Hydrogen Bond Networks in Protein Flexibility and Folding

Brandon M. Hespenheide,[1,2] Leslie A. Kuhn,[1,2,3] and M. F. Thorpe [2]

1. Department of Biochemistry, Michigan State University, East Lansing, MI 48824 (http://www.bch.msu.edu)
2. Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824 (http://www.pa.msu.edu)
3. *MolFlex* L.L.C., East Lansing, MI. (kuhn@agua.bch.msu.edu)

Abstract I

Hydrogen bonds play a critical role in stabilizing the secondary and tertiary structures of proteins by linking residues into primarily rigid substructures. We have used *FIRST*, a graph-theoretical algorithm, to identify flexible versus stable regions in proteins based on the locking of main- and side-chain dihedral angles by hydrogen bonds. Here we test the hypothesis that the folded structure of a protein contains encod ed information about the folding pathway within its hydrogen-bond net work. After hydrophobic collapse, the strongest hydrogen bonds are likely to form first and nucleate secondary and tertiary structures. These H-bonds will be harder to break and rearrange than weak ones, which are expected to play a role in subsequent folding and stabilization. This hypothesis is tested by incrementally removing the H-bonds in a folded protein, from weakest to strongest, as would happen for example if the temperature were raised. The order in which the rigid secondary and ter tiary substructures form, defined by the change in the bond network as a function of the strengths of H-bonds included, is compared to the order in which substructures form in hydrogen-exchange (HX) NMR experi ments. We correlate these results with HX NMR folding data for cyto chrome *c*. This analysis of protein H-bond networks also provides a use ful technique for identifying key hydrogen bonds crosslinking a protein.

Introduction II

Protein flexibility is an important structural aspect of protein function and folding. A few examples are:

- Molecular motors such as dynamin and actin:myosin
- Catalysis in HIV protease, dihydrofolate reductase, and ATP synthase
- Formation of stable secondary structures, such as -helices during protein folding

Here we present an analysis of protein flexibility and rigidity performed using a novel graph-theoreti cal algorithm, *FIRST* (Floppy Inclusion and Rigid Substructure Topography)[1,2,3]. We have used *FIRST* to identify the flexible and rigid regions in proteins, and to show the importance of the hydrogen bond network in protein flexibility and folding.

- *FIRST* represents a protein as bond network or 3-dimensional graph.
- The bond network contains all the structural information of the protein, such as fixed bond lengths and angles.
- *FIRST* quickly and accurately identifies rigid regions in proteins connected by flexible joints.
- The output of *FIRST* is a Rigid Cluster De composition (RCD) that describes the loca tion and size of the rigid and flexible regions in the protein.

Methods: Hydrogen Bond Dilution III

1. Hydrogens are added to the coordinates of all polar atoms in the protein structure using *WhatIf*.
2. Water molecules in the structure that are at least 50% buried are included in the structure.
3. *FIRST* calculates the covalent, hydrogen and salt-bridge bond network for the protein.
4. Each H-bond in the protein is assigned an energy according to Mayo, *et al.*[4], and H-bonds with an energy ≤ 0.1 kcal/mol are kept.
5. The RCD for the protein is calculated with all the H-bonds included. Then, thermal denaturation of the protein is simulated by removal of the weakest (highest energy) H-bond in the protein, and the RCD is recalculated.
6. RCD's are displayed in a 1-dimensional (1D) format (Figure 1) whenever removal of an H-bond causes a change in the number or size of main-chain rigid clusters.

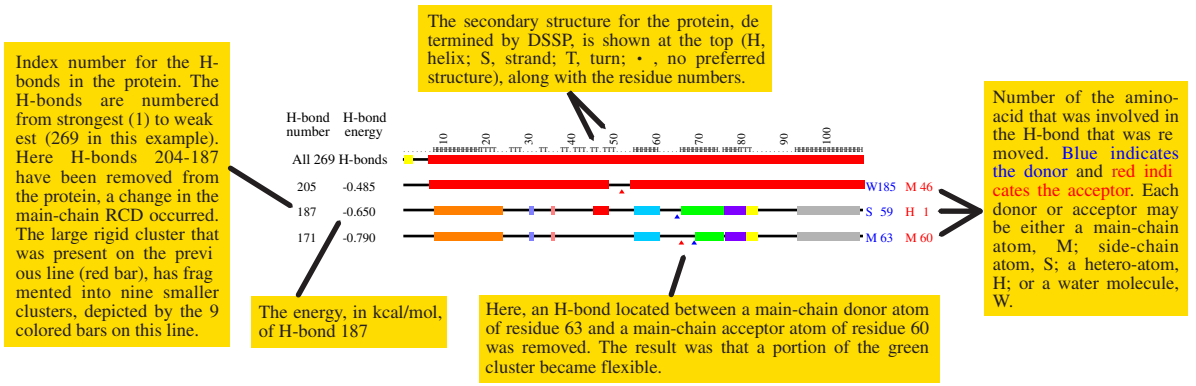

Figure 1. Describing how to interpret the 1D RCD's resulting from H-bond dilution in a protein. Rig id clusters are shown as colored bars. The main-chain is represented as a thin black line.