CavPred 1.0.0

Theoretical Background

Authors: Anna Basquet, Juana Muñoz, Ester Muñoz
Courses: SBI, PYT 2022/2023

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Contents

Abstract

CavPred 1.0.0 is a geometry-based method developed for the detection of ligand binding sites using grid and Voronoi methods. This program is written in Python and employs several algorithms such as grid-based clustering, density-based clustering, and Voronoi diagrams to identify cavities in protein structures that may potentially bind ligands. The grid method is used to partition the protein structure into a 3D grid, which is then clustered based on the density of grid points. On the other hand, the Voronoi method involves dividing the space around each atom in a protein into regions based on their distance from the atom. These regions are then used to create a Voronoi diagram, which can be used to identify potential binding sites. By combining these two methods, CavPred 1.0.0 is able to identify more accurate binding sites and provide a more complete picture of the ligand-protein interactions.

The advantages of geometry-based methods for the detection of ligand binding sites are numerous. Firstly, they are fast and efficient, making them ideal for high-throughput screening of protein structures. Secondly, they are not reliant on experimental data, which can be expensive and time-consuming to obtain. Thirdly, they are able to detect cavities that are not visible by other methods, such as those that are partially buried or have irregular shapes. Finally, these methods can provide insights into the structural and functional characteristics of proteins, which can be useful for drug discovery and protein engineering applications.

Introduction

Proteins are some of the most important elements for life and play a crucial role in numerous vital processes within living organisms by engaging in interactions with other molecules. Identifying the specific residues involved in these interactions is not only important for gaining a deeper understanding of protein function, but it also holds significant implications for the development of new drugs, agrochemical design, cancer mechanisms, drug formulation, and physiological regulation (1,2). Taking into account that traditional experimental methods for detecting such binding sites are costly and time-consuming, making computational approaches a more effective solution, numerous computational techniques for predicting ligand-binding have been created. However, there remains a need to enhance both the accuracy and efficiency of these predictions (3).

There are three categories of algorithms for predicting protein-ligand binding sites: (a) geometry-based methods, (b) energy-based methods, and (c) knowledge-based methods. Geometry-based methods assume that the binding sites are in crevices or cavities on the protein surface and include methods like POCKET and LIGSITE. Energy-based methods use van der Waals interactions to predict sites, such as Q-SiteFinder and, on the other hand Knowledge-based methods include statistical and machine learning methods and look for clusters of conserved residues to predict binding sites. However, one of the main principal problems detecting ligand binding is the structural information of proteins, often limited in situations where only the protein target sequence is known, and adequate 3D structure information is unavailable (4).

Grid Analysis

Grid analysis is a valuable tool for studying ligand binding sites. By dividing the binding site into a grid of evenly spaced points, researchers can measure the interaction energies between the ligand and the protein at each point. Additionally, grid analysis can provide insights into the binding mechanism, such as the specific residues that contribute to the binding energy and the conformational changes that occur upon ligand binding. The use of grid analysis in ligand binding studies has led to the discovery of new ligands, the optimization of existing ligands, and the development of new drugs with increased binding potency and specificity (5,6).

Voronoi Analysis

In addition, when atomic coordinates are available, the Voronoi description of proteins is a useful geometric tool that has been applied in a variety of settings, and consist on a Voronoi diagram that assigns a Voronoi cell, which is a three-dimensional, convex polyhedron that contains all points in space that are closer to a specific atom than to any other atom, to each atom (7). It has been employed in recent times to establish contacts within macromolecules, without the need for a distance cutoff . The criteria for determining whether two atoms are in contact is whether their Voronoi cells share a common facet. The same technique can also be applied to amino acid residues, allowing the definition of residue-residue contacts (8).

CavPred Pipeline

Based on this logic, many geometry-based algorithms have been developed to predict protein-ligand interactions; geometry-based algorithms are independent of protein sequence and structure, which means that they can be applied to proteins from different families and with different conformations. Also, these methods have high efficacy and efficiency in identifying ligand binding sites, allowing rapid and accurate identification of ligand binding sites in proteins (9). Geometry-based methods for the detection of ligand binding sites have several advantages. Then here we present CavPred 1.0.0, a geometry-based method written in python to determine ligand binding sites in a protein using and combining voronoi and grid approximation to determine ligand binding sites and providing the list of amino acids involved. CavPred takes the pdb file as input and generates a pdb with ligand binding sites on the protein in pdb color format as output for grid analysis and provides voronoi diagrams and list of amino acids to corroborate grid analysis output. All instructions related to installation, use and other clarifications were put in the CavPred 1.0.0 tutorial.

Bibliography

1. Le Zhanga JZC. Exploring the computational methods for protein-ligand binding site prediction. Comput Struct Biotechnol J [Internet]. 2020 Jan 1 [cited 2023 Apr 10];18:417–26. Available from: http://dx.doi.org/10.1016/j.csbj.2020.02.008

2. Guo YDJT. Identification of Protein–Ligand Binding Sites by Sequence Information and Ensemble Classifier. American Chemical Society [Internet]. 2017;57(12):3149–61. Available from: https://doi.org/10.1021/acs.jcim.7b00307

3. Jérémy Desaphy, Karima Azdimousa, Esther Kellenberger, Didier Rognan. Comparison and Druggability Prediction of Protein–Ligand Binding Sites from Pharmacophore-Annotated Cavity Shapes. American Chemical Society [Internet]. 2012;52(8):2287–99. Available from: https://doi.org/10.1021/ci300184x

4. Dai T, Liu Q, Gao J, Cao Z, Zhu R. A new protein-ligand binding sites prediction method based on the integration of protein sequence conservation information. BMC Bioinformatics [Internet]. 2011 Dec 14 [cited 2023 Apr 10];12(14):1–7. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S14-S9

5. Harigua-Souiai E, Cortes-Ciriano I, Desdouits N, Malliavin TE, Guizani I, Nilges M, et al. Identification of binding sites and favorable ligand binding moieties by virtual screening and self-organizing map analysis. BMC Bioinformatics [Internet]. 2015 Mar 21 [cited 2023 Apr 10];16(1):1–15. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0518-z

6. Fukunishi Y, Nakamura H. Prediction of ligand-binding sites of proteins by molecular docking calculation for a random ligand library. Protein Sci [Internet]. 2011 Jan [cited 2023 Apr 10];20(1):95. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3047065/

7. Feinstein J, Shi W, Ramanujam J, Brylinski M. Bionoi: A Voronoi Diagram-Based Representation of Ligand-Binding Sites in Proteins for Machine Learning Applications. Protein-Ligand Interactions and Drug Design [Internet]. 2021 [cited 2023 Apr 10];299–312. Available from: https://link.springer.com/protocol/10.1007/978-1-0716-1209-5_17

8. Cazals F, Proust F, Bahadur RP, Janin J. Revisiting the Voronoi description of protein–protein interfaces. Protein Sci [Internet]. 2006 Sep [cited 2023 Apr 10];15(9):2082. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2242599/

9. Lee HS, Im W. Ligand Binding Site Detection by Local Structure Alignment and Its Performance Complementarity. J Chem Inf Model [Internet]. 2013 Sep 9 [cited 2023 Apr 10];53(9). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3821077/