

# Free Energy Simulations Come of Age: Protein–Ligand Recognition

THOMAS SIMONSON\*,<sup>†</sup>  
 GEORGIOS ARCHONTIS,<sup>‡</sup> AND  
 MARTIN KARPLUS\*,<sup>§,||</sup>

*Laboratoire de Biologie et Génomique Structurales (CNRS), IGBMC, 1 rue Laurent Fries, 67404 Illkirch-Strasbourg, France, Department of Physics, University of Cyprus, PO 20537, CY 1678, Nicosia, Cyprus, Laboratoire de Chimie Biophysique, Institut Le Bel, 4 rue Blaise Pascal, Université Louis Pasteur, 67000 Strasbourg, France, and Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138*

Received October 10, 2001

## ABSTRACT

In recent years, molecular dynamics simulations of biomolecular free energy differences have benefited from significant methodological advances and increased computer power. Applications to molecular recognition provide an understanding of the interactions involved that goes beyond, and is an important complement to, experimental studies. Poisson–Boltzmann electrostatic models provide a faster and simpler free energy method in cases where electrostatic interactions are important. We illustrate both molecular dynamics and Poisson–Boltzmann methods with a detailed study of amino acid recognition by aspartyl-tRNA synthetase, whose specificity is important for maintaining the integrity of the genetic code.

## 1. Introduction

The calculation of free energies is among the most important applications of biomolecular simulations. It provides a direct link between the microscopic structure and fluctuations of a system and its most fundamental thermodynamic property, the free energy. Molecular dynamics are widely used for such calculations, even though only equilibrium averages, rather than the dynamics per se, are of interest. Applications of free energy simulations include ligand binding,<sup>11</sup> enzyme reaction paths,<sup>2</sup> protein stability,<sup>3</sup> cooperativity,<sup>4</sup> ion transport,<sup>5</sup>

protein folding,<sup>6</sup> free energy surfaces,<sup>7</sup> and electron transfer.<sup>8</sup> In some cases, the simulations have been used to predict thermodynamic data that were not available for a particular system. However, their main utility has been to obtain additional insights concerning the origin of free energy differences, in synergy with experiment.

Molecular dynamics free energy simulations (MDFE) were first applied to proteins in the mid-1980s.<sup>9,10</sup> Although the early work was plagued by technical difficulties and by the problem of sampling conformational space, the potential of such simulations for providing a molecular understanding of the free energy was recognized at that time.<sup>4</sup> Considerable progress has been made since then.<sup>11–14</sup> Recently, MDFE simulations have been improved by new treatments of long-range electrostatic interactions, by the solution of the outstanding technical problems, and by continued increases in computer power which make it possible to obtain converged results for many cases of interest.

In this article, we illustrate the recent progress by focusing on the important area of protein–ligand binding.<sup>13,14</sup> Noncovalent association between proteins and their ligands is a key element of the biochemistry and information flow in living systems. Many competing effects can contribute to protein–ligand binding.<sup>15</sup> Experimental studies often combine structure determination methods with point mutagenesis and thermodynamic measurements to obtain information on the binding.<sup>16</sup> However, there are considerable difficulties in the experimental analysis of longer-range electrostatic contributions, the cooperativity between protein residues, or disordered solvent, for example. Such effects can be determined from MDFE simulations, as an essential complement to experiments. An illustrative early example of complementary mutation experiments and simulations is the study of substrate recognition by tyrosyl-tRNA synthetase.<sup>17,18</sup>

The basic principle of a protein–ligand binding analysis by MDFE is explained in Figure 1.<sup>19</sup> For a review of the relevant statistical thermodynamics, see refs 15 and 20. Most applications focus on binding free energy differences between a series of ligands or protein mutants. Methods have also been developed to compute absolute binding free energies.<sup>21,22</sup> With fast computers and in favorable cases, it is possible to achieve reasonable conformational sampling and perform a detailed MDFE analysis of a single point mutation in a few days—the same time scale as for the X-ray analysis of a point mutant once the native protein structure is known. With current computer resources and the advances in MDFE methodology reviewed below, the application of free energy simulations to protein–ligand binding, as well as to other problems, can be said to have “come of age” (although “immature” examples still haunt the literature).

Thomas Simonson was born in Chicago, IL, in 1958. He studied physics in France at the Ecole Polytechnique and the University of Paris and spent three postdoctoral years at Yale University. He is currently Directeur de Recherches (Senior Research Scientist) at the CNRS in Strasbourg, France.

Georgios Archontis was born in Athens, Greece, in 1965. After undergraduate studies in physics at the University of Athens, he did his graduate work (Ph.D. in Biophysics, Harvard) and postdoctoral work (University Louis Pasteur, Strasbourg, France) with Martin Karplus. He is currently Lecturer in the Department of Physics, University of Cyprus.

Martin Karplus was born in Vienna, Austria. He was an undergraduate in chemistry and physics at Harvard University and did his graduate work with Linus Pauling at Cal Tech. After a postdoctoral year in Oxford with Charles Coulson, he joined the faculty of the University of Illinois in 1955. He moved to Columbia University in 1960 and to Harvard in 1966. He is now T. W. Richards Research Professor at Harvard and Professeur Conventionné at the University Louis Pasteur in Strasbourg, France.

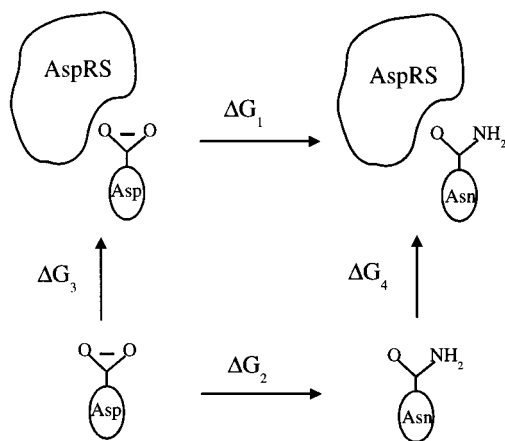
\* Corresponding authors.

<sup>†</sup> Laboratoire de Biologie et Génomique Structurales.

<sup>‡</sup> University of Cyprus.

<sup>§</sup> Institut Le Bel.

<sup>||</sup> Harvard University.



**FIGURE 1.** Thermodynamic cycle describing the binding of Asp and Asn to native aspartyl-tRNA synthetase (AspRS). The binding free energy change can be calculated from either the “alchemical”, horizontal legs or from the “chemical”, vertical legs:  $\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_4 - \Delta G_3$ .

Like experiments, the MDFE method requires a reversible (or near-reversible<sup>23,24</sup>) path between the initial and final states. The free energy change along the horizontal legs in Figure 1 can be obtained by a “thermodynamic integration”:

$$\Delta G(0 \rightarrow 1) = \int_0^1 \frac{dG}{d\lambda} d\lambda = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (1)$$

$H$  represents the energy function, which depends on  $\lambda$ , and the broken brackets represent an average over the ensemble corresponding to  $H(\lambda)$ . At either endpoint,  $\lambda = 0$  or  $1$ , the energy function is that of the native or mutant state; intermediate values correspond to “alchemical” states. The double free energy difference  $\Delta\Delta G$  (Figure 1) can also be obtained from the vertical legs of the cycle, which correspond to “chemical”, i.e., binding reactions, but the simulations are usually more difficult and costly.<sup>1,13,25</sup>

Given the computer-intensive nature of MDFE and the increasing interest in rapid calculations of binding constants (e.g. for drug design<sup>26</sup>), more approximate methods are being developed. One that is particularly promising is based on continuum electrostatics. Such methods estimate the free energy of the end-point states (native, mutant) directly, without passing through intermediate states.<sup>27</sup> A hydrophobic surface area term and/or side chain entropy terms are often added to the continuum electrostatic contribution.<sup>28,29</sup> We refer to this class of methods as the PBF approach (Poisson–Boltzmann free energy). They are of primary interest when electrostatic interactions make the dominant contribution to the binding or binding specificity.<sup>28,30–32</sup>

In this Account, we illustrate recent advances in both MDFE and PBF calculations and the results that can be obtained by a detailed study of amino acid recognition by the enzyme aspartyl-tRNA synthetase (AspRS).<sup>33–36</sup> Aminoacyl-tRNA synthetases (aaRS’s) catalyze the attachment of a specific amino acid (Asp in this case) to a

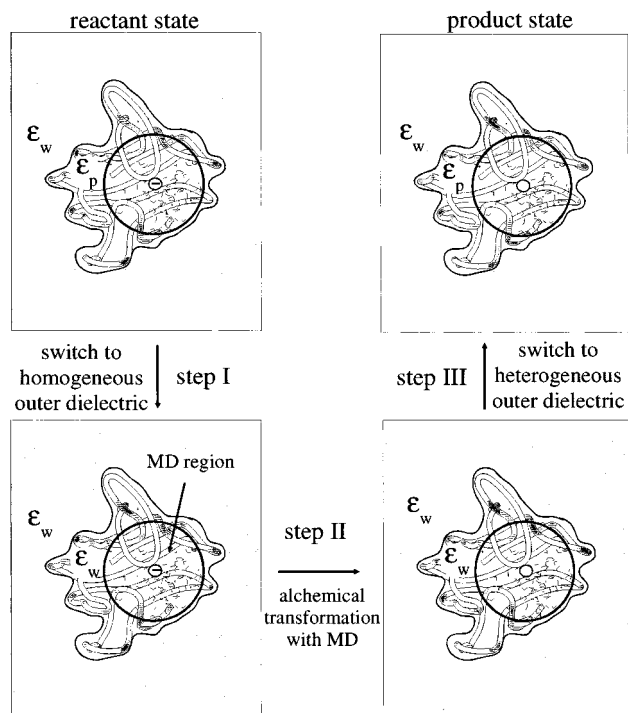
specific tRNA, establishing the amino acid–anticodon relation that defines the genetic code. Specificity for both the tRNA and the amino acid are needed to maintain the integrity of the code. Attempts have been made to engineer aaRS’s with modified specificities, particularly aaRS’s that bind nonnatural amino acids, which could lead to an expansion of the genetic code.<sup>37</sup> The AspRS study<sup>33–36</sup> analyzed the preferential binding of the negatively charged substrate Asp over the neutral analogue Asn with MDFE. To analyze a series of AspRS mutants, we employed a less expensive PBF approach.<sup>36</sup> The combined use of MDFE to study one binding process thoroughly, followed by MD and Poisson–Boltzmann calculations to study a series of related ligands or point mutations, is proposed as a paradigm for protein and ligand design.

## 2. Molecular Dynamics Free Energy Simulations

**2.1. New Electrostatics Treatments.** The key ingredients of an MDFE study are the completeness of conformational sampling and the accuracy of the force field, including the treatment of electrostatics. When the mutation of interest involves a significant rearrangement of charge, it is critical to treat electrostatic interactions accurately. One strategy is to use periodic boundary conditions, with a protein fully solvated in a large box of water, and to compute long-range interactions with either Ewald summation<sup>38</sup> or a continuum reaction field approach.<sup>39</sup> The main drawback of periodic boundaries is the cost associated with the large explicit solvent layer. The “shell” models of Warshel et al. are another possibility, where regions close to the mutation are treated in detail, including explicit solvent; more distant regions are treated as networks of polarizable dipoles, and the most distant regions are treated as a dielectric continuum.<sup>40–42</sup>

To treat the AspRS system accurately with reasonable amounts of computer time, we developed a new inhomogeneous continuum reaction field method (ICRF; Figure 2).<sup>33,35</sup> This method employs a spherical system, including part of the protein and some explicit solvent, initially surrounded by either vacuum<sup>33</sup> or a homogeneous dielectric medium.<sup>35</sup> The mutation is performed using MDFE. In a separate step, the finite model is transferred into the inhomogeneous environment formed by the complete protein and bulk solvent. The free energy for the transfer is obtained from continuum electrostatics.<sup>43</sup> Within the finite MD model, electrostatic interactions are treated without any truncation, with the help of a multipole approximation for distant groups,<sup>44,45</sup> similar to that in refs 41 and 42. A method was proposed recently to include the inhomogeneous reaction field during the MDFE step directly, eliminating the need for a transfer step.<sup>46</sup>

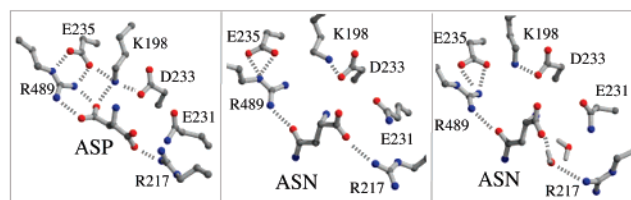
The ICRF approach is similar in spirit to the shell models of Warshel et al. but employs a continuum dielectric environment rather than polarizable dipoles. The efficiency of these methods makes them attractive alternatives to periodic boundary models and Ewald



**FIGURE 2.** Three-step ICRF scheme to compute the free energy change  $\Delta G$  associated with a local transformation in a macromolecule in bulk solution. In the reactant and product states (upper left, right), protein and solvent within a sphere are modeled in atomic detail; protein and solvent outside the sphere are modeled as two distinct dielectric media, with dielectric constants  $\epsilon_p$  and  $\epsilon_w$ , respectively. Step I: the dielectric constant of that portion of the macromolecule that lies outside a spherical inner region (“MD region”) is switched to  $\epsilon_w$ . Step II: the local transformation is carried out using MD. The transformation is schematized by the removal of a “-” sign near the center of the sphere. In this step, the region inside the sphere is treated in atomic detail with MD, while the region outside the sphere corresponds to a homogeneous dielectric medium; i.e., it has a uniform dielectric constant  $\epsilon_w$ . Step III: the dielectric constant of that portion of the macromolecule that lies outside the MD region is switched back to its original value  $\epsilon_p$ . The free energy changes for steps I and III are calculated with a Poisson–Boltzmann continuum model; the free energy change for step II is calculated with MD.

summation. They can be combined with a quantum mechanical treatment of a local region such as the enzyme active site<sup>2,40</sup> to calculate the lowering of the activation barrier; see ref 47 for a recent application to substrate-assisted catalysis by uracil-DNA glycosylase. In another application, dielectric relaxation in the AspRS active site was studied in detail.<sup>48,49</sup>

**2.2. Sampling and Computational Cost.** The cost of an MD simulation depends very much on the system and the goal of the study. If the goal is to interpret qualitatively an experimental number in terms of microscopic interactions, and if the systems of interest (e.g. native and mutant protein) are very similar, then only limited conformational sampling will be needed in most cases, and a few short runs with a small model may suffice. If the goal is to predict accurately and precisely an unknown free energy difference or if the transformation involves large confor-



**FIGURE 3.** Binding mode of Asp (left) and two binding modes of Asn to aspartyl-tRNA synthetase. Center: Asn makes a hydrogen bond (dashed line) to Arg217. Right: Asn is separated from Arg217 by a water layer. The Asn binding modes were calculated to have the same free energy. (This figure was produced with Molscript<sup>85</sup> and Raster3d.<sup>86</sup>)

mational changes, MD simulation can be much more costly. Not only are many long simulations needed but comparisons between different force fields may be necessary to assess the accuracy. One such study examined the binding of the native Raf protein and its Arg89Lys mutant to the signaling protein Ras.<sup>50</sup> Experimentally, only a lower bound (3 kcal/mol) was known for the reduction in binding due to the mutation. With use of multiple MD simulation runs (totaling  $\sim 20$  ns of MD) with three different force fields and careful exploration of the important conformers using biased sampling techniques, it was demonstrated that the reduction in binding is  $3 \pm 2$  kcal/mol, close to the experimental lower bound. Interestingly, the calculations showed that the weaker Raf(Arg89Lys) binding comes from a stronger solvation of the Lys89 side chain in the mutant protein in the unbound state. This is presumably a general effect, which contributes to the lower propensity of Lys, compared to Arg, to participate in protein–protein interfaces. The importance of solvation highlighted here has been demonstrated in other MD simulation studies, e.g. ref 18.

To understand the molecular mechanism of AspRS amino acid specificity, the binding free energies of the charged substrate Asp and the neutral analogue Asn were compared.<sup>33–36</sup> Neither the Asn binding constant nor the AspRS:Asn X-ray structure was known experimentally, so that predictions were sought. As discussed, this system involved a difficult electrostatic problem. In addition, as became clear from the simulations, it poses two conformational sampling problems. First, each ligand has three stable conformations in solution, corresponding to different side chain rotamers. To correctly estimate the Asp  $\rightarrow$  Asn free energy change in solution, all three conformers must be adequately sampled, using biased sampling techniques.<sup>51</sup> Second, the Asn ligand was found to have a large mobility in the binding pocket of AspRS and a different binding mode from Asp (Figure 3). Thus, extensive sampling was also needed for the AspRS:Asn complex. Nine MD simulation runs were done for the mutation in the complex, totaling 5 ns of dynamics and leading to an acceptable statistical uncertainty for this rather difficult problem ( $\pm 3$  kcal/mol out of a total binding free energy difference of 15 kcal/mol). Some of the results are described in section 2.4. These examples show that for difficult cases, and especially when a prediction is being made, a large number of simulations may be necessary.

**2.3 Bonded Contributions.** MDFF simulations have also been improved by the solution of several technical problems, including theoretical analyses of the van der Waals end-point singularity,<sup>52,53</sup> rigorous treatments of holonomic constraints,<sup>54,55</sup> methods to calculate absolute binding constants,<sup>13,21,22</sup> and a clarification of the significance of free energy components.<sup>56,57</sup>

Another advance is the in-depth analysis of stereochemical energy terms (bonds, angles) in free energy calculations.<sup>58–60</sup> When an alchemical mutation is made for a chemical group, e.g. Asp → Asn, two general approaches can be employed. A hybrid amino acid can be introduced, possessing both a native and a mutant side chain connected to the same backbone (dual topology method).<sup>33,34</sup> Alternatively, a single side chain can be introduced, resembling Asp (in this example) but possessing two “dummy” atoms attached to the side chain oxygen that is to be transformed into an NH<sub>2</sub> group. The stereochemistry must be gradually adjusted by modifying the parameters of certain bonded energy terms, such as the length and force constant of the side chain C–N bond in this case. Several early MDFF applications contained errors or uncertainty associated with these bonded contributions. A thorough discussion of these problems has been given.<sup>59,60</sup> It was shown that when correctly analyzed and with sufficient sampling, bonded terms can be calculated to a high accuracy, so that single and dual topology methods give consistent results. The bonded contributions were also analyzed analytically using the rigid rotor harmonic oscillator model.<sup>58</sup> This model yields an exact separation into “vibrational” contributions (associated with changing force constants), “Jacobian” factors (associated with changing geometric parameters and their effect on the motions of the bonded atoms themselves), and “potential-of-mean force” contributions (associated with the response of the environment to changing geometric parameters). The uncertainty that has plagued bonded free energy contributions now appears to be resolved.

**2.4. MDFF Yields Structures and Free Energy Components.** In experimental studies of molecular recognition, an important goal is to interpret the overall binding free energy in terms of specific structural groups, such as individual hydrogen bond partners, and specific physical effects, such as electrostatic forces. X-ray and NMR structure determination reveal the existence (though not the strengths) of short-range interactions between individual groups; long-range electrostatic interactions are difficult to determine in this way. Group contributions to the binding affinity can be estimated from a series of point mutations, as in alanine scanning experiments.<sup>16,61</sup> Double mutant cycles can be used to measure “coupling” between residues.<sup>61,62</sup> Dynamical changes can sometimes be detected, e.g. by NMR,<sup>63</sup> and related to entropic effects. These approaches, though very useful, have limitations. For example, when ligand binding is too weak or too strong, thermodynamic measurements are not possible,<sup>64</sup> so that a mutation that abolishes detectable binding cannot be characterized.<sup>50</sup> Solvent contributions are usu-

ally not measurable separately; they are often inferred by considering correlations of the binding free energy with side chain size or surface area for a series of mutations.<sup>65</sup>

An advantage of simulations is that a single computer “experiment” provides not only the binding free energy difference but also a dynamical structural model of the system (including solvent) and detailed information concerning the contributions of different system parts (e.g. individual amino acids) and different energy terms (e.g. electrostatics, van der Waals). Because molecular mechanics energy functions commonly take the form of sums over small groups of atoms and because eq 1 is linear,  $\Delta G$  can be expressed as a sum over groups of atoms.<sup>4</sup> While these free energy “components” depend on the integration path used to connect the end-point states,<sup>12,56,66</sup> they can provide useful insights;<sup>56</sup> moreover, the choice of the path provides an additional degree of freedom that can be used to obtain information about different physical processes (see below).

In the AspRS study,<sup>33–36</sup> the various improvements in MDFF methodology described above were included: correct treatment of long-range electrostatics (including the effect of bulk solvent); van der Waals end-point effects; bonding terms. The Asp → Asn free energy difference was calculated to be  $95 \pm 3$  kcal/mol in the protein complex and  $80 \pm 2$  kcal/mol in solution. The two free energy differences are mostly due to the loss of electrostatic interactions with the Asp side chain carboxylate. The protein is predicted to “solvate” the Asp side chain 15 kcal/mol more strongly than bulk solvent does, a striking example of charge stabilization, which contrasts with the view that electrostatic interactions are not important for binding.<sup>28,29</sup> A free energy component analysis<sup>4,56</sup> showed that the free energy difference in the complex arises mainly from electrostatic interactions with four charged side chains, Arg489/Glu235 and Lys198/Asp233, which form two dipolar groups that contact the Asp side chain directly (Figure 3). Thus, dipoles, rather than salt bridges per se, are used to specifically stabilize Asp, while large contributions from more distant charged groups cancel almost exactly. Interestingly, three distinct Asn binding modes were observed which all had essentially the same free energy (Figure 3).

Another study considered the effect of the Asn47Leu mutation on the copper oxidation potential of azurin.<sup>57</sup> This case is of particular interest because both legs of the thermodynamic cycle (analogous to Figure 1) could be calculated. The horizontal, “alchemical”, and vertical, “chemical”, runs agreed approximately but gave very different free energy components, because they correspond to different physical processes.

### 3. Poisson Boltzmann Free Energy Calculations (PBFE)

Continuum models are being increasingly used to study protein–ligand recognition.<sup>67</sup> Many studies have considered series of similar ligands or protein mutants and

focused on binding free energy differences. This leads to cancellation of most of the nonelectrostatic contributions. We use the AspRS study to illustrate this type of application.

**3.1. Extending the Scope of MD FE.** We extended the scope of the AspRS MD FE analysis by using a combination of MD simulations and Poisson–Boltzmann free energy calculations.<sup>36</sup> The results illustrate the power, but also some important limitations and pitfalls, of such methods. Several point mutations in the amino acid binding pocket were considered. For each mutant, an MD simulation with explicit solvent was performed to generate structural models. For each ligand (Asp, Asn), free energies of the bound and separated states were calculated by the finite-difference Poisson–Boltzmann method and subtracted. The difference between Asp and Asn gives an estimate of the electrostatic contribution to the binding free energy difference  $\Delta\Delta G$  between ligands, which should approximate the total  $\Delta\Delta G$ , because nonelectrostatic terms largely cancel.<sup>34</sup> PBFE, unlike MD FE, calculates  $\Delta\Delta G$  from the vertical legs of the thermodynamic cycle in Figure 1.

A crucial parameter in these calculations is the dielectric constant  $\epsilon$  of the solutes, assumed to be the same for all solutes and all states. It is difficult,<sup>68</sup> though not impossible,<sup>48,67,69</sup> to evaluate the dielectric “constant” of a protein; in fact it does not have to be spatially constant,<sup>69</sup> and its effective value may depend strongly on the set of atomic charges used (the force field) and the process considered.<sup>48</sup> Therefore, an empirical approach is preferred. The strategy used in this work was to adjust  $\epsilon$  to reproduce with PBFE the  $\Delta\Delta G$  obtained from MD FE for Asp/Asn binding to native AspRS (15 kcal/mol). This leads to  $\epsilon \approx 4$ , a reasonable value which has been used in many, though not all,<sup>30,70</sup> other studies. With the same value for  $\epsilon$ , good agreement was also found for the Lys198Leu mutant of AspRS, where the MD FE and PBFE  $\Delta\Delta G$ 's are 0.2 and 0.6 kcal/mol, respectively.

Another key ingredient is the set of structural models used in the calculations. It was very important to use, for each state, structures corresponding to that state. If the native structure was used to calculate properties of the Lys198Leu mutant protein, for example, large errors were obtained; specifically, the Asp binding decreased by 10 kcal/mol. It was also important to average results for each state over several structures, taken from a simulation of that state. Indeed, the PBFE results are very sensitive to the details of the structure, so that free energies from instantaneous structures (or the X-ray structure) can deviate by 4–5 kcal/mol from the ensemble average.<sup>36</sup>

The MD FE and PBFE free energy components were compared for the first time. With MD FE, solvent is treated explicitly (except for distant, bulk solvent, treated as a continuum in this work); with PBFE, all the solvent is treated implicitly. The free energy components for the protein residues are smaller in the PBFE case, because electrostatic shielding by the implicit solvent is “folded into” the protein components. Nevertheless, the four most important residues identified by MD FE also make the largest contributions to  $\Delta\Delta G$  with PBFE.

This system illustrates several principles that are important for protein–ligand binding. The mutations were found to induce structural shifts in the binding pocket that could not have been predicted from the native structure alone. Residues that were not mutated then made new interactions with the ligands and, consequently, made large contributions to the change in binding free energies. An example is the Lys198Leu mutation. It removes the strong Asp–Lys198 interaction and induces a shift of the Asp ligand (toward the right of Figure 3), strengthening its interactions with Asp233 and Arg217 and weakening its interaction with Arg489. Although the Lys198 contribution to the Asp binding free energy is  $-13$  kcal/mol in the native state, the Asp affinity is actually calculated to *increase* slightly due to the mutation. The lost Lys198 contribution is partly compensated by the new interactions; e.g. the Asp233 contribution to binding decreases from  $+5$  to  $-3$  kcal/mol. In addition, removal of the Lys198 positive charge decreases the electrostatic penalty for desolvating the binding site by 4 kcal/mol, making it easier to bind Asp (or Asn). The greater cost to desolvate charged Asp, compared to neutral Asn, contributes  $-5$  kcal/mol to  $\Delta\Delta G$  (disfavoring Asp binding). In cases such as these, the changes in the individual binding free energies, the double free energy differences  $\Delta\Delta G$ , and the specificity change, all depend on the details of the structural shifts and can be much larger than the direct contribution of the mutated residue(s) (as would be calculated by the very efficient, but sometimes inaccurate, method of “alanine scanning” in MD FE<sup>71</sup>).

**3.2. Other Applications.** We mention for reference several other recent studies that show that continuum electrostatics can yield important qualitative or semiquantitative information on protein–ligand binding. Hendsch and Tidor developed a novel PBFE component analysis<sup>32</sup> and applied it to GCN4 leucine zipper assembly. Electrostatic coupling between the binding of protons and small ligands to proteins has been studied.<sup>72,73</sup> Sensitivity analysis and charge optimization have been used to analyze or alter binding specificity.<sup>74,75</sup>

Two studies employed free energy functions consisting of a Poisson–Boltzmann electrostatic solvation term, plus additional contributions (a cavity term proportional to surface area, a molecular mechanics solute energy, and a harmonic or quasiharmonic solute vibrational entropy<sup>70,76</sup>). With this PBFE variant (“MM-PBSA”), reasonable agreement with experiment was obtained in several studies involving proteins and small ligands, using solute dielectrics of 1<sup>30</sup> or 4.<sup>77</sup> As in ref 36, MD simulations were performed for each ligand or protein variant. Other applications did not use separate structural models for each variant;<sup>71</sup> this can lead to large errors,<sup>78</sup> as suggested by the AspRS study (see above).

## 4. Conclusions

Free energy simulations, which were first applied to proteins in refs 9 and 10, represent an important application of molecular dynamics to analyze thermodynamic

properties that play a fundamental role in living systems. MDFE studies have provided many insights that were not available from experiment. An essential aspect of such studies is the decomposition of the total free energy into its components by the use of thermodynamic integration. The fact that these components depend on the choice of the integration path makes it possible to focus on specific aspects of the phenomena of interest.

We have entitled this article “Free Energy Simulations Come of Age”. This is appropriate because the technical problems that plagued earlier MDFE applications have been resolved by the careful work of many people and by increased computer power, which makes it possible to extend the simulation times into the range required for convergence. Of course, there are still sampling difficulties when the process under study involves large (unknown) conformational changes. Recent extensions of MDFE have included absolute binding free energies and entropies,<sup>79</sup> free energy profiles for unbinding reactions,<sup>25</sup> and reconstruction of such profiles from irreversible pulling experiments.<sup>24</sup> PBFE methods, on the other hand, will benefit from improved numerical methods<sup>67,80</sup> and from more sophisticated treatments of hydrophobic interactions.<sup>81,82</sup>

The future of free energy simulations as a tool in structural biology is bright. With the increased realization that the faster and simpler Poisson–Boltzmann methods (PBFE) can in many cases complement MDFE, particularly when combined with molecular dynamics sampling, the range of applications is expected to extend to many additional aspects of molecular biology. As high-resolution structures and other experimental data are obtained for new proteins and protein–protein complexes, such as those involved in cellular regulation, new questions will arise requiring answers that can only be provided by such simulations. Many of these concern only thermodynamic aspects, while others require an understanding of the dynamics.<sup>83</sup> Both thermodynamics and dynamics are a consequence of the potential energy surfaces involved. In this review, we have focused on the thermodynamics. Since reactions involve the motion on potential energy surfaces and the rates are largely governed by the free energy along the reaction path, corresponding MDFE and PBFE simulations will play an increasing role in these areas. Methods for finding reaction paths exist for the study of conformational changes, such as the T → R transition in hemoglobin or the closed → open transition in GroEL,<sup>84</sup> but the free energy changes along the path and the transition rates have still to be calculated. Other cases of interest are the profile for proton transfer in bacteriorhodopsin and the photosynthetic reaction center. Protein–protein interactions represent an intense field of experimental investigation, including transcription factors, kinases, and the immune system, where free energy simulations are needed to answer many questions. As molecular biology becomes really molecular (one might even use the term “chemical biology”), there will be an increasing need for free energy simulations. Of particular interest already is the decomposition of the free energy involved in the formation of microtubules and actin

filaments; a recent PBFE study of these systems<sup>80</sup> indicates one direction for the future.

*M.K. acknowledges partial support by the National Institutes of Health, the CNRS, and the Ministère de l'Éducation. T.S. acknowledges support from the CNRS and the Ministère de la Recherche. G.A. acknowledges support from the Cyprus Government.*

## References

- (1) Miyamoto, S.; Kollman, P. Absolute and relative binding free energy calculations of the interaction of biotin and its analogues with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins* **1993**, *16*, 226–245.
- (2) Schweins, T.; Langen, R.; Warshel, A. Why have mutagenesis studies not located the general base in ras p21? *Nat. Struct. Biol.* **1994**, *1*, 476–484.
- (3) Tidor, B.; Karplus, M. Simulation analysis of the stability mutant R96H of T4 lysozyme. *Biochemistry* **1991**, *30*, 3217–3228.
- (4) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden thermodynamics of mutant proteins: A molecular dynamics analysis. *Science* **1989**, *244*, 1069–1072.
- (5) Roux, B.; Karplus, M. Ion transport in a gramicidin channel: structure and thermodynamics. *Biophys. J.* **1991**, *59*, 4856–4868.
- (6) Boczeko, E. M.; Brooks, C. L. First principles calculation of the folding free energy of a three helix bundle protein. *Science* **1995**, *269*, 393–396.
- (7) Bartels, C.; Schaefer, M.; Karplus, M. Determination of equilibrium properties of biomolecular systems using multidimensional adaptive umbrella sampling. *J. Chem. Phys.* **1999**, *111*, 8048–8067.
- (8) Muegge, I.; Qi, P.; Wand, A. J.; Chu, Z.; Warshel, A. Reorganization energy of cytochrome *c* revisited. *J. Phys. Chem. B* **1997**, *101*, 825–836.
- (9) Wong, C.; McCammon, J. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3832.
- (10) Warshel, A.; Sussman, F.; King, G. Free energy changes in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* **1986**, *25*, 8368–8372.
- (11) Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395.
- (12) van Gunsteren, W.; Beutler, T.; Fraternali, F.; King, P.; Mark, A.; Smith, P. Computation of free energy in practice: choice of approximations and accuracy limiting factors. In *Computer simulation of biomolecular systems*; van Gunsteren, W., Weiner, P., Wilkinson, A., Eds.; Escrom Science Publishers: Leiden, The Netherlands, 1993; pp 315–348.
- (13) Lamb, M. L.; Jorgensen, W. Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449–457.
- (14) McCammon, J. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.* **1998**, *8*, 245–249.
- (15) Gilson, M.; Given, J.; Bush, B.; McCammon, J. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047–1069.
- (16) Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*; Freeman: New York, 1999.
- (17) Fersht, A.; Shi, J.; Knill-Jones, J.; Lowe, D.; Wilkinson, A.; Blow, D.; Brick, P.; Carter, P.; Waye, M.; Winter, G. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **1985**, *314*, 235–238.
- (18) Lau, F.; Karplus, M. Molecular recognition in proteins. Simulation analysis of substrate binding by a tyrosyl-tRNA synthetase mutant. *J. Mol. Biol.* **1994**, *236*, 1049–1066.
- (19) Tembe, B.; McCammon, J. Ligand-receptor interactions. *Comput. Chem.* **1984**, *8*, 281–283.
- (20) Simonson, T. Free energy calculations. In *Computational Biochemistry & Biophysics*; Becker, O., Mackerell, A., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker: New York, 2001; Chapter 9.
- (21) Roux, B.; Nina, M.; Pomes, R.; Smith, J. Thermodynamic stability of water molecules in the Bacteriorhodopsin proton channel: a molecular dynamics and free energy perturbation study. *Biophys. J.* **1996**, *71*, 670–681.
- (22) Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2702–2714.
- (23) Reinhardt, W.; Miller, M.; Amon, L. Why is it so difficult to simulate entropies, free energies, and their differences? *Acc. Chem. Res.* **2001**, *34*, 607–614.

- (24) Hummer, G.; Szabo, A. Free energy reconstruction from non-equilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658–3661.
- (25) Palma, R.; Himmel, M.; Brady, J. Calculation of the potential of mean force for the binding of glucose to benzene in aqueous solution. *J. Phys. Chem. B* **2000**, *104*, 7228–7234.
- (26) Ajai, and Murcko, M. Computational methods to predict binding free energy in ligand–receptor complexes. *J. Med. Chem.* **1995**, *38*, 4953–4967.
- (27) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144.
- (28) Hendsch, Z.; Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatics analysis. *Protein Sci.* **1994**, *3*, 211–226.
- (29) Frolloff, N.; Windemuth, A.; Honig, B. On the calculation of binding free energies using continuum methods: application to the MHC class I protein-peptide interactions. *Protein Sci.* **1997**, *6*, 1293–1301.
- (30) Chong, L.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. Molecular dynamics and free energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14330–14335.
- (31) Sharp, K. Calculation of HyHel10-lysozyme binding free energy changes: effect of ten point mutations. *Proteins* **1998**, *33*, 39–48.
- (32) Hendsch, Z.; Tidor, B. Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. *Protein Sci.* **1999**, *8*, 1381–1392.
- (33) Simonson, T.; Archontis, G.; Karplus, M. Continuum treatment of long-range interactions in free energy calculations. Application to protein–ligand binding. *J. Phys. Chem. B* **1997**, *101*, 8349–8362.
- (34) Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. Specific amino acid recognition by aspartyl-tRNA synthetase studied by free energy simulations. *J. Mol. Biol.* **1998**, *275*, 823–846.
- (35) Simonson, T. Electrostatic free energy calculations for macromolecules: a hybrid molecular dynamics/continuum electrostatics approach. *J. Phys. Chem. B* **2000**, *104*, 6509–6513.
- (36) Archontis, G.; Simonson, T.; Karplus, M. Binding free energies and free energy components from molecular dynamics and Poisson–Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol.* **2001**, *306*, 307–327.
- (37) Liu, D.; Schultz, P. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4780–4785.
- (38) Sagui, C.; Darden, T. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.
- (39) Oostenbrink, B.; Pitera, J.; van Lipzig, M.; Meerman, J.; van Gunsteren, W. Simulations of the estrogen receptor ligand-binding domain: affinity of natural ligands and xenoestrogens. *J. Med. Chem.* **2000**, *43*, 4594–4605.
- (40) Warshel, A. *Computer modelling of chemical reactions in enzymes and solutions*; John Wiley: New York, 1991.
- (41) Åqvist, J.; Luzhkov, V. Ion permeation mechanism of the potassium channel. *Nature* **2000**, *404*, 881–884.
- (42) Florian, J.; Goodman, M.; Warshel, A. Free energy perturbation calculations of DNA destabilization by base substitutions: the effect of neutral guanine–thymine, adenine–cytosine, and adenine–difluorotoluene mismatches. *J. Phys. Chem. B* **2000**, *104*, 10092–10099.
- (43) Resat, H.; McCammon, J. Free energy simulations: correcting for electrostatic cutoffs by use of the Poisson equation. *J. Chem. Phys.* **1996**, *104*, 7645–7651.
- (44) Stote, R.; States, D.; Karplus, M. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys.* **1991**, *88*, 2419–2433.
- (45) Lee, F.; Warshel, A. A local reaction field method for fast evaluation of long range interactions in molecular simulations. *J. Chem. Phys.* **1992**, *95*, 4366–4377.
- (46) Im, W.; Bernèche, S.; Roux, B. Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* **2001**, *114*, 2924–2937.
- (47) Dinner, A.; Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* **2001**, *413*, 752–755.
- (48) Simonson, T.; Archontis, G.; Karplus, M. A Poisson–Boltzmann study of charge insertion in an enzyme active site: the effect of dielectric relaxation. *J. Phys. Chem. B* **1999**, *103*, 6142–6156.
- (49) Archontis, G.; Simonson, T. Dielectric relaxation in an enzyme active site: molecular dynamics simulations interpreted with a macroscopic continuum model. *J. Am. Chem. Soc.* **2001**, *123*, 11047–11056.
- (50) Zeng, J.; Fridman, M.; Maruta, H.; Treutlein, H. R.; Simonson, T. Protein–protein recognition: an experimental and computational study of the R89K mutation in Raf and its effect on Ras binding. *Protein Sci.* **1999**, *8*, 50–64.
- (51) Tobias, D. J.; Brooks, C. L. Calculation of free energy surfaces using the methods of thermodynamic perturbation theory. *Chem. Phys. Lett.* **1987**, *142*, 472–476.
- (52) Simonson, T. Free energy of particle insertion. an exact analysis of the origin singularity for simple liquids. *Mol. Phys.* **1993**, *80*, 441–447.
- (53) Beutler, T.; Mark, A.; Schaik, R. v.; Gerber, P.; Gunsteren, W. v. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (54) Carter, E.; Ciccotti, G.; Hynes, J.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- (55) Otter, W. d.; Briels, W. The calculation of free energy differences by constrained molecular dynamics simulations. *J. Chem. Phys.* **1998**, *109*, 4139–4146.
- (56) Boreesch, S.; Archontis, G.; Karplus, M. Free energy simulations: the meaning of the individual contributions from a component analysis. *Proteins* **1994**, *20*, 25–33.
- (57) Boreesch, S.; Karplus, M. The meaning of component analysis: decomposition of the free energy in terms of specific interactions. *J. Mol. Biol.* **1995**, *254*, 801–807.
- (58) Boreesch, S.; Karplus, M. The Jacobian factor in free energy simulations. *J. Chem. Phys.* **1996**, *105*, 5145–5154.
- (59) Boreesch, S.; Karplus, M. The role of bonded terms in free energy simulations. 1. Theoretical analysis. *J. Phys. Chem. A* **1999**, *103*, 103–118.
- (60) Boreesch, S.; Karplus, M. The role of bonded terms in free energy simulations. 2. Calculation of their influence on free energy differences of solvation. *J. Phys. Chem. A* **1999**, *103*, 119–136.
- (61) Pons, J.; Rajpal, A.; Kirsch, J. Energetic analysis of an antibody/antigen interface: alanine scanning mutagenesis and double mutant cycles on the HyHEL-10/lysozyme interaction. *Protein Sci.* **1999**, *8*, 958–968.
- (62) Horovitz, A. Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Folding Des.* **1996**, *1*, R121–126.
- (63) Stone, M. NMR relaxation studies of the role of conformational entropy on protein stability and ligand binding. *Acc. Chem. Res.* **2001**, *34*, 379–388.
- (64) Xie, D.; Gulnick, S.; Erickson, J. Dissection of binding energy with native and ligand-bound protein stabilities: determining the affinity of ultratight-binding inhibitors of HIV-1 protease and its drug-resistant mutants. *J. Am. Chem. Soc.* **2000**, *122*, 11533–11534.
- (65) Sundberg, E. J.; Urrutia, M.; Braden, B. C.; Isern, J.; Tsuchiya, D.; Fields, B. A.; Malchiodi, E. L.; Tormo, J.; Schwarz, F. P.; Mariuzza, R. A. Estimation of the hydrophobic effect in an antigen–antibody protein–protein interface. *Biochemistry* **2000**, *39*, 15375–15387.
- (66) Simonson, T.; Brünger, A. T. Thermodynamics of protein–peptide binding in the ribonuclease s system studied by molecular dynamics and free energy calculations. *Biochemistry* **1992**, *31*, 8661–8674.
- (67) Simonson, T. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.
- (68) Schütz, C.; Warshel, A. What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins* **2001**, *8*, 211–217.
- (69) Simonson, T.; Perahia, D. Internal and interfacial dielectric properties of cytochrome c from molecular dynamics simulations in aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082–1086.
- (70) Srinivasan, J.; Cheatham, T.; Cieplak, P.; Kollman, P.; Case, D. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- (71) Massova, I.; Kollman, P. Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133–8143.
- (72) Mackerell, A.; Sommer, M.; Karplus, M. pH dependence of binding reactions from free energy simulations and macroscopic continuum electrostatics calculations: application to the 2′GMP/3′GMP binding to Ribonuclease T1 and implications for catalysis. *J. Mol. Biol.* **1995**, *247*, 774–807.
- (73) Spassov, V.; Bashford, D. Multiple-site ligand binding to flexible macromolecules: separation of global and local conformational change and an iteratively mobile clustering approach. *J. Comput. Chem.* **1999**, *20*, 1091–1111.
- (74) Lee, L.; Tidor, B. Barstar is electrostatically optimized for tight binding to barnase. *Nat. Struct. Biol.* **2001**, *8*, 73–76.

- (75) Wong, C.; Hunenberger, P.; Akamine, P.; Narayana, N.; Diller, T.; McCammon, J.; Taylor, S.; Xuong, N. Computational analysis of PKA–balanol interactions. *J. Med. Chem.* **2001**, *44*, 1530–1539.
- (76) Vorobjev, Y.; Almagro, J.; Hermans, J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with explicit solvent and an implicit solvent continuum model. *Proteins* **1998**, *32*, 399–413.
- (77) Wang, W.; Lim, W.; Jakalian, A.; Wang, J.; Luo, R.; Bayly, C.; Kollman, P. An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. *J. Am. Chem. Soc.* **2001**, *123*, 3986–3994.
- (78) Reyes, C.; Kollman, P. Investigating the binding specificity of U1A–RNA by computational mutagenesis. *J. Mol. Biol.* **2000**, *295*, 1–6.
- (79) Luo, R.; Gilson, M. Synthetic adenine receptors: direct calculation of binding affinity and entropy. *J. Am. Chem. Soc.* **2000**, *122*, 2934–2937.
- (80) Baker, N.; Sept, D.; Simpson, J.; Holst, M.; McCammon, J. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (81) Hummer, G. Hydrophobic force field as a molecular alternative to surface area models. *J. Am. Chem. Soc.* **1999**, *121*, 6299–6305.
- (82) Huang, D.; Chandler, D. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8324–8327.
- (83) Young, M.; Gonfloni, S.; Superti-Furga, G.; Roux, B.; Kuriyan, J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **2001**, *105*, 115–126.
- (84) Ma, J.; Sigler, P.; Xu, Z.; Karplus, M. A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* **2000**, *302*, 303–319.
- (85) Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.
- (86) Merritt, E.; Murphy, M. Raster 3D: a program for photorealistic molecular graphics. *Acta Crystallogr., Sect. D* **1994**, *50*, 869–873.

AR010030M