

Crystal Structure Prediction of Energetic Materials and a Twisted Arene with Genarris and GAtor

Immanuel Bier¹, Dana O'Connor¹, Yun Ting Hsieh¹, Wen Wen²,
Anna M. Hiszpanski³, T. Yong-Jing Han³, and Noa Marom^{1,2,4}

¹Department of Materials Science and Engineering, Carnegie
Mellon University, Pittsburgh, PA, 15213, USA.

²Department of Chemistry, Carnegie Mellon University,
Pittsburgh, PA, 15213, USA.

³Materials Science Division, Lawrence Livermore National
Laboratory, Livermore, CA, 94550, USA.

⁴Department of Physics, Carnegie Mellon University, Pittsburgh,
PA, 15213, USA.

Email address: nmarom@andrew.cmu.edu

1 TATB

1.1 Genarris

For three values of s_r , 0.65, 0.75, and 0.85, the Robust workflow for Genarris was used to generate an initial pool of structures to seed GAator. For each step of the Robust workflow, the space group and volume distribution are shown for the three s_r values of 0.65, 0.75, and 0.85 below in Figure S1, Figure S2, and Figure S3, respectively. The experimental space group, 2, is indicated by the green arrow in the space group distribution plots. The experimental volume, 425 Å³, is shown in red. Clearly, Genarris is able to generate structures in a diversity of space groups. However, as downselection proceeds, lower volume structures are filtered out.

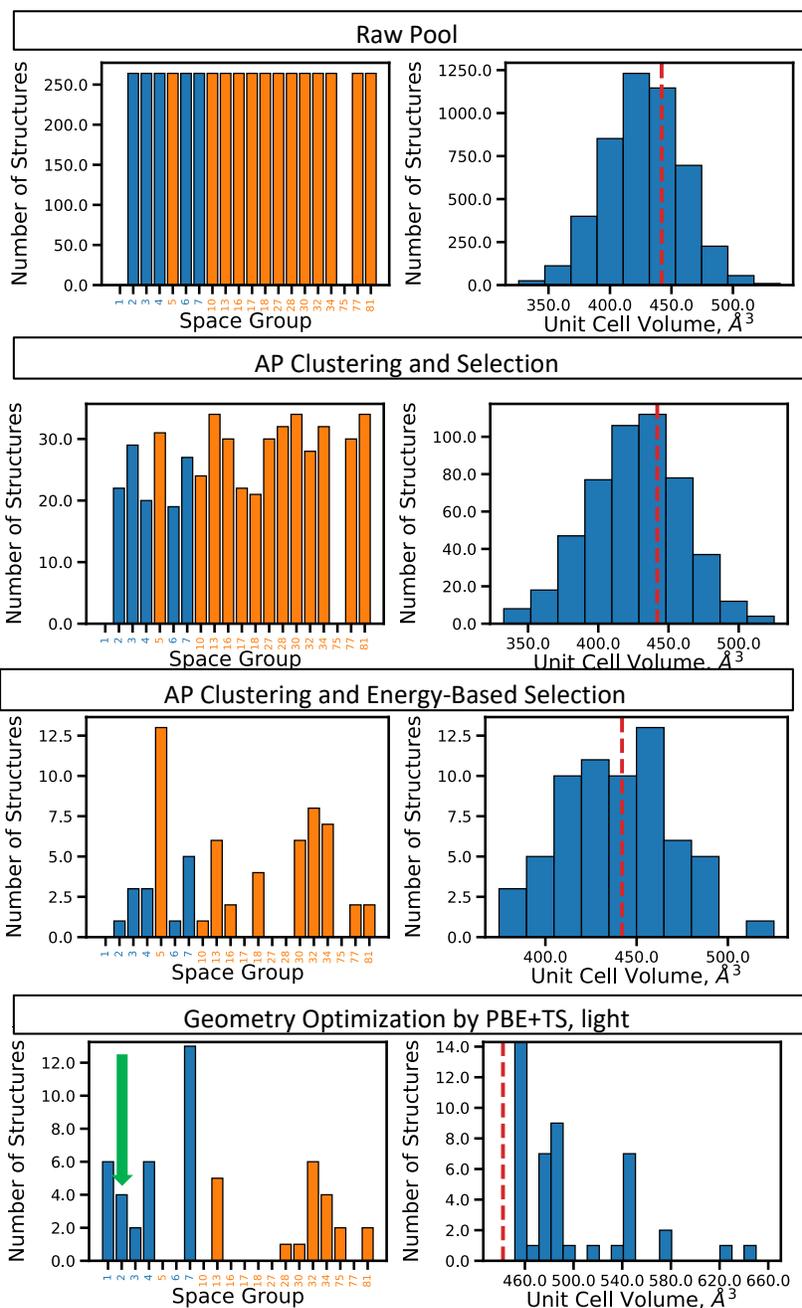


Figure S1: Space group distribution (left) and volume histograms (right) for each stage of the Genarris workflow for TATB with s_r of 0.65. The green arrow on the space group distribution plots indicates the experimental space group. Special positions are shown in orange and general positions are shown in blue. The red line on the volume histograms indicates the experimental volume.

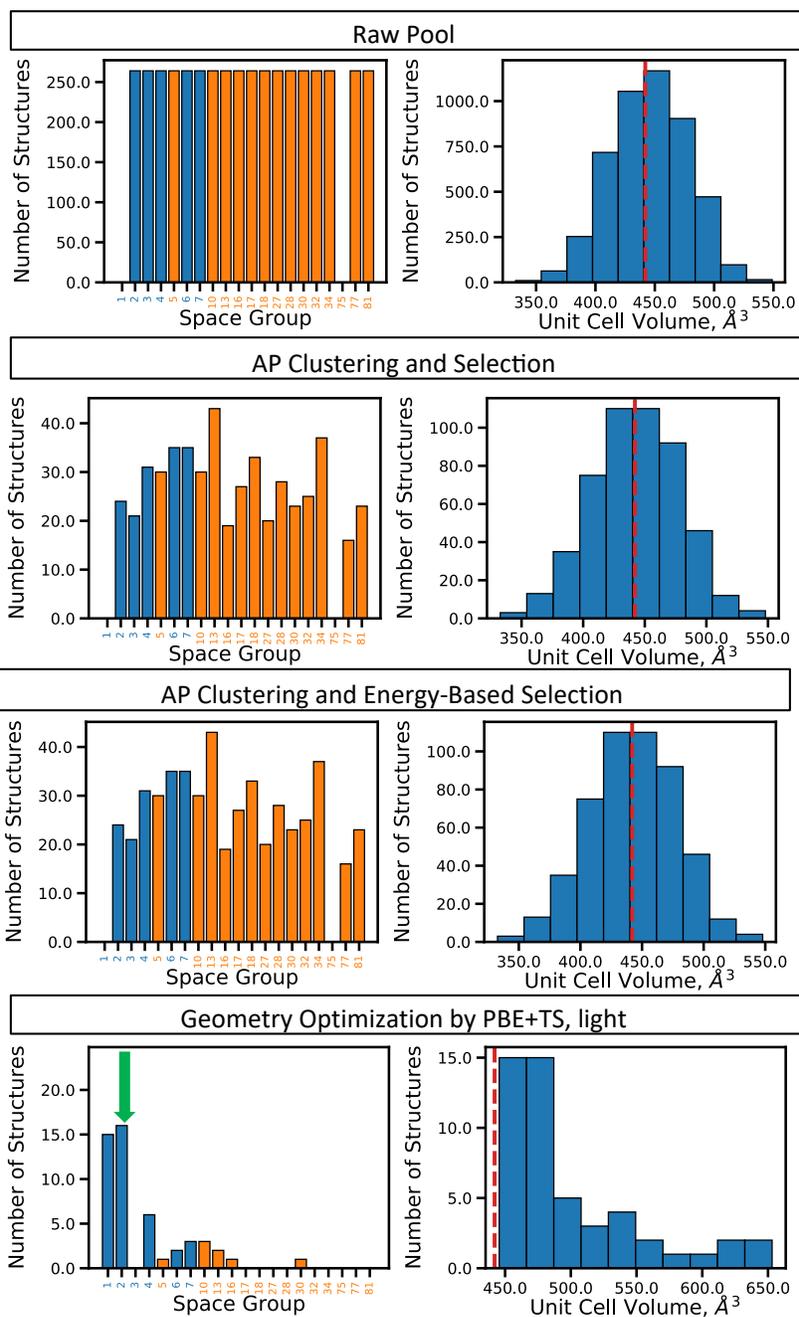


Figure S2: Space group distribution (left) and volume histograms (right) for each stage of the Genarris workflow for TATB with s_r of 0.75. The green arrow on the space group distribution plots indicates the experimental space group. Special positions are shown in orange and general positions are shown in blue. The red line on the volume histograms indicates the experimental volume.

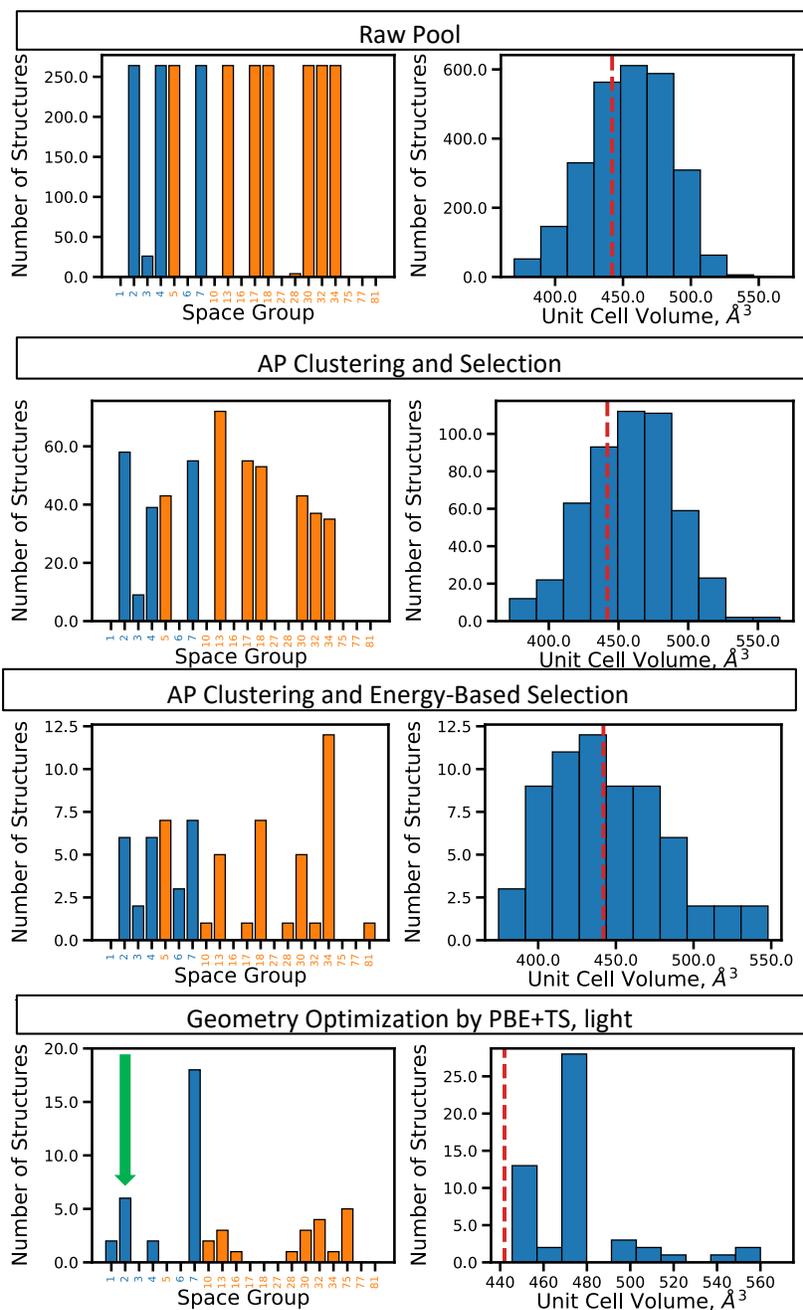


Figure S3: Space group distribution (left) and volume histograms (right) for each stage of the Genarris workflow for TATB with s_r of 0.85. The green arrow on the space group distribution plots indicates the experimental space group. Special positions are shown in orange and general positions are shown in blue. The red line on the volume histograms indicates the experimental volume.

1.2 GA Convergence

As a GA is not guaranteed to find the global minimum, convergence may be defined as when the GA is no longer producing unique, low energy structures over a large number of iterations. The average energy is used as a measurement of convergence; when the average energy either begins to increase or stagnates, the GA is said to have converged. While normally GAtor is run for about 350 iterations, TATB appears to have converged in a shorter amount of time. As can be seen in Figure S4, the average energy for all four GA runs began to increase around iteration 175. The energy fitness function at crossover probability 25% had the lowest average energy at a little less than 20 kJ/mol. The other three GA runs all converged to about the same average energy, between 20-25 kJ/mol. The energy fitness function at crossover probability 25% may have a lower average energy because the experimental basin, which is sampled heavily by the energy fitness function, is the lower energy basin. The three other GA runs may have sampled more from the higher energy basin.

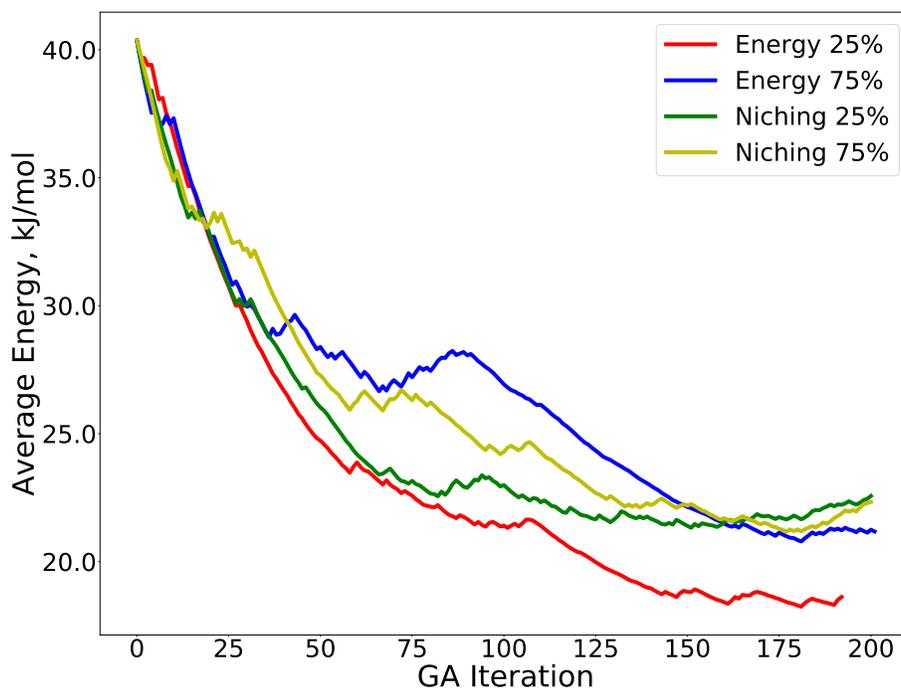


Figure S4: The average energy for the common population of each GA run for TATB. The energy fitness function at crossover probabilities 25% and 75% are shown in red and blue, respectively. The niching fitness function at crossover probabilities 25% and 75% are shown in green and yellow, respectively.

2 DATB

2.1 Genarris

Below are the space group and volume distribution for each step of Genarris' Robust workflow for DATB. Three values of s_r are used, 0.65, 0.75, and 0.85, which are shown in Figure S5, S6, and S7, respectively. The experimental space group is Pc (7) and the experimental volume is 436 \AA^3 .

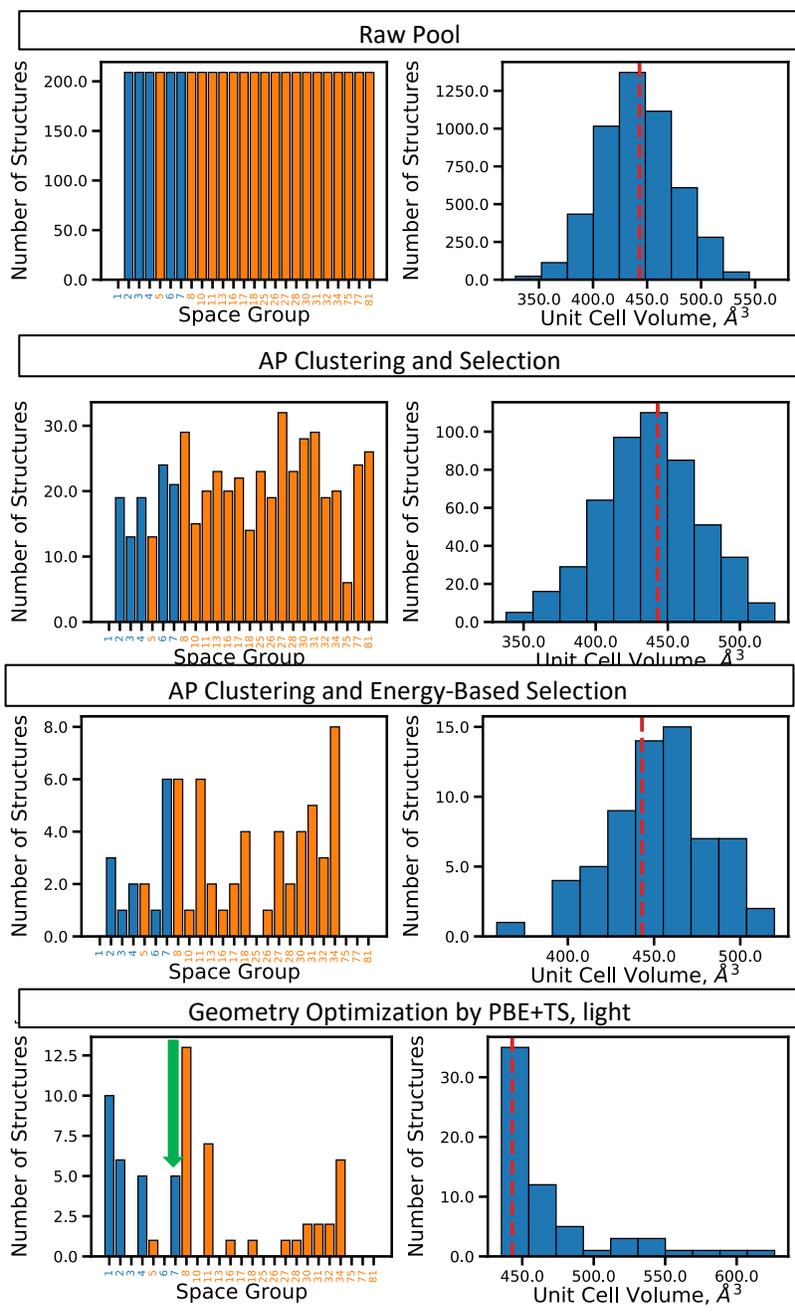


Figure S5: Space group distribution (left) and volume histograms (right) for each stage of the Genarris workflow for DATB with s_r of 0.65. The green arrow on the space group distribution plots indicates the experimental space group. Special positions are shown in orange and general positions are shown in blue. The red line on the volume histograms indicates the experimental volume.

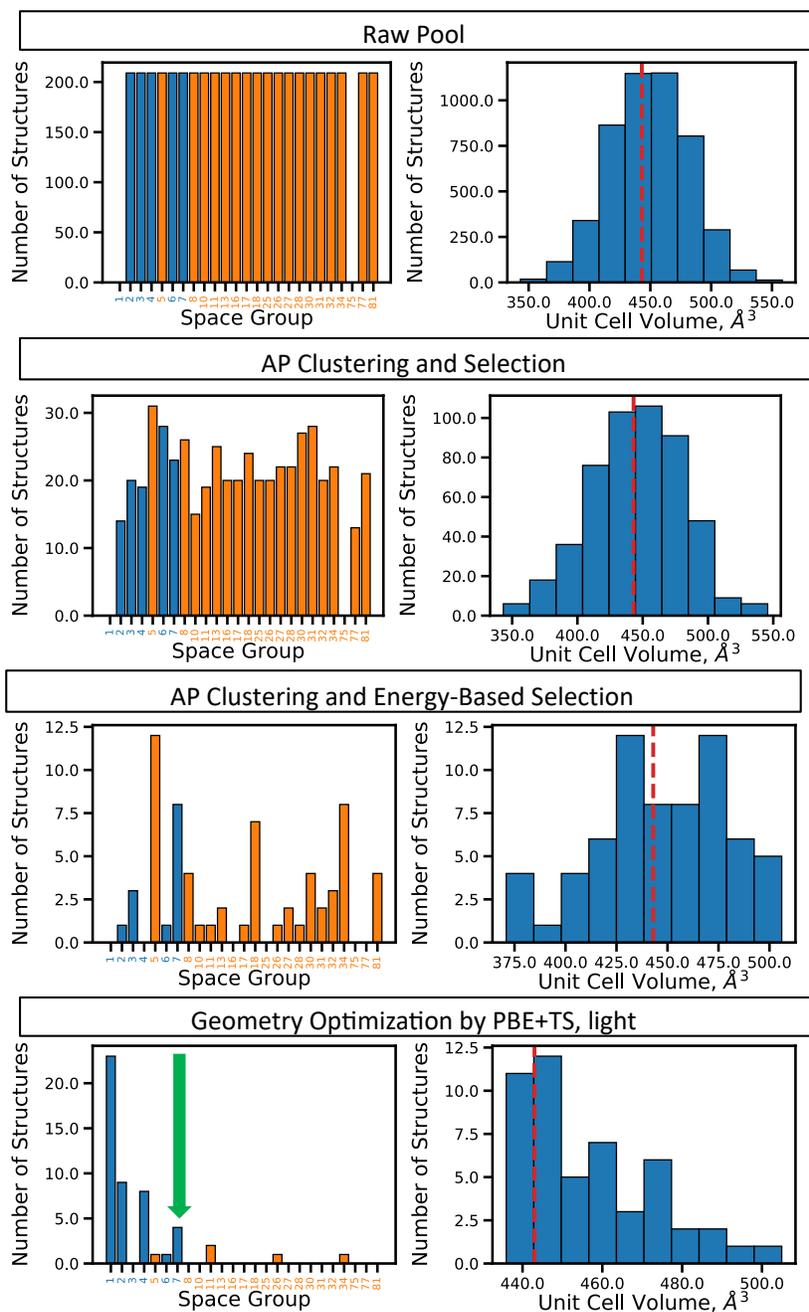


Figure S6: Space group distribution (left) and volume histograms (right) for each stage of the Genarris workflow for DATB with s_r of 0.75. The green arrow on the space group distribution plots indicates the experimental space group. Special positions are shown in orange and general positions are shown in blue. The red line on the volume histograms indicates the experimental volume.

2.2 GA Convergence

As stated previously, the convergence of the GA can be determined by examining the average energy of the common population. For DATB, all four GA runs converged in about 300-350 iterations, as seen below in Figure S8. Unlike TATB, the average energy only began to increase for the energy fitness function at crossover probability 25% when the GA was terminated. For the other three GA runs, the average energy appears to have converged when the GA was terminated. However, all four GA runs converged to about the same average energy, between 10-12 kJ/mol. Additionally, the niching fitness function at crossover probability 75% appears to have sampled heavily from higher energy portions of the configuration space. This may explain why this GA run took much longer to generate the experimental structure.

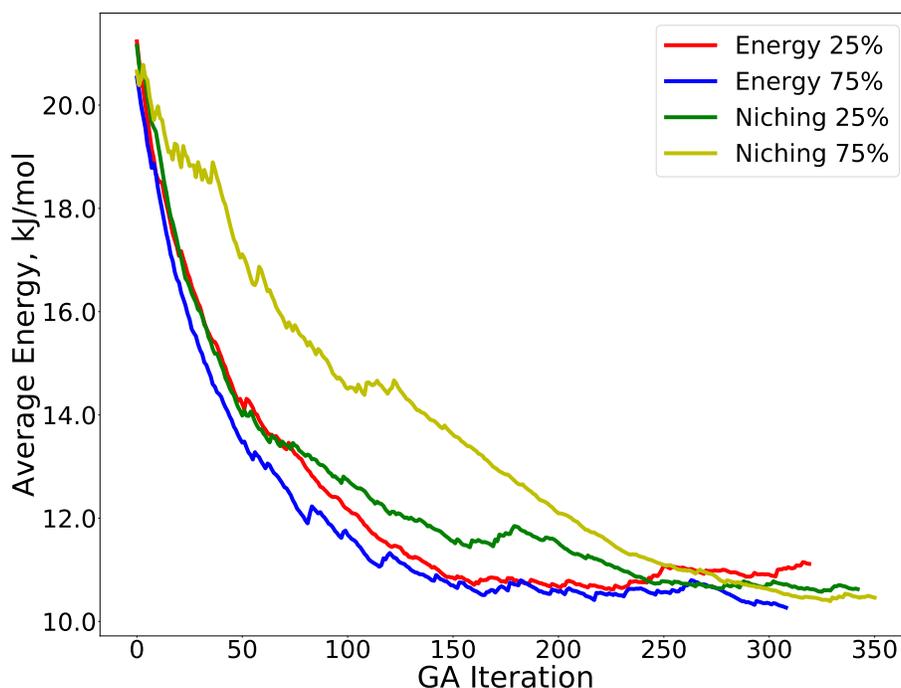


Figure S8: The average energy for the common population of each GA run for DATB. The energy fitness function at crossover probabilities 25% and 75% are shown in red and blue, respectively. The niching fitness function at crossover probabilities 25% and 75% are shown in green and yellow, respectively.

3 4,5-dimethylphenanthrene

3.1 Genarris

For this target, Genarris' Robust workflow was used with two values of s_r . The space group and volume distributions for both values of s_r , 0.70 and 0.75, are shown below in Figure S9 and S10, respectively. The experimental space group is $P2_1(4)$ and the experimental volume is 555 \AA^3 .

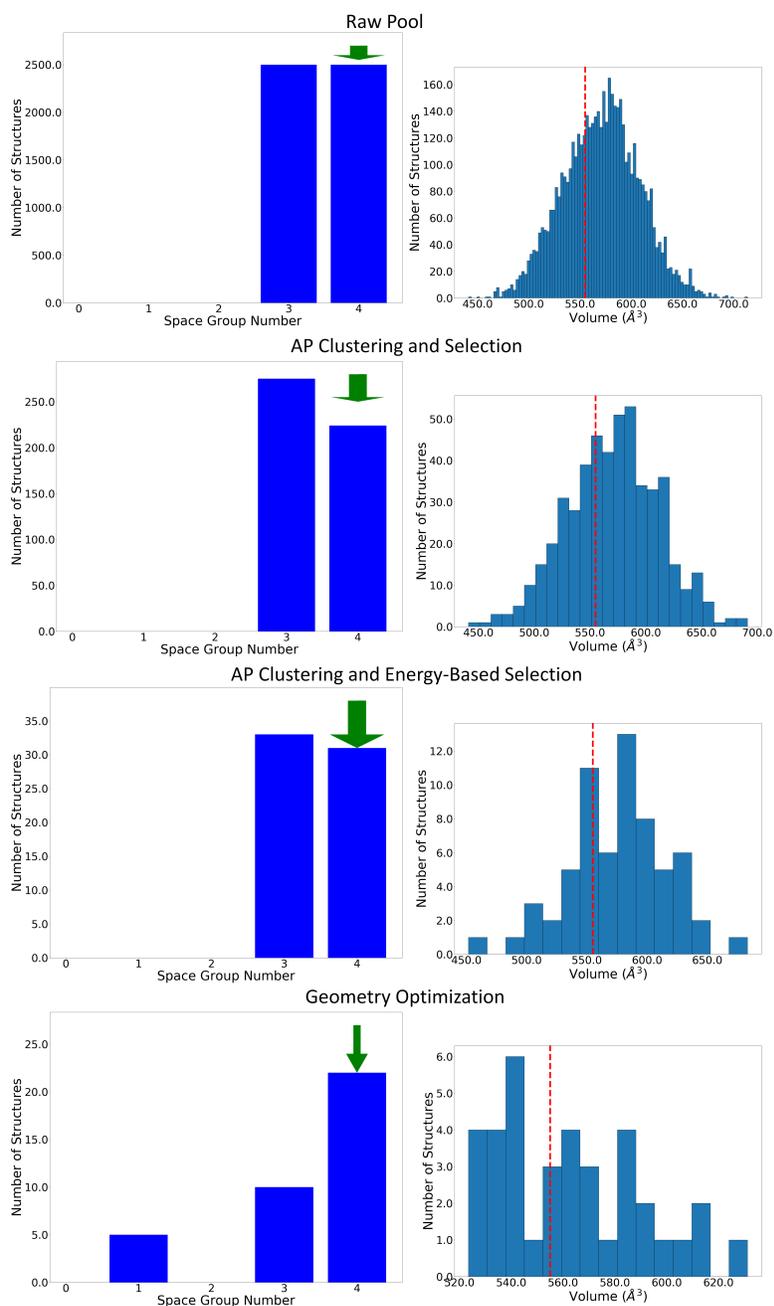


Figure S9: Space group distribution (left) and volume histograms (right) for each stage of the Genarris workflow for 4,5-dimethylphenanthrene with s_r of 0.70. The green arrow on the space group distribution plots indicates the experimental space group. Special positions are shown in orange and general positions are shown in blue. The red line on the volume histograms indicates the experimental volume.

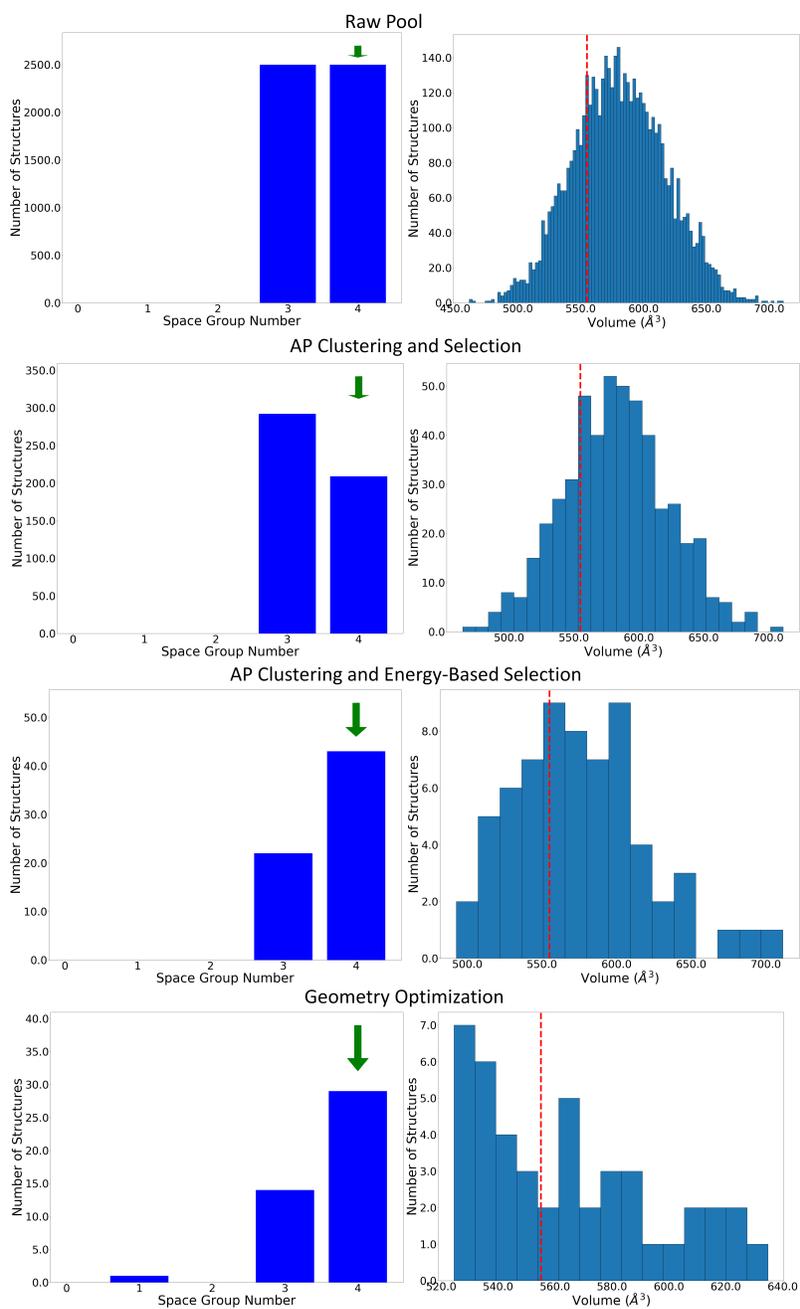


Figure S10: Space group distribution (left) and volume histograms (right) for each stage of the Genarris workflow for 4,5-dimethylphenanthrene with s_r of 0.75. The green arrow on the space group distribution plots indicates the experimental space group. Special positions are shown in orange and general positions are shown in blue. The red line on the volume histograms indicates the experimental volume.

3.2 GA Convergence

Again, GA convergence can be measured by the average energy of the common population for each GA run. As seen in Figure S11 below, three of the four GA runs converged to the same energy, between 26-28 kJ/mol. The energy-based fitness function at crossover probability 25% had the lowest average energy, around 24 kJ/mol. However, all four GA runs appear to have converged around the same time, around iteration 200.

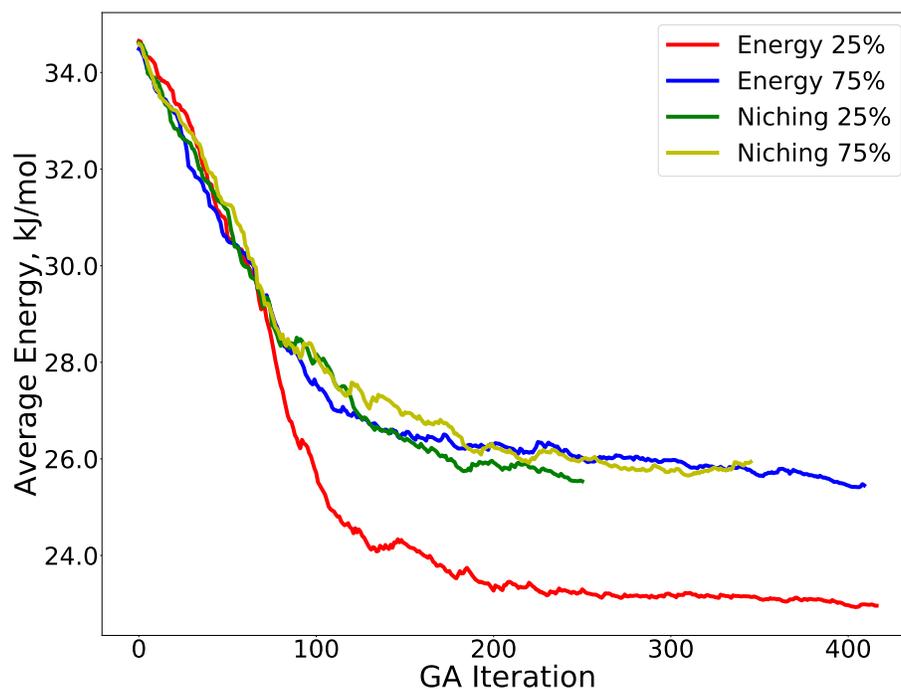


Figure S11: The average energy for the common population of each GA run for 4,5-dimethylphenanthrene. The energy fitness function at crossover probabilities 25% and 75% are shown in red and blue, respectively. The niching fitness function at crossover probabilities 25% and 75% are shown in green and yellow, respectively.

3.3 Energy-Density

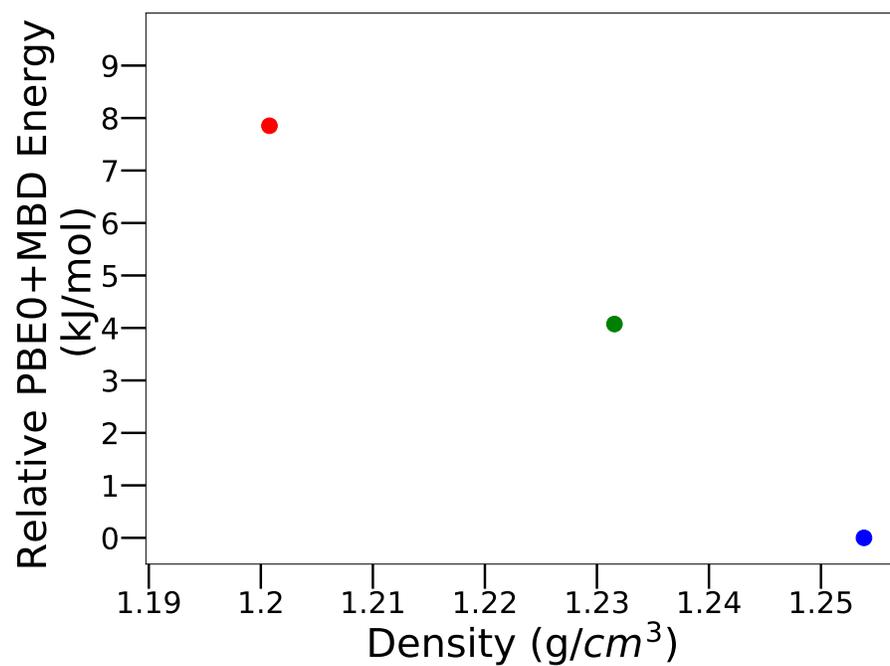


Figure S12: PBE0+MBD relative energy as a function of density for 4,5-dimethylphenanthrene structures produced by GAtor. Colored markers correspond to structures shown in the same colors in Figure 20

4 Principal Component Analysis

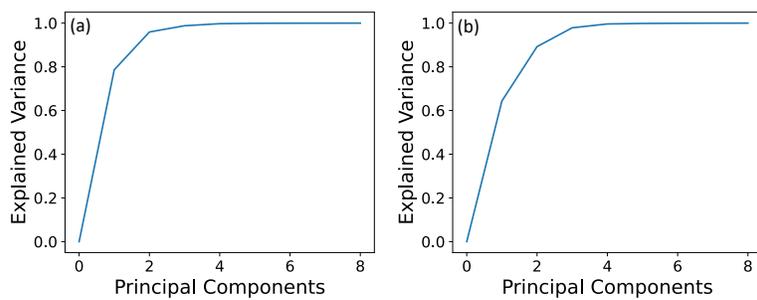


Figure S13: Explained variance as a function of the number of principal components for (a) TATB and (b) DATB. The explained variance of the first two components for the TATB crystal landscape are 0.78 and 0.96 respectively and the explained variance of the first two components for the DATB crystal landscape are 0.64 and 0.89 respectively.

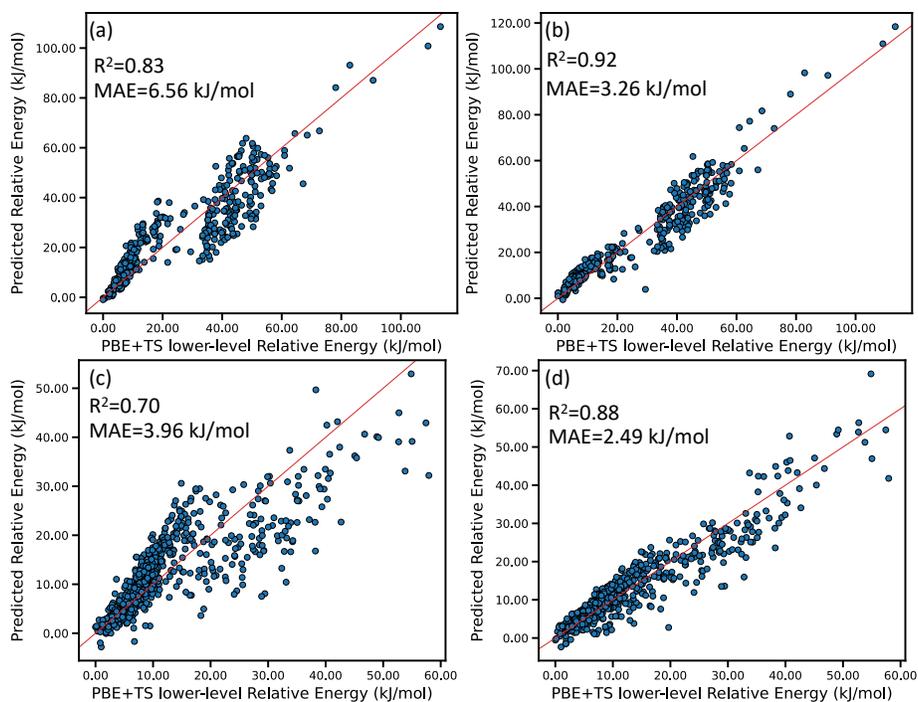


Figure S14: Correlation of predictions of the relatively energy using a linear model fit to (a)/(c) the first principal component and (b)/(d) both principal components for TATB and DATB respectively. Shown on each graph is a red line of $x = y$ and the R^2 value and mean-absolute error (MAE) of each model.

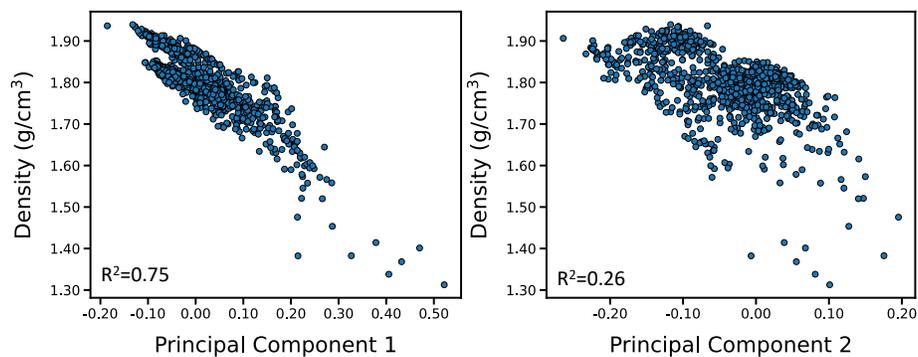


Figure S15: Correlation of the density with respect to the first and second principal components of each DATB and TATB structure. It's found that principal component 1 correlates well with the density of the structures whereas principal component 2 provides only weak correlation.

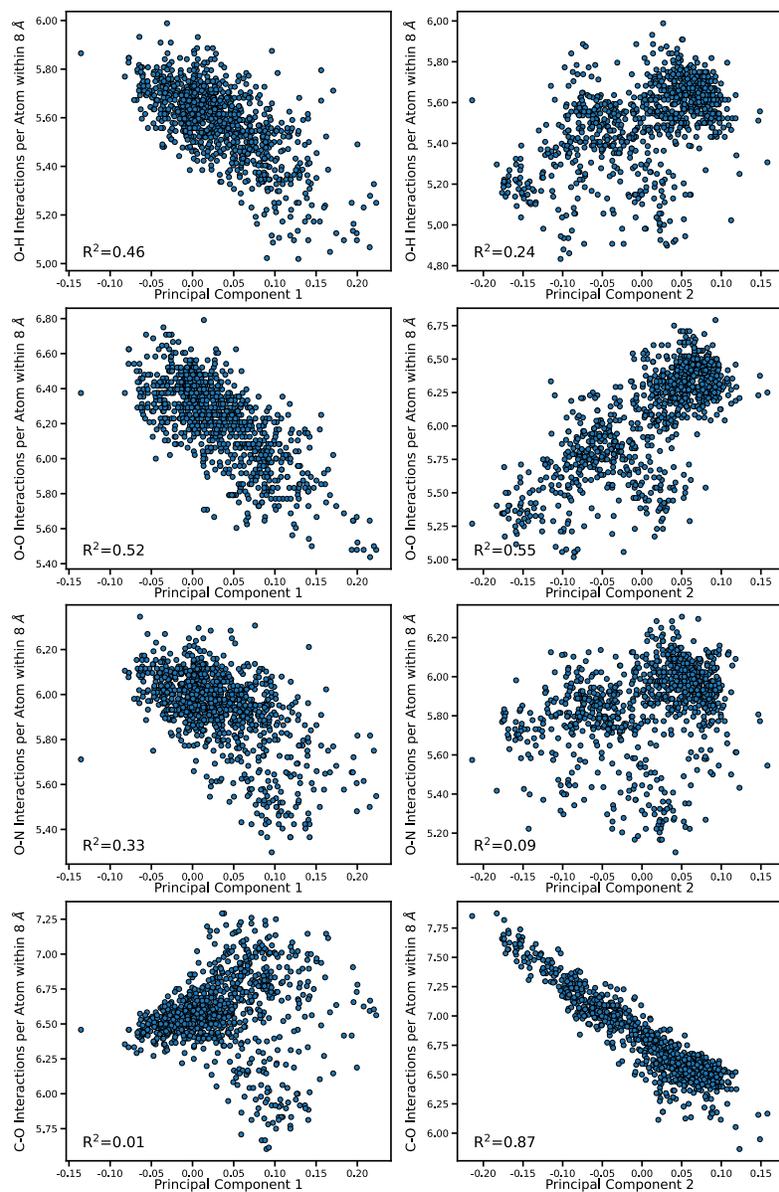


Figure S16: Correlation of the number of interactions per atom in the unit cell within the RSF cutoff radius for four different types of interactions with respect to the first and second principal component of each structure.

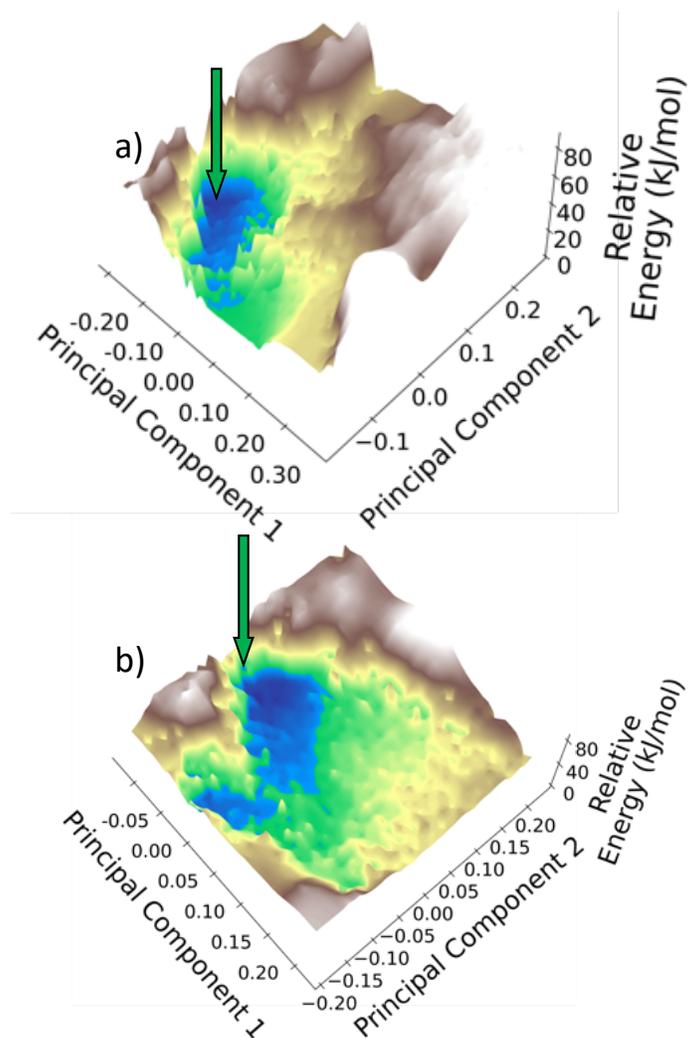


Figure S17: Full computed representation of the potential energy surfaces for TATB, in (a), and DATB in (b) using PCA of the RSF descriptors. The full PES was constructed by using all intermediate structures from DFT relaxations to construct approximate local minima and energy barriers. The location of the lowest energy, experimentally determined structure is indicated by a green arrow in each graph.