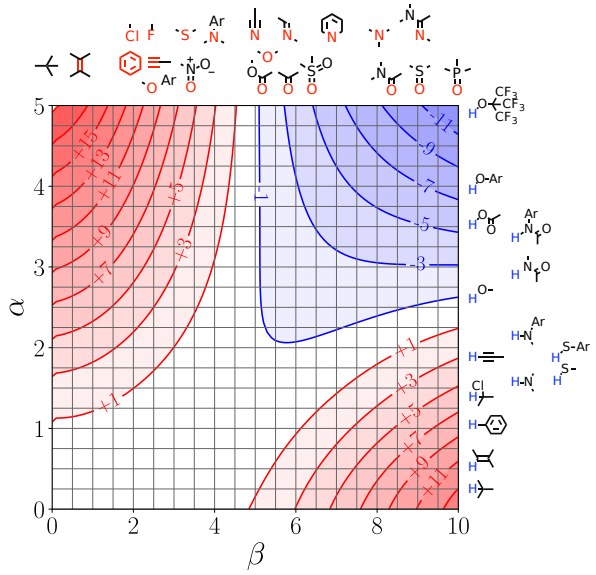


Fig. H.51: FGIP for 1,3-propanediol at 298K.

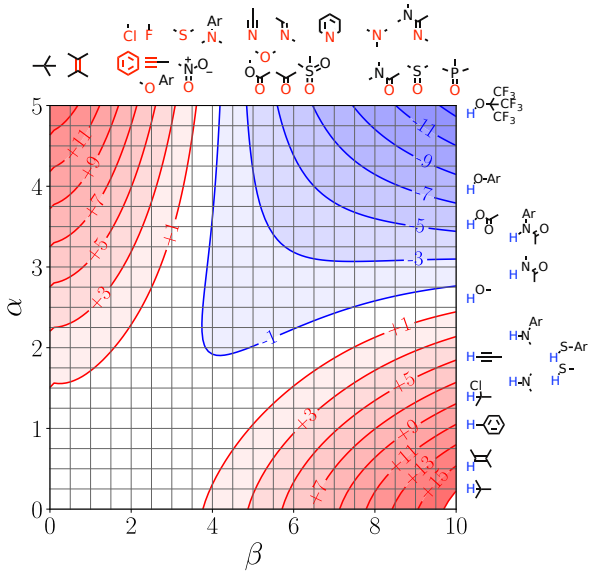


Fig. H.52: FGIP for 1,2-butanediol at 298K.

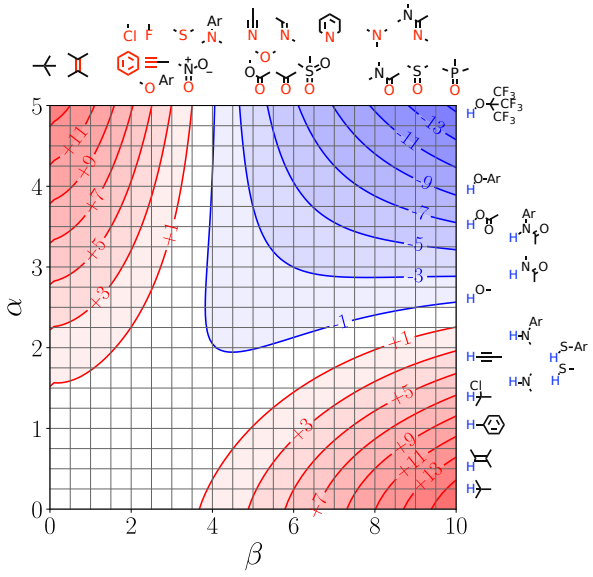


Fig. H.53: FGIP for (2R,3S)-2,3-butanediol at 298K.

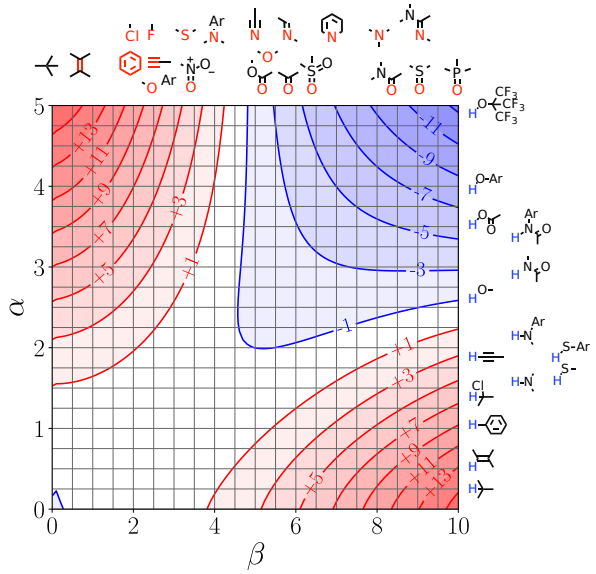


Fig. H.54: FGIP for 1,4-butanediol at 298K.

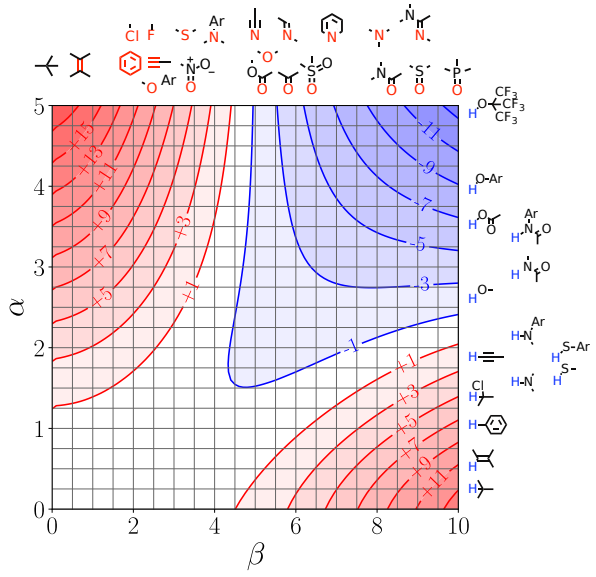


Fig. H.55: FGIP for 1,5-pentanediol at 298K.

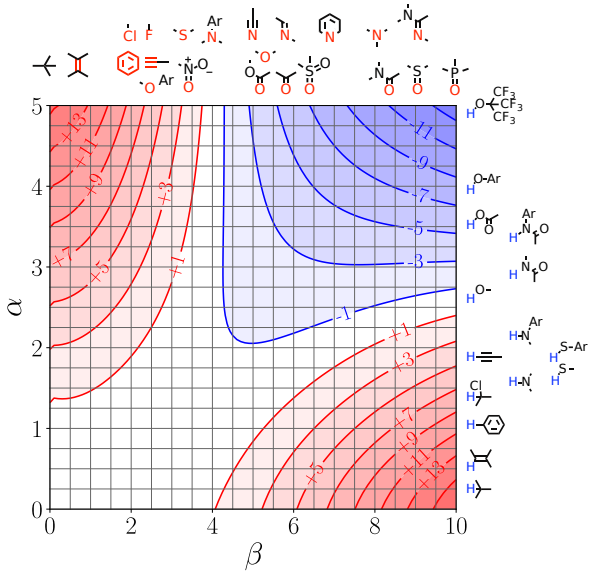


Fig. H.56: FGIP for diethylene glycol at 298K.

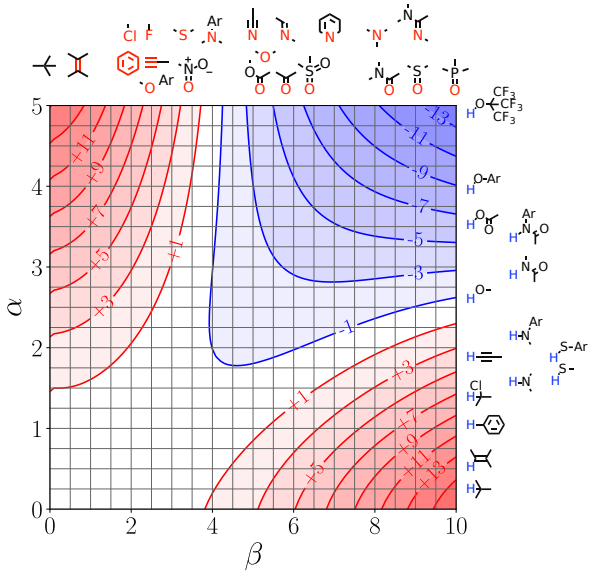


Fig. H.57: FGIP for triethylene glycol at 298K.

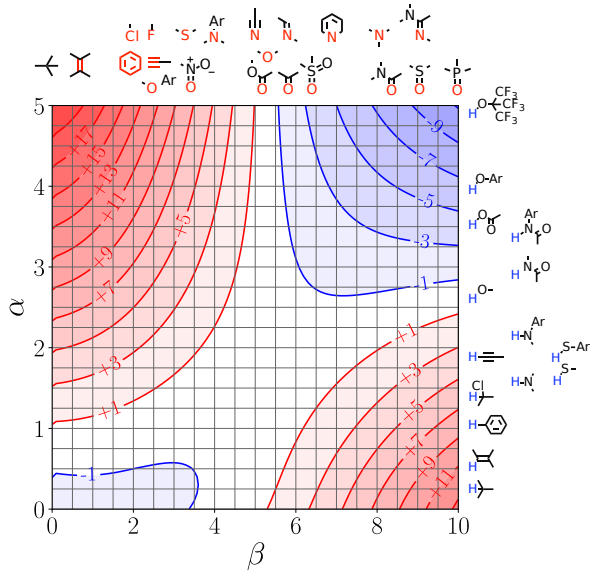


Fig. H.58: FGIP for glycerol at 298K.

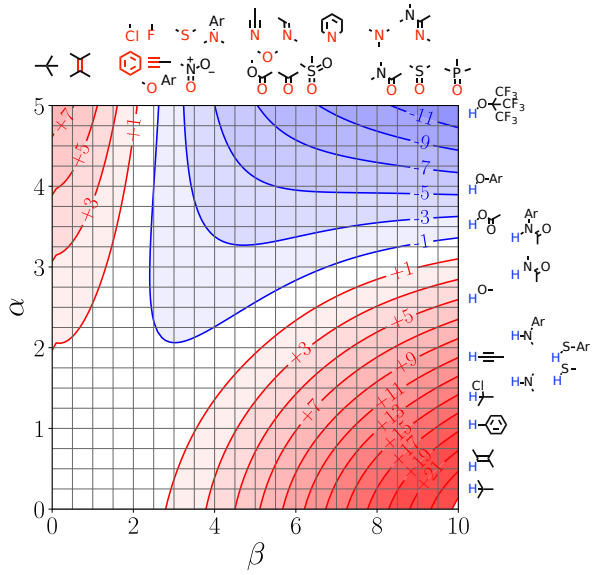


Fig. H.59: FGIP for phenol at 298K.

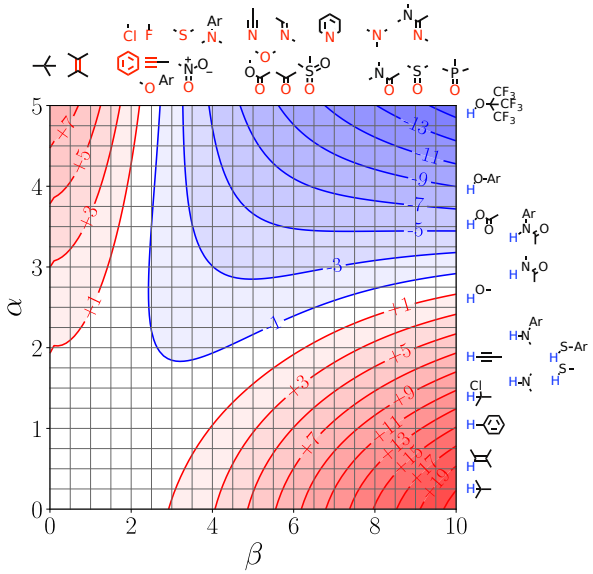


Fig. H.60: FGIP for ortho-cresol at 298K.

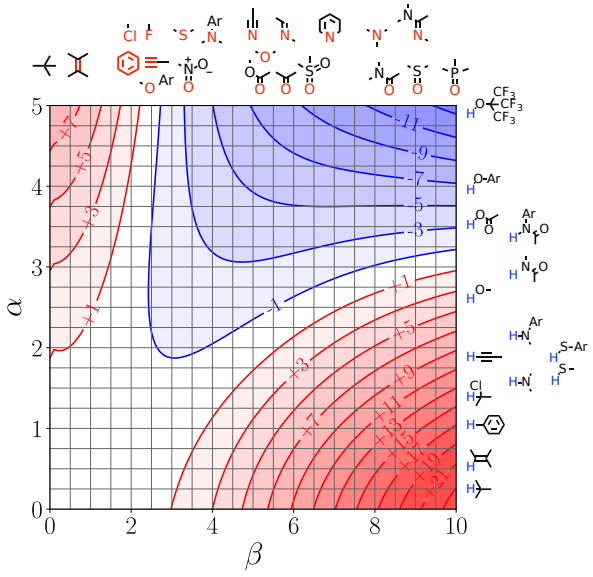


Fig. H.61: FGIP for meta-cresol at 298K.

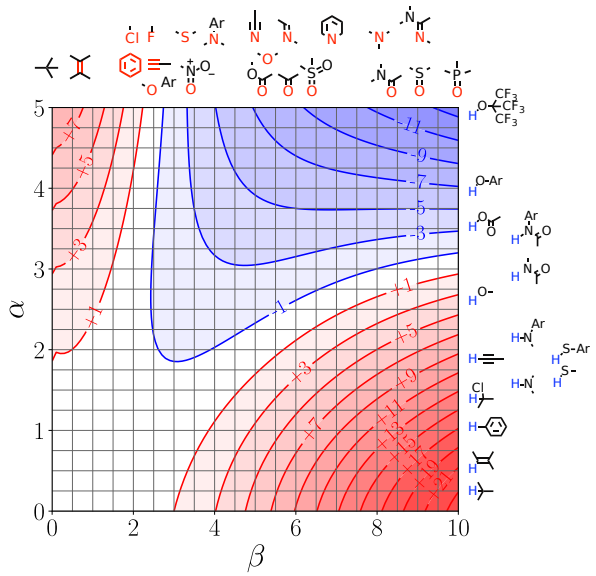


Fig. H.62: FGIP for para-cresol at 298K.

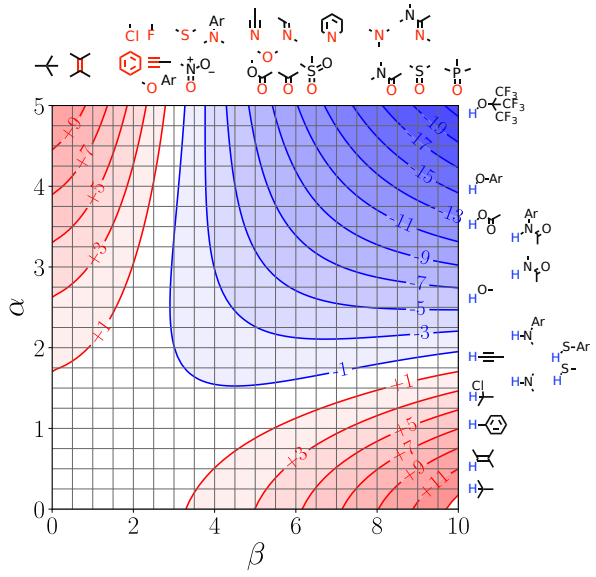


Fig. H.63: FGIP for 2-methoxyphenol at 298K.

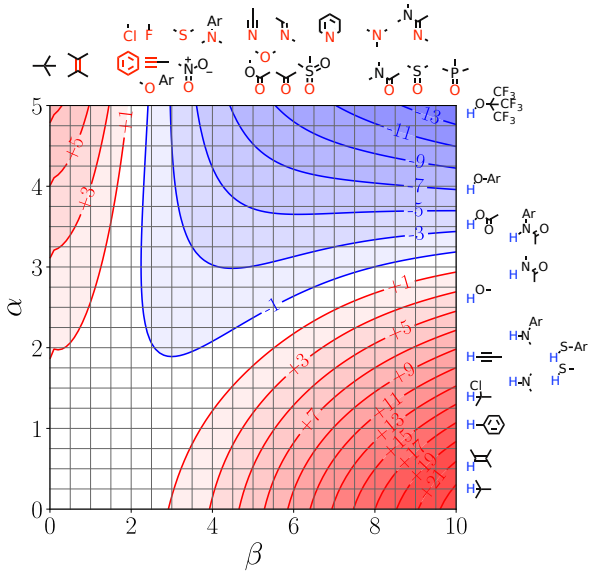


Fig. H.64: FGIP for 2,4-dimethylphenol at 298K.

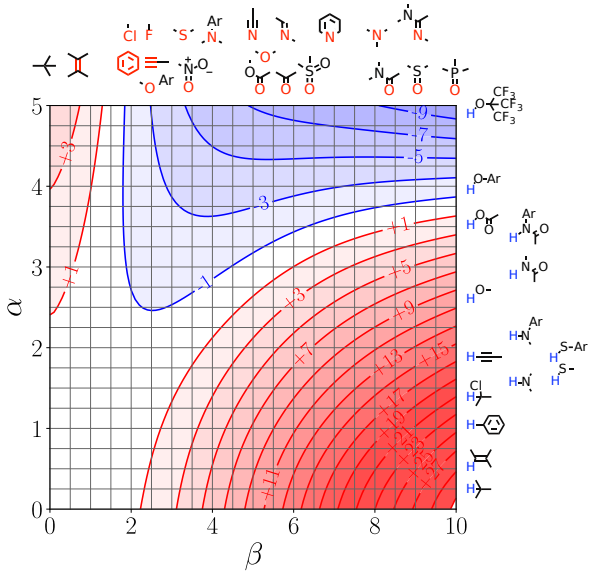


Fig. H.65: FGIP for 3-chlorophenol at 298K.

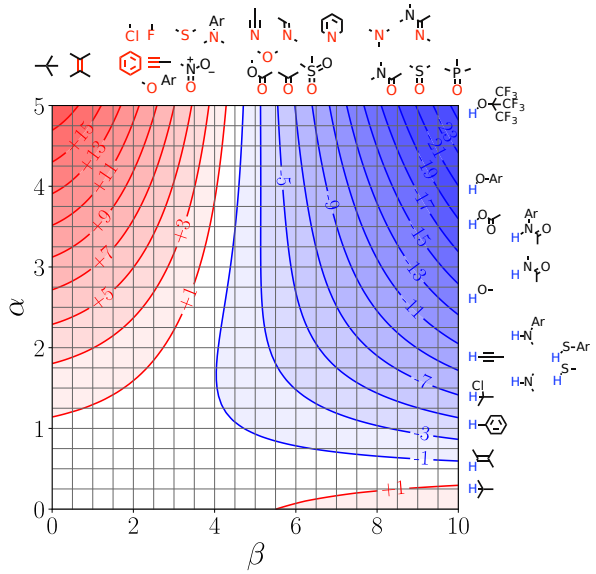


Fig. H.66: FGIP for diethyl ether at 298K.

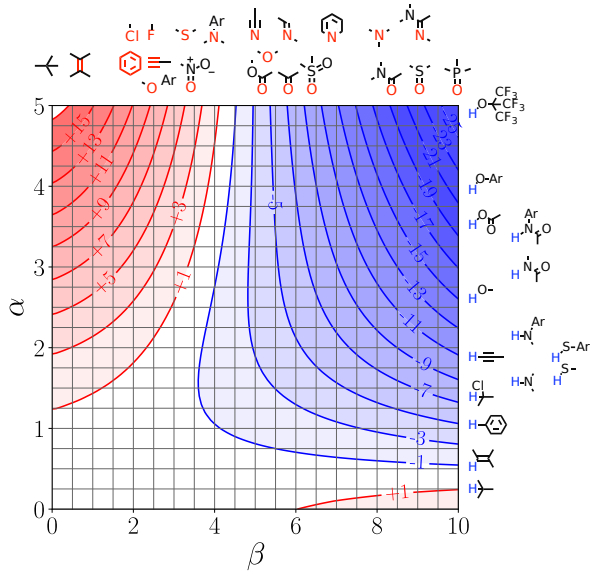


Fig. H.67: FGIP for di-n-propyl ether at 298K.

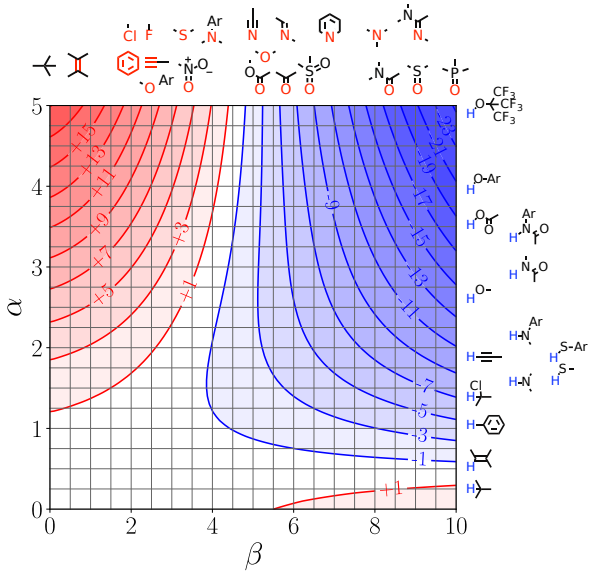


Fig. H.68: FGIP for diisopropyl ether at 298K.

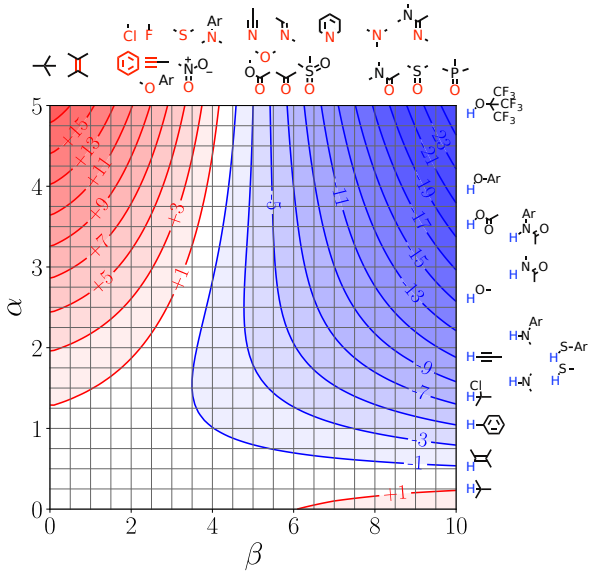


Fig. H.69: FGIP for dibutyl ether at 298K.

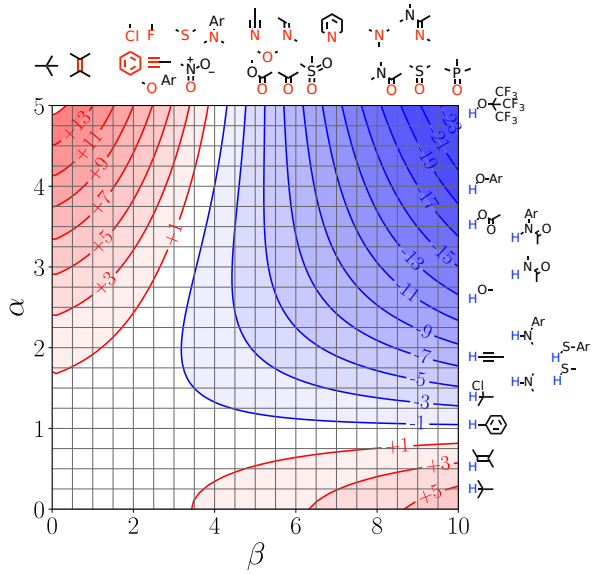


Fig. H.70: FGIP for bis(2-chloroethyl) ether at 298K.

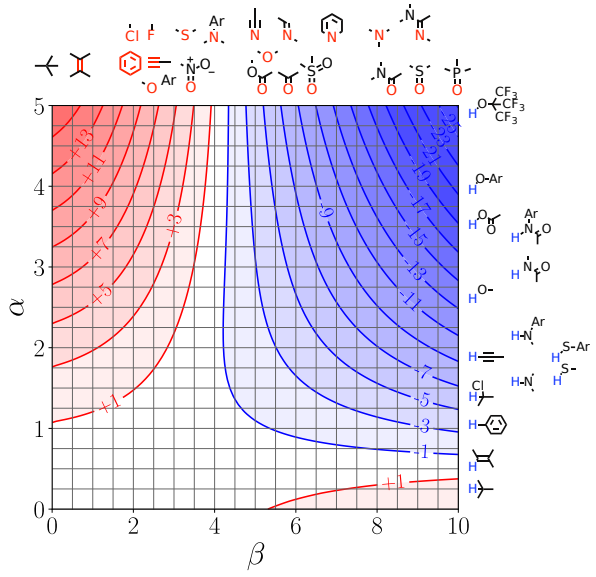


Fig. H.71: FGIP for 1,2-dimethyoxyethane at 298K.

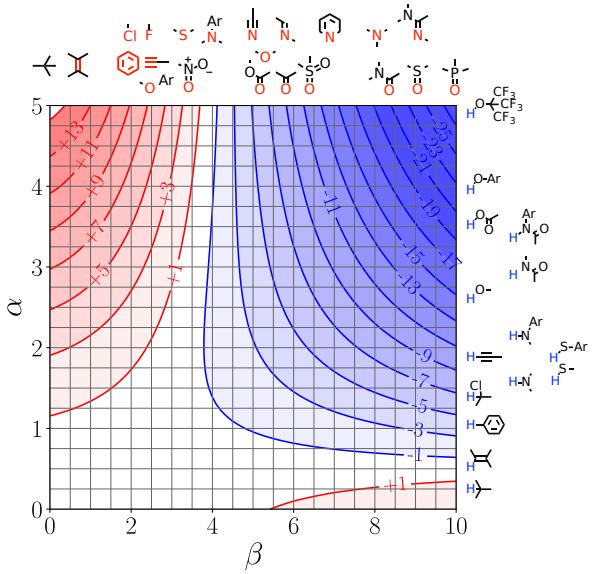


Fig. H.72: FGIP for diethylene glycol dimethyl ether at 298K.

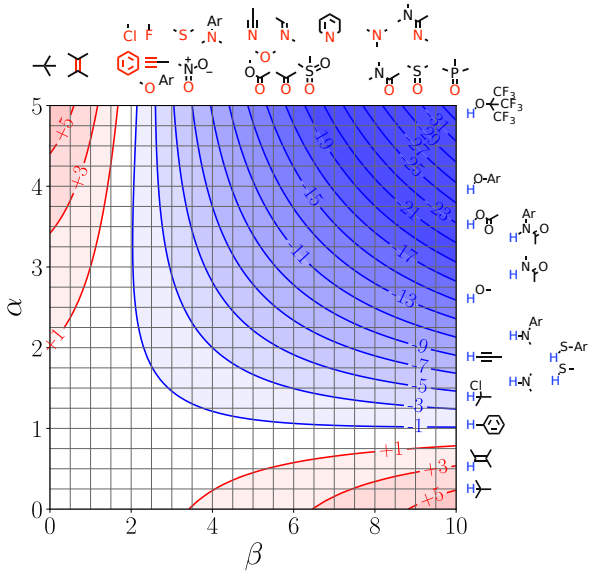


Fig. H.73: FGIP for furan at 298K.

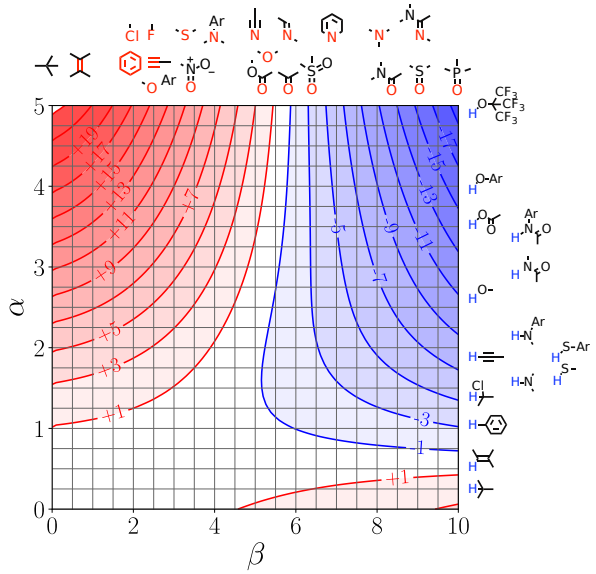


Fig. H.74: FGIP for tetrahydrofuran at 298K.

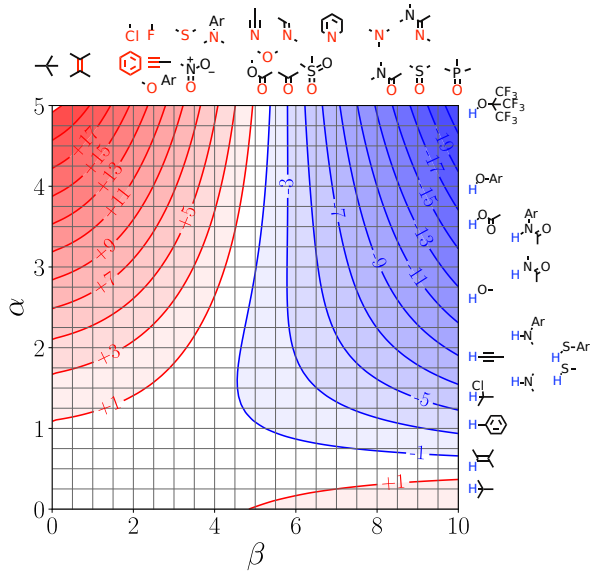


Fig. H.75: FGIP for 2-methyltetrahydrofuran at 298K.

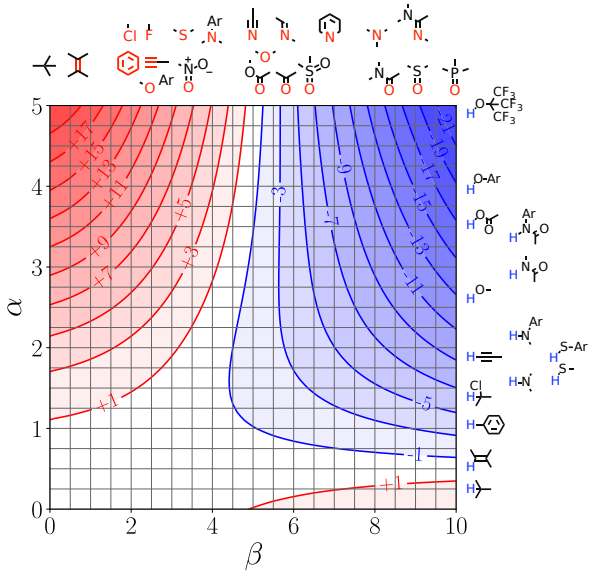


Fig. H.76: FGIP for tetrahydropyran at 298K.

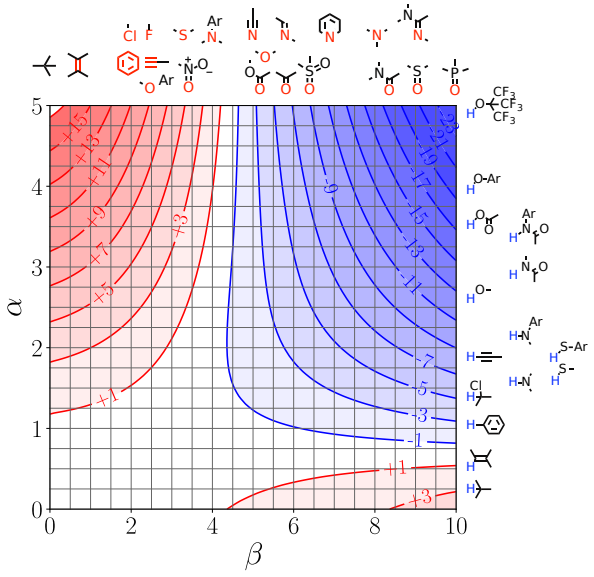


Fig. H.77: FGIP for 1,3-dioxane at 298K.

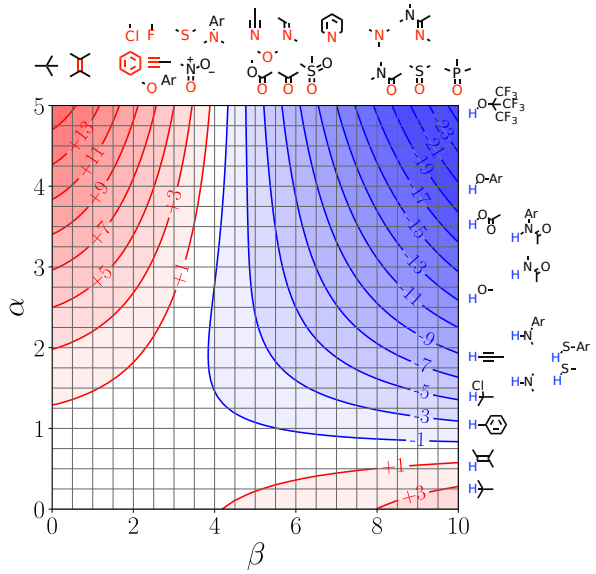


Fig. H.78: FGIP for 1,3-dioxolan at 298K.

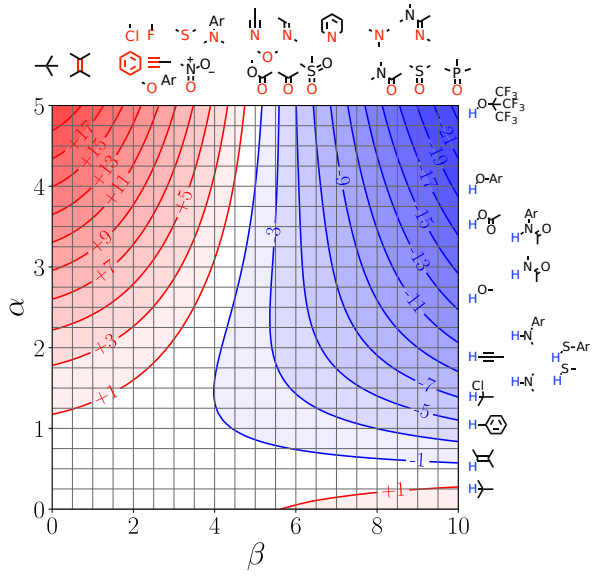


Fig. H.79: FGIP for 1,8-cineole at 298K.

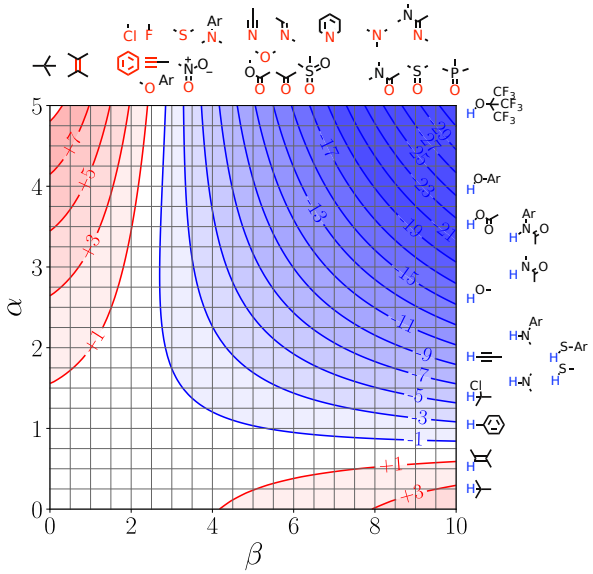


Fig. H.80: FGIP for anisole at 298K.

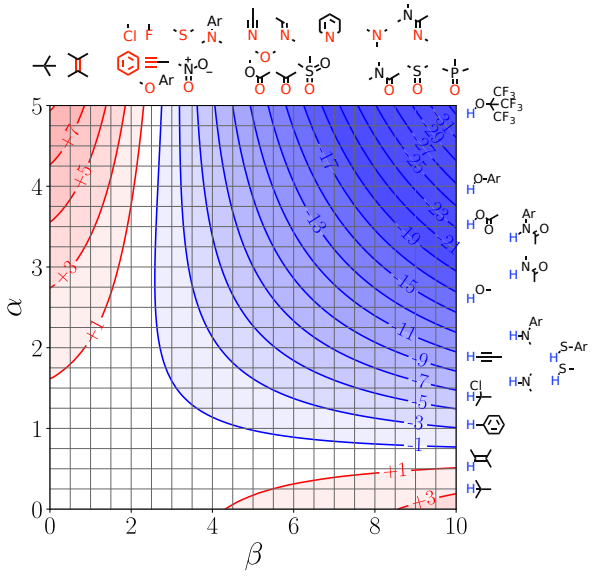


Fig. H.81: FGIP for ethyl phenyl ether at 298K.

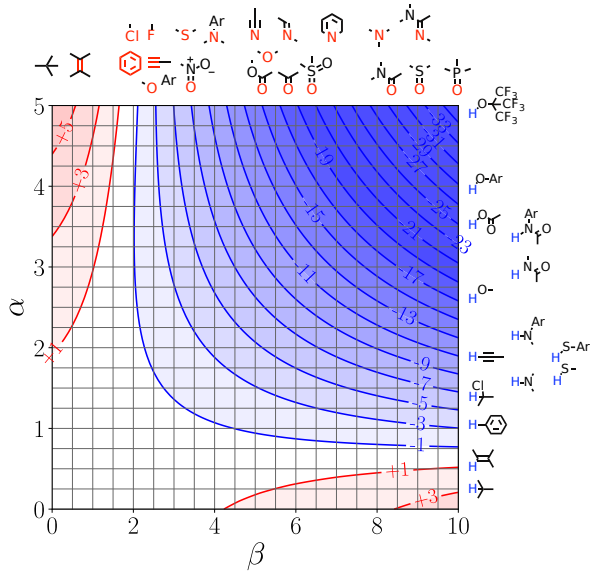


Fig. H.82: FGIP for diphenyl ether at 298K.

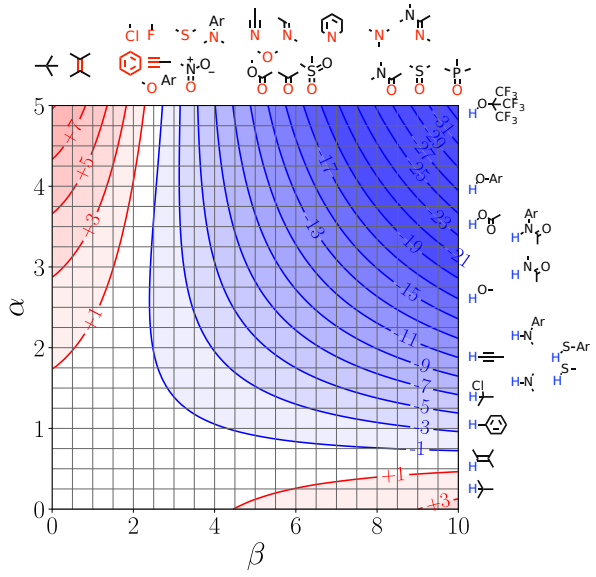


Fig. H.83: FGIP for dibenzyl ether at 298K.

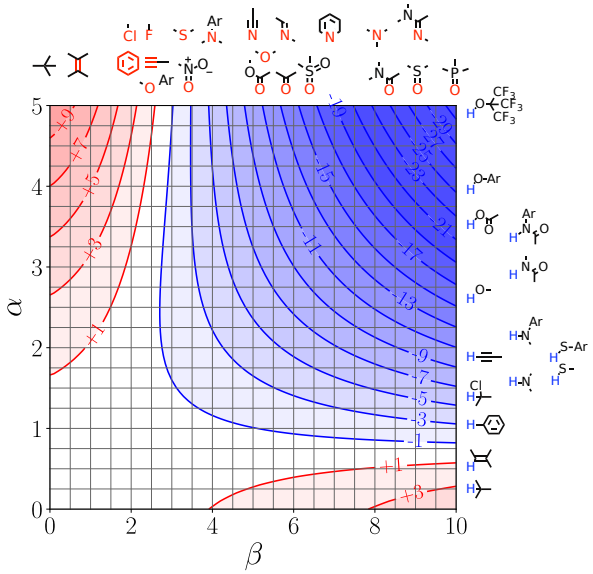


Fig. H.84: FGIP for 1,2-dimethoxybenzene at 298K.

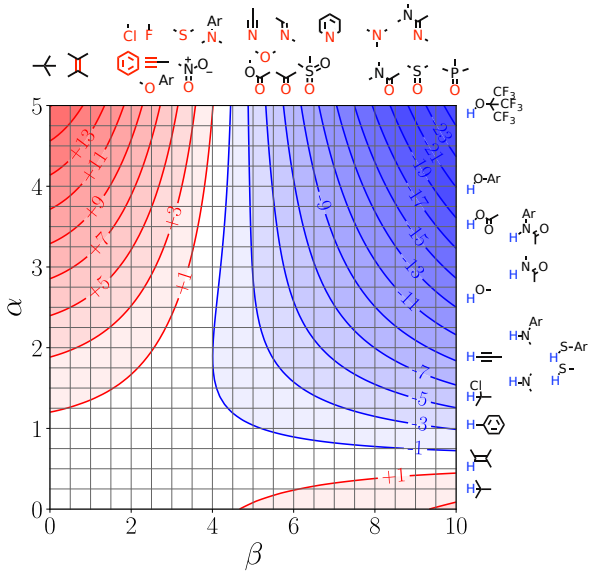


Fig. H.85: FGIP for methyl orthoformate at 298K.

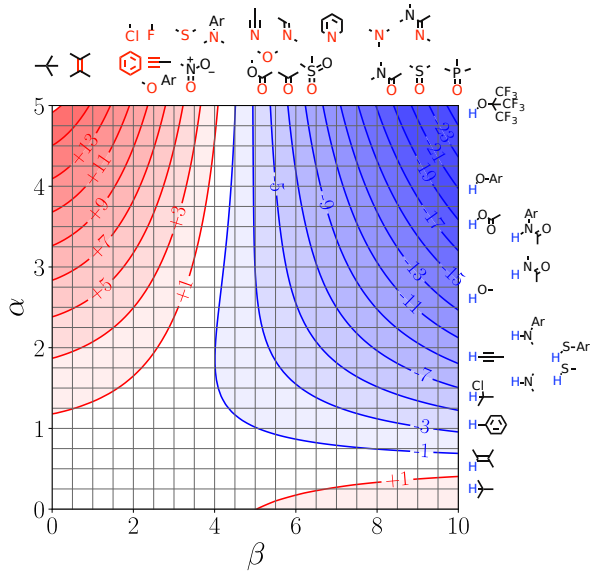


Fig. H.86: FGIP for methyl orthoacetate at 298K.

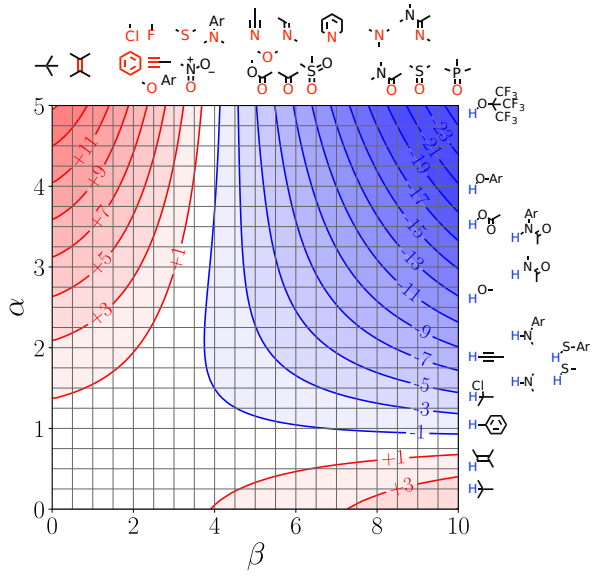


Fig. H.87: FGIP for propionaldehyde at 298K.

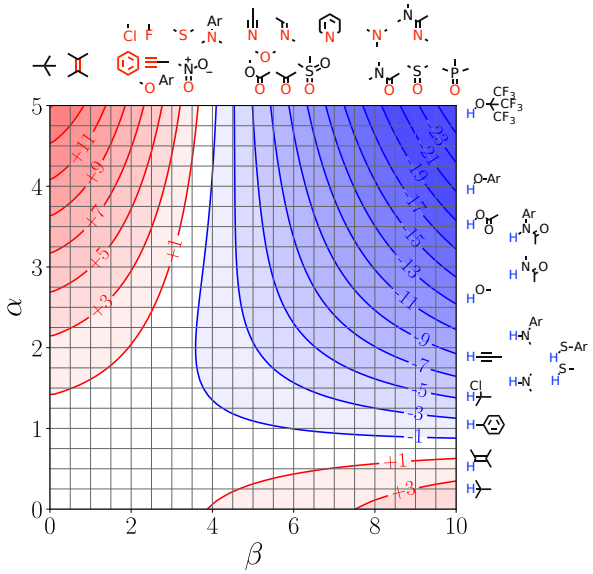


Fig. H.88: FGIP for butyraldehyde at 298K.

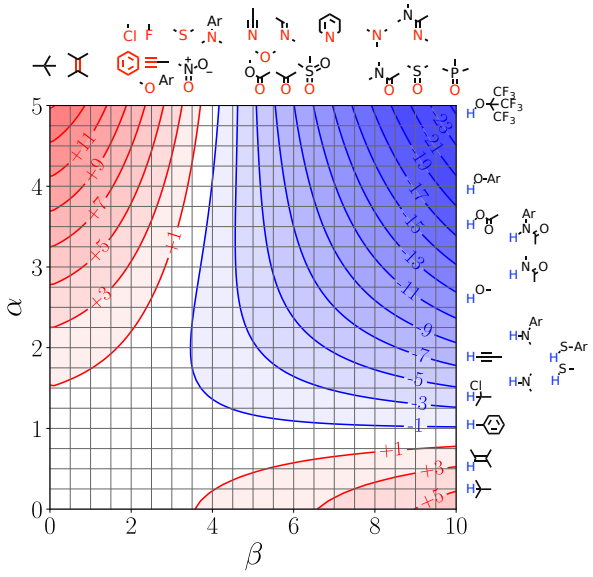


Fig. H.89: FGIP for benzaldehyde at 298K.

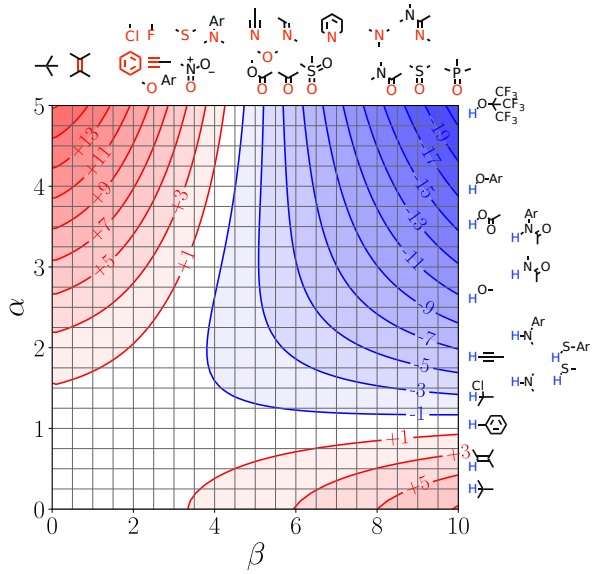


Fig. H.90: FGIP for p-methoxybenzaldehyde at 298K.

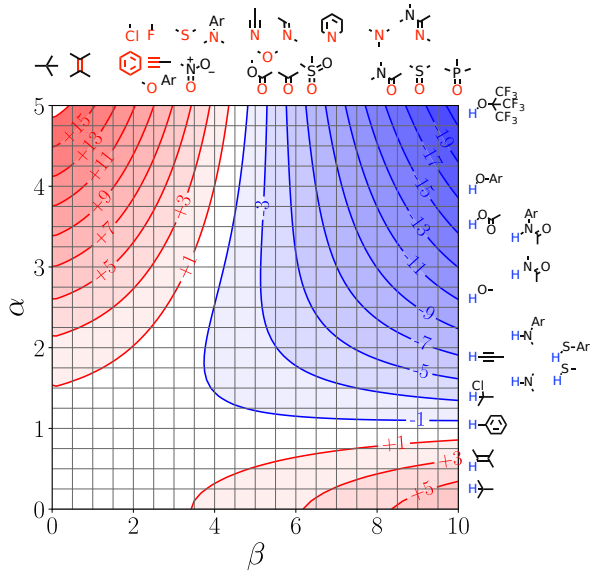


Fig. H.91: FGIP for cinnamaldehyde at 298K.

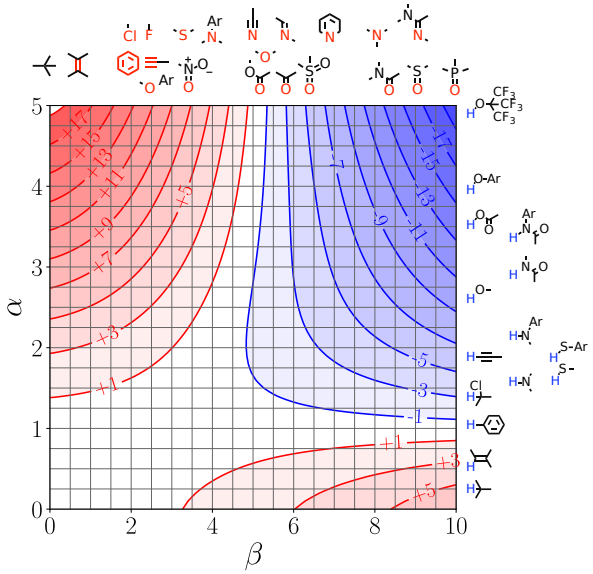


Fig. H.92: FGIP for acetone at 298K.

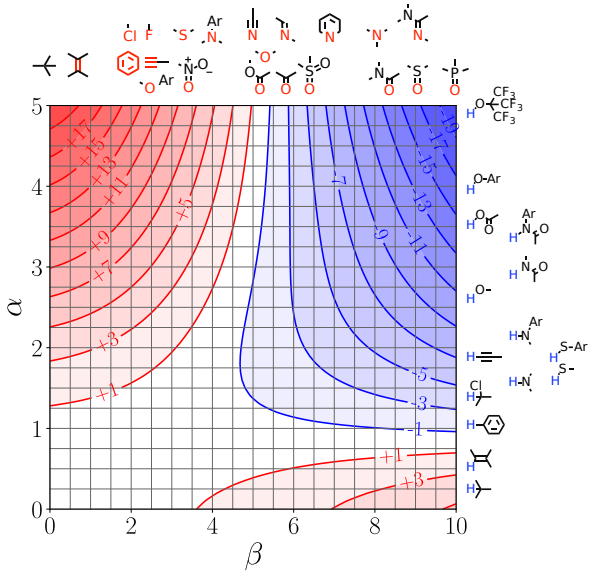


Fig. H.93: FGIP for 2-butanone at 298K.

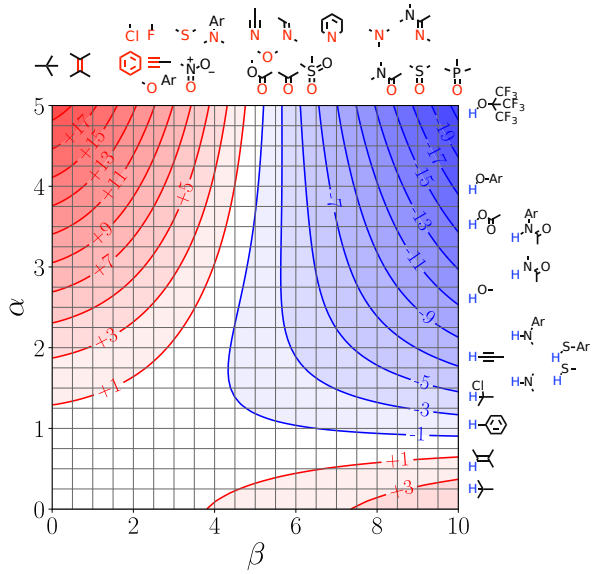


Fig. H.94: FGIP for 2-pentanone at 298K.

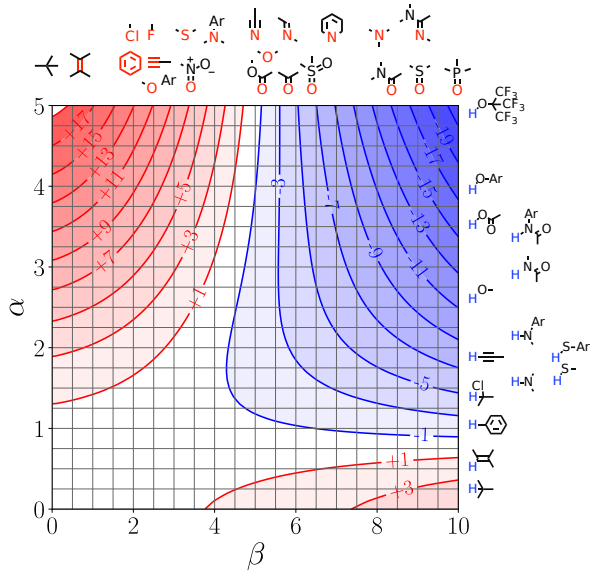


Fig. H.95: FGIP for 3-methyl-2-butanone at 298K.

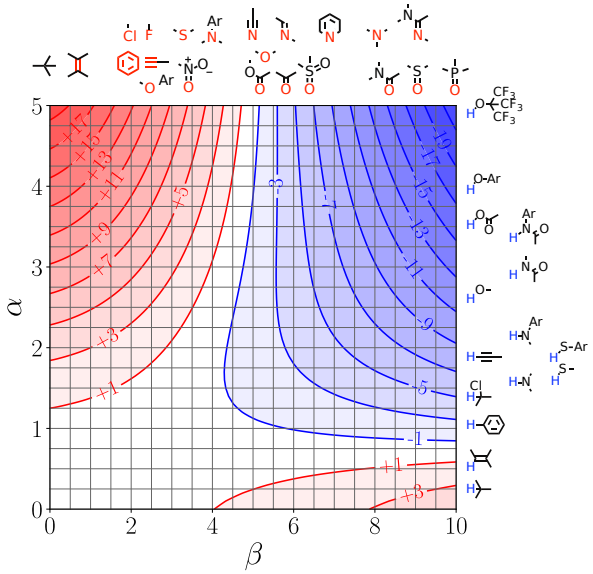


Fig. H.96: FGIP for 3-pentanone at 298K.

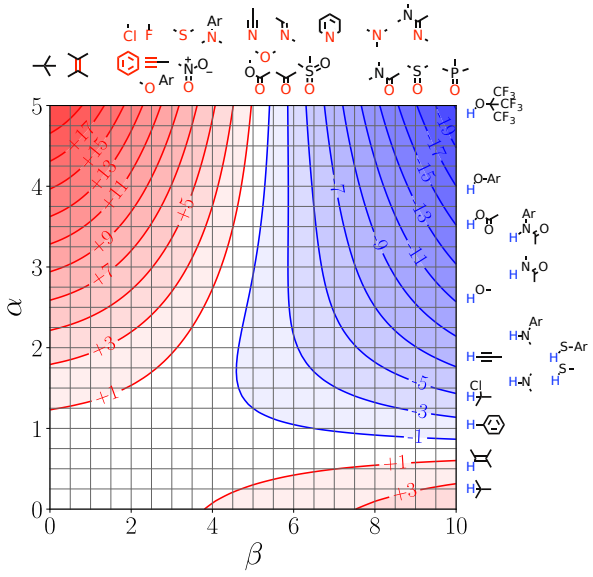


Fig. H.97: FGIP for cyclopentanone at 298K.

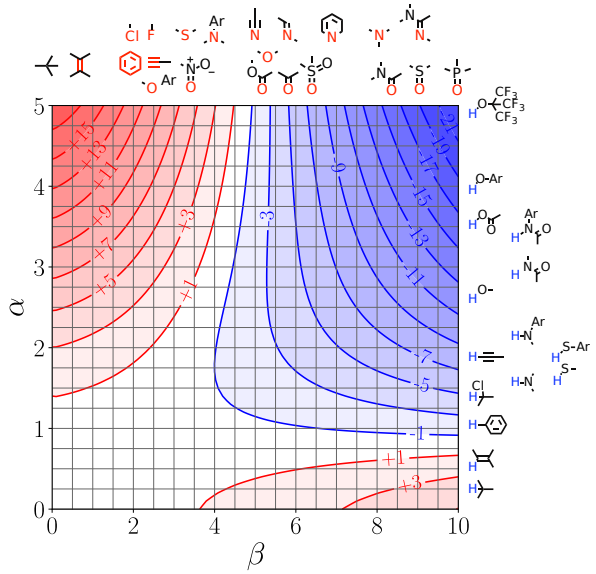


Fig. H.98: FGIP for 4-methyl-2-pentanone at 298K.

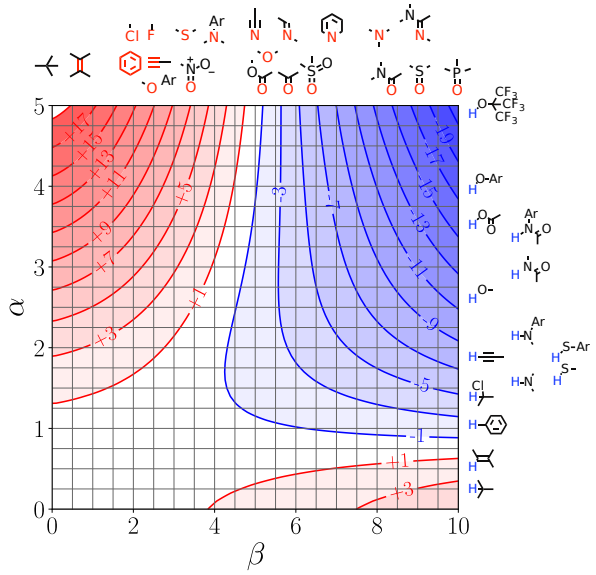


Fig. H.99: FGIP for 3,3-dimethyl-2-butanone at 298K.

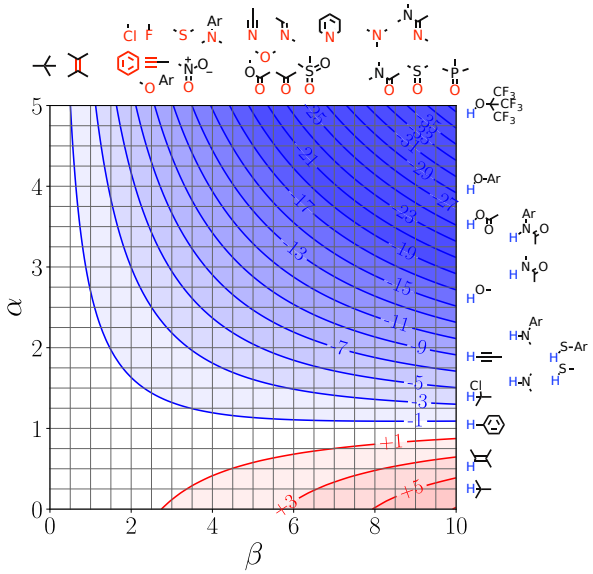


Fig. H.100: FGIP for perfluorooctane at 298K.

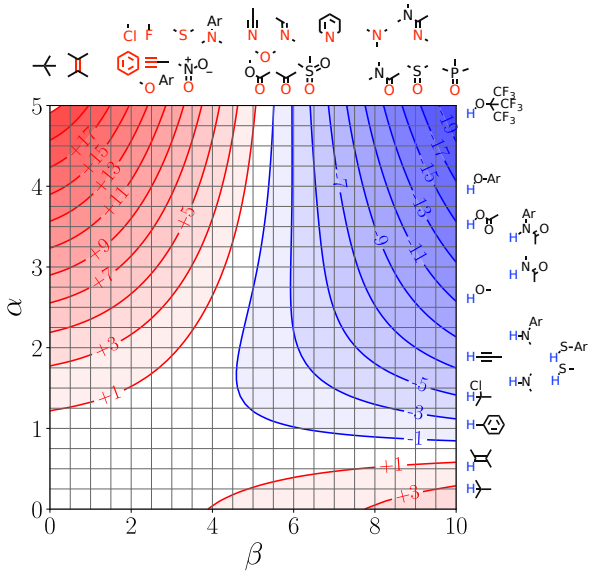


Fig. H.101: FGIP for cyclohexanone at 298K.

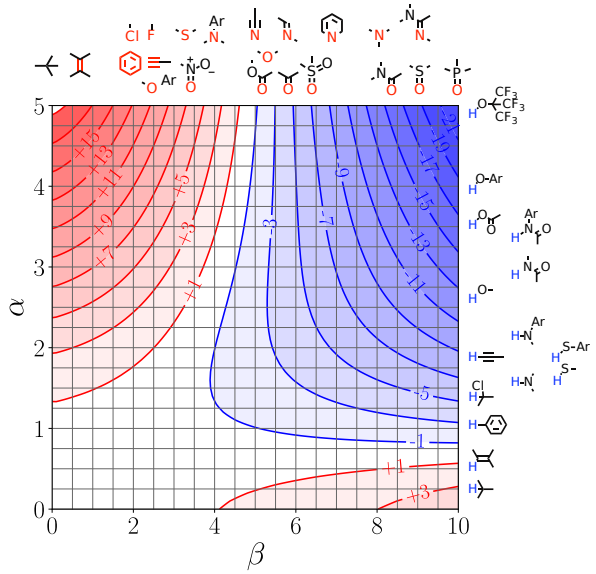


Fig. H.102: FGIP for 2-heptanone at 298K.

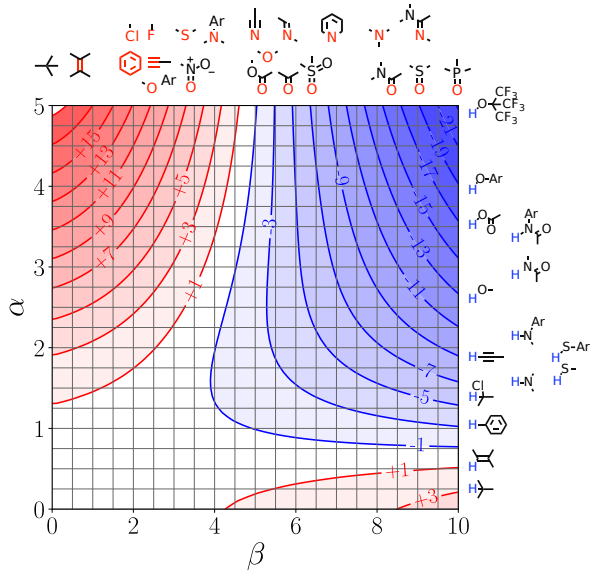


Fig. H.103: FGIP for 3-heptanone at 298K.

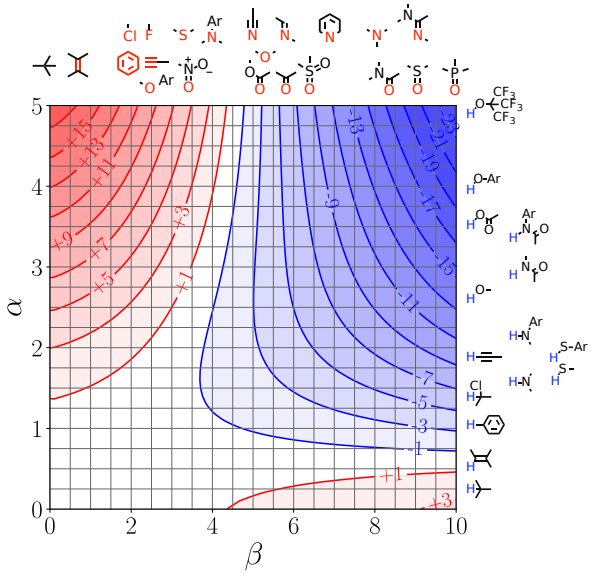


Fig. H.104: FGIP for 2,2,4,4-tetramethyl-3-pentanone at 298K.

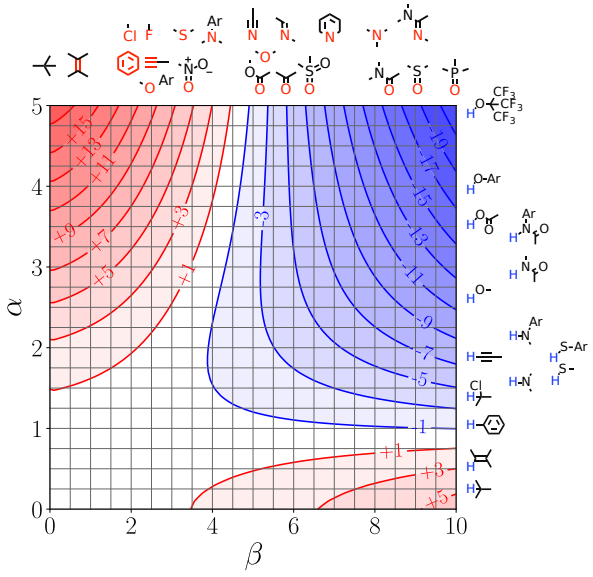


Fig. H.105: FGIP for acetophenone at 298K.

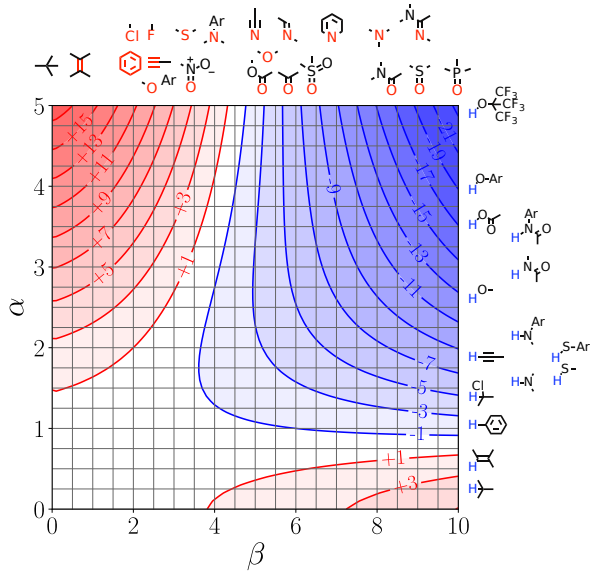


Fig. H.106: FGIP for ethyl phenyl ketone at 298K.

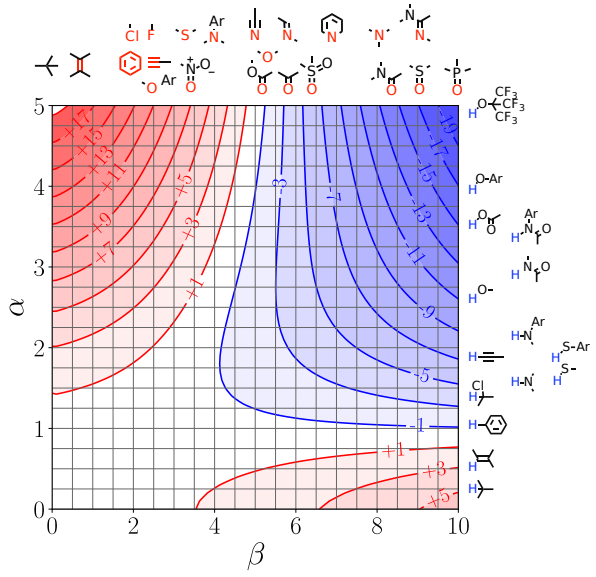


Fig. H.107: FGIP for benzyl methyl ketone at 298K.

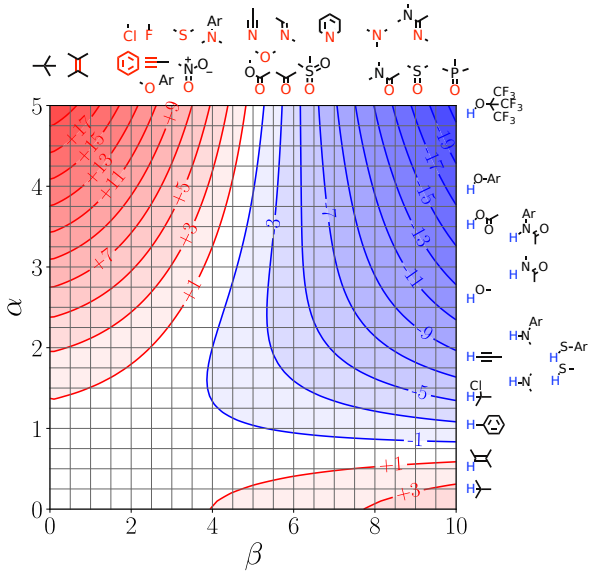


Fig. H.108: FGIP for 2,4,5-trimethylacetophenone at 298K.

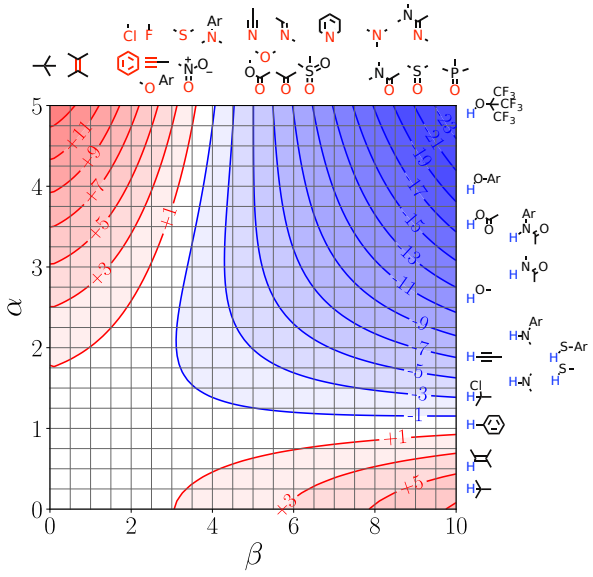


Fig. H.109: FGIP for p-chloroacetophenone at 298K.

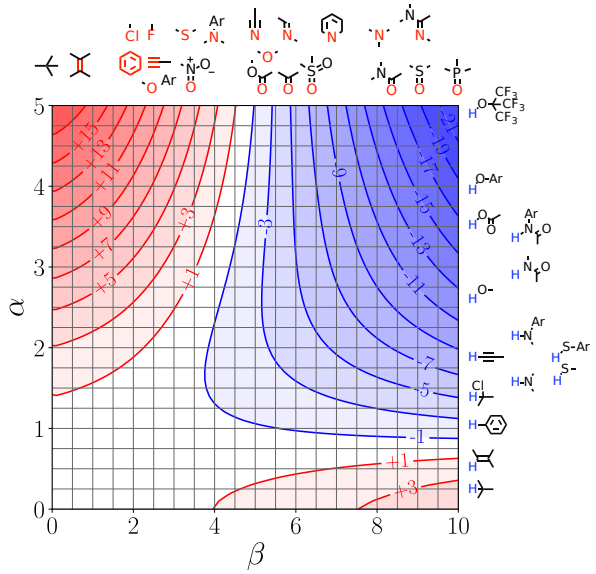


Fig. H.110: FGIP for diphenyl ketone at 298K.

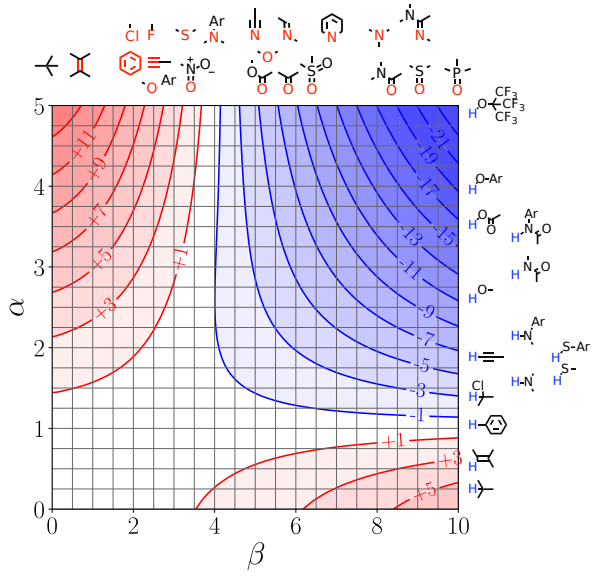


Fig. H.111: FGIP for 2,4-pentanedione at 298K.

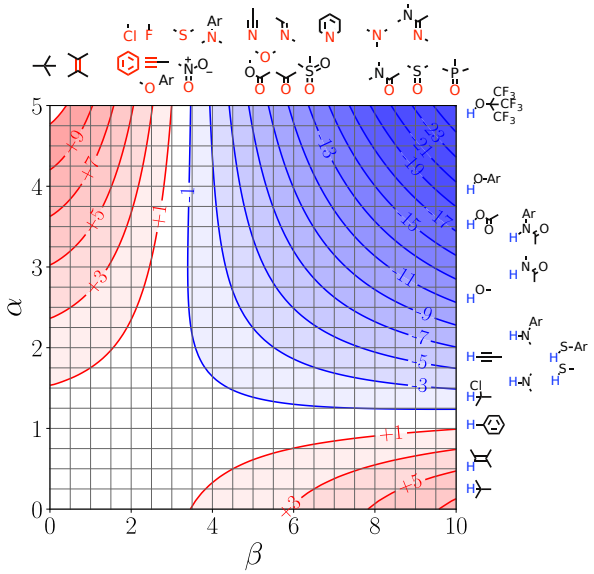


Fig. H.112: FGIP for 2,3-butanedione at 298K.

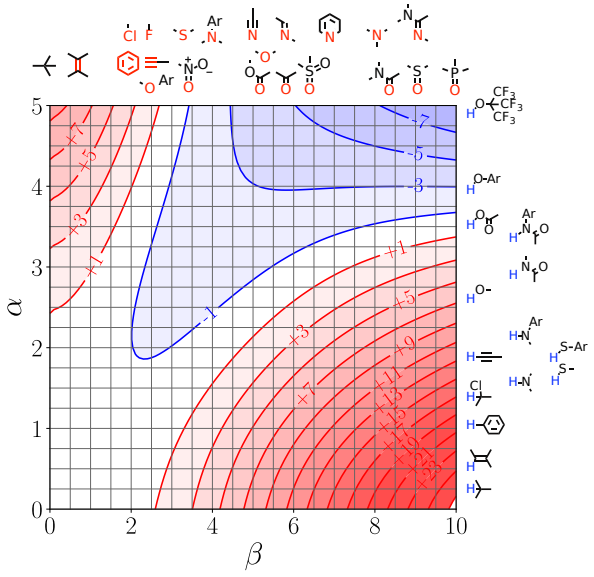


Fig. H.113: FGIP for formic acid at 298K.

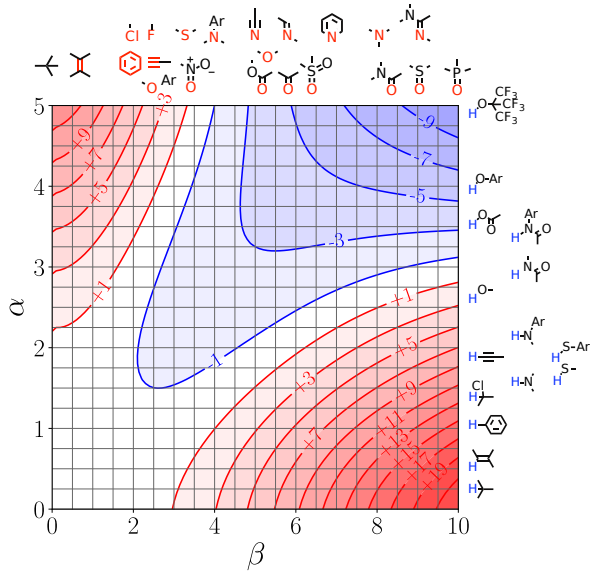


Fig. H.114: FGIP for acetic acid at 298K.

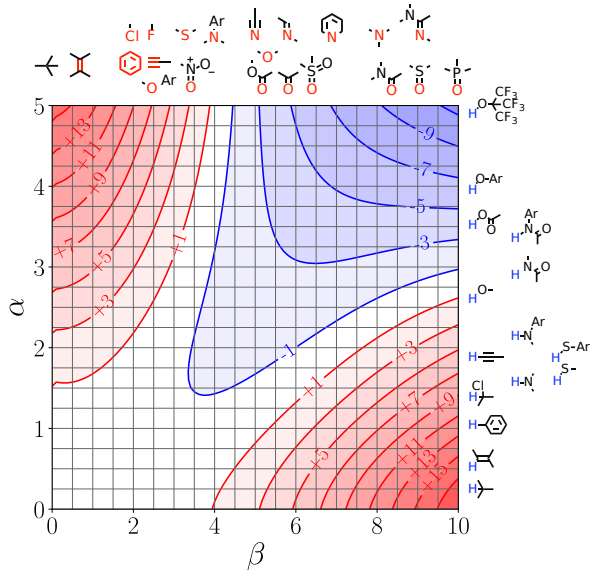


Fig. H.115: FGIP for propanoic acid at 298K.

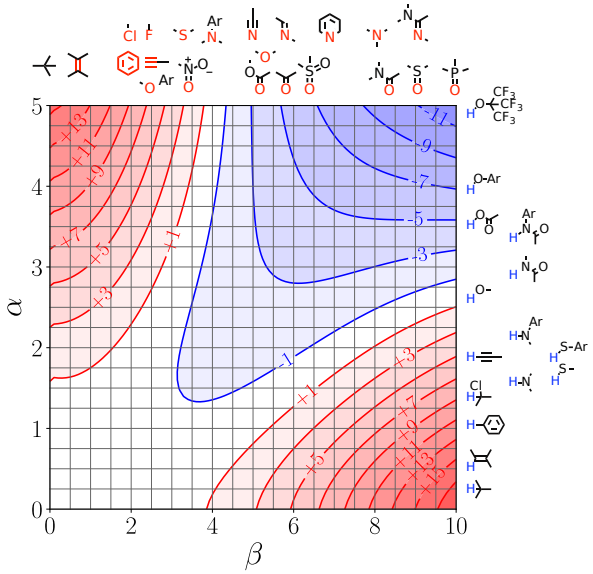


Fig. H.116: FGIP for butanoic acid at 298K.

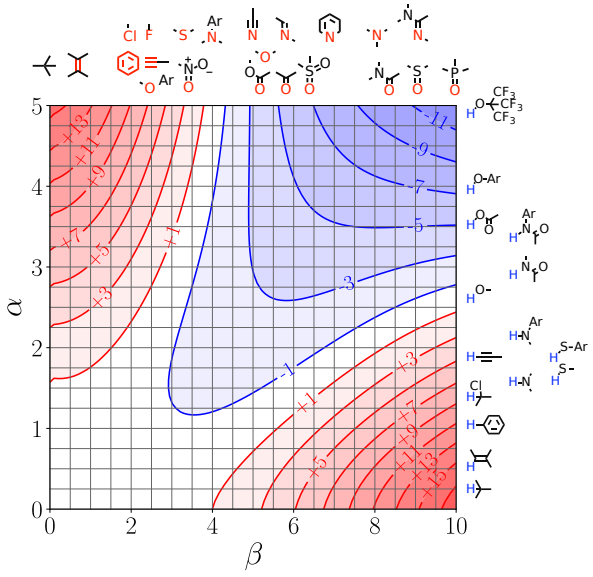


Fig. H.117: FGIP for pentanoic acid at 298K.

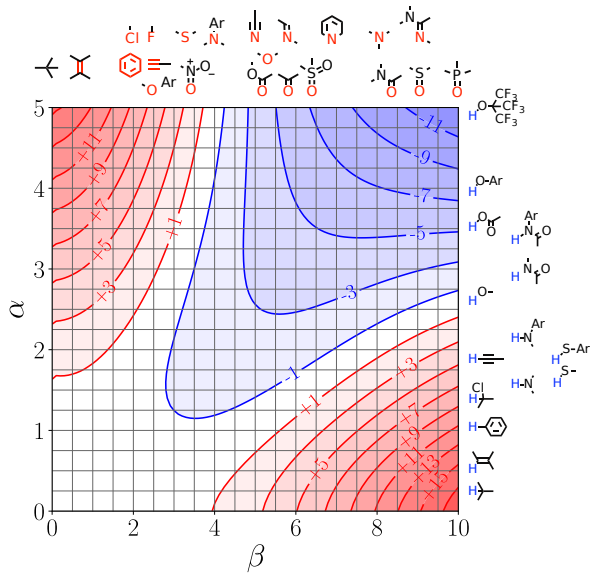


Fig. H.118: FGIP for hexanoic acid at 298K.

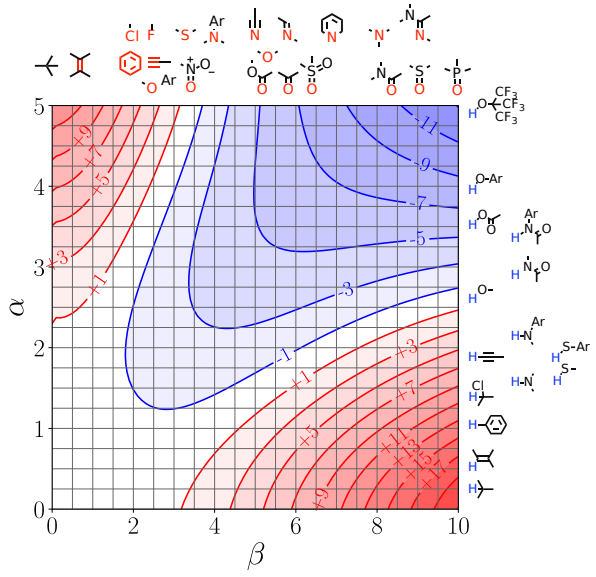


Fig. H.119: FGIP for heptanoic acid at 298K.

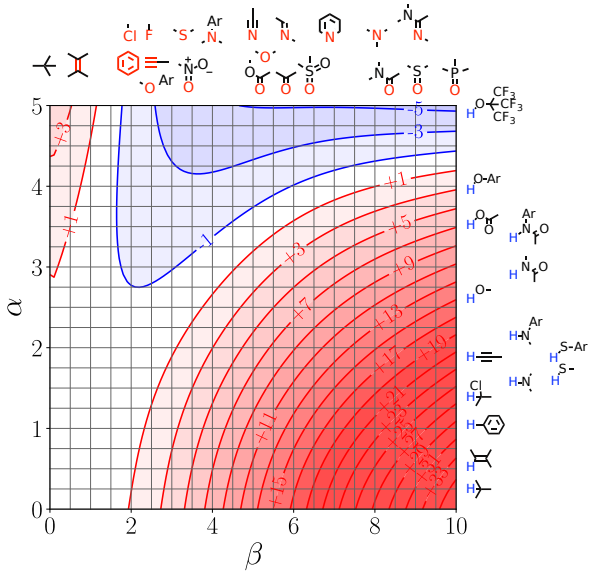


Fig. H.120: FGIP for dichloroacetic acid at 298K.

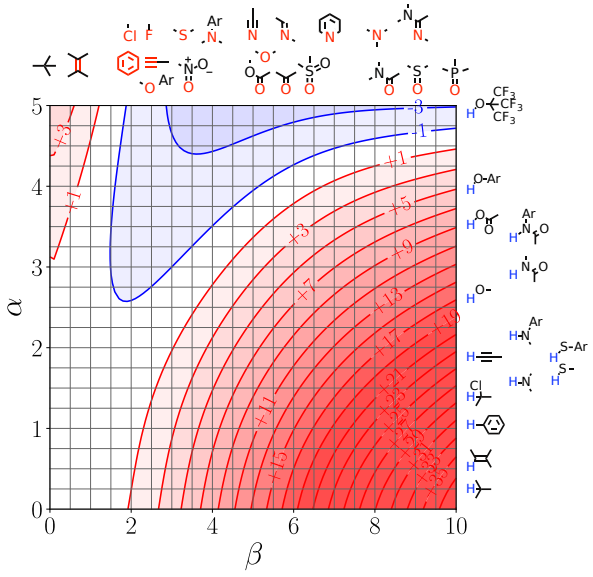


Fig. H.121: FGIP for trifluoroacetic acid at 298K.

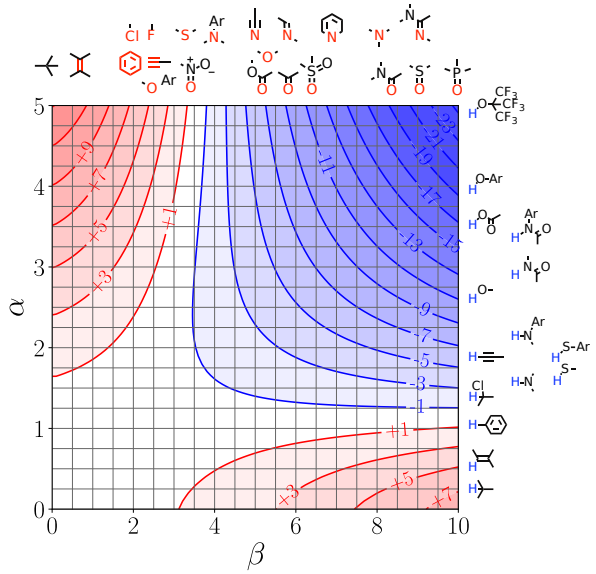


Fig. H.122: FGIP for acetic anhydride at 298K.

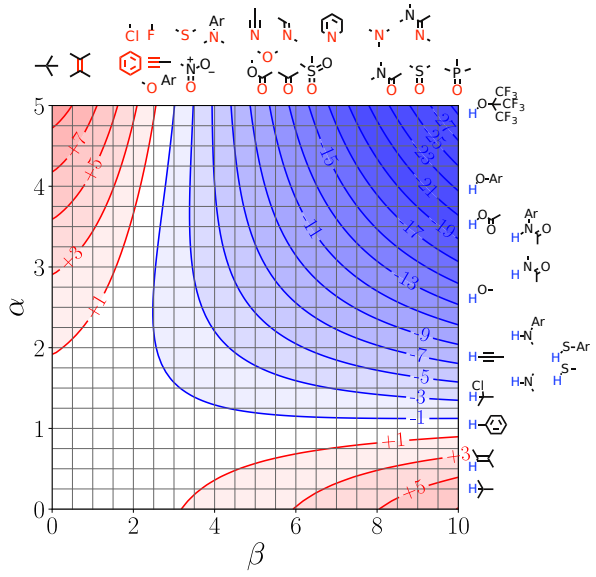


Fig. H.123: FGIP for benzoyl chloride at 298K.

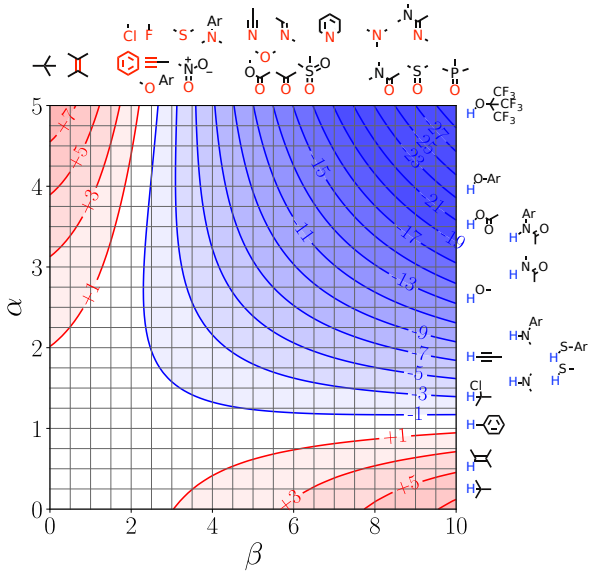


Fig. H.124: FGIP for benzoyl bromide at 298K.

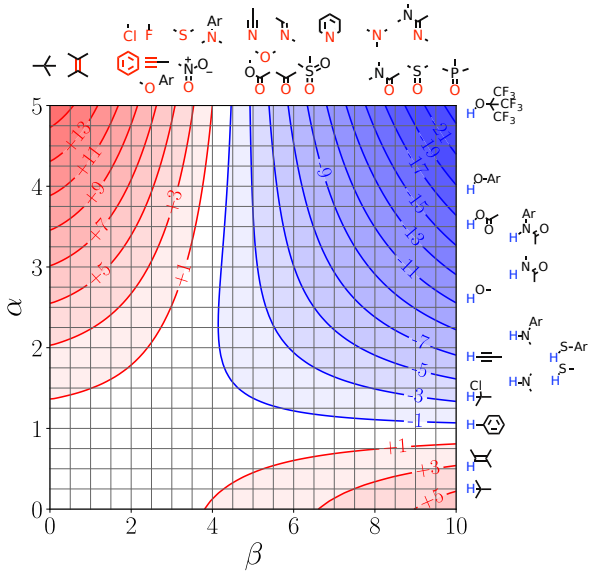


Fig. H.125: FGIP for methyl formate at 298K.

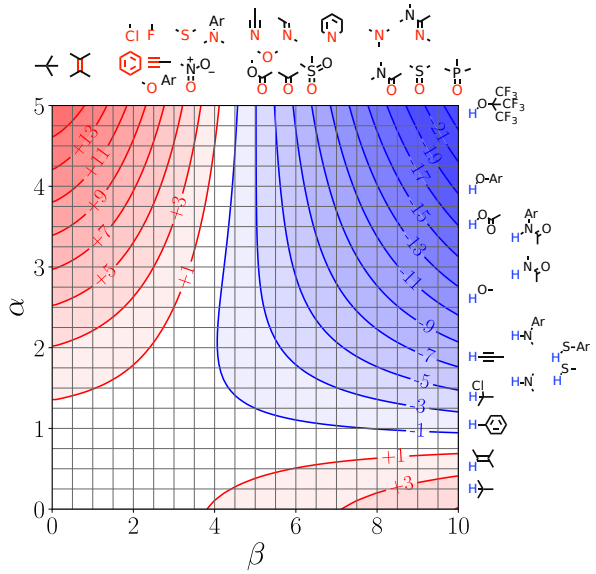


Fig. H.126: FGIP for ethyl formate at 298K.

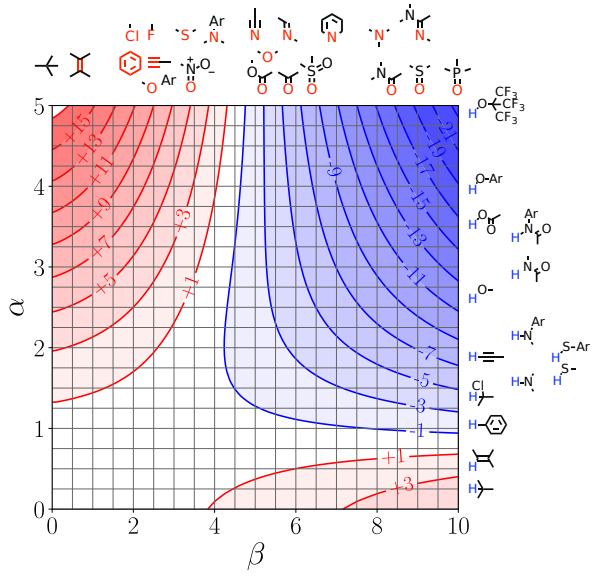


Fig. H.127: FGIP for methyl acetate at 298K.

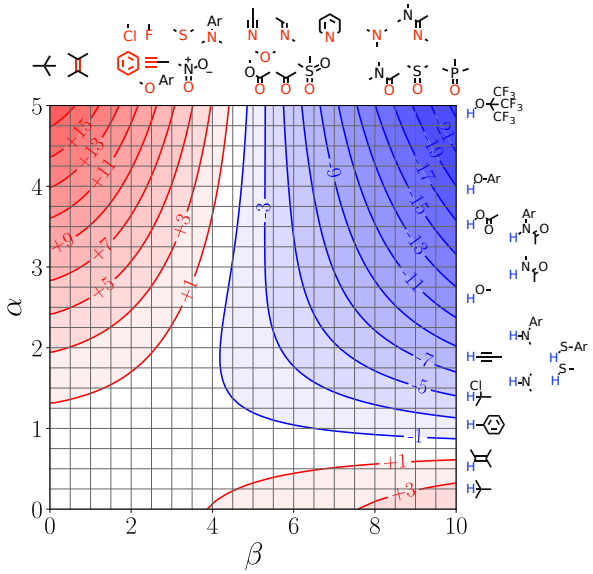


Fig. H.128: FGIP for ethyl acetate at 298K.

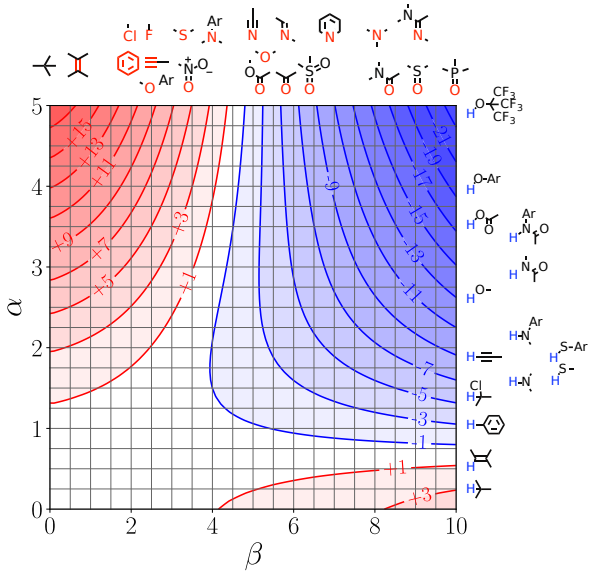


Fig. H.129: FGIP for n-propyl acetate at 298K.

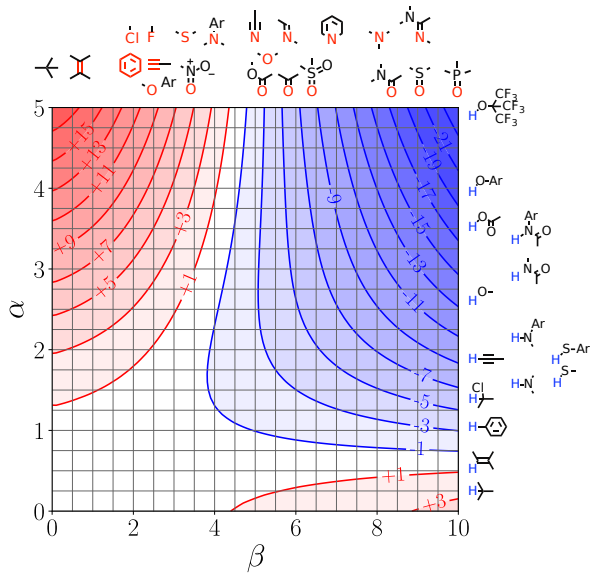


Fig. H.130: FGIP for butyl acetate at 298K.

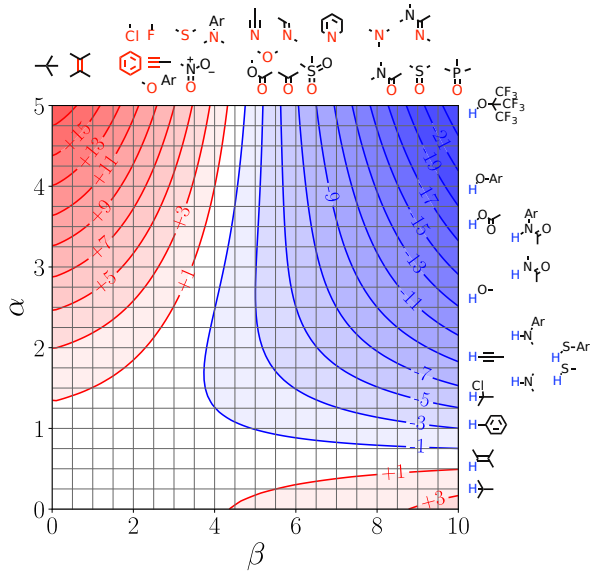


Fig. H.131: FGIP for isopentyl acetate at 298K.

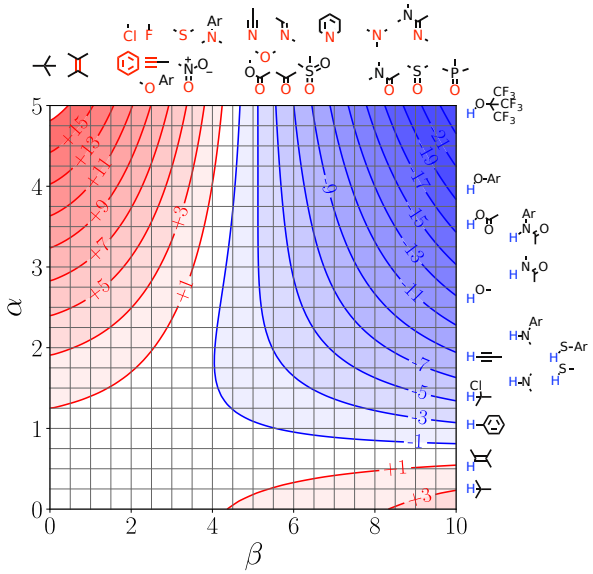


Fig. H.132: FGIP for methyl propionate at 298K.

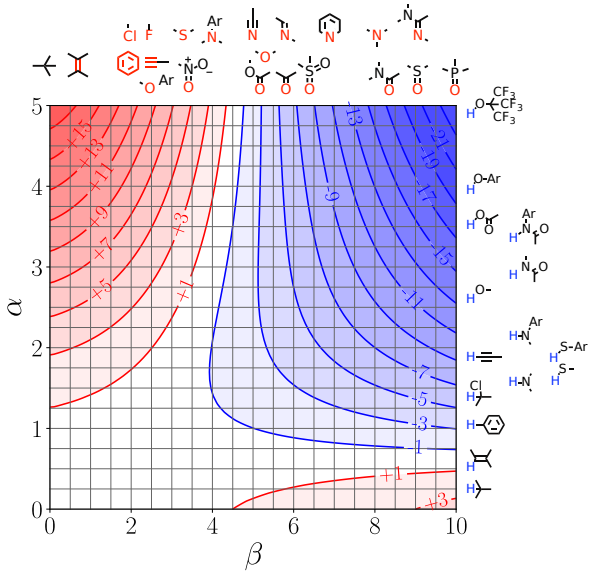


Fig. H.133: FGIP for ethyl propionate at 298K.

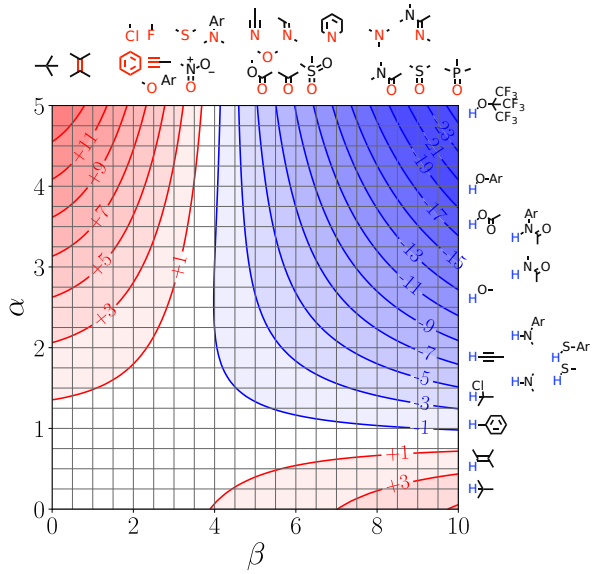


Fig. H.134: FGIP for dimethyl carbonate at 298K.

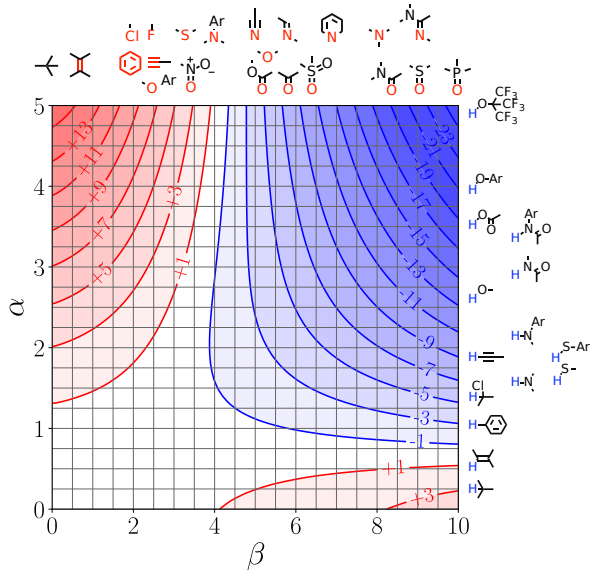


Fig. H.135: FGIP for diethyl carbonate at 298K.

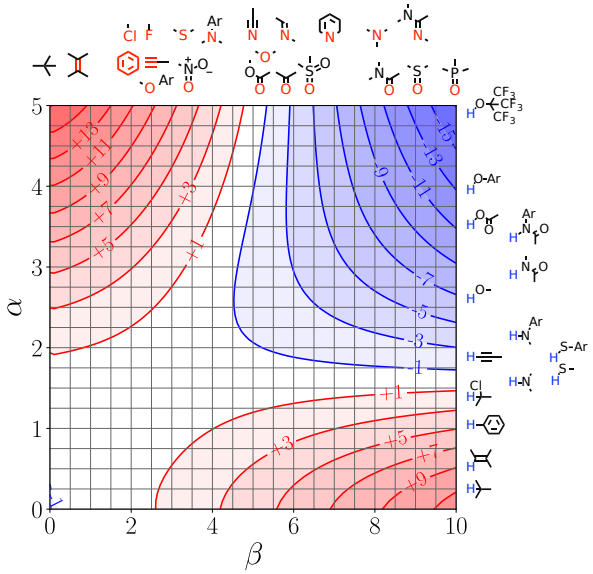


Fig. H.136: FGIP for ethylene carbonate at 298K.

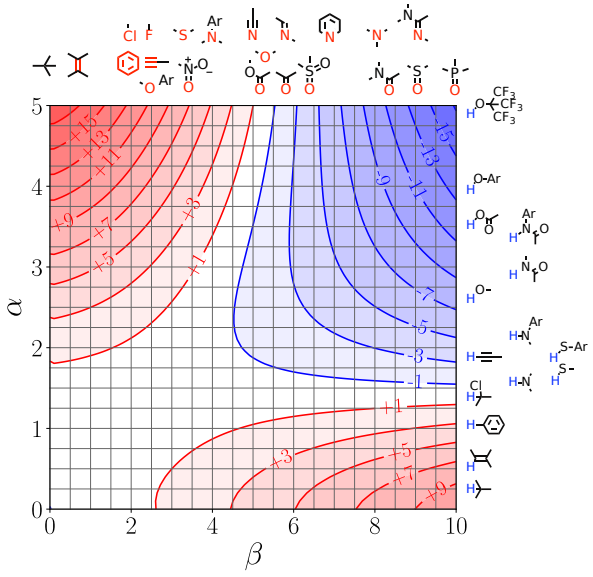


Fig. H.137: FGIP for 4-methyl-1,3-dioxolan-2-one at 298K.

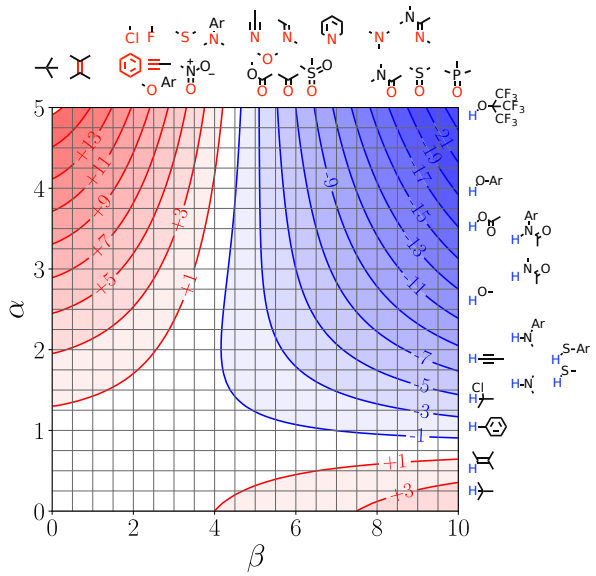


Fig. H.138: FGIP for diethyl malonate at 298K.

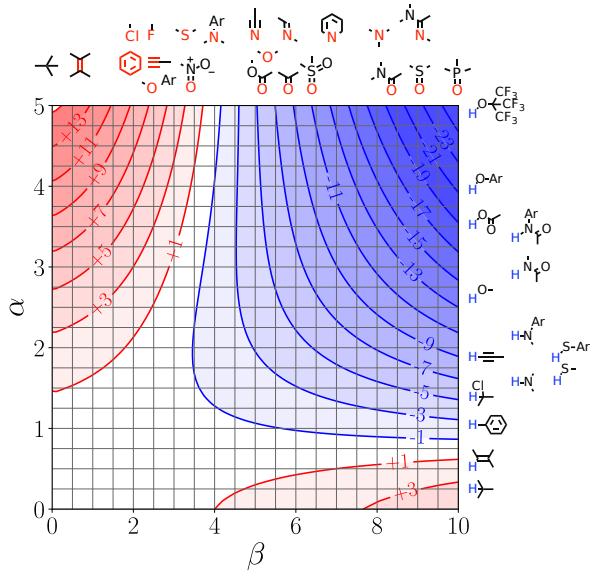


Fig. H.139: FGIP for methyl benzoate at 298K.

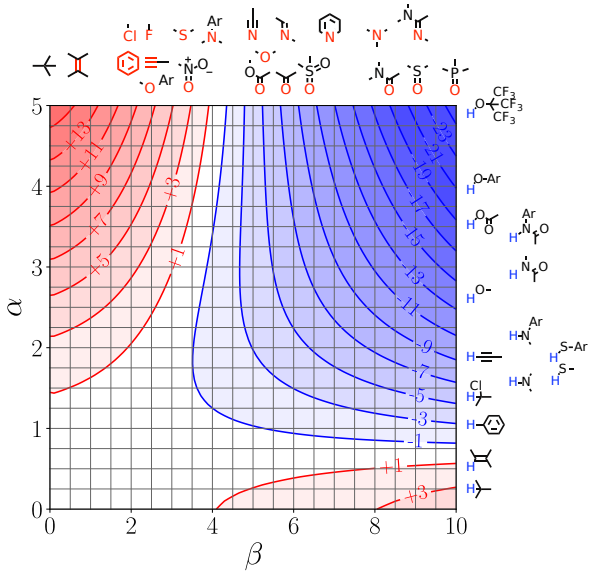


Fig. H.140: FGIP for ethyl benzoate at 298K.

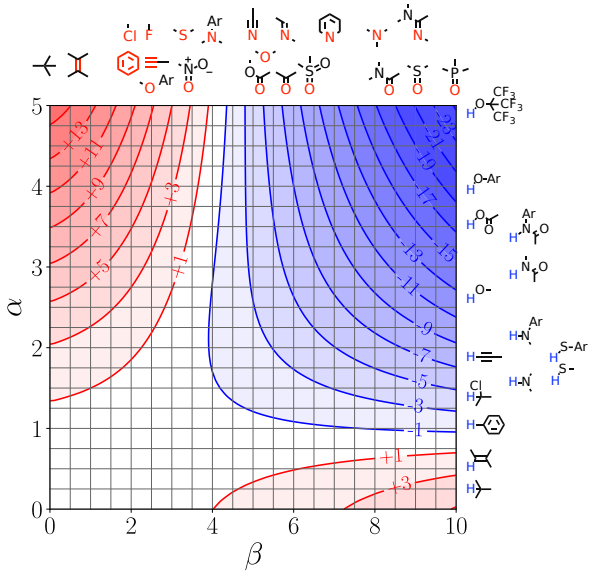


Fig. H.141: FGIP for dimethylphthalate at 298K.

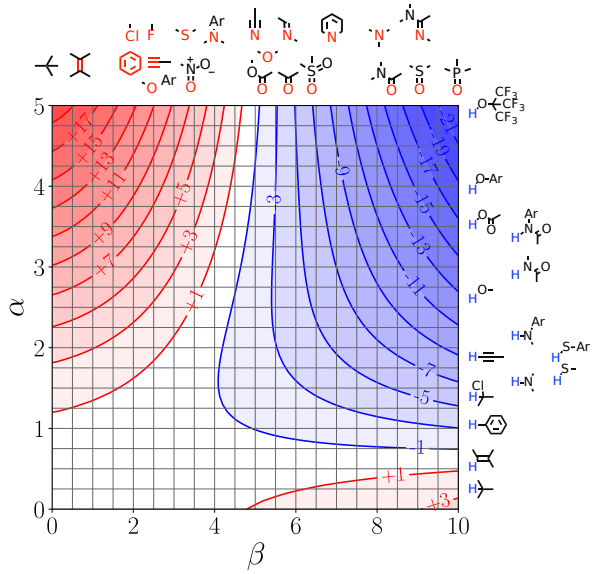


Fig. H.142: FGIP for di-n-butylorthophthalate at 298K.

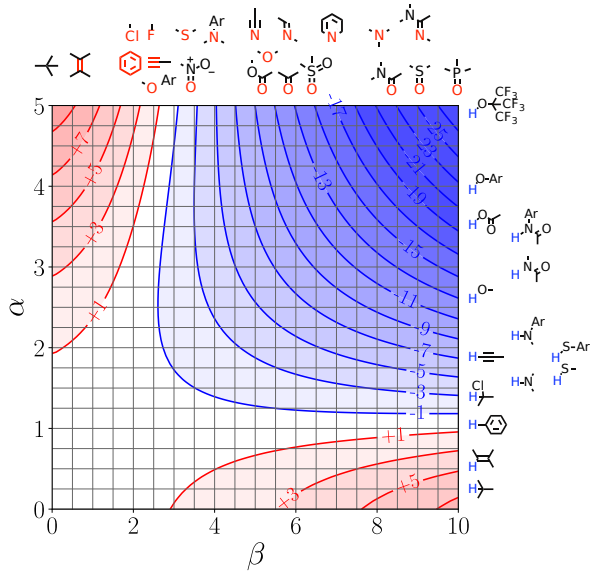


Fig. H.143: FGIP for ethyl chloroacetate at 298K.

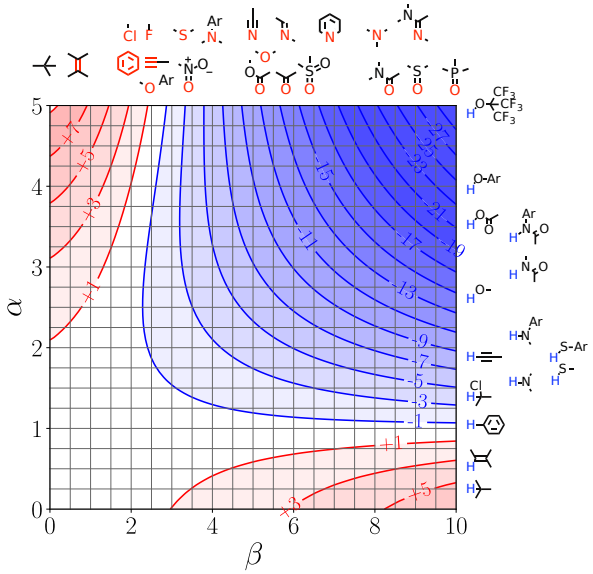


Fig. H.144: FGIP for ethyl trichloroacetate at 298K.

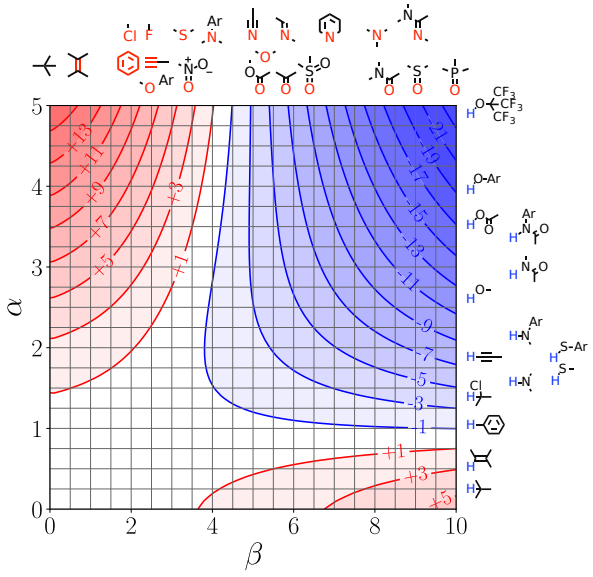


Fig. H.145: FGIP for ethyl acetoacetate at 298K.

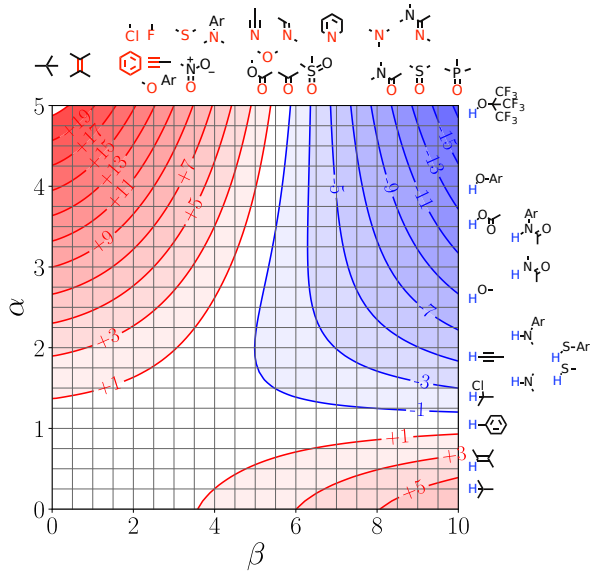


Fig. H.146: FGIP for gamma-butyrolactone at 298K.

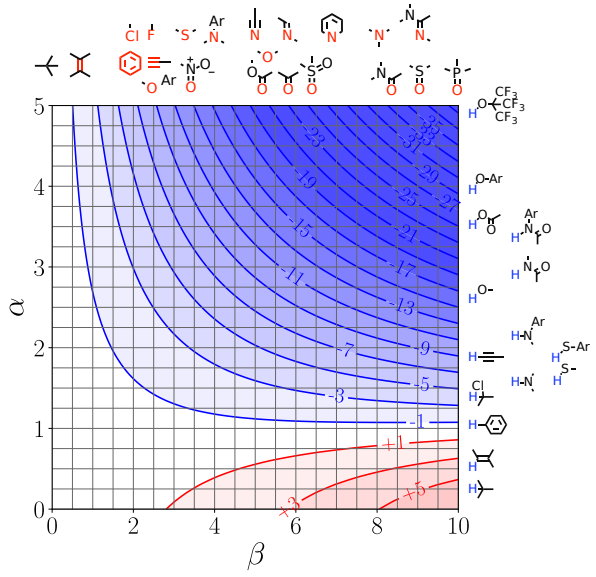


Fig. H.147: FGIP for n-perfluorohexane at 298K.

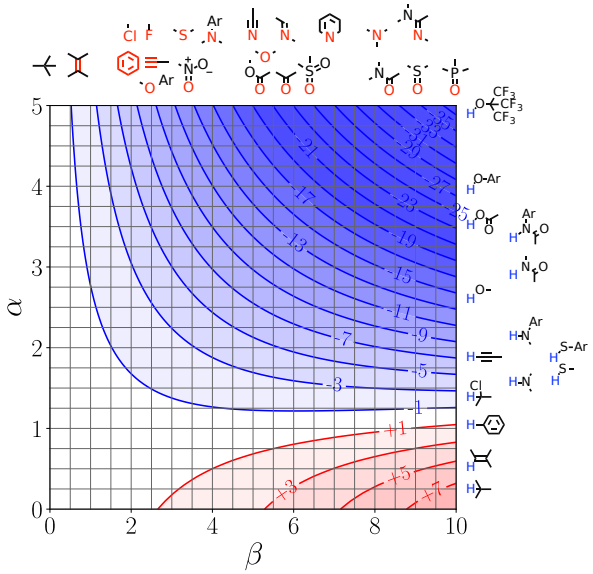


Fig. H.148: FGIP for perfluoromethylcyclohexane at 298K.

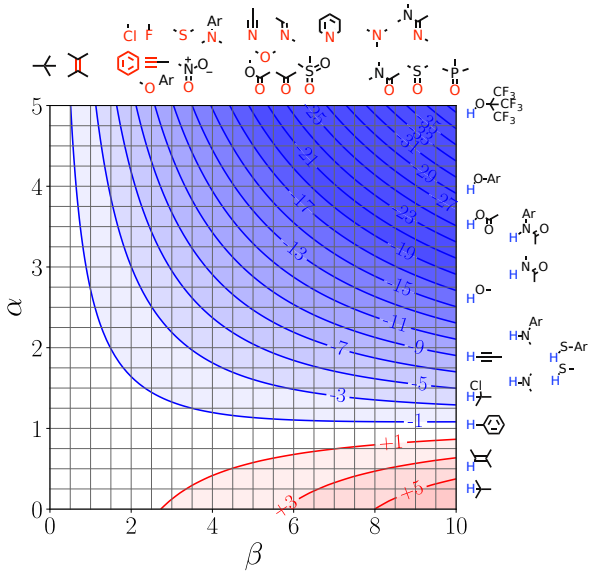


Fig. H.149: FGIP for perfluoroheptane at 298K.

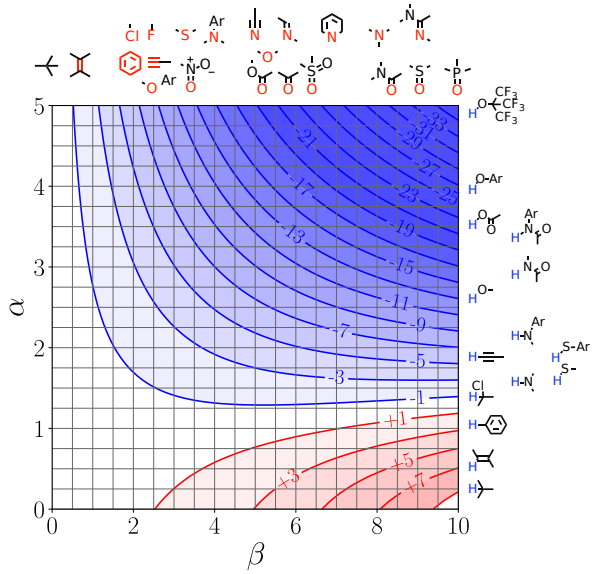


Fig. H.150: FGIP for cis-perfluorodecalin at 298K.

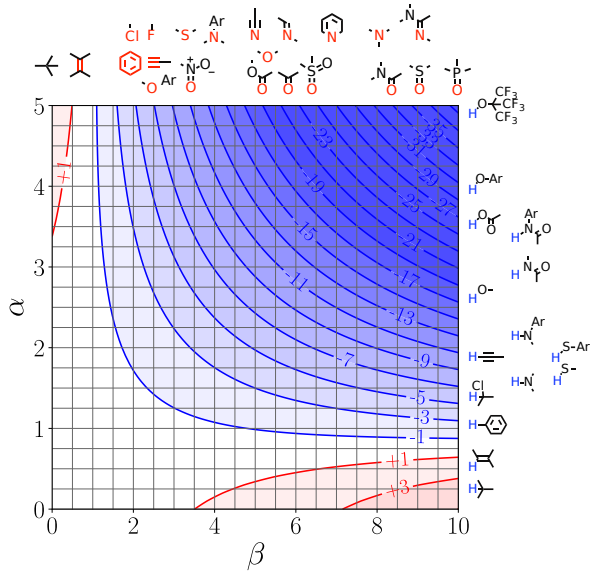


Fig. H.151: FGIP for fluorobenzene at 298K.

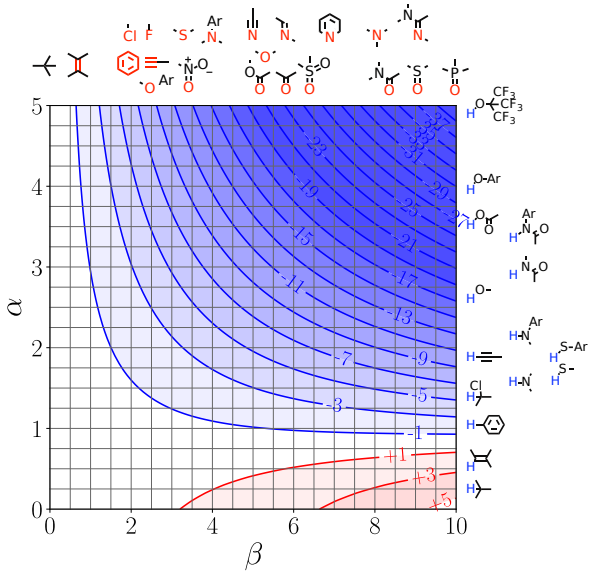


Fig. H.152: FGIP for hexafluorobenzene at 298K.

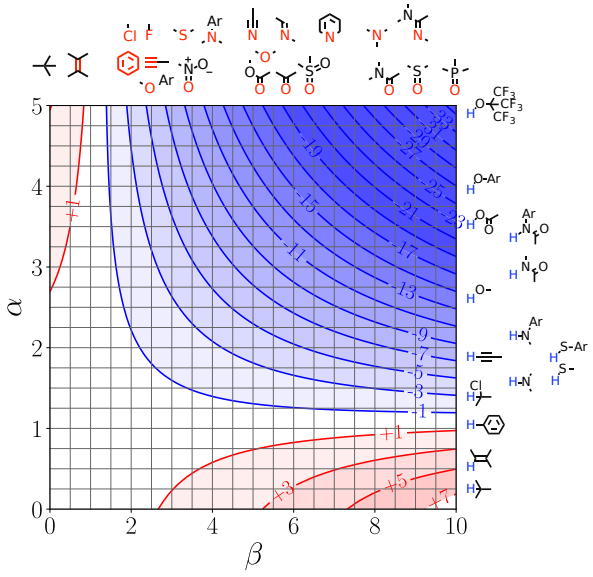


Fig. H.153: FGIP for 1,4-dichlorobutane at 298K.

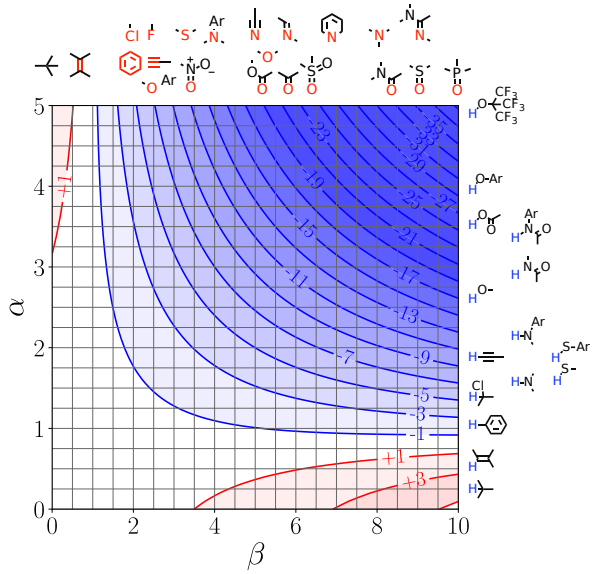


Fig. H.154: FGIP for chlorobenzene at 298K.

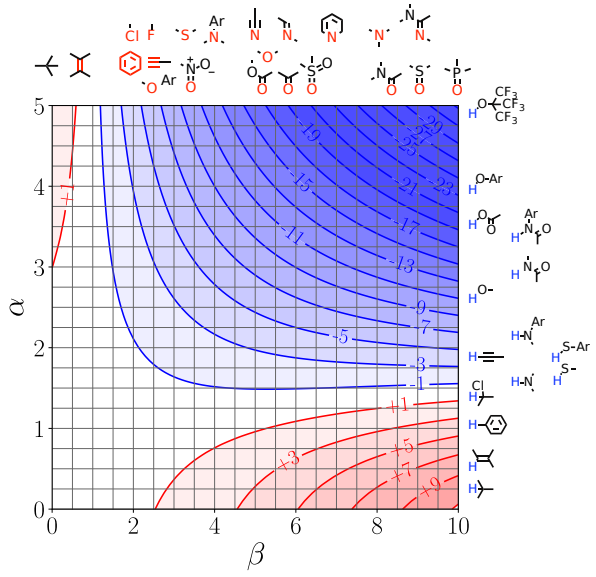


Fig. H.155: FGIP for dichloromethane at 298K.

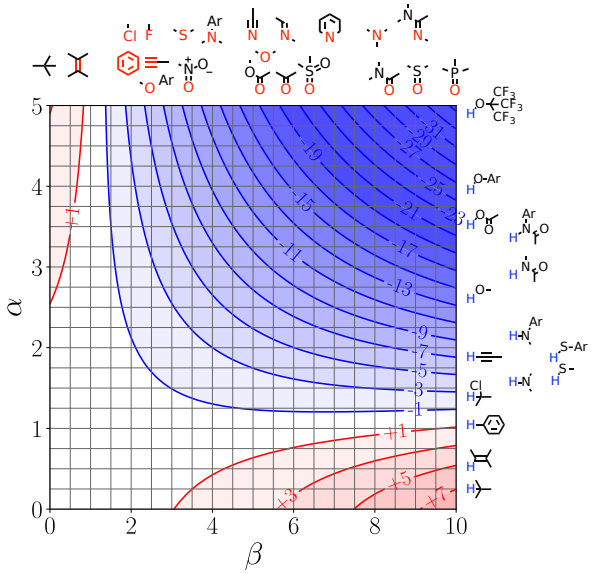


Fig. H.156: FGIP for 1,1-dichloroethane at 298K.

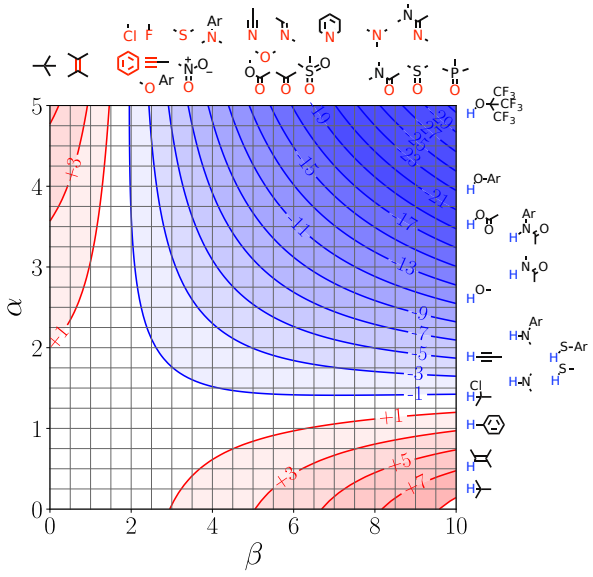


Fig. H.157: FGIP for 1,2-dichloroethane at 298K.

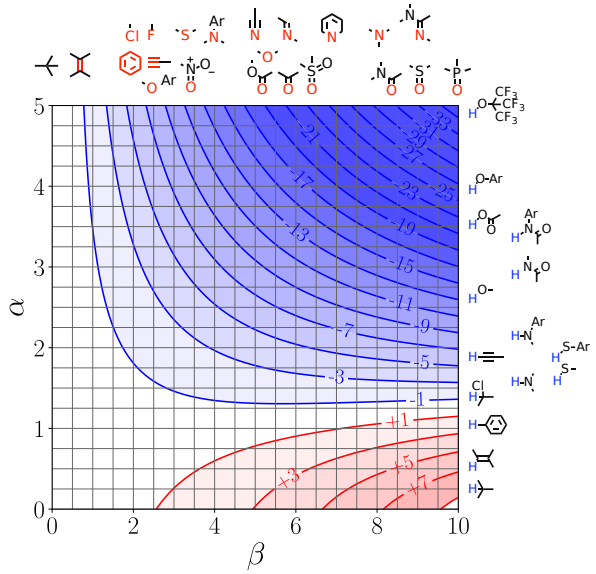


Fig. H.158: FGIP for trans-1,2-dichloroethylene at 298K.

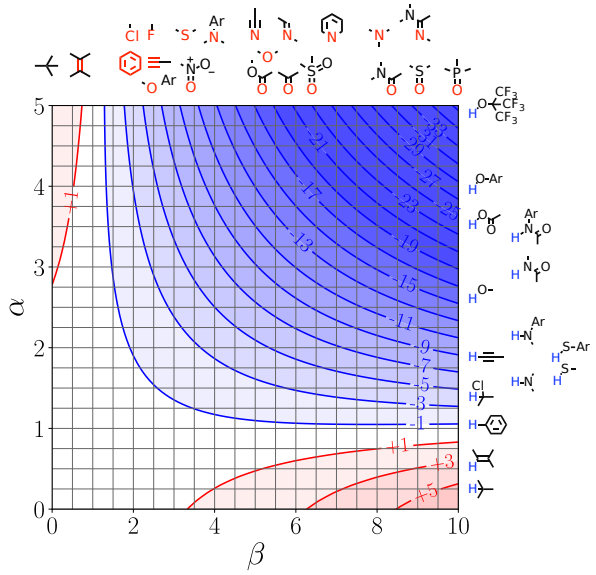


Fig. H.159: FGIP for ortho-dichlorobenzene at 298K.

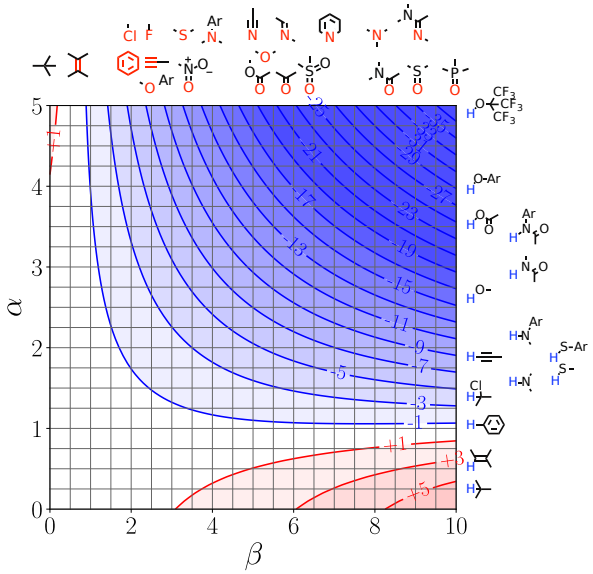


Fig. H.160: FGIP for meta-dichlorobenzene at 298K.

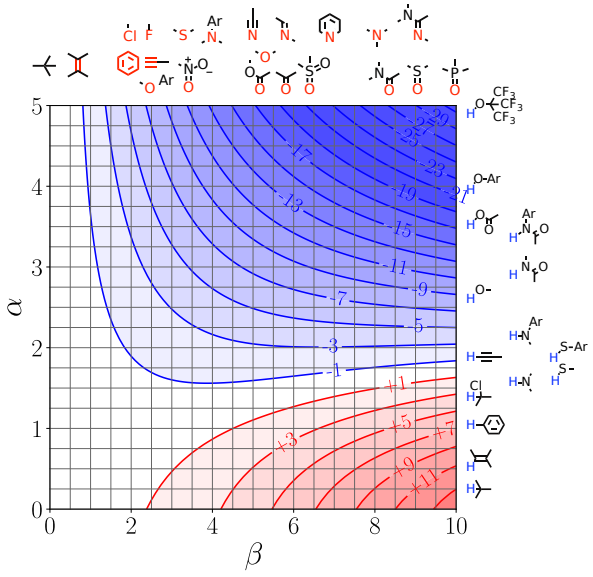


Fig. H.161: FGIP for chloroform at 298K.

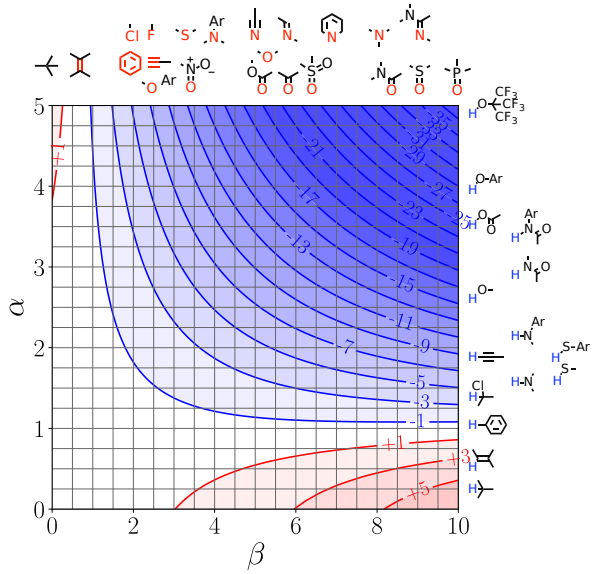


Fig. H.162: FGIP for 1,1,1-trichloroethane at 298K.

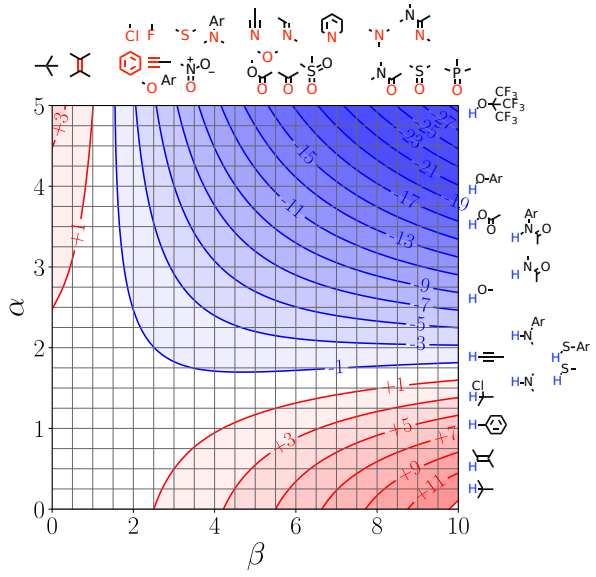


Fig. H.163: FGIP for 1,1,2-trichloroethane at 298K.

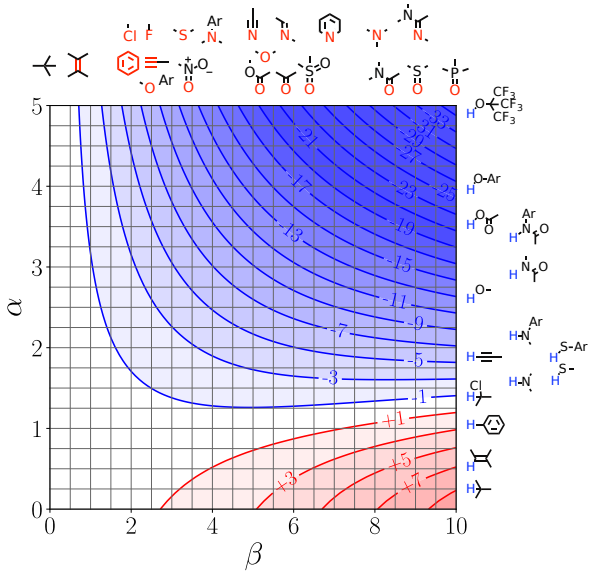


Fig. H.164: FGIP for trichloroethylene at 298K.

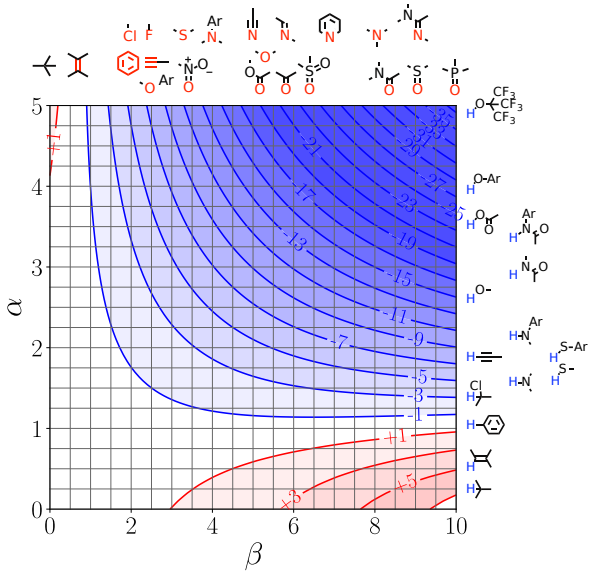


Fig. H.165: FGIP for 1,2,4-trichlorobenzene at 298K.

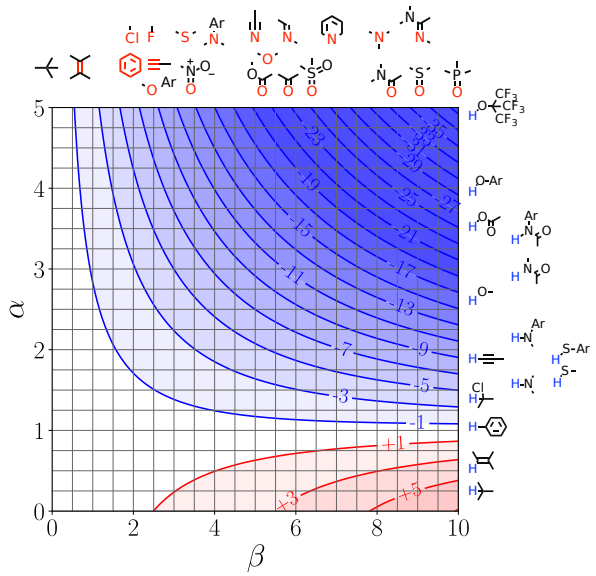


Fig. H.166: FGIP for carbon tetrachloride at 298K.

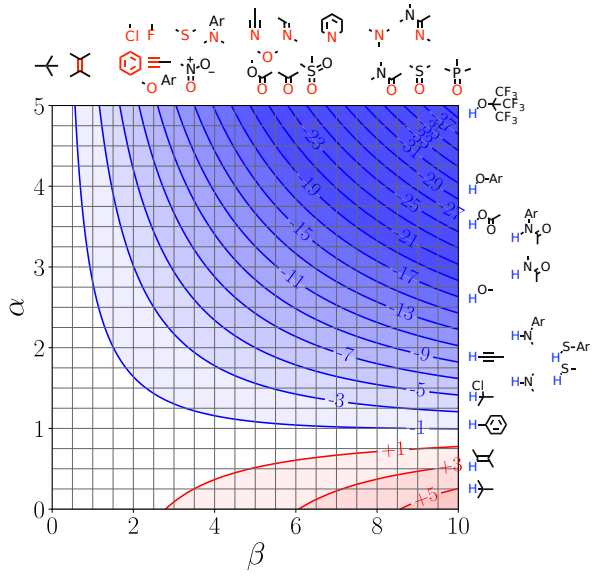


Fig. H.167: FGIP for tetrachloroethylene at 298K.

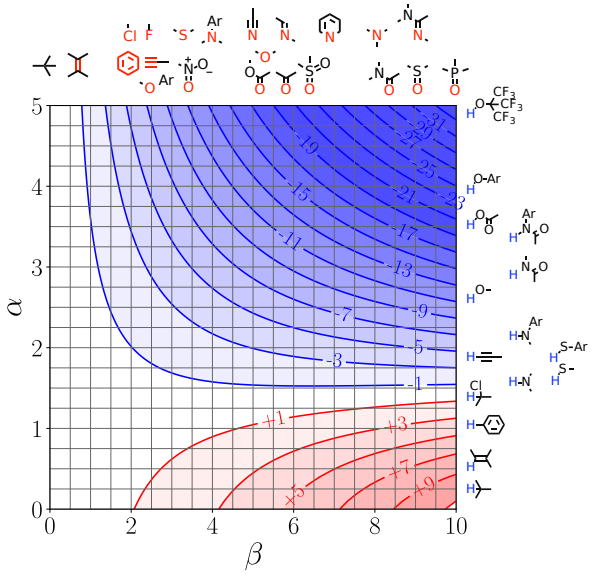


Fig. H.168: FGIP for 1,1,2,2-tetrachloroethane at 298K.

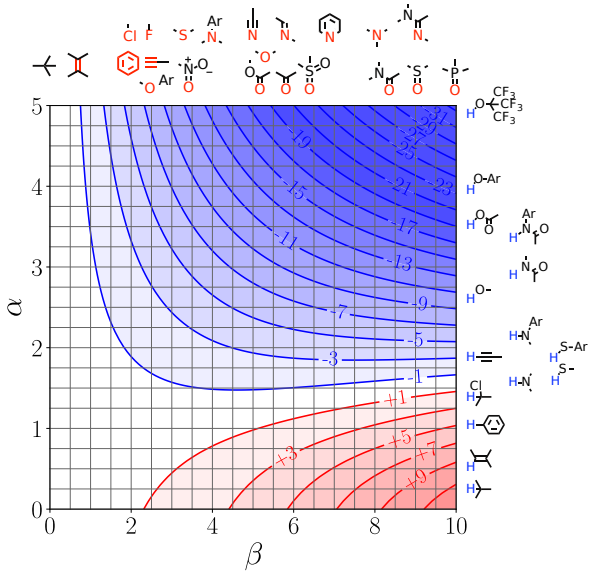


Fig. H.169: FGIP for pentachloroethane at 298K.

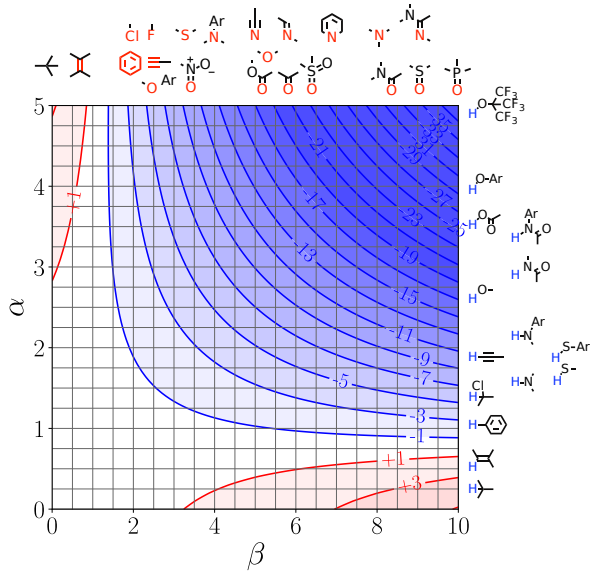


Fig. H.170: FGIP for 1-bromobutane at 298K.

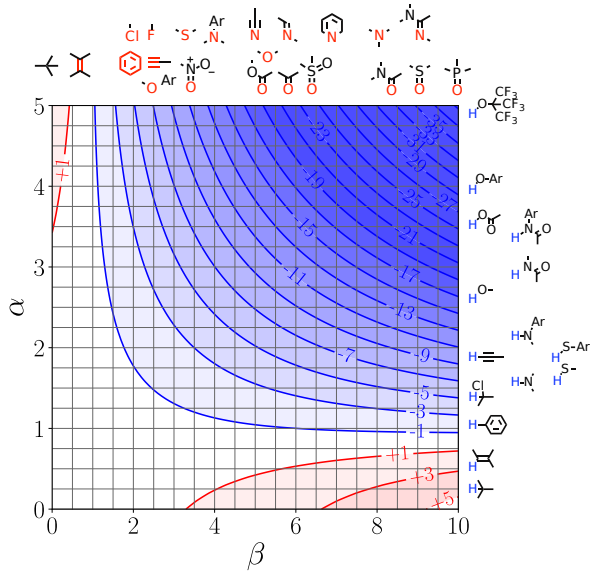


Fig. H.171: FGIP for bromobenzene at 298K.

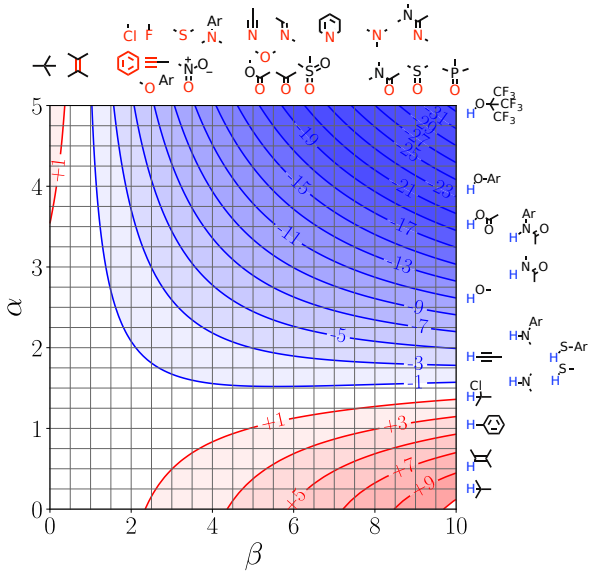


Fig. H.172: FGIP for dibromomethane at 298K.

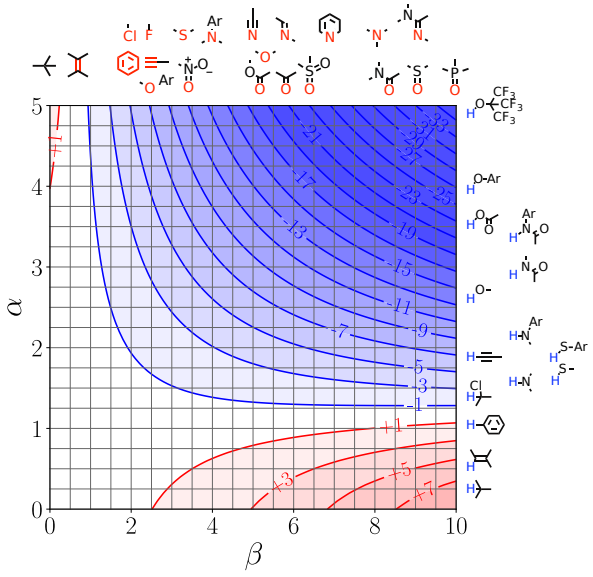


Fig. H.173: FGIP for 1,2-dibromoethane at 298K.

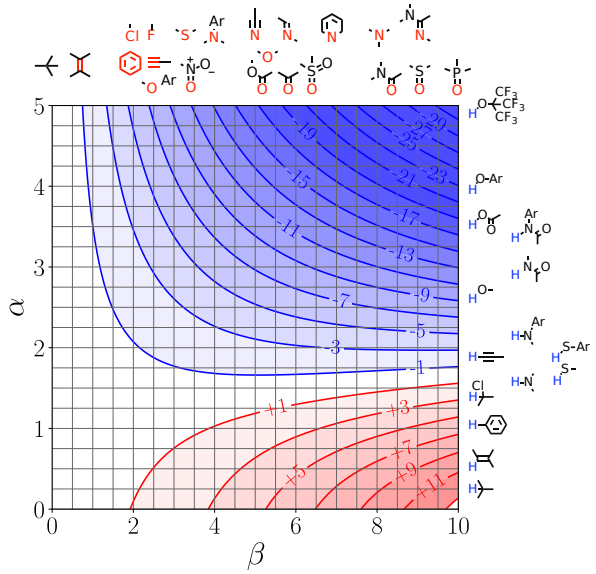


Fig. H.174: FGIP for bromoform at 298K.

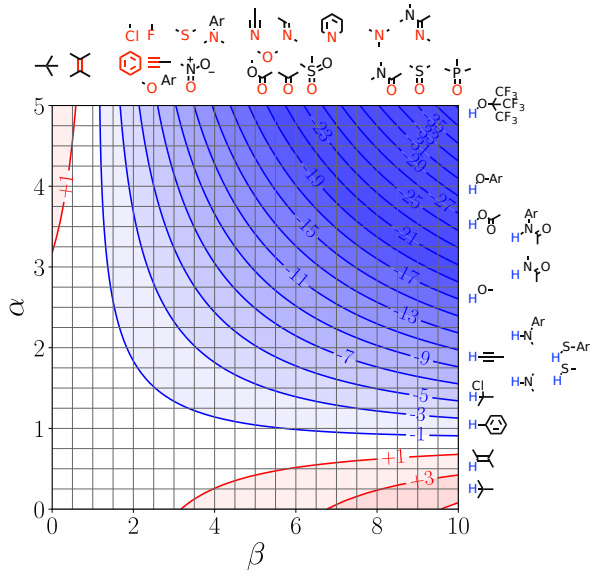


Fig. H.175: FGIP for n-butyl iodide at 298K.

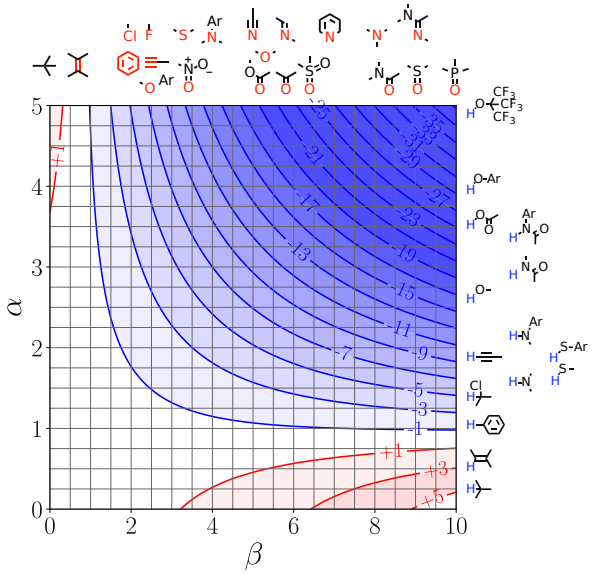


Fig. H.176: FGIP for iodobenzene at 298K.

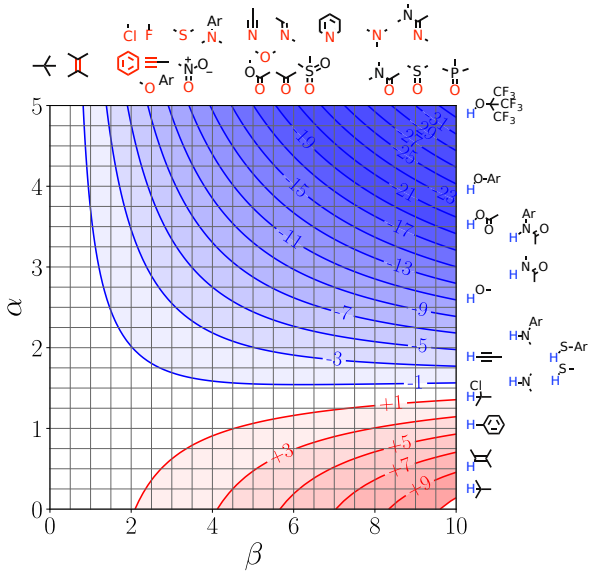


Fig. H.177: FGIP for methylene iodide at 298K.

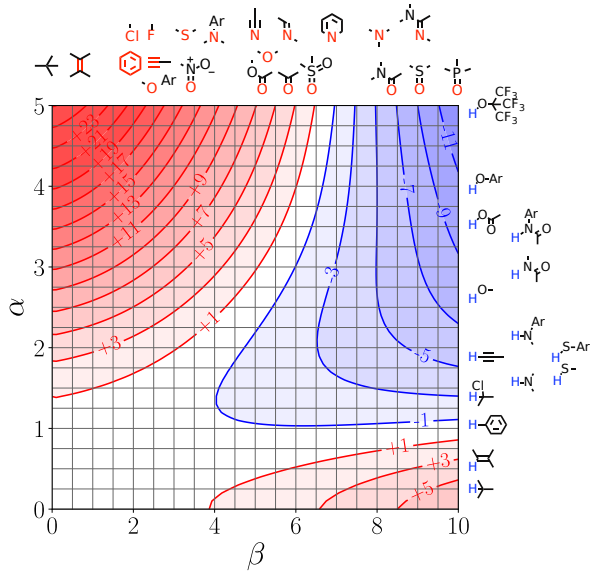


Fig. H.178: FGIP for n-butylamine at 298K.

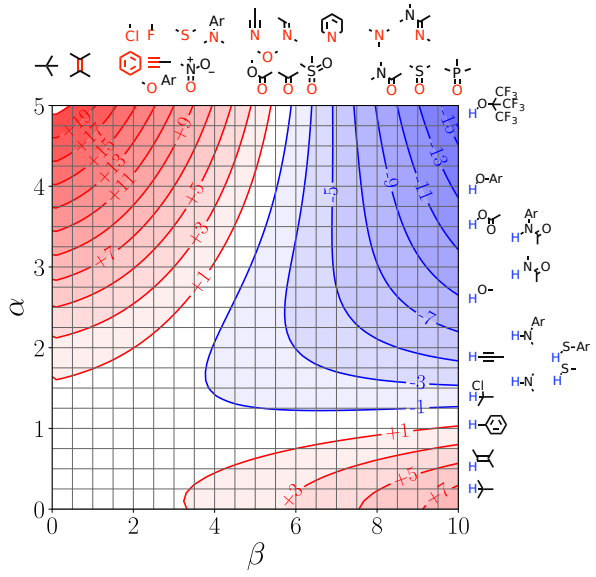


Fig. H.179: FGIP for benzylamine at 298K.

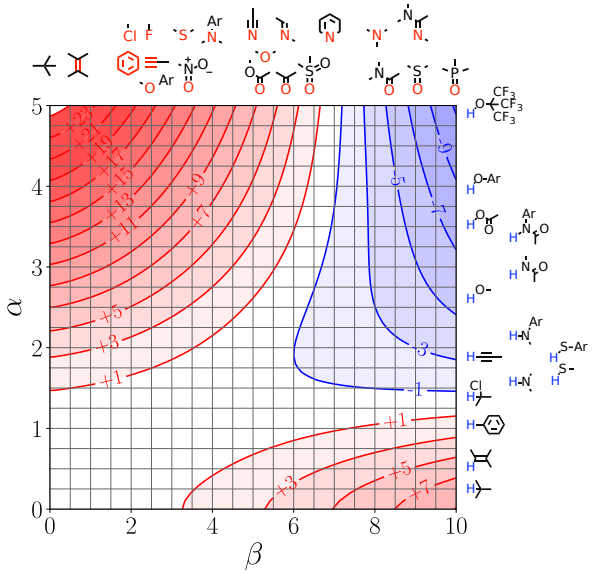


Fig. H.180: FGIP for ethylenediamine at 298K.

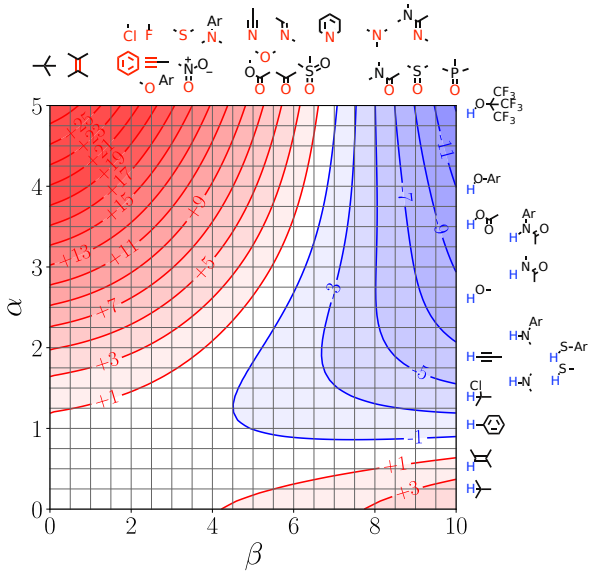


Fig. H.181: FGIP for diethylamine at 298K.

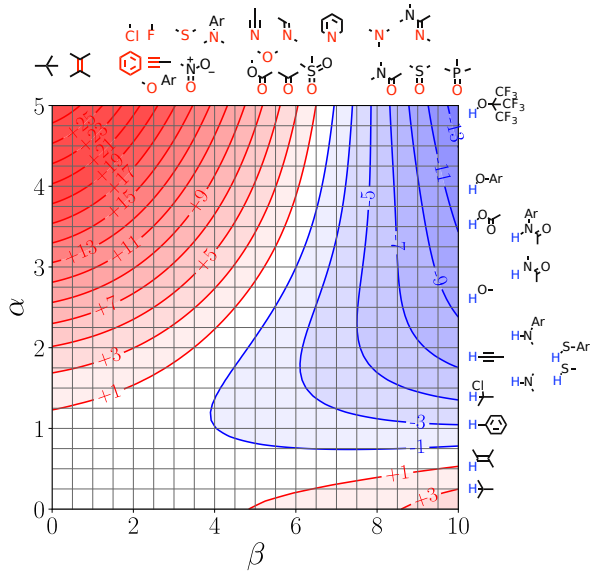


Fig. H.182: FGIP for di-n-butylamine at 298K.

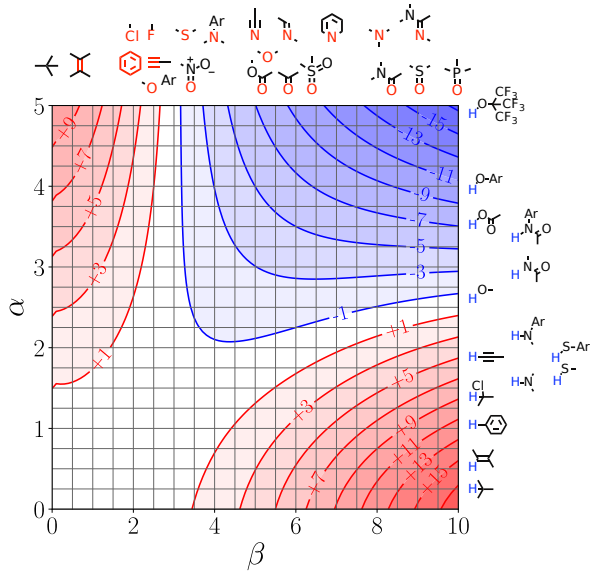


Fig. H.183: FGIP for pyrrole at 298K.

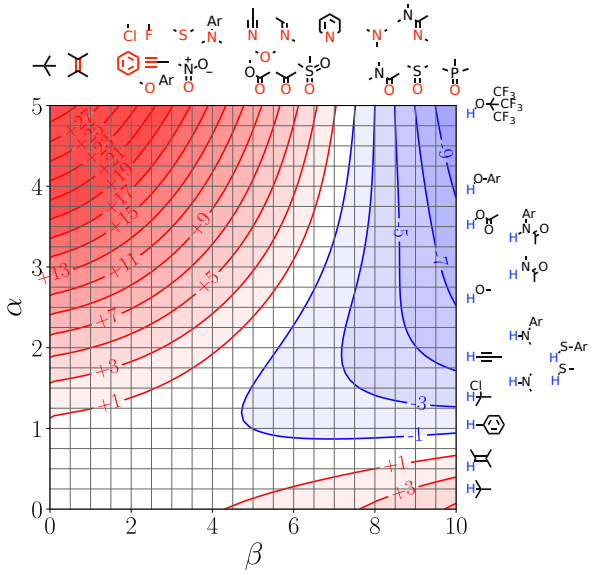


Fig. H.184: FGIP for pyrrolidine at 298K.

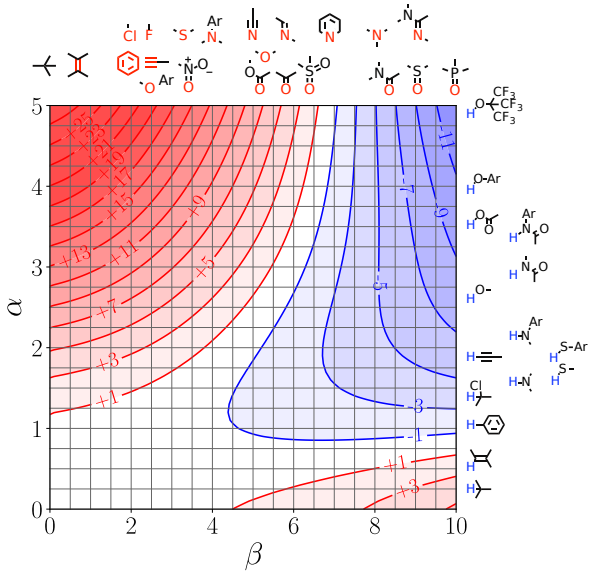


Fig. H.185: FGIP for piperidine at 298K.

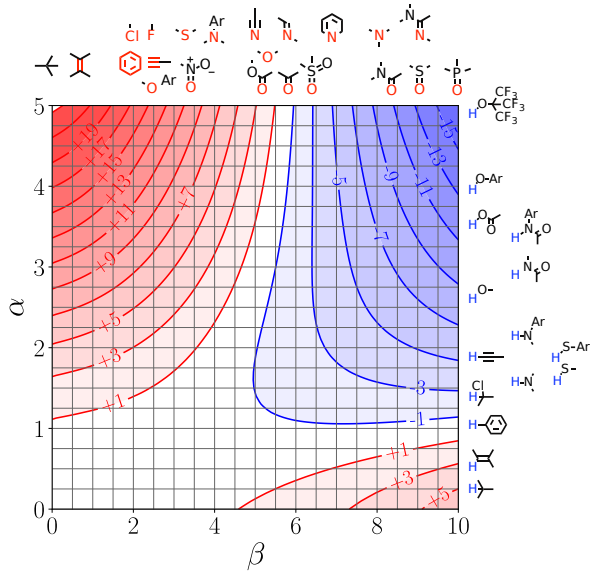


Fig. H.186: FGIP for morpholine at 298K.

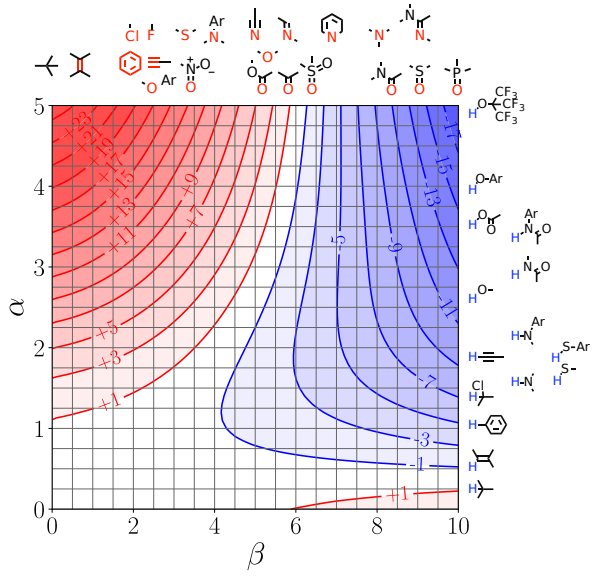


Fig. H.187: FGIP for triethylamine at 298K.

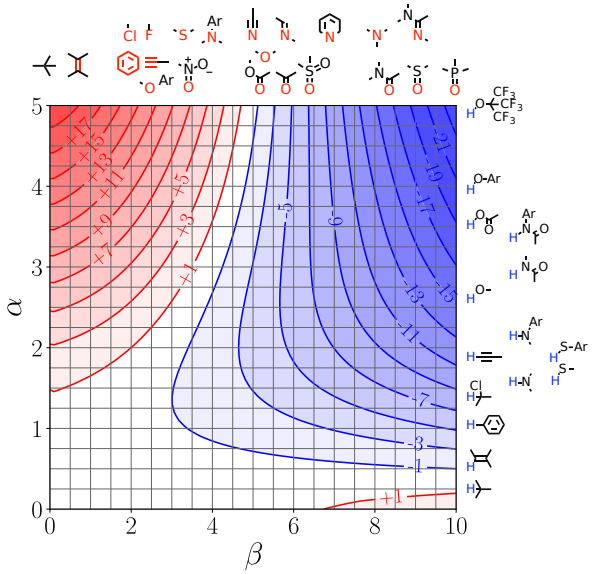


Fig. H.188: FGIP for tri-(n-butyl)amine at 298K.

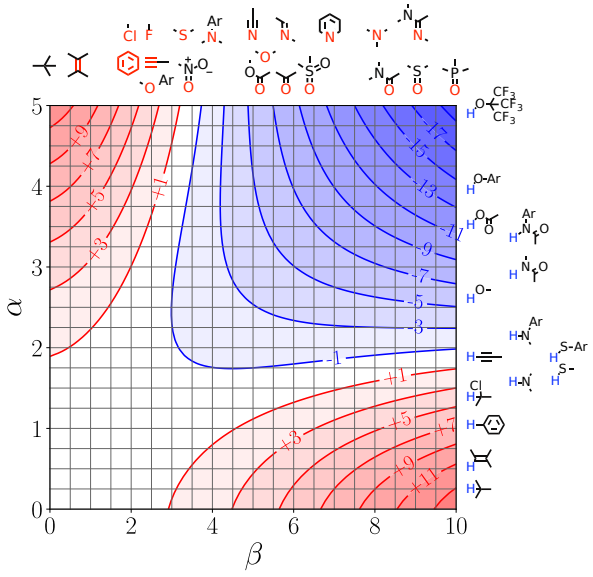


Fig. H.189: FGIP for aniline at 298K.

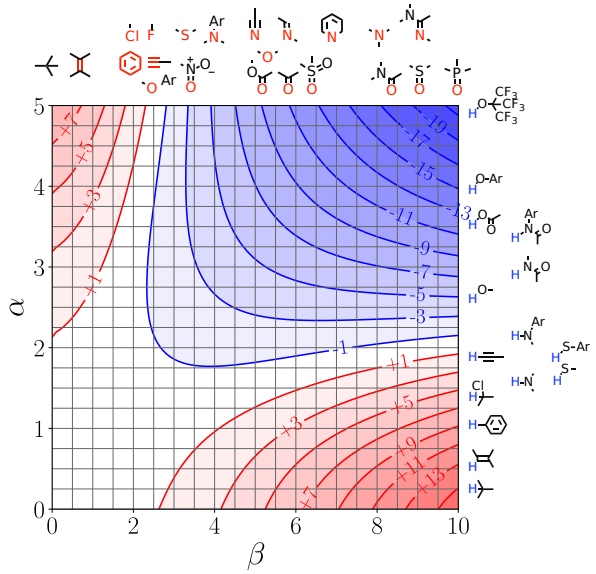


Fig. H.190: FGIP for o-chloroaniline at 298K.

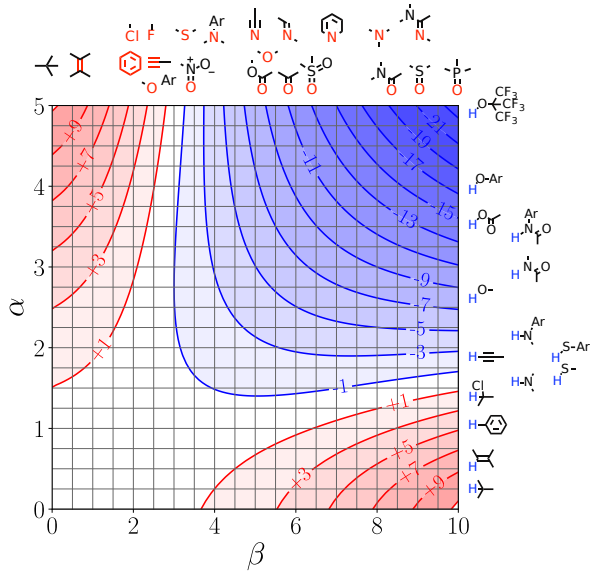


Fig. H.191: FGIP for methylphenylamine at 298K.

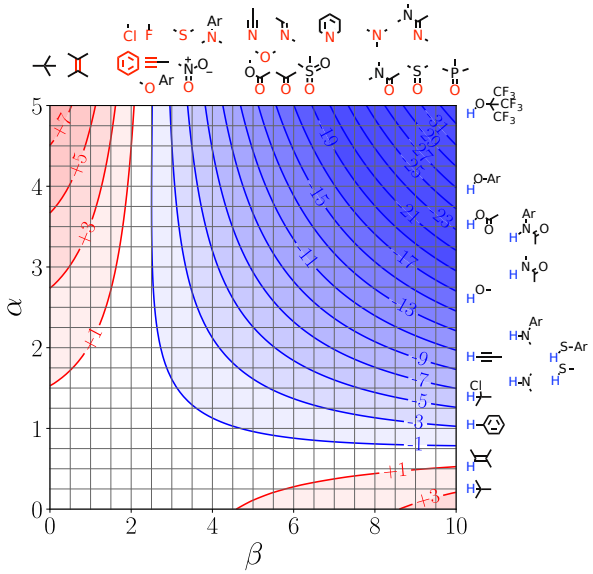


Fig. H.192: FGIP for N,N-dimethylaniline at 298K.

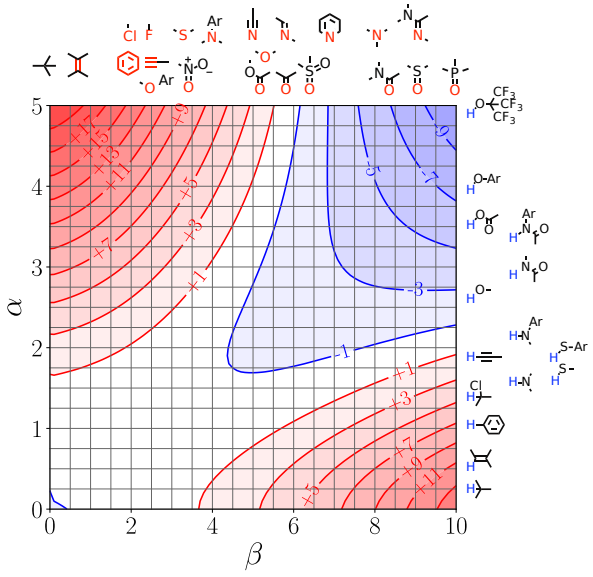


Fig. H.193: FGIP for aminoethanol at 298K.

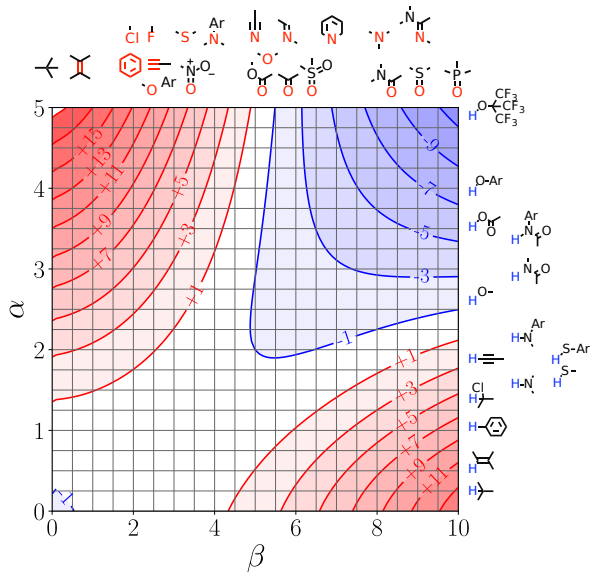


Fig. H.194: FGIP for diethanolamine at 298K.

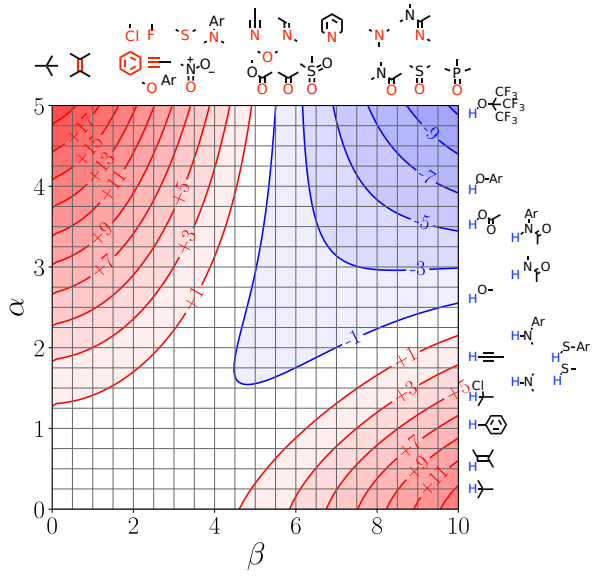


Fig. H.195: FGIP for triethanolamine at 298K.

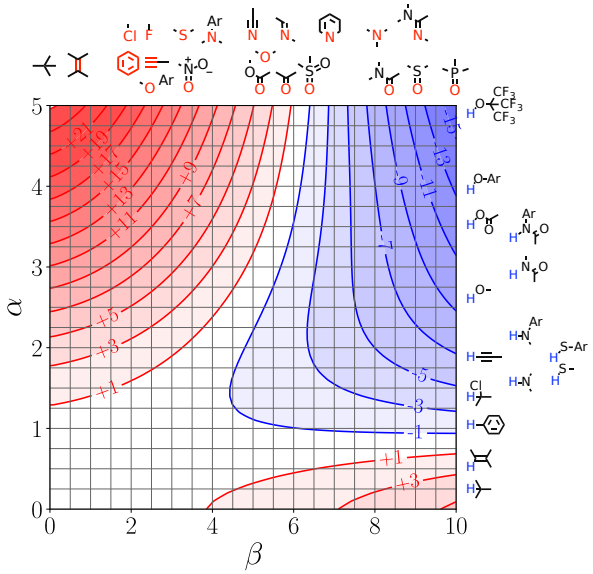


Fig. H.196: FGIP for pyridine at 298K.

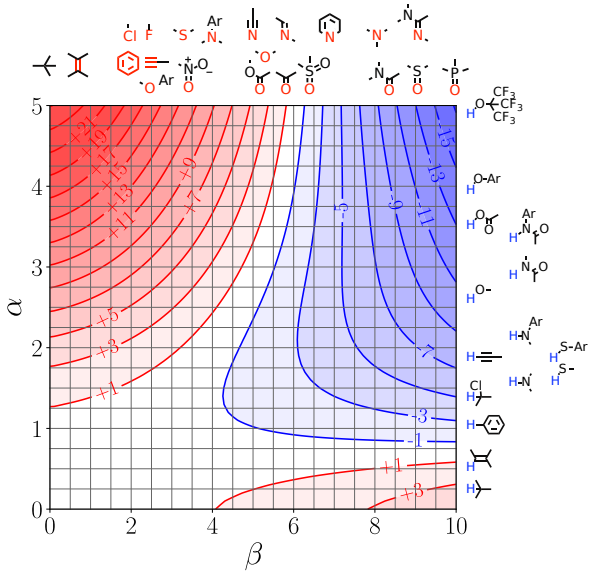


Fig. H.197: FGIP for 2-methylpyridine at 298K.

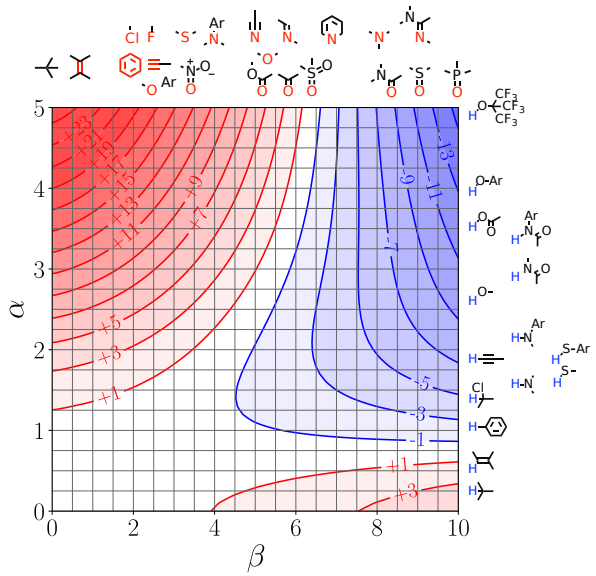


Fig. H.198: FGIP for 3-methylpyridine at 298K.

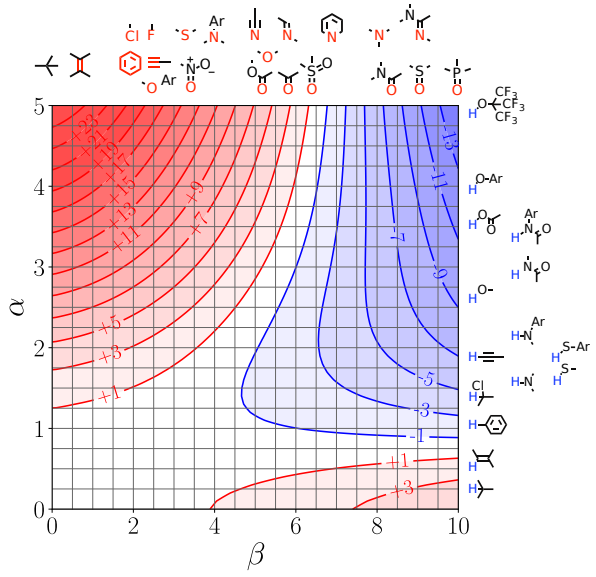


Fig. H.199: FGIP for 4-methylpyridine at 298K.

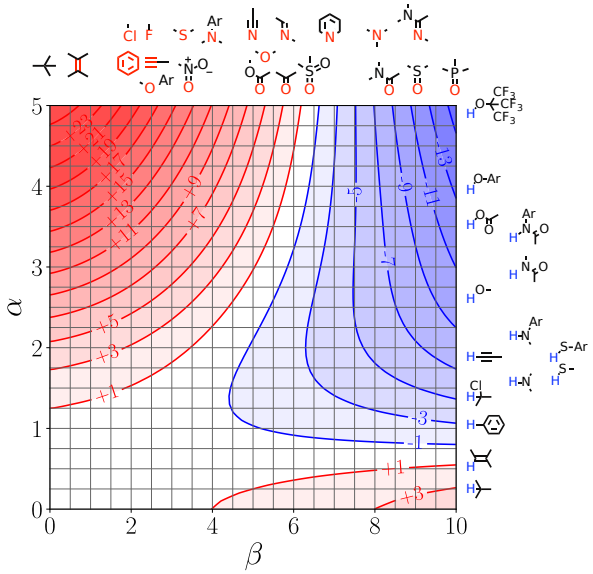


Fig. H.200: FGIP for 2,4-dimethylpyridine at 298K.

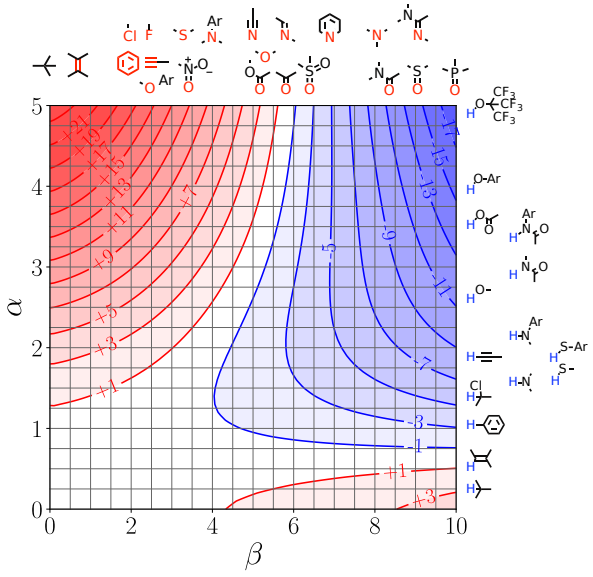


Fig. H.201: FGIP for 2,6-dimethylpyridine at 298K.

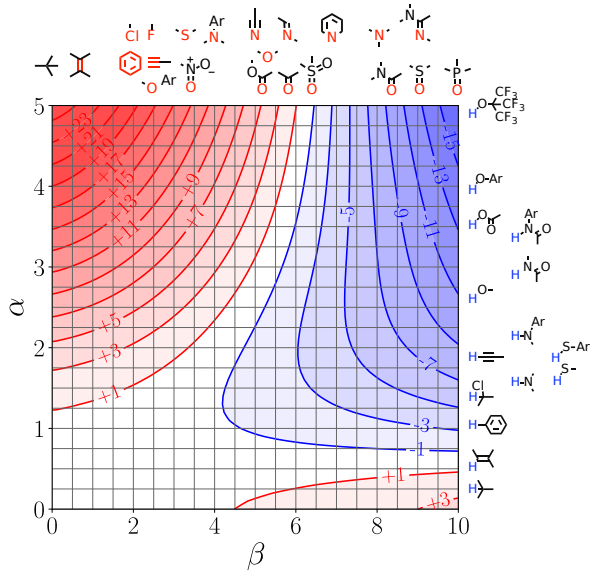


Fig. H.202: FGIP for 2,4,6-trimethylpyridine at 298K.

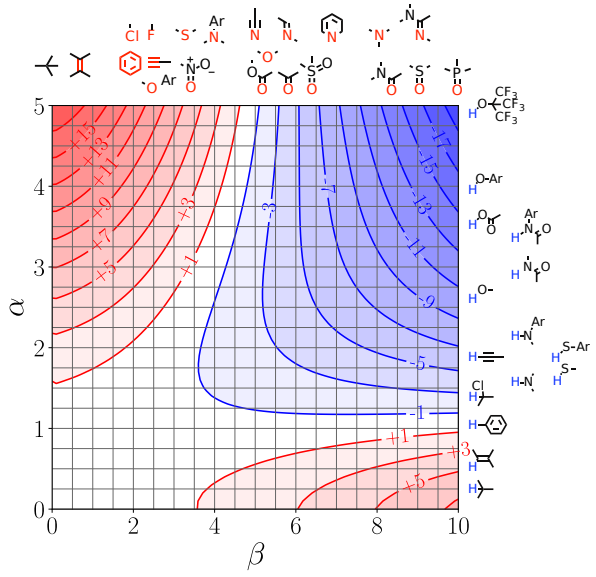


Fig. H.203: FGIP for 2-bromopyridine at 298K.

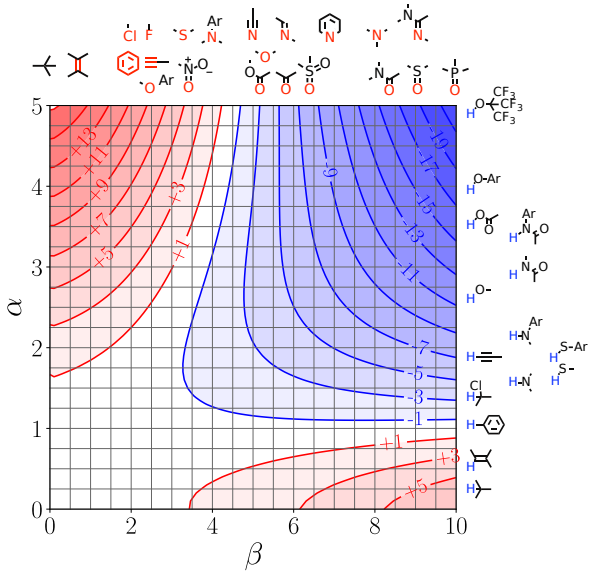


Fig. H.204: FGIP for 3-bromopyridine at 298K.

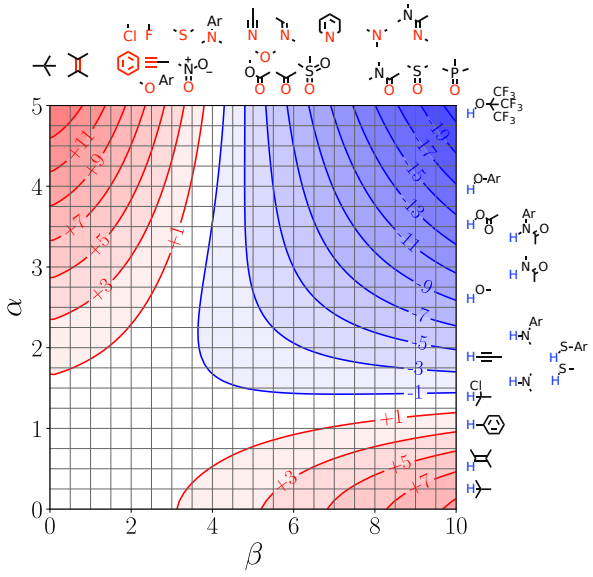


Fig. H.205: FGIP for 2-cyanopyridine at 298K.

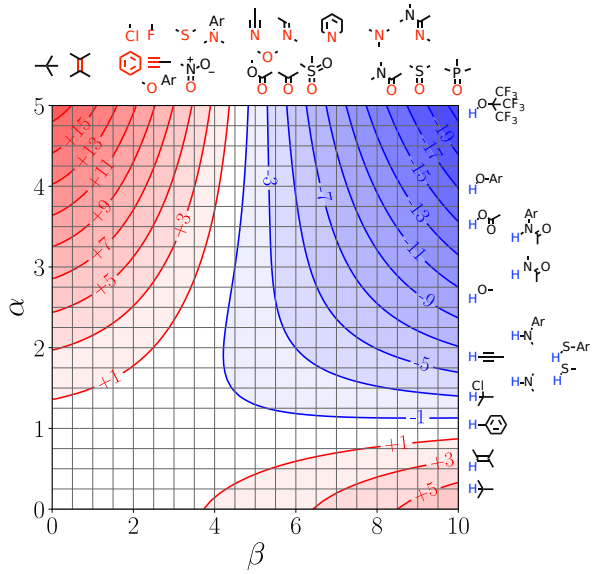


Fig. H.206: FGIP for pyrimidine at 298K.

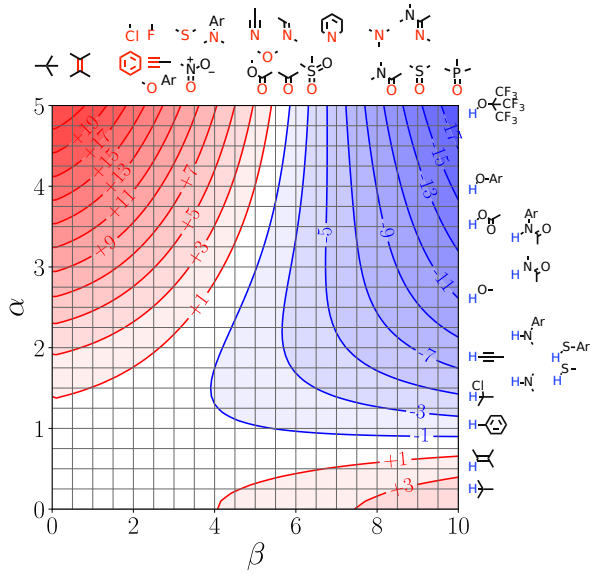


Fig. H.207: FGIP for quinoline at 298K.

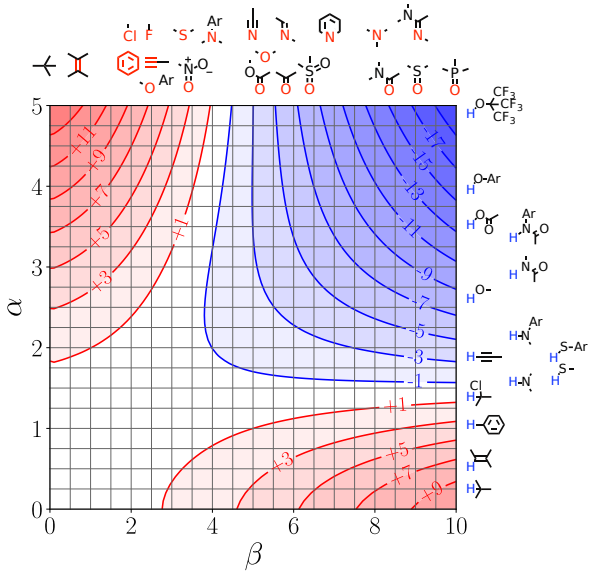


Fig. H.208: FGIP for acetonitrile at 298K.

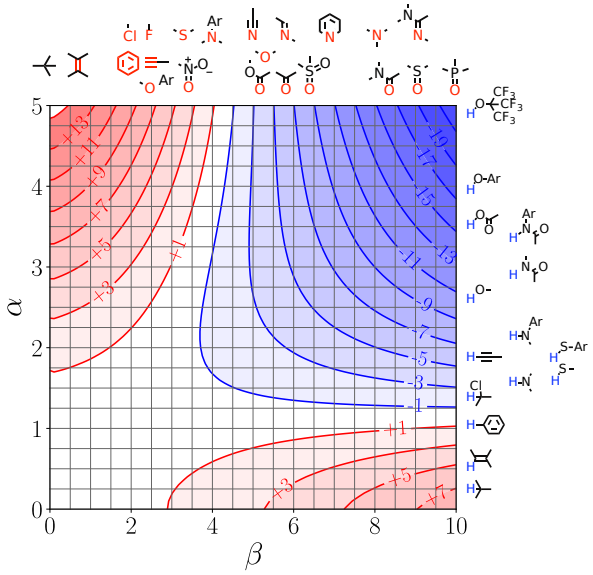


Fig. H.209: FGIP for propionitrile at 298K.

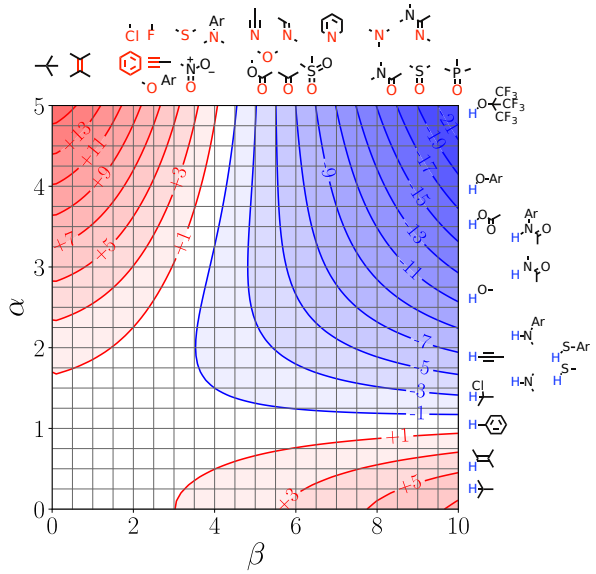


Fig. H.210: FGIP for n-butyronitrile at 298K.

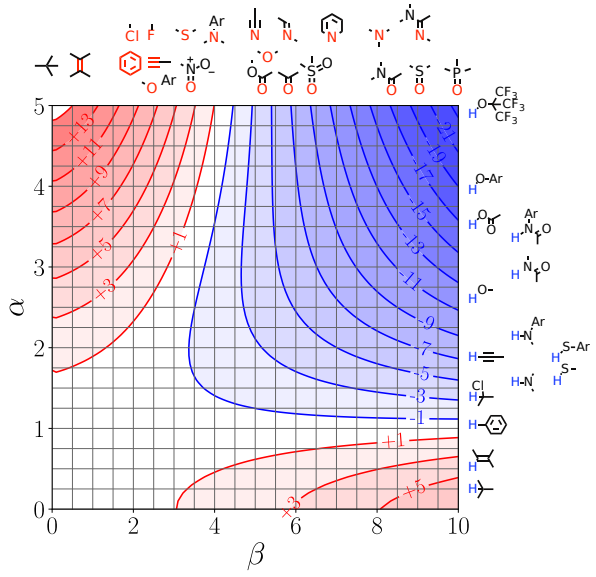


Fig. H.211: FGIP for 3-methylbutanenitrile at 298K.

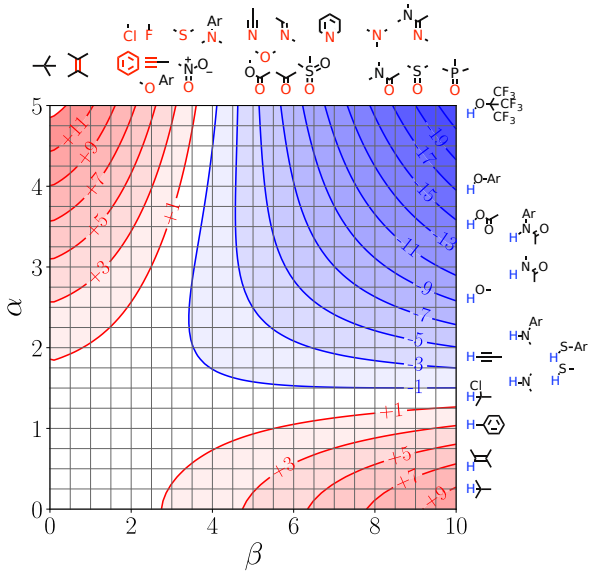


Fig. H.212: FGIP for acrylonitrile at 298K.

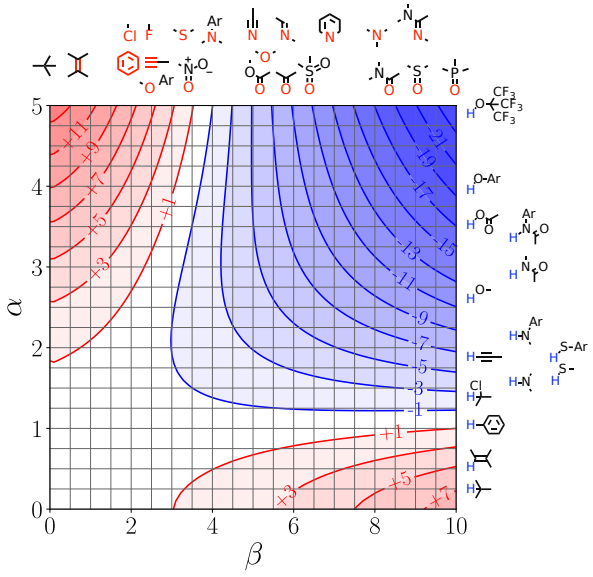


Fig. H.213: FGIP for phenylacetonitrile at 298K.

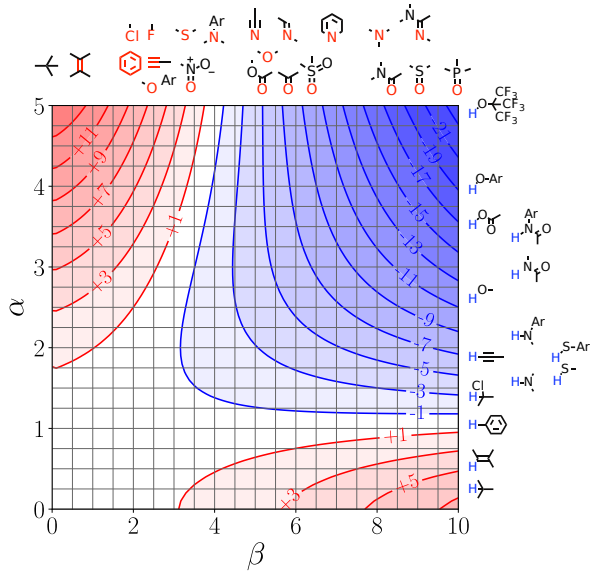


Fig. H.214: FGIP for benzonitrile at 298K.

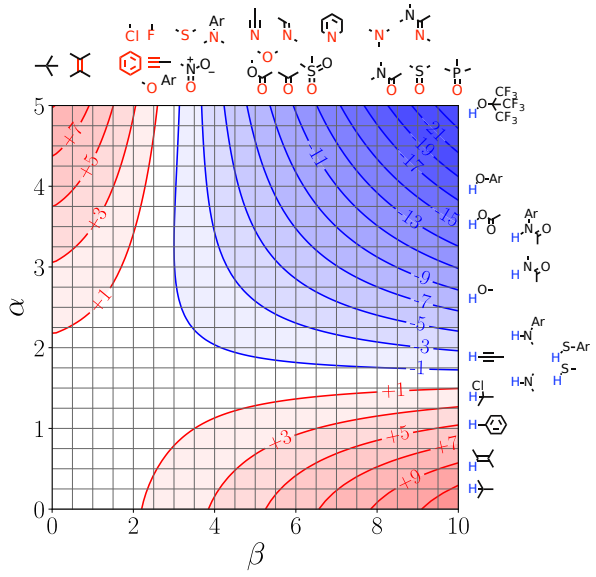


Fig. H.215: FGIP for nitromethane at 298K.

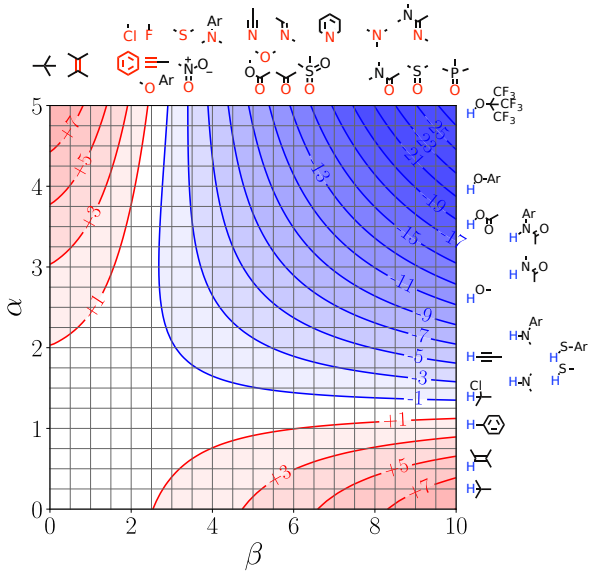


Fig. H.216: FGIP for nitroethane at 298K.

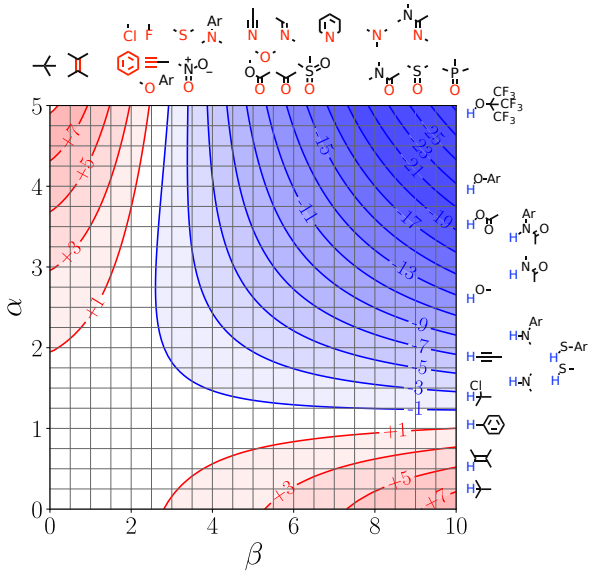


Fig. H.217: FGIP for 1-nitropropane at 298K.

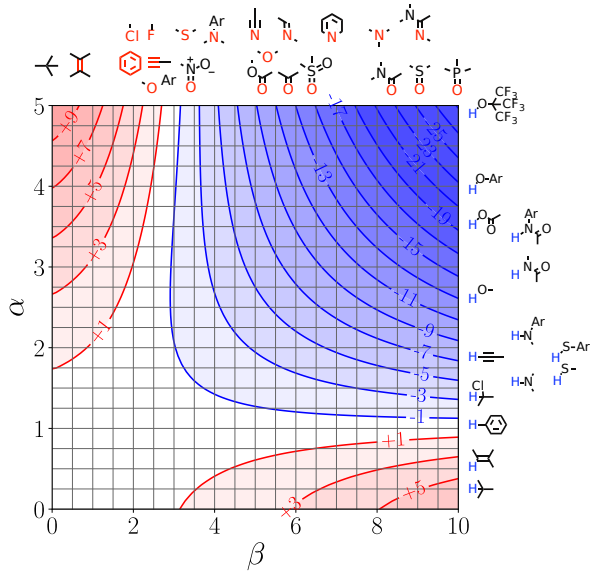


Fig. H.218: FGIP for 2-nitropropane at 298K.

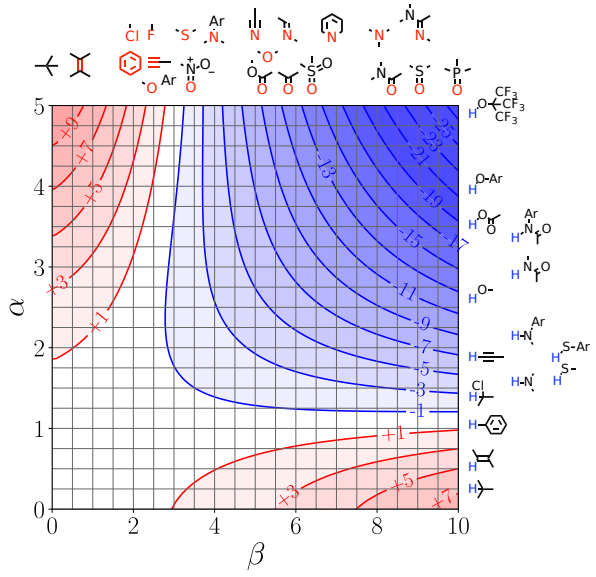


Fig. H.219: FGIP for nitrobenzene at 298K.

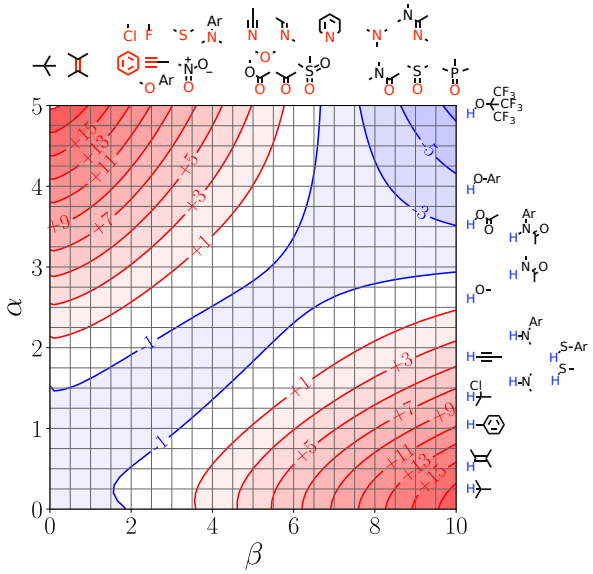


Fig. H.220: FGIP for formamide at 298K.

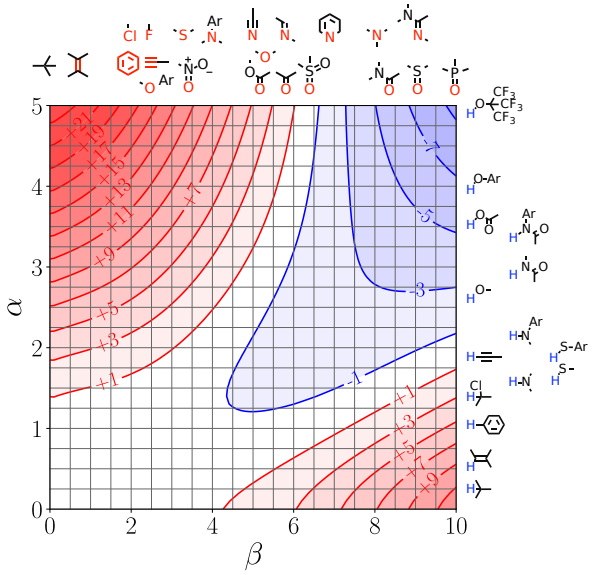


Fig. H.221: FGIP for N-methylformamide at 298K.

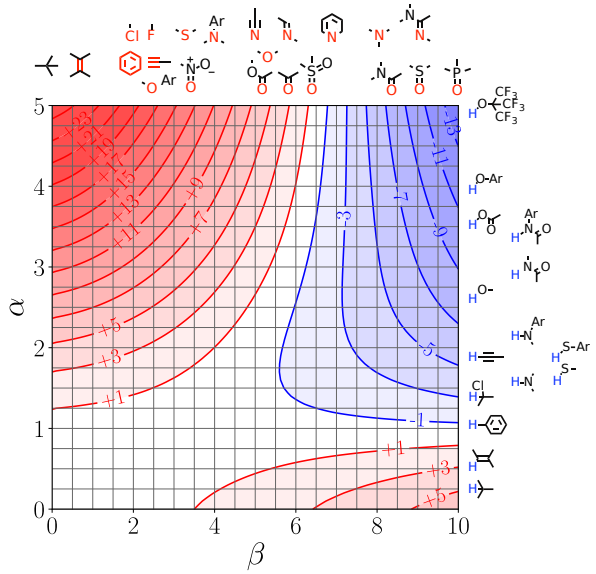


Fig. H.222: FGIP for N,N-dimethylformamide at 298K.

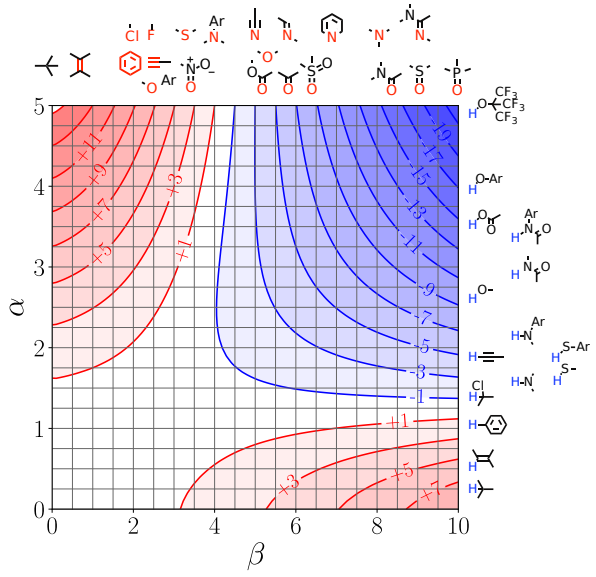


Fig. H.223: FGIP for N,N-dimethylthioformamide at 298K.

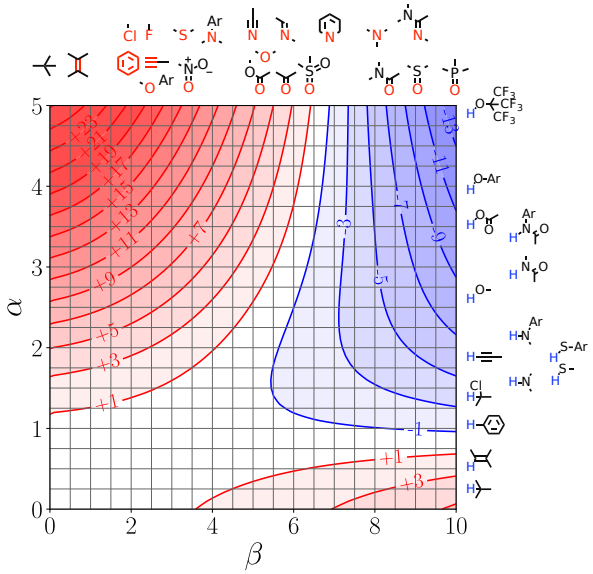


Fig. H.224: FGIP for N,N-diethylformamide at 298K.

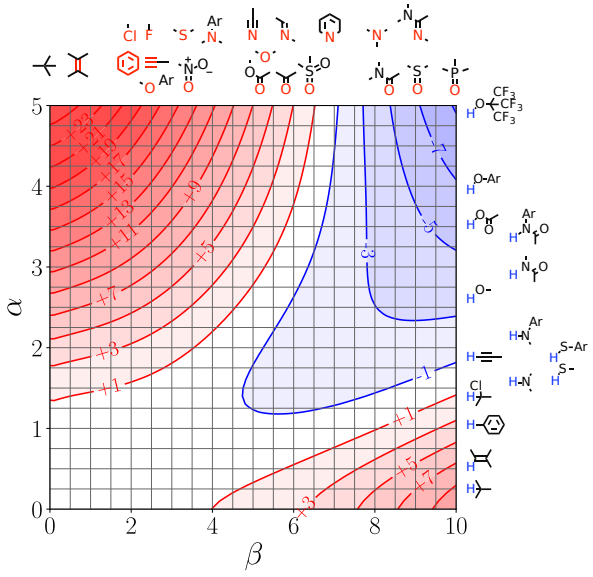


Fig. H.225: FGIP for N-methylacetamide at 298K.

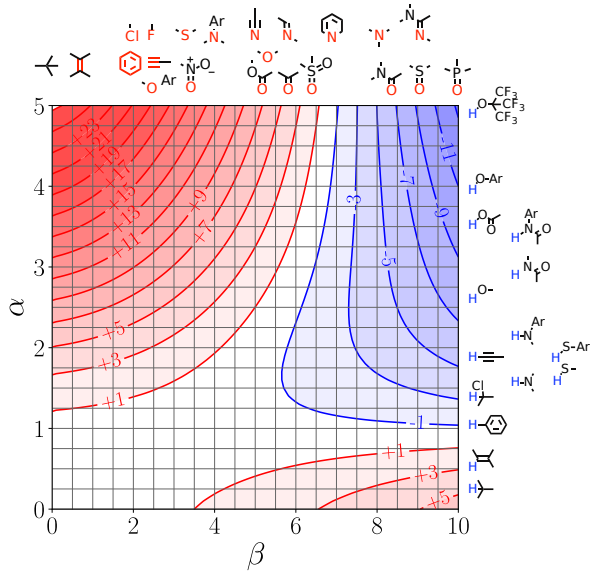


Fig. H.226: FGIP for N,N-dimethylacetamide at 298K.

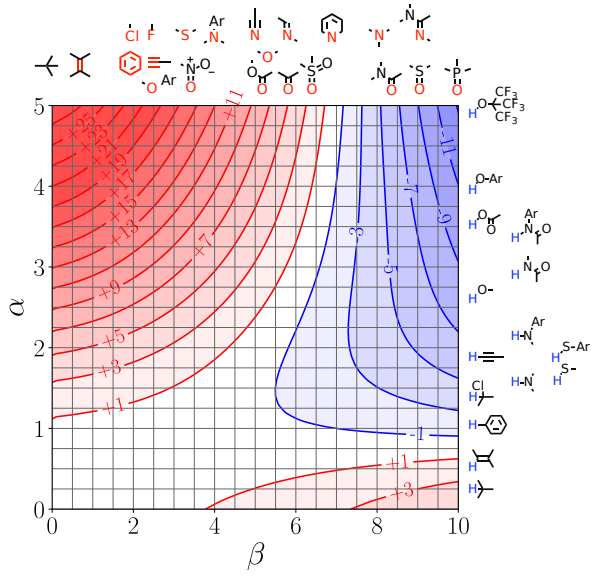


Fig. H.227: FGIP for N,N-diethylacetamide at 298K.

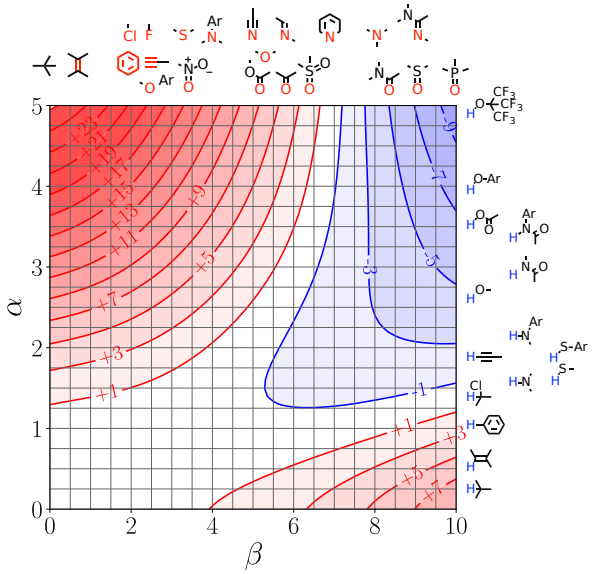


Fig. H.228: FGIP for 2-pyrrolidinone at 298K.

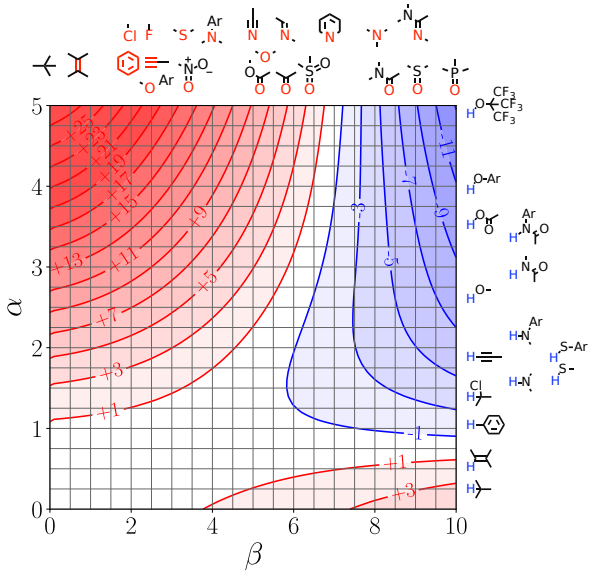


Fig. H.229: FGIP for N-methyl pyrrolidinone at 298K.

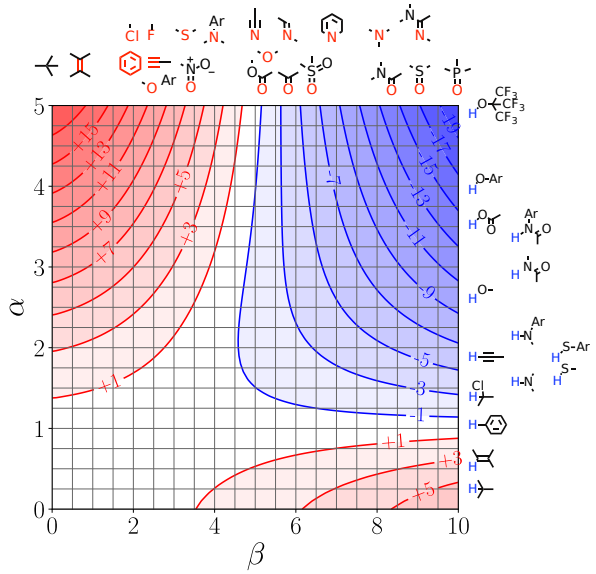


Fig. H.230: FGIP for 1-methyl-2-pyrrolidinethione at 298K.

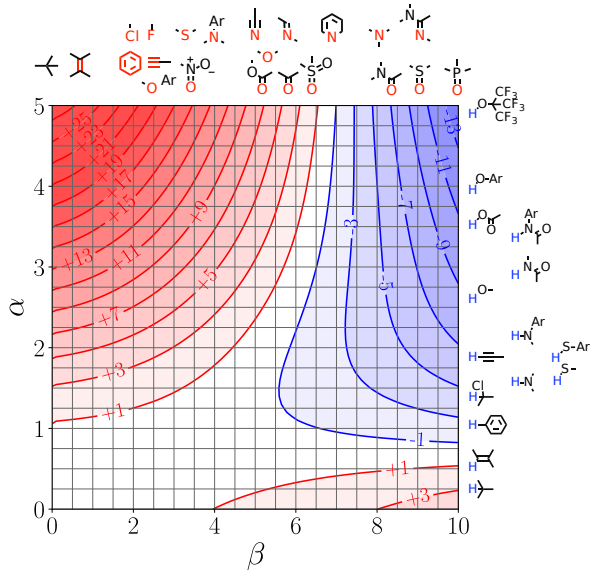


Fig. H.231: FGIP for tetramethylurea at 298K.

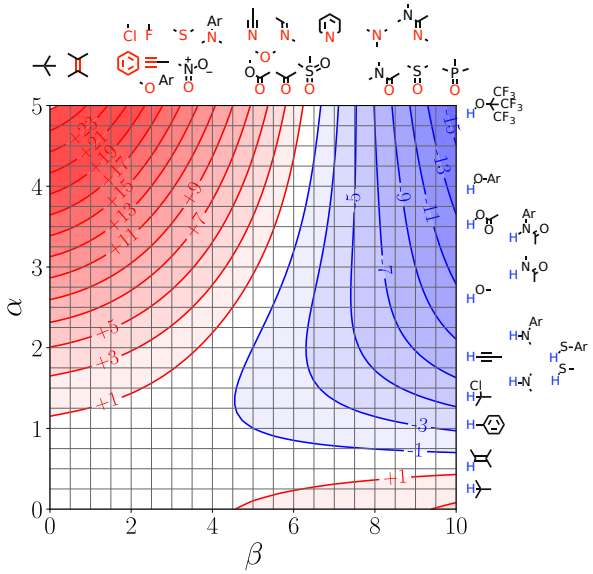


Fig. H.232: FGIP for tetraethylurea at 298K.

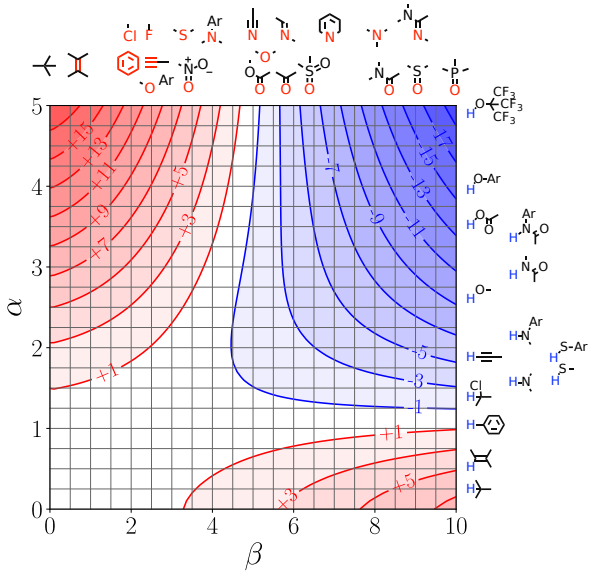


Fig. H.233: FGIP for dimethylcyanamide at 298K.

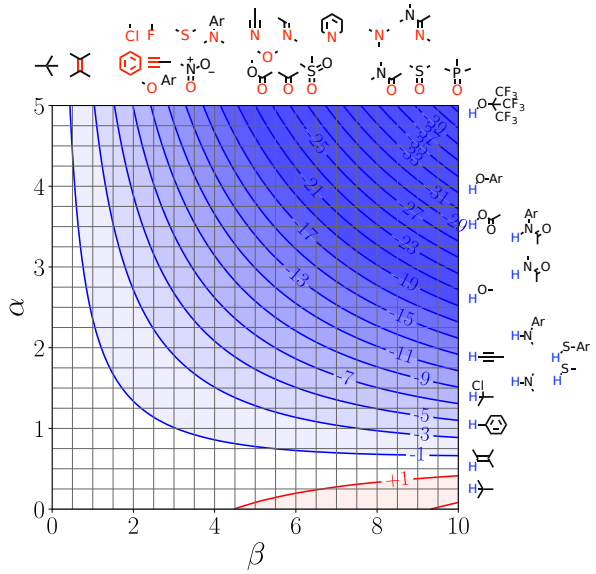


Fig. H.234: FGIP for carbon disulfide at 298K.

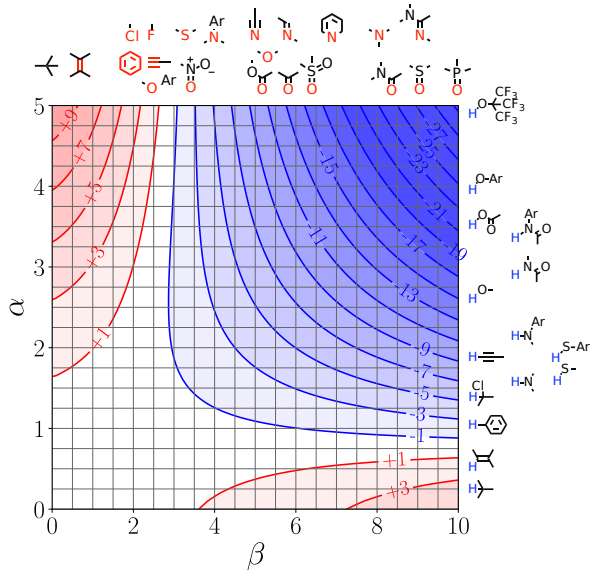


Fig. H.235: FGIP for dimethyl sulfide at 298K.

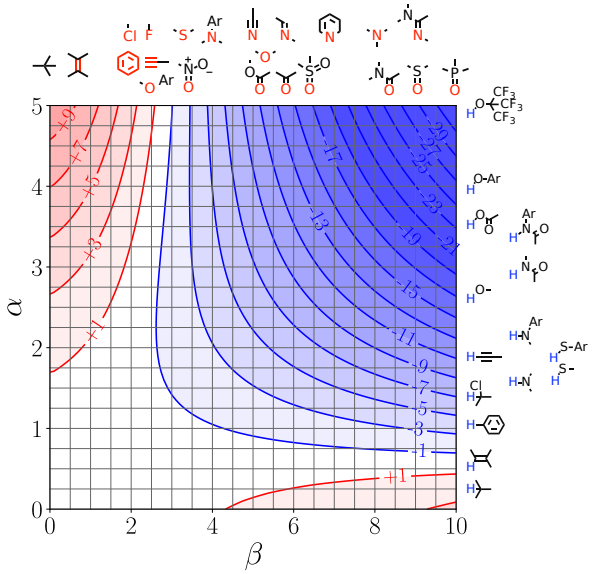


Fig. H.236: FGIP for diethyl sulfide at 298K.

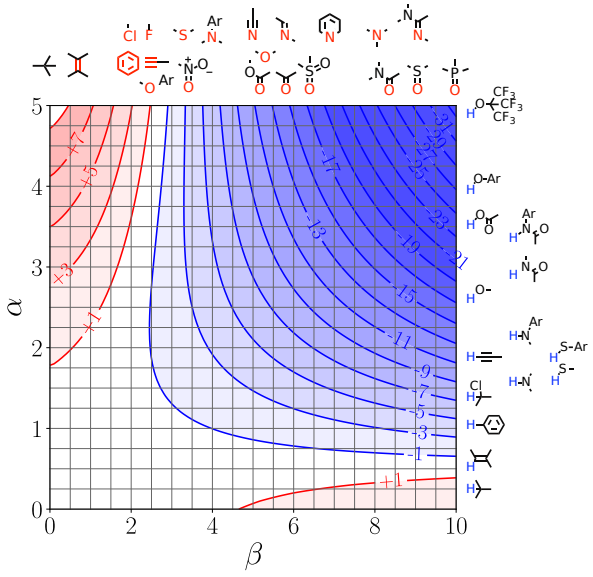


Fig. H.237: FGIP for diisopropyl sulfide at 298K.

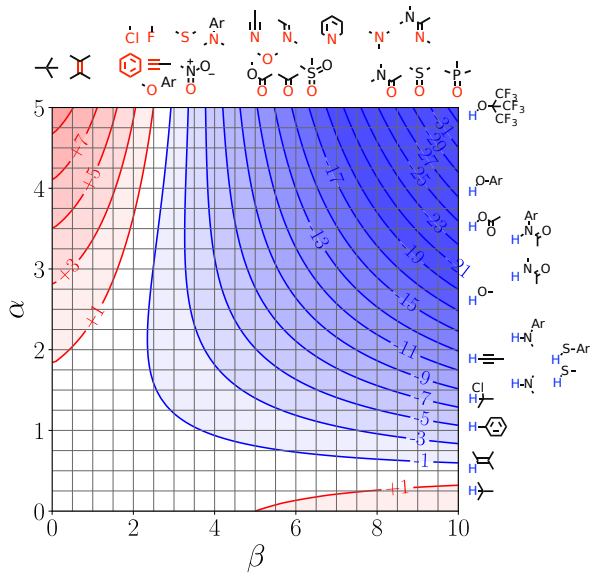


Fig. H.238: FGIP for dibutyl sulfide at 298K.

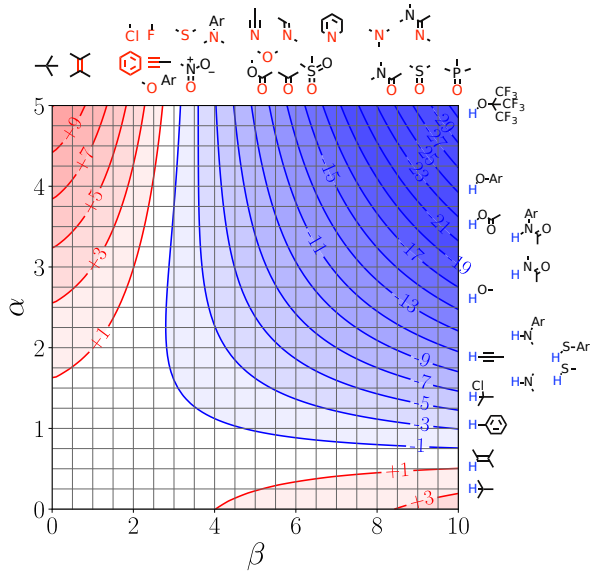


Fig. H.239: FGIP for tetrahydrothiophene at 298K.

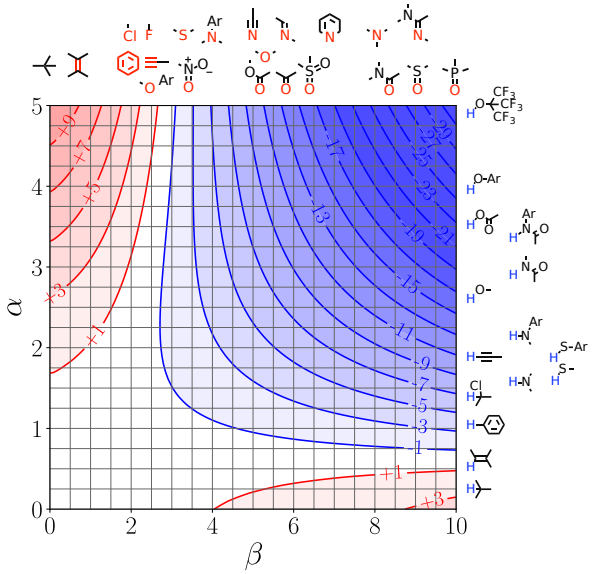


Fig. H.240: FGIP for thiane at 298K.

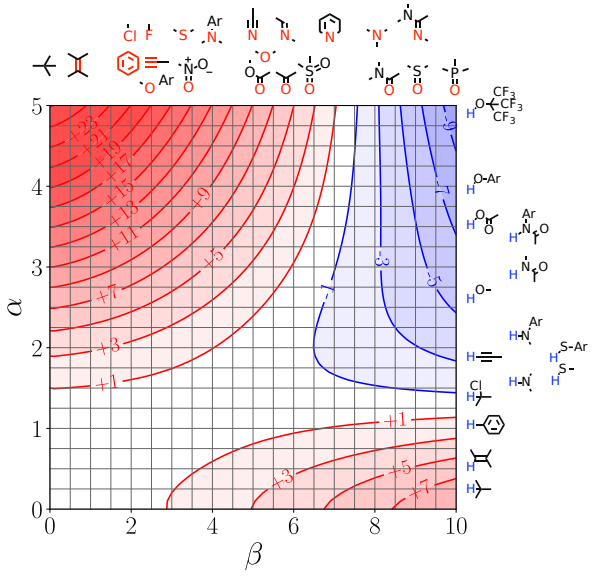


Fig. H.241: FGIP for dimethylsulfoxide at 298K.

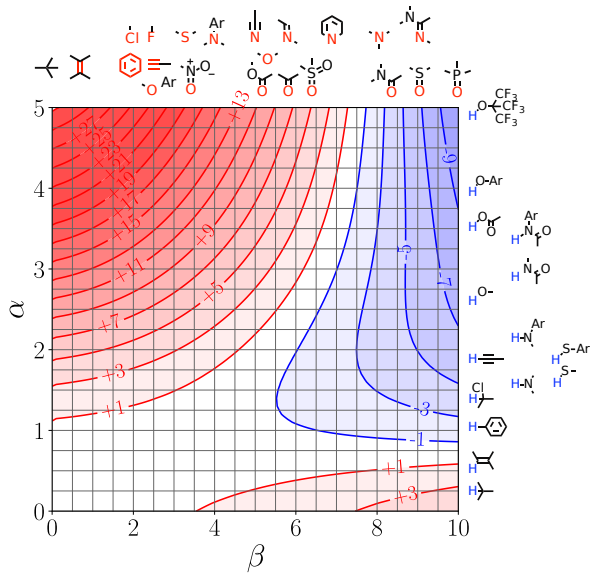


Fig. H.242: FGIP for dibutyl sulfoxide at 298K.

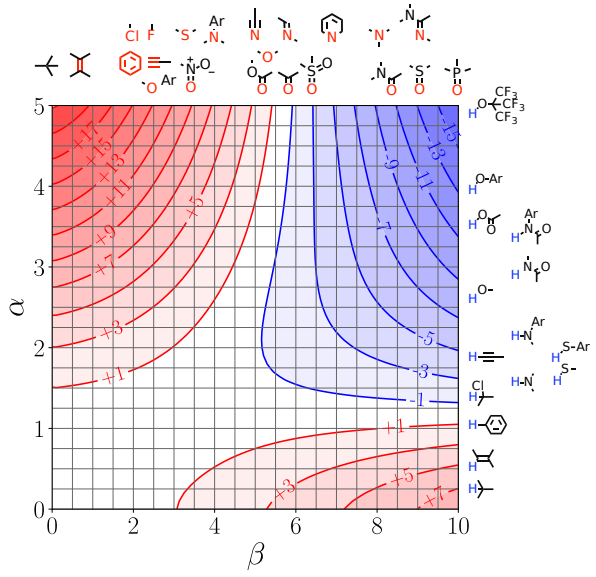


Fig. H.243: FGIP for sulfolane at 298K.

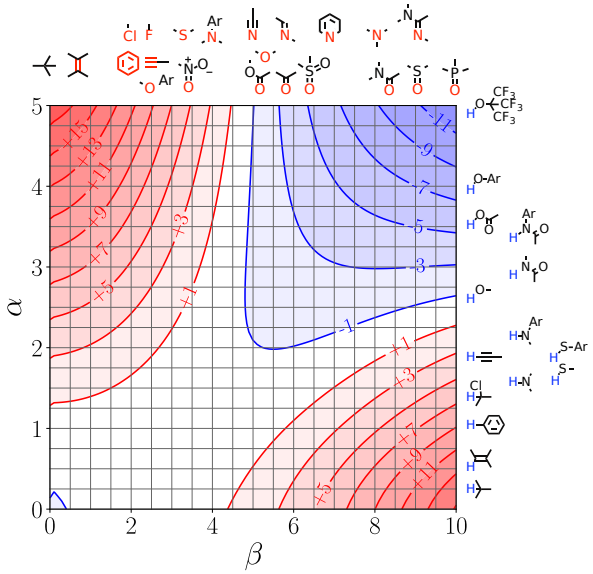


Fig. H.244: FGIP for thiobis(2-ethanol) at 298K.

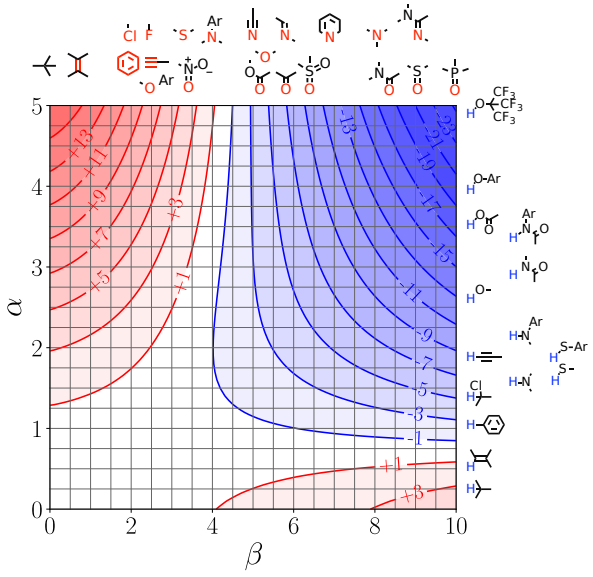


Fig. H.245: FGIP for diethyl sulfite at 298K.

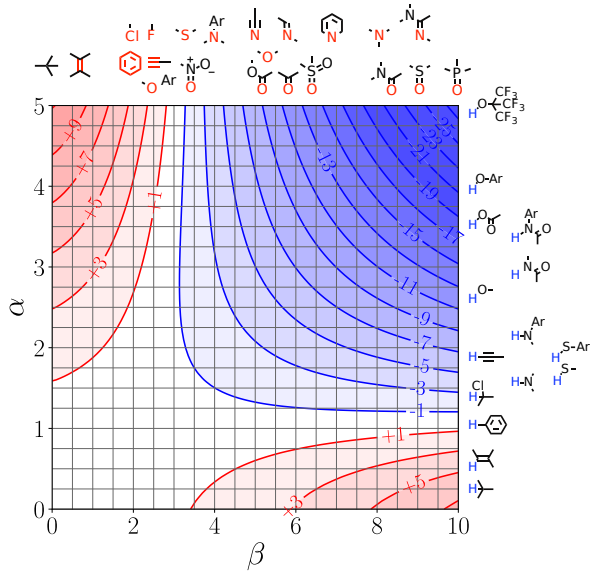


Fig. H.246: FGIP for dimethyl sulfate at 298K.

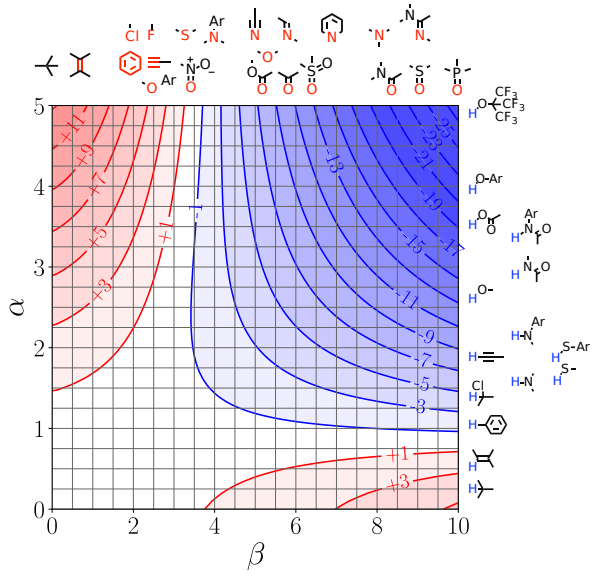


Fig. H.247: FGIP for diethyl sulfate at 298K.

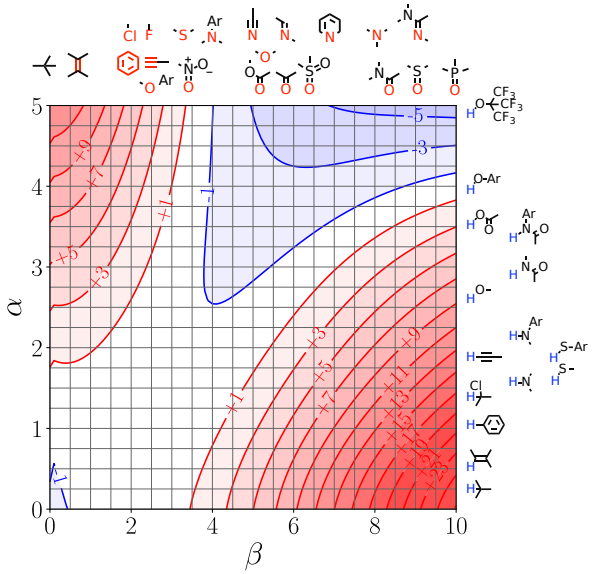


Fig. H.248: FGIP for methanesulfonic acid at 298K.

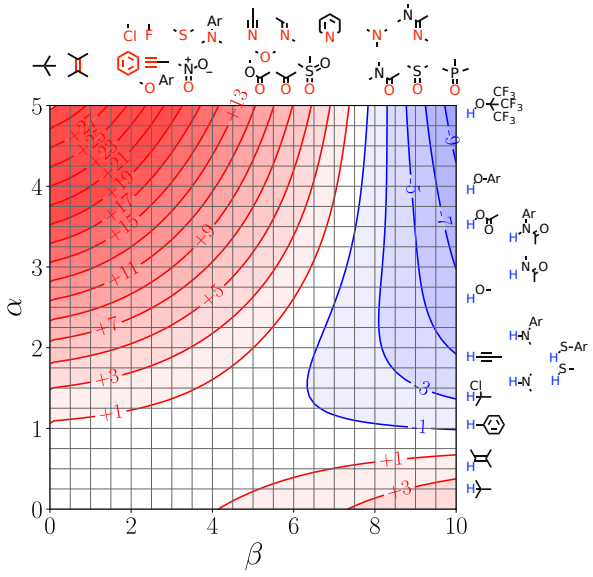


Fig. H.249: FGIP for trimethylphosphate at 298K.

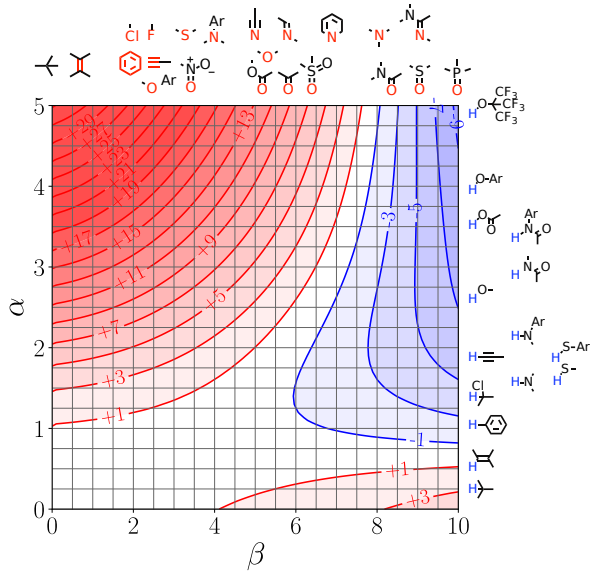


Fig. H.250: FGIP for triethylphosphate at 298K.

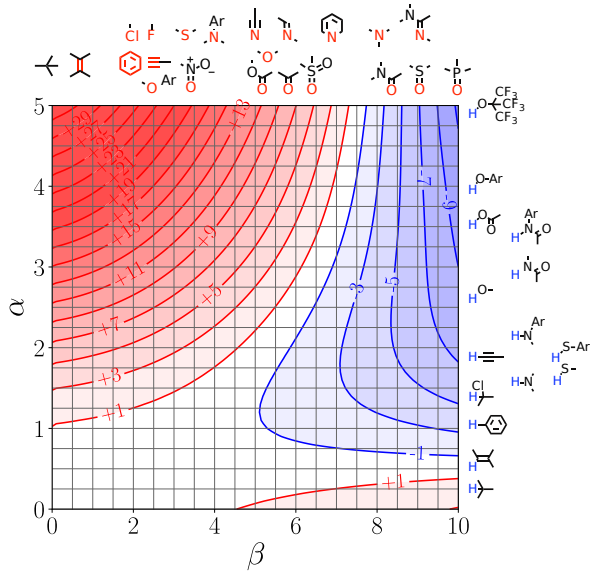


Fig. H.251: FGIP for tri-n-butylphosphate at 298K.

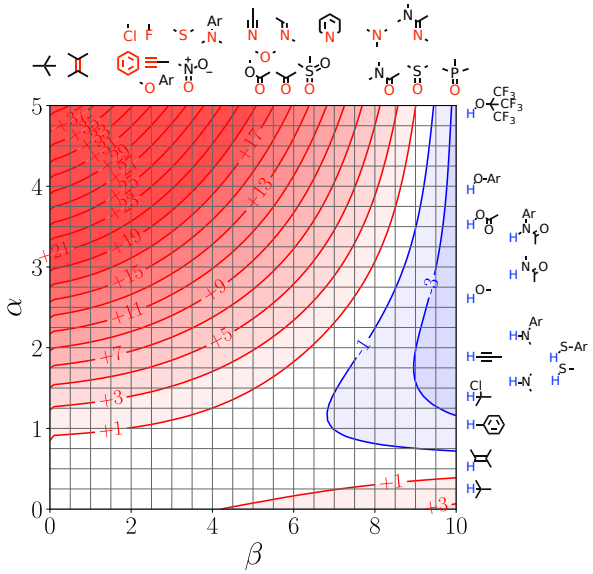


Fig. H.252: FGIP for hexamethylphosphoric triamide at 298K.

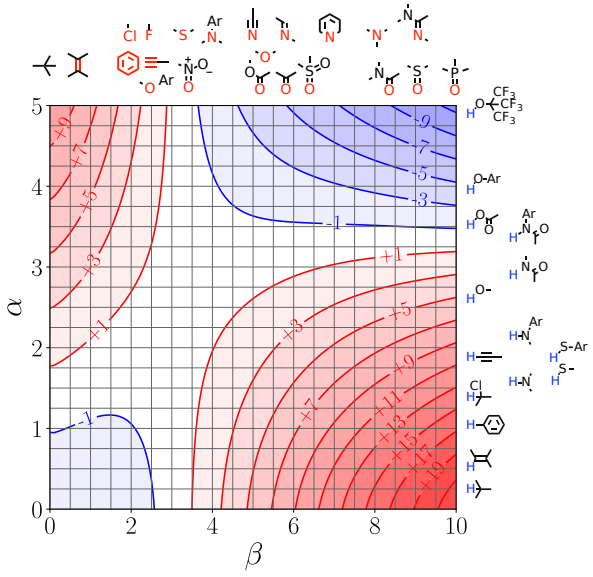


Fig. H.253: FGIP for hydrogen peroxide at 298K.

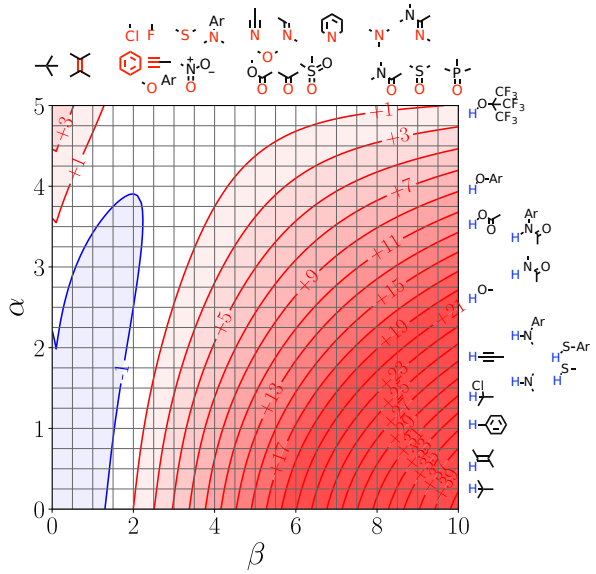


Fig. H.254: FGIP for hydrogen fluoride at 298K.

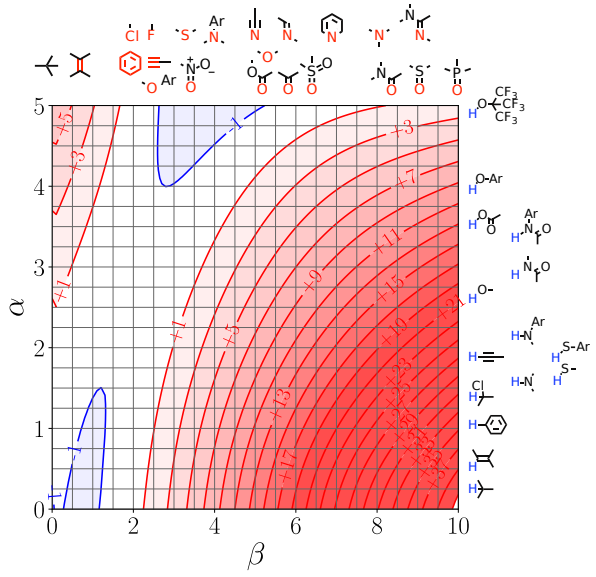


Fig. H.255: FGIP for sulfuric acid at 298K.

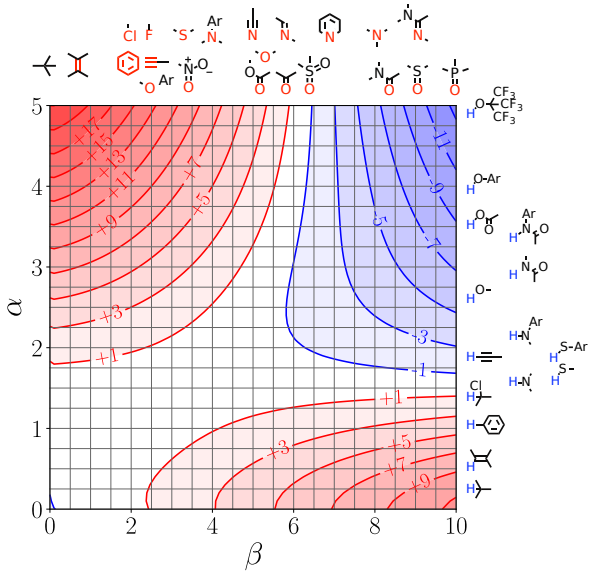


Fig. H.256: FGIP for ammonia at 298K.

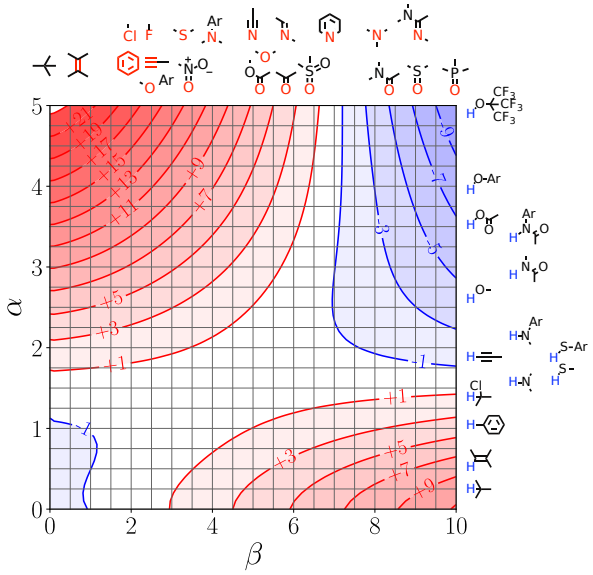


Fig. H.257: FGIP for hydrazine at 298K.

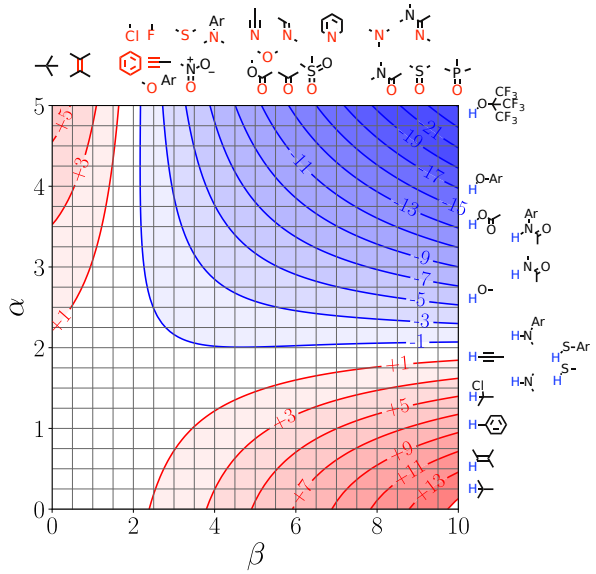


Fig. H.258: FGIP for sulfur dioxide at 298K.

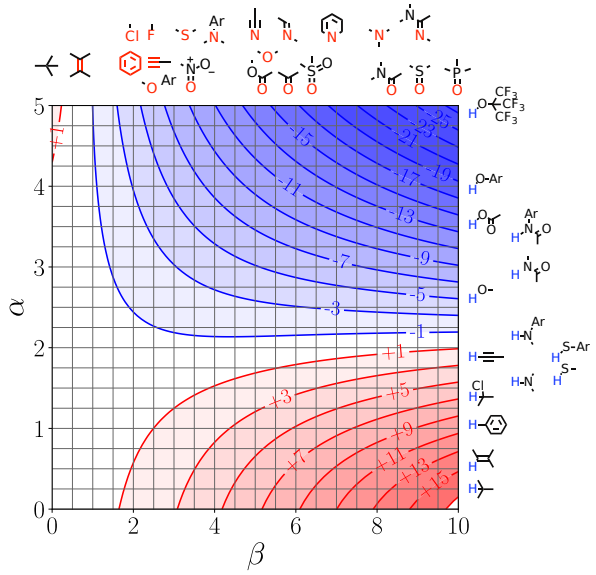


Fig. H.259: FGIP for thionyl chloride at 298K.

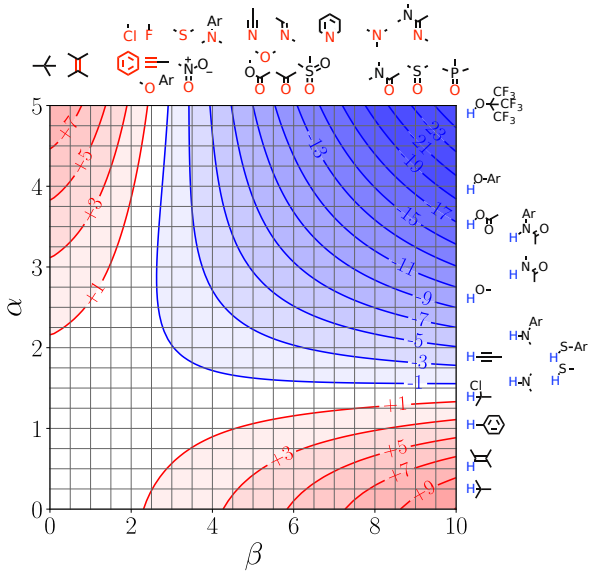


Fig. H.260: FGIP for phosphorus oxychloride at 298K.

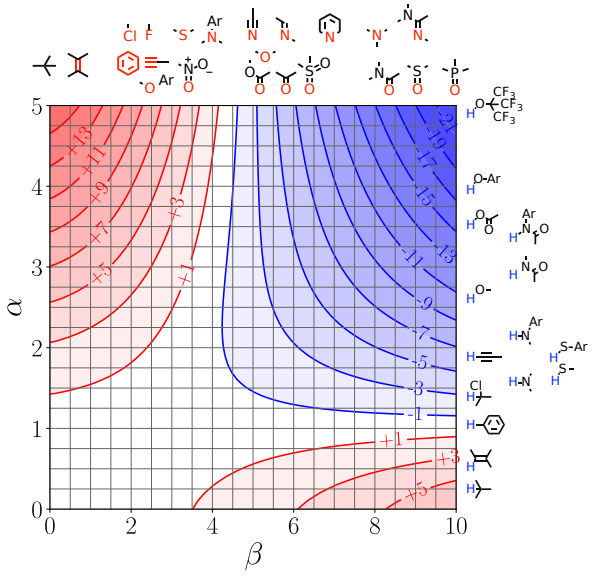


Fig. H.261: FGIP for dihydrolevoglucosenone at 298K.

H.2 Binary solvent mixture FGIPs

Volume fractions used to express solvent concentrations.

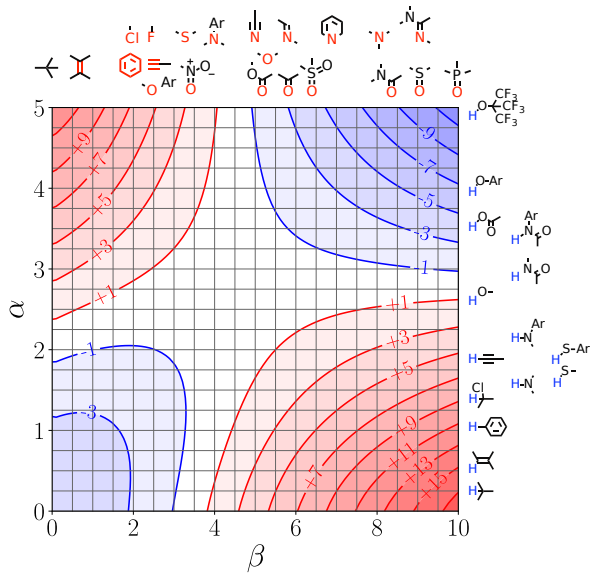


Fig. H.262: FGIP for 0.0% ethanol 100.0% water at 298K.

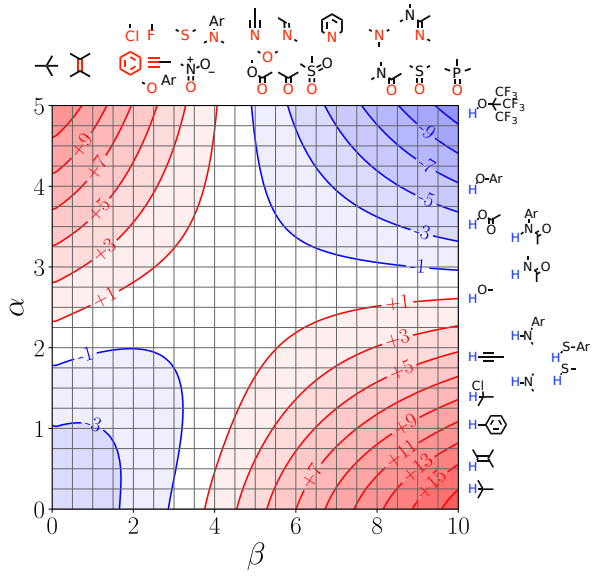


Fig. H.263: FGIP for 5.0% ethanol 95.0% water at 298K.

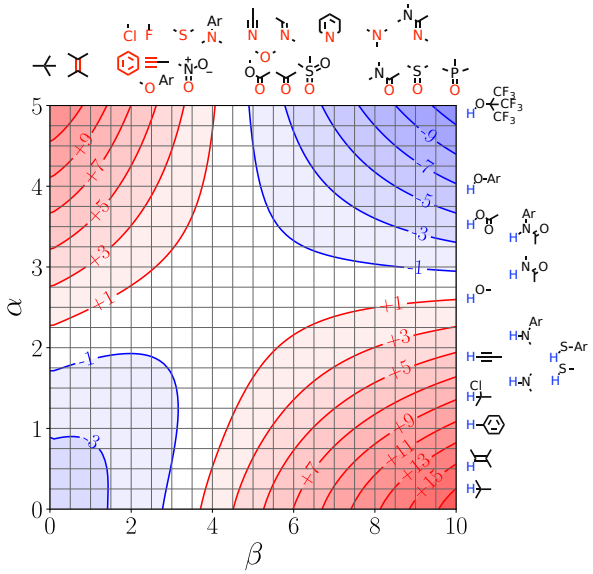


Fig. H.264: FGIP for 10.0% ethanol 90.0% water at 298K.

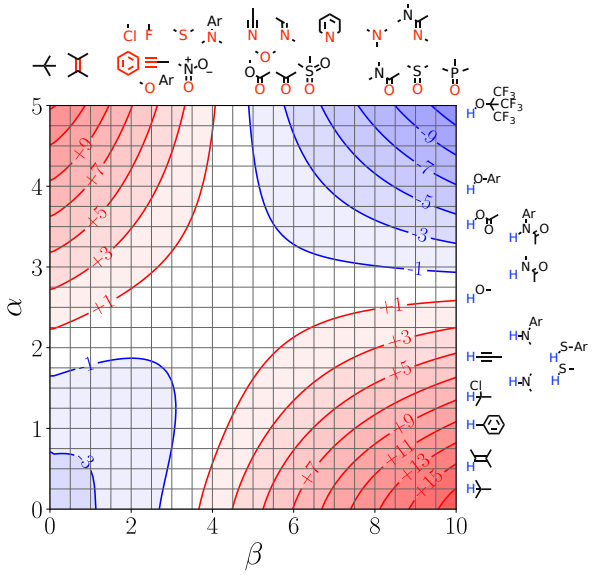


Fig. H.265: FGIP for 15.0% ethanol 85.0% water at 298K.

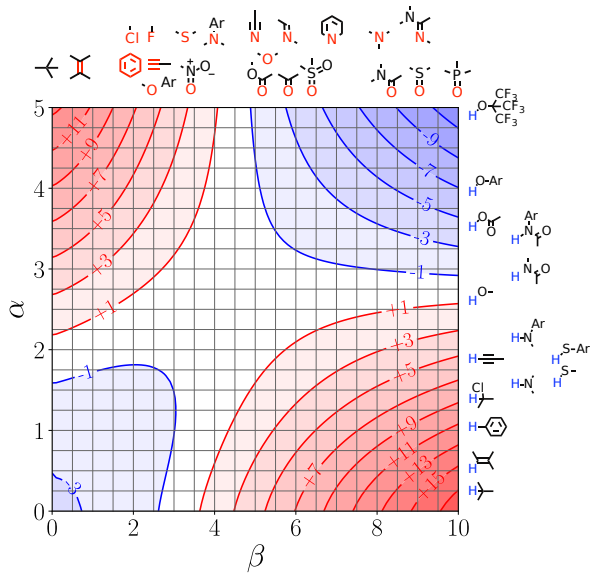


Fig. H.266: FGIP for 20.0% ethanol 80.0% water at 298K.

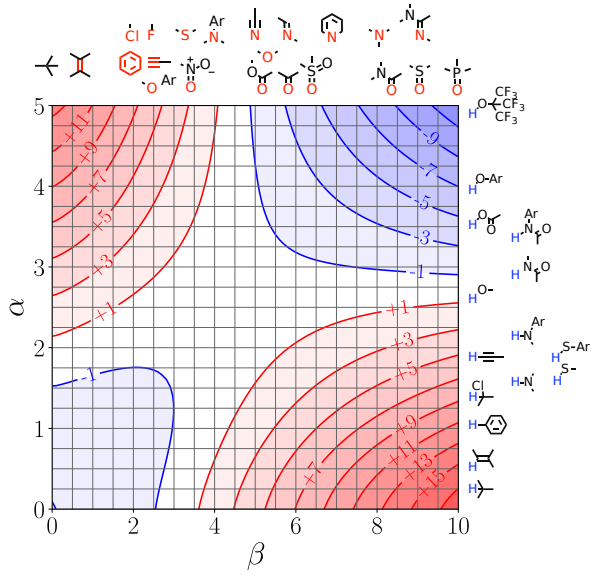


Fig. H.267: FGIP for 25.0% ethanol 75.0% water at 298K.

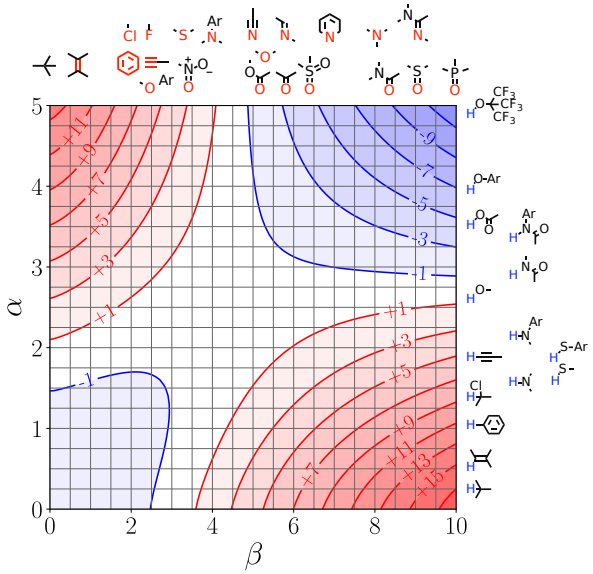


Fig. H.268: FGIP for 30.0% ethanol 70.0% water at 298K.

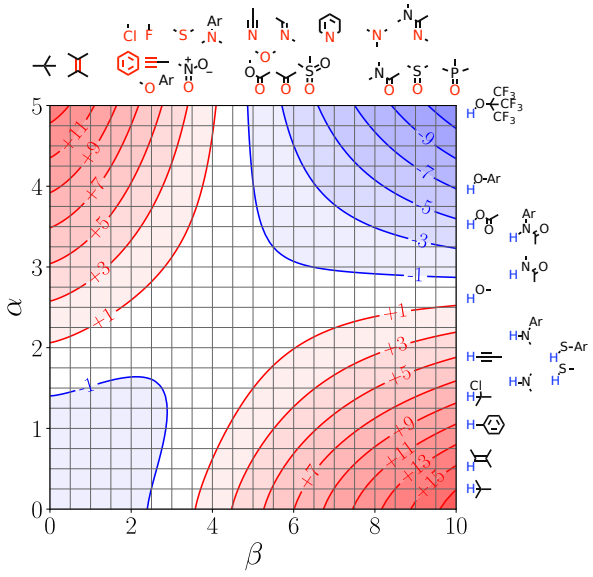


Fig. H.269: FGIP for 35.0% ethanol 65.0% water at 298K.

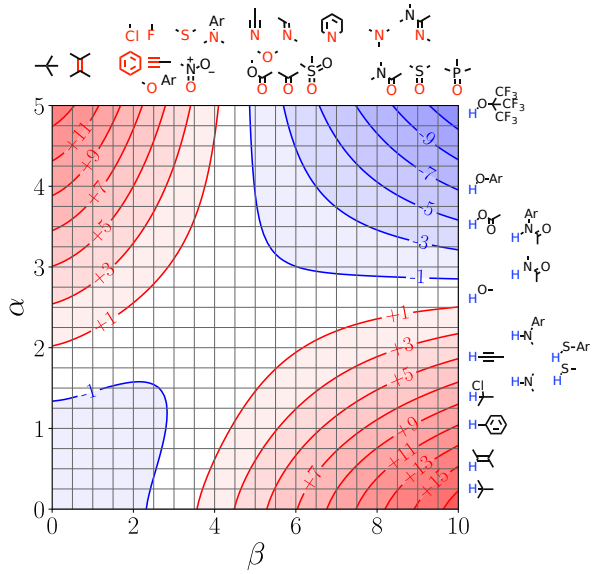


Fig. H.270: FGIP for 40.0% ethanol 60.0% water at 298K.

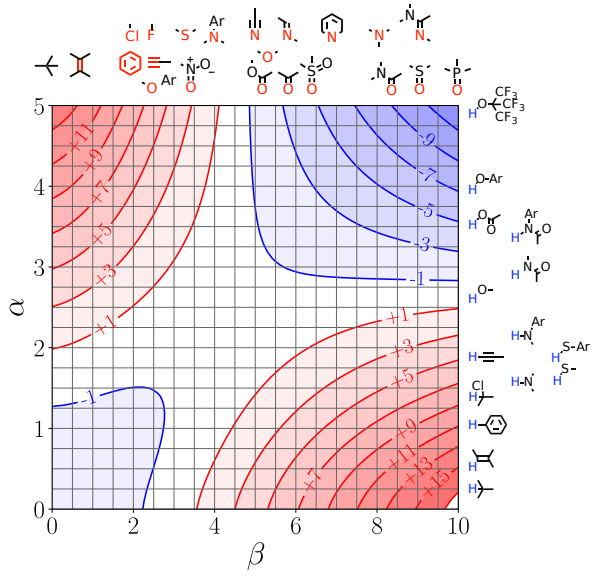


Fig. H.271: FGIP for 45.0% ethanol 55.0% water at 298K.

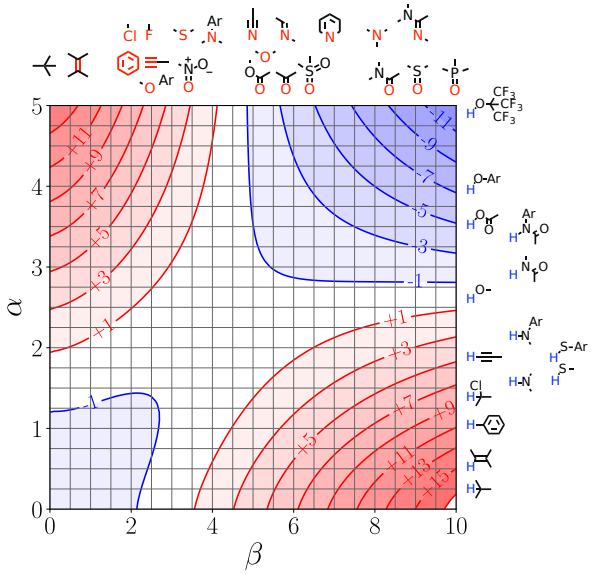


Fig. H.272: FGIP for 50.0% ethanol 50.0% water at 298K.

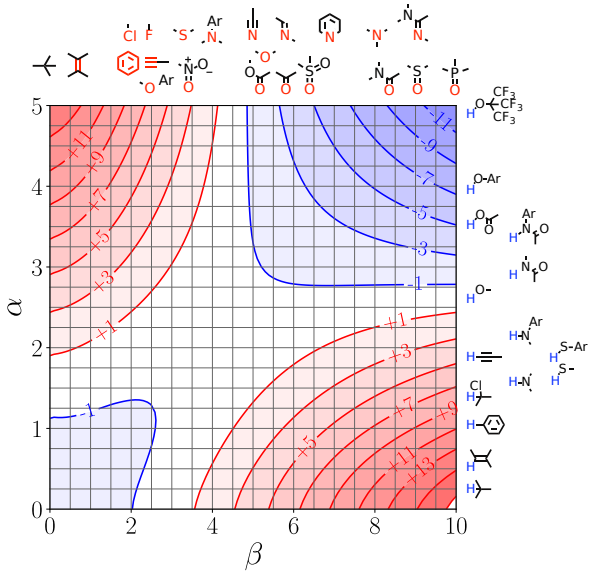


Fig. H.273: FGIP for 55.0% ethanol 45.0% water at 298K.

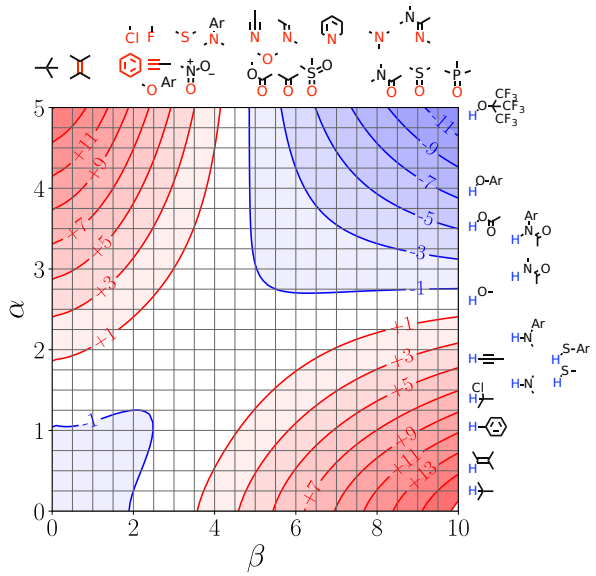


Fig. H.274: FGIP for 60.0% ethanol 40.0% water at 298K.

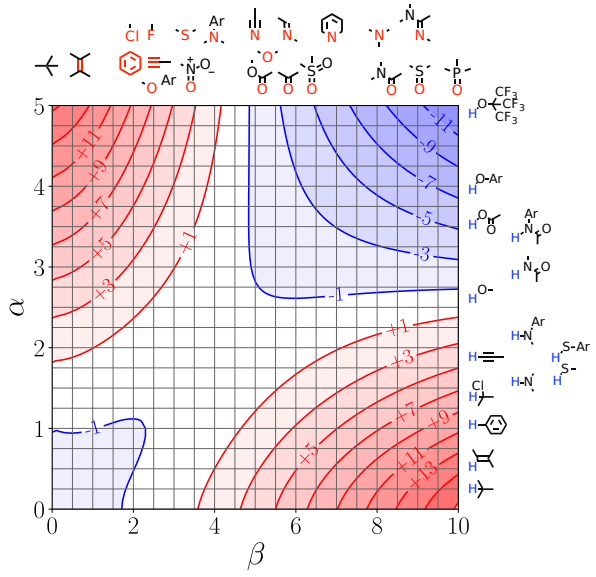


Fig. H.275: FGIP for 65.0% ethanol 35.0% water at 298K.

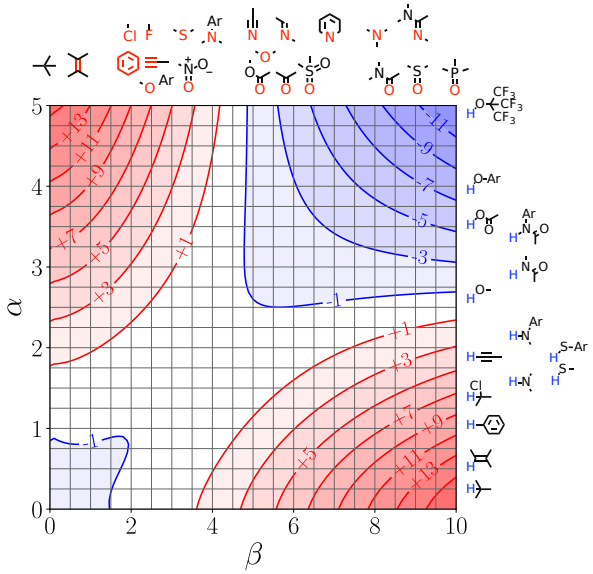


Fig. H.276: FGIP for 70.0% ethanol 30.0% water at 298K.

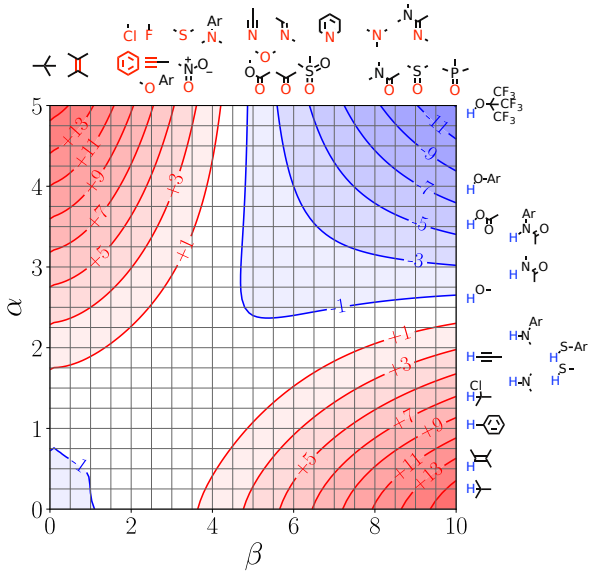


Fig. H.277: FGIP for 75.0% ethanol 25.0% water at 298K.

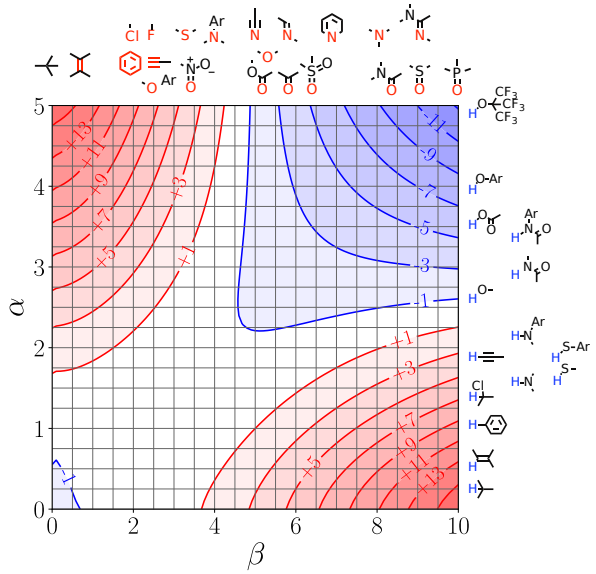


Fig. H.278: FGIP for 80.0% ethanol 20.0% water at 298K.

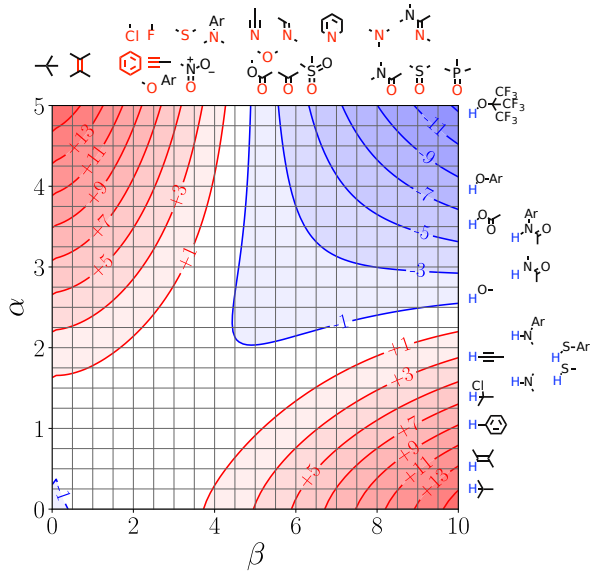


Fig. H.279: FGIP for 85.0% ethanol 15.0% water at 298K.

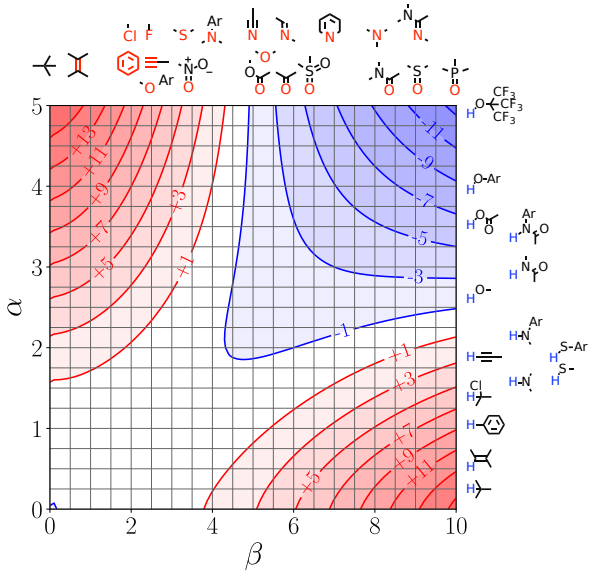


Fig. H.280: FGIP for 90.0% ethanol 10.0% water at 298K.

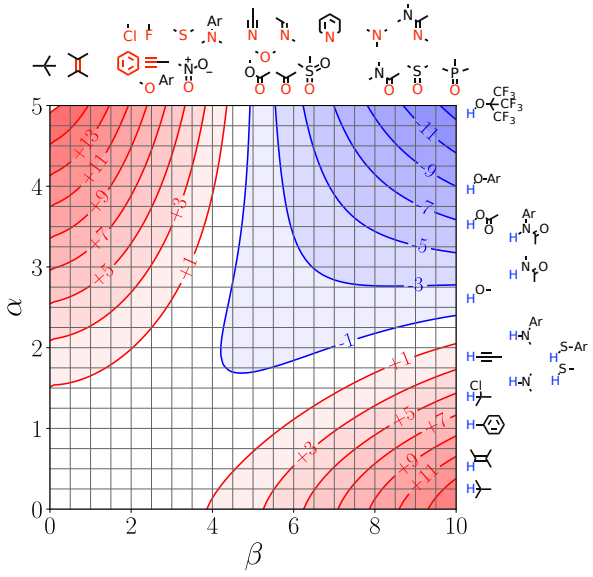


Fig. H.281: FGIP for 95.0% ethanol 5.0% water at 298K.

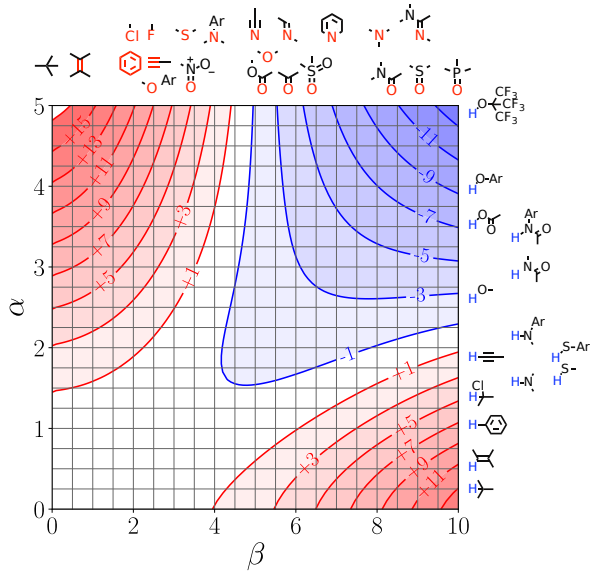


Fig. H.282: FGIP for 100.0% ethanol 0.0% water at 298K.

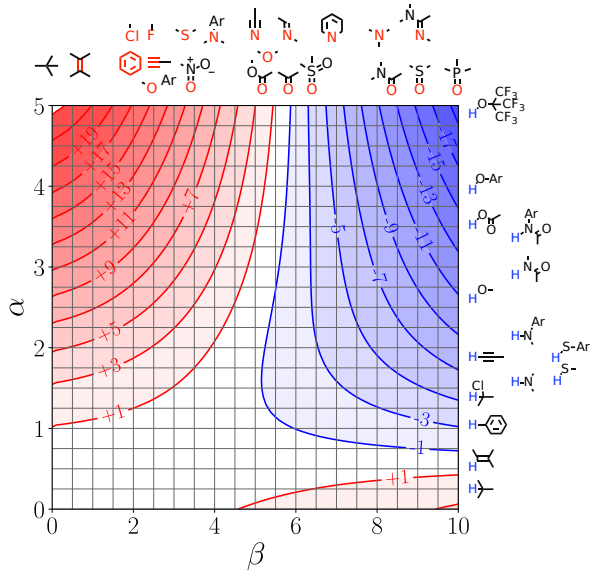


Fig. H.283: FGIP for 0.0% chloroform 100.0% tetrahydrofuran at 298K.

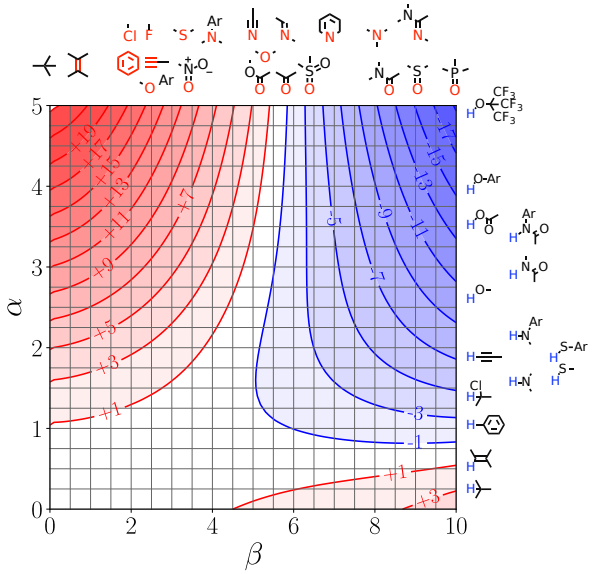


Fig. H.284: FGIP for 5.0% chloroform 95.0% tetrahydrofuran at 298K.

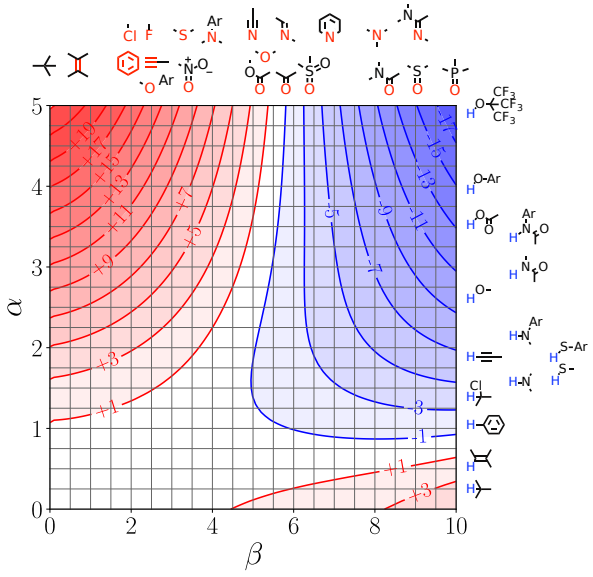


Fig. H.285: FGIP for 10.0% chloroform 90.0% tetrahydrofuran at 298K.

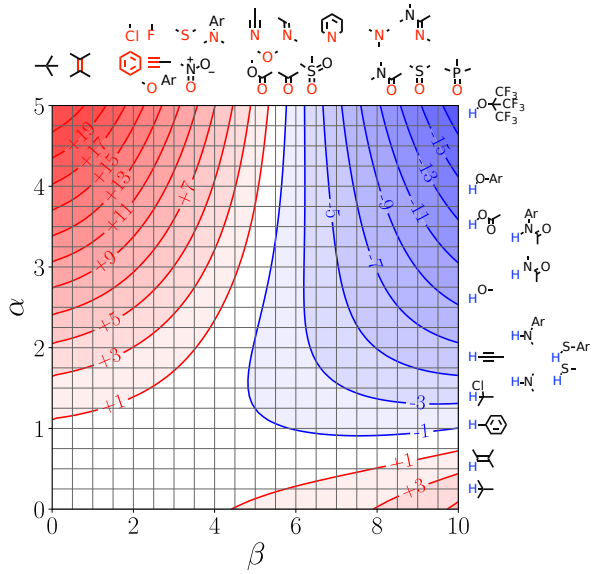


Fig. H.286: FGIP for 15.0% chloroform 85.0% tetrahydrofuran at 298K.

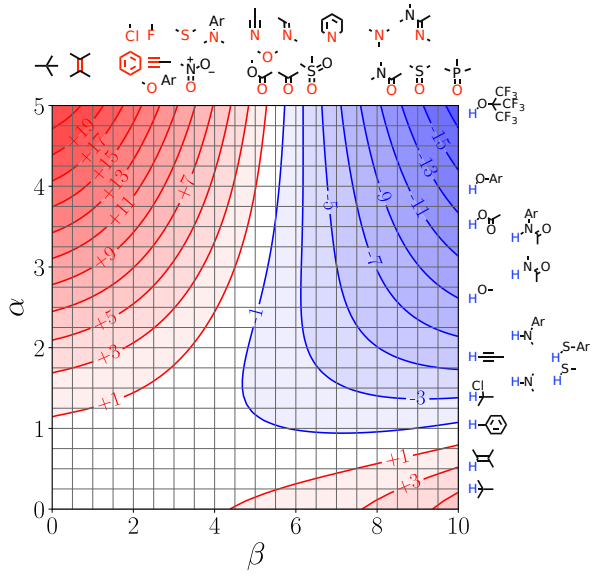


Fig. H.287: FGIP for 20.0% chloroform 80.0% tetrahydrofuran at 298K.

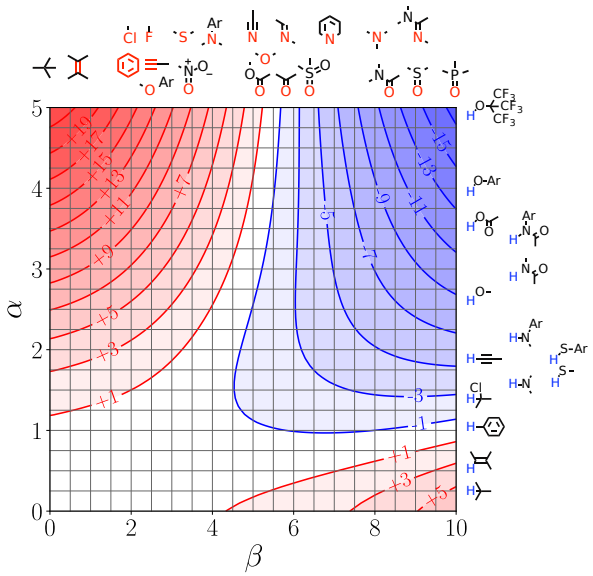


Fig. H.288: FGIP for 25.0% chloroform 75.0% tetrahydrofuran at 298K.

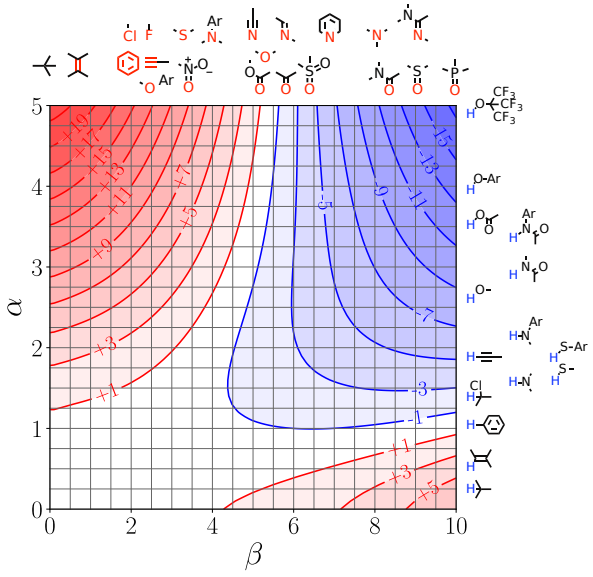


Fig. H.289: FGIP for 30.0% chloroform 70.0% tetrahydrofuran at 298K.

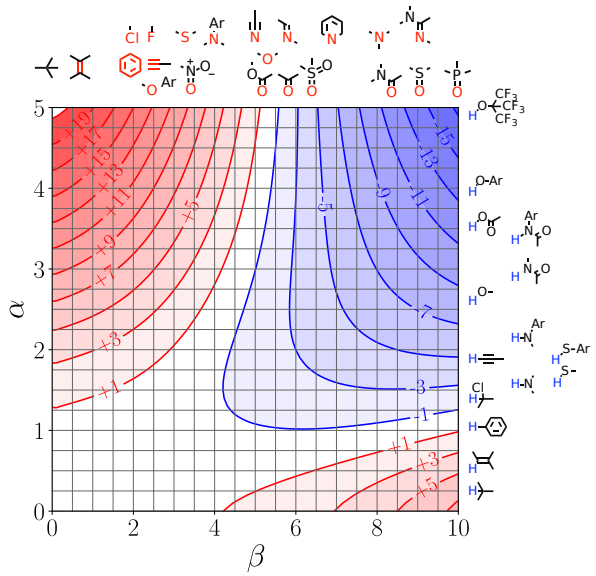


Fig. H.290: FGIP for 35.0% chloroform 65.0% tetrahydrofuran at 298K.

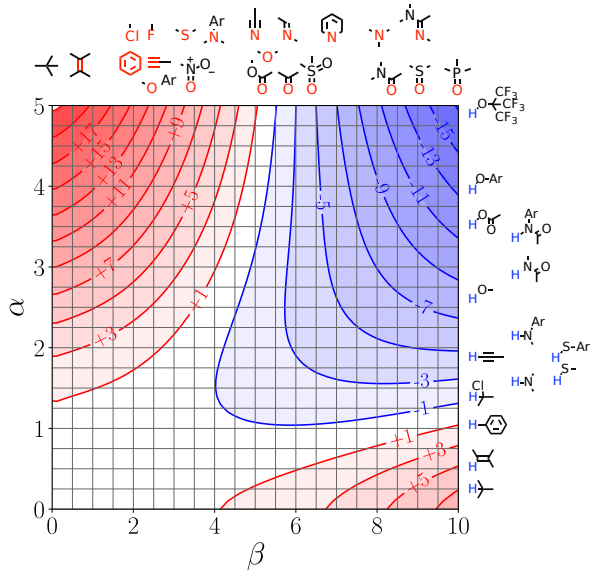


Fig. H.291: FGIP for 40.0% chloroform 60.0% tetrahydrofuran at 298K.

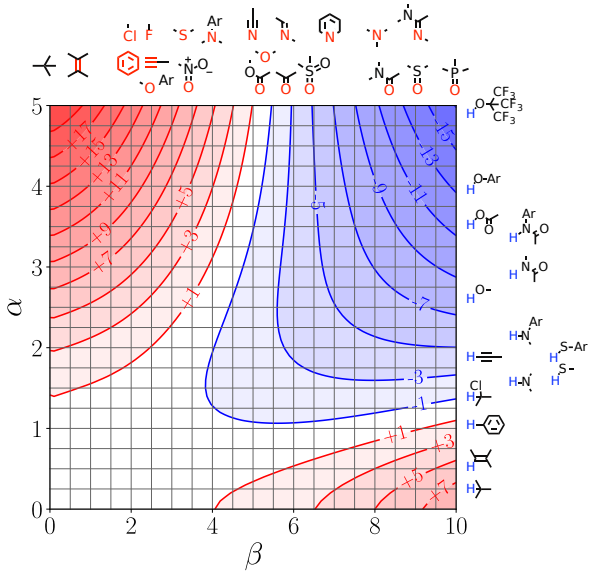


Fig. H.292: FGIP for 45.0% chloroform 55.0% tetrahydrofuran at 298K.

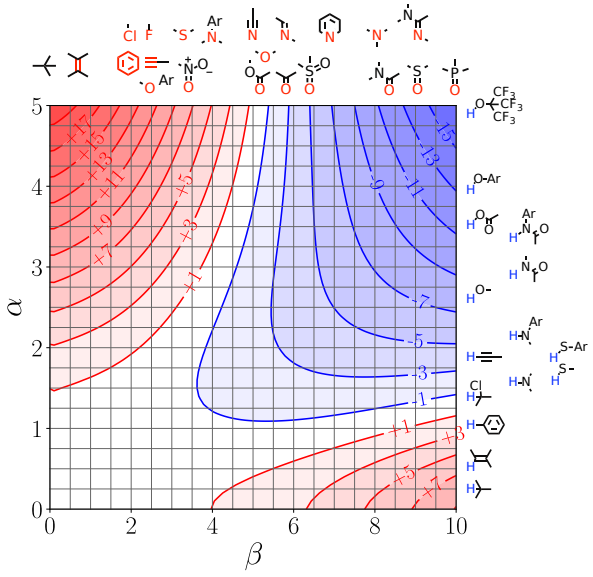


Fig. H.293: FGIP for 50.0% chloroform 50.0% tetrahydrofuran at 298K.

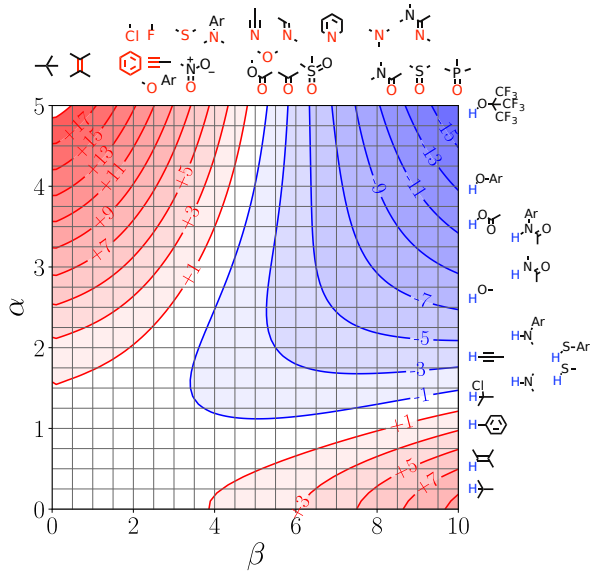


Fig. H.294: FGIP for 55.0% chloroform 45.0% tetrahydrofuran at 298K.

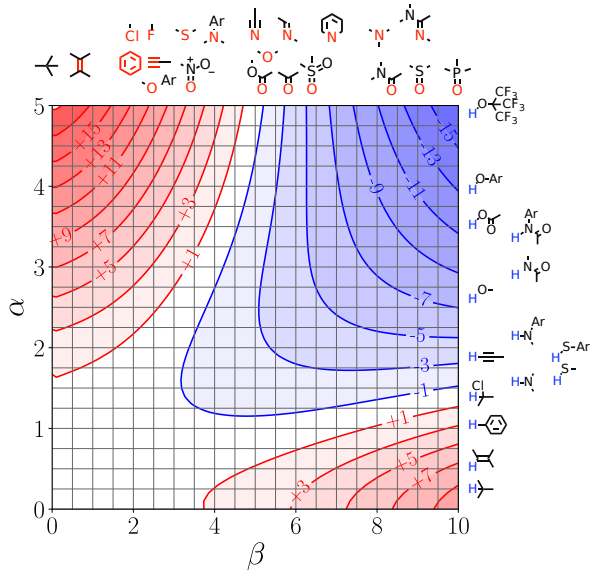


Fig. H.295: FGIP for 60.0% chloroform 40.0% tetrahydrofuran at 298K.

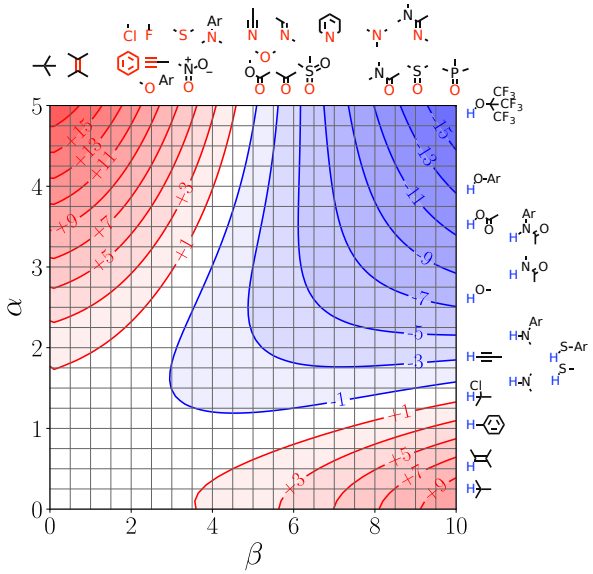


Fig. H.296: FGIP for 65.0% chloroform 35.0% tetrahydrofuran at 298K.

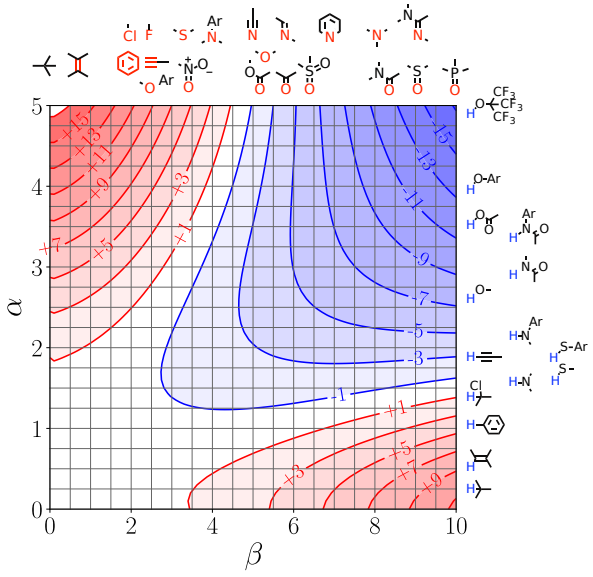


Fig. H.297: FGIP for 70.0% chloroform 30.0% tetrahydrofuran at 298K.

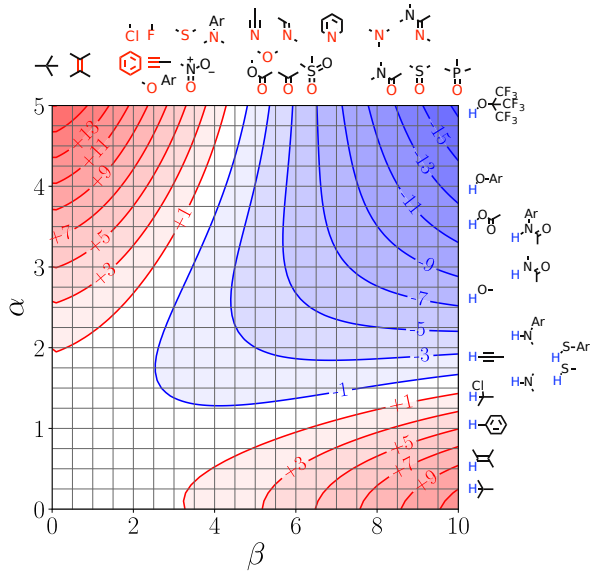


Fig. H.298: FGIP for 75.0% chloroform 25.0% tetrahydrofuran at 298K.

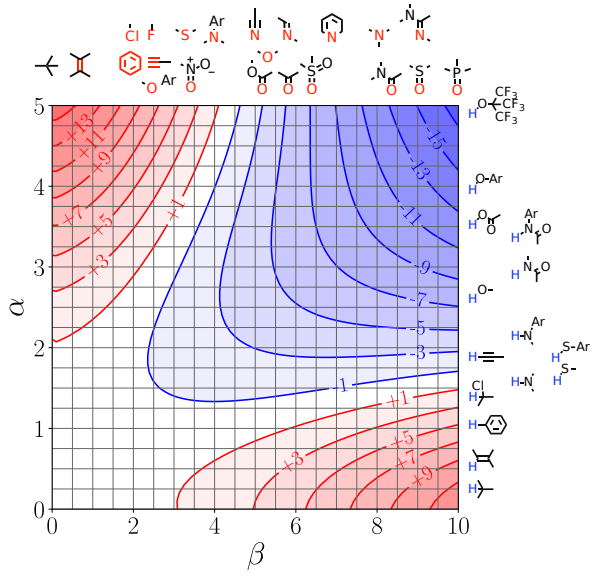


Fig. H.299: FGIP for 80.0% chloroform 20.0% tetrahydrofuran at 298K.

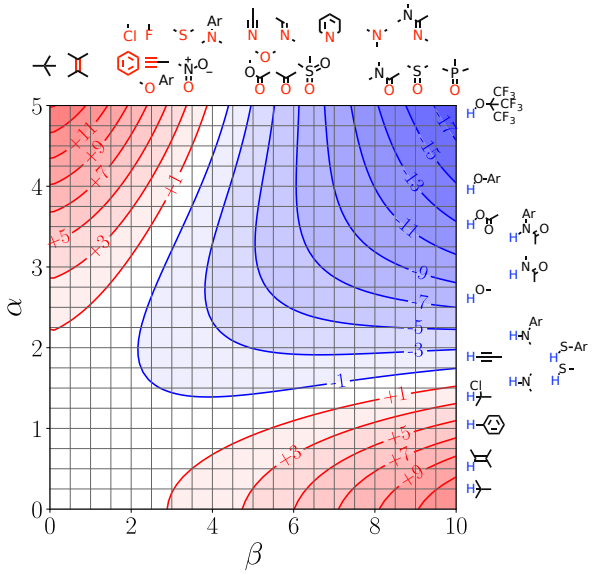


Fig. H.300: FGIP for 85.0% chloroform 15.0% tetrahydrofuran at 298K.

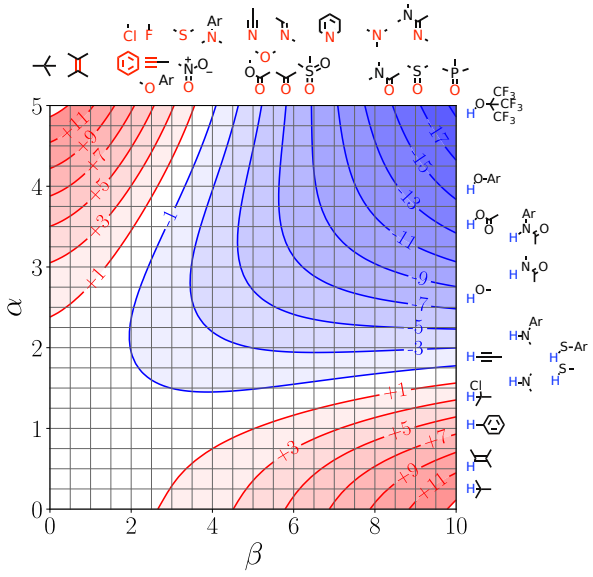


Fig. H.301: FGIP for 90.0% chloroform 10.0% tetrahydrofuran at 298K.

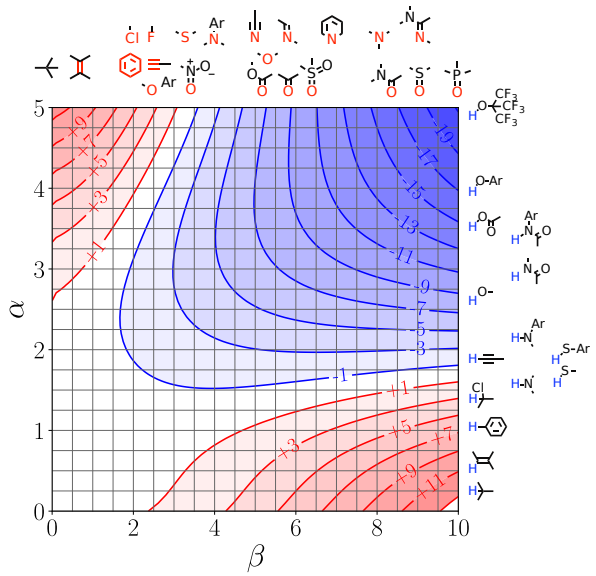


Fig. H.302: FGIP for 95.0% chloroform 5.0% tetrahydrofuran at 298K.

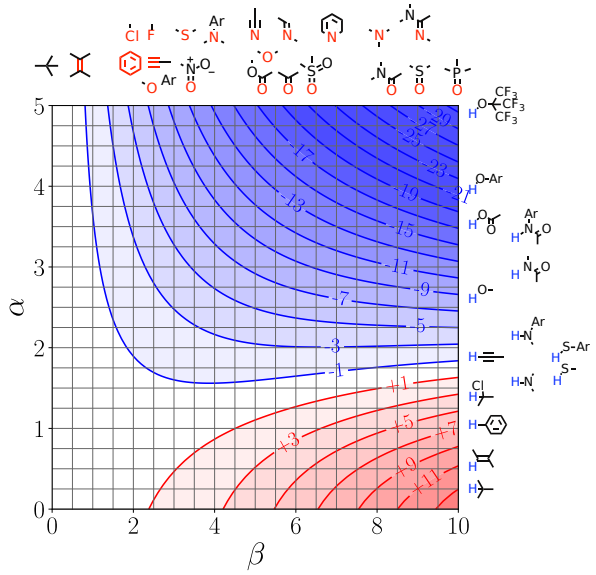


Fig. H.303: FGIP for 100.0% chloroform 0.0% tetrahydrofuran at 298K.

H.3 Calculation of γ from simulation

With the SSIMPLE approach, it is possible to calculate γ directly using equation (H.1), without needing to with $\Delta G^o_{SSIMPLE}$ defined in equation (5.23) and $\Delta\Delta G_{H\ bonds,FGIP}$ defined in equation (5.4).

$$\gamma = \Delta G^o_{SSIMPLE} - \Delta\Delta G_{H\ bonds,FGIP} \tag{H.1}$$

Figure H.304 shows the distribution of γ values for the same set of 261 pure solvents (details in G). Solute values of $\varepsilon_1 = 5.0$ and $\varepsilon_2 = -10.0$ were used to investigate γ from simulation, which should be within the tight binding region. The mean value is 10.53 kJ mol^{-1} with a standard deviation of 0.27 kJ mol^{-1}.

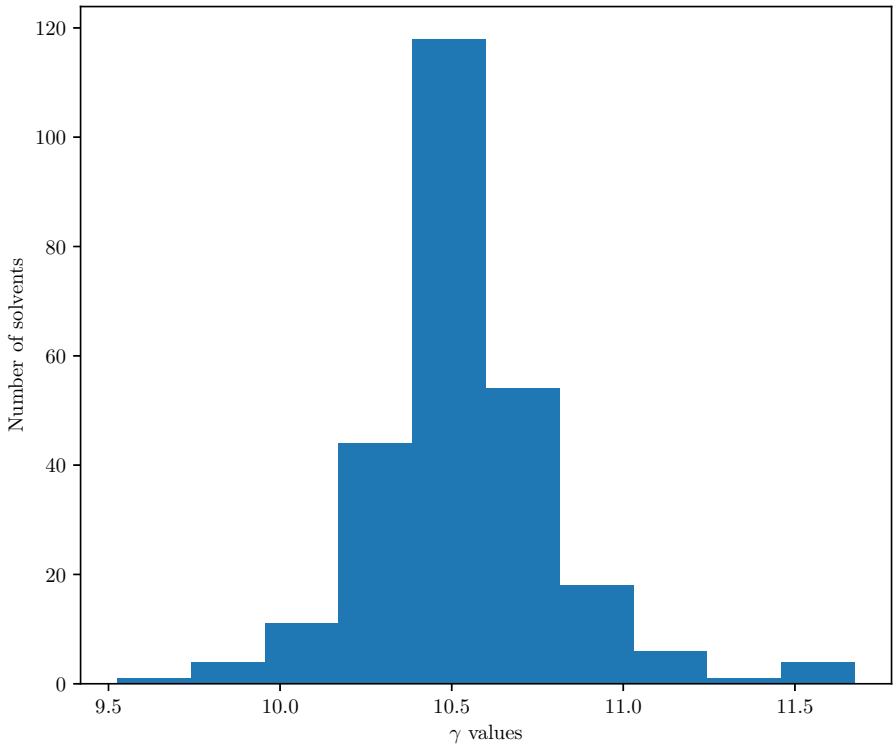


Fig. H.304: Plot of difference between calculated $\Delta\Delta G_{Hbonds}$ and ΔG^o, for 261 solvents, with solute SSIPs of $\varepsilon_1 = 5.0$ and $\varepsilon_2 = -10.0$. 10 bins were used to group values.

Appendix I

Vapour Liquid Equilibria data

I.1 Computational Implementation details

I.1.1 Calculation of interaction energy

The expression for E_0 in equation (7.3) cannot be inverted analytically. A cubic polynomial spline is fitted to values of this function over the range $-200 \leq \varepsilon_{i,298}\varepsilon_{j,298} < 0$, covering a wide range of possible interactions.

Once E_0 has been calculated, it is then used to find $\varepsilon_{i,T}\varepsilon_{j,T}$ at the temperature of interest, for the calculation of K_{ij} as in equation 5 in [158].

I.2 Data for Vapour Liquid equilibria

Information plotted in chapter 7 figures.

Molecule	T/K	$\Delta T/K$	$-\Delta H^*$/kJ mol^{-1}
Water	273.00	0.00	42.70
Water	278.00	5.00	42.52
Water	283.00	10.00	42.34
Water	288.00	15.00	42.15
Water	293.00	20.00	41.96
Water	298.00	25.00	41.77
Water	303.00	30.00	41.58
Water	308.00	35.00	41.39
Water	313.00	40.00	41.19
Water	318.00	45.00	41.00
Water	323.00	50.00	40.81
Water	328.00	55.00	40.62
Water	333.00	60.00	40.44
Water	338.00	65.00	40.26
Water	343.00	70.00	40.08
Water	348.00	75.00	39.90
Water	353.00	80.00	39.73
Water	358.00	85.00	39.56
Water	363.00	90.00	39.40
Water	368.00	95.00	39.25
Water	373.00	100.00	39.10
Methane	90.68	0.00	8.16
Methane	100.00	9.32	8.02
Methane	110.00	19.32	7.94
Methane	111.63	20.95	7.90
Methane	120.00	29.32	7.86
Methane	130.00	39.32	7.86
Methane	140.00	49.32	7.94
Methane	150.00	59.32	8.20
Argon	83.80	0.00	6.30
Argon	85.00	1.20	6.27
Argon	87.28	3.48	6.21
Argon	90.00	6.20	6.18
Argon	95.00	11.20	6.14
Argon	100.00	16.20	6.13
Argon	105.00	21.20	6.13
Argon	110.00	26.20	6.21
Argon	120.00	36.20	6.48

Table I.1 Data for figure 7.2. Experimental data from [278]

Guest InChIKey	Host	Experimental $\frac{K_{293}}{K_{323}}$	Calculated $\frac{K_{293}}{K_{323}}$
AATNZNJRDOVKDD-UHFFFAOYSA-N	methanol	2.083333	1.922781
AATNZNJRDOVKDD-UHFFFAOYSA-N	Alpha-naphthol	2.634052	2.627667
AJFDBNQQDYLMJN-UHFFFAOYSA-N	phenol	2.242511	2.356667
AVPYQKSLYISFPO-UHFFFAOYSA-N	phenol	1.645570	1.687495
AVQQQNCBBIEMEU-UHFFFAOYSA-N	phenol	2.314647	2.451078
AWJUIBRHMBBTKR-UHFFFAOYSA-N	phenol	3.009804	2.202193
BGNGWHSBYQYVRX-UHFFFAOYSA-N	phenol	2.165803	2.022277
BGTOWKSIORTVQH-UHFFFAOYSA-N	phenol	2.054422	1.930665
BLKXLEPPVDUHBY-UHFFFAOYSA-N	phenol	2.375375	2.475406
BLKXLEPPVDUHBY-UHFFFAOYSA-N	methanol	2.161765	1.849410
BLKXLEPPVDUHBY-UHFFFAOYSA-N	Alpha-naphthol	2.527837	2.475406
BLKXLEPPVDUHBY-UHFFFAOYSA-N	Indole	3.209302	1.973004
BSKHPKMHTQYZBB-UHFFFAOYSA-N	phenol	3.012097	2.244975
BUDQDWGNQVEFAC-UHFFFAOYSA-N	phenol	1.953488	1.616152
BXRFQSNOROATLV-UHFFFAOYSA-N	phenol	1.752066	1.602435
CDQSTBHGKNNPSY-UHFFFAOYSA-N	phenol	2.367113	2.311164
CQQUWTMMFMJEFE-UHFFFAOYSA-N	phenol	2.096330	2.119744
CSCPPACGZOOCGX-UHFFFAOYSA-N	phenol	2.048253	1.878353
DENRZWYUOJLTMF-UHFFFAOYSA-N	phenol	1.880952	1.732559
DHXVGJBLRPWPCS-UHFFFAOYSA-N	phenol	2.566914	1.895576
DKLYDESVXZKCFI-UHFFFAOYSA-N	phenol	2.277992	2.160457
DQWPFSLDHJDLRL-UHFFFAOYSA-N	phenol	2.779873	2.524951
DQWPFSLDHJDLRL-UHFFFAOYSA-N	methanol	2.012987	1.873379
DQWPFSLDHJDLRL-UHFFFAOYSA-N	Alpha-naphthol	2.791199	2.524951
DQWPFSLDHJDLRL-UHFFFAOYSA-N	Indole	2.200000	2.001689
DSSYKIVIOFKYAU-UHFFFAOYSA-N	phenol	1.829175	1.930665
DURPTKYDGMDSBL-UHFFFAOYSA-N	phenol	2.343284	1.763564
DVVGIUUJYPYENY-UHFFFAOYSA-N	phenol	2.634528	2.524951
DZBUGLKDJFMEHC-UHFFFAOYSA-N	phenol	2.871795	2.202193
FIQMHBFVRAXMOP-UHFFFAOYSA-N	phenol	2.887551	2.877208
FIQMHBFVRAXMOP-UHFFFAOYSA-N	methanol	2.292453	2.041030
FIQMHBFVRAXMOP-UHFFFAOYSA-N	Alpha-naphthol	3.032510	2.877208
FIQMHBFVRAXMOP-UHFFFAOYSA-N	Indole	2.106164	2.202821
FIQMHBFVRAXMOP-UHFFFAOYSA-N	Carbazole	2.043880	2.377488

Continued on next page

Guest InChIKey	Host	Experimental $\frac{K_{293}}{K_{323}}$	Calculated $\frac{K_{293}}{K_{323}}$
FIQMHBFVRAXMOP-UHFFFAOYSA-N	Pyrrole	2.066667	2.120375
FKNQCJSGGFJEIZ-UHFFFAOYSA-N	phenol	2.877698	2.266767
FKRCODPIKNYEAC-UHFFFAOYSA-N	phenol	1.990950	1.844537
FXHOOIRPVKKKFG-UHFFFAOYSA-N	phenol	2.205882	2.451078
GAEKPEKOJKCEMS-UHFFFAOYSA-N	phenol	2.064935	1.966632
GGSUCNLOZRCGPQ-UHFFFAOYSA-N	phenol	1.238095	1.747965
GGSUCNLOZRCGPQ-UHFFFAOYSA-N	methanol	1.307692	1.493333
GNOIPBMMFNIUFM-UHFFFAOYSA-N	phenol	2.752039	3.121656
GRIXINIGTYIHSN-UHFFFAOYSA-N	phenol	1.750000	1.966632
GXHHFMZALQKUFL-UHFFFAOYSA-N	phenol	1.952381	2.119744
HPMLGNIUXVXALD-UHFFFAOYSA-N	phenol	1.565217	1.414269
HUMNYLRZRPPJDN-UHFFFAOYSA-N	phenol	1.979592	1.732559
HUTDHWXNOSDRBN-UHFFFAOYSA-N	phenol	2.298246	2.333776
HZCDANOFLILNSA-UHFFFAOYSA-N	phenol	2.309568	2.333776
HZCDANOFLILNSA-UHFFFAOYSA-N	methanol	1.689655	1.780443
HZCDANOFLILNSA-UHFFFAOYSA-N	Alpha-naphthol	2.620312	2.333776
HZXRDAIBHPCJLE-UHFFFAOYSA-N	phenol	2.040000	1.913012
IAZDPXIOMUYVGZ-UHFFFAOYSA-N	phenol	3.585670	2.475406
ILVXOBCQQYKLDS-UHFFFAOYSA-N	phenol	3.089018	2.575698
IMFACGCPASFAPR-UHFFFAOYSA-N	phenol	3.010309	2.080030
IMFACGCPASFAPR-UHFFFAOYSA-N	methanol	1.708333	1.655746
IMFACGCPASFAPR-UHFFFAOYSA-N	Alpha-naphthol	2.557471	2.080030
ISXOBTBCNRIIQO-UHFFFAOYSA-N	phenol	3.009021	2.475406
ITQTTZVARXURQS-UHFFFAOYSA-N	phenol	2.833333	2.223452
JFDZBHWFFUWGJE-UHFFFAOYSA-N	phenol	1.956522	1.732559
JHIVVAPYMSGYDF-UHFFFAOYSA-N	phenol	2.251604	1.966632
JIRGTCBOQAJBHY-UHFFFAOYSA-N	phenol	2.178218	2.099764
JJHHIJFTHRNPIK-UHFFFAOYSA-N	phenol	2.505338	2.223452
JLNGEXDJAQASHD-UHFFFAOYSA-N	phenol	2.415423	2.288829
JLZWKZDHHREXTA-UHFFFAOYSA-N	phenol	2.467914	2.099764
JUJWROOIHBZHMG-UHFFFAOYSA-N	phenol	3.035533	2.160457
JUJWROOIHBZHMG-UHFFFAOYSA-N	methanol	1.666667	1.695356
JUJWROOIHBZHMG-UHFFFAOYSA-N	Alpha-naphthol	2.462687	2.160457
JUJWROOIHBZHMG-UHFFFAOYSA-N	Carbazole	1.777778	1.887608

Continued on next page

Guest InChIKey	Host	Experimental $\frac{K_{293}}{K_{323}}$	Calculated $\frac{K_{293}}{K_{323}}$
JUJWROOIHBZHMG-UHFFFAOYSA-N	Pyrrole	1.750000	1.741359
JYYNAJVZFGKDEQ-UHFFFAOYSA-N	phenol	2.968571	2.333776
KJPRLNWUNMBNBZ-QPJJXVBHSA-N	phenol	2.055556	1.861341
KSMVZQYAVGTKIV-UHFFFAOYSA-N	phenol	2.039841	1.717347
KWOLFJPFCHCOCG-UHFFFAOYSA-N	phenol	1.507246	1.913012
KXKVLQRXCPHEJC-UHFFFAOYSA-N	phenol	2.046229	1.811545
KYVIFDXEMABEDB-UHFFFAOYSA-N	phenol	2.464435	2.451078
LFSAPCRASZRSKS-UHFFFAOYSA-N	phenol	2.089184	1.913012
LOWMYOWHQMKBTM-UHFFFAOYSA-N	phenol	2.733195	2.500029
LRMLWYXJORUTBG-UHFFFAOYSA-N	phenol	2.871442	3.058469
LRMLWYXJORUTBG-UHFFFAOYSA-N	methanol	2.587097	2.125286
LRMLWYXJORUTBG-UHFFFAOYSA-N	Alpha-naphthol	3.209123	3.058469
LRMLWYXJORUTBG-UHFFFAOYSA-N	Indole	2.207506	2.304333
MHDVGSVTJDSBDK-UHFFFAOYSA-N	phenol	1.873984	1.717347
MJUJXFBTEFXVKU-UHFFFAOYSA-N	phenol	2.473644	2.403298
MJUJXFBTEFXVKU-UHFFFAOYSA-N	methanol	1.741935	1.814375
MJUJXFBTEFXVKU-UHFFFAOYSA-N	Alpha-naphthol	2.253301	2.403298
MJUJXFBTEFXVKU-UHFFFAOYSA-N	Indole	2.411765	1.931093
MTZQAGJQAFMTAQ-UHFFFAOYSA-N	phenol	1.791667	1.795353
MXHTZQSKTCCMFG-UHFFFAOYSA-N	phenol	1.285714	1.588898
MXHTZQSKTCCMFG-UHFFFAOYSA-N	methanol	1.259259	1.415734
MXHTZQSKTCCMFG-UHFFFAOYSA-N	Alpha-naphthol	1.266667	1.588898
MZBIWKMCTWJLPT-UHFFFAOYSA-N	phenol	2.838480	2.451078
MZBIWKMCTWJLPT-UHFFFAOYSA-N	methanol	1.946429	1.837610
MZBIWKMCTWJLPT-UHFFFAOYSA-N	Alpha-naphthol	2.466515	2.451078
MZBIWKMCTWJLPT-UHFFFAOYSA-N	Indole	1.936709	1.958885
NBBJYMSMWIIQGU-UHFFFAOYSA-N	phenol	1.969432	1.702326
NUJGJRNETVAIRJ-UHFFFAOYSA-N	phenol	1.689655	1.702326
NYYLZXREFNYPKB-UHFFFAOYSA-N	phenol	2.701170	2.601528
NYYLZXREFNYPKB-UHFFFAOYSA-N	methanol	2.129412	1.910247
NYYLZXREFNYPKB-UHFFFAOYSA-N	Alpha-naphthol	2.435610	2.601528
NYYLZXREFNYPKB-UHFFFAOYSA-N	Indole	2.345865	2.045840
OISVCGZHLKNMSJ-UHFFFAOYSA-N	phenol	2.993711	2.160457
PJGSXYOJTGTZAV-UHFFFAOYSA-N	phenol	1.907950	1.878353

Continued on next page

Guest InChIKey	Host	Experimental $\frac{K_{293}}{K_{323}}$	Calculated $\frac{K_{293}}{K_{323}}$
POLCUAVZOMRGSN-UHFFFAOYSA-N	phenol	1.941828	1.795353
PPMDSXRMQXBCJW-UHFFFAOYSA-N	phenol	2.606613	2.601528
QAUUDNIGJSLPSX-UHFFFAOYSA-N	phenol	1.479532	1.588898
QJUBWCXHSCFIHV-UHFFFAOYSA-N	phenol	2.554217	2.333776
QSJXEFYPDANLFS-UHFFFAOYSA-N	phenol	1.844340	1.644132
RDOXTESZEPMUJZ-UHFFFAOYSA-N	phenol	1.683333	1.523779
RHLIPLVNXYUJQV-UHFFFAOYSA-N	phenol	1.870968	2.080030
RTZKZFJDLAIYFH-UHFFFAOYSA-N	phenol	2.133333	1.811545
RVAQSYWDOSHWGP-UHFFFAOYSA-N	phenol	1.801477	2.181195
RVAQSYWDOSHWGP-UHFFFAOYSA-N	methanol	1.789474	1.705563
RVAQSYWDOSHWGP-UHFFFAOYSA-N	Alpha-naphthol	2.259669	2.181195
RVAQSYWDOSHWGP-UHFFFAOYSA-N	Indole	1.630435	1.800939
RWCCWEUUXYIKHB-UHFFFAOYSA-N	phenol	1.688985	1.811545
SECXISVLQFMRJM-UHFFFAOYSA-N	phenol	2.579114	2.451078
SMWDFEZZVXVKRB-UHFFFAOYSA-N	phenol	3.136612	2.181195
SNZSAFILJOCMFM-UHFFFAOYSA-N	phenol	3.560456	3.089886
SNZSAFILJOCMFM-UHFFFAOYSA-N	Indole	2.101124	2.321792
SUNMBRGCANLOEG-UHFFFAOYSA-N	phenol	1.575581	1.644132
SUSQOBVLVYHIEX-UHFFFAOYSA-N	phenol	2.208333	1.732559
SYBYTAAJFKOIEJ-UHFFFAOYSA-N	phenol	1.892276	1.827939
UBUCNCOMADRQHX-UHFFFAOYSA-N	phenol	1.523810	1.658399
UIQGEWJEWJMQSL-UHFFFAOYSA-N	phenol	1.873727	1.827939
UKROGNGFIXLRHI-UHFFFAOYSA-N	phenol	2.040359	2.139974
UMFJAHHVKNCGLG-UHFFFAOYSA-N	phenol	1.960000	1.895576
USIUVYZYUHIAEV-UHFFFAOYSA-N	phenol	1.967213	1.498811
VEUUMBGHMNQHGO-UHFFFAOYSA-N	phenol	1.486207	1.702326
VEZXCJBBBCKRPI-UHFFFAOYSA-N	phenol	2.040000	1.763564
VSYFZULSKMFUJJ-UHFFFAOYSA-N	phenol	2.682160	2.475406
WEVYAHXRMPXWCK-UHFFFAOYSA-N	phenol	2.120000	1.779360
WFJXYIUAMJAURQ-UHFFFAOYSA-N	phenol	2.520433	2.500029
WMFABESKCHGSRC-UHFFFAOYSA-N	phenol	1.515152	1.462237
WVLBCYQITXONBZ-UHFFFAOYSA-N	phenol	2.318124	2.451078
WVLBCYQITXONBZ-UHFFFAOYSA-N	methanol	1.709677	1.837610
WVLBCYQITXONBZ-UHFFFAOYSA-N	Alpha-naphthol	2.450719	2.451078

Continued on next page

Guest InChIKey	Host	Experimental $\frac{K_{293}}{K_{323}}$	Calculated $\frac{K_{293}}{K_{323}}$
WVLBCYQITXONBZ-UHFFFAOYSA-N	Indole	2.011494	1.958885
WYURNTSHIVDZCO-UHFFFAOYSA-N	phenol	2.476008	1.913012
XDOKVFJJVSLHDJ-UHFFFAOYSA-N	phenol	2.192751	2.288829
XDOKVFJJVSLHDJ-UHFFFAOYSA-N	methanol	1.500000	1.758436
XDOKVFJJVSLHDJ-UHFFFAOYSA-N	Alpha-naphthol	2.475789	2.288829
XDOKVFJJVSLHDJ-UHFFFAOYSA-N	Indole	1.981818	1.864193
XEKOWRVHYACXOJ-UHFFFAOYSA-N	phenol	1.958084	1.795353
XNLICIUVMPYHGG-UHFFFAOYSA-N	phenol	2.086059	1.861341
XZZNDPSIHUTMOC-UHFFFAOYSA-N	phenol	1.859410	2.119744
XZZNDPSIHUTMOC-UHFFFAOYSA-N	Indole	1.924528	1.764709
YEJRWHAVMIAJKC-UHFFFAOYSA-N	phenol	2.036145	1.930665
YFTHZRPMJXBUME-UHFFFAOYSA-N	phenol	2.574713	2.041290
YFTHZRPMJXBUME-UHFFFAOYSA-N	methanol	1.545455	1.636669
YFTHZRPMJXBUME-UHFFFAOYSA-N	Alpha-naphthol	2.663793	2.041290
YHQMSHVVGOSZEW-UHFFFAOYSA-N	phenol	2.606324	2.575698
YHQMSHVVGOSZEW-UHFFFAOYSA-N	Alpha-naphthol	2.684474	2.575698
YKOQQFDCCBKROY-UHFFFAOYSA-N	phenol	2.441016	2.266767
YLQBMQCUIZJEEH-UHFFFAOYSA-N	phenol	1.176471	1.390100
YMBSJQHEFLROIX-UHFFFAOYSA-N	phenol	2.272727	2.099764
YRKCREAYFQTBPV-UHFFFAOYSA-N	phenol	1.925267	1.779360
ZAFNJMIOTHYJRJ-UHFFFAOYSA-N	phenol	2.371571	1.844537
ZMANZCXQSJIPKH-UHFFFAOYSA-N	phenol	3.469466	2.223452
ZMANZCXQSJIPKH-UHFFFAOYSA-N	methanol	1.828571	1.726344
ZMANZCXQSJIPKH-UHFFFAOYSA-N	Alpha-naphthol	2.850785	2.223452
ZMXDDKWLCZADIW-UHFFFAOYSA-N	phenol	2.377273	2.202193
ZSSWXNPRLJLCDU-UHFFFAOYSA-N	phenol	2.939394	3.186262
ZSSWXNPRLJLCDU-UHFFFAOYSA-N	methanol	2.748503	2.183856
ZSSWXNPRLJLCDU-UHFFFAOYSA-N	Alpha-naphthol	3.468203	3.186262
ZSSWXNPRLJLCDU-UHFFFAOYSA-N	Indole	2.088235	2.375098
ZSSWXNPRLJLCDU-UHFFFAOYSA-N	Carbazole	2.330632	2.583100
ZSSWXNPRLJLCDU-UHFFFAOYSA-N	Pyrrole	1.946488	2.277467
ZTQSAGDEMFDKMZ-UHFFFAOYSA-N	phenol	2.192308	1.717347
ZWEHNKRNPOVVGH-UHFFFAOYSA-N	phenol	2.304251	1.861341

Table I.2 Data for figure 7.3

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
methane	51.16	32.76	5.5	190	3.728362
ethane	73.41	52.43	7.9	305	5.606174
propane	93.70	71.99	10.0	370	6.611560
n-butane	113.84	91.77	12.2	425	7.474120
n-pentane	133.89	111.30	14.3	470	8.182759
n-hexane	154.08	131.08	16.5	507	8.764723
n-heptane	174.06	150.64	18.6	540	9.287384
n-octane	194.20	170.31	20.8	569	9.747501
n-nonane	214.27	189.95	22.9	594	10.144051
n-decane	234.41	209.63	25.1	617	10.510160
n-undecane	254.43	229.25	27.2	639	10.862199
n-dodecane	274.59	248.96	29.4	658	11.165466
n-tridecane	294.21	268.37	31.5	676	11.453926
n-tetradecane	313.72	287.91	33.6	697	11.794546
n-pentadecane	333.21	307.23	35.6	707	11.950484
n-hexadecane	352.89	326.79	37.7	721	12.175071
n-heptadecane	372.43	346.09	39.8	733	12.366991
n-octadecane	392.14	365.67	41.9	745	12.559594
n-nonadecane	411.65	384.95	44.0	756	12.736195
n-eicosane	431.32	404.55	46.1	767	12.913277
n-heneicosane	450.92	423.85	48.2	782	13.158320
n-docosane	470.57	443.40	50.3	792	13.319574
n-tricosane	490.12	462.73	52.4	801	13.464518
n-tetracosane	509.78	482.31	54.5	810	13.609769
n-pentacosane	529.37	501.59	56.6	819	13.755450
n-hexacosane	549.12	521.09	58.7	827	13.884582
n-heptacosane	568.59	540.43	60.8	835	14.014036
n-octacosane	588.38	559.99	62.9	843	14.143689
n-nonacosane	607.89	579.36	65.0	850	14.256840
n-triacontane	627.62	598.83	67.1	857	14.370181
isobutane	111.72	91.62	11.9	408	7.175832
isopentane	129.63	111.02	13.9	460	8.009608
neopentane	127.48	111.27	13.6	434	7.556091
2-methylpentane	149.67	130.62	16.0	497	8.593049

Continued on next page

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
3-methylpentane	147.40	130.32	15.8	504	8.714877
2,2-dimethylbutane	143.09	130.28	15.3	489	8.455609
2,3-dimethylbutane	143.99	130.33	15.4	500	8.645684
2-methylhexane	169.75	150.24	18.2	530	9.116229
3-methylhexane	165.61	150.14	17.7	535	9.202442
3-ethylpentane	162.87	149.86	17.4	540	9.289044
2,2-dimethylpentane	160.21	149.75	17.1	520	8.945232
2,3-dimethylpentane	158.33	149.53	16.9	537	9.238142
2,4-dimethylpentane	164.77	150.10	17.6	520	8.944512
3,3-dimetmylpentane	158.33	149.53	16.9	536	9.220939
2,2,3-trimethylbutane	156.02	148.86	16.7	531	9.136345
2-methylheptane	189.81	169.87	20.3	559	9.576942
3-methylheptane	185.26	168.30	19.8	564	9.665338
4-methylheptane	185.17	168.25	19.8	562	9.631151
3-ethylhexane	182.67	169.49	19.5	565	9.680394
2,2-dimethylhexane	183.15	169.51	19.6	550	9.423359
2,3-dimethylhexane	181.18	169.20	19.4	563	9.646630
2,4-dimethylhexane	182.94	169.35	19.6	553	9.475031
2,5-dimethylhexane	185.52	169.71	19.8	550	9.423021
3,3-dimethylhexane	178.03	169.07	19.0	562	9.629721
3,4-dimethylhexane	179.48	168.94	19.2	569	9.749893
2-methyl-3-ethylpentane	179.60	168.79	19.2	567	9.715886
3-methyl-3-ethylpentane	172.99	168.45	18.5	576	9.870714
2,2,3-trimethylpentane	173.45	167.22	18.6	563	9.650109
2,2,4-trimethylpentane	177.76	169.25	19.0	544	9.320994
2,3,3-trimethylpentane	170.70	168.00	18.3	573	9.820109
2,3,4-trimethylpentane	173.42	167.33	18.5	566	9.701334
2,2,3,3-tetramethylbutane	166.13	167.06	17.8	568	9.736099
2-methyloctane	207.91	188.46	22.2	587	10.026651
3-methyloctane	203.54	188.35	21.8	590	10.078055
4-methyloctane	203.33	188.15	21.7	588	10.044184
3-ethylheptane	200.75	187.90	21.5	594	10.147044
2,2-dimethylheptane	201.91	188.46	21.6	577	9.855840
2,6-dimethylheptane	203.97	188.48	21.8	576	9.838730

Continued on next page

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
2,2,3-trimethylhexane	193.05	187.37	20.6	591	10.096577
2,2,4-trimethylhexane	192.99	187.67	20.6	574	9.805721
2,2,5-trimethylhexane	197.76	188.12	21.2	568	9.702587
2,3,3-trimethylhexane	191.19	187.19	20.4	599	10.233518
2,3,5-trimethylhexane	195.92	187.74	21.0	582	9.942285
2,4,4-trimethylhexane	191.04	187.44	20.4	582	9.942720
3,3,4-trimethylhexane	189.05	186.91	20.2	604	10.319364
3,3-diethylpentane	187.18	187.01	20.0	610	10.421721
2,2-dimethyl-3-ethylpentane	190.05	187.31	20.3	589	10.062497
2,4-dimethyl-3-ethylpentane	188.30	186.87	20.1	591	10.097318
2,2,3,3-tetramethylpentane	182.43	186.37	19.5	611	10.439790
2,2,3,4-tetramethylpentane	186.26	187.14	19.9	592	10.114002
2,2,4,4-tetramethylpentane	186.56	187.52	20.0	571	9.754685
2,3,3,4-tetramethylpentane	184.01	186.69	19.7	607	10.370955
2-methylnonane	227.70	207.93	24.4	609	10.375958
3-methylnonane	225.59	207.62	24.1	614	10.461531
4-methylnonane	225.54	207.58	24.1	619	10.546773
5-methylnonane	225.56	207.66	24.1	610	10.393329
2,7-dimethyloctane	223.72	207.77	23.9	600	10.222813
3,3,4-trimethylheptane	208.54	206.31	22.3	628	10.701745
3,3,5-trimethylheptane	208.84	206.62	22.3	609	10.377579
2,2,3,3-tetramethylhexane	202.00	205.84	21.6	623	10.617142
2,2,5,5-tetramethylhexane	211.43	207.47	22.6	581	9.899443
2,4-dimethyl-3-isopropyl-pentane	204.31	205.51	21.9	614	10.464183
cyclopropane	82.02	61.85	8.8	398	7.199997

Continued on next page

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
cyclobutane	100.05	80.02	10.7	460	8.158669
cyclopentane	116.22	98.21	12.4	512	8.970222
cyclohexane	131.30	117.07	14.0	553	9.605531
cycloheptane	146.70	136.36	15.7	604	10.425594
cyclooctane	160.28	155.09	17.1	647	11.116711
cyclononane	173.85	173.47	18.6	682	11.676863
cis-decahydronaphthalene	180.93	181.25	19.4	702	12.004001
trans-decahydronaphtmalene	184.38	181.76	19.7	687	11.746570
methylcyclopropane	102.21	81.52	10.9	437	7.741242
ethylcyclopropane	122.12	101.15	13.1	482	8.431477
cis-1,2-dimethylcyclopropane	120.72	100.95	12.9	484	8.467335
trans-1,2-dimethylcyclopropane	122.23	101.10	13.1	469	8.204283
methylcyclobutane	119.73	99.76	12.8	487	8.525117
ethylcyclobutane	138.17	119.29	14.8	527	9.146321
methylcyclopentane	134.97	117.96	14.4	533	9.255019
ethylcyclopentane	153.21	137.47	16.4	569	9.818444
1,1-dimethylcyclopentane	147.21	136.37	15.7	547	9.441695
cis-1,2-dimethylcyclopentane	149.25	137.10	16.0	565	9.750415
trans-1,2-dimethylcyclopentane	151.75	137.49	16.2	553	9.542302
cis-1,3-dimethylcyclopentane	153.75	137.60	16.4	551	9.507504
trans-1,3-dimethylcyclopentane	153.69	137.61	16.4	553	9.541988
n-propylcyclopentane	172.95	156.95	18.5	603	10.356618
isopropylcyclopentane	168.12	156.68	18.0	593	10.185444
1-methyl-1-ethylcyclopentane	164.55	156.38	17.6	582	9.997138

Continued on next page

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
cis-l-methyl-2-ethyl-cyclopentane	165.01	155.38	17.6	592	10.171073
trans-l-methyl-2-ethyl-cyclopentane	169.12	156.71	18.1	581	9.979267
cis-l-methyl-3-ethyl-cyclopentane	171.48	156.91	18.3	581	9.978848
trans-1-methyl-3-ethyl-cyclopentane	168.72	156.36	18.0	581	9.980003
1,1,2-trimetmylcyclopentane	161.46	155.29	17.3	570	9.793282
1,1,3-trimetmylcyclopentane	165.87	155.70	17.7	557	9.569089
l,cis-2,cis-3-trimethyl-cyclopentane	163.84	155.43	17.5	584	10.033518
l,cis-2,trans-3-trimethyl-cyclopentane	164.28	155.42	17.6	576	9.896094
l,trans-2,cis-3-trimethyl-cyclopentane	166.76	155.94	17.8	566	9.723209
l,cis-2,cis-4-trimethyl-cyclopentane	167.41	155.54	17.9	575	9.878660
l,cis-2,trans-4-trimetmyl-cyclopentane	165.88	155.55	17.7	576	9.895819
l,trans-2,cis-4-trimethyl-cyclopentane	168.70	155.86	18.0	564	9.689016
n-butylcyclopentane	191.00	175.16	20.4	621	10.629408
isobutylcyclopentane	186.43	175.24	19.9	624	10.680614
1-methyl-1-n-propyl-cyclopentane	182.96	174.78	19.6	606	10.373322
1,1-diethylcyclopentane	178.25	174.46	19.1	612	10.476594
cis-1,2-diethylcyclopentane	183.32	174.72	19.6	617	10.561723
1,1-dimethyl-2-ethyl-cyclopentane	177.32	174.61	19.0	594	10.168201
n-pentylcyclopentane	212.01	195.78	22.7	644	10.989081

Continued on next page

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
n-hexylcyclopentane	231.50	215.09	24.8	665	11.320784
n-heptylcyclopentane	251.14	234.52	26.9	684	11.621320
n-octylcyclopentane	270.70	253.95	29.0	701	11.890336
n-nonylcyclopentane	290.33	273.38	31.1	717	12.144392
n-decylcyclopentane	309.92	292.81	33.1	732	12.383150
n-undecylcyclopentane	329.54	312.25	35.2	746	12.606351
n-oodecylcyclopentane	349.17	331.67	37.3	759	12.813823
n-trrdecylcyclopentane	368.76	351.07	39.4	771	13.005413
n-tetradecylcyclopentane	388.43	370.57	41.5	782	13.180943
n-pentadecylcyclopentane	407.97	389.94	43.6	792	13.340434
n-hexadecylcyclopentane	427.59	409.39	45.7	803	13.517376
n-heptadecylcyclopentane	447.11	428.77	47.8	811	13.644419
n-octadecylcyclopentane	466.80	448.27	49.9	821	13.805575
n-nonadecylcyclopentane	486.34	467.64	52.0	829	13.933590
n-eicosylcyclopentane	505.98	487.08	54.1	836	14.045185
methylcyclohexane	146.62	136.37	15.7	572	9.873217
ethylcyclohexane	167.19	156.12	17.9	609	10.461499
1,1-dimethylcyclohexane	161.47	155.68	17.3	591	10.153241
cis-1,2-dimethylcyclohexane	162.19	155.69	17.3	606	10.410916
trans-1,2-dimethylcyclohexane	164.61	155.99	17.6	596	10.238465
cis-i,3-dimetmylcyclohexane	167.08	156.30	17.9	591	10.151905
trans-1,3-dimethylcyclohexane	164.41	155.97	17.6	598	10.272866
cis-i,4-dimethylcyclohexane	164.50	155.98	17.6	598	10.272844
trans-1,4-dimethylcyclohexane	167.20	156.43	17.9	590	10.134449
n-propylcyclohexane	186.80	175.55	20.0	639	10.936793
isopropylcyclohexane	181.32	175.36	19.4	636	10.885792

Continued on next page

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
n-butylcyclohexane	206.23	194.91	22.1	667	11.382879
isobutylcyclohexane	200.59	194.33	21.5	660	11.264304
sec-butylcyclohexane	198.37	194.74	21.2	650	11.093016
tert-butylcyclohexane	193.32	193.87	20.7	645	11.008987
1-methyl-4-isopropyl-cyclohexane	198.06	194.15	21.2	637	10.872026
n-pentylcyclohexane	225.81	214.32	24.2	670	11.406878
n-hexylcyclohexane	245.29	233.71	26.2	688	11.690164
n-heptylcyclohexane	264.93	253.14	28.3	706	11.975915
n-octylcyclohexane	284.56	272.55	30.4	723	12.246713
n-nonylcyclohexane	304.18	291.93	32.5	738	12.485305
n-decylcyclohexame	323.78	311.40	34.6	751	12.691408
n-undecylcyclohexane	343.42	330.82	36.7	764	12.898743
n-dodecylcyclohexane	362.96	350.21	38.8	776	13.090219
n-tridecylcyclohexane	382.57	369.58	40.9	787	13.265706
n-tetradecylcyclohexane	402.10	388.99	43.0	797	13.425079
n-pentadecylcyclohexane	421.64	408.36	45.1	807	13.585134
n-hexadecylcyclohexane	441.25	427.85	47.2	817	13.745713
n-meptadecylcyclohexane	460.82	447.29	49.3	825	13.873180
n-octadecylcyclohexane	480.45	466.77	51.4	833	14.001103
n-nonadecylcyclohexane	500.02	486.19	53.5	841	14.129456
n-eicosylcyclohexane	520.00	506.00	55.6	849	14.258062
ethylcycloheptane	181.78	175.08	19.4	640	10.954770
bicyclohexyl	218.38	220.30	23.4	727	12.369293
1-methyl-[cis-decahydro-naphthalene]	194.84	199.43	20.8	755	12.876985
1-methyl-[trans-decahydro-naphthalene]	198.81	200.01	21.3	743	12.671372
1-ethyl-[cis-decahydro-naphthalene]	209.97	218.53	22.5	729	12.405655
1-ethyl-[trans-decahydro-naphthalene]	214.03	219.22	22.9	714	12.149499

Continued on next page

Name	Surface Area/ Å^2	Volume/ Å^3	N	T_c/K	E_{exp}/kJ mol^{-1}
9-ethyl-[cis-decahydronaphthalene]	206.78	217.86	22.1	721	12.270400
9-ethyl-[trans-decahydro-naphthalene]	207.34	217.97	22.2	709	12.066034

Table I.3 Data for figure 7.4. Experimental data from [334]

T/K	$[vapour]_T$/M	$[liquid]_T$/M	c_T/M	Experimental E_{exp}/ kJ mol^{-1}	$1-\phi_b$	ϕ_b	$\frac{\phi_b}{1-\phi_b}$	Calculated E_{exp}/ kJ mol^{-1}
273.15	0.000269	55.5	55.5	71.72	5.10%	94.90%	18.61	100.255399
283.15	0.000522	55.5	55.5	74.03	5.67%	94.33%	16.65	89.926653
293.15	0.000960	55.5	55.5	73.35	6.25%	93.75%	15.00	81.208991
303.15	0.001690	55.3	55.3	70.60	6.85%	93.15%	13.59	73.786031
313.15	0.002840	55.1	55.1	66.76	7.47%	92.53%	12.38	67.414536
323.15	0.004600	54.9	54.9	62.53	8.11%	91.89%	11.34	61.905095
333.15	0.007220	54.6	54.6	58.16	8.75%	91.25%	10.43	57.107833
343.15	0.011000	54.3	54.3	53.97	9.41%	90.59%	9.63	52.904455
353.15	0.016300	54.0	54.0	50.21	10.07%	89.93%	8.93	49.200671
363.15	0.023500	53.6	53.7	46.72	10.74%	89.26%	8.31	45.918484
373.15	0.033200	53.2	53.3	43.53	11.43%	88.57%	7.75	42.995142
383.15	0.045900	52.8	52.9	40.79	12.11%	87.89%	7.26	40.380327
393.15	0.062300	52.4	52.5	38.25	12.80%	87.20%	6.81	38.029870
403.15	0.083200	51.9	52.0	35.99	13.50%	86.50%	6.41	35.908890
413.15	0.109000	51.4	51.6	33.96	14.19%	85.81%	6.05	33.987559
423.15	0.142000	50.9	51.1	32.14	14.89%	85.11%	5.71	32.240848
433.15	0.181000	50.4	50.6	30.51	15.60%	84.40%	5.41	30.647343
443.15	0.224000	49.9	50.1	29.01	16.30%	83.70%	5.14	29.194147
453.15	0.287000	49.3	49.6	27.70	17.00%	83.00%	4.88	27.850608
463.15	0.356000	48.7	49.0	26.50	17.71%	82.29%	4.65	26.618434
473.15	0.437000	48.0	48.5	25.39	18.41%	81.59%	4.43	25.480901
483.15	0.534000	47.4	47.9	24.39	19.11%	80.89%	4.23	24.427672
502.55	0.771000	46.0	46.7	22.55	20.47%	79.53%	3.88	22.590786
533.15	1.320000	43.6	44.9	20.37	22.61%	77.39%	3.42	20.155826
552.55	1.840000	41.7	43.5	19.08	23.97%	76.03%	3.17	18.832700
583.15	3.070000	38.4	41.4	17.48	26.11%	73.89%	2.83	17.027667
602.55	4.270000	35.6	39.8	16.40	27.50%	72.50%	2.64	16.006191
633.15	8.070000	29.3	37.4	15.13	29.77%	70.23%	2.36	14.543948
644.25	11.900000	24.2	36.1	14.48	30.72%	69.28%	2.25	13.994731
647.02	17.200000	17.2	34.4	13.46	31.19%	68.81%	2.21	13.738336

Table I.4 Data for Figures 7.5 and 7.6. Experimental data from [334]

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
(2R,3S)-2,3-butanediol	10.95	10.88
1,1,1,3,3,3-hexafluoro-2-propanol	9.56	10.44
1,1,1-trichloroethane	9.97	9.95
1,1,2,2-tetrachloroethane	9.45	8.78
1,1,2-trichloroethane	10.73	10.30
1,1-dichloroethane	11.81	11.96
1,2,3,4-tetrahydronaphthalene	7.30	7.07
1,2,4-trichlorobenzene	8.02	7.46
1,2-butanediol	11.09	10.88
1,2-dibromoethane	11.55	10.41
1,2-dichloroethane	12.59	12.30
1,2-dimethoxybenzene	7.83	7.49
1,2-dimethyoxyethane	9.57	10.13
1,2-propanediol	13.57	13.27
1,3,5-trimethylbenzene	7.17	7.29
1,3-dioxane	11.67	11.32
1,3-dioxolan	14.37	13.36
1,3-propanediol	13.79	13.09
1,4-butanediol	11.24	11.10
1,4-dichlorobutane	9.51	8.79
1,5-pentanediol	9.47	9.32
1,8-cineole	5.96	6.04
1-bromobutane	9.26	9.36
1-butanol	10.87	11.15
1-decanol	5.24	5.44
1-dodecanol	4.46	4.67
1-hexanol	7.99	8.19
1-methyl-2-pyrrolidinethione	11.66	9.11
1-nitropropane	11.17	11.24
1-octanol	6.31	6.54
1-pentanol	9.22	9.47
1-propanol	13.32	13.61

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
2,2,2-trifluoroethanol	13.82	14.67
2,2,4,4-tetramethyl-3-pentanone	5.77	6.12
2,2,4-trimethylpentane	6.24	6.59
2,3-butanedione	11.39	11.96
2,4,5-trimethylacetophenone	7.50	6.07
2,4,6-trimethylpyridine	7.52	7.61
2,4-dimethylphenol	8.32	8.08
2,4-dimethylpyridine	8.66	8.61
2,4-pentanedione	9.71	10.17
2,6-dimethylpyridine	8.57	8.58
2-bromopyridine	10.48	9.77
2-butanol	10.83	11.09
2-butanone	10.93	11.71
2-chloroethanol	15.48	13.99
2-cyanoethanol	14.64	13.86
2-cyanopyridine	10.39	10.33
2-ethoxyethanol	10.27	10.46
2-heptanone	7.11	7.53
2-methoxyethanol	12.61	12.66
2-methoxyphenol	9.09	8.75
2-methyl-1-propanol	10.76	11.12
2-methyl-2-butanol	9.13	9.47
2-methyl-2-propanol	10.54	11.22
2-methylbutane	8.51	9.44
2-methylpyridine	10.09	9.96
2-methyltetrahydrofuran	9.91	10.07
2-nitropropane	11.04	11.23
2-pentanone	9.30	9.84
2-phenylethanol	8.38	8.08
2-propanol	13.00	13.59
2-pyrrolidinone	13.01	12.23
3,3-dimethyl-2-butanone	8.00	8.55

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
3-bromopyridine	10.38	9.56
3-chlorophenol	9.87	9.30
3-heptanone	7.11	7.54
3-methyl-1-butanol	9.16	9.40
3-methyl-2-butanone	9.35	9.81
3-methylbutanenitrile	9.56	9.93
3-methylpyridine	10.24	10.00
3-pentanone	9.40	9.83
4-methyl-1,3-dioxolan-2-one	11.75	11.58
4-methyl-2-pentanone	7.96	8.55
4-methylpyridine	10.20	10.06
N,N-diethylacetamide	7.86	7.88
N,N-diethylformamide	8.98	8.99
N,N-dimethylacetamide	10.75	10.62
N,N-dimethylaniline	7.87	7.72
N,N-dimethylformamide	12.90	12.72
N,N-dimethylthioformamide	11.74	11.41
N-methyl pyrrolidinone	10.37	9.95
N-methylacetamide	12.99	13.15
N-methylformamide	16.92	16.62
acetic acid	17.39	16.89
acetic anhydride	10.54	10.90
acetone	13.51	14.54
acetonitrile	18.90	19.44
acetophenone	8.52	8.39
acrylonitrile	15.10	15.91
allyl alcohol	14.58	14.88
aminoethanol	16.58	15.42
ammonia	40.00	33.14
aniline	10.93	10.42
anisole	9.15	9.10
benzaldehyde	9.84	9.51

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
benzene	11.12	10.98
benzonitrile	9.70	9.62
benzoyl bromide	8.48	7.98
benzoyl chloride	8.62	8.36
benzyl alcohol	9.63	9.24
benzyl methyl ketone	7.57	7.48
benzylamine	9.16	8.74
bis(2-chloroethyl) ether	8.48	8.37
bromobenzene	9.48	9.03
bromoform	11.39	9.99
butanoic acid	10.82	11.17
butyl acetate	7.55	8.03
butyraldehyde	11.05	11.46
carbon disulfide	16.50	14.00
carbon tetrachloride	10.30	9.87
chlorobenzene	9.79	9.53
chloroform	12.40	11.63
cinnamaldehyde	7.95	7.78
cis-decalin	6.46	6.17
cis-perfluorodecalin	4.21	4.58
cyclohexane	9.20	8.96
cyclohexanol	9.68	9.04
cyclohexanone	9.60	9.38
cyclopentanone	11.24	11.00
di-n-butylamine	5.86	6.26
di-n-butylorthophthalate	3.75	3.94
di-n-propyl ether	7.27	8.00
dibenzyl ether	5.19	5.25
dibromomethane	14.33	12.59
dibutyl ether	5.87	6.42
dibutyl sulfide	5.73	5.99
dibutyl sulfoxide	5.13	5.88

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
dichloroacetic acid	12.12	11.27
dichloromethane	15.50	14.60
diethanolamine	10.42	10.03
diethyl carbonate	8.21	8.77
diethyl ether	9.55	10.64
diethyl malonate	6.56	6.98
diethyl sulfate	7.64	7.87
diethyl sulfide	9.22	9.47
diethyl sulfite	7.84	8.21
diethylamine	9.60	10.28
diethylene glycol dimethyl ether	7.00	7.30
diethylene glycol	10.52	10.11
dihydrolevoglucosenone	9.76	9.33
diisopropyl ether	7.03	8.00
diisopropyl sulfide	6.88	7.29
dimethyl carbonate	11.88	12.37
dimethyl sulfate	10.57	10.31
dimethyl sulfide	13.55	13.39
dimethylcyanamide	12.38	12.92
dimethylphthalate	6.13	6.25
dimethylsulfoxide	14.03	13.57
diphenyl ether	6.29	6.22
diphenyl ketone	6.08	5.96
ethanol	17.03	17.61
ethyl acetate	10.15	10.81
ethyl acetoacetate	7.85	8.15
ethyl benzoate	6.94	7.04
ethyl chloroacetate	10.26	9.32
ethyl formate	12.36	13.09
ethyl phenyl ether	7.87	7.90
ethyl phenyl ketone	7.53	7.43
ethyl propionate	8.66	9.22

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
ethyl trichloroacetate	7.23	7.35
ethylbenzene	8.13	8.23
ethylene carbonate	15.12	14.38
ethylene glycol	17.89	17.30
ethylenediamine	14.79	14.04
fluorobenzene	10.60	10.52
formamide	25.06	24.27
formic acid	26.38	23.46
furan	13.68	13.64
gamma-butyrolactone	13.07	12.70
glycerol	13.66	12.77
heptanoic acid	7.06	7.25
hexafluorobenzene	8.67	8.97
hexamethylphosphoric triamide	5.69	5.80
hexanoic acid	7.95	8.20
hydrazine	31.35	26.74
hydrogen fluoride	47.62	59.76
hydrogen peroxide	42.37	36.12
iodobenzene	8.94	8.42
isopentyl acetate	6.66	7.14
isopropylbenzene	7.14	7.29
meta-cresol	9.53	9.24
meta-dichlorobenzene	8.73	8.37
meta-xylene	8.11	8.24
methanesulfonic acid	15.37	14.71
methanol	24.57	25.13
methyl acetate	12.53	13.13
methyl benzoate	7.96	7.91
methyl formate	16.10	16.75
methyl orthoacetate	7.37	8.24
methyl orthoformate	8.89	9.41
methyl propionate	10.31	10.75

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
methylene iodide	12.41	10.29
methylphenylamine	9.17	8.98
morpholine	11.43	11.00
n-butyl iodide	8.73	8.72
n-butylamine	10.07	10.40
n-butyronitrile	11.38	11.79
n-decane	5.11	5.49
n-dodecane	4.43	4.73
n-heptane	6.78	7.30
n-hexadecane	3.40	3.68
n-hexane	7.60	8.25
n-octane	6.12	6.64
n-pentane	8.61	9.46
n-perfluorohexane	4.97	6.04
n-propyl acetate	8.65	9.23
nitrobenzene	9.74	9.22
nitroethane	13.91	13.76
nitromethane	18.52	17.86
o-chloroaniline	9.47	8.95
ortho-cresol	9.62	9.24
ortho-dichlorobenzene	8.84	8.39
ortho-xylene	8.26	8.21
p-chloroacetophenone	7.71	7.49
p-methoxybenzaldehyde	8.23	8.01
para-cresol	9.44	9.22
para-xylene	8.08	8.27
pentachloroethane	8.27	7.77
pentanoic acid	9.15	9.44
perfluoroheptane	4.51	5.34
perfluoromethylcyclohexane	4.45	5.84
perfluorooctane	4.03	4.80
phenol	11.39	10.79

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
phenylacetonitrile	8.65	8.41
phosphorus oxychloride	10.87	10.65
piperidine	10.06	9.70
propanoic acid	13.33	13.55
propionaldehyde	13.62	14.13
propionitrile	14.10	14.57
pyridine	12.36	11.78
pyrimidine	12.69	12.63
pyrrole	14.39	14.12
pyrrolidine	12.01	11.49
quinoline	8.44	8.00
styrene	8.66	8.64
sulfolane	10.50	9.99
sulfur dioxide	22.83	20.74
sulfuric acid	18.69	16.63
tetrachloroethylene	9.74	9.10
tetraethylurea	5.26	5.48
tetrahydrofuran	12.25	12.07
tetrahydropyran	10.18	10.03
tetrahydrothiophene	11.27	10.45
tetramethylsilane	7.33	8.66
tetramethylurea	8.31	8.32
thiane	9.63	8.97
thiobis(2-ethanol)	9.67	9.31
thionyl chloride	13.71	12.44
toluene	9.35	9.44
trans-1,2-dichloroethylene	12.85	12.41
tri-(n-butyl)amine	4.18	4.47
tri-n-butylphosphate	3.65	3.94
trichloroethylene	11.12	10.62
triethanolamine	7.51	7.34
triethylamine	7.15	7.66

Continued on next page

Molecule Name	Experimental concentration/ M	Calculated concentration/ M
triethylene glycol	7.46	7.33
triethylphosphate	5.87	6.25
trifluoroacetic acid	12.97	14.38
trimethylphosphate	8.67	8.89
water	55.35	55.39

Table I.5 Data for Figure 7.7. Experimental data from [332, 333]

T/K	$[vapour]_{exp}$ / M	$[liquid]_{exp}$ / M	$\frac{1}{2}c_{T,exp}$/ M	$[liquid]_{calc}$ / M	$[vapour]_{calc}$ / M	$\frac{1}{2}c_{T,calc}$/ M
273.15	0.000269	55.543889	27.772079	55.921448	0.000220	27.960834
283.15	0.000522	55.536117	27.768319	55.718894	0.000387	27.859641
293.15	0.000960	55.453517	27.727238	55.498655	0.000654	27.749655
303.15	0.001687	55.311078	27.656382	55.260457	0.001109	27.630783
313.15	0.002837	55.120656	27.561746	55.004054	0.001768	27.502911
323.15	0.004603	54.891367	27.447985	54.729227	0.002742	27.365984
333.15	0.007222	54.621528	27.314375	54.435764	0.004145	27.219954
343.15	0.010994	54.317128	27.164061	54.123450	0.006169	27.064809
353.15	0.016283	53.989850	27.003067	53.792070	0.008949	26.900509
363.15	0.023511	53.630233	26.826872	53.441374	0.012744	26.727059
373.15	0.033189	53.239633	26.636411	53.071069	0.017889	26.544479
383.15	0.045911	52.834572	26.440242	52.680819	0.024692	26.352755
393.15	0.062278	52.396072	26.229175	52.270191	0.033650	26.151920
403.15	0.083167	51.935639	26.009403	51.838671	0.045264	25.941967
413.15	0.109278	51.449856	25.779567	51.385591	0.060242	25.722916
423.15	0.141556	50.940361	25.540958	50.910149	0.079343	25.494746
433.15	0.181278	50.408817	25.295047	50.411299	0.103608	25.257454
443.15	0.223500	49.852433	25.037967	49.887763	0.134208	25.010985
453.15	0.286944	49.273222	24.780083	49.337927	0.172617	24.755272
463.15	0.355667	48.668906	24.512286	48.759768	0.220653	24.490210
473.15	0.437111	48.037667	24.237389	48.150740	0.280564	24.215652
483.15	0.533667	47.378094	23.955881	47.507651	0.355147	23.931399
502.55	0.770556	45.963067	23.366811	46.146876	0.555666	23.351271
533.15	1.324444	43.555900	22.440172	43.598612	1.107450	22.353031
552.55	1.843333	41.705239	21.774286	41.622068	1.701050	21.661559
583.15	3.069444	38.380350	20.724897	37.735193	3.210974	20.473084
602.55	4.273333	35.566933	19.920133	34.838635	4.495739	19.667187
633.15	8.072222	29.332394	18.702308	30.094315	6.596432	18.345374
644.25	11.916667	24.228333	18.072500	28.420030	7.289469	17.854750
647.02	17.216944	17.216944	17.216944	28.009141	7.453804	17.731472

Table I.6 Data for Figure 7.8

T/K	$[vapour]_{exp}$ / M	$[liquid]_{exp}$ / M	$\frac{1}{2}c_{T,exp}$/ M	$[liquid]_{calc}$ / M	$[vapour]_{calc}$ / M	$\frac{1}{2}c_{T,calc}$/ M
353.15	0.0345	8.58	4.31	8.307599	0.024775	4.166187
373.15	0.0588	8.34	4.20	8.062910	0.042826	4.052868
393.15	0.0948	8.08	4.09	7.808949	0.070142	3.939546
413.15	0.1460	7.82	3.98	7.542403	0.110040	3.826221
433.15	0.2160	7.53	3.87	7.259021	0.166765	3.712893
453.15	0.3130	7.22	3.77	6.953340	0.245777	3.599559
473.15	0.4450	6.88	3.66	6.618396	0.354057	3.486227
493.15	0.6250	6.50	3.56	6.245678	0.500093	3.372886
513.15	0.8920	6.03	3.46	5.825639	0.693448	3.259543
533.15	1.3200	5.40	3.36	5.349671	0.942725	3.146198
543.15	1.7100	4.91	3.31	5.089171	1.089883	3.089527
547.15	1.9500	4.63	3.29	4.980790	1.152925	3.066857
550.15	2.2100	4.34	3.27	4.897968	1.201743	3.049856
552.15	2.5100	4.04	3.27	4.842054	1.234984	3.038519
553.80	3.2500	3.25	3.25	4.795480	1.262859	3.029169

Table I.7 Data for Figure 7.9