11. Visualization and clustering of Protein Local Conformational Space using Geometric Invariant Theory

Ashish V. Tendulkar, Pramod P. Wangikar

Structures of peptide fragments from a protein can potentially occupy a vast conformational continuum. Here, we propose a general framework to visualize the protein local conformational space using geometric invariants as structure descriptors. We observe that the number of preferred local conformations is far less than that predicted previously.

Protein local conformations have conventionally been classified into alphahelix, Betastrand and loop. Helix and strand are characterized by their regularity in their backbone torsion angles while loops can potentially occupy a vast conformational continuum. Ramchandran plot was the first step in demarcating the feasible and infeasible regions of conformational space even for loops (1). More recently, loops have been systematically classified based on structural similarity (26).

Several of these loop classification methods begin by looking at the secondary structural elements that flank a loop on the two sides. Thus, the loops that join two betastrands are classified separately from those that join an alphahelix and a betastrand.
Regardless of the method of classification, loop classification has important implications in interpreting the electron density maps or in protein structure modeling (7, 8).
We have previously shown that the protein conformation space is biased in favor of a finite number of conformations (9). Here we present a visual map of the restricted protein local conformational space using geometric invariant theory. We have selected octapeptide as a unit of protein local conformation and represented each octapeptide with its Calpha geometry (10). We have drawn approximately 1.7 million overlapping octapeptides from ASTRAL95 dataset, version 1.67 (11). Each octapeptide is described with a suite of geometric invariants (9) followed by dimension reduction via Principal Component Analysis. Note that the closeness in the geometric invariant space guarantees that the two structures are superimposable without having to compute the superimposing transform(12). The conformational space is then visualized in the form of conditional bivariate probability distribution plots, which contain peaks of varying size, corresponding to the different preferred conformations. The peak corresponding to alphahelix is sharper and taller than that of betastrand. A Separate peak is identifiable for the kinked helix and several others for loops. The octapeptide fragments were subjected to kmeans clustering to detect clusters of similar geometry. The clusters and peaks in the conditional bivariate distribution plots share a one to one correspondence. We observe that the number of preferred local conformations is far less than that predicted previously.
It has been previously reported that the protein local conformation space is highly restricted; however, visualization of the conformational space has remained a challenge. Conventional methods of pairwise comparison and alignment of available protein structures are inadequate for the task of visualization of conformational space.

Unilateral representation of the local conformations using geometric invariants followed by dimension reduction via principal component analysis allowed us to achieve the visualization. The conditional bivariate distribution plots provide a visualmap of allowed and disallowed protein conformations. The peak size in conditional bivariate distribution indicates likelihood of the structure in a randomly selected natural protein. The method presented here can have applications in protein structure

prediction and validation. The current protein structure prediction algorithms search a vast protein conformational space using a computationally expensive energy minimization protocol. Visualizing the allowed and disallowed regions in the conformational space provides a useful method for eliminating the disallowed conformations with significant savings in computational time. Moreover, the peak size in the distribution is indicative of the likelihood of the structure
occurring in a randomly selected natural protein. This can be useful in checking the integrity of both predicted and experimentally deduced structures.

References:

1. G. N. Ramchandran, Ramkrishnan, C., Sasisekharan, V., Journal of Molecular Biology 7, 9599 (1963).
2. E. James MilnerWhite, R. Poet, Trends in Biochemical Sciences 12, 189192 (1987).
3. B. L. Sibanda, T. L. Blundell, J. M. Thornton, Journal of Molecular Biology 206, 759777 (1989/4/20, 1989).
4. B. L. Sibanda, J. M. Thornton. (Academic Press, 1991) pp. 5982.
5. J. Wojcik, J.P. Mornon, J. Chomilier, Journal of Molecular Biology 289, 14691490 (1999/6/25, 1999).
6. R. T. Wintjens, M. J. Rooman, S. J. Wodak, Journal of Molecular Biology 255, 235253 (1996/1/12, 1996).
7. M. B. Swindells, J. M. Thornton, Current Opinion in Biotechnology 2, 512519 (1991/8, 1991).
8. M. W. MacArthur, R. A. Laskowski, J. M. Thornton, Current Opinion in Structural Biology 4, 731737 (1994/10, 1994).
9. A. V. Tendulkar, A. A. Joshi, M. A. Sohoni, P. P. Wangikar, Journal of Molecular Biology 338, 611629 (2004/4/30, 2004).
10. T. J. Oldfield, R. E. Hubbard, ProteinsStructure Function and Genetics 18, 324337 (Apr, 1994).
11. S. E. Brenner, P. Koehl, R. Levitt, Nucleic Acids Research 28, 254256 (Jan 1, 2000).
12. A. V. Tendulkar, P. P. Wangikar, M. A. Sohoni, V. V. Samant, C. Y. Mone, Journal of Molecular Biology 334, 157172 (2003/11/14, 2003).