Technische Universität München
TUM School of Life Sciences

Molecular Modeling of Macrocyclic Inhibitors

Maximilian Meixner

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Mathias Wilhelm

Prüfer*innen der Dissertation:

1. Prof. Dr. Dmitrij Frischmann
2. Prof. Dr. Martin Zacharias

Die Dissertation wurde am 25.01.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 05.05.2023 angenommen.

Table of Contents

Summary

Macrocyclic inhibitors are an interesting class of therapeutics that are yet to be fully understood. Due to the structural constraint of their characteristic ring scaffold, the bound conformation of such ligands is often locked, increasing target specificity while lowering entropic binding costs. Despite exceeding the typical rule-of-5 space, some macrocycles obtain decent oral bioavailability through passive membrane permeability, attributed to environment-dependent conformational changes of the ring. These promising features contribute to their therapeutic potential. However, macrocycles proved to be challenging for computational modeling, particularly for structural methods like ligand conformational sampling and during pose generation of molecular docking calculations. Such methods must precisely reproduce the complex dynamics and flexibility of the macrocyclic ring for accurate structural description. Due to the interdependence of ring torsions, many sampling and docking algorithms exclude such moieties by default, leaving the ring conformation unaltered. Without compensation, these techniques fail to correctly represent the ligands conformational space, introducing computational artifacts and structural errors. In this work, a suitable strategy for combined conformational sampling and molecular docking of macrocycles was developed. The workflow is based on molecular dynamics simulations and includes a separate conformational sampling step before the actual docking procedure. In these simulations, the macrocycle conformational space was exhaustively explored employing optimized sampling conditions. To extract a subset of ring structures including near bioactive states, a classification scheme based on the sampled dihedral angles of ring torsions was developed for in-depth conformational analysis. Such conformer ensembles were then subjected to flexible receptor docking calculations performed with DynaDock to compensate the lack of ring flexibility during pose generation. Every initial pose was subsequently refined by optimized potential molecular dynamics simulations, allowing for the dynamic adaption of binding partners. This pipeline was tested in detailed case studies and evaluated in a benchmark study including 20 diverse macrocycles and their host proteins in bound and unbound states. It could be shown that with this dynamics-based approach highly accurate binding poses could be obtained, thereby overcoming major structural challenges in molecular modeling of macrocycles.

Zusammenfassung

Makrozyklische Inhibitoren sind eine interessante Klasse von Therapeutika, die in ihrer Vollständigkeit noch nicht verstanden sind. Aufgrund der charakteristischen Ringstruktur, die eine konformative Einschränkung darstellt, ist die gebundene Form solcher Liganden oft fixiert, was die Spezifizität zum Zielprotein erhöht und gleichzeitig die entropischen Bindungskosten senkt. Obwohl sie die Grenzen der Rule of Five (Faustregel für die orale Bioverfügbarkeit chemischer Substanzen) meist überschreiten, weisen manche Makrozyklen eine hohe orale Bioverfügbarkeit durch passive Membranpermeabilität auf, welche der umgebungsbedingten Anpassung der Ringstruktur zugeschrieben werden kann. Solche vielversprechenden Eigenschaften machen das therapeutische Potenzial der Makrozyklen aus. Für die Computermodellierung, insbesondere für strukturbasierte Methoden wie das konformative Sampling und die Platzierung des Liganden in der Proteinbindetasche (molekulares Docking), stellen sie jedoch eine besondere Herausforderung dar. Für eine exakte strukturelle Darstellung von Makrozyklen ist es notwendig, dass solche Methoden die komplexe Dynamik und Flexibilität des Rings berücksichtigen. Aufgrund der gegenseitigen Abhängigkeit von Ringtorsionen schließen viele Sampling und Docking Algorithmen die Ringstruktur automatisch aus und lassen somit die Konformation des Rings unverändert. Ohne einen entsprechenden Ausgleich können solche Methoden den Konformationsraum dieser Liganden nicht korrekt darstellen, was zu Rechenartefakten und Strukturfehlern führt. In dieser Arbeit wurde eine geeignete Strategie für die molekulare Modellierung von Makrozyklen entwickelt. Sie basiert auf Molekulardynamik Simulationen und beinhaltet ein separates Konformations-Sampling des Liganden vor dem eigentlichen Docking Schritt. Mit speziell für Makrozyklen optimierten Simulationsbedingungen konnte der Konformationsraum hinreichend gesampelt werden. Außerdem wurde ein torsionsbasiertes Klassifikationsverfahren entwickelt, um die erzeugten Konformationen im Detail zu analysieren und ein Ensemble an Ringstrukturen zu extrahieren, welches auch den gebundenen Zustand erfasst. Die Strukturen des Ensembles wurden dann als Startstrukturen für Docking Simulationen verwendet, die mit DynaDock durchgeführt wurden. Dadurch konnte der Mangel an Flexibilität des Rings, dessen Konformation während der Erzeugung von Bindeposen nicht verändert wurde, kompensiert werden. Die Molekulardynamik jeder Pose wurde anschließend unter Verwendung optimierter Potenziale simuliert, um eine dynamische Anpassung der

Bindungspartner zu ermöglichen. Diese Pipeline wurde in detaillierten Fallstudien getestet und in einer Benchmark-Studie mit 20 verschiedenen Makrozyklen und ihren Wirtsproteinen in gebundenem und ungebundenem Zustand evaluiert. Durch diesen dynamikbasierten Ansatz konnten die entsprechenden Bindungspositionen mit hoher Genauigkeit bestimmt und somit einige strukturelle Herausforderungen der molekularen Modellierung von Makrozyklen überwunden werden.

List of publications

First-author publications included in this dissertation:

Meixner, M., Zachmann, M., Metzler, S., Scheerer, J., Zacharias, M., & Antes, I. (2022). Dynamic Docking of Macrocycles in Bound and Unbound Protein Structures with DynaDock. *Journal of Chemical Information and Modeling*, *62*(14), 3426-3441.

Wang, S.*, **Meixner, M.*,** Yu, L., Zhuo, L., Karmann, L., Kazmaier, U., Vollmar, A. M., Antes, I., Zahler, S. (2021). Turning the Actin Nucleating Compound Miuraenamide into Nucleation Inhibitors. *ACS omega*, *6*(34), 22165-22172.

**Shared first authors*

Co-authored publications relevant to this dissertation:

Schütz, R., **Meixner, M.,** Antes, I., & Bracher, F. (2020). A modular approach to the bisbenzylisoquinoline alkaloids tetrandrine and isotetrandrine. *Organic & Biomolecular Chemistry*, *18*(16), 3047-3068.

Eyermann, B., **Meixner, M.,** Brötz-Oesterhelt, H., Antes, I., & Sieber, S. A. (2020). Acyldepsipeptide probes facilitate specific detection of caseinolytic protease P independent of its oligomeric and activity state. *ChemBioChem*, *21*(1-2), 235-240.

These publications appear in chronological order.

Other publications by the author (not included in this dissertation):

Nogueira, J. J., **Meixner, M.,** Bittermann, M., & González, L. (2017). Impact of lipid environment on photodamage activation of methylene blue. *ChemPhotoChem*, *1*(5), 178-182.

1. Introduction

According to the International Union of Pure and Applied Chemistry, a macrocycle (MC) is a cyclic macromolecule or a cyclic portion of a macromolecule, but the term is also used to describe cyclic molecules of low molecular mass[1]. Often conventionally defined as molecules with at least 12 ring atoms[2], the official definition, however, which is followed throughout this work, does not specify a minimum ring size for MCs. A steadily growing research interest in MCs, as illustrated in **Figure 1**, reflects their emerging significance in science. The plot shows the PDB* and PubMed† text search results for entries including the terms "macrocycle" or "macrocyclic" in groups of 5 years, as well as cumulatively since 1975, and elucidates a continuous increase in available structural data and scientific citations until today. The importance of MCs as a new class of therapeutics was fueled by successful clinical trials of natural products (NPs) in the 1990s. Compounds like sirolimus (rapamycin) and tacrolimus (FK506), which are included in the following, were approved for medical use shortly thereafter[3].

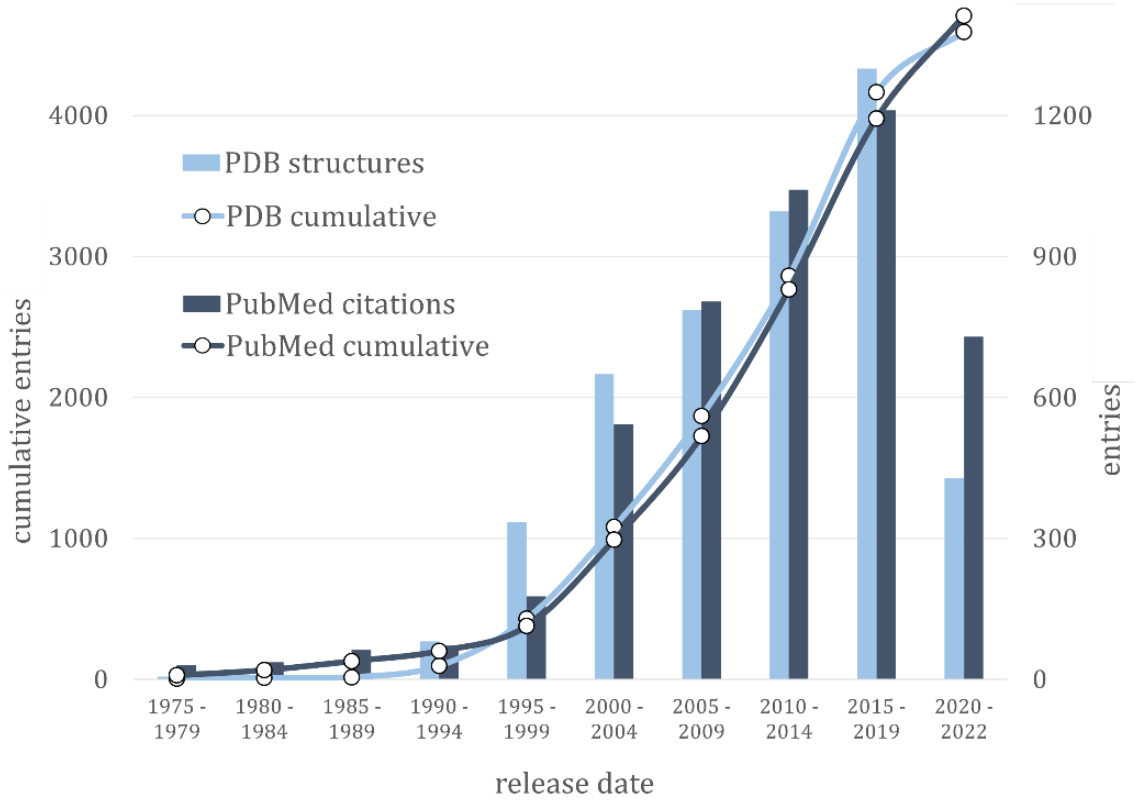

Figure 1. Entries in the PDB and PubMed databases including "macrocycle" or "macrocyclic" in groups of 5 years (columns, right axis) and their cumulative entries since 1975 (lines, left axis).

* Protein Data Bank, https://www.rcsb.org/

† https://pubmed.ncbi.nlm.nih.gov/

1.1. Examples of macrocycles in nature and their therapeutic use

All example NPs introduced in this section were computationally investigated in this work. A substantial overview of molecules is provided in "Macrocycles in Drug Discovery" by Jeremy Levin[4] and "Practical Medicinal Chemistry with Macrocycles" by Eric Mersault and Mark Peterson[5]. Natural MCs are secondary metabolites originating from plants and microbes[2]. Their complex structures exhibit antagonistic function and thus provide evolutionary advantage for the host. In fact, their existence and production via sophisticated biosynthesis pathways favored the hypothesis that the presence of secondary metabolites is intended rather than coincidental[6, 7].

Molecules of the ansamycin class consist of an aromatic moiety such as benzene or naphthalene, or derivatives thereof, that is incorporated in a MC ring. The benzoquinone ansamycin geldanamycin (GDM, **Figure 2A**), a secondary metabolite of the soil bacterium

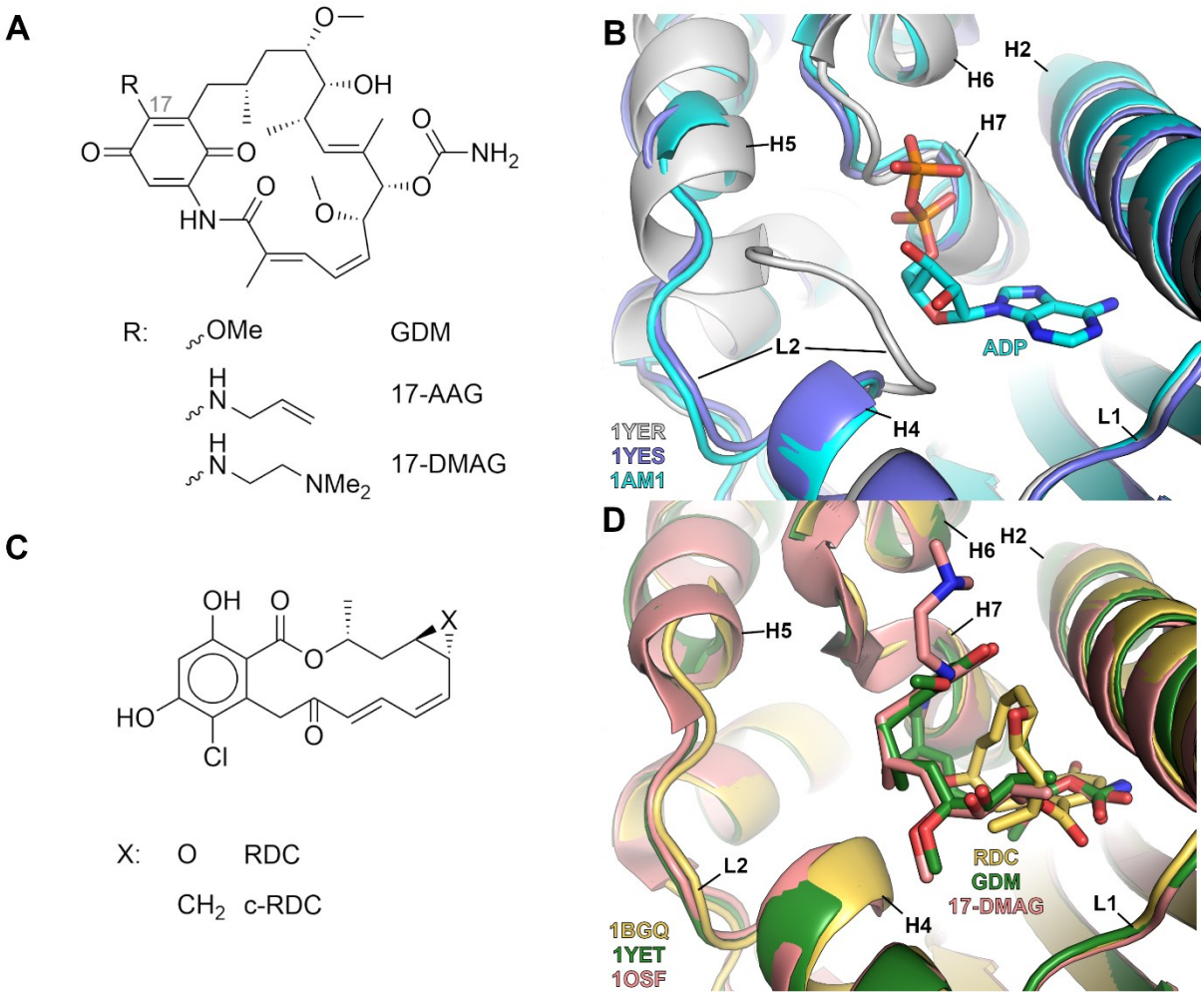

Figure 2. A) 2D structures of macrocyclic Hsp90 inhibitors geldanamycin (GDM) and derivatives. **B)** Crystal structure of the nucleotide binding site of Hsp90 bound by ADP (PDB-ID: 1AM1) and in two unbound forms with open and closed L2 loop (1YES and 1YER, respectively). **C)** Same as A) for radicicol (RDC) and the derivative c-RDC. **D)** Same as B) bound by GDM (1YET), RDC (1BGQ) and 17-DMAG (1OSF). Secondary structure labeling according to Stebbins *et al*[10].

Streptomyces hygroscopicus var. *geldanus* with antimicrobial activity was first characterized in 1970[8]. Its antiproliferative effects were demonstrated in 1994, as oncogenic cellular transformations caused by tyrosine kinases, could be indirectly reversed by GDM by inhibiting the heat shock protein 90 (Hsp90), a chaperone assisting the proper maturation and refolding of cell-cycle regulating protein kinases and other proto-oncogenic client proteins[9]. **Figure 2B** shows the nucleotide binding cleft in the N-terminal domain of Hsp90, characterized by the flexible L2 loop. GDM targets this binding site, thereby altering ATPase activity, and ultimately, chaperone function, which results in the degradation of Hsp90 substrates and indirect cell death[10, 11].

Radicicol (RDC, **Figure 2C**), a natural macrolide also known as monorden, exhibiting antifungal properties and produced by the fungus *Monosporium bonorden*[12], is another MC inhibitor of Hsp90, and like GDM, a potential candidate as antitumor therapeutic. The binding affinity of RDC for the same ATP binding site of Hsp90 is significantly stronger than that of GDM, with a K_d = 19 nM (K_d = 1200 nM for GDM), making it the most potent MC inhibitor among NPs[13]. Despite their promising antitumor activity *in vitro*, undesirable effects *in vivo* prevented both compounds from further proceeding to clinical testing. For GDM, the characteristic quinone moiety proved to be redox-active, producing superoxide radicals that led to fatal cell damage unrelated to Hsp90 inhibition[14]. Moreover, GDM caused hepatotoxicity, which was attributed to the electrophilic quinonimine species, obtained through oxidative metabolism, depleting detoxifying hepatic glutathione reservoirs, and eventually reacting with biomolecules[15, 16]. Attempting to improve the *in vivo* profile, structural modifications targeting position 17 in the quinone ring of GDM replaced its methoxy group with an allylamine moiety producing 17-AAG* (tanespimycin). Among many derivatives, 17-AAG proved to be the most potent with comparable antiproliferative effects as GDM on some cancer cell lines[17]. Additionally, 17-AAG showed increased inhibitory activity and stability *in vivo* compared to the parent compound[18]. Another promising derivative with a similar biological profile as 17-AAG but with improved aqueous solubility was 17-DMAG† (alvespimycin)[19]. Although both analogues entered clinical trials, adverse effects eventually led to their discontinuation. **Figure 2D** illustrates that both NPs and the derivative 17-DMAG indeed bind the same binding site in the open L2 state. However,

* 17-allylamino-17-demethoxygeldanamycin

† 17-dimethylaminoethylamino-17-demethoxygeldanamycin

GDM and RDC fill the cleft to different extents. Exemplarily for RDC, the epoxide moiety is prone for nucleophilic attack, inactivating the compound *in vivo*. A metabolically stable derivative is cycloproparadicicol (c-RDC), where the epoxide is replaced by a cyclopropyl ring. This functional group retains the strain on the MC scaffold, pre-determining the MC ring conformation that is crucial for recognition. Although the epoxide in RDC is involved in hydrogen bonding with the binding site of Hsp90, this interaction only stabilizes the bound complex and is not essential for its activity. Thus, c-RDC remains potent with a bearable drop in affinity[20].

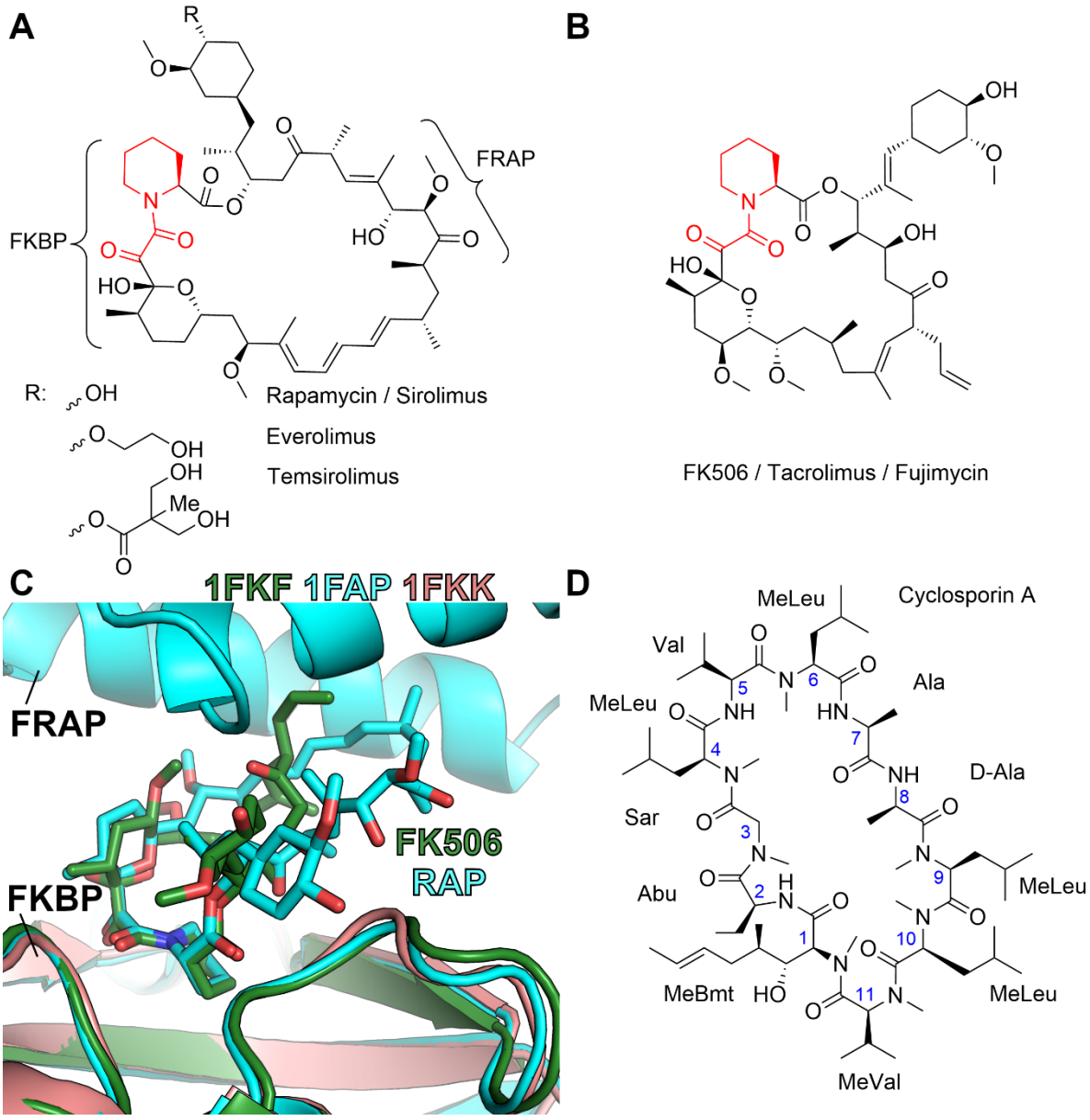

Figure 3. Immunosuppressive MCs. 2D structures of **A)** rapamycin (or sirolimus) and rapalogs (everolimus and temsirolimus) and **B)** FK506 (tacrolimus or fujimycin). **C)** Crystal structures of bound FK506 (PDB-ID: 1FKF) and rapamycin (1FAP) in complex with their immunophilin target FKBP, superimposed on the apo protein (1FKK). For the rapamycin-bound complex, the FRAP domain of the effector protein complex was co-crystallized. **D)** 2D structure of cyclic peptide cyclosporin A, labeling according to Loosli *et al*[60].

Another interesting approach for designing improved Hsp90 ligands was a combination of both RDC and GDM scaffolds, producing the chimeric inhibitor randamycin. This derivative proved to be biologically active, as shown by the degradation of Hsp90 clients in breast cancer cells and serves as a promising class of potential Hsp90 inhibitors, with new options for structural variations[21, 22].

The same bacterium that produces GDM also builds the antibiotic rapamycin (RAP, sirolimus, **Figure 3A**). This macrolide was first discovered in 1975 and exhibits antifungal properties[23]. However, in mammalian cells, RAP demonstrated immunosuppressive and antiproliferative effects[24]. The structurally related MC FK506 (tacrolimus or fujimycin, **Figure 3B**) produced by *Streptomyces tsukubaensis* competes with RAP for binding the immunophilin FK506-binding protein (FKBP, also FKBP12), a cytosolic *cis*/*trans* isomerase that catalyzes the rotation of amide bonds in peptidyl proline residues. Both NPs are strong inhibitors with sub-nanomolar binding affinities (K_d = 0.2 and 0.4 nm for RAP and FK506, respectively) and share the consensus binding motif, an α-keto homoprolyl amide (highlighted in red). **Figure 3C** illustrates the alignment of this group in the bound complexes, proving identical binding modes.

However, the sole inhibition of FKBP's rotamase activity could not be the reason for the compounds' immunosuppressive effects, evidenced by the fact that although binding the same cytosolic protein, both MCs seem to affect T-cell activation in different ways. This led to the hypothesis that the two molecules act as pro-drugs that form a gain-of-function complex with FKBP, which subsequently interacts with distinct signaling pathways involved in T-cell activation[25]. FK506-FKBP inactivates calcineurin, a cytosolic Ca^{2+} calmodulin-dependent serine/threonine phosphatase that activates NFAT (nuclear factor of activated T-cells), a transcription factor which initiates cytokine production. Thus, FK506 is associated with calcium dependent signaling and early events of antigen-induced activation of T-cells[26, 27]. In contrast, the RAP-FKBP complex allosterically binds the serine/threonine kinase mTOR (mechanistic target of rapamycin) as part of mTORC1 (mTOR complex 1), a multi-protein master regulator of the mTOR signaling pathway. mTORC1 is responsible for sensing environmental signals such as growth factors, nutrients, stress, energy and oxygen levels and reacts accordingly by phosphorylating and thus activating many downstream proteins, controlling cellular functions related to growth, macromolecule biosynthesis, cell cycle progression and metabolism[28]. In that way, RAP inhibits the cytokine-induced proliferation and antibody production of T- and B-lymphocytes without affecting lymphokine production

itself[27]. Additionally, RAP became a therapeutic candidate for cancer, where the mTOR pathway is de-regulated.

As one of the first MC drugs of natural origin, RAP denotes an impressive clinical history counting multiple indications, thereby illustrating remarkable therapeutic versatility. Due to its strong immunosuppressive effects, RAP was first marketed for the prevention of organ transplant rejection under the brand name Rapamune®, approved by the U.S. Food and Drug Administration (FDA) in 1999[29] and by the European Medicines Agency (EMA) in 2001[30]. Furthermore, Rapamune® was also approved in 2015 (2018) by the FDA (EMA) against LAM*, a progressive lung disease. More recently, RAP was approved by the FDA under the brand name Fyarro®[31] for adults with advanced malignant or metastatic PEComa† in 2021, and in 2022 as Hyftor™[32], the first topical formulation of RAP, indicated for facial angiofibroma associated with tuberous sclerosis. Analogs of RAP (called rapalogs), such as temsirolimus and everolimus (**Figure 3A**), approved by the FDA in 2007 and 2009, respectively, are used against several types of cancers[33, 34].

1.2. Characteristic features of macrocycles

The central characteristic of all MCs, the large ring system, builds a structural scaffold that constrains the compounds overall molecular geometry, resulting in structurally pre-organization of attached functional groups. This structural strain locks the bioactive conformation while concomitantly removing unfavorable conformations in solution. Thereby, the entropic barrier for target binding is lowered, improving the ligand's binding affinity, thus lowering its effective dosage[2]. Moreover, the removal of unfavorable conformations generally reduces potential off-target interactions and toxicity, which leads to improved target specificity[35].

Due to their structural topology and molecular size, MC compounds are not limited to classical inactivation of an enzyme's active site. Their advanced binding modes enable unique possibilities of inhibition, tackling low-druggability targets with challenging binding sites like large, featureless surfaces and protein-protein interactions (PPIs) that cannot be targeted by typical small molecules[36, 37]. As demonstrated by the immunosuppressive MCs (**Figure 3**), the large ring scaffold features dedicated regions for binding and modulating different target proteins. While RAP and FK506 share the characteristic FKBP-binding motif, the opposing side differentiates between effector

* Lymphangioleiomyomatosis

† Perivascular Epithelioid Cell tumor (sarcoma)

proteins. In this way, MCs form complex intermolecular interactions facilitating the mediation of PPIs through sophisticated binding modes. **Figure 4** shows MC inhibitors of the HCV NS3/4A protease, some of which are approved drugs (danoprevir and simeprevir). The shallow active site is highly surface-exposed and features multiple sub-sites[38, 39]. Here, the MC ring acts as central anchor from which one (NHN), two (30B) or three (TSV, 9H4) substituents extend in different directions, covering a large area of the protein. Thereby, the compounds themselves form a surface-like topology, matching the protein's shape. Such mode of action was also demonstrated for the NP erythromycin and its semi-synthetic derivatives clarithromycin and roxithromycin, approved antibiotic

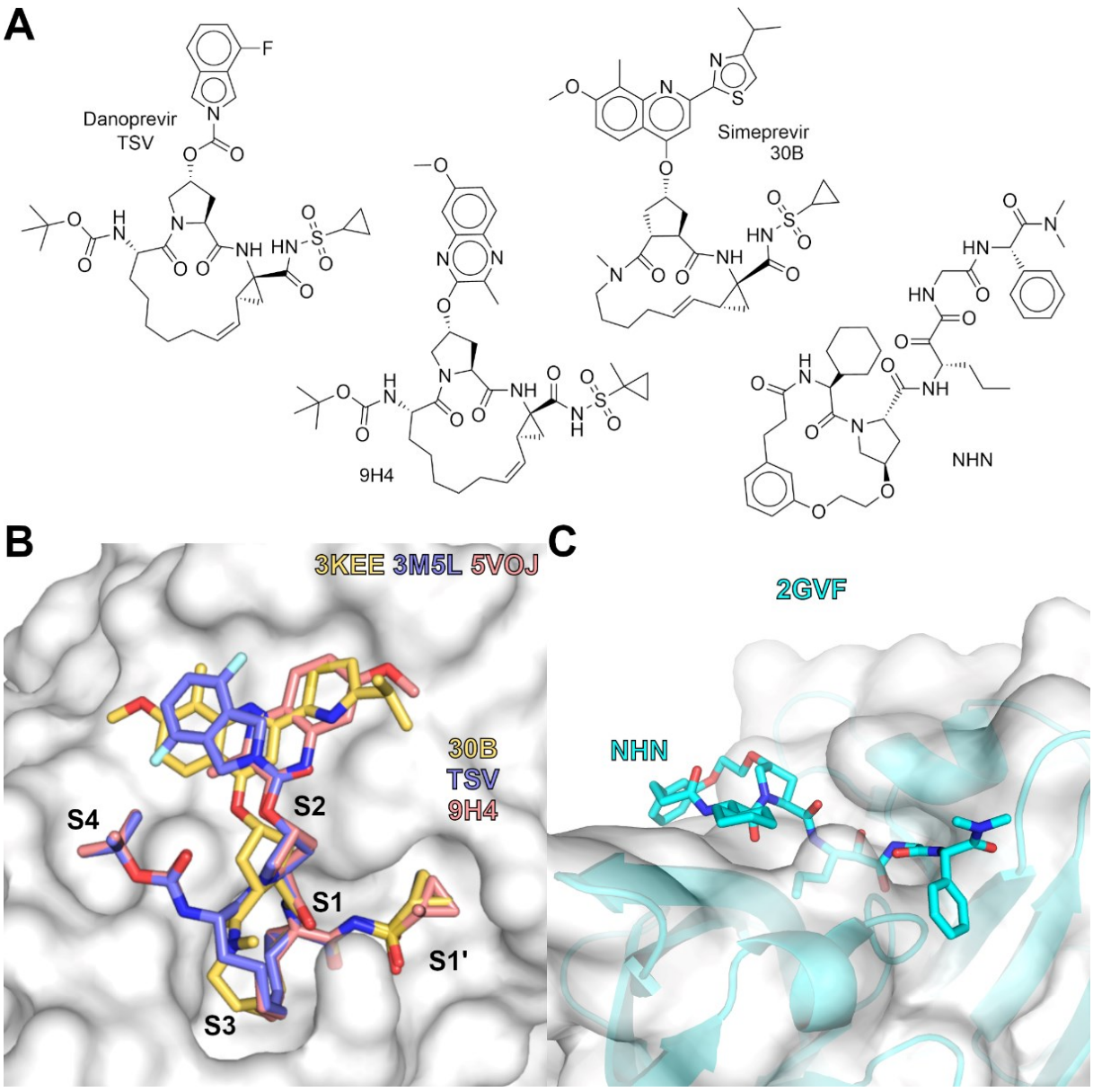

Figure 4. A) 2D structures of HCV NS3/4A protease inhibitors. **B)** Crystal structures of MC compounds 30B (PDB-ID: 3KEE), TSV (3M5L) and 9H4 (5VOJ) bound to the shallow surface-exposed binding site of the protease. Sub-sites on the protein surface are labeled according to Berger *et al*[39]. **C)** Crystal structure of bound NHN (2GVF), further demonstrating the sophisticated binding mode, stretching over large areas of the protein surface.

drugs. These MCs block the tunnel of the ribosomal peptidyl transferase of various bacteria through which nascent peptides exit the complex. By binding to the inner surface of its entrance, the pore is narrowed, and the release of synthesized peptides is prevented. This surface-modifying steric inhibition is distinct from the mode of action of non-cyclic inhibitors, which directly interact with the peptidyl transferase cavity[40].

Another aspect contributing to the therapeutic potential of some MCs is their decent oral bioavailability. Despite violation of Lipinski's rule-of-5, a guideline for small drug-like molecules to avoid poor absorption and permeation[41], MCs can exhibit enhanced passive membrane permeability. What might seem counterintuitive at first can be attributed to the ring structure that changes conformation depending on the environment. In- and outside of the cell, hydrogen bond donors and acceptors are exposed to the polar surrounding, but form intramolecular hydrogen bonds when entering the lipid bilayer of the membrane, thus shielding polar functional groups[2]. Linear peptides, albeit great therapeutic potential, suffer from poor membrane permeability since their backbone amides cause a large conformation-averaged polar surface area. Furthermore, they are prone to rapid degradation inside the cell, unlike their cyclized counterparts that often exhibit proteolytic stability[42]. The peptidic MC cyclosporin A (CsA, **Figure 3D**) produced by the fungus *Beauveria nivea*, is a membrane permeable immunosuppressive drug targeting the intracellular *cis*/*trans* peptidyl-prolyl isomerase cyclophilin. Many

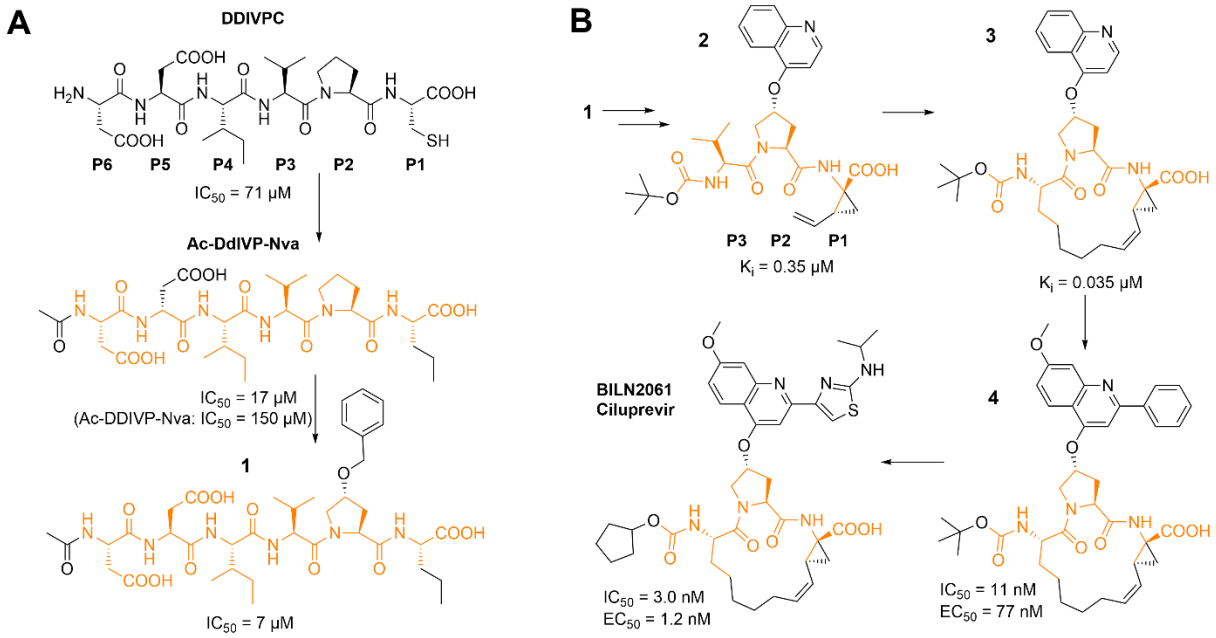

Figure 5. Drug design from **A)** linear product-based peptides to **B)** MC inhibitors of HCV protease NS3/4A. Structural remains of the original DDIVPC peptide highlighted in orange. P1-P6 label moieties that bind to the corresponding subsites on the protease surface (see **Figure 4B**). Values for binding affinities and potencies according to Tsantrizos *et al.*[49], Llinàs-Brunet *et al.*[51] and LaPlante and Llinàs-Brunet[52].

amide groups in CsA are N-methylated, while the remaining ones participate in intramolecular hydrogen bonds across the MC ring backbone in apolar environment[43]. This effect is known as chameleonicity and was more recently found in non-peptidic MC compounds[44]. Thus, new structural and physicochemical guidelines are needed for the chemical space beyond the rule-of-5 in order to purposefully design MC drug candidates with chameleonic properties[37].

BILN2061, a predecessor of TSV and related inhibitors (**Figure 4A**), is a great example for the design of a MC drug from a linear precursor (see **Figure 5**), that illustrates how macrocyclization improved pharmacodynamic and -kinetic properties. Initially,

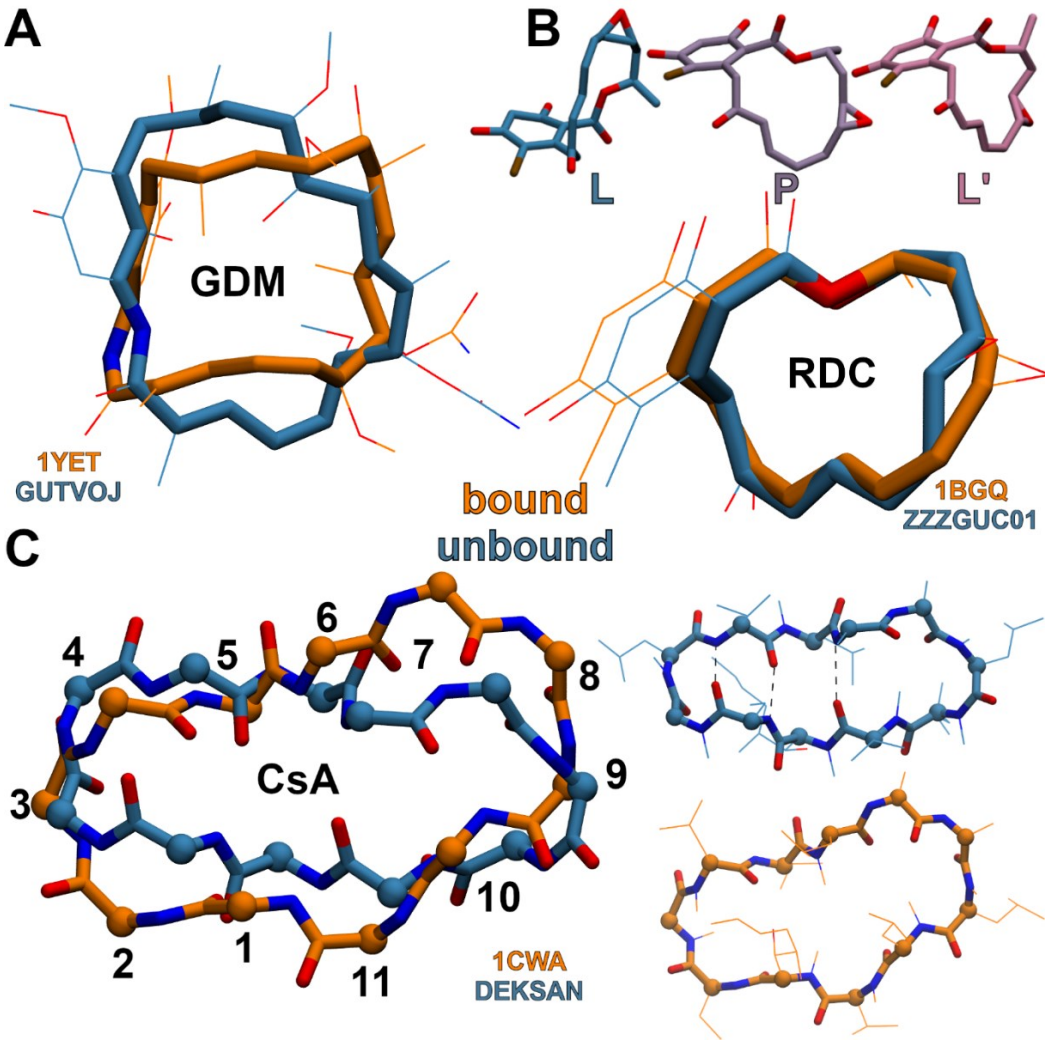

Figure 6. Conformational flexibility of MC ring structures comparing bound (orange) and unbound (cyan) crystal structures with their PDB and CSD identifiers, respectively, aligned on heavy atoms of the ring. **A)** geldanamycin (GDM). **B)** radicicol (RDC). Additionally, known unbound conformers L, P and L' according to Moulin *et al.*[59], here qualitatively reproduced by Meixner *et al.*[104] in similar *in silico* studies. **C)** Cyclosporin A (CsA) aligned on backbone heavy atoms only showing the backbone (left) with the Cα-atoms as spheres and their corresponding residue number as in **Figure 3D** for orientation. Both conformations are separately shown on the right, additionally with their side chain heavy atoms as lines. Characteristic intramolecular hydrogen bonds in the unbound conformer (referring to the most dominant conformation in apolar solvent) indicated by dashed lines.

compound design was based on the N-terminal product peptide DDIVPC (**Figure 5A**), which inhibited HCV activity[45]. First modifications included replacement of the chemically active cysteine P1 residue by norvaline (Nva) and addition of P6 acetyl (Ac) and D-Asp P5 (d) (Ac-DdIVP-Nva)[46]. To further reduce the peptidic character of the compound, investigations focused on side chain specific modifications, which revealed that benzylmethoxy-proline at P2 significantly improved potency (**1**). NMR experiments of free and bound inhibitors could show that the P1-P4 region established most contact with the protein, while sites ≥P5 seemed to be solvent exposed in the bound state, suggesting N-terminal truncation into smaller peptides[47, 48]. Additionally, molecular docking studies of **1** provided more structural detail about the shallow binding surface and the extended binding conformation of the compound. Thus, the binding mode could be further analyzed, confirming that the P1 side chain bound to the only sub-pocket in close proximity to the P3 site, which only indirectly contributed to binding, while the P2 benzyl moiety covered large portions of the protein surface[49]. In line with segmental flexibility sensing NMR spin relaxation experiments, it could be concluded that the peptide backbone conformation is extended in both bound and unbound state, while the side chains, especially in the P1 site, show significant loss of flexibility upon binding[47]. Since then, the design focused on reasonably rigidifying this region of the compound (e.g., P1 in **2**) to fix the observed binding mode and thereby increase potency due to lowered entropic costs. One attempt was intramolecular macrocyclization of a tripeptide via P1 and P3, exploiting their spatial proximity, to conserve the rigid side chain character of P1 and pre-determine a crucial *trans* P2-P3 backbone geometry, thereby locking the bound conformation (**2** → **3**), yielding a 10-fold increase in potency[49]. Conformational search calculations have shown that different moieties at the P2 proline site occupy distinct areas of conformational space, suggesting that a combination of aromatic ring systems would cover larger areas of the protein surface while concomitantly increasing the rotational barrier of the dihedral angle around the oxyl-prolyl bond, further reducing the compound's flexibility, and leading to the optimized compound **4**[50]. The crystal structure of the HCV NS3/4A protease in complex with **4** confirmed the successful rigidification supported by NMR data, which proved similar conformations of the ligand in bound and unbound state[49]. Based on **4**, cellular assays were performed to determine inhibitory activity in cell cultures. While this compound exhibited promising potency, further empirical optimizations of the N-terminal capping group and the aromatic P2 moiety eventually led to BILN2061. Due to the unique structure, BILN2061 is a potent and

competitive inhibitor *in vitro* and in cellular context, exhibiting decent oral bioavailability and metabolic stability in rats. Additionally, target specificity was confirmed by the lack of activity against multiple human serine and cysteine proteases[49, 51, 52]. With these promising drug-like properties, BILN2061 could be investigated in clinical trials, where oral administration proved efficient reduction of plasma viral load after 48 hours[53]. Since its discovery, BILN2061 became a prominent parent compound for additional ligand-based drug design studies, resulting in a family of potent inhibitors, some of which became approved HCV drugs. Moreover, this example elucidates the powerful potential of combining different methods such as extensive structure-activity relationship studies, NMR and X-ray crystallographic experiments, as well as molecular modeling investigations, for the successful design of MC inhibitors.

Although MCs are structurally constrained by their ring scaffold, they exhibit complex flexibility. In general, drug-like inhibitors often change conformation upon binding. Together with the concomitant structural adaption of the protein, this mechanism is known as induced-fit[54]. Thereby it was discovered that the bound state of the ligand is often different from the lowest energy conformation (global energy minimum), and even different from a local energy minimum state[55]. Such costly rearrangements and the associated loss of entropy of the ligand are compensated by favorable interactions of the complex[56]. **Figure 6** shows abovementioned MC ligands, for which multiple conformations were observed experimentally. GDM is a prominent example where the bound and unbound conformations are significantly different, as observed by the corresponding crystal structures (**Figure 6A**). While the unbound form develops a flat ring scaffold with a *trans* amide, the bound conformation is folded into a C-clamp shape (see also **Figure 2D**) and exhibits a *cis* amide bond[10, 57]. RDC binds as the lowest energy conformer in its *L* shape[13, 58]. Correspondingly, the bound and unbound conformations shown in **Figure 6B** are similar. A study of RDC analogs reported one inactive isomer with inverted stereochemistry of the epoxide group[59]. Although this derivative obtained the overall same ring conformations in solution (*L*, *P* and *L'* shape), the bioactive conformer was different from the lowest energy state for that species. This finding highlights the importance of the ring conformation for binding, which can be thus pre-determined by rigid proximal functional groups. Conformations of CsA (**Figure 6C**) were extensively studied over the years using NMR and X-ray crystallographic experiments as well as *in silico* approaches[60-63]. They are often described as *closed* and *open* depending on the environment. In nonpolar solvents (*closed*) the backbone of

residues 1-7 and 11 forms a twisted β-sheet with intramolecular hydrogen bonds across the ring involving all non-methylated amide protons, and residues 3 and 4 build a type II' β-turn. The remaining residues form a loop with a characteristic *cis* peptide bond between residues 9 and 10. In polar solvents, multiple conformations exist that are described as *open*, like the cyclophilin bound state, that exhibits all-*trans* peptide bonds and no intramolecular hydrogen bonds. As illustrated above, this conformational polymorphism is linked to the membrane permeability and oral availability of CsA. These examples prove that MCs are capable of significant rearrangements and conformational changes, albeit their structural constraints. In X-ray crystallographic studies, a protein-MC ligand complex, that knowingly forms by induced-fit mechanism, could be crystallized well only after incubation at 50°C for 10 minutes, likely due to the rigidity of the MC ring[64]. Another extensive NMR study proved that MC ligands with smaller ring sizes and associated higher rigidity obtained slower conformational exchange rates. One example MC, the crystal structure of which only showed one conformational species, resulted in more rapid exchange rates and NMR signal coalescence with increasing temperatures[65]. In these experiments, high temperatures were used to induce conformational changes of the MC ring, thereby evidencing increased energy barriers for interconversion between ring conformers.

1.3. Computational challenges for molecular modeling of macrocycles

The accurate structural representation of MCs is crucial for reproducing and predicting dynamic properties regarding many theoretical endpoints, such as conformational analysis, molecular docking, pharmacophore modeling, 3D-QSAR studies, physicochemical characterization, thermodynamic calculations, ligand based virtual screening and similarity searches[66]. However, the energetic increase of interconversion between ring conformers and the interconnectedness of ring torsions poses a unique challenge for structural approaches.

1.3.1. Conformational sampling

Many conformational sampling tools aim for the optimization of different parameters or outcomes of the method such as reproducing bioactive conformation, including the global minimum structure, sampling high quality and low-energy conformations, covering diverse areas of the conformational landscape or generating large numbers of conformers for fast screening of many different compounds. Historically, most conformational sampling tools were originally developed for small organic compounds and validated on

data sets and data bases including typical drug like molecules. In principle, the general workflow of conformer generation tools starts with identifying rotatable bonds of the molecule. Different conformations are then generated by changing dihedral angles of such torsions according to the implemented algorithm. Most tools follow the rigid rotor approximation, under which bond distances and angle values are kept fixed for simplicity. Often, generated conformations are then ranked, filtered or clustered to select a final set of representative structures[67].

Table 1 provides insights into some of the most popular conformational sampling tools. This list is by no means complete but gives a general overview of some common algorithmic strategies used to sample conformations of a molecule and includes the primary goal, for which the tool was originally designed. Various reviews thoroughly compared the performance of those and other tools[66 , 67, 84-88].

Torsions of ring moieties are usually identified in the first step of the workflow and, by most tools, excluded from dihedral modifications due to their interconnectedness, to

Table 1. Incomplete list of common 3D conformer generation tools.

CATALYST[68]	quasi-systematic: *fast*	fuzzy grid + FF[a] refinement	building conf. libraries	template library (≤ 9 atoms)
	Numerical: *best*	DG[b] + FF[a] refinement + poling[69]	large coverage	
CAESAR[70]	systematic	divide-and-conquer + recursive buildup	diversity, low energy	template library (DG[b])
MOE[71]	stochastic	Random Incremental Pulse Search[72]	detailed conf. analysis	
	systematic	all comb. of tors. angles (discrete intervals)		
	systematic: *Conformation Import*	buildup from fragment library	screening databases	
CORINA[73] **ROTATE**[74]	systematic	rule- and data-based	screening databases	template library (≤ 9 atoms)
OMEGA[75]	systematic	knowledge-based	high speed	template library
Balloon[76]	stochastic	DG[b], GA[c]	diversity, low energy	flip-of-fragments[77] for small rings
iCon[78]	systematic	knowledge-based	screening databases	initial generation by DG[b] and FF[a] refinement
Confab[79]	systematic	torsion-driven using compound library[80]	large coverage, low energy	none
ConfGen[81]	deterministic	physics-based	small subsets + bioactive state	template library
Macro-Model[82]	LMOD[d,][83]	eigenvector-/ vib. mode-following	exhaustiveness	explicit

[a] force field, [b] distance geometry, [c] genetic algorithm, [d] low-mode search

prevent the generation of unphysical ring structures. Especially for systematic searches, this lack of ring flexibility is often compensated by using a pre-defined set of possible ring conformations from a generated or existing template library. However, such libraries only cover the most essential small ring structures and fail to predict multiple conformations for more complex rings like MCs. From the tools in **Table 1**, Balloon incorporates a separate flip-of-fragment algorithm to generate multiple ring conformations. However, if a molecule has many ring systems, the performance easily suffers due to large memory requirements and poor complexity (increasing with the second power of number of ring atoms). By default, the number of possible flip-of-fragment operations is thus limited to a small number, even though the authors exemplarily show the generation of one MC conformer ensemble which includes a conformation similar to the bioactive state[76]. The LMOD strategy of MacroModel significantly differs compared to the other tools mentioned in **Table 1**, employing a normal mode analysis and following the low-frequency modes reaching different minima on the potential energy hypersurface[83].

Table 2. Incomplete list of MC-specific 3D conformer generation tools or general tools applicable to MC compounds.

tool	type	algorithm	goal	ring split
MMBS[89][a]	LMMD[b] + LLMOD[c]	stochastic MD[d] + eigenvector-following	low energy	no
BRIKARD[90][e]	enumerative	algebraic geometry, inverse kinematics; breadth-first	diversity, exhaustiveness, speed	no
ForceGen[f] **3D**[91]	deterministic	physics-based	generality, accuracy	no
Prime-MCS[92][g]	stochastic	coarse-grained rotors, dead-end elimination	diversity, speed, low energy	yes 2 linear fragments
ConfBuster[93]	systematic	all comb. of tors. angles	low energy	yes 1 linear fragment
CCDC[h] **conf. generator**[94]	stochastic	knowledge-, rule- and data-based; depth-first; incremental construction	small molecules, speed	yes, if ring is not found among templates
Confor-mator[95]	systematic	knowledge-based, incremental construction	accuracy, low energy	yes, if more than 9 ring atoms
Moloc[96]	shape-guided	Fourier-based	diversity, accuracy, screening	no

[a] MacroModel Baseline Search; [b] LowModeMD method; [c] large-scale LMOD; [d] Molecular Dynamics; [e] Builder for Recursive Inverse Kinematic Assembly and Ring Design; [f] Force Field Based Conformational Generation; [g] macrocycle conformational sampling; [h] Cambridge Crystallographic Data Centre

This does not require a discrimination between cyclic and acyclic portions of the molecule. However, due to hydration and solvent accessible surface area models and many different FFs implemented, this approach is computationally more costly.

Designed with a focus on small molecules, it has been shown that such tools are indeed inapplicable to MC compounds[97]. Furthermore, recent publications on the exploration of the unique conformational space of MCs highlighted the consistent need for advanced methods[98]. Without MC-specific techniques, the promising therapeutic potential of these molecules remains under-explored[99]. **Table 2** shows an overview of tools specifically developed for MCs within the last 8 years. A general distinction can be made between methods that split the MC ring and sample the linear part(s) before re-fusing, and those that keep the ring as is to generate conformers ("ring split" column). Some algorithms employ similar strategies as in **Table 1** with an adaptation for larger rings (e.g., MMBS, ConfBuster, Conformator, etc.). Others follow a more MC-guided strategy (e.g., BRIKARD, Moloc).

Compared to systematic or stochastic tools, molecular dynamics (MD, see chapter **2.1**) is another popular approach for conformational sampling. Simulating the molecular movement according to the atomic forces present at given conditions allows for studying the dynamics of the system including changes between different ligand conformations. Often for MD simulations of MCs, elevated temperatures are applied trying to overcome the increased interconversion barrier between ring conformations[100-104]. Some studies also used meta-dynamics[105] (MTD) or quantum mechanics[106] (QM) for this issue. Substantial reviews compared conformer generation strategies for MCs or included macrocycle data sets for evaluation[97, 103, 107-109]. Furthermore, specialized tools for cyclic peptides emerged[110-112], but were excluded from **Table 2**, since only non-peptide MCs were studied in this work.

1.3.2. Molecular docking

Here, we concentrate on protein-ligand docking for the prediction of a bound state of drug-like MC inhibitors and their host proteins. More recent comprehensive overviews on molecular docking were provided for example by Pagadala *et al.*[113] and Wang *et al.*[114]. A typical docking calculation consists of a pose sampling and scoring step. In the former, the rotational, translational and conformational degrees of freedom (DoFs) of the

Table 3. Incomplete list of popular docking programs giving their degree of incorporated ligand and protein flexibility, the treatment of large ring conformations, as well as reference to studies where this approach was applied to molecular docking of macrocyclic molecules.

program	ligand flexibility[a]	MC ring conf.	protein flexibility[a,b]	MC study
DOCK 6[115]	anchor-and-grow incremental construction	-	-	[116]
AutoDock 4[117]	GA[c]/Solis-Wets search	on-the-fly	limited side chains rotations	[116, 118]
AutoDock Vina 1.2.0[119]	Monte-Carlo/BFGS[d]	on-the-fly	limited side chains rotations	[116, 120]
Glide[121][e]	exhaustive rotamer pre-screening	templating	IFD protocol[122][f]	[99, 116]
GOLD[123][g]	GA[c]	corner flipping	limited side chains rotations	[124]
ICM[125][h]	Monte-Carlo search	on-the-fly	limited side chain refinement	[126]
MOE-Dock[71]	simulated annealing, tabu search	-	limited side chain refinement	[127]

[a] referring to the pose sampling step; [b] other than using a target structure ensemble; [c] genetic algorithm; [d] Broyden-Fletcher-Goldfarb-Shanno method; [e] grid-based ligand docking with energetics; [f] induced-fit docking protocol only in combination with Prime, limited to small backbone minimizations and side chain conformational changes; [g] Genetic Optimization for Ligand Docking; [h] internal coordinates mechanics

ligand are varied to generate different potential binding poses. For the latter, a score calculated by an energy function evaluates the strength of each pose and provides a rating for selecting the most likely. Since this work focused on structural accuracy during pose sampling, only the former step is introduced in detail. In principle, docking calculations can be categorized according to the level of molecular flexibility incorporated during pose generation. In rigid body docking, only the relative orientation but not the conformations of both binding partners change. Such calculations were based on the lock-and-key binding model, which assumes a geometric fit between the interacting partners without conformational changes upon binding. Flexible ligand docking or semi-flexible calculations sample different conformations only for the ligand. A fully flexible approach incorporates the highest level of flexibility, where both ligand and receptor conformational DoF are modified during pose sampling[128]. While by now most docking approaches follow a flexible ligand strategy, accounting for proper protein flexibility is still challenging due to the large number of conformational DoF. However, the consideration of conformational changes on the protein site is crucial for the accurate prediction of dynamic binding effects such as induced-fit and conformational selection phenomena, that involve structurally distinct bound (holo) and unbound (apo) protein states. Thus, many sampling algorithms with varying degrees of receptor and ligand

flexibility exist. **Table 3** lists features of commonly used docking programs. Ligand placement is performed employing either fragmentation and incremental construction, shape-based fitting, evolutionary algorithms, systematic or stochastic searches. Conformational flexibility of the ligand is thereby considered concomitantly by varying rotational DoF during the search or, e.g., in the case of Glide, *a priori* by sampling different ligand conformers ("ligand flexibility" in **Table 3**)[113]. For MC compounds, the conformational ring sampling problem introduced before, extends to the pose sampling step of flexible ligand docking calculations. Some docking tools attempt to internally generate ring conformations ("MC ring conf." in **Table 3**) on-the-fly by opening the ring, independently modifying its torsion angles and applying a linear potential forcing ring closure during docking(AutoDock, AutoDock Vina)[119]. A similar approach implemented in ICM uses internal coordinates (bond distances, bond angles and dihedral angles) converting the ligand into a tree representation. Ring structures are read linearly as if one ring bond is broken, allowing to modify the remaining internal (ring) coordinates and using that "broken" bond as a closure constraint[125]. Alternatively, docking tools use template libraries (Glide) or a corner flipping algorithm[129] (GOLD) to generate MC ring conformers. However, as this is still an active field of research, some docking tools simply ignore (MC) ring torsions and treat such moieties rigidly. To compensate this artifact, many MC docking studies perform an external conformational sampling step first, for example with MC-specific methods introduced in **Table 2**, to generate a set of ring conformers that is used as a starting ensemble for subsequent docking calculations, thereby indirectly accounting for ring flexibility[130].

Independent of the type of ligand, receptor flexibility is only incorporated by some docking programs ("protein flexibility" in **Table 3**) and often limited to side chain rotations[128]. Depending on the algorithm, torsional DoF of side chains are modified, e.g., by using a rotamer library during ligand pose search (AutoDock, AutoDock Vina, GOLD) or after an initial docking run by refinement (ICM, MOE-Dock) or re-placement and minimization (Glide), allowing the amino acid residues to adapt to the presence of the docked ligand, thereby simulating induced-fit binding. However, this approach, often called selective docking, is still inapplicable when the conformational changes upon binding exceed simple side chain movements. Exemplarily demonstrated for the backbone of the L2 loop in proximity to the nucleotide binding site of Hsp90 (see **Figure 2B**), that shows a displacement of up to 6.0 Å between holo and apo states[10]. Furthermore, this active site is able to host distinct (MC) ligands with varying binding

modes, emphasizing its structural adaptability[37]. Thus, neglecting protein flexibility for such systems can be fatal for molecular docking calculations, especially for docking into apo structures[131]. In the following, different strategies of flexible receptor docking are elaborated since this poses a major challenge for molecular modeling.

Ensemble docking accounts for full receptor flexibility by generating and docking into a set of different protein conformations, like using a starting ensemble of MC ligands to compensate the lack of ring flexibility as mentioned above. Such approaches typically follow the conformational selection hypothesis and rely on experimental structures (X-ray, NMR) or computational methods (homology modeling, MD). The ligand is then docked to all structures of the receptor ensemble, either in multiple runs, leaving the actual docking algorithm unaltered, or in a single run by combining the ensemble structures into a single model, e.g., for grid-based methods[132]. However, ensemble docking only accounts for protein flexibility implicitly since the single receptor conformations are still treated rigidly during docking. Moreover, multiple run approaches increase the risk of generating false positives and the computational costs compared to using a single receptor structure. However, this strategy could improve docking results for difficult targets[133], but success is highly dependent on the receptor ensemble, for which no clear guidelines are developed yet[134].

Among the important variations of this method is 4D docking, which uses a discrete variable that determines the current receptor conformation, that is changed similarly as the ligand DoFs rather than sequentially docking the ligand to all receptor conformations[135]. In contrast to that, composite structure docking creates a new receptor structure based on discrete alternative conformations among the structures in the ensemble, as implemented in the FlexE method[136].

Approaches specifically focusing on backbone conformational changes were introduced in RosettaLigand[137], MedusaDock[138] and BP-Dock (Backbone Perturbation-Dock)[139]. In the former, docking is performed in three steps from coarse-grained to high resolution, where receptor backbone flexibility is only incorporated in the last step. First, the ligand is placed based on shape compatibility. Secondly, the ligand and protein side chains are optimized using Monte-Carlo minimization and, lastly, a gradient-based minimization is performed for the whole ligand and protein side chain and backbone torsions. However, the authors reported average backbone movements of 0.45 Å, only corresponding to minor local conformational changes[137].

In MedusaDock, a sequential ensemble dock is performed for pre-constructed ensembles of receptor backbone conformations. However, the ensembles are based on small numbers of experimental structures. Although the authors claim to thus cover a wide variety of backbone conformations, they are biased and limited to the available PDB structures. For example, their urokinase system was represented with only one backbone structure, since all known bound receptor conformations were similar[138].

In contrast to that, BP-Dock generates the receptor conformations of the ensemble by perturbing protein residues with Brownian kicks (random external forces applied to Cα-atoms). Perturbation response scanning is used to calculate a fluctuation profile of the backbone, followed by energy minimization. New receptor conformations are thus created that potentially correspond to ligand-induced conformational perturbations. The promising study included unbound protein structures for evaluation and claims to enable large-scale backbone movements. However, the actual docking is performed with RosettaLigand and, even without additional backbone flexibility, reproduced the binding positions already quite well in terms of structural accuracy. Thus, the BP-Dock protocol could only show minor improvements in these cases[139].

As introduced above, including sufficient receptor flexibility is still a challenging task for molecular docking calculations, limiting their applicability. To tackle this issue, MD simulations are often used to support such calculations and either implemented in the docking algorithm or applied as a separate method in a pre- or post-processing step.

An example strategy, where MD was incorporated before and after the actual docking calculations is illustrated by works of Lin *et al*[140]. They introduced the relaxed-complex scheme, where ligand libraries were sequentially docked in many conformations of the unbound FKBP (1FKK, see **Figure 3C**) pre-generated with MD. While this improved the accuracy of the docked poses of two investigated ligands, the AutoDock scoring function failed to identify them. The docked complexes were thus re-scored in a follow-up study by Molecular Mechanics/Poisson-Boltzmann Surface Area (MMPBSA) calculations, which successfully identified the accurate poses as the complexes with the lowest free energies[141]. This study further highlighted the possible decrease in specificity when using generated protein ensembles. The simulated conformational changes in the active site might accidentally allow for the placement of non-binding ligands. At the same time, docking into receptor ensemble might identify new binding modes that would be missed when docking into a single static crystal structure[142]. Studying the same protein, Zacharias *et al.*[143] used MD and principal component analysis (PCA) to extract

collective DoFs (soft modes) of the apo state as possible receptor deformations during docking. It could be shown that simulating receptor flexibility in that way improved the docking results for the MC compound FK506.

A method, where MD is directly implemented into the docking algorithm is DynaDock[144]. During ligand placement in the active site, a user-defined overlap of protein and ligand atoms is allowed, that is subsequently refined by optimized potential MD (OPMD). In this refinement simulation, the non-bonded interactions (van der Waals and electrostatics) between receptor and ligand atoms are modified with the soft-core parameter α (see chapter **2.2**). Depending on the initial overlap, the value of α (between 0 and 1) is automatically determined and optimized over the course of the refinement until the overlap was fully resolved. Thereby, a value of 1 turns off non-bonded interactions completely and 0 corresponds to the standard physical potentials. In that way, the initial clashes during placement can be refined without causing strong repulsive forces, while the receptor atoms can smoothly adapt to the presence of the ligand, thereby explicitly accounting for full receptor flexibility.

Furthermore, MD is used in post-processing to investigate the stability of the docked complexes and solvation effects[145]. Since only true binding poses would result in a stable trajectory, such simulations help to identify the correct docking pose[146]. Studying the dynamics of binding poses of different ligands, MD simulations provide a structural explanation for experimentally observed differences in inhibition[147, 148].

Aiming for a fully dynamic docking approach, pure MD studies were conducted studying ligand binding. Among popular methods are enhanced MD techniques that introduce an additional bias in form of collective variables (MTD) or tempering (replica exchange MD, REMD; accelerated MD; multicanonical MD). In contrast, unbiased approaches include brute-force MD and discontinuous methods. These have been substantially reviewed before[149]. Here, they are not discussed in detail since this would exceed the scope of this introduction. Instead, a comparison to static docking is drawn in the following. Dynamic docking methods retrieve information about the mechanism of ligand binding. Various studies of different methods could not only reproduce the known ligand binding poses, but calculate binding-free energies[150] and association constants[151], identify unknown (allosteric) binding sites[152] and predict new binding mechanisms[153], as well as study desolvation effects, intermediate states and associated energy barriers[154]. Thus, such methods not only investigate the final binding pose but also the dynamic binding process and its observables. The major disadvantage of all dynamic docking

methods, however, is the high computational cost due to long simulation times. Static docking is thus more useful for screening large databases of ligands in moderate time, while more expensive dynamic docking methods are usually applied to study few compounds in more detail. Both types of methods can be complementary and are often performed together in drug discovery pipelines or as hybrid approaches.

Summarizing, the treatment of the MC ring during conformational sampling and docking pose generation is still an issue for computational methods. Additionally, although not limited to MC docking, the proper account of receptor flexibility during molecular docking calculations remains challenging.

1.4. Aims and structure of this study

The aim of this work was to develop a generally applicable pipeline for the accurate prediction of binding modes of MC compounds. In a previous study[102], an MD-based sampling and subsequent molecular docking protocol was successfully tested for 7 small MC inhibitors and their holo proteins. Our goal was to increase the data set to a larger number of diverse MC ligands, adjust and optimize the sampling and docking parameters of the pipeline and evaluate the new strategy in a benchmark-like study by docking in holo and apo protein states.

Therefore, we first explored the ligand conformational sampling step. In a first study of the total synthesis of the NP tetrandrine and its analog iso-tetrandrine[155] (chapter **3.1**), we computationally investigated MC intermediate structures of the final reaction of the pipeline. Long-term MD simulations performed under experimental conditions and statistical analysis of thus sampled pre-reactive conformational states were performed. From this, the ratio between conformers likely to result in tetrandrine and iso-tetrandrine could be assessed, which reproduced the experimentally observed product ratios.

The in-depth conformational analysis based on dihedral angles of MC ring torsions led to the development of a torsional classification scheme for MD-simulated MC conformers. With this, the structures obtained from simulations can be classified according to the collective distribution of sampled dihedral angles of MC ring torsions and thus reduced to a subset of ring conformers. This was the basis for two subsequent case studies, where ensembles of thus generated conformer subsets were docked, thereby exploring possible modifications of the docking step of the pipeline. In the first study, acyldepsipeptides were docked to caseinolytic protease P (chapter **3.2**)[147]. The second study investigated the binding mode of an actin-binding miuraenamide derivative (chapter **3.3**)[156]. In both

cases, our pipeline could successfully identify binding poses that explained experimentally observed differences compared to their parent compounds.

The knowledge and ideas gained from these studies were combined for the design of the final MD-based sampling and docking protocol, which was applied to 20 structurally diverse MC ligands and their host proteins in a benchmark study(chapter **3.4**)[104]. We thereby provided a robust pipeline for the accurate prediction of binding modes of MCs with DynaDock[144], overcoming major structural challenges of molecular modeling of MCs.

In the following, the theory behind the methods used throughout these studies is explained (chapter **2**) and the results are summarized in more detail (chapter **3**). A comprehensive discussion of the most important aspects of the computational work of all studies is presented (chapter **4**), followed by a conclusion (chapter **5**).

2. Theory

In the following, the theory behind the methods used throughout this work will be elaborated. This section will thus cover molecular dynamics and molecular docking with DynaDock. Regarding the former, more details can be found in standard works of J. Andrew McCammon and Stephen C. Harvey[157] ("Dynamics of proteins and nucleic acids"), Charles L. Brooks, Martin Karplus and B. Montgomery Pettitt[158] ("Proteins: A theoretical perspective of dynamics, structure and thermodynamics") and the substantial work of Tamar Schlick[159] ("Molecular Modeling and Simulation: an interdisciplinary guide"), the concepts of which shaped this section. For computational details and settings of employed programs, the material and methods sections of the individual publications contributing to this work provide the necessary information, and are therefore not listed here again.

2.1. Molecular Dynamics (MD) Simulations

MD constitutes a powerful theoretical method that describes the dynamics of a molecular system according to the physical forces that act on it under given conditions. These forces are obtained as the first derivative of the systems potential energy, which can be expressed as a function of atomic coordinates. The potential energy of a system is determined for a certain molecular geometry, the corresponding forces are obtained and, with classical equations of motion, the change of atomic positions with time, and thus the molecular motion, can be calculated. Simulating motion provides information about dynamic properties of the molecular system including changes in conformation and energy, which, with laws of statistical thermodynamics, can be related to macroscopic observations, thereby linking theory and experiment. For biomolecules, MD is an essential tool that supplements experimental findings. For example, simulating protein structures resolved by X-ray crystallography, which only resemble a static picture of one specific state, provides insights into structural transitions between thermally accessible states. Furthermore, the binding of a non-covalent inhibitors can be studied with MD by analyzing intermolecular interactions. Therefore, MD simulations unravel the complex and dynamic behavior of (macro)molecules, and thus help to better understand the relationship between structure and function to comprehend their biological activity.

2.1.1. Primary considerations

An accurate description of the motion of any molecular system is provided by the Schrödinger equation (1), that employs the Hamiltonian operator $\hat{H}$, which consists of kinetic E_k and potential energy E_p. In the notation of equation (2), these depend on the momentum $\tilde{P}$ and position vectors $\tilde{X}$ of all nuclei and electrons, respectively.

$$\hat{H}\psi_n = E_n\psi_n \tag{1}$$

$$\hat{H}(\tilde{P}, \tilde{X}) = E_k(\tilde{P}) + E_p(\tilde{X}) \tag{2}$$

The exact solution of this equation, however, is not feasible for large molecules. To simplify this expression, the Born-Oppenheimer approximation separates the motion of electrons from the motion of nuclei. Since electrons are lighter and, thus, faster, they instantaneously adapt to new positions of the heavier and slower nuclei. High level QM approaches provide *ab initio* or semi-empirical solutions under the Born-Oppenheimer approximation. For the size of biomolecules, these calculations are often still too demanding for practical applications. An even more approximated approach uses the principle of molecular mechanics (MM). In MM, the electronic motion is neglected, thereby reducing the number of variables, and resulting in a function for the potential energy that is only dependent on the position of nuclei. Since electrons are not explicitly accounted for in that approximation, changes in electronic structure, like electronic excitation, chemical bond breaking and formation, rearrangements of charge distribution and polarization effects cannot be modeled. Instead, the system is simply described as a mechanical body, where atoms are treated as point masses with point charges centered at the position of the nuclei, that move according to physical forces. Thus, the potential energy can be calculated by empirical functions that describe the molecular constitution and intra- and intermolecular interactions with classical potentials.

2.1.2. Principles of molecular mechanics

The MM approach is based on the principles of additivity and transferability. The former states that the potential energy of a molecular system is a sum of separable potentials derived from bonded and nonbonded mechanical forces. Bonded potentials include terms describing covalent bonds, bond angles and dihedral angles. Nonbonded terms include van der Waals and electrostatic interactions between pairs of nonbonded atoms. Transferability assumes that the potential functions developed from experimental data of representative model structures can in general be used for the prediction of larger (bio)molecules. For that, the concept of atom types is introduced with which

characteristic variations due to the chemical environment and general hybridization of atoms in a molecule are considered. For example, bond angles are highly sensitive to hybridization and even small deviations – caused by, e.g., lone pairs, ring strains or other chemical environments – can alter the overall geometry of the molecule. With increasing numbers of atom types, these environment dependent variations can be modeled, facilitating an accurate prediction of biomolecular structures.

2.1.3. Force Fields

The functional form of the beforementioned potentials describing bonded and nonbonded interactions is called force field (FF) and can be written as in equation (3)[160]. The individual contributions or terms in the FF are explained in the following.

$$E_p = \sum_{bonds} k_r\left(r - r_{eq}\right)^2 + \sum_{angles} k_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\varphi - \gamma)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] \tag{3}$$

In MM, two covalently bound atoms are described as two spheres connected by a spring. As illustrated by the first term of equation (3), a harmonic potential characterizes the bond stretching motion with a simple functional form. This represents the oscillation of the bond length r around a reference value r_{eq} with the force constant k_r of the respective bond vibration, where r_{eq} and k_r define parameters usually determined by experiment (X-ray crystallography) or higher level (QM-)calculations. The basic concept is deduced from Hooke's law which states that the restoring force $F(x)$ of a spring is proportional to the displacement x and its spring constant k, derived from the mass m on the spring and the resulting angular frequency ω of the oscillation, as shown in equation (4), with the potential energy $E(x)$ given in equation (5).

$$F(x) = -kx \quad , \qquad k = m\omega^2 > 0 \tag{4}$$

$$E(x) = \frac{k}{2}x^2 \tag{5}$$

A more correct description of a chemical bond is provided by the Morse potential[161] shown in equation (6), including the width and depth parameters S_m and D, respectively. A comparison of both potentials is shown in **Figure 7**.

$$E_{Morse}(r) = D\left\{1 - exp[-S_m(r - r_{eq})]\right\}^2 \tag{6}$$

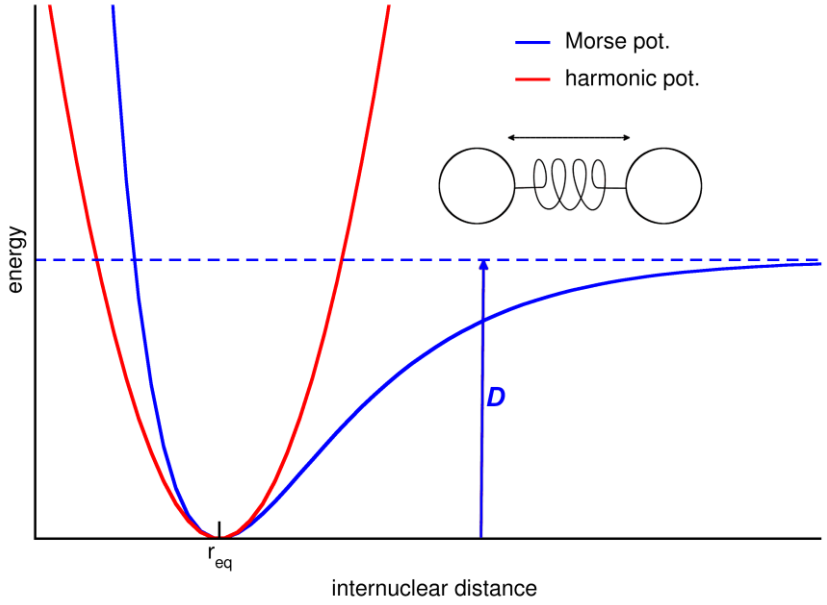

Figure 7. Schematic illustration of the harmonic potential and the Morse potential describing bond stretching motion between two covalently bound atoms. Both potentials are similar near the corresponding reference bond length r_{eq}, but differ for increasing internuclear distance.

Unlike the harmonic potential, the Morse potential reproduces the dissociation of two atoms with increasing internuclear distance expressed by the dissociation energy D in **Figure 7**. However, both potentials are similar for small deviations around the reference bond length. Since most MD simulations are performed at ordinary temperatures, where the atom bond lengths stay close to their equilibrium values, the harmonic potential is sufficient in representing chemical bonds and used by many FFs for its computational simplicity.

The second term in equation (3) considers bond angle interactions occurring between two adjacent covalent bonds. Like bond stretching, this motion is accounted for by a harmonic potential, where the spring is attached to the centers of the two covalent bonds. Correspondingly, the term includes the bond angle θ, its reference value θ_{eq} and the force constant k_θ. Additionally, some FFs include cross terms correlating bond stretching and angle bending motions, i.e., concomitantly adjusting the bond lengths as the angle between these bonds changes. Such correction terms are described, e.g., by so called Urey-Bradley potentials[162] and improve the fit between calculated and experimental vibrational frequencies.

Most crucial for the conformation of a molecule is the rotation around covalent bonds, given by the third term of equation (3). This bonded interaction occurs for the central bond of any four consecutively bound atoms $i-j-k-l$ and is expressed as the corresponding torsion angle (or dihedral angle) between the two planes spanned by the

atoms $i - j - k$ and $j - k - l$, illustrated in **Figure 8A**. For freely rotatable bonds, this motion is periodic and thus described by a periodic functional form employing the cosine function, where φ is the dihedral angle, V_n is the height of the torsional barrier and n and γ denote the periodicity and phase of this torsion, respectively.

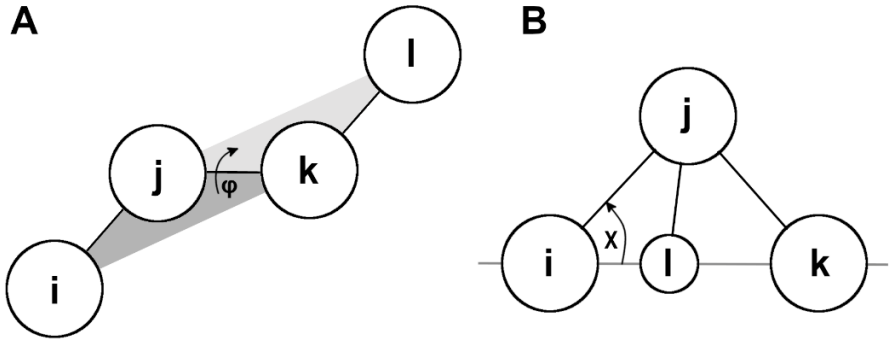

Figure 8. Illustration of a torsion angle φ (**A**) and an improper dihedral angle χ (**B**).

The origin of the rotational barrier of this motion is much debated, but can be attributed mainly to hyperconjugation, rather than pure steric repulsion[163]. The form of the torsional profile is highly dependent on the atoms in the sequence. Thus, the functional form must be able to model various shapes of different torsional potentials. This is achieved by employing a Fourier series of periodic functions as shown in equation (7), where the sum runs over the integer value n. The first three functions of this equation (n = 1, 2, 3) are exemplarily plotted in **Figure 9A**. Through combination of functions with varying n and V_n values, different torsional potentials can be modeled, illustrated by the more complex potential in **Figure 9B**.

$$\sum_n \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)] \qquad (7)$$

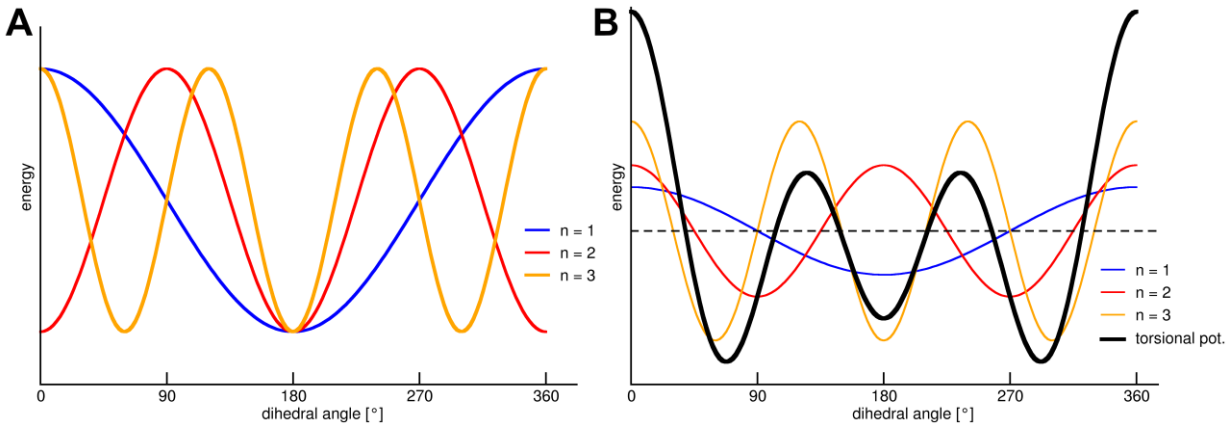

Figure 9. Schematic illustration of periodic functions (**A**) in the form of equation (7) for n = 1, 2, 3 with the same barrier height V_n and a local minimum at 180° for better comparison. Fourier series of such functions used to model more complex torsional potentials in (**B**).

With such series typical *cis/trans* and *trans/gauche* energy differences are reproduced, thereby implicitly accounting for steric, dipole-dipole and gauche effects. To evaluate appropriate parameters, torsional potentials of model compounds are calculated with *ab initio* calculations and, together with nonbonded interactions, fitted to experimental vibrational frequencies. Such parameter sets are then applied to similar atom sequences in larger molecules.

Furthermore, improper dihedral terms can be included in the FF. Illustrated in **Figure 8B**, an improper dihedral angle χ occurs between a plane spanned by three atoms (*i*, *k*, *l*) and a central atom (*j*) which is covalently bound to these. Often called out-of-plane motion, such terms are typically used to enforce planarity and chirality in regions of a molecule and described by a simple harmonic potential as shown in equation (8).

$$E(\chi) = \left(\frac{V'}{2}\right)\chi^2 \tag{8}$$

The last term in equation (3) accounts for nonbonded interactions and consists of two contributions: van der Waals and electrostatic interactions. The sum runs over pairs of atoms *i* and *j*, that are further than 3 bonds apart and within a certain cutoff distance to each other, mainly for computational convenience. Van der Waals interactions occur as an attractive force between two nonbonded atoms moving towards each other until a certain internuclear distance r_{eq}. Below this critical distance, the attraction is compensated by a strong repulsive contribution. The attractive part emerges from electron correlation in QM. For example, the electron fluctuation around a nucleus can generate a temporary dipole moment that will induce a complementary dipole moment in the adjacent atom. This dipole/induced-dipole interaction creates an attractive London or dispersion force and was first shown to decrease with the inverse sixth power of internuclear distance[164] ($-B_{ij}/R_{ij}^6$, red in **Figure 10**). The repulsive force originates from a combination of internuclear repulsion and electron-electron repulsion under the Pauli exclusion principle and, for computational reasons, was chosen to increase with the inverse 12th power of internuclear distance (A_{ij}/R_{ij}^{12}, blue). Both effects are combined in the Lennard-Jones potential that describes the nonbonded van der Waals interactions (black). The exact shape of this potential is dependent on the types of atoms, determined by the repulsive and attractive coefficients A_{ij} and B_{ij}, respectively.

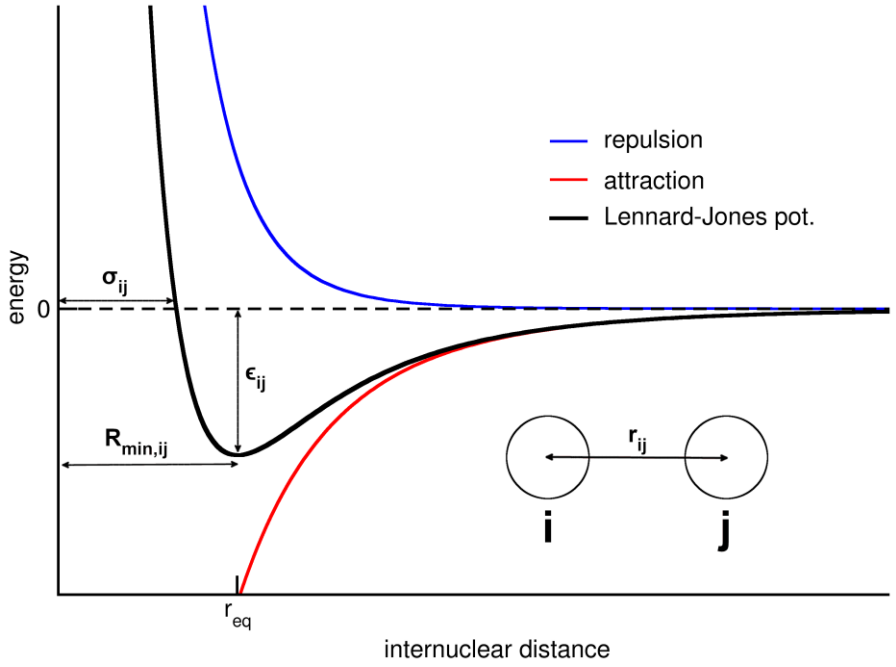

Figure 10. Lennard-Jones potential consisting of an attractive and a repulsive contribution, representing the van der Waals interactions between two nonbonded atoms as a function of their internuclear distance with equilibrium distance r_{eq}.

The Lennard-Jones potential is often written in the form of equation (9), where ϵ_{ij} corresponds to the potential well depth (see **Figure 10**), σ_{ij} indicates where the potential cuts the abscissa and $R_{min,ij}$ is the internuclear distance where the potential has its minimum (with $R_{min,ij} = \sqrt[6]{2}\sigma_{ij}$). Rearranging the Lennard-Jones potential in that way gives the definition of the coefficients A_{ij} and B_{ij} as shown in equation (10).

$$E_{vdw}(R_{ij}) = \sum_{i<j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] = \sum_{i<j} \epsilon_{ij}\left[\left(\frac{R_{min,ij}}{R_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{R_{ij}}\right)^{6}\right] \tag{9}$$

$$A_{ij} = 4\epsilon_{ij}{\sigma_{ij}}^{12} = \epsilon_{ij}{R_{min,ij}}^{12}$$
$$B_{ij} = 4\epsilon_{ij}{\sigma_{ij}}^{6} = 2\epsilon_{ij}{R_{min,ij}}^{6} \tag{10}$$

Lastly, the final term included in the nonbonded potential energy terms in equation (3) accounts for the electrostatic interactions. Charles de Coulomb first formulated a law for the force $F_{Coulomb}$ between two ions or partially charged groups decreasing with the inverse square of their distance, equation (11). This force is attractive, if the effective charges of the interacting particles q_i and q_j are of opposite sign, and repulsive, for the

same sign. From this, the Coulomb potential $E_{Coulomb}$ is derived, where the pre-factor is needed for conversion of the electrostatic charge units into kcal*mol^{-1} with the permittivity of vacuum ϵ_0. ϵ denotes the dielectric constant of the environment if the charges are exposed to a surrounding other than vacuum. In aqueous solution, for example, the charges are shielded by water. This screening effect is stronger when the two charges are further apart and vanishes (towards vacuum) when they approach each other. Such changes are accounted for by using more complex expressions for ϵ like sigmoidal functions allowing for distance-dependent changes in the electrostatic field.

$$F_{Coulomb} \propto \frac{q_i q_j}{r_{ij}^2}, \qquad E_{Coulomb}(R_{ij}) = \frac{1}{4\pi\epsilon_0} \cdot \frac{q_i q_j}{\epsilon R_{ij}} \tag{11}$$

Compared to the van der Waals interactions, the Coulomb potential, illustrated in **Figure 11**, is a long-range interaction since the decay is proportional to the inverse first power of the internuclear distance. The decrease of the Lennard-Jones potential is much more rapid, and therefore constitutes a short-range effect for nonbonded interactions.

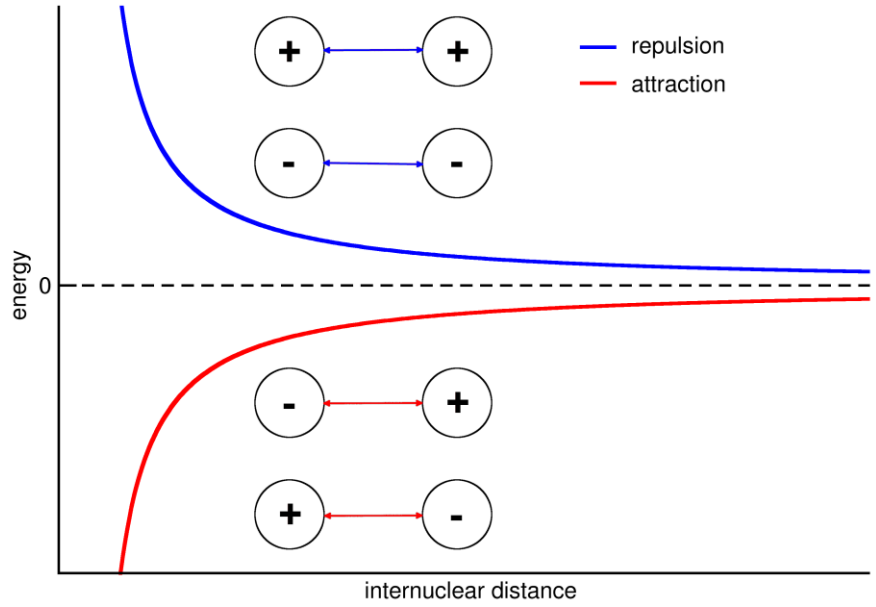

Figure 11. Coulomb potential according to equation (11), differentiating between an attractive and a repulsive interaction depending on the (partial) charges of the two particles.

Since electron movement, and thus charge redistribution, is not considered in MM, effective partial or net charges are assigned to all atoms reproducing the electrostatic potential calculated by more sophisticated QM methods. Because of that, conformation dependent changes in polarization or induced dipole moments cannot be modeled explicitly and are thus considered negligible.

Summarizing, the previous paragraphs introduced the standard, all-atom non-polarizable FF for calculating the potential energy of a system as a function of atomic coordinates. The simple functional formulation implicates a physical interpretation and, moreover, allows rapid calculation and derivation of biomolecular interactions and their relative stability. However, since there is no universal FF, the exact functional form and values of associated parameters must be chosen carefully for the system of interest. More limitations are caused by the drastic electrostatic approximations that, as mentioned above, fail to model higher electrostatic phenomena. More sophisticated polarizable FFs were developed to tackle this shortcoming. Furthermore, the computation of nonbonded energy terms define the bottleneck for such calculations since they must be evaluated for all pairs of atoms, and thus generate a complexity of $O(N^2)$, where N is the number of particles. For the dimensions of biomolecular simulations, this step becomes severely time consuming. However, different strategies were developed to reduce this complexity for feasibility.

2.1.4. Equations of motion and numerical integration

In classical mechanics, a system's motion follows underlying Newtonian laws and can be predicted for future time points. However, Newtonian dynamics are highly sensitive to initial conditions, meaning that two systems with almost identical initial conditions (initial coordinates and velocities) will diverge already after a few time steps. The system is thus both deterministic and chaotic. The motion can be calculated for a molecular geometry and its potential energy E_p, equation (3), with Newton's equation of motion. Accordingly, the force of the system is defined as the negative gradient (first partial derivative) of E_p as shown in equations (12) and (13).

$$F(X(t)) = -\nabla E_p(X(t)) = MA(t) \tag{12}$$

$$A(t) = \dot{V}(t) = \ddot{X}(t) \ , \qquad \nabla E_p(X)_i = \frac{\partial E_p(X)}{\partial \alpha_i} \ , \qquad \alpha_i = (x_i, y_i, z_i) \tag{13}$$

X is the collective vector of Cartesian coordinates of all atoms, V the corresponding velocity vector, A the acceleration, M the diagonal mass matrix. Dot notation indicates differentiation with respect to time t. For calculating the position and velocities for different points in time, this expression must be numerically integrated, for example, using the Verlet algorithm[165], which results in pairs of position and velocity $\{X^n, V^n\}$ for integers n at different time steps $n\Delta t$. In general, the basic position update results as a

two-step propagation, which means that for determining the positions after one step X^{n+1} the positions from the current step X^n and preceding step X^{n-1} must be saved:

$$X^{n+1} = 2X^n - X^{n-1} + \Delta t^2 A^n \tag{14}$$

Using a Taylor expansion around $X(t)$, the general definition of V^n can be derived as the central difference approximation:

$$V^n = \frac{(X^{n+1} - X^{n-1})}{2\Delta t} \tag{15}$$

Multiple variants, differing in numerical properties, exist to calculate the propagation of position and velocity like the Leapfrog scheme, the Position Verlet and Velocity Verlet variants. The former uses half steps for velocity and the corresponding propagation pairs $\{X^{n+1}, V^{n+1/2}\}$ can be written as:

$$\begin{aligned} X^{n+1} &= X^n + \Delta t V^{n+1/2} \\ V^{n+1/2} &= V^n + \frac{\Delta t}{2} A^n \end{aligned} \tag{16}$$

The other two variants formulate propagation pairs $\{X^{n+1}, V^{n+1}\}$ as whole steps of Δt, where the Velocity Verlet scheme is:

$$\begin{aligned} X^{n+1} &= X^n + \Delta t V^n + \frac{\Delta t^2}{2} A^n \\ V^{n+1} &= V^n + \frac{\Delta t}{2} (A^n + A^{n+1}) \end{aligned} \tag{17}$$

where X^{n+1} is obtained by combining equations (14) and (15). Lastly, the Position Verlet method is:

$$\begin{aligned} X^{n+1} &= X^n + \Delta t V^n + \frac{\Delta t^2}{2} A^{n+1/2} \\ V^{n+1} &= V^n + \Delta t A^{n+1/2} \end{aligned} \tag{18}$$

From equations (16), (17) and (18) it shows that the Leapfrog and Velocity Verlet schemes evaluate the force (via A, see equation (12)) at whole steps of Δt, but the Position Verlet variant at half steps. All variants recursively calculate future steps using information of the proceeding steps, resulting in a trajectory.

2.1.5. Practical considerations

A useful MD simulation of biomolecules requires further considerations, which will be introduced shortly. Before starting a simulation, the structure of the system must be properly prepared and minimized. After this initialization, the system is placed inside a space-filling box, called unit cell, which is surrounded by infinite copies of itself in all three dimensions under periodic boundaries. Only the unit cell is propagated throughout the simulation, so that a particle crossing the boundary is replaced by its periodic image. Under the minimum-image convention, every atom interacts only with the closest image of all other atoms within a spherical cutoff region. The unit cell must thus be large enough to avoid artifacts due to self-interaction. If the simulation is performed in explicit solvent, the unit cell is filled with solvent molecules, for example water, and ions. With that, the initial coordinates of all atoms of the system were defined. Initial velocities are then pseudo-randomly chosen so that the total kinetic energy of the system corresponds to the expected value at a target temperature. During the beginning of the simulation, an equilibration run should be performed to allow for proper relaxation until potential and kinetic energy converge to an equilibrium value. A thorough equilibration protocol, usually consisting of minimization and heat-up steps, is therefore useful for investigating the proper performance of periodic boundaries or artifacts due to initial placement of molecules. Constant energy simulations under the microcanonical ensemble (NVE, constant number of particles N, constant volume V, constant energy E) are only partly useful for biomolecular systems, where molecular properties are studied dependent on temperature (T) or pressure (P). Therefore, the canonical (NVT) or isothermic-isobaric ensembles (NPT) are more appropriate, which require the use of thermostats and barostats for temperature and pressure control, respectively. After the system is properly equilibrated, the production run is performed. Thereby, it is important to simulate the system for a sufficient amount of time to ensure meaningful statistical sampling of properties.

2.2. Molecular Docking with DynaDock

DynaDock is a docking tool and part of the DynaCell software published by Iris Antes in 2010[144]. Herein, the original implementation will be followed. Two steps define the docking procedure: (i) generation of initial poses during a broad sampling step, allowing for a certain atomic overlap between protein and ligand atoms. (ii) refinement of each

initial placement with a short OPMD simulation resolving the initial overlap and enabling dynamic adaption of binding partners.

In the broad sampling step, initial poses are generated within the binding site by random translation and rotation of the ligand and conformations are sampled by randomly modifying rotational DoFs of freely rotatable bonds. During this step, the receptor is kept rigid and a specified maximum overlap of van der Waals radii between any ligand and protein atom pair is allowed.

$$O = 1 - \frac{d_{12}}{r_1 + r_2} \tag{19}$$

The overlap O is defined as a ratio between the distance d_{12} between two atoms and their van der Waals radii r_1 and r_2. This criterion prevents the sampling of poses deep inside the protein, while accepting small clashes with protein residues around the binding site. A separate overlap value can be specified for ligand-ligand atom pairs. This might be useful for acceptable small clashes inside the ligand caused by random conformational sampling. Together with a maximum translation distance and maximum distance from protein parameter, the three-dimensional space for ligand placement can be defined depending on the shape of the binding site. The goal of the broad sampling step is to thoroughly sample the conformational space of the ligand while covering large areas of the binding site.

In the second step, each initial broad sampling pose is refined with a short OPMD simulation to resolve the atomic overlap. For this, soft-core potentials are employed for van der Waals and electrostatic interactions using the functional form of the Lennard-Jones and Coulomb potential by Taylor[166], shown in equations (20) and (21), respectively.

$$V_{LJ}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}^{12}}{\left(\boldsymbol{\alpha}^m\sigma_{ij}^6 + r_{ij}^6\right)^2}\right) - \left(\frac{\sigma_{ij}^6}{\left(\boldsymbol{\alpha}^m\sigma_{ij}^6 + r_{ij}^6\right)}\right)\right] \tag{20}$$

$$V_{Coul}(r) = \frac{(1-\boldsymbol{\alpha})^n q_i q_j}{4\pi\epsilon_0\left(\boldsymbol{\alpha} + r_{ij}^2\right)^{0.5}} \tag{21}$$

The soft-core parameter α scales both potentials and can take values between 1.0 and 0.0, where the latter corresponds to the normal physical potentials. The higher the value of α, the softer the potential and the lower the resulting energies and forces. Additional weighting parameters m and n were chosen to balance both contributions in a combined

potential and were finalized to *m=3* and *n=6* resulting in a suitable functional form. This way, for small interatomic distances of overlapping atoms, a large α value corresponds to a small repulsive term, softly forcing their separation without disrupting the system. The value for α is automatically determined depending on the initial overlap of the broad sampling pose and is optimized with respect to the potential energy of the system throughout the simulation. This cycle of optimization continues until α reaches zero and fully physical potentials are restored. Like that, the overlap is successfully refined, and the simulation continues for a user-defined number of steps for short equilibration of the system leading to the final docking pose.

3. Results

3.1. A modular approach to the bisbenzylisoquinoline alkaloids tetrandrine and isotetrandrine

Citation

Schütz, R., **Meixner, M.,** Antes, I., & Bracher, F. (2020). A modular approach to the bisbenzylisoquinoline alkaloids tetrandrine and isotetrandrine. *Organic & Biomolecular Chemistry*, *18*(16), 3047-3068.

https://doi.org/10.1039/D0OB00078G

Summary

Tetrandrine is a MC alkaloid of the plant *Stephania tetrandra* and a known inhibitor of calcium channels and P-gylcoproteins, making it a potential lead compound against viral infections or cancer. However, options for structural modifications of the natural product are sparse and a known total synthesis comprises more than 20 individual steps. In this study, a modular approach of 12 reaction steps was introduced, allowing for implementation of structural variations along the way. The synthesis strategy started from commercially available compounds and consisted of 4 key steps that, depending on their order, can direct the synthesis yielding either an equimolar mixture or preferably tetrandrine or its diastereomer iso-tetrandrine. Tetrandrine contains two stereocenters located within the MC ring with the same absolute configuration (1-R,1'-R or 1-S,1'-S), whereas iso-tetrandrine exhibits either 1-S,1'-R or 1-R,1'-S. In the final step of a proposed reaction route, the reactant, which was obtained as a racemic mixture of R- and S-configuration on the first stereo center, performed an intramolecular cyclization reaction under the formation of the second stereo center. Solely depending on the order of the previous reaction steps, the second stereo center was either built at the C-1 or C-1' atom and the reaction yielded a ratio of tetrandrine(**R**,**R**/**S**,**S**):iso-tetrandrine(**R**,**S**/**S**,**R**) of 29:71 and 87:13, respectively.

Based on these observations, molecular modeling studies were performed aiming for a structural explanation of the experimental ratios. The usual procedure would include, first, a MD-based conformational sampling step to generate appropriate starting structures, and second, the explicit calculation of the reaction paths using QM methods. As the second step is computationally demanding, such calculations for all possible reaction routes would have exceeded the scope of this study. Thus, we assessed the

resulting diastereomeric ratios based on the conformational stability of the MC intermediates of the corresponding reactions. Therefore, the four possible intermediates, differing in the configuration (S/R) and the location (C-1/C-1') of their first stereo center (C-1S, C-1R, C-1'S, C-1'R), were simulated at experimental conditions employing extensive MD simulations. A statistically significant number of potential starting conformations for the final reaction, where the respective atoms that form the second stereo center were reasonably close, were extracted as pre-reaction ensembles. Based on the conformation of the MC scaffold, the potential configuration (pre-S/pre-R) resulting at the second stereo center could be determined for every structure in these ensembles. Estimated computational ratios were thus obtained by comparing the number of structures in the pre-reaction ensembles that would result in tetrandrine (or iso-tetrandrine) with respect to the total number of pre-reaction structures. In case the second stereo center was built at C-1, the resulting computational ratio of tetrandrine(pre-**R**,**R**/pre-**S**,**S**):iso-tetrandrine(pre-**R**,**S**/pre-**S**,**R**) was 39:61. Correspondingly, for the case in which the second stereo center was built at C-1' in the final reaction, the computational ratio of tetrandrine(**R**,pre-**R**/**S**,pre-**S**):iso-tetrandrine(**R**,pre-**S**/**S**,pre-**R**) was 74:26. These estimates agreed very well with the trends of the experimental ratios. However, since no reaction quantities were explicitly calculated, the results could be only qualitatively compared, but are, nevertheless, interesting. Our investigations highlighted that the conformational distribution and stability of intermediates resembled the stereoselectivity of these cyclization reactions.

Contribution

Experimental work was performed by Ramona Schütz. Maximilian Meixner designed the theoretical part of this study and performed all computational simulations and statistical analysis and wrote the theoretical part of the manuscript. All authors reviewed the final manuscript.

3.2. Acyldepsipeptide probes facilitate specific detection of caseinolytic protease P independent of its oligomeric and activity state

Citation

Eyermann, B., **Meixner, M.,** Brötz-Oesterhelt, H., Antes, I., & Sieber, S. A. (2020). Acyldepsipeptide probes facilitate specific detection of caseinolytic protease P independent of its oligomeric and activity state. *ChemBioChem*, *21*(1-2), 235-240.

https://doi.org/10.1002/cbic.201900477

Summary

The caseinolytic protease is a homo tetradecameric complex of the serine peptidase caseinolytic protease P (ClpP), consisting of two heptameric rings forming a barrel shaped quaternary structure. The proteolytic activity of the complex depends on the conformation of the protein subunits varying between an inactive compressed state, an intermediate compact state, and a fully active extended state. Accompanied by the chaperone ClpX, which unfolds globular proteins, the ClpP complex breaks down the linear peptide chain. ClpX binds hydrophobic pockets on the surface of the ClpP complex in-between two ClpP subunits with a conserved IGF residue motif. Acyldepsipeptides (ADEPs) are small MC compounds mimicking this motif and therefore bind the same apical binding site, widening the pore of the barrel and facilitating access and breakdown of larger substrates resulting in uncontrolled digestion of cellular proteins. In this study, ADEP derived photoprobes were designed to label the ClpP complex independent of its conformation and activity state. For the detection via affinity-based protein profiling, the compounds were designed with an alkyne handle as well as a photoreactive diazirine group. The two photoprobes, called 266 and 288, only differed in the location of the diazirine group, which was incorporated in one of two proline residues of the MC. The proline residue facing towards the pore is methylated in ADEP compounds. For 288, the diazirine group replaced this methyl group while the methyl-proline is conserved in 266, where the diazirine moiety was attached to the second proline. Albeit these minor structural variations, significantly different activation, and labeling of ClpP was observed. Molecular modeling studies could provide a structural explanation to the experimental differences. The two molecules 266 and 288 were docked in the known ADEP binding site located on the subunit interface of ClpP and exhibited a similar binding mode as co-crystallized ADEP and ADEP1 compounds. For 288, the weak activity could be explained by the missing methyl group of the conserved methyl-proline in the ADEP

scaffold, which is crucial for activation of the ClpP complex. For 266, the diazirine group was solvent exposed. Accordingly, the labeling signal was buffered by solvent molecules explaining the weaker labeling signal while retaining ClpP activation.

Contribution

Experiments in this study were performed by Barbara Eyermann. Maximilian Meixner designed, conducted, and analyzed the computational part of the study and wrote the respective paragraphs of the manuscript. All authors reviewed the final version of the manuscript.

3.3. Turning the Actin Nucleating Compound Miuraenamide into Nucleation Inhibitors

Citation

Wang, S.*, **Meixner, M.*,** Yu, L., Zhuo, L., Karmann, L., Kazmaier, U., Vollmar, A. M., Antes, I., Zahler, S. (2021). Turning the Actin Nucleating Compound Miuraenamide into Nucleation Inhibitors. *ACS omega*, *6*(34), 22165-22172.

https://doi.org/10.1021/acsomega.1c02838

***Shared first authors**

Summary

The actin protein is the building block for dynamic filaments in the eucaryotic cell. Actin monomers can accumulate and form nuclei, which eventually grow into filaments. The 3D structure of a free actin monomer is distinctly different from an actin subunit incorporated in the filament (F-actin). Various actin-binding MCs are known, which can be classified into stabilizers and destabilizers. The myxobacterial MC miuraenamide A is an actin nucleator. In a previous study, the binding mode of miuraenamide could be predicted using a combination of molecular docking and MD approaches. Binding the characteristic macrolide binding cleft at the actin-actin interface of the nucleus, miuraenamide facilitates a tighter packing of subunits compared to unbound F-actin. Moreover, interactions between miuraenamide and the adjacent actin subunit caused a crucial shift of the DNase-binding loop (d-loop), prohibiting the decoration of filament-destabilizing proteins such as cofilin.

Based on these findings, derivatives were synthesized, some of which showed the opposite biological effect to the parent compound causing overall less filaments to be formed without altering the elongation rate of already existing filaments. In a case study, the binding mode of the derivative LK701 was investigated. While a stable binding pose could be predicted in a free actin monomer, the compound was unable to penetrate the macrolide binding cleft of a stable apo nucleus structure. In the latter, the cleft was occupied by the d-loop of the adjacent subunit, building characteristic longitudinal interactions that promote filament formation. These results confirmed that LK701 blocked the macrolide binding cleft of free actin subunits on a monomer level, prohibiting such proteins from forming new nuclei, while existing nuclei remained unaffected.

Comparing the binding modes of LK701 and miuraenamide revealed that the former, which missed a bulky phenyl substituent, bound deeper inside the macrolide binding cleft of an actin monomer, blocking important interaction sites for nucleus formation. It could thus be explained how minor structural changes could convert the nucleator parent compound into a nucleation inhibitor.

Contribution

Experiments were performed by Shuaijun Wang, Lushuang Yu and Ling Zhuo. Compounds were synthesized by Lisa Karmann and Uli Kazmaier. Maximilian Meixner was responsible for the computational part of the study. Together with Iris Antes he conceptualized the study. He performed and analyzed all calculations and simulations and wrote the respective chapter of the manuscript, which was carefully revised by all authors.

3.4. Dynamic docking of macrocycles in bound and unbound protein structures with DynaDock

Citation

Meixner, M., Zachmann, M., Metzler, S., Scheerer, J., Zacharias, M., & Antes, I. (2022). Dynamic Docking of Macrocycles in Bound and Unbound Protein Structures with DynaDock. *Journal of Chemical Information and Modeling*.

https://doi.org/10.1021/acs.jcim.2c00436

Summary

MCs hold great therapeutic potential and count many advantages due to their characteristic ring scaffold. The many experimental applications require suitable theoretical tools for the accurate prediction of binding modes of MC inhibitors. Most molecular docking programs, however, fail to describe the flexibility of the MC ring during the generation of docking poses. Additionally, common theoretical workflows are mainly based on the holo state of the receptor with the conformation of the binding site already adapted to a co-crystallized ligand. In this study, the bioactive states of 20 diverse MC inhibitors were predicted in holo and apo proteins. Therefore, we optimized our previous sampling and docking protocol. The dynamics of the ligands were separately simulated in a pre-sampling step with explicit solvent to generate different conformations. We investigated different sampling conditions to find the most appropriate settings for MCs. The sampled structures were thoroughly analyzed with our new torsional classification tool *ClassTor.py*, which classifies conformations based on the distribution of sampled dihedral angles of ring torsions. We could show that this created a more meaningful partition compared to the dihedral clustering approach of an existing analysis tool. From the classification results, ensembles of conformers were extracted and subjected to subsequent fully flexible docking calculations with DynaDock to compensate the lack of ring flexibility during pose generation. First, an initial broad sampling step was performed, where random ligand poses were generated inside the rigid binding site of the receptor, allowing for atomic overlap between protein and ligand atoms. A subsequent refinement step resolved the initial overlap using short, OPMD simulations considering full flexibility of receptor and ligand. This way, high accuracy prediction of binding poses in bound and unbound protein structures was achieved. Our results suggested that docking in apo structures required high level of receptor flexibility. We further highlighted that MD-based pre-sampling led to high quality MC (ring) conformers. Including a

conformation close to the bioactive state improved the chance for successful docking. Large flexible substituents binding to solvent-exposed binding sites were the main cause for remaining unsuccessful docking cases. However, such poses could be further improved by explicit MD simulations of the docked complexes.

Contribution

The idea of this study was based on previous work performed in the group of Iris Antes. The concept was designed and developed by Maximilian Meixner and Iris Antes. All simulations and calculations were conducted by Maximilian Meixner with help on two specific systems of the scope of a research internship project by the students Sebastian Metzler and Jonathan Scheerer, who were supervised by Maximilian Meixner. Furthermore, Maximilian Meixner conceptualized, wrote, and distributed the torsional classification tool *ClassTor.py*. Calculations involving DynaCell were performed by Maximilian Meixner and consulted by Martin Zachmann. The manuscript was written by Maximilian Meixner and reviewed by all authors.

4. Discussion

The discussion focuses on the most challenging points for molecular modeling of MC compounds as elucidated in the introduction and comprehensively compares the results of studies presented herein with recent and relevant publications in the field.

4.1. Conformational sampling

4.1.1. Predictive power of MD simulations

In the first study of this work, using extensive MD simulations we structurally explained the stereochemical preferences of MC intermediates during a final intramolecular cyclization reaction and successfully reproduced the experimentally observed diastereomeric ratio on a qualitative level by conformational sampling and analysis of pre-reaction ensembles[155]. We discussed that the addition of such extensive computational investigation could generally guide targeted synthesis towards the desired (stereochemical) product, if multiple possible pathways and their associated reaction mechanisms and intermediates are known, for example by estimating corresponding ratios in advance.

In the past, similar case studies have been conducted, where detailed MD simulations and conformational analysis were performed to identify stable and flexible parts of a MC[167]. From thus generated conformers, energy calculations could accurately estimate rotational barriers and reproduce spectroscopic data. With this, the absolute configuration of the bisbibenzyl MC isoplagiochin C could be unraveled for the first time.

In a sole computational investigation by Cleays and coworkers, semiempirical MD simulations could accurately reproduce the thermodynamic stability of MC alkene isomers featuring two stereocenters and a double bond[100]. Supported by high level torsional energy profiles of crucial ring dihedrals, large influence and minor effects on the MC ring conformation could be attributed to the chirality of the stereocenters and the configuration of the double bond, respectively.

In our work, like in these earlier studies, such simulations were performed to substantiate experimental findings. However, more recently, the aim shifted towards developing prediction models for specific experimental observations. In the work of Wang *et al.* in 2019, a predictive model for chromatographic elution orders and separation factors of enantiomers with a chiral polymer stationary phase was developed from MD simulations[168]. 10 racemic solutes in 3 different explicit solvents were separately simulated. Hydrogen bond lifetimes and aromatic π-π interactions between the drugs and

the stationary phase were quantified as averages over the MD simulation. The maximum hydrogen bond lifetime was found as a suitable metric to predict separation factors with a correlation factor of 0.85 and correct elution order in most cases. Remarkably, their model worked accurately without prior experimental information or fitting. Encouraged by these results, they improved their simulations in the following year by attaching the stationary phase to an amorphous silica, corresponding to the experimental setup[169]. This provided the advantage that more interactions could be formed in total, and larger drugs could interact across adjacent polymer strands. Based on such quantities, their model performed even more consistently and could overcome previous weaknesses.

Kim *et al.* investigated the product specificity of different monoterpene synthases[170]. This class of enzymes catalyzes the first step in the monoterpene biosynthesis by converting a precursor substrate either to an acyclic, a monocyclic or a bicyclic product. Such enzymes exhibit different product specificities, meaning that some produce only one such product, while others can generate multiple products from the same substrate. The conformation of a crucial carbocation intermediate seems to be responsible for either direct deprotonation resulting in a monocyclic product, or a second cyclization, yielding the bicyclic product. With a combination of high-level energy calculations and MD simulations, the group could successfully correlate crucial active site residues and differences in their nonbonded interactions to the intermediate with the experimentally observed product ratios, thus laying the foundations for models that could predict the product specificities of new monoterpene synthases in advance.

Together with these findings, our analysis highlighted the general potential of MD simulations for the development of prediction models, if the experimental objective and mechanistic insights are known and a suitable computational metric for quantification could be determined, which must be investigated for each case. More research is necessary that benchmark the incorporation of MD simulations in predictive models and investigate their applicability for MCs.

4.1.2. Torsion-based analysis of sampled structures

We developed the classification scheme *ClassTor.py* based on sampled ring torsion angles throughout MD-based conformational sampling of MCs to reduce the amount of obtained structures and analyze the conformational space in a meaningful way[104]. We could show that our procedure produced more qualitative partitions than dihedral clustering of existing methods. In short, the sampled dihedral angle values of selected MC ring torsions

were extracted and individually counted and convoluted with a standard Gaussian kernel to obtain a smooth spectrum-like distribution. Each spectrum was automatically characterized by detecting minima and maxima of the distribution, corresponding to less sampled and prominent dihedral angle values of this torsion, respectively. This information was used to divide each spectrum into labeled bins, declaring those minima and maxima as bin borders and bin centers, respectively. With this definition, each sampled structure could be classified by a specific combination of bin labels, and structures with the same combination were thus combined in the same class. Naturally, the obtained number of classes corresponds to the number of unique combinations of bin labels. Relying solely on the actual sampled torsion angles, this classification provides a meaningful analysis of occurring ring conformations and is thus more intuitive for MCs than Cartesian clustering, as employed in the previous pipeline[102]. From the classification results, representative structures of classes were extracted to a conformer ensemble, which was subjected to subsequent docking calculations. Our focus was to include a structure with a ring conformation close to the bound state for increased chance of successful docking. By choosing the centroid structures of the 10 highest populated classes, the final ensemble featured a bioactive-like conformer for most of the MCs in our data set, meaning the bound state was frequently visited during MD sampling and thus among the most prominent structures of the simulation. However, in the case of GDM, where the bound state is significantly different from abundant solution structures[62], the former, although sampled, was not found among the 10 highest populated classes. For such cases we implemented a selection of conformer subsets performed after initial classification that is optimized for structural diversity. By the user-defined subset size, the diverse ensemble is automatically generated by picking conformers based on the dissimilarity of their characterizing bin labels. For example, if two structures differ only in the bin label of one torsion of the MC ring, they will be classified into different classes, however since most ring torsions obtained the same bin labels, and thus related dihedral angle values, the two conformers might still be similar. The diverse selection thus considers only representatives of classes featuring many differences in their combined bin labels. We could show that this selection procedure resulted in a subset of conformers including the less sampled but structurally distinct bound-like conformation of GDM.

Another method optimized for diversity of MC structures used the Kennard-Stone algorithm and was applied in the abovementioned study by Claeys *et al*[100]. In their procedure, the two most dissimilar structures based on an RMSD value of geometric

distance matrices were selected first. Iteratively, the least similar structure, corresponding to the conformer with the largest pair distance to the previously chosen pool was added to the selection until a desired ensemble size was obtained. This way a uniform distribution was created that represented all parts of the sampled space equally, independent of their sampling frequency.

Focusing on torsion angle space, a former study clustered MD-sampled peptide conformations using dihedral angles of backbone torsions and showed that, unlike dihedral clustering, Cartesian clustering and dihedral PCA could not properly resolve all conformational states[171]. Furthermore, their dihedral clustering allowed for visualization of transition frequencies between clusters and could thus be applied to distinguish low-frequency transitions in folding events.

Substantiated by these examples, our results further highlighted the use of a diversity-oriented classification in torsion angle space for analysis of MC conformations. Due to the elevated rotational barriers of ring torsions, higher energy conformers might be sampled less frequently compared to local minimum structures. Therefore, a population-based selection might miss structurally distinct ring conformations. Thus, the option for creating a diverse conformer ensemble in our approach successfully accounted for that. Moreover, our approach is not limited to MC ring torsions but could also be applied to internal conformational changes in peptide and protein structures. However, the complexity of this procedure increases with the number of torsions included for classification. Thus, *a priori* knowledge about the system is required to limit this number to a manageable amount, for example when investigating a ligand-sized peptide or if a certain region of a protein is responsible for the investigated conformational changes. Future research should be conducted to compare this torsional classification with standard Cartesian clustering for conformational analysis of peptides and proteins.

4.1.3. The MC ring: sampling with MD or structure generation tools?

We showed that exhaustive MD simulations performed at elevated temperatures with carefully optimized sampling conditions (3x 1,000 ns, 600 K) produced high quality structural ensembles that featured conformers close to the bioactive state. The moderately increased temperature enabled overcoming elevated torsional barriers without sampling unphysical ring structures. Thus, the conformational space was extensively explored. However, this in-depth procedure required more computation time than structure generation algorithms of typical tools, as discussed before[104].

As introduced above, the characteristic ring scaffold of MC compounds poses a major challenge for conformational sampling methods. Highlighted by a recent study, the conformational space of MCs is still understudied mainly due to the increased rotational barrier of ring torsions, which accounts for their unique structures compared to linear counterparts[98]. Ebejer *et al.* performed a comparative study of freely available tools for conformer generation (Balloon, Confab, Frog2, RDKit, MOE), tested on a set of 708 small molecules[86]. MCs with rings larger than 7 atoms could not be handled by many methods, which failed to reproduce the experimental ring conformation. They exemplarily mentioned a 15-membered MC for which the generated conformers of all tools showed an average RMSD value of 2.79 Å relative to the crystal structure conformation, likely due to the challenging ring scaffold. They furthermore elaborated that for molecules with a central core and flexible extensions, which is often the case for MCs, all tools produced bad results. In general, increasing molecular flexibility caused less accurate structural ensembles. For example, molecules with 9 or more rotatable bonds showed average minimum RMSD values to the crystal structure reference above 1.0 Å by all toolkits. A similar study in the following year included a MC test set of 30 compounds, which exclusively featured 9 or more rotatable bonds per molecule[107], further elucidating the incompatibility of the former tools with MCs. In the latter study by Chen *et al.*, other mainstream methods for conformer generations were tested (MOE, LMOD approaches, MMBS [*]) for the reproduction of the crystal structure, the diversity of generated conformers and the location of the global energy minimum structure. They found that low-mode methods performed better than a stochastic search, and advanced settings were superior to default settings, especially for MCs. For example, increasing the energy window improved the diversity (in terms of compactness and extendedness, measured by radius of gyration, Rgyr) of resulting MC conformer ensembles, thereby stressing the increased interconversion barrier for such compounds. Furthermore, the MC-specific method MMBS (see **Table 2**) has per default an increased energy cutoff and, as a proof of principle, outperformed most of the other methods on the MC test set, even with its default settings, which reproduced the bioactive conformer with an accuracy below 2.0 Å in 97% of the cases. This finding confirmed the need of specialized strategies for complex molecules like MCs.

[*] called MD/LLMOD in that study or MacroModel (MC)[108] and Macrocycle Conformational Sampling (MCS)[91] in other studies. However, for consistency in this work and to prevent confusion with the distinct Prime-MCS algorithm, the term MMBS (MacroModel Baseline Search)[92] is used, as proposed by the authors.

In two recent studies, Poongavanam and colleagues first investigated if conformer generation tools (OMEGA, MOE, MMBS) could reproduce known MC conformations in different environments[108], and secondly, if cell permeability could be successfully predicted from OMEGA-sampled conformers[44]. In their first work, they concluded that OMEGA outperformed other methods in accuracy and coverage of conformational space and explained that this might be due to the algorithmic differences of the investigated methods. According to that, the fragmentation and reconstruction technique implemented in OMEGA was more suitable for MCs and superior even to the MC-specific low mode-based MMBS and MD-based MOE methods. To stress this even further, structural ensembles from extensive MD simulations were structurally compared with the OMEGA ensembles by Rgyr and polar surface area (PSA). Again, MD performed worse in exploration of conformational space and resulting ensembles were influenced by the polarity of the environment, e.g., ensembles sampled in apolar solvent were less diverse than those of water simulations. However, their MD simulations were limited to 20 ns, which – as results in our study suggested[104] – is insufficient for exhaustive conformational sampling of MCs and, under modern standards, too short for extensive ligand-only simulations. Nevertheless, their study highlighted the use of structural measures other than RMSD and the strengths and weaknesses of conformer generation tools, which, as introduced above, should be carefully considered for the intended objective. In the follow-up study, a machine learning prediction method for MC compounds was introduced that classified molecules as low-medium and highly permeable based on 2D and 3D descriptors. While 2D descriptors were faster to calculate and more useful for such predictions, the authors showed that a model based on 3D descriptors was successful only for rigid MCs, for which the conformations in solution could be accurately predicted. Flexible MCs, for which such models failed, obtained solution structures significantly above the global energy minimum (5-15 kcal/mol). Although the defined energy cutoff for OMEGA (25 kcal/mol) would have been sufficient for complete sampling, relevant conformations for flexible MCs could not be obtained, which the authors attributed to the inability of the FF, which is used for refinement of generated structures, to identify such conformations. However, the authors did not compare other FF- or MD-based methods for conformer generation in that study. Interestingly, they explained that conformers of rigid MCs were more accurately predicted because they consisted of only the core ring structure. The error for larger MCs was thus attributed not to the ring scaffold but to flexible side extensions anchored to this

core, for which biologically relevant conformations were more difficult to generate. This finding is in agreement with results of our combined sampling and docking study[104], which also highlights flexible substituents of the ring as one of the main structural challenge for molecular modeling of MCs.

Other studies that compare tools with MD were performed by Sindhikara and colleagues[92], the authors of Prime-MCS, comparing its performance with MMBS, MD and MOE, and Romero *et al.*[96] (Moloc, Prime-MCS, MacroModel, MOE, ETKDG, Conformator, CCDC, MD). Both used only short MD simulations (24 ns) for conformer generation, which again seemed insufficient for exhaustive exploration of MC conformer space. While the former work highlighted Prime-MCS as the superior method in terms of structural accuracy, conformational diversity and computational speed, the latter study mentioned MD among the most accurate software along MOE, MacroModel and Prime-MCS. In both cases, instead of only using RMSD values, Rgyr and torsional fingerprints were calculated to structurally evaluate the generated conformers.

A profound work from 2021 studied ensemble completeness for conformers of 7 small MCs in different charge states and solvents, generated by 3 tools (Prime-MCS, BIOVIA Best, Conformator) and extensive MD simulations (5x 100 ns)[103]. Their conformational maps obtained from PCA of ring torsions favored MD in coverage of sampled space. Furthermore, MD simulations produced distinct ensembles dependent on the MC, explicit solvent, and charge state, whereas conformers generated by tools were similar, independent of the (implicit) solvent.

MD methods, and others that perform FF refinement, are, of course, dependent on the employed FF. A previous study, that compared 3 different FFs concluded that the type of FF did not significantly influence the reproduction of the bioactive state as well as the conformational coverage, and that the extent of sampling and the treatment of electrostatics were more crucial[107]. Further application studies of MD for conformational sampling (of MCs) are also included in the next section. Hence, the previous drawback of the MD method described by Poongavanam *et al.* might have been due to insufficient sampling rather than the use of the FF *per se*[108]. Therefore, a thorough benchmark of different FFs for the conformational sampling of MCs would be needed, which could more properly judge their general applicability and suggest if further development is necessary. For example, in the case of cyclic peptides, a residue specific FF was developed[172]. Providing guidance, results from our study suggest suitable MD sampling conditions optimized for MC compounds[104].

To further compare the conformer ensembles obtained by our optimized MD sampling conditions, we generated 1,000 conformers with Prime-MCS and Balloon (DG option), respectively, and calculated RMSD values of MC ring atoms relative to the known bound state. From the combined MD simulations of each MC, we chose 1,000 equidistant frames constituting an ensemble. These supplemental results (**Figure A1** of the appendix) suggest that MD structurally reproduces the bioactive state of diverse MC ligands more accurately than Prime-MCS and Balloon, and often cover a greater range of conformations, especially for large rings like FK5, RBT and RAP. Although 1,000 conformers were requested for each tool, the actual ensemble size was consistently smaller. This might be due to the implemented algorithms, that are said to filter out redundant conformations. It must be mentioned that no torsional fingerprints or unique ring conformations were determined for this analysis, and thus, no comments can be made about the exhaustiveness of sampled space. However, Balloon found the fewest conformers for large rings (FK5, RBT, RAP: 4, 4 and 2 structures, respectively) or no conformer at all for structurally complex MCs (CY9, fused ring system). This confirms that MCs cannot be handled by Balloon, as previously indicated by others[97]. Surprisingly so, since no upper limit for the number of flip-of-fragment operations was set for the production of flexible ring conformations[76]. Moreover, when given a 3D structure as input, Balloon often failed to assign the correct atom types, which led to saturation of double bonds and, ultimately, to the production of wrong output conformers. In such cases, the SMILES code was used instead. To overcome these shortcomings, different settings were tested beforehand, but no parameter set could be found that worked equally well for all MCs. As described previously[103, 107] and mentioned by the authors of Balloon[76], advanced settings slow down the procedure, which is why the flip-of-fragment operation is limited to a small number by default. However, if enhanced settings of these tools take significantly more time to generate structures, their major advantage, fast computational speed, is lost.

In general, default settings seemed to be inapplicable to complex MCs, while appropriate settings (e.g., energy window) for capturing relevant ring conformations are unknown. Although many abovementioned studies compared advanced settings, clear guidelines are still missing. Since this is, in general, also the case for MD, our results suggest suitable and highly optimized MC-specific conditions for MD-based conformational sampling. Thus, the more time-consuming methods like MD are worth their computational demand in this case. However, if such computation-expensive approaches are no option, tools with

special MC treatment might be generally preferred over standard tools. But even newer conformer generators such as Conformator sometimes failed to predict MC ring conformation correctly, as shown for the structurally complex GDM[95]. This could be due to the unjustifiable fact that dihedral angles of ring torsions are sampled individually although such torsions are not independent. Such and related ring splitting algorithms have therefore no guarantee of producing valid ring conformations. Conformational sampling of MCs is still a field of active research, we hereby stressed that MCs need specialized workflows and strategies to overcome their unique structural challenges.

4.2. Molecular Docking

The following discussion covers a comparison of our studies with other approaches in terms of overall performance, incorporating ligand and receptor flexibility, as well as evaluation of docked poses. For easier reading, the relevant results are summarized first. Our updated pipeline for molecular docking of MCs, tested on 20 diverse ligands, included an exhaustive pre-sampling step to generate conformer ensembles[104]. From MD simulations with optimized sampling conditions, 10 starting structures were extracted by dihedral classification. Each conformer was flexibly docked with DynaDock[144] into the binding site of its host protein in holo and apo states. Docking was defined successful if the ligand position in the equilibrated X-ray structure could be reproduced with an accuracy below 2.0 Å, which was achieved in 95% (75%) of holo (apo) cases. Scoring was performed with MMGBSA (Amber) and compared to an equivalent score calculated with DynaCell. For successful docking poses in holo proteins, both scores could identify at least one such pose among the 10 top-scored poses in all but one system, while for apo docking calculations, DynaCell and Amber scores failed to rank such poses in only two and four cases, respectively. Additional post-docking MD simulations of unsuccessful apo docking poses quickly stabilized the initially docked complex, resulting in a more accurate pose that resembled the bound state in most cases. For an average sized system, pre-sampling of ligand conformations took 6 days, pose generation in the broad sampling step 2.8 hours and refinement of a single pose 6.2 hours on average.

In the recent past, five benchmark studies investigated molecular docking of MCs[99, 116, 124, 126, 127]. They either focused on comparing different conformational sampling strategies (MMBS, Prime-MCS, MCMM*, MCMMshort) producing distinct conformer ensembles which were subjected to the same docking algorithm (Glide)[99], or pre-

* MonteCarlo Multiple Minimum approach implemented in MacroModel

sampling with a single method (MMBS) and comparing the performance of different docking programs (Glide, AutoDock Vina, AutoDock 4, Autodock 3, DOCK)[116]. Furthermore, Anighoro *et al.* and Lam *et al.* tested ligand sampling and subsequent docking in MOE[127] and ICM[126], respectively, while Martin *et al.* investigated MT*/LMOD sampling in MacroModel and docking with GOLD[124].

In the first of these studies, published in 2016 by Anighoro *et al.*, 48 MC compounds were docked into their holo proteins[127]. MC ring flexibility was accounted for by using a pre-generated ensemble, which increased the docking success to more than 50% in contrast to using a random ligand structure or the lowest energy conformer, which could accurately reproduce the binding mode in only 27% and 41% of cases, respectively. Docked poses were only structurally evaluated by calculating RMSD values with respect to the crystallized reference discriminating three categories (≤1.0 Å, >1.0 Å and ≤2.5 Å, >2.5 Å). The major challenges were to extract a ligand conformer close to the bioactive state from the generated structures with MOE and to accurately account for van der Waals interactions, which dominated the shape-driven binding modes of MCs.

The following year, the study by Alogheli included 16 ligands in a test set and 31 in a validation set[99]. Glide flexible ring docking, where MC ring conformations were intrinsically sampled during docking, was compared to external generation of conformers with different sampling methods and subsequent docking. Docked poses were structurally and energetically evaluated by RMSD and the internal Glide score, respectively. Only the flexible ligand docking with Glide and rigid docking of MMBS-generated conformers produced average RMSD values below 2.0 Å for ligands in the validation set, while Prime-MCS sampling suffered from significant problems for larger MC ligands. When employing external sampling, computation times for the conformational search were higher than for the docking step, ranging from 139 to 33,579 minutes for the former and from 4 to 2,167 minutes for the latter.

In the same year, Castro-Alvarez published their benchmark study of docking programs investigating 20 MC complexes[116]. All programs except AutoDock 3 produced top-scored poses with mean RMSD values below 2.0 Å and lowest RMSD poses below 1.0 Å. Presented computation times varied between 5 minutes and 52 hours for docking all conformers of a system. Re-scoring with MMGBSA could improve weaknesses of some of the scoring functions, obtaining values close to experimental data.

* Mixed-torsion

In 2019, Martin *et al.* studied 41 MC compounds with GOLD, comparing intrinsic ligand sampling during docking with a more extensive pre-sampling and docking protocol[124]. Docking success was defined as reproduction of the crystallized binding mode with an accuracy below 2.0 Å. When GOLD flexible ligand docking was performed, binding modes could be successfully reproduced in 29.3% of cases, while using a pre-generated ensemble of conformers improved performance to 58.5%. Reported mean computing times ranged from 3 minutes for flexible docking to 7.5 hours for rigid docking of conformer ensembles excluding the pre-sampling step.

Finally, Lam *et al.* presented the most substantial benchmark set of 246 MC-protein complexes subjected to ICM-dock, which included flexible ligand sampling during docking[126]. Their procedure produced at least one pose within 2.0 Å among the top 5 (top 10) scored poses in 75% (81%) of cases. The median computation time was only 16 minutes. Furthermore, this group participated in the 2018 Drug Design Data Resource (D3R) Grand Challenge 4, which was a community-wide blind prediction challenge that featured – among other tasks – pose prediction of 20 ligands in beta secretase 1 (BACE-1), 19 of which were MC compounds[173]. The first stage (1A) of this challenge was a cross-docking setup, where the ligands were supposed to be docked into the active site without knowledge of the experimental binding pose and only provided with the FASTA sequence of the receptor and SMILES codes of the ligands. For the second stage (1B), the experimentally determined receptor conformations of these complexes were released, and ligand poses should be re-evaluated in a self-docking study. The challenge counted many submissions that resulted in corresponding publications on molecular docking of MC ligands, which will be included in the following discussion.

4.2.1. Performance

First, it needs to be stated that all beforementioned studies included different MC ligands in their data sets, which only allows for a qualitative comparison with our study. However, some prominent MC ligands overlapped among data sets (see below). Furthermore, for the most reasonable comparison possible, we included results of other studies with similar setups (external generation of ligand conformations and subsequent docking of conformer ensembles), even if other protocols might have obtained better results. Such differences are discussed afterwards. It must be kept in mind that, next to the varying programs and applied parameters, other preliminaries (like preparation of ligand and receptor structures, definition of the binding site, quality of structural alignment, etc.)

might have also influenced the docking procedure, and thus the final calculation of RMSD values of resulting poses. This section focuses on structural accuracy of the docked poses and scoring is included in the end of this discussion.

Table 4. RMSD values [Å] for the most accurate docking pose of data set-overlapping MC ligands of holo docking results of different studies. Mean value of lowest RMSD poses of all ligands of a data set is given below ("data set average"). Values highlighted in red if above 2.0.

sampling		MMBS					MMBS	Prime-MCS	MD	
docking		AD[a] Vina	AD[a]4	AD[a]3	DOCK	Glide	Glide		DynaDock	
ligand-ID	S1A	0.45	0.69	0.64	0.41	0.52				0.55
	LAR	0.47	0.61	0.82	0.59	0.36	0.21	1.56	1.74	1.00
	CY9	0.61	0.71	0.61	0.45	0.31	0.17	0.15	1.25	0.63
	GDM						0.34	0.59		0.62
	FK5						2.67	2.37		1.79
	RAP						0.51	7.24		1.31
	6QG								1.25	0.31
	06H								1.15	0.43
	27J								1.12	1.23
	ZER								1.64	0.84
	58T								0.83	0.53
data set average		0.83	0.97	1.46	0.93	0.92	1.14	2.55	1.28	1.06
reference		Castro-Alvarez *et al.*[116]					Alogheli *et al.*[99]		Ugur *et al.*[102][b]	Meixner *et al.*[104][c]

[a] AutoDock, [b] previous pipeline, [c] updated pipeline

In our approach, the mean RMSD value of the closest docking poses* for all ligands was 1.06 Å (1.97 Å) in holo (apo) systems and most accurate poses were below 2.0 Å in 95% (75%) of cases. In similar setups of other studies, accuracy was significantly lower. Martin *et al.* and Anighoro *et al.* reported success rates of only 58.5%†[124] and 52.1%[127], respectively, where, for the latter, the accuracy limit was even raised to 2.5 Å. For overlapping ligands across data sets, a more direct comparison can be drawn to the works of Alogheli *et al.* and Castro-Alvarez *et al.* For similar procedures included in their works, **Table 4** lists reported RMSD values of the most accurate docking pose of each ligand and the mean value over all ligands in the respective data sets. For the individual ligands, our

* minimum RMSD pose to the equilibrated bound complex

† here, success was defined if one of the top 3 scored poses obtained an RMSD value below 2.0 Å.

approach was comparable in structural accuracy to other methods. For large rings like FK5, our dynamics-based pipeline achieved accuracy below 2.0 Å, outperforming docking with Glide. For such ligands (FK5, RAP), Alogheli *et al.* reported significant problems when using Prime-MCS, a conformational sampling method specifically designed for MC compounds, leading to an average RMSD value of 2.55 Å for the most accurate docking pose[99]. Together with our studies, this highlighted the importance of accurate sampling of bioactive-like conformers for the success of the subsequent docking step. Furthermore, results by Castro-Alvarez *et al.* suggested that outdated docking methods, like AutoDock 3, achieved overall less accurate docking poses (mean RMSD of 1.46 Å)[116], making it less suitable for MCs compared to newer versions and other docking approaches including DynaDock.

Moreover, we included all seven small MC systems of the evaluation set of the previous pipeline in our benchmark to directly compare possible differences (see **Table 4**). While the old workflow already produced accurate docking poses below 2.0 Å for all systems (1.28 Å on average)[102], the optimized pipeline could improve all but one of those systems significantly. The average RMSD value over these seven systems employing the new pipeline was 0.71 Å[104]. The differences in both pipelines will be briefly outlined. In the old pipeline, MD-based conformational sampling of MC ligands was performed for 3x 250 ns at 370 K. Combined trajectories were clustered in Cartesian space obtaining five clusters. Representatives of these were used as starting structures for molecular docking calculations with DynaDock producing 2,000 initial broad sampling poses each. Of these, 250 poses per starting conformer were selected for OPMD refinement simulations. Successfully refined poses were clustered and ranked[102]. In contrast to that, our optimized ligand sampling parameters featured 3x 1,000 ns simulations at 600 K[104]. We could show that, for some MCs in our evaluation set, many new conformations were sampled within the first 1,000 ns (even after 250 ns) and that elevated temperatures enabled the sampling across torsional barriers leading to distinct MC conformations absent at 370 K. These optimized sampling conditions facilitated more exhaustive pre-sampling of conformers including the bioactive state of the MC inhibitors. Combined trajectories were classified in torsional space for more intuitive analysis and high-quality partitions. An ensemble size of 10 was used and each starting conformer produced 2,000 initial broad sampling poses that were clustered obtaining 100 remaining poses, which were then refined with OPMD. The last part of the refinement simulation was used for MMGBSA-based scoring and clustering, where the representative of the

highest populated cluster was appointed the final refined docking pose. These changes in our pipeline account for the detailed improvements in docking poses of six of the seven overlapping MC systems. Moreover, the new settings are generally applicable to MCs since we extended the data set for the benchmark study to cover a large variety of MCs including small, medium, and large MC rings with and without flexible substituents. Overall, MD pre-sampling and docking with DynaDock provided consistent and highly accurate results that rank among the best structural performances for this setup in holo structures.

18 out of 20 ligands in the study by Castro-Alvarez overlapped with compounds in the benchmark set by Lam *et al.*, which allows for a more direct performance comparison. For those, mean RMSD values of 1.4 Å, 1.59 Å, 2.54 Å, 1.4 Å and 1.35 Å for AutoDock Vina, AutoDock 4, AutoDock 3, DOCK and Glide were achieved[116], which were higher than for ICM-dock (1.2 Å)[126]. Thus, the internal coordinate scheme implemented in ICM constitutes a remarkable strategy with the capability of accurately sampling MC ring conformations during docking (further discussed in the following section), that also achieved the best performance with sub-angstrom accuracy in stage 1B of the blind docking challenge. In contrast to our and other benchmark studies, in this challenge only a single protein system was explored. Moreover, multiple co-crystallized structures existed in the PDB with very similar (template) ligands compared to those investigated in the challenge. For example, 340 complexes were found in the PDB that shared 95% sequence similarity to the receptor target sequence[174]. Furthermore, the structure of one target ligand only differed in a hydroxyl group compared to an existing co-crystallized structure[175]. Thus, many submissions and most of those with mean sub-angstrom accuracy relied on highly tailored protocols that incorporated tight system-specific constraints guiding the docking procedure. These included hydrogen bond constraints in the active site[176], restrained minimization[177], atomic property fields of template ligands[126], distance dependent penalties and filters that only included poses with significant overlap between common substructures of target and template ligands[118]. In addition, the entire algorithm of SkeleDock, a newly introduced docking web application, was designed for template-based procedures[178]. All these approaches require detailed information about the investigated system or a highly similar one. It can thus be assumed that their performances strongly depend on this knowledge. Therefore, it would be interesting for future challenges to stipulate MC systems with less accessible experimental information uncovering the limits of biased docking.

Interestingly, many results presented in studies that participated in the challenge agree well with key findings from our works: (i) the accuracy of ligand pre-sampling and the inclusion of bioactive-like MC (ring) conformers are crucial for successful docking[120, 175]. (ii) Structural errors in otherwise correctly placed poses could mainly be attributed to large flexible substituents of the ligand, not the MC ring[175]. This was also confirmed by other benchmark studies[99] and our own results, which further suggested that this was especially challenging for solvent-exposed binding sites[104]. (iii) During MD-based refinement of docked poses ligand moieties could adapt to the binding site, thereby overcoming weak initial placements resulting in higher accuracy[177].

Regarding our MD-heavy pipeline, it is worth further discussing the performance of this method employed by participants in the challenge. Sasmal *et al.* performed short post-docking MD simulations for 14 ns and evaluated possible changes of binding modes solely based on the first and last frames[175]. For 9 cases, the docked poses did not change significantly or slightly improved with respect to the bound reference. However, in the remaining 11 cases they drastically deteriorated. Overall, the mean RMSD value thus increased from 1.32 Å to 1.75 Å. Although a short minimization and equilibration of the initially docked poses were performed, relaxation of the complex could cause strong movements in the beginning of the production simulations. Furthermore, only considering the last frame of the simulation is arbitrary and not representative. Cluster analysis should have been performed to evaluate the most dominant structure of the simulation. In contrast to that, as part of our docking process, DynaDock performed a short OPMD simulation of each pose after initial placement to allow for induced-fit adaptions of the binding partners. The end of this refinement simulation was clustered to determine the final docking pose. From that, our long-term post-docking MD simulations carefully minimized, heated, and equilibrated the system. They furthermore stated different conditions between the reference structure (crystallized at low temperatures) and the docked complex (simulated at room temperature). We therefore also equilibrated the bound reference structure to assure similar treatment and conditions and to thus minimize possible structural artifacts from crystallization experiments. In our setup, a docking pose with an RMSD value of 0.0 Å to this equilibrated reference would be theoretically exact, allowing for more meaningful conclusions about the accuracy of the computational workflow.

Computational costs across these approaches were discussed before[104] and will be summarized herein. Compared to strategies where docking and/or conformer sampling

were performed by stochastic tools, our MD-driven pipeline was significantly more demanding in computation time and resources. However, such tools were often designed and optimized for speed, allowing fast screening approaches of many ligands of a database. In contrast, our workflow is focused on structural accuracy and thus more suitable for detailed predictions of few systems. Fast docking programs often generate many possible binding modes. The DynaDock strategy can be used for distinguishing and refining the true binding mode. We further highlighted that costly external generation of ligand conformations with MD is worth the investment, if the necessary computational resources are available.

4.2.2. Ligand flexibility

Important for dynamically complex molecules like MCs, ligand flexibility is mainly accounted for in two different ways regarding molecular docking: (i) generating ligand conformations by external sampling and selecting an ensemble of starting structures or (ii) intrinsically modifying ligand torsional DoFs during pose generation, called flexible ligand docking, which is challenging for MCs due to interdependence of ring bonds, as introduced earlier.

Tested in other benchmark studies, flexible ligand docking with AutoDock 4[116] and GOLD[124] produced worse results when directly compared to rigid docking of pre-generated conformers. GOLD's optional retrieval of template ring conformations from experimental CSD structures was inapplicable to larger MCs, since only rings consisting of less than 8 ring atoms were included. Alogheli *et al.* reported robustness for intrinsic sampling with Glide, which was comparable to other conformational sampling methods, and even outperformed Prime-MCS[99]. However, some rings (for example RAP) could not be processed by Glide and no docking poses could be obtained prohibiting unified treatment. Unfortunately, no explanation was provided for those cases. Other flexible ligand docking approaches like ICM-dock[126] or a newly GPU-adapted version of AutoDock[118] seemed more promising for MCs, producing accurate docking poses below 2.0 Å for ligands of the Grand Challenge 4, whereas the anchor-and-grow type of conformer generation algorithm implemented in DOCK seemed unsuitable[175]. However, another disadvantage of these methods is the lack of knowledge about the generated (ring) structures. Since conformations are sampled on-the-fly without optional output prior to docking, it cannot be judged if a bioactive-like ring conformer was generated apart from evaluating the final docking poses.

Pre-sampling of MC conformations in a separate step allows for more detailed structural analysis[177], but also faces the same challenges as discussed above. Moreover, in the context of subsequent docking, wrongly generated starting conformers from RDKit led to unsuccessful docking calculations[120]. In that case, the tool strongly favored *cis* configuration for an amide dihedral angle within the MC ring, despite the native *trans* configuration. The researchers could only force the right configuration with a highly tailored protocol, relying on human intervention and knowledge about the system. This result confirmed that poor external pre-sampling limits the performance of the docking step. OMEGA was another tool that showed significant problems with MC conformer generation during the Grand Challenge 4. Participants reported that an old version of the tool could not handle MCs well, missing near bioactive conformations, which decreased the docking accuracy[175]. Moreover, docking poses with high RMSD values originated from OMEGA conformers due to general sampling issues rather than the use of a conformer subset[174].

Thus, MC ligand flexibility is still not thoroughly accounted for in some flexible docking tools as well as in some stochastic conformational sampling tools, making this an active field of research[130]. Therefore, our optimized dynamics-based sampling and fully flexible docking pipeline meets the need for MC-specific methods and provides sufficient MC ligand flexibility.

4.2.3. Receptor flexibility

Abovementioned benchmark studies focused on rigid receptor dockings in the bound state of the protein. Such protocols are only useful for method evaluation and comparison and thus limited in their applicability. Accounting for receptor flexibility is crucial for more realistic setups and upcoming comparative studies should thus include apo states. Our benchmark study was the first of this kind focused on MC compounds[104], where we explicitly included receptor flexibility during OPMD refinement of the docking step. The significantly higher accuracy obtained by our approach even in the holo state compared to similar studies (see chapter above) suggests that incorporating receptor flexibility improves the docking performance also for the ligand-bound state.

Ensemble docking, as introduced earlier, is a common strategy to implicitly account for receptor flexibility during docking. A direct comparison of our approach to that strategy would be an interesting future study. For the Grand Challenge 4, biased docking approach with ICM-dock incorporated 14 receptor states in a 4D docking setup, which led to highly

accurate results[126]. In another study, all ligands were docked to handpicked ensemble of 11 receptor conformations[118]. As mentioned before, such approach is more costly when docking in a systematic fashion, and typically more suited for conformational selection phenomena, since each chosen receptor state is still treated rigidly during docking[128].

In contrast to that, for the submission by Basciu *et al.*[174] their previously published EDES approach[179] (ensemble docking with enhanced sampling) was applied. For that, three orthogonal inertia planes were defined across the binding site including its center of mass. The contacts between residues of each side of a plane and the Rgyr of the binding site were then employed as collective variables in MTD simulations. The latter was used to alter the shape of the binding site in a controlled manner throughout the simulation. With this, they generated 200 holo-like receptor structures *in silico* starting from apo states for subsequent unbiased ensemble docking calculations. Although their prediction results ranked worse compared to biased docking studies (stage 1A average minimum RMSD values of 2.48 Å and 2.28 Å for docking with AutoDock and HADDOCK, respectively), they achieved comparable results to docking in real holo states of the protein. Furthermore, their approach was generally disadvantaged by average-based rankings since it was intendedly designed to produce distinct shapes of the binding site. However, docking in the crystallized apo form of the protein resulted in poses with larger RMSD values of 3.78 Å (AutoDock), showing that incorporating receptor flexibility in that way significantly improved docking accuracy. A recent benchmark study developed a similar approach where apo structures were reliably converted to holo-like states employing MD simulations[180]. Docking in these refined structures achieved high accuracy with average RMSD values of 1.97 Å, comparable to self-docking and significantly lower compared to rigid apo docking results. These studies thus highly strengthened the validity of dynamics-based methods for incorporating receptor flexibility, agreeing with our own conclusions[104].

4.2.4. Evaluation of docking poses

Focusing on structural evaluation first, the RMSD value is still the most common metric, useful in benchmark studies, where structural accuracy is compared between different approaches. We intentionally chose a harsh cut-off value of 2.0 Å for docking success to show the high structural accuracy of our approach that we aimed for. However, such threshold is usually employed for typical small molecules[181], which is why other

studies employed more relaxed boundaries up to 2.5 Å[127] or 3.0 Å[124]. Large RMSD values can be obtained for minor errors even if most parts of the docked pose were predicted accurately. Some MCs are prone to this inherent weakness since their structural architecture comprises an MC core with multiple flexible substituents. Exemplarily demonstrated by Alogheli *et al.*, even if the core scaffold and two of the three extensions were placed correctly, the one wrongly predicted substituent caused an overall RMSD of 4.2 Å for that pose[99]. Most of our failed docking cases (JZB, JZC and NHN), could be explained in the same way[104], attributing these errors to the substituents rather than the MC ring. For conformational pre-sampling, we intendedly focused on the RMSD of ring atoms, since other studies showed that the RMSD of all heavy atoms can be insufficient in describing the ring conformation[96]. However, other studies included Rgyr to assess the extendedness or compactness of a conformation and torsional fingerprints to measure exhaustiveness of sampling or diversity of conformers[92, 182]. The latter is an interesting metric for ring structures, often employed when comparing structure generation methods and should be used in combination with MD simulations in future studies, since an arbitrary number of conformers can be generated by the latter without accounting for redundancy. For evaluating docking poses, the reproduction of native contacts might be a promising alternative to RMSD, as demonstrated by the MC peptide docking study of Zhang *et al*[183]. They exemplarily showed a docking pose where the protein-bound parts were docked correctly, while solvent-exposed extensions differed from the reference. Although reproducing 96% of native contacts, the pose obtained an RMSD value of 5.6 Å and would have been neglected by a strict RMSD cut-off. Thus, their metric could indeed overcome deficiencies of the RMSD like the dependence on structural alignment, which is of great importance when considering apo structures. Future studies should therefore consider this metric and test its potential to identify MC docking poses that might resemble the correct binding mode but obtained RMSD values larger than 2.0 Å due to solvent-exposed substituents. Such docking poses can also be evaluated by MD simulations in a post-processing step. We showed that docking poses quickly stabilized during the simulation, thereby confirming that the initial placement indeed resembled the bound state, even if the RMSD was slightly above 2.0 Å[104]. A similar strategy was employed in a docking protocol for antibody-carbohydrate complexes, where MD simulations were used to discriminate between real bound states and wrong docking poses produced by the Vina-CARB program[184]. Together, these results suggest that MD

simulations offer a useful way of evaluating docked poses over the course of a trajectory, instead of only considering a static pose based on a single RMSD value.

Besides structural measures docked poses can be evaluated energetically. Realistically, if the exact binding mode is unknown and no co-crystallized reference structure exists, scoring and ranking the docking poses based on their interaction energy might be the only consideration. Unfortunately, there is no ultimate scoring function that correlates exactly with structural measures. The development of suitable scoring functions is still an active field of research[185] and would thus exceed the scope of this work since we focused on structural accuracy. Nevertheless, we illustrated that MMGBSA is generally a suitable measure for ranking of MC docking complexes. A benchmark setup by Greenidge *et al.* has shown before that re-ranking poses docked by GOLD and Glide with MMGBSA in a post-processing step improved the identification of the correct pose, overcoming sensitivities of the scoring functions implemented in these programs[186]. Interestingly, implicit receptor flexibility accounted for by ensemble docking was indispensable for this pipeline. Furthermore, our workflow has the advantage that the trajectory of the OPMD simulations can be directly used for MMGBSA energy calculations (employing a single trajectory protocol) without the need of expensive post-processing simulations. In the Grand Challenge 4, affinity ranking and free energy prediction were additional tasks of both stages. Sasmal *et al.* reported near zero correlation between experimental rankings or values and MMGBSA, and thus argued that this method was neither suitable for pose ranking nor affinity prediction[175]. However, single trajectory MMGBSA calculations were performed with frames of a 10 ns production run after rigid receptor docking, during which some initially docked poses proved to be unstable. A connected study further highlighted the sensitivity of MMGBSA calculations to the initial conformation of the receptor and its protonation state[187]. Thus, this might be explained by significant structural rearrangements during the long simulation due to poor docking rather than scoring. Most likely, such poses produced unreliable scores which evened out the scores of stable poses. Therefore, we only used the frames of the last 5 ps of the refinement OPMD for MMGBSA calculations. This way, potential fluctuations of the beginning of such simulations due to induced-fit of the initial placement and the flexible receptor were excluded. Taking all results into account, we thus argue that MMGBSA calculations might be more meaningful for ranking of high-quality docking poses that resemble stable binding modes. Associated with that, we strongly advocate the incorporation of receptor

flexibility during docking to assure the generation of reliable poses which allow for improved scoring.

5. Conclusion

In this work the development of a suitable strategy for successful molecular modeling of MCs was presented. We first explored and optimized MD-based sampling of MC ligands. Our early case study of the MC tetrandrine and iso-tetrandrine intermediates revealed that few specific ring torsions determined the compound's overall conformation[155]. Detailed analysis of dihedral angles of such torsions helped to understand the distribution of sampled conformers throughout the simulation which could be correlated to experimental observations. This in-depth investigation eventually led to the development of a torsion-based classification scheme for MD-sampled conformations of MCs (*ClassTor.py*)[104].

In two following case studies, molecular docking of MCs was performed where different parameters could be tested to obtain structurally accurate docking poses[147, 156]. In the first, acyldepsipeptide derivatives were docked to caseinolytic protease P. Docking parameters could be validated for a test case of the parent compound, where a co-crystallized structure was available. Thus optimized, the docking of the investigated compounds predicted accurate binding modes which could explain major experimentally observed differences. In the second study, the miuraenamide derivative LK701 was studied. Earlier studies could determine the binding mode of the parent compound miuraenamide bound to an actin trimer nucleus and structurally explain its filament-stabilizing effects[188]. The derived compounds, however, showed an opposing biological outcome since less filaments were formed after LK701 treatment[156]. Our extensive computational investigation showed that LK701 binds to the actin monomer and cannot penetrate and bind an existing unbound nucleus. Blocking important longitudinal interaction sites in the macrolide binding cleft of single actin subunits, binding of LK701 likely prevents actin monomers from incorporation into existing nuclei and filaments. Our docking calculations confirmed that the minor structural differences between LK701 and its parent miuraenamide led to distinct binding modes explaining the observed differences in the experiments.

These three detailed case studies provided the basis for the final benchmark study, where we combined conformational sampling and molecular docking in a generally applicable pipeline for molecular modeling of MCs. The final workflow was based on a previous strategy in our group that was tested for seven small MCs[102]. We extended the data set to 20 highly diverse MC compounds ranging from 11 to 29 MC ring atoms with or without

large flexible substituents and further included proteins in their holo and apo states. We found suitable conditions for the ligand simulations accounting for their complex topology and dynamics. We developed a dihedral classification scheme to rationally analyze the sampled ring conformations which we used for the extraction of conformer ensembles. Settings of the docking procedure with DynaDock could be optimized and thus produced highly accurate and refined binding modes.

This pipeline thus solved some of the major structural challenges in molecular modeling of MCs. Missing ring flexibility during broad sampling of molecular docking calculations could be compensated by a conformer ensemble generated by MD-based ligand pre-sampling. Here, the extensiveness of ligand MD simulations was crucial for high quality structural ensembles that included a bioactive-like state of the MC rings which ultimately increased the chance for successful docking. Furthermore, during the pose refinement OPMD simulations MC rings could adapt conformation to the presence of the protein, corresponding to a switch from unbound to bound state in some cases. The refinement allowed for incorporation of full protein flexibility which enabled explicit simulation of dynamic phenomena like induced-fit binding, improving docking results in holo and apo states.

Insights from our studies suggested that large flexible substituents on the ring scaffold in combination with solvent-exposed binding sites on the protein remained troublesome and should be tackled in future studies. There, it should be investigated if small changes to our existing pipeline solve these remaining issues or if larger modifications of the strategy are necessary. In case of the former, it must be shown if simply increasing the number of initially generated broad sampling poses or modifications in torsional DoFs would be sufficient to accurately account for the larger conformational space of these flexible moieties. In case of the latter, a fragmentation-based docking approach could be an interesting alternative, where first the ring scaffold is placed, while adding the substituents afterwards. However, this entirely different procedure would pose new questions and might overly complicate the issue. In this regard, upcoming AI-based docking procedures could also be considered for MCs. An easier solution without fragmentation might be a general docking with our pipeline followed by additional refinements of only the substituents to allow for further adaption in these regions of the MC compounds. Moreover, alternative structural measures despite RMSD values must be tested for this issue, for example the reproduction of native contacts. Such measures could distinguish protein-bound parts of the compounds from solvent-exposed regions and

reveal their importance for accurate structural reproduction under consideration of their sophisticated binding modes. Furthermore, detailed investigations of scoring functions are necessary for MC compounds which could reveal the need of MC-specific functions.
Concluding, studies in this work provided helpful solutions for molecular modeling of MCs. With our sophisticated pipeline that incorporates high level of molecular flexibility, the dynamics and the complexity of the MC ring was accounted for in detail. Thereby, we redirected the focus in modeling of larger MC ligands to their flexible substituents bound to solvent-exposed binding sites, which remain challenging for molecular docking methods.

References

[1] McNaught, A.D. and A. Wilkinson, *Compendium of chemical terminology*. Vol. 1669. 1997: Blackwell Science Oxford.

[2] Driggers, E.M., S.P. Hale, J. Lee, and N.K. Terrett, *The exploration of macrocycles for drug discovery—an underexploited structural class.* Nature Reviews Drug Discovery, 2008. **7**(7): p. 608-624.

[3] Peterson, M., *The Evolution of Macrocycles in Drug Discovery: from Technologies to Drugs.* Am. Pharma. Rev., 2017: p. 343609.

[4] Levin, J.I., *Macrocycles in drug discovery*. Vol. 40. 2015: Royal Society of Chemistry.

[5] Marsault, E. and M.L. Peterson, *Practical Medicinal Chemistry with Macrocycles: Design, Synthesis, and Case Studies*. 2017: John Wiley & Sons.

[6] Williams, D.H., M.J. Stone, P.R. Hauck, and S.K. Rahman, *Why are secondary metabolites (natural products) biosynthesized?* Journal of natural products, 1989. **52**(6): p. 1189-1208.

[7] Stone, M. and D. Williams, *On the evolution of functional secondary metabolites (natural products).* Molecular microbiology, 1992. **6**(1): p. 29-34.

[8] DeBoer, C., P. Meulman, R. Wnuk, and D. Peterson, *Geldanamycin, a new antibiotic.* The Journal of antibiotics, 1970. **23**(9): p. 442-447.

[9] Whitesell, L., E.G. Mimnaugh, B. De Costa, C.E. Myers, and L.M. Neckers, *Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation.* Proceedings of the National Academy of Sciences, 1994. **91**(18): p. 8324-8328.

[10] Stebbins, C.E., A.A. Russo, C. Schneider, N. Rosen, F.U. Hartl, and N.P. Pavletich, *Crystal structure of an Hsp90–geldanamycin complex: targeting of a protein chaperone by an antitumor agent.* Cell, 1997. **89**(2): p. 239-250.

[11] Prodromou, C., S.M. Roe, R. O'Brien, J.E. Ladbury, P.W. Piper, and L.H. Pearl, *Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone.* Cell, 1997. **90**(1): p. 65-75.

[12] Delmotte, P. and J. Delmotte-Plaquee, *A new antifungal substance of fungal origin.* Nature, 1953. **171**(4347): p. 344-344.

[13] Roe, S.M., C. Prodromou, R. O'Brien, J.E. Ladbury, P.W. Piper, and L.H. Pearl, *Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin.* Journal of medicinal chemistry, 1999. **42**(2): p. 260-266.

[14] Dikalov, S., U. Landmesser, and D.G. Harrison, *Geldanamycin leads to superoxide formation by enzymatic and non-enzymatic redox cycling: implications for studies of Hsp90 and endothelial cell nitric-oxide synthase.* Journal of Biological Chemistry, 2002. **277**(28): p. 25480-25485.

[15] Supko, J.G., R.L. Hickman, M.R. Grever, and L. Malspeis, *Preclinical pharmacologic evaluation of geldanamycin as an antitumor agent.* Cancer chemotherapy and pharmacology, 1995. **36**(4): p. 305-315.

[16] Cysyk, R.L., R.J. Parker, J.J. Barchi, P.S. Steeg, N.R. Hartman, and J.M. Strong, *Reaction of geldanamycin and C17-substituted analogues with glutathione: product identifications and pharmacological implications.* Chemical research in toxicology, 2006. **19**(3): p. 376-381.

[17] Schulte, T.W. and L.M. Neckers, *The benzoquinone ansamycin 17-allylamino-17-demethoxygeldanamycin binds to HSP90 and shares important biologic activities*

with geldanamycin. Cancer chemotherapy and pharmacology, 1998. **42**(4): p. 273-279.
[18] Schnur, R., M. Corman, R. Gallaschun, B. Cooper, M. Dee, J. Doty, M. Muzzi, J. Moyer, and C. DiOrio, *Inhibition of the oncogene product p185erbB-2 in vitro and in vivo by geldanamycin and dihydrogeldanamycin derivatives.* Journal of medicinal chemistry, 1995. **38**(19): p. 3806-3812.
[19] Glaze, E.R., A.L. Lambert, A.C. Smith, J.G. Page, W.D. Johnson, D.L. McCormick, A.P. Brown, B.S. Levine, J.M. Covey, and M.J. Egorin, *Preclinical toxicity of a geldanamycin analog, 17-(dimethylaminoethylamino)-17-demethoxygeldanamycin (17-DMAG), in rats and dogs: potential clinical relevance.* Cancer chemotherapy and pharmacology, 2005. **56**(6): p. 637-647.
[20] Yamamoto, K., R.M. Garbaccio, S.J. Stachel, D.B. Solit, G. Chiosis, N. Rosen, and S.J. Danishefsky, *Total synthesis as a resource in the discovery of potentially valuable antitumor agents: cycloproparadicicol.* Angewandte Chemie, 2003. **115**(11): p. 1318-1322.
[21] Kyle Hadden, M., D.J. Lubbers, and B.S. J Blagg, *Geldanamycin, radicicol, and chimeric inhibitors of the Hsp90 Nterminal ATP binding site.* Current topics in medicinal chemistry, 2006. **6**(11): p. 1173-1182.
[22] Wang, M., G. Shen, and B.S. Blagg, *Radanamycin, a macrocyclic chimera of radicicol and geldanamycin.* Bioorganic & medicinal chemistry letters, 2006. **16**(9): p. 2459-2462.
[23] Vezina, C., A. Kudelski, and S. Sehgal, *Rapamycin (AY-22, 989), a new antifungal antibiotic I. taxonomy of the producing streptomycete and isolation of the active principle.* The Journal of antibiotics, 1975. **28**(10): p. 721-726.
[24] Sehgal, S.N., *Rapamune®(RAPA, rapamycin, sirolimus): mechanism of action immunosuppressive effect results from blockade of signal transduction and inhibition of cell cycle progression.* Clinical biochemistry, 1998. **31**(5): p. 335-340.
[25] Bierer, B.E., P.S. Mattila, R.F. Standaert, L.A. Herzenberg, S. Burakoff, G. Crabtree, and S.L. Schreiber, *Two distinct signal transmission pathways in T lymphocytes are inhibited by complexes formed between an immunophilin and either FK506 or rapamycin.* Proceedings of the National Academy of Sciences, 1990. **87**(23): p. 9231-9235.
[26] Wiederrecht, G., E. Lam, S. Hung, M. Martin, and N. Sigal, *The mechanism of action of FK-506 and cyclosporin A.* Annals of the New York Academy of Sciences, 1993. **696**: p. 9-19.
[27] Sehgal, S. *Sirolimus: its discovery, biological properties, and mechanism of action*. in *Transplantation proceedings*. 2003. Elsevier.
[28] Laplante, M. and D.M. Sabatini, *mTOR signaling in growth control and disease.* cell, 2012. **149**(2): p. 274-293.
[29] U.S. Food and Drug Administration. *Rapamune*. 1999 [cited 2022 July 11]; Available from: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=021083.
[30] European Medicines Agency. *Rapamune*. 2001 [cited 2022 July 11]; Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/rapamune.
[31] U.S. Food and Drug Administration. *Fyarro*. 2021 [cited 2022 August 26]; Available from: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=213312.

[32] U.S. Food and Drug Administration. *Hyftor*. 2022 [cited 2022 August 26]; Available from: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=213478.
[33] U.S. Food and Drug Administration. *Torisel*. 2007 [cited 2022 August 28]; Available from: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=022088.
[34] U.S. Food and Drug Administration. *Afinitor*. 2009 [cited 2022 August 28]; Available from: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=022334.
[35] Kim, Y.-k., M.A. Arai, T. Arai, J.O. Lamenzo, E.F. Dean, N. Patterson, P.A. Clemons, and S.L. Schreiber, *Relationship of stereochemical and skeletal diversity of small molecules to cellular measurement space.* Journal of the American Chemical Society, 2004. **126**(45): p. 14740-14745.
[36] Marsault, E. and M.L. Peterson, *Macrocycles are great cycles: applications, opportunities, and challenges of synthetic macrocycles in drug discovery.* Journal of medicinal chemistry, 2011. **54**(7): p. 1961-2004.
[37] Villar, E.A., D. Beglov, S. Chennamadhavuni, J.A. Porco, D. Kozakov, S. Vajda, and A. Whitty, *How proteins bind macrocycles.* Nature chemical biology, 2014. **10**(9): p. 723-731.
[38] Cummings, M.D., J. Lindberg, T.I. Lin, H. de Kock, O. Lenz, E. Lilja, S. Felländer, V. Baraznenok, S. Nyström, and M. Nilsson, *Induced-fit binding of the macrocyclic noncovalent inhibitor TMC435 to its HCV NS3/NS4A protease target.* Angewandte Chemie International Edition, 2010. **49**(9): p. 1652-1655.
[39] Berger, A. and I. Schechter, *Mapping the active site of papain with the aid of peptide substrates and inhibitors.* Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1970. **257**(813): p. 249-264.
[40] Schlünzen, F., R. Zarivach, J. Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath, and F. Franceschi, *Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria.* Nature, 2001. **413**(6858): p. 814-821.
[41] Lipinski, C.A., F. Lombardo, B.W. Dominy, and P.J. Feeney, *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Advanced drug delivery reviews, 2012. **64**: p. 4-17.
[42] Adessi, C. and C. Soto, *Converting a peptide into a drug: strategies to improve stability and bioavailability.* Current medicinal chemistry, 2002. **9**(9): p. 963-978.
[43] Rezai, T., J.E. Bock, M.V. Zhou, C. Kalyanaraman, R.S. Lokey, and M.P. Jacobson, *Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides.* Journal of the American Chemical Society, 2006. **128**(43): p. 14073-14080.
[44] Danelius, E., V. Poongavanam, S. Peintner, L.H. Wieske, M. Erdélyi, and J. Kihlberg, *Solution conformations explain the chameleonic behaviour of macrocyclic drugs.* Chemistry–A European Journal, 2020. **26**(23): p. 5231-5244.
[45] Llinas-Brunet, M., M. Bailey, G. Fazal, S. Goulet, T. Halmos, S. Laplante, R. Maurice, M. Poirier, M.-A. Poupart, and D. Thibeault, *Peptide-based inhibitors of the hepatitis C virus serine protease.* Bioorganic & Medicinal Chemistry Letters, 1998. **8**(13): p. 1713-1718.

[46] Llinàs-Brunet, M., M. Bailey, R. Déziel, G. Fazal, V. Gorys, S. Goulet, T. Halmos, R. Maurice, M. Poirier, and M.-A. Poupart, *Studies on the C-terminal of hexapeptide inhibitors of the hepatitis C virus serine protease.* Bioorganic & Medicinal Chemistry Letters, 1998. **8**(19): p. 2719-2724.

[47] LaPlante, S.R., D.R. Cameron, N. Aubry, S. Lefebvre, G. Kukolj, R. Maurice, D. Thibeault, D. Lamarre, and M. Llinas-Brunet, *Solution structure of substrate-based ligands when bound to hepatitis C virus NS3 protease domain.* Journal of Biological Chemistry, 1999. **274**(26): p. 18618-18624.

[48] Llinàs-Brunet, M., M. Bailey, G. Fazal, E. Ghiro, V. Gorys, S. Goulet, T. Halmos, R. Maurice, M. Poirier, and M.-A. Poupart, *Highly potent and selective peptide-based inhibitors of the hepatitis C virus serine protease: towards smaller inhibitors.* Bioorganic & medicinal chemistry letters, 2000. **10**(20): p. 2267-2270.

[49] Tsantrizos, Y.S., G. Bolger, P. Bonneau, D.R. Cameron, N. Goudreau, G. Kukolj, S.R. LaPlante, M. Llinàs-Brunet, H. Nar, and D. Lamarre, *Macrocyclic inhibitors of the NS3 protease as potential therapeutic agents of hepatitis C virus infection.* Angewandte Chemie, 2003. **115**(12): p. 1394-1398.

[50] Goudreau, N., D.R. Cameron, P. Bonneau, V. Gorys, C. Plouffe, M. Poirier, D. Lamarre, and M. Llinas-Brunet, *NMR structural characterization of peptide inhibitors bound to the Hepatitis C virus NS3 protease: design of a new P2 substituent.* Journal of medicinal chemistry, 2004. **47**(1): p. 123-132.

[51] Llinas-Brunet, M., M.D. Bailey, G. Bolger, C. Brochu, A.-M. Faucher, J.M. Ferland, M. Garneau, E. Ghiro, V. Gorys, and C. Grand-Maître, *Structure– activity study on a novel series of macrocyclic inhibitors of the hepatitis C virus NS3 protease leading to the discovery of BILN 2061.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1605-1608.

[52] LaPlante, S.R. and M. Llinàs-Brunet, *Dynamics and structure-based design of drugs targeting the critical serine protease of the hepatitis C virus-from a peptidic substrate to BILN 2061.* Current Medicinal Chemistry-Anti-Infective Agents, 2005. **4**(2): p. 111-132.

[53] Lamarre, D., P.C. Anderson, M. Bailey, P. Beaulieu, G. Bolger, P. Bonneau, M. Bös, D.R. Cameron, M. Cartier, and M.G. Cordingley, *An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus.* Nature, 2003. **426**(6963): p. 186-189.

[54] Koshland Jr, D., *Application of a theory of enzyme specificity to protein synthesis.* Proceedings of the National Academy of Sciences, 1958. **44**(2): p. 98-104.

[55] Nicklaus, M.C., S. Wang, J.S. Driscoll, and G.W. Milne, *Conformational changes of small molecules binding to proteins.* Bioorganic & medicinal chemistry, 1995. **3**(4): p. 411-428.

[56] Perola, E. and P.S. Charifson, *Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding.* Journal of medicinal chemistry, 2004. **47**(10): p. 2499-2510.

[57] Skrzypczak, N., K. Pyta, P. Ruszkowski, M. Gdaniec, F. Bartl, and P. Przybylski, *Synthesis, structure and anticancer activity of new geldanamycin amine analogs containing C (17)-or C (20)-flexible and rigid arms as well as closed or open ansa-bridges.* European Journal of Medicinal Chemistry, 2020. **202**: p. 112624.

[58] Cutler, H.G., R.F. Arrendale, J.P. Springer, P.D. Cole, R.G. Roberts, and R.T. Hanlin, *Monorden from a novel source, Neocosmospora tenuicristata: Stereochemistry and plant growth regulatory properties.* Agricultural and biological chemistry, 1987. **51**(12): p. 3331-3338.

[59] Moulin, E., V. Zoete, S. Barluenga, M. Karplus, and N. Winssinger, *Design, synthesis, and biological evaluation of HSP90 inhibitors based on conformational analysis of radicicol and its analogues.* Journal of the American Chemical Society, 2005. **127**(19): p. 6999-7004.
[60] Loosli, H.R., H. Kessler, H. Oschkinat, H.P. Weber, T.J. Petcher, and A. Widmer, *Peptide conformations. Part 31. The conformation of cyclosporin A in the crystal and in solution.* Helvetica Chimica Acta, 1985. **68**(3): p. 682-704.
[61] Mikol, V., J. Kallen, G. Pflügl, and M.D. Walkinshaw, *X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2· 1 Å resolution.* Journal of molecular biology, 1993. **234**(4): p. 1119-1130.
[62] Thepchatri, P., T. Eliseo, D.O. Cicero, D. Myles, and J.P. Snyder, *Relationship among ligand conformations in solution, in the solid state, and at the Hsp90 binding site: geldanamycin and radicicol.* Journal of the American Chemical Society, 2007. **129**(11): p. 3127-3134.
[63] Corbett, K.M., L. Ford, D.B. Warren, C.W. Pouton, and D.K. Chalmers, *Cyclosporin Structure and Permeability: From A to Z and Beyond.* Journal of Medicinal Chemistry, 2021. **64**(18): p. 13131-13151.
[64] Phan, J., Z.-D. Shi, T.R. Burke Jr, and D.S. Waugh, *Crystal structures of a high-affinity macrocyclic peptide mimetic in complex with the Grb2 SH2 domain.* Journal of molecular biology, 2005. **353**(1): p. 104-115.
[65] Bogdan, A.R., S.V. Jerome, K. Houk, and K. James, *Strained cyclophane macrocycles: Impact of progressive ring size reduction on synthesis and structure.* Journal of the American Chemical Society, 2012. **134**(4): p. 2127-2138.
[66] Hawkins, P.C., *Conformation generation: the state of the art.* Journal of chemical information and modeling, 2017. **57**(8): p. 1747-1756.
[67] Schwab, C.H., *Conformations and 3D pharmacophore searching.* Drug Discovery Today: Technologies, 2010. **7**(4): p. e245-e253.
[68] Sprague, P.W. and R. Hoffmann, *CATALYST pharmacophore models and their utility as queries for searching 3D databases.* Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry, 1997: p. 223-240.
[69] Smellie, A., S.L. Teig, and P. Towbin, *Poling: promoting conformational variation.* Journal of Computational Chemistry, 1995. **16**(2): p. 171-187.
[70] Li, J., T. Ehlers, J. Sutter, S. Varma-O'Brien, and J. Kirchmair, *CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration.* Journal of chemical information and modeling, 2007. **47**(5): p. 1923-1932.
[71] ULC, C.C.G., *Molecular Operating Environment (MOE)*. 2022.
[72] Ferguson, D.M. and D.J. Raber, *A new approach to probing conformational space with molecular mechanics: random incremental pulse search.* Journal of the American Chemical Society, 1989. **111**(12): p. 4371-4378.
[73] Sadowski, J., J. Gasteiger, and G. Klebe, *Comparison of automatic three-dimensional model builders using 639 X-ray structures.* Journal of chemical information and computer sciences, 1994. **34**(4): p. 1000-1008.
[74] Renner, S., C.H. Schwab, J. Gasteiger, and G. Schneider, *Impact of conformational flexibility on three-dimensional similarity searching using correlation vectors.* Journal of chemical information and modeling, 2006. **46**(6): p. 2324-2332.
[75] Hawkins, P.C., A.G. Skillman, G.L. Warren, B.A. Ellingson, and M.T. Stahl, *Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database.* Journal of chemical information and modeling, 2010. **50**(4): p. 572-584.

[76] Vainio, M.J. and M.S. Johnson, *Generating conformer ensembles using a multiobjective genetic algorithm.* Journal of chemical information and modeling, 2007. **47**(6): p. 2462-2474.
[77] Mekenyan, O., T. Pavlov, V. Grancharov, M. Todorov, P. Schmieder, and G. Veith, *2D-3D migration of large chemical inventories with conformational multiplication. Application of the genetic algorithm.* Journal of chemical information and modeling, 2005. **45**(2): p. 283-292.
[78] Poli, G., T. Seidel, and T. Langer, *Conformational sampling of small molecules with iCon: Performance assessment in comparison with OMEGA.* Frontiers in chemistry, 2018. **6**: p. 229.
[79] O'Boyle, N.M., T. Vandermeersch, C.J. Flynn, A.R. Maguire, and G.R. Hutchison, *Confab-Systematic generation of diverse low-energy conformers.* Journal of cheminformatics, 2011. **3**(1): p. 1-9.
[80] Huang, N., B.K. Shoichet, and J.J. Irwin, *Benchmarking sets for molecular docking.* Journal of medicinal chemistry, 2006. **49**(23): p. 6789-6801.
[81] Watts, K.S., P. Dalal, R.B. Murphy, W. Sherman, R.A. Friesner, and J.C. Shelley, *ConfGen: a conformational search method for efficient generation of bioactive conformers.* Journal of chemical information and modeling, 2010. **50**(4): p. 534-546.
[82] Mohamadi, F., N.G. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, and W.C. Still, *Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics.* Journal of Computational Chemistry, 1990. **11**(4): p. 440-467.
[83] Kolossváry, I. and W.C. Guida, *Low mode search. An efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides.* Journal of the American Chemical Society, 1996. **118**(21): p. 5011-5019.
[84] Boström, J., *Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools.* Journal of Computer-Aided Molecular Design, 2001. **15**(12): p. 1137-1152.
[85] Chen, I.-J. and N. Foloppe, *Conformational sampling of druglike molecules with MOE and catalyst: implications for pharmacophore modeling and virtual screening.* Journal of chemical information and modeling, 2008. **48**(9): p. 1773-1791.
[86] Ebejer, J.-P., G.M. Morris, and C.M. Deane, *Freely available conformer generation methods: how good are they?* Journal of chemical information and modeling, 2012. **52**(5): p. 1146-1158.
[87] Friedrich, N.-O., A. Meyder, C. de Bruyn Kops, K. Sommer, F. Flachsenberg, M. Rarey, and J. Kirchmair, *High-quality dataset of protein-bound ligand conformations and its application to benchmarking conformer ensemble generators.* Journal of Chemical Information and Modeling, 2017. **57**(3): p. 529-539.
[88] Friedrich, N.-O., C. de Bruyn Kops, F. Flachsenberg, K. Sommer, M. Rarey, and J. Kirchmair, *Benchmarking commercial conformer ensemble generators.* Journal of chemical information and modeling, 2017. **57**(11): p. 2719-2728.
[89] Watts, K.S., P. Dalal, A.J. Tebben, D.L. Cheney, and J.C. Shelley, *Macrocycle conformational sampling with MacroModel.* Journal of chemical information and modeling, 2014. **54**(10): p. 2680-2696.
[90] Coutsias, E.A., K.W. Lexa, M.J. Wester, S.N. Pollock, and M.P. Jacobson, *Exhaustive conformational sampling of complex fused ring macrocycles using inverse kinematics.* Journal of chemical theory and computation, 2016. **12**(9): p. 4674-4687.

[91] Cleves, A.E. and A.N. Jain, *ForceGen 3D structure and conformer generation: from small lead-like molecules to macrocyclic drugs.* Journal of computer-aided molecular design, 2017. **31**(5): p. 419-439.
[92] Sindhikara, D., S.A. Spronk, T. Day, K. Borrelli, D.L. Cheney, and S.L. Posy, *Improving accuracy, diversity, and speed with prime macrocycle conformational sampling.* Journal of chemical information and modeling, 2017. **57**(8): p. 1881-1894.
[93] Barbeau, X., A.T. Vincent, and P. Lagüe, *ConfBuster: Open-source tools for macrocycle conformational search and analysis.* Journal of Open Research Software, 2018. **6**(1).
[94] Cole, J.C., O. Korb, P. McCabe, M.G. Read, and R. Taylor, *Knowledge-based conformer generation using the Cambridge structural database.* Journal of chemical information and modeling, 2018. **58**(3): p. 615-629.
[95] Friedrich, N.-O., F. Flachsenberg, A. Meyder, K. Sommer, J. Kirchmair, and M. Rarey, *Conformator: a novel method for the generation of conformer ensembles.* Journal of chemical information and modeling, 2019. **59**(2): p. 731-742.
[96] Reyes Romero, A., A.J. Ruiz-Moreno, M.R. Groves, M. Velasco-Velázquez, and A. Dömling, *Benchmark of Generic Shapes for Macrocycles.* Journal of chemical information and modeling, 2020. **60**(12): p. 6298-6313.
[97] Bonnet, P., D.K. Agrafiotis, F. Zhu, and E. Martin, *Conformational analysis of macrocycles: finding what common search methods miss.* Journal of chemical information and modeling, 2009. **49**(10): p. 2242-2259.
[98] Diaz, D.B., S.D. Appavoo, A.F. Bogdanchikova, Y. Lebedev, T.J. McTiernan, G. dos Passos Gomes, and A.K. Yudin, *Illuminating the dark conformational space of macrocycles using dominant rotors.* Nature Chemistry, 2021. **13**(3): p. 218-225.
[99] Alogheli, H., G. Olanders, W. Schaal, P. Brandt, and A. Karlén, *Docking of macrocycles: comparing rigid and flexible docking in glide.* Journal of chemical information and modeling, 2017. **57**(2): p. 190-202.
[100] Claeys, D.D., T. Verstraelen, E. Pauwels, C.V. Stevens, M. Waroquier, and V.V. Speybroeck, *Conformational sampling of macrocyclic alkenes using a kennard–stone-based algorithm.* The Journal of Physical Chemistry A, 2010. **114**(25): p. 6879-6887.
[101] Witek, J., B.G. Keller, M. Blatter, A. Meissner, T. Wagner, and S. Riniker, *Kinetic models of cyclosporin A in polar and apolar environments reveal multiple congruent conformational states.* Journal of Chemical Information and Modeling, 2016. **56**(8): p. 1547-1562.
[102] Ugur, I., M. Schroft, A. Marion, M. Glaser, and I. Antes, *Predicting the bioactive conformations of macrocycles: a molecular dynamics-based docking procedure with DynaDock.* Journal of Molecular Modeling, 2019. **25**(7): p. 197.
[103] Seep, L., A. Bonin, K. Meier, H. Diedam, and A.H. Göller, *Ensemble completeness in conformer sampling: the case of small macrocycles.* Journal of cheminformatics, 2021. **13**(1): p. 1-22.
[104] Meixner, M., M. Zachmann, S. Metzler, J. Scheerer, M. Zacharias, and I. Antes, *Dynamic Docking of Macrocycles in Bound and Unbound Protein Structures with DynaDock.* Journal of Chemical Information and Modeling, 2022.
[105] Grimme, S., *Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations.* Journal of chemical theory and computation, 2019. **15**(5): p. 2847-2862.
[106] Gutten, O., D. Bim, J. Rezac, and L. Rulisek, *Macrocycle conformational sampling by DFT-D3/COSMO-RS methodology.* Journal of Chemical Information and Modeling, 2018. **58**(1): p. 48-60.

[107] Chen, I.-J. and N. Foloppe, *Tackling the conformational sampling of larger flexible compounds and macrocycles in pharmacology and drug discovery.* Bioorganic & medicinal chemistry, 2013. **21**(24): p. 7898-7920.
[108] Poongavanam, V., E. Danelius, S. Peintner, L. Alcaraz, G. Caron, M.D. Cummings, S. Wlodek, M. Erdelyi, P.C. Hawkins, and G. Ermondi, *Conformational sampling of macrocyclic drugs in different environments: can we find the relevant conformations?* ACS omega, 2018. **3**(9): p. 11742-11757.
[109] Olanders, G., H. Alogheli, P. Brandt, and A. Karlén, *Conformational analysis of macrocycles: comparing general and specialized methods.* Journal of computer-aided molecular design, 2020. **34**(3): p. 231-252.
[110] Thevenet, P., Y. Shen, J. Maupetit, F. Guyon, P. Derreumaux, and P. Tuffery, *PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides.* Nucleic acids research, 2012. **40**(W1): p. W288-W293.
[111] Singh, S., H. Singh, A. Tuknait, K. Chaudhary, B. Singh, S. Kumaran, and G.P. Raghava, *PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues.* Biology direct, 2015. **10**(1): p. 1-19.
[112] Jusot, M., D. Stratmann, M. Vaisset, J. Chomilier, and J. Cortés, *Exhaustive exploration of the conformational landscape of small cyclic peptides using a robotics approach.* Journal of chemical information and modeling, 2018. **58**(11): p. 2355-2368.
[113] Pagadala, N.S., K. Syed, and J. Tuszynski, *Software for molecular docking: a review.* Biophysical reviews, 2017. **9**(2): p. 91-102.
[114] Wang, Z., H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, and T. Hou, *Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power.* Physical Chemistry Chemical Physics, 2016. **18**(18): p. 12964-12975.
[115] Lang, P.T., S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, R.C. Rizzo, D.A. Case, T.L. James, and I.D. Kuntz, *DOCK 6: Combining techniques to model RNA–small molecule complexes.* Rna, 2009. **15**(6): p. 1219-1230.
[116] Castro-Alvarez, A., A.M. Costa, and J. Vilarrasa, *The performance of several docking programs at reproducing protein–macrolide-like crystal structures.* Molecules, 2017. **22**(1): p. 136.
[117] Morris, G.M., R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, and A.J. Olson, *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.* Journal of computational chemistry, 2009. **30**(16): p. 2785-2791.
[118] Santos-Martins, D., J. Eberhardt, G. Bianco, L. Solis-Vasquez, F.A. Ambrosio, A. Koch, and S. Forli, *D3R Grand Challenge 4: prospective pose prediction of BACE1 ligands with AutoDock-GPU.* Journal of computer-aided molecular design, 2019. **33**(12): p. 1071-1081.
[119] Eberhardt, J., D. Santos-Martins, A.F. Tillack, and S. Forli, *AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings.* Journal of Chemical Information and Modeling, 2021. **61**(8): p. 3891-3898.
[120] Kadukova, M., V. Chupin, and S. Grudinin, *Docking rigid macrocycles using Convex-PL, AutoDock Vina, and RDKit in the D3R Grand Challenge 4.* Journal of computer-aided molecular design, 2020. **34**(2): p. 191-200.
[121] Friesner, R.A., J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, and J.K. Perry, *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* Journal of medicinal chemistry, 2004. **47**(7): p. 1739-1749.

[122] Sherman, W., T. Day, M.P. Jacobson, R.A. Friesner, and R. Farid, *Novel procedure for modeling ligand/receptor induced fit effects.* Journal of medicinal chemistry, 2006. **49**(2): p. 534-553.
[123] Jones, G., P. Willett, and R.C. Glen, *Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation.* Journal of molecular biology, 1995. **245**(1): p. 43-53.
[124] Martin, S.J., I.-J. Chen, A.E. Chan, and N. Foloppe, *Modelling the binding mode of macrocycles: Docking and conformational sampling.* Bioorganic & medicinal chemistry, 2020. **28**(1): p. 115143.
[125] Abagyan, R., M. Totrov, and D. Kuznetsov, *ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation.* Journal of computational chemistry, 1994. **15**(5): p. 488-506.
[126] Lam, P.C.-H., R. Abagyan, and M. Totrov, *Macrocycle modeling in ICM: benchmarking and evaluation in D3R Grand Challenge 4.* Journal of Computer-Aided Molecular Design, 2019. **33**(12): p. 1057-1069.
[127] Anighoro, A., A.d.l.V. de León, and J. Bajorath, *Predicting bioactive conformations and binding modes of macrocycles.* Journal of computer-aided molecular design, 2016. **30**(10): p. 841-849.
[128] Antunes, D.A., D. Devaurs, and L.E. Kavraki, *Understanding the challenges of protein flexibility in drug design.* Expert opinion on drug discovery, 2015. **10**(12): p. 1301-1313.
[129] Jones, G., P. Willett, R.C. Glen, A.R. Leach, and R. Taylor, *Development and validation of a genetic algorithm for flexible docking.* Journal of molecular biology, 1997. **267**(3): p. 727-748.
[130] Allen, S.E., N.V. Dokholyan, and A.A. Bowers, *Dynamic docking of conformationally constrained macrocycles: methods and applications.* ACS chemical biology, 2016. **11**(1): p. 10-24.
[131] Wei, B.Q., L.H. Weaver, A.M. Ferrari, B.W. Matthews, and B.K. Shoichet, *Testing a flexible-receptor docking algorithm in a model binding site.* Journal of molecular biology, 2004. **337**(5): p. 1161-1182.
[132] Österberg, F., G.M. Morris, M.F. Sanner, A.J. Olson, and D.S. Goodsell, *Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock.* Proteins: Structure, Function, and Bioinformatics, 2002. **46**(1): p. 34-40.
[133] Park, S.-J., I. Kufareva, and R. Abagyan, *Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles.* Journal of computer-aided molecular design, 2010. **24**(5): p. 459-471.
[134] Evangelista Falcon, W., S.R. Ellingson, J.C. Smith, and J. Baudry, *Ensemble docking in drug discovery: how many protein configurations from molecular dynamics simulations are needed to reproduce known ligand binding?* The Journal of Physical Chemistry B, 2019. **123**(25): p. 5189-5195.
[135] Bottegoni, G., I. Kufareva, M. Totrov, and R. Abagyan, *Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking.* Journal of medicinal chemistry, 2009. **52**(2): p. 397-406.
[136] Claußen, H., C. Buning, M. Rarey, and T. Lengauer, *FlexE: efficient molecular docking considering protein structure variations.* Journal of molecular biology, 2001. **308**(2): p. 377-395.
[137] Davis, I.W. and D. Baker, *RosettaLigand docking with full ligand and receptor flexibility.* Journal of molecular biology, 2009. **385**(2): p. 381-392.

[138] Ding, F. and N.V. Dokholyan, *Incorporating backbone flexibility in MedusaDock improves ligand-binding pose prediction in the CSAR2011 docking benchmark.* Journal of chemical information and modeling, 2013. **53**(8): p. 1871-1879.
[139] Bolia, A., Z.N. Gerek, and S.B. Ozkan, *BP-Dock: a flexible docking scheme for exploring protein–ligand interactions based on unbound structures.* Journal of chemical information and modeling, 2014. **54**(3): p. 913-925.
[140] Lin, J.-H., A.L. Perryman, J.R. Schames, and J.A. McCammon, *Computational drug design accommodating receptor flexibility: the relaxed complex scheme.* Journal of the American Chemical Society, 2002. **124**(20): p. 5632-5633.
[141] Lin, J.H., A.L. Perryman, J.R. Schames, and J.A. McCammon, *The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme.* Biopolymers: Original Research on Biomolecules, 2003. **68**(1): p. 47-62.
[142] Schames, J.R., R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, *Discovery of a novel binding trench in HIV integrase.* Journal of medicinal chemistry, 2004. **47**(8): p. 1879-1881.
[143] Zacharias, M., *Rapid protein–ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: Binding of FK506 to FKBP.* Proteins: Structure, Function, and Bioinformatics, 2004. **54**(4): p. 759-767.
[144] Antes, I., *DynaDock: A new molecular dynamics-based algorithm for protein–peptide docking including receptor flexibility.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(5): p. 1084-1104.
[145] Rastelli, G., A.M. Ferrari, L. Costantino, and M.C. Gamberini, *Discovery of new inhibitors of aldose reductase from molecular docking and database screening.* Bioorganic & medicinal chemistry, 2002. **10**(5): p. 1437-1450.
[146] Cavalli, A., G. Bottegoni, C. Raco, M. De Vivo, and M. Recanatini, *A computational study of the binding of propidium to the peripheral anionic site of human acetylcholinesterase.* Journal of medicinal chemistry, 2004. **47**(16): p. 3991-3999.
[147] Eyermann, B., M. Meixner, H. Brötz-Oesterhelt, I. Antes, and S.A. Sieber, *Acyldepsipeptide probes facilitate specific detection of caseinolytic protease P independent of its oligomeric and activity state.* ChemBioChem, 2020. **21**(1-2): p. 235-240.
[148] Park, H., M.S. Yeom, and S. Lee, *Loop flexibility and solvent dynamics as determinants for the selective inhibition of cyclin-dependent kinase 4: comparative molecular dynamics simulation studies of CDK2 and CDK4.* Chembiochem, 2004. **5**(12): p. 1662-1672.
[149] Gioia, D., M. Bertazzo, M. Recanatini, M. Masetti, and A. Cavalli, *Dynamic docking: a paradigm shift in computational drug discovery.* Molecules, 2017. **22**(11): p. 2029.
[150] Gervasio, F.L., A. Laio, and M. Parrinello, *Flexible docking in solution using metadynamics.* Journal of the American Chemical Society, 2005. **127**(8): p. 2600-2607.
[151] Provasi, D., A. Bortolato, and M. Filizola, *Exploring molecular mechanisms of ligand recognition by opioid receptors with metadynamics.* Biochemistry, 2009. **48**(42): p. 10020-10029.
[152] Shan, Y., E.T. Kim, M.P. Eastwood, R.O. Dror, M.A. Seeliger, and D.E. Shaw, *How does a drug molecule find its target binding site?* Journal of the American Chemical Society, 2011. **133**(24): p. 9181-9183.
[153] Dror, R.O., H.F. Green, C. Valant, D.W. Borhani, J.R. Valcourt, A.C. Pan, D.H. Arlow, M. Canals, J.R. Lane, and R. Rahmani, *Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs.* Nature, 2013. **503**(7475): p. 295-299.

[154] Bisignano, P., S. Doerr, M.J. Harvey, A.D. Favia, A. Cavalli, and G. De Fabritiis, *Kinetic characterization of fragment binding in AmpC β-lactamase by high-throughput molecular simulations.* Journal of Chemical Information and Modeling, 2014. **54**(2): p. 362-366.
[155] Schütz, R., M. Meixner, I. Antes, and F. Bracher, *A modular approach to the bisbenzylisoquinoline alkaloids tetrandrine and isotetrandrine.* Organic & biomolecular chemistry, 2020. **18**(16): p. 3047-3068.
[156] Wang, S., M. Meixner, L. Yu, L. Zhuo, L. Karmann, U. Kazmaier, A.M. Vollmar, I. Antes, and S. Zahler, *Turning the Actin Nucleating Compound Miuraenamide into Nucleation Inhibitors.* ACS omega, 2021. **6**(34): p. 22165-22172.
[157] McCammon, J. and S. Harvey, *Dynamics of proteins and nucleic acids cambridge univ.* Press, New York, 1987.
[158] Brooks, C., M. Karplus, and B.M. Pettitt, *Adv. Chem. Phys.* Proteins, A Theoretical Perspective of Dynamics, Structure, and Thermodynamics, 1988. **71**.
[159] Schlick, T., *Molecular modeling and simulation: an interdisciplinary guide*. Vol. 2. 2010: Springer.
[160] Cornell, W.D., P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules.* Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
[161] Morse, P.M., *Diatomic molecules according to the wave mechanics. II. Vibrational levels.* Physical review, 1929. **34**(1): p. 57.
[162] Urey, H.C. and C.A. Bradley Jr, *The vibrations of pentatonic tetrahedral molecules.* Physical review, 1931. **38**(11): p. 1969.
[163] Mandal, D.K., *Stereochemistry and Organic Reactions: Conformation, Configuration, Stereoelectronic Effects and Asymmetric Synthesis*. 2021: Academic Press.
[164] Landau, L.D. and E.M. Lifshitz, *Quantum mechanics: non-relativistic theory*. Vol. 3. 2013: Elsevier.
[165] Verlet, L., *Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Physical review, 1967. **159**(1): p. 98.
[166] Taylor, R.D., P.J. Jewsbury, and J.W. Essex, *FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function.* Journal of computational chemistry, 2003. **24**(13): p. 1637-1656.
[167] Bringmann, G., J. Mühlbacher, M. Reichert, M. Dreyer, J. Kolz, and A. Speicher, *Stereochemistry of isoplagiochin C, a macrocyclic bisbibenzyl from liverworts.* Journal of the American Chemical Society, 2004. **126**(30): p. 9283-9290.
[168] Wang, X., D.W. House, P.A. Oroskar, A. Oroskar, A. Oroskar, C.J. Jameson, and S. Murad, *Molecular dynamics simulations of the chiral recognition mechanism for a polysaccharide chiral stationary phase in enantiomeric chromatographic separations.* Molecular Physics, 2019. **117**(23-24): p. 3569-3588.
[169] Wang, X., C.J. Jameson, and S. Murad, *Modeling enantiomeric separations as an interfacial process using amylose Tris (3, 5-dimethylphenyl carbamate)(ADMPC) polymers coated on amorphous silica.* Langmuir, 2020. **36**(5): p. 1113-1124.
[170] Kim, H., N. Srividya, I. Lange, E.W. Huchala, B. Ginovska, B.M. Lange, and S. Raugei, *Determinants of Selectivity for the Formation of Monocyclic and Bicyclic Products in Monoterpene Synthases.* ACS Catalysis, 2022. **12**: p. 7453-7469.
[171] Frickenhaus, S., S. Kannan, and M. Zacharias, *Efficient evaluation of sampling quality of molecular dynamics simulations by clustering of dihedral torsion angles and Sammon mapping.* Journal of computational chemistry, 2009. **30**(3): p. 479-492.

[172] Geng, H., F. Jiang, and Y.-D. Wu, *Accurate structure prediction and conformational analysis of cyclic peptides with residue-specific force fields.* The journal of physical chemistry letters, 2016. **7**(10): p. 1805-1810.
[173] Resource, D.D.D. *Grand Challenge 4*. 2018 [cited 2022 November 19]; Available from: https://drugdesigndata.org/about/grand-challenge-4.
[174] Basciu, A., P.I. Koukos, G. Malloci, A.M. Bonvin, and A.V. Vargiu, *Coupling enhanced sampling of the apo-receptor with template-based ligand conformers selection: performance in pose prediction in the D3R Grand Challenge 4.* Journal of Computer-Aided Molecular Design, 2020. **34**(2): p. 149-162.
[175] Sasmal, S., L. El Khoury, and D.L. Mobley, *D3R Grand Challenge 4: ligand similarity and MM-GBSA-based pose prediction and affinity ranking for BACE-1 inhibitors.* Journal of computer-aided molecular design, 2020. **34**(2): p. 163-177.
[176] Elisée, E., V. Gapsys, N. Mele, L. Chaput, E. Selwa, B.L. de Groot, and B.I. Iorga, *Performance evaluation of molecular docking and free energy calculations protocols using the D3R Grand Challenge 4 dataset.* Journal of computer-aided molecular design, 2019. **33**(12): p. 1031-1043.
[177] Kotelnikov, S., A. Alekseenko, C. Liu, M. Ignatov, D. Padhorny, E. Brini, M. Lukin, E. Coutsias, K.A. Dill, and D. Kozakov, *Sampling and refinement protocols for template-based macrocycle docking: 2018 D3R Grand Challenge 4.* Journal of computer-aided molecular design, 2020. **34**(2): p. 179-189.
[178] Varela-Rial, A., M. Majewski, A. Cuzzolin, G. Martínez-Rosell, and G. De Fabritiis, *SkeleDock: a web application for scaffold docking in PlayMolecule.* Journal of chemical information and modeling, 2020. **60**(6): p. 2673-2677.
[179] Basciu, A., G. Malloci, F. Pietrucci, A.M. Bonvin, and A.V. Vargiu, *Holo-like and druggable protein conformations from enhanced sampling of binding pocket volume and shape.* Journal of Chemical Information and Modeling, 2019. **59**(4): p. 1515-1528.
[180] Guterres, H., S.-J. Park, W. Jiang, and W. Im, *Ligand-binding-site refinement to generate reliable Holo protein structure conformations from Apo structures.* Journal of chemical information and modeling, 2020. **61**(1): p. 535-546.
[181] Cole, J.C., C.W. Murray, J.W.M. Nissink, R.D. Taylor, and R. Taylor, *Comparing protein–ligand docking programs is difficult.* Proteins: Structure, Function, and Bioinformatics, 2005. **60**(3): p. 325-332.
[182] Hawkins, P.C. and A. Nicholls, *Conformer generation with OMEGA: learning from the data set and the analysis of failures.* Journal of chemical information and modeling, 2012. **52**(11): p. 2919-2936.
[183] Zhang, Y. and M.F. Sanner, *Docking flexible cyclic peptides with AutoDock CrankPep.* Journal of chemical theory and computation, 2019. **15**(10): p. 5161-5168.
[184] Makeneni, S., D.F. Thieker, and R.J. Woods, *Applying pose clustering and MD simulations to eliminate false positives in molecular docking.* Journal of chemical information and modeling, 2018. **58**(3): p. 605-614.
[185] Shen, C., J. Ding, Z. Wang, D. Cao, X. Ding, and T. Hou, *From machine learning to deep learning: Advances in scoring functions for protein–ligand docking.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2020. **10**(1): p. e1429.
[186] Greenidge, P.A., C. Kramer, J.-C. Mozziconacci, and W. Sherman, *Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA.* Journal of chemical information and modeling, 2014. **54**(10): p. 2697-2717.

[187] El Khoury, L., D. Santos-Martins, S. Sasmal, J. Eberhardt, G. Bianco, F.A. Ambrosio, L. Solis-Vasquez, A. Koch, S. Forli, and D.L. Mobley, *Comparison of affinity ranking using AutoDock-GPU and MM-GBSA scores for BACE-1 inhibitors in the D3R Grand Challenge 4.* Journal of computer-aided molecular design, 2019. **33**(12): p. 1011-1020.
[188] Wang, S., A.H. Crevenna, I. Ugur, A. Marion, I. Antes, U. Kazmaier, M. Hoyer, D.C. Lamb, F. Gegenfurtner, and Z. Kliesmete, *Actin stabilizing compounds show specific biological effects due to their binding mode.* Scientific reports, 2019. **9**(1): p. 1-17.

List of Figures

List of Tables

Abbreviations

17-AAG	17-allylamino-17-demethoxygeldanamycin
17-DMAG	17-dimethylaminoethylamino-17-demethoxygeldanamycin
2D	two-dimension(al)
3D	three-dimension(al)
AD	AutoDock
ATP	adenosine triphosphate
BP-Dock	backbone perturbation-dock
BRIKARD	builder for recursive inverse kinematic assembly and ring design
CAESAR	conformer algorithm based on energy screening and recursive buildup
CCDC	Cambridge Crystallographic Data Centre
c-RDC	cyclopropa**r**adicicol
CsA	cyclosporin A
CSD	Cambridge Structural Database
DG	distance geometry
DoF(s)	degree(s) of freedom
EDES	ensemble docking with enhanced sampling
EMA	European Medicines Agency
FDA	U.S. Food and Drug Administration
FF(s)	force field(s)
FKBP(12)	FK506-binding protein (12)
ForceGen	force field based conformational generation
GA	genetic algorithm
GDM	geldanamycin
Glide	grid-based ligand docking with energetics
GOLD	genetic optimization for ligand docking
HBA	hydrogen bond acceptor
HBD	hydrogen bond donor
HCV	hepatitis C virus
Hsp90	heat shock protein 90
ICM	internal coordinates mechanics
LAM	lymphangioleiomyomatosis
LLMOD	large scale LMOD

LMMD	low mode MD method
LMOD	low mode search
MC(s)	macrocycle(s), macrocyclic
MD	molecular dynamics
MM	molecular mechanics
MMBS	MacroModel baseline search
MMGBSA	MM/generalized Born Surface Area
MMPBSA	MM/Poisson-Boltzmann Surface Area
MOE	molecular operating environment
MTD	meta-dynamics
mTOR(C)	mechanistic target of rapamycin (complex)
NFAT	nuclear factor of activated T-cells
NP	natural product
OPMD	optimized potential MD
PCA	principle component analysis
PDB	Protein Data Bank
PEComa	perivascular epithelioid cell tumor (sarcoma)
PPI	protein-protein interaction
Prime-MCS	Prime macrocycle conformational sampling
PSA	polar surface area
QM	quantum mechanics
RAP	rapamycin
RDC	radicicol
Rgyr	radius of gyration

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Dr. Iris Antes for giving me the opportunity to start my PhD project in the group of Theoretical Chemical Biology at the chair of Protein Modeling. I am thankful for her scientific guidance and feedback that helped me overcome the struggles I faced during projects. I always felt supported and appreciated by her. Unfortunately, she passed away in 2021. Her effort and kindness will be missed, and I will keep her in honorable memory.

I'd like to further thank Prof. Dr. Dmitrij Frishman, Prof. Dr. Martin Zacharias and my mentor Prof. Dr. Stephan A. Sieber for taking over supervision and providing helpful advice when I needed it the most. Without their undoubted support, I would not have been able to successfully finish my PhD and I am truly grateful for that.

I thank Prof. Dr. Stephan A. Sieber, Prof. Dr. Franz Bracher and Prof. Dr. Stefan Zahler as well as their groups and colleagues for fruitful collaborations and interesting discussions. Working together broadened my horizon and helped me develop valuable skills over the borders of disciplines.

Next, I thank my former group members Dr. Helmut Lutz, Dr. Chen Zheng, Dr. Okke Melse, Manuel Glaser, Martin Zachmann, Markus Schneider, Lukas Wietbrock, Simone Göppert and Martijn Bemelmans for helpful scientific discussions that shaped my way of thinking and working. And for more than needed quality time, off-work activities, and dinners. I am very happy I met you. You made this journey a pleasant experience. In this regard, I would also like to thank my students Sebastian Metzler and Jonathan Scheerer for their help on my projects. It was a pleasure teaching you. Further, I thank Drazen Jalsovec for fast technical support on all ends and Martina Rüttger for administrative help and an open ear in countless coffee breaks, your friendship is a precious gift.

Finally, I'd like to thank my family and friends for their emotional support during this turbulent time, especially my partner Tae Kyeong, who was there for me continuously and believed in me through good and bad times.

Appendices

Appendix 1: Supplemental results of macrocycle conformational sampling

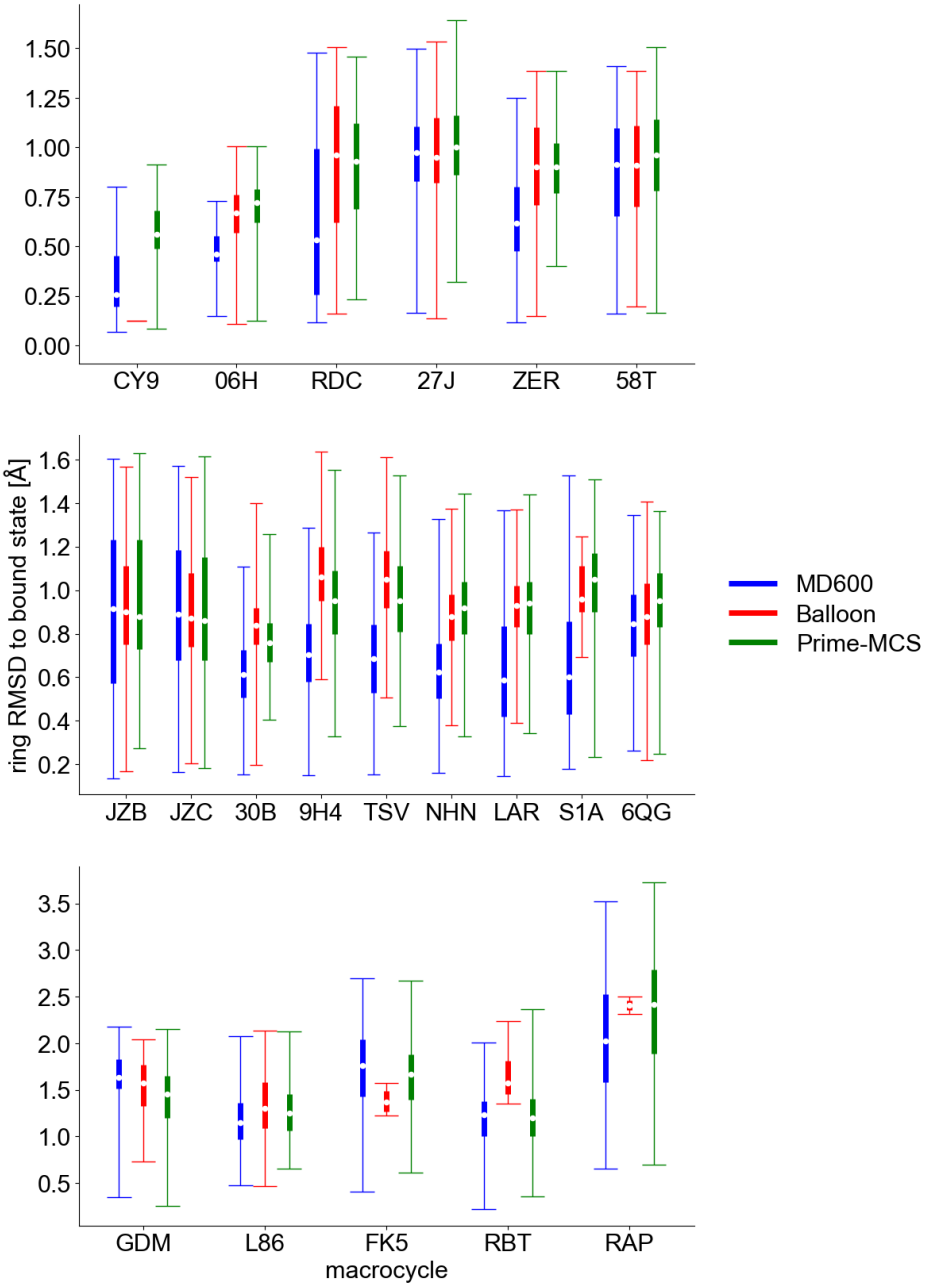

Figure A1. Comparison between 3 different conformer generation methods for the conformational sampling of MCs of the evaluation set by Meixner *et al.*[104] (chapter **3.4**): extensive MD simulations (blue: 3x1,000 ns, 600 K, explicit solvent), distance geometry algorithm of Balloon software (red) and Schrödinger's Prime Macrocycle Conformational Sampling tool (green). The MC ring atoms of generated structures were compared to the bound form in the holo state (ring RMSD, y-axis), boxplots mark the first and third quartile values of the conformer ensemble, whiskers extend to the overall minimum and maximum values and the median is indicated by a white dot.