THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 33, pp. 23792–23803, August 18, 2006

Molecular Dynamics Simulations Show That Bound Mg^{2+} Contributes to Amino Acid and Aminoacyl Adenylate Binding Specificity in Aspartyl-tRNA Synthetase through Long Range Electrostatic Interactions*[S]

Received for publication, March 27, 2006, and in revised form, June 12, 2006 Published, JBC Papers in Press, June 14, 2006, DOI 10.1074/jbc.M602870200

Damien Thompson[‡§1] **and Thomas Simonson**[‡2]

From the [‡]*Laboratoire de Biochimie, CNRS, UMR7654, Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France and* [§]*Tyndall National Institute, University College Cork, Cork, Ireland*

Molecular recognition between the aminoacyl-tRNA synthetase enzymes and their cognate amino acid ligands is essential for the faithful translation of the genetic code. In aspartyl-tRNA synthetase (AspRS), the co-substrate ATP binds preferentially with three associated Mg^{2+} cations in an unusual, bent geometry. The Mg^{2+} cations play a structural role and are thought to also participate catalytically in the enzyme reaction. Co-binding of the ATP·Mg_3^{2+} complex was shown recently to increase the Asp/Asn binding free energy difference, indicating that amino acid discrimination is substrate-assisted. Here, we used molecular dynamics free energy simulations and continuum electrostatic calculations to resolve two related questions. First, we showed that if one of the Mg^{2+} cations is removed, the Asp/Asn binding specificity is strongly reduced. Second, we computed the relative stabilities of the three-cation complex and the 2-cation complexes. We found that the 3-cation complex is overwhelmingly favored at ordinary magnesium concentrations, so that the protein is protected against the 2-cation state. In the homologous LysRS, the 3-cation complex was also strongly favored, but the third cation did not affect Lys binding. In tRNA-bound AspRS, the single remaining Mg^{2+} cation strongly favored the Asp-adenylate substrate relative to Asn-adenylate. Thus, in addition to their structural and catalytic roles, the Mg^{2+} cations contribute to specificity in AspRS through long range electrostatic interactions with the Asp side chain in both the pre- and post-adenylation states.

Specific molecular association is fundamental to many biochemical processes and is frequently used to transfer energy or information. Aminoacyl-tRNA synthetases (aaRSs)[3] are an important class of information-processing enzymes (1–4). Each aaRS catalyzes the aminoacylation of a specific tRNA by a cognate amino acid, establishing the genetic code (5–7). The amino acid (aa) and ATP react first to form an aminoacyl adenylate; in a second step, the amino acid is transferred to the tRNA. Some aaRSs have evolved a third, editing step where incorrect tRNA-aa products are hydrolyzed (8–10). Specificity for the aa and the tRNA can arise from different component steps, such as binding or release of the amino acid or binding or acylation of the tRNA. Furthermore, through their particular combination of reversible binding and irreversible reaction steps, aaRSs can use specificity in successive steps to amplify the overall effect (11). For example, amino acid specificity can be established in the aaRS·aa complex, the aaRS·aaAMP complex, or both.

The 20 aaRSs form two distinct classes of 10 members each (6). Below we have focused mainly on aspartyl-tRNA synthetase (AspRS), one of the best studied aaRSs. AspRS belongs to the aaRS class II, forming a subclass IIb with AsnRS and LysRS. Although aaRSs are generally very amino acid-specific, they have a complex evolutionary history (4, 12), which has led to a remarkable diversity in the modern enzymes. Within class IIb, for example, LysRS is very specific in yeast; but in *Escherichia coli*, it is more promiscuous (13). In *E. coli*, AspRS discriminates strongly against Asn, but more weakly against D-Asp (10). Several aaRSs achieve a high fidelity through their editing step. An ambiguous IleRS was constructed recently by deleting the IleRS editing domain (14); the resulting Ile/Val ambiguity actually led to a growth advantage in bacteria. As a last example, many archaebacteria lack AsnRS and produce $tRNA^{Asn}$-Asn by an indirect route; $tRNA^{Asn}$ is aspartylated by a "nondiscriminating" AspRS (which accepts both $tRNA^{Asp}$ and $tRNA^{Asn}$), and then the Asp moiety is amidated (1).

The mechanism of the two enzyme reactions, amino acid adenylation and tRNA aminoacylation, is qualitatively understood in both aaRS classes (1). For AspRS, crystal structures are available from several organisms, encompassing the three "kingdoms" of life and the whole reaction pathway: apoenzyme, complexes with Asp alone, ATP alone, aspartyl-adenylate alone (AspAMP) (15–20), and complexes with $tRNA^{Asp}$ present (21–23). Each substrate is recognized by side chains conserved throughout AspRSs (Asp side-chain recognition) or throughout all or most of class II (ATP,

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental material including Fig. SM1 and Table SM1.

[1] Supported in part by an Egide postdoctoral fellowship.

[2] To whom correspondence should be addressed: Laboratoire de Biochimie (CNRS, UMR7654), Dept. of Biology, Ecole Polytechnique, 91128 Palaiseau, France. Tel.: 33-169-33-38-81; Fax: 33-169-33-30-13; E-mail: thomas.simonson@polytechnique.fr.

[3] The abbreviations used are: aaRS, aminoacyl-tRNA synthetase; aa, amino acid(s); AspRS, aspartyl-tRNA synthetase; AMP, adenosine monophosphate; ATP, adenosine triphosphate; MD, molecular dynamics; MDFE, molecular dynamics free energy; PBFE, Poisson-Boltzmann free energy; PME, particle mesh Ewald; CRF, continuum reaction field; r.m.s., root mean square.

Asp backbone recognition). The active site is highly preorganized to receive the Asp and ATP substrates, with the apoenzyme largely superimposable on the various complex structures (22). Each substrate, taken separately, is almost exactly superimposable on the corresponding moiety in the known AspAMP or tRNA·Asp complexes. In all of class II, ATP binds in a very unusual, completely bent conformation, with three associated Mg^{2+} cations (or two in a few cases; see below). The principal cation coordinates both the reactive α-phosphate and the β-phosphate and is positioned by conserved side chains (24). The other two cations coordinate the β- and γ-phosphates on either side of the ATP.

In AspRS, the adenylation reaction can occur in the absence of tRNA. Once Asp and ATP are in place, the Asp backbone reacts with the α-phosphate through an in-line mechanism and a pentacoordinate transition state. Inversion of the α-phosphate is clearly seen when crystals of AspRS·ATP and AspRS·AspAMP are compared (19). As with most enzyme reactions, the exact role of each surrounding group is hard to establish. It is presumed that the principal Mg^{2+} cation helps activate the α-phosphate by withdrawing electrons and pulling its oxygens into a pentacoordinate geometry, helping to stabilize the transition state through electrostatic interactions. The other two cations neutralize the leaving pyrophosphate product and may also contribute to transition state stabilization.

We have focused here on the role of Mg^{2+} in Asp/Asn discrimination by AspRS from *E. coli*. AspRS specificity is a complex problem. Although the enzyme is preorganized, Asp binding does induce structural reorganization in two important regions: (i) a so-called histidine loop (residues 436–449 in *E. coli*) shifts and becomes more ordered, with His-448 making a hydrogen bond to Asp; (ii) a flexible "flipping loop" (residues 167–173 in *E. coli*) closes over the aa binding site, bringing the negative Glu-171 close to the Asp ligand. The flipping loop is conserved in eukaryal, eubacterial, and archaebacterial AspRS (17, 18, 23), and a corresponding mobile loop is found in AsnRS (27) and LysRS (28). The exact populations of the open and closed loop states, with and without bound Asp, are unknown. Another difficulty is that the substrates Asp and ATP are both charged. Therefore, both short and long range electrostatic interactions are expected to play a role in both binding and specificity. Amino acid binding might couple to proton binding or release by His-448 or His-449 and to Mg^{2+} binding or release by ATP.

AspRS specificity has been analyzed with a powerful combination of crystallography, site-directed mutagenesis, kinetic and thermodynamic experiments, and phylogenetic analyses. These methods have limitations, however. Conserved residues may contribute to binding, binding specificity, catalysis, or all three. Experimental assays based on catalytic activity (29, 30) usually become infeasible once a single essential residue is mutated. The strength of electrostatic interactions is very difficult to infer from crystal structures, because the complex dielectric environment within a solvated protein causes large deviations from a simple Coulomb's law (31). Crystallography does not reveal the ionization states of acidic and basic residues, and pK_a measurements are difficult for AspRS, which is a homodimer of 1180 residues. Most importantly, weakly populated states are difficult to study experimentally. One extensive mutation study explored changes in the apparent Asp and ATP dissociation constants spanning only 2 orders of magnitude, corresponding to a free energy span of less than 3 kcal/mol (26). Non-cognate complexes like AspRS·Asn could not be analyzed. Weakly populated states are invisible in a crystal structure. Thus, from the AspRS·ATP crystal structures, ATP binds preferentially with three associated Mg^{2+} cations (17). However, the crystallographic data do not reveal the exact occupancies of the three magnesium sites. A complex with ATP and only two cations, present in the crystal with a population of 10 or 20%, for example, would not have an appreciable effect on the electron density maps and would be impossible to infer from the crystallographic data.

Theoretical methods represent a valuable complementary tool that can resolve these difficulties (32–37). The specificity of Asp binding to AspRS is governed by the binding free energy difference between the cognate Asp and competitor ligands. This difference can be obtained from molecular dynamics free energy simulations (MDFE), which have matured enormously in recent years and have been used to study several aaRSs. Extensive studies (33–35) show that when the AspRS binding pocket is in the "open" state (open flipping loop), there is an enormous preference for Asp over Asn, thanks to a network of electrostatic interactions in the active site. A thermodynamic cycle (see below) was used to obtain binding free energy differences, and a group decomposition of the free energy was used to identity the residues determining amino acid binding specificity.

When the flipping loop closes (17, 23), the negative Glu-171 is brought close to the Asp ligand site. We recently predicted computationally (36) that this conformational change induces proton binding by the nearby His-448. The His-448 positive charge then accounts for most of the large, computed Asp/Asn discrimination. In another long range electrostatic effect, a substrate-assisted specificity was observed: co-binding of ATP increases the Asp/Asn discrimination further. In eukaryotic AspRSs, His-448 is absent, being replaced by an Arg that is more distant from the ligand site. In these organisms, the role of ATP as a mobile discriminator is therefore important, protecting against Asn binding. In the same study, the computational model was tested and validated by experimental measurements of Asp-stimulated pyrophosphate exchange and its inhibition by Asn (36).

The present article focuses on the precise stability of the ATP-associated Mg^{2+} cations in the AspRS structure and their role in the thermodynamics of Asp and Asn binding. We consider AspRS from both *E. coli* and the archaebacterium *Pyrococcus kodakaraensis*. We report free energy simulations that compare Asp and Asn binding to AspRS in the presence of either bound ATP·Mg_2^{2+} or bound ATP·Mg_3^{2+}. The level of Asp/Asn discrimination in each case was computed using two distinct, largely independent methods for the free energy changes. The first method, MDFE, alchemically transforms Asp into Asn during a series of molecular simulations with an explicit solvent representation (33). The second, Poisson-Boltzmann free energies (PBFE), models the ligand binding reactions using a continuum dielectric model of both protein

and solvent (35). MDFE was then used to compute the relative affinities of AspRS for ATP·Mg_2^{2+} and ATP·Mg_3^{2+} and to demonstrate that the ATP·Mg_2^{2+} complex has a negligible occupancy and does not play any role in the specificity. In the closely homologous *E. coli* LysRS, the ATP 3-cation complex is again strongly favored. Binding of the positively charged Lys substrate, however, is not affected by cation binding.

We also considered the post-adenylation, AspRS·tRNA·AspAMP complex. Our simulations predict a strongly bound Mg^{2+} cation that aids AspAMP recognition. We show that specificity is maintained in the post-adenylation state (11), with AspAMP binding more strongly than AsnAMP in the presence of one Mg^{2+} cation and $tRNA^{Asp}$. The Mg^{2+} cation boosts AspAMP binding specificity, in a long range electrostatic effect similar to that of ATP·Mg_3^{2+} in the preadenylation complex. Thus, the introduction of negative charge into AspRS, by conformational change (36) or tRNA binding, is compensated by histidine protonation (36) and/or cation binding to preserve Asp recognition.

Finally, in the supplemental material we have reported a survey of 238 x-ray structures of protein complexes with ATP or GTP taken from the Protein Data Bank, which sheds additional light on the role of the Mg^{2+} cations and supports the predicted 3-cation AspRS state. Indeed, we find that ATP binding in a completely bent conformation with three associated Mg^{2+} cations is a characteristic property of class II aaRSs.

EXPERIMENTAL PROCEDURES

Molecular Dynamics Simulations

Starting structures for AspRS with bound Asp and ATP were generated from a 2.6-Å resolution crystal structure of *E. coli* AspRS with bound aspartyl-adenylate, AspAMP (Protein Data Bank entry 1IL2) (23). The two tRNA ligands were removed. We considered protein residues within a 24-Å sphere centered on the γ-carbon of the adenylate ligand of one monomer of the 1IL2 dimer. We overlaid either Asp or Asn on the adenylate molecule and deleted the original ligand. ATP and its associated cations were positioned by taking ATP·Mg_3^{2+} from the 1.9-Å resolution *P. kodakaraensis* AspRS·ATP complex (Protein Data Bank entry 1B8A) (17) and building it into the 1IL2 structure so as to overlap with the AMP moiety of the original AspAMP ligand. Hydrogens were constructed with ideal stereochemistry. Protonation states of histidines were assigned by visual inspection, except for His-448 and His-449 in the active site, which were assigned earlier through extensive simulations (36). Orientations of His, Asn, and Gln side chains in the active site were taken from the crystal structure and verified by inspection (33). In addition to crystal waters, a 73-Å cubic box of water was overlaid, and waters overlapping the protein were removed. The final model contained the amino acid and ATP ligands, 357 protein residues, and around 10,000 waters, including 113 crystal waters. Periodic boundary conditions were assumed; *i.e.* the entire 73-Å box was replicated periodically in all directions.

We also generated AspRS·Asp·ATP complexes from the *P. kodakaraensis* AspRS·ATP complex (17). We built in Asp and Asn ligands by alignment with the *E. coli* structure (Protein Data Bank entry 1IL2; see above) and generated solvated 24-Å spheres centered on the ligand. We then used pK_a calculations (see below) to determine the protonation state of His-223, a histidine residue oriented into the ATP binding pocket. LysRS structures were generated from a 2.1-Å resolution *E. coli* LysRS·Lys·ATP·Mn_3^{2+} crystal structure (28). We took a 24-Å sphere centered on the γ-carbon of the Lys ligand, replaced the Mn^{2+} cations with Mg^{2+}, and then solvated the system. Histidine protonation states were assigned by visual inspection except for His-270, which points into the LysRS ATP-binding pocket. In both AspRS and LysRS, the protonation state of the histidine oriented into the ATP pocket is coupled to cation binding.

Structures for AspRS with bound AspAMP and tRNA were generated from the 2.4-Å resolution *E. coli* crystal structure 1C0A (18). Again, we considered a 24-Å sphere centered on the ligand γ-carbon and solvated the system. pK_a calculations were used to determine the predominant state of His-448, a histidine residue close to AspAMP. The cation associated with the AspAMP ligand phosphate group (15) was placed by structural alignment with the principal cation in the Asp·ATP·Mg_3^{2+} state (17). The cation occupies the space assigned to water molecule 1073 in the x-ray structure (18) and remains in the same octahedral binding mode throughout the simulations, coordinating the Glu-482 and Asp-475 carboxylates and two or three water molecules and remaining 4–5 Å away from the AspAMP phosphate group.

For both AspRS and LysRS, if the entire aaRS protein were included in the model, rather than a spherical subset, a far greater number of solvating waters would be needed. To reduce artifacts due to the protein truncation, protein groups between 20 and 24 Å from the center were harmonically restrained to their positions in the crystal structure. In this way, protein regions beyond 24 Å are accounted for structurally. They will also be accounted for thermodynamically, in a separate step (see below), where the free energy to reintroduce the missing protein groups is computed from a continuum electrostatic model. This hybrid, atomic/continuum approach is accurate, because earlier work on this system showed that both structures and free energies obtained with spherical subsets of 20-, 24-, or 28-Å radii were all similar (37).

All long range electrostatic interactions were computed efficiently by the particle mesh Ewald (PME) method. Four sodium counterions were included to reduce the formal charge of the system. One nanosecond of unrestrained molecular dynamics was performed (for each complex) at constant room temperature and pressure with a Nosé-Hoover algorithm following 200 ps of thermalization. The complexes modeled are *E. coli* (23) AspRS·Asp·ATP·Mg_3^{2+}, AspRS·Asp·ATP·Mg_2^{2+}, AspRS·Asn·ATP·Mg_3^{2+}, and AspRS·AsnATP·Mg_2^{2+}. 500-ps trajectories were also produced for each amino acid ligand in solution, Asp or Asn, solvated at the center of a box of water molecules, with the ligand γ-carbon weakly restrained at the origin throughout the dynamics.

Additional simulations were performed with a less expensive, spherical, continuum reaction field (CRF) (37, 38) model. It included the same protein residues as mentioned above,

along with the water molecules inside the 24-Å sphere (about 560 waters). Water and protein outside the 24-Å sphere were treated as a single, homogeneous, dielectric medium with a dielectric constant of 80 (38). Electrostatic interactions between atoms within the sphere were computed without any cutoff, using an efficient multipole approximation for distant groups (39). A multipolar expansion with 20 terms was used to approximate the reaction field due to the surrounding continuum (37, 38). Newtonian dynamics were used for the inner 20 Å of the sphere and Langevin dynamics for the outer region (20–24 Å), with a bath temperature of 293 K. The same four complexes as above were simulated for 3 ns each, as well as the following 10 complexes (3 ns each): AspRS·Asp (or Asn)·ATP·Mg_3^{2+} (or Mg_2^{2+}) complexes from *P. kodakaraensis* (17); and *E. coli* (28, 18) LysRS·Lys·ATP·Mg_3^{2+} (or Mg_2^{2+}), AspRS·AspAMP (or AsnAMP)·Mg^{2+}·tRNA, and AspRS·AspAMP (or AsnAMP)·tRNA complexes. The CHARMM22 force field (40) was used for the protein, ligands, and counterions. A slightly modified TIP3P model was used for the water (41). We used the CHARMM program (42), version c30b1, for all calculations.

Free Energy Calculations

The methods used to compute ligand binding free energy differences in AspRS have been described previously (33, 36, 43, 44). For Asp/Asn binding, we use the thermodynamic cycle shown in Fig. 1. We use the horizontal legs for simplified PBFE calculations and the vertical legs for more rigorous, alchemical, MDFE simulations. In the latter, the ligand Asp is reversibly transformed into Asn during a series of simulations; the corresponding work is derived from a thermodynamic integration formula (43). Similarly, to test the stability of the ATP·3-cation complex, the most weakly bound cation (see below) was reversibly transformed into a water molecule (as shown in Fig. 2). Related simulations, using the Poisson-Boltzmann linear response approximation (PB/LRA) method, were used to identify the pK_a, and hence the predominant state, of protein histidine residues important for ligand binding. See two recent studies (36, 45) for details of the method.

PBFE Computations—The electrostatic contribution to the ligand binding free energy was obtained by subtracting the electrostatic free energy in the ligand·protein complex and in the separate ligand and protein (35). Non-electrostatic contributions to the free energy were assumed to cancel. This should be a good approximation for the Asp/Asn differences, because the two ligands have the same size and bind in the same position (35). The electrostatic potential was obtained by numerically solving the Poisson-Boltzmann equation using a cubic grid and a finite difference algorithm, implemented in CHARMM (grid size, 144 Å; grid spacing, 0.4 Å). The protein·solvent boundary was defined by the molecular surface as in an earlier work (35). PBFE was performed at zero ionic strength. In an earlier study of AspRS (35), the Asp/Asn discrimination computed with zero and physiological ionic strength differed by just 0.3 kcal/mol. The structure of the protein·ligand complex was taken from the MD simulations (above). Given the longer sampling times possible with CRF and the close similarity between the CRF and PME trajectories in terms of binding pocket geometry and flexibility, we generally used structures from CRF simulations. Calculations were performed for multiple structures, sampled every 4 ps along the equilibrated 3 ns trajectory, for a total of 750 structures. The separate ligand and protein structures were obtained by simply discarding the unwanted partner. Thus, structural relaxation on separating protein and ligand was not explicitly included (though it is implicit in the dielectric constant). The solvent dielectric constant was set to 80. The solute dielectric constant was set to 4, based on extensive earlier comparisons between MDFE and PBFE calculations for AspRS (35). Proteins groups outside the 24-Å sphere (above) were neglected. This was a convenient and harmless approximation, because they have been shown to contribute less than 1 kcal/mol to the Asp/Asn binding free energy difference (37). We also used PBFE to estimate the binding strengths of each Mg^{2+} cation in AspRS·ATP and LysRS·ATP, Lys binding to LysRS, AspAMP/AsnAMP binding to AspRS·tRNA, and cation binding to AspRS·aaAMP·tRNA, with the same setup.

MDFE Simulations—Asp/Asn mutation runs were performed with the PME simulation setup. The ligand side-chain geometry, atom types, and charges were reversibly changed from the Asp values to the Asn values (33–36) over a series of ten 100-ps simulations or "windows."

Having identified the most weakly bound cation from PBFE calculations (see above), we also performed MDFE simulations to determine the stability of the ATP·3-cation complex in AspRS and LysRS, reversibly changing the third cation into a water molecule over a series of MD windows. This procedure was analogous to that used for the Asp/Asn mutations. The energy function can be expressed as a linear mixture of Mg^{2+} and water terms,

$$U(\lambda) = U_0 + (1 - \lambda)U_{\text{water}} + \lambda U_{Mg^{2+}} \qquad \text{(Eq. 1)}$$

where λ is a weight or "coupling parameter" and U_0 represents interactions between parts of the system other than the hybrid ligand. We gradually mutated Mg^{2+} into water by changing λ from 1 to zero. The successive weights were λ = 0.99, 0.95, 0.9, 0.8, 0.6, 0.4, 0.2, 0.1, 0.05, and 0.01 for Mg^{2+} and 1-λ for water. The derivative of the free energy, *G*, can be written,

$$dG/d\lambda = \langle U_{Mg^{2+}} - U_{\text{water}} \rangle_\lambda \qquad \text{(Eq. 2)}$$

where the brackets represent a time average over an MD trajectory performed with the energy function $U(\lambda)$. The free energy derivatives were computed at each λ value from a 100-ps MD simulation. Each run thus corresponds to 1.0 ns in total.

Equation 2 can be used to perform a group decomposition of the free energies (43, 46). Indeed, the term $U_{Mg^{2+}}$ is a sum over interactions between the cation and surrounding amino acids or waters, as is similarly true for U_{water}. Thus, the free energy derivative and, ultimately, the free energy can be viewed as a sum of group contributions. Such decompositions have proven useful for identifying the sources of binding affinity (44).

Correction for Distant Parts of the Protein—In the MD simulations above, protein groups outside of the 24-Å spherical region were only taken into account structurally through the application of harmonic restraints to protein groups near the 24-Å boundary. Their direct contribution to the free energy can be computed in a second step, where they are reversibly intro-

duced back into the system. The free energy for this step can be obtained using a continuum model (as detailed in Refs 37 and 47). A continuum model is appropriate because the groups are more than 24 Å away from the ligands. Refs. 37 and 47 show that the net contribution of this second step to the Asp/Asn binding free energy difference is less than 1 kcal/mol; this is less than the overall uncertainty of both MDFE and PBFE. Therefore, in this work, we did not actually compute the long range correction. We simply approximated it by zero and included its effect in the overall uncertainty estimate. Note that in this work, continuum electrostatics are used in three distinct ways, which should not be confused with each other: 1) in this second, free energy step; 2) for PBFE calculations (see above); and 3) in the MDFE simulations with CRF boundary conditions.

RESULTS

We will first summarize the most important structural and dynamic features of the amino acid, ATP, and tRNA binding sites in AspRS. Next, we will describe our calculations to compare Asp and Asn binding to AspRS in the presence of either ATP·Mg_3^{2+} or ATP·Mg_2^{2+} and calculations to compare ATP·Mg_3^{2+} and ATP·Mg_2^{2+} binding to both AspRS and LysRS. Finally, we report the calculations to compare AspAMP and AsnAMP binding to AspRS·tRNA in the presence of one or no co-bound Mg^{2+} cations.

Structure, Dynamics, and Solvation of the Amino Acid, ATP, and tRNA Binding Sites in AspRS

The first set of simulations were done for both Asp and Asn bound to AspRS, co-bound with ATP and either three or two cations. The CRF and PME methods yield very similar binding pocket structures and dynamics. For the AspRS·Asp·ATP·Mg_3^{2+} complex, the r.m.s. deviations from the starting crystal geometry were 0.7 and 1.4 Å for backbone and side-chain atoms, respectively. Fig. 3 illustrates the active site dynamics in a typical series of snapshots from the MD trajectory. Fig. 3 was prepared using Molscript (48) and rendered using Raster3D (49). The cognate Asp ligand makes a stable network of hydrogen bonds to several binding pocket residues, which agree very well with the available crystal structures (17, 18, 23) and previous MD simulations (33–35). From earlier pK_a calculations (45), His-448 is doubly protonated (36) and forms a salt bridge to the Asp side-chain carboxylate. The Asp side chain and backbone amino group are also stabilized by H-bonds to Arg-489, Lys-198, a buried water molecule W1, Glu-171 of the closed flipping loop, and Gln-195. The terminal carboxylate forms H-bonds with Arg-217 and, 70% of the time, with Gln-231.

ATP retained its fully bent geometry, characteristic of ATP binding to class II aaRSs (50), throughout the simulations. The active site dynamics shown in Fig. 3 are from the *E. coli* structure (23), with ATP·Mg_3^{2+} built in from the *P. kodakaraensis* structure (17) as described above under "Experimental Procedures." Here we focus on the dynamics of the ATP binding site in the *P. kodakaraensis* AspRS·ATP complex. The principal Mg^{2+} cation nearest the amino acid (labeled M1 in Table 4) is stabilized by H-bonds to the ATP α- and β-phosphates and also to Asp-475 and Glu-482. Two water molecules, or occasionally three depending on the strength of the M1/β-phosphate interaction, complete the M1 coordination sphere. One of these is a highly ordered water, always bridging M1 and Glu-482 in an interaction secondary to the direct M1-Glu-482 stabilization. Another cation, M2, is positioned between the ATP β-,γ-phosphates and Glu-482, with three waters also coordinated. M2 and Glu-482 also have a secondary water-mediated interaction, although here the water is only present ~50% of the time, with exchange in water molecules every few hundred ps. The ionized residues Asp-475 and Glu-482 are highly conserved in class II aaRSs; mutagenesis experiments indicate that both are functionally irreplaceable in AspRS (24).

The third cation, M3, also bridges the ATP β-,γ-phosphates but does not bind to protein. M3 remains on the more solvent-exposed side of the binding pocket, completing its octahedral coordination sphere with four water molecules. The coordinating waters have an average r.m.s. fluctuation of just 0.4 Å. The r.m.s. fluctuations of each of the cations are: M1 = 0.24 Å; M2 = 0.26 Å; M3 = 0.32 Å. Thus, M3 is predicted to have a crystallographic B-factor that is larger than M1 and M2, in agreement with the AspRS·ATP crystal structure (17). M3 has ~55 water molecules within a 9-Å sphere, compared with 40 for the more strongly bound cations, 25 for the amino acid ligand backbone ammonium group, and 100 for bulk water. Finally, the ATP γ-phosphate oxygen not pointing directly to either M2 or M3 has water-mediated interactions with M2 and M3 through the water molecules completing the cation coordination spheres. For M2, there is an exchange in water molecules on the sub-100-ps scale, so that only the water associated with M1 and Glu-482 above and the four waters coordinated to M3 are truly ordered.

When Asn replaces Asp in the amino acid binding site, earlier free energy simulations showed that His-448 loses its labile proton and becomes neutral (36). In other respects, the structure is very similar to the Asp complex. Lifetimes of H-bonds stabilizing the Asn backbone groups are reduced by around 10% compared with the Asp complex, whereas the Asn side-chain carbonyl binds strongly to Arg-489 but not to Lys-198. Overall, the mean number of H-bonds between the amino acid and AspRS is reduced from 10 (Asp complex) to 7 (Asn complex).

When the third cation in the ATP 3-cation complex (M3) is removed, neither ATP nor its binding pocket undergoes significant displacements. The r.m.s. deviations of ATP and its immediate environment from the starting geometry are 1.7 and 1.1 Å, respectively, similar to those seen for the 3-cation complex above (1.4 and 1.2 Å, respectively). Also, the terminal oxygens of the aa ligand remain within 3.5–4.9 Å of the ATP α-phosphorus in both the 2-cation and 3-cation systems, similar to what was observed in HisRS simulations (51). Overall, the MD structures of the 2- and 3-cation complexes have a comparable agreement with the available x-ray structures. This confirms that in the AspRS·ATP crystal structure, if a 2-cation complex were present in 10 or 20% of the unit cells, it would not have a noticeable effect on the observed electron density and could not be detected. Therefore, to determine the probability of a 2-cation complex, structural data is not enough; free energies must be computed. Finally, we note that the pK_a calculations (45) reported in Table 1 indicate that cation binding is

TABLE 1

Computed protonation states for the histidine next to the third Mg^{2+} cation in the AspRS·Asp·ATP, AspRS·Asn·ATP, and LysRS·Lys·ATP complexes: His-223 in *P. kodakaraensis* AspRS and His-270 in *E. coli* LysRS

The final two entries are for *E. coli* AspRS·aaAMP·Mg^{2+}·$tRNA^{Asp}$, and correspond to His-448, close to the adenylate ligand and the Mg^{2+} cation.

Complex	ΔG_{sol}	ΔG_{prot}	$\Delta\Delta G^a$	pK_a shift	His state
AspRS·Asp·ATP·Mg_3^{2+}	−10.6	−6.0	+4.6 (2.4)	+3.4	Neutral
AspRS·Asp·ATP·Mg_2^{2+}	−10.6	−20.4	−9.8 (2.4)	−7.2	Charged
AspRS·Asn·ATP·Mg_3^{2+}	−10.6	−4.1	+6.3 (2.2)	+4.6	Neutral
AspRS·Asn·ATP·Mg_2^{2+}	−10.6	−20.9	−10.3 (2.4)	−7.5	Charged
LysRS·Lys·ATP·Mg_3^{2+}	−10.6	+4.9	+15.5 (1.9)	+11.4	Neutral
LysRS·Lys·ATP·Mg_2^{2+}	−10.6	−11.7	−1.1 (2.2)	−0.8	Charged/neutral
AspRS·AspAMP·Mg^{2+}·$tRNA^{Asp}$	−10.6	−19.1	−8.5 (1.6)	−6.2	Charged
AspRS·AsnAMP·Mg^{2+}·$tRNA^{Asp}$	−10.6	−7.5	+3.1 (1.3)	+2.3	Neutral/charged

[a] $\Delta\Delta G = \Delta G_{prot} - \Delta G_{sol}$ (in kcal/mol; see Refs. 36 and 45 for more details) (sol, solvent; prot, protein). Uncertainty, shown in parentheses, is estimated from the deviation in computed ΔG_{prot} values, sampling every 4 ps along 0.5-ns segments.

coupled to nearby histidine charging. The ATP binding pockets of both AspRS and LysRS feature a histidine residue pointing toward the ATP γ-phosphate (17, 28). The structure and dynamics of the LysRS substrate binding pocket will be described in detail elsewhere. This histidine is neutral in the presence of ATP·Mg_3^{2+} but can gain a positive charge when ATP·Mg_3^{2+} is artificially changed to ATP·Mg_2^{2+} (Table 1). As shown in the energetic analysis below, this histidine charging does not fully compensate for the cost of unbinding the third cation, so that the state with three cations is strongly favored and always present.

Following our extensive analysis of AspRS substrate specificity in the preadenylation step (present work and Refs. 33–36), the last system we consider here involves AspAMP/AsnAMP specificity in the presence of co-bound tRNA in the post-adenylation step. Opening of the "flipping loop" accompanies tRNA binding, as shown in the x-ray structure (18). Apart from the loss of the Glu-171 interaction, the structure and dynamics of the AspAMP binding site are similar to those in the Asp·ATP·AspRS systems described above, in agreement with the x-ray data (17). The r.m.s. deviations of heavy atoms from the starting geometry in AspAMP and its immediate environment are 0.7 and 1.0 Å, respectively. AsnAMP is slightly more mobile than AspAMP in the binding pocket. The r.m.s. deviations of AsnAMP and its immediate environment from the starting geometry are both 1.1 Å. As shown in Table 1, His-448 loses its extra proton and becomes neutral in the presence of AsnAMP and a co-bound Mg^{2+} cation. This does not lead to significant changes in ligand or binding site structure; His-448 simply swings away from AsnAMP toward the solvent-exposed side of the AspRS pocket. The Mg^{2+} cation associated with the aaAMP phosphate maintains a strong coordination with Glu-482 and Asp-475 throughout the dynamics and has an r.m.s. fluctuation of only 0.30 Å, similar to the cations coordinating ATP in the preadenylation step (above). The co-bound $tRNA^{Asp}$ molecule does not undergo significant deviations from its starting geometry. The r.m.s. deviation of the tRNA from its x-ray position is 1.4 Å, with either AspAMP or AsnAMP co-bound. Finally, removing the Mg^{2+} cation associated with aaAMP causes only a slight increase in structural disorder. The r.m.s. deviations for AspAMP, tRNA, and their immediate environments increase by 0.1–0.2 Å when the cation is artificially removed.

TABLE 2

MDFE and PBFE binding free energy differences $\Delta\Delta G$ (kcal/mol) for Asp *versus* Asn in AspRS, alone and with co-bound ATP·Mg_3^{2+} and ATP·Mg_2^{2+}

PBFE $\Delta\Delta G$ values are also given for AspAMP *versus* AsnAMP binding in AspRS, with $tRNA^{Asp}$ and either one or zero co-bound cations. MDFE $\Delta\Delta G$ is computed from the vertical, alchemical legs in Fig. 1: $\Delta\Delta G = \Delta G_1 - \Delta G_4$. A positive sign corresponds to preferential Asp binding. ΔG_1 and ΔG_4 are computed by alchemically transforming Asp into Asn (or the reverse) during a series of simulations. The MDFE $\Delta\Delta G$ values are averages over four 1.0-ns runs; PBFE $\Delta\Delta G$ values are computed from 3 ns native MD trajectories, using the horizontal, binding legs in Fig. 1. $AspRS^+$ indicates a doubly protonated His-448.

Medium	MDFE $\Delta\Delta G$ (Asp → Asn)[a]	PBFE $\Delta\Delta G$ (Asp → Asn)[b]
AspRS	+3 (2)	
$AspRS^+$	+11 (2)	+10 (2)
AspRS·ATP·Mg_3^{2+}	+9 (3)	
$AspRS^+$·ATP·Mg_3^{2+}	+19 (3)	+18 (2)
AspRS·ATP·Mg_2^{2+}	+5 (1)	+5 (2)
$AspRS^+$·aaAMP·Mg^{2+}·$tRNA^{Asp}$		+9 (2)
$AspRS^+$·aaAMP·$tRNA^{Asp}$		+4 (3)
AspRS·aaAMP·Mg^{2+}·$tRNA^{Asp}$		+4 (2)

[a] MDFE values are from Ref. 36. Uncertainty (shown in parentheses) in MDFE data is obtained by first averaging each pair of forward/backward runs and then taking twice the deviation among pairs. The estimated uncertainty is much smaller than the difference (or "hysteresis") between forward and backward runs (see text).

[b] Uncertainty in PBFE data is estimated from the deviation in computed PBFE $\Delta\Delta G$ values, sampling every 4 ps along 1.0–3.0-ns segments.

Asp/Asn Discrimination by AspRS in the Presence of ATP and Either Three or Two Mg^{2+} Cations

We first compared the binding free energies of Asp and Asn in the presence of ATP with three associated cations. We compared MDFE results (some of them obtained earlier (36)) with the PBFE results obtained here. The data are summarized in Table 2. Each MDFE $\Delta\Delta G$ value was computed from two solution simulations and four protein simulations (of ten 100-ps windows each; see "Experimental Procedures"). In solution, one run was performed in each direction, mutating Asp to Asn or the reverse, yielding ΔG_1 (see Fig. 1). In the protein, two runs were performed in each direction, yielding ΔG_4 (Fig. 1). When ATP·Mg_3^{2+} is co-bound, $\Delta\Delta G = +19$ kcal/mol, favoring Asp. This decreases to +9 kcal/mol when His-448 is artificially maintained in the singly protonated state (36). Notice that His-448 is absent from most eukaryotic AspRSs, being replaced by an Arg that is more distant from the ligand site. PBFE free energy values are obtained from the vertical legs of the thermodynamic cycle in Fig. 1. In the three most important states for which both MDFE and PBFE simulations were performed, the PBFE estimates match the MDFE values to within ±2 kcal/mol.

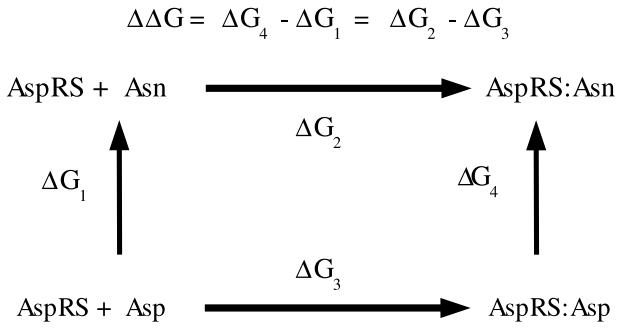

FIGURE 1. **Thermodynamic cycle used for computation of the amino acid binding free energy differences.** ΔG_1 and ΔG_4 refer to free energy changes for ligand mutation in solution and enzyme, respectively, and ΔG_2 and ΔG_3 are free energy changes for Asn and Asp binding to enzyme.

TABLE 3

Electrostatic (PBFE) contribution to the binding free energies for AspRS·Asp, LysRS·Lys, and AspRS·AspAMP complexation as a function of the number of bound Mg^{2+} cations

$LysRS^+$ denotes LysRS with His-270 positively charged; $AspRS^+$ is AspRS with His-448 positively charged. ΔG values have standard deviations of 1–3 kcal/mol and were computed from at least 250 MD snapshots sampled every 4 ps over multi-nanosecond MD trajectories.

Complex	Amino acid ligand binding ΔG
	kcal/mol
AspRS·Asp·ATP·Mg_3^{2+}	−38
AspRS·Asp·ATP·Mg_2^{2+}	−31
LysRS·Lys·ATP·Mg_3^{2+}	−61
$LysRS^+$·Lys·ATP·Mg_3^{2+}	−61
LysRS·Lys·ATP·Mg_2^{2+}	−60
$LysRS^+$·Lys·ATP·Mg_2^{2+}	−59
$AspRS^+$·AspAMP·$tRNA^{Asp}$	−35
$AspRS^+$·AspAMP·Mg^{2+}·$tRNA^{Asp}$	−48

This is comparable with the uncertainty in the MDFE values themselves.

When ATP is associated with only two cations, the Asp/Asn binding free energy difference decreases substantially, to $\Delta\Delta G = +5$ kcal/mol, compared with 9 kcal/mol with ATP·Mg_3^{2+} (Table 2). Agreement between MDFE and PBFE is good. Thus, even though the third Mg^{2+} cation is 11 Å away from the γ-carbon of the ligand side chain, it interacts strongly with the amino acid ligand and strongly favors the negative Asp over neutral Asn. Consistent with this finding, PBFE estimates of amino acid binding strengths in Table 3 show that removing the third cation strongly reduces Asp binding to AspRS. The results for LysRS are also shown in Table 3. We see that Lys binding to LysRS is not sensitive to the number of Mg^{2+} cations present, despite its positive charge and in contrast to the large effect of the cations in AspRS. This is shown by the near equality of the Lys binding free energies for the LysRS·Lys·ATP·3-cation complex (with either neutral or charged His-270) and for the 2-cation complex (with either neutral or charged His-270). In the next section, we show that in both AspRS and LysRS, ATP always binds with all three cations.

Stability of the ATP·Mg_3^{2+} Complex in AspRS and LysRS

Identifying the Most Weakly Bound Cation—X-ray crystal structures of ATP bound to AspRS and LysRS show that both AspRS and LysRS are among the class II synthetases that preferentially bind three divalent cations (17, 28). PBFE calculations reported in Table 4 were used to identify the most weakly bound cation in each enzyme. A solute dielectric constant of 4 was used, with structures sampled from the Asp·AspRS, Asn·AspRS, and Lys·LysRS MD trajectories.

TABLE 4

Electrostatic (PBFE) contribution to the binding free energies of each cation in AspRS·ATP·Mg_3^{2+} and LysRS·ATP·Mg_3^{2+} complexes and of the single remaining cation in AspRS·aaAMP·Mg^{2+}·$tRNA^{Asp}$

$LysRS^+$ denotes LysRS with the near-M3 His-270 positively charged, and $AspRS^+$ denotes AspRS with the near-M1 His-448 positively charged. ΔG values have standard deviations of 1–3 kcal/mol and were computed from at least 250 MD snapshots sampled every 4 ps over multi-nanosecond MD trajectories.

Complex	Cation binding ΔG		
	M1	M2	M3
	kcal/mol		
AspRS·Asp·ATP·Mg_3^{2+}	−104	−144	−96
AspRS·Asn·ATP·Mg_3^{2+}	−85	−135	−81
AspRS·Asp·ATP·Mg_3^{2+}[a]	−110	−138	−103
AspRS·Asn·ATP·Mg_3^{2+}[a]	−109	−137	−104
LysRS·Lys·ATP·Mg_3^{2+}	−132	−133	−88
$LysRS^+$·Lys·ATP·Mg_3^{2+}	−124	−130	−73
$AspRS^+$·AspAMP·Mg^{2+}·$tRNA^{Asp}$	−117		
$AspRS^+$·AsnAMP·Mg^{2+}·$tRNA^{Asp}$	−110		

[a] The starting structure is the *P. kodakaraensis* AspRS·ATP complex (see Ref. 17). Allother starting structures are from *E. coli* complexes.

From Table 4, it is clear that M3 is the least stable cation in all the complexes. In AspRS, the order of cation binding strengths is M2 > M1 > M3, irrespective of the amino acid ligand. M3 is on the solvent-exposed side of ATP. Its r.m.s. fluctuations are about 25% larger than those of M1 and M2. Of the three cations, M1, coordinating the ATP α,β-phosphates Asp-475 and Glu-482, is the most strongly stabilized by Asp rather than Asn binding. Its binding free energy is 19 kcal/mol lower with bound Asp than with bound Asn. M1 is the cation closest to the amino acid, and hence it is the most strongly affected by the net charge of the amino acid ligand. M2 is the most tightly bound cation, coordinating the β- and γ-phosphates and Glu-482. M2 is furthest from the amino acid binding site, and hence the least affected by the amino acid identity; its binding free energy is 9 kcal/mol lower with bound Asp than with bound Asn. M3, the most weakly bound cation, coordinates the β- and γ-phosphates but not protein. M3 is intermediate in distance from the amino acid pocket and intermediate in its energetics; its binding free energy is 15 kcal/mol lower with bound Asp than with bound Asn. This situation can be compared with the LysRS case. Table 4 shows clearly that in LysRS, M3 is again the most weakly bound cation. M1 and M2 have similar binding strengths; both are strongly stabilized by nearby LysRS glutamate residues.

Relative Stabilities of the ATP·Mg_3^{2+} and ATP·Mg_2^{2+} Complexes in Alchemical MDFE Simulations—ATP co-binding along with three divalent cations is supported but not proved by the AspRS·ATP and LysRS·ATP crystal structures (17, 28). Indeed, the crystallographic data cannot rule out a partial occupancy for the third cation. In other words, the third cation could be present in fewer than 100% of the crystal cells. From the data presented above, a partial occupancy, even as high as 80 or 90%, would have a significant effect on the Asp/Asn discrimination in AspRS. Therefore, we performed alchemical free energy simulations in which the third Mg^{2+} cation, M3, was reversibly transformed into a water molecule, both in the enzyme and alone in solution. This cation was chosen because it is the most

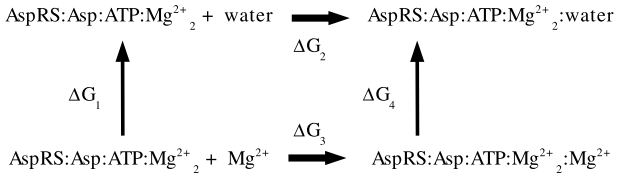

FIGURE 2. **Thermodynamic cycle for binding of the third cation in AspRS·Asp·ATP.** An analogous cycle was used for binding of the third cation in LysRS·Lys·ATP.

weakly bound of the three (Table 4). The data were analyzed with the thermodynamic cycle in Fig. 2. The double free energy difference $\Delta\Delta G$ represents the standard free energy for binding the third cation to the preformed AspRS·aa·ATP·Mg_2^{2+} complex.

Table 5 shows the computed free energy changes. We obtained large deviations between forward and backward runs in enzyme (*i.e.* where Mg^{2+} is mutated into water or the reverse; third column in Table 5). In fact, previous experience with charging free energy calculations in AspRS and other systems suggests that when forward and backward results are averaged, there is a significant compensation of errors. For example, when Asp is transformed into Asn in AspRS, there is a large forward/backward free energy hysteresis, but the forward/backward average changes by less than 4 kcal/mol when the run length is doubled (see supplemental material). Thus, estimated statistical uncertainty is obtained by first averaging pairs of forward/backward runs and then taking twice the standard deviation among these. For Mg^{2+} binding to $LysRS^+$, the standard error estimate (16 kcal/mol) is about half the computed $\Delta\Delta G$ (36 kcal/mol). Thus, even though the precision of the MDFE data is low, the predicted Mg^{2+} binding trends are very strong and can be taken as qualitatively accurate.

When the third cation binds, the nearby His protonation state can change, as shown in Table 1. To compare the most relevant states with 2 and 3 bound cations, the MDFE data must be combined with the computed protonation free energies (Table 1). Thus, when the third cation binds to $LysRS^+$, the nearby His-270 becomes deprotonated, and the free energy is lowered by an additional 16 kcal/mol (Table 1). The overall binding free energy of the third cation to LysRS is therefore estimated to be $-36 - 16 = -52$ kcal/mol, corresponding to the process Mg^{2+} + $LysRS^+$·Lys·ATP·Mg_2^{2+} → $LysRS^+$·Lys·ATP·Mg_3^{2+} → LysRS·Lys·ATP·Mg_3^{2+} + H^+. Notice that the estimated uncertainty in the computed protonation free energy is moderate (Table 1). Similarly, for AspRS, we consider the process Mg^{2+} + $AspRS^+$·Asp·ATP·Mg_2^{2+} → Mg^{2+} + AspRS·Asp·ATP·Mg_2^{2+} + H^+ → AspRS·Asp·ATP·Mg_3^{2+} + H^+. The total binding free energy is $+10 - 78 = -68$ kcal/mol. Finally, the four AspRS·aa·ATP·Mg_n^{2+} systems (aa = Asp or Asn; $n = 2$ or 3) form a thermodynamic cycle, allowing us to infer also the free energy for binding the third cation to the AspRS·Asn·ATP·Mg_2^{2+} complex. Given that binding of the third cation in AspRS·Asp·ATP is favored by 68 kcal/mol, the third cation in AspRS·Asn·ATP is favored by 59 kcal/mol.

The concentration of Mg^{2+} in the cell is on the order of 1 mM (52). The free energy changes reported above correspond to a standard state of 1 M Mg^{2+}. Thus, at physiological Mg^{2+} concentrations, the binding free energy is at most 3–5 kcal/mol lower than the standard state value. Overall, for both Asp and Asn, the third cation is much more stable bound to AspRS than alone in solution, and so ATP always binds to AspRS with all three associated Mg^{2+} cations; the Boltzmann probability of the 2-cation state is infinitesimal. The same qualitative result is found for LysRS; the 3-cation complex is favored by 52 kcal/mol in the standard state and by at least 47 kcal/mol at cellular concentrations. It is easy to show, using the data in Tables 1 and 5, that states in which the aaRS·aa·ATP complex binds only one or no Mg^{2+} cations are negligibly populated compared with even the aaRS·aa·ATP·Mg_2^{2+} states. Note that Mg^{2+} competition with monovalent ions is also insignificant, because the most abundant ion, potassium, has a concentration of about 150 mM. The standard binding free energy of K^+ can be estimated from electrostatic considerations to be weak, on the order of half that of the third Mg^{2+} cation. Thus, mixed complexes such as ATP·Mg_2^{2+}·K^+ are expected to be negligibly populated in both AspRS and LysRS.

TABLE 5

Stability of the third Mg^{2+} cation in AspRS and LysRS ATP·3-cation complexes using MDFE

Energies are in kcal/mol. ΔG corresponds to the mutation of Mg^{2+} into a water molecule. $LysRS^+$ denotes LysRS with the near-M3 His-270 positively charged. "Forward" runs transform Mg^{2+} into a water; "backward" runs transform a water into Mg^{2+}. ΔG in protein is computed for each system from the four ΔG values given, corresponding to runs from independent starting structures. A positive $\Delta\Delta G$ corresponds to preferential Mg^{2+} binding to the enzyme. Uncertainty (shown in parentheses) is obtained by first averaging each pair of forward/backward runs and then taking twice the deviation among pairs.

Medium	Direction of alchemical mutation run	ΔG ($Mg^{2+} \rightarrow$ water)	$\Delta\Delta G$
Solution	Forward	+430	
Solution	Backward	+438	
AspRS·Asp·ATP·Mg_2^{2+}	Forward	+534/+528	
AspRS·Asp·ATP·Mg_2^{2+}	Backward	+482/+502	+78 (10)
LysRS·Lys·ATP·Mg_2^{2+}	Forward	+511/+517	
LysRS·Lys·ATP·Mg_2^{2+}	Backward	+452/+468	+53 (8)
$LysRS^+$·Lys·ATP·Mg_2^{2+}	Forward	+500/+480	
$LysRS^+$·Lys·ATP·Mg_2^{2+}	Backward	+465/+434	+36 (16)

To identify the most important interactions stabilizing the third cation, we performed a group decomposition of the computed binding free energy based on Equation 2 (46, 43). The resulting free energy components are given in Table 6. A very large, favorable contribution (+367 kcal/mol) comes from the ATP·Mg_2^{2+} moiety itself, even though it is electrically neutral. Another +126 kcal/mol come from the conserved Glu-482/Asp-475 side chains, even though these are 6/9 Å away and are primarily involved in coordinating the other two cations. The contribution of solvent is small at −73 kcal/mol (opposing binding). The solvent contribution can be interpreted as a dielectric shielding of the negative charges that favor Mg^{2+} binding. Other protein groups make much smaller contributions. For example, the net contribution of the nearby salt bridge pair, Glu-219/Arg-225, is just +11 kcal/mol, and nearby Glu-171 and Arg-217 contribute just +14 kcal/mol. The Asp ligand contributes +45 kcal/mol. Overall, M3 binding is favored largely by ATP·Mg_2^{2+}, the conserved Glu-482/Asp-475 pair, and the Asp ligand. On balance, these groups stabilize cation·protein binding compared with cation solvation in bulk water (Tables 5 and 6).

TABLE 6

Free energy components for the stability of the third Mg^{2+} cation in the AspRS ATP·3-cation complex using alchemical MDFE simulations

Energies are in kcal/mol. ΔG corresponds to the mutation of Mg^{2+} into a water molecule; total ΔG = +528 kcal/mol (Table 5), comprising the sum of the ATP·Mg_2^{2+}, water, Asp ligand, and total protein contributions, together with +15 kcal/mol for the Poisson-Boltzmann continuum correction (see Ref. 33 for details) and +2 kcal/mol for van der Waals interactions. A positive component indicates stabilization of Mg^{2+} binding to the enzyme.

Group	Component
ATP·Mg_2^{2+}	+367
Water	−73
Asp ligand	+45
Asp-475	+48
Glu-482	+78
Glu-219	+32
Arg-225	−21
Glu-171	+37
Arg-217	−23

AspAMP/AsnAMP Discrimination in the Presence of tRNA and Either One or Zero Mg^{2+} Cations

To further probe the role of cations in AspRS substrate binding specificity, we performed simulations of AspAMP and AsnAMP binding to AspRS in the presence of a single Mg^{2+} cation and the co-substrate tRNA. This system corresponds to the post-adenylation step, immediately before tRNA aminoacylation. A Mg^{2+} cation is generally thought to be associated with the aaAMP substrate (15), although its position is not always explicitly determined in x-ray structures. In our MD simulations, the cation remained strongly coordinated to the conserved acidic residues Glu-482 and Asp-475 in the binding pocket. PBFE cation binding calculations (Table 4) confirm that the Mg^{2+} cation is very strongly held, with an estimated electrostatic contribution to the binding free energy of 100 kcal/mol. This is comparable with the two strongest cations in the AspRS·aa·ATP·Mg_3^{2+} complexes (Table 4). PBFE calculations reported in Table 2 show that this cation is crucial for AspAMP/AsnAMP discrimination. For AspRS with the co-bound cation and a positively charged His-448, we computed an AspAMP/AsnAMP binding free energy difference $\Delta\Delta G$ of +12 kcal/mol in favor of AspAMP. pK_a calculations (Table 1) indicate a preference for neutral His-448 when AsnAMP replaces AspAMP, with a 3 kcal/mol free energy gain when His-448 becomes neutral. We therefore subtracted this 3 kcal/mol (Table 1) from our PBFE $\Delta\Delta G$, giving a corrected $\Delta\Delta G$ of +9 kcal/mol in favor of AspAMP (Table 2). Artificially removing the cation reduces the AspAMP/AsnAMP difference strongly, with a $\Delta\Delta G$ of just +4 kcal/mol. Interestingly, if one keeps the cation but artificially switches His-448 to its neutral state, one observes the same net effect, with a $\Delta\Delta G$ of just +4 kcal/mol. This offers a further illustration of the coupling between cations in the ATP/AMP binding site and the ionizable His-448, near the amino acid side chain.

DISCUSSION

Validation of Simulation Methodology—Despite the power of modern structural biology, molecular recognition is too complex to be completely understood with experiments alone: crystal structures do not show electric fields. MD gives complementary information that is difficult to obtain experimentally. For the same reason, some of our predictions are hard to test directly. Nevertheless, several lines of argument strongly suggest that our main results are meaningful.

For the crucial, long range, electrostatic interactions, we used two recent, sophisticated methods, including one that combines an atomic representation of groups near the active site with a continuum representation of distant groups (37, 47). The simplified treatment of distant groups is quite accurate. Indeed, calculations that use atomic regions of different sizes give very similar free energy changes. We note that although substrate interactions with distant groups have been observed experimentally and discussed extensively, they are weak compared with the effects of interest here. For example, AspRS is a homodimer, and a certain cooperativity is found for Asp and ATP binding to the two active sites (53). But the two binding constants for ATP differ by a factor of only 10, corresponding to 1.4 kcal/mol of binding free energy; Mg^{2+} effects are much larger.

To obtain thermodynamic information, we did MDFE simulations. MDFE has given reasonable accuracy for several proteins where direct experimental comparisons were possible; see recent work on redox potentials (54), side-chain pK_a shifts (55), and protein·protein binding (56, 57). Its validity for AspRS was tested earlier by comparison with experimental measurements of Asp-stimulated pyrophosphate exchange (36). We made systematic comparisons with PBFE, an independent method that has been extensively calibrated and tested for AspRS, and found good agreement. We also explored the sensitivity of our results to system details by studying several AspRS variants: *E. coli* AspRS with and without tRNA and *P. kodakaraensis* AspRS. Considering the different amino acid ligands and the numbers of Mg^{2+} cations and His protonation states, we simulated 24 different systems for over 75 ns, which is several times the total simulation length in related studies (54, 55, 56).

Several predictions could be tested experimentally, in principle. For LysRS, one could measure the effect of Mg^{2+} on the Lys affinity (predicted here to be small) using an established fluorescence assay (58). For both AspRS and LysRS, we predicted that ATP is always bound with three Mg^{2+} cations, so that the rate constant, k_{cat}, for the chemical adenylation step should be independent of Mg^{2+} concentration. We look forward to these predictions being tested by others.

Complexity of the Amino Acid Binding Reaction—AspRS·aa recognition is especially complex because it involves "hidden" reorganization, substrate-assisted specificity, and long range interactions. Indeed, the binding pocket is largely preorganized, and rearrangement upon aa binding is mostly limited to the flipping loop, which brings its negative Glu-171 into the pocket. However, a hidden reorganization also occurs, with His-448 binding a proton to oppose the Glu-171 charge. This labile proton favors Asp binding and helps discriminate against Asn.

ATP·Mg_3^{2+} acts as another mobile discriminator, so that AspRS specificity is substrate-assisted. The most distant Mg^{2+} cation contributes 4 kcal/mol to the Asp/Asn binding free energy difference. The strength of the coupling between the aa side chain and this cation is remarkable, given that they are 11 Å apart. We note that analogous couplings have been observed

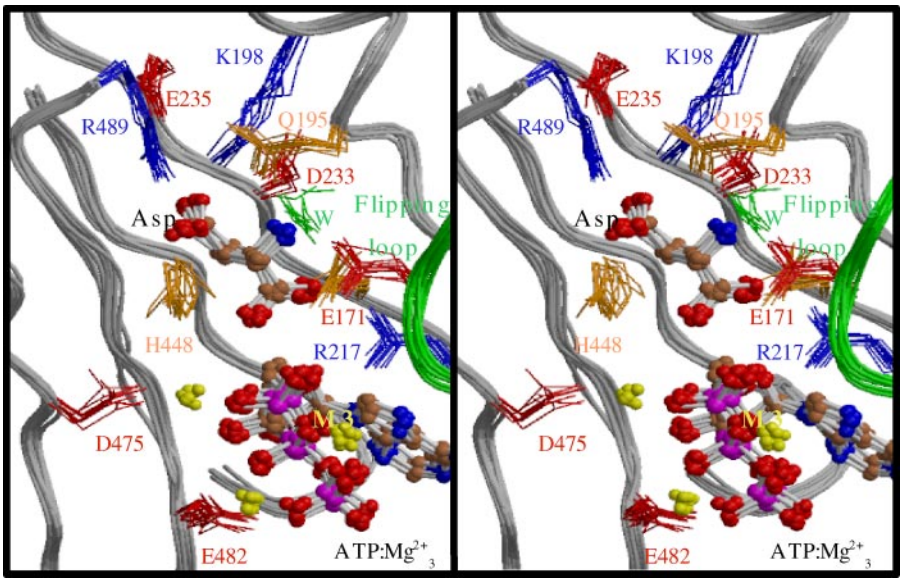

FIGURE 3. **Wall-eyed stereo view of 11 snapshots of the AspRS·Asp·ATP active site at 20-ps intervals, from the last 200 ps of the MD trajectory.** Asp and ATP are shown in *ball-and-stick* representation. Mg^{2+} cations are shown as *yellow spheres*. The most weakly bound cation, M3, is in front. Important binding pocket residues are labeled and colored according to charge: *red*, negative; *blue*, positive; *orange*, neutral. A stable water molecule is shown in *green* and labeled *W*. The protein backbone is shown in *tube* representation with flipping loop residues colored *green*. This figure was prepared using Molscript (48) and Raster3D (49).

between aa and ATP binding in other aaRSs, *e.g.* ThrRS (59) and MetRS (60). But the coupling here was significantly larger.

Another long range effect is the coupling between the labile His-448 proton and the third Mg^{2+} cation. Indeed, we predict (36) that when ATP·Mg_3^{2+} (but not ATP·Mg_2^{2+}) binds with Asn instead of Asp, His-448 will revert to its neutral form. Similarly, His-254 is predicted to be neutral when the third Mg^{2+} is present and positive when it is absent.

Mg^{2+} Binding Is Governed by Charge Balance and Long Range Interactions—The ATP phosphate groups are deprotonated at physiological pH, so that ATP interacts readily with cations. In solution, ATP usually binds one divalent cation. When complexed to a protein, it usually binds one or two (25). Yet our simulations show that the 3-cation form is overwhelmingly favored in AspRS. Given the role of the third cation, M3, in AspRS specificity, we want to understand its stability.

ATP binds preferentially to all 10 class II aaRSs with either three divalent cations or two cations and a positive protein side chain (see the data base analysis in the supplemental material). The fully bent, U-shaped geometry of ATP in AspRS is also characteristic of class II aaRSs and might be thought to play a role in M3 stabilization. Excluding aaRSs, only two of the 238 nonredundant NTP·protein complexes in the Protein Data Bank exhibit such a fully bent form (see supplemental material).

The most obvious M3 interactions (Fig. 3) are with the ATP β- and γ-phosphates and surrounding waters, and ATP·Mg_2^{2+} does make a very large contribution to M3 binding (+367 kcal/mol; Table 6). But these groups are not unique to AspRS or aaRSs, so that the stability of M3 is something of a puzzle. In fact, longer range interactions are at play. These were identified by the group decomposition of the binding free energy. A large contribution (+126 kcal/mol) comes from the conserved Glu-482 and Asp-475 side chains, 6 and 9 Å away from M3, respectively. Their negative charges, along with the four negative phosphate charges, exactly balance the charges of the Mg^{2+} cations. A large, unfavorable contribution (−73 kcal/mol) arises from solvent, corresponding to a dielectric shielding of the negative charges in the binding pocket. Finally, the Asp ligand itself contributes +45 kcal/mol, another example of long range coupling in the AspRS active site.

Mg^{2+} Contributes to AspRS Specificity in Synergy with Other Active Site Groups—This and earlier studies show that AspRS amino acid specificity arises from a complex network of electrostatic interactions involving conserved side chains and other factors. We have focused here on ATP and its associated cations found to play an important role. The simulations show that if the weakest bound Mg^{2+} cation is removed, the Asp/Asn discrimination is strongly reduced. The binding free energy difference $\Delta\Delta G$ drops from 9 ± 2 to 5 ± 2 kcal/mol. The Asp and Asn dissociation constants then differ by a factor of between $10^{-2.2}$ and $10^{-5.1}$. The range of values reflects the uncertainty in the MDFE estimation. If the third cation were absent 10% of the time, and we averaged correspondingly over the 2- and 3-cation states, the average Asp and Asn dissociation constants would differ by a factor of between $10^{-3.2}$ and $10^{-6.1}$. This is roughly comparable with the average error rate in protein synthesis, $0.03\% = 10^{-3.5}$.

This estimate assumes that His-448 is present. In fact, His-448 is conserved in prokaryotes but absent in eukaryotes. Fortunately, we found that AspRS is, in fact, protected against the 2-cation state, which is negligibly populated in all cases. By eliminating the 2-cation form and always binding ATP·Mg_3^{2+}, AspRS ensures a large Asp/Asn discrimination even in eukaryotes, where His-448 is absent.

The Mg^{2+} cations and the conserved side chains that coordinate them are required for AspRS activity, and the cations play both a structural and a catalytic role. It seems economical that the enzyme should also use them for a third purpose, to help ensure substrate specificity. In contrast, the evolutionarily related LysRS also binds ATP only in the 3-cation form, but artificially removing the third cation does not affect binding of the (positively charged) Lys substrate.

Fig. 4 summarizes the substrate binding specificity in AspRS as determined from the present simulations and earlier studies. Many different states are potentially available, the most important of which are shown in Fig. 4. Some of them produce a low Asp/Asn specificity, and these states are invariably unstable and negligibly populated. Overall, a delicate interplay between conformational shifts, ATP·Mg_3^{2+} binding, and histidine charging allows AspRS to combine a moderate Asp binding affinity with

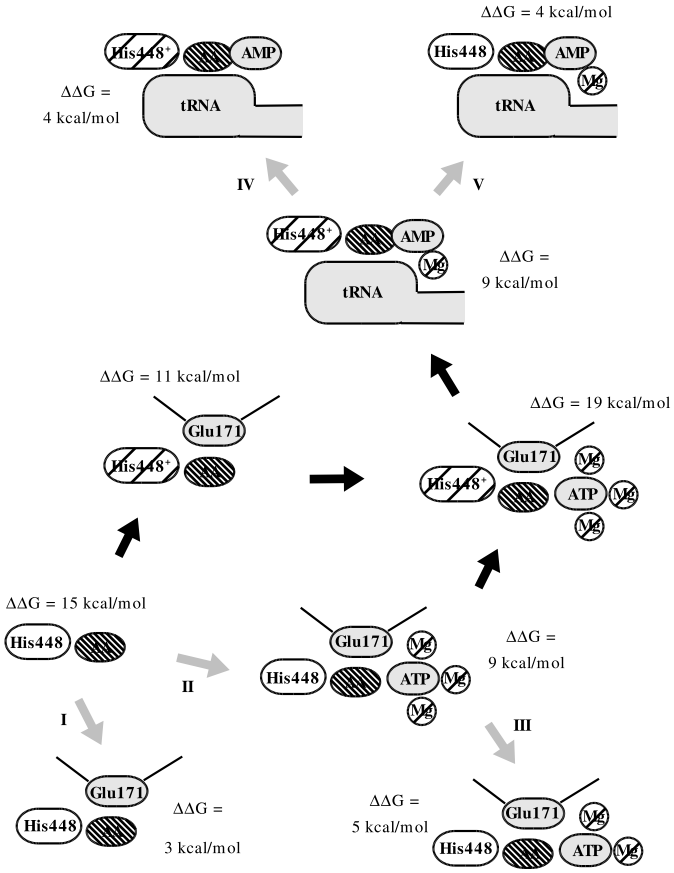

FIGURE 4. **Asp/Asn specificity in different AspRS states.** Each state is labeled with its associated Asp/Asn binding free energy difference ($\Delta\Delta G$) computed using MDFE, PBFE, or both (this work (Table 2) and earlier studies (33–46)). Selected groups are shown: the amino acid ligand (*AA*); nearby His-448; ATP and its cations; tRNA, aaAMP, and Glu-171 in states in which the flipping loop is closed (when the flipping loop is open, Glu-171 is not shown). The electric charge is shown by color or hatching: *gray* = negative, *hatched* = positive, *white* = neutral. The amino acid (*dense hatching*) can be negative or neutral. *Arrows* leading to inaccessible states are colored *gray* and labeled *I, II, III, IV*, and *V*. Following is a description of these labels and an explanation of why the less specific AspRS state is inaccessible in each case. *I*, loop closure is accompanied by histidine charging, with a calculated free energy of at least 17 kcal/mol in favor of histidine protonation (36). *II*, loop closure, even in the presence of positively charged ATP·Mg_3^{2+}, triggers histidine charging (36). Many organisms do not have this histidine, and so Asp/Asn binding without the charged histidine but with co-bound ATP·Mg_3^{2+} is an important step, for example in *P. kodakaraensis* (studied here). *III*, binding of the third cation is favored by 76 kcal/mol, so the Mg_2^{2+} state is unpopulated even at low cellular Mg^{2+} concentrations (Table 5). *IV*, the cation associated with the aaAMP ligand is always present (Table 4). *V*, with bound AspAMP, His-448 has a charging free energy of −8 kcal/mol (favoring the charged state), so Asp/Asn competition does not occur within the neutral His-448 state (Table 1).

a very strong Asp/Asn binding specificity. It is striking that different steps in the aminoacylation process feature a similar charge balance in the Asp binding pocket, as reflected in the computed Asp/Asn binding free energy differences. For example, both loop closure in the preadenylation step and tRNA binding in the post-adenylation step trigger His-448 charging, and both ATP and AspAMP have very strongly bound cations. These Mg^{2+} cations act as mobile discriminators in AspRS. In tandem with His-448 charging, they compensate electrostatically for the Asp binding penalties associated with loop closure and tRNA binding.

AspRS amino acid binding is thus a complex process involving switching of protonation states, substrate-assisted specificity, and long range, electrostatic interactions. We expect that as simulation methods become increasingly reliable and increasingly integrated within structural biochemistry, the general character of such effects in protein·ligand recognition will become more apparent.

Acknowledgments—Calculations were done at the CINES Supercomputer Center of the French Ministry of Research. We thank Pierre Plateau, Georgios Archontis, and Sylvain Blanquet for helpful discussions and Martin Karplus for the CHARMM program.

REFERENCES

1. Ibba, M., Francklyn, C., and Cusack, S. (eds) (2005) *Aminoacyl-tRNA Synthetases*, Landes Bioscience, Georgetown, TX
2. Arnez, J. G., and Moras, D. (1994) in *RNA-Protein Interactions* (Nagai, K., and Mattaj, I., eds) pp. 52–81, Oxford University Press, Oxford, UK
3. Meinnel, T., Mechulam, Y., and Blanquet, S. (1995) in *tRNA: Structure, Biosynthesis and Function* (Söll, D., and Raj Bhandary, T. L., eds) pp. 251–291, ASM Press, Washington, D. C.
4. Schimmel, P., and Ribas de Pouplana, L. (2000) *Trends Biochem. Sci.* **25,** 207–209
5. Pallanck, L., Pak, M., and Schulman, L. H. (1995) in *tRNA: Structure, Biosynthesis and Function* (Söll, D., and Raj Bhandary, T. L., eds) pp. 371–393, ASM Press, Washington, D. C.
6. Arnez, J. G., and Moras, D. (1997) *Trends Biochem. Sci.* **22,** 211–216
7. Francklyn, C. S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9650–9652
8. Baldwin, A. N., and Berg, P. (1966) *J. Biol. Chem.* **241,** 839–845
9. Fersht, A. (1986) in *Accuracy of Molecular Processes* (Kirkwood, T., Rosenberger, R., and Galas, D., eds) pp. 67–82, Chapman and Hall, New York
10. Soutourina, J., Plateau, P., and Blanquet, S. (2000) *J. Biol. Chem.* **275,** 32535–32542
11. Hopfield, J. J. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 4135–4139
12. Artymiuk, P. J., Rice, D. W., Poirrette, A. R., and Willett, P. (1994) *Nat. Struct. Biol.* **11,** 758–760
13. Jakubowski, H. (1999) *Biochemistry* **38,** 8088–8093
14. Pezo, V., Metzgar, D., Hendrickson, T. L., Waas, W. F., and Doring, V. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 8593–8597
15. Poterszmann, A., Delarue, M., Thierry, J. C., and Moras, D. (1994) *J. Mol. Biol.* **244,** 158–167
16. Delarue, M., Poterszman, A., Nikonov, S., Garber, M., Moras, D., and Thierry, J. C. (1994) *EMBO J.* **13,** 3219–3229
17. Schmitt, E., Moulinier, L., Fujiwara, S., Imanaka, T., Thierry, J. C., and Moras, D. (1998) *EMBO J.* **17,** 5227–5237
18. Eiler, S., Dock-Bregeon, A. C., Moulinier, L., Thierry, J. C., and Moras, D. (1999) *EMBO J.* **18,** 6532–6541
19. Sauter, C., Lorber, B., Cavarelli, J., Moras, D., and Giege, R. (2000) *J. Mol. Biol.* **299,** 1313–1324
20. Rees, B., Webster, G., Delarue, M., Boeglin, M., and Moras, D. (2000) *J. Mol. Biol.* **299,** 1157–1164
21. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991) *Science* **252,** 1682–1689
22. Briand, C., Poterszman, A., Eiler, S., Webster, G., Thierry, J. C., and Moras, D. (2000) *J. Mol. Biol.* **299,** 1051–1060
23. Moulinier, L., Eiler, S., Eriani, G., Gangloff, J., Thierry, J. C., Gabriel, K., McClain, W. H., and Moras, D. (2001) *EMBO J.* **20,** 5290–5301
24. Ador, L., Jaeger, S., Geslain, R., Martin, F., Cavarelli, J., and Eriani, G. (2004) *Biochemistry* **43,** 7028–7037
25. Frausto da Silva, J. J. R., and Williams, R. J. P. (2001) *The Biological Chemistry of the Elements*, Clarendon Press, Oxford
26. Cavarelli, J., Eriani, B., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J. C., and Moras, D. (1994) *EMBO J.* **13,** 327–337
27. Berthet-Colominas, C., Seignovert, L., Hartlein, M., Grotli, M., Cusack, S., and Leberman, R. (1998) *EMBO J.* **17,** 2947–2960

28. Desogus, D., Tadone, F., Brick, P., and Onesti, S. (2000) *Biochemistry* **39,** 8418–8425
29. Fersht, A. R. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman & Company, New York
30. Blanquet, S., Fayat, G., and Waller, J. P. (1974) *Eur. J. Biochem.* **44,** 343–351
31. Simonson, T. (2003) *Rep. Prog. Phys.* **66,** 737–787
32. Fothergill, M., Goodman, M. F., Petruska, J., and Warshel, A. (1995) *J. Am. Chem. Soc.* **117,** 11619–11627
33. Archontis, G., Simonson, T., Moras, D., and Karplus, M. (1998) *J. Mol. Biol.* **275,** 823–846
34. Archontis, G., and Simonson, T. (2001) *J. Am. Chem. Soc.* **123,** 11047–11056
35. Archontis, G., Simonson, T., and Karplus, M. (2001) *J. Mol. Biol.* **306,** 307–327
36. Thompson, D., Plateau, P., and Simonson, T. (2006) *ChemBioChem* **7,** 337–344
37. Simonson, T. (2000) *J. Phys. Chem. B* **104,** 6509–6513
38. Beglov, D., and Roux, B. (1994) *J. Chem. Phys.* **100,** 9050–9063
39. Stote, R. H., States, D. J., and Karplus, M. (1991) *J. Chim. Phys.* **88,** 2419–2433
40. MacKerrell, A. D., Bashford, D., Bellott, M., Dunbrack, D. L., Evanseck, J. D., and Field, M. J. (1998) *J. Phys. Chem. B* **102,** 3586–3616
41. Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R., and Klein, M. (1983) *J. Chem. Phys.* **79,** 926–935
42. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaninathan, S., and Karplus, M. (1983) *J. Comp. Chem.* **4,** 187–217
43. Simonson, T. (2001) in *Computational Biochemistry and Biophysics* (Becker, O. M., MacKerell, A. D., Roux, B., and Watanabe, M., eds) pp. 169–197, Marcel Dekker Inc., New York
44. Simonson, T., Archontis, G., and Karplus, M. (2002) *Acc. Chem. Res.* **35,** 430–437
45. Archontis, G., and Simonson, T. (2005) *Biophys. J.* **88,** 3888–3904
46. Gao, J., Kuczera, K., Tidor, B., and Karplus, M. (1989) *Science* **244,** 1069–1072
47. Simonson, T., Archontis, G., and Karplus, M. (1997) *J. Phys. Chem. B* **101,** 8349–8362
48. Kraulis, P. J. (1991) *Appl. Crystallogr.* **24,** 946–950
49. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr Sect. D Biol. Crystallogr.* **50,** 869–873
50. Cusack, S. (1997) *Curr. Opin. Struct. Biol.* **7,** 881–889
51. Arnez, J. G., Flanagan, K., Moras, D., and Simonson, T. (1998) *Proteins* **32,** 362–380
52. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) *Molecular Biology of the Cell*, p. 161, Garland Science, New York
53. Caverelli, J., Rees, B., Ruff, M., Thierry, J. C., and Moras, D. (1993) *Nature* **362,** 181–184
54. Simonson, T. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6544–6549
55. Simonson, T., Carlsson, J., and Case, D. A. (2004) *J. Amer. Chem. Soc.* **126,** 4167–4180
56. Zeng, J., Fridman, H., Maruta, H., Treutlein, H., and Simonson, T. (1999) *Protein Sci.* **8,** 50–64
57. Hénin, J., and Chipot, C. (2005) *J. Am. Chem. Soc.* **127,** 8478–8484
58. Hughes, S. J., Tanner, J. A., Hindley, A. D., Miller, A. D., and Gould, I. R. (2003) *BMC Struct. Biol. 2003* **3,** 5
59. Bovee, M. L., Pierce, M. A., and Francklyn, C. S. (2003) *Biochemistry* **42,** 15102–15113
60. Blanquet, S., Fayat, G., and Waller, J. P. (1975) *J. Mol. Biol.* **94,** 1–15