Biochemical and Biophysical Research Communications 436 (2013) 725–729

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Atom depth analysis delineates mechanisms of protein intermolecular interactions

Davide Alocci [a], Andrea Bernini [a], Neri Niccolai [a,b,*]

[a] Department of Biotechnology, Chemistry and Pharmacy, University of Siena, via A. Fiorentina 1, 53100 Siena, Italy
[b] SienaBioGrafix Srl, via A. Fiorentina 1, 53100 Siena, Italy

ARTICLE INFO

Article history:
Received 30 May 2013
Available online 17 June 2013

Keywords:
Protein–protein interactions
Structural layers
3D atom depth
Protein surface charge
Protein–protein network

ABSTRACT

The systematic analysis of amino acid distribution, performed inside a large set of resolved protein structures, sheds light on possible mechanisms driving non random protein–protein approaches. Protein Data Bank entries have been selected using as filters a series of restrictions ensuring that the shape of protein surface is not modified by interactions with large or small ligands. 3D atom depth has been evaluated for all the atoms of the 2,410 selected structures. The amino acid relative population in each of the structural layers formed by grouping atoms on the basis of their calculated depths, has been evaluated. We have identified seven structural layers, the inner ones reproducing the core of proteins and the outer one incorporating their most protruding moieties. Quantitative analysis of amino acid contents of structural layers identified, as expected, different behaviors. Atoms of Q, R, K, N, D residues are increasingly more abundant in going from core to surfaces. An opposite trend is observed for V, I, L, A, C, and G. An intermediate behavior is exhibited by P, S, T, M, W, H, F and Y. The outer structural layer hosts predominantly E and K residues whose charged moieties, protruding from outer regions of the protein surface, reorient free from steric hindrances, determining specific electrodynamics maps. This feature may represent a protein signature for long distance effects, driving the formation of encounter complexes and the eventual short distance approaches that are required for protein–protein functional interactions.

1. Introduction

Interactions among the molecular components of Life determine a huge variety of biochemical events hosted by Nature. Nowadays, the available structural information of biomolecules is large enough to contain already many relevant clues for deciphering, at the atomic level, mechanisms of biological processes. Indeed, all the structural features that are required for stabilizing protein adducts with nucleic acids, small molecules or other proteins, are already well known and updated by the continuous growth of information stored in the Protein Data Bank, PDB [1].

At the end of last century, a System Biology perspective has been added to Structural Biology and molecular mechanisms of protein–protein interactions, PPI, have been investigated with experimental and computational approaches. Thus, the formation of encounter complexes has been proposed as a preliminary step for PPI [2–5] also under conditions of macromolecular crowding [6].

It is apparent that, in the molecular crowd typical of biological fluids, the formation of encounter complexes cannot be driven simply by self-diffusion processes, suggesting that a protein–protein networking must be present. Electrostatic assistance through long range interactions to drive protein associations has been proposed for hub proteins [2], that is for proteins exhibiting multiple PPI, and the presence of high surface charge has been suggested as the main source of their enhanced social activity [7].

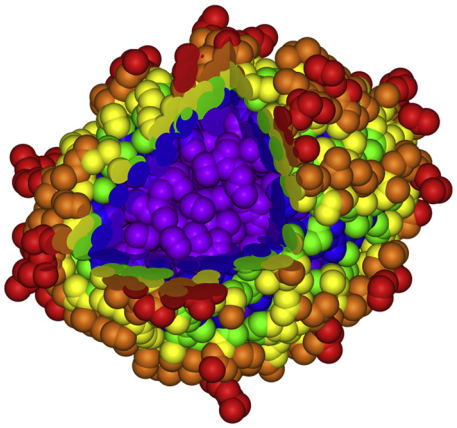
In order to find possible mechanisms which determine non-random translations between proteins and their eventual ligands, in the present report we have performed a systematic analysis on the population and distribution of amino acid residues on the surface of unbound proteins. Thus, to remove possible biases induced by protein–ligands complexation from all the available PDB protein structures, a reduced dataset of only protein singles, DOOPS, has been built.

To have a quantitative assessment of amino acid composition of protein surfaces, we used our original 3D atom depth analysis, similarly to what has been recently done to define protein cores [8]. A parameter dubbed depth index, D_i, has been used as a tool to classify DOOPS atoms into different structural layers. Recurrent presence of specific amino acids on protruding moieties of protein surface hosted by the outer structural layer has been here analyzed

* Corresponding author at: Department of Biotechnology, Chemistry and Pharmacy, University of Siena, via A. Fiorentina 1, I-53100 Siena, Italy. Fax: +39 (0) 577 234903.

E-mail addresses: alodavide@gmail.com (D. Alocci), andrea.bernini@unisi.it (A. Bernini), neri.niccolai@unisi.it (N. Niccolai).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.06.024



L_n	L_0	L_1	L_2	L_3	L_4	L_5	L_6
D_i	< 0.2	0.2 - 0.4	0.4 -0.6	0.6 – 0.8	0.8 – 1.0	1.0 – 1.2	> 1.2
% atoms	17,9	17,1	17,7	17,4	14,4	9,4	5,7
color	violet	indigo	blue	green	yellow	orange	red

Fig. 1. Spacefill representation of protein structural layers: rainbow coloring of the seven atom layers defined on the basis of atom depth indexes for the cyanobacterial bicarbonate transport protein, CmpA, ID PDB code 2I49, and the used D_i limits are shown in the inset. Atom percent presence in each structural layer of DOOPS proteins are also given. The same octants of L_1–L_6 structural layers have been removed to show L_0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

to explore specific mechanisms of long range protein–protein interactions.

2. Materials and methods

2.1. The DOOPS and sDOOPS datasets

The entire content of the Protein Data Bank has been scanned with a final updating on May 8, 2013 when 90,424 structures were available. In order to take into account only proteins with well defined secondary structure elements, only crystal structures have been selected with the following limitations provided by the PDB advanced search interface: (i) homologue removal at 95% identity, (ii) only one chain in the asymmetric unit, (iii) only proteins with monomeric biological assembly and (iv) only one entity in the asymmetric unit. Among all these restrictions, the last one considerably reduces the number of proteins in the dataset, as most of protein PDB structures refer to protein–ligand complexes. Thus, only 2410 PDB structures are included in the dataset of only protein singles, DOOPS, where only proteins whose surface shape is minimally influenced by molecular neighbors different than water should be present. In Fig. S1, we show a comparison between the content of DOOPS and the whole 2013_05 release of Swiss-Prot. In spite of the huge difference in the number of objects contained in the two datasets, they report very similar amino acid populations and protein chain length distributions, suggesting that data derived from DOOPS are statistically significant. In order to obtain a subset of DOOPS containing only proteins with a relevant involvement in PPI, DOOPS has been further refined by choosing in the PDB advanced search interface a PubMed selection with the following syntax: protein–protein interaction AND structure AND X-ray. Thus, 321 protein structures were obtained; their relevance in PPI has been manually checked through the STRING server at the URL: http://string-db.org. It turned out that 84% of the selected proteins are actually involved in PPI and 29% with more than 10 other protein partners. Then, these 321 structures define a DOOPS subset of social proteins, sDOOPS.

2.2. Atom depth calculations

Atom depth for all the 4,657,574 atoms contained in DOOPS have been calculated with the SADIC algorithm [9], by using the freely downloadable software at http://www.sbl.unisi.it. Atom depths are calculated as depth indexes, D_i, defined as:

$$D_i = 2V_i/V_0 \quad (1)$$

where V_i is the exposed volume of a sphere of radius r_0 centered on atom i and V_0 is the exposed volume of the same sphere when centered on an isolated atom. As shown in Fig. S1, DOOPS protein se-

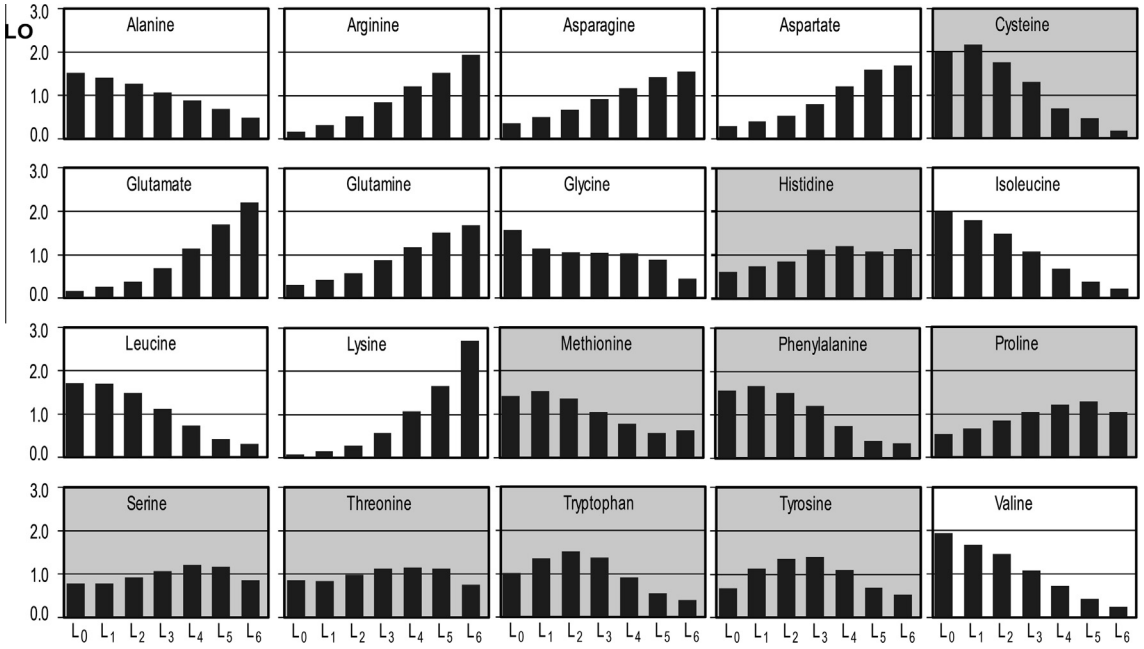


Fig. 2. Amino acid occupancy of structural layers: structural layer occupancy, LO, of the twenty natural amino acids normalized by their overall occurrence in DOOPS. Pale grey backgrounds highlight amino acids not having monotonic content changes at increasing L_n indexes.