PINSA - A, **64**, No. 2, 1998, pp.179-192
©Printed in India:

BIOMOLECULAR MODELLING AND SIMULATIONS - THE CURRENT STATUS

B JAYARAM

Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

(Received 9 June 1997; Accepted July 1997)

Introduction

Recall the ball and stick model of a 1, 2-disubstituted ethane introduced at high-school level these days. Ideas related to hindered rotation around a single bond, the instability of an eclipsed conformation relative to its staggered analogue emerges in a visually telling manner. The balls of course represent atoms and the sticks, the chemical bonds between atoms. A clash between two unconnected balls upon rotation around the axis of a stick, is interpreted as steric hindrance to rotation. This simple model provides a powerful tool to appreciate strain and stability associated with molecular conformations. It has been a valuable companion in synthetic and mechanistic organic chemistry and in other areas as well. In such a model, from the standpoint of intermolecular forces, there are no attractions between two unconnected atoms and no repulsions either, not even when the two atoms encounter each other–just that they cannot pass through each other. If this behaviour is to be formulated mathematically, one would say that the potential energy between two non-bonded atoms is zero for all distances greater than the sum of their radii and the energy goes to infinity if the distance is less than the above sum. This in the technical jargon is known as a hard sphere model which in a primitive way describes the interaction between two non-bonded atoms. The sticks which are not elastic imply that the force constants associated with the bonds and bond angles are infinitely large. A free rotation around a bond is indicative of a flat torsional energy surface. Some of these approximations can be easily removed. A soft sphere potential which includes van der Waals attractions and short range repulsions originating in Pauli's forces can be employed, as with a (12,6) Lennard-Jones potential or a Buckingham potential (exp,6) in lieu of the hard sphere model. A coulombic potential can be added to this if the two atoms carry charges, fractional or full. Similarly, the force constants associated with bond stretching and angle bending can be assigned more realistic values instead of infinities. The torsional energy surface can also be modified to represent reality more closely by introducing multiple minima indicating the conformational preferences of the molecule. Thus all the internal degrees of freedom of the molecule can be taken cognizance of.

$$E(\underline{X}^N) = (\tfrac{1}{2}) \sum_{\text{all bonds}} k_l (l - l_o)^2 + (\tfrac{1}{2}) \sum_{\text{all bond angles}} k_\theta (\theta - \theta_o)^2 + (\tfrac{1}{2}) \sum_{\text{all dihedral angles}} V_n (1 + \cos n(\omega - \omega_o)) + \sum_{\substack{\text{all non-bonded} \\ \text{pairs}}} [\varepsilon_{ij} \{(R_{ij}/r_{ij})^{12} - 2(R_{ij}/r_{ij})^6\} + (q_i q_j / Dr_{ij})] \quad \ldots(1)$$

In the above equation, the first term describes the penalty for bond stretch from its equilibrium value. Term k_l is the force constant for bond stretch and l_o the equilibrium bond length. The second term denotes the energy penalty for deforming the bond angle θ_o is the equilibrium value for the bond angle and k_θ the corresponding force constant. Both first and second terms are harmonic (i.e., Hooke's law type) functions. Cross terms of the stretch-bend genre can also be added if required. The third term, a Fourier series

representation, quantifies the energy as a function of the dihedral angle ω, V_n is the rotational barrier height, n is the periodicity of rotation ω_0 is the reference angle where the torsional energy is maximum. This function (expressed further as a sum of different terms with n = 1,2,3 in the Fourier series) captures the multiple energy minima of varying heights usually seen with dihedral angle variations. The non-bonded interactions are taken care of by the last term. Variable ε_{ij} is the well depth, R_{ij} is the minimum energy interaction distance, q_i and q_j are the partial charges on atoms i and j, r_{ij} is the distance between them and D is a dielectric function.

The consequence of all these refinements (eq. I) is that the ball and stick model hops out of our hands to become a more accurate molecular model easier handled by a computer. Qualitative arguments on stability and reactivity turn quantitative. Just as the original ball and stick model, computer modelling is but a tool to help interpret experiment at an atomic level, to suggest new experiments without the ordeals of the hit and trial approach and to generate new knowledge on the behaviour of the molecular system wherever experiments are difficult to perform.

The current day computational research in the realm of biomolecules is concerned with developing a molecular view of stability and recognition based on the laws of physics. This is an energy-based approach. Other approaches profiting from computational techniques involving artificial intelligence, distance geometry, computer graphics etc, have also found fruitful applications in deciphering biomolecular structure and function such as in modelling protein structures and in NMR structural refinements. The main focus of the following discussion will be on energy-based approaches.

A potential function[1] (eq. I) is a prescription for computing the energy of the molecular system if the cartesian coordinates (i.e., configuration X^N) of all the atoms/ions/particles of the system are specified. The energy computed is often referred to as single point energy, (i.e. the energy of the point X^N) in the configuration space of the molecular system. A force field comprises a potential function together with a listing of all the parameters required for the calculation of energy and forces, such as the force constants, the equilibrium values for the internal degrees of freedom each considered in isolation, the partial atomic charges on each of the atoms in varying chemical environments etc. The parameters come from extensive and careful calibrations against experiment or rigorous quantum mechanical calculations. AMBER[2], CHARMM[3], CVFF & CFF-91[4], ECEPP[5], GROMOS[6], MMx(x=1,2,3)[7], OPLS[8] are some of the popular force fields in vogue today. (Names such as AMBER and GROMOS are also used for a suite of programs which perform modelling and simulations.) Each of the above force fields has undergone considerable improvements over the years with a significant increase in accuracy and reliability of the computational predictions. An alternative to the potential function approach is to solve the Schrödinger equation for the system but this becomes computationally prohibitive as the system size increases. The choice between molecular and quantum mechanics rests on the system size and the level of rigour required in representing the system. For biomolecular systems consisting of 10^3-10^4 atoms, the former is the obvious choice although whenever a chemical bond is formed or an electron transfer occurs a quantum mechanical treatment in some form or another is indispensable.

In energy minimization (EM) protocols, the objective is to arrive at a structure that has the minimum most energy. Several methods such as the steepest descent, conjugate gradient, Newton-Raphson method have been proposed to accomplish this task[1]. Starting with an arbitrary structure one often ends up with the nearest energy minimum in the configuration space of the molecule. Global energy minimum is elusive but is reachable for small systems by an intelligent design of a series of minimizations. This—a structure with minimum most energy—is often the goal of molecular modelling exercises. This is supplemented with the hypothesis that the lower the energy the more stable is the mechanical system. Most of the questions related to structure and function can be converted to energy related problems. Conformational analysis is a stability problem which involves mapping the energy as a function of the internal coordinates of the molecule. The effect of a mutation on the stability of a protein can be converted into a problem seeking the minimum energy structures of the system before and after the mutation followed by a computation of the energy difference. The relative binding efficacy of a series of ligands to a receptor is a recognition problem. The minimum energy structure of the receptor-ligand complex with each ligand is generated. The binding energy is then computed as $\Delta E_{binding} = E_{complex} - (E_{receptor} + E_{ligand})$. This allows an ordering of the ligands based on the energetics which hopefully correlates with the binding constant data $\Delta G^{\circ} = -RT \ln K$; ($\Delta G^{\circ}$ is the change in

the standard free energy, R is the gas constant, T the absolute temperature and K the equilibrium constant). A serious theoretical limitation of the energy minimization studies besides the difficulty in attaining the global minimum, is the lack of contact with thermodynamics and the consequent loss of correspondence with experiment. This notwithstanding, energy minimization studies continue to be of considerable value providing qualitative insights into the problem on hand and prove to be decisive in favourable cases, especially when entropic effects and thermal averaging are expected to play only a minor role. Methods to compute heats of formation and entropies via a normal mode analysis have also been developed[1]. Molecular mechanics calculations are employed routinely for investigating structure-activity relationships particularly in the area of drug design.

The connection with thermodynamics is established *via* an ensemble averaging (as in Monte Carlo method) or a time averaging (as in molecular dynamics method) of the energies of all the configurations accessible to the system under specified conditions (constant temperature, volume and number of particles for instance). This gives the average internal energy U. $U = \langle E(\chi^N) \rangle$ where angular brackets denote Boltzmann averages. This with a pressure volume correction gives the enthalpy ($H = U + PV$). A further addition of the entropic term results in the Gibbs free energy of the system ($G = H - TS$). Molecular simulations *via* Monte Carlo (MC) method (in which the configurations of the system are generated according to Boltzmann distribution in a stochastic approach) or molecular dynamics (MD) method (where new configurations of the system are generated by solving equations of motion in a deterministic approach) mimic the system at equilibrium and both structural and thermodynamic properties of the system become accessible. Besides these MD also yields dynamic properties. The MC method due to Metropolis *et al.*[9], the molecular dynamics studies of Alder and Wainright[10] on simple liquids, those of Rahman and Stillinger[11] on liquid water and the pioneering theoretical studies of Scheraga and coworkers[12] on proteins, influenced the course of events quite significantly. Molecular dynamics (MD) studies on BPTI *in vacuo*[13] and alanine dipeptide in water[14] constitute some of the significant early attempts to apply the MD technique to biomolecular systems. Parallel to this evolved the MC methodology as applied to aqueous solutions of biomolecules[15,16]. The last decade has seen a welcome explosion in the numerous extensions and improvements of the above methodologies as applied to biomolecular systems [Ref. 17-23 and references therein] in several laboratories all over the globe.

Structure and conformation of biomolecules are sensitive to solvent and salt concentration. Modelling molecules under physiological conditions implies a consideration of the aqueous environment preferably along with the supporting electrolyte (145 mM NaCl for instance). The first step thus is to ensure that an accurate description of the structure and thermodynamics of solvent water is available. Several water models have been proposed, SPC[24] and TIP4P[25] being two of the popular choices. Simulations of aqueous solutions proceed *via* the solvation of biomolecule of interest with desired / sufficient number of explicit waters (TIP4P or any such water model) with their configurations, along with that of the biomolecule, generated by MD or MC simulations. Water organization around biomolecules has become an intense research area[26-28] with a growing theoretical evidence of its stabilizing influence and the concurrent experimental observations such as the identification of spine of hydration in the minor groove of a B-DNA crystal structure[29] and the observation that water trapped at the interface of two biomolecules (protein and DNA for instance) could contribute to specificity via water mediated hydrogen bonds[30]. Several methods of solvent analysis have been proposed of which the proximity criterion developed by Beveridge and coworkers[31], enables a quantification of the structure and energetics of solvent around multifunctional solutes, in a unique fashion, *via* quasi-component molecular distribution functions[32].

Salt effects can be included in a natural way in the simulations once the ion-water[33], ion-ion and ion-solute potential functions are available. The number of small ions considered is mostly restricted to maintaining electroneutrality in the solution and few simulations reported thus far could consider added salt effects. This is due to severe convergence problems in simulations with ions and issues related to sampling. Thus methodologies exist at least in principle, to deal with the environmental (solvent and salt) effects on the stability of biomolecules.

An outstanding problem of current interest to researchers in biomolecular area is that of molecular recognition. Protein folding is an intramolecular recognition problem and regulation of gene expression by certain proteins (protein-DNA recognition) is an intermolecular recognition problem. Two critical indices that a chemist desires out of a theoretical study of such events is the free energy of folding/binding and rates associated with these processes. Once the thermodynamic and kinetic parameters are specified, the focus shifts to an understanding of the origins of stability, spontaneity and speed at an atomic level. This is to help design molecules with altered stabilites and binding proclivities. A theoretical determination of these parameters requires structural data as input. The current status on structure determination and characterization of the dynamics of biomolecules is summarized below.

Structure & Dynamics of Biomolecules

Molecular dynamics simulations offer a viable approach to study the structure and internal motions of biomolecular systems. It is one's good fortune if X-ray structures (or NMR derived distance constraints) are available as points of departure for further theoretical studies. The Brookhaven Protein Data Bank[34] and the Nucleic Acid Data Base[35], in this regard, have been sources of rich structural information. De novo structure determination of isolated proteins from their primary sequences (protein folding) is an unsolved problem as yet. For DNA, one would expect that Watson-Crick type double helical structures could be built with any arbitrary base sequence. Single crystal studies on oligonucleotides point to several intrinsic sequence dependent structural variations (the DNA fine structure) within the overall double helix, which appear to have a functional role[36-38]. Arriving at the structure of a complex of two biomolecules continues to be a challenging problem in docking. Nonetheless, several simulation studies have appeared in the literature addressing structural issues with progressively increasing levels of confidence.

Protein Folding

Ideally, one would like to build a random structure from a given amino acid sequence and let the theory (or computer) generate the conformation with the minimum most free energy, considering naturally the surrounding medium, temperature etc. and this conformation would hopefully correspond[1] to the native structure of the protein. This is a hopeless approach[39]. Cast in the language of energy hyper-surface, folding is an attempt to locate the global minimum among *multiple minima*. Typical values of the free energy difference, ΔG, between the native and denatured states for globular proteins lie in the range of -5 to -15 kcal/mol[40] and the folding process occurs *in vitro* on a milliseconds-to-seconds time scale[41]. Both these factors compound the task for simulations. Some novel simulation strategies and numerous fancy mathematical techniques have been applied to gain some insights into the protein folding problem and these have been reviewed recently[1,12,20,39]. It is probably fair to say that with the currently available state of the art techniques, the tertiary structure of a polypeptide in solution with ten or less amino acid residues, can be predicted to a high degree of accuracy. Theoretical determinations of structures of [Met]enkephalin–a pentapetide, Gramicidin S–a decapeptide are some success stories[12]. Results on a 20-residue membrane-bound portion of a relatively small globular protein, melittin, are in general accord with experiment[12]. This is still far off from the folding of a full polypeptide chain with more than fifty residues. For larger systems, low resolution prediction of structure is feasible where the folding motif is known[42]. In this regard, besides the energy-based approaches, modelling by homology[1,12] is becoming a popular technique to get to the native structure. It is of course well known that secondary structure prediction schemes (Chou-Fasman rules and their several variants)[43,44] have been highly successful. It is the piecing together of the secondary structural units embedded in solvent that is problematic. We have no clearcut clue to the rules of protein-protein recognition beyond the traditional way of a decomposition of the interactions in terms of electrostatic, van der Waals, and hydrophobic contributions. (Glibly stated, protein folding is a 20 by 20 problem. Protein-DNA recognition in this parlance is a 20 by 4 problem. Both are unsolved. Watson and Crick solved the 4 by 4 problem).

DNA Fine Structure: A minimal set of internal variables of DNA comprises six backbone torsions, the sugar pucker and the glycosidic angle of the base (syn/anti) on each strand. Inter and intra-basepair degrees of freedom, groove depths and widths add to this list[36, 45-49]. Some of these variables are correlated. A combination of these parameters decides whether a polynucleotide is A-like or Z-like with several alphabets in between, with the B-form of course being the most stable under physiological conditions[36-38,46]. A goal of the simulations is to characterize and codify these base sequence dependent structural features. Simulations on DNA, unlike the protein case, must include counterions and solvent to reduce inter-phosphate repulsions[49-52]. The simulated system is highly charged and a proper treatment of the electrostatics is thus imperative. Additionally, a key question with DNA structure determination from simulations has been how to judge the accuracy of the results. Comparisons with crystal structure are fair if the environment in the simulation is also crystalline. DNA in aquo does not necessarily conform to the crystal structure. Progress made in NMR structural determinations[53] has been helpful in resolving this issue. These factors notwithstanding, a number of simulations were reported highlighting the local structural deviations from the canonical[54] and crystal structures. The current standard for MD simulations on DNA appears to be a nanosecond or longer simulation with Ewald summation for charge-charge interactions[55-58], a protocol that can be implemented only on a supercomputer. It is now feasible to delineate crystal packing effects from sequence induced structural effects and appreciate groove narrowing, helix bending, persistence length, ligand or protein induced kinks etc.[47,50]. DNA fine structure prediction appears to be nearing a solution.

Although, the above discussion deals only with proteins and DNA, extensions of the molecular simulation methodology to understand RNA folding[59] and the dynamics of lipid membranes constitute some of the more nascent and upcoming areas. MD simulations combined with computer graphics displays of the trajectories have vividly contributed to a growing awareness that molecular systems are dynamic and not static. Also, the MD method has become an integral part of the tool-chest of X-ray and NMR practioners pursuing structural refinement of biomolecules.

Free Energy

Whether or not an enzyme is stable or a reaction is feasible under given conditions are matters of free energy. Free energy may be rigorously defined[60] in terms of the following expression

$$A = kT \ln < e^{\beta E(\underline{X}^N)} >$$

A is the Helmholtz free energy, $E(\underline{X}^N)$ is the energy due to intermolecular interactions in a given configuration ($\underline{X}^N$) of the system, $\beta = 1/kT$, k is the Boltzmann constant, T is the absolute temperature and the angular brackets refer to (canonical ensemble) Boltzmann averages

An expansion of the above gives

$$A = < E(\underline{X}^N) > - \frac{\beta}{2!}\{< E(\underline{X}^N)>^2 - < E^2(\underline{X}^N) >\} + ...$$

Computationally speaking, each extra term in this expansion involves an order of magnitude more CPU time. Free energy is slower to converge than average internal energy ($E\underline{X}^N$). Alternatively stated,

$$A = kT \ln \int ... \int e^{\beta E(\underline{X}^N)} P(\underline{X}^N) d\underline{X}^N$$

where $P(\underline{X}^N)$ is the normalized Boltzmann probability of observing the system in configuration ($\underline{X}^N$). The right hand side involves a 3N-dimensional integral where N is the number of particles in the system. Examination of the above expression reveals that high energy regions, despite their low probability of occurrence, are important to free energy estimates. These are not adequately sampled (avoided in fact) in conventional molecular simulation (Metropolis Monte Carlo or molecular dynamics simulations of practical length). Another interesting catch is that to evaluate the free energy *via* the above expression, the normalized probabilities have to be known. But to obtain the normalization constant for the probabilities i.e. the

configurational partition function (Z), the free energy has to be known ($Z = e^{-A/kT}$). Free energy has to be known to evaluate free energy! Extension of the above arguments to the Gibbs free energy ($G = A + PV$) is straight forward.

Free energy is a global quantity that depends upon the extent of configuration space accessible to the molecular system. Mechanical properties such as internal energy U, radial distribution function $g(r)$ (related to the Fourier transform of the structure factor) etc. which can be defined on each configuration of the system, converge rapidly in molecular simulation conducted through a suitable sampling procedure. Statistical properties such as entropy and free energy on the other hand cannot be defined on a single configuration but rather depend upon the ensemble of configurations that the system can exhibit under the given external constraints (such as temperature, pressure, volume, and the number of particles in the system). Computation of absolute free energies of a molecular system is virtually impossible. Free energy computation is a tricky issue even though it is overwhelmingly important for chemical and biochemical systems. The problem with free energy predictions and the necessity for inventing byways is put forth elegantly in an early article by Valleau and Torrie[61].

Molecular Simulations

Conventional molecular modelling techniques (molecular mechanics applications) deal with single point energies or at best average internal energies/enthalpies. Free energy is out of bounds in most modelling exercises. Some excellent reviews on free energy studies are available now [60-65]. In most of the studies, free energy is typically considered as a sum of intramolecular and intermolecular contributions, gas phase free energy and solvation free energy, for example. The gas phase free energies, can be evaluated either *via ab initio* calculations, normal mode analysis or via estimates of configurational entropies in the quasi-harmonic approximation[66-68]. The intermolecular contributions such as due to solvation or ion atmosphere are evaluated via one of the several statistical mechanical procedures such as the thermodynamic integration, the probability ratio method, the potential of mean force approach or the statistical mechanical perturbation procedure[60]. The errors involved in such computations are small compared with experiment (on the order of 1 kcal/mol) especially on relative free energy estimates between any two systems which are similar in the sense of their phase spaces. This is extremely gratifying considering the *a priori* nature of these calculations. Studies are also in progress for fine tuning the simulation protocols and results. Absolute free energies are still elusive but frequently an appropriate thermodynamic cycle can be constructed[69,70] to obtain meaningful quantities. In a study of the relative affinity of two inhibitors I_1 and I_2 to an enzyme E, a suitable thermocycle would be

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

$$\begin{array}{ccc} & \Delta G_1 & \\ I_1 + E & \rightarrow & I_1E \\ \Delta G_3 \downarrow & & \downarrow \Delta G_4 \\ I_2 + E & \rightarrow & I_2E \\ & \Delta G_2 & \end{array}$$

It is difficult to compute the free energies of association, ΔG_1 and ΔG_2 of the enzyme with the inhibitors. The quantities ΔG_3 and ΔG_4 however, are computationally tractable. The processes 3 and 4 involve non-physical mutations but free energy being a state function, there is considerable latitude in the design of suitable computer experiments to obtain free energy differences. Applications of the free energy simulation methodology include calculations of relative free energies of solvation, relative pK_a values, conformational preferences in solution etc. Extensions to binding and molecular recognition have also been forthcoming. Some of the recent studies include binding in metalloporphyrin- ligand systems[71], metal ion catalyzed proton transfer in water[72], hydrogen bonding with imides and lactams in chloroform (chemical chameleons)[73], solvent effect on the anomeric equilibrium in D-glucose [74], polyelectrolyte effects in protein- DNA interactions[75], the "Z-phobicity" of A-T base pairs[76], association of K^+ :18-crown-6 complex in water[77] and

cis-trans isomerization of the N-methylated peptide bond of a bacterial collagenase inhibitor[78]. A recent free energy simulation study of Singh *et al.* [79] has unravelled the molecular origins of preferential binding of distamycin to DNA over its 2-Imidazole analogue. An examination of the molecular structures of the two drugs would suggest the contrary. Tidor[80] performed a free energy simulation analysis of the consequences of helix capping interaction in λ-cro. A mutation of Tyr 26 to Asp stabilizes the protein by -1.4 kcal/mol confirming the expectation that an acidic residue is preferred at the N terminal end of an α-helix. Merits and limitations of a decomposition of free energy in terms of specific interactions have recently been analyzed carefully by Karplus[81], van Gunsteren[82] and their coworkers.

Molecular simulations and free energy techniques may be essentially seen as a culmination of the pioneering efforts of Gibbs, Kirkwood, Zwanzig and others. Statistical mechanics has come to fruition in its applications to chemistry and biochemistry. The nature of the problems that confront the practitioners are very different now. To widen the scope of applicability of statistical mechanics, to understand structure and thermodynamics at a molecular level, one needs computationally efficient and accurate methods which can be routinely used by theoreticians or experimentalists on systems of interest. A severe limitation of the simulation techniques is that these are computationally expensive. A regular Metropolis Monte Carlo or a molecular dynamics simulation (seeking structural, dynamic and internal energy information) on a 216 particle liquid water system or a single ion in 215 waters takes about one hour on a Cray Y-MP. A simulation of protein-DNA complex with sufficient waters to provide two layers around the complex takes about 500 hours on Cray Y-MP. The same systems require ten times more CPU when a free energy related quantity is sought. In short, a single free energy determination can take any where from 10 to 1000 CPU hours on a supercomputer depending upon the level of detail and complexity of the problem which puts the free energy predictions starting from a molecular description of the system, beyond the reach of bench chemists and most theoreticians around the globe, without access to Cray. Ways have to be found to cut down on the computational time. If molecular modelling has to have its desired impact on chemical and pharmaceutical industry, development of suitable methodologies for an expeditious evaluation of free energies of molecular systems is essential.

Dielectric Continuum Method

Free energy on the other hand is accessible with relative ease *via* the dielectric continuum approaches, analytical or numerical. Analytical methods have a long history[83]. Born model for ion solvation and Onsager's reaction field approach to dipolar solvation found a generalization in Kirkwood's formulation of the Helmholtz free energy of an arbitrary charge distribution embedded in a dielectric continuum solvent with added salt[84]. Beveridge and Schnuelle[85] reported a concentric dielectric model which can in principle incorporate several layers of solvent with varying "dielectric constants" to account for saturation effects in calculating solvation free energies of arbitrary charge distributions with an overall spherical symmetry. This theory was subsequently extended to other geometries[86-88]. Extensions to Tanford-Kirkwood theory[89] were also reported[90] and results compared with those based on molecular simulations[91]. Solvation and salt effects in the stability of globular proteins, for instance, fall within the purview of such theories. The SATK (static accessibility Tanford- Kirkwood) model[92], which uses a simpler treatment of solvent than that of Beveridge and coworkers but an additional depth parameter has been extensively applied to study protein titration curves, merits and limitations of which are discussed in several places[92-93]. States and Karplus[94] developed an electrostatic continuum solvent model to treat hydrogen exchange behavior in proteins[95]. Molecular simulations have provided new insights into the success of continuum models and have repeatedly validated the underlying approximations in the continuum models[96-99]. The above theories deal with solvent and / or salt effects on an ion, a molecule or a macromolecule. Dielectric continuum theories for binding and interaction have progressed beyond the Coulomb's law only recently[100].

On the numerical front, there is a constant methodological influx for calculating electrostatic contribution to free energies of biomolecular systems. The protein dipole Lanqevin dipole (PDLD) approach[101], the finite boundary element method[102], the finite difference Poisson-Boltzmann (FDPB) method[103-112] and the finite element method[113] are some of the techniques to capture the electrostatics of a given system, the FDPB

method being the most widely used (Delphi package distributed by BIOSYM Technologies /MSI) for quantitative predictions[114-126]. All these are based on solutions (analytical/numerical) to the nonlinear Poisson- Boltzmann equation:

$$\nabla[D(r)\nabla\Phi(r)] - \bar{\kappa}^2 \sinh[\Phi(r)] + 4\pi \rho'(r) = 0$$

Φ in the above equation is the electrostatic potential, D(r) is the dielectric function, κ is a modified Debye-Huckel parameter and ρf is the fixed charge distribution of the molecular system.

The electrostatic potential obtained by solving the above equation is then used to calculate the free energies as a sum over the product of charges and potentials or as a volume integral of the product of electric field and dielectric displacement. In such studies, only the electrostatic component can be calculated. The results have to be supplemented with non-electrostatic contributions to free energy. Continuum methods deserve special attention for free energy estimates in view of the rapidity with which these calculations can be performed. The results however, are sensitive to the choice of parameters such as cavity or van der Waals radii, partial atomic charges etc. These parameters have to be tested against molecular simulations or experiment and particularly the consequences of neglecting molecular solvent have to be closely examined. Force field comparisons with continuum model solvent have been undertaken recently[127,128]. Some recent reviews[129-137] have summarized the existing theoretical tools to treat electrostatic effects in biological molecules. A recent article by Honig and Nicholls[138] sums up the current status on the applications of classical electrostatics to biomolecules.

Empirical Approaches

Other class of methods for estimating free energies comprise empirical techniques for probing solvation effects based on hydration shell volumes [68,139-142] and accessible surface areas[143-145]. These hinge upon the idea that solvation free energies could be estimated with a PV or γA type term where P and V, γ and A denote pressure, volume, surface tension and area respectively. In the excluded volume approach[141,142], hydration shell volumes excluded due to the encroachment of hydration spheres of any two atoms in a given conformation are calculated analytically or numerically. These excluded volumes are then multiplied by an appropriate free energy density parameter to obtain solvation free energies of a given molecular system in a specified conformation. In the accessible area approach[146-151], solvent accessibility (area) of each atom in the molecule is computed and multiplied by an atomic solvation parameter to obtain solvation free energies. This is a particularly useful technique to estimate the hydrophobic contribution to the total free energy of the system. In all these cases, a careful calibration of the parameters involved (free energy densities/atomic solvation parameters etc.) for a given class of compounds is required to obtain quantitatively correct results. The quest for linear free energy relationships in the context of host-guest chemistry is a field in itself[152].

Rate Constants

A number of biomolecular events are diffusion controlled. The rate constants for such reactions are on the order of $10^9 M^{-1}sec^{-1}$ [153]. Brownian dynamics simulations are a natural choice for monitoring macromolecular association in the nanosecond to microsecond regime and for determining the rate constants. Also, Brownian dynamics is appropriate for describing macromolecular motions over distances which are large compared to the solvent size and times that are long compared to the interval between successive solvent impacts. A select few applications[154-156] of this methodology are outlined below.

The enzyme superoxide dismutase (SOD) dismutes the negatively charged superoxide radical with a rate constant of $\sim 10^9 M^{-1}sec^{-1}$. This is extremely high considering that the substrate must react with a copper ion in the active site whose solvent exposure is one thousandth of the total surface area of the protein. Furthermore, the enzyme has a net negative charge (-4 at *p*H 7) suggesting that there should be a net repulsion between the enzyme and the substrate. This repulsion would decrease with increasing ionic strength leading to an increase in the rate. This is contrary to the experimental observations which show that the rate constant decreases by 30% when the ionic strength is increased from 0 to 150mM. Klapper *et al.*[104] considered the

detailed shape of the protein, the dielectric inhomogeneities in the solution and evaluated the electrostatic potentials of the protein using the finite difference Poisson-Boltzmann methodology. A curious focussing of the field lines emanating from the active site and jetting into the solution was noticed Sharp *et al.*[154] subsequently integrated these electric fields into Langevin equations of motion to obtain rate constants. The calculated trends matched perfectly with the experimental observations. These studies also led to certain conjectures on the design of more efficient enzymes to carry out the same task. Diffusion controlled association of cytochrome c (CYTC) and cytochrome c peroxidase(CYP) provides another interesting case. The electron transfer between CYTC and CYP is studied extensively to understand the heme catalyzed reductive clevage of O-O bonds and the coupling of this event to electron transfer in the electron-transport chain. Northrup and coworkers[155] carried out Brownian dynamics simulations on this system to account for the observed rates and to develop a molecular view of the steps in association.

Some of the regulatory proteins bind to their cognate operator (DNA) sequences on the genome with very large rate constants (10^9- 10^{10} M^{-1} sec^{-1}), von Hippel and coworkers[157] argued that such rates could not be due to a random search of the protein for the target site on the DNA. They suggested that protein would find the operator site *via* a search in a reduced dimensional configuration space. Brownian dynamics simulation on protein-DNA association[156] confirmed the above mechanism in addition to predicting rate constants which were consistent with experimental observations.

Kinetic studies are out of bounds in most cases except for diffusion controlled reactions where some structural data are available on the complex. The time scales for the kinetic investigations are mostly out of reach of the present day computer capacity. Simulations at this stage can provide some useful information at a molecular level, on diffusion controlled reactions.

Some Challenges and Perspectives

Docking: Molecular recognition involves docking of two molecules possibly in the presence of solvent and other solutes or salt. Can we dock two biomolecules into a geometrically and energetically *correct* structure based purely on the laws of physics as implemented on a computer ? The answer at this juncture is no. What is solicited from the theoretical studies is the structure of the complex and the free energy of binding as a precursor to probe the elements of recognition at an atomic level. This is quite a formidable problem. Most of the existing methods introduce either subjectivity or some bias. There is no reliable objective docking algorithm as yet. Automated searches require exhaustive computing power and there is no promise of a unique structural solution in a general case. The current status is amply described by some of the recent efforts in this direction which include docking of substrates to proteins using molecular dynamics simulations[158], application of genetic algorithms for docking actinomycin D with deoxyguanosine molecules[159] rational automatic search method (which introduces a hydrogen bond bias) for stable docking models of proteins and ligand[160] and hydrophobic docking to probe molecular recognition[161]. Some simplified force fields for a rapid calculation of the energetics and suitably designed penalty functions have to be built into simulation strategies to obtain the desired results.

Structure-based Drug Design

Earlier studies on drug-receptor systems were configured to monitor the static and dynamic interactions in the complexes[79,162]. Drug design is an extension of the docking and free energy determination problems with a further requirement of optimizing the ligand-receptor interactions *via* ligand modifications to obtain the desired binding constants. Computational strategies work best when the crystal/NMR structure of the receptor is already available. A survey of the range of problems to be tackled, methodologies and examples of drugs developed are presented recently[163-167]. Captopril, an antihypertensive drug is hailed as a successful progeny of computer aided drug design.

Overall, the last two decades have seen a phenomenal increase in the popularity of molecular modelling and a gradual incorporation of the simulation methods into the repertoire of both theoreticians and experimentalists alike probing the mysteries of nature at a molecular level. What we need besides increasing

computational power of some orders of magnitude and a wider access, are some methodological advances which enhance the predictive value of biomolecular modelling.

References

1 (a) G H Grant and W G Richards *Computational Chemistry* Oxford Chemistry Primers Vol.29 Oxford (1995);
(b) D B Boyd and K B Lipkowitz *J chem Edu* **59** (1982) 269-274;
(c) P J Cox *J chem Edu* **59** (1982) 275-277;
(d) U Burkert and N L Allinger *Molecular Mechanics* ACS Monogr No.177, ACM Washington (1982)

2 (a) S J Weiner, P A Kollman, D A Case, U C Singh, C Ghio, G Alagona, S Profeta Jr and P Weiner *J Am chem Soc* **106** (1984) 765-784;
(b) S J Weiner, P A Kollman, D T Nguyen and D A Case *J comp Chem* **7** (1986) 230-252;
(c) W D Cornell, P Cieplak, C I Bayly, I R Gould, K M Merz Jr, D M Ferguson, D C Spellmeyer, T Fox, J W Caldwell and P A Kollman *J Am chem Soc* **117** (1995) 5179-5197

3 (a) B R Brooks, R E Bruccoleri, B D Olafson, D J States, S Swaminathan and M Karplus *J Comp Chem* **4** (1983) 187-217;
(b) A D Mackerrell Jr, J Wiorkiewicz-Kuczera and M Karplus *J Am chem Soc* **117** (1995) 11946-11975

4 (a) J Maple, U Dinur and A T Hagler *Proc Natl Acad Sci (USA)* **85** (1988) 5350-5354;
(b) J R Maple, T S Thacher, U Dinur and A T Hagler, *Chem Des Automation News* **5** (1990) 5-10;
(c) S Lifson, A T Hagler and P Dauber *J Am chem Soc* **101** (1979) 5111;
(d) U Dinur and A Hagler *J Comp Chem* **16** (1995) 154-170

5 (a) F A Momany, L M Carruthers, R F McGuire and H A Scheraga *J phys Chem* **78** (1974) 1595-1620;
(b) F A Momany, R F McGuire, A W Burgess and H A Scheraga *J phy Chem* **79** (1975) 2361-2381;
(c) I Dunfield, A Burgess and H A Scheraga *J phys Chem* **82** (1978) 2609

6 (a) W F van Gunsteren and H J C Berendsen*Groningen Molecular Simulation (GROMOS) System Univ* Groningen, The Netherlands (1987);
(b) W F van Gunsteren, S R Billeter, A A Eising, P H Hunenberger, P Kruger, A E Mark, W R P Scott and I G Tironi *Biomolecular Simulation: The GROMOS96 Manual and User Guide* Univ Gromingen, The Netherlands (1996)

7 (a) N L Allinger *Adv phys org Chem* **13** (1976) 1;
(b) N L Allinger *J Am chem Soc* **99** (1977) 8127-8134;
(c) J H Lii and N L Allinger *J Am chem Soc* **111** (1989) 8566- 8588

8 (a)W L Jorgensen and J Tirado-Rives *J Am chem Soc* **110** (1988) 1657; *J Am chem Soc* **110** (1988) 4212;
(b) W L Jorgensen and J Pranata *J Am chem Soc* **112** (1990) 2008-2010;
(c) J Pranata, S G Wirschke and W L Jorgensen *J Am chem Soc* **113** (1991) 2810-2819;
(d) W L Jorgensen *J Am chem Soc* **103** (1981) 335-340

9 N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Tellerwier and E Teller *J chem Phys* **21** (1953) 1087

10 B J Alder and T E Wainright *J chem Phys* **33** (1960) 1439

11 A Rahman and F H Stillinger *J chem Phys* **53** (1971) 3336-3359

12 M Vasquez, G Nemethy and H A Scheraga *Chem Rev* **94** (1994) 2183-2239 and references therein

13 J A McCammon, B R Gelin and M Karplus *Nature* **267** (1977) 585

14 P J Rossky and M Karplus *J chem Soc* **101** (1979) 1913

15 (a) O. Matsuoka, E Clementi and M Yoshimine *J chem Phys* **64** (1976) 1351;
(b) E Clementi and G Corongiu *Ann N Y Acad Sci* **367** (1981) 83

16 D L Beveridge, P V Maye, B Jayaram, G Ravishanker and M Mezei *J biomol Struc Dyn* **2** (1984) 261-270

17 M P Allen and D J Tildesley *Computer Simulations of Liquids* Clarendon Press Oxford (1987)

18 J A McCammon and S C Harvey *Dynamics of Proteins and Nucleic Acids* Cambridge Univ Press Cambridge (1987)

19 P Kollman *Ann Rev phys Chem* **38** (1987) 303-316

20 (a) C Brooks, M Karplus and B M Pettitt *Adv chem Phys* **71** (1987) 1-259;
(b)M J Field, P A Bash and M Karplus *J Comp Chem* **11** (1990) 700-733

21 W F van Gunsteren and H J C Berendsen *Angew Chem int Ed (Engl)* **29** (1990) 992-1023

22 M Karplus and G A Petsko *Nature* **347** (1990) 631-638

23 B Jayaram and D L Beveridge *Ann Rev Biophys biomol Struct* **25** (1996) 367-394

24 (a) H J C Berendsen, J P M Postma, W F van Gunsteren and J Hermans in:*Intermolecular Forces* Ed B Pullman Reidel Dordrecht (1981) 331;
(b) H J C Berendsen, J R Grigera and T P Straatsma *J phys Chem* **91**(1987) 6269-6271

25 W L Jorgensen, J Chandsekhar, J P Madura, R W Impey and M L Klein *J chem Phys* **79** (1983) 926-935

26 D L Beveridge, S Swaminathan, G Ravishanker, J M Whitka, J Srinivasan, C Prevost, S Louise-May, D R Langley, F M DiCapua, P H Bolton In: *Topics in Molecular and Structural Biology* Vol.17 *Water and Biological Molecules* Ed E Westhoff Macmillan Press Ltd London (1993) 165-285

27 H M Berman *Curr Opin Struct Biol* **4** (1994) 345-350

28 C Tanford *The Hydrophobic Effect* John Wiley & Sons, New York (1980)

29 H R Drew and R E Dickerson *J mol Biol* **151** (1981) 535-556

30 Z Otwinowski, R W Schevitz, R G Zhang, C L Lawson, A Joachimiak , R Q Marmorstein, B F Luisi and P B Sigler *Nature* **335** (1988) 321-329

31 M Mezei and D L Beveridge, in: *Methods in Enzymology, 127:Protons and Water: Structure and Translocation* (Ed L Packer) Academic Press Inc (1986)

32 A Ben Naim *Water and Aqueous Solutions* Plenum Press New York (1974)

33 J Chandrasekhar, S F Smith and W L Jorgensen *J Am chem Soc* **106** (1985) 3049; **107** (1985) 154

34 F C Bernstein, T F Koetzle, G J B Williams, E F Meyer Jr, M D Brice, J R Rogers, O Kennard, T Shimanouchi, M Tasumi *J mol Biol* **112** (1977) 535

35 H M Berman, W K Olson, D L Beveridge, J Westbrook, A Gelbin, T Demeny, S H Hsieh, A R Srinivasan and B Schneider *Biophys J* **63** (1992) 751-759

36 R E Dickerson *Method Enzymol* **211** (1992) 67-111

37 S Neidle *DNA Structure and Recognition* Oxford Univ press (1994)

38 R R Sinden *DNA Structure and Function* Academic Press San Diego (1994)

39 K A Dill *Biochem* **29** (1990) 7133-7155 and references therein

40 C N Pace *Trends Biochem Sci* **15** (1990) 14-17

41 C R Matthews *Curr Opin Struc Biol* **1** (1991) 28-35

42 M Nilges and A T Brunger *Proteins: Structure, Function & Genetics* **15** (1993) 133

43 G E Schulz and R H Schirmer *Principles of Protein Structure* Springer-Verlag New York (1979)

44 C Branden and J Tooze *Introduction to Protein Structure* Garland Publishing Inc New York (1991)

45 W K Olson *in: Topics in Nucleic Acid Structure* Part-2 (Ed S Neidle) McMillan Press London (1982) 1-79

46 W Saenger *Principles of Nucleic Acid Structure* Springer-Verlag New York (1984)

47 (a) M Bansal *Biological Structure and Dynamics Proc Ninth Conversation State Univ* New York Albany (1995) (Ed R H Sarma and M H Sarma) Adenine Press (1996) Vol 1, 121-133;
(b) D Bhattacharrya and M Bansal *J Biomol Struct Dyn* **8** (1990) 539-572

48 S Mohan and N Yathindra *J biomol Struct & Dyn* **9** (1992) 1195-1211

49 (a) R Lavery *Adv Comp Biol* **1** (1994) 69-145;
(b) R Lavery and B Hartmann *Biophys Chem* **50** (1994) 33-45

50 D L Beveridge and G Ravishanker *Curr Opin Struct Biol* **4** (1994) 246-255

51 B Jayaram, N Aneja, E Rajasekaran, V Arora, A Dass, V Ranganathan and V Gupta *J scient ind Res* **53** (1994) 88-105

52 G S Manning *Q Rev Biophys* **11** (1978) 179-246

53 (a) S Arnott and D W L Hukins *J mol Biol* **81** (1973) 93;
(b) S Arnott, R Chandrasekaran, D L Birdsall, A G W Leslie and R L Ratlift *Nature* **283** (1980) 743

54 (a) R V Hosur, G Govil and H Todd Miles *Magnetic Reson Chem* **26** (1988) 927-944;
(b) R V Hosur, K V R Chary, A Saran, G Govil and H Todd Miles *Biopolym* **29** (1990) 953-959;
(c) I Russu *Trend Biotechnol* **9** (1991) 96-104

55 K J McConnell, R, Nirmala, M A Young, G Ravishanker and D L Beveridge *J Am chem Soc* **116** (1994) 4461-4462

56 T E Cheatham, III, J L Miller, T Fox, T A Darden, P A Kollman *J Am chem Soc* **117** (1995) 4193

57 D M York, W Yang, H Lee, T Darden, L G Pederson *J Am chem Soc* **117** (1995) 5001

58 M A Young, B Jayaram and D L Beveridge *J Am chem Soc* **119** (1997) 59-69

59 P Auffinger, S Louise-May, E Westhof *J Am chem Soc* **117** (1995) 6720-6726

60 D L Beveridge and F M Dicapua *Ann Rev Biophys Chem* **18** (1989) 431-492

61 J P Valleau and G M Torrie *in: Modern Theoretical Chemistry,* Vol 5 (Ed B J Berne) Plenum Press New York (1977) Ch-5, 169-194

62 M Mezei and D L Beveridge in: *Computer Simulations of Chemical and Biochemical Systems* (Eds D L Beveridge and W L Jorgensen *Ann N Y Acad Sci* **482** (1986) 1

63 W L Jorgensen *chem Res* **22** (1989) 184-189

64 J A McCommon *Curr Opin Struct Biol Accts***1** (1991) 196-200
65 P Kollman *Chem Rev* **93** (1993)2395-2417
66 M Karplus and J Kushick *Macromol* **14** (1981) 325
67 R M Levy, O Rojas and R A Freisner *J phys Chem* **88** (1984) 4233
68 B Jayaram, G Ravishanker and D L Beveridge *J phys Chem* **92** (1988) 1032-1034
69 B L Tembe and J A McCammon *Comp Chem* **8** (1984) 281-283
70 I Ghosh and J A mcCammon *Biophys J* **51** (1987) 637-641
71 M A Lopez and P A Kollman *J Am chem Soc* **111** (1989) 6212
72 J Aqvist *J phys Chem* **95** (1991) 4587
73 W L Jorgensen and D L Severance *J Am chem Soc* **113** (1991) 209
74 S Ha, J Gao, B Tidor, J W Brady and M Karplus *J Am chem Soc* **113** (1991) 1553
75 B Jayaram, F M DiCapua and D L Beveridge *J Am chem Soc* **113** (1991) 5211-5215
76 S B Singh, D A Pearlman and P A Kollman *J biomol Struct Dyn* **1** (1993) 303-311
77 L X Dang and P A Kollman *J phys Chem* **99** (1995) 55-58
78 B Gilquin, D Perahia and V Dive *Int J Peptide Protein Res* **43** (1994) 384-392
79 S B Singh, Ajay, D E Wemmer and P A Kollman *Proc Natl Acad Sci (USA)* **91** (1994) 7673-7677
80 B Tidor *Proteins* **19** (1994) 310-323
81 S Boresch, G Archontis and M Karplus *Proteins* **20** (1994) 25-33
82 A F Mark and W F van Gunsteren *J mol Biol* **240** (1994) 167-176
83 J Tomasi and M Persico *Chem Rev* **94** (1994) 2027-2094
84 C Tanford *Physical Chemistry of Macromolecules* Wiley New York (1961) Ch-7
85 D L Beveridge and G Schnuelle *J phys Chem* **79** (1975) 2562-2566
86 S W Harrison, J J Nolte and D L Beveridge *J phys Chem* **80** (1976) 2580
87 J S Gomez-Jeria and D Morales-Lagos *J phys Chem* **94** (1990)3790-3795
88 B Jayaram and D L Beveridge *J phys Chem* **94** (1990) 4666-4671
89 C Tanford and J G Kirkwood *J Am chem Soc* **79** (1957)5333-5339
90 B Jayaram and D L Beveridge *Biopolym* **27** (1988) 617-627
91 B Jayaram, M Mezei and D L Beveridge *J Am chem Soc* **110** (1988) 1691-1694
92 J B Matthew *Ann Rev Biophys, biophys Chem* **14** (1985) 387
93 S C Harvey *Proteins* **5** (1989) 78-92
94 D J States and M Karplus *J mol Biol* **197** (1987) 122
95 M Delepierre, C M Dobson, M Karplus, F M Poulsen, D J States and R E Wedin *J mol Biol* **197** (1987) 111-122
96 B Jayaram, R Fine, K Sharp and B Honig *J phys Chem* **93** (1989) 4320-4327
97 B Roux, H-A Yu and M Karplus *J phys Chem* **94** (1990) 4683-4688
98 C Lim, D Bashford and M Karplus *J phys Chem* **95** (1991) 5610
99 W L Jorgensen, J K Buckner, S E Huston and P J Rossky *J Am chem Soc* **109** (1987) 1891-1899
100 B Jayaram *J phys Chem* **98** (1994) 5773-5777
101 (a) J Aqvist and A Warshell *J Am chem Soc* **112** (1990) 2860-2868;
(b) A Warshell and S Russell *Q Rev Biophys* **17** (1984) 238-422
102 A A Rashin *J phys Chem* **93** (1989)4664-4669
103 J Warwicker and H C Watson *J mol Biol* **157** (1982) 671
104 I Klapper, R Hagstrom, R Fine, K Sharp and B Honig *Proteins* **1**(1986) 47
105 M K Gilson, K Sharp and B Honig *J Comp Chem* **9** (1987) 327-335
106 B Jayaram, K Sharp and B Honig *Biopolym* **28** (1989) 975-993
107 M E Davis and J A McCammon *J Comp Chem* **10** (1989) 386-391; **12** (1991) 904-912
108 A Nicholls and B Honig *J Comp Chem* **12**(1991) 435-445
109 B A Luty, M E Davis and J A McCammon *J Comp Chem* **13** (1992) 1114-1118
110 H Oberoi and N Allewell *Biophys J* **65** (1993) 48-55
111 M Holst and F Saied *J Comp chem* **14** (1993) 105-113

112 M Holst, R E Kozack, F Saied and S Subramaniam *Proteins* **18** (1994) 231-245

113 T J You and S C Harvey *J Comp Chem* **14** (1993) 484-501

114 M K Gilson and B H Honig *Nature* **330** (1987) 84-86

115 M K Gilson and B Honig *Proteins* **4** (1988) 7-18

116 S H Northrup, T G Wesnel, C F Meares, J J Wendoloski and J B Matthew *Proc natl Acad Sci (USA)* **87** (1990) 9503-9507

117 A Jean-Charles, A Nicholls, K Sharp, B Honig, A Tempzyck, T Henderson and W C Still *J Am chem Soc* **113** (1991) 1454-1455

118 R A Friedman and B Honig *Biopolymers* **32** (1992) 145-159

119 V Mohan M Davis, J A McCammon and B M Pettitt *J phys Chem* **96** (1992) 6428-6431

120 S W Rick and B J Berne *J Am chem Soc* **110** (1994) 3949-3954

121 E Rajasekaran, B Jayaram and B Honig *J Am chem Soc* **116** (1994) 8238-8240

122 M J Potter, M K Gilson and J A McCammon *J Am chem Soc* **116** (1994) 10298-10299

123 S T Simonson and A T Brunger *J phys Chem* **98** (1994) 4683-4694

124 V Misra, K Sharp, R Friedman and B Honig *J mol Biol* **238** (1994) 245

125 V Misra, J Hecht, K Sharp, R Friedman and B Honig *J mol Biol* **238** (1994) 264

126 A H Elcock and J A McCammon *J Am chem Soc* **117** (1995) 10161-10162

127 S B Dixit, R Bhasin, E Rajasekaran and B Jayaram *J chem Soc Faraday Trans* **93** (1997) 1105-1113

128 N Arora and B Jayaram *J Comp chem* (1997) (in press)

129 N K Rogers *Progr Biophys mol Biol* **48** (1986) 37

130 M E Davis and J A McCammon *Chem Rev* **90** (1990) 509-521

131 A A Rashin *J phys Chem* **94** (1990) 1725-1733

132 K A Sharp and B Honig *Ann Rev Biophys biophys Chem* **19** (1990) 301-332

133 N M Allewell and H Oberoi *Method Enzymol* **202** (1991) 3-19

134 D Bashford *Curr Opin Struct Biol* **1** (1991) 175-184

135 C J Cramer and D G Truhlar *Science* **256** (1992) 213

136 (a) K Sharp *Curr Opin Struct Biol* **4** (1994) 234-239;
(b) M K Gilson *Curr Opin Struct Biol* **5** (1995) 216

137 S R Gadre, S S Pundik and I H Srivastava *Proc Indian Acad sci (chem Sci)* **106** (1994) 303-314

138 B Honig and A Nicholls *Science* **268** (1995) 1144-1149

139 K D Gibson and H A Scheraga *Proc natl Acad Sci (USA)* **58** (1967) 420

140 L G Dunfield, A W Burgess and H A Scheraga *J phys Chem* **82** (1978) 2601

141 Z I Hodes, G Nemethy and H A Scheraga *Biopolymer* **18** (1979) 1565

142 (a) A J Hopfinger *Conformational Properties of Macromolecules* Academic Press, New York (1973) p-70;
(b) A J Hopfinger *Intermolecular Interactions and Biomolecular Organization* Wiley-Interscience New York (1973) 324

143 B Lee and F M Richards *J mol Biol* **55** (1971) 379

144 F M Richards *Ann Rev Biophys Bioengn* **6** (1977) 151-176

145 M L Connolly *Science* **221** (1983) 709-713

146 C Chothia *Nature* **254** (1975) 304-308

147 D Eisenberg and A D McLachlan *Nature* **319** (1986)199

148 K A Sharp, A Nicholls, R F Fine and B Honig *Science* **252** (1991) 106-109

149 M Wesson and D Eisenberg *Protein Sci* **1** (1992) 227-235

150 P K Ponnuswamy and M M Gromiha *J theor Biol* **166** (1994) 63-74

151 C A Schiffer, J W Caldwell, P A Kollman and R M Stroud *Mol Simul* **10** (1993) 121-149

152 D B Smithrud and F Diedrich *J Am chem Soc* **112** (1990) 339-343

153 P W Atkins *Physical Chemistry* ELBS (4th Ed) Oxford Univ Press (1990) Ch-28

154 K Sharp, R Fine and B Honig *Science* **236** (1987) 1460-1463

155 (a) S H Northrup, J C L Reynolds, C M Miller, K J Forrest and J O Boles *J Am chem Soc* **108** (1986) 8162-8170;
(b) S H Northrup, J O Boles and J C L Reynolds *Science* **241** (1988) 67-70

156 B Jayaram, A Das, N Aneja *J mol Struct (Theochem)* **361** (1996) 249-258

157 (a) O G Berg, R B Winter and P H von Hippel *Biochemistry* **20** (1981) 6929;
(b) R B Winter and P H von Hippel *Biochemistry* **20**(1981) 6948;
(c) R B Winter, O G Berg and P H von Hippel *Biochemistry* **20** (1981) 6961

158 A D Nola, D Roccatano and H J C Berendsen *Proteins* **19** (1994) 174-182

159 Y L Xiao and D E Williams *J phys Chem* **98** (1994) 7191- 7200

160 M Y Mizutani, N Tomioka and A Itai *J mol Biol* **243** (1994) 310-326

161 I A Vakser and C A flao *Proteins* **20** (1994) 320-329

162 V Kothekar and S Khurana *Indian J Biochem Biophys* **29** (1992) 330-335

163 M D Walkinshaw *Med Res Rev* **12** (1992) 317-372

164 I D Kuntz *Science* **257** (1992) 1078-1082

165 T R Krugh *Curr Opin struct Biol* **4** (1994) 351-364

166 P M Colman *Curr Opin Struct Biol* **4** (1994) 868-874

167 D DeCamp, R Ogden, I Kuntz and C S Craïk *in: Protein Engineering Principles and Practice* (Eds J L Cleland and C S Craik) Wiley-Liss Inc New York (1996) 467-505