Beyond Small Molecules: *How Advances in 3D Modelling Are Opening New Frontiers in Macrocyclic Drug Discovery*

While traditional small molecules excel at targeting buried active sites, they often struggle with flatter protein surface interactions. Macrocyclic compounds offer a promising alternative, binding with better affinity and selectivity to address previously "undruggable" targets. However, their larger size and structural complexity present unique computational challenges. This article explores how advanced 3D modelling capabilities enable more accurate prediction of macrocyclic conformations, providing practical considerations for effectively working with these promising therapeutic agents.

Macrocycles are an exciting class of compounds to tackle previously undruggable targets. In this article, we'll explore how advances in computational modelling of macrocycles are unlocking efficiency improvements in macrocycle drug discovery that were previously only available for small molecules.

Beyond the Traditional Small Molecule Therapeutic Space

For decades, Lipinski's Rule of Five has provided guidelines for the characteristics of compounds that are more likely to yield oral bioavailability.[1] These include a molecular weight threshold of 500 Da, which has somewhat limited the exploration of chemical space by disfavouring larger molecules.

While small molecules excel at binding to deep, well-defined pockets within proteins, they struggle with broader, flatter surfaces, such as those that characterise protein-protein interactions.[2] These are attractive therapeutic targets as they are fundamental to cell signalling and signal transduction.

Macrocycles, which are defined by ring structures containing 12 or more heavy atoms, bridge the gap between small molecules and biologics. They can bind to flatter, surface-level protein surfaces with high specificity, and despite surpassing the 500 Da threshold, can achieve good oral bioavailability. As a result, the industry has seen greater interest and investment in macrocyclic drug discovery.

Early concerns about metabolic stability were addressed by demonstrating that the constrained, cyclic architecture provides sufficient protection against proteolytic degradation.[3] Additionally, strategic incorporation of non-natural amino acids can further enhance stability whilst maintaining biological activity.

Recent macrocyclic peptides that have been approved by the FDA for therapeutic use include rezafugnin (Rezzayo®), which is used to treat candidemia and invasive candidiasis in adults, and Lurbinectedin (Zepzelca®), for the treatment of metastatic small cell lung cancer.

The Computational Challenge: Why Size and Flexibility Matter

This same size and flexibility that makes macrocycles a promising therapeutic agent, also makes them much more difficult to model using traditional computational methods. With a higher number of rotatable bonds, macrocycles can adopt a vast number of 3D conformations. This flexibility cannot be ignored as it impacts every aspect of a molecule's drug-like characteristics, including on- and off-target potency, and ADMET and physicochemical properties.

Traditional methods that work well for small molecules in modelling their 3D structure and how they bind to a protein target, as well as accurately predicting drug-like properties, falter here as they can't efficiently capture this complex conformational landscape.

As a result, the predictions are either much less accurate, too slow or require an unfeasibly high computational cost. For practical drug discovery, we need modelling that is both accurate and fast enough to integrate into design-make-test cycles.

The Three Pillars of Accurate 3D Macrocycle Modelling

Recent advances have transformed our ability to accurately predict macrocyclic behaviour, making it possible to routinely use 3D macrocycle modelling in hit-to-candidate workflows.

1. Efficient, High-quality and Comprehensive Conformational Sampling

High-quality conformational sampling aims to characterise the biologically relevant ensemble of low-energy 3D conformations, rather than just finding the single lowest energy structure. The bioactive conformation often differs from the global minimum, representing a higher-energy state that is stabilised when the molecule interacts with its target.

Traditionally, conformational ensembles have been generated using templates or pre-calculated torsional libraries, which limit the speed and generality of those methods. However, modern approaches use physically intuitive molecular movements to explore a molecule's conformational space without relying on rigid templates or pre-set torsional libraries. This allows for a more direct and unconstrained exploration of the chemical landscape based purely on molecular energetics in a tractable time and computational cost.

Such advances are highly transformative for macrocycles, and today we can routinely achieve conformational search accuracy for macrocycles that rivals what we see with smaller and simpler non-macrocyclic ligands.[5] Many medicinally relevant macrocycles with up to 21–23 rotatable bonds, once thought "too big to handle", can be explored efficiently at timescales compatible with modern drug discovery pipelines. As a result,

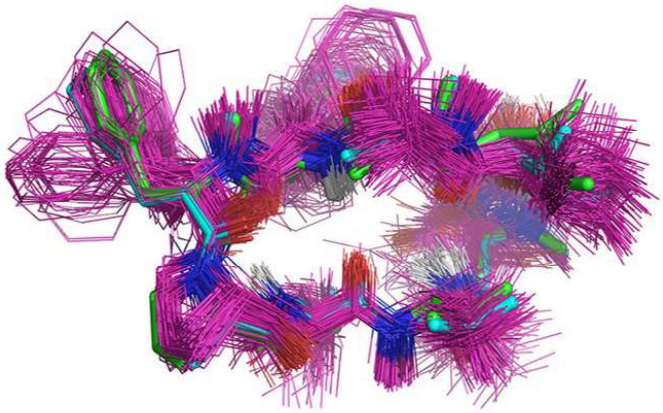


Figure 1. Conformational sampling of macrocyclic peptide, Aureobasidin A. As the size and flexibility of a macrocycle increases, the number of accessible conformers grows exponentially.

conformational sampling is no longer a limiting step, but a practical tool that can be integrated directly into early drug discovery pipelines where speed and reliability are critical.

2. Leveraging Biophysical Restraints

While unrestrained computational searches can be fruitful for very large and complex peptidic macrocycles, it is time-consuming and identifying a conformation close to the bound form becomes increasingly challenging with the increase in size. In these cases, integrating biophysical restraints from techniques such as nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, or cryogenic electron microscopy (Cryo-EM) can help focus the search on biologically relevant regions of conformational space. These biophysical restraints bridge the gap between purely theoretical sampling, narrowing down the vast available space and increasing the likelihood of finding the bioactive conformation efficiently.

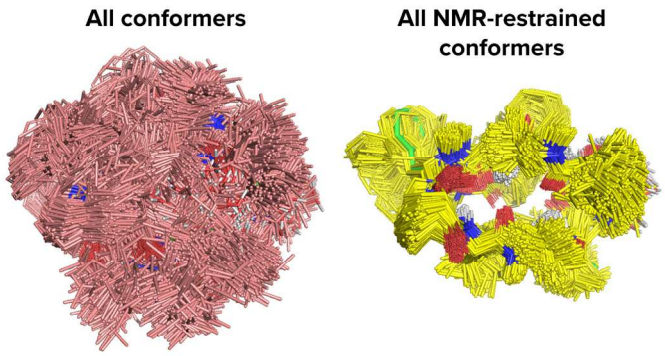


Figure 2. NMR-derived restraints used to narrow down the conformational search space of Aureobasidin A. Such restraints help to increase the likelihood of obtaining a bioactive conformation within the pool.

3. Understanding Molecular Strain

Another critical component of 3D macrocycle modelling is the accurate estimation of the bound ligand strain. This refers to the energetic penalty a molecule pays when its atoms or bonds adopt geometries in the bound state that deviate from their preferred, lowest-energy arrangement in the unbound state. While small, rigid molecules typically experience minimal strain differences in bound versus unbound states, the story is very different for macrocycles due to their size and flexibility.[6]

In the case of large macrocycles, their intramolecular interactions and geometries can be as significant as their ligand-protein interactions. As a result, even seemingly small molecular modifications can drastically change the overall backbone of the molecule, shifting the global energy minimum and introducing additional strain in the bound conformation.

Deriving accurate strain estimates directly from Protein Data Bank coordinates, however, remains challenging. Traditional methods for fitting ligand poses to electron densities determined by X-ray crystallography often rely on a single static conformation, which can exaggerate strain by forcing atoms into geometries that fail to reflect the molecule's intrinsic flexibility. Novel real-space refinement approaches address this limitation by integrating force field calculations with experimental data. By considering ensembles of low-energy conformations compatible with the observed electron density, these methods provide more realistic estimates of strain.[7] Thus, combining experimental structural data with advanced computational refinement enables the discovery and optimisation of macrocycles with realistic, low-strain conformations, supporting more efficient and reliable macrocyclic drug design.

How Does a Better Understanding of Macrocyclic Flexibility Improve Drug Discovery?

These advances unlock computational capabilities previously reserved for small molecules:

1. Precision Molecular Docking

Molecular docking predicts how a ligand would fit into the binding site of the target protein and provides a score that can be used to prioritise or rank compounds for further study. It's a valuable method for virtual screening and lead optimisation, while providing detailed insights into molecular interactions.

For large, flexible macrocycles, successful docking depends critically on accurately sampling ligand conformations, as discussed above. Starting with a diverse, low-energy ensemble of ligand conformations increases the likelihood of finding a geometry that is complementary to the protein pocket during docking. Additionally, it has been shown that refining confirmational searches with NMR or X-ray crystallography data gives a more accurate starting point for docking experiments, generating the correct bound poses of new compounds and thereby better estimating ligand strain and protein-ligand intermolecular binding energy.

Combining improved conformational search with the intelligent use of prior structural knowledge and enhanced scoring functions achieves pose prediction accuracies rivalling those for small molecules. Understanding the binding modes of compounds guides the design of new analogues with better binding affinity, reducing the number of compounds that must be synthesised and tested.

Furthermore, combining conventional docking scores for protein-ligand interactions with a rigorous assessment of strain enables active compounds to be identified and inactive compounds rejected prior to synthesis and testing. Illustrative applications have shown that this can save up to 90% of compound synthesis and testing.[8]

Moreover, understanding whether activity is driven by binding interactions with the protein or molecular strain informs strategies for further optimisation.

2. Enhanced Ligand-based Virtual Screening

3D ligand-based methods are based on the principle that molecules that adopt similar shapes and interactions in three dimensions are likely to show similar biological activities. Ligand-based methods are particularly valuable when a 3D structure is not available for the target protein, meaning that structure-based approaches such as docking cannot be applied. However, ligand-based methods can also complement structure-based virtual screening; the combination often provides better results than either individual approach.[10]

While the size, flexibility and complex conformational landscapes of macrocycles have traditionally made 3D similarity searches challenging, we can now effectively incorporate conformational ensembles and strain estimates to evaluate multiple accessible conformations. This dramatically increases the likelihood of identifying conformations that engage the protein effectively, even when screening chemically diverse molecules that can adopt alternative conformations to preserve key interactions.

The integration of experimental data further enhances these capabilities. Conformational restraints, derived by exploiting NMR data to identify low-energy solution ensembles of a lead compound, can be used to focus ligand-based molecular similarity optimisation. This approach has proven highly effective for predicting bound poses and prioritising lead-compound analogues.[8]

By moving beyond single, lowest-energy conformations to comprehensive ensemble-based approaches that consider shape, electrostatics and pharmacophoric features across multiple poses, these methods now identify active compounds that would have been missed by traditional single-conformation screening.

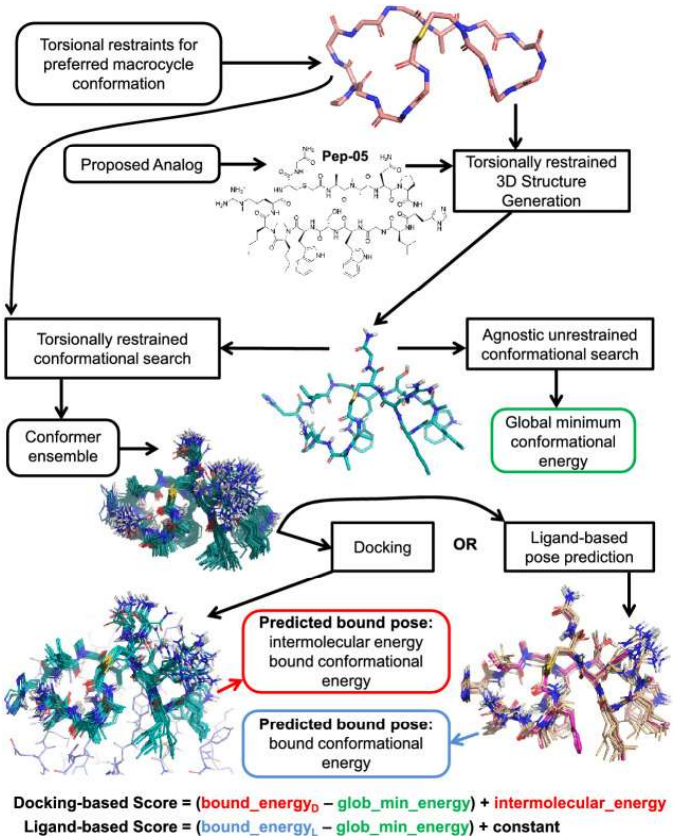
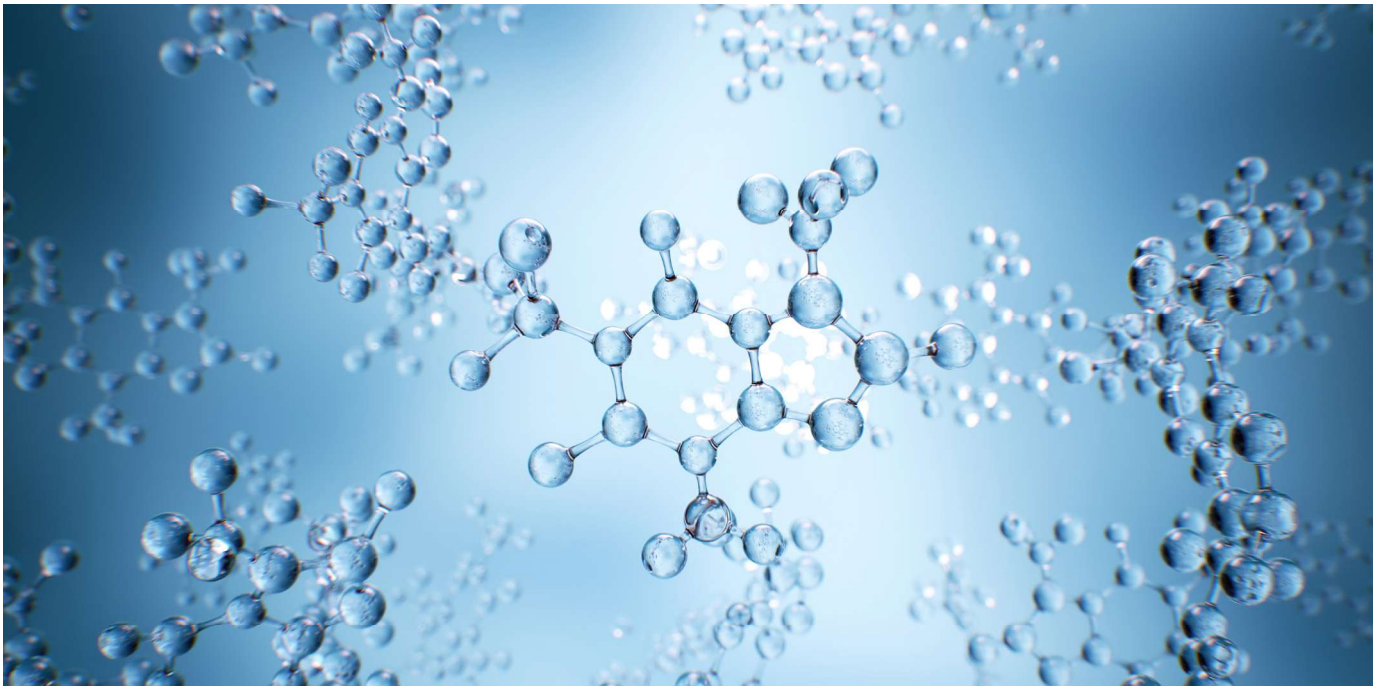


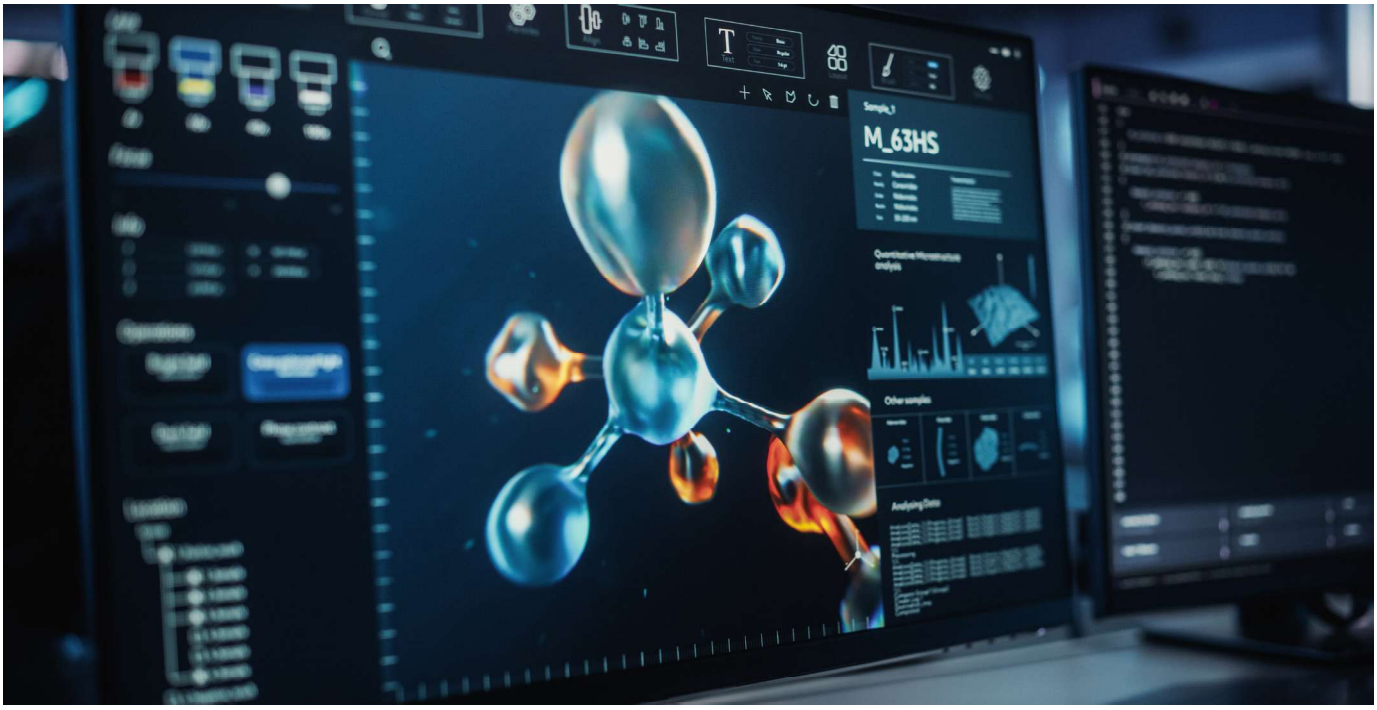
Figure 3. Schematic diagram showing how NMR-restrained macrocyclic conformational preferences can be exploited for enhanced binding pose prediction, either through structure-enabled molecular docking or ligand-based similarity screening experiments. Taken from Jain et al. 2023 (CC BY 4).[8]

Summary

Macrocycles provide a promising solution to a key challenge in drug discovery, addressing protein targets with large, flat binding surfaces that traditional small molecules cannot effectively modulate. These ring structures offer the binding affinity and specificity of biologics, whilst maintaining the potential for oral bioavailability and cell permeability.

Computational modelling of macrocycles has evolved from an intractable problem to a practical reality through three





critical advances. Firstly, modern sampling methods can now accurately and efficiently explore the complex conformational landscapes of these flexible molecules. Secondly, accurate strain estimation allows us to better understand and predict the energetic costs of binding, preventing wasted synthesis efforts on molecules that appear promising but would require prohibitive conformational changes. Lastly, the ability to effectively integrate experimental data from NMR, X-ray and cryo-EM provides real-world constraints that make it possible to model even particularly large and complex macrocycles.

These capabilities now enable drug discovery teams to apply the full toolkit of ligand and structure-based computational methods to macrocycles, to prioritise optimal compounds for synthesis and testing. We're now able to achieve similar accuracy as with predictions for small molecules, and importantly, with sufficient speed and efficiency to incorporate into a routine drug discovery workflow.

REFERENCES

1. Lipinski, C. A.et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 23, 3–25 (1997).
2. Dougherty, P. G., Qian, Z. & Pei, D. Macrocycles as protein-protein interaction inhibitors. Biochem. J. 474, 1109–1125 (2017).
3. Khatri, B., Nuthakki, V. R. & Chatterjee, J. Strategies to enhance metabolic stabilities. Methods Mol. Biol. 2001, 17–40 (2019).
4. Du, Y. et al. FDA-approved drugs featuring macrocycles or medium-sized rings. Arch. Pharm. 358, e2400890 (2025).
5. Cleves A.E and Jain, A N. ForceGen 3D structure and conformer generation: from small lead-like molecules to macrocyclic drugs J Comput Aided Mol Des. 31(5), 4190439 (2017)
6. Jain, A. N. et al. A Distributional model of bound ligand conformational strain: from small molecules up to large peptidic macrocycles. J. Med. Chem. 66, 1955–1971 (2023).
7. Jain, A. N. et al. XGen: Real-space fitting of complex ligand conformational ensembles to X-ray electron density maps J Med Chem. 63(18), 10509-10528 (2020)
8. Jain, A. N. et al. Complex peptide macrocycle optimization: combining NMR restraints with conformational analysis to guide structure-based and ligand-based design. J. Comput. Aided Mol. Des. 37, 11 (2023)
9. Cleves, A. E., Tandon, H. & Jain, A. N. Structure-based pose prediction: Non-cognate docking extended to macrocyclic ligands. J. Comput. Aided Mol. Des. 38, 33 (2024).
10. Cleves, A. E. & Jain, A. N. Structure- and ligand-based virtual screening on DUD-E+: performance dependence on approximations to the binding pocket. J. Chem. Inf. Model. 60, 4296–4310 (2020).

Matthew Segall

Matthew Segall, CEO, Optibrium, has an MSC in Computation from the University of Oxford and a PhD in Theoretical Physics from the University of Cambridge. He has led teams developing predictive models and intuitive decision-support and visualisation tools for drug discovery and has published over 40 peer-reviewed papers and book chapters. In 2009, he founded Optibrium, which develops ground-breaking AI software and services, that improve the efficiency and productivity of drug discovery.

Email: matt@optibrium.com

Himani Tandon

Himani Tandon, Principal Scientist, Optibrium, works in the research division at the Company, developing cutting-edge software solutions that support small-molecule and macrocycle design in drug discovery. Her work focuses on applying 3D structure-based and ligand-based design strategies for lead discovery and optimisation. Himani holds a PhD. in Computational Structural Biology and Bioinformatics from the Indian Institute of Science and completed her postdoctoral research at the MRC Laboratory of Molecular Biology.

Email: himani@optibrium.com