Supporting Information: Accelerating Structure Relaxation in Chemically Disordered Materials with a Chemistry-driven Model

Liying An[1,2,3], Huan Ma[1,2,3], Jinjia Liu[3*], Wenping Guo[3*], Xiaodong Wen[1,3,4,*]

[1]*State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan 030001, China*

[2]*University of Chinese Academy of Sciences, Beijing 100049, China*

[3]*National Energy Center for Coal to Liquids, Synfuels China Co., Ltd., Beijing 101400, China*

[4]*Beijing Advanced Innovation Center for Materials Genome Engineering, Industry−University Cooperation Base between Beijing Information S&T University and Synfuels China Co., Ltd., Beijing 100101, China*

1. Supplementing the main text with figures and equations

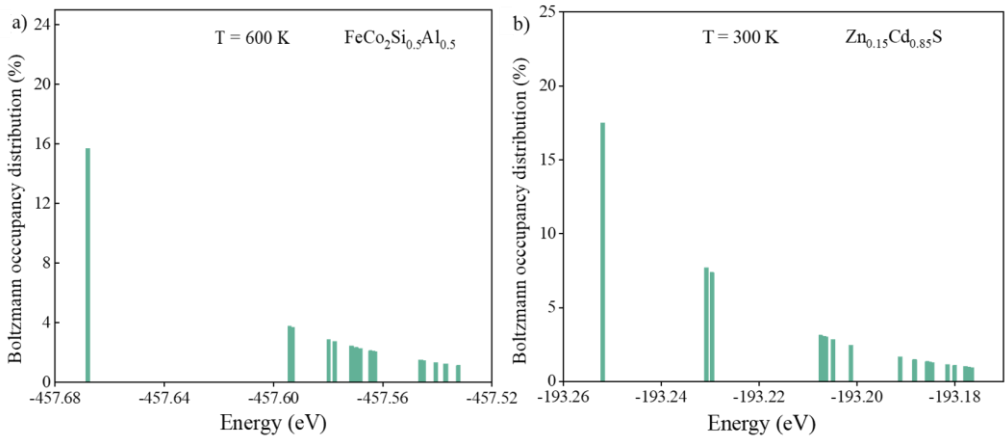

Figure S1 | Boltzmann distribution of low-energy structures at specific temperatures. a for $FeCoSi_{0.5}Al_{0.5}$ system and **b** for $Zn_{0.15}Cd_{0.85}S$ system.

Pearson coefficients correlation is determined as follows.

$$r = \frac{\sum_{i=1}^{n}(x-\tilde{x})(y-\tilde{y})/(n-1)}{\sqrt{\sum_{i=1}^{n}(x-\tilde{x})^2/(n-1)}\sqrt{\sum_{i=1}^{n}(y-\tilde{y})^2/(n-1)}} \quad (1)$$

Here, $\sum_{i=1}^{n}(x-\tilde{x})(y-\tilde{y})/(n-1)$ is the covariance of the sample, $\sqrt{\sum_{i=1}^{n}(x-\tilde{x})^2/(n-1)}$ is the standard deviation of variable x, $\sqrt{\sum_{i=1}^{n}(y-\tilde{y})^2/(n-1)}$ is the standard deviation of variable y.

2. SIESTA computation

In order to demonstrate the robustness and universality of the SBA method more efficiently and comprehensively, we utilized the the Spanish Initiative for Electronic Simulations with Thousands of Atoms code (SIESTA) program[1] with lower computational cost to perform calculations on the FCC-$Ni_{1-x}Mo_x$ alloy system and the Al_xCoNi_{6-x} system.

Computational details in SIESTA

We performed first-principles calculations using SIESTA, the general gradient approximation of the Perdew-Burke-Ernzerhof parameterization (GGA-PBE) was adopted for the exchange and correlation functions[2]. It is less computationally demanding thanks for the significant reduction of the basis size. All atoms were allowed to relax in struture optimization, the coresponding force torlerence and energy cutoff were set to 0.04 eV/Å and 500 Ry, respectively.

Rigid structure system

In this study, we utilized the FCC nickel lattice to substitute Mo atoms for Ni atoms using the supercell approximation method, with a supercell size of 2x2x2. the number of configurations increased exponentially as the proportion of Mo substitution increased. The configuration of $Ni_{28}Mo_4$ system with a total of 71 configurations is described as example. The ground state energies of these structures were arranged in ascending order, and the electrostatic energy and single-point energy corresponding to each structure are plotted sequentially. The Pearson coefficient[3]correlation between E_elec, E_sp and E_opt are 33.88% and 80.72%, respectively (Figure S2), which shows a misalignment in energy data and distortion in identifying low-energy structures. It is

observed that the coefficient between E_SBA_sp and E_opt is increased to 96.16%, after applying SBA to pre-optimized the unreasonable initial structures. The greater the match degree in energy information, the higher the precision in identifying low-energy conformations.

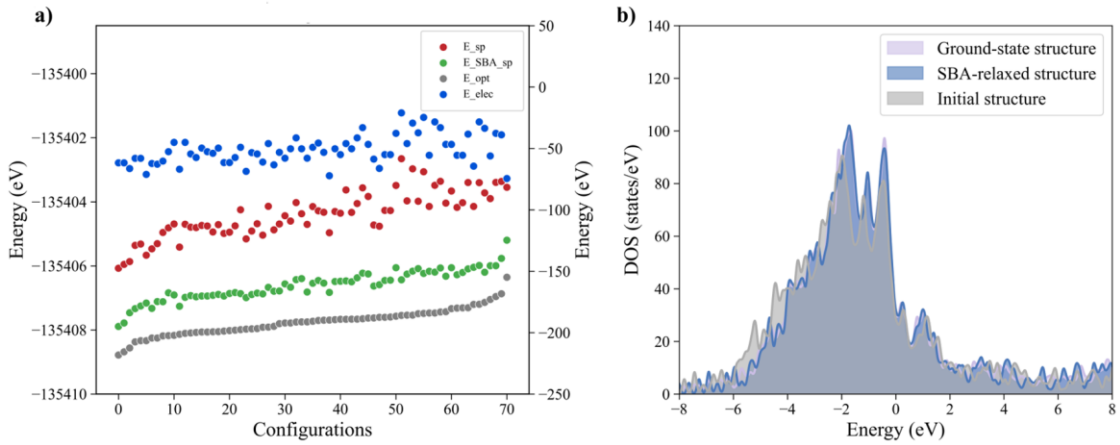

Figure S2 | Energy landscape and electronic structure analysis of the $Ni_{28}Mo_4$ system. a The energy landscape of different approaches, ground state energy (E_opt, grey circles), single-point energy (E_sp, red circles), electrostatic energy (E_elec, blue circles) and single-point energy after SBA relaxation (E_SBA_sp, green circles) in $Ni_{28}Mo_4$ system. the left axis is the energy range of E_elec, while right axis is for E_opt, E_sp and E_SBA_sp. **b** Comparison of the density of states the initial (light gray shading), SBA relaxed structure (light blue shading), and ground-state structure (light purple shading) in the $Ni_{28}Mo_4$ system.

After applying SBA to pre-optimized unreasonable initial structures, the averaged atomic force is reduced to 1.17 eV/Å from 1.91 eV/Å. To validate the rationality of SBA pre-optimization, we compared electronic properties such as the density of states (DOS) based on the structures from initial, standard optimized (ground-state), and pre-optimized by SBA. When comparing with the DOS of ground state, the Pearson coefficients of initial and SBA-relaxed are 97.44% and 99.9%, respectively. The latter perfect similarity indicates an excellent structure prediction ability of SBA.

While SIESTA is known for its fast first-principles calculations, we still consider the cost-benefit analysis of SBA screening. We defined 30 target structures to evaluate

the ranking capability in the $Ni_{28}Mo_4$ system. The proximity of the ROC curve to the top left corner indicates higher effectiveness in classifying true low-energy structures (Figure S3a). A more intuitive evaluation metric is the area under the ROC curve (AUC); the closer the AUC value is to 1, the better the method's performance. In this criterion, the AUC of SBA being near 1 indicates its ability to correctly classify all stable structures, making it a more desirable option for this system. In contrast, the performances of sp and elec are less satisfactory. Additionally, the lower time cost of SBA further supports the superiority of this method (Figure S3b).

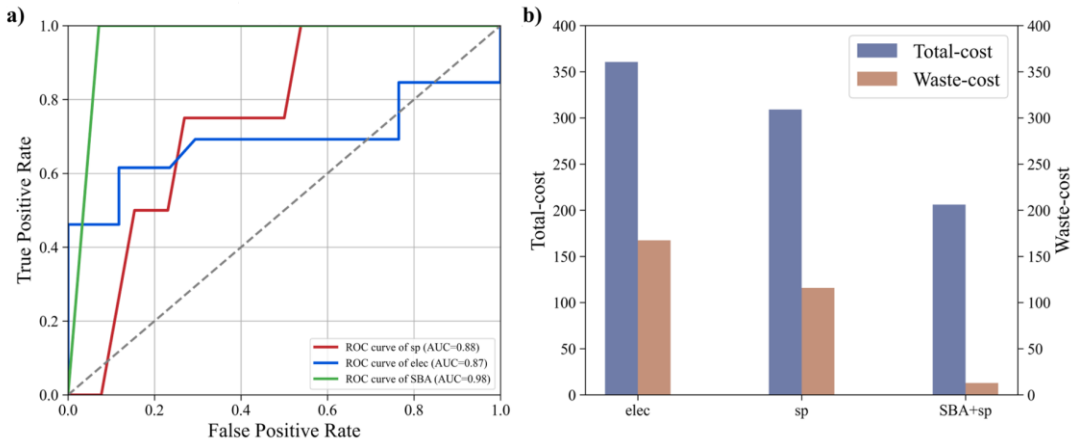

Figure S3 | Evaluation of stability classification and sampling cost in structural screening. a Receiver Operating Characteristic Curve. It displays the ability of elec, sp, and SBA to classify stable structures when applied to initial structures. **b** Sampling cost for screening low-energy structures in different methods.

Flexible structure system

In this study, the original cell of $AlCoNi_6$ was used to replace a group of Ni atoms with Al atoms using the supercell approximation method. In this system, the Pearson coefficients between E_elec, E_sp and E_opt are 73.10% and 89.29% (Figure S4a), respectively. After the application of SBA for pre-optimization of the structure, the match degree of E_SBA_sp is increased to 92.04%. This minimal gains in match degree attributes to the substantial lattice changes in relaxation, which leads to high difficulty in structural screening. However, the per atom force within the structure decreases from

2.17 eV/Å to 1.47 eV/Å, narrowing the gap with the ground state structure. The computational resources saved vary among different configurations with maximium savings of 60% as illustrtaed in Figure and the averaged computational resource saving exceeds 40%. (Figure S5b). Structural optimization with a more rational and stable structure enhances the convergence speed of relaxation (Figure S5b).

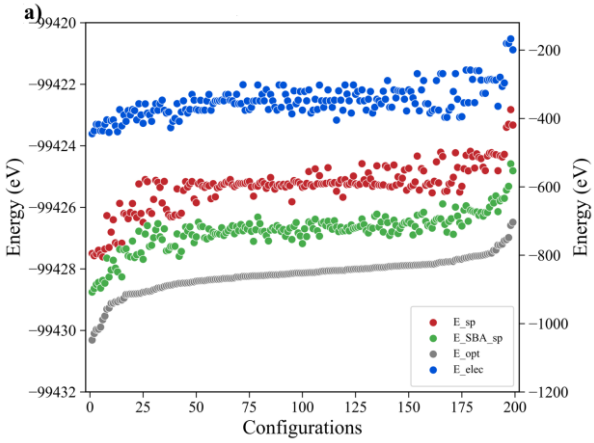

Figure S4 | Comparative evaluation of energy landscapes from different computational approaches. a The energies landscape derived from different approaches E_SBA_sp (green circles), E_sp (red circles), E_elec (blue circles) and E_opt(gray circles). The left y-axis is the range of E_elec, while right y-aixs is for E_opt, E_SBA_sp and E_sp. (199 inequivalent structures under the 2×2×1 supercell for $Al_{2.5}CoNi_{4.5}$ system).

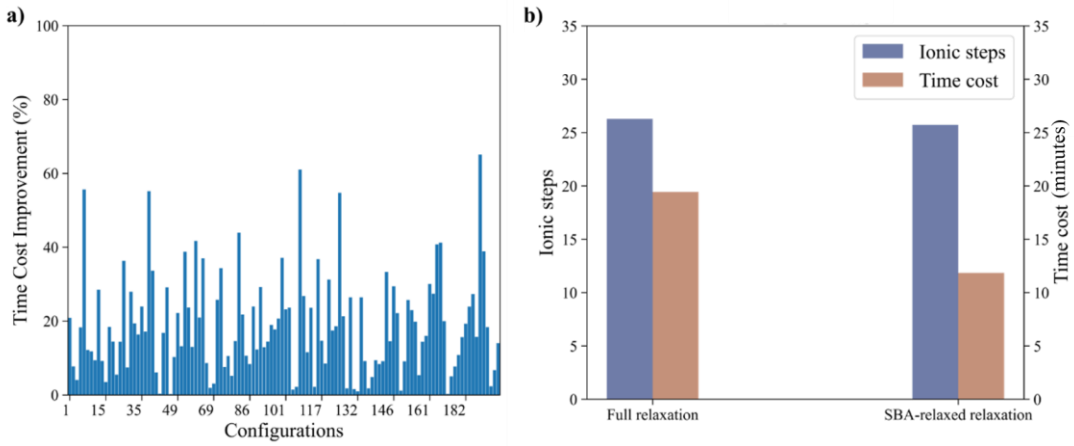

Figure S5 | Efficiency enhancement in structural optimization. a Time cost improvement of each configuration after SBA relaxation. **b** The comparison of

structural optimization convergence rate.

Table S1 | Efficiency of SBA in structural screening.

				Relaxation time (minutes)	
System	Subgraphs	Training set	Test	Full relaxation	SBA-relaxed
$Ni_{28}Mo_4$	49	29	71	6.45	3.78
$Al_{2.5}CoNi_{4.5}$	140	38	199	19.43	11.83

Performance in structure screening of SBA by comparing the relaxation time of test system and structure with traditional calculations by SIESTA.

References

1. Hirshman, S. P., Sanchez, R. & Cook, C. R. SIESTA: A scalable iterative equilibrium solver for toroidal applications. *Phys. Plasma* **18**, (2011).

2. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865-3868 (1996).

3. Zhang, W.-Y., Wei, Z.-W., Wang, B.-H. & Han, X.-P. Measuring mixing patterns in complex networks by Spearman rank correlation coefficient. *Physica A* **451**, 440-450 (2016).