A database for comparative analysis of paradigms for prospecting contacts in protein-protein interfaces

Martins, P. M. [1], Mayrink, V. D. [2], Silveira, S. A. [3], da Silveira, C. H. [4] , Lima, L. H. F. [5] , Melo-Minardi, R. C. [1]

[1] Department of Computer Science, UFMG, Belo Horizonte, Brazil, [2] Department of Statistics, UFMG, Belo Horizonte, Brazil, [3] Informatics Department, UFV, Viçosa, Brazil, [4] Advanced Campus at Itabira, UNIFEI, Itabira, Brazi, [5] Campus at Sete Lagoas, UFSJ, Sete Lagoas, Brazil.

Computing contacts in proteins is important to several types of studies from Bioinformatics to Structural Biology. An accurate computation of contacts is essential to the correctness and reliability of applications involving folding prediction, protein structure prediction, quality assessment of protein structures, network contacts analysis, thermodynamic stability prediction, protein-protein and protein-ligand interactions, docking and so forth. In this work, we built a large database of contacts using about 45,000 PDB files to compare three paradigms for contacts prospection at atomic level: distance-based only, distance and geometric-based (occlusion free) and distance and angulation-based.

The main contribution of this work is a critical evaluation of three different paradigms that can be used to compute contacts between protein atoms. We focused on protein-protein interfaces (multiprotein complex) and analysed four types of contacts, namely hydrogen bonds, aromatic stackings, hydrophobic and ionic (attractive) interactions. We scanned for possible contacts in the range from 0 to 7 Å. Our data showed the importance of a geometric approach to filter out spurious occluded contacts after about 3.5 Å for aromatic stackings, hydrophobic and ionic interactions. For hydrogen bonds the angulation criteria presented more reliable results at every distance in the considered interval. Furthermore, we find a recommended limits of distances and paradigms for each type of interaction considering that approximately 95% of such interactions are in the respective interval.

We provide the database with all computed contacts and the source codes used to populate such database. These resources are available at http://homepages.dcc.ufmg.br/~pmartins/capri/.

This work was supported by the Brazilian agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).