Molecular docking in organic, inorganic, and hybrid systems: a tutorial review

Madhuchhanda Mohanty[1] · Priti S. Mohanty[1,2]

Abstract

Molecular docking simulation is a very popular and well-established computational approach and has been extensively used to understand molecular interactions between a natural organic molecule (ideally taken as a receptor) such as an enzyme, protein, DNA, RNA and a natural or synthetic organic/inorganic molecule (considered as a ligand). But the implementation of docking ideas to synthetic organic, inorganic, or hybrid systems is very limited with respect to their use as a receptor despite their huge popularity in different experimental systems. In this context, molecular docking can be an efficient computational tool for understanding the role of intermolecular interactions in hybrid systems that can help in designing materials on mesoscale for different applications. The current review focuses on the implementation of the docking method in organic, inorganic, and hybrid systems along with examples from different case studies. We describe different resources, including databases and tools required in the docking study and applications. The concept of docking techniques, types of docking models, and the role of different intermolecular interactions involved in the docking process to understand the binding mechanisms are explained. Finally, the challenges and limitations of dockings are also discussed in this review.

Graphical abstract
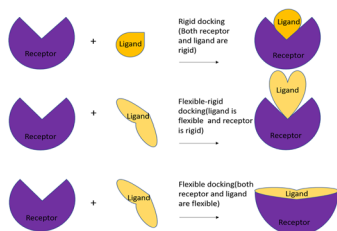
Keywords Density functional calculations · Donor–acceptor effects · Nanostructures · Natural products · Proteins

Introduction

Docking is a computational modelling technique. It predicts the best energy-minimized pose of one molecule to a second molecule when bound to each other to form a stable complex in a virtual way to complement the experimental method [1–4]. It is rapid and cost-effective compared to the trial-and-error methods using experimental studies. It can replace or provoke experiments. It also helps to explain and understand experiments and creates a map between the theoretical and practical aspects of molecular interaction. Sometimes using experimental methods, like solution NMR, X-ray diffraction, electron microscopy, solution scattering, and neutron diffraction, it may not be possible to obtain the structure of a stable complex of two different molecules docked together because of cost, research timeline limitation, blind and too many parameters to control, difficulty, impossible or dangerous to perform experiments, and unavailability of

✉ Madhuchhanda Mohanty
madhuchhanda.mohanty@kiitbiotech.ac.in

1   School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT), Deemed to be University, Bhubaneswar 751024, India

2   School of Chemical Technology, Kalinga Institute of Industrial Technology (KIIT), Deemed to be University, Bhubaneswar 751024, India

material, etc. So, using such an in-silico mechanism based on structural information it is possible to attempt to find the best match between two molecules to predict their intermolecular attraction or repulsion. Knowledge of preferred conformation might speculate the strength of association (or binding affinity) between two molecules and result in different possible complex structures that are ranked and grouped using scoring functions. The molecular docking study evaluates the binding mode of a molecule to the binding pocket of another molecule based on the minimum binding energy.

The binding of a small molecule or atom or ion (called a ligand) to a biomolecule (termed a macromolecule, receptor, or target) is a common computational study in numerical noncovalent interaction processes. Accurate and reliable prediction of docking results between protein and ligand is useful in modern structure-based drug design [5]. Interaction between biologically related macromolecules, namely DNA, RNA, protein, carbohydrate, lipid, enzyme, and small organic/inorganic molecules or atoms or ions play a central role in biological processes, including signal transduction, transport, cell regulation, gene expression control, enzyme inhibition, antibody-antigen recognition, and the assembly of multi-domain proteins. They interact with each other to perform a certain biological function together inside a biological cell. Different relative orientations at specific sites of interactive biological entities may alter their functions. So, a docking study is essential for predicting the pose of biomolecules to dock with each other more precisely. The interaction can occur between any two molecules, like DNA-ligand, RNA-ligand protein–protein, protein-RNA, protein-DNA, protein-drug, protein-nanoparticle, enzyme–substrate, enzyme-drug, etc.

Applications of molecular docking

There are numerous applications of molecular docking technique, namely prediction of lowest free binding energy-optimized structure of a receptor-ligand complex, differential binding of a ligand to different macromolecular receptors, geometry of a receptor-ligand stable complex, propose modification of a lead molecule to increase its efficiency for binding to a receptor and to act together performing a task, de novo design of candidate molecules based on docking studies and their library design, study on side effects of a candidate molecule because of its interaction with other molecules, interaction of a potential drug molecule with homologous proteins, understand a protein–protein interaction (PPI), assign putative roles to unknown proteins, understand the relationships between proteins that form multimolecular stable complexes like proteasome, molecular interaction study to understand various biological pathways (a number of interactions performed by numerous biological

molecules in a cell that result in certain other biomolecules and accomplishing different biological functions in a cell), and prediction of pollutants degraded by enzyme, etc.

The docking technique is also used extensively to investigate nanomaterials applications in different areas. Nanoparticles are solid colloidal particles (sizes vary from 10 to 1000 nm) having properties, such as increased surface-to-volume ratio, magnetism for bigger size particles, chemical stability, non-toxicity, biocompatibility, high saturation magnetization and high magnetic susceptibility, etc. These properties are the cause of its beneficial use in antibiotic treatment, antimicrobial (antiviral, antibacterial, antifouling, antifungal), nanocomposite coating, catalyst, and lubricants and thus are needed in biomedical applications such as targeted drug delivery, hyperthermia, photoablation therapy, bioimaging, biosensors, cell labelling, and gene delivery, etc. Many docking studies have been carried out to understand the molecular interaction between biomolecules and inorganic or organic (synthetic/natural) nanoparticles due to their uses in biomedical applications [2].
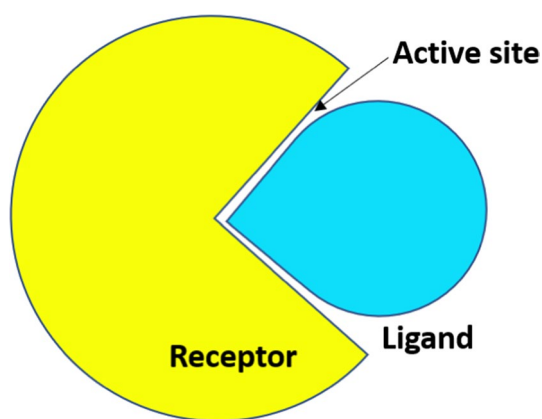
Databases and tools used in the docking study

The increased availability of three-dimensional structural information of molecules using different major reliable databases, protein structure prediction tools, macromolecular structure validation and quality assessment tools, chemical molecule drawing tools, etc. assist in preparing the molecules for docking. Rapid advance in the development of reliable computational docking tools, docking web servers contribute to solve the unknown problems by docking study between different types of molecules or atoms or ions interacting with each other. Chen has listed down docking databases and webservers in their review [6]. Although there are many accurate reliable docking algorithms emerging, limitations and challenges are still there because of imperfections in their scoring functions, handling protein flexibility, explicit water, etc. Docking software can be evaluated and their limitations can be overcome by benchmark testing with Critical Assessment of PRediction of Interactions (CAPRI) [7]. Various types of databases, software programs, and web servers used in different steps in the docking technique are listed in Table 1.

Molecules used in docking studies and binding energy calculation

When two molecules, in close proximity, favorably interact with each other, they bind to form a stable complex. The acceptor molecule is termed a receptor or target or macromolecule

Table 1 List of databases, software programs and web servers used in different steps in docking method

Type of database	Database
Macromolecular structures	Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) [22]
	Protein Data Bank in Europe (PDBe) [121]
	Protein Data Bank Japan (PDBj) [122]
	Biological Magnetic Resonance Data Bank (BMRB) [123]
	Electron Microscopy Data Bank (EMDB) [124]
	Molecular Modeling Database (MMDB) [125]
	Worldwide Protein Data Bank (wwPDB) [126]
	Electron Microscopy Data Bank (EMDB) [124]
Nucleic acid structures	Nucleic Acid Database (NDB) [127]
Chemical compound structures	PubChem [39]
	DrugBank [40]
	Chemical Entities of Biological Interest (ChEBI) [128]
	ChEMBL [129]
	ChemSpider [130]
	ZINC [131]
	eMolecules
Biomolecular complexes binding affinity	PDBbind [132]
Binding site of protein	LIGand Attachment SITE (LigASite) [45]
Protein–ligand interaction	BioLiP [133]
Protein–ligand binding affinity	BindingDB [134]
	MOAD [135]

Fig. 1 Receptor-ligand docked complex. Docking of a small molecule "Ligand" (cyan) to a bigger size molecule "Receptor" (yellow) to produce a stable docked complex. (Color figure online)

(though it is bigger in size compared to the second molecule) and the received molecule by the receptor is termed a ligand. The site where binding happens is known as the active site or binding site or binding pocket. The stable complex formed is termed a receptor-ligand complex as shown in Fig. 1. There are many different types of noncovalent interaction forces that cause the receptor and ligand to bind together to form a stable complex. Those are torsional, hydrophilic, hydrophobic, van der Waals, electrostatic, hydrogen-bonding, and desolvation, etc. So, the main goal was to find the most stable receptor-ligand complex with optimized geometry and minimum binding energy.

The energy score and energy terms of binding pose are calculated using scoring functions. The ongoing scoring functions applied in different docking algorithms are of four types as follows: i.e., force-field-based, empirical, knowledge-based, and machine-learning-based [8, 9]. Force-field based scoring functions use different non-covalent types of force fields to predict the protein–ligand interactions. Non-covalent interactions, such as van der Waals (ΔE_{VDW}), electrostatic ($\Delta E_{\text{electrostatic}}$), hydrogen bonding ($\Delta E_{\text{H-bond}}$), desolvation ($\Delta G_{\text{desolvation}}$) contribute towards the total binding energy. The extensively used generalized functional form of energy calculation in such type of scoring function is given as follows [9]:

$$\Delta G_{\text{binding}} = \Delta E_{\text{VDW}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{H-bond}} + \Delta G_{\text{desolvation}}$$

(1)

Second type of scoring function is empirical-based that calculates the fitness of binding between protein and ligand by adding the contribution from each energetic factor of the protein–ligand binding. In case of GOLD docking software program, the scoring function, ChemScore (the higher is the score the better is the docking result) is calculated as follows [10]:

$$\text{ChemScore} = S_{\text{H-bond}} + S_{\text{metal}} + S_{\text{lipophilic}} + P_{\text{rotor}} + P_{\text{strain}} + P_{\text{clash}} + P_{\text{covalent}} + P_{\text{constraint}}$$

(2)

where $S_{\text{H-bond}}$, S_{metal}, $S_{\text{lipophilic}}$ are rewarding scores for hydrogen bonding, coordinate bonding with metal ions and lipophilic contacts, respectively, whereas the P_{rotor}, P_{strain}, P_{clash}, P_{covalent} and $P_{\text{constraint}}$ are the penalties assigned for frozen rotatable bonds, ligand's internal strain energy, steric clashes between protein and ligand, covalent type docking, and restrained docking.

The third category of scoring function is knowledge-based scoring function which computes the binding score as the sum of statistical pairwise potentials between protein and ligand as follows:

$$A = \sum_{i}^{\text{lig}} \sum_{j}^{\text{prot}} \omega_{ij}(r)$$

(3)

where $w_{ij}(r)$ is the potential between any pair of two atoms i and j at a distance of r and is calculated by the Boltzmann inversion as follows:

$$\omega_{ij}(r) = -k_{\text{B}}T\ln\left[g_{ij}(r)\right] = -k_{\text{B}}T\ln\left[\frac{\rho_{ij}(r)}{\rho_{ij}^*}\right]$$

(4)

where $g_{ij}(r)$ is radial distribution function (RDF) of the atom pair $i-j$ separated at a distance of r, K_{B} is Boltzmann constant, T is temperature, $\rho_{ij}(r)$ stands for the density of the atom pair $i-j$ set apart at a distance of r, and ρ_{ij}^* denotes density of the atom pair $i-j$ in a reference state, in which the interactions between atoms are taken as 0.

The fourth category of scoring function is machine-learning (ML) based. It learns the functional form of binding affinity from the training data. The complicated functional form applies ML methods, namely deep neural network (DNN), convolutional neural network (CNN), graphical neural network (GNN), support vector machine (SVM), random forest (RF), eXtreme gradient boosting (XGB) [11–14]. Quantity structure activity relationship (QSAR) analysis can be useful to model biological, physicochemical, and pharmaceutical properties of the ligand. Similarly, the properties of the protein, and patterns in the protein–ligand interaction can be modelled using some potential descriptors and ML methods can be applied in QSAR analysis to attain a statistical model to calculate the protein–ligand binding score. Scoring functions are discussed more in detail elsewhere [8, 9, 15]. Here we are showing functional form of empirical free energy calculation in AutoDock.

In AutoDock docking tool, the net free energy calculated in a docking process is summation of different types of potentials, including van der Waals (ΔG_{vdW}), hydrogen bonding ($\Delta G_{\text{H-bond}}$), electrostatics (ΔG_{elec}), tortional (ΔG_{tor}),

and desolvation (ΔG_{sol}) [16]. The net change in free energy calculation in AutoDock is given as follows:

$$\Delta G = \Delta G_{\text{vdW}} \sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}\right) + \Delta G_{\text{H-bond}} \sum_{i,j}E(\theta)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right)$$
$$+ \Delta G_{\text{elec}} \sum_{i,j}(q_iq_j/\varepsilon(r_{ij})r_{ij} + \Delta G_{\text{sol}} \sum_{i,j}(S_iV_j + S_jV_i)e^{(-r_{ij}^2/2\sigma^2)} + \Delta G_{\text{tor}}N_{\text{tor}}$$

(5)

where ΔG are the coefficients of different potentials obtained from different studies of receptor-ligand complexes accompanied by known binding constants. The sum is carried out for each pair of ligand atom, i, and receptor atom, j and to each pair of atoms in the ligand separated by three or more bonds. The in-vacuo contributions consist of three following interaction energy terms: (i) Lennard–Jones (12–6 dispersion/repulsion); (ii) a directional 12–10 hydrogen bonding, where $E(\theta)$ the directional weight based on the angle, θ, between the probe and the target atom; and (iii) a screened Coulombic electrostatic potential [17, 18]. Detailed of their parameterizations have been described by Morris et al. [19]. A measure of the unfavorable entropy of ligand binding is added to the in-vacuo function. This entropy term is because of the restriction of conformational degrees of freedom and is proportional to the number of sp^3 bonds in the ligand, N_{tor} [20]. The functional forms of various scoring functions used in different docking algorithms are presented elsewhere [21].

Important steps in the docking technique

The steps in a docking procedure are explained in the following.

Receptor selection

Based on the input file format needed in a software tool (varies from one software to another), three-dimensional structural information of a receptor in any of the file formats, like pdb, mol2, sdf, smi, xyz, and cif, etc. is given as input. The pdb file format is accepted commonly in most of the docking software tools and web servers. High-resolution three-dimensional structural information of a receptor can be obtained by experimental methods, namely solution nuclear magnetic resonance (NMR), X-ray diffraction, electron microscopy, solution scattering, and neutron diffraction or downloaded from any structural database, such as, RCSB PDB [22]. Structural information of molecules if not available by experimental methods or structural databases can be obtained by any drawing tool like Avogadro [23]. For proteins, homology modeling or threading can be done to predict the structures of unknown proteins but gives less accurate results compared to structures obtained from PDB. Machine learning can also be useful for protein structure prediction [24]. The quality of the structure can be

checked using any validation software, such as PROCHECK [25].

Receptor preparation

After selecting the target macromolecule it is prepared for docking. Ligand, cofactor, ion, lipid, water, solvent, and other unnecessary molecules if present along with the receptor (usually these co-crystallized molecules are found in a database) are removed by any molecule editor program namely PyMOL, Chimera, AutoDock as presence of these molecules in the binding pocket of the target may obstruct the binding of the ligand [26, 27]. Schiebel and Barillari have elucidated the fascinating role of water in protein–ligand interaction [28, 29].

Despite the fact that involvement of water molecules in receptor–ligand interaction is not well understood, they work in either of the two ways as follows: (1) being removed from the binding cavity of receptor that results in increase in binding affinity between the receptor and the ligand as displacement of water molecules may encourage ligand to bind to its receptor; (2) remains in the binding pocket to make the interaction more stable because water molecules get involved in making hydrogen bonds with ligand and receptor at the same time forming hydrogen-bonding network which makes stable binding between them. Hydrated docking (water mediated protein ligand docking) can be performed by docking software suites AutoDock vina, GOLD, GLIDE [30–33]. These tools permit water molecules to be switched on or off during the docking process [34]. After removing water molecules, missing atoms are added, multiple or low occupancies are refined, chain breaks are repaired and included, hydrogen atoms are added to protonate the molecule which is necessary especially in the binding site though these may be involved in any type of interaction with the ligand. The accurate protonation of the target receptor molecule gives good docking result with correct prediction of van der Waals surface and dipole moment in the binding pocket. Programs like AutoDock, Reduce, Maestro (Schrödinger), PropKa, Chimera are used for protonation of receptor [26, 27, 35, 36]. Metal ionization states are corrected to deal with accurate formal charge and force field. Bond order to HET groups is counted. In case of proteins or peptides, some mutated amino acids may be found in structures deposited in PDB for stability, crystallization and other biochemical causes which can be changed to wild residues specifically in the binding pocket; capping termini is done with ACE (N-terminal capping acetyl group) and NME (C-terminal *N*-methyl amide capping group) residues. The stable conformation of the receptor is obtained by relaxing it that permits hydrogen atoms to move and freely minimized and heavy atoms to move sufficient enough to make strained bonds, angles, and clashes relaxed. Structural optimization can be done to minimize the energy to get the most stable conformation by programs, such as Avogadro, UCSF Chimera's Dock Prep, Schrödinger's Protein Preparation Wizard and molecular operating environment (MOE) programs [23, 37, 38]. The partial charges (also called as point charges) are added to each atom to predict the electrostatic potential around a receptor molecule because of Coulombic interaction between the point charges. The receptor should be taken as a flexible body to reflect the real-time analysis of ligand-receptor dynamics (ex: amino acid residues in the binding site may be involved in binding).

Ligand selection and preparation

Structural information of a ligand molecule or atom or ion can be obtained similarly from any structural database, including PubChem, DrugBank, and PDB etc. or drawn using any chemical molecule drawing tool, like ChemDraw, Avogadro, ChemIDplus, ACD/ChemSketch freeware, BIOVIA Draw, ChemSpider, and PubChem Sketcher, etc. [22, 23, 39–44]. They should be treated as flexible bodies as well except for ring conformations. The number of rotatable bonds is calculated. The more the number of rotatable bonds, the more difficult and time-consuming the docking process will be due to the increase in search space. The ligand geometry should be optimized to have a stable conformation by any structure optimization tool, such as Avogadro, UCSF Chimera's Dock Prep, NAMD tool and the atomic partial charges are then assigned to it by partial charges addition tool, such as UCSC Chimera (Table 2).

Docking

The ligand can be docked into the binding pocket of the receptor either using the known location of the active site (active-site docking) or searching for it (blind docking). If the structure of a holoprotein (protein bound with ligand) gets available then the binding site information can be useful for active-site docking. Otherwise, binding site prediction tools, namely LigASite [45] can be useful to predict the binding pocket information. In the molecular docking process, the search space to fit the ligand is explored considering the flexibility of both receptor and ligand.

Different approaches used to predict the binding site depend on (1) templates, (2) energy functions, (3) geometric considerations, and (4) machine learning [46]. The template-based approach predicts the binding site of a given protein using well-studied proteins having similar structures with known binding pocket information. The energy-based approach guesses the binding pocket information by using energetically favorable regions of the protein to bind the ligand in it. The geometry-based approach uses the geometry of the protein to find binding pocket of it. Machine

Table 2 List of software programs/web servers used for different steps in docking process

Aim	Software program/web server
Protein structure prediction	IntFOLD [136]
	Schrödinger Prime [137, 138]
	AlphaFold2 [24, 139]
	RaptorX [140, 141]
	Biskit [142]
	EsyPred3D [143]
	FoldX [144]
	Phyre and Phyre 2 [145]
	HHpred [146]
	MODELLER [147]
	CONFOLD2 [148]
	Molecular Operating Environment (MOE)
	ROBETTA [149]
	BHAGEERATH-H [150]
	SWISS-MODEL [151]
	Yasara[87]
	AWSEM-Suite (Associative memory Water mediated Structure and Energy Model) [152]
	trRosetta [153]
	I-TASSER [154]
	Rosetta@home [155]
	Abalone
Macromolecular structure validation and quality assessment	CheckMyMetal [156, 157]
	MolProbity [158]
	NQ-Flipper [159]
	Procheck [25, 160]
	ProSA-web [161]
	Verify3D Structure Evaluation Server
	WHAT IF [162]
	WHAT_CHECK [163]
Chemical molecule drawing	ChemDraw [41]
	Avogadro [23]
	ChemIDplus
	ACD/ChemSketch freeware [42]
	BIOVIA Draw [42]
	ChemSpider [42, 43]
	PubChem Sketcher [44]
	Marvinsketch
	OpenEye Omega [164]
	Schrodinger LigPrep
Molecule visualization	PyMOL [66]
	Schrodinger Maestro
	VMD [67]
	Avogadro [23]
	UCSF Chimera [26]
Molecule protonation	AutoDockTools (ADT) [100]
	Reduce [35]
	Schrodinger Maestro
	PropKa [36]

Table 2 (continued)	Aim	Software program/web server
	Structural optimization	UCSF Chimera [26]
		Avogadro [23]
		UCSF Chimera's Dock Prep [37]
		Schrödinger's Protein Preparation Wizard [38]
		Molecular Operating Environment (MOE)
	Binding site prediction	NAMD [165]
		GHECOM [166]
		Schrödinger SiteMap [167]
		FTMap [168, 168]
		COACH [170]
		SURFNET [171]
		Fpocket [172, 173]
		Depth [174]
		AutoSite [175]
		DeepSite [176]
		DeepCSeqSite [177]
	Computational docking	AutoDock [16, 19, 27, 100, 104, 109, 110]
		AutoDock Vina [55, 95]
		SwissDock [98]
		PatchDock [120]
		Hex [105]
		HADDOCK [178]
		GOLD [179]
		Schrodinger Glide [30]
		DOCK [180]
		HDock [181]
		ClusPro [96]
		NPDock [182]
		MIB2 [183, 184]
		RosettaDock [185]
		FlexX
		Surflex
		FRED/HYBRID
	Partial charges addition	UCSF Chimera [26]

learning approaches learn from existing vast amount of high-resolution protein structure data available in different structure databases and generalize competently to new data and predicts the binding pocket information more accurately than the other three approaches. The available structural protein–ligand binding data are collected and then normalized and used to guess the binding site information of the input protein by machine learning approaches, like shallow supervised learning algorithm, artificial neural network, convolutional neural network, and ensemble methods. Data acquisition and preprocessing, feature engineering, model development, training–testing, hyperparameter tuning and evaluation are the five important steps in machine learning approach to bind the binding site information of a given protein. A simplified view of machine learning based approach is shown in Fig. 5. Different methods to search for binding site information of a given protein is discussed more in detail [46].

There are three main approaches that treat the ligand as flexible used in various popular docking algorithms for searching conformational space [15] As follows: (i) systematic search (incremental construction, conformational search, databases), (ii) random/stochastic search (Monte Carlo, genetic algorithms, tabu search), (iii) simulation method (molecular dynamics, energy minimization). The energy score and energy terms of each binding pose are calculated using scoring functions. Search approaches are discussed in depth by Kitchen et al. [15].

A scoring function is a mathematical function that is used to predict the binding affinity between a receptor and

a ligand when they are bound to each other. The scoring function is used to evaluate and give a ranking of predicted possible ligand conformations in the docking result. Ideally, the best-scored ligand is the best binder having the highest binding affinity to the receptor. There are four categories of scoring functions used in different docking algorithms [8, 9] as follows: (i) force-field-based, (ii) knowledge-based, (iii) empirical, and (iv) machine-learning-based. First three (i, ii, and iii) categories are classical type scoring functions. Force-field based or physics-based scoring functions are calculated using interaction due to van der Waals, electrostatic, desolvation. The docking software, such as GoldScore, early versions of AutoDock and DOCK, employ force-field based scoring function [33, 47–49].

Knowledge-based or potential of mean force scoring function use statistical inspections of potentials obtained from experimentally found three-dimensional protein–ligand structures and predict the potential of mean force by inverse Boltzmann distribution. Examples of such type of scoring functions are PMF, DrugScore, SMoG and ITScore, KECSA [50–54].

Empirical or regression-based scoring function depends on list of weighted scoring terms contributed by different types of intermolecular interactions, such as van der Waals, hydrogen bonding, hydrophobic, desolvation, electrostatics, number of rotatable bonds (entropy), and many more (similar to force-field-based scoring function). The weighted scoring terms are derived by regression analysis of experimentally derived binding energy and three-dimensional structural information of known protein–ligand complexes (similar to Knowledge-based scoring function). Examples of empirical-based scoring functions are LUDI, Chem-Score (employed in GOLD), GlideScore (implemented in Glide), X-Score, F-Score, SCORE, Fresno, Vina, Lin_F9 [20, 55–60].

Machine-learning (ML) or descriptor-based scoring functions use machine learning techniques. The functional form of binding affinity learns from the existing data (training data). Docking tools using ML scoring functions give more accurate docking results compared to classical scoring functions-based docking tools. These ML scoring functions employ ML methods, such as deep neural network (DNN), convolutional neural network (CNN), graphical neural network (GNN), support vector machine (SVM), random forest (RF), eXtreme gradient boosting (XGB) [11–14]. Scoring functions are discussed more in detail elsewhere [8, 9, 15].

Evaluation and analysis of docking result

After the docking simulation ends, evaluation and analysis of the docking results are important steps to interpret the docking result. Best binding pose (i.e., orientation and geometry), closest to native structure with top score (highest docking score) is chosen based on the estimated lowest binding energy. In the docking result, all potential hydrogen bond acceptors and donors of the ligand should be satisfied [61], charged groups in the ligand should interact with oppositely charged groups of the receptor. The accuracy of the docking result closest to native structure can be verified by similarity checking with the experimental data if available. The certainty of the best binding pose chosen can be evaluated by the following approaches: (1) calculated root mean square distance, also called as root mean square deviation (RMSD, that is used to predict the correct geometry, orientation, and position of the ligand) value of the successful predicted pose of the ligand from its reference native pose must be less than 2 Å, and (2) comparisons of different types of protein–ligand intermolecular complex interactions, such as van der Waals (vdW), electrostatic, hydrogen bonding, tortional, and desolvation for the estimated pose and experimentally derived one [62–65]. The second approach, i.e., based on key interactions could be a better relevant approach to measure the docking accuracy. Combination of approaches with key interactions and shape complementarity between the receptor and ligand could also be helpful to verify the correctness of the docking result [62]. Accuracy of the obtained ligand conformations can be checked by visualizing and comparing the molecular interactions between the observed and experimental result [62]. Molecular three-dimensional structural visualization tools, namely PyMOL, VMD, UCSF Chimera can be useful to view and inspect the docking result [26, 66, 67]. The predicted binding affinity between a receptor and a ligand is measured by the calculated binding free energy to bind the two molecules, which can be estimated by different scoring functions used in different docking tools in terms of scores. These scores cannot be compared directly with the experimentally derived binding data [8]. An extensively used measure, Pearson correlation coefficient (PCC) can be applied to find the linear correlation between the experimentally derived binding data and scores predicted by docking tools for sets of test data [68, 69]. Alternative measure is by applying Spearman ranking correlation coefficient to rank the predicted and experimentally derived scores and then find the statistical relationship between the two ranked sets [70].

Virtual screening (a computational method used in drug discovery to find the best possible candidates among large number of molecules available to bind with a target molecule (e.g., protein or enzyme). The screening capability and predictive ability of good hits depends on the power of a scoring function used in molecular docking. The area under the receiver operating characteristic curve (AUC) and best 1% enrichment factor (EF1%) can be used to see the ranking performance of a scoring function [71]. How good the best candidates have been ranked is assessed by AUC which

gives a value 1.0 to the accurate ranking results and 0.5 for random ranking results. The ratio of percentage active compounds of top 1% of the ranked compounds and overall percentage of active compounds is termed as EF1%. Afterwards, EF1% can be normalized to have NEF1% by dividing estimated EF1% by the best EF1%. NEF% checks the quality of ranking performance of scoring function accurately by checking the value 1.0 means most of the active compounds are ranked in top 1% of the ranked compounds and the value 0 means none of the compounds are selected as active [72]. Comparative assessment of scoring functions (CASF) evaluates the performance of a scoring function [68, 69]. The scoring function gets analyzed by its scoring, ranking, docking, and screening power using by CASF benchmark. LIT-PCBA benchmarking datasets can be useful to evaluate the quality of a scoring function in machine-learning based virtual screening [73].

Moreover, molecular dynamics (MD) simulation can be used in the next step to affirm and refine the model of the docking complex [74–76].

The best docking result can be interpreted with all types of information that tells how, why and where the ligand binds with the given receptor. It includes 3D structural information of the receptor-ligand complex, details about the interacting residues, any type of contact within a certain distance Ex: 3, 3.5, 4 Å, polar contacts, π–π and π-cation interactions, binding affinity, docking score, net binding energy with contributions from different types of interaction potentials, such as van der Waals (vdW), electrostatic, hydrogen bonding, tortional, and desolvation, RMSD value from the reference conformation. Visualization of different type of interactions (hydrogen bonding, hydrophobic, ionic, aromatic, π-π, π-cation) and regions of contact the receptor makes with the ligand in the binding cavity aids in confirmation of the binding stability.

Types of docking

Docking study can be of three types, namely rigid docking, flexible-rigid docking, and flexible docking (based on the flexibility of the interacting molecules, receptor, and ligand) as shown in Fig. 2 [77]. Flexible docking gives more reliable and accurate results because the relative bond angle and bond length of molecules may vary. Pagadala and his

Fig. 2 Different types of docking studies based on flexibility of receptors/ligands considered in molecular interaction. (Color figure online)
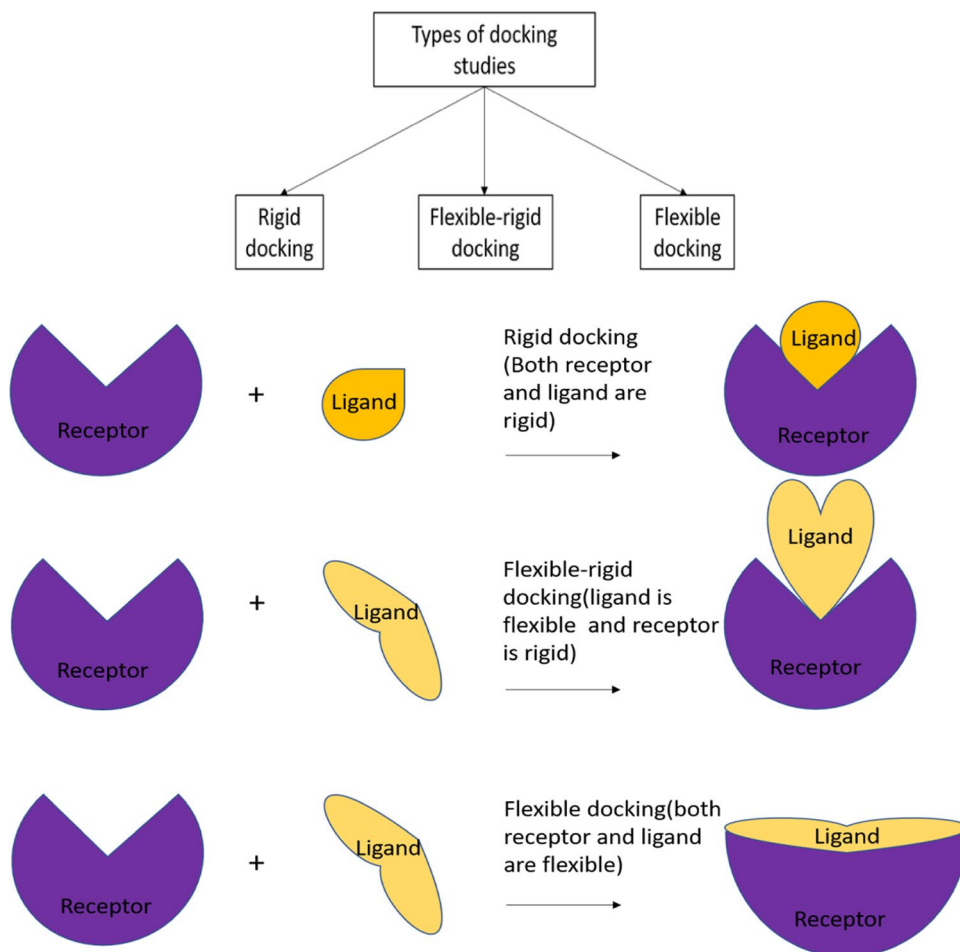
co-workers have presented a review on different molecular docking programs based on rigid and flexible docking [78].

Rigid docking

In rigid type docking, both ligand and receptor molecules are considered rigid bodies. Their shape is not changed and thus the internal geometry of each molecule is kept fixed. Its position can be varied and thus the translational and rotational degrees of freedom are only considered. This is an early docking method that can be carried out between macromolecules, such as two protein molecules and the result is less accurate and unreliable and thus is less frequently used in current docking studies. The lock-and-key principle can be applied in this method [79].

Flexible-rigid docking

It is a semi-flexible docking method. In this case, either ligand or receptor is taken as a rigid body. Usually, the shape of the receptor is kept fixed, and the conformation of the ligand is varied. This method gives more accurate and better reliable results than the rigid docking method and is thus frequently used.

Flexible (soft) docking

It is a fully flexible docking method, in which both ligand and receptor are considered as flexible bodies, i.e., an enumeration of rotations of the molecules (both receptor and ligand) is done to search for optimized conformation and orientation of the molecules to interact with each other. The molecule's shape can be varied by changing the torsion angles and rotatable bonds. This method results in the prediction of docked conformation with high accuracy that most probably resembles experimental results but may require heavy computational calculation and time.

In both semi-flexible and fully flexible docking methods, the induced-fit principle can be implemented, and the docking process becomes complicated when the interacting molecule has many conformational degrees of freedom [79].

Importance of structural information of receptor's binding pocket in docking study

A receptor interacts with a ligand at its binding site. But the docking study can be done with or without using the binding site information as shown in Fig. 3 [4].

Blind docking

Docking without taking into consideration the structural information of the binding or active site (where the ligand is expected to bind and interact with it) of the receptor molecule is called blind docking. If the ligand under study does not interact with the receptor in its binding site (the receptor may have a binding pocket where it interacts with its another partner ligand in nature), or the location of the binding site is unknown then the entire surface of the receptor is searched for the binding pocket to accommodate the ligand. If it is not possible to target the whole receptor molecule, then possible pockets can be searched to accommodate the ligand. This type of docking study is usually done to investigate the possible interaction between a novel protein and a ligand under study.

Fig. 3 **A** A yellow dashed line rectangular docking box is created around the receptor's whole surface. The whole surface is scanned for a possible binding pocket to dock with the ligand. The ligand is connected in the best binding site to make a stable complex. **B** The docking box is created around the receptor's binding site surface. The ligand is connected to the binding site to make a stable complex. (Color figure online)
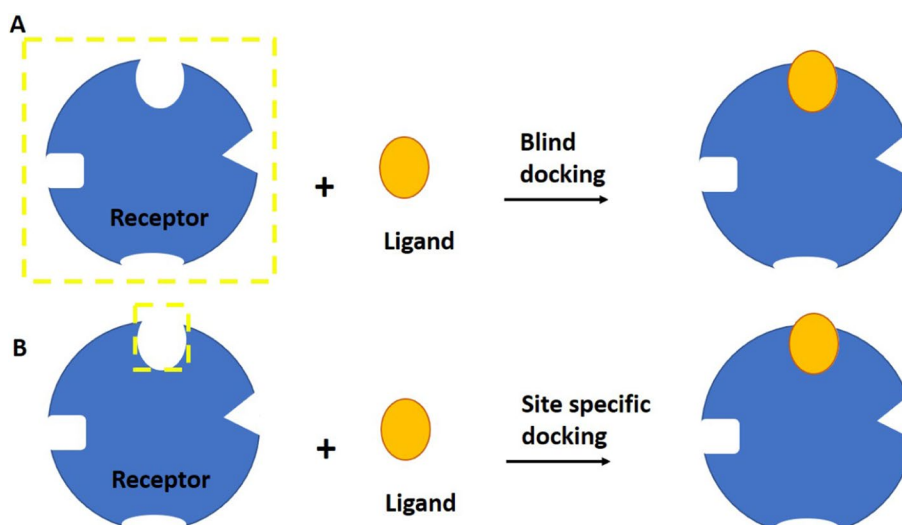
Active site docking

A docking study based on the binding site information of the receptor is called active site or site-specific docking. Previously found results about the position of the binding pocket, active residues of the receptor, and binding modes of known ligands in the binding site reduce the search area of interest and thus increase the speed and effectiveness of docking search. Sometimes for well-studied protein molecules, knowledge of the binding pocket information can be used, in case it is not known it can be retrieved by mapping the protein to its family of well-recognized proteins having similar structure and function with known active site information. In such a docking process, the ligand binds only in the active site and nowhere else.

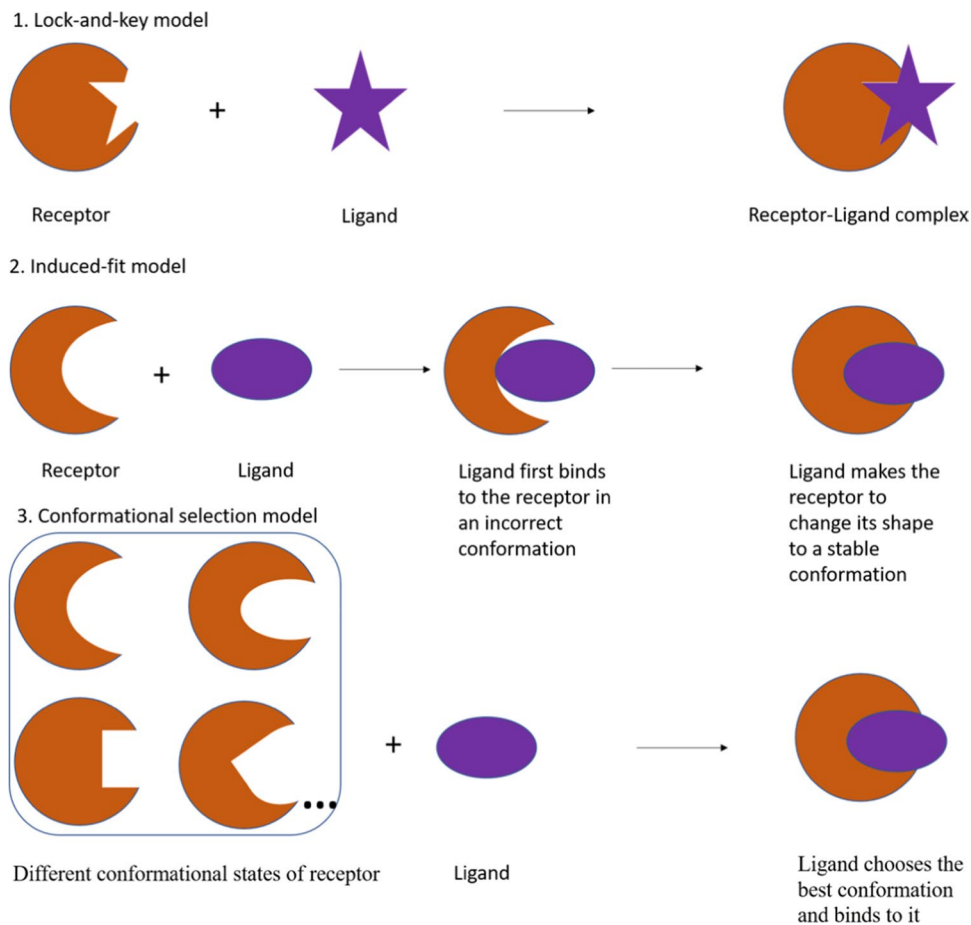## Models to understand docking study

In the molecular docking method, molecular recognition is processed by the computational simulation to find the best conformation of ligand and receptor based on different features and complementarity of both molecules that predicts the binding affinity and interaction between them [79]. There

are three basic models as shown in Fig. 4, which help to understand the docking simulation study more precisely.

Lock-and-key

It is the first proposed model suggested by Fisher to predict a docked complex of a rigid receptor and a rigid ligand [80]. In this type of rigid docking, the ligand is taken as a key and the receptor as a lock. The key tries to open the lock based on the shape complementarity between the receptor's active site and the ligand. The computation must try to find the accurate relative position and orientation of the "key" with respect to the lock which will unlock the "lock". This model is quite simple but may not reflect the dynamic behavior of molecules (both receptor and ligand). HINT (Hydropathic INTeractions) docking program uses the experimentally derived solvent partition, Log$P$ values between water and 1-octanol for modeled organic molecules to estimate and predict the score of hydropathic interactions between atoms of protein and ligand in the docking process [81]. Cozzini and co-workers, in their study have used this program to estimate the binding free energy of 53 protein–ligand complexes constructed by 17 known proteins of solved 3D structures, available in protein data bank (PDB) and hydrophobic/polar

Fig. 4 Different docking models: (1) Lock and key model, (2) Induced-fit model, (3) Conformational selection model. Among these three models, only induced-fit model induces a change in the shape of the binding pocket. (Color figure online)

1. Lock-and-key model

Receptor + Ligand → Receptor-Ligand complex

2. Induced-fit model

Receptor + Ligand → Ligand first binds to the receptor in an incorrect conformation → Ligand makes the receptor to change its shape to a stable conformation

3. Conformational selection model

Different conformational states of receptor + Ligand → Ligand chooses the best conformation and binds to it

ligands. They have demonstrated the hydropathic (lock-and-key) complementarity HINT interaction maps to ascertain the binding between protein and ligand at different pH (3.5, 4.5, and 5.5).

Induced-fit

Docking can be better understood by the "hand-in-glove" concept which considers the flexibility of the molecules. The molecules can change their shape to best fit each other while binding. Such type of binding is termed "induced-fit" binding. This model was proposed by Koshland to explain specificity of enzyme during protein synthesis with the concept that the structural changes in an enzyme get induced by a substrate while binding to it at its active site to form a stable enzyme-ligand complex [82]. The model hypothesizes the interface optimization of the enzyme by conformational modifications of its residues present in its active site to react with the substrate by physical interactions. Basically, the binding site of the enzyme is flexible in shape. Before binding, the substrate does not fit into the binding pocket of the enzyme and while binding, the active site of the enzyme undergoes a change in shape to complement the shape of the substrate. This model was backed up by the fact found in the structures of protein–ligand complex obtained in X-ray crystallography experimental method deposited in protein data bank (PDB) which validates the enclosure of ligands in the binding pockets of proteins and proposes covering up the ligands by the binding-site residues after the binding process starts [22, 83]. The difference between structural information of a protein attached with a ligand (holo) and without a ligand (apo) ascertains the conformational alteration of a protein [84]. The ligand binds to the protein conformation with least energy and following the initial binding process the protein undergoes a conformational change. So, the molecules undergo conformational changes to fit each other and minimize the net binding energy. The result obtained using this model best suits the real fact, i.e., the molecules and atoms show physical movements in nature. The ligand first binds to the receptor (with its original shape) and then causes the structural change in the receptor to accommodate it. Without the ligand, the structural changed state of the receptor may not exist. Flexible docking using the induced-fit model permits conformational changes while predicting the binding pose in the interaction between protein with other molecules ex: ligand, protein, or peptide [34, 85]. Fleksy program recognizes both ligand and receptor as flexible and can be used for flexible and induced-fit docking [86]. In this approach, structural ensemble of receptor molecule conformations is produced from backbone-dependent rotamer library in to which the ligand is docked, the best ranked (ranking based on a consensus scoring function which depends on docking scores

and molecular dynamics force field interaction energies) selected poses follow a refinement step using the Yasara program, which includes steepest descent minimization, follow a simulated annealing and finally the results with a list of minimized docked complexes are obtained [87]. In the minimization step, both receptor and ligand molecules are allowed to move. Using induced-fit approach Molegro Virtual Docker tool docks a ligand against a protein with user given constraints, option of removal of water molecules, structural alignment of protein and ligand, and ligand-based Graphics processing unit (GPU) screening [88]. FiberDock is a protein–protein docking refinement web server [89]. In the docking process it generates potential candidate docking complexes, models their backbone and side-chain residues based on their movement in the protein–protein interaction, performs refinement of the docking complexes and gives scores to them. The flexible refinement of candidate docking complexes follows induced-fit model. HADDOCK, a flexible protein-peptide docking software uses both the induced-fit and conformational selection mode approaches [90]. This program yields ensembles of favorable peptide structures and these conformations are further refined allowing movement in side-chain and backbone residues and then subjected to dock in the binding pocket of the target protein using the induced-fit approach.

Conformational selection or population shift

The conformational selection model was proposed by Monod, Wyman, and Changeux in 1975, which suggests that the unbound receptor exists in different conformational states based on relative free energy according to Gibbs distribution before binding to the ligand and the ligand chooses the best fit among them to form the final stable docked complex [91]. The difference between the induced-fit and conformation selection model is that in the case of the induced fit model, the ligand binds to the receptor before the structural change in the receptor, and in the conformation selection model the structural change of the receptor occurs before the ligand binding. This model is extensively used even though it requires more time and high-throughput computing facilities. HADDOCK protein-peptide docking software uses conformational selection and induced-fit approach [90].

Combination lock and key

In the docking study, both lock and key and induced-fit models represent a simplified view of the interaction between ligand and receptor. In real scenario, the interaction is a net result of many complicated interaction processes. In this view, Tripathi have proposed a combination of lock and key model [79]. Combined complementary features of receptor and ligand help to understand the stable interaction

between them. The more the complementary features fit, the better is the binding. The complementary features can be (1) geometric properties like shape, size, volume, surface area, bond lengths, bond angles, torsional angles, etc., and (2) physicochemical properties like solvation, electrostatic, hydrophobic, polar, nonpolar, van der Waals, etc. For molecules, energy-based features change, on the contrary, geometry-based features do not change a lot in three-dimensional space. Pharmacophoric chemical features, including hydrophobic centroids, aromatic rings, hydrogen bond acceptors or donors, cations, anions, etc., geometric features, and electronic features of molecules due to the presence of electrons contribute towards the net interaction between the molecules. Electronic features, like electrostatic, hydrophobic, van der Waals have varying field intensity in different spatial arrangements of the molecules and thus create a unique electric field whose strength depends on the distance between the atoms and molecules, and it varies from one point to another. Energy contributed by each type of molecular feature, namely pharmacophoric chemical, geometric, and intrinsic electronic towards the net interaction energy corresponds to distinct interaction between molecules. The pattern of these molecular features in three-dimensional space sets in molecular recognition.

This hybrid model is a novel approach and awaits development of new docking software using it.

Use of molecular docking in drug design

Molecular docking is an effective method in drug discovery [5]. There are several applications of molecular docking simulation in drug discovery, such as to explore the binding interaction between protein target and ligand, hit identification and optimization (virtual screening), predicting new disease related targets for existing drugs (drug repositioning), reverse screening for target fishing and profiling (prediction of targets by ligand-receptor complementarity), multi-receptor ligand design and repositioning, investigating the connection between different targets associated with a particular disease (polypharmacology), prediction of structural information those are responsible for effective receptor-ligand binding (ligand-target binding rationalization) [92]. It can be merged with large scale screening methods to recognize the binding pocket of a receptor, predict new targets for known ligand molecules and speculate the adverse drug effects.

Molecular docking approach can be incorporated with artificial intelligence means. Artificial intelligence has two subclasses namely, machine learning and deep learning which have remarkable applications in pharmaceutical industry. Machine learning (ML) [a subset of artificial intelligence (AI)] implements statistical methods to upgrade the

machines with experience and novel data, whereas deep learning (DL) (a subcategory of machine learning) empowers the machine to learn from data based on neural network. Machine learning and deep learning methods can be used to predict the binding pocket information of protein where the ligand is bound, binding affinity of the ligand for the protein and its binding orientation and geometry effectively.

ML algorithms are of the following two types: (1) supervised, (2) unsupervised [93]. Unsupervised machine learning algorithms are used to model the training data even if output data is unavailable. So, these algorithms are employed to group data based on feature similarity. On the other hand, in case of supervised machine learning algorithm, the available output is fed into the algorithm with the input data for training. Dhakal and his coworkers have categorized machine learning methods in 2 groups; (1) classical ML (non-deep learning method), (2) modern deep learning or simply deep learning method [46]. Different classical ML methods and modern deep learning methods for prediction of binding site, binding affinity between protein and its ligand, prediction and scoring of binding pose of protein and ligand are listed and summarized by Dhakal and his coworkers [46]. Machine learning learns the interlink between the physicochemical properties and the interactions between protein and ligand from the known binding complexes and implements statistical methods to predict the interactions of unknown protein–ligand complexes. Figure 5 shows the workflow of machine learning method to predict the interaction between protein and ligand [46].

The machine learning (ML) algorithms discovers the underlying hidden information from existing data. After training data set are fed into the algorithm, the ML model learns from it automatically. It can identify the hidden relationships and patterns lying in the data which is unidentifiable by experts. The ML model while training can be validated by validation data set to examine how well the model works. Now the model can be executed for a new test data and different assessment metrics can be used to access the attainment of the model prediction [46]. As this model learns from the data, the more the amount of data are fed into it, the better it works. There are several widely used datasets available, which have useful resources for training, validating, and testing data of ML models [46].

Deep learning (DL) method, the subcategory of machine learning method is getting more popularity because of its potential to detect the hidden complex relationships within the data. DL method can automatically extract biological features from raw data. DL neural networks can be employed for predicting pose of a ligand in a protein's binding pocket and ranking it [93]. Conventional neural networks can be utilized to visualize protein–ligand binding complex in terms of three-dimensional grid and
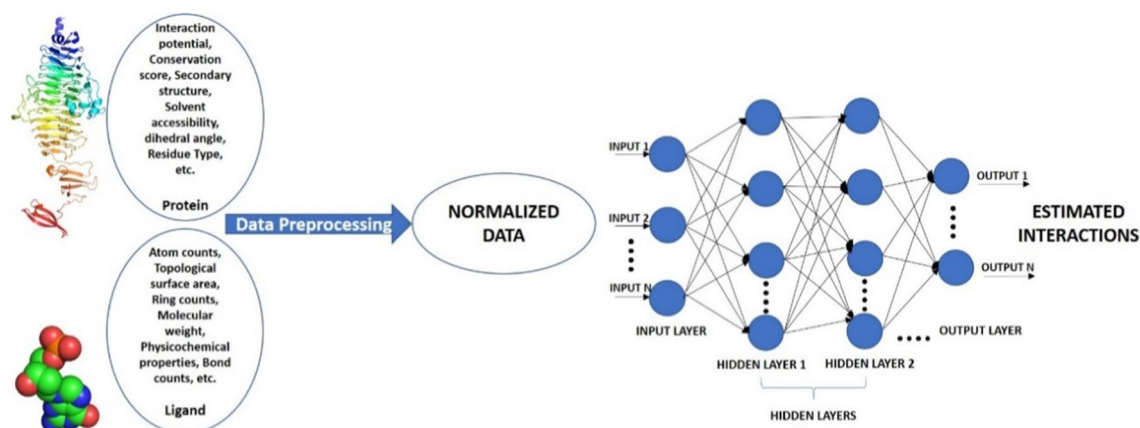
Fig. 5 Workflow of machine learning method in predicting protein–ligand interaction [46]. (Color figure online)

predict binding affinity [94]. DL scoring functions can give remarkable result compared to others. Therefore, together with the abundance of physicochemical properties of protein and ligand along with the modern DL technique a very high accuracy can be achieved. In future, with multitasking learning feature, advance DL method can predict binding site, binding affinity, binding pose simultaneously [46].

Docking in organic, inorganic, and hybrid systems

The molecular docking method can be applied to both organic and inorganic molecules. An organic molecule always contains carbon whereas a pure inorganic molecule does not contain carbon. The organic molecule can be a protein or non-protein or synthetic organic molecule. In the interaction study, the organic receptor and ligand molecule can be any synthetic organic molecule, such as synthetic organic nanoparticle, plastic products made from polymers (small repeating molecules), elastomers (flexible rubber material), medicine, organic dye, artificial sweeteners like stevia and equal, or bio-organic molecule, including carbohydrate, lipid, protein, enzyme, DNA, RNA, or organic nanoparticle derived from natural materials, etc. The inorganic receptor and ligand can be any molecule other than organic molecules and are not found in living beings, like ammonia, hydrogen sulfide, all metals, most elements (such as calcium) and metal nanoparticles, etc. In most of the docking tools, like AutoDock, AutoDock Vina, and ClusPro, etc., inorganic molecules are considered

only as ligands and organic molecules are taken as both receptors and ligands [19, 55, 95, 96]. Metal nanoparticle, atom, or ion can be taken as an inorganic ligand, but not as a receptor in a docking study. In case of binding study between protein-metal nanoparticle through molecular docking technique, smaller size protein is assumed as a receptor and the bigger size nanoparticle as a ligand. The types of molecular docking problems depend on the molecules under study which can be organic, inorganic, or hybrid. There are mainly three types of docking problems, namely organic-organic, organic–inorganic, hybrid receptor/ligand, and inorganic-inorganic.

Organic-organic

In this case, both receptor and ligand are organic molecules. Analyzing problems, like protein–protein, protein-DNA/RNA, protein-synthetic organic nanoparticle, protein-organic ligand, protein-organic drug, DNA/RNA-organic ligand, DNA/RNA-organic drug, enzyme-protein, enzyme-organic substrate, and synthetic organic dye-polymer moiety, etc. molecular interactions are examples of organic-organic intermolecular docking studies.

Biomolecules-organic ligand docking case studies

There are many studies over the past that focused on such type. Basu A have found a conformational change in the receptor, a chain of angiotensin-converting enzyme 2 (ACE2) after docking to SARS CoV2 spike protein fragment by ClusPro web server with binding energy − 779.8 kcal/mol as shown in Fig. 6A [96, 97]. The binding complex of SARS Cov2 spike protein and ACE2 protein have been taken as a therapeutic target for SARS CoV2 treatment with naturally
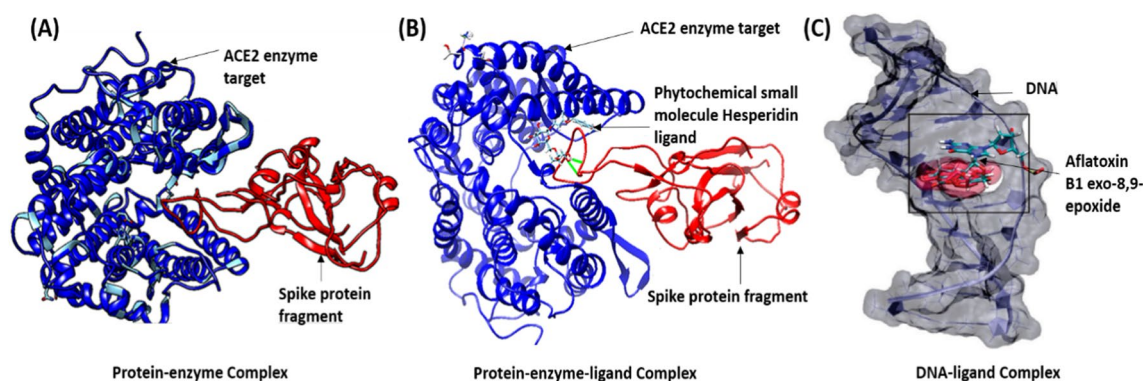
Fig. 6 **A** Docked complex structure of spike protein fragment of coronavirus (SARS-CoV2) and human receptor angiotensin-converting enzyme 2 (ACE2) using molecular docking webserver ClusPro 2.2 [96, 97]. Spike protein fragment (331–524) is shown in red cartoon view, and human ACE2 is shown in blue cartoon view. **B** Docked complex structure of phytochemical "hesperidin" (ligand) and protein-enzyme bound structure of spike protein fragment of coronavirus (SARS-CoV2) and angiotensin-converting enzyme 2 (ACE2) (receptor) using SWISSDOCK web server and EADock DSS [97, 98] Spike protein fragment (331–524) is shown in red cartoon view, hesperidin molecule is shown in cyan and red stick view and human ACE2 is shown in blue cartoon view. **C** Docked complex structure of DNA with aflatoxin B1 exo-8,9-epoxide using AutoDock 4.0 [100, 101, 109]. Aflatoxin B1 exo-8,9-epoxide is shown in red stick view and DNA is shown in violet surface view [97, 101]. (Color figure online)

available phytochemicals among which hesperidin was found to be the best match after performing docking studies with several phytochemicals. Docking of the complex structure (SARS CoV2 spike protein fragment and ACE2) and different phytochemicals were executed with SWISSDOCK web server and constructed by EADock DSS [98]. The binding affinity of different docking structures of ACE2 and phytochemicals, in presence and absence of the SARS CoV2 spike protein fragment was calculated using Dockthor web server [99].

In absence of spike protein fragment, they have performed the docking study between hesperidin (ligand) and ACE2 protein (receptor) and the binding affinity was found to be − 9.167 kcal/mol. In the second scenario, they have carried out the docking study in presence of spike protein fragment, in which case the previously obtained docking result from ClusPro web server i.e., the organic mixture of SARS Cov2 spike protein and ACE2 protein was taken as receptor and the organic molecule hesperidin was considered as ligand and the binding affinity reduced to − 8.639 kcal/mol. The docking result found is shown in Fig. 6B. Comparison between the binding affinity obtained in these two docking studies in absence and presence of spike protein fragment concluded that, due to the presence of hesperidin molecule the bound complex structure of ACE2 and spike protein fragment gained instability and thus they predicted that hesperidin may be showing antiviral activity in SARS CoV2 infection. In their work, Ricci and Paulo have performed docking between aflatoxin B1 exo-8,9-epoxide, and DNA (PDB id: 1MKL) using AutoDock 4.0 (Fig. 6C) [100, 101].

To understand the possible allosteric inhibiting ability of fourteen laboratory-synthesized aromatic compounds (derived from 2,3-dihydroquinazolin-4(1H)-one compound and considered as ligands) to replace 3,8-diamino-5-ethyl-6-phenylphenanthridinium bromide (Ethidium Bromide or EB) (also referred as ligand) from the calf thymus DNA (CT-DNA, also called as 1-BNA dodecamer or B-DNA) (PBD ID: 1BNA) (taken as receptor), Kumar has performed the molecular docking study between the ligands and the CT-DNA receptor by AutoDock Vina tool using Lamarckian genetic algorithm [16, 95, 102]. Among 15 ligands, 2-(3-nitrophenyl)-2,3-dihydroquinazolin-4(1H)-one) (G13) is found to show better binding affinity than other ligands based on estimated minimum binding energy with hydrogen bonding interaction that helps to understand the potential intercalative DNA binding of the compound. The binding energy in the molecular docking study was − 8.6 and − 7.4 kJ/mol for G13 and EB compounds, respectively. One hydrogen bond was found between G13 compound and nucleotide numbered 22 (DG22, where G stands for Guanine) of CT-DNA with hydrogen bond length 2.224 Å. No hydrogen bond was found between EB compound and CT-DNA. Molecular docking analysis proves that G13 binds to the smaller groove (called as minor grooves) of CT-DNA more strongly than EB that explains the replacing EB by G13 from the minor groove of the DNA. The docking complex results of calf thymus DNA (CT-DNA) with ligands G13 and EB are shown in Fig. 7A, B, respectively.
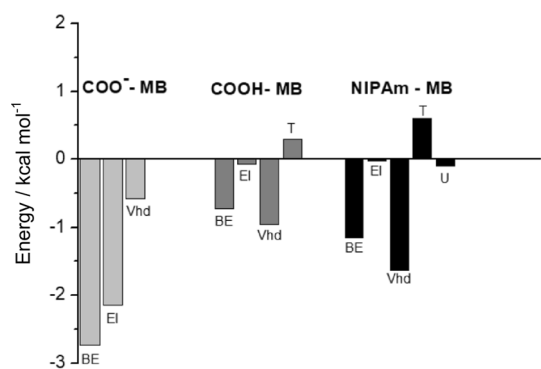
Fig. 7 **A** Docked complex result of calf thymus DNA (CT-DNA) and G13 **B** Docking complex result of CT-DNA and EB using AutoDock Vina molecular docking tool [95, 102]. The DNA is represented in in cartoon view and both the ligands (G13 and EB) are shown in sticks view. H-bond is shown in green dashed line [102]. (Color figure online)

Synthetic organic molecules-organic ligand docking case studies

Majority of the docking algorithms recognize biomolecules (e.g., DNA, RNA, protein and enzyme) exclusively as receptors. Other than the biomolecules, they do not support other molecules such as synthetic-organic, synthetic organic–inorganic hybrid, and inorganic molecules as receptors. In our recent work (ability of anionic polymerized hydrogel materials (PGM) (poly(*N*-isopropylacrylamide-*co*-acrylic acid) anionic microgels)) to separate cationic organic dye molecules, methylene blue (MB) from water has been studied [103]. The adsorption mechanism of methylene blue dye by organic polymer moieties on the PGM matrix has been understood through molecular docking simulation method using AutoDock 4.2 with simulated annealing algorithm [100, 104].

Docking has been performed between MB cationic dye (receptor) and constituting components i.e., monomeric units of the anionic microgel (ligands), such as carboxylate ion (COO⁻), the carboxylic acid (COOH), and *N*-isopropylacrylamide (NIPAM) to investigate the possible intermolecular interactions, like hydrogen bonding, electrostatic, van der Waals, desolvation, and torsional between them. The docking complex models between different pairs (COO⁻, MB), (COOH, MB), and (NIPAM, MB) found in this study

Fig. 8 Molecular docking results with least binding energy between methylene blue (MB) cation (recognized as receptor) and organic polymer moieties (identified as ligands), namely carboxylate ion (COO⁻) (deprotonated), carboxylic acid (COOH) (protonated), and *N*-isopropylacrylamide (NIPAM) using AutoDock 4.2 tool with simulated annealing algorithm [100, 103, 104]. The polymer moieties and MB cationic dye are represented in ball-and-stick models and the interatomic distances (within 3.5 Å) between them are displayed in yellow dotted lines. MB, COO⁻, COOH, and NIPAM are colored blue, gray, dark gray, and black, respectively [103]. (Color figure online)

Fig. 9 Column charts showing the total binding energy (BE), the contribution from different types of potential namely, electrostatics (EI), van der Waals-hydrophobic-desolvation (Vhd), tortional (T), and unbound (U) found in the docking result for the (COO⁻, MB), (COOH, MB), (NIPAM, MB) bound complexes [103]. (Color figure online)

are presented in Fig. 8. The net binding energy (BE) along with the contributed potentials due to electrostatic (EI), van der Waals-hydrophobic-desolvation (Vhd), tortional (T), and unbound (U) calculated for different docking complexes i.e., (COO⁻, MB), (COOH, MB), and (NIPAM, MB) are displayed with a column plot in Fig. 9. The total binding energy was found to be − 2.73 kcal/mol for the docking complex between MB cation and carboxylate anion (COO⁻), which is maximum compared to other docking complexes, such as (COOH, MB) and (NIPAM, MB).

Organic–inorganic

In this case, the receptor is organic type, and the ligand is inorganic type. The interaction between enzyme-inorganic substrate, enzyme-inorganic drug, protein-inorganic drug, protein-inorganic ligand, protein-metal atom or ion, protein-metal nanoparticle, DNA/RNA-inorganic drug, DNA/RNA-metal atom or ion, DNA/RNA-metal nanoparticle, polymer moiety-metal atom or ion come under the category of organic–inorganic intermolecular docking study.

Molecular docking of ligands having metal atoms or ions (e.g., metal nanoparticle) is challenging though the existing force fields do not support bigger size nanoparticles (having many metal atoms) and all types of metal atoms. Some studies, however, have been carried out using Brownian dynamics (BD) rigid-body docking, AutoDock, AutoDock Vina, and Hex software and are discussed here [55, 95, 100, 105, 106].

Biomolecules-inorganic ligand docking case studies

Aghili Z has predicted the molecular interaction between hen egg lysozyme protein and water-coated iron nanoparticle (Fe NP) by docking study using Hex 6.3 software tool with net binding energy − 230.92 kJ/mol as shown in Fig. 10A [105, 107]. In his work, he found the hydrogen bonding formation between water molecules surrounding the Fe NP and polar residues on the surface of lysozyme protein.

Molecular docking study between trimeric structure of VP6 protein and palladium ions (Pd(II)) obtained from PdCl$_3$(H$_2$O)⁻ at pH 5 has been demonstrated in Fig. 10B [108]. The study predicts the existence of binding between
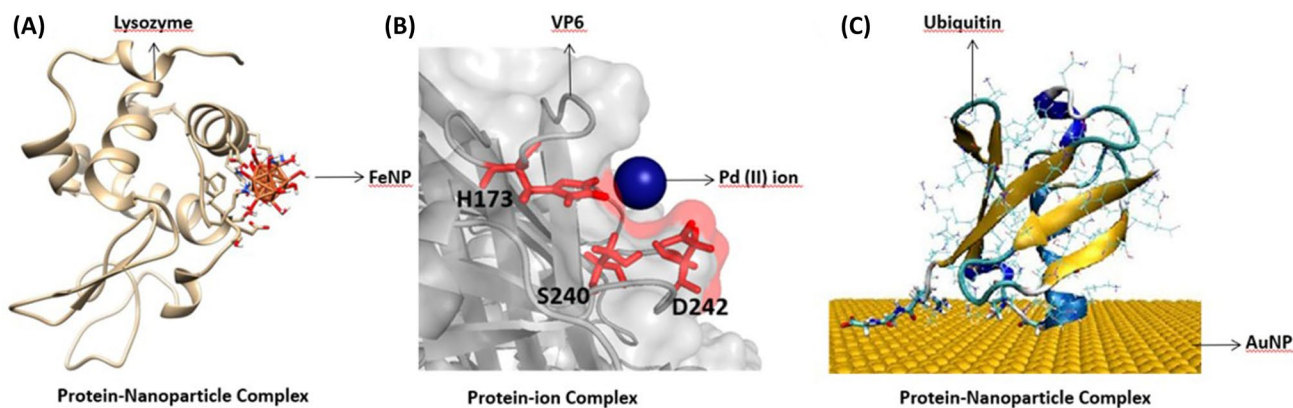

Fig. 10 A Docked complex structure of Hen egg white lysozyme (HEWL) protein in gold color cartoon view with water-coated iron nanoparticle (FeNP) in orange stick view using Hex 6.3 docking software tool [105, 107]. **B** Docked complex structure of VP6 trimer protein in grey cartoon and surface view with Pd(II) ion in blue sphere view at pH 5.0 using AutoDock 4.0 [100, 108]. The metal binding site is shown in red surface view containing polar residues His173, Ser240, and Asp242 viewed in red stick representation. **C** Docked complex structure of ubiquitin protein with bare neutral gold nano particle (AuNP) using Brownian dynamics (BD) rigid-body docking [106, 186]. The backbone of protein is shown in cartoon view and the residues those are in contact with the Au surface are shown in stick view and the rest of the atoms are shown in line view. The Au surface is shown in gold color net structure [107, 108, 186]. (Color figure online)

VP6 protein and Pd(II) ions which further helps in nucleation, growth, and stability in palladium nanoclusters formation. AutoDock 4.0 tool has been used to estimate the metal binding site with residues His173, Ser240, and Asp242 of VP6 for Pd(II) ions and binding energy (-2.0 kcal/mol at pH 5). Pd(II) ions were found to interact with charged and polar residues, including Asp, His, Ser, Gln, and Asn which results in hydrogen bond formation between Pd(II) ions and protonated side chains of Ser and Asp residues, Pd(II) ions and nitrogen atom of His173. Different types of interaction forces, like hydrogen bond formation between Pd(II) ions and polar residues, electrostatic force of attraction between Pd(II) ions and charged residues in the binding site of VP6 protein favor the stabilization and nucleation growth of Pd nanoparticle.

Brancolini G. in their work have used Brownian dynamics rigid-body docking simulation to understand the interaction between ubiquitin protein taken as rigid body and naked neutral gold nanoparticle (Au(111) NP) surface using protein surface docking method with SDA 6.0 as shown in Fig. 10C [106, 108]. The residues of ubiquitin that make good contact at distances ≤ 3 Å with Au(111) surface are found to be GLY35, PRO37, ARG74, GLY75, and GLY76. The RMSD value of docked complex structure with respect to the initial representative complex structure was found to be -1.31 Å. The total interaction energy of the resultant docked complex was found to be -29.12 kcal/mol at 300 K.

Synthetic organic molecules-inorganic ligand docking case studies

In our recent work, we have performed the molecular docking simulation study using the AutoDock 4.2.6 tool between different metal ions (Au^{3+}, Ag^+, Fe^{2+}) and atoms (Au^0, Ag^0, Fe^0) with polymer moieties (NIPAM, carboxylate ion (COO^-)) to understand the molecular interaction which binds the nanoparticle with the polymer matrix and results in stable microgel metal hybrids as shown in Fig. 11 [27, 49, 100, 104, 109–111].

The simulated annealing algorithm was used for all types of dockings. Ions were found to interact with COO^- more prominently than atoms. The net binding energy was found to be -9.28, -5.7, -1.46 kcal/mol for interaction between COO^- (taken as a receptor) and ligands, namely Fe^{2+}, Au^{3+}, and Ag^+, respectively. The electrostatic potential contributes the most towards the total binding energy among all types of potentials (van der Waals, hydrophobic, solvation, torsional, and electrostatic). Different types of potentials with the net binding energy found in docking studies for different complexes of polymer moieties (COO^-, NIPAM) with ions
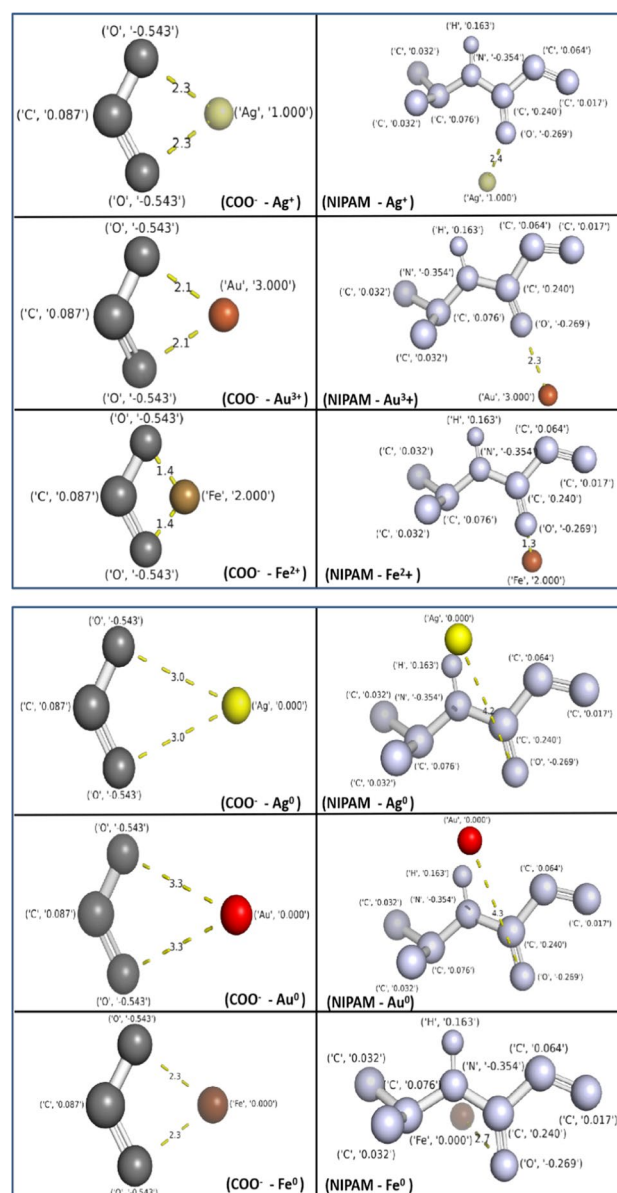

Fig. 11 Best docking results with minimum binding energy after docking study between polymer moieties (COO^-, NIPAM) with ions (Ag^+, Au^{3+}, and Fe^{2+}) and atoms (Ag^0, Au^0, and Fe.^{0}), respectively [111]. The polymer moieties, ions and atoms are represented in ball-and-stick, sphere, and sphere views, respectively. The distances between randomly chosen neighboring atoms of polymer moieties and ions are shown in yellow dashed line labelled in Angstrom unit (Å) [111]. (Color figure online)

(Ag^+, Au^{3+}, and Fe^{2+}) and atoms (Ag^0, Au^0, and Fe^0) are demonstrated using column charts in Fig. 12.

Hybrid receptor/ligand

From the above classification we understand the interacting molecules can be organic or inorganic. The macromolecule and/or ligand can be an organic–inorganic hybrid molecule,
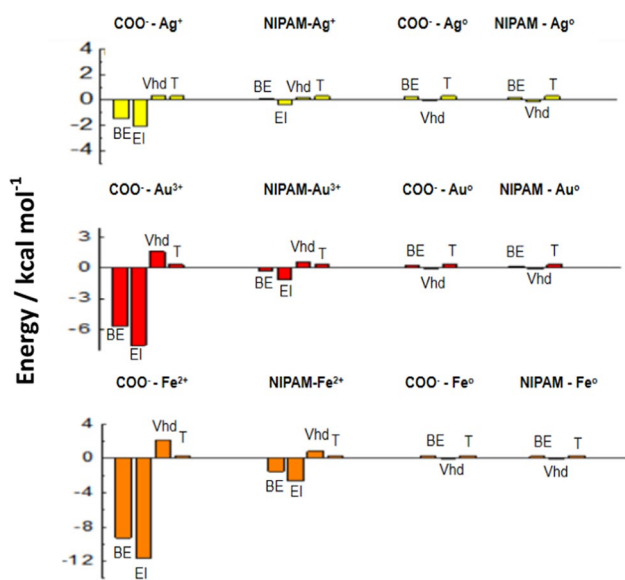
Fig. 12 Column charts that represent the contribution from different types of potential such as electrostatic (EI), van der Waals-hydrophobic (Vhd), and torsional(T) towards the net binding energy (BE) in docking studies between polymer moieties (COO⁻, NIPAM) with metal ions (Ag^+, Au^{3+}, and Fe^{2+}) and atoms (Ag^0, Au^0, and Fe^0) [111]. (Color figure online)
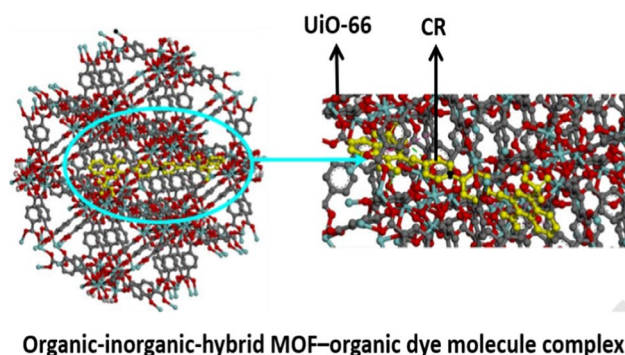

Organic-inorganic-hybrid MOF—organic dye molecule complex

Fig. 13 Organic–inorganic-hybrid MOF—organic dye molecule complex [112]. Complex structure of congo red dye (CR) (yellow color) with UiO-66 (a Zr (IV) based metal organic framework (MOF)) (gray and red) using ball-and-stick representation after docking study using AutoDock Vina [55, 95]. The active site is shown in cyan color ellipse [112]. (Color figure online)

such as metal–organic framework (MOF). MOFs are new advanced synthetic hybrid crystalline materials and can be synthesized from metal ions and organic linkers by coordination bonds. The adsorption property of MOF can be used to detect and remove organic pollutants, including dye, a drug from different environmental water samples. Different case studies are discussed below.

Due to the adsorption behavior of UiO-66 (a Zirconium (Zr) based metal organic framework (MOF)) it can be used to remove pollutants, like anionic dye molecules Congo

Organic-inorganic-hybrid MOF– organic dye molecule complex
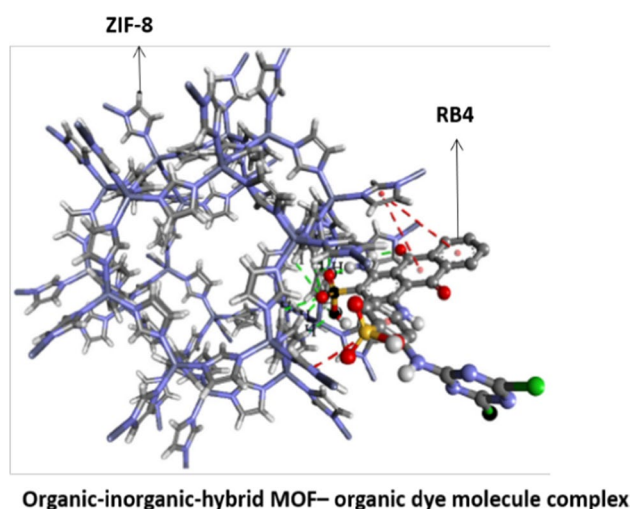
Fig. 14 Docked complex structure of zeolitic imidazole framework-8 (ZIF-8), a special class of metal organic framework (MOF) with organic molecule reactive blue-4 (RB4) ions using docking study with AutoDock Vina software [55, 95, 113]. ZIF-8 (receptor) and RB4 (ligand) are represented in stick and ball-and-stick models, respectively. H-bonds are represented with green and π–π interactions are shown with red dashed lines in the active site of ZIF-8 [113]. (Color figure online)

red (CR) from water [112]. To understand the absorption of Congo red dye by UiO-66, Panda et al. have performed molecular docking study between Congo red dye (CR) and UiO-66 by AutoDock Vina tool using Lamarckian genetic algorithm (LGA) and the docked complex result found in the docking study is shown in Fig. 13 [27, 55, 95]. The binding energy was found to be − 12.82 kcal/mol and the molecular interaction was because of two hydrogen bonds formed with bond length 2.68 and 2.66 Å and three hydrophobic interactions with bond length 4.89, 5.46, and 5.75 Å.

Panda J et al. have demonstrated that ZIF-8 MOF can be used to remove reactive blue-4 (RB4), a reactive triazine toxic dye from water because of the absorption property of ZIF-8 MOF for RB4 which helps in water treatment [113]. They have carried out a docking study between ZIF-8 MOF (receptor) and RB4 (ligand) using AutoDock Vina to understand the absorption process computationally and the docking result is shown in Fig. 14. The net binding energy was found to be a negative value (− 22.34 kcal/mol) which means the molecular interaction is an exothermic adsorption process. The favorable binding interactions were due to ten H-bonding having a bond length ranging from 1.87 to 2.8 Å between oxygen atom (O) of –SO_3 group in RB4 and hydrogen atoms (H) of –CH_3 group in ZIF-8 MOF and three π–π interactions having bond length 4.78, 5.2, and 5.8 Å.

Lu et al. in their work have used the molecular docking method using AutoDock 4 tool to investigate the
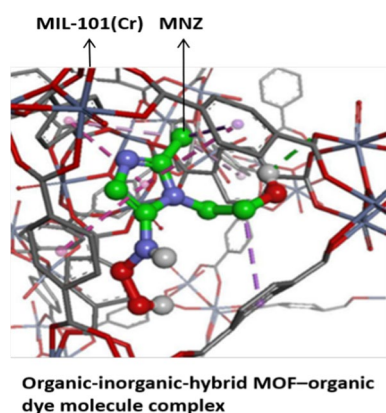
Fig. 15 Docked complex structure of porous chromium terephthalate, MIL-101(Cr) and metronidazole (MNZ) using AutoDock 4 [114]. The MIL-101(Cr) (receptor) and MNZ (ligand) are viewed in stick and ball-and-stick models, respectively. For ligand, carbon, oxygen, sulfur, nitrogen, and hydrogen atoms are marked in green, red, yellow, blue, and orchid, respectively. For receptor, oxygen atoms and Cr ions are marked in red and orchid, respectively. The π–π interactions between imidazole rings of MIL-101(Cr) and MNZ, and hydrogen bonds are represented by pink and green dashed lines respectively in the binding site of MIL-101(Cr) [114]. (Color figure online)

intermolecular interactions between porous chromium terephthalate, MIL-101(Cr) and widely used antibiotics named 5-nitroimidazoles (5-NDZs) to understand the adsorption mechanism [114]. MIL-101(Cr) is an organic–inorganic-hybrid metal organic framework (MOF) material constructed from metal ions and organic linkers and can be used for the trace analysis of organic pollutants 5-NDZs in different water samples. Among five antibiotics, metronidazole (MNZ) is chosen here to show the molecular interactions. The docking study using the Lamarckian genetic algorithm (LGA) have been carried out between MNZ (ligand) and MIL-101(Cr) (receptor), the docked complex is demonstrated in Fig. 15 and the calculated binding energy by Auto-Dock tool was found to be − 6.21 kcal/mol. The net binding energy is mostly by the electrostatic force of attraction and the other types of potentials that contribute to binding energy are the formation of hydrogen bonds, coordination bonds and different types of interactions, namely hydrophobic, π–π, and van der Waals. The contributions from all such interaction potentials contribute to the extraction property of MIL-101(Cr) MOF.

A highly porous zirconium metal–organic framework (MOF), such as UiO-66 (Universitetet i Oslo), which is composed of $[Zr_6O_4(OH)_4]$ clusters and 1,4-benzodicarboxylic acid struts (University of Liverpool, ChemTube3D) can be used as an adsorbent material. Adsorption of polycyclic aromatic hydrocarbon (PAH) (recognized as a hazardous and toxic pollutant), such as chrysene (CRY, Molecular formula: $C_{18}H_{12}$) (a product of coal tar) by UiO-66(Zr) helps in wastewater treatment by removal of the pollutant [115]. To
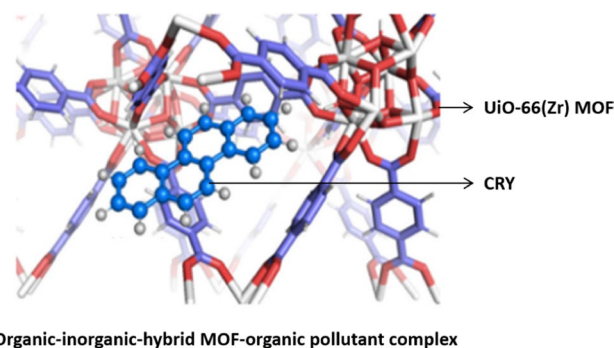

Fig. 16 Docked complex structure of UiO-66(Zr) metal organic framework (MOF) and chrysene (CRY), a toxic and hazardous polycyclic aromatic hydrocarbon (PAH) pollutant using AutoDock 4.2 [115]. UiO-66(Zr) MOF (receptor) and CRY (ligand) are shown in stick and ball-and-stick models, respectively. UiO-66(Zr) is marked in white, violet, and red color. Four benzene rings of CRY are marked in blue and light gray color [115]. (Color figure online)

understand the mechanism and driving forces of the adsorption process of CRY onto UiO-66(Zr) MOF surface, Zango and his co-workers have implemented molecular docking using AutoDock 4.2 Suite with Lamarckian genetic algorithm (LGA) as shown in Fig. 16 and the predicted binding energy was found to be − 2.32 kcal/mol, that indicates a stable absorption because of potential interactions on the surface.

The calculate energy due to different types of interaction potentials, such as van der Waals, hydrogen bonding and desolvation was − 3.1 kcal/mol and electrostatic was 0.79 kcal/mol. CRY was found to show electrostatic interaction with Zr^{4+} ion of the MOF.
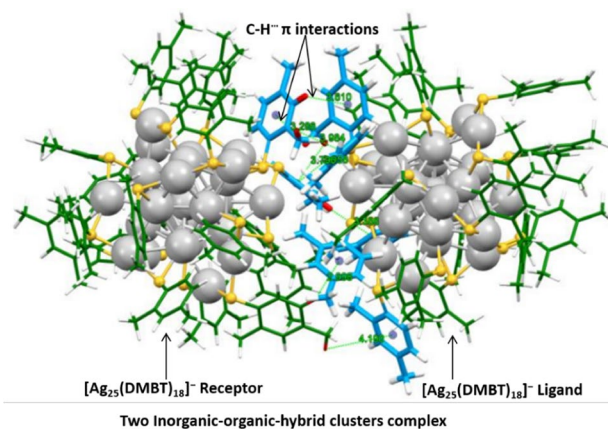

Fig. 17 Docked complex structure of two $[Ag_{25}(DMBT)_{18}]^-$ clusters using AutoDock 4.2 [100, 115]. Complex is shown in ball-and-stick model. Silver (Ag) and sulfur (S) atoms are shown in gray and yellow, respectively. C–H··· π interactions are viewed in green dotted lines. Hydrogen (H) atoms and benzene rings associated with these interactions are viewed in red and blue, respectively. Other benzene rings not associated with these interactions are shown in green [116]. (Color figure online)

To understand the cause of instant isotopic exchange reaction in silver nanoparticles cluster, Chakraborty et al. have carried out molecular docking study between two $[Ag_{25}(DMBT)_{18}]^-$ (DBMT for 2,4-dimethylbenzenethiol, which acts as a protecting ligand) clusters using Auto-Dock 4.2 with Lamarckian genetic algorithm [100, 116]. $[Ag_{25}(DMBT)_{18}]^-$ has been used both as receptor and ligand in the docking process. The binding energy was found to be -23.7 kcal/mol. The docking result with least binding energy is shown in Fig. 17.

The fluorescent cobalt oxide (CoO) umbelliferone nano-conjugate having anti-cancer activity can be used both as a drug and carrier. Ali et al. have conducted molecular docking studies applying the protein docking program, HEX 8.0.0 using Spherical Polar Fourier Correlations technique to find the binding interactions of the CoO-drug nanoconjugate with the B-DNA dodecamer (also called as Calf thymus DNA (CT-DNA)), a DNA duplex having the sequence (CGC GAATTCGCG)$_2$ (PDB ID: 1BNA) and human serum albumin (HSA) (PDB ID: 1H9Z) protein separately [117–119]. The most stable docking results with least binding energies of the CoO-drug nanoconjugate against the DNA molecule and interactions in the binding cavity of the CT-DNA are shown in Fig. 18A, B. The docking study predicted the
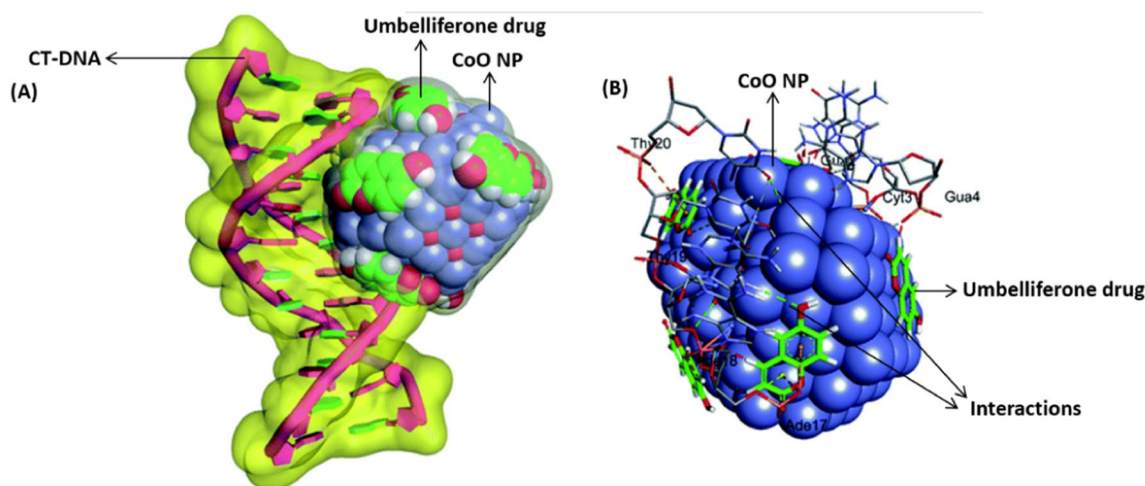

Fig. 18 A Molecular docking complex of calf thymus DNA (CT-DNA) (PDB ID: 1BNA) and cobalt oxide (CoO) umbelliferone drug nanoconjugate using HEX 8.0.0 [117, 118]. DNA is represented in carton and surface view. The drug nanoconjugate, a mixture of CoO nanoparticle (blue and red sphered view) and umbelliferone drug (green, red, and white sphered view). **B** Non-covalent interactions of DNA bases with the nanoconjugate depicted by dashed lines [117]. (Color figure online)
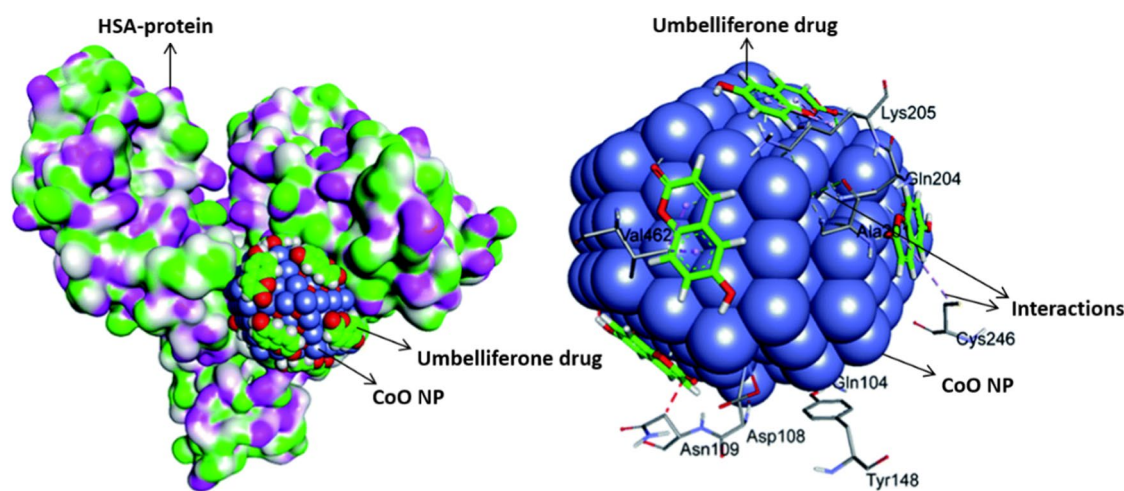

Fig. 19 A Docking result of human serum albumin (HSA) with the CoO-umbelliferone drug nanoconjugate by protein docking program, HEX 8.0.0 using Spherical Polar Fourier Correlations technique [117–119]. **B** Binding interactions of noncovalent type between neighboring amino acid residues in the binding pocket of HSA and the drug nanoconjugate shown in dashed lines [117]. (Color figure online)

existence of electrostatic, hydrophobic and hydrogen bonding intermolecular noncovalent interactions between them. Similarly, the most stable bound conformation of the docking result of the CoO-drug nanoconjugate and HSA protein with binding site interactions, such as hydrophobic, hydrogen bonding and metal acceptor are shown in Fig. 19A, B.

Inorganic–inorganic

In this case, both receptor and ligand in docking study are inorganic. The interaction study between any two chemical molecules that do not have any carbon-hydrogen bond comes under this category. To the best of our knowledge, there has not been any study carried out on inorganic receptor and ligand docking till now.

Challenges in molecular docking

There are many limitations and challenges in docking techniques [15]. A docking result predicted may not be accurate and match the result found in an experimental approach.

(i) Recognition of different types of molecular features which contribute to the interaction between them can be complex, difficult to understand and time-consuming to simulate in a computer. The docking technique tries to incorporate additional complexity in each step.

(ii) Number of conformational degrees of freedom of molecules makes it difficult and time-consuming to pose a ligand in the binding pocket of a macromolecule. A complex effective scoring function helps in this perspective.

(iii) Low resolution crystallographic structural information of molecules and molecules with unknown structural information makes it more difficult to find the correct binding pockets of receptors.

(iv) Flexibility of molecules and changes in the geometry of molecules that occur during the binding also complicate the finding.

(v) Hydrated docking is not possible always by most of the docking tools and thus involvement of water molecules in a macromolecule-ligand interaction is difficult to predict. Removing water molecules from the binding pocket of receptor before docking process is usually done by docking algorithms. But this prepossessing of receptor is not universally correct in many cases, such as if the water molecules are tightly bound or they are functionally active in the binding-site. In such cases hydrated docking is very much essential.

(vi) Existing force fields do not support all types of atoms. Finding the force field and some other required parameters is a challenging task in docking.

(vii) Most of the docking software tools accept only biomolecules, such as DNA, RNA, protein, enzyme, etc. as receptors. But other than biomolecules they do not recognize synthetic-organic, synthetic organic–inorganic hybrid, and inorganic molecules as receptors.

(viii) Also, for ligand, they accept a few metal atoms but not all. Similarly, few docking tools, including Auto-Dock, Hex, PatchDock recognize nanoparticles as ligands but many docking tools do not recognize all types of nanoparticles, like gold (Au), silver (Ag), iron (Fe), etc. [100, 105, 120]. Even a bigger size ligand having numerous atoms, such as metal nanoparticle having hundreds, or a greater number of atoms cannot be processed by a docking algorithm.

(ix) Ions are also supported by a few docking tools. When they are taken as ligands the charge gets neutralized automatically by the software. So, the user must manually check the charges carefully.

(x) Number of atoms and torsions of a molecule are also restricted in many docking tools.

Concluding remarks

From the above study, we conclude that although there are numerous docking tools and web servers available, they are specific to type of molecule to be accepted as input as receptor or ligand. That is why interaction studies of biomolecules receptors, namely protein, DNA, and RNA, etc. with organic, inorganic, or hybrid (natural or synthetic organic–inorganic mixture) ligands are commonly seen. Looking into the challenges and limitations we would say the best docking algorithm is probably the one made by a hybrid of many algorithms which will have novel search, scoring function, force field for all types of atoms, allowance of unlimited number of atoms and rotatable bonds (torsions) in a molecule etc. We have shown here some case studies in synthetic organic–inorganic, metal–organic hybrid-organic systems. Inorganic-inorganic docking studies are still missing which can be seen in future hopefully with better force fields and required parameters.

References

1. Dar AM, Mir S (2017) J Anal Bioanal Tech 8:356
2. Abdelsattar AS, Dawoud A, Helal MA (2021) Nanotoxicology 15:66
3. Gupta M, Sharma R, Kumar A (2018) Comput Biol Chem 76:210
4. Morris GM, Lim-Wilby M (2008) Methods Mol Biol 443:365
5. Meng X-Y, Zhang H-X, Mezei M, Cui M (2011) Curr Comput Aided Drug Des 7:146
6. Chen Y-C (2015) Trends Pharmacol Sci 36:78
7. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ (2003) Proteins 52:2
8. Yang C, Chen EA, Zhang Y (2022) Molecules 27:4568
9. Liu J, Wang R (2015) J Chem Inf Model 55:475
10. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) J Comput Aided Mol Des 11:425
11. Chen T, Guestrin C (2016) XGBoost. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, p 785
12. Cortes C, Vapnik V (1995) Mach Learn 20:273
13. Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, Wang L, Li C, Sun M (2020) AI Open 1:57
14. Schmidhuber J (2015) Neural Netw 61:85
15. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Nat Rev Drug Discov 3:935
16. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) J Comput Chem 19:1639
17. Mehler EL, Solmajer T (1991) Protein Eng 4:903
18. Goodford PJ (1985) J Med Chem 28:849
19. Morris GM, Goodsell DS, Huey R, Olson AJ (1996) J Comput Aided Mol Des 10:293
20. Bohm H-J (1994) J Comput Aided Mol Des 8:243
21. Haque R, Mohammad M, Islam M (2018) Mod Approach Drug Des 1:518
22. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, Christie CH, Dalenberg K, Di Costanzo L, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK, Guranović V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H, Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C, Tao YP, Voigt M, Westbrook JD, Young JY, Zardecki C, Zhuravleva M (2021) Nucleic Acids Res 49:D437
23. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) J Cheminform 4:17
24. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Nature 596:583
25. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283
26. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25:1605
27. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ (2016) Nat Protoc 11:905
28. Schiebel J, Gaspari R, Wulsdorf T, Ngo K, Sohn C, Schrader TE, Cavalli A, Ostermann A, Heine A, Klebe G (2018) Nat Commun 9:3559
29. Barillari C, Taylor J, Viner R, Essex JW (2007) J Am Chem Soc 129:2577
30. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) J Med Chem 47:1739
31. Forli S, Olson AJ (2012) J Med Chem 55:623
32. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS (2002) Proteins 46:34
33. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267:727
34. Huang Z, Wong CF (2009) J Phys Chem B 113:14343
35. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) J Mol Biol 285:1735
36. Bas DC, Rogers DM, Jensen JH (2008) Proteins 73:765
37. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID (2009) RNA 15:1219
38. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) J Comput Aided Mol Des 27:221
39. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE (2021) Nucleic Acids Res 49:D1388
40. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J (2006) Nucleic Acids Res 34:D668
41. Li Z, Wan H, Shi Y, Ouyang P (2004) J Chem Inf Comput Sci 44:1886
42. Hunter AD (1997) J Chem Educ 74:905
43. Pence HE, Williams A (2010) J Chem Educ 87:1123
44. Ihlenfeldt WD, Bolton EE, Bryant SH (2009) J Cheminform 1:20
45. Dessailly BH, Lensink MF, Orengo CA, Wodak SJ (2008) Nucleic Acids Res 36:D667
46. Dhakal A, McKay C, Tanner JJ, Cheng J (2022) Brief Bioinform 23:bbab476
47. Meng EC, Shoichet BK, Kuntz ID (1992) J Comput Chem 13:505
48. Makino S, Kuntz ID (1997) J Comput Chem 18:1812
49. Goodsell DS, Morris GM, Olson AJ (1996) J Mol Recognit 9:1
50. Muegge I, Martin YC (1999) J Med Chem 42:791
51. Gohlke H, Hendlich M, Klebe G (2000) J Mol Biol 295:337
52. DeWitte RS, Shakhnovich EI (1996) J Am Chem Soc 118:11733
53. Huang S-Y, Zou X (2006) J Comput Chem 27:1876
54. Zheng Z, Merz KM (2013) J Chem Inf Model 53:1073
55. Trott O, Olson AJ (2010) J Comput Chem 31:455
56. Yang C, Zhang Y (2021) J Chem Inf Model 61:4630
57. Rarey M, Kramer B, Lengauer T, Klebe G (1996) J Mol Biol 261:470
58. Wang R, Liu L, Lai L, Tang Y (1998) J Mol Model 4:379
59. Rognan D, Lauemoller SL, Holm A, Buus S, Tschinke V (1999) J Med Chem 42:4650
60. Wang R, Lai L, Wang S (2002) J Comput Aided Mol Des 16:11
61. Stanzione F, Giangreco I, Cole JC (2021) Prog Med Chem 60:273
62. Kroemer RT, Vulpetti A, McDonald JJ, Rohrer DC, Trosset JY, Giordanetto F, Cotesta S, McMartin C, Kihlén M, Stouten PF (2004) J Chem Inf Comput Sci 44:871
63. Plewczynski D, Łaźniewski M, Augustyniak R, Ginalski K (2011) J Comput Chem 32:742
64. Onodera K, Satou K, Hirota H (2007) J Chem Inf Model 47:1609
65. Perola E, Walters WP, Charifson PS (2004) Proteins 56:235
66. Seeliger D, de Groot BL (2010) J Comput Aided Mol Des 24:417
67. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14:33
68. Li Y, Su M, Liu Z, Li J, Liu J, Han L, Wang R (2018) Nat Protoc 13:666
69. Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, Wang R (2019) J Chem Inf Model 59:895
70. Ashtawy HM, Mahapatra NR (2012) IEEE/ACM Trans Comput Biol Bioinform 9:1301
71. Sunseri J, Koes DR (2021) Molecules 26:7369

72. Liu S, Alnammi M, Ericksen SS, Voter AF, Ananiev GE, Keck JL, Hoffmann FM, Wildman SA, Gitter A (2019) J Chem Inf Model 59:282
73. Tran-Nguyen V-K, Jacquemard C, Rognan D (2020) J Chem Inf Model 60:4263
74. Alonso H, Bliznyuk AA, Gready JE (2006) Med Res Rev 26:531
75. De Vivo M, Masetti M, Bottegoni G, Cavalli A (2016) J Med Chem 59:4035
76. Salmaso V, Moro S (2018) Front Pharmacol 9:923
77. Fan J, Fu A, Zhang L (2019) Quant Biol 7:83
78. Pagadala NS, Syed K, Tuszynski J (2017) Biophys Rev 9:91
79. Tripathi A, Bankaitis VA (2017) J Mol Med Clin Appl 2:19
80. Fischer E (1894) Ber Dtsch Chem Ges 27:2985
81. Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A (2002) J Med Chem 45:2469
82. Koshland DE (1958) Proc Natl Acad Sci USA 44:98
83. Teague SJ (2003) Nat Rev Drug Discov 2:527
84. Brylinski M, Skolnick J (2008) Proteins 70:363
85. Lexa KW, Carlson HA (2012) Q Rev Biophys 45:301
86. Nabuurs SB, Wagener M, de Vlieg J (2007) J Med Chem 50:6507
87. Land H, Humble MS (2018) Methods Mol Biol 1685:43
88. Bitencourt-Ferreira G, de Azevedo WF (2019) Methods Mol Biol 2053:149
89. Mashiach E, Nussinov R, Wolfson HJ (2010) Nucleic Acids Res 38:W457
90. Trellet M, Melquiond ASJ, Bonvin AMJJ (2013) PLoS ONE 8:e58769
91. Monod J, Wyman J, Changeux JP (1965) J Mol Biol 12:88
92. Pinzi L, Rastelli G (2019) Int J Mol Sci 20:4331
93. Torres PHM, Sodero ACR, Jofily P, Silva FP Jr (2019) Int J Mol Sci 20:4574
94. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G (2018) J Chem Inf Model 58:287
95. Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) J Chem Inf Model 61:3891
96. Alekseenko A, Ignatov M, Jones G, Sabitova M, Kozakov D (2020) Methods Mol Biol 2165:157
97. Basu A, Sarkar A, Maulik U (2020) Sci Rep 10:17699
98. Grosdidier A, Zoete V, Michielin O (2011) Nucleic Acids Res 39:W270
99. Guedes IA, Costa LSC, Dos Santos KB, Karl ALM, Rocha GK, Teixeira IM, Galheigo MM, Medeiros V, Krempser E, Custódio FL, Barbosa HJC, Nicolás MF, Dardenne LE (2021) Sci Rep 11:5543
100. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) J Comput Chem 30:2785
101. Ricci CG, Netz PA (2009) J Chem Inf Model 49:1925
102. Kumar G, Mogha NK, Masram DT (2021) ACS Appl Nano Mater 4:2682
103. Pany B, Ghosh Majundar A, Mohanty M, Fyis KP, Dey T, Tripathy G, Bhat S, Yamanaka J, Mohanty PS (2023) J Mol Liq 375:121267
104. Goodsell DS, Olson AJ (1990) Proteins 8:195
105. Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW (2010) Nucleic Acids Res 38:W445
106. Kokh DB, Corni S, Winn PJ, Hoefling M, Gottschalk KE, Wade RC (2010) J Chem Theory Comput 6:1753
107. Aghili Z, Taheri S, Zeinabad HA, Pishkar L, Saboury AA, Rahimi A, Falahati M (2016) PLoS ONE 11:e0164878
108. Carreño-Fuentes L, Bahena D, Palomares L, Ramírez O, José-Yacamán M, Plascencia-Villa G (2016) Metals 6:200
109. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) J Comput Chem 28:1145
110. Ravindranath PA, Forli S, Goodsell DS, Olson AJ, Sanner MF (2015) PLoS Comput Biol 11:e1004586
111. Bhol P, Mohanty M, Mohanty PS (2021) J Mol Liq 325:115135
112. Panda J, Sahu SN, Sahoo JK, Biswal SP, Pattanayak SK, Samantaray R, Sahu R (2020) Environ Nanotechnol Monit Manag 14:100340
113. Panda J, Sahoo JK, Panda PK, Sahu SN, Samal M, Pattanayak SK, Sahu R (2019) J Mol Liq 278:536
114. Lu N, Wang T, Zhao P, Zhang L, Lun X, Zhang X, Hou X (2016) Anal Bioanal Chem 408:8515
115. Zango ZU, Sambudi NS, Jumbri K, Bakar NHHA, Abdullah NAF, Negim EM, Saad B (2020) Chem Eng Sci 220:115608
116. Chakraborty P, Nag A, Natarajan G, Bandyopadhyay N, Paramasivam G, Panwar MK, Chakrabarti J, Pradeep T (2019) Sci Adv 5:eaau7555
117. Ali MS, Tabassum S, Al-Lohedan HA, Farah M, Al-Anazi K, Usman M (2019) RSC Adv 9:26503
118. Ritchie DW, Venkatraman V (2010) Bioinformatics 26:2398
119. Ritchie DW, Kemp GJ (2000) Proteins 39:178
120. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) Nucleic Acids Res 33:W363
121. Velankar S, Alhroub Y, Alili A, Best C, Boutselakis HC, Caboche S, Conroy MJ, Dana JM, van Ginkel G, Golovin A, Gore SP, Gutmanas A, Haslam P, Hirshberg M, John M, Lagerstedt I, Mir S, Newman LE, Oldfield TJ, Penkett CJ, Pineda-Castillo J, Rinaldi L, Sahni G, Sawka G, Sen S, Slowley R, da Silva AWS, Suarez-Uruena A, Swaminathan GJ, Symmons MF, Vranken WF, Wainwright M, Kleywegt GJ (2011) Nucleic Acids Res 39:D402
122. Kinjo AR, Bekker GJ, Suzuki H, Tsuchiya Y, Kawabata T, Ikegawa Y, Nakamura H (2017) Nucleic Acids Res 45:D282
123. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2007) Nucleic Acids Res 36:D402
124. Lawson CL, Patwardhan A, Baker ML, Hryc C, Garcia ES, Hudson BP, Lagerstedt I, Ludtke SJ, Pintilie G, Sala R, Westbrook JD, Berman HM, Kleywegt GJ, Chiu W (2016) Nucleic Acids Res 44:D396
125. Madej T, Addess KJ, Fong JH, Geer LY, Geer RC, Lanczycki CJ, Liu C, Lu S, Marchler-Bauer A, Panchenko AR, Chen J, Thiessen PA, Wang Y, Zhang D, Bryant SH (2012) Nucleic Acids Res 40:D461
126. wwPDB consortium (2019) Nucleic Acids Res 47:D520
127. Coimbatore Narayanan B, Westbrook J, Ghosh S, Petrov AI, Sweeney B, Zirbel CL, Leontis NB, Berman HM (2014) Nucleic Acids Res 42:D114
128. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M (2007) Nucleic Acids Res 36:D344
129. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) Nucleic Acids Res 40:D1100
130. Brumfiel G (2008) Nature 453:139
131. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) J Chem Inf Model 52:1757
132. Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) J Med Chem 48:4111
133. Yang J, Roy A, Zhang Y (2012) Nucleic Acids Res 41:D1096
134. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) Nucleic Acids Res 35:D198
135. Ahmed A, Smith RD, Clark JJ, Dunbar JB Jr, Carlson HA (2015) Nucleic Acids Res 43:D465
136. McGuffin LJ, Adiyaman R, Maghrabi AHA, Shuid AN, Brackenridge DA, Nealon JO, Philomina LS (2019) Nucleic Acids Res 47:W408
137. Jacobson MP, Friesner RA, Xiang Z, Honig B (2002) J Mol Biol 320:597

138. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA (2004) Proteins 55:351
139. Veit M, Gadalla MR, Zhang M (2022) Int J Mol Sci 23:13209
140. Wang S, Peng J, Ma J, Xu J (2016) Sci Rep 6:18962
141. Wang S, Li W, Liu S, Xu J (2016) Nucleic Acids Res 44:W430
142. Grünberg R, Nilges M, Leckner J (2007) Bioinformatics 23:769
143. Lambert C, Léonard N, De Bolle X, Depiereux E (2002) Bioinformatics 18:1250
144. Buß O, Rudat J, Ochsenreither K (2018) Comput Struct Biotechnol J 16:25
145. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) Nat Protoc 10:845
146. Söding J, Biegert A, Lupas AN (2005) Nucleic Acids Res 33:W244
147. Webb B, Sali A (2014) Methods Mol Biol 1137:1
148. Adhikari B, Cheng J (2018) BMC Bioinform 19:22
149. Kim DE, Chivian D, Baker D (2004) Nucleic Acids Res 32:W526
150. Jayaram B, Dhingra P, Mishra A, Kaushik R, Mukherjee G, Singh A, Shekhar S (2014) BMC Bioinform 15(Suppl 16):S7
151. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) Nucleic Acids Res 46:W296
152. Jin S, Contessoto VG, Chen M, Schafer NP, Lu W, Chen X, Bueno C, Hajitaheri A, Sirovetz BJ, Davtyan A, Papoian GA, Tsai MY, Wolynes PG (2020) Nucleic Acids Res 48:W25
153. Du Z, Su H, Wang W, Ye L, Wei H, Peng Z, Anishchenko I, Baker D, Yang J (2021) Nat Protoc 16:5634
154. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) Nat Methods 12:7
155. Das R, Qian B, Raman S, Vernon R, Thompson J, Bradley P, Khare S, Tyka MD, Bhat D, Chivian D, Kim DE, Sheffler WH, Malmström L, Wollacott AM, Wang C, Andre I, Baker D (2007) Proteins 69:118
156. Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB, Minor W (2017) Acta Crystallogr D Struct Biol 73:223
157. Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, Minor W (2014) Nat Protoc 9:156
158. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB 3rd, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC (2018) Protein Sci 27:293
159. Weichenberger CX, Sippl MJ (2006) Bioinformatics 22:1397
160. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) J Biomol NMR 8:477
161. Wiederstein M, Sippl MJ (2007) Nucleic Acids Res 35:W407
162. Vriend G (1990) J Mol Graph 8:52
163. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Nature 381:272
164. Hawkins PCD, Nicholls A (2012) J Chem Inf Model 52:2919
165. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, Buch R, Fiorin G, Hénin J, Jiang W, McGreevy R, Melo MCR, Radak BK, Skeel RD, Singharoy A, Wang Y, Roux B, Aksimentiev A, Luthey-Schulten Z, Kalé LV, Schulten K, Chipot C, Tajkhorshid E (2020) J Chem Phys 153:044130
166. Kawabata T (2019) Biophys Physicobiol 16:391
167. Halgren TA (2009) J Chem Inf Model 49:377
168. Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, Kozakov D, Vajda S (2012) Nucleic Acids Res 40:W271
169. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S (2015) Nat Protoc 10:733
170. Yang J, Roy A, Zhang Y (2013) Bioinformatics 29:2588
171. Laskowski RA (1995) J Mol Graph 13:323
172. Le Guilloux V, Schmidtke P, Tuffery P (2009) BMC Bioinform 10:168
173. Schmidtke P, Le Guilloux V, Maupetit J, Tuffery P (2010) Nucleic Acids Res 38:W582
174. Tan KP, Nguyen TB, Patel S, Varadarajan R, Madhusudhan MS (2013) Nucleic Acids Res 41:W314
175. Ravindranath PA, Sanner MF (2016) Bioinformatics 32:3142
176. Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, De Fabritiis G (2017) Bioinformatics 33:3036
177. Cui Y, Dong Q, Hong D, Wang X (2019) BMC Bioinform 20:93
178. Dominguez C, Boelens R, Bonvin AMJJ (2003) J Am Chem Soc 125:1731
179. Cole J, Nissink JWM, Taylor R (2005) Protein-ligand and docking virtual screening with GOLD. In: Shoichet B, Alvarez J (eds) Virtual screening in drug discovery. CRC Press, Boca Raton, p 379
180. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, Rizzo RC (2006) J Comput Aided Mol Des 20:601
181. Yan Y, Zhang D, Zhou P, Li B, Huang SY (2017) Nucleic Acids Res 45:W365
182. Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM (2015) Nucleic Acids Res 43:W425
183. Lin YF, Cheng CW, Shih CS, Hwang JK, Yu CS, Lu CH (2016) J Chem Inf Model 56:2287
184. Lu CH, Chen CC, Yu CS, Liu YY, Liu JJ, Wei ST, Lin YF (2022) Bioinformatics 38:4428
185. Ó Conchúir S, Barlow KA, Pache RA, Ollikainen N, Kundert K, O'Meara MJ, Smith CA, Kortemme T (2015) PLoS ONE 10:e0130433
186. Brancolini G, Kokh DB, Calzolai L, Wade RC, Corni S (2012) ACS Nano 6:9863