As published in BTi April/May 2006

Prediction of protein flexibility from geometrical constraints

by Daniel Seeliger and Dr Bert L. de Groot

Protein function is in many cases coupled to molecular flexibility. However prediction of protein flexibility with molecular dynamics simulations is computationally demanding. The CONCOORD method provides an efficient way to predict the "essential" degrees of freedom of a protein from a given 3-dimensional structure.

A protein is a many-body system with 3N degrees of freedom (where N is the number of atoms in the molecule). In most cases the greatest biological interest concerns only the low fre-

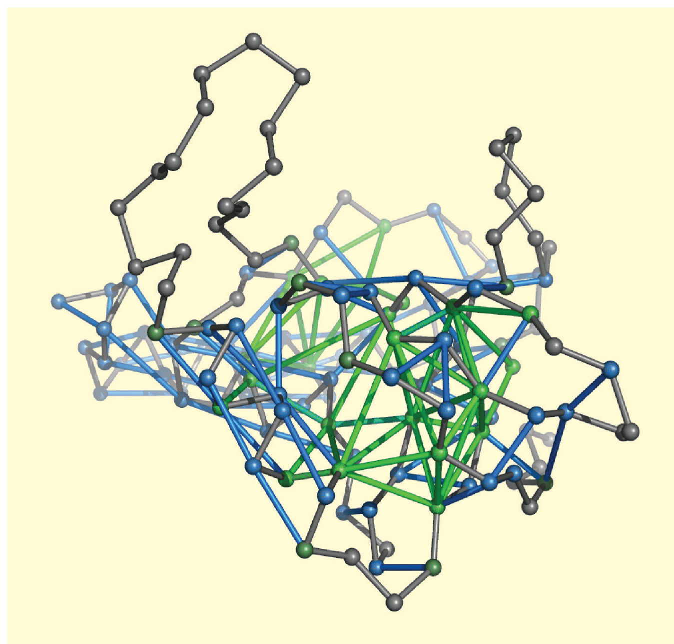

Figure 1. Geometrical constraints in staphylococcal nuclease (PDB 1ey4). Besides covalent connections, residues are connected via hydrogen bonds (blue) and interactions in the hydrophobic core (green). For clarity, amino acids are represented by their Cα-atom.

quency motions, the so called "essential" degrees of freedom [1, 2], which typically appear in timescales of nanoseconds to microseconds and play important roles in signalling, activation and enzyme function. Despite the rapid growth of computer power these timescales are not accessible with current molecular dynamic simulations systems. The CONCOORD system attempts to alleviate this sampling problem by predicting pro-

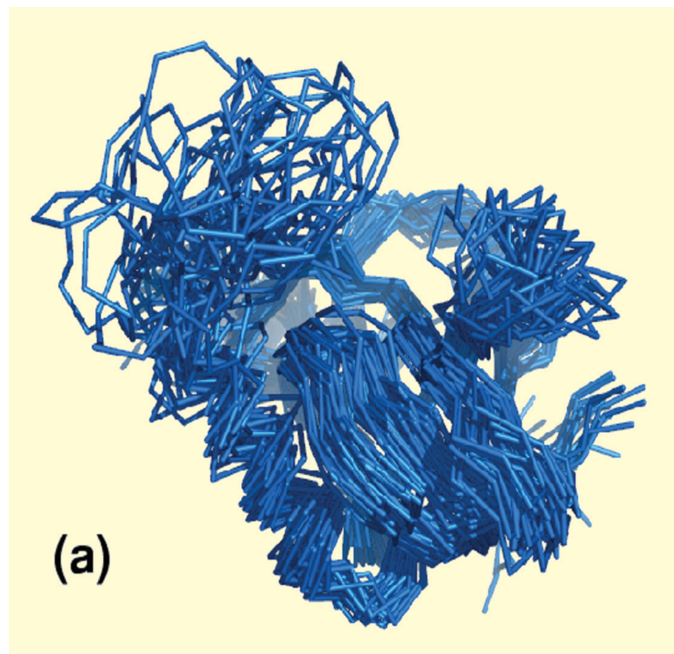

Figure 2. A CONCOORD ensemble of staphylococcal nuclease, computed from PDB 1ey4.

tein flexibility from geometrical constraints, thereby efficiently crossing local barriers in the rugged potential energy landscape.

The starting point is a 3D-structure of the protein. This structure is analysed for interactions which are translated into a set of geometrical constraints, based on which an ensemble of new structures is generated. The "essential" degrees of freedom can be obtained by analysing the generated ensemble using principal component analysis.

Geometrical Constraints

The 3-dimensional structure of a protein is determined by various interactions, such as covalent bonds, hydrogen bonds and the hydrophobic effect. CONCOORD performs a careful analysis of the input structure and evaluates interactions. Figure 1 shows a schematic diagram of interactions within a protein structure. The hydrogen bond framework (blue) and the hydrophobic core restrict the flexibility of the peptide chain (grey). Some regions of a protein structure are kept tightly together by a large number of interactions, whereas others, such as loop regions, which often function as binding sites, display

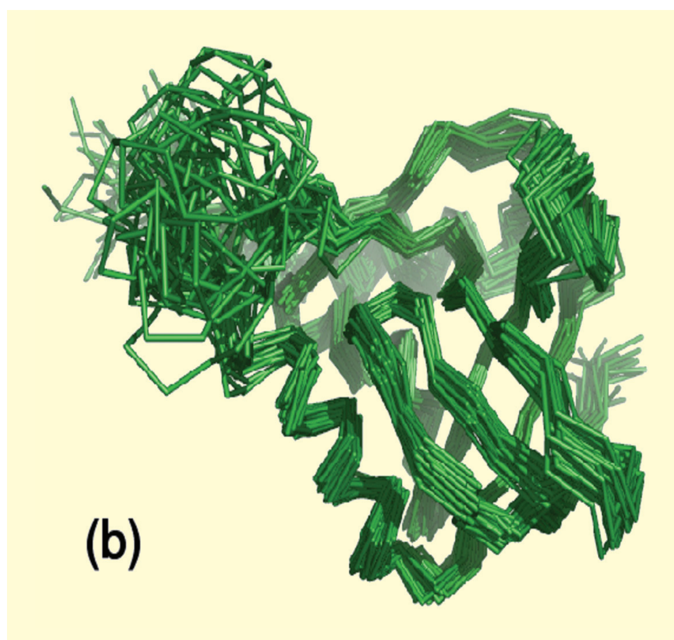

Figure 3. NMR ensemble of staphylococcal nuclease (PDB 1jor).

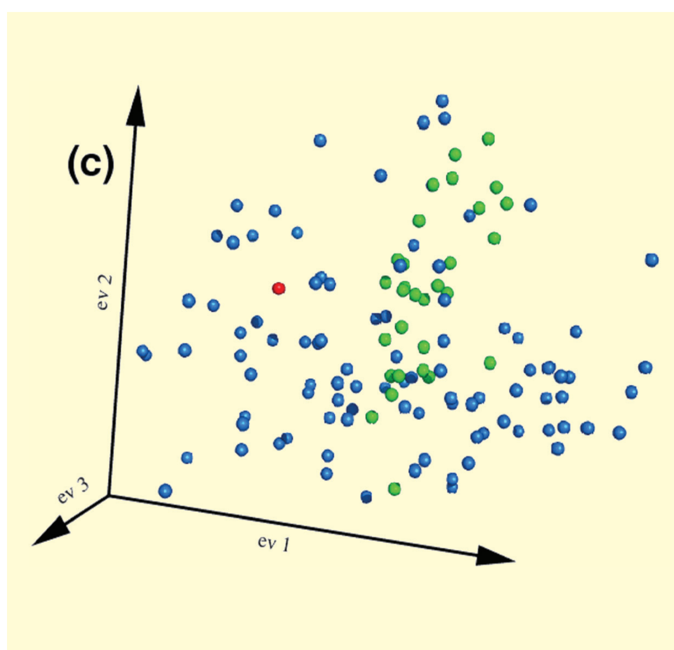

Figure 4. Principal components analysis was carried out on the CONCOORD ensemble. The axes represent the three eigenvectors with the largest eigenvalues and correspond to the main collective motions of the protein. Each dot represents a single structure. The red dot represents the X-ray structure (1ey4) that was used as input for CONCOORD. Blue dots represent CONCOORD structures and the green dots belong to the NMR ensemble.

enhanced flexibility. CONCOORD turns this information into geometrical constraints with upper and lower bounds, depending on the strength of the interaction, leading to a detailed atomic description of the protein comparable to a construction plan.

Structure Generation and Analysis

In a second step, CONCOORD uses the pre-defined constraints to generate a structure ensemble. The CONCOORD kernel is a SHAKE - based algorithm that iteratively corrects distances until the pre-defined constraints are fulfilled [3]. Starting from random coordinates guarantees complete independence between subsequently generated structures. Usually, a set of a few hundred structures suffices for a converged coverage of the available conformational space. Thus, the method does not suffer from sampling problems inherent to other techniques such as MD-simulations.

Depending on the size of the protein and the chosen parameters, an ensemble of several hundred structures can be obtained within a few hours or days. Figure 2 shows a CONCOORD ensemble that is generated from the structure PDB 1ey4 (staphylococcal nuclease). For comparison an NMR-ensemble of the same protein is shown in Figure 3. This shows that there is a favourable agreement with the experimentally measured flexibility. Highly ordered regions show little flexibility whereas the large loop appears in different conformations.

Structure ensembles are commonly analysed by principal component analysis (PCA). This is based on diagonalisation of the covariance matrix of atomic fluctuations; it extracts the "essential" degrees of freedom from an ensemble and allows a quantitative comparison of different ensembles along the major modes of collective fluctuation.Figure 4 shows a comparison of the CONCOORD ensemble and the NMR ensemble of staphylococcal nuclease. The PCA was carried out on the CONCOORD ensemble and both ensembles were projected on the first three eigenvectors that correspond to the three largest collective motions. The plot shows that the CONCOORD ensemble, started from the X-ray structure (red), samples the complete conformational space that is accessible by the NMR ensemble.

Application to Large Systems

Because of its computational efficiency, CONCOORD can be routinely applied to extract functionally relevant modes of flexibility for molecular systems that are beyond the size limitations of other atomistic simulation techniques such as molecular dynamics simulations. Application of the system to the chaperonin GroEL-GroES complex that contains

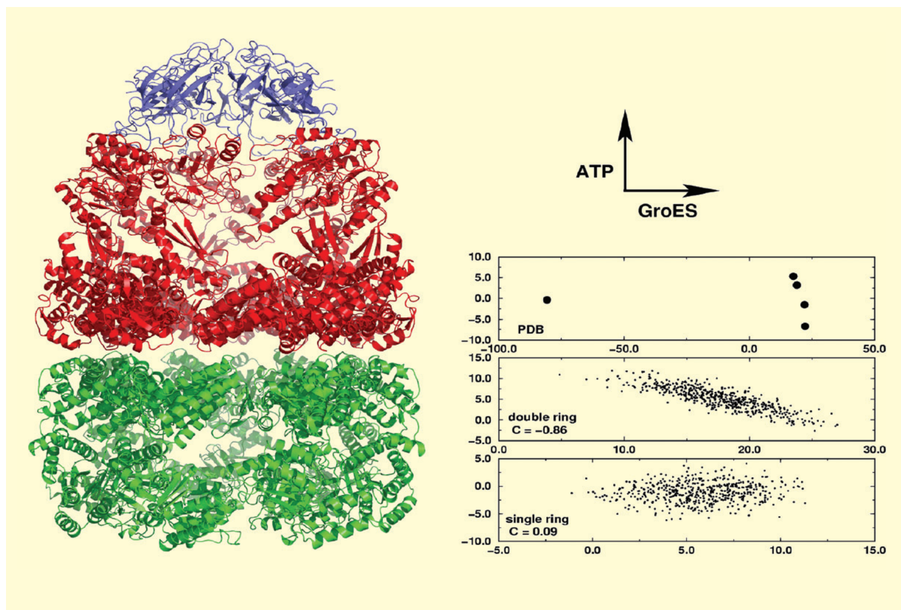

Figure 5. Asymmetric GroEL-GroES complex (left), together with CONCOORD simulation results (right). The co-chaperonin GroES is shown in blue, the trans ring of GroEL, bound to GroES, in red, and the cis-ring in green. A principal component analysis revealed two main structural transitions per GroEL ring, upon nucleotide binding (vertical axis in the right panels) and GroES binding (horizontal axis), respectively. In simulations of the double ring, but not in a single ring, these modes were found to be coupled, suggesting a coupling between intra-ring and inter-ring cooperativity.

more than 8000 amino acids revealed a novel form of coupling between intra-ring and inter-ring cooperativity [4]. Each GroEL ring displays two main modes of collective motion: the main conformational transition upon binding of the co-chaperonin GroES (coloured blue in Figure 5), and a secondary transition upon ATP binding [Figure 5, upper right panel]. CONCOORD simulations of a single GroEL ring did not show any coupling between these modes, whereas simulations of the double ring system showed a strict correlation between the two modes, thereby providing an explanation for how nucleotide binding is coupled to GroES affinity in the double ring, but not in a single ring.

References

1. de Groot B L, van Aalten DMF, Scheek RM, Amadei A, Vriend G and Berendsen HJC. PROTEINS: Structure, Function and Genetics 1997; 29: 240-251.

2. Amadei A, Linssen ABM and Berendsen HJC, PROTEINS: Structure, Function and Genetics 1993; 17: 412-425.

3. Ryckaert J-P, Cicotti G, Berendsen HJC. J Comput Chem 1977; 23: 327-341.

4. de Groot BL, Vriend G and Berendsen HJC. J MolBiol 1999; 286: 1241-1249.

The authors

Daniel Seeliger, Dipl. Chem.
Bert L. de Groot, Ph D.,
Computational biomolecular dynamics group,
Max-Planck-Institute for Biophysical Chemistry,
D-37077 Göttingen,
Germany.