



Research Article

Predicting fragment binding modes using customized Lennard-Jones potentials in short molecular dynamics simulations

Christopher Vorreiter¹ , Dina Robaa, Wolfgang Sippl^{*,2} 

Department of Medicinal Chemistry, Institute of Pharmacy, Martin-Luther-University of Halle-Wittenberg, Halle (Saale) 06120, Germany

ARTICLE INFO

Keywords:

Fragment-based drug discovery
Co-solvent molecular dynamics simulation
Hotspot analysis
Binding free energy calculations
Ligand docking

ABSTRACT

Reliable *in silico* prediction of fragment binding modes remains a challenge in current drug design research. Due to their small size and generally low binding affinity, fragments can potentially interact with their target proteins in different ways. In the current study, we propose a workflow aimed at predicting favorable fragment binding sites and binding poses through multiple short molecular dynamics simulations. Tailored Lennard-Jones potentials enable the simulation of systems with high concentrations of identical fragment molecules surrounding their respective target proteins. In the present study, descriptors and binding free energy calculations were implemented to filter out the desired fragment position. The proposed method was tested for its performance using four epigenetic target proteins and their respective fragment binders and showed high accuracy in identifying the binding sites as well as predicting the native binding modes. The approach presented here represents an alternative method for the prediction of fragment binding modes and may be useful in fragment-based drug discovery when the corresponding experimental structural data are limited.

1. Introduction

In Fragment-Based Drug Discovery (FBDD), molecules of limited size serve as starting points for the design of potent drugs [1]. Fragment screenings are usually carried out for a certain target in order to identify fragment hits that can be optimized and grown into lead structures [2]. In order to enable a structure-based approach in drug design, crystal structures of target proteins containing fragments co-crystallized within the binding pocket of interest are of high importance [3–5]. However, crystallization studies are complex, time-consuming and not always realizable leaving the binding mode of known fragmental binders unresolved [5,6]. Computer-Aided Drug Design (CADD) methods aim to fill the void of missing experimental data and predict the binding mode of respective compounds [7–9]. Fragments can be characterized by the so-called „rule of three“. The rule implies a molecular weight restriction to 300 Da, a limit of hydrogen bond donors and acceptors to three, respectively, and a ClogP value ≤ 3 . Fragment hits usually conform to the rule of three emphasizing its importance in fragment library design [10].

Co-solvent molecular dynamics (MD) simulations are a new and

evolving approach in structure-based drug design. The simulation of small organic molecules in the presence of a protein structure can give insight into potential binding sites [11,12]. In addition, favorable interactions within the binding spots can be detected, depending on the characteristics and functional groups of the used probe molecules [13]. The advantages of co-solvent MDs have been extensively demonstrated while applying the method to a variety of protein targets and for multiple purposes [14,15].

In our previous work, the performance of different probe molecules in sampling experimentally solved binding site conformations in Tudor domains was studied. It was concluded that especially hydrophobic co-solvents and probe molecules mimicking features of known inhibitors succeeded at this task [16]. The present study extends the work and investigates whether an optimized method is capable of simulating diverse fragment- to lead-sized probes while reproducing their resolved binding modes. Reported studies have shown that computationally less expensive methods like molecular docking do not deliver satisfying binding mode predictions for fragments [17–20]. Especially, blind docking approaches usually lead to wrong binding pose predictions mainly due to the incorrect identification of the binding sites [21,22]. A

* Corresponding author.

E-mail address: wolfgang.sippl@pharmazie.uni-halle.de (W. Sippl).

¹ orcid.org/0000-0002-6108-0994.

² orcid.org/0000-0002-5985-9261.

recent study reported the use of MD simulations to reproduce fragment binding modes [7]. However, in the reported study relatively high simulation times had to be chosen because only a single fragment molecule was simulated at a time.

In the current work, MD simulations were carried out using fragments in high concentration to allow the decrease of simulation time due to the increased probability of observable binding events. Filters, descriptors and QM/MM-GBSA calculations were applied in order to reliably predict the binding site as well as the correct binding pose in the investigated cases. A comparable approach using multiple fragments in the solvation phase has been described in a recent publication [9]. However, the fragment concentrations were chosen relatively low and multiple long MD simulations (≥ 200 ns) were necessary for observing the desired binding events. The method was not successful in all presented cases and was generally limited in fragment size. In contrast, we propose a method that takes a different approach and focuses more on multiple short MD simulations (20 ns) with slightly larger fragments at millimolar concentrations. The proposed workflow could be beneficial in cases where experimental elucidation of the binding modes of fragments or hit compounds has so far been unsuccessful.

In order to assess the performance of the proposed method, four epigenetic protein targets and associated fragment-binders were studied: the Tudor domain of SMN, the zinc finger ubiquitin-binding domain (ZnF-UBD) of HDAC6, the PWWP1 (Pro-Trp-Trp-Pro 1) domain of NSD3 and the YEATS domain of ENL were studied. All mentioned domains except HDAC6-ZnF-UBD are reader proteins that are capable of recognizing specific targets carrying post-translationally methylated lysine or arginine residues [23–25]. Meanwhile, HDAC6-ZnF-UBD plays a special role in lysosomal degradation by recruiting ubiquitinated protein aggregates to the respective degradation machinery [26]. SMN is primarily known to be involved in the pathogenesis of spinal muscular atrophy (SMA) [27], while the other targets are involved in cancer models, among others [25,28–43]. Therefore, all proteins represent attractive targets for fragment-based drug design. We selected target proteins with rather shallow binding pockets that are located near the surface to evaluate the performance of the *in silico* approach for rather challenging targets. For all investigated protein domains, fragment-like molecules which target the respective binding sites and hamper the binding to natural ligands were discovered [44–51].

2. Materials and methods

2.1. Protein preparation

All mentioned protein structures were retrieved from the Protein Data Bank (PDB) [52]. The structures were initially prepared using Schrödinger's Protein Preparation Wizard [53,54]. The preprocessing step included the following actions: assigning bond orders, adding missing hydrogen atoms, creating zero-order bonds to metals, filling in missing side chains, deleting water molecules, and capping the proteins' termini. If the prepared protein structure contained co-crystallized fragments, the respective protonation states were generated at pH 7 ± 2 . In the next step, the hydrogen bonding network was automatically optimized at pH 7. At last, the structures were relaxed during a restrained minimization step using the OPLS 2005 force field [55–57].

2.2. Fragment preparation

Schrödinger's LigPrep [58] was used to generate extended, low-energy fragment conformations. Thereby, all possible tautomers and protonation states were generated at pH 7 ± 2 using Epik [59–61], while specified chirality was retained. OPLS 2005 was chosen as the operating force field for minimization. Respective descriptors for testing the compliance with the rule of three were calculated using Marvin Sketch [62].

2.3. Simulation system setup

For each investigated fragment, 21 simulation systems were built using Packmol [63], respectively. The first system only included fragment molecules in mixture with TIP3P [64] water at the desired concentration of $0.2 \text{ mol} \cdot \text{l}^{-1}$ (Table S1). The residual 20 systems contained the protein of interest in ligand-unbound conformation surrounded by a fragment-water mixture (Table S1). The fragment positions were varied in each system by using random seeds in Packmol. If necessary, all systems were neutralized by adding either sodium or chloride ions. The systems were converted for usage in Amber with the aid of Amber Tools [65]. The proteins were prepared by pdb4amber [65] and parameterized according to the ff14SB force field [66]. In the case of HDAC6-ZnF-UBD, the zinc ions were parameterized using the Li/Merz ion parameters for + 2 to + 4 ions in TIP3P water (12–6 normal usage set) [67–72]. Antechamber [65] and Parmchk [65] were used to assign AM1-BCC [73, 74] atomic charges to the fragment molecules and to parameterize them according to the General Amber Force Field 2 [75]. At last, the periodic cuboid systems were reconstituted using Tleap [65]. The fragments' geometric centers were defined by CPPTRAJ [76]. MDGX [65] was used to place dummy atoms as close as possible to the identified geometric centers. To this end, vectors were defined that originate from the fragment heavy atoms. Finally, customized Lennard-Jones parameters were assigned to the dummy atoms using ParmEd [65].

2.4. Molecular dynamics simulations

All simulations were performed using Amber 22 [65]. The prepared systems were initially minimized in two steps. The first step included 1000 iterations of steepest descent, followed by 2000 steps of conjugate gradient minimization while restraining the proteins with a force constant of $10 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$. In the second step, the restraints were lifted and the systems underwent additional 2000 steps of steepest descent and 2000 steps of conjugate gradient minimization. Subsequently, the systems were heated to target temperature (300 K) through 100 ps of MD simulation while restraining the proteins again using the above-mentioned force constant. Thereby, constant volume periodic boundaries were applied and the velocities were randomly assigned. At target temperature, a pressure equilibration of the restrained systems was carried out, including 100 ps of simulation at a constant pressure of 1 bar. Afterwards, the systems were finally simulated for 20 ns at constant pressure without any restraints. Temperature equilibration and maintenance at 300 K was ensured by the Langevin thermostat using a collision frequency of 2 ps^{-1} and a time step of 2 fs. Moreover, all bonds involving hydrogen were constrained by the SHAKE algorithm [77]. Additionally, the Particle Mesh Ewald (PME) method [78,79] was applied; the nonbonded cutoff distance for long-range electrostatic interactions was varied as mentioned in the main text.

2.5. Trajectory analysis

First, all generated trajectories were prealigned considering the respective proteins' C_{α} atoms. The fragment ("mobility analysis") and protein RMSF values as well as the protein RMSD values were calculated by CPPTRAJ [76]. The production phases generated for a single protein were stripped down to the protein coordinates and concatenated into a pseudo-trajectory for the following binding site analysis. MDpocket (Fpocket) [80,81] was used for the detection of all binding pockets in the respective pseudo-trajectories. In the initial pocket exploration run, all transiently formed binding sites were mapped into a so-called frequency grid. MDpocket automatically extracted a pdb file out of this grid, including the grid points at an isocontour value of 0.5. These grid points were used to define the binding site locations and dimensions. In the subsequent pocket characterization run, the isolated binding pockets were further analyzed by identifying the pocket-constituting amino acid residues and by calculating pocket descriptors like the pocket volume

and the mean local hydrophobic density (MLHD). Representative fragment-protein structures (centroid frames) were extracted out of the individual production phases using the hierarchical agglomerative clustering algorithm provided by CPPTRAJ [76]. During the clustering process, the heavy-atom RMSD of the respective protein and fragment was used as distance metric.

2.6. QM/MM-GBSA calculations

The generated fragment binding poses were ranked by the QM/MM-GBSA interaction energies calculated by Amber 22 [65]. To this end, the extracted centroid frames were first minimized in implicit solvent using Sander [65], including 5000 steps of steepest descent and up to 5000 steps of conjugate gradient minimization. Afterwards, hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) calculations were carried out using the MMPBSA.py script [82] provided by Amber. The Generalized Born/Surface Area (GBSA) Model was applied for modeling the continuous solvent and calculating the solvation free energy [83,84]. More precisely, the HCT (Hawkins, Cramer, Truhlar) GB model was utilized [84–86]. Furthermore, all QM/MM calculations were carried out using the semi-empirical Hamiltonian PM6 [87] with added dispersion and hydrogen bond corrections (PM6-DH+) [88]. During the calculations, the fragment of interest and the pocket residues identified by MDpocket were defined as QM regions (Table S2). MM force fields were applied to the remaining part of the complex. ΔG_{bind} was computed based on Eqs. (1)–(3) [84,89,90].

$$G_{\text{gas}} = E_{\text{VDW}} + E_{\text{EL}} \quad (1)$$

$$G_{\text{solv}} = E_{\text{GB}} + E_{\text{SURF}} \quad (2)$$

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} = \Delta G_{\text{gas}} + \Delta G_{\text{solv}} + \Delta E_{\text{SCF}} \quad (3)$$

The QM/MM-GBSA calculations take into account the changes in the gas-phase free energy (ΔG_{gas}) and solvation free energy (ΔG_{solv}) upon complex formation. Additionally, changes in the Self-Consistent-Field Energy (ΔE_{SCF}) are considered. The gas-phase free energy (G_{gas}) is calculated by summing up the van der Waals (E_{VDW}) and electrostatic energies (E_{EL}). The polar solvation energy (E_{GB}) and the non-polar solvation energy (E_{SURF}) terms contribute to the solvation free energy (G_{solv}).

2.7. Further analysis of the MD-derived representative complexes

Phase VolCalc [91–93] was used to calculate the volume overlaps between the fragments of the centroid frames and the binding pockets identified by MDpocket. In this way, the low-mobility fragments could be assigned to the detected binding sites. The minimized fragment-protein complexes were imported into Schrödinger's Maestro [94] and aligned to the reference X-ray structures considering the C_{α} atoms of the binding site residues. Afterwards, heavy atom RMSD values were calculated with respect to the experimentally solved binding poses.

2.8. Docking study

Molecular docking was carried out using Molecular Operating Environment (MOE) [95] and the Schrödinger software suite. For blind and site-directed rigid body docking in MOE, the triangle matcher method and the London dG score were selected for ligand placement. For the refinement stage, the rigid receptor method and the GBVI/WSA dG score were chosen. Glide [96–99] as part of the Schrödinger software tools was used in standard precision (SP) mode to generate docking poses in rigid receptors. For blind docking in MOE, the whole unbound protein structure was defined as ligand binding site. Analogously, receptor grids were created in the Schrödinger software suite so that the midpoint cuboid covered the entire protein and the outer box featured additional 20 Å of side length for ligand placement. For site-directed

rigid body docking and induced fit docking, the coordinates of the superimposed co-crystallized fragments were used to specify the binding site locations (MOE) or to center the grid boxes (Schrödinger) whose dimensions were automatically adapted to ligand size. Induced fit docking in MOE was carried out by switching to the induced fit refinement method. Default settings were used including a cutoff distance of 6 Å and a tether weight of 10. The induced fit docking protocol [100–103] of Schrödinger implied the usage of Glide and Prime as software modules. Residues within 5 Å of the superimposed co-crystallized ligand were treated flexibly (default setting). For all docking runs, the prepared fragment structures were used as input and a maximum number of 20 output poses was specified. Heavy atom RMSD values were calculated for all generated docking poses with respect to the experimentally solved binding poses using Schrödinger's Maestro [94].

2.9. Molecular visualization and plot generation

All violin plots were generated using Matplotlib [104]. The graph visualizing the Lennard-Jones potential was created using GeoGebra [105]. All remaining plots were compiled by Microsoft Excel. All pictures showing 3D structures were generated by PyMol [106]. The 2D depictions of the fragments were exported from Marvin Sketch [62].

3. Results and discussion

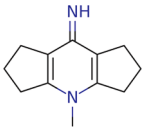
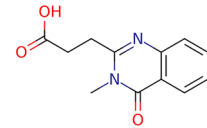
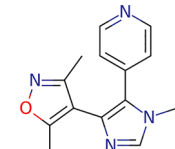
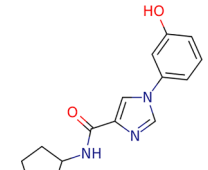
3.1. From co-solvents to fragments: adapting the cosolvent MD approach

In our previous work, we showed that crystallographically solved binding site conformations in Tudor domains could be successfully predicted using co-solvent MDs [16]. The study showed that hydrophobic co-solvents and probe molecules that mimic the characteristics of known binding partners were particularly successful in predicting the correct binding site and conformation. Hence, we sought to expand our studies by examining whether experimentally determined binding modes could be reproduced in the framework of short MD simulations. Since co-solvent MD simulations require mobile probe molecules which are able to probe the whole protein surface and stabilize within binding sites during relatively short simulation times, we focused in this work on binding modes of fragments, which are characterized by a restricted size and limited flexibility. More precisely, the investigated fragments predominantly conform to the so-called „rule of three“, a common rule for library design in fragment-based lead discovery (Table 1) [10].

In order to test the efficacy of the proposed method, four epigenetic targets were chosen for which crystal structures with co-crystallized fragments (reference X-ray structures) are available in the Protein Data Bank (PDB) (Table 1). All four targets show rather shallow binding pockets where the fragments are bound. The selected targets are: SMN-Tudor, HDAC6 ZnF-UBD, NSD3-PWWP1, and ENL-YEATS. For SMN-Tudor, HDAC6 ZnF-UBD, and NSD3-PWWP1, multiple crystal structures with co-crystallized fragments are available. Therefore, we focused on fragments showing the highest ligand efficiency (cpd. 1, cpd. 31, cpd. 9) [44,45,47]. Meanwhile, only one crystal structure of ENL-YEATS (PDB ID 8PJI) was co-crystallized with a fragment-like ligand (cpd. 10a) [48]. The objective was to simulate the respective protein structures in ligand-unbound conformation (Table 1) surrounded by an aqueous phase containing high concentrations of the selected fragments and observe to what extent the molecules reproduce the reference binding modes. In order to evaluate whether computationally less demanding methods show similar performance to the proposed MD-based method, three different docking approaches were performed for direct comparison: blind rigid body docking, site-directed rigid body docking and site-directed induced fit docking.

Table 1

Information about the selected fragments, their conformity to the rule of three and the retrieved PDB structures.

Fragment	Cpd. 1	Cpd. 31	Cpd. 9	Cpd. 10a
Target protein	SMN-Tudor	HDAC6 ZnF-UBD	NSD3-PWWP1	ENL-YEATS
PDB ID – unbound protein (initial MD coordinates)	1MHN	3C5K	6G3T	6HQ0
PDB ID – fragment-bound protein (ref. binding mode)	4QQ6	6CED	6G2C	8PJI
Fragment radius [Å]	3.5	4.8	4.1	6.6
Molecular weight [Da]	188	232	254	271
ClogP	1.98	0.82	1.16	2.25
H-bond acceptor count	2	4	3	3
H-bond donor count	1	0/1*	0	2
2D structures				

* Depending on the protonation state of the acidic group.

3.2. Preliminary studies: fragment distribution

In preliminary studies, the extent to which a homogenous fragment distribution could be achieved within the aqueous phase was investigated. Due to the larger size of the fragments compared to commonly used co-solvents, it was expected that the intermolecular attractive forces would dominate during the MD simulations, leading to fragment aggregation. The high probe concentration ($1 \text{ mol} \cdot \text{l}^{-1}$) used in our previous publication [16], which focused on rather small co-solvents, would therefore not be appropriate when using larger fragments. As a first measure to decrease clustering of the fragments, the probe concentration was reduced to $0.2 \text{ mol} \cdot \text{l}^{-1}$. In order to test the behavior of the fragments in water, MD simulations were carried out on systems

containing only water and fragment molecules in addition to the respective ions for neutralization. Despite the decrease in probe concentration, we still observed that the fragments formed aggregation clusters within the aqueous phase (Fig. 3). Therefore, an additional repulsive force had to be introduced to ensure a more homogenous fragment distribution.

3.3. Artificial Lennard-Jones potentials as a tool to enhance fragment repulsion

Diverse solutions to prevent probe aggregation in co-solvent MDs have already been reported in previous publications [107–109]. A common way is to place dummy atoms (DAs, massless particles) within

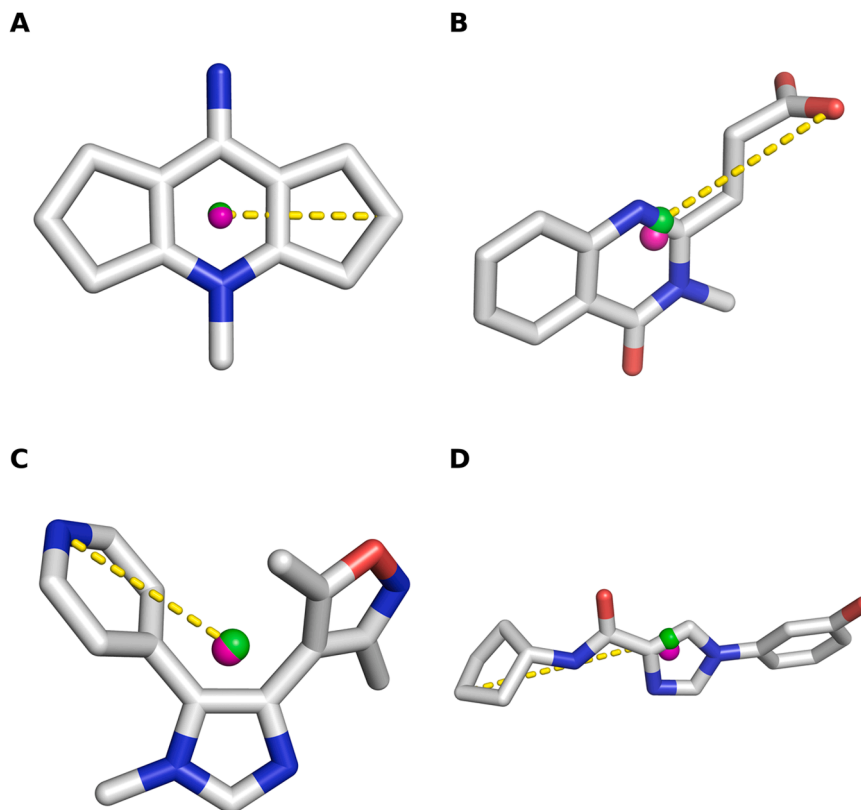


Fig. 1. Extended fragment conformations together with the identified geometric centers (magenta), the placed dummy atoms (green) and the determined fragment radii (dashed yellow lines). A: cpd. 1; B: cpd. 31; C: cpd. 9; D: cpd. 10a.

the probe molecules and to introduce artificial Lennard-Jones (LJ) potentials originating from these points. The DAs are assigned a separate LJ type so that the fragment-water and fragment-protein interactions remain unaffected.

All the herein used fragments are rather rigid structures or have a limited number of rotatable bonds. To place the DAs, the fully extended conformations were first generated using LigPrep and the geometric centers were determined using CPPTRAJ. MDGX, another component of AmberTools, was then used to place the DAs as close as possible to the geometric center by defining vectors between (rigid) fragment heavy atoms. Finally, the fragment radius r was defined by measuring the distance between the respective DAs and the farthest heavy atoms (Table 1, Fig. 1).

The LJ potential consists of a repulsive term dominating in near distances and an attractive term acting in far distances (Eq. (4)). The potential can be modeled using three parameters: R_{\min} , σ , and ϵ (Fig. 2). At the defined distance R_{\min} (equilibrium distance), the potential reaches a minimum and switches between repulsive and attractive. σ marks the particle distance at which the potential equals zero. R_{\min} and σ are related; therefore, only one of these parameters has to be defined. ϵ sets the potential well depth and affects the steepness of the curve [110–112].

$$V(x) = 4\epsilon \left[\left(\frac{\sigma}{x} \right)^{12} - \left(\frac{\sigma}{x} \right)^6 \right] = \epsilon \left[\left(\frac{R_{\min}}{x} \right)^{12} - 2 \left(\frac{R_{\min}}{x} \right)^6 \right] \quad (4)$$

To eliminate the attractive forces between the fragments, the attractive part of the LJ potential should be removed. In Amber, electrostatic interactions are usually split into short- and long-ranged components and calculated using the Particle Mesh Ewald (PME) method [78,79]. This approach uses a distance cutoff value that truncates the potential of the van der Waals (VDW) interactions in order to save computational cost [113]. Cutoff values between 8 and 10 Å are commonly used in MD simulations [66,79,114] since typical LJ potentials, e.g., describing VDW interactions, have already converged to zero at such distances [113]. In this study, the PME cutoff was set to R_{\min} of the newly introduced LJ potential in all MD simulations so that the repulsive part of the artificial LJ potential is completely covered while the attractive part at distances $> R_{\min}$ is discarded. Thereby, the PME cutoff was always chosen > 8 Å so that the repulsive and attractive parts of any other LJ potential within the simulation system still remained fully covered and unaffected by truncation. Since the newly introduced potentials were truncated at their minima, where the force acting on the dummy atoms equals zero, accuracy in energy calculation and energy conservation could be assumed. To prove the absence of major fluctuations or drifts in energy, the electrostatic, VDW, and total potential energies were exemplarily plotted for the fragment-water simulations (Fig. S1).

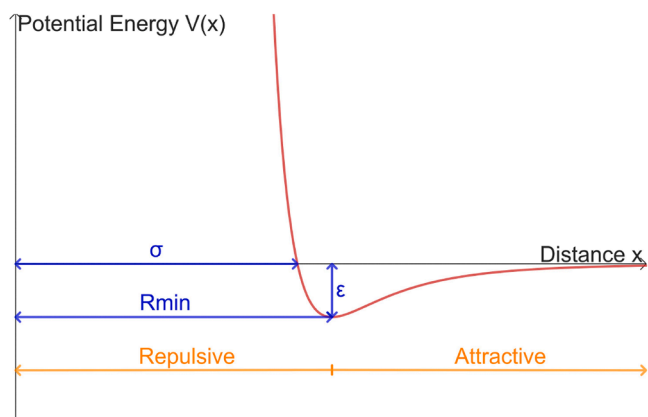


Fig. 2. The Lennard-Jones potential and its parameters.

3.4. Finding suitable parameters for the LJ potential

First, the LJ parameter σ was systematically varied and tested in MD simulations (20 ns) using the largest fragment, cpd. 10a, at target concentration ($0.2 \text{ mol} \cdot \text{l}^{-1}$). Radial distribution functions (RDFs) were calculated for the last 5 ns in order to check the distribution of the fragments within the aqueous phase. ϵ was constantly kept at $1 \text{ kcal} \cdot \text{mol}^{-1}$, while σ was gradually increased ($\sigma = 2r, 2r + 0.5, 2r + 1, 2r + 1.5, 2r + 2$ Å). The corresponding RDF plots reveal an equal probe distribution for all chosen σ values since all curves converge at $g(r) = 1$ at longer distances (Fig. 3). Ideally, we sought a σ value where $g(r)$ starts to increase at a distance $> 2r$ (the diameter of the fragment) since this would ensure that the fragment molecules are unable to come close to each other. This was fulfilled with σ values $\geq 2r + 1$ (Fig. 3). Since higher σ values would necessitate the use of larger PME cutoff values during the MD simulation which would unnecessarily increase the calculation times, a σ value of $2r + 1$ was chosen and subsequently tested for the remaining fragments. The respective RDF graphs confirm that the selected σ is suitable (Fig. 3). It is noteworthy that, unlike cpds. 1 and 31, the RDFs of cpds. 10a and 9 show prominent initial peaks. While the former cpds., 1 and 31, bear charged functional groups which would amplify the repulsion between the fragment molecules, cpds. 9 and 10a are uncharged. This can explain the observed initial peak in the respective RDF plots. Since the VDW attraction between the fragment molecules is not completely switched off, but just overridden by the repulsive LJ potential, a slight attraction (visible as an initial peak in the RDFs) is inevitable at close distances, especially with uncharged fragments.

3.5. MD simulations of protein-fragment systems

Based on the assigned LJ parameters (Table 2), the proteins of interest were simulated in the presence of the fragments in order to study whether the respective experimentally determined binding poses could be obtained. To this end, the systems were first prepared using Packmol; 20 systems (Table S1) per protein were built containing the unbound protein structure and the corresponding fragment at target concentration ($0.2 \text{ mol} \cdot \text{l}^{-1}$) (Table S1). In each system, the initial fragment positions were varied using random seeds in Packmol. All systems were simulated with Amber 22 for 20 ns and only the last 5 ns were considered as the production stage. The rationale behind selecting short simulation time was multifold. The high fragment concentration with changing initial positions increased the likelihood that a fragment molecule was initially situated in close proximity to the binding site of interest. Hence, the diffusion paths were assumed to be short. Moreover, the risk of observing extensive protein denaturation due to the high fragment concentrations was mitigated through the utilization of short simulation times. The first step in analyzing the generated trajectories included the verification of protein stability by calculating the C_{α} -RMSF and heavy atom RMSD values. The obtained RMSF plots reveal similar patterns to the corresponding B-factor plots (Fig. S2); high RMSF values could only be observed for flexible regions like loops or termini. Similarly, the RMSD plots of all simulated systems also indicate protein stability since the graphs quickly converge in all MD simulations (Fig. S3).

3.6. Identifying the binding pocket of interest

3.6.1. Fragment mobility analysis

As a first step in analyzing the obtained trajectories, the fluctuation of the fragments during the production stages was examined in order to identify potential candidates for binding mode reproduction. Low fluctuations (low-mobility fragments) are indicative of a higher ability to stably bind to the protein surface. Hence, RMSF values were calculated for each fragment in all systems to assess their mobility throughout the individual production phases. To filter out unstable or unbound

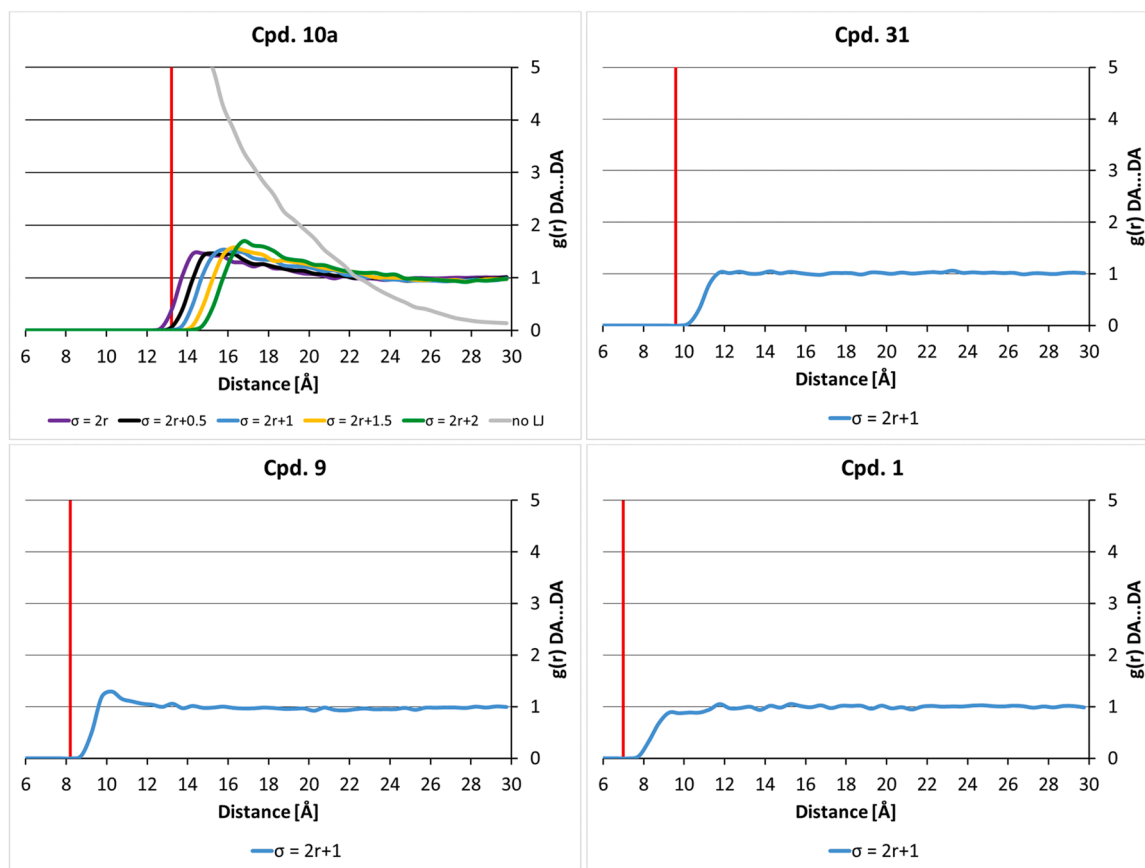


Fig. 3. Radial distribution functions (RDFs) calculated for the simulated systems containing fragment-water mixtures. The fragment diameter is indicated with a vertical red line. For Cpd. 10a, multiple simulations were carried out in which the LJ parameter σ was varied or the LJ potential was turned off.

Table 2

Determined Lennard-Jones parameters for each fragment.

Parameter	Cpd. 1	Cpd. 31	Cpd. 9	Cpd. 10a
ϵ [kcal * mol ⁻¹]	1.0	1.0	1.0	1.0
σ [Å]	8.0	10.6	9.2	14.2
R_{\min}/PME cutoff [Å]	9.0	11.9	10.3	15.9

Table 3

Low-mobility fragments (RMSF < 1.5 Å) assigned to the detected binding sites in SMN-Tudor via the calculated volume overlaps. The fragment binding poses within the pocket of interest (P2) are further characterized by RMSD values calculated with respect to the co-crystallized binding pose of cpd. 1 (PDB ID 4QQ6) as well as by the calculated QM/MM-GBSA interaction energies.

System no.	Fragment RMSF [Å]	Volume overlap [Å ³]		Fragment RMSD [Å]	QM/MM-GBSA [kcal * mol ⁻¹]
		P1	P2		
3	1.1	33	0	-	-
1	0.7	0	110	0.2	-24.4
17	1.4	0	103	0.8	-24.0
15	1.3	0	102	3.2	-23.6
11	0.8	0	92	0.3	-23.5
14	1.4	0	112	1.4	-22.8
10	1.5	0	103	2.2	-22.6
12	1.4	0	107	1.0	-21.7
6	1.1	0	118	3.1	-21.3
3	1.0	0	112	1.9	-21.2
16	1.3	0	108	1.6	-21.1
4	1.3	0	102	1.6	-20.7

Table 4

Low-mobility fragments (RMSF < 1.5 Å) assigned to the detected binding sites in HDAC6-ZnF-UBD via the calculated volume overlaps. The fragment binding poses within the pocket of interest (P2) are further characterized by RMSD values calculated with respect to the co-crystallized binding pose of cpd. 31 (PDB ID 6CED) as well as by the calculated QM/MM-GBSA interaction energies.

System no.	Fragment RMSF [Å]	Volume overlap [Å ³]			Fragment RMSD [Å]	QM/MM-GBSA [kcal * mol ⁻¹]
		P1	P2	P3		
3	0.7	0	195	0	0.9	-26.2
14	0.8	0	195	0	0.9	-19.6
16	1.1	0	194	0	3.4	-17.9
9	1.2	0	167	0	4.4	-16.8
10	0.8	0	196	0	3.4	-16.1
17	1.1	0	185	0	3.8	-12.9
11	0.9	0	165	0	5.0	-12.8
5	1.2	0	181	0	4.1	-12.4
2	0.9	0	152	0	3.8	-9.4
8	1.2	0	199	0	2.6	-9.0
20	0.9	0	196	0	3.7	-7.8
14	1.4	0	0	93	-	-
10	1.4	0	0	47	-	-

fragments, only fragment molecules showing RMSF values below 1.5 Å were further considered (Tables 3–6, S3, S4).

3.6.2. Pocket exploration and characterization

In order to identify pockets in the proteins, all production stages generated for a single protein were concatenated into pseudo-trajectories and subsequently analyzed by MDpocket. Additionally, pocket descriptors were calculated for all frames in the respective pseudo-trajectories.

Table 5

Low-mobility fragments (RMSF < 1.5 Å) assigned to the pocket of interest (P1) detected in NSD3-PWWP1 via the calculated volume overlaps. The fragment binding poses are further characterized by RMSD values calculated with respect to the co-crystallized binding pose of cpd. 9 (PDB ID 6G2C) as well as by the calculated QM/MM-GBSA interaction energies.

System no.	Fragment RMSF [Å]	Volume overlap [Å ³]						Fragment RMSD [Å]	QM/MM-GBSA [kcal * mol ⁻¹]
		P1	P2	P3	P4	P5	P6		
3	0.6	203	0	0	0	0	0	0.7	-27.3
20	1.3	192	0	0	0	0	0	0.6	-24.7
2	0.7	187	0	0	0	0	0	0.5	-24.6
18	0.7	198	0	0	0	0	0	1.0	-23.1
7	0.7	199	0	0	0	0	0	0.7	-22.0
15	1.0	185	0	0	0	0	0	3.5	-20.4
8	0.9	146	0	0	0	0	0	5.0	-20.2
6	0.9	128	0	0	0	0	0	6.2	-18.5
16	0.9	159	0	0	0	0	0	6.1	-17.6
11	0.9	183	0	0	0	0	0	5.9	-16.0
12	0.8	79	0	0	0	0	0	7.1	-15.9
8	1.1	96	0	0	0	0	0	7.4	-14.6
12	0.8	29	0	0	0	0	0	8.7	-14.4
10	1.4	163	0	0	0	0	0	5.1	-13.6
5	0.9	204	0	0	0	0	0	4.4	-11.5

Table 6

Low-mobility fragments (RMSF < 1.5 Å) assigned to the pocket of interest (P4) in ENL-YEATS via the calculated volume overlaps. The fragment binding poses are further characterized by RMSD values calculated with respect to the co-crystallized binding pose of cpd. 10a (PDB ID 8PJI) as well as by the calculated QM/MM-GBSA interaction energies.

System no.	Fragment RMSF [Å]	Volume overlap [Å ³]							Fragment RMSD [Å]	QM/MM-GBSA [kcal * mol ⁻¹]
		P1	P2	P3	P4	P5	P6	P7		
16	0.6	0	0	0	229	0	0	0	0.3	-33.5
2	1.1	0	0	0	235	0	0	0	0.4	-31.0
10	0.9	0	0	0	227	0	0	0	1.5	-25.5
5	1.0	0	0	0	227	0	0	0	1.6	-21.8
6	1.0	0	0	0	192	0	0	0	9.1	-18.1
14	1.5	0	0	0	70	0	0	0	11.0	-17.9

MDpocket identified two binding sites in SMN-Tudor, three sites in HDAC6-ZnF-UBD, six sites in NSD3-PWWP1, and seven sites in ENL-YEATS (Fig. 4). The additionally calculated pocket descriptors were used to further differentiate the detected cavities. Especially two descriptors seemed to be capable of indicating the binding pocket of interest: the pocket volume descriptor and the mean local hydrophobic density (MLHD) descriptor. The latter can be considered as an indicator for binding pocket druggability [115]. The correct binding site was always detected amongst the identified pockets, and showed notably higher values for pocket volume and MLHD than the other pockets (Fig. 5).

3.6.3. Binding pocket assignment

Our next step was aimed at filtering out all fragment molecules which do not bind in the detected pockets and localizing the actual binding site of the previously detected low-mobility fragments. To this end, Phase VolCalc was used to calculate volume overlaps between the pocket grid points and the fragments in the centroid structures extracted from the respective production phases. A table of fragment molecules assigned to the detected binding sites via the calculated volume overlaps was created for each protein (Tables 3–6, S3, S4). It is noteworthy that the highest volume overlap values are always linked to one binding site, namely the pocket of interest.

3.7. Identifying the binding pose of interest

3.7.1. Binding pose scoring

In the last step, we aimed to establish a score that is capable of highlighting the true fragment binding pose among all generated poses within the pocket of interest. To this end, QM/MM-GBSA calculations using Amber's MMPA.py application [82] were implemented. In

advance, all obtained centroid structures containing the representative fragment-protein complexes were minimized in implicit solvent. The pocket-forming residues detected by MDpocket were used to define the receptor QM region (Table S2). Besides, the respective fragment bound to the pocket of interest was treated quantum-mechanically. All single-point calculations were carried out using the minimized structures while applying the PM6-DH + Hamiltonian [87,88]. The resulting interaction energies are listed in Tables 3–6. Additionally, RMSD values were calculated for all minimized fragment poses with respect to the solved binding modes found in the reference X-ray structures to identify matching fragment poses (Tables 3–6).

With regard to the results obtained for SMN-Tudor, it is noteworthy that two binding pockets were detected on the protein surface (Fig. 4A). The majority of the low-mobility fragments were observed to bind to pocket two (P2), which corresponds to the pocket of interest (Table 3). The X-ray structure of SMN-Tudor co-crystallized with cpd. 1 (PDB ID 4QQ6) reveals that the fragment binds to an aromatic cage, constituted of W102, Y109, Y127, and Y130. It is sandwiched between W102 and Y130 undergoing pi-pi interactions. Cpd. 1 is additionally stabilized by cation-pi interactions with the surrounding aromatic residues. Moreover, a hydrogen bond is formed to N132 (Fig. 6A). The results of the simulations with cpd. 1 surrounding the apo protein structure 1MHN demonstrate that multiple fragment poses were obtainable within the aromatic cage. Most of them show RMSD values below 2 Å, suggesting a successful reproduction of the original binding mode. The top two scored poses as well as the fourth-ranked pose showed a good overlap with the experimentally solved binding mode which is evidenced by RMSD values < 1 Å (Table 3). Additionally, visual inspection confirmed the presence of all essential ligand-protein interactions (Figs. 7A, S4A, S4B). In contrast, the third ranked pose shows a higher RMSD value of 3.2 Å. Visual inspection of the respective fragment-protein complex

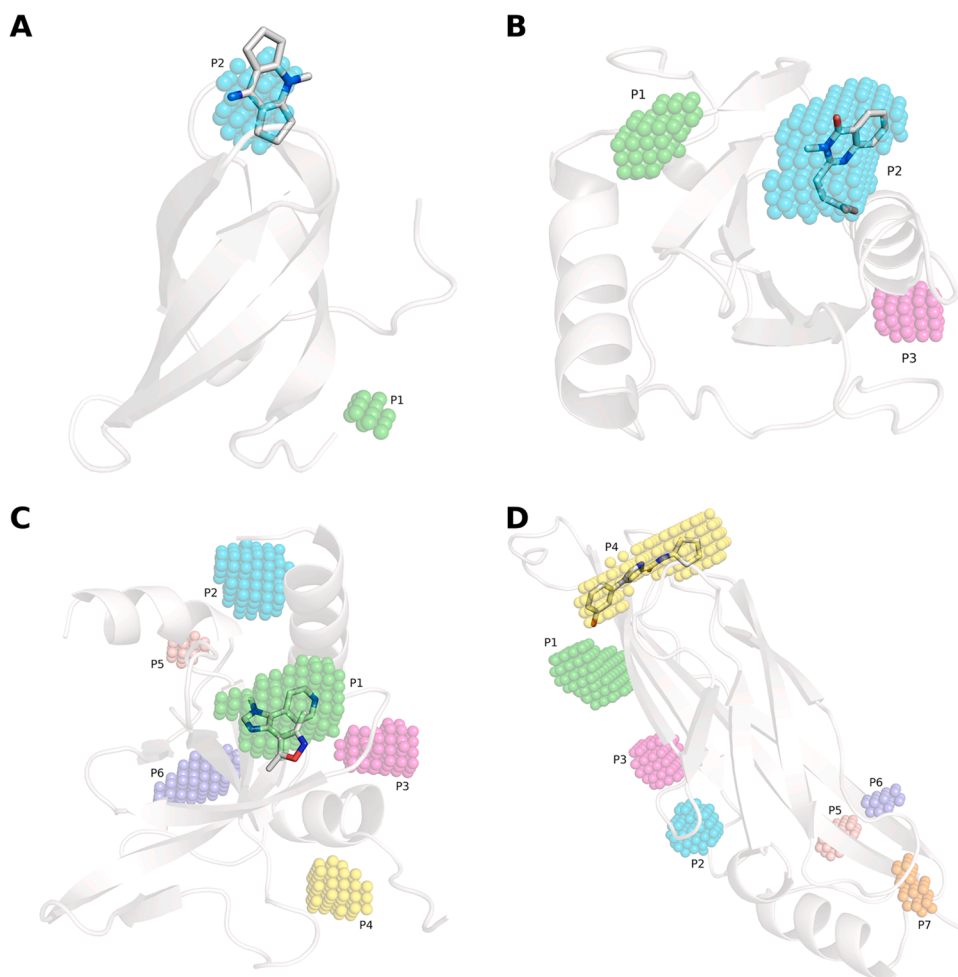


Fig. 4. Pocket frequency maps generated by MDpocket (isovalue = 0.5). The respective co-crystallized fragments are shown to indicate the binding pocket of interest. A: SMN-Tudor; B: HDAC6-ZnF-UBD; C: NSD3-PWWP1; D: ENL-YEATS.

revealed a flipped fragment orientation (Fig. 7B). However, the ligand still adopts the sandwiched position within the aromatic cage. Since all generated fragment poses show similar and typical interactions with the aromatic pocket, there is no significant difference in the calculated QM/MM-GBSA values (Table 3).

Three binding sites were detected in HDAC6-ZnF-UBD (Fig. 4B). Most of the low-mobility fragments were assigned to pocket two (P2), which was identified as the pocket of interest (Table 4). The experimentally solved binding mode of cpd. 31 (PDB ID 6CED) reveals that the fragment is sandwiched between R1155 and W1182 whereby the aromatic ring system is stabilized by pi-pi interactions with W1182. On the opposite side, the side chain of R1155 seems to be flexible since an alternative position is provided by the crystal structure. The fragment's carboxyl group forms hydrogen bonds to the backbone of R1155 and the phenolic hydroxyl group of Y1184 as well as a salt bridge to the side-chain of R1155 (Fig. 6B). The low RMSD values calculated for the top two scored poses suggest a successful binding mode reproduction (Table 4). The first pose reconstituted almost all crystallographically-observed ligand-receptor interactions, including the salt bridge between the fragment's carboxyl and the sidechain of R1155. The carboxylate group adopts a slightly different conformation and hence no hydrogen bond to Y1184 was formed (Fig. 8A). The second pose shows both hydrogen bonds to the backbone-carbonyl of R1155 and the side chain of Y1184, while the salt bridge to R1155 is absent (Fig. 8B) due to the different side chain conformation of R1155. This could explain the decreased QM/MM-GBSA interaction energy of the second pose as compared to the top-scored one. Overall, there is a significant difference

in the QM/MM-GBSA interaction energies calculated for native-like and nonnative binding modes (Table 4). This observation supports the hypothesis that the established score is able to differentiate between the generated poses and to highlight correct poses.

The trajectory analysis of simulations performed with NSD3-PWWP1 revealed six detectable pockets in the protein (Fig. 4C). Similar to the previously described findings, most of the low-mobility fragments were assigned to the pocket of interest, pocket one (P1, Tables 5, S3). The reference X-ray structure, including co-crystallized cpd. 9 (PDB ID 6G2C), demonstrates that the compound's binding mode is characterized by pi-pi interactions to Y281 and F312. Additionally, cpd. 9 is stabilized by a hydrogen bond to S314. Moreover, the isoxazole ring adopts a rather solvent-exposed position within the binding site (Fig. 6C). The five top-scored fragment poses showed RMSD values ≤ 1 Å with respect to the co-crystallized ligand proving a good overlap with the experimentally determined binding pose (Table 5). Visual inspection of the poses also confirmed the presence of essential ligand-receptor interactions (Figs. 9A, S5). In this case, we could also observe that the calculated QM/MM-GBSA interaction energy is effective in differentiating between native and non-native binding poses (Table 5).

The last of the investigated proteins, ENL-YEATS, revealed seven binding sites (Fig. 4D). The highest volume overlaps between fragments and binding sites were observed for pocket four (P4), the pocket of interest (Tables 6, S4). When focusing on the calculated pocket descriptors, it is observable that the volume and MLHD values of pocket one also tend to relatively high values (Fig. 5D). However, no low-mobility fragments showed any overlap with this particular pocket

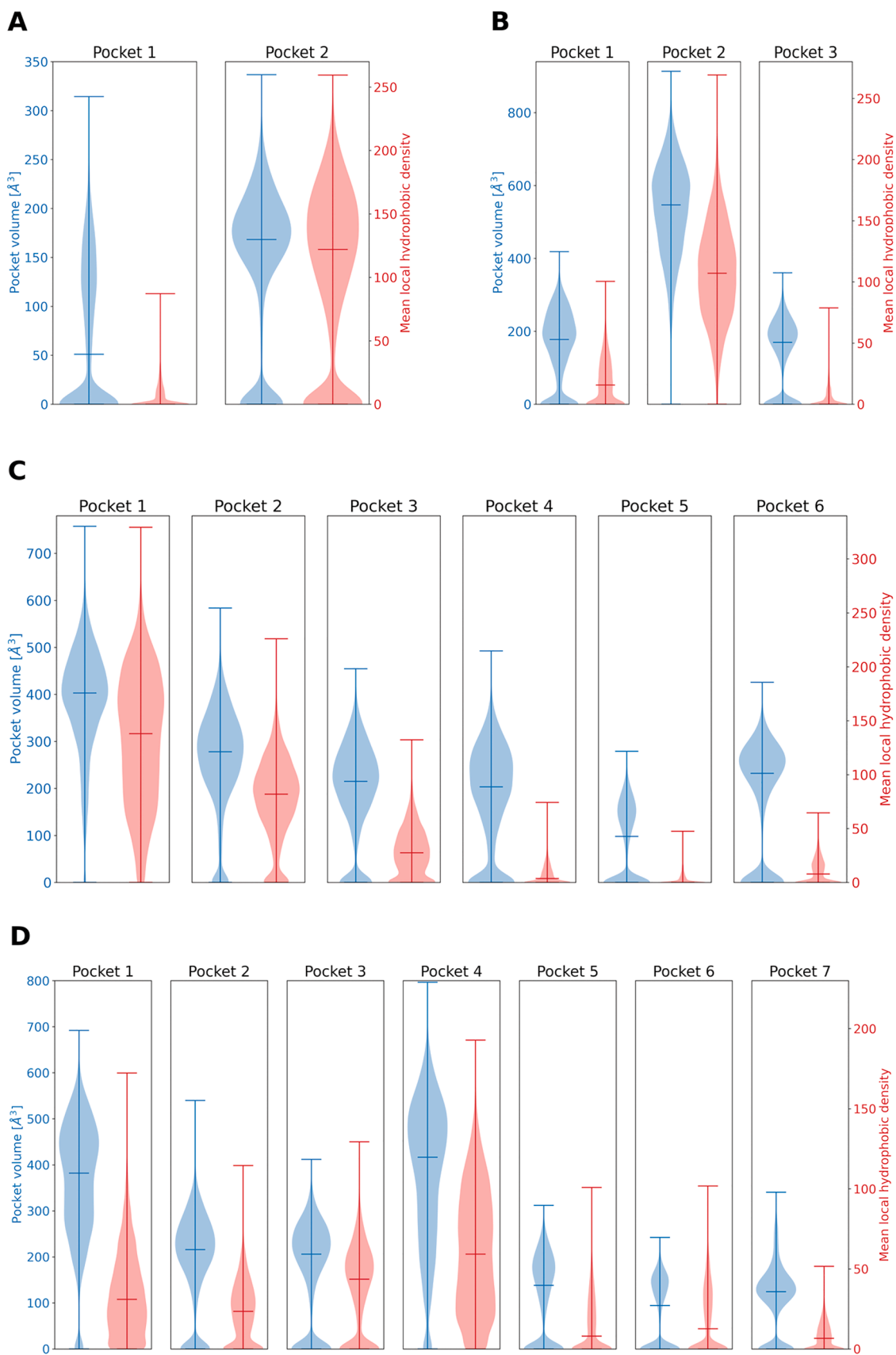


Fig. 5. Pocket volume and MLHD value distributions for all detected binding pockets. A: SMN-Tudor; B: HDAC6-ZnF-UBD; C: NSD3-PWWP1; D: ENL-YEATS.

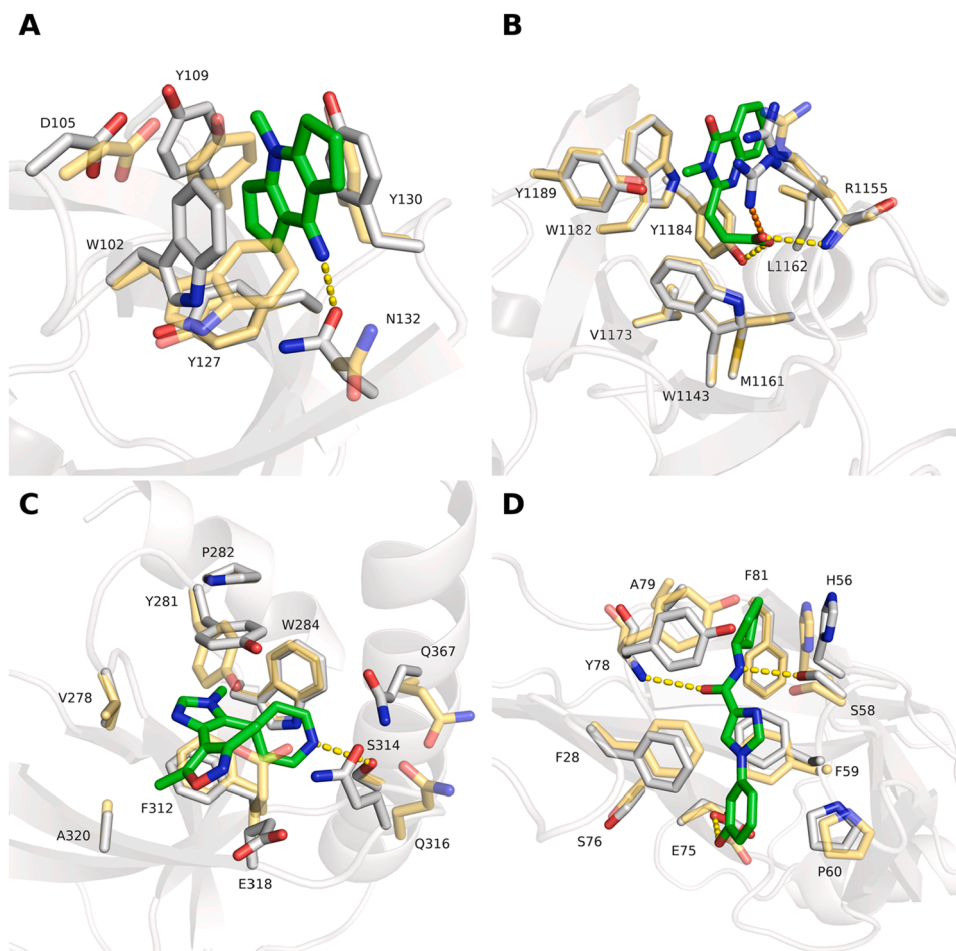


Fig. 6. Crystallographically solved fragment binding modes. Key hydrogen bonds and salt bridges are depicted with dashed yellow and orange lines, respectively. Site residue conformations of the unbound protein structures are shown with transparent yellow sticks for comparison. A: SMN-Tudor with cpd. 1 (PDB ID 4QQ6); B: HDAC6-ZnF-UBD with cpd. 31 (PDB ID 6CED); C: NSD3-PWWP1 with cpd. 9 (PDB ID 6G2C); D: ENL-YEATS with cpd. 10a (PDB ID 8PJI).

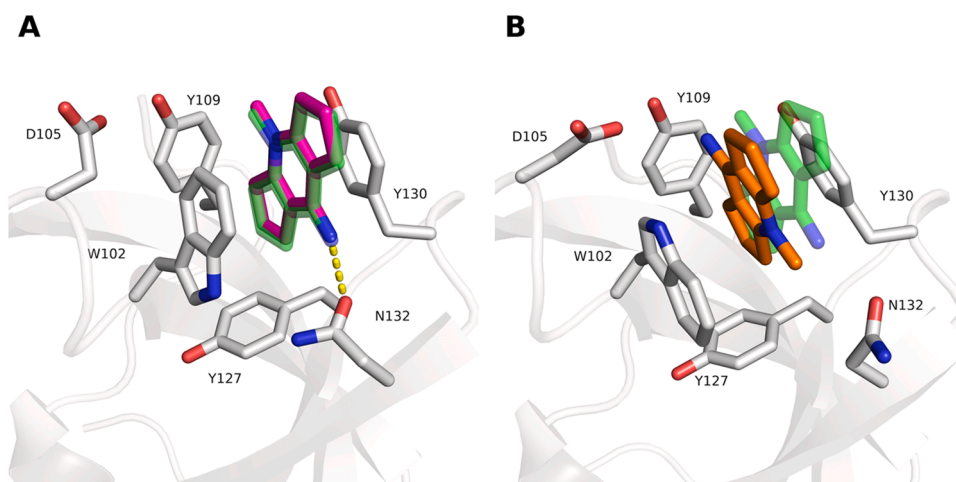


Fig. 7. Binding poses obtained for cpd. 1 (SMN-Tudor). The native binding mode, retrieved from PDB ID 4QQ6, is shown in transparent green sticks for comparison. Key hydrogen bonds are depicted with dashed yellow lines. A: Top-scored pose (magenta), B: third-ranked pose (orange).

(Tables 6, S4). Cpd. 10a, co-crystallized in PDB ID 8PJI, exhibits multiple ligand-protein interactions: the compound is stabilized within the binding site via pi-pi interactions to F28 and F59. Besides, the amide group of the ligand forms hydrogen bonds to the side chain of S58 and the backbone of Y78 and the phenolic hydroxyl group forms a hydrogen

bond to E75 (Fig. 6D). The original binding mode, including the described interactions, was successfully reproduced by the two top-scored fragments according to the calculated RMSD values (Table 6) and the visual inspection of respective complexes (Figs. 9B, S4C). Once more, the calculated QM/MM-GBSA interaction energies allow a clear

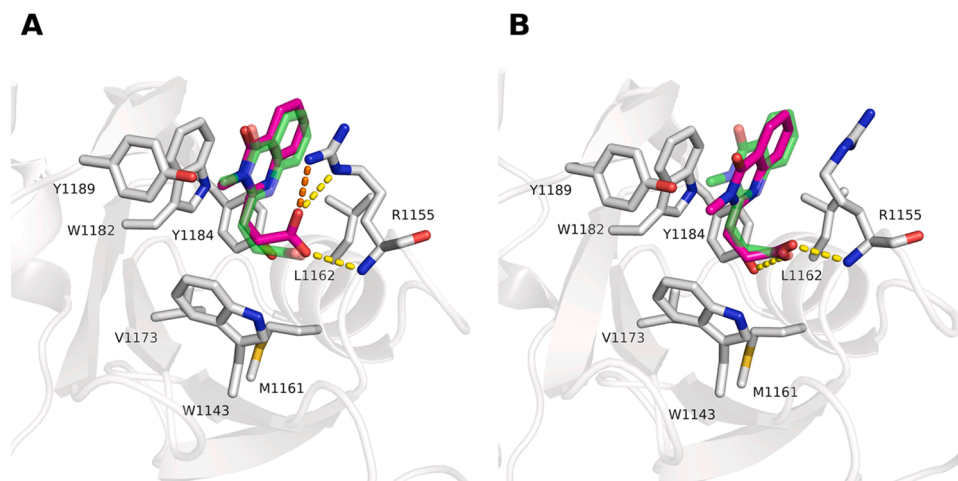


Fig. 8. Top-scored binding poses obtained for cpd. 31 (HDAC6-ZnF-UBD). The native binding mode, retrieved from PDB ID 6CED is shown in transparent green sticks for comparison. Key hydrogen bonds and salt bridges are depicted with dashed yellow and orange lines, respectively. A: Top-scored pose; B: second-ranked pose.

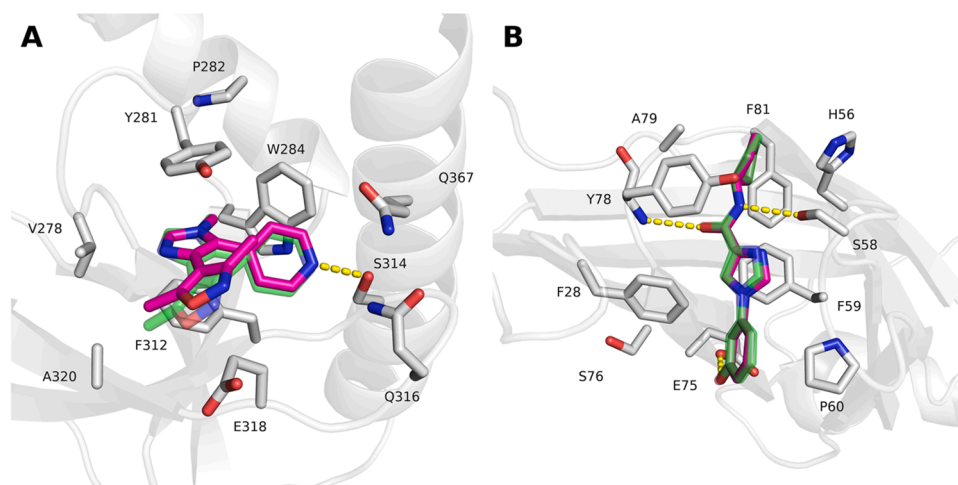


Fig. 9. Top-scored binding poses obtained for cpd. 9 (NSD3-PWWP1, A) and cpd. 10a (ENL-YEATS, B). The respective native binding modes, retrieved from PDB IDs 6G2C and 8PJI, are shown in transparent green sticks for comparison. Key hydrogen bonds are depicted with dashed yellow lines.

differentiation from the remaining generated poses (Table 6).

3.8. Binding mode prediction through molecular docking

In order to compare the performance of the presented workflow with conventional and computationally less expensive methods for binding mode prediction, molecular docking was carried out for the selected targets and fragments. In a blind docking study, Molecular Operating Environment (MOE) and Glide were tested for their ability to correctly predict the binding modes while specifying the whole proteins as potential binding region. RMSD values were calculated for the generated docking poses with respect to the fragment binding modes found in the reference co-crystal structures (Table S5). As shown in Table S5, high RMSD values were predominantly obtained which indicate that in most cases neither the binding site nor the binding pose were correctly predicted. Blind docking was only successful for cpd. 31 in HDAC6 ZnF-UBD as shown by RMSD values below 2 Å. Visual inspection of the low-RMSD binding poses revealed that docking using MOE failed to accurately orientate the carboxylate group within the binding pocket (Fig. 10A). Conversely, the obtained Glide docking pose showed the key hydrogen bond interactions between the carboxylate group and the neighboring residues as observed in the reference crystal structure (Fig. 11A). However, the salt bridge to R1155 was not observed. We visually

compared the experimentally solved unbound and fragment-bound binding site conformations in order to identify potential reasons for the rather poor blind docking performance in the other studied targets (Fig. 6). The comparison shows that in the unbound state the pockets are blocked or are less accessible. Meanwhile in the case of HDAC6 ZnF-UBD, where blind docking showed good performance, the binding site residues show a better overlap in the unbound and fragment-bound forms (Fig. 6B).

In the next step, we investigated whether rigid body docking was successful in binding mode reproduction when the binding site location was specified in the docking process. To this end, the binding sites were defined with the aid of the coordinates of the superimposed co-crystallized fragments. The docking performance for cpd. 31 in HDAC6 ZnF-UBD was comparable to the blind docking study: while Glide generated a similar docking pose (Fig. 11B), MOE failed to correctly place the fragment's carboxylate group (Fig. 10B). In the case of cpd. 10a in ENL-YEATS, RMSD values below 2 Å were obtained which suggests a successful binding mode reproduction (Table S5). However, visual inspection of the respective docking poses revealed that the hydrogen bond between the ligand's amide and the backbone of Y78 was not formed (Figs. 10C, 11C). The binding pocket in the apo protein structure does not offer enough space to correctly fit the fragment, which demonstrates the limitations of rigid body docking. For the

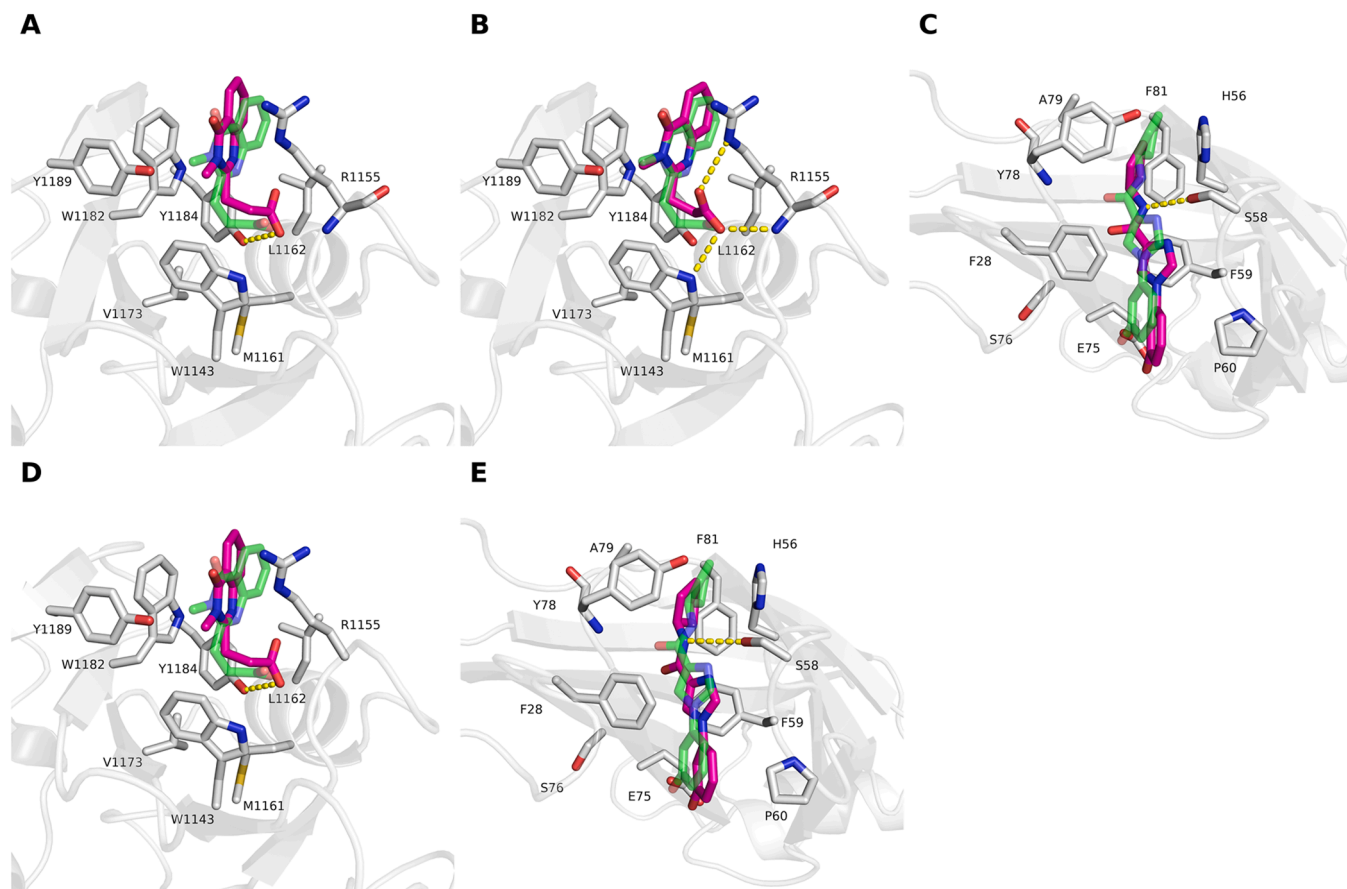


Fig. 10. Native-like binding poses generated by MOE. A: Blind docking pose of cpd. 31 in HDAC6-ZnF-UBD; B: site-directed docking pose of cpd. 31 in HDAC6-ZnF-UBD; C: site-directed docking pose of cpd. 10a in ENL-YEATS; D: induced fit docking pose of cpd. 31 in HDAC6-ZnF-UBD; E: induced fit docking pose of cpd. 10a in ENL-YEATS.

residual fragment inhibitors, cpds. 1 and 9, no native-like docking poses could be obtained in the respective binding sites (Table S5).

In the last step, it was assessed whether induced fit docking was able to generate the desired binding modes. For HDAC6 ZnF-UBD, ENL-YEATS, and their respective fragment binders, docking poses with RMSD values below 2 Å were generated (Table S5). MOE failed again at correctly placing the carboxylate group of cpd. 31 within the binding pocket (Fig. 10D). Schrödinger's induced fit workflow reproduced most of the key protein-ligand interactions (Fig. 11E). However, we did not observe any docking pose in which all of the described reference interactions were formed at the same time. The docking poses generated by MOE for cpd. 10a in ENL-YEATS lack key hydrogen bond interactions (Fig. 10E). Glide combined with Prime performed better and reproduced all the reference interactions (Fig. 11F). Moreover, Schrödinger's induced fit docking algorithm was the only tool that reproduced the binding mode of cpd. 1 in SMN-Tudor (Fig. 11D). The conformational change of W102 upon ligand binding was sampled correctly. However, the hydrogen bond to N132 was not formed. Meanwhile, in the case of cpd. 9 in NSD3-PWWP1 induced fit docking as well as rigid body docking were not successful in binding mode reproduction (Table S5).

4. Conclusions

In the present study, the extent to which fragments can be used instead of the usual probe molecules in co-solvent MD simulations for the prediction of binding modes was tested. To this end, fragments with known binding modes in four different epigenetic targets were studied. A major challenge lied in ensuring a homogenous fragment distribution within the simulated aqueous phase. To achieve satisfying distribution,

the fragment concentration was lowered ($0.2 \text{ mol} \cdot \text{l}^{-1}$) and an artificial repulsive LJ potential was introduced. Preliminary studies focusing on the simulation of fragment-water mixtures confirmed that suitable LJ parameters were chosen. In the next step, the fragments were simulated in the presence of their respective target proteins while applying the modified LJ parameters. The aim was to investigate to what extent a prediction of the right fragment binding site and binding mode is possible, assuming the case that both are unknown. The analysis of the obtained trajectories using MDpocket allowed the prediction of the right fragment binding site. Especially the calculated volume and MLHD descriptors turned out to be useful indicators for the prediction. Subsequently, filters were established for identifying the fragments that potentially reproduced the experimentally solved binding modes within the pocket of interest. Only fragments which showed low mobility during the production phase and a significant volume overlap with the identified pockets were considered for further assessment. Finally, scoring with QM/MM-GBSA was performed on the minimized fragment-protein complexes. We were able to show that the best scoring fragment-protein complexes had a high degree of agreement with the experimentally determined binding modes. This underlines the ability of the applied QM/MM-GBSA method to discriminate between correct and incorrect poses.

Predicting fragment binding modes and identifying their respective binding pockets is challenging. Several studies have, for example, shown that molecular docking shows weak performance in identifying correct fragment poses due to the intrinsically low binding affinities [17–20]. The current study also shows that conventional docking methods were not always successful in reproducing the experimentally determined binding modes of the selected fragments, particularly when the binding

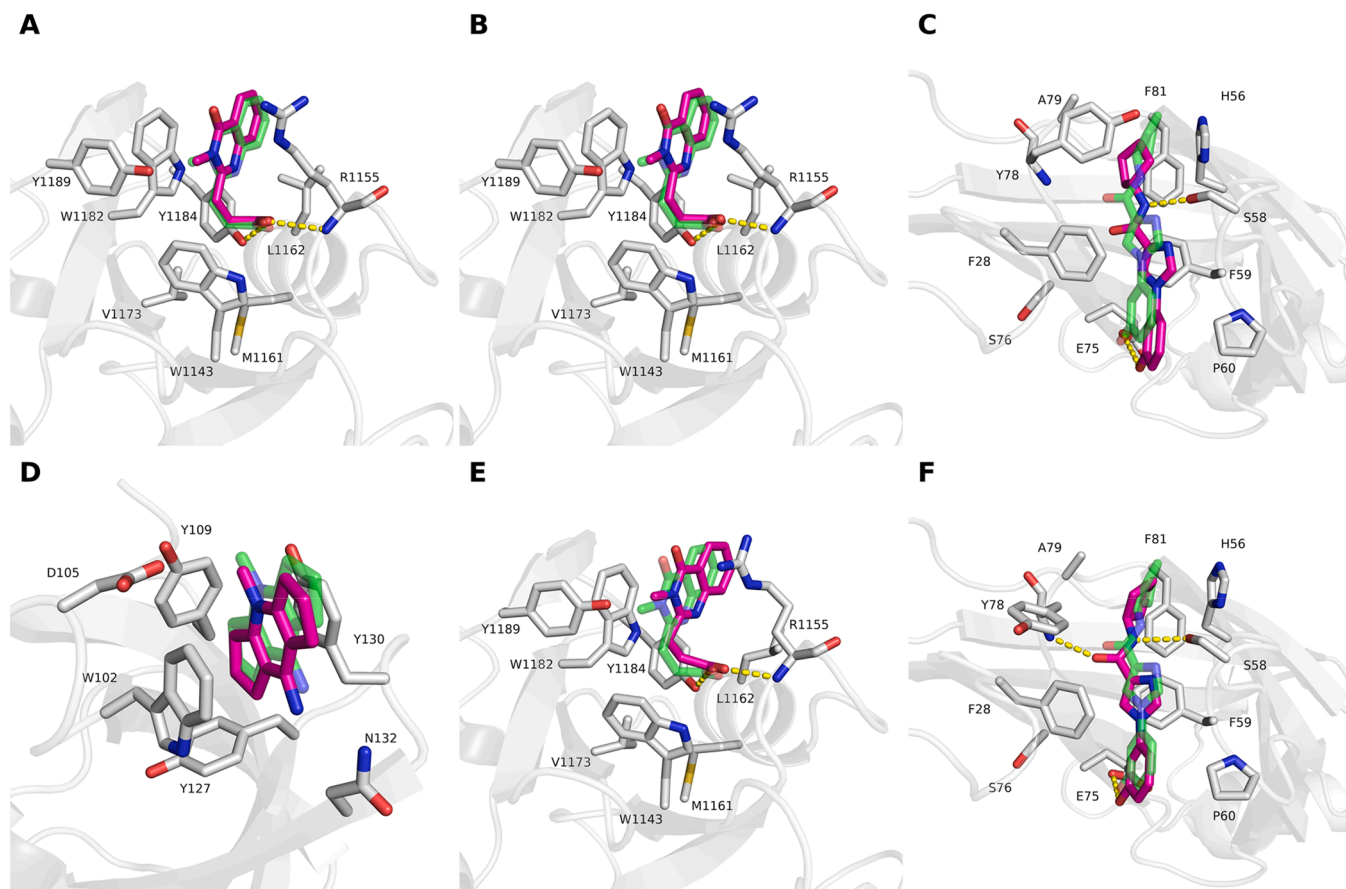


Fig. 11. Native-like binding poses generated by the Schrödinger software suite. A: Blind docking pose of cpd. 31 in HDAC6-ZnF-UBD; B: site-directed docking pose of cpd. 31 in HDAC6-ZnF-UBD; C: site-directed docking pose of cpd. 10a in ENL-YEATS; D: induced fit docking pose of cpd. 1 in SMN-Tudor; E: induced fit docking pose of cpd. 31 in HDAC6-ZnF-UBD; F: induced fit docking pose of cpd. 10a in ENL-YEATS.

site was not specified. It was also observed that the docking algorithms especially performed poorly with cryptic pockets like that of NSD3-PWWP1. The herein presented method can be implemented as a blind binding mode prediction for fragments where no crystallographic structures are available, which would be very useful for fragment-based drug design and lead optimization studies. It must be acknowledged that the developed methodology is computationally more expensive than docking approaches. An increase in fragment size and flexibility would necessitate extended MD simulations, higher PME cutoff values and, subsequently, increased calculation times. However, given the low performance of computationally cheaper methods like molecular docking, particularly with fragment-like molecules, our presented method represents a highly viable alternative. Compared to other MD methods that aim for fragment binding mode prediction [7,9], the herein presented method uses high fragment concentrations which allows a decrease of simulation times. Besides, the proposed approach is based on a high number of short MD replicas with changing initial fragment positions which increases the probability of desired binding events. Finally, the method is not restricted to small fragments and can be applied to lead-like molecules due to the flexible adaption of the PME cutoff.

CRedit authorship contribution statement

Dina Robaa: Writing – review & editing, Supervision, Formal analysis. **Christopher Vorreiter:** Writing – original draft, Conceptualization. **Wolfgang Sippl:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Author contributions

CV and DR did the computational studies and wrote the manuscript. WS supervised the experiments and revised the manuscript.

Data statement

All crystal structures used in the current work are retrievable from the RCSB protein data bank (<https://www.rcsb.org/>). The protein structures were prepared with the aid of the Protein Preparation Wizard (Schrodinger 2021–3), the fragment inhibitor structures were prepared with LigPrep (Schrodinger 2021–3). All systems for MD simulation were generated by Packmol (20.3.2, <https://m3g.github.io/packmol/>). MD simulations were carried out using Amber 22 software. AmberTools 22 was used for preparing the built systems for MD input and for analyzing the obtained trajectories. MDpocket (4.0, <https://github.com/Discngin e/fpocket>) executed the detection and characterization of binding pockets in the generated trajectories. QM/MM-GBSA calculations were performed by the MMPBSA.py script provided by Amber 22. Volume overlaps for binding pocket assignment were calculated using Phase VolCalc (Schrodinger 2021–3). The docking studies were carried out using Molecular Operating Environment (2019.01), Glide (Schrödinger 2021–3) and the Induced Fit Docking protocol of Schrödinger 2021–3. RMSD calculations for identifying fragment binding poses matching with the poses observed in the reference X-ray structures were executed in Schrödinger’s graphical interface Maestro (2021–3). Microsoft Excel and Matplotlib were used for plot generation. Figures were created using GeoGebra (<https://www.geogebra.org/>), PyMol (2.6.0a0) and Marvin Sketch (19.19.0 2019). The latter was also used for calculating

respective descriptors to test the fragments' compliance with the rule of three. The input scripts for Packmol and Amber as well as the generated trajectories and fragment binding poses are available upon request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge funding by the DFG Grant RU5433 (Project no. 495271833) and DFG SI868/26-1 (Project no. 468534282).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.csbj.2024.12.017](https://doi.org/10.1016/j.csbj.2024.12.017).

References

- [1] Ferenczy GG, Keseru GM. How are fragments optimized? A retrospective analysis of 145 fragment optimizations. *J Med Chem* 2013;56(6):2478–86.
- [2] Kirsch P, et al. Concepts and core principles of fragment-based drug design. *Molecules* 2019;24(23).
- [3] Carr R, Jhoti H. Structure-based screening of low-affinity compounds. *Drug Discov Today* 2002;7(9):522–7.
- [4] Erlanson DA, et al. Site-directed ligand discovery. *Proc Natl Acad Sci USA* 2000;97(17):9367–72.
- [5] Hartshorn MJ, et al. Fragment-based lead discovery using X-ray crystallography. *J Med Chem* 2005;48(2):403–13.
- [6] Schiebel J, et al. Six biophysical screening methods miss a large proportion of crystallographically discovered fragment hits: a case study. *ACS Chem Biol* 2016;11(6):1693–701.
- [7] Linker SM, et al. Fragment binding pose predictions using unbiased simulations and Markov-state models. *J Chem Theory Comput* 2019;15(9):4974–81.
- [8] Tang GW, Altman RB. Knowledge-based fragment binding prediction. *PLoS Comput Biol* 2014;10(4):e1003589.
- [9] Privat C, et al. Fragment dissolved molecular dynamics: a systematic and efficient method to locate binding sites. *Phys Chem Chem Phys* 2021;23(4):3123–34.
- [10] Congreve M, et al. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* 2003;8(19):876–7.
- [11] Seco J, Luque FJ, Barril X. Binding site detection and druggability index from first principles. *J Med Chem* 2009;52(8):2363–71.
- [12] Lexa KW, Carlson HA. Improving protocols for protein mapping through proper comparison to crystallography data. *J Chem Inf Model* 2013;53(2):391–402.
- [13] Raman EP, et al. Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J Chem Inf Model* 2013;53(12):3384–98.
- [14] Ghanakota P, Carlson HA. Driving structure-based drug discovery through cosolvent molecular dynamics. *J Med Chem* 2016;59(23):10383–99.
- [15] Sayyed-Ahmad A. Hotspot identification on protein surfaces using probe-based MD simulations: successes and challenges. *Curr Top Med Chem* 2018;18(27):2278–83.
- [16] Vorreiter C, Robaa D, Sippl W. Exploring aromatic cage flexibility using cosolvent molecular dynamics simulations – an in-silico case study of tudor domains. *J Chem Inf Model* 2024;64(11):4553–69.
- [17] Good AC, et al. Implications of promiscuous Pim-1 kinase fragment inhibitor hydrophobic interactions for fragment-based drug design. *J Med Chem* 2012;55(6):2641–8.
- [18] Herbst C, et al. Assessment of fragment docking and scoring with the endothepepsin model system. *Arch Pharm* 2024;357(6):e2400061.
- [19] Jacquemard C, et al. Binding mode information improves fragment docking. *J Cheminform* 2019;11(1):24.
- [20] Chachulski L, Windshugel B. LEADS-FRAG: a benchmark data set for assessment of fragment docking performance. *J Chem Inf Model* 2020;60(12):6544–54.
- [21] Brown WM, Vander Jagt DL. Creating artificial binding pocket boundaries to improve the efficiency of flexible ligand docking. *J Chem Inf Comput Sci* 2004;44(4):1412–22.
- [22] Grasso G, et al. Fragmented blind docking: a novel protein-ligand binding prediction protocol. *J Biomol Struct Dyn* 2022;40(24):13472–81.
- [23] Tripsianes K, et al. Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nat Struct Mol Biol* 2011;18(12):1414–20.
- [24] Zhang M, et al. Histone and DNA binding ability studies of the NSD subfamily of PWWP domains. *Biochem Biophys Res Commun* 2021;569:199–206.
- [25] Wan L, et al. ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature* 2017;543(7644):265–9.
- [26] Kawaguchi Y, et al. The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. *Cell* 2003;115(6):727–38.
- [27] Lefebvre S, et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 1995;80(1):155–65.
- [28] Sakuma T, et al. Aberrant expression of histone deacetylase 6 in oral squamous cell carcinoma. *Int J Oncol* 2006;29(1):117–24.
- [29] Wickstrom SA, et al. CYLD negatively regulates cell-cycle progression by inactivating HDAC6 and increasing the levels of acetylated tubulin. *EMBO J* 2010;29(1):131–44.
- [30] Lee YS, et al. The cytoplasmic deacetylase HDAC6 is required for efficient oncogenic tumorigenesis. *Cancer Res* 2008;68(18):7561–9.
- [31] Kanno K, et al. Overexpression of histone deacetylase 6 contributes to accelerated migration and invasion activity of hepatocellular carcinoma cells. *Oncol Rep* 2012;28(3):867–73.
- [32] Rey M, et al. HDAC6 is required for invadopodia activity and invasion by breast tumor cells. *Eur J Cell Biol* 2011;90(2-3):128–35.
- [33] Zuo Q, et al. HDAC6 and SIRT2 promote bladder cancer cell migration and invasion by targeting cortactin. *Oncol Rep* 2012;27(3):819–24.
- [34] Li Y, Shin D, Kwon SH. Histone deacetylase 6 plays a role as a distinct regulator of diverse cellular processes. *FEBS J* 2013;280(3):775–93.
- [35] Chen Y, et al. Identification of druggable cancer driver genes amplified across TCGA datasets. *PLoS One* 2014;9(5):e98293.
- [36] Kang D, et al. The histone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis. *Genes Chromosomes Cancer* 2013;52(2):126–39.
- [37] Rosati R, et al. NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). *Blood* 2002;99(10):3857–60.
- [38] Taketani T, et al. NUP98-NSD3 fusion gene in radiation-associated myelodysplastic syndrome with t(8;11)(p11;p15) and expression pattern of NSD family genes. *Cancer Genet Cytogenet* 2009;190(2):108–12.
- [39] Yang ZQ, et al. Transforming properties of 8p11-12 amplified genes in human breast cancer. *Cancer Res* 2010;70(21):8487–97.
- [40] French CA, et al. NSD3-NUT fusion oncoprotein in NUT midline carcinoma: implications for a novel oncogenic mechanism. *Cancer Discov* 2014;4(8):928–41.
- [41] Vougiouklakis T, et al. The NSD family of protein methyltransferases in human cancer. *Epigenomics* 2015;7(5):863–74.
- [42] Perlman EJ, et al. MLLT1 YEATS domain mutations in clinically distinctive favourable histology wilms tumours. *Nat Commun* 2015;6:10013.
- [43] Zhao D, et al. YEATS domain—a histone acylation reader in health and disease. *J Mol Biol* 2017;429(13):1994–2002.
- [44] Liu Y, et al. A small molecule antagonist of SMN disrupts the interaction between SMN and RNAP II. *Nat Commun* 2022;13(1):5453.
- [45] Ferreira de Freitas R, et al. Identification and structure-activity relationship of HDAC6 zinc-finger ubiquitin binding domain inhibitors. *J Med Chem* 2018;61(10):4517–27.
- [46] Harding RJ, et al. Small molecule antagonists of the interaction between the histone deacetylase 6 zinc-finger domain and ubiquitin. *J Med Chem* 2017;60(21):9090–6.
- [47] Bottcher J, et al. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nat Chem Biol* 2019;15(8):822–9.
- [48] Raux B, et al. Discovery of PFI-6, a small-molecule chemical probe for the YEATS domain of MLLT1 and MLLT3. *Bioorg Med Chem Lett* 2024;98:129546.
- [49] Moustakim M, et al. Discovery of an MLLT1/3 YEATS domain chemical probe. *Angew Chem Int Ed Engl* 2018;57(50):16302–7.
- [50] Ni X, et al. Structural Insights into interaction mechanisms of alternative piperazine-urea YEATS domain binders in MLLT1. *ACS Med Chem Lett* 2019;10(12):1661–6.
- [51] Heidenreich D, et al. Structure-based approach toward identification of inhibitory fragments for eleven-nineteen-leukemia protein (ENL). *J Med Chem* 2018;61(23):10929–34.
- [52] Berman HM, et al. The protein data bank. *Nucleic Acids Res* 2000;28(1):235–42.
- [53] Sastry GM, et al. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013;27(3):221–34.
- [54] Protein Preparation Wizard, Epik, Schrödinger, LLC, New York, NY, Impact, Schrödinger, LLC, New York, NY, Prime, Schrödinger, LLC, New York, NY, 2021.
- [55] Shivakumar D, et al. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J Chem Theory Comput* 2010;6(5):1509–19.
- [56] Jorgensen WLM, D.S., Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 1996;118(45):11225–36.
- [57] Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 1988;110(6):1657–66.
- [58] Schrödinger Release 2021-3: LigPrep, Schrödinger, LLC, New York, NY, 2021.
- [59] Greenwood JR, et al. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J Comput Aided Mol Des* 2010;24(6-7):591–604.
- [60] Shelley JC, et al. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* 2007;21(12):681–91.
- [61] Schrödinger Release 2021-3: Epik, Schrödinger, LLC, New York, NY, 2021.
- [62] Marvin 19.19.0 2019, ChemAxon, (<http://www.chemaxon.com>).
- [63] Martinez L, et al. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* 2009;30(13):2157–64.

- [64] Jorgensen WLC, J, Madura J, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–35.
- [65] Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, et al. Amber 2022. San Francisco: University of California; 2022.
- [66] Maier JA, et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 2015;11(8):3696–713.
- [67] Sengupta A, et al. Parameterization of monovalent ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *J Chem Inf Model* 2021;61(2):869–80.
- [68] Li P, Song LF, Merz Jr KM. Systematic parameterization of monovalent ions employing the nonbonded model. *J Chem Theory Comput* 2015;11(4):1645–57.
- [69] Li Z, et al. Systematic parameterization of divalent metal ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *J Chem Theory Comput* 2020;16(7):4429–42.
- [70] Li P, Merz Jr KM. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J Chem Theory Comput* 2014;10(1):289–97.
- [71] Li Z, et al. Parameterization of trivalent and tetravalent metal ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *J Chem Theory Comput* 2021;17(4):2342–54.
- [72] Li P, Song LF, Merz Jr KM. Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water. *J Phys Chem B* 2015;119(3):883–95.
- [73] Jakalian A, B.L.B, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem* 2000;21:132–46.
- [74] Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 2002;23(16):1623–41.
- [75] Wang J, et al. Development and testing of a general amber force field. *J Comput Chem* 2004;25(9):1157–74.
- [76] Roe DR, Cheatham 3rd TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 2013;9(7):3084–95.
- [77] Ryckaert JP, G.C, Berendsen HJC. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *J Comput Phys* 1977;23:327–41.
- [78] Darden T, D.Y, Pedersen L. Particle Mesh Ewald – an N.log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–92.
- [79] Essmann U, et al. A smooth particle mesh Ewald method. *J Chem Phys* 1995;103(19):8577–93.
- [80] Schmidtke P, et al. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* 2011;27(23):3276–85.
- [81] Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinform* 2009;10:168.
- [82] Miller 3rd BR, et al. MMPBSA.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput* 2012;8(9):3314–21.
- [83] Still WC, et al. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 1990;112(16):6127–9.
- [84] Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 2000;56(4):275–91.
- [85] Hawkins GD, Cramer CJ, Truhlar DG. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem* 1996;100(51):19824–39.
- [86] Hawkins GD, Cramer CJ, Truhlar DG. Pairwise solute descreening of solute charges from a dielectric medium. *Chem Phys Lett* 1995;246(1):122–9.
- [87] Stewart JJ. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J Mol Model* 2007;13(12):1173–213.
- [88] Korth M. Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields. *J Chem Theory Comput* 2010;6(12):3808–16.
- [89] Srinivasan J, et al. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J Am Chem Soc* 1998;120(37):9401–9.
- [90] Kollman PA, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 2000;33(12):889–97.
- [91] Dixon SL, et al. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput-Aided Mol Des* 2006;20(10):647–71.
- [92] Dixon SL, Smondryev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* 2006;67(5):370–2.
- [93] Schrödinger Release 2021-3: Phase, Schrödinger, LLC, New York, NY, 2021.
- [94] Schrödinger Release 2021-3: Maestro. Schrödinger, LLC, New York, NY, 2021.
- [95] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2019.
- [96] Friesner RA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 2006;49(21):6177–96.
- [97] Friesner RA, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;47(7):1739–49.
- [98] Halgren TA, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 2004;47(7):1750–9.
- [99] Schrödinger Release 2021-3: Glide, Schrödinger, LLC, New York, NY, 2021.
- [100] Farid R, et al. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg Med Chem* 2006;14(9):3160–73.
- [101] Sherman W, et al. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 2006;49(2):534–53.
- [102] Sherman W, Beard HS, Farid R. Use of an induced fit receptor structure in virtual screening. *Chem Biol Drug Des* 2006;67(1):83–4.
- [103] Schrödinger Release 2021-3: Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2021; Prime, Schrödinger, LLC, New York, NY, 2021.
- [104] Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9(3):90–5.
- [105] GeoGebra Rechner Suite, GeoGebra GmbH, (<https://www.geogebra.org/>).
- [106] PyMOL Molecular Graphics System, Version 2.6.0a0 Schrödinger, LLC.
- [107] Guvench O, MacKerell Jr AD. Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput Biol* 2009;5(7):e1000435.
- [108] Tan YS, et al. Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1. *Angew Chem Int Ed Engl* 2012;51(40):10078–81.
- [109] Beyens O, De Winter H. Preventing lipophilic aggregation in cosolvent molecular dynamics simulations with hydrophobic probes using plumed automatic restraining tool (PART). *J Cheminform* 2024;16(1):23.
- [110] Jones JE, Chapman S. On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proc R Soc Lond Ser A Contain Pap A Math Phys Character* 1997;106(738):441–62.
- [111] Lennard-Jones JE. Cohesion. *Proc Phys Soc* 1931;43(5):461.
- [112] Wang X, et al. The Lennard-Jones potential: when (not) to use it. *Phys Chem Chem Phys* 2020;22(19):10624–33.
- [113] Crowley M, et al. Adventures in improving the scaling and accuracy of a parallel molecular dynamics program. *J Supercomput* 1997;11(3):255–78.
- [114] Tian C, et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput* 2020;16(1):528–52.
- [115] Schmidtke P, Barril X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J Med Chem* 2010;53(15):5858–67.