Chapter 2

Virtual Screening for Binding of Phenylalanine Analogs to Phenylalanyl-tRNA Synthetase

Portions of this chapter are adapted from:
*Pin Wang, Nagarajan Vaidehi, David A. Tirrell and William A. Goddard III, J. Am. Soc. Chem., (**2002**), 124, 14442-14449*

Abstract

Although incorporation of non-natural amino acids provides a powerful means of controlling protein structure and function, experimental investigations of amino acid analogs for utilization by the protein biosynthetic machinery can be costly and time-consuming. In this chapter, we describe a computational protocol (HierDock) for predicting the relative energies of binding of phenylalanine analogs to phenylalanyl-tRNA synthetase (PheRS). Starting with the crystal structure of *Thermus thermophilus* PheRS without bound ligand, HierDock predicts the binding site of phenylalanine (Phe) within 1.1Å of that revealed by the crystal structure of PheRS co-crystallized with Phe. The calculated binding energies of Phe analogs in PheRS, using HierDock, correlate well with the translational activities of the same analogs in *Escherichia coli*. HierDock identifies *p*-fluorophenylalanine and 3-thienylalanine as especially good substrate for PheRS, in agreement with experiment. These results suggest that the HierDock protocol may be useful for virtual screening of amino acid analogs prior to experiment.

1. Introduction

In vivo expression systems provide the most powerful means of controlling the composition, folding, and assembly of proteins. The power of such systems would be enhanced significantly by the development of an expanded set of amino acids that exhibit good translational activity in conventional expression hosts.

Various laboratories have demonstrated the ability of the wild-type translational apparatus to incorporate non-natural amino acids containing fluorinated (*1-4*), unsaturated (*5-8*), electroactive (*9*), and other side chain functionalities (*10-13*). Nevertheless, the number of amino acids shown conclusively to exhibit translational activity *in vivo* is relatively small, and the range of chemical functionality that has been accessed by this method remains modest. The experimental techniques for preparing and testing amino acid analogs can be costly and time-consuming; consequently, we consider in this paper a computational protocol (HierDock) for screening amino acid analogs before experiment.

The recognition of the amino acid or analog by the appropriate aminoacyl-tRNA synthetase (aaRS) and attachment to the cognate tRNA constitute critical steps in the protein synthesis pathway, and manipulation of the aaRS has been shown to facilitate incorporation of non-natural amino acids into protein *in vivo* (*13-19*). To guard against incorporation of incorrect amino acids into proteins, aaRS must bind three substrates (amino acid, ATP, and cognate tRNA) with very high specificity (*20*). Bound amino acid is activated through the formation of an aminoacyl adenylate intermediate; subsequent transfer of the aminoacyl group yields the aminoacyl-tRNA (Eq. 1).

$$AARS + aa + ATP \Leftrightarrow [AARS : aa \sim AMP] + PPi$$
$$[AARS : aa \sim AMP] + tRNA^{aa} \rightarrow aa \sim tRNA^{aa} + AMP + AARS \quad \text{Eq. 1}$$

Here we assume that selective binding of the amino acid or analog to the aaRS is the step that determines the selectivity of amino acid incorporation into the growing polypeptide chain. This need not be the case: formation of the aminoacyl adenylate, attachment to tRNA, or editing by the aaRS could in principle constitute the product-determining step. However, selective binding is a necessary step, and we find herein that the calculated binding energies of a set of Phe analogs to PheRS correlate well with the translational activities of the analogs with the wild type PheRS *in vivo*. When possible, we compare the calculated difference in binding energies to the differential free energies of binding determined from experimentally measured kinetic parameters that describe the activation of amino acid substrate.

Predicting the binding site and binding energy for an amino acid or a novel amino acid analog can easily become an intractable process. There might be many possible binding sites for the analog (not just that occupied by the natural amino acid), and it is necessary to examine all such possibilities in order to determine if an analog will bind to the appropriate site. Even with the binding site specified there are many complications in predicting a binding free energy. There are often many possible configurations of the ligand and of the protein side chains, as well as uncertainty concerning solvation of the ligand and protein, and in identifying molecules and ions that might compete for the active site. In order to obtain a practical protocol for examining all possible sites, while predicting sufficiently accurate binding energies for the most favorable sites, we use the HierDock hierarchical strategy (*21-24*). HierDock starts with a coarse grain search of

conformations for the binding site, and ends with fine grain molecular dynamics (MD) optimization of the full ligand-protein complex, including solvation.

In this chapter, we apply the HierDock computational protocol to screening of amino acid analogs for binding to the corresponding aaRS. To validate this protocol we first predict the binding site of phenylalanine (Phe) in phenylalanyl-tRNA synthetase (PheRS) of *Thermus thermophilus (T. thermophilus)*. The calculated binding site of Phe in PheRS is within 1.1Å CRMS (coordinate root mean square error for all atoms) of the crystal structure. The HierDock procedure was then used to predict the binding site and to calculate the binding energies of various analogs of Phe to PheRS. We find that the calculated binding energies of Phe and its analogs correlate well with the results of experimental measurements of *in vivo* incorporation in *Escherichia coli* and of *in vitro* measurements of the differential free energy of binding. It should be noted that the binding energies were calculated for PheRS from *T. thermophilus* and compared to the experimental measurements of *in vivo* translational activities made on wild type *E. coli* PheRS. This is reasonable since the sequence identity between *E. coli* and *T. thermophilus* for PheRS is 43% with 94% identity in the binding site. The HierDock procedure also yields a predicted binding site for each of the Phe analogs, which allows assessment of the feasibility of activation of each bound substrate. We suggest that the methods described here could be used for virtual screening of amino acid analogs prior to experiment. Such virtual screening procedures could considerably speed the development of libraries of analogs likely to be incorporated into protein *in vivo*.

2. Methods

2.1. Computational Methodology

We use the HierDock procedure, which has been applied successfully to study the binding of ligands to globular (*22, 23*), and membrane-bound proteins (*21, 24, 25*). The HierDock ligand screening protocol follows a hierarchical strategy for examining conformations, binding sites and binding energies. The steps involve using coarse-grain docking methods to generate several conformations of protein/ligand complexes followed by molecular dynamics (MD) simulations including continuum solvation methods performed on a subset of good conformations generated from the coarse grain docking. Methods combining docking and MD simulations have been tested(*26*), but the main drawback of these tests was that only one protein/ligand complex structure was kept from the coarse grain docking methods for MD simulations. This is risky considering that the coarse grain methods do not have accurate scoring functions that include solvation. Free energy perturbation methods lead to accurate free energies of binding but are computationally intensive and not readily applicable to a wide variety of ligands(*27*). Our goal is to derive a fast hierarchical computational protocol that uses hierarchical conformation search methods along with different levels of scoring functions, that would allow screening of amino acid analogs for aminoacyl t-RNA synthetases. The steps in HierDock are as follows:

1) First we carry out a coarse grain docking procedure to generate a set of conformations for ligand binding. In this paper we used DOCK 4.0(*28*) to generate and score 30,000 configurations, of which 10% (3000) were ranked using the DOCK scoring function.

2) We then select the 20 best conformations for each ligand from DOCK and subject them to annealing molecular dynamics (MD) to further optimize the conformation in the

local binding pocket, allowing the atoms in both ligand and binding cavity (residues with an atom within 5Å of the binding ligand) to move. In this step the ligand and the binding cavity in the protein were heated and cooled from 50K to 600K in steps of 10K (0.05 ps at each temperature) for 5 cycles. At the end of each annealing MD cycle the best energy structure is retained. Annealing MD allows the protein cavity to readjust for interaction with the ligand. This fine-grain optimization was performed using MPSim (*29*) and a full atom forcefield (DREIDING) (*30*) and continuum solvation methods. We use the Surface Generalized Born (*31*) (SGB) continuum solvent method to obtain forces and energies resulting from the polarization of the solvent by the charges of the ligand and protein. This allows us to calculate the change in the ligand structure due to the solvent field and hence obtain accurate binding energies that take into account the solvation effects on the ligand/protein structure. For the annealing MD procedure, the charges for the ligand were derived using the charge equilibration (QEq) (*32*) method, while charges for the protein were taken from CHARMM22 (*33*). This procedure generated 5x20=100 good protein/ligand complexes for each ligand.

3) For the 100 structures generated by annealing MD simulations for each ligand, we minimized the potential energy (conjugate gradients) of the full ligand/protein complex in aqueous solution using SGB. This step of protein/ligand-complex optimization is critical to obtaining energetically good conformations for the complex (cavity + ligand). Then we calculated binding energies as the difference between the total energy of the ligand-protein complex in solvent and the sum of the total energies of the protein and the ligand separately in solvent. The energies of the protein and the ligand in solvent were calculated after independent energy minimization of the protein and the ligand separately in water. The energy calculations used the more accurate Poisson-Boltzmann (*34*) (PB)

solvation method to calculate solvation energies. The non-bond interaction energies were calculated exactly using all pair interactions.

Section 3 shows that this HierDock strategy provides a practical scheme for predicting ligand binding structures and relative binding energies of Phe and its analogs to PheRS. The calculated relative binding energies correlate well with *in vivo* incorporation results with wild type PheRS, reported previously (*2, 9, 13*).

2.2. Procedures for Screening Analogs and Application to PheRS

PheRS offers a good test case for HierDock since a range of Phe analogs has been evaluated with respect to *in vivo* translational activity with wild type PheRS. PheRS is an $\alpha_2\beta_2$ heterotetrameric enzyme. The crystal structure of the *T. thermophilus* variant without bound Phe has been reported by Mosyak and coworkers(*35*). We will denote this structure (PDB code: 1PYS) as No/PheRS, where "No" indicates the crystal structure of PheRS protein with *no ligand*. The entity before the slash in this notation indicates the ligand, after the slash is the protein. Reshetnikova *et al.* (*36*) determined the crystal structure of PheRS complexed with Phe from *T. thermophilus* at 2.7Å resolution (PDB code: 1B70). We will denote this structure as Phe/PheRS, where the underline indicates that the protein structure is that of the liganded protein. The CRMS (all atoms) between the protein part of Phe/PheRS and No/PheRS is 0.4 Å, indicating that binding the ligand does not cause significant reorganization of the protein.

To determine how well the HierDock procedure works for a system in which the binding site has not been determined, we predicted the structure of PheRS with bound Phe, starting with No/PheRS. This predicted structure is denoted as Phe/PheRS.

Although the binding site of Phe in PheRS is known, we did not use the explicit three-dimensional coordinates from the crystal structure in our simulations. This is because we wanted to validate how well our procedure would identify an unknown binding site and structure. This is a critical test for the procedure, since in many important applications one may not have access to a known crystal structure with the ligand bound.

We validated the forcefield used here by performing energy minimization on two crystal structures. We started with No/PheRS structure (shown in red in Figure 2-1) and added explicit hydrogens to all heavy atoms. To represent the solvation shell expected near (*33*) the charged residues, we neutralized the acidic residues by adding Na^+ counterions and basic residues by adding Cl^- counterions. The energy of this structure was then minimized with the conjugate gradient method while including solvation via the SGB continuum solvation model. We found a 0.8Å CRMS difference between the coordinates of all atoms in the minimized structure and the crystal structure (as shown in Figure 2-1), which is well within the resolution of the crystal structure (2.9Å). To further validate our forcefield we also performed similar minimization on the 1B70 (Phe/PheRS) structure. The CRMS difference in the coordinates of all atoms between the crystal structure and the forcefield-minimized structure is 0.7Å. The above two tests, summarized in Table 2-1A, indicate that the forcefield, charges, and solvation description are satisfactory. The notations used for the minimized and other structures of PheRS used or generated in this study, are given in Table 2-1B.

For all the docking studies we started with the *unliganded* No/PheRS(min) structure. This was to remove any bias that might be introduced by using the crystal structure with Phe bound. Since the α chain contains the binding site, we used just this

chain for the docking studies. Crystallographic water molecules were removed to allow the volume of the receptor site to be explored completely.

We then carried out a coarse grain search using "DOCK4.0". The docking site was limited to a 10Å cube (the box shown in Figure 2-2) in the region of the binding site for Phe. However, we made no use of the coordinates from the crystal structure with Phe bound. The DOCK search was done with the following controls:

1) Mapping possible binding regions: The negative image of the receptor molecular surface (using the Connolly method (*37*)) was filled with a set of overlapping spheres. A probe of 1.4 Å radius was used to generate a molecular surface. These spheres represent potential ligand binding sites. Sphere clusters were generated for the whole binding site using the program "Sphgen" from the DOCK suite of programs.

2) Defining regions for docking: The sphere-filled volume representing the void space or the potential binding region in the protein around the binding site was chosen from the description of the binding site(*36*) as shown in Figure 2-2.

3) Generating docked conformations of the receptor-ligand complexes: Starting with an arbitrary conformation of Phe, several orientations of Phe within the receptor were generated using DOCK 4.0. Here we used flexible docking with torsion minimization of ligands, a non-distance dependent dielectric constant of 1, and a cutoff of 10 Å for energy evaluation.

Using DOCK 4.0 to dock Phe to the No/PheRS(min) structure did not give properly bound structures since most of the structures generated were outside the binding site. This was because the binding cavity in No/PheRS(min) was too narrow for Phe to bind. To relax the binding cavity in No/PheRS(min) we docked Phe with reduced van

der Waals radii for the Phe atoms (by 50% of the original DREIDING radii) while keeping the van der Waals radii of the protein atoms unchanged. Phe was then found in the binding site. Using the scoring function from DOCK4.0 we then selected the best structure, and changed the vdW parameters back to the original values of the Phe atoms. Subsequently the best docked conformation of the protein/ligand complex was optimized using constant temperature and constant volume MD simulations on the protein cavity at 300K for 100ps. This procedure relaxed the binding site of Phe in the No/PheRS(min) structure. We denote this structure as No/PheRS(dyn), implying that PheRS structure is the one after 100ps of MD simulations. The CRMS difference of the binding cavity before and after MD is 2.1 Å, indicating rearrangement to accommodate the ligand. Figure 2-3 shows the relaxed binding cavity (No/PheRS(dyn) in yellow compared to the No/PheRS(min) in red. Figure 2-3 also compares No/PheRS(dyn) to Phe/PheRS(min) in green. The overall CRMS difference between these structures is only 1.1Å which is within the resolution of the crystal structure, 2.5Å. Such a relaxation procedure is applicable to proteins that do not undergo major conformational changes on ligand binding. Using the relaxed cavity predicted for No/PheRS(dyn), we then redocked Phe and each of the seven analogs shown in Scheme 1, using the same set of DOCK parameters. The analog structures were built using PolyGraf (MSI San Diego) and optimized in solution (SGB) using the DREIDING FF (*30*) and charges from charge equilibration (QEq)(*32*). The best 10-30 configurations for each analog in the binding region were selected using the energy scores from DOCK4.0 and used as input for the subsequent annealing MD step of HierDock. To select these best configurations, we allowed nonbond interactions from all atoms within 20Å of the binding region.

Next, we carried out the fine grain HierDock procedure to select the optimum configuration for binding each ligand. Thus the binding site was determined by considering the 100 best scoring structures of Phe in PheRS after MD annealing and optimizing the structures with minimization. This procedure was repeated to determine the binding site and binding energies of the seven Phe analogs shown in Scheme 1. As described below the calculated binding energies correlate well with the experimental *in vivo* results on incorporation of these analogs into recombinant proteins. Further more, we compared the calculated binding energies with the relative free energies of binding estimated from the kinetics of pyrophosphate exchange

2.3. Measurement of Relative Free Energies

The PheRS gene was cloned directly from *E. coli* genomic DNA with flanking primers encoded the restriction sites *SacI* and *HindIII* (primer 1: 5'-CAC CAC TGA CAC AAT GAG CTC AAC CAT GTC ACA TCT CG-3'; primer 2: 5'-CAT ATG GCT AGC AAG CTT CAT AGG TTC AAT CCC-3'). The resulting 3500 base-pair DNA fragment was gel-purified, digested with *SacI* and *HindIII*, and ligated into the expression plasmid pQE30 (Qiagen) to yield pQE-pheST, which encodes both the α and β subunits of *E. coli* PheRS. The integrity of the cloned gene was confirmed by DNA sequencing. The cloned enzymes contained the N-terminal leader sequence MRGSHHHHHHTDPHASST. pQE-pheST was transformed into XL-1 (Stratagene) to yield the expression strain. Protein expression was induced at OD_{600}=0.6 with 1 mM IPTG. After three hours, the cells were harvested. The enzyme was purified using Ni-NTA agarose resin under native conditions according to the manufacturer's instructions

(Qiagen), and protein was stored in Buffer A (50 mM Tris-HCl, 1 mM DTT)/50% glycerol. Aliquots were flash frozen and stored at –80 ^{o}C. The concentration of the purified enzyme was determined by absorbance at 280 nm under denaturing conditions.

Measurement of relative free energy of binding was performed by determining the kinetics of ATP-PP_i exchange assay (*38*). The assay buffer contained 30 mM HEPES, pH 7.4, 10 $MgCl_2$, 1 mM DTT, 2 mM ATP and 2 mM [^{32}P]-PP_i (0.5 TBq/mol). The enzyme concentration was 100 nM. The amino acid concentration varied depending on the activity of enzyme toward the substrate (Phe: 10-500 μM; others: 30 μM –10 mM). Aliquots (20 μL) of reaction mixture were quenched into 500 μL quench buffer (200 mM PP_i, 7% w/v $HClO_4$ and 3% activated charcoal). The charcoal was washed twice with wash buffer (10 mM PP_i, 0.5% $HClO_4$) and counted. The data reported here was the average value from duplicated experiments. The kinetic parameters were obtained by nonlinear regression fitting analysis according to Michaeli-Menten kinetics model.

3. Results and Discussion

3.1. Prediction of the Binding Site of Phe in PheRS

Figure 2-4 shows the predicted bound structure (Phe/PheRS) starting from the No/PheRS(min) structure. The substrate Phe forms seven H-bonds to the protein, the residues Trp149, Arg204, His178, Gln218, Ser180, Glu220, and His178. It is important to note that: 1) Gln218 and Glu220 make two hydrogen bonds with the amino group of Phe; 2) The carboxyl oxygen of Phe hydrogen bonds with the backbone amide group of Trp149; 3) The side chains of Phe258, Phe260, Val261, Gly282, Ala283, Gly284,

Phe315 and Gly316 form a hydrophobic pocket that sandwiches the phenyl ring of the ligand.

Table 2-2A lists the distances of the hydrogen bonds to Phe in Phe/PheRS. These hydrogen bond distances agree well with the Phe/PheRS(min) structure. These distances also agree well with the hydrogen bonds in the Phe/PheRS (pdb code: 1B70) crystal structure, column 5 in Table 2-2. The calculated hydrogen bonds in Phe/PheRS are typically shorter than in the Phe/PheRS structure, probably because the vibrations at finite temperature lead to an expansion in the anharmonic hydrogen bonds. Table 2-2B also shows the residues in the Phe/PheRS structure that make van der Waals contact with Phe.

Comparison of the predicted structure (Phe/PheRS) to the experimental Phe/PheRS(min) structure shows a CRMS difference of 1.1Å. The CRMS for the heavy atoms of the bound Phe (ligand only) between the predicted structure Phe/PheRS and Phe/PheRS(min) is 0.7Å. This excellent agreement with the crystal structure gives confidence in predicting the structures for Phe analogs bound to PheRS.

3.2. Calculation of Binding Energies for Phe Analogs

The ligands shown in Scheme 1 were docked in the same region as Phe and optimized using the same HierDock procedure. The calculated binding energies are shown in Table 2-3 and Figure 2-5. The analogs to the left of the line in Figure 2-5 are observed from *in vivo* experiments with conventional *E. coli* strains to be incorporated into recombinant protein. Analogs 2,4,6-trifluorophenylalanine and 3-pyrrolylalanine in Scheme 1 have not been tested for *in vivo* translational activity. Histidine and the other

Phe analogs from Scheme 1 are incorporated into protein in conventional expression strains. *p*-chloro-phenylalanine and *p*-bromo-phenylalanine have been shown to support protein synthesis in an *E. coli* strain outfitted with a mutant form of PheRS that exhibits relaxed substrate specificity (*13, 16*). The binding energies of the analogs to the left of the line in Figure 2-5 thus correlate well *in vivo* translational activity. The binding energies for these analogs are calculated to be larger than those of the analogs that are not incorporated experimentally. Most importantly, *p*-fluorophenylalanine and 3-thienylalanine, both of which have been shown experimentally to incorporate into recombinant proteins *in vivo,* are predicted to exhibit the highest binding energies. It is noteworthy that histidine, a natural amino acid, shows unfavorable binding energy compared to Phe. This correlation with experimental results suggests that this virtual screening procedure may be useful for screening analogs for other aaRS.

Table 2-3 analyzes the contributions to the binding energy from the Coulomb, van der Waals, and solvation terms. We see that desolvation and Coulomb interactions favor *p*-fluorophenylalanine, *p*-chlorophenylalanine, and *p*-bromophenylalanine. On the other hand, van der Waals interactions favor *p*-fluorophenylalanine, Phe, and 3-thienylalanine but are not as favorable for *p*-chlorophenylalanine or the bulky *p*-bromophenylalanine. Figure 2-6 shows the van der Waals surface (as dotted pink spheres) of the bromine atom in *p*-bromophenylalanine in its binding cavity as predicted in the present study. We see that the van der Waals surface of bromine clashes with that of methyl group of the Ala314 (green dots in Figure 2-6). This is consistent with the binding energy values, which show that the van der Waals interactions in the hydrophobic pocket of the binding cavity distinguish the various analogs. Differential solvation is also critical to these

predictions indicating that it is important to obtain fast and accurate predictions of solvation (for both structure optimization and scoring).

Table 2-3 indicates that Phe has a binding energy (computational) 3.79 kcal/mol more favorable than *p*-fluorophenylalanine and 9.40 kcal/mol better than that of 3-thienylalanine. In order to test these predictions, the experimental kinetics of pyrophosphate exchange were analyzed to yield the differential free energy of binding via Eq. 2

$$K_m^a / K_m^b = \exp[\Delta\Delta G/(RT)] \qquad \text{Eq. 2}$$

in which the superscript a refers to the cognate amino acid, and the superscript b refers to the amino acid analog. The quantity ΔΔG is the difference in the free energies of binding of the analog (ΔG^b) and the cognate amino acid (ΔG^a), i.e. $\Delta\Delta G = \Delta G^b - \Delta G^a$.(*39*) The kinetic analysis yields differential binding energies considerably smaller than those predicted computationally; *p*-fluorophenylalanine and 3-thienylalanine bind just 0.98 and 1.09 kcal/mol, respectively, more weakly than phenylalanine. Part of this difference arises because the energies from the theory are for minimized structures (0 K) while the experimental results are for ~300 K. Molecular dynamics studies at 300 K would lead to wide excursions of the molecules over the binding site, leading to less differentiation between the energies of the various bound states. Thus while the qualitative conclusions drawn from the computational work are of substantial value, quantitative prediction of differential binding energies will require further refinement of these computational methods.

3.3. Predicted Binding Sites for *p*-Fluorophenylalanine and 3-Thienylalanine

Comparison of the best predicted structures of Phe/PheRS with those of *p*-fluorophenylalanine and 3-thienylalanine in their respective binding sites shows that the overall CRMS difference between the binding site of *p*-fluorophenylalanine and Phe/PheRS is 0.4Å, identical to that for 3-thienylalanine. However the CRMS difference between the binding sites of 3-thienylalanine and *p*-fluorophenylalanine is 0.1Å, much less than the CRMS between Phe/PheRS and the corresponding binding sites of *p*-fluorophenylalanine and 3-thienylalanine. Tables 2-4A and 2-4B respectively, list all of the hydrogen bonds and van der Waals contacts that *p*-fluoro-phenylalanine and 3-thienylalanine make in their respective binding cavities. Figure 2-7 compares the binding pockets of *p*-fluorophenylalanine and 3-thienylalanine analogs to Phe/PheRS. It is seen from List 2 that the side chain of 3-thienylalanine makes a more favorable van der Waals contact to Phe260 (3.8Å) compared to *p*-fluorophenylalanine (4.3Å) with Phe260. However *p*-fluorophenylalanine makes close van der Waals contact with Ala314 (3.4Å) and Val261 (3.4Å) compared to Phe/PheRS. It is also seen that the amino terminus of 3-thienylalanine is farther from Arg204 and His178 compared to *p*-fluorophenylalanine. This could be due to the larger size of fluorine atom, which moves the *p*-fluorophenylalanine more toward Arg204 and His178.

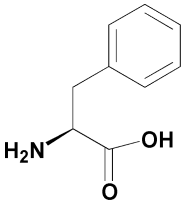
4. Conclusion

The HierDock protocol predicts the binding site for Phe in PheRS to within 0.7Å CRMS of the crystal structure, suggesting that it might correctly predict the binding site for Phe analogs in PheRS. Using this procedure we predicted the binding sites and

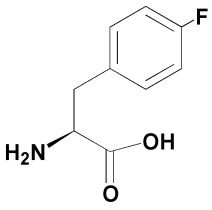
binding energies of seven Phe analogs. The two predicted to have the most favorable binding energies, *p*-fluorophenylalanine and 3-thienylalanine, are the only analogs that have been incorporated into recombinant proteins in an E. coli host harboring an unmodified PheRS. These results suggest that the HierDock protocol may be useful as a virtual screening tool for designing non-natural amino acid analogs for protein engineering. Further refinement will be required for accurate prediction of differential binding energies of amino acid substrates.

Scheme 1:

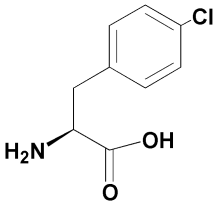
Phenylalanine and analogs: phenylalanine (**Phe**), *p*-fluorophenylalanine (**4Fphe**), *p*-chlorophenylalanine (**4Clphe**), *p*-bromophenylalanine (**4Brphe**), 2,4,6-trifluorophenylalanine (**TFphe**), 3-thienylalanine (**3TA**), 3-pyrrolylalanine (**3PA**), histidine (**His**).


H2N
OH
O

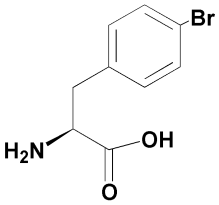
1


F
H2N
OH
O

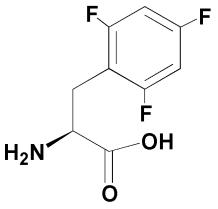
2


Cl
H2N
OH
O

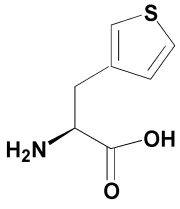
3


Br
H2N
OH
O

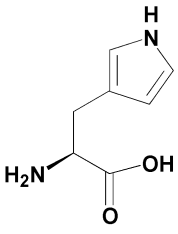
4


F
F
F
H2N
OH
O

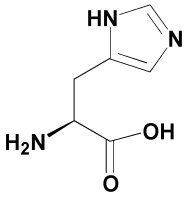
5


S
H2N
OH
O

6


H
N
H2N
OH
O

7


HN
N
H2N
OH
O

8

Table 2-1A. Comparison of Various Predicted and Experimental Structures for PheRS

System	CRMS (Å)
1PYS and 1PYS minimized	0.8
1B70 and 1B70 minimized	0.7
Predicted Phe/PheRS with 1B70 crystal structure	1.1

Table 2-1B. Notation Used to Denote the Different Ligand Protein Complex in This Paper. The Entity before the Slash in This Notation Indicates the Ligand and after the Slash is the Protein.

System	Notation
T. Thermophilus PheRS crystal structure with no bound Phe (pdb: 1PYS)	No/PheRS
Energy minimized (min) structure starting with the No/PheRS structure	No/PheRS(min)
No/PheRS(min) structure after MD simulations at 300K	No/PheRS(dyn)
Crystal structure of PheRS (*T. Thermophilus*) with bound Phe (pdb: 1B70)	Phe/PheRS
Forcefield energy minimized structure of Phe/PheRS	Phe/PheRS(min)
Predicted structure of Phe bound to PheRS starting from No/PheRS(dyn)	Phe/PheRS

Table 2-2A. Hydrogen Bond Distances and vdW Interactions in the Phe Binding Sites of PheRS/Phe Complexes

Substrate Phe	Hydrogen Bond Distance in Å			
	Residue atom	Phe/PheRS	Phe/PheRS(min)	Phe/PheRS[31]
Phe-O	Trp149 Nε1	2.97	2.97	3.32
Phe-O	Arg204 Nη1	3.07	2.87	3.28
Phe-O	Gln218 Nε2	2.93	3.02	3.46
Phe-O	His178 Nδ1	2.93	3.40	3.50
Phe-N	Glu220 Oε2	4.49	4.32	3.28
Phe-N	Gln218 Oε1	2.99	3.08	3.31

Table 2-2B. All Residues in van der Waals Contact (within 5.0Å) of Phe in the Phe/PheRS Structure.

VdW interaction	Glu220	Phe258	Phe260	Val261	Gly282	Ala283	Gly284	Ala314	Phe315	Gly316

Table 2-3. Binding Energy and Its Components for Phe Analogs Calculated from HierDock. The three cases marked with * have incorporated experimentally *in vivo* through the agency of the wild-type PheRS. (Phe: Phenylalanine; 4Fphe: *p*-fluorophenylalanine; 4Clphe: *p*-chloro-phenylalanine; 4Brphe: p-bromo-phenylalanine; TFphe: 2,4,6-trifluorophenylalanine; 3TA: 3-thienylalanine; 3PA: 3-pyrrolylalanine; His: histidine.)

	Phe	4Fphe*	3TA*	4Brphe	4Clphe	3PA	TFphe	His
ΔG(Kcal/mol)	-46.66	-42.87	-37.26	-35.82	-34.48	-31.81	-30.19	-30.97
ΔΔG(Kcal/mol)		3.79	9.40	10.84	12.18	14.85	16.47	15.69
ΔG(sol)(Kcal/mol)	-10.9	-1.13	1.25	0.31	0.04	7.79	6.65	-0.37
ΔG(coul)(Kcal/mol)	-2.84	-7.94	-3.15	-5.87	-6.75	-9.33	-5.44	-3.15
ΔG(vdW)(Kcal/mol)	-36.46	-34.54	-35.02	-26.98	-29.67	-31.99	-31.22	-28.86
ΔΔG(Kcal/mol)(exp)		0.98	1.09	NA	NA	NA	NA	NA
K_m (μM)(exp)	28±5[a)]	148±31	176±30					
k_{cat} (1/s)(exp)	0.14±0.01[a)]	0.02±0.002	0.09±0.007					

NA: Analogs not activated by wild type *E. coli* PheRS.

a): The K_m values reported here are in decent agreement with previously reported value, although k_{cat} values appear to be quite different, presumably due to the different methods of measuring concentrations of the enzyme.

Table 2-4A. Hydrogen Bond Distances of Phe, *p*-Fluoro-phenylalanine (4Fphe) and 3-Thienylalanine (3TA) Analogs in Their Respective Binding Sites.

Protein residue (atom)	Hydrogen bond distance (Å)		
	Phe	4Fphe	3TA
Arg204($N^{\eta 1}$)	3.07 (O -main)	3.64 (O-main)	4.03 (O-main)
Gln218($N^{\varepsilon 2}$)	2.99 (O-main)	2.89 (O-main)	2.93 (O-main)
His178(($N^{\delta 1}$)	2.93 (O-main)	2.87 (O-main)	4.10 (O-main)
Trp149($N^{\varepsilon 1}$)	2.97 (O-main)	2.95 (O-main)	3.02 (O-main)
Glu220($O^{\varepsilon 2}$)	3.04 (N-main)	4.86 (N-main)	3.88 (N-main)
Gln218($O^{\varepsilon 1}$)	2.99 (N-main)	2.93 (N-main)	3.60 (N-main)
Ser180(O^{γ})	4.18 (N-main)	4.11 (O-main)	5.78 (N-main)

Table 2-4B. List of van der Waals Contacts of the Side Chains of *p*-Fluoro-phenylalanine and 3-Thienylalanine, with the Side Chains of Residues in Their Respective Binding Sites. All vdW contacts within 5Å are shown. The residues with side chains within 3.2Å to 3.9Å of the ligand are marked with *.

4Fphe	Phe258*	Phe260	Val261*	Gly282	Ala283	Gly284	Ala314*	Phe315	Gly316*
3TA	Phe258	Phe260*	Val261*	Gly282	Ala283	Gly284	Ala314*	Phe315	Gly316*

Figure 2-1

Comparison of crystal structure without ligand (No/PheRS, 1pys) in red with the forcefield minimized structure in yellow. CRMS = 0.8Å.

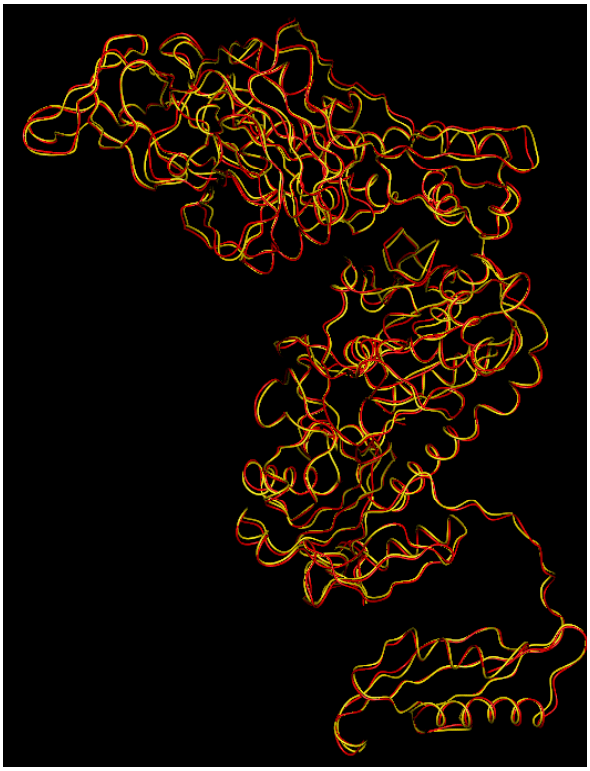


Figure 2-2

α-subunit of *T. thermophilus* PheRS from the crystal structure (No/PheRS, 1PYS). The dots show the potential binding sites and the box shows the binding region used for HierDock.

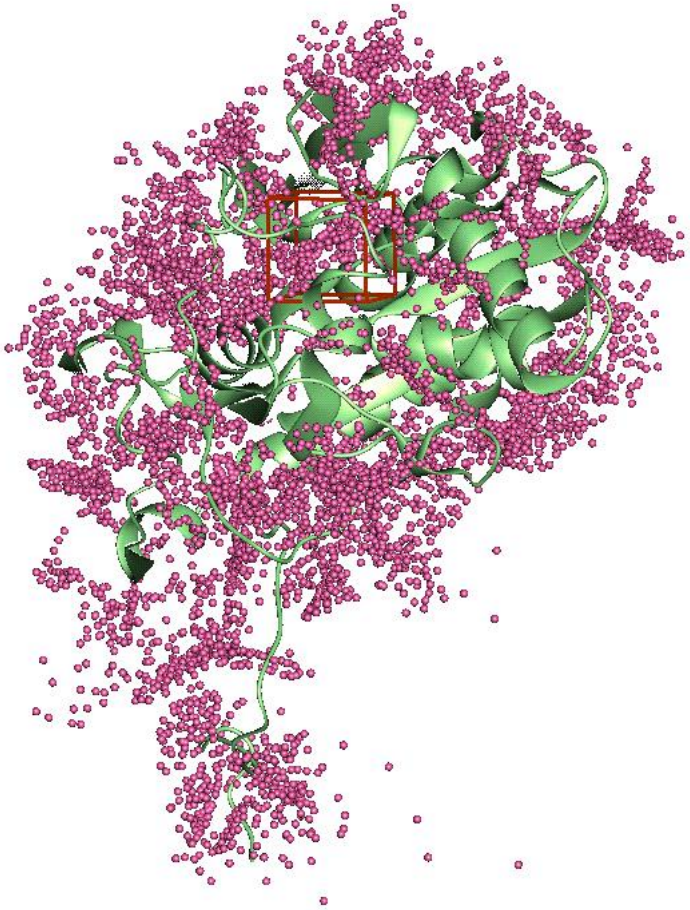


Figure 2-3

Comparison of binding pocket for PheRS from (a) No/PheRS, ligand-free crystal structure (red) of (b) minimized structure from No/PheRS (yellow), (c) the Phe/ PheRS crystal structure (1B70). (green).

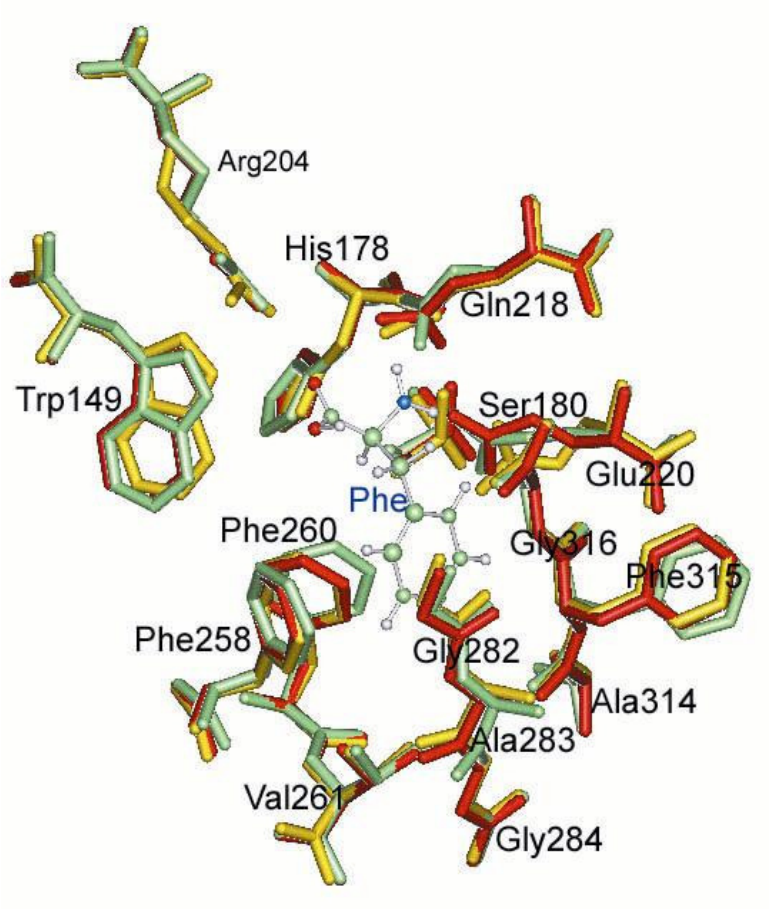

Arg204
His178
Gln218
Trp149
Ser180
Glu220
Phe
Phe260
Gly316
Phe315
Phe258
Gly282
Ala314
Ala283
Val261
Gly284

Figure 2-4

Predicted binding site of Phe in PheRS from applying HierDock to No/ PheRS. The predicted position of the Phe is shown as pink sticks.

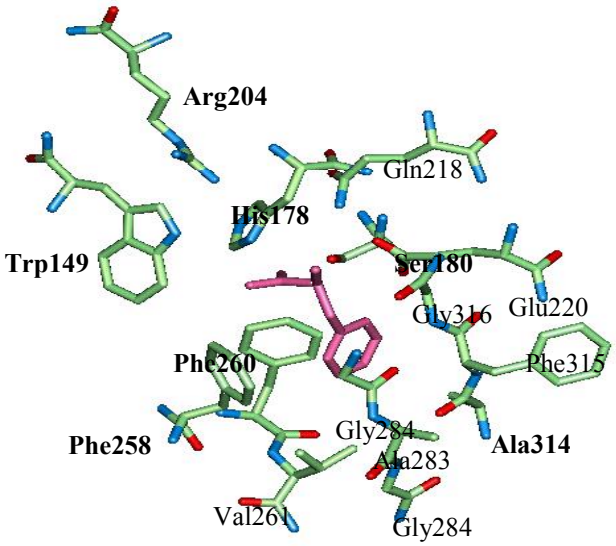

Arg204
Gln218
His178
Trp149
Ser180
Gly316
Glu220
Phe260
Phe315
Gly284
Phe258
Ala314
Ala283
Val261
Gly284

Figure 2-5

Calculated binding energies for analogs of Phe in PheRS. The analogs are shown in Scheme 1. The analogs to the left of the broken line have been observed experimentally to serve as phenylalanine surrogates in wild-type *E. coli* cells; those to the right have not. The abbreviations in the figure for the Phe analogs are: phenylalanine (**Phe**), *p*-fluoro-phenylalanine (**4Fphe**), *p*-chloro-phenylalanine (**4Clphe**), *p*-bromo-phenylalanine (**4Brphe**), 2,4,6-trifluoro-phenylalanine (**TFphe**), 3-thienylalanine (**3TA**), 3-pyrrolylalanine (**3PA**), histidine (**His**).

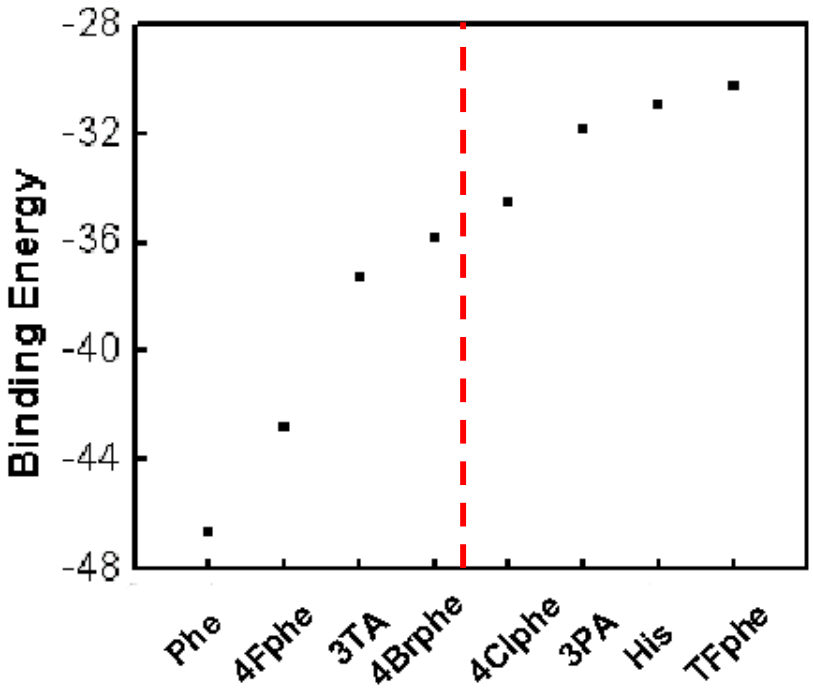

Binding Energy
-28
-32
-36
-40
-44
-48
Phe
4Fphe
3TA
4Brphe
4Clphe
3PA
His
TFphe

Figure 2-6

The van der Waals surface (bromine atom surface shown in pink) of *p*-bromo-phenylalanine clashing with the vdW surface of the side chain of Ala 314.

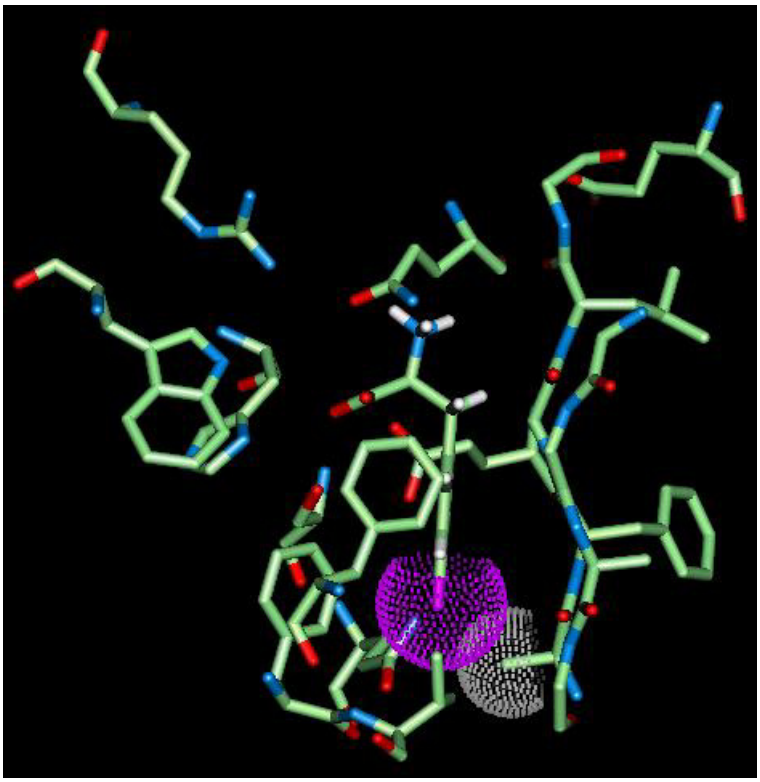
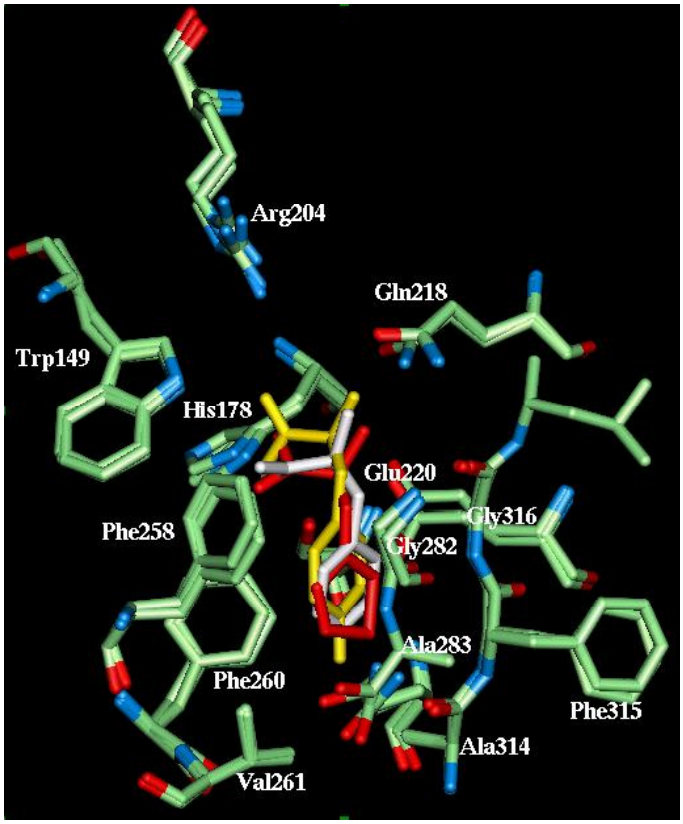


Figure 2-7

Comparison of binding pocket for phenylalanine (white) to that of *p*-fluorophenylalanine (yellow) and 3-thienylalanine (red).


Arg204
Gln218
Trp149
His178
Glu220
Gly316
Phe258
Gly282
Ala283
Phe260
Phe315
Ala314
Val261

5. Referencs and Notes

1. Richmond, M. H. (1963) *J. Mol. Biol. 6*, 284-294.

2. Yoshikawa, E., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1994) *Macromolecules 27*, 5471-5475.

3. Duewel, H., Daub, E., Robinson, V., and Honek, J. F. (1997) *Biochemistry 36*, 3404-3416s.

4. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angew. Chem., Int. Ed. Engl. 40*, 1494.

5. Kiick, K. L., and Tirrell, D. A. (2000) *Tetrahedron 56*, 9487-9493.

6. Deming, T. J., Fournier, M. J., Mason, T. L., and Tirrell, D. A. (1997) *J. Macromol. Sci. Pure Appl. Chem. A34*, 2143-2150.

7. van Hest, J. C. M., Kiick, K. L., and Tirrell, D. A. (2000) *J. Am. Chem. Soc. 122*, 1282-1288.

8. van Hest, J. C. M., and Tirrell, D. A. (1998) *FEBS Lett. 428*, 68-70.

9. Kothakota, S., Mason, T. L., Tirrell, D. A., and Fournier, M. J. (1995) *J. Am. Chem. Soc. 117*, 536-537.

10. Budisa, N., Steipe, B., Demange, P. Eckerskorn, C., Kellermann, J., and Huber, R. (1995) *Eur. J. Biochem. 230*, 788-796.

11. Cowie, D. B., and Cohen, G. N. (1957) *Biochim. Biophys. Acta 26*, 252-261.

12. Dougherty, M. J., Kothakota, S., Mason, T. L., Tirrell, D. A., and Fournier, M. J. (1993) *Macromolecules 26*, 1779-1781.

13. Sharma, N., Furter, R., Kast, P., and Tirrell, D. A. (2000) *FEBS Lett. 467*, 37-40.

14. Wang, L., Brock, A., Herberich, B. and Schultz, P. G. (2001) *Science 292*, 498.

15. Doring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crecy-Lagard, V., Schimmel, P. and Marliere, P. (2001) *Science 292*, 501-504.

16. Ibba, M., Kast, P. and Hennecke, H. (1994) *Biochemistry 33*, 7107-7112.

17. Kast, P., and Hennecke, H. (1991) *J. Mol. Biol. 222*, 99-124.

18. Ibba, M., and Hennecke, H. (1995) *FEBS Lett. 364*, 272-275.

19. Kiick, K. L., van Hest, J.C.M., and Tirrell, D.A. (2000) *Angew. Chem. Int. Ed. Engl. 39*, 2148-2152.

20. Arnez, J. G., and Moras, D. (1997) *Trends Biochem. Sci. 22*, 211-216.

21. Datta, D., Vaidehi, N., Floriano, W. B., Kim, K. S., Prasadarao, N. V., and Goddard III, W. A. (2002) *Proteins 50,* 213-220.

22. Datta, D., Vaidehi, N., Xu, X., and Goddard III, W. A. (2002) *Proc. Natl. Acad. Sci. USA 99*, 2636-2641.

23. Zhang, D., Vaidehi, N., Goddard III, W. A., Danzer, J. F., and Debe, D. (2002) *Proc. Natl. Acad. Sci. USA 99*, 6579-6584.

24. Vaidehi, N., Floriano, W. B., Trabanino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and Goddard III, W. A. (2002) *Proc. Natl. Acad. Sci. USA 99,* 12622-12627.

25. Floriano, W. B., Vaidehi, N., Goddard, W. A., Singer, M. S. and Shepherd, G. M. (2000) *P. Natl. Acad. Sci. USA 97*, 10712-10716.

26. Wang, J. M., Morin, P., Wang, W., and Kollman, P. A. (2001) *J. Am. Chem. Soc. 123*, 5221-5230s.

27. Kollman, P. A. (1993) *Chem. Rev. 93*, 2395-2417.

28. Ewing, T. A., and Kuntz, I.D. (1997) *J. Comput. Chem. 18*, 1175-1189.

29. Lim, K.-T., Brunett, S., Iotov, M., McClurg, R.B., Vaidehi, N., Dasgupta, S., Taylor, S., and Goddard III, W.A. (1997) *J. Comput. Chem. 18*, 501.

30. Mayo, S. L., Olafson, B.D., and Goddard III, W.A. (1990) *J. Phys. Chem. 94*.

31. Ghosh, A., Rapp, C.S., and Friesner, R.A. (1998) *J. Phys. Chem B 102*, 10983-10990.

32. Rappe, A. K., and Goddard III, W. A. (1991) *J. Phys. Chem 95*, 3358.

33. MacKerell, A. D., Bashford, D. et. al. (1998) *J. Phys. Chem. B 102*, 3586-3616.

34. Tannor, D. J., Marten, B., Murphy, R., Friesner, R.A., Sitkoff, D., Nicholls, A., Ringnalda, M.,. Goddard III, W.A., and Honig, B. (1994) *J. Am. Chem. Soc. 116*, 11875.

35. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M., and Safro, M. G.,. (1995) *Nat. Struc. Biol. 2*, 537.

36. Reshetnikova, L., Moor, N., Lavrik, O., and Vassylyev, D. G. (1999) *J. Mol. Biol. 287*, 555-568.

37. Connolly, M. L. (1983) *J. Appl. Crystallogr. 16*, 548-558.

38. Hendrickson, T. L., Nomanbhoy, T. K., and Schimmel, P. (2000) *Biochemistry 39*, 8180-8186.

39. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman and Co., New York. Fersht has pointed out that use of the ratio of Michaelis constants to calculate relative binding energies neglects the contribution of binding to catalysis. He recommends instead use of the ratio of specificity constants, $(k_{cat}/K_m)_a/(k_{cat}/K_m)_b$. However, the energies obtained from the HierDock method refer only to the binding event and do not take into account changes in the activation energy of the catalytic step. If we apply the Fersht treatment to the data in Table 2-3, the values of $\Delta\Delta G$ are 1.35 and 2.13 kcal/mole for 3-thienylalanine and *p*-fluorophenylalanine, respectively. The latter value is in excellent agreement with that reported by Gabius et al. (Gabius, H. J.; von der Haar, F.; and Cramer, F. *Biochemistry* **1983**, *22*, 2331-2339)