The SHAPES Program for the Reconstruction of Protein Envelopes from X-ray Solution Scattering Data

J Badger[1]
[1]Consultant, San Diego, CA
info1.dgtech@gmail.com

SHAPES is a computer program intended for the convenient, efficient and reliable reconstruction of protein molecular envelopes from x-ray solution scattering data (J.Badger, J. Appl. Cryst. (2019) 52, 937-944). The program represents the protein volume using a set of volume-filling elements ('beads') that interact with each other via a modified 6-12 Lennard-Jones potential. A Monte Carlo reconstruction process drives the beads to coalesce into a space equal to the expected partial specific volume of the target protein while maintaining a relatively uniform packing density. The final set of bead positions fits the pair-distribution function derived from x-ray scattering data. Convenience features built into the program include the capability of performing multiple reconstruction trials within a single run and the output of ready-to-use molecular volumes. SHAPES is written in python and is compatible with python 2 and 3 interpreters. It is freely available as an open source software under a GNU GPLv3 software license from http://saxs2shapes.com. In addition to providing transparency in program operation, the availability of source code allows for local adaptation, extension and experimentation with use and application of reconstruction methodologies. Although relatively fast when compared to most other available programs for ab initio envelope reconstruction, considerable acceleration may be achieved by adapting repetitive array-based operations to employ functions available in the numpy module (R. Von Dreele, personal communication). Trials of SHAPES with simulated and real data show that accurate and definitive molecular envelopes may be recovered from a wide range of structure types. However, reconstructions of protein envelopes from x-ray solution scattering data are inherently less reliable when the shape of the target protein or protein assembly is highly elongated or flattened. The difficulty in determining correct solutions from these types of sample manifests itself in the appearance of inconsistent solutions over replicate trials and in nonphysical solutions, where the volume-defining beads are broken into many clusters. Recent work with SHAPES includes an option for providing an additional interaction term between beads that tends to prevent them from dissolving into sub-clusters. Application of this option permits the reconstruction of protein shapes from some of the more problematic examples.