Robust and efficient analysis of biomacromolecular stability using ensembles of random network topologies

Christopher Pfleger, Holger Gohlke

Computational Pharmaceutical Chemistry, Institute of Pharmaceutical and Medicinal Chemistry, Heinrich Heine-University, Universitätsstr. 1, 40225 Düsseldorf

Biomacromolecules require a balance of flexibility and rigidity to achieve their diverse functional roles. Hence, it is desirable to have a precise knowledge about what can move and how. The flexibility for a given biomacromolecule can be analyzed within a few seconds by a graph theory-based approach as implemented in the FIRST program (*F*loppy *I*nclusions and *R*igid *S*ubstructure *T*opography)[1]. Previous studies have demonstrated that the approach is sensitive with respect to the structural information used as input [2,3]. This sensitivity problem can be overcome by analyzing an *ensemble of network topologies* rather than a single structure network. To do so, MD-generated conformations were used so far. This way, however, the efficiency of the FIRST approach is lost.

Here, we will present a more efficient alternative where an ensemble of network topologies is generated by fluctuating non-covalent constraints in a constraint network derived from a single input structure, with hydrogen bonds, salt-bridges, and hydrophobic tethers varied. This approach has been implemented into the Constraint Network Analysis (CNA) program package developed in our group. The CNA program functions as a front-end to the FIRST software and allows to I) set up a variety of constraint network representations for rigidity analysis, II) process the results obtained from FIRST, and III) calculate different indices for characterizing macroscopic and microscopic stability in biomacromolecules.

The approach of fluctuating networks was validated on a dataset of 38 lysozyme crystal structures with high structural quality as well as on four MD trajectories of lysozyme[4]. Remarkably, in almost all cases of the 38 lysozyme crystal structures, the predicted flexibility/rigidity characteristics are in good agreement with results obtained from ensembles of network topologies derived from the four MD trajectories. Each calculation requires less than 5 h of computing time on a standard workstation computer. The approach shall thus be valuable when it comes to investigating many proteins, e.g. in the case of analyzing the effect of mutations on protein stability.

[1] Jacobs, D.J., *et al.*, Proteins (2001), **44**(2): p. 150-65.
[2] Gohlke, H., *et al.*, Proteins: Struct., Funct., Bioinf. (2004), **56**: p. 322-37.
[3] Mamonova, T., *et al.*, Phys. Biol. (2005), 2: p. S137-47.
[2] Koller, A.N., *et al.*, Biophys. J. (2008), **95**(1), L04-6.