ELSEVIER

Biophysical Chemistry 111 (2004) 143–157

Biophysical Chemistry

www.elsevier.com/locate/bpc

Conformational energy landscape of the acyl pocket loop in acetylcholinesterase: a Monte Carlo-generalized Born model study

Louis Carlacci[a], Charles B. Millard[b], Mark A. Olson[b,*]

[a] *Army High Performance Computing Research Center, Network Computing Services, Inc., 1425 Porter Street, Frederick, MD 21702, USA*
[b] *Department of Cell Biology and Biochemistry, USAMRIID, 1425 Porter Street, Frederick, MD 21702, USA*

Received in revised form 14 May 2004; accepted 17 May 2004
Available online 24 June 2004

Abstract

The X-ray crystal structure of the reaction product of acetylcholinesterase (AChE) with the inhibitor diisopropylphosphorofluoridate (DFP) showed significant structural displacement in a loop segment of residues 287–290. To understand this conformational selection, a Monte Carlo (MC) simulation study was performed of the energy landscape for the loop segment. A computational strategy was applied by using a combined simulated annealing and room temperature Metropolis sampling approach with solvent polarization modeled by a generalized Born (GB) approximation. Results from thermal annealing reveal a landscape topology of broader basin opening and greater distribution of energies for the displaced loop conformation, while the ensemble average of conformations at 298 K favored a shift in populations toward the native by a free-energy difference in good agreement with the estimated experimental value. Residue motions along a reaction profile of loop conformational reorganization are proposed where Arg-289 is critical in determining electrostatic effects of solvent interaction versus Coulombic charging.
Published by Elsevier B.V.

Keywords: Structural reorganization; Free-energy calculation; Electrostatics; Hydration; Implicit solvent model; Entropy

1. Introduction

Structural details of the conformational repertoire available to the active-site gorge of acetylcholinesterase (AChE) are beginning to emerge from X-ray crystallographic studies of reaction products of the enzyme with organophosphate inhibitors [1,2]. Because of their variation in molecular size, these inhibitors probe the redistribution of populations of conformational isomers away from the native state of AChE. Smaller inhibitors, such as *O-iso*proplymethylphosphonofluoridate (sarin) and *O*-pinacolymethylphosphonofluoridate (soman), react with AChE without significant conformational change in the protein, whereas the more bulky di*iso*propylphosphorofluoridate (DFP) produces large-scale movement of ~ 5 Å in the main chain of a loop segment (residues 287–290) in the acyl pocket. The alternative loop conformation is triggered by disruption of a hydrophobic network in the reaction site, yet crystallographic *B*-factors for the unbound AChE structure suggest minimal structural plasticity in accommodating DFP. Comparison of reversible dissociation constants measured for the inhibitors [1,3–7] offers an experimental estimate of 4 kcal/mol as the free energy of loop reorganization and indicates weak dynamic connectivity between native and DFP-reacted macrostates.

Understanding conformational selection of ligand binding to AChE requires knowledge of energy landscapes. Energy landscapes provide useful insights into alternative and intermediate macrostates, folding and binding pathways, and for relating conformational fluctuations to biological function [8,9]. Here, we survey the energy landscape of the acyl pocket loop by using Monte Carlo (MC) simulation methods. For generating the native and alternative conformational ensemble, a computational strategy was applied of a combined simulated annealing and room temperature multiple-copy Metropolis sampling approach [10,11] with solvent polarization modeled by a generalized Born (GB) approximation. While free energies can be evaluated from microscopic simulations with explicit treatment of solvent by the application of linear response approximation [12–14], the calculations become computationally demanding for large macromolecular systems. Im-

* Corresponding author. Tel.: +1-301-619-4236; fax: +1-301-619-2348.
E-mail address: molson@ncifcrf.gov (M.A. Olson).

0301-4622/$ - see front matter. Published by Elsevier B.V.
doi:10.1016/j.bpc.2004.05.007

plicit solvent schemes based on GB models provide an efficient computational framework for sampling potential energy surfaces and have been successfully applied to a wide range of problems, including modeling of protein loops [15], protein structure predictions [16–18], molecular dynamics simulations of nucleic acids [19,20], and protein folding [21], among others. Formulations of GB models range from grid-based approaches (see, e.g., Refs. [22,23]) to analytical methods (e.g., Refs. [23,24]). We selected a simple analytical pairwise atomic summation scheme for obtaining the Born radii and its parameterization given by Qiu et al. [25]. Non-electrostatic components to the free energy of hydration were calculated by using a solvent accessibility (SA) model. Combined, our strategy of a MC and GB/SA approach predicts free-energy structures in an iterative fashion of annealing and optimization, and takes advantage of folding funnels in determining macrostates.

The effect of compensation among individual free-energy components was analyzed in ranking the basins. We applied this strategy one step further and dissected the electrostatic free energies into residue contributions. This provides a free-energy 'signature' for evaluating which residues of the loop make significant electrostatic contributions to the native and alternative states [14,26], and should aid in the design of new ligands that explore greater conformational permissiveness of the reaction site. For molecular docking of ligands to AChE, the culling of structures from thermal annealing into basins provides the correct starting approach. Central to modeling binding affinities, the calculation of reorganization free energies is key to reliable and accurate predictions of complex formation [12–14]. We believe that the experimental estimate of the free energy of structural reorganization for the acyl pocket loop of AChE offers an excellent benchmark for testing different methodological approaches to conformational analysis of protein structures.

2. Theory

Transitions between two loop conformations (denoted as $\mathcal{P}$ and $\mathcal{P}$*) located in different basins can be described by the reaction

$$\mathcal{P} \overset{\Delta G}{\rightleftharpoons} \mathcal{P}^*, \tag{1}$$

where ΔG is the free energy difference of the protein and solvent system, defined as a function of n degrees of freedom, $\zeta=[\zeta_1, \zeta_2, \ldots, \zeta_n]$. Stable states are determined by the minimum value of $G(\zeta)$, found by exploring over all possible values of ζ on the energy landscape. We define the free energy $G(\zeta)$ for a macrostate ξ as (see, e.g., the work of Dill [8])

$$G_{\mathrm{macro}}(\xi) = -k_{\mathrm{B}}T\ln\Omega = G_{\mathrm{micro}}(\xi) - k_{\mathrm{B}}T\ln g(\xi), \tag{2}$$

where Ω is the partition function, G_{micro} is the free energy of a single loop conformation taken to be the most probable microstate, the function $g(\xi)$ describes the multiplicity of microstates of similar energy in basin ξ, T is the absolute temperature and k_{B} is the Boltzmann constant. The conformational entropy is given by $S_{\mathrm{conf}}(\xi)=k_{\mathrm{B}}\ln g(\xi)$, and Eq. (2) becomes

$$G_{\mathrm{macro}}(\xi) = G_{\mathrm{micro}}(\xi) - TS_{\mathrm{conf}}(\xi). \tag{3}$$

The full energy landscape can be reduced in size by a projection-operator decomposition of the partition function, leading to a simulation volume of reduced degrees of freedom. The remaining degrees of freedom are assumed to provide a thermal reservoir and can be averaged out [27]. The free energy defined by Eq. (3) corresponds to an ensemble of conformations that collectively have some common subset of ζ. Examples of macrostates include reaction coordinate, cluster of dihedral angles, hydrophobic contacts, etc.

To determine G_{micro}, we partition the free energy for each conformation into separate contributions; for example, conformation $\mathcal{P}$ gives

$$G^{\mathcal{P}}_{\mathrm{micro}} = E^{\mathcal{P}}_{\mathrm{int}} + G^{\mathcal{P}}_{\mathrm{solv}} + G^{\mathcal{P}}_{\mathrm{cav}}, \tag{4}$$

where E_{int} is the internal energy of the protein, G_{solv} is the solvation term, and G_{cav} is the cavitation free energy. The internal energy is given by

$$E^{\mathcal{P}}_{\mathrm{int}} = E^{\mathcal{P}}_{\mathrm{local}} + E^{\mathcal{P}}_{\mathrm{vdW}} + W^{\mathcal{P}}_{\mathrm{ele}}, \tag{5}$$

where E_{local} is the bonded and torsional terms defined by a given force field, E_{vdW} is van der Waals (vdW) intraprotein interactions, and W_{ele} is the energetic cost of creating the charge distribution in an environment of unit dielectric.

Solvent effects are modeled as the sum of non-polar and electrostatic contributions

$$G^{\mathcal{P}}_{\mathrm{solv}} = G^{\mathcal{P}}_{\mathrm{s,np}} + \Delta G^{\mathcal{P}}_{\mathrm{s,ele}}, \tag{6}$$

determined for each microstate. Our treatment of the non-polar term is an effective SA atomic solvation potential developed by combining $G_{\mathrm{s,np}}$ with G_{cav}

$$G^{\mathcal{P}}_{\mathrm{SA}} = G^{\mathcal{P}}_{\mathrm{s,np}} + G^{\mathcal{P}}_{\mathrm{cav}} = \sum_i \sigma_i A^{\mathcal{P}}_i, \tag{7}$$

where for atom i, σ_i is a fitted scaling parameter [25] and A_i is the solvent accessible surface area.

The electrostatic component of solvation energy is calculated from a GB model [25] and is defined by the equations

$$\Delta G^{\mathcal{P}}_{\mathrm{s,ele}} = \Delta G^{\mathcal{P}}_{\mathrm{GB}} = \sum_i [(\Delta G^{\mathcal{P}}_{\mathrm{scr}})_i + (\Delta G^{\mathcal{P}}_{\mathrm{pol}})_i], \tag{8}$$

where the screening term is given by

$$(\Delta G^{\mathcal{P}}_{\mathrm{scr}})_i = -166.0\left(\frac{1}{\varepsilon_{\mathrm{p}}} - \frac{1}{\varepsilon_{\mathrm{w}}}\right)\sum_j \frac{q_i q_j}{f^{\mathrm{GB}}_{ij}(r_{ij})} \tag{9}$$

and the polarization term

$$(\Delta G^{\mathcal{P}}_{\mathrm{pol}})_i = -166.0\left(\frac{1}{\varepsilon_{\mathrm{p}}} - \frac{1}{\varepsilon_{\mathrm{w}}}\right)\frac{q_i^2}{\alpha_i}, \tag{10}$$

with f^{GB}_{ij} is a function which depends only upon atomic radii and interatomic distances r_{ij}, q_i and q_j are atomic charges for atom $i(j)$, α_i are Born radii, and ε_{p} and ε_{w} are dielectric constants for the protein and solvent, respectively.

Our estimate of the conformational entropy due to multiplicity of microstates in basin ξ follows the work of Vorobjev et al. [28]. This term can be estimated in the quasiharmonic approximation calculated from the covariance fluctuation matrix $\mathbf{C}$ [28–32]

$$\mathbf{C}^{\mathcal{P}}_{ij} = \langle(\mathbf{x}^{\mathcal{P}}_i - \langle\mathbf{x}^{\mathcal{P}}_i\rangle)(\mathbf{x}^{\mathcal{P}}_j - \langle\mathbf{x}^{\mathcal{P}}_j\rangle)\rangle_{\xi}, \tag{11}$$

where $\mathbf{x}$ are mass-weighted atomic coordinates of microstates of $\mathcal{P}$ sampled from ξ. Frequencies υ_i of the normal modes are defined by the eigenvalues λ_i

$$(2\pi\upsilon_i)^2 = k_{\mathrm{B}}T/\lambda_i, \tag{12}$$

and entropy $S_{\mathrm{conf}}(\xi)$ is approximated by [28]

$$TS_{\mathrm{conf}}(\xi) = \sum_i -k_{\mathrm{B}}T\ln[1-\exp(-h\upsilon_i/k_{\mathrm{B}}T)] + h\upsilon_i/[1-\exp(-h\upsilon_i/k_{\mathrm{B}}T)], \tag{13}$$

where h is Planck's constant.

The probability distribution function of normal mode m is obtained from the projections of the fluctuation of a microstate Cartesian coordinates onto a normal mode eigenvector m, which is given by

$$q_m = (\mathbf{x}^{\mathcal{P}} - \langle\mathbf{x}^{\mathcal{P}}\rangle)\cdot\mathbf{w}^{\mathcal{P}}_m, \tag{14}$$

where $\mathbf{w}^{\mathcal{P}}_m$ is the eigenvector of normal mode m. The probability distribution function is used to determine the probability that the fluctuation of the microstate Cartesian coordinates is found at a position q_m.

3. Computational methods

3.1. Protein simulation volume

Calculation of the ensemble of loop conformations started with the Cartesian coordinates of AChE taken from the crystallographic PDB file 2ACE [33]. Crystallographic waters were ignored in all calculations. For comparison purposes, the alternative loop conformation was taken from the 2DFP structure [1]. To model a common structure differing only in the loop conformation, the alternative loop was built into 2ACE by substitution of the native and optimized by a MC simulation (described below).

The geometry of the protein was generated in the ECEPP/2 reference frame; namely, bond lengths and bond angles were held fixed at experimental values for amino acids and dihedral angles may vary [34]. The exception was the peptide bond between the computed segment and the fixed part of the protein (see Section 3.3). The termini residues and the ionizable side chains were in the ionized state. The geometry of the loop and select side chains of the surrounding local region (defined below) were first generated from dihedral angles and then positioned in the protein by the use of rigid-body coordinates [11].

The simulation volume in the loop prediction calculations was generated from a shell of residues that were within 12-Å distance of residues 287–290 in the optimized 2ACE X-ray structure. Additional residues were included by graphical visualization to select interactions that may contribute to non-native states. The latter was beneficial in keeping the simulation size to a tractable problem. Listed in Table 1 are simulation volumes used for various stages of the calculation. The simulation volume consisted of the computed segment (CS) residues for which backbone and side-chain dihedral angles were allowed to vary, surrounding local region (SLR) residues for which no backbone movement was allowed, and SLR residues where side chains vary (CS + SLR). The simulation volume employed in calculations to predict the loop macrostates (SV I in Table 1) was smaller than the simulation volume employed in calculations to compute thermodynamic properties (SV II in Table 1). In calculations that employed simulation

Table 1
Simulation volume at various stages of the calculation

SV[a]	CS[b]	SLR[c]	CS + SLR[d]
SV I	287–289 Ile–Phe–Arg–Phe	69–72, 84, 85, 118–122, 151, 199–201, 226–227, 231–237, 240, 243–244, 275–286, 291–293, 327–337, 354, 355, 357–363, 391, 395, 396, 398–400, 432, 436, 439, 440, 442	Tyr-121, Trp-279, Leu-282, Phe-284, Ser-286, Phe-330, Phe-331
SV II	SLR segments of SV I	residues within 12 Å from a computed segment	none

[a] See text for information on the use of simulation volume (SV).
[b] Computed segment (CS) containing listed residues.
[c] Surrounding local region (SLR) residues.
[d] CS + SLR includes flexible side chains of SLR.

volume I, the computed segment was the four-residue loop segment of residues 287–290. In calculations that employed simulation volume II, the backbone and side-chain dihedral angles of all the residues of simulation volume I were allowed to vary.

3.2. Monte Carlo methods

The results of this study were generated from calculations to predict conformations of macrostates, to refine the predicted conformations, to compute room temperature thermodynamic properties and to find the energy minima of macrostates. Calculations to predict the macrostates consisted of multiple independent Monte Carlo simulated annealing (MCSA) runs starting from random conformations. Prediction calculations were designed to compute the lowest-energy state as well as alternative low-energy states, which is the principle reason for the use of the methods. Refinement calculations consisted of Metropolis MC runs starting from the microstate determined as the lowest-energy conformation (LEC) of a macrostate. The production calculation consisted of multiple-copy room temperature MC runs starting from the macrostate LECs of the refinement calculation. Multiple-copy low temperature MCSA runs starting from the macrostate LECs of the production calculation accomplished energy minimization.

Conformational search of the acyl pocket loop was computed by the use of the molecular mechanics program *Salp* [11]. Trial conformations were generated by perturbation of dihedral angles and rigid-body coordinates, and the total energy was computed. Acceptance or rejection of conformations was in accordance with the Metropolis criteria [35]. A MC run consisted of a number of steps (nstep), in which a specified number of trial conformations (nconf) were sequentially evaluated. In the MCSA prediction calculation, the Boltzmann temperature factor (*RT*, where *R* is the universal gas constant) was lowered at the end of each MC cycle, and a two exponential cooling schedule was used to decrease the value of *RT* [11]. Table 2 summarizes these parameter assignments for the MCSA prediction calculation. In exponential cooling from $RT = 7.0$ to 4.0, the folding simulation removed bad contacts and established important interactions. In exponential cooling from $RT = 4.0$ to 0.6, important interactions were optimized to improve the scoring of conformations. In the MC runs, *RT* was constant. The protocol parameters for the room temperature MC runs and MCSA energy minimization runs are given in the top of Table 3. Parameters describing the acceptance ratio are listed in the bottom of Table 3. Convergence of the energy minimization calculations was achieved when the total energy in successive iterations of MCSA runs changed by less than about 0.1%.

For the starting conformation, backbone and side-chain dihedral angles were randomly assigned from a rotamer library and backbone ω angles were set to 180. The loop *N*-terminal dummy residue was initially superimposed on the overlapping residue of the fixed part of the protein. Additional protocol parameters and methodology were employed as described in a previous study [11]. Details are given for parameters used to randomly generate a trial conformation; the use of a main-chain and side-chain dihedral angle rotamer library at high temperature factors of the MCSA; option to alter the rate that main-chain and side-chain dihedral angles were selected; and convergence criteria to stop a MCSA run when the *N*- and *C*-termini of the loop fail to properly connect with the rest of the protein.

Table 2
Summary of parameters for MCSA prediction runs

Parameter[a]	Equilibration stage	Prediction stage
RT main chain	7.0 to 4.0	4.0 to 0.6
RT side chain	5.0 to 2.0	2.0 to 0.6
nstep	15	10
nconf	4000 (10000 in step 1)	4000
Parameter[a]	**Equilibration and prediction stages**	
nprchk	400 (500 in step 1)	
pert	± 90° backbone, ± 120° side chain ± 45° Euler angles, ± 0.5 Å translations	
xacrat	35% for main-chain and 40% for side-chain trial conformations 35% for both Euler angle and translation trial conformations	
pertrd	1.5 for main-chain, side-chain and rigid-body coordinates	

[a] *RT* is the Boltzmann temperature factor in units of kcal/mol, where *R* is the universal gas constant and *T* is the absolute temperature; nstep is the number of steps; nconf is the number of trial conformations evaluated in a step; nprchk is the number of trials conformations evaluated before the acceptance ratio was checked; pert is the initial perturbation used to generate trial conformations; xacrat is the acceptance ratio limit; and pertrd is the factor used to lower the perturbation.

3.3. Energy determinations

The vdW, electrostatic and torsion potentials were evaluated with the OPLS united atom parameter set [36]. The use of ECEPP/2 standard geometries of amino acids is consistent with the development of the OPLS united atom parameter set. The equilibrium bond lengths and bond angles of amino acids in the OPLS set are from experimental values. For the backbone, the differences in bond lengths between the two geometric parameter sets are about 0.01 Å and the bond angle differences are about 1°. Except for a few side chains, the differences are about 0.02 Å and 2°, for bond lengths and bond angles, respectively. The difference in the C^{δ}–N^{ϵ} bond length of His is 0.07 Å. Bond angle differences are 5°, for two bond angles in the Trp ring, and one bond angle each in the side chains of Thr and Met. In the program *Salp*, which originated from

Table 3
Summary of parameters for room temperature Monte Carlo runs and MCSA energy minimization runs

Parameter[a]	Pre-equilibration	Equilibration	Production	Pre-equilibration	Equilibration + production
RT	0.7 to 0.6	0.6	0.6	0.6 to 0.2	0.2
nconf[b]					
I	10K[c]	620K	–	–	–
II	10K	50K	60K	–	–
III	–	18K	48K	–	–
IV	–	–	–	80K	320K

Parameter[a]	Equilibration and production
nprchk	200
nprchk	200
pert	± 2° backbone, ± 22.5° side chain, ± 0.5° Euler angles, ± 0.05 Å translations
xacrat	40% for main-chain and 45% for side-chain trial conformations 40% for both Euler angle and translation trial conformations
pertrd	1.4 for main-chain, side-chain and rigid-body coordinates

[a] Protocol parameters employed during various stages of low temperature MC runs (see Table 2 and text).
[b] I: Room temperature Metropolis MC refinement part of the prediction calculation. II: Room temperature Metropolis MC production calculation. III: Same as II but starting conformation was the lowest-energy microstate with the side-chain conformation of Trp-279 adjusted and optimized to assume the X-ray structure. IV: MCSA energy minimization calculation.
[c] 10K = 10,000 conformations.

ECEPP, the torsion potential employed a single Fourier term that corresponded to the barrier height. In the original OPLS parameter set, torsion potentials that have terms that biased the equilibrium states and a term for the barrier height are: C–S–S–C of disulfide bond, H–C–C–N and C–C–C–N of primary amine, and O–C–N–H of peptide bond. Test calculations that employ the full Fourier expansion indicated that the error in the present implementation is negligible.

A vacuum dielectric constant ($\varepsilon_p = 1$) was applied in all simulations. For protein segments that varied, the local region energy was the sum of all intra-segment and inter-segment energies plus the sum of all interaction energies with the surrounding local region segments. For the non-polar term of the solvation free energy defined by Eq. (7), solvent accessible surface area was computed with MSEED [37] with a 1.4-Å probe radius.

The term ΔG_{GB} of Eq. (8) was computed with $\varepsilon_p = 1$ and $\varepsilon_w = 80$ and the OPLS derived united atom radii by Qui et al. [25], which was developed and optimized to reproduce atom polarization energies for atoms of small organic molecules and biopolymers. The optimization was independent of partial atomic charges. In the present study, OPLS united atom charges were employed in the computation of ΔG_{GB}. Energy terms constant from the use of fixed bond lengths and bond angles were taken as atomic volume, self, 1–2 and 1–3 interactions. These terms were computed only once and by the use of the Cartesian coordinates of the optimized X-ray conformation of the protein. For the simulation volume case, small variations of the 1–2 and 1–3 contributions to the atom polarization energy due to deviations of the internal coordinates of non-fixed *N*- and *C*-terminal peptide bonds of segments were neglected. For fixed segments in the surrounding local region and segments outside the simulation volume, 1–4 contributions to the atom polarization energy were constant. The constant parts of the atom polarization energy for each trail conformation were added to the terms that vary; namely, interactions between segments that move and other segments in the simulation volume. Atoms in the simulation volume were considered in the GB solvent polarization energy calculation.

A weighted harmonic constraint [11] was added to Eq. (4) to force the loop to dock against the protein, and was based on the overlap of dummy residues appended to the *N*- and *C*-termini of the CS with identical residues immediately adjacent in the fixed SLR. Atoms overlapped were backbone heavy atoms. In the beginning of MCSA prediction simulations, the segment *N*- and *C*-terminal peptide bonds were severed from the rest of the protein to facilitate the removal of bad contacts in the initial loop conformation. In these calculations, the constraint weight was gradually introduced by using published protocols [11]. For MC runs and MCSA energy minimization runs, the weight was set to one and the loop-closing constraint was the sum of peptide bond stretching, angle bending and improper dihedral angle-bending energy terms for the termini of each computed segment [36].

3.4. Separation of microstates

For clustering the calculated structures and their scoring by Eq. (3), we differentiated the macrostates ξ by use of main-chain $\varphi\psi$ rotamer tags. Although not a requirement, clustering in $\varphi\psi$ space is consistent with our separation of soft and hard modes in the Monte Carlo approach of sampling only the torsional freedom. There are many

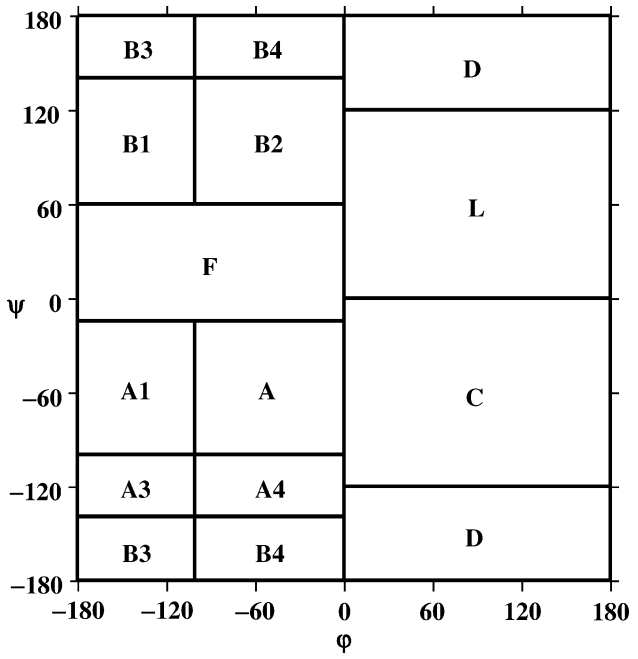

Fig. 1. Alphanumeric tags that specify clustering of microstates in φψ space. Symbols 'A' and 'L' denote right-handed and left-handed α-helix regions, respectively, and right-handed and left-handed twisted β-strand regions are given symbols 'B1' and 'B2', respectively. The remainder of the rotamer regions corresponds to regions that give rise to turn and hairpin conformations.

different strategies to clustering microstates, for example, k-means root-mean-square-deviation (rmsd) of coordinate space, among others. Our computational approach filters the initial culling of structures and projects the energy basins in φψ space by using the approximation that motions of the loop can be separated into motions between rotamer states and motions within a rotamer state, which never deviate much from the latter. Loop conformations were interactively selected until one cluster of conformations occurred in φψ-scatter plots, which were generated for each residue of the polypeptide [11]. For each cluster, an alphanumeric tag that specified conformations of φψ was assigned to each residue of the separated conformations. A basin of conformations with the same sequence of main-chain rotamer tags was generated and defined a given macrostate ξ. Side-chain conformational states were examined and their positional fluctuations, which provide the rugged nature of the basin, were combined with backbone motions in estimating the conformational entropy term via Eq. (13).

Fig. 1 illustrates approximate locations of the main-chain rotamer regions that were used in clustering conformations. The β-strand region and α-helix region of the Ramachandran plot are each divided into four rotamer regions. The right-handed and left-handed α-helix regions are given the symbols 'A' and 'L', respectively, and right-handed and left-handed twisted β-strand regions are given symbols 'B1' and 'B2', respectively. The remainder of the rotamer regions corresponds to regions that give rise to turn and hairpin conformations.

4. Results and discussion

We begin this section with a structural comparison of the native and alternative loop conformation of the acyl pocket as determined from X-ray crystallographic structures. This

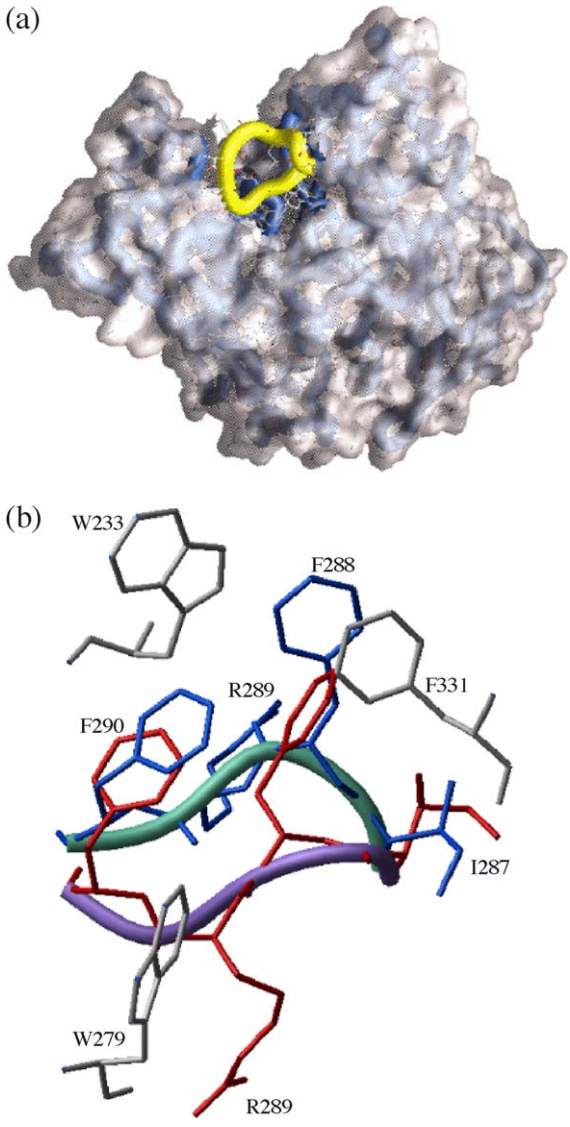

Fig. 2. (a) Molecular illustration of the crystallographic structure of the unbound AChE from *T. californica*, highlighting in yellow color the location of loop residues 279–291. The molecular surface is shown with the loop-region surface element carved out to illustrate the acyl pocket. (b) Superposition of the native four-residue loop segment (cyan backbone tube) and the reorganized loop (purple tube) from reaction of AChE with DFP. Native side chains for residues I287, F288, R289 and F290 are shown in blue and the alternative conformation in red. Several neighboring hydrophobic packing residues taken from the native structure are shown in gray.

comparison provides a basis for assessing the outcome of the structural predictions and determination of macrostates. This is followed by a presentation of the results of the MCSA calculations, and based on these predictions, we sketch out a reaction profile connecting the macrostates associated with the experimental conformations. Further details of the topology of the reaction profile are given from room temperature MC studies, optimization of the predicted states and their residue contributions. In the final section, we discuss residue motions along the profile surveyed by the simulations.

4.1. Structural comparison of loop conformations

Shown in Fig. 2 is a representation of the crystallographic structure of the unbound AChE from *T. californica*, highlighting the location of loop residues 279–291. Also presented is a superposition of the native loop conformation and the reorganized loop from reaction of AChE with DFP. We will focus our attention primarily on residues that show the largest variation between the two structures; namely, residues 287–290 (Ile–Phe–Arg–Phe). For the native conformation, the four-residue segment spans regions of the Ramachandran $\varphi\psi$ space corresponding to rotamer tags [B3,L,B3,B3] (see Fig. 1) and transitions to [A,B4,F,B4]. Backbone geometries of the two conformations differ by the 180° reorientation of the plane of the peptide bond between Ile-287 and Phe-288. The peptide hydrogen of Phe-288 is pointing inside the active-site gorge for the native conformation and outside the gorge upon displacement. Because of this reorientation, we designate the native conformation of residues 287–290 as the '*in*' conformation and the alternative conformation observed with DFP as the '*out*' conformation.

Side-chain conformations of Phe-288 and Arg-289 show significant difference between the two loop structures. For the *in* conformation, Phe-288 lies in a hydrophobic pocket and interacts via π-stacking with side chains of Phe-331 and Trp-233. The *out* conformation moves Phe-288 to outside the hydrophobic pocket. Residue Arg-289 is observed in the *in* conformation to be hydrogen bonded to the backbone carbonyl of Phe-288 and the side chain of His-398, while the *out* conformation shows Arg-289 at the top of the gorge and interacting with backbone carbonyls of Trp-279, Leu-282 and Phe-284.

4.2. MCSA predictions

Table 4 presents the results of the MCSA calculations. Before we discuss the results, a few comments are needed regarding the meaning of the various terms listed in Table 4. We rank each cluster by averaging the single-chain conformations that exhibits the lowest energy (designated as the LEC of that basin). The scoring of these microstates corresponds to Eq. (4) and $\Delta G_{\mathrm{micro}}$ is relative to rank 1 microstate. We further examine each basin by computing the rmsd for residues 287–290 between the LEC and the *in* and *out* loop conformations from the X-ray crystal structures.

From the large collection of ~2600 characterized loop conformations, the eight lowest-energy clustered microstates and their populations are informative about the shape of the energy landscape. The calculations determine *out*-like basins [A,B2,F,B1] and [A,B2,F,B2] are sampled more frequently by a factor >10 than the most energetically favorable *in* conformation [B3,L,B3,B3]. Examination of the crystallographic structure offers insight into why the simulations observed low population of the native basin. The *in* loop conformation, as noted above, is a complex network of cooperative hydrophobic interactions that stabilize native contacts within the acyl-binding pocket. Residues Phe-288 and Phe-290 form tight aromatic packing with Trp-233 and Phe-331, and the simulations starting from random

Table 4
Monte Carlo simulated annealing prediction of microstates

Rank[a]	ξ Basin type[b]	nconf[c]	$\Delta G_{\mathrm{micro}}$[d]	ξ Basin rotamer tags				LEC/X-ray rmsd (Å)			
								in		*out*	
				Ile	Phe	Arg	Phe	bb[e]	sc[f]	bb	sc
1	*out*-like	49	0.0	A	B2	F	B1	2.1	6.0	1.4	3.7
2	*out*-like	5	1.6	A	F	B2	B4	0.9	3.8	1.8	4.7
3	*out*-like	68	2.4	A	B2	F	B2	1.8	5.8	1.0	4.0
4	*extended*	4	3.8	B2	B2	F	B1	2.4	6.6	1.9	5.3
5	*out*-like	3	3.9	A	F	B1	B4	1.0	2.6	1.7	3.9
6	*in*	4	4.0	B3	L	B3	B3	0.3	1.5	2.1	4.7
7	*alternate*	24	12.0	A	B4	L	B4	1.2	1.9	1.9	4.5
8	*in*-like	21	13.6	B4	L	B1	B1	1.4	4.2	1.3	2.1

[a] Rank order of lowest-energy G_{micro} determined from a statistical average over microstates in basin ξ. A total of 2638 microstates were characterized.
[b] Main-chain structure type defined by rotamer tags.
[c] Number of conformations in basin.
[d] Relative free energy (kcal/mol) from rank 1.
[e] Backbone positional root-mean-square deviation (rmsd) for residues 287–290 between lowest-energy conformation (LEC) in basin ξ and X-ray structure of *in* and *out* conformations.
[f] Side-chain positional rmsd.

conformations are faced with the challenge of reinserting these two residues into their hydrophobic network. In the course of folding, the free-energy barrier to reconstituting the native conformation is the slow entropic search of many different simulation trajectories seeking entry to the narrow *in* basin. The computational approach of annealing a small subset of residues (see Table 1) rather than the total protein reflects the combinatorial nature of a large sampling problem and may have missed the structure-determining guidance of cooperativity via long-wavelength transitions embedded in realistic folding micropaths of the native basin, consequently a lower population. Simulations with increased surrounding protein volume show only marginal change in the distribution of loop conformations. A priori inclusion of correct pathways in a reduced and computationally tractable landscape is tantamount to determination of an efficient folding code (see, e.g., discussions by Chen and Dill [27]).

A modeled reaction profile extracted from the MCSA results is illustrated in Fig. 3. Given the lack of a sharp tilt in the energy landscape toward native *in* [B3,L,B3,B3] funnel, the success of sampling basins is simply proportional to their width. Greater population of *out*-like microstates indicates a surface topology of a much wider basin opening and allows many more possible incoming trajectories to reach the funnel. Basins with larger openings have greater conformational entropy and larger distribution of energies, although not necessarily greater basin stabilization as determined from exploring the complete funnel shape. We will return to this point in the next section. For now, four separate basins were found for the ensemble of *out* conformations and their energies of interconversion are greater than thermal fluctuations.

We should note that in our proposed profile, no attempt was made to evaluate free-energy barriers separating macrostates. Although it is possible to identify a reaction path between two structures and use umbrella sampling to calculate the free energy of transition (see, e.g., Rick et al. [38]), calculations for complex and large systems can render a potential of mean force that is strongly dependent on the choice of atoms along the reaction pathway. The end result is an ambiguous determination of the transition state free energy and calculations that suffer from convergence in obtaining configurational averages. Much work is still needed on developing a reliable computational methodology for treating a wide range of protein motions along complex reaction paths.

The predicted *in* conformation (Fig. 4a) reproduces the starting structure very well in terms of rmsd, with a displacement of 0.3 Å in the backbone position and 1.5 Å for side-chain positions. For the initial *out* conformation [A,B4,F,B4] taken from the crystallographic structure, the predictions observed several closely related basins sampled at high populations (conformations [A,B2,F,B1] and [A,B2,F,B2]). The crystallographic conformation contains significant strain energy from complex formation with DFP, and thus lacks a favorable population. Regularization of the predicted *out* conformations leads to a calculated rmsd from the crystal structure of ~ 1 Å for the backbone (Fig. 4b) and ~ 4 Å for side-chain conformations.

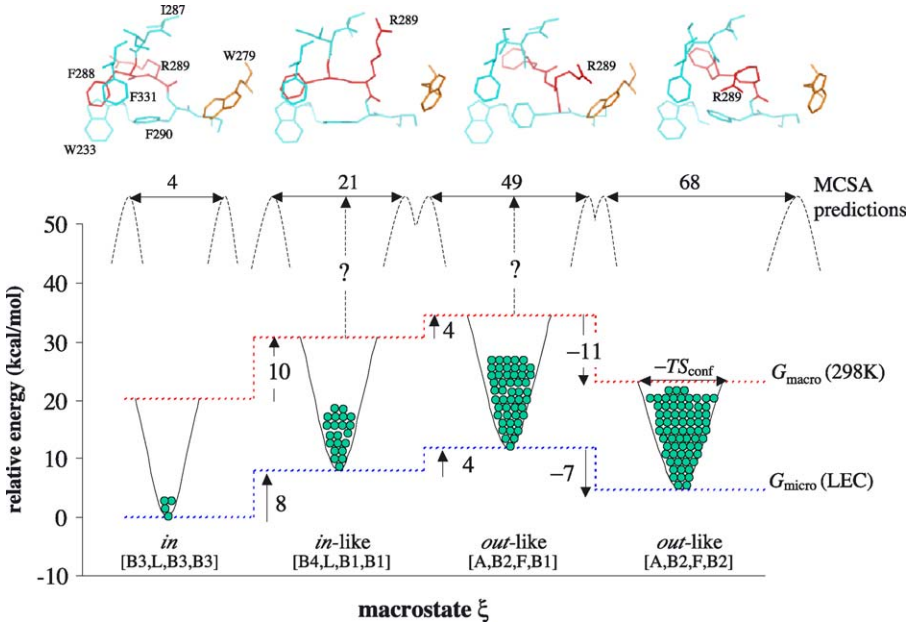

Fig. 3. Modeled reaction profile and schematic of the calculated energy landscape. Top picture depicts motion along the profile for the conformational transition *in* [B3,L,B3,B3] → *out*-like [A,B2,F,B2], highlighting key residues of the loop segment. Numbers in the MCSA predictions and balls in the basins correspond to populations of the microstates (Table 4). Energy differences among basins are taken from Tables 5 and 6; transition barriers are undetermined and designated by question marks. For each basin, the lowest-energy conformation (LEC) from MCSA energy minimization models the funnel minima, and conformational entropy approximates funnel width evaluated at room temperature (298 K).

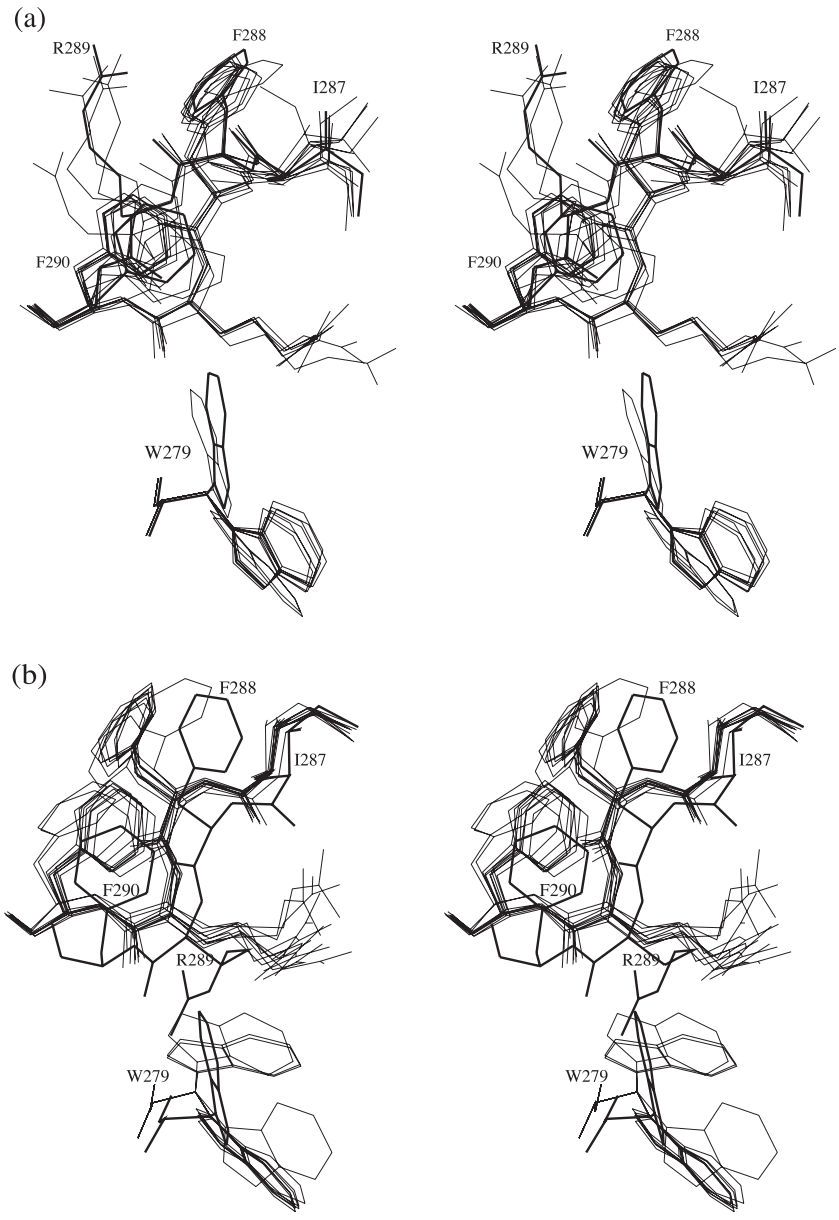

Fig. 4. (a) Superposition of the crystallographic native four-residue loop segment (thick line) with examples of predicted *in* and *in*-like structures taken from annealing simulations (thin lines). (b) Superposition of the crystallographic alternative loop segment (thick line) with predicted *out* and *out*-like structures (thin lines).

Further comparisons between the predicted and crystal structures outside the four-residue loop segment reveal repositioning of several side chains. The most interesting is the movement of Trp-279, where in the crystal structure, one face of the indole ring forms part of the gorge wall and the other is in close contact with a water molecule conserved among crystal structures of AChE from *T. californica* [39]. The simulations show the side chain of Trp-279 displaced toward greater solvent exposure for 3 out of 4 *in* [B3,L,B3,B3] conformations and predominate occurrence for *out* [A,B2,F,B1] and [A,B2,F,B2]. The high population of this alternative conformation may indicate possible deficiencies of implicit solvation schemes in modeling the effect of structural waters on protein conformational flexibility or inadequate conformational sampling. Such solvent effects appear as smoothing of bumps and valleys on energy landscapes from GB solvent interactions and have been observed in calculations of protein folding [21]. More detailed computational studies are needed to address these types of issues.

4.3. Thermal averages of microstates

While the results from the MCSA simulations provide insight into the funnel opening of the *in* and *out* energy basins, room temperature MC simulations allow the deter-

mination of free energies that govern macrostates. Simulation analysis of the lowest-energy microstates predicted from the MCSA study (Table 4) together with five additional conformations are listed in Table 5. These latter conformations are structures in which the indole ring of residue Trp-279 was observed to model the native side-chain conformation (denoted by an asterisk in Table 5). We designate the rank of macrostates by combining the ranking from Table 4 with the room temperature MC simulation results. For example, macrostate 6.4 is defined as the lowest-energy microstate ranked 4 in the MCSA prediction and ranked 6 at the end of thermal averaging by way of room temperature MC. As before, the difference $\Delta G_{\mathrm{macro}}$ is relative to rank 1.

The simulation results for the transition *in* [B3,L, B3,B3] → *out*-like [A,B2,F,B2] is predicted to have a value of $\Delta G_{\mathrm{macro}} \sim 3$ kcal/mol (illustrated in Fig. 3). This free-energy difference and its relative ranking in contrast with Table 4 show the importance of sampling beyond the determination of local minima found by annealing simulations. Initial structures calculated from annealing are low-resolution predictions and lack subtle rearrangement that can significantly lower the free energy. This is nicely illustrated by placement of Trp-279 in the less-solvent accessible native conformer (rank order 1) instead of the more populated conformation observed during annealing (rank order 12). Refinement allows the side chain of Trp-279 to collapse and stabilize the hydrophobic network of residues Phe-288 and Phe-290, thus moving deeper in the energy funnel by optimizing E_{local} rather than G_{solv}. Interdependence of Trp-279 with Phe-288 and Phe-290 observed in the simulations is consistent with the proposal of Millard et al. [1] that the loop movement is a structural device that conjoins the peripheral substrate binding site (Trp-279) with the active site (Phe-288 and Phe-290), and the calculations offer a free-energy basis for allosterism of these two functional sites in AChE.

An assessment of the accuracy of $\Delta G_{\mathrm{macro}}$ can be made from experimental data of measuring reversible association constants for complex formation between AChE and various ligands [1,3–7]. Comparison of binding sarin and DFP to *E. electricus* AChE shows a free energy penalty of ~ 4 kcal/mol for binding the larger ligand [5]. The proposal that the penalty is due primarily to loop reorganization is corroborated by site-directed mutagenesis of *Human* AChE at residues Phe-295 and Phe-297 in the loop segment and the effect on binding DFP [6,7]. These two phenylalanines are conserved among species of AChE and are equivalent to Phe-288 and Phe-290 in *Torpedo*. Substituting both residues with smaller aliphatic side chains alters the free energy of association for DFP by ~ 4 kcal/mol [6,7], making complex formation significantly more *favorable*. Modeling calculations suggest that the favorable increase in binding is the result of elimination of steric clashes between the bulky *iso*propropyl group of DFP and the phenylalanine rings, thus removing the energetic cost of displacing the loop.

Errors in the calculations reported in Table 5 are derived from the sampling distribution. The standard error of the mean free energy is the standard deviation of the free energy divided by the square root of the sample size; namely, $Z(\sigma G_{\mathrm{macro}})/\sqrt{N} = 1.6452(20.0 \text{ kcal/mol})/\sqrt{4000} = 0.5$ kcal/mol, where $Z = 1.6452$ is the number of standard deviations from the mean that is needed to claim with 95% confidence the value of the true mean. As a result of propagation of error analysis, the standard error of the difference between the *in* and *out* energy basins is $\sqrt{(0.5^2+0.5^2)} = 0.7$ kcal/mol. In an alternative method of computing the standard error of the mean, the sampled populations are divided into two groups of 2000. The

Table 5
Monte Carlo predictions of macrostates at 298 K

Rank[a]	ξ Basin type	$\Delta G_{\mathrm{macro}}$[b]	$\langle E_{\mathrm{local}} \rangle_\xi$	$\langle G_{\mathrm{solv}} \rangle_\xi$	TS_{conf}	ξ Basin rotamer tags			
						Ile	Phe	Arg	Phe
1.6*	*in*	0.0	−1612.4	−3916.2	131.7	B3	L	B3	B3
2.3*	*out*-like	2.8	−1660.5	−3862.3	134.8	A	B2	F	B2
3.4*	*extended*	10.8	−1713.8	−3806.8	128.4	B2	B2	F	B1
4.8*	*in*-like	10.4	−1702.5	−3814.2	133.4	B4	L	B1	B1
5.1*	*out*-like	14.1	−1676.1	−3835.7	134.5	A	B2	F	B1
6.4	*extended*	21.7	−1688.7	−3818.0	132.0	B2	B2	F	B1
7.1	*out*-like	23.7	−1708.6	−3796.0	132.2	A	B2	F	B1
8.3	*out*-like	26.4	−1689.4	−3813.7	130.9	A	B2	F	B2
9.5	*out*-like	28.4	−1561.6	−3941.9	128.6	A	F	B1	B4
10.2	*out*-like	29.8	−1533.6	−3965.5	131.5	A	F	B2	B4
11.7	*alternate*	31.9	−1519.0	−3982.8	126.8	A	B4	L	B4
12.6	*in*	33.8	−1580.6	−3912.9	133.2	B3	L	B3	B3
13.8	*in*-like	42.6	−1705.3	−3785.1	127.7	B4	L	B1	B1

[a] Rank order of macrostate ξ. Rank *x.y* designates order *x* determined from thermal averaging and *y* refers to ranking from Table 4 (see text). Asterisk denotes native side-chain conformation for residue Trp-279. MC simulations evaluated a total of 630,000 trial conformations for each macrostate and 4000 conformations were extracted and analyzed.

[b] Relative free energy (kcal/mol) from rank 1. Terms $\langle E_{\mathrm{local}} \rangle_\xi$ and $\langle G_{\mathrm{solv}} \rangle_\xi$ are statistical averages over microstates in basin ξ. Scatter in values of $G_{\mathrm{macro}}(\xi) = \langle G_{\mathrm{micro}} \rangle_\xi - TS_{\mathrm{conf}}(\xi)$ is on the order of ± 20.0 kcal/mol with a precision of ± 1.3 kcal/mol in $\Delta G_{\mathrm{macro}}$.

standard error of the mean free energy for the *in* state is 0.5 and the standard error of the mean free energy *out* state is 1.2. The standard error of the difference between the *in* and *out* states is 1.3 kcal/mol. Because the latter is larger, we report this value in Table 5 as an upper bound to the estimated error, yielding $\Delta G_{\text{macro}} \sim (3 \pm 1)$ kcal/mol.

A further observation from the results is that solvent shifts the distribution of conformational microstates between basins [B3,L,B3,B3] and [A,B2,F,B2] in favor of the *in* macrostate (by $\Delta G_{\text{solv}} \sim -54$ kcal/mol), whereas internal energy stabilizes the *out* macrostate ($\Delta E_{\text{local}} \sim 48$ kcal/mol difference). The computed internal energetic difference includes the L to B2 transition from a less-populated and marginally favorable left-handed helix in residue Phe-288 to an extended conformation. As illustrated below, the dominant effect on the conformational redistribution is, however, the compensation between solvent interactions from the GB model and Coulomb charging of the protein. This electrostatic compensation is key to accurate discrimination of native-like protein structures among a library of decoys (see, e.g., Vorabjev et al. [28]).

An important component in the determination of ΔG_{macro} is conformational entropy, where the quasi-harmonic approximation yields $T\Delta S_{\text{conf}} \sim 3$ kcal/mol. Combined with the annealing results of a broader basin opening, room temperature calculations establish that the *out* macrostate is more accessible to positional fluctuations of larger amplitude, thereby indicating a wider energy basin. Fig. 5a illustrates the distribution of microstates calculated for the two lowest-energy basins. Thermodynamic averages were taken from the histograms and Fig. 5b reports the convergence of the free energies as a function of number of trial microstates evaluated. A broader distribution is observed for the *out*-like macrostate and greater fluctuation in its convergence, both of which are consistent with the *in* basin being more stable.

For the *in* and *out* macrostates, Fig. 5c and d illustrates the probability distribution function along the normal mode associated with the largest eigenvalue. According to Kitao et al. [40], anharmonicity is most likely to be found in normal modes associated with the largest eigenvalues. For the two basins, the probability distribution functions along the normal modes are Gaussian like, and as a result provide sufficient validation of the clustering in order to obtain an estimate of the conformational entropy.

The excellent agreement between the simulation ΔG_{macro} and experimental results indicates that the model correctly ranks the predicted basins. We should note that scoring of conformations is dependent on the implicit solvent model used for sampling the potential energy surface. To illustrate this, we have rescored the *in* and *out* conformations by the Poisson equation using the APBS technique [41]. Fig. 6 reports a scatter plot of the two solvent models and clearly indicate differing results. The free energy ΔG_{macro} computed

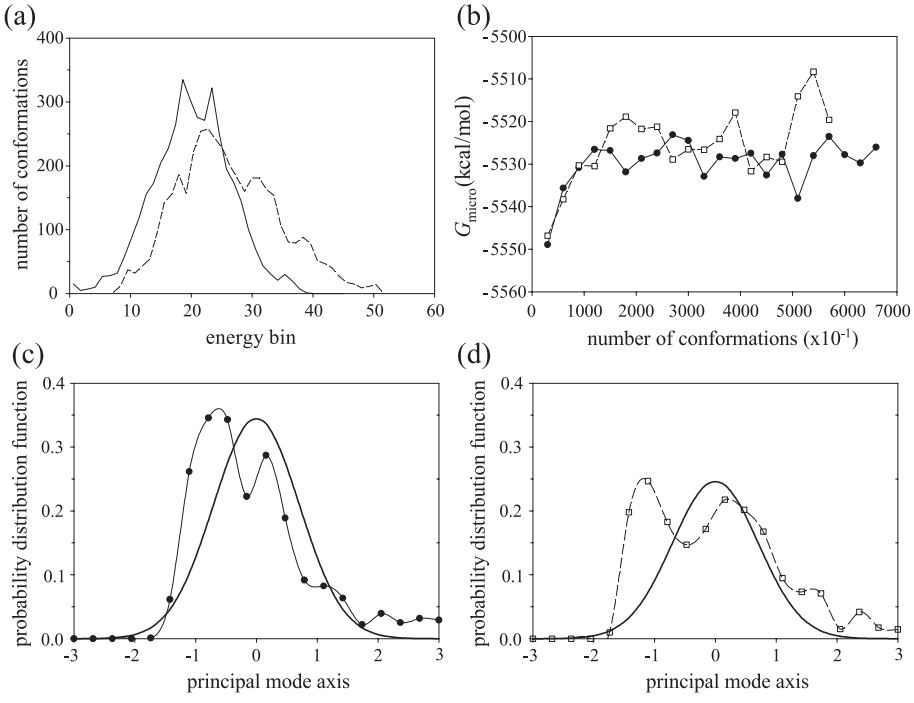

Fig. 5. (a) Distribution of microstates calculated from the two lowest-energy basins, *in* [B3,L,B3,B3] (solid line) and *out*-like [A,B2,F,B2] (dashed line). (b) Convergence of the free energies as a function of number of trail microstates evaluated; solid line fitted to solid circles, *in* [B3,L,B3,B3]; and dashed line fitted to open boxes, *out*-like [A,B2,F,B2]. (c) and (d) Probability distribution functions for the *in* (thin solid line) and *out*-like (dashed line) macrostates along the normal mode associated with the largest eigenvalues. Abscissa is scaled by the rmsd fluctuation, which for the *in* basin the fluctuation is 1.2 Å and for the *out* basin, the fluctuation is 1.6 Å. For comparison, Gaussian distributions are plotted (thick lines).

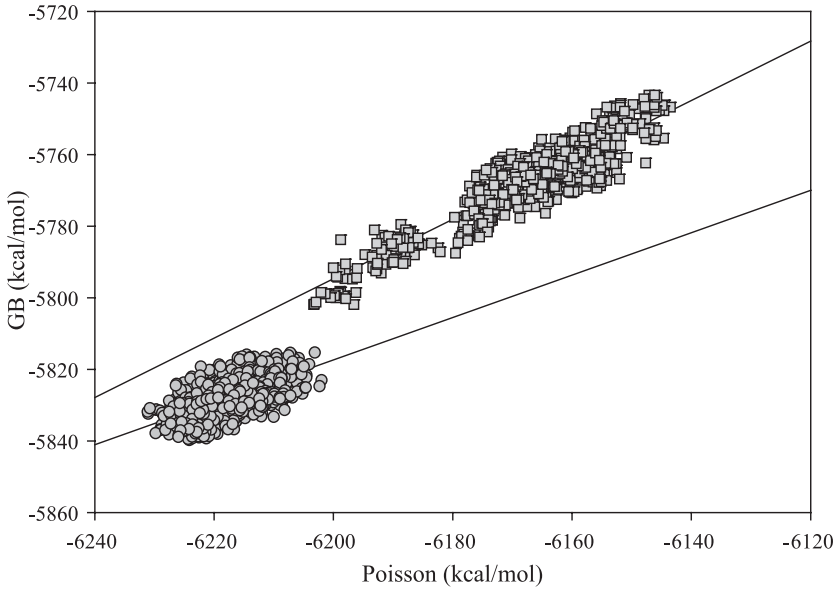

Fig. 6. Scatter plot of electrostatic solvation free energies computed with the Poisson equation [41] versus the GB model. The line slopes through the *in* (circles) and *out* (squares) basins are 0.7 and 1.0, respectively. The correlations and standard deviations are 0.7 and 4.1 kcal/mol, for the *in* basin, and 0.9 and 4.4 kcal/mol, for the *out* basin.

with the Poisson solvent model has the *in* macrostate less favorable than the *out* basin by 7.4 kcal/mol, which is inconsistent with data from experiment and the MC simulation. The inconsistency is due to reevaluation of conformations with a scoring function different than the one used in the generation of the ensemble. While the Poisson equation is more rigorous, the GB/SA method searches the potential energy surface that is described by an implicit reaction field of a five-parameter solvent model, rather than a search satisfied by the Poisson equation. Conformations found from the search are those determined by the offset of solvent polarization versus Coulomb charging for that particular model. The replacement of the solvent electrostatic term with the Poisson equation perturbs the energy landscape without allowing for conformational relaxation to re-optimize solvent polarization for the electric field. Further discussion of different solvent models on rescoring conformational ensembles and their inherent problems is given elsewhere (Olson, manuscript submitted for publication).

4.4. Optimization of microstates and residue contributions

Table 6 reports MCSA energy minimization of the highest-ranking microstates from thermal averaging. For the purpose of modeling the landscape topology, we approximate these microstates as funnel minima and thus ignore conformational entropy. The calculations show further energy separation between the *in* and *out*-like basins, with a free-energy difference of ~ 5 kcal/mol. The free-energy determinant of stabilization observed for the *in* microstate is similar to that calculated in Table 5, yet more pronounced. Solvent effects modeled from the GB term favor the *in* conformation by roughly 88 kcal/mol, while protein internal interactions advance redistribution to the *out*-like state by 84 kcal/mol. Negligible effect on scoring the two basins is energy from the solvent surface-area term of Eq. (7).

The next ranking microstate is the marginally populated [B2,B2,F,B1] conformation and exhibits a free energy ~ 6 kcal/mol greater than native basin. This particular microstate

Table 6
Optimization of microstates

Rank[a]	ξ Basin type	ΔG_{micro}[b]	E_{local}	G_{GB}	G_{SA}	*in* bb[c]	*out* bb[d]	ξ Basin rotamer tags			
								Ile	Phe	Arg	Phe
1.1.5*	*in*	0.0	− 8881.9	− 5848.6	110.0	0.6	3.5	B3	L	B3	B3
2.2.2*	*out*-like	4.7	− 8966.0	− 5760.6	110.8	2.5	1.7	A	B2	F	B2
3.3.4*	*ext*	5.7	− 9001.8	− 5723.2	110.1	3.2	2.9	B2	B2	F	B1
4.5.6*	*in*-like	8.1	− 9030.2	− 5692.8	110.5	2.0	2.1	B4	L	B1	B1
5.4.1*	*out*-like	11.9	− 9007.6	− 5711.2	110.2	2.7	2.6	A	B2	F	B1

[a] Rank order of lowest-energy conformer (LEC) in basin ξ. A conformer ranked *x.y.z* refers to microstate ranked *z* in the MCSA prediction calculation (Table 4), ranked *y* in thermal averaging (Table 5) and ranked *x* in optimization. Asterisk denotes native conformation for residue Trp-279.
[b] Relative free energy (kcal/mol) from microstate of rank 1.
[c] Backbone positional rmsd (in Å) between the LEC and X-ray *in* conformation.
[d] Backbone positional rmsd (in Å) between the LEC and X-ray *out* conformation.

is an extended conformation and loop movement to this basin will be considered an off-pathway transition (see discussion below). Alternative *in*-like [B4,L,B1,B1] and *out*-like [A,B2,F,B1] follow on the free-energy ranking.

Summarized in Table 7 are electrostatic energies for the optimized microstates decomposed in term of residue contributions. The free-energy components include W_{ele}, G_{scr} and G_{pol}, and their sum is given by G_{res}. The ranking of microstates follows those listed in Table 6.

The calculations reveal that residue Arg-289, which is conserved among species of AChE, plays a principal role in shaping the landscape of loop microstates. The *in* [B3, L,B3,B3] conformation positions Arg-289 at the side of the reactive-site gorge and charging the residue embedded in a charged medium of the protein is unfavorable ($W_{ele}=5$ kcal/mol), while solvent dielectric screening of charge–charge interactions is favorable ($G_{scr}=-56$ kcal/mol). The stability where $W_{ele}>0$ and $G_{scr}<0$ found for Arg-289 in the *in* microstate is distinctly reversed by a sizeable magnitude for the other microstates. Considerable dipolar reorganization takes place of the protein from reorientation of the positive charge of Arg-289, and for basin formation and stabilization, optimal placement of the positive charge is a determining factor. This effect is observed by the transition *in* [B3,L,B3,B3] → *out*-like [A,B2,F,B2], where reorganization of Arg-289 to a position top of the gorge yields a free-energy difference $\Delta G_{res} \sim 0$. Conformational search of the landscape for optimal G_{res} is by a reorder of W_{ele} versus G_{GB} between the two basins and is obtained with minimum net free-energy penalty for charge reorganization. An unfavorable effect of charge displacement is disruption of cation π-stacking interactions, which significantly increase the value of W_{ele}, resulting in destabilization of Phe-288 as new stabilizing cation interactions are formed.

A further interesting finding is that of the four residues in the loop, only Ile-287 shows a significant net electrostatic shift in favor of the *out*-like [A,B2,F,B2] state. This shift is achieved by solvent reorganization of the transitioned peptide backbone. Combined, the four loop residues contribute an electrostatic free energy that favors the *out*-like conformation by ~ -3 kcal/mol and similarly non-polar interactions are stabilized by ~ -8 kcal/mol (data not shown). Only the total term G_{scr} at the residue level shows stabilization of the *in* macrostate and the remaining protein structure outside the loop segment augments this term disfavoring the transition (see Table 6).

4.5. Residue motions along the reaction profile

Tables 4 and 5 indicate none of the alternative macrostates are dynamically connected to the two crystallographic conformations. This is consistent with low *B*-factors for loop residues determined from crystal structures of unbound AChE, as well as supported by the lack of observed large-scale loop movement in a 10-ns molecular-dynamics simulation of AChE [42]. As illustrated in Fig. 3, the transition from the *in* [B3,L,B3,B3] to the final *out*-like [A,B2,F,B2] conformation is modeled on a pathway initiated by the *in*-like [B4,L,B1,B1] macrostate. This basin is characterized structurally by the abrupt movement of Arg-289 calculated at $\sim$ 10-Å rmsd from native, and given the overall importance of this residue in shaping the loop landscape (Table 7), we designate this basin as the 'rate-limiting' step along the reaction profile. A net free-energy penalty of $\sim$ 10 kcal/mol is incurred at 298 K for reorganization and includes an increase of 2 kcal/mol in conformational entropy.

The next state on the modeled pathway is the peptide bond flip described by the *out*-like [A,B2,F,B1] macrostate. The energy difference from [B4,L,B1,B1] to this interme-

Table 7
Electrostatic free energies (kcal/mol) for residues 287–290

ξ Basin[a]	G_{res}[b]				W_{ele}			
	Ile	Phe	Arg	Phe	Ile	Phe	Arg	Phe
in	−35.6	−42.0	−118.0	−43.6	−41.2	−57.1	4.6	−52.7
out-like	−46.5	−37.9	−118.2	−39.3	−42.1	−29.0	−127.1	−45.6
ext	−37.5	−37.1	−112.2	−41.9	−38.7	−32.0	−140.8	−48.5
in-like	−42.1	−36.4	−115.5	−40.1	−43.5	−39.4	−164.2	−46.7
out-like	−40.5	−38.6	−108.5	−38.5	−34.8	−28.6	−157.1	−43.0
ξ Basin[a]	G_{pol}				G_{scr}			
	Ile	Phe	Arg	Phe	Ile	Phe	Arg	Phe
in	−1.0	−1.9	−67.1	−1.7	6.6	17.0	−55.5	10.8
out-like	−2.6	−1.1	−73.0	−1.1	−1.8	−7.8	81.9	7.4
ext	−1.9	−1.6	−65.1	−0.8	3.1	−3.5	93.7	7.4
in-like	−1.8	−3.1	−66.6	−1.1	3.2	6.1	115.3	7.7
out-like	−2.4	−1.9	−63.8	−1.0	−3.3	−8.1	112.4	5.5

[a] Rank order given by Table 6.
[b] Total electrostatic free energy calculated on a per-residue level for the optimized microstate in basin ξ and is defined as: $G_{res}=W_{ele}+G_{pol}+G_{scr}=W_{ele}+G_{GB}$.

diate is 4 kcal/mol for both the funnel minima and thermal averaging at 298 K, where the latter contains roughly 1 kcal/mol increase in entropy to the free energy. Additional reorganization cost is protein internal energy (Tables 5–7), where Arg-289 is further destabilized by an electrostatic penalty of $\sim$ 7 kcal/mol.

The final *out*-like [A,B2,F,B2] macrostate is reached by optimization of side-chain and backbone interactions of the four-residue segment, leading to a favorable free-energy change of 7 kcal/mol for the minima and 11 kcal/mol for room temperature. Our modeled reaction profile of *in* → *out* is advanced by entropy similar to protein unfolding, yet dissimilar in that electrostatic solvent effects promotes the native macrostate. As a result of high energies, the two intermediate states [B4,L,B1,B1] and [A,B2,F,B1] may be eliminated by concerted movement during loop reorganization, although it is very interesting that the simulations observed independent basins for the two key reaction steps.

The paucity of low-energy backbone basins has implications for conformational selection of new or modified ligands. Sarin and soman both stabilize the native *in* [B3,L,B3,B3] macrostate, while DFP the *out* macrostate or from simulations the optimized *out*-like [A,B2,F,B2]. The question is what remaining basins can be captured from redistributing conformational isomers? Our survey of the energy landscape and its projection in main-chain $\varphi\psi$ space suggests only the off-pathway [B2,B2,F,B1] or high-energy macrostates. Other than these states, what remains is side-chain conformational freedom of the basins. For example, the *out* basin is a coarse-grained collection of different side-chain rotamers; in particular, residues Trp-279 and Arg-289 show movement among microstates (see Fig. 4). Ligands can be designed that either stabilize the residues in specific conformations or alternatively promote conformational disorder among dynamically connected rotamers. In either case, the predictions suggest that the protein backbone conformation remains well defined. This conclusion can be easily tested by further crystallographic work on AChE.

Our prediction of macrostates and their lack of dynamic connectivity are also significant to the problem of molecular docking of DFP or similarly sized ligands to AChE. A computational approach of limited protein flexibility starting from native *in* [B3,L,B3,B3] state will fail to predict the correct conformational binding mode of DFP. The correct approach is the generation of an ensemble of microstates by high-temperature annealing and to cluster them into basins as starting structures for docking. These basins and their reorganization free energies are those predicted here. An active area of research is the application of implicit solvent models to determine free energies of complex formation. The crux of the problem is conformational sampling of dipolar relaxation of bound and unbound states, and the issue of scaling ε_p in implicit schemes to obtain accurate results [12–14,43]. For unbound protein states, we have demonstrated that relative changes in free energies can be accurately calculated without the need for scaling the dielectric constant. This success is due to sufficient configurational averaging of the basins and their smooth-funnel topology.

5. Conclusions

We investigated the energy landscape of the acyl pocket loop of AChE by use of a Monte Carlo sampling approach with a GB/SA solvent model. The computational strategy consisted of determining landscape topology of basin openings, thermal averages of microstates and funnel minima. A loop reorganization profile was constructed and motions of residues and their free energies were described along the profile. Residue Arg-289 was identified to play a critical role in basin stabilization and should be the focus of mutagenesis studies. The free-energy difference between native and the loop conformation selected by the inhibitor DFP was accurately predicted.

Acknowledgements

The authors thank Dr. M.S. Lee for the many valuable discussions. We gratefully acknowledge the Army High Performance Computing Research Center (AHPCRC)/University of Minnesota for a generous grant of computer time. The work of L.C. was supported by the AHPCRC under the auspices of the Department of Army, Army Research Laboratory cooperative agreement number DAAD19-03-D-0001. Research also sponsored by the U.S. Army Medical Research and Materiel Command Project number RIID 02-4-1R-069.

References

[1] C.B. Millard, G. Kryger, A. Ordentlich, H.M. Greenblatt, M. Harel, M.L. Raves, Y. Segall, D. Barak, A. Shafferman, I. Silman, J.L. Sussman, Biochemistry 38 (1999) 7032.
[2] C.B. Millard, G. Koellner, A. Ordentlich, A. Shafferman, L. Silman, J.L. Sussman, J. Am. Chem. Soc. 121 (1999) 9883.
[3] Y.C. Chiu, A.R. Main, W.C. Dauterman, Biochem. Pharmacol. 18 (1969) 2171.
[4] G.J. Hart, R.D. O'Brien, Biochemistry 12 (1973) 2940.
[5] A. Forsberg, G. Puu, Eur. J. Biochem. 140 (1984) 153.
[6] A. Ordentlich, D. Barak, C. Kronman, Y. Flashner, M. Leitner, Y. Segall, N. Ariel, S. Cohen, B. Velan, A. Shafferman, J. Biol. Chem. 268 (1996) 17083.
[7] A. Ordentlich, C. Kronman, D. Barak, D. Stein, N. Ariel, D. Marcus, B. Velan, A. Shafferman, FEBS Lett. 334 (1993) 215.
[8] K.A. Dill, Protein Sci. 8 (1999) 1166.
[9] S. Kumar, M. Buyong, C.-J. Tsai, N. Sinha, R. Nussinov, Protein Sci. 9 (2000) 10.
[10] L. Carlacci, A.S. Edison, Proteins 40 (2000) 367.
[11] L. Carlacci, Biopolymers 58 (2001) 359.
[12] I. Muegge, T. Schweins, A. Warshel, Proteins 30 (1998) 407.

[13] Y.Y. Sham, Z.T. Chu, H. Tao, A. Warshel, Proteins 39 (2000) 393.
[14] M.A. Olson, Biophys. J. 81 (2001) 1841.
[15] C.S. Rapp, R.A. Friesner, Proteins 35 (1999) 173.
[16] L. David, R. Luo, M. Gilson, J. Comp. Chem. 21 (2000) 295.
[17] N. Calimet, M. Schaefer, T. Simonson, Proteins 45 (2001) 144.
[18] Y. Liu, D.L. Beveridge, Proteins 46 (2002) 128.
[19] D. Williams, K. Hall, Biophys. J. 76 (1999) 3192.
[20] V. Tsui, D.A. Case, J. Am. Chem. Soc. 122 (2000) 2489.
[21] B.D. Bursulaya, C.L. Brooks III, J. Phys. Chem., B 104 (2001) 12378.
[22] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, J. Am. Chem. Soc. 112 (1990) 6127.
[23] M.S. Lee, F.R. Salsbury Jr., C.L. Brooks III, J. Chem. Phys. 116 (2002) 10606.
[24] M. Schaefer, M. Karplus, J. Phys. Chem. 100 (1996) 1578.
[25] D. Qiu, P.S. Shenkin, F.P. Hollinger, W.C. Still, J. Phys. Chem., A 101 (1997) 3005.
[26] M.A. Olson, in: J. Leszczynski (Ed.), Computational Chemistry: Reviews of Current Trends vol. 4, World Scientific Publishing, Singapore, 1999, p. 153.
[27] S.-J. Chen, K.A. Dill, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 646.
[28] Y.N. Vorobjev, J.C. Almagro, J. Hermans, Proteins 32 (1998) 399.
[29] A. Amadei, A.B.M. Linsen, H.J.C. Berendsen, Proteins 17 (1993) 412.
[30] B.R. Brooks, D. Janezic, M. Karplus, J. Comp. Chem. 16 (1995) 1522.
[31] D. Janezic, B.R. Brooks, J. Comp. Chem. 16 (1995) 1543.
[32] D. Janezic, R.M. Venable, B.R. Brooks, J. Comp. Chem. 16 (1995) 1554.
[33] M.L. Raves, M. Harel, Y.P. Pang, I. Silman, A.P. Kozikowski, J.L. Sussman, Nat. Struct. Biol. 4 (1997) 57.
[34] G. Némethy, M.S. Pottle, H.A. Scheraga, J. Phys. Chem. 87 (1983) 1883.
[35] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.
[36] W.L. Jorgensen, J. Tirado-Rives, J. Am. Chem. Soc. 110 (1988) 1657.
[37] G. Perrot, B. Cheng, K.D. Gibson, J. Vila, K.A. Palmer, A. Nayeem, B. Maigret, H.A. Scheraga, J. Comp. Chem. 13 (1992) 1.
[38] S.W. Rick, J.W. Erickson, S.K. Burt, Proteins 32 (1998) 7.
[39] G. Koellner, G. Kryger, C.B. Millard, I. Silman, J.L. Sussman, T. Steiner, Mol. Biol. 296 (2000) 713.
[40] A. Kitao, S. Hayward, N. Go, Proteins 33 (1998) 496–517.
[41] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037.
[42] K. Tai, T. Shen, U. Börjesson, M. Philippopoulous, J.A. McCammon, Biophys. J. 81 (2001) 715.
[43] M.A. Olson, L.T. Reinke, Proteins 38 (2000) 115.