European Journal of Medicinal Chemistry 91 (2015) 4–14

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Review article

# Molecular dynamics in drug design

Hongtao Zhao, Amedeo Caflisch*

*Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland*

ARTICLE INFO

*Article history:*
Received 10 June 2014
Received in revised form
31 July 2014
Accepted 3 August 2014
Available online 4 August 2014

*Keywords:*
Atomistic simulation
In silico screening
High-throughput docking
Bromodomains
Tyrosine kinases
Proteases

ABSTRACT

Molecular dynamics (MD) simulations are useful tools for structure-based drug design. We review recent publications in which explicit solvent MD was used at the initial or final stages of high-throughput docking campaigns. In some cases, MD simulations of the protein target have been carried out before docking to generate a conformer of the protein which differs from the available crystal structure(s). Furthermore, MD runs have been performed after docking to assess the predicted binding modes of the top ranking compounds as final filter in silico or to guide chemical synthesis for hit optimization. We present examples of in silico discoveries of tyrosine kinase inhibitors and bromodomain antagonists whose binding mode was predicted by automated docking and further corroborated by MD simulations with final validation by X-ray crystallography.



## 1. Introduction

While atomistic MD simulations are still computationally expensive for docking large libraries of compounds, their application for hit discovery and optimization is increasing steadily. Here we review in silico screening campaigns in which MD played a key role in the identification of small molecules that bind to protein targets. The focus is on explicit solvent MD simulations carried out in our research group to prepare a structure for docking or to assess predicted binding modes. The high-throughput campaigns performed in our group are summarized in Table 1, and representative hits are shown in Fig. 1 [1–15]. MD simulations were employed at the protein-preparation stage or final scoring step in the campaigns that led to the identification of the enzyme inhibitors **5**, **11**–**17** [5,11–14] and the bromodomain ligands **18**–**19** [15]. Several of the hits identified in silico, e.g., compounds **7** and **10**–**11**, have been advanced into series of potent and selective tyrosine kinase inhibitors which are promising pre-clinical candidates [11,16,17].

## 2. MD as a tool for mapping molecular fragments to binding sites

### 2.1. Pioneering studies and recent developments

Almost 30 years ago, Peter J. Goodford pioneered the use of molecular fragments to map protein binding sites by calculating the interaction energy on a grid around the protein surface [18]. In 1991, Andrew Miranker and Martin Karplus proposed energy minimization as a simple and efficient method for generating functionality maps, that is optimal positions and orientations of functional groups in the binding site of a protein target [19]. The method was called multiple copies simultaneous search (MCSS) because during minimization the interaction energy between multiple replicas of a molecular fragment was switched off so that each replica feels only the force field of the protein. Some of the biophysical methods for fragment-based lead identification, such as structure–activity relationship by nuclear magnetic resonance spectroscopy [20], are similar in principle to the MCSS approach. It is interesting to note that computational methods for generating functionality maps [18,19,21,22] have preceded experimental methods that report on the binding of small molecules. The experimental techniques include X-ray crystallography [23], nuclear magnetic resonance spectroscopy [24], surface plasmon resonance [25], mass spectrometry [26,27], substrate activity screening (where the fragments are substrates later converted into inhibitors [28–30]), and tethering [31,32].

Recently, MD simulations have been employed for determining the binding modes of small aliphatic and aromatic molecules into the oncoprotein BCL-6 [33] and isopropyl alcohol into five different proteins [34]. These studies were published 18 years after the minimization-based MCSS protocol of Miranker and Karplus, which in principle allowed also for binding site flexibility by a combination of MD and energy minimization. The MD protocol developed recently in MacKerell's group is called SILCS (Site-Identification by

* Corresponding author.
*E-mail address:* caflisch@bioc.uzh.ch (A. Caflisch).

http://dx.doi.org/10.1016/j.ejmech.2014.08.004
0223-5234/

**Table 1**
Structure-based virtual screening campaigns performed at the Department of Biochemistry of the University of Zurich during the decade 2005–2014.

| | Protein | Representative hits[a] | | | Hit rate[b] (%) | Scoring method | Ref. | PDB entry |
|---|---|---|---|---|---|---|---|---|
| | | No. | Affinity (μM) | LE | | | | |
| Proteases | β-secretase | **1** | 3.0 | 0.19 | 17 | LIECE[c] | [1] | NA |
| | β-secretase | **2** | 7.1 | 0.19 | 10 | LIECE | [2] | NA |
| | Plasmepsin | **3** | 2.0 | 0.25 | 32 | Consensus | [3] | NA |
| | NS3 protease | **4** | 40 | 0.33 | 5 | LIECE | [4] | NA |
| | NS3 protease | **5** | 2.8 | 0.34 | 40 | Filtering | [5] | NA |
| | Cathepsin B | **6** | 4.8 | 0.29 | 3 | Consensus | [6] | NA |
| Kinases | EphB4 | **7** | 1.5 | 0.32 | 19 | LIECE | [7] | 4GK3 [d] |
| | CDK2 | **8** | 7.8 | 0.32 | 3 | LIECE | [8] | NA |
| | EphB4 | **9** | 2.0 | 0.31 | 13 | Filtering | [9] | NA |
| | EphB4 | **10** | 0.3 | 0.35 | 44 | FFES[e] | [10] | 4P4C |
| | EphB4 | **11** | 5.2 | 0.30 | 67 | FFES | [11] | 4G2F |
| | Abl1, EphA3 | **12** | 3.9 | 0.22 | 25 | FFES | [12] | NA |
| | ZAP70 | **13** | 21 | 0.26 | 31 | FFES | [13] | NA |
| | ZAP70 | **14** | 14 | 0.25 | | | | NA |
| | JAK2 | **15** | 0.1 | 0.37 | | | | NA |
| | SYK | **16** | 23 | 0.21 | 9 | FFES | [14] | NA |
| | SYK | **17** | 18 | 0.22 | | | | NA |
| Bromodomains | BRD4 | **18** | 7.0 | 0.37 | 17 | FFES | [15] | 4PCE |
| | BRD4 | **19** | 7.5 | 0.37 | | | | 4PCI |
| | CREBBP | – | 5.0 | 0.37 | 12 | SEED [47] | [f] | 4TQN |

[a] Compound with highest measured affinity obtained directly from high-throughput docking and scoring. The chemical structures of these compounds are shown in Fig. 1 while the binding modes of compounds **18** and **19** are shown in Fig. 6(C) and (D), respectively. The ligand efficiency (LE) is the measured affinity divided by the number of non-hydrogen atoms and has units of kcal/mol per heavy atom.
[b] Percentage of active compounds i.e., compounds with measured affinity <100 μM divided by the number of molecules tested experimentally. The number of compounds tested ranges from 5 (in the 2nd campaign for NSGH protease) to 88 (in the 2nd campaign for β-secretase).
[c] LIECE: linear interaction energy model with continuum electrostatics [66].
[d] Hit optimization by chemical synthesis of derivatives resulted in low nM inhibitors [16,17] (see also PDB entries 4GK2 and 4GK4).
[e] FFES: force-field energy with electrostatic solvation evaluated by numerical solution of the finite-difference Poisson equation using the continuum dielectric approximation [10].
[f] Min Xu et al. unpublished results.

Ligand Competitive Saturation), and as in MCSS the attractive interactions between fragments are switched off. This simulation stratagem makes possible the use of a very high concentration even for hydrophobic fragments, which would otherwise aggregate in the simulation box [35]. We note *en passant* that this is an interesting example in which a simulation protocol allows one to study a molecular system under conditions that are not accessible by experiments.

### 2.2. Thermodynamics and kinetics of small molecule binding to proteins from MD simulations

We have proposed MD as a tool to analyze the free-energy surface and pathways of (un)binding of small molecules from/to proteins [36,37]. Because of the available crystal structures and measured binding affinities we have applied MD to the peptidyl-prolyl cis-trans isomerase called FKBP (the FK506 binding protein) and six ligands which have between four (dimethylsulphoxide) and eleven (5-diethylamino-2-pentanone) non-hydrogen atoms. Their affinity for FKBP is in the high μM to low mM range as measured by tryptophan fluorescence quenching assay [38]. For each ligand, a conformational space network [39] of the binding process was generated (Fig. 2). In a first step, the relative position and orientation saved along multiple trajectories were clustered according to a set of intermolecular distances. The clusters were considered as nodes of a network, and the direct transitions between these clusters observed during MD were the links of the network. Interestingly, the network analysis revealed multiple binding modes characterized by distinct intermolecular hydrogen bonds and hydrophobic contacts (Fig. 2). Moreover, the unbinding kinetics showed single-exponential time dependence which indicates that the barrier for full dissociation is significantly higher than the barriers between different binding modes. It is instructive to compare experimental and simulation approaches. The aforementioned biophysical techniques for the analysis of fragment binding to proteins have limitations in temporal and/or spatial resolution. In contrast, the MD simulations of (un)binding generate a complete picture of the free-energy surface and (un)binding pathways at atomic level of detail [36,37,40].

Bromodomains are α-helical bundles of approximately 110 residues, which bind acetylated lysine side chains mainly on histone tails [41]. Some of the 61 human bromodomains have been involved in cancer and inflammation. We have carried out MD simulations of two bromodomains (BAZ2B and CREBBP) to assess the structural stability of the six water molecules that seem to be conserved at the bottom of the acetyl-lysine binding site in most crystal structures of bromodomains [42]. The MD runs revealed that the occupancy of the structured water molecules is influenced by the flexibility of the loop connecting helices Z and A (Fig. 3). Additional simulations in the presence of high concentration of cosolvent (i.e., 440 mM of dimethylsulfoxide, methanol, or ethanol) revealed that some of the structured water molecules can be displaced transiently [42]. This observation is consistent with two recently disclosed crystal structures of the fifth bromodomain of human Poly-bromodomain containing protein 1 (PB1) in complex with hydroxyphenyl-propenone ligands (PDB codes 4Q0N and 4Q0O) which show that the phenyl ring of the ligand can replace some of the water molecules at the bottom of the binding pocket.

Two main observations emerge from our MD studies of molecular fragments and cosolvent binding to FKBP and bromodomains: (1) the presence of metastable states corresponding to multiple binding poses, and (2) the importance of solvent molecules in molecular recognition. These two features are fully captured by atomistic, explicit solvent MD simulations while they are difficult to

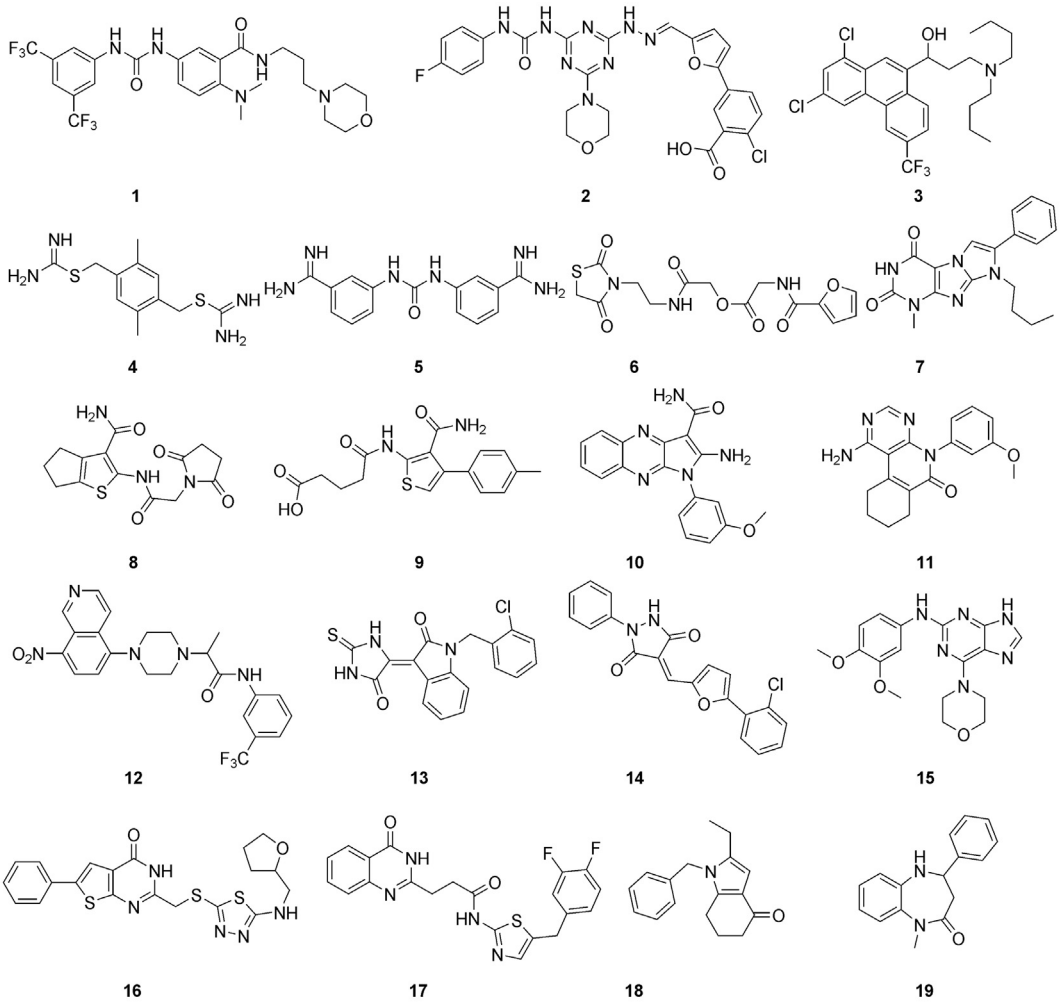


**Fig. 1.** Hits identified in the Caflisch group by high-throughput docking into β-secretase (**1** and **2**), plasmepsin II (**3**), West Nile virus NS3 protease (**4** and **5**), cathepsin B (**6**), EphB4 tyrosine kinase (**7** and **9**–**11**), CDK2 Ser/Thr kinase (**8**), Abl1 kinase (**12**), ZAP70/SYK kinase (**13**–**17**), and BRD4 bromodomain (**18** and **19**). See Table 1 for experimentally measured affinities, hit rates, scoring methods, and PDB codes.

treat with rigid-protein docking. On the other hand, docking is significantly less expensive computationally than MD simulations which, as mentioned in the Introduction, cannot be used for screening libraries with thousands or more compounds.

## 3. Selection of conformer by MD and preparation of binding site by MD-induced fit (MD-IF)

### 3.1. MD protocols for generating protein conformations for docking

The crystal structure of the protein target represents one of the many substates within the folded state. In most cases, the overall topology of the folded state is conserved in its substates but a different orientation of even a single side chain in the binding site can influence significantly the docking results. The McCammon group has developed a so-called relaxed complex scheme to account for protein flexibility in docking [43]. First, MD simulations of the apo protein are conducted to extensively sample the conformational space. The second phase involves docking to a large ensemble of MD snapshots [43]. Ensemble docking following the initial concept of relaxed complex scheme has resulted in the identification of several inhibitors, which would have not been possible by docking to the original crystal structure [44–46].

We have devised MD protocols to sample protein conformations different from those in the crystal structure. These conformations show alternative orientations of side chains in the active site and/or larger aperture of some of the pockets in the binding site of the natural ligand or substrate. The major difference between our MD-based protocol and the ensemble docking approach of McCammon and coworkers is that we use a (small) set of known inhibitors (or fragments) to select a single MD snapshot instead of an ensemble for high-throughput docking. The obvious advantage of doing so is the computational efficiency. On the other hand, the use of an ensemble of structures from MD simulations of the apo protein target is not biased by the knowledge of known inhibitors and might result in completely novel (e.g., allosteric) inhibitors.

### 3.2. Application to the West Nile virus nonstructural 3 protease (NS3pro)

To select a structure of the West Nile virus nonstructural 3 protease (NS3pro) for high-throughput docking three molecular fragments (benzene, methylguanidinium, and 2-phenylimidazoline) were docked into 100 snapshots saved every 0.01 ns along a 1-ns MD run started from the crystal structure [5]. Benzene was used, as it is the most common ring in small molecule drugs, while the other two

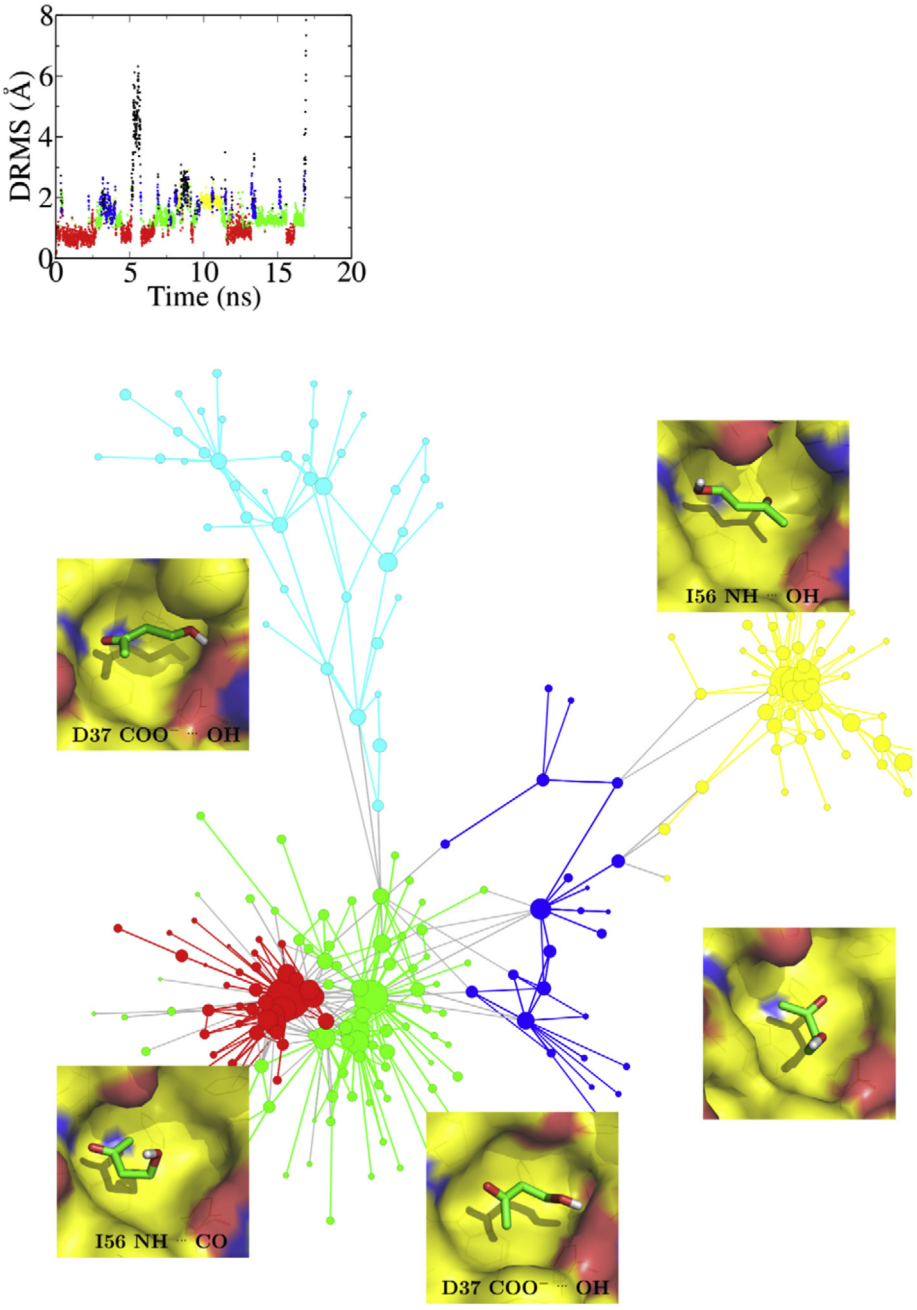


**Fig. 2.** Multiple binding modes of 4-hydroxy-2-butanone (BUT). The binding modes of BUT in the active site of FKBP, i.e., the subbasins within the bound state, are shown by different colors. (Top) Time series of distance root mean square deviation (DRMS) from the crystal structure of the BUT/FKBP complex for one of the 50 MD runs at 310 K. The majority of MD snapshots in the most populated subbasin (red) have a DRMS smaller than 1.0 Å. The interconversions between subbasins are evident. (Bottom) Network representation [39] of the bound state of BUT. Nodes and links are the conformations (i.e., clusters obtained by DRMS clustering) and direct transitions (within 4 ps), respectively, sampled in the 50 MD runs at 310 K. The size of each node is proportional to the natural logarithm of its statistical weight, and only nodes connected by at least one link of weight $\geq$5 are shown to avoid overcrowding. Links connecting pairs of nodes in the same subbasin have the same color of the subbasin, otherwise they are gray. In the insets close to each basin, the FKBP surface is colored according to atom type with carbon atoms surface in yellow while BUT is shown by sticks with carbon atoms in green. Reprinted with permission from Plos Computational Biology. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

fragments with formal charge of +1 were employed because of the negative electrostatic potential on the surface of the active site of NS3pro and previous knowledge on NS3pro inhibitors. The MD snapshot 29, which corresponded to the structure after 0.29 ns of the 1-ns run, was selected for in silico screening because the binding energy (calculated by SEED [47] which includes the electrostatic contribution to solvation) was very favorable for all of the three fragments (Fig. 4). A diversity set of nearly 19,000 molecules was docked into MD snapshot 29 by the program FFLD [48]. After filtering and visual inspection (of 480 predicted binding modes of 178 different compounds) only five molecules were tested in vitro and two of them showed low micromolar affinity (measured by an enzymatic assay, a tryptophan fluorescence assay, and NMR). Compound **5** had a ligand efficiency of 0.34 kcal/mol per non-hydrogen atom according to the enzymatic assay (Table 1) [5]. The hit rate of 40% (2 actives of 5 compounds tested) is very high.

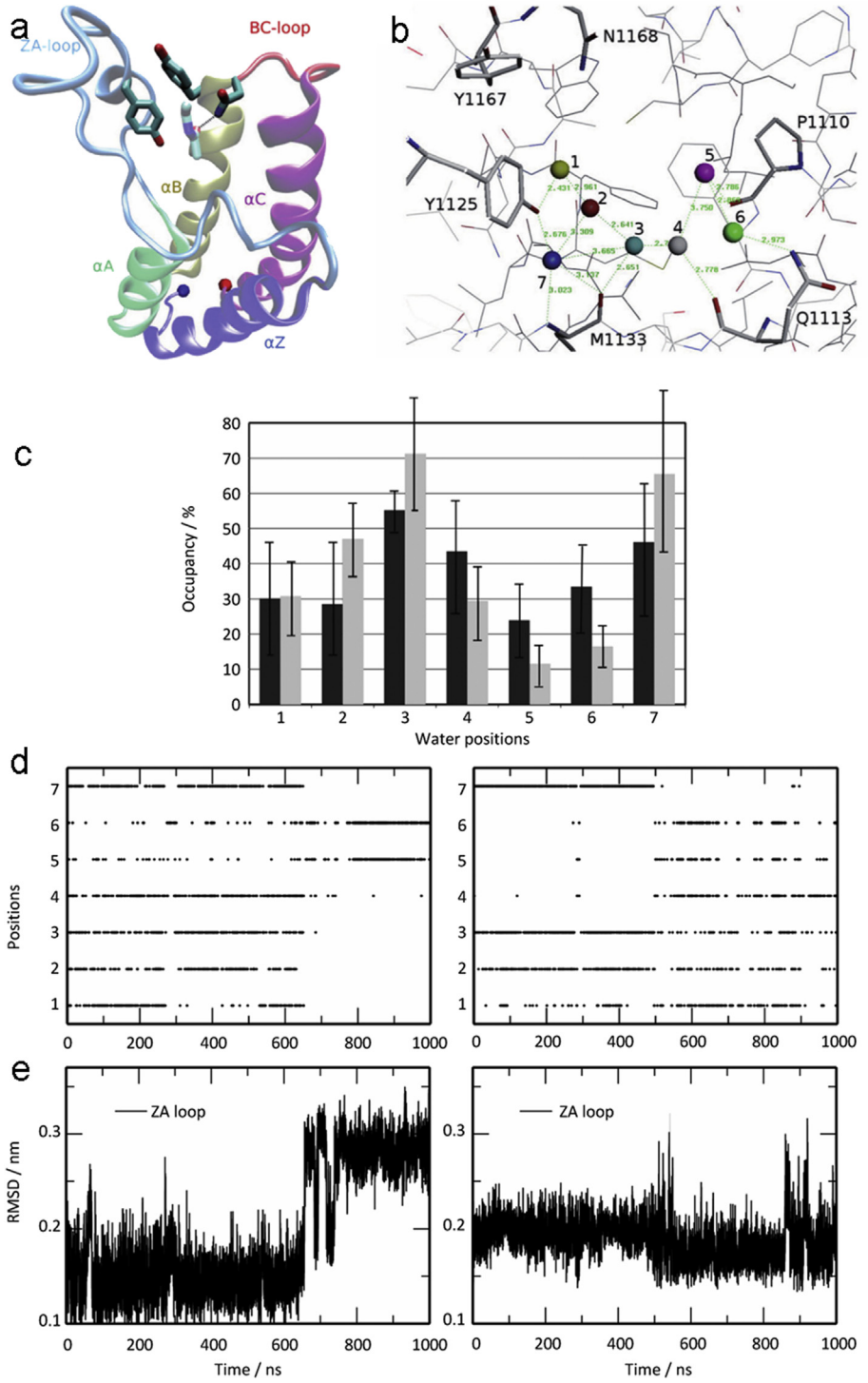


**Fig. 3.** Bromodomain structure and structured water molecules in the bottom of the acetyl-lysine binding site. (a) Ribbon illustration of the crystal structure of the complex between CREBBP and acetyl-lysine (PDB code: 3P1C). Each of the four α-helices and two binding site loops are displayed with a different color. The side chains of the conserved residues Tyr1125 of the ZA-loop, and Tyr1167–Asn1168 of the BC-loop are emphasized (sticks) together with the acetyl-lysine ligand (sticks, light colors). The N and C termini are shown with a blue and red sphere, respectively. (b) Positions of structured water molecules. (c) Percentage of water occupancy at each of the seven positions shown in panel b. Each bar shows the average and standard deviation calculated over four independent simulations for a total of 4 μs for each bromodomain (black, BAZ2B; gray CREBBP). (d) Time series of occupancy of positions of structured water molecules along one of the four 1 μs MD runs for BAZ2B (left) and CREBBP (right). (e) Time series of RMSD from the crystal structure of the carbon atoms in the ZA loop upon structural overlap of the carbon atoms of residues in nonflexible parts of the bromodomain. These two panels show results for the same MD runs shown in panel d. Reprinted with permission from Wiley-VCH Verlag GmbH & Co. KGaA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

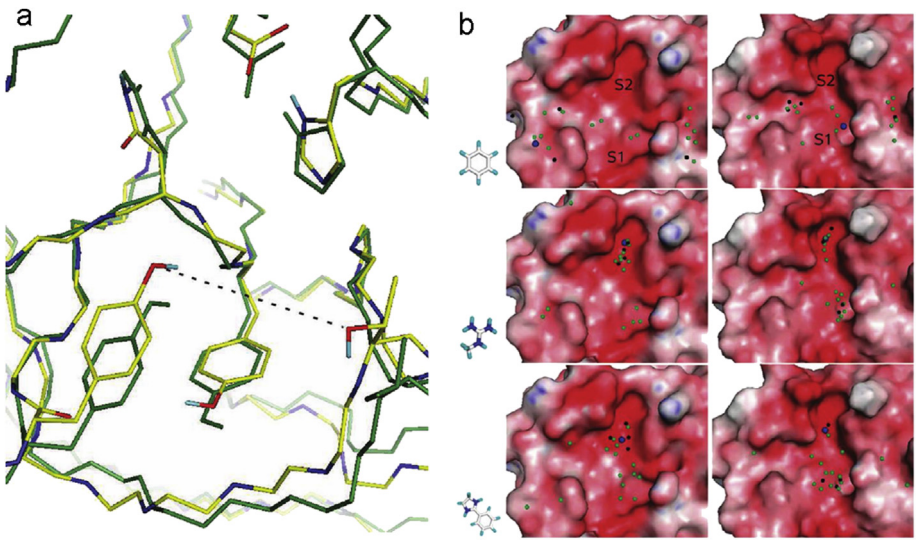


**Fig. 4.** (a) S1 pocket in the crystal structure (green) and MD snapshot 29 (colored by atom-type with carbon atoms in yellow) upon optimal overlap of the backbone atoms of WNV NS2B-NS3pro. The RMSD of backbone and side chain atoms of the S1 pocket is 0.3 Å and 1.3 Å, respectively. Note the decrease of the separation between the oxygen atoms of the side chains of Tyr161 and Thr132 (black dashed line). (b) Docking of fragments into the X-ray structure (left) and the MD conformation 29 (right). The geometrical centers of the best 20 poses of benzene (top), methylguanidinium (middle), and 2-phenylimidazoline (bottom) are represented by spheres. The large blue sphere is the best pose among all. This pose together with four black spheres represents the five poses with the most favorable binding energy as evaluated by SEED [47]. The surface of WNV NS2B-NS3pro is colored by electrostatic potential with red and blue for negative and positive potential, respectively. The figure was prepared using PyMOL (Schroedinger). Reprinted with permission from the American Chemical Society. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Importantly, the two active compounds could be docked into the crystal structure but they did not pass the filter on van der Waals efficiency. In other words, they are true positives in the structure of MD snapshot 29 and false negatives in the crystal structure.

### 3.3. Application to the Eph tyrosine kinases

The ATP binding site in protein kinase is located between the two lobes of the catalytic domain and shows pronounced plasticity particularly in two loops (activation loop or A-loop, and Gly-rich loop or G-loop) as well as in the orientation of some of the side chains (Fig. 5, right panels). Small molecule inhibitors can promote displacement of the loops and side chains, e.g., the so-called type II inhibitors which bind upon a large reorientation of the Phe side chain in the Asp–Phe–Gly triad segment (the DFG motif) preceding the A-loop. Such rearrangement results in an additional pocket called allosteric site, which is close to the ATP binding site and is usually occupied by hydrophobic moieties (e.g., trifluoromethylbenzene) in many potent and selective inhibitors of tyrosine kinases including several anti-cancer compounds used in the clinics since several years (e.g., nilotinib and sorafenib). The erythropoietin-producing hepatocellular (Eph) carcinoma receptor tyrosine kinases are an interesting family of anti-cancer drug targets because of the role of these receptors in angiogenesis and blood vessel remodeling which are essential for metastasis growth [49–52]. In a recent work, an induced fit of the active site of the tyrosine kinase EphA3 was generated by MD to accommodate a known type II inhibitor that could not be docked into the crystal structure of EphA3 [12]. First, the side chain of Tyr742 was rotated manually by about 110° around the $\chi_1$ angle to increase the aperture of the allosteric pocket (Fig. 5). The alternative orientation of the Tyr742 had been observed in a 500-ns sampling by MD (10 independent runs of 50 ns each) of the complex between the EphA3 kinase domain and trifluoromethylbenzene in the allosteric pocket. Second, a type II inhibitor was cross-docked manually into the structure of the complex of EphA3 and a different type II compound (PDB code 3DZQ). In this way, a fit in the ATP binding site of EphA3 was induced as the docked inhibitor was larger than the one in the 3DZQ structure. Third, a 2-ns MD run of the new complex was carried out with harmonic restraints on the $C_\alpha$ atoms excluding those in the G-loop and A-loop. Fourth, seven snapshots were selected from the second ns of the MD trajectory according to the presence of five key hydrogen bonds between the manually placed type II inhibitor and EphA3. Finally, flexible docking of four previously reported type II inhibitors was used to select the MD snapshot in which they had the most favorable calculated binding energy. This MD snapshot, called MD induced fit (MD-IF) structure, was used for high-throughput docking of a pharmacophore tailored library of about 175,000 compounds which yielded 3 actives out of 12 molecules tested in vitro, i.e., a hit rate of 25% (Table 1) [12]. It is important to note that the induced displacements of the loops in the ATP binding site of the EphA3 kinase and reorientation of the Tyr742 side chain are not attainable by a conventional (i.e., gradient-based) minimization of the energy.

### 3.4. Application to the ZAP70 tyrosine kinase

The zeta-chain associated protein kinase 70 kDa (ZAP70) is expressed mainly in T cells and natural killer cells, and is implicated in inflammatory and autoimmune diseases. Despite its importance as potential drug target very few inhibitors of ZAP70 have been reported as of today. We have used an MD-based protocol to induce a fit of the side chain of the so-called gatekeeper residue (Met414) and hydrophobic pocket next to it [13]. First, a potent Janus kinase 2 (JAK2) inhibitor was docked into the ATP binding site of ZAP70 by structural overlap of the kinase domains (PDB codes 3KCK and 1U59 for JAK2 and ZAP70, respectively). The steric clashes in the resulting complex of ZAP70 and the JAK2 inhibitor were then relaxed by a 2-ns MD simulation with harmonic restraints on the backbone atoms. Nineteen of the 500 snapshots saved during the second half of the MD run were prioritized as they showed close to ideal geometry for three key hydrogen bonds between the JAK2 inhibitor and ZAP70. The JAK2 inhibitor was then re-docked into each of these 19 MD snapshots, and the MD-IF snapshot with the

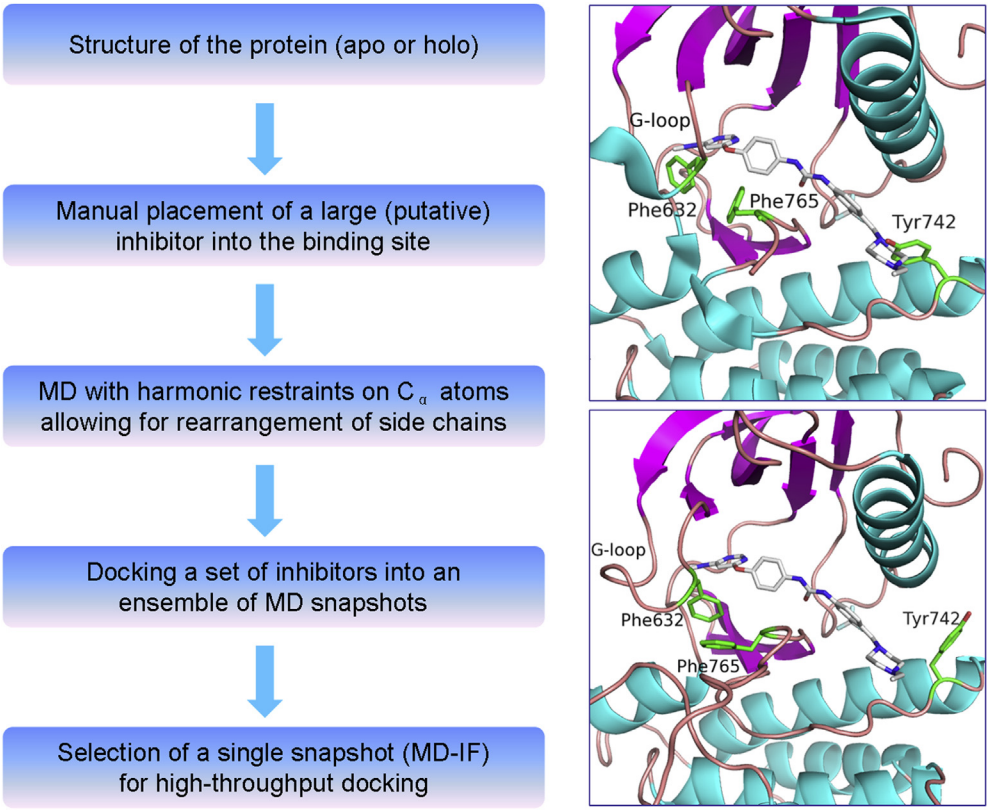


**Fig. 5.** (Left) Illustration of the MD-IF protocol. A set of known inhibitors usually cannot be docked as expected due to collapse of the binding site. The MD-IF protocol is meant to increase the aperture of the binding site for top ranking of a diversity of inhibitors by docking to one single conformer. (Top right) Manual cross-docking of an 80 nM inhibitor of the EphA3 kinase into the crystal structure of the EphA3 kinase from the complex with another inhibitor (3DZQ) resulted in severe clashes with the side chains of Tyr742, Phe765, and Phe632. (Bottom right) Automatic docking of the same inhibitor into the MD-IF structure results in a pose with very favorable interaction energy.

most favorable docking score was selected for in silico screening. High-throughput docking into this MD-IF snapshot of ZAP70 yielded 10 low-micromolar inhibitors corresponding to six distinct chemotypes. Interestingly, one of the ZAP70 inhibitors identified by docking into the MD-IF structure shows nanomolar affinity for JAK2 (compound **15** in Table 1).

In conclusion, the high hit rates of in silico screening campaigns based on an MD-IF structure of the target indicate that the approach is robust. In other words, the MD-induced increase of the aperture of sub-pockets in the binding site, with minor changes to the protein backbone, is likely to be useful in general.

## 4. MD validation of predicted binding mode

There are two types of applications of MD for validation of a binding mode obtained by docking (called also pose hereafter). The first uses MD simulations as a final filter after high-throughput docking and ranking (Subsections 4.1 and 4.3). This approach is intrinsically parallel as independent MD runs are carried out for a set of candidate ligands, and the trajectories can be analyzed in parallel. The second uses MD simulations of active compounds as a means of guiding chemical synthesis for hit (or lead) optimization (Subsections 4.1 and 4.2). The atomistic information obtained from MD sampling is particularly useful when the crystal structure of the complex is not available.

### *4.1. Application to the EphB4 tyrosine kinase*

A novel chemical class of inhibitors of the EphB4 tyrosine kinase was discovered in our group by high-throughput docking of a focused library of nearly 20,000 compounds followed by MD simulations for the final selection of compounds [11]. After docking and scoring, representatives of the four top ranking scaffolds were further investigated by multiple MD runs of 100 ns each. In the MD runs, two compounds sharing a pyrimidoisoquinolinone scaffold were involved in stable intermolecular hydrogen bonds with the hinge region while molecules representing other scaffolds showed rupture of these key hydrogen bonds within the first 10 ns (Figure S4 in the Supporting Information of Ref. [11]). Thus, we focused on six commercially available compounds bearing a pyrimidoisoquinolinone scaffold and four of them (e.g., compound **11**) showed inhibition of the EphB4 tyrosine kinase in the low micromolar range in two different enzymatic assays. Importantly, two independent MD simulations of 300 ns each (started from the pose obtained by docking) suggested that an additional hydroxyl group involved in two favorable (buried) hydrogen bonds with EphB4 would improve potency. This prediction was confirmed by the synthesis of a single derivative of one of the commercially available pyrimidoisoquinolinone, resulting in an improvement of the inhibitory potency from 8400 to 160 nM and a ligand efficiency of 0.39 kcal/mol per non-hydrogen atom. The definitive validation of the in silico predictions was obtained by the crystal structure of the complex between compound **11** and the EphA3 kinase, whose ATP-binding site is essentially identical to the one of the cognate EphB4 (PDB code 4G2F). The binding mode in the crystal structure is identical to the pose suggested by docking and validated in silico by MD (Fig. 6(B)) [11].

In another MD application, a 45-ns MD run of a low nanomolar inhibitor of the EphB4 tyrosine kinase [16] (compound 66 in Ref. [16]) was carried out to validate the binding mode obtained by

automatic docking for the prioritization of further chemical synthesis which aimed at improving cellular permeation. During the MD run the hydrogen bonds between compound 66 and the hinge region (residues 694–697 which connect the N-terminal and C-terminal lobes of the catalytic domain) were preserved as in the pose obtained by docking except for some transient ruptures of the hydrogen bond involving the carbonyl group of Met696 (Figure 5 in Ref. [16]). Overall, the 45-ns MD run of the complex of EphB4 with compound 66 confirmed that the major van der Waals contacts and polar interactions in the binding mode predicted by docking are stable. Recently, we have solved the crystal structure of the complex between compound 66 and the EphA3 kinase (PDB code 4GK2 [17]) which has provided again the definitive validation of the binding mode predicted by docking and further supported by the MD run (Fig. 6(A)).

### 4.2. Application to the SYK and ZAP70 tyrosine kinases

MD simulations were used recently to discriminate between alternative binding modes of a single-digit μM inhibitor of the ZAP70 kinase identified by high-throughput docking into a structure of the catalytic domain of the cognate SYK (spleen tyrosine kinase) with the C-helix-out orientation [14]. The inhibitor was stable in the binding mode predicted by docking along 1000-ns runs of MD. In contrast, in the four independent MD runs started from a pose in the ATP binding site of ZAP70 that does not require the C-helix-out displacement, the inhibitor (compound **17**, Fig. 1) dissociated fully within 150 ns, providing evidence that it occupies a hydrophobic pocket that is present only in the C-helix-out conformation. This MD simulation result is consistent with further experimental validation on a set of 16 derivatives of compound **17** (shown in the Supplementary Information of Ref. [14]) which can fit well into the ATP site by docking but are inactive at a test concentration of 50 μM. These results make clear that MD can effectively discriminate among putative binding modes and thus can facilitate medicinal chemistry endeavors.

### 4.3. Application to the first bromodomain of BRD4

We have carried out a fragment-based high-throughput docking campaign to identify small molecules that compete with the acetyl-lysine binding site in the first bromodomain of BRD4 [15]. Starting from a focused library of 665,184 commercially available molecules, nearly 5000 compounds survived the filters of predicted binding energy and ligand efficiency (Fig. 7). Selection of poses for MD validation was guided by predicted ligand efficiency as well as chemical diversity, rigidity, novelty and actual commercial availability. Multiple MD simulations of 55 compounds with different anchor fragments were performed starting from their docked poses into the first bromodomain of BRD4 to assess the main binding interactions, for example, the stability of the hydrogen bond with the conserved Asn140 which acts as hydrogen bond donor for the acetyl group of the natural ligand. The MD simulations revealed that the binding mode of some of the selected chemotypes was not

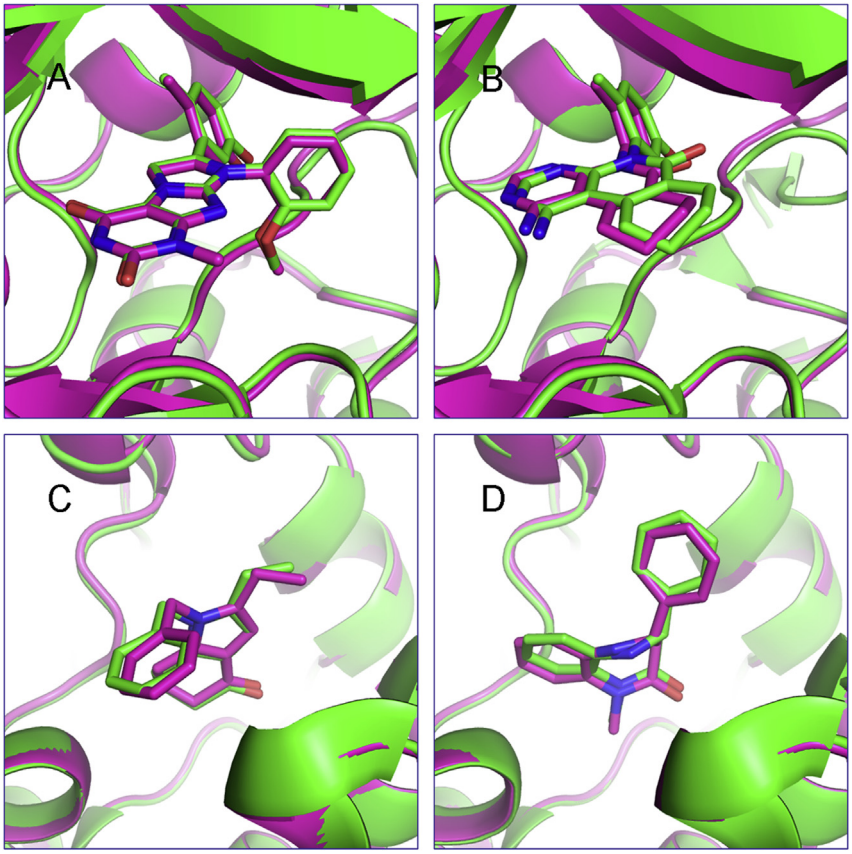


**Fig. 6.** Superposition of binding modes in crystal structures (green) and as predicted by means of docking followed by MD validation (magenta). The structural alignment is based on the $C_\alpha$ atoms of the protein target. (A) Docking pose of compound **3** of Ref. [17] into the EphB4 tyrosine kinase (PDB structure 2VWX used as target) and crystal structure of its complex with EphA3 (4GK2). (B) Docking pose of compound 7 of Ref. [11] into EphB4 (2VWX) and crystal structure of its complex with EphA3 (4G2F). (C) Docking pose of compound **18** into BRD4(1) (PDB structure 3MXF used as target) and crystal structure (4PCI). (D) Docking pose of compound **19** into BRD4(1) (3MXF) and crystal structure (4PCE). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

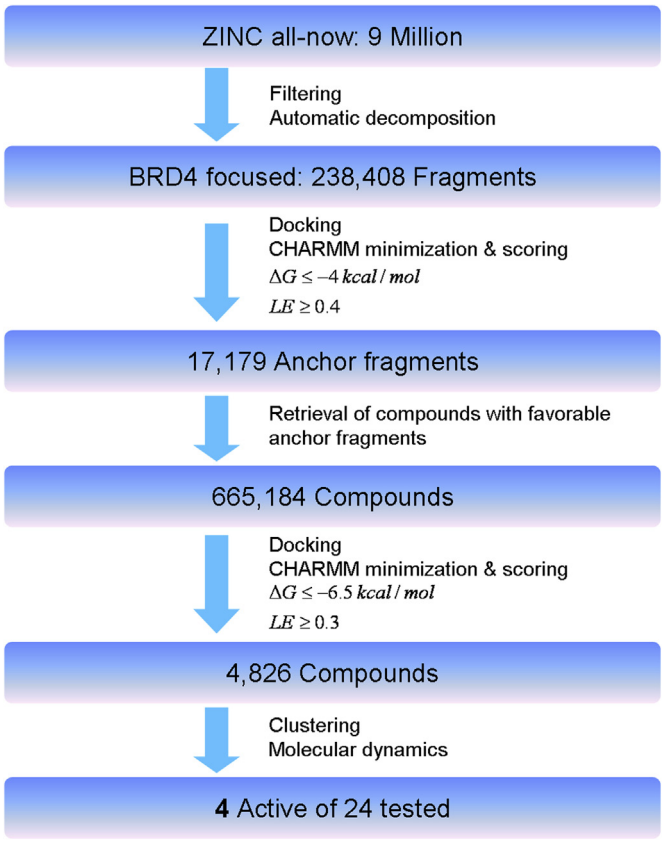


**Fig. 7.** Illustration of the ALTA procedure [7,11] for fragment-based in silico screening of BRD4(1) inhibitors. Fragment docking and flexible ligand docking were carried out with an in-house developed program [13,14] using the X-ray structure of BRD4(1) (PDB code 3MXF). Units of LE are kcal/mol per non-hydrogen atom. Reprinted with permission from Elsevier.

stable, and this information was used to reduce the number of compounds for in vitro validation. Finally, only 24 compounds were purchased and two of them showed an affinity in the single-digit μM range. After the paper was accepted we have solved the crystal structures of the first bromodomain of BRD4 in the complex with these two inhibitors, i.e., compounds **18** and **19** (Fig. 1; resolution of 1.29 Å and R-free factor of 0.165, PDB code 4PCE; resolution of 1.25 Å and R-free factor of 0.165, PDB code 4PCI). For both inhibitors, the binding mode predicted by docking and MD is identical to the one in the crystal structure (Fig. 6(C) and (D)).

## 5. MD for docking of known drugs and natural ligands

Brute force MD simulations have been employed by others to reproduce binding modes of ligands into proteins as observed in crystal structures [40,53]. We have used MD to shed light on the mechanism of drugs whose binding mode is unknown (Subsections 5.1 and 5.2). Furthermore, our MD simulations have revealed alternative binding mode of a natural ligand into an epigenetic target (Subsection 5.3).

### 5.1. Binding mode of the HIV-1 protease inhibitor darunavir

The anti-AIDS drug darunavir (Fig. 8) is a potent inhibitor of an essential protease of the human immunodeficiency virus type 1 (HIV-1). This protease is active as a homodimer of two identical 99-residue polypeptide chains. Among the dozen inhibitors of the HIV-1 protease used in the clinics, darunavir is one of the most interesting and useful because of the high genetic barrier to the development of viral resistance. Experimental evidence suggests that darunavir is able to prevent dimerization of the HIV-1 protease but the mechanism of dimerization inhibition and the binding mode(s) of darunavir to the monomeric state of the HIV-1 protease are unknown. Crucially, MD simulations for a total sampling of 11 μs and 1 μs with the CHARMM and AMBER force field, respectively, have provided evidence that the monomeric state of the HIV-1 protease is structurally stable [54]. It has to be mentioned that it is essentially impossible to fully verify this simulation result because the wild type HIV-1 protease exists only in the dimeric state at the concentrations required for biophysical characterization. On the other hand, mutants of the HIV-1 protease that destabilize the homodimeric interface can populate the monomeric state according to NMR spectroscopy data [55–57]. Moreover, NMR and size-exclusion chromatography indicate that the protease of HIV-2, another retrovirus which can cause AIDS, can assume a stable monomeric structure [58] (and its homodimeric structure is very similar to the one of the HIV-1 protease). Remarkably, spontaneous association and a major binding mode were observed during the MD simulations started with darunavir positioned randomly in the bulk solvent far away from the monomeric protease [54]. The binding mode is stabilized by favorable interactions between the apolar groups of darunavir and seven hydrophobic residues of the protease. In contrast, there are no interactions between darunavir and the catalytic aspartate (Asp25). Furthermore, the MD trajectories show stable interactions between darunavir and a segment of the protease called flap (residues 46–55), which are likely to sterically hinder the formation of the dimerization interface. Importantly, the MD results suggest that inhibition of HIV protease dimerization might be the predominant mechanism of action of darunavir particularly in the case of multiple mutations at the active site in multidrug resistant strains.

### 5.2. Binding mode of the JAK2 kinase inhibitors ruxolitinib and SAR302503

The atomic level knowledge of the binding mode of a kinase inhibitor is extremely helpful to interpret its mechanism-of-action, analyze its specificity, and rationalize possible mechanisms of resistance. In addition, it can be used to suggest points of chemical derivatization for improved potency and specificity. Ruxolitinib and SAR302503 (originally called TG101348) are two potent inhibitors of the tyrosine kinase JAK2 used in the clinics against myeloid malignancies (Fig. 8). Yet, their binding mode to JAK2 has not been reported. Recently, we have determined the binding mode of these two anticancer drugs by multiple MD runs [59]. Our MD simulations (cumulative sampling of 1.5 and 0.1 μs for Ruxolitinib and SAR302503, respectively) provide evidence that both drugs inhibit JAK2 by a so-called type I binding, in which the inhibitor targets the ATP-binding site in its active conformation and the DFG-motif at the base of the activation loop is in its inward-facing conformation. The simulation results indicate that the double-ring scaffold (7H-pyrrolo[2,3-d]pyrimidin) of Ruxolitinib is involved in two persistent hydrogen bonds with the hinge region of JAK2 (see Figure 1 of Ref. [59]). These two key interactions are preserved during all MD runs of Ruxolitinib. In contrast, the cyclopentane ring and propanenitrile, as well as the pyrazole ring, can vary their orientations with respect to the rigid double-ring system. The binding mode of SAR302503 predicted by the MD simulations is stabilized by two hydrogen bonds between the aminopyrimidin scaffold and the backbone polar groups of Leu932 in the hinge region (see Figure 1 of Ref. [59]). These hydrogen bonds are present in the crystal structure of the JAK2 kinase in complex with TG101209 (PDB code 4JI9 [60]), an inhibitor very similar to SAR302503 which was

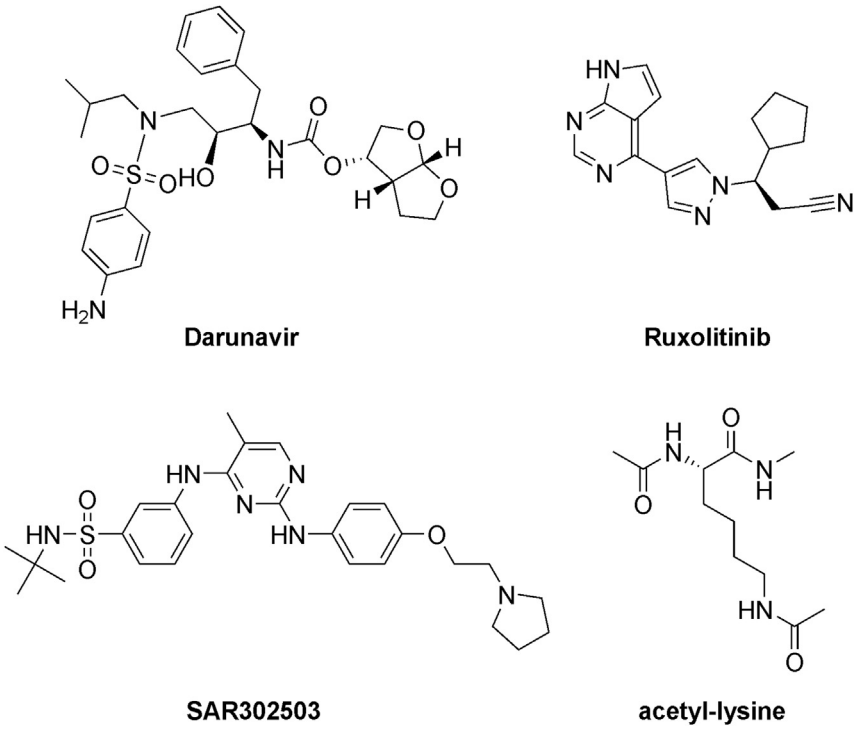


**Fig. 8.** Chemical structure of Darunavir, Ruxolitinib, SAR302503, and capped lysine with acetylated side chain (Kac).

published after the MD simulation study. Importantly, the binding modes obtained by MD simulations explain at atomic level of detail the effect of resistance-causing point mutations that have been observed in vitro [59]. Furthermore, they are very useful as a template to interpret mutations that are expected to arise in patients treated with these two inhibitors of JAK2.

### 5.3. Binding mode of the acetylated lysine side chain into human bromodomains

In another simulation study we have investigated the binding of the natural ligand acetyl-lysine (Kac, Fig. 8) to the second bromodomain of the human TAF1 protein (TAF1(2)) by multiple MD runs started from the fully unbound state [61]. This study has shed light for the first time on the binding process of a post-translationally modified amino acid side chain to an epigenetic reader module. Spontaneous binding of Kac to the TAF1(2) bromodomain was observed in two thirds of the 24 trajectories of 500 ns each. The MD simulations reproduced the binding mode observed in the co-crystal structure of Kac with the CREBBP bromodomain (PDB code 3P1C) whose Kac binding site is very similar to the one of TAF1(2) for which the structure of the complex has not been solved. Moreover, the MD trajectories revealed that this binding mode reversibly converted into a more buried binding state in which three of the six structured water molecules (mentioned above) are replaced by the amide group of the Kac side chain. In the more buried pose, the backbone carbonyl oxygen of the proline of the so-called WPF shelf (a Trp–Pro–Phe motif in the ZA loop) acts as acceptor for a hydrogen bond with the NH group of the Kac side chain, an interaction not observed in the crystal structures. Furthermore, a detailed analysis of the free energy surface and binding kinetics revealed that the inter-conversion between the two binding modes is an order of magnitude faster than the unbinding process. The existence of the second, more buried binding mode of Kac was further validated by control simulations with the capped tetrapeptide Kac–Gly–Gly–Kac to analyze potential discrepancies between the isolated Kac and a histone tail segment. Independent MD simulations were carried out with two different force fields and two bromodomains (BRD4(1) and CREBBP), and the two binding modes were observed in almost all runs [61]. It is interesting to note that after the discovery by means of MD simulations of the hydrogen acceptor role of the carbonyl group of the Pro in the middle of the WPF shelf [61], two experimental studies have reported inhibitors that exhibited a similar binding pose with a hydrogen bond to the carbonyl oxygen of the corresponding Pro residue of the BAZ2B and BRD4(1) bromodomains [62,63].

## 6. Conclusions

All-atom MD simulations of pharmaceutically relevant protein targets in their apo state or in complex with (putative) ligands have provided useful information at the beginning and final phase, respectively, of high-throughput docking campaigns. Explicit solvent MD runs of the apo structure are becoming more frequent for the selection of one or more snapshots for docking of large libraries of compounds. At the final stage of ranking, MD simulations starting from the predicted binding mode are performed for the in silico validation of the top ranking compounds. Moreover, atomistic MD simulations have provided insights into the mechanism of the antiviral drug darunavir (a dimerization inhibitor of the HIV protease, Subsection 5.1); have helped in the interpretation of the selectivity profile of two kinase inhibitors used in the clinics as anticancer drugs (Subsection 5.2); and have revealed that acetyl-lysine (the natural ligand of bromodomains) has an alternative binding mode which is more buried than the one observed in the available crystal structures (Subsection 5.3). These and several other examples (some of which are reviewed in section 4 of Ref. [64]) suggest that the applications of MD will play an even more important role in drug design in the future. There are three main obstacles to the identification of potent (i.e., nanomolar) inhibitors by high-

throughput docking: the very small chemical space of the libraries of compounds, the approximations used for scoring, and the use of a (mainly) rigid protein structure [65]. The MD protocols reviewed in this work and future developments of MD-based methods for in silico screening will help to remove the two latter obstacles.

## Acknowledgments

AC thanks the talented PhD students and postdoctoral associates who have developed methods for docking and scoring, and carried out the docking campaigns and MD simulations in his group at the University of Zurich. We thank Dr. Jing Dong for solving the crystal structures of several kinase/inhibitor and bromodomain/inhibitor complexes (PDB codes are listed in Table 1). We thank Dr. Dimitrios Spiliotopoulos for useful discussions and comments to the manuscript. The MD simulations reported by the group of AC were run on the Matterhorn and Schrödinger compute clusters at the University of Zurich. This work was supported in part by the Swiss National Science Foundation (Grant no. 315230_149897/1) (grant to AC).

## References

[1] D. Huang, U. Luthi, P. Kolb, K. Edler, M. Cecchini, S. Audetat, A. Barberis, A. Caflisch, J. Med. Chem. 48 (2005) 5108.
[2] D. Huang, U. Luthi, P. Kolb, M. Cecchini, A. Barberis, A. Caflisch, J. Am. Chem. Soc. 128 (2006) 5436.
[3] R. Freidman, A. Caflisch, ChemMedChem 4 (2009) 1317.
[4] D. Ekonomiuk, X.C. Su, K. Ozawa, C. Bodenreider, S.P. Lim, Z. Yin, T.H. Keller, D. Beer, V. Patel, G. Otting, A. Caflisch, D. Huang, PLoS Negl. Trop. Dis. 3 (2009) e356.
[5] D. Ekonomiuk, X.C. Su, K. Ozawa, C. Bodenreider, S.P. Lim, G. Otting, D. Huang, A. Caflisch, J. Med. Chem. 52 (2009) 4860.
[6] P. Schenker, P. Alfarano, P. Kolb, A. Caflisch, A. Baici, Protein Sci. 17 (2008) 2145.
[7] P. Kolb, C.B. Kipouros, D. Huang, A. Caflisch, Proteins 73 (2008) 11.
[8] P. Kolb, D. Huang, F. Dey, A. Caflisch, J. Med. Chem. 51 (2008) 1179.
[9] T. Zhou, A. Caflisch, ChemMedChem 5 (2010) 1007.
[10] H. Zhao, D. Huang, PLoS One 6 (2011) e19923.
[11] H.T. Zhao, J. Dong, K. Lafleur, C. Nevado, A. Caflisch, ACS Med. Chem. Lett. 3 (2012) 834.
[12] H. Zhao, D. Huang, A. Caflisch, ChemMedChem 7 (2012) 1983.
[13] H. Zhao, A. Caflisch, Bioorg. Med. Chem. Lett. 23 (2013) 5721.
[14] H. Zhao, A. Caflisch, Bioorg. Med. Chem. Lett. 24 (2014) 1523.
[15] H. Zhao, L. Gartenmann, J. Dong, D. Spiliotopoulos, A. Caflisch, Bioorg. Med. Chem. Lett. 24 (2014) 2493.
[16] K. Lafleur, D. Huang, T. Zhou, A. Caflisch, C. Nevado, J. Med. Chem. 52 (2009) 6433.
[17] K. Lafleur, J. Dong, D. Huang, A. Caflisch, C. Nevado, J. Med. Chem. 56 (2013) 84.
[18] P.J. Goodford, J. Med. Chem. 28 (1985) 849.
[19] A. Miranker, M. Karplus, Proteins 11 (1991) 29.
[20] S.B. Shuker, P.J. Hajduk, R.P. Meadows, S.W. Fesik, Science 274 (1996) 1531.
[21] H.J. Bohm, J. Comput. Aided Mol. Des. 6 (1992) 61.
[22] A. Caflisch, A. Miranker, M. Karplus, J. Med. Chem. 36 (1993) 2142.
[23] V.L. Nienaber, P.L. Richardson, V. Klighofer, J.J. Bouska, V.L. Giranda, J. Greer, Nat. Biotechnol. 18 (2000) 1105.
[24] M. Pellecchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T.L. James, S.W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe, G. Siegal, Nat. Rev. Drug. Discov. 7 (2008) 738.
[25] S. Geschwindner, L.L. Olsson, J.S. Albert, J. Deinum, P.D. Edwards, T. de Beer, R.H. Folmer, J. Med. Chem. 50 (2007) 5903.
[26] E.E. Swayze, E.A. Jefferson, K.A. Sannes-Lowery, L.B. Blyn, L.M. Risen, S. Arakawa, S.A. Osgood, S.A. Hofstadler, R.H. Griffey, J. Med. Chem. 45 (2002) 3816.
[27] D.A. Ockey, J.L. Dotson, M.E. Struble, J.T. Stults, J.H. Bourell, K.R. Clark, T.R. Gadek, Bioorg. Med. Chem. 12 (2004) 37.
[28] W.J. Wood, A.W. Patterson, H. Tsuruoka, R.K. Jain, J.A. Ellman, J. Am. Chem. Soc. 127 (2005) 15521.
[29] A.W. Patterson, W.J. Wood, M. Hornsby, S. Lesley, G. Spraggon, J.A. Ellman, J. Med. Chem. 49 (2006) 6298.
[30] A.W. Patterson, W.J. Wood, J.A. Ellman, Nat. Protoc. 2 (2007) 424.
[31] D.A. Erlanson, J.W. Lam, C. Wiesmann, T.N. Luong, R.L. Simmons, W.L. DeLano, I.C. Choong, M.T. Burdett, W.M. Flanagan, D. Lee, E.M. Gordon, T. O'Brien, Nat. Biotechnol. 21 (2003) 308.
[32] D.A. Erlanson, J.A. Wells, A.C. Braisted, Annu. Rev. Biophys. Biomol. Struct. 33 (2004) 199.
[33] O. Guvench, A.D. MacKerell Jr., PLoS Comput. Biol. 5 (2009) e1000435.
[34] J. Seco, F.J. Luque, X. Barril, J. Med. Chem. 52 (2009) 2363.
[35] E.P. Raman, W. Yu, O. Guvench, A.D. Mackerell, J. Chem. Inf. Model. 51 (2011) 877.
[36] D. Huang, A. Caflisch, ChemMedChem 6 (2011) 1578.
[37] D. Huang, A. Caflisch, PLoS Comput. Biol. 7 (2011) e1002002.
[38] P. Burkhard, P. Taylor, M.D. Walkinshaw, J. Mol. Biol. 295 (2000) 953.
[39] F. Rao, A. Caflisch, J. Mol. Biol. 342 (2004) 299.
[40] I. Buch, T. Giorgino, G. De Fabritiis, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 10184.
[41] P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J.P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Muller, T. Pawson, A.C. Gingras, C.H. Arrowsmith, S. Knapp, Cell 149 (2012) 214.
[42] D. Huang, E. Rossini, S. Steiner, A. Caflisch, ChemMedChem 9 (2014) 573.
[43] J.H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon, J. Am. Chem. Soc. 124 (2002) 5632.
[44] K.W. Lexa, H.A. Carlson, Q. Rev. Biophys. 45 (2012) 301.
[45] R. Baron, J.A. McCammon, Annu. Rev. Phys. Chem. 64 (2013) 151.
[46] A.H. Chan, J. Wereszczynski, B.R. Amer, S.W. Yi, M.E. Jung, J.A. McCammon, R.T. Clubb, Chem. Biol. Drug Des. 82 (2013) 418.
[47] N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhardt, A. Caflisch, Proteins 37 (1999) 88.
[48] N. Budin, N. Majeux, A. Caflisch, Biol. Chem. 382 (2001) 1365.
[49] R.H. Adams, Semin. Cell. Dev. Biol. 13 (2002) 55.
[50] R. Noberini, I. Lamberto, E.B. Pasquale, Semin. Cell. Dev. Biol. 23 (2012) 51.
[51] G. Martiny-Baron, P. Holzer, E. Billy, C. Schnell, J. Brueggen, M. Ferretti, N. Schmiedeberg, J.M. Wood, P. Furet, P. Imbach, Angiogenesis 13 (2010) 259.
[52] S.A. Mitchell, M.D. Danca, P.A. Blomgren, J.W. Darrow, K.S. Currie, J.E. Kropf, S.H. Lee, S.L. Gallion, J.M. Xiong, D.A. Pippin, R.W. DeSimone, D.R. Brittelli, D.C. Eustice, A. Bourret, M. Hill-Drzewi, P.M. Maciejewski, L.L. Elkin, Bioorg. Med. Chem. Lett. 19 (2009) 6991.
[53] Y. Shan, M.P. Eastwood, X. Zhang, E.T. Kim, A. Arkhipov, R.O. Dror, J. Jumper, J. Kuriyan, D.E. Shaw, Cell 149 (2012) 860.
[54] D.Z. Huang, A. Caflisch, J. Chem. Theory Comput. 8 (2012) 1786.
[55] R. Ishima, R. Ghirlando, J. Tozser, A.M. Gronenborn, D.A. Torchia, J.M. Louis, J. Biol. Chem. 276 (2001) 49110.
[56] J.M. Louis, R. Ishima, I. Nesheiwat, L.K. Pannell, S.M. Lynch, D.A. Torchia, A.M. Gronenborn, J. Biol. Chem. 278 (2003) 6085.
[57] R. Ishima, D.A. Torchia, S.M. Lynch, A.M. Gronenborn, J.M. Louis, J. Biol. Chem. 278 (2003) 43311.
[58] J.M. Louis, R. Ishima, A. Aniana, J.M. Sayer, Protein Sci. 18 (2009) 2442.
[59] T. Zhou, S. Georgeon, R. Moser, D.J. Moore, A. Caflisch, O. Hantschel, Leukemia 28 (2014) 404.
[60] M. Siu, R. Pastor, W. Liu, K. Barrett, M. Berry, W.S. Blair, C. Chang, J.Z. Chen, C. Eigenbrot, N. Ghilardi, P. Gibbons, H. He, C.A. Hurley, J.R. Kenny, S. Cyrus Khojasteh, H. Le, L. Lee, J.P. Lyssikatos, S. Magnuson, R. Pulk, V. Tsui, M. Ultsch, Y. Xiao, B.Y. Zhu, D. Sampath, Bioorg. Med. Chem. Lett. 23 (2013) 5014.
[61] A. Magno, S. Steiner, A. Caflisch, J. Chem. Theory Comput. 9 (2013) 4225.
[62] F.M. Ferguson, O. Fedorov, A. Chaikuad, M. Philpott, J.R. Muniz, I. Felletar, F. von Delft, T. Heightman, S. Knapp, C. Abell, A. Ciulli, J. Med. Chem. 56 (2013) 10183.
[63] X. Lucas, D. Wohlwend, M. Hugle, K. Schmidtkunz, S. Gerhardt, R. Schule, M. Jung, O. Einsle, S. Gunther, Angew. Chem. Int. Ed. Engl. 52 (2013) 14055.
[64] D. Spiliotopoulos, A. Caflisch, Isr. J. Chem. (2014), http://dx.doi.org/10.1002/ijch.201400009.
[65] D. Huang, A. Caflisch, J. Mol. Recognit. 23 (2010) 183.
[66] D. Huang, A. Caflisch, J. Med. Chem. 47 (2004) 5791.