PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A perturbative view of protein structural variation

Julián Echave* and Francisco M. Fernández

Instituto Nacional de Investigaciones Fisicoquímicas Teóricas y Aplicadas, Consejo Nacional de Investigación Científica y Técnicas & Universidad Nacional de La Plata, La Plata, Argentina

ABSTRACT

It was recently found that the lowest-energy collective normal modes dominate the evolutionary divergence of protein structures. This was attributed to a presumed functional importance of such motions, i.e., to natural selection. In contrast to this selectionist explanation, we proposed that the observed behavior could be just the expected physical response of proteins to random mutations. This proposal was based on the success of a linearly forced elastic network model (LFENM) of mutational effects on structure to account for the observed pattern of structural divergence. Here, to further test the mutational explanation and the LFENM, we analyze the structural differences observed not only in homologous (globin-like) proteins but also in unselected experimentally engineered myoglobin mutants and in wild-type variants subject to other perturbations such as ligand-binding and pH changes. We show that the lowest normal modes dominate structural change in all the cases considered and that the LFENM reproduces this behavior quantitatively. The collective nature of the lowest normal modes results in global conformational changes that depend little on the exact nature or location of the perturbation. Significantly, the evolutionarily conserved structural core matches the regions observed to be more robust with respect to mutations, so that the core would be more conserved even under unselected random mutations. In a word, the observed patterns of structural variation can be seen as the natural response of proteins to perturbations and can be adequately modeled using the LFENM, which serves as a common framework to relate *a priori* different phenomena.

Proteins 2010; 78:173–180.

Key words: protein evolution; structural divergence; ligand binding; mutation; perturbation; model; simulation; elastic network model.

INTRODUCTION

The evolutionary conservation of protein structure has been the subject of much research during the past 30 years. In contrast, structural divergence has received much less attention. In a significant piece of work, Ortiz and coworkers focused on structural differences between homologous proteins and showed that most evolutionary structural divergence is contained within a subspace spanned by a few low-energy collective normal modes.[1] The dominant contribution of the lowest normal modes to structural divergence was interpreted as being somehow related to their functional importance.[1,2] The underlying assumption in such function-based interpretations is that it is natural selection that drives structural changes along the lowest (functional) normal modes.

To gain further insight into the mechanism of protein structural divergence, we recently proposed a "Linearly Forced Elastic Network Model" (LFENM) that models mutations by adding a linear perturbative term to the coarse grained ENM Hamiltonian of the unperturbed protein.[3] This model predicted that even unselected random mutations would make the native structure change along the lowest normal modes. The LFENM results agreed quantitatively with those obtained for homologous proteins. Therefore, the suggestion was put forward that such behavior would be just what is expected from the response of proteins to random mutations. The reasoning behind this suggestion was of the "Ockham's Razor" type: if mutation is enough to account for the normal-mode pattern of structural variation, there is no need to resort to natural selection.

If the LFENM is essentially correct, then the lowest normal modes should dominate structural change not only for homologous proteins, but also for unselected mutants. More generally, the LFENM can be seen as a model of the structural effect not only of mutations but of any type of perturbation. Therefore, if the model is correct,

Additional Supporting Information may be found in the online version of this article.
Abbreviations: βGNM, beta gaussian network model; ENM, elastic network model; LFENM, linearly forced elastic network model; LRA, linear response approximation; PDB, Protein Data Base; RMSD, root mean square deviation.

The authors state no conflict of interest.
Grant sponsor: This work was supported by Consejo Nacional de Investigación Científica y Técnica and Agencia Nacional de Promoción Científica y Tecnológica.
*Correspondence to: Julián Echave; INIFTA, Suc. 4, C.C. 16, (1900) La Plata, Argentina.
E-mail: jechave@inifta.unlp.edu.ar
Received 4 April 2009; Revised 6 July 2009; Accepted 13 July 2009
Published online 20 July 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22553

nonmutational perturbations should also change the protein's structure mainly along the lowest normal modes.

In a previous brief report, we proposed the LFENM and compared its predictions with a normal mode analysis of the structural divergence of a set of homologous globin-like proteins. The aim of the present study is to see if the LFENM predictions hold also for unselected mutations and, more generally, other types of perturbation. To this end, in addition to a set of globin-like evolutionary related proteins, we studied a set of experimentally engineered mutants of sperm whale myoglobin and a set of unmutated wild-type variants that result from other perturbations such as ligand-binding, pH changes, and Fe oxidation state changes. We also present a detailed derivation of the model.

MATERIALS AND METHODS

Linearly forced elastic network model (LFENM)

The LFENM models the effect of a mutation upon a reference "wild-type" protein. Here we show the main equations.[3] A detailed derivation can be found in the Supporting Information Appendix.

ENM of the reference protein

We consider the backbone fluctuations of the reference "wild-type" protein around its equilibrium conformation to be described by a coarse-grained elastic network model (ENM), which represents the protein as a network of nodes placed at the α carbons (C_α) connected by springs.[4–8] In general, the ENM potential is of the form

$$V_{\mathrm{wt}} = \frac{1}{2}(\mathbf{r} - \bar{\mathbf{r}}_{\mathrm{wt}})^T \mathbf{K} (\mathbf{r} - \bar{\mathbf{r}}_{\mathrm{wt}}), \qquad (1)$$

where, for a protein of N sites, $\mathbf{r}$ is a column vector with $3N$ elements: the x, y, z coordinates of the N C_α, $\bar{\mathbf{r}}_{\mathrm{wt}}$ is the equilibrium structure, and $\mathbf{K}$ is the "stiffness" matrix, which represents the network's topology and spring force constants.

LFENM model of mutant proteins

We model a point mutation (amino acid replacement) by adding a linear perturbative term to the reference potential Eq. (1):

$$V_{\mathrm{mut}} = V_{\mathrm{wt}} - \mathbf{f}^T (\mathbf{r} - \bar{\mathbf{r}}_{\mathrm{wt}}), \qquad (2)$$

where $\mathbf{f}$ is a column vector with 3N elements that models the mutation. Since $\mathbf{f} = -\left(\frac{\partial V_{\mathrm{mut}}}{\partial \mathbf{r}}\right)_{\mathbf{r}=\bar{\mathbf{r}}_{\mathrm{wt}}}$, it can be interpreted as a force that drives the mutant's structure away from that of the wild type. Eq. (2) can be derived by expanding the ENM potential in Taylor series with respect to parameter variations and keeping only first and second order terms, as is shown in the Supporting Information Appendix.

Change of equilibrium structure

The equilibrium structure of the mutant $\bar{\mathbf{r}}_{\mathrm{mut}}$ is the value of $\mathbf{r}$ that minimizes V_{mut}. Using Eqs. (1) and (2) we find the structural variation due to the mutation:

$$\delta\bar{\mathbf{r}} \equiv \bar{\mathbf{r}}_{\mathrm{mut}} - \bar{\mathbf{r}}_{\mathrm{wt}} = \mathbf{K}^{-1}\mathbf{f}. \qquad (3)$$

This equation shows that the structural perturbation introduced by a mutation is related to the mutation ($\mathbf{f}$) and to the network of oscillators, via $\mathbf{K}^{-1}$. We should note here that $\mathbf{K}^{-1}$ is actually the pseudo inverse, because $\mathbf{K}$ has six zero eigenvalues, corresponding to translations and rotations, so that it is not invertible (Supporting Information Appendix).

Model parameters

To completely specify the model, we must choose $\mathbf{K}$ and $\mathbf{f}$. There are a number of alternative ENMs.[5,7,8] In the current implementation, we have used the "beta Gaussian Network Model" (βGNM), which builds $\mathbf{K}$ from the coordinates of C_α and C_β, and uses a higher force constant for consecutive residues than for non-consecutive ones.[8]

To calculate $\mathbf{f}$, given a mutation at a site l, each site i in contact with l is assigned a force $\mathbf{f}_{l\to i}$ directed along the l-i contact and site l is assigned a reaction force $\mathbf{f}_l = -\sum_{i\neq l} \mathbf{f}_{l\to i}$. This form of $\mathbf{f}$ is not arbitrary, as shown in the Supporting Information Appendix. To simulate random mutations, the magnitudes of $\mathbf{f}_{l\to i}$ were randomly picked from a uniform distribution in $[-f_{\max}, f_{\max}]$. The forces for different contacts were picked independently. Since $f_{\max}$ does not affect the results, we set $f_{\max} = 1$. We can think of the range $[-f_{\max}, f_{\max}]$ as a continuous approximation of the perturbations introduced by the $20 \times 19 = 380$ possible mutations, covering from mutations between physicochemically similar amino acids ($f \sim 0$) up to mutations between very different amino acids ($f \sim \pm f_{\max}$).

Datasets

LFENM

We used the entry 1a6m of the Protein Data Base (PDB) as reference protein. This contains the structure determined by x-ray diffraction of the 151 sites long sperm whale oxy-myoglobin crystallized at pH 7. Starting from 1a6m, we simulated $151 \times 100 = 15{,}100$ single-point mutants using random forces as described in the previous section. To take into account the heme into the ENM we placed five extra nodes at the positions of the heme's Fe and the four CH porphyrin atoms. The structure of each mutant was calculated using Eq. (3).

Random

As a reference model, a null model to compare the LFENM with, we used a dataset of 1510 simulated structures obtained by adding to the wild-type (reference) structure a vector of dimension 3N with random elements picked from a uniform distribution with values in $(-\xi, \xi)$ with $\xi = 0.1$.

Globin-like

This dataset includes 1a6m and 21 members of the superfamily of globin-like homologous proteins, as classified in the SCOP database.[9] It includes one neural globin, two truncated hemoglobins, and 19 globins. For details, see Supporting Information Table I.

Mutants

This dataset contains 119 sperm whale myoglobin mutant structures: 22 single mutants, 77 double mutants, 7 triple mutants, and 13 quadruple mutants. Members of this dataset may also differ from 1a6m in the ligand bound to Fe and/or pH. Also, in most cases, the heme iron is in its Fe(II) state, but there are some cases with Fe(III). Finally, in one case Fe is replaced by Mn(III) and in another by Cr(III) (for details, see Supporting Information Table II). The total number of mutations is 249 and their distribution is uneven (Supporting Information Figure): for most (108) cases the aspartic acid in position 122 is replaced by asparagine. Most of the rest of the mutations are at sites 29, 64, 68, and 67 (site numbers correspond to the reference protein 1a6m).

Wild-type variants

This dataset includes 1a6m and 48 structures that have the same (wild-type) sequence. Different members of this set have different ligands, pH, and/or Fe oxidation state. There are also 3 members with Co(II) replacing Fe(II). For details see Supporting Information Table III.

Analysis

Structural differences

For each dataset, we computed the structural variation with respect to the reference protein as follows. First, structures were aligned with the reference protein using the structural alignment program Multiple Alignment with Translations and Twists (Matt).[10] For the alignment of a given protein "p" with the reference protein, the structural core consists of aligned positions with no gaps. Once aligned, vectors of core C_α coordinates $\bar{\mathbf{r}}^{\mathrm{p}}_{\mathrm{core}}$ and $\bar{\mathbf{r}}^{\mathrm{ref}}_{\mathrm{core}}$ were obtained for protein p and the reference, respectively. Then the structural variation was calculated using

$$\Delta\bar{\mathbf{r}} = \bar{\mathbf{r}}^{\mathrm{p}}_{\mathrm{core}} - \bar{\mathbf{r}}^{\mathrm{ref}}_{\mathrm{core}}. \quad (4)$$

Normal mode analysis

The normal mode analysis of structural changes was performed by projecting the structural differences onto the normal modes of the reference protein. The normal modes $\mathbf{q}_n$ and eigenvalues λ_n of 1a6m were obtained solving the eigenvalue equation:

$$\mathbf{K}\mathbf{q}_n = \lambda_n \mathbf{q}_n. \quad (5)$$

There are 3N-6 nonzero eigenvalues, which correspond to the vibrational modes, numbered $n = 0,2,\ldots 3\mathrm{N} - 7$.

For a single protein pair with structural difference vector $\Delta\bar{\mathbf{r}}$, we calculated the normal mode contributions using:

$$P_n \equiv \frac{(\mathbf{q}_n^T \Delta\bar{\mathbf{r}})^2}{\sum_n (\mathbf{q}_n^T \Delta\bar{\mathbf{r}})^2}. \quad (6)$$

The dot product in this equation is calculated over the aligned positions only. The structural variations $\Delta\mathbf{r}$ were calculated using Eq. (4). The normal mode contributions calculated using Eq. (6) satisfy $\sum_n P_n = 1$. We also calculated the overall cumulative contribution of the lowest n normal modes:

$$Q_n \equiv \sum_{m=0}^{n} P_m. \quad (7)$$

For datasets, P_n and Q_n were averaged over all members.

Root mean square deviation (RMSD) profiles

To visualize the structural difference between a given protein and the reference protein in a cartesian-coordinates representation, we calculated the square deviation of each C_α:

$$||\Delta\bar{\mathbf{r}}_i||^2 = \Delta\bar{x}_i^2 + \Delta\bar{y}_i^2 + \Delta\bar{z}_i^2, \quad (8)$$

where $\Delta\bar{\mathbf{r}}_i = (\Delta\bar{x}_i\ \Delta\bar{y}_i\ \Delta\bar{z}_i)^T$ is the column vector of cartesian displacements of the i^{th} C_α with respect to the reference structure obtained from Eq. (4). Since we are interested in the relative variability of different sites, deviations obtained using Eq. (8) were renormalized so

that the sum over sites of square deviations adds up to 1. Root square deviations $\mathrm{RMSD}_i = \sqrt{||\Delta\bar{\mathbf{r}}_i||^2}$ were plotted as a function of i to obtain site-dependent structural variation profiles. For datasets of proteins, we averaged over the dataset:

$$\mathrm{RMSD}_i = \sqrt{\langle ||\Delta\bar{\mathbf{r}}_i||^2 \rangle}, \qquad (9)$$

where $\langle \cdot\cdot \rangle$ stands for averaging over the set of pairs.

RESULTS AND DISCUSSION

To see whether the lowest normal modes dominate structural variation, we calculated the structural differences between the members of each dataset and the reference protein 1a6m and projected them onto the normal modes of the latter. Figure 1 shows the normal-mode contributions averaged over each of the datasets considered: LFENM, globin-like, mutants, and wild-type variants. The LFENM (Fig. 1, top panel), predicts a rapid decrease of normal-mode contributions with increasing normal mode number. It was shown before that the normal-mode projections are proportional to the inverse eigenvalues.[3] It can be seen from Figure 1 that a similar behavior is found for the other three datasets. The Pearson correlation coefficients between the LFENM profile and the experimental profiles of Figure 1 are 0.76, 0.74, and 0.74 for the globin-like, mutants, and wild-type datasets, respectively. These are all significant with $P \ll 10^{-2}$ and they are not significantly different from each other.

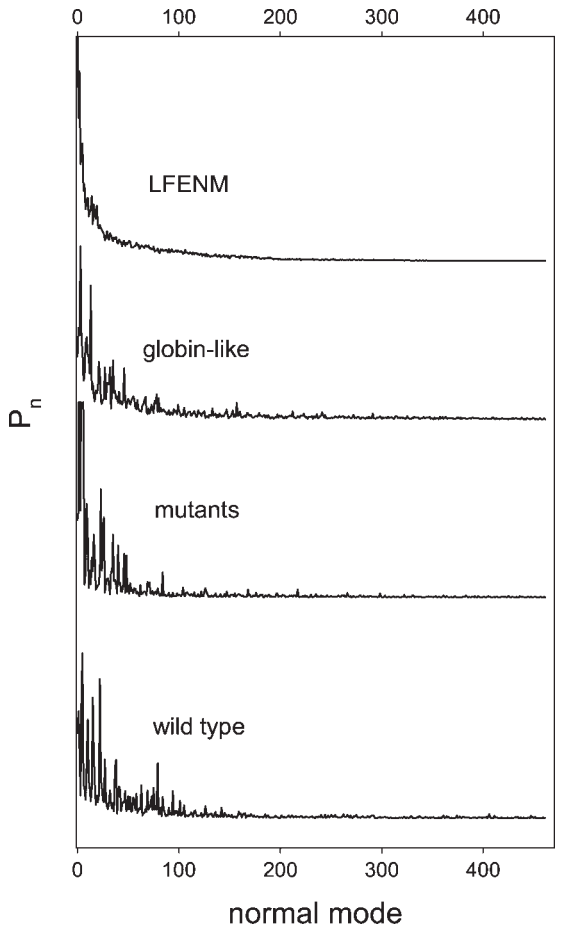

Figure 1

Normal mode patterns of structural variation: average projections. This figure plots normal mode contributions P_n as a function of normal mode number averaged over four datasets. From top to bottom: LFENM, globin-like, mutants, and wild-type variants. Normal modes correspond to the reference protein 1a6m.

Figure 2 shows the average cumulative contributions for the datasets of Figure 1 and also for the dataset obtained by introducing random structural perturbations. For any given n, the average contribution of the lowest n normal modes for the LFENM and the three experimental datasets is significantly higher than that expected for random structural changes. The structural variation of the backbone is dominated by the lowest normal modes for all three experimental datasets, in agreement with what is expected from the LFENM calculations.

To perform a more detailed comparison, for each protein of each dataset we calculated the cumulative contribution of the lowest 100 modes, Q_{100}, and estimated the probability density distribution for each dataset using kernel density estimation. We used 100 modes because it is the number necessary to describe approximately 85% of the LFENM structural variation. Results are shown in Figure 3. This figure shows that the experimental Q_{100} values lie in the range expected from the LFENM simulations and outside the range expected from random structural variations. A likelihood ratio test (not shown) results in discarding the (null) random model in favor of the LFENM model with $P \ll 10^{-2}$ for all cases. It is not expected that the experimental distributions are identical to the LFENM distribution. The model distribution is that expected from random mutations equally distributed among all sites. In contrast, the mutants dataset is highly biased: mutations at only five sites are significantly represented. Similarly, the wild-type dataset corresponds basically to changes in ligand (i.e., near the active site) and pH (i.e., at the few amino acids the protonation state of which will change with pH). In the case of globin-like proteins, which are very divergent, the Q_{100} distribution is shifted toward higher normal modes. This is an artifact due to alignment procedure, which introduces gaps at sites for which the structural divergence is too large. Despite these differences, the overall model–experiment agreement is very good.

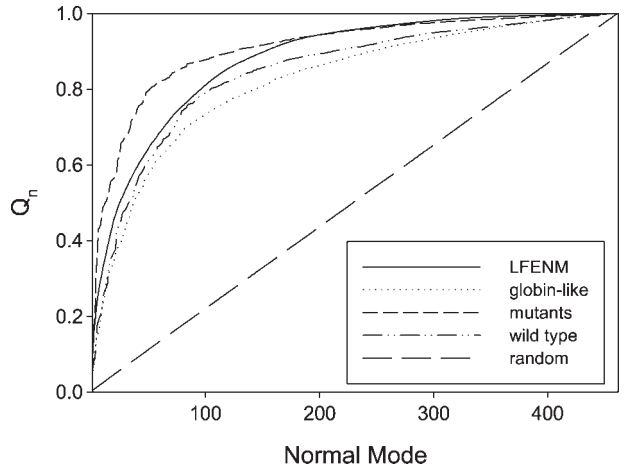

Figure 2

Normal mode patterns of structural variation: average cumulative contributions. This figure plots cumulative normal mode contributions Q_n averaged over LFENM, globin-like, mutants, and wild-type variants datasets. The null model of random structural changes is also included. Normal modes correspond to the reference protein 1a6m.

The Q_{100} values are correlated with RMSD, the magnitude of structural change. Figure 3 shows that the LFENM Q_{100} distribution is rather complex. The peaks of the distribution cannot be assigned to single sites. Accordingly, inspection of the Q_{100} values for the mutants and wild-type datasets does not lead to an easy correlation of Q_{100} with perturbed site. Even if there is

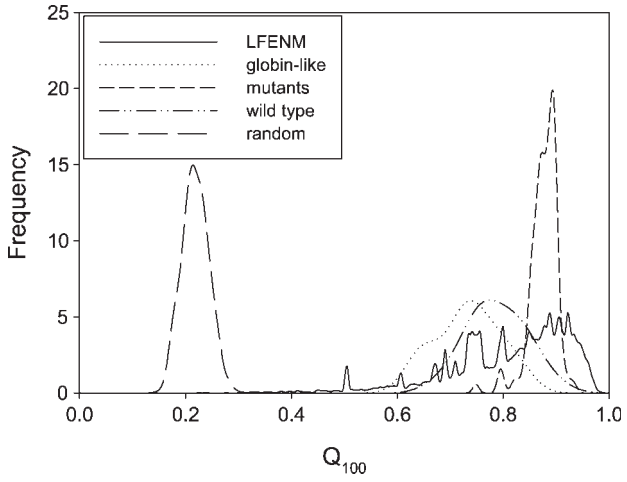

Figure 3

Normal mode patterns of structural variation: distributions of Q_{100}. This figure shows the distribution functions, obtained using the Kernel Density Estimation method, of Q_{100}, the projection onto the subspace spanned by the lowest 100 modes, for the following datasets: LFENM (15100 proteins), globin-like (20 proteins), mutants (119 proteins), wild-type variants (48 proteins), and random (1510 proteins).

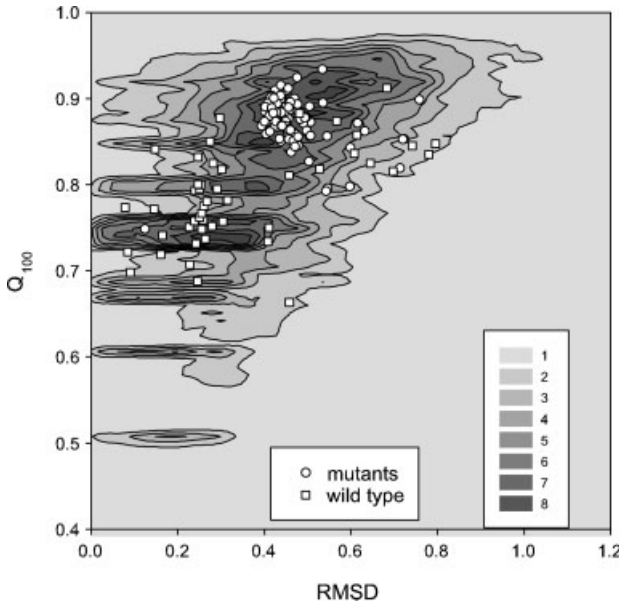

Figure 4

Correlation between normal mode contributions and magnitude of structural variation. This figure shows the two-dimensional distribution of (RMSD, Q_{100}) for the LFENM dataset and the (RMSD, Q_{100}) points for each of the members of the mutants and wild-type variants datasets. The LFENM RMSD values are proportional to f_{max}, which was scaled to make the average RMSD of the LFENM dataset equal to the average over the mutants and wild-type datasets.

no obvious interpretation of Q_{100} in terms of the perturbed site, there is a correlation between the overall magnitude of structural variation and Q_{100}. This can be seen in Figure 4, which shows the Q_{100} vs. RMSD for mutants and wild-type variants superimposed with the two-dimensional RMSD-Q_{100} distribution obtained using the LFENM. Perturbations that result in higher values of Q_{100} (i.e., more global changes) also result in larger structural variation (RMSD).

To complete the analysis, it is instructive to look at the structural variation using the more familiar representation of C_α cartesian coordinates. Figure 5 shows the RMSD profile predicted by the LFENM compared with those obtained using the globin-like, mutants, and wild-type variants datasets. The Pearson correlation coefficients between the LFENM profile and the experimental profiles are 0.53, 0.51, and 0.54 for the globin-like, mutants, and wild-type datasets, respectively (correlations calculated over aligned sites without gaps). These are all significant with $P < 10^{-2}$ and they are not significantly different from each other. We should note that these values are smaller than those obtained from the normal mode representation of the structural variation, discussed in the first paragraph of this section. However, this does not mean that the LFENM prediction is worse in the cartesian representation than in the normal mode one: the

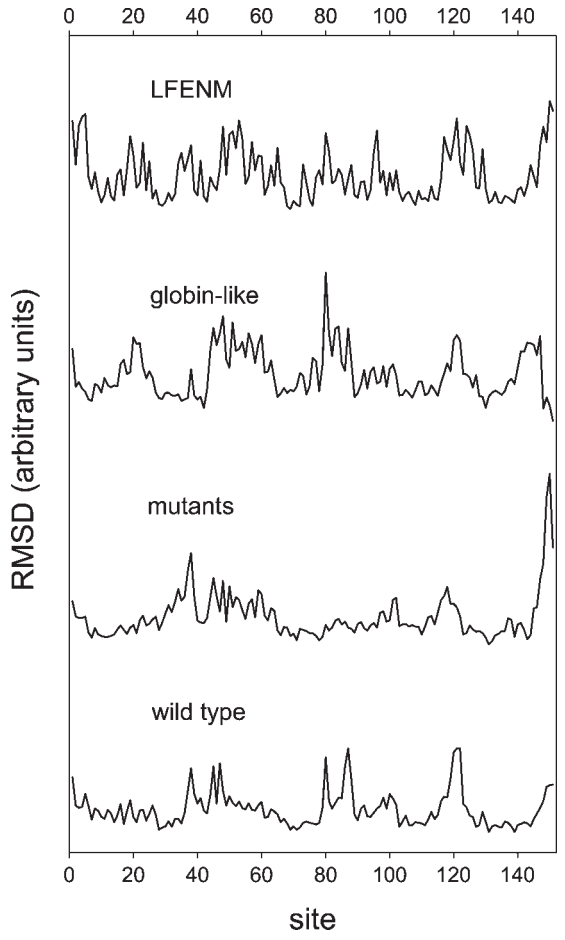

Figure 5

Patterns of structural variability along the backbone. This figure shows the C_α RMSD profiles for the LFENM, globin-like, mutants, and wild-type variants datasets. Site numbers correspond to the reference protein 1a6m.

Pearson correlation coefficient for datasets of 20 simulated mutants compared with the whole dataset of 15,100 mutants has an average value of 0.52 and a standard deviation of 0.08. This shows that the values obtained for the experimental datasets are within the range expected from the model.

To summarize, Figures 1–4 show that the lowest normal modes dominate structural variation for the globin-like, mutants, and wild-type variants datasets. The pattern predicted by the LFENM agrees with those obtained from three experimental datasets, which, in turn, are very similar to each other. These results support the idea that these patterns are just the expected response of protein structure to any type of perturbation such as mutations, ligand-binding, and pH changes.

Regarding protein evolution, the present results support the mutational explanation of the observed patterns of evolutionary structural divergence. The lowest normal modes would dominate such divergence even under random mutations, so that there is no need to resort to function-based selectionist interpretations to account for this dominance. In the cartesian representation, the dominance of the lowest normal modes, which are collective, is reflected in the global nature of the RMSD profiles shown in Figure 5. The patterns of structurally conserved/variable regions for the three experimental datasets and the LFENM predictions are similar, which means that the conserved regions (structural core) would be conserved even under random mutations with no natural selection.

The present results are consistent with previous findings on the structural changes resulting from mutations and ligand binding. To our knowledge, this is the first normal mode analysis of backbone structural changes due to mutation. However, it has been reported that local mutations at different protein sites result in global changes of the backbone conformation, the results being rather independent of the mutated site or the chemical nature of the mutation.[11] The present work accounts for this observation: mutations act via the lowest normal modes, which, being collective, result in global conformational changes. Regarding nonmutational perturbations, the results presented here are consistent with the now established fact that a few low-energy normal modes are enough to account for the backbone conformational changes resulting from ligand-binding, including protein-protein and protein-DNA binding.[12–18]

CONCLUSION

We showed that proteins subject to different perturbations including mutations, ligand-binding, and pH changes, among others, experiment structural changes of the backbone dominated by the lowest-energy normal modes. Since these modes are collective, the resulting conformational changes are global, even under local perturbations. The observed patterns are well reproduced by the linearly forced elastic network model (LFENM), which, therefore, provides a common framework to understand the role of the lowest normal modes in directing structural change due to *a priori* different causes: they are all just perturbations, and the protein responds in similar ways.

The similarity between evolutionary deformations and perturbational deformations supports a mutational interpretation of the observed patterns of evolutionary structural divergence. Mutations make protein structures evolve along the lowest energy normal modes, and the observed structurally conserved regions match those whose conformation is more robust with respect to mutations. Of course, this does not mean that natural selection plays no role in protein evolution. Using a

simple neutral-evolution model of selection, we found that the observed patterns of structural variation (i.e., the relative contributions of different normal modes) may reflect the mutational effects even under the effect of natural selection.[3] Moreover, if there were no natural selection, different sites would have the same evolutionary rates, which is not the case. Natural selection does operate, it is just not the reason behind the observed dominance of the lowest normal modes.

A related issue that deserves consideration is that of the functional importance of the lowest normal modes. Numerous studies successfully used normal modes to interpret protein function.[2] Moreover, it has been shown that functionally different homologous protein families cluster in different regions of normal-mode space.[1] Similarly, in this work we found that the distribution of Q_{100} for the globin-like dataset shown in Figure 3 is bimodal: the peak located at smaller values of Q_{100} is due to neuroglobins (one case) and truncated hemoglobins (two cases), whereas the peak at higher Q_{100} values corresponds to globins. Interpretations that regard this clustering as being obviously connected with the functional role of these modes implicitly assume adaptive natural selection.[1,2] However, since proteins diverge within a subspace spanned by the lowest normal modes even under unselected mutations or neutral selection,[3] different protein families might cluster in different regions of normal mode space because of phylogenetic history. Further studies would be needed to disentangle the effects of mutation and natural selection.

The LFENM is based on the ENM, which has been thoroughly validated. The normal mode analysis of experimental datasets does not depend on the assumptions of the LFENM, just on those of the ENM. The ability of elastic network models to produce accurate backbone normal modes has been shown in several assessments.[6–8,19–21] They have also proved useful in many applications aimed at gaining insight into the functional role of collective motions (for a comprehensive review see Ref. 2).

Despite the assumptions of the LFENM, it is validated by its agreement with experimental results. The LFENM adds only one extra parameter to the ENM, f_{max}, and the results reported are independent of its value. The agreement of LFENM predictions with the results obtained for the different experimental datasets validate the model. To further test the model, we repeated the computations by changing the model parameters (spring force constants and cutoff value to define contacts) by 20%, which had almost no effect on the results (not shown). A detailed all-atom model may be necessary to predict the effect of a single specific perturbation, such as binding of a specific ligand, or a specific point mutation, but the coarse grained LFENM has enough detail to produce averages and distributions in accordance with experimental observations.

Before finishing, we compare the LFENM with other related approaches. A linear response approximation (LRA) based on an all-atom representation of a protein subject to a linear perturbative force was used to successfully predict ligand-binding effects on structure.[22] The later approach did not propose a Hamiltonian to model the perturbed protein, and did not analyze the results in terms of normal modes, but is related to the LFENM in that for the LFENM Hamiltonian the LRA would be exact. The LRA formula to calculate the structural change expected from an external force for an ENM is identical to Eq. (3).[8] Another perturbational approach based on perturbing the spring force-constants has also been proposed.[23] Perturbing the force constants amounts to modifying the stiffness matrix in Eq. (1) and would not produce a change in conformation; therefore such an approach cannot be used to predict conformational variation. Rather than the force-constants, the LFENM perturbs the equilibrium lengths of the springs connecting the network's nodes, which results in a structural change. Finally, we mention that the LFENM is similar to a perturbative model developed independently to predict conformational changes using distance constraints,[24] the main difference being that, in our case, the perturbative term models mutations rather than distance constraints.

To finish, we would like to mention that the LFENM can be extended to study the evolutionary divergence of protein motions. Because the LFENM does not take into consideration changes of the stiffness matrix of the elastic network, the normal modes of the simulated mutants are identical to those of the wild type. It is known that the overall pattern of protein flexibility is evolutionarily conserved,[25] but not all modes are conserved in the same degree: the lowest normal modes are the most conserved.[26–29] Normal mode conservation patterns can be accounted for by an extension of the LFENM that takes into account changes of the stiffness-matrix due to structural perturbations in a self-consistent way, as we will show elsewhere.

ACKNOWLEDGMENTS

This article is dedicated to the memory of Angel Ortiz. We thank Ugo Bastolla and Angel Ortiz for inspiring discussions, Cristian Micheletti for providing his βGNM code, and Alejandra Leo-Macias and Angel Ortiz for providing their data on structural divergence. Both authors are members of CONICET.

REFERENCES

1. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. An analysis of core deformations in protein superfamilies. Biophys J 2005;88:1291–1299.
2. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.

3. Echave J. Evolutionary divergence of protein structure: the linearly forced elastic network model. Chem Phys Lett 2008;457:413–416.
4. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
5. Hinsen K, Kneller GR. A simplified force field for describing vibrational protein dynamics over the whole frequency range. J Chem Phys 1999;111:10766–10769.
6. Hinsen K, Petrescu AJ, Dellerue S, Bellissent-Funel MC, Kneller GR. Harmonicity in slow protein dynamics. Chem Phys 2000;261:25–37.
7. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
8. Micheletti C, Carloni P, Maritan A. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and gaussian models. Proteins 2004;55:635–645.
9. Murzin AG, Brenner SE, Hubbard T, Chothia C. Scop—a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
10. Menke M, Berger B, Cowen L. Matt: local flexibility aids protein multiple structure alignment. Plos Compu Biol 2008;4:e10.
11. Sinha N, Nussinov R. Point mutations and sequence variability in proteins. Redistributions of preexisting populations. Proc Natl Acad Sci USA 2001;98:3139–3144.
12. Reuter N, Hinsen K, Lacapere JJ. Transconformations of the SERCA1 Ca-ATPase: a normal mode study. Biophys J 2003;85: 2186–2197.
13. Xu CY, Tobi D, Bahar I. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T ↔ R2 transition. J Mol Biol 2003;333:153–168.
14. Zheng W, Brooks BR, Thirumalai D. Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. Biophys J 2007;93:2289–2299.
15. Keskin O. Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. BMC Struct Biol 2007;7:31.
16. Tobi D, Bahar I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. Proc Natl Acad Sci USA 2005;102:18908–18913.
17. Dobbins SE, Lesk VI, Sternberg MJE. Insights into protein flexibility: the relationship between normal modes and conformational change upon protein-protein docking. Proc Natl Acad Sci USA 2008;105:10390–10395.
18. Delarue M, Sanejouand YH. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J Mol Biol 2002;320:1011–1024.
19. Rueda M, Chacon P, Orozco M. Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. Structure 2007;15:565–575.
20. Hinsen K. Structural flexibility in proteins: impact of the crystal environment. Bioinformatics 2008;24:521–528.
21. Emperador A, Carrillo O, Rueda M, Orozco M. Exploring the suitability of coarse-grained techniques for the representation of protein dynamics. Biophys J 2008;95:2127–2138.
22. Ikeguchi M, Ueno J, Sato M, Kidera A. Protein structural change upon ligand binding: linear response theory. Phys Rev Lett 2005;94: 078102.
23. Zheng WJ, Brooks BR, Doniach S, Thirumalai D. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure 2005;13:565–577.
24. Zheng WJ, Brooks BR. Normal-modes-based prediction of protein conformational changes guided by distance constraints. Biophys J 2005;88:3109–3117.
25. Maguid S, Fernandez-Alberti S, Parisi G, Echave J. Evolutionary conservation of protein backbone flexibility. J Mol Evolut 2006;63: 448–457.
26. Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J. Exploring the common dynamics of homologous proteins. Application to the globin family. Biophys J 2005;89:3–13.
27. Maguid S, Fernandez-Alberti S, Echave J. Evolutionary conservation of protein vibrational dynamics. Gene 2008;422:7–13.
28. Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78: 2093–2106.
29. Zen A, Carnevale V, Lesk AM, Micheletti C. Correspondences between low-energy modes in enzymes: dynamics-based alignment of enzymatic functional families. Protein Sci 2008;17:918–929.