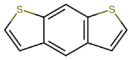
Latent Space Encoding of Molecular Crystal Structure

Wong King Ming Alex, University of Southampton
Email: kmw1g19@soton.ac.uk

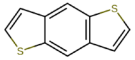
Skills 4 Scientists

Sponsored by: optibrium, dotmatics knowledge solutions

1. Introduction and motivation



Syn -BDT



Anti -BDT

- Computational methods for predicting crystal structures have been developed, with applications in screening molecules for polymorphism (the ability to crystallise in multiple crystal structures) and the design of materials.
- Current methods for crystal structure prediction sample the energy landscape for local minima by generating and lattice energy minimising trial crystal structures. We are investigating the possibility of deep generative methods for crystal structures.
- As an initial investigation, we are interesting in understanding the intrinsic dimensionality of the crystal structure landscape of small organic molecules, and how the dimensionality varies between space groups and with the energy range of crystal structures included in the analysis.

3. Methods/Data origin:

Data

- Crystal structure prediction was performed for two molecules, syn- and anti-benzodithiphene (BDT), using quasi-random sampling of the energy surface using the Global Lattice Energy Explorer software [1]. Crystal structures were generated in the 10 most frequently observed space groups for organic molecules: P21/c, P212121, P-1, P21, Pbca, C2/c, Pna21, Cc, Pca21 and C2.
- We have analysed the sets of crystal structures within 10 and 20 kJ/mol of the global energy minimum.

Descriptors

- As descriptors of the crystal structures, we use the atom centered symmetry functions proposed by Behler and Parrinello [2]. We use the modified version developed for the ANI-1 force field [3]. These describe the atomic environment with a series of radial and angular functions built up from the distribution of neighbouring atoms.

Analysis

- We have analysed the data using principal component analysis (PCA).
- We perform PCA separately on the symmetry function description of crystal structures in each space group, and on the entire set of low energy crystal structures.

4. Results:

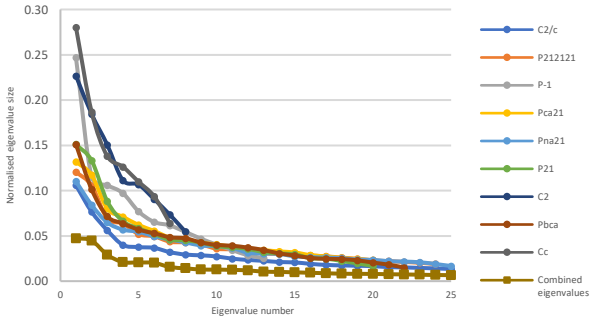


Fig 1: Plot of the eigenvalues for pca in lowest 10 kJmol-1 antibdtx1 in 10 most common space group (up to 25 dimensions)

	Total number of eigenvalues
10 kJ/mol anti-BDT	315
20 kJ/mol anti-BDT	1301
10 kJ/mol syn-BDT	267
20 kJ/mol syn-BDT	1349

Table 1: Table of number of eigenvalues in different energy windows

Number of eigenvalues above 1% between individual space group	P21/c	C2/c	P212121	P-1	Pca21	Pna21	P21	C2	Pbca	Cc	Combined
10 kJ/mol anti-BDT	26	35	24	13	20	27	20	8	22	7	15
20 kJ/mol anti-BDT	11	12	42	28	40	25	37	31	25	33	6
10 kJ/mol syn-BDT	29	32	19	12	16	26	11	16	21	6	17
20 kJ/mol syn-BDT	11	12	44	34	38	26	32	31	25	32	6

Table 2: Table of number of eigenvalues above 1% for each space group

5. Conclusion

- Different space groups have different intrinsic dimensionalities, although the dimensionality we get from counting structural degrees of freedom is almost the same
- There is a different in number of eigenvalues greater than 0.01 between space group.
- Similar relationship obtained between molecules when changing from 10 to 20 kJ/mol

6. Further works

- further analysis of PCA results and how the principal components relate to packing features in the crystal structures
- Attempts to build autoencoder based on the atom-centered symmetry functions, with the bottleneck dimension of the bottleneck chosen based on the dimensionality analysis.

7. Acknowledgement:

- Thanks to Prof Graeme Day, Rebecca J Clements and the research Group, AI3SD for funding and training and Iridis HPC

Reference:

[1] Convergence Properties of Crystal Structure Prediction by Quasi-Random Sampling
David H. Case, Josh E. Campbell, Peter J. Bygrave, and Graeme M. Day, J. Chem. Theory Comput. 2016, 12, 2, 910–924

*[2] Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Jörg Behler and Michele Parrinello, Phys. Rev. Lett. **98**, 146401*

*[3] ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, J. S. Smith, O. Isayev and A. E. Roitberg, **Chem. Sci.**, 2017,**8**, 3192-3203*

4:36
4:37

AI³ Science Network+

PSDS PHYSICAL SCIENCES DATA-SCIENCE SERVICE

RSC INTEREST GROUP CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP