



# Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery

Ashutosh Kumar and Kam Y. J. Zhang\*

Laboratory for Structural Bioinformatics, Center for Biosystems Dynamics Research, RIKEN, Yokohama, Japan

## OPEN ACCESS

### Edited by:

Honglin Li,  
East China University of Science and  
Technology, China

### Reviewed by:

Pedro Ballester,  
Institut National de la Santé et de la  
Recherche Médicale (INSERM),  
France  
Antreas Afantitis,  
NovaMechanics Ltd., Cyprus  
Xiaofeng Liu,  
East China University of Science and  
Technology, China

### \*Correspondence:

Kam Y. J. Zhang  
kamzhang@riken.jp

### Specialty section:

This article was submitted to  
Medicinal and Pharmaceutical  
Chemistry,  
a section of the journal  
Frontiers in Chemistry

Received: 10 April 2018

Accepted: 09 July 2018

Published: 25 July 2018

### Citation:

Kumar A and Zhang KYJ (2018)  
Advances in the Development of  
Shape Similarity Methods and Their  
Application in Drug Discovery.  
Front. Chem. 6:315.  
doi: 10.3389/fchem.2018.00315

Molecular similarity is a key concept in drug discovery. It is based on the assumption that structurally similar molecules frequently have similar properties. Assessment of similarity between small molecules has been highly effective in the discovery and development of various drugs. Especially, two-dimensional (2D) similarity approaches have been quite popular due to their simplicity, accuracy and efficiency. Recently, the focus has been shifted toward the development of methods involving the representation and comparison of three-dimensional (3D) conformation of small molecules. Among the 3D similarity methods, evaluation of shape similarity is now gaining attention for its application not only in virtual screening but also in molecular target prediction, drug repurposing and scaffold hopping. A wide range of methods have been developed to describe molecular shape and to determine the shape similarity between small molecules. The most widely used methods include atom distance-based methods, surface-based approaches such as spherical harmonics and 3D Zernike descriptors, atom-centered Gaussian overlay based representations. Several of these methods demonstrated excellent virtual screening performance not only retrospectively but also prospectively. In addition to methods assessing the similarity between small molecules, shape similarity approaches have been developed to compare shapes of protein structures and binding pockets. Additionally, shape comparisons between atomic models and 3D density maps allowed the fitting of atomic models into cryo-electron microscopy maps. This review aims to summarize the methodological advances in shape similarity assessment highlighting advantages, disadvantages and their application in drug discovery.

**Keywords:** molecular similarity, virtual screening, shape similarity, drug discovery, gaussian overlay, spherical harmonics, 3D Zernike descriptors

## INTRODUCTION

Molecular similarity is a key concept in drug discovery and has been routinely used in the discovery and design of new molecules. It is based on the notion that two molecules often share similar physical properties and biological function if they are structurally similar. This similarity principle has been widely utilized in early phases of drug development to discover new molecules. Virtual screening has been used to filter large databases of compounds to a smaller number based on this similarity principle. Molecular similarity has been also employed to optimize the potency and pharmacokinetic properties of lead compounds based on their structure–activity relationships.

There are two components of molecular similarity analysis (1) structural representations and (2) quantitative measurements of similarity between two structural representations. Many types of structural representations have been suggested to measure the similarity between two molecules. These include physiochemical properties, topological indices, molecular graphs, pharmacophore features, molecular shapes, molecular fields etc. Further, there are various methods to quantify the similarity between two structural representations, e.g., Tanimoto coefficient, Dice index, cosine coefficient, Euclidean distance, Tversky index etc. Among these, Tanimoto coefficient (Rogers and Tanimoto, 1960) is the most popular and widely used similarity measure. Based on the structural representation, molecular similarity approaches can be broadly classified into 2D or 3D similarity methods. The 2D similarity methods rely only on the 2D structural information and are among the fastest, efficient and most popular similarity search methods. Moreover, they do not rely on structural alignments for estimating the similarity between two molecules. These methods include substructure search, fingerprint similarity search and 2D descriptor-based methods. However, most of these methods are limited in their ability to enable scaffold hopping and provide no structural and mechanistic insights. To deal with the limitations associated with 2D similarity methods, several approaches were developed that account for 3D conformations of a molecule while performing similarity search. These methods include pharmacophore modeling, shape similarity, molecular field-based methods, 3D fingerprints among others. In recent years, ligand 3D shape-based similarity analysis has become a method of choice in increasing number of virtual screening campaigns. Several successful applications of shape similarity to discover new molecules have been published in the literature. The major advantage with shape-based virtual screening methods is that scaffold hopping can be conveniently accomplished and scaffolds other than the query can be identified.

In this review, we will summarize the development and application of various 3D shape similarity methods and will comment on their utility in drug discovery. We will first outline the classification and various types of 3D shape similarity methods highlighting their advantages and disadvantages. Later, we will describe various applications of 3D shape similarity methods in drug discovery.

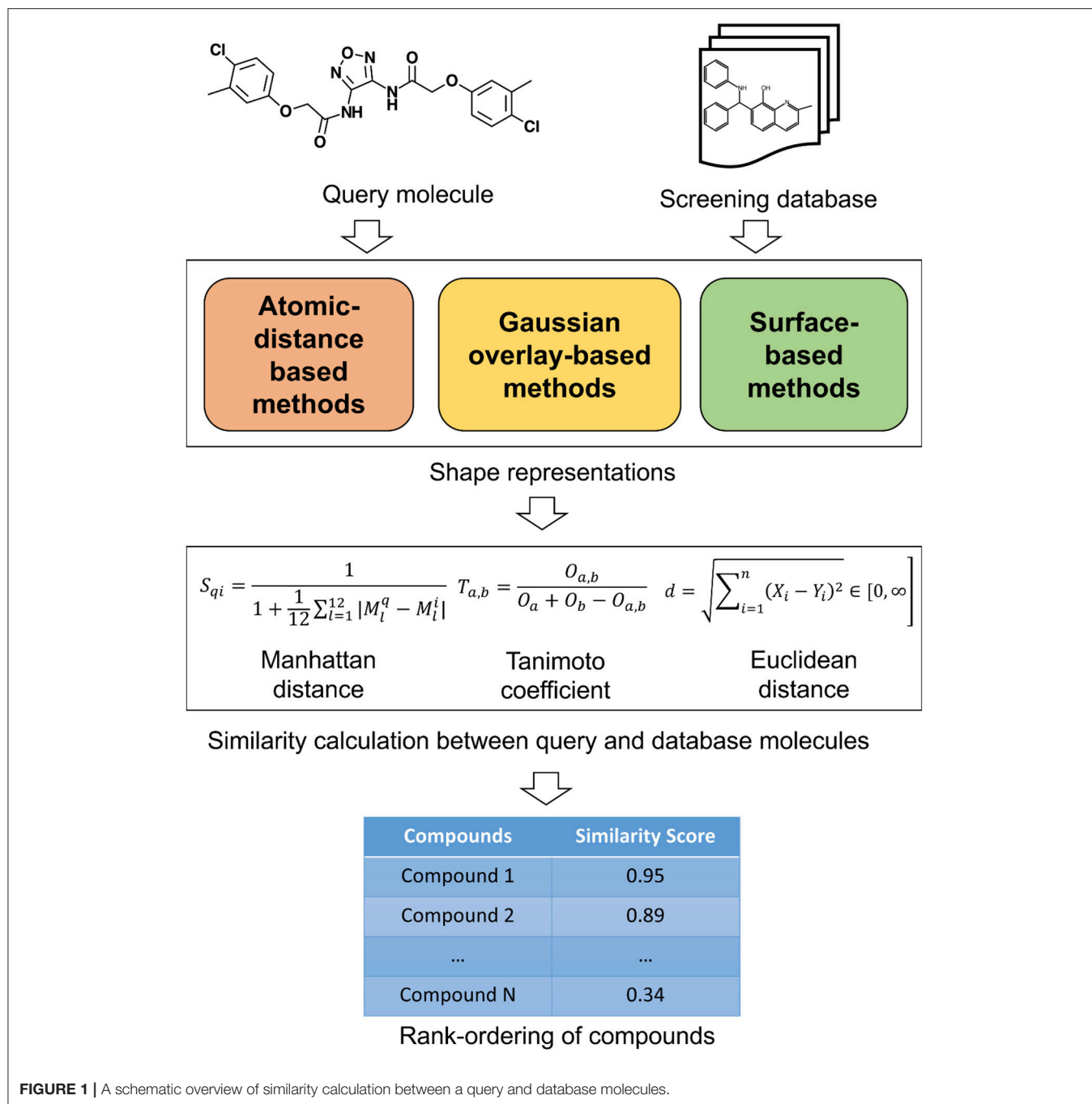
### 3D SHAPE SIMILARITY METHODS

The 3D shape has been widely recognized as a key determinant for the activity of small molecules and other biomolecules (Zauhar et al., 2003; Rush et al., 2005; Schnecke and Boström, 2006; Kortagere et al., 2009). The shape complementarity between ligand and receptor is necessary for bringing the receptor and ligand sufficiently close to each other so they can form critical interactions necessary for binding. Two molecules with similar shape are likely to fit the same binding pocket and thereby exhibiting similar biological activity. Shape comparison methods could be broadly classified as (1) Alignment-free or non-superposition methods and (2)

Alignment or superposition-based methods. Both of these methods have their own advantages and disadvantages. Alignment-free methods are independent of the position and orientations of molecules. As such, they are much faster and could be used to screen large compound databases. Alignment-based methods rely on finding the optimal superposition between the compounds. Alignment-based methods are highly effective in identifying shape similarities among the molecular structures but they are computationally expensive. These methods enable comparison of the surface properties such as hydrophobicity and polarity. Visualization is one of the advantages with the alignment-based methods and the similarity between two molecules can be displayed. This information is useful in the design of new molecules and to guide further optimization. However, a subpar molecular alignment may lead to errors in comparing two molecules. Apart from this broad classification, shape similarity methods could be classified based on the underlying representation of molecular shape. The similarity between these shape representations is evaluated by employing various similarity metrics. A schematic overview of the similarity calculation between a query and database molecules is given in **Figure 1**. In the following paragraphs, we will outline commonly utilized shape representations with their advantages and disadvantages. As this review is targeted toward a broader readership, we will only provide an overview of the methods. For algorithmic details and mathematics behind each method, original publications may be referred.

### Atomic Distance-Based Descriptors

These methods are based on the assumption that the shape of a molecule can be described by the relative positions of its atoms. The similarity between molecules can be then calculated by comparing the corresponding distributions of atomic distances. As these descriptors only require the computation of interatomic distances in compounds, these methods are faster compared to other shape comparison methodologies. Additionally, these methods do not require the alignment between two molecules for shape comparison. An overview of various atomic distance-based methods is given in **Table 1** highlighting their availability as well as their advantages and disadvantages. One of the earlier atomic distance-based shape comparison method was based on atom triplet distances (Bemis and Kuntz, 1992). This method considered each molecule as a collection of three atom sub-molecules. The atom triplet triangle perimeters were used to generate shape histograms which were then utilized to compare the shape of molecules. This method however has a few limitations. It is difficult to select bin size suitable for all molecules. Each molecule typically generates 300–500 atom triplets and storing them require large space especially when comparing a large database of molecules. To deal with this limitation, another atom triplet based molecular shape comparison method was developed where a 2,048 bits long single condensed triplet shape signature was employed to represent the entire set of triplets in each molecule (Nilakantan et al., 1993). A signature of the query molecule is first compared with the already stored signatures of database molecules. Then only the compounds with adequately similar signatures are compared



in detail by generating all triplets. Although this method was efficient but there was a risk of missing similar compounds due to the use of highly reduced signature representation. Another group developed molecular descriptors based on atom triplet triangles, angular information from surface point normal and local curvature to facilitate shape comparisons (Good et al., 1995). However, these descriptors have limited discriminating power and require large disk space for storage.

Ultrafast shape recognition (USR) (Ballester and Richards, 2007a,b; Ballester, 2011) is possibly the most popular atomic

distance-based method developed to overcome alignment and speed problems associated with shape similarity methods. This method also uses the relative positions of atoms to describe the shape of a molecule. The schematic overview of USR method is given in **Figure 2** along with an example of the shape similarity evaluation. USR calculates the distribution of all atom distances from four reference positions: the molecular centroid (ctd), the closest atom to molecular centroid (cst), the farthest atom from molecular centroid (fct) and the atom farthest away from fct (fff). Consecutively, the first three statistical moments (mean,

**TABLE 1** | Atomic distance based shape comparison methods.

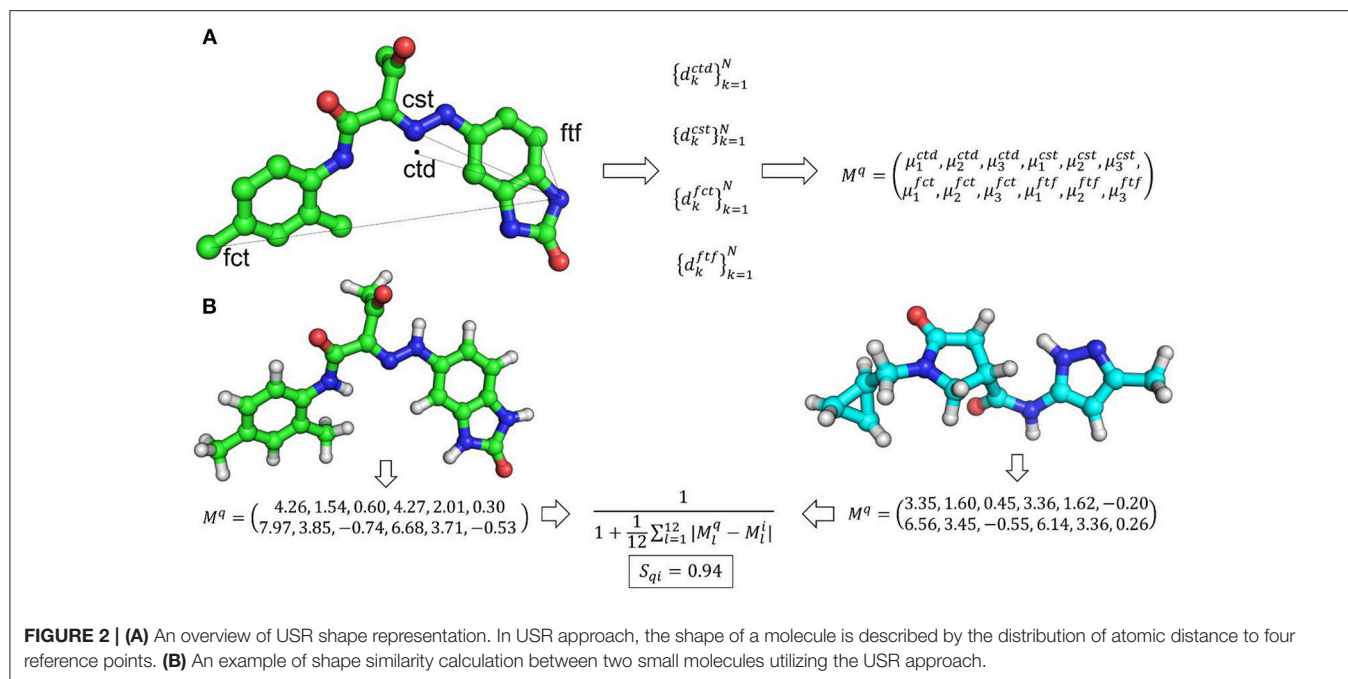
Method	Description	Availability	References
USR	Extremely fast shape comparison method. Webserver can screen about 55 million conformers in 1 s. Different functional groups and enantiomers not recognized.	A ligand-based virtual screening webserver, USR-VS is available at <a href="http://usr.marseille.inserm.fr">http://usr.marseille.inserm.fr</a>	Ballester and Richards, 2007a,b; Ballester, 2011; Li et al., 2016
USR+MACCS	Functional group information added to USR. Enantiomers not recognized.	Available on request	Cannon et al., 2008
CSR and USR:OptIso	Chiral shape recognition. Optical isomerism descriptors added to USR.	Developed by University of Oxford, UK. May be available from Oxford Drug Design company ( <a href="https://www.oxforddrugdesign.com">https://www.oxforddrugdesign.com</a> ), Another implementation USR:OptIso is available at <a href="https://code.google.com/archive/p/usrchirality/">https://code.google.com/archive/p/usrchirality/</a>	Armstrong et al., 2009; Zhou et al., 2010
Electroshape	Chiral shape recognition, include descriptor for charge and lipophilicity.	Developed by University of Oxford, UK. May be available from Oxford Drug Design company ( <a href="https://www.oxforddrugdesign.com">https://www.oxforddrugdesign.com</a> ), A similarity search webserver including Electroshape implementation is available at <a href="http://www.swissimilarity.ch">http://www.swissimilarity.ch</a>	Armstrong et al., 2010, 2011; Zoete et al., 2016
UFSRAT	Pharmacophoric constraints by including atom-type information.	Developed by University of Edinburgh. Server available at <a href="http://opus.bch.ed.ac.uk/ufsrat/index.php">http://opus.bch.ed.ac.uk/ufsrat/index.php</a>	Shave, 2010; Lim et al., 2011; Shave et al., 2015
USRCAT	Included CREDO atom-type information.	A python implementation of the method using RDKit toolkits is available from <a href="https://bitbucket.org/aschreyer/usrcat">https://bitbucket.org/aschreyer/usrcat</a>	Schreyer and Blundell, 2009, 2012; Li et al., 2016
ACPC	Method uses autocorrelation of partial charges. High throughput virtual screening possible. Cannot distinguish a molecule from its enantiomer.	Developed by Laboratory for Structural Bioinformatics, Centre for Biosystems Dynamics Research, RIKEN and is available from <a href="http://www.riken.jp/zhangiru/software.html">http://www.riken.jp/zhangiru/software.html</a> .	Berenger et al., 2014

variance, and skewness of distribution) are calculated from each of these distributions. Hence, each molecule has a vector of twelve descriptors to describe its 3D shape. Finally, the similarity between shapes of two molecules is calculated through an inverse of the Manhattan distance of these 12 values:

$$S_{qi} = \frac{1}{1 + \frac{1}{12} \sum_{l=1}^{12} |M_l^q - M_l^i|}$$

where  $M^q$  and  $M^i$  are vectors of shape descriptors for query and  $i^{\text{th}}$  molecule, respectively. The performance of USR was retrospectively compared with EigenSpectrum Shape Fingerprints (EShape3D) where better mean enrichment for USR was observed (Ballester et al., 2009). A retrospective comparison with three state-of-the-art shape similarity methods: EShape3D, shape signatures and ROCS revealed that USR is 1,546, 2,038, and 14,238 times faster than each one of them respectively (Ballester and Richards, 2007a). A web implementation of USR (USR-VS) is an extremely fast way of carrying out shape similarity calculations (Li et al., 2016). USR-VS is capable of screening 55 million 3D conformers per second and can calculate similarity scores for 94 million 3D conformers in about 2 s. This extremely fast speed is achieved as the features for all 3D conformers are preloaded into the memory. Moreover, the multi-threaded design of the webserver and alignment-free nature of USR method also contributed to such a high computational efficiency. A hardware implementation of USR has been shown to achieve two-fold speed gains over standard CPU based implementation

of USR (Morro et al., 2018). In this implementation, a computing technique, Spiking Neural Networks, has been adapted utilizing Field-Programmable Gate arrays to allow highly parallelized implementation of USR. Prospective application of USR in the identification of arylamine *N*-acetyltransferases, protein arginine deiminase 4 (PAD4), falcipain 2, phosphatases of regenerating liver (PRL-3), p53-MDM2 inhibitors and for phenotypic targets such as colon cancer cell lines established the real-world applicability of USR (Li et al., 2009; Ballester et al., 2010, 2012; Teo et al., 2013; Hoeger et al., 2014; Patil et al., 2014). As USR is an ultrafast, purely shape-based similarity method, several methods augmenting the original USR capabilities were developed. These include a method where USR was combined with MACCS key encoding the topological information of small molecules (Cannon et al., 2008). To clearly distinguish between enantiomers, methods complementing USR with optical isomerism descriptors were developed (Armstrong et al., 2009; Zhou et al., 2010). Electroshape, a USR variant appended partial charge and atomic lipophilicity (alogP) as additional molecular properties to account for electrostatics and lipophilicity along with shape recognition (Armstrong et al., 2010, 2011). A web implementation of Electroshape is available at SwissSimilarity (Zoete et al., 2016). AutoCorrelation of Partial Charges (ACPC) also utilized partial charges with atomic distances to measure similarity between two molecules (Berenger et al., 2014). The method uses an autocorrelation function and a point charge model to encode all atoms of a molecule into two vectors that are rotation translation invariant. Another implementation of



**FIGURE 2 | (A)** An overview of USR shape representation. In USR approach, the shape of a molecule is described by the distribution of atomic distance to four reference points. **(B)** An example of shape similarity calculation between two small molecules utilizing the USR approach.

USR method is Ultrafast Shape Recognition with Atom Types (UFSRAT) which introduced pharmacophoric constraints to USR by incorporating atom type information (Shave, 2010; Lim et al., 2011; Shave et al., 2015). UFSRAT is capable of very fast comparison of query molecule with small molecule libraries from several major chemical vendors via its webserver (Table 1). Application of UFSRAT method in the discovery of MDM2, PRL-3, FK506-Binding Protein 12, kynurenine 3-monooxygenase and 11 $\beta$ -hydroxysteroid dehydrogenase type 1 (11 $\beta$ HSD1) inhibitors demonstrated its utility in key areas of drug discovery such as cancer, Alzheimer's disease, inflammation and type-II diabetes. (Hoeger et al., 2014; Houston et al., 2015; Shave et al., 2015, 2018). Another similar implementation, USRCAT utilized CREDO atom types to encode pharmacophoric information to USR (Schreyer and Blundell, 2009, 2012). USRCAT not only retained USR abilities to retrieve hits with low structural similarity but also demonstrated improved performance over the original USR implementation.

Atomic distance or descriptor-based methods are widely used due to their ability to quickly compare the shapes of query molecules with large small molecule libraries. A fast comparison of a wide range of chemical space increases the chances of finding novel hits. These methods are not only computationally efficient but also have produced excellent hit rates as revealed from several successful prospective studies against a wide range of molecular and non-molecular targets. Moreover, they are also capable of retrieving chemical scaffolds which are different from the query molecule, thus allowing scaffold hopping. As atomic distance-based shape similarity approaches are alignment-free, the visual inspection of shape similarity may be sometimes challenging especially for molecules which have low structural similarity. Selection of the right query compound is a key component of atomic distance-based shape similarity methods and their

performance depends on optimal query selection. Hit rate can be improved by employing multiple queries and increasing the diversity of selected hits. Moreover, clustering based on shape similarity could be utilized to understand how different chemotypes arrange in binding pockets and thereby generating consensus queries (Pérez-Nueno et al., 2008; Pérez-Nueno and Ritchie, 2011) to improve virtual screening performance and reducing false positives.

## Atom-Centered Gaussian-Based Shape Similarity Methods

Among many methods of describing the molecular shape of a molecule, hard sphere (Connolly, 1985; Masek et al., 1993) and Gaussian sphere (Grant and Pickup, 1995; Grant et al., 1996) are two most widely adopted models. Both of these models describe the shape in terms of the volume of a molecule. Two molecules will possess similar shape if they have similar volume. Hard sphere model represents a molecule by a set of merged spheres where each sphere serves as an atom with its van der Waals radius. The volume of a molecule can be calculated by a formula that describes the union of a number of sets and their intersection. Although the analytical expression of the volume and its derivatives is reported in the original publication (Masek et al., 1993), it is not easy to implement as the formulas become very complicated with increasing number of intersections. Gaussian sphere model (Grant and Pickup, 1995, 1997; Grant et al., 1996) represents a molecule using a set of overlapping Gaussian spheres and measures the integral volume over all overlapping Gaussians. In this model, each intersection is expressed as the integral of a set of overlapping atom-centered Gaussian spheres and the volume of a molecule is described based on the inclusion-exclusion principle. Analytical expression for the volume calculation is given in the original publication

which describes highly accurate volume calculation up to sixth order intersections (Grant and Pickup, 1995). The authors also proposed comparing shapes of two molecules by numerically optimizing the overlap between two molecules (Grant et al., 1996).

Several methods based on Gaussian overlays were developed to measure the shape similarity between two molecules. An overview of these methods is presented in **Table 2**. Among these, Rapid Overlay of Chemical Structures (ROCS) is undoubtedly the most widely used method that utilizes Gaussian functions to measure the shape similarity between two molecules (Rush et al., 2005; Hawkins et al., 2007). ROCS algorithm is based on the original Gaussian overlay approach that finds and quantifies the maximum volume overlap between two molecules (Grant and Pickup, 1995; Grant et al., 1996). An overview of ROCS shape similarity calculation is given in **Figure 3**. However, to improve the efficiency of volume overlap calculations, it incorporated several modifications to the original implementation. ROCS ignores hydrogens for the volume calculations and uses equal radii for all heavy atoms. Furthermore, ROCS utilizes only the first order terms of shape density function. ROCS employs Tanimoto (Rogers and Tanimoto, 1960) and Tversky (Tversky, 1977) correlation coefficients as similarity metrics to calculate the overlap between two molecules which are defined as:

$$Tanimoto_{a,b} = \frac{O_{a,b}}{O_a + O_b - O_{a,b}}$$

$$Tversky_{a,b} = \frac{O_{a,b}}{O_{a,b} + \alpha O_a + \beta O_b}$$

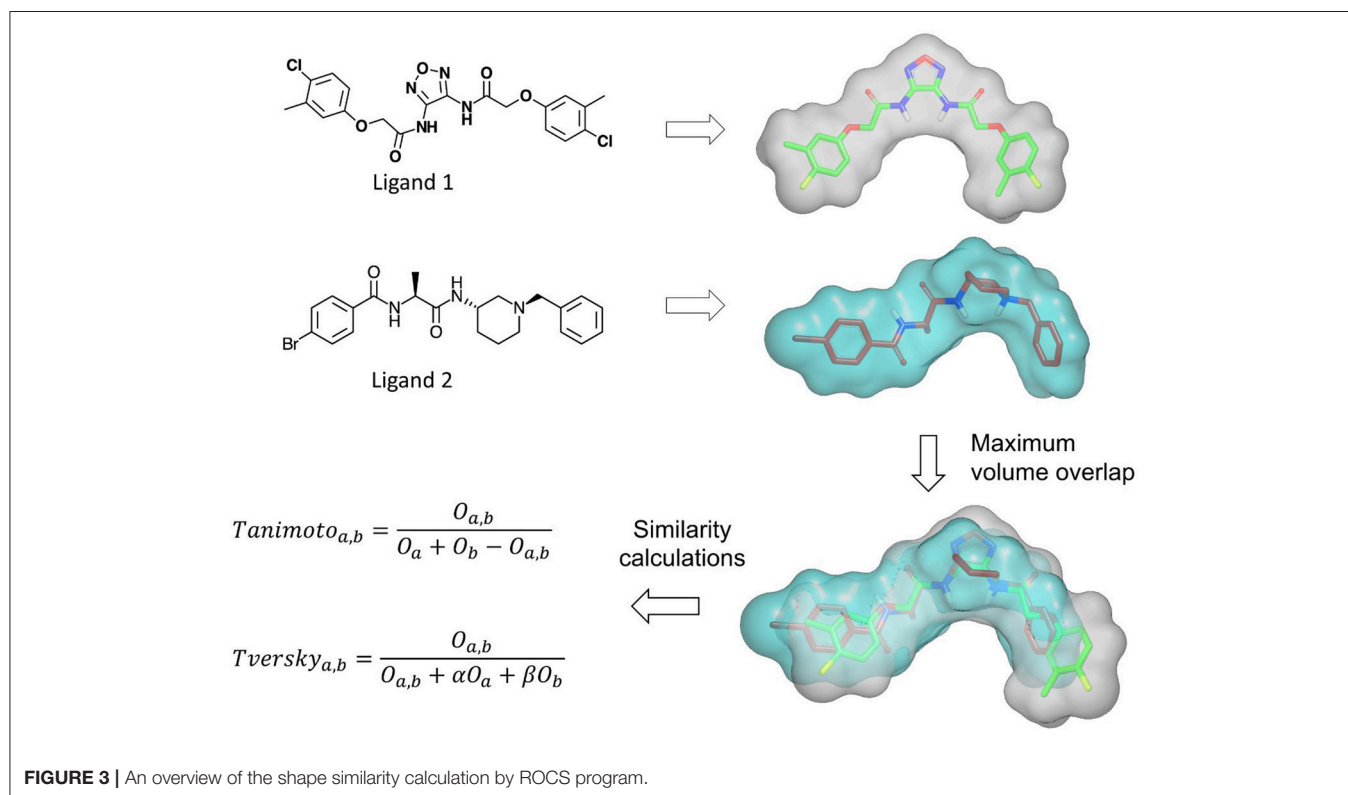
where  $O_{a,b}$  is the volume overlap between molecules  $a$  and  $b$ ,  $O_a$  is the volume of molecule  $a$  and  $O_b$  is the volume of molecule  $b$ .  $\alpha$  and  $\beta$  are parameters for Tversky index. ROCS also considers chemical complementarity by including the chemical features to improve shape-based superposition. ROCS has been successfully employed in various drug discovery campaigns such as in the identification of small molecules inhibitors (Kumar et al., 2014b), to scaffold hop from one chemical class to another (Kumar et al., 2016), to rescore docking generated poses (Kumar and Zhang, 2016a) and to predict binding poses and ranking of inhibitors (Kumar and Zhang, 2016b,c). ROCS can routinely perform shape and chemical feature comparisons of about 600–800 conformers per second on a modern CPU. Although this speed is reasonable for alignment-based shape similarity methods, it takes several hours to screen a moderately sized virtual screening library. To facilitate large scale shape comparison, e.g., to screen large small molecule libraries within minutes, FastROCS (<https://www.eyesopen.com/molecular-modeling-fastrocs>), a GPU implementation of ROCS has been developed that increased the shape comparison speed by about three orders of magnitude over its CPU implementation. FastROCS is capable of processing up to a million conformers per second on a single NVIDIA Tesla K20 GPU (<https://docs.eyesopen.com/toolkits/python/fastrocs/structure.html>). PAPER, an open source GPU implementation of ROCS algorithm, also demonstrated speed acceleration up to two orders of magnitude on an NVIDIA GeForce GTX 280 GPU over its open source CPU

implementation on a Intel Xeon E5345 CPU (Haque and Pande, 2010). MolShaCS is another method that engages Gaussian description of shape to evaluate molecular similarity between two molecules (Vaz de Lima and Nascimento, 2013). In addition to shape, MolShaCS utilizes Gaussian description of charge distribution to optimize overlays and similarity computations using Hodgkin's index (Hodgkin and Richards, 1987; Good et al., 1992). It was able to process 21 compounds per second, which seems to be a quite impressive speed for computers of that time. As Gaussian overlay based methods require precise alignment for the calculation of shape similarity, several groups employed approaches such as pharmacophore and field based methods to generate initial alignment. SHAFTS (*SHApe-FeaTure Similarity*) (Liu et al., 2011) adopted pharmacophoric point triplets and least square fitting to generate initial alignment. A weighted sum of pharmacophoric fit and volume overlap was then used to assess shape similarities. Phase Shape (Sastry et al., 2011) also employed the same concept of atom distribution triplets to generate initial alignments which were then refined by maximizing the volume overlap. Phase Shape is capable of performing shape comparisons of about 500 conformers per second. Reminiscent of Shape and Electrostatic Potential (ShaEP) (Vainio et al., 2009) also resembles SHAFTS and Phase Shape as it utilizes a hybrid approach that combined field-based methods with volumetric methods to estimate molecular similarity. ShaEP borrowed a graph matching algorithm to generate initial superposition. Molecular graphs represented shape and electrostatic potential at points close to molecular surface. The method then optimized the initial alignment by maximizing the volume overlap calculated through Gaussian functions. Another similar method, SimG (Cai et al., 2013), adopted downhill simplex method (Nelder and Mead, 1965) to evaluate the similarity in shape and chemical features of a molecule and a binding pocket or ligand. SimG shape similarity method possessed advantage over other methods described here in the sense that it is capable of performing shape similarity evaluations between a ligand and a binding pocket. SABRE method (Hamza et al., 2012, 2013) introduced two modifications to the original Gaussian overlay based shape similarity implementation. First, it utilized reduced chemical structures by removing the functional group not present in query to generate initial alignments. Reduced chemical structures were subsequently replaced by full structures and the initial alignments were refined by rigid-body translation and rotation using steepest descent to produce shape density overlap with the query. Secondly, to avoid bias for large sized ligands when using Tanimoto similarity metric, a new scoring function Hamza-Wei-Zhan (HWZ) score was developed. An extension to SABRE method enabled its utility in chemogenomics area (Wei and Hamza, 2014). Shapelets (Proschak et al., 2008) is unlike any other Gaussian overlay based shape comparison method. It describes the shape of a molecule by decomposing its surface into discrete patches. This 3D graph representation can then be used for either full or partial shape similarity evaluations.

In most Gaussian function based overlay methods shape density of a molecule is described as the sum of shapes of individual atoms which sometimes results in the overestimation of the volume, for example, in molecules where some atoms

**TABLE 2** | An overview of commonly used Gaussian overlay based shape comparison methods.

Method	Description	Availability	References
ROCS	Fast Gaussian overlay based shape comparison. Widely used shape based virtual screening tool. GPU version also available.	Developed by OpenEye Scientific Software ( <a href="https://www.eyesopen.com">https://www.eyesopen.com</a> ). Commercial.	Rush et al., 2005; Hawkins et al., 2007
PAPER	Accelerates large scale virtual screening experiments. Parallel implementation on NVIDIA GPUs.	Developed by Stanford University. Open source. Available from SimTK at <a href="https://simtk.org/projects/paper">https://simtk.org/projects/paper</a>	Haque and Pande, 2010
MolShaCS	Uses Gaussian description of shape and charge. Hodgkin like similarity metric. Molecules are considered rigid.	Developed by University of Sao Paulo, Brazil. Open source tool available at <a href="https://code.google.com/archive/p/molshacs/downloads">https://code.google.com/archive/p/molshacs/downloads</a>	Vaz de Lima and Nascimento, 2013
SHAFTS	It combines shape similarity with pharmacophoric features. Employs a hybrid similarity metric combining shape and chemical similarity. Suitable for large scale virtual screening.	Developed by Shanghai Key Laboratory of New Drug Design, East China University of Science & Technology, Shanghai, China. Available for download from <a href="http://lilab.ecust.edu.cn/home/resource.html">http://lilab.ecust.edu.cn/home/resource.html</a>	Liu et al., 2011
Phase Shape	Uses atom triplets to generate initial alignments which are refined by Gaussian overlay.	Developed by Schrodinger. ( <a href="https://www.schrodinger.com">https://www.schrodinger.com</a> ). Commercial.	Sastry et al., 2011
ShaEP	Generate consensus shape pattern based on structural features of known ligands.	Developed by Abo Akademi University, Finland. Free for Academics. Available from the Abo Akademi University at <a href="http://users.abo.fi/mivainio/shaep/index.php">http://users.abo.fi/mivainio/shaep/index.php</a>	Vainio et al., 2009
SimG	Uses downhill simplex method to evaluate shape and chemical similarity between two molecules. Comparison of ligand and binding pocket shape or chemical similarity is also possible.	Developed by Shanghai Key Laboratory of New Drug Design, East China University of Science & Technology, Shanghai, China. Available for download from <a href="http://lilab.ecust.edu.cn/home/resource.html">http://lilab.ecust.edu.cn/home/resource.html</a>	Cai et al., 2013
SABRE	Uses consensus shapes to generate initial alignments which are later refined by rigid-body rotations and translations.	Academic license is available on request	Hamza et al., 2012, 2013
WEGA	Uses a weighted Gaussian function to improve the accuracy of first order approximation. A GPU implementation (gWEGA) is also available for large scale virtual screenings.	Developed by Research Center for Drug Discovery, Sun Yat-sen University, China. Academic license is available on request at <a href="http://www.rcdd.org.cn/home/program.html">http://www.rcdd.org.cn/home/program.html</a> .	Yan et al., 2013



highly overlap with others in the vicinity. Weighted Gaussian algorithm (WEGA) method (Yan et al., 2013) puts forward a modification where a weight factor is introduced for every atom. This weight factor reflects the crowdedness of an atom with its neighbors. The shape density of a molecule is represented by the linear combination of weighted atomic Gaussian functions. Utilizing this modification, WEGA method demonstrated improved shape similarity and virtual screening performance. The speed of WEGA shape similarity calculations varies with the size of query and database compounds. For an average drug-like query, WEGA can process 1,000–1,500 conformations per second (Yan et al., 2013). A GPU implementation of this method (gWEGA) has also been developed that reported a virtual screening speed increase by two orders of magnitude on one NVIDIA Tesla C2050 GPU over its CPU implementation on a quad-core Intel Xeon X3520 CPU (Yan et al., 2014). Another WEGA derivative, HybridSim proposed a hybrid metric combining 2D fingerprints with WEGA shape similarity and demonstrated improved virtual screening performance over standalone 2D fingerprint and shape similarity methods (Shang et al., 2017).

Overall, atom-centered Gaussian-based shape similarity methods present many advantages over other shape similarity methods. Although not as fast as distance based methods, these methods are fast enough for large scale virtual screenings. The major advantage with atom-centered Gaussian-based shape similarity methods is the visualization. The visualization of shape similarity between two molecules is immensely helpful in deriving the structure activity relationship for the optimization and for scaffold hopping. A majority of these methods address the problem of ligand flexibility by utilizing conformational ensemble. However, in some cases it may not be trivial to sample all possible conformations, e.g., natural products. Moreover, several top performing conformational generation methods face difficulty in modeling the correct conformation of some molecules, e.g., macrocycles, peptidomimetics etc. Another limitation with these methods is that their performance highly depends upon the query molecule and choosing the right query is a critical component of a shape-based virtual screening campaign (Kirchmair et al., 2009). Despite these limitations, atom-centered Gaussian overlay based methods are the most widely used shape similarity methods. They have provided many successful examples demonstrating their utility in various areas of drug discovery which will be discussed later in this manuscript.

## Surface Based 3D Shape Similarity Comparison Methods

Molecular surface is another way of depicting the shape of a molecule. Comparison of molecular surfaces based on their shapes can reveal similarity in their physical and biological properties. There are many ways to describe the surface of a molecule. Precise definitions such as surface based on quantum mechanical wave functions are not practical especially for large molecules (Mezey, 2007). Surface definitions such as solvent-accessible surface (Lee and Richards, 1971; Connolly, 1983) and van der Waals surface are more practical and much easier to

calculate. Some studies employed alpha shapes (Edelsbrunner et al., 1983; Edelsbrunner and Mücke, 1994; Edelsbrunner, 1995) which is a coarse representation of Connolly surface (Connolly, 1983) to describe the shape of a molecule (Wilson et al., 2009). Alpha shapes of a set of points “S” are generalization of convex hull and utilize a parameter,  $\alpha$  to describe the shape with varying levels of details. For large  $\alpha$  values, the alpha shape is equivalent to convex hull and shape feature details such as concavities and voids started to appear with decrease in  $\alpha$  value. The alpha shape method has been applied to represent and compare shapes of 3D molecules (Wilson et al., 2009).

Shape signatures or shape histograms offer another representation of molecular shape that can be used to explore 3D volume of a molecule confined by the solvent accessible surface (Zauhar et al., 2003; Meek et al., 2006). Shape signatures are probability distribution histograms borrowed from a computer graphics technique, ray-tracing. In this method, a ray is initiated within a molecule bound by its solvent accessible surface. Propagation of a ray trace inside of the triangulated solvent accessible surface is recorded as probability distribution histograms. The histograms for query and any other molecule can be easily compared using the following metrics:

$$L_1^{1D} = \sum_i |H_i^1 - H_i^2|$$
$$L_1^{2D} = \sum_i \sum_j |H_{ij}^1 - H_{ij}^2|$$

where 1D represents the probability distribution of ray-trace lengths only while 2D represents ray-trace lengths in combination with additional molecular property such as electrostatic potential. Shape signature encodes shape, molecular size and surface charge distribution of a molecule and can be utilized to compare the histogram of a query molecule with the pre-generated histograms of small molecule libraries. The utility of shape signatures as a virtual screening approach has been demonstrated in several studies (Nagarajan et al., 2005; Wang et al., 2006; Hartman et al., 2009; Ai et al., 2014; Werner et al., 2014). As shape signature based similarity comparisons are fast and do not require the alignment of molecules, they are capable of screening millions of molecules in a short time. In addition to shape similarity, shape signatures also allow shape complementarity comparisons against a receptor binding pocket. Although shape similarity calculations with shape signature have been effectively used in many inhibitor discovery efforts, the high number of false positives is a concern especially for large and complex queries. To cope with these drawbacks, a few modifications to the original methods were reported. These include fragment-based shape signature (FBSS) (Zauhar et al., 2013) and inner distance shape signature (IDSS) (Liu et al., 2009, 2012). FBSS involves the generation and comparison of shape signatures for fragments in the molecules. IDSS utilizes inner distance which is the shortest path between landmark points within the molecular shape. IDSS has been shown to be especially useful in case of flexible molecules as it is insensitive to shape deformation of flexible molecules.



**TABLE 3** | An overview and availability of a few surface-based shape comparison methods.

Method	Description	Availability	References
SURFCOMP	Molecular surface is divided into patches and corresponding patches are identified using geometrically invariant descriptors and physicochemical properties.	Available on request.	Hofbauer et al., 2004
ParaFit	Performs 3D superposition and surface property comparison. Electronic surface properties are calculated using ParaSurf program. Spherical harmonics expansion coefficients of molecular surface are used.	Developed by CEPOS <i>in silico</i> Ltd. Commercial or Academic license can be obtained at <a href="http://www.ceposinsilico.de/">http://www.ceposinsilico.de/</a>	Mavridis et al., 2007
SHeMS	Uses spherical harmonics description of shape. Weights of spherical harmonics expansion coefficients are optimized using a genetic algorithm.	Developed by Shanghai Key Laboratory of New Drug Design, East China University of Science & Technology, Shanghai, China. Obtained by contacting Prof. Honglin Li at <a href="http://lilab.ecust.edu.cn/home/resource.html">http://lilab.ecust.edu.cn/home/resource.html</a>	Cai et al., 2012
HPCC	Combined spherical harmonics shape comparison with pharmacophoric features. Tanimoto similarity coefficients for shape and chemical similarity are added to evaluate similarity between two molecules.	Developed by Harmonic Pharma. May be available from <a href="https://www.harmonicpharma.com/oncology/">https://www.harmonicpharma.com/oncology/</a>	Karaboga et al., 2013
3DZD	Uses 3D Zernike descriptors which are extension of spherical harmonics. Rotation translation invariant.	Developed by Kihara Bioinformatics laboratory at Purdue University, USA. Several implementations of 3DZD are available either as standalone program or web-server at <a href="http://kiharaab.org/contact.php">http://kiharaab.org/contact.php</a>	Sael et al., 2008a, Venkatraman et al., 2009a

Several methods employed local surface shape similarity to align and estimate the similarity between molecules. One such method applied subgraph isomorphism to molecular surface comparison (Cosgrove et al., 2000). In this method, molecular surface was represented by patches of the same shape. Alignment between two molecules was obtained by using a clique-detection algorithm to obtain overlapping patches. Quadratic shape descriptors (Goldman and Wipke, 2000) exploited a similar concept where molecular surface was divided into a series of patches. Each patch was represented by geometrically invariant descriptors such as the normal, the shape index and the principle curvatures which were then used to identify similar patches. SURFCOMP (Hofbauer et al., 2004) further applied several filters such as surrounding shape and physicochemical properties to identify corresponding patches on surfaces of two molecules (Table 3).

Spherical harmonics (SH) based representations which are expansion of SH functions also allow quantitative description of molecular shapes (Max and Getzoff, 1988). In this representation, shapes are expressed as functions on a unit sphere. Each point on a unit sphere surface is described by its spherical coordinates  $(r, \theta, \phi)$  and setting  $f(\theta, \phi) = r$ , where  $r$  is a radial function encoding the distance of surface points from a chosen origin. This function can be determined by deriving an expansion of SH basis function given by:

$$r(\theta, \phi) = \sum_{l=0}^L \sum_{m=-l}^l c_{l,m} Y_l^m(\theta, \phi)$$

where  $Y_l^m(\theta, \phi)$  is the SH basis function for degree  $l$  and order  $m$ .  $c_{l,m}$  are coefficients of SH function.  $L$  is the chosen limit to get desired resolution of the surface. The number of terms in the function depends upon this limit as a value of  $L$ , which yields  $(L+1)^2$  terms. In general, SH are not rotation translation invariant as magnitude of  $c_{l,m}$  change based on the rotation of

$r(\theta, \phi)$ . Hence, prior alignment is necessary before comparing the shape of molecules. Efforts were also made to make SH rotation translation invariant (Kazhdan et al., 2003; Mak et al., 2008), however, these modifications increase the number of terms thereby increasing the complexity of SH.

About two decades ago, it was shown that SH functions could be applied to estimate the 3D molecular similarity between two macromolecules (Ritchie and Kemp, 1999). Since then, it has been successfully applied in virtual screening (Cai et al., 2002; Mavridis et al., 2007), protein structure comparisons (Tao et al., 2005; Gramada and Bourne, 2006), protein-ligand docking (Ritchie and Kemp, 2000; Lin and Clark, 2005; Yamagishi et al., 2006), binding pocket similarity comparison (Morris et al., 2005) etc. Additionally, several groups utilized variations of SH to compare the shapes of small molecules. The first implementation of SH to compare shapes of small molecules opened the way for many applications ranging from virtual screening to quantitative structure-activity relationship (QSAR) model building (Lin and Clark, 2005). SpotLight program utilizes SH to superpose and classify small molecules (Mavridis et al., 2007). To enable high throughput virtual screening, the vector interpretation of SH coefficients was used to construct rotation translation invariant fingerprints (RIFs) which were compared using a distance score (Mavridis et al., 2007). In this method, rotational invariance was gained by binning together the SH coefficients of the same order. This method was later developed as ParaFit (<http://www.ceposinsilico.de>) (Table 3). In another study, SH based molecular surface was decomposed and the norm of decomposition coefficients were used to describe the molecular shape (Wang et al., 2011). Norms of decomposition coefficients are partially rotation translation invariant enabling large scale comparison. The performance of this method was retrospectively demonstrated and was also prospectively applied in the discovery of cyclooxygenase-1 and cyclooxygenase-2 inhibitors. SHeMS method utilizes genetic algorithm to optimize the weights

of SH expansion coefficients for a reference set (Cai et al., 2012). Through optimization of weights, SHeMS demonstrated improved performance over original SH implementation and USR method. To facilitate measurement of similarity between sets of compounds, many shape similarity methods were complemented with physicochemical properties. Harmonic pharma chemistry coefficient (HPCC) method combined SH shape representation with pharmacophoric features (Karaboga et al., 2013). In HPCC method, SH surfaces are discretized as triangle meshes which are assigned pharmacophoric features. Tanimoto similarity for both shape and pharmacophore features is calculated separately between query and test molecules. A combo score is finally calculated by adding Tanimoto scores for shape and chemical overlay. HPCC method demonstrated improved performance for the combo approach over utilizing the shape alone.

In several studies, 3D-Zernike descriptors (3DZD) (Novotni and Klein, 2003), which are the extension of SH were employed to compare the shapes of molecules and cryoEM maps (Figure 4 and Table 3). 3DZD differs from SH in terms of their mathematical description. 3DZD can model molecular shape precisely as compared to SH which can only model single valued or star-shape surfaces. They are rotation translation invariant, whereas SH depends on the orientation of the molecule. Although rotation translation invariant SH descriptors have been developed (Kazhdan et al., 2003), the number of terms are much higher in SH descriptors. 3DZD is also suitable to represent other properties on molecular surfaces such as hydrophobicity and electrostatic potential (Sael et al., 2008a). In the drug discovery area, 3DZD was initially applied to compare shapes of protein molecules (Sael et al., 2008b; Figure 4A). Later, the concept was extended to measuring shape similarity and small molecules (Venkatraman et al., 2009a) and between binding pockets (Kihara et al., 2009; Venkatraman et al., 2009b; Figures 4B,C). In 3DZD method, 3D Zernike function is described as:

$$Z_{nl}^m(r, \theta, \phi) = R_{nl}(r)Y_l^m(\theta, \phi)$$

where  $Y_l^m(\theta, \phi)$  is the SH basis function while  $R_{nl}(r)$  is the radial function. Zernike moments are calculated using the following equation:

$$F_{nl}^m = \frac{3}{4\pi} \int f(x) \overline{Z_{nl}^m(x)} dx$$

As Zernike moments are not rotationally invariant, so to make them rotation translation invariant, they are expressed as norm  $F_{nl}^m$  which is known as 3DZD. Shape similarity between two molecules based on 3DZD is compared using the following metrics:

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \in [0, \infty]$$

$$\text{Pearson } r = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{n \sum X_i^2 - (\sum X_i)^2} \sqrt{n \sum Y_i^2 - (\sum Y_i)^2}} \in [-1, 1]$$

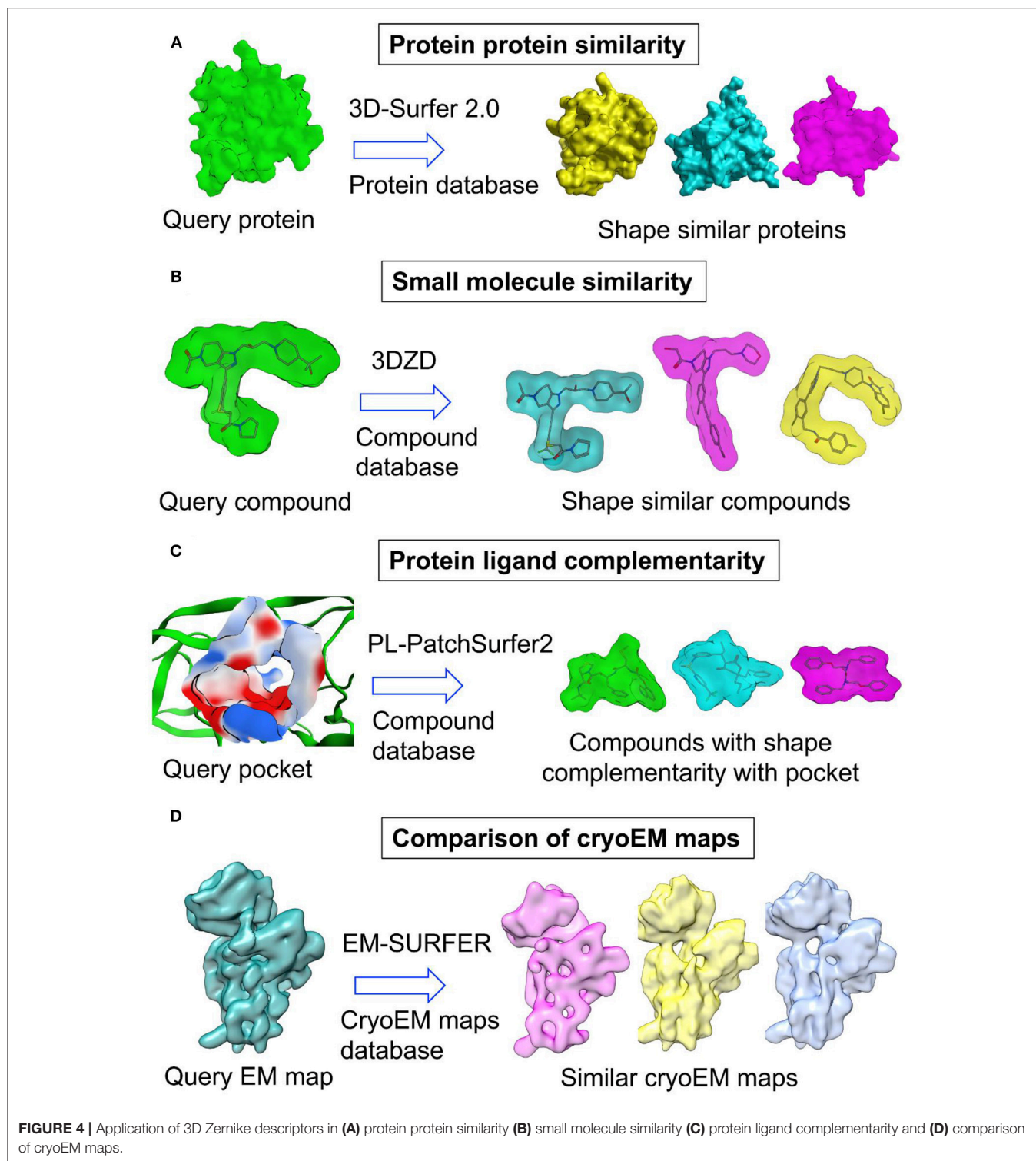
$$\text{Manhattan distance} = \frac{1}{1 + \frac{\sum_{i=1}^n |X_i - Y_i|}{N}} \in [0, 1]$$

Ligand 3D shape similarity comparison using 3DZD is fast and rotation translation invariant. As no alignment step is required for comparison, it can be utilized as a virtual screening tool to filter a database of compounds based on shape similarity with a query molecule.

Overall, surface-based shape similarity methods present attractive options for comparing the shapes of small molecules and macromolecules. They were quite successful in estimating the global and local similarities between macromolecules. However, most of these methods are still in infancy as far as small molecule shape comparison is concerned. Several reasons may have contributed to the lack of interest from researchers in accepting these methods as small molecule shape comparison tools. Surface-based methods such as SH and 3DZD are mathematically complex and involve inclusion of many terms to fully capture the shape of a molecule. Moreover, they are slow in comparison to atomic distance-based shape description and comparison methods while their accuracy in retrieving compounds similar in shape to a query does not match Gaussian overlay-based shape similarity methods. Further, while these methods capture very well the global shape of a molecule, the local shape similarity is not represented comprehensively which is very critical in comparing the shapes of small molecules. However, these methods present several new areas of shape comparison such as comparing shape of ligands with that of binding pockets which may be of immense utility for structure-based design.

## Other Shape Similarity Approaches

There are many other approaches of shape representation and methods of similarity measurement in addition to these described above. Another way of representing molecular shape is to use molecular descriptors. Several shape-based descriptors have been traditionally used to compare small molecules and develop QSAR models. These descriptors mostly represent shape implicitly with other properties such as size, symmetry and atom distribution. These include Weighted Holistic Invariant Molecular (WHIM) descriptors of shape (Gramatica, 2006), shape indices, descriptors for moments of the distribution of molecular volume (Mansfield et al., 2002). Most of molecular descriptors are alignment independent, however, some such as moments of the distribution of molecular volume require superposition of molecules. Comparative Molecular-Field Analysis (CoMFA) (Cramer et al., 1988) is a widely used technique to develop QSAR models and understand SAR for a series of compounds. CoMFA compares a set of molecules by placing them on a grid and calculating potential energy fields. The differences and similarities between molecules are then correlated with differences and similarities in their biological activities. As CoMFA requires molecules to be pre-aligned, the 3D shape similarity of molecules can be obtained based on potential energy fields. A modification of CoMFA approach, Comparative Molecular Moment Analysis (CoMMA) calculates geometric moments from the center of mass, center of charge and center of dipole of a molecule (Silverman and



Platt, 1996). However, superposition of molecules is not required in this approach. Shape of the molecules can also be inferred from structural descriptors such as molecular quantum numbers (MQNs) (Nguyen et al., 2009; van Deursen et al., 2010). The MQN represents counts for 42 structural features such as atom,

ring and bond types, polar groups and topology. MQN system has been used to effectively classify and visualize large libraries of organic molecules such as ZINC, GDB, and PubChem.

Volumetric aligned molecular shapes (VAMS) method (Koes and Camacho, 2014) uses data structures to represent and

compare shapes of 3D molecules. It applies inclusive and exclusive shape constraints to estimate the similarity in shapes of 3D molecules. In VAMS method, the shape of a molecule is represented by solvent-excluded volume calculated from its heavy atoms using a water probe of radius 1.4 Å. Volume is discretized on a grid of 0.5 Å resolution where each point on the grid represents a Voxel or 3D pixel. An oct-tree data structure is used to store voxelized volume. This method requires all the shapes to be pre-aligned to a standard reference coordinates. The conformations of the molecule are aligned using the moment of inertia of heavy atoms. Voxelized shapes are compared using Tanimoto similarity (Rogers and Tanimoto, 1960) where the ratio of number of voxels common in two shapes and number of voxels present in either of the shapes is measured. The performance of VAMS method as a standalone virtual screening tool is not better than many other shape similarity methods, e.g., ROCS, however, VAMS is reasonably fast and could perform a million shape comparisons in about 10 s. Hence, it may be used as a pre-filtering tool for other shape similarity methods. Fragment oriented molecular shape (FOMS) is the extension of VAMS method, where shapes are aligned using fragments (Hain et al., 2016).

## APPLICATION OF SHAPE SIMILARITY METHODS IN DRUG DISCOVERY

### Application in Virtual Screening

Shape similarity attempts to quantify the resemblance between two molecules utilizing several descriptions of molecular shape as described previously. This approach has been successfully utilized as a virtual screening tool to identify molecules similar to a given query from the library of chemicals. Several retrospective studies have been published demonstrating the utility of shape based similarity methods over 2D and other 3D similarity methods (Nagarajan et al., 2005; Renner and Schneider, 2006; Ballester et al., 2009; Giganti et al., 2010; Venkatraman et al., 2010; Ballester, 2011; Hu et al., 2012, 2016). Several studies also presented computational approaches to improve the performance and efficiency of shape comparison methods. One study recommended the selection of a suitable query and incorporation of chemical information such as pharmacophoric features of the query molecule to improve the performance of shape-based virtual screening (Kirchmair et al., 2009). Another study demonstrated that the application of a machine learning method, Support Vector Machine (SVM), to shape comparisons can significantly improve virtual screening efficiency (Sato et al., 2012). The need of automation was further suggested specially to carry out multiple query searches which ensure a diverse hit list (Kalászi et al., 2014).

Apart from retrospective tests, many prospective applications of shape similarity have been published in the literature. In numerous studies, it was employed as the only virtual screening approach to filter and prioritize compounds from a large library to a number small enough for biological testing (Rush et al., 2005; Boström et al., 2007; Freitas et al., 2008; Ballester et al., 2010, 2012; Kumar et al., 2012; Vasudevan et al., 2012; Sun et al.,

2013; Hoeger et al., 2014; Patil et al., 2014; Temml et al., 2014; Chen et al., 2016; Bassetto et al., 2017). Among these studies, the shape based identification of a compound active on colon cancer cell line is quite interesting (Patil et al., 2014). This study employed USR to screen a database of approved drugs. The top virtual screening hit displayed dose dependent inhibition of a colon cancer cell line. This study not only repurposed a known drug but also demonstrated the applicability of shape similarity methods for phenotypic screens, e.g., anti-bacterial or anti-fungal drug discovery where molecular target is often unknown. This is especially important considering the fact that most approved drugs come from phenotypic screens (Swinney and Anthony, 2011). In other investigations, it was combined with other ligand-based virtual screening methods or structure based approaches such as molecular docking. Among ligand-based approaches, shape similarity was frequently used in combination with electrostatic similarity. As electrostatic comparison between two small molecules requires precise alignment between them, shape matching was first performed and then followed by the electrostatic potential similarity calculations. This hierarchical combination was utilized to discover a wide variety of binders including enzyme inhibitors (Hevener et al., 2011), mRNA binders (Kaoud et al., 2012), chemical probes (Naylor et al., 2009), protein-protein interaction inhibitors (Boström et al., 2013), SUMO activating enzyme 1 inhibitors (Kumar et al., 2016), and Aurora kinase A inhibitors (Kong et al., 2018).

Although shape-based approaches demonstrated considerable success in ligand-based virtual screening studies, the true potential of the method was realized when it was combined with structure based methods in a hierarchical manner or in a parallel manner. To effectively use shape based virtual screening, several groups employed hierarchical virtual screening (Kumar and Zhang, 2015) where it was coupled with molecular docking. As shape matching calculations are comparatively faster than structure based virtual screening methods, it is generally used during initial steps of a hierarchical virtual screening protocol. This hierarchical combination of shape similarity with molecular docking has been successfully employed in the discovery of type II dehydroquinase inhibitors (Ballester et al., 2012) and that of MDM2 inhibitors (Houston et al., 2015), 11 $\beta$ -hydroxysteroid dehydrogenase 1 inhibitors (Xia et al., 2011), PPAR $\gamma$  partial agonists (Vidović et al., 2011), inhibitors of chemokine receptor 5 (CCR5)-N terminus binding to gp120 protein (Acharya et al., 2011), Grb7-based antitumor agents (Ambaye et al., 2013), fungal trihydroxynaphthalene reductase inhibitors (Brunskole Šveglj et al., 2011), non-steroidal FXR ligands (Fu et al., 2012; Wang et al., 2015), novel SIRT3 scaffolds (Salo et al., 2013), protein kinase CK2 inhibitors (Sun et al., 2013), SUMO conjugating enzyme inhibitors (Kumar et al., 2014a), and chemokine receptor type 4 inhibitors (Das et al., 2015). Combination of shape similarity methods with structure-based methods such as docking provide several advantages. Ultrafast shape comparison methods such as USR can very quickly filter large libraries for compounds that are similarly shaped as known inhibitors. Hence, the time required for docking could be drastically reduced by eliminating compounds that doesn't fit in the binding pocket. Moreover, in case of some proteins the inhibitor activity is driven

by key moieties in compounds, e.g., metal binding groups in case of metalloproteins, reactive functional groups in cysteine proteases, hinge binding groups in kinases etc. In these scenarios, docking will help in the prioritization of compounds based on the interactions they make with the binding pocket. Sometimes the difference in shape similarity scores for compounds is very small and it is challenging to cherry pick for biological assay. Here, docking of shape similarity hits could also help in the prioritization of compounds for purchase or chemical synthesis. However, the combination of shape similarity with molecular docking is not always advantageous especially for proteins with highly flexible binding pockets, multiple pocket conformations or homology models where accurate docking is challenging. A virtual screening scheme where USR hits were re-ranked using Autodock-Vina score produced no active hits as docking was performed in a quite different pocket conformation (Hoeger et al., 2014). In another study, shape-based virtual screening alone produced better hit rates than hierarchical combination of shape similarity and docking methods (Ballester et al., 2012). In numerous studies, shape similarity calculations along with molecular docking were complemented with other approaches such as 2D similarity search, pharmacophore modeling, electrostatic potential matching, machine learning and MM-PBSA method (Mochalkin et al., 2009; Alcaro et al., 2013; Poongavanam and Kongsted, 2013; Wiggers et al., 2013; Hamza et al., 2014a; Kumar et al., 2014b; Pala et al., 2014; Feng et al., 2015; Corso et al., 2016; Mangiatordi et al., 2017; Xia et al., 2017). The use of different virtual screening approaches in parallel has been previously suggested as different methods tend to identify different set of compounds and virtual screening hit rates could be improved by employing them in parallel manner (Sheridan and Kearsley, 2002). In parallel virtual screening, several methods are run independently and the top hits from each method is selected. Parallel combination of various ligand and structure based methods with shape similarity approaches was found to be productive especially in case of challenging targets (Swann et al., 2011; Langdon et al., 2013; Hoeger et al., 2014). A parallel virtual screening to identify inhibitors of PRL-3 employing several ligand and structure-based methods against the same screening library produced contrasting hit rates for different approaches (Hoeger et al., 2014). Many prospective applications suggest the utility of hierarchical or parallel combination of shape similarity approaches with other ligand and structure-based methods. However, no benchmark study demonstrating their utility has been published. A systematic study will help researchers to identify areas where the combination of several approaches will be better than employing shape based virtual screening methods alone.

One application of shape similarity methods is to hop from one chemical scaffold to another in order to improve the potency, selectivity, physicochemical properties and to create novel intellectual property positions (Hu et al., 2017). Shape similarity methods are capable of identifying several scaffolds which are structurally different from the query compounds and each scaffold may be pursued separately. Scaffold hopping is highly effective in rescuing the problematic leads that cannot be pursued further due to problems in selectivity, pharmacology and

pharmacokinetics. Both atomic distance-based and Gaussian-overlay shape similarity methods can effectively perform scaffold hopping as exemplified from several prospective studies. Among the first prospective application of shape similarity based methods in scaffold hopping, small molecule inhibitors of ZipA-FtsZ protein-protein interaction were identified (Rush et al., 2005). Some recent scaffold hopping applications include the identification of inhibitors of arylamine *N*-acetyltransferases (Ballester et al., 2010), type II dehydroquinase inhibitors (Ballester et al., 2012) sumoylation enzymes (Kumar et al., 2014b, 2016), anti-tubercular agents (Hamza et al., 2014b; Wavhale et al., 2017), anti-tumor agents (Ge et al., 2014), 11 $\beta$ HSD1 inhibitors (Shave et al., 2015), leucine zipper kinase inhibitors (Patel et al., 2015), kynurenine 3-monooxygenase inhibitors (Shave et al., 2018), and partial agonist of inositol trisphosphate receptor (Vasudevan et al., 2014). In addition to prospective application, rigorous benchmarking of shape similarity methods for their scaffold hopping capabilities is important. However, systematic benchmarking is challenging due to disagreement on the definition of scaffold. In one retrospective study, the scaffold hopping potential of atomic distance-based shape similarity method USRCAT has been demonstrated utilizing DUD-E dataset (Schreyer and Blundell, 2012). For the tested benchmark dataset, USRCAT was capable of identifying structurally dissimilar active hits that could not be retrieved by utilizing topological similarities. Shape similarity was also used to repurpose existing drugs for previously unknown activity (Vasudevan et al., 2012). Another application is *in silico* target fishing or the identification of protein targets of orphan chemical compounds. In one recent research, the target of anti-fungal macrocycle amidinoureas was identified following a shape similarity screening (Maccari et al., 2017). The representative structure from a series of macrocycle amidinoureas was used as a query to obtain most similar crystallographic ligand from all solved crystal structures. A prioritized list of targets based on similarity score and subsequent docking and enzymatic assay revealed *Trichoderma viride* chitinase as target of this class of compounds. Along the same line, retrospective studies showed that the combination of molecular shape and chemical structure similarity can reliably achieve biological target prediction (Abdulhameed et al., 2012; Gfeller et al., 2013). Additionally, shape similarity comparison based on spherical harmonics surface representation has been demonstrated that it can be used to predict drug promiscuity (Perez-Nueno et al., 2011). Furthermore, shape similarity comparisons could also be used to predict subtype selectivity of ligands (Kuang et al., 2016).

One important application of shape similarity methods in drug discovery is the clustering of known inhibitors of a protein target. As the performance of most shape-based methods highly depend on the selection of right query for the virtual screening (Kirchmair et al., 2009), special attention was paid toward the development of methods dealing with this problem. It has been reported that clustering of known inhibitors based on their shapes could help the identification of optimal query for virtual screening (Pérez-Nueno and Ritchie, 2011). Clustering of spherical harmonics-based consensus shapes assisted in the identification of ligands that bind to different regions in the

binding pocket of some protein targets such as CCR5 (Pérez-Nueno et al., 2008). Further, the clustering of molecular shapes also helped in the identification of promiscuous protein targets and ligands (Pérez-Nueno and Ritchie, 2011). Selection and use of high quality compound libraries is an important aspect of high throughput screening (HTS). However, testing a large number of compounds is not economically viable. *In silico*, mostly 2D similarity based, methods are commonly employed to generate a subset or focused set for HTS (Huggins et al., 2011; Dandapani et al., 2012). The limitation with 2D similarity methods is that they ignore inherent property such as the shape of a molecule. Use of shape-based clustering of large compound libraries for creating quality HTS library present several advantages. Clustering of molecular libraries based on atomic distance-based methods such as USR can achieve similar or significantly better computational efficiency as 2D fingerprint-based methods. Moreover, it will ensure maximum diversity with less number of compounds in HTS library.

Apart from employing ligand 3D shape similarity as a virtual screening method, several groups adopted it to improve the performance of other virtual screening methods. Molecular docking is one such method widely used in drug discovery. Although there has been significant progress in the development of molecular docking methods, challenges still remain both in sampling and scoring of binding poses within protein binding pockets. In the last few years, several methods were developed that utilized ligand 3D shape similarity to improve both sampling and scoring performance of molecular docking. The shape overlap with known crystallographic ligands for the target protein was utilized to guide ligand conformational sampling toward critical regions of protein binding site (Wu and Vieth, 2004). Other methods used shape similarity based alignment for the selection of reliable poses among many docking generated poses (Fukunishi and Nakamura, 2008, 2012; Anighoro and Bajorath, 2016; Kumar and Zhang, 2016a). Ligand 3D shape similarity was also a key component of many pose prediction methods where shape similarity with existing ligand bound crystal structures was utilized to predict binding poses of unknown ligands (Kelley et al., 2015; Huang et al., 2016; Kumar and Zhang, 2016b,c). Several of these methods demonstrated excellent retrospective and prospective performance. Moreover, shape similarity also facilitated the improvement in scoring and rank-ordering performance of a docking method. Several methods have reported improved virtual screening performance of a docking method when shape overlap with crystallographic ligands was employed to select the best binding pose of ligands in a screening library (Roy et al., 2015; Anighoro and Bajorath, 2016). Consideration of protein flexibility in molecular docking is a challenging problem and several methods have been developed to tackle it (B-Rao et al., 2009). Among these, receptor ensemble based methods demonstrated reasonable performance (Bottegoni et al., 2011) where the receptor ensemble is selected either from many crystallographic structures or from those generated by *in silico* methods such as molecular dynamics simulation. It has been shown previously that the selection of receptor ensemble based on binding pocket shape similarity is an effective way of considering receptor flexibility in molecular

docking (Osguthorpe et al., 2012). Further, one method suggested utilizing a single suitable receptor for each ligand in a screening library instead of docking all compounds to multiple receptor structures (Kumar and Zhang, 2018). It was also shown that single suitable receptor selection based on ligand 3D shape similarity is superior to 2D similarity based selection.

## Applications in Protein Structure Comparison

Evaluation of structural similarity between protein structures has many applications including but not limited to classification of protein structures, evolutionary relationship between protein structures, identification of templates for homology modeling, functional annotation, protein-protein interactions etc. Conventional methods for protein structure comparison are based on the alignment of protein atoms or residues. These methods require extensive rotational and translational sampling thereby limiting their utility for large scale protein structure comparisons. Several methods have been developed that utilize shape similarity to detect global or local similarity between protein structures. Classification of these methods also follows the previously described classification including Gaussian overlay based methods, surface-based methods using spherical harmonic descriptors, 3D Zernike descriptors etc. Among these, surface-based methods were developed previously to measure similarity between protein structures. Only later they were applied to the small molecule area. Several methods of protein structure comparison employed SH to represent shapes of protein structures (Tao et al., 2005; Gramada and Bourne, 2006; Konarev et al., 2016). Like SH, 3D Zernike based moments are also suitable to compare shapes of protein structures (Sael et al., 2008b; **Figure 4A**). Not only they were suitable to estimate the similarity between two proteins but also their rotation-translation invariant nature allows fast real-time search of similar proteins in structural databases such as PDB (La et al., 2009; Kihara et al., 2011; Xiong et al., 2014). A Gaussian mixture model based protein shape similarity method (Kawabata, 2008) also allows large scale comparisons of proteins with data from PDB and EMDB. This method has been implemented as Omokage search in PDB Japan (Suzuki et al., 2016; Kinjo et al., 2017). The server compares global shapes of proteins and results are obtained reasonably fast within 1 min after submission of a query. Large scale comparison of protein structures based on shape is useful in functional annotation, selection of templates for comparative modeling etc. An application of shape comparison method to protein classification has also been reported (Daras et al., 2006).

One important application of shape matching is the evaluation of similarity between protein binding pockets. This field is especially interesting as sequence and structural alignments are often not useful when comparing binding pockets of proteins with different folds. As protein binding pockets are much more conserved than protein structures (Gao and Skolnick, 2013), a reliable comparison between protein binding pockets is crucial for predicting protein functions, polypharmacology of ligands and for drug repurposing. Numerous methods based

on distinct structural representations as described previously were developed in the last decade. One such method employed spherical harmonics to represent and compare the shapes of protein binding pockets (Morris et al., 2005). This method was later extended to compare the shape of protein binding pockets with that of binding ligands (Kahraman et al., 2007). PocketMatch compares two binding pockets based on the sorted list of distances that captured chemical nature and 3D shape of the binding pocket (Yeturu and Chandra, 2008). Another method based on property-encoded shape distributions (PESD) combines the concept of shape distributions with the chemical environment of the binding pocket surface to effectively capture binding pocket similarities (Das et al., 2009). Pocket-Surfer utilizes pseudo-Zernike descriptors and 3D Zernike descriptors to represent and compare properties and 3D shapes of binding pockets (Chikhi et al., 2010). An extension of this method, Patch-Surfer searches local similarity by representing a binding pocket as amalgamation of segmented surface patches which are described by properties such as shape, electrostatic potential, concaveness and hydrophobicity (Sael and Kihara, 2012). Similarity between protein cavities was also measured by representing the pockets by pharmacophoric grid points and aligning them by optimizing their volume overlap (Desaphy et al., 2012).

Concept of pocket similarity was also extended to complementarity between binding pockets and ligands. This gave rise to a new virtual screening methodology based on shape complementarity between binding pockets and ligands. PL-Patch-Surfer2 program evaluates the compatibility between ligand and binding pocket by measuring the complementarity between ligand surface and local surface patches in the binding pocket (Shin et al., 2016a,b; **Figure 4C**). The program utilizes 3DZD to represent molecular shape while physicochemical properties are also mapped onto the surface. The method was evaluated on benchmark datasets and revealed better performance than two docking programs. Spherical harmonics expansion coefficients have also been employed in the approximation and comparison of binding pockets and ligand surfaces (Cai et al., 2002). The complementarity was demonstrated utilizing 35 protein-ligand complexes. Elekit adopted shape and electrostatic complementarity concept to discover small molecule inhibitors of protein-protein interactions (Voet et al., 2013). Elekit assesses the similarity between small molecules and protein ligands of a receptor protein based on the electrostatic potential values stored on a 3D grid.

## Applications in Fitting of Atomic Models Into Cryo-Electron Microscopy Maps

Recent developments in cryo-electron microscopy (cryo-EM) has helped researchers to overcome resolution barrier and provide structural and mechanistic insights into structures of difficult proteins and large protein assemblies. Most of these improvements came from the advances in sample preparation, electron detector technologies, improved microscope and computational data processing. Computational methods played an important part in particle picking, particle reconstruction,

building and fitting of structures into cryo-EM maps. In recent years, several methods were developed to improve building, fitting and refinement of protein structures in cryo-EM maps (Esquivel-Rodríguez and Kihara, 2013). Among these methods, a few methods employed shape similarity to fit atomic structures of protein subunits into the cryo-EM maps of multi-subunit proteins. One method, Gaussian Mixture macromolecule FITting (gmfit), utilizes Gaussian mixture models (GMM) to represent the shape of cryo-EM maps and atomic models (Kawabata, 2008). GMMs are probability distribution functions obtained by joining many 3D Gaussian functions. Initially, both the cryo-EM map and atomic models are first converted into GMM followed by the fitting of a single subunit GMM into the GMM of protein complex using random and gradient based local search. Finally, the fit between atomic models and cryo-EM map is obtained based on the position and orientation of GMM. This method is reasonably fast and can fit multiple subunits with reasonable accuracy. PDB Japan (<https://pdbj.org>) has implemented this method in its EM navigator utility to provide shape based structural similarity search against protein databases (Kinjo et al., 2017). Another method adopted a surface-based approach where 3DZD was used to represent and compare isosurface derived from low resolution cryo-EM maps of protein structures (Sael and Kihara, 2010; **Figure 4D**). It was demonstrated that 3DZD can distinguish proteins of different folds even at low resolution of 15 Å. A web-based platform for comparing cryo-EM maps was also developed by the same group (Esquivel-Rodríguez et al., 2015; Han et al., 2017). A similar method utilized 3D Zernike moments to search a database of protein structures for matching protein structures to a cryo-EM map (Yin and Dokholyan, 2011). EMLZerD method also utilized 3DZD to fit multiple structures in a cryo-EM map (Esquivel-Rodríguez and Kihara, 2012). The method generates hundreds of putative configurations of subunit arrangement using a protein-protein docking method. These configurations were later compared with a cryo-EM map using 3DZD and Euclidean distance. The biggest advantage of 3D Zernike moments methods is that they are rotation translation invariant and no computational expensive step of rigid body or flexible structural alignment is required. Moreover, these methods enable screening of proteins from structural databases such as PDB to find out models that can fit into a cryo-EM map.

## CONCLUSION AND FUTURE DIRECTIONS

3D shape similarity methods have contributed immensely to the overall acceptance of the computational virtual screening methods in drug discovery. Most shape similarity methods for shape comparison of small molecules and macromolecules took inspiration from the approaches developed to compare the shapes of 3D objects in computational geometry field. Several approaches were developed ranging from extremely fast atom distance-based methods to comparatively slower mathematically complex methods such as SH and 3DZD. Among all the 3D shape comparison methods, atomic distance-based and Gaussian overlay-based methods are the most widely used. These approaches possess several advantages over

surface-based methods. Atomic distance-based methods present an extremely fast way of quickly comparing the shapes of small molecules. This has facilitated the screening of very large libraries of millions of compounds within a few seconds. Moreover, screening large libraries increased the probability of finding novel chemical scaffolds. Furthermore, as most of these methods depend on shape rather than the underlying chemical structure, scaffold hopping can be conveniently achieved. Another possible application of these fast shape similarity evaluation methods would be the clustering of large chemical space to generate quality shape diverse HTS screening libraries. Although Gaussian overlay-based methods are slower than atomic-distance based methods, they are fast enough to allow high throughput virtual screening. GPU implementations of these methods is not very difficult as exemplified by the development of several GPU compatible programs such as FastROCS, PAPER, gWEGA etc. resulting in further increase in the processing speeds. Another advantage with Gaussian-based methods is that they allow visualization as they require alignment of molecules prior to shape similarity calculations. Visualization is helpful in understanding the features responsible for biological activity and critical for the optimization of a molecule especially for the molecules with low structural similarity with query compound. However, a suboptimal alignment can lead to errors in volume overlap calculations and thereby affecting similarity scores and visualization. As alignment is the key component of Gaussian overlay methods, efforts should be focused toward improving molecular alignment. Some of these methods employ chemical features to refine global overlays. As alignment is global optimization problem, molecular alignment could also be improved by employing fast local optimization methods. Both atomic distance-based and Gaussian overlay-based shape similarity methods handle ligand flexibility by employing the conformational ensemble. The performance thus indirectly depends upon conformation

generation methods. Current state-of-the-art conformation generation methods still struggle to generate near-native conformations of ligands such as peptidomimetics, macrocycles etc. Development of novel conformation generation approaches utilizing knowledge from experimental databases such as CSD and PDB will steer improvement in performance of shape-based virtual screening approaches. Surface based methods such as SH expansion coefficients and 3DZD are suitable for comparing macromolecules and atomic models with electron density maps, however, comparatively less efforts have been made toward utilizing them in small molecule area. One advantage with surface-based methods is that the protein ligand complementarity search is possible by comparing enclosed shapes of binding pockets and ligands. This will be handy in cases where ligand-based virtual screening methods could not be used due to the lack of active compounds. Finally, shape-based similarity could be used in combination with other ligand and structure-based approaches either in hierarchical or parallel manner to improve hit rate especially for difficult targets.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

We acknowledge RIKEN ACCC for the supercomputing resources at the Hokusai GreatWave supercomputer. The research in our laboratory was partially supported by Platform Project for Supporting Drug Discovery and Life Science Research [Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)] from AMED under Grant Number JP18am0101082. We thank members of our lab for help and discussions.

## REFERENCES

- Abdulhameed, M. D. M., Chaudhury, S., Singh, N., Sun, H., Wallqvist, A., and Tawa, G. J. (2012). Exploring polypharmacology using a ROCS-based target fishing approach. *J. Chem. Inf. Model* 52, 492–505. doi: 10.1021/ci2003544
- Acharya, P., Dogo-Isonagie, C., Lalonde, J. M., Lam, S. N., Leslie, G. J., Louder, M. K., et al. (2011). Structure-based identification and neutralization mechanism of tyrosine sulfate mimetics that inhibit HIV-1 entry. *ACS Chem. Biol.* 6, 1069–1077. doi: 10.1021/cb200068b
- Ai, N., Welsh, W. J., Santhanam, U., Hu, H., and Lyga, J. (2014). Novel virtual screening approach for the discovery of human tyrosinase inhibitors. *PLoS ONE* 9:e112788. doi: 10.1371/journal.pone.0112788
- Alcaro, S., Musetti, C., Distinto, S., Casatti, M., Zagotto, G., Artese, A., et al. (2013). Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches. *J. Med. Chem.* 56, 843–855. doi: 10.1021/jm3013486
- Ambaye, N. D., Gunzburg, M. J., Lim, R. C. C., Price, J. T., Wilce, M. C. J., and Wilce, J. A. (2013). The discovery of phenylbenzamide derivatives as Grb7-based antitumor agents. *ChemMedChem* 8, 280–288. doi: 10.1002/cmdc.201200400
- Anighoro, A., and Bajorath, J. (2016). Three-dimensional similarity in molecular docking: prioritizing ligand poses on the basis of experimental binding modes. *J. Chem. Inf. Model.* 56, 580–587. doi: 10.1021/acs.jcim.5b00745
- Armstrong, M. S., Finn, P. W., Morris, G. M., and Richards, W. G. (2011). Improving the accuracy of ultrafast ligand-based screening: incorporating lipophilicity into ElectroShape as an extra dimension. *J. Comput. Aided Mol. Des.* 25, 785–790. doi: 10.1007/s10822-011-9463-8
- Armstrong, M. S., Morris, G. M., Finn, P. W., Sharma, R., Moretti, L., Cooper, R. I. D. (2010). ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. *J. Comput. Aided Mol. Des.* 24, 789–801. doi: 10.1007/s10822-010-9374-0
- Armstrong, M. S., Morris, G. M., Finn, P. W., Sharma, R., and Richards, W. G. (2009). Molecular similarity including chirality. *J. Mol. Graph. Model.* 28, 368–370. doi: 10.1016/j.jmkgm.2009.09.002
- Ballester, P. J. (2011). Ultrafast shape recognition: method and applications. *Future Med. Chem.* 3, 65–78. doi: 10.4155/fmc.10.280
- Ballester, P. J., Finn, P. W., and Richards, W. G. (2009). Ultrafast shape recognition: evaluating a new ligand-based virtual screening technology. *J. Mol. Graph. Model.* 27, 836–845. doi: 10.1016/j.jmkgm.2009.01.001
- Ballester, P. J., Mangold, M., Howard, N. I., Robinson, R. L., Abell, C., Blumberger, J., et al. (2012). Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification. *J. R. Soc. Interface* 9, 3196–3207. doi: 10.1098/rsif.2012.0569
- Ballester, P. J., and Richards, W. G. (2007a). Ultrafast shape recognition for similarity search in molecular databases. *Proc. R. Soc. Math. Phys. Eng. Sci.* 463, 1307–1321. doi: 10.1098/rspa.2007.1823



- Ballester, P. J., and Richards, W. G. (2007b). Ultrafast shape recognition to search compound databases for similar molecular shapes. *J. Comput. Chem.* 28, 1711–1723. doi: 10.1002/jcc.20681
- Ballester, P. J., Westwood, I., Laurieri, N., Sim, E., and Richards, W. G. (2010). Prospective virtual screening with Ultrafast Shape Recognition: the identification of novel inhibitors of arylamine N-acetyltransferases. *J. R. Soc. Interface* 7, 335–342. doi: 10.1098/rsif.2009.0170
- Bassetto, M., Leysen, P., Neyts, J., Yerukhimovich, M. M., Frick, D. N., and Brancale, A. (2017). Shape-based virtual screening, synthesis and evaluation of novel pyrrolone derivatives as antiviral agents against HCV. *Bioorg. Med. Chem. Lett.* 27, 936–940. doi: 10.1016/j.bmcl.2016.12.087
- Bemis, G. W., and Kuntz, I. D. (1992). A fast and efficient method for 2D and 3D molecular shape description. *J. Comput. Aided Mol. Des.* 6, 607–628. doi: 10.1007/BF00126218
- Berenger, F., Voet, A., Lee, X. Y., and Zhang, K. Y. (2014). A rotation-translation invariant molecular descriptor of partial charges and its use in ligand-based virtual screening. *J. Cheminform.* 6:23. doi: 10.1186/1758-2946-6-23
- Boström, J., Berggren, K., Elebring, T., Greasley, P. J., and Wilstermann, M. (2007). Scaffold hopping, synthesis and structure-activity relationships of 5,6-diarylpyrazine-2-amide derivatives: a novel series of CB1 receptor antagonists. *Bioorg. Med. Chem.* 15, 4077–4084. doi: 10.1016/j.bmc.2007.03.075
- Boström, J., Grant, J. A., Fjellstrom, O., Thelin, A., and Gustafsson, D. (2013). Potent fibrinolysis inhibitor discovered by shape and electrostatic complementarity to the drug tranexamic acid. *J. Med. Chem.* 56, 3273–3280. doi: 10.1021/jm301818g
- Bottegoni, G., Rocchia, W., Rueda, M., Abagyan, R., and Cavalli, A. (2011). Systematic exploitation of multiple receptor conformations for virtual ligand screening. *PLoS ONE* 6:e18845. doi: 10.1371/journal.pone.0018845
- B-Rao, C., Subramanian, J., and Sharma, S. D. (2009). Managing protein flexibility in docking and its applications. *Drug Discov. Today* 14, 394–400. doi: 10.1016/j.drudis.2009.01.003
- Brunskole Švegelj, M., Turk, S., Brus, B., Lanišnik Rižner, T., Stojan, J., and Gobec, S. (2011). Novel inhibitors of trihydroxynaphthalene reductase with antifungal activity identified by ligand-based and structure-based virtual screening. *J. Chem. Inf. Model.* 51, 1716–1724. doi: 10.1021/ci2001499
- Cai, C., Gong, J., Liu, X., Gao, D., and Li, H. (2013). SimG: an alignment based method for evaluating the similarity of small molecules and binding sites. *J. Chem. Inf. Model.* 53, 2103–2115. doi: 10.1021/ci400139j
- Cai, C., Gong, J., Liu, X., Jiang, H., Gao, D., and Li, H. (2012). A novel, customizable and optimizable parameter method using spherical harmonics for molecular shape similarity comparisons. *J. Mol. Model.* 18, 1597–1610. doi: 10.1007/s00894-011-1173-6
- Cai, W., Shao, X., and Maigret, B. (2002). Protein-ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening. *J. Mol. Graph. Model.* 20, 313–328. doi: 10.1016/S1093-3263(01)00134-6
- Cannon, E. O., Nigsch, F., and Mitchell, J. B. (2008). A novel hybrid ultrafast shape descriptor method for use in virtual screening. *Chem. Cent. J.* 2:3. doi: 10.1186/1752-153X-2-3
- Chen, W. L., Wang, Z. H., Feng, T. T., Li, D. D., Wang, C. H., Xu, X. L., et al. (2016). Discovery, design and synthesis of 6H-anthra[1,9-cd]isoxazol-6-one scaffold as G9a inhibitor through a combination of shape-based virtual screening and structure-based molecular modification. *Bioorg. Med. Chem.* 24, 6102–6108. doi: 10.1016/j.bmc.2016.09.071
- Chikhi, R., Sael, L., and Kihara, D. (2010). Real-time ligand binding pocket database search using local surface descriptors. *Proteins* 78, 2007–2028. doi: 10.1002/prot.22715
- Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221, 709–713. doi: 10.1126/science.6879170
- Connolly, M. L. (1985). Computation of molecular volume. *J. Am. Chem. Soc.* 107, 1118–1124. doi: 10.1021/ja00291a006
- Corso, G., Alisi, M. A., Cazzolla, N., Coletta, I., Furlotti, G., Garofalo, B., et al. (2016). A novel multi-step virtual screening for the identification of human and mouse mPGES-1 inhibitors. *Mol. Inform.* 35, 358–368. doi: 10.1002/minf.201600024
- Cosgrove, D. A., Bayada, D. M., and Johnson, A. P. (2000). A novel method of aligning molecules by local surface shape similarity. *J. Comput. Aided Mol. Des.* 14, 573–591. doi: 10.1023/A:1008167930625
- Cramer, R. D., Patterson, D. E., and Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110, 5959–5967. doi: 10.1021/ja00226a005
- Dandapani, S., Rosse, G., Southall, N., Salvino, J. M., and Thomas, C. J. (2012). Selecting, acquiring, and using small molecule libraries for high-throughput screening. *Curr. Protoc. Chem. Biol.* 4, 177–191. doi: 10.1002/9780470559277.ch110252
- Daras, P., Zarpalas, D., Axenopoulos, A., Tzouvaras, D., and Strintzis, M. G. (2006). Three-dimensional shape-structure comparison method for protein classification. *IEEE/ACM Trans. Comput. Biol. Bioinformatics.* 3, 193–207. doi: 10.1109/tcbb.2006.43
- Das, D., Maeda, K., Hayashi, Y., Gavande, N., Desai, D. V., Chang, S. B., et al. (2015). Insights into the mechanism of inhibition of CXCR4: identification of piperidinylethanamine analogs as anti-HIV-1 inhibitors. *Antimicrob. Agents Chemother.* 59, 1895–1904. doi: 10.1128/AAC.04654-14
- Das, S., Kokardekar, A., and Breneman, C. M. (2009). Rapid comparison of protein binding site surfaces with property encoded shape distributions. *J. Chem. Inf. Model.* 49, 2863–2872. doi: 10.1021/ci900317x
- Desaphy, J., Azdimousa, K., Kellenberger, E., and Rognan, D. (2012). Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *J. Chem. Inf. Model.* 52, 2287–2299. doi: 10.1021/ci300184x
- Edelsbrunner, H., and Mücke, E. P. (1994). Three-dimensional alpha shapes. *ACM Trans. Graph.* 13, 43–72. doi: 10.1145/174462.156635
- Edelsbrunner, H. (1995). The union of balls and its dual shape. *Discrete Comput. Geom.* 13, 415–440. doi: 10.1007/BF02574053
- Edelsbrunner, H., Kirkpatrick, D., and Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* 29, 551–559. doi: 10.1109/TIT.1983.1056714
- Esquivel-Rodríguez, J., and Kihara, D. (2012). Fitting multimeric protein complexes into electron microscopy maps using 3D zernike descriptors. *J. Phys. Chem. B* 116, 6854–6861. doi: 10.1021/jp212612t
- Esquivel-Rodríguez, J., and Kihara, D. (2013). Computational methods for constructing protein structure models from 3D electron microscopy maps. *J. Struct. Biol.* 184, 93–102. doi: 10.1016/j.jsb.2013.06.008
- Esquivel-Rodríguez, J., Xiong, Y., Han, X., Guang, S., Christoffer, C., and Kihara, D. (2015). Navigating 3D electron microscopy maps with EM-SURFER. *BMC Bioinformatics* 16:181. doi: 10.1186/s12859-015-0580-6
- Feng, T., Chen, W., Li, D., Lin, H., Liu, F., Bao, Q., et al. (2015). Identification of novel JMJD2A inhibitor scaffold using shape and electrostatic similarity search combined with docking method and MM-GBSA approach. *RSC Adv.* 5, 82936–82946. doi: 10.1039/C5RA11896D
- Freitas, R. F., Oprea, T. I., and Montanari, C. A. (2008). 2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L. *Bioorg. Med. Chem.* 16, 838–853. doi: 10.1016/j.bmc.2007.10.048
- Fu, J., Si, P., Zheng, M., Chen, L., Shen, X., Tang, Y., et al. (2012). Discovery of new non-steroidal FXR ligands via a virtual screening workflow based on phase shape and induced fit docking. *Bioorg. Med. Chem. Lett.* 22, 6848–6853. doi: 10.1016/j.bmcl.2012.09.045
- Fukunishi, Y., and Nakamura, H. (2008). Prediction of protein-ligand complex structure by docking software guided by other complex structures. *J. Mol. Graph. Model.* 26, 1030–1033. doi: 10.1016/j.jmkgm.2007.07.001
- Fukunishi, Y., and Nakamura, H. (2012). Integration of ligand-based drug screening with structure-based drug screening by combining maximum volume overlapping score with ligand docking. *Pharmaceuticals* 5, 1332–1345. doi: 10.3390/ph5121332
- Gao, M., and Skolnick, J. (2013). A comprehensive survey of small-molecule binding pockets in proteins. *PLoS Comput. Biol.* 9:e1003302. doi: 10.1371/journal.pcbi.1003302
- Ge, H., Wang, Y., Zhao, W., Lin, W., Yan, X., and Xu, J. (2014). Scaffold hopping of potential anti-tumor agents by WEGA: a shape-based approach. *MedChemComm* 5, 737–741. doi: 10.1039/C3MD00397C
- Gfeller, D., Michielin, O., and Zoete, V. (2013). Shaping the interaction landscape of bioactive molecules. *Bioinformatics* 29, 3073–3079. doi: 10.1093/bioinformatics/btt540
- Giganti, D., Guillemain, H., Spadoni, J.-L., Nilges, M., Zagury, J.-F., and Montes, M. (2010). Comparative evaluation of 3D virtual ligand screening methods:

- impact of the molecular alignment on enrichment. *J. Chem. Inf. Model.* 50, 992–1004. doi: 10.1021/ci900507g
- Goldman, B. B., and Wipke, W. T. (2000). Quadratic shape descriptors. 1. rapid superposition of dissimilar molecules using geometrically invariant surface descriptors. *J. Chem. Inf. Comput. Sci.* 40, 644–658. doi: 10.1021/ci980213w
- Good, A. C., Ewing, T. J., Gschwend, D. A., and Kuntz, I. D. (1995). New molecular shape descriptors: application in database screening. *J. Comput. Aided Mol. Des.* 9, 1–12. doi: 10.1007/BF00117274
- Good, A. C., Hodgkin, E. E., and Richards, W. G. (1992). Utilization of gaussian functions for the rapid evaluation of molecular similarity. *J. Chem. Inf. Comput. Sci.* 32, 188–191. doi: 10.1021/ci00007a002
- Gramada, A., and Bourne, P. E. (2006). Multipolar representation of protein structure. *BMC Bioinformatics* 7:242. doi: 10.1186/1471-2105-7-242
- Gramatica, P. (2006). WHIM descriptors of shape. *QSAR Comb. Sci.* 25, 327–332. doi: 10.1002/qsar.200510159
- Grant, J. A., Gallardo, M. A., and Pickup, B. T. (1996). A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* 17, 1653–1666. doi: 10.1002/(SICI)1096-987X(19961115)17:14<1653::AID-JCC7>3.0.CO;2-K
- Grant, J. A., and Pickup, B. T. (1995). A gaussian description of molecular shape. *J. Phys. Chem.* 99, 3503–3510. doi: 10.1021/j100011a016
- Grant, J. A., and Pickup, B. T. (1997). “Gaussian shape methods,” in *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, eds W. F. Van Gunsteren, P. K. Weiner, and A. J. Wilkinson (Dordrecht: Springer), 150–176.
- Hain, E., Camacho, C. J., and Koes, D. R. (2016). Fragment oriented molecular shapes. *J. Mol. Graph. Model.* 66, 143–154. doi: 10.1016/j.jmkgm.2016.03.017
- Hamza, A., Wagner, J. M., Evans, T. J., Frasinuk, M. S., Kwiatkowski, S., Zhan, C.-G., et al. (2014a). Novel mycosin protease mycp1 inhibitors identified by virtual screening and 4D fingerprints. *J. Chem. Inf. Model.* 54, 1166–1173. doi: 10.1021/ci500025r
- Hamza, A., Wagner, J. M., Wei, N.-N., Kwiatkowski, S., Zhan, C.-G., Watt, D. S., et al. (2014b). Application of the 4D fingerprint method with a robust scoring function for scaffold-hopping and drug repurposing strategies. *J. Chem. Inf. Model.* 54, 2834–2845. doi: 10.1021/ci5003872
- Hamza, A., Wei, N.-N., Hao, C., Xiu, Z., and Zhan, C.-G. (2013). A novel and efficient ligand-based virtual screening approach using the HWZ scoring function and an enhanced shape-density model. *J. Biomol. Struct. Dyn.* 31, 1236–1250. doi: 10.1080/07391102.2012.732341
- Hamza, A., Wei, N. N., and Zhan, C. G. (2012). Ligand-based virtual screening approach using a new scoring function. *J. Chem. Inf. Model.* 52, 963–974. doi: 10.1021/ci200617d
- Han, X., Wei, Q., and Kihara, D. (2017). Protein 3D structure and electron microscopy map retrieval using 3D-SURFER2.0 and EM-SURFER. *Curr. Protoc. Bioinformatics* 60, 3.14.11–13.14.15. doi: 10.1002/cpbi.37
- Haque, I. S., and Pande, V. S. (2010). PAPER—Accelerating parallel evaluations of ROCS. *J. Comput. Chem.* 31, 117–132. doi: 10.1002/jcc.21307
- Hartman, I., Gillies, A. R., Arora, S., Andaya, C., Royapet, N., Welsh, W. J., et al. (2009). Application of screening methods, shape signatures and engineered biosensors in early drug discovery process. *Pharm. Res.* 26, 2247–2258. doi: 10.1007/s11095-009-9941-z
- Hawkins, P. C. D., Skillman, A. G., and Nicholls, A. (2007). Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* 50, 74–82. doi: 10.1021/jm0603365
- Hevener, K. E., Mehboob, S., Su, P.-C., Truong, K., Boci, T., Deng, J., et al. (2011). Discovery of a novel and potent class of *F. tularensis* enoyl-reductase (FabI) inhibitors by molecular shape and electrostatic matching. *J. Med. Chem.* 55, 268–279. doi: 10.1021/jm201168g
- Hodgkin, E. E., and Richards, W. G. (1987). Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem.* 32, 105–110. doi: 10.1002/qua.560320814
- Hoeger, B., Diether, M., Ballester, P. J., and Kohn, M. (2014). Biochemical evaluation of virtual screening methods reveals a cell-active inhibitor of the cancer-promoting phosphatases of regenerating liver. *Eur. J. Med. Chem.* 88, 89–100. doi: 10.1016/j.ejmech.2014.08.060
- Hofbauer, C., Lohninger, H., and Aszodi, A. (2004). SURFCOMP: a novel graph-based approach to molecular surface comparison. *J. Chem. Inf. Comput. Sci.* 44, 837–847. doi: 10.1021/ci0342371
- Houston, D. R., Yen, L.-H., Pettit, S., and Walkinshaw, M. D. (2015). Structure- and ligand-based virtual screening identifies new scaffolds for inhibitors of the oncoprotein MDM2. *PLoS ONE* 10:e0121424. doi: 10.1371/journal.pone.0121424
- Hu, B., Kuang, Z. K., Feng, S. Y., Wang, D., He, S. B., and Kong, D. X. (2016). Three-dimensional biologically relevant spectrum (brs-3d): shape similarity profile based on pdb ligands as molecular descriptors. *Molecules* 21:1554. doi: 10.3390/molecules21111554
- Hu, G., Kuang, G., Xiao, W., Li, W., Liu, G., and Tang, Y. (2012). Performance evaluation of 2D fingerprint and 3D shape similarity methods in virtual screening. *J. Chem. Inf. Model.* 52, 1103–1113. doi: 10.1021/ci300030u
- Hu, Y., Stumpfe, D., and Bajorath, J. (2017). Recent advances in scaffold hopping. *J. Med. Chem.* 60, 1238–1246. doi: 10.1021/acs.jmedchem.6b01437
- Huang, S. Y., Li, M., Wang, J., and Pan, Y. (2016). HybridDock: a hybrid protein–ligand docking protocol integrating protein- and ligand-based approaches. *J. Chem. Inf. Model.* 56, 1078–1087. doi: 10.1021/acs.jcim.5b00275
- Huggins, D. J., Venkiteraman, A. R., and Spring, D. R. (2011). Rational methods for the selection of diverse screening compounds. *ACS Chem. Biol.* 6, 208–217. doi: 10.1021/cb100420r
- Kahraman, A., Morris, R. J., Laskowski, R. A., and Thornton, J. M. (2007). Shape variation in protein binding pockets and their ligands. *J. Mol. Biol.* 368, 283–301. doi: 10.1016/j.jmb.2007.01.086
- Kalászi, A., Szisz, D., Imre, G., and Polgár, T. (2014). Screen3D: a novel fully flexible high-throughput shape-similarity search method. *J. Chem. Inf. Model.* 54, 1036–1049. doi: 10.1021/ci400620f
- Kaoud, T. S., Yan, C., Mitra, S., Tseng, C.-C., Jose, J., Taliaferro, J. M., et al. (2012). From *in silico* discovery to intracellular activity: targeting JNK–protein interactions with small molecules. *ACS Med. Chem. Lett.* 3, 721–725. doi: 10.1021/ml300129b
- Karaboga, A. S., Petronin, F., Marchetti, G., Souchet, M., and Maigret, B. (2013). Benchmarking of HPCC: a novel 3D molecular representation combining shape and pharmacophoric descriptors for efficient molecular similarity assessments. *J. Mol. Graph. Model.* 41, 20–30. doi: 10.1016/j.jmkgm.2013.01.003
- Kawabata, T. (2008). Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophys. J.* 95, 4643–4658. doi: 10.1529/biophysj.108.137125
- Kazhdan, M., Funkhouser, T., and Rusinkiewicz, S. (2003). “Rotation invariant spherical harmonic representation of 3D shape descriptors,” in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (Aachen).
- Kelley, B. P., Brown, S. P., Warren, G. L., and Muchmore, S. W. (2015). POSIT: flexible shape-guided docking for pose prediction. *J. Chem. Inf. Model.* 55, 1771–1780. doi: 10.1021/acs.jcim.5b00142
- Kihara, D., Sael, L., and Chikhi, R. (2009). Local surface shape-based protein function prediction using Zernike descriptors. *Biophys. J.* 96:650a.
- Kihara, D., Sael, L., Chikhi, R., Esquivel-Rodriguez, J. (2011). Molecular surface representation using 3D zernike descriptors for protein shape comparison and docking. *Curr. Protein Pept. Sci.* 12, 520–530. doi: 10.2174/138920311796957612
- Kinjo, A. R., Bekker, G. J., Suzuki, H., Tsuchiya, Y., Kawabata, T., Ikegawa, Y., et al. (2017). Protein data bank japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures. *Nucleic Acids Res.* 45, D282–D288. doi: 10.1093/nar/gkw962
- Kirchmair, J., Distinto, S., Markt, P., Schuster, D., Spitzer, G. M., Liedl, K. R., et al. (2009). How to optimize shape-based virtual screening: choosing the right query and including chemical information. *J. Chem. Inf. Model.* 49, 678–692. doi: 10.1021/ci8004226
- Koes, D. R., and Camacho, C. J. (2014). Shape-based virtual screening with volumetric aligned molecular shapes. *J. Comput. Chem.* 35, 1824–1834. doi: 10.1002/jcc.23690
- Konarev, P. V., Petoukhov, M. V., and Svergun, D. I. (2016). Rapid automated superposition of shapes and macromolecular models using spherical harmonics. *J. Appl. Crystallogr.* 49, 953–960. doi: 10.1107/S1600576716005793

- Kong, Y., Bender, A., and Yan, A. (2018). Identification of novel aurora kinase a (AURKA) inhibitors via hierarchical ligand-based virtual screening. *J. Chem. Inf. Model.* 58, 36–47. doi: 10.1021/acs.jcim.7b00300
- Kortagere, S., Krasowski, M. D., and Ekins, S. (2009). The importance of discerning shape in molecular pharmacology. *Trends Pharmacol. Sci.* 30, 138–147. doi: 10.1016/j.tips.2008.12.001
- Kuang, Z. K., Feng, S. Y., Hu, B., Wang, D., He, S. B., and Kong, D. X. (2016). Predicting subtype selectivity of dopamine receptor ligands with three-dimensional biologically relevant spectrum. *Chem. Biol. Drug Des.* 88, 859–872. doi: 10.1111/cbdd.12815
- Kumar, A., Ito, A., Hirohama, M., Yoshida, M., and Zhang, K. Y. J. (2014a). Identification of sumoylation inhibitors targeting a predicted pocket in Ubc9. *J. Chem. Inf. Model.* 54, 2784–2793. doi: 10.1021/ci5004015
- Kumar, A., Ito, A., Hirohama, M., Yoshida, M., and Zhang, K. Y. J. (2016). Identification of new SUMO activating enzyme 1 inhibitors using virtual screening and scaffold hopping. *Bioorg. Med. Chem. Lett.* 26, 1218–1223. doi: 10.1016/j.bmcl.2016.01.030
- Kumar, A., Ito, A., Takemoto, M., Yoshida, M., and Zhang, K. Y. J. (2014b). Identification of 1,2,5-oxadiazoles as a new class of SENP2 inhibitors using structure based virtual screening. *J. Chem. Inf. Model.* 54, 870–880. doi: 10.1021/ci4007134
- Kumar, A., Parkesh, R., Sznajder, L. J., Childs-Disney, J. L., Sobczak, K., and Disney, M. D. (2012). Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(cag)-containing transcripts. *ACS Chem. Biol.* 7, 496–505. doi: 10.1021/cb200413a
- Kumar, A., and Zhang, K. Y. (2015). Hierarchical virtual screening approaches in small molecule drug discovery. *Methods* 71, 26–37. doi: 10.1016/j.ymeth.2014.07.007
- Kumar, A., and Zhang, K. Y. (2016a). Application of shape similarity in pose selection and virtual screening in CSARDock2014 exercise. *J. Chem. Inf. Model.* 56, 965–973. doi: 10.1021/acs.jcim.5b00279
- Kumar, A., and Zhang, K. Y. (2016b). A pose prediction approach based on ligand 3D shape similarity. *J. Comput.-Aided Mol. Des.* 30, 457–469. doi: 10.1007/s10822-016-9923-2
- Kumar, A., and Zhang, K. Y. (2016c). Prospective evaluation of shape similarity based pose prediction method in D3R grand challenge 2015. *J. Comput. Aided Mol. Des.* 30, 685–693. doi: 10.1007/s10822-016-9931-2
- Kumar, A., and Zhang, K. Y. J. (2018). A cross docking pipeline for improving pose prediction and virtual screening performance. *J. Comput. Aided Mol. Des.* 32, 163–173. doi: 10.1007/s10822-017-0048-z
- La, D., Esquivel-Rodríguez, J., Venkatraman, V., Li, B., Sael, L., Ueng, S., et al. (2009). 3D-SURFER: software for high-throughput protein surface comparison and analysis. *Bioinformatics* 25, 2843–2844. doi: 10.1093/bioinformatics/btp542
- Langdon, S. R., Westwood, I. M., van Montfort, R. L. M., Brown, N., and Blagg, J. (2013). Scaffold-focused virtual screening: prospective application to the discovery of TTK inhibitors. *J. Chem. Inf. Model.* 53, 1100–1112. doi: 10.1021/ci400100c
- Lee, B., and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379–400. doi: 10.1016/0022-2836(71)90324-X
- Li, H., Huang, J., Chen, L., Liu, X., Chen, T., Zhu, J., et al. (2009). Identification of novel falcipain-2 inhibitors as potential antimalarial agents through structure-based virtual screening. *J. Med. Chem.* 52, 4936–4940. doi: 10.1021/jm801622x
- Li, H., Leung, K. S., Wong, M. H., and Ballester, P. J. (2016). USR-VS: a web server for large-scale prospective virtual screening using ultrafast shape recognition techniques. *Nucleic Acids Res.* 44, W436–W441. doi: 10.1093/nar/gkw320
- Lim, S. V., Rahman, M. B., and Tejo, B. A. (2011). Structure-based and ligand-based virtual screening of novel methyltransferase inhibitors of the dengue virus. *BMC Bioinformatics* 12:S24. doi: 10.1186/1471-2105-12-S13-S24
- Lin, J. H., and Clark, T. (2005). An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties. *J. Chem. Inf. Model.* 45, 1010–1016. doi: 10.1021/ci050059v
- Liu, X., Jiang, H., and Li, H. (2011). SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. method and assessment of virtual screening. *J. Chem. Inf. Model.* 51, 2372–2385. doi: 10.1021/ci200060s
- Liu, Y. S., Fang, Y., and Ramani, K. (2009). IDSS: deformation invariant signatures for molecular shape comparison. *BMC Bioinformatics* 10:157. doi: 10.1186/1471-2105-10-157
- Liu, Y. S., Wang, M., Paul, J. C., and Ramani, K. (2012). 3DMolNavi: a web-based retrieval and navigation tool for flexible molecular shape comparison. *BMC Bioinformatics* 13:95. doi: 10.1186/1471-2105-13-95
- Maccari, G., Deodato, D., Fiorucci, D., Orofino, F., Truglio, G. I., Pasero, C., et al. (2017). Design and synthesis of a novel inhibitor of T. viride chitinase through an *in silico* target fishing protocol. *Bioorg. Med. Chem. Lett.* 27, 3332–3336. doi: 10.1016/j.bmcl.2017.06.016
- Mak, L., Grandison, S., and Morris, R. J. (2008). An extension of spherical harmonics to region-based rotationally invariant descriptors for molecular shape description and comparison. *J. Mol. Graph. Model.* 26, 1035–1045. doi: 10.1016/j.jmgm.2007.08.009
- Mangiatordi, G. F., Trisciuzzi, D., Alberga, D., Denora, N., Iacobazzi, R. M., Gadaleta, D., et al. (2017). Novel chemotypes targeting tubulin at the colchicine binding site and unbiasing P-glycoprotein. *Eur. J. Med. Chem.* 139, 792–803. doi: 10.1016/j.ejmech.2017.07.037
- Mansfield, M. L., Covell, D. G., and Jernigan, R. L. (2002). A new class of molecular shape descriptors. 1. theory and properties. *J. Chem. Inf. Comput. Sci.* 42, 259–273. doi: 10.1021/ci000100o
- Masek, B. B., Merchant, A., and Matthew, J. B. (1993). Molecular shape comparison of angiotensin II receptor antagonists. *J. Med. Chem.* 36, 1230–1238. doi: 10.1021/jm00061a014
- Mavridis, L., Hudson, B. D., and Ritchie, D. W. (2007). Toward high throughput 3D virtual screening using spherical harmonic surface representations. *J. Chem. Inf. Model.* 47, 1787–1796. doi: 10.1021/ci7001507
- Max, N. L., and Getzoff, E. D. (1988). Spherical harmonic molecular surfaces. *IEEE Comp. Graph. Appl.* 8, 42–50. doi: 10.1109/38.7748
- Meek, P. J., Liu, Z., Tian, L., Wang, C. Y., Welsh, W. J., and Zauhar, R. J. (2006). Shape signatures: speeding up computer aided drug discovery. *Drug Discov. Today* 11, 895–904. doi: 10.1016/j.drudis.2006.08.014
- Mezey, P. G. (2007). “Molecular Surfaces,” in *Reviews in Computational Chemistry*, eds K. B. Lipkowitz and D. B. Boyd (New York, NY: VCH Publishers), 265–294.
- Mochalkin, I., Miller, J. R., Narasimhan, L., Thanabal, V., Erdman, P., Cox, P. B., et al. (2009). Discovery of antibacterial biotin carboxylase inhibitors by virtual screening and fragment-based approaches. *ACS Chem. Biol.* 4, 473–483. doi: 10.1021/cb9000102
- Morris, R. J., Najmanovich, R. J., Kahraman, A., and Thornton, J. M. (2005). Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics* 21, 2347–2355. doi: 10.1093/bioinformatics/bti337
- Morro, A., Canals, V., Oliver, A., Alomar, M. L., Galán-Prado, F., Ballester, P. J., et al. (2018). A stochastic spiking neural network for virtual screening. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 1371–1375. doi: 10.1109/TNNLS.2017.2657601
- Nagarajan, K., Zauhar, R., and Welsh, W. J. (2005). Enrichment of ligands for the serotonin receptor using the shape signatures approach. *J. Chem. Inf. Model.* 45, 49–57. doi: 10.1021/ci049746x
- Naylor, E., Arredouani, A., Vasudevan, S. R., Lewis, A. M., Parkesh, R., Mizote, A., et al. (2009). Identification of a chemical probe for NAADP by virtual screening. *Nat. Chem. Biol.* 5, 220–226. doi: 10.1038/nchembio.150
- Nelder, J. A., and Mead, R. (1965). A simplex method for function minimization. *Comput. J.* 7, 308–313. doi: 10.1093/comjnl/7.4.308
- Nguyen, K. T., Blum, L. C., van Deursen, R., Reymond, J. L. (2009). Classification of Organic Molecules by Molecular Quantum Numbers. *ChemMedChem* 4, 1803–1805. doi: 10.1002/cmdc.200900317
- Nilakantan, R., Bauman, N., and Venkataraghavan, R. (1993). New method for rapid characterization of molecular shapes: applications in drug design. *J. Chem. Inf. Comput. Sci.* 33, 79–85. doi: 10.1021/ci00011a012
- Novotni, M., and Klein, R. (2003). “3D zernike descriptors for content based shape retrieval,” in *Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications* (Seattle, WA).
- Osguthorpe, D. J., Sherman, W., and Hagler, A. T. (2012). Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chem. Biol. Drug Des.* 80, 182–193. doi: 10.1111/j.1747-0285.2012.01396.x
- Pala, D., Castelli, R., Incerti, M., Russo, S., Tognolini, M., Giorgio, C., et al. (2014). Combining ligand- and structure-based approaches for the discovery of new inhibitors of the epha2-ephrin-a1 interaction. *J. Chem. Inf. Model.* 54, 2621–2626. doi: 10.1021/ci5004619

- Patel, S., Harris, S. F., Gibbons, P., Deshmukh, G., Gustafson, A., Kellar, T., et al. (2015). Scaffold-hopping and structure-based discovery of potent, selective, and brain penetrant N-(1H-Pyrazol-3-yl)pyridin-2-amine inhibitors of dual leucine zipper kinase (DLK, MAP3K12). *J. Med. Chem.* 58, 8182–8199. doi: 10.1021/acs.jmedchem.5b01072
- Patil, S. P., Ballester, P. J., and Kerezi, C. R. (2014). Prospective virtual screening for novel p53-MDM2 inhibitors using ultrafast shape recognition. *J. Comput. Aided Mol. Des.* 28, 89–97. doi: 10.1007/s10822-014-9732-4
- Pérez-Nuño, V. I., and Ritchie, D. W. (2011). Using consensus-shape clustering to identify promiscuous ligands and protein targets and to choose the right query for shape-based virtual screening. *J. Chem. Inf. Model.* 51, 1233–1248. doi: 10.1021/ci100492r
- Pérez-Nuño, V. I., Ritchie, D. W., Borrell, J. I., and Teixidó, J. (2008). Clustering and classifying diverse hiv entry inhibitors using a novel consensus shape-based virtual screening approach: further evidence for multiple binding sites within the ccr5 extracellular pocket. *J. Chem. Inf. Model.* 48, 2146–2165. doi: 10.1021/ci800257x
- Perez-Nuño, V. I., Venkatraman, V., Mavridis, L., and Ritchie, D. W. (2011). Predicting drug promiscuity using spherical harmonic surface shape-based similarity comparisons. *Open Conf. Proc. J.* 2, 113–129. doi: 10.2174/2210289201102010113
- Poongavanam, V., and Kongsted, J. (2013). Virtual screening models for prediction of HIV-1 RT associated RNase H Inhibition. *PLoS ONE* 8:e73478. doi: 10.1371/journal.pone.0073478
- Proschak, E., Rupp, M., Derksen, S., and Schneider, G. (2008). Shapelets: possibilities and limitations of shape-based virtual screening. *J. Comput. Chem.* 29, 108–114. doi: 10.1002/jcc.20770
- Renner, S., and Schneider, G. (2006). Scaffold-hopping potential of ligand-based similarity concepts. *ChemMedChem* 1, 181–185. doi: 10.1002/cmdc.200500005
- Ritchie, D. W., and Kemp, G. J. L. (1999). Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *J. Comput. Chem.* 20, 383–395. doi: 10.1002/(SICI)1096-987X(199903)20:4<383::AID-JCC1>3.0.CO;2-M
- Ritchie, D. W., and Kemp, G. J. (2000). Protein docking using spherical polar Fourier correlations. *Proteins* 39, 178–194. doi: 10.1002/(SICI)1097-0134(20000501)39:2<178::AID-PROT8>3.0.CO;2-6
- Rogers, D. J., and Tanimoto, T. T. (1960). A computer program for classifying plants. *Science* 132, 1115–1118. doi: 10.1126/science.132.3434.1115
- Roy, A., Srinivasan, B., and Skolnick, J. (2015). PoLi: a virtual screening pipeline based on template pocket and ligand similarity. *J. Chem. Inf. Model.* 55, 1757–1770. doi: 10.1021/acs.jcim.5b00232
- Rush, T. S.III., Grant, J. A., Mosyak, L., and Nicholls, A. (2005). A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* 48, 1489–1495. doi: 10.1021/jm040163o
- Sael, L., and Kihara, D. (2010). Improved protein surface comparison and application to low-resolution protein structure data. *BMC Bioinformatics* 11:S2. doi: 10.1186/1471-2105-11-S11-S2
- Sael, L., and Kihara, D. (2012). Detecting local ligand-binding site similarity in nonhomologous proteins by surface patch comparison. *Proteins* 80, 1177–1195. doi: 10.1002/prot.24018
- Sael, L., La, D., Li, B., Rustamov, R., and Kihara, D. (2008a). Rapid comparison of properties on protein surface. *Proteins* 73, 1–10. doi: 10.1002/prot.22141
- Sael, L., Li, B., La, D., Fang, Y., Ramani, K., Rustamov, R., et al. (2008b). Fast protein tertiary structure retrieval based on global surface shape similarity. *Proteins* 72, 1259–1273. doi: 10.1002/prot.22030
- Salo, H. S., Laitinen, T., Poso, A., Jarho, E., and Lahtela-Kakkonen, M. (2013). Identification of novel SIRT3 inhibitor scaffolds by virtual screening. *Bioorg. Med. Chem. Lett.* 23, 2990–2995. doi: 10.1016/j.bmcl.2013.03.033
- Sastry, G. M., Dixon, S. L., and Sherman, W. (2011). Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *J. Chem. Inf. Model.* 51, 2455–2466. doi: 10.1021/ci2002704
- Sato, T., Yuki, H., Takaya, D., Sasaki, S., Tanaka, A., and Honma, T. (2012). Application of support vector machine to three-dimensional shape-based virtual screening using comprehensive three-dimensional molecular shape overlay with known inhibitors. *J. Chem. Inf. Model.* 52, 1015–1026. doi: 10.1021/ci200562p
- Schnecke, V., and Boström, J. (2006). Computational chemistry-driven decision making in lead generation. *Drug Discov. Today* 11, 43–50. doi: 10.1016/S1359-6446(05)03703-7
- Schreyer, A., and Blundell, T. (2009). CREDO: a protein-ligand interaction database for drug discovery. *Chem. Biol. Drug Des.* 73, 157–167. doi: 10.1111/j.1747-0285.2008.00762.x
- Schreyer, A. M., and Blundell, T. (2012). USRCAT: real-time ultrafast shape recognition with pharmacophoric constraints. *J. Cheminform.* 4, 27–27. doi: 10.1186/1758-2946-4-27
- Shang, J., Dai, X., Li, Y., Pistolozzi, M., and Wang, L. (2017). HybridSimVS: a web server for large-scale ligand-based virtual screening using hybrid similarity recognition techniques. *Bioinformatics* 33, 3480–3481. doi: 10.1093/bioinformatics/btx418
- Shave, S., Blackburn, E. A., Adie, J., Houston, D. R., Auer, M., Webster, S. P., et al. (2015). UFSRAT: ultra-fast shape recognition with atom types – the discovery of novel bioactive small molecular scaffolds for FKBP12 and 11βHSD1. *PLoS ONE* 10:e0116570. doi: 10.1371/journal.pone.0116570
- Shave, S., Mcguire, K., Pham, N. T., Mole, D. J., Webster, S. P., and Auer, M. (2018). Diclofenac identified as a kynurenine 3-monooxygenase binder and inhibitor by molecular similarity techniques. *ACS Omega* 3, 2564–2568. doi: 10.1021/acsomega.7b02091
- Shave, S. R. (2010). *Development of High Performance Structure and Ligand Based Virtual Screening Techniques*. Ph.D., The University of Edinburgh.
- Sheridan, R. P., and Kearsley, S. K. (2002). Why do we need so many chemical similarity search methods? *Drug Discov. Today* 7, 903–911. doi: 10.1160.10/S1359-6446(02)02411-X
- Shin, W. H., Bures, M. G., and Kihara, D. (2016a). PatchSurfers: two methods for local molecular property-based binding ligand prediction. *Methods* 93, 41–50. doi: 10.1016/j.ymeth.2015.09.026
- Shin, W.-H., Christoffer, C. W., Wang, J., and Kihara, D. (2016b). PL-PatchSurfer2: improved local surface matching-based virtual screening method that is tolerant to target and ligand structure variation. *J. Chem. Inf. Model.* 56, 1676–1691. doi: 10.1021/acs.jcim.6b00163
- Silverman, B. D., and Platt, D. E. (1996). Comparative Molecular Moment Analysis (CoMMA): 3D-QSAR without Molecular Superposition. *J. Med. Chem.* 39, 2129–2140. doi: 10.1021/jm950589q
- Sun, H., Xu, X., Wu, X., Zhang, X., Liu, F., Jia, J., et al. (2013). Discovery and design of tricyclic scaffolds as protein kinase ck2 (ck2) inhibitors through a combination of shape-based virtual screening and structure-based molecular modification. *J. Chem. Inf. Model.* 53, 2093–2102. doi: 10.1021/ci400114f
- Suzuki, H., Kawabata, T., and Nakamura, H. (2016). Omokage search: shape similarity search service for biomolecular structures in both the PDB and EMDB. *Bioinformatics* 32, 619–620. doi: 10.1093/bioinformatics/btv614
- Swann, S. L., Brown, S. P., Muchmore, S. W., Patel, H., Merta, P., Locklear, J., et al. (2011). A unified, probabilistic framework for structure- and ligand-based virtual screening. *J. Med. Chem.* 54, 1223–1232. doi: 10.1021/jm1013677
- Swinney, D. C., and Anthony, J. (2011). How were new medicines discovered? *Nat. Rev. Drug Discov.* 10:507. doi: 10.1038/nrd3480
- Tao, Z., Wei, C., Min, H., and Qunsheng, P. (2005). “A similarity computing algorithm for proteins,” in *Ninth International Conference on Computer Aided Design and Computer Graphics (CAD-CG'05)* (Hong Kong), 5.
- Temml, V., Voss, C. V., Dirsch, V. M., and Schuster, D. (2014). Discovery of new liver X receptor agonists by pharmacophore modeling and shape-based virtual screening. *J. Chem. Inf. Model.* 54, 367–371. doi: 10.1021/ci400682b
- Teo, C. Y., Rahman, M. B. A., Chor, A. L. T., Salleh, A. B., Ballester, P. J., and Tejo, B. A. (2013). Ligand-Based Virtual Screening for the discovery of inhibitors for Protein Arginine Deiminase Type 4 (PAD4). *Metabolomics* 3:1000118. doi: 10.4172/2153-0769.1000118
- Tversky, A. (1977). Features of similarity. *Psychol. Rev.* 84, 327–352. doi: 10.1037/0033-295X.84.4.327
- van Deursen, R., Blum, L. C., and Reymond, J. L. (2010). A searchable map of pubchem. *J. Chem. Inf. Model.* 50, 1924–1934. doi: 10.1021/ci100237q
- Vainio, M. J., Puranen, J. S., and Johnson, M. S. (2009). ShaEP: molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.* 49, 492–502. doi: 10.1021/ci800315d
- Vasudevan, S. R., Moore, J. B., Schymura, Y., and Churchill, G. C. (2012). Shape-based reprofiling of FDA-approved drugs for the H1 histamine receptor. *J. Med. Chem.* 55, 7054–7060. doi: 10.1021/jm300671m

- Vasudevan, S. R., Singh, N., and Churchill, G. C. (2014). Scaffold hopping with virtual screening from ip3 to a drug-like partial agonist of the inositol triphosphate receptor. *ChemBioChem* 15, 2774–2782. doi: 10.1002/cbic.201402440
- Vaz de Lima, L. A., and Nascimento, A. S. (2013). MolShaCS: a free and open source tool for ligand similarity identification based on Gaussian descriptors. *Eur. J. Med. Chem.* 59, 296–303. doi: 10.1016/j.ejmech.2012.11.013
- Venkatraman, V., Chakravarthy, P. R., and Kihara, D. (2009a). Application of 3D Zernike descriptors to shape-based ligand similarity searching. *J. Cheminform.* 1:19. doi: 10.1186/1758-2946-1-19
- Venkatraman, V., Pérez-Nueno, V. I., Mavridis, L., and Ritchie, D. W. (2010). Comprehensive comparison of ligand-based virtual screening tools against the DUD Data set reveals limitations of current 3D methods. *J. Chem. Inf. Model.* 50, 2079–2093. doi: 10.1021/ci100263p
- Venkatraman, V., Sael, L., and Kihara, D. (2009b). Potential for protein surface shape analysis using spherical harmonics and 3D Zernike descriptors. *Cell Biochem. Biophys.* 54, 23–32. doi: 10.1007/s12013-009-9051-x
- Vidović, D., Busby, S. A., Griffin, P. R., and Schürer, S. C. (2011). A combined ligand- and structure-based virtual screening protocol identifies submicromolar PPAR $\gamma$  Partial Agonists. *ChemMedChem* 6, 94–103. doi: 10.1002/cmdc.201000428
- Voet, A., Berenger, F., and Zhang, K. Y. (2013). Electrostatic similarities between protein and small molecule ligands facilitate the design of protein-protein interaction inhibitors. *PLoS ONE* 8:e75762. doi: 10.1371/journal.pone.0075762
- Wang, C. Y., Ai, N., Arora, S., Erenrich, E., Nagarajan, K., Zauhar, R., et al. (2006). Identification of previously unrecognized antiestrogenic chemicals using a novel virtual screening approach. *Chem. Res. Toxicol.* 19, 1595–1601. doi: 10.1021/tx060218k
- Wang, L., Si, P., Sheng, Y., Chen, Y., Wan, P., Shen, X., et al. (2015). Discovery of new non-steroidal farnesoid X receptor modulators through 3D shape similarity search and structure-based virtual screening. *Chem. Biol. Drug Des.* 85, 481–487. doi: 10.1111/cbdd.12432
- Wang, Q., Birod, K., Angioni, C., Grösch, S., Geppert, T., Schneider, P., et al. (2011). Spherical harmonics coefficients for ligand-based virtual screening of cyclooxygenase inhibitors. *PLoS ONE* 6:e21554. doi: 10.1371/journal.pone.0021554
- Wavhale, R. D., Martis, E. A. F., Ambre, P. K., Wan, B., Franzblau, S. G., and Iyer, K. R. (2017). Discovery of new leads against Mycobacterium tuberculosis using scaffold hopping and shape based similarity. *Biorg. Med. Chem.* 25, 4835–4844. doi: 10.1016/j.bmc.2017.07.034
- Wei, N. N., and Hamza, A. (2014). SABRE: ligand/structure-based virtual screening approach using consensus molecular-shape pattern recognition. *J. Chem. Inf. Model.* 54, 338–346. doi: 10.1021/ci4005496
- Werner, M. M., Li, Z., and Zauhar, R. J. (2014). Computer-aided identification of novel 3,5-substituted rhodanine derivatives with activity against Staphylococcus aureus DNA gyrase. *Biorg. Med. Chem.* 22, 2176–2187. doi: 10.1016/j.bmc.2014.02.020
- Wiggers, H. J., Rocha, J. R., Fernandes, W. B., Sesti-Costa, R., Carneiro, Z. A., Chelieski, J., et al. (2013). Non-peptidic cruzain inhibitors with trypanocidal activity discovered by virtual screening and *in vitro* assay. *PLoS Negl. Trop. Dis.* 7:e2370. doi: 10.1371/journal.pntd.0002370
- Wilson, J. A., Bender, A., Kaya, T., and Clemons, P. A. (2009). Alpha shapes applied to molecular shape characterization exhibit novel properties compared to established shape descriptors. *J. Chem. Inf. Model.* 49, 2231–2241. doi: 10.1021/ci900190z
- Wu, G., and Vieth, M. (2004). SDOCKER: a method utilizing existing x-ray structures to improve docking accuracy. *J. Med. Chem.* 47, 3142–3148. doi: 10.1021/jm040015y
- Xia, G., Xue, M., Liu, L., Yu, J., Liu, H., Li, P., et al. (2011). Potent and novel 11 $\beta$ -HSD1 inhibitors identified from shape and docking based virtual screening. *Biorg. Med. Chem. Lett.* 21, 5739–5744. doi: 10.1016/j.bmcl.2011.08.019
- Xia, J., Feng, B., Shao, Q., Yuan, Y., Wang, X., Chen, N., et al. (2017). Virtual screening against phosphoglycerate kinase 1 in quest of novel apoptosis inhibitors. *Molecules* 22:E1029. doi: 10.3390/molecules22061029
- Xiong, Y., Esquivel-Rodriguez, J., Sael, L., and Kihara, D. (2014). “3D-SURFER 2.0: web platform for real-time search and characterization of protein surfaces,” in *Protein Structure Prediction*, ed D. Kihara (New York, NY: Springer), 105–117.
- Yamagishi, M. E., Martins, N. F., Neshich, G., Cai, W., Shao, X., Beutrait, A., et al. (2006). A fast surface-matching procedure for protein–ligand docking. *J. Mol. Model.* 12, 965–972. doi: 10.1007/s00894-006-0109-z
- Yan, X., Li, J., Gu, Q., and Xu, J. (2014). gWEGA: GPU-accelerated WEGA for molecular superposition and shape comparison. *J. Comput. Chem.* 35, 1122–1130. doi: 10.1002/jcc.23603
- Yan, X., Li, J., Liu, Z., Zheng, M., Ge, H., and Xu, J. (2013). Enhancing molecular shape comparison by weighted gaussian functions. *J. Chem. Inf. Model.* 53, 1967–1978. doi: 10.1021/ci300601q
- Yeturu, K., and Chandra, N. (2008). PocketMatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics* 9:543. doi: 10.1186/1471-2105-9-543
- Yin, S., and Dokholyan, N. V. (2011). Fingerprint-based structure retrieval using electron density. *Proteins* 79, 1002–1009. doi: 10.1002/prot.22941
- Zauhar, R. J., Gianti, E., and Welsh, W. J. (2013). Fragment-based Shape Signatures: a new tool for virtual screening and drug discovery. *J. Comput. Aided Mol. Des.* 27, 1009–1036. doi: 10.1007/s10822-013-9698-7
- Zauhar, R. J., Moyna, G., Tian, L., Li, Z., and Welsh, W. J. (2003). Shape signatures: a new approach to computer-aided ligand- and receptor-based drug design. *J. Med. Chem.* 46, 5674–5690. doi: 10.1021/jm030242k
- Zhou, T., Lafleur, K., and Cafilisch, A. (2010). Complementing ultrafast shape recognition with an optical isomerism descriptor. *J. Mol. Graph. Model.* 29, 443–449. doi: 10.1016/j.jmkgm.2010.08.007
- Zoete, V., Daina, A., Bovigny, C., and Michielin, O. (2016). SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. *J. Chem. Inf. Model.* 56, 1399–1404. doi: 10.1021/acs.jcim.6b00174

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer XL and handling Editor declared their shared affiliation.

Copyright © 2018 Kumar and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.