doi:10.1006/jmbi.2000.4285 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **306**, 307–327

JMB

Binding Free Energies and Free Energy Components from Molecular Dynamics and Poisson-Boltzmann Calculations. Application to Amino Acid Recognition by Aspartyl-tRNA Synthetase

Georgios Archontis[1]*, Thomas Simonson[2]* and Martin Karplus[3,4]*

[1]*Department of Physics University of Cyprus PO 20537, CY 1678, Nicosia Cyprus*

[2]*Laboratoire de Biologie et Génomique Structurales (C.N.R.S.), I.G.B.M.C., 1 rue Laurent Fries, 67404, Illkirch-Strasbourg, France*

[3]*Laboratoire de Chimie Biophysique, Institut Le Bel, 4 rue Blaise Pascal, Université Louis Pasteur 67000, Strasbourg, France*

[4]*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St. Cambridge, MA 02138, USA*

Specific amino acid binding by aminoacyl-tRNA synthetases (aaRS) is necessary for correct translation of the genetic code. Engineering a modified specificity into aminoacyl-tRNA synthetases has been proposed as a means to incorporate artificial amino acid residues into proteins *in vivo*. In a previous paper, the binding to aspartyl-tRNA synthetase of the substrate Asp and the analogue Asn were compared by molecular dynamics free energy simulations. Molecular dynamics combined with Poisson-Boltzmann free energy calculations represent a less expensive approach, suitable for examining multiple active site mutations in an engineering effort. Here, Poisson-Boltzmann free energy calculations for aspartyl-tRNA synthetase are first validated by their ability to reproduce selected molecular dynamics binding free energy differences, then used to examine the possibility of Asn binding to native and mutant aspartyl-tRNA synthetase. A component analysis of the Poisson-Boltzmann free energies is employed to identify specific interactions that determine the binding affinities. The combined use of molecular dynamics free energy simulations to study one binding process thoroughly, followed by molecular dynamics and Poisson-Boltzmann free energy calculations to study a series of related ligands or mutations is proposed as a paradigm for protein or ligand design.

The binding of Asn in an alternate, ''head-to-tail'' orientation observed in the homologous asparagine synthetase is analyzed, and found to be more stable than the ''Asp-like'' orientation studied earlier. The new orientation is probably unsuitable for catalysis. A conserved active site lysine (Lys198 in *Escherichia coli*) that recognizes the Asp side-chain is changed to a leucine residue, found at the corresponding position in asparaginyl-tRNA synthetase. It is interesting that the binding of Asp is calculated to increase slightly (rather than to decrease), while that of Asn is calculated, as expected, to increase strongly, to the same level as Asp binding. Insight into the origin of these changes is provided by the component analyses. The double mutation (K198L,D233E) has a similar effect, while the triple mutation (K198L,Q199E,D233E) reduces Asp binding strongly. No binding measurements are available, but the three mutants are known to have no ability to adenylate Asn, despite the ''Asp-like'' binding affinities calculated here. In molecular dynamics simulations of all three mutants, the Asn ligand backbone shifts by 1-2 Å compared to the experimental Asp:AspRS complex, and significant side-chain rearrangements occur around the pocket. These could reduce the ATP binding constant and/or the adenylation reaction rate, explaining the lack of catalytic activity in these complexes. Finally, Asn binding to AspRS

Abbreviations used: aaRS, aminoacyl-tRNA synthetase; MDFE, molecular dynamics free energy; PBFE, Poisson-Boltzmann free energy; MD, molecular dynamics; AspRS, aspartyl-tRNA synthase.

E-mail addresses of the corresponding authors: archonti@ucy.ac.cy; simonson@igbmc.u-strasbg.fr; marci@tammy.harvard.edu

0022-2836/01/020307–21 $35.00/0

with neutral K198 or charged H449 is considered, and shown to be less favorable than with the charged K198 and neutral H449 used in the analysis.

Keywords: protein electrostatics; molecular recognition; aminoacyl-tRNA synthetases; protein engineering; dielectric continuum

*Corresponding authors

Introduction

Recognition between molecules, such as enzymes and substrates, ligands and receptors, or proteins and nucleic acids is an important element of the biochemistry and information flow in living systems (Janin, 1995; Westhof, 1997; Nagai & Mattaj, 1994). Specificity is needed to preserve the correctness of the biochemical pathways and the integrity of the information. It is often provided by non-covalent interactions between neighboring chemical groups, through hydrogen bonds, salt bridges, and tight packing of complementary molecular surfaces, although longer range electrostatic interactions also play a role, particularly in the formation of encounter complexes (Wade *et al*., 1994; Vijayakumar *et al*., 1998). An experimental approach to determine the importance of individual residues for the strength and specificity of binding is to delete them or introduce new ones by means of mutagenesis. This approach, now referred to as protein engineering (Fersht, 1999), has led to a better understanding of recognition phenomena, and to insights that can aid in the design of new ligands.

Protein engineering can also be performed with computer simulations. For several important classes of ligands, binding free energies have been calculated and/or the effect of point mutations studied. Molecular dynamics simulations, when properly used and interpreted, can provide accurate estimates of binding free energy changes associated with point mutations (Kollman, 1993). Also, models based on continuum electrostatics can often provide semi-quantitative binding free energies (Schaefer *et al*., 1998). An important advantage of computer simulations over experiments is the possibility of decomposing the overall free energies into contributions from microscopic interactions and determining their structural origins. Thus, in a protein:ligand complex, the binding free energy change associated with a point mutation can be decomposed into a sum of contributions from individual amino acid residues, the ligand, and the surrounding solvent (Gao *et al*., 1989, Boresch *et al*., 1994). While these decompositions are not unique, they provide a measure of the relative importance of each group for the process considered. Here, we show how a similar decomposition can be made with continuum electrostatics models and compare the two approaches. We then illustrate the method, by applying it to an important molecular recognition process, namely amino acid binding by an aminoacyl-tRNA synthetase.

Aminoacyl-tRNA synthetases (aaRSs) catalyze the first step in the translation of the genetic code, attaching a specific amino acid to a cognate tRNA molecule (Meinnel *et al*., 1995; Cusack, 1995; Arnez & Moras, 1997). A high specificity is needed to preserve the correct correspondence between the identity of the amino acid and the identity of the anticodon of the tRNA. The rate of erroneous amino acid incorporation into proteins is thought to be dominated by selection of incorrect amino acid residues by aaRSs (Arnez & Moras, 1997), while the rates for erroneous tRNA selection by aaRSs are much lower. For *Escherichia coli* aaRSs *in vitro*, for example, the ratio k_{cat}/K_M of the turnover number of the enzyme and the Michaelis constant of the tRNA is 10^5-10^8 times lower for aminoacylation of non-cognate tRNA than for cognate tRNA (Pallanck *et al*., 1995). The overall rate of erroneous amino acid incorporation is around 1 in 10^4. To understand the molecular basis of amino acid recognition is therefore of fundamental importance.

Several groups have engineered aaRSs and/or tRNAs with modified specificities. For tyrosyl-tRNA synthetase, extensive mutation experiments have been done to determine the contributions of various residues to the binding and catalysis (Fersht *et al*., 1985; Fersht, 1988), and increased Tyr specificity was obtained (de Prat Gay *et al*., 1993). Modified GlnRS's have been obtained by random mutagenesis that preferentially aminoacylate tRNA(Gln) with a Glu instead of a Gln residue (Agou *et al*., 1998). The identities of several tRNAs have also been modified successfully by mutagenesis (Normanly & Abelson, 1989). The ability to control and modify aaRS and tRNA specificity could be of considerable technological value. Thus, engineered pairs of aaRS and tRNA have been used to incorporate artificial amino acid residues into proteins *in vivo* (Ibba & Hennecke, 1995; Hohsaka *et al*., 1999; Liu & Schultz, 1999; Wang *et al*., 2000). Bacteria carry out related ''engineering'' when they develop drug resistance: the pathogen *Staphylococcal aureus* developed a second isoleucine-tRNA synthetase in response to a drug that targeted its original IleRS (Gilbart *et al*., 1993).

This article focuses on Asp and Asn binding to native and mutant AspRS's (from *E. coli*). Experimental attempts to engineer a modified specificity into aspartyl-tRNA synthetase (AspRS) have not been successful (Cavarelli *et al*., 1994). Asp binds to

AspRS with a milli- to micromolar dissociation constant (Eriani *et al*., 1990), corresponding to a standard binding free energy of around 6 kcal/mol. Although the active site sequences of AspRS and AsnRS are very similar (Archontis *et al*., 1998), native AspRS does not adenylate Asn detectably, and various single, double, or triple mutations in the binding pocket of AspRS that could be expected to improve Asn binding did not lead to any activity for Asn. From the experiments alone, it is difficult to rationalize these negative results, in part because the exact level of binding produced by a particular mutation has not been measured. Moreover, since three-dimensional structures of the Asn complexes have not been obtained, it is unclear whether Asn binding, if any, corresponds to a position and orientation suitable for reaction with ATP and/or tRNA, i.e. the assay method might reflect the absence of aminoacylation rather than lack of binding.

Computer modelling can be a powerful tool to aid in the interpretation of such experiments. Previously, we reported a detailed study of Asp and Asn binding to native AspRS, based on molecular dynamics free energy simulations (MDFE) (Archontis *et al*., 1998; Simonson *et al*., 1997). A useful feature of these calculations is the decomposition of the overall free energy change into group contributions (Gao *et al*., 1989; Boresch *et al*., 1994). The MDFE results thus provide a measure of the relative importance of different amino acids and different types of energy terms (e.g. electrostatics, van der Waals) in the differential binding. A combination of molecular dynamics (MD) simulations and Poisson-Boltzmann free energy calculations (PBFE) (Warwicker & Watson, 1982; Rogers, 1986) represents a less expensive alternative to MDFE, suitable for examining multiple active site mutations in an engineering effort (Srinivasan *et al*., 1998; Massova & Kollman, 1999). In this context, a component analysis of the Poisson-Boltzman free energy changes will also be valuable. It can serve to complement MDFE results if they are available, as in the present study. Component analyses of PBFEs have been performed by several others (Gilson & Honig, 1988; Davis & McCammon, 1991; Hendsch & Tidor, 1999). The present analysis is most similar to that of Hendsch & Tidor (1999), although it differs in the decomposition of some of the free energy terms.

The PBFE calculations for AspRS are first validated by their ability to reproduce the published MDFE results (Archontis *et al*., 1998) as well as some new MDFE results, reported here, for Asp/Asn binding to a mutant AspRS. The PBFE calculations are then used to extend the analysis of Asn binding to native or mutant AspRS. In each case, a molecular dynamics simulation is performed to provide structural models for the Poisson-Boltzmann calculations. Several processes are considered. First, we use PBFE to calculate the binding free energy of Asn to AspRS in a reverse, ''head-to-tail'' orientation. This orientation, observed in the homologous asparagine synthetase, places the Asn backbone carboxylate roughly in the position normally occupied by the Asp side-chain carboxylate. The previous MDFE calculations considered only the ''Asp-like'' orientation, where the Asn side-chain is in roughly the position of the Asp side chain. In principle, only the latter orientation could lead to erroneous misacylation of a tRNA(Asp) with Asn; i.e. if the head-to-tail orientation occurs, it would be inactive. Second, we calculate the binding free energies of Asp and Asn to three mutant AspRS's: K198L (*E. coli* numbering), the double mutant (K198L,D233E), and the triple mutant (K198L,Q199E,D233E). All three mutants have been examined experimentally (Cavarelli *et al*., 1994), and showed no ability to adenylate either Asp or Asn. Finally, we consider the effect on Asn binding of deprotonating the active site Lysl98 or protonating the active site His449 (Supplementary Material).

The structure of the native Asp:AspRS complex was determined recently (Schmitt *et al*., 1998), and structures of two aspartyl-adenylate:AspRS complexes are known (Schmitt *et al*., 1998; Poterszman *et al*., 1994). However, the structures of native AspRS complexed with an Asn residue, and of mutant AspRSs complexed with Asp or Asn are unknown. In our earlier work, the Asp:AspRS complex was modelled by building Asp into the active site of AspRS from *E. coli* (Archontis *et al*., 1998); the latter structure was determined by crystallography with an aspartyl-adenylate analogue in the active site. MD simulations then produced models of the Asp:AspRS complex in good agreement with experiment (average backbone rms deviation of 0.5 Å for the protein over 500 ps of simulation). The quality of the simulation model was partly due to a new treatment of electrostatic interactions, specifically adapted to stochastic boundary conditions (Simonson *et al*., 1997). For the Asn:AspRS complex, extensive MD and MDFE simulations were performed to identify the most stable Asn positions in the active site (Archontis *et al*., 1998; Simonson *et al*., 1997). Here, we use MD simulations with the same protocol to generate models of Asn bound to AspRS in the head-to-tail orientation, and of Asp or Asn bound to the mutant AspRS's in the ''normal'' orientation. These models are used as input for the Poisson-Boltzmann calculations. They also provide information about the flexibility of each structure and the detailed interactions stabilizing it.

This paper is organized as follows. The following section decribes theoretical development. The PBFE method is described in brief, and the present implementation is compared to recent related work (Massova & Kollman, 1999). The component analysis of Poisson-Boltzmann free energy differences is presented. The third section presents results. We first compare PBFE binding free energy differences to MDFE results. Good agreement is found (within ~1 kcal/mol) when a protein dielectric constant of four is used for PBFE. This is in

accord with earlier work (Sharp & Honig, 1991). The agreement suggests that it is useful to apply the PBFE approach to problems for which detailed experimental or MDFE data are not available, such as the binding of Asn in the head-to-tail orientation, and the binding of Asp and Asn to three AspRS mutants. In each case, the MD structures are discussed along with the PBFE free energy results. The fourth section is a Discussion. We consider the relation between the free energy component analysis and point mutagenesis experiments. We point out that the contribution of a given residue to the binding free energy is not simply related to the binding free energy change when that residue is mutated (as already discussed for MDFE calculations (Archontis *et al.*, 1998; Boresch & Karplus, 1995; Lau & Karplus, 1994)). In some cases, the mutation can induce structural changes whereby a different residue becomes more strongly involved in the binding (Fersht, 1988; Horovitz, 1996; Otzen & Fersht, 1999). Nevertheless, the component analysis provides a clear measure of the importance of such rearrangements and of the role of specific structural features in binding. The effect of His449 protonation and Lys198 deprotonation on binding is considered in the Supplementary Material.

Theoretical Development

Binding free energies from continuum electrostatics

A subset of the binding reactions considered in this work are shown in the thermodynamic cycle in Figure 1 . Each binding reaction is treated separately; i.e. a separate calculation is performed for the binding of each ligand (Asp, Asn) to each protein variant. This is in contrast to ''alchemical'' MDFE calculations, which typically model the vertical legs of the thermodynamic cycle. Thus, to compare Asp and Asn binding using MDFE, the ligand is changed gradually from Asp into Asn both in the protein and in solution.

While MDFE calculations explicitly include both electrostatic and non-electrostatic contributions, the PBFE approach (Gilson & Honig, 1988; Hendsch & Tidor, 1994) as used here includes only electrostatic contributions to the binding free energy. The protein and ligand are viewed as low dielectric bodies with embedded point charges, surrounded by high dielectric solvent. The free energies of the separated and bound states are calculated from the distribution of electrostatic potential, obtained by solving the Poisson or Poisson-Boltzmann equation numerically (see below).

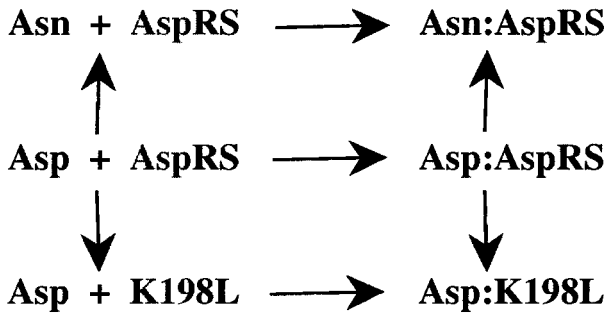

Figure 1. Thermodynamic cycles describing the binding of Asp and Asn to native AspRS and to the K198L mutant AspRS.

Related studies have combined continuum electrostatics with various non-electrostatic contributions (Srinivasan *et al.*, 1998; Froloff *et al.*, 1997; Schapira *et al.*, 1999). In particular, in a recent approach, contributions from the hydrophobic effect were estimated from a solvent accessible surface model and contributions from loss of side-chain entropy were estimated from an MD simulation by using a harmonic approximation (Vorobjev *et al.*, 1998; Srinivasan *et al.*, 1998; Massova & Kollman, 1999). For the system studied here, MDFE calculations showed (Archontis *et al.*, 1998) that the binding free energy differences between Asp and Asn are dominated by electrostatic contributions. Therefore other terms are not considered.

Component analysis of binding free energies

To obtain a more complete understanding of the interactions that determine the PBFE binding free energies, a component analysis can be performed. Compared to MDFE component analyses (Gao *et al.*, 1989; Boresch *et al.*, 1994), there are important differences. Usually, the processes considered are different: i.e. the alchemical transformation of the ligand or protein with MDFE, and the ''chemical'' binding reactions with PBFE (see Figure 1). Also, the treatment of solvent is explicit with MDFE and implicit with PBFE. The consequences of these differences are considered when the specific application is discussed. The following analysis is similar to that of (Hendsch & Tidor, 1999). It is presented for completeness, and because there are differences in the decomposition of some of the free energy terms. We consider only the usual, linear response limit of the continuum model, which is expected to be a good approximation for the system considered (Sharp & Honig, 1990). If non-linear effects were significant, as in the full (non-linear) Poisson-Boltzmann equation, the following decomposition would not apply.

We start from an expression for G_{pl}, the free energy of the protein:ligand complex in solution. It can be written:

$$G_{\mathrm{pl}} = \frac{1}{2}\sum_{\mathrm{i}} q_{\mathrm{i}} V_{\mathrm{i}}^{\mathrm{pl}} = \frac{1}{2}\sum_{\mathrm{i}\in\mathrm{lig}} q_{\mathrm{i}} V_{\mathrm{i}}^{\mathrm{pl}} + \frac{1}{2}\sum_{\mathrm{i}\in\mathrm{prot}} q_{\mathrm{i}} V_{\mathrm{i}}^{\mathrm{pl}} \qquad (1)$$

where the first sum is over all protein and ligand atoms; q_{i} is the partial charge of atom i; $V_{\mathrm{i}}^{\mathrm{pl}}$ is the total electrostatic potential on atom i in the complex, and the sums on the right are over ligand and protein atoms, respectively. From the linearity

of continuum electrostatics, the potential on atom i can be expressed as a sum over all ligand and protein atoms:

$$V_i^{pl} = \sum_j V_{j\to i}^{pl} = \sum_{j\in lig} V_{j\to i}^{pl} + \sum_{j\in prot} V_{j\to i}^{pl} \qquad (2)$$

where $V_{j\to i}^{pl}$ is the potential at atom i when only the partial charge q_j is present in the protein. Using the reciprocity relation $q_i V_{j\to i}^{pl} = q_j V_{i\to j}^{pl}$ (Landau & Lifschitz, 1980), we have:

$$G_{pl} = \frac{1}{2}\sum_{i\in lig, j\in lig} q_i V_{j\to i}^{pl} + \sum_{i\in prot, j\in lig} q_i V_{j\to i}^{pl} + \frac{1}{2}\sum_{i\in prot, j\in prot} q_i V_{j\to i}^{pl} \qquad (3)$$

The first and third sums on the right each include terms of the form $1/2 q_i V_{i\to i}$, representing the Born ''self-energy'' of each charge q_i (Schaefer & Froemmel, 1990).

To obtain the binding free energy ΔG_{bind}, we subtract the analogous expressions for the separated protein and ligand:

$$\Delta G_{bind} = G_{pl} - G_p - G_l = \sum_{i\in prot, j\in lig} q_i V_{j\to i}^{pl} + \frac{1}{2}\sum_{i\in lig, j\in lig} q_i [V_{j\to i}^{pl} - V_{j\to i}^{l}] + \frac{1}{2}\sum_{i\in prot, j\in prot} q_i [V_{j\to i}^{pl} - V_{j\to i}^{p}] \qquad (4)$$

The first term on the right represents direct interactions between the ligand and the protein residues in the complex, screened by solvent. It will be referred to as the ''direct interaction term'' The second term (''ligand desolvation term'') includes the change in intra-ligand interactions upon binding, due to changes in the ligand geometry or charge distribution, as well as changes in the interaction of the ligand with polarization charge in the surrounding dielectric media. If the ligand is assumed to have the same geometry and partial charges in the bound and free state (as below), then the intra-ligand interactions do not contribute, and this term arises entirely from ligand interactions with the polarization charge. It corresponds therefore to decreased ligand-solvent interactions. The third term on the right of equation (4) has an identical interpretation. If the protein is assumed to have the same structure in the bound and free state (as below), then this term represents the effect of displacing solvent from the binding site by inserting the ligand, decreasing the protein-solvent interactions. It will be referred to as the ''protein desolvation term''.

Solvent contributions are present implicitly in all three terms; this is in contrast to MDFE, where solvent is treated explicitly and is associated with its own free energy component. Furthermore, each protein or ligand group contributes to ΔG_{bind} through all three terms in equation (4). Thus, the protein contributes to the ligand desolvation term: even though the protein charges do not appear explicitly in this term, each protein residue occupies space around the ligand and contributes to its desolvation (Schaefer & Froemmel, 1990). In a similar way, the ligand contributes to the protein desolvation term, even though its charges do not appear explicitly.

If one compares the binding of two ligands that occupy exactly the same space in the active site, the protein desolvation terms will be identical for the two ligands. In the present application, the two ligands occupy very similar positions, and the protein desolvation terms are approximately equal in all cases. In a similar way, the ligand desolvation for Asp (or Asn) is approximately equal in all the protein variants. Therefore, we focus on the analysis of the direct interaction term, although we also present the desolvation contributions. The contribution ΔG_R^{DI} of residue R to the direct interaction term has the form:

$$\Delta G_R^{DI} = \sum_{i\in R, j\in lig} q_i V_{j\to i}^{pl} \qquad (5)$$

The quantity on the right can be obtained from a calculation of the electrostatic potential arising from the partial charges of the ligand at the positions of the partial charges q_i of the protein:ligand complex. Subtracting the results for the protein:Asp and protein:Asn complexes, we obtain the contribution of residue R to the direct interaction term in the binding free energy difference.

In systems where the desolvation terms are larger, it would be of interest to decompose these into group contributions as well. Although a decomposition as simple as the one based on equation (5) is not possible, a related method exists for the protein desolvation term. This term takes the form (equation 4) of a sum over all pairs i, j of protein atoms. The contribution of a given pair of residues R, R' can be defined by summing over those atom pairs where i belongs to R and j to R'. A possible approach is then to partition this pair contribution equally between the two residues R and R'. With this approach, the contribution ΔG_R^{PD} of residue R to the protein desolvation term has the form:

$$\Delta G_R^{PD} = \frac{1}{2}\sum_{i,j\in R} q_i [V_{j\to i}^{pl} - V_{j\to i}^{p}] + \frac{1}{2}\sum_{R'\neq R}\frac{1}{2}\sum_{i\in R, j\in R'} q_i [V_{j\to i}^{pl} - V_{j\to i}^{p}] \qquad (6)$$

The first term on the right is a ''self'' contribution of residue R. An example of this decomposition is reported below. A different decomposition was used by Hendsch & Tidor (1999), where the entire

R, R' coupling was included in both the R and the R' components, so that the residue components do not add up to the total effect due to ''double counting'', as they point out. The present partitioning is analogous to that which is implicit in the MDFE free energy decomposition (Boresch *et al.*, 1994).

An analysis of the contribution of protein groups to the ligand desolvation term in equation (4) has to be made somewhat differently. The protein charges do not appear explicitly in this term; rather the protein atoms contribute by occupying space around the ligand in the protein:ligand complex, replacing high-dielectric solvent by the lower-dielectric protein medium. Point mutations in the protein that do not significantly change the space occupied by protein will not contribute significantly to the ligand desolvation term. For this reason, no decomposition of the ligand desolvation term is attempted in the application below. On the other hand, if deletion mutations of the protein are of interest, especially large-scale changes where entire protein domains are removed or inserted, then the ligand desolvation term will vary substantially. In that case, a group decomposition of a different nature from the one used for the protein desolvation term is possible. The ligand desolvation term can be calculated including only selected protein groups. As more and more protein groups are included in the calculation, the change in the ligand desolvation term at each step can be attributed to the groups added at that step. Although this decomposition will depend on the order in which the protein groups are added, a ''natural'' order can be used, such as one based on increasing distance from the ligand.

Results

Validation of PBFE for AspRS ligand binding

To test the continuum model for the present system, double free energy differences are compared to MDFE results for two processes: binding of Asp and Asn to the native protein, and binding to the K198L mutant protein (see Figure 1).

Comparing PBFE to MDFE: binding to the native protein

To obtain the ligand binding free energies with the continuum approach, we calculate the free energies of the protein alone, the ligand alone, and the protein:ligand complex (see Methods). Subtracting the first two from the third gives the electrostatic contribution to the binding free energy. Results are given in Table 1 with protein/ligand dielectric constants ranging from one to four.

The binding free energy at zero ionic strength is seen to vary roughly as $1/\varepsilon_p$. A protein dielectric constant of 4, commonly used in PBFE calculations, gives good agreement with MDFE (Table 1). The resulting $\Delta\Delta G$ is 16.4 kcal/mol favoring Asp binding, compared to 15.3 kcal/mol with MDFE. The difference is within the uncertainty of both the MDFE calculations (±2.8 kcal/mol) and the PBFE calculation ($\sim \pm 3$ kcal/mol; see below). With an ionic strength corresponding to 0.1 M of monovalent counterions, the Asp and Asn binding are both slightly increased, but the binding free energy difference is almost unchanged.

Although non-electrostatic effects contribute to the binding free energy difference, a large cancellation is expected between the two ligands. In the 500 ps MD simulation of the Asn:AspRS complex, the Asn ligand samples several positions, already described in detail (Archontis *et al.*, 1998). The overall rms shift is 1.5-2 Å compared to the Asp position; the same conserved arginine residues (Arg489 and Arg217) anchor its backbone and side-chain at either end of the binding pocket, though less strongly than in the Asp case. Therefore, the structure is similar to the Asp:AspRS complex; the extent of solvent displacement from the binding interface and the amount of side-chain ordering are similar in the two systems.

Comparing PBFE to MDFE: binding to the K198L mutant

Lys198 is a key Asp recognition residue in AspRS. It hydrogen-bonds to the Asp side-chain and contributes tens of kcal/mol to the binding free energy difference between Asp and Asn; see (Archontis *et al.*, 1998) and Table 2. In asparaginyl-tRNA synthetase, this residue is a Leu or a Gly, and the Lys198Leu mutation of AspRS has been investigated experimentally in attempts to engineer Asn binding to AspRS (Cavarelli *et al.*, 1994; G. Eriani, personal communication). MDFE calculations were carried out here for the K198L mutation with bound Asp or Asn. The mutant protein was modelled by simply removing the net positive charge on the Lys198 side-chain and replacing it by a leucine-like distribution, shown in Figure 2 (to reflect this, the mutation is written with italics, K198*L*, in what follows). From MDFE, the double free energy difference $\Delta\Delta G_{bind} = \Delta G_{bind}(Asn) - \Delta G_{bind}(Asp)$ for binding to AspRS(K198*L*) is found to be 0.6 kcal/mol, as compared with a $\Delta\Delta G_{bind}$ of 15.3 kcal/mol for the native protein. The binding free energies of Asp and Asn to the K198*L* mutant were also estimated with the PBFE method. With $\varepsilon_p = 4$, the binding free energies are −11.4 kcal/mol and −11.2 kcal/mol for Asp and Asn, respectively, giving a $\Delta\Delta G_{bind}$ for binding of 0.2 kcal/mol, in very good agreement with MDFE (Table 1). The PBFE calculations lead to the somewhat surprising result, that the change in $\Delta\Delta G_{bind}$ relative to the wild-type includes a slight increase in Asp binding, along with the expected large increase in Asn binding. We discuss the Asn binding first.

The structures sampled in the Asn:AspRS (K198*L*) molecular dynamics simulation deviate from the initial, native-like structure, reaching an

Table 1. PBFE binding free energies of Asp and Asn to AspRS (kcal/mol)

Protein	ε_p	ΔG_{bind} (Asp)	ΔG_{bind} (Asn)	$\Delta\Delta G_{bind}$ (Asp → Asn)
Native	1	−38.1 (10.7)	-	-
Native	2	−18.8 (5.2)	15.5 (4.2)	34.3
Native	4	−9.3 (2.5)	7.1 (1.9)	16.4
Native	4[b]	−10.7 (2.6)	6.0 (2.1)	16.7
Native	MDFE			15.3 ± 2.8[a]
Head-to-tail Asn	4	-	−5.5 (2.4)	-
K198*L*	4	−11.4 (2.9)	−11.2 (2.4)	0.2
K198*L*	MDFE			0.6[c]
K198L,D233E	4	-	−7.5	-
K198L,Q199E,D233E	4	2.1	−10.2	−12.3
H449H$^+$ [d]	4	−14.8 (2.7)	7.5 (2.7)	22.3
H449H$^+$ H[e]	4	−14.8 (2.7)	7.1 (1.9)	21.9

Results averaged over appropriate MD simulations, using $N \sim$ 10-50 conformations spanning 100-200 ps in each case. Zero ionic strength, except where mentioned. Numbers in parentheses are standard deviations. If the conformations used for averaging are assumed to be statistically independent, the corresponding standard errors would be $\sqrt{N-1}\sim$3-7 times smaller.

[a] From (Archontis *et al.*, 1998).

[b] Ionic strength corresponding to 0.1 M monovalent counterions.

[c] This work.

[d] Binding to native AspRS with the active site His449 in its doubly protonated state.

[e] His449 in doubly protonated state in the Asp:AspRS complex, but in neutral state in the Asn:AspRS complex (see Supplementary Material).

rms deviation for backbone and C$^\beta$ atoms of 0.6 Å after a few picoseconds, then increasing gradually to 0.9 Å after 450 ps (data not shown). Typical snapshots are shown in Figure 3. Structural deviations in the active site are significant. The Asn ligand shifts rapidly during the first few picoseconds, then remains in a stable position with a 1.8 Å rmsd from that of the Asp ligand in the native Asp:AspRS complex. It is anchored by a side-chain hydrogen bond to Arg489, a salt bridge between its ammonium group and Asp233, and a salt bridge between its backbone carboxylate and Arg217. Arg489 and Asp233 are themselves linked through a water molecule nested in the back of the binding pocket (labelled W1 in Figure 3). The ligand ammonium also hydrogen bonds to Ser193,

Table 2. Residue contributions to binding free energies for native AspRS

	PBFE			MDFE	
Residue	Asp binding	Asn binding	Total $\Delta\Delta G_{bind}$	Total $\Delta\Delta G_{bind}$	vdW
Lys198	−13.2 (−19.5)	4.4 (4.2)	17.6 (23.7)	59.8	
Gln199	0.1 (0.1)	−0.1 (−0.1)	0.0 (−0.2) [−0.3][a]		
Asp233	5.3 (11.3)	−3.0 (−2.9)	−8.3 (−14.2) [−9.1][a]	−52.1	
Arg489	−20.0 (−26.3)	1.9 (1.8)	21.9 (28.1)	92.8	−1.9
Glu235	7.8 (13.3)	−2.4 (−2.4)	−10.2 (−15.8)	−54.1	
Arg217	−10.1 (−22.5)	−5.9 (−10.5)	4.3 (12.1)	34.6	
Asp475	2.7 (9.9)	−1.2 (−2.1)	−4.0 (−12.0)	−41.1	
Gln231	−0.4 (−0.8)	−0.5 (−0.5)	0.1 (0.3)		
His449	−1.6 (−1.9)	−0.7 (−1.0)	0.9 (0.9)	6.4	−0.4
Glu482	2.8 (10.5)	−0.3 (−0.8)	−3.2 (−11.3)	−28.3	
Arg537	−2.4 (−9.1)	0.2 (0.2)	2.7 (9.2)	23.8	
Asp175	1.1 (6.2)	0.4 (1.3)	−0.7 (−4.9)	−19.1	
Gln192	0.1 (0.6)	−0.1 (0.1)	−0.3 (−0.5)		
Gln195	−0.9 (−0.6)	−2.1 (−2.0)	−1.2 (−1.4)	−4.3	0.1
Glu219	1.0 (5.9)	0.2 (1.2)	−0.8 (−4.7)	−18.7	
Arg225	−1.6 (−7.4)	−0.1 (−0.6)	1.6 (6.8)	20.6	
Glu228	0.7 (8.1)	0.0 (0.4)	−0.6 (−7.7)	−15.2	
Arg245	−1.2 (−6.3)	0.3 (0.7)	1.4 (7.0)		
Asp282	1.2 (5.8)	0.0 (−0.2)	−1.3 (−6.1)		
His448	−1.2 (−1.5)	0.1 (0.7)	1.4 (2.2)		
Asp536	1.0 (5.5)	−0.1 (0.0)	−1.1 (−5.5)	−19.3	
Protein desolvation	5.5	4.8	−0.7	Solvent component	
Ligand desolvation	16.6	11.5	−5.1	−34.2	2.1

Free energy components (kcal/mol). The residue components in the upper part of the Table correspond to the direct interaction term in the binding free energy (ΔG_R^{DI} in equation 5). Values in parentheses are calculated with a dielectric of four throughout space (infinite protein). MDFE values are the sum of an electrostatic and a (small) van der Waals contribution. The van der Waals term is listed separately only when it is larger than 0.1 kcal/mol in magnitude.

[a] Side-chain contribution in square brackets; reported only if it differs from the total residue contribution by more than (±0.2) kcal/mol.

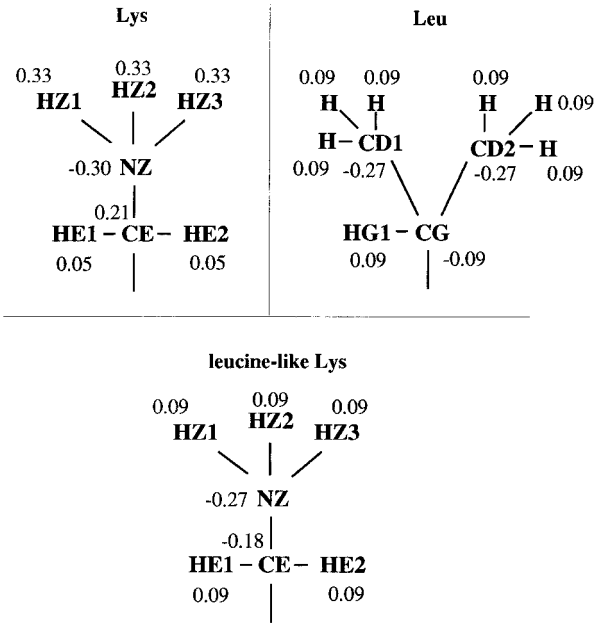

Figure 2. Atomic charges for lysine (top left), leucine (top right), and ''leucine-like'' lysine (bottom). The latter charges are used to model Leu198 in the K198*L* single mutant calculations.

as in the native Asp:AspRS complex, and Gln231. In the simulations of the native Asp:AspRS and Asn:AspRS complexes (Archontis *et al.*, 1998), this last interaction was replaced by a Gln195-ammonium hydrogen bond. The positional shift of the ligand and the surrounding side-chains, compared to the Asp:AspRS complex, may reduce the binding affinity of ATP and/or the reaction rate for the adenylation reaction. This could explain the lack of adenylation observed experimentally with this mutant (Cavarelli *et al.*, 1994), even if Asn were bound, as the simulations suggest. In particular, while the conserved Arg217 (found in all class II aaRSs) maintains a strong interaction with the ligand backbone carboxylate as in the native Asp:AspRS complex, it may not be optimally placed with respect to an ATP substrate or an adenylate product molecule. When an adenylate molecule was model-built into the active site by superimposing active site residues from an adenylate:AspRS crystal structure (Poterszman *et al.*, 1994), a bad steric contact was found between its α-phosphate and the Arg217 side-chain. To resolve this bad contact, it may be necessary to displace active-site groups.

In the Asp:AspRS(K198*L*) complex, the displacement of the ligand is even larger: 1.79 Å on average from the native position (Table 3). This displacement may explain the lack of activity of the mutant enzyme for Asp. A snapshot of the

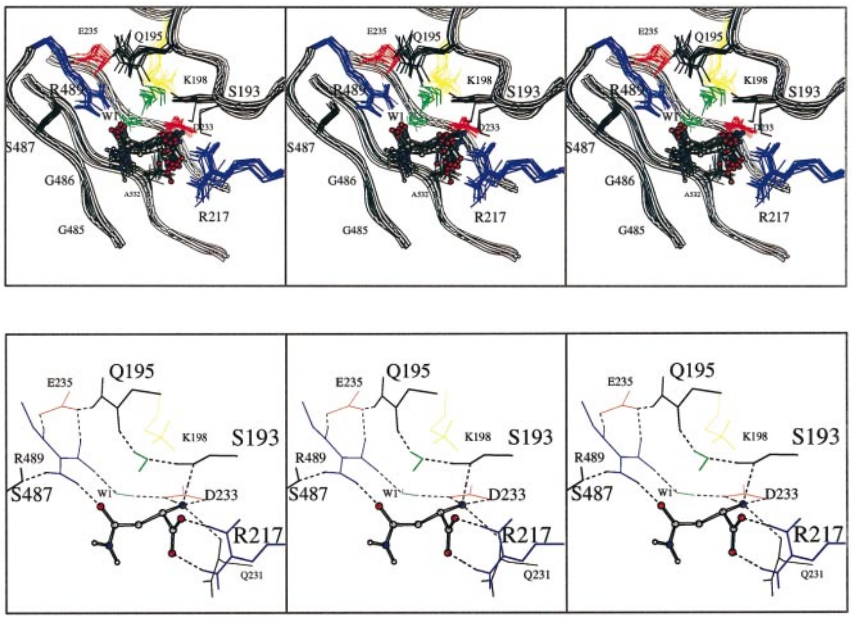

Figure 3. (a) Eleven snapshots of the active site region of the Asn:AspRS(K198*L*) complex at 20 ps intervals, covering the last 200 ps of the MD simulation. The left and middle panels provide wall-eyed stereo; the middle and right panels provide cross-eyed stereo. Important residues are labelled. The protein backbone is shown in tube representation. Positively charged residues are in blue, negatively charged are in red. The ''leucine-like'' residue 198 (see the text) is in yellow. The ligand is shown in ball-and-stick form. Aliphatic hydrogens and the ligand ammonium hydrogens are omitted. Two stable water molecules are shown (green), including one (W1) that bridges Arg489 and Asp233. The structure is remarkably stable, though a range of orientations are sampled by the ligand backbone carboxylate. This and the following Figures were prepared with the program MOLSCRIPT (Kraulis, 1991). (b) Close-up of the last snapshot from (a). Hydrogen bonds are shown as broken lines.

Table 3. Rms deviations from initial structure for MD simulations

Protein	Ligand	Simulation length (ps)	Overall rmsd[a]	Ligand rmsd[b]
Native	Asp	630	0.82	0.97 (1.02)
Native	Asn	500	1.55	1.65 (2.41)
Native(H449+)	Asp	400	0.80	1.58 (1.78)
Native(H449+)	Asn	400	0.76	1.94 (1.87)
Native	Asn[c]	500	1.55	1.65 (2.41)
K198*L*	Asp	600	1.02	1.11(1.33)
K198*L*	Asn	520	0.95	1.79 (1.20)
K198L,D233E	Asn	510	0.97	1.20 (1.21)
K198L,Q199E,D233E	Asp	530	1.32	1.48 (2.07)
K198L,Q199E,D233E	Asn	500	1.34	1.17 (0.87)

[a] Rms deviation (Å) for backbone and C^{β} atoms, averaged over the last 50 ps of the trajectory, excluding the restrained outer region (''buffer zone'', Simonson *et al*., 1997).
[b] Deviation of ligand (Å) from the position of Asp in the native Asp:AspRS structure (excluding hydrogens, backbone carboxylate oxygens, and the terminal side-chain atoms O^{δ} or N^{δ}). Values in parentheses represent the deviation of the backbone carboxylate carbon (where the adenylation is to occur).
[c] Head-to-tail orientation; all other Asn simulations have the original orientation (Archontis *et al*., 1998).

Asp:AspRS(K198*L*) complex from the trajectory is shown in Figure 4(a). A snapshot from the native Asp:AspRS trajectory is shown in Figure 4(b). The interactions in the two structures are similar but not identical. In the mutant complex, there is an intramolecular hydrogen bond between the ligand's ammonium group and its side-chain carboxylate, and the ligand backbone carboxylate interacts with three water molecules. This contrasts with the native trajectory, where for example there are only one to two water molecules interacting with the backbone carboxylate.

The PBFE calculations so far do not include any specific water molecules; all solvent is treated implicitly. The Asn:AspRS(K198*L*) complex has an ordered water molecule in the back of the binding pocket, bridging Arg489 and Asp233 (W1 in Figure 3). Since this water is highly ordered in the simulation, it may be appropriate to treat it as part of the low-dielectric ''protein'' medium. Table 4 gives results for this case. More generally, the presence of ordered waters (e.g. those observed in an X-ray structure) introduces a difficulty in defining precisely the boundary between solute and bulk solvent. To test the sensitivity of the calculations to the choice of boundary, Table 4 also gives results corresponding to different radii of the probe sphere used to construct the solute-solvent boundary. When the bridging water W1 is treated explicitly, ΔG_{bind}(Asn) changes by only ±0.2-0.4 kcal/mol (for probe radii between 1 and 2 Å). When no explicit solvent is included but the probe radius is

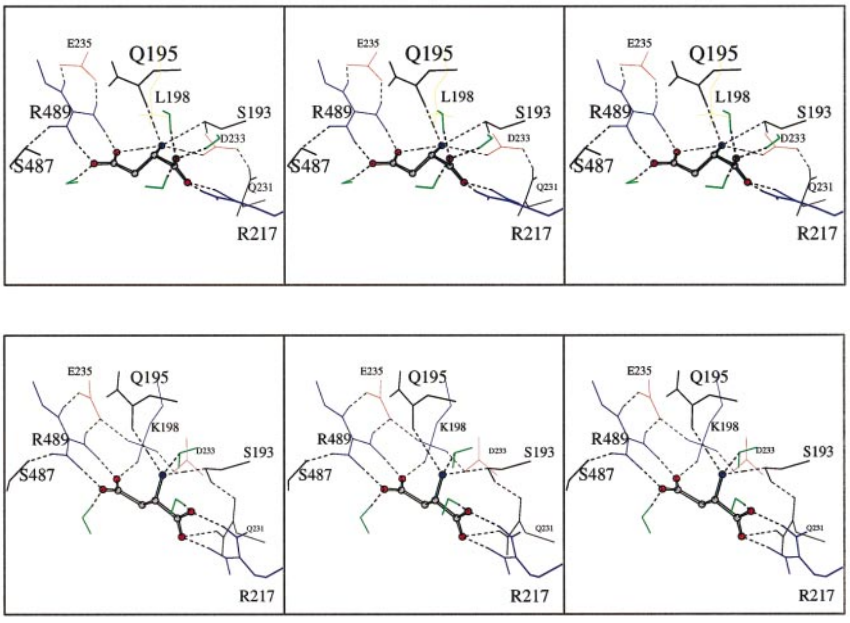

Figure 4. (a) A typical snapshot of the Asp:AspRS(K198*L*) complex. The viewpoint and mode of representation are the same as in Figure 3(b). Positively charged residues are in blue, negatively charged are in red. Selected waters are shown in green. Hydrogen bonds are shown as broken lines. (b) Analogous view of the Asp:AspRS native complex.

Table 4. Effect of explicit water and probe radius

	ΔG_{bind} (Asn)					
Probe radius (Å)	0	0.5	1	1.5	2	2.5
Implicit W1	−8.5	−8.5	−9.1	−10.7	−11.1	−14.2
Explicit W1	−9.2	−8.6	−9.5	−10.9	−10.9	−13.2

kcal/mol. Results for Asn binding to AspRS(K198*L*), averaged over five conformations, 20 ps apart. The water molecule W1 is shown in Figure 3.

varied from 0 to 2.5 Å, the binding free energy varies from −8.5 kcal/mol (0-0.5 Å probe) to −14.2 kcal/mol (2.5 Å probe). For intermediate probe radii of 1-2 Å, ΔG_{bind}(Asn) varies from −9.1 to −11.1 kcal/mol. Overall, it appears that an uncertainty of ±1-2 kcal/mol is associated with probe radius changes. The conformational averaging performed in this work may account in part for this moderate uncertainty level, since the solute-solvent boundary fluctuates slightly over time. The magnitude of the boundary fluctuation is approximately that of the atomic displacements, namely 0.5-1 Å over a 500 ps period.

PBFE free energy component analysis

As described in Theoretical Development, a component analysis can be performed for PBFE that is analogous to that used in MDFE. This analysis was applied to the native and K198*L* proteins. Because there are significant differences between the PBFE and MDFE component analyses, they are compared in Discussion and the complementary information they provide is described.

The PBFE free energy components (see equations (4)-(6)) of selected residues are reported in Tables 2 and 5 for the native protein. We discuss first the contributions to the direct interaction term (ΔG_R^{DI}, equation (5)); the smaller desolvation terms are discussed next. We focus on the differences $\Delta\Delta G_R^{DI}$ between the Asp and Asn components (e.g. fourth column in Table 2). The contributions of the entire residues, backbone and side-chain, are reported; in all cases, the side-chain contribution is predominant, and within 0.8 kcal/mol of the total residue contribution (within 0.2 kcal/mol for all but two residues). Only six residues make contributions $\Delta\Delta G_R^{DI}$ larger than 4 kcal/mol. The largest contributions come from Arg489 (21.9 kcal/mol) and Lys198 (17.6 kcal/mol), which have strong favorable interactions with the Asp ligand. The neighboring residues Asp233 and Glu235 make significant contributions (−8.3 and −10.2 kcal/mol) that partly cancel the previous two, due to their unfavorable interactions with the Asp ligand. Arg217 interacts strongly with the ligand backbone carboxylate of both Asp and Asn, but more strongly with Asp. Asp475 has a smaller, but still significant interaction disfavoring the Asp ligand. Additional contributions arise from the somewhat more distant Glu482 and Arg537. It is important to note that the ''explicit'' contribution ΔG_R^{DI} of a particular residue contains a large implicit contribution from the surrounding solvent and protein dielectric media. Indeed, ΔG_R^{DI} (equation (5)) involves the potentials $V_{j\to i}^{pl}$ produced by ligand charges on residue *R*; these potentials are strongly screened by the surrounding, polarizable, dielectric media. Most of the screening comes from the high-dielectric solvent; some comes from the protein medium. To isolate the effect of the solvent screening, the free energy components ΔG_R^{DI} are also calculated with a uniform dielectric constant of four throughout space; the results correspond to a system without solvent (infinite protein). For almost all the residues listed in Table 2, the values

Table 5. Residue pair contributions to protein desolvation free energy for AspRS:Asp

	Arg489	Glu235	Lys198	Asp233	Arg217	Other	Total
Arg489	**3.3**	−0.6	0.9	−0.5	−0.1		
Glu235	−0.6	**0.6**	−0.6	0.4	0.0		
Lys198	0.9	−0.6	**2.5**	−0.8	−0.1		
Asp233	−0.5	0.4	−0.8	**1.6**	0.0		
Arg217	−0.1	0.0	−0.1	0.0	**2.0**		
Other[a]	0.3	−0.1	0.1	−0.1	0.0		
Total[b]	3.4	−0.3	2.1	0.6	1.8	−1.9[c]	5.5[d]

Pairwise contributions to the protein desolvation part of the binding free energy (kcal/mol) for selected residues in the AspRS:Asp complex. Self-contributions of each residue are in bold type. The symmetry of the Table with respect to its diagonal reflects the reciprocity relation $q_i V_{j\to i} = q_j V_{i\to j}$ (see Methods).

[a] Coupling between the individual residue and all residues not listed in the Table.

[b] Total contribution of the individual residue, including self-contribution (ΔG_R^{pD} in equation 6).

[c] Contribution from all residues not listed in the Table.

[d] Total protein desolvation term for the AspRS:Asp complex.

without solvent have the same sign but are much larger in magnitude; they are one to two times larger for highly buried residues (Lys198, Arg489), and four to five times larger for more exposed ones (Asp475, Glu482).

Dielectric screening is also reflected in the desolvation terms (see equation (4)), also reported in Table 2. The ligand desolvation term is rather large: +16.6 kcal/mol for Asp binding and +11.5 kcal/mol for Asn binding, contributing −5.l kcal/mol to $\Delta\Delta G_{\text{bind}}(\text{Asp} \rightarrow \text{Asn})$. The negative sign corresponds to the greater cost to partially desolvate Asp, compared to Asn, due to the net charge on the former. For comparison, the solvation free energies of Asp and Asn (vapor to water transfer) are about −107 and −65 kcal/mol from PBFE calculations (data not shown).

The protein desolvation term measures the effect of replacing solvent in the active site by the lower-dielectric ligand. The resulting desolvation modifies the mutual interactions of the active site residues, reducing their screening by solvent. This term is positive and very similar in the Asp and Asn cases, contributing only −0.7 kcal/mol to the binding free energy difference $\Delta\Delta G_{\text{bind}}(\text{Asp} \rightarrow \text{Asn})$. As described in Theoretical Development, this term can be decomposed into components, based either on individual residues or on pairs of residues (equation (6)). Such a decomposition was made for the AspRS:Asp complex. The contributions of the most important residue pairs are reported in Table 5. The total residue contributions $\Delta\Delta G_R^{PD}$ are also reported. Each residue contributes a positive ''self'' term, which reflects the decreased screening of the residue's partial charges by solvent when the ligand binds. These vary from 3.3 kcal/mol for Arg489, whose solvent-accessibility decreases significantly upon ligand binding, to 0.6 kcal/mol for Glu235, whose solvent-accessibility is hardly affected by ligand binding. These changes can be compared to the total self-energies of the charged side-chains around the binding pocket; e.g. that of Lys198 is 53 kcal/mol. Pairs of residues with favorable interactions (e.g. Lys198-Asp233) contribute small negative terms favoring binding, since their interaction is reinforced upon ligand binding. Pairs with unfavorable interactions (e.g. Lys198-Arg489) contribute small positive terms. The total contributions of the residues listed range from −0.3 kcal/mol for Glu235 to 3.4 kcal/mol for Arg489, out of a total protein desolvation term of 5.5 kcal/mol. Overall, the protein desolvation components are much smaller than the direct interaction components, justifying our focus on the latter. Nevertheless, the ΔG_R^{PD} contributions from Table 5 should be added to the ΔG_R^{DI} term in Table 2 to get a better approximation to the total residue contribution to the binding free energy; e.g. for Arg489, which makes the largest residue contribution in the AspRS:Asp complex, the value is reduced to −16.6 kcal/mol.

A component analysis was also performed for the K198*L* mutant (Table 6). Almost all of the binding free energy difference arises from four residues and the desolvation terms. Asp233 interacts strongly with the Asn ammonium; Arg489 interacts strongly with the Asp side-chain; Glu235 repels the Asp side-chain and partially compensates the Arg489 contribution; Arg217 interacts favorably with the ligand backbone in both cases, but more so in the Asp case. The net result from all other residues is only −1.6 kcal/mol, indicating significant cancellation between many small contributions of opposite signs; a similar cancellation was found with MDFE for the native protein (Archontis *et al.*, 1998). The four above residues also made the largest contributions in the native protein (along with K198, which is now leucine-like), due to similar but not identical interactions (see Table 2). In fact, comparison of the four residues in the wild-type and the K198*L* mutant (considering only the direct interaction components ΔG_R^{DI}, equation (5)) shows that they contribute −17.0 and −29.2 kcal/mol, respectively, to ΔG_{bind} of Asp. Thus, the larger contribution of these residues in the K198*L* mutant essentially makes up for the contribution of Lys198 (−13.2 kcal/mol) in the wild-type. The different components in the K198*L* mutant correspond to somewhat different structural interactions (Figure 4): in K198*L*, Asp233 interacts directly with the ligand ammonium; the hydrogen bonds between the ligand and Arg489 are slightly longer (1.8 Å *versus* 1.7 Å); the ligand side-chain is farther away from Glu235, and Arg217 has strengthened one of its two hydrogen bonds to the ligand.

The effect of solvent on the residue contributions ΔG_R^{DI} can be characterized as above, by comparing the ΔG_R^{DI} to the values obtained without solvent (a dielectric of four throughout space). As with the native protein, the effect of solvent is to reduce the magnitude of the direct residue-ligand interactions. The protein desolvation term is of the same magnitude as for the native protein, but cancellation between the Asp and Asn complexes is poorer, and the overall term is rather large (5.6 kcal/mol). The ligand desolvation term is similar to that for the native protein above, contributing −5.4 kcal/mol to $\Delta\Delta G_{\text{bind}}$, so that the two desolvation terms cancel.

Applying PBFE to related binding processes

The above agreement between PBFE and MDFE free energies suggests that the PBFE method can be used to analyze related binding processes for which detailed experimental or MDFE results are not available. We consider first the binding of Asn to AspRS in a head-to-tail orientation found in the homologous asparagine synthetase. Next, we consider binding of Asp and Asn to a double and a triple mutant AspRS. The effect of protonating the active-site His449 (thought to be neutral) or deprotonating Lys198 (thought to be charged) is considered in the Supplementary Material.

Table 6. Residue contributions to binding free energies for mutant AspRSs

	K189*L*			K198L,Q199E,D233E		
Residue	Asp binding	Asn binding	Difference	Asp binding	Asn binding	Difference
Leu198	−0.1 (−0.3)	0.2 (0.1)	0.3 (0.4)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Gln/Glu199	0.2 (0.3)	0.1 (0.1)	−0.1 (−0.2)	1.8 (7.8)	−1.4 (−1.6)	−3.2 (−9.4)
Asp/Glu233	−3.2 (4.2)	−17.2 (−16.0)	14.0 (−20.4)	−4.4 (3.8)	−13.6 (−12.7)	−9.2 (−16.5)
Arg489	−17.5 (−24.2)	−2.8 (−3.6)	14.7 (20.6)	−7.6 (−15.3)	1.9 (1.8)	9.4 (17.1)
Glu235	4.3 (10.0)	−0.9 (−0.4)	−5.2 (−10.4)	1.1 (7.1)	−2.4 (−2.4)	−3.5 (−9.6)
Arg217	−12.8 (−25.3)	−6.7 (−10.4)	6.1 (15.0)	−4.0 (−13.8)	−6.9 (−11.5)	−2.9 (2.3)
Asp475	2.5 (9.6)	0.2 (0.7)	−2.2 (−9.0)	2.6 (10.0)	0.0 (−0.4)	−2.6 (−10.4)
Gln231	−0.1 (0.4)	−3.2 (−3.0)	−3.2 (−3.4)	−0.9 (−1.0)	0.0 (0.3)	0.9 (1.2)
His449	−3.5 (−3.9)	−0.2 (−0.6)	3.3 (3.4)	−0.7 (−1.3)	−1.1 (−0.9)	−0.4 (0.4)
Glu482	2.3 (9.4)	0.3 (1.2)	−2.0 (−8.3)	1.8 (7.3)	0.4 (0.8)	−1.4 (−6.5)
Arg537	−2.2 (−8.9)	0.0 (−0.5)	2.2 (8.4)	−2.1 (−7.8)	−0.4 (−1.2)	1.6 (6.6)
Asp175	1.5 (6.3)	0.0 (0.5)	−1.5 (−5.9)	1.4 (6.6)	0.1 (0.6)	−1.3 (−6.0)
Gln192	0.3 (0.9)	−0.1 (0.2)	−0.5 (−0.7)	1.4 (2.7)	0.5 (0.8)	−0.9 (−1.9)
Gln195	−1.6 (−1.4)	−0.3 (0.0)	1.3 (1.5)	−3.8 (−3.9)	−3.9 (−3.8)	−0.1 (0.1)
Glu219	0.7 (5.0)	0.3 (1.0)	−0.4 (−4.0)	1.4 (7.0)	0.3 (1.6)	−1.1 (−5.4)
Arg225	−1.2 (−5.9)	−0.3 (−0.7)	0.9 (5.2)	−1.4 (−6.7)	−0.4 (−1.2)	1.1 (5.5)
Glu228	0.7 (8.0)	−0.4 (−0.7)	−1.0 (−8.7)	0.8 (8.6)	0.0 (0.4)	−0.8 (−8.2)
Arg245	−1.1 (−6.2)	0.0 (−0.1)	1.1 (6.1)	−0.8 (−5.8)	0.1 (0.4)	0.9 (6.2)
Asp282	1.2 (6.2)	0.3 (0.7)	−0.9 (−5.5)	1.4 (6.3)	0.1 (0.3)	−1.3 (−6.0)
His448	−0.1 (−0.5)	0.1 (0.3)	0.1 (0.8)	−1.3 (−1.3)	0.0 (0.2)	1.3 (1.4)
Asp536	0.7 (5.0)	−0.6 (−0.6)	−1.3 (−5.7)	0.8 (5.1)	0.1 (0.5)	−0.7 (−4.6)
Protein desolvation	1.4	7.0	5.6	4.0	5.5	1.5
Ligand desolvation	18.3	12.9	−5.4	16.7	12.1	−4.6

Free energy components (kcal/mol). The residue components in the upper part of the Table correspond to the direct interaction term in the binding free energy (ΔG_R^{DI} in equation 5). Values in parentheses are calculated with a dielectric of four throughout space (infinite protein).

Effect of Asn orientation on binding

Both MDFE and the above PBFE calculations show that when Asn binds to AspRS in an Asp-like orientation, the binding is much weaker than that of an Asp residue. However, a different Asn orientation, not found by simulations starting with the Asp-like structure, may lead to stronger binding (Archontis *et al.*, 1998). An asparagine molecule was therefore built into the active site in the head-to-tail orientation, in which the Asn backbone carboxylate occupies approximately the position occupied by the Asp side-chain in the Asp:AspRS complex. This orientation is observed in the protein asparagine synthetase, whose active site has a high structural homology to AspRS (Nakatsu *et al.*, 1998). The most significant amino acid changes around the ligand binding pocket are Leu196, Glu199 and Thr236 in asparagine synthetase, in place of three positive residues (Lys, Arg, Arg) in AspRS. A 500 ps molecular dynamics simulation was performed with Asn in the head-to-tail orientation in wild-type AspRS, in which the Asn molecule remained in approximately the same position and the active site was stable overall (rmsd of 0.9 Å from the starting structure for backbone and C^{β} atoms). A typical snapshot from the trajectory is shown in Figure 5. The backbone carboxylate binds tightly to Arg489 and Lys198, in the same manner as the Asp side-chain in the native Asp:AspRS complex. Arg489 and Lys198 are themselves linked tightly through Glu235, and Lys198 and the ligand NH_3 group are linked through Asp233. The ligand side-chain does not contact Arg217, which now has direct interactions only with solvent, including six water molecules less than 3 Å away. A similar, fully solvated arrangement of Arg217 was observed in segments of earlier Asn:AspRS simulations (Archontis *et al.*, 1998).

The binding affinity of Asn in this head-to-tail orientation was calculated to be −5.5 kcal/mol, 12.6 kcal/mol better than in the normal orientation, but still 3.8 kcal/mol poorer than for Asp binding. A free energy difference of +3.8 kcal/mol corresponds to a probability of about 1.6 10^{-3} of erroneously binding an Asn instead of an Asp residue. However, the resulting Asn:AspRS structure is almost certainly unsuitable for the adenylation reaction to occur. Therefore, erroneous aminoacylation of tRNA(Asp) by Asn is eliminated not only because the protein penalizes Asn binding, but also because it penalizes ''chemically active'' Asn binding.

Binding of Asn to the double mutant AspRS(K198L,D233E)

Asn is calculated to bind to the double mutant AspRS(K198L,D233E) with a free energy of −7.5 kcal/mol, somewhat weaker than the K198*L* case above, but much stronger than binding to the native protein.

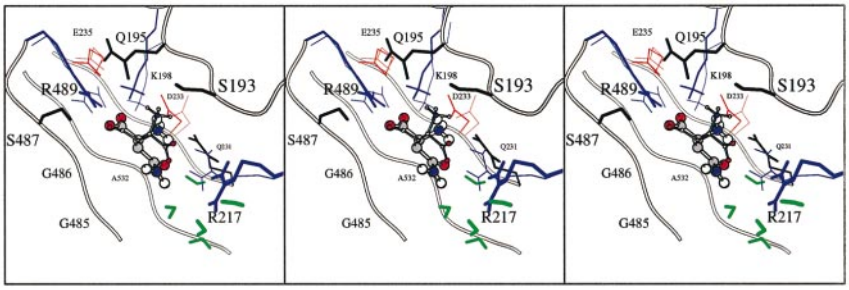

Figure 5. A typical snapshot of Asn bound to native AspRS in the ''head-to-tail'' orientation found in the homologous asparagine synthetase. The viewpoint and mode of representation are the same as in Figure 3. Positive charged residues are in blue, negatively charged are in red. The five water molecules closest to Arg217 are shown in green. The ligand and selected residues (R489, K198, D233, E235) from the native Asp:AspRS structure are also shown in thin lines, illustrating the similarity between the two complexes: ''head-to-tail'' Asn fits in the active site pocket without disturbing the surrounding side-chains.

The structures sampled in the simulation were described in brief by Archontis *et al*. (1998). They are similar to the K198*L* case above. However, there is a direct and short (1.7 Å) hydrogen bond between Arg489 and Glu233, instead of a water-mediated bond in K198*L*. The average rms deviation of the ligand from the Asp position in the native protein is 1.2 Å. Several changes occur around the ligand ammonium group: e.g. Gln195 hydrogen bonds to the ammonium instead of Gln231 above, and Ser193 hydrogen bonds to Glu233 instead of to the ammonium.

Binding of Asp to the double mutant protein was not examined.

Binding of Asp and Asn to the triple mutant AspRS(K198L,Q199E,D233E)

Asn binding to the triple mutant AspRS(K198L,Q199E,D233E) is similar to the single mutant K198*L*: $\Delta G_{bind}(\mathrm{Asn}) = -10.2$ kcal/mol. However, the presence of Glu199 leads to a much weaker Asp binding: $\Delta G_{bind}(\mathrm{Asp}) = +2.1$ kcal/mol, 11.4 kcal/mol less stable than to native AspRS. Therefore, the triple mutant is predicted to have reversed Asp/Asn binding affinities, as compared with native AspRS.

A typical snapshot from the Asn:protein complex is shown in Figure 6. The native aspartyl-adenylate:AspRS structure is shown for comparison. Although the Asn binding affinity is similar to that of Asp for the native protein, the ligand positions are very different in the two complexes. In the mutant MD structure, the orientation of the ligand is almost vertical with respect to the Arg489-Arg217 axis and to the β sheet underlying the active site (e.g. the Arg489-CZ...Asn-C^{γ}...Asn-C angle is 107°). In contrast, the Asp position in the native complexes lies approximately in this plane (e.g. the Arg489-CZ...Adenylate-C^{γ}...Adenylate-C angle is 154° for adenylate:AspRS; Poterszman *et al*., 1994). The average rmsd between the ligand positions in the two simulations is 1.17 Å. The Asn side-chain carbonyl no longer interacts directly with Arg489; rather, the side-chain makes hydrogen bonds to several waters, while its β-carbon packs against Arg489 and Glu233. These two residues form a salt bridge; Glu233 also makes a hydrogen bond to the ligand ammonium, similar to the

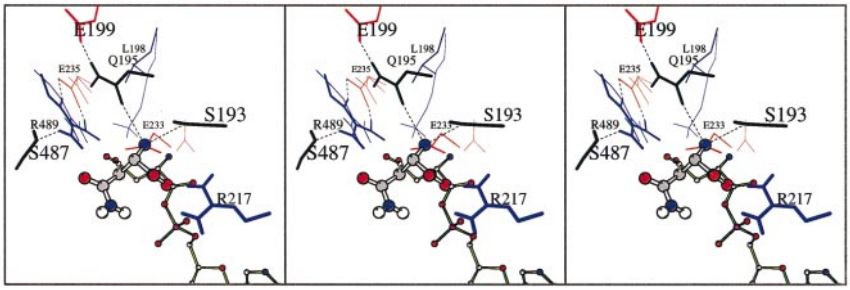

Figure 6. A typical snapshot of Asn bound to the triple mutant AspRS(K198L,Q199E,D233E). The viewpoint and mode of representation are the same as in Figures 3-5. Positively charged residues are in blue, negatively charged are in red. For comparison, selected groups from a native AspRS:aspartyl adenylate complex are shown in thin lines. The adenylate itself is shown in ball-and-stick mode with thin yellow bonds. The Asn ligand in the simulation is shown in ball-and-stick mode with thicker bonds. It is roughly perpendicular to the adenylate in the native complex.

AspRS(K198L,D233E) mutant case. Arg217 makes a single hydrogen bond to the Asn backbone carboxylate, instead of two in the native Asp:AspRS complex, and the Gln231 side-chain hydrogen bonds to both Glu233 and the ligand carboxylate. Although the backbone carboxylate has only shifted by 0.87 Å on average compared to the Asp:AspRS complex (Table 3), the structural rearrangements seen in Figure 6 could have a significant effect on both ATP binding and the rate of the adenylation reaction.

A free energy component analysis of the binding free energies was performed. The direct interaction contributions ΔG_R^{DI} (see equation (5)) of individual protein-ligand interactions to the binding free energies and the Asp/Asn differences are reported in Table 6. Six protein residues make large contributions to the binding free energy difference: Arg489 disfavors Asn by 9.4 kcal/mol, while Glu233 favors Asn by 9.2 kcal/mol. Arg217, Asp475, Glu235 and Glu199 all favor Asn by 2-3.5 kcal/mol. The large negative Glu233 component results in a stabilization of Asn; it reflects favorable interactions with the ammonium group of both Asp and Asn, diminished in the Asp case by repulsive interactions with the side-chain carboxylate. The large positive Arg489 component results in a stabilization of Asp; it arises mostly from favorable interactions with the ligand side-chain in the Asp state. As before, solvent reduces the direct contributions ΔG_R^{DI}, compared to the ''infinite protein'' situation. The protein desolvation term (equation (4)) makes again a moderate contribution to the individual binding free energies, which nearly cancels when the two ligands are compared. The ligand desolvation terms are similar to the native and K198*L* cases above, contributing −4.6 kcal/mol to $\Delta\Delta G_{bind}$.

Discussion

Meaning of component analysis: comparison of PBFE and MDFE results

It is interesting to compare and contrast the results of the PBFE free energy component analysis with the MDFE component analysis performed earlier by Archontis *et al*. (1998). The MDFE component analysis (Gao *et al*., 1989) refers to the alchemical transformation of Asp → Asn bound to the protein and Asp → Asn free in solution. This choice leads to the simplest MDFE simulations (Kollman, 1993) and is the most suitable for identifying important interactions in the bound and free systems (Boresch *et al*., 1994). In contrast, the PBFE analysis focusses on the actual binding reactions of Asp and Asn, which are more difficult to simulate with MDFE. The decomposition in equation (5) depends only on the endpoints considered; i.e. the individual components do not depend on the choice of the specific pathway, in contrast to the MDFE decomposition (Simonson & Brünger, 1992; Boresch *et al*., 1994; van Gunsteren *et al*., 1994). The same is true for the decomposition of the PBFE protein desolvation term (equation (6)) Nevertheless, a pathway can be constructed such that the desolvation terms and the direct interaction term can be associated with particular steps in the pathway (see Figure 2 by Gilson & Honig (1988)).

By use of the thermodynamic cycle (Figure 1), the MDFE and PBFE results for the double free energy difference $\Delta\Delta G_{bind}$ can be compared directly. As we have seen, the overall free energies obtained from MDFE and PBFE simulations are very similar if a protein dielectric constant of 4 is used in the latter. Correspondingly, it is possible to compare the components. In so doing, one must take into account the different treatment of solvent in the two sets of results. In the MDFE case, solvent contributions appear explicitly; in the PBFE case they are implicit, and appear as part of the individual ''residue'' free energy components. Since the MDFE solvent components total about −34 kcal/mol favoring Asn binding to AspRS, one expects the MDFE protein components to be significantly more positive than those obtained with PBFE. Indeed, the PBFE free energy components ΔG_R^{DI} have much smaller magnitudes than those from MDFE (see Tables 2 and 6). They fall in the same range as the free energy changes associated with double mutations, as discussed in the next section. Nevertheless, the same residues make the largest contributions in both cases, and the signs are the same. Moreover, if one takes the PBFE interaction free energy with a dielectric constant of one (i.e. multiplying by four the values in parentheses in Tables 2 and 6) the PBFE values are overall much more similar to the MDFE results in magnitude (though this is not true for all residues). The empirical choice of four for the protein dielectric constant with PBFE appears to give the most consistent results for binding phenomena (Gilson & Honig, 1988; Hendsch & Tidor, 1994). It is justified in part by the use of fixed structures in the calculations (i.e. structural relaxation of the protein in response to a mutation or ligand binding is treated as a rearrangement of polarization charge), although the primary justification is that it gives

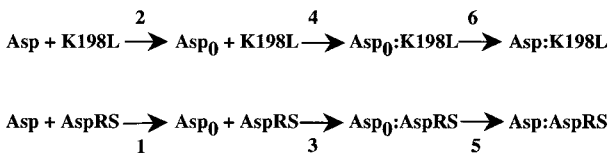

Figure 7. Thermodynamic cycle for analyzing residue contributions to the binding free energies (see the text). Binding is decomposed into three steps: first the ligand charges are removed in solution (Asp → Asp0), then the neutral analogue binds, then the charges are restored. The upper processes correspond to the mutant K198*L* protein.

agreement with the MDFE calculations. For MDFE, a dielectric of unity is justified by the fact that the parameterization of the charges used such a dielectric constant (see, for example, Mackerell *et al.* (1998)).

The role of residues having small free energy components was also discussed in the published MDFE study of the Asp → Asn transformation in AspRS (Archontis *et al.*, 1998). It was noted that residues that make small direct free energy contributions could nevertheless affect the overall free change indirectly, e.g. by stabilizing favorable conformations of other, directly interacting residues. This point can be made more explicit by considering the thermodynamic cycle in Figure 7. Binding of Asp is decomposed into three steps, involving a completely uncharged intermediate, Asp_0. For simplicity, we assume that the mutation (K198L) removes the atomic charges of the mutated residue *R*. With this assumption, the residue *R* contributes a very small component to the upper legs of the cycle (arising from van der Waals interactions only). The free energy changes for legs 1 and 2 cancel exactly. Since Asp_0 is completely uncharged, legs 3 and 4 correspond to the protein desolvation term, which cancels approximately for the present system (consider the PBFE protein desolvation terms for Asn binding to the native and K198L proteins above). As a consequence, the overall binding free energy change $\Delta G^{\mathrm{mut}}_{\mathrm{Asp}} - \Delta G^{\mathrm{nat}}_{\mathrm{Asp}}$ is approximately $\Delta G_6 - \Delta G_5$. Expressing these as sums over free energy components, we have:

$$\Delta G^{\mathrm{mut}}_{\mathrm{Asp}} - \Delta G^{\mathrm{nat}}_{\mathrm{Asp}} \approx \Delta G_6 - \Delta G_5 = -\Delta G^{\mathrm{R}}_5 + \sum_{R' \neq R} (\Delta G^{R'}_6 - \Delta G^{R'}_5) \quad (7)$$

where ΔG^{R}_5 is the free energy component of residue *R* for leg 5 of the cycle. The overall binding free energy change takes the form of a direct contribution from the residue being mutated and a sum of ''indirect'' contributions from all other residues, where interactions between these residues and the ligand appear in the form of differences between the native and mutant proteins.

Meaning of component analysis: relation to mutagenesis results

In AspRS, as in many other proteins, specific molecular recognition is achieved by non-covalent interactions between nearby chemical groups. To determine the importance of individual groups, an approach used in many experimental studies is to delete selected groups or introduce new ones by means of mutagenesis (Fersht, 1999, 1988; Horovitz, 1996). In simple cases, it has been shown that the resulting binding free energy change can be attributed to the modified groups; i.e. it is a measure of the direct interactions between the mutated side-chain and the ligand. In cases where three-dimensional information is not available, this approach is often used to infer possible interactions. For example, if the binding of an Asp ligand decreases significantly when a nearby Lys side-chain is deleted, direct interactions between the Lys side-chain and the ligand would be assumed to play an important role; conversely, if the Lys deletion mutant has a free energy similar to the wild-type, it is concluded that the Lys is not important for binding (but see below). In a similar way, when Gln199 is changed to a Glu residue in the present system, the large Asp binding free energy decrease would be attributed to unfavorable interactions of Asp with the Glu199 side-chain.

The free energy component analysis performed here for two ligands and three protein variants allows us to check this type of hypothesis. Since the binding of Asp and Asn to the native protein are very different, large effects are expected. As shown below, although direct interactions do play a role, the most important effect of the mutations in the present cases is to allow the formation of alternate structures, where strong interactions are made by other (unmutated) side-chains. The latter can then make much larger contributions to the binding free energy change than the mutated residues. Analogous observations have been made in several experimental studies (Fersht, 1988; Horoviz, 1996; Otzen & Fersht, 1999). This suggests that the magnitude of the ''interaction'' between an Asn ligand and a Lys, or an Asp ligand and a Glu, as measured, by such binding/mutagenesis experiments, can depend not only on their relative distances and orientations, but also on the nature of the surrounding pocket, the possibility for the ligand to shift its position, and the availability of other nearby groups with which strong interactions can be made.

To demonstrate this for the K198*L* and (K198L,Q199E,D233E) mutations, we first focus on the changes in Asp binding when the protein is mutated. We concentrate on the ''direct interaction'' term in the binding free energy and the corresponding components, $\Delta G^{\mathrm{DI}}_{\mathrm{R}}$, equation (5). The residue components in the protein desolvation term are smaller, and are expected to vary less from one protein variant or ligand to the next. We denote the protein variants by ''nat'', ''mut1'' and ''mut3'', and the binding free energy of e.g. Asp to nat by $\Delta G^{\mathrm{nat}}_{\mathrm{Asp}}$. The K198*L* mutation changes Asp binding by $\Delta G^{\mathrm{mut1}}_{\mathrm{Asp}} - \Delta G^{\mathrm{nat}}_{\mathrm{Asp}} = -2.1$ kcal/mol. The Lys198 free energy component in the native protein is −13.2 kcal/mol (or −11.1 kcal/mol including the protein desolvation term); the Leu198 component is −0.1 kcal/mol in the mutant. Therefore, the Lys198 free energy component for Asp binding in the native complex does not predict even the rough magnitude of the binding free energy change associated with its deletion. The small binding free energy change results from two main cancelling effects: the direct Asp-Lys198 interaction is lost in the mutant complex, but it is partly replaced by a direct interaction between

Table 7. Selected contributions to binding free energy differences

	Total	K/L198	D/E233	R489	E235	R217	Q/E199	LD[a]	PD[b]
$\Delta\Delta G^{nat}_{bind} = \Delta G^{nat}_{Asn} - \Delta G^{nat}_{Asp}$	16.4	**17.6**	−8.3	21.9	−10.2	4.3	0.0	−5.1	−0.7
$\Delta\Delta G^{mut1}_{bind} = \Delta G^{mut1}_{Asn} - \Delta G^{mut1}_{Asp}$	0.2	**0.3**	**−14.0**	14.7	−5.2	6.1	**−0.1**	−5.4	5.6
$\Delta\Delta G^{mut3}_{bind} = \Delta G^{mut3}_{Asn} - \Delta G^{mut3}_{Asp}$	−12.3	0.0	**−9.2**	9.7	−3.5	−2.9	**−3.2**	−4.6	1.5
$\Delta\Delta G^{mut1}_{bind} - \Delta\Delta G^{nat}_{bind}$	−16.2	**−17.3**	−5.7	−7.2	5.0	1.8	−0.1	−0.3	6.3
$\Delta\Delta G^{mut3}_{bind} - \Delta\Delta G^{mut1}_{bind}$	−12.5	−0.3	**4.8**	−5.0	1.7	−9.0	**−3.3**	−4.6	1.5

Free energy components (kcal/mol). Direct interaction components only (equation 5); the protein desolvation components (equation 6) are much smaller. The contributions of residues being mutated are in bold.

[a] Total ligand desolvation term.

[b] Total protein desolvation term.

Asp233 and the Asp ammonium, as the ligand shifts slightly away from the native position. Snapshots of the Asp:AspRS(K198*L*) and Asp:AspRS complexes are shown in Figure 4. Tables 2 and 6 show that several other residues and desolvation effects also play a role. It is clear that in this case, the binding free energy change $\Delta G^{mut1}_{Asp} - \Delta G^{nat}_{Asp}$ does not measure the magnitude of the Lys198-Asp interaction *per se*; rather it depends also on the availability of the nearby, partly ''unused'' Asp233 carboxylate. To demonstrate this experimentally, a double mutant cycle, K198L, D233E, and (K198L,D233E); would be appropriate. However, without the simulations (including the component analysis) or an experimental K198L structure, it would not be obvious what residues to mutate.

The effect on Asp binding of the Q199E and D233E mutations is similar (Table 7). The (Q199E,D233E) mutations lead to a much weaker binding affinity: $\Delta G^{mut3}_{Asp} - \Delta G^{mut1}_{Asp} = 13.5$ kcal/mol. Most of the difference arises from Arg489 and Arg217. The (Q199E,D233E) mutation leads to positional shifts in the pocket that modify the interactions of Arg489 and Arg217 with the ligand. Here again, the free energy components of Q199 and D233 in the initial state do not predict the binding free energy change. However, the free energy components in the final state do identify the important interactions for binding to the modified protein.

We consider next the discrimination between Asp and Asn. The most important free energy components are grouped in Table 7. The binding free energy differences between Asp and Asn are denoted e.g. $\Delta\Delta G^{nat}_{bind} = \Delta G^{nat}_{Asn} - \Delta G^{nat}_{Asp}$. Comparing the native and mutant (mut1) protein, the overall difference $\Delta\Delta G^{mut1}_{bind} - \Delta\Delta G^{nat}_{bind}$ is 0.2-16.4 = − 16.2 kcal/mol. Lys198 makes a large contribution (17.6 kcal/mol) to $\Delta\Delta G^{nat}_{bind}$, while Leu198 contributes negligibly (0.3 kcal/mol) to $\Delta\Delta G^{mut1}_{bind}$. The difference, −17.3 kcal/mol, is close to the overall effect, −16.2 kcal/mol. However, there are several other large contributions that cancel, e.g. from Asp233, Arg489, Glu235. These are related to positional changes in the pocket that could not be predicted without analyzing the structures of all four relevant complexes. The Asp233 contribution reflects a stronger interaction with the Asn ammonium group in the mutant protein, as described above. The contributing groups (Asp233 carboxylate, ligand ammonium) are present in all four states (Asp:nat, Asn:nat, Asp:mut1, Asn:mut1), but they interact differently in each case. The Glu235 contribution reflects a shift of the Asp ligand away from this residue in the mutant protein. The Arg489 contribution reflects a weak favorable interaction with Asn in the mutant, while strong favorable interactions with the Asp side-chain approximately cancel between the native and mutant protein. The relative magnitude of all these interactions therefore depends on the details of the structural shifts in the four states, and their cancellation in $\Delta\Delta G^{mut1}_{bind} - \Delta\Delta G^{nat}_{bind}$ must be viewed as fortuitous. Thus, in both the Asp binding and the difference between Asp and Asn, the mutation K198L plays an important role, as expected. However in Asp binding, the change produced by the mutation in other interactions essentially cancels the direct K198L effect, while in the difference between Asp and Asn, all the terms associated with the other (unmutated) residues cancel approximately.

Finally, comparing the binding specificities of mut1 and mut3, a similar complexity is seen. The overall specificity change is very large: $\Delta\Delta\Delta G = \Delta\Delta G^{mut3}_{bind} - \Delta\Delta G^{mut1}_{bind} = -12.3 - 0.2 = -12.5$ kcal/mol. The negative sign indicates that discrimination against Asn has been reduced (mut1 is non-specific, while mut3 favors Asn over Asp). Two mutations are made in going from mut1 to mut3: Q199E and D233E. The first leads to a direct contribution of −3.3 kcal/mol to $\Delta\Delta\Delta G$, mainly reflecting weak interactions between Glu199 and Asp (repulsive) and Asn (attractive) in mut3. The second mutation, which preserves the net charge, leads to a larger direct contribution of +4.8 kcal/mot. This mainly reflects strong interactions of Asp/Glu233 with the Asn ammonium in mut1 and mut3, which are stronger in mut1. Two residues that are not mutated make large, only partly cancelling contributions: both Arg489 and Arg217 strongly favor Asp binding to mut1, with smaller contributions in mut3. The weaker Arg489 contribution in mut3 reflects a positional shift of Asp away from Arg489, while the weaker Arg217 contribution in mut3 reflects a shift of Arg217 away from the ligand and into solvent (Figure 6).

Here, as in the nat → mut1 transformation, the main effect of the mutation is to allow or force the Asp and Asn ligands to shift into new positions, where new interactions are made and existing interactions are modified. The magnitude of the individual free energy differences, the double free energy differences $\Delta\Delta G_{bind}$, and the specificity change $\Delta\Delta\Delta G$ all depend on the details of the structural shifts. The specificity change is much larger than the direct contribution of the two mutated residues.

A combined molecular dynamics/continuum electrostatics approach for binding calculations

Continuum models are being used increasingly to study biomolecular recognition (Schaefer *et al*., 1998; Massova & Kollman, 1999), and they have proved surprisingly accurate for some applications. There are two important points concerning the use of continuum models, in general, and PBFE calculations in particular. First, it is essential to validate PBFE by comparison with experiment and/or MDFE calculations. As demonstrated above, when data from MDFE simulations are available for a given system, the continuum model can be parameterized (essentially by adjusting the protein dielectric constant) and tested, before applying it to related problems. Second, it is important to obtain PBFE results by averaging over a series of structures obtained by MD simulations of the system of interest; the latter can be done efficiently because no MDFE simulations are involved. By using MD simulations of all the variants of the system, structural changes due to the variations in the protein and/or ligand are included. Further, because the PBFE calculations are sensitive to the details of the structures, an average over a time series of structures is expected to give more meaningful results.

The MD simulations show that while the structures are all broadly similar, significant structural rearrangements do occur for the different protein variants and the different ligands. This confirms the necessity of using, for each state, structures representative of that state. Incorrect results are obtained, for example, if structures of the native protein are used to estimate the free energies of binding to the K198*L* protein: with the native structures, the binding free energies are 1.0 kcal/mol for Asp and −3.4 kcal/mol for Asn (instead of −11.4 and −11.2 kcal/mol with the K198*L* structures). The possibility of such a model bias should be examined for each application. This permits one to test the possibility of predicting accurately the effect of many different mutations from a simulation of a single reference state Massova & Kollman (1999).

Here, this recommendation was followed for the bound states, but not the unbound protein: the unbound state of the protein was modelled by simply removing the ligand from the structure of the relevant protein: ligand complex. This approximation was made (as in most applications) for a technical reason: it allows an exact cancellation of artificial grid effects from the binding free energies. This constraint could in principle be avoided, e.g. by using so-called boundary element formulations of the continuum model (Zauhar & Morgan, 1985; Juffer *et al*., 1991), or more complex implementations of the Poisson-Boltzmann approach that solve directly for the solvent reaction potential (Zhou *et al*., 1996). For the present system, crystal structures of the proteins from yeast, *Thermus thermophilus* and *Pyrococcus kodakareansis* (Schmitt *et al*., 1998; Poterszman *et al*., 1994; Sauter *et al*., 2000) are all available with and without bound aminoacid (or aminoacyl adenylate); in all cases the active sites are very similar, so that the approximation should be acceptable.

Engineering of Asn binding to AspRS

The engineering of modified specificities into aminoacyl-tRNA synthetases is an important technological goal, which has recently allowed the genetic code to be artificially extended to include unnatural amino acid residues (Liu & Schultz, 1999). The present work was motivated by unsuccessful experimental attempts to engineer Asn activity into AspRS (Cavarelli *et al*., 1994); it was not known whether the lack of activity was due to lack of Asn and/or ATP binding, or to binding in an inactive conformation. The calculations above strongly suggest that the simple ideas initially derived from sequence comparisons between AspRS and AsnRS were correct: the mutations considered above, proposed from these comparisons, should all lead to a reasonable level of Asn binding. In addition, the triple mutant is predicted to have strongly reduced Asp binding and therefore a reversed specificity, favoring Asn binding over Asp.

The calculations show that to bind Asp, the native and mutant proteins must overcome a desolvation penalty of about 16-18 kcal/mol (the ligand desolvation term in Tables 2 and 6). This is accomplished by providing strong ionic interactions with Arg489 and (in the native protein) with Lys198. While Asn is less costly to desolvate (11-12 kcal/mol), its interactions in the native pocket are distinctly unfavorable, unless it binds in the head-to-tail orientation. Deleting Lys198 is sufficient to bring Asn binding up to the level of native Asp binding, through a combination of factors, including more favorable interactions with both Arg489 and Asp/Glu233. Head-to-tail Asn binding is probably also present with the mutant proteins K198L and (K198L,D233E). However, the head-to-tail affinity is probably similar to the native case (just as the Asp binding is similar to the native case), and therefore 2-8 kcal/mol weaker than the ''normal'' orientation (see Table 1).

The experimental data show that none of these mutants adenylates Asn or Asp at an appreciable level. The MD simulations suggest that the structures of the Asp and Asn complexes with the

mutant proteins, while broadly similar to the native Asp:AspRS complex, do have significant differences that may affect ATP binding and/or the adenylation rate constants. It is clear that experimental measurements of the binding affinities are needed to test this result.

Conclusions

We have presented an approach that combines molecular dynamics (MD), molecular dynamics free energy simulations (MDFE), and Poisson-Boltzmann free energy simulations (PBFE) to obtain reliable results efficiently for a set of related protein:ligand complexes. The first step is to use MDFE and PBFE (averaged over trajectories) to study one binding process thoroughly, and validate the latter by comparison with the former for that system. This is then followed by MD and PBFE calculations to study a series of related ligands or mutations. Such a paradigm is particularly useful for protein or ligand design, e.g. for understanding a series of mutations of a given protein with the same ligand or a series of related ligands. An advantage over experimental studies is the possibility of decomposing the overall free energies into contributions from microscopic interactions. Such a component analysis is presented here for PBFE free energies, and compared for the first time with MDFE results for the same system. Several general principles for ligand binding are illustrated. In particular, when a mutation in the protein induces structural shifts in the binding pocket, residues that are not mutated can make new interactions with the ligands, and consequently make large contributions to the change in the binding free energies. In such cases, a large binding free energy change does not necessarily indicate a large direct interaction between the ligand and the mutated side-chain, either in the native or the mutant structure; conversely, a small free energy change does not demonstrate that the residue is not making a strong direct interaction.

The present study addressed several important aspects of structure-function relations in aspartyl-tRNA synthetases that are unresolved experimentally. We have suggested that the substrate analogue Asn binds to wild type AspRS predominantly in a head-to-tail orientation which is incompetent for adenylation. If so, amino acid specificity would be achieved in part by preventing Asn binding in an active conformation. While pK_a calculations for His449 were not conclusive (see Supplementary Material), the structures sampled in molecular dynamics simulations strongly suggest that this residue is neutral. Lys198 pK_a calculations (Supplementary Material) suggest that Asn binding to native AspRS with deprotonated Lys198 is even less probable than Asn binding with protonated Lys198. Finally, the calculations suggest that it should be possible to engineer Asn binding into AspRS by a single mutation (K198L), and to reverse the Asp/Asn binding specificity by the (K198L,Q199E,D233E) triple mutation. The resulting mutants have been shown experimentally to be inactive for adenylation of Asp and Asn. As the structures sampled in the simulations have significant differences from the native Asp:AspRS complex, we conclude that the lack of observed activity is likely to be due to structural deformations in the mutant complexes; i.e. we find that engineering a modified catalytic activity can be significantly more difficult than engineering a modified binding specificity.

Methods

Poisson-Boltzmann calculations

To obtain the ligand binding free energy with the Poisson-Boltzmann approach, we require (see equation (4)) the electrostatic potentials calculated with the protein alone, the ligand alone, and the protein:ligand complex. These are obtained by solving the Poisson or Poisson-Boltzmann equation on a three-dimensional grid with a standard finite-difference method. For the three states, we use the same three-dimensional grid and the same three-dimensional coordinates, so that the various species are localized in the same positions on the grid. This allows artificial contributions to the potential arising from the grid to be subtracted out exactly (Gilson & Honig, 1988). Finite-difference Poisson-Boltzmann calculations (FDPB) were done for multiple (usually ~10-50) structures taken from 400-600 ps molecular dynamics simulations of the protein:ligand complexes in solution, as described below. Averaging over multiple structures is expected to improve the precision and accuracy of the FDPB results (Vlijmen *et al.*, 1998, Massova & Kollman, 1999). For each structure, the FDPB equation was solved first on a large, coarse grid (0.8 Å spacing, 96 Å side), with screened Coulombic potentials on the grid boundary, then again on a smaller, finer grid (0.4 Å spacing, 48 Å side) with the coarse solution as boundary conditions (''focussing'' method (Gilson *et al.*, 1988)).

The solvent dielectric constant was set to 80. Those of the protein and ligand are taken to be equal; values of one to four were compared. Atomic radii and charges were taken from the CHARMM22 force field used in the MDFE simulations (Mackerell *et al.*, 1998). The radii correspond to the minimum in the van der Waals potential for each CHARMM22 atom type (Roux *et al.*, 1990). The protein-solvent boundary was defined by the molecular surface, constructed using a solvent probe radius of 2 Å. In selected cases, the effect of including one or two waters explicitly in the model was explored, as well as the effect of changing the solvent probe radius. Calculations were done with the UHBD program (Madura *et al.*, 1995). Most calculations only included protein atoms contained within a spherical region of 20 Å radius centered on the ligand side-chain, the rest of the protein being neglected; it is replaced by bulk solvent. A few calculations were done with the entire protein, demonstrating that this spatial cutoff introduced an error of only about 1 kcal/mol (~1%; data not shown) in the binding free energies. Most calculations used zero ionic strength; a few were carried out with an ionic strength corresponding to 0.1 M of monovalent counterions.

Molecular dynamics and MDFE simulations

For the Asp:AspRS and Asn:AspRS complexes, published MD simulations (Archontis *et al*., 1998; Simonson *et al*., 1997) were employed. In brief, a spherical model of 20 Å radius was used, including most of the enzyme active site domain, the ligand, and 384 water molecules, with stochastic boundary conditions (Brünger *et al*., 1984). The ligand is at the center of the sphere. No cutoff for electrostatic interactions was used; i.e. a multipole approximation was employed for distant groups (Stote *et al*., 1991). Ionized side-chains close to the boundary had their charges reduced so as to mimic their screening by bulk solvent, with scaling factors derived from a continuum solvent model (Simonson *et al*., 1997). The CHARMM22 force field was used (Mackerell *et al*., 1998) and the calculations were done with the CHARMM program (Brooks *et al*., 1983).

An additional simulation was performed for each ligand, with His449 in the doubly protonated state. All other simulations assume His449 to be singly protonated. For the Asn:AspRS complex, a simulation was performed with Asn in the head-to-tail orientation described in the Introduction. To obtain the initial Asn position, the crystal structure of asparagine synthetase (Nakatsu *et al*., 1998) was least-squares superimposed onto that of AspRS, using conserved active site residues as a reference. This brought the Asn ligand from the asparagine synthetase structure into a reasonable starting position with respect to the AspRS pocket; this position was used to initiate energy minimization followed by the MD simulation. For the complexes of Asp and Asn with the AspRS(K198L) and AspRS(K198L,Q199E,D233E) mutant proteins, and of Asn with AspRS(K198L,D233E), additional simulations were performed. The mutated residues were placed initially by superimposing conserved portions of the side chains on the native residues. This completely determines the position of Glu in the Q199E case. For the D233E and K198L cases, the initial orientation of the terminal portion of the side chain was random; energy minimization and MD equilibration with backbone restraints brought them into reasonable orientations; e.g. the Glu233 side-chain carboxylate was brought into approximately the same plane as the Asp233 one. All the new simulations used the same protocol as that used by Simonson *et al*. (1997). No additional water molecules were needed, as the mutations approximately conserve volume. All simulations were of 400-600 ps duration, including equilibration (see Table 3).

The K198L single mutation is modelled by a simplified procedure, i.e. the lysine atoms and connectivity are retained, but their charges are changed to leucine-like values (see Figure 2). To reflect this, the mutation is denoted K198L. On the other hand, for the double and triple mutations, the leucine side-chain is modelled in the usual way, with the leucine atoms and connectivity.

The K198L mutation was also studied by MDFE simulations, following the same protocol as the earlier Asp → Asn simulations (Archontis *et al*., 1998; Simonson *et al*., 1997). In brief, the Lys198 side-chain charges were transformed in a linear manner into the mutant charges, using ten equally spaced windows and a thermodynamic integration analysis. The total simulation time for the transformation was 500 ps. The transformation was performed in the both the Asp:AspRS and the Asn:AspRS complexes. Details will be presented elsewhere.

Acknowledgments

Discussions with G. Eriani, J. Gangloff, D. Kern and M. Schaefer are gratefully acknowledged. Comments by M. Schaefer, who has developed a similar decomposition of the electrostatic free energy for interpreting pH effects on ligand binding, were particularly helpful. M.K. acknowledges partial support by the National Institutes of Health, the CNRS, and the Ministere de 1'Education. TS acknowledges support from the CNRS through a ''Young Group Leader'' grant. GA acknowledges support from the Cyprus Government through the grant ''From Strong Interactions to Molecular Recognition: Theoretical and Computational Studies''.

References

Agou, F., Quevillon, S., Kerjan, P. & Mirande, M. (1998). Switching the amino acid specificity of an aminoacyl-tRNA synthetase. *Biochemistry,* **37**, 11309-11314.

Archontis, G., Simonson, T., Moras, D. & Karplus, M. (1998). Specific amino acid recognition by aspartyl-tRNA synthetase studied by free energy simulations. *J. Mol. Biol.* **275**, 823-846.

Arnez, J. & Moras, D. (1997). Structural and functional consideration of the aminoacylation reaction. *Trends Biochem. Sci.* **22**, 211-216.

Bashford, D. & Karplus, M. (1990). The pK_as of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry,* **29**, 10219-10225.

Boresch, S. & Karplus, M. (1995). The meaning of component analysis: decomposition of the free energy in terms of specific interactions. *J. Mol. Biol.* **254**, 801-807.

Boresch, S., Archontis, G. & Karplus, M. (1994). Free energy simulations: the meaning of the individual contributions from a component analysis. *Proteins: Struct. Funct. Genet.* **20**, 25-33.

Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S. & Karplus, M. (1983). Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comp. Chem.* **4**, 187-217.

Brünger, A. T., Brooks, C. L. & Karplus, M. (1984). Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Letters,* **105**, 495-500.

Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M. & Mitschler, A. *et al*. (1994). The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J.* **13**, 327-337.

Cusack, S. (1995). Eleven down and nine to go. *Nature Struct. Biol.* **2**, 824-831.

Davis, M. & McCammon, J. (1991). Solving the finite-difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. *J. Comp. Chem.* **10**, 386.

de Prat Gay, G., Duckworth, H. & Fersht, A. (1993). Modification of the amino acid specificity of tyrosyl-tRNA synthetase by protein engineering. *FEBS Letters,* **318**, 167-171.

Eriani, G., Dirheimer, G. & Gangloff, J. (1990). Aspartyl-tRNA synthetase from *Escherichia coli*: cloning and characterization of the gene, homologies of its translated amino acid sequence with asparaginyl- and

lysyl-tRNA synthetases. *Nucl. Acids Res.* **18**, 7109-7118.
Fersht, A. (1988). Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. *Biochemistry,* **27**, 1577-1580.
Fersht, A. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Freeman, New York.
Fersht, A., Shi, J., Knill-Jones, J., Lowe, D., Wilkinson, A. & Blow, D., *et al*. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature,* **314**, 235-238.
Froloff, N., Windemuth, A. & Honig, B. (1997). On the calculation of binding free energies using continuum methods: application to the MHC class I protein-peptide interactions. *Protein Sci.* **6**, 1293-1301.
Gao, J., Kuczera, K., Tidor, B. & Karplus, M. (1989). Hidden I thermodynamics of mutant proteins: a molecular dynamics analysis. *Science,* **244**, 1069-1072.
Gilbart, J., Perry, C. & Slocombe, B. (1993). *Antimicrob. Agents Chemother.* **37**, 32-38.
Gilson, M. & Honig, B. (1988). Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins: Struct. Funct. Genet.* **4**, 7-18.
Gilson, M., Sharp, K. & Honig, B. (1988). Calculating electrostatic interactions in bio-molecules: method an error assessment. *J. Comp. Chem.* **9**, 327-335.
Hendsch, Z. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatics analysis. *Protein Sci.* **3**, 211-226.
Hendsch, Z. & Tidor, B. (1999). Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. *Protein Sci.* **8**, 1381-1392.
Hohsaka, T., Ashizuka, Y., Sasaki, H., Murakami, H. & Sisido, M. (1999). Incorporation of two different nonnatural amino acids independently into a single protein through extension of the genetic code. *J. Am. Chem. Soc.* **121**, 12194-12195.
Horovitz, A. (1996). Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* **1**, R121-R126.
Ibba, M. & Hennecke, H. (1995). Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows *in vitro* and *in vivo* synthesis of proteins containing unnatural amino acids. *FEBS Letters,* **364**, 3272-3275.
Janin, J. (1995). Protein-protein recognition. *Prog. Biophy. Mol. Biol.* **64**, 145-166.
Juffer, A., Botta, E., van Keulen, B., van der Ploeg, A. & Berendsen, H. (1991). The electric potential of a macromolecule in a solvent: a fundamental approach. *J. Comp. Phys.* **97**, 144.
Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395.
Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
Landau, L. & Lifschitz, E. (1980). *Electrodynamics of Continuous Media*, Pergamon Press, New York.
Lau, F. & Karplus, M. (1994). Molecular recognition in proteins. Simulation analysis of substrate binding by a tyrosyl-tRNA synthetase mutant. *J. Mol. Biol.* **236**, 1049-1066.
Liu, D. & Schultz, P. (1999). Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl Acad. Sci. USA,* **96**, 4780-4785.
Mackerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, I. & Field, M. *et al*. (1998). An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B,* **102**, 3586-3616.
Madura, J., Briggs, J., Wade, R., Davis, M., Luty, B. & Ilin, A., *et al*. (1995). Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comp. Phys. Commun.* **91**, 57-95.
Massova, I. & Kollman, P. (1999). Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* **121**, 8133-8143.
Meinnel, T., Mechulam, Y. & Blanquet, S. (1995). Aminoacyl-tRNA synthetases: occurrence, structure and function. In *tRNA: Structure, Biosynthesis and Function*, pp. 251-291, ASM Press, Washington, DC.
Nagai, K. & Mattaj, I. (1994). *RNA Protein Interaction*, Oxford University Press, Oxford.
Nakatsu, T., Kato, H. & Oda, J. (1998). Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase. *Nature Struct. Biol.* **5**, 15.
Normanly, J. & Abelson, J. (1989). tRNA identity. *Annu. Rev. Biochem.* **58**, 1029-1049.
Otzen, D. & Fersht, A. (1999). Analysis of protein-protein interactions by mutagenesis: direct versus indirect effects. *Protein Eng.* **12**, 41-45.
Pallanck, L., Pak, M. & Schulman, L. (1995). tRNA discrimination in aminoacylation. In *tRNA: Structure, Biosynthesis and Function*, pp. 371-393, ASM Press, Washington, D.C.
Poterszman, A., Delarue, M., Thierry, J. & Moras, D. (1994). Synthesis and recognition of aspartyl-adenylate by *Thermus thermophilus* aspartyl-tRNA synthetase. *J. Mol. Biol.* **244**, 158-167.
Rogers, N. (1986). The modelling of electrostatic interaction in the function of globular proteins. *Prog. Biophy. Mol. Biol.* **48**, 37-66.
Roux, B., Yu, H. & Karplus, M. (1990). Molecular basis for the Born model of solvation. *J. Phys. Chem.* **94**, 4683-4688.
Sauter, C., Lorber, B., Cavarelli, J., Moras, D. & Giégé, R. (2000). The free yeast aspartyl-tRNA synthetase differs from the tRNA(Asp)-complexed enzyme by structural changes in the catalytic site, hinge region, and anti-codon binding domain. *J. Mol. Biol.* **299**, 1313-1324.
Schaefer, M. & Froemmel, C. (1990). A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J. Mol. Biol.* **216**, 1045-1066.
Schaefer, M., Vlijmen, H. V. & Karplus, M. (1998). Electrostatic contributions to molecular free energies in solution. *Advan. Protein Chem.* **51**, 1-57.
Schapira, M., Totrov, M. & Abagyan, R. (1999). Prediction of the binding energy for small molecules, peptides and proteins. *J. Mol. Recog.* **12**, 177-190.
Schmitt, E., Moulinier, L., Fujiwara, S., Imanaka, T., Thierry, J. & Moras, D. (1998). Crystal structure of aspartyl-tRNA synthetase from *Pyrococcus kodakaraensis KOD*: archaeon specificity and catalytic mechanism of adenylate formation. *EMBO J.* **17**, 5227-5237.

Sharp, K. & Honig, B. (1990). Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.* **94**, 7684-7692.

Sharp, K. & Honig, B. (1991). Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **19**, 301-332.

Simonson, T. & Brünger, A. T. (1992). Thermodynamics of protein-peptide binding in the ribonuclease S system studied by molecular dynamics and free energy calculations. *Biochemistry*, **31**, 8661-8674.

Simonson, T., Archontis, G. & Karplus, M. (1997). Continuum treatment of long-range interactions in free energy calculations. Application to protein-ligand binding. *J. Phys. Chem. B*, **101**, 8349-8362.

Srinivasan, J., Cheatham, T., Cieplak, P., Kollman, P. & Case, D. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **120**, 9401-9409.

Stote, R., States, D. & Karplus, M. (1991). On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys.* **88**, 2419-2433.

van Gunsteren, W., King, P. & Mark, A. (1994). Fundamentals of drug design from a biophysical standpoint. *Q. Rev. Biophys.* **27**, 435-481.

Vijayakumar, M., Wong, K., Schreiber, G., Fersht, A., Szabo, A. & Zhou, H. (1998). Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar. *J. Mol. Biol.* **278**, 1015-1024.

Vlijmen, H. V., Schaefer, M. & Karplus, M. (1998). Improving the accuracy of protein pK_a calculations: conformational averaging versus the average structure. *Proteins: Struct. Funct. Genet.* **33**, 145-158.

Vorobjev, Y., Almagro, J. & Hermans, J. (1998). 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: Struct. Funct. Genet.* **32**, 399-413.

Wade, R., Luty, B., Demchuk, E., Madura, J., Davis, M., Briggs, J. & McCammon, J. (1994). Simulation of enzyme-substrate encounter with gated active sites. *Nature Struct. Biol.* **1**, 65-69.

Wang, L., Magliery, T., Liu, D. & Schultz, P. (2000). A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* **122**, 5010-5011.

Warwicker, J. & Watson, H. (1982). Calculation of the electrostatic potential in the active site cleft due to α helix dipoles. *J. Mol. Biol.* **571**, 671.

Westhof, E. (1997). Nucleic acids. From self-assembly to induced-fit recognition. *Curr. Opin. Struct. Biol.* **7**, 305-309.

Zauhar, R. & Morgan, R. (1985). A new method for computing the macromolecular electric potential. *J. Mol. Biol.* **186**, 815.

Zhou, Z., Payne, P., Vasquez, M., Kuhn, N. & Levitt, M. (1996). Finite-difference solution of the Poisson-Boltzmann equation: complete elimination of self-energy. *J. Comp. Chem.* **11**, 1344-1351.

Edited by A. R. Fersht

(Received 5 June 2000; received in revised form 30 October 2000; accepted 31 October 2000)

http://www.academicpress.com/jmb

Supplementary Material for this paper is available on IDEAL