Comparative analysis of rigidity across protein families

S A Wells¹, J E Jimenez-Roldan¹² and R A Römer¹

¹ Department of Physics and Centre for Scientific Computing, University of Warwick, Coventry, CV4 7AL, UK
² Department of Systems Biology, University of Warwick, Coventry, CV4 7AL, UK
E-mail: r.roemer@warwick.ac.uk

Received 20 May 2009
Accepted for publication 17 August 2009
Published 22 September 2009
Online at stacks.iop.org/PhysBio/6/046005

Abstract
We present a comparative study in which ‘pebble game’ rigidity analysis is applied to multiple protein crystal structures, for each of six different protein families. We find that the main-chain rigidity of a protein structure at a given hydrogen bond energy cutoff is quite sensitive to small structural variations, and conclude that the hydrogen bond constraints in rigidity analysis should be chosen so as to form and test specific hypotheses about the rigidity of a particular protein. Our comparative approach highlights two different characteristic patterns (‘sudden’ or ‘gradual’) for protein rigidity loss as constraints are removed, in line with recent results on the rigidity transitions of glassy networks.

1. Introduction

It is a common goal in biophysics to represent the flexibility of a protein and study its large-scale motion without incurring the full computational cost of molecular dynamics simulations. One popular family of approaches is based on normal-mode analysis applied to a full or simplified representation of the protein structure [1–10], with the aim of representing large-scale conformation change in terms of a reduced set of low-frequency motions [11]. Another approach is to divide up the protein structure into relatively rigid sections or domains, connected together by flexible regions or ‘hinges’. This can be done using a variety of structure-based approaches [12–17].

In this paper we concern ourselves with the ‘pebble game’ [18], an integer algorithm for rigidity analysis. By matching degrees of freedom against constraints, it can rapidly divide a network into rigid regions and floppy ‘hinges’ with excess degrees of freedom. The program FIRST implements this algorithm for protein crystal structures [19]. The rigid units in a protein structure may be as small as individual methyl groups or large enough to include entire protein domains containing multiple secondary-structure units. The division of a structure into rigid units is referred to as a rigid cluster decomposition (RCD).

Rigidity analysis has been used to study phenomena such as virus capsid assembly [20] and protein folding [21, 22]. The coarse graining provided by a RCD also forms the basis of simulation methods aiming to explore the large-amplitude flexible motion of proteins: the ROCK algorithm [23] and more recently the FRODA geometric simulation algorithm [24], which has been applied in various studies of protein flexibility [25–29], and the rigidity-enhanced elastic network model [30].

The results of rigidity analysis on proteins depend upon the set of constraints that are included. The primary control on the constraints is an energy ‘cutoff’ set by the user, which determines the set of hydrogen bonds to include in the analysis (see section 2). However, previous studies using FIRST have used widely differing, sometimes contradictory, cutoff values and methods of constraint selection—we give a brief review of the situation in the appendix. This methodological issue not only makes it more difficult for scientists to adopt pebble-game rigidity analysis as a method but also raises issues in the interpretation of results. There is at present no clear guidance on the ‘correct’ choice of cutoff value; nor is it clear how comparable are the results of rigidity analysis using a given cutoff value on slightly different protein structures.

Hence the primary motivation for our study is to fill this gap by explicitly comparing the results of rigidity analysis in the context of six different protein families.
analysis on groups of very similar crystal structures. We concentrate particularly on eukaryotic cytochrome C while also considering five other proteins (hemoglobin, myoglobin, α-lactalbumin, trypsin and HIV-1 protease). For each protein structure we observe the pattern of rigidity loss during the progressive removal of hydrogen bonds, or ‘rigidity dilution’ [21, 22]. We define main-chain rigidity as a measure of the rigidity of the protein backbone in order to describe the rigidity loss during dilution. On the basis of this study we comment on the selection of cutoff values and the interpretation of rigidity analyses.

The second motivation for our study is to observe the pattern of rigidity loss during dilution. Previous studies on protein folding [21] have drawn comparisons between the folding transition of proteins and the rigidity transition of glassy networks. A recent study [31] found that the rigidity transition in glasses could display either first-order or second-order behaviour depending on the character of the constraint network. In the first case, a small change in the constraints causes a sudden transition from an entirely floppy state to one in which the entire system becomes rigid. In the second, rigidity develops in a percolating rigid cluster which initially involves only a small proportion of the network and then gradually increases in size as more constraints are introduced. Our data on rigidity dilution show that both types of transitions are possible in proteins, with four of our proteins typically displaying ‘gradual’ rigidity change and two (trypsin and HIV-1 protease) displaying ‘sudden’ rigidity change.

2. Materials and methods

2.1. Protein selection

We have chosen sets of proteins from the protein data bank (PDB) [32] to obtain similar crystal structures for our comparison, as summarized in table 1. We sought particularly (i) examples of the same protein from different organisms, e.g. cytochrome C proteins from multiple different eukaryotic mitochondria, and (ii) protein structures obtained under different conditions of crystallization, e.g. in complex with different ligands, proteins or substrates. In the present study we will only investigate non-membrane proteins because the default treatment of hydrogen bonds and hydrophobic tethers in FIRST is based on the assumption that the protein exists in a polar solvent (cytoplasm) rather than being within a hydrophobic or amphiphilic environment as for membrane-bound proteins. Proteins in a membrane environment can still be handled but this requires hand-editing of the constraint network. Rigidity analysis is best carried out on crystal structures with high resolution, so that we can have confidence in the accuracy of the atomic positions when constructing the hydrogen bond geometries. We therefore concentrated on x-ray crystal structures with resolutions of better than 2.5 Å.

From each PDB crystal structure we extracted a single protein chain, eliminating all crystal water molecules, but retaining important hetero groups such as the porphyrin/heme units of cytochrome C and hemoglobin. The PYMOL visualization software [33] proved very useful for this purpose.

2.2. Rigidity analysis and dilution

Rigidity analysis in FIRST involves balancing the degrees of freedom of the atoms in the protein structure against the constraints provided by the various forms of bonding that stabilize a protein. The covalent bonding between atoms is of course included, as are hydrophobic interactions between adjacent hydrophobic side chains. Hydrogen bonds are identified based on donor-hydrogen acceptor geometry; the ‘salt bridge’ interaction between adjacent, oppositely charged ionic groups is also identified. Non-specific long-range forces (such as general electrostatic and dispersion interactions) are not counted as constraints. This hierarchy of constraints is discussed in detail in the literature on FIRST [19].

The energy of each potential hydrogen bond in the processed structure is calculated in FIRST using the Mayo potential [35]; the distance-dependent part of this potential is shown in figure 1. For the dilution, FIRST performs an initial rigidity analysis including all bonds with energies of 0 kcal mol$^{-1}$ or lower; bonds are then removed in order of strength, gradually reducing, or ‘diluting’, the rigidity of the structure.

An example of this rigidity dilution for a given protein is shown in figure 2(a) for the 1HRC horse cytochrome C structure. The horizontal axis represents the protein’s linear primary structure. Flexible areas of the polypeptide sequence are shown as horizontal thin black lines while areas lying within a rigid cluster are shown as thicker coloured blocks. Colour is used to differentiate which residues belong to which rigid cluster. The three-dimensional protein fold makes it

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{hydrogen_bond_energy_graph.png}
\caption{Dependence of the hydrogen bond energy $E$ in FIRST on the donor–acceptor distance. The shaded region indicates how a distance variation of ±0.1 Å can lead to a variation in the bond energy of more than 1 kcal mol$^{-1}$.}
\end{figure}

We add the hydrogen atoms that are absent from x-ray crystal structures, using the REDUCE software [34] which also performs necessary flipping of side chains. After the addition of hydrogen atoms we renumbered the atoms using PYMOL again to produce files usable as input to FIRST [19, 21]. In the case of HIV protease we analysed the homodimer unit, as in [19].
Table 1. List of all the proteins, their organism of origin, PDB codes as well as the figures in which they appear.

| Protein     | Organism | PDB ID | Figure  | Comments                                      |
|-------------|----------|--------|---------|-----------------------------------------------|
| Cytochrome C| Horse    | 1HRC   | 4, 6(a) | Uncomplexed                                   |
|             |          | 1WEJ   |         | Complexed with antibody E8                    |
|             |          | 1U75   |         | Complexed with peroxidase                      |
|             |          | 1CRC   |         | At low ionic strength                         |
| Cytochrome C| Tuna     | 5CYT   | 5, 6(b) | Ferricytochrome                               |
|             |          | 1HS4   |         | 2FE:1ZN mixed-metal porphyrins                |
|             |          | 1HS5   |         | 2ZN:1FE mixed-metal porphyrins                |
|             |          | 1LFC   |         | Cobalt(III)-substituted                       |
|             |          | 1LFM   |         |                                               |
| Cytochrome C| Rice     | 1CCR   | 7(a)    |                                               |
|             | Bonito   | 1CYC   |         |                                               |
|             | Bacteria | 1A7V   |         |                                               |
|             | Tuna     | 1I55   |         |                                               |
|             | Yeast    | 1YCC   |         |                                               |
|             |          | 2YCC   |         |                                               |
| Myoglobin   | Horse    | 1DWR   | 7(b)    |                                               |
|             | Whale    | 1HJT   |         |                                               |
|             | Turtle   | 1LHS   |         |                                               |
| α-Lactalbumin| Baboon  | 1ALC   | 7(c)    |                                               |
|             | Human    | 1HML   |         |                                               |
|             | Goat     | 1HFC   |         |                                               |
|             | Human    | 1HFC   |         |                                               |
|             | Guinea pig| 1I54   |         |                                               |
|             | Cattle   | 1I55   |         |                                               |
|             |          | 1I54   |         |                                               |
|             |          | 1I55   |         |                                               |
|             |          | 1LFC   |         |                                               |
| Hemoglobin  | Human    | 1A3N   | 7(d)    | Deoxy                                         |
| (α chain)   |          | 2DN1   |         |                                               |
|             |          | 2DN2   |         |                                               |
|             |          | 2DN3   |         |                                               |
|             | Goose    | 1A4F   |         |                                               |
|             | Rice     | 1D8U   |         |                                               |
|             | Bacteria | 1DLW   |         |                                               |
|             | Alga     | 1DLX   |         |                                               |
|             | Cattle   | 1G09   |         |                                               |
|             | Worm     | 1GR7   |         |                                               |
|             | Clam     | 1MOH   |         |                                               |
| HIV-1 Protease| Virus  | 1HTG   | 7(e)    | homodimers with inhibitors bound              |
|             |          | 4HVP   |         |                                               |
|             |          | 7HVP   |         |                                               |
|             |          | 8HVP   |         |                                               |
|             |          | 9HVP   |         |                                               |
| Trypsin     | Salmon   | 1A0J   | 7(f)    |                                               |
|             | Cattle   | 1AQ7   |         |                                               |
|             |          | 1AUJ   |         |                                               |
|             | Pig      | 1AVW   |         |                                               |
|             | Pig      | 1AVX   |         |                                               |
|             | Cattle   | 1AZ8   |         |                                               |
|             | Rat      | 1BRA   |         |                                               |
|             |          | 1BRB   |         |                                               |
|             |          | 1BRC   |         |                                               |
|             | Cattle   | 1BTH   |         |                                               |
|             | Salmon   | 1BZB   |         |                                               |
|             | Human    | 1H4W   |         |                                               |
|             |          | 1HPT   |         |                                               |
|             | Cattle   | 1K11   |         |                                               |
|             |          | 1K15   |         |                                               |
|             |          | 1K1M   |         |                                               |
|             |          | 1K1N   |         |                                               |
|             |          | 1K10   |         |                                               |
|             |          | 1KIP   |         |                                               |
|             | Pig      | 1LDT   |         |                                               |
|             | Human    | 1TRN   |         |                                               |
|             |          | 2RA3   |         |                                               |
|             | Rat      | 3TGI   |         |                                               |
dilution of constraints by progressively lowering the cutoff energy for inclusion of hydrogen bonds in the constraint network. Each time the rigid cluster analysis of the main-chain α-carbon atoms (Cα) changes as a result of the dilution, a new line is drawn on the plot, labelled with the energy cutoff and with the network mean coordination for the protein at that stage. We should stress that the RCD is always performed over the entire protein structure (main-chain and side-chain atoms) and a dilution is performed for every hydrogen bond removed from the set of constraints, typically several hundred bonds for a small globular protein. The dilution plot is then a summary concentrating on the rigid-cluster membership of the Cα atoms defining the protein backbone.

2.3. Main-chain rigidity loss during dilution

Dilution plots of very similar protein structures can be compared directly as shown in figure 4. This form of comparison, however, becomes unwieldy when comparing large numbers of structures, and can obscure differences in the hydrogen bond energy scale. For glassy networks [31] the overall degree of rigidity of the structure was measured by the number of atoms in the largest spanning rigid cluster in a network with periodic boundary conditions. Since the protein is not a periodic structure, we measure its overall rigidity by considering how many of its residues are included in large rigid clusters.

In figure 3(a) we show the number $N_{\alpha}$ of Cα contained within the larger $N$ rigid clusters of the horse cytochrome C structure 1HRC, for which the total number of Cα atoms equals $N_{\alpha} = 105$. It is clear that only the first few rigid clusters (numbered 1–5) contain more than one Cα while higher numbered clusters do not contain more than one Cα and do not represent two or more residues forming a single rigid unit. In figure 3(b) we show the fraction $f_N$ of Cα contained in the first $N$ cluster, defined as

$$f_N(E) = \frac{1}{N_{\alpha}} \sum_{1}^{N} n_N(E)$$

for, e.g. those Cα lying within rigid clusters $N = 1$–5 and also 10. The inclusion of the first five rigid clusters captures the large-scale rigidity of the protein; the difference between $N = 5$ and $N = 10$ is minimal. We therefore use the $N = 5$ measure, $f_5(E)$, to quantify protein rigidity hereafter, which we will refer to as main-chain rigidity. We emphasize that we have also computed all results presented here for $N = 4$ and $N = 6$ with quantitatively similar and qualitatively identical results.

It is worth noting the ‘stepped’ appearance of our graphs. This is because a given pattern of rigidity persists as the cutoff is lowered until at a specific value it changes and a certain amount of rigidity is lost.

2.4. Structural comparison by RMSD

When dealing with slightly varying crystal structures of the same protein, we quantify the structural variation by aligning
Table 2. Root-mean-square deviation in Å for Cα positions among (a) four horse cytochrome C structures and (b) four tuna cytochrome C structures, showing the similarity of the structures.

| From \ to: | 1HRC | 1CRC | 1WEJ |
|-----------|------|------|------|
| 1CRC      | 0.32 | --   | --   |
| 1WEJ      | 0.318| 0.321|--   |
| 1U75      | 0.472| 0.53 | 0.572|

| From \ to: | 5CYT | 1I55 | 1I54 |
|-----------|------|------|------|
| 1I55      | 0.27 | --   | --   |
| 1I54      | 0.2668| 0.041|--   |
| 1LFM      | 0.286| 0.116| 0.087|

The Cα atoms of two structures and obtaining the root-mean-square deviation between Cα positions,

\[ d = \sqrt{\frac{1}{N_{\text{Ca}}} \sum_{i=1}^{N_{\text{Ca}}} d_{ii}^2} \]  

where \( d_{ii} \) is the distance between the Cα atoms of residue \( i \) in the aligned structures.

3. Results and discussion

3.1. Comparing rigidity of very similar proteins: cytochrome C

In figure 4 we show dilution plots for four mitochondrial cytochrome C structures obtained from horse crystallized under different conditions as detailed in table 1. The structural variations between these four structures are small (table 2a), the largest being 0.572 Å between 1U75 and 1WEJ; for comparison, Minany and Levitt [36] consider structures within \( d \approx 4\AA \) as ‘near-native’.

The patterns of rigidity loss shown in figure 4 appear quite similar on first inspection. The central portion of the protein sequence breaks up into smaller clusters and then becomes entirely flexible, while the rigidity of the two ends of the sequence, around residues 5–15 and 90–100, persists longer; due to this persistence, these portions (α-helical in the secondary structure) were identified in [22] as being the folding core of cytochrome C, in agreement with experimental evidence.

On closer inspection, however, we can see differences between the four structures in the cutoff energies in which changes in rigidity take place. For example, in structures 1HRC and 1WEJ, the terminal α-helical sequences remain rigid down to cutoff values below \(-3\) kcal mol\(^{-1}\), while in 1CRC and 1U75 these sequences are already largely flexible at a cutoff value of \(-2\) kcal mol\(^{-1}\). We plot the main-chain rigidity of these four proteins as a function of cutoff energy during dilution in figure 6(a). The differences in energy scale of the rigidity loss is now clearly visible. Note in particular that in the energy range around \(-0.1\) to \(-0.6\) kcal mol\(^{-1}\), two of the structures retain main-chain rigidity (\( f_\alpha > 0.9 \)) while the other two have already dropped to \( f_\alpha < 0.5 \).

In figure 5, we now consider mitochondrial cytochrome C structures (from tuna) which differ only in their heme-group metal content and are structurally very similar (RMSD values given in table 2b). We see that the dilution plots for the tuna protein have similar shapes and indeed are quite similar to those for the horse protein (figure 4). There are differences, however: in particular, in the structure 1I54 the α-helical region at residues 60–70 remains rigid to lower cutoff values than that at residues 90–100, which would disagree with the ‘folding core’ prediction of [22]. We would therefore argue that physical conclusions drawn from rigidity analysis should be based on the comparison of as many structures as possible if they are to be robust.

Once we plot main-chain rigidity as a function of cutoff energy we again observe differences in the energy scales at which rigidity is lost (figure 6(b)). The greatest discrepancy appears in the energy range from \(-1\) to \(-2\) kcal mol\(^{-1}\); here the 5CYT structure has \( f_\alpha \approx 0.4 \) while 1I55 has \( f_\alpha \approx 0.9 \),

Figure 3. (a) The number \( N_N \) of Cα atoms contained within rigid clusters (RC) \( N = 1, \ldots, 5 \) and 10 of the 1HRC structure. Smaller, higher-numbered clusters do not contain more than one Cα. (b) The fraction \( f \) of the protein’s Cα atoms contained within clusters 1 to N. The line corresponding to the \( N = 5 \) data has been shaded to show that the inclusion of rigid clusters 1 through 5 captures the large-scale rigidity of the protein.
although the structures differ by less than $d = 0.3$ Å in $C_{\alpha}$ RMSD.

### 3.2. Variability of energy scales and selection of cutoff values

It is clear from our investigation of cytochrome C structures that the rigidity analysis at a given cutoff value on very similar structures can easily produce different results. This is not because the dilution plots for these structures differ drastically in their shape, but rather because the cutoff energy at which a major change in rigidity takes place can differ by approximately 1 kcal mol$^{-1}$ between very similar structures. This sensitivity of cutoff energy scales to small structural variations is understandable if we consider, for example, the distance dependence for the hydrogen bond energy function [19]: we show in figure 1 that a variation in the donor–acceptor distance of only 0.1 Å can shift the hydrogen bond energy by around 1 kcal mol$^{-1}$. Thus while the hydrogen bond energy function is successful in distinguishing weaker from stronger bonds, its resolution is limited to approximately 1 kcal mol$^{-1}$.

This implies that exact values of the hydrogen bond cutoff energy cannot be easily transferred between different crystal structures. Rather, it is advisable to perform rigidity dilution on the specific protein structure(s) of interest and to observe how the rigidity changes as the weaker bonds are eliminated, and which portions of the structure are most stable, before selecting an appropriate cutoff value for further investigation of the rigidity/flexibility of the structure(s). While this is, in a sense, the implicit message of the wide variety of cutoff values used in previous studies (see the appendix) we believe the point should be made explicitly for the benefit of potential users of the method.

Figure 4. Dilution plots for four crystal structures of horse cytochrome C. The four structures are very similar to each other (see the text) and display similar patterns of rigidity loss. The central portion of the protein sequence breaks up into smaller clusters (e.g. close to $E = -1$ for 1HRC and $E = -0.7$ for 1WEJ) and then becomes entirely flexible, while the rigidity of the two ends of the sequence, around residues 5–15 and 90–105, persists longer; these portions are $\alpha$-helical in a secondary structure.
3.3. Molecular dynamics and structural variability

We note that the RMSD variations among the crystal structures in section 3.1 are comparable in magnitude to the variations seen in molecular dynamics simulations. Previous studies have applied rigid body analysis to frames from MD trajectories [37, 38]. It appears that individual frames in the course of an MD simulation can display RCDs that differ considerably from each other and from that of the input crystal structure [37] but that consistent results can be obtained by averaging...
Figure 7. Rigidity dilutions for different families of proteins: cytochrome C, myoglobin, α-lactalbumin, hemoglobin, HIV-1 protease and trypsin. We can see that proteins can display either a ‘gradual’ (a–d) or a ‘sudden’ (e, f) pattern of rigidity loss.

over the course of an MD trajectory [37, 38]. This is quite understandable, since any particular MD frame may catch a group of atoms momentarily quite far from their equilibrium geometry. Our sets of similar crystal structures, on the other hand, represent equilibrium positions, differing from one structure to another because of differences in the conditions of crystallization.

3.4. Patterns of rigidity loss

For the cytochromes that we have so far considered (3.1), the general pattern is one of gradual rigidity loss, particularly for $|E| > 1$. This indicates a hierarchy of stability in the rigid clusters, with some areas being rigidified by very weak hydrogen bonds, some by bonds of medium strength and some by the strongest bonds. This is reminiscent of the gradual or second-order rigidity transition observed in some glassy networks [31], specifically those with a wide diversity in their constraint distribution. Glassy networks with less diverse constraint networks, however, show a sudden, first-order-like rigidity transition in which the structure passes between largely flexible and largely rigid states on the addition or removal of only a few constraints.
In figure 7 we show the patterns of rigidity loss for six different families of proteins as listed in table 1. Our sample falls into two classes, those displaying a gradual pattern of rigidity loss (figure 7, (a) cytochrome C, (b) myoglobin, (c) lactalbumin and (d) hemoglobin) and those displaying a sudden loss of rigidity once weak bonds are eliminated (figure 7, (e) HIV-1 protease and (f) trypsin). For proteins in this second class, all the 25 structures that we examine display rapid loss of main-chain rigidity as weak bonds are removed and the main chain has become almost entirely flexible once the cutoff energy is reduced below $-2 \text{ kcal mol}^{-1}$. This indicates that the rigidity of clusters in these proteins is due to weaker hydrogen bonds and we do not see (as we do in the other four proteins) the persistence of rigid clusters bound by stronger hydrogen bonds.

The HIV-1 protease is a natural homodimer and we consider the rigidity of the dimer, as in [19]. For the other protein families our data are obtained from single protein chains; for example, for hemoglobin we analyse $\alpha$-hemoglobin chains. It should be clear that a protein chain treated in isolation always has fewer constraints then when treated as part of a complex, and indeed we find that the individual chains from the HIV-1 protease structure are even less rigid than the entire dimer (data in supplementary materials available at stacks.iop.org/PhysBio/6/046005). For the case of hemoglobin, we can confirm that the rigidity of the isolated A-chain seen in figure 7(d) differs only slightly from the rigidity of the same chain when analysed as part of the full tetrameric hemoglobin structure (data in supplementary materials available at stacks.iop.org/PhysBio/6/046005). Consideration of isolated HIV-1 protease monomers, or of full hemoglobin tetrameric complexes, thus does not alter their classification in terms of gradual or sudden loss of rigidity.

The comparison of these six protein families thus leads us to the conclusion that protein structures, such as glassy networks, can display two distinct patterns of rigidity loss depending on the diversity of their constraint networks. We have identified two families of proteins, HIV protease and trypsin, whose members display rapid loss of rigidity as weaker hydrogen bonds are eliminated, in contrast to four other families of proteins which display a gradual loss of rigidity indicating a hierarchy of hydrogen bond strengths in the constraints that maintain protein rigidity.

We have previously noted (section 3.1) that the portions of the crystal structure that retain rigidity longest during dilution are generally $\alpha$-helical. The HIV-1 protease and trypsin structures consist almost entirely of the $\beta$-sheet structure, in contrast to the other four families in our study all of which have mostly the $\alpha$ structure. Thus it seems that the most likely explanation for these two classes of behaviour during rigidity dilution lies in the difference of the secondary structure. This difference of pattern in rigidity loss between mostly $\alpha$ and mostly $\beta$ structures has not, it seems, been explicitly remarked on the first literature.

4. Conclusion and outlook

Our motivation in this study was twofold: to clarify a methodological issue in the use of rigidity analysis on protein structures, by determining the robustness of RCDs against small structural variations and the significance of the cutoff energy value, and to obtain an insight into the patterns of rigidity loss during hydrogen bond dilution, by comparison with the observed patterns in glassy networks.

On the first point, we find that there is considerable variation in the RCDs of structurally similar proteins during dilution. Figure 6, for example, shows that among a group of cytochrome C structures drawn from similar eukaryotic mitochondria, energy cutoffs in the range from 0 to $-2 \text{ kcal mol}^{-1}$ (such as have typically been used for FIRST/FRODA simulations of flexible motion [24, 25, 27, 28]) can produce a wide range of degrees of main-chain flexibility. We conclude that the results of rigidity analysis on individual crystal structures should not be over-interpreted as being ‘the’ RCD for a protein. The hydrogen bond energy function in FIRST is quite sensitive to small structural variations; while it successfully divides weaker from stronger bonds, it is not possible to identify a unique value of the hydrogen bond cutoff energy which can be applied to all protein structures to give meaningful results. Rather, each protein structure should first be subjected to rigidity dilution to produce a dilution plot; a suitable value of the cutoff energy can then be chosen to test a specific hypothesis about the rigidity and flexibility of the protein. Similarly, when physical significance is attached to the pattern of rigidity loss [22], then multiple similar examples of a given protein structure should be studied in order to be robust against structural variation.

On the second point, we find that proteins can display either gradual (second-order-like) or sudden (first-order-like) patterns of rigidity loss during dilution. We find sudden rigidity loss in two proteases, eukaryotic trypsin and viral HIV-1 protease. Both consist largely of the $\beta$-sheet secondary structure with little $\alpha$-helical content compared to the other proteins in our set, which may account for their different rigidity behaviour. Previous work [21] has emphasized the analogy between the rigidity transitions of proteins and of glassy networks; we have now found that the two distinct patterns of rigidity transition recently identified in glassy networks [31] are also seen in proteins.

Our results in this paper suggest several avenues for further enquiry. The rigidity of protein monomers extracted from complexes should be systematically compared with their rigidity within the complex, which will be affected by interchain interactions. The robustness of flexible motion simulations based on rigidity analysis using different cutoff values must also be investigated. A recent study of the flexible motion of myosin [29] found that the flexible motion of the myosin structure appeared qualitatively similar over a wide range of cutoff values covering both highly flexible and more rigid structures. This suggests that rigidity analysis retains its value as a natural coarse graining for simulations even if the rigidity behaviour during dilution is as variable as we have found.

Acknowledgments

We gratefully acknowledge discussions with R Freedman, T Pinheiro, D Roper and M F Thorpe. We also gratefully
acknowledge financial support from the Leverhulme Trust (SAW and RAR, grant F/00 215/AH), BBSRC (JEI) and EPSRC (JEI and RAR, EP/CO07042/1). We would like to thank two anonymous reviewers for their valuable comments and perspective.

Appendix. Cutoff values in previous studies using FIRST

Jacobs et al [19] comment that the results of FIRST analysis should not be sensitive to the typical ‘fluctuations known to occur within protein structures’. Their advice is that the cutoff should be at least −0.1 kcal mol⁻¹ in order to eliminate a large number of very weak hydrogen bonds with energies in the range from 0.0 to −0.1 kcal mol⁻¹, and that a natural choice is near the ‘room temperature’ energy of −0.6 kcal mol⁻¹. As we have seen in section 3.2, this criterion is not sufficient to avoid sensitivity to small structural variations.

Rader et al [21] consider the protein folding transition by monitoring (r) (mean number of bonded neighbours per atom) during rigidity dilution; they do not, however, comment on the hydrogen bond energy values. Hespenheide et al [22] identify the protein folding core with ‘the set of secondary structure that remain rigid the longest in the simulated denaturation’, without regard to the exact values of the cutoff energy at which rigidity is lost. Here the cutoff energy is used qualitatively to distinguish weaker from stronger bonds. In considering the rigidity of virus capsid protein complexes, Hespenheide et al [20] make use of a cutoff of −0.35 kcal mol⁻¹, a value chosen so that capsid protein dimers would be flexible while the inner ring of proteins in a pentamer of dimers would be rigid, and draw conclusions about the rigidity of other multimeric complexes. Meanwhile, Hemberg et al [26] use a different cutoff of −0.7 kcal mol⁻¹ in a study on the dynamics of capsid assembly.

The FRODA geometric simulation algorithm [24] makes use of the RCD generated by FIRST as a coarse-graining. Simulations of protein mobility using FIRST/FRODA have tended to use cutoff values that are systematically lower than in applications of FIRST alone; typically −1 kcal mol⁻¹ or lower [24, 25, 27–29], as cutoff values closer to zero seem to include too many constraints to allow large-scale motion to occur. In a paper on the combination of rigidity analysis and elastic network modelling, Gohlke et al [30] discuss RCDs of two protein crystal structures but do not specify a cutoff value, though the FRODA mobility simulations given in figure 3(a) of that paper were performed using a cutoff of −1.5 kcal mol⁻¹ and give an excellent match to experimental data from NMR ensembles.

References

[1] Canino L S, Shen T and McCammon 2002 Changes in flexibility upon binding: application of the self-consistent pair contact probability method to protein–protein interactions J. Chem. Phys. 117 9927–33
[2] Case D A 1999 Molecular dynamics and normal mode analysis of biomolecular rigidity Rigidity Theory and Applications ed M F Thorpe and P M Duxbury (New York: Kluwer/Plenum) pp 329–44
[3] Bahar I, Atilgan A R and Erman B 1997 Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential Fold. Des. 2 173–81
[4] Ming D, Kong Y, Wu Y and Ma J 2003 Substructure synthesis method for simulating large molecular complexes Proc. Natl Acad. Sci. USA 100 104–9
[5] Ming D, Kong Y, Wakil S J, Brink J and Ma J 2002 Domain movements in human fatty acid synthase by quantized elastic deformational model Proc. Natl Acad. Sci. USA 99 7895–9
[6] Tama F, Wriggers W and Brooks C L III 2002 Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory J. Mol. Biol. 321 297–305
[7] Halle B 2002 Flexibility and packing in proteins Proc. Natl Acad. Sci. USA 99 1274–9
[8] Tirion M M 1996 Large amplitude elastic motions in proteins from single-parameter atomic analysis Phys. Rev. Lett. 77 1905–8
[9] Tama F, Gadea F X, Marques O and Sanejouand Y H 2000 Building-block approach for determining low-frequency normal modes of macromolecules Proteins 41 1–7
[10] Delarue M and Sanejouand Y H 2002 Simplified normal mode analysis of conformational transitions in DNA-dependant polymerases: the elastic network model J. Mol. Biol. 320 1011–24
[11] Petrone P and Pande V S 2006 Can conformational change be described by only a few normal modes? Biophys. J. 90 1583–93
[12] Vihinen M, Torkkila E and Rikonen P 1994 Accuracy of protein flexibility predictions Proteins 19 141–9
[13] Holm L and Sander C 1994 Parser for protein folding units Proteins 19 256–68
[14] Zefus M H and Rose G D 1986 Compact units in proteins Biochemistry 25 5759–65
[15] Karplus P A and Schulz G E 1985 Prediction of chain exibility in proteins Naturwissenschaften 72 212–3
[16] Maoirov V and Abagyan R 1997 A new method for modeling large-scale rearrangements of protein domains Proteins 27 410–24
[17] Flores M, Echols N, Milburn D, Hespenheide B M, Keating K, Lu J, Wells S A, Yu E Z, Thorpe M F and Gerstein M 2005 The database of macromolecular motions: new features added at the decade mark Nucl. Acid Res. 34 D296–301
[18] Jacobs D J and Thorpe M F 1995 Generic rigidity percolation: the pebble game Phys. Rev. Lett. 75 4051–4
[19] Jacobs D J, Rader A J, Kuhn L A and Thorpe M F 2001 Protein flexibility predictions using graph theory PROTEINS: Struct., Func. and Gen. 44 150–63
[20] Hespenheide B M, Jacobs D J and Thorpe M F 2004 Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus J. Phys.: Condens. Matter 16 S5055–64
[21] Rader A J, Hespenheide B M, Kuhn L A and Thorpe M F 2002 Protein unfolding: rigidity lost Proc. Natl Acad. Sci. 99 3540–5
[22] Hespenheide B M, Rader A J, Thorpe M F and Kuhn L A 2002 Identifying protein folding cores: observing the evolution of rigid and flexible regions during unfolding J. Mol. Graph. Model. 21 195–207
[23] Thorpe M F, Lei M, Rader A J, Jacobs D J and Kuhn L A 2001 Protein flexibility and dynamics using constraint theory J. Mol. Graph. Model. 19 60–9
[24] Wells S A, Menor S, Hespenheide B M and Thorpe M F 2005 Constrained geometric simulation of diffusive motion in proteins Phys. Biol. 2 S127–36
[25] Jolley C C, Wells S A, Hespenheide B M, Thorpe M F and Fromme P 2006 Docking of photosystem I subunit c using a constrained geometric simulation J. Am. Chem. Soc. 128 8803–12
[26] Hemberg M, Yaliraki S N and Barahona M 2006 Stochastic kinetics of viral capsid assembly based on detailed protein structures Biophysical J. 90 3029–42

[27] Jolley C C, Wells S A, Fromme P and Thorpe M F 2008 Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations Biophys. J. 94 1613–21

[28] Macchiarulo A, Nuti R, Bellochi D, Camaioni E and Pellicciari R 2007 Molecular docking and spatial coarse graining simulations as tools to investigate substrate recognition, enhancer binding and conformational transitions in indoleamine-2,3-dioxygenase (ido) Biochim. Biophys. Acta. 1774 1058–68

[29] Sun M, Rose M B, Ananthanarayanan S K, Jacobs D J and Yengo C M 2008 Characterisation of the pre-force-generation state in the actomyosin cross-bridge cycle Proc. Natl Acad. Sci. 105 8631–6

[30] Gohlke H and Thorpe M F 2006 A natural coarse graining for simulating large biomolecular motion Biophys. J. 91 2115–20

[31] Sartbaeva A, Wells S A, Huerta A and Thorpe M F 2007 Local structural variability and the intermediate phase window in network glasses Phys. Rev. B 75 224204

[32] Berman H M, Westbrook J, Feng Z, Gililand G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 The protein data bank Nucl. Acids Res. 28 235–42

[33] DeLano W The Pymol molecular graphics system http://www.pymol.org

[34] Word J M, Lovell S C, Richardson J S and Richardson D C 1999 Asparagine and glutamine: Using hydrogen atoms contacts in the choice of side-chain amide orientation J. Mol. Biol. 285 1735–47

[35] Dahiyat B I, Gordon D B and Mayo S L 1997 Automated design of the surface positions of protein helices Protein Sci. 6 1333–7

[36] Minary P and Levitt M 2008 Probing protein fold space with a simplified model J. Mol. Biol. 375 920–33

[37] Mamonova T, Hespenheide B, Straub R, Thorpe M F and Kumikova M 2005 Protein flexibility using constraints from molecular dynamics simulations Phys. Biol. 2 S137–47

[38] Gohlke H, Kuhn L A and Case D A 2004 Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach PROTEINS: Struct., Func. Gen. 56 322–37