1. Introduction and motivation

- 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 interested 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.

4. Results:

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.

4.36

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

<table>
<thead>
<tr>
<th>Energy range</th>
<th>Total number of eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 kJ/mol anti-BDT</td>
<td>315</td>
</tr>
<tr>
<td>20 kJ/mol anti-BDT</td>
<td>1301</td>
</tr>
<tr>
<td>10 kJ/mol syn-BDT</td>
<td>267</td>
</tr>
<tr>
<td>20 kJ/mol syn-BDT</td>
<td>1349</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Individual space group</th>
<th>P21/c</th>
<th>C2/c</th>
<th>P212121</th>
<th>P-1</th>
<th>Pbca</th>
<th>Pna21</th>
<th>Pca21</th>
<th>P21</th>
<th>C2</th>
<th>Cc</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 kJ/mol anti-BDT</td>
<td>11</td>
<td>12</td>
<td>42</td>
<td>28</td>
<td>40</td>
<td>25</td>
<td>37</td>
<td>31</td>
<td>25</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td>20 kJ/mol syn-BDT</td>
<td>11</td>
<td>12</td>
<td>44</td>
<td>34</td>
<td>38</td>
<td>26</td>
<td>32</td>
<td>31</td>
<td>25</td>
<td>32</td>
<td>6</td>
</tr>
</tbody>
</table>

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

References: