About the Authors

Yi Xiong

Yi Xiong obtained his PhD degree in bioinformatics from Wuhan University, China, in 2011. He worked as a research assistant at the Department of Bioinformatics and Biostatistics in Shanghai Jiao Tong University, China, from 2009 to 2011. He is currently working as a postdoctoral research associate at the Department of Biological Sciences, Purdue University (IN, USA). He has published several scientific papers in peer-reviewed journals and conferences.

Xiaolei Zhu

Xiaolei Zhu obtained his PhD degree in 2009 from Peking University, China. He worked in University of Winsconsin-Madison, USA, in a postdoctorate position at the Department of Biochemistry from 2009 to 2011. He is now working in a postdoctoral position at the Department of Biological Sciences, Purdue University. He has published several scientific papers in peer-reviewed journals.

Daisuke Kihara

Daisuke Kihara is Associate Professor of the Department of Biological Sciences and the Department of Computer Science at Purdue University. He received his PhD degree from Kyoto University, Japan, in 1999. His research projects include structure- and sequence-based protein function prediction, computational drug design, protein tertiary structure prediction and protein docking. He has published over 85 research papers and book chapters. He is a member of The Protein Society, The Biophysical Society and The International Society of Computational Biology.

Chapter 15

Binding site identification in target proteins

Shape-based approaches for identifying & characterizing drug-binding sites 206
Binding pocket identification 208
Pocket comparison by Pocket-Surfer 210
Patch-surfer 212
Rapid shape-based ligand search 214

Yi Xiong, Xiaolei Zhu & Daisuke Kihara

When the tertiary structure of a target protein is known, computational methods can be used to identify potential drug-binding sites in the target protein. Furthermore, an identified binding site can be compared against known ligand-binding sites of proteins to characterize its physicochemical properties and binding ligands. In this chapter, we introduce a method for identification of ligand-binding sites in proteins, and two recent methods for ligand-binding site comparison on a large database of known binding sites as well as a rapid drug molecule search. These newly developed techniques will be useful for computer-aided drug design.

doi:10.4155/EBO.13.165

Virtual screening is a computational search of large libraries of chemical compounds for identifying a small set of compounds that would bind to a target.

The 3D Zernike descriptor is a series expansion of a 3D mathematical function, which represents a 3D object (e.g., pockets in a protein surface, global protein surface shape) in a compact and rotationally invariant fashion. It allows fast, real-time database search of the 3D object.

Shape-based approaches for identifying & characterizing drug-binding sites

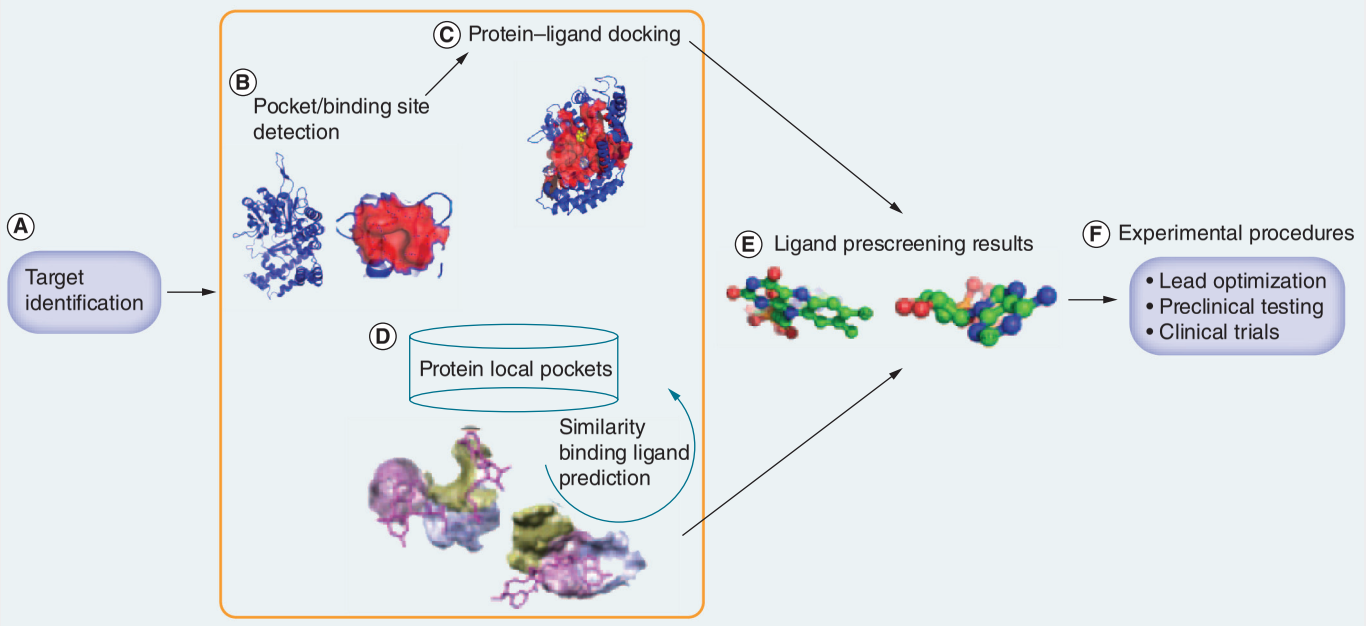
Developing an effective drug for a disease is a resource-consuming endeavor [1]. Computational methods can significantly reduce the financial cost and time for drug development. **Figure 15.1** illustrates that computational methods can be used in the drug development process if the tertiary structure of a target protein is known **(Figure 15.1A)**. The first step of the computational analysis would be identification of ligand-binding sites in the target protein by considering geometrical and physicochemical properties **(Figure 15.1B)**. Next, virtual screening can be performed to search chemical compounds in a large library that may bind to the binding pocket **(Figure 15.1C)** [2]. At the same time, fast binding pocket comparison methods, which predict binding ligands by finding similar known pockets, will be able to provide additional clues for developing drugs **(Figure 15.1D)**. Selected chemical compounds **(Figure 15.1E)** are then subject to lead optimization and experiments to examine binding affinity, preclinical testing (analysis of the bioactivity, safety and efficacy of the formulated lead on animals), and clinical trials **(Figure 15.1F)** [3].

In this chapter, we overview computational methods for pocket detection and comparison, which are developed in our group. We first review a ligand-binding site-finding program, named VisGrid [4]. VisGrid uses an intuitive metric called visibility to identify pockets in protein surfaces. In the subsequent sections, we describe fast binding-site comparison methods, Pocket-Surfer [5] and Patch-Surfer [6]. These methods use a compact molecular surface representation named 3D Zernike descriptor (3DZD), which enables a fast, real-time search against a few thousands of known pockets [7]. Since results of binding ligand prediction by the two methods will be finished in a couple of minutes, Pocket-Surfer and Patch-Surfer can be used as a prescreening step prior to the conventional *in silico* screening that actually performs ligand and protein docking computation [8]. Alternatively, results of Pocket-Surfer and Patch-Surfer can be used as orthogonal information to *in silico* screening results, since the former predict binding ligand from protein pocket

The ligand–protein interaction is governed by physical forces, including Van der Waals, electrostatic forces, hydrogen bonds and hydrophobic contacts.

The binding of ligands to a pocket can be predicted by identifying similar known pockets from a database. 3D Zernike descriptor, a mathematical series expansion of a 3D function, is an efficient protein surface representation that allows a fast database search.

Figure 15.1. The drug-discovery process.



(A) Target identification; **(B)** pocket/binding site detection; **(C)** protein–ligand docking; **(D)** protein local pockets comparison; **(E)** ligand prescreening results; and **(F)** experimental procedures.

Identifying a ligand-binding pocket is a first step for performing *in silico* drug screening. Potential ligand-binding sites in a target protein can be predicted by detecting geometrical pocket regions in protein surfaces.

Binding pockets of the same ligand may have different overall shape due to flexibility of the ligand molecule or binding of water molecules. Their similarity can be identified by comparing local surface regions of pockets.

comparison while the latter evaluates binding energy of compounds in a given binding pocket. In the last section, we show that 3DZD is also effective in performing fast database search of the 3D shapes of chemical compounds. The approaches described in this chapter are new in the field and not yet routinely used in actual drug development. We believe they have great potential for effective drug design, as they are unique in that they can perform 3D shape comparison of proteins and compounds.

Binding pocket identification

In most of the cases, a small chemical ligand molecule binds to a pocket (cavity) of a target protein. Thus, potential ligand-binding sites in a target protein can be well predicted by detecting geometrical pocket regions on a protein surface. Identifying binding pockets is also a first step for performing *in silico* drug screening, since a ligand–protein docking method usually limits its search space to a potential binding region. In the last decade, a wide variety of geometry-based algorithms have been proposed to detect the pockets, based on the assumption that a ligand tends to bind in the largest pocket in its target protein. SURFnet identifies the largest pocket as a sphere with maximum volume by placing a sphere between two atoms [9]. LIGSITE places a protein onto a 3D grid and identifies a binding pocket as solvent-accessible grids, which are enclosed on both sides by the protein [10]. CAST triangulates the surface atoms and groups triangles by merging small triangles to neighboring large triangles [11]. The pocket is then defined as a collection of empty triangles. PASS coats a protein surface with probe spheres, and identifies the pockets as probes that have large number of contacts with atoms [12]. Existing available methods are summarized in **Table 15.1**. Here, we introduce the VisGrid algorithm, which uses an intuitive idea of the visibility to characterize local geometric features of protein surfaces [4]. The visibility for a position on a protein surface is defined as the number of visible directions divided by the total number of possible directions. The total number of possible directions of a voxel is 26 (a cube of $3 \times 3 \times 3$ minus the center voxel) when the first surrounding layer of the voxel is considered, while it grows to 98 (a cube of $5 \times 5 \times 5$ minus a cube of $3 \times 3 \times 3$) when the second layer is taken into account. A direction is regarded as visible when a ray casted from the target voxel toward the direction does not encounter the protein surface for 20 steps.

Table 15.1. Available software for pocket identification, binding site comparison and chemical compound database search.

Category	Software	Feature	Web address
Pocket identification	SURFNET	Identifies a pocket with a large and a small probe sphere rolling on the protein surface	www.ebi.ac.uk/thornton-srv/software/SURFNET [104]
	LIGSITEcsc	Identifies a ligand-binding pocket as a spatially semiclosed region by protein surface and shows sequence conservation	http://projects.biotec.tu-dresden.de/pocket [105]
	CAST	Identifies a pocket using the Delaunay triangulation and the α complex	http://sts.bioengr.uic.edu/castp [106]
	PASS	Identifies a pocket as a surface region where small probe spheres fit into	www.ccl.net/cca/software/UNIX/pass/overview.html [107]
	VisGrid	Identifies a pocket as a surface position with a low visibility	http://kiharalab.org/VisGrid [101]
Binding site comparison	eF-Site	Compares shape and the surface electrostatic potential using a graph matching method	http://ef-site.hgc.jp [108]
	SitesBase	Compares atoms in pockets using geometric hashing	www.modelling.leeds.ac.uk/sb [109]
	Pocket-Surfer	3DZD-based method described in this chapter	http://kiharalab.org/pocket-surfer [103]
Chemical compound database search	SIMCOMP	Subgraph matching method	www.genome.jp/tools/simcomp [110]
	UNITY2D	Fingerprint-based method, a molecule is represented as a Boolean array of features of the molecule	A part of SYBYL package www.tripos.com [111]
	PubChem	Users can draw query molecules or specify formula	http://pubchem.ncbi.nlm.nih.gov [112]

The VisGrid algorithm works as follows: the first step is to map the protein onto a 3D grid (the grid size is set to 0.9 Å). Voxels that overlap with the protein atoms are marked as protein voxels while the rest of them are marked as empty. Protein voxels that are adjacent to an empty voxel are considered as protein surface voxels. Thus, each voxel in the model is categorized into one of the three types: surface, filled (but not surface) or empty. Among the surface voxels, voxels in a pocket are defined as those whose visibility is smaller than a defined cutoff. By contrast, voxels with a large visibility indicate that they are on the top of a protrusion. Voxels with a certain visibility or lower (i.e., voxels in pockets) are grouped according

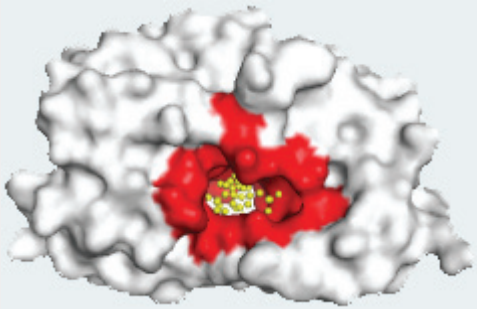
to their spatial proximity. If a voxel is farther than 2.0 Å to any other voxels that constitute a pocket, it belongs to a different new pocket. Although VisGrid is mainly used for identifying pockets, it can also detect protrusions, flat region and hollows (space inside of protein volume) on a protein surface.

Figure 15.2 shows a ligand-binding pocket in HIV-1 protease identified by VisGrid (red). The yellow molecule is an inhibitor (aminodiol inhibitor). When benchmarked on a large dataset of 5616 proteins with binding pockets, 83.5% of the binding pockets were identified by VisGrid as one of the three largest pockets on the surface. VisGrid is freely available for academic users at [101]. It is also incorporated at the 3D-Surfer website [102].

Pocket comparison by Pocket-Surfer

A pocket region identified in a target can be characterized by comparing it against known binding sites of proteins. If a pocket in a target protein is found to be similar to known binding pockets in terms of geometric shape and physicochemical properties, it is natural to speculate that similar ligands to the known binding pockets would bind to the query pocket. Local pocket comparison methods are also very useful for function prediction of proteins, because they do not rely on homologous relationship of proteins. A pocket can be naturally represented by the coordinates of residues/atoms in the pocket. The similarity of two pockets can be quantified by the root mean square deviation of corresponding residues/atoms [13–15] after identifying corresponding atoms by two pockets, by geometric hashing technique [16] or fingerprinting methods [17,18]. Alternatively, surface shape is used for pocket representation. Klebe *et al.* described the surface geometry and the electrostatic potential of binding sites using subgraph matching algorithm [19]. In the eF-Site method, Kinoshita *et al.* used a clique detection algorithm for finding similar local surfaces that are represented by graphs [20]. In addition, Guerra *et al.* used the spin-image, a 2D histogram representation of surface points, to describe the relative geometrical position between a point and the other points [21]. Another way to represent surface is based on mathematical moments. Our method, Pocket-Surfer, uses 3D Zernike moments to represent pocket surface [5].

Figure 15.2. Example of a ligand-binding pocket of HIV-1 protease (protein data bank ID: 1ODW) identified by VisGrid.



Pocket-Surfer searches similar binding pockets in a database for a query pocket by

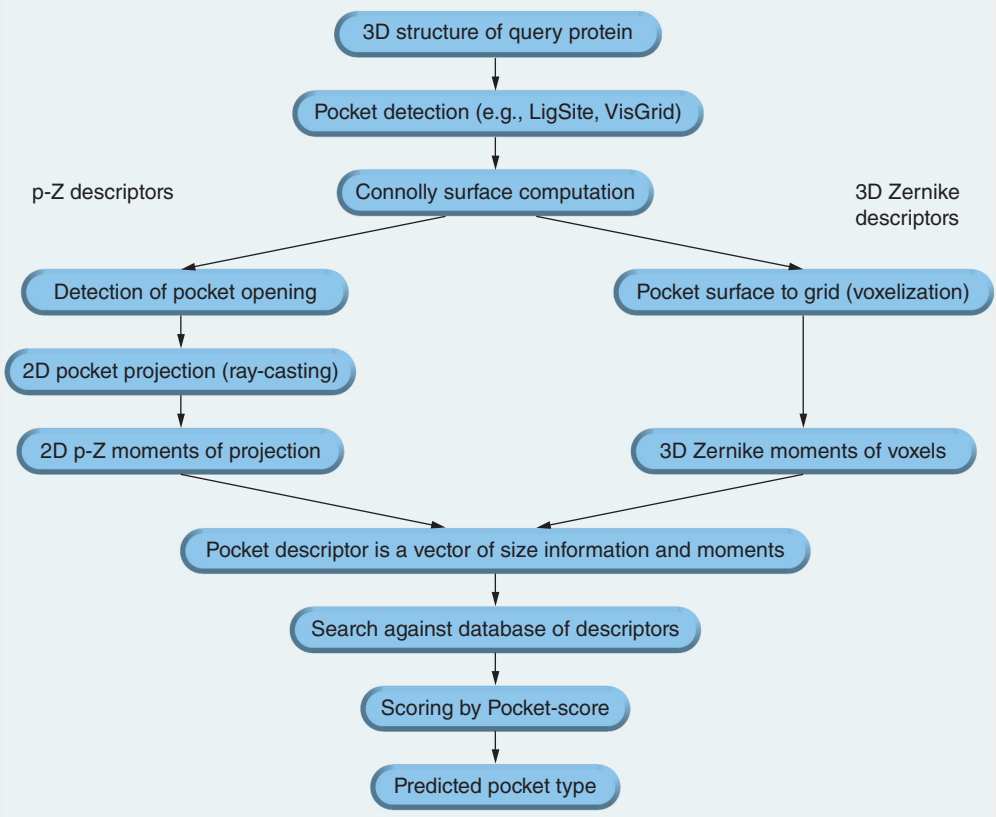
comparing shape and physicochemical properties [5]. It was originally developed for predicting binding ligands for a query protein. In the computational drug-design scenario, the method will be useful for finding natural ligand molecules or drugs that bind to similar pockets to the query pocket, because these molecules may be considered as drug leads for the target. Pocket-Surfer allows fast, real-time pocket searches due to the use of a mathematical compact and rotationally invariant representation of pockets – that is, 3DZD [7,22] and 2D Pseudo-Zernike (2DPZ) Moments. 3DZD and 2DPZ are series expansions of mathematical 3D or 2D functions, where the 3D or 2D functions represent a pocket that is mapped onto a 3D/2D grid.

The flowchart of Pocket-Surfer is shown in **Figure 15.3**. Given a query protein structure, a ligand-binding pocket is detected by a pocket identification program such as VisGrid or LIGSITE. Then, pocket surface is constructed from atoms in the pocket region. Next, the pocket surface is placed on a 3D grid. To represent a pocket surface on a 3D grid, a grid (voxel) is assigned a value of 1 if it is occupied by the protein and 0 otherwise. Values of physicochemical properties, such as electrostatic potentials and hydrophobicity scales, are also assigned to the surface voxels. The resulting voxels with assigned values are considered as a 3D function, which is represented with 3DZD.

In addition, we also developed a pocket comparison method using a 2D representation of pockets. In this scheme, a pocket is projected onto a 2D plane of a spherical panoramic picture. For each direction of (θ,ϕ) from the center of the pocket, the physical distance to the pocket wall is computed and mapped on the 2D picture. In addition to the distance, another picture that maps surface electrostatic potential is also prepared. The 2DPZ moments computed for these pictures are mathematically invariant to rotation around the axis from the center of the pocket to the pocket opening. This 2D representation of pockets relies on the fact that pockets can be quite reliably prealigned using their opening.

In both 3DZD and 2DPZ representations, a feature (either 3D shape or the surface electrostatic potential) of a pocket is represented by a vector of coefficients of the series expansion of the descriptors. **Figure 15.4** shows an example of the 3D and 2D representation of a flavin adenine dinucleotide-binding pocket of glutathione reductase and its 3DZD and 2DPZ representations. Since 3DZD and 2DPZ are vectors, similarity of the feature of pockets can be simply evaluated by the Euclidean distance between vectors. In addition to the two features, the size of pockets is also considered, which is represented by the longest distance from the center

Figure 15.3. The flowchart of Pocket-Surfer.



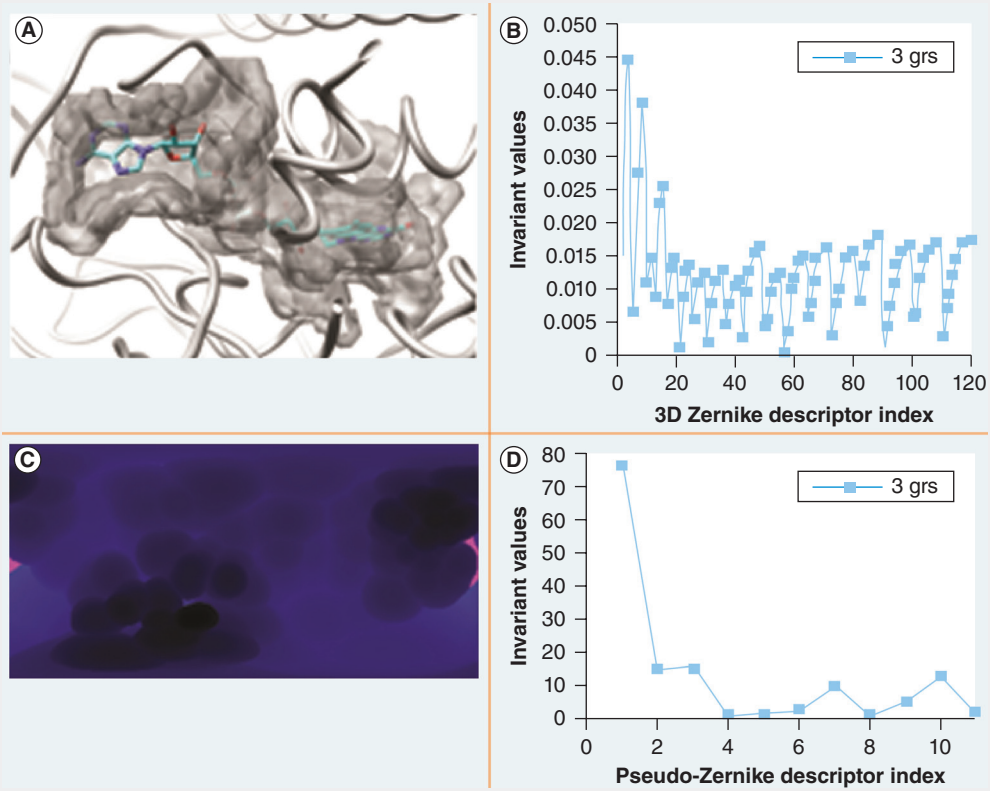
p-Z: Pseudo-Zernike.

to the wall. The performance of Pocket-Surfer was benchmarked by examining how well binding ligands of a query pocket can be predicted by searching and ranking similar known pockets to the query in a database. It was shown that our methods are favorably compared with similar methods [5,23]. Pocket-Surfer was implemented as a web server at [103].

Patch-Surfer

Pocket-Surfer describes the overall shape and physicochemical properties of ligand-binding pockets in protein surfaces [5]. Although it shows superior performance in binding ligand prediction to the other existing methods, it fails to identify pockets of the same ligand type if the pockets have significantly different shapes and properties. The shape and physicochemical

Figure 15.4. Examples of binding pocket representation by 3D Zernike descriptor and 2D Pseudo-Zernike.



(A) A flavin adenine dinucleotide-binding site of glutathione reductase (Protein Data Bank ID: 3GRS). **(B)** The 3D Zernike descriptor of the binding site. **(C)** The 2D picture of the pocket. The color scale shows the distance from the center to the pocket surface. The darker, the longer the distance is. **(D)** The 2D Pseudo-Zernike of the 2D picture.
Reproduced with permission from [23].

properties of pockets for the same ligand type can significantly vary due to several reasons, such as flexibility of the ligand molecules and binding of solvent molecules to the pockets [24]. Since such pockets still show consistent properties at local regions in the pockets, the similarity between pockets may be evaluated by combination of similarity of their local patch regions.

In the second pocket comparison method we developed, Patch-Surfer, a pocket is segmented into circular patches, each of which are

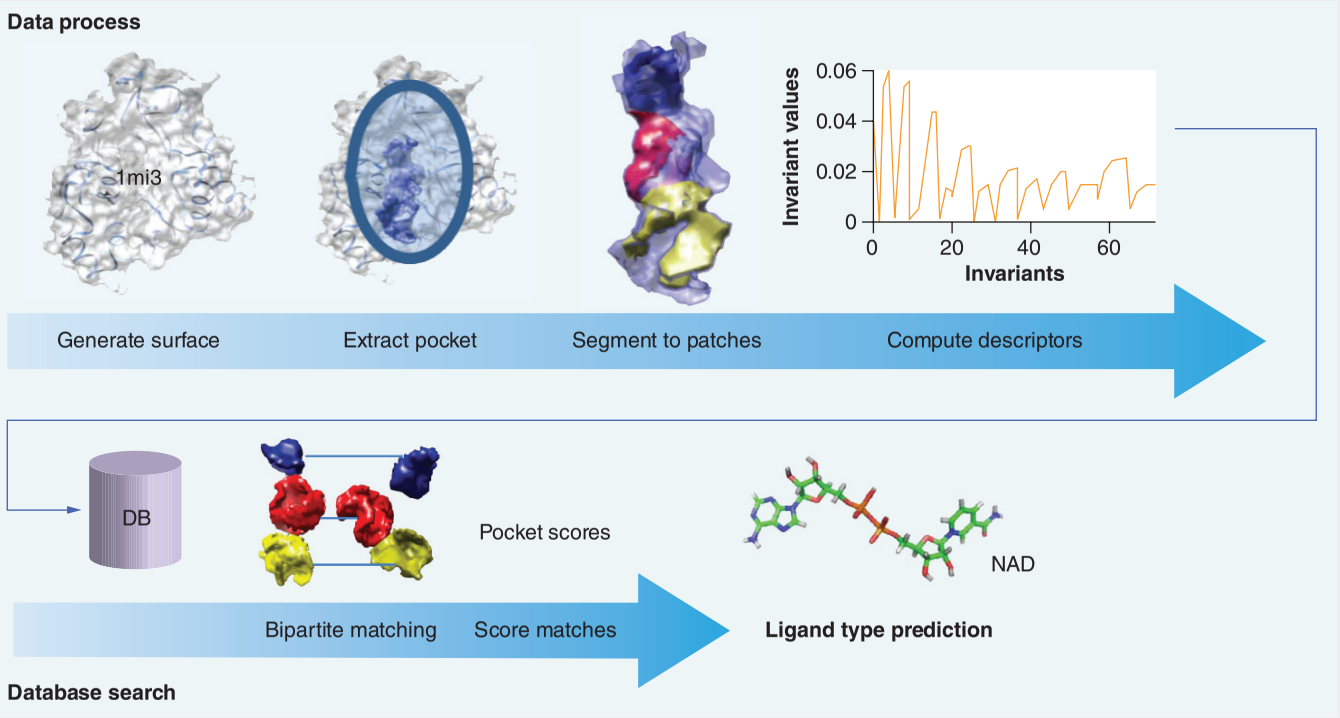
characterized by shape, surface electrostatic potential, hydrophobicity and concaveness [6]. In the same way as implemented in Pocket-Surfer, these four features are represented by the 3DZD. A pocket is represented by 30–60 overlapping surface patches. The flowchart of Patch-Surfer is shown in **Figure 15.5** [25]. First, the surface of a pocket region in a protein is extracted, which is then segmented into a group of surface patches. As mentioned above, they are characterized by four different surface properties. To quantify the similarity of two pockets, the similarity of all pairs of surface patches in the two pockets is computed, which is followed by matching of patches of similar properties from the two pockets. After the matched, patches are identified, the similarity of the two pockets is computed by combining three scores, which are the average similarity of the matched patches, the average relative position of the matched patches in each of the pockets, and the size of the two pockets (the number of patches in the pockets). For a query pocket potential binding ligands are predicted based on the list of the similar pockets retrieved from a database and ranked by the similarity to the query pocket.

Figure 15.6 shows illustrative examples of pocket pairs that are detected as similar to each other by Patch-Surfer. The first two examples, **Figures 15.6A & 15.6B**, are cases where pockets bind the same ligands but their overall shapes are different due to the different conformation of the ligand molecules. In the both cases of nicotinamide adenine dinucleotide **(Figure 15.6A)** and fructose-6-phosphate **(Figure 15.6B)**, the ligands bind in a stretched conformation to the targets on the left figure while in a more compact conformation on the right figure. Despite overall pocket shape difference, Patch-Surfer could identify them as similar pockets because equivalent patches in the two pockets that bind to the same chemical groups are matched (they are represented in the same color). The last example, **Figure 15.6C** shows a pair of pockets that bind different ligands, nicotinamide adenine dinucleotide or flavin adenine dinucleotide. When these two pockets are compared by Patch-Surfer, it detects adenosine-binding regions of the two pockets as similar local patches (shown in the same color). The recognition of binding regions of chemical groups does not always contribute to improvement in the prediction accuracy of binding ligands; however, it suggests that Patch-Surfer can be developed into a unique method for predicting chemical group-binding sites rather than predicting the entire ligand molecules.

Rapid shape-based ligand search

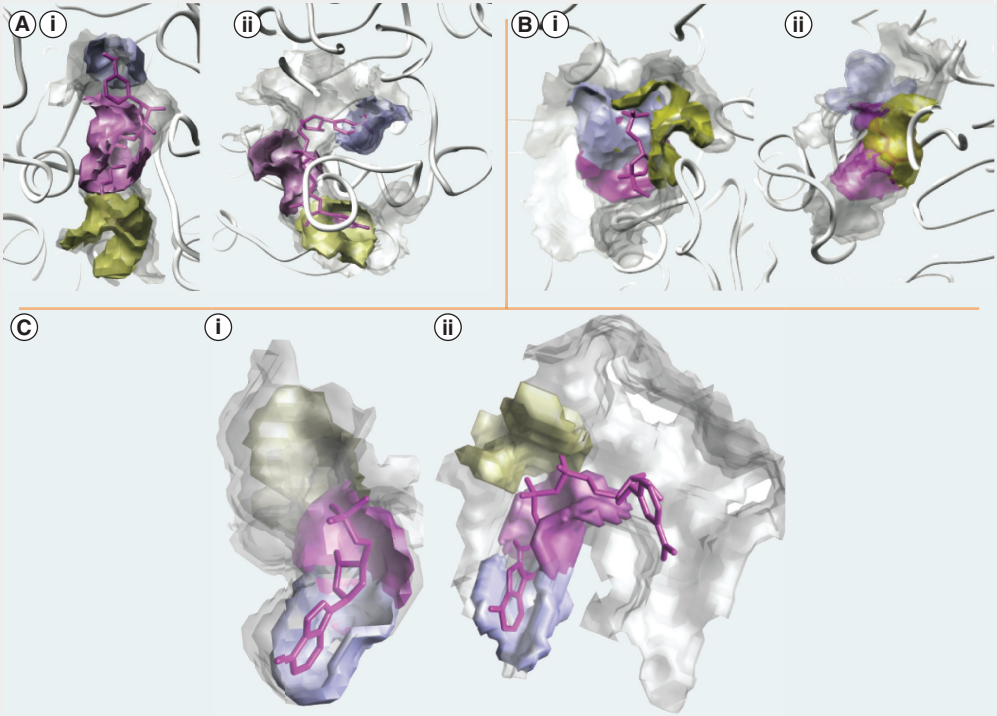
In the last section, we discussed the fact that the 3DZD is also effective for rapid chemical compound screening regarding their 3D shape

Figure 15.5. Flow chart of Patch-Surfer.



DB: Database; NAD: Nicotinamide adenine dinucleotide.

Figure 15.6. Examples of pockets that bind ligands in different conformations.

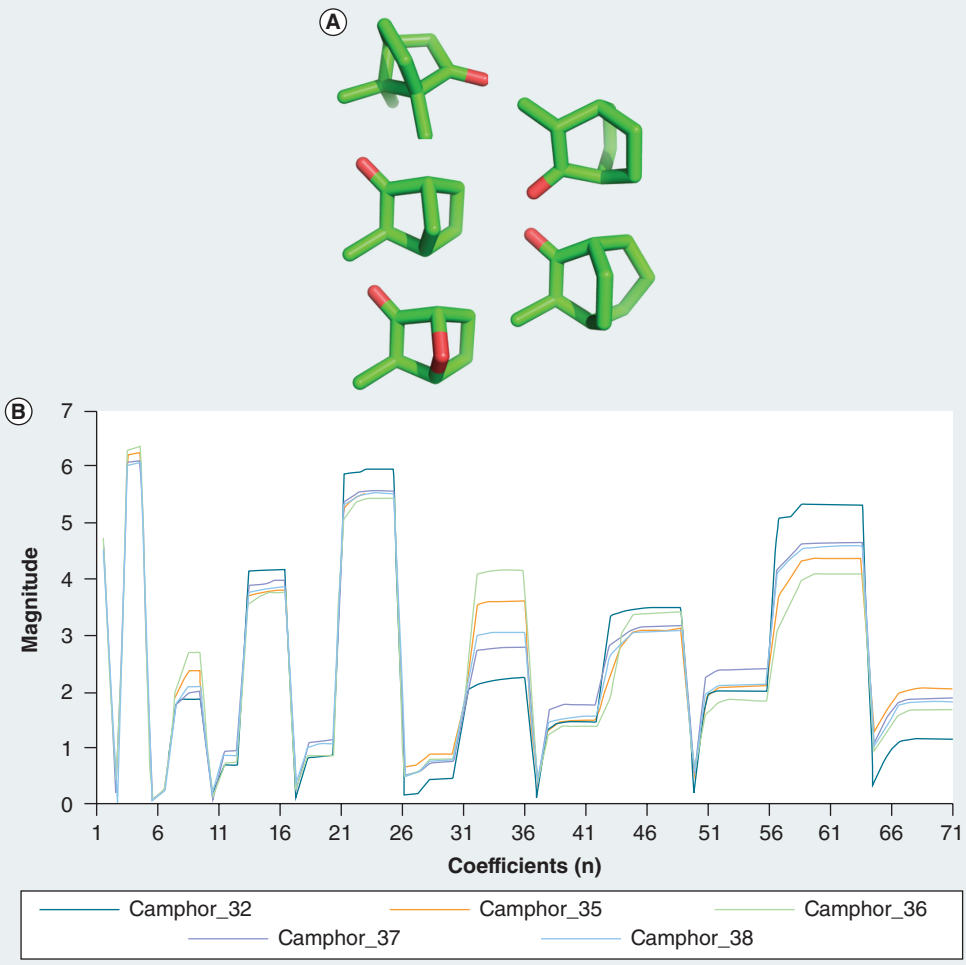


In a database search, Patch-Surfer was able to identify them as similar by matching pocket positions binding special chemical group. **(A)** A pair of nicotinamide adenine dinucleotide-binding proteins, **(A,i)** xylose reductase (protein data bank [PDB] ID: 1mi3) and **(A,ii)** sir2 enzyme (PDB ID: 1s7g). The root mean square deviation of the ligand molecules is 3.49 Å. **(B)** A pair of fructose-6-phosphate-binding proteins, **(B,i)** sucrose phosphate synthase (PDB ID: 2r66) and **(B,ii)** central glycolytic gene regulator (PDB ID: 3bxh). The root mean square deviation of the ligands is 1.02 Å. **(C)** Binding regions of adenosine moiety of **(C,i)** nicotinamide adenine dinucleotide (PDB ID: 1s7g) and **(C,ii)** flavin adenine dinucleotide (PDB ID: 1k87) detected by Patch-Surfer. Adapted from [6].

information. With the increasing availability of a chemical space of compounds, accurate identification of potential drug leads from a large database of compounds is a crucial step in the early stages of drug design. In general, two main factors determine the efficacy of the similarity-based searching methods for identifying drug leads, which are a molecular representation scheme and a similarity measure used. A variety of molecular representation methods have been proposed, including graph-based and fragment-based approaches. Among them, 3DZD is unique among others in that it truly captures 3D shape

information and still allows fast database screening [26]. 3DZD for a chemical molecule is computed in the same way as it is used to represent pockets. A chemical molecule is mapped on a 3D grid, where molecular surface is marked as 1 and 0 otherwise. Then, considering the 3D grid with 1 and 0 as a 3D function, 3DZD is computed. **Figure 15.7** shows examples of 3DZD for five molecules that have camphor smell. In our

Figure 15.7. Compounds in the camphor class.



(A) Atom composition and shape. **(B)** 3D Zernike descriptor invariants.
Reproduced from [27].

paper, we reported benchmark studies that compared 3DZD's performance of clustering similar compounds and also retrieving active drug molecules for several targets in comparison with other existing ligand comparison methods [27]. Through the benchmark studies, the strengths and weakness of 3DZD were highlighted. Since 3DZD represents the 3D surface shape of molecules, active molecules that do not have similar molecular structure but similar surface shape can clearly be better retrieved by 3DZD than other methods that consider atomic chemical structures of ligands. On the other hand, the weakness of 3DZD is that it is sensitive to alternative conformations of molecules, since they change overall surface shapes. Despite some observed weakness, 3DZD is a new attractive method for database screening of chemical molecules, because it is fast and directly compares the 3D shape of ligand molecules without considering molecular backbones. Thus, 3DZD can be effective for lead hopping for a target with a known ligand.

Acknowledgements

Y Xiong and X Zhu made an equal contribution to this article.

Financial & competing interests disclosure

This work has been supported by grants from the NIH (R01GM075004 and R01GM097528), National Science Foundation (EF0850009, IIS0915801 and DMS0800568), and the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-220-C00004). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Summary

- Computational methods can significantly reduce the costs of drug development.
- Ligand- and drug-binding sites of targets can be detected by identifying geometrical pocket regions in the target protein surface.
- Binding sites can be characterized by identifying similar pockets of known binding ligands from a protein structure database.
- The 3D Zernike descriptor has the unique ability for rapid searching of shape-based similar molecules on large databases of pockets.
- The 3D Zernike descriptor is also effective for shape-based ligand database screening.

References

1 Haupt VJ, Schroeder M. Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief. Bioinform.* 12(4), 312–326 (2011).

2 Lengauer T, Lemmen C, Rarey M, Zimmermann M. Novel technologies for virtual screening. *Drug Discov. Today* 9(1), 27–34 (2004).

3 Swinney DC, Anthony J. How were new medicines discovered? *Nat. Rev. Drug Discov.* 10(7), 507–519 (2011).

4 Li B, Turuvekere S, Agrawal M, La D, Ramani K, Kihara D. Characterization of local geometry of protein surfaces with the visibility criterion. *Proteins* 71(2), 670–683 (2008).

5 Chikhi R, Sael L, Kihara D. Real-time ligand binding pocket database search using local surface descriptors. *Proteins* 78(9), 2007–2028 (2010).

6 Sael L, Kihara D. Detecting local ligand-binding site similarity in nonhomologous proteins by surface patch comparison. *Proteins* 80(4), 1177–1195 (2012).

7 Novotni M, Klein R. 3D Zernike descriptors for content based shape retrieval. Presented at: *the 8th ACM Symposium on Solid Modeling and Applications.* Seattle, WA, USA, 216–225 (2003).

8 Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31(2), 455–461 (2010).

9 Laskowski RA. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.* 13(5), 323–330 (1995).

10 Hendlich M, Rippmann F, Barnickel G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* 15(6), 359–363, 389 (1997).

11 Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* 7(9), 1884–1897 (1998).

12 Brady GP Jr, Stouten PF. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* 14(4), 383–401 (2000).

13 Porter CT, Bartlett GJ, Thornton JM. The catalytic site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.* 32, 129–133 (2004).

14 Arakaki AK, Zhang Y, Skolnick J. Large-scale assessment of the utility of low-resolution protein structures for biochemical function assignment. *Bioinformatics* 20(7), 1087–1096 (2004).

15 Ferre F, Ausiello G, Zanzoni A, Helmer-Citterich M. SURFACE: a database of protein surface regions for functional annotation. *Nucleic Acids Res.* 32, 240–244 (2004).

16 Gold ND, Jackson RM. Fold independent structural comparisons of protein–ligand binding sites for exploring functional relationships. *J. Mol. Biol.* 355(5), 1112–1124 (2006).

17 Yeturu K, Chandra N. PocketMatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics* 9, 543 (2008).

18 Xiong B, Wu J, Burk DL, Xue M, Jiang H, Shen J. BSSF: a fingerprint based ultrafast binding site similarity search and function analysis server. *BMC Bioinformatics* 11, 47 (2010).

19 Schmitt S, Kuhn D, Klebe G. A new method to detect related function among proteins independent of sequence and fold homology. *J. Mol. Biol.* 323(2), 387–406 (2002).

20 Kinoshita K, Nakamura H. Identification of protein biochemical functions by similarity search using the molecular surface database eF-Site. *Protein Sci.* 12(8), 1589–1595 (2003).

21 Bock ME, Garutti C, Guerra C. Cavity detection and matching for binding site recognition. *Theor. Comput. Sci.* 408(2–3), 151–162 (2008).

22 Canterakis N. 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition. Presented at: *the 11th Scandinavian Conference on Image Analysis.* Greenland, Denmark, 85–93 (1999).

23 Chikhi R, Sael L, Kihara D. Protein binding ligand prediction using moments-based methods. In: *Protein*

Function Prediction for Omics Era. Kihara D (Ed.). Springer, 145–163 (2011).

24 Kahraman A, Morris RJ, Laskowski RA, Favia AD, Thornton JM. On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins. *Proteins* 78(5), 1120–1136 (2010).

25 Sael L, Kihara D. Binding ligand prediction for proteins using partial matching of local surface patches. *Int. J. Mol. Sci.* 11(12), 5009–5026 (2010).

26 Venkatraman V, Sael L, Kihara D. Potential for protein surface shape analysis using spherical harmonics and 3D Zernike descriptors. *Cell Biochem. Biophys.* 54(1–3), 23–32 (2009).

27 Venkatraman V, Chakravarthy PR, Kihara D. Application of 3D Zernike descriptors to shape-based ligand similarity searching. *J. Cheminformat.* 1, 19 (2009).

Websites

101 VisGrid. http://kiharalab.org/VisGrid

102 3D-Surfer 2.0. http://kiharalab.org/3d-surfer

103 Pocket Surfer. http://kiharalab.org/pocket-surfer

104 SURFnet. www.ebi.ac.uk/thornton-srv/software/SURFNET

105 LIGSITE. Pocket identification. http://projects.biotec.tu-dresden.de/pocket

106 Computed Atlas of Surface Topography of Proteins. http://sts.bioengr.uic.edu/castp

107 Putative Active Sites with Spheres. www.ccl.net/cca/software/UNIX/pass/overview.html

108 Electrostatic Surface of Functional Site. http://ef-site.hgc.jp

109 SitesBase. www.modelling.leeds.ac.uk/sb

110 SIMCOMP Search. www.genome.jp/tools/simcomp

111 Tripos. A Certara™ Company. www.tripos.com

112 The PubChem Project. http://pubchem.ncbi.nlm.nih.gov