Electronic Supplementary Material (ESI) for Digital Discovery.

Electronic Supplementary Information: Illuminating the property space in crystal structure prediction using Quality-Diversity algorithms

1 Hyperparameter Selection

The hyperparameters used in this work are based on previous work in the field as summarised below:

Parameter	Value	Reason	Reference
Number of atoms	24	The number used in the references is 48, because increasing the number of atoms increases the difficulty to find the ground state. However to balance the trade-off between difficulty and time required for an experiment a smaller size of 24 atoms was selected for this investigation.	[1–3]
Unit cell parameter limits	unit cell vectors: $2-60$ Å; angles: $0-\pi$	These were set to be less restrictive than the already considered 'loose' parameters $0-40$Å, and $20-120$, degrees used by Lonie and Zurek.	[3]
Maximum angle between unit cell vector and plane created by other 2 vectors	[20, 160] , [60, 120], [20, 160]	As above, this was set to be less restrictive than the example provided in the reference implementation of a genetic algorithm.	[4,5]
Unit cell volume	450 Å^3.	This was estimated based on the 483 Å^3 used by Lonie and Zurek in the reference.	[3]
Scaling factor between atomic radii	0.4	This parameter defines how closely atoms can be positioned in the unit cell. Set to the values used in references.	. [1,2]
Crossover operators and their probabilities	Operators: Strain and Permutation mutations with probabilities: 50%, 50%	The initial set up included crossover and soft mutation operators as set by Lykahov et al in the reference. These were not included because they did not guarantee an individual to be outputted and soft mutation also requires realistic individuals which could not be guaranteed in this work. Therefore for computational efficiency these were not included.	[2]
Unit cell splitting factors	2, 4	This is a parameter used to determine transnational symmetry when using the random structure generator. The values were set using the reference	[2]
Starting Population Size	20	The size used by the Lyakhov et al. is 10, but since our objective is also around diversity a larger starting point was used. This and larger population sizes were also tested as reported in table S2. as part of the MAP-Elites parameters search	[2]

Table S1: Hyperparameter selection for TiO_2 mapped to past work.

Additionally a brief summary of the hyperparameters set from a computational perspective are summarised below:

Parameter	Value Set	Values Tested	Reason
Number of niches	200	200, 500	200 niches was selected to increase the chances of competition between structures. The dynamics of the experiment were consistent across the three experiments so 200 was selected as a good trade off between speed and diversity of possible solutions.
Proportion of niches to be filled	0.1	0.1, 0.2	The higher proportion of niches didn't improve the average fitness obtained nor the number of reference matches found, so the lower threshold was selected to move to structure mutations quicker.
Number of random structures to initialise	20	20, 40, 80	20 was set as it limited the time spent on generating new structures, and if required would be called multiple times to filled the proportion of niches desired so on its own this parameter had limited impact.
Batch size per generation	100	20, 50, 100	This parameter defines how many individuals will be mutated per generation. Setting this to 100 meant that some individuals were mutated multiple times. This accelerated finding of reference matches.

Table S2: Default hyperparameter selection.

2 Selection of force threshold value

The maximum force acting on each of the reference structures was computed. This includes the forces acting on each atom as well as the stresses on the unit cell, as implemented and used for relaxation in `CHGNEt`. The results are presented in the histogram in Figure S1.

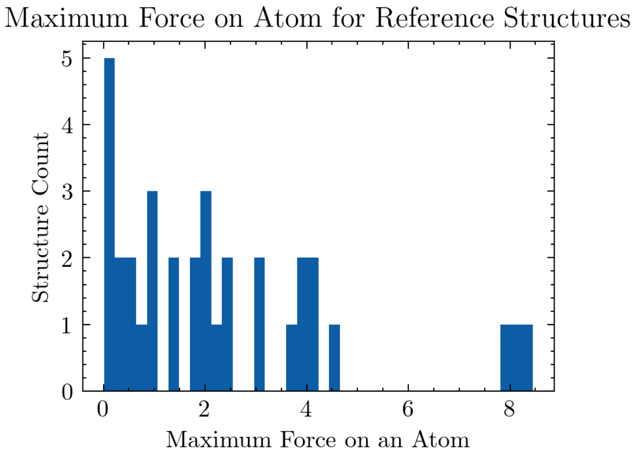
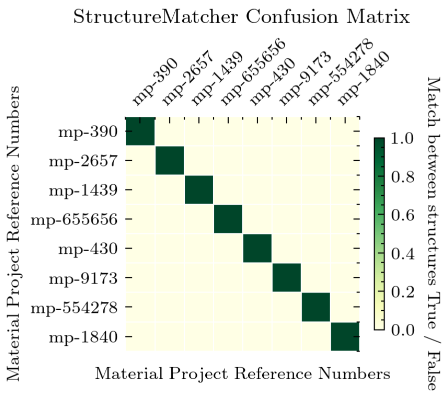


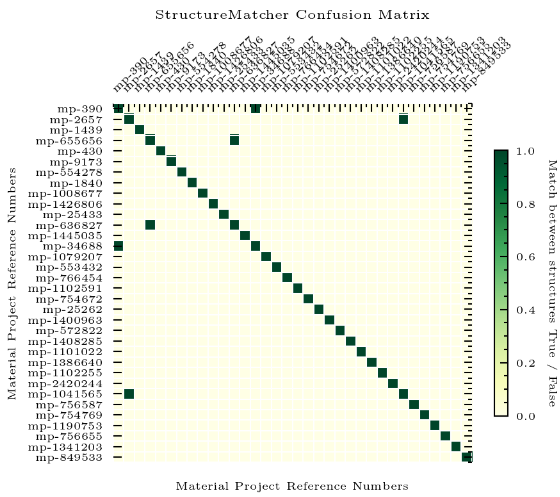
Figure S1: Histogram of the maximum force acting on each of the reference structures of TiO_2 computed using `CHGNet`.[6]

3 Identifying equivalent structures

Within the tolerances set in this work we observed that some structures would be equivalent to each other. This is demonstrated within Figure S2 below. There the confusion matrix was constructed by comparing all reference structures to each other. Structures considered equivalent to each other are marked with green. In Figure S2b we can observe that equivalent pairs are: *mp-390* and *mp-34688*, *mp-2657* and *1041565*.



(a) Matches between experimentally observed reference structures.



(b) Matches between all reference structures.

Figure S2: Correlation plot of matches between reference structures as generated using `StructureMatcher` from `pymatgen`.[7]

4 Sample results for C, SiC, SiO_2

Below in Figure S3 we demonstrate sample archives for C, SiC and SiO_2. We can observe similar trend as with TiO_2; the archives largely developed in the expected areas of the features space. With the exception of SiC the archives demonstrate that a wide set of solutions is demonstrated.

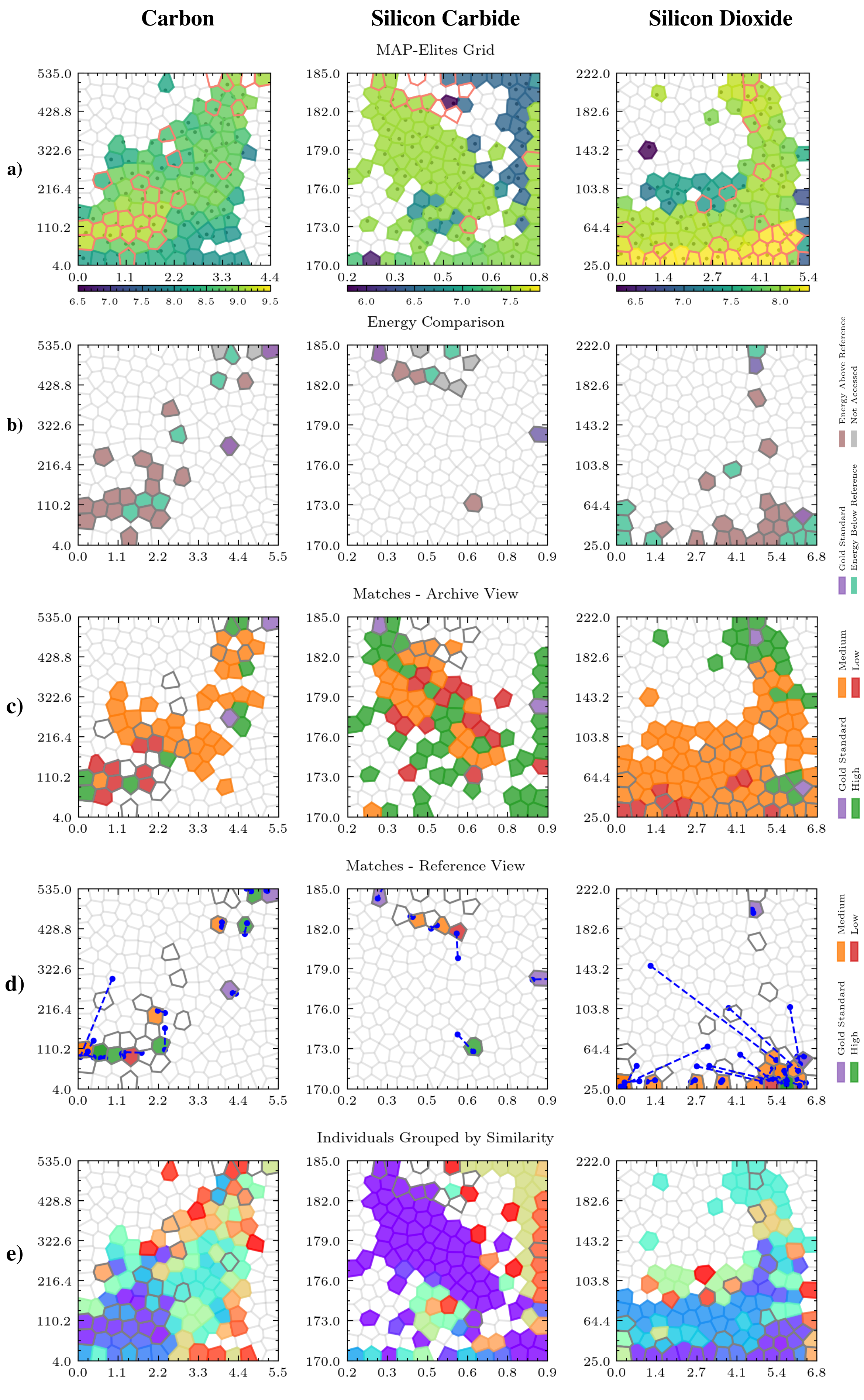


Figure S3: Sample results for C, SiC and SiO_2 experiments.

References

[1] Z. Falls, P. Avery, X. Wang, K. P. Hilleke and E. Zurek, The XtalOpt Evolutionary Algorithm for Crystal Structure Prediction, *The Journal of Physical Chemistry C*, 2021, **125**, 1601–1620.

[2] A. O. Lyakhov, A. R. Oganov, H. T. Stokes and Q. Zhu, New developments in evolutionary structure prediction algorithm USPEX, *Computer Physics Communications*, 2013, **184**, 1172–1182.

[3] D. C. Lonie and E. Zurek, XtalOpt: An open-source evolutionary algorithm for crystal structure prediction, *Computer Physics Communications*, 2011, **182**, 372–387.

[4] M. Van Den Bossche, H. Grönbeck and B. Hammer, Tight-Binding Approximation-Enhanced Global Optimization, *Journal of Chemical Theory and Computation*, 2018, **14**, 2797–2807.

[5] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, The atomic simulation environment—a Python library for working with atoms, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.

[6] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, *Nature Machine Intelligence*, 2023, **5**, 1031–1041.

[7] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, *Computational Materials Science*, 2013, **68**, 314–319.