Bioinformatics, 2026, **42**(2), btag027
https://doi.org/10.1093/bioinformatics/btag027
Published: 22 January 2026
Original Paper | Structural bioinformatics

OXFORD

De novo protein–ligand design including protein flexibility and conformational adaptation

Jakob Agamia and Martin Zacharias*

Center of Functional Protein Assemblies and Physics Department, Technical University Munich, Bavaria 85748, Germany
*Corresponding author. Center of Functional Protein Assemblies and Physics Department, Technical University Munich, Ernst-Otto-Fischer-Straße 8, Bavaria 85748, Germany. E-mail: zacharias@tum.de
Associate Editor: Lenore Cowen

Abstract

Motivation: **The rational design of chemical compounds that bind to a desired protein target molecule is a major goal of drug discovery. Most current molecular docking but also fragment-based buildup or machine learning-based generative drug design approaches employ a rigid protein target structure.**

Results: **Based on recent progress in predicting protein structures and complexes with chemical compounds, we have designed an approach, AI-MCLig, to optimize a chemical compound bound to a fully flexible and conformationally adaptable protein binding region. During a Monte Carlo (MC)-type simulation to randomly change a chemical compound, the target protein–compound complex is completely rebuilt at every MC step using the Chai-1 protein structure prediction program. Besides compound flexibility it allows the protein to adapt to the chemically changing compound. MC protocols based on atom-/bond-type changes or based on combining larger chemical fragments have been tested. Simulations on four test targets resulted in potential ligands that show very good binding scores comparable to experimentally known binders using several different scoring schemes. The MC-based compound design approach is complementary to existing approaches and could help for the rapid design of putative binders including induced fit of the protein target.**

Availability and implementation: **Datasets, examples, and source code are available on our public GitHub repository https://github.com/JakobAgamia/AI-MCLig and on Zenodo at https://doi.org/10.5281/zenodo.17800140.**

Introduction

Protein molecules and complexes formed between proteins and other types of biomolecules play key roles in basically all biological processes (Zahiri *et al.* 2020). Many drug design efforts aim at identifying or developing small organic molecules that specifically bind to target proteins and interfere with its function (Horvath 2010, Śledź and Caflisch 2018, Tang *et al.* 2024b). Such ligands can potentially inhibit protein function or influence interactions with biomolecule partners (Martino *et al.* 2021).

Computational approaches play an increasing role to screen for ligands that can bind to protein pockets with high affinity. Traditionally, *in silico* pharmacophore screening and molecular docking methods are used to select putative binders from large data bases (Śledź and Caflisch 2018, Liu *et al.* 2020, Mouchlis *et al.* 2021, Corrêa Veríssimo *et al.* 2025). Screening of large data bases with multibillion of compounds (Grygorenko *et al.* 2020, Lyu *et al.* 2023) is computationally demanding and faces the difficulty to accurately score putative binders (Schneider *et al.* 2020, Lyu *et al.* 2023, Sindt *et al.* 2025). An additional difficulty arises from the fact that most virtual screening methods employ rigid protein target structures, completely neglecting even small adaptations of the protein structure upon ligand binding (Beier and Zacharias 2010, Śledź and Caflisch 2018). In recent years, artificial intelligence (AI)-driven methods have gained popularity and allow the generation of entirely new compounds in *de novo* drug design (Liu *et al.* 2020, Schneider *et al.* 2020, Mouchlis *et al.* 2021, Lyu *et al.* 2023, Tang *et al.* 2024a). Aim of *de novo* compound design methods is to identify or newly generate chemical compounds (generative models) that fit to a given target protein pocket and bind with high affinity. A variety of AI-based methods are available ranging from generative adversarial networks, variational auto-encoders and more recently diffusion models (e.g. reviewed in Tang *et al.* 2024a). The methods are trained on large databases of known protein–ligand complexes (Liu *et al.* 2017, Zdrazil *et al.* 2023).

Typically, such approaches employ a defined and fixed (rigid) target pocket and can generate entirely new potential binders, thereby including also biases for solubility and chemical accessibility of the generated compounds (Xie *et al.* 2022, Tang *et al.* 2024b). However, one major drawback of both traditional molecular docking and generative models is the typical assumption

Received: 25 July 2025. **Revised:** 18 December 2025. **Accepted:** 8 January 2026

of a rigid protein target pocket that realistically may change shape and adapt to some degree to a bound ligand (Beier and Zacharias 2010, Śledź and Caflisch 2018).

AI-based structure prediction approaches such as AlphaFold2 (AF2) (Jumper *et al.* 2021) allow the accurate 3D modeling of proteins and protein complexes. More recently, the deep learning methodology was extended to not only accurately model protein structures but also complexes with organic molecules using AlphaFold3 (AF3) (Abramson *et al.* 2024) and related programs such as Boltz1 (Wohlwend *et al.* 2024) and Chai-1 (Chai Discovery 2024). In particular, Chai-1 allows quite rapid generation of protein 3D structures in complex with organic drug-like ligands.

Based on Chai-1, we have developed Monte Carlo (MC)-type simulation approaches in chemical space (termed AI-MCLig) to generate new organic compounds that potentially bind at a protein target region. The first protocol uses a set of atomistic (e.g. atom-type, bond-type) chemical changes to modify a simple starting structure (e.g. benzene). The second method recombines molecular fragments based on the BRICS (Degen *et al.* 2008) method. We demonstrate that the AI-MCLig approach can recover a desired ligand if structural similarity is used as target score. In case of *de novo* ligand generation for a desired protein target, for both protocols the ligand–protein complex is completely rebuilt (using the AI-based Chai-1 method at each MC step). Besides of including compound flexibility, it offers an important additional advantage of the method allowing for full conformational adaptation of the protein structure upon ligand modification. Whether or not the change is accepted, is based mostly on the change in the Chai-1 confidence score. On four examples we demonstrate that the MC method allows to rapidly generate new putative binders with favorable scoring including also alternative force field-based scoring. The method could complement traditional molecular docking as well as generative *de novo* drug design approaches that mostly employ rigid binding pockets.

Materials and methods

MC simulation approach

In the MC simulation, the ligand is built by modifying a current compound structure with elementary chemical steps (see below). The modifications are introduced using the rdkit package (Landrum 2023) which includes routines to ensure the chemical validity of the modified compound. The compound chemical SMILES (Weininger 1988) and the protein sequence are used with Chai-1 (Chai Discovery 2024) for AI-based rebuilding and evaluation (termed AI-MCLig approach). The scoring includes mostly the confidence score, a synthetic accessibility bias, a solubility bias and a drug-likeness bias (see paragraph below). The newly generated ligand structure and placement is accepted if the score improves, otherwise the structure is only accepted with a probability of

$$P(\Delta s) = \exp(-\beta \cdot \Delta s), \quad (1)$$

where Δs is the difference between the new score and current score and β is a factor to influence the probability of accepting a less favorable structure in the simulation. It corresponds to the inverse of an effective temperature of the MC-type simulation (small β translates to a broad sampling of compounds with various scores and high β is a very selective search accepting mostly well-scoring compounds).

The chemical steps used by the simulation are:

Adding small chemical groups (standard probability .3): In this step, a small chemical group (e.g. CH3, OH) is added at a randomly chosen accessible point in the molecule replacing another group (e.g. H-atom).

Removing atoms (standard probability .2): A randomly chosen atom in the molecule is removed.

Changing atom types (standard probability .1): The type of a randomly chosen atom is changed to a different type.

Adding atoms in chain (standard probability .15): An atom is added in between two atoms. While this step is not strictly necessary to reach all chemical structures, it helps to avoid local minima.

Changing bond types (standard probability .025): A randomly chosen bond is changed from single to double or vice versa.

Forming rings (standard probability .025): Chains of sufficient length to form a ring are identified. One of those is randomly chosen and a bond is formed between two atoms to form a ring structure.

Breaking up rings (standard probability .05): A random bond in a ring is selected and removed to break up a ring. If the ring was aromatic, the involved atoms and bonds are automatically set to nonaromatic.

Turning rings aromatic (standard probability .05): A randomly chosen nonaromatic ring is turned aromatic or vice versa.

Rearranging bonds (standard probability .1): In this step, an atom with more than two bonds is selected. One of the bonds is removed and reformed with a neighboring atom. This step is again not strictly necessary but helpful in avoiding local minima.

Which step is used each time is randomly chosen with probabilities for chemical changes as given above. The probabilities for each step were determined by testing different probability sets on three targets (bromodomain, p38 map kinase, and Pim-1 kinase). The set of probabilities used in the MC search correspond to the most efficient combination (see Fig. 1, available as supplementary data at *Bioinformatics* online, for comparison with other probabilities). It should be mentioned that since the simulations are time-consuming the number of tested parameter sets is limited. Hence, optimal parameter sets might also vary for different protein targets which will be investigated in future studies. Each 100th step, the structure is reset to the best structure found up to that point avoiding local minima. We term this type of MC simulation atomistic-step MC simulation.

Fragment-based MC simulation

The fragment-based MC simulation follows the same concept; however, instead of using basic chemical steps, the ligand is built by recombining molecular fragments based on rdkit's (Landrum 2023) implementation of the BRICS (Degen *et al.* 2008) method. During the simulation, entire chemical fragments are

added or removed from the molecule using the same criteria as in the atomistic-step MC procedure. To determine which fragments are compatible with each other, the rules from rdkit's (Landrum 2023) implementation of BRICS were followed. The two steps used by this simulation are:

Adding fragments: A placeholder in the molecule is chosen at random and an appropriate fragment is added at that position.
Removing fragments: A random bond between fragments is chosen. It is removed and replaced with the placeholder, which occupied that position before the bond.

Scoring of compounds during MC simulations

As mentioned before, the score for whether a change is kept or not is mostly based on the Chai-1 confidence score. However, three biases are introduced to the score to enforce desired properties:

Synthetic accessibility (SA) bias: The ligands resulting from the simulations should not only have good binding affinities but also be synthesized with reasonable effort. Therefore, a bias with the SA score (Ertl and Schuffenhauer 2009) is added. It should be noted that for these simulations, the SA score is considered maximal below a threshold. This is because attempting to enforce a very low SA score is unreasonable, as the ligand must have a certain amount of complexity to fit the binding site.
Solubility bias: To ensure the resulting molecules are soluble in water, a bias with the Estimated SOLubility (ESOL) score (Delaney 2004) is added.
Drug-likeness bias: The ligands should ideally have drug-like properties. Therefore rdkit's quantitative estimate of drug-likeness (QED) score, which is based on Bickerton *et al.*'s (2012) study, is added (neglecting molecular weight factor).

With these biases, the final score used for all simulations is derived as

$$\text{score} = 0.8 \cdot \text{Chai} - 1 + 0.1 \cdot \text{SA} + 0.05 \cdot \text{ESOL} + 0.05 \cdot \text{QED}, \quad (2)$$

with the SA and ESOL scores being normalized to also lie between zero and one. The contribution of the Chai-1 score is set to 0.8, as it has been found that a simulation with a smaller contribution often fails to maximize this score. The contribution of the SA score should only be high enough to hold the simulation under the given threshold. A value of 0.1 has been found to do this in almost all cases. The QED and ESOL scores only make small contributions. However, even with these low scoring contributions, we found that the QED and ESOL scores of the generated compounds remain in a reasonable range for most of the tested cases and still allow generation of a large variety of compounds. In the Supplementary Material, available as supplementary data at *Bioinformatics* online (see "Results and discussion" section), simulations with a variation of score compositions are included, demonstrating the deviations from our combination indicated above generally lead to worse results.

Molecular dynamics simulations and MMGBSA calculations

Some of the resulting ligands were additionally evaluated with molecular dynamics (MD) simulations and molecular mechanics generalized born surface area (MMGBSA) calculations (Miller *et al.* 2012). The initial coordinate files were generated using Chai-1 (Chai Discovery 2024). Ligand force field parameters were obtained using the antechamber program (Wang *et al.* 2006a, 2006b). The tleap module of the Amber package (Case *et al.* 2023, 2025) was used to solvate the protein–ligand complex using TIP3P water model (Jorgensen *et al.* 1983). The complex was energy minimized (2000 steps) and then heated within 1 ns to a temperature of 300 K including positional restraints on each nonhydrogen atom. Subsequently, an 20 ns unrestraint MD simulation using pmemd.cuda (Goetz *et al.* 2012, Le Grand *et al.* 2013, Salomon-Ferrer *et al.* 2013, Case *et al.* 2023, 2025) was performed. The first 2.5 ns of the trajectory were disregarded. The remaining trajectory (750 frames) was then processed by the MMGBSA tool of the Amber package (using igb = 5) (Miller *et al.* 2012, Case *et al.* 2023). The MDAnalysis package (Theobald 2005, Liu *et al.* 2010, Michaud-Agrawal *et al.* 2011, Van Der Walt *et al.* 2011, Gowers *et al.* 2016) was used for calculation of root-mean-square deviations.

Results and discussion

MC-based recovery of target ligands from simple starting compounds

For our MC-type searches in chemical space, it is important to assess, whether relevant chemical structures can in principal be reached using the employed MC steps. To this end, MC simulations were performed with the regular score being replaced with a dice similarity score, derived with rdkit (Landrum 2023). The dice score is calculated from atom-pair fingerprints (Carhart *et al.* 1985), as it has been found that this type of fingerprints can be calculated very efficiently in our MC simulations. The similarity was calculated between the current compound and a known target compound. The target compounds were taken from Liu *et al.* (2017), mostly from complexes of the bromodomain, serine/threonine-protein kinase Pim-1 and p38 map kinase. The same targets were subsequently also used for the design of new putative binders. Note that the dice similarity score for a compound with respect to a reference compound represents a very rough scoring landscape in chemical space with many local minima and barriers. Therefore, the atom-based simulations consisted of two stages. For each case, a simple benzene molecule was used as a starting compound. In the first stage, 30 simulations were run simultaneously for 5000 steps. Every 500 steps, the structure of all 30 simulations was set to the best structure found by all simulations. The second stage consisted of a single simulation with 10 000 MC steps and reset to the best structure found at every 100th step. The β factor was set to 50 for all simulations to provide high selectivity.

This factor leads to a reasonable acceptance rate of 0.1–0.2 to select efficiently among a very large number of possible chemical changes those few that lead in the right direction of similarity to the target compound.

A score of 1 indicates identity between MC generated structure and target structure (see Figs 2 and 3, available as supplementary data at *Bioinformatics* online). Especially for ligands with many ring structures, it is often very difficult to escape out of certain local minima. Nevertheless, the results indicate (Table 1) that for most cases the approach succeeds in producing exactly the desired target compound. For some of the more complicated compounds, the simulation got stuck in local minima that slightly differ from the target reference compound. Nevertheless, it indicates that even on a very rough score landscape, the MC procedure can reach or come very close to a desired target chemical structure. The probabilities for different types of chemical changes for each of the MC step are shown in the "MC simulation approach" section. Variations of these step parameters and of the effective simulation temperature were tested (Fig. 4, Tables 1 and 2, available as supplementary data at *Bioinformatics* online) but resulted in reduced performance compared to the standard parameters (see "Materials and methods" section).

For the MC search approach based on combining chemical fragments, the known ligand compound was broken down into fragments used in a simulation of 10 000 steps. Since the addition or removal of whole chemical fragments causes larger dice score changes compared to the atomistic-step MC approach, a smaller β factor of 5 was necessary to achieve a reasonable MC acceptance rate. As fragments can only be removed or added, often reaching a new structure may require acceptance of intermediates with a significantly reduced score. The standard probabilities for each type of step were .6 for adding fragments and .4 for removing fragments. The structure was reset to the best structure found at every 50th step. A random fragment of the target ligand was used as starting structure. All tested ligand compounds were successfully reassembled (Table 3, available as supplementary data at *Bioinformatics* online). Examples for the dice score versus MC steps are given in Fig. 5, available as supplementary data at *Bioinformatics* online, including different choices of β and probability parameters for adding or removing fragments.

Correlation of Chai-1 confidence score and ligand binding affinity

Similar to AF3 or Boltz1, the Chai-1 structure prediction program provides a pLDDT confidence score that is typically a very good measure for the reliability of a predicted protein structure. Such confidence score is also provided in case of a protein–ligand complex. However, it is not clear how well it correlates with experimental binding affinities for known complexes. We selected three protein cases for which structures of many complexes and corresponding binding affinities are available. To this end, the protein–ligand complexes for the bromodomain, serine/threonine–protein kinase Pim-1, and p38 map kinase were evaluated for a number of different ligands (data taken from PDBbind+ data set; Liu *et al.* 2017). The Chai-1 scores were compared to the known binding affinities (Fig. 1). While there is no direct correlation, a very high Chai-1 score does typically correspond to a high average binding affinity. Although the Chai-1 confidence score is far from ideal, the results suggest, however, that maximizing the Chai-1 score is a reasonable approach for finding ligands with potentially high binding affinities.

Nevertheless, it is also important to realize that the realistic scoring of protein–ligand complexes is still a major challenge and needs further improvement in future work. We also investigated the possibility to use other more force field-based scoring provided by the Autodock Vina docking approach (Trott and Olson 2010, Eberhardt *et al.* 2021) as an alternative to the Chai-1 confidence score. The correlation between the Autodock score and the binding affinity was similarly tested (Fig. 6, available as supplementary data at *Bioinformatics* online) and does not

Table 1 The target and resulting compounds in SMILES representation for atomistic-step MC simulations.

Target(SMILES) (from dB)	Final closest compound	Dice score
CNC(=O)c1[nH]c(c(c1CC)C(=O)C)C (4lzs)	CCc1c(C(=O)NC)[nH]c(C)c1C(C)=O	1.00
CNc1cnn(c(=O)c1Cl)C (5mli)	CNc1cnn(C)c(=O)c1Cl	1.00
O=C1NC(=O)/C(=C/c2cccc(c2)C(F)(F)F)/S1 (3vc4)	O=C1NC(=O)C(=Cc2cccc(C(F)(F)F)c2)S1	1.00
CCn1cnc2c1c(=O)[nH]c(=O)n2Cc1ccccc1 (6fnx)	CCn1cnc2c1c(P)nc(=O)n2Cc1ccccc1	0.905
CCNC(=O)C[C@@H]1N=C(c2ccc(cc2)Cl)c2c(-n3c1nnc3C)ccc(c2)OC (2yek)	CCN=NC1 = C2C=CC=C2C(OC)=CC1c1nc(C=CCl)c(C)n1C1 = CCC(=O)NC1	0.726
CC(=O)c1cc(c2n1cccc2)c1ccccn1 (4a9i)	CC(=O)c1cc(-c2ccccn2)c2ccccn12	1.00
SCc1ccc(cc1)C(=O)N1CCC(CC1)Cc1ccccc1 (3iw7)	O = C(c1ccc(CS)cc1)N1CCC(Cc2ccccc2)CC1	1.00
COc1cc2c(ncnc2cc1OC)Nc1cccc(c1)SC (1di9)	COc1cc2ncnc(Nc3cccc([SH]C)c3)c2cc1OC	1.00
CNC(=O)[C@H](Cc1ccccc1)NC(=O)[C@@H](CC(=O)NO)CC(C)C (1mnc)	CNC(=O)C(Cc1ccccc1)NC(=O)C(CC(=O)NO)CC(C)C	1.00
CCCCCN(C(=O)[C@@H](NC(=O)[C@H](Cc1ccccc1)NC(=O)C)CCC(=O)[O-])CCCCC (1a07)	CC=CC=CC1 = CC(=O)NC(C(=O)NC(CCC(=O)O)C(=O)N(CCCCC)CCCCC)C1	0.907

Benzene was used as starting compound in the atomistic-step MC approach and the dice similarity score was optimized in the MC search. The target SMILES correspond to bromodomain ligands extract from PDBbind (Liu *et al.* 2017).

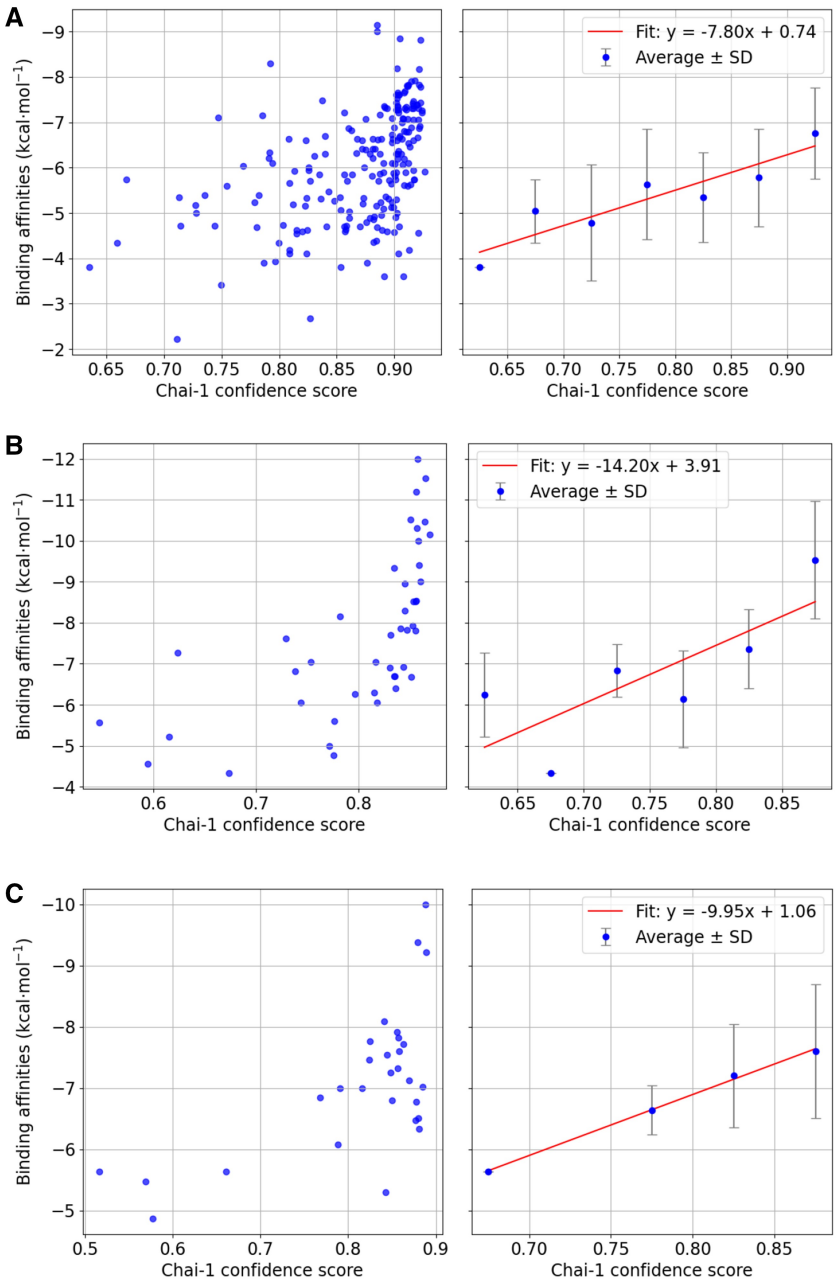

Figure 1 Illustration of the correlation between the experimental ligand-binding affinity and the Chai-1 confidence score. (A) Experimental binding affinity against the Chai-1 confidence score for different ligands of the bromodomain complex. (B) Experimental binding affinity against the Chai-1 confidence score for ligands of the serine/threonine-protein kinase pim-1 complex. (C) Binding affinity against the Chai-1 confidence score for ligands of the p38 map kinase complex.

appear to be much better than that of the Chai-1 score and when used in simulations the Chai-1 score generally produced more promising results than the auto dock score. Efforts to combine the Autodock and Chai-1 scores (as a sum of normalized Autodock score and Chai-1 score) did not result in much improved correlation (Fig. 7, available as supplementary data at *Bioinformatics* online). Hence, although not ideal, the Chai-1 confidence score (in combination with additional contributions, see "Materials and methods" section) was used for all following simulations.

De novo generation of bound ligands using atomistic-step MC simulation

The MC approach based on atomic changes of bound ligands was applied to the bromodomain, the serine/threonine–protein and the β-1 adrenergic receptor (a seven helix bundle GPCR protein). Each simulation consisted of 2000 MC steps, with a constant $\beta = 50$ and starting from benzene molecules. The probabilities for each type of changes in an MC step are the same as for the similarity-based simulation. For each protein

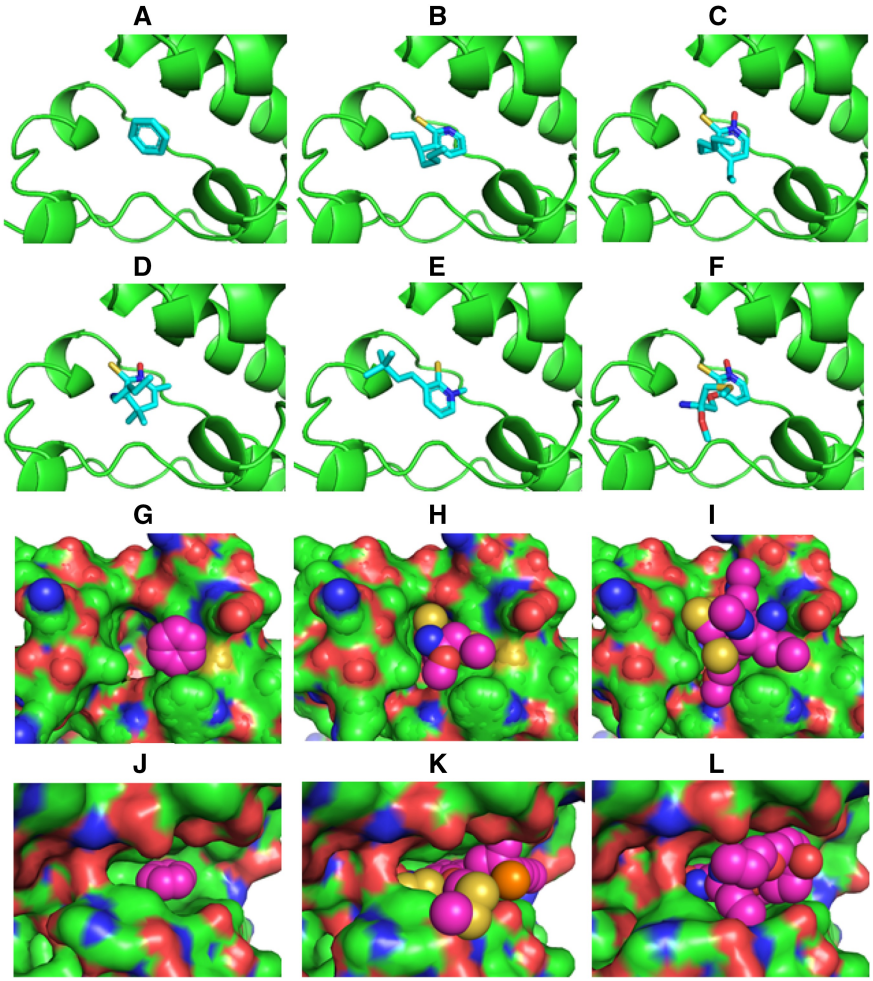

Figure 2 Illustration of the MC-based ligand design process. (A–F) Generated ligands (-stick model) after 0, 20, 40, 60, 80, and 100 MC steps for the bromodomain ligand binding pocket (protein as green cartoon). (G) Benzene (van der Waals spheres) start ligand placed in the bromodomain binding pocket. (H) same as (G) with the generated ligand after 200 MC steps. (I) same as (G) but after 2000 MC steps. Note that the generated ligand tightly fits into the binding pocket (protein represented as solvent accessible surface). (J–L) same as (G–I) but for the serine/threonine Pim-1 kinase target.

complex, we performed three MC simulations using different initial random seeds. Each MC run took roughly 20 h, slightly depending on protein size, on a PC with NVIDIA RTX4090 card.

In order to illustrate the progressive chemical changes of the ligand structure, every 20th accepted structure of the first 100 MC steps of the first simulation on the bromodomain complex is shown in Fig. 2. It also indicates how the MC approach progressively fills the protein pocket with sterically well-fitting ligands.

In each case, the finally generated compounds are bound at the target protein in the desired binding pocket and result in a high Chai-1 confidence score (Table 2). Since the MC-type simulation is a stochastic search method, it is not unexpected that the final compounds differ for each MC run. It is also compatible with the experimental observation that indeed many chemically distinct ligands can be found for the four target proteins that bind with nano- or micromolar affinity to the proteins. Encouragingly, the high final Chai-1 scores are also comparable to the scores for known ligands for each target that bind experimentally with highest affinity (compare Tables 2 and 3). As explained in the previous section, this should correspond, at least on average, to a very favorable binding affinity. During the MC run, the score improves initially within the first 250 steps quite rapidly and evolves further more gradually (illustrated in Fig. 3, see also Fig. 8, available as supplementary data at *Bioinformatics* online). After the initial phase, improvements that are simultaneously chemically feasible improve interactions and allow good packing without sterical overlap become increasingly difficult for the MC procedure. In the future, it might be possible to design smarter MC moves in chemical space to increase the probability of improvement at each MC step.

In the previous paragraph, we found only an average correlation of the Chai-1 score and the experimental ligand binding affinity for the test proteins. Hence, it is important to evaluate the generated compounds using alternative force field-based approaches. We used the widely employed molecular mechanics generalize born (MMGBSA) method (Miller *et al.* 2012) to estimate the mean interaction between compounds and target proteins. In an evaluation of a large number of protein–ligand complexes (Sindt *et al.* 2025), mean MMGBSA interaction

Table 2 Final generated compounds (SMILES code, some of the SMILES are split into two lines) resulting from the atomistic-step MC method for the bromodomain, p38 kinase, serine/threonine-protein kinase Pim-1, and the β-1 adrenergic receptor and the final scores.

Attempt		Score	Chai-1 score	SA score	ESOL score	QED score	MMGBSA score (kcal/mol)	Boltz-2 score (kcal/mol)
Results for bromodomain								
1	CCOC1(C)COc2cc(NC)n(OC)c (=S)c2N(C)C1	0.908	0.902	4.347	−2.954	0.857	−27.47±0.16	−8.484
2	C=CNCC1Cn2cc(C)c(C) c2C2 = C1c1c (nnc(C)c1CC)C2	0.919	0.917	4.405	−4.240	0.929	−29.79±0.10	−9.400
3	CCC1CCC(C)(N)N1CCCn1oc (=NC)n(C)c1 = O	0.918	0.908	4.798	−1.687	0.844	−19.19±0.13	−7.763
Results for p38 map kinase								
1	CC(CCN)c1noc2nnc(O)c- 2c1CCC1(C)C=CC=N1	0.885	0.874	4.623	−2.986	0.844	−40.20±0.14	−7.947
2	CCNC(CC(C)=C1C2 = C(CCC2)C (N)=C1 C(=S)COCCCCN)(ON(O)N (CN)CCO)C(=N)N	0.846	0.863	4.904	−2.330	0.030	−41.73±0.29	−7.790
3	CCCCC1CCC(F)(c2cnc(N)c(C(=O) O c3nnc(F)o3)c2N)O1	0.877	0.872	4.578	−3.684	0.720	−34.01±0.28	−9.388
Results for Pim-1 kinase								
1	CCOc1c2c(c(C)c(C)c1PCN)CN (CC(=O)N1C=NC (C)(N)C1 = O)C2	0.860	0.843	4.698	−2.311	0.659	−34.23±0.21	−7.422
2	C=C1C(C)=c2ccc3c (c2 = C1C1CCCC N1C)OC (CCN)N3O	0.876	0.857	4.490	−2.771	0.868	−32.54±0.14	−8.750
3	CN=c1ccop1CC(C)=c1c(O)c (CCCO) c(C)c(F)c1 = CCOC	0.854	0.837	4.956	−3.5001	0.754	−31.55±0.15	−7.151
Results for β-1 receptor								
1	C=CC(C)C1 = C(N)N(C)C = C(C(N) CCC(C)C)C1	0.849	0.838	4.490	−2.943	0.724	−32.16±0.27	−10.226
2	CCC(OC)C1 = C(C = C(CCN(C)C)C (C)O)NC(C)=C (C)C1N	0.851	0.844	4.714	−2.758	0.633	−47.61±0.52	−8.007
3	COC(C)Cc1c2c(c(C)n1C)CC(N)C (CC(N) C1CCC(C)C1)C2	0.875	0.865	4.871	−3.785	0.818	−56.57±0.19	−8.780

energies of −25 to −45 kcal/mol were typically found for high-affinity binders. The MMGBSA results for the three runs and for each protein target case are in this range (Table 2) and encouragingly are also of similar magnitude as MMGBSA results for experimentally evaluated protein binders (Table 3). Note also that the MMGBSA scores for the four test proteins and different ligands correlate qualitatively with experimentally known binding affinity (Table 3).

We also recalculated the MMGBSA score for the first 200 MC steps of one simulation for each target (Fig. 9, available as supplementary data at *Bioinformatics* online) and the calculated averages of the three runs on each protein (Fig. 3). Similar to the Chai-1 score, it improves initially quite rapidly but subsequently only gradually.

As another independent test, the generated compounds were also evaluated with the recent ligand–receptor binding affinity model of Boltz-2 (Passaro *et al.* 2025). This model returns a log (IC50) score, which estimates the binding affinity of ligand–receptor complexes and the accuracy of the model is considered to be similar to alchemical absolute binding free energies that are considered as most accurate binding prediction method (Passaro *et al.* 2025). The log(IC50) score was converted into a

Table 3 Experimental binding affinities, MMGBSA score, and Boltz-2 score for ligands of the bromodomain, the serine/threonine-protein kinase Pim-1 protein, the p38 map kinase protein, and the β-1 adrenergic receptor.

Complex	Protein	Ligand SMILES	Exp. binding affinity (kcal/mol)	MMGBSA score (kcal/mol)	Boltz-2 score (kcal/mol)
5mli	Bromodomain	CNc1cnn(c(=O)c1Cl)C	−3.90	−20.71±0.13	−7.930
4o7a	Bromodomain	Clc1cccc(c1)C1=C(Nc2ccc(c(c2)Cl)O)C(=O)NC1=O	−4.72	−23.11±0.13	−9.521
4hbw	Bromodomain	CCNS(=O)(=O)c1ccc2c(c1)CN(C(=O)N2)C	−5.32	−25.23±0.10	−8.97
5igk	Bromodomain	CCOC(=O)Nc1cc(nn2c1nnc2C)c1ccc(c(c1)NS(=O)(=O)C)C	−7.38	−33.77±0.12	−9.682
5khm	Bromodomain	COc1nnc2n1nc(cc2)N1CCC(CC1)c1ccc(cc1)OCC[N@H+]1CCN(C(=O)[C@H]1C)C	−8.3	−38.68±0.18	−9.428
3iw8	p38 kinase	O=C(Nc1scc(n1)[C@@H](COCc1ccccc1)[NH3+])Nc1ccc(c(c1)Cl)F	−4.87	−41.17±0.29	−7.812
3iw7	p38 kinase	SCc1ccc(cc1)C(=O)N1CCC(CC1)Cc1ccccc1	−5.64	−35.39±0.14	−6.701
6m95	p38 kinase	COc1cc(SC)ccc1C(=O)N1CCC(CC1)Cc1ccccc1	−6.85	−37.48±0.17	−7.943
3l8s	p38 kinase	O=C(Nc1snc(c1C(=O)N)OCc1c(F)cc(cc1F)Br)NCCCC[NH+]1CCCC1	−7.00	−44.22±0.18	−10.531
6ohd	p38 kinase	Cc1ccc(cc1c1cnc2c(c1)[nH]c(=O)n2C(C)(C)C)C(=O)Nc1nocc1	−9.38	−50.18±0.15	−10.862
4xh6	pim-1 kinase	COc1c(O)cc2c(c1O)c(=O)cc(o2)c1ccc(cc1)O	−5.57	−30.22±0.14	−7.502
5kcx	pim-1 kinase	C[NH+]1CCN(CC1)c1ccc(cc1)c1cc2c(n1C)nc(cc2Cl)C(=O)N	−6.70	−31.24±0.22	−9.378
4xhk	pim-1 kinase	O=C(c1csc(n1)c1c(F)cccc1F)Nc1cnccc1O[C@@H]1C[NH2+]CC1	−8.30	−34.81±0.11	−10.540
5v82	pim-1 kinase	Cc1cncc(n1)c1ncc2c(c1)n(nc2)c1cccc(n1)[C@H]1C[NH2+]CCC1(F)F	−10.15	−38.31±0.10	−12.372
4n70	pim-1 kinase	C[C@H]1CN(C[C@H]([C@@H]1O)[NH3+])c1cc[nH+]cc1NC(=O)c1ccc(c(n1)c1c(F)cccc1F)F	−12.00	−46.36±0.23	−12.872
3zpr	β-1 receptor	Cc1cc(nc2c1cccc2)N1CC[NH2+]CC1	−6.65	−34.36±0.17	−8.461
3zpq	β-1 receptor	C1[NH2+]CCN(C1)c1cccc2c1cc[nH]2	−7.17	−24.02±0.22	−9.795

pIC50 in kcal/mol. In addition, the Boltz-2 score assigns quite favorable predicted binding scores to the MC-generated compounds (Table 2).

One important advantage of the present MC-based ligand generation method is the full reconstruction of the protein partner allowing for conformational-induced fit adaptation of the protein. The variation of the protein structure during MC-based ligand generation depends significantly on the protein target (Fig. 3, see also Fig. 10, available as supplementary data at *Bioinformatics* online). No major structural changes were observed for the bromodomain protein. It indicates a relatively rigid binding pocket which varies little in structure upon binding different ligands. Indeed, this finding compares well with experimental structures of the bromodomain that change only little in complexes with different ligands (Fig. 11, available as supplementary data at *Bioinformatics* online). It is important to note that even small changes in conformation can be of critical importance for correctly placing and evaluation of protein–ligand complexes (a major problem of rigid protein–ligand docking approaches).

In contrast, the structural variations during the MC simulations are larger for the serin kinase and P38 cases (Fig. 3). This is also in line with the experimentally found larger structural variation in complex with different ligands (Fig. 11, available as supplementary data at *Bioinformatics* online).

Furthermore, it was investigated whether the ligands from the different MC runs bind to the protein in a similar fashion. To this end, the complexes of all the simulations were evaluated with LigPlot+ (Laskowski and Swindells 2011) (Fig. 12, available as supplementary data at *Bioinformatics* online). For the bromodomain protein, the results show high similarity, all ligands form similar sets of hydrogen bonds with protein residues and similar contacts to surrounding protein chains. While no chemical similarity was found, there is some overlap in the way the ligands bind to the proteins, as shown in Fig. 13, available as supplementary data at *Bioinformatics* online. To further investigate the similarities in the distribution of known and generated ligands, an UMAP (McInnes *et al.* 2018) analysis was employed through the ChemPlot implementation (Cihan Sorkun *et al.* 2022). The approach allows for comparison of compounds in a principal component type of chemical space representation. As shown in Fig. 14, available as supplementary data at *Bioinformatics* online, the distribution of the ligands is very diverse and, encouragingly, the ligands from the simulation are in all cases close to some clusters representing known ligands.

Finally, it was attempted to improve the SA, ESOL, and QED scores further, by adding an additional cleanup phase to the simulation. This consists of 100 simulation steps, with the score consisting only of the SA, ESOL, and QED scores. However, under the condition that the Chai-1 score never falls below a certain level ($>$ 0.02), the threshold was determined as the starting Chai-1 score minus 0.02. The probabilities for the types of steps were also adjusted. This cleanup phase was tested for the best resulting structure found for each protein. The results are shown

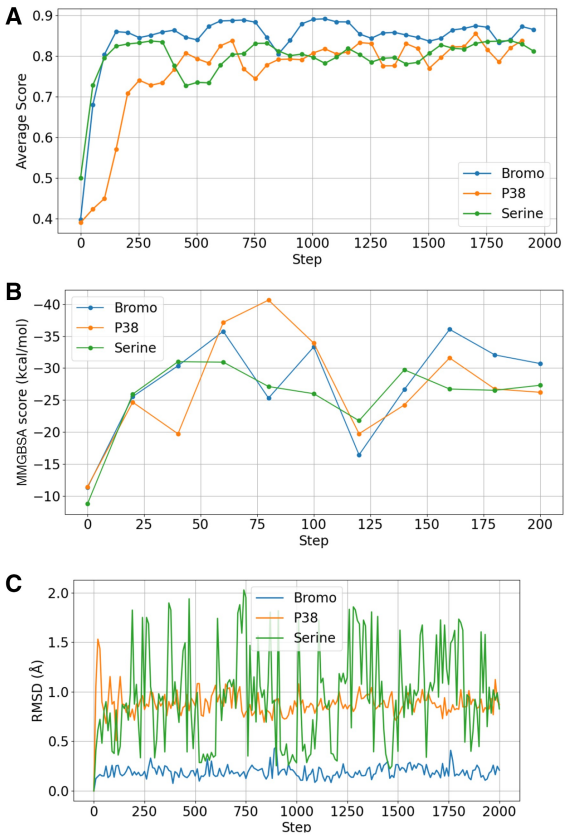

Figure 3 (A) The average of the score for every 50 steps during the basic MC simulation for one design run on each of the three protein targets. The average was taken as running windows over 100 MC steps. (B) Calculated MMGBSA score of three MC simulation runs for the first 200 MC steps. (C) Root-mean-square deviation (RMSD) of the nonhydrogen protein atoms forming the ligand binding pocket with respect to the start structure. The pocket protein atoms are defined as all atoms within 5 Å of the native ligand.

in Table 4, available as supplementary data at Bioinformatics online. The SA, ESOL, and QED scores improve successfully, without incurring a major drawback in terms of the primary Chai-1 score. It is still important to note that even with these improvements, the SA score is only an estimate and does not guarantee that resulting structures can be synthesized with reasonable effort. Further work may be required to ensure the structures resulting from the simulation are reasonable.

Fragment-based MC simulations

Instead of atomistic-step MC method, the MC search based on larger chemical fragments was also investigated. In a first attempt, some known ligands for the two proteins with high scores were selected. These ligands were broken down into fragments and recombined within a simulation of 500 steps, with $\beta = 5$. Hence, the expected outcome for this type of stochastic MC-based fragment assembly is to recover the known ligand (Table 5, available as supplementary data at *Bioinformatics* online). Interestingly, for all cases, compounds very similar to but not always exactly the same as the known ligands were recovered (Table 5, available as supplementary data at *Bioinformatics* online). It reflects also the stochastic character of the compound generation.

Additionally, the MC simulation approach was tested on a set of 50 randomly selected fragments from the ChEMBL database (Zdrazil *et al.* 2023) and fractured into BRICS fragments. These fragments were then used for an MC simulation of 1000 steps (Table 6, available as supplementary data at *Bioinformatics* online). Interestingly, the resulting scores are not as favorable as for the atomistic-step MC simulations. This is most likely due to the fact that the fragments were chosen randomly and may not contain the most complementary chemical groups in a sterically well- fitting fragment. In the future, it might be possible to design smaller, more elementary fragments that allow for a greater variety of combining chemical groups in different spatial arrangements.

However, it was also tested if fragment selections specifically suited for the selected target protein by decomposing known ligands result in improved performance. These fragments were then used in the same simulation setup as used for MC searches based on random fragments. Indeed, on average, this type of MC simulation resulted in significantly improved scores compared to using randomly selected fragments (Table 7, available as supplementary data at *Bioinformatics* online). The generated compounds indicate MMGBSA scores and Boltz-2 scores in the same range as known experimental binders to the target proteins. The course of the average MC confidence score and of the reevaluation using MMGBSA was also investigated (Fig. 15, available as supplementary data at *Bioinformatics* online). In both the MC simulations with random and specific fragments, the scores improved rapidly at the beginning to reach MC scores

around 0.7–0.8 and after 200 steps changed only gradually with significant fluctuations due to the relatively small $\beta = 5$ but also due the bigger changes caused by adding or removing whole fragments. The course of the MMGBSA score indicates similar trends (Fig. 15, available as supplementary data at *Bioinformatics* online).

Conclusion

Recent advances in AI allow for the rapid and accurate structure prediction of proteins and also complexes with organic ligands (Jumper *et al.* 2021, Chai Discovery 2024, Wohlwend *et al.* 2024). It is also possible to use similar generative AI-based methods to construct entirely new ligands employing, however, mostly rigid protein pockets as target input structures (Śledź and Caflisch 2018, Liu *et al.* 2020, Wang *et al.* 2022, Tang *et al.* 2024b). Based on the rapid Chai-1 structure prediction approach, we have developed an MC-based approach, termed AI-MCLig, that searches stochastically through chemical space to generate new potential binders for a given target protein structure. It allows also for control of synthesizability of compounds which will by subject of future efforts of improvement. The explicit rebuilding of the entire complex structure at every MC step allows for adaptation (induced fit) of the entire protein structure in response to an altered compound structure (and also of the ligand itself). This is a distinct advantage of our approach compared to existing construction methods that mostly construct ligands in a rigid protein environment (Liu *et al.* 2020).

For the targets in our study, we found that only the average confidence score of Chai-1 correlates with experimentally measured ligand binding affinity. The limited accuracy of the score used to rapidly evaluate generated ligands clearly limits the applicability of the MC approach. However, this is an issue that affects all available docking and generative ligand design methods and we therefore consider this as a separate issue, not disfavoring our approach. This aspect clearly needs future efforts for improvement. Nevertheless, our approach allowed in all cases to rapidly suggest new potential binders with MMGBSA scores and Boltz-2 scores that are in the same range as those calculated for experimentally known ligands suggesting that these generated ligands potentially represent realistic binders.

We consider our MC-based compound generation approach as complementary to the variety of fragment buildup and new reverse-diffusion-based approaches. The method can be quickly adapted to specific needs. For example, one can limit the search just to modify side chains or subsets of positions of a known ligand and keep a desired core or reference structure unmodified. The method can be especially useful to rapidly generate a large variety of potential binders for a given target protein structure.

Acknowledgements

The authors thank Niklas Halbwedl and Patrick Quoika for useful discussions.

Author contributions

Jakob Agamia (Data curation, Investigation, Software [lead], Formal analysis, Methodology, Writing—original draft, Writing—review & editing [equal]) and Martin Zacharias (Conceptualization, Funding acquisition, Project administration, Resources, Supervision [lead], Investigation, Methodology, Writing—original draft, Writing—review & editing [equal])

Supplementary material

Supplementary material is available at *Bioinformatics* online.

Conflict of interests

No competing interest is declared.

Funding

This work was financially supported by the Deutsche Forschungsgemeinschaft (grant DFG Za153/29-1) and by local compute cluster (partially funded by DFG INST 95/1610-1 FUGG). The authors acknowledge additional HPC resources provided by the Erlangen National High Performance Center (NHR@FAU, grant b118bb).

Data availability

Datasets, examples, and source code are available on our public GitHub repository https://github.com/JakobAgamia/AI-MCLig and on Zenodo at https://doi.org/10.5281/zenodo.17800140.

References

Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* 2024;**630**:493–500.

Beier C, Zacharias M. Tackling the challenges posed by target flexibility in drug design. *Expert Opin Drug Discov* 2010; **5**:347–59.

Bickerton G, Paolini G, Besnard J *et al.* Quantifying the chemical beauty of drugs. *Nat Chem* 2012;**4**:90–8.

Carhart RE, Smith DH, Venkataraghavan R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J Chem Inf Comput Sci* 1985;**25**:64–73.

Case DA, Aktulga HM, Belfon K *et al.* Ambertools. *J Chem Inf Model* 2023;**63**:6183–91.

Case DA, Aktulga HM, Belfon K *et al. Amber 2025*. San Francisco: University of California, 2025.

Chai Discovery team. Chai-1: decoding the molecular interactions of life. bioRxiv, https://doi.org/10.1101/2024.10.10.615955, 2024, preprint: not peer reviewed.

Cihan Sorkun M, Mullaj D, Koelman JMVA *et al.* Chemplot, a python library for chemical space visualization. *Chem Methods* 2022;**2**:e202200005.

Corrêa Veríssimo G, Salgado Ferreira R, Gonçalves Maltarollo V. Ultra-large virtual screening: definition, recent advances, and challenges in drug design. *Mol Inform* 2025; **44**:e202400305.

Degen J, Wegscheid C, Gerlach A *et al.* On the art of compiling and using 'drug-like' chemical fragment spaces. *ChemMedChem* 2008;**3**:1503–7.

Delaney JS. ESOL: estimating aqueous solubility directly from molecular structure. *J Chem Inf Comput Sci* 2004;**44**:1000–5.

Eberhardt J, Santos-Martins D, Tillack AF *et al.* Autodock vina 1.2.0: new docking methods, expanded force field, and python bindings. *J Chem Inf Model* 2021;**61**:3891–8.

Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminform* 2009;**1**:8–309.

Goetz AW, Williamson MJW, Xu DX *et al.* Routine microsecond molecular dynamics simulations with AMBER—part I: generalized born. *J Chem Theo Comput* 2012;**8**:1542–55.

Gowers R, Linke M, Barnoud J *et al.* Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations. In: *Proceedings of the 15th Python in Science Conference, SciPy*. Tacoma, WA, USA: Tacoma Convention Center, 2016, 98–105.

Grygorenko OO, Radchenko DS, Dziuba I *et al.* Generating multibillion chemical space of readily accessible screening compounds. *iScience* 2020;**23**:101873.

Horvath C. Comparison of preclinical development programs for small molecules (drugs/pharmaceuticals) and large molecules (biologics/biopharmaceuticals): studies, timing, materials, and costs. In: Gad SC (ed.), *Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing.* Wiley Press, 2010, 1–35

Jorgensen WL, Chandrasekhar J, Madura JD *et al.* Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;**79**:926–35.

Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with alphafold. *Nature* 2021;**596**:583–9.

Landrum G. Rdkit: open-source cheminformatics software. https://www.rdkit.org, https://doi.org/10.5281/zenodo.10633624 (28 January 2026, date last accessed).

Laskowski RA, Swindells MB. Ligplot+: multiple ligand–protein interaction diagrams for drug discovery. *J Chem Inf Model* 2011;**51**:2778–86.

Le Grand S, Götz AW, Walker RC. SPFP: speed without compromise - a mixed precision model for GPU accelerated molecular dynamics simulations. *Comp Phys Comm* 2013;**184**:374–80.

Liu P, Agrafiotis DK, Theobald DL. Fast determination of the optimal rotational matrix for macromolecular superpositions. *J Comput Chem* 2010;**31**:1561–3.

Liu X, IJzerman AP, van Westen GJ. Computational approaches for de novo drug design: past, present, and future. *Methods Mol Biol* 2021;**2190**:139–65.

Liu Z, Su M, Han L *et al.* Forging the basis for developing protein-ligand interaction scoring functions. *Acc Chem Res* 2017; **50**:302–9.

Lyu J, Irwin JJ, Shoichet BK. Modeling the expansion of virtual screening libraries. *Nat Chem Biol* 2023;**19**:712–8.

Healy J, McInnes L. Uniform manifold approximation and projection. *Nat Rev Methods Primers* 2024;**4**:82. https://doi.org/10.1038/s43586-024-00363-x

Martino E, Chiarugi S, Margheriti F *et al.* Mapping, structure and modulation of PPI. *Front Chem* 2021;**9**:718405.

Michaud-Agrawal N, Denning EJ, Woolf TB *et al.* Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* 2011;**32**:2319–27.

Miller BRI, McGee TDJ, Swails JM *et al.* Mmpbsa.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput* 2012;**8**:3314–21.

Mouchlis V, Afantitis A, Serra A *et al.* Advances in de novo drug design: from conventional to machine learning methods. *Int J Mol Sci* 2021;**22**:1676.

Passaro S, Corso G, Wohlwend J *et al.* Boltz-2: towards accurate and efficient binding affinity prediction. 2025. https://github.com/jwohlwend/boltz (28 January 2026, date last accessed).

Salomon-Ferrer RS-F, Goetz AW, Poole D *et al.* Routine microsecond molecular dynamics simulations with AMBER—part II: particle mesh Ewald. *J Chem Theory Comput* 2013;**9**:3878–88.

Schneider P, Walters WP, Plowright AT *et al.* Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov* 2020; **19**:353–64.

Sindt F, Bret G, Rognan D. On the difficulty to rescore hits from ultralarge docking screens. *J Chem Inf Model* 2025; **65**:5553–66.

Śledź P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol* 2018; **48**:93–102.

Tang X, Dai H, Knight E *et al.* A survey of generative ai for de novo drug design: new frontiers in molecule and protein generation. *Brief Bioinfo* 2024a;**25**:bbae338.

Tang Y, Moretti R, Meiler J. Recent advances in automated structure-based de novo drug design. *J Chem Inf Model* 2024b; **64**:1794–805.

Theobald DL. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Crystallogr A* 2005; **61**:478–80.

Trott O, Olson AJ. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;**31**:455–61.

Van Der Walt S, Colbert SC, Varoquaux G. The numpy array: a structure for efficient numerical computation. *Comput Sci Eng* 2011;**13**:22–30.

Wang J, Wang W, Kollman P *et al.* Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 2006a;**25**:247–60.

Wang J, Wolf R, Caldwell J *et al.* Development and testing of a general amber force field. *J Comput Chem* 2006b;**25**:1157–74.

Wang M, Wang Z, Sun H *et al.* Deep learning approaches for de novo drug design: an overview. *Curr Opin Struct Biol* 2022; **72**:135–44.

Weininger D. Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;**28**:31–6.

Wohlwend J, Corso G, Passaro S *et al.* Boltz-1: democratizing biomolecular interaction modeling. bioRxiv, 2024.

Xie W, Wang F, Li Y *et al.* Advances and challenges in de novo drug design using three-dimensional deep generative models. *J Chem Inf Model* 2022;**62**:2269–79.

Zahiri J, Emamjomeh A, Bagheri S *et al.* Protein complex prediction: a survey. *Genomics* 2020;**112**:174–83.

Zdrazil B, Felix E, Hunter F *et al.* The ChEMBL database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucl Acids Res* 2023; **52**:1180–92.

Bioinformatics, 2026, 42, 1–12
https://doi.org/10.1093/bioinformatics/btag027
Original Paper