

Computer-aided molecular design of solvents for accelerated reaction kinetics

Heiko Struebing[†], Zara Ganase[†], Panagiotis G. Karamertzanis[†], Eirini Sioumkrou[†], Peter Haycock[‡],
Patrick M. Piccione[^], Alan Armstrong[‡], Amparo Galindo[†], Claire S. Adjiman^{†*}

[†] Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom. [‡] Department of Chemistry, Imperial College London, London SW7 2AZ, United Kingdom. [^] Process Studies Group, Syngenta, Jealott's Hill International Research Centre, Bracknell, Berkshire RG42 6EY, United Kingdom.

ABSTRACT: Solvents crucially alter the rates and selectivity of liquid-phase organic reactions, often hindering the development of new synthetic routes or, if chosen wisely, facilitating routes by improving rates and selectivities. To address this challenge, a systematic methodology is proposed that quickly identifies improved reaction solvents by combining quantum mechanical computations of the reaction rate constant in a few solvents with a computer-aided molecular design (CAMD) procedure. The approach allows the identification of a high-performance solvent within a very large set of possible molecules. The validity of our CAMD approach is demonstrated through application to a classical nucleophilic-substitution reaction for the study of solvent effects, the Menschutkin reaction. The results are successfully validated via in-situ kinetic experiments. A space of 1341 solvents is explored *in silico*, but requiring quantum mechanical calculations of the rate constant in only 9 solvents, and uncovering a solvent that increases the rate constant by 40%.

What is the best solvent for a given chemical reaction? Given that the rate and selectivity of chemical reactions can vary by several orders of magnitude in different solvents,^{1,2} this question has important ramifications for the exploration of novel reaction routes and the development of industrial processes.^{3,4} When investigating new liquid-phase reactions, it is essential to find a solvent that promotes the desired reaction without excessive catalyst deactivation, side-product formation, or solubility limitations. Indeed, a poor solvent choice can result in a missed opportunity to investigate novel chemistry or catalysts. At the process-development level, the problem of solvent choice is

compounded by the numerous safety, environmental and process constraints that must be satisfied. Yet, few tools exist to support this decision, especially when it affects reaction kinetics, and researchers are often left to choose on the basis of qualitative chemical knowledge and/or extensive and costly experimental investigations.

Advances in the understanding of liquid-phase reactions remain a topic of intense academic interest⁵ and practical relevance, as illustrated by identification of the development of solvent selection techniques as a key priority area by the ACS Green Chemistry roundtable.⁶ A very promising avenue of research in this direction is the development of Computer-Aided Molecular Design (CAMD) techniques. CAMD offers systematic methodologies, typically in the form of algorithms, to identify chemical species/molecular structures that perform a chosen function best (e.g., maximize the rate of a given reaction). The resulting molecular designs can be used to guide experiments in an otherwise huge space of possibilities. CAMD techniques have been widely applied in the context of solvent design for separations, and have had a significant impact on academic and industrial practice.⁷ In a batch extractive distillation for fine chemicals processing, for example, CAMD has been shown to identify a solvent that makes it possible to meet purity requirements, and that minimizes the time required for separation, with gains in throughput/productivity of at least 15%.⁸ The extension of CAMD techniques to reactions has the potential to facilitate the task of finding good solvents, and to uncover unexpected choices. The key objective in using such an approach is to identify choices of solvent molecules that offer good overall performance and meet the plethora of criteria that constrain solvent choice. These molecules may in some cases be novel, but the use of existing molecules makes the implementation of the results more straightforward. Thanks to the availability of simple predictive models, e.g., UNIFAC,⁹ many physical properties that affect reactions, such as the effect of solvent choice on solubility, can already be incorporated.⁴ However, accounting for chemical properties, such as the impact of solvents on reaction rate constants, remains challenging. A methodology for solvent design for enhanced reaction kinetics, QM-CAMD, is proposed here to address this issue. It relies on the integration of quantum mechanical (QM) rate-constant calculations into a CAMD framework.^{10,11} The approach allows the rapid exploration of a solvent design space consisting of thousands of

potential molecules and leads to a shortlist of promising solvents that can then be assessed experimentally.

Initial attempts to develop systematic approaches for the identification of the best reaction solvents were based on chemometrics,¹² and multivariate analysis.¹³⁻¹⁸ The use of these techniques is however limited by the need for large amounts of data to ensure statistical significance, and the fact that the information obtained for one reaction cannot be transferred reliably to another reaction, even within the same class.¹⁹ Some promising CAMD alternatives have recently been put forward.^{4,20} However, solvent effects on the kinetics of the reaction (via changes in the activation energy barrier) are either neglected or treated empirically within these approaches. An approach which incorporates kinetics as a final step of solvent design has also been proposed,²¹ but it requires the reaction rate to be computed quantum mechanically for every solvent that meets basic physical properties. Here, we fully integrate kinetics with other design considerations and we reduce the computational cost by using a less expensive “surrogate” (approximate) model that replaces the quantum mechanical calculations. This surrogate model is improved iteratively. As a result, only a small number of QM calculations are performed during the course of the QM-CAMD algorithm, although more than one thousand candidate solvents are considered.

Results

Systematic approach to solvent design. The overall QM-CAMD methodology is illustrated in Figure 1 and is presented here in the context of the identification of kinetic-enhancing solvents for a Menshutkin reaction.²² This class of S_N2 reaction has been studied extensively experimentally e.g.^{23,24} and computationally, e.g.,²⁵⁻²⁸ as the rate constant depends strongly on the solvent used. It provides an excellent test of a new methodology for solvent design, because there is sufficient knowledge available to test that the results generated by the computational design approach are physically and chemically meaningful. The specific reaction studied, shown in Figure 1, is that of phenacyl bromide **1** with pyridine **2** to form the phenacyl pyridinium bromide salt **3**. Its kinetics, which are second-order, have been investigated by a number of authors, e.g.,²⁹⁻³³.

Before describing the methodology and its application, we verify experimentally the validity of the computational approach used in Step 2 to obtain liquid-phase reaction rate constants, based on

conventional transition state theory (CTST)^{34,35} and a continuum solvation model, as accurate calculations remain challenging.³⁶ The SMD continuum solvation model³⁷ is known to predict energies of solvation well, although larger deviations can be noted for alcohols. Here we compute reaction rate constants using SMD and QM density functional theory (DFT), using B3LYP/6-31+G(d), and compare them to values obtained using ¹H NMR *in situ* kinetic experiments in five aprotic solvents (toluene, chloroform, tetrahydrofuran (THF), acetone and acetonitrile), as well as from conductance experiments.^{32,33} The measured and predicted rate constants (Figure 2) were found to be within one order of magnitude of each other. Perhaps more importantly, the trend (the ranking of solvents in order of increasing rate constant) is predicted well by the computational method, despite the quantitative differences. This indicates that the DFT+SMD approach chosen is appropriate to identify solvents which promote the rate of the selected reaction. Although other models can yield better agreement with experiments (e.g., M05-2X/6-31G(d), see SI), the QM-CAMD approach does not require high accuracy to generate useful results, provided that qualitative trends are reliable. The effect of model uncertainty was studied extensively¹¹ on a precursor of the approach based on experimental rate constants¹⁰ and showed that consistent solvent choices can be obtained for different values of the uncertain model parameters.

Given the suitability of DFT calculations, the challenge that remains to be addressed with the QM-CAMD methodology is the use of this technique to explore a large space of possible solvents without explicitly carrying out thousands of expensive DFT calculations. In the QM-CAMD approach, as in many CAMD methods, a “solvent” is defined by its molecular composition in terms of atom groups of varying size such as CH₃, CH₂, OH, or CH₂NO₂. Although this is a coarse description of molecular structure, there exists a broad class of well-established predictive techniques, known as group-contribution (GC) methods, which can predict with good accuracy the bulk properties of thousands of molecules from knowledge of the properties of a much smaller number of groups.³⁸ By building on this and using the continuum solvation concept, we ensure that no experimental input is needed during the course of the QM-CAMD algorithm. Furthermore, by using GC methods, we include existing as well as entirely novel solvent molecules in the design space, allowing truly innovative solutions to

emerge. The chosen atom groups and physicochemical constraints, such as valency and physical property ranges, are the only limits on the design space.

In Step 1 of the QM-CAMD algorithm (cf. Figure 1), the design objective and constraints are stated. Here, the objective is to maximize the reaction rate constant k at 298 K, and the only design constraint is that the solvent designed should be liquid at this temperature. This is achieved by setting bounds on its normal boiling and melting points, as predicted by group contribution techniques.³⁹ An initial set of six diverse solvents, with varying polarity and functional groups, is also required. Here, we use the first six solvents listed in Table 1 in order of increasing dielectric constant.⁴⁰

In Step 2, the reaction rate constants at 298 K in specific solvents are computed based on CTST using DFT+SMD. The bulk solvent properties required as input to the SMD model, namely the Abraham acidity, basicity and polarity descriptors, the macroscopic surface tension and the dielectric constant, are predicted using the group contribution concept.⁴¹

In Step 3, a surrogate model is derived from the QM rate constants, k^{QM} . A surrogate model is a simplified model that must have two key properties: (i) it must be applicable to a very large number of solvents; (ii) it must be inexpensive to evaluate, thereby alleviating the computational cost of embedding QM calculations directly in the solvent design algorithm.⁴² Here, we use the solvatochromic equation,^{43,44} a well-recognized and widely used linear free energy relationship that can be applied to a range of properties, including reaction rate constants. The solvatochromic equation used relates the logarithm of the rate constant k to five properties of the solvent via six coefficients that are characteristic of the reaction being investigated:

$$\log k = c_0 + c_A A + c_B B + c_S S + c_\delta \delta + c_H \delta_H^2,$$

where A , B , S , δ and δ_H^2 are solvent properties. A , B and S are the Abraham descriptors (hydrogen bond acidity, hydrogen bond basicity and polarizability/dipolarity, respectively)⁴⁵⁻⁴⁷, δ is a chemical class indicator, specifying the presence of aromatic or halogen atoms and δ_H^2 is the cohesive energy density of the solvent (the square of its Hildebrand solubility parameter).⁴⁴ The five solvent properties are computed using predictive GC techniques.^{10,39,41} The GC techniques used here are simple equations

that have been developed on the basis of extensive experimental data sets and that allow the prediction of the five relevant properties for solvents for which no data are available. The average absolute errors in the GC methods are low, with values 0.017 for A , and 0.043 for B , 0.065 for S (A , B , S range from 0 to 1.2) and a value of 1.13 MPa^{0.5} for δ_H , which ranges from 10 to 50 MPa^{0.5} for a wide range of potential solvents. The GC method for calculating δ is exact. The reaction-specific coefficients c_i , $i=0, A, B, S, \delta, H$, are obtained by linear regression. This can be based on experimentally-determined rate constants in a few solvents e.g.^{10,11} or, as is done here, based on a set of calculated k^{QM} values.

Every time the algorithm passes through Step 3, the reaction-specific coefficients, c_i , are re-estimated using all known k^{QM} values at that point. At the first iteration only 6 values are available, and one value is added at each subsequent iteration, corresponding to the latest new solvent identified. Thus, the number of data points used in the regression of the coefficients grows as the algorithm progresses and leads to an increasingly accurate surrogate model.

In Step 4, the solvent design problem is posed as a mathematical optimization problem in which the molecular structure of the solvent that maximizes the reaction rate constant is sought. The problem formulation follows the standard structure of optimization-based CAMD problems.⁷ The solvent's molecular structure, as defined by the atom groups it contains, is represented by integer variables denoting the number of functional groups of different types (e.g., CH₃, CH₂, OH). A set of chemical feasibility and complexity constraints is imposed to ensure that only chemically permissible combinations of the functional groups are generated when solving the optimization problem. In addition, a set of constraints linking molecular structure to physical properties is included. This consists of the GC techniques for the relevant solvent properties (A , B , S , δ , δ_H^2 , melting and boiling points)^{10,39,41} and the solvatochromic equation for the reaction rate constant. The problem solved here is linear and contains both continuous and binary (0-1) variables, making it a mixed-integer linear problem (MILP) that can readily be solved with standard solvers. Due to the use of the solvatochromic equation, the MILP is an approximation of the full, nonlinear, design problem, which may have multiple local solutions; the QM-CAMD approach can thus identify a high-performance

solvent but this solution cannot be guaranteed to be the very best one. The likelihood of finding the very best solution can be increased by using different sets of initial solvents.

For this work, we use a set of 38 functional groups, listed in Supplementary Table 1. This set includes a wide range of functional groups, enabling the design of non-polar, polar aprotic and protic solvents. It yields a design space consisting of 1341 possible combinations of the atom groups, within the limits imposed on molecular complexity (e.g., maximum number of atom groups in the solvent, maximum number of certain functional groups). Some of the possible group combinations represent more than one solvent molecule, but differences in the physical properties of the isomeric solvents are not taken into account in the computations and can be investigated in the post-design phase of the approach. A candidate solvent, i.e., the molecular structure of the best solvent in the design space, for the current solvatochromic equation, is found by the MILP algorithm.

In Step 5, a convergence criterion is applied. If the candidate solvent from Step 4 has not previously been identified, the algorithm returns to Step 2, where the reaction rate constant for the best solvent found in Step 4 is computed with the DFT+SMD approach and added to the list of k^{OM} values. This additional information is to be used to regress new coefficients for the solvatochromic equation. As iterations proceed, the statistical significance of the coefficients increases and the solvatochromic equation becomes increasingly reliable. If the termination criterion is met, the algorithm proceeds to Step 6.

In Step 6, the solvent with the largest value of k^{OM} among the values computed is checked against any criteria not explicitly included in the design problem formulation (e.g., chemical stability or reactivity). If it meets these criteria it is reported as a high-performance solvent and the results are verified via in-situ kinetic experiments. Otherwise, the next best solvent among those identified during the course of the algorithm, and which meets any additional criteria, is reported.

Application to a bimolecular reaction. To apply the proposed QM-CAMD methodology to the reaction scheme in Figure 1, the initial set of six solvents used in Step 2 consists of toluene, chlorobenzene, ethyl acetate, THF, acetone and acetonitrile. The computed rate constants and the solvent dielectric constants are shown in Table 1. The final iteration of the algorithm is also shown in the Table. At every iteration, the solvatochromic equation is updated in Step 3 (cf. Supplementary

Information for details of each iteration). As additional k^{QM} values are included in the regression, the rate constant values predicted by the solvatochromic equation become increasingly reliable and the algorithm converges after five iterations, with nitromethane reported as the best solvent. This implies that a solvent space of over 1341 molecules has been explored with only nine k^{QM} calculations (six in iteration 1, one in iterations 2, 3 and 4, respectively, and none in the final iteration). Given that each k^{QM} calculation requires the optimization of the structure and free energies of the two reactants and of the transition state, and that the CAMD problem of Step 4 with the use of the solvatochromic equation is solved in less than 1 CPU second on an Intel Xeon 1.60 GHz CPU, the proposed approach allows a very efficient exploration of the design space, identifying a high-performance solvent in a matter of hours.

The best solvent identified by the algorithm is nitromethane. The same experimental procedure as previously is carried out to determine the reaction rate constant in deuterated nitromethane; it is found to be $3.4 \cdot 10^{-3} \text{ dm}^3 \text{ mol}^{-1} \text{ s}^{-1}$ at 298 K. This is an improvement of 40% over the measured rate constant for acetonitrile, the best initial solvent. Although nitromethane is an established solvent, it is unlikely to be prioritised over more common dipolar aprotic solvents for inclusion in experimental screening. Given that nitromethane and acetonitrile have very similar dielectric constant (Table 1), the QM-CAMD approach has played a crucial role in recognising the superior performance of nitromethane. The benefit of using nitromethane instead of one of the initial solvents in a batch reactor is illustrated in Figure 3, in terms of product concentration, for an initial concentration of 0.4 mol dm^{-3} for **1** and of 0.8 mol dm^{-3} for **2**.

Discussion

The QM-CAMD methodology presented here leads to the rapid identification of a solvent that accelerates the reaction kinetics. Its application to a classic reaction for the study of solvent effects, the Menschutkin reaction, has shown that it is indeed able to achieve the objective of finding a solvent that increases the reaction rate constant, among a large design space of over 1300 solvents. Only nine quantum mechanical evaluations are required, hence making the methodology computationally affordable. The enhancement predicted in terms of the new solvent has been confirmed experimentally via detailed in-situ ^1H NMR kinetic experiments. The proposed computational

approach can help to place solvent selection on a more systematic footing, reducing the amount of experimental work required and allowing a quantitative evaluation of rates in different solvents, allowing solvents with identical dielectric constants, such as acetonitrile and nitromethane, to be assessed. The optimization framework at the core of the approach is ideally suited to tackle more complex cases, including solvent effects on selectivity and the inclusion of other important factors in reactor design such as solubility and toxicity, which make the identification of reaction solvents more challenging.^{4,48,49} This promising QM-CAMD methodology can be further refined, e.g., by introducing an improved surrogate model, and applied to challenging problems in organic chemistry. One particularly promising application area is for reaction systems with competing, yet similar, reactions.² In such cases, an understanding of the reaction mechanism is not sufficient to identify solvents in which the reaction of interest is fast, but where the side reactions proceed at a slow or negligible rate. Instead, existing mechanistic knowledge could readily be used within the proposed QM-CAMD approach to identify solvents that promote the desired reaction pathway and increase selectivity.

Methods

Experimental methods. To determine the reaction rate constants, four experiments at 298 K with different initial concentrations were carried out for each deuterated solvent, using trimethoxybenzene as an internal standard. NMR spectra were recorded on a Bruker AV500 (¹H 500 MHz). To ensure that the data collected was quantitative, T₁-relaxation times were determined and the time between rf pulses was set to be greater than five times the longest T₁ value. The use of four runs ensures that the values obtained for the measured rate constants are statistically significant and leads to very tight confidence intervals on the regressed rate constants (two to three orders of magnitude smaller than the rate constants). In each experiment, sufficient time was allowed to reach a conversion of at least 50%. As many peaks as available from the ¹H NMR spectra were used to extract the experimental rate constant. In all cases this included at least two of the 4 phenacyl bromide peaks (3 aromatic peaks and 1 CH₂ peak), at least one of the 3 pyridine peaks and as many of the 7 product peaks available as possible (except in THF and toluene where the product was insoluble and thus could not be followed by liquid-state NMR). Molar concentrations for the two reactants and, when feasible, for the product

were obtained from the measured peaks and internal standard concentrations. The molar concentrations were then used to determine the rate constants, based on second-order kinetics, i.e.,

$$\frac{d[\mathbf{1}]}{dt} = -k[\mathbf{1}][\mathbf{2}],$$

$$\frac{d[\mathbf{2}]}{dt} = -k[\mathbf{1}][\mathbf{2}],$$

$$\frac{d[\mathbf{3}]}{dt} = k[\mathbf{1}][\mathbf{2}],$$

where t is time, k is the rate constant in $\text{dm}^3 \text{mol}^{-1} \text{s}^{-1}$, $[\mathbf{i}]$ denotes the concentration of species \mathbf{i} (understood as a function of time). As an example, the initial concentrations used for $\mathbf{1}$ and $\mathbf{2}$ for the reaction in nitromethane-d₃ are shown in Supplementary Table 2. The initial product concentration is always set to 0. The estimates of the rate constant were obtained by applying nonlinear least squares parameter estimation within the gPROMS software (<http://www.psenterprise.com>). A high quality of fit was achieved in all cases, as illustrated in Supplementary Figs 1 to 3.

Computational methods. The overall approach to computing reaction rate constants in liquid phase from a basic description of the solvent in terms of its composition in atom groups is illustrated in Supplementary Fig 4. The rate constant k^{QM} , is computed using transition state theory, according to the following equation, which applies to the second-order Menshutkin reaction:

$$k^{QM} = \kappa \frac{k_B T}{h} \frac{1}{c^{o,L}} \left(\frac{q_{(AB)^\ddagger}^{o,IG}}{q_A^{o,IG} q_B^{o,IG}} \right) \exp\left(\frac{-\Delta^\ddagger E^{el}}{RT}\right) \exp\left(\frac{-\Delta^\ddagger \Delta G^{o,solv}}{RT}\right),$$

where κ is the transmission coefficient, calculated according to the Wigner tunnelling correction factor, k_B is the Boltzmann constant, T is the temperature, h is the Planck constant, $c^{o,L}$ is the standard state liquid phase concentration (here, 1 mol dm^{-3}), $q_i^{o,IG}$ is the ideal gas partition function, (excluding the electronic energy as is standard practice) for component i , where $i = A$ (reactant A), B (reactant B) or $(AB)^\ddagger$ (transition state), R is the gas constant, E^{el} is the gas phase electronic energy, $\Delta G^{o,solv}$ is the standard state free energy of solvation and Δ^\ddagger indicates that an activation energy barrier is being calculated, i.e.,

$$\Delta^\ddagger E^{el} = E_{(AB)^\ddagger}^{el} - E_A^{el} - E_B^{el},$$

$$\Delta^\ddagger \Delta G^{o,solv} = \Delta G_{(AB)^\ddagger}^{o,solv} - \Delta G_A^{o,solv} - \Delta G_B^{o,solv}.$$

The gas phase minimum energy structures for the reactants, the gas phase transition state structures and the gas phase partition functions are computed with Gaussian 09,⁵⁰ using the B3LYP/6-31+G(d) functional. For the calculation of the free energies of solvation, the structures are re-optimized using the same functional, the SMD continuum solvation model³⁷ and intrinsic Coulomb radii. Gas phase vibrational frequencies are used. The bulk solvent properties required in the SMD model are obtained using group contribution methods.⁴¹ All structures obtained are reported in the Supplementary Information, together with further details of the computational methodology and justification of the choices made.

The MILPs are solved by the CPLEX solver, accessed via the GAMS software

(<http://www.gams.com>).

References

- 1 Reichardt, C. & Welton, T. *Solvents and Solvent Effects in Organic Chemistry*. (Wiley-VCH, 2011).
- 2 Maki, B. E., Patterson, E. V., Cramer, C. J. & Scheidt, K. A. Impact of solvent polarity on N-heterocyclic carbene-catalyzed α -protonations of homoenolate equivalents. *Org. Lett.* **11**, 3942-3945 (2009).
- 3 Carlson, R. Designs for explorative experiments in organic synthetic chemistry. *Chemometr. Intell. Lab.* **73**, 151-166 (2004).
- 4 Gani, R., Jiménez-González, C. & Constable, D. J. C. Method for selection of solvents for promotion of organic reactions. *Comput. Chem. Eng.* **29**, 1661-1676 (2005).
- 5 Otto, R. *et al.* Single solvent molecules can affect the dynamics of substitution reactions. *Nature Chem.* **4**, 534-538 (2012).
- 6 Jiménez-González, C. *et al.* Key green engineering research areas for sustainable manufacturing: A perspective from pharmaceutical and fine chemicals manufacturers. *Org. Process Res. Dev.* **15**, 900-911 (2011).
- 7 Achenie, L. E. K., Gani, R. & Venkatasubramanian, V. *Computer Aided Molecular Design : Theory and Practice*. (Elsevier, 2003).
- 8 Giovanoglou, A., Barlatier, J., Adjiman, C. S., Pistikopoulos, E. N. & Cordiner, J. L. Optimal solvent design for batch separation based on economic performance. *AIChE J.* **49**, 3095-3109 (2003).
- 9 Fredenslund, A., Jones, R. L. & Prausnitz, J. M. Group-contribution estimation of activity-coefficients in nonideal liquid-mixtures. *AIChE J.* **21**, 1086-1099 (1975).
- 10 Folić, M., Adjiman, C. S. & Pistikopoulos, E. N. Design of solvents for optimal reaction rate constants. *AIChE J.* **53**, 1240-1256 (2007).
- 11 Folić, M., Adjiman, C. S. & Pistikopoulos, E. N. Computer-aided solvent design for reactions: Maximizing product formation. *Ind. Eng. Chem. Res.* **47**, 5190-5202 (2008).
- 12 Carlson, R. & Carlson, J. E. *Design and Optimization in Organic Synthesis*. (Elsevier, 2005).
- 13 Koppel, I. A. & Palm, V. A. in *Advances in Linear Free Energy Relationships* (eds N. B. Chapman & J. Shorter) Ch. 5, (Plenum Press, 1972).
- 14 Kamlet, M. J., Abboud, J. L. & Taft, R. W. The solvatochromic comparison method. 6. The π^* scale of solvent polarities. *J. Am. Chem. Soc.* **99**, 6027-6038 (1977).

- 15 Taft, R. W., Pienta, N. J., Kamlet, M. J. & Arnett, E. M. Linear solvation energy relationships. 7. Correlations between the solvent-donicity and acceptor-number scales and the solvatochromic parameters, ρ^* , ρ , and σ . *J. Org. Chem.* **46**, 661-667 (1981).
- 16 Taft, R. W., Abboud, J. L. M. & Kamlet, M. J. Linear solvation energy relationships. 12. The δ term in the solvatochromic equations. *J. Am. Chem. Soc.* **103**, 1080-1086 (1981).
- 17 Kamlet, M. J., Abboud, J.-L. M., Abraham, M. H. & Taft, R. W. Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, ρ^* , ρ and σ , and some methods for simplifying the generalized solvatochromic equation. *J. Org. Chem.* **48**, 2877-2887 (1983).
- 18 Taft, R. W., Abboud, J.-L. M., Kamlet, M. J. & Abraham, M. H. Linear solvation energy relations. *J. Solution Chem.* **14**, 153-186 (1985).
- 19 Bini, R., Chiappe, C., Mestre, V. L., Pomelli, C. S. & Welton, T. A rationalization of the solvent effect on the Diels-Alder reaction in ionic liquids using multiparameter linear solvation energy relationships. *Org. Biomol. Chem.* **6**, 2522-2529 (2008).
- 20 Gani, R., Gómez, P. A., Folić, M., Jiménez-González, C. & Constable, D. J. C. Solvents in organic synthesis: Replacement and multi-step reaction systems. *Comput. Chem. Eng.* **32**, 2420-2444 (2008).
- 21 Stanescu, I. & Achenie, L. E. K. A theoretical study of solvent effects on Kolbe-Schmitt reaction kinetics. *Chem. Eng. Sci.* **61**, 6199-6212 (2006).
- 22 Menshutkin, N. A. Über die Affinitätskoeffizienten der Alkylhaloide und der Amine. *Z. Physik. Chem.* **6**, 41-57 (1890).
- 23 Abraham, M. H. & Grellier, P. L. Substitution at saturated carbon. Part XX. The effect of 39 solvents on the free energy of Et_3N , EtI , and the $\text{Et}_3\text{N-EtI}$ transition state. Comparison with solvent effects on the equilibria $\text{Et}_3\text{N} + \text{EtI} \rightleftharpoons \text{Et}_4\text{N}^+\text{I}^-$ and $\text{Et}_3\text{N} + \text{EtI} \rightleftharpoons \text{Et}_4\text{N}^+ + \text{I}^-$. *J. Chem. Soc. Perk. T. 2*, 1735-1741 (1976).
- 24 Lassau, C. & Jungers, J. L'influence du solvant sur la réaction chimique. La quaternation des amines tertiaires par l'iodure de méthyle. *Bull. Soc. Chim. Fr.* **7**, 2678-2685 (1968).
- 25 Castejon, H. & Wiberg, K. B. Solvent effects on methyl transfer reactions. 1. The Menshutkin Reaction. *J. Am. Chem. Soc.* **121**, 2139-2146 (1999).
- 26 Acevedo, O. & Jorgensen, W. L. Exploring solvent effects upon the Menshutkin reaction using a polarizable force field. *J. Phys. Chem. B* **114**, 8425-8430 (2010).
- 27 Chuang, Y.-Y., Cramer, C. J. & Truhlar, D. G. The interface of electronic structure and dynamics for reactions in solution. *Int. J. Quantum Chem.* **70**, 887-896 (1998).
- 28 Su, P., Wu, W., Kelly, C. P., Cramer, C. J. & Truhlar, D. G. VBSM: A solvation model based on valence bond theory. *J. Phys. Chem. A* **112**, 12761-12768 (2008).
- 29 Pearson, R. G., Langer, S. H., Williams, F. V. & McGuire, W. J. Mechanism of the reaction of α -haloketones with weakly basic nucleophilic reagents. *J. Am. Chem. Soc.* **74**, 5130-5132 (1952).
- 30 Barnard, P. W. C. & Smith, B. V. The Menshutkin reaction - a group experiment in a kinetic-study. *J. Chem. Educ.* **58**, 282-285 (1981).
- 31 Forster, W. & Laird, R. M. The mechanism of alkylation reactions. 1. The effect of substituents on the reaction of phenacyl bromide with pyridine in methanol. *J. Chem. Soc. Perk. T. 2*, 135-138 (1982).
- 32 Hwang, J. U., Chung, J. J., Yoh, S. D. & Jee, J. G. Kinetics for the reaction of phenacyl bromide with pyridine in acetone under high-pressure. *B. Kor. Chem. Soc.* **4**, 237-240 (1983).
- 33 Halvorsen, A. & Songstad, J. The reactivity of 2-bromo-1-phenylethanone (phenacyl bromide) toward nucleophilic species. *J. Chem. Soc., Chem. Commun.*, 327-328 (1978).
- 34 Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107-115 (1935).
- 35 Evans, M. G. & Polanyi, M. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.* **31**, 875-894 (1935).
- 36 Jalan, A., Ashcraft, R. W., West, R. H. & Green, W. H. Predicting solvation energies for kinetic modeling. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **106**, 211-258 (2010).

- 37 Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378-6396 (2009).
- 38 Poling, B. E., Prausnitz, J. M. & O'Connell, J. P. *The Properties of Gases and Liquids*. (McGraw-Hill, 2000).
- 39 Marrero, J. & Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibr.* **183-184**, 183-208 (2001).
- 40 Winget, P., Dolney, D. M., Giesen, D. J., Cramer, C. J. & Truhlar, D. G. *Minnesota Solvent Descriptor Database*, <<http://comp.chem.umn.edu/solvation/mnsddb.pdf>> (2010).
- 41 Sheldon, T. J., Adjiman, C. S. & Cordiner, J. L. Pure component properties from group contribution: Hydrogen-bond basicity, hydrogen-bond acidity, Hildebrand solubility parameter, macroscopic surface tension, dipole moment, refractive index and dielectric constant. *Fluid Phase Equilibr.* **231**, 27-37 (2005).
- 42 Sheldon, T. J., Folić, M. & Adjiman, C. S. Solvent design using a quantum mechanical continuum solvation model. *Ind. Eng. Chem. Res.* **45**, 1128-1140 (2006).
- 43 Abraham, M. H., Doherty, R. M., Kamlet, M. J., Harris, J. M. & Taft, R. W. Linear solvation energy relationships. Part 37. An analysis of contributions of dipolarity-polarisability, nucleophilic assistance, electrophilic assistance, and cavity terms to solvent effects on t-butyl halide solvolysis rates. *J. Chem. Soc. Perk. T. 2*, 913-920 (1987).
- 44 Abraham, M. H., Doherty, R. M., Kamlet, M. J., Harris, J. M. & Taft, R. W. Linear solvation energy relationships. Part 38. An analysis of the use of solvent parameters in the correlation of rate constants, with special reference to the solvolysis of t-butyl chloride. *J. Chem. Soc. Perk. T. 2*, 1097-1101 (1987).
- 45 Kamlet, M. J., Abraham, M. H., Doherty, R. M. & Taft, R. W. Solubility properties in polymers and biological media. 4. Correlation of octanol/water partition coefficients with solvatochromic parameters. *J. Am. Chem. Soc.* **106**, 464-466 (1984).
- 46 Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **22**, 73-83 (1993).
- 47 Abraham, M. H. Application of solvation equations to chemical and biochemical processes. *Pure Appl. Chem.* **65**, 2503-2512 (1993).
- 48 Gao, C., Govind, R. & Tabak, H. H. Application of the group contribution method for predicting the toxicity of organic chemicals. *Environ. Toxicol. Chem.* **11**, 631-636 (1992).
- 49 Stefanis, E., Constantinou, L. & Panayiotou, C. A group-contribution method for predicting pure component properties of biochemical and safety interest. *Ind. Eng. Chem. Res.* **43**, 6253-6261 (2004).
- 50 Frisch, M. J. *et al. Gaussian09 Revision A.1.* (Gaussian Inc, 2009)

Acknowledgments

The authors gratefully acknowledge funding from the Engineering and Physical Sciences Research Council (EPSRC) (EP/E016340, EP/J003840/1, EP/J014958/1); funding from EPSRC & Syngenta (CASE award); access to computational resources and support from the High Performance Computing Cluster at Imperial College London. We are thankful to Donna Blackmond for valuable discussions.

Author contributions

H.S. contributed to the QM-CAMD methodological development and carried out all computations, with the help of P.G.K. and E.S., Z.G. performed and analyzed all the experiments, with the help of P.H. for the NMR measurements, P.M.P. provided input for the interpretation of the kinetic data,

A.A., A.G. and C.S.A. designed the methodology, supervised the project and prepared the manuscript.

All authors edited the manuscript.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permission information is available online at <http://www.nature.com/reprints>. Correspondence should be addressed to C.S.A.: c.adjiman@imperial.ac.uk

Competing financial interests

The authors declare no competing financial interests.

List of figures

Figure 1 – Computer-aided molecular design algorithm, illustrated for a Menshutkin reaction. **a**, The reaction studied, combining phenacyl bromide (**1**) and pyridine (**2**) to produce phenacyl pyridinium bromide (**3**). **b**, The QM-CAMD solvent design algorithm. **c**, The free energy curve illustrates the calculation of the rate constant in Step 2, with the transition state structure corresponding to the top of the curve. **d**, The concentration vs time plot shows experimental data obtained via ^1H NMR in-situ monitoring of the Menshutkin reaction in nitromethane: green triangles: pyridine; red squares: phenacyl bromide; purple crosses: phenacyl pyridinium bromide; blue diamonds: mole balance (sum of phenacyl bromide and phenacyl pyridinium bromide concentrations).

Figure 2 - Comparison of QM-calculated and experimental rate constants in several solvents. The calculations and measurements show good qualitative agreement, making the rate prediction method suitable as a basis for solvent design. The QM calculations (first bar, dark blue) are performed using B3LYP/6-31+G(d) and SMD. The experimental rate constants determined in this work (second bar, red) are obtained from ^1H NMR in situ experiments. The calculated confidence intervals for these data are two to three orders of magnitude smaller than the measured rate constants. Rate constants determined by conductance measurements in other work (third bar, light blue. Acetone³²; Acetonitrile³³) are also reported for comparison, highlighting the difficulty in obtaining reliable kinetic data in solution. All values are obtained at 298 K.

Figure 3 – Time evolution of product concentration in different solvents. The calculated concentration of the product, **3**, is plotted as a function of time in a batch reactor for different solvents at 298 K, based on experimental rate constants and initial concentration of **1** of 0.4 mol dm^{-3} and of **2** of 0.8 mol dm^{-3} .

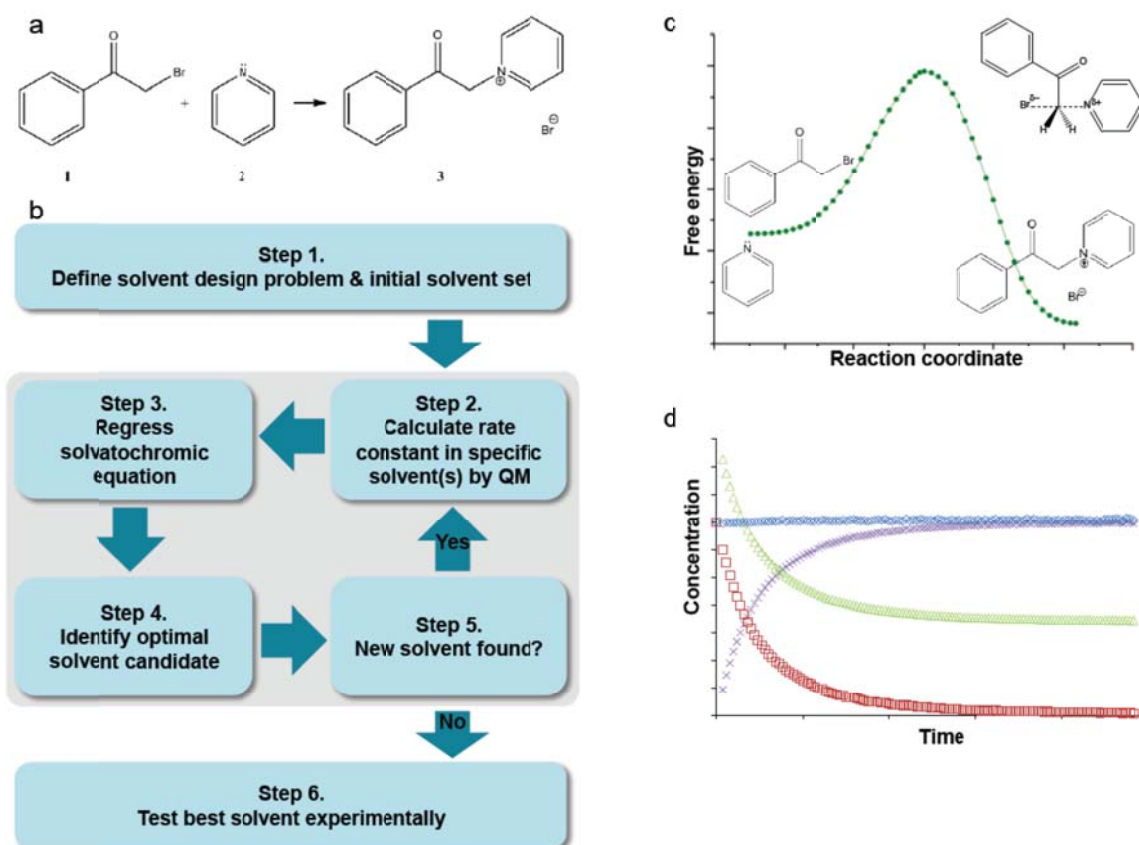


Figure 2 – Computer-aided molecular design algorithm, illustrated for a Menshutkin reaction. a, The reaction studied, combining phenacyl bromide (1) and pyridine (2) to produce phenacyl pyridinium bromide (3). b, The QM-CAMD solvent design algorithm. c, The free energy curve illustrates the calculation of the rate constant in Step 2, with the transition state structure corresponding to the top of the curve. d, The concentration vs time plot shows experimental data obtained via ^1H NMR in-situ monitoring of the Menshutkin reaction in nitromethane: green triangles: pyridine; red squares: phenacyl bromide; purple crosses: phenacyl pyridinium bromide; blue diamonds: mole balance (sum of phenacyl bromide and phenacyl pyridinium bromide concentrations).

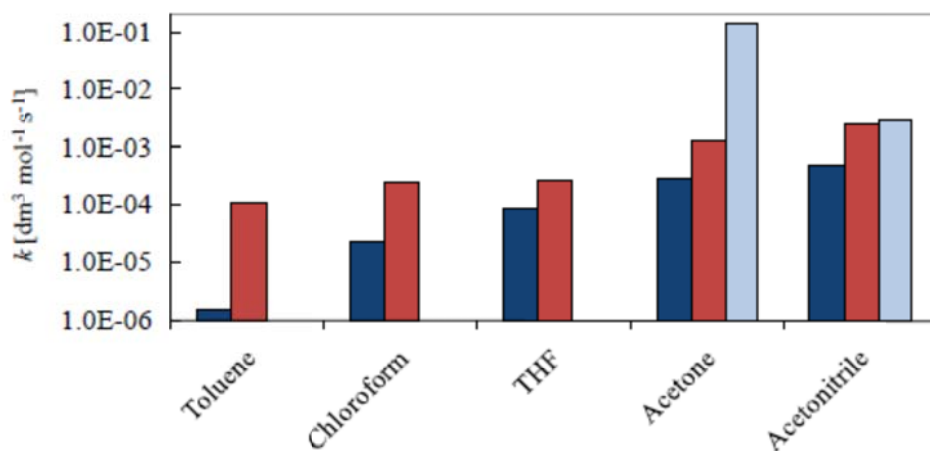


Figure 2 - Comparison of QM-calculated and experimental rate constants in several solvents. The calculations and measurements show good qualitative agreement, making the rate prediction method suitable as a basis for solvent design. The QM calculations (first bar, dark blue) are performed using B3LYP/6-31+G(d) and SMD. The experimental rate constants determined in this work (second bar, red) are obtained from ^1H NMR in situ experiments. The calculated confidence intervals for these data are two to three orders of magnitude smaller than the measured rate constants. Rate constants determined by conductance measurements in other work (third bar, light blue. Acetone³²; Acetonitrile³³) are also reported for comparison, highlighting the difficulty in obtaining reliable kinetic data in solution. All values are obtained at 298 K.

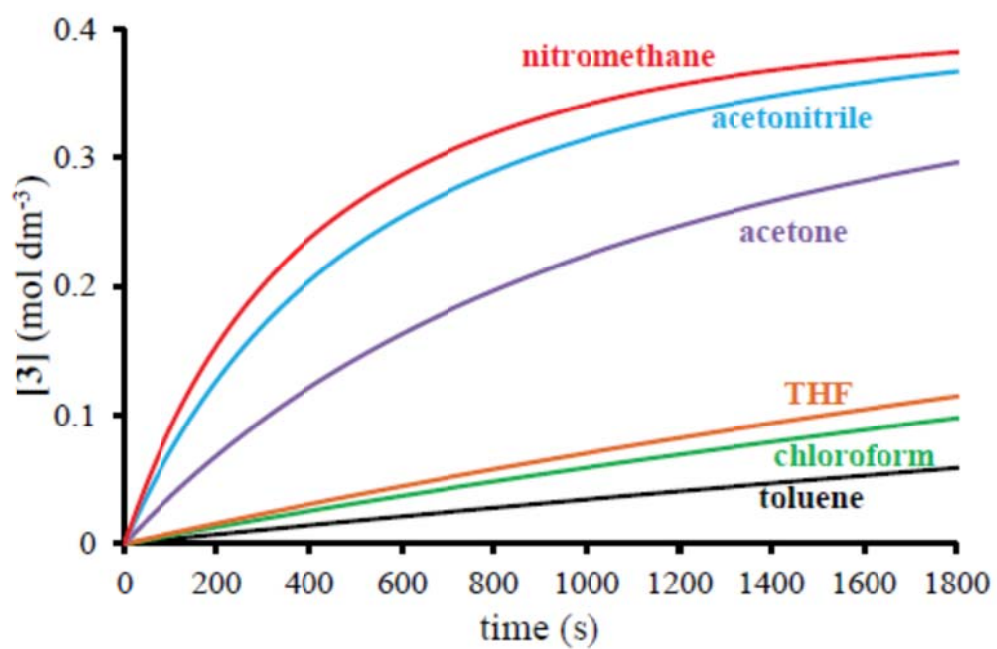


Figure 3 – Time evolution of product concentration in different solvents. The calculated concentration of the product, **3**, is plotted as a function of time in a batch reactor for different solvents at 298 K, based on experimental rate constants and initial concentration of **1** of 0.4 mol dm⁻³ and of **2** of 0.8 mol dm⁻³.

Tables

Table 1. Key properties for the six initial solvents and the best solvent identified, nitromethane.

Solvent	Dielectric constant [*]	Rate constant k^{QM} ^{**} ($\text{dm}^3 \text{mol}^{-1} \text{s}^{-1}$)
Toluene	2.38	$1.61 \cdot 10^{-6}$
Chlorobenzene	5.70	$3.94 \cdot 10^{-5}$
Ethyl acetate	5.99	$4.88 \cdot 10^{-5}$
THF	7.43	$8.70 \cdot 10^{-5}$
Acetone	20.49	$2.86 \cdot 10^{-4}$
Acetonitrile	35.69	$4.70 \cdot 10^{-4}$
<i>Nitromethane</i>	<i>36.56</i>	<i>$6.09 \cdot 10^{-4}$</i>

^{*} Experimental dielectric constant at 298 K⁴⁰

^{**} Rate constants calculated by QM (B3LYP/6-31+G(d)) at 298 K