JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

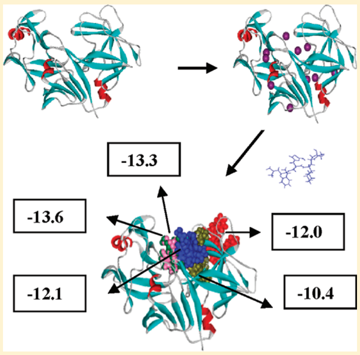
AADS - An Automated Active Site Identification, Docking, and Scoring Protocol for Protein Targets Based on Physicochemical Descriptors

Tanya Singh,[†,‡] D. Biswas,[‡] and B. Jayaram*[,†,‡,§]

[†]Department of Chemistry, [‡]Supercomputing Facility for Bioinformatics & Computational Biology, and [§]School of Biological Sciences, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India

S *Supporting Information*

ABSTRACT: We report here a robust automated active site detection, docking, and scoring (*AADS*) protocol for proteins with known structures. The active site finder identifies all cavities in a protein and scores them based on the physicochemical properties of functional groups lining the cavities in the protein. The accuracy realized on 620 proteins with sizes ranging from 100 to 600 amino acids with known drug active sites is 100% when the top ten cavity points are considered. These top ten cavity points identified are then submitted for an automated docking of an input ligand/candidate molecule. The docking protocol uses an all atom energy based Monte Carlo method. Eight low energy docked structures corresponding to different locations and orientations of the candidate molecule are stored at each cavity point giving 80 docked structures overall which are then ranked using an effective free energy function and top five structures are selected. The predicted structure and energetics of the complexes agree quite well with experiment when tested on a data set of 170 protein−ligand complexes with known structures and binding affinities. The *AADS* methodology is implemented on an 80 processor cluster and presented as a freely accessible, easy to use tool at http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp.



INTRODUCTION

The human genome project and developments in functional genomics are promising to present researchers with a number of clinically important targets. Attempts to generate three-dimensional structures of the target proteins are moving equally fast.[1,2] We have been focusing on providing freely accessible computational tools for developing reliable *in silico* suggestions of candidate molecules[3] against biomolecular targets (www.scfbio-iitd.res.in). Here, we introduce an automated version of active site (potential ligand binding site) detection, docking, and scoring methodology for any target protein.

Proteins contain binding sites which are used by the natural ligands/substrates and allosteric regulatory sites. An automated determination of the potential ligand binding site/active site, the binding pose of the candidate molecule, and its binding free energy are very demanding computationally but essential not only to understand molecular recognition events in the natural and diseased states but also to generate new leads for the target.

Active Site Identification. Many computational strategies have been developed in order to detect active sites in target proteins. POCKET[4] was one of the first grid based geometric methods to discover the active site. An extension to POCKET was made by LIGSITE,[5] which makes the algorithm less dependent on how the protein is oriented in the three-dimensional grid. Further extensions to LIGSITE were reported subsequently such as LIGSITE^{cs} in which the surface-solvent-surface events are monitored using the protein's Connolly surface[6] rather than the protein solvent protein events and LIGSITE^{csc}[7] in which the pockets identified by the surface-solvent-surface events are reranked by the degree of conservation of the surface residues involved which improved the ranking of the top ranked pocket from 67 to 75% when tested on a data set of 210 protein ligand complexes from PDB and from 69 to 79% for a test set of 48 bound protein−ligand complexes. PocketPicker[8] was another extension of LIGSITE which calculates the buriedness-index of grid points. SURFNET is another pocket identifier in which a sphere is placed between all pairs of protein atoms, so that the two atoms are on opposite sides on the surface of the sphere. If the sphere contains any other atoms, it is reduced in size until it contains no other atoms. Only spheres with a radius of 1−4 Å are kept. This results in a number of separate groups of interpenetrating spheres, both inside the protein and on its surface, which corresponds to its pocket sites. SURFNET was tested on a data set of 67 enzyme-ligand complexes.[9] The ligand was found to be bound in the largest cavity in 83% of the cases. The PHECOM[10] uses small and large spheres to define the pocket size and depth, respectively, and performs better than SURFNET, but the computational time taken was more. Another geometry based active site identifier is APROPOS[11] in which a family of shapes is generated and by comparing these shapes, ligand binding cavity is detected on the surface of the protein. The algorithm has been shown to have a success of 95% on a data set of proteins with one subunit, though the accuracy dropped when protein complexes

Received: April 30, 2011
Published: August 30, 2011

ACS Publications © 2011 American Chemical Society

were tested. CAST[12] uses an algorithm similar to APROPOS. The algorithm has been tested on a data set of 51 out of 67 SURFNET data set,[9] and an accuracy of 74% was achieved. CAST is also available as a Web server CASTp.[13] Fpocket[14] is based on alpha spheres and Voronoi tessellation. VOIDOO,[15] PASS,[16] LiGandFit,[17] methods of Delaney,[18] Del Carpio,[19] Cavity Search,[20] Kleywegt et al.,[21] Masuya and Doi,[22] Xie et al.,[23] Kim et al.,[24] Bock et al.,[25] and PocketDepth[26] are a few other attempts at active site identification. In all the above cases, the main criterion taken to identify the pocket in a protein is geometry. Huang and Schroeder compared the performance of CAST, LIGSITE, LIGSITE^{cs}, LIGSITE^{csc}, PASS, and SURFNET on a data set of 48 proteins with bound and unbound structures and 210 nonredundant proteins with bound structures using the same evaluation criteria.[5] Considering only the top three predictions, the results showed that the above geometric methods could achieve a success rate of 71−77% for the 48 unbound structures and 80−87% for the 210 bound structures. LIGSITE^{csc} achieved a success rate of 71 and 75% for the 48 unbound and 210 bound structures respectively in the prediction of topmost cavity point. SURFNET-Consurf[27] combines geometric method with sequence knowledge. MetaPocket metaserver[28] combines several of the geometry based methods by clustering and reranking the top three predicted pocket sites from LIGSITE^{cs}, PASS, Q-Site Finder,[29] and SURFNET. Meta Pocket improved the prediction accuracy from 70 to 75% for 210 bound structures.

Although the largest pocket tends to frequently correspond to the observed ligand binding site, this rule cannot be generalized, and different studies have tackled this problem.[6,27,30−32] There are some energy based methods like GRID,[33] Q site Finder,[29] desolvation based free energy models,[34] and CS-Map.[35] In GRID algorithm, the molecular target is superposed on a 3D grid, and interaction energy is then calculated between a molecular probe and the target. The interaction energy comprises Lennard-Jones, Columb, and directional hydrogen bond energy terms.[36−39] A similar algorithm is used to that described for Pocket Picker[8] in which grid points in close proximity to the protein surface are selected and buried-ness indices are calculated using 30-directional scanning around probes placed at each grid point. Another energy based method Pocketome[40] is quite similar to Q Site finder. In this method, small organic molecules are used as probes and the scoring potential includes a solvation term. Another energy based method finds consensus sites for organic solvent molecules by employing variable chemical potential grand canonical Monte Carlo simulations.[41,42] In these computationally intensive simulations, a relative chemical potential difference is set between a protein in a simulation cell and a bath of organic solvent. The high affinity sites on the protein for a particular small organic compound are identified at the end of the simulation. Also, there are methods to identify molecules at interfaces[43] and liquid interface in particular, from atomic coordinates.[44] Binding site detection algorithm which emphasizes the advantage of energy based methods[45] has also been reported. EasyMIFs generates molecular interaction field which is used as an input by SITEHound[46] to identify binding sites in a protein structure. A new method of circular variance to characterize molecular structures has also been reported.[47] ProBis[48] is a Web server which detects binding sites in a protein target based on local structural alignment. The Web server ProBis[49] compares a protein against a database of proteins and determines structurally similar regions by performing local structural alignment.

Desolvation based free energy method has also been reported[50] and so also a method based on the spatial distribution of hydrophobicity in a protein molecule, using a fuzzy-oil-drop model.[51] In the sticky spot method,[52] the protein surface is coated with a collection of molecular fragments that could bind with the protein. Each fragment serves as an alignment point for the atoms in a ligand and is scored to represent the probe's affinity for the protein. The probes are then clustered by accumulating their affinities. The high affinity clusters are identified as the stickiest portions of the protein surface. The stickiest portion identified is then used for docking the ligand to the protein.

Comparative modeling studies can also identify ligand binding sites since the sites are often highly conserved.[53−57] Cavbase describes and compares protein binding pockets on the basis of their geometrical and physiochemical properties.[58,59] Conserved residues in proteins have been used in the identification of ligand binding sites.[60] IsoCleft[61] is a graph-matching-based method which compares large sets of atoms obtained from native binding site and discriminates those proteins that bind similar ligands based on local 3D atomic similarities. A molecular interaction field approach is also used in mapping and selecting the active site.[62] Statistical analyses[63] of the protein−ligand contacts and neural networks[64,65] have also been harnessed to identify active sites in proteins. There are some surface property based approaches which have been used for predicting protein−protein interactions, including the use of support vector machines.[66] Last but not least, there is often an adaption of the pocket geometry to the formation of a complex with the ligand − the so-called induced fit − which is also considered.[66−70] Structure to function transition based on active site information has also been reported.[71]

If the lead discovery process has to be automated from a crystal or homology or *ab intio* built structure, a robust method is needed which can detect active sites with 100% accuracy. We address this issue in this contribution and propose a method and protocol with which this goal is realizable.

Docking and Scoring. One aims to discover ligands that will bind to the target protein with high affinity and specificity, once the drug active sites are identified, in computer aided drug design, before embarking on the time and cost intensive experimental work. Docking algorithms are used to predict how a ligand interacts with the binding site of a receptor. The docking algorithms are generally comprised of two methods: a search algorithm to generate all possible configurations of the ligand molecule in the active site of the protein and an efficient scoring function to evaluate how well the ligand interacts with the protein.[72,73] The first docking algorithm for small molecules was developed by Kuntz et al.[74] Subsequent application of the docking algorithm helped in identification of new leads against HIV-1 protease illustrating the potential of computer aided drug discovery.[75] There are over 60 docking programs and more than 30 scoring functions to date. Some of the known docking programs are listed in Supplementary Table 1.[76] Please[125] see refs 156−163 for a recent review of the docking and scoring strategies. While speed is essential for effective virtual high throughput screening of large libraries, accuracy is critical for lead optimization. The expectations are that the docking and scoring algorithms would generate structures which are nearly superposable on the cocrystal structures if available (essentially giving near zero root-mean-square deviation) with the computed binding free energies correlating well with experiment and thus helping in the design of lead molecules.[164,165]

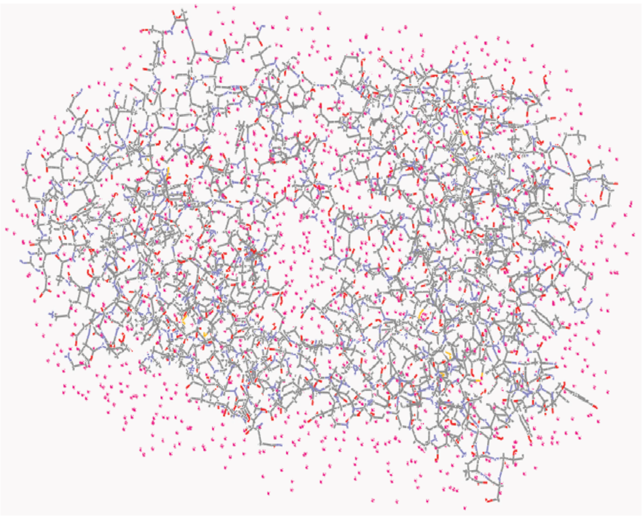


Figure 1. Average coordinates of grid points sorrounded by protein atoms from two side clustering in 4 Å sphere in protein with PDB ID 1A4K.

The above calls for a combination of a sturdy active site identifier, which works in conjunction with an efficient docking and scoring strategy in an automated mode. We present here one such protocol. Details of the methodology and results of validation on a large number of systems together with a brief description of the Web-utilities are presented below.

METHODS

Active Site Finder. The three-dimensional structure of a protein is taken in the Protein Data Bank format.[1] Hydrogen atoms are added to the protein through the tleap module of AMBER.[166] All the protein atoms are assigned their van der Waals radii, and the whole structure is mapped onto a three-dimensional grid with a resolution of 1 Å. The three-dimensional array so generated has two distinguishable parts: one part occupied by the protein atoms and the other part representing the unoccupied region. From the unoccupied regions, search in different directions is carried out to find out which vacant regions are bounded by the protein from at least two sides. The points which are within a distance of 4 Å from each other are clustered assuming them to be in a sphere of 4 Å and the average coordinates of all these points are stored. Figure 1 represents all these averaged points for the protein with PDB ID 1A4K. The number of points in each of these 4 Å clusters is also stored considering it to be an approximate volume of these clusters. Then to check which points lie close to the protein surface, two concentric spheres are generated around the above generated averaged points and the number of protein atoms above a cutoff limit (in our protocol we take it to be greater than 150) trapped inside the two spheres are counted. Cavity points satisfying the previous criteria are noted and finally clustered in a 10 Å radius and averaged to generate a reference point representing the cavity position in the protein along with the approximate volume of the cavity around that reference point. Here the volume contains a sum of all the points in the 4 Å clusters contained in the cavity of 10 Å. The final cavity points are shown in Figure 2 for protein with PDB ID 1A4K. Thus each pocket is represented by a single cavity point. We then sort these cavities in the descending order of their volumes. The amino acid residues which are lining the reference cavity points are noted. The number of hydrogen bond acceptors, donors, aromatic rings, and hydrophobic atoms among these residues around the cavity point are also counted. We store the maximum number of hydrogen bond donor atoms, hydrogen bond acceptor atoms, hydrophobic atoms, and aromatic rings among all the cavities identified. The values of these properties for the cavities generated in protein along with the Cartesian coordinates of the cavities shown in Figure 2 for the protein with PDB ID 1A4K are specified in Table 1. A score is then generated using the following formula

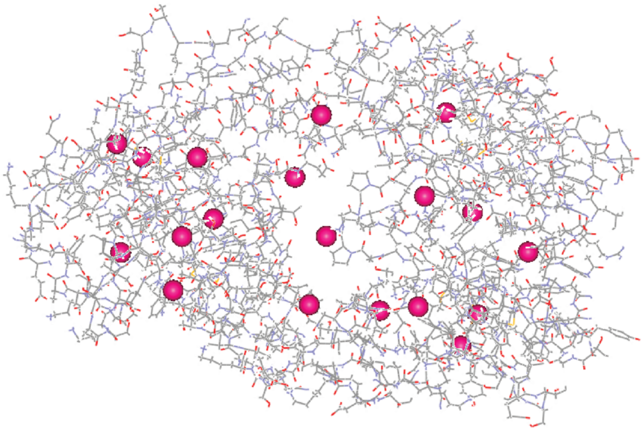


Figure 2. Cavity points generated by active site finder in protein with PDB ID 1A4K.

$$\mathrm{Score}_j = \left(\sum_{i=1..n} (X_{ij}/X_i^{\max}) \right)/n \qquad (1)$$

where Score_j is the fuzzy score that conveys the likeliness of the j^{th} cavity for a particular protein to be an actual ligand binding site; X_{ij} is any of the (n) properties considered such as approximate volume, number of hydrogen bond donors, number of hydrogen bond acceptors, or number of ring structures and number of hydrophobic groups in the j^{th} cavity of the protein, and $X_i^{\max}$ is the maximum value for the corresponding parameter in the protein. Here the value of n is equal to 5. The value of the score lies in the (0, 1) interval; 1.0 is the highest score for a cavity to be called, a ligand binding site. The cavity with the maximum volume and maximum number of hydrogen bond donors, acceptors, rings, and hydrophobic groups gets the maximum score. The above score is used for ranking the cavity points. The algorithm for an automated detection of the active site is shown in the form of a flowchart in Supplementary Figure 1. The greater the volume of a site the greater is the chance of a small molecule to bind there. However this is not always true as we find that the largest site is not always the ligand binding site. Similarly, the presence of a large number of hydrogen bond donors, hydrogen bond acceptors, aromatic rings, and hydrophobic groups in a cavity provides a greater opportunity for the small organic molecule to bind there. Keeping these issues in mind, the above fuzzy score function is developed to detect active sites in proteins. The methodology, is fast - taking less than a minute on a single processor - and, as the results indicate, is foolproof and does not involve any training which makes it applicable to diverse protein targets.

Docking. The top ten cavity points generated through the above algorithm act as reference points where the candidate drug

Table 1. Coordinates of Cavity Points Generated by Active Site Finder in Protein 1A4K, along with an Approximate Volume, Number of Hydrogen Bond Acceptors, Hydrogen Bond Donors, Hydrophobic Atoms, and Aromatic Rings Present in the Respective Cavity

cavity points	X coordinate	Y coordinate	Z coordinate	volume	Hbond acceptor	Hbond donor	ring structures	hydrophobic groups
Cavity 1	9.782	26.954	22.947	1354	11.00	14.00	8.00	15.00
Cavity 2	22.384	25.552	12.322	548	7.00	10.00	6.00	42.00
Cavity 3	21.539	12.537	5.081	521	3.00	14.00	9.00	31.00
Cavity 4	−7.026	46.097	33.571	425	9.00	12.00	4.00	39.00
Cavity 5	−5.911	37.936	42.633	353	7.00	13.00	5.00	35.00
Cavity 6	7.567	18.982	5.465	311	7.00	13.00	6.00	22.00
Cavity 7	6.347	36.601	37.502	294	16.00	18.00	4.00	16.00
Cavity 8	8.544	43.906	30.879	269	11.00	16.00	5.00	18.00
Cavity 9	7.913	39.551	14.810	217	10.00	13.00	3.00	17.00
Cavity 10	15.969	21.759	30.659	193	10.00	11.00	2.00	24.00
Cavity 11	4.732	31.523	41.787	172	15.00	16.00	1.00	11.00
Cavity 12	−1.335	34.023	20.030	156	9.00	11.00	1.00	6.00
Cavity 13	23.033	14.715	20.709	128	4.00	6.00	7.00	24.00
Cavity 14	9.743	15.588	22.408	124	9.00	5.00	2.00	11.00
Cavity 15	3.679	50.405	37.458	99	4.00	6.00	2.00	10.00
Cavity 16	11.599	6.501	16.653	88	4.00	3.00	2.00	10.00
Cavity 17	16.203	8.999	13.632	82	0.00	0.00	5.00	12.00
Cavity 18	11.541	31.496	2.221	66	0.00	0.00	3.00	13.00
Cavity 19	−3.264	35.224	31.067	51	2.00	2.00	2.00	9.00

molecule can be docked. The following steps are involved for docking the candidate drug molecules at the reference points:[155] (a) Preparation of the protein and the candidate drug molecule, (b) Translation of the drug molecule to the reference cavity points, (c) Grid Generation, (d) Generation of Monte Carlo configurations of the candidate drug molecule in the cavity points, and (e) Collection of eight low energy configurations for each reference cavity point. We describe each step in detail below.

(a). Preparation of the Protein and the Candidate Drug Molecule. The hydrogen added protein molecule is prepared in a force field compatible manner.[167] The ligand is considered in its input pose/conformation for the calculation. Hydrogens are added to the ligand molecule through the xleap module of AMBER,[166] maintaining the ionization state as reported in the literature (or as input by the user) which is then geometry optimized through the AM1 procedure followed by calculation of partial charges of the ligand by AM1-BCC procedure.[168] The GAFF force field parameter[169] is then used to assign atom types,[170] bond angle, dihedral, and van der Waals parameters for the ligand.

(b). Translation of the Candidate Molecule to the Reference Cavity Points. The center of mass of the above prepared candidate molecule is calculated and then translated to each of the top ten cavity points detected through the above active site finder algorithm. A cube with each side measuring 20 Å and centered on each of the 10 reference points is then created with a uniform grid size of 1 Å inside the cube. Only those grid points in each cube which are not occupied by the protein atoms (described in (c) below) are then considered for further calculations. This provides all possible accessible translation points surrounding any given cavity point.

(c). Grid Generation. A cubic grid of 1 Å resolution is pregenerated around the protein, and the grid points occupied by the protein side chains are stored. While searching for the spatial positions around the reference cavity points, the clash module helps in identifying the appropriate translation points on the 20 Å cubic grid. These translation points are the grid points in the cubic grid which are not occupied by protein atoms as explained in (b) above. The number of clashes is calculated in the cube at each translation point, and the best translation points with minimum number of clashes are selected. Interaction energy of each ligand atom with protein atoms within a specified cutoff distance are calculated using a scoring function comprising the electrostatic and van der Waals interactions and hydrophobic contributions as described by the equation below

$$E = \sum(E_{el} + E_{vdw} + E_{hpb}) \tag{2}$$

Here E is the sum over all ligand atoms. E_{el} is the electrostatic component of the energy, E_{vdw} is the van der Waals component between the protein and ligand atoms,[171−173] and E^{hpb} is the hydrophobic component.[174] This gives an approximate interaction energy.

(d). Generation of Monte Carlo Configurations of the Candidate Drug Molecule in the Cavity Points. At each cavity site, several (10^3) configurations are generated via a six-dimensional rigid body Monte Carlo methodology in the space of accessible grid points (translation points). This results in many ligand configurations which are scored based on the above-mentioned scoring function. These Monte Carlo runs are carried out concurrently at each of the 10 cavity points.

(e). Collection of Eight Low Energy Configurations for Each Cavity Point. About 8 low energy structures from each Monte Carlo run are collected giving a total of 80 candidate structures for the protein−ligand complex.

Scoring. The above selected 80 docked structures are energy minimized using the sander module of AMBER[166] and scored using a previously developed in-house scoring function christened *Bappl*[167] which embeds an effective free energy function. It is an all atom energy based empirical scoring function comprising electrostatics, van der Waals, hydrophobicity, and loss of conformational entropy of protein side chains upon ligand binding and is of the following form

$$\Delta G = \alpha(E_{el}) + \beta(E_{vdw}) + (\textstyle\sum_{A=1}^{22}\sigma_A \Delta A_{LSA}) + \lambda(\Delta S_{CR}) + \delta \qquad (3)$$

ΔG is the binding free energy (in kcal/mol), E_{el} is the electrostatic component, and E_{vdw} is the van der Waals component. ΔA_{LSA} is the loss in surface area of the atom type - *Bappl* defines 22 atom types[167] - σ_A is the atomic desolvation parameter in kcal/mol/Å^2 for an atom type, ΔS_{CR} is the loss in conformational entropy and δ is a constant. The electrostatic contribution is computed through Coulomb's law with a sigmoidal dielectric function. The van der Waals contribution is calculated using the (12, 6) Lennard-Jones potential between the protein and the ligand atoms.[167,175]

The above function has been validated[167] previously on a data set of 161 protein ligand complexes which yielded a correlation coefficient of 0.92 for the predicted binding free energies against the experimental binding affinities. The scoring function has been Web enabled at http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp. A comparative evaluation of the *Bappl* scoring function is given in Supplementary Table 2.

On the basis of the energy ranking by *Bappl*, the five best docked structures are reported one of which is expected to be native-like. A schematic representation of a parallel implementation of the docking/scoring methodology on an 80 processor cluster is shown in Supplementary Figure 2.

CALCULATIONS

We verified the accuracies of the Active Site Finder initially on 48 bound complexes used as a bench mark in the literature[6,8,14] for efficiency evaluation. The study was further extended to 572 additional protein—ligand complexes summing to a total of 620 complexes reported in the protein data bank. PDB IDs of the 620 protein—ligand complex data set are provided in Supplementary Table 3. In order to assess different methods on the same data set, a common criterion is needed for judging accuracies. The active site finder gives Cartesian coordinates of the geometric center of each of the ten cavity sites detected. Similarly, for other softwares, the center of mass of the site was taken into consideration. Thus for the above validation and to be consistent with the previous practices reported in the literature, a cavity point detected is considered to be a hit if the point is within 4 Å from any of the atoms of the ligand molecule in the crystal structure.

The top ten cavity points detected by the above algorithm are considered as the reference points for docking and scoring. A blind docking was performed with candidate molecules at each of the ten reference points. Eight best energy ranked structures are stored corresponding to each reference point generating a total of 80 docked structures for each candidate molecule. These docked structures are then scored on the basis of binding free energies using *Bappl* scoring function, and five structures are collected. This process is repeated with all the 170 protein—ligand complexes with known experimental structures and binding free energies. The experimental binding free energies for these complexes are available in the public domain databases like LPDB[194] and PLD.[195] For the purposes of assessing the docked structures vis-à-vis native, the ligand molecule and the amino acid residues surrounding it up to a distance of 6 Å in the crystal structure are superposed and a root-mean-square deviation (RMSD) is calculated. This is a more sensitive test than all atom RMSDs which generally tend to be small.

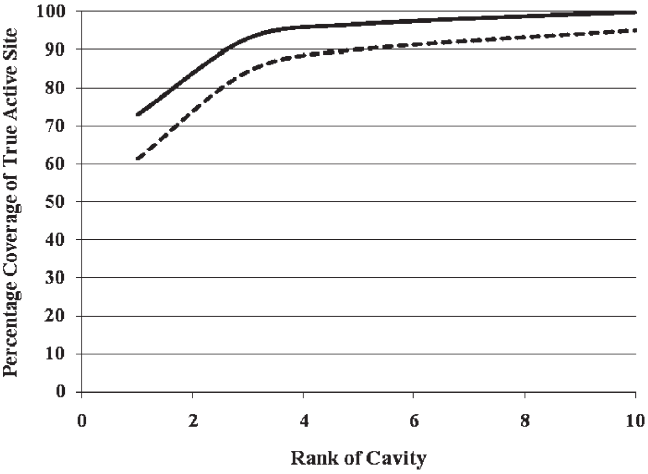


Figure 3. Rank of the cavity point versus cumulative percentage of true active site coverage. It is seen that true active site is captured with 90% and 95% accuracy by the top five and top ten cavity points with a distance constraint of 4 Å (dashed line) and with 98% and 100% accuracy with a distance constraint of 7 Å (bold continuous) respectively in 620 proteins (see Supplementary Table 3 for a list of proteins).

RESULTS AND DISCUSSION

Active Site Identification. The algorithm was tested on a data set of 620 protein—ligand complexes comprising different classes and sizes of proteins. The top five cavity points capture the true active site in 90% cases, and the top ten cavity points detect the true active site with 95% accuracy as shown in Figure 3 with a 4 Å criterion described under calculations. However, if the distance is increased to 7 Å, the prediction accuracies increase to 98% and 100% for the top five and top ten cavity points generated respectively as shown in Figure 3. Note that the focus here has been on capturing the active site 100% of the time and 7 Å distance is not a cause for concern if the docking is able to restore the ligand to its native like pose and location (as the results of docking described below indicate).

A comparison of the results was performed with different softwares available in public domain with the 48 protein—ligand complex data set (Supplementary Table 4) employed by others previously, and the prediction accuracies are shown in Table 2 on the lines of Huang and Schroder[7] and Weisel et al.[8] for their software Ligsite^{csc} and Pocket Picker respectively and Vincent Le Guilloux et al. for their software Fpocket.[14] The 48 protein ligand complex along with PDB IDs and the ranking of active site finder on the data is shown in Supplementary Table 4. The prediction accuracies of different softwares shown in Table 2 correspond to this 48 complex data set. The active site finder reported in the present work compares quite well in terms of detection accuracies (92% when top 3 points are considered). For the above comparison the distance constraint was 4 Å as mentioned. More interestingly, if one takes a distance constraint of 7 Å, then the active site finder protocol reported here captures the true active site with 100% accuracy in top ten points. Note that this accuracy is retained not just for 48 complexes but for all the 620 protein—ligand complexes studied (Figure 3). In short, the active site finder protocol reported here performs better than other softwares

Table 2. Prediction Accuracies (in %) of the Active Site Finder Shown along with Results from Different Softwares on 48 Bound Protein−Ligand Complexes[14]

Sl. no.	softwares	Top1	Top3[a]	Top5	Top 10[b]
1	SCFBIO(Active Site Finder)	67	92	98	100
2	Fpocket	83	92	-	-
3	PocketPicker	72	85	-	-
4	LiGSITE^{cs}	69	87	-	-
5	LIGSITE	69	87	-	-
6	CAST	67	83	-	-
7	PASS	63	81	-	-
8	SURFNET	54	78	-	-
9	LIGSITE^{csc}	79	-	-	-

[a] 92 under Top3 means active site within cut off distance of 4 Å is captured by one of the top3 cavity points detected in 92% of cases. [b] For Top 10 the distance cut off is 7 Å.

if one considers three or more (maximum being ten) predicted sites, our main aim being detecting the true active site with 100% accuracy for the purposes of automation. Even though the topmost cavity detected by active site finder returns 67% accuracy, the docking protocol as discussed below scores the ligand best at its true native site with the top ranked structure with 90% accuracy (see below). Fpocket gave a prediction accuracy of 83% for the topmost point and 92% for the top three detected points on the benchmark data set as clear from Table 2. However, the prediction accuracies of Fpocket on a data set[14] of 85 protein−ligand complexes was 67% and 82% for the Top 1 and Top 3 predicted sites, respectively, with a distance constraint of 4 Å. Fpocket has been trained on a data set of proteins with defined binding sites in order to determine parameters used in the program. However, there is no such training involved in the *AADS* methodology which gives an advantage of its transferability to diverse proteins. The program has been validated on a data set of 620 protein−ligand complexes containing different classes of proteins giving a prediction accuracy of 100% for the top ten points detected.

Docking and Scoring. The calculated RMSDs for 170 complexes with the topmost ranked structure are shown in Figure 4. The RMSDs are less than 2 Å in 90% of the cases (Figure 4). However when we considered the top five reported structures, at least one of them had an RMSD less than 2 Å in 95% of the cases (Figure 5). RMSD for the remaining 5% was within 4 Å as can be seen in Figure 5 for the top five structures reported. However, if an RMSD calculation is done considering only the main chain atoms, then they were within 2 Å for the remaining 5% as well signifying that the overall pose of the ligand was similar to the native ligand.

The calculated binding free energies for the topmost docked structure correlate well with experimental binding free energies (correlation coefficient ∼ 0.82) (Figure 6). Notice that this number is a little less than 0.92 achieved by *Bappl*. However, results in Figure 6 are from blind docking studies in an automated mode with just the tertiary structural information of the protein and no inputs of the binding site information, which is extremely encouraging.

In a real-world application of the method, a researcher would like prediction of a site nearest to the native active site which could be used for designing new molecules. The Monte Carlo docking algorithm described in section 2 of Methods brings the ligand molecule closer to the native active site and scores the candidate molecule, best at its true active site. However, in some cases the sites predicted by the above methodology can also be used to discover new allosteric sites.

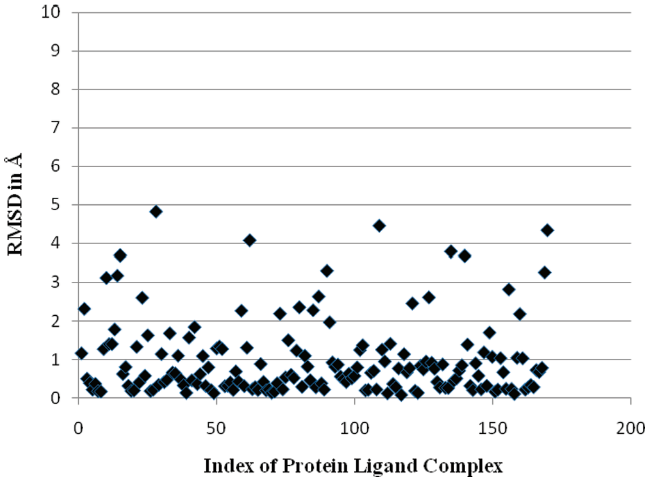


Figure 4. Root Mean Square Deviation between the crystal structure and the top ranked docked structure for the 170 protein−ligand complex data set.[167]

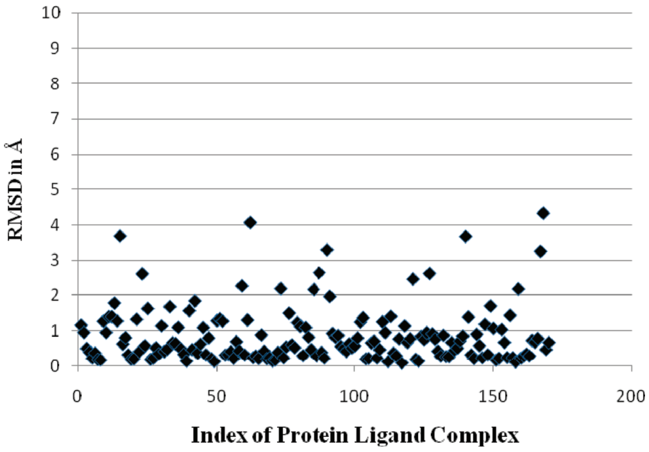


Figure 5. Root Mean Sqaure Deviation between the crystal structure and one of the top five docked structures for the 170 protein−ligand complex data set.[167]

The efficiency of the active site finder in combination with docking is shown in Figure 7 which reports the RMSDs before docking and after docking for the top docked structure in each of the 170 complexes studied. It could be seen that even if the cavity point was away up to 7 Å from the center of mass of the native ligand before docking, the RMSD for the top ranked docked structure lies within 2 Å. This gives us enough confidence in the possibility of automating the entire process of active site identification, docking, and scoring.

On the choice of the ligand to be docked for a protein target, we have recently developed a physicochemical descriptor based methodology (G. Mukherjee and B. Jayaram, manuscript in preparation; url: http://www.scfbio-iitd.res.in/software/drugdesign/raspd.jsp) and created an option to scan a million compound library rapidly without docking starting with the results of

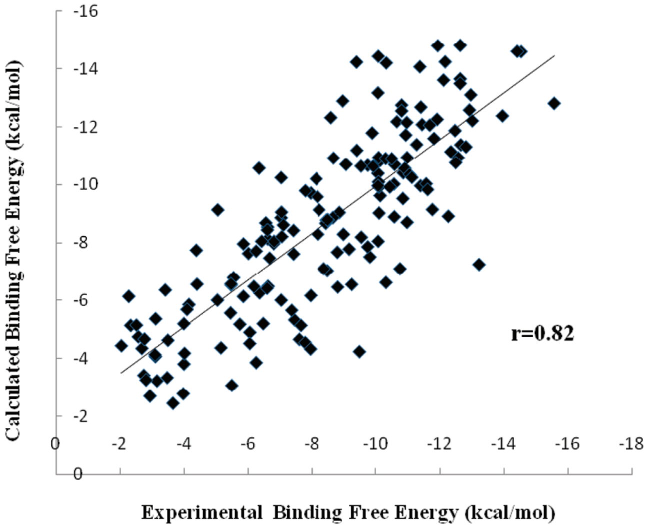


Figure 6. Correlation between experimental and predicted binding free energies of the top ranked docked structures in 170 protein−ligand complexes.

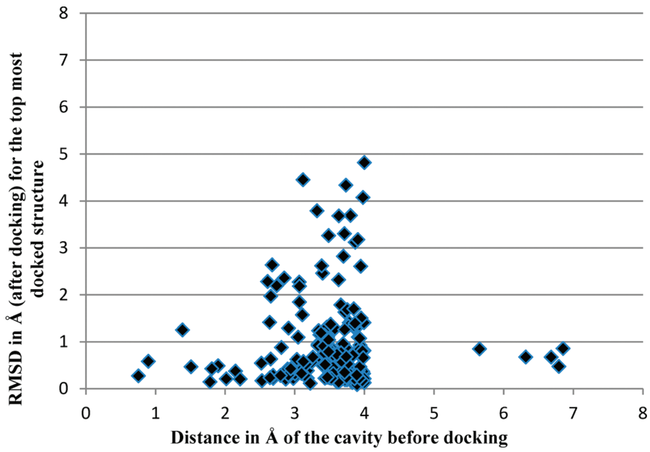


Figure 7. Distance of the cavity point from the center of mass of the native ligand before docking shown against the Root Mean Square Deviation between the native structure and the top ranked docked structure for the 170 protein−ligand complex data set.[167]

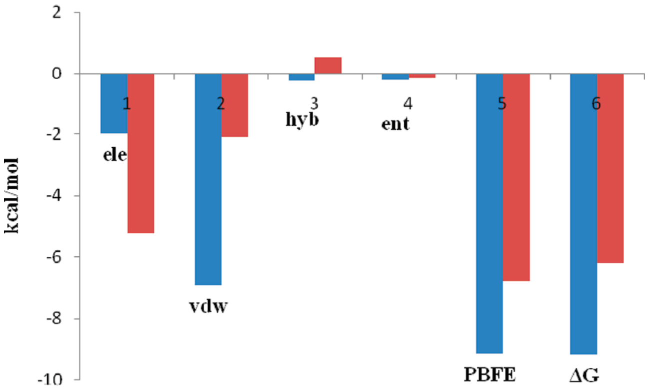


Figure 8. A consensus view of the relative magnitude of the energy components contributing to the binding free energy in 170 protein−ligand complexes. Abbreviations: ele: electrostatic component, vdw: van der Waals component, hyb: hydrophobic component, ent: rotational translational entropy, PFBE: predicted binding free energy, ΔG: experimental binding free energy. Note that the positive contributions are unfavorable and negative contributions are favorable. The targets for which van der Waals component dominates are shown in blue and those where electrostatics dominates are shown in red.

the active site finder, to sort out promising hits which can form inputs for the present docking and scoring protocol.

A component-wise analysis of the binding free energies of 170 complexes comprising 55 unique proteins targets reveals a few clear trends of use in ligand choice/design. Figure 8 depicts a consensus view of the diverse energy components contributing to the binding free energy in the 170 systems studied. For some of the targets, electrostatics was found to be more favorable than the van der Waals. The targets for which electrostatics was more favorable were grouped into one class (indicated in red), and ones for which van der Waals is dominant were grouped into another class (shown in blue). The results are somewhat similar to the trends seen with DNA binding proteins of different classes.[196,197] See in particular Figure 2 of ref 198. Such analyses which are provided together with the scoring function[199−201] can help in optimizing hit molecules.

To understand the sensitivity of the results to the input conformation of the ligand, the conformation of the small molecule was changed randomly around the rotatable bonds and about 1000 conformers were generated distinct from the native pose. The algorithm eliminates physically unrealistic conformers of the ligand, for example conformers with overlapping atoms. These conformers were minimized using the GB continuum solvent model for water using the AMBER force field.[166] At least one of the top ten conformers scored energy-wise was close to the bound conformer to within an RMSD of 2 Å. The ten conformers were submitted for docking and scoring through *Bappl* scoring function. The minimization step before scoring through *Bappl* protocol helps to take the ligand close to its local potential energy minimum. It may be recalled that docking generates 80 structures of the ligand corresponding to various cavities and orientations of the ligand in the protein which are put through minimization and binding free energy estimates. In our experience, a successful convergence on a solution can be obtained by generating conformer ensemble containing at least one conformer that is close to probably <1 Å RMSD to the actual solution. The bioactive conformer is not necessarily found at the global energy minimum in the energy landscape. Further fine-tuning of the conformer generator is in progress. We propose to integrate an option to generate low energy conformations of the small molecule at each of the 80 sites in an automated mode, followed by further minimization of the complex and binding free energy estimates and ranking, in the subsequent versions of *AADS*.

We envision that given a target protein and a database of small molecules, the *AADS* methodology will predict the binding sites in the protein with 100% accuracy, dock the molecules, and score them at all the ten detected potential binding sites and capture the experimental location and binding affinity in an automated mode. The methodology is robust enough to bracket potential lead molecules which can be further optimized to yield molecules with high affinity against the target protein.

The issues which are yet to be integrated into this automated protocol are (i) a rigorous consideration of the flexibility/dynamics of the ligand and active site residues of the target and (ii) explicit solvent and salt effects. We envisage that the five docked structures reported for each protein-candidate molecule complex could be put through molecular dynamics simulations and rigorous free energy calculations to develop structural, dynamic, and energetic perspectives on the protein-candidate molecule binding.[201−204]

Presented with (a) the sequence of amino acids of a target protein from genomic information, the computers would (b) generate three-dimensional structures of the protein in their appropriate oligomeric state, (c) identify the drug active sites, (d) narrow down the search space of ligands from a large library of synthesizable compounds/generate multiple conformations of each ligand molecule, (e) dock each ligand (potential candidate drug) and (f) score and rank candidate molecules, and (g) optimize the candidates and (g) transmit the best candidates to a medicinal chemist for synthesis and testing. The work reported here addresses steps (c), (e), and (f) in the above envisioned automated lead discovery pipe-line.

Brief Description of the Web-Utilities. (a) *AADS*: This version of the active site finder detects ten cavity points in a protein based on the physicochemical properties of the functional groups lining the cavities in the target protein. A rigid docking of the uploaded candidate drug molecule at the ten cavity points is performed in an automated mode. Five docked structures along with their binding free energy values in kcal/mol are e-mailed back to the user. The above program has been Web-enabled at the following link http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp.

(b) Another version of the active site finder was created in which the algorithm detects all the possible cavity points in a protein based on the volume of the respective cavity and also lists out the amino acid residues lining the cavity points. On the basis of the biochemical information from the literature, that is the amino acid residues involved in the biochemical activity of the protein, the user can select the cavity of interest and dock the candidate ligand molecule at that point. Four docked structures along with their binding free energies in kcal/mol are e-mailed back to the user. The above program has been Web-enabled at the following link http://www.scfbio-iitd.res.in/dock/ActiveSite.jsp.

CONCLUSIONS AND PERSPECTIVES

A robust automated active site identification protocol is formulated. The method supplements geometric information of the active sites with physicochemical properties of amino acid residues lining the active sites. The top ten cavity points identified capture the true active sites 100% of the time. All ten cavity points detected are used to dock and score ligands in an automated mode. The predicted structures and the energetics are in good accord with experiment. All the stages of the computational protocol presented, involve atomic level descriptions of the systems and no training thus assuring transferability and generality. The methodologies are configured in a high performance computing (HPC) environment, requiring 15 min on an 80 processor cluster for each protein−ligand complex. The computational methods are maturing to a point where automated structure based lead discovery is within the realm of feasibility in the near future.

ASSOCIATED CONTENT

S Supporting Information. Citations to some docking program reported in the literature; comparative evaluation of *Bappl* scoring function; PDB ids of protein data set to validate Active Site Identifier; comparison of different active site identification tools available in public domain; protocol for Active Site Identification; flowchart of the docking and scoring protocol. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +91 11 2659 1505, +91 11 2659 6786. Fax: +91 11 2658 2037. E-mail: bjayaram@chemistry.iitd.ac.in. Website: www.scfbio-iitd.res.in.

ACKNOWLEDGMENT

This work is supported by grants from the Department of Biotechnology, Govt. of India. The authors are thankful to Mr. Bharat Lakhani, Ms. Vandana, and Abhinav Mathur for Web-enabling the active site finder and the current version of *Sanjeevini* for automated docking and scoring.

REFERENCES

(1) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(2) Shenoy, S. R.; Jayaram, B. Proteins: sequence to structure and function- current Status. *Curr. Protein Pept. Sci.* **2010**, *11*, 498–514.

(3) Shaikh, S. A.; Jain, T.; Sandhu, G.; Latha, N.; Jayaram, B. From drug target to leads- sketching, A physicochemical pathway for lead molecule design in silico. *Curr. Pharm. Des.* **2007**, *13*, 3454–3470.

(4) Levitt, D.; Banaszak, L. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graphics* **1992**, *10*, 229–234.

(5) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graphics Modell.* **1997**, *15*, 359–363.

(6) Connolly, M. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983a**, *16*, 548–558.

(7) Huang, B.; Schroeder, M. LIGSITE^{csc}: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.* **2006**, *6*, 19.

(8) Weisel, M.; Proschak, E.; Schneider, G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem. Cent. J.* **2007**, *1*, 1–17.

(9) Laskowski, R.; Luscombe, N.; Swindells, M.; Thornton, J. Protein clefts in molecular recognition and function. *Protein Sci.* **1996**, *5*, 2438–2452.

(10) Kawabata, T.; Go, N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins* **2007**, *68*, 516–529.

(11) Peters, K. P.; Fauck, J.; Frommel, C. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J. Mol. Biol.* **1996**, *256*, 201–13.

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

(13) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **2006**, *34*, 116–118.

(14) Guilloux, L. V.; Schmidtke, P.; Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinf.* **2009**, *10*, 168.

(15) Kleywegt, G. J.; Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 178–185.

(16) Brady, G.; Stouten, P. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 383–401.

(17) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289–307.

(18) Delaney, J. S. Finding and filling protein cavities using cellular logic operations. *J. Mol. Graphics* **1992**, *10*, 174–177.

(19) Del Carpio, C. A.; Takahashi, Y.; Sasaki, S. A new approach to the automatic identification of candidates for ligand receptor sites in proteins: (I). Search for pocket regions. *J. Mol. Graphics* **1993**, *11*, 23–29.

(20) Ho, C. M.; Marshall, G. R. Cavity search: an algorithm for the isolation and display of cavity-like binding regions. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 337–354.

(21) Kleywegt, G. J. Recognition of spatial motifs in protein structures. *J. Mol. Biol.* **1999**, *285*, 1887–1897.

(22) Masuya, M.; Doi, J. Detection and geometric modeling of molecular surfaces and cavities using digital mathematical morphological operations. *J. Mol. Graphics* **1995**, *13*, 331–336.

(23) Xie, L.; Bourne, P. E. A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. *BMC Bioinf.* **2007**, *8*, 4–9.

(24) Kim, D.; Cho, C. H.; Cho, Y.; Ryu, J.; Bhak, J.; Kim, D. S. Pocket extraction on proteins via the Voronoi diagram of spheres. *J. Mol. Graphics Modell.* **2008**, *26*, 1104–1112.

(25) Bock, M. E.; Garutti, C.; Guerra, C. Effective labeling of molecular surface points for cavity detection and location of putative binding sites. *Comput. Syst. Bioinf. Conf.* **2007**, *6*, 263–274.

(26) Kalidas, Y.; Chandra, N. Pocket Depth: A new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.* **2008**, *161*, 31–42.

(27) Glaser, F.; Morris, R.; Najmanovich, R.; Laskowski, R.; Thornton, J. A method for localizing ligand binding pockets in protein structures. *Proteins* **2006**, *62*, 479–488.

(28) Huang, B. MetaPocket: a meta approach to improve protein ligand binding sites prediction. *Omics* **2009**, *13*, 325–330.

(29) Laurie, A.; Jackson, R. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **2005**, *21*, 1908–1916.

(30) Nayal, M.; Honig, B. On the nature of Cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins: Struct., Funct., Bioinf.* **2006**, *6*, 892–906.

(31) An, J.; Totrov, M.; Abagyan, R. Comprehensive identification of "druggable" protein ligand binding sites. *Genome Inform.* **2004**, *15*, 31–41.

(32) Zhong, S.; MacKerell, A. D. J. Binding response: a descriptor for selecting ligand binding site on protein surfaces. *J. Chem. Inf. Model.* **2007**, *47*, 2303–2315.

(33) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

(34) Coleman, R. G.; Sharp, K. A. Travel depth, a new shape descriptor for macromolecules: application to ligand binding. *J. Mol. Biol.* **2006**, *362*, 441–58.

(35) Landon, M. R.; Lancia, D. R., Jr.; Yu, J.; Thiel, S. C.; Vajda, S. Identification of hot spots within druggable binding regions by computational solvent mapping of proteins. *J. Med. Chem.* **2007**, *50*, 1231–1240.

(36) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083–1094.

(37) Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148–156.

(38) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140–147.

(39) Morita, M.; Nakamura, S.; Shimizu, K. Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures. *Proteins* **2008**, *73*, 468–479.

(40) An, J.; Totrov, M.; Abagyan, R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol. Cell. Proteomics* **2005**, *4*, 752–761.

(41) Guarnieri, F.; Mezei, M. Simulated annealing of chemical potential: a general procedure for locating bound waters. application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J. Am. Chem. Soc.* **1996**, *118*, 8493–8494.

(42) Clark, M.; Guarnieri, F.; Shkurko, I.; Wiseman, J. Grand canonical Monte Carlo simulation of ligand-protein binding. *J. Chem. Inf. Model.* **2006**, *46*, 231–242.

(43) Partay, L. B.; Hantal, G.; Jedlovszky, P.; Vincze, A.; Horvai, G. A new method for determining the interfacial molecules and characterizing the surface roughness in computer simulations. Application to the liquid–vapor interface of water. *J. Comput. Chem.* **2008**, *29*, 945–956.

(44) Williard, A. P.; Chandler, D. Instantaneous Liquid interfaces. *J. Phys. Chem. B* **2010**, *114*, 1954–1958.

(45) Ghersi, D.; Sanchez, R. Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures. *J. Struct. Funct. Genomics* **2011**, *12*, 109–117.

(46) Ghersi, D.; Sanchez, R. EasyMIFS and SiteHound: a toolkit for the identification of ligand-binding sites in protein structures. *Bioinformatics* **2009**, *25*, 3185–3186.

(47) Mezei, M. A new method for mapping macromolecular topography. *J. Mol. Graphics Modell.* **2003**, *21*, 463–472.

(48) Janez, K.; Dusanka, J. ProBiS: A web server for detection of structurally similar protein binding sites. *Nucleic Acids Res.* **2010**, *38*, 436–440.

(49) Janez, K.; Dusanka, J. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26*, 1160–1168.

(50) Coleman, R. G.; Salzberg, A. C.; Cheng, A. C. Structure-based identification of small molecule binding sites using a free energy model. *J. Chem. Inf. Model.* **2006**, *46*, 2631–2637.

(51) Brylinski, M.; Prymula, K.; Jurkowski, W.; Kochanczyk, M.; Stawowczyk, E.; Konieczny, L.; Roterman, I. Prediction of functional sites based on the Fuzzy Oil Drop model. *PLoS Comput. Biol.* **2007**, *3*, 94.

(52) Ruppert, J.; Welch, W.; Jain, A. N. Automatic identification and representation of protein binding sites for molecular docking. *Protein Sci.* **1997**, *6*, 524–533.

(53) Pupko, T.; Re, R. B.; Mayrose, I.; Glaser, F.; Ben, T. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* **2002**, *18*, 71–77.

(54) de Rinaldis, M.; Ausiello, G.; Cesareni, G.; Helmer-Citterich, M. Three-dimensional profiles: a new tool to identify protein surface similarities. *J. Mol. Biol.* **1998**, *284*, 1211–1221.

(55) Armon, A.; Graur, D.; Ben-Tal, N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.* **2001**, *307*, 447–63.

(56) Glaser, F.; Pupko, T.; Paz, I.; Bell, R. E.; Bechor-Shental, D.; Martz, E.; Ben-Tal, N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **2003**, *19*, 163–164.

(57) Brylinski, M.; Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci.* **2008**, *105*, 129–134.

(58) Kuhn, D.; Weskamp, N.; Schmitt, S.; Hullermeier, E.; Klebe, G. From the similarity analysis of protein cavities to the functional classification of protein families using cavbase. *J. Mol. Biol.* **2006**, *359*, 1023–1044.

(59) Kuhn, D.; Weskamp, N.; Hullermeier, E.; Klebe, G. Functional classification of protein kinase binding sites using Cavbase. *ChemMedChem* **2007**, *2*, 1432–1447.

(60) Guharoy, M.; Chakrabarti, P. Conserved residue clusters at protein-protein interfaces and their use in binding site identification. *BMC Bioinf.* **2010**, *11*, 286.

(61) Najmanovich, R.; Kurbatova, N.; Thornton, J. Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics* **2008**, *24*, 105–111.

(62) Kumar, A.; Ghosh, I. Mapping Selectivity and Specificity of Active Site of Plasmepsins from Plasmodium falciparum Using Molecular Interaction Field Approach. *Protein Pept. Lett.* **2007**, *14*, 569–574.

(63) Taroni, C.; Jones, S.; Thornton, J. M. Analysis and prediction of carbohydrate binding sites. *Protein Eng.* **2000**, *13*, 89–98.

(64) Gutteridge, A.; Bartlett, G. J.; Thornton, J. M. Using a neural network and spatial clustering to predict the location of active sites in enzymes. *J. Mol. Biol.* **2003**, *330*, 719–734.

(65) Stahl, M.; Taroni, C.; Schneider, G. Mapping of protein surface cavities and prediction of enzyme class by a self-organizing neural network. *Protein Eng.* **2000**, *13*, 83–88.

(66) Bradford, J. R.; Westhead, D. R. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics* **2005**, *21*, 1487–1494.

(67) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895–2907.

(68) Bhinge, A.; Chakrabarti, P.; Uthanumallian, K.; Bajaj, K.; Chakraborty, K.; Varadarajan, R. Accurate detection of protein:ligand binding sites using molecular dynamics simulations. *Structure* **2004**, *12*, 1989–1999.

(69) Yang, A. Y.; Källblad, P.; Mancera, R. L. Molecular modelling prediction of ligand binding site flexibility. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 235–250.

(70) Murga, L. F.; Ondrechen, M. J.; Ringe, D. Prediction of interaction sites from apo 3D structures when the holo conformation is different. *Proteins* **2008**, *72*, 980–992.

(71) Ondrechen, M. J.; Clifton, J. G.; Ringe, D. THEMATICS: A simple computational predictor of enzyme function from structure. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12473–12478.

(72) Irwin, J. J.; Shoichet, B. K.; Mysinger, M. M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated Docking Screens: A Feasibility Study. *J. Med. Chem.* **2009**, *52*, 5712–5720.

(73) DesJarlais, R. L.; Cummings, M. D.; Gibbs, A. C. Virtual docking: how are we doing and how can we improve? *Front. Drug Des. Discovery* **2007**, *3*, 81–103.

(74) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. J. *Mol. Biol.* **1982**, *161*, 269–288.

(75) Desjarlais, R. L.; Seibel, G. L.; Kuntz, D.; Furth, P. S.; Alvarez, J. C.; Ortiz De Montellano, P. R.; Decamp, D. L.; Babe, L. M.; Craik, C. S. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.

(76) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **2008**, *153*, 7–26.

(77) Yamada, M.; Itai, A. Development of an efficient automated docking method. *Chem. Pharm. Bull.* **1993**, *41*, 1200–1202.

(78) Mizutani, M. Y.; Tomioka, N.; Itai, A. Rational automatic search method for stable docking models of protein and ligand. *J. Mol. Biol.* **1994**, *243*, 310–326.

(79) Mizutani, M. Y.; Takamatsu, Y.; Ichinose, T.; Nakamura, K.; Itai, A. Effective handling of induced-fit motion in flexible docking. *Proteins: Struct., Funct., Genet.* **2006**, *63*, 878–891.

(80) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(81) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in autodock. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 34–40.

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

(83) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. Detailed analysis of grid-based molecular docking: a case study of CDOCKER—a CHARMm-based MD docking algorithm. *J. Comput. Chem.* **2003a**, *24*, 1549–1562.

(84) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III Assessing energy functions for flexible docking. *J. Comput. Chem.* **1998b**, *19*, 1612–1622.

(85) Lawrence, M. C.; Davis, P. C. CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins: Struct., Funct., Genet.* **1992**, *12*, 31–41.

(86) Taylor, J. S.; Burnett, R. M. DARWIN: a program for docking flexible molecules. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 173–191.

(87) Clark, K. P.; Jain, A. N. Flexible ligand docking without parameter adjustment across four ligand−receptor complexes. *J. Comput. Chem.* **1995**, *16*, 1210–1226.

(88) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. Flexible ligand docking using a genetic algorithm. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 113–130.

(89) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424–440.

(90) Kang, X.; Shafer, R. H.; Kuntz, I. D. Calculation of ligand-nucleic acid binding free energies with the generalized-born model in DOCK. *Biopolymers* **2004**, *73*, 192–204.

(91) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput.- Aided Mol. Des.* **2006**, *20*, 601–619.

(92) Irwin, J. J.; Shoichet, B. K.; Mysinger, M. M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated Docking Screens: A Feasibility Study. *J. Med. Chem.* **2009**, *52*, 5712–5720.

(93) Hart, T. N.; Read, R. J. A multiple-start Monte Carlo docking method. *Proteins: Struct., Funct., Genet.* **1992**, *13*, 206–222.

(94) Vieth, M.; Cummins, D. J. DoMCoSAR: a novel approach for establishing the docking mode that is consistent with the structure—activity relationship. Application to HIV-1 protease inhibitors and VEGF receptor tyrosine kinase inhibitors. *J. Med. Chem.* **2000**, *43*, 3020–3032.

(95) Schafferhans, A.; Klebe, G. Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* **2001**, *307*, 407–427.

(96) Grosdidier, A.; Zoete, V.; Michielin, O. EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins: Struct., Funct., Genet.* **2007**, *67*, 1010–1025.

(97) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Johnson, A. P. eHiTS: an innovative approach to the docking and scoring function problems. *Curr. Protein Pept. Sci.* **2006**, *7*, 421–435.

(98) Pang, Y. P.; Perola, E.; Xu, R.; Prendergast, F. G. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* **2001**, *22*, 1750–1771.

(99) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J. Comput. Chem.* **2003**, *24*, 1637–1656.

(100) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 88–105.

(101) Budin, N.; Majeux, N.; Caflisch, A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.* **2001**, *382*, 1365–1372.

(102) Kolb, P.; Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.* **2006**, *49*, 7384–7392.

(103) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47*, 435–449.

(104) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

(105) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein−ligand docking predictions. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 17–28.

(106) Clausen, H.; Buning, C.; Rarey, M.; Lengauer, T. FLEXE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–395.

(107) Zhao, Y.; Sanner, M. F. FLIPDock: docking flexible ligands into flexible receptors. *Proteins: Struct., Funct., Bioinf.* **2007**, *68*, 726–737.

(108) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J. Comput.- Aided Mol. Des.* **1994**, *8*, 153–174.

(109) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.

(110) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997**, *272*, 106–120.

(111) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.

(112) Li, H.; Li, C.; Gui, C.; Luo, X.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. GAsDock: a new approach for rapid flexible docking based on an improved multi-population genetic algorithm. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4671–4676.

(113) Yang, J. M.; Chen, C. C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 288–304.

(114) Tietze, S.; Apostolakis, J. GlamDock: development and validation of a new docking tool on several thousand protein−ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 1657–1672.

(115) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(116) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553.

(117) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609–623.

(118) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor., R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.

(119) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449–462.

(120) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: a protein−protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.

(121) Floriano, W. B.; Vaidehi, N.; Zamanakos, G.; Goddard, W. A., III HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J. Med. Chem.* **2004**, *47*, 56–71.

(122) Trabanino, R. J.; Hall, S. E.; Vaidehi, N.; Floriano, W. B.; Kam, V. W. T.; Goddard, W. A., III First principles predictions of the structure and function of G-protein-coupled receptors: validation for bovine rhodopsin. *Biophys. J.* **2004**, *86*, 1904–1921.

(123) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994b**, *15*, 488–506.

(124) Totrov, M.; Abagyan, R. Flexible protein−ligand docking by global energy optimization in internal coordinates. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 215–220.

(125) Diller, D. J.; Merz, K. M., Jr. High throughput docking for library design and library prioritization. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 113–124.

(126) Wu, S. Y.; McNae, I.; Kontopidis, G.; McClue, S. J.; McInnes, C.; Stewart, K. J.; Wang, S.; Zheleva, D. I.; Marriage, H.; Lane, D. P.; Taylor, P.; Fischer, P. M.; Walkinshaw, M. D. Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure* **2003**, *11*, 399–410.

(127) Sobolev, V.; Wade, R. C.; Vriend, G.; Edelman, M. Molecular docking using surface complementarity. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 120–129.

(128) Fradera, X.; Kaur, J.; Mestres, J. Unsupervised guided docking of covalently bound ligands. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 635–650.

(129) Liu, M.; Wang, S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput.- Aided Mol. Des.* **1999**, *13*, 435–451.

(130) Thomsen, R.; Christensen, M. H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49*, 3315–3321.

(131) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, 363–367.

(132) Tøndel, K.; Anderssen, E.; Drabløs, F. Protein Alpha Shape (PAS) Dock: A new gaussian-based score function suitable for docking in homology modelled protein structures. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 131–144.

(133) Joseph-McCarthy, D.; Thomas, B. E., IV; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 172–188.

(134) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: pharmacophorebased protein−ligand docking. *J. Med. Chem.* **2004**, *47*, 6804–6811.

(135) Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: an FFTbased protein docking program with pairwise potentials. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 392–406.

(136) Korb, O.; Stutzle, T.; Exner, T. E. PLANTS: application of ant colony optimization to structure-based drug design. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Brussels, 2006; pp 247−258.

(137) Trosset, J. Y.; Scheraga, H. A. PRODOCK: software package for protein modeling and docking. *J. Comput. Chem.* **1999**, *20*, 412–427.

(138) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547–562.

(139) Seifert, M. H. J. ProPose: steered virtual screening by simultaneous protein−ligand docking and ligand−ligand alignment. *J. Chem. Inf. Model.* **2005**, *45*, 449–460.

(140) Pei, J.; Wang, Q.; Liu, Z.; Li, Q.; Yang, K.; Lai, L. PSI-DOCK: towards highly efficient and accurate flexible ligand docking. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 934–946.

(141) Jackson, R. M. Q-fit: a probabilistic method for docking molecular fragments by sampling low energy conformational space. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 43–57.

(142) McMartin, C.; Bohacek, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333–344.

(143) Morley, S. D.; Afshar, M. Validation of an empirical RNA−ligand scoring function for fast flexible docking using RiboDocks. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 189–208.

(144) Meiler, J.; Baker, D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 538–548.

(145) Burkhard, P.; Taylor, P.; Walkinshaw, M. D. An example of a protein ligand found by database mining: description of the docking method and its verification by a 2.3A° X-ray structure of a thrombin−ligand complex. *J. Mol. Biol.* **1998**, *277*, 449–466.

(146) Wu, G.; Vieth, M. SDOCKER: a method utilizing existing X-ray structures to improve docking accuracy. *J. Med. Chem.* **2004**, *47*, 3142–3148.

(147) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discovery Des.* **2000**, *20*, 171–190.

(148) Zavodszky, M. I.; Kuhn, L. A. Side-chain flexibility in protein—ligand binding: the minimal rotation hypothesis. *Protein Sci.* **2005**, *14*, 1104–1114.

(149) Alberts, I. L.; Todorov, N. P.; Dean, P. M. Receptor flexibility in de novo ligand design and docking. *J. Med. Chem.* **2005**, *48*, 6585–6596.

(150) Chen, H. M.; Liu, B. F.; Huang, H. L.; Hwang, S. F.; Ho, S. Y. SODOCK: Swarm optimization for highly flexible protein—ligand docking. *J. Comput. Chem.* **2007**, *28*, 612–623.

(151) Fradera, X.; Knegtel, R. M. A.; Mestres, J. Similarity-driven flexible ligand docking. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 623–636.

(152) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.

(153) Jain, A. N. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306.

(154) Choi, V. YUCCA: an efficient algorithm for small-molecule docking. *Chem. Biodiversity* **2005**, *2*, 1517–1524.

(155) Gupta, A.; Gandhimathi, P.; Sharma, P.; Jayaram, B. ParDOCK: An All Atom Energy Based Monte Carlo Docking Protocol for Protein-Ligand Complexes. *Protein Pept. Lett.* **2007**, *14*, 632–46.

(156) Sauton, N.; Lagorce, D.; Villoutreix, B. O.; Iteva, M. A. MS-DOCK: Accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinf.* **2008**, *9*, 184.

(157) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006**, *65*, 15–26.

(158) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Peter Johnson, A. eHiTS: An Innovative Approach to the Docking and Scoring Function Problems. *Curr. Protein Pept. Sci.* **2006**, *7*, 000–000.

(159) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.

(160) Coupez, B.; Lewis, R. A. Docking and scoring--theoretically easy, practically impossible? *Curr. Med. Chem.* **2006**, *13*, 2995–3003.

(161) Kroemer, R. T. Structure-Based Drug Design: Docking and Scoring. *Curr. Protein Pept. Sci.* **2007**, *8*, 312–328.

(162) Rester, U. Dock around the Clock — Current Status of Small Molecule Docking and Scoring. *QSAR Comb. Sci.* **2006**, *25*, 605–615.

(163) Lee1, H. S.; Choi, J.; Yoon, S. Evaluation of Advanced Structure-Based Virtual Screening Methods for Computer-Aided Drug Discovery. *Genom. Informatics* **2007**, *5*, 24–29.

(164) Sundriyal, S.; Khanna, S.; Saha, R.; Bharatam, P. V. Metformin and glitazones: does similarity in biomolecular mechanism originate from tautomerism in these drugs? *J. Phys. Org. Chem.* **2008**, *21*, 30–33.

(165) Langley, D. R.; Walsh, A. W.; Baldick, C. J.; Eggers, B. J.; Rose, R. E.; Levine, S. M.; Kapur, A. J.; Colonno, R. J.; Tenney, D. J. Inhibition of hepatitis B virus polymerase by entecavir. *J. Virol.* **2007**, *81*, 3992–4001.

(166) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheathem, J. E., III; et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

(167) Jain, T.; Jayaram, B. An all atom energy based computational protocol for predicting binding affinities of protein—ligand complexes. *FEBS Lett.* **2005**, *579*, 6659–6666.

(168) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(169) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(170) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(171) Narang, P.; Bhushan, K.; Bose, S.; Jayaram, B. A computational pathway for bracketing native-like structures for small alpha helical globular proteins. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2364–2375.

(172) Arora, N.; Jayaram, B. Energetics of base pairs in B-DNA in solution: An appraisal of potential functions and dielectric treatments. *J. Phys. Chem. B* **1998**, *102*, 6139–6144.

(173) Arora, N.; Jayaram, B. Strength of hydrogen bonds in alpha helices. *J. Comput. Chem.* **1997**, *18*, 1245–1252.

(174) Jayaram, B.; Beveridge, D. L. Grand canonical monte carlo simulation studies on aqueous solution of NaCl ana NaDNA: Excess chemical potentials and source of nonideally in electrolyte and polyelectrolyte solutions. *J. Phys. Chem.* **1991**, *95*, 2506–2516.

(175) Jain, T.; Jayaram, B. Computational protocol for predicting the binding affinities of Zinc containing metalloprotein-ligand complexes. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 1167–1178.

(176) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule database. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.

(177) Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: A computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.* **2001**, *22*, 1750–1771.

(178) Momany, F. A.; Rone, R. Validation of a general purpose QUANTA_3.2/CHARMm_ force field. *J. Comput. Chem.* **1992**, *13*, 888–900.

(179) Morris, G. M.; et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(180) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein—ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.

(181) DeWitte, R. S.; Shakhnovich, E. I. SMoG: de novo design method based on simple, fast and accurate free energy estimates. Methodology and supporting evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733–11744.

(182) Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. BLEEP: potential of mean force describing protein—ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comput. Chem.* **1999**, *20*, 1177–1185.

(183) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein—ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.

(184) Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledgebased energy function for protein—ligand, protein—protein and protein—DNA complexes. *J. Med. Chem.* **2005**, *48*, 2325–2335.

(185) Wang, R.; Liu, L.; Lai, L.; Tang, Y. SCORE: A new empirical method for estimating the binding affinity of a protein— ligand complex. *J. Mol. Model.* **1998**, *4*, 379–394.

(186) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(187) Bohm, H. J. Prediction of binding constants of protein—ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309–323.

(188) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

(189) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

(190) Head, R. D.; et al. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959–3969.

(191) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: A novel scoring function for predicting binding affinities. *J. Mol. Graphics Modell.* **2005**, *23*, 395–407.

(192) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.

(193) Friesner, R. A.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(194) Roche, O.; Kiyama, R.; Brooks, C. L., III Ligand—protein database: linking protein—ligand complex structures to binding data. *J. Med. Chem.* **2001**, *44*, 3592–3598.

(195) Puvanendrampillai, D.; Mitchell, J. B. O. Protein ligand database (PLD): additional understanding of the nature and specificity of protein—ligand complexes. *Bioinformatics* **2003**, *19*, 1856–1857.

(196) Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Beveridge, D. L. Free Energy Analysis of Protein-DNA Binding: The EcoRI Endonuclease - DNA Complex. *J. Comput. Phys.* **1999**, *151*, 333–357.

(197) Jayaram, B.; McConnell, K.; Dixit, S. B.; Das, A.; Beveridge, D. L. Free-Energy component analysis of 40 protein-DNA complexes: A consensus view on the thermodynamics of binding at the molecular level. *J. Comput. Chem.* **2002**, *23*, 1–14.

(198) Jayaram, B.; Jain, T. The role of water in Protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 343–61.

(199) Shaikh, S. A. B.; Jayaram, A Swift all-atom energy based computational protocol to predict DNA ligand binding affinity and ΔTm. *J. Med. Chem.* **2007**, *50*, 2240–2244.

(200) Shaikh, S. A.; Ahmed, S. R.; Jayaram, B. A molecular thermodynamic view of DNA-drug interaction: A case study of 25 minor groove binders. *Arch. Biochem. Biophys.* **2004**, *429*, 81–99.

(201) Kalra, P.; Reddy, T. V.; Jayaram, B. Free energy component analysis for drug design: A case study of HIV-1 protease-inhibitor binding. *J. Med. Chem.* **2001**, *44*, 4325–4338.

(202) Reddy, M. R.; Erion, M. D. Calculation of relative binding free energy differences for fructose 1,6-bisphosphatase inhibitors using the thermodynamic cycle perturbation approach. *J. Am. Chem. Soc.* **2001**, *123*, 6246–6252.

(203) Gilson, M. K.; Given, A. J.; Bush, B. L.; Mc Cammon, J. A. The statistical thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047–1069.

(204) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181–9183.