Atomic density distributions in proteins: structural and functional implications

Sotirios Touliopoulos, Nicholas M. Glykos

Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, Alexandroupolis, 68100, Greece

Abstract

Atomic density is the number of atoms per Å^3 (atoms/Å^3). In a protein structure it is a measure of proximity between atoms. A protein's atomic density distribution shows how well packed is a structure and it may include information on potentially identifying proteins with special folding patterns. We examine the atomic density distributions derived from 21,304 protein structures and show that statistically significant differences between those distributions are present. Current efforts focus on identifying putative patterns connecting the structure and function of those proteins with their corresponding atomic density distributions.

Methods

- 21304 protein entries were downloaded from the PDB with a list obtained from the culling server PISCES3 with the criteria: maximum 2.2 Å resolution, maximum R-factor 0.25 and 50% identity cut–off.
- Hydrogen atoms were added using the riding model as implemented in the program OpenBabel.
- A cutoff for proteins with more than 50 aminoacids was applied.
- All computational scripts are done in Python & R programming languages.

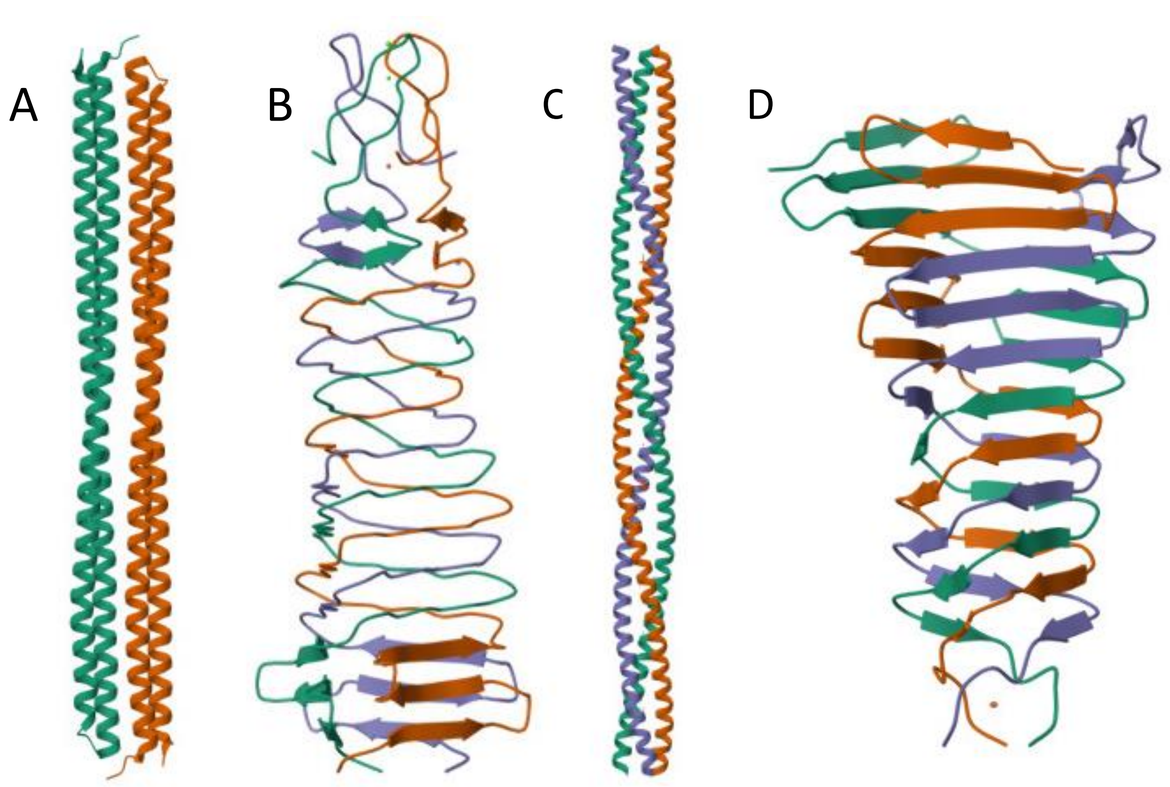

Figure 3: Structures from proteins with uncommon atomic density distributions.
(A) Dimeric Coiled-coil protein (PDB ID: 4KE2). (B) Beta-sheet protein (PDB ID: 3QR7).
(C) Trimeric Coiled-coil protein (PDB ID: 2QIH). (D) Beta-sheet protein (PDB ID: 4S36)

Introduction

Atomic density is calculated by dividing the number of atoms in a hypothetical sphere, with the volume of the sphere. Each sphere comes with a certain radius in Angstrom values (Å), with the atom's position as the center. A superposition of all 21304 atomic density distributions, together with their corresponding mean and associated standard deviations are shown in Figure 1. Several protein structures deviate significantly and systematically from the average behavior and —not unexpectedly— include proteins with characteristic structures such as elongated coiled-coils and beta-helices (PDB entries: 4KE2, 3QR7). These proteins can be seen as outliers in Figure 1. Their corresponding structure appears in Figure 3. Hierarchical clustering of the atomic density distributions (Figure 2) indicated that a far distinct cluster occurs. It supports the existence of a protein group with uncommon atomic density distributions. Search for persistent patterns in this cluster's proteins might be an indication for structural implication of atomic density.

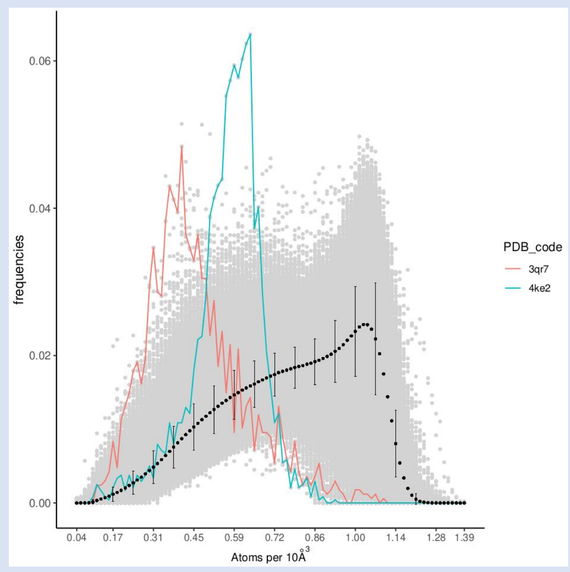

Figure 1: Raw Data from superimposed distributions

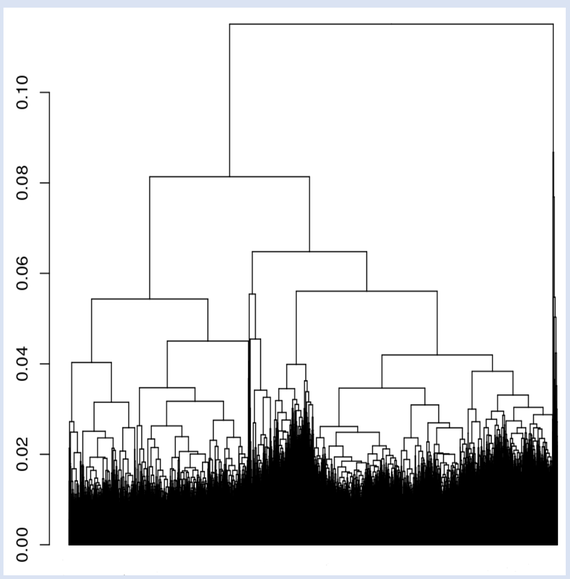

Figure 2: Hierarchical Cluster Dendrogram

Future Work

We aim to access gene ontology (GO) terms of the proteins classified together and check whether a functional implication appears. Different clustering methods (kmeans, hdbscan), ways to compute distance between distributions (euclidean, pearson), application of different radius cutoffs (6Å–9Å) and calculation of density based on atomic weight will remove bias from our approach and enhance validity.

References

[1] Sonavane S, Chakrabarti P (2008) Cavities and Atomic Packing in Protein Structures and Interfaces. PLOS Computational Biology 4(9): e1000188

[2] Jernigan RL, Kloczkowski A. Packing regularities in biological structures relate to their dynamics. Methods Mol Biol. 2007;350:251-76. doi: 10.1385/1-59745-189-4:251. PMID: 16957327; PMCID: PMC203970

[3] G, W. & RL, D. PISCES: a protein sequence culling server. Bioinformatics 19, 1589–1591 (2003

[4] O'Boyle, N.M., Banck, M., James, C.A. *et al.* Open Babel: An open chemical toolbox. *J Cheminform* **3**, 33 (2011). https://doi.org/10.1186/1758-2946-3-33

Contact Information

Sotirios Touliopoulos, email: s.v.touliopoulos@gmail.com , Nicholas M. Glykos, email: glykos@mbg.duth.gr
Structural and Computational Biology Laboratory, Dept. of Molecular Biology and Genetics, DUTH