Proceedings of the ***Conference of the Hellenic Society for Computational Biology and Bioinformatics (HSCBB23)***

Thessaloniki, Greece, 10-12 October 2023

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

∗ Correspondence to: s.v.touliopoulos@gmail.com, glykos@mbg.duth.gr

Abstract—**Atomic density is the number of atoms per Å^3 (atoms / Å^3). It 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. Atomic density in a protein structure is a measure of proximity between protein's 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. In this preliminary report we examine atomic density distributions derived from 21.304 protein structures and show that statistically significant differences between those distributions are present. Several protein structures deviate significantly and systematically from the average behaviour and —not unexpectedly— include proteins with characteristic structures such as elongated coiled-coils and beta-helices (PDB entries: 4ke2, 2qr7). These proteins can be seen as outliers in Figure 1. 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. Current efforts focus on identifying putative patterns connecting the structure and function of those proteins with their corresponding distributions. We aim to access gene ontology (GO) terms of the proteins classified together and check whether a functional implication appears too. 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**

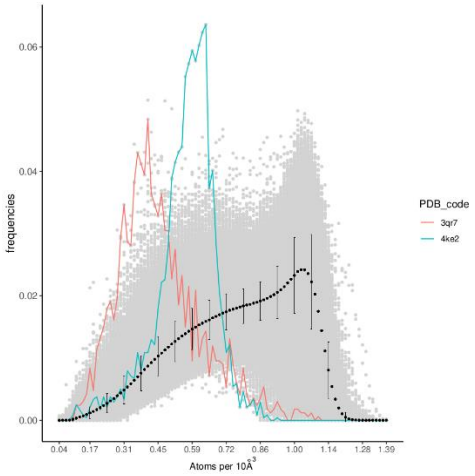

Figure 1. Raw data and outliers of atomic density distirbutions

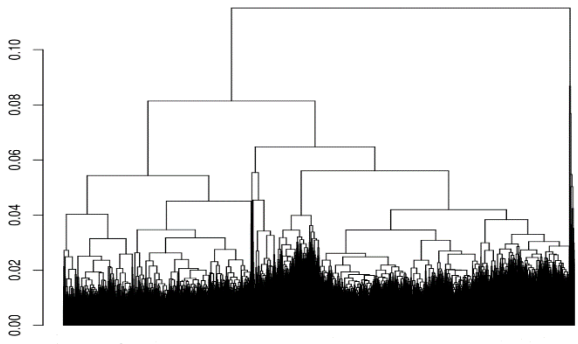

Figure 2. Hierarchical Cluster Dendrogram

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: PMC2039702