

# Permuting input for more effective sampling of 3D conformer space

Giorgio Carta<sup>1</sup>, Valeria Onnis<sup>1</sup>, Andrew S Knox,<sup>2</sup> Daren Fayne<sup>1</sup>, David G Lloyd<sup>1</sup>

<sup>1</sup>Molecular Design Group, School of Biochemistry and Immunology TCD, Dublin. <sup>2</sup>School of Pharmacy and Pharmaceutical Sciences, TCD, Dublin.

## Introduction

SMILES strings and other classic 2D structural formats offer a convenient way to represent molecules as a simplistic connection table, with the inherent advantages of ease of handling and storage. Depending on the algorithm employed, the atoms of a connection table defining a molecule can be ordered differently. Upon conversion to 3D coordinates they result in the production of ostensibly the same molecule. In this work we show how different permutations of a SMILES string can affect conformer generation, affecting reliability and repeatability of the results. Furthermore, we propose a novel procedure for the generation of conformers, taking advantage of the permutation of the input strings – both SMILES and other 2D formats, leading to more effective sampling of conformation space in output, and also implementing fingerprint and principal component analyses step to post process and visualise the results.

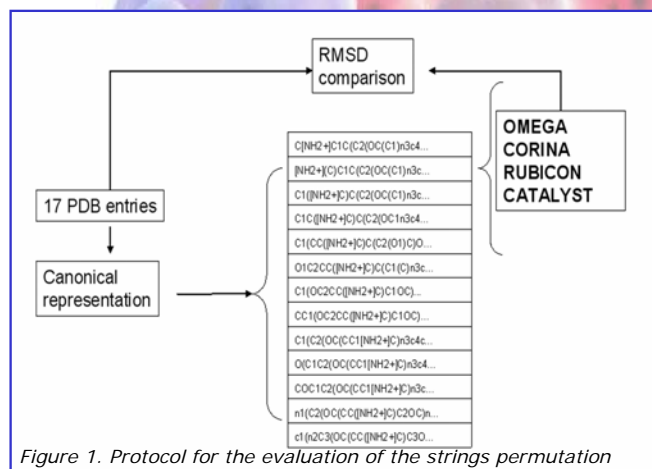


Figure 1. Protocol for the evaluation of the strings permutation

## Methods

As outlined in Figure 1 above, seventeen crystal structures representing a wide variety of different structural characteristics such as size, number of rotatable bonds and features (resolution  $\leq 2.5$  Å, except for 1P93 (2.7 Å) the glucocorticoid complex) were extracted from the PDB. Most of the structures were selected from the list used to validate FlexX 13/Gold 14 and also MASC 6.

The 17 co-crystallised ligands were reduced to canonicalised SMILES strings and the Permsmi utility was subsequently used to produce ~ 5.5 permutations of each SMILES string.

Each SMILES string was used as 2D input for generation of 3D conformers using OMEGA, CORINA, RUBICON & CATALYST.

### Experiment A:

The space sampled by creating multiple conformers of the same SMILES string representation using OMEGA versus the space sampled by conformers generated from permuted SMILES strings the following was undertaken:

50 conformers of each of the 17 SMILES representations were produced using OMEGA. A total of 2406 conformers resulted.

Permsmi was utilised to permute each canonicalised string and single conformers produced for all until the total no. confs. ~2400.

Both sets of conformers were superposed by heavy atoms with the co-crystallised ligands and RMSD analysis carried out.

### Experiment B:

An unlimited number of conformers was generated for each SMILES string using OMEGA to ensure full space sampling, and compared with the Permuted set from above using PCA analysis.

## Results

**Experiment A:** Table below depicts that minimum RMSD values (the closest conformers to the crystal structure) are derived from a permuted string rather than the canonical one which would be normally used in a virtual screening process.

	1aq1	1d1q	1glq	1abe	1azm	1cbx	4dfr	1ebp	1hyt	1mrk	1p93	1phf	4phv	1poc	1i7i	1sij	1tpp
Can <sup>a</sup>	1.82	0.3	1.52	0.61	D	0.42	1.7	0.53	0.68	0.8	1.3	0.24	2.41	3.41	1.06	0.39	0.3
P1 <sup>b</sup>	1.26	0.52	1.73	0.68	D	0.68	0.76	0.74	0.45	0.95	1.3	0.24	1.72	2.1	1.18	0.39	0.3
P2	1.77	0.32	1.44	0.37	D	0.49	1.98	1.19	1	0.71	1.64	0.67	2.38	1.98	1.26	0.18	0.3
P3	0.86	0.29	1.72		D	0.68	1.96	0.61	0.45	1.06	1.26	0.24	2.28	1.97	1.04	0.46	0.3
P4	1.27		1.17				1.17	0.66		0.83	1.61		1.99	1.73	0.78	0.19	
P5	1.51		1.27				1.98				1.25		2.57	2.49	0.95		
P6	1.43		1.66				1.15						2.41	3.44	1.1		
P7							1.09						1.86				
P8													2.45				
P9													2.57				

**Experiment B:** From the canonical input, OMEGA produced 2239 conformers. The set originating from the permuted input comprised 2326 conformers. In Figure 2, the potential energy has been calculated and it is clear the span originating from the permuted strings input is larger. Moreover, members of this ensemble are energetically much closer to the energy calculated for bioactive conformation than those members of the ensemble originating from the canonical input.

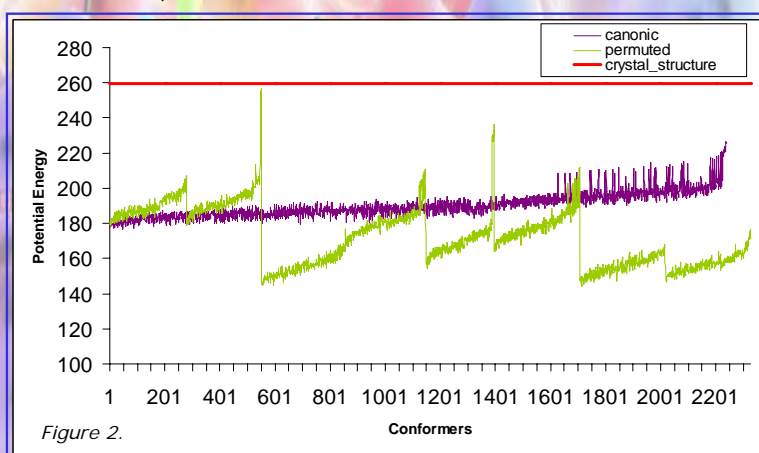


Figure 2.

Figure 3 shows the plotted result of the PCA. It depicts clearly how the conformational space explored by the conformers originating from permuted input is diverse and closely surrounds the bioactive space.

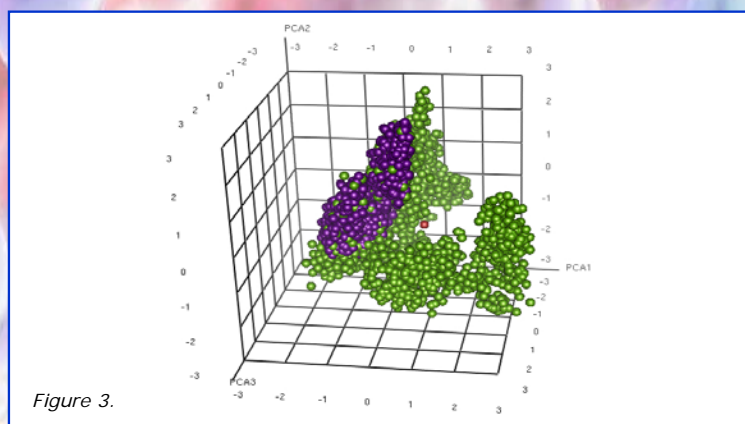


Figure 3.

In green the conformers generated from the 8 permutations of 1DFR keeping all the parameters in OMEGA as default (with the exception of  $-\text{max\_conf}$ , that was set to 310 to generate a comparable number of final conformers with the canonicalised set. In red is showed the PCA coordinates calculated for the crystal structure.

## Conclusions

Conformers produced from permuted SMILES strings exhibit a smaller RMSD value, when compared to the X-ray structure, than do conformers generated from the canonical SMILES string. Moreover superior conformational sampling arises from use of permuted input in term of potential energy and as evidenced from the PCA.

## References:

Permuting input for more effective sampling of 3D conformer space  
Carta G., Onnis V., Knox J. S. A., Fayne D., Lloyd G. D. *JCAM in press*