Analysis of a 10-ns Molecular Dynamics Simulation of Mouse Acetylcholinesterase

Kaihsu Tai,* Tongye Shen,[†] Ulf Börjesson,* Marios Philippopoulos,* and J. Andrew McCammon[‡]

*Department of Chemistry and Biochemistry, [†]Department of Physics, and [‡]Howard Hughes Medical Institute and Departments of Chemistry and Biochemistry and of Pharmacology, University of California, San Diego, La Jolla, California 92093 USA

ABSTRACT   A 10-ns molecular dynamics simulation of mouse acetylcholinesterase was analyzed, with special attention paid to the fluctuation in the width of the gorge and opening events of the back door. The trajectory was first verified to ensure its stability. We defined the gorge proper radius as the measure for the extent of gorge opening. We developed an expression of an inter-atom distance representative of the gorge proper radius in terms of projections on the principal components. This revealed the fact that collective motions of many scales contribute to the opening behavior of the gorge. Covariance and correlation results identified the motions of the protein backbone as the gorge opens. In the back-door region, side-chain dihedral angles that define the opening were identified.

INTRODUCTION

Acetylcholinesterase (AChE) is the enzyme responsible for the termination of signaling in cholinergic synapses (such as the neuromuscular junction) by degrading the neurotransmitter acetylcholine. AChE has a gorge, 2 nm deep, leading to the catalytic site. Molecular dynamics (MD) simulations (McCammon and Harvey, 1987; Brooks et al., 1988) have shown breathing motions of this gorge (Gilson et al., 1994; Wlodek et al., 1997; Zhou et al., 1998). They also showed that an alternative portal providing access to the catalytic site is present in AChE. This back door, as it is called, may facilitate rapid solvent and product removal (Gilson et al., 1994; Tara et al., 1999).

In this paper, we collect a 10-ns trajectory of mouse AChE (mAChE) using MD simulation. We define and calculate the time series for the proper radius of the gorge and for the back-door opening events, as observed in our 10-ns MD trajectory. We also perform principal component analysis for the trajectory. Then we look for correlations among the results from our analyses with a view to finding the important collective motions in AChE responsible for its opening behaviors.

MATERIALS AND METHODS

Molecular dynamics simulation

The previous 1-ns MD simulation of mAChE (Tara et al., 1999) was extended to afford a trajectory of 10.8 ns. The set-up procedure has been described in the previous paper and is summarized below. The crystal structure of mAChE in complex with fasciculin 2 (Fas2) (Bourne et al., 1995) (Protein Data Bank identification code: 1MAH) was used to model unliganded mAChE. Fas2 was removed from the structure, and seven residues in mAChE that are missing in 1MAH (residues 258–264) were repaired by modeling with Insight II (Molecular Simulations, San Diego, CA) to give the initial structure. The protein was solvated in a cubic box (9.6-nm edges) of pre-equilibrated water molecules. To neutralize the -10 charge of mAChE, nine sodium ions were placed in the solvent at ~0.5 nm from the protein surface, and one sodium ion was placed in the active site near Trp86 and His447. The simulation system had a total of 8,289 solute atoms and 75,615 solvent atoms.

The simulation was performed in the isothermal-isobaric ensemble. The solvent and solute were separately coupled to temperature reservoirs of 298.15 K with coupling times of 0.1 ps. Pressure was restrained to 1 atm with a coupling time of 0.4 ps. All minimization and MD simulation steps were performed using NWChem version 3.0 (Straatsma et al., 2000) with the AMBER 94 force field (Cornell et al., 1995). Long-range electrostatic interactions were calculated using particle-mesh Ewald summation (Darden et al., 1993). Bond lengths between hydrogens and heavy atoms were constrained using SHAKE (Ryckaert et al., 1977). The simulation was performed on 32 processors of a Cray T3E parallel supercomputer at the San Diego Supercomputer Center over a period of 3 years, consuming a total of ~200 processor-months of supercomputer time. Frames were collected at 1-ps intervals for the simulation length of 10.8 ns, giving 1.08×10^4 frames. The first 700 frames of this trajectory were considered the equilibration phase and not used in the main analysis for reasons described in Results; only the subsequent 10,000 frames (10 ns) were used in the main analysis. Thus, the last 300 ps of the previous trajectory (Tara et al., 1999) are the first 300 ps of the present 10-ns trajectory admitted for analysis.

Proper radius of the gorge

To characterize the degree of the gorge opening with a single variable, we defined the gorge proper radius ρ for the conformation of each snapshot as the maximum radius of a spherical ligand that can go through the gorge from outside the protein to reach the bottom. Equivalently, it is the maximum probe radius that still generates a solvent-accessible surface with a continuous topology.

In our algorithm, we first generate the Shrake and Rupley surface with a testing probe sphere of given radius; then we try to determine whether the surface generated by the atom O_γ in residue Ser203 (one of the residues in the active site; the bottom of the gorge) is topologically continuous with the surface of the bottleneck region (represented by the atoms Leu76 $C_{\delta 1}$, Trp286 C_β, and Tyr72 O_H) for that probe radius. Using a binary search algorithm to decide what will be the next probe radius, we can determine ρ with desired precision. We started with a test value of 160 pm and assumed ρ to be bounded by 80 pm and 240 pm. With six iterations in the binary search algorithm, we achieved a final discretization of 5 pm. We calculated ρ for each snapshot, and the results were collected as a time series $\rho(t)$.

Back-door opening events

In addition to the gorge, MD results from this and previous (Tara et al., 1999) simulations of mAChE also showed an alternative opening occasionally large enough for at least water molecules to pass. This opening, named the back door (Gilson et al., 1994), is conjectured to assist in releasing solvent or reaction products. The back door is formed by the residues Trp86, Gly448, Tyr449, and Ile451. Instead of using multiple probe radii to calculate the proper radius, as in the case of the gorge, we used a single probe radius of 140 pm to test whether it is open or closed. We always blocked the gorge entrance while probing for back-door opening events. (Similarly, we blocked the back-door region in the gorge calculation described above.) Each frame that the algorithm reported to have a back door opening was verified visually using Insight II. The result is expressed as a time series:

$$\Upsilon(t) := \begin{cases} 0 & \text{if closed} \\ 1 & \text{if open} \end{cases} \quad (1)$$

Principal components

Principal component analysis

Considering only the α-carbons, the N-residue trajectory can be considered as a vector function of time t, namely, $r(t) = [r_{1x}(t), r_{1y}(t), \ldots, r_{Nz}(t)]^T$ of size $f = 3N$, containing the Cartesian coordinates at time t for residue 1 in the x direction, residue 1 in the y direction, up to residue N in the z direction.

The ij entry C_{ij} of the covariance matrix $\mathbf{C}$ is the covariance of the positions for two degrees of freedom i and j, namely, $C_{ij} = \langle (r_i(t) - \langle r_i \rangle_t)(r_j(t) - \langle r_j \rangle_t) \rangle_t$ where $\langle \cdot \rangle_t$ is the time average over the whole trajectory. Principal component analysis (PCA) (García, 1992; Amadei et al., 1993; Balsera et al., 1996; Wlodek et al., 1997) diagonalizes $\mathbf{C}$ by solving $\Lambda = \mathbf{T}^T\mathbf{C}\mathbf{T}$, so as to obtain the diagonal matrix Λ with the diagonal entries being the eigenvalues ranked by magnitude. The cth column of the transformation matrix $\mathbf{T}$ is the cth eigenvector $\mathbf{v}_c$, that is, $\mathbf{T} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_f]$.

In our analysis, the algorithm for PCA of the Cartesian coordinates of the α-carbon atoms was implemented in the Java programming language, with the JAMA matrix package (The MathWorks, Natick, MA, and National Institute of Standards and Technology, Gaithersburg, MD).

The first five residues in the MD simulation (residues 4–8) and the last five residues (539–543) were removed before the PCA to avoid terminal motions excessively dominating the analysis. Therefore, each frame contained the Cartesian coordinates for $N = 530$ α-carbon atoms (Leu9 to Lys538), or $f = 3N = 1590$ degrees of freedom. Each frame in the 10-ns trajectory is superimposed to the first frame using quaternion fitting before being admitted into the PCA. This removes the motions in the rotational and translational degrees of freedom.

Using the transformation matrix $\mathbf{T}$, we can convert between the projections along the principal components and the Cartesian coordinates using $(\Delta r(t) := r(t) - \langle r \rangle_t = \mathbf{T}\mathbf{p}(t)$, where $\mathbf{p}(t) = [p_1(t), p_2(t), \ldots, p_f(t)]^T$ is a

vector of size f; the cth entry thereof is the projection along the principal component c.

Back-projecting: expressing a C_α-C_α distance in terms of projections along principal components

Knowing the rules for matrix multiplication, we can solve for each of the entries in Δr. For example, the deviation from average for residue 124 in the x direction is $\Delta r_{124x} = v_{1,124x}p_1 + v_{2,124x}p_2 + \ldots + v_{f,124x}p_f$ where $v_{i,124x}$ is the (scalar) entry in the ith eigenvector that corresponds to residue 124 in the x direction, etc. (For simplicity in notation, we omit explicitly writing out that the rs and ps are indeed functions of time.)

The squared distance between residues 124 and 338 (see Results for the reason for this choice of residues) can then be written as

$$[d_{124,338}(t)]^2 = (r_{124x} - r_{338x})^2 + (r_{124y} - r_{338y})^2 + (r_{124z} - r_{338z})^2$$

$$= (\langle r_{124x} \rangle_t - \langle r_{338x} \rangle_t + \Delta r_{124x} - \Delta r_{338x})^2 \quad (2)$$

$$+ (\langle r_{124y} \rangle_t - \langle r_{338y} \rangle^t + \Delta r_{124y} - \Delta r_{338y})^2$$

$$+ (\langle r_{124z} \rangle_t - \langle r_{338z} \rangle_t + \Delta r_{124z} - \Delta r_{338z})^2.$$

We define $\delta_{i,x} := v_{i,124x} - v_{i,338x}$ and $\xi_x := \langle r_{124x} \rangle_t - \langle r_{338x} \rangle_t$, and similarly for y and z; then we can write

$$r_{124x} - r_{338x} = \xi_x + \delta_{1,x}p_1 + \delta_{2,x}p_2 + \cdots + \delta_{f,x}p_f$$

$$= \xi_x + \sum_{c=1}^{f} \delta_{c,x}p_c,$$

and similarly for y and z. The square of these values is

$$(r_{124x} - r_{338x})^2 = \xi_x^2 + 2\sum_{c=1}^{f} \xi_x \delta_{c,x}p_c$$

$$+ \sum_{c_2=1}^{f} \sum_{c_1=1}^{f} \delta_{c_1,x}\delta_{c_2,x}p_{c_1}p_{c_2},$$

etc. By defining

$$S := \xi_x^2 + \xi_y^2 + \xi_z^2 = \langle d_{124,338} \rangle_t^2 \quad (3)$$

$$R_c := 2(\xi_x\delta_{c,x} + \xi_y\delta_{c,y} + \xi_z\delta_{c,z}) \quad (4)$$

$$Q_{c_1c_2} := \delta_{c_1,x}\delta_{c_2,x} + \delta_{c_1,y}\delta_{c_2,y} + \delta_{c_1,z}\delta_{c_2,z}, \quad (5)$$

we can rewrite Eq. 2 as

$$[d_{124,338}(t)]^2 = S + \sum_{c=1}^{f} R_c p_c + \sum_{c=1}^{f} \sum_{c_1=1}^{f} Q_{c_1c_2}p_{c_1}p_{c_2}. \quad (6)$$

We can calculate the time-independent S, $\mathbf{R} := [R_c]$, and $\mathbf{Q} := [Q_{c_1c_2}]$ simply by looking at the average structure $\langle r \rangle_t$ and the transformation matrix $\mathbf{T}$. Note that the square of the distance, $[d_{124,338}(t)]^2$, is not simply a linear combination of the projections $p_c(t)$, but also has cross-terms in the product of two projections $p_{c1}(t)p_{c2}(t)$.

Alternatively, this squared distance can be equated to

$$[d_{124,338}(t)]^2 = \sum_{\alpha=x,y,z} \left(\xi_\alpha + \sum_{c=1}^{f} \delta_{c,\alpha}p_c \right)^2 \quad (7)$$
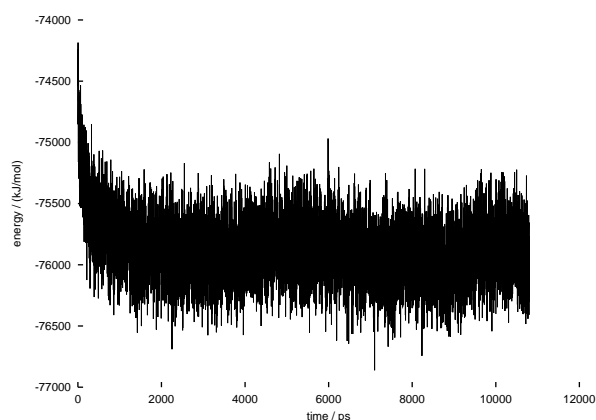
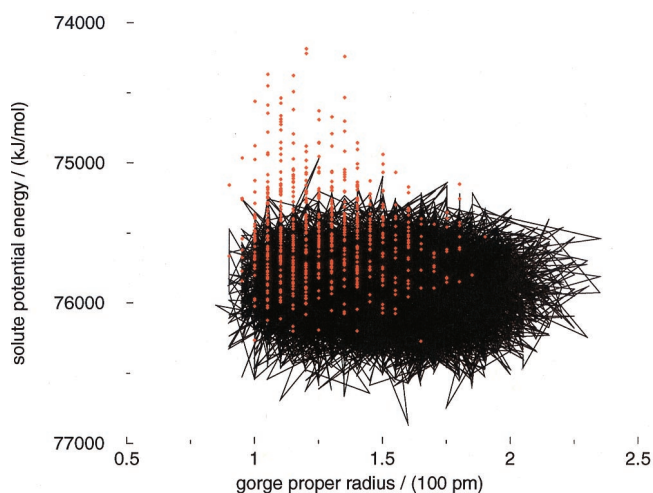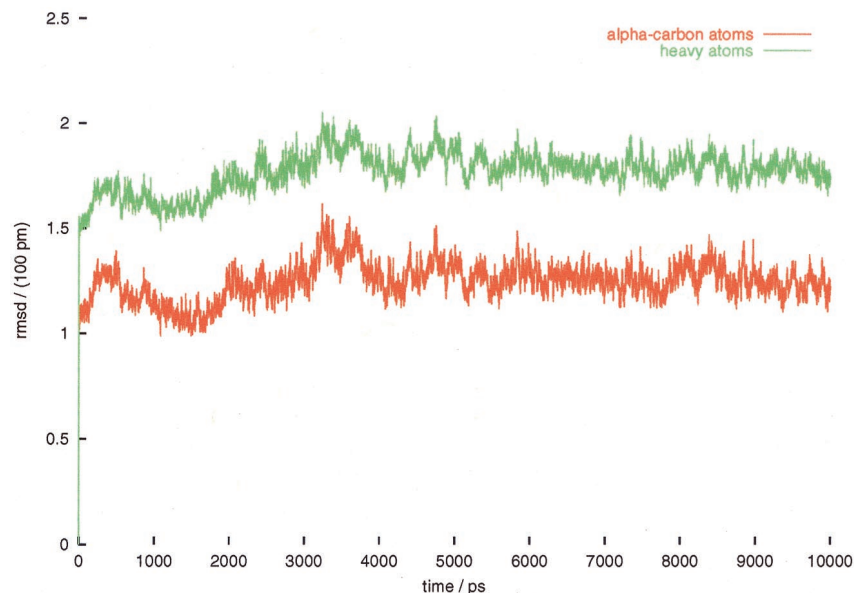FIGURE 1   The solute potential energy for the 10.8-ns trajectory.

FIGURE 2   The solute potential energy versus the gorge proper radius ρ. The red dots are the first 700 ps of trajectory. The black dots are the following 10 ns of trajectory, with the adjacent frames connected with lines.

Equations 6 and 7 have equivalent parts: the squared average distance $S = \langle d_{124,338} \rangle_t^2$ is $\Sigma_\alpha \xi_\alpha^2$; the linear part $\Sigma_c R_c p_c$ is $2\Sigma_\alpha \xi_\alpha (\Sigma_c \delta_{c,\alpha} p_c)$; and the cross-term part $\Sigma_{c2} \Sigma_{c1} Q_{c1c2} p_{c1} p_{c2}$ is $\Sigma_\alpha [(\Sigma_c \delta_{c,\alpha} p_c)^2]$. Because of this last correspondence, we expect to see that the cross-term part will always be larger than zero; this is indeed what we see in the Results.

RESULTS

Stability of trajectory and mobility of residues

Fig. 1 shows the solute potential energy of the 10.8-ns trajectory. After the first 700 ps, the solute potential energy ended its decreasing trend and fluctuated around 75.9 MJ mol^{-1}. This is one of our reasons for discarding the first 700 ps of trajectory and starting the equilibrium statistics thereafter.

Another justification for discarding the first 700 ps comes from Fig. 2, where we plot the solute potential energy with the gorge proper radius ρ. The points from the first 700-ps cluster are distinct from the rest of the trajectory for their high energies and small gorge radii. This suggests that the mAChE molecule relaxes from the conformation in the crystal structure 1MAH with higher energy and smaller gorge proper radius to one that has a wider gorge and is lower in energy.

The root mean square deviation (rmsd) from the crystal structure through the 10-ns trajectory is shown in Fig. 3. The rmsd for the α-carbon atoms kept stable around 120 pm, and that for all heavy atoms was maintained at 170 pm.

The temperature, the total energy, the mass density, and the volume during the 10-ns trajectory remained stable. The energy components were inspected to ensure the stability of the trajectory.

The isotropic temperature (B) factor can be calculated from the mean square fluctuation (msf) (McCammon and Harvey, 1987; Brooks et al., 1988) using the equation

FIGURE 3   Time dependence of the root mean square deviations (rmsd) for the 10-ns mAChE MD trajectory, using the mAChE part in the 1MAH crystal structure as reference. Red, rmsd for α-carbon atoms; green, rmsd for heavy atoms.

FIGURE 4 The *B* factor for each residue from the mAChE part in the 1MAH crystal structure. Black, crystal structure; red, calculated from the msf in the 10-ns MD trajectory.
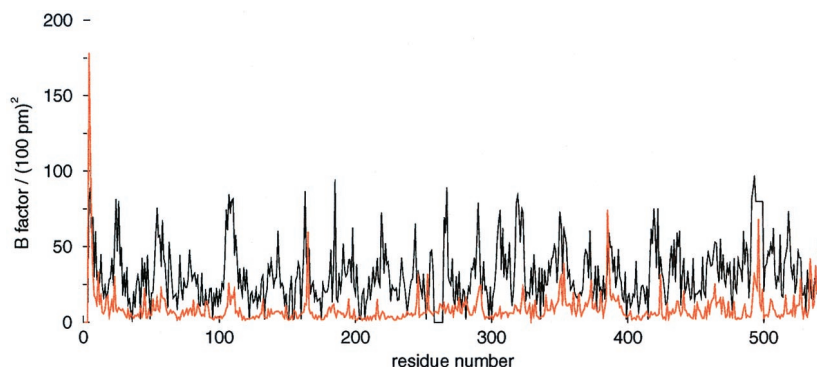
$$B = \frac{8\pi^2}{3}(\text{msf}). \qquad (8)$$

The *B* factor for each residue from the mAChE part in the 1MAH crystal structure and that calculated from the msf in the MD simulation are shown in Fig. 4.

The time-averaged rmsd for each residue, using the mAChE part in the 1MAH crystal structure as the reference, was also calculated; it showed similar features as the simulated *B* factor (data not shown). Overall, we do not observe outstanding fluctuations in the gorge region over other parts of the protein.

Back-door opening events

In the 10,000 frames analyzed, only 78 had back-door openings (Fig. 5). We compared $\Upsilon(t)$ with the dihedral angle values in the residues that form the back door, namely, Trp86, Gly448, Tyr449, and Ile451. The most important ones are shown in Fig. 6.

Correlation and covariance analyses

The 10-ns time series for the gorge proper radius $\rho(t)$ is shown in Fig. 7. The average of the proper radius is 152 pm. The probability distribution over all observed values of ρ is not Gaussian, as previously described (Shen et al., 2001).

We define the correlation between the *i*th degree of freedom, $r_i(t)$, and the gorge proper radius $\rho(t)$ to be

$$\frac{\langle(r_i(t) - \langle r_i \rangle_t)(\rho(t) - \langle \rho \rangle_t)\rangle_t}{\sqrt{\langle(r_i(t) - \langle r_i \rangle_t)^2\rangle_t \langle(\rho(t) - \langle \rho \rangle_t)^2\rangle_t}}. \qquad (9)$$

We calculated a correlation vector with three entries thus defined for the α-carbon atom of each residue (Fig. 8). Also, we calculated the average correlation vector within each of the secondary structure elements (α-helices and β-strands; Fig. 9).

Similarly, the average velocity covariance is defined as

$$\langle(r_i(t) - r_i(t - 1\ \text{ps}))(\rho(t) - \rho(t - 1\ \text{ps}))\rangle_t. \qquad (10)$$

The average velocity covariance vector is shown in Fig. 10.

Principal component analysis and back-projecting

We found that the distance between Phe338 $C_{\epsilon 2}$ and Tyr124 O_H and the gorge proper radius are highly correlated, with a correlation coefficient of 0.91 (Fig. 11); the correlation coefficient between the distance Phe338 C_α-Tyr124 C_α and the gorge proper radius is 0.55.

We calculated the time-dependent contributions to $[d_{124,338}(t)]^2$, namely, the linear part $\Sigma_c R_c p_c(t)$ and the cross-term part $\Sigma_{c1}\Sigma_{c2} Q_{c1c2} p_{c1}(t)p_{c2}(t)$ (Fig. 12). Note that sometimes the linear part and the cross-term part have comparable magnitudes but opposite directions, for example, around 8.1 ns.
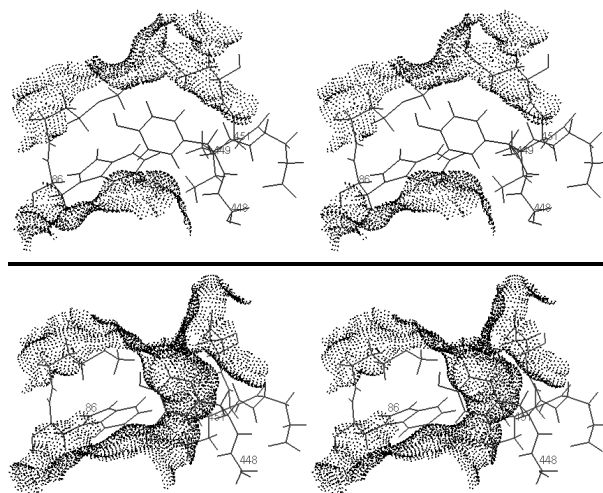

FIGURE 5 Stereo views of two conformations in the back door region. In each conformation, the outside of the protein is on the top; Trp86 is on the lower left, Gly448 on the lower-right, and Tyr449 in the middle. The Connolly surface generated with a probe radius of 140 pm is shown. (*Top*) The frame at 291 ps, showing a closed back door. In this frame, χ_2 of Trp86 is 114°, ψ of Gly448 is 86°, χ_1 of Tyr449 is $-121°$, χ_1 of Ile451 is $-72°$, and χ_2 of Ile451 is 161°. (*Bottom*) The frame at 8996 ps, showing an open back door. In this frame, χ_2 of Trp86 is 120°, ψ of Gly448 is 65°, χ_1 of Tyr449 is $-122°$, χ_1 of Ile451 is $-64°$, and χ_2 of Ile451 is 163°.

FIGURE 6 Selected dihedral angles through the 10 ns that are important to the opening of the back door. The red lines at the bottom mark the back door opening events. The dots are the dihedral angles: χ_2 of Trp86 (*green*); ψ of Gly448 (*blue*); χ_1 of Tyr449 (*pink*); χ_1 of Ile451 (*cyan*); and χ_2 of Ile451 (*brown*).
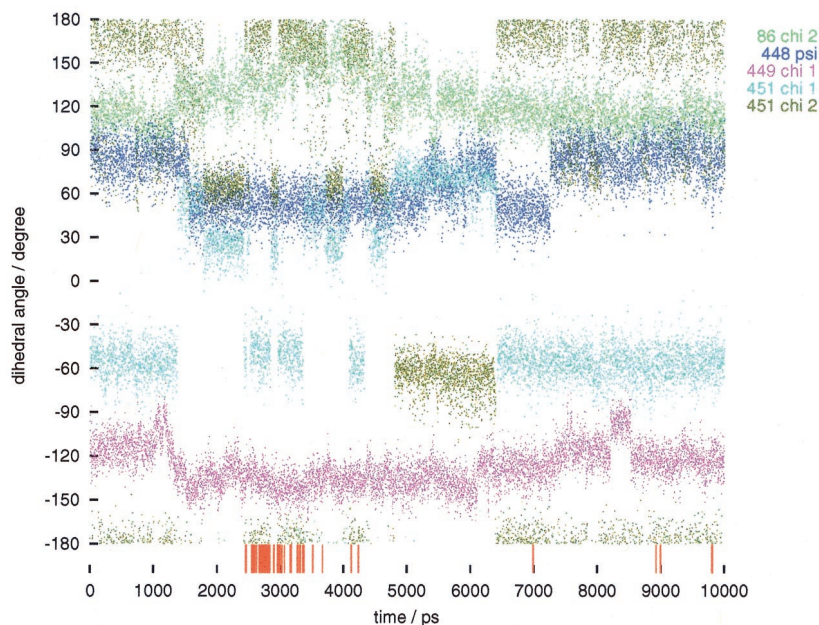
DISCUSSION

B factors

The crystallographic *B* factors include a variety of contributions (e.g., crystal contacts and static disorder in the crystal) in addition to that from the dynamical fluctuations, and these additional contributions are expected to be substantial for a structure of modest resolution (Karplus and McCammon, 1981). The crystallographic resolution is 0.32 nm (3.2 Å) for the structure 1MAH (Bourne et al., 1995). Generally, the crystallographic *B* factors were larger than those observed in MD because the simulation sampled far fewer conformations of the protein than crystallography; the termini, however, exhibited larger *B* factors in the MD simulation where the protein is solvated in water rather than packed in a crystal.
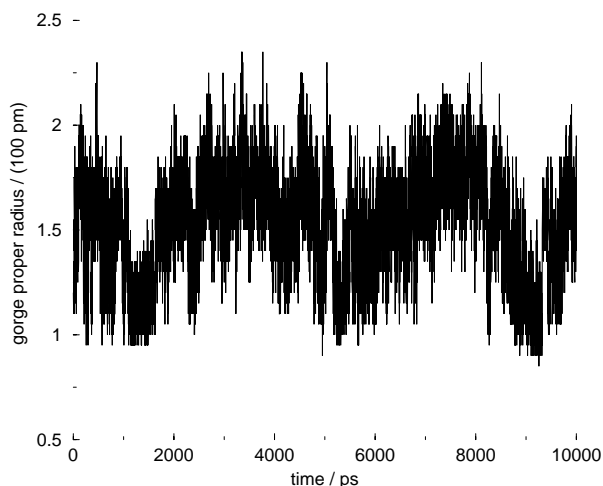

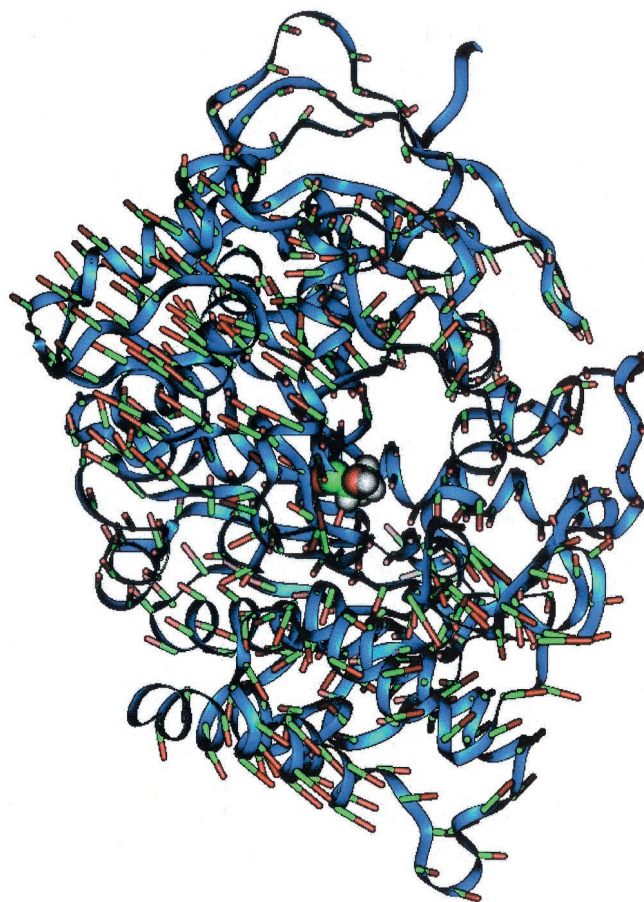FIGURE 7 The 10-ns time series for the gorge proper radius, $\rho(t)$.

FIGURE 8 The correlation vector for each residue displayed on the initial structure (see text for definition). The vectors point from green to red; the lengths are in arbitrary units. The AChE molecule is shown in the conventional orientation: the viewer looks down the gorge, with the N-terminus on the top of the figure and the C-terminus at the lower left corner. The active site Ser203 is shown in the space-filling model.

FIGURE 9 The correlation vector within each secondary structure element displayed on the initial structure (see text for definition). The α-helices are shown as green cylinders; the β-strands are shown in yellow. The vectors point from green to red; the lengths are in arbitrary units. The AChE molecule is shown in the conventional orientation: the viewer looks down the gorge, with the N-terminus on the top of the figure and the C-terminus at the lower left corner. The active site Ser203 is shown in the space-filling model. Helix 14 is marked with numerals 14.
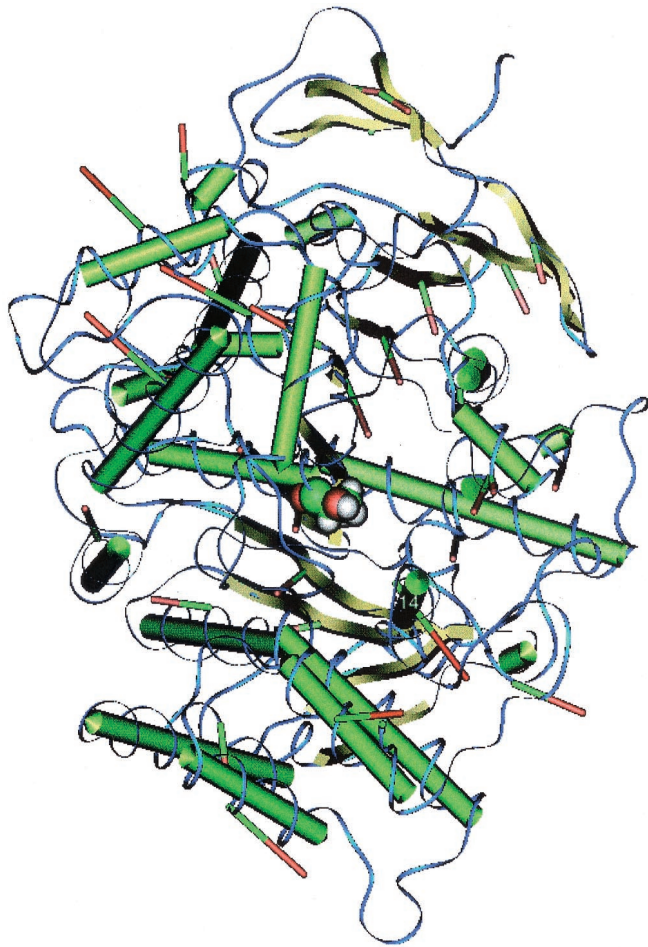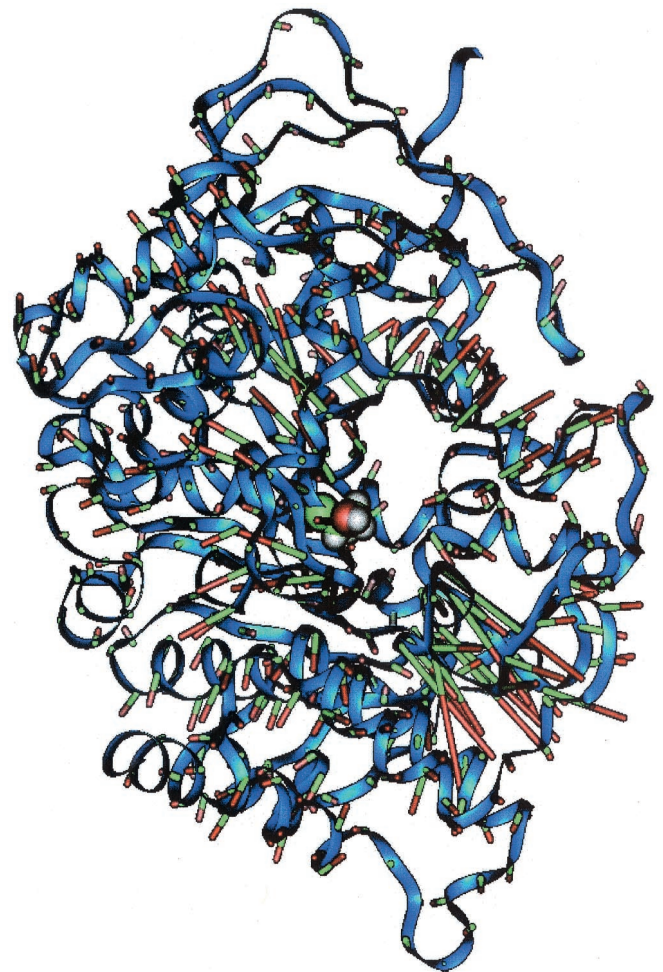
FIGURE 10 The average velocity covariance vector for each residue displayed on the initial structure (see text for definition). The vectors point from green to red; the lengths are in arbitrary units. The AChE molecule is shown in the conventional orientation: the viewer looks down the gorge, with the N-terminus on the top of the figure and the C-terminus at the lower left corner. The active site Ser203 is shown in the space-filling model.

In the original crystal structure (Bourne et al., 1995), loop II of Fas2 interacts with the peripheral anionic site of mAChE, thereby blocking substrate access to the active site; loop I of Fas2 fits in a crevice near the entrance of the gorge. Our simulation starts from this crystal structure, with the inhibitor Fas2 removed. However, we do not see larger B factors in MD than in the crystal structure for the residues involved in contact with Fas2; this is probably due to the same reasons mentioned in the previous paragraph. Nor were any outstanding deviations observed in the Fas2 contact interface in the time-averaged rmsd for each residue.

Access to the active site

Unlike the previous 0.5-ns MD simulation on the *Torpedo californica* AChE (TcAChE) complexed with tacrine

(Wlodek et al., 1997), our simulation showed a non-Gaussian probability distribution of the gorge proper radius as previously described (Shen et al., 2001). In addition, our average radius, 152 pm, was smaller than that in the tacrine-complexed TcAChE simulation (~190 pm). The time series $\rho(t)$ almost never reached the 240 pm; for a substrate as large as acetylcholine, AChE hardly allowed it any spontaneous access to the active site during the 10-ns trajectory.

The presence of the ligand tacrine in the gorge may be the reason for the larger proper radius in the tacrine-complexed TcAChe simulation. Favorable electrostatic interactions between AChE and its substrate may help in overcoming the difficulty of squeezing through a narrow gorge. Note that the catalysis of AChE occurs on a millisecond time scale; as long as the frequency of gorge opening is not so low that the substrate diffuses away from the gorge entrance before commitment to catalysis, rarity of opening in this nano-
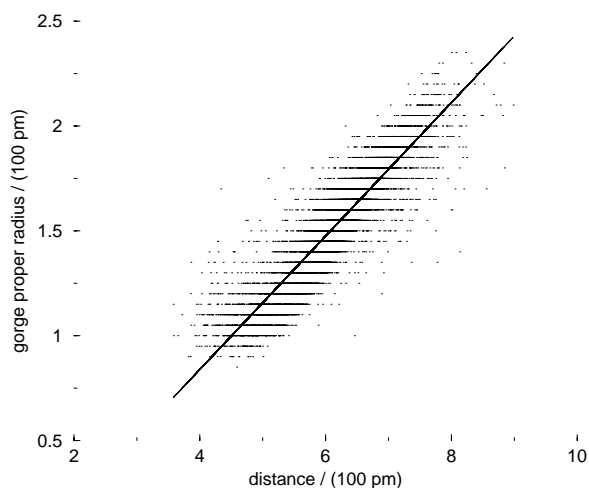
FIGURE 11 The gorge proper radius ρ versus the distance between Phe338 $C_{\epsilon 2}$ and Tyr124 O_H. The correlation coefficient is 0.91.

second simulation does not preclude diffusion-controlled kinetics.

Back-door opening events

The proposition for the existence of the back door has been based on visual observation of the crystal structure (Ripoll et al., 1993) and supported by conformations sampled by simulations of TcAChE (Gilson et al., 1994; Faerman et al., 1996). In our simulation, although only 78 frames out of our 10-ns trajectory had back-door opening events, the back-door opening observed here aligned sequentially with that observed in the MD simulation of TcAChE: Trp86, Gly448, and Tyr449 in mAChE correspond respectively to Trp84, Gly441, and Tyr442 in TcAChE. This is a remarkable fact;
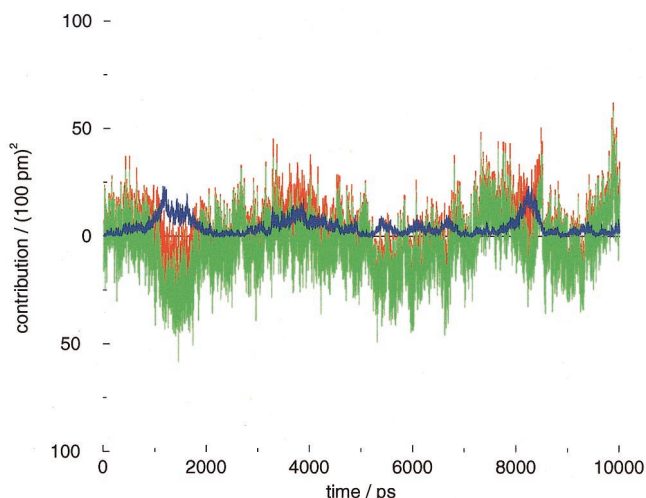

FIGURE 12 Contribution of linear ($\Sigma_c R_c p_c(t)$, green) and cross $\Sigma_{c1}\Sigma_{c2} Q_{c1c2} p_{c1}(t) p_{c2}(t)$, blue) terms to the square distance $[d_{124,338}(t)]^2$ and the sum of these two (red). See text for definitions.

this alternative opening, named the back door, is not limited to a single species but has been observed by MD simulations in at least two species.

The selected dihedral angles in the region, as shown have preferred values when the back door opens. For most of the opening events, e.g., around 3 ns, χ_2 of Trp86 acquires higher values around 150° than the usual values around 110°. Back-door opening is more likely to happen when ψ of Gly448 assumes a value around 45° rather than around 80° and when χ_1 of Tyr449 is around $-140°$ rather than $-120°$. Four major types of rotamers for the χ_1 and χ_2 of Ile451 are observed: 1) ($\chi_1 \approx -60°$, $\chi_2 \approx \pm 180°$), or ($-gauche$, $trans$); observed most often in this trajectory, this seems to be almost a necessary, but not sufficient, condition for back-door opening; 2) ($\chi_1 \approx 30°$, $\chi_2 \approx 60°$), or ($+gauche$, $+gauche$); observed sporadically in the first half of the trajectory, it is interspersed between opening events; 3) ($\chi_1 \approx 60°$, $\chi_2 \approx -60°$), or ($+gauche$, $-gauche$); observed around 5 ns and 6 ns, this rotamer does not give any opening events; 4) ($\chi_1 \approx -60°$, $\chi_2 \approx 60°$), or ($-gauche$, $+gauche$); observed after 7 ns, this rotamer also does not give any opening events.

It is difficult to be conclusive about preferred dihedral angle values for back-door opening for the following reasons. We have only a small number of opening events. In addition, some of the opening events in the last 2 ns of the simulation have violations of preferred dihedral angle values; for example, the two conformations shown in Fig. 5 have subtly different values for these dihedral angles, but the one at 8996 ps has an open back door but that at 291 ps does not.

Site-directed mutagenesis experiments on human AChE (Kronman et al., 1994) and TcAChE (Faerman et al., 1996) do not support significant participation of back-door traffic in catalytic activities. On the other hand, there is recent experimental evidence from two different methods supporting the back-door hypothesis: product clearance through the back door is implied by the crystal structure of a carbamoylated TcAChE (Bartolucci et al., 1999); one inhibitory monoclonal antibody of *Electrophorus* AChE is reported to bind to the region of the back door (Simon et al., 1999). It is our view that one has to be cautious while attempting to interpret and reconcile these data concerning the back door, and take into consideration the time scale of each method of observation. Indeed, even this simulation of 10 ns covers only a fraction of a thousandth of a catalytic cycle of AChE. Despite consistently observing conformations with an open back door in several MD simulations, we cannot ascertain details of the involvement of the back door in AChE catalytic function.

Correlation and covariance analyses

We have found that the porcupine plots (Figs. 8–10) are very helpful in visualizing the concerted motions between

parts of the protein and the functionally interesting motion, the gorge opening. The correlation vectors in Fig. 8 reveal how much each residue moves in concert with the gorge. Note that a correlation vector does not show how much the residue fluctuates or how flexible it is, but how it moves from its average position when the gorge changes in proper radius. The moiety of AChE that includes the gorge (the half of the protein that is closer to the viewer in the figures) has remarkable concerted motion with the gorge proper radius. The residues in this moiety generally have correlation vectors pointing away from the gorge. These residues apparently move away from the gorge when the gorge opens. Some residues that are in the other moiety (closer to the base of the gorge and further from the viewer) have correlation vectors that are generally smaller. The discrimination of these concerted motions is even more pronounced in Fig. 10: the residues that construct the gorge itself have the largest average velocity covariance vectors, whereas those farthest from the gorge have the smallest average velocity covariance vectors. We introduce Fig. 9 as an easy way to visualize and interpret how the secondary structure elements contribute to the opening of the gorge. For example, helix 14 (Gly335 to Tyr341) is close to the active site gorge; it has a large vector in this figure pointing away from the gorge, which means it moves away when the gorge opens. This helix includes Phe338, one of the two residues whose separation correlates well with the gorge proper radius, as described above. Most of the α-helices in the same moiety as the gorge move away from the gorge; those further from the gorge show little concerted motion. The helices and two large β-sheets (B and C) in the other moiety, which is at the bottom of the gorge, generally have upward motion toward the gorge as the latter opens.

We speculate that inhibitors such as Fas2 may decrease the likelihood of gorge opening by restricting groups on the surface of AChE that have large concerted motions, in addition to sterically occluding the entrance to the gorge. Other allosteric inhibition mechanisms are also possible. These remain to be investigated and confirmed by our current work on the MD simulation of the complex of Fas2 and mAChE. We understand that in vivo AChE exists in cholinergic synapses as oligomers indirectly anchored on the post-synaptic membrane. These attachments to the AChE monomer are not represented in this simulation; they may also modulate the opening behavior of the gorge.

Contributions of the principal components in gorge opening

Several convincing pieces of evidence support a complex opening behavior of the gorge. Starting from a naive guess that one single principal component or just a few, representing a large collective motion in the protein, has a dominant contribution in the gorge opening behavior, we shall now enumerate evidence to demonstrate the contrary.

There is a high correlation (0.91) between the Phe338 $C_{\epsilon2}$-Tyr124 O_H distance and the gorge proper radius; this distance alone may be used to predict the gorge proper radius with high confidence. The corresponding α-carbon distance has a lower correlation coefficient (0.55). The side chains seem to contribute significantly to the opening of the gorge. As these particular side chains project directly toward the gorge and actually constitute the bottleneck much of the time, it makes sense that the correlation of the side-chain distance and the gorge radius will be particularly strong. A predictive model of the gorge width with high accuracy will thus have to include the side chains into consideration; for that purpose, it is inappropriate to concentrate only on large backbone motions and ignore those of the side chains. On the other hand, the largest fluctuations in the gorge radius can be expected to reflect movement of backbone segments, and we investigate these in the current work. Ligands larger than acetylcholine do bind to the active site, including a number of important inhibitors, and with the backbone correlation analysis we are beginning to explore the mechanism of such processes.

The eigenvalues from the PCA, ranked by magnitude, decrease rapidly: the 15th largest eigenvalue is less than one-tenth (0.1) of the largest, and the 113th largest is less than one-hundredth (0.01) of the largest. Despite this, there is not one principal component that has an outstandingly high correlation with $\rho(t)$; in other words, no one principal component dominates the opening of the gorge. The naive guess expects dismissible cross-term parts $(\Sigma_{c2}\Sigma_{c1}Q_{c1c2}p_{c1}p_{c2})$ in the back-projecting expression (Eq. 6); in addition, it expects the linear part $(\Sigma_c R_c p_c(t))$ to have one dominant term with a large coefficient R_c, while all other terms have small coefficients. Contrary to this expectation, when ranked by magnitude, there was not a dominant minority of terms that stand out in the list of the R_c coefficients; indeed, none of the $R_c p_c(t)$ terms in the linear part are dominant (data not shown). It is not possible to single out principal components that dominate the width of the gorge.

Furthermore, we saw in Fig. 12 that a model that ignores the contribution of either the cross-term part or the linear part to the gorge opening behavior is not justified. Both the linear part and the cross-term part contribute significantly to fluctuations in the width of the gorge. In sum, the naive guess that a few principal components dominate the gorge is not justified.

We conjecture that this finding is related to our fractal dynamics model previously proposed (Shen et al., 2001). In that model, the time-dependent fluctuation of gorge proper radius reveals fractal dynamics that lacks a characteristic time scale over the several orders of magnitude being observed. We suggested that this behavior is caused by a hierarchy of motions on different time scales acting together. It is likely that these motions are related to our present results, where many motions on different length

FIGURE 13 The solute potential energy projected onto a histogram of the first two principal components.
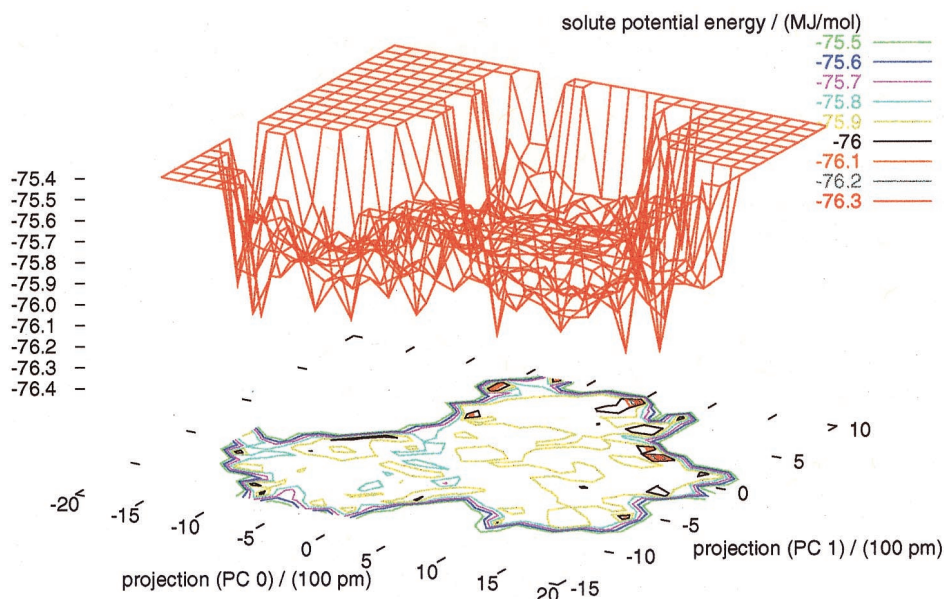
scales (as revealed by PCA) have a similar effect on the gorge behavior.

As a side note, we did attempt to perform energy landscape analysis (Becker, 1997a,b, 1998) by projection of the solute potential energy onto the first two principal components obtained from PCA (Fig. 13). For a protein as large as AChE, our 10,000 frames did not provide enough sampling to show a meaningful energy landscape. Comparison between the solute potential energy and the gorge proper radius was also inconclusive, likely for the same reason.

CONCLUSIONS

The 10-ns MD simulation of mAChE was equilibrated and stable as shown by various properties and energy components and gave reasonable B factors.

The small fraction of total number of frames in which we observed an open back door severely limits the statistics of our analysis. Back-door opening seems to prefer certain dihedral angle values for some of the residues in the region.

We developed the porcupine plots to visualize the concerted motions between residues in AChE and the gorge width. The plots show that residues lining the gorge have the largest velocity covariance with the gorge, and residues in the same moiety as the gorge move concertedly away from the gorge when it opens. This clearly reveals the residues involved in gorge opening and gives the magnitude and direction of such involvement.

We have found that the Phe338 $C_{\epsilon 2}$-Tyr124 O_H distance is highly predictive of the gorge proper radius; the corresponding α-carbon distance has a moderate correlation. We have not investigated side-chain motions extensively in this paper, but side chains will be important in a predictive model of the gorge opening.

Principal component analysis showed that the gorge proper radius is determined by motions on many different length scales; this is likely to be related to the fractal dynamics model we previously proposed.

REFERENCES

Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. *Proteins Struct. Funct. Genet.* 17:412–425.

Balsera, M. A., W. Wriggers, Y. Oono, and K. Schulten. 1996. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* 100:2567–2572.

Bartolucci, C., E. Perola, L. Cellai, M. Brufani, and D. Lamba. 1999. "Back door" opening implied by the crystal structure of a carbamoylated acetylcholinesterase. *Biochemistry.* 38:5714–5719.

Becker, O. M. 1997a. Quantitative visualization of a macromolecular potential energy "funnel." *J. Mol. Struct. (Theochem).* 398–399: 507–516.

Becker, O. M. 1997b. Geometric versus topological clustering: an insight into conformational mapping. *Proteins.* 27:213–226.

Becker, O. M. 1998. Principal coordinate maps of molecular potential energy surfaces. *J. Comp. Chem.* 19:1255–1267.

Bourne, Y., P. Taylor, and P. Marchot. 1995. Acetylcholinesterase inhibition by fasciculin: crystal structure of the complex. *Cell.* 83:503–512.

Brooks, III, C. L., M. Karplus, and B. M. Pettitt. 1988. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics, Vol. LXXI of Wiley Series on Advances in Chemical Physics. Wiley-Interscience, New York.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

Darden, T., D. York, and L. Pedersen, L. 1993. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

Faerman, C., D. Ripoll, S. Bon, Y. Le Feuvre, N. Morel, J. Massoulié, J. L. Sussman, and I. Silman. 1996. Site-directed mutants designed to test back-door hypothesis of acetylcholinesterase function. *FEBS Lett.* 386: 65–71.

García, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.

Gilson, M. K., T. P. Straatsma, J. A. McCammon, D. R. Ripoll, C. H. Faerman, P. H. Axelsen, I. Silman, and J. L. Sussman. 1994. Open "back door" in a molecular dynamics simulation of acetylcholinesterase. *Science.* 263:1276–1278.

Karplus, M., and J. A. McCammon. 1981. The internal dynamics of globular proteins. *CRC Crit. Rev. Biochem.* 9:293–349.

Kronman, C., A. Ordentlich, D. Barak, B. Velan, and A. Shafferman. 1994. The "back door" hypothesis for product clearance in acetylcholinesterase challenged by site-directed mutagenesis. *J. Biol. Chem.* 269: 27819–27822.

McCammon, J. A., and S. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, UK.

Ripoll, D. R., C. H. Faerman, P. H. Axelsen, I. Silman, and J. L. Sussman. 1993. An electrostatic mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* 90:5128–5132.

Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comp. Phys.* 23: 327–341.

Shen, T. Y., K. Tai, and J. A. McCammon. 2001. Statistical analysis of the fractal gating motions of the enzyme acetylcholinesterase. *Phys. Rev. E.* 63:041902.

Simon, S., A. Le Goff, Y. Frobert, J. Grassi, and J. Massoulié. 1999. The binding sites of inhibitory monoclonal antibodies on acetylcholinesterase. *J. Biol. Chem.* 274:27740–27746.

Straatsma, T. P., M. Philippopoulos, and J. A. McCammon. 2000. NWChem: exploiting parallelism in molecular simulation. *Comp. Phys. Commun.* 128:377–385.

Tara, S., T. P. Straatsma, and J. A. McCammon. 1999. Mouse acetylcholinesterase unliganded and in complex with huperzine A: a comparison of molecular dynamics simulations. *Biopolymers.* 50:35–43.

Wlodek, S. T., T. W. Clark, L. R. Scott, and J. A. McCammon. 1997. Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. *J. Am. Chem. Soc.* 119:9513–9522.

Zhou, H.-X., S. T. Wlodek, and J. A. McCammon. 1998. Conformation gating as a mechanism for enzyme specificity. *Proc. Natl. Acad. Sci. U.S.A.* 95:9280–9283.