Applications of a New Program for the Reconstruction of Protein Envelopes from Solution Scattering Data

John Badger

Delta G Technologies, 4360 Benhurst Ave, San Diego, CA 92122, USA, info1.dgtech@gmail.com

The protein molecular envelopes that are reconstructed from x-ray solution scattering data may be used: (*i*) to match the shapes of proteins in solution to structures determined using x-ray crystallography, (*ii*) to model oligomeric forms or large proteins with the structures of individual protein units or known subdomains, (*iii*) to discover conformational forms that may occur in solution but which are not amenable to capture in crystal samples. For all these purposes we seek the most reliable reconstruction of the molecular envelope possible, containing the maximum appropriate level of detail. Although deviations in the protein shapes derived from solution scattering data from crystallographic models may indicate functionally important conformational variants, discrepancies may also result from artifacts, noise or over-interpretation of reconstructions.

A new program for the convenient, efficient and reliable reconstruction of protein molecular envelopes from solution scattering data, SHAPES, has been developed with these concepts in mind. The SHAPES program represents the protein volume using point scattering centers that interact via a modified 6-12 Lennard-Jones potential. Compared to programs that model protein shapes using points on a preset grid, the productive use of intensity data is not limited to the low-q regime. Furthermore, since with SHAPES the parameterization of interactions between scattering centers is not directly based on the concept of 'dummy amino acids' the number of scattering centers used to model the protein volume need not be equal to the expected number of amino acids in the target structure. For this reason, applications to large structures, that include many amino acids, does not require a very large numbers of scattering centers and are not limited by excessive run times. Many trials of SHAPES with simulated and real data show that accurate and definitive molecular envelopes are recovered across a wide range of structure types.

Concerns with solution scattering methodologies for the reconstruction of protein molecular volumes include the extent to which the resulting molecular envelopes represent unique solutions and the characterization of error in the reconstructions. As noise in the intensity data increases, replicated structure solutions runs typically produce increasingly dissimilar solutions compatible with the data and this ambiguity complicates comparisons with crystal structures. Using SHAPES, several strategies have been evaluated for utilizing multiple reconstruction solutions and comparing them to atomic models.