Docking by Monte Carlo Minimization with a Solvation Correction: Application to an FKBP–Substrate Complex

AMEDEO CAFLISCH

Biochemisches Institut der Universität Zürich, CH-8057 Zürich, Switzerland and Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02140

STEFAN FISCHER,* and MARTIN KARPLUS†

Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, F-6700 Strasbourg, France and Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02140

Received 30 April 1996; accepted 19 August 1996

ABSTRACT

A Monte Carlo docking procedure that combines random displacements of the substrate and protein side chains with minimization of the enzyme–substrate complex is described and applied to finding the binding mode of the blocked tetrapeptide *N*-acetyl-Leu-Pro-Phe-methylamide to the FK506 binding protein (FKBP). The tetrapeptide, an analog of the preferred FKBP substrate, and the FKBP binding site are flexible during the docking procedure. The twisted-imide transition-state form of the substrate is used during docking. The enzyme charges are scaled individually to account for solvent screening of specific binding site residues during the Monte Carlo sampling. To evaluate the relative binding free energies of the resulting structures, a rapid method for calculating polar and nonpolar solvation effects is introduced. Accurate electrostatic solute–solvent energies are calculated by solving the finite-difference linearized Poisson–Boltzmann equation; nonpolar contributions to the stability of the different conformers are estimated by the free energy of cavity formation, which is obtained from the molecular surface, and the solute–solvent van der Waals energy, which is calculated with a continuum approach. In the conformation of the enzyme–substrate complex with the lowest free energy, the tetrapeptide is bound as a type VIa proline turn with solvent accessible ends to permit longer polypeptide chains to act as substrates. Except for the imide carbonyl, which is involved in polar interactions with aromatic side chains of the FKBP binding

*Present address: Pharmaceutical Research, F. Hoffmann–La Roche Ltd., CH-4070 Basel, Switzerland.

†Author to whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 18, No. 6, 723–743 (1997)

CCC 0192-8651 / 97 / 060723-21

site, all of the seven potential hydrogen bond donors or acceptors of the tetrapeptide are satisfied. The FKBP binding site has a similar conformation in the substrate complex as in the FKBP-FK506 cocrystal structure, except for the predicted reorientation of the Tyr 82 hydroxyl, which plays an important role in substrate binding. The present model for the FKBP–substrate complex is in agreement with the recently determined crystal structure of a cyclic peptide–FK506 hybrid bound to FKBP and supports the structure obtained previously by iterative model building. In addition, it is consistent with the observed effects of FKBP point mutations on the enzyme activity. The approach described here should be useful, in general, for the prediction of the structure of a molecule in solution or as part of a complex. It provides for the effective sampling of conformational space and for the inclusion of solvent effects.

Introduction

One of the general problems in understanding enzyme mechanisms at the molecular level is the difficulty of obtaining, by experimental means, the structure of the bound substrate, and even more so, the structure of the transition state. Thus, theoretical models for such species can play an important role. The present work is concerned with the development of an efficient combined Monte Carlo/Minimization method (MCM) that includes solvation effects for modeling the structure of a bound substrate or transition state. The method is illustrated by applying it to determine the structure of the enzyme–substrate complex of the peptidylprolyl isomerase FKBP, the FK506 binding protein[1–3]; the substrate considered is the blocked tetrapeptide *N*-acetyl-Leu-Pro-Phe-methylamide (Fig. 1).[4] The choice of this system was motivated by its intrinsic interest, the existence of the crystal structure of a cyclic peptide–FK506 hybrid bound to FKBP,[5] and the need to confirm the substrate binding geometry predicted by iterative model building in an analysis of the prolyl isomerase mechanism of FKBP.[6]

Peptidylprolyl isomerases (PPIases) catalyze the interconversion between cis and trans rotamers of peptidylprolyl amide bonds in peptide substrates and are widely distributed in biological systems. Two well-characterized classes of PPIases exist: the FKBPs[1,2] and the cyclosporin A binding proteins (cyclophilins). Both are defined in terms of their artificial ligands, the natural products FK506 and cyclosporin A, that are macrocyclic immunosuppressants that inhibit the rotamase activity of their respective class, while leaving the other class unaffected. The therapeutic value of these compounds and their importance as templates for immunosuppressive drugs and as probe molecules for cytoplasmic signal transduction have spurred great interest in understanding their interactions with the binding proteins and the effect of the complex on the immune system.[7,8] Although there is compelling experimental evidence that inhibition of the PPIase activity is not related to immunosuppression,[9] the mechanism of the rotamase activity is of considerable interest. In addition, PPIases are important for accelerating the folding of certain proteins[10,11] for which the rate-limiting step involves the trans to cis isomerization of proline peptide bonds.[12] The blocked tetrapeptide used in this study, an analog of the preferred substrate,[13,14] is docked onto human FKBP-12 (107 residues and 12 kDa), where the suffix 12, which is dropped henceforth, distinguishes it from other FKBPs by its molecular weight. FK506 binds to FKBP in a deep hydrophobic pocket, which consists of several aromatic and aliphatic side chains. Its ''floor'' consists of the indole ring of the unique Trp 59. The substrate is docked with the proline imide bond constrained near the cis–trans transition state, because it is expected that the enzyme binds this distorted form most strongly. Furthermore, cis and trans conformers can be easily obtained from the twisted imide form by letting the bond relax into either of the two planar imide forms.[6]

The present method is an improved version of a previously developed approach for docking.[15,16] In this approach the protein was kept rigid and each cycle of the MCM consisted of random perturbations of the flexible dihedral angles of the ligand, followed by a phase of conventional energy minimization and application of the Metropolis criterion to the resulting structure. Such a combined MCM method was applied to oligopeptide folding by Li and Scheraga,[17,18] and with the addition of a thermalization procedure, by Caflisch et al.[19] Im-

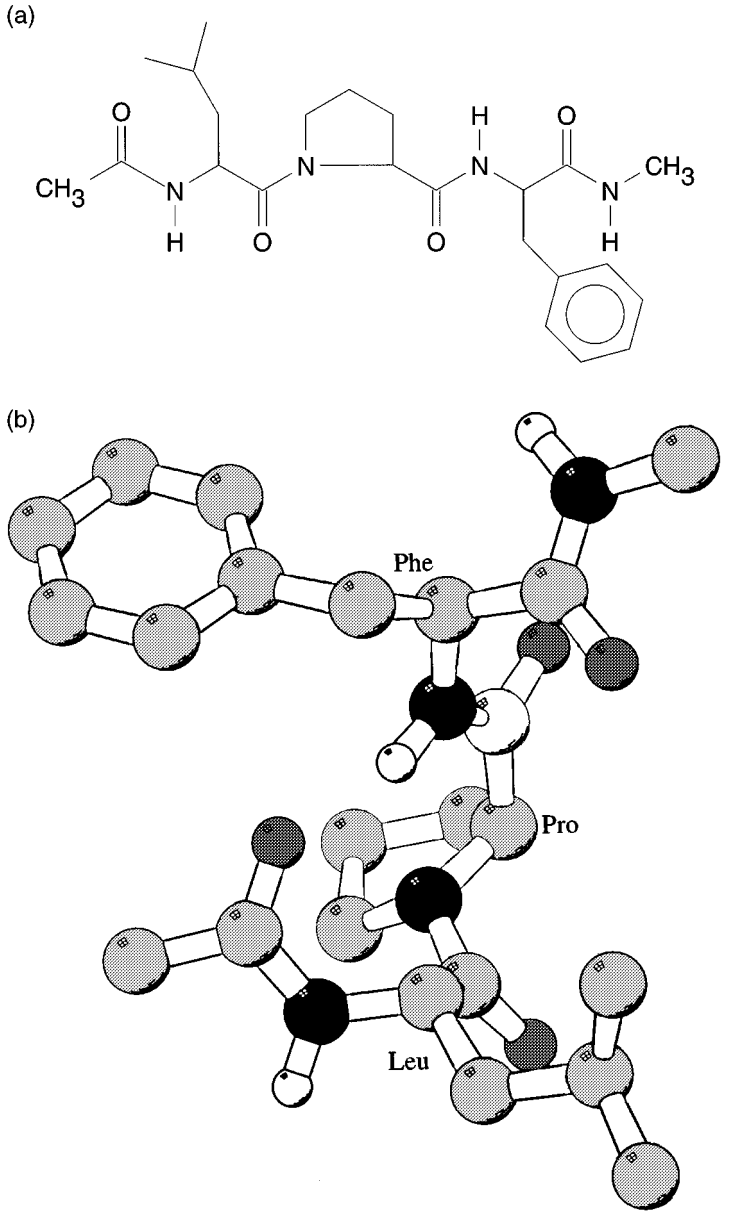

FIGURE 1. (a) Structure of the *N*-acetyl-Leu-Pro-Phe-methylamide substrate. (b) Twisted syn imide form of the *N*-acetyl-Leu-Pro-Phe-methylamide substrate, found to be the lowest free energy conformer (89th saved structure in MCM run S1). Polar hydrogens are shown in white, carbons and oxygens in gray, and nitrogens in black. Figure 1b was made with the Molscript program.[4]

provements introduced in this article are: the use of a flexible enzyme binding site, whose side chains are submitted to random perturbations; random rigid body translations and rotations of the ligand in addition to perturbations in the torsion angles of the ligand; scaling of the enzyme charges to account for solvent screening effects during the docking procedure; and estimation of the free energy of solvation of the resulting conformers by applying a continuum solvent model that accounts for the hydrophilic and hydrophobic effects. The first two improvements result in more extensive sampling of the conformational space, while the latter two yield a more accurate treatment of the relative binding free energy of the different bound structures generated by the algorithm. This two-step approach for estimating solute–solvent contributions makes it possible to preserve the rapidity of the calculation resulting from the use of a simple molecular mechanics potential energy function

in the docking procedure, while evaluating the candidate structures with a slower and more accurate postprocessing procedure.

The lowest free energy enzyme–substrate complex has the tetrapeptide bound as a type VIa proline turn with an intrasubstrate hydrogen bond; the terminal methyl groups are exposed, as required if the turn were part of a protein chain. This binding mode is in agreement with the effect of point mutations on the catalytic activity of FKBP. The proposed structure also displays the essential features of the recently solved crystal structure of a cyclic peptide–FK506 hybrid bound to the active site of FKBP.[3,5]

The Monte Carlo docking procedure and the approach used to evaluate the free energy of the resulting conformers are described in the Computational Methods section. The energy values of these conformers and the correlation between the scaled Coulombic energy and the continuum dielectric value are given in the Results, as well as a detailed analysis of the lowest free energy structures. A brief comparison of the present work with other docking algorithms and a recent study by Zacharias et al.[20] is presented in the concluding Discussion.

Computational Methods

SYSTEM SETUP

This section describes the potential energy function and the degree of freedom used during the MC docking procedure. The coordinates of the enzyme were taken from the crystal structure of the FKBP–FK506 complex.[21,22] Neutral blocking groups were added to the model substrate *N*-acetyl-Leu-Pro-Phe-methylamide (Fig. 1) to avoid artifacts from charged peptide termini. The standard CHARMM[23] all-hydrogen empirical energy function (PARAM 22) was used for the FKBP and the substrate. Intrasolute Coulombic interactions were computed with a distance-dependent dielectric parameter, and a cutoff at 12 Å was used with a shifted function.[23] Intrasolute van der Waals interactions were computed with a 6–12 Lennard–Jones potential, and a cutoff at 12 Å was used with a switching function.[23] The proline imide torsional potential was modified to hold the imide bond near either of its two possible twisted imide transition states.[6] This was achieved by constraining the virtual dihedral angle ζ (C_α^{Leu}–O^{Leu}–C_δ^{Pro}–C_α^{Pro}) to $\pm 90°$.[24]

Standard protonation states were used for all residues except for the histidines. The pK_a values of His 25, His 87, and His 94 were determined with the continuum dielectric model and finite-difference technique[25] to be 6.15, 6.84, and 6.79, respectively; these values were essentially unaffected by the binding of the substrate. Thus, for the model structure, which corresponds to pH 7, singly protonated His were used. The electrostatic interaction energy between the side chains of each of the histidines and the rest of the system was determined. It was found that the ε-protonated residue was most stable in all three cases;[6] this form was used in the present study.

Solvent shielding of the electrostatic interactions between FKBP and the substrate was taken into account during the MC docking procedure. To introduce this in an efficient way, the charges of each binding site residue i were scaled by a factor ε_i equal to the ratio of the CHARMM Coulombic interaction between the individual FKBP residue i and the entire substrate, to the corresponding screened interaction calculated by numerical solution of the finite-difference Poisson–Boltzmann (PB) equation ($\varepsilon_{int} = 1$, $\varepsilon_{ext} = 78.5$; details are given later). The ratio ε_i can be considered as a residue-dependent dielectric constant, which accounts for the difference in the solvent shielding of the interactions with the different residues.[25] The values of ε_i are the same as those used previously.[6] To test the sensitivity of the ε_i to local structural changes, 100 conformers generated by MC docking were examined. The resulting scaling factors ε_i were found to be similar. The charges on protein residues having Coulombic interaction with the substrate smaller in absolute value than 0.1 kcal/mol were not scaled.

During the minimization phase of MC docking, the substrate and 23 residues of FKBP delimiting the binding pocket (i.e., residues 25–27, 36–39, 42, 46–48, 54–56, 59, 81–82, 86–87, 90–91, 97, and 99)[21,22] were free to move. The remaining FKBP residues were kept fixed.

MC DOCKING

A FORTRAN program was developed to generate a CHARMM input script for the MCM calculation. It performs the MC perturbations on a set of internal coordinates and on the translational and rotational degrees of freedom of the ligand and applies the Metropolis criterion to the structure obtained after minimization. This combines the

advantages of the Metropolis MC method[26] in global optimization and that of the conjugate gradient method[27,28] in local optimization. The set of internal coordinates that undergoes MC perturbation can be specified so that this script generator can be applied to any molecular assembly. Eight starting conformations, chosen as described below, were submitted to the MCM procedure[15–19] to increase the search space in a simple way.

Each MCM cycle consisted of a perturbation in the conformation of the molecular complex, i.e., in the dihedral angle around rotatable bonds and in the relative orientation of protein and substrate, followed by energy minimization. First, a number n of torsion angles τ_i $(i = 1, \ldots, n)$ were perturbed, where n is a variable chosen with a probability 2^{-n}.[18] All other random numbers used in this work were taken from uniform distributions. The n torsion angles were randomly selected from the specified set of variable internal coordinates, i.e., the dihedral angles around the rotatable bonds of the substrate and of the side chains of FKBP residues Tyr 26, Phe 46, Val 55, Ile 56, Tyr 82, Ile 91, and Leu 97. These side chains, which have at least one atom within 8 Å of one or more FK506 atoms, contribute to delimiting the binding site but are likely to tolerate some reorientation without destroying the cohesion of the binding pocket. Torsional angles around amide bonds, proline and aromatic ring bonds, and those resulting in methyl group rotations were not included. The selected torsion angles τ_i were changed randomly by $-180° < \Delta\tau_i < 180°$. To increase conformational sampling of the substrate, its flexible dihedrals were perturbed twice as often as those of the binding site side chains. The position and orientation of the substrate with respect to the enzyme was perturbed by a random rigid body translation smaller than 0.35 Å along a randomly chosen vector and a random rigid body rotation by an angle smaller than 60° around a randomly selected axis through the substrate center of the geometry. In each MC cycle, the energy minimization phase was started by subjecting the perturbed structure of the enzyme–substrate complex to 10 steps of steepest descent minimization to remove steric clashes. If the resulting total energy of the complex was smaller than some absolute threshold (200 kcal/mol in the present case), the optimization was continued by conjugate gradient minimization until either the root mean square (RMS) of the energy gradient reached a value of 0.2 kcal/mol Å or a maximum number of minimization steps was performed (200 in this study). The threshold used to decide whether to proceed with the conjugate gradient minimization after steepest descent was introduced to improve the computational efficiency, because test runs showed that structures minimized by 200 conjugate gradient steps were always rejected by the Metropolis criterion if their absolute energy after the preliminary 10 steepest descent steps was larger than 150 kcal/mol. The conjugate gradient minimization found a nearby local minimum in the energy basin reached by a random perturbation of the complex. The Metropolis criterion[26] at a temperature of 500 K was then applied to the local minimum. The 500 K temperature used for the Metropolis criterion accepts increased energies (on the order of 1 kcal/mol) with reasonable frequency. This resulted in an efficient sampling of a large portion of the conformational space available to the substrate in the binding site (see below).

To direct the sampling toward conformations characterized by optimal binding and intrasubstrate energies, the Metropolis criterion was applied to the sum of the intrasubstrate and enzyme–substrate interaction energies. The intraenzyme energy was not included in the Metropolis criterion to prevent rearrangements of the enzyme groups from dominating the docking process. It is important to note that the enzyme conformations were not significantly strained during the docking procedure, because the total energy, i.e. the sum of intraenzyme, intrasubstrate, and enzyme–substrate energies, was minimized by the conjugate gradient algorithm. It was found that for the *in vacuo* simulations of the type used here, there was an artificial tendency of the binding site to close so as to improve the van der Waals and Coulombic interactions within the enzyme if the self-energy term was included in the Metropolis criterion. This neglect was justified by a comparison of the free solution structure[29] and the FK506 bound structure,[21,22] which showed only small differences in the region of the binding site. In other applications where there are larger changes in the structure of the enzyme on substrate binding, the enzyme self-term has to be included in the Metropolis criterion to direct the sampling toward significant regions of the configurational space.

To optimize the coverage of configurational space, a set of significantly different initial conformations of the complex, the two possible transition state forms of the substrate, syn (clockwise twisted form of the imide bond) and anti (counterclock-

wise twisted form), were each placed into the FK506 binding site in four different ways by manually orienting them with respect to two axes in the binding pocket. Around the first axis, which goes through the enzyme and exits the binding pocket orthogonal to the Trp 59 indole ring (Fig. 2a), the substrate was oriented with its N-terminus either near the Asp 37 side chain (henceforth called the ''up'' orientation) or near the backbone of residue Val 55 (''down'' orientation). For each of these two orientations, the substrate was rotated about an axis that is orthogonal to the first axis and goes from the Asp 37 side chain to the Val 55 side chain (Fig. 2a), with the imide carbonyl pointing either toward the interior of the FKBP binding side (''in'' orientation) or toward the solvent (''out'' orientation). Hence, a total of eight initial conformations of the enzyme–substrate complex were generated to insure thorough sampling; four of them had the syn imide bond twist, S1–S4, and four had the anti twist, A1–A4. Both syn and anti conformers were included because cis to trans transitions of the stiff amide bonds were not allowed. A series of MC runs using the different starting substrate orientations resulted in a more efficient sampling of conformational space than performing one long run with short thermalization periods. Moreover, with multiprocessor workstations, it was more efficient to do separate MCM calculations simultaneously (coarse grained parallelization) than to use a parallelized program and do one calculation on several processors.

Two thousand MCM cycles were performed for each of six of the eight initial substrate orientations. For each of the remaining two, the up and down orientations of the syn substrate conformer with the imide carbonyl pointing toward the floor of the binding site (S1 and S3), 6100 MCM cycles were performed. For these two MCM runs, a thermalization period of 50 MCM cycles at a Metropolis temperature of 2×10^4 K was introduced after 2000 and 4050 cycles to investigate if a long run with short thermalization periods resulted in a better sampling of conformational space than a series of short runs starting from the various different initial conformations. In each MCM run, conformations characterized by an energy (sum of the intrasubstrate and enzyme–substrate interaction energies) smaller than the lowest value found up to the current MCM cycle were saved. This resulted in 355 conformations that were subjected to further minimization (RMS of the energy gradient smaller than 0.005 kcal/mol Å) before estimating their solvation free energy as described below.

SOLVATION FREE ENERGY

The effect of the solvent on the relative stability of the bound conformers resulting from the docking procedure was estimated by a continuum treatment for polar and nonpolar interactions. Because the goal was to find which of the minima of the complex generated by the MC docking procedure was most stable, it was necessary only to calculate the relative free energies of the different enzyme–substrate complex structure. This was much simpler than evaluating the free energy of binding, which would have required the free energy of the free enzyme and the unbound ligand.

To account for hydrophilic and hydrophobic effects, the free energy of solvation of a given conformer was obtained by decomposing solvation into a 3-step process: formation of a cavity in the bulk solvent, introduction of dispersive centers into the cavity, and introduction of permanent partial charges into the cavity. The free energy change of step 1 is related to the change of the molecular surface area on binding[30]; volume dependent ($p\Delta V$) terms are expected to be negligible. The dispersive interactions between solvent and solute (step 2) were computed from a continuum van der Waals representation of the solvent medium. The energy terms associated with steps 1 and 2 were assumed to be independent of the permanent charges of the polar groups of the solute and were termed nonpolar contributions to solvation. Finally, in step 3 the continuum dielectric effect of the solvent on the electrostatic energy of the solute was computed by numerical solution of the finite-difference linearized (LPB) equation.[31] The sum of the three terms provided an estimate of the free energy of solvation of each bound conformer complex. It was added to the intrasolute energy evaluated with the CHARMM potential ($\varepsilon = 1$) for each conformer.

Nonpolar Solvation Free Energy

The nonpolar contribution to the solvation free energy of a molecule in aqueous solution consists of the unfavorable (positive) free energy of cavity formation, G_{cavity}, and the favorable solute–solvent dispersion energy, $G_{\text{vdW, solvat}}$.[32] G_{cavity} was estimated from the molecular surface (MS), which is defined as the solvent-excluding surface of the solute. This corresponds to the area mapped out by the surface of a spherical probe of radius 1.4 Å rolled over the van der Waals surface of the solute. The two components of the MS[30] are the convex

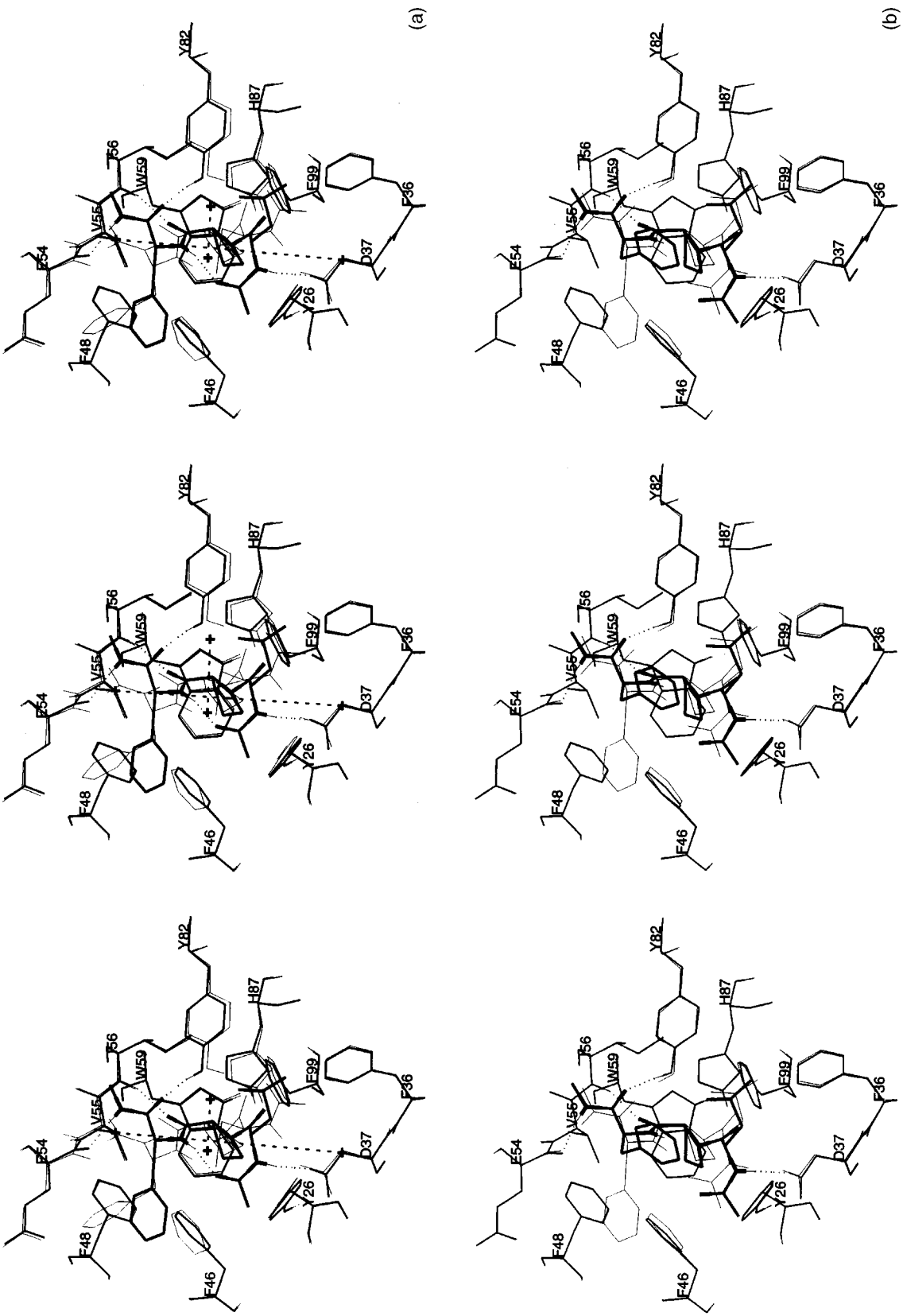

FIGURE 2. Stereoplot of the conformers characterized by the lowest free energy. Aliphatic and aromatic hydrogens are omitted. (a) S1:89, the substrate is in thick lines (thicker main chain N and O atoms, dotted hydrogen bonds) and FKBP in medium lines. The S1:1 conformation is also given in thin lines for comparison. Rotation around the two dashed lines connecting the crosses defines the four orientations of the substrate within the FKBP binding site: up, down, in, and out (see text); (b) S1:81 (thick lines, dotted hydrogen bonds) and S1:89 (thin lines).

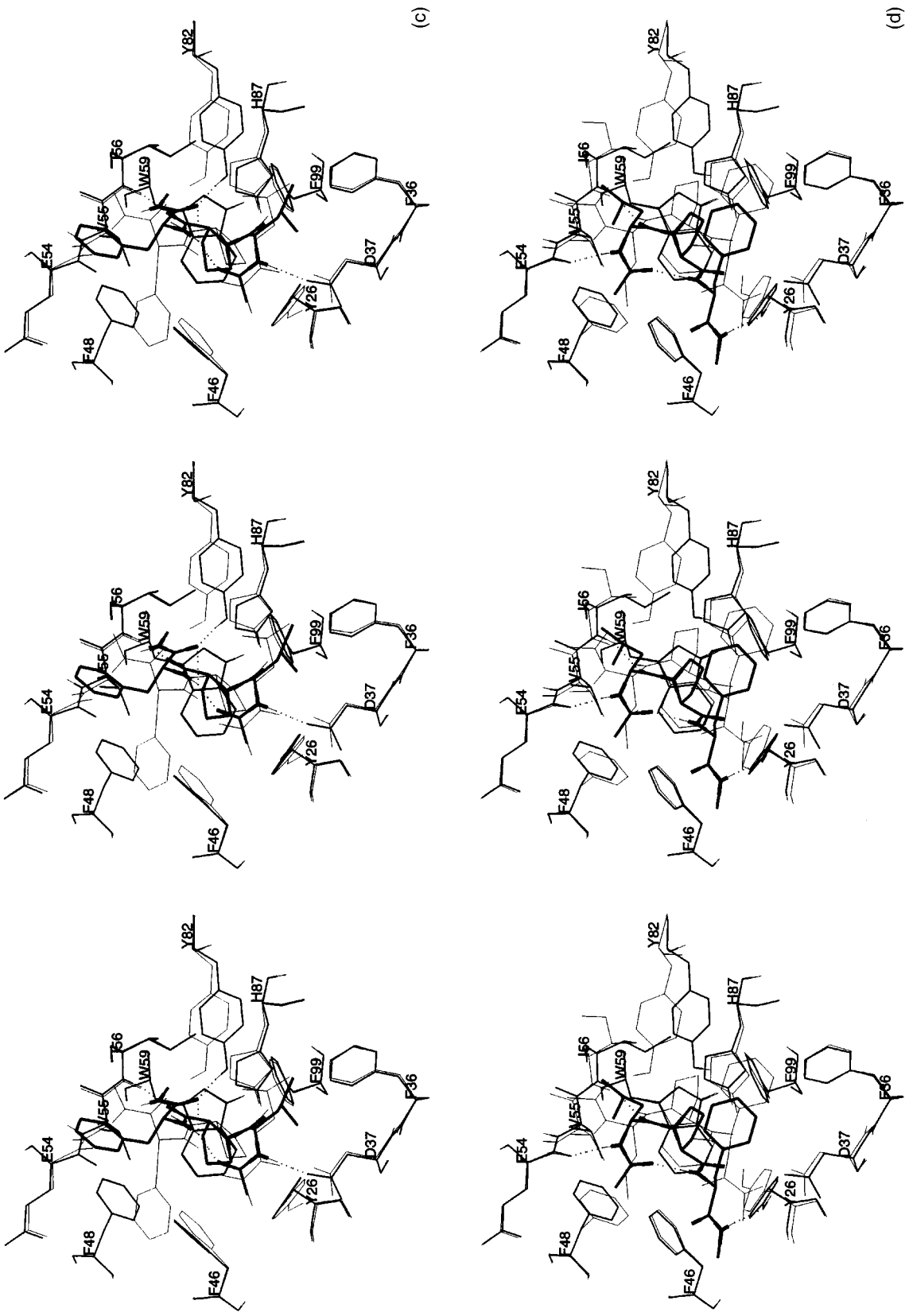

FIGURE 2. (Continued) (c) S1:41 (thick lines, dotted hydrogen bonds) and S1:89 (thin lines); (d) S3:1 (thick lines, dotted hydrogen bonds) and S3:77 (thin lines). In all stereofigures, the view is cross-eyed for the left pair of images and wall-eyed for the right pair.

solute/probe contact areas (contact MS) and the concave solute reentrant surface/probe ''contact'' areas (interstitial MS), which were calculated numerically with the program CODISP.[33] The resulting MS areas were multiplied by respective ''surface tensions:'' 0.104 kcal/mol Å^2 for the contact MS and 0.096 kcal/mol Å^2 for the interstitial MS. These values were parameters derived from a fit to the solvent contribution to the potential of mean force of the methane dimer as a function of the methane–methane distance, as obtained from simulations of methane association in water.[34] These parameters yielded a good correlation (RMS deviation of the error of 1.9 kcal/mol) with experimental values of the free energy of hydration for 19 weakly polar compounds.

$G_{vdW,solvat}$, the solute–solvent dispersion interaction, was calculated with a continuum model of the solvent, in which a van der Waals energy density was assigned to the bulk solvent, based on the CHARMM van der Waals parameters of a given solute atom and water and on the experimental density of water at room temperature.[33] For each solute atom, this energy density was integrated numerically over all solvent space. Summing this result over all solute atoms yielded the solute–solvent dispersion interaction energy.

Electrostatic Solvation Free Energy

Polarization of the solvent by the partial charges of the solute affects the electrostatic energy of a molecular assembly in two ways: the interactions among the solute partial charges are screened (interaction term) and the solvent reaction field interacts directly with each solute charge (self-term). The continuum electrostatic free energy of solvation calculated by solution of the PB equation was the sum of the interaction and self-terms.[35]

Studies have shown that application of the LPB equation yields a good estimate of the electrostatic free energies of solvation in macromolecules.[25,31,35,36] The LPB differential equation is approximated by a set of finite-difference equations on a grid.[37] The latter are solved on a computer by iterative adjustments of the value of the potential at each grid point. In this study, the UHBD program[38–40] was utilized for solving the finite-difference LPB equation. The partial charges and atomic radii of the CHARMM all-hydrogen potential were used for the LPB calculations. Recent results showed that the solvation free energy of models of polar and ionizable side chains, calculated with the finite-difference method and the CHARMM all-hydrogen parameter set, agreed well with the experiment.[41]

UHBD places the charges on a grid according to the trilinear weighting method.[42] First, a grid of $40 \times 40 \times 50$ points was used along with a grid spacing of 2.0 Å; this yielded a layer of solvent (high dielectric) of at least 20 Å around the structure of the complex (low dielectric). Coulombic potentials for each point charge were used to set the boundary potential. A second *focused* calculation was performed with a grid of $80 \times 80 \times 100$ points and a grid spacing of 0.5 Å; the potential obtained from the unfocused calculation was used for the boundary potential. Focusing dramatically reduces the error originating from the distribution of the charges on the grid points.[43,44] The protein dielectric constant was set to 1.0, which is consistent with the value used for the parametrization of the CHARMM charges. A dielectric constant of 78.5 was assigned to the continuum solvent medium. The molecular surface of the low dielectric region was delimited by applying a solvent probe of 1.4 Å radius. Furthermore the permittivity was linearly interpolated at the midpoints between grid points intersecting the dielectric boundary (dielectric boundary smoothing), because this reproduced the potential near the discontinuity region more accurately and has been shown to improve convergence.[40,44] Values of 298 K for the temperature, 100 mM for the ionic strength (corresponding to physiological conditions), and 2.0 Å for the Stern layer (ion exclusion layer) were used.

As a convergence criterion for the finite-difference LPB calculations, the iterative algorithm was stopped when the norm of the residual vector (or $|\mathbf{Ax} - \mathbf{b}|$, where $\mathbf{x}$ contains the field values at the grid points, and the matrix $\mathbf{A}$ and vector $\mathbf{b}$ originate from the conversion of the PB equation into a system of linear equations as described in Press et al.,[37] reached a value smaller than 10^{-6}. This required on average 100 and 270 iterations of the diagonal-scaling preconditioned conjugate gradient algorithm for the first and second (*focused*) calculation, respectively.

Finite-difference LPB calculations were used to adjust the enzyme–substrate electrostatic interaction energy during the MCM calculations and to evaluate the electrostatic free energy of solvation for the minima obtained by MCM. To compute the FKBP–substrate electrostatic free energy of interaction for evaluating the scaling factors, the sub-

strate atoms were charged and the protein was considered as a neutral region of low dielectric, which displaced the solvent. The electrostatic free energy of interaction between an FKBP residue i and the substrate is

$$G_{i,\mathrm{int}} = \sum_{j=1}^{N_i} q_j \phi_j, \qquad (1)$$

where N_i is the number of atoms in FKBP residue i, q_j is the charge of atom j on residue i, and ϕ_j is the electrostatic potential generated by the substrate at the location of atom j, calculated by solving the finite-difference LPB equation. No factor 1/2 appears, because the partial charges generating the electrostatic field reside on the atoms of the substrate, while the q_j belong to FKBP atoms. The CHARMM Coulombic interaction between FKBP residue i and the entire substrate (distance dependent dielectric, 12 Å shifted cutoff) was divided by $G_{i,\mathrm{int}}$ if the former value was larger in absolute value than 0.1 kcal/mol. This ratio, ε_i, represents an effective residue-specific dielectric constant[25] and was used to scale the partial charges of the enzyme residues during the MC docking runs.

For each minimum-energy conformation obtained by MCM, the evaluation of the electrostatic free energy was divided into two parts, the continuum electrostatic free energy of solvation $G_{\mathrm{elect,solvat}}$ (screening and self-energy) and the true intrasolute vacuum Coulombic energy, E_{elect}. The former is calculated by subtracting both the finite-difference approximation to the Coulombic interaction energy between charged atoms *in vacuo* and the interaction energy of each atom with its own potential (this finite contribution arises from the discretization of the atom charges onto a grid) from the total electrostatic energy of the system calculated by the finite-difference LPB echnique.[45] These three energy terms were calculated on the same grid to obtain consistent results. When the solute dielectric is 1.0 this yields the same result as the usual (and computationally more expensive) method of performing two finite-difference calculations; the first with the low dielectric solute in a high dielectric continuum and the second with the low dielectric solute *in vacuo* ($\varepsilon = 1.0$). The value $\varepsilon = 1.0$ for the solute was appropriate for the CHARMM charges. E_{elect} was calculated with the CHARMM program using unscaled charges, a dielectric constant of 1.0, no cutoff, and exclusion of 1–2 and 1–3 terms.

TOTAL FREE ENERGY ESTIMATE

The present treatment was concerned with finding the optimum conformation and position of a substrate in the binding site. As discussed earlier, this means that it was necessary only to determine the differences in free energy of the system with different bound forms of the substrate to determine the most stable bound conformation. The total free energy was estimated for the 355 conformers of the FKBP–substrate complex by use of the equation

$$G = E_{\mathrm{bond}} + G_{\mathrm{cavity}} + (E_{\mathrm{vdW}} + G_{\mathrm{vdW,solvat}}) + (E_{\mathrm{elect}} + G_{\mathrm{elect,solvat}}). \qquad (2)$$

G_{cavity}, $G_{\mathrm{vdW,solvat}}$, $G_{\mathrm{elect,solvat}}$, and E_{elect} were described earlier and E_{bond} and E_{vdW} are, respectively, the bonded terms (bonds, angles, and torsions), and the pairwise Lennard–Jones interactions of the standard CHARMM potential energy.

Changes in the intrasolute entropy of binding were neglected. These were assumed to be sufficiently similar for the different conformers that they did not contribute to the ΔG values, which were the only quantities of interest for determining the relative free energies of different enzyme–substrate complexes.

COMPUTATION TIMES

About 60 h CPU time on a single processor of a SGI 4D/340 workstation were required for 2000 cycles of MC docking. The evaluation of the two components of the nonpolar contribution took about 4 CPU min per conformation on a DEC Alpha 3000. Solution of the LPB equation required about 8 CPU min per structure on a CONVEX C220. Part of the PB calculations were performed on a DEC Alpha 3000; on this machine the incomplete Cholesky conjugate gradient algorithm[38] converged in about 4 min after about one-third of the iterations required for the diagonal-scaling conjugate gradient algorithm, which is simpler and more vectorizable.

Results

MCM

About one-half of the randomly perturbed structures had an energy smaller than the 200 kcal/mol threshold after the preliminary steepest descent minimization and were subjected to conju-

gate gradient minimization. On average, 120 conjugate gradient iterations were required for convergence to 0.2 kcal/mol Å and about 25% of the minimizations were stopped after reaching the maximum number of steps (200). Nearly 20% of the structures that underwent conjugate gradient minimization were accepted on the basis of the Metropolis criterion. In all, 254 structures for syn substrate conformers and 101 structures for anti conformers were saved and subjected to a full free energy evaluation. Table I lists the three lowest energy conformers from the four syn runs (S1–S4) and the four anti runs (A1–A4); in addition, the first (S1) or last (all others) are listed.

While the 50 cycles of thermalization (performed every 2000 MCM cycles in the S1 and S3 runs) yielded significant changes in the substrate structure and in the orientation of the binding site side chains, they did not result in complete reorientation of the substrate with respect to the binding pocket. For example, the thermalization did not allow the conversion of an up substrate to the down orientation (as defined previously). This confirmed the necessity to perform separate MCM docking runs with the different starting orientations of the substrate as described earlier.

The range of structures sampled in the MC simulations is illustrated by the maximum RMS differences in the position of the substrate in the various calculations. For the 89 minima determined starting with the S1:1 structure, the maximum backbone difference from the lowest minimum (S1:89) was found for S1:1 (2.61 Å) and the maximum side chain difference was S1:36 (5.34 Å). Thus, within a given region, the MC sampling covered a significant portion of the conformational space. Figure 3 shows the backbone and side chain deviations from S1:89 as a function of the number of the minimum. It can be seen that after S1:57, the backbone remained in the neighborhood of S1:89, but the side chains continued to have relatively large fluctuations until near the end of the run. Comparing the minima from all runs, the largest difference in substrate coordinates from S1:89 was obtained for A3:13 with an RMS deviation of 7.15 Å; the largest difference from S1:89 within the S1 set was 3.93 Å (S1:34). The flexible portions of the protein that were also optimized had significantly smaller displacements; for example, the largest deviation from S1:89 for all 355 minima was 1.07 Å for minimum S1:32.

Figure 4a shows plots of some of the calculated energies during the third part of run S1 (from cycle 4101 to cycle 6100), while Figure 4b shows the corresponding results for S2. Similar behavior was found for the other runs. A good correlation was seen between the energy used in the Metropolis criterion (sum of the intrasubstrate and enzyme–substrate contributions) and the enzyme–substrate interaction energy along all MCM trajectories. This implies that the substrate strain played a minor role in distinguishing between the substrate minima. The intrasubstrate energy ranged from 54.6 to 73.4 kcal/mol for the S1 minima and from 57.2 to 69.4 kcal/mol for the S2 minima. The internal strain in the tetrapeptide substrate originated mainly from the dihedral energy term.

For the 89 conformers saved in run S1, the enzyme–substrate interaction energy ranged from -80.1 to -57.7 kcal/mol; the average value and standard deviation were -71.8 ± 5.2 kcal/mol. For the 54 conformers from the S2 run, values ranging from -72.8 to -38.3 kcal/mol and an average of -58.1 ± 9.7 kcal/mol was found. Similar energy ranges were obtained in the other MCM runs. That the binding energy was always negative means that the attractive interactions dominated. This was due to the hydrophobic character of both the substrate side chains and the FKBP binding site and from the minimization phase that allows optimal electrostatic and van der Waals intermolecular energies. The relatively large ranges of values and standard deviation for the enzyme–substrate interaction energy was in accord with the fact that substantially different binding modes were obtained by the MCM procedure.

While the fluctuations in the Metropolis (intrasubstrate plus substrate-enzyme interaction) energy had amplitudes of up to 5.0 kcal/mol (kT at 500 K is 1.0 kcal/mol), the total energy of the enzyme–substrate complex showed larger oscillations than both the Metropolis energy and the intrasubstrate contribution. These arose mainly from the intraenzyme energy, which was not included in the energy used for the Metropolis criterion. Nevertheless, many of the successive structures saved during each MCM run had an improving total energy (diamonds on broken line in Fig. 4a, b). In all runs, the total energy of the accepted structures never exceeded -75 kcal/mol, because excessively large values of the intraenzyme energy were prevented by the minimization phase of the MCM cycle. The energy of the complex ranged from -145.0 to -98.9 kcal/mol for the S1 minima and from -129.7 to -86.6 kcal/mol for the S2 minima.

TABLE I.
Lowest Free Energy Conformations.

MCM Run and Minimum Number	Intrasolute Energy						Solvation Energy					
	FKBP		Substrate		FKBP, Substrate		Continuum Elect		Solute Solvent Dispersion, $G_{vdW,solvat}$	Cavity Formation Energy, G_{cavity}	G[f]	Figure
	Bond[a]	Nonb[b]	Bond[a]	Nonb[b]	vdWaals[c]	Elect[d]	FKBP, Substrate[e]	$g_{elect,solvat}$				
S1 (up, in)												
89[g]	158.9	−736.3	51.0	9.3	−45.1	−44.1	−20.6	−1058.4	−511.3	360.8	0.0	2a
81	−2.5	−0.6	−1.4	−0.1	3.0	−0.3	1.1	3.3	−0.6	0.0	0.6	2b
41	−2.4	−7.1	−1.4	−3.7	4.8	12.6	4.3	7.4	−0.8	1.2	10.4	2c
1	−2.9	−8.6	2.1	8.7	5.6	8.8	3.4	2.8	−3.0	2.4	15.8	2a
S2 (up, out)												
5	−1.7	−2.1	−0.2	0.2	16.5	12.2	5.7	2.5	−5.5	3.1	25.0	
2	−2.1	−7.9	0.7	2.6	13.7	18.5	9.2	3.5	−7.5	5.8	27.3	
3	−0.2	−11.2	1.2	−2.7	20.0	24.7	10.6	1.1	−10.1	7.1	29.6	
54	3.4	60.0	−4.9	1.2	8.3	1.7	1.2	−22.2	−1.6	−0.2	45.7	
S3 (down, in)												
1	−8.1	−10.5	1.1	−3.6	6.5	24.6	12.7	9.7	0.1	0.2	19.8	2d
5	−2.4	3.0	0.3	−7.4	7.7	13.7	6.7	6.5	1.6	−0.7	22.2	
8	−1.9	9.8	−6.2	1.1	6.1	5.3	2.6	7.4	2.6	−0.7	23.5	
77	4.3	16.2	−4.7	1.1	7.2	−9.5	−3.8	12.8	1.6	−0.5	28.4	2d
S4 (down, out)												
1	−2.2	1.0	2.9	4.2	9.2	15.7	5.8	−5.7	−1.4	−0.5	23.2	
24	3.8	−10.7	−4.9	−1.1	9.7	6.1	2.8	22.1	0.9	−2.0	23.4	
31	2.0	7.6	−2.4	−0.8	4.5	−2.1	−2.3	18.6	2.5	−3.5	26.5	
34	6.5	15.6	−1.9	−0.7	6.5	−4.3	−3.5	21.9	1.9	−2.4	43.0	
A1 (up, in)												
13	4.1	0.9	4.8	2.7	10.0	0.1	1.2	10.3	−4.5	4.3	32.7	
15	4.1	1.2	4.7	2.7	10.0	0.1	1.2	10.5	−4.4	4.3	33.1	
6	3.6	4.5	12.9	−4.6	5.3	16.4	7.9	−0.2	−0.7	1.5	38.5	
32	10.7	66.6	4.9	3.3	10.1	−6.3	−1.7	−15.2	−2.9	3.3	74.5	
A2 (up, out)												
1	−4.9	−19.5	9.4	0.9	15.0	23.9	11.0	4.0	−11.7	9.1	26.0	
5	−6.0	−17.8	7.8	−4.0	7.2	32.6	15.6	11.0	−4.2	2.3	28.9	
2	−5.6	−19	6.9	0.1	15.2	31.9	14.1	3.4	−8.7	8.5	32.7	
30	3.0	31.4	12.5	1.3	3.1	9.3	4.2	−4.7	0.9	−0.9	55.8	
A3 (down, in)												
1	−3.8	0.1	7.4	2.5	8.3	15.6	7.5	1.2	−2.9	0.5	28.6	
2	−4.0	0.0	7.4	2.4	8.2	16.1	7.4	1.2	−3.2	0.7	28.8	
12	2.5	4.9	7.0	2.9	8.8	6.1	3.2	4.5	−1.2	1.0	36.5	
15	14.0	17.6	7.6	2.5	8.0	4.8	2.4	0.7	−2.8	0.7	53.1	
A4 (down, out)												
4	−2.3	−18.3	7.5	2.9	7.0	12.3	5.6	15.1	0.9	−0.4	24.7	
12	0.3	−10.5	2.4	0.2	4.6	19.3	7.4	8.8	1.9	−0.7	26.1	
17	−1.3	8.6	0.6	1.5	7.3	10.3	6.5	3.6	−0.1	−0.2	30.3	
24	5.0	2.5	0.8	0.2	8.7	13.9	7.3	9.5	2.6	−1.6	41.6	

Energy values in kcal / mol are listed for the three lowest free energy conformations obtained in each MCM run. The last conformer saved during each run is also shown in each fourth line (except for S1, for which the first conformation is listed in the fourth line).
[a]Bonded energy terms (bonds, angles, and torsions) calculated with CHARMM.
[b]Nonbonded energy terms (van der Waals and vacuum electrostatic) calculated with CHARMM.
[c]CHARMM van der Waals interaction energy between FKBP and substrate.
[d]CHARMM vacuum electrostatic interaction energy between FKBP and substrate.
[e]LPB free energy of interaction between FKBP and substrate.
[f]Sum of energy values in all preceding columns except for the LPB free energy of interaction (eighth column).
[g]Conformation S1:89 (first line), which has the lowest total free energy, is taken as the reference. Except for the S1:89 conformer, for which the absolute energies are listed, all energies listed are relative to the value found for conformer S1:89.

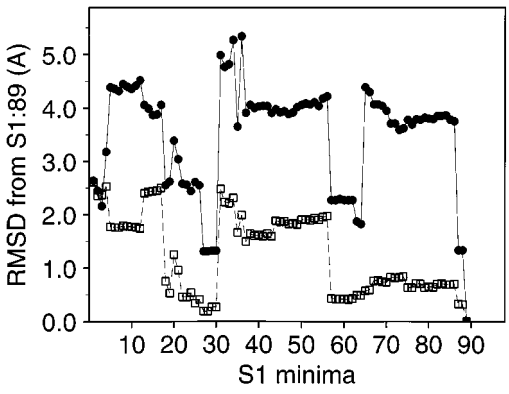

FIGURE 3. Substrate coordinates RMS deviation from conformer S1:89 plotted for the minima saved during run S1. (●) Side chains and (□) main chain.

FREE ENERGY RESULTS

For each of the 355 minima the value of the free energy was estimated with eq. (2). The Coulombic intermolecular energy used during sampling (calculated by CHARMM with scaled enzyme charges) was compared with the electrostatic free energy of interaction computed by numerical solution of the finite-difference LPB equation. For most of the minima, the scaled Coulombic energy value was up to 50% more favorable than the LPB free energy. However, the correlation, which is most important for the relative ranking of the conformers, was very high (Fig. 5a, b); the correlation coefficients (R) were 0.962 (254 syn minima), 0.970 (101 anti minima), and 0.963 (all minima). Further, the electrostatic free energy of interaction (evaluated *a posteriori* by the LPB approach) generally improved during the runs (see Fig. 4). This provided further support for the use of the residue-dependent dielectric constant.

As a basis of comparison, for 600 test structures of the λ repressor–operator complex, Zacharias and coworkers[20] found a correlation coefficient of 0.86 between finite-difference electrostatic energies and the approximation they used during sampling, which was based on an induced polarization approach.

The 89th and last minimum saved during S1 (called S1:89 henceforth) had the lowest total free energy. There was only one minimum with a free energy value within 10 kcal/mol of S1:89 and only three minima within 15 kcal/mol. Overall, the complexes generated during S1 had more favorable free energy than any of the minima from the seven other orientations. Nineteen among the 20 structures with the lowest free energies were obtained during S1; the remaining one was minimum 1 of S3 (S3:1), whose free energy was 19.8 kcal/mol higher than that of S1:89.

In the S1 orientation of the substrate, the improvement of the Metropolis energy (intrasubstrate energy plus enzyme–substrate interaction energy) progressed in parallel to the improvement in total free energy, whereas there was no such correlation in the runs with the other seven substrate orientations. The intraenzyme nonbonded (bonded) energy contributions deteriorated by 8.6 (2.9) kcal/mol in the run S1 in going from S1:1 to S1:89. In the other cases they deteriorated by 64.4 (9.3) kcal/mol in S2, 26.7 (12.4) kcal/mol in S3, 14.6 (8.7) kcal/mol in S4, 76.3 (5.2) kcal/mol in A1, 50.9 (7.9) kcal/mol in A2, 17.5 (17.8) kcal/mol in A3, and 6.9 (5.9) kcal/mol in A4. These results indicate that the enzyme binding site was able to accommodate the twisted imide substrate without any major deterioration of its self-energy only if the latter had an orientation relatively close to the one characterized by the lowest free energy. Different orientations of the substrate were optimized at the expense of a major loss in the self-energy of the enzyme binding site. This resulted mainly from a deterioration of the Coulombic interaction between side chains of the binding pocket.

To evaluate the influence of individual energy contributions to the total free energy difference among different bound conformers [eq. (2)], the range of each energy term was computed for the 20 best conformations. The largest variations occurred for the vacuum electrostatic interaction energy between the substrate and the enzyme (values ranging from −48.1 kcal/mol for S1:85 to −19.5 kcal/mol for S3:1) and the intraenzyme nonbonded energy (values ranging from −747.7 kcal/mol for S1:4 to −720.0 kcal/mol for S1:26). Ranges of 15.1, 14.5, and 13.5 kcal/mol were found for the FKBP bonded energy, the intrasubstrate nonbonded energy, and the continuum electrostatic free energy of solvation, respectively. The remaining contributions, substrate bonded energy, CHARMM intermolecular van der Waals energy, and the nonpolar solvation energy varied by no more than 8.0 kcal/mol among the 20 best conformations. For all 355 minima the greatest variability was found for the intraenzyme nonbonded energy (values ranging from −756.5 kcal/mol for A2:4 to −667.8 kcal/mol for S2:45) and the intermolecular vacuum electrostatic energy (values ranging from −53.6 kcal/mol for S3:77 to −1.7 kcal/mol for S3:50).

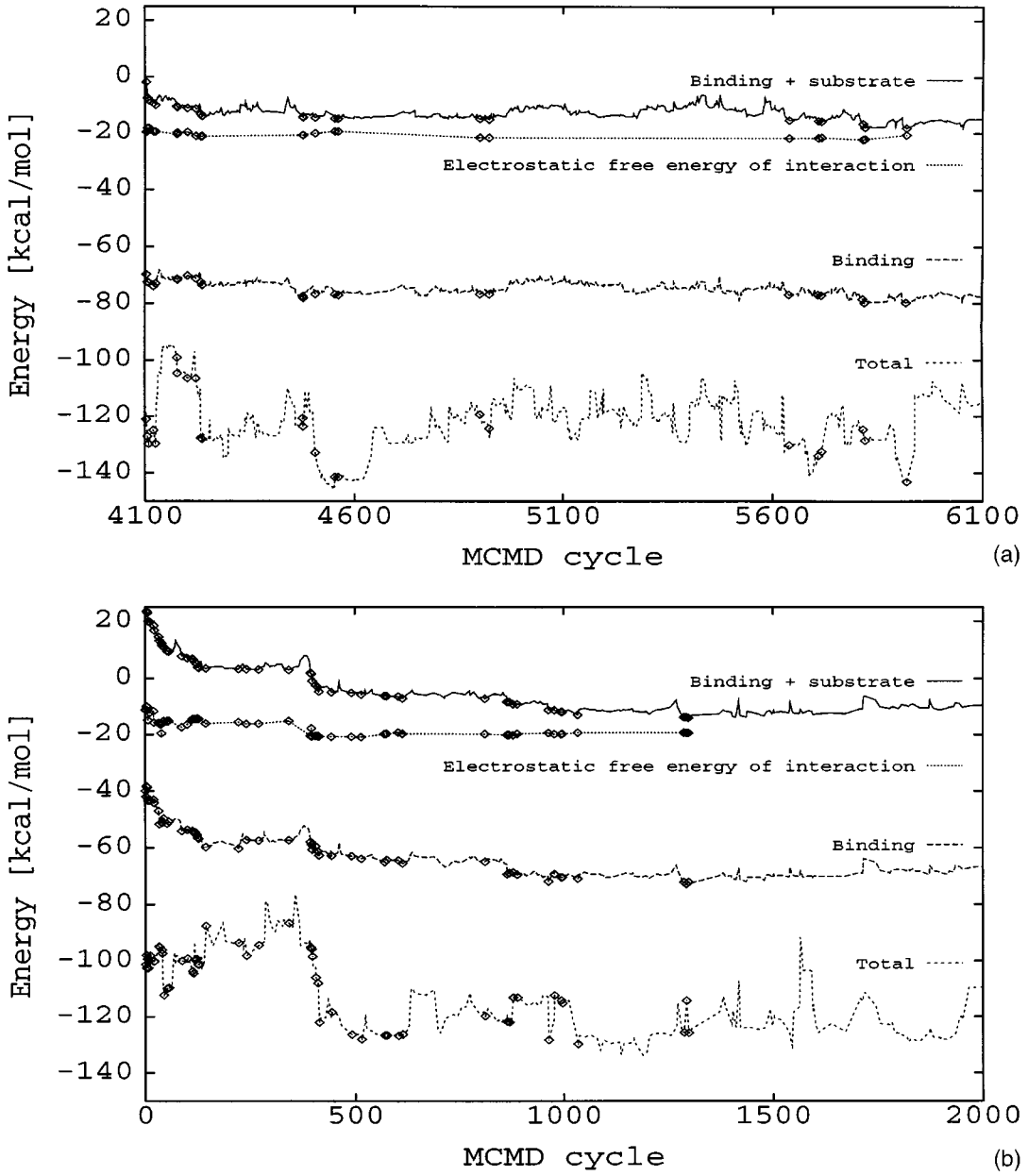

FIGURE 4. Energy values of the accepted conformations plotted as a function of Monte Carlo cycles. (—) Metropolis energy (intrasubstrate plus binding energy), (– – –) binding energy, (---) total energy. The diamonds correspond to the conformations that were saved, i.e., those characterized by the lowest Metropolis energy found up to a given MCM cycle. The dotted line is used for clarity sake and does not correspond to any calculated value, because the electrostatic free energy of interaction was calculated (by numerical solution of the finite-difference LPB equation) only for the conformations saved, i.e., those corresponding to the diamonds. (a) Third part of MCM run S1; (b) MCM run S2.

For the 355 minima a very good correlation ($R = 0.959$) was found between the intermolecular vacuum electrostatic energy (unscaled charges, constant dielectric, no cutoff) and the electrostatic free energy of interaction calculated by the LPB equation. The approximation used for the electrostatic free energy during MCM runs (scaled charges on FKBP residues, distance dependent dielectric, and 12 Å cutoff) yielded a very slightly better correlation ($R = 0.963$). Correspondingly, the range of the approximated energy with the scaled interaction (from -41.6 to -4.4 kcal/mol) was similar although larger than the LPB energy range (from -24.5 to -3.5 kcal/mol, see Fig. 5a, b). The vac-

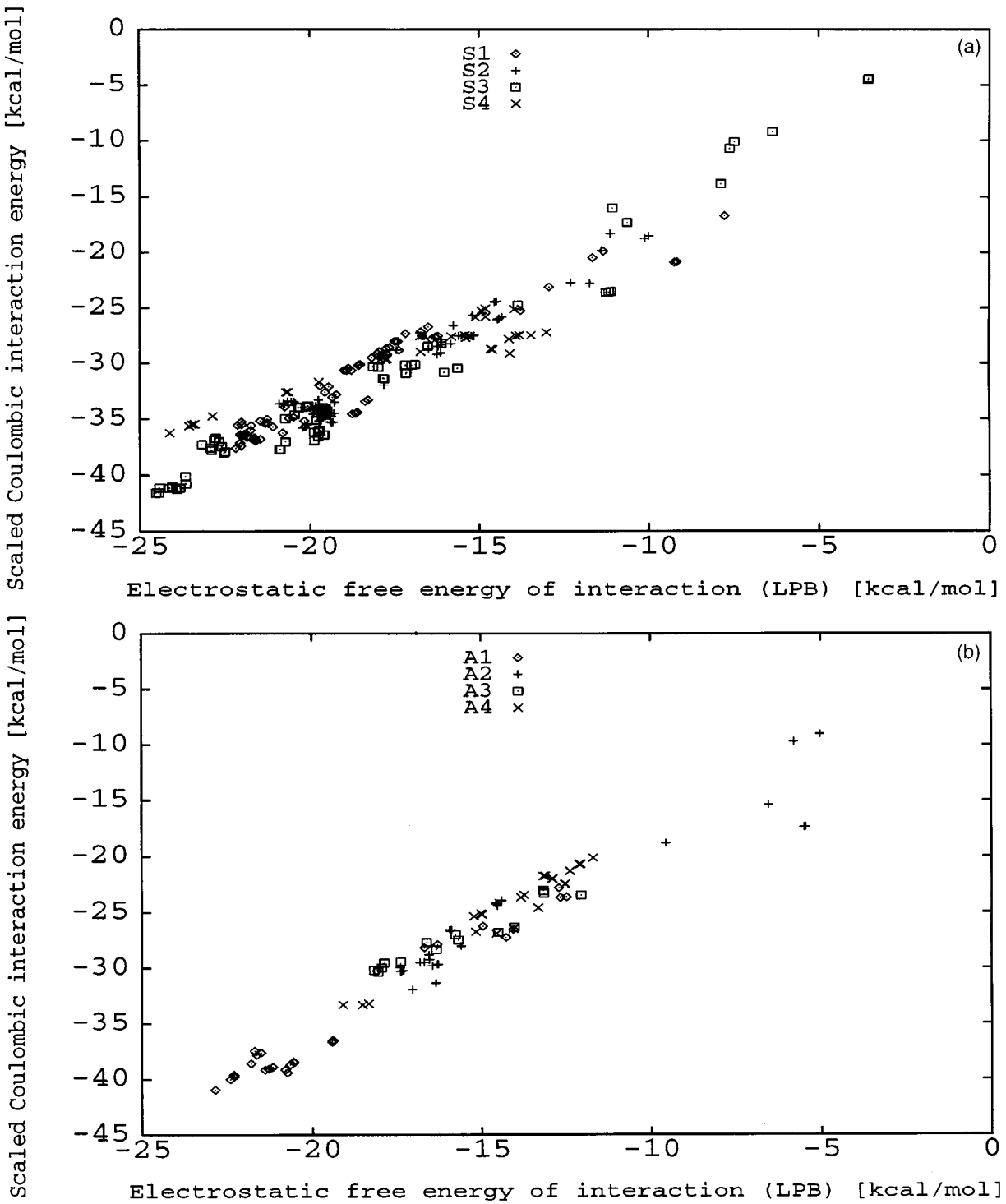

FIGURE 5. Correlation of the electrostatic contribution to the binding energy, as calculated by solving the finite-difference LPB equation or as calculated by using the residue-dependent dielectric constant (scaled FKBP charges in the Coulomb energy. (a) Minima obtained with the syn twisted amide conformer of the substrate. (b) Minima obtained with the anti twisted amide conformer of the substrate.

uum electrostatic energy range (from −53.6 to −1.7 kcal/mol) was still greater. Thus, introduction of the scaled charges for the MCM runs gave a somewhat better approximation than the Coulombic energy.

The solute–solvent dispersion interaction and the free energy of cavity formation showed a significantly smaller variation. For the 355 minima obtained by MCM, the former ranged from −524.8 to −506.7 kcal/mol, the latter from 357.3 to 369.9 kcal/mol, and their sum from −157.5 to −146.8 kcal/mol. For the 20 best conformers, the solute–solvent dispersion energy ranged from −514.3 to −510.3 kcal/mol, the cavitation potential from 359.7 to 363.5 kcal/mol, and their sum from −151.1 to −149.2 kcal/mol. This small variability was the result of the fact that all the conformers saved during docking by MCM had the tetrapeptide substrate almost equally buried in the binding pocket. The dispersion and cavity terms showed a high anticorrelation ($R = -0.892$). This is in accord with the observation that a complex with a partially exposed substrate will have more dispersive interactions with the solvent than a complex with a more buried substrate. The anticorrelation is responsible for the applicability of simpler models of the hydrophobic effect that neglect the dispersion term, and compensate for its absence by using a smaller value of the surface area parameter, common values are usually on the order of 0.007–0.025 kcal/Å^2.[46,47]

MINIMUM ENERGY CONFORMATIONS

Conformers with Lowest Free Energy

The last minimum energy conformation saved during run S1 (S1:89, Figs. 1b and 2a, thick lines) had the lowest total free energy among the 355 minima. It was found at step 5918 out of 6100; no lower energy minima were obtained after that. The tetrapeptide was bound as a type VIa proline turn with an intrasubstrate hydrogen bond between the C═O group of the *N*-acetyl moiety and the N–H of Phe. This brings the Phe N–H within hydrogen-bonding distance of the lone pair of the pyramidalized N of the substrate Pro. The S1:89 conformation was essentially identical to the one obtained by manual model building.[6] The latter was used in determining the catalytic mechanism of FKBP. The proline ring position relative to the FKBP binding site was very similar to that of the pipecolinyl ring of FK506[21,22] and interacted favorably with the indole of Trp 59.

The FKBP binding site obtained after the MC simulations was similar to the starting structure obtained from the complex with FK506 (see Fig. 2a). A small but significant difference (see Discussion) was the orientation of the Tyr 82 hydroxyl, which pointed outward when interacting with the Phe carbonyl group of the substrate, rather than inward as it does when interacting with the C8 imide carbonyl of FK506. There was also a rotation of about 30° of the Phe 46 aromatic ring, which moved to accommodate the phenyl ring of the substrate.

All of the seven potential hydrogen bond donors or acceptors of the tetrapeptide substrate were satisfied in S1:89, except for the imide carbonyl. There were four intermolecular hydrogen bonds, in addition to the aforementioned intrasubstrate hydrogen bond. Two of these formed a short stretch of antiparallel twisted β sheet between the enzyme and substrate backbones; they went from the proline carbonyl to the N–H of Ile 56 and from the C-terminal N–H of the tetrapeptide to the carbonyl of Glu 54. These two hydrogen bonds had electrostatic interaction energies of −3.3 and −3.0 kcal/mol, respectively, as calculated from the interaction of the C_α–NH group with the C═O group in the LPB calculation. The values were somewhat smaller than the vacuum CHARMM values (−4.9 and −4.8 kcal/mol, respectively). The other two hydrogen bonds were from the N–H of Leu to the side chain carboxyl of Asp 37 (−4.1 kcal/mol) and from the C═O of Phe to the hydroxyl group of Tyr 82 (−5.1 kcal/mol); the CHARMM values were −9.4 and −8.5 kcal/mol, respectively. All four enzyme groups that were partners of these hydrogen bonds to the substrate were involved in hydrogen bonds to FK506[21,22] and to the peptide mimic[4] in the corresponding cocrystal structure. Three of the five hydrogen bonds formed a chain across the binding site (those involving the Asp 37 side chain, the Ile 56 N–H, and the intramolecular hydrogen bond), with the dipole moments of all amide groups aligned favorably along the chain. The orientation of the N-terminal amide and the Phe amide of the substrate, required for making the hydrogen bonds with Asp 37 and Ile 56, stabilized the intrasubstrate hydrogen bond, which is characteristic of the type VIa turn structure of the tetrapeptide. Only the imide carbonyl of the substrate did not participate in strong hydrogen bonds, although it participated in weak polar interactions with the aromatic C–H groups of the aromatic triad Tyr 26, Phe 36, and Phe 99.[48] The LPB equation yielded an electro-

static free energy of interaction between FKBP and the imide carbonyl (partial charges of 0.51 on C and −0.51 on O) of −2.2 kcal/mol. In addition to the electrostatic interaction, S1:89 also had the best enzyme–substrate van der Waals energy (−45.1 kcal/mol) among the 355 minima; this originated mainly from the interactions of the Phe side chain of the substrate with the FKBP Phe 46 and Phe 48 aromatic rings, and from the optimal van der Waals interactions of the Leu side chain of the substrate with a pocket formed by the Phe 36, Tyr 82, His 87, and Phe 99 side chains of FKBP.

The conformer S1:81 had the second lowest free energy. It was only 0.6 kcal/mol higher than S1:89 (Table I). Figure 2b shows that S1:81 (thick lines) had a substrate main chain conformation very similar to that of S1:89 (thin lines). The main difference is at the *N*-acetyl terminus, which is slightly tilted toward the solvent in S1:81. The backbone RMS deviation between S1:81 and S1:89 was 0.63 Å. The FKBP binding site conformation was almost identical in S1:81 and S11:89. Also, the pattern of intermolecular hydrogen bonds was exactly the same. The most significant difference consisted of the orientation of the Phe side chain, which in S1:81 pointed toward the His 87 imidazole ring and the Leu side chain of the substrate, while in S1:89 it was oriented toward the aromatics rings of Phe 46 and Phe 48. As a consequence, S1:81 had 3.0 kcal/mol less intermolecular van der Waals stabilization than S1:89. This was compensated by the intrasubstrate van der Waals energy, which was more favorable by −2.5 kcal/mol in S1:81 than in S1:89, because of the additional interactions between the substrate Leu and Phe side chains in the former. In addition, the S1:81 continuum electrostatic free energy of solvation was 3.3 kcal/mol less favorable than the one of S1:89. This was partly compensated by the more favorable bonded energy [see eq. (2)] of the substrate and FKBP in S1:81; the values relative to S1:89 were of −1.4 kcal/mol for the substrate and −2.5 kcal/mol for FKBP. The Phe N–H still formed a hydrogen bond with the lone pair of the pyramidalized imide nitrogen as in S1:89. However, the distance between the hydrogen of the Phe amide and the oxygen of the N-terminus acetyl was too large for a hydrogen bond, 3.53 Å in S1:81 compared to 2.00 Å in S1:89. Hence, the intrasubstrate Coulombic energy was 2.4 kcal/mol less favorable in S1:81 than in S1:89. S1:81 and S1:89 had the same cavity formation energy, because the contact and reentrant surfaces were almost identical in the two conformations. The solute–solvent dispersion energy was 0.6 kcal/mol more favorable in S1:81 than in S1:89. This originated from the different orientation of the substrate Phe side chain, which was slightly more solvent exposed in S1:81 than in S1:89 (Fig. 2b).

S1:41, the conformer with the third lowest free energy, differed from S1:89 by 10.4 kcal/mol. It is shown in Figure 2c (thick lines) superimposed on the S1:89 structure (thin lines). The substrate main chain conformation was similar to S1:89 except for a rotation of the ϕ angle of Phe by about 45°. This resulted in a different position of the C-terminal methylamide, whose N–H group in S1:41 was not involved in a hydrogen bond with the backbone C═O of Glu 54, as in S1:89, but pointed toward the phenyl ring of the substrate. In addition, in S1:89 the substrate Phe side chain packed against the aromatic rings of Phe 46 and Phe 48, while in S1:41 it was directed more toward the solvent. The loss of one intermolecular hydrogen bond and the increased solvent exposure of the substrate Phe side chain accounted for the less favorable intermolecular energy (loss of 12.6 kcal/mol in *vacuum* electrostatic interaction and 4.8 kcal/mol in van der Waals stabilization with respect to S1:89) and continuum electrostatic free energy of solvation ($\Delta G_{\text{elect, solvat}}$ = 7.4 kcal/mol). This was partly compensated by a −7.1 kcal/mol more favorable intraenzyme energy and by a −3.7 kcal/mol more favorable intrasubstrate nonbond energy in S1:41 relative to S1:89. Also, the substrate Leu side chain in S1:89 was completely buried in the aromatic pocket formed by the Phe 36, Tyr 82, His 87, and Phe 99 side chains; in S1:41 one of the C_δ methyl groups of Leu was partially exposed.

Figure 2a–c shows that in the three conformations characterized by the lowest free energy (i.e., S1:89, S1:81, and S1:41), the N- and C-terminal methyl groups of the substrate were solvent accessible. Thus, a protein loop could bind with the same conformation as any one of these three tetrapeptide conformers. Moreover, the FKBP residues implicated in binding were in agreement with the observed effects of FKBP point mutations on the PPIase catalysis. Specifically, mutation of eight residues distal to the FK506 binding pocket (S8A, S38A, S39A, S67A, S77A, C22A, T75A, and T96A) did not alter rotamase activity.[49] These resides were not in contact with the substrate. Also the mutations H87A[50] and H87V[51] had no effect. The imidazole ring of His 87 was involved in a hydrophobic contact with the Leu side chain of the substrate, which would be expected to tolerate such mutations. The mutant D37V was inactive in

accord with the essential role of the Asp 37 side chain in binding and catalysis. The mutation of Tyr 82 to Phe was expected to alter the trans to cis catalysis, but no data are available.

Minima Found in Run S3

Some of the minima of run S3 are discussed here to show the relative importance of the different energy contributions and to obtain a clearer understanding of the sampling of conformational space by the various runs. A similar picture arose from the analysis of the remaining minima (i.e., S2, S4, and A1–A4); they are not presented to save space.

As mentioned above, the first minimum obtained in S3 (S3:1, Fig. 2d, thick lines) had the 19th lowest total free energy (19.8 kcal/mol above S1:89) and was the only non-S1 minimum among the 20 lowest free energy conformers. S3:1 had a poor electrostatic free energy of interaction (−7.9 kcal/mol), which originated from the fact that only three of its seven polar groups were involved in hydrogen bonds with FKBP (dotted lines in Fig. 2d). S3:77, the last conformer saved in the S3 MCM run (Fig. 2d, thin lines), had a total free energy difference [eq. (2)] 28.4 kcal/mol above S1:89 and ranked 69th among the 355 minima. It had the third best electrostatic free energy of interaction (−24.4 kcal/mol). Five among the seven polar groups of the substrate participated in hydrogen bonds with the enzymne in S3:77 versus only three in S3:1. Two of these, from the Neu N–H to the C═O of Glu 54 and from the imide carbonyl to the N–H of Ile 56, formed a short stretch of parallel twisted β sheet between the enzyme and substrate main chains, which, *mutatis mutandis*, is similar to the antiparallel β sheet in S1:89. The remaining three FKBP–substrate hydrogen bonds were from the C═O at the *N*-acetyl terminus to the Tyr 82 hydroxyl, from the C═O of Pro to the hydroxyl of Tyr 26, and from the C-terminal amide to the side chain carboxyl of Asp 37. The two polar groups of the substrate not involved in hydrogen bonds with FKBP, the N–H and C═O of Phe, were accessible to solvent.

The effect of the special choice of the Metropolis criterion energy (i.e., exclusion of the intraenzyme energy) was made clear by comparing S3:1 to S3:77 (Fig. 2d). To better accommodate the substrate and improve the enzyme–substrate interaction energy, some FKBP binding site side chains underwent major rearrangements during the S3 run. In S3:77 the Tyr 82 side chain moved toward the top of the binding site and its hydroxyl rotated by 180° to donate a hydrogen to the C═O of the substrate *N*-acetyl terminus. In S3:1 the *N*-acetyl C═O was involved in an intrasubstrate hydrogen bond with the Phe N–H, which was lost in S3:77. The intraenzyme hydrogen bond between the O–H of Tyr 26 and the side chain of Asp 37 was also lost in S3:77 because of the displacement of the Tyr 26 aromatic ring, which resulted in the additional intermolecular hydrogen bond with the C═O of the substrate Pro. The C-terminal N–H of the substrate made a hydrogen bond with the charged Asp 37 side chain in S3:77, while it donated its hydrogen to the side chain oxygen of Tyr 26 in S3:1. In addition, the Ile 56 side chain rotated around χ_1 by about 180° and minor displacements were found for the Phe 48, Trp 59, and His 87 side chains. These structural changes resulted in improvements in binding energy ($\Delta E = -33.4$ kcal/mol) and substrate energy ($\Delta E = -1.1$ kcal/mol) at the expense of the FKBP energy ($\Delta E = 39.1$ kcal/mol).

Concluding Discussion

A MCM procedure for docking ligands was implemented. It is rapid and includes the effect of solvent shielding of polar interactions. The resulting set of low energy structures was ranked by estimating their relative binding free energies with continuum models for the polar and nonpolar interactions. The electrostatic free energy of solvation was computed by solving the finite-difference LPB equation. To estimate the nonpolar contributions, the free energy of cavity formation was obtained from the molecular surface, while the solute–solvent dispersion energy was calculated with a model in which a continuous Lennard–Jones energy density was assigned to the bulk solvent. The electrostatic and hydrophobic solvation free energies can be used to complement intrasolute energies calculated by a vacuum molecular mechanics force field. The method was illustrated by the docking of an analogue of the preferred substrate, the blocked tetrapeptide *N*-acetyl-Leu-Pro-Phe-methylamide, in the binding site of FKBP. During the docking procedures both the tetrapeptide substrate and the FKBP binding site residues were flexible. Random perturbations were selected to alter the dihedral angles of the enzyme and the substrate and the relative orientation of the two molecules. To explore a large portion of the available conformational space, eight MCM docking

runs were done starting from four different orientations of the syn and anti substrate conformers.

Because the molecular complex was expected to mimic the catalytic transition state, it was assumed that the binding energy and intrasubstrate energy would play a dominant role in determining the structure. Consequently, to better focus on the important attributes of the complex, the intraenzyme contribution to the energy was not taken into account in the Metropolis criterion. This approach worked well for the present case but may not be appropriate for all molecular complexes. For other cases, the Metropolis criterion could be based on the total energy or a linear combinations of the different energy contributions, with weights determined by test calculations.

An improvement with respect to most previous studies was the use during the MC docking of charges for the enzyme binding site residues that were scaled individually to account for solvent screening effects evaluated by a preliminary PB calculation. The validity of this approximation was demonstrated by the high correlation between the resulting approximate electrostatic interaction energy (Coulombic energy with scaled FKBP charges and distance-dependent dielectric) and the actual electrostatic free energy of interaction (finite-difference LPB energy) of the 355 MCM minima sampled during docking.

Most of the recently published docking methods[52–54] make use of a rigid structure for the receptor. A degree of flexibility was introduced by Leach.[55] He allowed some rotatable bonds of the binding site side chains to take values from a data base of high-resolution protein structures. The MCM procedure introduces ligand flexibility and binding site flexibility and takes into account solvation effects. Solvation effects were also introduced by Zacharias et al.[20] who used an MC conformational search to investigate the effects of an operator mutation in the λ repressor–operator complex. They used the finite-difference LPB approach and the loss in solvent accessible surface to postprocess the conformations of the complex generated by the MC sampling. Only the dihedral angles of one side chain (Lys 4) in the λ arm were perturbed (bonds and bond angles were kept fixed) and because of the absence of a minimization step, only very small changes of the rotatable bonds were allowed. Although their search method is very efficient for local sampling, it is not expected to sample portions of conformational space that are significantly distant from the starting conformation. There are three advantages in the docking approach used in the present work with respect to the search procedure of Zacharias et al.[20] First, a more global search of conformational space is performed. Second, the approximate solvent screening effect in the electrostatic binding energy used during sampling yields a better correlation with the finite-difference LPB free energies of interaction determined in the postprocessing. Finally, a more accurate evaluation of the hydrophobic effect is included in postprocessing the minima sampled during MC docking.

From the analysis it is evident that a significant amount of the conformational space accessible to the FKBP-*N*-acetyl-Leu-Pro-Phe-methylamide complex was sampled in the present study. While it can never be claimed that the global minimum of the free energy was found, the evidence suggests that the S1:89 and the S1:81 conformers correspond to the structure of the complex between FKBP and its preferred substrate in the transition-state conformation. All nonimide donor and acceptor groups of the substrate were involved in hydrogen bonds, good intermolecular van der Waals energy was present, and the termini were solvent exposed.

In S1:89, the conformer characterized by the lowest total free energy, the tetrapeptide was bound as a type VIa proline turn with an intrasubstrate hydrogen bond, in which the C=O group of the *N*-acetyl moiety interacted with the main chain N–H of Phe. The substrate was found to fit into the FKBP binding pocket by positioning its imide group and proline ring onto the corresponding moieties of FK506, i.e., the C9 α-keto carbonyl and the six-membered pipecolinyl ring.[22] All of the seven potential donors or acceptors of the tetrapeptide were satisfied, except for the imide carbonyl which was involved in polar interactions with the aromatic side chains of FKBP residues Tyr 26, Phe 36, and Phe 99. The structure of the FKBP binding site was similar to its conformation in the complex with FK506. The main difference was the orientation of the Tyr 82 hydroxyl, which pointed outward when interacting with the Phe carbonyl group of the substrate rather than inward when interacting with the C8 imide carbonyl of FK506. This key change in the binding pocket was found by the MCM docking algorithm and then used in the catalytic mechanism study described by Fischer et al.[6] The second best conformer (S1:81) showed a similar binding mode with the same intermolecular hydrogen bonds, the main differences being the weakening of the intrasubstrate hydrogen bond and the orientation of the Phe side

chain. Four essential features of S1:89 were consistent with the crystal structure of a cyclic peptide–FK506 hybrid bound to FKBP[5]: the type VIa turn structure; the stretch of antiparallel β sheet between the peptide–FK506 hybrid and residues 54–56 of FKBP; the hydrogen bond between the hydroxyl group of Tyr 82 and the second amide down chain of the macrocyclic ring instead of to the pipecolinyl imide carbonyl as in the FKBP-FK506 cocrystal structure; and the lack of an FKBP hydrogen-bond donor group in the vicinity of the imide carbonyl. Essentially the same features were also found in S1:81, while the first, third, and fourth features were present in S1:41, the conformer with the third lowest free energy. In these three conformers, the N- and C-terminal methyl groups of the substrate were exposed to the solvent. Thus, a protein loop could bind with the same conformation as the one involved in the binding of the tetrapeptide substrate.

These results provide further support to the previously suggested mechanism for rotamase catalysis by FKBP.[6] The catalytic mechanism, which was solved by calculation of the reaction pathway,[56] involved lowering of the activation energy by a combination of desolvation of the imide carbonyl, ground-state destabilization, substrate autocatalysis, and preferential transition-state binding.

Because the methodology used to evaluate the solvation correction was relatively fast (it required a few minutes of CPU time) when compared to the minimization procedure, it would be possible to calculate it at the end of every conjugate gradient minimization and then incorporate it into the energy used for the Metropolis criterion. The utility of this methodological refinement is being examined (J. Apostolakis and A. Caflisch, in preparation).

While the MCM procedure has been mainly used to dock oligopeptide ligands into protein binding sites of known 3-dimensional structure, it could also be utilized to predict the conformation of a loop on a protein surface or to predict the readjustments needed upon mutations of one or more residues. In general, the combination of the MCM approach and the methodology for the accurate evaluation of solute–solvent interactions will be of benefit for the prediction of any molecular structure whenever extensive sampling of conformational space has to be performed and the effect of the solvent cannot be neglected. In addition, the computational approach described here might be useful for refining the structure of a molecular complex obtained by X-ray crystallography, particularly when only low resolution data are available.

Acknowledgments

This work was supported by a grant from the National Institutes of Health. A. Caflisch was supported by the Swiss Federal Office of Public Health (Nationales Aids-Forschungs-Programm, Grant 3139-043652.95) and the Swiss National Science Foundation (Schweizerischer Nationalfonds Grant 3100-043423.95). We are indebted to Michael Sommer who calculated the pK_a values of the histidines. We thank Joannis Apostolakis and Dr. Michael Schaefer for helpful discussions, Dr. J. A. McCammon for providing the 4.1 release of the UHBD program, and Dr. A. Šali for version 1.1 of the ASGL program used to generate Figure 3. The calculations were performed on a DEC Alpha 3000, a CONVEX C220, a SGI 4D/340, and an IBM RISC/6000 550. Inquiries concerning the FORTRAN program used to generate the MCM input script for CHARMM should be addressed to A. Caflisch at the Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland.

References

1. J. J. Siekierka, H. Y. Hung, M. Poe, C. S. Lin, and N. S. Sigal, *Nature*, **341,** 755 (1989).
2. M. W. Harding, A. Galat, D. E. Uehling, and S. L. Schreiber, *Nature*, **341,** 758 (1989).
3. J. Clardy, Proc. Natl. Acad. Sci. USA, **92,** 56 (1995).
4. P. Kraulis, J. Appl. Crystallogr., **24,** 946 (1991).
5. Y. Ikeda, L. W. Schultz, J. Clardy, and S. L. Schreiber, J. Am. Chem. Soc., **116,** 4143 (1994).
6. S. Fischer, S. Michnick, and M. Karplus, *Biochemistry*, **32,** 13830 (1993).
7. S. L. Schreiber, *Science*, **251,** 283 (1991).
8. S. L. Schreiber, M. W. Albers, and E. J. Brown, Acc. Chem. Res., **26,** 412 (1993).
9. S. L. Screiber and G. R. Crabtree, Immunol. Today, **13,** 136 (1992).
10. G. Fischer, B. Wittmann–Liebold, K. Lang, T. Kiefhaber, and F. X. Schmid, *Nature*, **337,** 476 (1989).
11. M. Tropschug, E. Wachter, S. Mayer, E. R. Schönbrunner, and F. X. Schmid, *Nature*, **346,** 674 (1990).
12. J. F. Brandts, H. R. Halvorson, and M. Brennan, *Biochemistry*, **14,** 4953 (1975).
13. R. K. Harrison and R. L. Stein, *Biochemistry*, **29,** 1684 (1990).
14. M. W. Albers, C. T. Walsh, and S. L. Schreiber, J. Org. Chem., **55,** 4984 (1990).
15. A. Caflisch, P. Niederer, and M. Anliker, Proteins: Struct. Funct. Genet., **13,** 223 (1992).

16. A. Caflisch, A. Miranker, and M. Karplus, J. Med. Chem. **36,** 2142 (1993).
17. Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84,** 6611 (1987).
18. Z. Li and H. A. Scheraga, J. Mol. Struct. Theochem. **179,** 333 (1988).
19. A. Caflisch, P. Niederer, and M. Anliker, Proteins: Struct. Funct. Genet., **14,** 102 (1992).
20. M. Zacharias, B. A. Luty, M. E. Davis, and J. A. McCammon, J. Mol. Biol., **238,** 455 (1994).
21. G. D. Van Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, and J. Clardy, *Science*, **252,** 839 (1991).
22. G. D. Van Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, and J. Clardy, J. Mol. Biol. **229,** 105 (1993).
23. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4,** 187 (1983).
24. S. Fischer, R. L. Dunbrack, and M. Karplus, J. Am. Chem. Soc. **116,** 11931 (1994).
25. D. Bashford and M. Karplus, *Biochemistry*, **29,** 10219 (1990).
26. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys., **21,** 1087 (1953).
27. M. R. Hestenes and E. Stiefel, J. Res. Natl. Bur. Stand., **49,** 409 (1952).
28. R. Fletcher and C. M. Reeves, Comp. J., **7,** 149 (1964).
29. S. W. Michnick, M. K. Rosen, T. J. Wandless, M. Karplus, and S. L. Schreiber, *Science*, **252,** 836 (1991).
30. R. M. Jackson and M. J. E. Sternberg, Protein Eng., **7,** 371 (1994).
31. J. Warwicker and H. C. Watson, J. Mol. Biol., **157,** 671 (1982).
32. J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak, Faraday Symp. Chem. Soc., **17,** 55 (1982).
33. S. Fischer, A. Caflisch, and M. Karplus, unpublished results.
34. D. E. Smith, L. Zhang, and A. D. J. Haymet, J. Am. Chem. Soc., **114,** 5875 (1992).
35. M. K. Gilson and B. H. Honig, Proteins: Struct. Funct. Genet., **4,** 7 (1988).
36. M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, Comp. Phys. Commun., **62,** 187 (1991).
37. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran*, Cambridge University Press, New York, 1992, p. 753.
38. M. E. Davis and J. A. McCammon, J. Comp. Chem., **10,** 386 (1989).
39. M. E. Davis and J. A. McCammon, J. Comp. Chem., **11,** 401 (1990).
40. M. E. Davis and J. A. McCammon, J. Comp. Chem., **12,** 909 (1991).
41. C. Lim, D. Bashford, and M. Karplus, J. Phys. Chem., **95,** 5610 (1991).
42. D. T. Edmonds, N. K. Rogers, and M. J. E. Sternberg, Mol. Phys. **52,** 1487 (1984).
43. M. K. Gilson, K. A. Sharp, and B. H. Honig, J. Comp. Chem., **9,** 327 (1988).
44. V. Mohan, M. E. Davis, J. A. McCammon, and B. M. Pettitt, J. Phys. Chem., **96,** 6428 (1992).
45. B. A. Luty, M. E. Davis, and J. A. McCammon, J. Comp. Chem., **13,** 768 (1992).
46. C. Chothia, *Nature*, **248,** 338 (1974).
47. W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc., **112,** 6127 (1990).
48. G. R. Desiraju, Acc. Chem. Res., **24,** 290 (1991).
49. S. T. Park, R. A. Aldape, O. Futer, M. T. DeCenzo, and D. J. Livingston, J. Biol. Chem., **267,** 3316 (1992).
50. D. Yang, M. K. Rosen, and S. L. Schreiber, J. Am. Chem. Soc., **115,** 819 (1993).
51. R. A. Aldape, O. Futer, M. T. DeCenzo, B. P. Jarrett, M. A. Murcko, and D. J. Livingston, J. Biol. Chem., **267,** 16029 (1992).
52. I. D. Kuntz, E. C. Meng, and B. K. Shoichet, Acc. Chem. Res., **27,** 117 (1994).
53. K. P. Clark and Ajay, J. Comp. Chem., **16,** 1210 (1995).
54. R. S. Judson, Y. T. Tan, E. Mori, C. Melius, E. P. Jaeger, A. M. Treasurywala, and A. Mathiowetz, J. Comp. Chem., **16,** 1405 (1995).
55. A. R. Leach, J. Mol. Biol., **235,** 345 (1994).
56. S. Fischer and M. Karplus, Chem. Phys. Lett., **194,** 252 (1992).