Published in final edited form as:

Trends Pharmacol Sci. 2009 March ; 30(3): 138–147. doi:10.1016/j.tips.2008.12.001.

The Importance of Discerning Shape in Molecular Pharmacology

Sandhya Kortagere[†], Matthew D. Krasowski[§], and Sean Ekins^{‡,§,*}

[†]Department of Microbiology and Immunology, Drexel University College of Medicine, Philadelphia, PA 19129, USA

[§]Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15261, USA

[‡]Collaborations in Chemistry, Jenkintown, PA 19046, USA

[§]Department of Pharmaceutical Sciences, University of Maryland, Baltimore, MD 21202, USA

^{*}Department of Pharmacology, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA

Abstract

Shape is a fundamentally important molecular feature that often determines the fate of a compound in terms of molecular interactions with preferred and non-preferred biological targets.

Complementarity of binding in small molecule-protein, peptide-receptor, antigen-antibody and protein-protein interactions is key to life and survival, but also to targeting molecules with bioactivity. We review the application of shape in various biological systems such as substrate recognition, ligand specificity / selectivity and antibody recognition in the context of computational methods such as docking, quantitative structure activity relationships, classification models and similarity search algorithms. These *in silico* pharmacology methods have recently demonstrated the importance and applicability of determining molecular shape in drug discovery, virtual screening and predictive toxicology. The results from recently published studies show that shape and shape-based descriptors are at least as useful as other traditional molecular descriptors.

Keywords

Antibody; Depth; Descriptors; Dopamine receptors; Molecular shape; Nuclear hormone receptor; Pharmacophore

Introduction

Form follows function - that has been misunderstood. Form and function should be one, joined in a spiritual union.

Frank Lloyd Wright, 1908

It has been long recognized that determining molecular shape (Box 1) and changes in this property is essential in order to understand, molecules involved in chemical reactions ¹. Enzymes can differentiate between functional groups in a molecule through shape recognition; also, natural products produced through biosynthetic pathways involve shape recognition for selective oxidation ². More recently, chiral recognition was shown at the single-molecule level to involve mutually induced conformational adjustments ³. At lower concentrations of

* To whom correspondence should be addressed. Collaborations in Chemistry, 601 Runnymede Ave, Jenkintown, PA 19046. Phone: 269-930-0974. Fax: 215-481-0159. ekinssean@yahoo.com.

magnesium ions, ribozymes from *Escherichia coli* and *Bacillus subtilis* recognize cloverleaf shape RNAs rather than hairpin shape RNAs, indicating shape recognition⁴. Chemical shape interaction has a key role in the senses⁵: smell (via hundreds of olfactory receptors), sight (via receptors responsible for the perception of color), and taste (via receptors responsible for the perception of bitter, sour, sweet, salt and umami), and all of these receptors are G-protein coupled receptors. Various studies have reinforced the idea that molecular shape plays a major role in biological activity⁶ and references therein). Addition of other features to molecular shape would be expected to increase specificity such as complementary electrostatic or steric interactions.

Irrespective of the type of definition used, the essence of shape is therefore very useful in describing molecule(s) by themselves or the nature of interactions between molecules. Hence, the study of shape in molecular pharmacology has gained importance due to its applicability in drug design process *in silico* techniques (Table 1) widely employed to decrease the costs of drug discovery and development. These computational methods can enable rapid comparisons between small molecules, or small molecules with protein receptor sites, mainly based on their shape and other properties such as electrostatics. This review explores the various definitions of shape generally used when describing a molecule or interaction between molecules and provides examples of biological systems where the concept of shape plays a major role. The applicability of shape in *in silico* techniques is also detailed along with future developments for *in silico* pharmacology. Several published studies illustrate that shape-based methods and descriptors in various classification and other modeling schemes are as useful as traditional molecular properties, like 2D descriptors⁷⁻⁹.

Visualizing molecular shape in biological systems

There are many examples of applications of computational methods using shape (Table 2). Simple molecular shape analysis by determining the van der Waals volume of active and inactive compounds can be insightful for enzymes, for example, in elucidating substrate recognition by *Pseudomonas fluorescens* N3 dioxygenase¹⁰ and in visualizing differences in inhibitors for human cytochrome P450 (CYP) 51¹¹. Shape descriptors (Box 2) have been found to be important in some recent computational models. For example, a model for protein-protein interaction inhibitors found the shape descriptor SHP2 at the top of the decision tree¹². Sammon and Kohonen mapping human ether-ago-go potassium ion channel models contained Wiener and Balaban index descriptors, suggesting that molecular shape or topological characteristics were important for binding to this ion channel¹³.

Ligand-protein interactions

The co-evolution of molecule protein interactions with regard to shape has been explored in the study of nuclear hormone receptors (NHRs) that recognize bile salts^{14, 15}. Bile salts are the main end-metabolites of cholesterol in vertebrate animals. Evolutionarily early vertebrates such as jawless fish (lampreys and hagfish) use planar 5 α ('allo') bile salts¹⁵. In contrast, many other vertebrates, including humans and most mammals, use 5 β bile salts that have a 'bend' at the junction of the A and B steroid rings. Cross-species comparisons of the selectivity of the farnesoid X receptor (FXR; 'bile acid receptor') for structurally diverse bile salts showed that FXR has changed selectivity for bile salts from preference for 5 α (flat) bile salts ('ancestral' pattern in sea lamprey and zebrafish) to a preference for 5 β bile salts in humans and mice ('recent' pattern). Computational homology models predicted that this selectivity change was mediated by altering the shape and size of the ligand binding pocket¹⁵. Using similar computational approaches, vertebrate liver X receptors (LXR, 'oxysterol receptor') have also been found to diverge from invertebrate LXR in their ligand specificity¹⁶. Analyses of cross-species differences in receptor binding sites can also be inferred using a ligand-based approach such as a pharmacophore. For example, pregnane X receptors (PXR) (broad specificity nuclear

hormone receptor (NHR) involved in regulation of liver metabolism¹⁷), show significant cross-species differences in ligand specificity, with a broadening of ligand specificity from teleost fish to mammals and birds^{14, 18}. These NHRs represent robust model systems to explore the co-evolution of receptors and ligands in terms of shape and size.

In a recent study on identifying highly selective dopamine D4 receptor agonists and antagonists, shape and charge complementarity between the ligand and the receptor microdomain was found to play a major role in the functioning of the 1,4-disubstituted aromatic piperidines and piperazine inhibitors (Figure 1)¹⁹. Thus, the presence of structurally compatible regions between the receptor and its ligand is required for the functioning of these compounds.

Antibody-antigen interactions

Immunoassays based on antibodies are widely used in clinical medicine²⁰. Common applications include drug of abuse (DOA) screening, endocrinology testing, and therapeutic drug monitoring (TDM). Immunoassays are also employed as sensors for the detection of chemical warfare agents such as nerve gases and environmental pollutants²¹. Immunoassays may have either narrow specificity for a single target compound (e.g., a drug, vitamin, toxin, or hormone) or broader specificity for a group of structurally related target compounds (e.g., benzodiazepines, opiates). Antibodies use molecular shape to recognize the antigen and this may present problems with similar shaped but distinct molecules. Immunoassays are limited by the occurrence of false positives (or 'cross-reactive' compounds), defined as a positive result in the absence of the target compound(s). False positives are a particular limitation of DOA screening assays. For example, many drugs with structural similarity to amphetamine and methamphetamine such as pseudoephedrine and bupropion can cross-react with amphetamine screening tests. There have been some studies of the three-dimensional structure of antibodies bound to drugs that are the target of DOA screening tests or TDM assays, providing insight into the specificity of antibody-drug interactions. For example, in X-ray crystallographic structures of antibodies complexed with cocaine²² (additionally supported by molecular modeling studies²³), the antibody interacts with all portions of the target molecule.

There have been few efforts to use computational methods to predict and identify cross-reactive compounds for clinically used immunoassays. In initial studies, we have looked at molecular shape in terms of 2D similarity of test compounds to that of the antigen, using the MDL public keys fingerprint descriptors. A database of frequently used Food and Drug Administration (FDA)-approved drugs derived from the Clinician's Pocket Drug Reference supplemented with drugs of abuse and drug metabolites (n = 813) important in clinical toxicology was used for searching. We found that 'within-class' true positives for three urine toxicology screening assays (barbitures, benzodiazepines, and tricyclic antidepressants) tend to have high similarity to the target compound (Figure 2A). Particularly for benzodiazepines, however, some within-class compounds (e.g., clobazam and clonazepam) have lower similarity to the target compound diazepam, reflecting the diversity of structural modifications on the basic benzodiazepine core structure. Similarity methods provide a means of quantitatively rationalizing why compounds like clobazam and clonazepam generate 'false negative' results in an assay where diazepam was used to generate the assay antibodies.

In addition, other computational approaches could be used such as 3D methods that require the development of a pharmacophore or pharmacophoric pattern representing the arrangement of the chemical features and distances between them important for binding^{24, 25} (e.g. alignment of desipramine and amitriptyline). The advantage of the 3D similarity approach is that it is able to find structural matches that may not look similar in 2D, but possess the key features for 3D mapping. For example, a pharmacophore for amitriptyline and desipramine that was used to search the same database of over 800 drugs and metabolites retrieved over 150 hits including many non-tricyclic compounds (Figure 3). The number of hits can be filtered further to 79 by

addition of a van der Waals surface around one of the molecules accounting for shape of the molecule (Figure 3B, 3C). The approaches described above represent unique ways to search for potential cross reacting compounds based on shape that corresponds to some degree with the antibody-antigen interaction.

Shape representations in *in silico* methods

Given the enormous domain of shape-based features, it is challenging to characterize shape in algorithms for drug discovery. Several docking and scoring algorithms (Table 1) typically use shape either in implicit form or explicitly. In an explicit representation, the shape of the ligand is used to position it in the binding site of the target (usually a protein) with multiple conformer evaluations. This feature can be used to guide the rest of the docking process. Examples of the many available docking algorithms that use an explicit shape method are DOCK²⁶, GOLD²⁷, GLIDE¹² and AutoDock²⁸. Although some of these methods may also use shape implicitly through their scoring functions (the reader is referred to the individual references for further detail on each method). In an implicit representation, the shape of a molecule is used as a first level of filter or in a “screening” mode to identify if the molecule fits into the target or not. The soft docking method²⁹ and PUZZLE³⁰ (which use the shape complementarity between the surface of two interacting molecules as a filter) and many other programs use various representations for shape implicitly.

Another way to describe shape is to represent it in the form of molecular descriptors. Many different molecular shape descriptors have been proposed so far in the literature for small molecules and polymers (Box-2). It had been suggested over twenty years ago that a descriptor should obey all of the following properties to be useful, namely it should: a) describe local shape b) be invariant to coordinate transformations and c) enable determination of shape complementarity between 2 or more compounds³¹. One group³² added an additional requirement that the shape descriptor must provide a means of positioning the shape that it describes in a few canonical orientations. These shape descriptors can be used together with a distance metric such as the Tanimoto coefficient and/or quantitative structure activity relationship (QSAR) and quantitative structure property relationship (QSPR) algorithms (Table 2).

Similar to these descriptor notations is another collection of shape representations called the “shape catalog”. For a set of active compounds, a database or catalog of fingerprints can be created by considering all possible shapes that arise from the conformations of these compounds³³. The main advantage of this catalog, apart from its application as a repository of molecular shapes, is that it can be customized by the user for the desired level of similarity cutoff between the compounds^{34, 35}.

A minimal description of molecular shape may also be rendered as a pharmacophore or pharmacophore fingerprints^{24, 25}, whereby the angle and distance between key molecular features imparts some degree of information on size and shape required for favorable molecular interactions. Another approach that can incorporate shape details (of both ligand and indirectly the protein binding site) is pseudoreceptor modelling, which represents a protein binding site using one or more molecules or conformations, and has been recently discussed in detail³⁶. Pseudoreceptor models can be used for searching databases for molecules with complementarity to the pseudoreceptor model.

Applications of shape descriptors to molecular pharmacology

Shape descriptors have been applied to many pharmacologically relevant problems as briefly summarized below. Shape descriptors are highly useful for classification purposes³⁷ and can be modified to act as weights to scoring functions⁹. Further, shape descriptors have been also

shown to be applicable to clustering molecules into groups that share similar overall shapes and hence to find analogs of lead compounds in a drug discovery application.

Shape similarity searches can be performed at two levels of complexity namely: global shape analysis and local shape analysis³⁸. If a complete structure of one molecule is matched with another, it is called a global shape analysis method. Examples of algorithms that use global shape analysis are the distance geometry method³⁹ and the molecular shape analysis (MSA) method that uses steric and van der Waals volumes as a shape descriptor⁴⁰. If a sub-structure of a complete molecule matches with another molecule then such an analysis is called local shape analysis. Many algorithms have incorporated the local shape analysis to identify compounds that not only share a part of their core structure with the query molecule but also other scaffolds, that may have shape similarity and not necessarily structural similarity³⁸. Be it the local or global similarity, the issue narrows down to finding “neighborhood” molecules using many clues such as a receptor structure⁴¹ or QSAR based classification results⁴². 3D similarity or classification studies can be performed using the following global alignment and alignment free methods.

Alignment-based methods

One approach to classifying compounds based on shape is to overlay them (alignment-based) using a set of guide points. An example is the shape-based and ligand centric approach called rapid overlay of chemical structures (ROCS, Openeye Scientific Software, Sante Fe) which performs overlays of conformers of a molecule of interest^{37, 43-47}. The overlays are performed quickly as the molecules are described as atom-centered Gaussian functions and conformers are compared using the Tanimoto coefficient. Adding chemical feature information is also possible and has been found useful for improving virtual screening results. ROCS has generally been found to perform as well as docking and other virtual screening methods (or better in the case of most targets)^{37, 38, 40} although with some exceptions⁴⁶. ROCS has also been used to discriminate between cruzain and cathepsin L inhibitors, as well as replicating the X-ray conformation of a known cruzain inhibitor⁴³.

Alignment-free methods

Quantum chemical derived molecular descriptors are normally considered computationally costly; however, fragment-based approaches including transferable atom equivalent (TAE) descriptors surmount this limitation^{48, 49} and are alignment free. The TAE descriptors have been used with machine learning methods to model 26 absorption, distribution, metabolism, excretion and toxicity related datasets demonstrating good internal model validation statistics in most cases⁵⁰. These descriptors have also been used to describe ligands and binding sites of proteins in the Protein Database⁵¹. Using a k nearest neighbor (kNN) pattern recognition approach and variable selection, the active site structure could identify its complementary ligand after screening 1% of the database in over 90% of cases when a representative family protein was present in the training set. The alignment free TAE descriptors can also be used to generate property-encoded surface translation (PEST) descriptors using a ray tracing approach reflected on the inside of the electron density isosurface. This provides 2D histograms of distances versus surface property, and each bin of the histogram becomes a molecular descriptor. These descriptors have been used with an olfactory database to train a genetic algorithm to distinguish between musk and non-musk compounds⁵². A very similar approach to PEST is the alignment free Shape Signatures method which results in compact histograms that encode for molecular shape, shape and polarity or other properties to produce signatures⁵³. This method has been used for database similarity searching in a series of drug discovery applications⁵⁴. The approach can also be used to build enriched or biased databases of small molecules by customizing the screening database to a particular drug target such as G-protein coupled receptors⁴⁹. For every molecule, the heights of the corresponding

normalized molecular shape, polarity and shape signature bins comprise sets of molecular descriptors which have been used for machine learning models of the 5-HT_{2B} receptor, the human ether-a-go-go potassium ion channel ⁷, and blood-brain barrier penetration ⁸. These models were also evaluated against and in combination with additional descriptors in the MOE suite of software (Chemical Computing Group, Montreal, Canada). It was found that Shape Signature descriptors describing molecular shape and charge slightly outperformed shape descriptors alone. These descriptors have been used recently to build classification models for PXR, as ⁹ well as a hybrid docking and molecular descriptor-based approach, coupling the GoldScore with other shape-based scoring functions.

Bioisosterism

The concept of bioisosterism (i.e. molecules with similar shapes may share similar biological activities) has been applied to drug discovery ^{55, 56}. In a study identifying angiotensin-II analogs, four query molecules were used to search for bioisosteres in a database of ~1000 compounds. Based on the search results, 425 compounds were synthesized and tested for angiotensin II inhibition. Of these 425 compounds, only 63 compounds that were identified by shape similarity search as being most similar to any of the four query structures were found to be active ⁵⁷. A good correlation between the shape scores and the inhibitory activity was found among all 425 compounds. However, algorithms that implement this concept have often been criticized for using the shape descriptor derived from a single representative conformer to search databases of large numbers of compounds. To overcome this limitation, many groups have now resorted to using at least 3-10 unique conformers and also to perform enrichment studies on the virtual libraries of compounds as an initial filter. Although successful stories like the one above validate bioisosterism, this hypothesis does not necessarily hold true with other types of descriptor or even some targets ^{58, 59}.

These studies and others (Table 2) using different computational methods illustrate the broad applicability of shape-based descriptors which are of value to the pharmaceutical and environmental sciences fields. Shape can adequately describe the ligand-protein interaction indirectly and be used for screening databases to obtain good enrichments with active compounds. Most of these algorithms also have the advantage of enabling fast comparisons ⁶ and are broadly applicable.

Concluding remarks

Shape is a fundamentally important molecular feature important for describing ligands interacting with receptors, ion channels, enzymes and transporters and an array of other proteins and complex biological processes. The shape of the protein or a (pseudo)receptor binding site can also be used to find molecules with complementarity or by using the crystal structure conformation of the ligands in the PDB as a shape-based search query. This might point to potential off-targets or alternate targets that may represent repurposing opportunities for known drugs. Shape-based approaches have many potential areas for development in the future as applied to *in silico* pharmacology ⁶⁰ (Box 3).

In the same way that an architect like Frank Lloyd Wright had an excellent appreciation of shape, thus it appears that his concept of *form and function* is generally applicable to shape and its role in molecular pharmacology.

Box 1. Shape and depth

The **shape** (OE. *sceap* Eng. *created thing*) of an object located in some space refers to the part of space occupied by the object as determined by its external boundary — abstracting from other aspects the object may have such as its color, content, or the object's position

and orientation in space, and its size. Shape can also be more generally defined as "the appearance of something, especially its outline". However, the definition of shape in molecular pharmacology does encompass structural features like depth, size and surface. As can be visualized by a simple Tupperware toy in which small shapes fit into complementary holes (Box Figure A), the importance of shape comes into play when a protein packs against another protein (Box Figure B) or when a small molecule is desired to fit into a binding site on a protein surface (Box Figure C), where size and shape complementarity may be essential in addition to favorable electrostatic and steric interactions, to fulfill the "lock and key" or "induced fit" hypotheses ⁶³.

Analogous to the term shape is "depth" which can be defined as the distance of an atom from the nearest surface water molecule ⁶⁴ that is usually applicable when describing shape based features in proteins such as grooves in DNA or binding pockets buried in membranes ⁶⁵. However, the term depth is more abstract and provides only a limited quantitative measure of shape.

Box 1. Examples to illustrate the importance of shape

Figure A. simplified example to demonstrate how a protein represented by the Tupperware toy, could have multiple potential binding sites of different shape and size, into which specific shaped molecules may fit. B. Analogous to the Tupperware toy is an example of favorable protein-protein interactions that makes the homotetrameric structure of turkey beta adrenergic receptor. C. An example of a protein with multiple binding sites: surface representation of dihydroorotate dehydrogenase of *Plasmodium falciparum* with bound

inhibitor A26 colored red, flavin mononucleotide colored yellow, orotic acid colored green and N8E colored blue.

Box 2. Shape as a molecular descriptor

Many different molecular shape descriptors have been proposed so far in the literature for small molecules and polymers. A review of all the different topological indices and their application to drug discovery is discussed in ⁶⁶ and is beyond the scope of this current review. Fragment or substructure based indices (also called Free-Wilson-analysis) are the 2D descriptors commonly used to describe molecular shape ⁶¹. Field based descriptors and others such as Shape Signatures ⁵³, Zernicke descriptors ⁶⁷, local intersection volume ⁶⁸ and path-space ratio ⁶⁹ use 3D information of the molecule and are generally more efficient and computationally intensive forms describing the molecule. Field methods in general have been broadly classified as quantum mechanics (QM) based descriptors (PEST and TAE) ⁷⁰ and non-QM methods such as Comparative Molecular Field Method (CoMFA) ⁴². Shape descriptors represent only an essence of the molecular shape by reducing the three dimensions to a set of numbers. Hence, these descriptors cannot be qualitatively used to ascertain ligand atoms responsible for hydrogen bonding to protein donor atoms.

Box 3. Future uses of molecular shape

The following represent some potential future uses of molecular shape:

1. Masking molecular shape - developing prodrugs that have improved physicochemical processes such as improved absorption or BBB permeability, or cloaking small molecules in nucleic acids and proteins ⁷¹.
2. Mimicking molecular shape - using computational methods to find new bioisosteres ^{55, 56}.
3. Molecular phylogeny - while protein binding sites tend to have geomorphic or other shape descriptions ⁷², small molecules could be described as rods, cyclic, cylindrical, spherical, curved, T-shaped, Y-shaped, crosses, asymmetrical etc, which can be used for simple classifications.
4. Molecular shape chimeras - molecules could be developed that can change the shape of the protein by inducing order to disorder changes such as that seen when rifampicin binds to human PXR ⁷³, for example.
5. Molecular shape confidentiality - Molecular shape descriptors could be used to share molecules between groups while the encryption key is confidential, building on the approaches used with structural fragments ⁷⁴ and similarity based approaches ⁷⁵.
6. Development of faster methods - required for mining shapes of multiple molecules simultaneously and potential creation of multiconformer shape databases consisting of tens to hundreds of millions of compounds ^{6, 76, 77}. To date the databases searched have been much smaller and there has been little discussion of time required to create them or limitations of these algorithms ⁶. Many of the studies currently benchmark against ROCS in terms of speed of molecular shape comparisons ⁶. This may also require the creation of freely available standardized test databases for future shape-based evaluations.
7. Shape fragment based discovery - the above computational descriptors of shape and corresponding algorithms could be used to develop methods for shape based

drug design analogous to fragment based drug discovery⁷⁸. This may also enable analysis of marketed drugs based on molecular shapes rather than as fragments⁷⁹ (e.g. answer the question of whether some molecular shapes more drug-like)

Acknowledgments

SE kindly acknowledges Penelope Ekins for lending her Tupperware toy and Accelrys for providing Discovery Studio. MDK is supported by K08-GM074238 from the National Institutes of Health. We gratefully acknowledge our collaborators from many of the studies described.

References

1. Mezey, PG. Shape in Chemistry. VCH Publishers Inc; 1993.
2. Hili R, Yudin AK. Making carbon-nitrogen bonds in biological and chemical synthesis. *Nat Chem Biol* 2006;2:284–287. [PubMed: 16710330]
3. Lingenfelder M, et al. Tracking the chiral recognition of adsorbed dipeptides at the single-molecule level. *Angew Chem Int Ed Engl* 2007;46:4492–4495. [PubMed: 17443762]
4. Ando T, et al. Bacterial ribonuclease P reaction is affected by substrate shape and magnesium ion concentration. *Nucleic Acids Res Suppl* 2003;293–294. [PubMed: 14510496]
5. Atkins, PW. Molecules. Scientific American Library; 1987.
6. Ballester PJ, Richards WG. Ultrafast shape recognition to search compound databases for similar molecular shapes. *J Comput Chem* 2007;28:1711–1723. [PubMed: 17342716]
7. Chekmarev DS, et al. Shape signatures: new descriptors for predicting cardiotoxicity in silico. *Chem Res Toxicol* 2008;21:1304–1314. [PubMed: 18461975]
8. Kortagere S, et al. New Predictive Models for Blood-Brain Barrier Permeability of Drug-like Molecules. *Pharm Res* 2008;25:1836–1845. [PubMed: 18415049]
9. Kortagere S, et al. Hybrid scoring and shape based classification approaches for human pregnane X receptor. submitted. 2008
10. Di Gennaro P, et al. Specificity of substrate recognition by *Pseudomonas fluorescens* N3 dioxygenase. The role of the oxidation potential and molecular geometry. *J Biol Chem* 1997;272:30254–30260. [PubMed: 9374510]
11. Ekins S, et al. Three-Dimensional Quantitative Structure-Activity Relationship Analysis of Human CYP51 Inhibitors. *Drug Metab Dispos* 2007;35:493–500. [PubMed: 17194716]
12. Neugebauer A, et al. Prediction of protein-protein interaction inhibitors by chemoinformatics and machine learning methods. *J Med Chem* 2007;50:4665–4668. [PubMed: 17705363]
13. Ekins S, et al. Insights for human Ether-a-Go-Go-Related Gene Potassium Channel inhibition using recursive partitioning, Kohonen and Sammon mapping Techniques. *J Med Chem* 2006;49:5059–5071. [PubMed: 16913696]
14. Krasowski MD, et al. Evolution of the pregnane X receptor: adaptation to cross-species differences in biliary bile salts. *Mol Endocrinol* 2005;19:1720–1739. [PubMed: 15718292]
15. Reschly EJ, et al. Evolution of the bile salt nuclear receptor FXR in vertebrates. *J Lipid Res* 2008;49:1577–1587. [PubMed: 18362391]
16. Reschly EJ, et al. Ligand specificity and evolution of liver X receptors. *J Steroid Biochem Mol Biol* 2008;110:83–94. [PubMed: 18395439]
17. Synold TW, et al. The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. *Nat Medicine* 2001;7:584–590.
18. Ekins S, et al. Evolution of pharmacologic specificity in the Pregane X Receptor. *BMC Evol Biol* 2008;8:103. [PubMed: 18384689]
19. Kortagere S, et al. Certain 1,4-disubstituted aromatic piperidines and piperazines with extreme selectivity for the dopamine D4 receptor interact with a common receptor microdomain. *Mol Pharmacol* 2004;66:1491–1499. [PubMed: 15448188]
20. Kricka, LJ. Principles of immunochemical techniques. In: Burtis, CA., et al., editors. *Tietz textbook of clinical chemistry and molecular diagnostics*. 4. Elsevier: Saunders; 2006.

21. Arduini F, et al. Fast, sensitive and cost-effective detection of nerve agents in the gas phase using a portable instrument and an electrochemical biosensor. *Anal Bioanal Chem* 2007;388:1049–1057. [PubMed: 17508205]
22. Larsen NA, et al. Crystal structure of a cocaine-binding antibody. *J Mol Biol* 2001;311:9–15. [PubMed: 11469854]
23. Paula S, et al. Three-dimensional structure-activity relationship modeling of cocaine binding to two monoclonal antibodies by comparative molecular field analysis. *J Mol Biol* 2003;325:515–530. [PubMed: 12498800]
24. Langer, T.; Hoffman, RD. Pharmacophores and pharmacophore searches. Wiley-VCH; 2006.
25. Guner, OF., editor. Pharmacophore, perception, development, and use in drug design. University International Line; 2000.
26. Ewing TJ, et al. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 2001;15:411–428. [PubMed: 11394736]
27. Jones G, et al. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997;267:727–748. [PubMed: 9126849]
28. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. *Proteins* 1990;8:195–202. [PubMed: 2281083]
29. Jiang F, Kim SH. “Soft docking”: matching of molecular surface cubes. *J Mol Biol* 1991;219:79–102. [PubMed: 2023263]
30. Helmer-Citterich M, Tramontano A. PUZZLE: a new method for automated protein docking based on surface shape complementarity. *J Mol Biol* 1994;235:1021–1031. [PubMed: 7507171]
31. Connolly ML. Shape complementarity at the hemoglobin alpha 1 beta 1 subunit interface. *Biopolymers* 1986;25:1229–1247. [PubMed: 3741993]
32. Goldman BB, Wipke WT. QSD quadratic shape descriptors. 2. Molecular docking using quadratic shape descriptors (QSDock). *Proteins* 2000;38:79–94. [PubMed: 10651041]
33. Lancot JK, et al. Using ensembles to classify compounds for drug discovery. *J Chem Inf Comput Sci* 2003;43:2163–2169. [PubMed: 14632468]
34. Putta S, et al. A novel shape-feature based approach to virtual library screening. *J Chem Inf Comput Sci* 2002;42:1230–1240. [PubMed: 12377013]
35. Srinivasen J, et al. Evaluation of a novel shape-based computational filter for lead evolution: application to thrombin inhibitors. *J Med Chem* 2002;45:2494–2500. [PubMed: 12036357]
36. Tanrikulu Y, Schneider G. Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nat Rev Drug Discov* 2008;7:667–677. [PubMed: 18636071]
37. Ebalunode JO, et al. Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *J Chem Inf Model* 2008;48:889–901. [PubMed: 18396858]
38. Robinson DD, et al. Partial molecular alignment via local structure analysis. *J Chem Inf Comput Sci* 2000;40:503–512. [PubMed: 10761157]
39. Crippen GM. Distance geometry approach to rationalizing binding data. *J Med Chem* 1979;22:988–997. [PubMed: 490543]
40. Hopfinger AJ. A general QSAR for dihydrofolate reductase inhibition by 2,4-diaminotriazines based upon molecular shape analysis. *Arch Biochem Biophys* 1981;206:153–163. [PubMed: 7212714]
41. Babine RE, Bender SL. Molecular Recognition of Protein-ligand Complexes: Applications to Drug Design. *Chem Rev* 1997;97:1359–1472. [PubMed: 11851455]
42. Cramer RD, et al. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 1988;110:5959–5967.
43. Freitas RF, et al. 2D QSAR and similarity studies on cruzain inhibitors aimed at improving selectivity over cathepsin L. *Bioorg Med Chem* 2008;16:838–853. [PubMed: 17996450]
44. Hawkins PC, et al. Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* 2007;50:74–82. [PubMed: 17201411]
45. Kirchmair J, et al. Fast and efficient in silico 3D screening: toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J Chem Inf Model* 2007;47:2182–2196. [PubMed: 17929799]

46. McGaughey GB, et al. Comparison of topological, shape, and docking methods in virtual screening. *J Chem Inf Model* 2007;47:1504–1519. [PubMed: 17591764]
47. Moffat K, et al. A comparison of field-based similarity searching methods: CatShape, FBSS, and ROCS. *J Chem Inf Model* 2008;48:719–729. [PubMed: 18351728]
48. Breneman C, et al. New developments in PEST shape/property hybrid descriptors. *J Comp-Aided Mol Des* 2003;17:231–240.
49. Whitehead C, et al. Transferable atom equivalent multicentered multipole expansion method. *J Comput Chem* 2003;24:512–529. [PubMed: 12594794]
50. Ekins, S., et al. Novel applications of Kernel-partial least squares to modeling a comprehensive array of properties for drug discovery. In: Ekins, S., editor. *Computational Toxicology: Risk assessment for pharmaceutical and environmental chemicals*. Wiley-Interscience; 2007. p. 403–432.
51. Oloff S, et al. Chemometric analysis of ligand receptor complementarity: identifying Complementary Ligands Based on Receptor Information (CoLiBRI). *J Chem Inf Model* 2006;46:844–851. [PubMed: 16563016]
52. Lavine BK, et al. Electronic van der Waals surface property descriptors and genetic algorithms for developing structure-activity correlations in olfactory databases. *J Chem Inf Comput Sci* 2003;43:1890–1905. [PubMed: 14632438]
53. Zauhar RJ, et al. Shape signatures: a new approach to computer-aided ligand- and receptor-based drug design. *J Med Chem* 2003;46:5674–5690. [PubMed: 14667221]
54. Nagarajan K, et al. Enrichment of ligands for the serotonin receptor using the Shape Signatures approach. *J Chem Inf Model* 2005;45:49–57. [PubMed: 15667128]
55. Jennings A, Tennant M. Selection of molecules based on shape and electrostatic similarity: proof of concept of “electroforms”. *J Chem Inf Model* 2007;47:1829–1838. [PubMed: 17824684]
56. Olesen PH. The use of bioisosteric groups in lead optimization. *Curr Opin Drug Discov Devel* 2001;4:471–478.
57. Cramer RD, et al. Prospective identification of biologically active structures by topomer shape similarity searching. *J Med Chem* 1999;42:3919–3933. [PubMed: 10508440]
58. Cramer RD, et al. Bioisosterism as a molecular diversity descriptor: steric fields of single “topomeric” conformers. *J Med Chem* 1996;39:3060–3069. [PubMed: 8759627]
59. Patterson DE, et al. Neighborhood behavior: a useful concept for validation of “molecular diversity” descriptors. *J Med Chem* 1996;39:3049–3059. [PubMed: 8759626]
60. Ekins S, et al. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol* 2007;152:9–20. [PubMed: 17549047]
61. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*. Wiley-VCH; 2000.
62. Willett P. Similarity-based approaches to virtual screening. *Biochem Soc Trans* 2003;31:603–606. [PubMed: 12773164]
63. Lill MA. Multi-dimensional QSAR in drug discovery. *Drug Discov Today* 2007;12:1013–1017. [PubMed: 18061879]
64. Varrazzo D, et al. Three-dimensional computation of atom depth in complex molecular structures. *Bioinformatics* 2005;21:2856–2860. [PubMed: 15827080]
65. Coleman RG, Sharp KA. Travel depth, a new shape descriptor for macromolecules: application to ligand binding. *J Mol Biol* 2006;362:441–458. [PubMed: 16934837]
66. Ciubotariu D, et al. Molecular van der Waals space and topological indices from the distance matrix. *Molecules* 2004;9:1053–1078. [PubMed: 18007504]
67. Mak L, et al. An extension of spherical harmonics to region-based rotationally invariant descriptors for molecular shape description and comparison. *J Mol Graph Model* 2008;26:1035–1045. [PubMed: 17905617]
68. Verli H, et al. Local intersection volume: a new 3D descriptor applied to develop a 3D-QSAR pharmacophore model for benzodiazepine receptor ligands. *Eur J Med Chem* 2002;37:219–229. [PubMed: 11900866]
69. Edvinsson T, et al. Path-space ratio as a molecular shape descriptor of polymer conformation. *J Chem Inf Comput Sci* 2003;43:126–133. [PubMed: 12546545]

70. Urbano-Cuadrado M, et al. New quantum mechanics-based three-dimensional molecular descriptors for use in QSSR approaches: application to asymmetric catalysis. *J Chem Inf Model* 2007;47:2228–2234. [PubMed: 17960899]
71. Hof F, Rebek J Jr. Molecules within molecules: recognition through self-assembly. *Proc Natl Acad Sci U S A* 2002;99:4775–4777. [PubMed: 11880604]
72. Lee M, et al. Shapes of antibody binding sites: qualitative and quantitative analyses based on a geomorphic classification scheme. *J Org Chem* 2006;71:5082–5092. [PubMed: 16808494]
73. Chrencik JE, et al. Structural disorder in the complex of human pregnane X receptor and the macrolide antibiotic rifampicin. *Mol Endocrinol* 2005;19:1125–1134. [PubMed: 15705662]
74. Trepalin S, Osadchiy N. The centroidal algorithm in molecular similarity and diversity calculations on confidential datasets. *J Comput Aided Mol Des* 2005;19:715–729. [PubMed: 16331405]
75. Kaiser D, et al. Similarity-based descriptors (SIBAR)--a tool for safe exchange of chemical information? *J Comput Aided Mol Des* 2005;19:687–692. [PubMed: 16249834]
76. Fontaine F, et al. Fast 3D shape screening of large chemical databases through alignment-recycling. *Chem Cent J* 2007;1:12. [PubMed: 17880744]
77. Mansfield ML, et al. A new class of molecular shape descriptors. 1. Theory and properties. *J Chem Inf Comput Sci* 2002;42:259–273. [PubMed: 11911695]
78. Siegal G, et al. Integration of fragment screening and library design. *Drug Discov Today* 2007;12:1032–1039. [PubMed: 18061882]
79. Siegel MG, Vieth M. Drugs in other drugs: a new look at drugs as fragments. *Drug Discov Today* 2007;12:71–79. [PubMed: 17198975]
80. Wang CY, et al. Identification of previously unrecognized antiestrogenic chemicals using a novel virtual screening approach. *Chem Res Toxicol* 2006;19:1595–1601. [PubMed: 17173372]

Glossary

CoMFA	Comparative molecular field analysis ⁴² is a widely used 3D-QSAR method that is alignment dependent. This method is useful for suggesting steric and electrostatic interactions between molecule and protein but is not used widely for virtual screening
Decision tree	is a predictive data mining method with a tree-like structure consisting of leaves which represent classifications and branches representing molecular features or descriptors that lead to the classifications ¹³
Docking Methods	such as DOCK ²⁶ , GOLD ²⁷ , GLIDE ¹² and AutoDock ²⁸ , soft docking method ²⁹ and PUZZLE ³⁰ etc are generally computational methods to place a (small) molecule into a binding site (normally a protein) in a manner such that optimal molecule-protein interactions result
<i>k</i> nearest neighbor	A method for classifying molecular bioactivity / properties based on similarity of molecular descriptors for a test molecule to those of the most similar molecules (nearest neighbors) in a training set
MDL public keys	Molecular Design Limited (MDL now Symyx) created a key based fingerprint which uses a pre-defined set of definitions (relating to structural features). A fingerprint is created based on pattern matching of the structure to this set of 166 keys
Molecular descriptors	are the end result of an experimental or mathematical procedure transforming chemical information into a useful number ⁶¹
Pharmacophore	the essential molecular features that contribute to bioactivity
QSAR and QSPR	Quantitative structure activity relationship and quantitative structure property relationship are algorithmic methods to relate molecular descriptors with bioactivity (e.g. K_i values) or biological /

	physicochemical property (e.g logP) data, respectively. These methods will provide a quantitative prediction once descriptors for a new molecule are input
ROCS	rapid overlay of chemical structures, enables search for similarity of molecular 3D shape. A recent study has suggested this method is slower than the Ultrafast shape recognition method ⁶
Sammon and Kohonen Mapping	are dimensionality reduction techniques frequently used with cheminformatics data. Sammon non-linear maps are unbiased in reproducing the topology and structure of the data space. Kohonen maps are a type of neural networks known as self-organizing networks consisting of artificial neurons connected by a distance dependent function. These neurons self-organize until their pairwise neighborhoods represent the correct topology of the original data set ¹³
Scaffold hopping	use computational methods to find novel compounds with similar bioactivity or physicochemical property to a known reference ligand but with different structural chemotype/s
Shape SignaturesTAE, PEST descriptors	all use a ray-tracing algorithm to explore the volume enclosed by the surface of a molecule, using the output to construct compact histograms ^{48, 49, 53}
Tanimoto coefficient	a measure of molecular similarity frequently used for comparing molecular fingerprint descriptors ⁶²

Figure 1. Models to show favorable and unfavorable ligand-protein interactions

A. structural model of the wild type Dopamine D4 receptor with bound ligand L750,667 (stick model and colored atom type with C-green, O-red, N-blue, H-white, I-purple)¹⁹. The residues F2.61 and V3.28 in the second and third transmembrane domain (seen as orange sticks) aid in favorable interactions with the ligand. B. structural model of the wild type Dopamine D2 receptor with bound ligand L750,667. The residues V2.61 and F3.28 residues in TM2 and TM3 poses steric hindrance to the ligand.

Figure 2. Using 2D-Structural similarity to understand immunoassay false positives

A. 2D similarity of compounds in the database (FDA approved drugs and some metabolites) to the target compounds of several DOA immunoassays. B. Similarity of five compounds to the target compound desipramine (target of some tricyclic antidepressant screening assays) calculated by MDL public keys using the Tanimoto similarity measure. Of the five compounds, clomipramine is the most similar (Tanimoto similarity = 0.750) while ibuprofen is the least similar (Tanimoto similarity = 0.146).

Figure 3. Using 3D pharmacophores to understand immunoassay false positives

A. 3D-pharmacophore of amitriptyline and desipramine (yellow) generated with Catalyst Discovery Studio 2.0 (Accelrys, San Diego, CA) to illustrate how these could represent the shape of the antigen. B. Mapping of protriptyline (Fit value 3.964 and shape similarity 0.786) the best fitting molecule to the pharmacophore features and inside the van der Waals shape (shaded volume) after performing the pharmacophore-shape search. C. Mapping of clomipramine (Fit value 3.926 and shape similarity 0.761) to the pharmacophore features and inside the van der Waals shape. The cyan blobs represent hydrophobic features and the red blob represents a positive ionizable feature.

Table 1

Computational methods incorporating shape

Method	Examples	Uses	Pros and Cons
Docking using explicit shape	DOCK ²⁶ , GOLD ²⁷ , GLIDE ¹² and AutoDock ²⁸	Small molecule – protein interactions and protein-protein interactions	Value in database screening is linked with scoring algorithm used. Docking may be poorer at finding actives than ligand based methods ⁴⁶
Docking using implicit shape	soft docking method ²⁹ , PUZZLE ³⁰	Small molecule – protein interactions and protein-protein interactions	As above but the shape complementarity may be a useful filter.
Molecular shape descriptors	Topological torsions, Wiener indices, atom pair representations for small fragments ⁶¹ , Shape Signatures ⁵³ , Zernicke descriptors ⁶⁷ , local intersection volume ⁶⁸ , path-space ratio ⁶⁹ , ROCS ^{37, 43-47} , TAE ^{48, 49} , PEST ⁵² , Shape Signatures ⁵⁴ and others ⁷⁰ .	Use with Molecular similarity indices, QSAR, QSPR, machine learning methods for classification	Descriptor use may depend widely on data modeled, algorithm used and combination of descriptors evaluated. Some descriptors may require longer to calculate than others. Applicability domain of models may also be a limitation depending on the training set and the coverage of chemical space.
Pharmacophores	DISCO, GASP, Catalyst, Apex 3D, GALAHAD, Phase ^{24, 25}	Database searching, QSAR	The method is applicable when there is no target 3D structure available. Pharmacophore searching is faster than docking but likely not as fast as descriptor based methods for filtering compound libraries. Potential to find structural matches that may not look similar in 2D, but possess the key features for 3D mapping.
Pharmacophore fingerprints	ChemX/ ChemDiverse, PharmPrint, OSPREYS, 3D Keys and Tuptlets ^{24, 25}	Use with Molecular similarity indices, QSAR, QSPR, machine learning methods for classification	Fingerprints are useful for very rapid comparisons of molecules and similarity analyses.
Pseudoreceptor models	QUASAR, RAPTOR, and many others see previous review ³⁶ .	Exploration of ligand receptor interactions	As with pharmacophores this method has utility when there is no target 3D structure available. Potential for finding novel binding modes. Alignment dependent.
Molecular Field Analysis	CoMFA and XED ^{24, 25}	QSAR, QSPR	Requirement for molecular alignment is a limitation and may also restrict use of such methods for database searching.

Table 2

Examples showing the importance of shape

Role of shape	Examples
Substrate recognition	<i>Pseudomonas fluorescens</i> N3 dioxygenase ¹⁰ , human cytochrome CYP51 ¹¹
Classification models	Protein-protein interaction inhibitor decision tree ¹² , Sammon and Kohonen mapping hERG models ¹³ , PEST descriptors to distinguish musk and non-musk compounds ⁵² , models of the 5-HT _{2B} receptor, the hERG ion channel, blood brain barrier penetration ^{31, 32} and PXR ⁹
QSAR models	ADME/Tox related datasets using TAE descriptors ⁵⁰
Virtual screening	ROCS used to discriminate between cruzain and cathepsin L inhibitors ⁴³ , Comparisons of different methods ^{37, 44-47} , Shape signatures used to find novel antiestrogens ⁸⁰ and enrich screening for serotonin ligands ⁵⁴
Co-evolution of molecule–protein interactions	NHRs (pregnane X receptor, farnesoid X receptor and liver X receptor) ^{15, 16, 18}
Ligand-receptor interactions	Dopamine D4 receptor ¹⁹
Immunoassay cross reactivity	Using 2D similarity searching and 3D pharmacophore searching for barbitures, benzodiazepines, and tricyclic antidepressants
Bioisosterism	Angiotensin-II analogs ⁵⁷
Masking molecular shape	Compartmentalization of reagents ⁷¹
Inducing protein disorder	Rifampicin binding to human PXR ⁷³
Molecule encryption	Extending methods using structural fragments ⁷⁴ and similarity based approaches ⁷⁵