METHODOLOGY

Preventing lipophilic aggregation in cosolvent molecular dynamics simulations with hydrophobic probes using Plumed Automatic Restraining Tool (PART)

Olivier Beyens[1] and Hans De Winter[1*]

Abstract

Cosolvent molecular dynamics (MD) simulations are molecular dynamics simulations used to identify preferable locations of small organic fragments on a protein target. Most cosolvent molecular dynamics workflows make use of only water-soluble fragments, as hydrophobic fragments would cause lipophilic aggregation. To date the two approaches that allow usage of hydrophobic cosolvent molecules are to use a low (0.2 M) concentration of hydrophobic probes, with the disadvantage of a lower sampling speed, or to use force field modifications, with the disadvantage of a difficult and inflexible setup procedure. Here we present a third alternative, that does not suffer from low sampling speed nor from cumbersome preparation procedures. We have built an easy-to-use open source command line tool PART (Plumed Automatic Restraining Tool) to generate a PLUMED file handling all intermolecular restraints to prevent lipophilic aggregation. We have compared restrained and unrestrained cosolvent MD simulations, showing that restraints are necessary to prevent lipophilic aggregation at hydrophobic probe concentrations of 0.5 M. Furthermore, we benchmarked PART generated restraints on a test set of four proteins (Factor-Xa, HIV protease, P38 MAP kinase and RNase A), showing that cosolvent MD with PART generated restraints qualitatively reproduces binding features of cocrystallised ligands.

Keywords Molecular dynamics, Cosolvent simulations, Metadynamics, Conformational sampling, Drug design

Introduction

Cosolvent molecular dynamics have recently gained interest as a tool for structure-based drug design using a fragment-based approach. Cosolvent molecular dynamics simulations are molecular dynamics (MD) simulations in which a user-specified amount of the water solvent molecules surrounding a protein target are replaced by so-called cosolvent molecules, where these cosolvent molecules are also termed probes. The positions of the cosolvent molecules throughout the trajectory are then analyzed a posteriori to create density maps of where the probes most often reside during the molecular dynamics simulations. A variety of probe types and mixes of probe types have been used throughout literature, such as water-benzene-isopropane, water-isopropanol or water-isopropanol-acetamide-acetate-isopropylamine [1–5]. A more comprehensive review of the different type of probe mixtures used can be found in the miniperspective written by Ghanakota and Carlson [6]. Recently, a variety of enhanced sampling techniques have also been used in combination with cosolvent molecular dynamics to further enhance performance. In SWISH, Hamiltonian

*Correspondence:
Hans De Winter
hans.dewinter@uantwerpen.be
[1] Laboratory of Medicinal Chemistry, Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium

replica exchange is combined with a benzene cosolvent, leading to a methodology that is promising for discovering cryptic pockets [7]. In accelerated ligand-mapping MD, accelerating molecular dynamics is combined with benzene cosolvent, leading to a methodology in which occluded binding pockets can be discovered as well [8].

An interesting application to structure-based drug design consists of building pharmacophore models from the calculated probe densities. Pharmacophore models are three-dimensional maps of chemical features thar are considered important for ligand binding [9]. By screening a library of molecules for that three-dimensional arrangement of chemical features, novel inhibitor scaffolds for protein targets can be discovered. These cosolvent-based pharmacophore models outperform more conventional docking approaches, as was shown by the SILCS-Pharm approach [10]. Another workflow that combines cosolvent molecular dynamics with pharmacophore modeling is the Pharmmaker tool [11].

As hydrophobic and aromatic groups are often included in pharmacophore screenings, it is of interest to include these hydrophobic probes in cosolvent molecular dynamics [12]. However, a major concern in simulations with hydrophobic probe molecules, such as benzene, isopropane or isobutane, is that phase separation can occur by aggregation of these hydrophobic probes. Bakan et al. [3] mentioned that isobutane is one the most common fragment types in approved drugs, however isobutane was not included in the probe mix due to the insolubility of isobutane and risk for aggregation. The Site Identification by Ligand Competitive Saturation (SILCS) approach by Guvench et al. [2, 10, 13–18] provides a working solution to this aggregation risk by including an artificial repulsion term between the hydrophobic probes. This artificial repulsion term is incorporated by adapting the corresponding non-bonded force field parameters. However, modifying force fields is a rather complex procedure, and additionally force field cut-offs need to be adapted from the standards for which the force field was parametrized [19]. Another possibility to avoid hydrophobic probe aggregation is to use simulations with lower probe concentrations (<0.2 M) [8, 20]. However, for applications where faster convergence in probe density maps is desired, the low sampling speed stemming from the low concentrations might be detrimental. Consequently, there is a need for a flexible and easy to setup methodology that generates intermolecular repulsion interactions between hydrophobic probes in molecular dynamics simulations, and in this context we developed a workflow using PLUMED-based restraints [21–23]. In this paper we present the "Plumed Automatic Restraining Tool" (PART), a Python command line tool that prepares a PLUMED input file to integrate in cosolvent molecular

dynamics workflows with hydrophobic probes without aggregation issues. In this manuscript, we detail the PART methodology, show a benchmark of a restrained versus unrestrained simulation and we demonstrate that cosolvent MD simulations with restraints generated by PART may reproduce key ligand features. The PART program, together with some usage examples is freely available on GitHub at the following link: https://github.com/UAMC-Olivier/PART.

Methods

The methodology of the tool development consists of four parts, namely an explanation of the tool code, a comparison between unrestrained and restrained simulations to show that artificial repulsion is necessary at high hydrophobic probe concentration, a benchmark to show that known ligand features are reproduced by cosolvent simulations with intermolecular repulsion terms generated by PART, and lastly a benchmark on the slowdown of the simulation by including PART restraints.

Plumed Automatic Restraining Tool (PART)

The Plumed Automatic Restraining Tool is a Python command line tool that generates a PLUMED input file to prevent aggregation of user-defined probe molecules using the existing "many restraints" module of PLUMED [21–23]. The general methodology consists of calculating all distances between the defined molecules, and then calculating the to-be-applied artificial repulsion using a bias function, which is user customizable if desired. The input consists of a Protein Data Bank (PDB) or GROMACS (GRO) structure file of the system with cosolvent molecules and a description of the molecule types to which the restraining potential needs to be applied. It is possible to add a restraining potential between molecules of the same type (such as benzene-benzene), or between molecules of different types (such as benzene-propane), or any combination of these. The calculation of the intermolecular distances is point-based. The user can either specify a central atom of the probe to be used as the point on which the distance calculation is based, or specify a list of atoms from which the center of mass (COM) will be computed. PART can easily be integrated in a cosolvent molecular dynamics workflow as depicted in Fig. 1.

PART will first parse the user-specified PDB or GRO file using a custom structure file parser implemented in the PART code. PART will then build all necessary atom groups and write the requested center of mass calculation to the PLUMED file. As a final step, the tool will write out all distance calculation statements, all restraint statements, and print statements of the bias values. The restraints statements make use of lower walls, of which the limit and harmonic potential terms are

Fig. 1 Schematic visualization of the inclusion of PART in a cosolvent molecular dynamics workflow

user-customizable if desired using optional command line flags. The default parameter values for the artificial repulsion potential V are an energy constant k of approximately 0.5 kcal mol^{-1} Å^{-4}, a wall location a of 8.0 Å, a scaling factor s of 1.0 and an exponent e of 4.0 with x being the intermolecular distance in Å (Eq. 1):

$$V = k\left(\frac{x-a}{s}\right)^e \ if \ x < a, V = 0 \ if \ x \geq a \qquad (1)$$

After running PART, the PLUMED file can be further modified by the user to include additional restraints, such as protein root-mean-square deviation restraints, if desired.

Comparison between restrained and unrestrained cosolvent MD simulations

The Factor Xa protein (PDB: 1FJS) [24] structure was prepared by removing ligands and glycerol molecules, while retaining crystal waters and ions. Side chain flips were analyzed with Reduce [25] and protonation states of the ionizable groups with PROPKA [26, 27]. Occurrence of disulfide bridges was checked manually. Molecular dynamics engine of choice was GROMACS 2021.3 [28, 29] patched with PLUMED 2.7.2 [21–23], the chosen force field was CHARMM36m (July 2021 version) [19, 30]. Consequently, the prepared PDB file of the protein was converted to GROMACS format, and the system was placed in a periodic dodecahedral box, where the box edges are at least 12 Å away from any protein element. A workflow tool involving the PART tool was used for adding the cosolvent molecules, counterions and original TIP3P [31] water molecules at ten different randomly generated starting locations. The chosen cosolvent molecules were the same as in the SILCS multiple fragment types methodology [16], namely benzene, propane,

methanol, formamide, acetaldehyde, acetate and methylammonium. A PART generated PLUMED file with the intermolecular repulsion terms between benzene-benzene, propane-benzene, propane-propane and acetate-methylammonium was also generated at this stage. Force field parameters for the cosolvent molecules were generated using CGenFF (CGenFF version 4.6, CGenFF program version 2.5) [32–35]. The concentration of each cosolvent molecule type was set to 0.25 M.

After system preparation, the system was minimized in two stages: a first stage with flexible waters and without hydrogen bond constraints, followed by a second stage with rigid waters and with hydrogen bond constraints. Throughout the second minimization stage and all subsequent simulations, hydrogen bonds were constrained using the LINCS algorithm [36]. Minimization was set to steepest descent with a maximum number of 50,000 steps and the default convergence criterion. Minimization step size was 0.01 Å for the first stage and 0.1 Å for the second. The system was then initialized at a temperature of 300 K equilibrated in three stages: one in the NVT ensemble with positional restraints on the protein heavy atoms, one in the NPT ensemble, again with positional restraints, and one in the NPT ensemble without positional restraints. The force constant for positional restraints was approximately 2.4 kcal mol^{-1} Å^{-2}. The NVT equilibration was carried out over 0.5 ns, while both NPT stages of the equilibration were simulated for 1 ns.

The actual production runs in the NPT ensemble were simulated for 100 ns for each of the ten different randomly generated starting coordinates, leading to a total simulation length of 1 μs. Throughout all simulations, force field cut offs of 12 Å were used, where the potential is smoothly switched to zero between 10 Å and 12 Å. Long range electrostatics were treated using

Table 1 Overview of the analyzed ligands in the ligand feature reproduction benchmark

Protein system	PDB codes of analyzed ligands
Factor-Xa	1FJS [24], 1EZQ [46], 1MQ5 [47], 1Z6E [48]
HIV protease	1G2K [40], 1DMP [49], 1B6K [50], 1D4L [51]
P38 MAP kinase	1OUY [41], 1W84 [52], 1A9U [53], 1BL7 [53], 1DI9 [54], 1WBW [52]
RNase A	1O0H [55], 1O0O [55], 1O0M [54], 1QHC [56], 6PVV* [43], 6PVX* [43]

*Not included in the original SILCS [16] benchmark set

Particle Mesh Ewald (PME) [37] method. Temperature coupling in all simulations was carried out using the V-rescale thermostat [38] with a time constant of 0.1 ps, with the reference temperature set to 300 K. Protein and non-protein elements were coupled separately. Pressure coupling was executed by the C-rescale barostat [39] with a time constant of 2.0 ps and with a reference pressure of 1 bar in all NPT ensemble simulations.

The above simulation protocol was applied twice to Factor Xa, once with the PART generated PLUMED restraints and once without. For the simulation with PART restraints, the PART default potential was applied for both NPT equilibrations and for the production simulations. For the NVT equilibration, a softer potential (parameters: $k \approx 0.02$ kcal mol-1 Å^{-2}, $a = 8.0$ Å, $s = 1.0$, $e = 2.0$) was applied as to not introduce large forces to cosolvent molecules close to each other.

During simulation analysis, radial distribution functions (RDFs) were calculated by binning the individual intermolecular distances over the MD trajectories of the ten simulation replicas. Chosen distance calculation points were the COM of benzene, the central carbon atom of propane, the nitrogen atom of methylammonium, the carboxylic carbon atom of acetate and the carbon atom of formamide.

We also analyzed the percentage of ionic organic fragments participating in ion-ion interactions and the percentage of aggregated lipophilic molecules. To examine this, we computed the distances between the molecules of interest by calculating the intermolecular distance matrices. For each MD frame, the number of molecules participating in interactions was then calculated based on these intermolecular distance matrices. The criterium for ionic interactions was a distance below 4 Å. The criterium for a hydrophobic molecule to participate in lipophilic aggregation was modelled as one intermolecular distance to another hydrophobic molecule below 7 Å. Finally, the average fraction of molecules participating in interactions was calculated as the average of interacting molecules over all the MD frames from the ten replicas. The distance calculations were performed using PLUMED.

### Ligand feature reproduction benchmarks

We benchmarked the PLUMED restraints generated by PART on the results found in a SILCS benchmark by Raman et al. on four protein systems, namely Factor-Xa (PDB:1FJS) [24], HIV protease (PDB:1G2K) [40], P38 MAP kinase (PDB:1OUY) [41] and RNase A (PDB:1JVT) [42]. An identical cosolvent mixture as in the previously detailed restrained versus unrestrained MD simulations was used. Raman et al. [16] compared

SILCS fragment affinity maps to important binding features of cocrystallised ligands to the four benchmark proteins, of which we analyzed a selection of ligands as well. Additionally, we analyzed whether our simulations can reproduce key ligand features of a set of recently published crystal structures of RNase A [43]. An overview of the analyzed ligand–protein structures can be found in Table 1. Protein preparation and simulation protocols for the three extra proteins were the same as for the previously detailed restrained Factor Xa benchmark, with one exception: for the RNase A system, an additional restraint on the Cα carbons with a force constant of approximately 0.01 kcal mol^{-1} Å^{-2} was added during the second NPT equilibration and the MD production to prevent protein unfolding. Visualization was performed using PyMol (Version 2.4.1, Schrödinger LLC) and density analyses were performed using MDAnalysis [44, 45]. To allow comparison with previous executions of this benchmark [16], we calculated the densities of five atom groups: hydrophobic (propane and benzene carbon atoms), hydrogen bond donor (formamide and methanol donor hydrogen atoms), hydrogen bond acceptor (formamide, methanol and acetaldehyde oxygen atoms), negatively charged (acetate oxygen atoms) and positively charged (methylammonium polar hydrogens). The atom densities were calculated as an average over the ten replicas per protein. These groupwise summed atom density grids were converted to grid free energies (GFEs) using the below described formula, where R being the ideal gas constant, T is the temperature, n is the calculated density for the grid voxel under study and $n_{expected}$ is the expected density, calculated by dividing the total number of atoms of the atom group in the box by the average box volume throughout the simulation (Eq. 2):

$$GFE = -RT\ln(\frac{n}{n_{expected}}) \qquad (2)$$

Performance benchmarks

We have calculated a performance test of the PART restraints in which we measure the percentage of the total simulation time going to PART restraints as a function of different hardware architectures and different probe mixtures. We tested four hardware architectures, namely 16, 64 and 128 cores of an AMD EPYC 7H12 CPU and 12 cores of an AMD Epyc 7402 CPU combined with one Nvidia A100 GPU. The system setup was similar to the Factor-Xa benchmark, with as only difference the probe concentrations. We tested seven probe mixtures in total. The first mixture was set up as an exact copy of the Factor-Xa benchmark. The second and third system used only propane as cosolvent, at concentrations of 0.25
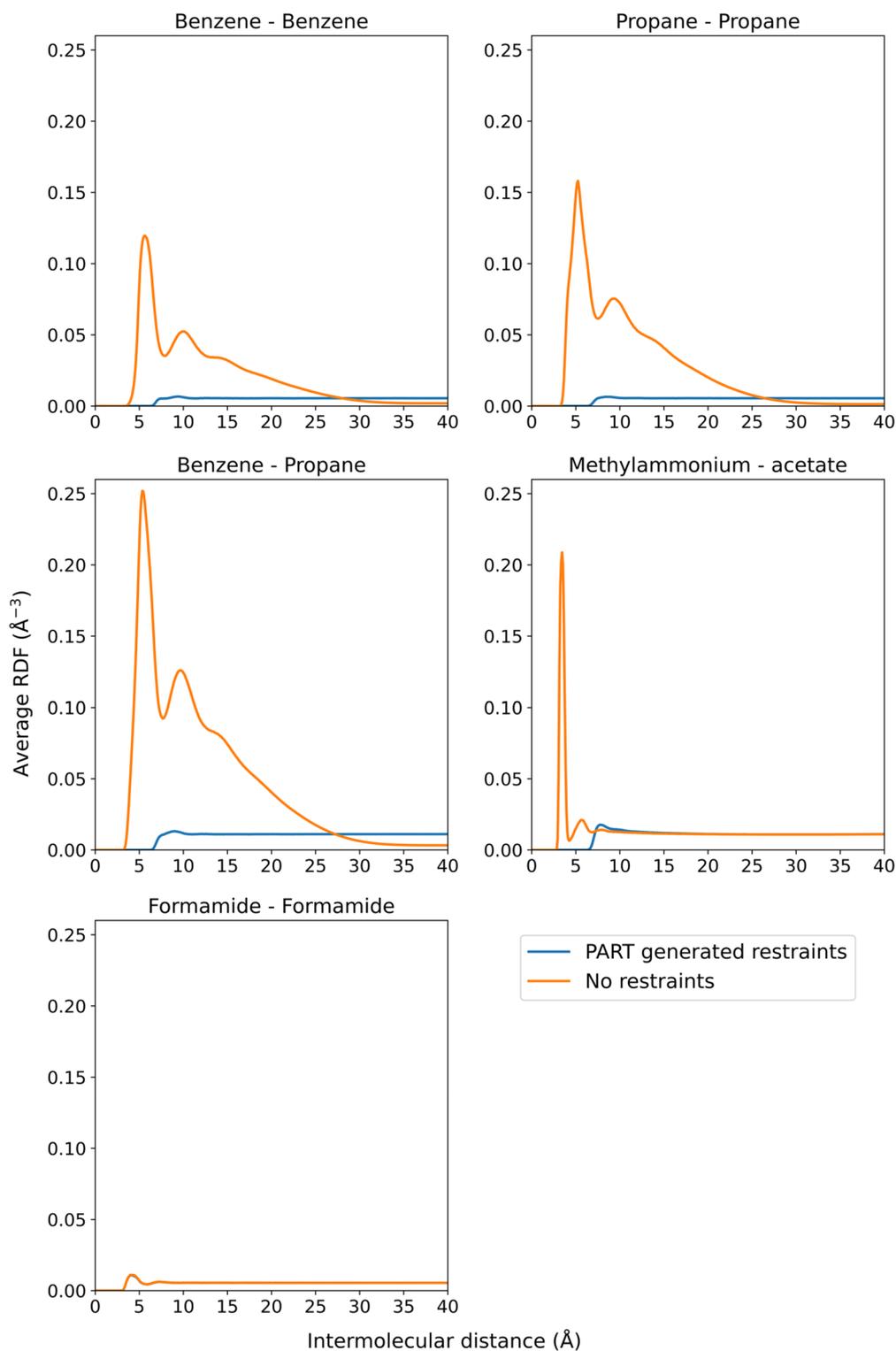
Fig. 2 Radial distribution functions for a selection of probe pairs for the Factor-Xa simulations with (blue curves) and without (orange curves) PART restraints. Radial distribution functions were calculated as an average over the ten MD replicas of the simulations with and without PART restraints. As a reference, the RDF for formamide is also shown, a molecule for which aggregation does not take place, hence the simulations with and without PART generated restraints have the same, overlapping RDF curve
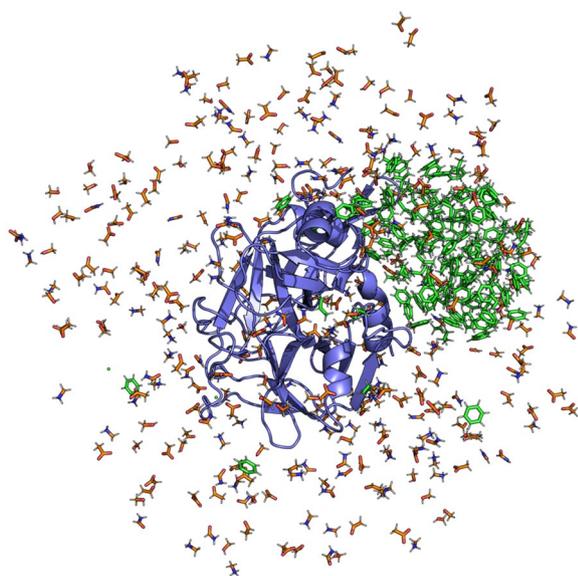
Fig. 3 Illustration of lipophilic aggregation between hydrophobic probes (green) during the Factor-Xa protein (purple) simulation without PART restraints. Water soluble probes are shown in orange

M and 0.5 M respectively. The fourth and fifth system used only benzene as cosolvent, also at concentrations of 0.25 M and 0.5 M. Comparing the propane and benzene only mixtures allows to study the influence of single atom versus COM-based distance calculations, as for propane a single atom was used and for benzene the COM of all benzene atoms. To study the influence of the number of atoms in a COM simulation, we made a sixth and seventh experiment in which we use the previously described benzene only mixtures, but with the COM calculation based on the benzene carbon atoms only. We used short MD simulations of 100 ps for the scaling tests.

Results and discussion

Comparison between restrained and unrestrained simulations

Figure 2 shows the radial distribution functions (RDFs) between a selection of cosolvent molecule types. These RDFs were calculated for the Factor-Xa mixture without PART generated restraints and for the Factor-Xa mixture with PART generated restraints, as described earlier. If aggregation takes place, RDF curves show peaks at low intermolecular distances and have less occurrences of higher intermolecular distances. Following these criteria, aggregation is clearly taking place between benzene-benzene, propane-benzene, and propane-propane pairs. Visual inspection of the trajectory also confirms this lipophilic aggregation (Fig. 3). This analysis indicates that intermolecular distance restraints between hydrophobic molecule types are necessary at higher concentrations. From the intermolecular distance matrices and by counting the number of interactions, the average fraction of intermolecular interactions between the fragments could be calculated (see Methods section). The average fraction of hydrophobic molecules participating in lipophilic aggregation was 88%, composed of 90% of propane interacting with other hydrophobic molecules, and 86% of benzenes interacting with other hydrophobic molecules. As a reference, Fig. 2 also shows the RDF for formamide, a molecule for which aggregation does not take place as it is water soluble and not formally charged.

When analyzing the ionic interactions between acetate (negative formal charge) and methylammonium (positive formal charge), a simple conclusion is less obvious. Due to the nature of these interactions, the ionic fragments form pairs rather than large aggregates. Pair formation will result in a peak at low intermolecular distances in an RDF curve, but the RDF values at longer intermolecular distances are less impacted since the ionic pairs are still randomly distributed in the simulation box and do not aggregate. The average fraction of ionic fragments
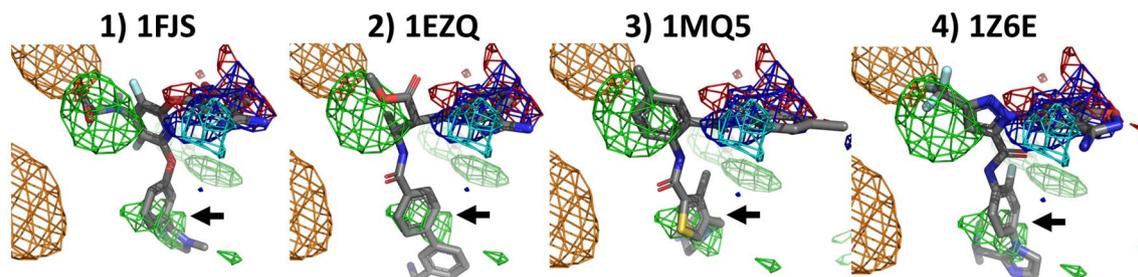
Fig. 4 Overlap between Factor-Xa ligands and the calculated densities from a Factor-Xa cosolvent MD simulation with PART generated restraints. All grid points with GFE values of less or equal to −1.5 kcal/mol are shown as a mesh (green: hydrophobic, dark blue: donor, red: acceptor, orange: negatively charged, cyan: positively charged). The black arrow highlights a hydrophobic density mesh that overlaps with a ring feature in all four ligands
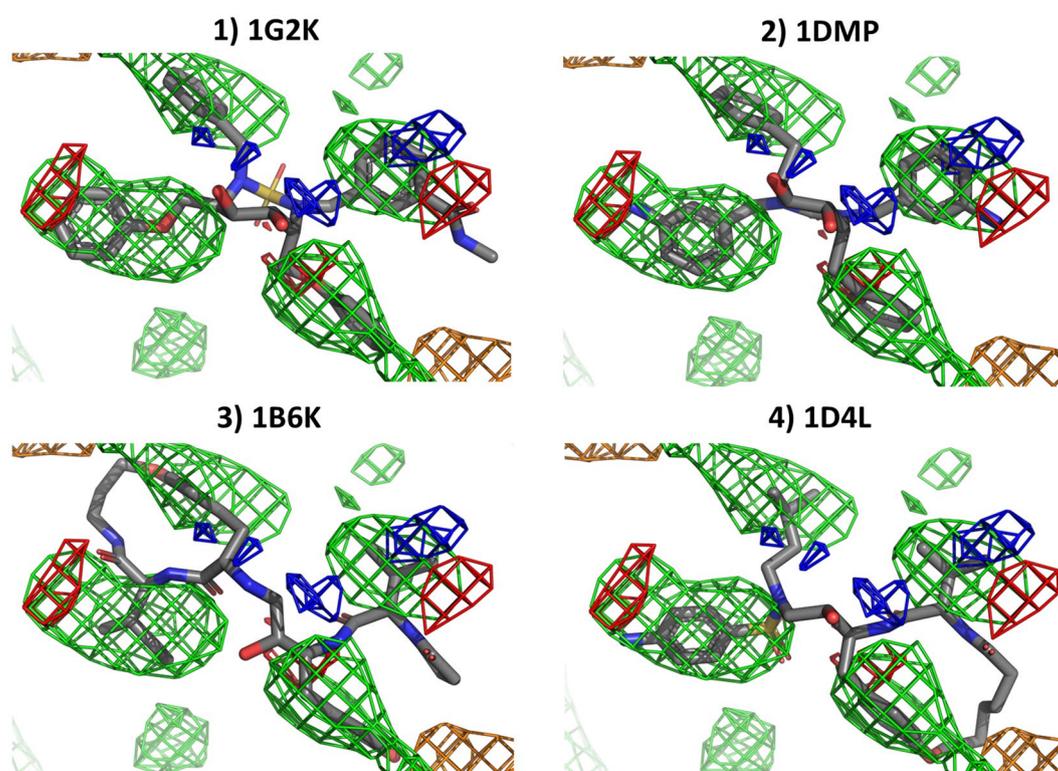
Fig. 5 Overlap between HIV protease ligands and the calculated densities from a HIV protease cosolvent MD simulation with PART generated restraints. All grid points with GFE values ≤ −1.5 kcal/mol are shown as a mesh (green: hydrophobic, dark blue: donor, red: acceptor, orange: negatively charged, cyan: positively charged)

making ion-ion interactions is 26.4%. As this is not as detrimental as compared to lipophilic aggregation, it is up to the user to decide whether it is worthwhile to include restraints between oppositely charged fragment types.

Ligand feature reproduction benchmarks

Figure 4 shows the fragment densities in Factor-Xa, as calculated from the cosolvent MD trajectory, and overlaid with several ligand structures. Positively charged fragment densities overlap nicely with ligand benzamidine groups, while negatively charged fragment density overlaps with the 1FJS ligand carbonic acid group. The trifluoromethyl group of the 1Z6E ligand is reproduced by the hydrophobic fragments, while hydrophobic rings in all four ligands (see black arrow) are also located in hydrophobic density sites.

HIV protease fragment densities (Fig. 5) also reproduce some key hydrophobic ligand features, including all four phenyl groups from the 1G2K and 1DMP ligands and the lipophilic macrocycle of the last two ligands. Hydrogen bond donor and hydrogen bond acceptor features are in general divided between spurious and non-spurious (such as the amide oxygen in the 1G2K ligand) sites. Of

note is that none of the fragments producing the donor and acceptor maps are directly influenced by PART generated restraints.

In the P38 MAP kinase benchmark, shown in Fig. 6, the hydrophobic map type corresponds to the ring systems in the ligands, as indicated by the two black arrows. The secondary amine of the 1OUY and 1BL7 ligand is also positioned inside the positive fragment density. Interestingly, the pyridine nitrogen atom of 1W84 and 1A9U also overlaps with an acceptor density, as indicated by the red arrow, although again acceptor densities are not influenced by PART.

Negatively charged phosphate groups are in general close to, or inside of, negatively charged fragment densities in the RNase A benchmark, as illustrated in Fig. 7. The aromatic system of the 6PVV ligand overlaps with calculated hydrophobic/aromatic fragment densities. Additionally, donor and acceptor densities overlap with donor and acceptor locations in the 6PVX ligand, as indicated by the blue and red arrows.

In general, we conclude from the above results that important ligand features can be qualitatively reproduced by cosolvent MD simulations with PART generated restraints. If using a high concentration of hydrophobic
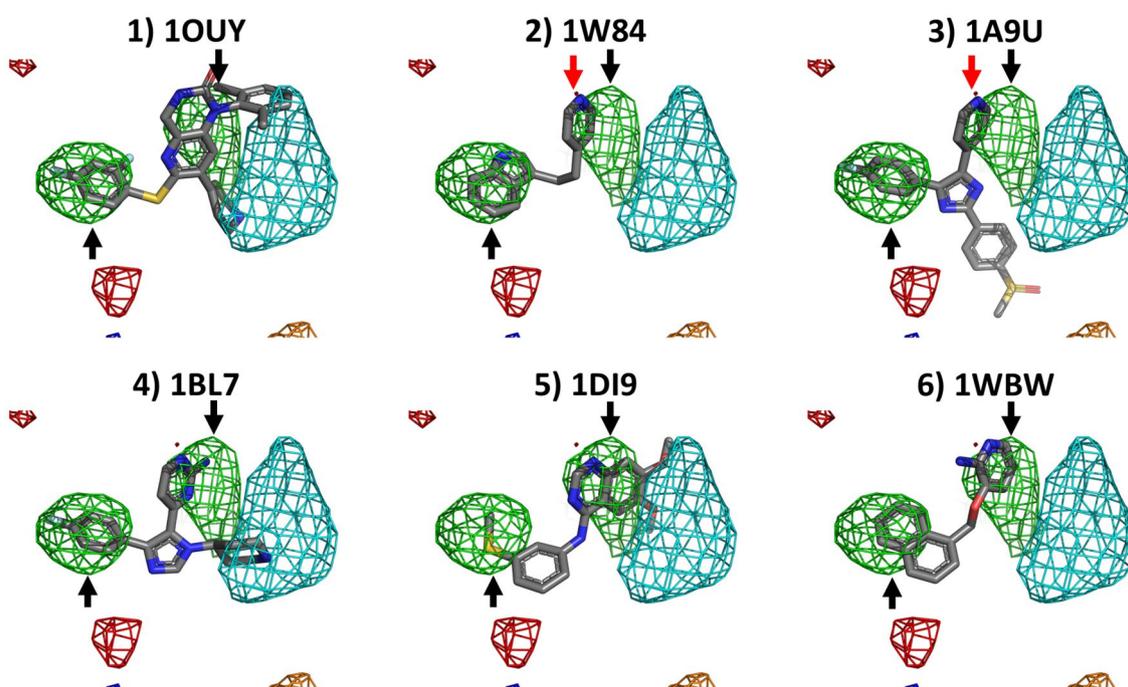
Fig. 6 Overlap between P38 MAP kinase ligands and the calculated densities from a HIV protease cosolvent MD simulation with PART generated restraints. All grid points with GFE values ≤ −1.5 kcal/mol are shown as a mesh (green: hydrophobic, dark blue: donor, red: acceptor, orange: negatively charged, cyan: positively charged). The black arrow highlights a hydrophobic density mesh that overlaps with the ring systems in the ligands. The red arrow highlights an acceptor density that overlaps with the pyridine nitrogen atom of 1W84 and 1A9U
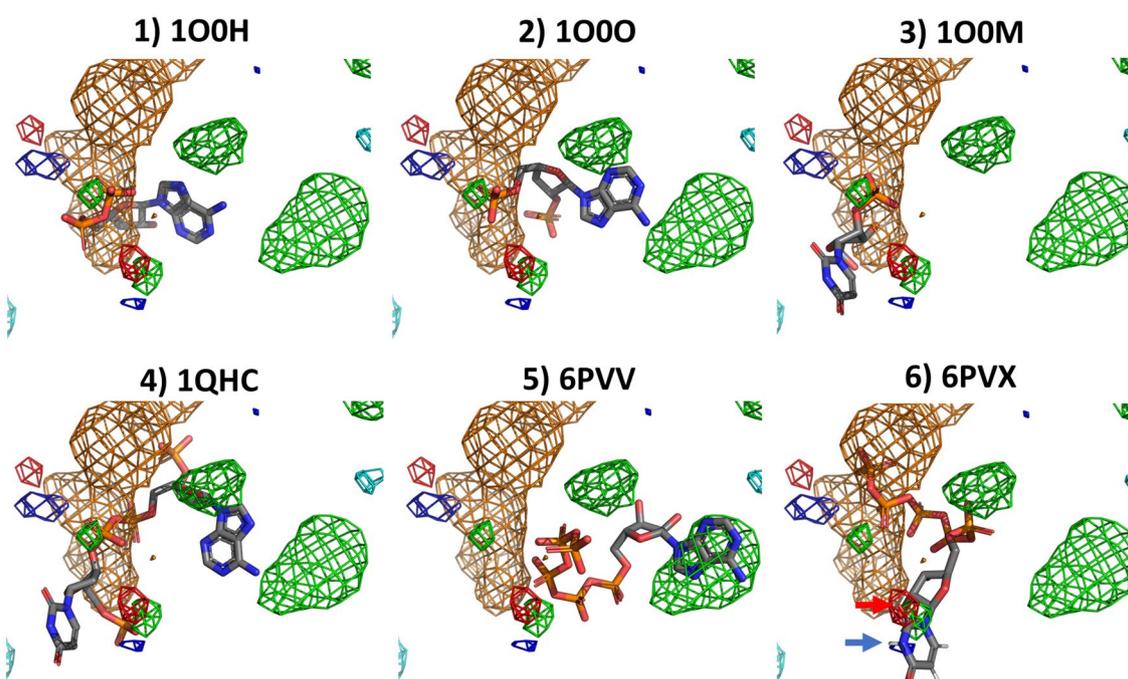
Fig. 7 Overlap between RNase A ligands and the calculated densities from a HIV protease cosolvent MD simulation with PART generated restraints. All grid points with GFE values ≤ −1.5 kcal/mol shown as a mesh (green: hydrophobic, dark blue: donor, red: acceptor, orange: negatively charged, cyan: positively charged). The red arrow highlights a hydrogen bond acceptor density that overlaps with an acceptor feature of the 6PVX ligand. The blue arrow highlights a donor density that overlaps with a donor feature of the 6PVX ligand

Table 2 Overview of the PART performance benchmark showing the percentage of the total simulation time going to PART restraints as a function of different hardware architectures and different probe mixtures

	16 CPU cores	64 CPU cores	128 CPU cores	1 A100 GPU
Mix 1 PLUMED time (%)	13	22	36	71
Mix 2 PLUMED time (%)	3	6	22	43
Mix 3 PLUMED time (%)	5	12	28	50
Mix 4 PLUMED time (%)	8	12	28	62
Mix 5 PLUMED time (%)	12	21	36	66
Mix 6 PLUMED time (%)	4	10	26	50
Mix 7 PLUMED time (%)	10	18	34	68

Mix 1 is a copy of the Factor-Xa benchmark where all benzene atoms where used to calculate the COM. Mix 2 is the Factor-Xa system where the only cosolvent is propane at a concentration of 0.25 M and Mix 3 is a copy of Mix 2 but with a probe concentration of 0.5 M. Mix 4 is the Factor-Xa system where the only cosolvent is benzene at a concentration of 0.25 M. Mix 5 is a copy of Mix 4 but with a probe concentration of 0.5 M. In Mix 4 and Mix 5, the COM of benzenes was calculated using all benzene atoms. Mix 6 and Mix 7 are a copy of Mix 4 and Mix 5 respectively, but now calculating the COM using only the benzene carbon atoms

probes, intermolecular restraints will ensure that fragments densities accumulate in the appropriate pockets at an expected sampling speed. These results are in line with benchmarks done using other methodologies such as SILCS [16].

Performance benchmarks
The results from the performance benchmarks are shown in Table 2. Clearly the PLUMED part of the calculation does not scale as well over increasing cores as the GROMACS part of the calculation. When making use of a GPU, the GROMACS part of the calculation can become faster than PLUMED part of the calculation, severely impacting total simulation speed. Consequently, we advise users of PART to make use of CPU partitions, while using one full CPU per simulation and run multiple replicas in parallel. We note that if future PLUMED code is better optimized for the calculations used in PART, then these scaling test results might change.

Conclusion
We have shown that PART is a new easy-to-use alternative for current methodologies for cosolvent MD simulations involving hydrophobic cosolvent molecules. The main advantages of PART are short setup times (especially compared to force field modifications where these modifications need to be validated), a restraint potential that can be easily modified, and that the recommended force field cutoffs can be used. Regarding the last point, in competing technologies such as SILCS, force field cutoffs are modified to make sure that the restraint potential introduced by force field modifications have the correct shape. Such force field cut-off changes are not required with PART-generated restraint potentials.

The main disadvantage of PART is that an extra PLUMED overhead time is introduced to the calculation.

It is up to the user to decide which arguments are more important when choosing a methodology.

Additionally, we have demonstrated that restrained simulations increase the effective concentration of hydrophobic probes by preventing aggregation. Cosolvent MD simulations involving PART-generated restraints on four benchmarked proteins also reproduce known ligand features qualitatively.

Author contributions
OB developed and benchmarked the software. HDW supervised the research. Both authors contributed to writing the publication. Both authors read and approved the final manuscript.

Availability of data and materials
Project name: PART. Project home page: https://github.com/UAMC-Olivier/PART. Archived version: N/A. Operating system(s): Platform independent. Programming language(s): Python. Other requirements: MD engine of choice patched with PLUMED, numpy module for Python. License: MIT license.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Not applicable.

References

1. Seco J, Luque FJ, Barril X (2009) Binding site detection and druggability index from first principles. J Med Chem 52(8):2363–2371. https://doi.org/10.1021/jm801385d
2. Guvench O, Mackerell AD (2009) Computational fragment-based binding site identification by ligand competitive saturation. PLoS Comput Biol 5(7):e1000435. https://doi.org/10.1371/journal.pcbi.1000435
3. Bakan A et al (2012) Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. J Chem Theory Comput 8(7):2435–2447. https://doi.org/10.1021/ct300117j
4. Ghanakota P, Carlson HA (2016) Moving beyond active-site detection: MixMD applied to allosteric systems. J Phys Chem B 120(33):8685–8695
5. Yang C-Y, Wang S (2010) Computational analysis of protein hotspots. ACS Med Chem Lett 1(3):125–129
6. Ghanakota P, Carlson HA (2016) Driving structure-based drug discovery through cosolvent molecular dynamics. J Med Chem 59(23):10383–10399. https://doi.org/10.1021/acs.jmedchem.6b00399
7. Comitani F, Gervasio FL (2018) Exploring cryptic pockets formation in targets of pharmaceutical interest with SWISH. J Chem Theory Comput 14(6):3321–3331. https://doi.org/10.1021/acs.jctc.8b00263
8. Tze-Yang Ng J, Tan YS (2022) Accelerated ligand-mapping molecular dynamics simulations for the detection of recalcitrant cryptic pockets and occluded binding sites. J Chem Theory Comput 18(3):1969–1981. https://doi.org/10.1021/acs.jctc.1c01177
9. Yang S-Y (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discovery Today 15(11–12):444–450
10. Yu W et al (2015) Pharmacophore modeling using site-identification by ligand competitive saturation (SILCS) with multiple probe molecules. J Chem Inf Model 55(2):407–420. https://doi.org/10.1021/ci500691p
11. Lee JY et al (2020) Pharmmaker: pharmacophore modeling and hit identification based on druggability simulations. Protein Sci 29(1):76–86. https://doi.org/10.1002/pro.3732
12. Dixon SL et al (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Computer-Aided Mol Design 20(10–11):647–671. https://doi.org/10.1007/s10822-006-9087-6
13. Faller CE, et al (2015) Site Identification by Ligand Competitive Saturation (SILCS) simulations for fragment-based drug design. In. Springer New York, pp 75–87. https://doi.org/10.1007/978-1-4939-2486-8_7
14. MacKerell AD Jr et al (1864) 2020) Identification and characterization of fragment binding sites for allosteric ligand design using the site identification by ligand competitive saturation hotspots approach (SILCS-Hotspots. Biochimica et Biophysica Acta BBA General Subjects 4:129519
15. Raman EP et al (2011) Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. J Chem Inf Model 51(4):877–896. https://doi.org/10.1021/ci100462t
16. Raman EP et al (2013) Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. J Chem Inf Model 53(12):3384–3398. https://doi.org/10.1021/ci4005628
17. Ustach VD et al (2019) Optimization and evaluation of site-identification by ligand competitive saturation (SILCS) as a tool for target-based ligand optimization. J Chem Inf Model 59(6):3018–3035. https://doi.org/10.1021/acs.jcim.9b00210
18. Yu W et al (2014) Site-Identification by Ligand Competitive Saturation (SILCS) assisted pharmacophore modeling. J Comput Aided Mol Des 28(5):491–507. https://doi.org/10.1007/s10822-014-9728-0
19. Huang J, Mackerell AD (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem 34(25):2135–2145. https://doi.org/10.1002/jcc.23354
20. Tan YS et al (2012) Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1. Angew Chem Int Ed 51(40):10078–10081. https://doi.org/10.1002/anie.201205676
21. (2019) Promoting transparency and reproducibility in enhanced molecular simulations. Nat Methods 16 (8):670–673. https://doi.org/10.1038/s41592-019-0506-8
22. Bonomi M et al (2009) PLUMED: a portable plugin for free-energy calculations with molecular dynamics. Comput Phys Commun 180(10):1961–1972. https://doi.org/10.1016/j.cpc.2009.05.011
23. Tribello GA et al (2014) PLUMED 2: new feathers for an old bird. Comput Phys Commun 185(2):604–613. https://doi.org/10.1016/j.cpc.2013.09.018
24. Adler M et al (2000) Preparation, characterization, and the crystal structure of the inhibitor ZK-807834 (CI-1031) complexed with factor Xa,. Biochemistry 39(41):12534–12542. https://doi.org/10.1021/bi001477q
25. Word JM et al (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 285(4):1735–1747
26. Chresten RS et al (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. J Chem Theory Comput 7:2284–2295. https://doi.org/10.1021/ct200133y
27. Mats HMO et al (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 7:525–537. https://doi.org/10.1021/ct100578z
28. Abraham MJ et al (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1:19–25
29. Van Der Spoel D et al (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718. https://doi.org/10.1002/jcc.20291
30. Huang J et al (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14(1):71–73. https://doi.org/10.1038/nmeth.4067
31. Jorgensen WL et al (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
32. Vanommeslaeghe K, et al (2009) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry:NA-NA. https://doi.org/10.1002/jcc.21367
33. Vanommeslaeghe K, MacKerell AD Jr (2012) Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J Chem Inf Model 52(12):3144–3154
34. Vanommeslaeghe K, Raman EP, MacKerell AD Jr (2012) Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. J Chem Inf Model 52(12):3155–3168
35. Yu W et al (2012) Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. J Comput Chem 33(31):2451–2468
36. Hess B et al (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472
37. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N·log (N) method for Ewald sums in large systems. J Chem Phys 98(12):10089–10092
38. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126(1):014101. https://doi.org/10.1063/1.2408420
39. Bernetti M, Bussi G (2020) Pressure control using stochastic cell rescaling. J Chem Phys 153(11):114107. https://doi.org/10.1063/5.0020514
40. Schaal W et al (2001) Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors. J Med Chem 44(2):155–169. https://doi.org/10.1021/jm001024j
41. Fitzgerald CE et al (2003) Structural basis for p38α MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity. Nat Struct Mol Biol 10(9):764–769
42. Vitagliano L et al (2002) Reversible substrate-induced domain motions in ribonuclease A. Proteins Struct Function Genet 46(1):97–104. https://doi.org/10.1002/prot.10033
43. Shepard SM et al (2019) Nucleoside tetra-and pentaphosphates prepared using a tetraphosphorylation reagent are potent inhibitors of ribonuclease A. J Am Chem Soc 141(46):18400–18404
44. Naveen M-A et al (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 32:2319–2327. https://doi.org/10.1002/jcc.21787,issue=10
45. Richard JG, et al (2016) MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference:98–105
46. Maignan S et al (2000) Crystal structures of human factor Xa complexed with potent inhibitors. J Med Chem 43(17):3226–3232

47. Adler M et al (2002) Crystal structures of two potent nonamidine inhibitors bound to factor Xa. Biochemistry 41(52):15514–15523
48. Quan ML et al (2005) Discovery of 1-(3'-Aminobenzisoxazol-5'-yl)-3-trifluoromethyl-N-[2-fluoro-4-[(2'-dimethylaminomethyl) imidazol-1-yl] phenyl]-1 H-pyrazole-5-carboxyamide hydrochloride (Razaxaban), a highly potent, selective, and orally bioavailable factor Xa inhibitor. J Med Chem 48(6):1729–1744
49. Hodge CN et al (1996) Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. Chem Biol 3(4):301–314
50. Martin JL et al (1999) Molecular recognition of macrocyclic peptidomimetic inhibitors by HIV-1 protease. Biochemistry 38(25):7978–7988
51. Tyndall JD et al (2000) Synthesis, stability, antiviral activity, and protease-bound structures of substrate-mimicking constrained macrocyclic inhibitors of HIV-1 protease. J Med Chem 43(19):3495–3504
52. Gill AL et al (2005) Identification of novel p38α MAP kinase inhibitors using fragment-based lead generation. J Med Chem 48(2):414–426
53. Wang Z et al (1998) Structural basis of inhibitor selectivity in MAP kinases. Structure 6(9):1117–1128
54. Shewchuk L et al (2000) Binding mode of the 4-anilinoquinazoline class of protein kinase inhibitor: X-ray crystallographic studies of 4-anilino-quinazolines bound to cyclin-dependent kinase 2 and p38 kinase. J Med Chem 43(1):133–138
55. Leonidas DD et al (2003) High-resolution crystal structures of ribonuclease A complexed with adenylic and uridylic nucleotide inhibitors. Implications for structure-based design of ribonucleolytic inhibitors. Protein Sci 12(11):2559–2574
56. Leonidas DD et al (1999) Toward rational design of ribonuclease inhibitors: high-resolution crystal structure of a ribonuclease A complex with a potent 3', 5'-pyrophosphate-linked dinucleotide inhibitor. Biochemistry 38(32):10287–10297

Publisher's Note