# Modelling the binding mode of macrocycles: docking and conformational sampling

Sarah J. Martin[a, 1], I-Jen Chen[b], A.W. Edith Chan[a], Nicolas Foloppe[b*]

[a]*Wolfson Institute for Biomedical Research, University College London, Gower Street, London, WC1E 6BT, UK*

[b]*Vernalis (R&D) Ltd., Granta Park, Abington, Cambridge CB21 6GB, UK*

[1]Present address: Charles River Laboratories, Chesterford Research Park, Saffron Walden, CB10 1XL, UK

[*]To whom correspondence should be addressed:
N. Foloppe email: n.foloppe@vernalis.com
Phone (direct): + (44) (0) 1223 895 338
Fax: + (44) (0) 1223 895 556

**KEYWORDS**: Drug Discovery; Computational Chemistry; Conformers; Docking; Macrocycle; Molecular Recognition

**ABBREVIATIONS:** CCDC: Cambridge Crystallographic Data Centre; CSD: Cambridge Structural Database; GA: Genetic Algorithm; GUI: Graphical User Interface; LowModeMD: search method combining low-mode moves and molecular dynamics; MT/LMOD: mixed torsional/Low-mode; MOE: Molecular Operating Environment; MW: molecular weight; NMR: Nuclear Magnetic Resonance; NRot: Oprea Number of Rotatable bonds; OPLS: Optimized Potential for Liquid Simulations force-field; PDB: Protein Data Bank; RMSD: Root Mean Square Deviation; SAR: Structure-Activity Relationship; SBDD: Structure-Based Drug Design; SD: Standard Deviation; TPSA: Mean Topological Polar Surface Area; 2D: two-dimensional; 3D: three-dimensional

**ABSTRACT:** Drug discovery is increasingly tackling challenging protein binding sites regarding molecular recognition and druggability, including shallow and solvent-exposed protein-protein interaction interfaces. Macrocycles are emerging as promising chemotypes to modulate such sites. Despite their chemical complexity, macrocycles comprise important drugs and offer advantages compared to non-cyclic analogs, hence the recent impetus in the medicinal chemistry of macrocycles. Elaboration of macrocycles, or constituent fragments, can strongly benefit from knowledge of their binding mode to a target. When such information from X-ray crystallography is elusive, computational docking can provide working models. However, few studies have explored docking protocols for macrocycles, since conventional docking methods struggle with the conformational complexity of macrocycles, and also potentially with the shallower topology of their binding sites. Indeed, macrocycle binding mode prediction with the mainstream docking software GOLD has hardly been explored. Here, we present an in-depth study of macrocycle docking with GOLD and the ChemPLP scores. First, we summarize the thorough curation of a test set of 41 protein-macrocycle X-ray structures, raising the issue of lattice contacts with such systems. Rigid docking of the known bioactive conformers was successful (three top ranked poses) for 92.7% of the systems, in absence of crystallographic waters. Thus, without conformational search issues, scoring performed well. However, docking success dropped to 29.3 % with the GOLD built-in conformational search. Yet, the success rate doubled to 58.5% when GOLD was supplied with extensive conformer ensembles docked rigidly. The reasons for failure, sampling or scoring, were analyzed, exemplified with particular cases. Overall, binding mode prediction of macrocycles remains challenging, but can be much improved with tailored protocols. The analysis of the interplay between conformational sampling and docking will be relevant to the prospective modelling of macrocycles in general.

## 1. INTRODUCTION

Macrocycles are natural[1, 2] or synthetically made[3, 4] compounds featuring at least one "large" cyclic moiety. Here, the term macrocycle will refer to the entire compound, unless specified otherwise. Macrocycles encompass diverse chemical classes, including a broad array of natural products, cyclic peptides, peptidomimetics and cyclodextrins[1, 5]. The size and structural complexity of macrocycles present numerous synthetic challenges[1, 3] . In addition, their physico-chemical properties typically lie outside of the conventional drug-like space[6, 7]. Despite those complications, macrocyclic compounds have a long history of contributing to pharmacology and medicine[1, 5]; some are important orally administered drugs, for example antibiotic erythromycin and immunosuppressant cyclosporine A[7]. Also, recent developments have spurred renewed interest in macrocycles for drug discovery[1, 3, 4, 8, 9].

The current impetus is motivated by the realisation that many potential drug targets are difficult to modulate with small conventional drug-like chemotypes, based on structural and energetic considerations[10, 11]. Thus, interfaces mediating protein-protein interactions tend to present flat/shallow surfaces (instead of deep binding clefts) for binding to compounds[11, 12]; binding with sufficient affinity and selectivity to such surfaces seems to call for larger chemotypes, a good number of which are macrocyclic[6, 13, 14]. An analysis of macrocycle-binding sites suggested that some would be poorly or not druggable with conventional drug-like compounds[6]. Pharmaceutical targets of macrocycles include numerous proteins[3, 15], DNA[16] and RNA[10, 17]. Importantly, the cyclization of macrocycles can pre-organize their conformation for more potent and selective binding[1, 5, 18, 19]. Also, cyclic peptides resist proteolytic degradation better than their linear counterparts, since cyclization imparts a molecular shape that does not fit in the endopeptidase binding sites[5]. However, most macrocycles retain significant flexibility[20-22], the basis for what has been termed the "chameleonic properties" of macrocycles[9]. This refers to the ability to undergo conformational changes which expose polar groups in aqueous solution, but bury them while crossing lipidic membranes[9, 23]. Thus, more rational approaches to the synthesis and design of therapeutic macrocycles are being developed[3, 4, 8, 9, 24].

Computational chemistry will play a growing role towards a more efficient elaboration of macrocycles, via the generation and analysis of conformers[21, 25-31], and docking to binding sites[22, 32-35]. Due to their many conformational degrees of freedom, obtaining a thorough and relevant conformational ensemble remains far from trivial for macrocycles[21, 25, 36]. However,

modern computational resources have greatly eased this challenge, which has also been addressed by new algorithms[20, 27-29, 31] or adapted protocols[21, 26]. An early study tackling the conformer generation for macrocycles found that low-mode based methods handle macrocyclic topologies robustly, and typically produce at least one conformer within 2.0 Å of the reference bioactive X-ray structures[21]; frequently a conformer within 1.0 Å of the X-ray structure was generated[21]. Further studies in the same vein followed[26-30, 37], with different computational conformer generators and extended test sets. All these investigations confirmed that conformers close to the bound X-ray reference can usually be obtained for macrocycles; also, new conformational sampling methods tend to be computationally faster than the low-mode based approaches[27-29]. The relevant point is that all these algorithms usually produce at least one conformer representative of the bioactive state, which can be exploited when attempting to infer the macrocycle binding mode.

Being able to visualize how a compound binds to its target is of great help to improve this ligand via structure-based drug design (SBDD)[38-40]. X-ray crystallography is the predominant technique to observe such binding mode, however not all proteins or their complexes crystallize[41], and X-ray models have limitations[42]. Computational docking is another approach to gain insights about a compound binding mode. This requires a model of the binding site (e.g. from crystallography or homology modelling), and a docking software to build and score multiple putative binding models of the compound[43-45]. Docking has become an established component of SBDD[39], for virtual screening or propose binding modes for compounds of particular interest. Mainstream docking softwares include AutoDock[45], DOCK[46], Glide[47] , GOLD[43, 48, 49], and others[50-52]. Thus, the general principles underpinning computational docking are well-documented[50, 53]. The receptor site is frequently approximated by a single fixed structure[49, 53]. On the other hand, the compound translational and rotational degrees of freedom are explored, as well as its internal conformations. The docked poses are scored[48, 52, 54] and ranked for their goodness of fit to the receptor. The intention is that the best scored poses can inform about the compound true binding mode. So, docking success is frequently assessed by the ability of optimally ranked docked poses to reproduce the bioactive structure.

Thus, the two essential ingredients of a docking protocol are sampling of the compound degrees of freedom, and scoring of the resulting interactions with the receptor. The X-ray binding mode of moderately-sized drug-like compounds is usually visited during

docking[52, 55], consistent with their limited flexibility. Then, conformational sampling is performed concurrently with docking, and success tends to be limited by the scoring of protein-ligand interactions rather than sampling[55]. However, studies noted a drop in docking success as the ligand flexibility increased, across investigated programs[43, 52, 53, 56, 57]. Efficient sampling of the compound during docking remains a particular challenge with macrocycles[22, 55]. Indeed, the docking algorithms and default protocols were not developed to handle the numerous degrees of freedom in macrocycles, nor the complications arising from ring closure with complex cyclic topologies[22, 58, 59]. Indeed, a study of 10 docking programs found that none could recover the X-ray binding mode of a Gyrase B macrocyclic inhibitor of moderate size, this failure being attributed to the conformational complexity of the macrocycle[55]. Some docking algorithms can be adapted to handle ring closure with macrocycles[59], however the sampling issue remains.

Another strategy with macrocycles is to perform adequate conformational sampling prior to docking, followed by rigid docking of this conformational ensemble[22]. However, the potential benefits and trade-offs of this strategy have rarely been examined in detail, despite a growing number of case studies involving docking of macrocycles[22]. A few studies recently investigated the rigid docking of pre-generated conformers for sets of macrocycles, with[34] or without[32, 33] comparing to direct docking with the tested docking engines. These questions have not been investigated specifically with the widely used program GOLD[50, 51], which has otherwise benefited from continuous development and benchmarking[48, 49, 60] since the presentation of the program[43, 61] and its genetic algorithm (GA) for compound sampling. GOLD ability to reproduce and score favorably the X-ray binding modes of small drug-like compounds is widely recognized[49, 52, 55, 56, 60], therefore it is important to investigate how GOLD fares with binding mode prediction of macrocycles.

The present work addresses this question by testing GOLD with various sampling protocols on a thoroughly curated test set of macrocycle-protein systems; the compilation of this set is discussed. The results show that conformational sampling is a limiting factor if relying on the docking engine alone. However, docking of good-quality pre-generated conformational ensembles offers marked improvement. Thus, the existing docking tools can provide very valuable insights when combined with extensive sampling and analysis, despite undeniable difficulties specific to macrocycles and their binding sites.

## 2. METHODS

The docking performance of the GOLD software was studied with 41 carefully selected protein-bound macrocycle systems and three docking protocols. The three docking approaches investigate how increasing conformational sampling of the macrocycles influences the docking success, defined as producing a well-ranked docking pose within 2.0 Å of its X-ray counterpart.

### 2.1. Selection of test systems.

The test systems were assembled from two subsets of macrocycle-protein complex X-ray structures available from the Protein Data Bank (PDB)[62]. Both subsets were selected based on the chemical diversity of the macrocycles and of their protein targets (hence diversity of binding sites), and the quality of the X-ray structures. The chemical diversity took into account the flexibility of each compound, characterized by the Oprea number of rotatable bonds (NRot), which accounts for flexibility in cyclic moieties[63]. The selected systems were solved at a crystallographic resolution ≤ 2.1Å, and typically much better. Only systems with structure factors available from the PDB were selected, and it was checked that the X-ray binding modes for those macrocycles are consistent with the 2Fo–Fc electron density maps contoured at the 1σ level with software PrimeX[64]. The final combined test set contains 41 distinct macrocycles.

The first subset was drawn from 30 systems carefully selected from the PDB before 2013[21], already used as benchmark in conformational sampling studies of macrocycles[21, 26, 29]. PDB entry 3OMJ was excluded since it involves a macrocycle bound to DNA, while the present work concentrates on docking to proteins. Thus, 29 systems were kept from the initial 30 (Table S1). For those 29 macrocycles, most of the X-ray models are an excellent fit to the observed 2Fo–Fc electron density. However, for systems 3BXS, 2HFK, 2CD8 and 2WI9, the fit of the macrocycle to the density is better in chain A than in chain B, so the macrocycle of chain A and associated site were used as the X-ray reference. With 2HFK chain A and 2C7X, the density is only of moderate quality, albeit supporting the macrocycle position. With 2WI9 chain A, the electron density is weaker, but provides the overall macrocycle orientation. With 1BKF, a small solvent-exposed part of the macrocycle has broken density, but there is clear density for most of the compound. Most of the cyclodextrin of 2XFY fits in the electron density, however for one sugar unit there is little

density. With 3M5L, the fluorinated ring of the macrocycle ligand has double-occupancy in the X-ray structure; the ring orientation with the fluorine pointing towards Asp1168 was kept as the reference ligand. When systems consisted of multiple subunits, monomer A was used, especially since this convenient selection happened to be supported by inspection of the crystallographic electron densities for some systems. When the first subset was assembled, preference was given to compounds where flexibility resides primarily in the macrocyclic substructure, instead of the pendant side-chains. This criterion was also applied when selecting the second subset.

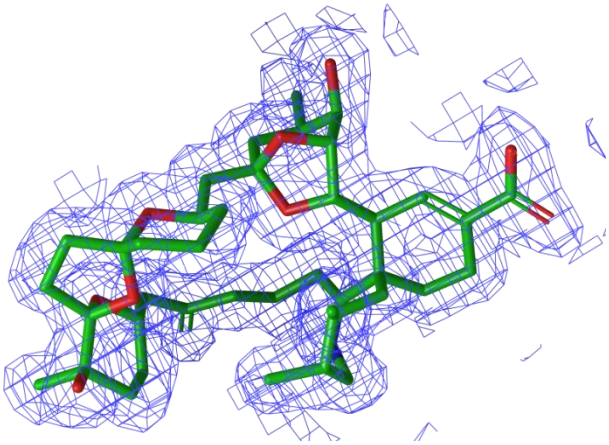
A second subset of 12 macrocycle-protein complexes was added, from PDB entries released after 2013 (Table 1). The selection criteria for these 12 systems were even more stringent than for the first subset. Thus, macrocycles covalently linked (e.g. 5E0J, 1.20Å resolution) or with dual occupancy (e.g. 4DPI, 1.90Å resolution) were discarded since they present complications for comparison to the docking results. Figure 1 illustrates the excellent fit to the electron density for representative examples of the second subset. In addition to crystallographic quality and diversity of compounds and protein targets, systems in the second subset had to i) be free of crystallization agents in close proximity with the macrocycle, and ii) be free of lattice contact between the macrocycle in its biological unit and symmetry-related non-biological interfaces. These two criteria are relevant to docking studies in general, but are particularly pertinent with macrocyclic ligands, when they bind to shallow and exposed binding surfaces[6, 13]. Such open sites increase the risk that some macrocycle binding modes reported by crystallography may be influenced by lattice contacts. Figure 2 shows systems which seemed promising for docking studies (good X-ray resolution and electron density), but were excluded since the macrocycle is involved in crystal packing. In addition, the second subset prioritized compounds relevant to drug discovery, intentionally excluding excessively large macrocycles. That recognizes that prospective docking of macrocycles which are too large and complex is currently unlikely to be productive. Indeed, medicinal chemistry efforts are likely to prioritize macrocycles of moderate size. Thus, NRot for compounds in the second subset ranges from 8 to 24, with a median of 13.

**Table 1. List of macrocycle test systems included in the second subset.**

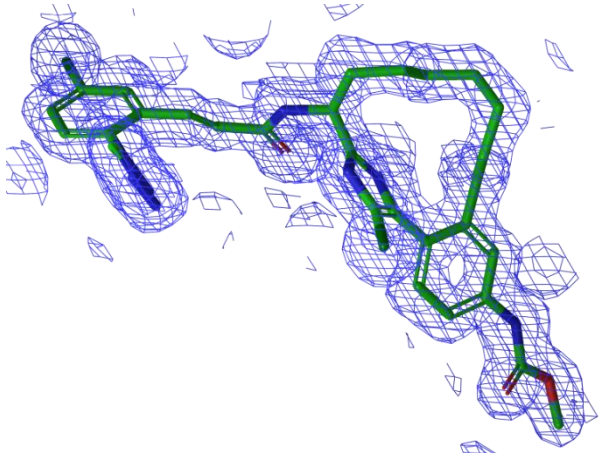
| PDB entry[a] | PDB lig. code[b] | MW[c] | NRot[d] | Res. (Å)[e] | Protein[f] |
|---|---|---|---|---|---|
| 4CLJ | 5P8 | 407.4 | 8 | 1.66 | Human Anaplastic Lymphoma Kinase |
| 4DPF | 0LG | 649.8 | 18 | 1.80 | BACE-1 |
| 4KE1 | 1R6 | 556.7 | 14 | 1.91 | BACE-1 |
| 4P3P | 2CR | 488.6 | 19 | 2.10 | Threonyl-tRNA Synthetase |
| 4X7Z | ZM3 | 682.9 | 24 | 1.44 | Sugar O-Methyltransferase |
| 4XHE | 40P | 711.9 | 10 | 1.90 | Acetylcholine-binding protein |
| 4YLA | ILV | 301.4 | 8 | 1.40 | Indole prenyltransferase |
| 5L30 | 70A | 616.7 | 12 | 1.73 | Factor VIIa |
| 5L7H | 6QG | 398.5 | 10 | 1.84 | Mineralocorticoid Receptor |
| 5TJX | GBT | 527.6 | 16 | 1.41 | Plasma kallikrein |
| 5TKS | 7DL | 609.5 | 15 | 1.55 | FACTOR XIA |
| 5TO8 | 7FM | 561.6 | 11 | 1.98 | Protein-tyrosine kinase 2-beta |

[a]Protein Data Bank entry for the macrocycle-protein complex. [b]Macrocycle ligand code in the PDB. [c]Molecular Weight. [d]Number of Oprea rotatable bonds. [e]Crystallographic resolution. [f]Protein to which the macrocycle is bound.

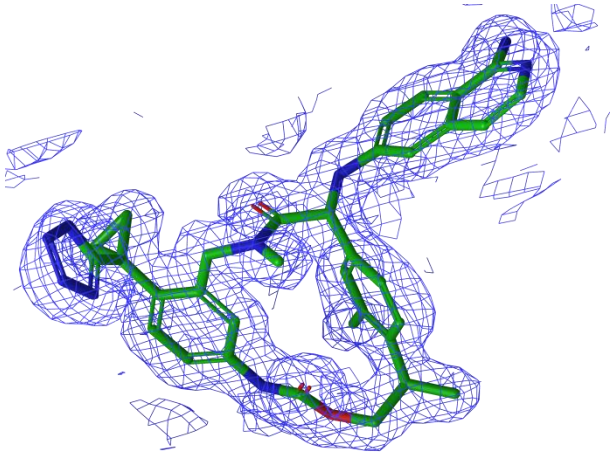
Figure 1



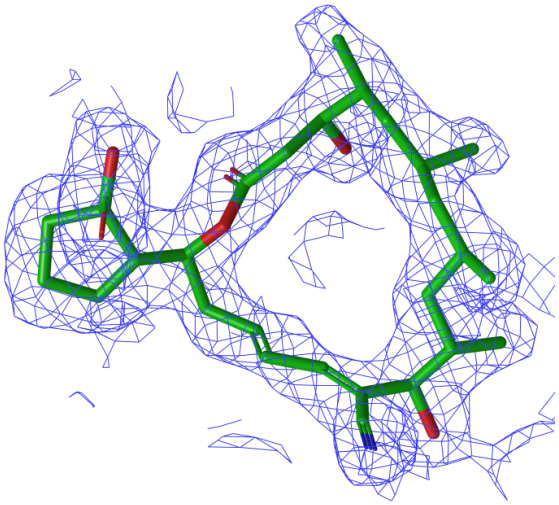
4XHE, NRot 10



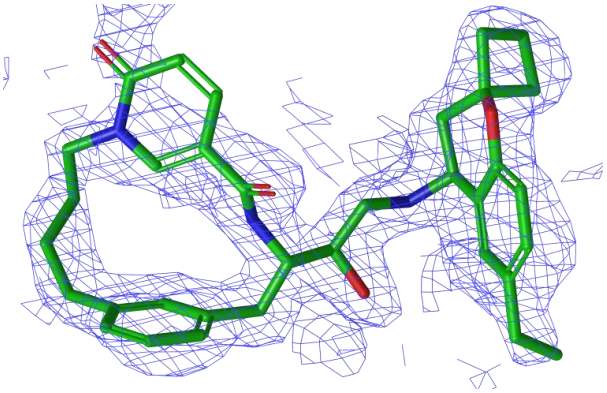
5TKS, NRot 15



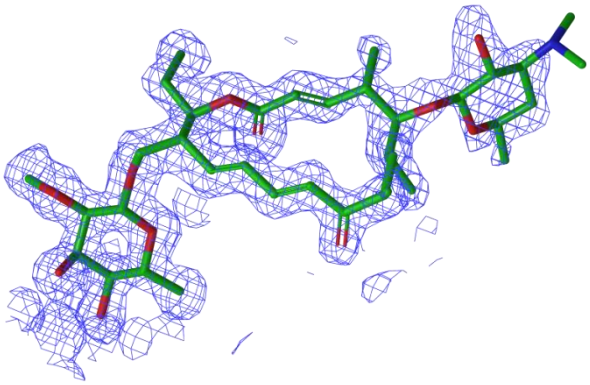
4L30, NRot 12



4P3P, NRot 19



4KE1, NRot 14



4X7Z, NRot 24

**Figure 1**. Examples of X-ray structures of macrocycles bound to proteins in the test set. The 2Fo-Fc electron density around the compound is shown as a blue mesh, contoured at the 1σ level, rendered with the software PrimeX. The PDB entry code is given under the compound, as well as its Oprea number of rotatable bonds (NRot). The binding mode of every compound is well-defined by the observed electron density from X-ray analysis.

Figure 2

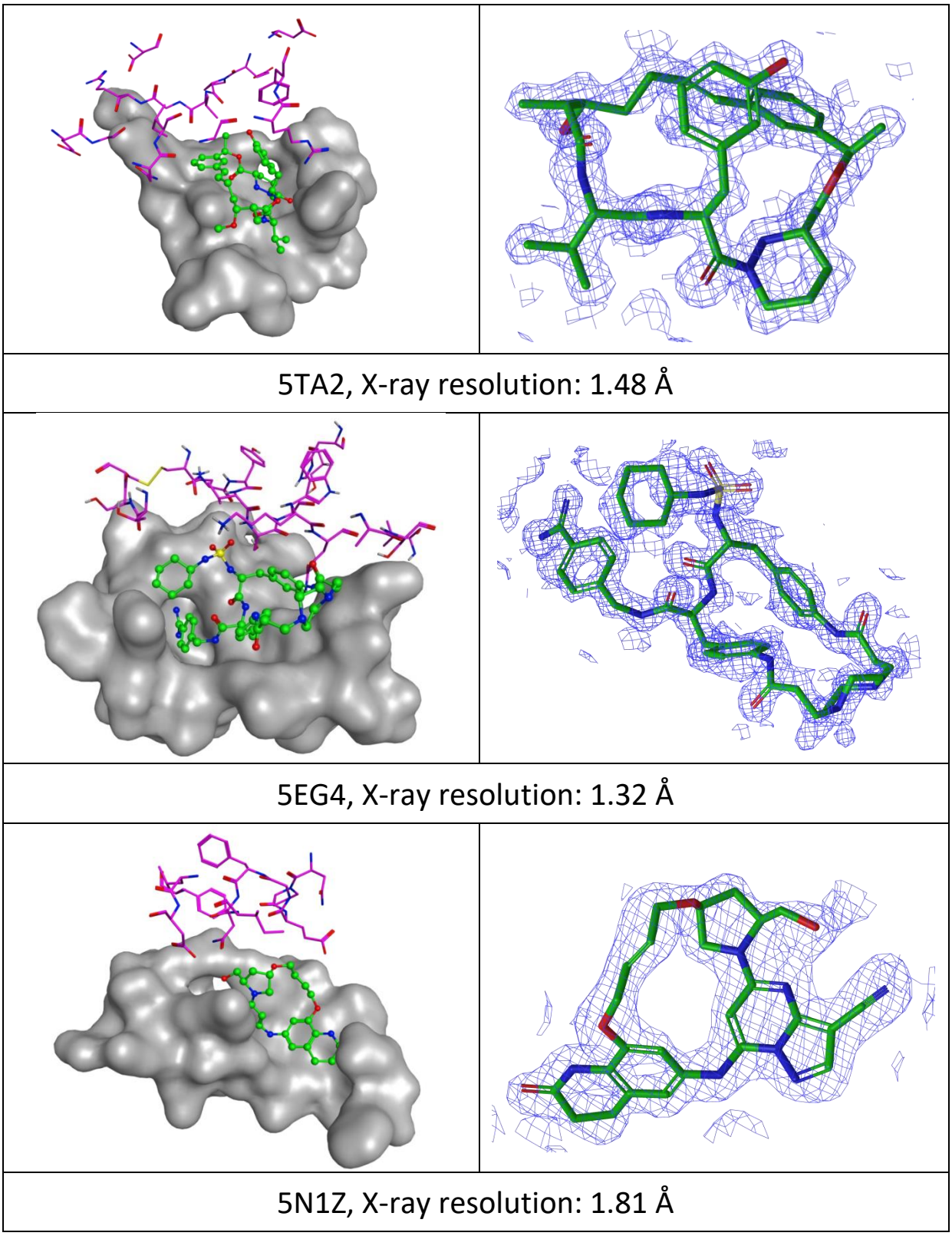



**Figure 2**. Examples of X-ray structures of macrocycles bound to a protein, and involved in lattice contacts with symmetry-related molecules. The left panels show the macrocycle (green carbons) and a grey surface around its protein binding site (6Å shell around the macrocycle). The magenta protein residues are from a symmetry-related molecule making

contacts with the macrocycle (distance ≤ 4Å). The right-hand side panels show the 2Fo-Fc electron density contoured at the 1$\sigma$ level for the same macrocycle. The PDB entry code and crystallographic resolution are given under each pair of panels. Systems where the macrocycle is involved in lattice contacts were excluded from the second subset of test systems. See main text for discussion of these issues.

**2.2 Preparation of proteins for docking.**

Each protein X-ray structure was prepared for docking using GOLD (version 5.5)[43, 49], and its set-up wizard, with the default settings. Deviating from default options would require additional information unlikely to be available when exploring a novel system, for example regarding non-standard protonation states or water-mediated binding interactions. Thus, it was deemed important to test the docking protocols primarily with the information supplied by the structure of the protein in isolation, without being supplemented by biasing details deduced from the compound-protein complex. This is consistent with the default settings of the GOLD set-up wizard.

Since GOLD uses an "all-atom" model for docking, hydrogen atoms were added by GOLD to satisfy unfilled atom valences. Protonation and tautomeric states were kept as assigned by the GOLD set-up wizard. Thus, standard protonation states at pH=7 were assigned for the protein residues; that is basic and acidic side-chains were protonated and deprotonated, respectively. For example, both catalytic Asp25 of the HIV protease (PBDs 3BXS, 1B6M), and Asp32 and Asp228 in the aspartate dyad of BACE1 (PDBs 4DPF, 4KE1, were deprotonated. When hydrogens were already present in the PDB entry (e.g. entry 5TKS), they were stripped prior to the hydrogen assignment performed by GOLD.

The docking site for each system was defined using the default GOLD cavity detection mode and the protein region within 6.0 Å of the macrocycle X-ray coordinates. This cavity detection algorithm defines the binding site by including only protein atoms within the concave binding region. Although macrocycles might sometimes bind to protein surface patches which may not be classically concave, this was retained for consistency with default settings. There was no indication that it adversely affected the results.

The macrocycle ligand was removed from the docking site prior to docking. The extracted macrocycle X-ray coordinates were used as reference for comparison to the docked poses. All crystallographic water molecules were stripped from all structures, despite awareness that some ligand-protein contacts are water-mediated. Keeping selected water molecules would involve arbitrary choices, and could artificially improve the docking outcomes. That would not be representative of prospective docking work. Also that faciliates comparison with previous benchmarking work on small molecules with GOLD[49]. Crystallization agents were also removed prior to docking (e.g. glycerol, sulfate, DMSO,

acetates). Co-factors (including hemes and nucleotides) and metals in the vicinity of the macrocycle were kept in the docking region since they were considered part of the receptor.

Several systems had metal ions in the binding region: Fe (1Q5D, 1Z8O, 2C7X, 2CD8, 2WI9, 2XBK), Mg (1S9D, 4X7Z) and Zn (1LD8, 4P3P). GOLD automatically determines the coordination geometries of metal centers using predefined geometry templates. Then, the unoccupied coordination points are available for potential coordination interactions with the docked ligand in a "pseudo-hydrogen bonding" fashion[65]. For example, GOLD assigned an octahedral coordination geometry to the iron atoms in 1Q5D, 2C7X, 2CD8, 2WI9, 2XBK, and identified 4 points of interaction between the iron and heme, sometimes with an additional interaction to a cysteine residue. Thus, GOLD added up to two fitting points between the iron and the docked compound, allowing interactions with electronegative/acceptor atoms of the docked ligands. The iron atom in 1Z8O already has a full coordination sphere including a bound oxygen molecule, therefore no fitting points between the iron and docked ligand were added for 1Z8O.

**2.3 Preparation of the macrocycles for docking.**

For consistency, the preparation of the macrocycles was similar to that used in a previous study with the subset of 29 macrocycles[21]. All compounds were initially prepared with the software MOE[66]. First, hydrogen atoms were removed when they were present in the compound X-ray structure (e.g. PDBs 5TKS and 5L30). Then, bond orders were assigned and checked based on information from publications and the PDB. Assignment of the bond orders implicitly defines the tautomers. For the nitrile and vicinal double bond in 4P3P, the correct bond orders were inconsistent with the X-ray bond lengths and were regularized, to allow the rigid docking of this X-ray conformer to proceed (see below).

The X-ray conformers were then exported from MOE, with the stereochemistry of chiral centers and double bonds flagged and maintained. Then, hydrogens were added with the program Maestro from Schrödinger[67], and protonation states handled. Common titratable groups were assigned standard protonation states at pH = 7, e.g. carboxylic acids were deprotonated and aliphatic amines were protonated. The imine in the macrocycle from PDB entry 4XHE was treated as protonated, consistent with several compound-based p$K_a$ computational estimates, performed with programs MarvinSketch (Chemaxon), Epik (Schrödinger)[68] and Jaguar (Schrödinger)[69, 70], which gave calculated p$K_a$s of 9.3, 9.1 and 8.3,

respectively. We stress that these are compound-based p$K_a$ determinations, independent of the X-ray structure 4XHE. Each macrocycle was represented by the same protonation and tautomer state across all docking experiments.

Preparation of the 3D conformers for input into GOLD was adapted to the conformational sampling approach tested during docking. For rigid docking of the X-ray conformers (protocol I), these conformers were extracted from the X-ray structure without 3D to 2D conversion. However, the position of those conformers was randomized in translation and rotation within a 100Å sphere with a Python script provided by Schrödinger; this randomization maintained the internal conformation of the X-ray bioactive structures. This was to allow rigid docking of the X-ray conformers without positional bias arising from their starting experimental location/orientation.

For docking involving conformational sampling (GOLD-standalone or assisted with MT/LMOD), memory of the X-ray bioactive conformation needed to be erased. Thus, 3D to 2D conversion was performed when exporting the macrocycle topologies from MOE, and fresh 3D conformers were generated. With complex molecules, MOE occasionally outputs the incorrect E/Z stereochemistry to 2D structures for double bonds, so those were all checked and corrected when required. The resulting 2D conformers were converted to 3D with the program LigPrep from Schrödinger[67], after addition of the hydrogens. All conformer generation performed outside of GOLD used the OPLS2005 force-field[71, 72], consistent with previous work[21]. Only one 3D conformer per compound was generated by LigPrep, maintaining the input tautomer/protonation state and stereochemistry; this was checked by examining each 3D conformer generated by LigPrep. The cis/trans conformations of amide and ester groups generated by LigPrep were consistent with the X-ray structures. Otherwise, the 3D conformers generated by LigPrep differed from their X-ray counterpart, i.e. they were given different conformations and displaced away from the binding site. Therefore, the LigPrep-generated 3D conformers were used as input conformers for flexible docking with GOLD-standalone (protocol II).

For MT/LMOD-assisted GOLD docking (protocol III), the macrocycle conformers were generated with OPLS2005 and the Mixed torsional/low-mode (MT/LMOD) method implemented in MacroModel[73]. This was performed with the same parameters already reported to enhance the MT/LMOD conformational sampling[21], including a generalized Born solvation model, an energy window of 15 kcal/mol, a maximum of 10000 search steps

(combined with 400 torsional steps per rotatable bond) and an RMSD threshold of 0.25 Å for removal of duplicate conformers. Before a conformer was tested for duplication, it was energy minimized in MacroModel with up to 3000 steps of Polak-Ribiere conjugate gradient, to an energy gradient of 0.05 $kJmol^{-1}/Å$ or less. These conformers were then input into GOLD to be docked rigidly.

**2.4 Overview of GOLD docking.**

The present study investigated three docking approaches with GOLD version 5.5 (from the Cambridge Crystallographic Data Centre, CCDC[74]), with increased conformational sampling of the macrocycles. Each compound was only docked to its associated X-ray structure binding site, i.e. no docking was attempted to non-native sites.

First, we review briefly general aspects of GOLD docking, and its default settings which were used unless stated otherwise. By default, each ligand is flexibly docked with a maximum of 10 runs, using the GOLD genetic algorithm (GA) for orientational and intramolecular conformational searches[43, 61]; the initial orientation is randomly varied each time. Default automatic GA Search and Efficiency options were used, which determine the optimal number of genetic operations per ligand for the crossover, migration and mutation parameters. GOLD allows limited flexibility of the protein side-chains, for the terminal single bonds linked to a hydrogen-bond donor or acceptor; for example, the orientation of hydroxyl groups of serine or threonine side-chains is varied [61]. Other parts of the protein were kept fixed. GOLD considers the ligand flexibility and binding site volume when specifying the number of genetic operations, to balance effectiveness with running time. By default, GOLD early termination rules finish docking (before completing 10 GA runs) as soon as the top three scoring solutions are within 1.5 Å RMSD of each other. This saves computational costs, and assumes that further runs are likely to produce similar outcomes. The docked poses were produced and ranked using the GOLD default scoring function ChemPLP[54], which outperformed other GOLD scoring functions for speed and accuracy in benchmarking studies[49, 54]. The ChemPLP overall fitness scores are positive numbers, the greater the better. The X-ray structure of the macrocycle ligand was the reference for RMSD calculations against docked poses (without fitting of docked onto X-ray), using non-hydrogen atoms only. All dockings were completed on HP Pavilion Desktop PC Dell OptiPlex GX620 workstation (Intel Pentium D processor, 8 GB memory, Microsoft Windows 10

operating system). Below, we detail how increasingly thorough macrocycle conformational sampling was investigated with three docking protocols.

### 2.5 Docking protocol I: Rigid docking of the macrocycle X-ray conformers.

The first and simplest approach was rigid docking of the compound X-ray conformer. This tests the ability to retrieve the X-ray binding mode by docking only the correct ligand conformer. It assesses if the scoring function can recognize the experimental binding mode, given the correct conformational information, without being confounded by compound conformational sampling. This is not intended to be representative of actual docking applications, but it is meaningful to investigate the relative roles of scoring and internal conformational sampling with respect to docking success.

For protocol I, the X-ray coordinates of the macrocycles were first rigidly randomized in rotation and translation before input into GOLD (section 2.3). Those displaced X-ray conformers were then rigidly docked to their binding site, by selecting the “fix all ligand rotatable bonds” in the GOLD set-up GUI. The default number of 10 docking runs per ligand was selected, and the “early termination rules” enabled. Application of those rules did not adversely affect the success rate of Protocol I.

### 2.6 Docking protocol II: Flexible docking using GOLD standalone.

The second approach was flexible docking with GOLD-standalone, which assesses the familiar docking regime in GOLD when applied to macrocycles. In addition to the rotational and translational degrees of freedom explored in protocol I, protocol II tests the impact of internal conformational sampling of the compound when performed exclusively by the GOLD GA (“GOLD-standalone”). Thus, the success of this approach requires effective exploration of conformers, in addition to scoring of protein-ligand interactions.

Before docking, the starting position and internal conformation of the macrocycles were randomized with the 2D to 3D conversion with LigPrep (section 2.3); however, the stereochemistry at the chiral centers and double bonds were kept as in the X-ray structure, to maintain the identity of the compounds. The bond lengths and valence angles are not varied by GOLD. With protocol II, the sampling was increased above default settings to give GOLD-standalone a better chance of success. So, the number of docking runs per ligand was doubled from 10 (default) to 20, and early termination rules disabled. This affords more

internal and orientational sampling of the compound; more docking poses can also be examined after docking.

### 2.7 Docking protocol III: Rigid docking of pre-generated MT/LMOD conformers.

The third approach was MT/LMOD-assisted docking with GOLD**.** This approach recognizes that the increased size and flexibility of macrocycles (compared to smaller drug-like compounds), coupled with ring closure constraints, makes conformational exploration challenging when performed during the docking[55, 56]. Thus, protocol III exploits an enhanced conformational search known to perform reasonably well with macrocycles, carried out outside of GOLD with MacroModel and MT/LMOD (section 2.3). This sampling method is well adapted to macrocyclic topologies and provides a broad exploration of their conformers. Those conformers were then rigidly docked with GOLD, referred to as MT/LMOD-assisted GOLD docking. The number of docking runs per MT/LMOD conformer was set to 10 (default), and "early termination" enabled. Due to the sheer number of poses generated by this protocol, only the 3 best scored solutions per input conformer were written to disk, to allow manageable data handling and analysis.

### 2.8 Evaluation of docking performance.

Each docking pose was scored using the ChemPLP scoring function[54] implemented in GOLD[49]. For each docking experiment (a given compound and docking protocol), the three poses with the best (most positive) ChemPLP fitness scores are referred to as the "top scoring poses". The performance of each docking protocol was assessed by its ability at reproducing the ligand X-ray structure in the three top scoring poses, across all compounds. Performance when considering solely the top scoring pose is also reported. Comparison between a docking pose and its X-ray counterpart used their root-mean-square deviation (RMSD) between non-hydrogen atoms; this was done with GOLD (without fitting between the compared atom sets). The RMSD calculated by GOLD takes the ligand symmetry into account. A docking pose was deemed to reproduce the X-ray binding mode if their pairwise RMSD was within 2.0 Å, a criterion widely used in docking studies[34, 46, 49, 53-55, 59]. Thus, a docking experiment was considered successful when at least one of the top three scoring poses had RMSD ≤ 2.0 Å to the compound X-ray structure, and a failure otherwise (Table 2). RMSDs are a widely used metric in docking studies, but their interpretation is size

dependent[75]; a RMSD of 2.0 Å represents a better performance for larger than smaller compounds[75]. So, RMSDs between 2.0 and 3.0 Å have been annotated as "Fail-Marginal" (Table 2).

The assessment of performance also considered if the X-ray conformer was visited during a docking experiment. This was also investigated with pairwise RMSDs calculated between non-hydrogen atoms of docked and X-ray conformers. However, those RMSDs were calculated after rigid fit of the compared conformers (with script rmsd.py from Schrödinger), since they aim to capture only differences in internal coordinates. Thus, two types of RMSDs are mentioned in Results and Discussion, with and without rigid fit; notations were not adapted to distinguish between them, as it is hoped that it is unambiguous from the context.

To assess if docking success correlated with features/descriptors of binding sites, those sites were analyzed with the program SiteMap[76] from Schrödinger. The binding site of interest was defined by picking the corresponding compound in the X-ray structure, including a (default) surrounding 6.0 Å buffer region; this is consistent with the definition of the docking region used with GOLD. SiteMap returns a number of descriptors for the size, polarity, enclosure and solvent exposure of the site. SiteMap was primarily calibrated to detect putative binding sites, but it also calculates a "duggability" score, Dscore; the calibration of druggability is notoriously difficult, especially since what constitutes an undruggable site is ill-defined. The author of Dscore "somewhat arbitrarily"[76] assigned sites with Dscore values smaller than 0.83 as "undruggable", those having Dscore values between 0.83 and 0.98 as "difficult", and those having larger Dscore values as "druggable". So, these Dscore ranges should certainly not be interpreted as hard cutoffs. When mentioned in section 3, the Dscore values are only intended as relative indicators, to be considered in the context of other factors.

## 3. RESULTS AND DISCUSSION.

In the following, four-letter PDB entry codes are used to identify the systems. Section 3.1 presents the curated test set of 41 systems and the trade-offs one faces when trying to balance larger sample size with high-quality and representative diversity. Then, the docking results are reported by order of increasingly sophisticated protocol, that is i) Rigid docking of X-ray conformers (protocol I, section 3.2), ii) GOLD-standalone flexible docking

(protocol II, section 3.3), and iii) MT/LMOD-assisted GOLD docking (protocol III, section 3.4). Table 2 lists the docking outcomes for every system per protocol, and the number of conformers generated by MT/LMOD for each compound. Table 3 provides an overall summary of those results. According to the ChemPLP scores, the best models correspond to the most positive scores, literally the top-scored docking poses. A straightforward approach to declare success with a docking experiment is when the best-scored pose "reproduces" the X-ray reference within a chosen RMSD threshold. Here, we have deliberately adopted a more lenient criterion by considering that any of the three top-ranked docked poses which "reproduces" the X-ray reference results in success. Thus, in the following, "top-ranked" refers to the three top scored poses. The number three is arbitrary but small enough to avoid inflating success rates unduly. This compromise is reasonable since docking of a macrocycle is likely to be pursued when this compound has elicited particular interest. Then, focusing on the three top poses is conservative, since a good number of docking poses would be analyzed, probably in the context of additional data (e.g. SAR, NMR). The X-ray structure was considered "reproduced" by a docked pose when they deviated by less than 2.0 Å RMSD, a commonly adopted threshold. Due to the size of the macrocycles, a 2.0 Å criterion is more stringent than with smaller drug-like compounds[75]; thus, RMSDs between 2.0 and 3.0 Å for any of the three top poses were considered "marginally failed" here, since those docking models still capture many features of the X-ray binding mode. Even approximate models can be insightful; for instance, identifying the key functionalities involved or not in direct receptor binding could help simplify the macrocycle, or suggest moieties to modify to improve ADMET properties.

Another issue is whether any of the docked conformers came close enough to their bioactive counterpart in terms of internal coordinates, since that is a prerequisite for successful pose prediction. This was analyzed with RMSDs obtained after best-fit of the docked conformers, contrary to the RMSDs assessing binding mode reproduction (section 2.8).

**Table 2. Results for individual test systems, with the three investigated docking protocols.**

| PDB entry[a] | Rigid docking of X-ray conformers (Protocol I) | GOLD Standalone docking (Protocol II) | MT/LMOD-assisted GOLD docking (Protocol III) | No of MT/LMOD conformers[b] |
|---|---|---|---|---|
| 1B6M | Success | Success | Fail | 3597 |
| 1BKF | Fail | Fail | Fail | 3503 |
| 1BXO | Success | Fail | Success | 4085 |
| 1LD8 | Success | Fail | Success | 76 |
| 1Q5D | Success | Success | Success | 2881 |
| 1QY8 | Success | Fail | Success | 638 |
| 1S9D | Success | Success | Success | 875 |
| 1UU3 | Success | Fail | Success | 487 |
| 1W96 | Success | Success | Success | 2442 |
| 1WAW | Success | Fail | Success | 2302 |
| 1Z8O | Success | Fail | Fail | 2473 |
| 2C7X | Success | Fail | Fail-Marginal | 2143 |
| 2CD8 | Success | Fail | Success | 938 |
| 2HFK | Fail | Fail | Fail | 385 |
| 2HW2 | Success | Success | Success | 1663 |
| 2IYA | Success | Fail | Success | 3184 |
| 2PGJ | Success | Fail | Fail | 3521 |
| 2Q0U | Success | Fail | Fail-Marginal | 3846 |
| 2V52 | Success | Fail | Success | 369 |
| 2VW5 | Success | Fail | Fail-Marginal | 2127 |
| 2WER | Success | Fail | Success | 423 |
| 2WI9 | Success | Success | Fail-Marginal | 2533 |
| 2XBK | Success | Fail-Marginal | Success | 3156 |
| 2XFY | Fail-Marginal | Fail | Fail | 5574 |
| 3BXS | Success | Fail | Fail | 543 |
| 3EKS | Success | Success | Success | 216 |
| 3FAP | Success | Success | Success | 4246 |
| 3INX | Success | Fail | Success | 4220 |
| 3M5L | Success | Fail | Fail-Marginal | 2960 |
| 4CLI | Success | Fail | Fail | 11 |
| 4DPF | Success | Success | Success | 3009 |
| 4KE1 | Success | Success | Success | 2306 |
| 4P3P | Success | Fail | Fail-Marginal | 3154 |
| 4X7Z | Success | Fail | Fail | 3161 |
| 4XHE | Success | Success | Success | 799 |
| 4YLA | Success | Fail-Marginal | Fail-Marginal | 137 |
| 5L30 | Success | Success | Fail | 723 |
| 5L7H | Success | Fail | Success | 414 |
| 5TJX | Success | Fail | Success | 2154 |
| 5TKS | Success | Fail | Success | 2296 |
| 5TO8 | Success | Fail | Success | 279 |

[a]Protein Data Bank entry for the macrocycle-protein complex. [b]Number of MT/LMOD generated conformers per macrocycle. Docking success was when one of the three top scored docked pose was within 2 Å of the X-ray reference. Otherwise, a docking experiment is considered to have failed, although an RMSD of 2-3 Å for one of the three top-scoring poses is annotated 'Fail-marginal'.

**Table 3. Summary of overall results for the three investigated GOLD docking protocols.**

| [a]Performance | | | | |
|---|---|---|---|---|
| | | [d]Rigid docking of X-ray conformers (Protocol I) | [d]GOLD Standalone docking (Protocol II) | [d]MT/LMOD-assisted GOLD docking (Protocol III) |
| [b]Full docking | [e]Success | 38 (92.7%) | 12 (29.3 %) | 24 (58.5%) |
| | [e]Failure | 2 (4.9%) | 27 (65.9 %) | 10 (24.4%) |
| | [e]Fail-Marginal | 1 (2.4%) | 2 (4.9 %) | 7 (17.1%) |
| [c]Conf. sampling | [f]Success | [g]NA | [h]78% | 100% |
| [i]Categories of docking result | | | | |
| [i]Category number | [k]Count | [d]Rigid docking of X-ray conformers (Protocol I) | [d]GOLD Flexible (Default) docking (Protocol II) | [d]MT/LMOD-assisted GOLD docking (Protocol III) |
| 1 | 15 | Success | Fail | Success |
| 2 | 11 | Success | Fail | Fail |
| 3 | 9 | Success | Success | Success |
| 4 | 3 | Success | Success | Fail |
| 5 | 3 | Fail | Fail | Fail |

[a]Performance refers to the result of a [b]full docking experiment, or to its internal [c]conformational sampling component. The [d]docking protocols and their [e]outcome categorization are as described in Methods. [f]Success with the conformation sampling is defined as at least one computed compound conformer being within 2.0 Å of the X-ray. [g]Not applicable since only the X-ray conformer was docked in this protocol. [h]Based on the docking poses generated at the end of each GA run.

The docking results fall into one of five [i]categories, in terms of success/failure across the three docking protocols. [k]Count is the number of systems per category.

### 3.1 Curation of the set of test systems.

The present work devoted significant efforts to the curation of a test set of protein-macrocycle complexes for docking. The full set of 41 protein-ligand complexes was assembled from a subset of 29 systems previously explored for ligand-centric conformational sampling[21], supplemented by 12 additional very carefully selected protein-ligand complexes (section 2.1).

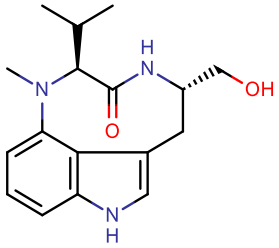
Instead of relying on an automated procedure controlled by a few descriptors (e.g. the resolution of the X-ray structures), we adopted an approach where every selected system was carefully inspected, to apply criteria of diversity and crystallographic quality. This approach is laborious and resulted in a moderately-sized set of 41 systems, since many inspected systems were discarded for some reason, as explained in Methods. Yet, it is reassuring that the first subset of 29 compounds was found by others to match in diversity and molecular complexity a much larger macrocycle set[29, 30].

Examples of the studied macrocycles are presented in Figure 3 (the remaining compounds are shown in Figure S1 of Supplementary Information), ordered by increasing value of their Oprea number of rotatable bonds (NRot), to convey pictorially the range of their complexity. Thus, Figure 3 shows the compounds with the lowest (4YLA, ILV, NRot = 8) and highest (2XFY, ACX, NRot = 30) nominal flexibility, and illustrates the chemical diversity and biochemical relevance of the investigated compounds. Those include approved drugs (e.g. Tacrolimus 1BKF; Lorlatinib 4CLI; Danoprevir 3M5L), other compounds issued from medicinal chemistry programs (e.g. BACE-1 inhibitors 4KE1 and 4DPF, HIV protease inhibitors 3BXS and 1B6M), natural products (e.g. antibiotics Narbomycin 2C7X and Rifampin 2HW2; phytotoxin Pinnatoxin 4XHE), a cyclodextrin (2XFY), and ligands involved in investigations of enzymatic mechanisms (e.g. 4YLA and 3BXS). When possible, ligands with limited side-chains were preferred (e.g. 4CLI, 1S9D, 3BXS, 2XFY) to emphasize the macrocyclic character in the tests.

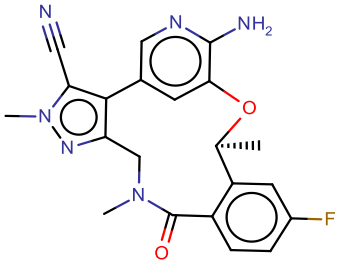
The diversity of the compounds is further documented by the distributions for relevant properties (Figure 4): molecular weight (panel A), NRot (panel B), and the size of the component macrocyclic moieties (panel C). The test set was selected so that these properties cover a broad range, balanced by the intention to emphasize macrocyclic moieties of moderate size, since they are more relevant to medicinal chemistry efforts. Basic statistics on the properties are given in the caption of Figure 4, e.g. the median is 15.0

for both NRot and the number of atoms forming the macrocyclic ring moiety. The minimal ring size of macrocyclic moieties has varied across studies[34]. Incidentally, many synthetic macrocycles are likely to be less complex than the larger compounds in the present test set. Maintaining a balance between the protein targets was also a concern. For example, we refrained from adding more Hsp90 systems to the test set since it already contains three heat shock protein 90.

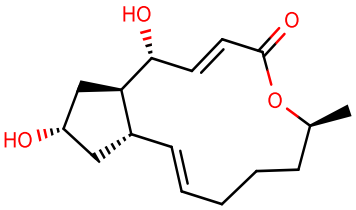
Figure 3



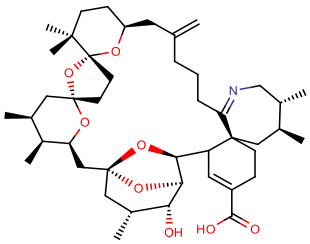
4YLA, ILV, NRot 8



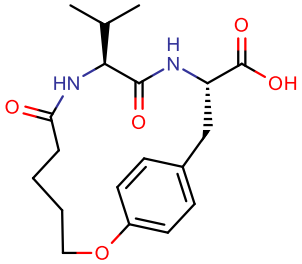
4CLI, 5P8, NRot 8



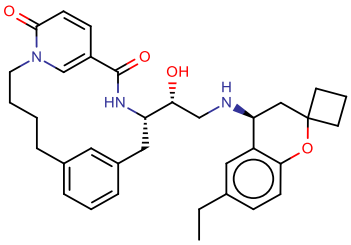
1S9D, AFB, NRot 9



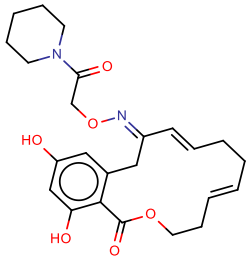
4XHE, 40P, NRot 10



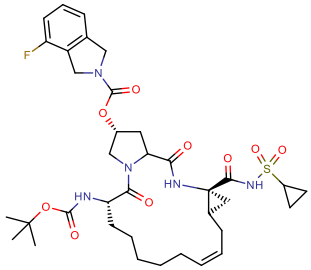
3BXS, DRS, NRot 13



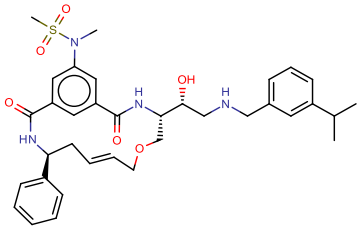
4KE1, 1R6, NRot 14



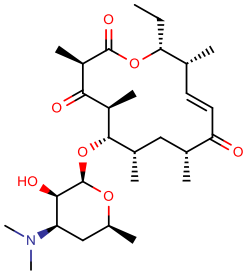
3INX, JZC, NRot 15



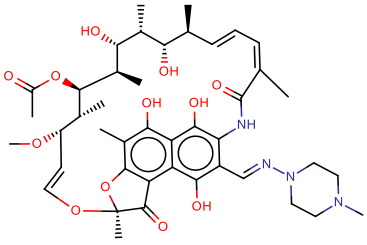
3M5L, TSV, NRot 16



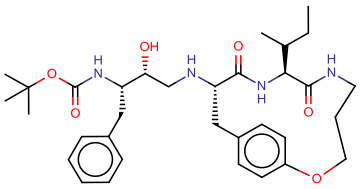
4DPF, 0LG, NRot 18



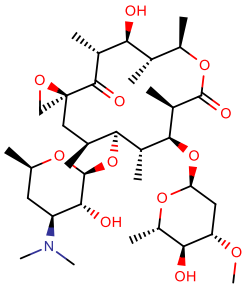
2C7X, NRB, NRot 18



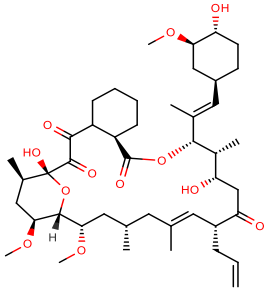
2HW2, RFP, NRot 20



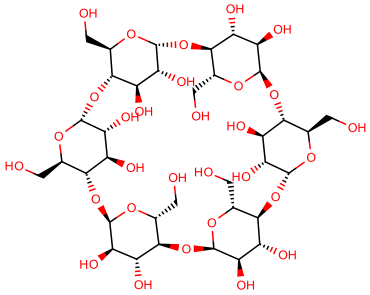
1B6M, PI6, NRot 20



2IYA, ZIO, NRot 23



1BKF, FK5, NRot 26



2XFY, ACX, NRot 30

**Figure 3.** Examples of two-dimensional structures of the studied ligands, arranged by increasing NRot values. Each ligand is annotated with a four-letter PDB entry code, followed by the PDB three-letter ligand code, and the compound Oprea number of rotatable bonds. The two-dimensional structures of the other investigated ligands are given in Figure S1 of Supplementary Information.

Figure 4

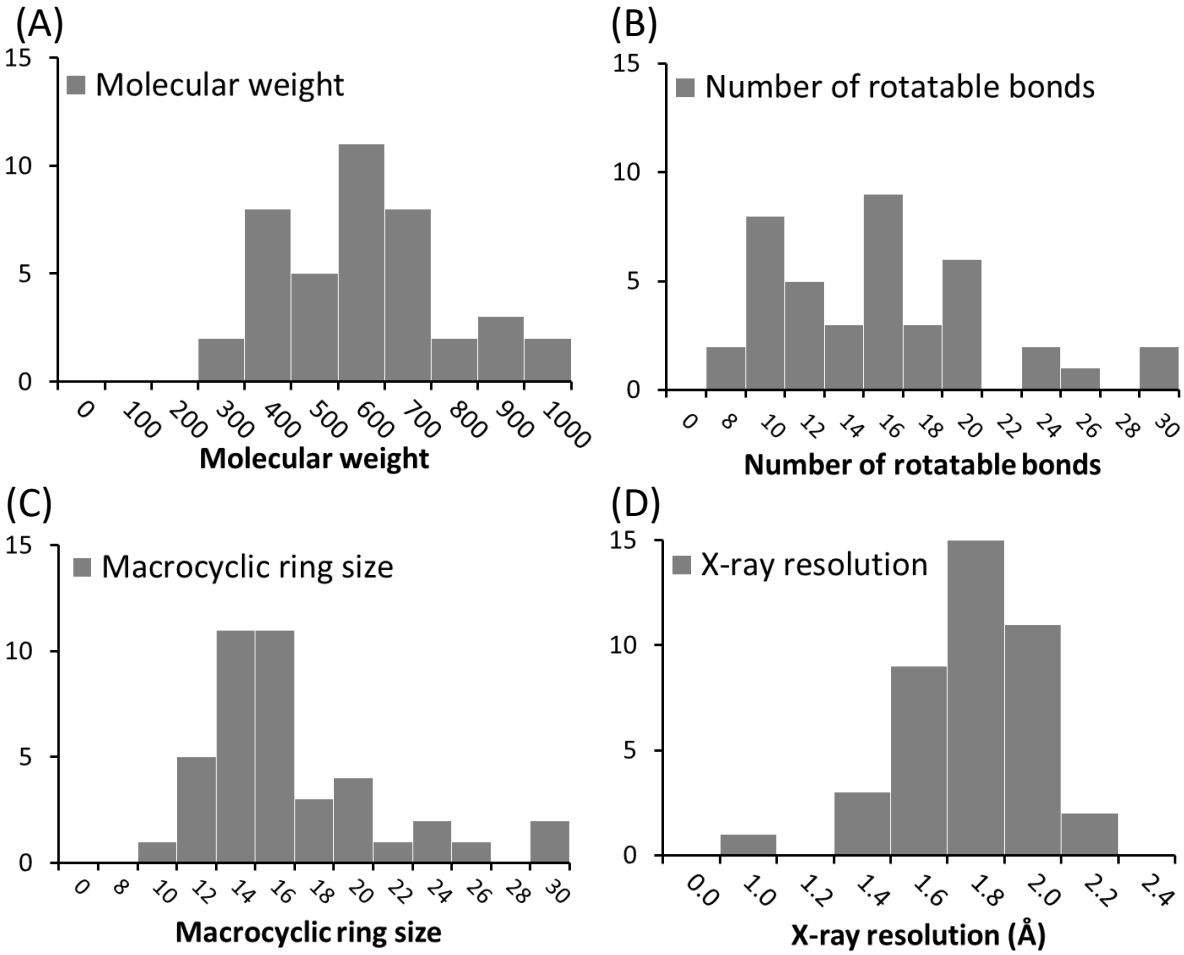



**Figure 4.** Histograms of distributions of selected properties for the 41 compounds included in the docking test set. For every panel, the Y-axis gives the frequencies, and the bins are identified by their bounds as labelled on the X-axis. (A) Compound molecular weight (Median: 520.7, Mean: 551.6, Standard deviation (SD): 175.8), (B) Number of Oprea rotatable bonds (Median: 15.0, Mean: 15.5, SD: 5.5), (C) Number of atoms forming the macrocyclic ring moiety as part of the whole compound (Median: 15.0, Mean: 16.5, SD: 4.6), (D) Resolution (Å) of the X-ray structures (Median: 1.75 Å, Mean: 1.69 Å, SD: 0.23 Å).

The binding sites were examined with the program SiteMap from Schrödinger[76], which returns numerical values for properties characterizing the site as a whole. The SiteMap descriptors for individual systems are given in Table S2 of Supplementary Informatiom. Dscore is an overall drugabbility score, reported alongside descriptors "volume", "exposure", "enclosure", "contact", "phobic" (hydrophobic character), "philic" (hydrophilic character), and "balance" (ratio of phobic over philic). Dscore ranges from 0.84 (1BKF) to 1.36 (5L7H), which is a broad range on the Dscore scale[76], indicative of diversity. The Dscore median and mean are 1.05 and 1.07, respectively. Seven systems (1B6M, 1BKF, 1BXO, 1LD8, 1Q5D, 1QY8, 1S9D) are in the range considered "difficult" (0.84 ≤ Dscore ≤ 0.98). However, the majority of the Dscore values (34 out of 41) fall in the range (Dscore > 0.98) associated with druggable sites[76], re-inforcing the notion that the selected systems are representative of drug discovery situations. Druggable sites imply the potential for sufficient ligand-protein intermolecular interactions, a sound premise to focus the docking tests on the complexity of the compounds. In other words, on average, docking failures are unlikely to reflect extremely adverse sites, but instead shortcomings in the treatment of the compounds. Overall, the selected 41 systems give a fair coverage of the type of situations that one may encounter when working with macrocycles.

Particular attention was given to the fit of the macrocycle X-ray models to the associated electron density (Figures 1-2). Thus, some systems solved at a nominally promising crystallographic resolution (≤ 2.0 Å) were discarded, in view of weak or insufficient electron density (e.g. PDB entry 5HI5, 1.8 Å resolution, not shown). Inspection of the candidate systems was performed in presence of the symmetry related molecules of the crystal lattice. This indicated that it is not uncommon for bound macrocycles to be in contact with symmetry-related molecules, in addition to the primary protein target (Figure 2). That is not unexpected, since macrocycles can target open and shallow protein surfaces[13, 14], which are more prone to lattice contacts than buried conventional enzyme active sites. For example, in subset one, 5 systems out of 29 present contacts between the macrocycle and the crystal lattice: 1BKF, 1BXO, 2Q0U, 3INX and 3M5L (the contacts in 1BXO, 3INX, and 3M5L are minor). Such systems were kept in subset one, to maintain the identity of subset one with prior studies; also reproduction of such binding modes by docking in absence of the lattice would imply that they are not strongly influenced by the lattice.

However, macrocycles in contact with symmetry-related proteins at non-biological interfaces were excluded from the new second docking subtest, even when there was a convincing fit between the macrocycle and the electron density (Figure 2). Indeed, contacts with the surrounding lattice may affect the compound binding mode, which would obscure comparison to the docking output. Lattice contacts may also influence the target protein binding site and affect the macrocycle binding mode indirectly. Thus, we suggest that this be considered when devising future docking studies for macrocycles. However, including or not systems based on the present criteria (fit to electron density and lattice contacts) is somewhat subjective, since it requires an overall judgement taking multiple factors into account. There is the danger of being overly conservative when eliminating systems which could be largely relevant, despite uncertainties. Based on graphical inspection and chemical intuition, there is no strong obvious indication that the macrocycles X-ray binding modes shown in Figure 2 are dominated by lattice contacts. The contacts between those macrocycle and their primary protein targets appear sensible, however with greater uncertainty about the peripheral, lattice-exposed, substituents. That is why it was considered prudent to down-prioritize such systems in subset two, but this approach may have to be revisited as more information becomes available. Thus, assembling growing and suitable docking test sets for macrocycles will be an ongoing learning process.

**3.2 Rigid docking of X-ray conformers (protocol I).**

Protocol I rigidly docked the X-ray conformation of the compounds, after randomization in translation and rotation. Importantly, this tests the GOLD configurational search without the complications of internal conformational sampling. Thus, it tests if the GOLD translational and rotational sampling visits a docking pose close to its X-ray counterpart. That was a sensible starting point, since initially it was not known if the GOLD translational/rotational sampling would perform as well with macrocycles in shallow/open binding sites (e.g. compound PTX in 2Q0U, or compound ZM3 in 4X7Z) as with smaller drug-like compounds in clefts. If the experimental structure is visited, protocol I tests if the scoring function ranks it as one of the top three docked poses (successful outcome). Protocol I is a strong test of the scoring scheme (ChemPLP) since, by supplying and keeping the correct input conformer, uncertainties arising from internal conformational sampling are eliminated. Systems for which this simplified docking approach fails can hardly be

expected to succeed with more demanding docking conditions. So, it is not intended to represent actual docking applications, but instead it provides a baseline comparison for docking protocols which tackle conformational sampling of the ligand.

Tables 2 and 3 show that rigid docking of the X-ray conformers was successful with 38 systems (92.7%) out of 41; this success rate was unchanged if considering only the top scoring pose (Table S3 of Supplementary Information). The success rates across protocols are also compared graphically in Figure 5. This was obtained with the GOLD default number of 10 docking runs. This high success rate implies that, given the correct input conformers for the macrocycle and its binding site, i) GOLD is very likely to visit the experimental binding mode, and ii) the X-ray binding mode scores favorably, being frequently retrieved in one of the top three poses. This is encouraging since it implies that, if the X-ray conformer can be computationally generated, it is likely that GOLD can at least generate a bioactive-like pose, and possibly score it favorably. Examples of top-ranked docking poses obtained with protocol I are shown in Figure 6, alongside their counterpart from protocols II and III. Figure 6 also illustrates the five categories of results (Table 3), discussed later.

The three systems where protocol I failed were 1BKF (tacrolimus), 2HFK and 2XFY (cyclodextrin), for which the closest RMSD to the X-ray structure in the three top poses were 9.5 Å, 5.6 Å and 2.7 Å, respectively. The cyclodextrin of 2XFY, the largest compound in the set, only failed marginally, reflecting the repetition of a same sugar building block in the cyclodextrin. The failure of 1BKF arises from the very open and shallow binding site of tacrolimus, primarily located on a monomer but also involving the interface of the biologically relevant protein dimer. Docking was performed with the relevant dimer, and resulted in top-scored poses straddling the dimer interface, rather than the X-ray observed binding site; docking to a monomer alone was successful, confirming the influence of the interface on the docking outcome. With 2HFK (chain A), the position of the macrocycle is supported by the crystallographic density, but with only loose interactions to the protein (e.g. no direct compound-protein hydrogen-bonds), in a site larger than the compound and partially filled with waters. Also, the macrocycle contacts a dimethylsulfoxide crystallization agent which was removed during preparation of the site for docking. Thus, neither polar anchoring points nor shape recognition can guide a scoring function to the 2HFK experimental binding mode; the docking poses were displaced from their experimental counterpart, driven by non-cognate contacts with the protein. The three systems which

failed with protocol I also failed with protocols II and III (category 5 in Table 3), consistent with expectations.

Figure 5

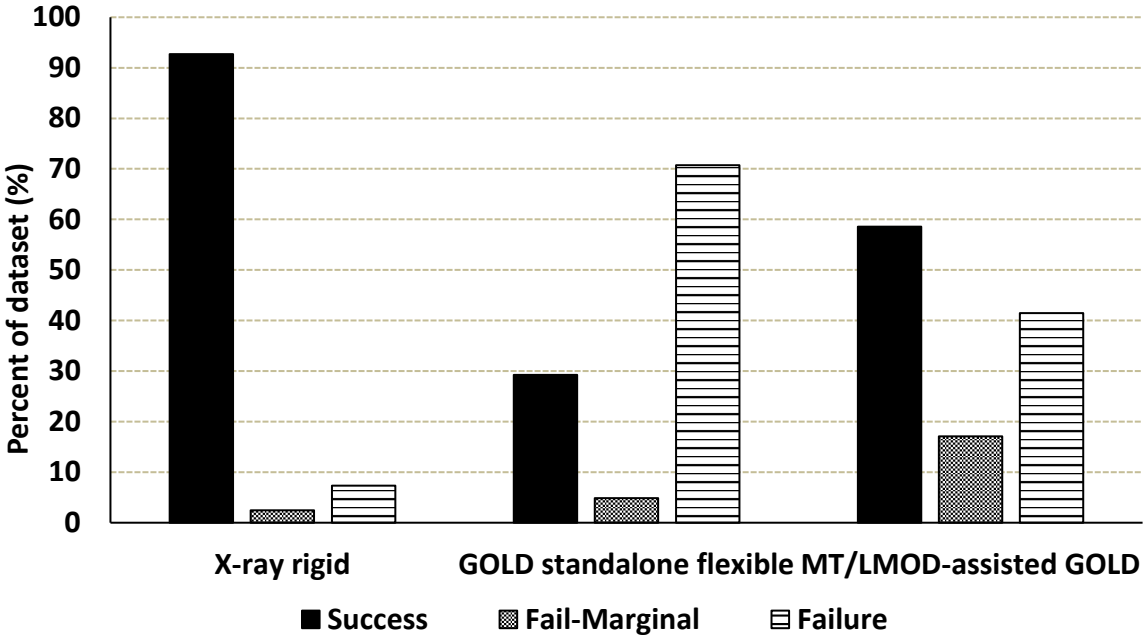
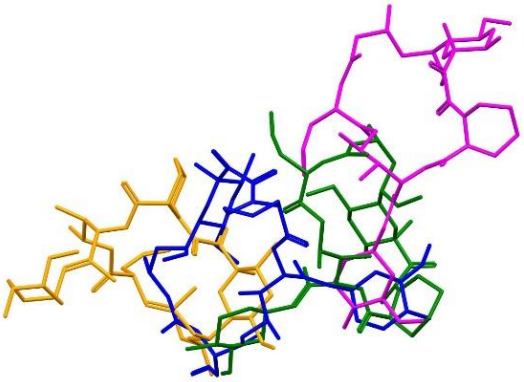
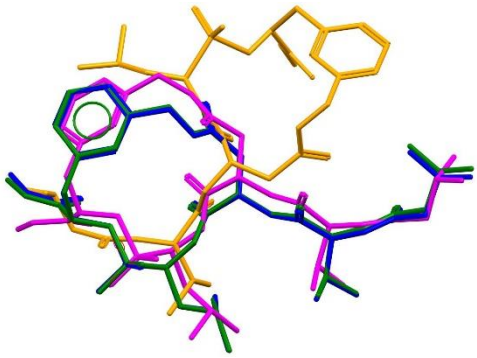



**Figure 5**. Summary of performance with the three investigated docking protocols, for all studied macrocycles. The percent (Y-axis) of Success (black bars), Fail-Marginal (grey bars), and Failure (striated) are given as bar charts for each docking protocol (X-axis). The count of failures includes the cases noted as Fail-Marginal, so the bars add to more than 100% for a given protocol. Briefly, Success was when one of the three top scored docked poses was within 2.0 Å (RMSD) of the X-ray reference; otherwise, the outcome was considered a Failure, although RMSDs of 2-3 Å were annotated as Fail-Marginal.

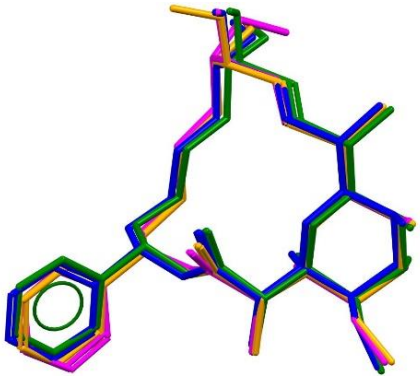
Figure 6



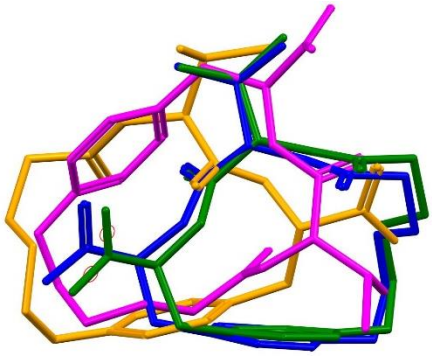
(A) 1BKF F, F, F NRot 25



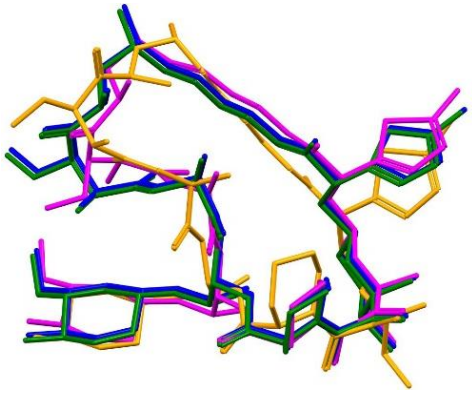
(B) 1BXO S, F, S NRot 20



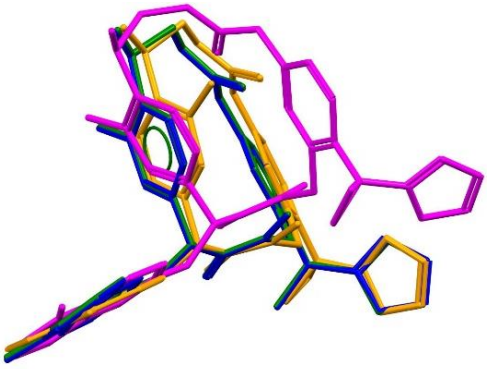
(C) 1W96 S, S, S NRot 16



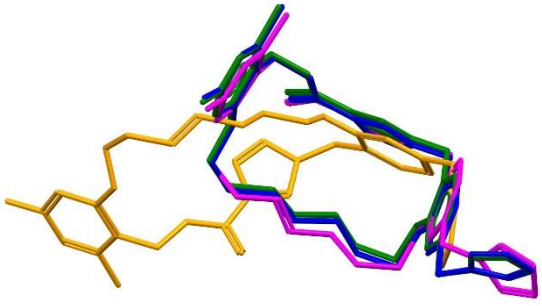
(D) 3BXS S, F, F NRot 13



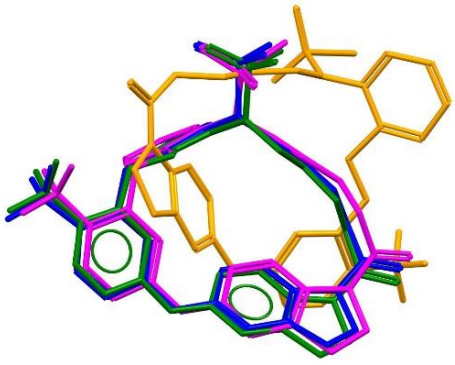
(E) 3FAP S, S, S NRot 29



(F) 5L30 S, S, F NRot 12



(G) 5TJX S, F, S NRot 16



(H) 5TO8 S, F, S NRot 11

**Figure 6.** Illustration of various docking outcomes regarding success (S) or failure (F), by comparison of the X-ray structure (green) to the closest docked pose (by RMSD) from the three top-ranked poses for i) protocol I: rigid re-docking of the X-ray conformer (blue), ii) protocol II: GOLD standalone docking (gold), iii) protocol III: MT/LMOD-assisted GOLD docking (magenta). The systems are identified and sorted by their PDB entry code, labelled in green like the depicted X-ray structure. The annotations S or F refer to success/failure with protocols I, II and III in that order, and colored accordingly. The systems were selected to illustrate the five result categories listed in Table 3, e.g. system 1BXO (panel B) belongs to category 1 (S, F, S: successful protocol I, failed protocol II, successful protocol III). NRot (number of rotatable bonds) is also given. For clarity, hydrogen atoms are not shown.

Even with the correct input X-ray conformations, the high success rate with protocol I was surprising, considering that i) all crystallographic waters were removed prior to docking, and ii) the protonation states were not adjusted based on the observed binding modes. Most of the studied systems involve water-mediated contacts between compound and protein; contacts between compound and water are rather extensive with some systems. Figure 7 shows example of X-ray structures from the test set, for which docking protocols I, II and III were successful, despite the compound being in direct contact with water molecules. Some of these waters mediate hydrogen-bonds to the protein (not shown), while the influence of other waters on the binding mode may be more diffuse. This situation is common with the investigated systems, and could be expected to prevent successful docking. However, this intuition appears refuted by the present results, where protocol I typically recognizes the bioactive binding mode without inclusion of the waters. Such success may be guided by overall shape complementarity between the macrocycle and its cognate site, and probably a few decisive polar interactions, even when details of the water structure are absent. This is encouraging for prospective applications, when one is unlikely to be able to inform docking with relevant water-mediated contacts. The importance of shape complementarity for successful reproduction of macrocycle binding modes by docking has already been noted[32]. However, shape recognition is evidently artificially biased in protocol I, by supplying both the experimentally observed compound and protein conformations. Indeed, protocol I gave the highest docking success rate, which degraded when having to infer the compound conformations in protocols II and III (Table 3).

Figure 7

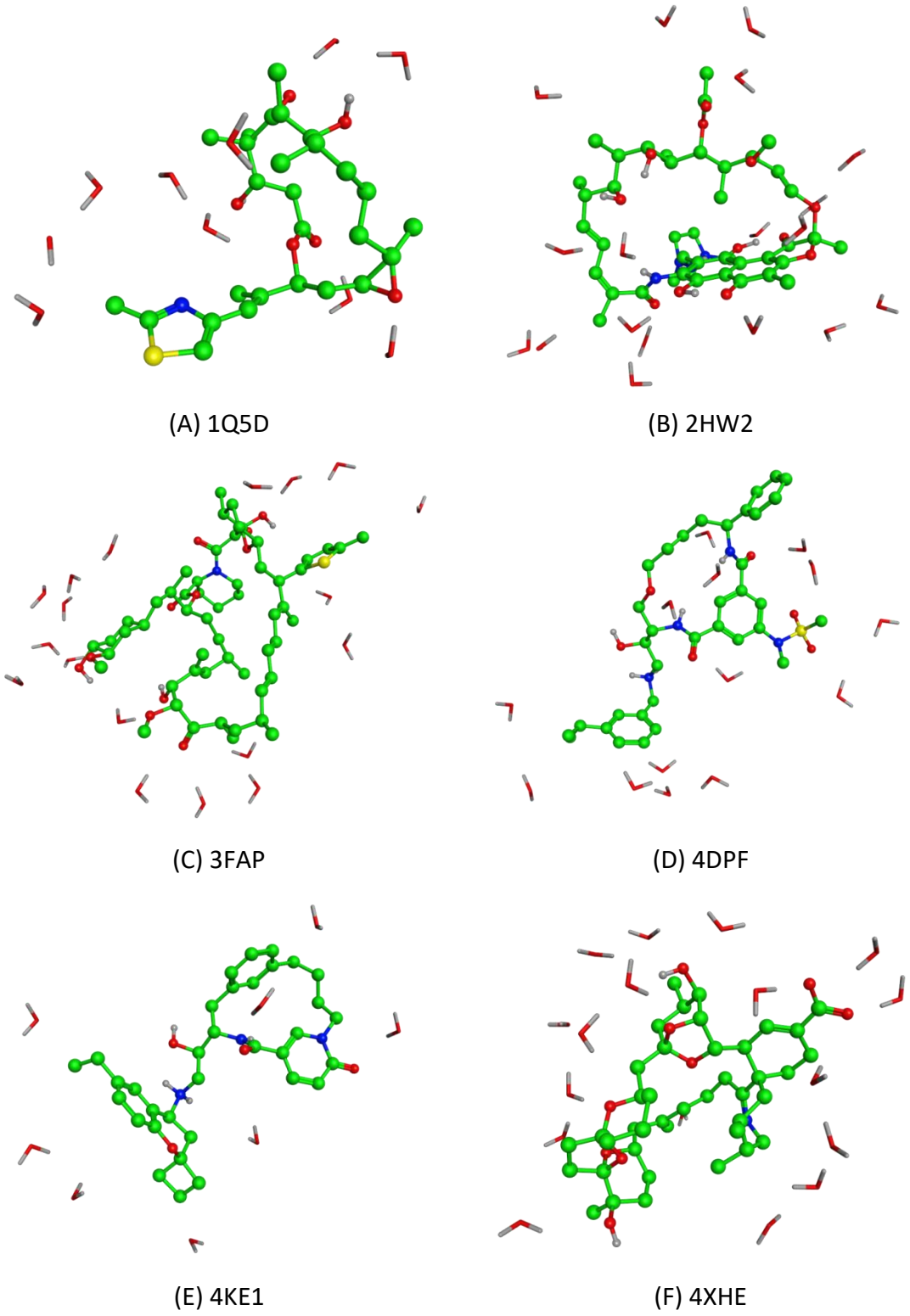



**Figure 7.** Examples of compounds for which GOLD docking was successful in protocols I, II and III, despite multiple contacts between compound and water molecules in the X-ray structures (identified by PDB entry codes). Panels (A) to (F) show the X-ray structures of studied compounds (ball & sticks, green carbons), and every crystallographic water within 3.5 Å of the compound, after addition of hydrogen atoms with the Protonate 3D function of software MOE.

### 3.3 Flexible docking with GOLD-standalone (protocol II).

Building on the previous section, protocol II samples the compound internal conformations, in addition to its rotational and translational degrees of freedom. Compound conformational sampling is an essential part of prospective docking applications, where the bioactive conformer is unknown. With protocol II, the conformers were generated exclusively by the GOLD genetic algorithm at default settings for mutations and recombinations. So, protocol II is how GOLD would typically be used for binding mode prediction, as a standalone tool. That is, success requires GOLD-standalone conformational sampling of the compound, together with the ability to score favourably (among top three poses) a bioactive-like docking mode. With protocol II, the number of docking runs per compound was increased from 10 (default) to 20, to give GOLD a better chance of success.

The results for protocol II (Tables 2-3, Figure 5) show that GOLD-standalone was only successful with 12 of the 41 systems (29.3%), a large drop from the 92.7% success rate with protocol I. If one considers only the top-scoring pose, the success rate of protocol II drops further to 24.4%, since systems 2WI9 and 3FAP change from ”Success” to “Fail-Marginal” (Table S3). Examples of top-ranking docking models obtained with protocol II are shown in Figure 6. The three systems (1BKF, 2HFK, 2XFY) which failed with protocol I also failed with protocol II. Of the 38 systems which successfully docked with protocol I, 26 failed with protocol II. If one monitors the top five or top ten scored poses (instead of the top three), the number of systems with a pose within 2.0 Å of X-ray increases to 14 (34.1%) and 17 (41.5%), respectively. So, examining a broader pool of top ranked poses increased modestly the retrieval of experimental binding modes, while adding noise with mis-docked poses. Thus, the three top-ranked poses are enriched in bioactive-like poses, and tests focusing on those poses strike a reasonable balance.

Having to build the ligand conformers is the main difference between protocols I and II, so it is a key factor behind the large drop in docking success rate from 92.7% to 29.3%. This is consistent with the limited sampling of rings by GOLD by corner flipping [43]. However, conformational exploration and scoring are not independent during population evolution under a genetic algorithm[48], so it is not trivial to dissect precisely the contributions of searching versus scoring to the outcome. The compounds in the docking failures had mean NRot of 15.2 and a mean macrocyclic ring size of 16.4, while the corresponding values for docking successes were 16.7 and 16.0, respectively. Thus, failure versus success could not

be rationalized in terms of compound flexibility or size. Docking successes correspond to slightly less polar compounds, with mean topological polar surface area (TPSA) of 124 Å$^2$, than docking failures (TPSA = 144 Å$^2$). That appears at variance with other tests of GOLD[43, 53], but consistent with another[56]; such observations may depend on the scoring function[56]. One can speculate that more polar macrocycles tend to be associated with more solvent-exposed groups, which are prone to artefactual interactions with the receptor in absence of a robust treatment of solvation during docking; for example, see the 4X7Z case in section 3.4 below.

The present results contrast with the largely more successful GOLD binding mode predictions reported with smaller conventionally drug-like compounds[49, 55]. One of those studies[49] found an 81% success rate in terms of an X-ray-like (RMSD ≤ 2.0 Å) pose being ranked best by ChemPLP. This cannot be extrapolated to the macrocycles since 70.7% of them failed by the more lenient success criterion of the present study (three top-ranked poses). Of course, comparing pose prediction performance for macrocycles and smaller compounds based on a same RMSD criterion of 2.0 Å is difficult, since a given RMSD cut-off is more demanding for larger than smaller compounds[75]. That is why we annotated any of the three top docked poses with RMSD versus X-ray between 2.0 and 3.0 Å as "Fail-Marginal" (Tables 2-3 and S3). Only two systems (2XBK and 4YLA) failed marginally with protocol II; if those were considered successes, the success rate would increase from 29.3% to 34.1%. Overall, protocol II confirms that docking of macrocycles with a conventional approach results in mediocre performance[22, 55], but it is the first time that it is investigated systematically with GOLD.

In principle, one would like to know if the bioactive conformations were visited by the genetic algorithm during the search. However, the user has no convenient way to output the configurations explored by GOLD during its search. Thus, as previously done[49], the present work can only consider the conformers output at the end of the docking runs, to estimate if the X-ray conformation was visited within 2 Å. Thus, the RMSDs between all poses output by protocol II and their X-ray counterpart were calculated after rigid fit of the docked conformers onto the X-ray. Such rigid fit was of course only used to compare the internal conformations between docked and X-ray (and not docking success rates, when computed poses are kept in place). Based on those rigid fits, GOLD-standalone generated at least one internal conformation within 2 Å of X-ray for 32 out of the 41 systems (78%).

Those 32 systems include the 12 with docking success, and 20 with docking failure. That is, 69% of the protocol II failures (20 out of 29) visited the X-ray internal conformation within 2 Å, without yielding a docking success. The percentage of macrocycles for which GOLD-standalone visited the X-ray conformer is maybe surprisingly high, however it was influenced by the size and flexibility of the compound. The 32 compounds with at least one computed bioactive-like conformer (RMSD ≤ 2.0 Å) had a mean NRot of 14.4 and a mean macrocyclic ring size of 15.9, while the equivalent values for 9 other compounds (RMSD > 2.0 Å) were 19.3 and 18.4, respectively. Thus, greater flexibility did hinder retrieval of the bioactive conformer, and therefore docking success. The degradation of binding mode prediction with increasing ligand flexibility has been noted before, typically with non-macrocyclic chemotypes[52, 54, 56, 57]. Smaller macrocyclic moieties can also suffer from conformational sampling limitations with GOLD, as was the case with the smallest macrocycle 4YLA, despite that compound having only 8 rotatable bonds. None of the output GOLD conformers for 4YLA recovered a ring puckering mimicking the X-ray reference, which controls the orientation of the substituents, in turn affecting the binding mode. This further justifies the inclusion of this smaller macrocycles in the test set. GOLD has an option to exploit experimentally determined ring conformations from the Cambridge Structural Database as ring templates[77, 78]; 1274 templates were available when the present work was performed, but they did not contain rings of more than 7 atoms; hence this feature of GOLD is not yet suitable for macrocycles[78]. In principle, GOLD could use ring templates obtained outside the Cambridge Structural Database, maybe generated by theoretical means. Such approach gave good results in a study of macrocycle docking with Glide[34], and it might strengthen GOLD-standalone capabilities in the future.

Of course, visiting the bioactive internal geometry is not sufficient for favorable scoring and docking success. There were five failures (1BXO, 2C7X, 3INX, 3M5L, 5TO8) for which GOLD-standalone produced a docking pose (hence a conformer) within 2.0 Å of X-ray, but which did not score well enough. Interestingly, success may depend on how close the bioactive conformation is reproduced within 2.0 Å. The closest average RMSDs after rigid fit of the docked conformers onto X-ray were 1.8 Å for failures, 1.7 Å for marginal failures and 1.0 Å for successes. That suggests that a particularly close match between experimental and computed conformers matters for recognition of the correct binding mode. That is not surprising since the scoring of intermolecular contacts mediating molecular recognition is

highly sensitive to small distance differences (e.g. from a favorable van der Waals contact to a clash). Thus, generating conformers which are only crudely bioactive-like may not be sufficient, and a finer sampling may be required. That conformational sampling limited the success of GOLD-standalone docking is consistent with the results obtained with protocol III (next section).

**3.4 MT/LMOD-assisted GOLD docking (protocol III).**

With protocol III, the compound conformers were generated outside of GOLD with the Mixed torsion/Low-mode (MT/LMOD) approach, with enhanced search parameters reported earlier[21]. This MT/LMOD method is geared towards generating conformers close to the bioactive X-ray structure for large and flexible compounds[21]. Under MT/LMOD the whole covalent structure has flexibility (including bond lengths and valence angles). MT/LMOD produced a conformer within 2.0 Å of the X-ray reference (internal coordinates) for all investigated macrocycles (Table 3). Actually, MT/LMOD generated a conformer within 1.5 Å of X-ray for all 41 compounds, and within 1.0 Å for 33 compounds. This contrasts with GOLD-standalone, where only 9 compounds had a conformer within 1.0 Å of X-ray. The number of MT/LMOD conformers per compound is in Table 2. Each MT/LMOD conformer was submitted to 10 rigid docking runs with GOLD, so the sampling per compound was very thorough with protocol III. Since 10 rigid docking runs were sufficient for the high success rate of protocol I, it seemed sensible to apply the same docking regime to a putative bioactive-like conformer embedded in the MT/LMOD ensemble.

Protocol III produced a top-ranked pose within 2.0 Å of the X-ray reference for 24 (58.5%) systems (Tables 2-3, Figure 5). This is twice the success rate obtained with protocol II (29.3%). In addition, 7 systems (17.1%) failed marginally with protocol III; if those were considered successes, the success rate would increase from 58.5% to 75.6% (34.1% equivalent with protocol II). If one uses the top five or top ten poses (rather than top three) to estimate the success of protocol III, the number of successes increases from 24 (58.5%) to 27 (65.9%) and 28 (68.3%), respectively. As with protocol II, these increases are marginal and confirm that the three top-ranked poses represent a sensible trade-off to evaluate docking success. Actually, the success rate for protocol III was the same when considering the three top-ranked poses or only the single top-scoring pose (Table S3). This further strengthens the value of protocol III over protocol II, since the success of protocol II

degraded slightly when considering only a single top-scoring pose. Of course, the success rates obtained with protocols II and III benefit from using a protein binding site known to be matched to the docked ligand. Thus, less flattering success rates are expected in prospective applications, when uncertainties on the detailed structure of the binding site come into play.

Based on the success/failure outcomes per protocol, one can analyze the results in terms of five categories (Table 3). The largest category (Count = 15) corresponds to docking success, failure and success, for protocols I, II and III, respectively (e.g. panels B, G, H in Figure 6). That is, docking failures with protocol II turned successes with protocols I and III for 15 systems (36.6%). The second largest category (Count = 11) corresponds to docking failure with protocols II and III (e.g. panel D in Figure 6), highlighting that modelling the binding mode of macrocycles remains very challenging. Nine (~22.0%) systems were docking successes across all protocols (category 3, panels C, E in Figure 6). As noted above, only 3 systems failed with all protocols (category 5, panel A in Figure 6). Some top-ranked poses were remarkably close to X-ray (Figure 6), especially with protocol III (e.g. magenta poses with 1BXO, 1W96, 3FAP, 5TJX, 5TO8). Instances where protocol III failed are also shown (e.g. 1BKF, 3BXS, 5L30). Figure 6 also illustrates that compound flexibility (NRot) does not easily explain success/failure.

The factors determining success or failure with protocol III are intricate and proved difficult to tease out; this was analyzed in terms of properties of the binding sites or the ligands. The compounds which succeeded or failed had mean NRot of 14.9 or 16.3, respectively, a mean macrocyclic ring size of 16.3 or 16.7 and TPSA of 128 $Å^2$ or 152 $Å^2$, respectively. Thus, more flexible and polar compounds were maybe more prone to fail. The effect of compound polarity on the outcome of protocol III echoes that observed with protocol II; system 4X7Z discussed below gives an example where a solvent-exposed polar group was mis-docked by forming instead spurious interactions with the protein.

Section 3.1noted that the tested sites are not concentrated at the “difficult” end of the druggability spectrum, according to the SiteMap analysis and Dscore values. This strengthens the interpretation of the different performance between protocols II and III in terms of ligand treatment (i.e. conformational sampling), rather than properties of the sites as a potential confounding factor. Of the seven “difficult” systems with Dscore ≤ 0.98 (1B6M, 1BKF, 1BXO, 1LD8, 1Q5D, 1QY8, 1S9D), five (71.4%) were successful with protocol III

(1BXO, 1LD8, 1Q5D, 1QY8, 1S9D). Table 4 reports the average Dscore values categorized by systems where protocol III succeeded or failed. On average, the values of the SiteMap descriptors were similar for these two categories. It is maybe counterintuitive that the successful sites had a slightly larger volume than those which failed. So, no trend from the SiteMap analysis could be discerned in terms of the docking outcome with protocol III. Thus, we turned to the examination of the docking output for particular systems, which was sometimes more informative than overall site descriptors.

**Table 4.** Characterization of the binding sites with SiteMap descriptors[a], categorized by docking success or failure with protocol III.

| Docking outcome[b] | Dscore | size | volume (Å$^3$) | exposure | enclosure | contact | phobic | philic | balance[c] |
|---|---|---|---|---|---|---|---|---|---|
| Success | 1.1 ± 0.1 | 180.3 ± 76.0 | 494.5 ± 213.1 | 0.5 ± 0.1 | 0.8 ± 0.1 | 1.0 ± 0.1 | 1.1 ± 0.8 | 1.0 ± 0.3 | 1.4 ± 1.5 |
| Failure | 1.1 ± 0.1 | 150.1 ± 55.7 | 451.8 ± 200.8 | 0.5 ± 0.1 | 0.8 ± 0.1 | 1.0 ± 0.1 | 1.0 ± 0.5 | 1.0 ± 0.2 | 1.2 ± 0.9 |

[a]The descriptor names are literally as given by SiteMap. The mean ± standard deviation of every descriptor is given. The mean was calculated over all systems falling in one of two categories defined by the docking outcome[b] with protocol III, success or failure. [c]balance is the ratio of phobic divided by philic for individual systems.

Protocol III produced a docking pose within 2.0 Å of X-ray for all systems except 2HFK (40 out of 41, 97.6%), consistent with internal geometries within 2.0 Å of X-ray supplied for all compounds by MT/LMOD. For 2HFK (Dscore = 1.14), the closest RMSD between docked (protocol III) and X-ray was 2.4 Å (rank 1101), not a surprise when considering the nature of that system, as discussed in section 3.2. The other 40 systems indicate adequate exploration of rotations and translations by GOLD, even in sites which are broad and open (e.g. 1BKF, see below). For 16 systems, the bioactive-like docked pose within 2.0 Å of X-ray were not ranked among the three top-ranked poses by ChemPLP. For example, with 1BKF, 1B6M and 4X7Z, the three top-ranked poses differed from their X-ray references by RMSDs > 8.0 Å, but poses with RMSD ≤ 2.0 Å were observed at ranks 11, 27 and 27, respectively. Thus, success of MT/LMOD-assisted docking was primarily limited by scoring rather than sampling, not a trivial point for that class of systems. This echoes the state-of-the-art in docking of smaller compounds where scoring uncertainties are commonly identified as the main limitation[50, 52, 56, 60].

Visualization of top-ranked mis-docked poses can sometimes suggest why they spuriously scored better than their bioactive-like counterpart. With 4X7Z (Dscore = 1.05), the X-ray structure (chain A) of the macrocycle includes an amine moiety linked to a six-membered aliphatic ring which protrudes into the solvent. In the top-ranked poses, the entire 4X7Z compound makes contacts with the protein, including the moiety which should be mostly solvent-exposed, resulting in artificially favorable protein-ligand interaction scores. This was also an issue with 1BKF, where a large fraction of the compound is solvent-exposed, in a particularly broad and shallow site (Dscore = 0.84). Indeed, 1BKF is a homodimer with two instances of the compound in the vicinity of each other, so removal of both compounds for docking created a particularly open and challenging docking surface. Indeed, 1BFK docking failed with all protocols. Yet, the X-ray binding mode was retrieved by protocol III at rank 11 (RMSD = 1.6 Å), which may be considered a fair outcome. System 1B6M is the HIV-protease (Dscore = 1.06), also a homodimer, where the asymmetric macrocycle straddles the interface of the two protein monomers forming the binding site. The top-ranked poses are, broadly speaking, flipped relative to their 1B6M X-ray counterpart, illustrating the difficulty of capturing/scoring the subtle balance of interactions favoring an asymmetric experimental binding orientation in an HIV protease binding site.

With 1B6M, the failure of protocol III is unrelated to our choice of standard protonation states for the (deprotonated) catalytic aspartic acids (see Methods), since the X-ray structure indicates negatively charged carboxylates bridged by a cationic amine in the inhibitor. However, keeping standard protonation states probably contributed to the failure of protocols II and III with HIV protease 3BXS (Dscore = 1.11), where the reference compound (Chain A) bridges the catalytic aspartic acids with another carboxylic acid. At least one of those carboxylic groups (ligand and/or protein) must be protonated in the experimental conditions, but such bespoke preparation of this system was considered unhelpful in the present work. Another complication with 3BXS is that its binding site contains two copies of the compound contacting each other (both removed for docking), but adopting different binding modes. Thus, systems like 1BKF, 2HFK and 3BXS are genuinely extremely challenging, and possibly unrealistically hard for successful docking. Keeping such systems in the test sets, however, has the benefit to remind practioners of those difficulties.

The HIV protease 1B6M was one of three intriguing systems (1B6M, 2WI9, 5L30) which were successes with protocols I and II, but failures with protocol III (category 4, Table 3). 2WI9 (Dscore = 1.25) only failed marginally with protocol III, the compound overall orientation being recovered in the three top-ranked poses, and the macrocyclic moiety positioned close to the X-ray. There was little actual difference in outcome between protocols II and III for 2WI9, both yielding top-ranked models which would be operational for qualitative inferences. With 5L30, the moderately-sized macrocycle is an inhibitor of the coagulation factor VIIa, with specific recognition interactions, e.g. the naphtylamine is well buried in a pocket and hydrogen-bonds Asp189, while the tetrazole hydrogen-bonds Lys60. The 5L30 site is somewhat open (Dscore = 0.98) but appears reasonable for docking work. Accordingly, the three top-ranked poses from protocol II were close to the X-ray reference (Panel F, Figure 6). In contrast, the first pose within 2.0 Å of X-ray from protocol III was at rank 385, while in better ranked poses the tetrazole frequently hydrogen-bonded Lys204 instead of Lys60. This may result from the top-ranked poses having a macrocyclic moiety with a cis-amide rather than the X-ray observed trans-amide, a consequence of this being allowed during the MT/LMOD sampling. The conformation of amide bonds in macrocycles is a source of uncertainty when modelling their binding mode, since their preference for trans is not absolute[23]. Thus, 5L30 is an example where allowing a cis amide among the MT/LMOD

conformers, and derived docking models, presented a noisy background, diluting the chances of successful scoring.

Valid docked poses embedded among many other poses may degrade the performance of protocol III with other systems, and probably contributes to the poorer performance of protocol III compared to protocol I. Otherwise, one might expect protocol III to perform closer to protocol I, since a MT/LMOD conformer close to X-ray (internal geometry within 1.5 Å of X-ray for all compounds, within 1.0 Å for 33 compounds) was input into GOLD for all compounds. Alternatively, it is plausible that even small deviations from the compound X-ray conformation detract from retrieval of its correct binding mode, since molecular recognition involves precise geometries. This would be consonant with the observation that docking successes with protocol II had average internal geometries within 1.0 Å of X-ray (section 3.2). Yet, it is remarkable that with a majority of systems (58.5%), the ChemPLP scores could identify a "correct" binding mode among a large number of candidates. In addition, in drug-discovery practice, a bioactive-like pose for a molecule of special interest may well be useful even if not scored very highly, since additional experimental information (SAR[22], NMR data[22, 23], protein mutants) may help discern a relevant docking model.

Overall, the present results indicate that, when applying GOLD to macrocycles, thorough conformational sampling obtained before their rigid docking has advantages over conformational exploration performed concurrently with docking. That supports a broader perspective[22], which advocated generation of macrocycle conformers separately from their subsequent docking. That is also consonant with two recent studies[32, 33]. Indeed, rigid docking of macrocycles with MOE, after generating their conformers with LowModeMD, gave meaningful binding modes in more than 50% of the cases[32]. In addition, rigid docking of pre-generated conformers with MD/LLMod was the approach chosen when comparing AutoDock, DOCK and Glide with 20 macrolides[33]. Yet, those two studies[32, 33] implicitly assumed that docking of pre-generated macrocycle conformers is preferable, since comparison with direct docking relying solely on the tested docking engines was not attempted. However, another investigation[34] found that direct Glide flexible docking with prior ring-templating was as accurate as rigid docking of pre-generated macrocycle conformers; since the former was faster than the latter, direct Glide flexible docking was recommended[34]. Ring-templating with GOLD is currently not available for macrocycles (see

above), and may not be a general solution since it requires the specification of templates for every investigated chemotype. In contrast, docking of thoroughly pre-generated conformers can be applied to any compound class and is easily deployed. The computational performance of protocol III is addressed below.

### 3.5 Computational performance and related issues.

The present docking calculations were all obtained with a commodity desktop PC (see Methods), so there is little computational barrier to performing such experiments. Indeed, the docking compute times per system were 1-2 minutes, and 2-3 minutes, with protocols I and II, respectively. With protocol III, the median and mean docking compute times were 4.7 and 7.5 hours (standard deviation = 10.4 hours). To this, one must add the compute time for the MT/LMOD generation of conformers, which was several hours per system (precise statistics are not available for the MT/LMOD calculations since they were performed on various hardwares, alongside other calculations). Therefore, the MT/LMOD-assisted docking with GOLD was vastly more computationally demanding than GOLD-standalone docking, including the larger storage requirements for the output docking poses. Analysis of the output of protocol III could also be time consuming in prospective applications, and would be facilitated by its categorization in structural families. It remains that the MT/LMOD-assisted docking is certainly tractable on the much longer time scale of drug discovery. The superior docking success rate with protocol III (58.5%) over protocol II (29.3 %) is worth the additional computational effort when one seeks information about a macrocycle of particular interest.

Devising ways to significantly reduce the compute times of protocol III is beyond the scope of this work. However, we briefly discuss how this may be approached. The large variation in the compute times with protocol III (standard deviation above) reflects the spread in the number of input conformers (every conformer was docked), modulated by the size the binding site. Since the extent of the docking region was mapped at default settings, it can hardly be adjusted for computational speed-up. On the other hand, the conformer generation offers opportunities for efficiency gains. First, the computation of the conformers itself could be faster, for instance with methods such as ForceGen[29] and distance geometry algorithms[25], which do not have to generate the conformers sequentially in a trajectory, and therefore can be parallelized. Also, future work should investigate the

impact of reducing the number of input conformers on the docking success. Reducing the number of input conformers with protocol III should reduce the overall docking time, and hopefully mitigate the noise in the docking output. Fewer conformers can of course be obtained with low-mode based methods[21], e.g. by increasing the RMSD cutoff to remove duplicate conformers, by decreasing the number of search steps, or by decreasing the allowed energy window. However, restricting those search parameters degrades the rate of retrieval of the X-ray conformer[21, 26, 79] , and may lower the docking success rate. Indeed, the present results suggest that the presence of a conformer rather close (say, RMSD closer to 1.0 Å than 2.0 Å) to the bioactive structure favors successful docking; so the granularity of the conformational coverage is relevant. Thus, the acceptable tradeoff between fewer input conformers and docking success would have to be determined for every protocol generating the input conformers. This could take advantage of the recently developed conformer generators ForceGen[29] or Conformator[31]. ForceGen reported a retrieval rate of the bioactive conformer comparable to that of enhanced MT/LMOD or LowModeMD, but with markedly fewer conformers[29].

## 4. CONCLUSIONS

As explained in Introduction, there is a regain of interest in macrocycles for medicinal chemistry. Work with macrocycles faces challenges, including with existing docking engines and protocols, which must be re-evaluated specifically with macrocycles[32-34, 59]. Thus, we embarked on an extensive study of macrocycle docking with the mainstream software GOLD. Several aspects of the study are of general interest to the modelling of macrocycle binding modes, including the thorough curation of a set of 41 test systems. The criteria combined to compile the full set included diversity of both protein targets and macrocycle chemotypes, crystallographic resolution, fit of the X-ray models to the electron density, and relevance to medicinal chemistry. A SiteMap analysis supported the diversity of the systems and confirmed that they offer a balance range for docking tests. For a new subset of 12 systems, even more stringent criteria were applied, i.e. we excluded systems with co-crystallization agents in the binding site, and macrocycles in contact with the crystallographic lattice. However, we do not claim that this conservative approach is essential or superior; it may be too strict since it rejects systems with otherwise acceptable

properties. We consider the compilation of the test set an ongoing learning experience, which raises issues requiring special attention with “beyond the rule of 5” compounds.

The present docking tests investigated modelling of the binding mode of macrocycles with GOLD, the established ChemPLP scoring function, and detailed attention to the conformational sampling. When bypassing entirely the complications of conformational sampling, Protocol I yielded a surprisingly high success rate of 92.7%. Thus, when not confounded by internal sampling limitations, a well-balanced scoring scheme like ChemPLP frequently recognizes the correct binding mode, even in extended and shallow sites. However, the success rate degraded dramatically to 29.3% in protocol II, when conformers had to be produced by GOLD. This is clearly poorer than success rates of 70-80% reported with smaller drug-like compounds[46, 49, 50, 80], including with GOLD and the same scoring function[49]. This lesser success with macrocycles is attributable to their increased flexibility under the constraint of ring closure. It is reasonable to expect a similar behavior with other docking engines geared towards smaller drug-like compounds[22, 59], especially if also developed around a genetic algorithm such as rDock[80]. Importantly, a direct comparison between protocols II and III showed that the docking performance can be much improved with a practical workaround addressing the conformational search. Using rigid docking of conformers pre-generated with a method adapted to macrocyles[21], protocol III yielded a success rate of 58.5%, clearly better than GOLD-standalone docking. That is remarkable since it was obtained without assistance from water-mediated interactions or protonation states inferred from the reference X-ray structures. It is also notable that the success rates of protocols I and II were the same whether considering the three top-scoring poses or only the top-scoring pose. Still, the protein structure matching the binding mode and the stereochemistry of the chiral centers were supplied, contributing to the success rate of protocol III. Without knowledge of the bound protein structure, prospective modelling of macrocycles is more challenging, and could be combined with an ensemble of receptor structures[81, 82]. Protocol III is a little more involved than conventional docking, but remains eminently tractable.

The performance of protocol III demonstrates that GOLD and ChemPLP scores can handle macrocycles productively, when supplemented with an adequate input collection of conformers. Pre-generating the conformers with the enhanced MT/LMOD sampling from MacroModel was a natural choice, since it performs favorably at retrieving the macrocycle

bioactive conformers[21]. If MacroModel is not available, one can pre-generate the conformers with alternative means, including LowModeMD[20, 32] or ForceGen[29]. A sampling method likely to retrieve the bioactive conformer with fewer conformers ought to be favored; it shortens the subsequent docking computations, and may curb the noise in the docking output, facilitating the analysis. Ring-templating in GOLD, informed by X-ray-derived macrocycle conformations from the Cambridge Structural Database, is not available but is sometimes proposed as an interesting future development. This would have to meet the formidable challenge of crystallizing chemically diverse macrocycles in the many states representing their conformational freedom. The present results show that computationally generated conformers offer a credible and convenient alternative.

Overall, the present work adds to recent studies[22, 32-34], which allow to be optimistic about the ability of docking studies to contribute to SBDD with macrocycles. Clearly, docking of a macrocycle can suggest valuable working hypotheses for its binding mode. However, in most instances, inferring with confidence the binding mode of a flexible macrocycle solely by computational means remains a hard problem, especially when the precise matching conformation of the binding site is unknown. Therefore, we envisage that docking studies of macrocycles will be most productive in conjunction with bespoke experimental input.

**Supplementary Information:**
i) Table S1 lists the 29 systems in subset 1 of the test set.
ii) Table S2 lists the SiteMap descriptors for the 41 systems in the test set.
iii) Table S3 lists the docking results per system when considering only the single top scored pose.
iv) Figure S1 shows the two-dimensional structures of the studied ligands, which are not shown in the main text.

## References


1. Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The exploration of macrocycles for drug discovery - an underexploited structural class. *Nature Rev. Drug Discov.* **2008,** 7, 608-624.
2. Harvey, A. L. Natural products in drug discovery. *Drug Discovery Today* **2008,** 13, 894-901.
3. Marsault, E.; Peterson, M. L. Macrocycles Are Great Cycles: Applications, Opportunities and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.* **2011,** 54, 1961-2004.
4. Heinis, C. Tools and rules for macrocycles. *Nat. Chem. Biol.* **2014,** 10, 696-698.
5. Yudin, A. K. Macrocycles: lessons from the distant past, recent developments, and future directions. *Chem. Sci.* **2015,** 6, 30-49.
6. Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; APorco, J.; Kozakov, D.; Vajda, S.; Adrian Whitty, A. How proteins bind macrocycles. *Nature Chemical Biology* **2014,** 10, 723-731.
7. Giordanetto, F.; Kihlberg, J. Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? *J. Med. Chem.* **2014,** 57, 278-295.
8. Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobson, M. P. Conformational Flexibility, Internal Hydrogen Bonding, and Passive Membrane Permeability: Successful in Silico Prediction of the Relative Permeabilities of Cyclic Peptides. *J. Am. Chem. Soc.* **2006,** 128, 14073-14080.
9. Whitty, A.; Zhong, M.; Viarengo, L.; Beglov, D.; Hall, D. R.; Vajda, S. Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. *Drug Discov. Today* **2016,** 21, 712-717.
10. Foloppe, N.; Matassova, N.; Aboul-ela, F. Towards the discovery of drug-like RNA ligands? *Drug Discovery Today* **2006,** 11, 1019-1026.
11. Whitty, A.; Kumaravel, G. Between a rock and a hard place? *Nature Chemical Biology* **2006,** 2, 112-118.
12. Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007,** 450, 1001-1009.
13. Doak, B. C.; Zheng, J.; Dobritzsch, D.; Kihlberg, J. How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *J. Med. Chem.* **2016,** 59, 2312−2327.
14. Dougherty, P. G.; Qian, Z.; Pei, D. Macrocycles as protein–protein interaction inhibitors. *Biochemical Journal* **2017,** 474, 1109-1125.
15. Mallinson, J.; Collins, I. Macrocycles in new drug discovery. *Future Med. Chem.* **2012,** 4, 1409-1438.
16. Granzhan, A.; Largy, E.; Nicolas Saettel, N.; Teulade-Fichou, M.-P. Macrocyclic DNA-Mismatch-Binding Ligands: Structural Determinants of Selectivity. *Chem. Eur. J.* **2010,** 16, 878 - 889.
17. Hansen, J. L.; Ippolito, J. A.; Ban, N.; Nissen, P.; Moore, P. B.; Steitz, T. A. The Structures of Four Macrolide Antibiotics Bound to the Large Ribosomal Subunit. *Molecular Cell* **2002,** 10, 117-128.
18. Lawson, A. D. G.; MacCoss, M.; Heer, J. P. Importance of Rigidity in Designing Small Molecule Drugs To Tackle Protein−Protein Interactions (PPIs) through Stabilization of Desired Conformers. *J. Med. Chem.* **2018,** 61, 4283−4289.
19. Balazs, A. Y. S.; Carbajo, R. J.; Davies, N. L.; Dong, Y.; Hird, A. W.; Johannes, J. W.; Lamb, M. L.; McCoull, W.; Raubo, P.; Robb, G. R.; Packer, M. J.; Chiarparin, E. Free

Ligand 1D NMR Conformational Signatures to Enhance Structure Based Drug Design of a Mcl-1 inhibitor (AZD5991) and other Synthetic Macrocycles. *J. Med. Chem.* **2019,** In press.
20. Labute, P. LowModeMD-Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. *J. Chem. Inf. Model.* **2010,** 50, 792-800.
21. Chen, I.; Foloppe, N. Tackling the conformational sampling of larger flexible compounds and macrocycles in pharmacology and drug discovery. *Bioorganic & Medicinal Chemistry* **2013,** 21, 7898-7920.
22. Allen, S. A.; Dokholyan, N. V.; Bowers, A. A. Dynamic Docking of Conformationally Constrained Macrocycles: Methods and Applications. *ACS Chem. Biol.* **2016,** 11, 10−24.
23. Witek, J.; Keller, B. G.; Blatter, M.; Meissner, A.; Wagner, T.; Riniker, S. Kinetic Models of Cyclosporin A in Polar and Apolar Environments Reveal Multiple Congruent Conformational States. *J. Chem. Inf. Model.* **2016,** 56, 1547−1562.
24. Over, B.; Matsson, P.; Tyrchan, C.; Artursson, P.; Doak, B. C.; Foley, M. A.; Hilgendorf, C.; Johnston, S. E.; Lee, M. D.; Lewis, R. J.; McCarren, P.; Muncipinto, G.; Norinder, U.; Perry, M. W. D.; Duvall, J. D.; Kihlberg, J. Structural and conformational determinants of macrocycle cell permeability. *Nat. Chem. Biol.* **2016,** 12, 1065-1074.
25. Bonnet, P.; Agrafiotis, D. K.; Zhu, F.; Martin, E. Conformational analysis of macrocycles: finding what common search methods miss. *J. Chem. Inf. Model.* **2009,** 49, 2242-2259.
26. Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. Macrocycle Conformational Sampling with MacroModel. *J. Chem. Inf. Model.* **2014,** 54 2680-2696.
27. Sindhikara, D.; Spronk, S. A.; Day, T.; Borrelli, K.; Cheney, D. L.; Posy, S. L. Improving Accuracy, Diversity, and Speed with Prime Macrocycle Conformational Sampling. *J. Chem. Inf. Model.* **2017,** 57, 1881-1894.
28. Coutsias, E. A.; Lexa, K. W.; Wester, M. J.; Pollock, S. N.; Jacobson, M. P. Exhaustive Conformational Sampling of Complex Fused Ring Macrocycles Using Inverse Kinematics. *J. Chem. Theory Comput.* **2016,** 12, 4674-4687.
29. Cleves, A. E.; Jain, A. N. ForceGen 3D structure and conformer generation: From small lead-like molecules to macrocyclic drugs. *Journal Comput. Aided Mol. Des.* **2017,** 31, 419-439.
30. Cleves, A. E.; Jain, A. N. ForceGen 3D Structure and Conformer Generation: Update for Surflex Platform v4.2 (rev1: v4.226). www.biopharmics.com/Public/ForceGen-v4.2.pdf (Accessed July 27, 2019).
31. Friedrich, N.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. Conformator: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.* **2019,** 59, 731−742.
32. Anighoro, A.; de la Vega de Leon, A.; Bajorath, J. Predicting bioactive conformations and binding modes of macrocycles. *J Comput Aided Mol Des* **2016,** 30, 841-849.
33. Castro-Alvarez, A.; Costa, A. M.; Vilarrasa, J. The Performance of Several Docking Programs at Reproducing Protein–Macrolide-Like Crystal Structures. *Molecules* **2017,** 22, 136-149.
34. Alogheli, H.; Olanders, G.; Schaal, W.; Brandt, P.; Karlén, A. Docking of Macrocycles: Comparing Rigid and Flexible Docking in Glide. *J. Chem. Inf. Model.* **2017,** 57 190-202.
35. Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-large library docking for discovering new chemotypes. *Nature* **2019,** 566, 224-229.

36. Hawkins, P. C. D. Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017,** 57, 1747-1756.
37. Poongavanam, V.; Danelius, E.; Peintner, S.; Alcaraz, L.; Caron, G.; Cummings, M. D.; Wlodek, S.; Erdelyi, M.; Hawkins, P. C. D.; Ermondi, G.; Kihlberg, J. Conformational Sampling of Macrocyclic Drugs in Different Environments: Can We Find the Relevant Conformations? *ACS Omega* **2018,** 3, 11742−11757.
38. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. **2013,** 5, e201302011
39. Śledź, P.; Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol.* **2018** 48, 93-102.
40. Cummings, M. D.; Sekharan, S. Structure-Based Macrocycle Design in Small-Molecule Drug Discovery and Simple Metrics To Identify Opportunities for Macrocyclization of Small-Molecule Ligands. *J. Med. Chem.* **2019** 62, 6843-6853.
41. Savitsky, P.; Bray, J.; Cooper, C. D. O.; Marsden, B. D.; Mahajan, P.; Burgess-Brown, N. A.; Gileadi, O. High-throughput production of human proteins for crystallization: The SGC experience. *Journal of Structural Biology* **2010,** 172 3-13.
42. Davis, A. M.; St-Gallay, S. A.; Kleywegt, G. J. Limitations and lessons in the use of X-ray structural information in drug design. *DDT* **2008,** 13, 831-841.
43. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997,** 267, 727-748.
44. Lorber, D. M.; Shochiet, B. K. Flexible ligand docking using conformational ensembles. *Protein Science* **1998,** 7, 938-950.
45. Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comp. Chem.* **2010,** 31, 455-461.
46. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J Comput Chem.* **2015,** 36, 1132-1156.
47. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004,** 47, 1739-1749.
48. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins* **2003,** 52, 609-623.
49. Liebeschuetz, J. W.; Cole, J. C.; Korb, O. Pose Prediction and Virtual Screening Performance of GOLD Scoring Functions in a Standardized Test. *J. Comput. Aided. Mol. Des.* **2012,** 26, 737-748.
50. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-Ligand Docking: Current Status and Future Challenges. *Proteins* **2006,** 65, 15-26.
51. Chen, Y.-C. Beware of docking! *Trends in Pharmacological Sciences* **2015,** 36 78-95.
52. Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys.Chem.Chem.Phys.* **2016,** 18, 12964-12975.
53. Perola, E.; Walters, W. P.; Charifson, P. S. A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance. *Proteins* **2004,** 56, 235-249.

54. Korb, O.; Stutzle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009,** 49, 84-96.
55. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006,** 49, 5912-5931.
56. LI, X.; LI, Y.; CHENG, T.; LIU, Z.; WANG, R. Evaluation of the Performance of Four Molecular Docking. Programs on a Diverse Set of Protein-Ligand Complexes. *J Comput Chem* **2010,** 31, 2109-2125.
57. Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.* **2004,** 47, 45-55.
58. Foloppe, N.; Chen, I. Conformational sampling and energetics of drug-like molecules. *Current Medicinal Chemistry* **2009,** 16, 3381-3413.
59. Forli, S.; Botta, M. Lennard-Jones Potential and Dummy Atom Settings to Overcome the AUTODOCK Limitation in Treating Flexible Ring Systems. *J. Chem. Inf. Model.* **2007,** 47, 1481-1492.
60. Chaput, L.; E., S.; Elisée, E.; Iorga, B. I. Blinded evaluation of cathepsin S inhibitors from the D3RGC3 dataset using molecular docking and free energy calculations. *J Comput Aided Mol Des* **2019,** 33, 93-103.
61. Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995,** 245, 43-53.
62. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000,** 28, 235-242.
63. Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.* **2000,** 14, 251-264.
64. Bell, J. A.; Ho, K. L.; Farid, R. Significant reduction in errors associated with non-bonded contacts in protein crystal structures: Automated all-atom refinement with PrimeX. *Acta Cryst. D* **2012,** 68, 935-952.
65. CCDC. GOLD User Guide. A Component of the GOLD Suite. 5.5 Release. www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/748e2e78b0384715a462f17d38dfff5c.pdf (Accessed July 27, 2019).
66. CCG. Molecular Operating Environment (MOE) Software. www.chemcomp.com (Accessed July 27, 2019).
67. Schrödinger. Software Suite. www.schrodinger.com (Accessed 27 July 2019).
68. Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. *J. Comp-Aid. Mol. Des.* **2007,** 21, 681-691.
69. Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *International Journal of Quantum Chemistry* **2013,** 113, 2110-2142.
70. Bochevarov, A. D.; Watson, M. A.; Greenwood, J. R.; Philipp, D. M. Multiconformation, Density Functional Theory-Based pKa Prediction in Application to Large, Flexible Organic Molecules with Diverse Functional Groups. *J Chem Theory Comput* **2016,** 12, 6001-6019.
71. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** 118, 11225-11236.

72. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; R.M., L. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comp. Chem.* **2005,** 26, 1752–1780.
73. Mohamadi, F.; Richards, M. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comp. Chem.* **1990,** 11, 440 - 467.
74. www.ccdc.cam.ac.uk (Accessed July 27, 2019).
75. Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J. Chem. Inf. Model.* **2012,** 52, 2919-2936.
76. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009,** 49, 377-389.
77. Cottrell, S. J.; Olsson, T. S. G.; Taylor, R.; Cole, J. C.; Liebeschuetz, J. W. Validating and Understanding Ring Conformations Using Small Molecule Crystallographic Data. *J. Chem. Inf. Model.* **2012,** 52, 956-962.
78. CCDC. Sampling experimentally observed ring conformations during protein-ligand docking. www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/GOLD_sampling_ring_conformational_space.pdf (Accessed July 27, 2019).
79. Foloppe, N.; Chen, I. Energy windows for computed compound conformers: covering artefacts or truly large reorganization energies? *Future Med. Chem.* **2019,** 11, 97-118.
80. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLOS Computational Biology* **2014,** 10, e1003571.
81. Evangelista, W.; Weir, R. L.; Ellingson, S. R.; Harris, J. B.; Kapoor, K.; Smith, J. C.; Baudry, J. Ensemble-based docking: From hit discovery to metabolism and toxicity predictions. *Bioorganic & Medicinal Chemistry* **2016,** 24, 4928-4935.
82. Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, O.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophysical Journal* **2018,** 114, 2271-2278.

# Supplementary Information

for

**Modelling the binding mode of macrocycles: docking and conformational sampling**

Sarah J. Martin[a,1], I-Jen Chen[b], A.W. Edith Chan[a], Nicolas Foloppe[b*]

[a]*Wolfson Institute for Biomedical Research, University College London, Gower Street, London, WC1E 6BT, UK*

[b]*Vernalis (R&D) Ltd., Granta Park, Abington, Cambridge CB21 6GB, UK*

[1]Present address: Charles River Laboratories, Chesterford Research Park, Saffron Walden, CB10 1XL, UK

[*]To whom correspondence should be addressed:
N. Foloppe email: n.foloppe@vernalis.com
Phone (direct): + (44) (0) 1223 895 338
Fax: + (44) (0) 1223 895 556

**Table S1. List of macrocycle test systems included in the first subset.**

| PDB entry[a] | PDB lig. code[b] | MW[c] | NRot[d] | Res. (Å)[e] | Protein[f] |
|---|---|---|---|---|---|
| 1B6M | PI6 | 596.8 | 20 | 1.85 | HIV-1 Protease |
| 1BKF | FK5 | 804 | 25 | 1.60 | FK506 Binding Protein |
| 1BXO | PP7 | 638.7 | 20 | 0.95 | Penicillopepsin |
| 1LD8 | U49 | 435.5 | 9 | 1.80 | Farnesyltransferase |
| 1Q5D | EPB | 507.7 | 16 | 1.93 | P450 Epoxidase |
| 1QY8 | RDI | 370.8 | 12 | 1.85 | GRP94 |
| 1S9D | AFB | 280.4 | 9 | 1.80 | ADP-Ribosylation Factor 1 |
| 1UU3 | LY4 | 468.5 | 11 | 1.70 | 3-Phosphoinositide Dependent Protein Kinase-1 |
| 1W96 | S1A | 520.7 | 16 | 1.80 | Acetyl-Coenzyme A Carboxylase |
| 1WAW | RIG | 672.7 | 20 | 1.75 | Chitinase |
| 1Z8O | DEB | 386.5 | 14 | 1.70 | 6-Deoxyerythronolide B Hydroxylase |
| 2C7X | NRB | 509.7 | 18 | 1.75 | Cytochrome P450 Monooxygenase |
| 2CD8 | PXI | 453.6 | 16 | 1.70 | Cytochrome P450 Monooxygenase |
| 2HFK | E4H | 296.4 | 11 | 1.79 | Pikromycin Thioesterase |
| 2HW2 | RFP | 823 | 20 | 1.45 | Rifampin ADP-Ribosyl Transferase |
| 2IYA | ZIO | 687.9 | 23 | 1.70 | Oleandomycin Glycosyltransferase |
| 2PGJ | N1C | 542.3 | 10 | 1.71 | ADP-Ribosyl Cyclase 1 |
| 2Q0U | PXT | 857 | 14 | 1.45 | Actin |
| 2V52 | LAB | 395.5 | 10 | 1.45 | G-Actin |
| 2VW5 | BC6 | 502.6 | 16 | 1.90 | ATP-Dependent Molecular Chaperone Hsp82 |
| 2WER | RDC | 364.8 | 9 | 1.60 | ATP-Dependent Molecular Chaperone Hsp82 |
| 2WI9 | 1D2 | 341.5 | 17 | 2.00 | Cytochrome P450 Hydroxylase PIKC |
| 2XBK | XBK | 649.7 | 20 | 1.95 | PIMD Protein |
| 2XFY | ACX | 972.8 | 30 | 1.21 | Beta-Amylase |
| 3BXS | DRS | 362.4 | 13 | 1.60 | HIV-1 Protease |
| 3EKS | CY9 | 507.6 | 11 | 1.80 | Actin-5C |
| 3FAP | ARD | 980.3 | 29 | 1.85 | FK506-Binding Protein |
| 3INX | JZC | 442.5 | 15 | 1.75 | Heat Shock Protein Hsp 90-Alpha |
| 3M5L | TSV | 729.8 | 16 | 1.25 | HCV NS3/4A |

[a]Protein Data Bank entry for the macrocycle-protein complex. [b]Macrocycle ligand code in the PDB. [c]Molecular Weight. [d]Number of Oprea rotatable bonds. [e]Crystallographic resolution. [f]Protein to which the macrocycle is bound.

**Table S2. Characterization of the binding sites of the 41 test systems with SiteMap descriptors[a].**

| PDB entry | size | Dscore | volume(Å$^3$) | exposure | enclosure | contact | phobic | philic | balance |
|---|---|---|---|---|---|---|---|---|---|
| 1B6M | 198 | 1.06 | 506.61 | 0.37 | 0.81 | 1.03 | 0.92 | 1.13 | 0.82 |
| 1BKF | 50 | 0.84 | 156.75 | 0.71 | 0.63 | 0.79 | 1.03 | 0.61 | 1.67 |
| 1BXO | 142 | 1.03 | 459.62 | 0.53 | 0.75 | 0.95 | 0.52 | 1.10 | 0.47 |
| 1LD8 | 137 | 1.06 | 562.86 | 0.63 | 0.78 | 0.88 | 0.64 | 1.07 | 0.60 |
| 1Q5D | 226 | 1.27 | 452.76 | 0.29 | 0.93 | 1.20 | 2.95 | 0.70 | 4.21 |
| 1QY8 | 168 | 1.03 | 396.17 | 0.33 | 0.86 | 1.18 | 1.00 | 1.31 | 0.76 |
| 1S9D | 233 | 1.11 | 451.73 | 0.41 | 0.88 | 1.21 | 1.76 | 1.11 | 1.59 |
| 1UU3 | 133 | 1.12 | 510.04 | 0.53 | 0.79 | 1.02 | 1.46 | 0.91 | 1.61 |
| 1W96 | 56 | 0.99 | 192.08 | 0.69 | 0.72 | 0.95 | 1.89 | 0.44 | 4.29 |
| 1WAW | 168 | 0.94 | 576.93 | 0.51 | 0.79 | 1.06 | 0.24 | 1.45 | 0.16 |
| 1Z8O | 207 | 1.27 | 480.54 | 0.30 | 0.94 | 1.15 | 2.40 | 0.72 | 3.33 |
| 2C7X | 246 | 1.16 | 775.87 | 0.52 | 0.77 | 0.98 | 1.53 | 0.75 | 2.05 |
| 2CD8 | 185 | 1.25 | 650.33 | 0.45 | 0.89 | 1.10 | 1.97 | 0.70 | 2.84 |
| 2HFK | 144 | 1.14 | 561.83 | 0.51 | 0.82 | 1.06 | 1.13 | 0.89 | 1.26 |
| 2HW2 | 237 | 1.08 | 703.15 | 0.46 | 0.85 | 1.15 | 1.11 | 1.12 | 0.99 |
| 2IYA | 223 | 0.95 | 348.49 | 0.36 | 0.95 | 1.23 | 0.19 | 1.72 | 0.11 |
| 2PGJ | 149 | 1.01 | 408.17 | 0.62 | 0.75 | 0.92 | 0.39 | 1.15 | 0.34 |
| 2Q0U | 113 | 1.05 | 283.66 | 0.46 | 0.75 | 0.95 | 0.55 | 1.04 | 0.53 |
| 2V52 | 139 | 1.05 | 311.44 | 0.42 | 0.76 | 0.97 | 0.82 | 1.08 | 0.76 |
| 2VW5 | 204 | 1.05 | 733.68 | 0.57 | 0.81 | 1.05 | 0.73 | 1.15 | 0.64 |
| 2WER | 134 | 1.05 | 401.65 | 0.64 | 0.69 | 0.87 | 0.72 | 0.93 | 0.78 |
| 2WI9 | 195 | 1.25 | 797.48 | 0.53 | 0.85 | 1.02 | 1.77 | 0.62 | 2.88 |
| 2XBK | 416 | 1.15 | 1038.60 | 0.42 | 0.84 | 1.08 | 1.52 | 0.92 | 1.66 |
| 2XFY | 142 | 1.02 | 637.29 | 0.67 | 0.77 | 0.96 | 0.39 | 1.16 | 0.34 |
| 3BXS | 130 | 1.11 | 263.42 | 0.34 | 0.82 | 1.06 | 1.31 | 0.98 | 1.34 |
| 3EKS | 91 | 1.03 | 327.57 | 0.49 | 0.73 | 1.01 | 0.75 | 0.94 | 0.80 |
| 3FAP | 259 | 1.13 | 932.96 | 0.53 | 0.77 | 0.96 | 1.16 | 0.83 | 1.39 |
| 3INX | 242 | 1.09 | 511.41 | 0.44 | 0.80 | 1.06 | 1.07 | 1.01 | 1.06 |
| 3M5L | 72 | 0.85 | 152.98 | 0.64 | 0.63 | 0.79 | 0.61 | 1.00 | 0.61 |
| 4CLI | 65 | 1.00 | 278.17 | 0.61 | 0.74 | 0.93 | 1.22 | 0.61 | 2.01 |
| 4DPF | 248 | 1.00 | 593.39 | 0.44 | 0.79 | 1.03 | 0.52 | 1.26 | 0.41 |
| 4KE1 | 239 | 0.97 | 588.59 | 0.48 | 0.78 | 1.01 | 0.43 | 1.34 | 0.32 |
| 4P3P | 163 | 1.03 | 424.63 | 0.45 | 0.81 | 1.02 | 0.98 | 1.21 | 0.81 |
| 4X7Z | 152 | 1.05 | 427.04 | 0.51 | 0.77 | 1.00 | 1.08 | 1.09 | 0.99 |
| 4XHE | 145 | 1.05 | 710.01 | 0.69 | 0.74 | 0.94 | 0.57 | 1.02 | 0.56 |
| 4YLA | 208 | 1.12 | 471.63 | 0.44 | 0.92 | 1.15 | 1.26 | 1.14 | 1.10 |
| 5L30 | 114 | 0.98 | 321.39 | 0.66 | 0.70 | 0.98 | 0.39 | 1.14 | 0.34 |
| 5L7H | 141 | 1.36 | 236.33 | 0.26 | 0.97 | 1.22 | 3.14 | 0.49 | 6.37 |
| 5TJX | 151 | 1.03 | 308.36 | 0.54 | 0.69 | 0.97 | 0.86 | 0.98 | 0.87 |
| 5TKS | 76 | 0.90 | 163.27 | 0.59 | 0.74 | 1.01 | 0.73 | 1.13 | 0.65 |
| 5TO8 | 139 | 1.08 | 440.07 | 0.58 | 0.77 | 1.00 | 0.92 | 0.99 | 0.92 |

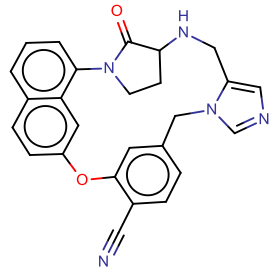
[a]The descriptor names are literally as given by SiteMap. balance is the ratio of phobic divided by philic.

**Table S3. Results for individual test systems based only on the top-scored pose, for the three investigated docking protocols.**

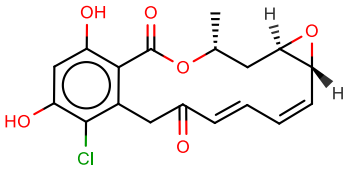
| PDB entry[a] | Rigid docking of X-ray conformers (Protocol I) | GOLD Standalone docking (Protocol II) | MT/LMOD-assisted GOLD docking (Protocol III) | No of MT/LMOD conformers[b] |
|---|---|---|---|---|
| 1B6M | Success | Success | Fail | 3597 |
| 1BKF | Fail | Fail | Fail | 3503 |
| 1BXO | Success | Fail | Success | 4085 |
| 1LD8 | Success | Fail | Success | 76 |
| 1Q5D | Success | Success | Success | 2881 |
| 1QY8 | Success | Fail | Success | 638 |
| 1S9D | Success | Success | Success | 875 |
| 1UU3 | Success | Fail | Success | 487 |
| 1W96 | Success | Success | Success | 2442 |
| 1WAW | Success | Fail | Success | 2302 |
| 1Z8O | Success | Fail | Fail | 2473 |
| 2C7X | Success | Fail | Fail-Marginal | 2143 |
| 2CD8 | Success | Fail | Success | 938 |
| 2HFK | Fail | Fail | Fail | 385 |
| 2HW2 | Success | Success | Success | 1663 |
| 2IYA | Success | Fail | Success | 3184 |
| 2PGJ | Success | Fail | Fail | 3521 |
| 2Q0U | Success | Fail | Fail-Marginal | 3846 |
| 2V52 | Success | Fail | Success | 369 |
| 2VW5 | Success | Fail | Fail-Marginal | 2127 |
| 2WER | Success | Fail | Success | 423 |
| 2WI9 | Success | Fail-Marginal | Fail-Marginal | 2533 |
| 2XBK | Success | Fail-Marginal | Success | 3156 |
| 2XFY | Fail-Marginal | Fail | Fail | 5574 |
| 3BXS | Success | Fail | Fail | 543 |
| 3EKS | Success | Success | Success | 216 |
| 3FAP | Success | Fail-Marginal | Success | 4246 |
| 3INX | Success | Fail | Success | 4220 |
| 3M5L | Success | Fail | Fail-Marginal | 2960 |
| 4CLI | Success | Fail | Fail | 11 |
| 4DPF | Success | Success | Success | 3009 |
| 4KE1 | Success | Success | Success | 2306 |
| 4P3P | Success | Fail | Fail-Marginal | 3154 |
| 4X7Z | Success | Fail | Fail | 3161 |
| 4XHE | Success | Success | Success | 799 |
| 4YLA | Success | Fail-Marginal | Fail-Marginal | 137 |
| 5L30 | Success | Success | Fail | 723 |
| 5L7H | Success | Fail | Success | 414 |
| 5TJX | Success | Fail | Success | 2154 |
| 5TKS | Success | Fail | Success | 2296 |
| 5TO8 | Success | Fail | Success | 279 |

[a]Protein Data Bank entry for the macrocycle-protein complex. [b]Number of MT/LMOD generated conformers per macrocycle. For this table, success is defined when the top scored docked pose was within 2 Å of the X-ray reference. Otherwise, a docking experiment is considered to have failed, although an RMSD of 2-3 Å for one of the three top-scoring poses is annotated 'Fail-marginal'. The results summarized in Table S3 differ from those in Table 2 only for protocol II and systems 2WI9 and 3FAP, which change from Success to Fail-Marginal when using only the top scored pose to define success.

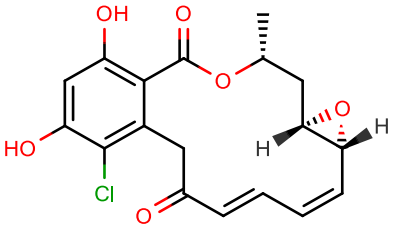
Figure S1



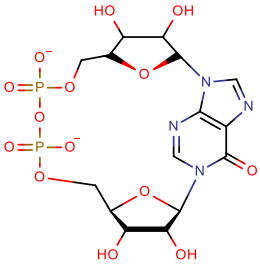
1LD8, U49, NRot 9



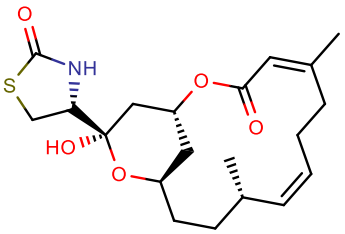
1QY8, RDI, NRot 9



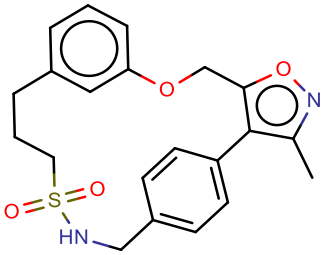
2WER, RDC, NRot 9



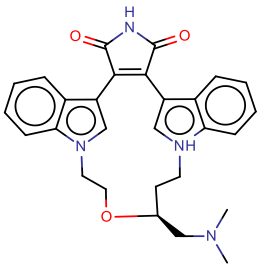
2PGJ, N1C, NRot 10



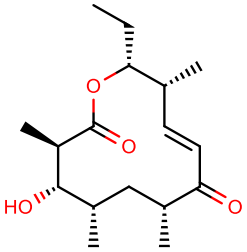
2V52, LAB, NRot 10



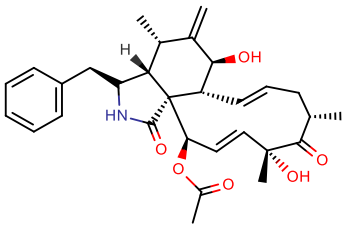
5L7H, 6QG, NRot 10



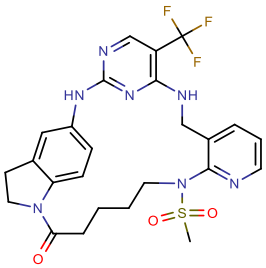
1UU3, LY4, NRot 11



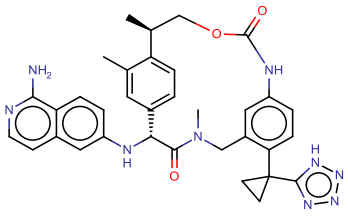
2HFK, E4H, NRot 11



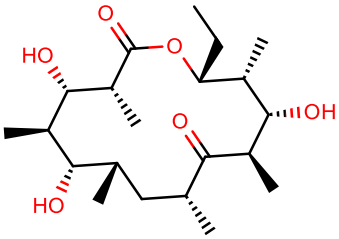
3EKS, CY9, NRot 11



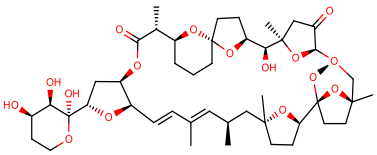
5TO8, 7FM, NRot 11



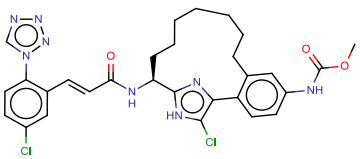
5L30, 70A, NRot 12



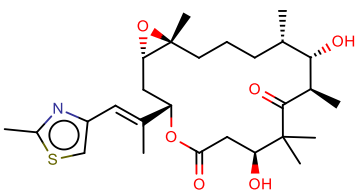
1Z8O, DEB, NRot 14



2Q0U, PXT, NRot 15



5TKS, 7DL, NRot 15



1Q5D, EPB, NRot 16

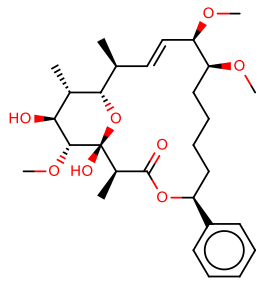
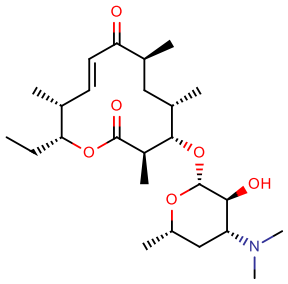
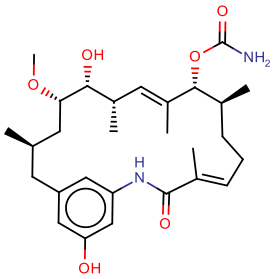
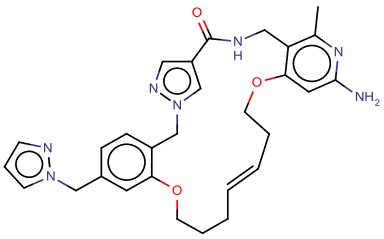
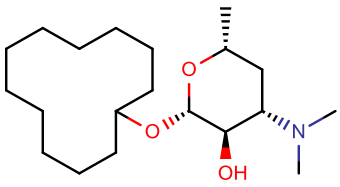
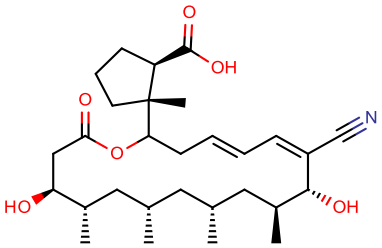
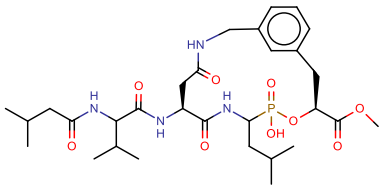
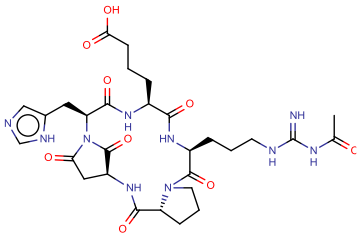
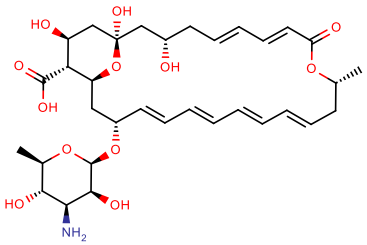
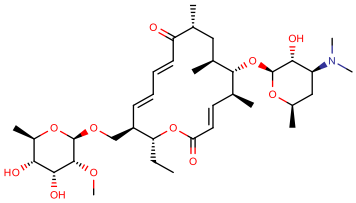
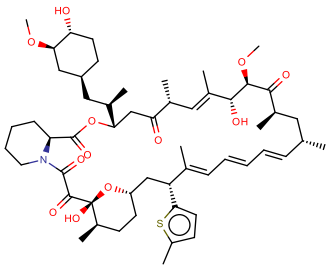
| | | |
|---|---|---|
|  1W96, S1A, NRot 16 |  2CD8, PXI, NRot 16 |  2VW5, BC6, NRot 16 |
|  5TJX, GBT, NRot 16 |  2WI9, 1D2, NRot 17 |  4P3P, 2CR, NRot 19 |
|  1BXO, PP7, NRot 20 |  1WAW, RIG, NRot 20 |  2XBK, XBK, NRot 20 |
|  4X7Z, ZM3, NRot 24 |  3FAP, ARD, NRot 29 | |

**Figure S1.** Two-dimensional structures for 26 of the 41 studied ligands, arranged by increasing NRot values (Oprea number of rotatable bonds). Each ligand is annotated with a four-letter PDB entry code, followed by the PDB three-letter ligand code. The two-dimensional structures of the other investigated compounds are shown in Figure 3 of the main text.