INVESTIGATION OF CALCULATED HYDROGEN BONDING PARAMETERS

TARAVAT GHAFOURIAN

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

September 1996

SCHOOL OF PHARMACY & CHEMISTRY
LIVERPOOL JOHN MOORES UNIVERSITY

Acknowledgments

I would like to express my most sincere thanks to Professor John Dearden for his help and advice with all aspects of this work.

I am very grateful to Dr. Mark Cronin, for his help and advice throughout the project, especially his help with the computing.

I am grately indebted to Dr. Jeff Morris of Zeneca for his contribution to this investigation.

I would also like to thank Dr. Chris Rostron and Dr. Gaynor Bresnen for their help during this investigation.

Thanks also to my fellow research students, Judy, Alison, Jane, and Thuy for making the laboratory a constructive work place.

I am grateful to the Ministry of Health and Medical Education of Iran for financial support.

I wish to express my appreciation to my parents and sisters who have given me much support during the course of my studies.

Finally, I should like to thank my husband Ali for his tolerance and help.

Abstract

Hydrogen bonding is an important interaction which controls solubility, partitioning and transport of drugs and is an important force in drug-receptor interactions. Therefore H-bonding parameters need to be used in QSAR stuies. Because of the difficulty of the measurement, they have appeared in QSAR equations mostly as indicator variables. In this work, the prime objective has been to devise readily accessible H-bonding parameters by means of theoretical chemistry.

Because of the electrostatic nature of this bond, different electrostatic descriptors have been examined, including dipole moments, atomic charges, electrotopological indices, electrostatic potentials and the similarity of electrostatic potentials. Among these, atomic charge and electrostatic potentials have shown good correlations with experimental H-bonding donor and acceptor abilities ($\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$).

Atomic charge parameters, Q_H and Q_{MN}, are the atomic charge on the most positively charged hydrogen atom in a H-bond donor and the atomic charge on the most negatively charged heteroatom (or average of the atomic charge in an aromatic system in a molecule which does not have a heteroatom capable of H-bonding) in a H-bond acceptor, respectively. These parameters have been calculated by different quantum mechanical and empirical methods. Electrostatic potential derived atomic charges have also been examined.

Electrostatic potential (ESP) parameters, ESP^+ and ESP^-, are the highest and the lowest electrostatic potential on the connolly solvent accessible surface of the molecule, which represent H-bonding donor and acceptor abilities, respectively.

Because the correlations of $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ with electrostatic descriptors showed a family dependent behaviour, the second most important contribution to the H-bonding energy, charge transfer energy, was also incorporated in correlations. E_{LUMO} of H-bond donor (energy of the lowest unoccupied molecular orbital, as a measure of the electron-accepting ability) and E_{HOMO} of H-bond acceptor (energy of the highest occupied molecular orbital, as a measure of the electron-donating ability) were used for this purpose.

The results showed that ESP parameters were superior to atomic charge parameters in prediction of H-bonding abilities. Among atomic charge parameters calculated by different methods, those calculated by MNDO and AM1 semiempirical methods were the best. E_{HOMO} and E_{LUMO} calculated by these two methods were also the best in correlations with H-bonding abilities.

Finally, these parameters were used to replace the experimental H-bonding parameters and indicator variables in a number of QSAR equations and their use was shown to be successful in these equations.

Molecular mechanical interaction energies with a H-bond donor and a H-bond acceptor probe, in the GRID program, could not parametrise H-bonding abilities.

CONTENTS

Chapters 1-5 constitute the introduction to the various aspects of this thesis.

Chapter 6 briefly outlines the aims of the study.

Chapters 7-12 deal with the different approaches used in the investigation; each of these chapters comprise methods, results and discussions, and conclusions.

Chapter 13 contains the general overall conclusions from the work.

List of Figures

List of Tables

1. Quantum mechanics

The application of quantum mechanics to chemistry requires the approximate solution of Schrödinger's wave equation:

$$H\psi = W\psi \qquad (1.1)$$

where the eigenvalue W is the total energy of the system under study and the eigenfunction ψ is the wavefunction. If the wavefunction ψ is normalized to unity,

$$\int \psi^2 \, d\nu = 1 \qquad (1.2)$$

then ψ^2 can be interpreted physically as an electron probability density function. Assuming fixed nuclei, the Hamiltonian operator H is given by:

$$H = \sum_i T(i) + \sum_{i,\mu} V_{ne}(i,\mu) + \sum_{i<j} V_{ee}(i,j) + \sum_{\mu<\nu} V_{nn}(\mu,\nu) \qquad (1.3)$$

where $T(i)$ are the kinetic energy operators for individual electrons i, $V_{ne}(i, \mu)$ are coulombic attractions between electrons i and nuclei μ, $V_{ee}(i, j)$ are the electron-electron repulsion operators, and $V_{nn}(\mu, \nu)$ are the nuclear-nuclear repulsions. The two-electron terms $V_{ee}(i, j)$ make the wave equation 1.1 difficult to solve because they link together the coordinates of all the electrons into a system of multidimensional partial differential equations that cannot, in general, be solved.

Since ψ and W are unknown, the usual practice is to approximate ψ by a function Φ of known form and then calculate an approximate energy E as:

$$E = \int \Phi H \Phi d\nu / \int \Phi^2 d\nu \qquad (1.4)$$

The function Φ can be chosen on the basis of physical interpretability and computational convenience. The variation theorem guarantees that for an approximate function Φ, the calculated approximate energy E is an upper bound to the true energy W, or $E \geq W$. A particularly meaningful and useful form for the choice of Φ is the molecular orbital wavefunction:

$$\Phi_{MO} = \phi_1(1)\phi_1(2)\ldots\phi_n(2n) \qquad (1.5)$$

expressed as a product of molecular orbitals ϕ_i that are functions of the coordinates of individual electrons. The MOs ϕ_i distribute individual electrons throughout the entire nuclear framework of the molecule rather than localising them in particular bonds or on atoms. The Pauli exclusion principle stipulates that no more than two electrons can be assigned to the same ϕ_i.

The approximate total energy E is minimised by a process called the self-consistent-field (SCF) method, which results in the expression:

$$E = \Sigma\varepsilon_1 - V_{ee} + V_{nn} \qquad (1.6)$$

where ε_i are orbital energies associated with individual MOs ϕ_i and the summation is over all electrons. The quantities V_{ee} and V_{nn} are the total potential energies of electron-electron and nuclear-nuclear coulombic repulsions. SCF method uses an iterative procedure. The procedure should converge, that is the changes in the solutions should get smaller as the iteration progresses. The SCF orbitals of many atoms were determined in the period 1930-1950 mainly according to procedures

developed by Hartree and Fock and therefore they are usually known as Hartree-Fock orbitals. In the Hartree SCF method, which is now rarely used, the electron repulsion is that given by a classical electrostatic picture of the electron density in the atom or molecule. In the Hartree-Fock SCF method there is an additional term in the electron repulsion associated with the antisymmetry of the wavefunction (exchange of electrons).

Since it becomes increasingly difficult to obtain accurate solutions of the Schrödinger equation as the number of variables increases, any method which allows us to reduce the number that have to be considered at any time is of considerable value. In the Born-Oppenheimer approximation the motion of the electrons and that of the nuclei are considered separately. The reason this can be done is that the nuclei are much heavier than the electrons and to a good approximation the electrons immediately adjust their positions to follow the nuclei as they move. In other words, the wavefunction of the electrons depends on the position of the nuclei, but not on the momenta of the nuclei.

1.1. LCAO approximation

The standard procedure for obtaining the eigenfunctions of an operator is to expand the eigenfunction in a set of known functions. The most widely used method is an expansion of a molecular orbital as a linear combination of atomic orbitals. This is commonly referred to as the LCAO approximation. Suppose we could approximate

the true Hamiltonian operator H by a sum of effective one-electron operators $H_{eff}(i)$:

$$H \sim H' = \sum_i H_{eff}(i) \qquad (1.7)$$

the sum being over all electrons i. An obvious way to form H' is by neglecting entirely the electron-electron repulsion terms $V_{ee}(i, j)$ in H in equation 1.3 and then grouping the remaining one-electron kinetic energy operators $T(i)$ and the nuclear-nuclear repulsion and electron-nuclear attraction terms as $H_{eff}(i)$. The calculations with MO theory in which electron-electron repulsion is not specifically included in the Hamiltonian, are referred to as independent-electron models. The model is moderately satisfactory for transition metal complexes and organic hydrocarbons. Most independent-electron calculations are empirical rather than *ab initio*, that is experimental data are used to determine the Hamiltonian integrals (Murrell et al, 1985). Instead one might choose $H_{eff}(i)$ so as to account for the effect of interelectronic repulsions in some average way. The MO wavefunction Φ_{MO} is an eigenfunction of the approximate Hamiltonian H':

$$H'\Phi_{MO} = E'\Phi_{MO} \qquad (1.8)$$

because each term $H_{eff}(i)$ contains only the coordinates of electron i and therefore $H_{eff}(i)$ operates only on an individual MO ϕ_i. This permits the many-electron wave equation 1.8 to be separated into identical, independent, one-electron eigenvalue equations:

$$H_{eff}(i)\phi_1(i) = \varepsilon_1\phi_1(i) \qquad (1.9)$$

where ε_l is the energy of an electron in orbital ϕ_l. Then:

$$E' = \Sigma\varepsilon_i \qquad (1.10)$$

where the sum is over all electrons. If we approximate ϕ_i by a linear combination of atomic orbitals χ_r from the various atoms that constitute the molecule:

$$\phi_i = \sum_r c_{ri}\chi_r \qquad (1.11)$$

The collection of atomic orbitals χ_r is called the basis set, from which the same number of MOs (ϕ_i) can be constructed. For convenience and without any loss of generality we require the atomic orbitals to be normalised:

$$\int \chi_r^2 \, dv = 1 \qquad (1.12)$$

but not necessarily orthogonal, so that in general:

$$\int \chi_r \chi_s dv = S_{rs} \neq 0 \qquad (1.13)$$

where S_{rs} is the overlap between the two atomic orbitals χ_r and χ_s. Using the LCAO expansion of equation 1.11 to calculate an approximate orbital energy ε_i gives:

$$\varepsilon_i = \frac{\int \phi_i H_{eff} \phi_i dv}{\int \phi_i^2 dv} = \frac{\sum_{r,s} c_{ri}c_{si} \int \chi_r H_{eff} \chi_s dv}{\sum_{r,s} c_{ri}c_{si} \int \chi_r \chi_s dv} = \frac{\sum_{r,s} c_{ri}c_{si} H_{rs}}{\sum_{r,s} c_{ri}c_{si} S_{rs}} \qquad (1.14)$$

where S_{rs} is the overlap integral above and:

$$H_{rs} = \int \chi_r \, H_{eff} \, \chi_s \, dv \qquad (1.15)$$

Applying the variation theorem to ε_i, each coefficient c_{ri} can be varied to minimize

the orbital energy ε_l, which leads to a set of n homogeneous, linear equations,

$$\sum_{r}^{n} c_{rl}(H_{rs} - \varepsilon_l S_{rs}) = 0, \qquad s = 1,2,\ldots,n. \qquad (1.16)$$

These equations have nontrivial solutions for the coefficient c_{rl} only if the $n \times n$ determinant formed from the quantities in the parentheses in 1.16 is zero:

$$|H_{rs} - \varepsilon S_{rs}| = 0 \qquad (1.17)$$

Equation 1.17 is the well-known secular equation or secular determinant. If the quantities H_{rs} and S_{rs} were known, the secular determinant could be solved for n values of ε_l, which could then be substituted back into the homogeneous linear equations 1.16 to obtain n sets of coefficients (c_{rl}), a different set for each ϕ_l and corresponding ε_l. The total energy E' is the sum of orbital energies $\Sigma\varepsilon_l$, where the summation is over all electrons with no more than two electrons in the same MO. The integrals that occur in equation 1.17 are of two types, those that occur on the diagonal of the determinant, H_{rr}, and those off the diagonal, $H_{rs}(r\neq s)$. The diagonal terms are generally referred to as the Coulomb integrals; the off-diagonal integrals are called the resonance integrals.

The integral H_{rr} can be interpreted as being related to the energy of an electron in an atomic orbital χ_r and, therefore, H_{rr} can be approximated by the experimental ionisation potential (IP) of an electron in the atomic orbital χ_r on the atom involved. Assuming that χ_r and χ_s are on different atoms, as they usually are, we can argue that H_{rs} $(r\neq s)$ should be some kind of interaction energy between atoms r and s. If the distance between the two atoms is large, then the interaction H_{rs} is small. At distances

approximating those of chemical bonds, H_{rr} might be related to the energy of a bond between the two atoms. A very simple choice is to make H_{rs} directly proportional to the overlap S_{rs}:$H_{rs} = KS_{rs}$, where K is a number to be chosen to make calculated results agree with those from experiment. Since S_{rs} is normally positive and H_{rs} is a bond energy, lower or more stable than the energy of the separated atoms, K must be negative. Other prescriptions in which H_{rs} is formed by averaging H_{rr} and H_{ss} have also been used.

Slater type atomic orbitals may be used as χ_r to evaluate the overlap integrals S_{rs}. For Slater-type-orbitals (STOs) the two-electron integrals are generally difficult to calculate, particularly if the four basis functions are on three or four different atomic centres. In general an STO basis is commonly used only for calculations on diatomic molecules as for these all integrals are one-centre or two-centre and the necessary integration is relatively simple. For polyatomic molecules it has been found that a much more convenient basis is one which uses the Gaussian-type orbitals (GTOs). Although a GTO is a poor representation of an atomic orbital a relatively small number of such functions in linear combination is quite good.

There are broadly speaking three levels of basis that are commonly used in SCF calculations, particularly for *ab initio* methods. The simplest (computationally cheapest but poorest by the variational criterion) is called a minimal basis; it consists of one variational basis function (STO or GTO contraction) for each type of atomic orbital that is occupied by electrons in the ground state of the atom. Many calculations in the

literature have employed a minimal basis but the predictive capability of such calculations, in respect of both energy and equilibrium geometry, is not high. STO-3G is the most commonly used minimal basis set for *ab initio* methods. An STO-3G basis set has only as many orbitals as necessary to accommodate the electrons of the neutral atom (Hehre et al, 1986).

The second common level of calculation uses two variationally independent basis function for each occupied atomic orbital of the atom, that is, twice as many functions as the minimal basis. This is called the double-zeta basis. The SCF geometries obtained with such a basis are usually in good agreement with experimental geometries and the relative energies of isomeric structures are usually in the correct order. Split valence basis sets have only the valence orbitals split in this manner in contrast to the double zeta basis sets which also have two exponents for core electron orbitals. Split valence basis sets are a considerable improvement over minimum basis set and use of a 3-21G basis set is a reasonable compromise for large molecular systems generally yielding good results.

The third level of calculation adds functions of higher angular momentum than are needed to represent the SCF ground states of the individual atoms; *p* functions for hydrogen, *d* functions for *B...F*, for example. These extra functions are called polarisation functions because they are used to represent the polarisation (distortion of the electron cloud) of the individual atoms in the presence of an external applied electric field (e.g. the distortion of a hydrogen atom can be represented by mixing

the *1s* orbital with a *2p* orbital). SCF calculations with a double-zeta basis plus polarisation functions are generally a good approximation to the Hartree-Fock limit.

1.2. Hartree-Fock SCF method

SCF orbitals are derived from a one-electron Hamiltonian which contains the average effect of electron repulsion and, as the name implies, this repulsion is consistent with that calculated from the orbitals which are produced. In the Hartree SCF method, which is now rarely used, the electron repulsion is that given by a classical electrostatic picture of the electron density in the atom or molecule. In the Hartree-Fock SCF method there is an additional term in the electron repulsion associated with the antisymmetry of the wavefunction (exchange of electrons). The derivation of the Hartree-Fock orbitals is mathematically rather complex and it gives rise to the following operator for a determinant wavefunction with spin orbitals $\psi_a \ldots \psi_n$ occupied:

$$F = H^c + \sum_{s=a}^{n} (J^{(s)} - K^{(s)}) \qquad (1.18)$$

where H^c is the core integral in which a two-electron integral is separated into a product of two one-electron integrals; $J^{(s)}$ and $K^{(s)}$ are matrix elements of (two-electron) coulomb and exchange operators. As F matrix requires a knowledge of the spin-orbital wavefunctions before it can be constructed, yet these wavefunctions are the eigenfunctions of F, it is clear that the orbitals can be determined only by an iteration to self-consistency. To obtain the eigenfunctions of the F operator, the F matrix is diagonalised at each cycle of the iterative procedure until self-consistency

is reached. The matrix elements of this operator take different forms depending on the spins associated with these spin-orbitals. Expansion of the F matrix elements for a closed-shell system (RHF operator) in which all occupied orbitals have two electrons of opposite spin leads to:

$$F_{\mu\nu} = H^{c}_{\mu\nu} + \sum_{r=a}^{n} \sum_{\rho} \sum_{\sigma} C_{r\rho} C_{r\sigma} \{2\langle \mu\rho | \nu\sigma \rangle - \langle \mu\rho | \sigma\nu \rangle\} \qquad (1.19)$$

where the following abbreviation has been used for the two-electron integrals:

$$\langle \mu\rho | \nu\sigma \rangle = \int\int \phi_{\mu}(1)\phi_{\rho}(2) r_{12}^{-1} \phi_{\nu}(1)\phi_{\sigma}(2) d\nu_1 d\nu_2 \qquad (1.20)$$

In this equation ϕ_{μ}, ϕ_{ρ}, ϕ_{ν} and ϕ_{σ} are atomic orbitals, and r_{12} is the distance between the two electrons. Equation 1.20 is generally referred to as the Roothaan equation although it was derived independently by Lennard-Jones and Hall (Murrell et al, 1985).

In all SCF calculations, the major computational problem is the evaluation of the electron repulsion integrals and their repetitive use to build up the F matrix elements (Eq. 1.19) in the iterative process. Two computationally different types of methods have emerged, *ab initio* and semiempirical.

1. Ab initio methods

In *ab initio* methods, all one- and two-electron integrals are retained and are calculated. Improvements in these methods include more efficient algorithms for solution of integrals, use of more complete forms of the Hamiltonian function, for

example including electron correlation terms and use of more accurate basis sets. In most *ab initio* methods, all electrons are explicitly included. In attempts to extend these methods to larger systems, modifications in which inner or core electrons are replaced by an "effective core potential" have been developed. *Ab initio* methods have the advantage of not requiring any parametrisation and therefore can be used for all types of systems. It is also much easier to identify failings of these methods and improve them (Loew & Burt, 1990). The advantage that *ab initio* methods have over semiempirical methods of calculation for predictive purposes is not that they may in any stage of development be superior in their predictions but that they have the potential to be superior. The variation theorem provides a rigorous route by which *ab initio* calculations can be improved and predictions can be tested to see whether they are unchanged on making an improved calculation. In many semiempirical models there is no such route for improvement and the reliability of a prediction can be judged only by the number of times the model has proved correct and the number of times it has proved incorrect. The reason for this is that the variation theorem establishes lower limits to the exact eigenvalues of a Hamiltonian. It does not establish limits relating the exact or approximate eigenvalues of a model Hamiltonian to the exact eigenvalues of the exact Hamiltonian. Thus semiempirical calculations can give energies which are higher or lower than the exact energies. Improving a model may move the calculated energies up or down; improving an *ab initio* calculation will always move the energies down and closer to the exact energy (Murrell et al, 1985).

2. Semiempirical methods

It is possible to approximate H_{rr} semiempirically, using atomic spectral data and various assumptions, and avoiding the explicit specification of $H_{eff}(i)$. Many approximate theories have been developed, some of which have used numerical approximations to the electron repulsion integrals, particularly approximations in which many were taken as zero. Others have combined these approximations with empirical values for the integrals so as to improve the predictive capability of the theories.

Hoffmann (1963) introduced the extended Hückel (EH) model. In Hückel π-electron theory only π orbitals are considered and σ orbital are completely ignored. However, the essential features of the Hoffmann treatment are that the atomic orbital basis consists of all the valence atomic orbitals of the component atoms. In this method each H_{rr} is taken as the ionisation potential for the appropriate electronic state in the appropriate isolated atom. The S_{rs} are computed according to equation 1.13 in which the χ_r are chosen to be Slater wavefunctions. The H_{rs}, normally called the exchange or resonance integrals, are approximated by the expression:

$$H_{rs} = \frac{1}{2}[k(H_{rr} + H_{ss})S_{rs}] \qquad (1.21)$$

where the k is a calculated constant which reproduces some experimental value (e.g. the rotational barrier height in ethane). One must solve equation 1.9 for each trial conformation of the molecule tested. Stable conformations of the molecule are those for which equation 1.10 yields a relative minimum in total energy. This technique

minimises the bonded and nonbonded energies simultaneously (Hopfinger, 1973). In Hückel π-electron theory it is assumed that the overlap integrals between orbitals on different atoms are zero. It is natural therefore that the extension of Hückel theory to an SCF model should retain the zero-overlap assumption and the two-electron integrals $\langle\mu\rho|\nu\sigma\rangle$, in which either ϕ_μ and ϕ_ν are on different atoms and/or ϕ_ρ and ϕ_σ are on different atoms, should be taken to zero. The only non-zero two-electron integrals retained in such calculations are of the type $\langle\mu\rho|\mu\rho\rangle$ which represents the repulsion of the electron densities $\phi_\mu^2(1)$ and ϕ_ρ^2. This development, usually referred to as the zero-differential-overlap (ZDO) approximation, was introduced into π-electron theory by Pariser and Parr and by Pople in 1953 (PPP method) (Pariser & Parr, 1953; Pople, 1953), and into an all-electron theory by Pople, Santry and Segal in 1965. The ZDO SCF theories use only valence orbitals as their bases. Many approximate theories based on these models have subsequently been developed (Murrell et al, 1985).

The Pople SCF MO method (Pople & Beveridge, 1970) is a valence-orbital model in which the valence electrons are assumed to move in a core composed of the nuclei and inner-shell electrons, the MOs being written as linear combinations of all the valence-shell atomic orbitals of the contributing atoms, and in which overlap and three- and four-centre integrals are neglected. The remaining integrals can be treated as parameters, subject to the restrictions that the total number of parameters must be kept within bounds, and that their values must be physically reasonable. The results of calculations by this method are sensitive to the choice of coordinate axes. Two

solutions to this problem are:

1- In the CNDO (Complete Neglect of Differential Overlap) approximation, all orbitals are assumed to be spherically symmetrical in calculating electron repulsion integrals, while the one-electron resonance integrals $\beta_{\mu\nu}^{c}$ are given by:

$$\beta_{\mu\nu}^{c} = \beta_0 S_{\mu\nu} \qquad (1.22)$$

where β_0 is a constant for the type of bond in question. The effect of orbital shape thus appears only through these resonance integrals, the value of the overlap integral $S_{\mu\nu}$ between a *p* atomic orbital ϕ_μ and some other atomic orbital ϕ_ν depending on the orientation of ϕ_μ relative to ϕ_ν. The basis set functions (i.e. the valence-shell atomic orbitals of the contributing atoms) are all mutually orthogonal, for the orbitals of a given atom form an orthogonal set, and we neglect all overlap between atomic orbitals of different atoms.

2- The NDDO (Neglect of Diatomic Differential Overlap) approximation solves the problem in a more complicated way, by retaining all three- and four-orbital integrals in which the overlap is between atomic orbitals of the same atom. Integrals involving overlap between atomic orbitals of different atoms are still set equal to zero. The NDDO model is much more complicated as it involves a large number of three- and four-orbital integrals. In PNDO (Partial Neglect of Differential Overlap) approximation, the number of integrals of the NDDO method has been minimised. In the NDDO or PNDO approximations, the *F* matrix is given by the full Roothaan expression but the three- and four-orbital integrals vanish if they involve overlap

between atomic orbitals of different atoms.

INDO (Intermediate Neglect of Differential Overlap) is a further approximation introduced by Pople et al (1967). This differs from CNDO only by the inclusion of one-centre electron repulsion integrals involving differential overlap. Most such integrals vanish through symmetry, the only ones to be considered being of the type $\langle\mu\rho|\mu\rho\rangle$ when ϕ_μ and ϕ_ρ are two different atomic orbitals of a given atom. The object of this method initially was to estimate geometries and dipole moments of molecules, not heats of formation; an open-shell version of INDO approximation was also used to calculate spin coupling constants in the esr spectra of radicals. The simpler version of INDO approach called MINDO (Modified INDO) has been adopted to calculate heats of formation. In this treatment, the various integrals are estimated in a manner similar to that used in the π approximation, the parameters being chosen to fit the observed heats of formation of selected molecules (Dewar, 1969).

In the CNDO and INDO approximations, the repulsion integrals between any atomic orbital of atom A and any atomic orbital of atom B are set equal, regardless of whether these atomic orbitals are of *s*, $p\sigma$, or $p\pi$ type. This simplification is essential if the results of calculation are to be invariant for rotation of the coordinate axes. The integrals are not in fact equal and in NDDO they are not assumed to be equal. Accordingly, some procedures for estimating the NDDO repulsion integrals have been introduced; MNDO (Modified Neglect of Diatomic Overlap) (Dewar & Thiel, 1977), AM1 (Austin Model 1) (Dewar et al, 1985) and MNDO-PM3 (Modified Neglect of

Diatomic Overlap, Parametric Method number 3) (Stewart, 1989a, b; 1991) are all MNDO-type approximations. These three methods together with MINDO/3 are available in the Mopac program (Stewart, 1990). All four semiempirical methods contain sets of parameters. For MINDO/3 atomic and diatomic parameters exist, while MNDO, AM1 and PM3 methods use only single-atom parameters. Not all of the parameters are optimised for all methods; the values of some parameters are obtained from experiment (not optimised) and some of the parameters are not used in certain methods. The methods all use a minimum basis set consisting of a maximum of one atomic orbital for each angular quantum number. The normal basis set for any atom consists of one *s* and three *p* orbitals. Three- and four-centre integrals and also all overlap integrals arising from the overlap of two different atomic orbitals are neglected.

One-centre two-electron integrals in the MNDO and AM1 methods are derived mostly from experimental data on isolated atoms and only a few have been obtained by optimisation to fit molecular properties. In the PM3 method, the values of one-centre two-electron integrals were optimised to reproduce experimental molecular properties.

Unlike the NDDO method, in which two-centre two-electron integrals are evaluated from analytical formulas, in the MNDO-type methods these integrals are calculated using a semiempirical model which takes the correlation effects into account. The model is based on the concept that the two-centre repulsion integrals $\langle\mu\nu|\lambda\sigma\rangle$ represent the energy of interaction between the charge distributions $e\phi_\mu\phi_\nu$ at atom A

and $e\phi_\lambda\phi_\sigma$ at atom B (elementary charge e) which classically are equal to the sum over all interactions between the multipole moments M_{lm} of the two charge distributions, the subscripts l and m specifying the order and orientation of the multipoles. MINDO/3 coulomb and exchange integrals , in marked contrast to the other MNDO-type methods, are simple and the integral is a function of the atom types and the interatomic distance only.

The one-centre one-electron integral represents the energy an electron in an atomic orbital would have if all electrons were removed from the system. This is approximated by adding on to the one-electron energy of the atomic orbital in the fully ionised atom the potential due to all the other nuclei in the system. The one-electron energy is obtained parametrically.

The two-centre one-electron integral (resonance integral) $H_{\mu\nu}$ is approximated using the overlap integral $S_{\mu\nu}$. This violates the NDO approximation, but since resonance integrals are large, this integral is retained. This is the origin of "Modified" in the MNDO and MINDO/3 names. Within MNDO, AM1, and PM3, $H_{\mu\nu}$ is approximated by:

$$H_{\mu\nu} = S_{\mu\nu}\tfrac{1}{2}(\beta_\mu + \beta_\nu) \qquad (1.23)$$

while MINDO/3 has a very different form:

$$H_{\mu\nu} = S_{\mu\nu}\beta_{AB}(I_\mu + I_\nu) \qquad (1.24)$$

This use of a diatomic parameter is the most distinctive difference between the

MNDO/AM1/PM3 philosophies and that of MINDO/3.

The core-core repulsion integrals can be calculated from simple electrostatics. In MINDO/3 method, in order to take into account the decreasing screening of the nucleus by the electrons as the interatomic distance becomes very small, an additional term is added to the basic core-core repulsion. The MNDO approximation to the screening effect is similar to that of MINDO/3 in practice, but has a different functional form. In both methods O-H and N-H interactions are treated differently. AM1 and PM3 modifications to the core-core term are the same as that for MNDO with addition of an extra term to reduce the excessive core-core repulsions just outside bonding distances. The additional term may be considered as a van der Waals attraction term. The extra terms define spherical Gaussian functions; PM3 has two Gaussians per atom, while AM1 has between two and four.

MNDO is the oldest of the three (MNDO, AM1, PM3) methods. PM3 is a re-parametrisation of the MNDO method in which the AM1 form of the core-core interaction is used. In parametrising MNDO and AM1, only a very few molecules could be used. PM3 was parametrised using a radically different optimisation procedure.

1.3. Application of quantum mechanical methods

In the ten years during which quantum mechanics was first applied to chemistry,

many fundamental problems of long standing were solved. The covalent bond, the periodic system of the elements, the mechanism of biomolecular reactions, the existence of free radicals, van der Waals forces, the magnetic properties of matter and the conduction of electricity by metals are only a few of the phenomena that were finally explained in atomic terms using quantum mechanics (Kauzmann, 1957).

Ab initio and semiempirical methods can be used in the calculation of energy conformation profiles, calculation of explicit electronic properties of individual compounds, characterisation of model chemical/biochemical reactions and model intermolecular complex formation. The molecular descriptors which can be calculated by these methods are electronic properties such as net atomic charges, dipole and higher moments, ionisation potentials, electron affinities, molecular electrostatic potentials, chemical reactivity properties (electrophilicities and nucleophilicities) and also conformational energies (Loew & Burt, 1990). Among the interactions which are difficult to characterise theoretically by any means other than those involving quantum mechanics are calculation of bond lengths and angles (and therefore determination of the molecular structure of simple molecules), calculation of force constants, calculation of hydrogen bonding functions and characterisation of torsional potential functions (Hopfinger, 1973).

Quantum mechanical methods can be useful in characterising drug-receptor complexes by calculating complex geometry, and stability. They can also be employed in characterising chemical/biochemistry reactions by calculating reaction mechanisms

(identification of reactants, intermediates, transition states and products) and also enthalpies and entropies of activation and reaction (Loew & Burt, 1990).

2. Molecular mechanics

Molecular mechanics is a method of calculating the potential energy of an isolated molecule or system of interacting molecules as a function of their nuclear coordinates. The molecular mechanical model considers a molecule to be a collection of atoms held together by classical forces. The atoms are treated as classical particles under the influence of the molecular mechanical potential or force field. The force field is a set of simple analytically differentiable functions of the nuclear coordinates that yields a potential energy for the molecule with respect to a hypothetical strain-free state. The strain-free state is one in which all bond lengths, angles and torsions are at their 'natural' or minimum energy values and nonbonded atoms are at infinite separation.

The molecular mechanical potential allows us to calculate the relative energy of different conformations of a molecule with little computational effort. Since the terms in the potential function are analytically differentiable, the gradients of potential energy with respect to coordinates, which constitute the forces on the atoms, are easily obtained. This allows one to use standard numerical optimisation methods to minimise the energy of the system, resulting in the location of a local minimum or in a few cases the global minimum energy structure. The term 'molecular mechanics' is generally synonymous with such energy minimisation using an analytical potential. In molecular mechanics, the electron distribution is implicit in the force field in which the nuclei are allowed to move, while quantum mechanics is concerned with the explicit calculation of the electron distribution in a fixed nuclear field. A consequence

of this difference is that bond-making and -breaking cannot be simulated by molecular mechanics unless suitable analytic representations of the event can be developed. The information that must be supplied in the quantum mechanics method is simply the nuclear and net charge, quantum mechanical state multiplicity and appropriate basis functions. Molecular mechanics, on the other hand, requires that all atoms be classified into distinct types that are recognised as different by the force field. In this method, the bonding topology must be specified, equilibrium values and force constants must be supplied for all valence terms, and all atom types must have nonbonded interaction parameters specified. In addition, electrostatic information in the form of atomic partial charges or bond dipoles is generally required (Seibel & Kollman, 1990). Molecular mechanics calculations in general, particularly when there are very few different atoms in a molecule or polymer, give very sensible indications of preferred conformations. These calculations are inexpensive to perform, but they are not complete alternatives to quantum mechanical calculations except within a narrow range of requirememns. The molecular mechanics method demands only a small fraction of the computing time required for a quantum mechanical calculation and for many applications to questions of conformation of large molecules, molecular mechanics has become the method of choice (Richards, 1983).

The molecular mechanical force field is an analytic description of the potential energy surface of a molecule. In reality every molecule has a unique force field but to a very good approximation the force field can be broken down into components that are transferable between molecules. The force field is parametrised against experimental

data for a given class of molecule, and is subsequently used to predict the structural and energetic properties of related molecules.

2.1. Potential functions and parametrisation of force fields

There are three general categories of data that may be used in force field parametrisation: structure, energy and vibrational frequencies. Force fields were first developed in the area of vibrational analysis in order to analyze and predict vibrational spectra. The early molecular mechanical force fields focused on structure and energy, employing modifications of the vibrational force fields, but did not effectively reproduce spectra. The consistent force field (CFF) method simultaneously reproduces structure, energy and vibrational frequencies from the same set of equations (Lifson & Warshel, 1968). This method uses all relevant and available experimental data of whole families of molecules in order to select the best potential energy functions and to determine their constant parameters by a least-squares procedure, such as to obtain a best fit to the experimental data (Lifson et al, 1979).

The developers of a force field usually have a particular class of problems that they are interested in treating well. The potential functions will be selected with the goals of the force field in mind, and parametrising and testing will be performed against a set of compounds representative of those the developers are interested in. This is a natural consequence of the difficulty of creating a truly general force field. Force fields in general use today tend to be focused towards either small molecules or

macromolecules. A small molecule force field will generally use a more elaborate potential, since computation time is not as critical. Macromolecule force fields tend to use simpler potentials that can be evaluated quickly, and place more emphasis on electrostatic interactions and hydrogen bonding. Typical force fields contain terms of potential energy due to bond stretch, angle bending, torsions, van der Waals interactions and Coulombic interaction:

$$E_{total} = E_{stretching} + E_{bending} + E_{torsion} + E_{vanderWaals} + E_{Coulomb} \quad (2.1)$$

In addition to the terms in equation 2.1, a term is usually included to account for out-of-plane distortion of sp^2 centres. A hydrogen bond function may also be used, and valence cross terms that take into account, for example, the change in bond force constants as angles are deformed may be employed. Each of the energy terms has preferential equilibrium positions (bond lengths, bond angles, dihedral angles, Van der Waals interaction distances, etc.) and force constants that are either experimentally known or theoretically estimated and used to associate energetic penalties with each individual deviation (Cohen et al, 1990). Atoms in a molecular mechanical force field are classified into distinct types and their combinations dictate the bond, angle and dihedral types that must be parametrised. The accuracy of the calculation will depend both upon the design of the force field and upon how well it is parametrised (Bays, 1992).

The parametrisation of the force field relies heavily on experimental data collected over the years, and judicious choice must be made in sorting out good and bad data

because the reliability of the method can be no better than the data used for parametrisation. Force fields have been parametrised to give excellent geometries, relative conformational energies, heats of formation, crystal packing arrangements, and even transition state structures and reactivities (Boyd & Lipkowitz, 1982). Various molecular mechanics programmes with their own distinct blend of potential functions have emerged.

MM2 (Allinger, 1977) is a hydrocarbon force field which has been improved over the MM1 force field (Allinger, 1976; Allinger & Chung, 1976) by incorporating two key parameters of the V1 and V2 in the torsional potentials. These terms cancel out exactly in the torsion of a symmetrical molecule such as ethane. However, in a less symmetrical molecule such as butane, they do not necessarily cancel out. MM2 potential, in addition to the energy components in equation 2.1, includes a stretch-bend term to account for the fact that as a bond angle is compressed, the force constants of the associated bonds decrease. This method was originally developed for hydrocarbons but has since been parametrised for a variety of organic functionalities.

CHARMM (Brooks et al, 1983) is a highly flexible computer program which uses empirical energy functions to model macromolecular systems. The fundamental unit used in CHARMM is the atom, but for large systems some or all hydrogens are combined with neighbouring heavy atoms to which they are bound. This combining of atoms is referred to as the 'extended atom representation'. Empirical energy terms, in addition to those used in equation 2.1, are improper torsion energy (which is

designed both to maintain chirality about a tetrahedral extended atom and to maintain planarity about certain planar atoms with a quadratic distortion potential) and hydrogen bonding energy. There is also the option of including different types of constraints (like dihedral constraint to maintain a certain local conformation) in the energy when manipulating the structure through minimisation or dynamics.

Weiner et al (1984) have developed a force field (AMBER) for simulation of nucleic acids and proteins. They have obtained equilibrium bond lengths and angles from microwave, neutron diffraction, and prior molecular mechanical calculations, torsion constants from microwave, NMR, and molecular mechanics studies, nonbonded parameters from crystal packing calculations, and atomic charges from the fit of a partial charge model to electrostatic potentials calculated by *ab initio* quantum mechanical theory. The energy terms used in this force field are the same as in equation 2.1 with the addition of a weak hydrogen bond term between hydrogen-bonding hydrogens and H-bond acceptor atoms.

MM2 is the current standard for small-molecule work, but is a poor choice for macromolecules. AMBER and CHARMM are similar and are the standard for macromolecules, but give only qualitative results on small molecules. The AMBER all-atom force field is used for calculations involving small molecule-macromolecule interactions. Most of the major software systems provide facilities for automatically assigning the appropriate atom types and parameters, but there is considerable variation in the quality and quantity of the parameters available. It is always prudent

to calibrate unfamiliar software with some well-known test cases (Cohen et al, 1990).

COSMIC molecular mechanics potentials have been developed with the aim of handling a wide variety of chemical entities. To this end, the force fields have been simplified as far as possible in order to allow users to modify and extend them to suit their problems and atom types have been kept to a minimum. Because of this generality, some structural features may not be well represented without specific changes to the force field parameters (Vinter et al, 1987). Introduction of some modification to the COSMIC force field (Morley et al, 1991) has greatly increased both its versatility and the accuracy of calculated conformational energies.

Assuming that all the necessary parameters are available for a given molecule, relative total strain energies can be calculated for estimating rotation or inversion barriers, preferred conformations, the energy required to achieve a specific conformation and so on.

2.2. Minimisation methods

It may be sufficient to calculate the potential energy of a single fixed conformation or map the energy as a function of one or more degrees of freedom in the molecule. Usually we wish to determine a low energy structure for the molecule under consideration. This involves finding a minimum on the potential energy hypersurface. Molecular mechanics energy minimisation involves successive iterative computations,

where an initial conformation is submitted to full geometry optimisation. All parameters defining the geometry of the system are modified by small increments until the overall structural energy reaches a local minimum. The goal is to reach a local minimum within the minimum amount of time. The more sophisticated methods use the first and occasionally the second derivatives of the energy function for guiding the minimisation (Cohen et al, 1990). The simplest first derivative method is the steepest descent minimiser, which moves atoms directly down the energy gradient. It is robust and has the advantage of quickly improving bad starting geometries, but suffers from poor convergence near the minimum. A better first derivative method is the conjugate gradient minimiser. It stores information on the direction of previous moves in order to predict better movement directions. This results in somewhat better convergence properties than steepest descent, although it is less tolerant of poor starting geometry. Many programmes use both minimisers, starting with steepest descent then switching to conjugate gradient when a suitable average gradient is reached. Second derivative methods are much more efficient in locating a minimum, although they have the least tolerance of poor starting geometry. The full Newton-Raphson method requires the storage and inversion of a $3N \times 3N$ second derivative matrix, the 3N dimensions corresponding to the 3N degrees of freedom of the molecule. Because of the large storage requirement, the full Newton-Raphson method is not widely used in molecular mechanics. A modification known as the block-diagonal Newton-Raphson method is used in MM2, where only a 3×3 portion of the second derivative matrix is stored for each atom. Some information about the curvature of the energy hypersurface is lost, but the savings in storage and matrix

inversion time are significant (Seibel & Kollman, 1990).

Energy minimisation can proceed either in internal coordinates (the variables explicitly considered are the bond lengths, bond angles, and dihedral angles) or, as is more often the case, in cartesian coordinates. An advantage of minimising in internal coordinates is that cooperative movements of several atoms or groups are well simulated in such treatments; moreover, since the degrees of freedom of the chemical structures are natural, the risk that the molecules are trapped in a false minimum is greatly reduced.

2.3. Conformational searching

Conformational analysis is a method of computational chemistry that allows a calculated relative energy to be associated with each conformation of a molecule. Conformational energy can be calculated by molecular mechanics (or alternatively with quantum mechanics). Those conformations of a molecule that are low in energy are the most likely to be adopted. There are two ways of searching for thermodynamically stable molecular states, that is, of carrying out a conformational analysis. One is to systematically vary each of the degrees of freedom (bond lengths, bond angles, torsional rotations) and to calculate the corresponding conformational energy. This scanning approach allows an investigator approximately to locate all stable states for a few degrees of structural freedom. The second means of seeking stable structures is to minimise energy as a function of degrees of freedom. An energy minimisation can be carried out for a large number of degrees of structural freedom

(Hopfinger, 1985). However, no minimisation method can guarantee finding the absolute lowest energy structure - the global minimum. Energy minimisation will stop at the first local minimisation encountered, without reaching other much deeper, more stable minima. For molecules with a very small number of rotatable bonds an exhaustive search of conformational space is possible, but naive approaches quickly become intractable as the number of rotatable bonds increases. Improvements to exhaustive search methods centre on not searching areas that are strictly inaccessible. Other methods for exploring conformational space include Monte Carlo methods (Paine & Scheraga, 1985) and distance geometry approaches (Weiner et al, 1983). Once various starting conformations have been generated, they can be subjected to molecular dynamics minimisation to move them to the nearest local energy minimum. Molecular dynamics (McCammon, 1990) has been used for conformational searching, but it is not very computer-time efficient in moving far from the initial geometry. It is useful for searching for local minima in a limited area of conformational space (Seibel & Kollman, 1990).

2.4. Intermolecular interactions

As a result of ever increasing interest in, and application of, intermolecular energy functions there have been numerous studies in which such functions were developed. The functions have been derived primarily by fitting directly to experimental thermodynamic and structural data on pure organic liquids, liquid water, and aqueous solutions of organic molecules and ions representative of peptide constituents

(Jorgensen, 1986). It has become apparent that the precise structural information available for the crystal state provides a rich source of information for deriving energy parameters (Lifson et al, 1979).

The general features of intermolecular potentials are very well known. At short distances the potential is repulsive and decreases roughly as an exponential in the separation. The repulsion arises from the Coulombic nuclear-nuclear interaction and the electron-electron Coulombic overlap interaction summing up to a greater repulsive value than the attractive interaction between the nuclear core and electrons on the two different atoms. In practice the interactions can be represented by nonbonded empirical potential functions. The van der Waals, electrostatic and hydrogen bonding potentials are normally considered (Goodford, 1989).

The Lennard-Jones potential function,

$$E_{ij} = A/d^{12} - B/d^{6} \qquad (2.2)$$

may be used to describe the interaction between two non-bonded atoms *i* and *j* which are separated by a distance *d*. *A* and *B* are parameters which determine the size of the attraction $(-B/d^6)$ and repulsion (A/d^{12}) between the atoms, and the most negative value of the interaction energy occurs when $d^6 = 2A/B$.This particular value of the distance d is the optimal separation between the atoms according to the Lennard-Jones function, and may be treated as the sum of two additive moieties r_i and r_j. The traditional approach has been to use each moiety as a radius in order to draw spheres around the corresponding atomic nucleus; interatomic distance is optimal when the

spheres are just touching, and r_i and r_j are often called the van der Waals radii of the atoms.

There is an electrostatic component of the interaction energy when the atoms *i* and *j* possess an electrostatic charge. This component vanishes if the atoms are separated by an infinite distance, and increases as they approach one another. Its exact magnitude depends on the local environment (which is sometimes described by means of a dielectric constant) and upon the charge distribution in the molecules containing the atoms.

The hydrogen bonding function is a direction-dependent potential function in which the optimal distance is dependent upon the relative orientations of the atoms as they approach one another. Hence it is not appropriate to use the ordinary van der Waals radii r_i and r_j in order to describe hydrogen bonding atoms.

When studying the interaction of two molecules, such as a protein and a drug, the ducking method is used. With conventional methods of computation and graphical display, every molecule is treated as an agglomeration of atoms, and each atom has its own particular properties, which might include a van der Waals radius, an electrostatic charge and a set of bond properties. The unified computer-graphics approach uses a similar representation for the first interacting molecule, but only one atom or group at a time is considered for the other molecule (probe) (Goodford, 1989). In a programme called GRID (Goodford, 1984) the interaction of a probe

group with a molecule (or protein of a known structure) is computed at sample positions throughout and around the molecule, giving an array of energy values. This strategy is also used in 3D-QSAR studies called CoMFA (Comparative Molecular Field Analysis) (Cramer et al, 1988; Kim & Martin, 1991; Kim, 1992; Kim, 1993; Kim et al, 1993) using the data analysis method of partial least squares (PLS) (Wold et al, 1984).

3. Electrostatic approach to intermolecular interactions

There are a number of simple approaches to analysing intermolecular complex formation. They range from a strictly empirical approach which uses experimental enthalpies of complex formation to derive empirical parameters characteristic of each acid and base, to more semiquantitative conceptual approaches such as the Mulliken two-determinant charge transfer model (Mulliken & Person, 1969), which relates the strength of electron donor-acceptor complexes to the ionisation potential of the electron donor and the electron affinity of the electron acceptor; and Allen's H-bond model (Allen, 1975) which focuses on the ionisation potential of the electron donor and the bond dipole of the proton donor as the key features of the hydrogen bond (Kollman, 1978).

In the variational molecular orbital methods, the interaction energy is obtained as the difference between the energy of the super-molecule AB and that of the two isolated partners: $\Delta E = E_{AB} - (E^{\circ}_{A} + E^{\circ}_{B})$. Energies E_{AB} and ΔE depend on the particular conformation assumed by AB (distance between the partners and their mutual orientation). The origin of the forces between molecules is electromagnetic, arising from the charges on the electrons and nuclei of the atoms and molecules. A partition of ΔE into electrostatic, polarisation, exchange, and charge transfer terms which was proposed by Coulson (1957) is intuitively simple and accepted by many researchers. When atoms or molecules with closed shells of electrons approach each other so that substantial overlap of the charge clouds occurs, the energy increases (*exchange*

repulsion energy). This repulsive interaction has a similar origin, but is of the opposite sign, to the attraction which may arise in valence interactions when open shell atoms come together with the formation of a chemical bond. At larger separations, when there is negligible overlap, further electrostatic interactions may occur. If the species involved carry a net electrical charge then clearly there is a (long-range) coulombic interaction. Even if there is no net charge, the symmetry of the distributions of positive and negative charge on each molecule may often still lead to a direct *electrostatic energy*. Furthermore, the electrical field resulting from the charge distribution of one molecule may induce small changes in the electronic distribution of a nearby molecule. This also leads to an interaction energy, generally called an *induction* or *polarisation energy*. Finally there remains a rather subtle source of attractive forces due to the *dispersion effect*, in which instantaneous dipoles associated with the rapid movements of electrons in one molecule are correlated with those in a neighbour. This interaction is often the major source of attractive forces and is present for all types of molecule.

For some classes of molecules, other types of interaction energy may arise, which can be of considerable importance. These may resemble the valence interactions characteristic of atoms with partially filled electron shells, and include *charge transfer interactions* (Rigby et al, 1986).

It was found (Kollman, 1977) using the Morokuma (1971) component analysis that the electrostatic energy was a very good guide in determining the minimum energy

structural parameters, with the exception of the molecule-molecule separation for dimers and trimers including van der Waals molecules, charge-transfer complexes, ionic associations, and radical complexes. To predict the separation, all energy components are important (Umeyama and Morokuma, 1976).

In electron donor-acceptor (or charge transfer) complexes there is a fairly linear relationship between ΔE and E_{es}. In addition, the electrostatic contribution is responsible for the mutual orientation of the partners near the equilibrium geometry (Umeyama et al, 1977).

A first order charge density function which includes also the nuclear charges is:

$$\gamma_A(r_1, R_\alpha) = -\rho_A(r_1) + \sum_{\alpha \in A} Z_\alpha \delta(r_1 - R_\alpha) \qquad (3.1)$$

where $\rho_A(r_1)$ represents the diagonal element of the first-order electronic density matrix of the molecule A, and R_α is the vector defining the position of the nucleus α of A, having a charge Z_α, with respect to an arbitrary coordinate system (the same coordinates are employed also for molecule B).

By using such notation, the electrostatic interaction energy between molecules A and B can be calculated by:

$$E_{es}(AB) = \iint \frac{\gamma_\alpha(r_1, R_\alpha)\,\gamma_\beta(r_2, R_\beta)}{|r_1 - r_2|} dr_1 dr_2 \qquad (3.2)$$

whose value depends on the distance and the mutual orientations of the molecules A and B, i.e. on the values assumed by the sets of vectors R_α and R_β. It is evident that

the computation of E_{es} requires a computational effort noticeably smaller than the complete calculation of ΔE.

Equation 3.2 may be written in the following form:

$$E_{es}(AB) = \int V_A(r_2)\,\gamma_\beta(r_2, R_\beta)\,dr_2 \qquad (3.3)$$

where

$$V_A(r_2) = \int \frac{\gamma_A(r_1, R_\alpha)}{|r_1 - r_2|}\,dr_1 . \qquad (3.4)$$

Here V_A is the molecular electrostatic potential. Within the limits of the electrostatic approximation, one obtains directly from $V_A(r)$ the value of the interaction between A and a point charge placed at the point r (Scrocco and Tomasi, 1978).

3.1. Molecular electrostatic potentials (ESP)

The electric potential at point r in the vicinity of a given molecule is the electric force acting on a unit positive charge at that point caused by the nuclei and the electrons of the molecule. The electric potential can be evaluated on a grid defined outside the nominal van der Waals radii of the atoms in the molecule plus the radii of atoms in the approaching group (Williams & Yan, 1988). The potential does not take into account the properties of an attacking entity, nor does it reflect the polarisations and distortions that may occur in the course of an interaction. Nevertheless, the electrostatic potential provides certain well-defined information that can permit important insight into the reactive behaviour of molecules (Politzer et al, 1982).

The electrostatic potentials may be useful from two points of view; they should provide, on the one hand, a visualisation of the features of molecular charge distribution, i.e. comparisons and relationships among different molecules or among similar chemical groups placed in different chemical frames, and on the other hand, an approximate picture of the capability of the molecule in question to interact with other chemical species. The more correct the first order approximation, the sharper this picture becomes. It is particularly well suited for regions at medium or large distances from the molecule where reaction channels begin to assume a definite shape (Scrocco & Tomasi, 1973).

An obvious use for electrostatic potential is in the evaluation of the electrostatic energy of interaction between two molecules. If molecule A has the potential V_A and molecule B has a charge density ρ_B the energy of interaction is:

$$E_{AB} = \int V_A(r)\,\rho_B(r)\,dv = \int V_B(r)\,\rho_A(r)\,dv \qquad (3.5)$$

where the second term shows the underlying symmetry of the relationship (Hall, 1985).

The applications of the electrostatic potential cut across the traditional division between thermodynamic effects on the one hand and kinetic/mechanistic effects on the other. It means that if the electrostatic potentials at some particular sites in a series of related molecules are related to their interaction energies with a given molecule B, then they are also a measure of the relative activation energies and hence relative rates of reaction of these molecules with B. The electrostatic potential of a molecule

is indicative of what an approaching species encounters when it first comes into the neighbourhood of the molecule, before significant polarisation of the latter has taken place. The potential is accordingly relevant to the question of what is the most favoured path of approach and the preferred region of initial attack. In this manner it reveals mechanistic aspects of a reaction (Politzer et al, 1985).

Iso-energy maps for the interaction potential with an approaching proton show the existence of well defined attractive and repulsive regions and the presence of distinct potential wells in the neighbourhood of the nucleophilic portions of the molecules. This electrostatic model can be applied to select with remarkable accuracy the most probable sites for protonation or electrophilic attacks in molecules. Bonaccorsi et al employed electrostatic potentials in the neighbouring space around some three membered ring molecules (1970) and formamide (1972) to evidence the molecular sites more likely subject to electrophilic attacks.

An approach to reactivity and in particular to protonation problems can be provided by the study of the electrostatic potential created by a molecule in the surrounding space. Of course the electrostatic term is only one of several contributions to the energy associated with an interaction and cannot, in general, be taken as a measure of its energy of interaction with some species. An investigation by Ghio and Tomasi (1973) on the protonation process of three-membered ring molecules using the SCF LCAO MO method showed that there was a linear relationship between ΔE_{SCF} of protonation (the difference between energy of protonated and non-protonated

molecules) and corresponding electrostatic potential minima. They concluded that electrostatic potential was sufficient to obtain an ordering of the protonation energies in different chemical sites and to obtain reliable representations of the proton approaching paths.

Scrocco and Tomasi (1978) have reviewed various studies of the interactions between neutral molecules and cations and concluded that the electrostatic potential does not always represent an index sufficient for the protonation processes; in many cases better results are achieved by including also the polarisation term in the calculation; the charge transfer term becomes more important when one is interested in comparing proton reactivities of different chemical groups, for example of amines with ketones.

ESP has been found to be a useful tool in understanding and rationalising H-bond energies and geometries. Kollman et al (1975) evaluated ESP for proton donors and proton acceptors at a reasonable "representative" point in space and concluded that ESP was far more satisfactory than the Mulliken population in predicting *ab initio* calculated H-bond ability.

Molecular electrostatic potential has been widely used in structure-activity studies in order to find similarities between molecules which interact with the same biological receptor. Petrongolo et al (1978) applied *ab initio* calculated ESP to rationalise the neuroleptical activity of chlorpromazine while promazine, which has rather similar conformation and gross atomic population to chlorpromazine, is an inactive

compound. A method and a computer program to determine the similarity between two ESP distributions have been presented by Manaut and co-workers (Manaut et al, 1991; Sanz et al, 1993).

The total variance of the electrostatic potential on the molecular surface (which reflects the interaction tendency of the molecule) has been quantitavely related to solubility (Politzer et al, 1992; Politzer et al, 1993) and the enhancement factor ($E = y_2P/P_2^{sat}$, where y_2 and P_2^{sat} are the solubility and vapour pressure of the solute and P is the pressure of the system)(Murray et al, 1993) of molecules in supercritical fluids. Electrostatic potential can also be useful in LSER relationships by prediction of solvent hydrogen bond acceptor (Murray et al, 1991) and donor (Murray & Politzer, 1991) parameters, β and α respectively, and also solute hydrogen bond acidity and basicity, α^H_2 and β^H_2 respectively (Murray & Politzer, 1992).

Binding of some simple anions, e.g. Cl^-, to nucleic acid bases (Pullman et al, 1977) and protonated nucleic acid bases (Goldblum & Pullman, 1978) has been investigated using molecular electrostatic potentials. The negative ion has been considered as a point negative unit charge and its electrostatic interaction energy with the molecule at every point of space has been measured as inverse ESP (-V). A method has been developed to overcome the problem of the exchange repulsion occurring as a result of the overlapping of the electron clouds of the anion and the molecule.

Politzer et al (1982) demonstrated that the electrostatic potential can be used to

predict and interpret nucleophilic process, provided that the molecules are examined in states of distorted geometries, which already anticipate somewhat the approach of the nucleophile. Nucleophilic process can also be analysed using the electrostatic potential of the ground-state undistorted molecule by computing the potential on the three-dimensional surface of the molecule that corresponds to a constant electronic density of 0.002 electron/bohr^3; the relative magnitudes of the positive electrostatic potentials in various regions on this surface reveal the sites most susceptible to nucleophilic attack (Sjoberg & politzer, 1990).

Calculated electrostatic potentials have been used to examine in detail the effects of amino and nitro substituents in activating or deactivating the benzene ring toward electrophilic attack (Politzer et al, 1984).

Molecular electrostatic potentials can be calculated quantum mechanically or by molecular mechanics methods. The electrostatic potential at any point r in space may be expressed quantum mechanically by the following equation:

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \sum \frac{\rho(r')\,dr'}{|r' - r|} \qquad (3.6)$$

in which Z_A is the charge on nucleus A, located at R_A, and $\rho(r)$ is the electrostatic density function of the molecule. The first term represents the nuclear contribution and the second term the electronic contribution. Those regions that have high nuclear contributions will yield positive ESP, corresponding to repulsive interaction energies with point positive charges, and those with higher electron contributions will yield

negative ESP, corresponding to attractive energies. V(r) is a real physical property, rigorously defined and experimentally measurable, for example by scattering experiments. Electrostatic fields calculated by theoretical methods have been found to agree well with such experimental results (Loew & Burt, 1990).

In the quantum mechanical methods, there are three factors that determine the computational effort necessary to obtain, via the electrostatic potential, the information on the chemical behaviour of a molecule: the time necessary to calculate the wavefunction, the number of points *r* where *V(r)* is calculated, and the time necessary to calculate *V(r)* at a single point.

In large molecules the calculation of the wavefunction can be very time-consuming. For this reason a large portion of the calculations of *V* for molecules of organic or bio-organic interest have been thus far performed on semiempirical wavefunctions (Scrocco & Tomasi, 1978). Within the framework of the MO-LCAO approximation, equation 3.6 can be rewritten in terms of the basis set of atomic orbitals χ, $P_{\mu\nu}$ being the element $\mu\nu$ of the first-order density matrix:

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \sum_\mu \sum_\nu P_{\mu\nu} \int \frac{\chi_\mu(r')\chi_\nu(r')}{|r - r'|} dr' \qquad (3.7)$$

Giessner-Prettre and Pullman employed semiempirical CNDO (1972) and INDO (1974) wavefunctions to calculate ESPs. They showed that in order to obtain quantitatively good isopotential maps from these wavefunctions one must transform

the eigenvectors from the orthogonal basis set into a regular Slater basis and introduce all the integrals. However, they concluded that although gross features of the potential maps were generally obtainable with CNDO or INDO functions, some care and considerable experience were required for fine distinctions, even after deorthogonalisation (Giessner-Prettre & Pullman, 1975).

The comparative analysis of molecular electrostatic potentials computed from the semiempirical and *ab initio* wavefunctions stated that the characteristics of MNDO-derived isopotential maps can be related to those determined from the *ab initio* wavefunction calculated at the 6-31G* level (Luque et al, 1990). Electrostatic potentials calculated from AM1 wavefunctions are in qualitative agreement with those calculated from *ab initio* STO-3G, and they are also useful for deriving atomic charges for use in molecular dynamics studies (Ferenczy et al, 1990).

The characteristics of MNDO and AM1 methods imply certain requirements for computing molecular electrostatic potentials: the semiempirical wavefunction must be deorthogonalised to describe adequately the essential details of isopotential maps; each Slater type orbital of the minimal basis set employed in both MNDO and AM1 methods is fitted to four Gaussian functions to facilitate the calculation of integrals in equation 3.7; and since MNDO and AM1 methods are valence electron methods, the "core" effective charge is used to evaluate the nuclear electrostatic term (Orozco & Luque, 1990).

In a comparative study of electrostatic potential maps obtained with different basis sets, it was concluded that basis sets of at least double zeta quality ought to be used in order to obtain reliable details in the potential around a molecule (Almlöf & Støgård, 1974). However, quantitave comparison of ESP distributions from MNDO, AM1, STO-3G, 3-21G, 4-31G, 6-31G, 4-31G*, 6-31G*, and 6-31G** wavefunctions (using cluster analysis of the similarity matrices) showed that the ESP distributions in these methods are similar in a reliable range (Rodriguez et al, 1993).

The second factor, the number of points, depends largely on the complexity of the molecule and on its topology. Sometimes it may be sufficient to know V at some selected points. There are strategies for minimising the number of points to get a 2D map from 3D information (Scrocco & Tomasi, 1978).

The third factor is the time necessary to compute $V(r)$ at a given point. Due to the relatively long time needed for numerical integrations of equation 3.4, the calculation of $V(r)$ for large molecules of interest becomes hardly feasible. For the usual Gaussian basis sets, the derivation of V by integration of the Poisson equation can reduce the computation time (Srebrenik et al, 1973).

Another kind of approximation of $V(r)$ relies on the multipole expansions of this function. One-centre and many-centre multipolar expansions are possible. Many-centre multipolar expansions are more promising (Scrocco & Tomasi, 1978) and have been used by Rein (1973) who chose to expand the charge distribution and V about the

atomic centres. Some studies use only the monopoles of the atom-centred multipole expansion (net atomic charges)(Dunfield et al, 1978). It is also possible to expand the electronic part of V on the charge centres of the localised molecular orbitals (Bonaccorsi et al, 1974; Lavery et al, 1982). These expansion procedures have been used within the context of a scheme in which V for a large biological system, such as a segment of DNA, is obtained by dividing it into subunits, computing wavefuntions, electronic densities and electrostatic potentials for the individual subunits, and then superposing these potentials to produce V for the whole system (Pullman & Berthod, 1978). The subunits are created by breaking appropriate single bonds and introducing hydrogen atoms to saturate the resulting free valencies. Tests of this approach indicate that if the subunits are chosen carefully, the resulting superposed potential should be a good approximation to that computed directly for the entire system; the perturbing effects of the added hydrogens are relatively insignificant (Pullman & Berthod, 1978).

The concept of representing localised molecular orbitals by multipole expansions has the desirable feature that such expansions could conceivably be transferred from one molecule to another, provided of course that the localised molecular orbital is reasonably valid for both molecules. The possible transferability of such 'group potentials' could save a great deal of computational time. This would also make it possible to obtain at least qualitative representations of V for very large systems that cannot presently be treated by other means (Politzer et al, 1985).

In another method, molecular electrostatic potentials were obtained as the first-order interaction energy between a unit positive charge and the molecule using the semiempirical MNDO and AM1 Hamiltonians. The atomic charge distributions derived from such electrostatic potentials are of similar quality to those obtained as expectation values of the coulomb potential operator using semiempirical wavefunctions derived from MNDO or AM1 density matrices. The method requires trivial amounts of computer time and therefore should be suitable for large biological molecules (Cummins & Gready, 1990).

3.2. Atomic point charge

Partial atomic charges are a direct result of the electron distribution and are intuitively appealing for the purpose of understanding physical and chemical behaviour. Although the description of atomic charges is shrouded in uncertainty, the use of atomic charges in empirical potential energy functions provides an extremely powerful tool. There is an ever-increasing supply of empirical and semiempirical methods for the calculation of atomic charges (Dixon & Jurs, 1992).

3.2.1. Empirical methods (topology-based) charge schemes:

Smith et al (1951) developed an atomic charge scheme for alkyl halides that utilised bond polarisabilities (from molecular refraction data), Pauling's covalent radii, and empirically adjusted effective nuclear charges (parametrised on molecular dipole

moments).

Del Re (1958) distributed the bonding electrons in a bond on the basis of a simple MO-LCAO treatment of the localised bond. MO coefficients were derived from a Hamiltonian matrix whose elements were a simple function of the electronegativity differences between bonded atoms.

Sanderson (1960) used a linear dependence of electronegativity on charge and the principle of electronegativity equalisation to assign charges in diatomic and other simple molecules.

Allinger and Wuesthoff (1977) extended the treatment of Smith et al (1951) by including the effects of nonadjacent bonds, (the original model considered only atoms bound to a common atom), and they obtained generally good agreement between the observed and calculated dipole moments of some haloalkanes and haloketones. In this extension, unlike the original formalism of Smith et al (1951), the derived charges are dependent on the atom taken as the origin.

The concept of partial equalisation of orbital electronegativities (PEOE) was introduced by Gasteiger and Marsili (1980). They used a quadratic dependence of electronegativity on atomic charge to transfer electron density from weakly electronegative atoms to strongly electronegative atoms.

Mullay (1986) applied the principles of electronegativity equalisation and charge conservation to each of the N bonds in a molecule to generate N simultaneous equations whose solution yielded the atomic charges.

Another empirical procedure for the rapid evaluation of partial atomic charges has been given by Abraham and coworkers (1982, 1988). This scheme is based on two fundamental chemical concepts:

1- the inductive effect in saturated molecules

2- Hückel molecular orbital calculations (HMO)(Streitweiser, 1961)

The inductive effect operates via atomic electronegativity and polarisability, and the Hückel scheme operates through the appropriate coulomb and resonance integrals. The first step is evaluation of atomic π charges q^{π} from the diagonal elements of the Hückel density matrix. The Hückel parameters h and k are used to distinguish among various types of atoms that can conjugate with a π system. They refer to the basic HMO relations:

$$\alpha_i = \alpha_0 + h_i\beta_0 \qquad (3.8)$$

$$\beta_{ij} = k_{ij}\beta_0, \qquad (3.9)$$

where α_0 and β_0 are the standard coulomb and resonance integrals for a carbon $2p_z$ orbital and h_i and k_{ij} are modifications to these integrals for any conjugated atom i.

The subscript j refers to an adjacent atom in the π system.

The iterative σ calculation is then initiated, whereby charge is induced on an atom i by each of its bonded neighbours j according to a one-bond α effect, a two-bond β effect, and a three bond γ effect. In any iteration, the σ charge that develops on atom i due to α effect exerted by j is given by:

$$q_\alpha^\sigma = \frac{(X_j - X_i)}{a_{ij}} \qquad (3.10)$$

Here, X_i and X_j are Mulliken atomic orbital electronegativities for atoms i and j, respectively, and a_{ij} is an adjustable bonding parameter whose value depends on the types of the two atoms involved. The β and γ effects also act through the i-j bond and are due to all atoms that are β and γ to atom i, with j being the α atom. In this case, charge is transferred from j to i according to:

$$q_\beta^\sigma(j \rightarrow i) = \sum_{k\beta} \frac{(\chi_{k\beta} - \chi_H)}{b} P_i(q_i) \qquad (3.11)$$

$$q_\gamma^\sigma(j \rightarrow i) = \sum_{m\gamma} \frac{(\chi_{m\gamma} - \chi_H)}{5b} P_i(q_i) \qquad (3.12)$$

where X_H is the atomic orbital electronegativity of hydrogen and b is assigned a value of 198.4 in all instances. The quantity $P_i(q_i)$ is a charge-dependent polarisability for atom i, given by:

$$P_i = [1 + 3(q_i^0 - q_i)] P_i^0 \qquad (3.13)$$

where P^0_i is a Miller-Savchik (1979) atomic hybrid component polarisability and q^0_i is a fixed standard state charge defined for atom i according to its electronegativity. The atomic polarisability for the current cycle depends on how much q_i, the total $\sigma + \pi$ charge from the previous cycle, deviates from its standard state value. Thus, in the first iteration, all previous σ charges are 0 and $q_i = q^\pi_i$. New σ charges q^σ_i are computed from α, β, and γ effects and $q_i = q^\pi_i + q^\sigma_i$ is sorted for use in the next cycle. The process is continued until the total charge on each atom remains constant over two successive iterations. The scheme contains a number of empirical parameters, many of which are adjustable and have been optimised by Abraham and coworkers to yield atomic charges that reproduce experimental gas phase electric dipole moments.

This scheme was extended (Abraham & Smith, 1989; Abraham et al, 1991) to include charge calculations for nitrates, nitriles, sulfides, thiols, thiophenes, and sulfoxides and as an added complexity, the ϖ-technique (Streitweiser, 1961) was used in the HMO calculation for π charges.

Dixon and Jurs (1992) made a few simple additions to Abraham's charge calculation scheme to obtain atomic charges in a variety of ionic species and found the results quite consistent with many intuitive concepts and considerably more appealing than the results of either MNDO or AM1 calculations.

3.2.2. Quantum mechanical methods

Atomic charge, unfortunately, is not a property which can be directly determined from the Hartree-Fock wave function. Some scheme must be adopted to divide the total electronic charge among the atoms in a molecule. The computationally simple technique of Mulliken population analysis (1955) is, and has long been, most commonly used (Chirlian & Francl, 1987) and it is the method used to apportion the charge on each atom in the *ab initio* treatment. This analysis is explained briefly below.

The square of a normalised molecular orbital ϕ_l can be interpreted as a probability density function for one electron. Integration of ϕ_l^2 over all space gives a total probability of unity for finding the electron:

$$\sum_r c_{rl}^2 + 2\sum_{r<s} c_{rl}c_{sl}S_{rs} = 1 \qquad (3.14)$$

The squared terms c^2_{rl} are related to the amount of charge on each atom r and the cross terms $2c_{rl}c_{sl}S_{rs}$ measure the amount of charge in the region between atoms r and s from the molecular orbital ϕ_l. Mulliken (1955) has defined the overlap population or bond order in terms of the cross terms. It might seem appropriate to define the atomic populations or charge densities as simply the squared terms c^2_{rl}, but in order to include all of the electron charge, Mulliken chose to include part of the overlap population as well, dividing it equally between atoms r and s. Therefore, let $q_r(l)$ be

the gross atomic population on atom r due to molecular orbital ϕ_l:

$$q_r(l) = n_l[c_{rl}^2 + \sum_{r<s} c_{rl}c_{sl}S_{rs}] \qquad (3.15)$$

The total gross atomic population or charge density is the sum of $q_r(l)$ for all molecular orbitals (Gimarc, 1979):

$$q_r = \sum_l q_r(l) \qquad (3.16)$$

It should be stressed that the Mulliken approach is an approximation; it gives a reasonably good picture of the electron distribution in a molecule but is far from perfect. Williams and Yan (1988) have explained some problems of this method. Better methods have been suggested for dividing up the overlap populations into orbital populations. These methods allocate charge to a molecule's atoms based on physical criteria rather than simply by the equal partitioning method of Mulliken population analysis. However, none of these treatments is completely satisfactory (Chirlian & Francl, 1987).

Unlike net atomic charges, the molecular electrostatic potential is a rigorously defined quantum mechanical property and some initial efforts to represent the extramolecular electric potentials by intramolecular point charges were made by Kollman (1978). Momany (1978) determined net atomic charges by fitting the classical electrostatic coulomb potential to the potential obtained from molecular orbital calculations for formaldehyde, methanol, and formic acid. The magnitude and direction of the

experimental dipole moments of the molecule were used in the fitting procedure.

Wiberg (1980) calculated the total charge distribution at every point in space from the appropriate *ab initio* wavefunction for some substituted methanes.

Optimised net atomic charges (potential-derived charges) were obtained by Cox and Williams (1981). In their method, the molecular electrostatic potentials at the grid points were fitted by a set of point charges located at the nuclei. The criterion of fit was the sum-of-squares function defined as follows:

$$R = \sum_{i}^{m} w_i |V_i^0 - \sum_{j}^{n-1} q_j \, r_{ij}^{-1} + (\sum_{j}^{n-1} q_j) \, r_{in}^{-1} - Z r_{in}^{-1}|^2 \qquad (3.17)$$

where V^0_i is the calculated QM electrostatic potential at point i, q_j is the net charge on atom j (a variable), r_{ij} is the distance between atom j and the ith grid point, m is the number of grid points, n is the number of atoms, Z is the net charge in the molecule (which is zero for nonionic species), and W_i is the statistical weight for each point (which was taken as unity). The last two terms of R reflect the condition that the net atomic charges must sum to Z. To find the minimum value of R, the first partial derivatives of R with respect to each of the (n - 1) independent net atomic charges were obtained, and the resulting linear equations solved in the usual least-squares fashion. Singh and Kollman (1984) discussed the question of including the location and charge of lone pairs.

Chirlian and Francl (1987) presented a new non-iterative and rapid algorithm for fitting atomic charges to molecular electrostatic potentials. This avoided many of the disadvantages of nonlinear least squares fit procedures such as the need for initial guess charge, iterative solution procedures and possible convergence problems.

Two methods of generating atomic charges appropriate to variable molecular conformations were proposed by Reynolds et al (1992). The first method involves determining a single ESP and constraining the charges to reproduce the dipole at an alternative geometry. The second method involves determining the ESP at appropriate conformations and weighting the ESP for each conformation according to the appropriate Boltzmann factor. The main use of these multiple conformation ESP derived charges is likely to be in Monte Carlo or molecular dynamics simulations where the ability of these methods to search conformational space is matched by the ability of the multiple conformation ESP derived charges to yield the correct electrostatic properties in these conformations.

Atomic point charges were determined using the semiempirical molecular orbital method MNDO in conjunction with the ESP fitting technique and it was found that the approach was able reasonably to reproduce ESPs as well as point charges relative to 6-31G* *ab initio* calculations (Besler et al, 1990). Merz (1992) compared MNDO ESP derived point charges with AM1 and PM3 ESP derived point charges for a large data base, and found that MNDO correlated well with 6-31G* ESP derived point charges, while AM1 and PM3 did so quite poorly.

4. Hydrogen bonding

According to Pimentel and McClellan (1960), a H-bond exists between a functional group A-H and an atom or a group of atoms B when:

a) there is evidence of bond formation (association or chelation),

b) there is evidence that this new bond linking A-H and B specifically involves the hydrogen atom already bonded to A.

H-bonding is an interaction of intermediate energy between complete chemical bonding and weak van der Waals interactions. Indeed, hydrogen bonds are chemical bonds, but the relative weakness of the bonds gives them distinctive properties that warrant a distinguishing name.

In a hydrogen bond hydrogen is bonded to more than one other atom, for instance, to two atoms named X and Y. The hydrogen bond can be symmetric or asymmetric depending on whether the proton is located midway between the terminal atoms or closer to one of them. Symmetric hydrogen bonds are quite rare, the ion FHF^- being the classical example. They are very strong; the dissociation energy of FHF^- to FH + F^- is of the order of 160 kJ mol^{-1} although the exact value is uncertain. In the majority of cases, the hydrogen bonds are asymmetric with energies being of the order of 5-10 kJ mol^{-1}. In asymmetric hydrogen bonds the stronger bond of the hydrogen will be written X-H and termed a normal X-H bond, but the weaker bond will be written Y···H and termed a *hydrogen bond* (abbreviated H-bond in this thesis) (Bratož, 1966).

The presence of H-bonding may be detected in several ways:

1. association, as in water, or as in acetone-water liquid mixtures;

2. crystal structure determination, showing the presence of the H atom between two other electronegative atoms;

3. sublimation energy;

4. change in characteristic vibration frequency, as e.g. O-H, between the monomer molecule and its dimer, or H-bonded polymer;

5. enhanced intensity of vibrational transitions in the infra-red which are often much stronger in the H-bonded polymer than in the monomer;

6. a broadening of these infra-red vibrational bands, which is noticeable even in gaseous, hydrogen-bonded dimers;

7. change of vibration frequency on compression. Presumably the compression is most effective in shortening the longer half of the bond;

8. effect on certain electronic transitions in a molecule such as pyridine, when the nitrogen atom is or is not hydrogen bonded as, for example, to water (Coulson, 1957).

Most H-bonded systems have been detected in solution or in crystal and factors such as solvation or long-range crystal forces may complicate the analysis of the bonding. Fewer systems have been studied in the gas phase but their numbers include some of the simplest molecules and they are therefore important for study.

The H-bond is said to be *intra*-molecular, e.g. in salicylaldehyde, or *inter*-molecular, e.g. in carboxylic acid dimers, depending on whether or not the atoms X and Y

belong to the same molecule. In the case of intermolecular association, one can distinguish between *self*-association and *mixed* association, if the complexing molecules are of the same type (e.g. water) or different types (e.g. thiamine and adenine), respectively. H-bonding gives rise to a specific interactions between atoms or functional groups in which: the strength is higher than that of dispersion forces alone; it is directed along a X-H bond; and it demonstrates some angular dependencies (Lippert 1975).

Oxygen, nitrogen, fluorine and chlorine are the best known H-bonding atoms. Evidence shows that the older view that the atoms involved in H-bonding must be highly electronegative is undesirably restrictive; suitable H-bond donors include the halogen activated C-H, the acetylenic C-H, and the S-H groups. A study by Taylor and Kennard (1982) provided conclusive evidence of the existence of C-H···O H-bonds in crystals. H-bonding bases include aromatic systems, possibly even the boron atom. Accepting these as members of the H-bonding family will aid in the recognition of other new examples and, more importantly, will guide us formulating a useful theory of the H-bond (Pimentel & McClellan, 1960).

4.1. The nature of hydrogen bonding

In spite of the voluminous experimental data available on the hydrogen bond, understanding of the interaction within the framework of quantum mechanical theory is far from satisfactory. It is now clear that the hydrogen bond has contributions from

the following forces:

1) electrostatic interaction (E_{es})

2) delocalisation energy (charge transfer) (E_{ct})

3) dispersion forces (polarisability forces) (E_{pol})

4) exchange repulsion forces (E_{ex}).

The fact that all known H-bonds are between electronegative elements suggests that there are electrostatic contributions in the energy of these bonds. The success of the electrostatic model in predicting the correct H-bond energies in certain systems is possibly due to the fact that the other terms are not significant in those systems (Murthy & Rao, 1970) or the algebraic sum of the other three forces is zero (Tsubomura, 1954). Although the electrostatic contribution to the hydrogen bond is important, it cannot represent the whole phenomenon. Large increase in intensity of infra-red absorption implies that during the motion of the proton there are larger fluctuations of charge when the H-bond is formed than when it is not formed. This seems to be possible only if charge can move on to and away from the farther atom during the vibrations of H; this is referred to as a delocalisation of electrons. Both the electrostatic and delocalisation forces would tend to shorten the total length of the H-bond. It is evident therefore that there must be some repulsive force (E_{ex}) which acts when the atoms approach too closely together. This force is exceedingly difficult to calculate accurately. Dispersion forces due to the high polarisability of the unshared pairs of electrons on the both heavy atoms involved in H-bonding play some part in the potential between non-bonded atoms and between the inner-shell and non-bonding

electrons of atoms which are bonded together (Coulson, 1957).

The decomposition of interaction energy in "normal" hydrogen bonds shows that, at distances larger than that at equilibrium between the partners the interaction energy is dominated by the E_{es} term. Near the equilibrium position the term numerically larger is E_{es} followed by E_{ex} which is of opposite sign. The other attractive terms (E_{pol} and E_{ct}) have in general a value that partially compensates that of the exchange term, and E_{es} alone represents a reasonable estimate (generally in excess) of ΔE (interaction energy). At shorter distances all contributions to ΔE become larger, and it is no longer possible to find evidence that one particular term is the dominant one (Scrocco & Tomasi, 1978).

In weak H-bonds (for example those whose proton donor group is CH), the results of the decomposition of ΔE are analogous, but near the equilibrium distance the charge-transfer term is also relatively important. The magnitude of E_{ct} does not change substantially in passing from strong to weak complexes, and its relative weight is larger when ΔE is lower. For strong hydrogen bonds, involving ionic partners, the importance of E_{es} is greater over the whole range of distances. The conformational potential energy surface, which determines the directionality of the H-bond and the other changes in the mutual orientation of the partners at the equilibrium distance, is essentially controlled by the electrostatic term alone (Scrocco & Tomasi, 1978) and electrostatic potential gives a good approximate of the hydrogen bonding direction (Legon, 1990).

The theory of normal hydrogen bonded complexes follows quite closely that of donor-acceptor complexes. Calculations show that there is a migration of electrons on H-bond complexing which is towards the hydrogen atom donor and there is a large electrostatic attraction in the no-bond state of the hydrogen bond. There is now general agreement that this electrostatic energy is a large contribution to the binding energy and for weak complexes is probably dominant. The charge transfer contribution to the bond energy will depend on the ionisation potential of the base and the electron affinity of the acid. There are conflicting views about the amount of the charge transfer energy in any one complex but it appears to increase in proportion to the Coulomb energy as the strength of complex increases (Murrell et al, 1985).

There is no general correlation between H-bonding energies (ΔH_f) and energies of proton transfer (ΔH_i) in solution (Arnett et al, 1974). On the other hand, a correlation between the free energy changes for H-bonding formation in solution and the gas phase protonic acidity or basicity of proton donor or proton acceptor has been established for OH···O and OH···N systems (Zeegers-Huyskens, 1986a).

4.2. Geometrical properties of H-bonding

The geometry of H-bonded complexes is a matter of considerable interest. Theoretical studies and, on the experimental side, crystal structure determination provide direct information about such questions of configuration.

One of the most obvious manifestations of hydrogen bonding is that the H···O length calculated in crystals is significantly shorter than the sum of the van der Waals radii for O and H (2.8 Å). Typical values lie between 1.8 to 2.0 Å for N-H···O bonds and 1.6 to 1.8 Å for O-H···O bonds (Desiraju, 1989).

In an analysis of the geometries of one hundred O-H···O hydrogen bonds in crystal structures, Ceccarelli and co workers (1981) found that twenty-five of the H-bonds could be described as bifurcated. An examination of the geometries of such bonds suggests that the O···H distace is increased, and the O-H···O angle reduced, relative to the values typical for non-bifurcated H-bonds.

Aside from the A-H distance, the angular orientations are of importance. One of the most crucial questions is the evidence for the extent of deviation from linearity of the A-H···B bond and the angle between the H···B line and the bond made by B to its adjacent atom. The concept of "lone pairs" of electrons, originating in the valence-bond theory of molecular structure, has had some success in rationalising directionality in hydrogen bonding (Pimentel & McClellan, 1960).

Schneider (1955), on the basis of point charge models for the water and hydrogen fluoride and also on the basis of a consideration of known H-bonded structures, suggested that the directional properties of the hydrogen bond are determined largely by the directional properties of the lone pair orbital, and that the strongest bonds result when the H-bond direction is collinear with the lone pair orbital direction.

The facts that hydrogen fluride molecules form zig-zag rather than linear chains in the crystal, that HCN crystals contain linear HCN···HCN chains, and that hydrogen bonds involving carbonyl "lone pair" donors often have C=O···H angles near 120° are all evidence in support of the view that the proton forming the hydrogen bond is approaching a lone pair of electrons. However, the lone-pair theory does not allow us to obtain a better and more detailed understanding of the key features of H-bond (Kollman, 1971). For carbonyl bases the H-bond should make an angle of 120° to the C=O bond direction. In crystalline formic acid this angle is 122°, but in crystalline acetic acid the angle is 144°. The deviation from 120° is also observed in the crystal structures of amides.

Legon (1990) has reviewed some gas-phase investigations of H-bonding (the results of rotational spectroscopy) and suggested some rules for H-bonding complex geometries. The gas-phase equilibrium geometry of a H-bonded dimer B···HA can be obtained by assuming that the axis of the HA molecule coincides with the supposed axis of a nonbonding electron pair as conventionally envisaged. Evidently, this rule is electrostatic in origin if it is assumed that the positive end H of the HA molecule seeks out the direction of greatest electron density (the axis of the n-pair) on the molecule B. However the directing effect of the n-pairs becomes less dominating as the hydrogen bond becomes weaker. When B carries no n-pairs but only π-bonding pairs, the geometry can be predicted by assuming that the axis of the HA molecule intersects the internuclear axis of the atoms forming the π-bond and is perpendicular to the plane of symmetry of the π-bond. When the acceptor molecule B carries both nonbonding and π-bonding electron pairs (e.g. HCN), the nonbonding pairs dictate the angular geometry of the complex

(Legon & Millen, 1982).

If the allene molecule acts as H-bond acceptor, the H-bond complex has an L-shaped geometry and the H end of the H-bond donor (HF) might move with facility from one of the four equivalent positions to another (Legon & Willoughby, 1988).

On the other hand, some other investigations suggest that a distinct preference for H-bonding in the direction of lone pairs does not exist. According to Murrell et al (1985) in almost all cases the hydrogen atom lies along the line of centres of the two heavy atoms although the energy required to displace it from this line must be very small (Bevan et al, 1980). Kroon and co-workers (1975), on the basis of a statistical and quantum mechanical analysis on O-H···O hydrogen bonding in molecular crystals, concluded that a distinct preference was neither observed nor calculated for H-bonding in the direction of one of the acceptor lone pairs. The range of accessible dimer geometries was determined largely by the classical coulomb energy.

In a survey of O-H···O hydrogen bond geometries determined by neutron diffraction, the mean O-H···O valence angle was found to be 167.1° and the shorter O···H bonds were more linear. A preferred direction of H-bonding with respect to the acceptor oxygen atom, which was in, or close to, the plane containing the oxygen lone pair orbitals, existed, but there was no evidence of a preferred direction within that plane (Ceccarelli et al, 1981).

4.3. The importance of hydrogen bonding (incorporation of hydrogen bond in QSAR and LSER equations)

Only a few examples are needed to illustrate the broad application of the principles of hydrogen bonding. The structures of many organic crystals and inorganic crystalline hydrates are determined by hydrogen bonding. Among the important fields of application of H-bonding are adsorption, catalysis, dyeing, kinetics and enzyme activity (Pimentel & McClellan, 1960).

The helical structures of proteins and DNA are fixed by H-bonding (Abraham et al, 1989a). The H-bond is a ubiquitous element of molecular recognition (Neder & Whitlock, 1990). Experiments on engineered enzymes, modified inhibitors and synthetic DNA duplexes indicated that an individual uncharged H-bond contributed some 2.1 to 7.5 kJ mol^{-1} to the binding energy and a factor of two to twenty to specificity (Fersht, 1987).

H-bonding is an important interaction in the control of drug activity by its effect on the solubility and partitioning of drugs and drug receptor interactions, Hence it is appropriate to incorporate its parameters into quantitative structure-activity relationships (QSARs), which are mathematical equations relating biological activity to physico-chemical and structural parameters as an aid to correlating biological activitiy with physicochemical properties (Dearden, 1990). The Hansch equation (1969) which for more than two decades has formed the foundation of QSARs in biology is:

$$\log(RBR) = -a(logP)^2 + b\log P + cE + dS + e \qquad (4.1)$$

This attempts to express relative biological response (RBR) in terms of a set of physical variables representing hydrophobic, electronic, and steric effects. Here *P* is the partition coefficient, *E* is an electronic effect (commonly represented by Hammett's σ value) and *S* is a steric interference term. Log *P* values measured in different solvent pairs are linearly correlated only when the organic solvents have similar physical properties, in particular, similar H-bonding capacity (El Tayar et al, 1991). These restrictions prompted Seiler (1974) to define the parameter I_H as a measure of H-bonding capacity of the solutes which is calculated as the difference between the 1-octanol-water partition coefficient and the cyclohexane-water partition coefficient. Use of the I_H parameter has led to a new physicochemical model in the design of brain-penetrating H_2 histamine receptor antagonists (Young et al, 1988).

Octanol/water partition coefficients, which are considered to model blood/lipid partition, are influenced by some major properties including H-bonding forces (Leo et al, 1976).

Fujita et al (1977) have shown that, when the relative H-bonding effect of drugs on phases involved in the binding at the site of biological action differs from that in the 1-octanol-water partitioning phases used as the reference to estimate π values, it is in fact expressible by an indicator variable and they have used this indicator variable in QSARs (Kamoshita et al, 1979).

Furthermore, because of the importance of H-bonding in solute-solvent interactions, its terms have been used in linear solvation energy relationships (LSERs). These are equations that can correlate many solubility- and solvent-dependent properties with linear

combinations of free energy or enthalpy contributions by three types of terms: cavity term, dipolar/polarisability term and H-bonding terms (Kamlet et al, 1982; Kamlet et al, 1983; Taft et al, 1985; Kamlet et al, 1986; Kamlet et al, 1987; Kamlet et al, 1988a,b). The cavity term is the measure of the free energy necessary to separate the solvent molecules (overcome solvent-solvent interactions) to provide a suitably sized cavity for the solute. The endoergic cavity term depends firstly on the solute molar volume V_2 taken as its molecular weight divided by its liquid density at 20°C, and secondly on the solvent Hildebrand solubility parameter, defined by $\delta_H = [(\Delta H_v - RT)/V_1]^{1/2}$, where ΔH_v is the molar heat of vaporisation and V_1 is the solvent molar volume (Kamlet et al, 1984). The exoergic solute-solvent dipolar/polarisability interactions are measured by $s(\pi^* + d\delta)$ term; π^* scale is an index of dipolarity/polarisability, which measures the ability of the medium to stabilise a dipole or a charge by virtue of its dielectric effect (Kamlet et al, 1977). Values of π^* for "select solvents", nonchlorinated nonprotonic aliphatic solvents with a single dominant bond dipole, have been shown to be generally proportional to molecular dipole moments (Abboud et al, 1977). The δ term is a polarisability correction parameter, with δ taken as 0.5 for polyhalogenated solvents, 1.0 for aromatic solvents, and zero for all others (Abraham et al, 1988c). The α scale of hydrogen bond donor acidities measures the solvent's ability to partially donate a proton in a solvent to solute hydrogen bond (Taft & Kamlet, 1976; Taft & Kamlet, 1979). The β scale of hydrogen bond acceptor basicities describes the solvent's ability to accept a proton (donate an electron pair) in a solute to solvent hydrogen bond (Kamlet & Taft; 1976; Taft et al, 1982). Rather than being based on solvent effects on single indicators, the solvatochromic parameters have been obtained by averaging multiple normalised solvent effects on a variety of properties involving many diverse types of indicators.

4.4. Thermodynamic properties of hydrogen bonds

Since a H-bond is generally formed in an equilibrium reaction, the thermodynamic equations are applicable. Reliable values of thermodynamic functions of H-bonds are derived from the equilibrium constant, K, and its variation with temperature. The experimental techniques vary only in their approach to finding the concentration or pressure values needed to determine K. The basic relations are:

$$K = \frac{activity\ of\ product}{activity\ of\ reactant} = \frac{[A\text{-}H\cdots B]}{[AH]\ [B]} \qquad (4.2)$$

$$\Delta G = -RT\ lnK \qquad (4.3)$$

$$\left(\frac{\partial \ln K}{\partial T}\right)_p = \frac{\Delta H}{RT^2} \qquad (4.4)$$

$$\Delta G = \Delta H - T\Delta S \qquad (4.5)$$

In equation 4.2 it is common to use concentration or pressure and to adjust the experimental conditions such that these quantities are nearly equal to activity and fugacity; then no appreciable error is involved. It is important to remember that the units of K influence both ΔS and ΔG, and that values of K (or ΔS and ΔG) are not directly comparable unless the units are the same (Pimentel & McClellan, 1960).

Applying the thermodynamic functions of H-bond complex formation is a reliable way of providing a quantitative description of H-bonding. There are three main approaches to estimate the thermodynamic functions of H-bond: a) calculative (quantum chemical

calculations, Monte-Carlo studies or molecular dynamics) b) correlative c) experimental (Raevsky et al, 1991).

4.4.1. Theoretical calculations

There have been many attempts to calculate ΔH independent of the equilibrium constant. The difficulty of a complete theoretical treatment of the H-bond unfortunately requires approximations. The uncertainties thus introduced deprive the calculations of predictive value. The usual approximations are based on some sort of electrostatic model, with computation of electrostatic, dispersion and repulsive contributions by the methods of classical physics (Pimentel & McClellan, 1960).

The early electrostatic theory was based on a model in which the dipole moment of a molecule is represented as formal charges on the atoms. Other studies have considered four electrons explicitly, two from the A-H bond and two from the B lone pair. The electrons are located on the AH···B line in such a way as to give the correct values of the AH bond and the B lone pair dipole moments. In elaborated electrostatic theories charge distribution is represented more carefully (Bratož, 1966). Schneider (1955) assumed that the dominant term in the interaction energy could be the interaction of the proton with the lone pair dipole and used the centroid of charge of the hybridised lone pair orbital as a rough measure of the relative donor properties of lone pair orbitals. Kollman et al (1975) proposed the use of electrostatic potential directly as an empirical index to predict the value of H-bonding interaction energy.

The need for quantum mechanical theories of hydrogen bonding was recognised as early as 1947 when the first theory of this type was published by Sokolov. Three groups of techniques, namely valence-bond (VB) theories, charge-transfer (CT) theories, and SCF-MO and CI theories, are applied to study H-bonded systems. There is no doubt that these theories permit the zero-order, or qualitative, understanding of the phenomena connected with hydrogen bonding. Nevertheless, not all of them are suitable for quantitative calculations. The VB theory is useless in this sense and the CT theory does not lend itself to detailed calculations. The use of the SCF-MO and CI methods is more appropriate. Unfortunately, the systems linked by H-bonds are nearly always too large to be treated in a reasonably complete way. FHF^- and $(H_2O)_2$, which are the simplest systems containing a symmetrical and an asymmetrical H-bond respectively, can be studied completely with the help of the nonempirical SCF-MO and CI methods. Approximate SCF-MO theory can be used to study other characteristic H-bonded systems. Such calculations allow a study of the effects of different functional groups on strength and other properties of a H-bond and also give a quantitative estimation of the OH bond stretch, the νOH frequency shift, and intensity increment on association (Bratož, 1966).

The strength and other properties of H-bonds vary considerably when going from one electronic state of the system to another. As this subject is most easily studied by the help of ultraviolet spectroscopy, one usually expresses the results in terms of frequency shifts, intensity increments, etc. The H-bond is easier to handle theoretically in the exited states of the system than in its ground state. This is due to the fact that a number of well-established π electron theories exist that are applicable to this problem (Bratož, 1966).

Kollman et al (1974) carried out *ab initio* molecular orbital studies (using an STO-3G basis set) on complexes of HCl and HF with a number of proton acceptors. They concluded that in order to predict the infrared spectrum of H-bonded complexes at least semi-quantitatively, a more sophisticated basis set was required.

In a study of H-bonding properties of water using extended Hückel (EHT) and semiempirical LCAO-SCF (CNDO/2) methods by Murthy and Rao (1968a), the enthalpy of formation of hydrogen bond dimers calculated by the CNDO/2 method was in better agreement with the experimental value. Later, based on detailed CNDO/2 calculations on the H-bonds in several donor-acceptor systems of varying strength, the same authers concluded that the results of the dissociation energy and proton potential functions are better than those obtained by the EHT method (Murthy and Rao, 1970).

Murrell and Van Duijneveldt (1967) have used the perturbation method to calculate various contributions to H-bond energy. A perturbation method, where the intermolecular energy is calculated directly, would be preferable to methods where the intermolecular energy is calculated as a difference between the theoretical values of the total energy and the sum of the energies of two separated molecules.

4.4.2. The correlative approach

This approach is based on using various sets of parameters (in particular, H-bond scales) and especially analytical relationships. Development of these empirical calculations of H-bond thermodynamic functions began in the mid 1960's. Two approaches were introduced

in that time: multiplicative-additive and multiplicative (Raevsky et al, 1992). The additive-multiplicative scheme for enthalpy calculations of donor-acceptor interactions was suggested by Drago and coworkers (Drago & Wayland, 1965; Drago et al, 1971; McMillan & Drago, 1972; Drago, 1973; Guidry and Drago ,1973; Drago et al, 1977; Kroeger & Drago, 1981). Drago's equation is presented as follows:

$$-\Delta H = E_A E_B + C_A C_B \qquad (4.6)$$

where E represents an electrostatic energy factor and C a covalent energy factor and subscripts A and B refer to Lewis acid and Lewis base respectively. Then a more sophisticated form was proposed:

$$-\Delta H = e_A e_B + c_A c_B + t_A t_B \qquad (4.7)$$

in which each term corresponds to particular types of intermolecular interactions : e_A e_B, to electrostatic interaction; c_A c_B, to covalent interaction; t_A t_B, to charge transfer interaction. This scheme gives good results for ΔH calculations (Kroeger & Drago, 1981).

The multiplicative approach or 'factor rule' is based on constancy and mutual independence of donor and acceptor functions (factors)of interacting molecules (Jogansen, 1971a; Jogansen, 1971b). According to Jogansen's approach, the enthalpy of H-bond formation is proportional to the product of i-factor of Lewis acid (P_i) and j-factor of Lewis base (E_j) as follows:

$$\Delta H_{ij} = \Delta H_{11} P_i E_j \qquad (4.8)$$

where dimensionless Pi and Ej characterise the relative H-bonding ability of compounds and the coefficient ΔH_{11} is the enthalpy of the standard complex formation between two arbitrary compounds with $P_i = E_j = 1$ (Raevsky et al, 1992).

This equation was rewritten by Raevsky (1987) to the more convenient form:

$$\Delta H_{ij} = |\Delta H_{11}|\ E_i E_j + K_0 \qquad (4.9)$$

where E_i and E_j have opposite signs. This latest equation demonstrates clearly the role of H-bonding partners. It was also shown that for some classes of strong donors and acceptors, the following form of the above equation had to be used:

$$\Delta H_{ij} = |\Delta H_{11}|\ E_i E_j + K_0 \qquad (4.10)$$

As the application of enthalpy was not enough for adequate description of H-bonding complex formation, the multiplicative approach was applied to ΔG values (Raevsky et al, 1989):

$$\Delta G_{ij} = |\Delta G_{11}|\ C_i C_j \qquad (4.11)$$

Here ΔG_{11} is the free energy of complex formation between the standard donor and acceptor. These studies showed that both ΔG and ΔH values can be described quite well on the basis of the multiplicative approach and that there should be two separate scales for ΔG and ΔH.

Using phenol as a standard proton donor and hexamethylphosphoramide as a standard proton acceptor, Raevsky et al (1992), devised the following equations for a data set of 163 proton donors and 195 proton acceptors:

$$\Delta H = 4.96 E_a E_d \qquad (4.12)$$

$$\Delta G = 2.43 C_a C_d + 5.70 \qquad (4.13)$$

Subscripts a and d denote acceptor and donor factors respectively. Both ΔG and ΔH have the units of kJ mol^{-1}.

As there were poor correlations between C_a and E_a and also between C_d and E_d, they concluded that correct accounting for H-bonding in drug design investigations requires acceptance of both enthalpy and free energy contributions.

4.4.3. Experimental approach

Since the introduction of the original concept of hydrogen bonding, a great deal of experimental effort has been directed towards understanding the energies of H-bond formation. With the development of solution calorimetric techniques, it is now possible to test the relationship between accurately measured thermodynamic functions and spectroscopic shifts. The more reliable calorimetric enthalpies strongly support the validity of such linear relationships within a given class of donor molecules.

Gas phase enthalpies and spectroscopic shifts for a wide variety of H-bonding acids and bases would be most beneficial in extracting the important contributions to the strength of a H-bond. Unfortunately limitations in volatility of reactants and products, laborious experimental procedures involving gas pressure changes and large experimental uncertainties in gas phase spectroscopic techniques have resulted in only isolated studies of a few acid-base systems (Sherry, 1976). Various workers have shown, however, that solution enthalpies measured in non-coordinating, inert solvents, such as carbon

tetrachloride and the aliphatic hydrocarbons, approximate the gas phase values (Eplay and Drago, 1967).

The common calorimetric procedure involves injection of a small volume of an acid as a neat liquid or as a concentrated solution into a solution of the base and the measured heat is corrected for the heat of solution or dilution of the acid. Other widely used experimental methods are infrared and nuclear magnetic resonance techniques.

Several linear enthalpy-spectroscopic shift relationships have been observed using calorimetrically determined enthalpies (Sherry, 1976). There are also inverse correlations between the heats of formation of H-bonds ($-\Delta H^{\circ}_{HB}$) and the difference between the proton affinity of the O^{-} anion and the proton affinities of the bases (ΔPA) (Zeegers-Huyskens, 1986b).

In the solvent tetrachloromethane, a plot of enthalpy against Gibbs energy of H-bonding complexation of some substituted phenols with N-methyl pyrrolidinone was linear with a positive slope (Abraham et al, 1986) exactly as was observed for numerous series of phenols against various bases in non-polar solvents such as benzene, cyclohexane or tetrachloromethane (Murthy & Rao, 1968b; Joesten & Schaad, 1974) and also gas phase complexation of carboxylic acids with iodide ion (Caldwell & Kebarle, 1984). However, in the case of 1,1,1-trichloroethane (TCE) the corresponding plot of enthalpy against Gibbs energy had a smaller slope that was negative (Abraham et al, 1988b). It was suggested that the involvement of the dipolar solvent TCE in the complexation reaction was the reason for this observation and because solute-solvent interactions could lead to

significant effects on enthalpies and entropies of complexation but not on Gibbs energies of complexation, the latter parameter would be the most useful one to use in any construction of a scale of solute H-bond acidity and basicity (Abraham et al, 1988b).

4.5. Hydrogen bonding parameters

H-bonding is an important interaction in the control of drug activity by its effect on the solubility and partitioning of drugs and on drug receptor interactions, Hence it is appropriate to incorporate its parameters into quantitative structure-activity relationships (QSARs), which are mathematical equations relating biological activity to physico-chemical and structural parameters (Dearden, 1990). Consequently, there have been a number of attempts to quantify H-bonding ability of compounds.

The first attempt to devise a hydrogen bonding parameter was made by Seiler (1974), who used the differences between octanol-water and cyclohexane-water log *P* values to develop group contributors (I_H) to hydrogen bonding. Clearly, Seiler's I_H values cannot distinguish between proton-donor and proton-acceptor. Seiler reported I_H values for 21 substituents.

A problem with this and most other substituent-based approaches is that the hydrogen bonding ability of a given substituent is not independent of the reminder of the molecule. Thus a whole-molecule approach is to be preferred, but this necessitates knowledge of two experimentally measured log *P* values for each compound studied. This approach was recently used by Young et al (1988) who correlated blood-brain barrier penetration with

the difference between octanol-water and cyclohexane-water log *P* values.

A further drawback of the Seiler approach is that the difference between octanol-water and cyclohexane-water log *P* values must be a function of polarity as well as hydrogen bonding, and hence I_H values can not be regarded as pure hydrogen bonding parameters.

Moriguchi (1975) assumed that log *P* contained volume and polarity components, and calculated the polarity component E_w as the difference in octanol-water log *P* value of a polar compound and that of a non-polar compound of the same molecular volume (Moriguchi actually used parachor as his volume term). Although Moriguchi called E_w a polarity term, he showed that it correlated well with hydrogen bond strength. We now know that log *P* can be factored into volume, polarity and hydrogen bonding terms (4), so that E_w must contain both polarity and hydrogen bonding terms.

Allen (1975) has proposed an empirical formula for H-bonding energy, based upon analysis of molecular orbital calculations and experiment, which is:

$$\Delta E = K\mu_{AH} \times \Delta I_B/R \qquad (4.14)$$

ΔI_B is the ionisation potential of the hydrogen atom acceptor measured relative to that of the isoelectronic inert gas atom. Calculations show that systems like FH···Ne have binding energies that can be explained solely by dispersion forces. μ_{AH} is the dipole moment of the AH bond. *R* is the internuclear distance between atoms A and B, and *K* is a constant (units charge^{-1}).

Fujita et al (1977) devised a hydrogen bonding indicator variable which simply took the

value of unity if a molecule or substituent was capable of forming a hydrogen bond, and of zero if it was incapable of doing so. This method can be adapted to distinguish between proton-donor and proton-acceptor ability. Because of its simplicity it has been to date the most widely-used hydrogen bonding descriptor.

Charton and Charton (1982) modified the Fujita approach by using the number of hydrogen bonds that a molecule or substituent was capable of forming. Thus $-NH_2$ would score 2 as a proton-donor and 1 as a proton-acceptor, whilst -OH would score 1 as a proton-donor and 2 as a proton-acceptor, since there are two lone pairs of electrons on the oxygen atom.

Yang et al (1986) devised two hydrogen bonding parameters; HB_1 is very similar to the Charton and Charton parameter, save that -OH is regarded as accepting only one hydrogen bond, and certain groups (e.g. OCF_3) are treated as non-hydrogen bond acceptor. For hydrogen bond donors, however, the values of HB_1 are identical to the Charton and Charton parameter. HB_2 is calculated by taking average enthalpy values for each type of hydrogen bond (e.g. OH···O), multiplying by the number of such bonds and scaling by 0.1. For example, HB_2 for -OH is calculated by taking ΔH for O-H···O as 6.05 kcal mol^{-1}, multiplying by 2 (since the -OH group can act as both proton-donor and proton-acceptor), and scaling by 0.1: $HB_2 = 6.05 \times 2 \times 0.1 = 1.21$. Yang et al reported HB_1 and HB_2 values for 144 substituents.

The pk_{HB} scale, which is the logarithm of the acceptor equilibrium constant for H-bond formation with p-fluorophenol as standard donor in the solvent tetrachloromethane at

25°C, was developed by Taft et al (1969). They used an n.m.r. methodology to determine pk_{HB} values.

Taft and Kamlet applied solvatochromic methodology to devise a solvent H-bonding donor parameter α (Taft & Kamlet, 1976; Taft & Kamlet, 1979) and a solvent H-bonding acceptor parameter β (Kamlet & Taft, 1976; Taft et al, 1982). Their method was based on the measurement of u.v. shifts for probe and reference molecules, relative to an inert standard solvent, in a series of neat organic liquids. These solvent parameters, unfortunately, cannot predict the behaviour of the compounds when one-to-one contact is involved (Kamlet et al, 1982) and their methodology can deal only with solvents. This limitation excludes nearly all compounds of direct interest to the medicinal chemist and important classes of functional groups. For compounds that are capable of self-association, α_m and β_m are corresponding monomer H-bond acidities and basicities determined by solvatochromic methods; it is assumed that for compounds that are not capable of self-association, $\alpha_m = \alpha$ and $\beta_m = \beta$ (Abboud et al, 1985; Kamlet et al, 1986).

Abraham et al (1989a) measured H-bonding equilibrium constants for a large and varied selection of proton donors against a standard proton acceptor (N-methylpyrrolidinone) and of proton acceptors against a common donor (4-nitrophenol) in the solvent 1,1,1-trichloroethane. They were used to create the $\log K_\alpha$ and $\log K_\beta$ scales of proton donor and acceptor ability which are explicitly targeted to the needs of the medicinal chemist in the context of potential drug-receptor interactions.

Scales of solute H-bond acidity (Abraham et al, 1988a; Abraham et al, 1989b) and solute

H-bond basicity (Abraham et al, 1989c; Abraham et al, 1990) have been constructed using equilibrium constants (as log$K$ values) for complexation of a series of acids against a given reference base and a series of bases against a given reference acid in dilute solution in tetrachloromethane. A system of forty-five linear equations was constructed for acidity parameter:

$$\log K^{i} = L_B \log K^{Hi}_{A} + D_B \qquad (4.15)$$

$\log K^{i}$ refers to log K values for a series of acids against a given reference base. L_B and D_B characterise the base and log K^{H}_{A} characterises the acid. These equations were solved using the observation that all the lines in the equation intersect at a given point where log K = log K^{H}_{A} = -1.1 with K on the molar scale, and some log K^{H}_{A} values were obtained. These values were transformed into α^{H}_{2} values through the equation:

$$\alpha^{H}_{2} = (\log K^{H}_{A} + 1.1)/4.636 \qquad (4.16)$$

Similar to the construction of H-bond acidity scale, for development of the basicity parameter, thirty-four linear equations of log K values for the complexation of different bases (i) against 34 reference acids were assembled as:

$$\log K^{i} = L_A \log K^{Hi}_{B} + D^{A} \qquad (4.17)$$

The equations were solved to yield L_A and D_A values that characterise the acids and log K^{H}_{B} values that characterise the base; all the 34 equations intersect at a point where log K = -1.1 with K on the molar scale. The log K^{H}_{B} values were transformed into a more convenient scale through:

$$\beta_2^H = (\log K_B^H + 1.1)/4.636 \qquad (4.18)$$

α^H_2 and β^H_2 values, that refer specifically to solute H-bond complexation at 298 K in CCl_4, can be combined in a general equation that can be used to predict a vast number of hitherto unknown log *K* values (Abraham et al, 1988d):

$$\log K = m\alpha_2 . \beta_2 + c \qquad (4.19)$$

where *m* and *c* may depend on the solvent (and also on the standard state) but are independent of acid and base.

As α^H_2 and β^H_2 values referred to 1:1 complexation, it was by no means obvious that such values were relevant to the solvation situation in which a solute was surrounded by solvent molecules. Therefore Abraham et al (1991b) set up a number of multiple regression equations for general solvation and "back-calculated" the *effective* or *summation* solute H-bond parameters ($\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$). In most of the cases $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ values followed closely the original H-bond α^H_2 and β^H_2 values but there were exceptions (in general multifunctional solutes were exceptions (Abraham, 1993)).

The strength of H-bond from rotational spectroscopy (in gas phase) can be measured by the quadratic force constant K_σ associated with the H-bond stretching mode ν_σ which gives a measure of the restoring force per unit infinitesimal extension of the H-bond. k_σ has been determined for a wide variety of H-bond complexes and a comparison of K_σ within a series of the complexes gives a measure of the relative strength of the H-bond along the series. In fact, for H-bond complexes that are not too strongly bound (where

the interaction between the components can be described without invoking significant charge redistribution), k_σ can be expressed by the empirical equation:

$$k_\sigma = cEN \qquad (4.20)$$

where E and N are numbers associated with the molecules HX and B, respectively and c is a constant having the value 0.25 Nm^{-1}. In the electrostatic model for H-bonding, the quantities E and N have been called "limiting gas-phase electrophilicities and nucleophilicities" respectively (Legon & Millen, 1987).

5. Quantitative Structure-Activity Relationships (QSAR)

The relationship between chemical structure and biological activity has drawn the attention of many investigators since the late nineteenth century. On the basis of the assumption that drugs with similar structure will have similar biological responses, QSAR is an attempt to rationalise and quantify the relationships between the biological activity of chemicals and their physicochemical properties. This assumes that measured physicochemical properties contain information about the structure of the compound and that this information can be used to explain biological effects.

One of the main tenets of QSAR is that all the compounds used in a study should exert their biological effect by the same mechanism, otherwise poor correlations will be observed. Since it is usually extremely difficult to determine precise mechanisms of action, the assumption is usually made that members of a congeneric series act by the same mechanism, and hence QSAR studies are usually confined to congeneric series.

The first workers systematically to study relationships between chemical structure and biological activity were Meyer (1899) and Overton (1897) who, independently, showed that narcotic potency of general anaesthetics on tadpole was proportional to the distribution coefficients of the compounds between water and olive oil. Ferguson (1939) formulated a concept linking narcotic activity, partition coefficient and thermodynamics. Ferguson declared that, when in a state of equilibrium, simple thermodynamic principles could be applied to drug activities, and so the important parameter to consider for the correlation of narcotic activities was the relative saturation of the substance in the applied phase. This has become known as Ferguson's principle.

5.1. The Hansch approach

There was no significant work in the area until the work of Hansch et al (1962) on the relationship between structure and plant growth regulating activity of phenoxyacetic acids was published. They proposed that the activity was a function of the Hammett σ-constant (Hammett, 1940) and a new substituent constant π. This is shown in equation 5.1.

$$\log 1/C = a + b\sigma_k + c\pi_k \qquad (5.1)$$

The term log 1/C is the biological activity, where C is the molar concentration required to give a standard, predetermined response from the biological system.

The Hammett σ-constant is a substituent constant derived from the ionisation constant , K_H, of benzoic acid and that of an appropriate benzoic acid derivative, denoted by k, as shown in equation 5.2.

$$\log K_k - \log K_H = \rho\sigma_k \qquad (5.2)$$

Substituents with $\sigma_k>0$ are electron withdrawing while those with $\sigma_k<0$ are electron-donating. Hammett found that this effect of substitution on benzoic acid ionisation could be extended to a large number of organic reactions through what were termed linear free energy relationships, LFERs.

π_k is a substituent constant analogous to σ_k and is related to the 1-octanol/water partition coefficient, P, for a compound by equation 5.3. Hansch suggested that 1-octanol be used as a model for lipoidal phases in the biological system. Other solvents, such as hexane, chloroform and ether, have been used, but 1-octanol is the solvent of choice (Clark & Moos, 1990).

$$\log P_k - \log P_H = \pi_k \qquad (5.3)$$

Substituents with $\pi_k>0$ are said to be lipophilic (relative to the substituent H) and those with $\pi_k<0$ are hydrophilic.

It was found necessary sometimes to add a second-order term in π to equation 5.2 to give satisfactory agreement between observed and predicted values of activity (Hansch, 1963). This leads to equation 5.4, which has been termed the Hansch model for QSAR.

$$\log 1/C = a + b\sigma_k + c\pi_k + d\pi_k^2 \qquad (5.4)$$

The theoretical origin of the second-order term has been shown to result from a mechanism in which there is differential transport to the active site, and this transport is related to the relative lipophilicity of the membranes of the series (Penniston et al, 1969): compounds of low partition coefficient do not partition well into lipid membranes, and thus reach the site of action only at a low rate; on the other hand compounds of high partition coefficient, whilst partitioning well into lipid membranes, do not partition well from there to the next aqueous compartment and also reach the site of action at a low rate. Compounds of intermediate partition coefficient, being able to partition reasonably well both into and out of lipid membranes, are thus more active.

There have been several proposals made to improve the modelling of structure with activity which are essentially alternatives to the Hansch model, and the two most notable of these are by Martin (1978) and Kubinyi (1977). Martin's model is essentially hyperbolic whereas Kubinyi's model is termed bilinear.

Unger and Hansch (1973) proposed the criteria described below which must be

considered before one identifies a best correlation equation for a set of congeners.

a. Selection of independent variables. The widest possible number of independent variables must be examined. The parameters selected should be essentially independent of each other as an aid in rationalisation of the mechanism of drug actions.

b. Justification of the choice of independent variables. In the best correlation equations, each term must validated by an appropriate statistical procedure. It is advantageous to examine regression analyses with all possible combinations of independent variables and then to use a forward selection procedure with sequential *F* tests to identify the best equation, generally that with the lowest standard deviation and all terms significant (usually over 95% level). (Nowadays statistical procedures such as stepwise regression is used which have eliminated the need for this).

c. Principle of parsimony. All things being equal, one should accept the simplest model.

d. Number of independent variable terms. According to the suggestion of Topliss and Costello (1972) one should have at least five to six data points per variable in order to minimise the risk of chance correlations.

e. Physical organic significance. The best correlation equation should be rationalised in terms of the principles of known physical organic and biomedicinal chemistry (Fujita, 1990).

5.2. The biological activity data

In order to find a good QSAR correlation, it is essential that the biological data are as accurate as possible, and of a consistent form; for example , a response to one concentration cannot be compared with a response to a different concentration.

The biological activities for the compounds under study should come from dose-response data. This is not always possible, and relative response data from a single treatment of the same concentration are frequently reported for each member of a series. This is much less expensive than using dose-response data but can introduce error into the data if the measurement to which the data are normalised contains error. If classification methods are to be applied to the data, only binary data (e.g. active/inactive) are necessary (Dunn, 1990).

5.3. The chemical descriptor data

For a compound to trigger a biological response when administered to a living organism, a number of processes must occur. Firstly the compound must dissolve in a body fluid, if it is not already in solution. Secondly it must be transported from the site of administration to the site of action (receptor site). Thirdly it must bind to the receptor, often in a quite specific manner, in order to initiate the biological response. Each of these processes depends on certain physico-chemical properties of the compound. Dissolution is related to the solubility of the compound, whilst transport is governed largely by its partitioning behaviour between lipid and aqueous phases. Binding of the compound to the receptor will depend upon the forces of interaction between the two, and upon the complementarity of size and shape between the two (Dearden, 1990).

Thus the biological response to a xenobiotic can be considered to be controlled by three broad classes of physico-chemical property, hydrophobic, electronic and steric. In general, two types of descriptor can be used. The first are descriptors which are derived from a consideration of the total structure of the compound, such as log *P*, boiling point, molecular weight, etc. If the compounds in the series are analogues of

a parent structure, substituent constants such as Hammett σ, Hansch π, group molar refractivity MR, etc., can be used.

5.3.1. Hydrophobic parameters

It has been demonstrated that the partitioning behaviour of a compound is the one factor above all others that controls the ability of a xenobiotic to produce a biological response in an organism. Since QSARs are free energy (ΔG) relationships, the common logarithm of the partition coefficient is used in correlation analysis.

Log *P*: Partition coefficient is the ratio of concentrations at equilibrium of a solute distributed between two immiscible liquid phases; the concentration in the more hydrophobic phase is, by convention, the numerator. Apparent partition coefficient or distribution coefficient (D) applies to the ratio of total concentrations, including associated and ionised species.

Hansch substituent constant π: Most of the work in QSARs has been based on the substituent constants quoted in a dimensionless form π. Values can be used to calculate 1-octanol/water partition coefficients in the same way as Hammett constants can be used to estimate dissociation constants.

Fujita et al (1964), in a more extensive study of π, showed that the π value of a given aromatic substituent varied somewhat with the nature of other substituents. These variations indicate that π values are not strictly additive, being dependent on the nature of the remainder of the molecule. Dunn et al (1983) concluded that the lack of additivity was due almost entirely to hydrogen bonding. It is moreover a fact that reasonably good estimates can be made of partition coefficients using published π

values for simple molecules, provided that only a few π values are involved (Hansch and Leo, 1979).

Fragmentation constants f: There is a fundamental flaw in the π value approach to the calculation of log *P* that it incorrectly assumes the log P value of hydrogen to be zero. Thus if many π values are summed then the calculated log P value will be appreciably in error. Nys and Rekker (1974) devised an alternative approach to the calculation of log *P* by factoring observed log *P* values of large number of compounds to give hydrophobic fragmental constants, f values. It assumes additivity, but attempts to take account of constitutive effects by introducing correction factors. Leo and co-workers (1975), in a fragmental approach, determined the partition coefficients of a number of small molecules, including hydrogen, from which they were able to obtain fragmental constant values. They also found it necessary to introduce correlation factors, for such things as chain branching, fragments attached to aromatic rings and numbers of bonds between fragments. Both fragmental methods have been computerised.

Chromatographic R_m values: In order to avoid practical difficulties often presented by the direct determination of the partition coefficient, chromatographic parameters related to partition coefficient have been used in some QSAR studies (Biagi et al, 1991). R_f value of chromatography is related to partition coefficient through:

$$P \propto [(1/R_f) - 1] \qquad (5.5)$$

Bate-Smith and Westall (1950) defined a parameter R_m as:

$$R_m = \log[(1/R_f) - 1] \qquad (5.6)$$

R_m values have been used as a substitute for partition coefficients in QSAR investigations.

HPLC capacity factor: Log P is directly proportional to log K, where K is the capacity factor of the column as defined in equation 5.7.

$$K = \frac{t_x - t_0}{t_0} \qquad (5.7)$$

Where t_x is the time taken for a specific compound to elute from the column, and t_0 is retention time of a non-retained compound. This parameter has also been used in QSAR studies (Barbato et al, 1991).

Solubility: Aqueous solubility is clearly a measure of hydrophilicity, and thus an inverse relationship is to be expected between solubility and partition coefficient (Hansch et al, 1968).

5.3.2. Electronic parameters

Intermolecular interaction forces control both the extent and strength of drug-receptor interactions. They are of several types including ion-ion, ion-dipole, ion-induced dipole, dipole-dipole, dipole-induced dipole, instantaneous dipole-induced dipole and hydrogen bonding. All of them depend on the electron distribution of a molecule or substituent, and polarisability. Electronic forces also determine to a large extent the rate of metabolism of a compound, since such forces affect bond order (Dearden, 1990).

Hammett constant σ: The method of calculation of σ has been explained earlier in this chapter. Hammett's and equivalent equations are said to be linear free energy relationships (LFER); this constant is related to free energy because equilibrium constants (K) are logarithmically related to free energy (ΔG) through the van't Hoff equation 5.8 in which *R* is the gas constant and *T* is temperature.

$$\Delta G = -2.303RT \log K \qquad (5.8)$$

The free energy of a transition involving a given molecule is assumed to be the sum of the free energies of its substituent groups. σ is therefore also additive.

Hammett obtained σ values for meta- and para-substituents in an aromatic ring; consistent values could not be obtained for ortho-substituents. σ_m and σ_p values differ from each other because of the differing inductive and resonance contributions in the two positions.

Hammett substituent constants can be used only for nuclear aromatic substituents and their effects upon the side-chain groups. Taft (1956) devised a set of electronic substituent constants σ^* for aliphatic substituents; these were obtained experimentally from ester hydrolysis rate constants.

Several attempts have been made to factor σ to its resonance and inductive components. Swain and Lupton's F (field effect) and R (resonance effect) (Swain et al, 1983) values and Charton's three parameters of σ_I (inductive substituent constant), σ_d (a resonance effect term) and σ_e (sensitivity of the substituent to change in electronic demand by the active site) (Charton, 1987), are examples of such factors.

Molar refractivity MR: MR models volume and it has been used in many QSAR correlations as a volume term. However it is also proportional to electron polarisability (the ability of electrons to be polarised in the presence of an electric field). MR is thus clearly an electronic as well as a steric property. Grieco et al (1978) pointed out that since MR models both steric and electronic effects, it is difficult to interpret the MR term in a QSAR correlation. Abraham et al (1990c) devised "excess molar refractivity" (ΔMR) as the difference between the experimental MR value and

that for an alkane having the same V_x (since there is an excellent correlation between MR values of alkanes and their V_x values, but other compounds have higher MR values and are outliers from this line). It is possible that this is a better measure of polarisability than is MR itself (Dearden et al, 1991).

Hydrogen bonding: Hydrogen bonding parameters have been discussed in detail in chapter 4. Some examples of QSARs containing H-bonding parameters will be given in section 5.5.

pK_a: Hammett substituent constants correlate closely with pK_a. pK_a also controls the extent of ionisation of a compound and thus affects the apparent partition coefficient (D). Therefore this parameter has been used in QSARs in both roles.

Dipole moment μ: Dipole moment might be said to be a measure of hydrophilicity, since it is a virtual prerequisite for aqueous solubility. It has been used in QSAR extensively as experimentally determined values. However, gas-phase dipole moments can be calculated using molecular orbital theory.

Solvatochromic parameters: These parameters were initially derived from solvent effects on electronic spectra. Kamlet and Taft and their co-workers developed these parameters using the assumption that solubility in a given solvent is controlled by three factors- the solute size, a dipolar/polarisability term (π^*), and hydrogen bond donor (α) and acceptor (β) terms. The first is taken to be molar volume (V)/100 (the molecular weight divided by the liquid density at 20°C), and the others are derived by various spectroscopic and/or chromatographic techniques (Kamlet et al, 1986a). These parameters have been successfully used in a series of equations termed linear solvation energy relationships (LSER) which have been explained in chapter 4 and are

widely used in QSAR.

NMR chemical shifts: NMR shift is a sensitive indicator of a local electronic effect within a molecule, and can be used to probe individual atomic or group interactions. Chemical shifts have been correlated with various interaction forces including H-bonding (Pimentel & McClellan, 1960), and have been used in some QSAR studies (Koehler et al, 1988).

Quantum chemical parameters: Since all properties of a molecule are related to its electron distribution, it is not surprising that properties obtained through quantum mechanical calculations have been used in QSAR. These parameters can be obtained relatively easily using different quantum mechanical methods (explained in chapter 1) which are available in computer software packages. A necessity for these calculations is finding the best conformation for the molecule, as quantum mechanical properties are sensitive to the conformation used. The assumption is usually made that the most probable conformation is that corresponding to the global energy minimum, and so minimisation must be carried out prior to property calculation. Minimisations can be performed using molecular mechanics or more sophisticated quantum mechanics computer programs. The minimisations (and also property calculations) are normally carried out on the isolated gas-phase molecule, which is hardly realistic. However introduction of solvents into the situation greatly increases the complexity of the calculations, and gas-phase minimisations seem to give acceptable conformations that often agree closely with those determined experimentally (Richards, 1983).

The most extensively used quantum mechanical properties in QSAR comprise atomic charge, frontier electron density, HOMO and LUMO energies, superdelocalisability,

dipole moment, and electrostatic potentials. The most usual way of using atomic charges in QSAR is to take as a parameter the charge on a particular atom. This may be an atom common to the whole set of molecules being examined or may be an atom of a substituent group. An alternative is to sum the modulus of atomic charges over the whole or part of the molecule, to yield a measure of the polar interaction of which the molecule is capable (Dearden et al, 1989). Another approach is to use the difference of charge (or its modulus) across a given bond, perhaps in a common functional group (Dearden & Nicholson, 1986).

Frontier electron density, being related to the frontier orbital, is useful when dealing with very localised interactions. Values relating to both the highest occupied and the lowest unoccupied molecular orbital can be calculated (Kier, 1971). The frontier electron theory was originally developed to explain the difference in reactivity at each position in an aromatic hydrocarbon. It is based on the intuitive idea that the reaction should occur at the position of the greatest density of the electrons in the frontier orbitals; HOMO in an electrophilic reaction, LUMO in a nucleophilic reaction and both of these in a radical reaction (Richards, 1983). The frontier electron density strictly permits only a comparison of reactivities at different positions within the same molecule. In order to extend this concept for use over a series of molecules, a further quantity, F, may be considered as a weighted frontier orbital:

$$F = f_r/\varepsilon \qquad (5.9)$$

where f_r is the frontier electron density, and ε is the energy of the appropriate frontier orbital.

HOMO and LUMO energies are the energy of the highest occupied and of the lowest unoccupied molecular orbitals respectively. The former represents the ease with which an electron can be donated by the molecule and is thus related to the ionisation

potential; the latter is a measure of the ease with which a molecule will accept an electron (and therefore related to electron affinity). Both terms clearly can model intermolecular interactions as well as reactivities; in particular they model charge transfer interactions well (Murrell et al, 1985).

Superdelocalisability is defined as the sum of the frontier electron densities on an atom divided by the HOMO or LUMO energy. Both nucleophilic (HOMO) and electrophilic (LUMO) superdelocalisability values can be calculated, as can the third type relating to free radical attack (Fukui et al, 1954).

Electrostatic potentials have been explained in details in chapter 3.

5.3.3. Steric parameters

The size and the shape of molecules and substituents are important in biological activity. A bulky substituent may shield a polar group, thereby reducing a molecule's affinity for water and/or increasing its affinity for a lipoidal phase. Size may be a barrier to the passage of molecules through aqueous channels in membranes, and size and shape are often extremely important in drug-receptor binding. Generally drug and receptor should be complementary in shape in order to exert a biological response.

Since size is an additive property, most steric parameters can be used as either substituent or whole molecule values. It is necessary to distinguish between parameters that model size or bulk alone, and those that contain shape information.

a. Bulk parameters:

Molecular weight: Molecular weight (MW) is the simplest measure of size and for that reason has been widely used in QSAR. It is often observed to be collinear with log *P*, but this is only because many high MW compounds are also very hydrophobic (Lien & Wang, 1980).

Molecular volume: Molar volume (MV) is defined as molecular weight /density (MW/ρ). Experimental determination of MV is a tedious procedure. Molecular volume can be calculated using the van der Waals radii of atoms for a substituent or the whole molecule. Another method of calculation is to use a computer program that rolls a probe sphere over the molecular surface defined by van der Waals radii, to give a cavity surface volume (Connolly, 1983).

Surface area: van der Waals radii can be used for summation of atomic surface areas to calculate approximate surface area. A better approach is calculating solvent accessible surface area by rolling a probe sphere over the van der Waals surface of the molecule (Connolly, 1985). The latter has been widely used in QSAR studies.

Taft's steric substituent constant E_s: Its calculation depends on the fact that acid hydrolysis of esters RCOOR′ is determined almost completely by steric factors and is defined by the following equation (Taft, 1956):

$$E_s = \log\ (k_R/k_{Me}) \qquad (5.10)$$

The use of this parameter in QSAR is limited by the experimental difficulties in obtaining the physicochemical data upon which E_s values are based.

Van der Waals dimensions: van der Waals volume (V_W) and radius (r_V) represent the

actual dimensions of the group. Since chemical groups are rarely symmetrical, the van der Waals radius depends on the axis along which it is measured, and three types are defined, $r_{v(min)}$, the minimum radius, $r_{v(max)}$, the maximum radius and $r_v\|$, which is the distance the group produces from the bulk of the parent molecule. Sometimes the mean of the three radii ($r_{v(av)}$) is used (James, 1988).

Charton's steric constants: Charton (1983) introduced a corrected van der Waals radius U, in which the minimum van der Waals radius of the substituent ($r_{v(min)}$) is corrected for the corresponding radius for hydrogen:

$$U = r_{v(min)} - r_{vH} \qquad (5.11)$$

Molar Refractivity MR: This has been discussed in the section on electronic parameters. It is very widely used in QSAR correlations, usually as a volume term.

The parachor: The parachor is molar volume (MV) which has been corrected for forces of intermolecular attraction by multiplying by the fourth root of surface tension. It has fallen out of favour recently as a QSAR parameter.

b. Shape parameters:

Sterimol parameters: Verloop (Verloop et al, 1976) described a new set of steric parameters which defined the dimensions of a substituent in five directions: L, which is the length along the main axis of the energy-minimised molecule or substituent, and B_1-B_4, which gives the widths of the group in four directions, 90° to each other and perpendicular to the L axis. Cross sectional dimensions increase from B_1 to B_4. A further parameter, B_5, has been developed more recently and represents the maximum width of the substituent. The parameters have received very wide usage in QSAR, and have often proved very effective.

The kappa index: This index is derived from molecular connectivity theory. The index is based on the count of 2-bond fragments in a hydrogen-suppressed graph relative to the maximum number possible in the isomeric star graph (i.e. with maximal branching) and the minimum number in the isomeric linear graph (i.e. no branching). The index is normalised to the number of atoms in each molecule (Kier, 1985).

Minimal steric difference (MSD): This parameter assesses the difference between molecules in terms of the parts which do not overlap when one energy-minimised molecule is placed on top of the other. The MSD is defined as the number of unsuperimposable atoms when a molecule is superimposed atom by atom upon a standard molecule that is presumed to be close to an ideal fit to its receptor (Simon et al, 1984).

Shape similarity index: Molecular similarity provides a quantitative measure of one molecule looking rather like another. It is expressed as an index with a range from zero to unity, which represents identity, and can be used in QSAR studies. The index can be obtained for similarity of shape as well as for similarity of electrostatic potentials and electrostatic field. As introduced by Meyer and Richards (1991), the two molecules being compared are superimposed in a 3D grid with points inside the van der Waals surface being assigned the value unity, and outside the value zero. It is possible to avoid the problems of numerical integration by using gaussian functions to represent shape (Good et al, 1992). Despite its simplicity there are a number of problems in this approach. One arises when investigating the compounds binding into a receptor site; then it may only be one side of the molecules which is relevant and the whole molecule similarity will not be appropriate. The other difficulty is to decide the conformation which is to be used, especially when dealing with flexible molecules.

3D QSAR, CoMFA: The basic assumption in the original CoMFA methodology is that a suitable sampling of the steric and electrostatic fields surrounding a set of drug molecules might provide all the information necessary for the understanding their observed biological properties (Cramer et al, 1988). CoMFA electrostatic and steric descriptors have been investigated for their applicability to describe the corresponding physico-chemical parameters used in traditional QSAR as well as to substantiate their use in 3D QSAR (Kim, 1991 & 1992a-c). A CoMFA analysis consists of the following steps: establishing the conformation of each molecule, superimposing the molecules, calculating for each the interaction energies with suitable probes at many points in a lattice, performing a statistical analysis of the relationship between the interaction energies and the property of interest, and displaying the 3D QSAR coefficient contour map.

5.4. Data analytic methods for QSAR studies

Once biological data and chemical descriptor data are available, the next step is exploring the relationship between the two data blocks. Selecting the appropriate method is a function of the type of information required and also the nature of the data which are to be analysed. Multiple regression analysis is usually used to correlate a given biological activity with molecular parameters. If one selects a few parameters for correlation, it is a relatively easy matter to decide which combination of them gives the best correlation: The standard procedure is to rely on the correlation coefficient and the standard error. But this is not adequate, since the inclusion of any additional parameter will raise the correlation coefficient. Therefore, it is recommended to include the variance ratio, which will fall if a non-significant parameter is included in the correlation. Standard error of a coefficient can also indicate its significance; it should be considerably smaller than its coefficient. In a

multiple regression, the confidence in a parameter can be assessed by dividing the coefficient by the standard error. The resulting ratio (*t*-ratio) can then be compared with the limiting Student's *t* value for a specific probability level and degrees of freedom.

The most significant result from a QSAR study is a predictive model. This makes model validation an important part of QSAR research. The predictability of a model should be tested on compounds that were not used in its derivation as is done with the jack-knife methods. It is best to select a test set from the set of active congeners prior to model development and use these compounds for model validation. A simpler procedure is to use the technique known as cross-validation, in which one compound is removed from the data set, the QSAR is developed from the remaining compounds and is used to predict the activity of the one that was left out; the procedure is then repeated until each compound has been left out in turn. A cross-validated correlation coefficient is obtained which is a much better measure of the predictive ability of the QSAR, although Wold (1991) notes that there are still some precautions to be observed.

The availability of molecular and quantum mechanical computer software to generate very large numbers of parameters for each compound studied. One then needs to select from among these the parameters that will best model the biological activity; this is usually done by the use of step-wise multiple regression or best sub-sets regression.

If the number of columns (parameters) exceeds the number of rows (compounds), techniques such as multiple regression analysis may not be used. Even where there are more rows than columns, care must be exercised if chance correlations are to be

avoided (Stouch & Jurs, 1986). Data reduction in so-called "over-square" matrices also reduces computer time.

There are several methods to carry out data reduction which are called multivariate analysis. Principal components analysis is a multivariate technique which is concerned with relationships between parameters, and attempts to combine them to form a lesser number of independent variables which describe the system as adequately as the original parameters. The new variables, which are called principal components (PCs), are orthogonal (independent of one another), and are combinations of the old values. The PCs themselves have no physical significance, but can be correlated with the original variables to see which they best represent.

Cluster analysis is used to classify physico-chemical properties into groups according to similarity of properties. A correlation matrix operation is often employed for this purpose; those pairs of variables having correlation coefficients in excess of a predecided value are considered to be related. The threshold chosen depends on the probability level required; a correlation coefficient ≥ 0.7 is a frequently chosen cut-off.

A similar method to principal component analysis is factor analysis, in which the components are rotated in multi-dimensional space in order to aid interpretation. Factor analysis is used on matrices in which the value of an element is influenced by more than one factor. For example, the value can be dependent on a factor which is specific to the value and a factor which is a function of the row to which the value belongs. This could occur if, for example, a range of topical preparations were submitted to a panel for subjective grading, and if there were a bias by the panel members towards the first samples examined. The values would then be the sum of two effects, $f+\varepsilon$, in which f is a factor characteristic of the order in which the samples

were examined and ε is specific to the sample (James, 1988).

Partial least squares regression is a refinement of principal components analysis. An advantage of PLS regression is that it can be used in the case in which there are more independent variables than compounds. The ability of PLS to handle both multivariate activity and structural descriptor data makes it the method of choice in QSAR when biological data are obtained in several biological test systems. This technique carries out the formation of principal components and the multiple regression in a single step, and is designed to give maximal correlation between the PCs and the dependent variable (Dunn et al, 1984).

5.5. Importance of H-bonding parameters in QSAR studies

H-bonding is an important interaction in most physico-chemical aspects of biological activity, affecting such processes as solubility, partitioning and receptor binding. It would therefore be expected to appear as a significant parameter in numerous QSAR correlations, and this is in fact the case. The following are some examples of incorporation of H-bonding parameters in QSAR equations:

1. Growth inhibition of Tetrahymena pyriformis by phenols (Schultz et al, 1987):

$\log (1/IC_{50}) = 0.685 \log P + 0.944 F + 0.337 HB_d - 1.376$

$n = 29 \quad r = 0.954 \quad s = 0.045$

HB_d = indicator variable for H-bond donor ability

2. Anticonvulsant activity of benzyl N,N-dimethylcarbamates in mice (Yamagami et al, 1982):

$-\log ED_{50} = 0.761 \log P - 0.209 (\log p)^2 - 0.316 \sigma^o - 0.179 HB + 2.952$

n = 18 s = 0.099 r = 0.951

HB = indicator variable for H-bond acceptor ability in substituents

3. Antifungal activity of N-substituted phenylsuccinimides (Takayama & Fujinami, 1979):

$pI_{50} = 0.723 \Sigma\pi_{3,5} + 1.464 \Sigma\sigma^o + 0.894 \Sigma E_s^{2,6} + 0.671 E_s^m + 0.345 E_s^p - 0.543 HB + 3.690$

n = 61 s = 0.293 r = 0.952

HB = indicator variable for H-bond acceptor ability in substituents

4. The association equilibrium constant with bovine erythrocyte acetylcholinesterase of substituted phenyl N-methylcarbamates (Nishioka et al, 1977):

$\log (1/K_d) = 1.399 \pi_{2,3} + 0.306 \pi_4 + 1.659 \sigma^o (\rho>0) - 1.784 \sigma^o (\rho<0) + 0.168 E_s + 0.770 F + 1.358 HB + 0.072$

n = 53 s = 0.238 r = 0.947

HB = indicator variable for H-bond acceptor ability in substituents

5. Anticonvulsant activity of aralkyl and alkyl carbamates in mice (Tanaka et al, 1985):

$-\log ED_{50} = 0.648 \log P - 0.196 (\log p)^2 - 3.331 \sigma_I - 0.547 I_A - 0.194 HB + 3.233$

n = 46 s = 0.134 r = 0.913

HB = indicator variable for H-bond acceptor ability in substituents

6. Soil absorbtion coefficients of polar compounds (organic pollutants) (Sabljic, 1987):

$\log K(OM/W) = 0.365 \log P + 0.0175 MR - 0.385 HBD + 0.513$

n = 128 s = 0.276 r = 0.935

HBD = indicator variable for H-bond donor ability

7. Toxicity of aliphatic toxicants versus guppies (Leegwater, 1989):

$\log 1/LC_{50} = 0.705 \log P + 0.0337 MR - 0.459 HBD - 5.29$

$n = 33 \quad s = 0.250 \quad r = 0.985$

HBD = indicator variable for H-bond donor ability

8. Fungicidal activity of methyl N-phenylcarbamates (Takahashi et al, 1988)

$pI_{50} = 1.075 \Sigma\pi_{o,m} + 0.632 \pi_p + 0.590 B_5^m - 0.087 (B_5^m)^2 + 0.379 B_5^{m'} + 0.295 HB_p$

$+ 2.363$

$n = 69 \quad s = 0.346 \quad r = 0.942$

HB_p = indicator variable for H-bond acceptor ability in p-substituents

Hydrogen bonding parameters are also incorporated in LSER equations which correlate large numbers of solubility and solvent-dependent properties.

9. Correlation of octanol/water partition coefficients of nonprotonic aliphatic solutes with solvatochromic parameters (Kamlet et al, 1984):

$\log P = 2.66 V/100 - 0.96 \pi^* - 3.38 \beta + 0.24$

$n = 47 \quad s = 0.18 \quad r = 0.991$

β = solute H-bond acceptor ability

10. Water solubility of non-HBD liquid solutes (Taft et al, 1985a):

$-\log S_w = 3.40 V/100 - 0.41 \pi^* - 5.30 \beta_m$

$n = 92 \quad s = 0.16 \quad r2=0.986$

β_m = monomer solute H-bond acceptor ability

11. Narcotic effects of organic nonelectrolytes to the tadpole (Kamlet et al, 1988b):

$\log C = -4.87\ V_I/100 - 0.48\ \pi^* + 4.57\ \beta_m - 0.65\ \alpha_m - 0.67$

$n = 39 \quad s = 0.168 \quad r = 0.9899$

β_m = monomer solute H-bond acceptor ability

α_m = monomer solute H-bond donor ability

12. Induction of general anesthesia in animals (Abraham et al, 1991a):

$\log (1/EC_{50}) = 0.87 - 0.53\pi^*_2 + 0.46R_2 - 4.25\beta^H_2 + 4.00V_X$

$n = 27 \quad s = 0.20 \quad r = 0.9923$

β^H_2 = solute H-bond acceptor ability

6. Aims of the study

The solvatochromic H-bonding parameters of Abraham et al (1989; 1990) are successful in correlating large numbers of diverse properties, but the difficulty with them is that they are experimentally derived and consequently empirical, thus precluding the use of LSER (linear solvation energy relationship) equations for *a priori* predictions and estimations of solute/solvent properties (Famini et al, 1992). Although there are tables of these parameters and predictive relationships to help in their estimation, their values for complex molecules are not easily found. The difficulty in generating these variables has greatly discouraged the application of this quantitative structure-activity relationship method (Hickey & Passino-Reader, 1991).

In this investigation, work has been done toward finding H-bonding parameters calculated from a knowledge of quantum chemistry. They are easily obtainable and they also make it possible to predict activity of compounds *a priori*. These descriptors correlate well with experimental parameters and they should be applicable to QSARs involving H-bonding parameters, and to LSERs.

Because of the importance of electrostatic contribution to the H-bonding energy and the fact that there is a good correlation between electrostatic energy and total interaction energy of some H-bonded complexes (Scrocco & Tomasi, 1978), the electrostatic theory of H-bonding was first considered. In doing so, electrostatic descriptors including atomic charges, dipole moment, electrotopological state indices

and electrostatic potentials of molecules were examined.

Although H-bonding is mainly an electrostatic interaction, it has also contributions from other forces among which the most dominant is charge transfer energy. Charge transfer energy depends on the ionisation potential of the base and the electron affinity of the acid. In this work, energies of the highest occupied and the lowest unoccupied molecular orbitals of the H-bond acceptor and H-bond donor respectively were used to quantify charge transfer contribution to the H-bonding energy.

Finally, Molecular Discovery programmes (Great, Grin, Grid, Emin) were used to calculate the lowest interaction energies between various acids with a common base and also between different bases and a common H-bond acid, using molecular mechanics methods.

7. Atomic charge parameters calculated by CNDO method

As hydrogen bonding is mainly electrostatic in nature (Murrell et al, 1985), it seemed reasonable that electrostatic interactions could model H-bonding ability in compounds. The usual (and simplest) representation of the electrostatic properties of a system is through atom centred point charges. In this chapter charges were calculated using the CNDO method.

7.1. Methods and experimental data

Molecules were chosen for study subject to the availability of the experimental hydrogen bonding values. The COSMIC force field was used to minimise the energy of each molecule under study. A semiempirical molecular orbital method, CNDO (Pople & Beveridge, 1970), was used to optimise geometries and calculate partial atomic charges. MINITAB data analysis software was used to carry out regression analysis between calculated parameters and H-bonding experimental data taken from the literature. Semiempirical calculations and statistical analysis were performed on the university's VAX mainframe.

The experimental data (experimental H-bonding parameters), namely log K_α and log K_β values (Abraham et al, 1989a), α^H_2 (Abraham et al, 1989b; Abraham) and β^H_2 values (Abraham et al, 1990; Abraham) and also $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ values (Abraham, 1993) were used for regression analysis with atomic charges.

Atomic charge on the most positive hydrogen atom in the molecule was used as its H-bonding donor (HBD) ability (Q_H). If there was more than one hydrogen bonding group in a molecule (e.g. in 4-aminobenzoic acid) the charges on the hydrogen atoms were summed; it was assumed that the $-NH_2$ group formed only one hydrogen bond.

In order to model H-bonding acceptor (HBA) ability, the atomic charge on the most negatively charged atom in the molecule which was also capable of H-bonding, Q_{MN}, was calculated.

7.2. Results and discussion

7.2.1. Comparison of α^H_2 and β^H_2 values with atomic charge parameters

The first series studied was some aniline derivatives for which melting points had been predicted (Dearden, 1991). Q_H values were atomic charges on the hydrogens of the amino groups of the anilines, except for those anilines which have a hydrogen donating group as the substituent, i.e. hydroxyl and carboxyl substituted anilines, in which the Q_H values were the sum of the charges of hydrogens in the two H-bond donor groups. Q_{MN} values were mostly atomic charge on amino nitrogen of substituted anilines; however, when aniline had the substituents, -OMe, -OH, -COOH, -COOMe, $-NO_2$, -OEt, the most negative atom was the oxygen atom of the substituent. In the case of anilines with a H-bond accepting group as the substituent, Q_{MN2} values are sum of the charges on the two H-bond accepting atoms. Table 7.1 contains α^H_2, β^H_2,

Q_H, Q_{MN}, Q_{MN2} for these compounds.

The CNDO method calculates a large positive charge on the bromine atom, therefore it could not be summed with the charge on the nitrogen atom to form Q_{MN2} value.

In order to avoid the influence of intramolecular H-bonding, 2-substituted anilines were initially excluded from the correlation analysis. The CNDO-calculated charge parameter (Q_H) correlated well with α^H_2-values:

$$\alpha^H_2 = 3.98\ Q_H - 0.05 \qquad (7.1)$$

$n = 29 \quad s = 0.018 \quad r = 0.947 \quad F = 235$

Here n is the number of data points, s the standard error of the estimate, r the correlation coefficient and F the variance ratio.

Because standard error of α^H_2 is about 0.02 (Abraham, 1989b), the s value of the correlations between α^H_2 and theoretical parameters should ideally be about (or less than) this value.

When including 2-substituted aniline derivatives in the correlations, the coefficient did not change significantly, and such inclusion gave rise to the following equation:

$$\alpha^H_2 = 3.44\ Q_H + 0.000 \qquad (7.2)$$

$n = 39 \quad s = 0.074 \quad r = 0.935 \quad F = 258$

Table 7.1. α^H_2 and β^H_2 (Abraham, personal communication); Q_H, Q_{MN} and Q_{MN2} (calculated by CNDO method); T_m(K) (melting point from Dearden, 1991); dipole moment (μ) (Debye) calculated by CNDO method for some substituted anilines

Substituent	T_m(K) (Observed)	α^H_2	β^H_2	Q_H	Q_{MN}	Q_{MN2}	μ
H	266.7	0.26	0.38	0.084	-0.208	-	1.80
2-F	244.5	0.30	0.32	0.083	-0.201	-0.402	3.42
2-Cl	271.1	0.30	0.33	0.090	-0.212	-0.378	4.34
2-Br	305	0.30	0.33	0.059	-0.212	*	3.92
2-Me	249.3	0.23	0.38	0.080	-0.216	-	1.62
2-Et	230	0.23	0.38	0.079	-0.216	-	1.53
2-OMe	279.2	0.23	0.35	0.078	-0.221	-0.426	3.25
2-OH	447	0.86	0.41	0.218	-0.253	-0.458	3.38
2-NH_2	375.5	0.39	0.57	0.163	-0.215	-0.428	3.19
2-COOH	419.5	0.72	0.59	0.242	-0.311	-0.504	6.16
2-NO_2	344.5	0.19	0.55	0.088	-0.362	-0.550	5.93
3-Cl	262.7	0.33	0.29	0.085	-0.215	-0.384	4.45
3-Br	291.5	0.33	0.27	0.075	-0.221	*	1.22
3-Me	242.6	0.23	0.40	0.081	-0.217	-	1.59
3-Et	209	0.23	0.40	0.081	-0.217	-	1.57
3-$CH_2C_6H_5$	312	0.24	0.39	0.081	-0.217	-	1.80
3-OMe	272	0.25	0.71	0.082	-0.218	-0.435	2.46
3-OH	395	0.86	0.60	0.218	-0.249	-0.466	2.52
3-NH_2	336.5	0.52	0.76	0.165	-0.218	-0.432	2.65
3-CN	326.5	0.38	0.80	0.084	-0.216	-0.378	1.75
3-COOH	447	0.80	0.80	0.237	-0.310	-0.498	4.01
3-COOMe	312	0.30	0.85	0.081	-0.314	-0.480	3.85
3-NO_2	387	0.40	0.72	0.089	-0.362	-0.575	6.07
4-F	272.2	0.28	0.36	0.081	-0.213	-0.419	2.33
4-Cl	345.5	0.30	0.34	0.086	-0.217	-0.391	3.35
4-Br	339.4	0.31	0.34	0.068	-0.213	*	2.74
4-Me	316.7	0.23	0.42	0.079	-0.216	-	1.81
4-Et	268.1	0.23	0.42	0.079	-0.216	-	1.83
4-iPr	210	0.23	0.42	0.079	-0.216	-	1.84
4-tBu	290	0.23	0.43	0.078	-0.215	-	1.87
4-$CH_2C_6H_5$	307.5	0.34	0.40	0.080	-0.216	-	1.47
4-OMe	330.2	0.23	0.71	0.081	-0.219	-0.426	3.00
4-OEt	275.4	0.23	0.71	0.081	-0.229	-0.436	2.99
4-OH	457	0.86	0.60	0.213	-0.251	-0.458	3.05
4-NH_2	419	0.52	0.76	0.160	-0.214	-0.421	1.74
4-NHMe	309	0.43	0.76	0.161	-0.208	-0.388	1.68
4-CN	359	0.40	0.80	0.081	-0.207	-0.377	3.92
4-COOH	461.5	0.85	0.80	0.239	-0.323	-0.496	5.77
4-NO_2	421.5	0.42	0.72	0.089	-0.365	-0.570	5.31

*: Anilines containing bromine atom; -: Q_{MN2} values are equal to the corresponding Q_{MN} value because there is only one H-bond accepting group in the molecule.

In the case of H-bonding proton acceptor ability (H-bonding basicity) the following correlations were obtained for 3- and 4-substituted anilines (eq. 7.3) and for all the anilines (eq. 7.4):

$$\beta^H_2 = -1.97\ Q_{MN} + 0.093 \qquad (7.3)$$

n = 29 s = 0.174 r = 0.482 F = 8.2

$$\beta^H_2 = -1.78\ Q_{MN} + 0.101 \qquad (7.4)$$

n = 39 s = 0.167 r = 0.462 F = 10

When using Q_{MN2} values, the correlation with β^H_2 is a little better than that of the Q_{MN}:

$$\beta^H_2 = -1.18\ Q_{MN2} + 0.143 \qquad (7.5)$$

n = 29 s = 0.133 r = 0.749 F = 32.9

$$\beta^H_2 = -0.973\ Q_{MN2} + 0.173 \qquad (7.6)$$

n = 39 s = 0.146 r = 0.636 F = 25.1

From correlations 7.1-7.6 it can be seen that the Q_H parameter is more successful than the Q_{MN} and Q_{MN2} parameters in predicting the corresponding H-bonding experimental values for aniline derivatives.

To examine the suitability of Q_H parameter in other classes of compounds, the charge parameter was calculated for a combined set of different types of H-bond acids including alcohols, phenols, amines, carboxylic acids, amides and other compounds using the CNDO method. The results have been tabulated in Table 7.2.

Table 7.2. α^H_2 values (calculated from log K^H_A values taken from Abraham et al (1989b) using the relationship: $\alpha^H_2 = (\log K^H_A + 1.1)/4.636$) and Q_H values calculated by the CNDO method for some H-bond donors

Compound	α^H_2	Q_H	μ
Water	0.353	0.147	2.100
Methanol	0.367	0.145	1.910
Ethanol	0.333	0.143	1.987
Propan-1-ol	0.316	0.142	1.975
Butan-1-ol	0.330	0.141	1.980
Propan-2-ol	0.325	0.136	1.949
t-Butyl alcohol	0.320	0.134	2.030
3-Ethyl-2,4-dimethylpentan-3-ol	0.246	0.135	1.904
Di-t-butylmethanol	0.269	0.140	1.848
3-Isopropyl-2,2,4,4-tetramethylpentan-3-ol	0.196	0.134	1.834
Me_3SiOH	0.393	0.174	2.887
2,2,2-Trifluoroethanol	0.567	0.158	3.682
2,2,2-Trichloroethanol	0.500	0.153	4.579
2,2,2-Tribromoethanol	0.478	0.114*	3.694
2,2,2,3-Tetrafluoropropan-1-ol	0.532	0.153	2.263
Hexafluoropropan-2-ol	0.771	0.174	3.044
2,2,2-Trifluoro-1,1-bis(trifluoromethyl)ethanol	0.862	0.184	1.905
Phenol	0.596	0.145	1.742
Thiophenol	0.074	-0.026*	2.609
2-Methoxyphenol	0.261	0.145	2.814
2-Methylphenol	0.519	0.145	1.597
2-Isopropylphenol	0.536	0.145	1.568
2-t-Butylphenol	0.500	0.144	1.566
3-Methylphenol	0.572	0.144	1.598
3-Fluorophenol	0.676	0.149	3.074
3-Chlorophenol	0.693	0.149	3.483
3-Bromophenol	0.699	0.138*	1.319
3-Trifluoromethylphenol	0.721	0.150	4.080
3-Nitrophenol	0.785	0.153	6.547
4-Methoxyphenol	0.573	0.143	3.035
4-Methylphenol	0.569	0.144	1.738
4-s-Butylphenol	0.572	0.143	1.745
4-t-Butylphenol	0.558	0.143	1.800
4-Fluorophenol	0.629	0.146	1.943
4-Chlorophenol	0.670	0.149	2.727
4-Bromophenol	0.674	0.134*	3.748
4-Iodophenol	0.679*		
4-Acetylphenol	0.723	0.149	4.267
4-Cyanophenol	0.787	0.148	3.233
4-Nitrophenol	0.824	0.155	5.150

Table 7.2. Continued

Compound	α^H_2	Q_H	μ
2,6-Dimethylphenol	0.390	0.145	1.662
3,4-Dimethylphenol	0.559	0.143	1.677
2-Methyl-6-t-butylphenol	0.366	0.151	1.793
3,4-Dichlorophenol	0.743	0.153	4.086
3,5-Dichlorophenol	0.774	0.153	2.430
2,4,6-Trimethylphenol	0.374	0.145	1.688
3,4,5-Trichlorophenol	0.817	0.156	4.258
Pentafluorophenol	0.764	0.170	2.055
Pentachlorophenol	0.553	0.172	1.667
Pentabromophenol	0.499	0.104*	2.114
1-Naphthol	0.608	0.147	1.517
2-Naphthol	0.612	0.145	1.453
N,N-Dibenzylhydroxylamine	0.453	0.131	3.223
4-Chloroperoxybenzoic acid	0.378	0.189	2.122
Hept-1-yne	0.127	0.087	0.636
Chloroform	0.197	0.073	2.896
1,1-Dinitroethane	0.394	0.039	6.625
Ammonia	0.434	0.078	1.978
Cyanic acid	0.558	0.173	2.665
Thiocyanic acid	0.751	-0.012*	2.833
N-Nitromethylamine	0.593	0.116	5.441
N-Nitropropylamine	0.569	0.114	5.642
N-Nitrobutylamine	0.568	0.114	5.668
N-Nitrocyclohexylamine	0.539	0.112	6.027
2-Cyano-N-nitroethylamine	0.738	0.140	2.248
N,3,3,3-Tetranitropropylamine	0.775	0.124	3.562
Ethyl N-nitrocarbamate	0.615	0.166	3.722
Aniline	0.264	0.077	1.887
2-Nitroaniline	0.368	0.089	6.193
3-Nitroaniline	0.398	0.086	6.572
4-Nitroaniline	0.421	0.084	4.034
4-Bromoaniline	0.308	0.064*	3.909
2-Chloro-4-nitroaniline	0.453	0.088	3.791
4-Chloro-2-nitroaniline	0.445	0.092	5.170
2-Aminopyridine	0.318	0.082	1.560
3-Aminopyridine	0.348	0.077	1.079
4-Aminopyridine	0.409	0.080	2.509
Diphenylamine	0.324	0.069	1.704
N,O-Dibenzylhydroxylamine	0.374	0.061	1.691
N-Methylacetamide	0.383	0.126	3.198
Pyrrole	0.408	0.098	2.012
Tetrachloropyrrole	0.722	0.143	4.702

Table 7.2. Continued

Compound	α^H_2	Q_H	μ
Tetrabromopyrrole	0.694	-0.006*	1.986
Tetraiodopyrrole	0.602*		
Indole	0.436	0.105	2.105
5-Fluoroindole	0.468	0.107	3.823
Carbazole	0.469	0.119	1.620
Maleimide	0.497	0.137	1.617
Succinimide	0.493	0.145	1.851
Isobutyl alcohol	0.311	0.142	1.968
Neopentyl alcohol	0.325	0.143	1.965
t-Pentyl alcohol	0.316	0.132	1.971
2-Chloroethanol	0.346	0.145	2.374
2-Fluoroethanol	0.396	0.147	1.907
Hexachloropropan-2-ol	0.645	0.169	2.234
Benzyl alcohol	0.392	0.140	1.820
Pentafluorobenzyl alcohol	0.466	0.151	3.405
1,1,1-Trichloro-2-methylpropan-2-ol	0.400	0.150	2.808
1,1,1-Trifluoro-2-methylpropan-2-ol	0.467	0.151	1.990
1,1,1,3,3,3-Hexafluoro-2-methylpropan-2-ol	0.655	0.168	0.508
1,1,1,3,3,3-Hexafluoro--2-trichloromethylpropan-2-ol	0.743	0.180	1.853
2-Chlorophenol	0.650	0.150	3.887
2-Cyanophenol	0.738	0.150	5.056
3-Ethylphenol	0.548	0.144	1.577
3-Dimethylaminophenol	0.520	0.145	1.502
3-Methoxyphenol	0.591	0.147	3.126
3-Cyanophenol	0.772	0.147	4.502
4-Ethylphenol	0.546	0.143	1.823
4-Propylphenol	0.546	0.143	1.827
4-Isopropylphenol	0.551	0.143	1.821
4-Octylphenol	0.547	0.143	1.821
4-Phenylphenol	0.595	0.144	1.689
4-Trifluoromethylphenol	0.723	0.151	2.810
2,3-Dimethylphenol	0.533	0.145	1.476
2,4-Dimethylphenol	0.532	0.144	1.628
2,5-Dimethylphenol	0.532	0.145	1.807
3,5-Dimethylphenol	0.567	0.144	1.781
4-Methyl-2-t-butylphenol	0.565	0.143	1.593
3-Methyl-6-t-butylphenol	0.554	0.144	1.805
2,4-di-t-butylphenol	0.545	0.143	1.675
4-Nitro-3-trifluoromethylphenol	0.955	0.156	5.332
2,6-Dichlorophenol	0.321	0.165	4.662
3,5-Di(trifluoromethyl)phenol	0.815	0.155	2.687

Table 7.2. Continued

Compound	α^H_2	Q_H	μ
2,3,5-Trimethylphenol	0.520	0.144	1.682
3,4,5-Trimethylphenol	0.546	0.143	1.833
4-Bromo-2,6-dimethylphenol	0.463	0.136*	1.736
2,6-Dichloro-4-nitrophenol	0.704	0.171	4.708
3-Chloroperbenzoic acid	0.387	0.189	3.584
4-t-Butylperbenzoic acid	0.314	0.186	3.224
Trifluoroacetic acid	0.951	0.188	1.860
Trichloroacetic acid	0.947	0.182	1.191
Dichloroacetic acid	0.899	0.178	3.388
Pentafluorobenzoic acid	0.888	0.178	1.098
2-Bromobenzoic acid	0.642	0.160*	0.615
Chloroacetic acid	0.818	0.174	0.667
Benzoic acid	0.588	0.169	1.623
Acetic acid	0.580	0.169	1.184
Hexanoic acid	0.471	0.167	1.227
Trimethylacetic acid	0.514	0.169	1.186
1,2-Dichloroethane	0.095	0.033	3.734
1,1,1-Trichloroethane	0.011	0.030	2.051
Dichloromethane	0.129	0.056	1.609
Bromoform	0.170	-0.093*	2.712
Bromodichloromethane	0.123	0.013*	2.269
1,1,2-Trichloroethene	0.119	0.054	1.498
1,2-Dibromo-1,1-difluoroethane	0.140	-0.019*	3.122
1,2-Dichloro-1-fluoroethane	0.175	0.051	1.941
1-Chloro-1,1,2-trifluoro-2-iodoethane	0.186*		
1,2-Dichloro-1,2-difluoroethane	0.201	0.033	1.935
1,1,2-Trichloro-2,2-difluoroethane	0.207	0.079	1.958
1-Bromo-2-chloro-1,1,2-trifluoroethane	0.209	0.010*	5.343
1-Bromo-1-chloro-2,2,2-trifluoroethane	0.224	0.022*	4.221
2,2-Dichloro-1,1-difluoroethyl methyl ether	0.166	0.076	1.156
2-Chloro-1,1,2-trifluoro -ethyl difluoromethyl ether	0.194	0.046	1.957
3-Chloro-3-methylbut-1-yne	0.151	0.099	2.571
Trimethylsilylethyne	0.132	0.077	0.657
Triethylsilylethyne	0.132	0.079	2.034
3-Chloropropyne	0.186	0.103	2.134
3-Bromopropyne	0.186	0.064*	0.841
t-Butylethyne	0.127	0.086	0.685
Benzoylethyne	0.194	0.104	3.325
Phenylethyne	0.116	0.092	0.162
Pentamethyl(prop-2-ynyl)phosphoric triamide	0.123	0.095	4.567

Table 7.2. Continued

Compound	α^H_2	Q_H	μ
N,N,N´N´-Tetramethyl-N˝-benzyl-N˝-			
-prop-2-ynylphosphoric triamide	0.123	0.095	4.769
Prop-2-ynyl bis(piperidino)phosphinate	0.138	0.086	4.603
Prop-2-ynyl bis(diethylamido)phosphinate	0.136	0.085	4.826
Prop-2-ynyl bis(dimethylamido)phosphinate	0.153	0.086	4.594
But-2-ynyl bis(dimethylamido)phosphinate	0.129	0.088	4.226
Prop-2-ynyl bis(dibutylamido)phosphinate	0.153	0.086	4.906
N˝-Ethyl-N,N,N´,N´,-tetramethyl-N˝-			
-prop-2-ynyl-phosphoric triamide	0.112	0.092	4.186
Prop-2-ynyl bis(aziridino)phosphinate	0.175	0.087	3.515
Prop-2-ynyl bis(morpholino)phosphinate	0.179	0.088	4.688
S-Prop-2-ynyl bis(dimethylamido)			
-thiophosphinate	0.181	0.088*	3.328
Diethyl prop-2-ynyl phosphate	0.205	0.101	3.542
O-Prop-2-ynyl bis(dimethylamido)			
-thiophosphinate	0.151	0.089*	1.937
Butyl sulphide	-0.018	0.018*	1.045
Isopropyl sulphide	-0.018	0.016*	2.029
t-Butyl sulphide	-0.018	0.004*	2.888
Thioacetamide	0.576	0.100*	2.889
N-Methylaniline	0.173	0.078	1.670
N-Phenylurethane	0.357	0.085	5.378
Propynonitrile	0.339	0.117	3.441
2-Aminopyrimidine	0.272	0.090	2.841
4-Aminopyrimidine	0.371	0.085	1.662
5-Aminopyrimidine	0.384	0.076	2.482
α-Naphthylamine	0.313	0.077	1.782
β-Naphthylamine	0.347	0.076	1.815
α-Heptafluoronaphthol	0.679	0.170	2.183
β-Heptafluoronaphthol	0.773	0.169	1.985

*: Compounds which have not been used in equation 7.7.

Unfortunately, as is clear from the following relationship, a single equation could not relate α^H_2 values to Q_H values very successfully.

$$\alpha^H_2 = 4.23\ Q_H - 0.079 \qquad (7.7)$$

$n = 164$ $s = 0.151$ $r = 0.718$ $F = 174$

However, after dividing up the whole set of compounds into different classes, good correlations within the families were found to exist:

Alcohols: $\alpha^H_2 = 11.7\ Q_H - 1.31$ (7.8)

$n = 27$ $s = 0.043$ $r = 0.970$ $F = 389.8$

3- and 4-substituted phenols: $\alpha^H_2 = 23.6\ Q_H - 2.82$ (7.9)

$n = 35$ $s = 0.042$ $r = 0.926$ $F = 200.0$

Amines (aromatic and aliphatic): $\alpha^H_2 = 5.68\ Q_H - 0.096$ (7.10)

$n = 30$ $s = 0.073$ $r = 0.863$ $F = 81.5$

Aliphatic carboxylic acids: $\alpha^H_2 = 26.0\ Q_H - 0.096$ (7.11)

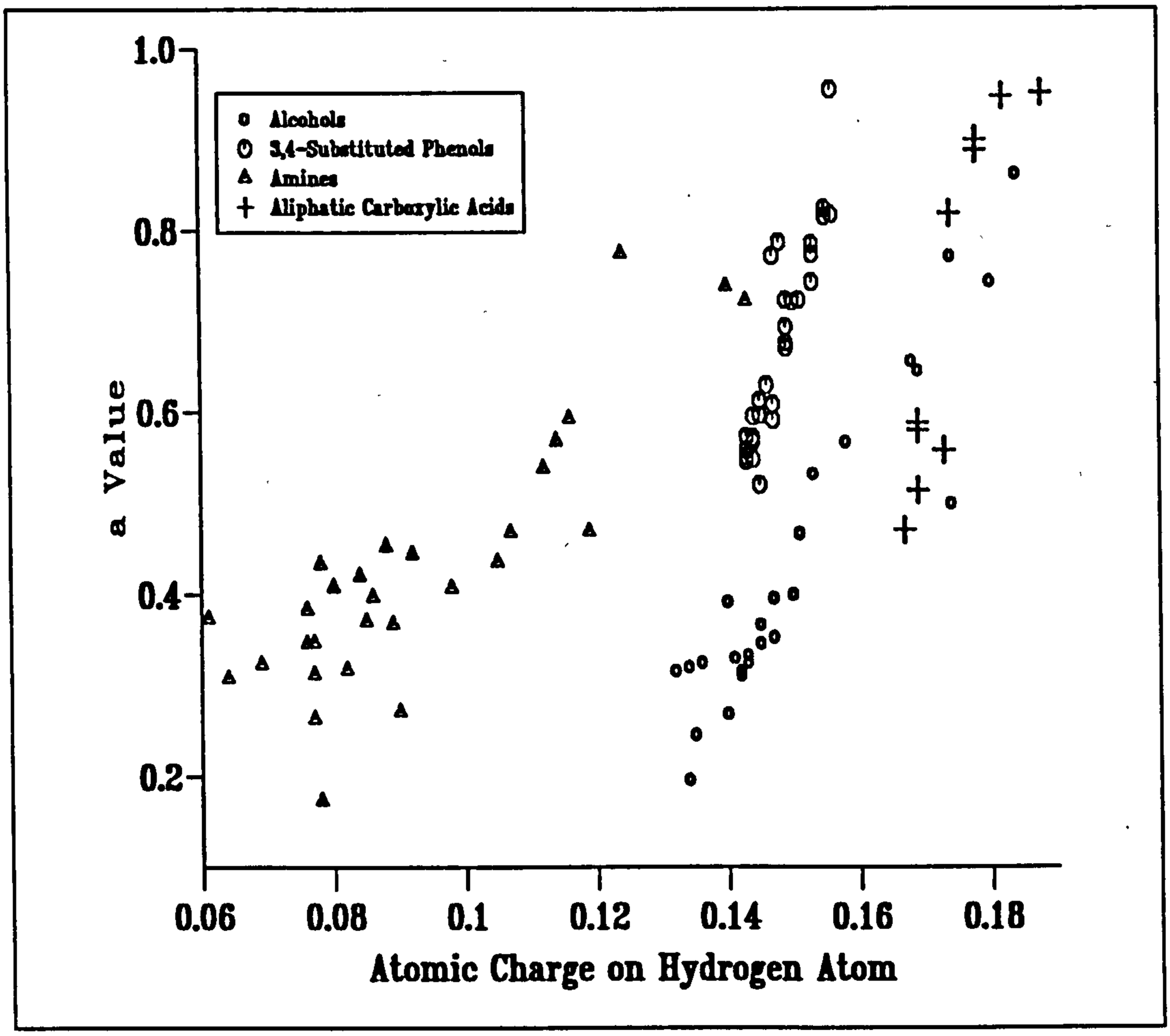
$n = 10$ $s = 0.089$ $r = 0.903$ $F = 35.2$

It should be noted that in these equations iodine-containing structures (because of the lack of parametrisation) and bromine-containing structures (because CNDO method

calculates an unreasonable positive charge for bromine, for instance in bromoalkanes) have been excluded. The molecules containing sulphur were also deleted from regression analysis because this method predicted hydrogen atoms connected to sulphur atoms, e.g. in thiocyanic acid and thiophenol, to have negative charge.

Although 2-substituted anilines did not change the relationship between α^H_2 and Q_H in the regression analysis for anilines (eq. 7.1 and 7.2), 2-substituted phenols had to be excluded from the regression analysis. The difference between aniline and phenol is in aniline having two hydrogen atoms connected to the nitrogen; it can be assumed that when one of the hydrogens is engaged in an intramolecular H-bonding the other hydrogen is still available for intermolecular H-bonding.

Taft and Kamlet (Kamlet et al, 1981, 1983, 1985) have many times drawn attention to the existence of family dependent (FD) properties, i.e., properties that are readily correlated *via* solvatochromic parameters within compound sets of similar type but require the addition of a new term if dissimilar sets are to be incorporated. Abraham et al (1989d) have explained that some basicity dependent properties (BDPs) show family dependent behaviour. A rather special area of BDPs is that of H-bonding. Here the very different regression coefficients in equations (7.8-7.11) show that a successful global correlation between α^H_2-values and Q_H is not possible and that charge parameters are family dependent properties. Plot 7.1 clearly shows that there are a number of lines for different families of acids.



Plot 7.1. Family dependent behaviour of acidity dependent properties.

For 215 different types of proton acceptors (including ketones, amines, amides, esters, ethers etc.), atomic charges were calculated using the CNDO method. The results of calculations (Q_{MN} values) together with the corresponding β^H_2 values are listed in Table 7.3. Q_{MN} values in this table are the most negative atomic charge in the molecule; when the molecule had more than one H-bond acceptor group (like dioxane), only the highest negative charge was listed in the table and further necessary statistical corrections will be discussed later. Q_{MN} values for halogenated alkanes are atomic charges on the halogen atoms. For aromatic structures which do not have a heteroatom substituent, the average of the atomic charges on the carbon atoms of the ring is taken as the charge parameter. For esters the most negative charge resided on the carbonyl oxygen.

Table 7.3. Q_{MN} values calculated by the CNDO method and the corresponding β^H_2 values calculated by equation: $\beta^H_2 = (\log K^H_B + 1.1)/4.636$, with $\log K^H_B$ values taken from Abraham et al (1990a) for some H-bond bases

Compound	β^H_2	Q_{MN}	Q_{MN2}
1-Chlorobutane	0.106	-0.159	-
2-Chloro-2-methylpropane	0.189	-0.188	-
1-Bromobutane	0.202	+0.238*	-
2-Bromo-2-methylpropane	0.167	+0.235*	-
Benzene	0.146	+0.006	-
Toluene	0.142	-0.010	-
1,2-Dimethylbenzene	0.162	-0.004	-
1,3-Dimethylbenzene	0.175	-0.026	-
1,4-Dimethylbenzene	0.179	-0.008	-
1,3,5-Trimethylbenzene	0.201	-0.035	-
1,2,4,5-Tetramethylbenzene	0.203	-0.018	-
Hexamethylbenzene	0.258	-0.017	-
Naphthalene	0.212	-0.007	-
Phenanthrene	0.218	-0.008	-
Chlorobenzene	0.110	-0.166	-
Bromobenzene	0.074	+0.275*	-
Dimethyl ether	0.433	-0.205	-
Diethyl ether	0.450	-0.226	-
Dipropyl ether	0.444	-0.233	-
Di-isopropyl ether	0.457	-0.245	-
Dibutyl ether	0.419	-0.229	-
Di(t-butyl) ether	0.375	-0.255	-
Ethyl t-butyl ether	0.495	-0.241	-
Trimethylene oxide (oxetane)	0.538	-0.220	-
Tetrahydrofuran	0.510	-0.230	-
1,4-Dioxane	0.475	-0.219	-0.438
Tetrahydropyran	0.477	-0.227	-
Diphenyl ether	0.244	-0.234	-
Dibenzyl ether	0.388	-0.225	-
Anisole	0.260	-0.222	-
1,8-Cineole	0.513	-0.260	-
Benzaldehyde	0.415	-0.238	-
Propanone	0.497	-0.265	-
Butanone	0.481	-0.272	-
Pentan-3-one	0.440	-0.285	-
4-Methylpentan-2-one	0.451	-0.273	-
Cyclopentanone	0.526	-0.263	-
Cyclohexanone	0.523	-0.274	-
Mesityl oxide	0.499	-0.294	-
Piperidone	0.536	-0.354	-
Hexafluoropropanone	0.195	-0.196	-
Acetophenone	0.511	-0.271	-
Benzophenone	0.459	-0.280	-

Table 7.3. Continued

Compound	β^H_2	Q_{MN}	Q_{MN2}
2,6-Dimethyl-4-pyrone	0.779	-0.349	-0.540
Flavone	0.653	-0.330	-0.533
4-Methoxyacetophenone	0.526	-0.282	-0.501
Methyl formate	0.379	-0.265	-
Methyl acetate	0.398	-0.303	-
Ethyl acetate	0.446	-0.319	-
Vinyl acetate	0.398	-0.311	-
2-Dimethylamino-3,3-dimethylaziridine	0.775	-0.191	-
N,N-Dimethylaminonitrile	0.560	-0.175	-
Nitrobenzene	0.341	-0.166	-
Acetonitrile	0.439	-0.162	-
1-Cyanobutane	0.441	-0.167	-
Chloroacetonitrile	0.337	-0.143	-0.246
Trichloroacetonitrile	0.168	-0.117	-
Benzonitrile	0.423	-0.166	-
Phenylacetonitrile	0.406	-0.170	-
t-Butylamine	0.712	-0.213	-
Diethylamine	0.704	-0.180	-
Di-isopropylamine	0.667	-0.194	-
Cyclohexyldimethylamine	0.700	-0.159	-
Triethylamine	0.669	-0.177	-
Tripropylamine	0.583	-0.186	-
Tributylamine	0.597	-0.184	-
Triallylamine	0.536	-0.155	-
Aniline	0.378	-0.208	-
Benzylamine	0.625	-0.196	-
Dibenzylamine	0.549	-0.182	-
Tribenzylamine	0.308	-0.166	-
N,N-Dimethylbenzylamine	0.596	-0.145	-
3-Aminotoluene	0.395	-0.214	-
4-Aminotoluene	0.421	-0.213	-
N,N-Dimethylaniline	0.351	-0.158	-
N,N-Diethylaniline	0.414	-0.213	-
3-Fluoroaniline	0.303	-0.217	-0.427
4-Fluoroaniline	0.362	-0.213	-0.419
3-Chloroaniline	0.288	-0.215	-0.384
4-Chloroaniline	0.338	-0.217	-0.391
3-Bromoaniline	0.274	-0.221*	-
4-Bromoaniline	0.336	-0.213*	-
3-Iodoaniline	0.288*		
4-Iodoaniline	0.312*		

Table 7.3. Continued

Compound	β^H_2	Q_{MN}	Q_{MN2}
3-Methoxyaniline	0.397	-0.218	-0.435
4-Methoxyaniline	0.454	-0.219	-0.426
N,N-Dimethylformamide	0.663	-0.238	-
N,N-Diethylformamide	0.672	-0.326	-
N-Methylacetamide	0.715	-0.354	-
N,N-Dimethylacetamide	0.730	-0.349	-
N,N-Diethylacetamide	0.730	-0.354	-
N,N-Dicyclohexylacetamide	0.766	-0.357	-
N-Acetylpiperidine	0.733	-0.351	-
N,N-Dimethylpropanamide	0.708	-0.356	-
N,N-Diethylpropanamide	0.689	-0.361	-
N,N-Dicyclohexylpropanamide	0.715	-0.368	-
N-Propionylpiperidine	0.717	-0.350	-
N,N-Diethylbutanamide	0.704	-0.357	-
N-Butyrylpiperidine	0.714	-0.359	-
Tetramethylurea	0.743	-0.393	-
1,1,1-Trifluoro-N,N-dimethylacetamide	0.455	-0.284	-
1-Chloro-N,N-dimethylacetamide	0.612	-0.338	-0.451
1-Chloro-N,N-diethylacetamide	0.621	-0.336	-0.446
1,1-Dichloro-N,N-diethylacetamide	0.539	-0.327	-0.413
1-Chloro-N,N-dicyclohexylacetamide	0.610	-0.351	-0.476
N-Chloroacetylpiperidine	0.618	-0.340	-0.457
N,N-Diphenylacetamide	0.642	-0.347	-
N,N-Diphenylpropanamide	0.615	-0.354	-
N,N-Diphenylbutanamide	0.627	-0.352	-
N,N-Diphenylchloroacetamide	0.540	-0.337	-0.278
N,N-Dimethylbenzamide	0.674	-0.349	-
N,N-Diethylbenzamide	0.700	-0.348	-
N,N-Dicyclohexylbenzamide	0.719	-0.365	-
N-Benzoylpiperidine	0.704	-0.352	-
N,N-Diphenylbenzamide	0.601	-0.345	-
N,N-Diethyl-4-nitrobenzamide	0.614	-0.348	-0.579
N,N-Dicyclohexyl-4-nitrobenzamide	0.616	-0.370	-0.581
4-Nitro-N,N-diphenylbenzamide	0.512	-0.335	-0.563
Pyridine	0.625	-0.145	-
2-Methylpyridine	0.625	-0.165	-
3-Methylpyridine	0.620	-0.142	-
4-Methylpyridine	0.655	-0.155	-
2,4-Dimethylpyridine	0.644	-0.175	-
2,6-Dimethylpyridine	0.638	-0.186	-
2,4,6-Trimethylpyridine	0.693	-0.194	-

Table 7.3. Continued

Compound	β^H_2	Q_{MN}	Q_{MN2}
2-Ethylpyridine	0.601	-0.170	-
2-t-Butylpyridine	0.497	-0.171	-
2-Fluoropyridine	0.432	-0.203	-0.388
2-Chloropyridine	0.450	-0.161	-0.291
3-Chloropyridine	0.488	-0.159	-0.298
2-Bromopyridine	0.435	-0.213*	-
3-Bromopyridine	0.508	-0.137*	-
4-(N,N-Dimethylamino)pyridine	0.859	-0.160	-0.319
Pyridine N-oxide	0.809	-0.403	-
N-Methylpyrrolidin-2-one	0.765	-0.274	-
N-Phenylpyrrolidin-2-one	0.631	-0.284	-
N-Methyl-2-pyridone	0.764	-0.398	-
N-Methylimidazole	0.805	-0.162	-0.296
2-Aminopyrimidine	0.610	-0.221	-0.430
Pyridazine	0.636	-0.077	-0.154
Pyrimidine	0.526	-0.077	-0.154
3-Methyl-4-pyrimidone	0.637	-0.396	-0.501
N-Methylmorpholine	0.607	-0.223	-0.369
Pyrazine	0.480	-0.115	-0.230
1,4-Diazabicyclo[2,2,2]octane	0.806	-0.154	-0.308
Nicotine	0.687	-0.190	-0.330
3-(N,N-Diethyl)nicotinamide	0.707	-0.353	-0.499
1,3-Dimethyluracil	0.617	-0.389	-
Quinoline	0.633	-0.165	-
N-(2-Chlorophenyl)pyrrolidin-2-one	0.696	-0.348	-0.514
N-(2-Methoxyphenyl)pyrrolidin-2-one	0.729	-0.351	-0.573
N-(3-Methylphenyl)pyrrolidin-2-one	0.635	-0.282	-
N-(3-Chlorophenyl)pyrrolidin-2-one	0.560	-0.346	-0.515
N-(3-Methoxyphenyl)pyrrolidin-2-one	0.624	-0.346	-0.568
N-(4-Methylphenyl)pyrrolidin-2-one	0.649	-0.348	-
N-(4-Ethylphenyl)pyrrolidin-2-one	0.651	-0.350	-
N-(4-Chlorophenyl)pyrrolidin-2-one	0.573	-0.346	-0.515
N-(4-Methoxyphenyl)pyrrolidin-2-one	0.670	-0.351	-0.574
Diethyl sulphide	0.285	-0.098	-
Ethyl methyl sulphide	0.242	-0.086	-
Dibutyl sulphide	0.290	-0.105	-
Di-t-butyl sulphide	0.286	-0.160	-
Tetrahydrothiophene	0.264	-0.104	-
Dimethyl sulphoxide	0.775	-0.321	-
Di-isopropyl sulphoxide	0.789	-0.337	-
Dibutyl sulphoxide	0.785	-0.331	-

Table 7.3. Continued

Compound	β^H_2	Q_{MN}	Q_{MN2}
Diphenyl sulphoxide	0.667	-0.325	-
Di-p-tolyl sulphoxide	0.694	-0.329	-
Tetrahydrothiophene S-oxide	0.770	-0.327	-
Diphenyl sulphone	0.512	-0.326	-
Sulpholane	0.523	-0.324	-
Diethyl sulphite	0.415	-0.304	-
Trimethylphosphine oxide	0.980	-0.301	-
Triethylphosphine oxide	1.017	-0.303	-
Triphenylphosphine oxide	0.919	-0.311	-
Dimethyl phosphite	0.720	-0.285	-
Diethyl phosphite	0.742	-0.275	-
Di-isopropyl phosphite	0.774	-0.297	-
Dimethyl ethylphosphonate	0.811	-0.305	-
Diethyl methylphosphonate	0.825	-0.308	-
Diethyl ethylphosphonate	0.830	-0.315	-
Diethyl isopropylphosphonate	0.823	-0.317	-
Di-(1-chloropropyl) methylphosphonate	0.786	-0.310	-
Diethyl chloromethylphosphonate	0.761	-0.299	-
Diethyl dichloromethylphosphonate	0.701	-0.291	-
Diethyl trichloromethylphosphonate	0.646	-0.292	-
Trimethyl phosphate	0.762	-0.306	-
Triethyl phosphate	0.792	-0.317	-
Tributyl phosphate	0.771	-0.317	-
Triphenyl phosphate	0.624	-0.304	-
Ethyl isothiocyanate	0.224	-0.131	-
Methyl thiocyanate	0.359	-0.062	-
Ethyl thiocyanate	0.366	-0.070	-
Tetramethylthiourea	0.514	-0.256	-
O-Methyl-N,N-dimethylthiocarbamate	0.416	-0.305	-
N,N-Dimethylthioacetamide	0.492	-0.338	-
N,N-Dimethylthiobenzamide	0.476	-0.314	-
N,N-Dimethylamino(thioxo)acetonitrile	0.368	-0.280	-0.396
N,N-Dimethylmethanesulphinamide	0.736	-0.323	-
N,N-Dimethylbezenesulphinamide	0.684	-0.323	-
N,N-Dimethyltoluene-p-sulphinamide	0.685	-0.326	-
N-Methylmethanesulphonamide	0.508	-0.328	-
N,N-Dimethylmethanesulphonamide	0.517	-0.330	-
N,N-Dimethylbenzenesulphonamide	0.530	-0.337	-
N,N-Dimethyltoluene-p-sulphonamide	0.546	-0.338	-
Hexamethylphosphoramide	1.000	-0.365	-
Diethyl N,N-dimethylaminophosphonate	0.844	-0.335	-

Table 7.3. Continued

Compound	β^H_2	Q_{MN}	Q_{MN2}
Tributylphosphine sulphide	0.548	-0.009*	-
Trioctylphosphine sulphide	0.566	-0.011*	-
Triethyl thiophosphate	0.392	-0.234*	-
Hexamethylthiophosphoramide	0.519	-0.137*	-
Diethyl selenide	0.268*		
Dibutyl selenide	0.285*		

*:These compounds have not been used in the regression analyses; -:Q_{MN2} value for these molecules are equal to the Q_{MN} value (They have only one H-bond accepting group).

In order to investigate the usefulness of the Q_{MN} parameter in a combined set of compounds, correlation analysis with β^H_2 was carried out for the compounds in Table 7.3:

$$\beta^H_2 = -1.19\ Q_{MN} + 0.254 \qquad (7.12)$$

n = 198 s = 0.149 r = 0.613 F = 122.5

In this equation structures containing bromine, iodine and selenium atoms have not been used (because of the lack of correct parametrisation in the CNDO program). Q_{MN} values in this equation are only the highest negative charge (when the sum of the negative charges of heteroatoms in a molecule like dioxane was used the quality of the equation did not change). Because the general correlation (eq. 7.12) was not satisfactory, correlations within families were also examined, but the results were not as good as those of proton donors and good regressions were found only for sets of very closely related bases:

Sulphoxides, sulphites and sulphinamides:

$$\beta^H_2 = -11.0\ Q_{MN} - 2.88 \qquad (7.13)$$

n = 10 s = 0.060 r = 0.857 F = 22

Phosphites and phosphonates: $\beta^H_2 = -2.47\ Q_{MN} + 0.035$ (7.14)

n = 10 s = 0.027 r = 0.851 F = 24

Ketones: $\beta^H_2 = -1.67\ Q_{MN} + 0.027$ (7.15)

$n = 11$ $s = 0.039$ $r = 0.918$ $F = 48$

The ester group possesses several sites available for H-bonding, namely the two lone pairs of electrons of the carbonyl group and those of the methoxy group. In low temperature argon matrices both carbonyl- and methoxy-sites can be involved in H-bond formation with water or HCl. However in solutions at room temperature in equilibrium conditions, the participation of the methoxy group in H-bonding is completely negligible. There is evidence that bonding to the Z lone pair is preferred despite apparent steric hindrance and the same is likely for amides (Huyskens et al, 1987). *Ab initio* molecular orbital calculations of amide group shows that only the carbonyl oxygen is a good π H-bond donor (Johansson et al, 1974). Nevertheles, here despite the fact that carbonyl oxygen had the highest negative atomic charge, the atomic charge on the ethereal oxygen of esters (Table 7.4) and nitrogen of amides were also used as the charge parameter in the regression analysis for these two classes of compounds. Unfortunately there was no relationship between the atomic charge on the nitrogen of amides (Q_N) and their β^H_2 values. In the case of esters, only four esters were present in Table 7.3, for which the following relationship was obtained.

$\beta^H_2 = -1.61$ (charge on sp3 oxygen) + 0.049 (7.16)

$n = 4$ $s = 0.011$ $r = 0.949$ $F = 18$

When the atomic charge on the carbonyl oxygen was used, it was possible to

incorporate esters, amides and carboxylic acids with ketones and aldehydes in the correlation between β^H_2 and Q_{MN}, while Q_{MN2} values were not successful in correlation analyses for these compounds.

Structures containing carbonyl groups (carboxylic acids, amides, esters, aldehydes, ketones):

$$\beta^H_2 = -1.79\ Q_{MN} + 0.021 \qquad (7.17)$$

$n = 65 \quad s = 0.082 \quad r = 0.717 \quad F = 66.6$

Deleting those molecules with more than one functional group from the above correlation gave rise to the following slightly better equation:

$$\beta^H_2 = -2.12\ Q_{MN} - 0.0814 \qquad (7.18)$$

$n = 47 \quad s = 0.092 \quad r = 0.731 \quad F = 52.8$

For ethers and amines there was no relationship between β^H_2 and Q_{MN} (or Q_{MN2}) and even when ethers or amines which also have other functional groups in the structure were deleted there was no correlation. The reason could be that, unlike structures containing carbonyl groups, here the heteroatom is connected to two or three other atoms. Therefore conformation of the molecules has a much larger influence on the availability of the lone pair electrons.

Table 7.4. Atomic charge on the carbonyl oxygen in esters (Q_{MN} value) and the atomic charge on the ethereal oxygen

Ester	Atomic charge on ethereal oxygen	Atomic charge on carbonyl oxygen
Methyl formate	-0.207	-0.265
Methyl acetate	-0.223	-0.303
Ethyl acetate	-0.244	-0.319
Vinyl acetate	-0.210	-0.311

7.2.2. Correlation of charge parameters with log K_α and log K_β

Abraham et al (1989a) measured the equilibrium constants of H-bond formation of different classes of hydrogen bond donors (HBDs) against N-methylpyrrolidinone and hydrogen bond acceptors (HBAs) against p-nitrophenol. For the first time they used 1,1,1-trichloroethane as the solvent, a more polar solvent than tetrachloromethane (α^H_2 and β^H_2 values were calculated from equilibrium constants in tetrachloromethane). Their scales (log K_α and log K_β) were compared with charge parameters calculated by CNDO (Tables 7.5 & 7.6).

H-bond donors in Table 7.5 are alcohols, substituted phenols, carboxylic acids, amines, amides, sulphonamides and some sophisticated heterocycles. Q_H values in these molecules were the atomic charges on the hydrogen atoms connected to an electronegative atom. Azoles (compounds 56-59) can be in different tautomeric forms. While for tetrazole and 1,2,4-triazole the dominant form is well established as structure 59 and 56a in scheme 7.1, there remain ambiguities for 1,2,3-triazole

(Elguero et al, 1976) so both forms as displayed in the scheme were built for structures 57 and 58, and atomic charges were calculated for both forms; in regression analyses the average of Q_H values for the two tautomers were used (Table 7.5). The following very poor equation was resulted for all the H-bond donors in the table.

$$\log K_\alpha = 11.6\ Q_H - 0.005 \qquad (7.19)$$

n=60 r = 0.366 s = 0.775 F = 8.95

Even after deleting 2-substituted phenols, the correlation is still poor (r = 0.448). Abraham et al (1989a) have suggested that in the correlation of log K_α with an enthalpy-related parameter (or a free energy property for which the entropy of binding remains substantially constant) like $\Delta\nu_{C=O}$ (the i.r. carbonyl shift for N-methylpyrrolidinone on H-bond complex formation), a good relationship can be found only in the absence of some extra entropic constraints (e.g. steric or stereoelectronic constraints). Thereby they have explained the deviation of some H-bond donors in Table 7.5 from the equation between log K_α and $\Delta\nu_{C=O}$. Assuming the same thermodynamic status for the charge parameter as $\Delta\nu_{C=O}$, the same deviants were deleted from the correlation between log K_α and Q_H. These are 2-substituted phenols (for steric reasons), oximes, lactams, triazoles, tetrazole, carboxylic acids, sulphonamides and acylsulphonamide mainly for stereoelectronic reasons and also the aromatic amines. The plot of log K_α against Q_H for the rest of the H-bond donors (Figure 7.2) showed that Q_H values of acetanilides (except for 4′-N,N-diethylaminoacetanilide) and chloroform predict lower log K_α values for them. In case

of acetanilides it seems that the CNDO calculated Q_H values are the reverse of the expected inductive order. Chloroform is the only carbon acid in the list. There is evidence from gas-phase equilibria measurements that the carbon acids have lower AHX^- binding energies than would be expected from their gas-phase acidity, i.e. they do not fit on the same linear relationship with alcohols and carboxylic acids (Caldwell & Kebarle, 1984).

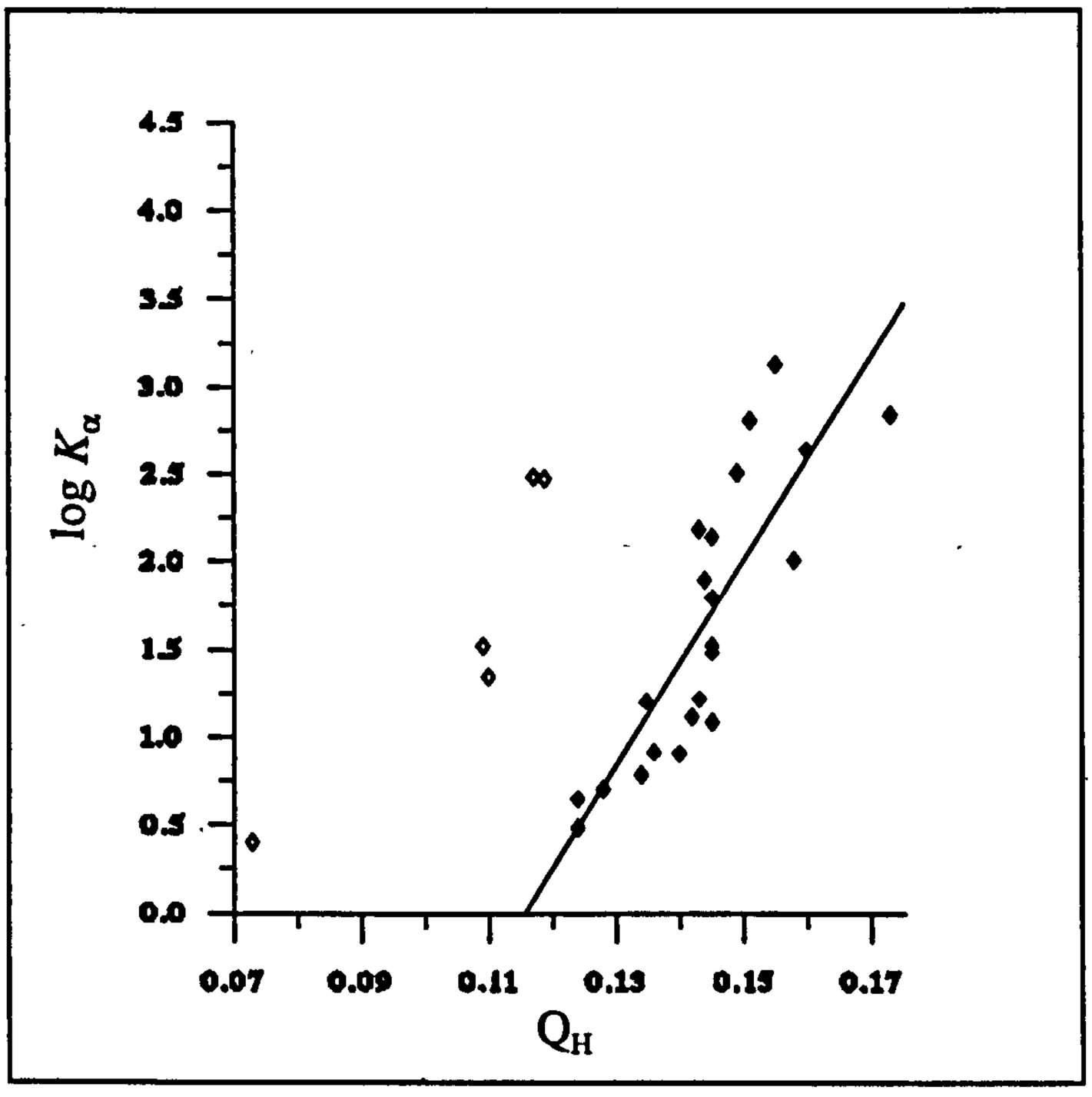


Figure 7.2. Plot of log K_α against Q_H; ♦: compounds which have been used in equation 7.20, ◊: acetanilides and chloroform.

Deleting these compounds results in the following equation:

$$\log K_\alpha = 58.4\ Q_H - 6.75 \qquad (7.20)$$

$$n = 22 \quad s = 0.444 \quad r = 0.822 \quad F = 22.9$$

Within families there were some good correlations:

Alcohols: $\log K_\alpha = 50.2\ Q_H - 5.93$ (7.21)

n = 10 s = 0.183 r = 0.961 F = 96.9

3,4-Substituted phenols: $\log K_\alpha = 105\ Q_H - 13.2$ (7.22)

n = 8 s = 0.186 r = 0.933 F = 40.4

The following equation for amines is not as good as equations 7.21 and 7.22. This could be because of the complex structures of the amines used in this correlation.

Amines (No. 34-38 & 53-59): $\log K_\alpha = 38\ Q_H - 2.22$ (7.23)

n = 12 s = 0.735 r = 0.657 F = 7.6

For amides (including sulphonamides) there was no correlation. However after deleting the sulphonamides a rather poor regression between $\log K_\alpha$ and Q_H resulted:

Amides: $\log K_\alpha = 27.2\ Q_H - 2.73$ (7.24)

n = 10 s = 0.490 r = 0.641 F = 4.9

The CNDO method calculates a more positive charge on the H-bond donor hydrogen of 4′-N,N-diethylaminoacetanilide than for example acetanilide and 3′-trifluoromethyl-4′-nitroacetanilide, which seems to be unreasonable. This could explain the resulting poor correlation for amides.

H-bond acceptors in Table 7.6 include alcohols, ethers, ketones, esters, amides, sulphoxides, sulphones, sulphonamides, phosphine oxide, phosphate, amines, nitriles, and heterocycles including some sophisticated ones. Many of the structures have more

than one potential H-bond accepting functional groups. It has been seen that the acylsulphonamide (structure 94) probably forms H-bonds to both acceptor moieties (Abraham et al, 1989a) and potential ambiguities of a similar sort attach to compounds 67, 100,112, 117, 139, 140, 142, 143 and especially the heterocycles 128-130, 132 and 136. For example for tetrazole (136) we do not know whether its acceptor abilities are confined to one nitrogen or are a function of all three. In Table 7.6 Q_{MN} values are atomic charges on only one of such H-bond accepting groups in the molecule (the heteroatom which has the highest atomic charge). Q_{MN2} values are the sum of the charges in the H-bond accepting atoms which are on different functional groups.

Unfortunately there was no correlation between log K_β and Q_{MN} or Q_{MN2}. Another set of Q_{MN2} values, in which for ester and amide groups the average of the atomic charges on the carbonyl oxygen and the charge on the sp^3 oxygen (for esters) and the sp^3 nitrogen (for amides) were used, were also examined in the regression analyses. This type of Q_{MN2} value was not successful either. The following equation is an example of the unsuccessful general regressions:

$$\log K_\beta = -1.24\ Q_{MN} + 1.64 \qquad (7.25)$$

$$n = 90 \quad r = 0.122 \quad s = 0.800 \quad F = 1.33$$

Unfortunately there was no correlation for any of the individual classes of compounds i.e. ethers and alcohols, ketones, esters, amides, amines (not even when only the simple structures with just one functional group on them were used).

Table 7.5. Log K_α values from Abraham et al (1989a) and Q_H values calculated by the CNDO method

Compound	log K_α	Q_H
(1) Methanol	1.48	0.145
(2) Ethanol	1.21	0.143
(3) Propan-1-ol	1.11	0.142
(4) Hexan-1-ol	1.20	0.135
(5) Propan-2-ol	0.91	0.136
(6) t-Butyl alcohol	0.78	0.134
(7) $PhCH_2OH$	0.90	0.140
(8) $ClCH_2CH_2OH$	1.08	0.145
(9) CF_3CH_2OH	2.00	0.158
(10) $(CF_3)_2CHOH$	2.83	0.173
(11) Phenol	2.14	0.145
(12) 2-Methylphenol	1.75	0.145
(13) 2,6-Dimethylphenol	1.08	0.145
(14) 2-Isopropylphenol	1.95	0.145
(15) 2,6-Di-isopropylphenol	0.00	0.147
(16) 2-t-Butylphenol	1.85	0.144
(17) 2,6-Di-t-butylphenol	0.00	0.150
(18) 2-Chlorophenol	2.33	0.150
(19) 2,6-Dichlorophenol	0.98	0.165
(20) 2-Cyanophenol	2.69	0.150
(21) 3-N,N-Dimethylaminophenol	1.79	0.145
(22) 3-Methylphenol	1.89	0.144
(23) 3-Isopropylphenol	1.89	0.144
(24) 3-Chlorophenol	2.50	0.149
(25) 4-Methoxyphenol	2.18	0.143
(26) 4-Trifluoromethylphenol	2.80	0.151
(27) 4-Nitrophenol	3.12	0.155
(28) a	0.98	0.134
(29) a	1.11	0.132
(30) Acetic acid	2.04	0.169
(31) Pivalic acid	1.77	0.169
(32) Benzoic acid	2.07	0.169
(33) Trifluoroacetic acid	3.55	0.188
(34) a	0.60	0.107
(35) a	0.60	0.068
(36) 4-Nitro-N-methylaniline	0.73	0.075
(37) a	1.00	0.079
(38) 2-Aminobenzothiazole	1.10	0.091
(39) CF_3CONH_2	1.52	0.145
(40) $C_6H_{13}NHCOC_6H_{13}$	0.64	0.124

Table 7.5. Continued

Compound	log K_α	Q_H
(41) MeNHCOBu^t	0.70	0.128
(42) Acetanilide	1.34	0.110
(43) 4′-N,N-Diethylaminoacetanilide	0.48	0.124
(44) 3′-Chloro-4′-nitroacetanilide	2.48	0.117
(45) 3′-Trifluoromethyl-4′-nitroacetanilide	2.47	0.119
(46) Thioacetanilide	1.52	0.109
(47) a	1.10	0.138
(48) $(CF_3CO)_2NH$	2.63	0.160
(49) Toluene-p-sulphonamide	1.15	0.156
(50) N-Benzyltoluene-p-sulphonamide	0.90	0.134
(51) N-(2-Naphthyl)toluene-p-sulphonamide	1.18	0.161
(52) $C_7H_{15}CONHSO_2Me$	1.00	0.154
(53) Pyrrole	0.95	0.098
(54) Indole	1.15	0.105
(55) a	1.20	0.105
(56) a	1.99	0.102
(57) a	2.18	0.109
(58) a	2.71	0.121
(59) a	3.55	0.111
(60) Chloroform	0.40	0.073

a: For structure see Scheme 7.1.

Table 7.6. Log K_β values from Abraham et al (1989a), Q_{MN} and Q_{MN2} values calculated by the CNDO method

Compounds	log K_β	Q_{MN}	Q_{MN2}
(2) Ethanol	1.41	-0.260	-
(5) Propan-2-ol	1.36	-0.267	-
(6) t-Butyl alcohol	1.45	-0.276	-
(61) Dibutyl ether	1.28	-0.229	-
(62) t-Butyl methyl ether	1.46	-0.233	-
(63) Tetrahydrofuran	1.69	-0.230	-
(64) Anisole	0.30	-0.222	-
(65) $MeO(CH_2)_2OMe$	1.69	-0.213	-0.423
(66) 1,4-Dioxane	1.28	-0.219	-0.438
(67) 1,4-Thioxane	1.06	-0.229	-0.322
(68) 1,3-Dioxolane	0.70	-0.236	-0.472
(69) Acetone	1.61	-0.266	-
(70) Pentan-3-one	1.50	-0.285	-

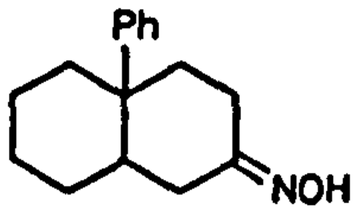
Table 7.6. Continued

Compounds	log K_β	Q_{MN}	Q_{MN2}
(71) $MeCOPr^i$	1.52	-0.281	-
(72) $MeCOBu^t$	1.44	-0.276	-
(73) Pr^iCOPr^i	1.39	-0.294	-
(74) Cyclohexanone	1.70	-0.274	-
(75) Acetophenone	1.46	-0.271	-
(76) Ethyl acetate	1.43	-0.319	-
(77) γ-Butyrolacetone	1.67	-0.303	-
(78) Dihydro-2(*3H*)-thiophenone	1.32	-0.224	-
(79) Dimethylformamide	2.81	-0.238	-
(80) Diethylformamide	2.73	-0.326	-
(81) $Bu^tCON(Me)Bu^t$	2.53	-0.358	-
(82) Dimethylthioacetamide	1.76	-0.338	-
(83) N-Methylpyrrolidinone	3.12	-0.351	-
(84) N-Dimethylbenzamide	2.82	-0.349	-
(85) Tetramethylurea	3.19	-0.393	-
(86) Tetramethylthiourea	1.96	-0.396	-
(87) a	2.38	-0.354	-
(88) $PhOCONMe_2$	2.09	-0.389	-
(89) N-Methylmaleimide	1.67	-0.296	-0.590
(90) N-Methylquinol-4-one	4.00	-0.293	-0.443
(91) Dimethyl sulphoxide	3.06	-0.321	-
(92) Tetramethylenesulphone	1.61	-0.325	-
(93) $PhSO_2N(Me)CH_2Ph$	1.36	-0.334	-
(94) a	0.99	-0.309	-0.614
(95) Triphenylphosphine oxide	3.85	-0.311	-
(96) Triethyl phosphate	3.17	-0.308	-
(97) Isopropylamine	2.84	-0.210	-
(98) Benzylamine	2.36	-0.196	-
(99) Allylamine	2.63	-0.202	-
(100) $CN(CH_2)_2NH_2$	1.74	-0.199	-0.365
(101) $CF_3CH_2NH_2$	1.01	-0.217	-0.399
(102) Pyridine	2.52	-0.145	-
(103) 2-Methoxypyridine	1.28	-0.221	-0.424
(104) 2-Fluoropyridine	1.41	-0.203	-0.388
(105) 2-Chloropyridine	1.48	-0.161	-0.291
(106) 2-cynopyridine	1.00	-0.148	-0.290
(107) 3-Methylpyridine	2.65	-0.142	-
(108) 3-Fluoropyridine	1.82	-0.201	-0.323
(109) 3-Chloropyridine	1.77	-0.159	-0.298
(110) 3- Bromopyridine	1.76	-0.137*	-
(111) 3-Cyanopyridine	1.41	-0.163	-0.304

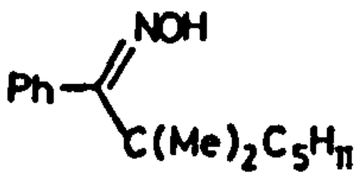
Table 7.6. Continued

Compounds	log K_β	Q_{MN}	Q_{MN2}
(112) 3-N,N-Dimethylcarbamoylpyridine	2.76	-0.359	-0.505
(113) 4-methylpyridine	2.78	-0.155	-
(114) 3,4-Dimethylpyridine	3.06	-0.152	-
(115) 4-Methoxypyridine	2.87	-0.221	-0.392
(116) 4-N,N-Dimethylaminopyridine	3.54	-0.160	-0.319
(117) 4-Acetylpyridine	2.20	-0.262	-0.394
(118) Pyrazine	1.46	-0.115	-0.230
(119) Pyrimidine	1.67	-0.171	-0.342
(120) Pyridazine	2.53	-0.077	-0.154
(121) Isoxsazole	1.06	-0.122	-0.202
(122) Oxazole	1.67	-0.178	-0.322
(123) 2,4,5-Trimethyloxazole	2.65	-0.219	-0.396
(124) Thiazole	1.90	-0.104	-0.181
(125) Benzothiazole	1.76	-0.124	-0.234
(126) 1-Methylpyrazole	2.22	-0.117	-0.197
(127) 1-Methylimidazole	3.68	-0.162	-0.296
(128) 1-Benzyl-1,2,4-triazole	2.38	-0.188	-0.431
(129) 1-Phenethyl-1,2,3-triazole	2.56	-0.094	-0.177
(130) 1-Methylbenzotriazole	2.17	-0.106	-0.195
(131) a	3.37	-0.157	-0.393
(132) a	0.57	-0.259	-0.547
(133) a	2.51	-0.104	-0.278
(134) a	1.98	-0.088	-0.264
(135) a	0.79	-0.149	-0.298
(136) a	1.99	-0.162	-0.334
(137) a	1.51	-0.191	-0.139
(138) Me_2C=NoPh	1.10	-0.179	-0.132
(139) a	2.90	-0.224	-0.423
(140) Me_2NCN	2.00	-0.175	-0.143
(141) Acetonitrile	1.23	-0.162	-
(142) $MeOCH_2CN$	1.04	-0.199	-0.340
(143) $MeO(CH_2)_2CN$	1.28	-0.217	-0.380
(144) $ClCH_2CN$	0.61	-0.143	-0.246
(145) PhCN	1.06	-0.166	-
(146) 4-Methoxybenzonitrile	1.32	-0.219	-0.388
(147) 4-Chlorobenzonitrile	0.92	-0.157	-0.310

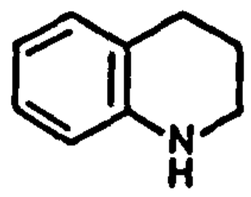
a: For structure see Scheme 7.1; -: Q_{MN2} value is equal to the corresponding Q_{MN} value (there is only one H-bond accepting group in the molecule); *: molecules containing bromine atom.



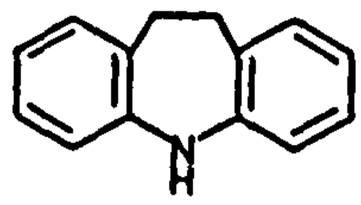
(28)



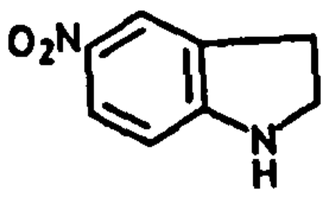
(29)



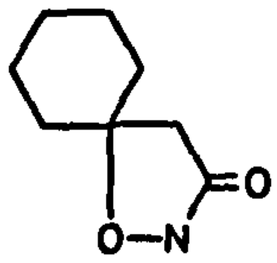
(34)



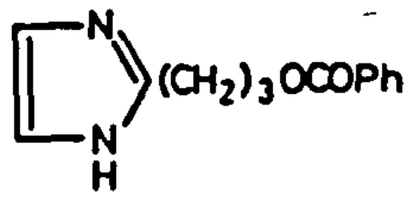
(35)



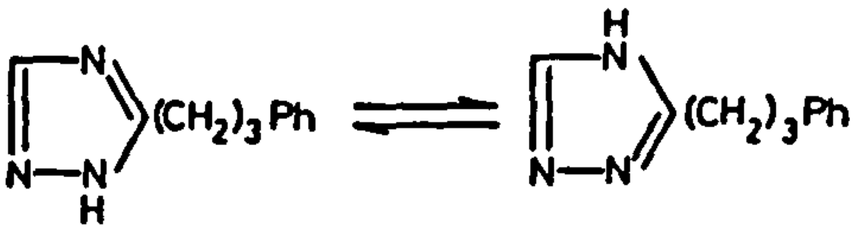
(37)



(47)



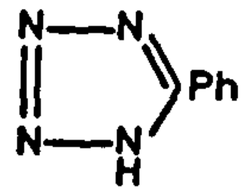
(55)



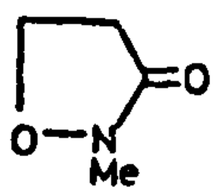
(56a) (56b)

(57)

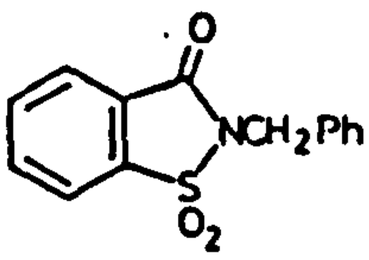
(58)



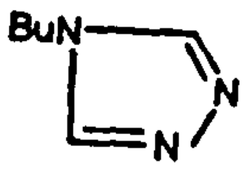
(59)



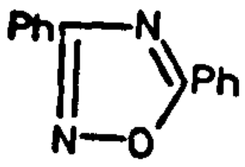
(87)



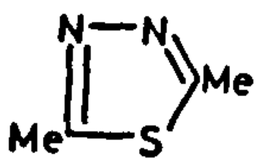
(94)



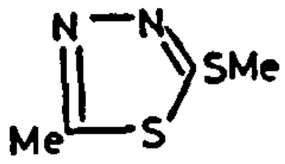
(131)



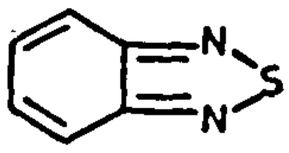
(132)



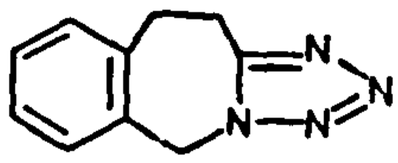
(133)



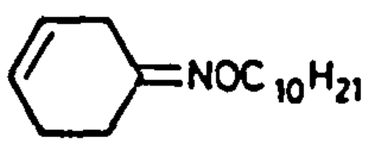
(134)



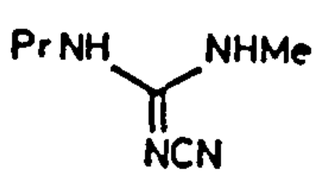
(135)



(136)



(137)



(139)

Scheme 1

7.2.3. Comparison of charge parameters with effective (summation) α^H_2 and β^H_2 values ($\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$)

119 compounds were chosen from the literature (Abraham, 1993) so that the set contained different classes of acids and bases. Charge calculations were performed using the CNDO method, and the charge parameters were obtained. The results are tabulated in Table 7.7. For all the compounds listed in the table (including those which are not H-bond donors, e.g. butanone) the following equation was obtained after deletion of the outliers which were the compounds containing bromine, 2-substituted phenols and water.

$$\Sigma\alpha^H_2 = 4.29\ Q_H - 0.0863 \qquad (7.26)$$

$n = 102 \quad s = 0.104 \quad r = 0.932 \quad F = 665$

From the above equation some other compounds, for which a negative Q_H value had been calculated, have also been excluded. These compounds are trimethylamine, benzene, naphthalene, phenanthrene, 1,4-dioxane and pyrazine. Figure 7.3 is the plot of this equation. It can be noticed in this Figure that for a group of compounds with low H-bond acidity ($\Sigma\alpha^H_2<0.10$) the Q_H values vary but the $\Sigma\alpha^H_2$ values remain almost constant. These compounds, most of which are not H-bond donors at all, should be analysed separately from the H-bond donors. After deleting the compounds with $\Sigma\alpha^H_2$ lower than 0.10, the following relationship resulted:

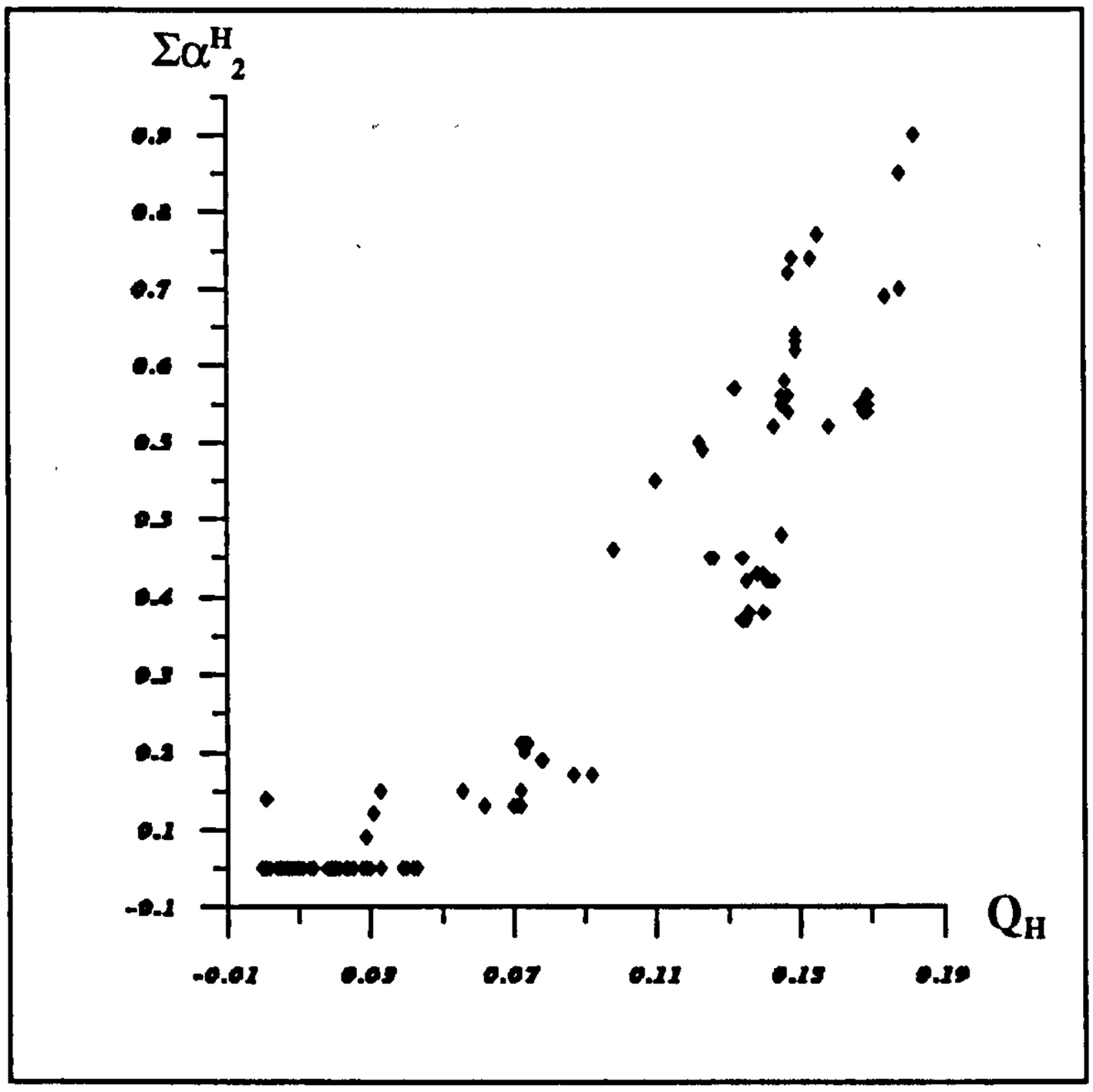


Figure 7.3. Graph of $\Sigma\alpha^H_2$ of the H-bond donors used in equation 7.26 against the corresponding Q_H values.

$$\Sigma\alpha^H_2 = 5.43\ Q_H - 0.241 \qquad (7.27)$$

$n = 56 \quad s = 0.120 \quad r = 0.849 \quad F = 140$

In case of H-bond basicity, there was not a good correlation:

$$\Sigma\beta^H_2 = -1.14\ Q_{MN} + 0.179 \qquad (7.28)$$

$n = 111 \quad s = 0.1914 \quad r = 0.482 \quad F = 32.7$

In this equation compounds with positive Q_{MN} values and also those containing bromine have not been used. Even after deleting the 2-substituted phenols, the correlation does not improve:

$$\Sigma\beta^H_2 = -1.18\ Q_{MN} + 0.177 \qquad (7.29)$$

n = 106 s = 0.1928 r = 0.495 F = 33.8

Here again within families some reasonable relationships between parameters exist.

3,4-Substituted phenols: $\Sigma\alpha^H_2 = 21.3\ Q_H - 2.47$ (7.30)

n = 13 s = 0.0553 r = 0.794 F = 18.8

$$\Sigma\beta^H_2 = -19.3\ Q_{MN} - 4.54 \qquad (7.31)$$

n = 13 s = 0.0801 r = 0.621 F = 6.9

Q_{MN2} values do not correlate with $\Sigma\beta^H_2$ values of phenols.

Alcohols: $\Sigma\alpha^H_2 = 9.75\ Q_H - 0.992$ (7.32)

n = 12 s = 0.0265 r = 0.929 F = 63.1 p = 0.0

2,2,2-Trifluoroethanol, because it is a much stronger H-bond donor than the other alcohols, has a large influence on equation 7.32; deletion of this alcohol results in the following equation:

$$\Sigma\alpha^H_2 = 6.64\ Q_H - 0.562 \qquad (7.33)$$

n = 11 s = 0.0240 r = 0.730 F = 10.3 p = 0.011

Alcohols, ethers, sulphides and thiols:

$$\Sigma\beta^H_2 = -1.31\ Q_{MN} + 0.150 \qquad (7.34)$$

n = 21 s = 0.0561 r = 0.875 F = 62.3 p = 0.0

Water was an outlier and has been excluded from all the equations (eqs. 7.32, 7.33

and 7.34). Dioxane is an outlier from the last equation and using its Q_{MN} value leads to an equation with an r value of 0.821. Using the Q_{MN2} value for dioxane improves the statistics of the resulting equation (r = 0.884) but the correlation is greatly affected by this compound because of the large difference between the charge value of this base and that of the others.

Carboxylic acids: $\Sigma\alpha^H_2 = 24.2\ Q_H - 3.47$ (7.35)

n = 12 s = 0.0374 r = 0.961 F = 120.6 p = 0

$\Sigma\beta^H_2 = -2.30\ Q_{MN} - 0.325$ (7.36)

n = 8 s = 0.0519 r = 0.775 F = 9.1 p = 0.024

Benzoic acid and methylbenzoic acids had a higher negative atomic charge on the carbonyl oxygen (Q_{MN} value) than did aliphatic acids, for example, acetic acid. On the other hand the $\Sigma\beta^H_2$ value for acetic acid is higher than that of benzoic acids. Therefore benzoic acids were outliers and have been excluded from the correlation analysis (eq. 7.36).

Amines: $\Sigma\alpha^H_2 = 25.5\ Q_H - 1.72$ (7.37)

n = 9 s = 0.0273 r = 0.783 F = 9.5 p = 0.022

The correlation for primary and secondary amines (eq. 7.37) exists only after excluding ammonia. The atomic charge on the hydrogen of ammonia (Q_H value) in comparison with the alkylamines is overestimated by the CNDO method (Table 7.7). Pyrrole is also excluded from the equation because it is a much stronger H-bond donor than the rest of the amines and therefore its inclusion, although leads to a better

correlation with r = 0.926, highly influences the equation. For H-bond acceptor ability of amines the following equation resulted:

$$\Sigma\beta^{H}_{2} = -1.92\ Q_{MN} + 0.301 \qquad (7.38)$$

$$n = 15 \quad s = 0.0915 \quad r = 0.672 \quad F = 10.7 \quad p = 0.006$$

Ammonia was also an outlier from the correlation for H-bond acceptor ability of amines and is not included in eq. 7.38. Correlations of Q_{MN2} values of amines (with $\Sigma\beta^{H}_{2}$) were also examined but they were not successful.

Amides: The CNDO calculated Q_H values for N-methylamides are higher than those of the corresponding non-substituted amides (Table 7.7), which is opposite to the inductive order observed in solution. Therefore there is no correlation between $\Sigma\alpha^{H}_{2}$ and Q_H for amides. For H-bond acceptor ability of amides the following equation resulted after omitting N,N-dimethylformamide:

$$\Sigma\beta^{H}_{2} = -4.34\ Q_{MN} - 0.826 \qquad (7.39)$$

$$n = 8 \quad s = 0.0459 \quad r = 0.804 \quad F = 10.9 \quad p = 0.016$$

Table 7.7. MO parameters calculated by CNDO method (Q_H & Q_{MN} and E_{HOMO} & E_{LUMO}) and $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ values for some H-bond acids and bases

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}(eV)	E_{LUMO}(eV)
Hept-1-yne	0.12	0.10	0.087	-0.155	-11.777	2.806
Dichloromethane	0.10	0.05	0.056	-0.090	-14.022	1.521
Trichloromethane	0.15	0.02	0.073	-0.065	-14.126	0.718
1,2-Dichloroethane	0.10	0.11	0.033	-0.125	-14.063	2.123
1,1,1-Trichloroethane	0.00	0.09	0.030	-0.090	-13.821	0.797
1-Chlorobutane	0.00	0.10	0.024	-0.159	-13.647	2.833
Tribromomethane*	0.15	0.06	-0.093	0.112	-12.477	-8.240
Diethyl ether	0.00	0.45	0.013	-0.226	-14.025	7.203
Di-n-propyl ether	0.00	0.45	0.005	-0.233	-13.747	7.067
Di-n-butyl ether	0.00	0.45	0.005	-0.229	-13.881	6.961
Propanone	0.04	0.49	0.029	-0.265	-13.217	3.804
Butanone	0.00	0.51	0.023	-0.272	-12.844	4.003
Cyclopentanone	0.00	0.52	0.020	-0.263	-12.945	3.652
Cyclohexanone	0.00	0.56	0.014	-0.274	-12.411	4.074
Methyl formate	0.00	0.38	0.006	-0.265	-14.428	4.506
Methyl acetate	0.00	0.45	0.039	-0.303	-13.742	4.332
Ethyl acetate	0.00	0.45	0.043	-0.319	-13.742	4.449
Vinyl acetate	0.00	0.43	0.040	-0.311	-13.320	4.109
Acetonitrile	0.07	0.32	0.031	-0.162	-15.892	5.842
1-Cyanobutane	0.00	0.36	0.019	-0.167	-14.632	5.802
Ammonia	0.14	0.62	0.078	-0.235	-16.025	8.169
Diethylamine	0.08	0.69	0.072	-0.180	-13.018	7.263
Methylamine	0.16	0.58	0.074	-0.195	-14.262	7.584
Ethylamine	0.16	0.61	0.073	-0.206	-13.862	7.423
n-Propylamine	0.16	0.61	0.073	-0.209	-13.486	7.119
n-Butylamine	0.16	0.61	0.072	-0.209	-13.274	6.966
Dimethylamine	0.08	0.66	0.072	-0.164	-13.424	7.328
Di-n-propylamine	0.08	0.69	0.071	-0.184	-13.151	7.021
Di-n-butylamine	0.08	0.69	0.070	-0.184	-13.124	6.939
Trimethylamine*	0.00	0.67	-0.009	-0.139	-12.746	7.328
Triethylamine	0.00	0.79	0.007	-0.177	-11.845	6.876
Formamide	0.62	0.60	0.132	-0.326	-13.587	5.437
Acetamide	0.54	0.68	0.123	-0.356	-13.092	5.029
Propionamide	0.55	0.68	0.122	-0.356	-12.909	4.993
N-Methylformamide	0.40	0.55	0.134	-0.324	-12.798	5.304
N-Methylpropanamide	0.40	0.71	0.125	-0.354	-12.526	4.912
N-Methylacetamide	0.40	0.72	0.126	-0.354	-12.547	4.947
N,N-Dimethylformamide	0.00	0.74	0.000	-0.238	-12.164	5.216
N,N-Dimethylacetamide	0.00	0.78	0.029	-0.349	-12.019	4.855
Acetic acid	0.61	0.44	0.169	-0.322	-14.107	4.359

Table 7.7. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}(eV)	E_{LUMO}(eV)
Hexanoic acid	0.60	0.45	0.167	-0.329	-13.336	4.640
Chloroacetic acid	0.74	0.36	0.174	-0.314	-13.889	2.123
Dichloroacetic acid	0.90	0.27	0.178	-0.301	-13.881	0.999
Trichloroacetic acid	0.95	0.28	0.182	-0.261	-13.916	0.433
Formic acid	0.75	0.38	0.178	-0.280	-14.858	4.523
Propanoic acid	0.60	0.45	0.169	-0.328	-13.652	4.645
Butanoic acid	0.60	0.45	0.168	-0.330	-13.486	4.697
2-Methylbenzoic acid	0.60	0.34	0.168	-0.337	-12.879	2.163
3-Methylbenzoic acid	0.59	0.38	0.168	-0.331	-12.917	2.155
4-Methylbenzoic acid	0.60	0.38	0.168	-0.336	-12.555	2.131
Water*	0.82	0.35	0.147	-0.293	-17.780	9.034
Methanol	0.43	0.47	0.145	-0.247	-15.399	7.540
Ethanol	0.37	0.48	0.143	-0.260	-14.893	7.347
Propan-1-ol	0.37	0.48	0.142	-0.264	-14.439	7.034
Propan-2-ol	0.33	0.56	0.136	-0.267	-14.319	7.189
Butan-1-ol	0.37	0.48	0.141	-0.264	-14.207	6.874
Hexan-1-ol	0.37	0.48	0.135	-0.259	-13.565	6.830
2,2,2-Trifluoroethanol	0.57	0.25	0.158	-0.227	-15.954	5.826
Cyclopentanol	0.32	0.56	0.135	-0.271	-13.949	6.514
Cyclohexanol	0.32	0.57	0.134	-0.271	-13.252	6.667
Prop-2-en-1-ol	0.38	0.48	0.140	-0.250	-14.319	4.762
trans-But-2-en-1-ol	0.38	0.48	0.138	-0.253	-13.274	4.376
Ethylthiol*	0.00	0.24	0.033	-0.060	-12.253	1.976
n-Propylthiol*	0.00	0.24	0.033	-0.069	-12.150	2.035
n-Butylthiol*	0.00	0.24	0.033	-0.070	-12.087	2.057
Diethyl sulphide	0.00	0.32	0.019	-0.098	-11.736	2.952
Di-n-Butyl sulphide	0.00	0.32	0.018	-0.114	-11.440	3.020
Trimethyl phosphate	0.00	1.00	0.000	-0.306	-14.746	0.971
Triethyl phosphate	0.00	1.06	0.021	-0.317	-14.166	1.184
Tri-n-butyl phosphate	0.00	1.21	0.013	-0.317	-13.690	1.197
Benzene*	0.00	0.14	-0.006	0.006	-13.889	3.992
Toluene	0.00	0.14	0.010	-0.019	-12.926	3.826
o-Xylene	0.00	0.16	0.010	-0.018	-12.526	3.714
m-Xylene	0.00	0.16	0.011	-0.026	-12.583	3.750
p-Xylene	0.00	0.16	0.009	-0.018	-12.207	3.676
1,3,5-Trimethylbenzene	0.00	0.19	0.011	-0.035	-12.594	3.763
Hexamethylbenzene	0.00	0.21	0.008	-0.017	-11.630	3.472
Phenylethyne	0.12	0.24	0.092	-0.142	-10.955	2.310
Naphthalene*	0.00	0.20	-0.005	-0.007	-11.453	2.250
Phenanthrene*	0.00	0.26	-0.006	-0.008	-11.151	2.125
Chlorobenzene	0.00	0.07	0.011	-0.166	-12.877	2.814

Table 7.7. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}(eV)	E_{LUMO}(eV)
Bromobenzene*	0.00	0.09	-0.026	-0.070	-12.596	-8.602
Benzaldehyde	0.00	0.39	0.007	-0.238	-13.124	2.008
Acetophenone	0.00	0.48	0.025	-0.271	-12.681	2.106
Benzophenone	0.00	0.50	0.004	-0.280	-12.120	1.880
Benzonitrile	0.00	0.33	0.001	-0.166	-13.102	2.773
Benzylamine	0.10	0.72	0.072	-0.196	-12.996	3.823
Acetanilide	0.50	0.67	0.110	-0.347	-11.513	3.772
Benzoic acid	0.59	0.40	0.169	-0.332	-13.119	2.139
Phenol	0.60	0.30	0.145	-0.253	-12.452	3.848
2-Fluorophenol*	0.61	0.26	0.156	-0.244	-12.428	3.527
3-Fluorophenol	0.68	0.17	0.149	-0.250	-12.672	3.557
4-Fluorophenol	0.63	0.23	0.146	-0.252	-12.131	3.401
2-Chlorophenol*	0.32	0.31	0.150	-0.241	-12.370	2.844
3-Chlorophenol	0.69	0.15	0.149	-0.249	-12.610	2.732
4-Chlorophenol	0.67	0.20	0.149	-0.249	-12.155	2.871
2-Bromophenol*	0.35	0.31	0.119	-0.265	-11.753	-8.732
3-Bromophenol*	0.70	0.16	0.138	-0.262	-11.731	-8.672
4-Bromophenol*	0.67	0.20	0.134	-0.261	-11.644	-8.593
2-Methoxyphenol*	0.22	0.52	0.145	-0.245	-12.000	3.747
3-Methoxyphenol	0.59	0.39	0.147	-0.252	-12.082	3.902
4-Methoxyphenol	0.57	0.48	0.143	-0.254	-11.402	3.690
2-Cyanophenol*	0.74	0.33	0.150	-0.248	-12.387	2.748
3-Cyanophenol	0.77	0.28	0.147	-0.250	-12.602	2.664
4-Cyanophenol	0.79	0.29	0.148	-0.250	-12.115	2.898
2-Nitrophenol*	0.05	0.37	0.196	-0.346	-12.917	0.721
3-Nitrophenol	0.79	0.23	0.153	-0.244	-13.296	-1.184
4-Nitrophenol	0.82	0.26	0.155	-0.245	-13.211	-1.088
1-Naphthol	0.61	0.37	0.147	-0.253	-10.847	2.343
2-Naphthol	0.61	0.40	0.145	-0.254	-11.140	2.201
Benzyl alcohol	0.33	0.56	0.140	-0.251	-13.328	3.690
Thiophenol	0.09	0.16	0.001	-0.065	-11.745	1.992
N,N-Dimethylbenzenesulphonamide	0.00	0.86	0.028	-0.337	-12.321	1.265
Tetrahydrofuran	0.00	0.48	0.002	-0.230	-14.164	6.645
1,4-Dioxane*	0.00	0.64	-0.009	-0.219	-13.040	6.585
Pyrrole	0.41	0.29	0.098	-0.071	-11.938	4.936
Pyrazine*	0.00	0.62	-0.007	-0.115	-12.009	3.162
Pyrimidine	0.00	0.65	0.000	-0.171	-12.678	3.118
Thiazole	0.00	0.45	0.042	-0.104	-12.226	2.191

*: Compounds not included in equation 7.26.

7.2.4. Replacement of H-bonding experimental parameters and indicator variables with atomic charge parameters

Because the ultimate aim of this work was the incorporation of calculated hydrogen bonding parameters into QSARs and LSERs, in order to find out the validity of these parameters, other H-bonding parameters were replaced with charge parameters in equations taken from the literature. Some examples are given below.

1- The bacterial growth inhibition activities of a set of 22 pyridine derivatives have been quantified by Schultz and Moulton (1985). The original equation was improved when the hydrogen bonding acceptor indicator variable ,H_a, was replaced with Q_{MN} calculated by CNDO method. The values of the original parameters and also Q_{MN} values are listed in Table 7.8.

$$\log BR = 0.0055MR - 0.37H_a + 1.988 \qquad (7.40)$$

$n = 20 \quad s = 0.36 \quad r = 0.842 \quad F = 20.7$

$$\log BR = 0.0055\ MR - 1.71\ Q_{MN} + 2.03 \qquad (7.41)$$

$n = 19 \quad s = 0.340 \quad r = 0.868 \quad F = 24.5$

In equation 7.41 4-bromopyridine has not been incorporated. The reason is the limitation of the CNDO method in the calculation of the atomic charge for bromine. It should be noted that deletion of the H-bond acceptor parameter from eq. 7.40 leads

to a correlation with r = 0.791. Thus the use of Q_{MN} leads to a very considerable improvement in the correlation.

2- Fungicidal activities of methyl N-X-phenylcarbamates have been represented by the following equation (Takahashi et al, 1988):

$$pI_{50} = 0.632\ \pi_p + 1.075\ \Sigma\pi_{O,M}^{a} + 0.590\ B_5^{M} - 0.087\ (B_5^{M})^2 - 0.379\ B_5^{M'} + 0.295\ HB_p + 3.247 \qquad (7.42)$$

n = 69 s = 0.346 r = 0.942

in which HB_p is a hydrogen bonding acceptor indicator variable in para position. Replacement with charge parameter did not,in this case, improve the correlation:

$$pI_{50} = 0.558\ \pi_p + 0.245\ B_5^{M} + 0.250\ B_5^{M'} - 1.13\ Q_{MN} + 1.00\ \Sigma\pi_{O,M}^{a} - 0.0323\ (B_5^{M})^2 + 3.25 \qquad (7.43)$$

n = 69 s = 0.360 r = 0.936

Since the *t-ratio* for the hydrogen bonding parameter coefficient had the lowest value among the descriptors in the both equations, it was likely that the presence of this parameter was not significant and this was the reason why the replacement of HB_p with the charge parameter did not improve the correlation. The equation resulting from the deletion of hydrogen bonding parameter confirmed this:

$$pI_{50} = 0.657\ \pi_P + 0.252\ B_5^{M} + 0.244\ B_5^{M'} + 0.989\ \Sigma\pi_{O,M}^{a} - 0.0359\ (B_5^{M})^2 + 3.31 \qquad (7.44)$$

n = 69 s = 0.372 r = 0.930

Values of the parameters used in the equations are listed in Table 7.9. Because the H-bonding acceptor ability in the para position only is significant, Q_{MN} in this table is the most negative atomic charge on the heteroatom of the substituent, if any, in that position. In compounds without a substituent in the para position and those with an alkyl substituent, i.e. where no H-bond acceptor ability was present in position 4, Q_{MN} has been given a zero value.

Of the 69 compounds listed in the table only 28 of them have substituents at position 4. Regression analysis was carried out for the 28 compounds and the following equations resulted:

$$pI_{50} = 0.885\ \pi_P + 0.430\ B_5^{M'} + 0.762\ \Sigma\pi_{O,M}^{a} + 0.087\ (B_5^{M})^2 + 0.447\ HB_p + 3.03$$

$$n = 28 \qquad s = 0.333 \qquad r = 0.939 \qquad (7.45)$$

$$pI_{50} = 0.870\ \pi_P + 0.447\ B_5^{M'} + 0.698\ \Sigma\pi_{O,M}^{a} + 0.092\ (B_5^{M})^2 - 2.29\ Q_{MN} + 2.923$$

$$n = 28 \qquad s = 0.334 \qquad r = 0.938 \qquad (7.46)$$

3- Upper respiratory tract irritation of male Swiss OF_1 mice by airborne chemicals has been well correlated by the following equation for the toxicity of nonreactive compounds (Abraham et al, 1990):

$$-\log FRD_{50} = 0.60 + 1.35\pi^{*}_{2} + 3.19\alpha^{H}_{2} + 0.77\ \log L_{16} \qquad (7.47)$$

$$n = 39 \quad r = 0.990 \quad s = 0.10$$

Replacement of α^H_2 give rise to a comparable equation (excluding bromobenzene because of the lack of parametrisation):

$$-\log FRD_{50} = 0.76 + 0.94\pi^*_2 + 7.68Q_H + 0.76\ \log L_{16} \qquad (7.48)$$

n = 38 r = 0.985 s = 0.13

4-Melting points of a series of 42 anilines can be predicted by the following equation (Dearden, 1990):

$$T_m = 331 + 181\alpha - 38.7\pi + 8.62MR - 62.1B_2 - 27.4I_3 \qquad (7.49)$$

n = 37 r = 0.931 s = 26.2

There are 3 anilines with iodine substituents in the list (Table 7.1) for which CNDO calculation is not feasible. Deleting these three, leads to the following equations:

$$T_m = 330 + 180\alpha - 38.6\pi + 8.57MR - 60.8B_2 - 27.7I_3 \qquad (7.50)$$

n = 39 r = 0.934 s = 26.9

$$T_m = 359 + 592Q_H - 35.9\pi + 9.30MR - 77.2B_2 - 23.6I_3 \qquad (7.51)$$

n = 39 r = 0.919 s = 29.9

Equation 7.51, in which Q_H has been used instead of α, has good statistics and the replacement is moderately successful.

Table 7.8. Parameters used in equations 7.40 and 7.41 (pyridine derivatives)

Substituent(s)	MR	H_a	log BR	Q_{MN}
1 H	1.03	0.0	1.673	0.000
2 4-CH_3	5.65	0.0	2.105	0.000
3 4-CH_2CH_2	10.30	0.0	2.703	0.000
4 4-$CHCH_2$	10.99	0.0	4.068	-0.016
5 4-CL	6.03	0.0	2.138	-0.161
6 4-BR	8.88	0.0	2.690	*
7 4-NO_2	7.36	1.0	3.409	-0.228
8 4-CN	6.33	1.0	2.181	-0.038
9 4-$COCH_3$	11.18	1.0	2.165	-0.262
10 4-CHO	6.88	1.0	2.841	-0.227
11 4-COC_6H_5	30.33	1.0	2.907	-0.272
12 4-$OCOCH_3$	12.47	1.0	2.186	-0.312
13 4-NH_2	5.47	1.0	2.561	-0.220
14 4-OH	2.85	1.0	1.413	-0.252
15 4-$N(CH_3)_2$	15.55	1.0	2.365	-0.160
16 4-CH_2OH	7.19	1.0	1.671	-0.255
17 4-COOH	6.93	1.0	1.614	-0.324
18 4-CHNOH	10.28	1.0	2.453	-0.214
19 4-$CONH_2$	9.81	1.0	1.985	-0.349
20 4-C_6H_5	25.36	0.0	3.664	-0.006
21 4-$CH_2C_6H_5$	30.01	0.0	3.676	-0.013
22 4-$C(CH_3)_3$	19.62	0.0	3.164	0.000

Table 7.9-Parameters used in equations 7.42-7.44 (methyl N-X-phenylcarbamates)

	Substituent(s)	PI_{50}	π_O	π_M	π_P	B_5^M	$B_5^{M'}$	HB_p	π_{OM}^a	Q_{MN}
1	H	3.37	0.00	0.00	0.00	1.00	0.00	0	0.00	0.000
2	2-F	2.90	-0.09	0.00	0.00	0.00	0.00	0	-0.09	0.000
3	2-Cl	3.37	0.38	0.00	0.00	0.00	0.00	0	0.38	0.000
4	2-Br	3.67	0.50	0.00	0.00	0.00	0.00	0	0.50	0.000
5	2-I	3.94	0.69	0.00	0.00	0.00	0.00	0	0.69	0.000
6	2-CH_3	2.89	-0.06	0.00	0.00	0.00	0.00	0	-0.06	0.000
7	2-OCH_3	3.51	0.09	0.00	0.00	0.00	0.00	0	0.09	0.000
8	2-OC_2H_5	4.37	0.65	0.00	0.00	0.00	0.00	0	0.65	0.000
9	2-OC_3H_7(N)	4.70	1.21	0.00	0.00	0.00	0.00	0	1.21	0.000
10	2-CF_3	3.84	0.38	0.00	0.00	0.00	0.00	0	0.38	0.000
11	2-NO_2	3.97	0.32	0.00	0.00	0.00	0.00	0	0.32	0.000
12	2-$COCH_3$	3.56	0.53	0.00	0.00	0.00	0.00	0	0.53	0.000
13	3-F	3.58	0.00	0.50	0.00	1.35	0.00	0	0.50	0.000
14	3-Cl	4.70	0.00	0.83	0.00	1.80	0.00	0	0.83	0.000
15	3-Br	4.69	0.00	1.03	0.00	1.95	0.00	0	1.03	0.000
16	3-I	5.33	0.00	1.44	0.00	2.15	0.00	0	1.44	0.000
17	3-CH_3	4.54	0.00	0.47	0.00	2.04	0.00	0	0.47	0.000
18	3-C_2H_5	4.95	0.00	0.97	0.00	3.17	0.00	0	0.97	0.000
19	3-OH	2.66	0.00	-0.72	0.00	1.93	0.00	0	-0.72	0.000
20	3-OCH_3	3.82	0.00	0.17	0.00	3.07	0.00	0	0.17	0.000
21	3-OC_2H_5	4.33	0.00	0.59	0.00	3.36	0.00	0	0.59	0.000
22	3-$OCHF_2$	4.29	0.00	0.48	0.00	3.61	0.00	0	0.48	0.000
23	3-OC_6H_5	4.87	0.00	1.43	0.00	5.89	0.00	0	1.43	0.000
24	3-SCH_3	4.46	0.00	0.66	0.00	3.26	0.00	0	0.66	0.000
25	3-CF_3	4.51	0.00	0.99	0.00	2.61	0.00	0	0.99	0.000
26	3-NO_2	4.20	0.00	0.20	0.00	2.44	0.00	0	0.20	0.000
27	3-CN	3.75	0.00	0.02	0.00	1.60	0.00	0	0.02	0.000
28	3-$COCH_3$	3.35	0.00	-0.12	0.00	3.13	0.00	0	-0.12	0.000
29	3-$COOC_2H_5$	3.99	0.00	0.60	0.00	4.41	0.00	0	0.60	0.000
30	4-F	3.26	0.00	0.00	0.20	0.00	0.00	0	0.00	-0.205
31	4-Cl	4.04	0.00	0.00	0.82	0.00	0.00	0	0.00	-0.169
32	4-CH_3	3.29	0.00	0.00	0.52	0.00	0.00	0	0.00	0.000
33	4-C_2H_5	3.72	0.00	0.00	0.97	0.00	0.00	0	0.00	0.000
34	4-C_4H_9(N)	4.39	0.00	0.00	1.93	0.00	0.00	0	0.00	0.000
35	4-OCH_3	2.95	0.00	0.00	-0.06	0.00	0.00	1	0.00	-0.222
36	4-OC_2H_5	4.11	0.00	0.00	0.36	0.00	0.00	1	0.00	-0.235
37	4-OC_3H_7(N)	4.59	0.00	0.00	0.97	0.00	0.00	1	0.00	-0.239
38	4-$OCHF_2$	4.00	0.00	0.00	0.36	0.00	0.00	1	0.00	-0.210
39	4-SCH_3	3.17	0.00	0.00	0.57	0.00	0.00	0	0.00	-0.116
40	4-CF_3	4.00	0.00	0.00	0.97	0.00	0.00	0	0.00	0.000
41	4-NO_2	4.31	0.00	0.00	0.50	0.00	0.00	1	0.00	-0.234
42	4-CN	3.68	0.00	0.00	0.07	0.00	0.00	1	0.00	-0.171

Table 7.9. Continued

	Substituent(s)	PI_{50}	π_O	π_M	π_P	B_5^M	$B_5^{M'}$	HB_p	π_{OM}^a	Q_{MN}
43	3,4-Cl_2	5.11	0.00	0.83	0.82	1.80	0.00	0	0.83	-0.142
44	3-CF_3,4-Cl	5.28	0.00	0.99	0.82	2.61	0.00	0	0.99	-0.141
45	3-F,4-CH_3	3.83	0.00	0.50	0.52	1.35	0.00	0	0.50	0.000
46	3-Cl,4-CH_3	4.73	0.00	0.83	0.52	1.80	0.00	0	0.83	0.000
47	3-Cl,4-OCH_3	3.94	0.00	0.83	-0.06	1.80	0.00	1	0.83	-0.210
48	3-Cl,4-OC_2H_5	4.46	0.00	0.83	0.36	1.80	0.00	1	0.83	-0.220
49	3-Cl,4-$OCHF_2$	4.60	0.00	0.83	0.36	1.80	0.00	1	0.83	-0.214
50	3-CH_3,4-OC_2H_5	4.45	0.00	0.47	0.36	2.04	0.00	1	0.47	-0.232
51	3-OCH_3,4-OC_2H_5	4.29	0.00	0.17	0.36	3.07	0.00	1	0.17	-0.222
52	3,4-$(OC_2H_5)_2$	5.54	0.00	0.59	0.36	3.36	0.00	1	0.59	0.000
53	3,5-F_2	3.92	0.00	1.00	0.00	1.35	1.35	0	1.00	0.000
54	3,5-Cl_2	6.44	0.00	1.66	0.00	1.80	1.80	0	1.66	0.000
55	3,5-Br_2	6.54	0.00	2.06	0.00	1.95	1.95	0	2.06	0.000
56	3,5-I_2	7.00	0.00	2.88	0.00	2.15	2.15	0	2.88	0.000
57	3,5-$(CH_3)_2$	5.29	0.00	0.94	0.00	2.04	2.04	0	0.94	0.000
58	3,5-$(C_2H_5)_2$	6.14	0.00	1.94	0.00	3.17	3.17	0	1.94	0.000
59	3-Cl,5-Br	6.12	0.00	1.86	0.00	1.95	1.80	0	1.86	0.000
60	3-Cl,5-CH_3	5.83	0.00	1.30	0.00	2.04	1.80	0	1.30	0.000
61	3,5-$(OCH_3)_2$	5.07	0.00	0.34	0.00	3.07	3.07	0	0.34	0.000
62	3,5-$(CF_3)_2$	5.98	0.00	1.98	0.00	2.61	2.61	0	1.98	0.000
63	3,5-$(NO_2)_2$	4.54	0.00	0.40	0.00	2.44	2.44	0	0.40	0.000
64	3-Cl,5-NO_2	5.27	0.00	1.03	0.00	2.44	1.80	0	1.03	0.000
65	3,5-Cl_2,4-F	5.89	0.00	1.66	0.20	1.80	1.80	0	1.66	-0.181
66	3,4,5-Cl_3	5.50	0.00	1.66	0.82	1.80	1.80	0	1.66	-0.117
67	3,5-Cl_2,4-CH_3	5.89	0.00	1.66	0.52	1.80	1.80	0	1.66	0.000
68	3,5,-Cl_2,4-OCH_3	5.67	0.00	1.66	-0.06	1.80	1.80	1	1.66	-0.210
69	3,5-$(CH_3)_2$,4-OC_2H_5	5.93	0.00	0.94	0.36	2.04	2.04	1	0.94	-0.216

Table 7.10- Parameters used in equations 7.47 and 7.48

Compounds	-logFRD'$_{50}$	δ_2	π^*_2	α^H_2	β^H_2	log L^{16}	V_x	Q_{MN}	Q_H
2-Propanone	3.01	0	0.71	0.04	0.50	1.760	0.547	-0.265	0.029
But-1-ene-3-one	6.67	0	0.70	0.00	0.48	2.330	0.645	-0.272	0.029
2-Butanone	3.36	0	0.67	0.00	0.48	2.287	0.688	-0.272	0.023
2-Pentanone	3.61	0	0.65	0.00	0.48	2.755	0.829	-0.275	0.036
Mesityloxide	5.60	0	0.70	0.00	0.55	3.300	0.927	-0.293	0.030
Cyclohexanone	4.51	0	0.76	0.00	0.52	3.616	0.861	-0.274	0.014
2-Hexanone	3.98	0	0.65	0.00	0.48	3.262	0.970	-0.267	0.028
4-Methyl-2-pentanone	3.88	0	0.65	0.00	0.48	3.050	0.970	-0.271	0.027
3,3-Dimethyl-2-butanone	3.64	0	0.65	0.00	0.48	2.887	0.970	-0.276	0.029
2-Heptanone	4.44	0	0.65	0.00	0.48	3.760	1.111	-0.275	0.030
4-Heptanone	4.35	0	0.65	0.00	0.48	3.820	1.111	-0.282	0.013
5-Methyl-2-hexanone	4.30	0	0.65	0.00	0.48	3.670	1.111	-0.274	0.028
2-Octanone	4.71	0	0.65	0.00	0.48	4.257	1.252	-0.276	0.032
5-Methyl-3-heptanone	4.51	0	0.65	0.00	0.48	4.200	1.251	-0.278	0.020
5-Nonanone	4.95	0	0.65	0.00	0.48	4.640	1.392	-0.276	0.032
2,5-Dimethyl-4-heptanone	4.88	0	0.65	0.00	0.48	4.180	1.392	-0.288	0.019
2-Undecanone	5.83	0	0.65	0.00	0.48	5.760	1.647	-0.275	0.026
Methanol	2.99	0	0.40	0.37	0.41	0.922	0.308	-0.247	0.145
Ethanol	3.21	0	0.40	0.33	0.44	1.485	0.449	-0.260	0.143
1-Propanol	3.71	0	0.40	0.33	0.45	2.097	0.590	-0.264	0.142
2-Propanol	3.69	0	0.40	0.32	0.47	1.821	0.590	-0.267	0.136
1-Butanol	4.29	0	0.40	0.33	0.45	2.601	0.731	-0.264	0.141
2-Methyl-1-Propanol	4.13	0	0.40	0.33	0.45	2.399	0.731	-0.264	0.143
1-Pentanol	4.60	0	0.40	0.33	0.45	3.106	0.872	-0.264	0.141
3-Methyl-1-butanol	4.52	0	0.40	0.33	0.45	3.011	0.872	-0.258	0.136
1-Hexanol	5.01	0	0.40	0.33	0.45	3.610	1.013	-0.259	0.135
4-Methyl-2-pentanol	4.76	0	0.40	0.32	0.47	3.400	1.013	-0.271	0.141
1-Heptanol	5.39	0	0.40	0.33	0.45	4.115	1.154	-0.260	0.143
1-Octanol	5.71	0	0.40	0.33	0.45	4.619	1.295	-0.260	0.143
2-Ethyl-1-hexanol	5.74	0	0.40	0.33	0.45	4.500	1.295	-0.261	0.144
Prop-2-en-1-ol	7.18	0	0.45	0.33	0.41	1.996	0.547	-0.250	0.140
But-2-en-1-ol	6.44	0	0.45	0.33	0.41	2.500	0.688	-0.252	0.138
Toluene	3.86	1	0.55	0.00	0.14	3.344	0.857	-0.019	0.010
Phenol	5.16	1	0.72	0.60	0.36	3.856	0.775	-0.253	0.145
Chlorobenzene	4.36	1	0.71	0.00	0.09	3.640	0.839	-0.166	0.011
Bromobenzene	4.78	1	0.79	0.00	0.09	4.305	0.891	0.000	0.000
1,2-Dichlorobenzene	5.13	1	0.80	0.00	0.03	4.405	0.961	-0.138	0.017
2-Chlorotoluene	4.63	1	0.67	0.00	0.08	4.160	0.980	-0.172	0.018
Acetophenone	5.38	1	0.90	0.00	0.51	4.483	1.014	-0.271	0.025
2-Xylene	4.23	1	0.51	0.00	0.17	3.937	0.998	-0.018	0.010
4-Xylene	4.27	1	0.51	0.00	0.17	3.858	0.998	-0.018	0.009
β-Chloroethylbenzene	5.47	1	0.70	0.00	0.25	4.600	1.121	-0.146	0.024
Styrene	4.62	1	0.55	0.00	0.18	3.908	0.955	-0.048	0.014
Ethylbenzene	4.24	1	0.53	0.00	0.15	3.765	0.998	-0.014	0.003
α-Methylstyrene	4.95	1	0.55	0.00	0.18	4.322	1.118	-0.075	0.013
4-Vinyltoluene	6.20	1	0.55	0.00	0.20	4.480	1.096	-0.052	0.013
4-Divinylbenzene	5.49	1	0.55	0.00	0.20	4.900	1.194	-0.048	0.014

7.3. Conclusion

As the results of regression analyses for anilines (Table 7.1) and other families of H-bond donors (Table 7.2) show, α^H_2 values correlate well with the charge parameter Q_H calculated by the CNDO method; in addition there are excellent relationships within families. However, the relationship between β^H_2 and Q_{MN} is not as good.

Although for anilines the Q_{MN2} parameter correlates slightly better than does Q_{MN} with β^H_2, it was not successful in any other group of H-bond acceptors and Q_{MN} values have been used for the rest of the regressions. One explanation for this could be that in anilines substituents are separated by an aromatic ring, while in a compound like chloroacetonitrile the chlorine atom is very close to the nitrile group and although there is opportunity for additional H-bonding, at the same time its inductive effect reduces the H-bond acceptor ability of the nitrile group. Accordingly, summation of charges was examined only in systems where the functional groups were separated by a benzene ring; unfortunately this procedure also failed (for example in correlations of $\Sigma\beta^H_2$ for substituted phenols).

Correlations with $\log K_\alpha$ and $\log K_\beta$ show less success compared with regressions with α^H_2 and β^H_2. This may be attributable to the more dipolar solvent (1,1,1-trichloroethane) used to determine the solute equilibrium constant of H-bond formation, considering that the charge parameters are calculated in vacuum. In this solvent, unlike complexations in the gas phase and in the non-polar solvents CCl_4,

benzene and cyclohexane, a plot of enthalpy against Gibbs energy of H-bond complexation has a small negative slope. This is due to the unfavourable enthalpy of desolvation which, in quite polar solvents, can approach or exceed the favourable enthalpy of H-bond formation (Abraham et al, 1988b).

$\Sigma\alpha^H_2$ correlates better than does α^H_2 with Q_H calculated by CNDO. On the other hand, β^H_2 in comparison with $\Sigma\beta^H_2$ has better correlation with Q_{MN} calculated by CNDO. This can be deduced from the following correlations for H-bond acids and bases for which both α^H_2 & β^H_2 values and $\Sigma\alpha^H_2$ & $\Sigma\beta^H_2$ values are available:

$$\alpha^H_2 = 4.97\ Q_H - 0.151 \qquad (7.52)$$

n = 41 s = 0.154 r = 0.799 F = 69.1

$$\Sigma\alpha^H_2 = 5.34\ Q_H - 0.201 \qquad (7.53)$$

n = 41 s = 0.151 r = 0.824 F = 82.6

$$\beta^H_2 = -1.38\ Q_{MN} + 0.158 \qquad (7.54)$$

n = 31 s = 0.139 r = 0.771 F = 42.7

$$\Sigma\beta^H_2 = -1.84\ Q_{MN} + 0.112 \qquad (7.55)$$

n = 31 s = 0.204 r = 0.740 F = 35.2

A drawback of the CNDO method used for charge calculation throughout this chapter is that it can not deal with iodine and bromine or with selenium atoms.

Both charge parameters work well in QSARs, indicating their validity when used in conjunction with other parameters in these correlations. Thus the use of calculated hydrogen bonding parameters to replace indicator variables and the solvatochromic parameters α^H_2 and β^H_2 appears valid, and should enable the wider and better incorporation of such parameters in QSAR studies.

8. Quantification of covalent contribution of H-bonding

Although H-bonding is mainly an electrostatic interaction, it also has contributions from other forces, among which the more dominant one is charge transfer energy. It has been concluded from atomic orbital calculations that the disagreement between the order of the experimentally found proton-accepting strengths of some proton-acceptors and the electrostatically predicted ones may be explained by taking into account the delocalisation energies (Tsubomura, 1954). Furthermore, in an *ab initio* molecular orbital study (Kollman et al, 1974), the fact that HCl forms a stronger H-bond to proton acceptors than does HF (despite the higher Mulliken population on the hydrogen of HCl) has been explained by the significantly greater charge transfer energy in H-bond complexes of HCl and also their higher acidity.

The family dependent behaviour of basicity (or acidity) dependent properties, which is responsible for the difficulty of constructing a general scale for H-bonding (and an example of which was seen in chapter 7 in the correlations between experimental H-bonding parameters and charge parameters (Figure 7.1)) is a result of the varying blend of electrostatic and charge transfer forces that is involved in any donor-acceptor combination. Maria et al (1987) have characterised this blend by angle θ; this is a measure of the electrostatic:covalent bonding ratio in the H-bonding complex.

Because charge transfer energy depends on the ionisation potential of the base and the electron affinity of the acid, it was decided to quantify it with HOMO and LUMO

energy values, using HOMO in conjunction with Q_{MN} for H-bonding acceptor ability, and LUMO with Q_H for H-bonding donor ability. HOMO is the highest occupied molecular orbital and LUMO is the lowest unoccupied molecular orbital. If the HOMO energy is high then it is easy to remove an electron from that orbital (a better charge transfer base). If the LUMO energy is low then it will readily accept an electron from another species (i.e. a better charge transfer acid) (Murrell et al, 1985).

Sabatino et al (1980) have used the SCFMO *ab initio* (STO-3G) formalism to show that for a sample of nine solvents, a relation exists between AN and DN parameters (acceptor and donor numbers of Gutmann (1978)) and the energies of LUMO and HOMO, respectively. Chastrette et al (1985), in order to take into account solute/solvent acid-base interactions, have used HOMO and LUMO energies calculated by EHT method as a parameter in the classification of solvents using a multivariate statistical treatment of quantitative solvent parameters.

8.1. Methods and experimental data

HOMO and LUMO energies (E_{HOMO} and E_{LUMO}), for H-bond acids and bases for which $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ values were available, were calculated by the CNDO method (Table 7.7). Atomic charge parameters and E_{HOMO} & E_{LUMO} were also calculated for 38 compounds which act as bioluminescence inhibitors in *Photobacterium phosphoreum* (listed in Kamlet et al, 1986). Regression analyses were carried out using Minitab.

8.2. Results and discussion

8.2.1. Comparison of $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ with E_{LUMO} and E_{HOMO} respectively

The results of the CNDO calculations are listed in Table 7.7. Correlation between $\Sigma\alpha^H_2$ and E_{LUMO} for compounds listed in the table, even after deletion of the compounds with bromine atom and 2-substituted phenols which are the obvious outliers, is not good; nevertheless this parameter improves the correlation between $\Sigma\alpha^H_2$ and Q_H (eq. 7.26) resulting in the following equation:

$$\Sigma\alpha^H_2 = 4.15\ Q_H - 0.023\ E_{LUMO} + 0.004 \qquad (8.1)$$

$n = 102 \quad s = 0.093 \quad r = 0.947 \quad F = 430$

In this equation *t*-ratio for the coefficient of E_{LUMO} is significant ($t = 5.14$, $p = 0.000$). After deleting the compounds with $\Sigma\alpha^H_2$ lower than 0.10 (as done in chapter 7 in the correlation with Q_H- equation 7.27), the following relationships resulted:

$$\Sigma\alpha^H_2 = -\ 0.049\ E_{LUMO} + 0.681 \qquad (8.2)$$

$n = 56 \quad s = 0.199 \quad r = 0.487 \quad F = 16.8$

$$\Sigma\alpha^H_2 = 4.96\ Q_H - 0.030\ E_{LUMO} - 0.055 \qquad (8.3)$$

$n = 56 \quad s = 0.102 \quad r = 0.897 \quad F = 109$

Figure 8.1 is the graph of $\Sigma\alpha^H_2$ against E_{LUMO} for the compounds used in equation 8.2. It is clear that the five carbon acids are outliers and deleting them results in the following correlation:

$$\Sigma\alpha^H_2 = -\ 0.073\ E_{LUMO} + 0.831 \qquad (8.4)$$

n = 51 s = 0.127 r = 0.791 F = 81.7

Figure 8.2 is the graph of $\Sigma\alpha^H_2$ against predicted $\Sigma\alpha^H_2$ values by equation 8.3.

For H-bond acceptors listed in Table 7.7, as was seen in chapter 7, although there are good correlations within the families, the general correlation between $\Sigma\beta^H_2$ and Q_{MN} is not good (eqs. 7.28 & 7.29). Using E_{HOMO} does not improve these equations, and in fact, there is no correlation between $\Sigma\beta^H_2$ and E_{HOMO}. The only relationship is after exclusion of aromatic structures:

$$\Sigma\beta^H_2 = -1.49\ Q_{MN} + 0.048\ E_{HOMO} - 0.793 \qquad (8.5)$$

n = 69 s = 0.189 r = 0.553 F = 14.5

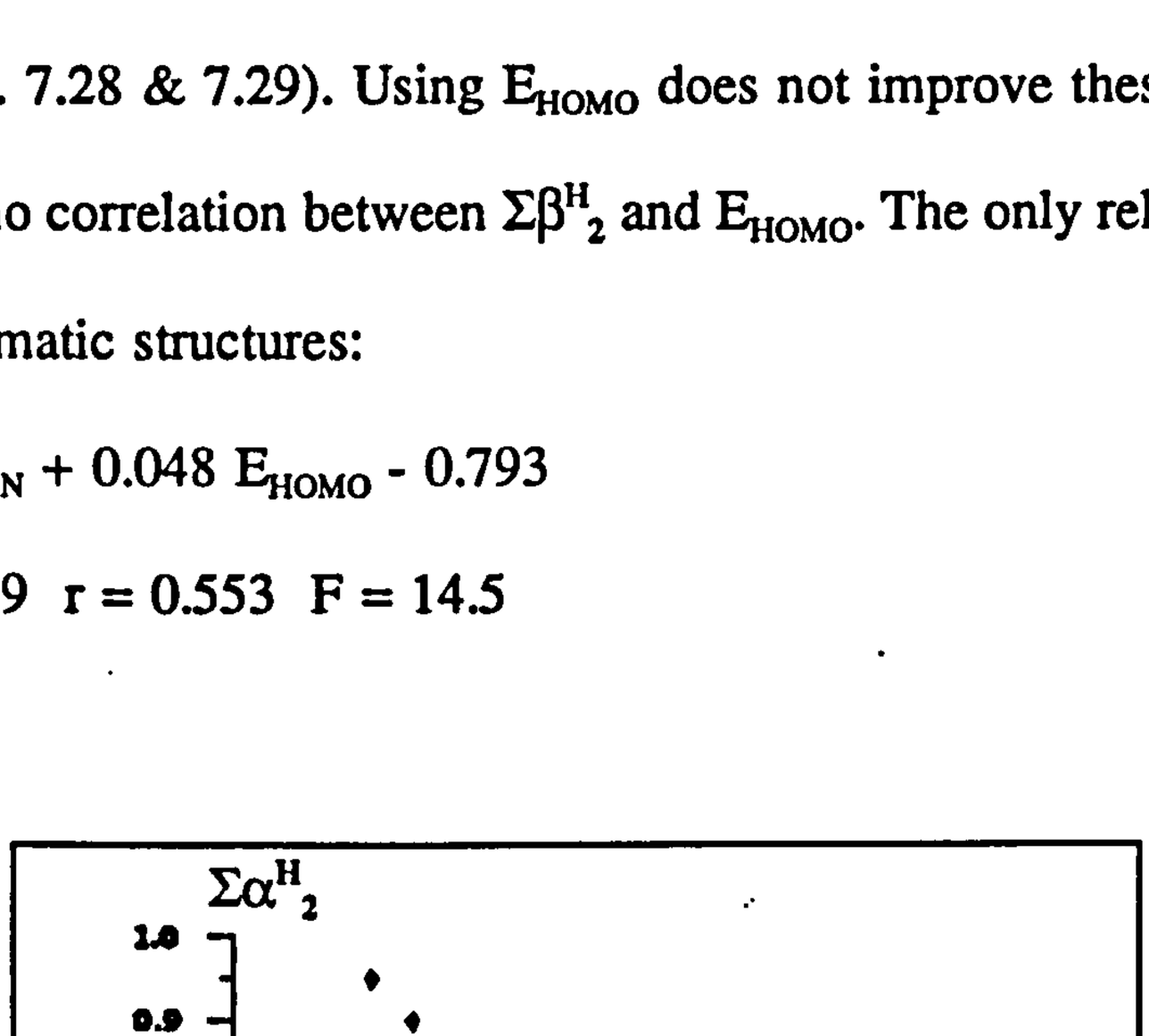


Figure 8.1. Relation between $\Sigma\alpha^H_2$ and E_{LUMO}; ♦: compounds which have been used in equation 8.4, ◊:carbon acids which are outliers.

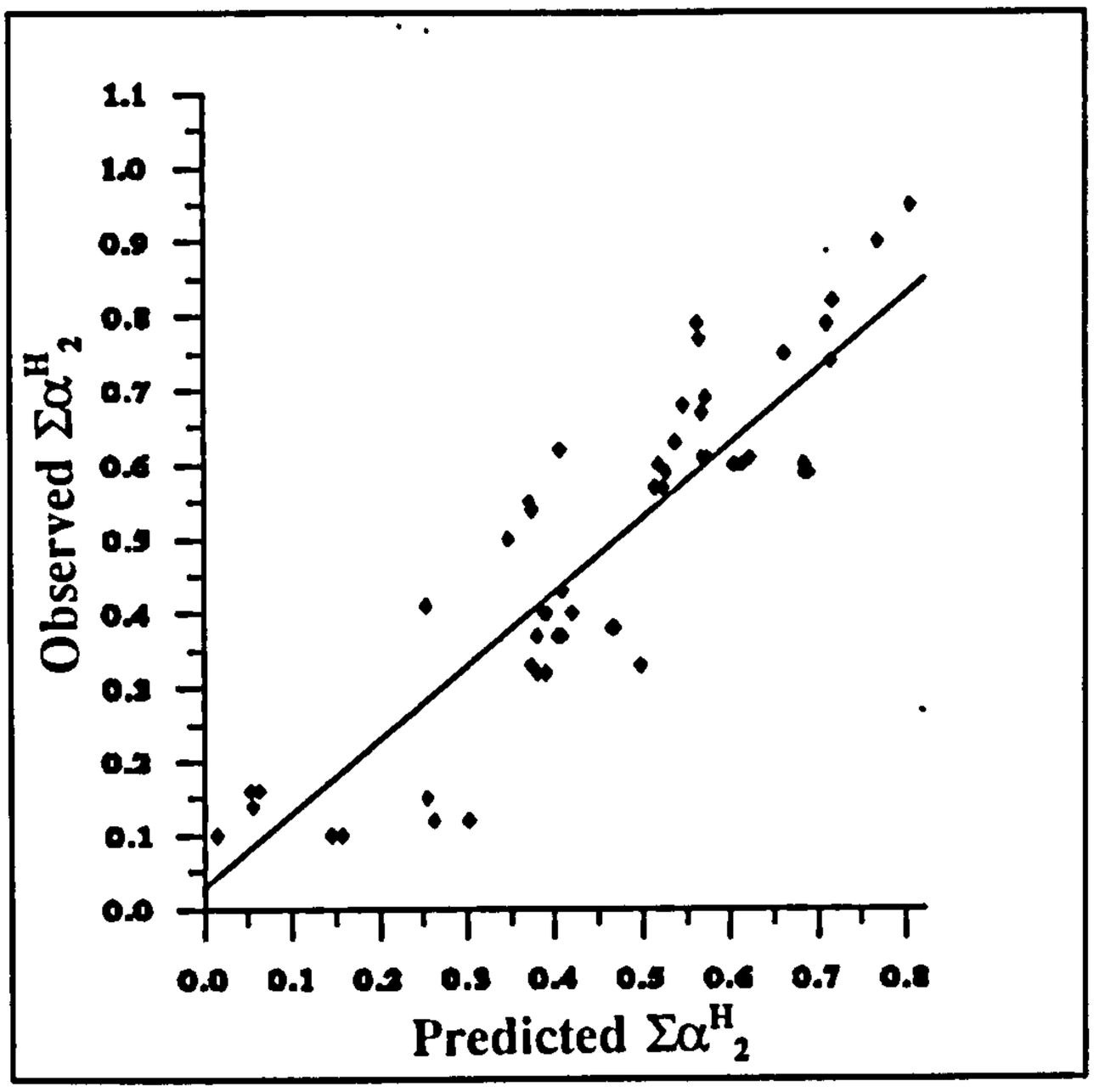


Figure 8.2. Plot of observed $\Sigma\alpha^H_2$ against predicted $\Sigma\alpha^H_2$ values from equation 8.3.

In equation 8.5 the t-ratio for the coefficients of Q_{MN} and E_{HOMO} are 5.27 (p=0.000) and 2.29 (p=0.025) respectively.

In different families some reasonable correlations were found:

3,4-Substituted phenols: $\Sigma\alpha^H_2 = - 0.036\ E_{LUMO} + 0.767$ (8.6)

$n = 13$ $s = 0.065$ $r = 0.699$ $F = 10.5$ $p = 0.008$

$\Sigma\beta^H_2 = 0.092\ E_{HOMO} + 1.41$ (8.7)

$n = 13$ $s = 0.074$ $r = 0.686$ $F = 9.8$ $p = 0.010$

Amides: There was no correlation between $\Sigma\alpha^H_2$ and E_{LUMO} for amides. The following correlation resulted for H-bond basicity of amides after deletion of N-

methylformamide and acetamide. Acetamide is the only aromatic amide in the list and conjugation has raised its HOMO energy to a higher value than its H-bonding ability suggests.

$$\Sigma\beta^{H}_{2} = 0.092\ E_{HOMO} + 1.41 \qquad (8.8)$$

$$n = 7 \quad s = 0.013 \quad r = 0.977 \quad F = 102.8$$

Alcohols: The alcohols with conjugated structures (benzyl alcohol, prop-2-en-1-ol and but-2-en-1-ol) and also trifluoroethanol fall out of the line of the correlation between $\Sigma\alpha^{H}_{2}$ and E_{LUMO}. After deleting these outliers the correlation is:

$$\Sigma\alpha^{H}_{2} = 0.199\ E_{LUMO} - 1.02 \qquad (8.9)$$

$$n = 9 \quad s = 0.052 \quad r = 0.950 \quad F = 64.9$$

In this equation E_{LUMO} has a positive coefficient, which means that the order of the E_{LUMO} values for alcohols is opposite to that of the acidity in solution.

For alcohols and ethers (except water which was an outlier), the following is the relationship between $\Sigma\beta^{H}_{2}$ and E_{HOMO}:

$$\Sigma\beta^{H}_{2} = 0.075\ E_{HOMO} + 1.55 \qquad (8.10)$$

$$n = 17 \quad s = 0.059 \quad r = 0.713 \quad F = 15.5 \quad p = 0.001$$

C*arboxylic acids:* After deletion of the aromatic acids (benzoic acids) the correlation between $\Sigma\alpha^{H}_{2}$ and E_{LUMO} improves to:

$$\Sigma\alpha^{H}_{2} = - 0.072\ E_{LUMO} + 0.957 \qquad (8.11)$$

$$n = 8 \quad s = 0.060 \quad r = 0.921 \quad F = 33.6 \quad p = 0.001$$

However there is no correlation between $\Sigma\beta^{H}_{2}$ and E_{HOMO} for carboxylic acids.

Amines: Aromatic structures were again outliers and have been excluded from the following equation:

$$\Sigma\beta^{H}_{2} = \ 0.081\ E_{HOMO} + 1.73 \qquad (8.12)$$

$$n = 11 \quad s = 0.032 \quad r = 0.850 \quad F = 26.2$$

For H-bond donors a correlation between $\Sigma\alpha^{H}_{2}$ and E_{LUMO} did not exist.

In all these classes when attempting to use charge parameter and frontier orbitals energy together, the resulting correlation is not improved and the t value of the parameters are statistically insignificant. The reason could be that in most cases there are good correlations between the two parameters. On the other hand, when regression is studied for more than one family, in some cases a two parameter correlation is successful. For example, the following equation is for alcohols, ethers, sulphides and thiols:

$$\Sigma\beta^{H}_{2} = \ 0.056\ E_{HOMO} - 1.97\ Q_{MN} + 0.77 \qquad (8.13)$$

$$n = 22 \quad s = 0.038 \quad r = 0.945 \quad F = 79$$

Photoelectron spectroscopy (PES) measures, in a rather direct way, the energies of filled orbitals (ionisation potentials, IPs) (Fleming, 1978). Graffeuil et al (1974) showed that the order of experimentally determined IPs can be reproduced by the

CNDO/2 method for alkylamines. In Table 8.1 ionisation potentials (measured by PES) and E_{HOMO} values calculated by CNDO are listed for some of the compounds used in the regressions. Although the order of the E_{HOMO} values does not follow that of the IPs in all the cases, nevertheless there is a good correlation between them, showing the reliability of the CNDO method in calculating this parameter:

$$IP = 0.686\ E_{HOMO} - 0.632 \qquad (8.14)$$

$n = 37 \quad s = 0.627 \quad r = 0.850 \quad F = 91$

When using E_{HOMO} and E_{LUMO}, although they have been proved successful in most cases in perturbation theory (Fukui, 1971), we are neglecting all the other orbital interactions. These other interactions are generally less energetically profitable than the HOMO/LUMO interactions, but there are many more of them. If other factors intervene to make the best HOMO/LUMO interaction energetically difficult to take advantage of, the interactions of lower orbitals than the HOMO (and higher orbitals than the LUMO) can become influential in determining the interaction. Secondly, there are factors of which frontier orbital theory takes little or no account, such as factors which affect the entropy (e.g. solvent effects) and steric effects. In the third place, the CNDO molecular orbital method is itself a simplification. These could explain why the correlations of E_{HOMO} with the measured H-bond basicity are so poor.

Table 8.1. Ionization potentials (IP) from PES (Stewart, 1989) and E_{HOMO} calculated by the CNDO method

Compounds	E_{HOMO}(eV)	IP (eV)
Dichloromethane	-14.022	-11.30
Trichloromethane	-14.126	-11.48
Tribromomethane	-12.477	-10.50
Diethylether	-14.025	-9.60
Propanone	-13.217	-9.72
Methyl formate	-14.428	-11.02
Methyl acetate	-13.742	-10.60
Acetonitrile	-15.892	-12.21
Ammonia	-16.025	-10.85
Methylamine	-14.262	-9.60
Ethylamine	-13.862	-9.50
Dimethylamine	-13.274	-8.93
Trimethylamine	-12.746	-8.54
Formamide	-13.587	-10.50
Acetic acid	-14.107	-10.80
Formic acid	-14.858	-11.51
Propanoic acid	-13.652	-10.50
Butanoic acid	-13.486	-9.80
Water	-17.780	-12.62
Methanol	-15.399	-10.96
Ethanol	-14.893	-10.60
Propan-1-ol	-14.439	-10.00
Ethylthiol	-12.253	-9.21
n-Propylthiol	-12.150	-9.19
n-Butylthiol	-12.087	-9.15
Benzene	-13.889	-9.25
Toluene	-12.926	-8.82
Naphthalene	-11.453	-8.15
Chlorobenzene	-12.877	-9.31
Bromobenzene	-12.596	-9.25
Benzaldehyde	-13.124	-9.70
1-Naphthol	-10.847	-7.80
2-Naphthol	-11.140	-7.90
Thiophenol	-11.745	-8.47
Pyrrole	-11.938	-8.21
Pyrazine	-12.009	-9.90
Pyrimidine	-12.678	-9.73

8.2.2. Introduction of the MO parameters into a LSER equation

The usefulness of the MO derived parameters was investigated by exchanging the older scales for them in a LSER equation. Inhibition of bioluminescence in *Photobacterium phosphoreum* (the Microtox test) as a function of toxicant properties was given by Kamlet et al (1986):

$$\log EC_{50} \text{ (in } \mu\text{mol/L)} = 7.61 - 4.11V/100 - 1.54\pi^* + 3.94\beta - 1.51\alpha_m \quad (8.15)$$

$$n = 38 \quad r = 0.987 \quad s = 0.28$$

where V is the solute molar volume and π^*, β, and α_m are the solvatochromic parameters that measure dipolarity/polarisability, hydrogen bonding basicity, and hydrogen bonding acidity of the solute (toxicant) respectively. This equation applies to compounds that act by a non-reactive toxicity mechanism. In Table 8.2 values of the original parameters and atomic charge parameters and also energies of the frontier orbitals are listed. The correlation of log EC_{50} with V and π^* (without the hydrogen bonding parameters), has a correlation coefficient of 0.876.

Table 8.2. Data used in correlation of Microtox Test Results

Toxicant	log EC_{50}	V/100	π^*	β	α_m	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Methanol	6.36	0.405	0.40	0.42	0.33	0.145	-0.247	-15.399	7.540
Ethanol	5.98	0.584	0.40	0.45	0.33	0.143	-0.260	-14.893	7.347
1-Propanol	5.16	0.748	0.40	0.45	0.33	0.142	-0.264	-14.439	7.034
2-Propanol	5.76	0.765	0.40	0.51	0.33	0.136	-0.267	-14.319	7.189
1-Butanol	4.54	0.915	0.40	0.45	0.33	0.141	-0.264	-14.207	6.874
2-Methyl-1-propanol	4.35	0.920	0.40	0.45	0.33	0.142	-0.263	-14.460	7.023
3-Pentanol	4.23	1.073	0.40	0.51	0.33	0.136	-0.271	-13.780	6.988
1-Hexanol	2.71	1.256	0.40	0.45	0.33	0.135	-0.259	-13.565	6.830
1-Heptanol	1.93	1.414	0.40	0.45	0.33	0.141	-0.264	-13.911	6.874
1-Octanol	1.62	1.575	0.40	0.45	0.33	0.141	-0.264	-13.829	6.860
2-Decanol	0.87	1.907	0.40	0.51	0.33	0.134	-0.270	-13.573	6.887
Acetone	5.57	0.734	0.71	0.48	0.00	0.031	-0.269	-13.238	3.918
2-Butanone	4.85	0.895	0.67	0.48	0.00	0.023	-0.272	-12.844	4.003
4-Methyl-2-pentanone	2.90	1.253	0.65	0.48	0.00	0.028	-0.273	-12.618	4.000
2-Octanone	2.14	1.563	0.65	0.48	0.00	0.030	-0.275	-12.419	4.049
Ethyl acetate	4.84	0.978	0.55	0.45	0.00	0.043	-0.319	-13.742	4.449
Ethyl propionate	3.84	1.146	0.47	0.46	0.00	0.025	-0.324	-13.339	4.699
Diethyl ether	4.88	1.046	0.27	0.47	0.00	0.014	-0.226	-14.025	7.203
Di-*n*-butyl ether	2.68	1.694	0.24	0.46	0.00	0.005	-0.229	-13.881	6.961
Dimethylformamide	5.43	0.774	0.88	0.69	0.00	0.011	-0.326	-12.164	5.216
CH_3CCl_3	2.90	0.996	0.49	0.10	0.00	0.030	-0.090	-13.821	0.797
$CHCl=CCl_2$	3.16	0.897	0.53	0.10	0.00	0.057	-0.082	-13.296	0.868
$ClCH_2CH_2Cl$	4.05	0.787	0.81	0.10	0.00	0.033	-0.125	-14.063	2.123
$CHCl_2CHCl_2$	1.70	1.052	0.95	0.10	0.13	0.062	-0.089	-13.766	0.852
Benzene	3.31	0.989	0.59	0.10	0.00	-0.005	0.005	-13.889	3.992
Toluene	2.29	1.163	0.54	0.11	0.00	0.010	-0.019	-12.926	3.826
o-Xylene	1.94	1.329	0.47	0.13	0.00	0.010	-0.018	-12.526	3.714
Chlorobenzene	2.12	1.118	0.71	0.07	0.00	0.011	-0.166	-12.877	2.814
1,3-Dichlorobenzene	1.35	1.226	0.80	0.03	0.00	0.027	-0.151	-12.996	2.093
1,2,3-Trichlorobenzene	1.14	1.334	0.85	0.03	0.00	0.021	-0.127	-13.100	1.339
1-CH_3-3,4-$C_6H_3Cl_2$	0.94	1.437	0.75	0.07	0.00	0.016	-0.142	-12.504	2.060
Phenol	2.63	0.989	0.75	0.33	0.61	0.145	-0.253	-12.452	3.848
o-Cresol	2.28	1.163	0.75	0.37	0.50	0.145	-0.255	-12.093	3.717
4-*t*-Bu-C_6H_4OH	0.15	1.698	0.75	0.37	0.58	0.143	-0.254	-11.736	3.853
2,4-Dimethylphenol	1.55	1.345	0.75	0.41	0.50	0.144	-0.255	-11.592	3.720
4-Nitrophenol	1.97	1.150	1.17	0.52	1.00	0.155	-0.244	-13.211	-1.088
Pyridine	4.51	0.905	0.87	0.47	0.00	0.000	-0.145	-12.629	3.592
6-Methyl-5-hepten-2-one	2.14	1.476	0.70	0.48	0.00	0.025	-0.270	-12.286	3.973
Cyclohexanol	3.06	1.140	0.40	0.51	0.33	0.134	-0.271	-13.249	6.667
Cyclohexanone	2.28	1.135	0.75	0.53	0.00	0.014	-0.274	-12.411	4.074
5-Methyl-2-hexanone	3.93	1.286	0.65	0.48	0.00	0.022	-0.270	-12.678	3.918
2-Decanone	1.70	1.894	0.63	0.48	0.00	0.024	-0.273	-12.648	4.016

For this set of toxicants α_m and β have the following relationships with Q_H & E_{LUMO} and Q_{MN} calculated by CNDO method. (Because the *t* statistic showed that the inclusion of E_{HOMO} was not significant, it was not used to describe β values in this case).

$$\alpha_m = 3.93\ Q_H - 0.0302\ E_{LUMO} + 0.0294 \qquad (8.16)$$

$n = 42 \quad r = 0.925 \quad s = 0.092$

$$\beta = 1.78Q_{MN} - 0.022 \qquad (8.17)$$

$n = 42 \quad r = 0.857 \quad s = 0.093$

We replaced solvatochromic hydrogen bonding parameters with Q_{MN} and Q_H & E_{LUMO}, which resulted in the following equation:

$$\log EC_{50} = 7.20 - 4.25V/100 - 1.08\pi^* - 5.40Q_{MN} - 7.19Q_H + 0.175E_{LUMO} \qquad (8.18)$$

$n = 38 \quad r = 0.973 \quad s = 0.40$

In this equation the *t* statistic for π^* coefficient is only 1.83. On the other hand, it is significant when the parameter E_{LUMO} is not used in the equation. This observation is due to the relatively good correlation between E_{LUMO} and π^*. The Pearson correlation coefficient matrix for the parameters used in both the original and above equations is given below:

	$\log EC_{50}$	V/100	π^*	β	α_m	Q_H	Q_{MN}	E_{LUMO}
$\log EC_{50}$	.	-0.88	-0.32	0.44	-0.13	0.03	-0.30	0.46
V/100		.	-0.01	-0.04	0.05	-0.05	-0.01	-0.07
π^*			.	-0.23	0.12	-0.19	0.16	-0.79
β				.	0.35	-0.39	-0.85	0.61
α_m					.	0.88	-0.36	0.11
Q_H						.	-0.46	0.39
Q_{MN}							.	-0.51
E_{LUMO}								.

Stepwise regression analysis omitted π^* from the relationship and yielded the following equation:

$$\log EC_{50} = 6.25 - 4.19V/100 - 4.56Q_{MN} - 7.33Q_H + 0.272E_{LUMO} \qquad (8.19)$$

$n = 38 \quad r = 0.971 \quad s = 0.42$

If instead of the compounds which have not been used in the equation given in the original paper (eq. 8.15), the outliers of the correlations with the theoretical parameters are deleted, the resulting equation is better:

$$\log EC_{50} = 6.15 - 4.19V/100 - 3.98Q_{MN} - 8.92Q_H + 0.336E_{LUMO} \qquad (8.20)$$

$n = 38 \quad r = 0.975 \quad s = 0.38$

Thus the use of E_{HOMO} and E_{LUMO} values in conjunction with charge parameters could be very useful in QSAR studies, particularly when α^H_2 and β^H_2 values are not available.

8.3. Conclusion

From the correlation analyses, it can be concluded that the energies of the frontier orbitals can be quite useful in conjunction with the charge parameters in prediction of H-bonding abilities of different classes of compounds. However, within closely related families, these two parameters cannot be employed together, because of the good correlation between Q_H and E_{LUMO} and also between Q_{MN} and E_{HOMO} within the families.

In the general regressions (correlations across families of compounds), the correlation for H-bond donors is quite successful, while correlation for H-bond acceptors is not very good (like correlations with atomic charge parameters in chapter 7).

There are good correlations between $\Sigma\alpha^H_2$ and E_{LUMO} and also between $\Sigma\beta^H_2$ and E_{HOMO} within families. However atomic charge parameters are superior in that, in order to find a good relationship, there is no need for aromatic structures to be excluded (compare for example, eq.7.32 with eq. 8.9 for alcohols or eq. 7.38 with eq. 8.12 for amines). Furthermore, in some families there is no correlation with frontier orbital energies, while the correlation with atomic charge is good (e.g. H-bond

acceptor ability of carboxylic acids and H-bond donor ability of amines).

In conclusion, the incorporation of the E_{LUMO} as well as charge parameters was found to be useful in QSAR studies.

9. Atomic charge parameters and energy of the frontier orbitals calculated by semiempirical methods in MOPAC program (MNDO, AM1 and PM3 methods)

In chapters 7 and 8, the CNDO calculated atomic charges and E_{HOMO} and E_{LUMO} values were used to predict the experimentally measured H-bond abilities of compounds. Here, other semiempirical methods (which are implemented in the MOPAC program (Stewart, 1990)) have been used to calculate the parameters. These methods are more sophisticated than CNDO, taking into account lone-pair/lone-pair repulsions; therefore a properly parametrised MNDO type model should perform better than an equivalent CNDO model. However, these methods are all parametric approaches and their quality depends not only on the theoretical framework but also on the set of the parameters used in them.

9.1. Methods and experimental data

The program COBRA in Oxford Molecular (OM) Package was used to carry out conformational analysis on the structures for which $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ values were available. The lowest energy conformation was further minimised by COSMIC force field (PIMMS program, OM) and then MOPAC 6.0 was used for the calculation of parameters. Atomic charges and HOMO and LUMO energies were calculated by the three semiempirical methods in the program, namely MNDO, AM1 and PM3. These programs were all running on a Silicon Graphics computer. The highest atomic charge on a hydrogen atom and the most negative atomic charge on a heteroatom in a

molecule (Q_H and Q_{MN} values respectively) were selected. Data were then transferred to the MINITAB statistical program (running on VAX) to carry out statistical analyses.

9.2. Results and discussion

The results of these calculations are listed in Tables 9.1, 9.2 and 9.3.

The Q_H value for the phenols studied by those methods was the atomic charge on the hydroxylic hydrogen. For nitrophenols, the most negative atoms by all the methods were nitro oxygens. For the rest of the phenols the charge on the most negatively charged atom (Q_{MN} value) resided on the hydroxylic oxygen except for the methoxyphenols for which the MNDO method calculated a more negative atomic charge on the methoxy oxygen.

All semiempirical methods underestimate the barrier to rotation of a peptide bond. A molecular mechanics correction can be added which increases the barrier (to 14 .Kcal/mol in N-methyl acetamide)(MOPAC Manual). Atomic charges calculated with or without these corrections were sometimes different (Table 9.1-9.3). The most positively charged hydrogen in the amide structures, regardless of the method used, was the hydrogen connected to the nitrogen atom (amides having two substituents on the nitrogen were excluded from correlations of H-bond donors). On the other hand, the most negatively charged atom in these compounds turned out to be different in

different calculation methods. In both with or without molecular mechanics (MM) corrections, the MNDO method calculated the most negative charge in the molecule to be on the nitrogen atom while in the PM3 method the carbonyl oxygen of the amide group was the most negative atom. AM1 method put the most negative charge on the nitrogen atom of the amides provided that it had a free hydrogen atom connected to the nitrogen. Otherwise, in amides with two substituents, the oxygen atom was the most negatively charged. Acetanilide was an exception to this rule; this molecule is the only amide in which the nitrogen atom is connected to the aromatic ring, providing the opportunity for nitrogen to donate electron through resonance.

In the alcohols investigated, the most positive atomic charge was calculated by all the methods, to be on the hydroxylic hydrogen, and the most negative atom in these compounds was the hydroxylic oxygen.

The most positive hydrogen atom in carboxylic acids was, as expected, the acidic hydrogen and the most negative heteroatom in all the acids and esters studied was the carbonyl oxygen, by all the methods.

The PM3 method underestimated the atomic charge on the nitrogen atoms and on the hydrogen atoms connected to these nitrogens, so that sometimes carbon atoms in the molecules were more negatively charged than nitrogen atoms. In amines, the atomic charge on the nitrogen atoms and the atomic charge on the hydrogens connected to the nitrogens were used as Q_{MN} and Q_H values respectively although sometimes in the

PM3 method they were not the most negative and the most positive charges in the structure.

The regression analyses between H-bonding experimental parameters ($\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$) and selected molecular orbital parameters (Q_H, Q_{MN}, E_{HOMO}, E_{LUMO}) were performed across the different classes of compounds in Tables 9.1-9.3 (including alcohols, amines, amides, carboxylic acids, esters, ethers, phenols, phosphates, sulphides, thiols) and also within the classes.

Table 9.1. MO parameters calculated by MNDO method

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Dichloromethane	0.10	0.05	0.0552	-0.1599	-12.4853	0.08096
Trichloromethane	0.15	0.02	0.0879	-0.1126	-12.9203	-0.67880
1,2-Dichloroethane	0.10	0.11	0.0488	-0.1969	-12.4154	-0.07969
1,1,1-Trichloroethane	0.00	0.09	0.0361	-0.1169	-12.7890	-0.56726
1-Chlorobutane	0.00	0.10	0.0304	-0.2197	-12.0741	0.86637
Tribromomethane	0.15	0.06	0.0856	-0.0314	-11.8621	-0.55645
Diethyl ether	0.00	0.45	0.0070	-0.3417	-10.9075	3.25456
Di-n-propyl ether	0.00	0.45	0.0023	-0.3456	-10.8158	3.15585
Di-n-butyl ether	0.00	0.45	0.0108	-0.3445	-10.9071	3.02592
Propanone	0.04	0.49	0.0167	-0.2840	-10.7521	0.65951
Butanone	0.00	0.51	0.0172	-0.2858	-10.6914	0.68751
Cyclopentanone	0.00	0.52	0.0362	-0.2770	-10.6080	0.70245
Cyclohexanone	0.00	0.56	0.0294	-0.2809	-10.5671	0.71642
Methyl formate	0.00	0.38	0.0657	-0.3292	-11.3684	1.02583
Methyl acetate	0.00	0.45	0.0263	-0.3572	-11.4593	0.90312
Ethyl acetate	0.00	0.45	0.0258	-0.3573	-11.4117	0.94237
Vinyl acetate	0.00	0.43	0.0848	-0.3494	-9.6663	0.50820
Ammonia	0.14	0.62	0.0756	-0.2268	-11.1899	4.33988
Diethylamine	0.08	0.69	0.1111	-0.3306	-10.0375	3.03639
Methylamine	0.16	0.58	0.0963	-0.2863	-10.5356	3.70699
Ethylamine	0.16	0.61	0.0940	-0.2735	-10.5329	3.45200
n-Propylamine	0.16	0.61	0.0939	-0.2743	-10.5281	3.37155
n-Butylamine	0.16	0.61	0.0944	-0.2766	-10.4560	3.20601
Dimethylamine	0.08	0.66	0.1164	-0.3538	-10.0480	3.32573
Di-n-propylamine	0.08	0.69	0.1124	-0.3367	-10.0099	2.98303
Di-n-butylamine	0.08	0.69	0.1112	-0.3351	-10.0228	2.91473
Trimethylamine	0.00	0.67	-0.0136	-0.4322	-9.6139	2.94999
Triethylamine	0.00	0.79	0.0008	-0.4170	-9.5076	2.62340
Formamide^m	0.62	0.60	0.1843	-0.4301	-10.6950	1.51728
Formamide^n	0.62	0.60	0.1548	-0.3616	-10.8411	1.15389
Acetamide^m	0.54	0.68	0.1844	-0.4213	-10.6075	1.34577
Acetamide^n	0.54	0.68	0.1498	-0.3429	-10.7580	0.97937
Propionamide^m	0.55	0.68	0.1847	-0.4196	-10.5986	1.39298
Propionamide^n	0.55	0.68	0.1847	-0.4196	-10.5986	1.39234
N-methylformamide^m	0.40	0.55	0.1890	-0.4546	-10.3794	1.42633
N-methylformamide^n	0.40	0.55	0.1846	-0.4446	-10.4081	1.36910
N-methylpropionamide^m	0.40	0.71	0.1757	-0.4415	-10.2654	1.34582
N-methylpropionamide^n	0.40	0.71	0.1715	-0.4333	-10.2840	1.31190
N-Methylacetamide^m	0.40	0.72	0.1750	-0.4434	-10.2713	1.29773
N-Methylacetamide^n	0.40	0.72	0.1686	-0.4291	-10.3142	1.22572
N,N-Dimethylformamide^m	0.00	0.74	0.0558	-0.4781	-10.1100	1.39220
N,N-Dimethylformamide^n	0.00	0.74	0.0558	-0.4781	-10.1100	1.39220
N,N-Dimethylacetamide^m	0.00	0.78	0.0204	-0.4648	-10.0465	1.21024
N,N-Dimethylacetamide^n	0.00	0.78	0.0204	-0.4648	-10.0465	1.21024
Acetic acid	0.61	0.44	0.2162	-0.3663	-11.5714	0.85103

Table 9.1. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Hexanoic acid	0.60	0.45	0.2163	-0.3662	-11.4636	0.90322
Chloroacetic acid	0.74	0.36	0.2232	-0.3398	-11.8694	-0.20194
Dichloroacetic acid	0.90	0.27	0.2285	-0.3303	-12.1320	-0.65584
Trichloroacetic acid	0.95	0.28	0.2328	-0.2991	-12.3578	-1.04370
Formic acid	0.75	0.38	0.2160	-0.3693	-11.7400	0.96031
Propanoic acid	0.60	0.45	0.2165	-0.3662	-11.4934	0.90374
Butanoic acid	0.60	0.45	0.2163	-0.3663	-11.4789	0.90317
2-Methylbenzoic acid	0.60	0.34	0.2142	-0.3617	-9.6402	-0.24074
3-Methylbenzoic acid	0.59	0.38	0.2167	-0.3780	-9.6530	-0.51896
4-Methylbenzoic acid	0.60	0.38	0.2142	-0.3617	-9.7264	-0.29770
Water	0.82	0.35	0.1628	-0.3256	-12.1913	5.44336
Methanol	0.43	0.47	0.1804	-0.3293	-11.4146	3.79527
Ethanol	0.37	0.48	0.1798	-0.3233	-11.2964	3.51491
Propan-1-ol	0.37	0.48	0.1787	-0.3239	-11.2410	3.25269
Propan-2-ol	0.33	0.56	0.1780	-0.3197	-11.2053	3.33790
Butan-1-ol	0.37	0.48	0.1785	-0.3238	-11.2312	3.19394
Hexan-1-ol	0.37	0.48	0.1785	-0.3237	-11.2170	3.14950
2,2,2-Trifluoroethanol	0.57	0.25	0.2020	-0.2973	-12.3771	1.42649
Cyclopentanol	0.32	0.56	0.1792	-0.3142	-11.1069	3.09956
Cyclohexanol	0.32	0.57	0.1785	-0.3194	-11.0846	3.06616
Prop-2-en-1-ol	0.38	0.48	0.1788	-0.3227	-10.3465	0.88857
trans-But-2-en-1-ol	0.38	0.48	0.1785	-0.3221	-9.9655	0.72986
Ethylthiol	0.00	0.24	-0.0305	0.0556	-9.7380	1.87992
n-Propylthiol	0.00	0.24	-0.0302	0.0532	-9.7303	1.89074
n-Butylthiol	0.00	0.24	-0.0302	0.0531	-9.7298	1.88579
Diethyl sulphide	0.00	0.32	0.0130	0.0267	-9.5208	1.65844
Di-n-butyl sulphide	0.00	0.32	0.0270	0.0209	-9.5116	1.66261
Trimethyl phosphate	0.00	1.00	0.0040	-0.6354	-11.2055	-0.86795
Triethyl phosphate	0.00	1.06	0.0076	-0.6421	-11.1161	-0.83555
Tri-n-butyl phosphate	0.00	1.21	0.0138	-0.6377	-11.0878	-0.84206
Benzene	0.00	0.14	0.0593	-0.0593	-9.3906	0.36749
Toluene	0.00	0.14	0.0602	-0.0606	-9.2816	0.24940
o-Xylene	0.00	0.16	0.0600	-0.0619	-9.2296	0.19096
m-Xylene	0.00	0.16	0.0601	-0.0620	-9.2398	0.19936
p-Xylene	0.00	0.16	0.0592	-0.0619	-9.1832	0.13386
1,3,5-Trimethylbenzene	0.00	0.19	0.0588	-0.0633	-9.2348	0.20262
Hexamethylbenzene	0.00	0.21	-0.0260	-0.0668	-9.0391	0.04038
Naphthalene	0.00	0.20	0.0599	-0.0475	-8.5714	-0.33095
Phenanthrene	0.00	0.26	0.0623	-0.0422	-8.4901	-0.47076
Chlorobenzene	0.00	0.07	0.0773	-0.1106	-9.6227	-0.13084
Bromobenzene	0.00	0.09	0.0739	-0.0515	-9.5502	-0.08911
Benzaldehyde	0.00	0.39	0.0715	-0.2941	-9.7265	-0.39292
Acetophenone	0.00	0.48	0.0640	-0.2818	-9.6678	-0.07220
Benzophenone	0.00	0.50	0.0644	-0.2785	-9.5863	-0.09371
Benzylamine	0.10	0.72	0.0950	-0.2749	-9.4996	0.05640

Table 9.1. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Acetanilide[m]	0.50	0.67	0.1806	-0.3894	-9.2254	0.15684
Acetanilide[n]	0.50	0.67	0.1765	-0.3768	-9.0790	0.16020
Benzoic acid	0.59	0.40	0.2142	-0.3628	-9.7684	-0.23374
Phenol	0.60	0.30	0.1930	-0.2467	-8.8825	0.25086
2-Fluorophenol	0.61	0.26	0.2064	-0.2422	-9.1463	-0.21489
3-Fluorophenol	0.68	0.17	0.1970	-0.2434	-9.2180	-0.20263
4-Fluorophenol	0.63	0.23	0.1957	-0.2448	-9.0069	-0.17272
2-Chlorophenol	0.32	0.31	0.1976	-0.2312	-9.1616	-0.17338
3-Chlorophenol	0.69	0.15	0.1968	-0.2437	-9.2224	-0.20964
4-Chlorophenol	0.67	0.20	0.1969	-0.2429	-9.1452	-0.17989
2-Bromophenol	0.35	0.31	0.1975	-0.2325	-9.0905	-0.12040
3-Bromophenol	0.70	0.16	0.1962	-0.2452	-9.1513	-0.16448
4-Bromophenol	0.67	0.20	0.1967	-0.2430	-9.0911	-0.13186
2-Methoxyphenol	0.22	0.52	0.2069	-0.3055	-8.6399	0.17799
3-Methoxyphenol	0.59	0.39	0.1929	-0.2864	-8.6971	0.17943
4-Methoxyphenol	0.57	0.48	0.1934	-0.2925	-8.8307	0.06521
2-Nitrophenol	0.05	0.37	0.2145	-0.3308	-9.7503	-0.95955
3-Nitrophenol	0.79	0.23	0.2014	-0.3275	-9.7321	-0.93110
4-Nitrophenol	0.82	0.26	0.2035	-0.3307	-9.8473	-0.82426
1-Naphthol	0.61	0.37	0.1956	-0.2501	-8.3128	-0.31369
2-Naphthol	0.61	0.40	0.1935	-0.2463	-8.4863	-0.39426
Benzyl alcohol	0.33	0.56	0.1798	-0.3230	-9.5195	0.08857
Thiophenol	0.09	0.16	-0.0347	0.1311	-9.6251	-0.15456
N,N-Dimethylbenzenesulphonamide	0.00	0.86	0.0918	-0.6778	-10.2126	-1.58937
Tetrahydrofuran	0.00	0.48	0.0208	-0.3275	-10.7749	3.10153
1,4-Dioxane	0.00	0.64	0.0168	-0.3305	-10.5518	2.97507
Pyrrole	0.41	0.29	0.1991	-0.2245	-8.5689	1.26282
Pyrazine	0.00	0.62	0.0968	-0.1828	-10.0219	-0.41745
Pyrimidine	0.00	0.65	0.1188	-0.2722	-10.3760	-0.29823
Thiazole	0.00	0.45	0.1224	-0.1951	-9.8840	-0.36227

m = with molecular mechanics correction; n = without molecular mechanics corrections.

Table 9.2. MO parameters calculated by AM1 method

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Dichloromethane	0.10	0.05	0.1284	-0.0765	-11.3860	0.59604
Trichloromethane	0.15	0.02	0.1580	-0.0408	-11.7718	-0.30712
1,2-Dichloroethane	0.10	0.11	0.1193	-0.1083	-11.4172	0.68224
1,1,1-Trichloroethane	0.00	0.09	0.1122	-0.0498	-11.9952	-0.26540
1-Chlorobutane	0.00	0.10	0.1025	-0.1272	-11.1325	1.51116
Tribromomethane	0.15	0.06	0.1726	0.0642	-11.0690	-0.74858
Diethyl ether	0.00	0.45	0.0824	-0.2828	-10.3925	2.97944

Table 9.2. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Di-n-propyl ether	0.00	0.45	0.0925	-0.2776	-10.3696	3.02172
Di-n-butyl ether	0.00	0.45	0.0940	-0.2822	-10.3885	2.88453
Propanone	0.04	0.49	0.1022	-0.2922	-10.6688	0.84368
Butanone	0.00	0.51	0.1026	-0.2915	-10.5191	0.87389
Cyclopentanone	0.00	0.52	0.1153	-0.2891	-10.4606	0.90797
Cyclohexanone	0.00	0.56	0.1056	-0.2896	-10.2946	0.91912
Methyl formate	0.00	0.38	0.1312	-0.2939	-11.2702	1.07634
Methyl acetate	0.00	0.45	0.1152	-0.3516	-11.4101	1.09760
Ethyl acetate	0.00	0.45	0.1147	-0.3552	-11.2485	1.14755
Vinyl acetate	0.00	0.43	0.1654	-0.3415	-9.9004	0.64900
Ammonia	0.14	0.62	0.1293	-0.3879	-10.4639	4.22839
Diethylamine	0.08	0.69	0.1538	-0.3124	-9.2668	3.22313
Triethylamine	0.00	0.79	0.0766	-0.2771	-8.9750	2.88470
Methylamine	0.16	0.58	0.1416	-0.3494	-9.7676	3.81769
Ethylamine	0.16	0.61	0.1440	-0.3527	-9.6878	3.63873
N-Propylamine	0.16	0.61	0.1441	-0.3535	-9.6801	3.57843
N-Butylamine	0.16	0.61	0.1389	-0.3372	-9.8223	3.63664
Dimethylamine	0.08	0.66	0.1522	-0.3046	-9.4027	3.48196
Di-n-propylamine	0.08	0.69	0.1511	-0.3060	-9.3052	3.15527
Di-n-butylamine	0.08	0.69	0.1525	-0.3094	-9.2810	3.11016
Trimethylamine	0.00	0.67	0.0692	-0.2661	-9.1207	3.19210
Formamide[m]	0.62	0.60	0.2213	-0.4481	-10.6680	1.56959
Formamide[n]	0.62	0.60	0.2213	-0.4478	-10.6688	1.56710
Acetamide[m]	0.54	0.68	0.2234	-0.4427	-10.5421	1.51451
Acetamide[n]	0.54	0.68	0.2233	-0.4427	-10.5422	1.51455
Propionamide[m]	0.55	0.68	0.2234	-0.4394	-10.5200	1.55845
Propionamide[n]	0.55	0.68	0.2232	-0.4390	-10.5197	1.55520
N-Methylformamide[m]	0.40	0.55	0.2302	-0.4021	-10.0795	1.51432
N-Methylformamide[n]	0.40	0.55	0.2295	-0.4013	-10.0818	1.50469
N-Methylpropionamide[m]	0.40	0.71	0.2229	-0.3906	-9.9033	1.54677
N-Methylpropionamide[n]	0.40	0.71	0.2182	-0.3827	-9.9215	1.49529
N-Methylacetamide[m]	0.40	0.72	0.2223	-0.3945	-9.9132	1.51065
N-Methylacetamide[n]	0.40	0.72	0.2218	-0.3936	-9.9156	1.50551
N,N-Dimethylformamide[m]	0.00	0.74	0.1209	-0.3615	-9.6175	1.51201
N,N-Dimethylformamide[n]	0.00	0.74	0.1209	-0.3615	-9.6175	1.51201
N,N-Dimethylacetamide[m]	0.00	0.78	0.1044	-0.3654	-9.5582	1.40344
N,N-Dimethylacetamide[n]	0.00	0.78	0.1044	-0.3654	-9.5582	1.40344
Acetic acid	0.61	0.44	0.2429	-0.3622	-11.6226	0.97101
Hexanoic acid	0.60	0.45	0.2421	-0.3613	-11.3999	1.01577
Chloroacetic acid	0.74	0.36	0.2489	-0.3374	-11.6124	0.13373
Dichloroacetic acid	0.90	0.27	0.2543	-0.3297	-11.7149	-0.15755
Trichloroacetic acid	0.95	0.28	0.2587	-0.3006	-11.9668	-0.81242
Formic acid	0.75	0.38	0.2417	-0.3570	-11.8202	0.95762
Propanoic acid	0.60	0.45	0.2422	-0.3610	-11.4571	1.01998
Butanoic acid	0.60	0.45	0.2422	-0.3613	-11.4618	1.01743

Table 9.2. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
2-Methylbenzoic acid	0.60	0.34	0.2447	-0.3690	-9.7167	-0.43070
3-Methylbenzoic acid	0.59	0.38	0.2452	-0.3656	-9.7446	-0.42658
4-Methylbenzoic acid	0.60	0.38	0.2458	-0.3679	-9.8342	-0.47344
Water	0.82	0.35	0.1914	-0.3828	-12.4641	4.41802
Methanol	0.43	0.47	0.1954	-0.3261	-11.1356	3.77772
Ethanol	0.37	0.48	0.1965	-0.3295	-10.8780	3.56524
Propan-1-ol	0.37	0.48	0.1947	-0.3259	-10.9720	3.59872
Propan-2-ol	0.33	0.56	0.1960	-0.3283	-10.9222	3.48352
Butan-1-ol	0.37	0.48	0.1948	-0.3254	-10.9784	3.54939
Hexan-1-ol	0.37	0.48	0.1947	-0.3256	-10.9475	3.46680
2,2,2-Trifluoroethanol	0.57	0.25	0.2163	-0.2917	-11.9627	1.38043
Cyclopentanol	0.32	0.56	0.1963	-0.3251	-10.8109	3.39808
Cyclohexanol	0.32	0.57	0.1976	-0.3272	-10.6528	3.34357
Prop-2-en-1-ol	0.38	0.48	0.1982	-0.3236	-10.3643	1.04642
trans-But-2-en-1-ol	0.38	0.48	0.1967	-0.3237	-9.7732	1.00648
Ethylthiol	0.00	0.24	0.1110	-0.0269	-8.9552	0.86765
n-Propylthiol	0.00	0.24	0.1120	-0.0245	-8.9503	0.88357
n-Butylthiol	0.00	0.24	0.1120	-0.0241	-8.9472	0.87669
Diethyl sulphide	0.00	0.32	0.0954	0.0471	-8.4440	0.86450
Di-n-butyl sulphide	0.00	0.32	0.1107	0.0541	-8.5062	0.89759
Trimethyl phosphate	0.00	1.00	0.0901	-1.0620	-11.5965	0.68377
Triethyl phosphate	0.00	1.06	0.0879	-1.0749	-11.4108	0.74278
Tri-n-butyl phosphate	0.00	1.21	0.0963	-1.0700	-11.2291	0.72890
Benzene	0.00	0.14	0.1301	-0.1301	-9.6530	0.55461
Toluene	0.00	0.14	0.1304	-0.1200	-9.3256	0.52050
o-Xylene	0.00	0.16	0.1295	-0.1096	-9.1741	0.52986
m-Xylene	0.00	0.16	0.1308	-0.1097	-9.1918	0.53209
p-Xylene	0.00	0.16	0.1318	-0.1097	-9.0543	0.48957
1,3,5-Trimethylbenzene	0.00	0.19	0.1306	-0.0996	-9.1541	0.56767
Hexamethylbenzene	0.00	0.21	0.0820	-0.0675	-8.7505	0.57494
Naphthalene	0.00	0.20	0.1328	-0.1055	-8.7066	-0.26928
Phenanthrene	0.00	0.26	0.1338	-0.0948	-8.6178	-0.40449
Chlorobenzene	0.00	0.07	0.1459	-0.1139	-9.5595	0.15657
Bromobenzene	0.00	0.09	0.1457	-0.1252	-9.6015	0.06014
Benzaldehyde	0.00	0.39	0.1541	-0.2891	-10.0043	-0.43400
Acetophenone	0.00	0.48	0.1560	-0.2982	-9.9362	-0.36384
Benzophenone	0.00	0.50	0.1514	-0.2898	-9.8734	-0.48069
Benzylamine	0.10	0.72	0.1425	-0.3333	-9.5692	0.33361
Acetanilide[m]	0.50	0.67	0.2272	-0.3507	-8.7739	0.33127
Acetanilide[n]	0.50	0.67	0.2268	-0.3518	-8.7801	0.32979
Benzoic acid	0.59	0.40	0.2457	-0.3654	-10.0841	-0.47214
Phenol	0.60	0.30	0.2167	-0.2510	-9.1145	0.39976
2-Fluorophenol	0.61	0.26	0.2296	-0.2430	-9.2724	0.01541
3-Fluorophenol	0.68	0.17	0.2214	-0.2467	-9.3696	0.02571
4-Fluorophenol	0.63	0.23	0.2198	-0.2466	-9.0939	0.06125

Table 9.2. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
2-Chlorophenol	0.32	0.31	0.2211	-0.2367	-9.1879	0.06621
3-Chlorophenol	0.69	0.15	0.2211	-0.2480	-9.2949	0.03772
4-Chlorophenol	0.67	0.20	0.2202	-0.2470	-9.1255	0.09144
2-Bromophenol	0.35	0.31	0.2221	-0.2358	-9.2453	-0.01343
3-Bromophenol	0.70	0.16	0.2212	-0.2477	-9.3400	-0.04862
4-Bromophenol	0.67	0.20	0.2216	-0.2449	-9.1946	0.01972
2-Methoxyphenol	0.22	0.52	0.2347	-0.2480	-8.7800	0.39376
3-Methoxyphenol	0.59	0.39	0.2176	-0.2498	-8.8760	0.36277
4-Methoxyphenol	0.57	0.48	0.2175	-0.2522	-8.8458	0.27233
2-Nitrophenol	0.05	0.37	0.2673	-0.3697	-9.9016	-1.19056
3-Nitrophenol	0.79	0.23	0.2255	-0.3539	-9.9364	-1.15298
4-Nitrophenol	0.82	0.26	0.2291	-0.3622	-10.0704	-1.06758
1-Naphthol	0.61	0.37	0.2200	-0.2525	-8.4504	-0.25369
2-Naphthol	0.61	0.40	0.2186	-0.2514	-8.6425	-0.34600
Benzyl alcohol	0.33	0.56	0.1986	-0.3193	-9.7073	0.28372
Thiophenol	0.09	0.16	0.1455	-0.1330	-9.0057	-0.07126
N,N-Dimethylbenzenesulphonamide	0.00	0.86	0.1539	-0.9314	-10.1559	-0.60638
Tetrahydrofurane	0.00	0.48	0.0919	-0.2832	-10.1985	3.10224
1,4-Dioxane	0.00	0.64	0.0979	-0.2704	-10.1993	2.85208
Pyrrole	0.41	0.29	0.2415	-0.1816	-8.6575	1.37852
Pyrazine	0.00	0.62	0.1714	-0.1024	-10.2502	-0.32532
Pyrimidine	0.00	0.65	0.1924	-0.1662	-10.5822	-0.22825
Thiazole	0.00	0.45	0.1913	-0.1011	-9.7026	-0.21172

m = with molecular mechanics correction; n = without molecular mechanics corrections.

Table 9.3. MO parameters calculated by PM3 method

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Dichloromethane	0.10	0.05	0.0747	-0.0212	-10.5818	0.52033
Trichloromethane	0.15	0.02	0.1044	0.0162	-10.8384	-0.11752
1,2-Dichloroethane	0.10	0.11	0.0754	-0.0512	-10.6842	0.54027
1,1,1-Trichloroethane	0.00	0.09	0.0739	0.0142	-10.7505	-0.06851
1-Chlorobutane	0.00	0.10	0.0654	-0.0773	-10.4139	1.22478
Tribromomethane	0.15	0.06	0.1378	0.0094	-10.8351	-1.17105
Diethyl ether	0.00	0.45	0.0460	-0.2688	-10.4813	2.86955
Di-n-propyl ether	0.00	0.45	0.0589	-0.2620	-10.5247	2.91542
Di-n-butyl ether	0.00	0.45	0.0620	-0.2668	-10.4935	2.70426
Propanone	0.04	0.49	0.0628	-0.3140	-10.7731	0.79199
Butanone	0.00	0.51	0.0689	-0.3118	-10.6588	0.80536
Cyclopentanone	0.00	0.52	0.0792	-0.3139	-10.6045	0.84170
Cyclohexanone	0.00	0.56	0.0716	-0.3075	-10.4622	0.84253
Methyl formate	0.00	0.38	0.0875	-0.3344	-11.1505	1.05174

Table 9.3. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Methyl acetate	0.00	0.45	0.0686	-0.3786	-11.2645	1.01227
Ethyl acetate	0.00	0.45	0.0685	-0.3832	-11.2424	1.05169
Vinyl acetate	0.00	0.43	0.1305	-0.3701	-9.9548	0.57765
Ammonia	0.14	0.62	-0.0023	0.0067	-9.7044	3.33541
Methylamine	0.16	0.58	0.0344	-0.0292	-9.4033	3.10619
Ethylamine	0.16	0.61	0.0414	-0.0367	-9.3809	3.02864
n-Propylamine	0.16	0.61	0.0543	-0.0362	-9.3912	2.97977
n-Butylamine	0.16	0.61	0.0595	-0.0347	-9.4906	3.04540
Dimethylamine	0.08	0.66	0.0418	-0.0553	-9.2172	2.90770
Di-n-propylamine	0.08	0.69	0.0455	-0.0646	-9.1847	2.68286
Di-n-butylamine	0.08	0.69	0.0463	-0.0659	-9.1888	2.65692
Trimethylamine	0.00	0.67	0.0375	-0.0716	-9.0609	2.71557
Diethylamine	0.08	0.69	0.0475	-0.0711	-9.1430	2.73640
Triethylamine	0.00	0.79	0.0431	-0.0821	-9.0015	2.47543
Formamide[m]	0.62	0.60	0.0655	-0.3954	-9.8459	1.36077
Formamide[n]	0.62	0.60	0.0558	-0.3755	-10.0671	1.07700
Acetamide[m]	0.54	0.68	0.0660	-0.3936	-9.7504	1.29451
Acetamide[n]	0.54	0.68	0.0519	-0.3668	-10.0860	0.95581
Propionamide[m]	0.55	0.68	0.0663	-0.3930	-9.7361	1.31887
Propionamide[n]	0.55	0.68	0.0517	-0.3633	-10.0860	0.96249
N-Methylformamide[m]	0.40	0.55	0.0976	-0.3870	-9.5575	1.26958
N-Methylformamide[n]	0.40	0.55	0.0822	-0.3653	-9.8142	0.99414
N-Methylpropionamide[m]	0.40	0.71	0.0733	-0.3742	-9.5404	1.16727
N-Methylpropionamide[n]	0.40	0.71	0.0654	-0.3607	-9.7237	1.02782
N-Methylacetamide[m]	0.40	0.72	0.0794	-0.3850	-9.4333	1.30277
N-Methylacetamide[n]	0.40	0.72	0.0623	-0.3601	-9.7645	0.98540
N,N-Dimethylformamide[m]	0.00	0.74	0.0773	-0.3609	-9.5429	1.04719
N,N-Dimethylformamide[n]	0.00	0.74	0.0773	-0.3609	-9.5429	1.04719
N,N-Dimethylacetamide[m]	0.00	0.78	0.0649	-0.3646	-9.4742	1.00565
N,N-Dimethylacetamide[n]	0.00	0.78	0.0649	-0.3646	-9.4742	1.00565
Acetic acid	0.61	0.44	0.2266	-0.3982	-11.4374	0.92635
Hexanoic acid	0.60	0.45	0.2254	-0.3946	-11.3481	0.94314
Chloroacetic acid	0.74	0.36	0.2301	-0.3724	-10.8299	0.12039
Dichloroacetic acid	0.90	0.27	0.2330	-0.3605	-10.8911	-0.10950
Trichloroacetic acid	0.95	0.28	0.2357	-0.3391	-10.9401	-0.45849
Formic acid	0.75	0.38	0.2239	-0.3966	-11.5636	0.96910
Propanoic acid	0.60	0.45	0.2258	-0.3954	-11.3324	0.95701
Butanoic acid	0.60	0.45	0.2256	-0.3950	-11.3566	0.94245
2-Methylbenzoic acid	0.60	0.34	0.2242	-0.3936	-9.7940	-0.25433
3-Methylbenzoic acid	0.59	0.38	0.2281	-0.4042	-9.8079	-0.48746
4-Methylbenzoic acid	0.60	0.38	0.2282	-0.4055	-9.8654	-0.52780
Water	0.82	0.35	0.1793	-0.3586	-12.3165	4.06162
Methanol	0.43	0.47	0.1810	-0.3087	-11.1367	3.50778
Ethanol	0.37	0.48	0.1834	-0.3122	-10.8956	3.33409
Propan-1-ol	0.37	0.48	0.1812	-0.3086	-11.1227	3.25152

Table 9.3. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
Propan-2-ol	0.33	0.56	0.1840	-0.3102	-11.0773	3.27702
Butan-1-ol	0.37	0.48	0.1814	-0.3085	-11.1390	3.17321
Hexan-1-ol	0.37	0.48	0.1814	-0.3083	-11.1372	3.08117
2,2,2-Trifluoroethanol	0.57	0.25	0.1971	-0.2755	-12.2368	0.91867
Cyclopentanol	0.32	0.56	0.1839	-0.3060	-10.9422	3.22560
Cyclohexanol	0.32	0.57	0.1853	-0.3089	-10.8998	3.14805
Prop-2-en-1-ol	0.38	0.48	0.1839	-0.3068	-10.4790	0.87238
trans-But-2-en-1-ol	0.38	0.48	0.1828	-0.3079	-9.8302	0.86936
Ethylthiol	0.00	0.24	0.0815	-0.0251	-9.2665	0.37326
n-Propylthiol	0.00	0.24	0.0838	-0.0183	-9.2641	0.37678
n-Butylthiol	0.00	0.24	0.0834	-0.0205	-9.2687	0.36956
Diethyl sulphide	0.00	0.32	0.0675	-0.0525	-8.8548	0.36377
Di-n-butyl-sulphide	0.00	0.32	0.0842	-0.0417	-8.9740	0.41244
Trimethyl phosphate	0.00	1.00	0.0174	-0.8508	-10.8199	0.23105
Triethyl phosphate	0.00	1.06	0.0433	-0.8534	-10.8368	0.21635
Tri-n-butyl phosphate	0.00	1.21	0.0594	-0.8559	-10.8146	0.19826
Benzene	0.00	0.14	0.1021	-0.1021	-9.7489	0.39459
Toluene	0.00	0.14	0.1048	-0.0978	-9.4427	0.37647
o-Xylene	0.00	0.16	0.1045	-0.0928	-9.2848	0.38981
m-Xylene	0.00	0.16	0.1077	-0.0934	-9.3079	0.39488
p-Xylene	0.00	0.16	0.1049	-0.0934	-9.1819	0.35958
1,3,5-Trimethylbenzene	0.00	0.19	0.1084	-0.1144	-9.2667	0.42728
Hexamethylbenzene	0.00	0.21	0.0525	-0.0707	-8.8617	0.47118
Naphthalene	0.00	0.20	0.1053	-0.0830	-8.8264	-0.41469
Phenanthrene	0.00	0.26	0.1115	-0.0753	-8.7424	-0.53136
Chlorobenzene	0.00	0.07	0.1167	-0.1020	-9.3876	0.06303
Bromobenzene	0.00	0.09	0.1229	-0.0932	-9.8258	-0.06173
Benzaldehyde	0.00	0.39	0.1198	-0.3168	-10.0482	-0.48586
Acetophenone	0.00	0.48	0.1172	-0.3121	-10.0090	-0.32720
Benzophenone	0.00	0.50	0.1149	-0.3040	-9.9310	-0.41233
Benzylamine	0.10	0.72	0.1065	-0.1013	-9.4533	0.14283
Acetanilide[m]	0.50	0.67	0.0726	-0.3667	-8.8329	0.17145
Acetanilide[n]	0.50	0.67	0.0658	-0.3526	-9.0085	0.15924
Benzoic acid	0.59	0.40	0.2286	-0.4036	-10.1343	-0.53825
Phenol	0.60	0.30	0.1961	-0.2275	-9.1752	0.28743
2-Fluorophenol	0.61	0.26	0.2055	-0.2157	-9.3992	-0.06187
3-Fluorophenol	0.68	0.17	0.1996	-0.2235	-9.4641	-0.06149
4-Fluorophenol	0.63	0.23	0.1984	-0.2215	-9.2675	-0.04781
2-Chlorophenol	0.32	0.31	0.1995	-0.2158	-9.0404	0.02418
3-Chlorophenol	0.69	0.15	0.1988	-0.2240	-9.2365	-0.00710
4-Chlorophenol	0.67	0.20	0.1984	-0.2233	-9.0048	0.05176
2-Bromophenol	0.35	0.31	0.1955	-0.2023	-9.3088	0.02727
3-Bromophenol	0.70	0.16	0.2001	-0.2221	-9.4098	-0.12671
4-Bromophenol	0.67	0.20	0.1990	-0.2228	-9.3169	-0.02405
2-Methoxyphenol	0.22	0.52	0.2074	-0.2233	-8.8615	0.31582

Table 9.3. Continued

Compound	$\Sigma\alpha^H_2$	$\Sigma\beta^H_2$	Q_H	Q_{MN}	E_{HOMO}	E_{LUMO}
3-Methoxyphenol	0.59	0.39	0.1959	-0.2260	-8.9469	0.27935
4-Methoxyphenol	0.57	0.48	0.1966	-0.2265	-8.9920	0.17524
2-Nitrophenol	0.05	0.37	0.2598	-0.6096	-9.9043	-1.22531
3-Nitrophenol	0.79	0.23	0.2017	-0.5959	-9.9625	-1.18081
4-Nitrophenol	0.82	0.26	0.2061	-0.6050	-10.1669	-1.08240
1-Naphthol	0.61	0.37	0.1992	-0.2267	-8.5366	-0.36470
2-Naphthol	0.61	0.40	0.1971	-0.2273	-8.7179	-0.45091
Benzyl alcohol	0.33	0.56	0.1839	-0.3051	-9.8179	0.16730
Thiophenol	0.09	0.16	0.1187	-0.1034	-9.3424	-0.24438
N,N-Dimethylbenzenesulphonamide	0.00	0.86	0.1213	-0.8203	-9.6595	-0.70024
Tetrahydrofuran	0.00	0.48	0.0566	-0.2729	-10.2923	3.25190
1,4-Dioxane	0.00	0.64	0.0529	-0.5132	-10.4477	2.84023
Pyrrole	0.41	0.29	0.1421	-0.2370	-8.9183	1.10978
Pyrazine	0.00	0.62	0.1295	-0.1081	-10.1530	-0.44424
Pyrimidine	0.00	0.65	0.1389	-0.1180	-10.2894	-0.40819
Thiazole	0.00	0.45	0.1550	-0.0566	-10.0041	-0.60536

m = with molecular mechanics correction; n = without molecular mechanics corrections.

9.2.1. MNDO method

a. Correlations within families of compounds

Phenols: Regression analyses showed that the 2-substituted phenols were outliers and could not be analysed with other phenols. This can be explained by the ability of such molecules to form intramolecular H-bonding, and by steric hindrance of the substituent which is not manifested in the atomic charges. The following equation correlates $\Sigma\alpha^H_2$ and Q_H of 3- and 4-substituted phenols.

$$\Sigma\alpha^H_2 = 23.1Q_H - 3.87 \qquad (9.1)$$

$n = 13 \quad r = 0.964 \quad s = 0.0208 \quad F = 146$

Although there was quite a good correlation between $\Sigma\alpha^H_2$ and E_{LUMO},

$$\Sigma\alpha^H_2 = 0.62 - 0.183E_{LUMO} \qquad (9.2)$$

$n = 13 \quad r = 0.829 \quad s = 0.0440 \quad F = 24$

the introduction of E_{LUMO} as the second descriptor to the first equation did not improve the regression and the t-ratio for E_{LUMO} in such a equation was only 0.17.

For H-bonding acceptor ability, both Q_{MN} and E_{HOMO} were significant in the prediction of $\Sigma\beta^H_2$ values (t-ratios are 6.97 and 8.66 respectively):

$$\Sigma\beta^H_2 = 0.267E_{HOMO} - 2.82Q_{MN} + 1.94 \qquad (9.3)$$

$n = 13 \quad r = 0.944 \quad s = 0.0385 \quad F = 40.9 \quad p = 0.000$

When correlations were attempted with the individual parameters alone, 3-nitrophenol and 4-nitrophenol were outliers. The extra resonance in nitro group should increase the E_{HOMO} but the other property of this group is the electron-withdrawing effect. It is known (Fleming, 1978) that $-CF_3$ and conjugated groups like -CHO, -CN, $-NO_2$

which are also electron withdrawing; when substituted on a benzene ring, reduce the ionisation potential measured by PES. In fact the E_{HOMO} calculated by the MNDO method follow the trend of experimental IP values, and nitrophenols have exceptionally low E_{HOMO} values (Figure 9.1). This discrepancy is cancelled out if Q_{MN} for nitro oxygen is incorporated as shown by equation 9.3.

The following is the predictor equation for H-bond acceptor ability of 3- and 4-substituted phenols (with the two outliers excluded):

$$\Sigma\beta^{H}_{2} = 0.238\ E_{HOMO} - 3.44\ Q_{MN} + 1.53 \qquad (9.4)$$

$n = 11 \quad r = 0.954 \quad s = 0.0385 \quad F = 41$

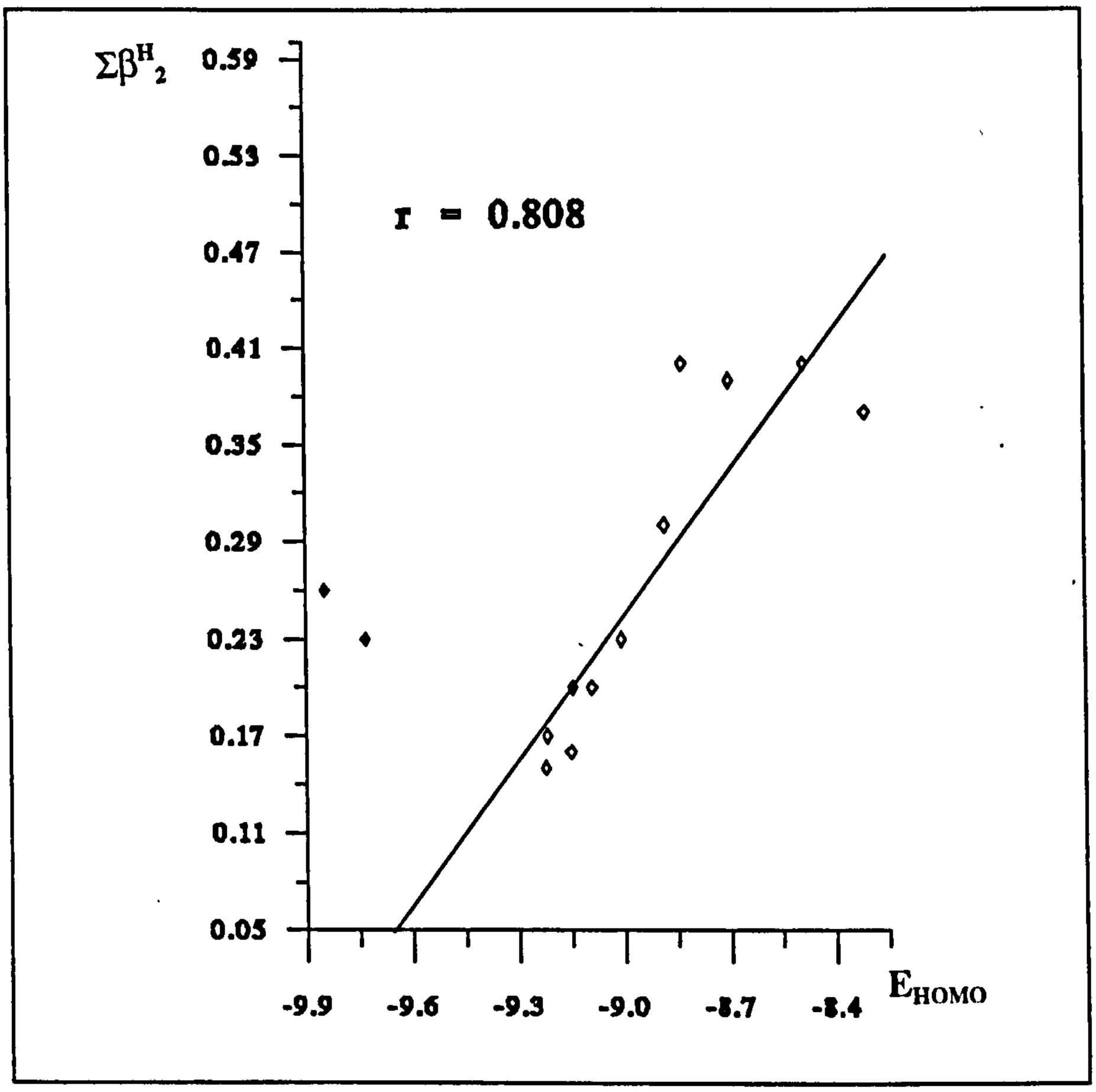


Figure 9.1. The relationship between $\Sigma\beta^{H}_{2}$ and E_{HOMO} of 3- and 4-substituted phenols(◊); 3- and 4-nitrophenols (♦) are outliers.

Amides: Among the amides studied, N-methylformamide has always been an outlier from the equations. The Q_{MN} value for this amide is unreasonably lower than that of amides like N-methylacetamide and N-methylpropanamide, both with and without molecular mechanics correction. Applying molecular mechanics corrections, the relationship between $\Sigma\alpha^H_2$ and Q_H is:

$$\Sigma\alpha^H_2 = 18.5Q_H - 2.84 \qquad (9.5)$$

$n = 6 \quad r = 0.942 \quad s = 0.0329 \quad F = 31.6 \quad p = 0.005$

N-methylformamide has not been included in the regression. For this compound the MNDO method overestimated Q_H. In this case E_{LUMO} did not correlate with $\Sigma\alpha^H_2$ and could not improve the correlation with Q_H.

H-bonding acceptor ability of amides can be expressed by the following equations (molecular mechanics correction included):

$$\Sigma\beta^H_2 = 0.146 - 1.26Q_{MN} \qquad (9.6)$$

$n = 8 \quad r = 0.653 \quad s = 0.0439 \quad F = 4.5 \quad p = 0.079$

$$\Sigma\beta^H_2 = 0.199E_{HOMO} + 2.77 \qquad (9.7)$$

$n = 7 \quad r = 0.915 \quad s = 0.0250 \quad F = 25.9 \quad p = 0.004$

Acetanilide has been excluded from equation 9.7 and N-methylformamide from

equations 9.6 and 9.7. Acetanilide is a pronounced outlier, and this could be due to the benzene ring in its structure; it is the only aromatic amide in the list. Conjugation reduces the distance between HOMO and LUMO energies, by increasing the energy of the highest occupied molecular orbital and decreasing the energy of the lowest unoccupied molecular orbital (Vollhardt, 1987).

For amides, correlations between charge parameters (calculated after molecular mechanics corrections) and experimental H-bonding parameters (eqs. 9.5 and 9.6), are in fact equation lines which connect primary, secondary and tertiary amides together; within each class of amides the alkyl chain shows an electron withdrawing effect, opposite to the order observed in solution. However because this effect is small and therefore, for example, the difference between (values of atomic charge parameters of) acetamides and propionamides is insignificant compared with that between secondary and tertiary amides, we can still find correlations (the only exception from this rule is N-methylformamide).

The Q_H and LUMO energy, calculated without molecular mechanics adjustments, cannot model the H-bond donor ability of amides. The order of the amides from the highest to the lowest Q_H values without MM corrections is:

propionamide > N-methylformamide > N-methylpropionamide > N-methylacetamide > formamide > acetamide.

This order indicates that N-methylated amides have higher Q_H values than do non-substituted ones (except for propionamide). This is opposite the order expected from the known inductive effect of the alkyl groups in the solution. MM corrections in fact give the substituted amides a lower Q_H value than have the non-substituted ones, except for N-methylformamide which was an outlier from the equation.

On the other hand, Q_{MN} and E_{HOMO} calculated by this method (without MM corrections) can be used to predict $\Sigma\beta^H_2$ values reasonably well:

$$\Sigma\beta^H_2 = 0.327 - 0.895Q_{MN} \qquad (9.8)$$

$n = 8 \quad r = 0.810 \quad s = 0.0340 \quad F = 11.5 \quad p = 0.015$

Again, N-methylformamide has been excluded from the equation. E_{HOMO} could not improve the regression (because of the good correlation between Q_{MN} and E_{HOMO} for these compounds) but omitting acetanilide and N-methyl formamide, it gives an excellent correlation with $\Sigma\beta^H_2$:

$$\Sigma\beta^H_2 = 2.44 + 0.167E_{HOMO} \qquad (9.9)$$

$n = 7 \quad r = 0.922 \quad s = 0.0241 \quad F = 28.2 \quad p = 0.003$

Clearly, Q_{MN} and E_{HOMO} calculated without MM corrections have better correlations with $\Sigma\beta^H_2$ than those calculated after MM corrections (without MM corrections, propionamides have a lower Q_{MN} value than have acetamides).

Alcohols: The H-bond donor ability of alcohols has been plotted against E_{LUMO} in Figure 9.2. Leaving out those alcohols which have a double bond in the structure, the relationship with E_{LUMO} is:

$$\Sigma\alpha^{H}_{2} = 0.206E_{LUMO} - 0.317 \qquad (9.10)$$

n = 9 r = 0.982 s = 0.0318 F = 188 p = 0.000

Surprisingly, E_{LUMO} has a positive coefficient in this equation. Even after deleting water, which is far from the rest of the alcohols in the plot and has a large influence to the equation, the slope of E_{LUMO} is still positive, which cannot be explained:

$$\Sigma\alpha^{H}_{2} = 0.118E_{LUMO} - 0.029 \qquad (9.11)$$

n = 8 r = 0.792 s = 0.0242 F = 10.1 p = 0.019

Adding the parameter Q_H to the equation did not improve the regression and the t-ratio for it was not significant. Furthermore, there is a poor correlation between $\Sigma\alpha^{H}_{2}$ and Q_H alone:

$$\Sigma\alpha^{H}_{2} = 30.3Q_H - 5.06 \qquad (9.12)$$

n = 8 r = 0.660 s = 0.0242 F = 4.6 p = 0.075

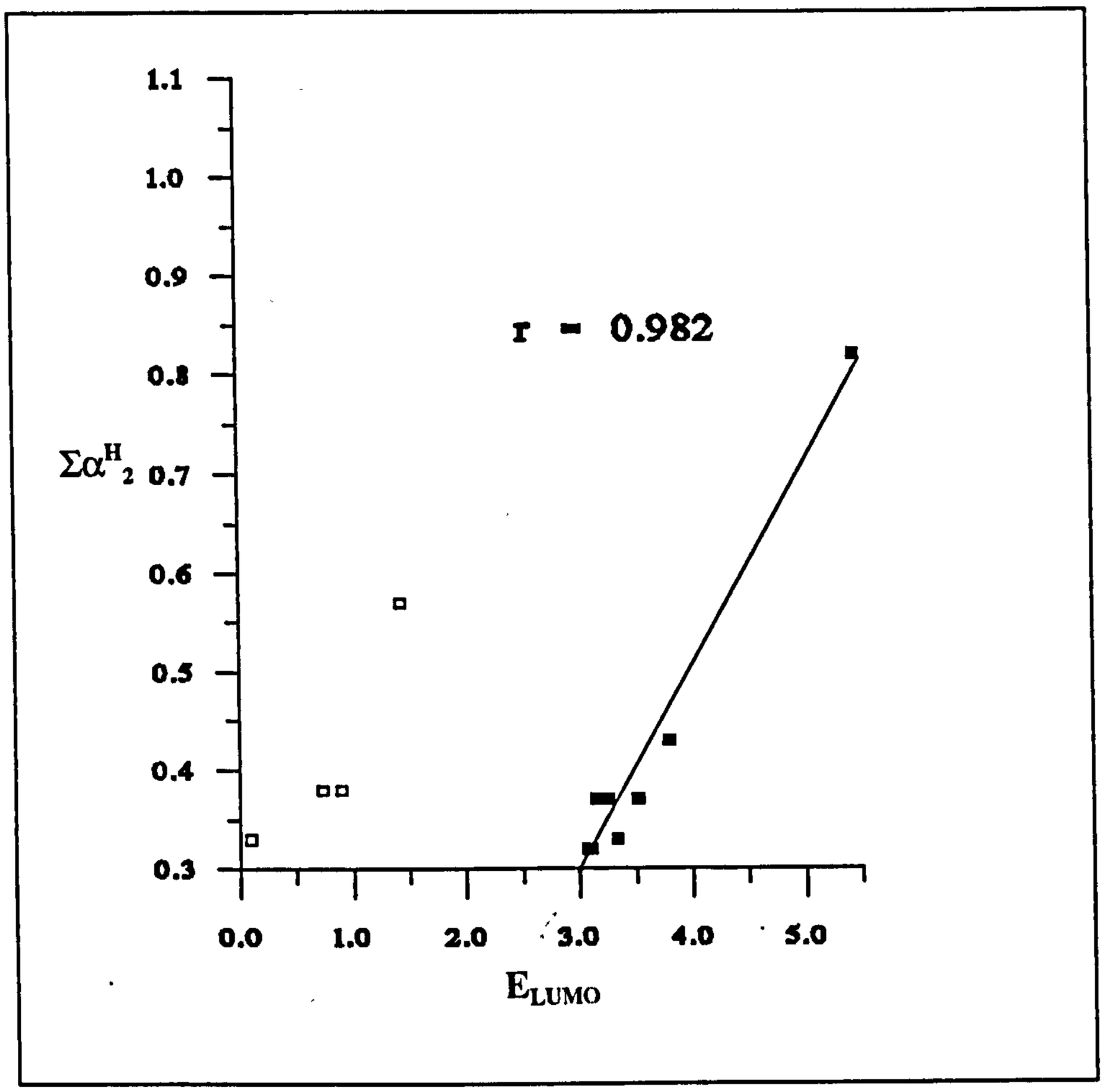


Figure 9.2. The plot of $\Sigma\alpha^H_2$ against E_{LUMO} of alcohols without π-bond in their structures (■); alcohols containing a double bond (□) are outliers.

Figure 9.3 demonstrates the plot of H-bond basicity of these compounds ($\Sigma\beta^H_2$) against E_{HOMO}. Here again structures containing π-bond are outliers, as is clear in Figure 9.3, and leaving them out gives rise to the equation:

$$\Sigma\beta^H_2 = 0.180E_{HOMO} + 2.53 \qquad (9.13)$$

$n = 9 \quad r = 0.896 \quad s = 0.0322 \quad F = 28.3 \quad p = 0.000$

Ethers can also be added to the alcohols:

$$\Sigma\beta^{H}_{2} = 0.147E_{HOMO} + 2.12 \qquad (9.14)$$

n = 15 r = 0.779 s = 0.0600 F = 20.1 p = 0.000

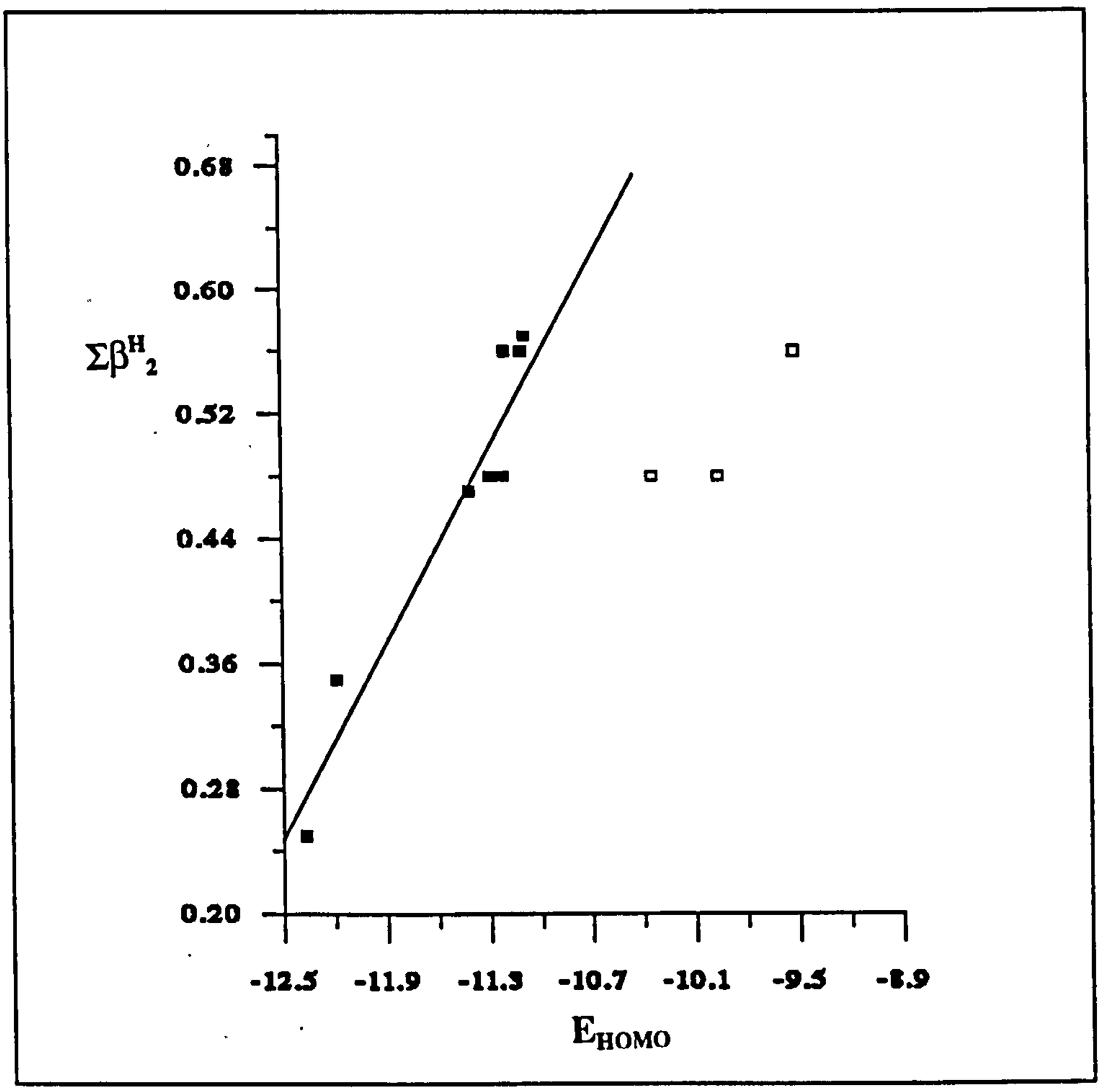


Figure 9.3. The plot of $\Sigma\beta^{H}_{2}$ against E_{HOMO} of alcohols without double bond in the structure (■); alcohols containing π-bonds (□) are outliers.

In the correlation with Q_{MN}, after deleting 2,2,2-trifluoroethanol, H-bond basicity apparently increases with decreasing negative point charge (with a very low r value of 0.377 for the equation). In fact charges calculated by MNDO method for the hydroxylic oxygens of the alcohols do not follow the order expected by the inductive electron donor effect of the alkyl substituents; for example methanol has a lower (more negative) Q_{MN} value than ethanol. The range of variations of the atomic charge on the hydroxyl oxygens of these alcohols, except for 2,2,2-trifluoroethanol, is quite

small (between -0.3293 and -0.3142) and therefore this inconsistency can be assumed to be attributed to errors of the approximation used. In confirmation of this, inclusion of 2,2,2-trifluoroethanol, which is a considerably stronger H-bond acceptor and has much more negative charge on the oxygen than do the other alcohols, changes the sign of the Q_{MN} coefficient. The other (and more important) explanation could be the observations that in the gas-phase an alkyl substituent can also act as an electron-withdrawing group (Brauman and Blair, 1968).

In Figure 9.4 $\Sigma\beta^H_2$ has been plotted against atomic charge on the hydroxyl oxygens of alcohols including 2,2,2-trifluoroethanol.

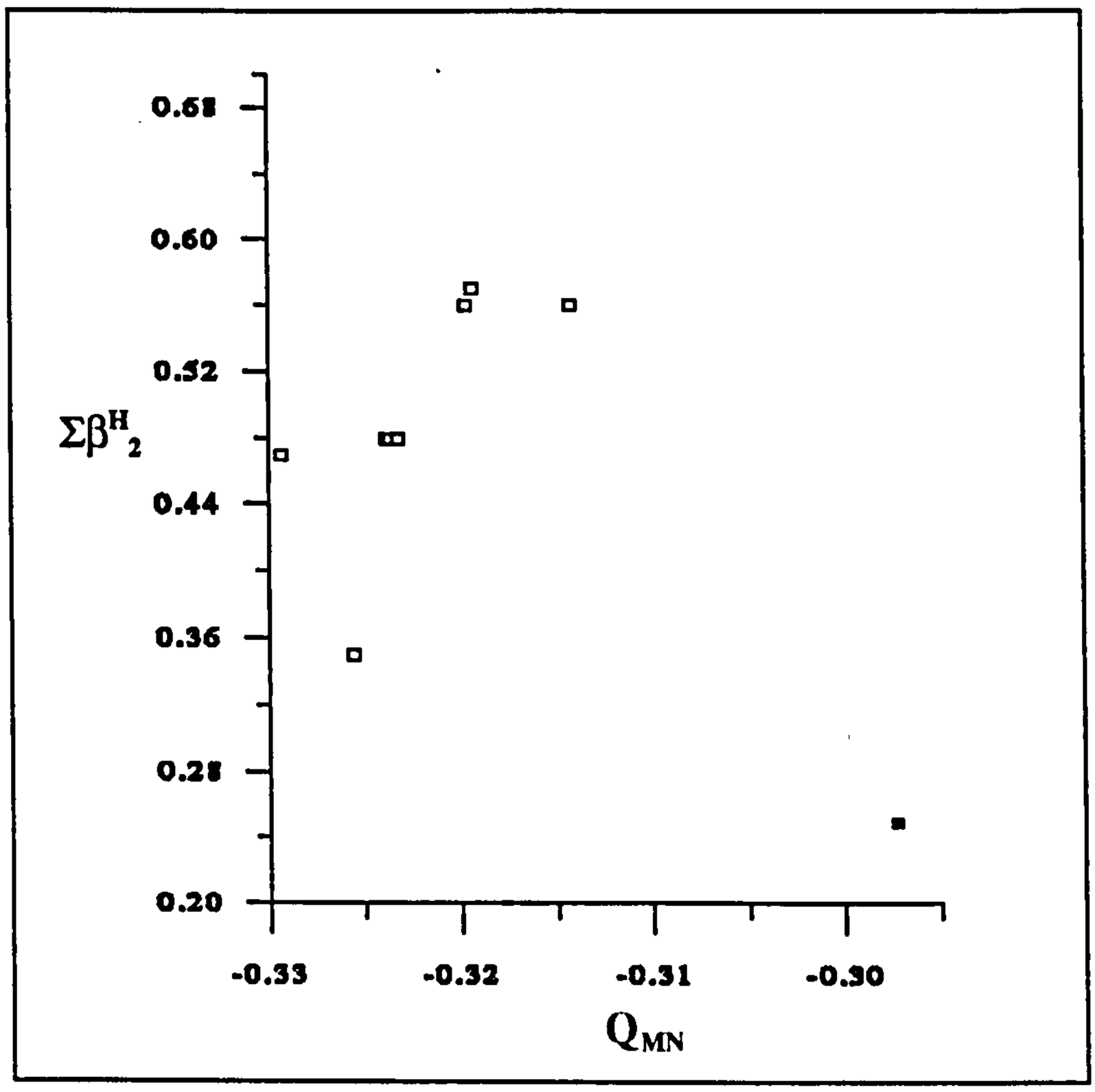


Figure 9.4. The plot of $\Sigma\beta^H_2$ against Q_{MN} of alcohols (□) including 2,2,2-trifluoroethanol (■).

Carboxylic acids: The H-bond donor ability can be expressed by the following relationship:

$$\Sigma\alpha^H_2 = 19.7Q_H - 3.64 \qquad (9.15)$$

n = 12 r = 0.934 s = 0.0482 F = 69 p = 0.000

The correlation with E_{LUMO} is poor (r = 0.510) and the inclusion of this parameter does not improve the correlation. However after deleting aromatic acids (benzoic acids) the following correlation exists:

$$\Sigma\alpha^H_2 = 0.770 - 0.156E_{LUMO} \qquad (9.16)$$

n = 8 r = 0.908 s = 0.0645 F = 28.1 p = 0.002

For H-bond acceptor ability, the esters were included in the regression analyses. At first there was a rather poor correlation with Q_{MN} only (r = 0.665), but as the plot of $\Sigma\beta^H_2$ and E_{HOMO} (Figure 9.5) clearly shows, compounds with conjugated systems (benzoic acids and vinyl acetate) were outliers and omitting them gave rise to the following equations:

$$\Sigma\beta^H_2 = -0.498 - 2.56Q_{MN} \qquad (9.17)$$

n = 11 r = 0.832 s = 0.0403 F = 20.3 p = 0.000

$$\Sigma\beta^H_2 = 0.192E_{HOMO} + 2.64 \qquad (9.18)$$

n = 11 r = 0.902 s = 0.0314 F = 39.4 p = 0.000

$$\Sigma\beta^{H}_{2} = 0.134E_{HOMO} - 1.15Q_{MN} + 1.56 \qquad (9.19)$$

$n = 11$ $r = 0.939$ $s = 0.0266$ $F = 29.6$ $p = 0.000$

t-ratios in the last equation for E_{HOMO} and Q_{MN} are 3.56 ($p = 0.007$) and 2.12 ($p = 0.067$) respectively.

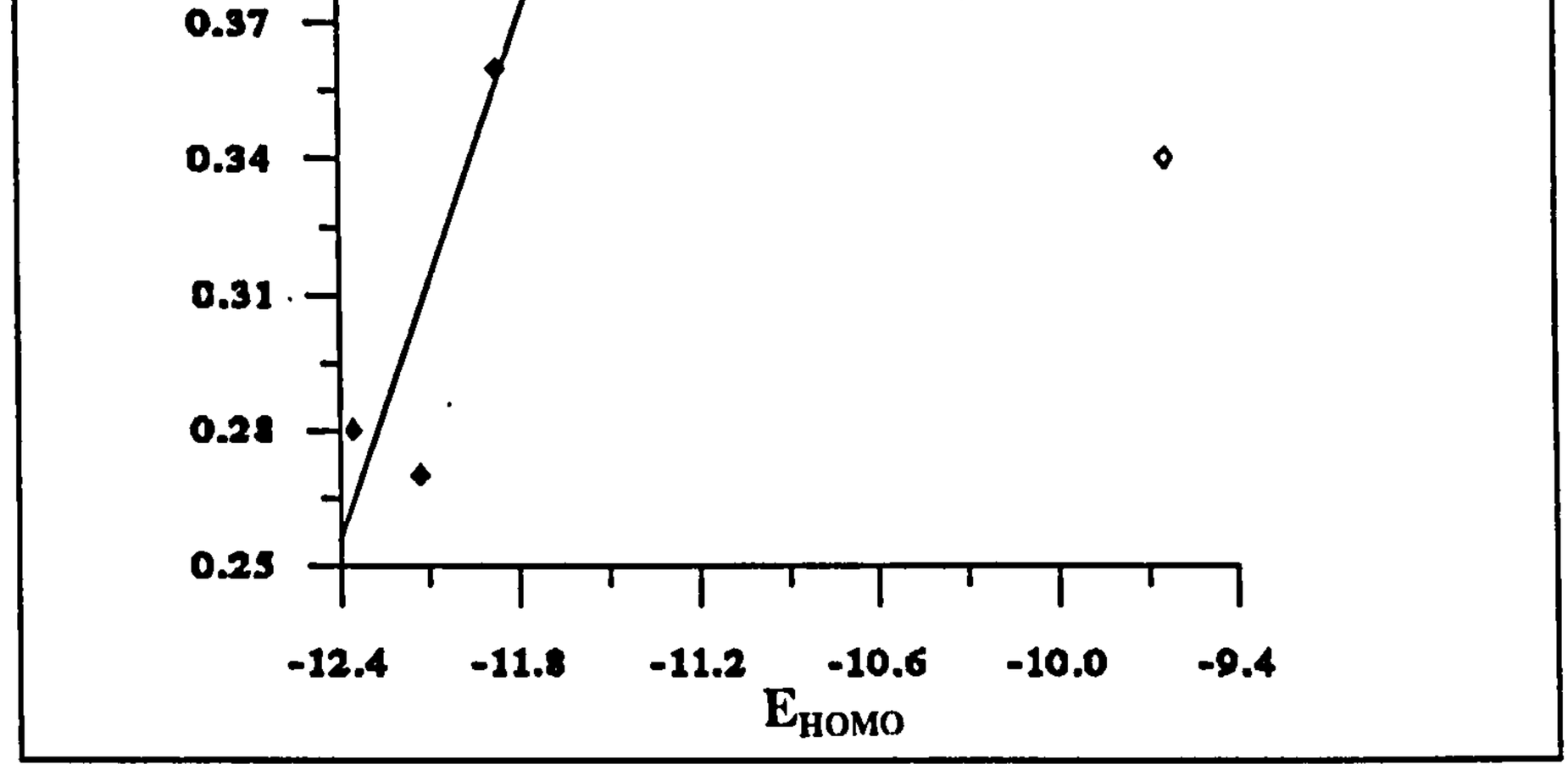


Figure 9.5. The plot of $\Sigma\beta^{H}_{2}$ against E_{HOMO} of non-conjugated carboxylic acids (♦); conjugated acids (◊) are outliers.

Amines: The MNDO method calculated more negative charge on the nitrogen of the tertiary amines than on that of the secondary amines, whilst ammonia had the least negative nitrogen. This is the order expected from the inductive electron donor effect of alkyl groups in solution and it is in agreement with the order of the gas-phase

basicities of amines (determined by ion cyclotron resonance spectroscopy) (Brauman et al, 1971). On the other hand atomic charges on the hydrogens connected to the nitrogen (the most positive hydrogens in the amines) were more positive in the secondary amines than in the primary amines and ammonia had the lowest Q_H value. This is opposite to the order predicted by the known inductive effect of alkyl groups in solution (electron-donating effect). However gas-phase ordering of acidities of amines shows a similar disparity with the order expected from alkyl inductive effect in solution (Brauman & Blair, 1971). There are a number of indications in the literature that an alkyl group can stabilise negative charge as well as positive charge. Schubert et al (1962), in a study of the effect of *p*-alkyl substituents on the energy of electronic transitions of phenol, anisole, aniline and N,N-dimethylaniline, demonstrated that alkyl groups can function either as electron donors or electron acceptors relative to hydrogen, depending on the nature of the electron demand on the alkyl group.

Accordingly, H-bond acceptor ability could be reasonably (but not well) described by the MNDO calculated parameters while H-bond donor ability could not:

$$\Sigma\beta^H_2 = 0.0916E_{HOMO} + 1.59 \qquad (9.20)$$

$n = 14 \quad r = 0.769 \quad s = 0.0371 \quad F = 17.4 \quad p = 0.001$

$$\Sigma\beta^H_2 = 0.425 - 0.737Q_{MN} \qquad (9.21)$$

$n = 15 \quad r = 0.691 \quad s = 0.0570 \quad F = 12 \quad p = 0.004$

Thiazole has been excluded from equation 9.20 and pyrrole from equations 9.20 and 9.21. Thiazole and pyrrole are both aromatic structures therefore it is not surprising for them to be outliers from equation 9.20; deletion of the rest of the aromatic

structures does not change this equation significantly. The atomic charge on the nitrogen of pyrrole is -0.2245. This amount of atomic charge is higher than expected, because of the possibility of the conjugation in the molecule which can lead to a positive charge on the nitrogen atom, as is shown in Figure 9.6.

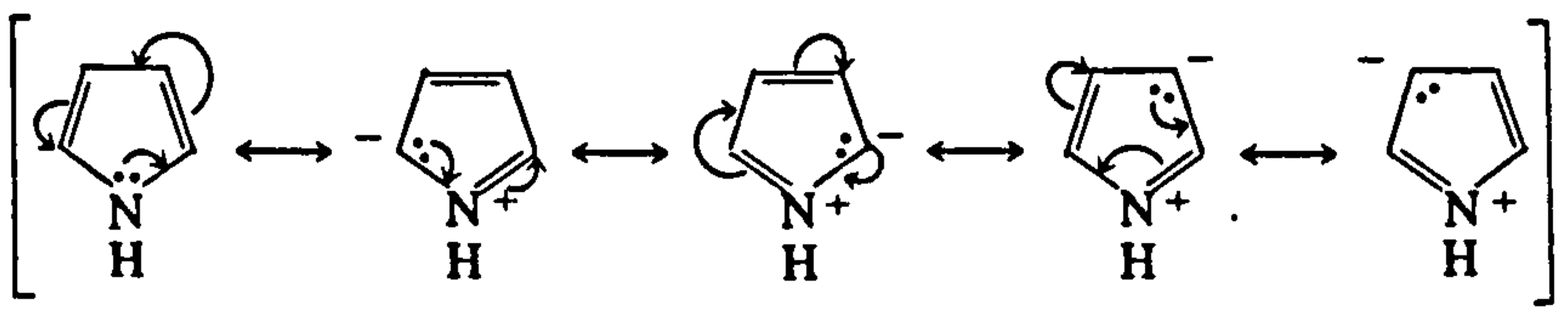


Figure 9.6. Resonance structures of pyrrole

b. Correlations across the different classes of compounds:

For 111 different structures, the following is the relationship between $\Sigma\alpha^H_2$ and Q_H:

$$\Sigma\alpha^H_2 = 3.06Q_H - 0.0925 \qquad (9.22)$$

$n = 111 \quad r = 0.865 \quad s = 0.1434 \quad F = 326 \quad p = 0.000$

E_{LUMO} did not improve the equation and there was no correlation between $\Sigma\alpha^H_2$ and E_{LUMO}.

Figure 9.7 is the plot of $\Sigma\alpha^H_2$ against Q_H. In the plot, it is clear that there are two groups of compounds. For the first group which are quite weak H-bond acids (or are not H-bond donors at all) the slope is much lower compared to the second group which are stronger H-bond acids. Therefore, the correlation cannot be very realistic.

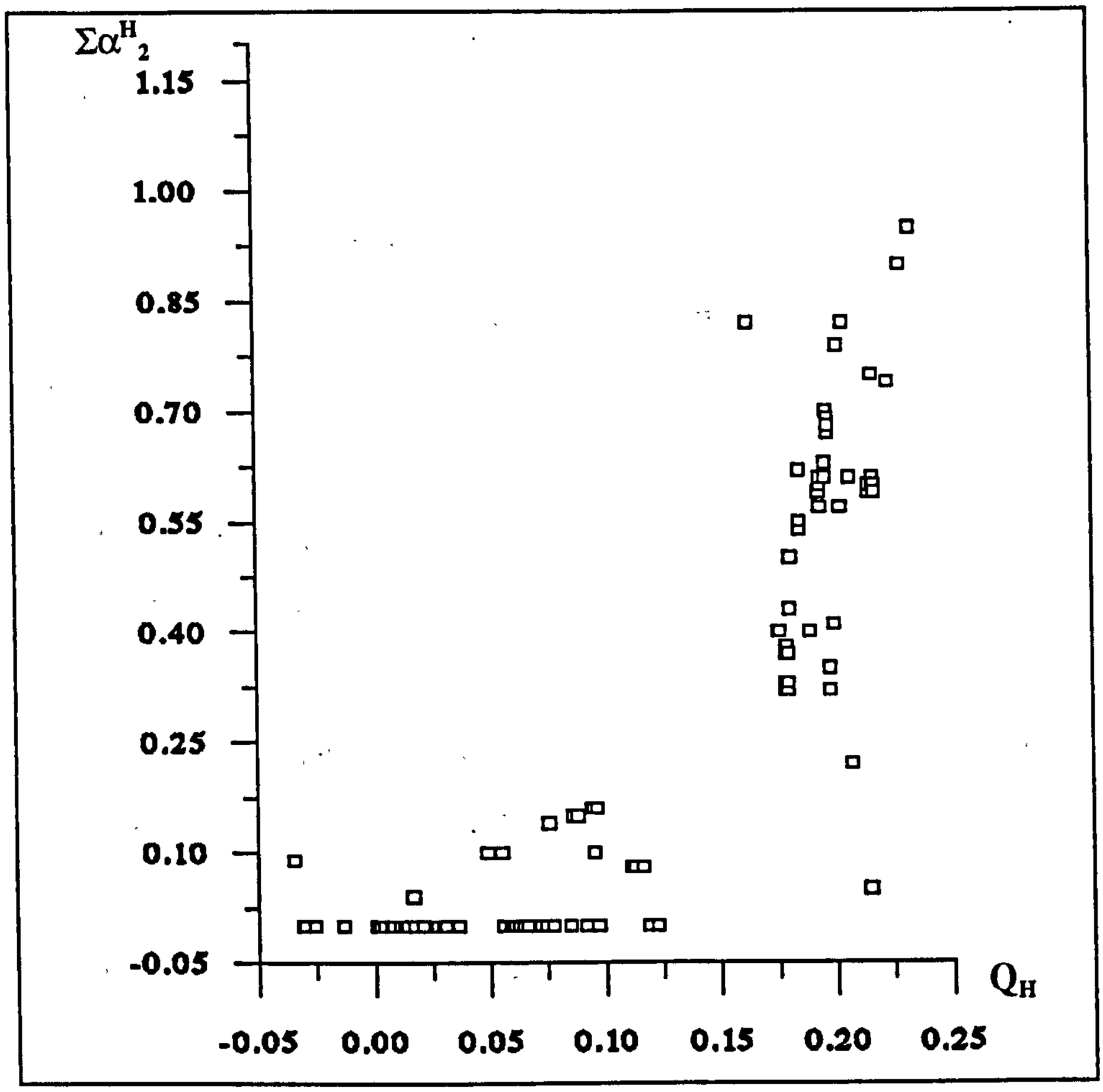


Figure 9.7. The plot of $\Sigma\alpha^H_2$ against Q_H for all the 111 compounds.

Stronger H-bond donors (the compounds which theoretically were capable of hydrogen donation, including alcohols, amides and amines with free hydrogen connected to the nitrogen, carboxylic acids, phenols) were selected to be studied separately. Regression analyses showed that 2-substituted phenols and water were outliers. 2-Substituted phenols were also outliers from the equation for phenols alone. In these phenols the short distance between the phenolic hydroxyl group and some substituents in ortho position can allow intramolecular H-bonding. Furthermore, the substituents in such a position can account for steric shielding of the hydroxyl group which limits the access of the H-bond pair. These two properties do not affect the corresponding Q_H or E_{LUMO} values meaning that the Q_H and E_{LUMO} of 2-substituted

phenols are similar to those of the 3- and 4-substituted phenols (Table 9.1) but their H-bond donor abilities are much less.

Water is an unusual molecule and its high polarity (μ = 1.84D) might be the reason for its exceptionally higher H-bond donor ability than its Q_H and E_{LUMO} values would predict. After omitting the above outliers the following equation resulted:

$$\Sigma\alpha^H_2 = 3.51Q_H - 0.0509E_{LUMO} - 0.086 \qquad (9.23)$$

n = 55 r = 0.918 s = 0.0905 F = 138.6 p = 0.000

It is also possible to include thiols and thiophenol as well as the halogenated hydrocarbons some of which have $\Sigma\alpha^H_2$ more than 0.10, in the regressions:

$$\Sigma\alpha^H_2 = 2.88Q_H - 0.0529E_{LUMO} + 0.0274 \qquad (9.24)$$

n = 65 r = 0.920 s = 0.1011 F = 176 p = 0.000

The difference between this equation and the former one is not statistically significant (t=0.176). Figure 9.8 shows the plot between $\Sigma\alpha^H_2$ and predicted $\Sigma\alpha^H_2$ values by equation 9.23.

The H-bond acceptor ability of the compounds listed in Table 9.1 can be expressed by:

$$\Sigma\beta^H_2 = -1.18Q_{MN} + 0.106 \qquad (9.25)$$

n = 111 r = 0.771 s = 0.1434 F = 160 p = 0.000

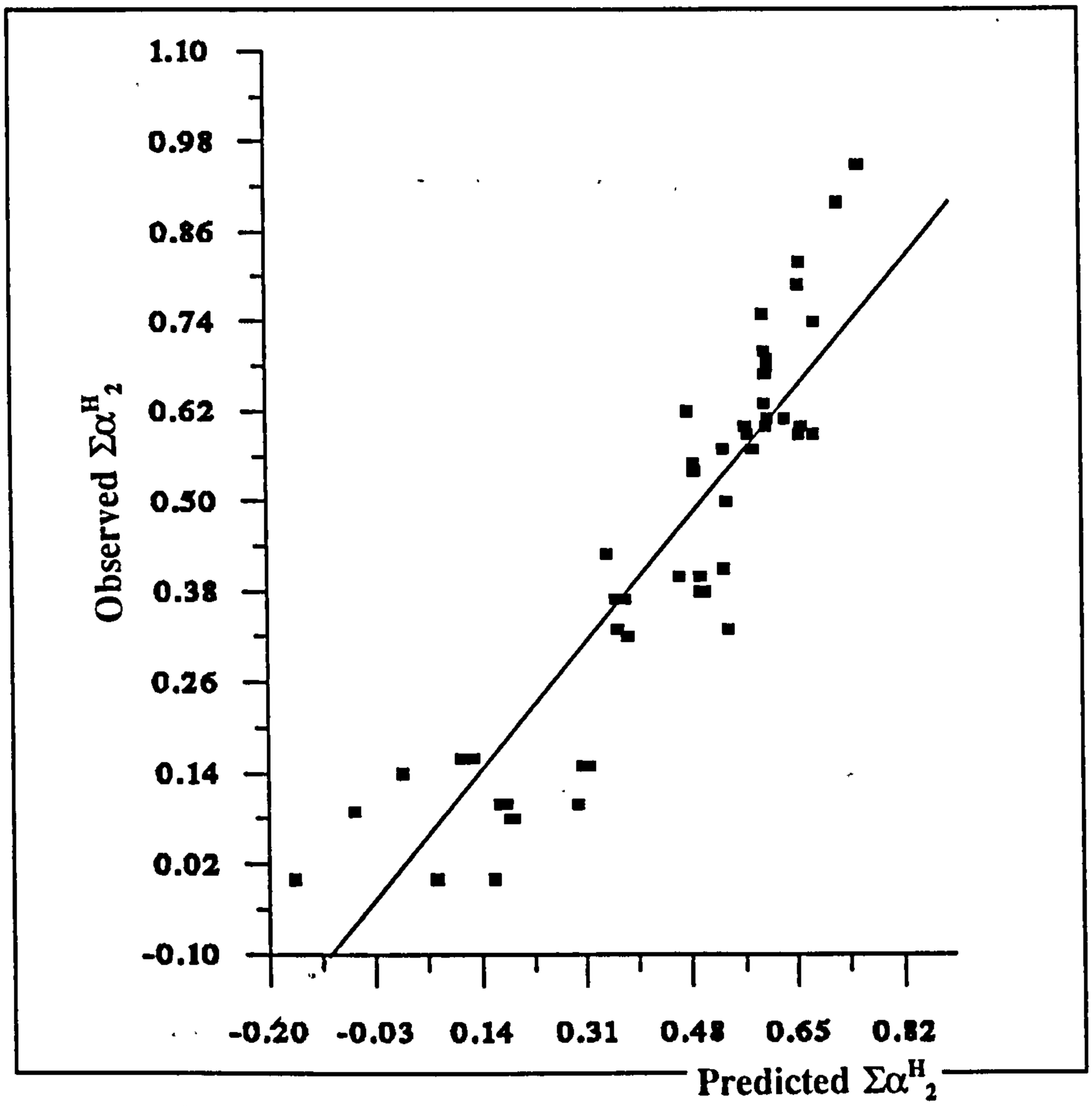


Figure 9.8. The graph between $\Sigma\alpha^H_2$ and predicted $\Sigma\alpha^H_2$ by equation 9.23.

There was no correlation with E_{HOMO}. The plot between $\Sigma\beta^H_2$ and E_{HOMO} (Figure 9.9) showed that there seemed to be two groups of compounds. For the first group, which contained compounds with E_{HOMO} values lower than about -10.5, $\Sigma\beta^H_2$ increased with increasing E_{HOMO}. The second group comprised the structures for which E_{HOMO} values were higher than about -10.5 and their graph of $\Sigma\beta^H_2$ against E_{HOMO} was rather scattered. Looking at Table 9.1 it was obvious that compounds which possessed resonance systems were in the second group but simple non-resonance structures were in the first group. Dividing up these compounds, the following equation was obtained for the first group:

$$\Sigma\beta^{H}_{2} = 0.135E_{HOMO} - 1.23Q_{MN} + 1.60 \qquad (9.26)$$

$n = 65 \quad r = 0.919 \quad s = 0.0914 \quad F = 170 \quad p = 0.000$

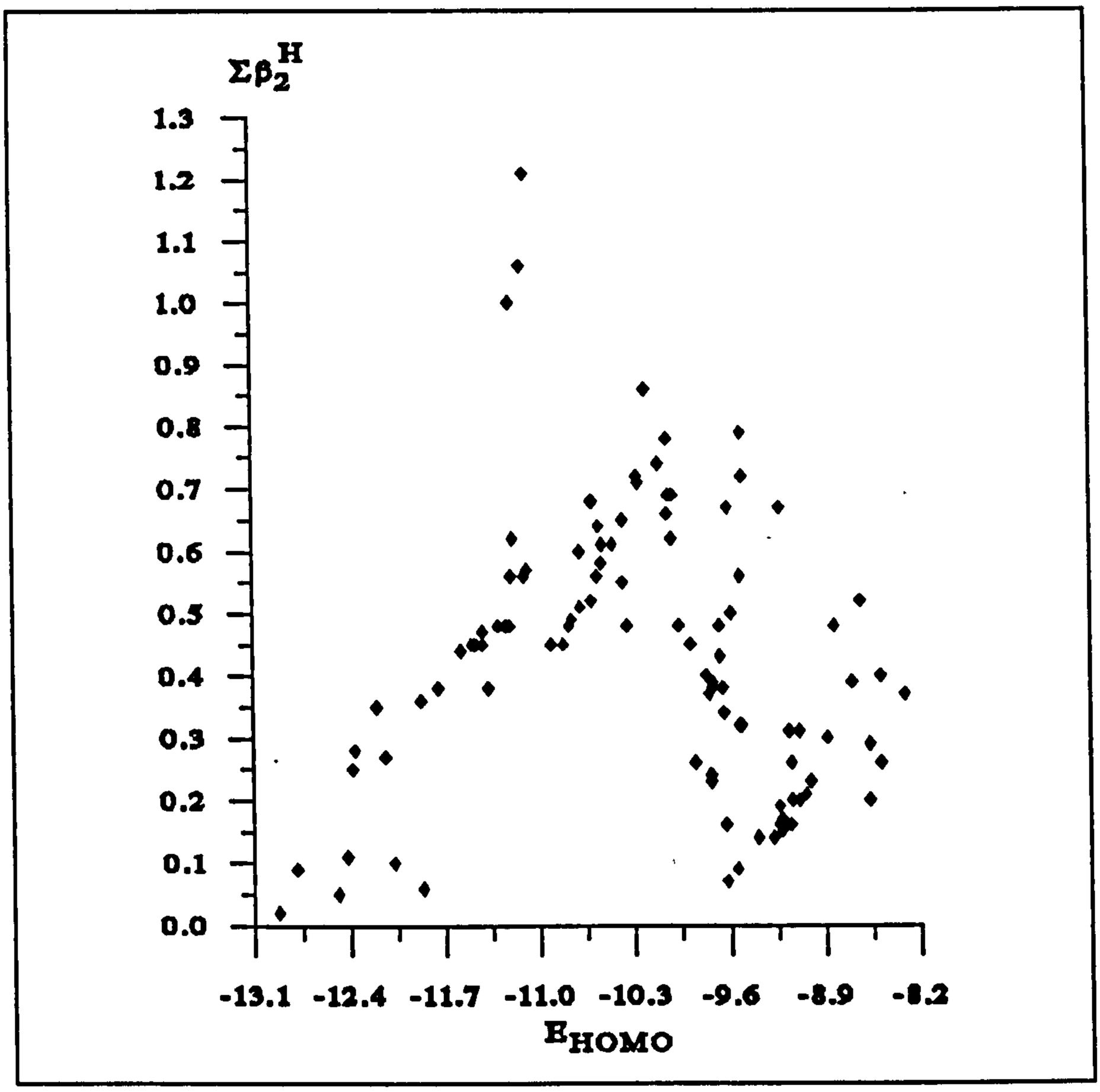


Figure 9.9. The plot between $\Sigma\beta^{H}_{2}$ and E_{HOMO} calculated by the MNDO method.

The 65 compounds in this equation consist of ethers, ketones, amines, amides, carboxylic acids, alcohols, phosphates, halogenated hydrocarbons, thiols and sulphides, provided that they do not have resonance structures. Phosphates are a little out of the line of the equation (Figure 9.10), which may be because of the poor parametrisation of the program for these compounds which also have slightly delocalised structures. Ammonia is also an outlier from this equation. If we delete the three phosphates and ammonia from equation 9.26, the correlation improves to:

$$\Sigma\beta^H_2 = 0.139E_{HOMO} - 1.03Q_{MN} + 1.69 \qquad (9.27)$$

$n = 61 \quad r = 0.945 \quad s = 0.0635 \quad F = 243 \quad p = 0.000$

Both Q_{MN} and E_{HOMO} are significant in the equations 9.26 and 9.27 ($p = 0.000$). Figure 9.11 is the plot of $\Sigma\beta^H_2$ versus the corresponding predicted values by equation 9.27.

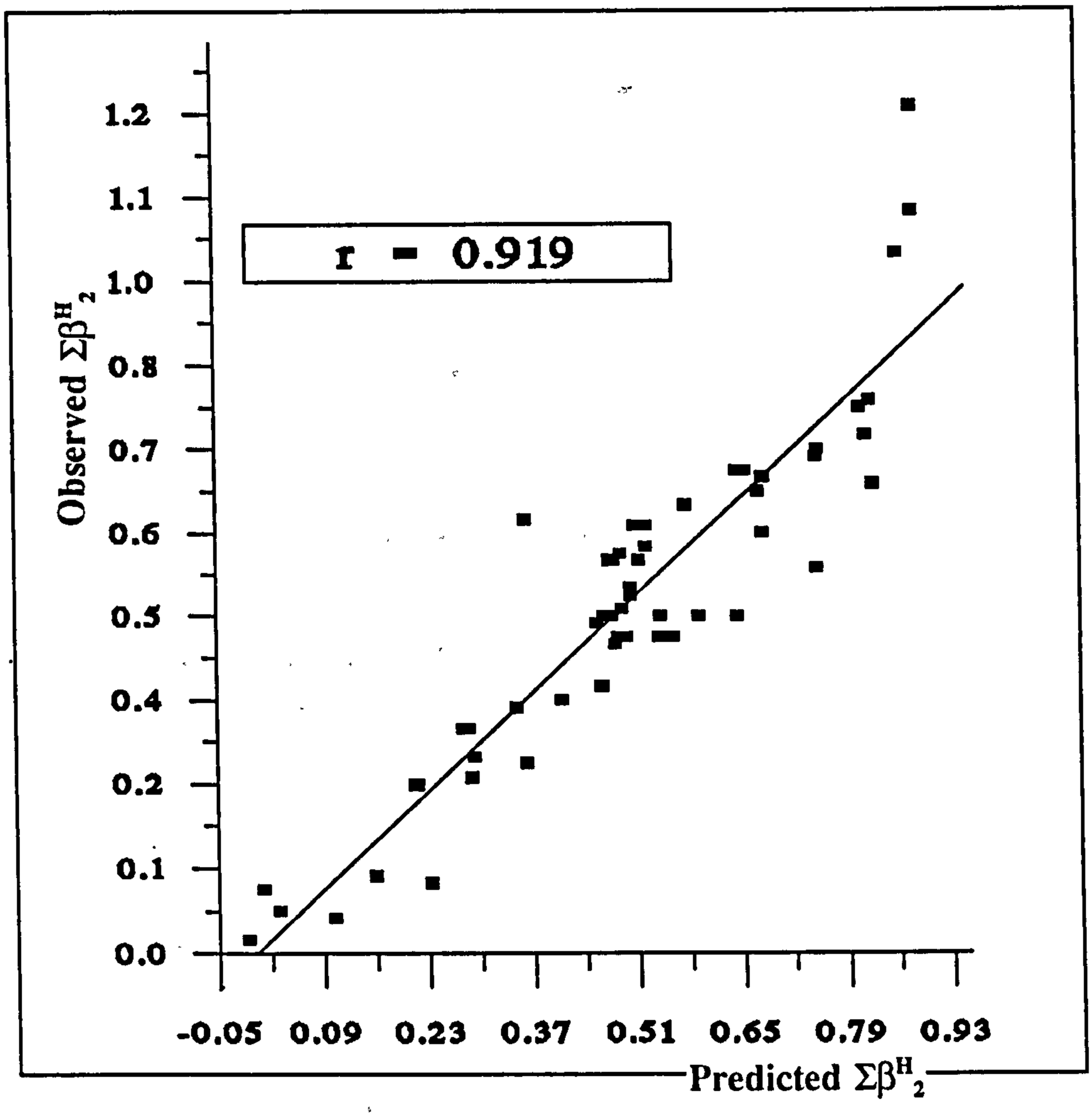


Figure 9.10. The plot between $\Sigma\beta^H_2$ and predicted $\Sigma\beta^H_2$ by equation 9.26.

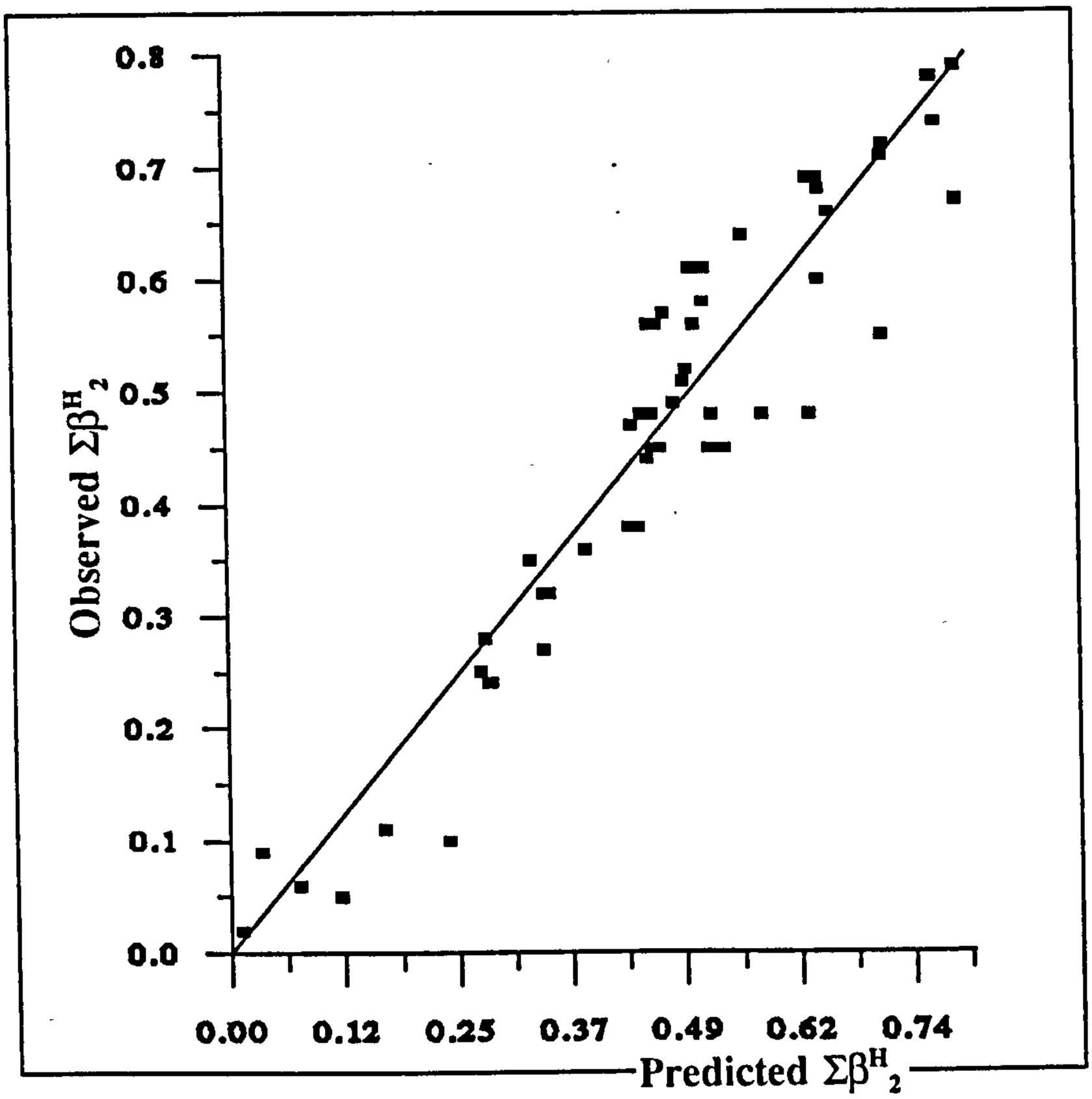


Figure 9.11. The plot between $\Sigma\beta^H_2$ and predicted $\Sigma\beta^H_2$ by equation 9.27.

9.2.2. *AM1 method*

a. Correlation within families

Phenols: Correlations of AM1 calculated parameters with H-bond experimental parameters were very similar to those using MNDO; nitrophenols were out of line in the correlation for H-bond acceptor ability and 2-substituted phenols were outliers from both H-bonding donor and acceptor equations. The equations are given below:

$$\Sigma\alpha^H_2 = 21.4Q_H - 4.05 \qquad (9.28)$$

n = 13 r = 0.959 s = 0.0223 F = 126.2 p = 0.000

$$\Sigma\alpha^H_2 = 0.648 - 0.127E_{LUMO} \quad (9.29)$$

$n = 13 \quad r = 0.819 \quad s = 0.0451 \quad F = 22.4 \quad p = 0.000$

When both parameters were used in a single equation to describe $\Sigma\alpha^H_2$, the t-ratio for E_{LUMO} was insignificant, with the correlation coefficient of the resulting equation being the same as the r of the correlation with only the Q_H parameter. This is again because of the high correlation between these two parameters.

The H-bond acceptor ability of phenols can be described by the following equation if the two nitrophenols are excluded from the equation. The reason for excluding these phenols has been given in the MNDO section.

$$\Sigma\beta^H_2 = 0.177E_{HOMO} - 21.7Q_{MN} - 3.53 \quad (9.30)$$

$n = 11 \quad r = 0.885 \quad s = 0.0601 \quad F = 14.5 \quad p = 0.002$

<u>Amides:</u> The only valid equation here was:

$$\Sigma\beta^H_2 = 0.110E_{HOMO} + 1.81 \quad (9.31)$$

$n = 7 \quad r = 0.902 \quad s = 0.0269 \quad F = 21.7 \quad p = 0.006$

In this equation E_{HOMO} values used were calculated after MM correction to the rotation barrier. However, a similar equation resulted when the corrections were not applied:

$$\Sigma\beta^H_2 = 0.111E_{HOMO} + 1.82 \quad (9.32)$$

$n = 7 \quad r = 0.903 \quad s = 0.0267 \quad F = 22.1 \quad p = 0.005$

The Q_{MN} values calculated by the AM1 method show reverse the order expected according to the inductive effect of alkyl groups in solutions. Therefore it is not

surprising that with increasing H-bond acceptor ability ($\Sigma\beta^H_2$), Q_{MN} value increases (the negative charge decreases) and that this parameter could not improve the correlations with E_{HOMO} parameter.

Q_H values of primary and secondary amides calculated by AM1 method show a very narrow range of variations (between 0.2213-0.2302 after MM corrections, and between 0.2181-0.2295 without MM corrections), with an irregular ordering. Therefore the H-bond donor ability of amides could not be expressed by this parameter. E_{LUMO} has a positive coefficient in correlation with $\Sigma\alpha^H_2$ (acetanilide is an outlier):

After MM corrections: $\Sigma\alpha^H_2 = 2.38E_{LUMO} - 3.16$ (9.33)

n = 6 r = 0.631 s = 0.0843 F = 2.6 p = 0.179

Without MM corrections: $\Sigma\alpha^H_2 = 2.91E_{LUMO} - 3.95$ (9.34)

n = 6 r = 0.895 s = 0.0484 F = 16.1 p = 0.016

In equations 9.33 & 9.34 E_{LUMO} has an unexpected positive slope, for which there is no explanation.

Alcohols: Q_H values calculated in AM1 method for different alcohols are not in the expected order; for example methanol has less charge on the hydroxylic hydrogen than have propan-2-ol, cyclopentanol, cyclohexanol and ethanol. The order of the Q_H values for the alcohols by AM1 method (Table 9.2) is as follows:

2,2,2-Trifluoroethanol > Benzyl alcohol > Prop-2-en-1-ol > Cyclohexanol > *trans*-But-2-en-1-ol > Ethanol > Cyclopentanol > Propan-2-ol > Methanol > Butan-1-ol > Propan-1-ol > Hexan-1-ol

Hence, these charges obviously could not predict H-bond donor ability. The same is applicable for H-bond acceptor ability which cannot be predicted by Q_{MN} parameter. The amount of negative charge on the oxygen of these alcohols are in the following order:

Ethanol > Propan-2-ol > Cyclohexanol > Methanol > Propan-1-ol > Hexan-1-ol > Butan-1-ol > Cyclopentanol > *trans*-But-2-en-1-ol > Prop-2-en-1-ol > Benzyl alcohol > 2,2,2-Trifluoroethanol

The oxygen atoms of propan-2-ol, cyclopentanol, and cyclohexanol, which are connected to a carbon with two inductive electron donor alkyl groups, should have more point charge than the oxygen of ethanol which is connected to a carbon atom with only one such group, but the charges calculated by AM1 show the reverse order. 2,2,2-Trifluoroethanol is the weakest H-bond acceptor among these alcohols, with a $\Sigma\beta^H_2$ value far less than the other alcohols. Its Q_{MN} value is the highest with a large distance from that of the other alcohols, giving it a large influence on the regression between $\Sigma\beta^H_2$ and Q_{MN}. This could explain the observation that $\Sigma\beta^H_2$ has the following relationship with Q_{MN} values, but deleting 2,2,2-trifluoroethanol, which is well removed from the other alcohols in the plot, destroys the correlation:

$$\Sigma\beta^H_2 = -7.06Q_{MN} - 1.79 \qquad (9.35)$$

$n = 12 \quad r = 0.836 \quad s = 0.0490 \quad F = 23.2 \quad p = 0.000$

Figure 9.12 is the plot of $\Sigma\beta^H_2$ against Q_{MN} of alcohols. As it is clear in the plot, the $\Sigma\beta^H_2$ values for alcohols, except for 2,2,2-trifluoroethanol, are quite close to each other and so are the Q_{MN} values; therefore the differences between Q_{MN} values of the alcohols could be related to the errors of the AM1 approximation causing the

observed order of the Q_{MN} values.

After deleting alcohols and ethers with π-bonds, the following is the equation between $\Sigma\beta^H_2$ and E_{HOMO} for non-resonance ethers and alcohols:

$$\Sigma\beta^H_2 = 0.099E_{HOMO} + 1.55 \qquad (9.36)$$

n = 15 r = 0.673 s = 0.0707 F = 10.8 p = 0.006

There is no correlation between $\Sigma\alpha^H_2$ and E_{LUMO}.

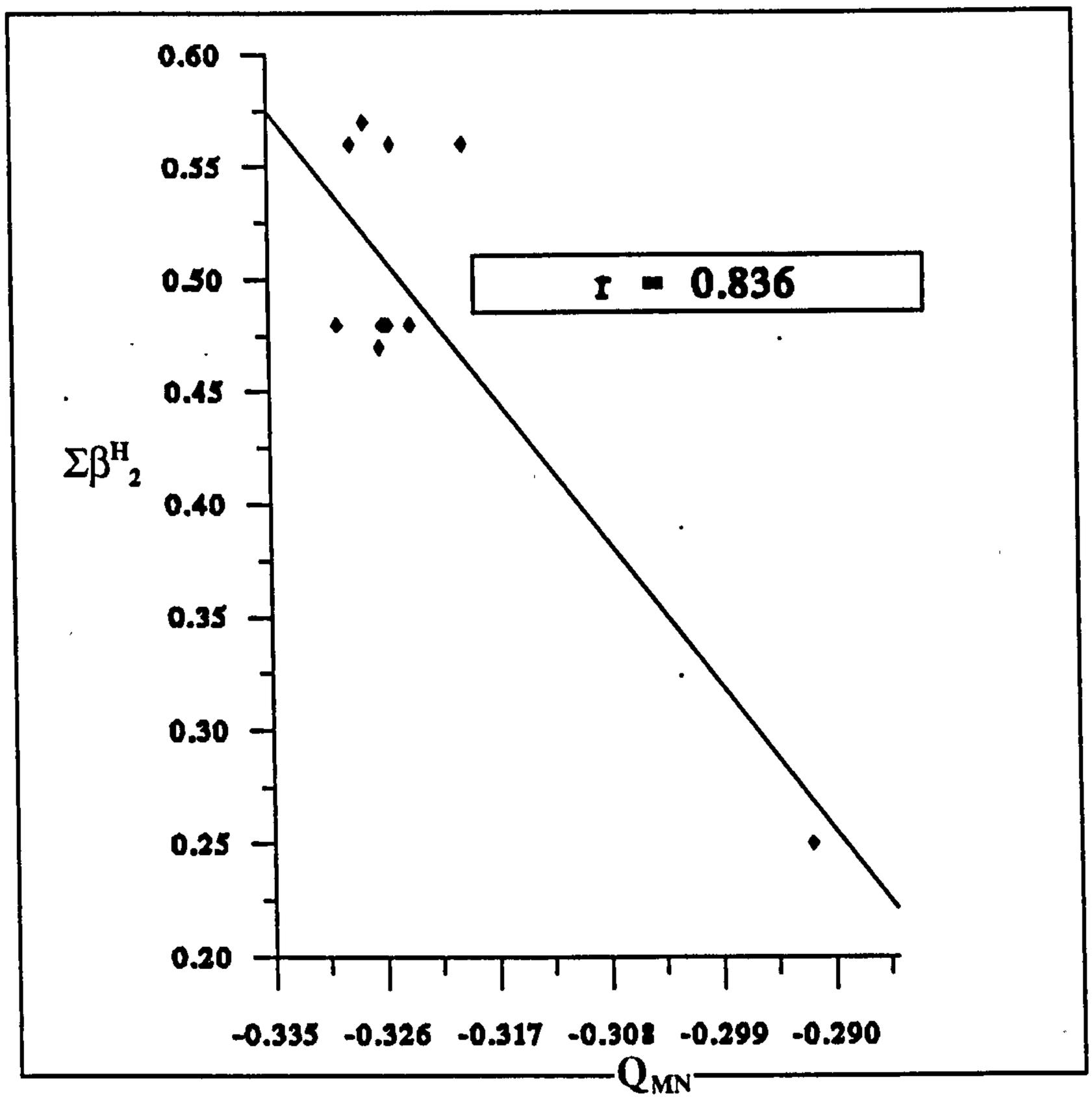


Figure 9.12. The plot between $\Sigma\beta^H_2$ and Q_{MN} of alcohols.

Carboxylic Acids: The H-bond ability of carboxylic acids can be reasonably correlated by AM1 molecular orbital parameters:

$$\Sigma\alpha^H_2 = 20.8Q_H - 4.43 \quad (9.37)$$

n = 12 r = 0.858 s = 0.0694 F = 28 p = 0.000

$$\Sigma\beta^H_2 = -2.16Q_{MN} - 0.382 \quad (9.38)$$

n = 12 r = 0.711 s = 0.0460 F = 10.2 p = 0.009

Deletion of the resonance structures (four benzoic acid derivatives) gave rise to better correlations with the atomic charges. There are also good relationships with HOMO and LUMO energies for the non-aromatic structures. The relationships are:

$$\Sigma\alpha^H_2 = 19.7Q_H - 4.15 \quad (9.39)$$

n = 8 r = 0.919 s = 0.0607 F = 32.6 p = 0.001

$$\Sigma\alpha^H_2 = 0.815 - 0.185E_{LUMO} \quad (9.40)$$

n = 8 r = 0.922 s = 0.0595 F = 34.1 p = 0.001

Because E_{LUMO} is highly correlated with Q_H (r = 0.992), it cannot improve the one parameter equation.

$$\Sigma\beta^H_2 = -3.07Q_{MN} - 0.677 \quad (9.41)$$

n = 8 r = 0.897 s = 0.0364 F = 24.7 p = 0.003

$$\Sigma\beta^H_2 = 0.310E_{HOMO} + 3.99 \quad (9.42)$$

n = 8 r = 0.797 s = 0.0497 F = 10.4 p = 0.018

Non-resonance esters can also be analysed together with carboxylic acids only when the two parameters are employed:

$$\Sigma\beta^{H}_{2} = 0.176E_{HOMO} - 1.63Q_{MN} + 1.87 \qquad (9.43)$$

$n = 11 \quad r = 0.909 \quad s = 0.0322 \quad F = 18.9 \quad p = 0.001$

Amines: The negative atomic charges on the nitrogens of the amines show the order: primary amines > secondary amines > tertiary amines. For example the magnitudes of negative charge are in the following order for the methylamines: ammonia> methylamine > dimethylamine > trimethylamine. This is opposite to the order of intrinsic proton affinities (from the gas-phase proton-transfer reaction in the mass spectrometer (Munson, 1965), and from *ab initio* (Hehre & Pople, 1970) and CNDO (Graffeuil et al, 1974) molecular orbital computations), but aggrees with the order of Li^{+} affinities of methylamines (and ammonia) (Regis & Corset, 1973). Pullman and Brochen (1975), using *ab initio* calculations, showed that the pure electrostatic attraction for the proton decreased but both pure polarisation and pure charge transfer effects increased upon successive methylation; the continuous increase in the total binding energy upon progressive methylation was brought about by the increase in the two latter components of the energy. AM1 calculated Q_{MN} values for amines, in fact, correctly follow the order of the electrostatic attraction.

In comparison of Q_H values of the amines, primary amines have less point charge on the hydrogen connected to the nitrogen than do the secondary amines. Furthermore, methylamines are less charged on the hydrogen than are the ethyl amines etc. Considering these facts, it is not surprising that the AM1 method cannot model the H-bonding ability of the amines. Neither E_{LUMO} and E_{HOMO} calculated by the AM1 method has any correlation with H-bonding abilities of amines..

b. Correlations across different classes of the compounds

For the whole 111 compounds of Table 9.2 the following equation shows the relationship between H-bond donor ability and AM1 calculated atomic charge:

$$\Sigma\alpha^H_2 = 4.49Q_H - 0.502 \qquad (9.44)$$

n = 111 r = 0.853 s = 0.1497 F = 290 p = 0.000

E_{LUMO} was not significant. The plot between $\Sigma\alpha^H_2$ and Q_H calculated by AM1 (Figure 9.13) shows that, like Q_H calculated by the MNDO method, there are two groups of compounds.

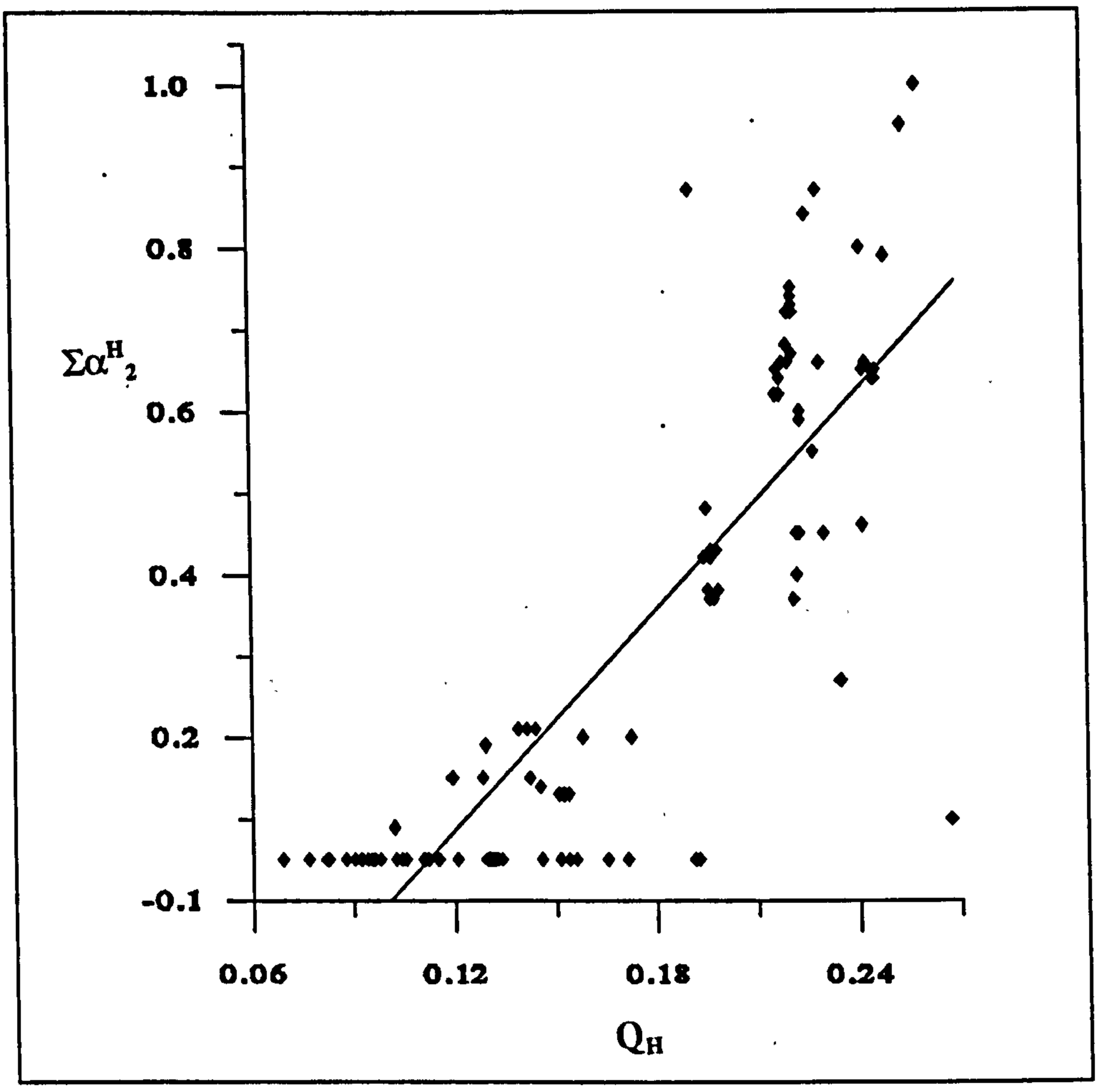


Figure 9.13. The plot between $\Sigma\alpha^H_2$ and Q_H calculated by the AM1 method.

As was done for MNDO calculated parameters, the group of compounds which are theoretically capable of H-bonding (phenols, alcohols, amines and amides which have free hydrogen connected to the nitrogen, carboxylic acids, thiophenol, thiols), were selected and analysed separately. When the outliers were omitted from regression analysis (i.e. 2-substituted phenols, water, pyrrole), the following equation was obtained:

$$\Sigma\alpha^H_2 = 4.42Q_H - 0.0403E_{LUMO} - 0.384 \qquad (9.45)$$

n = 54 r = 0.917 s = 0.0917 F = 134.7 p = 0.000

In this equation t-ratios are significant for both Q_H and E_{LUMO} (p = 0.000).

Thiophenol and thiols can also be included in the correlation:

$$\Sigma\alpha^H_2 = 4.96Q_H - 0.0304E_{LUMO} - 0.511 \qquad (9.46)$$

n = 58 r = 0.933 s = 0.0907 F = 183.6 p = 0.000

In this equation t-ratios are 13.95 (p=0.000) and 3.33 (p=0.002) for Q_H and E_{LUMO} respectively. The plot of this equation is shown in Figure 9.14.

H-bond acceptor ability of all the 111 compounds had the following relationship with Q_{MN} values, and E_{HOMO} was not significant.

$$\Sigma\beta^H_2 = 0.167 - 0.925Q_{MN} \qquad (9.47)$$

n = 111 r = 0.765 s = 0.1452 F = 153.5 p = 0.000

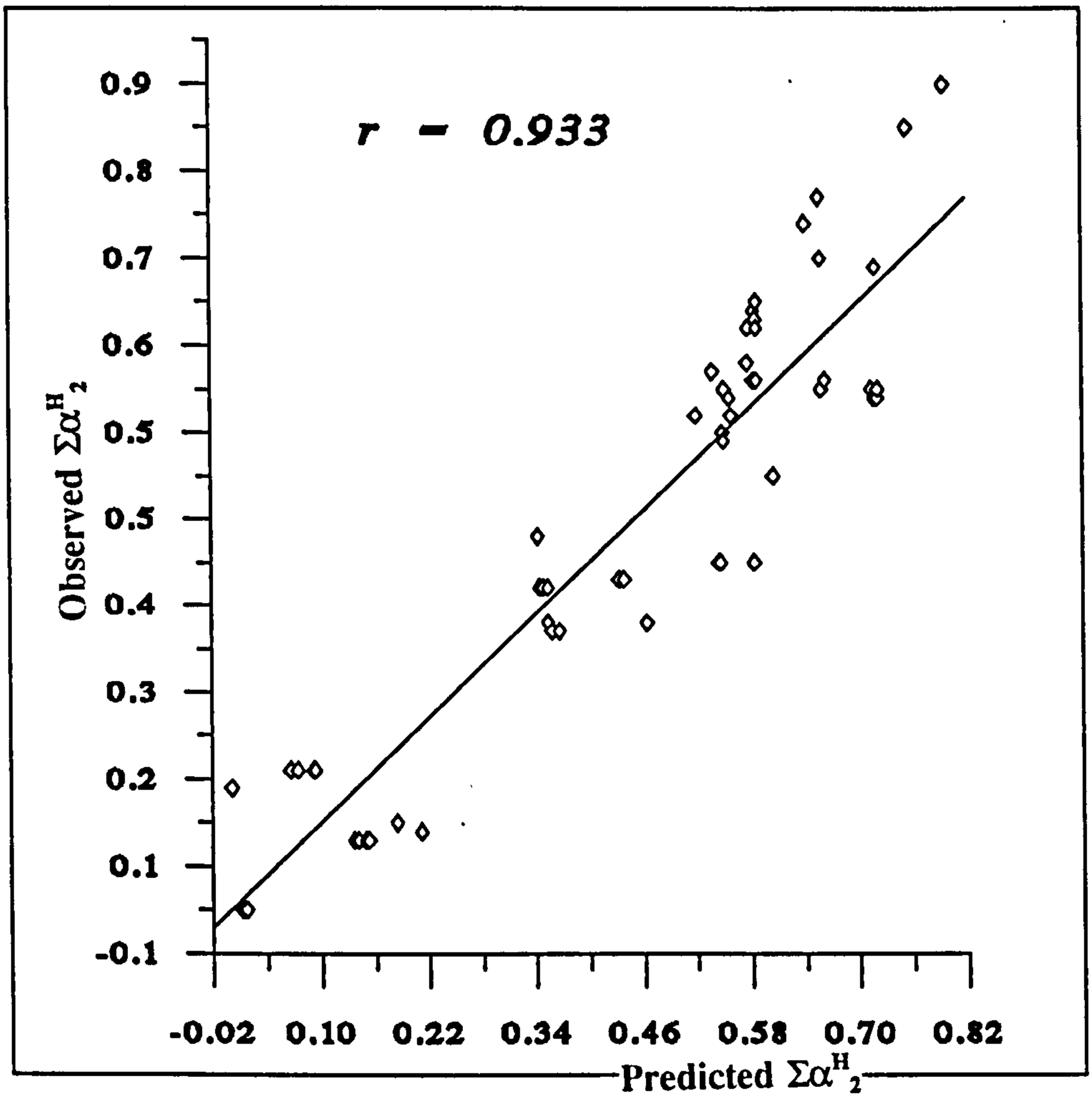


Figure 9.14. The plot of observed $\Sigma\alpha^H_2$ against predicted $\Sigma\alpha^H_2$ (by equation 9.46).

The plot between $\Sigma\beta^H_2$ and AM1 calculated E_{HOMO} values was very similar to that between $\Sigma\beta^H_2$ and MNDO calculated E_{HOMO} and so the resonance and non-resonance structures were separated. The following is the equation for non-resonance structures:

$$\Sigma\beta^H_2 = 0.125E_{HOMO} - 1.06Q_{MN} + 1.48 \qquad (9.48)$$

$n = 65 \quad r = 0.953 \quad s = 0.0704 \quad F = 307 \quad p = 0.000$

Even if phosphates, which are far from the rest of the compounds in the plot (Figure 9.15), are deleted, a comparable equation results:

$$\Sigma\beta^H_2 = 0.127E_{HOMO} - 1.28Q_{MN} + 1.44 \qquad (9.49)$$

$n = 62$ $r = 0.950$ $s = 0.0607$ $F = 271$ $p = 0.000$

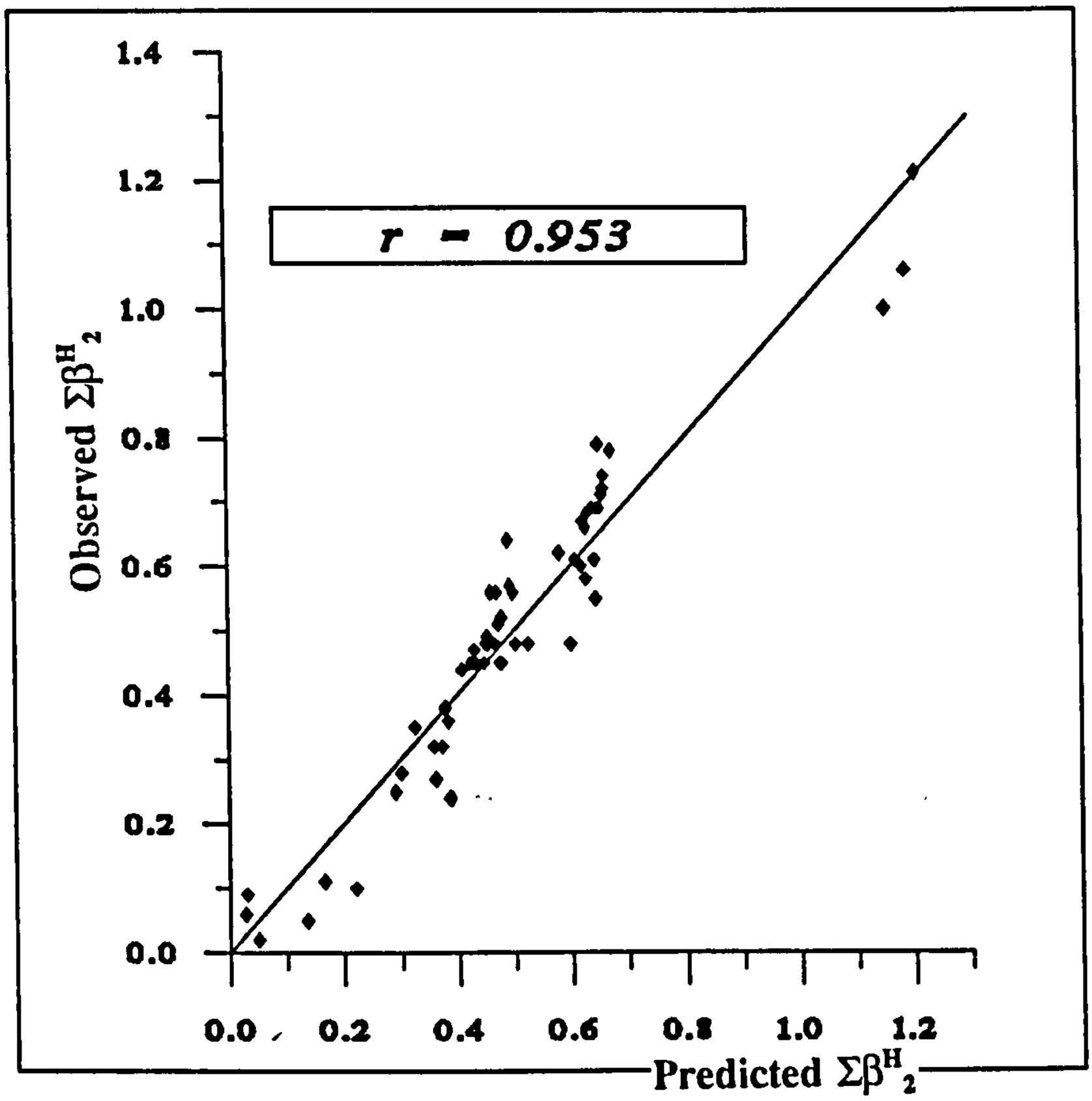


Figure 9.15. The plot between observed $\Sigma\beta^H_2$ and predicted $\Sigma\beta^H_2$ of H-bond bases, including the three phosphates, by equation 9.48.

Figure 9.16 shows the plot of $\Sigma\beta^H_2$ against its predicted values. In this equation the coefficient of E_{HOMO} has not changed significantly ($t = 0.247$). However, the coefficient of the Q_{MN} has reduced significantly ($t = 3.419$). As with MNDO calculated atomic charges, the AM1 method probably has not been parametrised properly to calculate atomic charges of the phosphates. The presence of both parameters is highly significant in both of the equations.

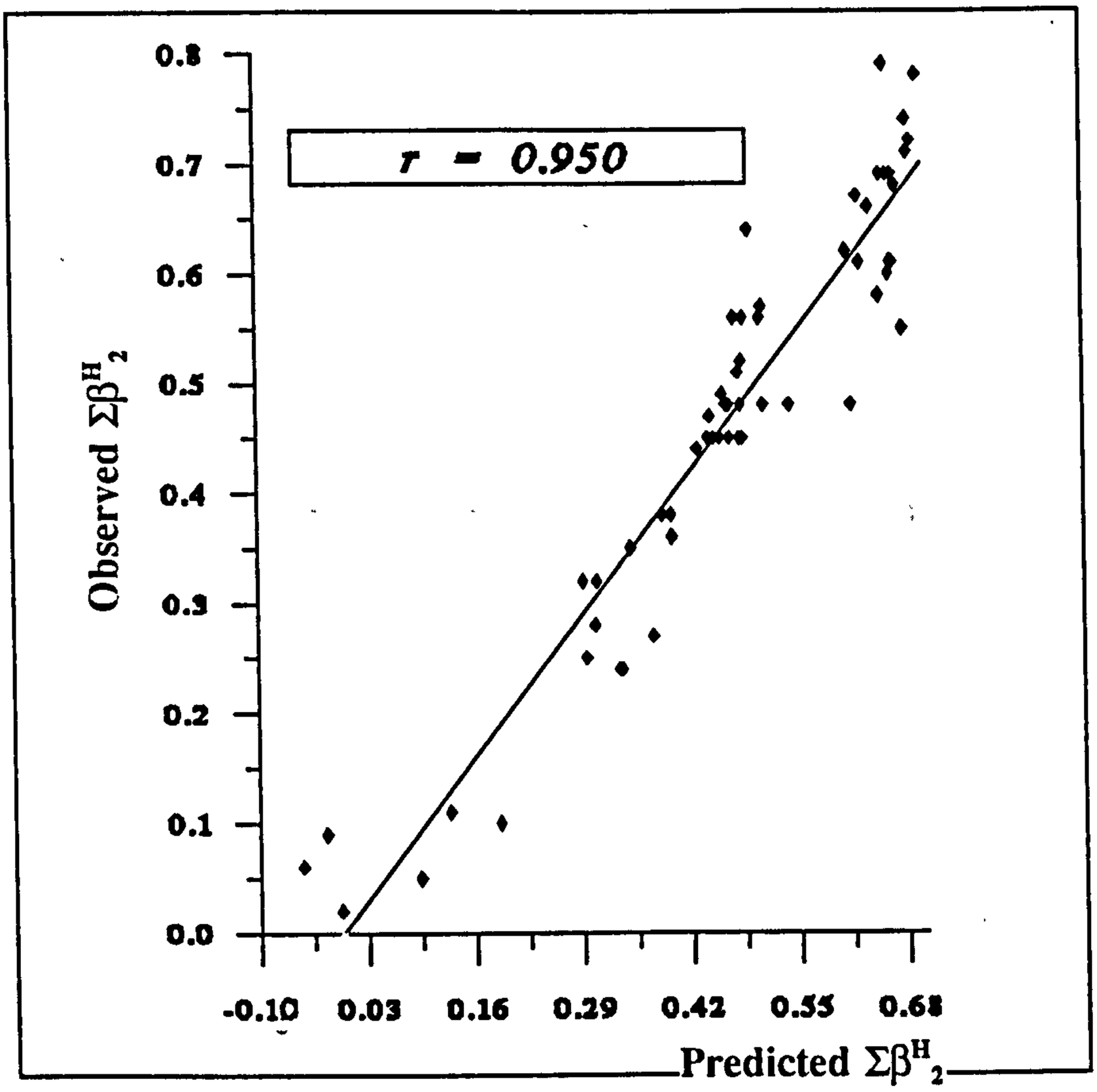


Figure 9.16. The plot between observed $\Sigma\beta^H_2$ and predicted $\Sigma\beta^H_2$ of H-bond bases excluding the three phosphates by equation 9.49.

9.2.3. *PM3 method*

a. Correlations within the families

Phenols: The following correlations were obtained for 3- and 4-substituted phenols:

$$\Sigma\alpha^H_2 = 25.6Q_H - 4.44 \qquad (9.50)$$

n = 13 r = 0.920 s = 0.309 F = 60.2 p = 0.000

$$\Sigma\alpha^H_2 = 0.639 - 0.127E_{LUMO} \qquad (9.51)$$

n = 13 r = 0.791 s = 0.0481 F = 18.4 p =0.001

E_{LUMO} and Q_H are quite highly correlated for this set of compounds (r = 0.819), so they could not be used as the two parameters of a multiple regression.

$$\Sigma\beta^H_2 = 0.285E_{HOMO} - 0.653Q_{MN} + 2.72 \qquad (9.52)$$

n = 13 r = 0.731 s = 0.0796 F = 5.7 p = 0.022

After deletion of the nitrophenols (for the reasons explained earlier) the correlations are:

$$\Sigma\beta^H_2 = -41.8Q_{MN} - 9.12 \qquad (9.53)$$

n = 11 r = 0.793 s = 0.0742 F = 15.3 p = 0.004

$$\Sigma\beta^H_2 = 0.295E_{HOMO} + 2.96 \qquad (9.54)$$

n = 11 r = 0.738 s = 0.0822 F = 10.8 p = 0.009

Although Q_{MN} and E_{HOMO} are not correlated, using both of the parameters in a multiple regression (after deletion of nitrophenols) was not satisfactory and the t-ratio for E_{HOMO} in such an equation was quite low.

Amides: Apparently, H-bond acceptor ability ($\Sigma\beta^H_2$) increases with increasing Q_{MN} values calculated after MM corrections to the amide bond rotation barrier. Comparing the atomic charges on the most negative atom in amides (the carbonyl oxygen), the order is:

formamide > acetamide > propanamide > N-methylformamide > N-methylacetamide > N-methylpropionamide > N,N-dimethylacetamide > N,N-dimethylformamide (from the highest negative Q_{MN} value to the lowest).

It is clear that Q_{MN} values increase (the amount of negative charge decreases) with addition of alkyl substituents to the nitrogen or carbonyl carbon, and this can explain the positive slope of Q_{MN} in the correlation with $\Sigma\beta^H_2$. Because the Q_{MN} values calculated by the MNDO method (Q_{MN} resides on the nitrogen atom of amides) correlates with $\Sigma\beta^H_2$ values, although here the most negatively charged atoms are the carbonyl oxygens, the charges on the nitrogen atoms were also examined in correlation with $\Sigma\beta^H_2$. These charges had no correlation with H-bond acceptor abilities.

A similar incorrect order is observed for Q_H values of amides (the charge on the hydrogen connected to the nitrogen) in which the more substituents the amide has, the higher is its Q_H value.

The following relationships with E_{HOMO} and E_{LUMO} resulted after deleting the outlier acetanilide from both of the relationships and N-methylformamide from equation 9.55:

$$\Sigma\beta^H_2 = 0.312E_{HOMO} + 3.7 \qquad (9.55)$$

n = 7 r = 0.870 s = 0.0306 F = 15.6 p = 0.011

$$\Sigma\alpha^H_2 = 1.72E_{LUMO} - 1.73 \qquad (9.56)$$

n = 6 r = 0.953 s = 0.0785 F = 58.9 p = 0.000

In this correlation E_{LUMO} unexpectedly has a positive coefficient.

The charge parameters calculated without MM corrections had no correlation with H-bond acidity or basicity. The only correlation is with E_{HOMO} values:

$$\Sigma\beta^H_2 = 0.185E_{HOMO} + 2.52 \qquad (9.57)$$

n = 7 r = 0.856 s = 0.0322 F = 13.6 p = 0.014

Alcohols: For H-bond basicity the following relationship for alcohols and ethers exists:

$$\Sigma\beta^{H}_{2} = 0.0807E_{HOMO} - 0.739Q_{MN} + 1.13 \qquad (9.58)$$

$n = 18 \quad r = 0.777 \quad s = 0.0576 \quad F = 11.4 \quad p = 0.001$

The t-ratios for Q_{MN} and E_{HOMO} are 2.93 ($p = 0.010$) and 3.85 ($p = 0.002$) respectively.

Deleting the π-bond structures improves the correlation:

$$\Sigma\beta^{H}_{2} = 0.103E_{HOMO} - 0.728Q_{MN} + 1.38 \qquad (9.59)$$

$n = 15 \quad r = 0.820 \quad s = 0.0570 \quad F = 12.3 \quad p = 0.001$

The H-bond donor ability of alcohols, after deleting water which was an outlier, showed the following relationship with Q_H:

$$\Sigma\alpha^{H}_{2} = 11.5Q_{H} - 1.74 \qquad (9.60)$$

$n = 12 \quad r = 0.728 \quad s = 0.0490 \quad F = 11.3 \quad p = 0.007$

But the plot (Figure 9.17) showed that 2,2,2-trifluoroethanol, because it was located far from the other points, unduly influenced the equation; deleting this point changed the correlation drastically so that, in the absence of this alcohol, the correlation between $\Sigma\alpha^{H}_{2}$ and Q_H had a negative slope with a r value of 0.734. The negative slope may be spurious, and arise mainly because of scatter. on the other hand, alkyl groups may have the inductive effect opposite to that in solution (and in the case of Q_{MN} values 2,2,2-trifluoroethanol has influenced equation 9.59).

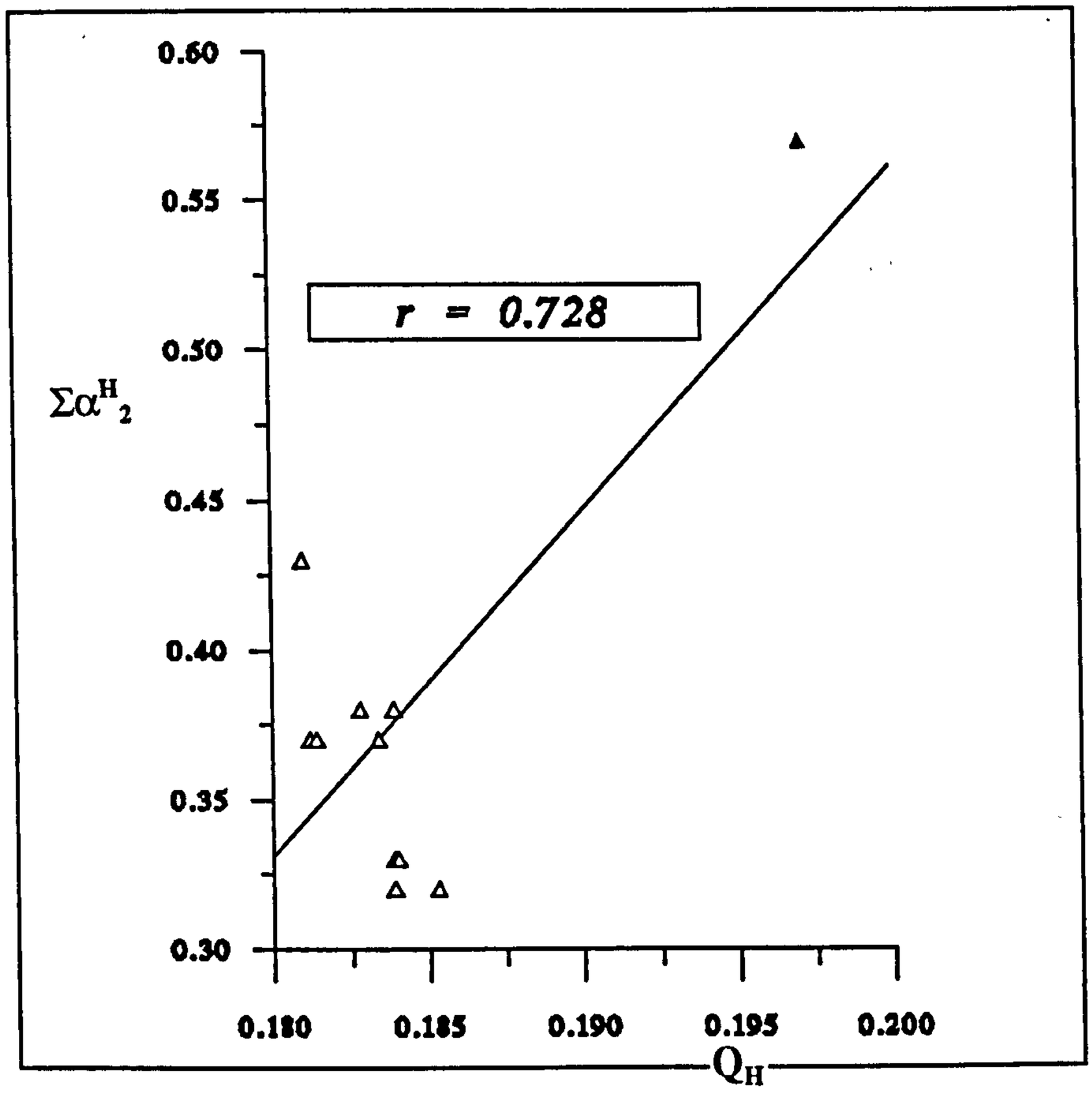


Figure 9.17. The plot between $\Sigma\alpha^H_2$ and Q_H of alcohols (△) including 2,2,2-trifluoroethanol (▲).

$\Sigma\alpha^H_2$ does not correlate with E_{LUMO} at all, and even after deleting the alcohols with π-bonds, there is unexpectedly a positive relationship between $\Sigma\alpha^H_2$ and E_{LUMO} with r value of 0.617.

Carboxylic acids: H-bond abilities of carboxylic acids had the following relationships with atomic charge parameters (there was no correlation with HOMO and LUMO energies at this stage):

$$\Sigma\alpha^H_2 = 28.0Q_H - 5.71 \qquad (9.61)$$

n = 12 r = 0.775 s = 0.0855 F = 15 p = 0.003

$$\Sigma\beta^H_2 = -2.28Q_{MN} - 0.503 \qquad (9.62)$$

n = 12 r = 0.744 s = 0.0437 F = 12.4 p = 0.006

The correlations of H-bond donor and acceptor ability of carboxylic acids with PM3 calculated parameters improved when the acids with resonance structures were deleted:

$$\Sigma\alpha^H_2 = 29.1Q_H - 5.91 \qquad (9.63)$$

n = 8 r = 0.860 s = 0.0785 F = 17 p = 0.006

$$\Sigma\alpha^H_2 = 0.836 - 0.219E_{LUMO} \qquad (9.64)$$

n = 8 r = 0.906 s = 0.0653 F = 27 p = 0.002

(Q_H and E_{LUMO} were highly correlated with a correlation coefficient of r = 0.981).

$$\Sigma\beta^H_2 = -3.13Q_{MN} - 0.810 \qquad (9.65)$$

n = 8 r = 0.903 s = 0.0354 F = 26.5 p = 0.002

Inclusion of E_{HOMO} did not improve the correlation with Q_{MN}.

If non-resonance esters are added to the correlation the following equation results:

$$\Sigma\beta^H_2 = -2.09Q_{MN} - 0.392 \qquad (9.66)$$

n = 11 r = 0.704 s = 0.0517 F = 8.8 p = 0.016

Amines: The PM3 method underestimates atomic charge on the nitrogen atoms and in some molecules it calculates the carbon atoms to be more negative than the nitrogen atom. Ignoring these irregularities, atomic charges on the nitrogen atoms of amines were used even if they were not the most negative atoms in the molecules. These charges showed the correct order in terms of the point charge on the nitrogen

atom; the primary amines had less negative charge than the secondary amines and the tertiary amines had the most negative nitrogen atoms. But atomic charge on the hydrogens connected to the nitrogens of the secondary amines were more positive than those of the primary amines, which is not what was anticipated. Accordingly, this method failed to model H-bond donor ability of the amines. The following is the expression for the H-bond acceptor ability of amines:

$$\Sigma\beta^H_2 = 0.131E_{HOMO} - 1.03Q_{MN} + 1.82 \qquad (9.67)$$

$n = 16 \quad r = 0.914 \quad s = 0.0501 \quad F = 33 \quad p = 0.000$

The PM3 method calculates an unreasonable atomic charge of +0.3634 on the nitrogen of pyrrole and this molecule has greatly influenced the above correlation.

For ammonia the atomic charge on the nitrogen atom is considerably overestimated (+0.0067) and for thiazole, considering the Q_{MN} value of the other amines, it is underestimated. Therefore their deletion gives rise to:

$$\Sigma\beta^H_2 = 0.134E_{HOMO} - 1.86Q_{MN} + 1.80 \qquad (9.68)$$

$n = 13 \quad r = 0.926 \quad s = 0.0235 \quad F = 30.2 \quad p = 0.000$

T-ratios for this equation are 6.87 and 6.63 for Q_{MN} and E_{HOMO} respectively.

b. Correlations across the different families of the compounds

For all the 111 compounds listed in the table the correlation of H-bond acidity with E_{LUMO} parameter was not successful. On the other hand the following is the relationship with Q_H parameter:

$$\Sigma\alpha^H_2 = 3.22Q_H - 0.149 \quad (9.69)$$

n = 111 r = 0.741 s = 0.1923 F = 132.9 p = 0.000

The plot of this equation (Figure 9.18) shows that this equation is not very reliable, because there are obviously two groups of compounds in the plot. If those structures which are not H-bond acids are omitted, the relationship for alcohols excluding water, 3- and 4-substituted phenols, carboxylic acids, amines and amides which have free hydrogen connected to the nitrogen is:

$$\Sigma\alpha^H_2 = 1.55Q_H - 0.0763E_{LUMO} + 0.322 \quad (9.70)$$

n = 55 r = 0.857 s = 0.1174 F = 71.8 p = 0.000

The following equation results if thiols and thiophenol are included:

$$\Sigma\alpha^H_2 = 2.25Q_H - 0.0453E_{LUMO} + 0.150 \quad (9.71)$$

n = 59 r = 0.797 s = 0.1505 F = 48.8 p = 0.000

The coefficients in equation 9.72 are clearly different from those in equation 9.71. Therefore, it can be concluded that the PM3 calculated Q_H and E_{LUMO} values for thiols and thiophenol cannot estimate the correct H-bonding abilities for and they are outliers from correlation 9.71. The plot for equation 9.71 is in Figure 9.19.

Compounds with and without resonance structures can also be separated:

H-bond acids without resonance effect in the structure:

$$\Sigma\alpha^H_2 = 1.44Q_H - 0.127E_{LUMO} + 0.478 \quad (9.72)$$

n = 32 r = 0.938 s = 0.0881 F = 102 p = 0.000

H-bond acids with resonance structures:

$$\Sigma\alpha^{H}_{2} = 2.66Q_{H} - 0.11E_{LUMO} + 0.054 \qquad (9.73)$$

n = 23 r = 0.761 s = 0.1092 F = 13.7 p = 0.000

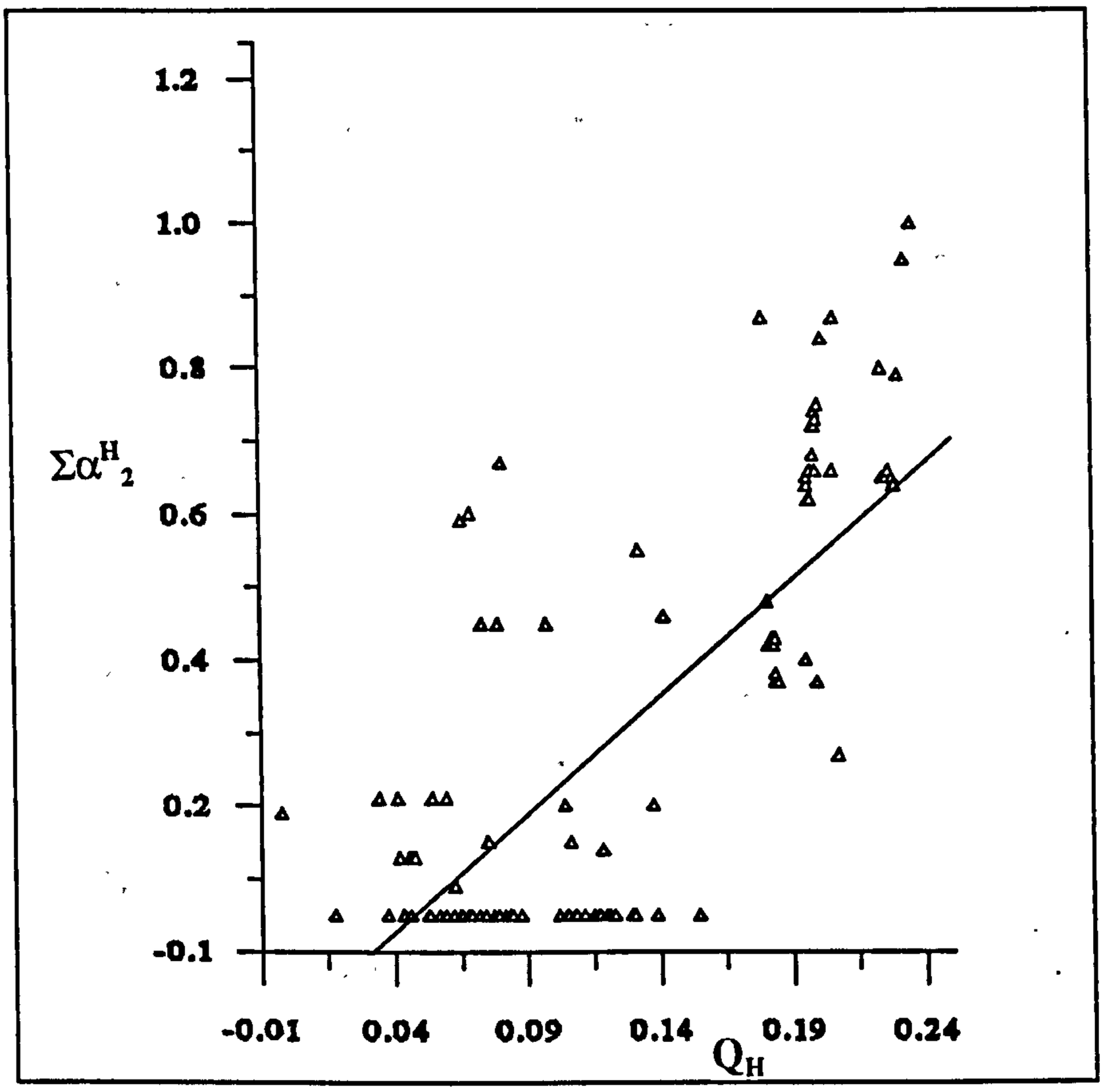


Figure 9.18. The plot of $\Sigma\alpha^{H}_{2}$ against Q_H calculated by the PM3 method of all the 111 compounds listed in Table 9.3.

The H-bond acceptor ability of all the compounds correlated only poorly with charge:

$$\Sigma\beta^{H}_{2} = -0.710Q_{MN} + 0.257 \qquad (9.74)$$

n = 111 r = 0.554 s = 0.1876 F = 48.2 p = 0.000

There was no correlation with E_{HOMO} at this stage. However, after the deletion of the

resonance structures the following equation was obtained:

$$\Sigma\beta^{H}_{2} = 0.179E_{HOMO} - 1.01Q_{MN} + 2.08 \qquad (9.75)$$

$n = 65 \quad r = 0.823 \quad s = 0.1320 \quad F = 65 \quad p = 0.000$

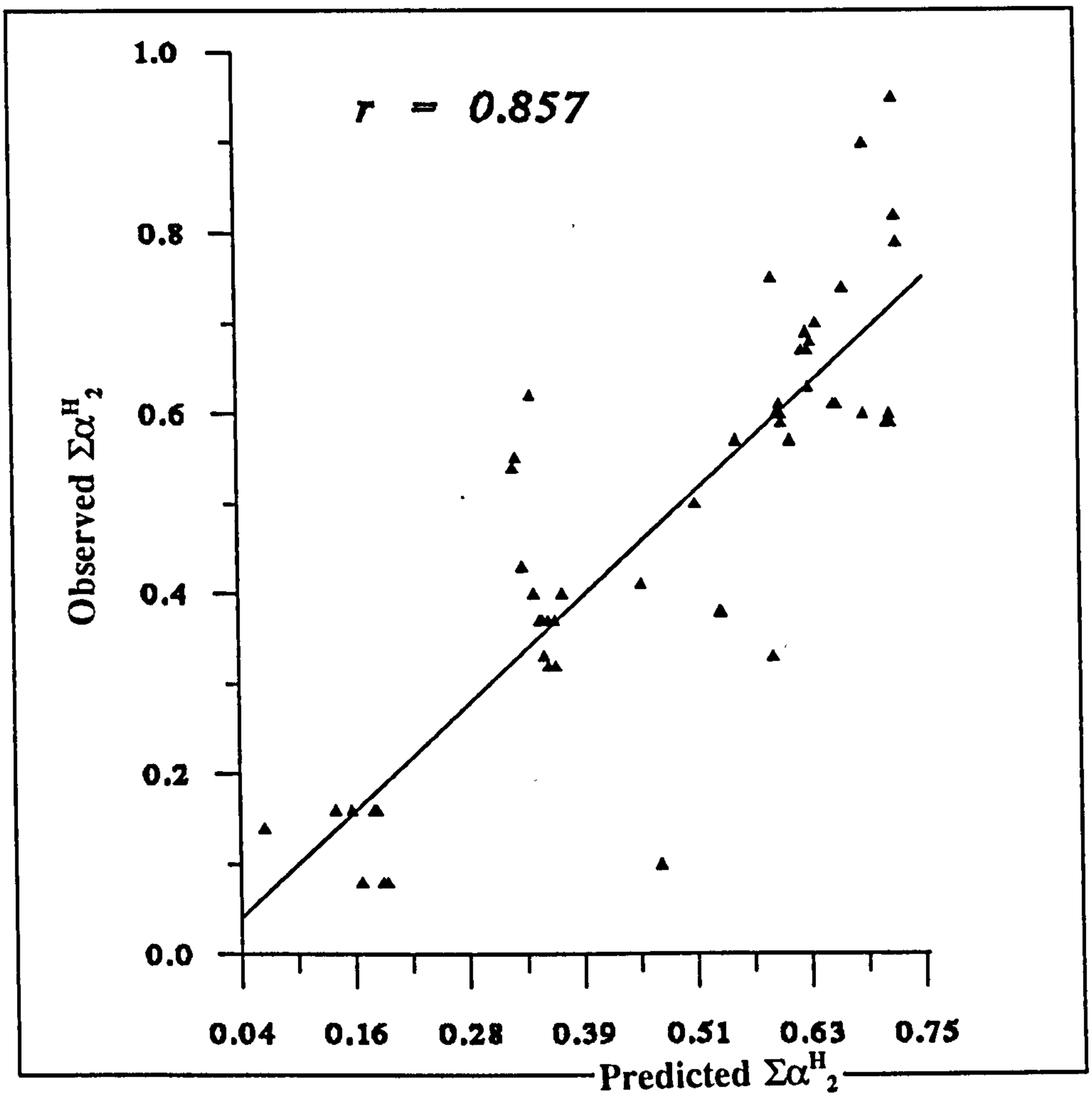


Figure 9.19. The plot $\Sigma\alpha^{H}_{2}$ against predicted $\Sigma\alpha^{H}_{2}$ by equation 9.70 for acids.

Phosphates had a large influence on the above equation (see Figure 9.20) when they were deleted the equation changed to:

$$\Sigma\beta^{H}_{2} = 0.170E_{HOMO} - 0.876Q_{MN} + 2.01 \qquad (9.76)$$

$n = 62 \quad r = 0.736 \quad s = 0.1313 \quad F = 35 \quad p = 0.000$

which obviously is not as good as the former equation. The reason, as is shown in Figures 9.20 and 9.21, is the distance of the phosphates from the rest of the compounds. In this equation ammonia was an outlier. The following equation resulted from deletion of this compound:

$$\Sigma\beta^H_2 = 0.170E_{HOMO} - 0.924Q_{MN} + 2.00 \qquad (9.77)$$

n = 61 r = 0.754 s = 0.1277 F = 38 p = 0.000

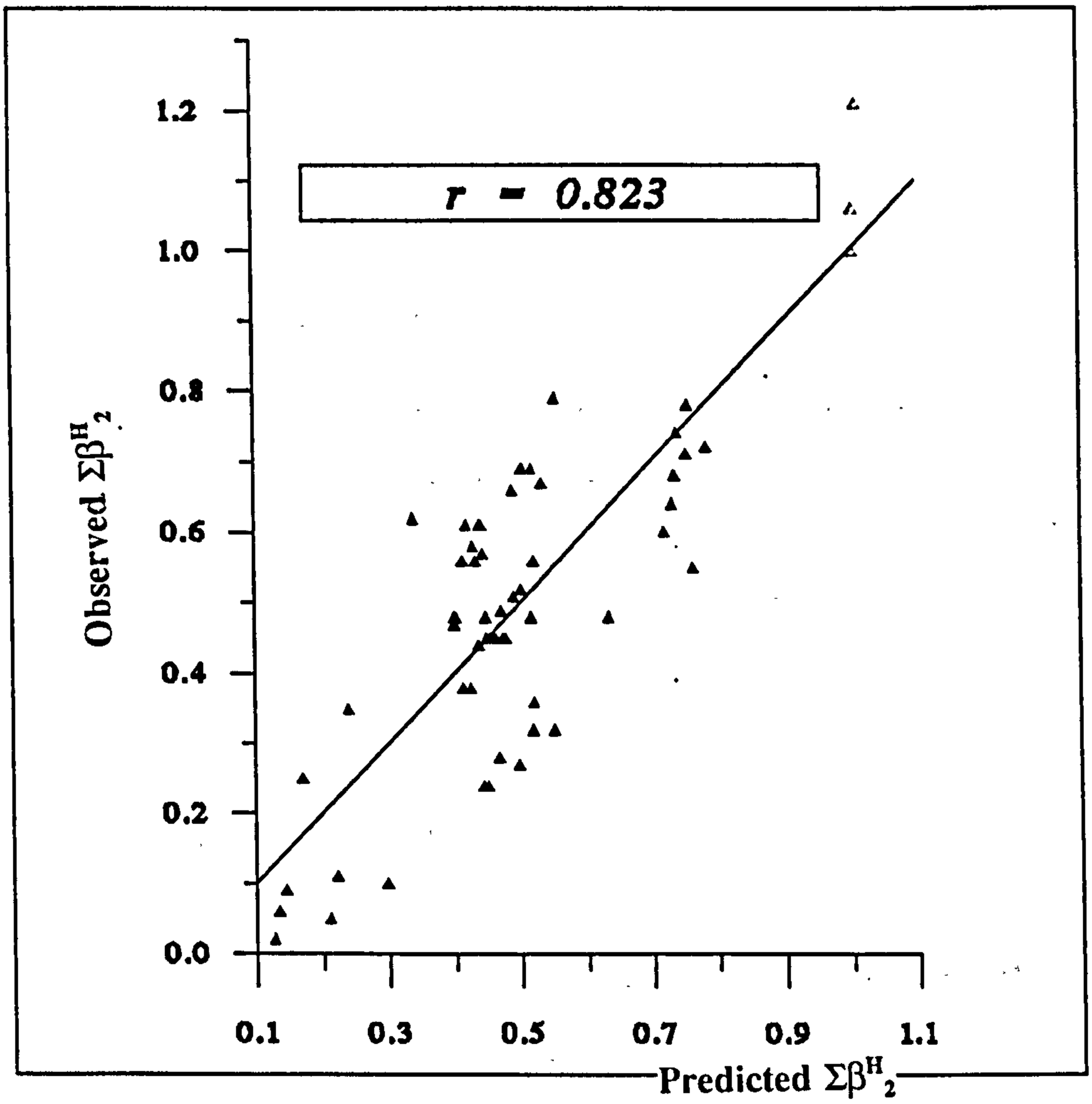


Figure 9.20. The plot of observed $\Sigma\beta^H_2$ against predicted $\Sigma\beta^H_2$ by equation 9.75 (including phosphates); ∆:phosphates, ▲:rest of the compounds.

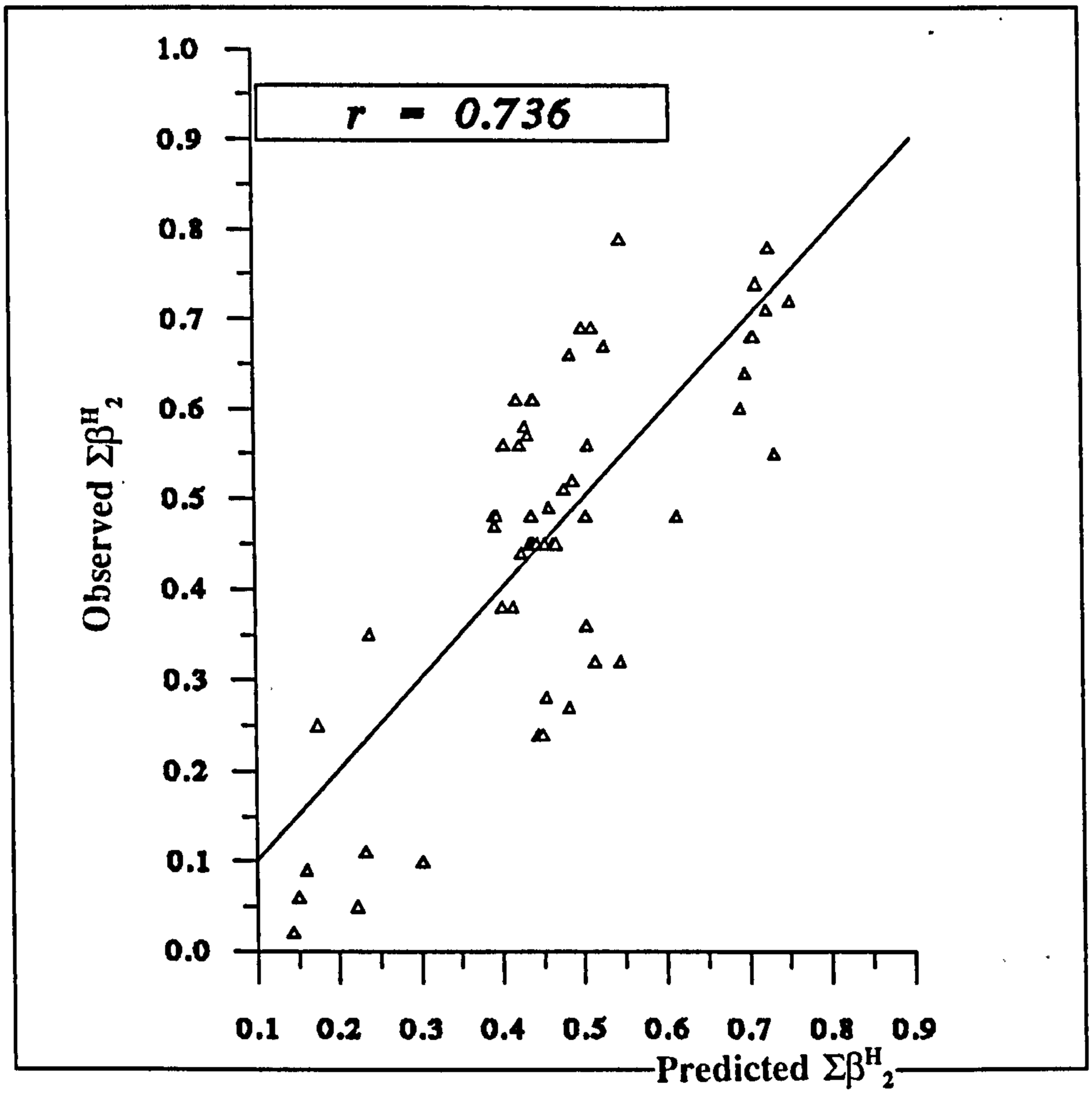


Figure 9.21. The plot of $\Sigma\beta^H_2$ against predicted $\Sigma\beta^H_2$ by equation 9.76 (excluding phosphates).

9.2.4. Comparison between the three methods

In order to compare the three methods of MNDO, AM1 and PM3, Stewart (1990) has listed some measured dipole moments (μ_{exp}) and also the dipole moments calculated by these methods. A correlation between these values showed that AM1 calculated dipole moments have the best agreement with the experimental values:

$$\mu_{exp} = 0.928\mu_{AM1} + 0.113 \qquad (9.78)$$

n = 123 s = 0.455 r = 0.858 F = 337

$$\mu_{exp} = 0.842\mu_{PM3} + 0.229 \quad (9.79)$$

n = 123 s = 0.497 r = 0.827 F = 262.3

$$\mu_{exp} = 0.757\mu_{MNDO} + 0.342 \quad (9.80)$$

n = 123 s = 0.578 r = 0.757 F = 162.2

E_{HOMO} calculated by these method can also be compared with the experimental values (ionisation potentials, IP). The following equation resulted from correlations with IPs listed in Table 8.1:

AM1: $IP = 0.938\ E_{HOMO} - 0.212$ (9.81)

n = 36 s = 0.416 r = 0.931 F = 219.4

MNDO: $IP = 0.872\ E_{HOMO} - 0.656$ (9.82)

n = 36 s = 0.447 r = 0.919 F = 185

PM3: $IP = 0.990\ E_{HOMO} + 0.23$ (9.83)

n = 36 s = 0.572 r = 0.864 F = 100.1

Clearly, AM1 method predicts best the IPs.

A summary of the correlations across the different classes of compounds is given in Table 9.4 and 9.5. In the correlations of the H-bond acids using the AM1 method, pyrrole was an outlier and has been omitted from regressions. Its inclusion worsens the statistics of the equation to r = 0.908 and s = 0.1052. In the PM3 method as well as the MNDO method, deletion of pyrrole does not change the equations significantly. By and large, it can be concluded from Table 9.4 that in correlations of a single parameter with $\Sigma\alpha^{H}_{2}$, MO parameters calculated by the AM1 method have produced equations with the best statistics, with the MNDO method being the second best. However, in a multiple regression, parameters calculated by MNDO method work best.

Table 9.4. Comparing the statistical results of the correlations between $\Sigma\alpha^H_2$ and the parameters calculated by different MO methods

$\Sigma\alpha^H_2 = A\ (Q_H) + B\ (E_{LUMO}) + C$

Method	Parameters	A	B	C	n	r	s	F
MNDO	Q_H	4.86	-	-0.390	55	0.880	0.107	182.6
	E_{LUMO}	-	-0.113	0.618	55	0.793	0.137	89.7
	Q_H&E_{LUMO}	3.51	-0.051	-0.086	55	0.918	0.090	138.6
AM1	Q_H	5.82	-	-0.729	54	0.896	0.101	212.3
	E_{LUMO}	-	-0.111	0.632	54	0.796	0.138	89.5
	Q_H&E_{LUMO}	4.42	-0.040	-0.384	54	0.917	0.092	134.7
PM3	Q_H	2.45	-	0.088	55	0.759	0.147	71.9
	E_{LUMO}	-	-0.119	0.618	55	0.763	0.146	73.7
	Q_H&E_{LUMO}	1.55	-0.076	0.322	55	0.857	0.117	71.8

Table 9.5. Comparing the statistical results of the correlations between $\Sigma\beta^H_2$ and the parameters calculated by different MO methods

$\Sigma\beta^H_2 = A\ (Q_{MN}) + B\ (E_{HOMO}) + C$

Method	Parameters	A	B	C	n	r	s	F
MNDO	Q_{MN}	-1.64	-	-0.045	57	0.739	0.129	66.2
	E_{HOMO}	-	0.207	2.770	57	0.894	0.086	218.8
	Q_{MN}&E_{HOMO}	-0.71	0.163	2.050	61	0.931	0.071	173.9
AM1	Q_{MN}	-1.41	-	0.055	57	0.729	0.131	62.6
	E_{HOMO}	-	0.170	2.300	57	0.767	0.122	78.9
	Q_{MN}&E_{HOMO}	-1.28	0.127	1.440	62	0.950	0.061	271.0
PM3	Q_{MN}	-	-	-	-	-	-	-
	E_{HOMO}	-	-	-	-	-	-	-
	Q_{MN}&E_{HOMO}	-0.92	0.170	2.00	61	0.754	0.128	38.0

For H-bond bases, in order to compare the methods, we have used the equations in

which phosphates are excluded. Thiols and sulphides were outliers from equations with only one parameter and have been deleted from such correlations. Although it has been shown in the previous sections that ammonia is an outlier from equations of both MNDO and PM3 calculated parameters, even after deleting this outlier from equations of these two methods, the AM1 method still gives the best results and MNDO is the second best in multiple regression equations. However, in correlations with a single parameter, MNDO calculated parameters are the best.

For phenols Q_H and E_{LUMO} are highly correlated with each other, while, Q_{MN} and E_{HOMO} are not. The results for this compounds (Table 9.6) show that the MNDO calculated Q_H and E_{LUMO} come first in the prediction of the $\Sigma\alpha^H_2$ values and those calculated by AM1 and PM3 are the second and the third respectively. On the other hand, both E_{HOMO} and Q_{MN} calculated by AM1 method are better predictors of the H-bond acceptor ability in comparison with PM3 and MNDO methods. However, when using both parameters in a multiple regression, the MNDO calculated parameters give the best correlation.

Table 9.6. Comparison between the MO parameters calculated by three different methods in correlating with $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ of 3- and 4-substituted phenols

Methods	Q_H	E_{LUMO}	Q_H & E_{LUMO}	Q_{MN}	E_{HOMO}	Q_{MN} & E_{HOMO}
MNDO	n = 13 r = 0.964 s = 0.021	n = 13 r = 0.829 s = 0.044	-	n = 11 r = 0.746 s = 0.081	n = 11 r = 0.808 s = 0.072	n = 11 r = 0.954 s = 0.039
AM1	n = 13 r = 0.959 s = 0.022	n = 13 r = 0.819 s = 0.045	-	n = 11 r = 0.838 s = 0.066	n = 11 r = 0.831 s = 0.068	n = 11 r = 0.885 s = 0.060
PM3	n = 13 r = 0.920 s = 0.031	n = 13 r = 0.791 s = 0.048	-	n = 11 r = 0.793 s = 0.074	n = 11 r = 0.738 s = 0.082	-

Table 9.7 shows the relationships of H-bond experimental acidity and basicity of amides with MO parameters calculated by different methods. As was the case for phenols, MNDO calculated parameters seem to work best here; it is the only method by which charge parameters have correlations with experimental H-bond parameters. In the AM1 and PM3 methods, charge parameters either have an order opposite to that suggested by the inductive effect of alkyl groups in solution or do not follow any particular order. E_{HOMO} values calculated by all the methods have good relationships with $\Sigma\beta^H_2$, while E_{LUMO} has a positive slope in the relationships with $\Sigma\alpha^H_2$ (which exists only when E_{LUMO} is calculated by AM1 or PM3 methods).

Table 9.7. Comparison between the MO parameters calculated by three different methods in correlating with $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ of amides

Parameters	Q_H			E_{LUMO}			Q_{MN}			E_{HOMO}		
	n	r	s	n	r	s	n	r	s	n	r	s
MNDO(MM)	6	0.942	0.033	-			8	0.653	0.044	7	0.915	0.025
MNDO	reverse order			-			8	0.810	0.034	7	0.922	0.024
AM1(MM)	-			6	0.631	0.084	reverse order			7	0.902	0.027
AM1	-			6	0.895	0.048	reverse order			7	0.903	0.027
PM3(MM)	reverse order			7	0.953	0.079	reverse order			7	0.870	0.031
PM3	-			-			-			7	0.856	0.032

-: where there is no correlation.

For alcohols, Q_H and Q_{MN} values calculated by all the three methods showed that alkyl groups have an electron-withdrawing inductive effect (except for MNDO calculated Q_H values which have a poor positive relationship with $\Sigma\alpha^H_2$). If 2,2,2-trifluoroethanol is included, the coefficients on Q_{MN} and Q_H have the correct signs, but excluding this alcohol, which is the only alcohol with a strong electron-withdrawing substituent and therefore with dominant difference in its H-bonding strength with the rest of the

alcohols, the wrong signs for the coefficients on Q_{MN} and Q_H parameters is obtained. The E_{LUMO} for these structures has not been successful for correlation with $\Sigma\alpha^H_2$ in either method; in the MNDO and PM3 methods, unexpectedly, E_{LUMO} has a positive slope in correlation with $\Sigma\alpha^H_2$; in the AM1 method there is no correlation between them at all. Because of the narrow range of the experimental and also the theoretical parameters, the inadequacy of the theoretical parameters in this particular class of compounds has not affected the good correlations between the different classes. Among these methods, E_{HOMO} calculated by the MNDO method correlates best with $\Sigma\beta^H_2$ of single bonded ethers and alcohols with r value of 0.779 in comparison with AM1 method (r = 0.673) and PM3 (r = 0.663).

A summary of the results of correlations for carboxylic acids (and esters in the case of the regressions for basicity) is shown in Table 9.8. Clearly, Q_H of the carboxylic acids, calculated by these three methods, shows very good correlations with H-bond acidity. The best method here is the MNDO method, with the AM1 method being the second best. Deletion of resonance structures improves the correlations for PM3 and AM1 methods but the MNDO method is still the best. In correlation between $\Sigma\beta^H_2$ and Q_{MN}, when aromatic acids are included, correlations are poor and for carboxylic acids alone the best correlation is with PM3 calculated charges (r = 0.775); the AM1 method gives the second best (r = 0.711) with MNDO giving the poorest correlation (r = 0.665). However, after deleting the four benzoic acids, it is clear in the table that the MNDO method is the best and AM1 is the second best. LUMO and HOMO energies (calculated by any of the three methods) have no correlation with $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ unless the aromatic structures are deleted (as in Table 9.7). E_{LUMO} calculated by the AM1 method is better than those calculated by the two methods of MNDO and AM1. E_{HOMO} calculated by PM3 has no correlation with $\Sigma\beta^H_2$ and MNDO calculated E_{HOMO} is the best.

Table 9.8. Comparison between the MO parameters calculated by three different methods in correlating with $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ of carboxylic acids (and esters in basicity analyses)

Parameters	Q_H			E_{LUMO}			Q_{MN}			E_{HOMO}		
	n	r	s	n	r	s	n	r	s	n	r	s
MNDO	12	0.934	0.048	8	0.908	0.065	11	0.832	0.040	11	0.902	0.031
AM1	12	0.858	0.069	8	0.922	0.060	11	0.714	0.051	11	0.704	0.052
PM3	12	0.744	0.086	8	0.906	0.065	11	0.704	0.052		-	

Q_H and E_{LUMO} calculated by all the semiempirical methods have little or no correlation with $\Sigma\alpha^H_2$ of amines. In all the methods, the order of Q_H values are opposite to that expected according to the inductive effect of alkyl groups in solution. However, the gas-phase acidities of amines are not according to the inductive effect of alkyl groups either (Brauman & Blair, 1971). The PM3 method underestimates the atomic charge on the nitrogen atoms but these charges, like those calculated by the MNDO method, show the correct order. Therefore MNDO and PM3 calculated Q_{MN} and E_{HOMO} correlate with $\Sigma\beta^H_2$ (although there are some outliers).

H-bond basicity of some aromatic rings without any other hydrogen bonding group in their structure can be predicted by E_{HOMO} values while Q_{MN} is not successful:

MNDO: $\Sigma\beta^H_2 = 0.135E_{HOMO} + 1.40$ (9.84)

$n = 11$ $r = 0.895$ $s = 0.0252$ $F = 36.3$ $p = 0.000$

AM1: $\Sigma\beta^H_2 = 0.133E_{HOMO} + 1.38$ (9.85)

$n = 11$ $r = 0.893$ $s = 0.0254$ $F = 35.7$ $p = 0.000$

PM3: $\Sigma\beta^H_2 = 0.124E_{HOMO} + 1.31$ (9.86)

$n = 11$ $r = 0.815$ $s = 0.0328$ $F = 17.8$ $p = 0.002$

9.3. Conclusion

It can be concluded from the correlations and graphs given in this chapter that the parameters calculated by semiempirical methods in MOPAC program are useful in prediction of experimental H-bonding abilities. The AM1 and MNDO methods give the better correlations in families and also between the families than the PM3 method. One important disadvantage of the PM3 method is underestimation of negative charge on nitrogen atoms and of positive charge on the hydrogens connected to them (this is the reason for the poor statistics of general correlations obtained for this method).

The effect of alkyl substitution on charge parameters in different families is controversial. In amines, amides and alcohols these groups seem to have an electron-withdrawing effect in AM1 and PM3 methods. The MNDO calculation of Q_{MN} values for amines and amides, and Q_H values of alcohols and amides show an electron-donating inductive effect for alkyl groups.

Within families of compounds, in order to find any relationship with E_{LUMO} and E_{HOMO}, it is always necessary to separate out the conjugated structures. In the correlation with E_{HOMO} (and Q_{MN}) for different classes of compounds, the aromatic structures had to be separated, although this is not required in correlation with E_{LUMO} (or Q_H).

Finally, in correlations within families, the energies of frontier orbitals and charge parameters cannot be used together because they are highly correlated (an exception from this rule is Q_{MN} and E_{HOMO} values of phenols). Across the different families, on the other hand, it is essential to use both of the parameters in order to improve the prediction of H-bonding ability and to reduce the family dependent behaviour.

10. Molecular electrostatic potentials as hydrogen bonding descriptors

In previous chapters atomic charge was used as an electrostatic parameter to predict hydrogen bonding abilities, and good correlations with H-bonding experimental parameters both within and across families of compounds were found. Unfortunately, atomic charge is not a defined physical property and it cannot explain the directionality of hydrogen bonding.

An alternative strategy for deriving electrostatic descriptors is to use molecular electrostatic potential (ESP). Unlike net atomic charge, the molecular electrostatic potential is a rigorously defined quantum mechanical property. Electrostatic potential at each point r around a molecule is the electrostatic interaction energy between the molecule and a point charge placed at the point r; the value of the ESP reflects the effects of all the charges present in the molecule. ESP has been evaluated for proton donors and proton acceptors at a reasonable "representative" point in space to predict *ab initio* calculated hydrogen bonding energy (Kollman et al, 1975). Murray et al (Murray & Politzer, 1991&1992; Murray et al, 1991)) employed electrostatic potential for the prediction of H-bond ability. They found relationships between solute as well as solvent H-bond acidity and ESP on the surfaces defined by the 0.002 electron/bohr^3 contour of the STO-5G electronic density. On the other hand, they used the electrostatic potential local minima (V_{min}) to correlate with solute and solvent H-bond acceptor ability. For a set of heterocycles with nitrogen as the H-bond acceptor, V_{min} was shown to be an excellent predictor of H-bond basicity and the ability of ESP and also magnitude of the electric field strength at points along the lone pair axis have been maximised at specific distances from the nitrogen to fit the experimental hydrogen bond basicity data (Kenny, 1994).

In this study ESPs on the surface of the molecule were intended to be used as H-

bonding descriptors (since it is the part of the molecule that interacts with other molecules). For small molecules the van der Waals surface gives a good representation of the outer surface and overall shape. But, for larger molecules, most of the van der Waals surface is buried in the interior. Richards (1977) presented a definition of a suitable surface. This molecular surface consists of two parts: the contact surface and the reentrant surface. The contact surface is that part of the van der Waals surface of the atoms that is accessible to a probe sphere representing a solvent molecule. The reentrant surface comes from the inward-facing surface of the probe sphere when it is simultaneously in contact with more than one atom. Connolly (1983&1985) has presented a computer algorithm for calculating this surface. Connolly's solvent accessible surface is created by rolling a probe sphere over the molecule. The resulting surface contour is made up of pieces of spheres and tori that join at circular arcs. The spheres, tori and arcs are defined by analytical expressions in terms of the atomic coordinates, van der Waals radii and the probe radius.

10.1. Methods and experimental data

111 different compounds, including carboxylic acids, phenols, alcohols, ethers, and amines for which Abraham's experimental hydrogen bonding parameters (Abraham, 1993) were available, were used in this study.

The global minimum energy conformation for each molecule was found by conformational analysis using the Cobra program in the Oxford Molecular (OM) software. A further COSMIC minimisation was also performed for molecules under study in Pimms (OM). Using the MAD program (OM) atomic charges were calculated by the two available methods, namely the Gasteiger (Gasteiger et al, 1980) and Abraham methods (Abraham & Grant, 1988; Abraham & Haworth, 1988; Abraham & Smith, 1988&1989). The most positive charge on the hydrogen atom (Q_H) and the

most negative charge on the heteroatoms (Q_{MN}) of a molecule were used as measures of its H-bonding acidity and basicity respectively. Furthermore, the charges were used to calculate molecular electrostatic potentials. The probe size 1.05 Å was used to generate a Connolly surface with the scaling factor of 1 to the van der Waals radius, and molecular electrostatic potentials for points on this surface were calculated using the procedure described by Giessner-Prettre and Pullman (1972). The dot density used was 10 dot/Å^2. The most negative and the most positive ESPs on this surface for each molecule were chosen to represent its H-bonding basicity and acidity respectively.

The energy of the highest occupied molecular orbital (E_{HOMO}) and the lowest unoccupied molecular orbital (E_{LUMO}) were calculated in MOPAC 6.0 using the semiempirical method of MNDO. Regression analyses were carried out using the MINITAB statistical package.

10.2. Results and discussion

In Figure 10.1, an example of electrostatic potentials on the Connolly surface has been shown for two different conformations of 2-methoxyphenol.

For the compounds studied the highest and the lowest electrostatic potentials on their Connolly surfaces (ESP^+ and ESP^- respectively) were found to be in the vicinity of the atoms of the most positive and the most negative atomic point charge. These atoms were, respectively, hydrogen atoms connected to heteroatoms, and heteroatoms which were capable of hydrogen bonding. In aromatic structures which did not contain any of the heteroatoms capable of H-bonding (N, O, F), the location of the lowest ESP was above the plane of the aromatic ring; however, the highest ESP was in the plane of the ring. Dimethylbenzenesulphonamide had its lowest ESP around the oxygen atoms. For substituted phenols the ESP^+s calculated from both Abraham and Gasteiger charges

were about the hydroxyl hydrogen. For ESPs calculated using Abraham charges, although the hydroxyl oxygen of the nitrophenols was more negatively charged than the nitro oxygens (-0.42 and -0.32 respectively), the ESP^-s were between the nitro oxygens and the ESP^-s in the remaining phenols were in the vicinity of the hydroxyl oxygen. The reason could be the co-operative effects of the two nitro oxygens in the region. In ESPs calculated from Gasteiger charges the ESP^-s were near the hydroxyl oxygen with the exception of the methoxy phenols; ESP^-s in methoxy phenols were associated with the methoxy oxygen. In thiophenol although the ESP^- was above the plane of the ring and not close to the sulphur, the ESP^+ was close to the hydrogen connected to the sulphur.

The results of the computations (ESP^+, Q_H, E_{LUMO}) and also experimental H-bond donor ability of Abraham ($\Sigma\alpha^H_2$) for hydrogen bonding acids are listed in Table 10.1. Table 10.2 gives the values of the experimental H-bond acceptor ability and also the computational parameters (ESP^-, Q_{MN}, E_{HOMO}) for hydrogen bonding bases.

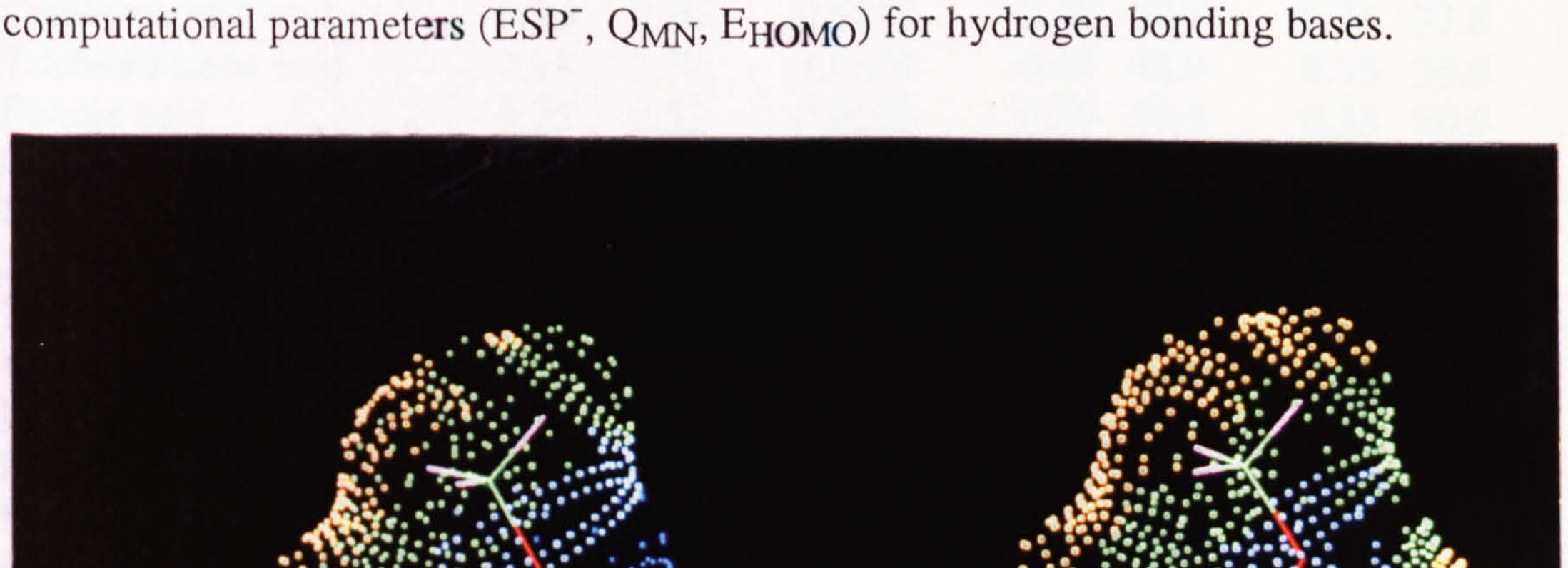


Figure 10.1. Electrostatic potentials on the Connolly surface of 2-methoxyphenol

Table 10.1. $\Sigma\alpha^H_2$ values and other parameters for hydrogen bonding acids; predicted $\Sigma\alpha^H_2$ values are the values predicted by equation 10.4

Compound	Exptl. $\Sigma\alpha^H_2$	Predicted $\Sigma\alpha^H_2$	E_{LUMO}	Gasteiger		Abraham	
				Q_H	ESP^+	Q_H	ESP^+
Diethylamine	0.08	0.13	4.3399	0.17	17.9	0.28	22.1
Methylamine	0.16	0.13	3.7070	0.16	16.6	0.27	21.4
Ethylamine	0.16	0.14	3.4520	0.16	17.5	0.27	21.4
n-Propylamine	0.16	0.15	3.3716	0.16	19.8	0.27	21.5
n-Butylamine	0.16	0.15	3.2060	0.16	21.2	0.27	21.4
Dimethylamine	0.08	0.18	3.3257	0.17	15.6	0.27	22.6
Di-n-propylamine	0.08	0.19	2.9830	0.17	21.3	0.27	22.4
Di-n-butylamine	0.08	0.20	2.9147	0.17	23.4	0.27	22.4
Formamide	0.62	0.42	1.5173	0.17	34.0	0.20	27.7
Acetamide	0.54	0.36	1.3458	0.16	24.8	0.20	25.5
Propionamide	0.55	0.34	1.3930	0.16	24.9	0.20	24.9
N-Methylformamide	0.40	0.32	1.4263	0.18	33.9	0.20	24.5
N-Methylpropionamide	0.40	0.35	1.3458	0.18	28.6	0.20	25.2
N-Methylacetamide	0.40	0.37	1.2977	0.18	28.1	0.20	25.8
Acetic acid	0.61	0.48	0.8510	0.29	27.2	0.33	28.8
Hexanoic acid	0.60	0.48	0.9032	0.29	30.0	0.33	29.0
Chloroacetic acid	0.74	0.64	-0.2019	0.29	44.6	0.33	32.4
Dichloroacetic acid	0.90	0.70	-0.6558	0.29	54.8	0.33	33.8
Trichloroacetic acid	0.95	0.77	-1.0437	0.29	48.9	0.33	35.6
Formic acid	0.75	0.53	0.9603	0.29	34.5	0.33	30.5
Propanoic acid	0.60	0.48	0.9037	0.29	27.5	0.33	28.8
Butanoic acid	0.60	0.48	0.9032	0.29	27.8	0.33	29.0
2-Methylbenzoic acid	0.60	0.59	-0.2407	0.29	30.3	0.33	30.9
3-Methylbenzoic acid	0.59	0.60	-0.5190	0.29	29.5	0.33	30.8
4-Methylbenzoic acid	0.60	0.59	-0.2977	0.29	30.0	0.33	30.9
Methanol	0.43	0.38	3.7953	0.28	28.2	0.33	29.5
Ethanol	0.37	0.38	3.5149	0.28	28.0	0.33	29.1
Propan-1-ol	0.37	0.38	3.2527	0.28	29.2	0.33	28.9
Propan-2-ol	0.33	0.37	3.3379	0.28	31.9	0.33	28.6
Butan-1-ol	0.37	0.39	3.1939	0.28	31.4	0.33	29.1
Hexan-1-ol	0.37	0.39	3.1495	0.28	33.3	0.33	29.1
2,2,2-Trifluoroethanol	0.57	0.60	1.4265	0.28	114.4	0.33	33.6
Cyclopentanol	0.32	0.37	3.0996	0.28	31.2	0.33	28.4
Cyclohexanol	0.32	0.38	3.0662	0.28	33.0	0.33	28.7
Prop-2-en-1-ol	0.38	0.44	0.8886	0.28	25.4	0.33	27.5
trans-But-2-en-1-ol	0.38	0.49	0.7299	0.28	28.9	0.33	29.0
Ethylthiol	0.00	-0.02	1.8799	0.14	19.3	0.13	13.8
n-Propylthiol	0.00	-0.02	1.8907	0.14	20.3	0.13	14.1
n-Butylthiol	0.00	-0.01	1.8858	0.14	22.5	0.13	14.2
Benzylamine	0.10	0.24	0.0564	0.16	20.8	0.27	19.9

Table 10.1. Continued

Compound	Exptl. $\Sigma\alpha^H_2$	Predicted $\Sigma\alpha^H_2$	E_{LUMO}	Gasteiger Q_H	Gasteiger ESP^+	Abraham Q_H	Abraham ESP^+
Acetanilide	0.50	0.49	0.1568	0.19	33.7	0.21	28.3
Benzoic acid	0.59	0.60	-0.2337	0.29	29.3	0.33	31.0
Phenol	0.60	0.61	0.2509	0.29	33.2	0.33	32.3
2-Fluorophenol	0.61	0.37	-0.2149	0.29	56.3	0.33	23.8
3-Fluorophenol	0.68	0.73	-0.2026	0.29	51.6	0.33	35.5
4-Fluorophenol	0.63	0.73	-0.1727	0.29	52.3	0.33	35.5
2-Chlorophenol	0.32	0.33	-0.1734	0.29	41.3	0.33	34.9
3-Chlorophenol	0.69	0.71	-0.2096	0.29	43.1	0.33	34.8
4-Chlorophenol	0.67	0.62	-0.1799	0.29	58.4	0.33	32.0
2-Bromophenol	0.35	0.09	-0.1204	0.29	37.2	0.33	34.6
3-Bromophenol	0.70	0.68	-0.1645	0.29	39.7	0.33	34.0
4-Bromophenol	0.67	0.70	-0.1319	0.29	38.1	0.33	34.5
2-Methoxyphenol	0.22	0.35	0.1780	0.29	32.4	0.33	23.7
3-Methoxyphenol	0.59	0.70	0.1794	0.29	39.3	0.33	34.9
4-Methoxyphenol	0.57	0.67	0.0652	0.29	37.1	0.33	34.0
2-Nitrophenol	0.05	0.25	-0.9596	0.29	23.5	0.33	18.8
3-Nitrophenol	0.79	0.84	-0.9311	0.29	39.8	0.33	37.9
4-Nitrophenol	0.82	0.87	-0.8243	0.29	37.8	0.33	39.0
1-Naphthol	0.61	0.66	-0.3137	0.29	36.1	0.33	33.1
2-Naphthol	0.61	0.66	-0.3943	0.29	35.1	0.33	32.9
Benzyl alcohol	0.33	0.51	0.0886	0.28	29.6	0.33	28.9
Thiophenol	0.09	0.04	-0.1546	0.15	22.9	0.13	13.3
Pyrrole	0.41	0.05	1.2628	0.18	17.6	0.20	12.3

Table 10.2. $\Sigma\beta^H_2$ and other parameters for hydrogen bonding bases; predicted $\Sigma\beta^H_2$ values are the values predicted by equation 10.27

Compound	Exptl. $\Sigma\beta^H_2$	Predicted $\Sigma\beta^H_2$	E_{HOMO}	Gasteiger		Abraham	
				Q_{MN}	ESP^-	Q_{MN}	ESP^-
Dichloromethane	0.05	0.05	-12.4853	-0.01	19.7	-0.123	-8.0
Trichloromethane	0.02	0.02	-12.9203	-0.01	24.5	-0.080	-3.5
1,2-Dichloroethane	0.11	0.11	-12.4154	-0.01	21.4	-0.169	-6.6
1,1,1-Trichloroethane	0.09	0.09	-12.7890	0.01	23.3	-0.077	-7.0
1-Chlorobutane	0.10	0.10	-12.0741	0.01	13.2	-0.178	-10.7
Tribromomethane	0.06	0.06	-11.8621	0.01	13.6	-0.060	-2.2
Diethyl ether	0.45	0.45	-10.9075	-0.39	-25.7	-0.259	-22.4
Di-n-propyl ether	0.45	0.45	-10.8158	-0.39	-21.7	-0.258	-22.1
Di-n-butyl ether	0.45	0.45	-10.9071	-0.39	-19.6	-0.258	-21.9
Propanone	0.49	0.49	-10.7521	-0.29	-22.6	-0.342	-29.2
Butanone	0.51	0.51	-10.6914	-0.29	-21.0	-0.341	-29.3
Cyclopentanone	0.52	0.52	-10.6080	-0.29	-19.4	-0.341	-29.1
Cyclohexanone	0.56	0.56	-10.5671	-0.29	-18.3	-0.341	-29.3
Methyl formate	0.38	0.38	-11.3684	-0.33	-28.6	-0.377	-26.4
Methyl acetate	0.45	0.45	-11.4593	-0.32	-12.4	-0.373	-22.5
Ethyl acetate	0.45	0.45	-11.4117	-0.32	-10.7	-0.373	-22.9
Vinyl acetate	0.43	0.43	-9.6663	-0.29	-12.8	-0.371	-22.6
Diethylamine	0.69	0.69	-11.1899	-0.34	-13.3	-0.435	-27.3
Methylamine	0.58	0.58	-10.5356	-0.38	-12.8	-0.630	-29.2
Ethylamine	0.61	0.61	-10.5329	-0.38	-12.6	-0.620	-30.1
n-Propylamine	0.61	0.61	-10.5281	-0.38	-9.4	-0.620	-29.9
n-Butylamine	0.61	0.61	-10.4560	-0.38	-8.4	-0.620	-28.8
Dimethylamine	0.66	0.66	-10.0480	-0.35	-15.6	-0.450	-26.1
Di-n-propylamine	0.69	0.69	-10.0099	-0.34	-9.6	-0.430	-26.9
Di-n-butylamine	0.69	0.69	-10.0228	-0.34	-7.5	-0.430	-26.8
Trimethylamine	0.67	0.67	-9.6139	-0.32	-18.0	-0.270	-21.0
Triethylamine	0.79	0.79	-9.5076	-0.31	-14.9	-0.260	-22.4
Formamide	0.60	0.60	-10.6950	-0.32	-22.7	-0.430	-26.0
Acetamide	0.68	0.68	-10.6075	-0.32	-22.5	-0.420	-27.3
Propionamide	0.68	0.68	-10.5986	-0.32	-22.2	-0.420	-27.5
N-Methylformamide	0.55	0.55	-10.3794	-0.29	-21.4	-0.350	-25.6
N-Methylpropionamide	0.71	0.71	-10.2654	-0.29	-21.0	-0.340	-27.7
N-Methylacetamide	0.72	0.72	-10.2713	-0.29	-21.4	-0.421	-27.6
N,N-Dimethylformamide	0.74	0.74	-10.1100	-0.28	-19.1	-0.420	-25.8
N,N-Dimethylacetamide	0.78	0.78	-10.0465	-0.28	-21.4	-0.421	-26.9
Acetic acid	0.44	0.44	-11.5714	-0.38	-13.0	-0.373	-21.3
Hexanoic acid	0.45	0.45	-11.4636	-0.38	-10.2	-0.373	-21.2
Chloroacetic acid	0.36	0.36	-11.8694	-0.38	-2.7	-0.360	-18.1
Dichloroacetic acid	0.27	0.27	-12.1320	-0.38	4.7	-0.359	-13.8
Trichloroacetic acid	0.28	0.28	-12.3578	-0.38	9.0	-0.352	-16.0
Formic acid	0.38	0.38	-11.7400	-0.38	-14.4	-0.380	-19.9

Table 10.2. Continued

Compound	Exptl. $\Sigma\beta^H_2$	Predicted $\Sigma\beta^H_2$	E_{HOMO}	Gasteiger		Abraham	
				Q_{MN}	ESP⁻	Q_{MN}	ESP⁻
Propanoic acid	0.45	0.45	-11.4934	-0.38	-13.0	-0.380	-21.4
Butanoic acid	0.45	0.45	-11.4789	-0.38	-12.5	-0.380	-21.3
2-Methylbenzoic acid	0.34	0.34	-9.6402	-0.37	-9.3	-0.330	-18.6
3-Methylbenzoic acid	0.38	0.38	-9.6530	-0.37	-10.8	-0.370	-18.1
4-Methylbenzoic acid	0.38	0.38	-9.7264	-0.37	-10.2	-0.370	-19.1
Methanol	0.47	0.47	-11.4146	-0.44	-20.8	-0.458	-25.8
Ethanol	0.48	0.48	-11.2964	-0.44	-21.9	-0.452	-26.6
Propan-1-ol	0.48	0.48	-11.2410	-0.44	-19.4	-0.452	-26.3
Propan-2-ol	0.56	0.56	-11.2053	-0.43	-16.5	-0.446	-27.2
Butan-1-ol	0.48	0.48	-11.2312	-0.44	-17.6	-0.452	-26.1
Hexan-1-ol	0.48	0.48	-11.2170	-0.44	-16.0	-0.452	-26.1
2,2,2-Trifluoroethanol	0.25	0.25	-12.3771	-0.43	34.2	-0.432	-16.2
Cyclopentanol	0.56	0.56	-11.1069	-0.43	-17.5	-0.450	-26.7
Cyclohexanol	0.57	0.57	-11.0846	-0.43	-16.1	-0.450	-26.7
Prop-2-en-1-ol	0.48	0.48	-10.3465	-0.43	-20.3	-0.450	-24.6
trans-But-2-en-1-ol	0.48	0.48	-9.9655	-0.43	-18.8	-0.450	-25.3
Ethylthiol	0.24	0.24	-9.7380	-0.21	-6.9	-0.240	-10.9
n-Propylthiol	0.24	0.24	-9.7303	-0.21	-8.5	-0.240	-10.8
n-Butylthiol	0.24	0.24	-9.7298	-0.21	-4.9	-0.240	-10.8
Diethyl sulphide	0.32	0.32	-9.5208	-0.16	-4.6	-0.226	-12.7
Di-n-butyl sulphide	0.32	0.32	-9.5116	-0.16	-0.6	-0.225	-12.6
Trimethyl phosphate	1.00	1.00	-11.2055	-0.29	-14.1	-0.337	-20.7
Triethyl phosphate	1.06	1.06	-11.1161	-0.29	-15.2	-0.337	-20.8
Tri-n-butyl phosphate	1.21	1.21	-11.0878	-0.29	-7.9	-0.337	-20.4
Benzene	0.14	0.14	-9.3906	-0.08	-11.6	-0.091	-20.8
Toluene	0.14	0.14	-9.2816	-0.07	-9.9	-0.086	-21.8
o-Xylene	0.16	0.16	-9.2296	-0.04	-8.3	-0.080	-22.9
m-Xylene	0.16	0.16	-9.2398	-0.04	-9.8	-0.080	-22.8
p-Xylene	0.16	0.16	-9.1832	-0.04	-8.3	-0.080	-22.8
1,3,5-Trimethylbenzene	0.19	0.19	-9.2348	-0.03	-11.2	-0.075	-23.8
Hexamethylbenzene	0.21	0.21	-9.0391	-0.04	-1.9	-0.060	-27.2
Naphthalene	0.20	0.20	-8.5714	-0.07	-10.3	-0.073	-20.3
Phenanthrene	0.26	0.26	-8.4901	-0.07	-9.0	-0.059	-19.7
Chlorobenzene	0.07	0.07	-9.6227	-0.05	3.5	-0.108	-13.7
Bromobenzene	0.09	0.09	-9.5502	-0.06	0.3	-0.088	-12.5
Benzaldehyde	0.39	0.39	-9.7265	-0.30	-20.0	-0.366	-26.0
Acetophenone	0.48	0.48	-9.6678	-0.29	-18.3	-0.361	-26.9
Benzophenone	0.50	0.50	-9.5863	-0.28	-12.8	-0.370	-25.2
Benzylamine	0.72	0.72	-9.4996	-0.37	-6.0	-0.609	-25.7
Acetanilide	0.67	0.67	-9.2254	-0.28	-20.5	-0.417	-29.6
Benzoic acid	0.40	0.40	-9.7684	-0.37	-11.0	-0.374	-18.9
Phenol	0.30	0.30	-8.8825	-0.40	-17.0	-0.384	-23.3

Table 10.2. Continued

Compound	Exptl. $\Sigma\beta^H_2$	Predicted $\Sigma\beta^H_2$	E_{HOMO}	Gasteiger		Abraham	
				Q_{MN}	ESP^-	Q_{MN}	ESP^-
2-Fluorophenol	0.26	0.26	-9.1463	-0.40	4.7	-0.420	-20.9
3-Fluorophenol	0.17	0.17	-9.2180	-0.40	-3.5	-0.384	-20.6
4-Fluorophenol	0.23	0.23	-9.0069	-0.40	0.8	-0.386	-20.3
2-Chlorophenol	0.31	0.31	-9.1616	-0.40	-8.3	-0.377	-22.9
3-Chlorophenol	0.15	0.15	-9.2224	-0.40	-9.2	-0.384	-20.5
4-Chlorophenol	0.20	0.20	-9.1452	-0.40	-30.4	-0.385	-23.0
2-Bromophenol	0.31	0.31	-9.0905	-0.40	-6.7	-0.420	-14.0
3-Bromophenol	0.16	0.16	-9.1513	-0.40	-11.7	-0.384	-20.9
4-Bromophenol	0.20	0.20	-9.0911	-0.40	-12.5	-0.385	-20.9
2-Methoxyphenol	0.52	0.52	-8.6399	-0.40	-18.3	-0.381	-23.8
3-Methoxyphenol	0.39	0.39	-8.6971	-0.40	-16.0	-0.384	-21.1
4-Methoxyphenol	0.48	0.48	-8.8307	-0.40	-18.9	-0.384	-21.4
2-Nitrophenol	0.37	0.37	-9.7503	-0.40	-13.1	-0.420	-20.0
3-Nitrophenol	0.23	0.23	-9.7321	-0.40	-9.5	-0.384	-23.9
4-Nitrophenol	0.26	0.26	-9.8473	-0.40	-11.9	-0.383	-23.0
1-Naphthol	0.37	0.37	-8.3128	-0.40	-12.0	-0.380	-23.9
2-Naphthol	0.40	0.40	-8.4863	-0.40	-14.9	-0.384	-22.7
Benzyl alcohol	0.56	0.56	-9.5195	-0.43	-17.9	-0.444	-24.3
Thiophenol	0.16	0.16	-9.6251	-0.17	-7.9	-0.204	-15.8
N,N-Dimethylbenzene-sulphonamide	0.86	0.86	-10.2126	-0.20	-18.2	-0.263	-45.6
Tetrahydrofuran	0.48	0.48	-10.7749	-0.39	-25.4	-0.258	-22.0
1,4-Dioxane	0.64	0.64	-10.5518	-0.38	-20.2	-0.252	-16.8
Pyrrole	0.29	0.29	-8.5689	-0.30	-13.1	-0.132	-23.7
Pyrazine	0.62	0.62	-10.0219	-0.27	-15.5	-0.281	-16.0
Pyrimidine	0.65	0.65	-10.3760	-0.25	-12.0	-0.327	-16.0
Thiazole	0.45	0.45	-9.8840	-0.26	-14.7	-0.352	-10.3

In Table 10.3 the highest and the lowest electrostatic potentials on the surface of 2-nitrophenol, when different dot densities are used, are given. Clearly, ESP^+s do not increase (or ESP^-s do not decrease) significantly with increasing dot density from 10 to 90 dot/$Å^2$.

Table 10.3. ESP^+ and ESP^- for 2-nitrophenol in different dot densities

Dot density	ESP^+	ESP^-
5.0	44.540	-36.263
7.0	46.047	-36.483
8.0	45.392	-35.965
9.0	45.392	-36.607
10.0	45.392	-36.598
15.0	45.392	-36.040
18.0	45.958	-36.555
20.0	45.981	-36.639
30.0	45.601	-36.643
40.0	45.965	-36.603
50.0	46.025	-36.643
90.0	46.095	-36.598

10.2.1. Hydrogen bonding acidity

For H-bond donors, the relationships of H-bond acidity with ESP^+ originating from both Gasteiger charges and Abraham charges were analysed. The energy of the lowest unoccupied molecular orbital (E_{LUMO}) was also used to describe the charge transfer contribution to the hydrogen bonding energy. The results of statistical analyses have been tabulated in Table 10.4 and Table 10.5. In Table 10.4 the correlations have been

reported for all the hydrogen bond donors including 2-substituted phenols, which have always been outliers from relationships between $\Sigma\alpha^H_2$ and atomic charge. The correlations in Table 10.5 do not include 2-substituted phenols. Comparing these two tables it is seen that all the correlations improved after deletion of 2-substituted phenols, with correlations containing atomic charge as the electrostatic descriptor showing the highest improvement.

The Abraham method calculates a Q_H value for pyrrole which, in comparison with other secondary amines, is lower than expected; because of the conjugation in its structure which puts negative charge in carbon atoms of the ring and positive charge on the nitrogen, it would be expected to be higher. Consequently the ESP^+ calculated from Abraham charges is unrealistically low and this compound has been excluded from correlations containing parameters calculated by the Abraham method. ESP^+ calculated by the Gasteiger method for 2,2,2-trifluoroethanol is excessively high and has been excluded from equations containing ESP^+ calculated from Gasteiger charges.

Table 10.4. Correlations between $\Sigma\alpha^H_2$ and theoretical parameters for all hydrogen bond donors

$$\Sigma\alpha^H_2 = A\ (Q_H \text{ or } ESP^+) + B\ (E_{LUMO}) + C$$

Method	Parameter	Eq	A	B	C	n	r	s	F
Abraham	Q_H	(10.1)	2.06	-	-0.164	62	0.513	0.212	21
	Q_H, E_{LUMO}	(10.2)	1.57	-0.0811	0.070	62	0.711	0.175	30
	ESP^+	(10.3)	0.0338	-	-0.490	62	0.857	0.127	166
	ESP^+, E_{LUMO}	(10.4)	0.0290	-0.0504	-0.301	62	0.905	0.106	133
Gasteiger	Q_H	(10.5)	2.70	-	-0.223	63	0.657	0.184	46
	Q_H, E_{LUMO}	(10.6)	1.97	-0.0596	0.023	63	0.734	0.167	35
	ESP^+	(10.7)	0.0198	-	-0.168	62	0.769	0.157	87
	ESP^+, E_{LUMO}	(10.8)	0.0162	-0.0426	0.072	62	0.802	0.148	53

Table 10.5. Correlations between $\Sigma\alpha^H_2$ and theoretical parameters for hydrogen bond donors excluding 2-substituted phenols

$$\Sigma\alpha^H_2 = A\ (Q_H \text{ or } ESP^+) + B\ (E_{LUMO}) + C$$

Method	Parameter	Eq	A	B	C	n	r	s	F
Abraham	Q_H	(10.9)	2.19	-	-0.180	57	0.571	0.203	26
	Q_H, E_{LUMO}	(10.10)	1.71	-0.0982	0.078	57	0.825	0.140	59
	ESP^+	(10.11)	0.0362	-	-0.564	57	0.885	0.115	200
	ESP^+, E_{LUMO}	(10.12)	0.0292	-0.0512	-0.305	57	0.926	0.095	162
Gasteiger	Q_H	(10.13)	2.99	-	-0.271	58	0.737	0.166	66
	Q_H, E_{LUMO}	(10.14)	2.14	-0.0747	0.025	58	0.847	0.132	70
	ESP^+	(10.15)	0.0196	-	-0.159	57	0.782	0.154	87
	ESP^+, E_{LUMO}	(10.16)	0.0141	-0.0582	0.085	57	0.834	0.138	62

It is clear that using ESP^+ is far more satisfactory than using atomic charge in the quantification of the electrostatic energy of the total H-bond energy. Although the atomic charges calculated by the Gasteiger method are better predictors of H-bonding strength than are those calculated by the Abraham method, the resulting ESPs show the reverse order. Figure 10.2 shows the plot of $\Sigma\alpha^H_2$ against the predicted $\Sigma\alpha^H_2$ values from equation 10.4, in which the independent variables are ESP^+ and E_{LUMO}.

To examine in more detail how the use of ESP^+ can model hydrogen bond donor ability, we considered different classes of compounds separately.

Alcohols There are 12 alcohols in the H-bond donors and both of the charge calculation methods yield identical atomic charge values on the hydroxyl hydrogen. This leads to the following relationship of calculated ESP^+ (Abraham method) with $\Sigma\alpha^H_2$:

$$\Sigma\alpha^{H}_{2} = 0.0411 ESP^{+} - 0.822 \qquad (10.17)$$

$n = 12 \quad r = 0.887 \quad s = 0.0330 \quad F = 37$

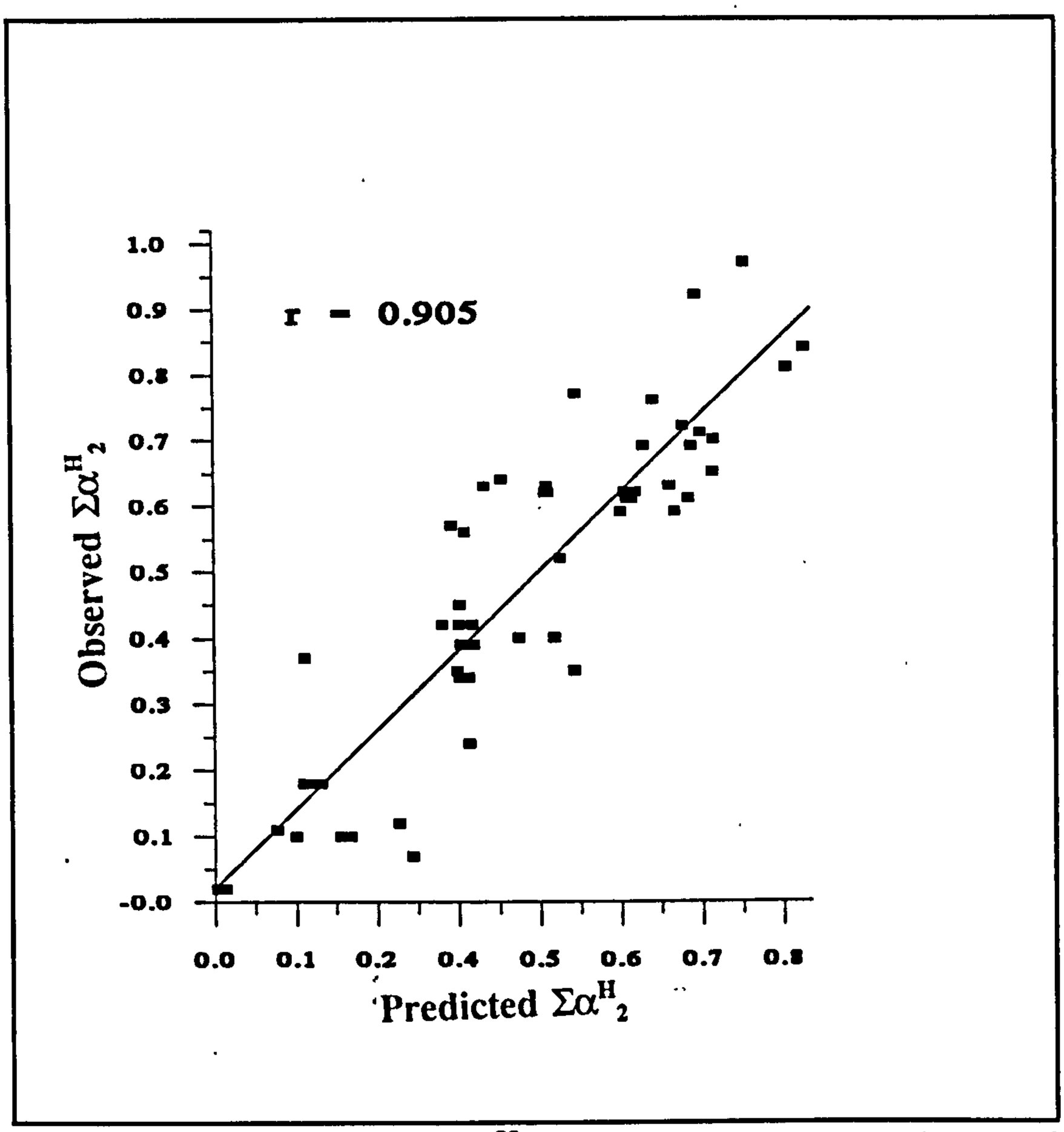


Figure 10.2. The graph between $\Sigma\alpha^{H}_{2}$ and the predicted values by equation 10.4.

In equation 10.17, prop-2-en-1-ol, with an ESP^{+} value of 27.5, is an outlier. Considering the atomic charge of 0.33 on the hydroxyl hydrogen of all the alcohols, the ESP^{+} value for prop-2-en-1-ol is unreasonably low. This could be related to the presence of a π-bond in the molecule, towards which, in the conformation used, the hydroxyl hydrogen is directed (Figure 10.3). The negative zone resulting from this double bond orbital can interfere with the interaction between the proton probe and the hydroxyl hydrogen, and lower the interaction energy in that area. A proton acceptor

approaching this alcohol will be confronted with the situation in which electron density of the double bond will reduce the attraction of the hydroxyl hydrogen; however, in the calculation of ESP^+, instead of a dipolar group a completely charged particle (a proton) is approaching the molecule and therefore the effect of the π-bond electron density is intensified.

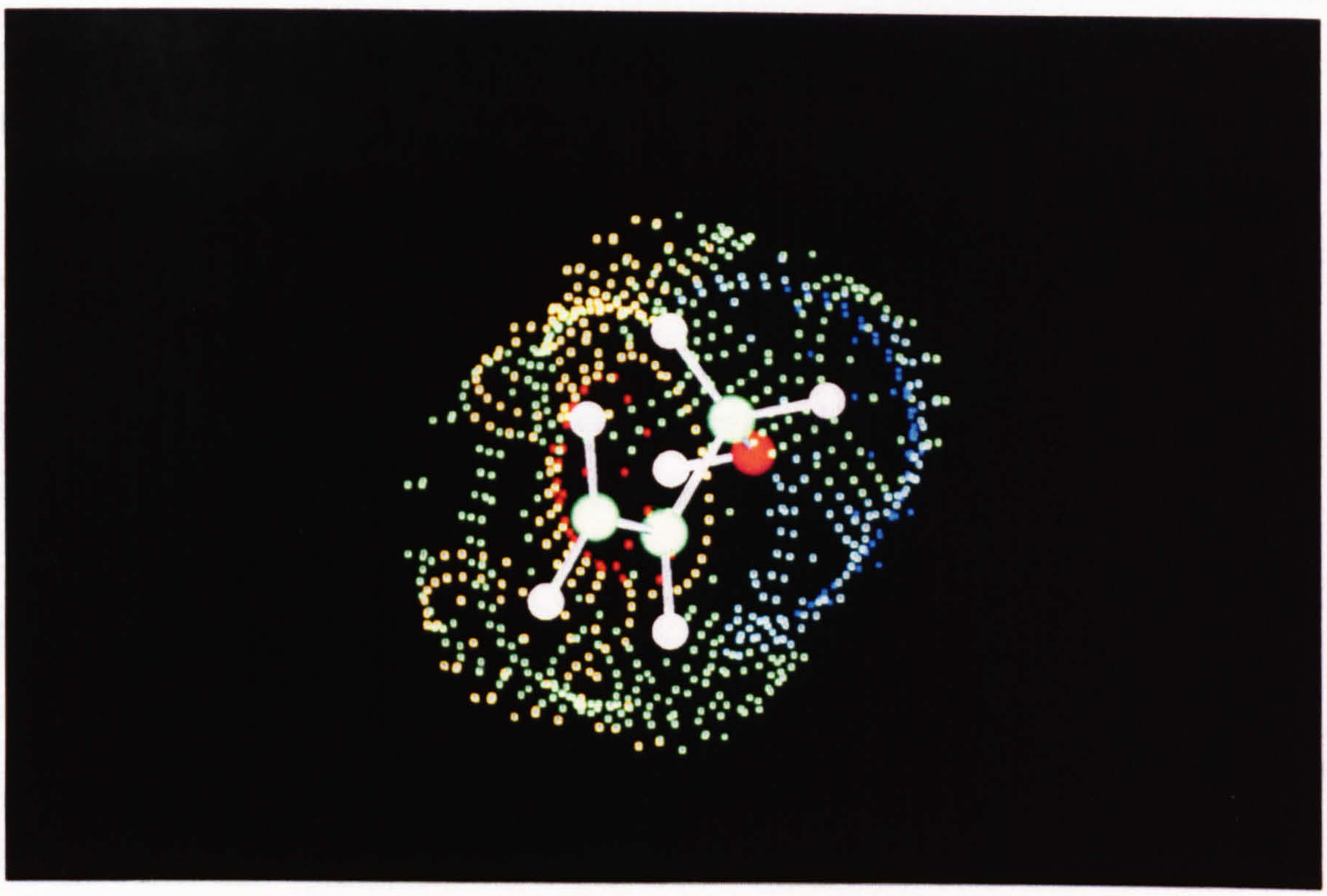


Figure 10.3.The conformation of prop-2-en-1-ol which has been used in equation 10.17.

Unfortunately, the Cobra conformation analysis program does not carry out energy minimisation, but can calculate only the molecular mechanics energy for each conformation and sort the conformations according to the energy. This study has found that it often happens that several of the conformations suggested by Cobra have energies approaching that of the minimum energy conformation after minimisation by the COSMIC force field. Five other conformations of this alcohol obtained from Cobra program (c2, c3, c4, c5, c6) were analysed; after atomic charge calculation by the

Abraham method and energy minimisation by COSMIC, the final conformations had total energies of 2.873, 2.889, 3.293, 2.870 and 3.293 kcal/mol respectively. The ESP^+ values for these conformations after recalculating the atomic charges were 28.6, 28.4, 30.6, 28.6 and 30.6 kcal/mol respectively. The conformation of prop-2-en-1-ol used in the regression analysis (c1) had a COSMIC energy of 2.870 kcal/mol. This is about the same amount of energy as the conformers c2, c3 and c5 have and therefore each conformation is likely to exist in the solution in which the experimental $\Sigma\alpha^H_2$ has been measured; hence the ESP^+ of any of the conformations c2, c3 and c5 (28.6, 28.4, 28.6) can be used in the regression analyses.

After deletion of this alcohol there is a better relationship between $\Sigma\alpha^H_2$ and ESP^+ (equation 10.18) and even if 2,2,2-trifluoroethanol, which has much higher H-bond acidity than the other alcohols and therefore exerts a large influence on the equation, is excluded, the correlation still has good statistics (equation 10.19).

$$\Sigma\alpha^H_2 = 0.0475ESP^+ - 1.02 \qquad (10.18)$$

$n = 11 \quad r = 0.956 \quad s = 0.0222 \quad F = 95$

$$\Sigma\alpha^H_2 = 0.101ESP^+ - 2.57 \qquad (10.19)$$

$n = 10 \quad r = 0.910 \quad s = 0.0152 \quad F = 39$

The correlation including prop-2-en-1-ol (c2 or c5 conformation) is:

$$\Sigma\alpha^H_2 = 0.0844ESP^+ - 2.08 \qquad (10.20)$$

$n = 11 \quad r = 0.785 \quad s = 0.0220 \quad F = 14.5$

ESP^+ calculated from Gasteiger charges has no relationship with $\Sigma\alpha^H_2$. There is no correlation between $\Sigma\alpha^H_2$ and Q_H calculated by Abraham's or Gasteiger's method as

atomic charges on the hydroxyl hydrogen are the same for all the alcohols.

Phenols In our previous studies where we used atomic point charges to quantify the electrostatic contribution to hydrogen bonding, 2-substituted phenols have always been outliers and therefore excluded from the equations. The possibility in these compounds of intramolecular hydrogen bonding and steric hindrance, which are not reflected in the atomic charges, is probably responsible for this observation. Here, using ESP as the electrostatic descriptor seems to be much more efficient. The reason is that electrostatic potential at each point in space around the molecule reflects the effects of all charges present in the molecule, and also it is affected by the steric situation of the heteroatom. On the other hand, ESP^+ and ESP^- are greatly influenced by conformation, as was the case in the alcohols. 2-Substituted phenols, as expected, were extremely sensitive to the conformation used. Table 10.6 shows the energies of different conformations of some 2-substituted phenols calculated using the COSMIC force field and also the ESP^+ and ESP^- values calculated from their Abraham charges. In this table the first conformation for each phenol is the one selected by the Cobra conformation analysis program as the lowest energy conformation. Table 10.6 shows that clearly this is not always the case, and after minimisation, the second (or the third) conformation has lower energy than has the first, for some of the phenols.

For all the substituted phenols (with the lowest energy conformations after COSMIC energy minimisation) the following equation shows the correlation between the experimental H-bonding acidity and ESP^+ calculated from Abraham charges:

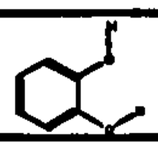
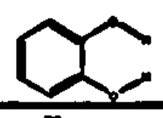
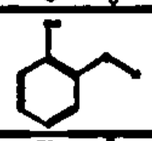
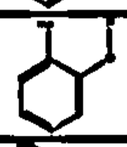
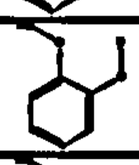
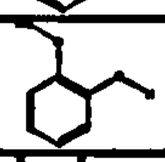
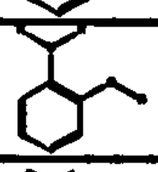
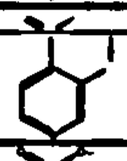
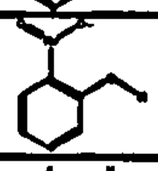
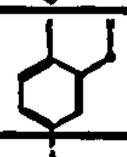
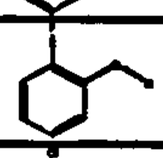
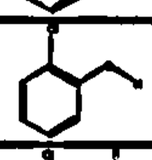
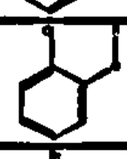
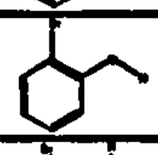
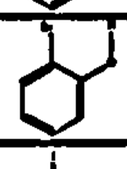
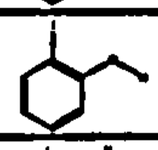
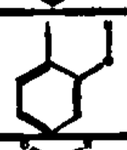
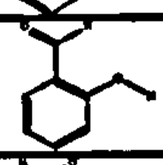
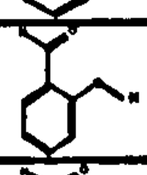
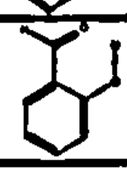
$$\Sigma\alpha^H_2 = 0.0249ESP^+ - 0.200 \qquad (10.21)$$

$n = 18 \quad r = 0.853 \quad s = 0.1082 \quad F = 42.7$

This correlation is in fact as good as that for 3- and 4-substituted phenols alone. One

particular anomaly was observed: the ESP^+ value for 4-chlorophenol was found to be 32.0, which in comparison with 3-chlorophenol and also other 3- and 4-halogenated phenols (ESP^+ = 34.0 - 35.5) seems to be unrealistic.

Table 10.6. Conformational energies of 2-substituted phenols

2-Substituted phenols	Conformation	Cosmic energy (kcal/mol)	ESP^+ (kcal/mol)	ESP^- (kcal/mol)
o-Cresol		2.510	37.3	-25.3
		6.967	31.3	-42.0
2-Methylphenol		-1.335	31.9	-24.0
		-1.152	30.0	-23.5
2-Methoxyphenol		6.495	23.7	-23.8
		10.177	32.6	-37.8
2-Aminophenol		-2.793	31.7	-30.1
2-Nitrophenol		1.651	18.8	-20.0
		6.572	38.4	-31.4
2-Fluorophenol		7.085	23.8	-20.9
		9.762	35.1	-31.5
2-Chlorophenol		6.044	34.9	-22.9
		3.230	22.6	-21.1
2-Bromophenol		5.146	34.6	-14.0
		2.528	14.8	-21.4
2-Iodophenol		4.208	33.3	-17.0
		3.511	16.8	-21.8
2-Hydroxybenz-		0.398	38.5	-27.6
aldehyde		5.159	31.3	-41.4
		-2.733	32.6	-24.8

As with the alcohols, there is no correlation between $\Sigma\alpha^{H}_{2}$ and ESP^{+} calculated from Gasteiger charges or Q_H calculated by any of the methods. The reason for the latter is that the value of the atomic charge on the hydroxyl hydrogen is the same for all the substituted phenols.

Carboxylic acids The point charge on the hydroxyl hydrogen of all the acids was constant using both methods, but because of the different steric situations, different ESP^{+} values were obtained for different acids, which could predict the H-bonding ability of the acids reasonably well. Equations 10.22 and 10.23 are the correlations with ESP^{+} calculated from Abraham and Gasteiger charges respectively.

$$\Sigma\alpha^{H}_{2} = 0.0529ESP^{+} - 0.961 \qquad (10.22)$$

$n = 12 \quad r = 0.866 \quad s = 0.0676 \quad F = 30$

$$\Sigma\alpha^{H}_{2} = 0.0128ESP^{+} + 0.236 \qquad (10.23)$$

$n = 12 \quad r = 0.938 \quad s = 0.0471 \quad F = 72$

Nitrogen acids The atomic charge on the hydrogen connected to the nitrogen for all the amines is almost constant using the Abraham method. The Gasteiger method gives two different point charge values, with primary amines having a lower charge than secondary amines, which is the opposite of the order expected from the inductive effect of alkyl groups in solution. In the case of the amides, again the Abraham method puts the same amount of the charge on the hydrogen connected to the nitrogen for all the compounds. The Gasteiger method, as for amines, calculates different Q_H values for amides depending on whether or not they are substituted; these atomic charges clearly cannot have any relationship with $\Sigma\alpha^{H}_{2}$ because of the erroneous order of the atomic charge values. Electrostatic potentials calculated from the better charge

calculation method of Abraham do not show a good relationship with $\Sigma\alpha^{H}_{2}$ of amines or amides (r = 0.434 and 0.490 for amines and amides respectively). This could be explained by the fact that the nitrogen atom is connected to two other atoms or groups while the oxygen atom is connected to only one group. This makes the ESP^{+} of nitrogen acids more sensitive to the conformation of the molecule.

10.2.2. Hydrogen bonding basicity

The H-bond acceptor abilities ($\Sigma\beta^{H}_{2}$) of the compounds listed in Table 10.2 show no correlation with Abraham or Gasteiger charges, or the electrostatic potentials calculated from them. The plot between $\Sigma\beta^{H}_{2}$ and E_{HOMO} calculated by MNDO method in chapter 9 (Fig 9.9) showed that there were two groups of compounds. In the first group (non-aromatic structures), E_{HOMO} values were lower than about -10.5 eV, and $\Sigma\beta^{H}_{2}$ increased with increasing E_{HOMO}. The second group comprised the structures for which E_{HOMO} values were higher than about -10.5 eV and the plot of $\Sigma\beta^{H}_{2}$ against E_{HOMO} was rather scattered (compounds with aromatic structures). The compounds were therefore divided into two groups, and Table 10.7 shows the results obtained for the first group (non-resonance):

Phosphates were outliers, and have been excluded from the correlations in Table 10.7.

Clearly, in both the Abraham and Gasteiger methods, ESP^{-}s are much better than are atomic charges for prediction of hydrogen bonding basicity. ESP^{-} values calculated from Abraham atomic charges are better predictors of H-bonding basicity than are those calculated from Gasteiger atomic charges.

Table 10.7. Correlations between $\Sigma\beta^H_2$ and theoretical parameters for compounds without resonance

$$\Sigma\beta^H_2 = A\ (Q_{MN} \text{ or } ESP^-) + B\ (E_{HOMO}) + C$$

Method	Parameter	Eq	A	B	C	n	r	s	F
Abraham	Q_{MN}	(10.24)	-1.0200	-	0.103	60	0.657	0.146	44
	Q_{MN}, E_{HOMO}	(10.25)	-0.8180	0.1070	1.340	60	0.804	0.116	52
	ESP^-	(10.26)	-0.0225	-	-0.020	60	0.855	0.101	158
	ESP^-, E_{HOMO}	(10.27)	-0.0191	0.0717	0.837	60	0.904	0.083	128
Gasteiger	Q_{MN}	(10.28)	-0.9690	-	0.166	60	0.621	0.156	35
	Q_{MN}, E_{HOMO}	(10.29)	-0.8120	0.1170	1.500	60	0.802	0.120	49
	ESP-	(10.30)	-0.0104	-	0.357	60	0.751	0.131	71
	ESP^-, E_{HOMO}	(10.31)	-0.0082	0.0632	1.070	60	0.785	0.124	43

For the second group of bases there was no correlation with electrostatic potentials or atomic charges calculated by either method. Correlations were therefore sought within chemical classes. For phenols, there were no correlations between $\Sigma\beta^H_2$ and ESP^- or Q_{MN}. This is not very surprising, because in the plot between $\Sigma\beta^H_2$ and E_{HOMO} phenols are in the group of compounds (resonance structures) for which no correlation was observed.

Very good correlations were found with ESP^-s resulting from the two charge calculation methods for carboxylic acids and esters, from which methyl formate is an outlier. The atomic charge on the most negative heteroatom (carbonyl oxygen) of this ester is unrealistically high, by both charge calculation methods, since it would be expected to be lower than those for the remainder (for example methyl acetate). This could explain the excessive ESP^-s for this ester. The relationships after excluding methyl formate are given by equations 10.32 and 10.33 for Abraham and Gasteiger calculated ESP^-s respectively:

$$\Sigma\beta^{H}_{2} = -0.0224ESP^{-} - 0.0473 \quad (10.32)$$

$n = 15 \quad r = 0.943 \quad s = 0.021 \quad F = 105$

$$\Sigma\beta^{H}_{2} = 0.330 - 0.00746ESP^{-} \quad (10.33)$$

$n = 15 \quad r = 0.840 \quad s = 0.034 \quad F = 31$

There is no correlation with atomic charges calculated by Gasteiger method and the correlation with Abraham charges is poor (equation 10.34).

$$\Sigma\beta^{H}_{2} = -3.13Q_{MN} - 0.756 \quad (10.34)$$

$n = 15 \quad r = 0.675 \quad s = 0.0467 \quad F = 11$

If ketones are included in these correlations, although there are no correlations with Q_{MN}, the relationships with ESP^{-} are found to be:

$$\Sigma\beta^{H}_{2} = 0.0855 - 0.0154ESP^{-} \quad (10.35)$$

$n = 21 \quad r = 0.931 \quad s = 0.028 \quad F = 123$

$$\Sigma\beta^{H}_{2} = 0.328 - 0.0086ESP^{-} \quad (10.36)$$

$n = 21 \quad r = 0.867 \quad s = 0.039 \quad F = 57,$

for ESP^{-} calculated from Abraham and Gasteiger methods, respectively.

Benzaldehyde could not be incorporated in the above equation because the charge on the carbonyl oxygen was unrealistically high; this is because the carbonyl group is similar to that of methyl formate (carbonyl group connected to a hydrogen), and as with methyl formate, its most negative atomic charge is unexpectedly higher than that of acetophenone.

Ethers were analysed together with alcohols. The parameters calculated by the Gasteiger method had no correlation with $\Sigma\beta^{H}_{2}$. In order to find a correlation with parameters calculated by the Abraham method, dioxane and 2,2,2-trifluoroethanol had to be excluded from the regression. Dioxane was an outlier because it has two symmetrical oxygens capable of forming hydrogen bonds, while we had used only the negative ESP value near to one of the oxygens. 2,2,2-trifluoroethanol, because of the electron attraction of the three fluorine atoms, is a much weaker H-bond acceptor and thus is quite different from the rest of the compounds, exerting a large influence on the regression equation. The equation after deleting these two compounds is:

$$\Sigma\beta^{H}_{2} = 0.146 - 0.0138ESP^{-} \qquad (10.37)$$

$$n = 15 \quad r = 0.669 \quad s = 0.032 \quad F = 9.7$$

This equation clearly is not as good as the equation for H-bond acceptors containing sp^2 oxygen (compounds with carbonyl groups). The reason again relates to the dependence of the ESPs on the conformation of the compounds; because sp^3 oxygens are connected to more groups than are sp^2 oxygens, their ESPs are affected by different conformations more than are those of sp^2 containing structures. If 2,2,2-trifluoroethanol and also dioxane (after doubling the ESP^- value which represents the H-bond acceptor ability of one of the oxygens) are included, the relationship is:

$$\Sigma\beta^{H}_{2} = -0.0085 - 0.0200ESP^{-} \qquad (10.38)$$

$$n = 17 \quad r = 0.871 \quad s = 0.042 \quad F = 47$$

The regression between $\Sigma\beta^{H}_{2}$ and Q_{MN} for ethers and alcohols has a correlation coefficient of 0.514.

For amines there was no correlation between $\Sigma\beta^{H}_{2}$ and ESP^- calculated by either

method.

10.2.3. Replacing H-bonding parameter with ESPs in QSAR equations

1- In the QSAR for upper respiratory tract irritation by airborne chemicals in mice (Abraham et al, 1990), the H-bonding parameter (α^H_2) can be replaced by ESP^+:

$$-\log FRD'_{50} = -0.69 + 0.77\delta_2 + 2.81\pi^*_2 + 4.93\,\alpha^H_2 + 2.82V_x \qquad (10.39)$$

$n = 39 \quad s = 0.136 \quad r = 0.985$

$$-\log FRD'_{50} = -0.59 + 1.23\delta_2 + 0.44\pi^*_2 + 0.0874\,ESP^+ + 2.74V_x \qquad (10.40)$$

$n = 39 \quad s = 0.148 \quad r = 0.981$

The results of ESP^+ calculations are listed in Table 10.8 for these chemicals.

2- The QSAR equation for Microtox toxicity of some non-reactive toxicants (Kamlet et al, 1986) is:

$$\log EC_{50} = 7.61 - 4.11\,V/100 - 1.54\,\pi^* + 3.94\,\beta - 1.51\,\alpha_m \qquad (10.41)$$

$n = 38 \quad s = 0.28 \quad r = 0.987$

From this equation four toxicants are excluded (Table 10.9). In case of deletion of the same compounds from correlation analysis, replacement of H-bond parameters in this equation (α_m and β) by ESPs and E_{LUMO}, results in the following equation:

$$\log EC_{50} = 7.21 - 4.12\,V/100 - 0.054\,ESP^- - 0.041\,ESP^+ + 0.292\,E_{LUMO} \qquad (10.42)$$

$n = 38 \quad s = 0.46 \quad r = 0.964$

π^* was not statistically significant and has not been used in equation 10.42.

On the other hand, when deleting the outliers of correlation with the theoretical parameters, the following equation is obtained:

$$\log EC_{50} = 7.36 - 3.99\ V/100 - 0.076\ ESP^{-} - 0.064\ ESP^{+} + 0.203\ E_{LUMO} \quad (10.43)$$

$n = 38 \quad s = 0.41 \quad r = 0.970$

Clearly the replacement of experimental H-bonding parameters with ESPs in these QSARs has resulted in the equations of comparable statistics. These results show the ability of elctrostatic potentials to serve as easily obtainable H-bonding descriptors in the QSAR equations.

Table 10.8. ESP^+ values for airborn chemicals used in equation 10.40

Toxicant	ESP^+
2-Propanone	18.9
But-1-ene-3-One	21.9
2-Butanone	19.3
2-Pentanone	19.4
Mesityl oxide	21.8
Cyclohexanone	16.0
2-Hexanone	18.8
4-Methyl-2-pentanone	18.7
3,3-Dimethyl-2-butanone	17.2
2-Heptanone	18.6
4-Heptanone	15.5
5-Methyl-2-hexanone	19.5
2-Octanone	19.2
5-Methyl-3-heptanone	15.1
5-nonanone	17.8
2,6-Dimethyl-4-heptanone	14.6
2-Undecanone	19.5
Methanol	29.5
Ethanol	29.1
1-Propanol	28.9
2-Propanol	28.6
1-Butanol	29.2
2-Methyl-1-propanol	28.8
1-Pentanol	29.3
3-Methyl-1-butanol	29.1
1-Hexanol	29.1
4-Methyl-2-pentanol	28.1
1-Heptanol	29.3
1-Octanol	29.3
2-Ethyl-1-hexanol	29.0
Prop-2-en-1-ol	27.5
But-2-en-1-ol	29.0
Toluene	9.3
Phenol	32.3
Chlorobenzene	12.7
Bromobenzene	12.2
1,2-Dichlorobenzene	15.5
2-Chlorotoluene	12.2
Acetophenone	21.7
2-Xylene	9.1
4-Xylene	8.8
β-Chloroethylbenzene	13.4
Styrene	10.9
Ethylbenzene	9.4
α-Methylstyrene	9.8
4-Vinyltoluene	10.7
4-Divinylbenzene	11.2

Table 10.9. E_{HOMO} and E_{LUMO} (calculated by MNDO method) and ESPs for toxicants used in correlations of Microtox test results

Toxicant	ESP^+	ESP^-	E_{HOMO}	E_{LUMO}
Methanol	29.5	-25.8	-11.415	3.795
Ethanol	29.1	-26.6	-11.296	3.515
1-Propanol	28.9	-26.3	-11.241	3.253
2-Propanol	28.6	-27.2	-11.205	3.338
1-Butanol	29.1	-26.1	-11.231	3.194
2-Methyl-1-propanol	28.8	-26.4	-11.179	3.134
3-Pentanol	28.2	-27.5	-11.097	3.082
1-Hexanol	29.1	-26.1	-11.217	3.150
1-Heptanol	29.3	-26.0	-11.213	3.114
1-Octanol	29.3	-26.0	-11.211	3.086
2-Decanol	28.2	-27.3	-11.126	3.018
Acetone	18.9	-29.2	-10.752	0.660
2-Butanone	19.3	-29.3	-10.691	0.688
4-Methyl-2-pentanone	18.7	-29.2	-10.673	0.691
2-Octanone	19.2	-29.1	-10.668	0.693
Ethyl acetate	15.9	-22.9	-11.412	0.942
Ethyl propionate	15.0	-28.4	-11.343	0.986
Diethyl ether	9.7	-22.4	-10.908	3.255
Di-*n*-butyl ether	9.8	-21.9	-10.907	3.026
Dimethylformamide	25.2	-25.8	-9.618	1.512
1,1,1-Trichloroethane	18.4	-7.0	-12.789	-0.567
1,1,2-Trichloroethene	22.7	-5.9	-10.644	-0.537
1,2-Dichloroethane	17.4	-6.6	-12.415	-0.080
1,1,2,2-Tetrachloroethane	20.9	-2.8	-12.737	-0.871
Benzene	9.5	-20.8	-9.391	0.367
Toluene	9.3	-21.8	-9.282	0.249
o-Xylene	9.1	-22.9	-9.230	0.191
Chlorobenzene	12.7	-13.8	-9.623	-0.131
1,3-Dichlorobenzene	17.1	-5.5	-9.895	-0.527
1,2,3-Trichlorobenzene	20.1	-8.7	-10.103	-0.815
3,4-Dichlorotoluene	16.2	-10.5	-9.747	-0.582
Phenol	32.3	-23.3	-8.883	0.251
o-Cresol	37.3	-25.3	-8.666	0.152
4-*t*-Butylphenol	34.3	-19.8	-8.813	0.207
2,4-Dimethylphenol	34.4	-21.1	-8.788	0.159
4-Nitrophenol	39.0	-23.0	-9.847	-0.824
Pyridine	13.1	-18.7	-9.687	0.005
6-Methyl-5-hepten-2-one	19.2	-28.9	-9.819	0.595
Cyclohexanol	28.7	-26.7	-11.085	3.066
Cyclohexanone	16.0	-29.3	-10.567	0.716
5-Methyl-2-hexanone	19.5	-29.3	-10.683	0.682
2-Decanone	18.9	-29.2	-10.670	0.696

10.3. Conclusions

The results of the regression analyses showed that electrostatic potentials calculated by both Abraham and Gasteiger method are better than their original charges for prediction of $\Sigma\alpha^{H}_{2}$ and $\Sigma\beta^{H}_{2}$ values. On the other hand, electrostatic potentials are dramatically dependent on conformation used and therefore care must be taken concerning the choice of conformation. The lowest energy conformation gives a reliable representation of what happens in reality, leading to good correlations between ESP values and $\Sigma\alpha^{H}_{2}$ and $\Sigma\beta^{H}_{2}$ values.

The methods used to calculate atomic charges do not yield accurate values; for instance atomic charges on the hydroxyl group of the phenols are the same for different substituted phenols. If better charge calculation methods such as the MNDO semiempirical method are used, this may lead to superior ESPs. Unfortunately, there was no way of enterng MO calculated charges as the input file to the program MAD.

11. Other electrostatic parameters

11.1. Dipole moments

Kamlet at al (1982) found that formation constants of hydrogen bonded complexes of a series of H-bond acceptors with the H-bond donors diphenylamine, 4-bromoaniline, and 5-fluoroindole in CCl_4 solvent, chloroform in cyclohexane solvent, and tri-*n*-butylammonium ion in *o*-dichlorobenzene solvent are correlated with the dipolarity/polarisability parameter π^*, and consequently, with the dipole moments μ, suggesting the importance of contributions from dipole/dipole interactions. They also explained the family-dependent behaviour observed in the correlations between log K_f values for complexes of 5-fluoroindole with H-bond acceptors in CCl_4 and pK_{HB} (H-bond acceptor parameter) (Mitsky et al, 1972) by the dipole/dipole contribution to the free energies of formation.

It was seen in the correlations of anilines (chapter 7) that some 4-substituted anilines (e.g. 4-aminoaniline) were outliers from equations and it was supposed that the dipole moment of compounds (μ) might play a role in their H-bond donor ability. It would also be interesting if dipole moment could help to produce a general correlation with H-bond abilities (rather than separate equations for different families of compounds).

Therefore, dipole moments were calculated by CNDO method (the method which had been used to calculate atomic charges in anilines), and regression analysis performed.

Two examples are given here:

1) 3- and 4-substituted anilines (the dipole moments are listed in Table 7.1):

$$\alpha^H_2 = 0.0609\ \mu + 0.218 \qquad (11.1)$$

n = 29 s = 0.191 r = 0.397 F = 0.505

2) For H-bonding acids consisting of different types of acids (table 7.2), there was no correlation ($r^2 = 0.003$).

When using dipole moments in conjunction with charge parameters, the correlation coefficient of the multiple regression increased a little, but the F statistic dropped and the maximum t-ratio for the coefficient of dipole moment was 2.36 in the case of aniline derivatives.

After separating different families of H-bond acids in table 7.2, for amines the following equations resulted:

$$\alpha^H_2 = 4.55\ Q_H - 0.005 \qquad (11.2)$$

n = 27 s = 0.065 r = 0.758 F = 35

$$\alpha^H_2 = 0.0303\ \mu + 0.298 \qquad (11.3)$$

n = 27 s = 0.082 r = 0.555 F = 11.6

$$\alpha^H_2 = 0.0142\ \mu + 3.83\ Q_H + 0.0122 \qquad (11.4)$$

$n = 27 \quad s = 0.062 \quad r = 0.792 \quad F = 21$

In equation 10.4, t-ratios are 4.62 ($p = 0.000$) and 1.88 ($p = 0.072$) for Q_H and μ respectively. It is clear from these examples that dipole moment does not correlate with H-bonding abilities. This conclusion can also be made from the observation that amines are better H-bond bases than nitriles, despite their smaller dipole moments. The reason is that μ is only the very crudest representation of the charge distribution in a molecule. Group or bond moments might give better correlation; Allen (1975) suggested the ionisation potential of the electron donor and the bond dipole of the proton donor as the key features of the hydrogen bond.

11.2. Electrotopological state index

Recently Hall and Kier (Hall et al, 1991) introduced a novel approach to the representation of molecular structure information based on an atomic level index derived from chemical graph theory (Kier & Hall, 1986). This index combines both electronic character and the topological environment of each skeletal atom in a molecule and thereby defines the electrotopological state of each atom. The electrotopological state index (E-state) value for a skeletal atom encodes information about the electronegativity, π and lone pair electron content, topological status and the environment of an atom within a molecule; thus the E-state may also be considered a measure of atomic electronic accessibility (Hall et al, 1993). In general, atoms with

higher E-state values are rich in π and lone-pair electrons and are of mantle topology in the molecule. As a consequence, the interactions involving these atoms might be quite strong, e.g. electrostatic or hydrogen bonding. With E-state values of an intermediate range, it is expected that dipolar forces are implied or the atom is partially buried in the molecule and therefore sterically less accessible to interactions across space. Lower E-state values correspond to the dominant propensity for dispersion interactions (Kier & Hall, 1992).

In a QSAR study of inhibition of MAO by hydrazides, it was shown (Hall et al, 1993) that the QSAR model based on E-state indexes is significantly superior to the one based on molecular orbital parameters calculated by AM1 method; the MO parameters studied were atomic partial charge for each atom, dipole moment, ΔH_f and ionisation potential.

Therefore, correlations of E-state indices with an experimental H-bond acceptor parameter ($\Sigma\beta^H_2$) were examined. 119 H-bond acceptors (compounds listed in Table 10.2) were used in this study. E-state indexes were calculated using the software package Molconn-X2. The highest E-state value in each molecule, belonging to an atom capable of H-bond acceptance, was selected as a probable reflection of H-bond acceptor ability of that molecule.

Regression analysis revealed a low correlation coefficient of r = 0.395 for correlation with $\Sigma\beta^H_2$ values. On the other hand, these E-state values correlated relatively well

with the atomic charges on the most negatively charged atom (Q_{MN}):

$$\text{E-state} = -28.4\ Q_{MN} + 1.35 \qquad (11.5)$$

$n = 84 \quad s = 1.48 \quad r = 0.875 \quad F = 267.8$

Correlations within families were also examined. The only good correlation was for alcohols and ethers:

$$\Sigma\beta^{H}_{2} = 0.0102\ \text{E-state} + 0.398 \qquad (11.6)$$

$n = 8 \quad s = 0.003 \quad r = 0.986 \quad F = 211.2$

In this equation 2,2,2-trifluoroethanol and propan-2-ol have not been included. The E-state calculations resulted in a higher E-state value for a fluorine atom than for oxygen or nitrogen atoms; for example in 3-fluorophenol, the E-state value for oxygen is 8.87 compared with 12.01 for fluorine. This sort of irregularity can explain the lack of the correlation with $\Sigma\beta^{H}_{2}$.

11.3. Similarity index

Molecular similarity provides a quantitative measure of how closely one molecule resembles another. Expressed as an index with a range from zero to unity, which represents identity, the index may have obvious utility in structure-activity studies. It may also be helpful in finding the best orientational superimposition of two different molecules, which is of importance in mapping receptors.

Carbo (Carbo et al, 1980) and Hodgkin (Hodgkin & Richards, 1987) have introduced two different formulae to calculate similarity. Both these formulae are based on quantum mechanics and they compare electron density which may be derived from wave functions. The Oxford Molecular Program ASP (Automated Similarity Package), computes molecular similarity indices by these methods and it provides different options of molecular property in terms of which the similarity can be calculated: electrostatic potentials, electrostatic field and molecular shape.

In a congeneric series of compounds like substituted phenols where there is a common H-bonding substituent, the similarity of the electrostatic potentials around the H-bonding substituents was correlated with $\Sigma\alpha^{H}_{2}$ and $\Sigma\beta^{H}_{2}$.

a. For a set of substituted phenols, the similarity of the ESPs around the hydrogen of the hydroxyl group and hydroxyl group to those of the phenol and 4-nitrophenol (the strongest H-bond donor among the phenols), were calculated in terms of Carbo and Hodgkin indices. The ASP program was used for the Gaussian calculation of similarity. This program optimises the similarity index by changing angles, torsion angles and translation step by step. The carbo and Hodgkin indices obtained by these methods did not correlate with $\Sigma\alpha^{H}_{2}$ or $\Sigma\beta^{H}_{2}$.

b. In the QSAR of the bacterial growth inhibition acitivities of a set of pyridine derivatives, the H-bonding acceptor indicator variable for the substituent on the position 4 (H_a) has been used (Schultz & Moulton, 1985). In this study the ESP

similarity indices for the substituent were calculated; pyridine, nitropyridine and also 4-$CONH_2$-pyridine were used as the lead compounds against which similarity was calculated. The following equation was obtained in which *cp* is the Carbo similarity index with pyridine as the lead compound:

$$\log BR = -4.988\ cp + 3.83 \qquad (11.7)$$

$n = 20 \quad s = 0.524 \quad r = 0.740 \quad F = 21.8$

Unfortunately the ASP program failed to optimise the Hodgkin similarity index and calculated the single point index which did not correlate with log BR at all. Because $\Sigma\beta^H_2$ values were not available for the set, it cannot be determined that the *cp* in equation 11.7 represents the H-bonding ability. The relationship between log BR and H_a and MR are:

$$\log BR = -0.531\ H_a + 2.73 \qquad (11.8)$$

$n = 20 \quad s = 0.589 \quad r = 0.423 \quad F = 3.9$

$$\log BR = 0.0591\ MR + 1.72 \qquad (11.9)$$

$n = 20 \quad s = 0.398 \quad r = 0.791 \quad F = 30$

The failure of similarity indices in prediction of H-bonding ability could be for a number of reasons. In similarity calculations there is always the problem of alignment: here, the similarity calculations for a series of compounds resulted in different indices when different coordinates were used in the ASP program. For our purpose, the

choice of the lead compound is also important: when the lead compound is not the strongest (or the weakest) H-bonding compound, the similarity indices cannot symbolise the H-bonding ability.

11.4. Electrostatic potential derived (PD) atomic charges

The simplest procedure to represent accurately the molecular charge distribution is to use Mulliken populations from quantum mechanical calculations, but these are based on a simplified model of describing the electron distribution and often yield rather different multipole moments for the molecule from those calculated from the actual wavefunction. A second approach is to use the molecular electrostatic potential evaluated at points in space around the molecule as a guide and to fit this to point-charge models (Singh & Kollman, 1984). The accuracy of the model depends on how well the electric potential is fitted, and on the accuracy of the electric potential itself. The accuracy of the calculation of the electric potential in turn depends on the quality of the wavefunction (Williams & Yan, 1988). In a study of electrostatic interaction energies in some H-bonded systems (Ray et al, 1985), it was shown that the PD model is superior to the Mulliken atomic charges obtained from STO-3G wave functions.

Orozco and Luque (1990) calculated Mulliken and PD point charges using semiempirical (MNDO and AM1) and *ab initio* (STO-3G and 6-31G*) wave functions. Their results showed the usefulness of semiempirical wave functions to compute PD

charges at a low computational cost. MNDO electrostatic charges reproduced the sophisticated *ab initio* 6-31G* PD charges as well as the experimental dipoles excellently, the AM1 PD charges showing poorer ability to do so.

In this study the ZDO approximation (Giessner-Prettre & Pullman, 1972) was used to calculate PD point charges in the RATTLER program (Oxford Molecular). The accuracy of the charges derived from ZDO electrostatic potentials and those derived from the ESP calculated following deorthogonalisation (as in MOPAC program) are essentially equivalent (Rattler Manual). These charges were then compared with the electrostatic potentials in the prediction of H-bonding parameters $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$.

11.4.1. Methods

The SMILES codes of the molecules (listed in Tables 11.1 and 11.2) were given to the COBRA conformation analysis program. The lowest energy conformations were selected for further COSMIC minimisation in the program PIMMS. The minimised structures were then imported to the program RATTLER which calculates electrostatic potential-derived atomic charges using the vectors, geometry and dipoles of the molecule provided by a MOPAC 6.0 output file (in the MOPAC program, the MNDO semiempirical method was used). In order to calculate PD atomic charges, ESPs were calculated on three layers with 0.2Å interval from each other. A scaling factor of 1.6 was applied to the van der Waals radii to create the dimension of the innermost surface. The dot density on these surfaces was 1 dot/Å^2, with the total number of

points for each molecule being between 250-5000.

The atomic charge on the most positively charged hydrogen atom (Q_H) and also the atomic charge on the most negatively charged heteroatom or the average of the charges on the carbon atoms of the aromatic systems (Q_{MN}) for each molecule were selected. The MNDO method was used to calculate the energies of the highest occupied and the lowest unoccupied molecular orbital (E_{HOMO} and E_{LUMO} respectively). In order to compare the efficiency of these charges with that of ESPs in the prediction of H-bonding ability, the charges were subjected to the ESP calculation in the MAD program (the parameters used in this program have been explained in chapter 10). The highest and the lowest electrostatic potential on the Connolly surface of the each molecule which had been generated, using a probe radius of 1.05 Å, were selected (ESP^+ and ESP^-).

The resulting descriptors were correlated against H-bonding donor and acceptor abilities, $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$ (Abraham, 1993).

11.4.2. Results and discussion

The results of the calculations for H-bond donors (Q_H, ESP^+, and E_{LUMO}) together with $\Sigma\alpha^H_2$ for 109 compounds are tabulated in Table 11.1. Table 11.2 is $\Sigma\beta^H_2$, Q_{MN}, ESP^- and E_{HOMO} values for 109 H-bond acceptors.

The choice of the scaling factor which was used in RATTLER is justified by the findings that the calculated point-charge models are rather insensitive to which shell(s) is used (the range examined was 1.2-2.0 times the van der Waals radius) (Singh & Kollman 1984).

PD charges have also been shown to be unaffected by the total number of points, especially when the fitting is performed with more than 100 points, and also by the total number of layers, particularly when three or more layers are considered (Orozco and Luque, 1990).

Although the ESP^- was always near the carbonyl group of carboxylic acids, esters and amides, the PD Q_{MN} for all the carboxylic acids and two of the esters (ethyl acetate, vinyl acetate) resided on the ethereal oxygen, and in all the primary amides it was on the nitrogen. In all the substituted phenols Q_{MN} was on the oxygen of the OH group, but ESP^- for nitrophenols was near the nitro group.

The following equations were obtained for the correlation between $\Sigma\alpha^H_2$ and calculated parameters for H-bond donors.

$$\Sigma\alpha^H_2 = 2.66\ Q_H - 0.397 \qquad (11.10)$$

$n = 63 \quad s = 0.1925 \quad r = 0.616 \quad F = 37.3$

$$\Sigma\alpha^H_2 = 2.05\ Q_H - 0.073\ E_{LUMO} - 0.127 \qquad (11.11)$$

$n = 63 \quad s = 0.1619 \quad r = 0.754 \quad F = 39.5$

$$\Sigma\alpha^{H}_{2} = 0.030\ ESP^{+} - 0.406 \qquad (11.12)$$

n = 63 s = 0.1055 r = 0.902 F = 266.1

$$\Sigma\alpha^{H}_{2} = 0.026\ ESP^{+} - 0.039\ E_{LUMO} - 0.259 \qquad (11.13)$$

n = 63 s = 0.0919 r = 0.928 F = 185.8

In equations 11.10-11.13, t-ratios were significant for all the parameters (p = 0.000). In these correlations 2-substituted phenols were also included. Deleting these phenols improved the correlations with all parameters, especially correlations with PD charges which, in comparison with ESP^{+} correlations, are poor:

$$\Sigma\alpha^{H}_{2} = 2.72\ Q_{H} - 0.401 \qquad (11.14)$$

n = 58 s = 0.1866 r = 0.650 F = 41.1

$$\Sigma\alpha^{H}_{2} = 2.03\ Q_{H} - 0.090\ E_{LUMO} - 0.077 \qquad (11.15)$$

n = 58 s = 0.1334 r = 0.843 F = 67.5

$$\Sigma\alpha^{H}_{2} = 0.032\ ESP^{+} - 0.487 \qquad (11.16)$$

n = 58 s = 0.0899 r = 0.931 F = 362.8

$$\Sigma\alpha^{H}_{2} = 0.028\ ESP^{+} - 0.033\ E_{LUMO} - 0.317 \qquad (11.17)$$

n = 58 s = 0.0810 r = 0.945 F = 230.4

For H-bond acceptors the correlations between $\Sigma\beta^H_2$ and the calculated parameters were:

$$\Sigma\beta^H_2 = -0.695\ Q_{MN} + 0.401 \qquad (11.18)$$

n = 109 s = 0.1558 r = 0.726 F = 119.1

$$\Sigma\beta^H_2 = -0.019\ ESP^- - 0.054 \qquad (11.19)$$

n = 109 s = 0.1255 r = 0.832 F = 241.3

E_{HOMO} had no correlation with $\Sigma\beta^H_2$, unless the aromatic structures were deleted (see chapter 8):

$$\Sigma\beta^H_2 = -0.801\ Q_{MN} + 0.054 \qquad (11.20)$$

n = 63 s = 0.1581 r = 0.735 F = 71.6

$$\Sigma\beta^H_2 = 0.079\ E_{HOMO} - 0.715\ Q_{MN} + 0.96 \qquad (11.21)$$

n = 63 s = 0.1456 r = 0.785 F = 48.1

$$\Sigma\beta^H_2 = -0.021\ ESP^- - 0.090 \qquad (11.22)$$

n = 63 s = 0.1162 r = 0.867 F = 184.3

$$\Sigma\beta^H_2 = 0.077\ E_{HOMO} - 0.020\ ESP^- + 0.793 \qquad (11.23)$$

n = 63 s = 0.0982 r = 0.908 F = 141.9

All the parameters in equations 11.18-11.23 were significant ($p = 0.000$).

In all the correlations of H-bond acidity and basicity so far, electrostatic potentials showed statistically better relationships than did the atomic charge parameters.

Correlations within families

In phenols the correlations for H-bond acidity were generally better than those for H-bond basicity:

$$\Sigma\alpha^{H}_{2} = 15.6\ Q_H - 4.75 \qquad (11.24)$$

$n = 18 \quad s = 0.1522 \quad r = 0.679 \quad F = 13.7 \quad p = 0.002$

$$\Sigma\alpha^{H}_{2} = 0.021\ ESP^{+} - 0.090 \qquad (11.25)$$

$n = 18 \quad s = 0.0897 \quad r = 0.902 \quad F = 69.5$

In these equations 2-substituted phenols were included. Deleting these phenols surprisingly did not improve the correlation with PD charge, but the already good correlation with ESP^{+} (eq. 11.25) changed to an even better equation:

$$\Sigma\alpha^{H}_{2} = 0.017\ ESP^{+} + 0.037 \qquad (11.26)$$

$n = 13 \quad s = 0.0269 \quad r = 0.940 \quad F = 83.0$

H-bond acceptor ability of phenols ($\Sigma\beta^H_2$) did not have any correlation with Q_{MN} or ESP^- and even after deleting the 2-substituted phenols the only correlation found was:

$$\Sigma\beta^H_2 = 0.233\ E_{HOMO} - 0.012\ ESP^- + 2.10 \qquad (11.27)$$

$n = 13 \quad s = 0.0735 \quad r = 0.776 \quad F = 7.6$

The t-ratios for E_{HOMO} and ESP^- in this equations were 3.88 ($p = 0.003$) and 2.46 ($p = 0.034$).

Alcohols: For this family there was no correlation with PD Q_H values. On the other hand, the correlation with ESP^+ was good:

$$\Sigma\alpha^H_2 = 0.017\ ESP^+ - 0.132 \qquad (11.28)$$

$n = 12 \quad s = 0.0232 \quad r = 0.946 \quad F = 85.2$

For H-bond acceptor ability, ethers were also included in the correlation. There was no correlation with PD Q_{MN} and the values of these charges for alcohols showed the reverse the expected order by methyl electron donating inductive effect. However, the following equation is the correlation between $\Sigma\beta^H_2$ and ESP^-:

$$\Sigma\beta^H_2 = -0.021\ ESP^- - 0.090 \qquad (11.29)$$

$n = 16 \quad s = 0.0396 \quad r = 0.858 \quad F = 39$

Amines: For H-bond basicity there was no correlation with either Q_{MN} or ESP^-. The

correlations for H-bond acidity of primary and secondary amines with ESP^+ were very good (eqs. 11.30 & 11.31; pyrrole was excluded from the latter because of its large influence in the equation), but correlation with PD Q_H (11.32) was poor and existed only when pyrrole was deleted:

$$\Sigma\alpha^H_2 = 0.029\ ESP^+ - 0.440 \qquad (11.30)$$

n = 10 s = 0.0256 r = 0.971 F = 129.6

$$\Sigma\alpha^H_2 = 0.020\ ESP^+ - 0.270 \qquad (11.31)$$

n = 9 s = 0.0179 r = 0.910 F = 34

$$\Sigma\alpha^H_2 = 2.21\ Q_H - 0.541 \qquad (11.32)$$

n = 9 s = 0.0327 r = 0.656 F = 5.3 p = 0.055

Amides: As with amines, there was no correlation with either Q_{MN} or ESP^-; if acetanilide and N-methylformamide were deleted there is a correlation with Q_{MN} with a positive coefficient (r = 0.791), which shows the ordering of Q_{MN} values to be the reverse of the order expected from the inductive effect of alkyl groups in solution.

H-bond acidity af primary and secondary amides, after deletion of the outlier (N-methylformamide), had the following correlations with ESP^+ and Q_H:

$$\Sigma\alpha^H_2 = 0.088\ ESP^+ - 2.02 \qquad (11.33)$$

n = 6 s = 0.0440 r = 0.894 F = 15.9 p = 0.016

$$\Sigma\alpha^H_2 = 3.11\ Q_H - 0.238 \qquad (11.34)$$

n = 6 s = 0.0350 r = 0.934 F = 27.5 p = 0.006

Carboxylic acids: In correlations with $\Sigma\alpha^H_2$, PD Q_H was not successful unless E_{LUMO} was also used in the correlation analyses:

$$\Sigma\alpha^H_2 = 7.55\ Q_H - 0.156\ E_{LUMO} - 2.16 \qquad (11.35)$$

n = 12 s = 0.0716 r = 0.865 F = 13.4 p = 0.002

In equation 11.35 t-ratios for Q_H and E_{LUMO} were 4.17 (p = 0.000) and 4.67 (p = 0.000) respectively. The correlation with ESP^+ was good:

$$\Sigma\alpha^H_2 = 0.031\ ESP^+ - 0.338 \qquad (11.36)$$

n = 12 s = 0.0388 r = 0.958 F = 111.3

$$\Sigma\alpha^H_2 = 0.029\ ESP^+ - 0.033\ E_{LUMO} - 0.262 \qquad (11.37)$$

n = 12 s = 0.0322 r = 0.974 F = 83.9

In this equation t-ratios for ESP^+ and E_{LUMO} were 11.04 (p = 0.000) and 2.36 (p = 0.043) respectively.

H-bond baiscity for this family had the following correlations with ESP^- and Q_{MN} after deleting the four benzoic acids:

$$\Sigma\beta^H_2 = -0.018\ ESP^- - 0.159 \qquad (11.38)$$

n = 8 s = 0.0163 r = 0.980 F = 146.4

$$\Sigma\beta^{H}_{2} = -3.21\ Q_{MN} - 1.74 \qquad (11.39)$$

$n = 8 \quad s = 0.0441 \quad r = 0.844 \quad F = 14.9$

11.4.3. Conclusion

In conclusion, despite the known advantages of PD charges over Mulliken charges, and also despite the fact that these charges like electrostatic potentials are affected by steric factors, electrostatic potentials still are much better predictors of H-bond abilities than are potential derived charge parameters both across and within families (except for H-bond acidity of amides).

Table 11.1. $\Sigma\alpha^H_2$, E_{LUMO} calculated by the MNDO method, PD Q_H calculated by RATTLER, and ESP^+ calculated by MAD for some H-bonding acids

Compound	$\Sigma\alpha^H_2$	E_{LUMO}	Q_H	ESP^+
Diethylamine	0.08	4.3399	0.29	17.8
Methylamine	0.16	3.7070	0.31	22.0
Ethylamine	0.16	3.4520	0.31	21.7
n-Propylamine	0.16	3.3716	0.31	21.6
n-Butylamine	0.16	3.2060	0.29	20.0
Dimethylamine	0.08	3.3257	0.29	18.6
Di-n-propylamine	0.08	2.9830	0.28	17.4
Di-n-butylamine	0.08	2.9147	0.29	17.5
Formamide	0.62	1.5173	0.26	30.1
Acetamide	0.54	1.3458	0.25	28.8
Propionamide	0.55	1.3930	0.26	29.0
N-Methylformamide	0.40	1.4263	0.25	23.7
N-Methylpropionamide	0.40	1.3458	0.21	28.0
N-Methylacetamide	0.40	1.2977	0.20	27.8
Acetic acid	0.61	0.8510	0.39	32.2
Hexanoic acid	0.60	0.9032	0.37	30.6
Chloroacetic acid	0.74	-0.2019	0.39	36.3
Dichloroacetic acid	0.90	-0.6558	0.38	38.4
Trichloroacetic acid	0.95	-1.0437	0.38	41.0
Formic acid	0.75	0.9603	0.40	34.4
Propanoic acid	0.60	0.9037	0.39	32.0
Butanoic acid	0.60	0.9032	0.38	30.7
2-Methylbenzoic acid	0.60	-0.2407	0.36	29.9
3-Methylbenzoic acid	0.59	-0.5190	0.36	28.0
4-Methylbenzoic acid	0.60	-0.2977	0.36	29.7
Methanol	0.43	3.7953	0.33	30.3
Ethanol	0.37	3.5149	0.34	29.4
Propan-1-ol	0.37	3.2527	0.32	28.6
Propan-2-ol	0.33	3.3379	0.33	27.7
Butan-1-ol	0.37	3.1939	0.33	28.5
Hexan-1-ol	0.37	3.1495	0.32	28.4
2,2,2-Trifluoroethanol	0.57	1.4265	0.34	41.0
Cyclopentanol	0.32	3.0996	0.33	28.0
Cyclohexanol	0.32	3.0662	0.33	27.7
Prop-2-en-1-ol	0.38	0.8886	0.34	28.1
trans-But-2-en-1-ol	0.38	0.7299	0.33	28.1
Ethylthiol	0.00	1.8799	0.18	18.0
n-Propylthiol	0.00	1.8907	0.18	17.9
n-Butylthiol	0.00	1.8858	0.18	18.2
Benzylamine	0.10	0.0564	0.31	20.0
Acetanilide	0.50	0.1568	0.25	27.9
Benzoic acid	0.59	-0.2337	0.37	29.7
Phenol	0.60	0.2509	0.34	32.8

Table 11.1. Continued

Compound	$\Sigma\alpha^{H}_{2}$	E_{LUMO}	Q_H	ESP^+
2-Fluorophenol	0.61	-0.2149	0.35	26.2
3-Fluorophenol	0.68	-0.2026	0.34	36.4
4-Fluorophenol	0.63	-0.1727	0.34	36.4
2-Chlorophenol	0.32	-0.1734	0.34	24.2
3-Chlorophenol	0.69	-0.2096	0.34	36.4
4-Chlorophenol	0.67	-0.1799	0.34	37.2
2-Bromophenol	0.35	-0.1204	0.33	12.7
3-Bromophenol	0.70	-0.1645	0.34	35.7
4-Bromophenol	0.67	-0.1319	0.35	36.8
2-Methoxyphenol	0.22	0.1780	0.32	19.1
3-Methoxyphenol	0.59	0.1794	0.33	33.9
4-Methoxyphenol	0.57	0.0652	0.34	33.6
2-Nitrophenol	0.05	-0.9596	0.33	17.2
3-Nitrophenol	0.79	-0.9311	0.35	44.8
4-Nitrophenol	0.82	-0.8243	0.35	45.3
1-Naphthol	0.61	-0.3137	0.33	32.6
2-Naphthol	0.61	-0.3943	0.35	33.0
Benzyl alcohol	0.33	0.0886	0.33	27.5
Thiophenol	*0.09*	*-0.1546*	*0.17*	16.0
Pyrrole	0.41	1.2628	0.22	*28.8*

Table 11.2. $\Sigma\beta^{H}_{2}$, E_{HOMO} calculated by the MNDO method, PD Q_{MN} calculated by RATTLER, and ESP^{-} calculated by MAD for some H-bonding bases

Compound	$\Sigma\beta^{H}_{2}$	E_{HOMO}	Q_{MN}	ESP^{-}
Dichloromethane	0.05	-12.4853	-0.12	-9.7
Trichloromethane	0.02	-12.9203	-0.02	-8.3
1,2-Dichloroethane	0.11	-12.4154	-0.19	-7.6
1,1,1-Trichloroethane	0.09	-12.7890	-0.02	-10.2
1-Chlorobutane	0.10	-12.0741	-0.20	-12.2
Tribromomethane	0.06	-11.8621	0.12	-5.9
Diethyl ether	0.45	-10.9075	-0.47	-24.5
Di-n-propyl ether	0.45	-10.8158	-0.55	-26.4
Di-n-butyl ether	0.45	-10.9071	-0.53	-25.6
Propanone	0.49	-10.7521	-0.47	-30.1
Butanone	0.51	-10.6914	-0.48	-30.3
Cyclopentanone	0.52	-10.6080	-0.48	-30.2
Cyclohexanone	0.56	-10.5671	-0.48	-30.8
Methyl formate	0.38	-11.3684	-0.55	-35.9
Methyl acetate	0.45	-11.4593	-0.58	-32.4
Ethyl acetate	0.45	-11.4117	-0.62	-33.4
Vinyl acetate	0.43	-9.6663	-0.59	-30.7
Diethylamine	0.69	-11.1899	-0.76	-31.4
Methylamine	0.58	-10.5356	-0.90	-33.5
Ethylamine	0.61	-10.5329	-0.88	-34.7
n-Propylamine	0.61	-10.5281	-0.89	-34.5
n-Butylamine	0.61	-10.4560	-0.83	-34.0
Dimethylamine	0.66	-10.0480	-0.72	-30.0
Di-n-propylamine	0.69	-10.0099	-0.82	-31.6
Di-n-butylamine	0.69	-10.0228	-0.80	-31.7
Trimethylamine	0.67	-9.6139	-0.46	-25.3
Triethylamine	0.79	-9.5076	-0.54	-27.7
Formamide	0.60	-10.6950	-0.64	-37.0
Acetamide	0.68	-10.6075	-0.63	-37.7
Propionamide	0.68	-10.5986	-0.65	-37.8
N-Methylformamide	0.55	-10.3794	-0.55	-36.1
N-Methylpropionamide	0.71	-10.2654	-0.54	-37.2
N-Methylacetamide	0.72	-10.2713	-0.52	-36.7
N,N-Dimethylformamide	0.74	-10.1100	-0.53	-35.4
N,N-Dimethylacetamide	0.78	-10.0465	-0.53	-36.0
Acetic acid	0.44	-11.5714	-0.67	-32.2
Hexanoic acid	0.45	-11.4636	-0.66	-33.1
Chloroacetic acid	0.36	-11.8694	-0.66	-28.6
Dichloroacetic acid	0.27	-12.1320	-0.63	-23.6
Trichloroacetic acid	0.28	-12.3578	-0.63	-23.6
Formic acid	0.38	-11.7400	-0.68	-31.4
Propanoic acid	0.45	-11.4934	-0.68	-32.6
Butanoic acid	0.45	-11.4789	-0.67	-33.2

Table 11.2. Continued

Compound	$\Sigma\beta^{H}_{2}$	E_{HOMO}	Q_{MN}	ESP^{-}
2-Methylbenzoic acid	0.34	-9.6402	-0.67	-34.8
3-Methylbenzoic acid	0.38	-9.6530	-0.64	-33.7
4-Methylbenzoic acid	0.38	-9.7264	-0.66	-34.0
Methanol	0.47	-11.4146	-0.62	-26.7
Ethanol	0.48	-11.2964	-0.60	-26.9
Propan-1-ol	0.48	-11.2410	-0.59	-27.1
Propan-2-ol	0.56	-11.2053	-0.57	-27.7
Butan-1-ol	0.48	-11.2312	-0.59	-27.2
Hexan-1-ol	0.48	-11.2170	-0.60	-27.3
2,2,2-Trifluoroethanol	0.25	-12.3771	-0.48	-22.2
Cyclopentanol	0.56	-11.1069	-0.59	-27.1
Cyclohexanol	0.57	-11.0846	-0.59	-27.8
Prop-2-en-1-ol	0.48	-10.3465	-0.60	-27.8
trans-But-2-en-1-ol	0.48	-9.9655	-0.60	-27.5
Ethylthiol	0.24	-9.7380	-0.36	-13.2
n-Propylthiol	0.24	-9.7303	-0.37	-13.4
n-Butylthiol	0.24	-9.7298	-0.37	-13.3
Diethyl sulphide	0.32	-9.508	-0.34	-13.9
Di-n-Butyl sulphide	0.32	-9.5116	-0.36	-14.8
Trimethyl phosphate	1.00	-11.2055	-0.83	-46.8
Triethyl phosphate	1.06	-11.1161	-0.90	-48.8
Tri-n-butyl phosphate	1.21	-11.0878	-0.87	-48.3
Benzene	0.14	-9.3906	-0.05	-9.7
Toluene	0.14	-9.2816	-0.05	-9.8
o-Xylene	0.16	-9.2296	-0.05	-10.2
m-Xylene	0.16	-9.2398	-0.05	-9.8
p-Xylene	0.16	-9.1832	-0.05	-10.1
1,3,5-Trimethylbenzene	0.19	-9.2348	-0.04	-9.8
Hexamethylbenzene	0.21	-9.0391	-0.04	-10.7
Naphthalene	0.20	-8.5714	-0.05	-9.7
Phenanthrene	0.26	-8.4901	-0.05	-9.8
Chlorobenzene	0.07	-9.6227	-0.10	-8.1
Bromobenzene	0.09	-9.5502	-0.02	-3.8
Benzaldehyde	0.39	-9.7265	-0.48	-30.2
Acetophenone	0.48	-9.6678	-0.48	-31.0
Benzophenone	0.50	-9.5863	-0.53	-32.6
Benzylamine	0.72	-9.4996	-0.89	-34.3
Acetanilide	0.67	-9.2254	-0.70	-37.8
Benzoic acid	0.40	-9.7684	-0.66	-34.1
Phenol	0.30	-8.8825	-0.55	-22.4
2-Fluorophenol	0.26	-9.1463	-0.51	-24.0
3-Fluorophenol	0.17	-9.2180	-0.54	-28.4
4-Fluorophenol	0.23	-9.0069	-0.55	-28.7
2-Chlorophenol	0.31	-9.1616	-0.56	-18.3

Table 11.2. Continued

Compound	$\Sigma\beta^{H}_{2}$	E_{HOMO}	Q_{MN}	ESP^{-}
3-Chlorophenol	0.15	-9.2224	-0.53	-18.2
4-Chlorophenol	0.20	-9.1452	-0.55	-17.9
2-Bromophenol	0.31	-9.0905	-0.57	-18.9
3-Bromophenol	0.16	-9.1513	-0.53	-19.0
4-Bromophenol	0.20	-9.0911	-0.55	-18.3
2-Methoxyphenol	0.52	-8.6399	-0.50	-23.1
3-Methoxyphenol	0.39	-8.6971	-0.53	-21.5
4-Methoxyphenol	0.48	-8.8307	-0.56	-26.5
2-Nitrophenol	0.37	-9.7503	-0.38	-22.1
3-Nitrophenol	0.23	-9.7321	-0.55	-31.8
4-Nitrophenol	0.26	-9.8473	-0.54	-31.9
1-Naphthol	0.37	-8.3128	-0.53	-18.3
2-Naphthol	0.40	-8.4863	-0.56	-21.6
Benzyl alcohol	0.56	-9.5195	-0.59	-27.1
Thiophenol	0.16	-9.6251	-0.32	-13.7
N,N-Dimethylbenzene-sulphonamide	0.86	-10.2126	-0.80	-46.7
Tetrahydrofuran	0.48	-10.7749	-0.46	-26.8
1,4-Dioxane	0.64	-10.5518	-0.45	-21.8
Pyrrole	0.29	-8.5689	-0.17	-20.3
Pyrazine	0.62	-10.0219	-0.55	-25.4
Pyrimidine	0.65	-10.3760	-0.94	-32.4
Thiazole	0.45	-9.8840	-0.57	-30.3

12. Interaction energies at grid points around a molecule

The docking method is used when studying the interaction of two molecules, such as a protein and a drug. The docking procedure would be fairly straightforward, were it not necessary to take account of the different types of interaction. With conventional methods of computation and graphical display, every molecule is treated as an agglomeration of atoms, and each atom has its own particular properties, which might include a van der Waals radius, an electrostatic charge and a set of bond properties (molecular mechanics representation of molecules). The unified computer-graphics approach uses a similar representation for the first interacting molecule (the target), but only one atom or group at a time is considered from the other molecule. Such a group is called a probe.

Program GRID is a computational procedure initially designed for determining energetically favourable binding sites on molecules of known structure. It may be used to study individual molecules such as drugs, molecular arrays such as membranes or crystals, and macromolecules such as proteins. The procedure is to construct a three-dimensional orthogonal grid of points throughout and around the target molecule. Computations are then carried out to determine the energetic interactions of the chosen probe with the target, when the probe is located at the first position on the grid. The most favourable interaction is determined by trying various hydrogen bonding orientations for this type of probe in that position, and the best energy value (i.e. the most negative energy, corresponding to the greatest attraction

between probe and target) is assigned to the first grid point. The whole process is then repeated with the probe at the next point on the grid, and is continued point by point until an energy value (E_{xyz}) has been assigned for the probe at each grid position (Goodford, 1989). The dimensions of the array of points are determined so that all points on the first XY plane are outside the molecule, and the computed energy values are therefore small when the probe is in this plane. However, subsequent planes start to intersect the macromolecule, and large positive energies due to Lennard-Jones repulsion term may then be calculated for any grid point that happens to be near an atom. Other points lie in the interatomic spaces, and modest negative energies would then correspond to favourable interactions between the probe and the target molecule. These would be partly due to the attraction term of the Lennard-Jones function, partly to electrostatic effects, and partly to H-bond interactions.

The non-bonded interaction energy E_{xyz} of the probe at each xyz position on the GRID program is calculated as the sum of different components:

$$\Sigma E_{xyz} = \Sigma E_{lj} + \Sigma E_{el} + \Sigma E_{hb} \qquad (12.1)$$

Each individual term in the summations relates to one pairwise interaction between the probe at position xyz and a single "extended" atom of the molecule. The summations extend over all "extended" atoms of *the target molecule*. ΣE_{lj} is the Lennard-Jones function:

$$E_{lj} = A/d^{12} - B/d^{6} \qquad (12.2)$$

In this equation d is the distance between a pair of non-bonded atoms whose Lennard-Jones energy E_{lj} is described by the parameters A and B. When d is small, the A/d^{12} term generates a dominating repulsion corresponding to a large positive value of E_{lj}. This effectively defines a minimum separation that can be apportioned between the atoms, giving each of them a nominal radius, and thus determining a molecular surface.

ΣE_{el} is the electrostatic interaction energy; it does not diminish rapidly with distance. However, the magnitude of E_{el} is critically sensitive to the spatial dielectric behaviour of the environment. In the program GRID, it is assumed that a planar interface separates a homogeneous target-molecule phase of dielectric ζ from a homogeneous solution of dielectric ε.

$$E_{el} = \frac{pq}{K\zeta}\left[\frac{1}{d} + \frac{(\zeta-\varepsilon)/(\zeta+\varepsilon)}{(d^2 + 4s_p s_q)^{1/2}}\right] \qquad (12.3)$$

In this equation p and q are the electrostatic charges on the probe group and the pairwise target-molecule atom that are separated by a distance d, and K is a combination of geometrical factors and natural constants. s_p and s_q are the nominal depth of probe and each target-molecule atom in the target molecule.

ΣE_{hb} is a direction-dependent hydrogen bond function:

$$E_{hb} = [C/d^6 - D/d^4] \cos^m \theta \qquad (12.4)$$

When two identical atoms are interacting, the tabulated values for C and D determine their interatomic separation d_{min} at the bottom of the curve. If the atoms are of different types, the geometric mean of their individual D values and the arithmetic mean of their d_{min} separations are used and the appropriate C value is calculated from D and d_{min}. If the target molecule donates a hydrogen bond, then the bond direction is determined by the hydrogen position as computed from the heavy atom structure of the target molecule. θ is the angle DHP where D is the molecule donor atom, H is the hydrogen, and P is the probe accepting the hydrogen bond. The term m is normally 4, but the whole E_{hb} term is set to zero when $\theta \leq 90°$. If the probe group donates the bond, it is assumed that the probe can orient itself in order to form the most effective H-bond interaction with the acceptor atom of the target, and the cos θ term is set to unity (Goodford, 1985).

This program has been used in 3D QSAR studies using CoMFA, in which the congeneric series of molecules under study are superimposed and then the interaction energy between the molecules and a probe is calculated in the grid points around the molecules (Kim, 1993).

In the present study the GRID program was used to calculate the most negative (the most favourable) interaction energy between a suitable probe with a defined property, and H-bonding molecules.

12.1. Methods

The PDB format of molecules coordinates were used in the GRIN program which combines these coordinates with the parameters needed for energy calculation listed in table GRUB. The output of the GRIN program was used for GRID calculation. Grid points were generated around the molecule with the maximum distance of 4.5 Å in each direction. The distance between grid points surrounding the molecule were set to 0.2 Å (Goodford, personal communication). Carbonyl oxygen and amide nitrogen (connected to one hydrogen) were used as probes for H-bond donors and acceptors respectively. The dielectric constant of the media was set to that of water (80). The minimum electrostatic, H-bonding and Lennard-Jones interaction energies were obtained from the output file of GRID. This file was then used in the program MINIM which interpolates between the grid points to get a better estimate of the total minimum energy. The total minimum energy was obtained from the output file of MINIM.

12.2. Results and discussion

The results of calculations (total (E_T), electrostatic (E_Q), H-bonding (E_{HB}) and Lennard-Jones (E_{LJ}) minimum interaction energies) are listed in Table 12.1 (for H-bond donors) and Table 12.2 (for H-bond acceptors).

The regression analysis in MINITAB showed the following correlations between $\Sigma\alpha^H_2$

with E_T and E_{HB} (interaction energies with carbonyl oxygen):

$$\Sigma\alpha^H_2 = -0.0983 - 0.134\ E_T \qquad (12.5)$$

n = 115 s = 0.1820 r = 0.780 F = 175.9

$$\Sigma\alpha^H_2 = 0.0197 - 0.120\ E_{HB} \qquad (12.6)$$

n = 115 s = 0.1756 r = 0.797 F = 197.1

However the graphs showed that both equations are invalid because the correlation coefficients are obtained from two clusters of compounds, compounds with a H-bond donor hydrogen atom and those without it. When the compounds without a H-bond donor were separated out, there was no correlation with any of the energy components or the total energy for H-bond donors even when the multiple regression analyses using all the energy components were examined.

Interaction energies with the probe amide nitrogen connected to one hydrogen atom did not have any correlation with $\Sigma\beta^H_2$. Even for a single family of H-bond acceptors (phenols), there was not a successful correlation with any of the energy components. This could be due to the inconsistency of the distances from the molecules at which the minimum interaction energy happens. Unlike the calculation of electrostatic potentials, here it was not possible to calculate the interaction energies on a van der Waals surface of the molecule. The other explanation could be the empirical formulae which were used in this program to calculate the energy components, as molecular

mechanics methods are recommended only for large molecules because they demand only a fraction of the computing time required for a quantum mechanical calculation. In GRID, molecules are represented as collections of "extended" atoms (except where the hydrogen atoms are capable of H-bonding). Thus a methyl group is treated as a single entity with a van der Waals radius which is somewhat larger than the normal value for a carbon atom. This single extended atom replaces four real atoms for computational purposes, so that the size and duration of all the computations is significantly reduced (Goodford, 1985). All these approximations, which are intended for large molecules, reduce the accuracy of the calculations.

The dielectric constant of the environment could be set to a lower value. But such a change would affect (lower) only the electrostatic interaction energy (for example E_Q of methanol increased from -0.04 to -3.69 when ε was changed from 80 to 1), which has been studied in details in previous chapters (electrostatic potentials).

Table 12.1. Minimum interaction energies with carbonyl oxygen resulting from GRID (and MINIM)

Compound	$\Sigma\alpha^H_2$	E_T	E_Q	E_{LJ}	E_{HB}
Hept-1-yne	0.12	-0.820	-0.01	-0.78	0.00
Dichloromethane	0.10	-0.777	-0.01	-0.78	0.00
1,2-Dichloroethane	0.10	-0.846	-0.01	-0.86	0.00
1,1,1-Trichloroethane	0.00	-0.989	-0.05	-1.04	0.00
1-Chlorobutane	0.00	-0.747	-0.07	-0.75	0.00
Diethyl ether	0.00	-0.751	0.00	-0.75	0.00
Di-n-propyl ether	0.00	-0.833	0.00	-0.79	0.00
Di-n-butyl ether	0.00	-0.976	-0.01	-0.85	0.00
Propanone	0.04	-0.678	-0.01	-0.67	0.00
Butanone	0.00	-0.768	-0.11	-0.76	0.00
Cyclopentanone	0.00	-0.887	-0.09	-0.88	0.00
Cyclohexanone	0.00	-0.931	-0.08	-0.93	0.00
Methyl formate	0.00	-0.553	-0.06	-0.55	0.00
Methyl acetate	0.00	-0.747	-0.06	-0.74	0.00
Ethyl acetate	0.00	-0.772	-0.06	-0.78	0.00
Vinyl acetate	0.00	-0.816	-0.04	-0.81	0.00
Acetonitrile	0.07	-0.663	-0.02	-0.67	0.00
1-Cyanobutane	0.00	-0.952	-0.04	-0.93	0.00
Diethylamine	0.08	-4.857	0.00	-0.55	-4.47
Methylamine	0.16	-4.206	-0.01	-0.36	-3.95
Ethylamine	0.16	-4.378	-0.01	-0.46	-3.95
n-Propylamine	0.16	-4.406	-0.01	-0.49	-3.95
n-Butylamine	0.16	-4.429	-0.01	-0.50	-3.95
Dimethylamine	0.08	-4.798	-0.01	-0.49	-4.43
Di-n-propylamine	0.08	-5.028	0.00	-0.66	-4.36
Di-n-butylamine	0.08	-5.047	-0.03	-0.65	-4.39
Trimethylamine	0.00	-0.650	0.00	-0.64	0.00
Triethylamine	0.00	-0.935	-0.04	-0.89	0.00
Formamide	0.62	-5.083	-0.05	-0.50	-4.71
Acetamide	0.54	-4.995	-0.05	-0.54	-4.61
Propionamide	0.55	-5.080	-0.04	-0.73	-4.54
N-Methylformamide	0.40	-4.356	-0.02	-0.67	-3.88
N-Methylpropionamide	0.40	-4.459	-0.10	-0.96	-3.80
N-Methylacetamide	0.40	-4.304	-0.03	-0.82	-3.81
N,N-Dimethylformamide	0.00	-0.843	-0.02	-0.84	0.00
N,N-Dimethylacetamide	0.00	-0.964	-0.01	-0.95	0.00
Acetic acid	0.61	-4.205	-0.01	-0.57	-3.80
Hexanoic acid	0.60	-4.006	-0.08	-0.72	-3.62
Chloroacetic acid	0.74	-4.390	-0.03	-0.62	-3.99
Dichloroacetic acid	0.90	-4.603	-0.05	-0.93	-4.19
Trichloroacetic acid	0.95	-4.877	-0.09	-1.07	-4.39
Formic acid	0.75	-4.395	-0.02	-0.42	-4.04
Propanoic acid	0.60	-4.042	-0.01	-0.70	-3.72

Table 12.1. Continued

Compound	$\Sigma\alpha^H_2$	E_T	E_Q	E_{LJ}	E_{HB}
Butanoic acid	0.60	-4.125	-0.07	-0.80	-3.67
2-Methylbenzoic acid	0.60	-3.894	-0.11	-0.99	-3.58
3-Methylbenzoic acid	0.59	-3.887	-0.18	-0.89	-3.58
4-Methylbenzoic acid	0.60	-3.917	-0.14	-0.93	-3.59
Methanol	0.43	-4.014	-0.01	-0.35	-3.83
Ethanol	0.37	-4.092	-0.01	-0.47	-3.77
Propan-1-ol	0.37	-4.105	-0.02	-0.46	-3.74
Propan-2-ol	0.33	-4.171	-0.01	-0.51	-3.74
Butan-1-ol	0.37	-4.097	-0.06	-0.51	-3.73
Hexan-1-ol	0.37	-4.302	-0.05	-0.61	-3.71
2,2,2-Trifluoroethanol	0.57	-5.005	-0.15	-0.60	-4.55
Cyclopentanol	0.32	-4.234	-0.04	-0.72	-3.73
Cyclohexanol	0.32	-4.210	-0.02	-0.63	-3.73
Prop-2-en-1-ol	0.38	-4.097	-0.01	-0.51	-3.74
trans-But-2-en-1-ol	0.38	-4.036	0.00	-0.53	-3.68
Ethylthiol	0.00	-0.714	0.00	-0.72	0.00
n-Propylthiol	0.00	-0.777	0.00	-0.78	0.00
n-Butylthiol	0.00	-0.810	0.01	-0.78	0.00
Diethyl sulphide	0.00	-0.829	0.01	-0.83	0.00
Di-n-Butyl sulphide	0.00	-1.072	-0.01	-0.99	0.00
Trimethyl phosphate	0.00	-1.137	-0.01	-1.15	0.00
Triethyl phosphate	0.00	-1.324	-0.12	-1.23	0.00
Tri-n-butyl phosphate	0.00	-1.461	-0.12	-1.18	0.00
Benzene	0.00	-0.838	0.00	-0.85	0.00
Toluene	0.00	-0.897	-0.01	-0.89	0.00
o-Xylene	0.00	-0.963	-0.02	-0.95	0.00
m-Xylene	0.00	-0.961	-0.01	-0.90	0.00
p-Xylene	0.00	-0.958	-0.01	-0.90	0.00
1,3,5-Trimethylbenzene	0.00	-1.014	0.00	-0.90	0.00
Hexamethylbenzene	0.00	-1.182	-0.01	-0.97	0.00
Phenylethyne	0.12	-0.983	-0.02	-0.95	0.00
Naphthalene	0.00	-1.061	0.00	-1.03	0.00
Phenanthrene	0.00	-1.276	0.01	-1.08	0.00
Chlorobenzene	0.00	-0.960	-0.03	-0.96	0.00
Bromobenzene	0.00	-1.013	-0.02	-1.02	0.00
Benzaldehyde	0.00	-0.968	-0.09	-0.92	0.00
Acetophenone	0.00	-1.004	-0.09	-0.93	0.00
Benzophenone	0.00	-1.484	-0.40	-1.09	0.00
Benzonitrile	0.00	-1.017	-0.06	-1.01	0.00
Benzylamine	0.10	-4.725	-0.04	-0.89	-3.85
Acetanilide	0.50	-4.458	-0.05	-0.96	-3.71
Benzoic acid	0.59	-3.883	-0.18	-0.90	-3.59
Phenol	0.60	-3.866	-0.01	-0.70	-3.49
2-Fluorophenol	0.61	-4.090	0.02	-0.92	-3.72

Table 12.1. Continued

Compound	$\Sigma\alpha^{H}_{2}$	E_T	E_Q	E_{LJ}	E_{HB}
3-Fluorophenol	0.68	-4.043	-0.19	-0.87	-3.62
4-Fluorophenol	0.63	-4.102	-0.09	-0.63	-3.67
2-Chlorophenol	0.32	-4.058	-0.11	-1.03	-3.64
3-Chlorophenol	0.69	-4.076	-0.06	-1.02	-3.65
4-Chlorophenol	0.67	-4.051	-0.05	-0.82	-3.64
2-Bromophenol	0.35	-4.030	0.02	-1.07	-3.61
3-Bromophenol	0.70	-4.018	-0.04	-1.06	-3.60
4-Bromophenol	0.67	-4.023	-0.04	-1.06	-3.61
2-Methoxyphenol	0.22	-3.987	-0.04	-0.67	-3.59
3-Methoxyphenol	0.59	-4.006	-0.02	-0.61	-3.58
4-Methoxyphenol	0.57	-4.024	-0.02	-0.87	-3.60
2-Cyanophenol	0.74	-4.049	-0.07	-0.89	-3.61
3-Cyanophenol	0.77	-4.032	-0.05	-0.96	-3.62
3-Cyanophenol	0.79	-4.040	-0.08	-1.01	-3.63
2-Nitrophenol	0.05	-4.075	-0.60	-1.23	-3.63
3-Nitrophenol	0.79	-4.133	-0.32	-1.27	-3.65
4-Nitrophenol	0.82	-4.090	-0.36	-1.34	-3.67
1-Naphthol	0.61	-3.859	-0.05	-0.65	-3.44
2-Naphthol	0.61	-3.879	-0.06	-0.97	-3.48
Benzyl alcohol	0.33	-4.386	-0.06	-0.87	-3.65
Thiophenol	0.09	-0.982	0.02	-0.99	0.00
N,N-Dimethylbenzenesulphonamide	0.00	-1.416	-0.31	-1.42	0.00
Tetrahydrofuran	0.00	-0.745	-0.01	-0.73	0.00
1,4-Dioxane	0.00	-0.754	-0.05	-0.75	0.00
Pyrrole	0.41	-4.047	0.00	-0.77	-3.66
Pyrazine	0.00	-1.029	0.00	-1.04	0.00
Pyrimidine	0.00	-1.039	-0.03	-1.06	0.00
Thiazole	0.00	-1.000	-0.01	-1.01	0.00

Table 12.2-Minimum interaction energies of amide nitrogen connecting to one hydrogen atom resulting from GRID (and MINIM)

Compound	$\Sigma\beta^H_2$	E_T	E_Q	E_{LJ}	E_{HB}
Hept-1-yne	0.10	-1.171	0.00	-1.11	0.00
Dichloromethane	0.05	-1.091	-0.01	-1.09	0.00
1,2-Dichloroethane	0.11	-1.203	0.00	-1.20	0.00
1,1,1-Trichloroethane	0.09	-1.457	-0.02	-1.47	0.00
1-Chlorobutane	0.10	-1.058	-0.02	-1.05	0.00
Diethyl ether	0.45	-4.016	0.00	-0.98	-3.14
Di-n-propyl ether	0.45	-4.086	0.00	-0.93	-3.13
Di-n-butyl ether	0.45	-4.261	0.00	-1.08	-3.13
Propanone	0.49	-5.311	0.00	-0.69	-4.77
Butanone	0.51	-5.439	-0.03	-1.06	-4.77
Cyclopentanone	0.52	-5.370	0.00	-0.78	-4.78
Cyclohexanone	0.56	-5.386	-0.03	-0.86	-4.76
Methyl formate	0.38	-5.138	-0.01	-0.70	-4.69
Methyl acetate	0.45	-5.256	-0.01	-0.82	-4.69
Ethyl acetate	0.45	-5.489	0.00	-0.98	-4.73
Vinyl acetate	0.43	-5.380	-0.01	-0.96	-4.74
Acetonitrile	0.32	-4.400	0.00	-0.95	-3.76
1-Cyanobutane	0.36	-4.457	-0.01	-1.21	-3.82
Diethylamine	0.69	-4.042	0.00	-1.02	-3.01
Methylamine	0.58	-4.405	0.00	-0.51	-4.00
Ethylamine	0.61	-4.670	0.00	-0.67	-4.00
n-Propylamine	0.61	-4.711	0.00	-0.72	-4.00
n-Butylamine	0.61	-4.755	0.00	-0.71	-4.00
Dimethylamine	0.66	-4.219	0.00	-0.77	-3.65
Di-n-propylamine	0.69	-4.616	0.00	-0.99	-3.62
Di-n-butylamine	0.69	-4.445	0.00	-1.03	-3.41
Trimethylamine	0.67	-5.805	0.00	-0.80	-5.06
Triethylamine	0.79	-6.375	-0.01	-1.25	-5.06
Formamide	0.60	-5.387	0.00	-0.90	-4.72
Acetamide	0.68	-5.424	0.00	-0.91	-4.77
Propionamide	0.68	-5.469	-0.01	-0.96	-4.74
N-Methylformamide	0.55	-5.351	0.00	-0.88	-4.57
N-Methylpropionamide	0.71	-5.548	-0.03	-1.20	-4.78
N-Methylacetamide	0.72	-5.435	-0.01	-1.16	-4.75
N,N-Dimethylformamide	0.74	-5.454	-0.01	-1.01	-4.74
N,N-Dimethylacetamide	0.78	-5.464	0.00	-1.18	-4.75
Acetic acid	0.44	-5.251	0.00	-0.69	-4.68
Hexanoic acid	0.45	-5.622	-0.03	-0.97	-4.75
Chloroacetic acid	0.36	-5.214	-0.01	-0.86	-4.62
Dichloroacetic acid	0.27	-5.530	-0.01	-1.16	-4.53
Trichloroacetic acid	0.28	-5.746	-0.02	-1.50	-4.45
Formic acid	0.38	-5.079	-0.01	-0.67	-4.57
Propanoic acid	0.45	-5.344	0.00	-0.75	-4.72

Table 12.2. Continued

Compound	$\Sigma\beta^H_2$	E_T	E_Q	E_{LJ}	E_{HB}
Butanoic acid	0.45	-5.387	-0.02	-0.84	-4.74
2-Methylbenzoic acid	0.34	-5.442	0.01	-1.10	-4.68
3-Methylbenzoic acid	0.38	-5.568	-0.02	-1.13	-4.77
4-Methylbenzoic acid	0.38	-5.496	-0.02	-1.11	-4.73
Methanol	0.47	-4.931	0.00	-0.51	-4.55
Ethanol	0.48	-5.221	0.00	-0.65	-4.60
Propan-1-ol	0.48	-5.286	-0.01	-0.70	-4.63
Propan-2-ol	0.56	-5.373	0.00	-0.80	-4.63
Butan-1-ol	0.48	-5.344	-0.02	-0.73	-4.64
Hexan-1-ol	0.48	-5.725	-0.02	-1.03	-4.66
2,2,2-Trifluoroethanol	0.25	-4.814	-0.02	-0.96	-3.91
Cyclopentanol	0.56	-5.469	-0.01	-1.02	-4.64
Cyclohexanol	0.57	-5.506	-0.01	-0.98	-4.64
Prop-2-en-1-ol	0.48	-5.301	0.00	-0.78	-4.63
trans-But-2-en-1-ol	0.48	-5.383	0.00	-0.81	-4.68
Ethylthiol	0.24	-1.008	0.00	-1.01	0.00
n-Propylthiol	0.24	-1.102	0.00	-1.10	0.00
n-Butylthiol	0.24	-1.146	0.00	-1.10	0.00
Diethyl sulphide	0.32	-1.172	0.00	-1.16	0.00
Di-n-Butyl sulphide	0.32	-1.505	0.00	-1.36	0.00
Trimethyl phosphate	1.00	-4.554	0.00	-1.34	-3.40
Triethyl phosphate	1.06	-4.954	-0.03	-1.55	-3.46
Tri-n-butyl phosphate	1.21	-5.325	-0.04	-1.66	-3.49
Benzene	0.14	-1.198	0.00	-1.20	0.00
Toluene	0.14	-1.289	0.00	-1.26	0.00
o-Xylene	0.16	-1.383	-0.01	-1.34	0.00
m-Xylene	0.16	-1.368	0.00	-1.26	0.00
p-Xylene	0.16	-1.374	0.00	-1.27	0.00
1,3,5-Trimethylbenzene	0.19	-1.458	0.00	-1.27	0.00
Hexamethylbenzene	0.21	-1.702	0.00	-1.40	0.00
Phenylethyne	0.24	-1.419	-0.01	-1.36	0.00
Naphthalene	0.20	-1.524	0.00	-1.48	0.00
Phenanthrene	0.26	-1.804	0.00	-1.54	0.00
Chlorobenzene	0.07	-1.368	-0.01	-1.37	0.00
Bromobenzene	0.09	-1.447	0.00	-1.45	0.00
Benzaldehyde	0.39	-5.535	-0.02	-1.24	-4.81
Acetophenone	0.48	-5.506	-0.03	-1.34	-4.79
Benzophenone	0.50	-5.761	-0.02	-0.99	-4.84
Benzonitrile	0.33	-4.488	-0.02	-1.43	-3.84
Benzylamine	0.72	-5.450	-0.01	-1.37	-4.08
Acetanilide	0.67	-5.766	0.01	-1.45	-4.73
Benzoic acid	0.40	-5.489	0.01	-0.85	-4.77
Phenol	0.30	-5.364	0.00	-1.23	-4.82
2-Fluorophenol	0.26	-5.268	0.01	-1.34	-4.55

Table 12.2. Continued

Compound	$\Sigma\beta^H_2$	E_T	E_Q	E_{LJ}	E_{HB}
3-Fluorophenol	0.17	-5.304	-0.04	-0.77	-4.71
4-Fluorophenol	0.23	-5.230	-0.03	-0.74	-4.67
2-Chlorophenol	0.31	-5.533	-0.03	-1.46	-4.70
3-Chlorophenol	0.15	-5.299	-0.02	-1.44	-4.70
4-Chlorophenol	0.20	-5.271	-0.02	-1.19	-4.67
2-Bromophenol	0.31	-5.633	-0.01	-1.53	-4.72
3-Bromophenol	0.16	-5.362	-0.01	-1.51	-4.70
4-Bromophenol	0.20	-5.310	-0.02	-1.50	-4.70
2-Methoxyphenol	0.52	-5.614	-0.01	-1.27	-4.75
3-Methoxyphenol	0.39	-5.303	-0.01	-1.20	-4.75
4-Methoxyphenol	0.48	-5.300	0.00	-0.85	-4.72
2-Cyanophenol	0.33	-5.933	-0.02	-1.26	-4.71
3-Cyanophenol	0.28	-5.338	-0.01	-0.84	-4.71
4-Cyanophenol	0.29	-5.298	-0.02	-1.44	-4.70
2-Nitrophenol	0.37	-5.534	-0.01	-1.52	-4.58
3-Nitrophenol	0.23	-5.381	-0.04	-0.91	-4.69
4-Nitrophenol	0.26	-5.254	-0.03	-1.06	-4.67
1-Naphthol	0.37	-5.518	-0.02	-0.80	-4.82
2-Naphthol	0.40	-5.418	-0.02	-1.46	-4.83
Benzyl alcohol	0.56	-5.985	-0.02	-1.32	-4.70
Thiophenol	0.16	-1.405	-0.01	-1.40	0.00
N,N-Dimethylbenzenesulphonamide	0.86	-5.227	-0.01	-1.91	-3.47
Tetrahydrofuran	0.48	-3.900	0.00	-0.93	-3.13
1,4-Dioxane	0.64	-4.071	0.00	-1.05	-3.13
Pyrrole	0.29	-1.090	0.00	-1.09	0.00
Pyrazine	0.62	-5.537	0.00	-1.45	-4.82
Pyrimidine	0.65	-5.579	-0.01	-0.88	-4.83
Thiazole	0.45	-5.630	0.00	-1.40	-4.93

13. General conclusion

In general, atomic charge parameters, Q_H and Q_{MN}, were successful in predicting experimental H-bonding parameters of α^H_2, β^H_2, log K_α, log K_β, $\Sigma\alpha^H_2$ and $\Sigma\beta^H_2$, but family dependent properties were observed. Correlations with log K_α and log K_β were the poorest which could be due to the more dipolar solvent of 1,1,1-trichloroethane used in the measurement of these parameters compared with the solvent tetrachloromethane used to obtain the other parameters.

The family dependent behaviour of basicity (or acidity) dependent properties is said to be a result of the varying blend of electrostatic and charge transfer forces that is involved in any donor-acceptor combination (Maria et al, 1987). Therefore energies of the frontier orbitals were used to quantify charge transfer contribution to the H-bonding energy. When E_{HOMO} and E_{LUMO} where used together with charge parameters in a multiple regression, the two parameters were able to predict the H-bonding abilities of different classes of compounds and the separation of different families was not necessary. However, in order to find correlations for H-bond acceptors, it was necessary to delete aromatic structures.

In correlations within families, the energies of frontier orbitals and charge parameters cannot be used together because they are highly correlated. However, there were good correlations between $\Sigma\alpha^H_2$ and E_{LUMO} and also $\Sigma\beta^H_2$ and E_{HOMO} within families. In most cases, correlations with atomic charge parameters were superior to the

relationships with energies of frontier orbitals within families. For correlations with energies of frontier orbitals, aromatic structures had to be separated.

Four different semiempirical methods have been used to calculate atomic charges and energies of the frontier orbitals (chapters 7, 8 and 9). In addition, two other classical charge calculation methods were also examined (chapter 10). The charges calculated by quantum mechanical methods were superior to those calculated by classical methods, since, the classical methods calculated the same amount of charge on, for example, oxygen of all substituted phenols or hydroxyl hydrogen of different alcohols; therefore, quantum mechanical charges correlated better with H-bonding abilities.

Among the four semiempirical methods, Q_H and E_{LUMO} values of AM1 and MNDO method gave better correlations than did the other methods. The CNDO method is better than the PM3 method when correlating Q_H alone, or Q_H together with E_{LUMO}. E_{LUMO} values of CNDO method are the poorest in correlation with $\Sigma\alpha^H_2$. For H-bond acceptors, again AM1 and MNDO calculated atomic charges and energies of the frontier orbital correlated best with $\Sigma\beta^H_2$; the PM3 calculated parameters are better than those calculated by the CNDO method.

By replacing different H-bonding parameters by atomic charges and also energies of HOMO and LUMO in QSAR equations, it was shown that these theoretically derived parameters are useful in QSAR studies.

The highest and the lowest electrostatic potentials (ESP^+ and ESP^-) were used as an alternative to Q_H and Q_{MN} as electrostatic descriptors of H-bonding abilities. ESPs were better descriptors of H-bonding abilities than were the charges calculated by classical methods used in calculation of the ESPs. ESPs calculated from Abraham charges were better predictors of H-bonding abilities than those calculated from Gasteiger charges. The ESP^+ calculated by Abraham method are even better than the Q_H calculated by MNDO and AM1 methods; ESP^+ resulted from Gasteiger charges are better than PM3 charges (compare the results of chapter 7, chapter 9 (Tables 9.4) and chapter 10 (Table 10.5)). Although ESP^-s were better than Q_{MN}s calculated by semiempirical methods, incorporating E_{HOMO}, Q_{MN} and E_{HOMO} calculated by AM1 and MNDO method gave much better correlation than did ESP^- and E_{HOMO} (compare the results of chapter 7, chapter 9 (Tables 9.5) and chapter 10 (Table 10.7)). ESPs were also used in QSAR equations successfully.

One difficulty with ESPs is the dependence of ESP^+ and ESP^- on the conformation of the molecule used in the calculation; this is a result of the dependence of ESPs on the steric situation of the point in which ESP is being calculated. Therefore care must be taken concerning the choice of conformation.

Another method of charge calculation which was also used in this thesis was calculation of atomic charges from electrostatic potential. Electrostatic potential derived (PD) charges inherited the dependence on the conformation and being affected by the steric factors of ESPs, but ESPs still are much better predictors of H-bonding

abilities both across and within families.

When analysing the relationships between experimental H-bonding parameters and the theoretical parameters, the effect of alkyl substitution was observed to be controversial in different families and with different methods. In amines, amides and alcohols, alkyl groups seem to have an electron-withdrawing effect in AM1 and PM3 methods, which is opposite to their known inductive effect in solution. The MNDO calculation of Q_{MN} values for amines and amides, and Q_H values of alcohols and amides, show an electron-donating inductive effect for alkyl groups. The charges calculated by the CNDO method show an electron-donating effect for alkyl groups in alcohols and amines.

The fact that dipole moments cannot parametrise H-bonding abilities shows that atomic charges are the simplest method of representing the molecular charge distribution which can quantify H-bonding abilities.

The parameters studied in this investigation are more successful in prediction of H-bond donor ability than of H-bond acceptor ability. Different atomic polarisabilities of H-bond acceptor atoms, which have not been parametrised, might be the reason for this observation; in H-bond acceptance different atoms and atomic orbitals are involved, while, in H-bond donation only atomic orbitals of the hydrogen atom are responsible.

14. References

Abboud J.L., Kamlet M.J. and Taft R.W., *J. Am. Chem. Soc.*, **99** (1977) 8325-7.

Abboud J.-L.M., Sraidi K., Guiheneuf G., Negro A., Kamlet M.J. and Taft R.W., *J. Org. Chem.*, **50** (1985) 2870-3.

Abraham M.H., personal communication.

Abraham M.H., *Chem. Soc. Rev.*, (1993) 73-83.

Abraham M.H., Duce P.P., Grellier P.L., Prior D.V., Morris J.J. and Taylor P.J., *Tetrahedron Letters*, **29** (1988b) 1587-90.

Abraham M.H., Duce P.P., Prior D.V., Barratt D.G., Morris J.J. and Taylor P.J., *J. Chem. Soc. Perkin Trans. 2*, (1989a) 1355-75.

Abraham M.H., Duce P.P., Prior D.V., Schultz R.A., Morris J.J. and Taylor P.J., *J. Chem. Soc. Faraday Trans. 1*, **84** (1988a) 865-9.

Abraham M.H., Duce P.P., Schultz R.A., Morris J.J., Taylor P.J. and Barratt D.G., J. Chem. Soc. Faraday Trans. 1, 82 (1986) 3501-14.

Abraham M.H., Grellier P.L., Abboud J.-L.M., Doherty R.M. and Taft R.W., *Can. J. Chem.*, **66** (1988x) 2673-86.

Abraham M.H., Grellier P.L., Prior D.V., Duce P.P., Morris J.J. and Taylor P.J., *J. Chem. Soc. Perkin Trans. 2*, (1989b) 699-711.

Abraham M.H., Grellier P.L., Prior D.V., Morris J.J. and Taylor P.J., *J. Chem. Soc. Perkin Trans. 2*, (1990a) 521-9.

Abraham M.H., Grellier P.L., Prior D.V., Morris J.J., Taylor P.J., Laurence C. and Berthelot M., *Tetrahedron Letters*, **30** (1989c) 2571-4.

Abraham M.H., Grellier P.L., Prior D.V., Morris J.J., Taylor P.J., Maria P.-C. and Gal J.-F., *J. Phys. Org. Chem.*, **2** (1989d) 243-54.

Abraham M.H., Grellier P.L., Prior D.V., Taft R.W., Morris J.J., Taylor P.J., Laurence C., Berthelot M., Doherty R.M., Kamlet M.J., Abboud J.-L.M., Sraidi K., and Guiheneuf G., *J. Am. Chem. Soc.*, **110** (1988c) 8534-6.

Abraham M.H., Lieb W.R. and Franks N.P., *J. Pharm. Sci.*, **80** (1991a) 719-24.

Abraham M.H., Whiting G.S., Alarie Y., Morris J.J., Taylor P.J., Doherty R.M., Taft R.W. and Nielsen G.D., *Quant. Struct.-Act. Relat.*, **9** (1990b) 6-10.

Abraham M.H., Whiting G.S., Doherty R.M. and Shuely W.J., *J. Chem. Soc. Perkin Trans. 2*, (1990c) 1451-60.

Abraham M.H., Whiting G.S., Doherty R.M. and Shuely W.J., *J. Chromatogr.*, **587** (1991b) 213-28.

Abraham R.J. and Grant G.H., *J. Comp. Chem.*, **9** (1988) 244-56.

Abraham R.J., Grant G.H., Haworth I.S., and Smith P.E., *J. Comp.-Aided Mol. Design*, **5** (1991) 21-39.

Abraham R.J., Griffiths L., and Loftus p., *J. Comp. Chem.*, **3** (1982) 407-16.

Abraham R.J. and Haworth I.S., *J. Comp.-Aided Mol. Design*, **2** (1988) 125-35.

Abraham R.J. and Smith P.E., *J. Comp. Chem.*, **9** (1988) 288-97.

Abraham R.J. and Smith P.E., *J. Comp.-Aided Mol. Design*, **3** (1989) 175-87.

Allen L.C., *J. Am. Chem. Soc.*, **97** (1975) 6921-40.

Allinger N.L., *Adv. Phys. Org. Chem.*, **13** (1976) 1-82.

Allinger N.L., *J. Am. Chem. Soc.*, **99** (1977) 8127-34.

Allinger N.L. and Chung D.Y., *J. Am. Chem. Soc.*, **98** (1976) 6798-803.

Allinger N.L. and Wuesthoff T., *Tetrahedron*, **33** (1977) 3-10.

Almlöf J. and Støgård Å, *Chem. Phys. Lett.*, **29** (1974) 418-20.

Arnett E.M., Mitchel E.J. and Murty T.S.S.R., *J. Am. Chem. Soc.*, **96** (1974) 3875-91.

Barbato F., La Rotonda M.I. and Morrica P., In: *QSAR: Rational Approaches to the Design of Bioactive Compounds*, eds. Silipo C. and Vittoria A., Elsevier Science Publishers B.V., Amsterdam, 1991, pp. 99-102.

Bate-Smith E.C. and Westall R.G., *Biochim. Biophys. Acta*, **4** (1950) 427-40.

Bays J.P., *J. Chem. Education*, **69** (1992) 209-15.

Besler B.H., Merz Jr. K.M., and Kollman P.A., *J. Comp. Chem.*, **11** (1990) 431-9.

Bevan J.W., Legon A.C., Millen D.J. and Rogers S.C., *Proc. R. Soc. Lond. A*, **370** (1980) 239-55.

Biagi G.L., Recanatini M., Barbaro A.M., Guerra M.C., Sapone A., Borea P.A. and Pietrogrande M.C., In: *QSAR: Rational Approaches to the Design of Bioactive Compounds*, eds. Silipo C. and Vittoria A., Elsevier Science Publishers B.V., Amsterdam, 1991, pp. 83-90.

Bonaccorsi R., Cimiraglia R., Scrocco E. and Tomasi J., *Theoret. Chim. Acta (Berl.)*, **33** (1974) 97-103.

Bonaccorsi R., Pullman A., Scrocco E. and Tomasi J., *Chem. Phys. Lett.*, **12** (1972) 622-4.

Bonaccorsi R., Scrocco E. and Tomasi J., *J. Chem. Phys.*, **52** (1970) 5270-84.

Boyd D.B. and Lipkowitz K.B., *J. Chem. Education*, **59** (1982) 269-74.

Bratož S., *Adv. Quantum Chem.*, **3** (1966) 209-37.

Brauman J.I. and Blair L.K., *J. Am. Chem. Soc.*, **90** (1968) 6561-2.

Brauman J.I. and Blair L.K., *J. Am. Chem. Soc.*, **93** (1971) 3911-4.

Brauman J.I., Riveros J.M. and Blair L.K., *J. Am. Chem. Soc.*, **93** (1971) 3914-6.

Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S. and Karplus M., *J. Comput. Chem.*, **4** (1983) 187-217.

Caldwell G. and Kebarle P., *J. Am. Chem. Soc.*, **106** (1984) 967-9.

Carbo R., Leyda L. and Arnau M., *Int. J. Quantum Chem.*, **17** (1980) 1185-9.

Ceccarelli C., Jeffrey G.A. and Taylor R., *J. Mol. Structure*, **70** (1981) 255-71.

Charton M., In: *Steric Effects in Drug Design*, eds. Charton M. and Motoc I., Springer, Berlin, 1983, pp. 57-91.

Charton M., *Prog. Phys. Org. Chem.*, **16** (1987) 287-315.

Charton M. and Charton B.I., *J. Theoret. Biol.*, **99** (1982) 629-44.

Chastrette M., Rajzmann M., Chanon M. and Purcell K.F., *J. Am. Chem. Soc.*, **107** (1985) 1-11.

Chirlian L.E. and Francl M.M., *J. Comp. Chem.*, **8** (1987) 894-905.

Cohen N.C., Blaney J.M., Humblet C., Gund P. and Barry D.C., *J. Med. Chem.*, **33** (1990) 883-94.

Connolly M.J., *J. Appl. Cryst.*, **16** (1983) 548-58.

Connolly M.J., *J. Am. Chem. Soc.*, **107** (1985) 1118-24.

Coulson C.A., *Research (London)*, **10** (1957) 149-59.

Cox S.R. and Williams D.E., *J. Comp. Chem.*, **2** (1981) 304-23.

Cramer,III R.D., Patterson D.E. and Bunce J.D., *J. Am. Chem. Soc.*, **110** (1988) 5959-67.

Cummins P.L. and Gready J.E., *Chem. Phys. Lett.*, **174** (1990) 355-60.

Dearden J.C., In: *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*, ed. Karcher W. and Devillers J., Kluwer Academic Publishers, Dordrecht, 1990, pp. 25-60.

Dearden J.C., *Sci. Tot. Environ.*, **109/110** (1991) 59-68.

Dearden J.C., Bradburne S.J.A. and Abraham M.H., In: *QSAR: Rational Approaches to the Design of Bioactive Compounds*, eds. Silipo C. and Vittoria A., Elsevier Science Publishers B.V., Amsterdam, 1991, pp. 143-50.

Dearden J.C. and Ghafourian T., *J. Pharm. Pharmacol.*, **45** (1993) 1143.

Dearden J.C. and Ghafourian T., in: *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*, eds. Sanz F., Giraldo J. and Manaut F., Prous Science Publishers, 1995, pp. 117-9.

Dearden J.C., Gregg C.N. and Nicholson R.M., in: *QSAR: Quantitative Structure-Activity Relationships in Drug Design*, ed. Fauchere J.L., A.R. Liss, New York, 1989, pp. 353-6.

Dearden J.C. and Nicholson R.M., *Pestic. Sci.*, **17** (1986) 305-10.

Del Re G., *J. Chem. Soc.*, (1958) 4031-40.

Desiraju G.R., *Crystal Engineering, The Design of Organic Solids*, Elsevier, Amsterdam, 1989, pp. 115-73.

Dewar M.J.S., *The Molecular Orbital Theory of Organic Chemistry*, McGraw-Hill, New York, 1969.

Dewar M.J.S. and Thiel W., *J. Am. Chem. Soc.*, **99** (1977) 4899-907.

Dewar M.J.S. and Thiel W., *J. Am. Chem. Soc.*, **99** (1977) 4907-17.

Dewar M.J.S., Zoebisch E.G., Healy E.F. and Stewart J.J.P., *J. Am. Chem. Soc.*, **107** (1985) 3902-9.

Dixon S.L. and Jurs P.C., *J. Comp. Chem.*, **13** (1992) 492-504.

Drago R.S., *Structure and bonding 15*, Springer-Verlag, Heidelberg, 1973.

Drago R.S. and Wayland B.B., *J. Am. Chem. Soc.*, **87** (1965) 3571-7.

Drago R.S., Parr L.B. and Chamberlain C.S., *J. Am. Chem. Soc.*, **99** (1977) 3203-9.

Drago R.S., Vogel G.C. and Needham T.E., *J. Am. Chem. Soc.*, **93** (1971) 6014-26.

Dunfield L., Burgess A. and Scheraga H., *J. Phys. Chem.*, **82** (1978) 2609-16.

Dunn W.J., In: *Drug Discovery Technologies*, eds. Clark C.R. and Moos W.H., Ellis Horwood Ltd. publishers, Chichester, 1990, pp 22-43.

Dunn W.J., Johansson E. and Wold S., *Quant. Struct.-Act. Relat.*, **2** (1983) 156-63.

Dunn W.J., Wold S., Edlund U., Hellberg S. and Gasteiger J., *Quant. Struct.-Act. Relat.*, **3** (1984) 131-7.

El Tayar N., Tsai R.-S., Testa B., Carrupt P.-A., Leo A., *J. Pharm. Sci.*, **80** (1991) 590-8.

Elguero J., Marzin C., Katritzky A.R. and Linda P., *The Tautomerism of Hetrocycles*, Academic Press, New York, 1976.

Epley T.D. and Drago R.S., *J. Am. Chem. Soc.*, **89** (1967) 5770-3.

Famini G.R., Penski C.A. and Wilson L.Y., *J. Phys. Org. Chem.*, **5** (1992) 395-408.

Ferenczy G.G., Reynolds C.A. and Richards W.G., *J. Comp. Chem.*, **11** (1990) 159-69.

Fersht A.R., *TIBS*, **12** (1987) 301-4.

Fleming I., *Frontier Orbitals and Organic Chemical Reactions*, John Wiley & Sons, London, 1978.

Fujita T., In: *Comprehensive Medicinal Chemistry, Vol. 4, Quantitative Drug Design,*

ed. Ramsden C.A., Pergamon Press Plc, Oxford, 1990, pp. 497-560.

Fujita T., Iwasa J. and Hansch C., *J. Am. Chem. Soc.*, **86** (1964) 5175-80.

Fujita T., Nishioka T. and Nakajima M., *J. Med. Chem.*, **20** (1977) 1071-81.

Fukui K., *Accts. Chem. Res.*, **4** (1971) 57-64.

Fukui K., Yonezawa T. and Nagata C., *Bull. Chem. Soc. Japan*, **27** (1954) 423-7.

Gasteiger J. and Marsili M., *Tetrahedron*, **36** (1980) 3219-28.

Ghio C. and Tomasi J., *Theoret. Chim. Acta (Berl.)*, **30** (1973) 151-8.

Giessner-Prettre C. and Pullman A., *Theoret. Chim. Acta (Berl.)*, **25** (1972) 83-8.

Giessner-Prettre C. and Pullman A., *Theoret. Chim. Acta (Berl.)*, **33** (1974) 91-4.

Giessner-Prettre C. and Pullman A., *Theoret. Chim. Acta (Berl.)*, **37** (1975) 335-9.

Gimarc B.M., *Molecular Structure and Bonding*, Academic Press, New York, 1979.

Goldblum A. and Pullman B., *Theoret. Chim. Acta (Berl.)*, **47** (1978) 345-7.

Good A.C., Hodgkin E.E. and Richards W.G., *J. Chem. Inf. Comput. Sci.*, **32** (1992) 188-91.

Goodford P.J., *J. Med. Chem.*, **28** (1985) 849-57.

Goodford P.J., In: *Computer-Aided Molecular Design*, ed. Richards W.G., Technical Services Ltd., Oxford, 1989, pp. 147-55.

Graffeuil M., Labarre J.-F. and Leibovici C., *J. Mol. Struct.*, **22** (1974) 97-108.

Grieco C., Silipo C. and Vittoria A., *Il Farm. Ed. Sci.*, **34** (1978) 433-64.

Guidry R.H. and Drago R.S., *J. Am. Chem. Soc.*, **95** (1973) 759-63.

Gutmann V., *The Donor-Acceptor Approach to Molecular Interactions*, Plenum Press, New York, 1978.

Hall G.G., *Adv. Atomic Mol. Phys.*, **20** (1985) 41-63.

Hall L.H., Mohney B. and Kier L.B., *Quant. Struct.-Act. Relat.*, **10** (1991) 43-51.

Hall L.H., Mohney B.K. and Kier L.B., *Quant. Struct.-Act. Relat.*, **12** (1993) 44-8.

Hammett L.P., *Physical Organic Chemistry*, McGraw-Hill, New York, 1940.

Hansch C., *Acc. Chem. Res.*, **2** (1969) 232-9.

Hansch C. and Fujita T., *J. Am. Chem. Soc.*, **86** (1963) 1616-26

Hansch C. and Leo A., *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.

Hansch C., Maloney P.P., Fujita T. and Muir R.M., *Nature*, **194** (1962) 178-80.

Hansch C., Quinlan J.E. and Lawrence G.L., *J. Org. Chem.*, **33** (1968) 347-50.

Hehre W.J., Radom L. and Schleyer P.V.R., *Ab Initio Molecular Orbital Theory*, Wiley Interscience, New York, 1986.

Hehre W.J. and Pople J.A., *Tetrahedron Lett.*, (1970) 2959-62.

Hickey J.P. and Passino-Reader D.R., *Environ. Sci. Technol.*, **25** (1991) 1753-60.

Hodgkin E.E. and Richards W.G., *Int. J. Quantum Chem. Biol. Symp.*, **14** (1987) 105-10.

Hoffmann R., *J. Chem. Phys.*, **39** (1963) 1397-412.

Hopfinger A.J., *Conformational Properties of Macromolecules*, Academic Press, London, 1973, pp. 94-8.

Hopfinger A.J., *J. Med. Chem.*, **28** (1985) 1133-9.

Huyskens P.L., Marshal H. and Zeegers-Huyskens Th., *J. Mol. Struct.*, **158** (1987) 379-85.

James K.C., in: *Introduction to the Principles of Drug Design*, ed. Smith H.J., Wright, London, 1988, pp. 240-64.

Joesten M.D. and Schaad L.J., *Hydrogen Bonding*, Marcel Dekker, New York, 1974.

Johansson A., Kollman P., Rothenberg S. and McKelvey J., *J. Am. Chem. Soc.*, **96** (1974) 3794-800.

Jogansen A.V., *Theor. i Exper. Khim.*, **7** (1971a) 302-11.

Jogansen A.V., *Theor. i Exper. Khim.*, **7** (1971b) 312-7.

Jorgensen W.L., *J. Phys. Chem.*, **90** (1986) 1276-84.

Kamlet M.J., Abboud J.-L.M., Abraham M.H. and Taft R.W., *J. Org. Chem.*, **48** (1983) 2877-87.

Kamlet M.J., Abboud J.-L.M. and Taft R.W., *J. Am. Chem. Soc.*, **99** (1977) 6027-38.

Kamlet M.J., Abboud J.-L.M. and Taft R.W., *Prog. Phys. Org. Chem.*, **13** (1981) 485-630.

Kamlet M.J., Abraham M.H., Doherty R.M. and Taft R.W., *J. Am. Chem. Soc.*, **106** (1984) 464-6.

Kamlet M.J., Dickinson C., Gramstad T. and Taft R.W., *J. Org. Chem.*, **47** (1982) 4971-5.

Kamlet M.J., Doherty R.M., Abboud J.-L.M., Abraham M.H. and Taft R.W., *J. Pharm. Sci.*, **75** (1986) 339-49.

Kamlet M.J., Doherty R.M., Abraham M.H., Carr P.W., Doherty R.F. and Taft R.W., *J. Phys. Chem.*, **91** (1987) 1996-2004.

Kamlet M.J., Doherty R.M., Abraham M.H., Marcus Y. and Taft R.W., *J. Phys. Chem.*, **92** (1988a) 5244-55.

Kamlet M.J., Doherty R.M., Abraham M.H. and Taft R.W., *Quant. Struct.-Act. Relat.*, **7** (1988b) 71-8.

Kamlet M.J., Doherty R.M., Veith G.D., Taft R.W. and Abraham M.H., *Environ. Sci. Technol.*, **20** (1986) 690-5.

Kamlet M.J., Gal J.-F., Maria P.C. and Taft R.W., *J. Chem. Soc. Perkin Trans. 2*, (1985) 1583-9.

Kamlet M.J. and Taft R.W., *J. Am. Chem. Soc.*, **98** (1976) 2886-92.

Kamoshita K., Ohno I., Fujita T., Nishioka T., Nakajima M., *Pest. Biochem. Physiol.*, **11** (1979) 83-103.

Kauzmann W., *Quantum Chemistry*, Academic Press, New york, 1957.

Kenny P.W., *J. Chem. Soc. Perkin Trans. 2*, (1994) 199-202.

Kier L.B., *Molecular Orbital Theory in Drug Research*, Academic Press, New York, 1971.

Kier L.B., *Quant. Struct.-Act. Relat.*, **4** (1985) 109-16.

Kier L.B. and Hall L.H., *Molecular Connectivity in Structure-Activity analysis*, John Wiley, London, 1986.

Kier L.B. and Hall L.H., *Adv. Drug Res.*, **22** (1992) 1-38.

Kim K.H., In: *QSAR: Rational Approaches to the Design of Bioactive Compounds*, eds. Silipo C. and Vittoria A., Elsevier Science Publishers B.V., Amsterdam, 1991, 151-4.

Kim K.H., *Med. Chem. Res.*, **2** (1992a) 22-7.

Kim K.H., *Quant. Struct.-Act. Relat.*, **11** (1992b) 127-34.

Kim K.H., *Quant. Struct.-Act. Relat.*, **11** (1992c) 309-17.

Kim K.H., *J. Comput.-Aided Mol. Design*, **7** (1993) 71-82.

Kim K.H., Greco G., Novellino E., Silipo C. and Vittoria A., *J. Comput.-Aided Mol. Design*, **7** (1993) 263-80.

Koch R., *Acta Hydrochim. Hydrobiol.*, **10** (1982) 535-49.

Koehler M.G., Hopfinger A.J. and Seydel J.K., *Theochem*, **179** (1988) 319-32.

Kollman P.A., *J. Am. Chem. Soc.*, **99** (1977) 4875-94.

Kollman P.A., *J. Am. Chem. Soc.*, **100** (1978) 2974-84.

Kollman P., Johansson A. and Rothenberg S., *Chem. Phys. Lett.*, **24** (1974) 199-202.

Kollman P.A., McKelvey J., Johansson A. and Rothenberg S., *J. Am. Chem. Soc.*, **97** (1975) 955-65.

Kroeger M.K. and Drago R.S., *J. Am. Chem. Soc.*, **103** (1981) 3250-62.

Kroon J., Kanters J.A., Van Duijneveldt-Van de Rijdt J.G.C.M., Van Duijneveldt F.B. and Vliegenthart J.A., *J. Mol. Structure*, **24** (1975) 109-29.

Kubinyi H., *J. Med. Chem.*, **20** (1977) 625-9.

Lavery R., Etchebest C. and Pullman A., *Chem. Phys. Lett.*, **85** (1982) 266-70.

Leegwater D.C., *Aquat. Toxicol.*, **15** (1989) 157-68.

Legon A.C., *Chem. Soc. Rev.*, **19** (1990) 197-237.

Legon A.C. and Millen D.J., *Discuss. Faraday Soc.*, **73** (1982) 71-8.

Legon A.C. and Millen D.J., *J. Am. Chem. Soc.*, **109** (1987) 356-8.

Legon A.C. and Willoughby L.C., *Chem. Phys. Lett.*, **143** (1988) 214-20.

Leo A., Hansch C. and Jow P.Y.C., *J. Med. Chem.*, **19** (1976) 611-5.

Leo A., Jow P.Y.C., Silipo C. and Hansch C., *J. Med. Chem.*, **18** (1975) 865-8.

Lewis T.P., *Tetrahedron*, **25** (1969) 4117-26.

Lien E.J. and Wang P.H., *J. Pharm. Sci.*, **69** (1980) 648-50.

Lifson S., Hagler A.T. and Dauber P., *J. Am. Chem. Soc.*, **101** (1979) 5111-41.

Lifson S. and Warshel A., *J. Chem. Phys.*, **49** (1968) 5116-29.

Loew G.H. and Burt S.K., In: *Comprehensive Medicinal Chemistry, Vol. 4, Quantitative Drug Design*, ed. Ramsden C.A., Pergamon Press Plc, Oxford, 1990, pp 105-21.

Luque F.J., Illas F. and Orozco M., *J. Comp. Chem.*, **11** (1990) 416-30.

Manaut F., Lozoya E. and Sanz F., In: *QSAR: Rational Approaches to the Design of Bioactive Compounds*, eds. Silipo C. and Vittoria A., Elsevier Science Publishers B.V., Amsterdam, 1991, pp. 339-42.

Maria P.-C., Gal J.-F., de Franceschi J. and Fargin E., *J. Am. Chem. Soc.*, **109** (1987) 483-92.

Martin Y.C., *Quantitative Drug Design*, Marcel Dekker, New York, (1978).

McCammon J.A., In: *Comprehensive Medicinal Chemistry*, ed. Hansch C., Vol. 4, ed. Ramsden C.A., Pergamon Press, 1990, pp. 139-52.

McMillin D.R. and Drago R.S., *Inorg. Chem.*, **11** (1972) 872-9.

Merz Jr. K.M., *J. Comp. Chem.*, **13** (1992) 749-67.

Meyer A.Y. and Richards W.G., *J. Comput.-Aided Mol. Design*, **5** (1991) 427-9.

Meyer H.H., *Arch. Exp. Pathol. Pharmakol.*, **42** (1899) 109-18.

Miller K.J. and Savchik J.A., *J. Am. Chem. Soc.*, **101** (1979) 7206-13.

Mitsky J., Joris L. and Taft R.W., *J. Am. Chem. Soc.*, **94** (1972) 3442-5.

Momany F.A., *J. Phys. Chem.*, **82** (1978) 592-601.

Moriguchi I., *Chem. Pharm. Bull. Japan*, **23** (1975) 247-57.

Morley S.D., Abraham R.J., Haworth I.S., Jackson D.E., Saunders M.R. and Vinter J.G., *J. Comput.-Aided Mol. Design*, **5** (1991) 475-504.

Morokuma K., *J. Chem. Phys.*, **55** (1971) 1236-44.

Mullay J., *J. Am. Chem. Soc.*, **108** (1986) 1770-5.

Mulliken R.S., *J. Chem. Phys.*, **23** (1955) 1833-40.

Mulliken R.S. and Person W.B., *Molecular Complexes: A Lecture and Reprint Volume*, Wiley-interscience, New York, N.Y., 1969.

Murray J.S., Lane P., Brink T. and Politzer P., *J. Phys. Chem.*, **97** (1993) 5144-8.

Murray J.S. and Politzer P., *J. Org. Chem.*, **56** (1991) 6715-7.

Murray J.S. and Politzer P., *J. Chem. Research(S)*, (1992) 110-1.

Murray J.S., Ranganathan S. and Politzer P., *J. Org. Chem.*, **56** (1991) 3734-7.

Murrell J.N., Kettle S.F.A., Tedder J.M., *The Chemical Bond*, John Wiley & Sons, 1985.

Murrell J.N. and Van Duijneveldt F.B., *J. Chem. Phys.*, **46** (1967) 1759-72.

Murthy A.S.N. and Rao C.N.R., *Chem. Phys. Lett.*, **2** (1968a) 123-5.

Murthy A.S.N. and Rao C.N.R., *Appl. Spectrosc. Rev.*, **2** (1968b) 69-191.

Murthy A.S.N. and Rao C.N.R., *J. Mol. Structure*, **6** (1970) 253-82.

Neder K.M. and Whitlock H.W., *J. Am. Chem. Soc.*, **112** (1990) 9412-4.

Nishioka T., Fujita T., Kamoshita K. and Nakajima M., *Pestic. Biochem. Biophysiol.*, **7** (1977) 107-21.

Nys G.G. and Rekker R.F., *Eur. J. Med. Chem. (Chim. Ther.)*, **9** (1974) 361-75.

Orozco M. and Luque F.J., *J. Comp. Chem.*, **11** (1990) 909-23.

Overton E., *Z. Physik. Chem.*, **22** (1897) 189-209.

Paine G.H. and Scheraga H.A., *Biopolymers*, **24** (1985) 1391-436.

Pariser R. and Parr R.G., *J. Chem. Phys.*, **21** (1953) 466-71, 767-76.

Penniston J.T., Beckett L., Bentley D.L. and Hansch C., *Mol. Pharmacol.*, **5** (1969) 333-41.

Petrongolo C., Preston H.J.T. and Kaufman J.J., *Int. J. Quantum Chem.*, **13** (1978) 457-68.

Pimentel G.C. and McClellan A.L., *The Hydrogen Bond*, Freeman and company, San Francisco and London, 1960.

Politzer P., Abrahmsen L. and Sjoberg P., *J. Am. Chem. Soc.*, **106** (1984) 855-60.

Politzer P., Landry S.J. and Warnhelm T., *J. Phys. Chem.*, **86** (1982) 4767-71.

Politzer P., Lane P., Murray J.S. and Brink T., *J. Phys. Chem.*, **96** (1992) 7938-43.

Politzer P., Laurence P.R. and Jayasuriya K., *Environ. Health Persp.*, **61** (1985) 191-202.

Politzer P., Murray J.S., Lane P. and Brink T., *J. Phys. Chem.*, **97** (1993) 729-32.

Pople J.A., *J. Chem. Phys.*, **21** (1953) 2234-5.

Pople J.A. and Beveridge D.L., *Approximate Molecular Orbital Theory*, McGraw-Hill, New York, 1970.

Pople J.A., Beveridge D.L. and Dobosh P.A., *J. Chem. Phys.*, **47** (1967) 2026-33.

Pullman A. and Berthod H., *Theoret. Chim. Acta (Berl.)*, **48** (1978) 269-77.

Pullman B., Goldblum A. and Berthod H., *Biochim. Biophys. Res. Commun.*, **77** (1977) 1166-9.

Raevsky O.A., In: *QSAR in Drug Design and Toxicology*, eds. Hadzi D. and Jerman-Blazic B, Elsevier, Amsterdam, 1987, pp. 31-6.

Raevsky O.A., Grigor'ev V. and Solov'ev V., *Khim.-Pharm. Zhurn.*, **23** (1989) 1294-300.

Raevsky O.A., Veniamin Yu.G., Kireev D.B. and Zefirov N.S., *Quant. Struct.-Act. Relat.*, **11** (1992) 49-63.

Ray N.K., Shibata M., Bolis G. and Rein R., *Int. J. Quantum Chem.*, **27** (1985) 427-37.

Rein R., *Adv. Quantum Chem.*, **7** (1973) 335-9.

Reynolds C.A., Essex J.W., and Richards W.G., *J. Am. Chem. Soc.*, **114** (1992) 9075-9.

Richards F.M., *Ann. Rev. Biophys. Bioeng.*, **6** (1977) 151-76.

Richards W.G., *Quantum Pharmacology*, Butterworth & Co, London, 1983, pp. 159-61.

Rigby M., Smith E.B., Wakeham W.A. and Maitland G.C., *The Forces Between Molecules*, Clarendon Press, Oxford, 1986.

Rodriguez J., Manaut F. and Sanz F., *J. Comp. Chem.*, **14** (1993) 922-7.

Sabatino A., La Manna G. and Paolini L., *J. Phys. Chem.*, **84** (1980) 2641-5.

Sabljic A., *Environ. Sci. Technol.*, **21** (1987) 358-66.

Sanderson R.T., *Chemical Periodicity*, Reinhold, New York, 1960.

Sanz F., Manaut F., Rodriguez J, Lozoya E. and Lopez-de-Brinas E., *J. Computer-Aided Mol. Design*, **7** (1993) 337-47.

Schneider W.G., *J. Chem. Phys.*, **23** (1955) 26-30.

Schubert W.M., Murphy R.B. and Robins J., *Tetrahedron*, **17** (1962) 199-214.

Schultz T.W., Riggin G.W. and Wesley S.K., In: *QSAR in Environmental Toxicology - II*, ed. Kaiser K.L.E., D. Reidel, Dordrecht, 1987, pp. 333-45.

Schultz T.W. and Moulton B.A., *Ecotoxicol. Environ. Saf.*, **10** (1985) 97-111.

Scrocco E. and Tomasi J., *Topics in Current Chemistry*, No. **42**, Springer-Verlag, Berlin, 1973, p. 25.

Scrocco E. and Tomasi j., *Adv. Quantum Chem.*, **2** (1978) 115-93.

Seibel G.L. and Kollman P.A., In: *Comprehensive Medicinal Chemistry*, ed. Hansch C., Vol. 4, ed. Ramsden C.A., Pergamon Press, 1990, pp. 125-38.

Seiler P., *Eur. J. Med. Chem.*, **9** (1974) 473-9.

Sherry A.D., In: *The Hydrogen Bond, Vol. 3, Dynamics, Thermodynamics and Special Systems*, eds. Schuster P., Zundel G. and Sandorfy C., North-Holland Publishing Company, Amsterdam, 1976.

Singh U.C. and Kollman P.A., *J. Comp. Chem.*, **5** (1984) 129-45.

Sjoberg P. and Politzer P., *J. Phys. Chem.*, **94** (1990) 3959-61.

Smith R.P., Ree T., Magee J.L., and Eyring H., *J. Am. Chem. Soc.*, **73** (1951) 2263-8.

Sokolov N.D., *Dokladi*, **58** (1947) 611-6.

Srebrenik S., Weinstein H. and Pauncz R., *Chem. Phys. Lett.*, **20** (1973) 419-23.

Stewart J.J.P., *J. Comput. Chem.*, **10** (1989) 209-20.

Stewart J.J.P., *J. Comput. Chem.*, **10** (1989) 221-64.

Stewart J.J.P., *J. Computer-Aided Mol. Design.*, **4** (1990) 1-105.

Stewart J.J.P., *J. Comput. Chem.*, **12** (1991) 320-41.

Stouch T.R. and Jurs P.C., *Quant. Struct.-Act. Relat.*, **5** (1986) 57-61.

Streitweiser Jr A., *Molecular Orbital Theory for Organic Chemists*, Wiley, London, 1961.

Swain C.G., Unger S.H., Rosenquist N.R. and Swain M.S., *J. Am. Chem. Soc.*, **105** (1983) 492-502.

Taft R.W., In: *Steric Effects in Organic Chemistry*, ed. Newman M.S., Wiley, New York, 1956, pp. 556-675.

Taft R.W., Abboud J.-L.M., Kamlet M.J. and Abraham M.H. *J. Solution Chem.*, **14** (1985a) 153-86.

Taft R.W., Abraham M.H., Doherty R.M. and Kamlet M.J., *Nature*, **313** (1985b) 384-6.

Taft R.W., Gramstad T. and Kamlet M.J., *J. Org. Chem.*, **47** (1982) 4557-63.

Taft R.W., Gurka D., Joris L., Schleyer P. von R. and Rakshys J.W., *J. Am. Chem. Soc.*, **91** (1969) 4801-8.

Taft R.W., Kamlet M.J., *J. Am. Chem. Soc.*, **98** (1976) 2886-94.

Taft R.W., Kamlet M.J., *J. Chem. Soc. Perkin Trans. 2*, (1979) 1723-9.

Takahashi J., Kirino O., Takayama C., Nakamura S., Noguchi H., Kato T. and Kamoshita K., *Pestic Biochem. Phisiol.*, **30** (1988) 262-71.

Takayama C. and Fujinami A., *Pestic Biochem. Phisiol.*, **12** (1979) 163-71.

Tanaka M., Horisaka K., Yamagami C., Takao N. and Fujita T., *Chem. Pharm. Bull.*, **33** (1985) 2403-10.

Taylor R. and Kennard O., *J. Am. Chem. Soc.* **104** (1982) 5063-70.

Topliss J.G. and Costello R.J., *J. Med. Chem.*, **15** (1972) 1066-8.

Tsar User Guide Issue 4, Oxford Molecular Ltd., 1993.

Tsubomura H., *Bull. Chem. Soc. JPN.*, **27** (1954) 445-50.

Umeyama H. and Morokuma K., *J. Am. Chem. Soc.*, **98** (1976) 4400-4.

Umeyama H. and Morokuma K. and Yamabe S., *J. Am. Chem. Soc.*, **99** (1977) 330-43.

Unger S.H. and Hansch C., *J. Med. Chem.*, **16** (1973) 745-9.

Verloop A., Hoogenstraaten W. and Tipker J., in: *Drug Design, Vol 7*, ed. Ariens E.J., Academic Press, New York, 1976, pp. 165-207.

Vinter J.G., Davis A. and Saunders M.R., *J. Comput.-Aided Mol. Design*, **1** (1987) 31-51.

Vollhardt K.P.C., *Organic Chemistry*, W.H. Freeman and Co., New York, 1987, pp.864, 1191.

Weiner S.J., Kollman P.A., Case D.A., Singh U.C., Ghio C., Alagona G., Profeta S. and Weiner P., *J. Am. Chem. Soc.*, **106** (1984) 765-84.

Weiner P.K., Profeta S., Wipff G., Havel T., Kuntz I.D., Langridge R. and Kollman P.A., *Tetrahedron*, **39** (1983) 1113-21.

Wiberg K.B., *J. Am. Chem. Soc.*, **102** (1980) 1229-37.

Williams D.E. and Yan J.-M., *Adv. Atomic Mol. Phys.*, **23** (1988) 87-130.

Wold S., *Quant. Struct.-Act. Relat.*, **10** (1991) 191-3.

Wold S., Ruhe A., Wold H. and Dunn W.J., *III SIAM J. Sci. Stat. Comput.*, **5** (1984) 735-8.

Yamagami C., Sonoda C., Takao N., Tanaka M., Yamada J., Horisaka K and Fujita T., *Chem. Pharm. Bull.*, **30** (1982) 4175-80.

Yang G., Lien E.J. and Guo Z., *Quant. Struct.-Act. Relat.*, **5** (1986) 12-8.

Young R.C., Mitchel R.C., Brown T.H., Ganellin R., Griffiths R., Jones M., Rana K.K., Saunders D., Smith I.R., Sore N.E., Wilks T.J., *J. Med. Chem.*, **31** (1988) 656-71.

Zeegers-Huyskens Th., *J. Mol. Liquids*, **32** (1986a) 191-207.

Zeegers-Huyskens Th., *J. Mol. Struct. (Theochem)*, **135** (1986b) 93-103.

Poster Presentations:

Dearden J.C. and Ghafourian T.
A hydrogen bond parameter from molecular orbital theory: comparison with α values
Presented at: 130th British Pharmaceutical Conference, Reading, UK, September 1993.

Dearden J.C. and Ghafourian T.
Investigation of calculated hydrogen bonding parameters for QSAR
Presented at: 10th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling, Barcelona, Spain, September, 1994.

Publications:

Dearden J.C. and Ghafourian T., A hydrogen bond parameter from molecular orbital theory: comparison with α values, J. Pharm. Pharmacol., 45 (1993) 1143.

Dearden J.C. and Ghafourian T.
Investigation of calculated hydrogen bonding parameters for QSAR, In: QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications, eds. Sanz F. Giraldo J. and Manaut F., Prous Science Publishers, 1995, pp. 117-9.

Seminars:

School of Pharmacy Postgraduate Research Seminars, John Moores University, 1993, 1994 and 1995.