CHAPTER 6

FLEXIBILITY IN BIOMOLECULES

M.F. Thorpe*, Mykyta Chubynsky, Brandon Hespenheide, Scott Menor
Physics & Astronomy Department, Arizona State University
Tempe, AZ 85287
**Email: mft@asu.edu*

Donald J. Jacobs
Department of Physics & Astronomy, California State University Northridge Northridge, CA 91330

Leslie A. Kuhn and Maria I. Zavodszky
Department of Biochemistry & Molecular Biology, Michigan State University East Lansing, MI 48824

Ming Lei
Department of Biochemistry, Brandeis University
Waltham, MA 02454

A.J. Rader
Center for Computational Biology & Bioinformatics
University of Pittsburgh, PA 15261

Walter Whiteley
Department of Mathematics & Statistics
York University, Toronto, Canada

In this chapter we review recent theoretical and computational work on the flexibility of biomolecules. This approach uses constraint theory and includes all the constraints in a biomolecule that are important at room temperature. A rigid region decomposition determines the rigid regions (both stressed and unstressed) and the flexible regions that separate them. Enzymes usually have a rigid core for stability and flexible regions for functionality. The rigid region decomposition can be used as input to for a Monte Carlo dynamics in which the flexible regions are allowed to move, consistent with the constraints. Results are illustrated with the proteins HIV protease and barnase.

1. Introduction

The relationship between the structural flexibility and dynamics of biomolecules and function is one of the key areas of research in modern biological science.

While structural flexibility has long been known to play an essential role in function, experimental and also computational data for many biomolecules has been slow in coming due to the difficulty in observing the motion over long time periods, especially for large systems. Drawing upon a range of disciplines, including biology, chemistry, physics, mathematics, and computer science, rigidity theory now offers a means to predict the flexibility inherent within a biomolecule, given only a single static three-dimensional structure. Knowledge of the rigid and flexible regions of a biomolecule can provide insight into its function, allow detailed exploration of the ensemble of conformations available for a given state, and provide a means to predict changes in structural flexibility as local environmental conditions such as temperature and pH change.

Advances in mathematical rigidity theory and computational algorithms, together with a representation of molecular forces as mathematical constraints, have resulted in the development of two programs, *FIRST* (Floppy Inclusions and Rigid Substructure Topography) and *ROCK* (Rigidity Optimized Conformational Kinetics). The *FIRST* software can decompose a static protein structure into rigid and flexible regions, and track changes in these regions during simulated thermal denaturation. The program *ROCK* extends the results of *FIRST* by exploring the ensemble of conformations accessible to the flexible regions in a structure, keeping the rigid regions stationary. Additionally, *ROCK* can be used in a directed dynamics mode to identify a conformational pathway between two distinct, known, structures of a single biomolecule. In the next two sections, we give examples of the results obtained by this approach.

2. Constraint Theory

For the last 30 years, there has been an evolving geometric and combinatorial theory of structural rigidity[1-3]. This work builds upon and extends a body of work spanning the last century, including that of James Clerk Maxwell[4], and a host of engineers studying the statics (and first-order kinematics) of bar and joint frameworks. Within this setting, it is useful to define a framework as a set of edges (E) and vertices (V) that define a simple graph (no loops or multiple edges) G = (V, E) and a configuration $\boldsymbol{p}$ of points for the vertices, together written as G($\boldsymbol{p}$). The edges of the graph represent distances between points that are constrained to remain constant in any transformation or motion of the framework.

A more abstract combinatorial theory of generic rigidity describes the rigidity properties of a given graph for ‘almost all’ configurations $\boldsymbol{p}$. Choosing a configuration at random will give the necessary generic behavior with probability 1, and these graphs, which lack any special symmetries, are classified as generic.

For realizations of a graph in the plane, there is a simple combinatorial criterion for identifying minimal graphs which are generically rigid. Known in the last century, but first proven by Laman[2] in 1970 - *A generic network in 2d with V sites and E bonds (defining a graph) does not have a redundant bond iff no subset of the network containing v sites and e bonds defining a subgraph violates* $e \leq 2v-3$. This criterion, in turn, leads to fast (almost linear) time combinatorial algorithms for decomposing a given graph into rigid components as well as predicting the degrees of freedom (dimension of the space of non-trivial motions) for the graph and its subgraphs. A clear implementation of this algorithm in 2D was suggested by Hendrickson[5], and subsequently developed into the *2D pebble game* algorithm by Thorpe and Jacobs[6]. Such a combinatorial algorithm for decomposing a graph into rigid and non-rigid subgraphs has several important advantages. It is fast and it is stable compared to the slower numerical evaluations of the rank of the rigidity matrices.

The underlying projective and combinatorial theory for representing molecular frameworks extends to a wider variety of structures with distance constraints as edges. One extension studies frameworks with larger 'bodies' and hinges (removing up to five degrees of freedom between bodies) as constraints, known as body-bar graphs[7-9]. There are several surprises in this extension. The biggest is that the theory of body-bar frameworks does have a good combinatorial theory leading to extensions of the *pebble game* in 3D. These are fast polynomial time algorithms for decomposing a generic 3D graph into rigid components, as well as the total degrees of freedom for non-rigid pieces. Essentially, this extension searches for six edge-disjoint spanning trees in a modified graph with vertices for the bodies and five edges for each hinge (corresponding to the five constraints imposed by the hinge). At a geometric level there is a complete first-order projective theory for these structures. [To date, this full blown projective theory has only been sketched in the literature[10], but the theory is well developed and should receive a proper exposition soon].

At the combinatorial level, the *Molecular Framework Conjecture* of Tay and Whiteley[8,10,12] is central to all our studies. There is overwhelming evidence for this conjecture, including comparing the combinatorial predictions and the actual rank computations for dynamical matrices. The conjecture, and the associated *3D pebble game*, has been verified explicitly for all structures studied with up to ~ 700 atoms, using costly matrix diagonalization techniques as a standard for comparison. The *pebble game* is an integer algorithm which is why it is so fast – handling a typical protein in less a minute in real time. These results use extensions of the projective and combinatorial techniques in references[7,8] and all

evidence points to the algorithms used in the *3D pebble game* being exact[13,14]. It would be irresponsible and also unnecessary to wait for a strict mathematical proof before proceeding with applications.

3. The Pebble Game and *FIRST*

The *3D pebble game* algorithm for frameworks is embodied in the *FIRST* software for measuring rigidity in biomolecular structures[15] and an example is shown in Figure 1. The fundamental step in the application is how to represent the microscopic forces in a molecule as distance constraints in a body-bar graph. For example, to fully model proteins, it is important to represent the prevalent and structurally crucial non-covalent interactions, in addition to the covalent bonds and angles. These non-covalent interactions and additional bond-rotational constraints include hydrogen bonds, salt bridges, hydrophobic tethers, and double bonds. With these representations, we have performed flexibility analyses for many proteins from different structural classes in their functional native state, including inter-domain hinge motions (lysine-arginine-ornithine binding protein and dihydrofolate reductase), loop or flap motions (HIV protease, cytochrome *c*, dihydrofolate reductase), and grip-like motions involving the concerted curling of multiple segments of the protein (adenylate kinase)[1,15-17]. These results indicate that the set of covalent and non-covalent interactions that we model as distance constraints is sufficient for reproducing the experimentally observed flexibility in these protein structures. Such information on native-state flexibility can be valuable, for example when studying enzyme mechanisms involving motion in parts of the protein.

An accurate methodology for computing native-state flexibility in proteins has an interesting extension to the prediction of protein folding pathways. This extension is based on the concept that as a protein folds, specific non-covalent bonds form, and remain formed throughout the remainder of the folding reaction. This suggests that the network of bonds in a native-state protein contains sub-networks corresponding to substructures formed along the folding pathway. Conversely, as a protein *un*folds, one might expect these early formed substructures to remain structurally stable the longest. We have simulated the effect of thermally induced unfolding on a protein by assigning an energy, or better a free energy change, to every non-covalent bond and then removing them from the flexibility calculation in order of energy, or free-energy. Changes in structural rigidity are tracked as the bonds are removed. The results of this simulation, referred to as *hydrogen bond dilution*, can be displayed graphically by mapping the rigidity results for the protein main chain onto a 1D line

(sequence). The results of hydrogen bond dilution for barnase are shown in Figure 2, where the shaded bars represent rigid main-chain bonds, and different shades indicate mutually exclusive rigid regions, often referred to as rigid clusters. The thin black line represents flexible main-chain bonds. This figure shows onset of flexibility as the protein is denatured, and also shows how the size of the largest rigid cluster, decreases until a point at which it fragments into small rigid clusters and flexible bonds. This result represents a plausible unfolding pathway for barnase, and as such, can be verified by comparison to experiment.

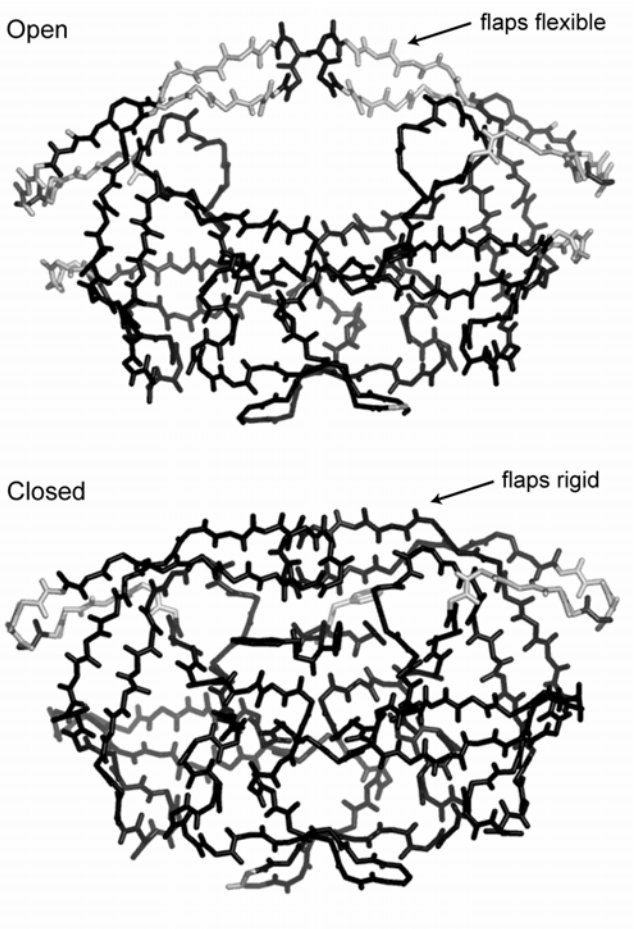

Figure 1. Showing a rigid region decomposition. The protein HIV protease is a dimer that consists of two polypeptide chains, shown on the left and right sides of each panel above, each containing 99 amino acids, and is an important part of the HIV virus. The open form shows the protease in the native state where it functions as a chemical scissors, whereas this function is inhibited in the closed form where the flaps are pinned against the inhibitor[11].The black regions are rigid.

For barnase, and a many other proteins, we have shown that those secondary structures that remain mutually rigid the longest during our denaturation simulation correspond well to the secondary structures that form the folding core[18,19]. Experimentally, the folding core is identified by using a combination of hydrogen-deuterium exchange and 2D NMR[20]. Such results of *FIRST* demonstrate that given only a single static protein structure, it is possible to obtain information about a dynamic process such as folding and predict the most stable regions of the protein. However, given that *FIRST* can also identify the flexible regions of a protein, it is possible to explore alternative conformations of a protein by moving the flexible regions in such a way that is consistent with the internal bond lengths and angles.

There have been significant studies in computational geometry using a variety of techniques, including extensions of first-order motions, to consider paths followed by flexible linkages. Some of this work has focused on small linkages connected with motion planning in robotics (or similar scaled biological problems such as the necks of birds). Other recent studies explore the 3D motions for a variety of structures, which include the equivalent of body-bar frameworks, or linkages for polymers and related structures[21]. These flexible linkage representations include polygonal chains, which are bonds linked end-to-end to form a linear polymer, and can include bond-coordination angle constraints coupled with dihedral rotations. In many cases, rather than generating algorithms for solving these problems in full generality, the complexity of the problems are confirmed, many of which are probably NP hard. Within the larger space of possible motions permitted by the basic constraints of bond lengths and angles, biochemists have used simulations that select appropriate paths by following an energy landscape to 'steer' the dynamic motion. These molecular dynamics simulations require potential-energy functions to apply classical physics equations of motion to the atoms in a molecule. While molecular dynamics remains the state-of-the-art for examining the details of molecular motion, the computational complexity involved and short time step limits its usefulness to examine the low-frequency modes associated with the large-scale motions of a biomolecule; motion that typically occurs in the microsecond to second regime.

The *3D pebble game* [6, 22, 23] algorithm, as currently implemented in *FIRST***,** can identify all the structurally rigid and flexible regions within a protein of hundreds of residues in a few seconds. Computationally, *FIRST* generates a directed graph of the covalent and non-covalent constraints present within the protein. This graph corresponds to a body-bar network of distance constraints that represent the physical forces present in the protein, and *FIRST* determines for every bond in this network whether it is part of a rigid cluster or an under-

constrained region composed of flexible bonds. In addition, the number of independent degrees of bond-rotational freedom, or floppy modes, associated with each under-constrained region is also determined. The concept of an independent flexible bond is important, as it is often found that the rotatable bonds in a flexible region are coupled to each other. For example, a ring of seven bonds is flexible, but contains only a single independent degree of freedom (DOF). Using *FIRST*, it is possible to determine whether a given flexible bond is independent, or part of a larger group of flexible bonds referred to as a *collective motion*. The conformational space available to the set of independent flexible bonds and collective motions can then be explored using the program *ROCK*.

The original *3D pebble game* algorithm as encoded in *FIRST* allows for only two types of bonds to be modeled, a rotatable bond or a non-rotatable bond, corresponding to a constraint with five bars or six bars respectively, within the body-bar formalism. This representation is sufficient for representing covalent bond networks, but has led to complications when attempting to model weaker, less-specific interactions such as those associated with the hydrophobic effect. This was rectified by replacing a single 5-bar constraint with a series of 5-bar constraints, which is explicitly included in the calculation through the use of pseudo-atoms. This modeling scheme reduces the impact of an interaction on network rigidity because each pseudo-atom introduces one more DOF to the system than the associated constraints remove (it is important to note that a series of pseudo-atoms still represents a single, real, microscopic interaction such as a hydrogen bond). Figure 3 depicts the impact a chain of pseudo-atoms has on the rigidity of a structure. For example, a single 5-bar constraint between a pair of atoms, as is used to represent a covalent bond, removes 5 degrees of freedom from the network (1 DOF is removed for each bar in the network). In contrast, a chain of 4 pseudo-atoms, shown at the bottom of Figure 3, removes only 1 DOF [the total loss of DOF comes from (4 pseudo-atoms x 6 DOF per pseudo-atom) minus (5 constraints x 5 bars per constraint) = –1 DOF]. Interestingly, a chain of 5 pseudo-atoms between atoms removes zero DOF from the system, and has absolutely no affect on the network rigidity. By taking advantage of the 4 different sized chains of pseudo-atoms shown in Figure 3, we can represent a range of microscopic forces, the physical interpretation being that a constraint removing the least number of DOF from a system has the least rigidifying effect on the structure.

While the results of the *3D pebble game* are not affected by the inclusion of pseudo-atoms, they have no physical interpretation. Additionally, because they

are explicitly included as part of the input structure, they increase the size of the system and hence the computation time

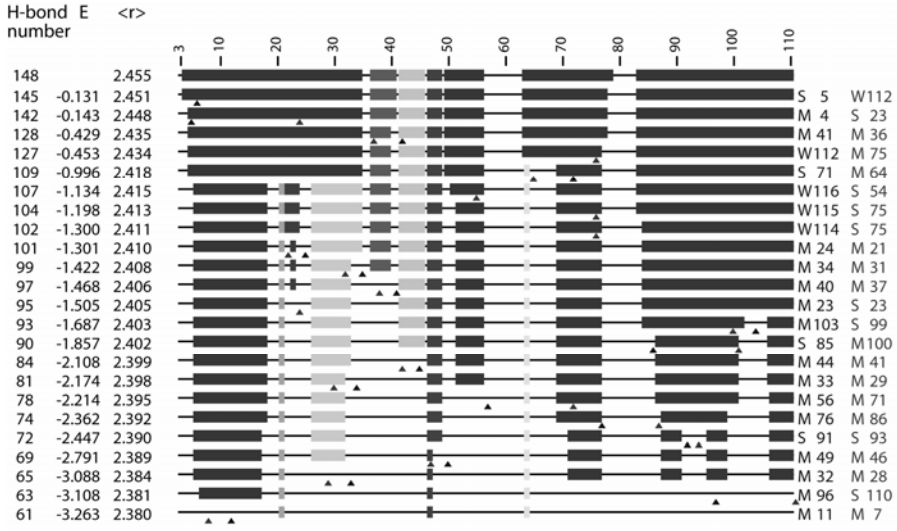

Figure 2. Hydrogen bond dilution results for barnase. Each line represents *FIRST* flexibility results for the main-chain bonds of barnase for a given concentration of hydrogen bonds (the total hydrogen bonds in the protein for a given line is listed in the first column on the right). The shaded bars indicate rigid bonds, and different shades represent different rigid clusters. Flexible bonds are shown as a thin black line. The top line represents the "native state" of the protein, and depicts a largely rigid structure consisting of a predominant rigid cluster. As stronger hydrogen bonds are removed from the structure (the energies of a given bond are listed in the second column on the left), the protein becomes more and more flexible. A key feature for this protein is the presence of a rigid core that persists until the protein becomes completely flexible.

We have recently eliminated the need for psuedoatoms due to a new implementation of the *3D pebble game* algorithm. This new algorithm allows for any number of bars between 1 and 6 to be placed between a pair of atoms, and this number directly corresponds to the degrees of freedom removed from the system. The mapping between a chain of psuedoatoms and the new 'fewer-bars' representation is shown in Figure 3. Not only is the new implementation faster, but it is more intuitive. While *more* psuedoatoms implies *less* affect on the rigidity, in the new algorithm the number of bars used to model a microscopic force is directly proportional to the effect on the network rigidity. The equivalence between a chain of pseudo-atoms and fewer bars has been demonstrated both mathematically, and in practice.

We have taken advantage of the new 3*D pebble game* algorithm to model the effect of hydrophobicity in protein structures. Hydrophobic interactions are known to contribute significantly to protein stability and are generally believed

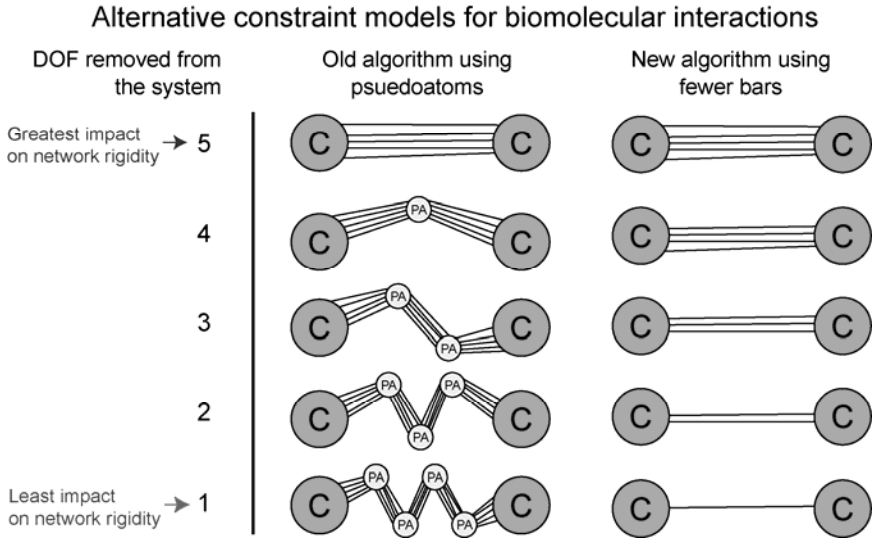

Figure 3. Modeling the affect of a carbon-carbon constraint on network rigidity. The more DOF a constraint or series of constraints removes, the more rigid the structure will be. Previously, pseudo-atoms (spheres) were included in a structure to attenuate the number of DOF removed by a bond between two atoms because the algorithm required 5 bars for every constraint. Non-intuitively, the higher the number of pseudo-atoms used, the smaller the affect on network rigidity. The new pebble game algorithm has been modified to allow 1-6 bars. Now, the number of bars directly corresponds to the number of DOF removed from the system. Not only does this have a clear physical interpretation, it is computationally less complex, reducing the run time for *FIRST*.

to be critical in driving the protein folding process[24]. Hydrophobicity reflects the tendency of the system to optimize entropy by folding hydrophobic groups that cannot form hydrogen bonds to the interior of the protein. This allows water molecules to associate randomly with each other and with polar groups on the protein surface, rather than being forced to become ordered relative to each other when presented with a hydrophobic surface of atoms. We model this tendency for hydrophobic atoms, principally carbon and sulfur atoms within proteins, to remain relatively near one another rather than unfolding to interact with the solvent. These hydrophobic contacts can be thought of as slippery, loosely constraining the local motion.

Hydrophobic interactions are identified geometrically between a pair of carbon and/or sulfur atoms[18]. The resulting constraint, which we refer to as a hydrophobic tether, is modeled with 2 bars. The decision to use 2 bars was the result of computing native-state flexibility and folding-core data in many proteins with an exhaustive sampling of geometric criteria and 4-, 3-, 2-, and 1-bar constraint representations. The 2-bar representation gave the best correlation to experiment. A physical interpretation of this 2-bar model is that it restricts the maximum distance between the two hydrophobic atoms, while allowing them to slide with respect to one another.

The most important of the microscopic forces that we model as constraints are hydrogen bonds and salt bridges. These non-covalent bonds are identified according to geometrical rules[16]. For PDB entries lacking polar hydrogen atom positions, the *What If* software package[25] is used to define hydrogen atom positions optimal for hydrogen bonding. Water molecules present in the input file are included in the analysis. In our recent research, we have only included water molecules if they are entirely buried within the structure, determined by the software *PRO_ACT*[26], as there is no mechanism in *FIRST* to identify these waters. Bonds between the protein and ligands, including metals and other ions, are treated as covalent bonds if so specified in the PDB file (or if they are within covalent bonding distance); otherwise, their polar and hydrophobic atoms are subject to the same rules as protein atoms for determining non-covalent interactions with the protein. Each potential hydrogen bond is assigned an energy using a modified Mayo potential, which evaluates the favorability of the bond based on a combination of distance and angular functions. We have modified the potential by strengthening the angular dependence on the donor–H–acceptor angle, so that it must be $\geq 120°$ for the bond to receive a favorable (negative) energy[27]. This avoids including non-physical H-bonds with angles near 90° (e.g., between C=O(i) and NH(i+3), rather than NH(i+4), in α-helices).

4. Dynamics using *ROCK*

While *FIRST* analysis can quickly identify flexibility in a biomolecule, one drawback is that atoms do not actually move – the analysis is based on statics, and so shows potential or virtual motion, rather than actual motion. While the atoms in the flexible regions have the possibility to move in various collective motions, *FIRST* does not give the amplitude of the motion. Some motions are restrained from having large amplitude by the constraints within a single flexible region, and the motion may also be restricted by collisions with adjacent regions.

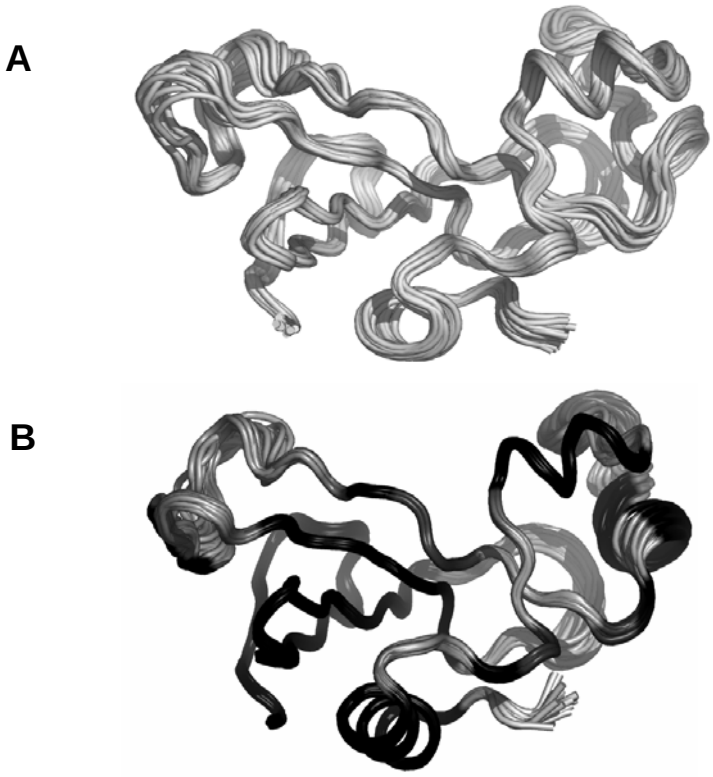

Figure 4. Flexibility in the native-state structure of barnase. The solid tubes represent the main chain of the protein. **A**. Result of *ROCK* dynamics using the X-ray structure **(1A2P)** of barnase[32]. The rigid regions are determined to be rigid from *FIRST* analysis, and as such, are kept fixed. The flexible regions are used to generate the 20 alternative conformations created by *ROCK* which are shown superimposed. Interestingly, the rigid helix on the right is flanked by two flexible regions, and *ROCK* explores a range of motion in which this helix moves as a rigid unit. **B**. Superposition of 20 conformers of barnase **(1BNR)** as determined by NMR spectroscopy[31]. The dynamic nature of NMR experiments results in many possible structures that fit the observed data. Those regions of the structure that overlap well imply little difference in the local conformation among the 20 structures. In contrast, those regions of the structure that can adopt many different conformations, as shown in the upper left portion of the panel, indicate structural flexibility. A comparison between the predicted structural flexibility in panel A and the experimentally observed flexibility shown in panel B reveals a good correspondence. The flexible regions in A correspond well to the thick regions in panel B, and the rigid regions in A align with the thin regions in panel B[29].

In order to clarify these points, the program *ROCK* (Rigidity Optimized Conformational Kinetics) has been developed. *ROCK*[28,29] uses *FIRST* as input and then makes Monte Carlo moves within the individual flexible regions, while maintaining ring closure. A typical flexible region, containing say 100 atoms,

will have very many rings, involving both covalent and non-covalent interactions. This is because rings consist of hydrogen bonds and hydrophobic tethers as well as covalent bonds, and so the rings themselves form a dense interlocking network. In addition to maintaining ring closure, the Monte Carlo moves must respect the hard spheres associated with the van der Waals radii, and also the Ramachandran constraints on the main-chain dihedral angles. After each attempted move in *ROCK*, the motions of adjacent flexible regions are checked for collisions. The conformations produced by *ROCK* show the diffusive motion of the protein [http://www.pa.msu.edu/ ~lei/Research/ *ROCK*/Proteins/HIV-1/HIV-1.html]. Thus the technique is complimentary to molecular dynamics which is good for time scales of up to a millisecond for small proteins. *ROCK* does not use a potential, and so does not give as accurate a picture as MD. However, by effectively freezing out the high frequency motions, which have little to do directly with biological function, the large diffusive motions can be visualized, giving the researcher a good sense of the possible motions. As an application, *ROCK* can be used to study directed dynamics by driving a protein from a known initial state towards a known target state, such as the ligand bound and unbound conformations of an enzyme. This allows a possible pathway(s) to be determined, through a set of conformationally allowed intermediate steps, which can be amplified later by more detailed optimization methods using realistic potentials.

ROCK can produce various conformations from a single static structure. This is useful when trying to find the flexible regions from X-ray crystallographic data when NMR data is not available. Using B-values alone is not very informative as no correlation information is obtained from the B-values which associate an amplitude with single atoms. Figure 4 shows a comparison of the flexibility of barnase using a single X-ray crystallographic structure (PDB: 1A2P) and *ROCK*, together with NMR data (PDB: 1BNR). The NMR figure is a superposition of the best 20 fits and contains sufficient measured constraints that the ‘spread” represents the flexibility of the protein.

It is important to ascertain such flexibility in order to understand function. The set of conformational states determined by *ROCK* is also useful in docking studies[29] which are facilitated by having a partially flexible interface because this adds a favorable entropic contribution to the free energy gain associated with the docking. Such studies are important in the early stages of finding drug candidates.

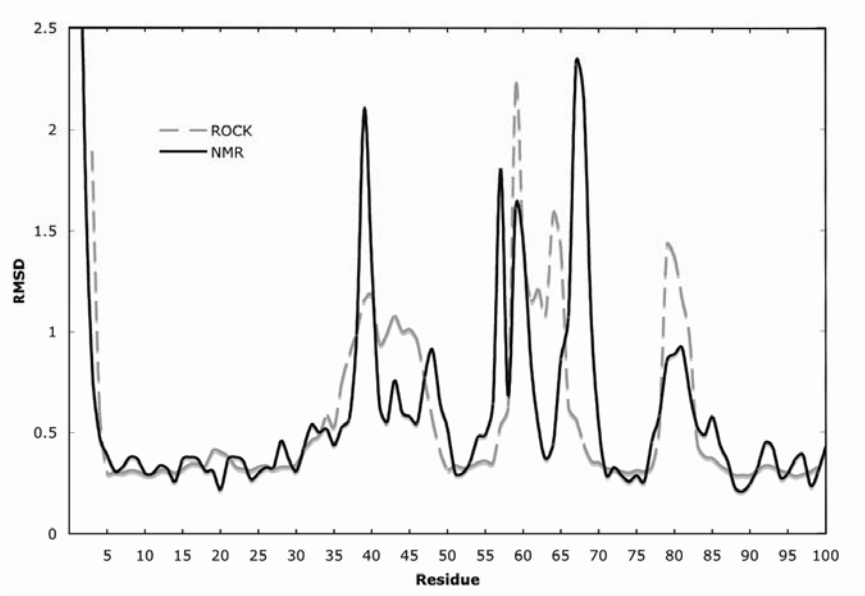

Figure 5. Showing a comparison for the two sets of aligned conformations shown in Figure 4 for barnase. The quantity RMSD is the root mean square deviation from the average structure in each case, measured in Angstroms, and plotted against residue number. The NMR structure (1BNR)[31] and X-ray structure (1A2P)[32] are from high resolution experiments. A background has been added to the *ROCK* results of an RMSD of 0.3Å which represents the high frequency motions, which are explicitly excluded from *ROCK*. To get the result displayed as *ROCK* above, the square deviations have been added[29].

5. Conclusions

In this brief review, we have outlined various methods that can be used to study flexibility and the associated motion in biomolecules. Much of the work here has been previously published, and so is only summarized. Full details can be found in the publications cited involving the authors of this paper in various combinations. The software used in this paper to study flexibility, including *FIRST* and *ROCK* can be found on the *Flex*web site at *Flex*web.asu.edu.

Acknowledgments

This work has been inspired by the approaches to science used by Sir Roger Elliott during a career in which he introduced many new methods and insights into the study disorder in non-crystalline systems. Such approaches have propagated through his many graduate students, like Mike Thorpe, and now

through the students of those students (grand-students) and recently some great-grand-students. The work described here is a natural evolution of the ideas and approaches that Roger has pioneered. It has been an honor to be asked to include this work in this volume that celebrates Roger's seventy-fifth birthday. This work was supported by the National Institutes of Health with grant R01 GM 67249-01.

References

1. Jacobs DJ, Rader AJ, Kuhn LA and Thorpe MF. 2001. *Protein flexibility predictions using graph theory*. Proteins, 44:150-65.
2. Laman G. 1970. *On graphs and rigidity of plane skeletal structures*. J. Eng. Mathematics, 4:331-340.
3. Graver J, Servatius B and Servatius H, Graduate Studies in Mathematics. 1993: Am. Math. Soc.
4. Maxwell JC. 1864. *On reciprocal figures and diagrams of forces*. Phil. Mag., 27:250-261.
5. Hendrickson B. 1992. Siam Journal of Computing, 21:65.
6. Jacobs DJ and Thorpe MF. 1996. *Generic rigidity percolation in two dimensions*. Physical Review. E. Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 53:3682-3693.
7. Whiteley W. 1988. *The union of matroids and the rigidity of frameworks*. SIAM J. Disc. Math., 1:237-255.
8. Tay T-S and Whiteley W. 1984. *Recent progress in the rigidity of frameworks*. Structural Topology, 9:31-38.
9. White N and Whiteley W. 1987. *The algebraic geometry of motions of bar and body frameworks.* SIAM J. Alg. Disc. Math., 8:1-32.
10. Crapo H and Whiteley W. 1982. *Statics of frameworks and motions of panel structures: a projective geometric introduction.* Structural Topology, 6:43-82.
11. Thorpe M.F.2003 *Protein Folding, HIV and Drug Design* Physics and Technology Forefronts. APS News, February.
12. Whiteley W. 1984. *Infinitesimal Motions of a Bipartite Framework.* Pacific Journal of Mathematics, 110:233-255.
13. Jacobs DJ. 1998. *Generic rigidity in three-dimensional bond-bending networks.* Journal of Physics a-Mathematical and General, 31:6653-6668.
14. Whiteley W, *Rigidity of molecular structures: generic and geometric analysis,* in Rigidity theory and applications, M.F. Thorpe and P.M. Duxbury, Editors. 1999, Academic/Kluwer. p. 21-46.
15. Jacobs DJ, Kuhn LA and Thorpe MF, *Flexible and rigid regions in proteins,* in Rigidity theory and applications, M.F. Thorpe and P.M. Duxbury, Editors. 1999, Academic/Kluwer. p. 357-384.
16. Thorpe MF, Hespenheide BM, Yang Y and Kuhn LA. 2000. *Flexibility and critical hydrogen bonds in cytochrome c*. Pac Symp Biocomput, 191-202.
17. Thorpe MF, Lei M, Rader AJ, Jacobs DJ and Kuhn LA. 2001. *Protein flexibility and dynamics using constraint theory.* J Mol Graph Model, 19:60-9.
18. Hespenheide BM, Rader AJ, Thorpe MF and Kuhn LA. 2002. *Identifying protein folding cores from the evolution of flexible regions during unfolding.* J Mol Graph Model, 21:195-207.
19. Rader AJ and Bahar I. 2004 *Folding core predictions from network models of proteins,* Polymer 655-668.
20. Li R and Woodward C. 1999. *The hydrogen exchange core and protein folding.* Protein Sci, 8:1571-90.

21. Soss M and Toussaint GT. 2000. *Geometric and computational aspects of polymer reconfiguration*. Journal of Mathematical Chemistry, 27:303-318.
22. Thorpe MF, Jacobs DJ, Chubynsky MV and Rader AJ, *Generic rigidity of network glasses, in Rigidity theory and applications*, M.F. Thorpe and P.M. Duxbury, Editors. 1999, Kluwer Academic. p. 239-277
23. Jacobs DJ and Thorpe MF. 1995. *Generic rigidity percolation: The pebble game.* Physical Review Letters, 75:4051-4054.
24. Dill KA. 1990. *Dominant forces in protein folding.* Biochemistry, 29:7133-55.
25. Vriend G. 1990. *WHAT IF: a molecular modeling and drug design program.* J Mol Graph, 8:52-6, 29.
26. Williams MA, Goodfellow JM and Thornton JM. 1994. *Buried waters and internal cavities in monomeric proteins.* Protein Sci, 3:1224-35.
27. Rader AJ, Hespenheide BM, Kuhn LA and Thorpe MF. 2002. *Protein unfolding: rigidity lost.* Proc Natl Acad Sci U S A, 99:3540-5.
28. Lei Ming, Kuhn LA, Zavodszky MI and Thorpe MF, 2004 *Sampling Protein Conformations and Pathways* Journal of Computational Chemistry **25**, 1133–1148
29. Zavodszky MI, Lei Ming, Thorpe MF, Day AR and Kuhn LA, 2004 *Modeling Correlated Main-Chain Motions in Proteins for Flexible Molecular Recognition* Accepted by Proteins: Structure, Function and Genetics
30. Menor S and Thorpe MF 2004. unpublished.
31. Bycroft M., Ludvigsen S., Fersht AR, Poulsen FM. 1991.*Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy*. Biochemistry **30** 8697.
32. Mauguen Y, Hartley RW, Dodson EJ, Dodson GG, Bricogne G, Chothia C, Jack A. 1982 *Molecular structure of a new family of ribonucleases*. Nature **297**, 162.