

Supporting Information:

Evolutionary Niching in the GAtor Genetic Algorithm for Molecular Crystal Structure Prediction

Farren Curtis,^{†,‡} Timothy Rose,[‡] and Noa Marom^{*,‡,†,¶}

[†]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

[‡]*Department of Materials Science and Engineering, Carnegie Mellon University,
Pittsburgh, PA 15213, USA*

[¶]*Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

E-mail: nmarom@andrew.cmu.edu

S.1 Cluster Histograms Generated from Runs that Used the Diverse Initial Pool

Fig. S.1. presents an analysis of the final clusters generated by the GA runs that were started from the diverse initial pool and used the lattice parameter descriptor (panel a), the RDF descriptor (panel b), and the RCD descriptor (panel c). For each run that utilized evolutionary niching, AP was used to cluster the final population of structures with respect to the same descriptor used in the GA. The resulting distributions are shown in blue. AP was then used to predict the cluster labels of the structures generated in the energy-based control run, shown in red. The distribution of initial pool structures contained in each cluster is shown in grey. The cluster assigned to the experimental structure is indicated by a green arrow. Additionally, the average and standard deviation of the relative energies of the structures in each cluster is plotted in orange. For all three descriptors, the use of a cluster-based fitness function suppressed the sampling of clusters that are over-sampled in the control run.

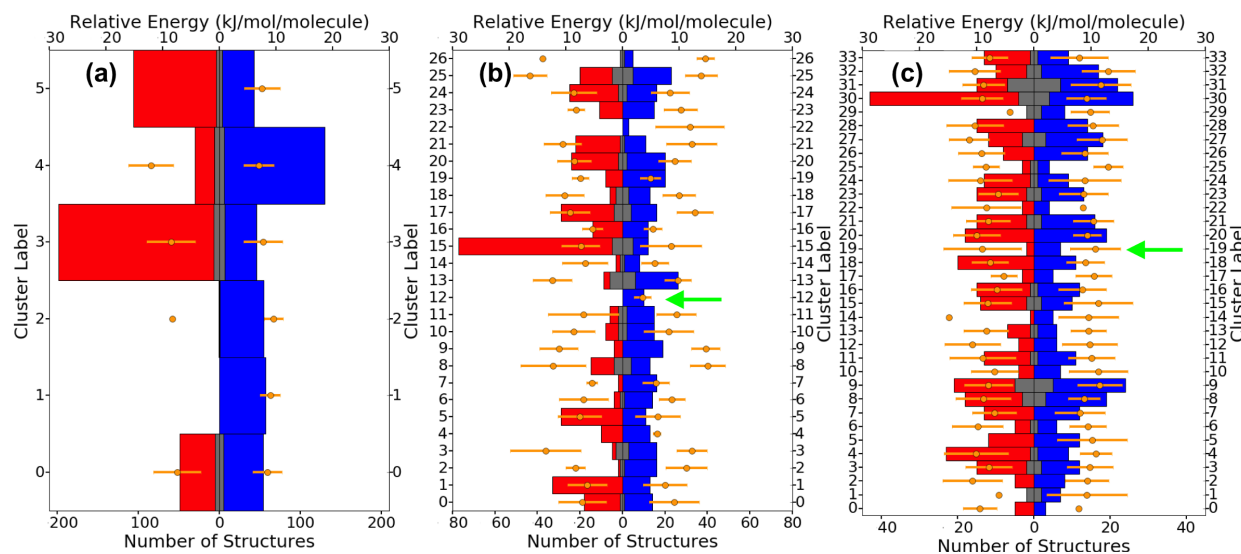


Figure S.1: The distribution of clusters produced by different GA runs initialized from the diverse initial pool. Histograms corresponding to runs that used evolutionary niching are shown in blue and histograms corresponding to the control run are shown in red. The structures are clustered with respect to (a) L, (b) RDF, and (c) RCD. In all plots the average relative energy and standard deviation for each cluster are shown in orange. For the runs that generated the experimental structure, the cluster assigned to the experimental structure is indicated by a green arrow.