A new class of models for computing receptor–ligand binding affinities

Michael K Gilson, James A Given and Martha S Head

Models for predicting the binding affinities of molecules in solution are either very detailed, making them computationally intensive and hard to test, or very simple, and thus less informative than one might wish. A new class of models that focus on the predominant states of the binding molecules promise to capture the essential physics of binding at modest computational cost.

Address: Center for Advanced Research in Biotechnology, National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, MD 20850, USA.

E-mail: gilson@indigo14.carb.nist.gov

Electronic identifier: **1074-5521-004-00087**

Chemistry & Biology February 1997, **4**:87–92

"*... there is a balance between admitting enough of the complexity of reality into a problem for it to be interesting, while keeping the chunk of reality simple enough or small enough so that it can be modeled. A completely understood (or 'reduced') problem is boring, but a realistically complex one is frustrating*" (M. Reppy, quoted in [1]).

The noncovalent association of molecules in solution is of central importance in biology and chemistry, and predicting binding affinities is a long-standing problem in computational chemistry. A new class of computer models for predicting binding affinities is now emerging from research in a number of laboratories. These 'predominant-states' models treat solvent implicitly and sample over only a modest number of solute degrees of freedom. This makes it possible to approximate the binding affinity as a sum of contributions from the low energy conformations of the complex and the free molecules. It is anticipated that these models will be computationally efficient while still capturing the essential physical chemistry of binding. This article outlines the motivation, form and expectations for this class of models.

Why seek new models of noncovalent association?

Living matter is neither liquid nor solid. It consists of a broad dispersion of organic molecules in an aqueous medium, organized into highly specific noncovalent complexes. To understand the physical chemistry of biological systems, we must first understand the noncovalent association of molecules.

In an important subset of biological association reactions, a small ligand, L, binds to a larger receptor molecule, R:

$$R + L \rightleftharpoons RL \qquad \text{Equation 1}$$

The greater the affinity of the two molecules for each other, the lower the standard free energy of binding, ΔG°. Examples of receptor–ligand binding include the recognition of substrates by enzymes, the action of biological transmitters at their protein receptors, the inhibition of enzymes by drugs and, in nonbiological systems, the recognition of guest molecules by synthetic hosts. Thus, models for receptor–ligand binding are useful in many areas, including enzyme engineering, drug design, and the design of synthetic hosts for use in chemical separations. Considerable progress has been made in developing computational methods for predicting receptor–ligand affinities. There is still much room for improvement in these methods, however.

Most existing methods of computing free energies of binding are either very detailed or highly simplified [2]. The detailed methods, free-energy simulations, typically involve molecular dynamics simulations in which many solute and solvent molecules are treated as mobile. They therefore involve conformational sampling over thousands of degrees of freedom [3–10]. Because these methods represent the system in detail, it is reasonable to expect them to be quite accurate. For the same reason, however, they are computationally intensive and are subject to convergence problems [11,12]. This, makes it difficult to assess the accuracy of free-energy simulations by systematically comparing their results with experimental data.

The simplified methods of computing free energies of binding are essentially phenomenological in nature. They avoid convergence problems by using a single, rigid conformation of the ligand–receptor complex [2,13–21], and involve no conformational sampling at all. The free energy of binding is taken to be a sum of energy components associated with ligand–receptor interactions, such as hydrogen bonds and hydrophobic contacts. Additional terms, analyzed elsewhere [22], account for the changes in solute entropy due to the loss of mobility upon binding. The various energy components are typically not derived explicitly from the underlying statistical thermodynamics [22], and the parameters of the models are often determined from binding data in a statistical manner. Thus These models can be difficult to interpret in physical

terms. Also, it seems inevitable that the accuracy of their predictions will be sharply limited by the neglect of conformational flexibility.

Is it possible to devise models of binding that are grounded in statistical thermodynamics, that allow for conformational flexibility, and that are nonetheless computationally tractable? Below, we outline a class of models that has these properties.

A middle ground for models of binding

Between the extremes of free-energy simulations and energy-component models, there lies a broad middle ground with ample room for models of intermediate complexity that would retain many of the advantages of the two approaches described above. One class of models, which we term 'predominant-states models', combines the judicious selection of a modest set of degrees of freedom with aggressive conformational sampling to identify the low-energy conformations of the complex and of the free molecules. These conformations can be used to estimate the standard free energy of binding (see below). The number of explicit degrees of freedom requiring sampling is kept small by using implicit models of the solvent. In addition, only those parts of the receptor and ligand that are most likely to change conformation upon binding are treated as flexible. Such methods should allow computations of binding affinities to converge in a matter of hours or days. Models of this type capture more of the physical chemistry of binding than energy-component models and cost less computer time than current free-energy simulations. The efficiency of these models will make it possible to define their accuracy and range of applicability through extensive comparisons with experiment. It should therefore be possible to establish well-characterized tools for molecular design and the interpretation of experimental data. The following sections discuss the theoretical basis for predominant-states models of binding, the formulation and validation of specific models, and the expected obstacles and gains associated with this approach.

Theoretical basis for predominant-states binding models

Statistical thermodynamics dictates that the affinity of a ligand for a receptor is directly related to configuration integrals that range over all possible conformations of the separate molecules and of the complex (for review, see [22]). The standard free energy of binding can be written to within a constant as:

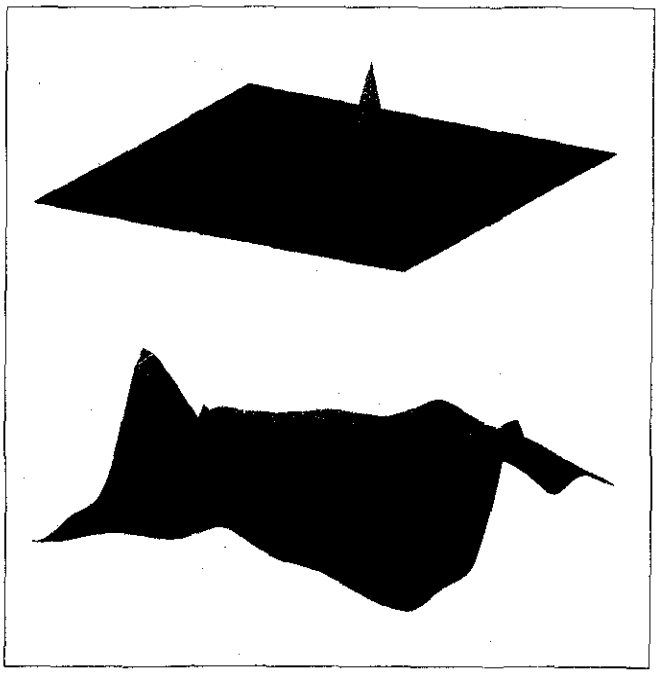
$$\Delta G^\circ = -RT \ln \left(C^\circ \frac{Z_{RL}}{Z_R Z_L}\right) \qquad \text{Equation 2}$$

where C° is the standard concentration, usually 1 M, R is the gas constant, T is the temperature and Z_R, Z_L and Z_{RL} are the configuration integrals of the receptor, the ligand, and the complex, respectively. A configuration integral Z is the integral of a Boltzmann factor over all conformations of the molecule or complex:

$$Z = \int e^{-(U(r)+W(r))/RT} dr \qquad \text{Equation 3}$$

Here, $U(r)$ and $W(r)$ are the potential energy and solvation energy, respectively, of the molecule or complex in a given conformation specified by coordinates r. It should be noted that for the complex, the coordinates r specify not only the conformation of the receptor and the ligand, but also the relative position and orientation of the two molecules. The solvent term, $W(r)$, is a potential of mean force that incorporates the influence of solvent molecules. It implicitly includes additional configuration integrals over solvent degrees of freedom [22]. The largest contributions to the configuration integrals in Equation 3 correspond to conformations with large Boltzmann factors; that is, conformations of low energy $U + W$. For biomolecules, the binding affinity often depends upon the pH of the solvent; in other words, upon the chemical potential of the proton. In such cases, the thermodynamic link between binding and protonation can be accounted for in Equation 2 by using a binding polynomial [22–24]. The binding polynomial essentially replaces the configuration integrals of R, L and RL by sums of configuration integrals over protonation states [22].

Figure 1



A few of the possible conformations of a receptor–ligand complex contribute most of the binding energy. Bottom, a schematic two-dimensional surface representing the energy as a function of the conformation of the complex. Although the energy surface is complicated, the corresponding Boltzmann factor, top, is dominated by a small set of low-energy conformations, the predominant states.

Although the standard free energy of binding is written compactly in Equation 2, this quantity is difficult to evaluate because the integrals range over an enormous number of possible conformations. This problem may be addressed by fixing many internal coordinates of the ligand–receptor system in a single predominant conformation and treating only the remaining coordinates as flexible. This is a good approximation as long as the probability distribution of the fixed coordinates does not change much upon binding [22]. It is usually necessary to treat the entire small molecule as flexible, although much of a large receptor may be treated as rigid. Bond lengths and angles may also be treated as fixed if their change upon binding is small [22]. Thus intractable configuration integrals over hundreds or thousands of degrees of freedom may often be replaced by tractable integrals over a few dozen essential degrees of freedom.

Some receptors do change conformation significantly upon binding a ligand, making it difficult to compute the binding energy. However, in many applications, such as computer-aided drug design, it suffices to compute the relative affinity of different molecules for the same receptor. Because the bound conformation of the receptor is often similar for a range of ligands [25], it will be a good approximation to keep much of the receptor rigid when computing relative affinities in these cases.

The limited integrals that remain after these simplifications may be estimated by a set of methods that we call predominant-states approximations [26,27]. These assume that the integrals are dominated by contributions from a set of low-energy conformations (Fig. 1) which can be identified by optimization methods. If N important energy minima are identified, then the configuration integral Z can be approximated as a sum of contributions from the energy minima. Furthermore, Monte Carlo methods can be used to account for conformations outside the N energy minima [27].

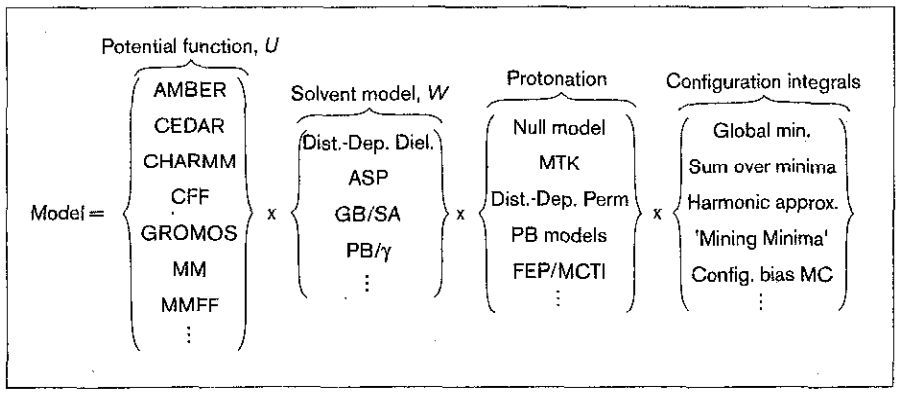
Systematic construction of tractable models of binding

Predominant-states models for computing the standard free energy of binding involve identifying the major minima of the energy function $U + W$ and evaluating their contributions to the configuration integrals of the receptor, the ligand and the complex. The four chief components of such models are as follows: a means of estimating the potential energy, U; a means of estimating the solvation energy, W; a treatment of protonation equilibria; and a means of evaluating the configuration integrals, Z. More than one implementation is already available for each component and a specific model of binding can be constructed from any combination of these implementations (Fig. 2). It should be noted that each component of a specific binding model can be tested independently because these components are physically meaningful. Below we discuss each of the four components.

The potential energy function, U

The methods for estimating the potential energy, U, of a molecule are currently the most mature of the model components listed below. A number of research groups have developed potential energy functions for use in molecular simulations; these include the AMBER [28], CEDAR [29,30], CHARMM [31], CFF [32–34], GROMOS [35], MM [36–38] and MMFF [39–43] force fields. Some of these have been substantially revised and improved since their first releases and further enhancements and extensions will undoubtedly be forthcoming. Although our own prototype binding calculations use the CHARMM energy

Figure 2



Construction of models of binding from some specific implementations of the four components defined in the text. Potential function citations are in the text; Dist.-Dep. Diel., distance dependent dielectric function; ASP, atomic solvation parameters [45]; GB/SA, generalized Born/surface area model [49]; PB/γ, Poisson–Boltzmann/ surface tension model [48]; Null model [53,62]; MTK, modified Tanford–Kirkwood model [54]; Dist.-Dep. Perm., distance-dependent permittivity model [65]; PB Models, Poisson–Boltzmann models of titration (citations in text); FEP/MCTI, free energy perturbation or multiconfiguration thermodynamic integration free-energy simulations (citations in text); Global Min., global minimum; Harmonic Approx., harmonic approximation [76]; 'Mining Minima', predominant-states method cited in [27]; Config. bias MC, configurational bias Monte Carlo (citations in text).

function, it will be important to test the performance of various energy functions in binding calculations.

The solvation energy function, W

Implicit solvent models suppress the degrees of freedom of the solvent and make them implicit in the solvation term, *W*. This term is essentially the work of transferring the solute, in a given conformation, into solvent [22]. Implicit solvent models have improved dramatically in the last 5–10 years and are currently the subject of intensive research in several laboratories. Existing implicit solvent models include: hydration shell and related models [44–46]; models based upon a combination of electrostatic and nonpolar solvation terms [47–50]; and the SMx models [51], which include contributions from several types of solute–solvent interaction. Many of these models provide good agreement with the measured free energies of solvation for small molecules, suggesting that implicit models of solvent can serve well in models of binding. In fact, it is somewhat surprising that so many disparate solvent models perform well for small molecules. This could be a result of the simplicity of the shapes of small molecules. Calculations of the binding interaction between two larger molecules may test solvent models more stringently, because such complexes often possess cracks and crevices that can sequester solvent molecules. Ultimately, it might prove necessary to employ a hybrid model of the solvent in which a small number of explicit solvent molecules occupy specific binding sites while the bulk of the solvent is treated implicitly [22,52].

Protonation equilibria

Models for protonation equilibria have been established at various levels of detail and computational speed (for review, see [53]). These include models based upon free energy simulations [54–57], upon the linearized Poisson–Boltzmann (PB) equation [58–64], and upon distance-dependent solvent permittivities [65]. Perhaps the simplest relevant model assumes that the pK_a of a titratable group in a biomolecule is the same as that of a chemically similar group in bulk solvent. This 'null model' is often used by default, and is fairly accurate [62]. Indeed, the null model is more accurate than some computationally intensive alternatives [53], highlighting the need to validate computer models of biomolecules.

Evaluation of configuration integrals

Predominant-states methods estimate configuration integrals as sums of contributions from energy minima. These methods can take advantage of the ingenious algorithms that have been devised for identifying energy wells on the multidimensional energy surfaces of small molecules and proteins. Predominant-states methods may be implemented at several levels of approximation.

In the most accurate predominant-states methods, the contributions of the low energy valleys surrounding the major energy minima are evaluated in detail [27]. The sum of these contributions is an estimate of the overall configurational integral. A supplementary contribution from conformations outside the major energy wells may be computed by means of a Metropolis Monte Carlo calculation [27]. This approach has been shown to be tractable and to yield numerically accurate configuration integrals for isolated alanine polypeptides and cyclic urea inhibitors of HIV-1 protease [27]. Our unpublished results also show that the algorithm is tractable for calculations involving a cyclic urea in the binding site of HIV-1 protease [66], and a nonpolar ligand in a cavity in T4 lysozyme [67]. Another accurate algorithm for computing configuration integrals by sampling within energy minima has been successfully applied to several small-molecule systems (I. Kollossvary, personal communication). A less detailed predominant-states approximation that assumes all energy wells to be harmonic has yielded promising results in a study of inhibitor binding by thrombin [68].

If one makes the simplifying assumptions that all energy minima are similar in shape and that they are separated by high energy barriers, there is no need to examine the shapes of the energy wells, and the ratios of configuration integrals in Equation 2 may be computed in terms of sums of the Boltzmann factors at the minima. Such an approximation has been used in calculations for a host–guest system [69] and for the interactions of analytes with chromatographic surfaces [70]. Finally, it might be assumed that only the global energy minimum for each molecular species makes an important contribution to the free energy. This approximation, which makes it unnecessary to find any other energy minima, may be appropriate for rigid molecules with a single binding mode. In this limit, the predominant-states model essentially becomes a simple energy-component model of binding.

So far, we have emphasized the use of predominant-states methods for computing ΔG°, but two other approaches deserve to be mentioned. First, configuration integrals for systems of modest size can be computed by configurational bias Monte Carlo methods [71–73]. Such an approach has been used to compute configuration integrals for alkanes up to C_{25} in a zeolite [74], and a configurational bias method has recently been used to identify stable conformations of cyclic peptides [75]. Second, thermodynamic integration methods [3–10] could be used to compute ΔG° with a continuum solvent model and a small set of solute degrees of freedom. These methods might prove somewhat inefficient, because they require the system to be equilibrated for a series of artificial states between the starting and final states of interest. The predominant-states and configurational bias Monte Carlo methods require calculations for only the free and bound states.

Validation

Any computationally tractable model of binding involves approximations and therefore requires validation. The models discussed here have two advantages with regard to validation. First, they are computationally efficient and will permit extensive and systematic comparison with experimental data. The statistics obtained in such studies not only indicate the reliability of the model, but can also reveal unexpected trends that motivate further scientific study (see, e.g. [62,63]). Second, the present models are based upon statistical thermodynamics, and their components (e.g. the solvent model, W) can be developed and tested independently. As a consequence, the strengths and weaknesses of a given model can be analyzed in a meaningful way.

Conclusions

This article has outlined a class of models of binding that have a clear foundation in statistical thermodynamics, and yet are computationally tractable. A central feature of these models is the use of thorough conformational sampling over a modest number of essential degrees of freedom. This approach yields converged results in tolerably short computational times. The models are physically interpretable, because they are assembled from well-defined components that can be tested. In addition, the efficiency of these models will enable statistically significant validation studies to be carried out. It should therefore be possible to provide users with practical and well-characterized computational tools. These will be valuable in a range of applications, including structure-based drug design, host–guest chemistry, and enzyme engineering.

Acknowledgements

We thank the National Institute of Health (GM 54053), the National Institute of Standards and Technology and the Advanced Technology Program for support. H.S.R. Gilson provided useful comments.

References

1. Hoffman, R. *The Same and Not the Same.* Columbia University Press, New York, NY, USA.
2. Ajay & Murcko, M.A. (1995). Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* **38**, 4953–4967.
3. Wong, C. & McCammon, J.A. (1986). Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* **108**, 3830–3832.
4. Bash, P.A., Singh, U.C., Langridge, R. & Kollman, P.A. (1987). Free energy calculations by computer simulation. *Science* **236**, 564–568.
5. Beveridge, D.L. & DiCapua, F.M. (1989). Free energy via molecular simulation: application to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431–492.
6. Lybrand, T. (1990). Computer simulation of biomolecular systems using molecular dynamics and free energy perturbation methods. *Rev. Comput. Chem.* **1**, 295–320.
7. Straatsma, T.P. & McCammon, J.A. (1992). Computational alchemy. *Annu. Rev. Phys. Chem.* **43**, 407–435.
8. Kollman, P. (1993). Free energy calculations – applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417.
9. Warshel, A., Tao, H., Fothergill, M. & Chu, Z.-T. (1994). Effective methods for estimation of binding energies in computer-aided drug design. *Isr. J. Chem.* **34**, 253–256.
10. Marrone, T.J., Briggs, J.M. & McCammon, J.A. (1997). Structure-based drug design: computational advances. *Annu. Rev. Pharm. Tox.*, in press.
11. Mitchell, M.J. & McCammon, J.A. (1991). Free energy difference calculations by thermodynamic integration: difficulties in obtaining a precise value. *J. Comput. Chem.* **12**, 271–275.
12. Balbes, L.M., Mascarella, S.W. & Boyd, D.B. (1994). A perspective of modern methods in computer-aided drug design. *Rev. Comput. Chem.* **5**, 337–379.
13. Chothia, C. & Janin, J. (1975). Principles of protein–protein recognition. *Nature* **256**, 705–708.
14. Andrews, P.R., Craik, D.J. & Martin, J.L. (1984). Functional group contributions to drug-receptor interactions. *J. Med. Chem.* **27**, 1648–1657.
15. Erickson, H.P. (1989). Cooperativity in protein-protein association. The structure and stability of the actin filament. *J. Mol. Biol.* **206**, 465–474.
16. Novotny, J., Bruccoleri, R.E. & Saul, F.A. (1989). On the attribution of binding energy in antigen-antibody complexes MCPC 603, D1.3 and HyHEL-5. *Biochemistry* **28**, 4735–4749.
17. Searle, M.S., Williams, D.H. & Gerhard, U. (1992). Partitioning of free energy contributions in the estimation of binding constants: residual motions and consequences for amide-amide hydrogen bond strengths. *J. Am. Chem. Soc.* **114**, 10697–10704.
18. Horton, N. & Lewis, M. (1992). Calculation of the free energy of association for protein complexes. *Prot. Sci.* **1**, 169–181.
19. Murphy, K.P., Xie, D., Garcia, K.C., Amzel, L.M. & Freire, E. (1993). Structural energetics of peptide recognition: angiotensin II/antibody binding. *Proteins* **15**, 113–120.
20. Bohm, H.-J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **8**, 243–256.
21. Weng, Z., Vajda, S. & DeLisi, C. (1996). Prediction of protein complexes using empirical free energy function. *Prot. Sci.* **5**, 614–626.
22. Gilson, M.K., Given, J.A., Bush, B.L. & McCammon, J.A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.*, in press.
23. Wyman, J. (1965). The binding potential, a neglected linkage concept. *J. Mol. Biol.* **11**, 631–644.
24. Schellman, J.A. (1975). Macromolecular binding. *Biopolymers* **14**, 999–1018.
25. Appelt, K., *et al.*, & White, J. (1991). Design of enzyme inhibitors using iterative protein crystallographic analysis. *J. Med. Chem.* **34**, 1925–1934.
26. Gilson, M.K. (1993). Multiple-site titration and molecular modelling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins* **15**, 266–282.
27. Head, M.S., Given, J.A. & Gilson, M.K. (1997). 'Mining Minima': direct computation of conformational free energy. *J. Phys. Chem.*, in press.
28. Weiner, S.J., *et al.* & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784.
29. Ferro, D., McQueen, J. & Hermans, J. (1980). Energy minimizations of rubredoxin. *J. Mol. Biol.* **136**, 1–18.
30. Hermans, J., Berendsen, H., van Gunsteren, W. & Postma, J. (1984). A consistent empirical potential for water–protein interactions. *Biopolymers* **23**, 1513–1518.
31. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
32. Maple, J., *et al.*, & Hagler, A. (1994). Derivation of Class II force fields. 1. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comput. Chem.* **15**, 162–182.
33. Hwang, M., Stockfisch, T. & Hagler, A. (1994). Derivation of Class II force fields. 2. Derivation and characterization of a Class II forcefield, CFF93, for the alkyl functional group and alkane molecules. *J. Am. Chem. Soc.* **116**, 2515–2525.
34. Maple, J., Hwang, M., Stockfisch, T. & Hagler, A. (1994). Derivation of Class II force fields. 3. Characterization of a quantum force field, for alkanes. *Isr. J. Chem.* **34**, 195–231.
35. van Gunsteren, W.F. & Berendsen, H.J.C. (1987). In *Groningen Molecular Simulation (GROMOS) Library Manual.* Biomos, Nijenborgh 16, 9747 AG Groningen, The Netherlands.
36. Lii, J. & Allinger, N. (1991). The MM3 force field for amides, polypeptides and proteins. *J. Comput. Chem.* **12**, 186–199.
37. Allinger, N., Kuohsiang, C. & Lii, J. (1996). An improved force field (MM4) for saturated hydrocarbons. *J. Comput. Chem.* **17**, 642–668.
38. Allinger, N.L., Xuefeng, Z. & Bergsma, J. (1994). Molecular mechanics parameters. *Theochem. J. Mol. Struct.* **118**, 69–83.
39. Halgren, T. (1996). Merck Molecular Force Field. I. Basis, form, scope

parameterization, and performance of MMFF94. *J. Comput. Chem.* **17**, 490–519.
40. Halgren, T. (1996). Merck Molecular Force Field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **17**, 520–552.
41. Halgren, T. (1996). Merck Molecular Force Field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **17**, 553–586.
42. Halgren, T. (1996). Merck Molecular Force Field. IV. Conformational energies and geometries for MMFF94. *J. Comput. Chem.* **17**, 587–615.
43. Halgren, T. (1996). Merck Molecular Force Field. V. Extension of MMFF94 using experimental data, additional computational data and empirical rules. *J. Comput. Chem.* **17**, 616–641.
44. Kang, Y.K., Gibson, K.D., Nemethy, G. & Scheraga, H.A. (1988). Free energies of hydration of solute molecules. 4. Revised treatment of the hydration shell model. *J. Phys. Chem.* **92**, 4739–4742.
45. Eisenberg, D. & McLachlan, A.D. (1986). Solvation energy in protein folding and binding. *Nature* **319**, 199–203.
46. Stouten, P.F.W., Frommel, C., Nakamura, H. & Sander, C. (1993). An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simulat.* **10**, 97–120.
47. Gilson, M.K. & Honig, B. (1988). Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies and conformational analysis. *Proteins* **4**, 7–18.
48. Honig, B., Sharp, K. & A.-S Yang. (1993). Macroscopic models of aqueous solutions: biological and chemical applications. *J. Phys. Chem.* **97**, 1101–1109.
49. Still, W.C., Tempczyk, A., Hawley, R.C. & Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129.
50. Davis, M.E. (1994). The inducible multipole solvation model: a new model for solvation effects on solute electrostatics. *J. Chem. Phys.* **100**, 5149–5158.
51. Cramer, C. & Truhlar, D. (1995). Continuum solvation models: classical and quantum-mechanical implementations. *Rev. Comput. Chem.* **6**, 1–72.
52. Beglov, D. & Roux, B. (1994). Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063.
53. Gilson, M.K. (1997). Modeling protonation equilibria in biomolecules. In *Computer Simulations of Biomolecular Systems*, Escom, in press.
54. Warshel, A., Sussman, F. & King, G. (1986). Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* **25**, 8368–8372.
55. Kenneth M. & Merz, J. (1992). Determination of pK_as of ionizable groups in proteins: the pK_a of Glu 7 and 35 in hen egg white lysozyme and Glu 106 in human carbonic anhydrase II. *J. Am. Chem. Soc.* **113**, 3572–3575.
56. Levy, R.M., Belhadj, M. & Kitchen, D.B. (1991). Gaussian fluctuation formula for electrostatic free-energy changes in solution. *J. Chem. Phys.* **95**, 3627–3633.
57. A.D. MacKerell, Jr, Sommer, M.S. & Karplus, M. (1995). pH-dependence of binding reactions from free energy simulations and macroscopic continuum electrostatic calculations: application to 2′GMP/3′GMP binding to ribonuclease T1 and implications for catalysis. *J. Mol. Biol.* **247**, 774–807.
58. Matthew, J.B., Gurd, F.R.N., Garcia-Moreno, B.E., Flanagan, M.A., March, K.L. & Shire, S.J. (1985). pH-Dependent processes in proteins. *Biochemistry* **18**, 91–197.
59. Bashford, D. & Karplus, M. (1990). pK_as of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **29**, 327–335.
60. Yang, A.-S., Gunner, M.R., Sampogna, R., Sharp, K. & Honig, B. (1993). On the calculation of pK_as in proteins. *Proteins* **15**, 252–265.
61. Oberoi, H. & Allewell, N.M. (1993). Multigrid solution of the non-linear Poisson-Boltzmann equation and calculation of titration curves. *Biophys. J.* **65**, 48–55.
62. Antosiewicz, J., McCammon, J.A. & Gilson, M.K. (1994). Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **238**, 415–436.
63. Demchuk, E. & Wade, R.C. (1996). Improving the dielectric approach to calculating pK_as of ionizable groups in proteins. *J. Phys. Chem.* **100**, 17373–17387.
64. Gibas, C.J. & Subramaniam, S. (1996). Explicit solvent models in protein pK_a calculations. *Biophys. J.* **71**, 138–147.
65. Mehler, E.L. (1996). Self-consistent, free energy based approximation to calculate pH dependent electrostatic effects in proteins. *J. Phys. Chem.* **100**, 16006–16018.
66. Lam, P.Y.S., *et al.*, & Erickson-Viitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **263**, 380–384.
67. Morton, A., Baase, W.A. & Matthews, B.W. (1995). Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* **34**, 8564–8575.
68. Wang, J. & Purisima, E.O. (1996). Analysis of thermodynamic determinants in helix propensities of nonpolar amino acids through a novel free energy calculation. *J. Am. Chem. Soc.* **118**, 995–1001.
69. Lipkowitz, K.B. & Zegarra, R. (1989). Theoretical studies in molecular recognition: Rebek's cleft. *J. Comput. Chem.* **10**, 595–602.
70. Lipkowitz, K.B., Demeter, D.A., Zegarra, R., Larter, R. & Darden, T. (1988). A protocol for determining enantioselective binding of chiral analytes on chiral chromatographic surfaces. *J. Am. Chem. Soc.* **110**, 3446–3452.
71. Rosenbluth, M.N. & Rosenbluth, A.W. (1955). Monte Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.* **23**, 356–359.
72. Frenkel, D., Mooij, G. & Smit, B. (1992). Novel scheme to study structural and thermal properties of continuously deformable molecules. *J. Phys.-Condens. Matter* **4**, 3053–3076.
73. Dodd, L., Boone, T. & Theodorou, D. (1993). A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. *Mol. Phys.* **78**, 961–996.
74. Maginn, E.J., Bell, A.T. & Theodorou, D.N. (1995). Sorption thermodynamics, siting, and conformation of long n-alkanes in silicalite as predicted by configurational-bias Monte Carlo integration. *J. Phys. Chem.* **99**, 2057–2079.
75. Deem, M.W. & Bader, J.S. (1996). A configurational bias Monte Carlo method for linear and cyclic peptides. *Mol. Phys.* **87**, 1245–1260.
76. Wang, J., Szewczuk, Z., Shi-Yi Yue., Tsuda, Y., Konishi, Y. & Purisima, E.O. (1995). Calculation of relative binding free energies and configurational entropies: a structural and thermodynamic analysis of the nature of non-polar binding of thrombin inhibitors based on hirudin. *J. Mol. Biol.* **253**, 473–492.