The influence of conformational fluctuations on enzymatic activity: modelling the functional motion of β-secretase

2005 J. Phys.: Condens. Matter 17 S1581

(http://iopscience.iop.org/0953-8984/17/18/014)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **17** (2005) S1581–S1593 doi:10.1088/0953-8984/17/18/014

The influence of conformational fluctuations on enzymatic activity: modelling the functional motion of β-secretase

M Neri[1], **M Cascella**[1,2] **and C Micheletti**[1]

[1] SISSA and INFM, Via Beirut 2-4, I-34014 Trieste, Italy
[2] EPFL, Lausanne, Switzerland

Received 1 October 2004, in final form 15 November 2004
Published 22 April 2005
Online at stacks.iop.org/JPhysCM/17/S1581

Abstract

Considerable insight into the functional activity of proteins and enzymes can be obtained by studying the low energy conformational distortions that the biopolymer can sustain. We carry out the characterization of these large scale structural changes for a protein of considerable pharmaceutical interest, the human β-secretase. Starting from the crystallographic structure of the protein, we use the recently introduced β-Gaussian model to identify, with negligible computational expenditure, the most significant distortions occurring in thermal equilibrium and the associated timescales. The application of this strategy helps us to gain considerable insight into the putative functional movements and, furthermore, allows us to identify a handful of key regions in the protein which have an important mechanical influence on the enzymatic activity despite being spatially distant from the active site. The results obtained within the Gaussian model are validated through an extensive comparison against an all-atom molecular dynamics simulation.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In the past few decades several experimental and theoretical studies have pointed out that, although proteins possess an atomic density comparable to that of crystalline solids, they are much more flexible that the latter [1–3]. This unusual elasticity, whose origin putatively resides in the neat secondary and tertiary protein organization [4], is a prerequisite for biological functionality. In fact, in order to carry out their biological tasks, proteins and enzymes need to sustain conformational distortions where groups of several amino acids are significantly displaced from the reference native states. The timescales associated with such configurational changes, which may occur over several nanoseconds, are also long compared to those of atomic motions [5].

From a computational point of view, the most direct way of observing these rearrangements would be by recourse to a molecular dynamics simulation [6]. Present implementations of this scheme allow one to follow the dynamical evolution of a large protein (of a few hundred residues) in its surrounding solvent for about one nanosecond. This timescale is large enough for gaining considerable insight into several dynamical aspects of large proteins but may be inadequate for characterizing accurately the functional movements mentioned before [7].

Several studies have attempted to bridge the gap between the timescales of feasible MD simulations and those of biologically relevant protein movements by resorting to a mesoscopic rather than a microscopic approach [8]. A key contribution in this framework has been the observation that the overdamped dynamics of a protein in its solvent can be described as occurring in an effective quadratic potential [9–13]. This observation was further supported by Tirion who pointed out that, in a normal mode analysis of protein vibrations, the complicated classical force field could be replaced by harmonic couplings with the same spring constants [8]. These results stimulated a variety of studies where the elastic properties of proteins have been described by means of coarse-grained models where amino acids are replaced by effective centroids (corresponding to the C_α and/or C_β atoms) and the energy function is reduced to harmonic couplings between pairs of spatially close centroids. These approaches have been found to be in accord with both experimental and MD characterization of the overall protein elastic behaviour [14–18]. In particular, the recently introduced β-Gaussian model has been shown to be quite effective in identifying, with a modest computational effort, the most important conformational changes that a protein can undergo in thermal equilibrium [19]. These vibrational modes are here investigated for a protein, human β-secretase (denoted as BACE hereafter) which is an important representative of the pepsins, a family of enzymes which are capable of cleaving a peptidic substrate through a chemical reaction (hydrolysis) involving an aspartic dyad. It is important to point out that a distinctive feature of this enzyme is the fact that the two aspartates are located on opposites sides of the interface between the two lobes. Recently, the BACE, a membrane-anchored extracellular protein, has been the subject of several experimental investigations since it constitutes a key target for drugs used in treatments of Alzheimer's disease (AD) [20, 21]. The onset of this neuro-degenerative disease is related to the formation of β-amyloid plaques in brain tissues. Recent investigations have demonstrated that the β-amyloid peptides are the product of the cleavage of a precursor protein operated by the BACE. These results have motivated a large body of scientific investigations aimed at clarifying the cleavage mechanisms of the BACE and its differences in functionality with respect to typical members of the pepsin family. Interestingly, from a structural point of view, see figure 1, the core region of the BACE presents only minor differences from typical pepsin members. Most of the structural changes are instead located at the surface of the protein, in the form of six loop insertions, and at the C-terminus, by a 35-residue long extension [22].

As is visible in figure 1, where the BACE is shown together with a bound inhibiting peptide, a long cleft of 35 Å is present at the interface of the two lobes (highlighted with different colours). Interestingly, a very small number of residue pairs are found in contact at the lobe interface and they are mostly constituted by the aspartic acids involved in the catalytic activity. From analogy with other enzymes [23, 24, 19] one may anticipate that also the BACE catalytic action depends not only on the favourable chemical interaction of the substrate and the aspartic dyads but also on the possibility that the conformational fluctuations of the enzyme itself may modulate the activity. We shall therefore undertake the task of characterizing in detail the elastic response of the BACE by resorting to the β-Gaussian model. An independent term of comparison for the robustness of these findings will be provided by an all-atom molecular dynamics simulation. It is found *a posteriori* that the two methods, which are very different in spirit, provide a consistent picture for the largest protein rearrangements occurring in thermal

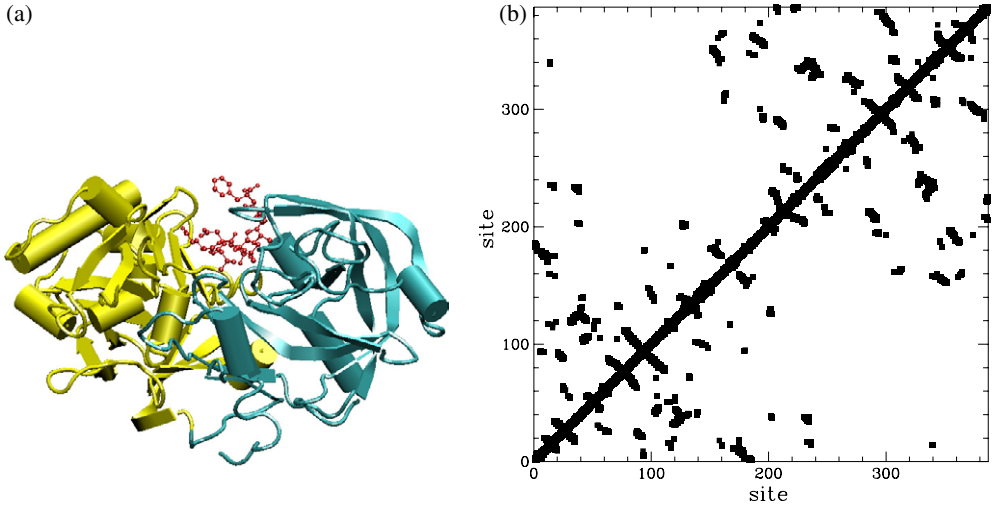

Figure 1. (a) The structure of the human β-secretase dimer (PDB file 1fkn) complexed with an inhibiting peptide (represented with a wire frame). (b) The contact matrix, Δ, associated with the native (crystallographic) structure of the enzyme.

equilibrium. Finally, we characterize the extent to which the motion of a substrate bound to the enzyme is correlated with the latter. By these means we identify a limited number of amino acids in the enzyme which have a strong mechanical bearing of the conformational fluctuations of the enzyme despite the lack of spatial proximity.

The program which implements the β-Gaussian model is available, upon request, from the authors.

2. The β-Gaussian model

Since the main objective is the modelling of the large scale fluctuations in a protein, it is convenient to reduce the spatial degrees of freedom of the biopolymer through a coarse-graining where a two-particle representation is used for each amino acid. Besides the C_α atom, an effective C_β centroid is employed to capture, in the simplest possible way, the side-chain orientation in a given amino acid (except for Gly for which only the C_α atom is retained). This reduced structural representation affects the form of the effective Hamiltonian associated with the coarse-grained structural representation. Arguably, the simplest energy function for the system can be constructed by assuming that all centroids in the protein (C_α and/or C_β) whose separation is smaller than a given interaction distance, R, interact through the same pairwise potential, V [8]. The information of which centroids are in interaction in the native state is aptly summarized in the native contact matrix Δ_{ij}^{XY} which takes on the values of 1 (0) if the native separation of the particles of type X and Y, belonging respectively to residues i and j, is below (above) R. The contact map of the BACE considered in this study is shown in figure 1. The system energy function evaluated on a trial structure, Γ, can then be written as

$$\mathcal{H}(\Gamma) = \mathcal{H}_{BB}(\Gamma) + \mathcal{H}_{\alpha\alpha}(\Gamma) + \mathcal{H}_{\alpha\beta}(\Gamma) + \mathcal{H}_{\beta\beta}(\Gamma) \tag{1}$$

where

$$\begin{aligned}
\mathcal{H}_{BB}(\Gamma) &= K \sum_{i} V(d_{i,i+1}^{CA-CA}) \\
\mathcal{H}_{\alpha\alpha}(\Gamma) &= \sum_{i<j} \Delta_{ij}^{CA-CA} V(d_{i,j}^{CA-CA}) \\
\mathcal{H}_{\alpha\beta}(\Gamma) &= \sum_{i,j} \Delta_{ij}^{CA-CB} V(d_{i,j}^{CA-CB}) \\
\mathcal{H}_{\beta\beta}(\Gamma) &= \sum_{i<j} \Delta_{ij}^{CB-CB} V(d_{i,j}^{CB-CB}).
\end{aligned} \tag{2}$$

In expressions (2) we indicated with d_{ij}^{XY} the actual separation of the particles in the trial structure, Γ. The indices i and j run over all integer values ranging from 1 up to the protein length, N. To account for the protein chain connectivity we have introduced in (2) a 'backbone' energy term which mimics the higher strength of the bond for consecutive amino acids with respect to non-covalent contact interactions. Consistently with previous studies [19] we have set R equal to 7.5 Å and $K = 1$.

The general pairwise Hamiltonian of equation (1) is subject to an important requirement since it must guarantee that the assigned reference conformation is at the global energy minimum. The customary way to accomplish this [14, 17, 18] is to assume that the potentials V^{X-Y} appearing in (2) attain their global minimum in correspondence of the native separation of the centroids. For small fluctuations around the native structure, the interaction energy of two centroids, i and j, can then be expanded in terms of the deviations from the native distance vector, $\mathbf{r}_{ij}$. If we indicate the deviation vector as $\mathbf{x}_{ij}$, so that the total distance vector is $\mathbf{d}_{ij} = \mathbf{r}_{ij} + \mathbf{x}_{ij}$, we can approximate the pairwise interaction as

$$V(d_{ij}) \approx V(r_{ij}) + \frac{k}{2} \sum_{\mu,\nu} \frac{r_{ij}^{\mu} r_{ij}^{\nu}}{r_{ij}^{2}} x_{ij}^{\mu} x_{ij}^{\nu} \tag{3}$$

where μ and ν denote the Cartesian components, x, y and z, and k is the second derivative of V at its minimum. The expansion of equation (3) brings about a dramatic simplification of Hamiltonian (1) which, in fact, acquires a quadratic dependence in terms of the deviation vectors, $\mathbf{x}$. However, a further simplification of the energy function can be achieved by exploiting the fact that the coordinates of the C_α centroids encode, in an almost unique way, the positions of all other atoms in the protein, and hence also of the C_β atoms [25, 26]. The construction scheme adopted in the β-Gaussian model (denoted as β-GM in the following) defines the location of the ith C_β through a co-planar version of the Park and Levitt construction rule [27]:

$$\mathbf{r}_{CB}(i) = \mathbf{r}_{CA}(i) + l \frac{2\mathbf{r}_{CA}(i) - \mathbf{r}_{CA}(i+1) - \mathbf{r}_{CA}(i-1)}{|2\mathbf{r}_{CA}(i) - \mathbf{r}_{CA}(i+1) - \mathbf{r}_{CA}(i-1)|} \tag{4}$$

where $l = 3$ Å. Thus, the degrees of freedom can be reduced to only those of the C_αs. The linear relationship between the C_α and C_β coordinates allows us to recast the Hamiltonian as follows:

$$\mathcal{H} = \tfrac{1}{2} k \sum_{ij,\mu\nu} x_{i,\mu} \mathcal{M}_{ij,\mu\nu} x_{j,\nu}, \tag{5}$$

where $x_{i,\mu}$ is the deviation of ith C_α along the μ axis and $\mathcal{M}$ is a $3N \times 3N$ symmetric matrix. The elastic response of the system is uniquely dictated by the eigenvalues and eigenvectors of $\mathcal{M}$.

In thermal equilibrium, each amino acid moves under the action of the quadratic Hamiltonian (5) subject to a viscous friction originating from interactions with the surrounding

solvent as well as with the rest of the protein [28]. The viscous hindrance of the motion is so important that the protein dynamics becomes severely overdamped [10–13, 28]. In this case it seems appropriate to describe the dynamics of the amino acids within a Langevin framework [30, 29]:

$$\gamma_i \dot{x}_{i,\mu}(t) = -k \sum_{j,\nu} \mathcal{M}_{ij,\mu\nu} x_{j,\nu}(t) + \eta_{i,\mu}(t), \tag{6}$$

where γ_i is the viscous friction coefficient and $\eta_{i,\mu}(t)$ is a stochastic noise satisfying the relations [29, 30]

$$\langle \eta_{i,\mu}(t) \rangle = 0, \qquad \langle \eta_{i,\mu}(t) \eta_{j,\nu}(t') \rangle = \delta_{ij}\delta_{\mu\nu}\delta(t-t')2k_{\mathrm{B}}T\gamma_i. \tag{7}$$

Within this dynamical framework it is possible to calculate exactly how correlations among the displacements of various pairs of residues decay as a function of time. We start by considering the case where the various viscous coefficients in equation (6) take on the same value, γ. In this case one has [31]

$$\langle x_{i,\mu}(t) x_{j,\nu}(t+\Delta t) \rangle = \frac{k_{\mathrm{B}}T}{k} \sum_l v_i^l v_j^l \frac{1}{\lambda_l} \mathrm{e}^{-\lambda_l \frac{k}{\gamma}\Delta t}, \tag{8}$$

where the average $\langle \cdot \rangle$ in equation (8) denotes the usual canonical thermodynamics average. The vector $\mathbf{v}^l$ and the scalar λ_l are, respectively, the lth eigenvector and the lth eigenvalue of the matrix $\mathcal{M}$ ordered in such a way that $\lambda_l < \lambda_{l+1}$. It should be noted that since the Hamiltonian is invariant under rotations and translations, it will always possess (at least) six eigenvectors associated with eigenvalues equal to zero that are obviously excluded from the sum in expression (8) and related ones [19, 32]. The eigenvectors $\{\mathbf{v}^l\}$ represent therefore the independent modes of structural relaxation in the protein, while the associated decay times are given by

$$\tau_l = \frac{\gamma}{k\lambda_l}. \tag{9}$$

Of particular interest, for the purpose of identifying the concerted large scale structural fluctuations occurring in a protein, is the degree of correlation of pairs of residues at equal times. This information is summarized in the covariance matrix, $\mathcal{C}$, whose elements are obtained by $\Delta t = 0$ in (8)

$$\mathcal{C}_{ij,\mu\nu} \equiv \langle x_{i,\mu} x_{j,\nu} \rangle = \frac{k_{\mathrm{B}}T}{k} \sum_l \frac{v_i^l v_j^l}{\lambda_l} = \frac{k_{\mathrm{B}}T}{k} \mathcal{M}^{-1}_{i,j,\mu,\nu}. \tag{10}$$

Due to the fact that this matrix contains the full three-dimensional information about pair correlations its linear size is $3N$. Typically it is important only to quantify the relative degree of correlation of any two residues; in this case one can consider the normalized reduced covariance matrix (of linear size N) which is defined as

$$\mathcal{C}_{ij} = \frac{\langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle}{\sqrt{\langle |\mathbf{x}_i|^2 \rangle \langle |\mathbf{x}_j|^2 \rangle}} = \frac{\mathcal{C}_{ij,\nu\nu}}{\sqrt{\mathcal{C}_{ii,\nu\nu}\mathcal{C}_{jj,\mu\mu}}} \tag{11}$$

where summation over repeated indices is implied. It is important to stress that the elements of the covariance matrix are thermodynamic averages reflecting equilibrium properties of the system and, hence, are independent of the friction coefficients. Through equation (10) one sees that the same set of vectors $\{\mathbf{v}^l\}$ as describes the modes of relaxation also describes the independent modes of structural fluctuations in thermal equilibrium. The contribution of each mode to the pair correlation is inversely proportional to λ_l, thus establishing the intuitive result that the modes associated with the longest relaxation times are those responsible for the largest

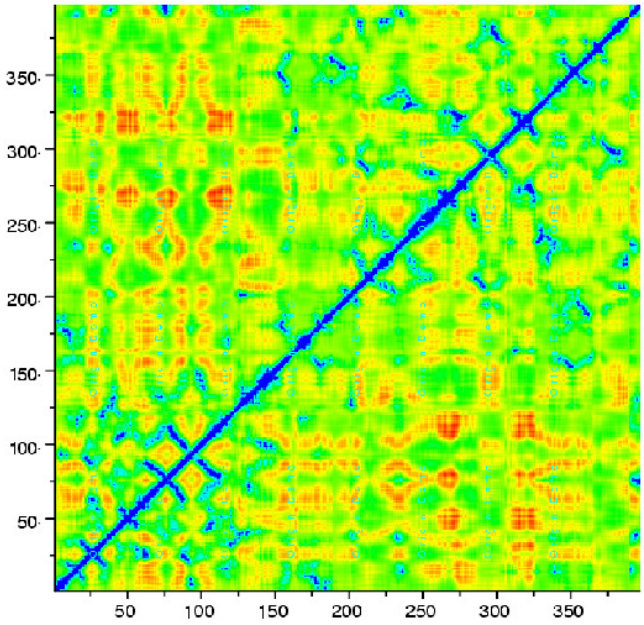

Figure 2. Reduced covariance matrices for BACE, as obtained from β-Gaussian model computations. Entries with values close to 1 (strong positive correlation) are shown in black (blue online), while the grey patches (red online) indicate anti-correlated regions.

structural fluctuations. The equivalence of the vectors describing the modes of relaxation and structural fluctuations is a consequence of having used the same viscous coefficient for all amino acids in the protein. When this simplifying assumption is not made one has that the modes and times of relaxation are determined by diagonalizing instead the symmetric matrix

$$\tilde{\mathcal{M}}_{i,j,\mu,\nu} = \frac{\mathcal{M}_{i,j,\mu,\nu}}{\sqrt{\gamma_i \gamma_j}}. \tag{12}$$

On the other hand, the modes of structural fluctuation are the same as before, as can be ascertained by calculating directly the thermodynamic average $\langle x_{i,\mu} x_{j,\nu} \rangle$ with the canonical weight associated with Hamiltonian (5).

3. Conformational fluctuations of the BACE

We have applied the above-mentioned analysis to the native structure of the BACE shown in figure 1(a). The associated normalized reduced covariance, computed through the β-GM, is shown in figure 2. The comparison of the covariance matrix with the contact map of figure 1(b) shows that the highest positive correlations are observed, as expected, in the correspondence of contacting residues. More interesting is the presence of negative correlations which signal important mechanical couplings between regions that are not in spatial proximity [23, 24, 19]. Inspection of figure 2 reveals a significant degree of anti-correlation among residues on one lobe (especially 40–55 and 100–114) and those on the other one (especially 260–272 and 305–320). The fact that the conformational fluctuations of the two lobes are, on average, directed in opposite directions suggests that the enzymatic functional modes may be based on an opening/closing mechanism of the lobes.

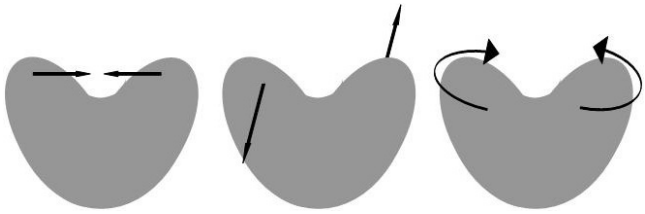

Figure 3. Pictorial representations of the first three modes (from left to right) which describe the largest conformational distortions in the BACE.

This insight can be considerably refined by examining the individual contribution of the various eigenvectors of $\mathcal{M}$ to the pair correlations. It is apparent from the decomposition of equation (10) that the weight of the eigenvectors is inversely proportional to the corresponding eigenvalues. For this reason we shall now describe the structural deformations encoded by the first three eigenvectors which, alone, are responsible for a good fraction of the overall residue mobility. By superimposing on the native structure the distortion associated with the first mode one can ascertain that it involves the movement of the lobes along opposite directions with respect to the plane identified by the cleft of the enzyme (each lobe moving almost rigidly while the connecting regions between the lobes are almost static). In the second mode the anti-correlated motion of the lobes is still visible but this time the motion occurs mostly parallel to the cleft plane, resulting in a shear deformation, as depicted in figure 3, while in the third mode, the two lobes appear to rotate in opposite directions. In all three modes the high mobility of the exposed loop ranging from Val309 to Asp317 is very noticeable.

So far we have summarized the results obtained with the β-GM which relies on several simplifying assumptions (including the use of the same friction coefficient for all residues, a limitation that will be removed later). It is therefore important to verify the validity of the conclusions reached here against independent terms of reference. Ideally, one would like to compare the model results against direct experimental determination of the quantities of interest here, such as the correlation of residue motion. However, this is not currently feasible in a direct way and hence such detailed information can only be obtained from dynamical simulations with all-atom interaction potentials. We therefore carried out a MD simulation of the BACE in explicit solvent. The simulated system was constituted by the protein immersed in a water box of size 75 Å $\times$ 87 Å $\times$ 90 Å to which nine sodium counter-ions were added to ensure the overall charge neutrality. The whole system, composed of about 48 000 atoms, underwent 20 ns of MD simulations (after relaxation) carried out at 300 K with the GROMACS program [33, 34]. The Amber force field (Parm98 [35]) was used to describe the protein and the counter-ions, whilst the TIP3P model was used for water. Particle mesh Ewald routines were used to treat long range electrostatic interactions. A cut-off of 12 Å was used for the van der Waals interactions and the real part of the electrostatic interactions.

We carried out the comparison of the β-GM and molecular dynamics through a series of steps which include the comparison of the overall mobility of the various amino acids, of corresponding entries of the covariance matrices and of the largest eigenvectors of the covariance matrix. These comparisons are here exploited not only to check the consistency of the two approaches, but also as a means to assign the β-GM parameter k from the MD comparison.

The first quantity that we shall compare is the mean square fluctuation which summarizes the overall mobility of any given residue. Within the β-GM this quantity is given by the diagonal elements of $\mathcal{M}^{-1}$:

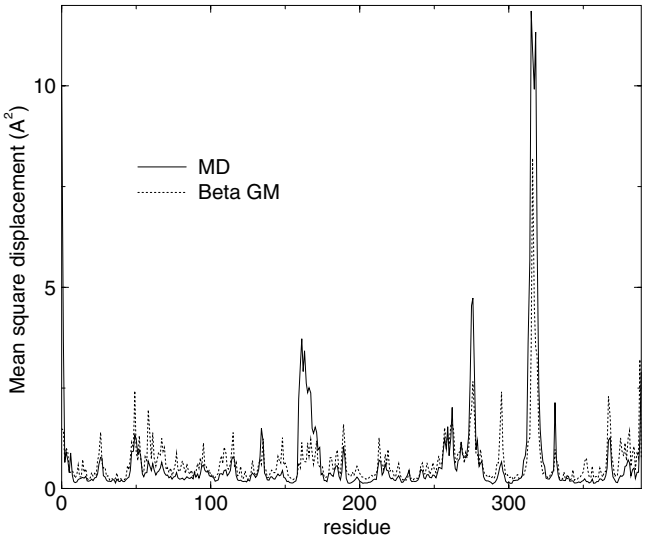

Figure 4. Mean square fluctuations of the various residues in the BACE obtained from MD and calculated from the β-GM after an optimal choice of the harmonic coupling parameter, k.

$$\langle \mathbf{x}_i \cdot \mathbf{x}_i \rangle = \sum_{\alpha} \mathcal{M}^{-1}_{ii,\alpha,\alpha}, \tag{13}$$

while, in the context of molecular dynamics, the thermodynamic average of equation (13) is aptly replaced by the time average over the simulation run. It should also be noted that, usually, it is not easy to ascertain whether the simulated trajectory is sufficiently long that thermodynamic averages can be legitimately replaced with dynamical ones. A practical way of checking *a posteriori* the validity of this ergodicity assumption is to check the consistency of the essential eigenspaces pertaining to different parts of the simulated trajectories. Indeed, this type of analysis shows that when the simulated time span greatly exceeds the typical timescale associated with the slow dynamical modes then the essential subspaces are robustly identified [36]. Besides this strategy, a recent theoretical study has introduced some valuable quantitative criteria by which it is possible to decide whether a given MD trajectory is too short to meet the ergodicity requirements [7]. These criteria are, again, based on the analysis of the essential dynamical spaces. When the dynamical sampling of phase space is insufficient, the eigenvectors of the essential spaces have distinctive cosine-like shapes. Conversely, it has been shown that when the cosine content of the relevant dynamical eigenspaces is negligible, the ergodicity assumption appears justified [7]. In the present study we have adopted both these criteria: the total simulation time span of 20 ns is about thirty times bigger than the longest autocorrelation time found in the system (see the last section); in addition the cosine content of the top five essential eigenvectors was, on average, below 10%.

It must be noted that, within the β-GM, the mean square displacements are proportional to $k_{\mathrm{B}}T/k$. The comparison of the MD and model mean square displacements offers, therefore, the opportunity to set the effective value of k by, e.g., matching the average mean square displacements in the two cases. This criterion yields, for this particular protein, the value $k^{-1} = 0.7\ \text{Å}^2/k_{\mathrm{B}}T$. The profiles for the residues mobility according to MD and the β-GM (having fixed k to the value mentioned above) are shown in figure 4.

As mentioned before, figure 4 reveals the high mobility of the exposed loop spanning residues 309–317. Aside from this, the two profiles have a good degree of correlation since

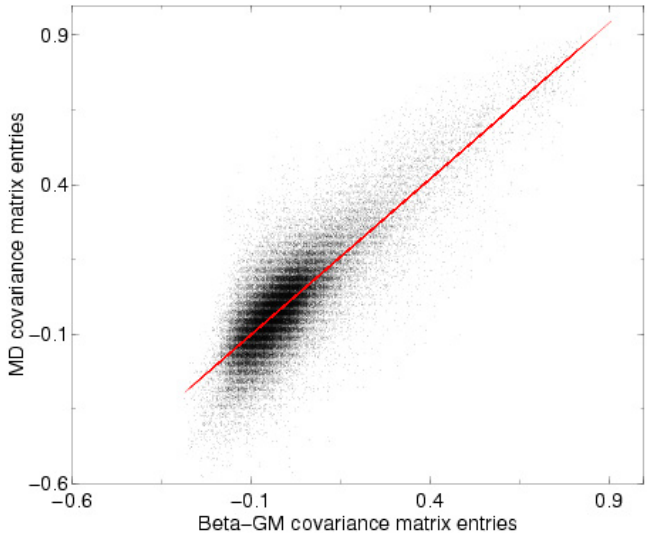

Figure 5. A scatter plot of corresponding entries of the reduced covariance matrices obtained by the β-Gaussian model and MD simulations of BACE. The interpolating line, determined by linear regression, is also shown.

the linear correlation coefficient is 0.78 while the more stringent Kendall (non-parametric) coefficient is 0.61 [37, 38].

Having ascertained the consistency of the overall residue mobility of MD and β-GM we analyse the agreement of the covariance matrices which convey information on pair correlations. We carry out the comparison in two stages: first by comparing corresponding entries of the matrices and subsequently by measuring the overlap between the significant subspaces of the matrices.

The degree of accord of the normalized reduced covariance matrices of the MD and β-GM is conveniently summarized through a scatter plot of corresponding entries of the two matrices. The results are summarized in figure 5 where we have deliberately omitted the diagonal entries of the matrix which are equal to 1 in both cases. Since the length of the BACE approaches 400 residues, the data set shown in figure 5 is constituted by more than 7.5×10^4 distinct data points. The linear correlation coefficient of the MD and β-GM data is $r = 0.83$. This represents a strikingly high value given the huge number of degrees of freedom. However, a precise quantification of its statistical significance cannot be conveyed since the data are manifestly not distributed in a binormal fashion and the Kendall non-parametric analysis is inapplicable to the excessive number of data points [19]. We conclude the discussion of the scatter plot of figure 5 by mentioning that, in the case of perfect correlation of two normalized (adimensional) covariance matrices, the data would align along the diagonal of the graph in figure 5. Interestingly, the best-fit line lies very close to the diagonal having a slope of $s = 1.04$, a fact which further testifies to the overall viability of the Gaussian scheme.

As a final test, we wish to measure the consistency of the most important eigenspaces of the two matrices, that is the eigenvectors associated with the largest eigenvalues of the covariance matrix (for β-GM this is equivalent to considering the *smallest* eigenvalues of $\mathcal{M}$). In fact, these eigenspaces are the ones that describe the most significant modes of distortion of the molecule in the three-dimensional space. Various strategies have been proposed for evaluating the accord of the significant eigenspaces of the covariance matrices [11, 39, 7]. Here we have

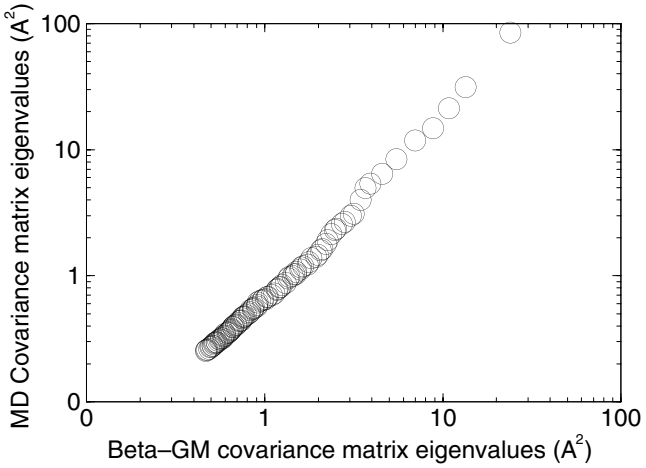

Figure 6. A scatter plot of equally ranking eigenvalues of the MD and β-GM covariance matrices.

adopted the simple strategy of considering the top 10 eigenvectors obtained from MD and the β-GM and calculating the modulus of the scalar product of any pair of vectors coming from the two sets. It is found that the overlap between the two essential eigenspaces is considerable since about 75% of the norm of the first four eigenvectors of the β-GM is projected onto the first four eigenvectors of MD. Overall the root mean square inner product, suggested by Amadei *et al* [36], among the two sets was 0.6 which denotes a non-trivial correlation, although a precise assessment of its statistical significance is difficult to ascertain [36, 7].

It is also interesting to evaluate the agreement of corresponding (i.e. equally ranking) eigenvalues of the covariance matrices. This analysis is carried out in the scatter plot of figure 6 where, for the β-GM, the ith largest ranking eigenvalue of the covariance matrix was taken as the inverse of the ith smallest eigenvalue of $\mathcal{M}^{-1}$ multiplied by the previously found coefficient $k_{\mathrm{B}}T/k$. Since the eigenvalues span a few orders of magnitude, the plot is presented in a log–log format and reveals a relationship that rather than being simply linear is better described by a power-law dependence with an exponent of about 1.4.

After having characterized the accord of the conformational fluctuations occurring in thermal equilibrium within the all-atom MD and the β-GM we focus on the characterization of the mechanical coupling between the BACE enzyme and the substrate (peptide) on which it acts. This analysis is accomplished by applying the Gaussian model not to the isolated enzyme but to the enzyme/substrate complex shown in figure 1. The crucial questions to formulate in this context relate to the possible existence of non-trivial mechanical couplings between the substrate and residues in the enzyme. For other enzymes, in particular the HIV-1 protease, it has been shown that several residues of HIV-1 protease despite being far away from the substrate have an important mechanical bearing on the latter [19, 23]. It is by virtue of this mechanical coupling that it is possible to rationalize the emergence of mutations causing drug resistance in correspondence with sites far from the cleavage region (e.g. Met63Ile, Met46Ile-Leu and Leu47Val) [40–42]. In fact, as the explicit MD calculation has shown, the high degree of such coupling between such sites and the cleavage is such that the detailed chemical identity of the former strongly influences the substrate binding affinity of the latter [24]. Therefore, mechanical couplings can be so important for enzymatic catalysis that mutations at a small number of key enzymatic sites (even if distant from the active site) can dramatically alter the enzyme reactivity. We have therefore undertaken a similar analysis here and calculated

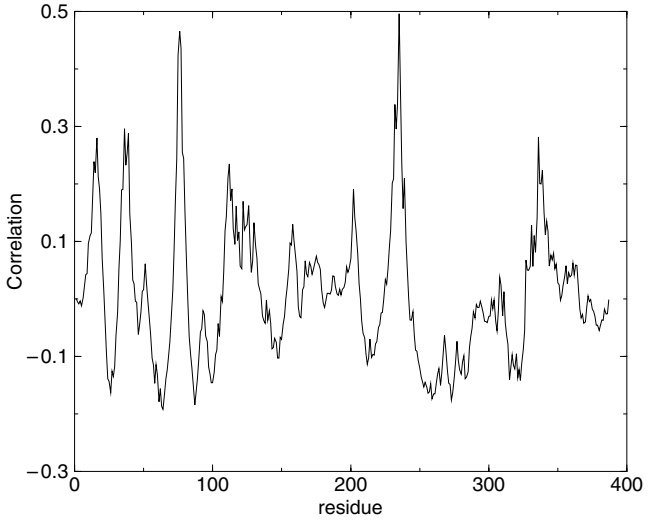

Figure 7. The curve indicates the degree of correlation, calculated with β-GM, of a residue in the middle of the inhibiting peptide and the 387 residues that constitute the BACE.

the degree of correlation of the displacement of residues in the substrate and all residues in the BACE. The correlation profiles are nearly identical for all residues in the substrate and a representative profile is shown in figure 7.

From the profile a strong positive correlations is discernible in correspondence with residues 72 and 230 (numbering according to the PDB file). This reflects the close spatial proximity of the substrate and these contacting residues. The crucial feature is however contained in the peaks of negative correlations which occur in correspondence with residues 23, 61, 83, 255, 269, 317. These residues are very distant from the substrate (25–30 Å) and yet are able to influence the conformational fluctuations of the substrate bound to the active site. These sites are located at the outward face of each of the two lobes, in close analogy to the key mechanical sites found for HIV-1 protease. This fact provides additional evidence in support of the steering of the substrate towards a reactive conformation being achieved not only through finely tuned chemical couplings with the enzymes but also through subtle mechanical influence of sites far apart from the active region.

We conclude our analysis by discussing the possibility of gaining insight through the β-GM into the timescales involved in the most important conformational fluctuations. To do so it is necessary to have a quantitative estimate for the effective friction coefficients appearing in equation (6). It has been pointed out before [6, 11] that the effective friction coefficients experienced by the various amino acids in a protein depend not only on the interaction with the solvent but are severely affected by the protein density itself, in their neighbourhood. In particular, Hinsen *et al* [28] have recently provided a phenomenological characterization of the viscous coefficient which they found to have an approximately linear relationship in terms of the local atomic density. We have translated this phenomenological relationship to one in terms of the number of contacts, n_i^c, in which residue i takes part (calculated with a cut-off distance of 7.5 Å) and thus arrived at the phenomenological relationship

$$\gamma_i = Bn_i^c \tag{14}$$

where $B = 1.53 \times 10^3$ amu ps^{-1}. We have therefore calculated the various modes of relaxation in this new context where the friction coefficient is not the same for all residues. Strikingly we

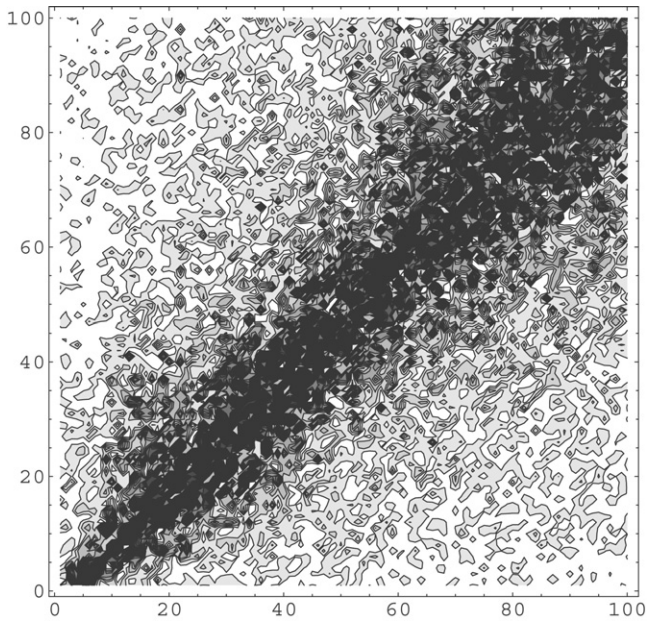

Figure 8. Density plot for the modulus of the scalar product of the eigenvectors of matrix $\mathcal{M}$ (rank index given on the x axis) against those of matrix $\tilde{\mathcal{M}}$ (rank index given on the y axis). Entries with values close to 1 [0] are shown in black [white].

have found that there is a good correspondence between the modes of relaxation (which depend on γ_i and reflect dynamical properties) and the eigenvectors which describe conformational fluctuations in thermal equilibrium. This correspondence is obviously exact in the case when γ_i is the same for all sites but is not otherwise expected in the case of heterogeneity of the friction coefficient. The degree of accord of the two sets of eigenvectors is shown in figure 8 and indeed the thick cloud around the diagonal highlights their good correspondence. The timescales which control the decay of conformational fluctuations in a dynamical trajectory are given by the inverse of the eigenvalues of $\tilde{\mathcal{M}}$ divided by k. Therefore, we obtain that the slowest relaxation time calculated within the Gaussian model has a decay time of 1.2 ns. The analysis of the decay of autocorrelation in the whole MD trajectory indicates 0.7 ns as the slowest relaxation time of the system. It appears, therefore, that the simple Gaussian approach can correctly identify the timescales of the system autocorrelation within a factor of two. Therefore, despite the several approximations which are at the basis of the β-GM, the latter appears to be useful also for estimating, with a modest computational expenditure, the correct order of magnitude of the system autocorrelation time. This may be exploited to obtain a preliminary indication of the lapse of time that needs to be covered in an all-atom MD simulation to ensure that the system has sufficient time to explore significantly different regions of the phase space.

4. Conclusions

We have studied the conformational fluctuations of BACE with the purpose of clarifying the effect of conformational fluctuations on the capability of enzymes to steer the substrate in a favourable reactive conformation. In analogy with results obtained on a different class of enzymes [23, 24, 19], for this enzyme also it is found that a few sites that are located remotely

from the active site have a strong mechanical influence on the enzyme–substrate couplings. The analysis was carried out using a simple and computationally inexpensive Gaussian network scheme. Several aspects of the results obtained within the simplified coarse-grained model have been validated and tested against an extensive all-atom MD simulation.

Acknowledgments

We are indebted to Paolo Carloni and Luca Marsella for several illuminating discussions and suggestions. We acknowledge support from INFM-Democritos and MIUR Cofin 2003.

References

[1] Levitt M, Sander C and Stern P S 1985 *J. Mol. Biol.* **181** 423–47
[2] Horiuchi T and Go N 1991 *Proteins: Struct. Funct. Genet.* **10** 106–16
[3] ben Avraham D 1993 *Phys. Rev.* B **47** 14559–60
[4] Micheletti C, Lattanzi G and Maritan A 2002 *J. Mol. Biol.* **321** 909–21
[5] Garcia-Viloca M, Gao J, Karplus M and Truhlar D G 2004 *Science* **303** 186–95
[6] Karplus M 2002 *Acc. Chem. Res.* **35** 321–3
[7] Hess B 2002 *Phys. Rev.* E **65** 031910
[8] Tirion M M 1996 *Phys. Rev. Lett.* **77** 1905–8
[9] Noguti T and Go N 1982 *Nature* **296** 776–8
[10] Brooks B and Karplus M 1985 *Proc. Natl Acad. Sci. USA* **82** 4995–9
[11] Hinsen K 1998 *Proteins: Struct. Funct. Genet.* **33** 417–29
[12] McCammon J A, Gelin B R, Karplus M and Wolynes P G 1976 *Nature* **262** 325–6
[13] Swaminathan S, Ichiye T, van Gusteren W and Karplus M 1982 *Biochemistry* **21** 5230–41
[14] Bahar I, Atilgan A R and Erman B 1997 *Folding Des.* **2** 173–81
[15] Bahar I, Atilgan A R, Demirel M C and Erman B 1998 *Phys. Rev. Lett.* **80** 2733–6
[16] Jernigan R L, Demirel M C and Bahar I 1999 *Int. J. Quantum Chem.* **75** 301–12
[17] Doruker P, Atilgan A and Bahar I 2000 *Proteins: Struct. Funct. Genet.* **40** 512–24
[18] Atilgan A R, Durell S R, Jernigan R L, Demirel M C, Keskin O and Bahar I 2001 *Biophys. J.* **80** 505–15
[19] Micheletti C, Carloni P and Maritan A 2004 *Proteins* **55** 635–45
[20] Sinha S and Lieberburg I 1999 *Proc. Natl Acad. Sci. USA* **96** 11049–52
[21] Lin X, Koelsch G, Wu S, Downs D, Dashti A and Tang J 2000 *Proc. Natl Acad. Sci. USA* **97** 1456–60
[22] Hong L, Koelsch G, Lin X, Wu S, Terzyan S, Ghosh A K, Zhang X C and Tang J 2000 *Science* **290** 150–3
[23] Piana S, Carloni P and Parrinello M 2002 *J. Mol. Biol.* **319** 567–83
[24] Piana S, Carloni P and Rothlisberger U 2002 *Prot. Sci.* **11** 2393–402
[25] De Mori G M S, Micheletti C and Colombo G 2004 *J. Phys. Chem.* B **108** 12267–70
[26] De Mori G M S, Colombo G and Micheletti C 2004 *Proteins* at press
[27] Park B and Levitt M 1996 *Proteins: Struct. Funct. Genet.* **258** 367–92
[28] Hinsen K, Petrescu A, Dellerue S, Bellissent-Funel M and Kneller G R 2000 *Chem. Phys.* **261** 25–37
[29] Doi M 1996 *Introduction To Polymer Physics* 1st edn (Oxford: Clarendon)
[30] Howard J 2001 *Mechanics of Motor Proteins and the Cytoskeleton* (Sunderland, MA: Sinauer Associates)
[31] Chandrasekhar S 1943 *Rev. Mod. Phys.* **15** 1–89
[32] Micheletti C, Cecconi F, Flammini A and Maritan A 2002 *Prot. Sci.* **11** 1878–87
[33] Berendsen H J C, van der Spoel D and Vandrunen R 1995 *Comput. Phys. Commun.* **91** 43–56
[34] Lindahl E, Hess B and van der Spoel D 2001 *J. Mol. Mod.* **7** 306–17
[35] Cornell W D, Cieplak P, Bayly C I, Gould I R, Caldwell J W and Kollman P A 1995 *J. Am. Chem. Soc.* **117** 5179–97
[36] Amadei A, Ceruso M A and Nola A D 1999 *Proteins: Struct. Funct. Genet.* **36** 419–24
[37] Halle B 2002 *Proc. Natl Acad. Sci. USA* **99** 1274–9
[38] Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1999 *Numerical Recipes* (Cambridge: Cambridge University Press)
[39] Amadei A, Linssen A B M and Berendsen H J C 1993 *Proteins* **17** 412–25
[40] Brown A J L, Korber B T and Condra J H 1999 *AIDS Res. Hum. Retroviruses* **15** 247–53
[41] Boucher C 1996 *AIDS* **10** S15–9
[42] Molla A *et al* 1996 *Nat. Med.* **2** 760–6