Biophysical Reviews (2022) 14:1423–1447
https://doi.org/10.1007/s12551-022-01015-8

REVIEW

Computer simulation of molecular recognition in biomolecular system: from in silico screening to generalized ensembles

Yoshifumi Fukunishi[1] · **Junichi Higo**[2,3] · **Kota Kasahara**[4]

Received: 22 September 2022 / Accepted: 6 November 2022 / Published online: 28 November 2022

Abstract

Prediction of ligand-receptor complex structure is important in both the basic science and the industry such as drug discovery. We report various computation molecular docking methods: fundamental in silico (virtual) screening, ensemble docking, enhanced sampling (generalized ensemble) methods, and other methods to improve the accuracy of the complex structure. We explain not only the merits of these methods but also their limits of application and discuss some interaction terms which are not considered in the in silico methods. In silico screening and ensemble docking are useful when one focuses on obtaining the native complex structure (the most thermodynamically stable complex). Generalized ensemble method provides a free-energy landscape, which shows the distribution of the most stable complex structure and semi-stable ones in a conformational space. Also, barriers separating those stable structures are identified. A researcher should select one of the methods according to the research aim and depending on complexity of the molecular system to be studied.

Keywords Free-energy landscape · Energy basin · Molecular binding · Conformation sampling · Thermodynamic integration · Weighted ensemble analysis method · Enhanced sampling · Drug discovery

Introduction

Computations together with X-ray, NMR, and electron microscopy have been used to study the tertiary structure of biologically important proteins and to develop drugs (Kyogoku et al. 2003). Haruki Nakamura and his group have contributed to development of computational approaches and the PDBj database (https://pdbj.org/). As known, PDB is the starting point to study a single biomolecular system and structural genomics, and those studies contribute to development of drug-discovery technologies.

The human genome includes 23,000 coding genes (International Human Genome Sequencing Consortium 2001; Venter et al. 2001). Data-driven deep learning models based on the Protein Data Bank, such as Alpha fold (Jumper et al. 2021; Mosalaganti et al. 2022), succeeded to predict precise 3D protein structures from the amino-acid sequences. Recent 76% of human-protein tertiary structures was predicted (Porta-Pardo et al. 2022). The mouse genome project elucidated the time-dependent RNA expression in each organ from embryo, ES cell, and mature mouse. The genes and other sequence data were annotated in FANTOM activities (Kawai et al. 2001; Abugessaisa et al. 2021). The ENCODE project showed the gene expression in each organ of human, and the time-dependent and organ-dependent RNA expression data were published as human cell atlas, brain atlas etc. (Regev et al. 2017; Kita et al. 2021). These approaches have indicated that transcription factors are coded in about 2000 genes (10% of genes) and that 1000 promotors exist on our genome. The transcription factors bind to other proteins and form functional transcription-factor complexes. Then, these complexes bind selectively to the promotors, and finally this selective

✉ Yoshifumi Fukunishi
lakeoesa@fc4.so-net.ne.jp

[1] Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26, Aomi, Koto-Ku, Tokyo 135-0064, Japan

[2] Graduate School of Information Science, University of Hyogo, 7-1-28 Minatojima Minamimachi, Chuo-Ku, Kobe, Hyogo 650-0047, Japan

[3] Research Organization of Science and Technology, Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan

[4] College of Life Sciences, Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan

binding controls pathways, which consist of functionally related proteins (Khambata-Ford et al. 2003; Babu et al. 2004). The KEGG pathway database includes about 500 pathways, and response of RNA expression patterns against 1000 chemicals were archived in the Broad Institute as a connectivity map (Kanehisa et al. 2021; Lamb et al. 2006; Musa et al. 2018). These progresses in research give new definitions of diseases, healthy, and ageing states of life. The combination of data-driven protein-complex modelling and genome-wide association study (GWAS) elucidates the structures and functions of organelles, nuclear pore, transcription factors, and membrane systems (Uffelmann et al. 2021; Mosalaganti et al. 2022).

The inter-disciplinary studies reveal multiple pathway control by a combination of approved medicines. One of the successes from the inter-disciplinary studies is "chimeric antigen receptor T cell" (CAR-T cell) therapies. The genetic engineering has enabled designing artificial antibodies targeting specific antigens and these artificial genes introduced in the T cells isolated from the patient's blood. These personalized medicines have succeeded mainly in cancer treatments. Although CAR-T cell therapies are always facing a risk of un-controlled proliferation of the CAR-T cells, some studies suggested how to control CAR-T by high-selective kinase inhibitors (Mestermann et al. 2019). Since the aging and healthy states are clearly distinguished by transcriptome and pathway analysis, some rational anti-senescence therapies have been proposed by using a pair of high-selective kinase inhibitor and Bcl-xL inhibitors (Fig. 1) (Campisi et al. 2019; Kirkland and Tchkonia 2020; Gasek et al. 2021; Shafqat et al. 2022).

The novel therapies mentioned above suggest that the state-of-the-art technology can be developed based on atomic-level interactions between the high-selective drug molecules and target proteins in solvent. This review, therefore, focusses on the computations to study the molecular interactions. How can small or medium-sized drugs and proteins bind to their target molecules selectively? The previous studies showed that each cell expresses only several thousand genes and that the produced proteins are localized in organelles (i.e., nucleus, mitochondria, endoplasmic reticulum, etc.) divided by membranes in a cell where innumerable molecules are crowded (Delarue et al. 2018; Mourão et al. 2014). As described above, these proteins bind selectively to their binding partner and form functional complexes.

The protein surfaces are mainly hydrophilic to avoid aggregation. Recently, "cryptic site" was proposed as a case of the selective binding mechanisms (Cimermancic et al. 2016; Beglov et al. 2018; Vajda et al. 2018). The cryptic site is hidden in the apo form and opened in the holo form, which is an example of polymorphism. This type of molecular recognition mechanism is understood by combination of database analysis and molecular simulations as will discussed later. The molecular simulation becomes more important when we study a binding mechanism between an intrinsically disordered protein (or domain) and its binding partner. Because a conformational motion of the intrinsically

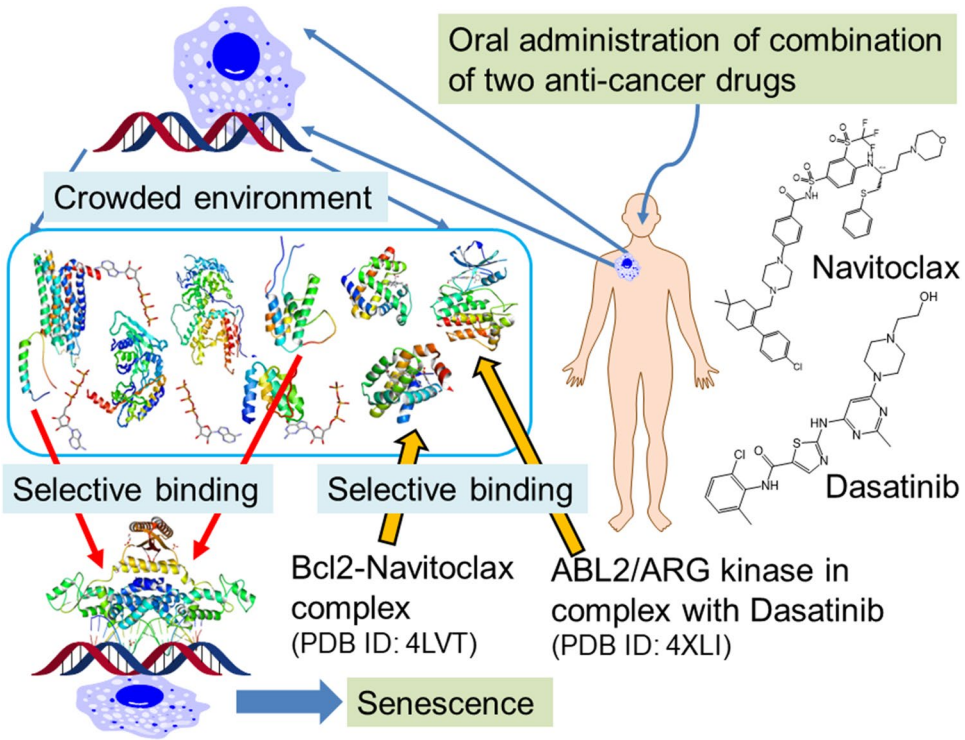

Fig. 1 Schematic representation of systems biology, medication by drug molecules, and difficulty. Genome, transcriptome, proteome, pathway analysis, and atomic-level molecular simulation enable us to find and analyze new understandings of life and new medications

disordered protein is considerably large and complicated, a more efficient sampling method (enhanced sampling) is required, as discussed later.

Many organs (mainly brain) secrete chemicals and peptides for inter-organ cross talks. The secretions of molecules (e.g., adrenalin, histamine, insulin, endothelin) mainly work for signal transduction as stress response. In our body, most of these molecules are generated from stocked materials like amino acids, lipids, and nucleic acid. For instance, adrenalin, histamine, and dopamine are respectively generated from Phe, His, and Trp, and their chemical formulas are similar mutually. Nonetheless, the secreted chemicals are selectively recognized by their receptors (Joedicke et al. 2018). To uncover such a high selective binding mechanism, an atom-based approach is mandatory.

As mentioned above, the pathway is controlled by the protein–ligand complex formation, and then a molecule, which binds to a pathway-relating protein, can be a drug candidate. In this review, we focused on various the computational approaches from in silico (virtual) screening to enhanced sampling (generalized ensemble) to elucidate the molecular-recognition mechanism. Before that, however, we present in the next chapter a simple and fundamental framework to consider the most thermodynamically stable complex structure and semi-stable complex structures.

Stable states

Before explaining the complex-structure prediction methods actually, we mention the complex formation fundamentally. Suppose that ligand and receptor are distant to each other in solution at a physiological temperature. Molecular binding is a process where the ligand approaches the ligand-binding site of the receptor and eventually the native complex is formed. In a conformational space, the binding is a process where the system's conformation moves from a high free-energy region to a low free-energy one and finally the conformation falls in the lowest free-energy basin (native complex basin) (Fig. 2a). Contrarily, if the lowest free energy is marginally lower than the others, the system may exhibit a fuzzy complex state (Fig. 2b) (Tompa and Fuxreiter 2008), and consequently a single complex conformation is not determined experimentally because the conformation is fluctuating among multiple conformations. In this case, the aim of the computation is to find these multiple basins.

A free-energy landscape shows distribution of low free-energy basins (Fig. 2a, b). The system's conformation $\boldsymbol{r}$ is originally a multi-dimensional quantity expressed as: $\boldsymbol{r} = [x_1, y_1, z_1, \cdots, x_N, y_N, z_N]$ where $[x_i, y_i, z_i]$ is the Cartesian coordinates of the i th atom and N is respectively

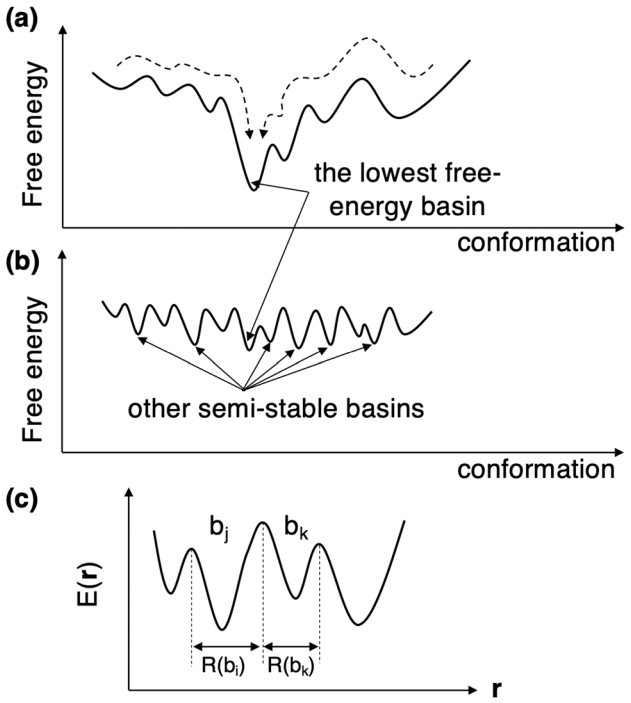

Fig. 2 **a** Scheme of free-energy landscape. X-axis represents molecular conformation one-dimensionally, although it is high dimensional originally. Y-axis represents free energy (PMF) of conformation at physiological temperature. The lowest free energy is remarkably lower than the others. Broken lines show complex formation process. **b** System with multiple complex basins, whose free energies are similar mutually. The conformation fluctuates among the basins. **c** Potential energy surface $E(\boldsymbol{r})$, where $\boldsymbol{r}$ is position of the system. Two basins b_j and b_k are mentioned in text, whose territories are $R(b_i)$ and $R(b_k)$, respectively

the number of constituent atoms of the system (biological molecules, solvent molecules and other atoms in the system). Denoting the potential energy of a conformation (microscopic state) as $E(\boldsymbol{r})$, the statistical weight (thermodynamic weight) at thermal equilibrium assigned to $\boldsymbol{r}$ at a temperature T is given formally by

$$\rho(\boldsymbol{r}) \propto \exp\left[-\frac{E(\boldsymbol{r})}{R_{gas}T}\right] \tag{1}$$

where R_{gas} is the gas constant (the energy unit is kcal/mol). We omit a kinetic energy in Eq. 1 to make explanation simple. The normalization factor (partition function) is also omitted because N, system's volume and T are constant here. A fractional free energy G_{b_j} assigned to a basin j (denoted as b_j in Fig. 2c) is defined by

$$G_{b_j} = -R_{gas}Tln\left[\int_{R(b_j)} \rho(\boldsymbol{r})d\boldsymbol{r}\right] \tag{2}$$

The multi-dimensional integral is taken in a region $R(b_j)$ (i.e., territory of b_j), which is occupied by microscopic states belonging to b_j. Figure 2c is presented so that b_j is more stable than b_k at equilibrium: $G_{b_j} < G_{b_k}$.

Equation 2 is a formal expression to assess the stability of each basin. However, this multidimensional integral is not achievable for many biological systems because the high-dimensional space is fractioned into basins of complicated shapes. Instead, a ratio G_{b_j}/G_{b_k} is computable numerically by an enhanced sampling simulation, whereas each of the fractional free energies G_{b_j} and G_{b_k} is not computable. Although rigorous determination of territory $R(b_j)$ is difficult, $\exp[-\frac{E(\boldsymbol{r})}{R_{gas}T}]$ around an inter-basin boundary is small. Therefore, an error caused by uncertainty of $R(b_j)$ may be negligible.

It is helpful to convert the position $\boldsymbol{r}$ defined in the Cartesian-coordinate space to a low-dimensional position $\boldsymbol{q}$, refer to as "reaction coordinate": $\boldsymbol{q} = [q_1, q_2, \ldots, q_n]$, where n is dimensionality of the reaction-coordinate space ($n < N$). Note that the function form of $\boldsymbol{q} = \boldsymbol{q}(\boldsymbol{r})$ is known for the coordinate conversion. Accordingly, the weight $\rho(\boldsymbol{r})$ is converted to $P(\boldsymbol{q})$ as:

$$P(\boldsymbol{q}) = \int \rho\left(\boldsymbol{r}'\right)\delta(\boldsymbol{q}(\boldsymbol{r}), \boldsymbol{q}(\boldsymbol{r}'))d\boldsymbol{r}' \tag{3}$$

where $\delta\left(\boldsymbol{q}(\boldsymbol{r}), \boldsymbol{q}(\boldsymbol{r}')\right)$ is a delta function that is non-zero only when $\boldsymbol{r}'$ is involved in a range $\boldsymbol{q} - d\boldsymbol{q} \le \boldsymbol{q}(\boldsymbol{r}') \le \boldsymbol{q} + d\boldsymbol{q}$ set in the reaction-coordinate space: $\int_{-\infty}^{\infty}\delta(\boldsymbol{q}(\boldsymbol{r});\boldsymbol{q}(\boldsymbol{r}'))d\boldsymbol{r}' = 1$. In a real sampling, the number of sampled conformations is finite. Then, $\delta(\boldsymbol{r} - \boldsymbol{r}')$ is replaced by a function $D\left(\boldsymbol{q}(\boldsymbol{r});\boldsymbol{q}(\boldsymbol{r}')\right)$ in Eq. 3: $D\left(\boldsymbol{q}(\boldsymbol{r});\boldsymbol{q}(\boldsymbol{r}')\right) = v$ in a range of $\boldsymbol{q} - \Delta\boldsymbol{q} \le \boldsymbol{q}(\boldsymbol{r}') \le \boldsymbol{q} + \Delta\boldsymbol{q}$ and $D\left(\boldsymbol{q}(\boldsymbol{r});\boldsymbol{q}(\boldsymbol{r}')\right) = 0$ outside the range with condition of $\int_{-\infty}^{\infty}D(\boldsymbol{q}(\boldsymbol{r});\boldsymbol{q}(\boldsymbol{r}'))d\boldsymbol{r}' = 1$. Then, $P(\boldsymbol{q}) = \sum_i w_i D(\boldsymbol{q}(\boldsymbol{r});\boldsymbol{q}(\boldsymbol{r}_i'))$, where w_i is a statistical weight assigned to the i th snapshot determined from the sampling.

A force $\boldsymbol{F}(\boldsymbol{q})$ acting on the system at $\boldsymbol{q}$ in the reaction-coordinate space is expressed formally as:

$$\boldsymbol{F}(\boldsymbol{q}) = \left[F_1(\boldsymbol{q}), F_2(\boldsymbol{q}), \ldots, F_n(\boldsymbol{q})\right] \tag{4}$$

where the i th element F_i is the force acting on the system at $\boldsymbol{q}$ parallel to the q_i axis and defined as:

$$F_i(\boldsymbol{q}) = \boldsymbol{e}_i \bullet \left[\int \delta(\boldsymbol{r} - \boldsymbol{r}')\boldsymbol{f}(\boldsymbol{r}')\rho\left(\boldsymbol{r}'\right)d\boldsymbol{r}'\right] \tag{5}$$

where $\boldsymbol{e}_i$ is the unit vector parallel to the q_i axis, and $\boldsymbol{f}(\boldsymbol{r})$ is the force acting on the system at $\boldsymbol{r}$ in the Cartesian space: $\boldsymbol{f}(\boldsymbol{r}) = -grad[E(\boldsymbol{r})]$, where derivatives are calculated with respect to the Cartesian coordinates $\boldsymbol{r}$. Equation 5 indicates that $\boldsymbol{F}(\boldsymbol{q})$ is related to the thermal average of force $\boldsymbol{f}(\boldsymbol{r})$ at $\boldsymbol{q}$ because the thermodynamic weight $\rho(\boldsymbol{r})$ is used for averaging $\boldsymbol{f}(\boldsymbol{r})$. Therefore, $\boldsymbol{F}(\boldsymbol{q})$ is called a "mean force." Then a potential function computed from a line integral of $\boldsymbol{F}(\boldsymbol{q})$ is called "potential of mean force" (PMF) (Tuckerman 2010). However, instead of executing the line integral, PMF is computable directly from $P(\boldsymbol{q})$ as:

$$PMF(\boldsymbol{q}) = -R_{gas}Tln[P(\boldsymbol{q})] \tag{6}$$

The fractional free energy G_{b_j} is computed by integrating $P(\boldsymbol{q})$ in its territory $R(b_j)$:

$$\begin{aligned} G_{b_j} &= -R_{gas}Tln\left[\int_{R(b_j)}\exp[-\tfrac{PMF(\boldsymbol{q})}{RT}]d\boldsymbol{q}\right] \\ &= \int_{R(b_j)}P(\boldsymbol{q})d\boldsymbol{q} \\ &= \sum_i^{n_j} w_i^{b_j}, \end{aligned} \tag{7}$$

where $w_i^{b_j}$ is a statistical weight assigned to the i th snapshot in b_j, and n_j is the number of snapshots in b_j. Although it is difficult to calculate $w_i^{b_j}$ by a conventional MD simulation in a wide conformational space, a generalized ensemble method provides $w_i^{b_j}$ naturally.

Here, we define the word "free-energy landscape" clearly. Originally, the word "free energy" is used to express an entire statistical property of the system: $G = -R_{gas}Tln[Z]$. The term Z is the, so-called, partition function defined as $Z = \int^{\text{entire}}\rho(\boldsymbol{r})d\boldsymbol{r} = \int^{\text{entire}}P(\boldsymbol{q})d\boldsymbol{q}$, where the integral is taken over the entire conformational space. On the other hand, the word "free-energy landscape" is usually used to show the spatial patterns of the probability $P(\boldsymbol{q})$ or $PMF(\boldsymbol{q})$ in the reaction-coordinate space. Therefore, the free-energy landscape may be called "PMF landscape" or "probability landscape." We note that the formulations of the free energy and PMF have a similarity: When $P(\boldsymbol{q})$ in Eq. 6 is replaced by Z, PMF becomes G.

Now we outline the computational methods and their limits of applications. If the intra-molecular deformation in each of receptor and ligand is small upon binding, a simple in silico docking is useful: A chemically stable ligand structure and the receptor's apo form are combined as building brocks to generate various complex poses. As explained later, the ligand conformational varieties caused by rotatable-bond rotations are considered in the in silico docking. Then, the plausibility of each pose is assessed by a physical interaction energy or an empirically introduced scoring function, which is given later. Because this procedure can be done very quickly, many ligands can be tested by repeating this procedure (high-throughput screening). Details are explained later.

If the receptor undergoes a large intra-molecular deformation during the complex formation, preparation of various receptor's conformations (ensemble) in advance is useful: The docking procedure is performed between the ligand and many conformations in the receptor's ensemble. This procedure is called "ensemble docking" (Carlson et al. 1999; Amaro et al. 2018; Falcon et al. 2019). If this

procedure works well, the complex formation accords probably to "conformation selection" (Bosshard 2001; James and Tawfik 2003; Yamane et al. 2010). The ensemble is generated from the receptor's apo form using conventional molecular dynamics (MD), Monte-Carlo (MC) sampling, or enhanced sampling (generalized ensemble method).

If both the receptor and ligand undergo large conformational deformations during the complex formation, preparation of a ligand's conformational ensemble as well as the receptor's conformational ensemble may be useful. However, this procedure might be inefficient when the generated ensembles do not contain ligand and receptor conformations appropriate for constructing the lowest free-energy form (the native complex structure). This suggests that the complex formation accords with the "induced fit" (Monod et al. 1965; Spolar and Record 1994). Furthermore, a difficulty appears when conformational fluctuations (i.e., entropy) and a solvent effect contribute to the complex stability.

An extreme case is found in an intrinsic disordered segment binding to its binding partner (Wright and Dyson 1999). This segment is disordered in the unbound state and may fold in a tertiary structure when binding to the partner (coupled folding and binding) (Dyson and Wright 2005; Sugase et al. 2007). To predict the complex structure, all molecules should be involved in a single system using a completely flexible model. Therefore, a powerful sampling method, a generalized ensemble method (an enhanced sampling method), is required.

It is fundamentally interesting to distinguish the population selection and the induced fit in the complex formation (Hammes et al. 2009). Many works argued the population-selection vs induced-fit problem (Okazaki and Takada 2008; Hammes et al. 2009; Silva et al. 2011; Bucher et al. 2011; Vogt and Cera 2012; Nussinov et al. 2014; Ravasio et al. 2019; Vauquelin and Maes 2021). A generalized-ensemble study by Nakamura and his coworkers (Higo et al. 2011) reproduced a coupled folding and binding phenomena, which is expressed by an intrinsically disordered segment NRAF/REST binding to the paired amphipathic helix (PAH) domain of mSin3B (Nomura et al. 2005). The study concluded that the population selection and the induced fit works together in a coupled manner. It is natural to consider that the binding mechanism depends on the system because of the variety of the biological system.

Receptor-ligand docking and in silico screening

Receptor–ligand docking software that predicts the receptor–ligand complex structures and the binding free energies ΔG, has been a key technology of the in silico (virtual) drug screenings and the rational drug designs from 1990, and still now a number of reports has been published on the receptor–ligand docking programs and the combinations of them (Pagadala et al. 2017; Salmaso and Moro 2018; Amaro et al. 2018; Bender et al. 2021; Pinzi and Rastelli 2019). Haruki Nakamura and his coworkers are developers of docking software (sievgene/myPresto) and a basic method for docking study (Fukunishi et al. 2005). Part of his work is now available as "myPresto program suite" (https://www.mypresto5.jp/en/) where about 20 programs can be downloaded under a LGPL v2 license. A member of myPresto software developers is allowed to use them under FreeBSD license.

Before starting the docking procedure, the ligand-binding site must be indicated (this identification is discussed later). In general, a docking method consist of two or three steps. The first step is the ligand-allocation scheme on the receptor surface around the indicated ligand-binding site and gives many receptor–ligand complex structure candidates ("docking poses"). The second step is an evaluation of the docking poses by applying a scoring function, which estimates roughly the ΔG values of given docking poses and selects some probable or stable docking poses. The third step is the re-scoring of the selected poses by using a more precise scoring function than the rough scoring function used above. The final docking poses correspond to ΔG values.

Usually, the scoring function is classified into three types (Li et al. 2019): physico-chemical, knowledge-based, and empirical scoring functions. We focus on the docking methods based on the physico-chemical scoring function because this scoring function can incorporate readily new elements, such as boron (Soriano-Ursúa et al. 2014) and silicon (Franz and Wilson 2013). The popular docking software based on the physico-chemical scoring is Dock (Kuntz et al. 1982), AutoDock (Goodsel et al. 1996), Glide (Friesner et al. 2004; Halgren et al. 2004) etc. The knowledge-based scoring function is calculated from a pairing distribution function between two atom-groups recorded in a database. The popular docking software based on the knowledge-based scoring function are GOLD (Jones et al. 1997; Verdonk et al. 2003). The empirical one is from parameter tiffing in a function to reproduce the experimental ΔG (Pereira et al. 2016; Ragoza et al. 2017). The popular docking software based on the empirical base is GOLD chemscore, FlexX (Rarey et al. 1996), PRO-LEAD (Baxter et al. 1998), rDock(Ruiz-Carmona et al. 2014) etc.

Binding-free energy estimation is the core technology of docking software

Receptor–ligand binding free energy is one of the major factors determining the activities of drug molecules, signal transductions, and many other physiological phenomena. Both the molecular simulation and experimental

approaches can give the ΔG values. Suppose that one receptor has only one ligand-binding site (or region) and that one receptor molecule can bind only one ligand molecule to form one receptor-ligand complex. Let the ligand-binding region (R_B) be clearly distinguishable from the other region (R_U) in the conformational space. Equations 3 and 7 gives ΔG ($= G_{R_B} - G_{R_U}$) as follows,

$$\Delta G = -R_{gas}Tln\left[\frac{\int_{R_B} P(\boldsymbol{q})d\boldsymbol{q}}{\int_{R_U} P(\boldsymbol{q})d\boldsymbol{q}}\right] = -R_{gas}Tln\left[\frac{P_B}{P_U}\right] \tag{8}$$

where P_B and P_U are the probabilities of the bound (R_B) and unbound (R_U) states, respectively: $P_B = \left[\int_{R_B} \rho(\boldsymbol{r})d\boldsymbol{r}\right]$ and $P_U = \left[\int_{R_U} \rho(\boldsymbol{r})d\boldsymbol{r}\right]$ using Eq. 3.

On the other hand, the popular experimental methods for ΔG evaluation are the isothermal titration calorimetry (ITC) and the surface plasmon resonance (SPR) experiments that give the binding constant K_a (Rich and Myszka. 2007; Wiseman et al. 1989). When the system is in the equilibrium state under the standard condition (298 K and 1 atm), the dissociation constant $K_D(= 1/K_a)$ gives the standard molar Gibbs free energy change of binding ΔG^0 as follows,

$$\Delta G^0 = R_{gas}Tln\left[\frac{K_D}{C^0}\right] \tag{9}$$

where C^0 is a reference concentration of 1 mol/L (Gilson et al. 1997; Deng and Roux 2009).

Since the ΔG value depends on the experimental conditions (temperature, pressure, and the other experimental conditions), ΔG^0 is useful for comparing the stability of complexes among multiple receptors and ligands measured from different experiments. Therefore, ΔG^0 is adopted in the scoring functions.

Besides the binding-constant observation experiments, there have been many experimental methods, which provide the binding affinity of ligand. Namely, the half maximal inhibitory concentration (IC_{50}), percent inhibition, inhibition constant (K_i) etc. These quantities could be somehow translated to the binding free energy differences by using the Cheng-Prusoff equation (Yung-Chi and Prusoff 1973) and the other equations. These affinity data have been useful for developing the scoring functions.

As mentioned, the accurate calculation of ΔG is still a very time-consuming and expensive task. On the other hand, in silico screening is usually applied to many compounds. Preparation of large ligand library and usage of many computation nodes are effective to increase the efficiency of the in silico screening (Gentile et al. 2022; Gorgulla et al. 2020; Lyu et al. 2019). Therefore, approximation of ΔG with maintaining a certain accuracy is crucially important for docking software.

Receptor–ligand docking as supporting tool for X-ray crystallography

In 1983, Kuntz group published the first docking program DOCK for assisting the X-ray crystallographic coordinates of small molecules in the protein–ligand complex (Kuntz et al. 1982). The docking procedure of DOCK was following four steps. (1) DOCK puts various size of spheres that represent the ligand on the receptor surface to search the ligand-binding position with avoiding atomic conflictions. (2) The second step is a trial-and-error ligand-docking cycle. DOCK locates the ligand molecule in various conformations on the predicted ligand-binding position indicated by Step 1. (3) DOCK evaluates the stability of each conformation by applying a binding-enthalpy function. (4) Finally, DOCK selects the candidate most-stable receptor-ligand conformation. DOCK estimates the binding enthalpy (ΔE) by Eq. 10 instead of Eq. 11 which is the classical force field for an MD simulation.

$$\Delta E = \sum_{i\in rec}\sum_{j\in rec} 4\varepsilon_{ij}\left\{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^9 - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right\} + 332.0\sum_{i\in rec}\sum_{j\in rec}\frac{q_iq_j}{4{r_{ij}}^2} \tag{10}$$

or

$$\Delta E = \sum_{i\in rec}\sum_{j\in rec} 4\varepsilon_{ij}\left\{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right\} + 332.0\sum_{i\in rec}\sum_{j\in rec}\frac{q_iq_j}{r_{ij}} \tag{11}$$

where subscripts i and j designate the i th atom of receptor and the j th atom of ligand, respectively. The parameters r_{ij}, σ_{ij}, and ε_{ij} are the inter-atomic distance (Å), van der Waals (vdW) radius (Å), and a coefficient for the vdW interaction assigned to the atom pair of i and j. The parameters q_i and q_j are atomic partial charges (atomic unit) assigned to atoms i and j, respectively. The number "332.0" is to set the energy in kcal/mol unit. The first and second terms of Eq. 10 correspond respectively to the soft-core vdW interaction and the Coulomb interaction in implicit-water solvent. In general, a receptor–ligand complex with a strong affinity shows good interface complementarity. A slight coordinate error of the ligand causes atomic conflicts, which result in a strong repulsion and a large error in the docking score. Therefore, most receptor–ligand docking programs have adopted the soft-core vdW potential.

In the Coulomb interaction term, $4r_{ij}$ represents an effective dielectric constant ε_{eff}. Assuming that ε_{eff} depends on the distance R_p from the protein surface to the ligand, the simplest form is $\varepsilon_{eff} = 4R_p$ (Mallik et al. 2002). In Eq. 10, R_p is approximated by r_{ij}.

Because the initial version of DOCK was designed for crystal-structure analysis, this version did not involve an entropy term. DOCK has been modified frequently last few decades, and the current ligand-allocation scheme and the scoring function are different from those of the initial version.

Toward receptor-ligand docking in implicit aqueous solvent

Present docking software is designed to the receptor–ligand docking in implicit water solvent at the room temperature. Equation 12 is one of the AutoDock scoring functions, and many other scoring functions are similar to this function more or less (Goodsel et al. 1996). Note that ΔG in Eq. 12 is regarded as an approximation of the free energy change caused by the ligand-receptor complex formation. The enthalpy part of ΔG is expressed by the first three terms: the receptor–ligand vdW, hydrogen-bonding, and Coulomb interactions. The entropy part of ΔG consists of the fourth and last terms. The fourth term represents the entropy loss of ligand molecule in binding: the ligand can have multiple conformations in bulk water (unbound state), although it is fixed to a single conformation in the binding site. The last term represents the receptor–ligand hydrophobic interaction in water.

$$\begin{aligned}\Delta G = f_{vdW}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}}+\frac{B_{ij}}{r_{ij}^{6}}\right)+f_{hbond}\left\{\sum_{i,j}\left(\frac{C_{ij}}{r_{ij}^{12}}+\frac{D_{ij}}{r_{ij}^{10}}\right)+E_{hbond}\right\}\\ +f_{elec}\sum_{i,j}\frac{q_iq_j}{\varepsilon(r_{ij})r_{ij}}+\Delta G_{tor}N_{tor}+f_{sol}\sum_{i,j}\left(S_iV_j+S_jV_i\right)\exp\left[-\frac{r_{ij}^2}{2\sigma^2}\right],\end{aligned} \quad (12)$$

where f_{vdW}, f_{hbond}, f_{ele}, f_{sol}, and f_{vdW} are respectively fitting coefficients for terms of vdW, hydrogen bond, Coulomb, entropy-loss, and the dehydration free energy to remove a hydration shell from the receptor–ligand interface. A classical MD force field gives the values of coefficients A_{ij}, B_{ij}, C_{ij}, and D_{ij}. E_{hbond}, q_k $(k = i \text{or} j)$, $\varepsilon\left(r_{ij}\right)$, ΔG_{tor}, and N_{rot} are respectively a correction term for a hydrogen bond, an atomic charge for atom k, the distance-dependent dielectric constant for atom pair i and j, the entropy loss with respect to a rotatable bond, and the number of rotatable bonds in the ligand. S_k, V_k, and σ are respectively an atomic-solvation parameter for atom k, an occupied atomic volume for atom k, and an average vdW radius of heavy atom except hydrogen atom. The fitting coefficients are determined to reproduce the experimental ΔG values from many receptor–compound complexes.

The parameters in the scoring function ΔG are determined from a thermodynamic cycle, which is a well-known cycle in computational field as shown in Fig. 3a and b. Figure 3a shows that ΔG is given by the molecular interaction in vacuum (E_1) and the aqueous solvation free energies (ΔG^2_{solv} and ΔG^3_{solv}). Since classical force field gives the E_1 value, the unknown factor is only the solvation free energy. Figure 3b shows that

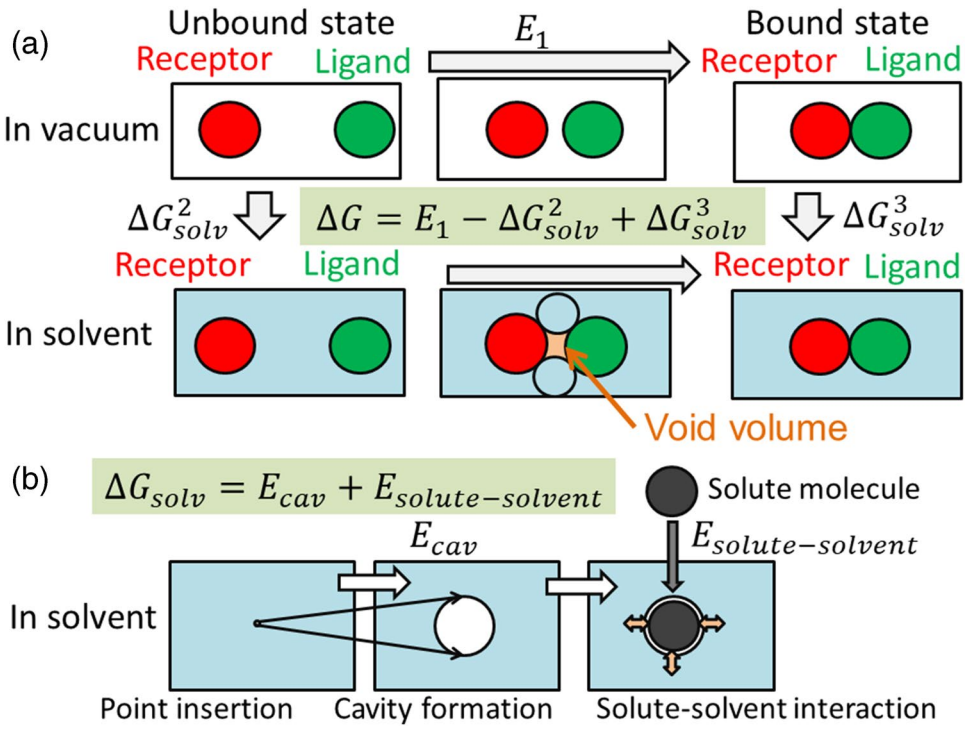

Fig. 3 Schematic representation of molecular interaction energy calculation. **a** Thermodynamic cycle to calculate ΔG. E_1, ΔG^2_{solv} and ΔG^3_{solv} are transfer energies and arrows represent the transfer directions. Boxes in blue represent solvent water. Red, green, and blue circles represent the receptor, ligand, and water molecules, respectively. Void volume is colored in orange between the receptor, ligand, and water molecules. **b** Schematic representation of scaled-particle theory. Point insertion, cavity formation, and solute–solvent interaction energy are shown. Small arrows in orange represent the solute–solvent interaction

the solvation process consists of the cavity formation and solute–solvent interaction processes. Broadly speaking, the works for cavity formation and the short-range solute–solvent interaction energy are approximately proportional to the surface area of the cavity. The fitting parameters of the surface area are determined to reproduce the experimental solvation free energy values of various compounds. Thus, the physico-chemical docking scoring function is, in general, a combination of surface area term and the classical force field used in the conventional molecular simulation.

Estimation of cavity formation energy in aqueous solvent

The scaled particle theory (SPT) has been one of the basic theories for solvation energy calculation and most of the docking software adopts the SPT or variations of SPT (Pierotti 1976). The original SPT explains the solvation of one spherical particle in the solvent and the SPT was extended to estimation of solvation energy of receptor-ligand systems. Namely, the result obtained by the SPT shows that the solvation free energy is approximately proportional to the solute surface area. By replacing the radius of spherical solute by the solvent-accessible surface area, the approximation formula of solvation free energy given by SPT is extended for solvation of polyatomic molecules. Finally, the approximation formula is extended to estimation of solvation-free energy of the receptor-ligand systems.

In SPT, the solvation consists of two processes (Fig. 3b). The first process is a vacuum cavity formation for insertion of solute in the solvent, and the second process is calculation of a solute–solvent interaction when the solute exists in the cavity. The cavity formation energy E_{cav} is approximated by a polynomial as follows.

$$E_{cav} = c_0 + c_1 R + c_2 R^2 + c_3 R^3 + c_4 R^4 + \cdots, \tag{13}$$

where R is the radius of the cavity, and c_k $(k = 1, 2, \ldots)$ is a coefficient assigned to each term. This equation is an expansion of a general equation $E_{cav} = -R_{gas} T ln[\rho]$, where ρ is the atomic packing factor.

Each term of Eq. 13 has its own physico-chemical significancy although we do not explain in detail: See paper by Pierotti (1976) for instance. When the solvent is water at pressure 1 atm and when it consists of a spherical-rigid water model, SPT shows that the radius of solvent molecule and the density of solvent determine the c_0, c_1, c_2 and c_3 values and the other higher order coefficients are zero, and the third term $(c_2 R^2)$ is dominant. Then, Eq. 13 is rewritten as

$$E_{cav} \approx c_2 R'^2 = c_{surf} \times ASA, \tag{14}$$

where c_{surf} and ASA are the coefficient of atomic surface tension and the solvent-accessible surface area (ASA) of the given solute, respectively. R is replaced by R', which is sum of R and the vdW radius of a water molecule. While ASA is a quantity difficult to be computed for a solute of general shape, Richmond provided an analytical computation method (Richmond 1984). However, the computation was still time consuming by a computer. Then, Stouten et al. proposed a simple and fast approximation method without conditional branch (Stouten et al. 1993) as follows.

$$E_{cav} = f_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i \right) \exp\left[-\frac{r_{ij}^2}{2\sigma^2} \right] \tag{15}$$

Note that this expression appears in the last term of Eq. 12. Now, this approximation and the variations have been widely used, e.g., AutoDock (Goodsel et al. 1996) and sievgene (Fukunishi et al. 2005).

Solvation free energy

Major inter-molecular interactions for biomolecules are the vdW and Coulomb interactions. Ooi et al. assumed that the vdW interaction is a short-range interaction and that the major contribution of the electrostatic interaction is from the first hydration shell (Ooi et al. 1987). Then, both E_{cav} and the solute–solvent interaction energy, $E_{solute-solvent}$, are proportional to ASA approximately, and the solvation free energy ΔG_{solv} is given simply as

$$\Delta G_{solv} = E_{cav} + E_{solute-solvent} \approx c \times ASA, \tag{16}$$

where c is, so-called, an atomic solvation parameter. Dividing ASA into contribution from individual atoms, Eq. 16 is transformed as

$$\Delta G_{solv} \approx \sum_{i=1}^{N} c_i \times ASA_i, \tag{17}$$

where c_i and ASA_i are the atomic solvation parameter and the ASA of the i th atom, respectively, and N is the number of atoms in the receptor and ligand. The parameter c_i depends on the atomic partial charge and vdW parameter of each atom. Various modified ASA methods have been proposed (Kang et al. 1987).

To improve the accuracy of ΔG_{solv}, the electrostatic energy was further considered because this energy is long range by nature. The Poisson-Boltzmann (PB) equation provides the electrostatic energy in a cell of the 3D real space consisting of multiple small volumes with different dielectric

constants. On the other hand, the PB equation should be solved in a large cell to consider the long-range property of the electrostatic energy (Gilson et al. 1988). Although Nakamura et al. succeeded in solving the PB equation precisely in a small cell, the computation was still time consuming (Nakamura and Nishida 1987; Nakamura 1988). The generalized Born (GB) method is a fast approximation method for the electrostatic energy, which is designed to mimic the results from the PB equation (Hawkins et al. 1996; Onufriev et al. 2002). In the framework of the GB method, the short-range interaction is only the vdW interaction (E_{vdW}), and ΔG_{solv} is given as

$$\Delta G_{solv} = E_{cav} + E_{vdW} + E_{Coulomb} \approx c\sum_{i=1}^{N} ASA_i + E_{GB}, \tag{18}$$

where $E_{Coulomb}$ is the electrostatic energy and E_{GB} the approximated electrostatic energy from the GB equation. The atomic surface tension parameter c_i in Eq. 17 is constant in Eq. 18, which is set to 10 cal/mol/Å^2 in aqueous solvent in general. The combination of the PB equation and SPT is called a PBSA method, and Eq. 18 is called a GBSA (generalized-Born accessible-surface area) method. Currently, the GBSA method with a quantum mechanics (QM) method in the reaction field has succeeded in reproducing the solvation free energies and pKa for various solutes (Irisa et al. 1995; Cramer and Truhlar 2008).

As mentioned in Eq. 18, the atomic surface tension parameter c is constant. However, the solvent structure and dynamics (entropy and enthalpy) depend on the site around protein (Suzuki et al. 1997; Assaf and Nau 2018; Salis and Ninham 2014; Nakamura et al. 1988; Lumry and Rajender 1970; Freire 2008; Kabir et al. 2003). This means that c is not constant. Still now, the solvent structure on the solute–solvent interface and the change of entropy and enthalpy upon the receptor-ligand binding are unclear.

Additional effect not included in many scoring functions: void-volume effect

The effect of a void volume to ΔG is not considered in Eq. 12. The void is defined by a volume between the Conolly and vdW surfaces in a system consists of solutes and water molecules (Fig. 3a). The contribution of the void volume to free energy is well explained by physics and the behavior of PMF by changing the void volume was computed accurately (Rashin 1989 and 1990; Fukunishi and Suzuki 1996; Gallicchio et al. 2000; Trzesniak et al. 2007). However, the estimation of the void volume contribution to ΔG is time consuming and ignored in many docking programs.

Ligand conformation generation and force field

Before starting ligand docking, most of docking software prepares the multiple conformations of ligand with respect to rotatable bonds. Designating the number of rotatable bonds as N_{rot} and supposing that the number of energy minima regarding the bond rotation is three, e.g., trans, gauche$^+$ and gauche$^-$, the number of possible stable conformations is $3^{N_{rot}}$. If the receptor-ligand complex adopts only one binding pose, the ligand selects one conformer out of the $3^{N_{rot}}$ ones and the ligand loses entropy of $ln\left[3^{N_{rot}}\right]$.

The force fields (FFs) for small compounds were estimated from X-ray diffraction data and infrared (IR) spectrums. A GF matrix (or FG) method translates the IR spectrum to the force constants of the bonds, angles, bond-angle cross terms of the molecule (Wilson 1941; Boyd 1968). Allinger et al. constructed the force fields of small compounds and developed the MM2/MMP2 and MM3 programs. MMP2 calibrates the force fields around aromatic rings by the semi-empirical QM (Allinger 1976, 1977; Allinger et al. 1994). The Quantum Chemistry Program Exchange (QCPE) distributed the programs of MM2/MMP2, ECEPP and many program-source codes to computer chemists all over the world by free (Halgren 1996a, 1996b; Boyd 2013). Halglen et al. applied the high-level ab-initio QM to many compounds and developed MMFF94 force field (Halgren 1996a, 1996b). Namely, MMFF94 force field parameters were derived from 500 molecular structures optimized at the HF/6-31G* level, 475 structures optimized at the MP2/6-31G* level, 380 MP2/6-31G* structures and 1450 structures partly derived from MP2/6-31G* geometries. The MM2/MMFF94 force fields and MM2 software have been widely used still now. Since the research purpose and force-field formula are different between the small chemical compounds and proteins, several groups have developed new force fields like the general AMBER force field (GAFF) and CHARMm, which are applicable to various biological systems including protein, RNA, DNA and drug molecules (Wang et al. 2004; Zhu et al. 2012; Kumar et al. 2020).

These force fields can be used for the conformer generation of small molecules, and conformation generators CONCORD, Corina and CONFLEX have been developed (Gasteiger et al. 1990; Osawa et al. 1989; Kotev et al. 2005). Whereas conformer generation of chain structures is easy, treatment of ring puckering (conformer generation of cyclic structure) is a difficult problem (Cremer and Pople 1975; Cremer 1990). Recent cluster analysis revealed that the number of ring conformers is considerably smaller than $3^{N_{rot}}$, that the possible torsional angles of the ring main chain are limited, and that the number of

typical ring conformers increases slowly with increasing the ring member atoms (Friedrich et al. 2019; Chan et al. 2021). Such study may enable the fast ring-conformer generation including macrocycles and cyclic peptides. However, the conformers of bi-cyclic compounds are still unclear. Steric hindrance restricts the rotation around rotatable bonds (atropisomers). Use of atropisomers is an effective technique to increase the binding affinities of drug molecules, although atropisomers make the estimation of the entropy change, ΔS, upon binding difficult (Toenjes and Gustafson 2018).

Problem omitted in this chapter

In this chapter, we did not explain the solvation accompanying with quantum effects. Namely, charge transfer complexes, covalent drugs, halogen bonding, S–O interaction, metal bindings and so on. Currently, some docking programs can be applied to the covalent drugs, halogen bonding and metal binding. Chemical reactions of approved covalent drugs are mainly mild with targeting OH and SH groups of receptors and improve the drug potency and clearance (Bauer 2015).

Heavy halogen atoms (I, Br, and Cl) of ligand molecules can bind to both positively and negatively charged atoms. Each halogen atom of the molecules has one chemical bond in general and the opposite side of the halogen atom becomes positively charged, while the other part is done negatively (sigma hole effect). Thus, halogen atoms can bind to both positively and negatively charged atoms. The halogen bond is an electrostatic interaction and could be estimated in the framework of classical force fields (Harder et al. 2016).

S–O interaction is an intra-molecule π orbital interaction between sulfur and oxygen atoms. S–O interaction is useful to fix the ligand conformer to the active coordinates for increasing the receptor-ligand binding free energy. S–O interaction is a quantum effect, and there is no classical force field that can represent this effect currently (Nagao et al. 1998).

One of the most important interactions is the metal bindings, since many enzymes contain soft metal atoms, e.g., Zn, Cu, and Fe that can change the number of valence electrons with a small energy change in the reaction centers. In general, lone-pair electrons of ligand bind to metal atoms of enzyme, and additional point charges representing the lone-pair electrons of ligand atoms can reproduce the metalloprotease–ligand complex structure. But the covalency of the metal binding makes the binding energy prediction difficult comparing to the vdW and Coulomb interactions.

Docking scores as descriptors of the receptors and ligands: ensemble receptor-ligand docking and other applications

Numerous receptor-ligand associations and dissociations support life activities. Therefore, the receptor-ligand docking results can be useful descriptors for predicting various biological phenomena. Let think about the docking results among all receptors and all compounds in our body, where each docking result is a pair of the docking score (or ΔG) and the docking pose. Figure 4a is an interaction table between five receptors and five ligand molecules, and Fig. 4c is an interaction table between five conformers of a single receptor and the five ligand molecules. Assume that a matrix element $d(i,j)$ is assigned to the i th receptor and the j th compound in Fig. 4a. Similarly, a matrix element $d(i,j)$ is assigned to the i th conformer of a single receptor and the j th compound in Fig. 4c. The element $d(i,j)$ represents a pair of the docking score s_{ij} and the docking pose. The docking pose is described in many ways: The 3D Cartesian coordinates of the receptor-ligand complex structure, vector representations that represent receptor-ligand interactions, 3D and 4D grid representations that represent the distribution of ligand's substructures and so on (Deng et al. 2004; Fujita and Orita 2008). Here after, we call the docking results among the multiple-receptor structures and multiple-compounds summarized in Fig. 4a and b as "interaction table" and discuss some applications of the tables.

Let the first receptor in Fig. 4a be the target receptor. The red solid-line frame in Fig. 4a represents the docking results of the ligands to the target receptor. Sorting $d(1,j)$ ($j = 1$ to 5) in descending order of s_{1j} is the conventional docking screening by the score. Note that once the set of multiple docking data is described as a matrix, we can apply various mathematical matrix operations to the matrix. These operations correspond to the chemical applications of the interaction table. In the following section, we introduce some applications of the interaction table.

Docking screening for choosing target-selective molecules based on the interaction table

Because many kinds of receptors exist in our body, drug molecules must bind specifically to its target receptor. Otherwise, the low selectivity and off-target binding may cause a side effect or adverse effect. Suppose that

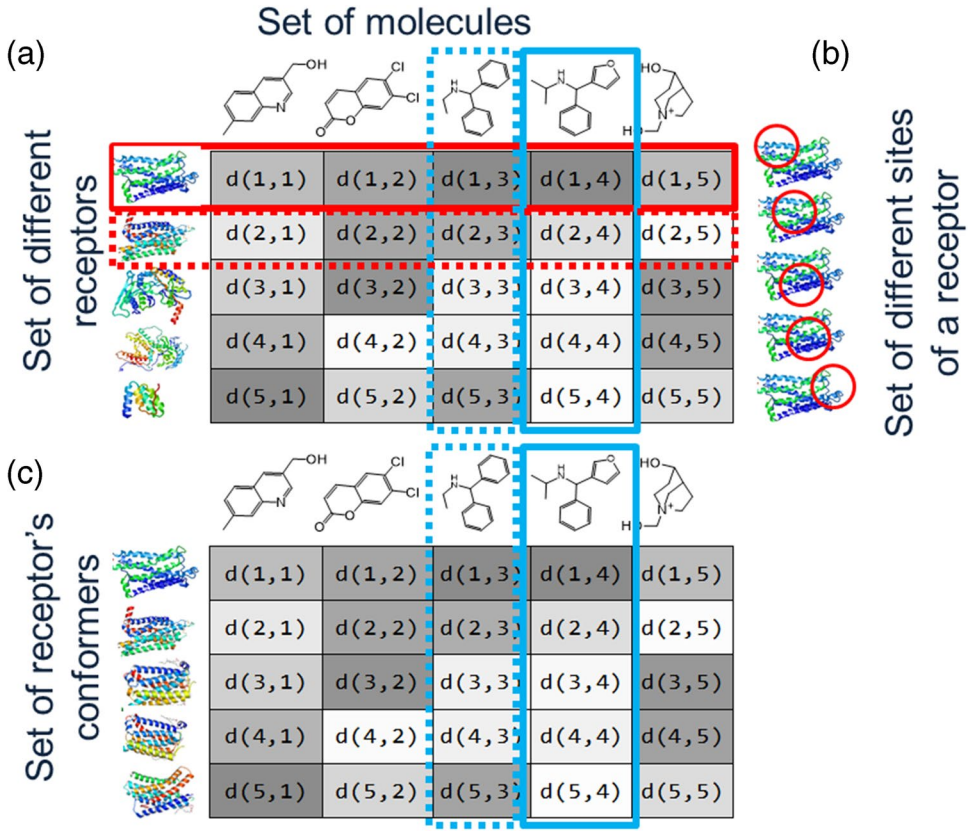

Fig. 4 **a** Interaction table. A matrix element $d(i,j)$, is docking results (docking score and docking pose) between the i th receptor and the j th ligand molecule obtained from the multiple-target screening and MASC scoring methods. In this table, i or $j = 1, \ldots, 5$. Ligand molecules in blue solid-line frame and in blue dotted-line frame are similar to each other. Receptors in red solid-line frame and in red dotted-line frame are similar to each other. In this panel, a darker tone assigned to $d(i,j)$ represent a higher score (higher affinity). **b** Red-colored circles indicate partial areas of a single receptor. Ligand docking is restricted in the red circles. **c** Interaction table for ensemble docking. Elements of this table are the docking results between conformers of a single target receptor in ensemble and a set of ligand molecules

Molecules A and B bind to a target receptor, and that the binding energy for Molecule A is stronger than that for Molecule B. If we select the molecule with the strongest binding energy, Molecule A should be the hit compound. However, if Molecule A bind to other receptors much more strongly than to the target receptor and if Molecule B does not bind to other receptors, we should select Molecule B as the hit compound. This example suggests that the in silico screening for a target receptor needs some additional docking studies for other receptors.

The multiple target screening (MTS) method is an in silico screening method to choose the target-selective molecules (Fukunishi et al. 2006b). This method is simple. Each molecule in the compound database (DB) is docked to proteins of a protein set (each column of the interaction table: Fig. 4a). The molecules with a higher score to the target than that to the other receptors are considered as target-selective molecules.

Figure 4a exemplifies that the fourth compound by the blue solid-line frame is the candidate hit molecule for the first receptor indicated by the red solid-line frame. If the rank of the docking score between the target receptor and the ligand represents the target-selectivity to the ligands, we can select the ligand whose docking score to the target receptor is top-ranked as candidate hit molecule.

The other method is to use a deviation of docking score (MASC score) instead of the intact docking score (Vigers and Rizzi 2004). The MASC scoring method assumes that each ligand molecule has its own average docking score to a set of receptors. Target-selective molecules should show strong docking scores to the target receptor, while the same molecules should show weak docking scores to the other receptor. Thus, the target-selectivity of a ligand molecule should correspond to the z-score (deviation) of the docking score to the target receptor among the docking scores to many receptors. The MASC scoring method selects the highest z-score molecules as the hit compounds.

Improvement of docking results by machine-learning approaches based on the interaction table

Similar receptors tend to bind to similar ligands: Subtypes of receptors belonging to the same protein family bind to the same or similar ligand (i.e., Kinase family, GPCR family etc.). Thus, we can expect that a weighted average of the docking scores of over multiple receptors

is more reliable than that based on a single receptor structure. The weight for each receptor depends on the degree of the similarity among the receptors. Then an averaged docking score, S_i^a, between a receptor a and a ligand i is defined as

$$S_i^a = \sum_b w_b^a s_i^b \tag{19}$$

where s_j^b is the docking score for complex of receptor a and ligand i (i.e., the score in the interaction table), and w_b^a is a weight representing the contribution of s_j^b to the averaged docking score S_i^a. The determination of w_a^b can be achieved by machine-learning approaches when teaching data sets are available, which are experimental assay results (Fukunishi et al. 2006a; Fukunishi 2009).

In fact, the Nakamura group applied the docking-score QSAR method to 107 kinase assay results registered in ChEMBL database, made the prediction models for the 107 kinases based on total 20,000 ligand molecules, and reported that the average error of ΔG prediction was 0.7 kcal/mol (Fukunishi and Nakamura 2012; Fukunishi et al. 2017). Interestingly, they started the study from a general equation for S_i^a (i.e., $S_i^a = f(\{w_b^a s_j^b\})$) and concluded that the simple linear equation (i.e., Eq. 19) is a good expression for S_i^a.

ChEMBL and PubChem are the most widely used public molecular-interaction repositories (Kim et al. 2021; Gaulton et al. 2017). ChEMBL31 includes the 20 million molecular interactions among 2.3 million chemicals and 15,072 proteins. PubChem does the 297 million bioactivities of 112 million compounds. Many prediction models have been developed based on these repositories (Fukunishi 2009).

Similarity searches of molecules, receptor binding sites and ligand-based drug screening based on interaction table

The interaction table determines both the similarities among receptors and those among ligand molecules. If the structures of the a th and b th receptors are similar, the vector for the a th receptor $V^a = \left\{ s_1^a, s_2^a, s_3^a, \ldots, s_{N_{mol}}^a \right\}$ is similar to that for the b th receptor $V^b = \left\{ s_1^b, s_2^b, s_3^b, \ldots, s_{N_{mol}}^b \right\}$, where N_{mol} is the number of ligand molecules in the interaction table. Thus, the ensemble of the vectors can be used for clustering the receptor ligand-binding sites. Similarly, if the structures of the i th and j th ligands are similar, the vector for the i th ligand molecule $U_i = \left\{ s_i^1, s_i^2, s_i^3, \ldots, s_i^{N_{rec}} \right\}$ is similar to that for the j th receptor $U_j = \left\{ s_j^1, s_j^2, s_j^3, \ldots, s_j^{N_{rec}} \right\}$, where N_{rec} is the number of receptors in the interaction table. Similarity of ligand molecules is useful for the ligand-based in silico drug screening (Fukunishi et al. 2005, 2006c).

Descriptors for docking poses: pharmacogram method and paring propensity of substructures

The docking poses are useful descriptors. The simplest 1D descriptor of docking pose is a SIFt vector. The original SIFt is a digitalized amino-acid sequence of the receptor in that the residues contacting to the ligand are set to 1 and the other residues to 0. There are many variations based on the original SIFt vector (Deng et al. 2004).

Fujita and Orita introduced a 4D grid or multi-color 3D grid descriptor to represent the docking pose (see Fig. 5) and developed an in silico drug screening method called "pharmacogram method" (Fujita and Orita 2008). Receptor–ligand binding depends on the receptor-specific pharmacophore. "Pharmacophore" is a spatial distribution of steric and electronic features that contribute mainly to the optimal receptor–ligand interactions. The docking screening methods sort the ligand molecules by their docking scores, and the higher-ranked molecules are rich of actual active molecules rather than the lower-ranked molecules in many cases. The docking poses of the top-ranked molecules are likely to have important receptor–ligand interactions that could be a part of the pharmacophore. Thus, we can predict the pharmacophore based on the docking screening.

Figure 5 illustrates the procedure of the pharmacogram method. The receptor structures are set to the same position. We put the receptor-ligand complex structure in the 3D grid box G and the ligand-binding site is put at the center of the box. The box G consists of sub cells that are divided by the grids and the box G is described as a matrix G (= $G(m,i,j,k)$ where m, i, j and k are integers: i, j and k specify the position of sub cell, and m does the type of atom involved in the grid). Here $G(m,i,j,k) = 1$ and $G(m,i,j,k) = 0$ mean that the mth atom type (or substructure) exists in the sub-cell (i,j,k) and not, respectively. The indexes of sub cell (i, j, k) correspond to the (x, y, z) coordinates in the Cartesian space.

We make the matrix G for each docking pose of the top-ranked molecules of the docking screening (Fig. 5a, b). And the average G matrix of these top-ranked molecules should give the pharmacophore of the target receptor (see Fig. 5c). Finally, the docking poses are evaluated and re-ranked to show how much the obtained pharmacophore is satisfied (Fig. 5d).

Most of the current deep-learning type docking methods adopt the pharmacogram type (Ragoza et al. 2017) or GOLD type pose descriptors (Pereira et al. 2016) to scoring the docking poses.

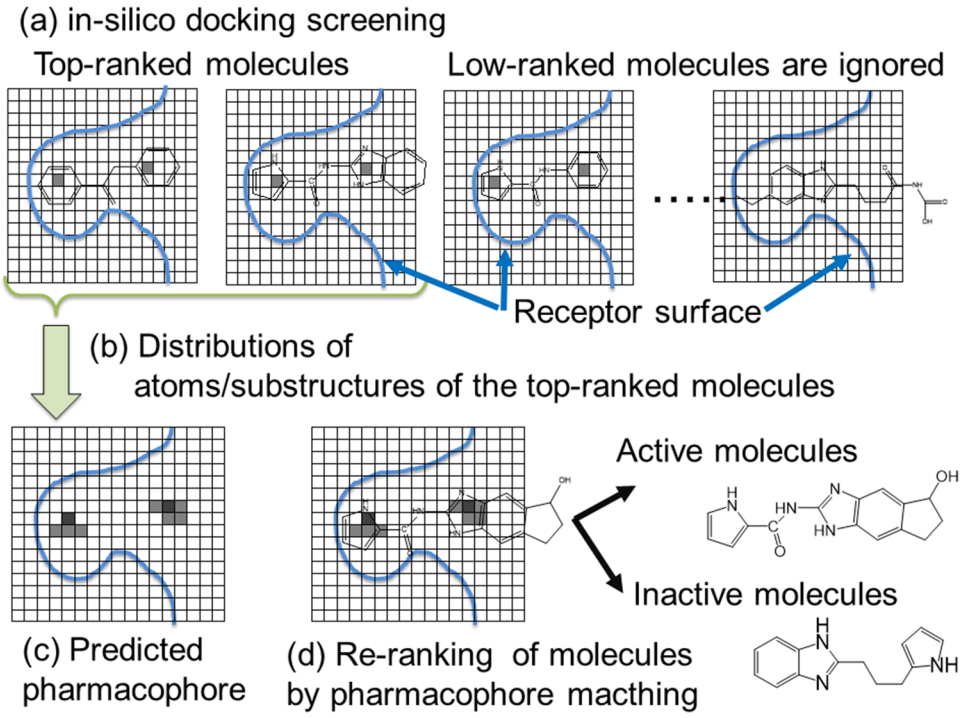

Fig. 5 Schematic representation of the pharmacogram method and the grid-type descriptor of docking pose. Although the grids are presented three-dimensionally originally, this figure is presented two-dimensionally

Prediction of ligand-binding sites of receptors based on the interaction table

Because the ligand-binding sites of enzymes and receptors are at a concave and hydrophobic, the sequences are conserved and the ligand-binding propensity of amino-acid residues at the ligand-binding sites shows clear trends mostly. An aromatic large residue is likely to bind to the ligand, although small residues are not: The trend of ligand-binding propensity is Trp > Phe > Tyr > His > Arg > … > Gly (Soga et al. 2007). Most of the ligand-binding site (pocket) prediction methods show high prediction accuracy by using these steric features and the amino-acid sequence information.

The conservation of amino-acid sequences and the 3D receptor–ligand complex structures in PDB suggest that the pocket shapes are classified into a limited number of shapes (so-called "pocketome") (Kufareva et al. 2012). The PoSSuM database summarized pairs of the receptor's pockets and their ligands (Ito et al. 2012).

Receptor–ligand docking should find the ligand-binding sites of the target receptor, to which native-ligand-like molecules binds. As mentioned in section "Docking screening for choosing target-selective molecules based on the interaction table," a wide variety of molecules can bind to receptors, regardless of binding energy. The MolSite method replaces the various receptors in Fig. 4a by the various sites of the target receptor described in Fig. 4b. Then, the Molsite method performs receptor–ligand docking of a set of small compounds including small drug molecules to the various sites of the receptor surface and predicts the site that exhibits the strongest docking scores (Fukunishi and Nakamura 2011).

Ensemble receptor–ligand docking

Structural dynamics are essential to realize the functions of enzymes and receptors. Namely, the ligand-binding sites change the structures during the ligand association and dissociation. These dynamics include the population shift, induced fit, local folding (coupled folding and binding), and so on. The definition of ΔG (Eq. 12) suggests that the docking results obtained from multiple receptor structures are closer to the reality than that from a single receptor structure.

Ensemble docking is a procedure where one ligand molecule binds to an ensemble of multiple receptor conformers to improve the accuracies of docking pose and binding activity prediction. Figure 4c shows the interaction matrix for the ensemble-docking screening. The ensemble of receptor structures can be obtained from various MD simulations such as the conventional MD simulation of receptor, generalized ensemble simulation, co-solvent MD simulations, and experiments (X-ray crystallography, liquid NMR, cryo-EM etc.).

The ensemble docking was considered when Kuntz's group developed the first docking program DOCK (Kuntz et al. 1982; Meng et al. 1992; Ferrari et al. 2004). The early ensemble docking replaced the grid potential from a single receptor structure by an average over multiple grid potentials of receptor structures, and then the docking using the

averaged grid potential was performed. This method did not increase the docking calculation cost. Currently, the ensemble-docking score is computed from the Boltzmann-weighted or simple average, and each docking score is obtained from the grid potential of each receptor structure of the ensemble (Knegtel et al. 1997).

A problem of ensemble docking is how to select the most suitable receptor structures from many structures in the ensemble since the in silico screening of millions or billion compounds are time consuming (Mohammadi et al. 2022). However, it is likely that a small number of receptor–ligand complex structures with strong binding energies contribute to a major part of the ΔG. The key point is a careful structural clustering of receptor conformers to decrease the number of candidate structures when some experimental active ligand molecules are available: When such active molecules are available, machine learning methods (i.e., random forest, naïve Bayesian model, deep learning) can make a rule for selection.

Molsite is also useful for the ensemble docking (Fukunishi et al. 2010). The Molsite method predicts receptor's surface sites that are likely to bind to ligand-like and drug-like molecules. Then, the predicted receptor sites are replaced by the receptor conformers in the ensemble. Suitable conformers are elected from conformers with high docking scores.

Remained problems: cryptic site

Each cell expresses several thousands of genes and many proteins produced by those genes are crowded in the cell. These proteins may interact randomly and conflict mutually (crowding effect). In this situation, exposed hydrophobic surfaces of proteins may cause non-selective protein–protein bindings. To avoid such bindings, the surfaces of the proteins are almost hydrophilic. Recently, binding sites that are exposed only when binding to a ligand or that appear transiently in an apo form have been investigated (Cimermancic et al. 2016; Beglov et al. 2018; Vajda et al. 2018). Such binding sites are called "cryptic sites" and may be one of mechanics of forming functional protein–protein complex structures like transcription factor complexes (Bekker et al. 2021b; Iida et al. 2020).

The conventional pocket prediction methods were not so useful to find the cryptic sites with using an apo form of a receptor. On the other hands, since the cryptic sites are functional, the amino-acid sequences around cryptic sites are conserved, and MD simulations show that the cryptic sites are transiently appear in 100–1000 ns at a room temperature (Frembgen-Kesner and Elcock 2006; Guo et al. 2016).

Iida et al. found that the ligand-binding propensity of amino-acid at the cryptic site is different from that of the conventional ligand-binding site. Namely, Tyr and Phe are the most popular in the cryptic site, although Trp is the most popular in the conventional ligand-binding site (Iida et al. 2020). With analyzing PDB statistically and using information from MD simulations, they proposed a "cryptic-site index" that provides the propensity of each amino-acid to be in the cryptic site. The cryptic site index showed that the aromatic residues (Tye > Phe > His) except Trp tend to be in the cryptic sites. In many cases, several 100 ns MD simulations at the room temperature are enough to find the cryptic sites at positions predicted by the cryptic-site index values. Some previous works showed that the chance of opening the cryptic sites increases with increasing the vdW interaction between the solvent water molecules and receptor atoms (SWISH method) (Oleinikovas et al. 2016). The combination of co-solvent MD, the cryptic site index, and ensemble docking may make the drug screening effective when the ligand-binding site is the cryptic site.

Enhanced sampling methods and molecular binding

Energy basins distribute in the conformational space and energy barriers hinder the inter-basin conformational transitions. As mentioned, when ensemble docking does not work because of the large conformational deformations/fluctuations of biomolecules during the complex formation, a powerful sampling method is required. One way is to use an MD-specialized computer such as ANTON (Shaw et al. 2008; Shaw et al. 2014) or MDGRAPE (Ohmura et al. 2014), MD Engine (Toyoda et al. 1999), or Express5800/MD server (Ohtaki et al. 2008). The other is to use an enhanced sampling (generalized ensemble) algorithm. In this review, we explain the latter because anyone can use a general-purpose computer.

To increase the sampling efficiency by algorithm, a generalized ensemble method such as a multicanonical method or a replica exchange method was proposed (Higo et al. 2012). The multicanonical algorithm was proposed first to study a physical system, a spin system, with using a MC simulation (Berg and Neuhaus 1992), applied to conformational motions of a biological system (Hansmann & Okamoto 1993; Kidera 1995), and incorporated in MD (Hansmann et al. 1996; Nakajima et al. 1997; Bartels and Karplus 1998). Similarly, the replica-exchange algorithm was developed to study a spin system using MC (Hukushima and Nemoto 1996) and applied to a biological system with using MD (Sugita and Okamoto 1999). Around the same time, several sampling methods, which have some similarity with the multicanonical or replica exchange methods, have been developed (Torrie and Valleau 1977; Paine and Scheraga 1985; Swendsen and Wang 1986; Mezei 1987; Lee 1993; Fukunishi et al. 1996; Iba et al. 1998; Wang and Landau

2001; Darve and Pohorille 2001; Laio and Parrinello 2002; Fukunishi et al. 2002; Hamelberg et al. 2004; Deng and Roux 2009; Moritsugu et al. 2010; Itoh and Okumura 2013; Peter and Shea 2014; Dasgupta et al. 2016; Kasahara et al. 2018; Ekimoto and Ikeguchi 2018; Higo et al. 2020b).

The enhanced sampling has a high efficiency to overcome energy barriers and importantly the method can assign a statistical weight equilibrated at a physiological temperature to any snapshot. Therefore, the resultant ensemble is equivalent to an equilibrated ensemble (canonical ensemble). By clustering the snapshots, one can identify basins in the conformational space, which means that Eq. 7 is computable. Suppose that the conformational space consists of three basins b_1, b_2, and b_3. Then, the free-energy ratio of F_{b_1} and F_{b_2} is expressed as:

$$\frac{F_{b_1}}{F_{b_2}} = \frac{\sum_i^{n_1} w_i^{b_1}}{\sum_i^{n_2} w_i^{b_2}}, \tag{20}$$

where $w_i^{b_j}$ and n_j were defined in Eq. 7. The normalization factor, which was omitted in Eq. 1, is cancelled out in Eq. 20.

Suppose that basin b_3 was sampled insufficiently (or not sampled at all). Even so, the ratio $\frac{F_{b_1}}{F_{b_2}}$ is computable correctly if b_1 and b_2 are sampled sufficiently. However, $\frac{F_{b_3}}{F_{b_j}}$ $(j = 1, 2)$ is computed inaccurately because of the insufficient data of b_3. Importantly, one may not notice this inaccuracy even after the simulation has finished. We note that such insufficiency occurs usually in minor basins fortunately. However, if the basin is a major one, main results from the sampling become misleading.

In enhanced sampling, a single or multiple reaction coordinates are introduced, which can be energy, temperature, Hamiltonian, other structural parameters (such as intermolecular distance or radius of gyration), or a virtual quantity for instance. In brief, the sampling is enhanced along the reaction coordinates by adding a bias potential along the reaction-coordinate axes or by controlling transition probability between different reaction-coordinate positions. The variation of the reaction coordinate(s) can be either continuous or discrete.

Application of the enhanced sampling to molecular binding is increasing (Sinko et al. 2013). Two-dimensional (temperature and Hamiltonian) replica exchange sampling was combined to the Rosetta docking (Zhang et al. 2015). The replica-exchange method was applied to poses obtained from Rosetta to detect stable complex conformations (poses) (Wang et al. 2017). To develop a drug for SARS-CoV-2, the in silico screening followed by MD simulation was applied to many existing drugs, and then metadynamics was applied to remaining poses to select better drug candidates (Kumawat et al. 2021; Namsani et al. 2021). Binding poses from ensemble docking were assessed by metadynamics to screen out false positives (Dandekar et al. 2021). However, it is still difficult to apply the enhanced sampling to many systems because this method requires a long computation time. Even so, enhanced sampling is useful to obtain details of the molecular binding process. Amyloid aggregation process was investigated by a replica-permutation method (Itoh and Okumura 2021). Metadynamics was applied to a protein–ligand binding phenomenon that accompanies an induced-fit conformational change (Zhao et al. 2021).

Nakamura and his coworkers introduced a generalized ensemble method, a multi-dimensional virtual-system coupled molecular dynamics (mD-VcMD) (Hayami et al. 2019), and then the Genetic Algorithm (GA) was incorporated to mD-VcMD, which was named "GA-mD-VcMD" (Higo et al. 2020b). In this method, the entire multidimensional reaction-coordinate space is divided into many small pieces (zones). The conformation (phase point) moves freely only in a zone for a while, and occasionally the phase point transitions to another zone using an inter-zone transition probability, which is defined by a user. This method was applied to some biological systems to elucidate ligand-receptor binding mechanisms and produce free-energy landscapes (Higo et al. 2019, 2020a, 2021; Hayami et al. 2021).

Here we introduce a study of GA-mD-VcMD applied to a middle-sized flexible drug, bosentan, binding to a GPCR molecule, human endothelin receptor type B (hETB) (Higo et al. 2022). Figure 6a illustrates the initial conformation of the simulation, where bosentan is far from hETB. The hRTB has a long N-terminal tail fluctuating largely in solution, and the root of the tail is located near the entrance of the gate of the binding pocket. The binding site is at the bottom of the pocket. Figure 6b demonstrates the resultant spatial density, $\rho_{MCb}(\boldsymbol{r})$, of the bosentan's centroid at position $\boldsymbol{r}$. The density was normalized so that $\rho_{CMb}(\boldsymbol{r})$ at the highest density position is 1.0. Apparently, the highest-density spot (region with $\rho_{CMb} \geq 0.5$; red-colored contours) corresponded to the bosentan's position in the native complex (crystal structure) (Shihoya et al. 2017). The density was still high in the binding pocket ($\rho_{CMb} \geq 0.1$). Although the density decreased with the ligand being apart from the binding pocket, this figure indicates that hETB and membrane affected bosentan even in a region far from hETB ($\rho_{CMb} \geq 0.0004$). Subsequent analyses showed that this long-range effect is caused by contacts of bosentan to the N-terminal tail of hETB.

The binding mechanism of this system is summarized as follows. First, bosentan and the N-terminal tail of hETB are fluctuating in solution (Fig. 7a). Then, the tip of the N-terminal tail of hETB captures bosemtan via nonspecific attractive interactions (Fig. 7b), which is called "fly casting" (Shoemaker et al. 2000; Sugase et al. 2007; Arai 2018). Next, bosentan slides occasionally from the tip to the root of the N-terminal tail (ligand–sliding) (Fig. 7c). During this sliding, bosentan passes the gate of the binding

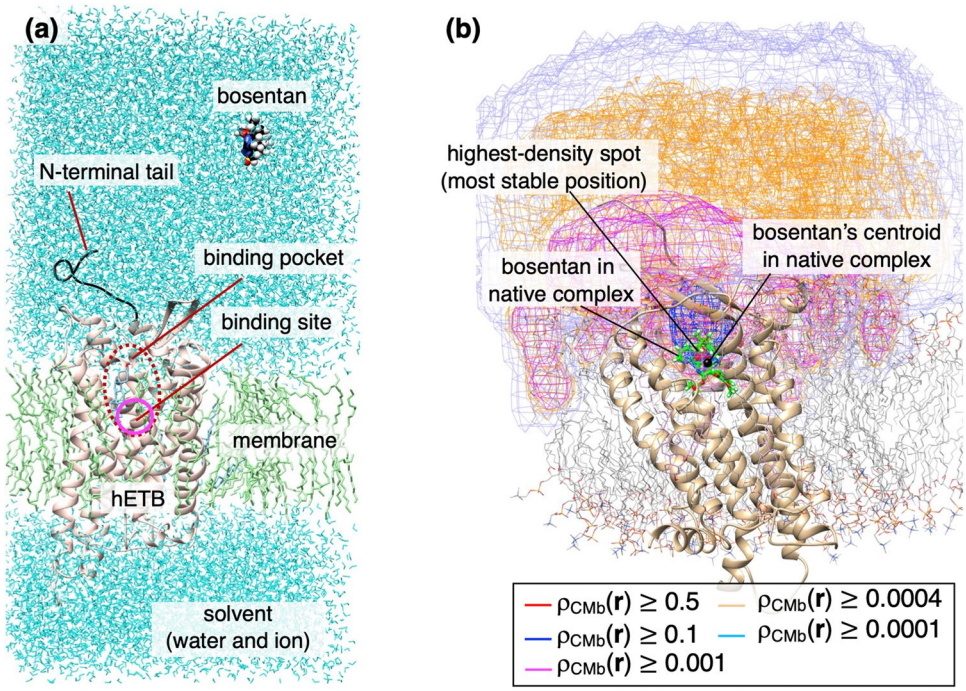

Fig. 6 **a** Initial conformation of simulation consisting of hETB, bosentan, membrane (cholesterol and POPC lipid molecules), and solvent (water molecules and ions). All molecules are flexible. **b** Spatial density of bosentan's centroid $\rho_{CMb}(\boldsymbol{r})$ at position $\boldsymbol{r}$. Iso-density surface is presented by five differently colored contours (see inset). Green-colored stick model and black sphere are, respectively, bosentan and the bosentan's centroid position in the native complex experimentally determined (PDB ID: 5xpr)

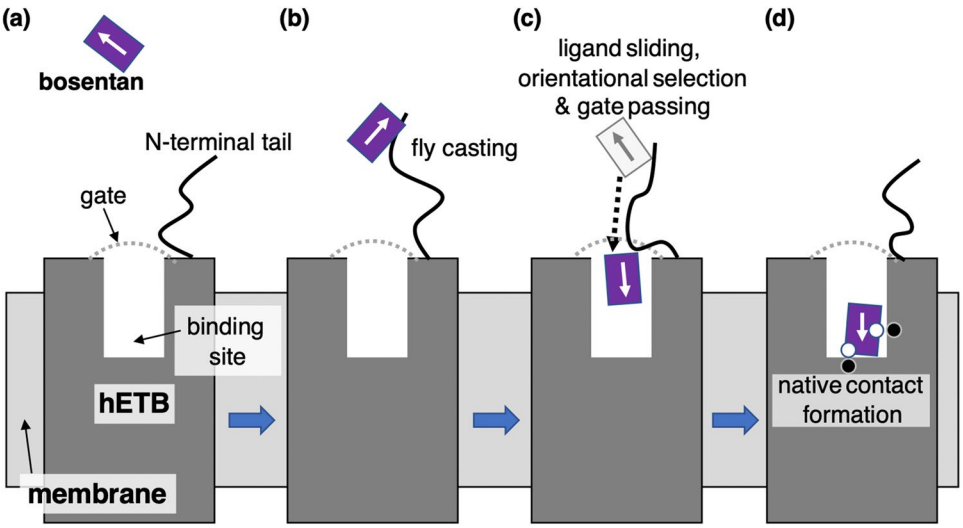

Fig. 7 Bosentan-hETB binding process follows panels as $(\boldsymbol{a}) \rightarrow (\boldsymbol{b}) \rightarrow (\boldsymbol{c}) \rightarrow (\boldsymbol{d})$. Arrow assigned to bosentan indicates its molecular orientation. Gate of the binding pocket is shown by gray dotted line. Ligand binding site is at the bottom of the binding pocket. Native attractive contacts are shown by pairs of open and filled small spheres in panel $(\boldsymbol{d})$

pocket, which accompanies rapid reduction of the molecular orientational variety of bosentan. This molecular orientational reduction, called a "orientational selection," is categorized to the population selection (Bosshard 2001; James and Tawfik 2003; Yamane et al. 2010), and consequently molecular orientations suitable for moving in the binding pocket toward the binding site are selected. Furthermore, this gate passing corresponds to overcoming a free-energy barrier in a free-energy landscape. When bosentan has reached the bottom of the pocket, attractive inter-molecular contacts are formed (formation of native contacts), which is the most thermodynamically stable complex (Fig. 7d). Details for this mechanism is reported in the paper (Higo et al. 2022).

Although the enhanced sampling (generalized ensemble) methods can assign a statistical weight to snapshots as mentioned above, the sampling requires a long simulation to obtain data that guarantee statistics accurate enough. One can perform multiple short runs instead of the long simulation, where the runs are distributed widely in the conformational space (Higo et al. 2009; Ikebe et al. 2011). However, the number of runs should be large when the system is complicated. For instance, the bosentan–GPCR simulation mentioned above, we performed 2000 runs. Therefore, it is still

difficult to perform the enhanced sampling for many computational researchers. We, however, believe that the applicability of the enhanced sampling methods increases because the computer power is increasing rapidly and steadily.

Binding free energy along a pathway: local sampling

As explained, the enhanced sampling method explores the conformational space widely with searching free-energy basins (binding poses). We refer to this approach as "global sampling" in this paper. The global sampling searches major basins to understand the binding process. Practically, on the other hand, the free-energy differences among the basins and the heights of free-energy barriers are not always estimated accurately when the computed system is large and complicated and when the simulation length is short. We suppose that the meshed area of Fig. 8a as well as the whole area of Fig. 8b are regions to be sampled by the global sampling.

If PMF is calculated along a pathway (line) in the real space, the volume to be sampled decreases drastically comparing with that sampled by the global sampling. We refer to this approach as a "line sampling," which is an extreme case of local sampling. Of course, the line sampling cannot discover basins out of the pathway. However, when the two conformations are set from the native complex and an unbound conformation, this method is useful to estimate the binding free energy.

Figure 8a presents schematically three pathways p_1, p_2, and p_3, each of which connects the most stable ligand position m_1 (the native-complex position) and a position m_5 in the unbound state. Figure 8b is a free-energy (PMF) landscape presented in the reaction-coordinate space. Remember that PMF is a quantity assigned to a position $\boldsymbol{q}$ (Eq. 6): $PMF = PMF(\boldsymbol{q})$. Then, the change of PMF from m_1 to m_5 is defined as $\Delta G = PMF(\boldsymbol{q}_{m_5}) - PMF(\boldsymbol{q}_{m_1})$, where $\boldsymbol{q}_{m_5}$ and $\boldsymbol{q}_{m_1}$ are respectively the positions of m_1 and m_5 in Fig. 8b in the reaction-coordinate space. In theory, ΔG is independent of the pathway. We, however, note that ΔG does not equivalent to the binding free-energy (free-energy difference between the native complex state and the full unfolded state). The free energy of the native-complex state is contributed by many conformations in the native-complex basin around m_1 (Eq. 7). Similarly, the free energy of the unbound state is contributed by many conformations in the unbound state. Furthermore, the binding free energy is measured in a solution that contains many identical receptors and identical ligands. Therefore, some corrections should be applied to ΔG. Wo do not explain the corrections in this paper. See a paper (Fukunishi 2009) for instance.

In fact, the free-energy profile was computed by a thermodynamic integration (Kirkwood 1935; Gelman and Meng 1998), a thermodynamic perturbation method (Zwanzig 1954; Beveridge and DiCapua 1989; Merz and Kollman 1989) or a weighted histogram analysis method (WHAM) (Kumar et al. 1992; Bartels 2000).

Practically, pathway setting is crucial to keep the accuracy of PMF in the line sampling. Problem of line sampling is that an appropriate pathway is unknown a priori. The pathway p_1 in Fig. 8a is simply set by a straight line between m_1 and m_5 in the real space. The corresponding pathway p_1 in the reaction-coordinate space is not necessarily straight, although it is straight in Fig. 8b. We

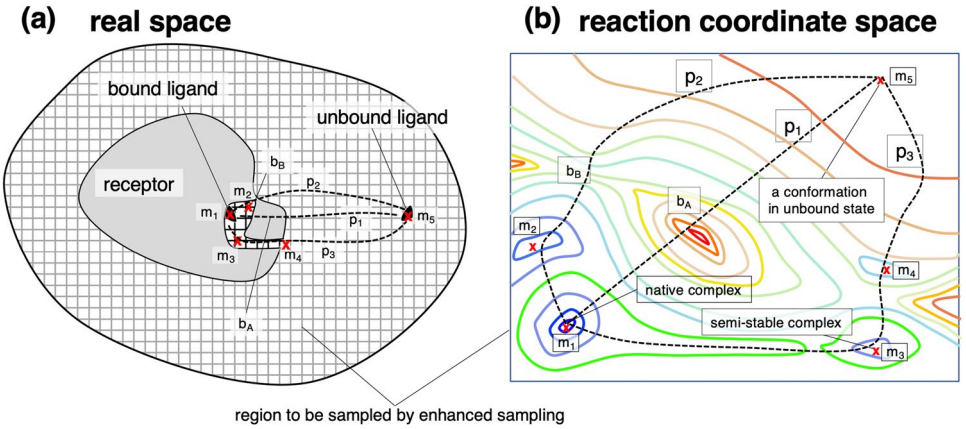

Fig. 8 **a** Binding/dissociating pathways in real space. Although the space is three-dimensional, it is presented two-dimensionally. Red-colored "×" labeled by m_i ($i = 1, \ldots, 5$) is as follows: m_1 is the most stable position of ligand (native-complex position), m_2, m_3 and m_4 are semi-stable positions, and m_5 is a conformation in the unbound state. Ligands at m_1 and m_5 are shown as "bound ligand" and "unbound ligand," respectively. Three ligand binding/dissociating pathways (p_1, p_2 and p_3) are indicated by broken lines, which connect m_1 and m_5. Meshed area is region to be sampled by global sampling. Labels b_A and b_B indicates positions of energy barriers along p_1 and p_2, respectively. **b** Free-energy landscape in reaction-coordinate space. Blue to red contour lines correspond to low to high PMF values. Meaning of p_i, b_A, b_B, "×" and m_i are the same as those in panel **a**

prepared Fig. 8 so that a high energy barrier b_A exists in p_1. Therefore, when the phase point is near b_A, a very strong force acts on the system, which causes a large numerical error in the resultant PMF. If the pathway is shifted slightly to a direction toward which the force decreases, then the numerical error decreases. By repeating this pathway resetting, the pathway may reach the pathway p_2 finally, because the barrier b_B in p_2 corresponds to a saddle point of PMF along p_2 (Fig. 8b). Therefore, the pathway resetting will not provide the pathway p_3, which is the best pathway, along which no remarkable barriers exist.

Nakamura and his coworkers proposed a method to escape high energy barriers in setting the pathway (Fukunishi et al. 2003). This sampling method consists of iterative simulations. An iteration (say iteration M) starts from the last conformation of iteration $M-1$, and the sampling is limited around the initial conformation by applying a restraint potential around the initial conformation: Sampling is localized around the initial conformation (effective range for the restraint potential is given by user). Furthermore, a repulsive potential is added at the vicinity of conformations sampled during the iteration. Besides, the repulsive potential is usually a Gaussian centered at the sampled conformations. With proceeding the iteration, the repulsive potential is accumulated in a low potential-energy region, and this region is gradually eliminated from sampling. This means that the simulation trajectory is not trapped in the low potential-energy region. On the other hand, very instable (high potential energy) regions (barrier b_A for instance) are also eliminated from sampling because of its high potential energy. When the next iteration (iteration $M+1$) is initiated, the repulsive potentials accumulated in iterations 1 to M are used in iteration $M+1$ from the beginning. Thus, the conformation does not return to a stable region, which was sampled in iterations 1 to M. The first iteration usually starts from a stable conformation (the native complex structure) and sampling continues till the phase point reaches an unbound conformation. By repeating the iterations and connecting the generated trajectories by the WHAM (Kumar et al. 1992; Bartels 2000), one can obtained a line in the 3D space, which connect the bound conformation to the unbound conformation. This method, named a "filling potential" method, is a procedure to escape conformational trapping and detour around high energy regions.

Although the filling-potential method produces a binding pathway along which rapid energy changes do not occur, the pathway looks like a random-work trajectory involving winding or loop-like curves. This may cause an unnecessarily computation. Then, Nakamura and his coworkers proposed a method to smoothen the random-like pathway by connecting the initial and final conformations by a linear combination of Legendre polynomials: Smooth-reaction path generation (SRPG) method (Fukunishi et al. 2009).

The idea of the filling potential is categorized in a Taboo search (Fred 1986). Around the same time, similar sampling methods to the filling potential method were proposed: Local elevation (Huber et al. 1994), conformational flooding (Grubmüller 1995), Wang–Landau sampling (Wang and Landau 2001), metadynamics (Laio and Parrinello 2002), and accelerated molecular dynamics (Hamelberg et al. 2004).

As explained above, the line sampling cannot discover out-of-pathway basins. Contrarily, the global sampling requires a high computational cost although it can discover various basins. To compensate the drawbacks of the two approaches, Nakamura and his coworkers proposed a local sampling method (Bekker et al. 2017). First, a cylinder is set in the system so that it covers both the ligand-binding site of receptor and an unbound position of ligand in solvent. Then, a multicanonical MD simulation is performed within the cylinder to obtain a free-energy landscape. Next, a low free-energy pathway is set by connecting the native-complex state and an unbound conformation in the resultant landscape. Note that the cylinder can be replaced by a body of an arbitral shape to define an appropriate pathway.

This method saves a computational time because the sampling is restricted in a volume enough to define the appropriate pathway. This method was applied to some systems: A ligand cyclin-dependent kinase 2 binding to a aminopyrazole inhibitor, yielding a binding free-energy error of 0.5 kcal/mol to the experimental value (Bekker et al. 2017), a medium-sized ligand 3MR binding to β-secretase 1 (error of 0.4 kcal/mol) (Bekker et al. 2019), and a peptide (about 10 residues long) from the amyloid-β peptide binding to an antibody solanezumab (error of 1.3 kcal / mol) (Bekker et al. 2020b). This procedure was also used to predict appropriate binding poses of some systems: Inhibitor binding to the N-terminal domain of heat-shock protein 90 (Bekker et al. 2020a), the Asian-dominant allele human leukocyte antigen binding to an HIV-1 Nef protein epitope (Bekker and Kamiya 2021), antagonist alprenolol binding to a GPCR, β2-adrenergic receptor (Bekker et al. 2021a), and two medium-sized inhibitors (ABT-737 and WEHI-539) binding to the cryptic site of Bcl-xL (Bekker et al. 2021b).

Conclusions

We reviewed various molecular binding methods from in silico screening to generalized ensemble methods. The in silico screening is a high throughput procedure because this method can provide binding poses of many ligand-receptor systems in a short time interval. This method is effective when the conformational change upon molecular binding is negligible in both the ligand and receptor. When a large conformational deformation occurs in receptor, ensemble

docking becomes useful because the ensemble may involve the deformed conformation of the receptor. When both the ligand and receptor are deformed considerably upon binding, a generalized ensemble method (global sampling) is useful. This approach, however, requires a considerable computational time. The line sampling or local sampling are methods to reduce the computation cost and to focus on a restricted region essential for the molecular binding process.

Nakamura has been contributing to the life science databases (PDB, eF-site, HitPredict etc.) and biomolecular simulation algorithms for understanding the life systems. His works follow the "Algorithms + Data Structures = Programs" and the DIKW pyramid (Data, Information, Knowledge, and Wisdom hierarchy) (Wirth 1976; Rowley 2007). Many people know about the Wirth's book and the DIKW Pyramid, but few spend their lives exploring it.

Funding The work was supported by Development of innovative drug discovery technologies for middle-sized molecules project from AMED (Y.F. and J.H), Project Focused on Developing Key Technology for Discovering and Manufacturing Drugs for Next-Generation Treatment and Diagnosis from AMED (Y.F. and J.H), and the Cooperative Research Program of the Institute for Protein Research, Osaka University, CR-21–05 (J.H). The work was supported by the HPCI System Research Project (hp220015 (J.H.) and hp220022 (K.K.)), and JSPS KAKENHI Grant No. 21K06052 (J.H.).

Declarations

Ethics approval This article does not contain any studies with human participants or animals performed by the author.

Conflict of interest The authors declare no competing interests.

References

Abugessaisa I et al (2021) FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. Nucleic Acids Res 49:D892–D898. https://doi.org/10.1093/nar/gkaa1054

Allinger NL (1976) Calculation of molecular structure and energy by force-field methods. In: Gold V, Bethell D (eds) Advances in physical organic chemistry, vol 13. Academic Press, London, pp 1–82. https://doi.org/10.1016/S0065-3160(08)60212-9

Allinger NL (1977) Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J Am Chem Soc 99:8127–8134. https://doi.org/10.1021/ja00467a001

Allinger NL, Zhou X, Bergsma J (1994) Molecular mechanics parameters. J Mol Struc 312:69–83. https://doi.org/10.1016/S0166-1280(09)80008-0

Amaro RE, Baudry J, Chodera J, Demir Ö, McCammon JA, Miao Y, Smith JC (2018) Ensemble docking in drug discovery. Biophys J 114:2271–2278. https://doi.org/10.1016/j.bpj.2018.02.038

Arai M (2018) Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. Biophys Rev 10:163–181. https://doi.org/10.1007/s12551-017-0346-7

Assaf KI, Nau WM (2018) The chaotropic effect as an assembly motif in chemistry. Angew Chem Int Ed 57:13968–13981. https://doi.org/10.1002/anie.201804597

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14:283–291. https://doi.org/10.1016/j.sbi.2004.05.004

Bartels C (2000) Analyzing biased Monte Carlo and molecular dynamics simulations. Chem Phys Let 331:446–454. https://doi.org/10.1016/S0009-2614(00)01215-X

Bartels C, Karplus M (1998) Probability distributions for complex systems: adaptive umbrella sampling of the potential energy. J Phys Chem B 102:865–880. https://doi.org/10.1021/jp972280j

Bauer RA (2015) Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. Drug Discov Today 20:1061–1073. https://doi.org/10.1016/j.drudis.2015.05.005

Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD (1998) Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins 33:367–382. https://doi.org/10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W

Beglov D, Hall DR, Wakefield AE, Luo L, Allen KN, Kozakov D, Whitty A, Vajda S (2018) Exploring the structural origins of cryptic sites on proteins. Proc Natl Acad Sci USA 115:E3416–E3425. https://doi.org/10.1073/pnas.1711490115

Bekker G-J, Kamiya N (2021) N-terminal-driven binding mechanism of an antigen peptide to human leukocyte antigen-A*2402 elucidated by multicanonical molecular dynamic-based dynamic docking and path sampling simulations. J Phys Chem B 125:13376–13384. https://doi.org/10.1021/acs.jpcb.1c07230

Bekker G-J, Kamiya N, Araki M, Fukuda I, Okuno Y, Nakamura H (2017) Accurate prediction of complex structure and affinity for a flexible protein receptor and its inhibitor. J Chem Theory Comput 13:2389–2399. https://doi.org/10.1021/acs.jctc.6b01127

Bekker G-J, Araki M, Oshima K, Okuno Y, Kamiya N (2019) Dynamic docking of a medium-sized molecule to its receptor by multicanonical MD simulations. J Phys Chem B 123:2479–2490. https://doi.org/10.1021/acs.jpcb.8b12419

Bekker G-J, Araki M, Oshima K, Okuno Y, Kamiya N (2020a) Exhaustive search of the configurational space of heat-shock protein 90 with its inhibitor by multicanonical molecular dynamics based dynamic docking. J Comput Chem 41:1606–1615. https://doi.org/10.1002/jcc.26203

Bekker G-J, Fukuda I, Higo J, Kamiya N (2020b) Mutual population-shift driven antibody-peptide binding elucidated by molecular dynamics simulations. Sci Rep 10:1406. https://doi.org/10.1038/s41598-020-58320-z

Bekker G-J, Araki M, Oshima K, Okuno Y, Narutoshi N (2021a) Accurate binding configuration prediction of a G-protein-coupled receptor to its antagonist using multicanonical molecular dynamics-based dynamic docking. J Chem Inf Model 61:5161–5171. https://doi.org/10.1021/acs.jcim.1c00712

Bekker G-J, Fukuda I, Higo J, Fukunishi Y, Kamiya N (2021b) Cryptic-site binding mechanism of medium-sized Bcl-xL inhibiting compounds elucidated by McMD-based dynamic docking simulations. Sci Rep 11:5046. https://doi.org/10.1038/s41598-021-84488-z

Bender BJ, Gahbauer S, Luttens A, Lyu J, Webb CM, Stein RM, Fink EA, Balius TE, Carlsson J, Irwin JJ, Shoichet BK (2021) A practical guide to large-scale docking. Nat Protoc 16:4799–4832. https://doi.org/10.1038/s41596-021-00597-z

Berg BA, Neuhaus T (1992) Multicanonical ensemble: A new approach to simulate first-order phase transitions. Phys Rev Lett 68:9–12. https://doi.org/10.1103/PhysRevLett.68.9

Beveridge DL, DiCapua FM (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem 18:431–492. https://doi.org/10.1146/annurev.bb.18.060189.002243

Bosshard HR (2001) Molecular recognition by induced fit: how fit is the concept? News Physiol Sci 16:171–173. https://doi.org/10.1152/physiologyonline.2001.16.4.171

Boyd RH (1968) Method for calculation of the conformation of minimum potential-energy and thermodynamic functions of molecules from empirical valence-force potentials--Application to the cyclophanes. J Chem Phys 49:2574–2583. https://doi.org/10.1063/1.1670456

Boyd DB (2013) Quantum chemistry program exchange, facilitator of theoretical and computational chemistry in pre-internet history. In: Strom ET, Wilson AK (eds) Pioneers of Quantum Chemistry, ACS Symposium Series 1122. American Chemical Society, Washington, DC, pp 221–273. https://doi.org/10.1021/bk-2013-1122.ch008

Bucher D, Grant BJ, McCammon JA (2011) Induced fit or conformational selection? The role of the semi-closed state in the maltose binding protein. Biochemistry 50:10530–10539. https://doi.org/10.1021/bi201481a

Campisi J, Kapahi P, Lithgow GJ, Melov S, Newman JC, Verdin E (2019) From discoveries in ageing research to therapeutics for healthy ageing. Nature 571:183–192. https://doi.org/10.1038/s41586-019-1365-2

Carlson HA, Masukawa KM, McCammon JA (1999) Method for including the dynamic fluctuations of a protein in computer-aided drug design. J Phys Chem A 103:10213–10219. https://doi.org/10.1021/jp991997z

Chan L, Hutchison GR, Morris GM (2021) Understanding ring puckering in small molecules and cyclic peptides. J Chem Inf and Model 61:743–755. https://doi.org/10.1021/acs.jcim.0c01144

Cimermancic P et al (2016) CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. J Mol Biol 428:709–719. https://doi.org/10.1016/j.jmb.2016.01.029

Cramer CJ, Truhlar DG (2008) A universal approach to solvation modeling. Acc Chem Res 41:760–768. https://doi.org/10.1021/ar800019z

Cremer D (1990) Calculation of puckered rings with analytical gradients. J Phys Chem 94:5502–5509. https://doi.org/10.1021/j100377a017

Cremer D, Pople JA (1975) General definition of ring puckering coordinates. J Am Chem Soc 97:1354–1358. https://doi.org/10.1021/ja00839a011

Dandekar BR, Sinha S, Mondal J (2021) Role of molecular dynamics in optimising ligand discovery: Case study with novel inhibitor search for peptidyl t-RNA hydrolase. Chem Phys Impact 3:100048. https://doi.org/10.1016/j.chphi.2021.100048

Darve E, Pohorille A (2001) Calculating free energies using average force. J Chem Phys 115:9169–9183. https://doi.org/10.1063/1.1410978

Dasgupta B, Nakamura H, Higo J (2016) Flexible binding simulation by a novel and improved version of virtual-system coupled adaptive umbrella sampling. Chem Phys Lett 662:327–332. https://doi.org/10.1016/j.cplett.2016.09.059

Delarue M et al (2018) mTORC1 controls phase separation and the biophysical properties of the cytoplasm by tuning crowding. Cell 174:338–349. https://doi.org/10.1016/j.cell.2018.05.042

Deng Y, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. Phys Chem B 113:2234–2246. https://doi.org/10.1021/jp807701h

Deng Z, Chuaqui C, Singh J (2004) Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein–ligand binding interactions. J Med Chem 47:337–344. https://doi.org/10.1021/jm030331x

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208. https://doi.org/10.1038/nrm1589

Ekimoto T, Ikeguchi M (2018) Multiscale molecular dynamics simulations of rotary motor proteins. Biophys Rev 10:605–615. https://doi.org/10.1007/s12551-017-0373-4

Falcon WE, Ellingson SR, Smith JC, Baudry J (2019) Ensemble docking in drug discovery: How many protein configurations from molecular dynamics simulations are needed to reproduce known ligand binding? J Phys Chem B 123:5189–5195. https://doi.org/10.1021/acs.jpcb.8b11491

Ferrari AM, Wei BQ, Costantino L, Shoichet BK (2004) Soft docking and multiple receptor conformations in virtual screening. J Med Mhem 47:5076–5084. https://doi.org/10.1021/jm049756p

Franz AK, Wilson SO (2013) Organosilicon molecules with medicinal applications. J Med Chem 56:388–405. https://doi.org/10.1021/jm3010114

Fred G (1986) Future paths for integer programming and links to artificial intelligence. Comput Oper Res 13:533–549. https://doi.org/10.1016/0305-0548(86)90048-1

Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? Drug Discov Today 13:869–874. https://doi.org/10.1016/j.drudis.2008.07.005

Frembgen-Kesner T, Elcock AH (2006) Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. J Mol Biol 359:202–214. https://doi.org/10.1016/j.jmb.2006.03.021

Friedrich NO, Flachsenberg F, Meyder A, Sommer K, Kirchmair J, Rarey M (2019) Conformator: a novel method for the generation of conformer ensembles. J Chem Inf Model 59:731–742. https://doi.org/10.1021/acs.jcim.8b00704

Friesner RA et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430

Fujita S, Orita M (2008) Method of searching for ligand. WIPO IP Portal. https://patentscope2.wipo.int/search/en/detail.jsf?docId=WO2008035729. Accessed 21 Nov 2022

Fukunishi Y (2009) Structure-based drug screening and ligand-based drug screening with machine learning. Comb Chem High Throughput Screen 12:397–408. https://doi.org/10.2174/138620709788167890

Fukunishi Y, Nakamura H (2011) Prediction of ligand-binding sites of proteins by molecular docking calculation for a random ligand library. Protein Sci 20:95–106. https://doi.org/10.1002/pro.540

Fukunishi Y, Nakamura H (2012) Statistical estimation of the protein-ligand binding free energy based on direct protein-ligand interaction obtained by molecular dynamics simulation. Pharmaceuticals 5:1064–1079. https://doi.org/10.3390/ph5101064

Fukunishi Y, Suzuki M (1996) Reproduction of the potential of mean force by a modified solvent-accessible surface method. J Phys Chem 100:5634–5636. https://doi.org/10.1021/jp9517615

Fukunishi Y, Tateishi T, Suzuki M (1996) Octane/water interfacial tension calculation by molecular dynamics simulation. J Colloid Interface Sci 180:188–192. https://doi.org/10.1006/jcis.1996.0288

Fukunishi H, Watanabe O, Takada S (2002) On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. J Chem Phys 116:9058–9067. https://doi.org/10.1063/1.1472510

Fukunishi Y, Mikami Y, Nakamura H (2003) The filling potential method: A method for estimating the free energy surface for protein-ligand docking. J Phys Chem B 107:13201–13210. https://doi.org/10.1021/jp035478e

Fukunishi Y, Mikami Y, Nakamura H (2005) Similarities among receptor pockets and among compounds: analysis and application to in silico ligand screening. J Mol Graph Model 24:34–45. https://doi.org/10.1016/j.jmgm.2005.04.004

Fukunishi Y, Kubota S, Nakamura H (2006a) Noise reduction method for molecular interaction energy: application to in silico drug screening and in silico target protein screening. J Chem Inf Model 46:2071–2084. https://doi.org/10.1021/ci060152z

Fukunishi Y, Mikami Y, Kubota S, Nakamura H (2006b) Multiple target screening method for robust and accurate in silico ligand screening. J Mol Graph Model 25:61–70. https://doi.org/10.1016/j.jmgm.2005.11.006

Fukunishi Y, Mikami Y, Takedomi K, Yamanouchi M, Shima H, Nakamura H (2006c) Classification of chemical compounds by protein−compound docking for use in designing a focused library. J Med Chem 49:523–533. https://doi.org/10.1021/jm050480a

Fukunishi Y, Mitomo D, Nakamura H (2009) Protein-ligand binding free energy calculation by the Smooth Reaction Path Generation (SRPG) method. J Chem Inf Model 49:1944–1951. https://doi.org/10.1021/ci9002156

Fukunishi Y, Ohno K, Orita M, Nakamura H (2010) Selection of in silico drug screening results by Using Universal Active Probes (UAPS). J Chem Inf Model 50:1233–1240. https://doi.org/10.1021/ci100108p

Fukunishi Y, Yamasaki S, Yasumatsu I, Takeuchi K, Kurosawa T, Nakamura H (2017) Quantitative Structure-Activity Relationship (QSAR) models for docking score correction. Mol Inform 36:1600013. https://doi.org/10.1002/minf.201600013

Gallicchio E, Kubo MM, Levy RM (2000) Enthalpy−entropy and cavity decomposition of alkane hydration free energies: numerical results and implications for theories of hydrophobic solvation. J Phys Chem B 104:6271–6285. https://doi.org/10.1021/jp0006274

Gasek NS, Kuchel GA, Kirkland JL, Xu M (2021) Strategies for targeting senescent cells in human disease. Nature Aging 1:870–879. https://doi.org/10.1038/s43587-021-00121-8

Gasteiger J, Rudolph C, Sadowski J (1990) Automatic generation of 3D-atomic coordinates for organic molecules. Tetrahedron Computer Methodology 3:537–547. https://doi.org/10.1016/0898-5529(90)90156-3

Gaulton A et al (2017) The ChEMBL database in 2017. Nucleic Acids Res 45:D945–D954. https://doi.org/10.1093/nar/gkw1074

Gelman A, Meng X-L (1998) Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Statist Sci 13:163–185. https://doi.org/10.1214/ss/1028905934

Gentile F, Yaacoub JC, Gleave J, Fernandez M, Ton AT, Ban F, Stern A, Cherkasov A (2022) Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. Nat Protoc 17:672–697. https://doi.org/10.1038/s41596-021-00659-2

Gilson MK, Sharp KA, Honig BH (1988) Calculating the electrostatic potential of molecules in solution: method and error assessment. J Comput Chem 9:327–335. https://doi.org/10.1002/jcc.540090407

Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 72:1047–1069. https://doi.org/10.1016/S0006-3495(97)78756-3

Goodsel DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 9:1–5. https://doi.org/10.1002/(sici)1099-1352(199601)9:1<1::aid-jmr241>3.0.co;2-6

Gorgulla C et al (2020) An open-source drug discovery platform enables ultra-large virtual screens. Nature 580:663–668. https://doi.org/10.1038/s41586-020-2117-z

Grubmüller H (1995) Predicting slow structural transitions in macromolecular systems: Conformational flooding. Phys Rev E 52:2893–2906. https://doi.org/10.1103/PhysRevE.52.2893

Guo Z, Thorarensen A, Che J, Xing L (2016) Target the more druggable protein states in a highly dynamic protein–protein interaction system. J Chem Inf Model 56:35–45. https://doi.org/10.1021/acs.jcim.5b00503

Halgren TA (1996a) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17:490–519. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P

Halgren TA (1996b) Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J Comput Chem 17:520–552. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<520::AID-JCC2>3.0.CO;2-W

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 47:1750–1759. https://doi.org/10.1021/jm030644s

Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 120:11919–11929. https://doi.org/10.1063/1.1755656

Hammes GG, Chang Y-C, Oas TG (2009) Conformational selection or induced fit: A flux description of reaction mechanism. Proc Nat Acad Sci USA 106:13737–13741. https://doi.org/10.1073/pnas.0907195106

Hansmann UAE, Okamoto Y (1993) Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem. J Comput Chem 14:1333–1338. https://doi.org/10.1002/jcc.540141110

Hansmann UHE, Okamoto Y, Eisenmenger F (1996) Molecular dynamics, Langevin and hydrid Monte Carlo simulations in a multicanonical ensemble. Chem Phys Lett 259:321–330. https://doi.org/10.1016/0009-2614(96)00761-0

Harder E et al (2016) OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. J Chem theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864

Hawkins GD, Cramer CJ, Truhlar DG (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. J Phys Chem 100:19824–19839. https://doi.org/10.1021/jp961710n

Hayami T, Higo J, Nakamura H, Kasahara K (2019) Multidimensional virtual-system coupled canonical molecular dynamics to compute free-energy landscapes of peptide multimer assembly. J Comput Chem 40:2453–2463. https://doi.org/10.1002/jcc.26020

Hayami T, Kamiya N, Kasahara K, Kawabata T, Kurita J, Fukunishi Y, Nishimura Y, Nakamura H, Higo J (2021) Difference of binding

modes among three ligands to a receptor mSin3B corresponding to their inhibitory activities. Sci Rep 11:6178. https://doi.org/10.1038/s41598-021-85612-9

Higo J, Kamiya N, Sugihara T, Yonezawa Y, Nakamura H (2009) Verifying trivial parallelization of multicanonical molecular dynamics for conformational sampling of a polypeptide in explicit water. Chem Phys Lett 473:326–329. https://doi.org/10.1016/j.cplett.2009.03.077

Higo J, Nishimura Y, Nakamura H (2011) A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations. J Am Chem Soc 133:10448–10458. https://doi.org/10.1021/ja110338e

Higo J, Ikebe J, Kamiya N, Nakamura H (2012) Enhanced and effective conformational sampling of protein molecular systems for their free energy landscapes. Biophysical Rev 4:27–44. https://doi.org/10.1007/s12551-011-0063-6

Higo J, Kasahara K, Wada W, Dasgupta B, Kamiya N, Hayami T, Fukuda I, Fukunishi Y, Nakamura H (2019) Free-energy landscape of molecular interactions between endothelin 1 and human endothelin type B receptor: fly-casting mechanism. Protein Eng Des Sel 32:297–308. https://doi.org/10.1093/protein/gzz029

Higo J, Kawabata T, Kusaka A, Kasahara K, Kamiya N, Fukuda I, Mori K, Hata Y, Fukunishi Y, Nakamura N (2020a) Molecular interaction mechanism of a 14-3-3 protein with a phosphorylated peptide elucidated by enhanced conformational sampling. J Chem Inf Model 60:4867–4880. https://doi.org/10.1021/acs.jcim.0c00551

Higo J, Kusaka A, Kasahara K, Kamiya N, Hayato I, Xie Q, Takahashi T, Fukuda I, Mori K, Hata Y, Fukunishi Y (2020b) GA-guided mD-VcMD: a genetic-algorithm-guided method for multi-dimensional virtual-system coupled molecular dynamics. Biophys Physicobiol 17:161–176. https://doi.org/10.2142/biophysico.BSJ-2020008

Higo J, Takashima H, Fukunishi Y, Yoshimori A (2021) Generalized–ensemble method study: A helix-mimetic compound inhibits protein–protein interaction by long-range and short-range intermolecular interactions. J Comput Chem 42:956–969. https://doi.org/10.1002/jcc.26516

Higo J, Kasahara K, Bekker G-J, Ma B, Sakuraba S, Iida S, Kamiya N, Fukuda I, Kono H, Fukunishi Y, Nakamura H (2022) Fly casting with ligand sliding and orientational selection supporting complex formation of a GPCR and a middle sized flexible molecule. Sci Rep 12:13792. https://doi.org/10.1038/s41598-022-17920-7

Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: A method for improving the searching properties of molecular dynamics simulation. J Comput-Aided Mol Des 8:695–708. https://doi.org/10.1007/BF00124016

Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Japan 65:1604–1608. https://doi.org/10.1143/JPSJ.65.1604

Iba Y, Chikenji G, Kikuchi M (1998) Simulation of lattice polymers with multi-self-overlap ensemble. J Phys Soc Japan 67:3327–3330. https://doi.org/10.1143/jpsj.67.3327

Iida S, Nakamura HK, Mashimo T, Fukunishi Y (2020) Structural fluctuations of aromatic residues in an apo-form reveal cryptic binding sites: implications for fragment-based drug design. J Phy Chem B 124:9977–9986. https://doi.org/10.1021/acs.jpcb.0c04963

Ikebe J, Umezawa K, Kamiya N, Sugihara T, Yonezawa Y, Takano Y, Nakamura H, Higo J (2011) Theory for trivial trajectory parallelization of multicanonical molecular dynamics and application to a polypeptide in water. J Comput Chem 32:1286–1297. https://doi.org/10.1002/jcc.21710

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. https://doi.org/10.1038/35057062

Irisa M, Takahashi T, Nagayama K, Hirata F (1995) Solvation free energies of non-polar and polar solutes reproduced by a combination of extended scaled particle theory and the Poisson-Boltzmann equation. Mol Phys 85:1227–1238. https://doi.org/10.1080/00268979500101791

Ito J-I, Tabei Y, Shimizu K, Tsuda K, Tomii K (2012) PoSSuM: a database of similar protein–ligand binding and putative pockets. Nucleic Acids Res 40:D541–D548. https://doi.org/10.1093/nar/gkr1130

Itoh SG, Okumura H (2013) Replica-permutation method with the Suwa–Todo algorithm beyond the replica-exchange method. J Chem Theory Comput 9:570–581. https://doi.org/10.1021/ct300791 9

Itoh SG, Okumura H (2021) Promotion and inhibition of Amyloid-β peptide aggregation: Molecular Dynamics Studies. Int J Mol Sci 22:1859. https://doi.org/10.3390/ijms22041859

James LC, Tawfik DS (2003) Conformational diversity and protein evolution – a 60-year-old hypothesis revisited. Trends Biochem Sci 28:361–368. https://doi.org/10.1016/S0968-0004(03)00135-X

Joedicke L et al (2018) The molecular basis of subtype selectivity of human kinin G-protein-coupled receptors. Nat Chem Biol 14:284–290. https://doi.org/10.1038/nchembio.2551

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748. https://doi.org/10.1006/jmbi.1996.0897

Jumper J et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Kabir SR, Yokoyama K, Mihashi K, Kodama T, Suzuki M (2003) Hyper-mobile water is induced around actin filaments. Biophys J 85:3154–3161. https://doi.org/10.1016/S0006-3495(03)74733-X

Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M (2021) KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 49:D545–D551. https://doi.org/10.1093/nar/gkaa970

Kang YK, Nemethy G, Scheraga HA (1987) Free energies of hydration of solute molecules. 1. Improvement of the hydration shell model by exact computations of overlapping volumes. J Phys Chem 91:4105–4109. https://doi.org/10.1021/j100299a032

Kasahara K, Shiina M, Higo J, Ogata K, Nakamura H (2018) Phosphorylation of an intrinsically disordered region of Ets1 shifts a multi-modal interaction ensemble to an out-inhibitory state. Nucleic Acids Res 46:2243–2251. https://doi.org/10.1093/nar/gkx1297

Kawai J et al (2001) Functional annotation of a full-length mouse cDNA collection. Nature 409:685–690. https://doi.org/10.1038/35055500

Khambata-Ford S, Liu Y, Gleason C, Dickson M, Altman RB, Batzoglou S, Myers RM (2003) Identification of promoter regions in the human genome by using a retroviral plasmid library-based functional reporter gene assay. Genome Res 13:1765–1774. https://doi.org/10.1101/gr.529803

Kidera A (1995) Enhanced conformational sampling in Monte Carlo simulations of proteins: Application to a constrained peptide. Proc Nat Acad Sci USA 92:9886–9889. https://doi.org/10.1073/pnas.92.21.9886

Kim S et al (2021) PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res 49:D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Kirkland JL, Tchkonia T (2020) Senolytic drugs: from discovery to translation. J Intern Med 288:518–536. https://doi.org/10.1111/joim.13141

Kirkwood JG (1935) Statistical mechanics of fluid mixtures. J Chem Phys 3:300–313. https://doi.org/10.1063/1.1749657

Kita Y, Nishibe H, Wang Y, Hashikawa T, Kikuchi SS, Mami U, Yoshida AC, Yoshida C, Kawase T, Ishii S, Skibbe H, Shimogori

T (2021) Cellular-resolution gene expression profiling in the neonatal marmoset brain reveals dynamic species-and region-specific differences. Proc Nat Acad Sci USA 118:e2020125118. https://doi.org/10.1073/pnas.2020125118

Knegtel RM, Kuntz ID, Oshiro CM (1997) Molecular docking to ensembles of protein structures. J Mol Biol 266:424–440. https://doi.org/10.1006/jmbi.1996.0776

Kotev MI, Goto H, Ivanov PM (2005) Molecular mechanics (CONFLEX/MM3) search/minimization study of the conformations of ornoside and escuside. J Mol Struct 748:9–16. https://doi.org/10.1016/j.molstruc.2005.03.016

Kufareva I, Ilatovskiy AV, Abagyan R (2012) Pocketome: an encyclopedia of sall-molecule binding sites in 4D. Nucleic acids Res 40:D535–D540. https://doi.org/10.1093/nar/gkr825

Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13:1011–1021. https://doi.org/10.1002/jcc.540130812

Kumar A, Yoluk O, MacKerell AD Jr (2020) FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules. J Comput Chem 41:958–970. https://doi.org/10.1002/jcc.26138

Kumawat A, Namsani S, Pramanik D, Roy S, Singh JK (2021) Integrated docking and enhanced sampling-based selection of repurposing drugs for SARS-CoV-2 by targeting host dependent factors. J Biomol Struct Dyn. https://doi.org/10.1080/07391102.2021.1937319

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161:269–288. https://doi.org/10.1016/0022-2836(82)90153-x

Kyogoku Y, Fujiyoshi Y, Shimada I, Nakamura H, Tsukihara T, Akutsu H, Odahara T, Okada T, Nomura N (2003) Structural genomics of membrane proteins. Acc Chem Res 36:199–206. https://doi.org/10.1021/ar0101279

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci USA 99:12562–12566. https://doi.org/10.1073/pnas.202427399

Lamb J et al (2006) The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 313:1929–1935. https://doi.org/10.1126/science.1132939

Lee J (1993) New Monte Carlo algorithm: Entropic sampling. Phys Rev Lett 71:211–214. https://doi.org/10.1103/PhysRevLett.71.211

Li J, Fu A, Zhang L (2019) An overview of scoring functions used for protein-ligand interactions in molecular docking. Interdiscip Sci 11:320–328. https://doi.org/10.1007/s12539-019-00327-w

Lumry R, Rajender S (1970) Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous properly of water. Biopolymers 9:1125–1227. https://doi.org/10.1002/bip.1970.360091002

Lyu J et al (2019) Ultra-large library docking for discovering new chemotypes. Nature 566:224–229. https://doi.org/10.1038/s41586-019-0917-9

Mallik B, Masunov A, Lazaridis T (2002) Distance and exposure dependent effective dielectric function. J Comput Chem 23:1090–1099. https://doi.org/10.1002/jcc.10104

Meng EC, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. J Comput Chem 13:505–524. https://doi.org/10.1002/jcc.540130412

Merz KM Jr, Kollman PA (1989) Free energy perturbation simulations of the inhibition of thermolysin: prediction of the free energy of binding of a new inhibitor. J Am Chem Soc 111:5649–5658. https://doi.org/10.1021/ja00197a022

Mestermann K, Giavridis T, Weber J, Rydzek J, Frenz S, Nerreter T, Mades A, Sadelain M, Einsele H, Hudecek M (2019) The tyrosine kinase inhibitor dasatinib acts as a pharmacologic on/off switch for CAR T cells. Sci Transl Med 11:eaau5907. https://doi.org/10.1126/scitranslmed.aau5907

Mezei M (1987) Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. J Comput Phys 68:237–248. https://doi.org/10.1016/0021-9991(87)90054-4

Mohammadi S, Narimani Z, Ashouri M, Firouzi R, Karimi-Jafari MH (2022) Ensemble learning from ensemble docking: revisiting the optimum ensemble size problem. Sci Rep 12:1–15. https://doi.org/10.1038/s41598-021-04448-5

Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: A plausible model. J Mol Biol 12:88–118. https://doi.org/10.1016/S0022-2836(65)80285-6

Moritsugu K, Terada T, Kidera A (2010) Scalable free energy calculation of proteins via multiscale essential sampling. J Chem Phys 133:224105. https://doi.org/10.1063/1.3510519

Mosalaganti S et al (2022) AI-based structure prediction empowers integrative structural analysis of human nuclear pores. Science 376:eabm9506. https://doi.org/10.1126/science.abm9506

Mourão MA, Hakim JB, Schnell S (2014) Connecting the dots: the effects of macromolecular crowding on cell physiology. Biophys J 107:2761–2766. https://doi.org/10.1016/j.bpj.2014.10.051

Musa A, Ghoraie LS, Zhang SD, Glazko G, Yli-Harja O, Dehmer M, Haibe-Kains B, Emmert-Streib F (2018) A review of connectivity map and computational approaches in pharmacogenomics. Brief Bioinformatics 19:506–523. https://doi.org/10.1093/bib/bbw112

Nagao Y, Hirata T, Goto S, Sano S, Kakehi A, Iizuka K, Shiro M (1998) Intramolecular nonbonded S···O interaction recognized in (Acylimino) thiadiazoline derivatives as angiotensin II receptor antagonists and related compounds. J Am Chem Soc 120:3104–3110. https://doi.org/10.1021/ja973109o

Nakajima N, Nakamura H, Kidera A (1997) Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. J Phys Chem B 101:817–824. https://doi.org/10.1021/jp962142e

Nakamura H (1988) Numerical calculations of reaction fields of protein-solvent systems. J Phys Soc Japan 57:3702–3706. https://doi.org/10.1143/JPSJ.57.3702

Nakamura H, Nishida S (1987) Numerical calculations of electrostatic potentials of protein-solvent systems by the self consistent boundary method. J Phys Soc Japan 56:1609–1622. https://doi.org/10.1143/JPSJ.56.1609

Nakamura H, Sakamoto T, Wada A (1988) A theoretical study of the dielectric constant of protein. Protein Eng Des Sel 2:177–183. https://doi.org/10.1093/protein/2.3.177

Namsani S, Pramanik D, Khan MA, Roy S, Singh JK (2021) Metadynamics-based enhanced sampling protocol for virtual screening: case study for 3CLpro protein for SARS-CoV-2. J Biomol Struct Dyn. https://doi.org/10.1080/07391102.2021.1892530

Nomura M, Uda-Tochio H, Murai K, Mori N, Nishimura Y (2005) The neural repressor NRSF/REST binds the PAH1 domain of the Sin3 corepressor by using its distinct short hydrophobic helix. J Mol Biol 354:903–915. https://doi.org/10.1016/j.jmb.2005.10.008

Nussinov R, Ma B, Tsai C-J (2014) Multiple conformational selection and induced fit events take place in allosteric propagation. Biophys Chem 186:22–30. https://doi.org/10.1016/j.bpc.2013.10.002

Ohmura I, Morimoto G, Ohno Y, Hasegawa A, Taiji M (2014) MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations. Phil Trans R Soc A 372:20130387. https://doi.org/10.1098/rsta.2013.0387

Ohtaki A et al (2008) Structure and molecular dynamics simulation of archaeal prefoldin: the molecular mechanism for binding and recognition of nonnative substrate proteins. J Mol Biol 376:1130–1141. https://doi.org/10.1016/j.jmb.2007.12.010

Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. Proc Natl Acad Sci USA 105:11182–11187. https://doi.org/10.1073/pnas.080252410

Oleinikovas V, Saladino G, Cossins BP, Gervasio FL (2016) Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. J Am Chem Soc 138:14257–14263. https://doi.org/10.1021/jacs.6b05425

Onufriev A, Case DA, Bashford D (2002) Effective Born radii in the generalized Born approximation: the importance of being perfect. J Comput Chem 23:1297–1304. https://doi.org/10.1002/jcc.10126

Ooi T, Oobatake M, Nemethy G, Scheraga HA (1987) Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Nat Acad Sci USA 84:3086–3090. https://doi.org/10.1073/pnas.84.10.3086

Osawa E, Goto H, Oishi T, Ohtsuka Y, Chuman T (1989) Application of molecular mechanics to natural product chemistry. Pure Appl Chem 61:597–600. https://doi.org/10.1351/pac198961030597

Pagadala NS, Syed K, Tuszynski J (2017) Software for molecular docking: a review. Biophys Rev 9:91–102. https://doi.org/10.1007/s12551-016-0247-1

Paine GH, Scheraga HA (1985) Prediction of the native conformation of a polypeptide by a statistical-mechanical procedure. I. Backbone structure of enkephalin. Biopolymers 24:1391–1436. https://doi.org/10.1002/bip.360240802

Pereira JC, Caffarena ER, Dos Santos CN (2016) Boosting docking-based virtual screening with deep learning. J Chem Info Model 56:2495–2506. https://doi.org/10.1021/acs.jcim.6b00355

Peter EK, Shea J-E (2014) A hybrid MD-kMC algorithm for folding proteins in explicit solvent. Phys Chem Chem Phys 16:6430–6440. https://doi.org/10.1039/C3CP55251A

Pierotti RA (1976) A scaled particle theory of aqueous and nonaqueous solutions. Chem Rev 76:717–726. https://doi.org/10.1021/cr60304a002

Pinzi L, Rastelli G (2019) Molecular docking: shifting paradigms in drug discovery. Int J Mol Sci 20:4331. https://doi.org/10.3390/ijms20184331

Porta-Pardo E, Ruiz-Serra V, Valentini S, Valencia A (2022) The structural coverage of the human proteome before and after AlphaFold. PLoS Comput Biol 18:e1009818. https://doi.org/10.1371/journal.pcbi.1009818

Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR (2017) Protein–ligand scoring with convolutional neural networks. J Chem Info Model 57:942–957. https://doi.org/10.1021/acs.jcim.6b00740

Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261:470–489. https://doi.org/10.1006/jmbi.1996.0477

Rashin AA (1989) Electrostatics of ion-ion interactions in solution. J Phys Chem 93:4664–4669. https://doi.org/10.1021/j100348a051

Rashin AA (1990) Hydration phenomena, classical electrostatics, and the boundary element method. J Phys Chem 94:1725–1733. https://doi.org/10.1021/j100368a005

Ravasio R, Flatt SM, Yan L, Zamuner S, Brito C, Wyart M (2019) Mechanics of allostery: contrasting the induced fit and population shift scenarios. Biophys J 117:1954–1962. https://doi.org/10.1016/j.bpj.2019.10.002

Regev A et al (2017) Science forum: the human cell atlas. elife 6:e27041. https://doi.org/10.7554/eLife.27041

Rich RL, Myszka DG (2007) Higher-throughput, label-free, real-time molecular interaction analysis. Anal Biochem 361:1–6. https://doi.org/10.1016/j.ab.2006.10.040

Richmond TJ (1984) Solvent accessible surface area and excluded volume in proteins: analytical equations for overlapping spheres and implications for the hydrophobic effect. J Mol Biol 178:63–89. https://doi.org/10.1016/0022-2836(84)90231-6

Rowley J (2007) The wisdom hierarchy: representations of the DIKW hierarchy. J Info Sci 33:163–180. https://doi.org/10.1177/0165551506070706

Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD (2014) rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. PLoS Comput Biol 10:e1003571. https://doi.org/10.1371/journal.pcbi.1003571

Salis A, Ninham BW (2014) Models and mechanisms of Hofmeister effects in electrolyte solutions, and colloid and protein systems revisited. Chem Soc Rev 43:7358–7377. https://doi.org/10.1039/C4CS00144C

Salmaso V, Moro S (2018) Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. Front Pharmacol 9:923. https://doi.org/10.3389/fphar.2018.00923

Shafqat S, Chicas EA, Shafqat A, Hashmi SK (2022) The Achilles' heel of cancer survivors: fundamentals of accelerated cellular senescence. J Clin Investig 132:e158452. https://doi.org/10.1172/JCI158452

Shaw DE et al (2008) Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 51:91–97. https://doi.org/10.1145/1364782.1364802

Shaw DE et al (2014) Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In: Kellenberger P (ed) SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society, Piscataway, NJ, pp 41–53. https://doi.org/10.1109/SC.2014.9

Shihoya W, Nishizawa T, Yamashita K, Inoue A, Hirata K, Kadji FMN, Okuta A, Tani K, Aoki J, Fujiyoshi Y, Doi T, Nureki O (2017) X-ray structures of endothelin ETB receptor bound to clinical antagonist bosentan and its analog. Nat Struct Mol Biol 24:758–764. https://doi.org/10.1038/nsmb.3450

Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc Natl Acad Sci USA 97:8868–8873. https://doi.org/10.1073/pnas.160259697

Silva D-A, Bowman GR, Sosa-Peinado A, Huang X (2011) A role for both conformational selection and induced fit in ligand binding by the LAO protein. PLOS Comput Biol 7:e1002054. https://doi.org/10.1371/journal.pcbi.1002054

Sinko W, Lindert S, McCammon JA (2013) Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. Chem Biol Drug Des 81:41–49. https://doi.org/10.1111/cbdd.12051

Soga S, Shirai H, Kobori M, Hirayama N (2007) Use of amino acid composition to predict ligand-binding sites. J Chem Inf Model 47:400–406. https://doi.org/10.1021/ci6002202

Soriano-Ursúa MA, Das BC, Trujillo-Ferrara JG (2014) Boron-containing compounds: chemico-biological properties and expanding medicinal potential in prevention, diagnosis and therapy. Expert Opin Ther Pat 24:485–500. https://doi.org/10.1517/13543776.2014.881472

Spolar RS, Record MTJ (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science 263:777–784. https://doi.org/10.1126/science.8303294

Stouten PF, Frömmel C, Nakamura H, Sander C (1993) An effective solvation term based on atomic occupancies for use in protein simulations. Mol Simul 10:97–120. https://doi.org/10.1080/08927029308022161

Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021–1025. https://doi.org/10.1038/nature05858

Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151. https://doi.org/10.1016/S0009-2614(99)01123-9

Suzuki M, Shigematsu J, Fukunishi Y, Kodama T (1997) Hydrophobic hydration analysis on amino acid solutions by the microwave dielectric method. J Phys Chem B 101:3839–3845. https://doi.org/10.1021/jp962543u

Swendsen RH, Wang JS (1986) Replica Monte Carlo simulation of spin-glasses. Phys Rev Lett 57:2607–2609. https://doi.org/10.1103/PhysRevLett.57.2607

Toenjes ST, Gustafson JL (2018) Atropisomerism in medicinal chemistry: challenges and opportunities. Future Med Chem 10:409–422. https://doi.org/10.4155/fmc-2017-0152

Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends in Biochem Sci 33:2–8. https://doi.org/10.1016/j.tibs.2007.10.003

Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys 23:187–199. https://doi.org/10.1016/0021-9991(77)90121-8

Toyoda S, Miyagawa H, Kitamura K, Amisaki T, Hashimoto E, Ikeda H, Kusumi A, Miyakawa N (1999) Development of MD Engine: high-speed accelerator with parallel processor design for molecular dynamics simulations. J Comput Chem 20:185–199. https://doi.org/10.1002/(SICI)1096-987X(19990130)20:2<185::AID-JCC1>3.0.CO;2-L

Trzesniak D, Kunz APE, van Gunsteren WF (2007) A comparison of methods to compute the potential of mean force. Chem Phys Chem 8:162–169. https://doi.org/10.1002/cphc.200600527

Tuckerman ME (2010) Statistical mechanics: theory and molecular simulation. Oxford University Press, Oxford

Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D (2021) Genome-wide association studies. Nature Reviews Methods Primers 1:1–21. https://doi.org/10.1038/s43586-021-00056-9

Vajda S, Beglov D, Wakefield AE, Egbert M, Whitty A (2018) Cryptic binding sites on proteins: definition, detection, and druggability. Curr Opin Chem Biol 44:1–8. https://doi.org/10.1016/j.cbpa.2018.05.003

Vauquelin G, Maes D (2021) Induced fit versus conformational selection: from rate constants to fluxes… and back to rate constants. Pharmacol Res Perspect 9:e00847. https://doi.org/10.1002/prp2.847

Venter JC et al (2001) The sequence of the human genome. Science 291:1304–1351. https://doi.org/10.1126/science.1058040

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein–ligand docking using GOLD. Proteins 52:609–623. https://doi.org/10.1002/prot.10465

Vigers GP, Rizzi JP (2004) Multiple active site corrections for docking and virtual screening. J Med Chem 47:80–89. https://doi.org/10.1021/jm030161o

Vogt AD, Cera ED (2012) Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. Biochemistry 51:5894–5902. https://doi.org/10.1021/bi3006913

Wang F, Landau DP (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. Phys Rev Lett 86:2050–2053. https://doi.org/10.1103/PhysRevLett.86.2050

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174. https://doi.org/10.1002/jcc.20035

Wang H, Liu H, Cai L, Wang C, Lv Q (2017) Using the multi-objective optimization replica exchange Monte Carlo enhanced sampling method for protein–small molecule docking. BMC Bioinfo 18:327. https://doi.org/10.1186/s12859-017-1733-6

Wilson EB Jr (1941) Some mathematical methods for the study of molecular vibrations. J Chem Phys 9:76–84. https://doi.org/10.1063/1.1750829

Wirth N (1976) Algorithms + data structures = programs. Prentice-Hall, Englewood Cliffs, NJ

Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem 179:131–137. https://doi.org/10.1016/0003-2697(89)90213-3

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. J Mol Biol 293:321–331. https://doi.org/10.1006/jmbi.1999.3110

Yamane T, Okamura H, Nishimura Y, Kidera A, Ikeguchi M (2010) Side-chain conformational changes of transcription factor PhoB upon DNA binding: a population-shift mechanism. J Am Chem Soc 132:12653–12659. https://doi.org/10.1021/ja103218x

Yung-Chi C, Prusoff WH (1973) Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol 22:3099–3108. https://doi.org/10.1016/0006-2952(73)90196-2

Zhang Z, Schindler CEM, Lange OF, Zacharias M (2015) Application of enhanced sampling Monte Carlo methods for high-resolution protein-protein docking in Rosetta. PLOS ONE 10:e0125941. https://doi.org/10.1371/journal.pone.0125941

Zhao Q, Capelli R, Carloni P, Lüscher B, Li J, Rossetti G (2021) Enhanced sampling approach to the induced-fit docking problem in protein–ligand binding: the case of mono-ADPRibosylation hydrolase inhibitors. J Chem Theory Comput 17:7899–7911. https://doi.org/10.1021/acs.jctc.1c00649

Zhu X, Lopes PE, MacKerell AD Jr (2012) Recent developments and applications of the CHARMM force fields. Wiley Interdiscip Rev Computational Molecular Science 2:167–185. https://doi.org/10.1002/wcms.74

Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. J Chem Phys 22:1420–1426. https://doi.org/10.1063/1.1740409