J Comput Aided Mol Des (2010) 24:605–611
DOI 10.1007/s10822-010-9342-8

Tautomers and reference 3D-structures: the orphans of *in silico* drug design

Timothy Clark

Received: 28 January 2010 / Accepted: 16 March 2010 / Published online: 27 March 2010

Abstract The importance of calculating not only the correct tautomer, but also the correct protonation state and conformation in 3D modeling applications is emphasized. Above all, identifying and characterizing the most stable form of a ligand under physiological conditions is seen to be the key to successful 3D modeling. Modeling strategies that make use of the performance of modern hardware can employ physically more appropriate models than most currently in use and still be easily applicable to large numbers of compounds. Because the performance of quantitative structure–property relationships is likely to be limited by the available training and validation data, we must either find new sources of such data or resort to explicit modeling, which can partly be parameterized using definitive ab initio calculations for reference data such as gas-phase proton affinities.

Keywords Tautomers · QSPR · Force fields · pK_a

Introduction

The first sentence of the abstract of a 2003 article on tautomers says it all:

> Tautomers are often disregarded in computer-aided molecular modeling applications. Little is known about the different tautomeric states of a molecule and they are rarely registered in chemical databases [1].

This was published 6 years before this perspective was written. Note that the comment refers to tautomerism in cheminformatics. The experimental investigation of tautomerism, particularly of heterocyclic compounds, has long been well established [2, 3], but this does not necessarily mean that an adequate basis for knowledge-based cheminformatics exists, as will be discussed below. One might think that the situation described in the above quotation would start alarm bells ringing throughout the drug-design community, but the reaction has been strictly limited [4–6]. Some tools exist [6–8] for predicting the correct tautomer of a given compound (i.e., the one that predominates in neutral aqueous solution), but the subject of reliable tautomer prediction is only now gradually moving to the forefront of our concerns. This is most likely because predicting tautomers is difficult and because 2D-techniques have been getting along just fine without considering tautomers explicitly. It is hardly necessary to point out here that tautomerism can turn H-bond donors into acceptors and vice versa. However, it can also change physical properties [6], conformations in solution, membrane permeability, solubility, distribution and many other properties close to our hearts. Above all, an energetically competitive tautomer that is not the major form in solution may be the one that binds in the specific environment of a receptor (and the same molecule may bind as another tautomer in another receptor).

History has tried to teach us about tautomers. Watson and Crick only made their breakthrough in deriving the structure of the double helix after Jerry Donohue told them

T. Clark (✉)
Computer-Chemie-Centrum, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 25, 91052 Erlangen, Germany
e-mail: Tim.Clark@chemie.uni-erlangen.de;
Tim.Clark@port.ac.uk

T. Clark
Centre for Molecular Design, University of Portsmouth, Mercantile House, Hampshire Avenue, Portsmouth PO1 2EG, UK

which tautomers to use for the four bases [9]. However, even today we do not really know how to predict the relative stabilities of the base tautomers in solution for unknown systems. Similarly, tetracycline was introduced in 1949 and derivatives have been on the market ever since. However, after approximately 1 year's work using density-functional theory (DFT) and high level ab initio calculations, we were forced to apply an ominous "zwitterion correction" to bring our calculated stabilities of tetracycline tautomers and conformations into line with the little that is known experimentally [10].

The purpose of this article is to propose not only a strategy for calculating tautomers reliably and effectively, but also to consider the more general problem of the applicability of 3D-molecular structures in a given environment, the energetic penalties involved with using a non-equilibrium form and possible changes in structure between different environments. This fundamental concern for 3D-molecular design encompasses questions such as solvation, acidity and basicity and conformational flexibility. Above it, it requires that we develop a sensible strategy to consider all these factors in drug design. We are accustomed to considering the receptor as a moving target (but then generally fix its structure), but often ignore the chameleon-like character of the ligand completely. Above all, we usually have no idea of the energetic baseline to which all our calculations refer. I will outline some thoughts on possible ways out of this dilemma and a general strategy for using 3D structures in databases and drug design. The two major proposals will be that advances in hardware and software will allow us to consider far more compute-intensive techniques than we generally use at the moment and that we must consider new sources of data for constructing new, reliable calculational techniques. An integral part of any new approach to 3D-modeling must, however, be to implement a rational strategy for storing and using 3D-information.

Fundamentals

An accurate technique for calculating pK_a and pK_b values is the essence of any tautomer prediction technique. Note that this does not mean that we intend to calculate differences in pK_a or pK_b to determine tautomeric ratios, but the information that we need to create predictive models is contained, if anywhere, in pK_a and pK_b data, which are far more prevalent than those on tautomer ratios. MoKa [11], which is probably state-of-the-art for pK_a predictions, claims an accuracy of ± 0.4 pK_a-units for "normal" molecules and ± 0.7 for "unusual" ones. The lower of these two values would give ± 0.6 pK units for the equilibrium between two tautomers and the higher ± 1.0 pK units. Thus, it should be clear that we cannot really expect to be able to predict percentages of individual tautomers in solution, although TauThor, which is based on MoKa, apparently does this quite well [6]. However, this is not important as we do not really need to know the exact constitution of a given compound in water. What we do need to know is the free-energy penalty that we incur if we dock a different tautomer to the predominant one in solution. This is the key to 3D modeling that has often been neglected; all quantitative 3D modeling must be related to the most stable form in solution as the energy baseline. This means that we not only need to know the correct protonation state(s) and tautomer(s), but also the most stable conformation(s) in solution. This is evident from the equilibrium equation for binding to a receptor, for instance:

$$\begin{aligned}\mathrm{Ligand}^{\mathrm{global}}_{\mathrm{H_2O}} + \mathrm{Receptor}_{\mathrm{H_2O}} &\rightleftarrows \mathrm{Ligand}^{\mathrm{biological}}_{\mathrm{H_2O}} \\ + \mathrm{Receptor}_{\mathrm{H_2O}} &\rightleftarrows \left(\mathrm{Ligand}^{\mathrm{biological}} : \mathrm{Receptor}\right)_{\mathrm{H_2O}}\end{aligned}$$

where $\mathrm{Ligand}^{\mathrm{global}}$ is the global minimum (tautomer, conformation, protonation state) of the ligand in water and $\mathrm{Ligand}^{\mathrm{biological}}$ represents the biologically active (also tautomer, conformation and protonation state) structure of the ligand. The subscript H_2O indicates that the compound or complex is solvated in a physiological environment (pH, ionic concentrations, temperature). Of course, the above equilibrium neglects possible conformational, protonation or tautomeric changes in the receptor for simplicity. Exactly the same considerations, however, apply to the receptor as to the ligand. Thus, one prerequisite for real 3D modeling is knowledge of the free-energy difference between $\mathrm{Ligand}^{\mathrm{global}}_{\mathrm{H_2O}}$ and $\mathrm{Ligand}^{\mathrm{biological}}_{\mathrm{H_2O}}$. This is a tall order as it requires accurate knowledge of the relative energies of tautomers and conformations and of pK_as and pK_bs.

Models and data-limitations

In the following, we will distinguish between explicit models (i.e., those in which the quantity of interest is calculated directly in one or more steps) and quantitative structure–property relationships (QSPRs). QSPR approaches suffer from the relative paucity of accurate experimental data. Note that what experimentalists consider to be adequate data is not sufficient for constructing global predictive QSPR models. There are tens of thousands of reliable data for a property for which this is possible, $\log P_{OW}$. This is far from the relative paucity of data on tautomeric equilibria. Published models for a variety of properties (for instance, pK_a [11], solubility [12, 13] and $\log P_{OW}$ [14]) all achieve a maximum accuracy of ± 0.4–0.5 log units, almost irrespective of the descriptors or interpolation technique used. This strongly suggests that the quality of the models is limited by the data and not by the model itself. We have recently demonstrated this for a

three-class classifier model for kinetic aqueous solubility [12]. The maximum possible classification accuracy for a published dataset [15] is calculated to be of the order of only 70% if a standard experimental error of 0.45 log units is assumed. Thus, the accuracy of QSPR approaches is limited, but TauThor [6] suggests that it could be adequate for useful tautomer prediction. The more serious problem is that the available data is also very limited, so that potential QSPR models would necessarily be quite local and not widely applicable. Using related data sources is useful (e.g., TauThor relies on MoKa, which is parameterized for pK_a, for which there are far more data than for tautomer concentrations) but QSPR models depend strongly on the available data and can only become universal tools if they are based on very large amounts of consistent, very high-quality data. This limitation was emphasized by the authors of TauThor [6], even though MoKa was parameterized using 25,000 data but has, for instance, no entries for 2-hydroxypyridines. The 2-pyridone/2-hydroxypyridine tautomerism therefore cannot be treated accurately. Combined approaches can also alleviate this situation, as, for instance, shown by SPARC [8]. It is therefore useful to consider explicit models, which do not suffer from such limitations.

For explicit calculations of pK_a or relative tautomer stabilities, two components of the calculation must function with an accuracy that is currently unprecedented in computational drug design. The first is simply the gas-phase proton affinity of the base and the second is the difference in solvation energy in water between the base and its conjugate acid. However, this approach has problems; a recent review [16] indicates that mean absolute deviations of the order of 2–3 pK_a units are realistic for explicitly calculated pK_as based on continuum solvation models.

In principle, gas-phase proton affinities can be calculated "with chemical accuracy" (i.e., ± 1 kcal mol^{-1} or less) for drug-sized molecules or fragments thereof using extrapolation techniques based on DFT and ab initio calculations. These techniques, such as the G2 [17–19], G3 [20–22], and G4 [23, 24] procedures proposed by the late John Pople and other more recent methods [25], are computationally expensive and require significant memory and disk resources, but they are essentially automatic, so that the cheap component of computational chemistry, the hardware, can be used to maximum advantage. The major advantage of such calculations is that they deliver proton affinities that are essentially free of inconsistency errors [26]. The technique itself may give some (hopefully small) systematic errors for some classes of compound, but random variations caused by experimental inconsistency can be ruled out. An additional advantage is that the conformation is known exactly in such a calculation. A significant problem with experimental data for many physical properties is that many are for flexible molecules for which the conformation(s) is/are not known. This reduces the effective resolution of this data for parameterization of 3D-models by the variation in the target property between different conformations of the molecule. Experimental gas-phase proton affinities are at best suitable for validation purposes. They are usually measured by an equilibration ladder, which simply places the proton affinity of the compound being investigated between two neighbors in the proton-affinity scale [27]. As the measurements are time-consuming, data for relatively few compounds, few of them drug-like, are available. Thus, experimental uncertainty and scarcity of data make the situation for parameterized calculations difficult if the parameterization is based on experimental data.

Clearly, extrapolation methods will not be fast enough for routine use in *in silico* drug design for a very long time. We therefore need a fast technique that can reproduce the results of such calculations in a small fraction of the time and if necessary also reproduce the systematic errors of the extrapolation techniques. Here, there is no need to reinvent the wheel. Lou Allinger's beautiful work on the MM*n* force fields [28], which culminated in MM4 [29–32], demonstrated just how accurate force fields can be. Furthermore, modified MM2 force fields for a variety of different types of carbocations have been parameterized successfully, so that it is clear that strong electronic effects such as hyperconjugation can be treated well within sophisticated force fields [33–36]. The danger of overtraining remains small as long as additional force-field components are based on real physical effects such as hyperconjugation, especially as the random noise that leads to overtraining would be absent from the training data. Drug-like molecules typically include cyclic conjugation so that a π-only molecular-orbital (MO) technique such as that used in MMP2 [29, 37] would be needed. These techniques already exist and have shown that they are capable of the accuracy required. They must simply be extended to drug-like molecules.

The technology for implementing a force field that could calculate proton affinities accurately therefore exists and has done for some time. The force field would be slower than those typically used in drug design today, but would be well within the capabilities of modern hardware for large-scale applications. However, there may also be a case, for instance, for a polarizable force field [38–40] or one that uses multipole electrostatics [41, 42] in order to obtain the detailed electrostatics required for continuum solvation calculations. Whatever the final form of this "small molecule" force field, it would serve the dual purpose of delivering reliable pK_a and pK_b data (and hence reliable gas-phase prediction of relative tautomer energies) and providing the means of carrying out accurate conformational searches for flexible molecules. Given these important applications, it is quite remarkable how little attention has been paid to small-molecule force fields over

the past decades. The Merck force field [43–47] is one of the few examples where significant effort has been devoted to developing a force field for drug-like molecules. This alone indicates how far current techniques that allow the ligand to be flexible in docking or superposition have removed themselves from sound physical principles. It seems trivial to point out that the conformational (strain) energy of the ligand is essential for calculating binding free energies to receptors, but it has traditionally been ignored or treated very approximately with a "docking" force field.

The second component of an accurate calculational technique for pK_a and pK_b (and hence also the relative stabilities of tautomers) is a reliable and accurate solvation model. Technically, such models present little challenge. They can be based on any polarizable continuum technique that gives reliable electrostatics and generally consist of a continuum electrostatic term plus a surface-dependent catch-all term for the cavity energy, dispersion contributions etc. Many excellent parameterized techniques exist that fit the experimental data well. They may be based on ab initio [48], DFT [49, 50] or semiempirical molecular orbital [51] calculations. Some use conventional polarizable continuum model (PCM) theory [52], some generalized Born theory [53] and some the perfect shielding approximation [54]. Some ab initio or DFT-based techniques [55, 56] calculate the cavity energy explicitly using the Pierotti approach [57], although Mecke et al. [58, 59] recently, introduced a very promising alternative approach that takes anisotropic geometries into account. Perhaps the most promising combination of some degree of physical realism and a flexible parameterization is to use classical PCM theory with a surface-dependent term and a respectable quantum mechanical technique such as a moderate level of DFT [60, 61]. Such calculations on the force-field optimized geometries would be economical enough for routine use on large databases.

Thus, the problem with solvation free energies is not the availability of suitable calculational techniques. It is the available data. Experimental free energies of solvation in water are relatively sparse and, at least for ions, of doubtful quality. As for aqueous solubility [13], all models are based on the same data and those that are not grossly over-trained or poorly fitted all perform similarly in fitting the experimental data. It remains to be shown, but it is unlikely that current solvation models would be accurate enough to calculate acceptable pK_as and pK_bs from accurate gas-phase proton affinities.

Once again, we must look for an alternative source of solvation-energy data. Here we can use the abundant pK_a data. Note that this does not mean calculating tautomeric ratios as differences in pK_as or pK_bs; the strategy is rather to use pK_a and pK_b data, which is plentiful, in combination with accurate calculated proton affinities to derive differences in solvation energies for parameterizing a solvation model. Experimental pK_as combined with accurate gas-phase proton affinities give accurate differences in solvation energies between the protonated and unprotonated forms. These data can be "anchored" using some reliable free energies of solvation to construct a new dataset that has the advantage that it is focused on the property that interests us. The complete model (proton affinity plus solvation energy) is in effect parameterized to reproduce pK_a values, but one of the components (the proton affinity) is taken from a separate and reliable source. This strategy promises to give robust and versatile models.

Conformations

Given a reliable and accurate "small molecule" force field, the problem of finding the most stable conformation of a given tautomer in solution requires an adequate conformational sampling technique and an accurate solvation model. Sampling techniques, both for determining tautomers and conformations, are not the subject of this perspective, but a variety of techniques that give a population of low-energy conformations or tautomers exist. More important is that the sampling should ideally occur using energies including solvation. This is clearly not feasible using the DFT-based PCM technique suggested above and would also be very time-consuming if the force field were able to give accurate electrostatics on which to base the PCM calculations. The most practical solution at present is to assume that there is little difference between gas phase geometries and those in solution and to use the former for conformational searches followed by PCM calculations on the gas phase geometries of the most stable conformers. This procedure is unlikely to introduce large errors and final relative energies of the conformers could be obtained by a limited number of optimizations in the PCM environment. This approach would be especially effective if a fairly conventional force field were used to identify likely candidates for the most stable conformation and a polarizable force field with multipole electrostatics were used for the final PCM optimizations.

An alternative refinement of the candidate conformations would be to use umbrella sampling [62] in conventional molecular dynamics simulations to drive torsional angles between candidate conformations in order to obtain free-energy differences between conformations in explicit solvent. The accuracy of this approach depends critically on the quality of the force field electrostatics and the water model.

Calculating and using 3D-structures

The success of 2D-modeling has often rendered more expensive 3D techniques unnecessary, but the fact that

3D-techniques do not dominate pharmaceutical modeling is also because there is no consensus on how to use 3D data and methods. Discussion often centers on how many low-energy conformations to store and search, whether they should simply be the *n* most stable or whether they should be as diverse as possible etc. If we assume that 2D-structures implicitly encode all possible conformations, we can conclude that storing all possible 3D conformations should give us the 2D result at a far higher cost. This is clearly not the way to proceed. The arguments given above suggest that it is most important that we know the energy of the most stable structure in solution. All other conformations and tautomers must be seen relative to this energy, so the most stable structure and its characteristics must be stored for virtual screening. This requires that the conformational search step be performed for every molecule prior to storing it in a screening dataset or database, so that this step becomes critical. However, given that at the time of writing an adequately configured dual quad-core Nehalem compute node costs less than $6,000, computationally intensive pre-processing of large databases becomes feasible.

However, we have neglected updating our methods as ever more powerful hardware becomes available. We still use methods that, if we are honest, we know don't work but that require little computational power. If anything, we use modern hardware to do the same calculations as ever on more and more compounds. Force fields are an excellent example. We could start developing small molecule force fields of the quality of, say MM4 tomorrow without the need for any development work except for the parameterization. MM4 was published in 1996. The Merck Molecular Force Field (MMFF) [43–47] embodies the approach advocated here and has also been extended to include a GB/SA solvent model [63]. MMFF remains one of the very few force fields that can be considered to be equally applicable to proteins and small drug-like molecules. This is quite remarkable considering that a major industry depends almost exclusively on drug–protein interactions. The traditional protein force fields have been developed continuously using very standard physical principles, but force fields for ligands generally have the quality of GAFF [64]. An exception is the work done for the HYDE scoring function, which used MMFF for protein and ligand [65].

Perhaps it is time to design new physical principles for "slower" force fields designed to treat proteins and ligands equally well. If the new force field is ten times slower than current ones, Moore's law suggests that we will be back to the same speed in 3.5 years and 100 times will only take 7 years. This is not simply a question of reparameterizing. Most current force fields use an atomic monopole (net atomic charge) model for the electrostatics. This is clearly not adequate [66]. The binding effect of a *para*-chloro substituent on a phenyl ring, for instance, requires that parts of the chlorine surface be negative and other parts positive in order to take the halogen-bonding contribution into account [67]. Such situations can be represented with multipole models [42], which are only moderately more expensive than monopole models. However, electron–electron repulsion is also important [68], so that more sophisticated (nucleus + electrons) electrostatic models may be required. This can be an advantage as electrostatic models that are close enough to the correct physics can be set up using data from quantum mechanical calculations. This removes one of the major problems of force-field parameterization, the linear dependency between the two-center terms (Coulomb and van der Waals).

Conclusions

We have neglected tautomerism for too long, but taking it into account will force us to think about other aspects of 3D drug design. These all relate to using the correct structure (tautomer, protonation state and conformation) and being able to relate the free energy of the structure used to that of the system (both ligand and receptor) under physiological conditions. This is doable, but requires that we rethink the physical models behind the methods that we use and use the power of modern hardware to improve them, rather than just to calculate more molecules.

References

1. Pospisil P, Ballmer P, Scapozza L, Folkers G (2003) Tautomerism in computer-aided drug design. J Recept Signal Transduct Res 23:361–371
2. Elgueor J, Marzin C, Katritzky AR, Lind P (1975) The tautomerism of heterocycles. Adv Heterocycl Chem Suppl 1:1–656
3. Beak P (1997) Energies and alkylations of tautomeric hetericyclic compounds: old problems—new answers. Acc Chem Res 10: 186–192
4. Oellien F, Cramer J, Bayer C, Ihlenfeldt WD, Selzer PM (2006) The impact of tautomer forms on pharmacophore-based virtual screening. J Chem Inf Model 46:2342–2354
5. Seidel T, Wolber G, Langer T (2008) Tautomerism in structure-based 3D pharmacophore modeling. Chem Cent J 2(suppl):11
6. Milletti F, Storchi L, Sforna G, Cross S, Cruciani G (2009) Tautomer enumeration and stability prediction for virtual screening on large chemical databases. J Chem Inf Model 49:68–75
7. Marvin C (2010) http://www.chemaxon.com/product/tautomer.html. Accessed 13 Jan 2010
8. SPARC (2010) http://sparc.chem.uga.edu/sparc/. Accessed 15 March 2010
9. Watson JD (1980) In: Stent GS (ed) The double helix; a personal account of the discovery of the structure of DNA, a norton critical edition. W. W. Norton, New York
10. Othersen OG, Waibel R, Lanig H, Gmeiner P, Clark T (2006) An SCRF-DFT and NMR comparison of tetracycline and 5a, 6-anhydrotetracycline in solution. J Phys Chem B 110:24766–24774

11. Milletti F, Storchi L, Sforna G, Cruciani G (2007) J Chem Inf Model 47:2172–2181
12. Kramer C, Beck B, Clark T (2010) Insolubility classification with accurate prediction probabilities using a MetaClassifier. J Chem Inf Model 50:404–414
13. Clark T (2003) Modelling the chemistry: time to break the mould? In: Ford M, Livingstone D, Dearden J, van der Waterbeemd H (eds) EuroQSAR 2002: designing drugs and crop protectants. Blackwell, Oxford, pp 111–121
14. Kramer C, Beck B, Clark T (2010) A surface-integral model for $\log P_{OW}$. J Chem Inf Model 50:429–436. doi:10.1021/ci900377e
15. Kramer C, Heinisch T, Fligge T, Beck B, Clark T (2009) A consistent kinetic solubility dataset for early-phase drug discovery. ChemMedChem 4:1529–1536
16. Ho J, Coote ML (2010) A universal approach for continuum solvent pK_a calculations: are we there yet? Theor Chem Acc 125:3–21
17. Curtiss LA, Raghavachari K, Trucks GW, Pople JA (1991) Gaussian-2 theory for molecular energies of first- and second-row compounds. J Chem Phys 94:7221–7230
18. Curtiss LA, Raghavachari K, Pople JA (1993) Gaussian-2 theory using reduced Møller–Plesset orders. J Chem Phys 98:1293–1298
19. Curtiss LA, Raghavachari K, Pople JA (1993) The accurate determination of enthalpies of formation. Chem Phys Lett 214:183–185
20. Curtiss LA, Raghavachari K, Redfern PC, Rassolov V, Pople JA (1998) Gaussian-3 (G3) theory for molecules containing first and second-row atoms. J Chem Phys 109:7764–7776
21. Curtiss LA, Redfern PC, Raghavachari K, Rassolov V, Pople JA (1999) Gaussian-3 theory using reduced Møller–Plesset order. J Chem Phys 110:4703–4709
22. Baboul AG, Curtiss LA, Redfern PC, Raghavachari K (1999) Gaussian-3 theory using density functional geometries and zero-point energies. J Chem Phys 110:7650–7657
23. Curtiss LA, Redfern PC, Raghavachari K (2007) Approximating correlation effects in multiconfigurational self-consistent field calculations of spin–spin coupling constants. J Chem Phys 126:084108
24. Curtiss LA, Redfern PC, Raghavachari K (2007) Gaussian-4 theory using reduced order perturbation theory. J Chem Phys 127:124105
25. Martin JML (2005) Computational thermochemistry: a brief overview of quantum mechanical approaches. Ann Rep Comput Chem 1:31–43
26. Danikiewicz W (2009) How reliable are gas-phase proton affinity values of small carbanions? A comparison of experimental data with values calculated using Gaussian-3 and CBS compound methods. Int J Mass Spectr 285:86–94
27. Dixon DA, Lias SG (1987) Absolute values of gas proton affinities and basicities of molecules: a comparison between theory and experiment. Mol Struct Energ 2:269–314
28. Burkert U, Allinger NL (1982) Molecular mechanics, ACS Monograph 177. American Chemical Society, Washington, DC
29. Allinger NL, Chen K, Lii J-H (1996) An improved force field (MM4) for saturated hydrocarbons. J Comput Chem 17:642–668
30. Nevins N, Chen K, Allinger NL (1996) Molecular mechanics (MM4) calculations on alkenes. J Comput Chem 17:669–694
31. Nevins N, Lii JH, Allinger NL (1996) Molecular mechanics (MM4) calculations on conjugated hydrocarbons. J Comput Chem 17:695–729
32. Allinger NL, Chen K, Katzenellenbogen JA, Wilson SR, Anstead GM (1996) J Comput Chem 17:747–755
33. Reindl B, Clark T, Schleyer PVR (1996) A new method for empirical force field calculations on localized and delocalized carbocations. J Comput Chem 17:1406–1430
34. Reindl B, Clark T, Schleyer PVR (1996) Empirical force-field and ab initio calculations on delocalized open chain cations. J Comput Chem 18:28–44
35. Reindl B, Clark T, Schleyer PVR (1998) Empirical force field and ab initio calculations on allyl cations. J Comput Chem 18: 533–551
36. Reindl B, Clark T, Schleyer PVR (1998) Modern molecular mechanics and ab initio calculations on benzylic and cyclic delocalized cations. J Phys Chem A 102:8953–8963
37. Sprague JT, Tai JC, Yuh Y, Allinger NL (1987) The MMP2 calculational method. J Comput Chem 8:581–603
38. Warshel A, Kato M, Pisliakov AV (2007) Polarizable force fields: history, test cases, and prospects. J Chem Theor Comput 3:2034–2045
39. Patel S, Brooks CL III (2006) Fluctuating charge force fields: recent developments and applications from small molecules to macromolecular biological systems. Mol Sim 32:231–249
40. Halgren TA, Damm W (2001) Polarizable force fields. Curr Opin Struct Biol 2001(11):236–242
41. Devereux M, Plattner N, Meuwly M (2009) Application of multipolar charge models and molecular dynamics simulations to study stark shifts in inhomogeneous electric fields. J Phys Chem A 113:13199–13209
42. Plattner N, Bandi T, Doll JD, Freeman DL, Meuwly M (2008) MD simulations using distributed multipole electrostatics: structural and spectroscopic properties of CO- and methane-containing clathrates. Mol Phys 106:1675–1684
43. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17:490–519
44. Halgren TA (1996) Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J Comput Chem 17:520–552
45. Halgren TA (1996) Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J Comput Chem 17:553–586
46. Halgren TA, Nachbar RB (1996) Merck molecular force field. IV. conformational energies and geometries for MMFF94. J Comput Chem 17:587–615
47. Halgren TA (1996) Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. J Comput Chem 17:616–641
48. Tomasi J, Persico M (1994) Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. Chem Rev 94:2027–2094
49. Kelly CP, Cramer CJ, Truhlar DG (2005) SM6: a density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute–water clusters. J Chem Theor Comput 1:1133–1152
50. Thompson JD, Cramer CJ, Truhlar DG (2005) Density-functional theory and hybrid density-functional theory continuum solvation models for aqueous and organic solvents: universal SM5.43 and SM5.43R solvation models for any fraction of Hartree-Fock exchange. Theor Chem Acc 113:107–131
51. Giesen DJ, Hawkins GD, Liotard DA, Cramer CJ, Truhlar DG (1997) A universal model for the quantum mechanical calculation of free energies of solvation in non-aqueous solvents. Theor Chem Acc 98:85–109
52. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105:2999–3094
53. Viswanadhan VN, Ghose AK, Wendoloski JJ (2000) Estimating aqueous solvation and lipophilicity of small organic molecules: a comparative overview of atom/group contribution methods. Perspect Drug Discov Design 19:85–98
54. Klamt A, Schüürmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2: 799–805
55. Soteras I, Curutchet C, Bidon-Chanal A, Orozco M, Luque FJ (2005) Extension of the MST model to the IEF formalism: HF

and B3LYP parametrizations. J Mol Struct (THEOCHEM) 727: 29–40
56. Curutchet C, Orozco M, Luque FJ (2001) Solvation in octanol: parametrization of the continuum MST model. J Comput Chem 22:1180–1193
57. Pierotti RA (1976) A scaled particle theory of aqueous and nonaqueous solutions. Chem Rev 76:717–726
58. Hansen-Goos H, Roth R, Mecke K, Dietrich S (2007) Solvation of proteins: linking thermodynamics to geometry. Phys Rev Lett 99:128101
59. Hansen-Goos H, Mecke K (2009) Fundamental measure theory for inhomogeneous fluids of nonspherical hard particles. Phys Rev Lett 102:018302
60. Marenich AV, Cramer CJ, Truhlar DJ (2009) 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
61. Cramer CJ, Truhlar DJ (2008) A universal approach to solvation modeling. Acc Chem Res 41:760–768
62. Torrie G, Valleau J (1974) Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid. Chem Phys Lett 28:578–581
63. Cheng A, Best SA, Merz KM, Reynolds CH (2000) GB/SA water model for the Merck molecular force field (MMFF). J Mol Graph Model 18:273–282
64. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174
65. Reulecke I, Lange G, Albrecht J, Klein R, Rarey M (2008) Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the HYDE scoring function. ChemMedChem 3:885–897
66. Politzer P, Murray JS, Concha MC (2008) σ-hole bonding between like atoms; a fallacy of atomic charges. J Mol Model 14: 659–665
67. Clark T, Hennemann M, Murray JS, Politzer P (2007) Halogen bonding: the σ-hole. J Mol Model 13:291–296
68. Wesolowski TA, Warshel A (1993) Frozen density functional approach for ab initio calculations of solvated molecules. J Phys Chem 97:8050–8053