Protein Engineering, Design & Selection vol. 19 no. 5 pp. 195–203, 2006
Published online March 3, 2006 doi:10.1093/protein/gzl001

Fidelity of seryl-tRNA synthetase to binding of natural amino acids from HierDock first principles computations

Christopher L.McClendon[1], Nagarajan Vaidehi, Victor Wai Tak Kam, Deqiang Zhang[2] and William A.Goddard III[3]

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

[1]Present address: Graduate Group in Biophysics, University of California San Francisco, San Francisco, CA, USA

[2]Present address: Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, USA

[3]To whom correspondence should be addressed.
E-mail: wag@wag.caltech.edu

Seryl-tRNA synthetase (SerRS) charges serine to tRNA^{Ser} following the formation of a seryl adenylate intermediate, but the extent to which other non-cognate amino acids compete with serine to bind to SerRS or for the formation of the activated seryl adenylate intermediate is not known. To examine the mechanism of discrimination against non-cognate amino acids, we calculated the relative binding energies of the 20 natural amino acids to SerRS. Starting with the crystal structure of SerRS from *Thermus thermophilus* with seryl adenylate bound, we used the HierDock and SCREAM (Side-Chain Rotamer Energy Analysis Method) computational methods to predict the binding conformation and binding energy of each of the 20 natural amino acids in the binding site in the best-binding mode and the activating mode. The ordering of the calculated binding energies in the activated mode agrees with kinetic measurements in *yeast* SerRS that threonine *will compete* with serine for formation of the activated intermediate while alanine and glycine *will not compete* significantly. In addition, we predict that asparagine *will compete* with serine for formation of the activated intermediate. Experiments to check the accuracy of this prediction would be useful in further validating the use of HierDock and SCREAM for designing novel amino acids to incorporate into proteins and for determining mutations in aminoacyl-tRNA synthetase design to facilitate the incorporation of amino acid analogs into proteins.

Keywords: aminoacyl-tRNA synthetase/fidelity of protein synthesis/HierDock/SCREAM/seryl-tRNA synthetase

Introduction

Protein engineering using non-natural amino acids with *in vivo* expression systems promises to enable many exciting applications ranging from new biomaterials to therapeutics (van Hest and Tirrell, 2001). However, it is essential to design the non-natural amino acid analogs and the mutant aminoacyl-tRNA synthetases (aaRSs) that incorporate these analogs in the growing biopolymer (Wang and Schultz, 2002; Link *et al*., 2003), such that none of the 20 natural amino acids would compete significantly with the non-natural analog for incorporation. However, even for the 20 wild-type aaRSs, the experimental rate of misincorporation of the non-cognate amino acids is known only for a few cases (Freist *et al*., 1998). Thus, it would be useful to develop reliable computational tools to predict the likelihood of the natural amino acids that compete to bind and aminoacylation by wild-type or mutant aaRSs. We have previously proposed the use of the HierDock first principles computational method for this purpose (Wang *et al*., 2002; Kekenes-Huskey *et al*., 2003; Datta *et al*., 2004). Here, we have applied HierDock2.0 to study potential misincorporation of the 19 non-cognate amino acids by seryl-tRNA synthetase (SerRS). This is a good test since there is some experimental information available to compare with our results.

The biochemistry of the wild-type aaRSs is well known, but many questions remain, including questions about the thermodynamics and kinetics of the processes involved in charging tRNAs with cognate amino acids. The recognition by aaRSs of the appropriate cognate amino acid involves up to four identification steps (Freist *et al*., 1998).

- *First, selective binding*: the amino acid and a molecule of ATP bind to the active site of the aaRS, sometimes leading to a change in conformation.
- *Second, selective activation*: the aaRS catalyzes the formation of a covalent bond between the amino acid and ATP, forming an aminoacyl adenylate complexed with the aaRS and removing inorganic pyrophosphate.
- *Third, pre-transfer proofreading*: if misactivation of a non-cognate amino acid occurs, the aaRS may hydrolytically cleave the aminoacyl adenylate complex.
- *Fourth, post-transfer proofreading*: If a non-cognate aminoacyl adenylate has survived, the aaRS may hydrolytically cleave the aminoacyl-tRNA complex.

In *Thermus thermophilus* SerRS, referred to as *ttSerRS* hereafter, binding of ATP or seryl adenylate leads to the stabilization of a motif 2 loop that interacts with the tRNA acceptor stem. Thus, the presence of the seryl adenylate may encourage binding of tRNA at the active site. Specificity for serine binding to *ttSerRS* is achieved through (Belrhali *et al*., 1994)

- hydrogen bonds between the side-chain hydroxyl group of serine and Glu279
- hydrogen bonds between the side-chain hydroxyl group of serine and Thr380
- exclusion of large side chains with the small size of the binding pocket.

In *yeast* SerRS, seryl-tRNA is thought to optimize the binding site for serine to aid in the discrimination against non-cognate amino acids (Gruic-Sovulj *et al*., 2002). Model building studies (Belrhali *et al*., 1994) in *yeast* SerRS have

suggested that threonine, glycine, and alanine might be competitors to serine, but of these three, only threonine shows significant competition in a standard pyrophosphate exchange reaction (Gruic-Sovulj *et al*., 2002).

In the HierDock procedure for predicting incorporation of non-natural amino acids, we assume that the *selective binding* step is a necessary but not sufficient condition for amino-acylation. Thus we would design a mutant aaRS so that it would bind much stronger to the non-natural amino acid than to any of the 20 natural amino acids. As a step towards a better structural–energetic mechanism for amino acid discrimination by SerRS, we will predict the binding of the 20 natural amino acids to SerRS (Zhang *et al*., 2002; Kekenes-Huskey *et al*., 2003; Datta *et al*., 2004). Such computational results can be compared directly with experimental binding studies, but such studies are rarely available. More common is data on the rates of misincorporation of non-cognate amino acids. This is a more ambiguous test of the computation since a strongly binding non-cognate amino acid could be prevented from incorporation by one of the other three mechanisms discussed above.

In this paper we start with the experimental crystal structure of *ttSerRS* and dock serine to this structure using HierDock procedure, to predict the serine bound to *ttSerRS*. The predicted structure gives a coordinate RMS deviation (CRMS) in the serine ligand heavy atoms of 0.87 Å from the corresponding atoms in the crystal structure. All the 20 natural amino acids were docked to *ttSerRS* and their relative binding energies calculated in the best-docked conformation and in the activated mode. Our results suggest that the competitors with Ser for formation of the aminoacyl adenylate intermediate bound to SerRS are Thr and Asn, in good agreement with experimental results (Gruic-Sovulj *et al*., 2002) in which Thr is known to compete with Ser for the formation of the aminoacyl adenylate intermediate. We further predict that Asn is a competitor that can be directly tested by a standard pyrophosphate exchange reaction. This validates the use of HierDock to screen competitors for the uptake of an amino acid analog. These techniques should also be useful for competitor screening while designing mutant aaRSs for the uptake of a particular amino acid analog.

Methods

Force fields

Unless specified otherwise all calculations for ligand–protein complexes used the DREIDING force field (FF) (Mayo *et al*., 1990) with charges from CHARMM22 (MacKerell *et al*., 1998) [The DREIDING FF with CHARMM22 charges has been validated for many docking studies in proteins (Vaidehi *et al*., 2002; Wang *et al*., 2002; Datta *et al*., 2003; Kekenes-Huskey *et al*., 2003; Floriano *et al*., 2004b), and for molecular dynamics (MD) simulations (Brameld and Goddard, 1998; Brameld *et al*., 1998). The non-bond interactions were calculated using the Cell Multipole Method (Ding *et al*., 1992) as implemented in MPSim (Lim *et al*., 1997).

For the 20 natural amino acids, the FF charges were obtained using the Jaguar (2000 'Jaquar 4.1' Schrodinger, Inc., Portland, Oregon) *ab initio* electronic structure suite with the Hartree–Fock (HF) level of quantum mechanics (QM) and the 6-31G** basis set (Krishnan *et al*., 1980). We used the Jaguar implementation of the Poisson–Boltzmann (Tannor *et al*., 1994) continuum solvation method (PBF) to calculate the self-consistent (Mulliken) charges in water. For serHx-AMP, we used the B3LYP (Slater, 1974; Becke 1993) flavor of density functional theory with the 6-31G** basis set to obtain the charges for an arbitrary conformation.

For the ligand–protein complex, we used the Surface Generalized Born (SGB) solvation method (Ghosh *et al*., 1998) for optimizing the structure of the ligand–protein complex and for calculating the binding energy. For the HierDock calculations we used the Analytical Volume Generalized Born (AVGB) continuum solvation method (Zamanakos, 2002) to calculate the solvation energies.

The energy of SerRS–serHx–AMP–xtal was minimized by the MPSim program using conjugate gradients. This minimized structure differs from the crystal structure (with hydrogens added by Biograf) by a CRMS (all-atom coordinate root mean square difference) of 0.59 Å. The CRMS of the atoms in the binding site (defined as all protein residues with atoms within 5 Å of any atom of the ligand) is 0.53 Å and while the heavy atoms in the ligand differ by CRMS of 0.79 Å from the crystal.

SCREAM for placement of side-chain conformations

In placing different ligands in a binding site, it is often useful to scan a complete set of amino acid side-chain conformations for the new ligand and for the other residues in the binding pocket to find the most compatible conformations. There are many approaches used for this [including SCWRL (Bower *et al*., 1997), SCAP (Xiang and Honig, 2001)], but we decided to use a method compatible with our DREIDING FF, SCREAM (Side-Chain Rotamer Energy Analysis Method) (V. W. T. Kam, N. Vaidehi and W. A. Goddard III, manuscript in preparation). SCREAM evaluates the interaction energy of each side-chain rotamer from a side-chain rotamer library (1478 rotamers based on 1.0 Å resolution) with the rest of the protein using Equation (1) from the all-atom DREIDING FF

$$E_k = \sum_{i,j}\left(\frac{q_i q_j}{4\pi\varepsilon r_{ij}} + D_e\left(\left(\frac{r_m}{r_{ij}}\right) - 2\left(\frac{r_m}{r_{ij}}\right)^6\right) + D_{\mathrm{HB}}\left(5\left(\frac{r_{\mathrm{HB}}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{\mathrm{HB}}}{r_{ij}}\right)^{10}\right)\cos^4\theta\right) \quad (1)$$

where i sums over all atoms of the target residues while j sums over with all atoms of nearby protein residues, q_i denotes the partial charge on atom i, r_{ij} is the distance between atoms i and j, r_m and D_e are the van der Waals distance and well depth for atom pair i and j, r_{HB} and D_{HB} are hydrogen bond distance and well depth for atom pair i and j (assuming that one is an acceptor atom and the other is a donor atom), and θ is the angle between the bond of the donor atom to the bridging hydrogen atom from the line between the donor and acceptor atoms. To avoid over penalizing clashes, the van der Waals radii are reduced to 90% of the standard values in the DREIDING FF.

Allowing only the side-chain atoms of the target residue to move, each rotamer conformation is minimized, and the most energetically favorable rotamer is chosen and saved for subsequent steps.

Structure preparation

Starting with the crystal structure (Belrhali *et al*., 1994) of *ttSerRS* with serHx-AMP bound (PDB code: 1SES) from the Protein Data Bank, we selected one monomer of the dimer and added hydrogens using BioGraf. In addition, we added a Na^+ or Cl^- counterion for each charged residue not involved in a salt bridge. This structure will be referred to as SerRS–serHx–AMP–xtal.

We then minimized (using conjugate gradients with MPSim) the energy of the SerRS–serHx–AMP–xtal structure using SGB solvation. Next, to optimize the interactions between the ligand and the protein, we carried out annealing MD using the DREIDING FF while keeping the protein fixed but making the ligand movable. The annealing temperature started at 50 K, increased to 300 K, and then decreased to 50 K with annealing increments of 50 K every 0.5 ps. At the end of the dynamics run, the energy was minimized again using MPSim and SGB solvation. Then, we optimized the side-chain conformations of Thr225 and then Thr380 using the SCREAM side-chain replacement program as described above. Note that this procedure did not significantly change the heavy atom positions for these residues, for they remained within 0.09 and 0.08 Å CRMS of their previous positions, respectively. However, we noted that the hydroxyl hydrogens of these residues achieved better hydrogen bond interactions with the ligand.

The full ligand–protein complex was again minimized, yielding what we denote as the serHx-AMP/SerRS(FF) structure. The all-atom CRMS between the SerRS–serHx–AMP–xtal structure and the serHx-AMP/SerRS(FF) structure is 0.47 Å. Additionally, the CRMS of the residues containing atoms in the binding site (within 5 Å of the ligand) is 0.42 Å and the CRMS of the heavy atoms in the ligand is 0.59 Å.

HierDock protocol

The HierDock hierarchical docking strategy for docking ligands into proteins has previously been tested and validated for predicting ligand binding sites and ligand binding energies for many globular (Datta *et al*., 2002, 2003; Zhang *et al*., 2002; Floriano *et al*., 2004b) and membrane-bound proteins (Floriano *et al*., 2000, 2004a; Vaidehi *et al*., 2002; Freddolino *et al*., 2004; Hall *et al*., 2004; Kalani *et al*., 2004; Trabanino *et al*., 2004; Hummel *et al*., 2005). These studies show that the HierDock strategy of coarse-grain docking followed by fine grain minimizing leads to efficient and accurate predictions for ligand binding to proteins.

Herein we use an improved version of HierDock referred to HierDock 2.0 and described in Vaidehi *et al*. (2002). The various steps involved in HierDock 2.0, as used in this study are as follows:

1. *Mapping the binding region*: The negative of the molecular surface of the minimized serHx-AMP/SerRS(FF) structure (with waters and counterions removed) was used to define the region within the receptor over which to sample the various ligand conformations. The void region around the serine moiety of the ligand binding region in serHx-AMP/SerRS(FF), created by cleaving the C–N bond of the serHx and replacing the N with an O, was mapped with spheres generated using the Sphgen program in DOCK 4.0 (Ewing *et al*., 2001).
2. *Protein grid calculation*: The docking step uses an energy grid for the protein contribution to the interaction energy. This grid is calculated only once per target using the program Grid which is part of the DOCK 4.0 package. This grid used a box constructed with a 10 Å margin around the sphere centers defining the docking region.
3. *Level 0: Coarse-grain conformational search*: DOCK 4.0 was used to dock each ligand into the binding site of the target protein to generate an ensemble of 150 000 conformations. We used the options of flexible docking, energy scoring, simplex minimization before scoring, and 150 000 maximum scored conformations. The DOCK 4.0 scoring function calculates Coulomb and van der Waals interaction energies for all atoms of the ligand using the protein grid. As described above, we used Mulliken charges for the amino acid ligands and DREIDING (Mayo *et al*., 1990) van der Waals radii. The van der Waals radii for the ligands (but not for the protein) were reduced by 25% to allow closer contacts to the target protein. For each ligand the 3000 best conformations (by DOCK scores) was kept for subsequent analysis.
4. *Buried surface filter*: For each ligand, the percentage of ligand buried surface area was evaluated for each of the 3000 conformations from Level 0. Conformations with <75% of buried surface area were eliminated, while the remaining ligand conformations were scored by their DOCK scores. The best 50 conformations per ligand were carried to the next step.
5. *Level 1: Ligand minimization*: For each of the 50 conformations per ligand from Step 3, we minimized the structure of the ligand for 25 conjugate-gradient steps while keeping the protein atoms fixed, using the all-atom DREIDING FF for ligand and protein with no solvation. The five lowest energy conformations were selected for subsequent steps.
6. *Level 2: Ligand–protein full minimization*: The five best ligand/protein conformations from Step 5 were minimized as in Level 1 but, this time, allowing all atoms to move (protein and ligand) for 100 conjugate-gradient steps or until the RMS force is <0.2 kcal/mol/Å.
7. For each of the five structures from Step 5 obtained for each of the ligands, we calculated the solvation energy using the AVGB continuum solvation method (Zamanakos, 2002) and subtracted this from the total energy given by the FF. The best structure was chosen as the one with the binding energy calculated by Equation (2)

$$BE_1 = ES\ (\text{ligand}) + ES\ (\text{protein}) - ES\ (\text{ligand/protein complex}) \qquad (2)$$

   Here ES (A) indicates the FF energy of species A including the AVGB solvation energy of species A. The conformation of the bound ligand with the best-binding energy from Equation (2) was selected for further analysis.
8. The complex and the most favorable binding energy from Step 7 above was minimized with solvation forces from SGB to yield what we will refer to as the best-docked complex. The binding energy for the best-docked complex was then calculated (Equation 2), by separating the ligand and the protein from this complex and minimizing the free protein and ligand for 10 steps using conjugate-gradient minimization and SGB solvation.

Binding energy calculation of the 20 natural amino acids in the conformation that is activated by the protein

For a bound amino acid other than Ser to be activated by *ttSerRS* (to form the aminoacyl adenylate complexed with the aaRS, Step 2 above), it is likely that the orientation of the backbone of the amino acid must be the same as that of serine, since the backbone is the part of the bound amino acid that must be activated. We refer to this backbone conformation, presumably required for the formation of the aminoacyl adenylate, as the '*activated mode*'. This backbone conformation assumes that the Cβ of the amino acid side chain is to be placed in SerRS's side-chain binding pocket, for the other atoms bound to the Cα are fixed. This is a reasonable assumption because side chains placed elsewhere could get in the way of ATP binding or catalysis. Using HierDock to predict the best-docked conformation for each amino acid in the serine binding region of the serHx-AMP/SerRS(FF) structure, we found a slightly different preferred binding conformation for each ligand backbone. Since we cannot be sure that these modified backbone conformations would lead to the activation by SerRS and promote the subsequent catalysis, we used the following procedure to obtain the binding energy for each of the 20 natural amino acids in the activated mode.

The predicted structure of serine bound to SerRS from Step 8 above was used as a starting point for these calculations. The backbone atoms of the target amino acid were given the same coordinates as those of serine. This should ensure a conformation consistent with the activated mode. Then SCREAM (see *SCREAM for placement of side-chain conformations*) was used to generate the optimum side-chain conformations for the target amino acid in the binding site. Here, the ligand side chain was minimized with SGB solvation forces. Then, while the backbone of the ligand and all residues of hydrogen bonding to the N- and C-terminal ends of the ligand were kept fixed, the remaining atoms in the protein–ligand complex were minimized with SGB solvation forces. The binding energy for each complex was then calculated using Equation (2) and SGB solvation energies. All 20 natural amino acid side chains were placed using this technique, including serine, with the serine conformation remaining within a CRMS of 0.67 Å from the best-docked conformation.

Results and discussion

Prediction of binding site of serine in ttSerRS using the HierDock protocol

To prepare each of the 20 natural amino acids for docking, we started with the zwitterion form in the extended conformation and assigned Mulliken charges from HF QM with Poisson–Boltzmann solvation, as previously described in Methods. Then, we carried out Steps 1–7 of the HierDock procedure (see Methods) for each of the 20 natural amino acids.

The best-docked structure of ser/SerRS from *HierDock Protocol* was chosen as the predicted structure of serine bound to SerRS and is referred to as ser/SerRS(HierDock). The CRMS between the heavy atoms of the serine in ser/SerRS(HierDock) structure and the corresponding atoms of the Ser/SerRS crystal structure is 0.87 Å (Figure 1).

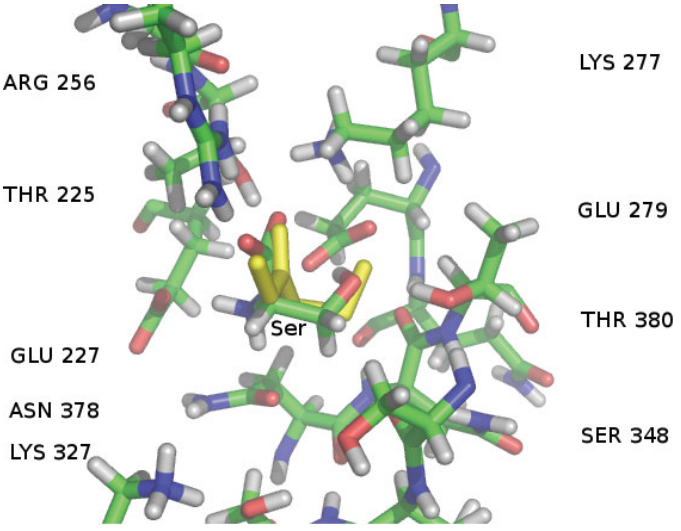

Fig. 1. The heavy atoms of the best-docked conformation of serine bound to SerRS and the corresponding atoms of the serHx-AMP ligand (shown in yellow) from the crystal structure are within a CRMS of 0.87 Å, suggesting that serine binds to SerRS in a conformation close to that of serine in the activated aminoacyl adenylate. Note that the crystal ligand has an intervening amide nitrogen between its carbonyl carbon and its terminal oxygen, and that these two oxygen accept hydrogen bonds from Arg 256.

Table I. Hydrogen bond distances (Å) in the serine binding site of SerRS for the predicted ser/SerRS(HierDock) structure, for the serHx-AMP/SerRS(FF) structure, and for the 1SES crystal structure

Ligand atom	Protein atom	ser/SerRS (HierDock) structure	serHx-AMP/SerRS (FF) structure	Crystal structure (Belrhali *et al.*, 1994)
N(amino)	Glu227 O	2.90	2.86	3.09
N(amino)	Asn378 O	2.90	2.90	3.58
N(amino)	Glu279 O	2.85	2.83	2.67
O(alcohol)	Thr380 O	2.90	2.91	2.77
O(alcohol)	Glu279 O	2.92	2.85	2.59
O(carboxyl)	Thr225 O	2.84	2.88	3.01
O(carboxyl)	Arg256 N	3.87	3.37	3.93
O^{XT}	Arg256 N	2.97	2.85[a]	3.18[a]

[a]Indicates that the terminal oxygen in both the ligand in the crystal structure and in the serHx-AMP/SerRS(FF) structure are separated from the carbonyl carbon by an intervening –NH group. The similarities between the distances in the ser/SerRS(HierDock) and serHx-AMP/SerRS(FF) structures suggests that the best-docked conformation of serine is in the activated mode since it makes contacts similar to those made by the serine moiety of the non-hydrolyzable seryl-AMP analog from the crystal structure.

The amino group of the ser/SerRS(HierDock) structure forms hydrogen bonds with Glu227, Asn378 and Glu279, while the alcohol group forms hydrogen bonds with Thr380 and Glu279, and the carboxylate group hydrogen bonds with Thr225 and Arg256. The lengths of these hydrogen bonds are given in Table I. Note that the ser/SerRS(HierDock) structure leads to all the important hydrogen bond interactions found in the serHx-AMP/SerRS(FF) and crystal structures (Table I and Figure 2), showing that, for serine, HierDock leads to a similar binding mode as observed in the crystal structure.

Binding energy calculation for docked structures of the 20 natural amino acids

The binding energy of the best-docked structure of each amino acid bound to SerRS was calculated as described in

HierDock Protocol. For serine and threonine, we found that the activated mode conformation had a more favorable binding energy than the best-docked conformation, presumably due to the fact that the coarse-grain conformational search (see Step 3 of the HierDock protocol above) used an electrostatic potential on a grid without the explicit hydrogen bond term as afforded by the side-chain placement program SCREAM (see discussion on *Binding energy calculation of the 20 natural amino acids in the conformation that is activated by the protein*). We will refer to the best-binding conformation as the one with the highest binding energy, whether it be the best-docked conformation or the activated mode conformation. Thus, for these two amino acids, the activated conformation and not the best-docked conformation is the best-binding conformation. The binding energies of the natural amino acids, relative to that of serine, are shown in Figure 3, and suggest that asparagine, tyrosine, aspartic acid and arginine would all compete with serine to bind to SerRS.

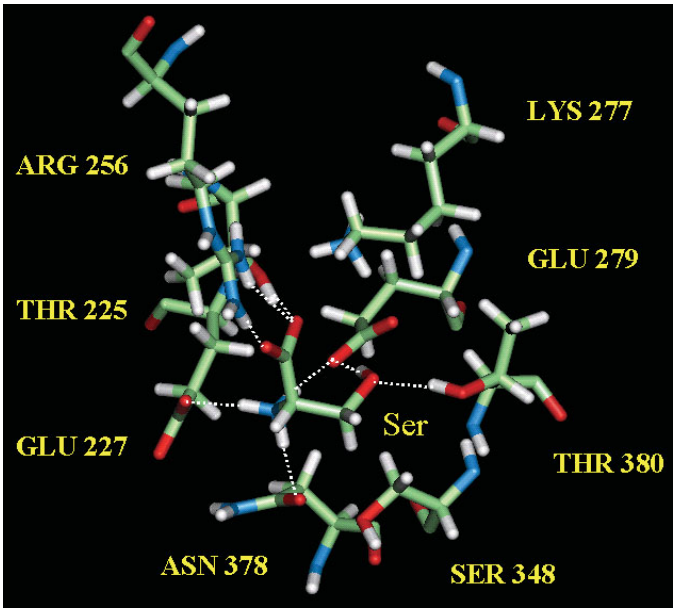

Fig. 2. Hydrogen bonds made between the best-docked conformation of Ser and SerRS. The heavy atom hydrogen bond distances are in Table I. The hydroxyl group of the serine ligand hydrogen bonds with Glu279 and Thr380, the same residues that hydrogen bond with this group in the crystal structure, suggesting that the serine side chain in the Ser/SerRS complex binds in the same side-chain binding pocket as in the crystal structure of SerHx-AMP bound to SerRS.

We find (Figure 4) that asparagine binds to SerRS better than serine. However, asparagine binds in a conformation different than serine, a conformation that we believe would not be activated by SerRS, for the amino group of the asparagine ligand is in a significantly different position in the binding pocket than the amino group of the serine ligand. The asparagine side chain is stabilized by hydrogen bonds to the side chains of Ser348 and Ser350, as well as by an intramolecular hydrogen bond between the side chain and the carboxyl group. Figure 4 shows the position of asparagine in its binding site along with the position of the phosphate group (shown in yellow) of the serHx-AMP ligand from serHx-AMP/SerRS(FF) structure. It is probable that the side chain of the asparagine ligand in this conformation would hinder the formation of the asparaginyl adenylate complex because it clashes with the PO_4 group of the AMP, as shown by van der Waals surfaces and by the N–O non-bonded interaction distance of 2.3 Å (Figure 4).

We predict that tyrosine would also compete with serine to bind to SerRS. However, we find that the best-docked conformation of tyrosine (Figure 5A) is not in the activated mode. The side chain of tyrosine is stabilized by a hydrogen bond to the N_δ on His204, but its N-terminal amino group only makes two instead of three strong hydrogen bonds, and only one of the oxygens in its C-terminus hydrogen bonds to the protein. Moreover, the side-chain conformation of tyrosine is flipped outside the binding pocket compared with serine in the same binding pocket.

We also predict aspartic acid to be a competitor with serine to bind to SerRS. But here again we conclude that, in the best-docked conformation, aspartic acid would not be activated by SerRS. The side chain is stabilized by salt bridges with Lys277 and by a hydrogen bond with the side chain of Ser350, while the zwitterion groups of the aspartic acid ligand are placed similarly to those in the ser/SerRS(HierDock) structure (Figure 5B). However, the side

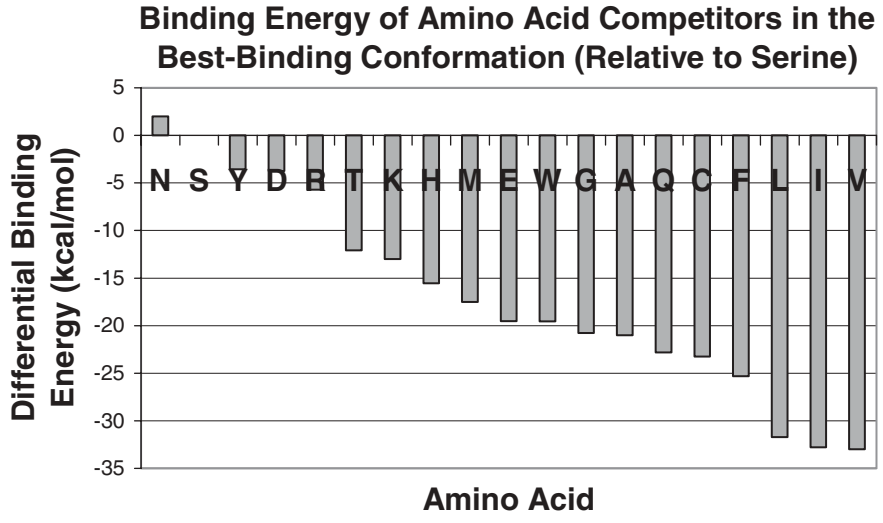

Fig. 3. Binding energies of amino acids to SerRS in the best-binding conformation, relative to that of serine (proline showed a significantly lower differential binding energy and so was not included in the figure). Asparagine, tyrosine, aspartic acid and arginine are predicted to compete with serine to bind to SerRS.

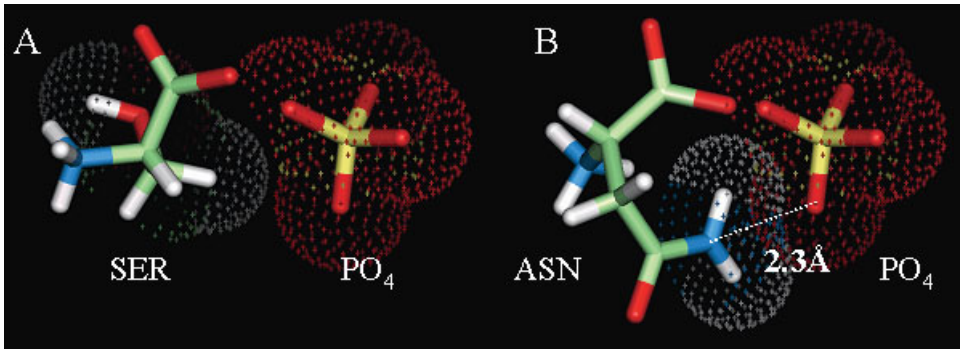

Fig. 4. (**A**) Comparison between the best-docked conformation of serine and the phosphate group of serHx-AMP from serHx-AMP/SerRS(FF). The side chain of serine does not clash with the phosphate group, as such a clash would be expected to sterically interfere with the adenylation reaction. (**B**) Comparison between the best-docked conformation of asparagine and the phosphate group of serHx-AMP from serHx-AMP/SerRS(FF). The steric clash between these two groups, shown by the overlap of the van der Waals surfaces and the N–O distance of 2.3 Å, suggests that the docked conformation of asparagine would inhibit the formation of asparaginyl adenylate. Not shown is the amino N–P distance of 3.2 Å.

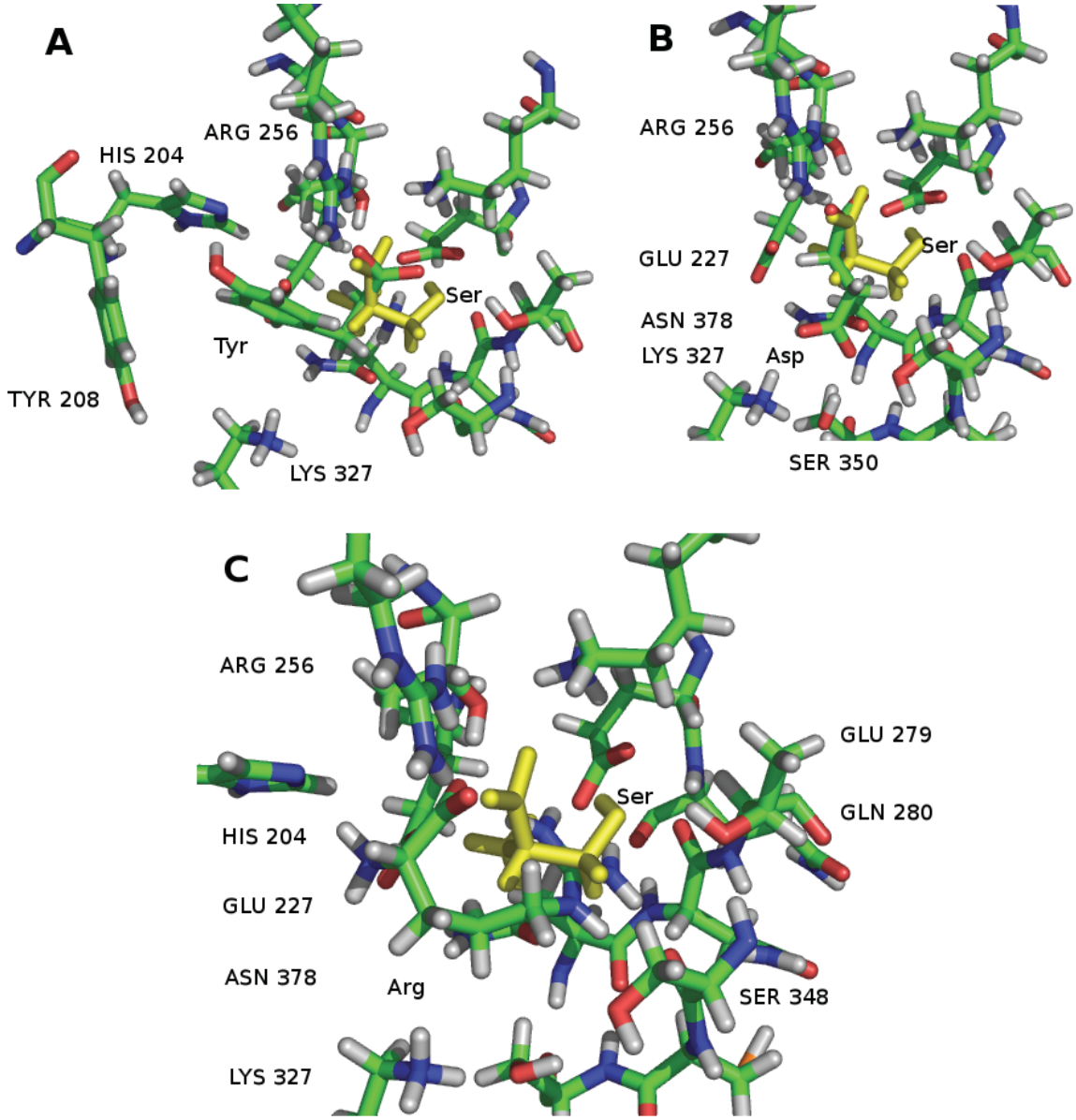

Fig. 5. The best-docked conformations of (**A**) tyrosine, (**B**) aspartic acid and (**C**) arginine bound to SerRS. In each case, the best-docked conformation of serine is superimposed as a solid yellow stick model. The side chains of tyrosine and aspartic acid are flipped outside of the serine side-chain binding pocket and would likely hinder the binding of ATP. Arginine's side chain is in the serine binding pocket but its N-terminal nitrogen is found 3.90 Å away from that of serine. Assuming the only conformations activated will be those that have a similar backbone conformation as serine (and thus also place their Cβs in SerRSs side-chain binding pocket), these best-docked conformations would not be activated by SerRS.

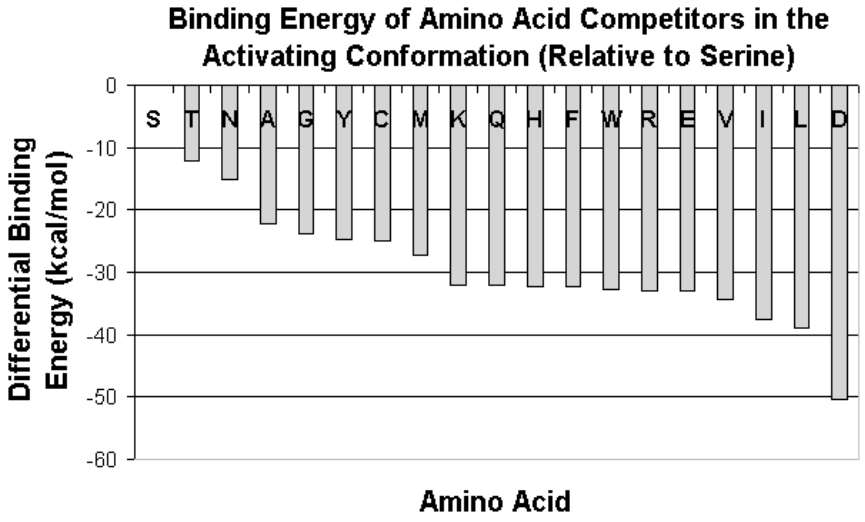

Fig. 6. Binding energies of amino acids to SerRS in the activated mode, relative to that of serine (proline showed a significantly lower differential binding energy and so was not included in the figure). Asparagine is close to threonine in binding energy, and since threonine is known to be misactivated by SerRS, we predict that asparagine would also be misactivated by SerRS. Note that alanine and glycine have significantly lower binding energies than threonine or asparagine.

chain is placed outside the binding pocket of the serine side chain, and would likely hinder the binding of ATP.

We also predict that arginine will compete with serine to bind to SerRS. However, we conclude that arginine will not compete with serine for the formation of the aminoacyl adenylate intermediate because the binding pocket is too small to accommodate its long side chain (Figure 5C). We see that the arginine side chain is stabilized by a salt bridge to Glu279. Because of this salt bridge and the long alkyl chain of arginine, the backbone zwitterion groups of the arginine ligand cannot make the same interactions with SerRS as can those of the serine ligand; thus, we believe this conformation of arginine will not be activated by SerRS.

In summary, our HierDock calculations suggest that asparagine, tyrosine, aspartic acid and arginine compete with serine to bind to SerRS, but that these best-binding configurations would not be activated by SerRS. These predictions could be tested by experiments that measure the binding constants of these natural amino acids to SerRS. We predict that such experiments would lead to the ordering in Figure 3.

Competitors for formation of the aminoacyl adenylate in SerRS

Placing each amino acid in the activated mode leads to the calculated relative binding energies shown in Figure 6. The five amino acids with the highest binding energies in the activated mode are serine, threonine, asparagine, alanine and glycine. Threonine has been shown experimentally to compete with serine for formation of the aminoacyl adenylate in SerRS, while alanine and glycine have been shown to not compete significantly (Gruic-Sovulj *et al*., 2002). Our calculations are consistent with these experimental results, as the binding energies (relative to serine) of alanine and glycine to SerRS in the activated mode are significantly lower than those of threonine and asparagine (Figure 6). As we have calculated that the binding energy of asparagine to SerRS in the activated mode is close to that of threonine and significantly higher than that of alanine and glycine (Table II), we predict that asparagine will also compete with serine in activation by SerRS. The extent to which asparagine competes with serine in activation by SerRS has not yet been tested experimentally.

Table II. Comparison of binding energies of amino acids to SerRS in the best-binding conformation and in the activated mode

Amino acid ligand	Binding energy (kcal/mol) Best binding	Activated mode
Ser	57.64	57.64
Thr	45.55	45.55
Asn	59.64	42.54
Ala	36.63	35.52
Gly	36.87	33.78
Tyr	54.08	32.88
Cys	34.39	32.57
Met	40.13	30.32
Lys	44.64	25.55
Gln	34.82	25.47
His	42.09	25.30
Phe	32.33	25.27
Trp	38.08	24.91
Arg	51.92	24.73
Glu	38.11	24.72
Val	24.65	23.40
Ile	24.86	20.03
Leu	25.94	18.73
Asp	53.89	7.21
Pro	16.60	−99.92

The amino acids are ordered by the binding energy in the activated mode. Since Asn in the activated mode has a binding energy close to that of threonine, we predict that Asn is misactivated by SerRS at a rate slightly less than but comparable with threonine.

The binding energy of the 20 amino acids in the activated mode shows that threonine binds to SerRS 12.09 kcal/mol less favorably than serine. The hydroxyl group of the threonine side chain makes hydrogen bonds to Glu279 and Thr380. This is expected, since threonine's side chain is similar to serine's and threonine has been shown to be misactivated by yeast SerRS (Gruic-Sovulj *et al*., 2002). Our calculations show that the threonine side chain methyl group has an unfavorable steric clash with the backbone of Ser348

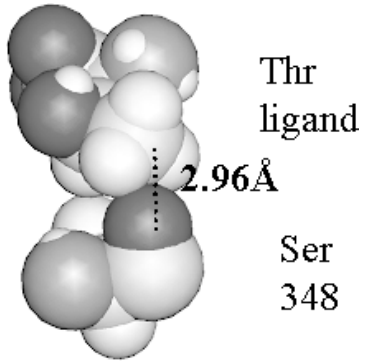

Fig. 7. van der Waals surface of the activated conformation of threonine with the van der Waals surface of Ser348, prepared using QUANTA. The side chain methyl of threonine clashes with the carbonyl oxygen of Ser348, at a distance of 2.96 Å from the methyl carbon to the carbonyl oxygen. Since serine does not suffer this clash, SerRS thus uses the small size of its amino acid side chain pocket to reduce misactivation by amino acids larger than serine.

Table III. Non-bond pairwise interaction energies (using Equation 1) between the serine ligand and protein residues within 5.0 Å in the ser/SerRS(activated mode) structure, compared with the interaction energies between these residues and the ligands of the thr/SerRS(activated mode) and asn/SerRS(activated mode) structures

Non-bond interaction energy (kcal/mol) SerRS		Ligand		
Residue		Serine	Threonine	Asparagine
HIS	204	0.95	0.96	0.97
GLY	224	0.83	0.84	0.89
THR	225	−3.00	−3.13	−3.40
GLU	227	−18.69	−18.57	−19.09
ARG	256	−13.89	−14.27	−14.37
LYS	277	−10.87	−10.95	−14.75
GLU	**279**	**−25.51**	**−25.67**	**−3.79**
GLN	280	−0.50	−0.61	−0.91
LYS	327	2.58	2.40	3.19
SER	348	−1.67	2.29	−2.85
CYS	349	−0.01	−0.20	−0.07
SER	350	−0.18	−0.36	−0.22
ASN	**378**	**−12.83**	**−11.53**	**−24.63**
ASN	**379**	**−1.68**	**0.25**	**−6.42**
THR	380	−11.52	−11.05	−12.54

Threonine's interactions with SerRS are similar to serine's, but threonine does not interact with Ser348 as favorably as does serine or asparagine (Ser348 is highlighted in blue in figures). Asparagine interacts less favorably with Glu279 than do serine and threonine, but interacts more favorably with Asn378 and Asn379 than do serine and threonine (Glu279, Asn378 and Asn379 are in bold).

(Figure 7), which contributes to the fact that threonine binds less favorably than serine (compare Ser and Thr in Table III).

Asparagine differs from serine and threonine in its hydrogen bonding to SerRS in the activated mode. The NH_2 group of the asparagine side chain is stabilized by hydrogen bonds to the backbone oxygen of Asn379 and the side chain oxygen of Asn378, while the carbonyl oxygen of the asparagine side chain is stabilized by a hydrogen bond from Thr380. The non-bond interaction energy (Equation 1) between protein residues within 5.0 Å of the binding pocket and the ligands in the ser/SerRS(activated mode) and asn/SerRS(activated mode) structures are shown in Table III. The difference in the stability of the activated conformations of serine and asparagine in SerRS is dominated by the difference between the strength of the hydrogen bond of the alcohol side chain of serine with charged Glu279 compared and the strength of the extra hydrogen bond of the side chain of asparagine with neutral Asn378. The binding energies for the 20 amino acids to SerRS in the best-docked conformation and in the activated mode are compared in Table II.

Conclusion

We find that HierDock leads to useful predictions of the binding for the 20 natural amino acids to SerRS. The predicted binding conformation of serine bound to *ttSerRS* is within 0.87 Å CRMS (heavy atoms) of the crystal structure.

HierDock finds that several natural amino acids (tyrosine, aspartic acid and arginine) could compete with serine to bind to SerRS, but the fact that their backbone binding conformations differ substantially from the serine binding conformation suggests that these other ligands will not lead to activation by SerRS. Our ordering of binding energies (Figure 2) could be tested by competitive binding experiments.

Considering the activated mode suitable for the formation of the aminoacyl adenylate, we find that serine is most favorable followed by threonine and asparagine and then by glycine and alanine. These results are consistent with the experimental evidence that yeast SerRS misactivates threonine at a significant rate but does not misactivate glycine and alanine. The prediction that SerRS misactivates asparagine at a rate comparable with threonine has not been tested experimentally. These results suggest that SerRS discriminates against non-cognate amino acids by the way it stabilizes the serine side chain with hydrogen bonds and favorable Coulombic interactions and by the way it excludes amino acids with larger side chains via a small side-chain binding pocket. Further calculations and experiments could re-design SerRS to incorporate non-natural amino acids; based on the present work, such design efforts would be well advised to consider the non-natural amino acids in the activated mode, as our calculations in the activated mode show good agreement with experimental measurements of activation by SerRS.

Acknowledgements

We thank Mr Ismet Caglar Tanrikulu for much assistance in developing the methodology and software tools to aid in calculating the binding energy in the activating mode. We also thank Dr Wely B.Floriano, Mr Peter Freddolino and Dr Rene J.Trabanino for additional assistance and valuable advice. Mr C.L.M. was funded in part by Mr and Mrs Donald M.Alstadt as part of the Summer Undergraduate Research Fellowship (SURF) program at Caltech. This research was also funded by CSEM (NSF MRSEC) and NIH. Molecular graphics presented in this paper were prepared using QUANTA (1998 Molecular Simulations, Inc.) and PyMOL (DeLano, 2002). The facilities of the Materials and Process Simulation Center at Caltech (MSC) used for this work were supported by grants from DURIP-ARO DURIP-ONR, IBM-SER and the Beckman Institute.

References

Becke,A.D. (1993) *J. Chem. Phys.*, **98**, 5648–5652.
Belrhali,H. *et al*. (1994) *Science* **263**, 1432–1436.
Bower,M.J., Cohen,F.E. and Dunbrack,R.L. (1997) *J. Mol. Biol.*, **267**, 1268–1282.
Brameld,K.A. and Goddard,W.A. (1998) *J. Am. Chem. Soc.*, **120**, 3571–3580.
Brameld,K.A., Shrader,W.D., Imperiali,B. and Goddard,W.A. (1998) *J. Mol. Biol.*, **280**, 913–923.

Datta,D., Vaidehi,N., Xu,X. and Goddard,W.A. (2002) *Proc. Natl Acad. Sci. USA* **99**, 2636–2641.
Datta,D., Vaidehi,N., Floriano,W.B., Kim,K.S., Prasadarao,N.V. and Goddard,W.A. (2003) *Proteins* **50**, 213–221.
Datta,D., Vaidehi,N., Zhang,D.Q. and Goddard,W.A. (2004) *Protein Sci.*, **13**, 2693–2705.
DeLano,W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA. http://www.pymol.org
Ding,H.Q., Karsawa,N. and Goddard,W.A. III (1992) *J Chem Phys.*, **97**, 4309–4315.
Ewing,T.J., Makino,S., Skillman,A.G. and Kuntz,I.D. (2001) *J. Comput. Aided Mol. Des.* **15**, 411–428.
Floriano,W.B., Vaidehi,N., Goddard,W.A., Singer,M.S. and Shepherd,G.M. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 10712–10716.
Floriano,W.B., Vaidehi,N. and Goddard,W.A. (2004a) *Chem. Senses*, **29**, 269–290.
Floriano,W.B., Vaidehi,N., Zamanakos,G. and Goddard,W.A.,III (2004b) *J. Med. Chem.*, **47**, 56–71.
Freddolino,P.L., Kalani,M.Y., Vaidehi,N., Floriano,W.B., Hall,S.E., Trabanino,R.J., Kam,V.W. and Goddard,W.A. (2004) *Proc. Natl Acad. Sci. USA*, **101**, 2736–2741.
Freist,W., Sternbach,H., Pardowitz,I. and Cramer,F. (1998) *J. Theor. Biol.*, **193**, 19–38.
Ghosh,A., Rapp,C.S. and Friesner,R.A. (1998) *J. Phys. Chem. B*, **102**, 10983–10990.
Gruic-Sovulj,I., Landeka,I., Soll,D. and Weygand-Durasevic,I. (2002) *Eur. J. Biochem.*, **269**, 5271–5279.
Hall,S.E., Floriano,W.B., Vaidehi,N. and Goddard,W.A. (2004) *Chem. Senses*, **29**, 595–616.
Hummel,P., Vaidehi,N., Floriano,W.B., Hall,S.E. and Goddard,W.A. (2005) *Protein Sci.*, **14**, 703–710.
Kalani,M.Y., Vaidehi,N., Hall,S.E., Trabanino,R.J., Freddolino,P.L., Kalani,M.A., Floriano,W.B., Kam,V.W. and Goddard,W.A.III (2004) *Proc. Natl Acad. Sci. USA*, **101**, 3815–3820.
Kekenes-Huskey,P.M., Vaidehi,N., Floriano,W.B. and Goddard,W.A. (2003) *J. Phys. Chem. B*, **107**, 11549–11557.
Krishnan,R., Binkley,J.S., Seeger,R. and Pople,J.A. (1980) *J. Chem. Phys.*, **72**, 650–654.
Lim,K.T., Brunett,S., Iotov,M., McClurg,R.B., Vaidehi,N., Dasgupta,S., Taylor,S. and Goddard,W.A. (1997) *J. Comput. Chem.*, **18**, 501–521.
Link,A.J., Mock,M.L. and Tirrell,D.A. (2003) *Curr. Opin. Biotechnol.*, **14**, 603–609.
MacKerell,A.D. *et al*. (1998) *J. Phys. Chem. B*, **102**, 3586–3616.
Mayo,S.L., Olafson,B.D. and Goddard,W.A. (1990) *J. Phys. Chem.*, **94**, 8897–8909.
Slater,J.C. (1974) *Quantum Theory of Molecules and Solids*, Vol. 4. McGraw-Hill, New York.
Tannor,D.J., Marten,B., Murphy,R., Friesner,R.A., Sitkoff,D., Nicholls,A., Ringnalda,M., Goddard,W.A. and Honig,B.J. A. C. S. (1994) *J. Am. Chem. Soc.*, **116**, 11875–11882.
Trabanino,R.J., Hall,S.E., Vaidehi,N., Floriano,W.B., Kam,V.W. T. and Goddard,W.A. (2004) *Biophys. J.*, **86**, 1904–1921.
Vaidehi,N., Floriano,W.B., Trabanino,R., Hall,S.E., Freddolino,P., Choi,E.J., Zamanakos,G. and Goddard,W.A. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 12622–12627.
van Hest,J.C.M. and Tirrell,D.A. (2001) *Chem. Commun.*, 1897–1904.
Wang,L. and Schultz,P.G. (2002) *Chem. Commun.*, 1–11.
Wang,P., Vaidehi,N., Tirrell,D.A. and Goddard,W.A. (2002) *J. Am. Chem. Soc.*, **124**, 14442–14449.
Xiang,Z.X. and Honig,B. (2001) *J. Mol. Biol.*, **311**, 421–430.
Zamanakos,G. (2002) A fast and accurate analytical method for the computation of solvent effects in molecular simulations. Thesis (PhD) PQ#3045468 available at http://resolver.caltech.edu/caltechETD:etd-04062005-082441.
Zhang,D.Q., Vaidehi,N., Goddard,W.A., Danzer,J.F. and Debe,D. (2002) *Proc. Natl Acad. Sci. USA*, **99**, 6579–6584.

Received August 22, 2005; revised December 24, 2005; accepted January 31, 2006

Edited by Dieter Söll