Fidelity of Phenylalanyl-tRNA Synthetase in Binding the Natural Amino Acids

Peter M. Kekenes-Huskey, Nagarajan Vaidehi, Wely B. Floriano, and William A. Goddard III*

Materials and Process Simulation Center, California Institute of Technology, Pasadena, California 91125

Received: March 10, 2003; In Final Form: June 17, 2003

Aminoacyl-tRNA synthetases guard the fidelity of cognate amino acid incorporation during protein biosynthesis; for example, phenylalanyl-tRNA synthetase (PheRS) activates and transfers only Phe to its tRNA. Since we are interested in using a computational protocol to identify nonnatural amino acids that are incorporated by wild-type PheRS, it is critical to understand the fidelity of PheRS in binding the 20 natural amino acids. To this end, HierDock, a computational protocol for predicting binding sites and relative binding affinities, was used for testing the natural amino acids in PheRS. Scanning the entire ligand-accessible protein surface for the best binding region, we find that HierDock correctly identifies the active site of Phe in PheRS and predicts Phe within 0.61 Å RMSD of the crystal structure. HierDock also successfully shows PheRS discriminates for Phe, as the noncognate amino acids bind less favorably in the binding site of Phe. However, we find that Met, Cys, and Tyr bind competitively but at positions distant from the Phe binding site. This result corroborates in vitro measurements of aminoacyl adenylate formation, which show Met competes with Phe at the amino acid binding stage. We predict that the binding site of Met would not activate PheRS, as the noncognate amino acid cannot establish suitable hydrogen bonds with the PheRS reaction center. These results validate the use of HierDock in predicting the binding sites of the cognate amino acids in PheRS. The HierDock procedure calculates the discrimination of aminoacyl-tRNA synthetases at the stage of binding the cognate amino acid and offers a molecular level understanding of the mistakes made in protein biosynthesis that are not readily uncovered through experiments. This technique is also useful for predicting the binding of a selected nonnatural amino acid analogue, thereby indicating whether the molecule would be incorporated into a wild-type aminoacyl-tRNA synthetase.

Introduction

The integrity of translation in protein synthesis relies on the precise delivery of amino acids to the ribosomes, according to the coded mRNA sequence.[1] The integration of amino acids into a newly synthesized polypeptide chain is reliant on three crucial steps: (1) cell uptake of the amino acids, (2) association with the cognate tRNA via the respective aminoacyl-tRNA synthetase, (3) recognition by the ribosome and the appropriate elongation factors therein.

The cellular uptake of amino acids has been shown experimentally to be indiscriminate in this process, whereas the latter steps are more selective. Of the selective steps, the unification of a given amino acid with its cognate tRNA is critical for the precise and accurate translation of the genetic code[2] to preclude the incorporation of incorrect amino acids into the protein.

The 3′-esterification of a tRNA with its respective amino acid yields the aminoacyl-tRNA complex, which delivers the amino acid to the site of protein synthesis. This reaction is mediated by a class of high-fidelity enzymes known as aminoacyl-tRNA synthetases (aaRSs), which, like the tRNA monomers, are unique to each of the 20 amino acids. The esterification reaction occurs in a two-step process, in which an adenosine triphosphate (ATP) serves as an intermediary in the aminoacylation of tRNA:[2]

$$\text{aa} + \text{ATP} + \text{aaRS} \leftrightarrow \text{aaRS}\cdot\text{aa-AMP} + \text{PP}_\text{i}$$

$$\text{aaRS}\cdot\text{aa-AMP} + \text{tRNA} \leftrightarrow \text{aaRS} + \text{aa-tRNA} + \text{AMP}$$

Each transformation involves multiple steps such as the aaRS recognition and binding of its cognate amino acid and mechanisms for correcting errors utilizing differences in physical binding and chemical proofreading.[3] Thus, the steps involved in the transfer of an aminoacyl group to the tRNA can be written as (1) binding of an amino acid and ATP to the aaRS, (2) conformational changes in the aaRS induced by binding and formation of the aminoacyl adenylate complex, (3) proofreading of misactivated noncognate aminoacyl adenylate complexes, and (4) transfer of the aminoacyl group to the tRNA and proofreading of the product.

The physical binding of the amino acid and ATP to aaRS achieved in steps 1 and 2 includes a conformational change in the aaRS upon binding. The chemical proofreading mechanism is accomplished in steps 3 and 4, which involve hydrolysis of two reaction products: either the aminoacyl adenylate complex in the pretransfer proofreading or the aminoacyl-tRNA in the posttransfer proofreading. Each step serves to increase the specificity in the recognition by the aaRS of its cognate amino acid, while discriminating more efficiently against noncognate amino acids. However, the degree to which each aaRS uses the specificity-enhancing steps varies considerably over the 20 naturally occurring amino acids and depends on the specific aaRS. For example, tyrosyl-tRNA synthetase has the highest specificity in the first binding step,[4] whereas isoleucyl-tRNA synthetase[4] achieves maximum discrimination in the pretransfer proofreading step. It is known that the rate of mischarging the tRNA[5] is much lower than the rate of misactivating the noncognate amino acid[6] by the aaRS. Hence, it is critical to understand the molecular mechanism by which the amino acids are recognized by aaRS.

* To whom correspondence should be addressed. Phone: (626) 395-2731. E-mail: wag@wag.caltech.edu.

10.1021/jp034607o CCC: $25.00
Published on Web 09/23/2003

Each aaRS has a binding cavity unique to its cognate amino acid. The active site generally contains hydrophilic residues that stabilize the amino and carboxyl termini of the bound amino acid with additional residues (hydrophobic or hydrophilic depending on the nature of the cognate amino acid) defining a cavity that further stabilizes the side chain[7] of the preferred amino acid, constraining it to an ideal location for further catalysis. The proper positioning of the carboxylate terminus of the amino acid facilitates reaction with a bound ATP structure to give the anticipated aminoacyl adenylate product. Several groups have reengineered aaRS or tRNA by performing mutations that modify specificities for the cognate amino acid,[8] its competitor,[9] or an amino acid analogue.[10−14]

Phenylalanyl-tRNA synthetase (PheRS) is a class II aaRS, which is the category of aaRSs whose members do *not* undergo large conformational changes upon substrate binding. PheRS is tetrameric, $(\alpha\beta)_2$, consisting of two domains each with α and β regions, the former of which is involved in amino acid binding. X-ray crystallographic structures of PheRS from *Thermus thermophilus* including a bound phenylalanine (Phe) and a Phe-AMP analogue have recently been solved at resolutions of 2.7 and 2.8 Å, respectively.[15]

Our primary interest is to determine the molecular mechanism underlying the fidelity of PheRS to binding of the 20 natural amino acids. Our goal is to understand how the various residues comprising the ligand binding domain (LBD) contribute to the recognition of the cognate amino acid[16] to streamline the engineering of nonnative amino acid incorporation. To validate our procedures for doing this, we report in this paper the application of the HierDock computational protocol[17] to predict the discrimination of PheRS in binding of the 20 natural amino acids and forming an aminoacyl adenylate complex. Though the binding site of Phe in PheRS is experimentally established, we have used the HierDock procedure to search the entire protein surface to validate this computational method and suggest its applicability in similar trials.

HierDock utilizes a hierarchical strategy based on a coarse-grain conformational search with DOCK 4.0 coupled with successively finer grained molecular mechanics (MM) optimization for a certain percentage of these best performing conformations including solvation and an all-atom force field using MPSim. Previous studies with a test set of 37 cocrystallized structures demonstrated that DOCK 4.0 alone could only predict 10 of the appropriate ligand conformations within a coordinate RMS (CRMS) of 2.0 Å, while the HierDock protocol predicted all 37 within this criterion. Previous applications on membrane-bound olfactory receptors,[18] G-protein-coupled receptors,[19] and analogues of Phe binding to apo-PheRS[20] showed HierDock to be effective in predicting binding sites and binding affinities. The current study on the PheRS globular protein serves as an additional test for the HierDock protocol.

Searching the entire protein for a favorable binding region, we find that HierDock predicts the correct binding region of Phe in PheRS. The predicted binding conformation of Phe in PheRS agrees with the crystal result within 0.61 Å CRMS. Predicting the binding site and energy of the 19 other natural amino acids to PheRS, we find that Phe has the best binding energy.

Methods

Preparation of Structures. The zwitterions of all 20 amino acids were selected for the ligand pool, as this configuration is the most stable for the pH conditions in the blood. These were optimized in the extended conformation at the Hartree−Fock level of theory with a 6-31G** basis set,[21] including solvation according to the Poisson−Boltzmann equation[22] using the Jaguar[23] computational suite. The Mulliken charges ascertained from this calculation were retained for the subsequent molecular mechanics simulations. Although in this case the amino acid ligands could be fitted with CHARMM22 charges, we instead applied quantum-mechanically derived charges to ensure a general method, which could later be applied to a host of ligand types.

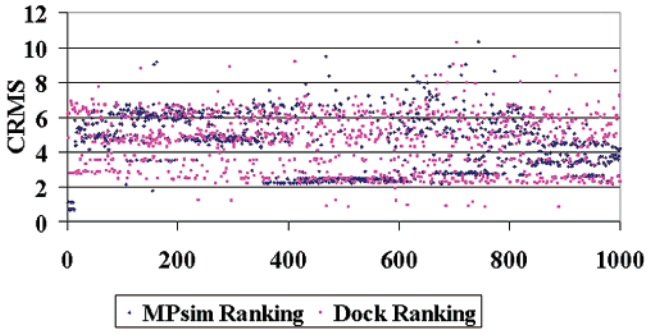

Figure 1. CRMS of the ligand-docked structures with the crystal Phe structure vs DOCK 4.0 and MPSim ranked conformations (best ranked on the left).

The PheRS structure obtained from Reshetnikova et al. (PDB code 1B70) included the α and β domains, a bound phenylalanine ligand, crystal waters, and a magnesium(II) ion. Both domains, as well as the additional components, were assigned CHARMM22[24] charges, with the nonpolar hydrogen charges summed onto the heavy atoms, according to the parameters set forth in the DREIDING force field (FF).[25] The δ-nitrogen for His178 was protonated, following hydrogen bond analysis with WHATCHECK.[26] The entire protein was neutralized by adding counterions (Na^+ and Cl^-) to the charged residues (D, E, R, K) and subjected to a minimization of the potential energy using conjugate gradients (MPSim) to an RMS force of 0.1 kcal/mol. The CRMS for this procedure, 0.39 Å, is well within experimental error and demonstrates the proficiency of our FF used in present studies. The crystal waters and other bound molecules were removed for docking to maximize the searchable surface of the protein.

The SGB continuum solvation model provided the basis for quantifying desolvation effects of the selected ligands. A dielectric constant of 2.5 was employed for inside the protein and a dielectric constant of 80.4 was used for the solvent in all calculations. The solvation energies were ascertained from single-point energies of gas-optimized structures.

HierDock Protocol. Without any bias from the X-ray crystal data, we assumed the binding site of PheRS was unknown and docked Phe into all defined regions of the protein to determine the most suitable for ligand binding. We employed the hierarchical docking protocol,[18−20,27] which has been shown to efficiently scan an entire receptor for all possible binding sites without any structural bias. This scanning procedure is based on docking via DOCK 4.0[28,29] and coupled with fine-grain MM techniques. The coarse-grain docked complex structures generated are scored with FF and differential solvation, which effectively filters the docked complexes to isolate the top contenders. As demonstrated in Figure 1, the CRMS for the DOCK 4.0 generated structures from the crystal structure for the ligand vary erratically with the DOCK 4.0 rank, whereas filtering with MPSim optimization appears to group the low CRMS (to the crystal) structures to the top rankers. Here we show that a two-step hierarchical approach is essential to get the best structures to the top. This hierarchical approach involves

keeping at least 10% of DOCK 4.0 generated structures with further optimization with an all-atom FF and solvation. The scanning is conducted with a ligand of known binding affinity to the receptor for the isolation of a plausible binding site. Successive docking of the entire ligand pool to this site clarifies the relative affinity of each member with respect to the known binding substrate. This docking protocol is comprised of the following steps.

(1) Mapping of possible binding regions. A probe of 1.4 Å radius was used to trace a 4 dots/Å negative image of the protein molecular surface, according to Connolly's method.[30] The resulting data were used to generate clusters of overlapping spheres with the SPHGEN[31] program. These spheres serve as the basis for the docking method.

(2) Definition of the docking region. The pockets of ligand-accessible empty space of the receptor surface represented by spheres were divided into 13 10 Å wide overlapping cubes, which covered the entire protein surface except for sequences Val86−His101, Leu154−Gly171, and Pro288−Gly302, those regions unavailable for ligand binding. For each region we retained 1000 docked Phe configuraions and selected the 100 with the lowest DOCK energies. We found that site A had the greatest binding energy with over 20 low-energy docked conformations, whereas two boxes had two low-energy structures, and the others none. Thus, we picked site A as the binding region.

(3) Generation of docked conformations for the ligand−receptor complex. The orientations of the ligand in the receptor were generated by DOCK 4.0, using flexible docking with torsional minimization of the ligand, a continuum dielectric of 80.4, and a distance cutoff of 10 Å for the evaluation of energy. A total of 1000 conformations were generated and ranked according to the DOCK 4.0 scoring function, and the top 100 structures were kept for further optimization.

(4) MPSim optimization of the complexes. The top-ranking DOCK structures were subjected to further optimization, using the more accurate DREIDING full-atom FF with SGB solvation. The first stage of gas-phase optimization utilized a fully flexible ligand with a fixed protein, followed by a single-point energy calculation of the solvation using a non-distant-dependent dielectric of 80.4. A buried surface calculation for the ligands with a minimum threshold of 75% selected only those structures that were sufficiently buried within the protein. The 10 lowest energy conformations underwent further all-atom gas optimization with a single-point solvation energy calculation to screen for the best binding candidate.

(5) Selection of the most probable binding site and best configurations. The top five conformations with the lowest energy score were selected for subsequent full ligand−protein optimization with solvation.

(6) Docking of the ligand pool into the binding site. Steps 3−5 were repeated for each member of the ligand pool to obtain relative binding affinities.

(7) Ranking of ligand affinities. The relative binding energies for the best ligand conformations are defined as the difference between the ligand in protein versus in solution given by

$$\Delta\Delta G_{\text{calcd}} = \Delta G(\text{protein} + \text{ligand}) - \Delta G(\text{protein}) - \Delta G(\text{ligand}) \quad (1)$$

where ΔG(protein + ligand) is the free energy for the protein−ligand complex, ΔG(protein) is the free energy for the protein, and ΔG(ligand) is the free energy for the ligand alone. Since the structure optimizations included solvation forces using the SGB continuum solvent approximation with the experimental dielectric constant, we consider that the calculated energies are free energies.[32] The amino acids can then be ranked according to binding affinities to determine which ligands have the highest affinity for the binding site.

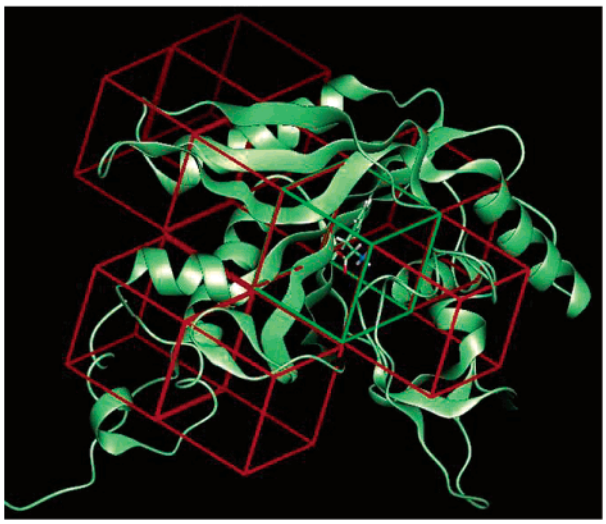

Figure 2. The 13 potential binding regions used for HierDock in phenylalanyl-tRNA synthetase. The green box denotes the HierDock-predicted binding region for PheRS.

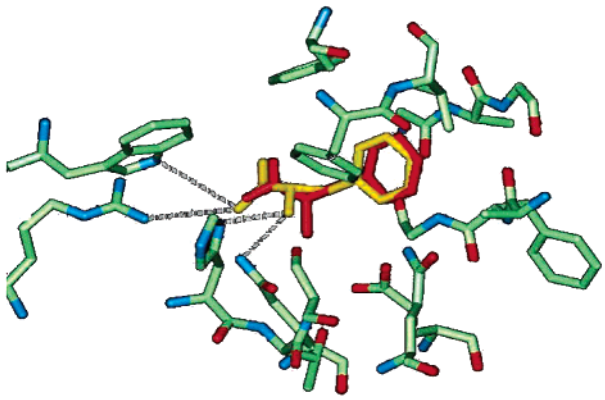

Figure 3. Superposition of the predicted (yellow) and crystal (red) Phe structure binding in PheRS.

Results

Docking PheRS with Phe and the Location of the Binding Site. The initial conjugate gradient minimization on the crystal structure of PheRS with Phe bound (PDB code 1B70) yielded a structure whose CRMS for all heavy atoms between Phe/PheRS(crystal) and Phe/PheRS(FF) is 0.39 Å, which is well within the resolution of the crystal structure. The entire negative surface of the α domain of the minimized PheRS was partitioned into 13 potential docking regions, as indicated by the boxes in Figure 2. The top 100 Phe conformations from DOCK 4.0 were scored using the DREIDING FF with SGB solvation. Of the 13 boxes covering the PheRS surface, we found that region 1 (the highlighted green box in Figure 2) contains the lowest energy conformations of Phe, which agrees with experimental data. The structures in this binding region were minimized using MPSim[33] with all-atom flexibility to optimize the ligand−protein complex. The most favorable conformation of bound Phe from this procedure (shown in Figure 3) differs from the crystal data by CRMS = 0.61 Å. We denote our predicted cocrystal structure of Phe in PheRS as Phe/PheRS(predicted).

Here we use primarily the vertical binding energy to discriminate between potential binding candidates, which is

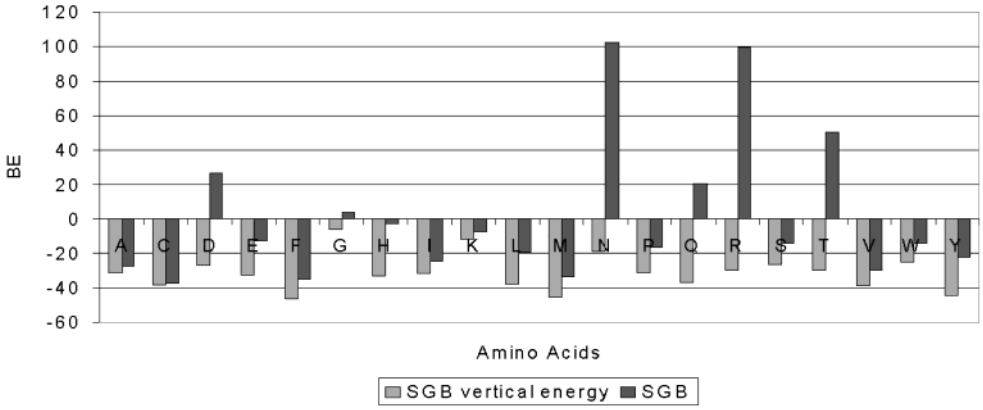

Figure 4. Vertical (light gray) and adiabatic (dark) binding energies for all 20 amino acids in PheRS, based on MPSim all-atom calculation. The vertical binding energies suggest that Cys would be a competitor of Phe for this binding site, which has been confirmed exprimentally (however, Cys is not incorporated into the growing protein).

		Critical Hydrogen Bonds of Bound Phe		
Phe	Protein	predicted	crystal	minimized crystal
N	O_γ – Ser 180	2.6	3.6	4.0
N	$O_{\varepsilon 1}$ – Gln 218	4.3	3.3	3.1
N	$O_{\varepsilon 2}$ – Glu 220	4.7	4.0	4.3
O	$N_{\varepsilon 1}$ – Trp 149	4.2	3.3	3.0
O	$N_{\delta 1}$ – His 178	3.9	3.5	3.4
O	$N_{\eta 2}$ – Arg 204	3.8	3.3	2.9
O	$N_{\varepsilon 2}$ – Gln 218	2.9	3.5	3.0

Figure 5. Hydrogen bond distances (Å) of Phe ligand to participating residues of *T. thermophilus* PheRS. Column 3 has the currently calculated values, while column 4 has the values measured in the crystal structure[15] and column 5 the values in the minimized crystal structure.

defined as the difference in binding energy between the Phe/PheRS(predicted) complex in solution and the individual components (free protein and ligand in solution) without allowing for subsequent relaxation of the individual components. This reflects the maximum possible binding energy between the ligand and substrate. As a basis of comparison, in Figure 4 we include the adiabatic binding energy, which is equivalent to the vertical binding energy but allows for relaxation of the protein and ligand components upon dissolution of the complex. We find that nonbond terms (vdW and electrostatic) dominate the binding energy (compared to valence terms), with the Coulombic term as the largest term, for which the established hydrogen bonds contribute greatly. The hydrogen bond distances (Figure 5) formed by the carboxyl terminus of the Phe ligand with PheRS in Phe/PheRS(predicted) compare well with both Phe/PheRS(crystal) and Phe/PheRS(FF) experimental data, as follows:

(i) The N−OG bond to Ser180 (2.6 Å calculated vs 3.56 Å in the crystal) and the O−NE2 bond to Gln218 are shorter (2.9 Å calculated vs 3.46 Å in crystal). (ii) The N−OE1 Gln218 (4.3 Å vs 3.31 Å), N−OE2 Glu220 (4.7 Å vs 3.97 Å), and O−NE1 Trp149 (4.2 Å vs 3.32 Å) interactions are longer. (iii) The hydrogen bonds to His178 (3.9 Å vs 3.5 Å) and Arg204 (3.8 Å vs 3.28 Å) are roughly similar. (iv) The side chains of Phe258, Phe260, Val261, Gly282, Ala283, Gly315, and Gly316 form a hydrophobic pocket, in which the phenyl ring of the liganded Phe is stabilized.

Thus, the worst agreement is with the hydrogen bond formed by the amino terminus of Phe to Trp149, which is 0.9 Å longer than in the crystallographic structure.

Docking All Natural Amino Acids. The procedure for determining the binding mode of Phe was repeated for the other 19 natural amino acids. In each case we considered just region

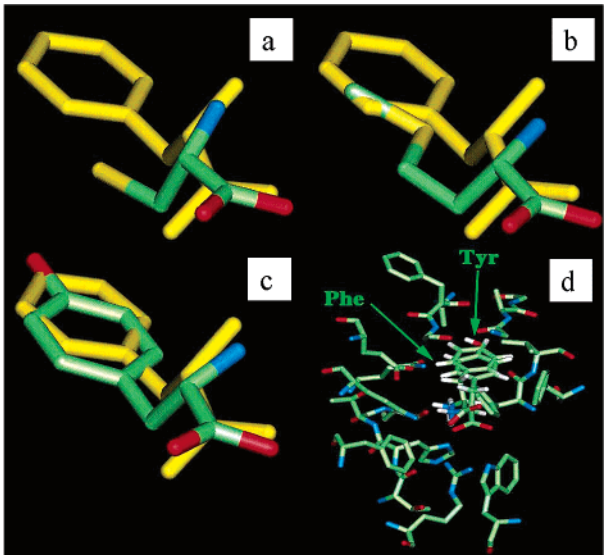

Figure 6. Predicted conformation of various competitors (full color) overlaid onto the predicted Phe (yellow) structure: (a) Cys binding conformation compared to that of Phe; (b) Met compared to Phe; (c) Tyr compared to Phe; (d) binding site of Tyr in PheRS compared to that of Phe.

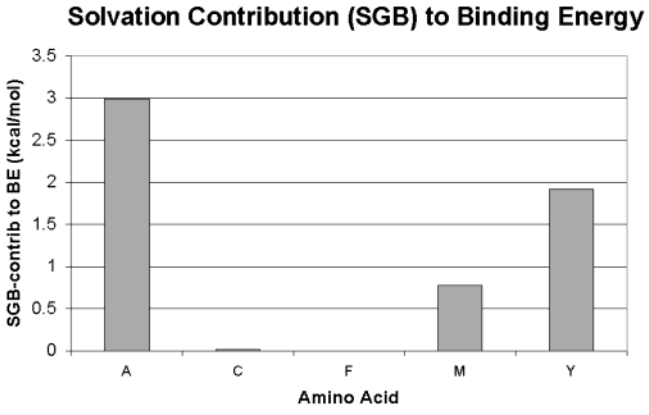

Figure 7. Relative reduction of ligand binding energy due to differential solvation (SGB). F is taken as the reference. Note that A and Y are especially excluded by this term.

1, but sampled 1000 configurations with DOCK and kept the best 100. Then we did MPSim minimization with the DREIDING FF, Q_M charges, and SGB solvation. The binding energies calculated for the best conformation of each amino

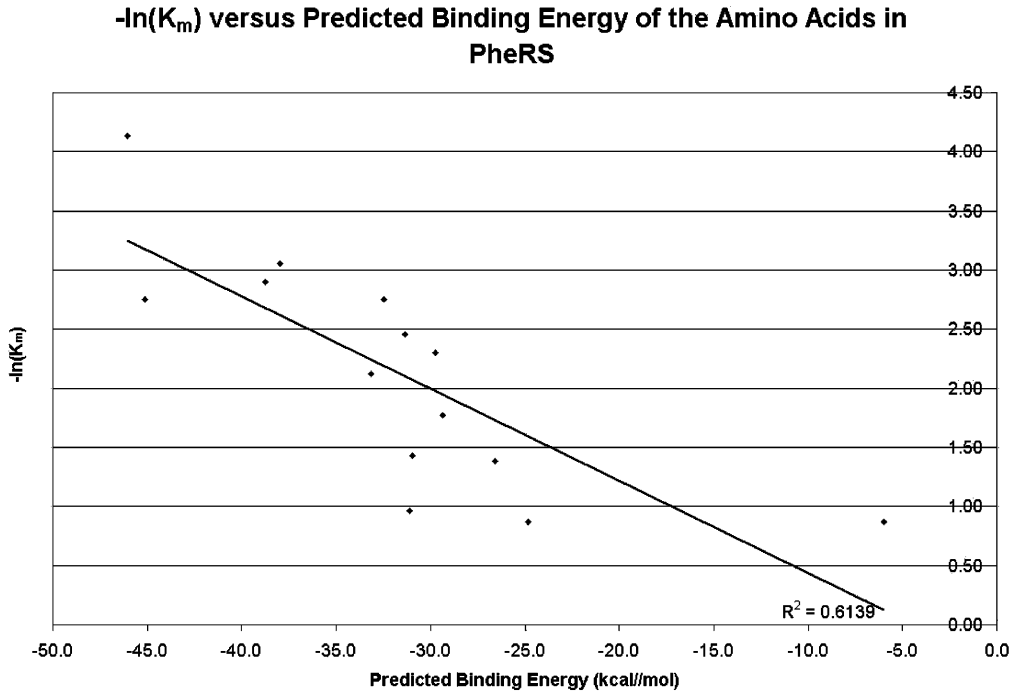

Figure 8. Comparison of the experimental K_m values (310 K) for the 13 neutral amino acids to PheRS determined by Friest[36] with the predicted binding energies (0 K) from HierDock. The R^2 of 0.61 goes to 0.68 if the Gly point is excluded. Note, the experimental values are for *S. cerevisiae*, while the predictions are for *T. thermophilus*.

acid are shown in Figure 4. We see here that Phe has the best binding energy of all 20 amino acids, although the binding energies of cysteine, methionine, and tyrosine are very close.

To provide a qualitative description of the relative positions of the amino acids in the PheRS binding site, the best scoring amino acids, Ala, Cys, Phe, Met, and Tyr, are compared to Phe in (Figure 6). These amino acids have roughly the same orientation, but the burial into the hydrophobic pocket and the backbone positions vary considerably from those of the Phe structure. This perhaps accounts for the relative preference in desolvation energy for Phe with respect to the noncognate amino acids, for which Ala and Tyr are especially excluded by this term by at least approximately 2.0 kcal/mol (Figure 7). Thus, the CRMS over the backbone atoms of bound Met differs from that of bound Phe by 1.4 Å. We expect that this orientation of the carboxyl and amino termini is critical for the subsequent reaction to form the aminoacyl adenylate complex.

To determine the ability of the HierDock protocol in ranking the amino acid binding energies with respect to experimental data, we have included a plot (Figure 8), as well as the raw data (Figure 9), of the experimentally determined Michaelis constant, K_m,[34] for PheRS in *Sacchromyces cerevisae* versus our predicted binding energies for the amino acids in *T. thermophilus* PheRS. Though the PheRS sequences of both species are certainly highly homologous, variations in the nature of the residues comprising the binding site lead to substrate binding energies that could potentially vary. Nevertheless, after removing several of the data points corresponding to charged and polar residues (Cys, Lys, Glu, Tyr, Ser) from the data set, we were able to obtain a rough correspondence of 0.61, which grew to 0.68 if Gly was also removed. The less than ideal correspondence, despite using different species, could indicate the shortcomings of the model in accurately predicting the binding energies of polar molecules, as well as suggest that refinement of the force field is necessary for more accurate ranking of binding affinities. On the other hand, this correlation indicates that we are making progress in the right direction.

Calculated Binding Energy vs. Experimental K_m			
AA	BE_{calc} (kcal/mol)	K_m (mM)	-ln(K_m)
F	-46.0	0.02	4.14
M	-45.1	0.06	2.75
V	-38.7	0.06	2.90
L	-38.0	0.05	3.06
H	-33.1	0.12	2.12
E	-32.5	0.06	2.75
I	-31.3	0.09	2.45
P	-31.0	0.24	1.43
T	-29.7	0.10	2.30
R	-29.3	0.17	1.77
D	-26.6	0.25	1.39
W	-24.8	0.42	0.87
G	-6.0	0.42	0.87
A	-31.1	0.38	0.97
N	-19.1	not reported	N/A
Excluded Polar and Charged Residues			
C	-38.4	0.72	0.33
K	-11.5	0.12	2.12
Q	-36.8	0.02	4.14
S	-26.4	1.10	-0.10
Y	-44.5	0.21	1.56

Figure 9. Predicted binding energies from HierDock of the 20 amino acids in PheRS compared with the experimentally determined K_m values of Friest (for yeast PheRS).

Binding Energies of Natural Amino Acids in the Active Conformation. The specificity of the aaRS receptor is expected to be quite dependent on ligand orientation, especially at the reaction center for the formation of the aminoacyl adenylate complex. The docking procedure discussed above gives the best possible binding orientation for each amino acid although this orientation may not lead to activation of PheRS. Hence, to recognize competitors to binding and activation of Phe among the natural amino acids, we considered a second series of predictions for the 19 noncognate amino acids in which we built each residue onto the Phe backbone. By this method, we determine the suitability of each competing amino acid in the binding mode necessary for activating and charging PheRS,

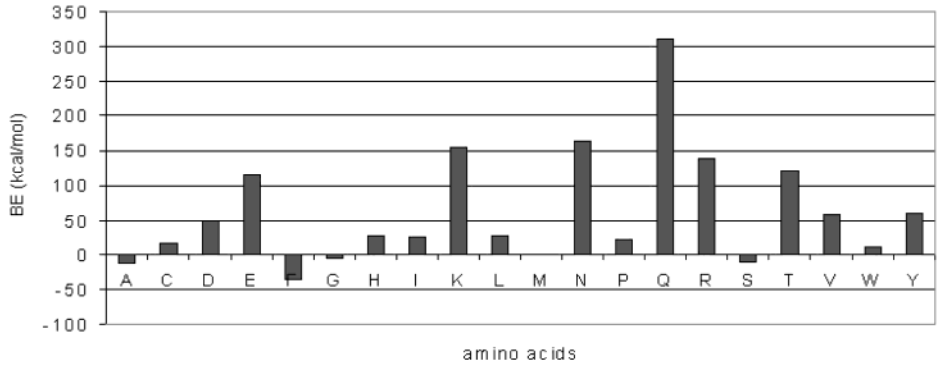

Figure 10. Binding energies of 20 amino acids using the conformations generated from perturbation of the Phe binding site in PheRS.

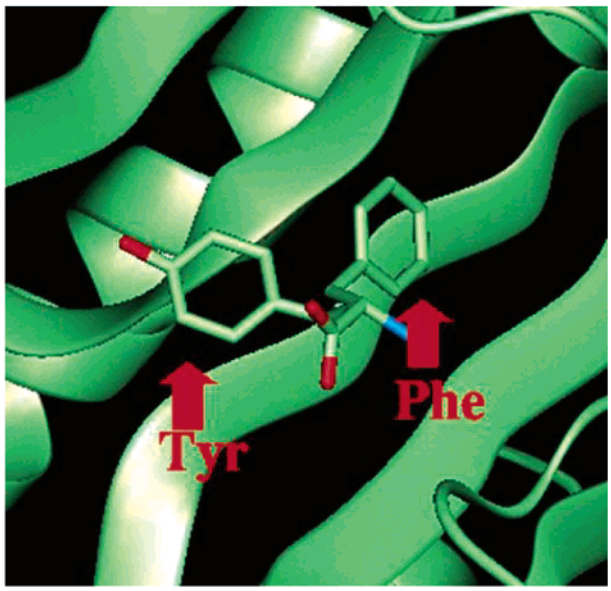

Figure 11. Perturbed Tyr after optimization superimposed onto the predicted Phe structure.

which is dominated by the ligands' interaction with the hydrophobic residues comprising the LBD. Although the noncognate amino acids may not bind in this conformation, it is useful to identify the unfavorable energies involved in binding such ligands in the metabolically active conformation of Phe. These results are summarized in Figure 10, where we find that Phe has the highest affinity for the binding site by a margin of ~20 kcal/mol. As an illustrative example, the perturbation of Phe into the predicted tyrosine site generated a highly unfavorable binding energy of +120 kcal/mol, thereby suggesting that Phe is exclusive for its predicted binding site. Furthermore, perturbation of Tyr into the Phe backbone and successive minimization led to the extrusion of Tyr from the LBD, therefore indicating the effectiveness of the hydrophobic residues in discriminating the analogues of the phenyl ring (Figure 11).

Component Analysis of the Binding Energy Contribution from Residues within the Binding Pocket. To provide insight into the nature of the binding of a ligand to a protein active site, it is useful to analyze the contributions of the neighboring residues to the ligand binding energy. To obtain this information, we calculate pairwise the summation of interactions between all atoms of the ligand and all atoms of each individual residue comprising the binding pocket. Since the valence terms (e.g., torsional strain) did not contribute significantly to the binding energy, we focus primarily on the nonbond terms.

The pairwise contributions of the Coulombic, van der Waals (vdW), and hydrogen bond terms for the residues constituting

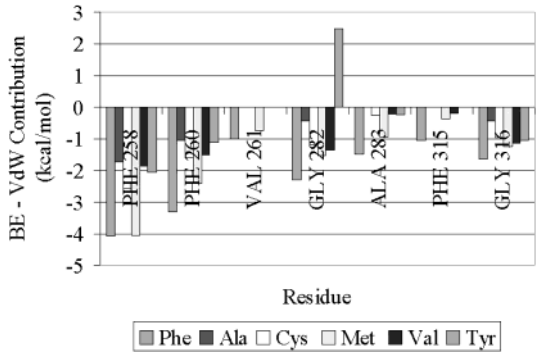

Figure 12. van der Waals contributions to the binding energy of Phe in PheRS per residue comprising the hydrophobic pocket of the PheRS binding cavity.

the PheRS binding cavity were ascertained for the above-mentioned Phe, Met, Cys, and Tyr perturbations. These data indicate that bound amino acid has the most favorable nonbond interactions with His178, Trp149, Gln218, Phe258, Phe260, Arg204, Gly282, Gly316, Ala283, Glu220, and Ser180, which contribute most of the overall binding energy. The van der Waals and electrostatic interactions are responsible for much of the nonbond stability, but hydrogen bonding involving His178, Trp149, Gln218, Arg204, and Ser180 provides further stability. Of the selected amino acids Phe yielded the most negative binding energy, with the vast majority of this stability arising from the van der Waals interactions. The electrostatic terms, although significant in the overall binding energy, varied little among the amino acids selected for this trial. The nature of this perturbation study is that all differences arise from the backbone of Phe, leading to identical hydrogen bond contributions for each case. The side chains of potentially hydrogen bonding residues, such as Tyr, did not lead to significant hydrogen bonding, since no hydrogen donors or acceptors in the binding pocket were close enough to interact. Consequently, the variations in binding energy arise primarily from the hydrophobic portion of the active site, in which the side chain of a given amino acid is stabilized. Residues Phe258, Phe260, Val261, Gly282, Ala283, Phe315, and Gly316 are the key residues constituting this binding region. The preference for Phe in this pocket is clearly demonstrated in Figure 12, with the phenyl rings of residues Phe258 and Phe260 providing much of the stabilizing effect. Methionine also appears to be stabilized to a similar degree by these phenyl rings; however, its less favorable interactions with the additional hydrophobic residues reduce its binding affinity in this region. These results suggest that Gly282 plays some role in discrimination against incorrect amino acids, most notably for tyrosine, for which we find a positive van der Waals energy.

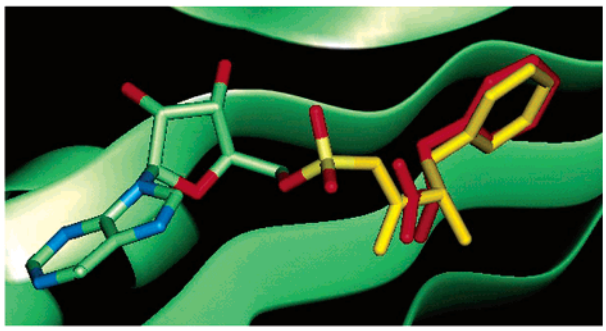

Figure 13. Predicted Phe conformation (yellow) overlaid onto crystal Phe (red), showing the relation to the location of the crystal AMP (full color).

Predicted Amino Acid Oxygen to AMP Phosphorus Distances	
$P_{AMP} - O_{aa}$	Distance (Å)
O_1 – Ala	2.1
O_2 – Ala	3.3
O_1 – Cys	2.5
O_2 – Cys	3.2
O_1 – Phe	3.1
O_2 – Phe	1.3
O_1 – Met	2.6
O_2 – Met	3.2
O_1 – Val	3.2
O_2 – Val	2.5
O_1 – Tyr	2.9
O_2 – Tyr	2.3
Expt'l	*1.6*

Figure 14. Phosphorus (AMP) to oxygen (amino acid) distances. The only distance satisfactory to form a covalent bond with the O from AMP is for the Phe structure.

aa-AMP Complex Binding. A critical step in the linkage of an amino acid with its respective tRNA is the generation of the aa-AMP complex. This requires the amino acid recognition site, or more specifically the carboxyl terminus, to be in a position conducive to reaction with the AMP phosphorus atom. To examine this, we superimposed the experimental crystallographic AMP coordinates[15] onto the best predicted amino acid/PheRS structures (Figure 13) and measured the resulting amino acid oxygen to AMP phosphorus distances (Figure 14) using SCRWL.[35] The experimental data indicate that this distance should be ~1.6 Å; thus, we find a reasonable O−P bond distance only for the predicted Phe structure (1.3 Å).

Discussion

Comparison of the Predicted Phe Structure to the Crystal Data. A decisive test of the accuracy of the HierDock protocol is whether it can locate the binding cavity and reproduce the available crystal data within experimental error. A suitable binding pocket must establish good hydrogen bonds with the recognition site for the amino acid (that is, the carboxyl and amino termini) as well as provide favorable stabilizing interactions with the side chain. After an exhaustive search of the entire molecular surface of PheRS with Phe, the greatest number of lowest energy conformations resided in region 1, thereby isolating the region as the most favorable binding site. This finding corroborates the available experimental data.

The 100 docked conformations from the preceding step were subjected to further optimization, allowing all atoms of the ligand and protein to optimize and including SGB solvation. The lowest energy structure differs from the crystal data by CRMS = 0.61 Å, which is in good agreement with the X-ray data. This demonstrates the effectiveness of the HierDock methodology, which filters out the nonnative conformations. Comparison of the predicted structure with the crystallographic structure in the binding pocket reveals that the side chain of Phe is almost identical in position. However, the amino and carboxyl termini differ by a torsion of roughly 15−20°, which may stretch some hydrogen bond distances for the Phe (Figure 3).

These results suggest that the HierDock protocol is suited for locating the binding site and the binding conformation, leading to excellent agreement with the crystal data. The scoring function in HierDock effectively selects the most suitable conformations, making it useful for predictive studies. Thus, combining coarse-grain docking with molecular mechanics optimization and solvation provides a powerful and efficient strategy for predicting the binding site and providing insight into the mechanics of ligand recognition and binding affinities. These components of the HierDock protocol should be valuable in studies for which the binding site characteristics of a protein are unknown.

Binding Mode and Energies of Other Amino Acids in PheRS. The binding energy results (Figure 4) obtained for the independent docking of all 20 amino acids to the binding site show that Phe has the best binding energy in PheRS. However, the binding energies for Cys, Met, and Tyr are similar to that for Phe within 5.0 kcal/mol. This suggests that some amino acids may compete with Phe in binding, although we find that they have a binding mode different from that of Phe. Since the proper orientation with respect to the reaction center is probably necessary for formation of the aa-AMP complex, it thus may be that the positioning of critical functional groups of the amino acid is the ultimate factor in discriminating activation of Cys, Met, and Tyr with respect to Phe. Indeed Freist et al. have measured the rate of charging the t-RNA with Phe and also the rate of formation of the aminoacyl adenylate complex for the natural amino acids in *yeast* PheRS.[27] Analysis of their experimental Michaelis constant data shows that Met is a competitor in the step of amino acid binding to PheRS, which agrees with our results in Figures 8 and 9.

We also find that Cys, like Met, achieves a favorable binding affinity but has the incorrect recognition site placement, and hence is not likely to participate in reaction with auxiliary substrates. Of these predicted competitors, we find that only tyrosine shares a conformation similar to that predicted for Phe. However, Tyr is unable to wedge into the hydrophobic pocket to the same degree as Phe, thereby leaving the side chains exposed to the solvent and the amino and carboxyl groups in poor positions for substrate interactions. Thus, to ensure binding in the proper orientation, we also calculated the binding energies of all amino acids when forced to have the same binding mode as Phe in PheRS.

Although several amino acids have good binding energies elsewhere in the protein, such binding is not expected to result in activation. Since both the binding mode and binding affinities are determinants in aaRS selectivity, each amino acid was perturbed into the predicted Phe backbone to compute their relative affinities in the activation site. This perturbation generated the most promising results, indicating that Phe was favored by over 20 kcal/mol over the competing amino acids. To avoid any bias, we also performed the perturbation of the Tyr binding mode into Phe. Though the predicted backbones of Phe and Tyr differ only slightly in position, this perturbation led to an exceedingly unstable binding energy for Phe. These cross perturbations suggest the binding mode for Phe is highly specific and confers little flexibility in the positioning of the amino acid.

To avoid biasing the results, the Phe was perturbed into the binding mode of Tyr. As expected, the hydrogen bond terms will be comparable for both amino acids. In this case, however, both electrostatic and van der Waals contributions play important roles in the stabilization of the Tyr amino acid. Namely, the predicted binding mode for Tyr gains some degree of stability via an electrostatic interaction with Glu220 of approximately −1.595 kcal/mol in comparison to −0.134 kcal/mol for Phe. Nevertheless, the van der Waals terms dominate the difference in binding energy between Phe and Tyr. Tyr clearly has a more favorable van der Waals interaction with residues Ala283, Phe260, Glu220, Phe258, and Gly282. The energy terms for residues Phe258 and Gly282 for the Phe perturbation, 7.780 and 9.013 kcal/mol, respectively, indicate a possible steric clash that would prevent a Phe ligand from binding in this mode.

Additionally, as reported in (Figure 7), the solvation energy of Phe in the PheRS binding site is preferred over that of Tyr, which would arise from the cost of desolvating the phenol group. Thus, Tyr is discriminated from the binding site presumably due to vdW clashes as well as an unfavorable solvation energy.

According to these results, it could be assumed that the predicted binding mode for Phe is exclusive for Phe and, moreover, that there is little flexibility in its preferred binding orientation. On the basis of these perturbation data, PheRS appears to distinguish between Phe and other falsely bound structures by discrimination at the actual preferred binding site of Phe, primarily through van der Waals interactions in the hydrophobic pocket. Hence, amino acid recognition gains its specificity by discriminating within the hydrophobic pocket, whereas recognition of a given amino acid by its amino and carboxyl tails is rather unspecific.

Residue analysis of the hydrophobic binding pocket provides insight into the receptor's preference for Phe. On the basis of the hydrogen bond data for Phe, the backbone appears to make good bonds with His178, Trp149, and Gln218, with compromised contacts with Arg204 and Ser180. Given that the amino acids perturbed into the predicted Phe orientation share a common backbone, the hydrogen bond contributions will be identical in each case. The contributions to the total binding energy vary most widely for the van der Waals component. Thus, the van der Waals term, especially within the hydrophobic binding pocket, serves as a discriminating factor in the distinguishing of amino acids.

Phe258, Phe260, Val261, Gly282, Ala283, Phe315, and Gly316 are the key residues comprising the hydrophobic portion of the binding site and are essential in the pocket's specificity for Phe. In this regard, the van der Waals contributions for each residue are most favorable for Phe, especially for residues Phe260, Gly282, Ala283, and Phe315. Residues Phe258, Val261, and Gly316 contribute relatively favorable van der Waals terms for methionine. Thus, it is assumed that the interaction with the remaining residues results in the exclusion of Met in the precise binding mode of Phe.

A remarkable trait of PheRS is its ability to distinguish between Phe and Tyr, which differ only by the para substitution of a hydroxyl group on the phenyl ring. In Figure 12, the disparities in van der Waals energies between Phe and Tyr are apparent, mostly for residues Phe258, Phe260, Gly282, and Ala283. An apparent clash of the van der Waals surfaces of bound Tyr and Gly282 could be the basis of tyrosine's exclusion from the PheRS binding pocket. This clash with glycine, or ultimately the protein backbone, suggests the folding of the PheRS pocket prevents the misacylation of tyrosine. It is also worth noting that, upon optimization of the perturbed Tyr structure, the phenol ring is expelled from the hydrophobic pocket (Figure 11). The most stable conformation of the tyrosine perturbation appears to be with the phenol group situated in the hydrophilic region of the pocket, versus burial in the hydrophobic region as noted for Phe. Presumably, this property could be the basis of selectivity for Phe versus Tyr.

Facility of aa-AMP Complex Formation for All Amino Acids. The positioning of the recognition sites has been stressed as an important requisite to favorable interactions with adjoining substrates. In the aaRS case, adequate hydrogen bond interactions with the bound amino acid termini should position the carboxyl group of the ligand for facile reaction with AMP. In the formation of the aa-AMP complex, the carboxyl oxygen of the amino acid attacks the innermost phosphorus atom of AMP via nucleophilic substitution.[2] This displaces a pyrophosphate group, thereby creating an oxygen−phosphorus bond which bridges the original amino acid and AMP molecules.

To establish the reliability of our results, we compared these O−P bond distances for the highest ranking ligands from the independent docking calculations. The predicted structure for Phe places the termini in the optimal position for reaction with AMP and results in an O−P bond distance that is only marginally smaller than the experimental case. For the competing amino acids, the bond lengths are in excess of that observed in the crystal, which would inhibit any nucleophilic attack on the AMP phosphorus group, or at least result in a very unstable bond.

Conclusions

We have demonstrated the HierDock procedure, which combines coarse-grain docking with fine-grain molecular dynamics with all-atom FF and solvation, to be very useful in locating the active site by searching the whole protein. For PheRS, a class II aaRS, the binding site of Phe has been predicted with a CRMS of 0.61 Å from the crystal structure after a search of the whole protein for the binding site. We predict hydrogen bond distances of Phe to Ser180, Gln 218, Glu220, Trp149, His178, and Arg204 in the binding site that are similar to the experimental data.

We have also calculated the binding energies of the 20 natural amino acids in PheRS. These energies show that Phe binds most favorably with a binding energy of −35.065 kcal/mol, while the close competitors such as Met, Cys, and Tyr do not bind in the same binding site as Phe. As an example, the Met site is 1.4 Å away from the Phe binding site and thus would fail in making the critical hydrogen bonds that would normally activate PheRS. Met indeed has been shown to be a competitor at the step of amino acid binding from in vitro rate measurements.[34] The binding sites predicted for Met, Cys, and Tyr have been shown not to facilitate the formation of AMP complexes. The HierDock procedure allows one to calculate the discrimination of aaRS at the stage of binding the amino acid and formation of the aminoacyl adenylate complex. Such procedures are very important to offer a molecular level understanding of the mistakes made in protein biosynthesis that are not readily uncovered through experiments. The accuracy of the binding site location and perturbation analysis of the HierDock protocol could prove it to be a powerful tool in cases for which the recognizable substrates for a receptor are unknown. Likewise, this hierarchical strategy could be useful in the screening of nonnatural amino acid analogues for aaRS, prior to synthesis.

Acknowledgment. The National Institutes of Health Grant 1 RO1-GM62523-01 generously supported this work.

Supporting Information Available: Atomic positions in Cartesian coordinates for the minimized amino acids and PheRS protein. This material is available free of charge via the Internet at http://pub.acs.org.

References and Notes

(1) Ibba, M.; Curnow, A. W.; Söll, D. *Trends Biochem. Sci.* **1997**, *22*, 39−42.

(2) Ibba, M.; Söll, D. *Annu. Rev. Biochem.* **2000**, *69*, 617−650.

(3) Freist, W.; S. H.; Pardowitz, I.; Cramer, F. *J. Theor. Biol.* **1998**, *193*, 19−38.

(4) Freist, W.; S. H. *Eur. J. Biochem.* **1988**, *177*, 425−433.

(5) Pallanck, L. P., M.; Schulman, L. In *In tRNA: Structure, Biosynthesis and Function*; ASM Press: Washington, DC, 1995; pp 371−393.

(6) Arnez, J.; Moras, D. *Trends Biochem. Sci.* **1997**, *22*, 211−216.

(7) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme catalysis and Protein Folding*; Freeman: New York, 1999.

(8) Fersht, A.; Shi, J.; Knill-Jones, J.; Lowe, D.; Wilkinson, A.; Blow, D.; et al. *Nature* **1985**, *314*, 235−238.

(9) Agou, F.; Quevillon, S.; Kerjan, P.; Mirande, M. *Biochemistry* **1998**, *37*, 11309−11314.

(10) Ibba, M. a. H., H. *FEBS Lett.* **1995**, *364*, 3272−3275.

(11) Hoshaka, T.; Ashizuka, Y.; Sasaki, H.; Murakami, H.; Sisidi, M. *J. Am. Chem. Soc.* **1999**, *121*, 12194−12195.

(12) Kothakota, S. Y. E.; Murphy, O. J.; Mason, T. I.; Tirrell, D. A.; Fournier, M. J. *J. Polym. Sci., Part A: Polym. Chem.* **1995**, *33*, 1267−1274.

(13) Kiick, K. L.; T. D. A. *Tetrahedron* **2000**, *56*, 9487−9493.

(14) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *20*, 5010−5011.

(15) Reshetnikova, L.; Moor, N.; Lavrik, O.; Vassylyev, D. *J. Mol. Biol.* **1997**, *287*, 555−568.

(16) Fersht, A. *Biochemistry* **1988**, *27*, 1577−1580.

(17) Floriano, W.; Vaidehi, N.; Goddard, W. A., III. Manuscript in preparation.

(18) Floriano, W.; Vaidehi, N.; Goddard, W. A., III; Singer, M.; Shepherd, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10712−10716.

(19) Vaidehi, N.; Floriano. W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **2002**.

(20) Wang, P. V., N.; Tirrell, D. A.; Goddard, W. A., III. *J. Am. Chem. Soc.* **2002**.

(21) Krishnan, R.; Binkley, J. S.; Seuger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650−654.

(22) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nichols, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875−11882.

(23) Jaguar 3.5, Schrödinger Inc., Portland, OR, 1998. Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J.-M.; Goddard, W. A., III; Donnelly, M. N.; Ringnalda, R. E. *J. Chem. Phys.* **1994**, *101*, 4028−4041.

(24) MacKerell, A. D.; Bashford, D.; Bellot, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fisdher, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Sclenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(25) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. *J. Phys. Chem. B* **1994**, *94*, 8897−8909.

(26) Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272−282.

(27) Datta, D.; Vaidehi, N.; Xu, X.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2636−2641.

(28) Ewing, T. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411−428.

(29) Ewing, T. A.; Kuntz, I. D. *J. Comput. Chem.* **1997**, *18*, 1175−1189.

(30) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.

(31) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E *J. Mol. Biol.* **1982**, *161*, 269−288.

(32) Hendsch, Z. S. T., B. *Protein Sci.* **1999**, *8*, 1381−1392.

(33) Lim, K. T.; Iotov, M.; McClurg, R. B.; Vaidehi, N.; Dasgupta, S.; Taylor, S.; Goddard, W. A., III. *J. Comput. Chem.* **1997**, *18*, 501−521.

(34) Freist, W.; Sternbach, H.; Cramer, F. *Eur. J. Biochem.* **1996**, *240*, 526−531.

(35) Bower, M.; Cohen, F.; Dunbrack, R. *J. Mol. Biol.* **1997**, *267*, 1268−1280.

(36) Freist, W.; Sternbach, H.; Cramer, F. *Eur. J. Biochem.* **1996**, *240*, 526−531.