

Rigidity-dilution behaviour of proteins.

Background: the crystal structure of a protein can be viewed as a network of bodies (atoms) subject to constraints (covalent bonds, hydrogen bonds, and hydrophobic interactions). Rigidity analysis (Jacobs and Thorpe 1995; Jacobs *et al.* 2001) then indicates that portions of the structure are highly constrained (rigid) while other are flexible, in the sense that covalent bonds are free to rotate. The rigidity of the structure depends on the set of constraints that are included. Specifically, the hydrogen bonds in the structure can be assigned an estimated energy based upon their geometry (Thorpe *et al.* 2001); hydrogen bonds are included only if their energy lies below a specified cut-off value, and the rigidity of the structure is a function of that cutoff. By systematically varying the cutoff from a small value (so that almost all possible hydrogen bonds are included as constraints) to a large (so that almost all hydrogen bonds are eliminated), we can obtain a "dilution plot" which indicates how the rigidity varies as hydrogen bonds are removed. Previous work (Rader *et al.* 2002) has indicated that such a dilution plot can provide information on the folding and unfolding of the protein, by identifying a "folding core" where rigidity persists longest during the dilution.

Most studies using this rigidity analysis approach have been performed on relatively small, globular proteins (a notable exception is a study of the assembly of a virus capsid, Hesperheide *et al.* 2004). For such proteins, the loss of rigidity during dilution is generally gradual. Initially the protein forms a single large rigid cluster; this cluster shrinks as constraints are removed and eventually splits into two clusters, which in turn shrink and split as the cutoff changes. For proteins containing hinges (which allow large-scale opening/closing motions, as for example in adenylate kinase) a division into multiple rigid clusters is of course necessary to allow the hinge motion to occur. In some cases, however, for example in some antibodies (Dr. Wells, personal observation), the loss of rigidity is far more sudden, the structure transforming from a single rigid block to complete flexibility within a very narrow range of cutoff values. This might, for example, indicate a difference between enzymatic proteins, whose function requires a specific, repeatable motion, and antibodies whose function is to have a particular shape and display a particular binding motif. It appears that no systematic study on this subject has been performed.

The tools required for a systematic investigation of the rigidity-dilution properties of different classes of proteins are readily available and easy to use. Dr. Wells has experience in this type of investigation and can provide training and guidance to the student. This miniproject that would be well within the capabilities of an intelligent MSc student and would almost certainly lead to a publication. **If the student is interested in pursuing the work further, there is scope for a PhD project studying the flexible motion of proteins and its dependence on rigidity.**

Programme of work: we propose that the student should: (i) assemble a set of proteins on which to perform rigidity analysis, with examples from different classes of proteins; (ii) prepare the structures for analysis and perform rigidity dilutions; and (iii) analyse the results in terms of the variation of rigidity with cutoff value. Classes of proteins to examine would include globular, enzymatic proteins such as lysozyme, adenylate kinase, dihydrofolate reductase; transport proteins such as cytochrome C, myoglobin; motor and sensory proteins such as myosin, rhodopsin, GFP; antibodies and immunoglobins; fibrous or structural proteins such as collagen, actin, virus capsid proteins; and any others that the student may identify as interesting cases. Structures are obtainable online from the RCSB protein databank (PDB) (<http://www.rcsb.org/pdb>); general criteria for selection would be that the structure should be obtained from X-ray crystal diffraction at good resolution (e.g. better than 2.5 Angstroms), as NMR structures and low-resolution crystal structures are generally not suitable for rigidity analysis.

Hydrogen atoms (which are generally not resolved in X-ray structures) can be added using free tools on the MolProbity website (<http://molprobity.biochem.duke.edu>). Rigidity analysis itself can be carried out using the software FIRST (<http://flexweb.asu.edu>). Dr. Wells is experienced in the use of MolProbity and FIRST and will provide training for the student. These steps require no more than a few minutes per structure. Analysis of the results will involve quantifying how rigidity varies during dilution. Several metrics suggest themselves, including: (i) the fraction of the main chain that lies within the largest rigid cluster; (ii) the fraction of the main chain that lies within any large rigid cluster (i.e. larger than a covalently rigid unit); (iii) the number of large rigid clusters in the protein. The behaviour of these metrics during dilution will reveal whether or not different classes of proteins have characteristically different rigidity properties. Visualisation of structures during rigidity dilution is best performed using the PyMOL software (<http://pymol.sourceforge.net>) as FIRST outputs scripts for this programme.

Skills to be learned: the student will become familiar with the use of the RCSB Protein Data Bank as a resource to obtain protein structures and literature references, and the MolProbity website to assess the quality of a structure and add hydrogens to X-ray crystal structures. This knowledge will be useful in any study involving protein structures. The student will learn the basis of rigidity theory in its application to proteins and will become an experienced user of FIRST. This complements conventional approaches such as molecular dynamics and elastic-network modelling, as well as providing a natural coarse-graining for protein structures (Gohlke and Thorpe 2006) that is useful in multi-scale modelling (Wells *et al.* 2005) and provides valuable insight into the protein structure-

function relationship (Costa and Yaliraki 2006).

Resources required: the PDB and MolProbity sites are freely accessible online, and the FIRST software is free to download for academic use. Rigidity information from FIRST is visualised using the PyMOL viewing software; a free version is available online. FIRST runs under Linux; the student will therefore need access to an account on a Linux machine. This can be provided by CSC. A small consumables budget will be required for office supplies, e.g. printing.

Literature review: The references provided in this proposal will provide the basis for the literature review. The starting reference for the student will be "*Protein unfolding: rigidity lost*" by Rader *et al.* (PNAS 99, 2002). The paper itself describes the use of the FIRST software for rigidity analysis and the dependence of the rigidity on the hydrogen-bond energy cutoff value. Its cited references (42 in number) cover background material, in rigidity theory and in protein crystal structure studies, while the papers that cite it (65, as of September 2007) provide a survey of the applications of FIRST to different protein structures.

References:

1. Costa and Yaliraki, "Role of rigidity on the activity of proteinase inhibitors and their peptide mimics", *Journal of Physical Chemistry B* 110 pp. 18981-18988, 2006.
2. Gohlke and Thorpe, "A natural coarse-graining for simulating large biomolecular motion", *Biophysical Journal* 91 pp. 2115-2120, 2006.
3. Jacobs and Thorpe, "Generic rigidity percolation- the pebble game", *PRL* 75 pp. 4051-4054, 1995.
4. Jacobs et al., "Protein flexibility predictions using graph theory", *Proteins- structure, function and genetics* 44 pp. 150-165, 2001.
5. **Rader *et al.*, "Protein unfolding; Rigidity lost", PNAS 90 pp. 3540-3545, 2002.**
6. Thorpe et al., "Protein flexibility and dynamics using constraint theory", *Journal of Molecular Graphics and Modelling* 19 pp. 60-69, 2001.
7. Wells et al., "Constrained geometric simulation of diffusive motion in proteins", *Physical Biology* 2 pp. S127-S136, 2005.

Consumables budget: Not more than £50 for office supplies.