Peptide Macrocycles
Molecules in many dimensions

R Lewis
Cresset User Group Meeting
Cambridge 21 June 2018
Molecules in Many dimensions

- 1D: the molecular formula
- 2D: the smiles string (possibly with stereochemistry and tautomers)
- 3D: a first realisation of the ‘shape’ (possibly with atropisomers, which seem to be coming more frequent)
- 4D: the family of conformations (database ‘standard’)
- 5D: the Boltzmann population in a specified solvent

In most cases, we can make reasonable assumptions and approximations for small molecules
Peptide Macrocycles

• Macrocycles benefit from preorganisation and (sometimes) lower clearance
  – Exceptionally aids permeation too

• We can select through genetic enhancement the peptide macrocycle with the best biochemical binding affinity for a target
  – www.peptidream.com

• How does this translate to more physiological conditions to validate the approach
  – Cellular activity or better

• How can we sample the conformational space of the macrocycle?

• How can we rank the conformations without resorting to long simulations?
  – And in different solvents that mimic water and membrane
Permeability

• “Conformational flexibility has been proposed to significantly affect drug properties outside the rule-of-5”
  – Impact of dynamically exposed polarity on permeability and solubility of chameleonic drugs beyond the rule of 5, Sebastiano et al. J. Med. Chem., 2018, 10.1021/acs.jmedchem.8b00347
  – Caco-2 correlates with 3D PSA
  – Intramolecular h-bonding in apolar environment
    – Different conformations (PSA) in water/membranes?
  – Poster child is cyclosporin A
Cyclosporin in Chloroform and Water
Very large cyclic libraries through redesigned codon tables

<table>
<thead>
<tr>
<th>Macroyclic Natural AAs &gt;10^{12} peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
</tr>
<tr>
<td>MePhe</td>
</tr>
<tr>
<td>Leu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Macroyclic NMe AAs &gt;10^{12} peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
</tr>
<tr>
<td><strong>MePhe</strong></td>
</tr>
<tr>
<td>Leu</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Macroyclic Bulky side chains &gt;10^{12} peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
</tr>
<tr>
<td>MeGly</td>
</tr>
<tr>
<td>Ile</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stapled Natural AAs &gt;10^{12} peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
</tr>
<tr>
<td>Phe</td>
</tr>
<tr>
<td>Stop</td>
</tr>
</tbody>
</table>

**10mer**
11mer
12mer
13mer
14mer
16-mer staple 4:8, i+4 staple 8:12, i+4
MDM2: Novel pharmacophores

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC50 Mdm2 FRET</td>
<td>393 pM</td>
<td>3.8 uM</td>
<td>&gt;10 uM</td>
</tr>
<tr>
<td>IC50 Mdm4 FRET</td>
<td>2.3 uM</td>
<td>4.2 uM</td>
<td>&gt;10 uM</td>
</tr>
<tr>
<td>SJSA1</td>
<td>&gt;30 uM</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IC50 Mdm2 BRET</td>
<td>-</td>
<td>&gt;30 µM</td>
<td>&gt;30 µM</td>
</tr>
<tr>
<td>IC50 Mdm4 BRET</td>
<td>-</td>
<td>&gt;30 µM</td>
<td>&gt;30 µM</td>
</tr>
</tbody>
</table>
You wait for ages....

• Even if you get nothing else from this talk, at least the references are current!

• Several good papers published in last 6 months
  – Just as I was doing this work 😞
Sampling and Scoring

• Strangely reminiscent of the docking/scoring problem
• Can we sample the conformational space of the macrocycle?
• Can we rank the conformations without resorting to long simulations?
• This work looks only at the backbone conformation as an indicator of performance
  – The side chains are very important but there are too many energetic minima
• Most work in this area focusses on one aspect or the other
Macrocycle CA Methodology

• MD – sampling issues due to barriers between conformer families – “Rare events”

• LowModeMD – better sampling but slow
  – Labute, JCIM, 2010, 50, 792
  – Predicting bioactive conformations and binding modes of macrocycles, Anighoro et al., JCAMD, 2016, 30, 841.

• Accelerated MD
  – Peptidic macrocycles – conformational sampling and thermodynamic characterisation Kameik et al., JCIM, 2018, 10.1021/acs.jcim.8booo97
  – Max of 12 rotatable bonds, only simulated in water
Macrocycle CA Methodology

• Knowledge-based
  – Heirarchical flexible peptide docking by conformer generation and ensemble docking Zhou et al. JCIM, 2018, 10.1021/acs.jcim.8b00142

• Kinematics
  – Exhaustive conformational sampling of complex fused ring macrocycles using inverse kinematics Coutsias et al JCTC 2016, 12, 4674

• Mixed methods
Distance Geometry

• Good sampling but need to improve scoring
• Can handle any molecular topology
• Can be quite fast
  – Need to work out when additional embeddings are unlikely to bring more conformers
  – Rate of discovery metric
• Can get trapped by systems with close contacts
  – Prescreen with corina to identify such systems
• OpenEye have a more sophisticated hybrid with attention paid to h-bonding from sidechains (ICCS 2018)
Metrics

• Score by number of ‘reasonable’ conformers generated
  – Radius of gyration, PMI ratios

• Is a good rmsd match to the reference structure found in the ensemble?
  – Various figures given from 2.0 Å to 3.0 Ångström
  – OK for retrospective study, but not helpful if you want to predict.

• Is one of the lowest energy conformer the closest to the native
  – Success here is the real goal!
  – Hosseinzadeh et al., Science 2017, 358, 1461. get there by making their designs and confirming by NMR (70% success).
Secret Sauce – fast QM

• Xtb with implicit solvent
    – a DFTB-based universal semiempirical method optimized for Geometries, Frequencies, and Non-covalent interactions
  – Reliable and performant identification of low-energy conformers in gas-phase and water Cavasin et al. JCIM, 2018, 10.1021/acs.jcim.8b00151
    – GFN-xTB best performance to benchmark, with OPLS3 in solvent a good second in solvent
  – Towards accurate conformational energies of smaller peptides and medium-sized macrocycles. Rezac et al. JCTC, 2018, 14, 1254
    – DFTB accurate to 3-5 kcal/mol

• Can use Cosmo but timings are slower
  – Macrocycle conformational sampling by DFT-D3/Cosmo-RS Gutten et al. JCIM, 2018, 58, 48
    – 1K – 100K cpu hours
Workflow

- Input structures (SDF/SMI) (w. Stereochem)
- Corina
- CA using DG
- Filter by RMSD
- Optimise by QM
- Filter by E
- Spread over multiple cores
  Then combine
  For H2O and CHCl3
Reference sets

- Kindly provided by Paul Hawkins
- BIRD (biologically Interesting Molecule Reference Dictionary)
  - 57 Structures
- CSD
  - 342 structures selected by flexscore
  - Needed some clean up for nitro groups
- MAC
  - 60 structures from Watts et al. JCIM, 2014, 54, 2680

There could be confounding effects due to crystal packing/docking effects
• Corina can generate structures for almost everything, but not necessarily good ones.
Results from DG

• Conversion rate
  – CSD 217 (63%), Mac 55 (91%), Bird 44 (77%)

• 296/316 with a conformation in h2o within 2.5 Ang of reference
Speed and ring size

- Usually quick enough (over multithreaded runs) but occasionally struggles

- Next, will use Corina as a seed and have fewer attempts (2000*Ncores)
Sampling via Span metric
Seems Ok but the Rgyr seems too large

- Span is a metric for sampling
- There is already a 2 Ang rmsd filter for new conformations
Many good matches also are near the Energy minimum – CSD not so.
Some compounds are just hard, with many equivalent minima (YACWIK)

Some do not have strong h-bond networks to limit freedom
RMSD and solvents

- Do find very different conformations in h2o and chcl3
- Evidence of chameleon behaviour
Explicit examples

- Successes
  - VABCUW

- Failures
  - YACWIK
Conclusions

• Distance Geometry linked to QM does have promise in addressing the sampling and scoring issues of peptidic macrocycles, especially with unusual cyclisations.

• Solvent handling could be better
  – Use as seed for explicit solvent methods, especially MD?

• Algorithm aborts and parameter optimisations need to be added
  – Written python/rdkit calling GFN-xtb 2017.6
Acknowledgements

- Michael Schaefer
- Rainer Wilcken
- Peter Ertl
- Andreas Marzinzik
- Prof S. Grimme
- Andrew Dalke for the original contrib code