Protein Engineering vol.14 no.1 pp.1–6, 2001

Conformational change of proteins arising from normal mode calculations

F. Tama and Y.-H. Sanejouand[1]

Laboratoire de Physique Quantique, UMR 5626 of CNRS, IRSAMC, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France

[1]To whom correspondence should be addressed. Present address: Centre de Recherche Paul Pascal, UPR 8641 of CNRS, Avenue Albert Schweitzer, 33600 Pessac, France. E-mail: sanejouand@crpp.u-bordeaux.fr

A normal mode analysis of 20 proteins in 'open' or 'closed' forms was performed using simple potential and protein models. The quality of the results was found to depend upon the form of the protein studied, normal modes obtained with the open form of a given protein comparing better with the conformational change than those obtained with the closed form. Moreover, when the motion of the protein is a highly collective one, then, in all cases considered, there is a single low-frequency normal mode whose direction compares well with the conformational change. When it is not, in most cases there is still a single low-frequency normal mode giving a good description of the pattern of the atomic displacements, as they are observed experimentally during the conformational change. Hence a lot of information on the nature of the conformational change of a protein is often found in a single low-frequency normal mode of its open form. Since this information can be obtained through the normal mode analysis of a model as simple as that used in the present study, it is likely that the property captured by such an analysis is for the most part a property of the shape of the protein itself. One of the points that has to be clarified now is whether or not amino acid sequences have been selected in order to allow proteins to follow a single normal mode direction, as least at the very beginning of their conformational change.

Keywords: collective motion/degree of collectivity/hinge bending/protein model/shear motion

Introduction

In many proteins, large conformational transitions involve the relative movement of almost rigid structural elements. These domain motions are important for a variety of protein functions, including catalysis and regulation of activity. For example, in citrate synthase, a two-domain protein, coenzyme A binding induces an 18° rotation of the small domain around an axis close to residue 274, which represents the hinge (Remington *et al.*, 1982; Hubert and Bennett, 1983; Wiegand and Remington, 1986). One consequence of this motion is the closure of the cleft between the two domains in which the substrate binding site lies. As in most other cases, e.g. hexokinase (Bennet and Steitz, 1980) and phage T4 lysozyme (Faber and Matthews, 1990), such a hinge bending motion has been found by X-ray crystallography.

One of the best suited theoretical methods for studying collective motions in proteins is normal mode analysis (NMA), which leads to the expression of the dynamics in terms of a superposition of collective variables, namely the normal mode coordinates. These coordinates have been used to analyze molecular dynamics trajectories through the quasi-harmonic approximation (Karplus and Kuschick, 1981; Levy *et al.*, 1984; Teeter and Case, 1990; Hayward and Go, 1995), to integrate the equations of atomic motion with large time steps (Elezgaray and Sanejouand, 1998, 2000) and to sample larger portions of the configurational space (Amadei *et al.*, 1993). However, the idea that normal mode theory may be an accurate tool for studying protein conformational changes comes from the fact that in several cases, e.g. hexokinase (Harrison, 1984), lysozyme (Brooks and Karplus, 1985; Gibrat and Go, 1990) and citrate synthase (Marques and Sanejouand, 1995), the largest amplitude motion obtained with this theory, that is, that with the lowest frequency, was found to compare well with the conformational change observed by crystallographers in these proteins upon ligand binding. In the case of hemoglobin, it is the second lowest frequency mode of the T form which was found to be very similar to the observed difference between the T and R forms (Perahia and Mouawad, 1995).

The reasons underlying these successes of NMA are not fully understood. First, proteins are known to fold and function in a water environment, within a narrow range of pH, temperature, ionic strength, etc., while standard NMA is performed *in vacuo*. Second, standard NMA requires a preliminary energy minimization which drifts the atoms of the protein up to a few ångströms away from their position in the crystallographic structure. As a consequence, the structure studied with standard NMA is always a distorted one. Finally, NMA is based on a severe small displacements approximation, which amounts to supposing that a protein behaves like a solid does at low temperature, whereas it is well known that a protein is a somewhat flexible polymer, undergoing many local conformational transitions at room temperature.

Recent results have shed some light on this paradox. It was shown that using a single parameter hookean potential for taking into account pairwise interactions between neighboring atoms yields results in good agreement with those obtained when NMA is performed with standard semi-empirical potentials, as far as low-frequency normal modes are concerned (Tirion, 1996; Hinsen, 1998). The use of the same kind of highly simplified potential, but including only one point mass per residue in the model, yields low-frequency normal modes also in good agreement with those obtained using standard NMA (Hinsen, 1998). Moreover, when the interactions between closely located α-carbon pairs are described by an elastic network model, protein crystallographic temperature B factors are found to be accurately predicted (Bahar *et al.*, 1997). Again, this means that low-frequency normal modes are well described with such a model, since it is well known that modes with frequencies under 30 cm^{-1} are responsible for most of the amplitude of atomic displacements, as they can be estimated from the knowledge of B factors (Levy *et al.*, 1982; Swaminathan *et al.*, 1982).

© Oxford University Press

Table I. Proteins studied, with the type of motion they perform and the pdb codes of the corresponding pair of 'open' and 'closed' crystallographic structures

Protein	No. of residues	Motion type[a]	Pdb codes
Adenylate kinase	218	H	1aky, 2ak3
Alcohol dehydrogenase	373	S	8adh, 6adh
Annexin V	317	H	1avr, 1avh
Aspartate aminotransferase	401	S	9aat, 1ama
Calmodulin	144	H	1cll, 1ctr
Che Y protein	128	U	3chy, 1chn
Citrate synthase	858	S	5csc, 6csc
Dihydrofolate reductase	159	S	4dfr, 5dfr
Diphtheria toxin	523	H	1ddt, 1mdt
Enolase	436	H	3enl, 7enl
HIV-1 protease	99	H	1hhp, 1ajx
Immunoglobulin	418	H	1hil, 1him
Lactoferrin	691	H	1lfh, 1lfg
LAO binding protein	238	H	2lao, 1lst
Maltodextrin binding protein	370	H	1omp, 1anf
Seryl-tRNA synthetase	421	H	1ses, 1set
Thymidylate synthase	264	S	3tms, 2tsc
Triglyceride lipase	265	H	3tgl, 4tgl
Triose phosphate isomerase	250	H	3tim, 6tim
Tyrosine phosphatase	278	H	1ypt, 1lyts

[a]H, S and U stand for hinge bending, shear motion and unclassified, respectively. This classification of motion types comes from the Molecular Movements Database (Gerstein and Krebs, 1998).

Hence, results obtained with NMA in the field of low-frequency protein dynamics seem to be of very good quality even when most atomic details are simply ignored. However, up to now, low-frequency normal modes obtained with simple models have mainly been compared with those obtained with standard NMA as a whole, and not on a one-to-one basis. Detailed comparisons of such modes with large amplitude conformational changes of proteins remain to be performed. To do so, a study of proteins of various sizes undergoing conformational changes of various types and amplitudes was undertaken (see Table I). The results obtained confirm that a lot of information on the nature of the conformational change can be found in a single low-frequency normal mode of the open form of a protein. Moreover, it is shown that simple protein models are adequate for performing this kind of analysis.

Methods

The studied protein set

For this analysis, 20 proteins were chosen for which at least two significantly different X-ray conformations are known. They were picked within the 'Database of Macromolecular Movements' where conformational changes are classified into five main types, referenced as predominantly shear, predominantly hinge, not hinge or shear, involving partial refolding of the structure, or unclassified (Gerstein and Krebs, 1998). Except for the Che Y case, all proteins considered in the present study undergo conformational changes of one of the two first kinds (see Table I for the pdb codes of the corresponding pairs of conformers), mainly because these are more frequent and/or better characterized.

Normal mode calculations

Normal mode theory. The small displacements of atomic coordinate i, $r_i(t)$ in the vicinity of a stationary point of the potential energy surface are given by Goldstein (1950):

$$r_i(t) = \frac{1}{\sqrt{m_i}} \sum_j^{3N} C_j a_{ij} \cos(\omega_j t + \phi_j)$$

where m_i is the mass of the corresponding atom, a_{ij} the ith coordinate of the normal mode j and $\nu_j = \omega_j/2\pi$ the corresponding frequency. C_j, a_{ij}, ω_j^2 and ϕ_j are obtained as follows: ω_j^2 is the jth eigenvalue of the $3N\times3N$ mass-weighted second derivatives of the potential energy matrix, and a_{ij} is the ith coordinate of eigenvector j. C_j and ϕ_j, the amplitude and phase of mode j, respectively, are determined once the coordinates and the velocities of the system at $t = 0$ are known.

Simplified potential. Within the frame of the approach proposed by Tirion (1996), the standard detailed potential energy function is replaced by

$$E_p = \sum_{d_{ij}^0 < R_c} C(d_{ij} - d_{ij}^0)^2 \quad (1)$$

where d_{ij} is the distance between atoms i and j, d_{ij}^0 being the distance between these two atoms in the given studied crystallographic structure. C, the strength of the potential, is a phenomenological constant assumed to be the same for all interacting pairs. Note that this energy function was designed in such a way that for any chosen configuration of any system the total potential energy, E_p, is a minimum of the function. Thus, with such an approach, by definition, NMA does not require any prior energy minimization.

Note that in Equation 1, the sum is restricted to atom pairs separated by less than R_c, which is an arbitrary cut-off parameter. In all normal mode calculations hereafter, a cutoff of 8 Å has been used and, as proposed by Bahar *et al.* (1997), only the C_α atoms have been taken into account. Such a model is adequate for studying backbone motions, which in turn is sufficient for characterizing low-frequency normal modes of large proteins. Moreover, it allows the study of proteins of large size on common workstations, using small amounts of CPU time, since with this simple model the matrix to be diagonalized is a $3N\times3N$ matrix, where N is the number of residues of the protein.

Comparison with experiment

Overlap. I_j, the overlap between $\overrightarrow{\Delta r} = \{\Delta r_1, \cdots, \Delta r_i, \cdots, \Delta r_{3N}\}$, the conformational change observed by crystallographers, and the jth normal mode of the protein is a measure of the similarity between the direction of the conformational change and the one given by mode j. It is obtained as follows (Marques and Sanejouand, 1995):

$$I_j = \frac{\left|\sum^{3N} a_{ij}\Delta r_i\right|}{\left[\sum^{3N} a_{ij}^2 \sum^{3N} \Delta r_i^2\right]^{1/2}} \quad (2)$$

where $\Delta r_i = r_i^o - r_i^c$, r_i^o and r_i^c being the ith atomic coordinate of the protein in the 'open' and 'closed' crystallographic structure, respectively. A value of one for the overlap means that the direction given by normal mode j is identical with $\overrightarrow{\Delta r}$. From a practical point of view, $\overrightarrow{\Delta r}$ is calculated after both crystallographic conformations of the protein were superimposed, using standard fitting procedures. These pairs of conformations are referred to as 'open' or 'closed' because many conformational changes considered in the present study occur in enzymes, in which an active pocket site is being closed as a consequence of substrate binding.

Correlation. The correlation coefficient c_j measures the similarity of the patterns of the atomic displacements in the conformational change and in mode j. It is obtained as follows:

$$c_j = \frac{1}{N} \frac{\sum^{N} (A_{ij}-\bar{A}_j)(\Delta R_i-\overline{\Delta R})}{\sigma(A_j)\sigma(\Delta R)} \quad (3)$$

where A_{ij} and ΔR_i are, respectively, the amplitudes of the displacement of atom i in the mode j and in the conformational change, $\bar{A}_j$ and $\overline{\Delta R}$ being the corresponding average displacements, while $\sigma(A_j)$ and $\sigma(\Delta R)$ are the corresponding root mean square values. A value of one for c_j means that both patterns of atomic displacements are identical.

The degree of collectivity of a protein motion. A measure of how collective a protein motion is was proposed by Bruschweiler (1995). In the present study, it was used in order to estimate the degree of collectivity of each conformational change considered, reflecting the number of atoms which are significantly affected during the conformational change. This degree of collectivity, κ, is defined as being proportional to the exponential of the 'information entropy' embedded in vector $\overrightarrow{\Delta R} = \{\Delta R_1, \cdots, \Delta R_i, \cdots, \Delta R_N\}$:

$$\kappa = \frac{1}{N} \exp\left(-\sum^{N} \alpha\Delta R_i^2 \log\alpha\Delta R_i^2\right) \quad (4)$$

where the sum is over the C_α atoms of the protein and where α is a normalization factor chosen so that $\sum^{N} \alpha\Delta R_i^2 = 1$. κ is analogous to Boltzmann's W in $S = k\log W$ and gives the effective number of non-zero ΔR_i^2. It is confined to the interval between N^{-1} and 1. If $\kappa = 1$, the conformational change is maximally collective and all ΔR_i^2 are identical. In the limit of extreme local motion, where the conformational change involves only a few atoms, κ is minimal ($\kappa = 1/N$ when only one atom is involved in the conformational change). κ values have been calculated for the conformational changes of all proteins considered in the present study (see Tables IV and V).

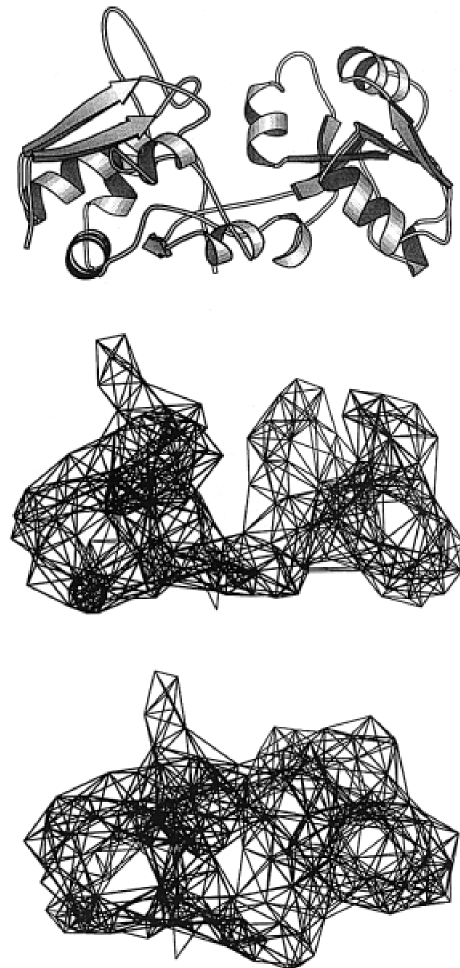

Figure 1. The simple protein model. The open (top and middle) and closed (bottom) forms of the lysine–arginine–ornithine (LAO) binding protein, as usually shown (top), and as modeled with the potential energy function considered in the present study, that is, when pairs of C_α atoms closer than 8 Å are supposed to be linked together by harmonic springs (middle and bottom). This figure was drawn using the Molscript program (Kraulis, 1991).

Results and discussion

Comparison with standard NMA

In the seminal work of Tirion in which simplified potentials such as the one that we use were introduced, it was mentioned that, using an all-atom model, tests performed on a periplasmic maltodextrin binding protein indicate that the slowest modes do indeed closely map the open form into the closed form (Tirion, 1996), but the corresponding details were not given. In further work by Hinsen (1998), a simplified potential was used together with a simplified protein model, with one point mass per residue as in ours, and, in the case of crambin, lysozyme and aspartate transcarbamylase, normal modes thus obtained were found to compare well with those obtained using standard detailed potentials and models (Hinsen, 1998). Next, using the same simple potential and model, it was shown that in the case of citrate synthase and aspartate transcarbamylase, low-frequency normal modes lead to essentially the same domain identification as when the corresponding pair of crystallographic conformations are compared (Hinsen *et al*., 1999).

Although all these results strongly suggest that simple potentials and models yield low-frequency normal modes as accurate as those obtained using standard detailed potentials and models, no direct comparison between such approximate low-frequency normal modes and an experimentally known conformational change has yet been published. In Table II, the overlap (see Equation 2) of the mode found to be the most involved in the conformational change, that is, the one with the largest overlap value (Ma and Karplus, 1997), is given in

Table II. Overlap of the mode the most involved in the conformational change of proteins of various sizes, when standard detailed (DP) or simplified (SP) potentials are used (in all five cases the open form of the protein was considered)

Protein	Overlap	
	DP	SP
Adenylate kinase	0.53	0.62
Enolase	0.31	0.33
Triglyceride lipase	0.26	0.30
Citrate synthase	0.85	0.83
Tyrosine phosphatase	0.22	0.55

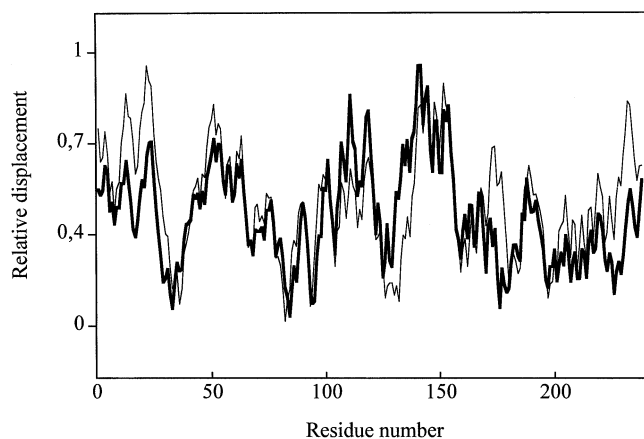

Figure 2. Comparison of the conformational change of the LAO binding protein with the normal mode the most involved. C_α displacements as a function of the residue number, when the open (2lao) and closed (1lst) crystallographic structures (Oh *et al.*, 1993) are compared (thick line), or when the atoms are displaced along the corresponding normal mode direction (thin line). For the sake of comparison, both kinds of atomic displacements have been normalized.

Table III. Overlap of the mode the most involved in the conformational change of proteins of various sizes and motion types, when the 'open' or 'closed' conformations are considered

Protein	Overlap	
	Open	Closed
Citrate synthase	0.83	0.57
Calmodulin	0.50	0.37
Che Y protein	0.32	0.34
Dihydrofolate reductase	0.72	0.64
Diphtheria toxin	0.58	0.37
Enolase	0.33	0.30
LAO binding protein	0.84	0.40
Triglyceride lipase	0.30	0.17
Maltodextrin binding protein	0.86	0.77
Thymidylate synthase	0.56	0.40

the case of five proteins, when the modes are calculated with the model described in the Methods section (see Equation 1 and Figure 1), or when they are calculated using standard detailed potentials and models, as done within the frame of a previous methodological study (Tama *et al.*, 2000).

The overlaps obtained with the simplified model are found to be almost equivalent to those obtained with standard approaches. In fact, in the former case, the overlap values are even slightly larger in four among the five cases considered. The most significant increase is observed in tyrosine phosphatase, for which no mode with a significant overlap was found using standard approaches (0.22 is a value that can be obtained when many random vectors of this size are compared; data not shown). This result may reflect one of the main advantages of the method proposed by Tirion, namely that no energy minimization has to be performed prior to the normal mode calculation. Thus, here, the structure studied is the crystallographic structure itself, whereas with standard approaches it may lie at a C_α r.m.s. distance of up to 2–3 Å; for the proteins in Table II, the C_α r.m.s. distance lies within a 1.2–1.9 Å range (Tama *et al.*, 2000).

The above results are in line with the hypothesis that, as far as the calculation of dynamic properties of proteins with NMA is concerned, simple potentials and protein models perform as well as detailed semi-empirical models (Tirion, 1996; Bahar *et al.*, 1997; Hinsen, 1998). Hereafter, advantage is taken of the ease of use of simple models in order to address other issues in a quantitative manner, namely through the study of a significant number of proteins.

NMA performs better with open forms

First, a potential such as the one we use (see Equation 1) is a description of a protein as a set of harmonic springs linked together, as illustrated in Figure 1 in the case of the lysine–arginine–ornithine (LAO) binding protein. The fact that NMA performs well using such a description (see Figure 2 and Tables II and III) suggests that the property captured by NMA may for the most part be a property of the shape of the protein itself. If this point is correct, NMA should perform better with 'open' than with 'closed' forms, since in the former the domains of the protein are, by definition, more separated, that is, better defined as far as their shape is concerned (see Figure 1). Indeed, in Table III it is shown that the normal mode most involved in the conformational change, as obtained when studying an open conformation of a protein, almost always yields a better description of the direction of the observed conformational change than that obtained when studying a closed form, the corresponding overlap being significantly larger in eight cases out of 10. Using standard detailed potentials and models, such a result had already been noted in the case of citrate synthase, when the normal modes of this protein are calculated for the open (Tama *et al.*, 2000) and for the closed (Marques and Sanejouand, 1995) forms, and also in the case of the open (see Table II) and closed forms of adenylate kinase. In the latter case, the overlap of the mode most involved in the conformational change is 0.37. Here again, this value is found to be close to that obtained with the simplified model used in the present study, that is, 0.38 (F.Tama and Y.-H.Sanejouand, unpublished results).

For the two remaining cases, the Che Y protein and enolase, their conformational changes happen to be rather localized, that is, with a small collectivity value (see Equation 4) and, as shown hereafter, it is likely that in such cases NMA cannot perform well, at least as far as overlap values are concerned.

NMA performs better with highly collective motions

In Table IV, values of κ, the degree of collectivity of atomic motions (see Equation 4) are given for each conformational change studied, as well as the overlap of the mode the most involved found in the case of the open forms of the studied proteins. It appears that for a degree of collectivity larger than 0.18, there is always one normal mode with a large overlap value with the conformational change (larger than 0.5). This result makes sense since normal modes of low frequency are known to be highly collective motions. In the case of rather localized motions, that is, when κ is <0.18, the direction of the conformational change is rarely well described with a single normal mode. Indeed, in the four cases with $\kappa < 0.14$, no mode is found with an overlap with the conformational change >0.33. Such results suggest that only highly collective conformational changes may occur along a direction well described by a single normal mode.

NMA also performs well with more localized motions

However, it appears that important correlation values can be obtained (see Equation 3), even for motions with a low degree of collectivity, as in the case of triglyceride lipase (see Table IV). In Figure 3, the experimental conformational change of

Table IV. Overlap of the mode the most involved in the conformational change, when the open conformations are considered, as a function of the degree of collectivity (κ) of the conformational change

Protein	Collectivity	Overlap[a]
Calmodulin	0.68	0.50 (5)
LAO binding protein	0.68	0.84 (2)
Maltodextrin binding protein	0.67	0.86 (2)
Diphtheria toxin	0.48	0.58 (1)
Lactoferrin	0.48	0.60 (1)
Aspartate aminotransferase	0.38	0.70 (7)
Immunoglobulin	0.37	0.51 (1)
Alcohol dehydrogenase	0.30	0.69 (2)
Citrate synthase	0.30	0.83 (3)
Adenylate kinase	0.27	0.62 (3)
Dihydrofolate reductase	0.26	0.72 (1)
HIV-1 protease	0.23	0.66 (8)
Thymidylate synthase	0.18	0.56 (1)
Che Y protein	0.17	0.32 (2)
Triose phosphate isomerase	0.15	0.46 (6)
Tyrosine phosphatase	0.14	0.55 (7)
Annexin V	0.13	0.33 (1)
Enolase	0.10	0.33 (2)
Seryl-tRNA synthetase	0.10	0.32 (18)
Triglyceride lipase	0.07	0.30 (1)

[a]The corresponding mode number is given in parentheses.

Table V. Correlation between a normal mode motion and a conformational change, as a function of the degree of collectivity (κ) of the conformational change

Protein	Collectivity	Correlation[a]
Calmodulin	0.68	0.53 (10)
LAO binding protein	0.68	0.71 (2)
Maltodextrin binding protein	0.67	0.78 (2)
Diphtheria toxin	0.48	0.66 (3)
Lactoferrin	0.48	0.68 (6)
Aspartate aminotransferase	0.38	0.69 (7)
Immunoglobulin	0.37	0.50 (13)
Alcohol dehydrogenase	0.30	0.58 (2)
Citrate synthase	0.30	0.83 (3)
Adenylate kinase	0.27	0.67 (3)
Dihydrofolate reductase	0.26	0.68 (1)
HIV-1 protease	0.23	0.80 (8)
Thymidylate synthase	0.18	0.75 (1)
Che Y protein	0.17	0.59 (6)
Triose phosphate isomerase	0.15	0.79 (6)
Tyrosine phosphatase	0.14	0.83 (15)
Annexin V	0.13	0.50 (22)
Enolase	0.10	0.31 (7)
Seryl-tRNA synthetase	0.10	0.68 (6)
Triglyceride lipase	0.07	0.71 (12)

[a]The number of the mode considered, that is, the mode found to have the largest correlation with the conformational change, is given in parentheses.

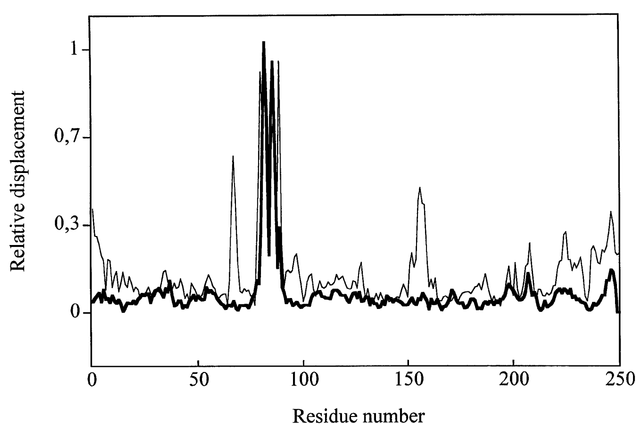

Figure 3. Comparison of the conformational change of triglyceride lipase with the normal mode with the highest correlation. C_α displacements as a function of the residue number, when the open (3tgl) and closed (4tgl) crystallographic structures (Derewenda *et al.*, 1992a,b) are compared (thick line), or when the atoms are displaced along the corresponding normal mode direction (thin line). For the sake of comparison, both kinds of atomic displacements have been normalized.

triglyceride lipase is shown. As expected from the small κ value of 0.07, this motion is localized in a small stretch of residues (~10). However, while no mode with a large overlap value was found either in our study (see Table IV) or in a study performed with standard NMA (Jaaskelainen *et al.*, 1998), the motion corresponding to the normal mode with the best correlation does indeed describe correctly the conformational change of triglyceride lipase, importance displacements being observed in the same stretch of residues. Similar results are obtained in the case of triose phosphate isomerase, tyrosine phosphatase and seryl-tRNA synthetase. Thus, in these cases also, there is some information on the conformational change of the protein lying in the low-frequency normal modes of the open form, but this information is about the amplitudes of the displacements of the atoms and not about the direction of their motion. This result can be understood since when atomic displacements in a small stretch of residues are as large as ~10 Å, as in the case of triglyceride lipase, the corresponding motion is not expected to be as linear as it is in a domain motion. So, in such cases, normal modes describing correctly where the atoms will be at the end of the conformational change should be rare, for the simple reason that this direction may vary all along the conformational change. Such an explanation may also help in understanding why in the case of calmodulin, whose conformational change has a very large degree of collectivity, namely, 0.68, the mode the most involved has a rather poor 0.5 overlap with the conformational change. Indeed, calmodulin is a special case, since it has a dumbbell-like shape in the open form, and a rather globular one in the closed form. Thus, atomic motions during this conformational change are likely to follow a curvilinear path, the direction of the motion varying as the conformational change proceeds.

A single normal mode can carry a lot of information

For 10 out of the 20 proteins studied, a lot of information on the nature of the conformational change is found in a single normal mode, namely its direction and the pattern of the atomic displacements in the protein, since the same normal mode presents both the best overlap and the best correlation with the conformational change. However, most often, this single normal mode is *not* the lowest frequency mode, as found, for instance, when the closed form of citrate synthase was studied using standard semi-empirical potentials (Marques and Sanejouand, 1995), but is often one of the three lowest frequency ones (see Tables IV and V).

Conclusion

Large conformational transitions are important for a variety of protein functions, including catalysis and regulation of activity. Most of these motions have been probed by X-ray crystallography. However, it is often difficult to obtain the crystallographic structures of the two forms of a protein, that is, both the free form and the form of the protein–ligand (or substrate) complex. As a consequence, conformational changes

are more often suspected than they are described at the residue level. Thus, theoretical tools able to give information on the kind of conformational change a protein can undergo would be welcome. What we have shown in the present study is that a lot of information on the nature of the conformational change is often carried by a single low-frequency normal mode of the open form of the protein, as is obtained when normal mode analysis is performed using very simple protein models. Thus, seeking such a normal mode could help in checking a hypothesis about the kind of conformational change a given protein undergoes in order to perform its function. At a more fundamental level, our results raise the possibility that protein sequences may have been designed, through evolutionary processes, so as to allow the protein to follow the direction of a single normal mode, at least at the very beginning of its conformational change. In other words, proteins may take advantage of one of the collective motions they are able to perform, because of the shape they happen to have, as a starting point for their large amplitude functional motions.

References

Amadei,A., Linssen,A. and Berendsen,H. (1993) *Proteins*, **17**, 412–425.
Bahar,I., Atilgan,A. and Erman,B. (1997) *Fold. Des.*, **2**, 173–181.
Bennet,W. and Steitz,T. (1980) *J. Mol. Biol.*, **140**, 210–230.
Brooks,B.R. and Karplus,M. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 4995–4999.
Bruschweiler,R. (1995) *J. Chem. Phys.*, **102**, 3396–3403.
Derewenda,U., Brzozowski,A., Lawson,D. and Derewenda,Z. (1992a) *Biochemistry*, **31**, 1532–1541.
Derewenda,Z., Derewenda,U. and Dodson,G. (1992b) *J. Mol. Biol.*, **227**, 818–839.
Elezgaray,J. and Sanejouand,Y.-H. (1998) *Biopolymers*, **46**, 493–501.
Elezgaray,J. and Sanejouand,Y.-H. (2000) *J. Comput. Chem.*, **21**, 1274–1282.
Faber,H. and Matthews,B. (1990) *Nature*, **348**, 263–266.
Gerstein,M. and Krebs,W. (1998) *Nucleic Acids Res.*, **26**, 4280–4290.
Gibrat,J. and Go,N. (1990) *Proteins*, **8**, 258–279.
Goldstein,H. (1950) *Classical Mechanics*. Addison-Wesley, Reading, MA.
Harrison,W. (1984) *Biopolymers*, **23**, 2943–2949.
Hayward,S. and Go,N. (1995) *Annu. Rev. Phys. Chem.*, **46**, 223–250.
Hinsen,K. (1998) *Proteins*, **33**, 417–429.
Hinsen,K., Thomas,A. and Field,M.J. (1999) *Proteins*, **34**, 369–382.
Hubert,R. and Bennett,W. (1983) *Biopolymers*, **22**, 261–279.
Jaaskelainen,S., Verma,C.S., Hubbard,R.E., Linko,P. and Caves,L.S. (1998) *Protein Sci.*, **7**, 1359–1367.
Karplus,M. and Kushick,J. (1981) *Macromolecules*, **14**, 325.
Kraulis,P. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Levy,R., Perahia,D. and Karplus,M. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 1346–1350.
Levy,R., Karplus,M., Kushick,J. and Perahia,D. (1984) *Macromolecules*, **17**, 1370.
Ma,J. and Karplus,M. (1997) *J. Mol. Biol.*, **274**, 114–131.
Marques,O. and Sanejouand,Y.-H. (1995) *Proteins*, **23**, 557–560.
Oh,B., Pandit,J., Kang,C., Nikaido,K., Gokcen,S., Ames,G. and Kim,S. (1993) *J. Biol. Chem.*, **268**, 11348–11355.
Perahia,D. and Mouawad,L. (1995) *Comput. Chem.*, **19**, 241–246.
Remington,S., Weigand,G. and Huber,R. (1982) *J. Mol. Biol.*, **158**, 111–152.
Swaminathan,S., Ichiye,T., Van Gunsteren,W. and Karplus,M. (1982) *Biochemistry*, **21**, 5230–5241.
Tama,F., Gadea,F., Marques,O. and Sanejouand,Y.-H. (2000) *Proteins*, **41**, 1–7.
Teeter,M. and Case,D. (1990) *J. Phys. Chem.*, **94**, 8091–8097.
Tirion,M. (1996) *Phys. Rev. Lett.*, **77**, 1905–1908.
Wiegand,G. and Remington,S. (1986) *Annu. Rev. Biophys. Biophys. Chem.*, **15**, 97–117.

Received April 10, 2000; revised October 24, 2000; accepted November 9, 2000