ACS NANO

Editors' Choice
Open Access

www.acsnano.org

Inverse Design Method with Enhanced Sampling for Complex Open Crystals: Application to Novel Zeolite Self-assembly

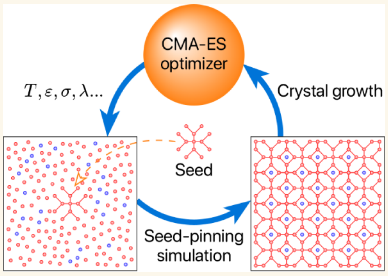
Chaohong Wang,* Alberto Pérez de Alba Ortíz,* and Marjolein Dijkstra*

Cite This: *ACS Nano* 2025, 19, 17423−17437

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: **Optimizing the design and synthesis of complex crystal structures presents pivotal opportunities and challenges in materials design. While recent computational advances in inverse design have proven effective for simpler crystals, their extension to intricate structures such as zeolites remains challenging. In this work, we introduce an efficient and robust inverse design workflow specifically tailored for the predictive design of a broad range of complex phases. By integrating an evolutionary parameter optimization strategy with enhanced sampling molecular dynamics simulations, this approach effectively surmounts the high energy barriers that typically hinder self-assembly in these complex structures. We apply this inverse design workflow to facilitate the efficient self-assembly of target zeolite frameworks in an efficient coarse-grained model of a tetrahedral network-forming component and a structure-directing agent. Using this method, we not only successfully reproduce the self-assembly of known structures like the Z1 and SGT zeolites and Type-I clathrates but also uncover previously unknown optimal design parameters for SOD and CFI zeolites. Remarkably, our approach also leads to the discovery of an uncatalogued framework, which we designate as Z5. Our methodology not only enables the screening and optimization of self-assembly protocols but also expands the possibilities for discovering hypothetical structures, driving innovation in materials design and offering a robust tool for advancing crystal engineering in complex systems.**



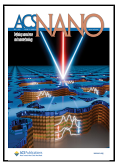
KEYWORDS: *inverse design, enhanced sampling, evolution strategy, open crystals, zeolites, self-assembly, simulations*

INTRODUCTION

Zeolites are microporous aluminosilicate crystals characterized by open three-dimensional framework structures composed of corner-sharing TO_4 tetrahedra, where T typically represents silicon (Si) or aluminum (Al) in varying ratios. Zeolites are extensively used in numerous industrial and technological applications, including as catalysts[1−3] molecular sieves[3,4] adsorbents for environmental protection[5] and even as biomedical materials.[6] This variety of applications is enabled by the vast diversity of specific pore shapes and sizes with which zeolites can be synthesized. To produce such zeolites, hydrothermal synthesis (HS) is the most common route.[7] In this method, a precursor such as tetraethylorthosilicate is mixed with water and a so-called structure directing agent (SDA), typically a tetraalkylammonium cation or another organic cation, under specific temperature and pressure conditions. The SDAs interact with the framework precursors, playing a crucial role in guiding the polymerization, self-aggregation, crystallization, and growth of the zeolite.[8] By adjusting the composition of T atoms and SDAs, as well as the thermodynamic conditions, one can control the metastability of different polymorphs and promote the formation of a specific zeolite framework. Hundreds of synthetic and natural zeolite frameworks have been cataloged by the Structure Commission of the International Zeolite Association (IZA-SC).[9] Still, more than four million hypothetical zeolites remain unrealized in experiments, according to databases built by varying unit cell parameters, density, T atom positions and compositions using Monte Carlo search methods.[10−12]

The potential applications of these hypothetical zeolites drive the search for novel strategies to fine-tune hydrothermal synthesis. Producing zeolites with desired properties, including hypothetical ones, requires identifying the appropriate T elements and their ratios, as well as optimizing the thermodynamic conditions, and selecting the optimal SDAs in the most effective compositions. These agents are crucial for guiding the framework components to form specific topologies. In this endeavor, computational approaches have become

Received: December 5, 2024
Revised: April 5, 2025
Accepted: April 7, 2025
Published: May 1, 2025



https://doi.org/10.1021/acsnano.4c17597
ACS Nano 2025, 19, 17423−17437

increasingly crucial for understanding and controlling zeolite self-assembly. Molecular dynamics (MD) simulations have provided insights into the zeolite self-assembly process, while artificial intelligence (AI) and optimization algorithms have advanced the design of stable zeolite frameworks. Here, we combine accelerated MD simulations with evolutionary computing techniques to facilitate the self-assembly of target frameworks within a coarse-grained zeolite model.

Efficiently screening for optimal conditions for the self-assembly of a target zeolite framework is prohibitively expensive when relying on systematic experimentation or forward modeling. The exploration of such combinatorial design parameter spaces has motivated the surge of inverse design methods. Inverse design is a bottom-up approach that involves adjusting the properties of building blocks, such as particle shape and particle interactions, as well as the thermodynamic conditions, to achieve the desired properties of the self-assembled material. Many inverse design frameworks have been proposed in recent decades.[13] For instance, Torquato and coworkers used inverse statistical mechanical methods to develop isotropic potentials that facilitate the formation of various colloidal lattices.[13−15] Other researchers have also designed potentials and conditions to identify quasicrystals through inverse design techniques.[16,17] However, porous materials such as zeolites and metal−organic frameworks (MOFs) pose unique challenges for inverse design due to their relatively complex topologies compared to simpler crystalline materials. Some studies have identified novel thermodynamically stable porous materials by leveraging existing data sources, such as the IZA-SC, and employing artificial intelligence, without considering the self-assembly mechanisms. For instance, porous materials have been designed using generative adversarial neural networks[18,19] variational autoencoders[20] diffusion models[21] and other machine learning methods.[22] While previous work has employed optimization strategies to design the self-assembly of open crystals[23] to the best of our knowledge, a generic, robust inverse design framework capable of optimizing the self-assembly of complex zeolitic frameworks does not currently exist.

An iterative inverse-design framework for zeolite self-assembly requires computational efficiency that is difficult to achieve with atomistic models. Recently, Molinero *et al.* proposed a novel and concise coarse-grained model for zeolites consisting of tetrahedral network-former T particles and structure-directing agent S particles.[24] The T particles are parametrized based on the mW water model[25] which in turn is a special parametrization of the Stillinger-Weber (SW) model.[26] The T-S model considers two-body interactions between T-T, T-S and S-S pairs, as well as an additional three-body interaction between T particles to capture tetrahedrality. This coarse-grained T-S model has been reported to reproduce several known structures, including the SGT zeolite framework and the sII clathrate, and a zeolite analog of the FIR-30[27] metal−organic framework dubbed Z1[24,28] along with various other phases.[29] Due to the complexity of phases introduced by many-body interactions, binary components, and a broad range of adjustable interaction parameters, this model has the potential to reproduce a wide diversity of zeolite frameworks.[30] The T-S model is also highly tunable, with free parameters such as particle size, interaction strength and range, tetrahedrality, making it well-suited for the inverse design of the self-assembly of desired zeolites.

In this work, we devise an innovative inverse design workflow capable of identifying optima in a multidimensional parameter space for the self-assembly of desired zeolite frameworks. To our knowledge, this is the first framework to robustly deliver self-assembly for such a variety of target zeolite phases. Our workflow utilizes the coarse-grained T-S model to represent interactions between T and S particles and to propagate their dynamics. To determine promising parameter values in the T-S zeolite model, we employ an evolutionary approach — the covariance matrix adaptation evolution strategy (CMA-ES)[31]— which is a robust, population-based, gradient-free optimizer. The parameters we optimize correspond to the effective size and the attraction strength between T-T, T-S and S-S pairs of particles, the strength of the tetrahedrality in T particles, and the temperature. To ensure that the identified potential parameter values can lead to self-assembly into the target framework, the fitness of the optimizer is determined from a nucleation perspective. To accelerate the nucleation and growth of our framework even further, we employ enhanced sampling techniques, i.e., our own variation of seed-pinning[32,33] to favor nucleation events that are typically rare in standard MD simulations. To this end, we introduce a seed, i.e., a small crystallite of the desired zeolite into a fluid mixture and monitor its growth to evaluate the fitness of the parameters. To determine this tendency to grow, we measure fluctuations in the environment similarity order parameter[34] with respect to the target framework as cataloged in the IZA-SC database. Additionally, we implemented a novel algorithmic approach to reduce the computational cost of sampling and measuring nucleus size fluctuations, thereby increasing the efficiency of the inverse design.

This paper is organized as follows. In the Results section, we present our results for the Z1 and SGT zeolite frameworks, as well as newly discovered parameters for SOD, sI clathrate, CFI, and a novel, previously uncatalogued, framework Z5. Next, we provide our conclusions in the Conclusions section. Finally, in the Methods section, we describe each of the methodology's elements, including the T-S zeolite model, the inverse design workflow, the CMA-ES optimizer, the environment similarity order parameter, and the seed-pinning method.

RESULTS

Reproducing Self-assembly Parameters for Known Target Phases.

Framework-Type Z1 Zeolite. We start our investigation by reverse-engineering the framework-type Z1 within a coarse-grained zeolite model consisting of a binary mixture of T and S particles. Our goal is to identify the optimal set of design parameters that facilitates the self-assembly of Z1 within this mixture using the protocol as outlined in Methods section.

The zeolite Z1, as examined in ref 29, has a T particle structure, derived from the Zn atom positions in the unit cell of the MOF FIR-30.[35] We obtain the coordinates of the 120 Zn atoms in the unit cell of MOF FIR-30 from the Supporting Information of ref 27. These 120 Zn atoms are represented by the red particles in Figure 1a. We then scale the coordinates so that the nearest neighbor distance between Zn atoms is set to 3.3 Å, matching the typical T-T interparticle distance. To convert the 3-coordinated particles at the surface of the triangular channels to the 4-coordinated particles as observed in Z1, we add 16 additional T particles, represented by the green particles in Figure 1a. This adjustment brings the total number of T particles in the FIR-30 unit cell to 136, closely resembling the Z1

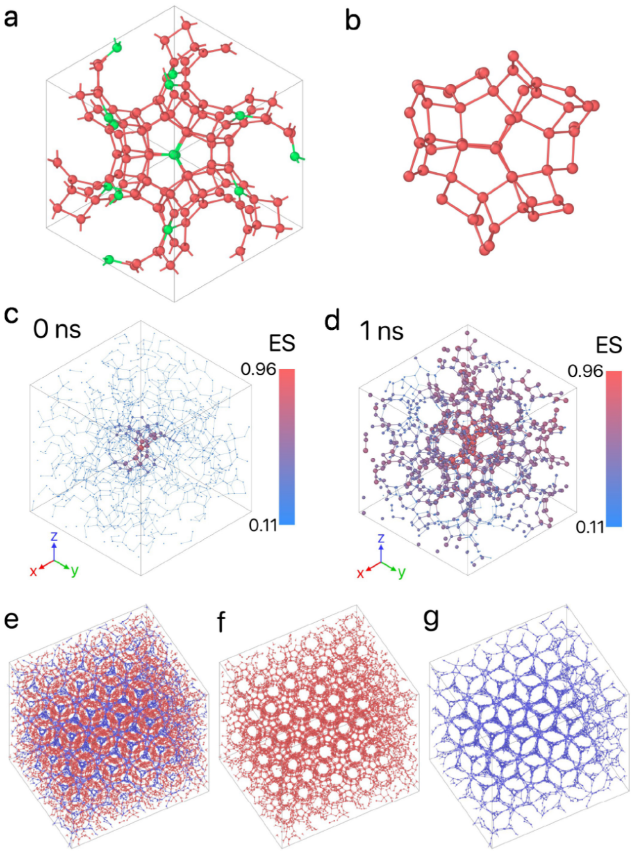


Figure 1. (a) The unit cell of the MOF FIR-30 framework,[35] consisting of 120 Zn atoms of FIR-30 (shown in red) and 16 additional particles (shown in green).[29] This FIR-30 unit cell, composed of 136 particles, is used to create a supercell, with coordinates scaled so that the neighbor distance of the Zn atoms is 3.3 Å, matching the typical T-T interparticle distance. (b) A crystalline seed of 50 particles is extracted from this FIR-30 supercell. (c) The initial configuration at 0 ns and (d) the final configuration at 1 ns of a seed-pinning simulation for the highest-fitness sample from the 29th generation of the inverse design protocol for Z1. The radius of the shown T particles is proportional to their environment similarity (ES) to the target framework, as indicated by the colorbar. For clarity, S particles are not shown. (e−g) The Z1 framework obtained from an unbiased self-assembly simulation using the highest-fitness solution from the 29th generation. (e) The intertwined structure of T and S particles. (f) The porous structure formed by T particles. (g) The gyroid network formed by S particles.

structure, which contains 132 particles. Using this modified unit cell, we create a supercell of FIR-30, from which we extract a crystalline seed consisting of 50 particles as shown in Figure 1b for the seed-pinning simulations. Additionally, we identify the unique environments for evaluating the environment similarity order parameter. The total number of unique environments of Z1 is as large as 136 with typically 55−75 particles in each environment.

For our inverse design protocol, we consider the following seven design parameters, ε_{TT}, ε_{TS}, ε_{SS}, σ_{TS}, σ_{SS}, λ, and temperature T. The values of these parameters are sampled at each generation from a multivariate Gaussian distribution according to the CMA-ES algorithm as outlined in the Methods section. For each set of parameters, also called sample, we run a short MD simulation of 1 ns, where a crystalline seed is immersed in a disordered fluid using the seed-pinning method. This short but effective sampling time is chosen to balance computational efficiency with the need to evaluate a vast number of samples across multiple generations. A typical example of the initial and final configurations of such a seed-pinning simulation is shown in Figure 1c,d. The environment similarity order parameter is then used to evaluate the fitness of each sample. This information is used by the CMA-ES optimizer.

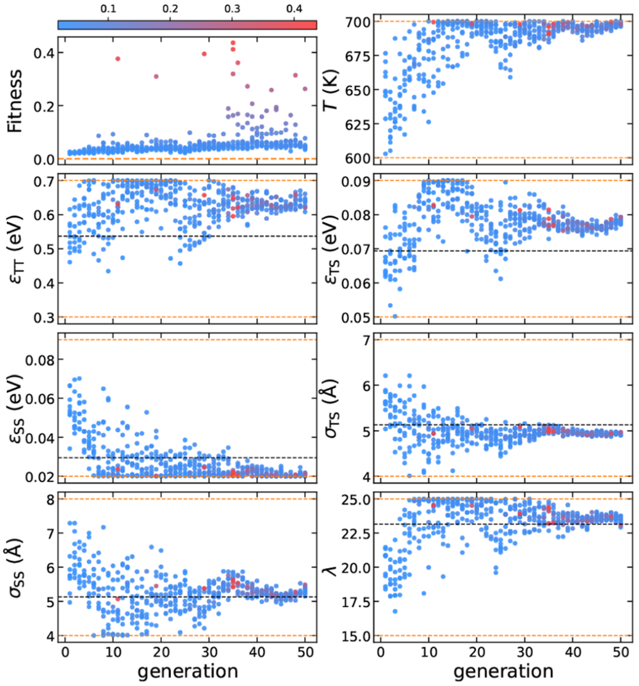


Figure 2. Evolution of the design parameters for Z1, with each point colored according to its fitness. The color bar is normalized based on the minimum and maximum values of fitness. The dashed orange lines represent the parameter boundaries, while the dashed black lines correspond to the values of ref 36.

We run the evolution strategy for 50 generations, and present the results of this reverse-engineering process in Figure 2. This maximum number of generations is determined arbitrarily as a compromise between limiting computational cost while still producing successful parameters for self-assembly. Although the average fitness value increases gradually over time, we identified several high-fitness samples that stand out throughout the process. Additionally, the evolution of the design parameters is shown in Figure 2, indicating that the inverse design protocol converges toward a specific range of parameter values, where high-fitness values are found. For instance, in the high-fitness solutions, σ_{TS}, σ_{SS}, ε_{SS}, and λ converged to 5.0 Å, 5.3 Å, 0.022 eV, and 24.3, respectively—values close to the previously reported 5.13 Å, 5.13 Å, 0.029 eV, and 23.15, respectively.[36] However, the interaction strengths ε_{TT}, ε_{TS}, and the temperature T, converged to 0.653 eV, 0.081 eV, and 698 T, which are significantly higher than those reported in ref 36 indicated by the horizontal black dashed lines in Figure 2.

To validate our inverse-design protocol, we perform unbiased self-assembly simulations using the top eight highest-fitness parameter combinations. Three of these parameter combinations, listed in Table 1, spontaneously nucleate and grow into Z1, as shown in Figure 1e−g. Compared to the parameters reported in ref 36, our solutions exhibit larger ε_{TT} and ε_{TS}, which

Table 1. High-Fitness Solutions That Spontaneously Self-assemble into the Framework-Type Z1 Zeolite within 50 ns

Solution in	T (K)	ε_{TT} (eV)	ε_{TS} (eV)	ε_{SS} (eV)	σ_{TS} (Å)	σ_{SS} (Å)	λ	Fitness
Gen 11	699	0.630	0.083	0.023	4.94	5.07	24.52	0.38
Gen 19	698	0.672	0.080	0.020	5.05	5.45	24.51	0.31
Gen 29	698	0.657	0.081	0.024	5.09	5.38	23.91	0.39
Ref 36	<654	0.537	0.069	0.029	5.13	5.13	23.15	

explains their ability to self-assemble into Z1 at elevated temperatures. Additionally, the self-assembly simulations showed no mesophases during the nucleation pathway of Z1, aligning with a recently published rescaled zeolite model where the temperature is below 636 K.[36] Overall, our optimized design parameters closely match previous results, as shown in Table 1.

Cage-Type SGT Zeolite. We now use our inverse design protocol to find the optimal set of design parameter values that facilitate the self-assembly of the cage-type SGT zeolite within the coarse-grained zeolite model of T and S particles. SGT is a typical high-purity silica zeolite with a predominantly tetrahedral structure and, according to our classification method, falls under the cage-type zeolites. We obtain the coordinates of the T particles for SGT from the IZA database[37] and present the structure in Figure 3a). We assume that each large cage contains one S particle. Given that the SGT unit cell consists of 64 T particles and 4 large cages, the composition χ_T is set at $64/(64 + 4) \approx 0.94$. For the inverse design process, we use a small seed of 20 T particles, as shown in Figure 3b, chosen as the minimal size necessary to induce nucleation. In addition, we identify the unique environments for evaluating the environment similarity order parameter. The number of unique environments for SGT is 32 as shown in Table 8. We follow the same inverse design protocol as described in the Methods section for the inverse design of Z1, and consider again seven design parameters, ε_{TT}, ε_{TS}, ε_{SS}, σ_{TS}, σ_{SS}, λ, and T, sampled from a multivariate Gaussian distribution according to the CMA-ES algorithm. For each set of parameters, we run a short MD simulation of 1 ns using the seed-pinning method. Figure 3c,d shows the initial and final configurations for the parameter combination corresponding to generation 44 of a seed-pinning simulation. The fitness of each sample is evaluated using the environment similarity order parameter, and these fitness values are then used by the CMA-ES optimizer. We run the evolution strategy for 50 generations, with the results for the evolution of the design parameters presented in Figure 4.

The evolution of the design parameters reveals high-fitness solutions across a broader range of values compared to Z1, suggesting that the self-assembly conditions for SGT are less restrictive. Additionally, we observe that the inverse design protocol converges for only a few design parameters, within a specific range where high fitness values are found. For example, T, σ_{TS}, and λ converge to 686 K, 5.4 Å, and 24.7, respectively, within the high-fitness solutions, while ε_{TT}, ε_{TS}, ε_{SS}, and σ_{SS} exhibit considerable variability.

To validate the wide variability in the parameter values of the high-fitness solutions, we run unbiased self-assembly simulations using the top ten highest-fitness solutions. We identify six solutions that spontaneously crystallize into SGT within 50 ns. We present the corresponding parameter values in Table 2. Previously reported parameters that facilitated SGT self-assembly match those of Z1, except for a higher χ_T.[29] Note that these parameters are based on the T-S model without rescaling the T-T distance.[36]

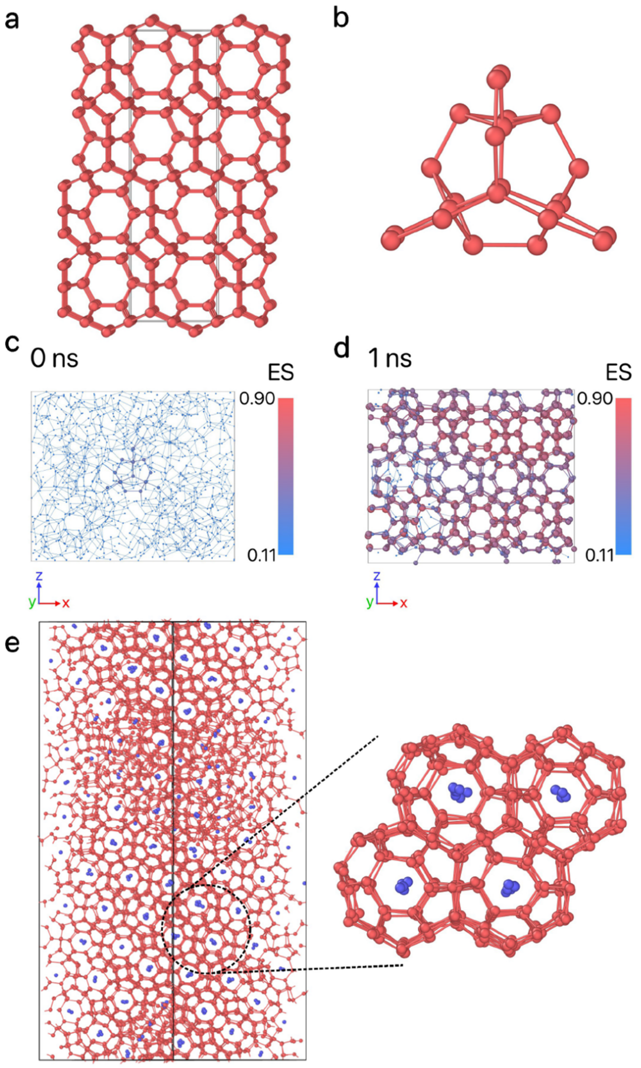


Figure 3. (a) The structure of the perfect SGT framework.[37] (b) A crystalline seed composed of 23 T particles. (c) The initial configuration is at 0 ns, and (d) the final configuration is at 1 ns of a seed-pinning simulation for the highest-fitness sample from the 44th generation of the inverse design protocol for SGT. The radius of the shown T particles is proportional to their environment similarity (ES) to the target framework, as indicated by the colorbar. For clarity, S particles are not shown. (e) The SGT framework obtained from an unbiased self-assembly simulation using the highest-fitness solution from the 44th generation. This configuration shows an AB random stacking structure.

Our optimized values for ε_{TT} and λ in the high-fitness solutions closely match previous results[29] which is expected, as SGT is primarily a pure silica zeolite with strong tetrahedral interactions between T particles. However, the values for ε_{SS} and σ_{SS} differ significantly. The optimized σ_{SS} values range from 5.09 to 7.61 Å, accompanied by a gradual increase in ε_{SS} from 0.044 to 0.079 eV. The larger interaction distance, σ_{SS}, necessitates a

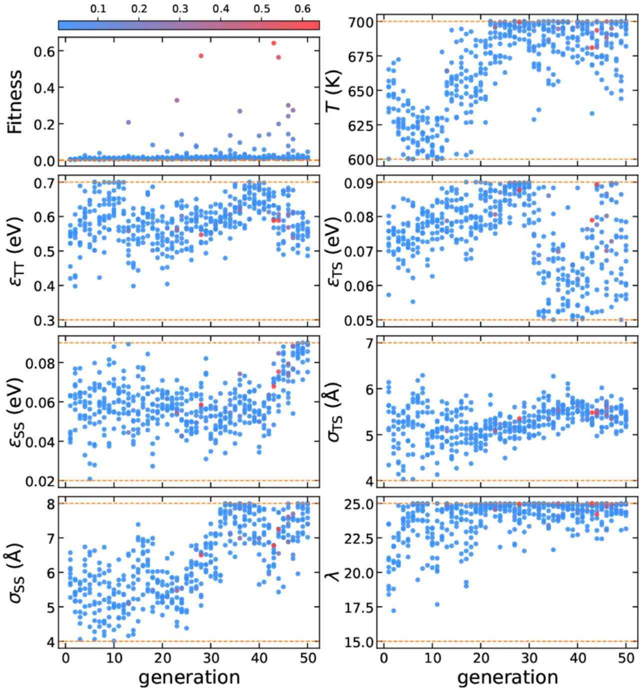


Figure 4. Evolution of the design parameters for SGT, with each point colored according to its fitness. The color bar is normalized based on the minimum and maximum values of fitness. The dashed orange lines represent the parameter boundaries.

stronger interaction strength, ε_{SS}, to achieve an effective shaping effect. This indicates that rather than narrow optimization valleys, there are broader channels in the optimization space, where trade-offs between parameters allow various types of S particles to guide the formation of the same T particle framework.

Discovering Self-assembly Parameters for Known Target Phases. *Cage-Type SOD Zeolite and Byproducts.* After the successful reproduction of Z1 and SGT, our inverse design protocol has proven its effectiveness in targeting frameworks known to self-assemble within the T-S model. We now aim to design frameworks that have not yet been reproduced using the T-S model. We start with SOD, one of the simplest zeolite frameworks, characterized by only six unique environments. The SOD unit cell contains 12 T particles and includes 2 cages, resulting in a T particle composition of $\chi_T = 12/(12 + 2) \approx 0.86$. The perfect SOD crystal and the crystalline seed used for seed-pinning are shown in Figure 5a,b, respectively.

In Figure 6, we present the evolution of the fitness values during the optimization process for SOD self-assembly. We use a seed of size 23 T particles. Similar to the trends observed in the evolution of Z1 and SGT, a consistent high fitness across all samples within a generation is not achieved due to the rare event nature of nucleation. However, from generation 25 onward, a significant number of samples per generation exhibits a high fitness. In comparison to the optimized parameters for Z1 and SGT, the value of ε_{TT} decreases, indicating that a weaker tetrahedral interaction strength facilitates the formation of 4-membered and 6-membered rings in SOD.[39] Additionally, σ_{SS} converges to approximately 8 Å, which is close to the distance between S particles in the SOD framework. Two high-fitness

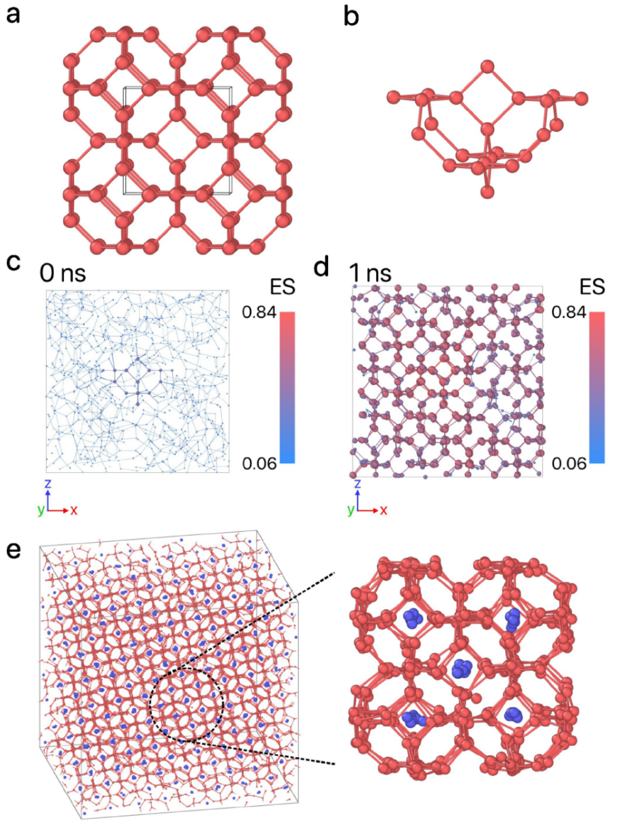


Figure 5. (a) The structure of the perfect SOD framework.[38] The black box indicates the unit cell. (b) A crystalline seed composed of 23 T particles. (c) The initial configuration at 0 ns and (d) the final configuration at 1 ns of a seed-pinning simulation for the highest-fitness sample from the 40th generation of the inverse design protocol for SOD. The radius of the shown T particles is proportional to their environment similarity (ES) to the target framework, as indicated by the colorbar. For clarity, S particles are not shown. (e) The SOD framework obtained from an unbiased self-assembly simulation using the high-fitness solution from the 40th generation. To our knowledge, there are no previously reported parameters for SOD in the T-S model.

Table 2. High-Fitness Solutions That Spontaneously Self-assemble into the Cage-Type SGT Zeolite within 50 ns

Solution in	T (K)	ε_{TT} (eV)	ε_{TS} (eV)	ε_{SS} (eV)	σ_{TS} (Å)	σ_{SS} (Å)	λ	Fitness
Gen 13	664	0.558	0.079	0.044	5.08	5.09	24.85	0.21
Gen 23	696	0.562	0.081	0.055	5.09	5.47	24.55	0.33
Gen 36	695	0.619	0.070	0.074	5.54	6.99	25.00	0.27
Gen 43	681	0.588	0.079	0.068	5.48	6.78	25.00	0.64
Gen 44	694	0.589	0.089	0.075	5.49	7.25	24.21	0.56
Gen 46	688	0.568	0.070	0.079	5.43	7.61	24.76	0.30
Ref 29		0.537	0.069	0.029	5.13	5.13	23.15	

https://doi.org/10.1021/acsnano.4c17597
ACS Nano 2025, 19, 17423−17437

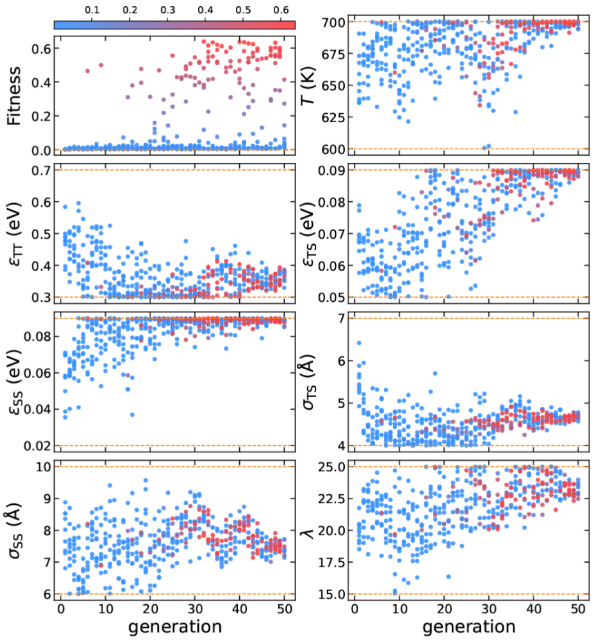


Figure 6. Evolution of the design parameters for SOD, with each point colored according to its fitness. The color bar is normalized based on the minimum and maximum values of fitness. The dashed orange lines represent the parameter boundaries.

samples were obtained in the 40th and 41st generations, as shown in Table 3. Both of these samples show very similar parameter values and successfully produced a SOD framework in an unbiased self-assembly simulation. Figure 5c shows the nucleated SOD framework using the parameters from the 40th generation. When comparing these optimized SOD parameters to those for ZI and SGT, we observe significant differences in ε_{TT} across the different frameworks. This observation aligns with the literature, which indicates that the selection of framework is primarily governed by the interaction between T particles, with ε_{TT} playing a key role in determining the phase behavior of this model.[40] Similarly, σ_{TS} is optimized to geometrically support the target framework, generally matching the radius of the inscribed sphere in the SOD cage when a single S particle is present.

During the inverse design of SOD, when using a seed of 19 rather than 23 T particles, we unexpectedly discovered the formation of the sI clathrate—also known by the zeolite code name MEP. As shown in Figure 7, the evolution of the fitness yielded only two high-fitness solutions. In unbiased self-assembly simulations, both of these solutions resulted in the formation of the sI clathrate rather than the intended SOD structure. Silica clathrates form a distinct subset of zeolites characterized by rings that are too small to allow the free movement of guest molecules within the crystal. However, it is precisely this property that makes them highly researched for gas storage applications. Molinero previously succeeded in reproducing sII clathrate (also known as MTN) in previous work.[29]

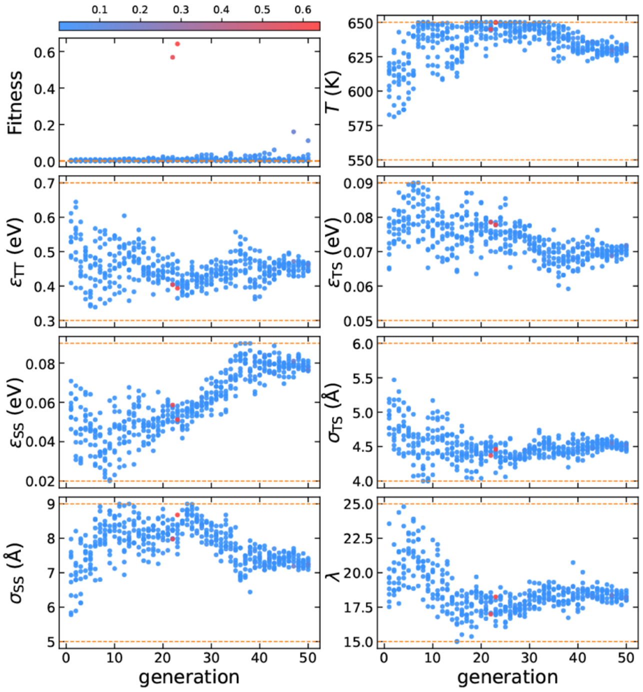


Figure 7. Evolution of the design parameters for SOD with a crystalline seed of 19 T particles and a lower temperature range. All the points are colored according to their fitness. The color bar is normalized based on the minimum and maximum values of fitness. The dashed orange lines represent the parameter boundaries.

The structure of the sI clathrate shows that there is one S particle in each cage, regardless of whether it is a small cage, 5^{12}, or a large cage, $5^{12}\,6^2$, as shown in Figure 8c. Consequently, the T particle composition, $\chi_T = 46/(46 + 8) \approx 0.85$, is similar to that of SOD. Compared to the parameters optimized for SOD, the main differences lie in ε_{SS} and λ. The lower value of ε_{SS} facilitates the formation of an anisotropic distribution of S particles in the clathrate structure (Table 4).

Hole-Type CFI Zeolite. We now turn our attention to the inverse design of CFI within our coarse-grained T-S model. CFI is a porous framework with one-dimensional channels, i.e., a hole-type framework. To our knowledge, no T-S model parameter sets have been optimized for the self-assembly of CFI. The CFI unit cell consists of 32 T particles and assumes the presence of 2 S particles, resulting in a composition of $\chi_T = 32/(32 + 2) \approx 0.94$.[41] The structure can be characterized by slices perpendicular to the channel direction. Therefore, for the inverse design, we employ a crystalline seed consisting of a single-particle-thick layer that spans the entire simulation box rather than a spherical seed, as illustrated in Figure 9b. Furthermore, due to the significant differences in S−S distances within and between the channels, a lower value of ε_{SS}, i.e., the strength of the isotropic two-body interaction between S particles, is required. Therefore, we reduce the optimization range of ε_{SS} from 0.02−0.09 to 0.01−0.05 eV. We again use the same inverse design recipe as described previously.

Table 3. High-Fitness Solutions That Spontaneously Self-assemble into the Cage-Type SOD Zeolite within 50 ns

Solution in	T (K)	ε_{TT} (eV)	ε_{TS} (eV)	ε_{SS} (eV)	σ_{TS} (Å)	σ_{SS} (Å)	λ	Fitness
Gen 40	700	0.403	0.090	0.089	4.75	8.03	23.36	0.64
Gen 41	698	0.382	0.089	0.087	4.66	8.12	22.65	0.63

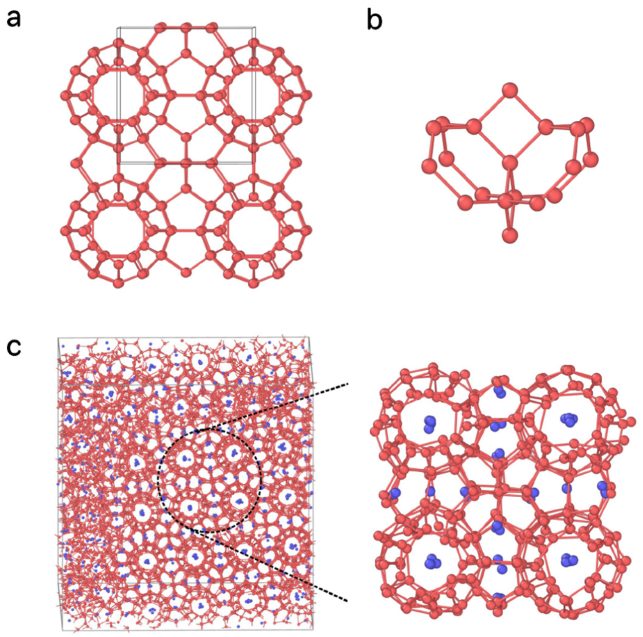


Figure 8. (a) The structure of the perfect sI clathrate. (b) A crystalline seed composed of 19 T particles. (c) The sI clathrate obtained from an unbiased self-assembly simulation using the high-fitness solution from the 22nd generation and a zoom-in diagram.

We present the evolution of fitness values during the optimization process for CFI self-assembly in Figure 10. We observe a clear trend of increasing fitness with stronger tetrahedral interactions, as indicated by the continuous rise in both ε_{TT} and λ. This is consistent with the highly tetrahedral nature of the CFI framework. We select the highest fitness sample from the 30th generation, see Table 5. Compared to the interaction parameters for cage-type zeolites, e.g., SGT and SOD, a higher value of $\sigma_{TS} \approx 6$ Å, which closely matches the radius of the largest 14-membered ring, provides strong geometric support for the CFI framework. An unbiased simulation using this parameter set successfully self-assembles into the CFI framework, revealing a two-step nucleation pathway for its crystallization, as shown in Figure 9e−i. As illustrated in Figure 9g, parallel strings of S particles initially guide the formation of a hexagonal mesophase, where T particles arrange into a locally disordered hexagonal structure that retains long-range order. A nucleation site subsequently forms within this structure. As shown in Figure 9i, the nucleus rapidly grows, transforming the entire crystal into the CFI framework within 1 ns (see Supporting Information).

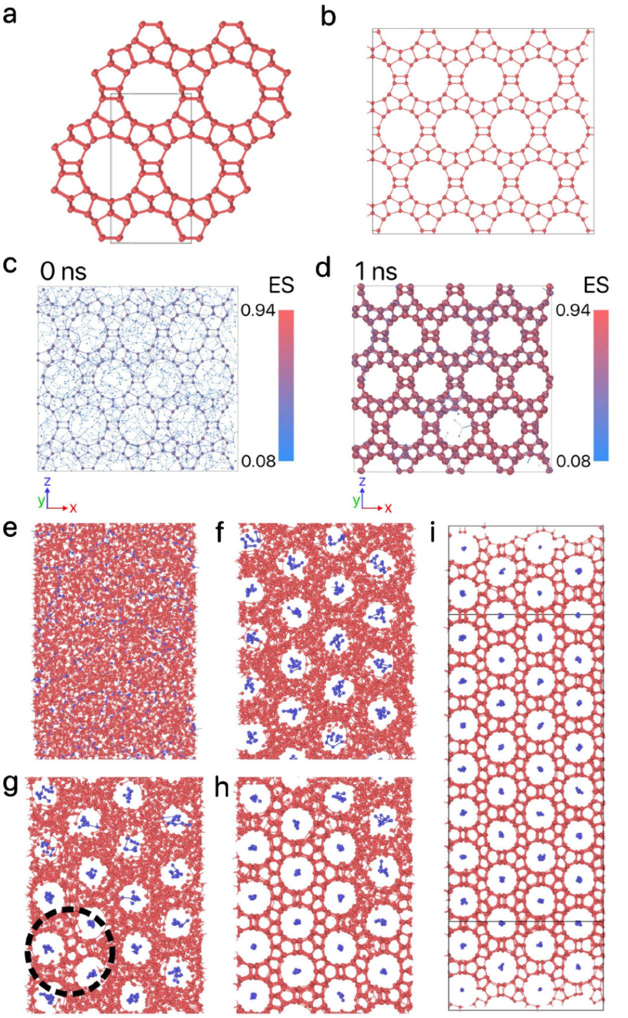


Figure 9. (a) A partial view of the perfect CFI structure. (b) A crystalline seed consisting of a single-particle-thick layer. (c) The initial configuration at 0 ns and (d) the final configuration at 1 ns of a seed-pinning simulation for the highest-fitness sample from the 30th generation of the inverse design protocol for CFI. The radius of the shown T particles is proportional to their environment similarity (ES) to the target framework, as indicated by the colorbar. (e−h) The CFI structure obtained from an unbiased self-assembly simulation using the highest-fitness sample from the 30th generation. (e) An amorphous mixture of T and S particles. (f) A metastable hexagonal mesophase, where chains of S particles support the channels. (g) A nucleation site forms within this structure, highlighted by the dashed circle. (h) Growth of the CFI framework. (i) The fully developed CFI framework spans the entire system. To our knowledge, there are no previously reported parameters for CFI in the T-S model.

Discovering Self-assembly Parameters for New Phases. *Hole-Type AFI Zeolite and Novel Z5 Zeolite.* As demonstrated by the discovery of the sI clathrate structure during the inverse design of the SOD zeolite, the unbiased self-assembly test can occasionally yield a polymorph distinct from the targeted framework when using seed-pinning. In this section, we report the discovery of an uncataloged framework during the inverse design of the hole-type AFI framework.[42] AFI is a one-dimensional porous zeolite characterized by its large 12-membered ring as shown in Figure 11a. The AFI unit cell contains 24 T particles and assumes the presence of 2 S particles, resulting in a composition of $\chi_T = 24/(24 + 2) \approx 0.92$. Similar to the CFI case, we use a crystalline seed of a single-particle-thick

Table 4. High-Fitness Solutions That Spontaneously Self-assemble into the sI Clathrate within 50 ns

Solution in	T (K)	ε_{TT} (eV)	ε_{TS} (eV)	ε_{SS} (eV)	σ_{TS} (Å)	σ_{SS} (Å)	λ	Fitness
Gen 22	645	0.404	0.079	0.059	4.37	7.98	16.99	0.57
Gen 23	650	0.394	0.078	0.051	4.46	8.68	18.25	0.64

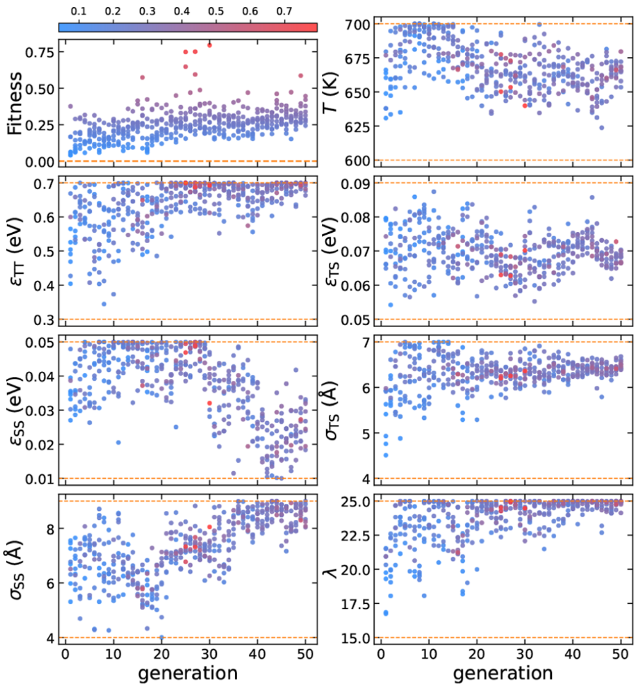


Figure 10. Evolution of the design parameters for CFI, with each point colored according to its fitness. The color bar is normalized based on the minimum and maximum values of fitness. The dashed orange lines represent the parameter boundaries.

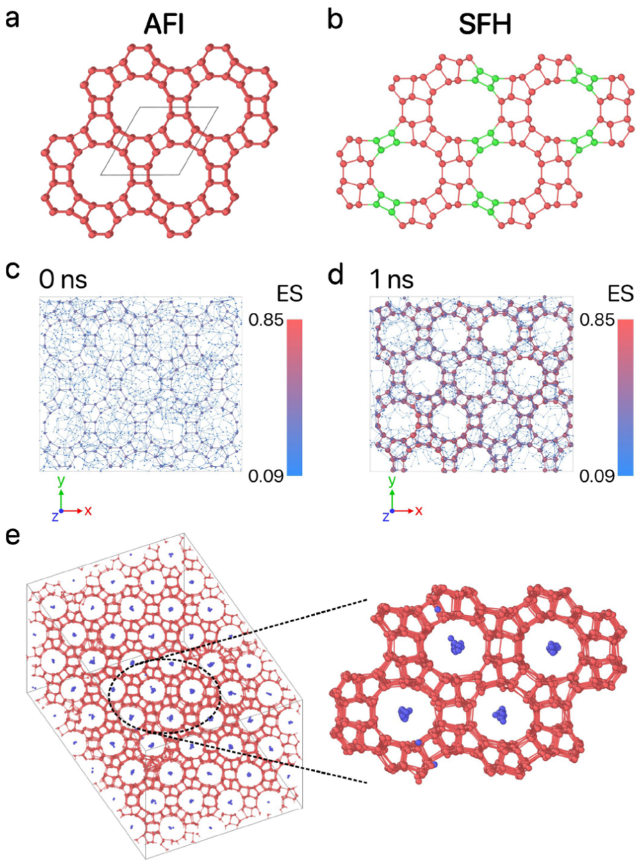


Figure 11. (a) The structure of the targeted AFI framework. (b) The structure of the SFH framework, which is the closest match to the Z5 framework. The rings formed by green particles in the SFH framework are absent in the Z5 framework. (c) The initial configuration at 0 ns and (d) the final configuration at 1 ns of a seed-pinning simulation for the highest-fitness sample from the 47th generation of the inverse design protocol for AFI. The radius of the shown T particles is proportional to their environment similarity (ES) to the target framework, as indicated by the colorbar. (e) The newly discovered Z5 framework, assembled using the highest-fitness solution from the 19th generation.

layer in our inverse design simulations. We follow again our inverse design protocol to optimize the seven design parameters for the self-assembly of AFI. The number of unique environments is $N_E = 24$. The evolution of the design parameters, as shown in Figure 12, suggests that some optimal parameter values may lie outside the predefined ranges, i.e., T, ε_{TT}, ε_{SS}, and λ have reached upper or lower boundaries, and the fitness has reached a plateau. In an additional run, with expanded parameter ranges for ε_{SS} and λ, we find there is no significant improvement in fitness. We selected the highest-fitness parameter combination, as presented in Table 6.

When tested in an unbiased self-assembly simulation, this parameter set leads to the discovery of a new zeolite framework not cataloged in IZA, as illustrated in Figure 11e. This new porous zeolite, which we label Z5, features the largest 12-membered ring, consistent with the target AFI framework. However, the substructures within the T framework differ significantly. A more similar T substructure can be found in the IZA database, i.e., the SFH framework[43] as shown in Figure 11b. However, the key difference is that the SFH framework contains two additional rectangular rings along the large 12-membered ring, which are absent in the Z5 framework. Furthermore, unlike the two-step nucleation pathway observed for the CFI framework, the crystallization of Z5 occurs without any intermediate mesophases.

The fact that the final structure obtained from unbiased self-assembly simulations differs from that obtained in biased seeding simulations, despite using the same parameters, suggests the existence of multiple nucleation pathways. These pathways may diverge or converge before or after reaching the seed size, which limits our control over growth in unbiased simulations. For example, in Figure 11c,d, we show that the layers adjacent to the AFI seed exhibit an AFI-like structure with no Z5-like formations, whereas the unbiased test results in the emergence of a Z5 structure. The diversity of space groups introduced by the tetrahedral many-body interactions also introduces potential competition between frameworks, particularly those with similar pore diameters, such as in AFI, CFI, and Z5. In our case, during the final 30 generations of optimization, the AFI structure consistently maintains a slightly lower potential energy than Z5, by approximately 0.3 k_BT per particle, allowing for close competition between the two phases.

Table 5. High-Fitness Solution That Spontaneously Self-assembles into the Hole-Type CFI Zeolite within 200 ns

Solution in	T (K)	ε_{TT} (eV)	ε_{TS} (eV)	ε_{SS} (eV)	σ_{TS} (Å)	σ_{SS} (Å)	λ	Fitness
Gen 30	640	0.693	0.070	0.032	6.36	8.05	24.47	0.79

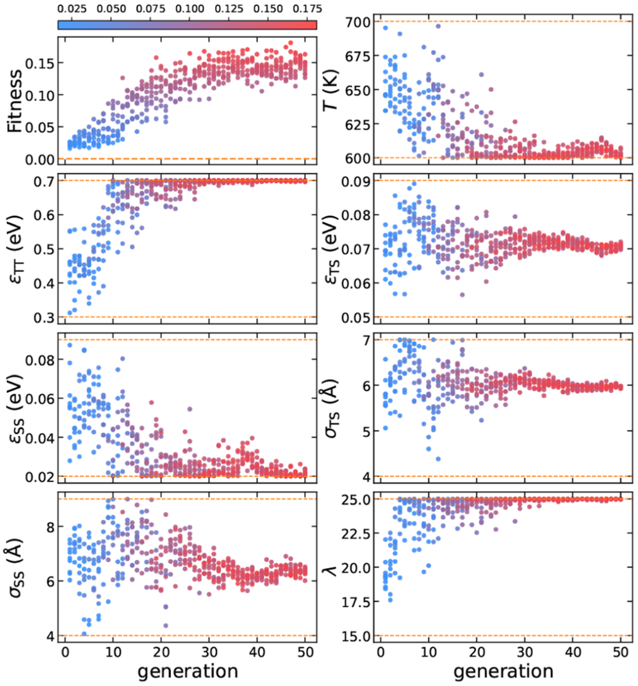


Figure 12. Evolution of the design parameters for the hole-type AFI zeolite framework, with each point colored according to its fitness. The color bar is normalized based on the minimum and maximum values of fitness. The dashed orange lines represent the parameter boundaries.

CONCLUSIONS

In conclusion, we have developed and presented an efficient and robust inverse design workflow capable of identifying optimal interaction parameters and thermodynamic conditions for the self-assembly of complex crystal structures. In this method, the design parameters are iteratively optimized using the covariance matrix adaptation evolution strategy, and the seed-pinning technique is implemented using steered molecular dynamics simulations to accelerate the sampling of nucleation events of the target structure. To evaluate the fitness of a sample, the growth of the target structure is monitored using the environment similarity order parameter with respect to the target phase. In this work, we have applied this inverse design workflow to facilitate the self-assembly of target zeolite frameworks within a coarse-grained model for silica and a structure-directing agent. Using this method, we have successfully optimized the design parameters for the spontaneous self-assembly of a known framework-type Z1 and a known cage-type SGT zeolite. Interestingly, the interaction parameters and thermodynamic conditions differ from those identified in previous work.[24,27−29] Additionally, we reproduced a cage-type SOD and a hole-type CFI zeolite as obtained from the IZA database. Remarkably, we also discovered the sI clathrate structure and a zeolite framework not cataloged in the IZA database during the inverse design of the SOD and AFI zeolite, respectively. Thus, the unbiased self-assembly simulations can occasionally yield polymorphs distinct from the targeted frameworks in seed-pinning simulations.

To summarize, our methodology not only enables the screening of synthesis protocols but also facilitates the discovery of undiscovered, hypothetical zeolites. More specifically, our inverse design protocol may give insights in the required coarse-grained interaction parameters of the structure directing agent, i.e., the size and attraction strength of the organic cation, to favor the self-assembly of a specific zeolite. Using a coarse-to-fine-grained mapping approach such as a generative adversarial network[44] the precise molecular details of the structure directing agent may be obtained. This will be explored in future work. Similarly, a subsequent extension of our framework could explore nonspherical shapes, e.g., rods, for the coarse-grained S particles.

The accelerating role of seed-pinning is essential, as earlier versions of our method without enhanced sampling could only achieve successful self-assembly of Z1, but no other frameworks, within similar computational budgets. Additionally, we note that selecting the shape and size of seeds for different frameworks involves some subjectivity, and the value of κ requires fine-tuning to ensure that the seeds do not melt. In this study, we fixed the composition, which is reasonable for cage-like frameworks. However, for pore structures, a strict number of S particles in a unit cell is not necessary. Given that zeolite synthesis involves careful control of SDA concentrations[45] including composition as a design parameter would be a logical extension. Our self-assembly simulations also indicate that higher fitness values do not consistently yield better performance; successful self-assembly can still occur with solutions that have lower fitness values. Additionally, the optimization process is inherently stochastic, so identical parameters and initial conditions may still lead to different final solutions.

Finally, our inverse design framework is inherently adaptable to other self-assembling systems such as other complex open crystals, metal−organic frameworks, quasicrystals, and liquid crystals. Many of these systems exhibit multiple competing self-assembly pathways, which our framework can address by tracking both target phases and byproducts. The key requirements for implementing our methodology are accurate structural data to define environment similarities and a computationally efficient particle interaction model for enhanced sampling MD. As structural data becomes increasingly accessible for a wide range of molecules and materials, the impact and applicability of our framework will grow. These contributions underscore the robustness and versatility of our approach, paving the way for future research in self-assembling materials and enabling the development of theoretical models and practical applications.

METHODS

In this section, we describe the key elements of our inverse design workflow. First, we present the coarse-grained T-S model used to determine particle interactions and compute forces for running molecular dynamics (MD) simulations. This model is parametrized by the range and strength of interactions, as well as the strength of tetrahedrality and temperature. Second, we

Table 6. High-Fitness Solution That Spontaneously Self-assembles into the Z5 Framework within 200 ns

Solution in	T (K)	ε_{TT} (eV)	ε_{TS} (eV)	ε_{SS} (eV)	σ_{TS} (Å)	σ_{SS} (Å)	λ	Fitness
Gen 19	604	0.699	0.070	0.022	5.95	6.82	24.93	0.16

explain how these model parameters are iteratively optimized using the covariance matrix adaptation evolution strategy (CMA-ES). Since CMA-ES requires a fitness function to guide the optimization, we show how we measure proximity to a target zeolite structure using the environment similarity order parameter.[34] To further speed-up the inverse-design pipeline, we introduce a seed-pinning technique that accelerates the sampling of nucleation events, allowing us to rapidly assess fluctuations in the environment similarity order parameter-based fitness. Finally, we provide specific details on the simulation parameters and protocol.

Zeolite Model. The coarse-grained zeolite model proposed by Molinero *et al.* is a binary mixture consisting of T and S particle types. The T particles, which exhibit tetrahedral interactions, represent the tetrahedrally coordinated atoms, such as silicon (Si), aluminum (Al), or other heteroatoms. The S particles, on the other hand, represent the structure-directing agent (SDA), typically an organic cation. The interactions between these particles are described by the Stillinger-Weber (SW) potential[26] and the potential energy U of the system reads

$$U = \sum_{i<j} \phi^{(2)}_{\alpha\beta}(r_{ij}) + \sum_{i,j\neq i,k>j,k\neq i} \phi^{(3)}_{\mathrm{TT}}(r_{ij}, r_{ik}, \theta_{ijk}) \tag{1}$$

where $r_{ij} = |\mathbf{r}_{ij}|$ is the distance between particle i and j with $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, θ_{ijk} is the angle between $\mathbf{r}_{ij}$ and $\mathbf{r}_{ik}$, and $\phi^{(2)}_{\alpha\beta}$ represents the two-body interaction applied to both T and S species

$$\phi^{(2)}_{\alpha\beta}(r_{ij}) = A\varepsilon_{\alpha\beta}\left[B\left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^p - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^q\right]\varphi_{\alpha\beta}(r_{ij}) \tag{2}$$

where $\alpha\beta$ denotes the particle pair types TT, TS, or SS, and where

$$\varphi_{\alpha\beta}(r_{ij}) = \exp\left(\frac{\sigma_{\alpha\beta}}{r_{ij} - a\sigma_{\alpha\beta}}\right) \tag{3}$$

The three-body interaction $\phi^{(3)}_{\mathrm{TT}}$ applied to solely the T particle species is denoted by

$$\phi^{(3)}_{\mathrm{TT}}(r_{ij}, r_{ik}, \theta_{ijk}) = \lambda\varepsilon_{\mathrm{TT}}(\cos\theta_{ijk} - \cos\theta_0)^2\psi(r_{ij})\psi(r_{ik}) \tag{4}$$

with

$$\psi(r_{ij}) = \exp\left(\frac{\gamma\sigma_{\mathrm{TT}}}{r_{ij} - a\sigma_{\mathrm{TT}}}\right) \tag{5}$$

The three-body interaction term favors the formation of a tetrahedral structure when $\theta_0 = 109.47°$. The exponential terms φ and ψ ensure that both the potential and its derivatives smoothly approach zero at a cutoff distance of $r = a\sigma$. The tetrahedrality parameter λ determines the strength of the tetrahedral interaction. In our work, the values of A, B, p, q, γ, and a are consistent with those in the original SW potential[26] shown in Table 7.

To achieve an average T-T bond length that matches closer to the Si−Si distance in zeolites and also aligns with Molinero's recent works[36,46] we use the value of $\sigma_{\mathrm{TT}} = 2.7275$ Å. The values of $\varepsilon_{\mathrm{TT}}$ and λ are related to the specific elements and composition of the framework components, meaning that the framework particles can exhibit varying bonding strength or tetrahedral character depending on the Si/Al ratio or the presence of other heteroatoms.[40] Similarly, $\varepsilon_{\mathrm{SS}}$ and σ_{SS} depend on the SDA. The values of $\varepsilon_{\mathrm{TS}}$ and σ_{TS} reflect the interaction strength and range between the framework components and the SDA. The parameters related to the S particles can be tuned by modifying the chemistry of the SDA in experiments. Therefore, the design parameters include $\varepsilon_{\mathrm{TT}}$, $\varepsilon_{\mathrm{TS}}$, $\varepsilon_{\mathrm{SS}}$, σ_{TS}, σ_{SS}, λ, and temperature T.

Table 7. Interaction Parameters of the SW Potential[26] Described by $\phi^{(2)}_{\alpha\beta}$ in eq 2 and $\phi^{(3)}_{\mathrm{TT}}$ in eq 4

Parameter	Value	Parameter	Value
A	7.049556277	B	0.6022245584
p	4	q	0
γ	1.2	a	1.8

Inverse Design Workflow. To optimize the design parameter values of the coarse-grained T-S model for the self-assembly of a target zeolite, we implement an iterative process described in this section. Our inverse design workflow involves three key steps: (i) selecting a set of design parameter values from the CMA-ES optimizer (see Methods section on Covariance Matrix Adaptation Evolutionary Strategy); (ii) initiating an MD simulation in the disordered fluid phase with these parameters, using the seed-pinning method to accelerate the nucleation and growth of the target phase (see Methods section on Seed-Pinning via Steered Molecular Dynamics); and (iii) calculating the fitness value from the resulting MD trajectory—specifically, the environment similarity order parameter relative to the target zeolite—and feeding this information back to the optimizer to refine the design parameter values for the next iteration. In this study, we choose as design parameters, temperature T, interaction strengths between different species $\varepsilon_{\mathrm{TT}}$, $\varepsilon_{\mathrm{TS}}$, $\varepsilon_{\mathrm{SS}}$, interaction ranges between different species σ_{TS}, σ_{SS}, and the strength of tetrahedrality λ. The following subsections provide an in-depth explanation of these three steps, as schematically illustrated in Figure 13.

Covariance Matrix Adaptation Evolutionary Strategy. In this work, we use the covariance matrix adaptation evolution strategy (CMA-ES) to optimize the fitness value, which

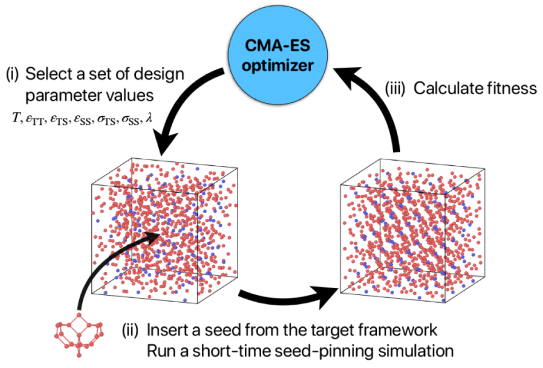


Figure 13. Schematic of the inverse design method. The workflow consists of the following steps: (i) selecting a set of design parameter values using the CMA-ES optimizer; (ii) inserting a seed of the target zeolite into a fluid phase and initiating a short-time molecular dynamics (MD) simulation, using the seed-pinning method; (iii) determining the fitness value by monitoring the environment similarity with respect to the target zeolite during the MD simulations. This fitness value is then fed back into the CMA-ES optimizer, which refines the design parameter values for the next iteration.

measures how closely the system resembles the target phase. CMA-ES is a well-established, population-based, gradient-free stochastic optimization algorithm for real-valued, nonconvex, and nonlinear functions.[31] The method operates without requiring gradient information and has been applied in inverse design studies.[16,17,40]

CMA-ES is an iterative algorithm that samples from a multivariate Gaussian distribution and adapts the mean vector and covariance matrix at each iteration to reach the optimal solution. In each iteration, often referred to as a generation, the algorithm draws n samples from a d-dimensional multivariate Gaussian distribution, where d represents the number of design parameters. The fitness function is then evaluated for these samples, and the outcomes are ranked in descending order. The top k samples are selected as the best candidates and form the set $\boldsymbol{X}$. Based on these best candidates, the algorithm updates the mean vector $\boldsymbol{m} \in \mathbf{R}^d$ and the covariance matrix $\mathbf{C} \in \mathbf{R}^{d\times d}$ to adjust the Gaussian distribution toward the optimal solution. This iterative process continues until a convergence criterion is met. Below is a brief description of each step in an iteration.

(1) A population of $n = 4 + \lfloor 3 \ln d \rfloor$ random samples[47] is drawn from a normal distribution, where d represents the dimension of the parameter space, i.e., the number of design parameters. Each sample or individual is generated as $\boldsymbol{x}_i^{(g+1)} \sim \mathcal{N}\left(\boldsymbol{m}^{(g)}, (\sigma^{(g)})^2 \mathbf{C}^{(g)}\right)$, where $\boldsymbol{m}^{(g)}$ is the mean vector, $\sigma^{(g)}$ is the step size, and $\mathbf{C}^{(g)}$ is the covariance matrix at generation g.

(2) Each sample $\boldsymbol{x}_i^{(g+1)}$ is evaluated using the fitness function $f\left(\boldsymbol{x}_i^{(g+1)}\right)$, and the samples are ranked according to their fitness values.

(3) The mean vector $\boldsymbol{m}^{(g+1)}$ is then updated as a weighted sum of the top k samples with the highest fitness

$$\boldsymbol{m}^{(g+1)} = \boldsymbol{m}^{(g)} + c_{\mathrm{m}} \sum_{i=1}^{k} w_i \left(\boldsymbol{x}_{i:n}^{(g+1)} - \boldsymbol{m}^{(g)}\right) \tag{6}$$

where $\boldsymbol{x}_{i:n}^{(g+1)}$ denotes the i-th ranked sample, w_i is the respective weight, and c_{m} is a learning rate, typically set to 1.

(4) The covariance matrix $\mathbf{C}^{(g)}$ is updated to adapt the shape of the distribution

$$\begin{aligned} \boldsymbol{C}^{(g+1)} &= (1 - c_1 - c_k \sum_{i=1}^{n} w_i)\boldsymbol{C}^{(g)} \\ &+ c_1\, \boldsymbol{p}_c^{(g+1)} {\boldsymbol{p}_c^{(g+1)}}^{\top} \\ &+ c_k \sum_{i=1}^{n} w_i \boldsymbol{y}_{i:n}^{(g+1)} {\boldsymbol{y}_{i:n}^{(g+1)}}^{\top}, \end{aligned} \tag{7}$$

where c_1 and c_k are learning rates for the covariance matrix adaptation, $\boldsymbol{y}_{i:n}^{(g+1)} = \left(\boldsymbol{x}_{i:n}^{(g+1)} - \boldsymbol{m}^{(g)}\right)/\sigma^{(g)}$, and $\boldsymbol{p}_c^{(g+1)}$ is the evolution path given by

$$\boldsymbol{p}_c^{(g+1)} = (1 - c_c)\boldsymbol{p}_c^{(g)} + \sqrt{c_c(2 - c_c)\mu_{\mathrm{eff}}}\, \frac{\boldsymbol{m}^{(g+1)} - \boldsymbol{m}^{(g)}}{c_m \sigma^{(g)}} \tag{8}$$

with $\mu_{\mathrm{eff}} = 1/\sum_{i=1}^{k} w_i^2$, and c_c the decay rate for the cumulation of the path.

(5) The step size $\sigma^{(g)}$ is adjusted based on the evolution path $\boldsymbol{p}_\sigma^{(g+1)}$ to control the overall scale of the search

$$\sigma^{(g+1)} = \sigma^{(g)} \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\left\|\boldsymbol{p}_\sigma^{(g+1)}\right\|}{\mathrm{E}\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|} - 1\right)\right) \tag{9}$$

where c_σ and d_σ are learning rates for the step-size adaptation, and $\mathrm{E}\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|$ represents the expected length of a random vector sampled from a standard normal distribution.

This process is repeated iteratively until a stopping criterion is met, such as reaching the maximum number of generations, achieving a target fitness value, or observing stagnation in the optimization progress. Please refer to ref 48 for a detailed description of this algorithm.

In this work, since there are seven design parameters, the initial population size is $9 = 4 + \lfloor 3 \ln 7 \rfloor$. The design parameters are normalized based on their upper and lower bounds. As initial guess, we set the values of the vector $\boldsymbol{m}^{(0)}$ in the middle between the upper and lower bounds for each design parameter. The initial step size $\sigma_{(0)}$ is set to 0.16, such that 99.7% of the samples lie within a range of $3\sigma \approx 0.5$ from the center of the normalized design space, providing wide coverage in the initial generation. We performed this algorithm using a Python implementation[49] and we use its default setting for the rest of the parameters. While CMA-ES can be used for global optimization by restarting with different initial guesses and population sizes, we exclude this from the scope of our current study due to the high computational cost. A global optimization of the design parameters, accompanied by full free-energy calculations of the nucleation process, could be performed as a refinement step following the implementation of our strategy.

Seed-Pinning via Steered Molecular Dynamics. In this section, we present the enhanced sampling methods used to accelerate nucleation and growth of the zeolite in the MD simulations. Nucleation is typically a rare event within affordable time scales in MD simulations. To better capture nucleation events within limited simulation times, various enhanced sampling methods have been developed, e.g., umbrella sampling[50] forward flux sampling[51] seeding[32,33] and variational umbrella seeding[52] among others. The seeding approach involves inserting a small seed or crystalline cluster into the fluid medium. This crystal seed helps to overcome the high free-energy barrier associated with nucleation. This method is straightforward to implement and can be used to calculate nucleation rates in accordance with classical nucleation theory.[53−55] However, in an inverse design setting, where the simulation parameters are initially far from optimal for zeolite self-assembly, the seed will often melt in the vast majority of simulations, providing little information about the fitness. To overcome this challenge, we propose a variant of the seeding approach called the seed-pinning method. Similar to the traditional seeding approach, a crystal seed is placed in the amorphous phase. The difference is that, similar to ref 56, a restraining potential is applied to the seed size to prevent the seed from melting while allowing growth. This method of allowing only seed growth significantly reduces the nucleation barrier. Samples with favorable design parameters will continue to grow, while those with unfavorable design parameters will

Table 8. Simulation Set-Up and Design Parameter Boundaries for Each Framework[a]

Framework	N_U	N_E	N_S	N_{seed}	κ	χ_T	T (K)	ε_{TT} (eV)	ε_{TS} (eV)	ε_{SS} (eV)	σ_{TS} (Å)	σ_{SS} (Å)	λ
Z1	132	136*	1056	50	5	0.74	[600, 700]	[0.3, 0.7]	[0.05, 0.09]	[0.02, 0.09]	[4, 7]	[4, 8]	[15, 25]
SGT	64	32	1024	23	20	0.94	[600, 700]	[0.3, 0.7]	[0.05, 0.09]	[0.02, 0.09]	[4, 7]	[4, 8]	[15, 25]
SOD	12	6	768	23	20	0.86	[600, 700]	[0.3, 0.7]	[0.05, 0.09]	[0.02, 0.09]	[4, 7]	[6, 10]	[15, 25]
AFI	24	24	1920	192	10	0.92	[600, 700]	[0.3, 0.7]	[0.05, 0.09]	[0.02, 0.09]	[4, 7]	[4, 9]	[15, 25]
CFI	32	16	2048	256	5	0.94	[600, 700]	[0.3, 0.7]	[0.05, 0.09]	[0.01, 0.05]	[4, 7]	[4, 9]	[15, 25]

[a] Here, N_U represents the number of T particles in a unit cell, N_E denotes the number of unique environments, and N_S is the total number of T particles within the seed-pinning simulation box. N_{seed} indicates the seed size, κ denotes the dimensionless constant used in eq 12, T the temperature, ε_{TT}, ε_{TS}, and ε_{SS} the interaction strength between TT, TS, and SS particle pairs, respectively, and σ_{TS} and σ_{SS} the interaction scale between TS and SS particle species, respectively. *: the number of environments for Z1 is determined from scaled FIR-30 unit cells, which contain a larger number of particles.

fluctuate around the restrained nucleus size. By monitoring these fluctuations the fitness of the sample can be evaluated. In the following, we will discuss how to measure these fluctuations and apply the restraint to the seed size.

Due to the intricate structure of zeolite frameworks, traditional methods like Steinhardt bond-orientational order parameters[57] are insufficient for distinguishing zeolite-like T particles from amorphous T particles. Some level of distinction can be achieved by calculating bond-orientational order parameters for T particles with specific coordination numbers[24] but this is not straightforward to generalize across all framework types. To solve this problem, we employ an order parameter called environment similarity (ES)[34] to identify zeolite-like T particles. To evaluate the ES, an environment is defined as the closest n neighbors of a given particle. ES compares the current environment, χ, around a particle i to a reference environment, χ_0, using a normalized kernel $\tilde{k}$ defined as

$$\tilde{k}_{\chi_0}(\chi) = \frac{1}{n}\sum_{i\in\chi}\sum_{j\in\chi_0}\exp\left(-\frac{|\mathbf{r}_i - \mathbf{r}_j^0|^2}{4\sigma^2}\right) \tag{10}$$

where $\mathbf{r}_i$ are the coordinates of the particles, σ is a broadening parameter to account for thermal fluctuations, and n is the number of particles in the environment χ. By definition, $\tilde{k}_{\chi_0}(\chi_0) = 1$. In general, a particle can have multiple distinct environments in a crystal of a specific lattice, denoted as $\chi_1, \chi_2, ..., \chi_{N_E}$, where N_E denotes the number of possible environments. In this case, we adopt a best-match strategy to determine the environment, with the kernel defined as the maximum similarity among the N_E reference environments

$$\tilde{k}_X(\chi) = \frac{1}{\eta}\log\left(\sum_{l=1}^{N_E}\exp(\eta\tilde{k}_{\chi_l}(\chi))\right) \tag{11}$$

The largest value of $\tilde{k}_{\chi_l}(\chi)$ among the environments $\chi_l \in X$ is selected as the ES when η is sufficiently large. Therefore, the ES order parameter ranges between 0 and 1.

The data for the unique environments is extracted from the IZA[58] and CDCC databases[59] using the Environment Finder tool.[60] In general, we set a cutoff of 10 Å to search for unique environments, which is sufficient to capture the structural features needed for accurate zeolite identification. The number of distinct environments N_E are listed in Table 8 and vary from 6 for SOD to 136 for Z1. The number of particles n in an environment typically varies from 55 to 75. Considering the notable fluctuations at high temperatures, the value of σ is set to 0.5 Å.

It is worth noting that this order parameter is not rotationally invariant. This means that even if the identified particles have the same structure as the reference environment, a high ES value will not be obtained if the orientation is inconsistent. As a result, our protocol is limited to zeolites that are aligned with the chosen references. However, we address this by starting with an aligned seed and applying restraints to it to preserve that alignment. To further restrain the seed size, we bias the number of particles N' with an ES order parameter greater than a threshold, $\tilde{k}_\chi(\chi) > k'$, using a rational switching function to ensure that N' is continuous and differentiable.

First, we insert a seed composed solely of T particles, extracted from the target zeolite framework, into the amorphous mixture system. For all cases, a seed is cut to include as many T-T bonds as possible while keeping the pores open. We then apply a harmonic potential to the seed particles to restrain the ES order parameter[50]

$$\beta V(t) = \frac{1}{2}\kappa(N'_{\rm seed}(t) - N_{\rm seed})^2 \tag{12}$$

where κ is a dimensionless constant, $N_{\rm seed}$ is the initial number of T particles in the seed, and $N'_{\rm seed}$ is the number of T particles from the original seed that retains an ES order parameter above a given threshold, k'. This bias is only applied to the seed particles that originally formed the seed, which has two advantages: first, the T particles that are not originally in the seed are free to either remain in the amorphous phase or nucleate, and second, the computational cost for calculating the ES and its gradient is invested only on a reduced number of T particles. For the fitness calculation, the ES is evaluated for all particles, but this quantity does not need to be computed at every time step. While it is in principle possible to calculate the nucleation free energy ΔG using seed-pinning, our goal is to enhance sampling around the nucleation barrier and use the fluctuations to assess the fitness for a given set of design parameters. Once a high fitness is achieved, the optimal design parameters are tested in a self-assembly simulation without any seed or bias.

Simulation Protocol. In this study, we categorize all zeolite frameworks into three types based on channel dimensionality: cage-type (channel dimensionality = 0), where S particles are confined within cages and cannot move freely; hole-type (channel dimensionality ≥ 1), where S particles are mobile; and framework-type, which includes both cages and holes (channels). This workflow was first applied to reproduce a framework-type Z1, and a cage-type SGT[37] zeolite, as documented in previous studies.[24,27−29] The Z1 zeolite structure is based on the structure of the MOF FIR-30[35] see for more details section on Z1. Subsequently, we focused on reproducing a cage-type SOD[38] and hole-type CFI[41] and AFI obtained from the IZA database.

For cage-type and framework-type zeolites, the number of S particles per cage is estimated based on geometric constraints, typically assuming one S particle per cage. A supercell is then constructed for each zeolite using unit cells from the IZA database, typically containing about a thousand T particles to balance computational efficiency and finite-size effects. Next, a seed is extracted from the supercell. The pseudospherical seed is selected based on two key principles: first, it must remain open rather than forming a closed cage, allowing the passage of S and T particles; second, its shape and size are tailored to the target zeolite framework. Table 8 presents the specific settings for each framework. While larger seeds may promote crystal growth in MD simulations without bias, crystallization often does not occur in self-assembly tests using the same design parameters due to a high nucleation barrier ΔG. This is expected, as a smaller seed more closely resembles unbiased nucleation conditions from a purely fluid phase. We have also tested even smaller seeds, but in those cases, no nucleation events occurred during our short MD runs. In general, the seed size is initially set to contain approximately a quarter of the T particles in the target unit cell and is increased only if nucleation does not occur during evolution. We do not systematically search for the critical nucleus size due to computational constraints when designing multiple different phases. Therefore, the seed should be kept as small as possible while still effectively inducing nucleation. The seed is then embedded in a simulation box with dimensions identical to the supercell, and with additional T and S particles randomly placed according to the target composition. Subsequently, a one-nanosecond MD simulation is performed in the isobaric−isothermal (NPT) ensemble at a pressure of p = 0 bar and a time step of 5 fs. The pressure is controlled isotropically in all three Cartesian coordinates, with the masses of both T and S particles set to 28.0855 u, equivalent to the mass of silicon. The average fraction of T particles with an environment similarity (ES) $\tilde{k}_{\chi}(\chi) > 0.5$ during the last quarter of the simulation trajectory is used as the fitness value for the optimizer, which then initiates the next iteration. This approach allows the fitness to be interpreted as the fraction of T particles in a target-like environment within the system. Fluctuations around the pinned seed size-whether positive or negative-indicate the nucleation free-energy gradient, providing insight into how favorable nucleation is from the small seed, thereby providing a valid fitness estimation. While the fitness estimation remains noisy during short evolution runs due to the stochastic nature of nucleation, extending the sampling time would significantly increase computational costs. Since CMA-ES is robust to noise, we opt not to extend the sampling time.

After 50 rounds of MD simulations, high-fitness solutions are tested in self-assembly simulations using larger system sizes, without the use of any seeds or biases, to assess their effectiveness. In these tests, the pressure is controlled independently along each of the three Cartesian coordinates. These unbiased self-assembly tests are essential. Given the stochastic nature of the nucleation event from the pinned seed, there is no minimum fitness value that can guarantee that nucleation will occur in an unbiased run, or that it will reach the same target phase. Therefore, in our pipeline, a high fitness parameter combination found during evolution is always validated by an unbiased test.

All simulations are conducted using LAMMPS (version 2 Aug 2023)[61] with biasing implemented through PLUMED (version 2.8.3).[62,63] Snapshots are rendered using the OVITO software.[64]

ASSOCIATED CONTENT

Data Availability Statement

A GitHub repository of this project is available at https://github.com/MarjoleinDijkstraGroupUU/zeolites-inverse-design.git.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsnano.4c17597.

> Configuration files of SOD framework, sI clathrate, CFI framework, and Z5 framework; self-assembly movies of SOD framework, sI clathrate, CFI framework, and Z5 framework (ZIP)

AUTHOR INFORMATION

Corresponding Authors

Chaohong Wang − *Soft Condensed Matter & Biophysics, Debye Institute for Nanomaterials Science, Utrecht University, 3584 CC Utrecht, the Netherlands;* orcid.org/0000-0003-4533-5464; Email: c.wang2@uu.nl

Alberto Pérez de Alba Ortíz − *Soft Condensed Matter & Biophysics, Debye Institute for Nanomaterials Science, Utrecht University, 3584 CC Utrecht, the Netherlands; Computational Soft Matter, van't Hoff Institute for Molecular Sciences and Informatics Institute, University of Amsterdam, Amsterdam 1098 XH, the Netherlands;* orcid.org/0000-0003-4776-5521; Email: a.perezdealbaortiz@uva.nl

Marjolein Dijkstra − *Soft Condensed Matter & Biophysics, Debye Institute for Nanomaterials Science, Utrecht University, 3584 CC Utrecht, the Netherlands;* orcid.org/0000-0002-9166-6478; Email: m.dijkstra@uu.nl

Complete contact information is available at: https://pubs.acs.org/10.1021/acsnano.4c17597

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.W., A.P.d.A.O., and M.D. gratefully acknowledge SURF (www.surf.nl) for the computational resources, the national supercomputer Snellius, and is also grateful for the financial support from the China Scholarship Council (CSC). A.P.d.A.O. and M.D. acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. ERC-2019-ADG 884902, SoftML).

REFERENCES

(1) Weitkamp, J. Zeolites and catalysis. *Solid State Ionics* **2000**, *131*, 175−188.

(2) Corma, A. Inorganic solid acids and their use in acid-catalyzed hydrocarbon reactions. *Chem. Rev.* **1995**, *95*, 559−614.

(3) Corma, A. From microporous to mesoporous molecular sieve materials and their use in catalysis. *Chem. Rev.* **1997**, *97*, 2373−2420.

(4) Sherman, J. D. Synthetic zeolites and other microporous oxide molecular sieves. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3471−3478.

(5) Wang, S.; Peng, Y. Natural zeolites as effective adsorbents in water and wastewater treatment. *Chem. Eng. J.* **2010**, *156*, 11−24.

(6) Serati-Nouri, H.; Jafari, A.; Roshangar, L.; Dadashpour, M.; Pilehvar-Soltanahmadi, Y.; Zarghami, N. Biomedical applications of zeolite-based materials: A review. *Mater. Sci. Eng.* **2020**, *116*, 111225.

(7) Cundy, C. S.; Cox, P. A. The hydrothermal synthesis of zeolites: Precursors, intermediates and reaction mechanism. *Microporous Mesoporous Mater.* **2005**, *82*, 1−78.

(8) Lobo, R. F.; Zones, S. I.; Davis, M. E. Structure-direction in zeolite synthesis. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1995**, *21*, 47−78.

(9) Ch. Baerlocher, B. M.; Brouwer, D.; McCusker, L. Database of Zeolite Structures. 2024; https://www.iza-structure.org/databases/.

(10) Earl, D. J.; Deem, M. W. Toward a database of hypothetical zeolite structures. *Ind. Eng. Chem. Res.* **2006**, *45*, 5449−5454.

(11) Deem, M. W.; Pophale, R.; Cheeseman, P. A.; Earl, D. J. Computational discovery of new zeolite-like materials. *J. Phys. Chem. C* **2009**, *113*, 21353−21360.

(12) Pophale, R.; Cheeseman, P. A.; Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys* **2011**, *13*, 12407−12412.

(13) Torquato, S. Inverse optimization techniques for targeted self-assembly. *Soft Matter* **2009**, *5*, 1157−1173.

(14) Rechtsman, M. C.; Stillinger, F. H.; Torquato, S. Optimized interactions for targeted self-assembly: application to a honeycomb lattice. *Phys. Rev. Lett.* **2005**, *95*, 228301.

(15) Marcotte, É.; Stillinger, F. H.; Torquato, S. Communication: Designed diamond ground state via optimized isotropic monotonic pair potentials. *J. Chem. Phys.*, **2013**, *138*, 6. .

(16) Coli, G. M.; Boattini, E.; Filion, L.; Dijkstra, M. Inverse design of soft materials via a deep learning−based evolutionary strategy. *Sci. Adv.* **2022**, *8* (3), No. eabj6731.

(17) Bedolla-Montiel, E. A.; Lange, J. T.; Pérez de Alba Ortíz, A.; Dijkstra, M. Inverse design of crystals and quasicrystals in a non-additive binary mixture of hard disks. *J. Chem. Phys.*, **2024**, *160*, 24. .

(18) Kim, B.; Lee, S.; Kim, J. Inverse design of porous materials using artificial neural networks. *Sci. Adv.* **2020**, *6* (1), No. eaax9324.

(19) Zheng, X.; Chen, T.-T.; Jiang, X.; Naito, M.; Watanabe, I. Deep-learning-based inverse design of three-dimensional architected cellular materials with the target porosity and stiffness using voxelized Voronoi lattices. *Sci. Technol. Adv. Mater.* **2023**, *24* (1), 2157682.

(20) Yao, Z.; Sánchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O. K.; Snurr, R. Q. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nat. Mach. Intell* **2021**, *3*, 76−86.

(21) Park, J.; Gill, A. P. S.; Moosavi, S. M.; Kim, J. Inverse design of porous materials: a diffusion model approach. *J. Mater. Chem. A* **2024**, *12*, 6507−6514.

(22) Gandhi, A.; Hasan, M. F. Machine learning for the design and discovery of zeolites and porous crystalline materials. *Curr. Opin. Chem. Eng.* **2022**, *35*, 100739.

(23) Piñeros, W. D.; Truskett, T. M. Designing pairwise interactions that stabilize open crystals: Truncated square and truncated hexagonal lattices. *J. Chem. Phys.*, **2017**, *146*. .

(24) Kumar, A.; Molinero, V. Two-Step to One-Step Nucleation of a Zeolite through a Metastable Gyroid Mesophase. *J. Phys. Chem. Lett.* **2018**, *9*, 5692−5697.

(25) Molinero, V.; Moore, E. B. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B* **2009**, *113*, 4008−4016.

(26) Stillinger, F. H.; Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B* **1985**, *31*, 5262.

(27) Wang, F.; Fu, H.-R.; Zhang, J. Homochiral Metal−Organic Framework with Intrinsic Chiral Topology and Helical Channels. *Cryst. Growth Des.* **2015**, *15*, 1568−1571.

(28) Dhabal, D.; Bertolazzo, A. A.; Molinero, V. Coarse-Grained Model for the Hydrothermal Synthesis of Zeolites. *J. Phys. Chem. C* **2021**, *125*, 26857−26868.

(29) Kumar, A.; Nguyen, A. H.; Okumu, R.; Shepherd, T. D.; Molinero, V. Could Mesophases Play a Role in the Nucleation and Polymorph Selection of Zeolites? *J. Am. Chem. Soc.* **2018**, *140*, 16071−16086.

(30) Dhabal, D.; Bertolazzo, A. A.; Molinero, V. What Is the Smallest Zeolite That Could Be Synthesized? *Angew. Chem., Int. Ed.* **2022**, *61* (29), No. e202205095.

(31) Hansen, N.; Ostermeier, A. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Comput.* **2001**, *9*, 159−195.

(32) Bai, X.-M.; Li, M. Calculation of Solid-Liquid Interfacial Free Energy: A Classical Nucleation Theory Based Approach. *J. Chem. Phys.* **2006**, *124* (12), 124707.

(33) Espinosa, J. R.; Vega, C.; Valeriani, C.; Sanz, E. Seeding approach to crystal nucleation. *J. Chem. Phys.* **2016**, *144* (3), 034501.

(34) Piaggi, P. M.; Parrinello, M. Calculation of Phase Diagrams in the Multithermal-Multibaric Ensemble. *J. Chem. Phys.* **2019**, *150* (24), 244119.

(35) https://www.ccdc.cam.ac.uk/structures/Search?Compound=FIR-30&DatabaseToSearch=Published.

(36) Bertolazzo, A. A.; Dhabal, D.; Lopes, L. J. S.; Walker, S. K.; Molinero, V. Unstable and Metastable Mesophases Can Assist in the Nucleation of Porous Crystals. *J. Phys. Chem. C* **2022**, *126*, 3776−3786.

(37) https://europe.iza-structure.org/IZA-SC/framework.php?ID=219.

(38) https://europe.iza-structure.org/IZA-SC/framework.php?ID=221.

(39) Noya, E. G.; Zubieta, I.; Pine, D. J.; Sciortino, F. Assembly of Clathrates from Tetrahedral Patchy Colloids with Narrow Patches. *J. Chem. Phys.* **2019**, *151* (9), 094502.

(40) Kumar, A.; Zare, M.; Molinero, V. Assembly of Zeolitic Crystals From a Model of Mesogenic Patchy Nanoparticles. *J. Phys. Chem. C* **2019**, *123*, 971−978.

(41) https://europe.iza-structure.org/IZA-SC/framework.php?ID=62.

(42) https://europe.iza-structure.org/IZA-SC/framework.php?ID=25.

(43) https://europe.iza-structure.org/IZA-SC/framework.php?ID=214.

(44) Stieffenhofer, M.; Wand, M.; Bereau, T. Adversarial reverse mapping of equilibrated condensed-phase molecular structures. *Mach. Learn.: sci. Technol.* **2020**, *1*, 045014.

(45) Gómez-Hortigüela, L.; Camblor, M. A. Introduction to the zeolite structure-directing phenomenon by organic species: general aspects. *Insights Into The Chemistry Of Organic Structure-Directing Agents In The Synthesis Of Zeolitic Materials* **2017**, *175*, 1−41.

(46) Chu-Jon, C.; Bertolazzo, A.; Banik, S.; Sankaranarayanan, S.; Molineroa, V. Interzeolite transformation through cross-nucleation. *Mater. Chem.* **2024**.

(47) Loshchilov, I.; Schoenauer, M.; Sebag, M.; Hansen, N. Maximum likelihood-based online adaptation of hyper-parameters in CMA-ES. *Int. Conf. Parallel Problem Solving Nature* **2014**, *8672*, 70−79.

(48) Hansen, N. The CMA evolution strategy: A tutorial. *arXiv Preprint arXiv: 1604.00772*; arXiv. 2016.

(49) Hansen, N.; Akimoto, Y.; Baudis, P. CMA-ES/pycma on Github. Zenodo. **2019**. .

(50) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(51) Allen, R. J.; Valeriani, C.; ten Wolde, P. R. Forward flux sampling for rare event simulations. *J. Phys.: condens. Matter* **2009**, *21*, 463102.

(52) Gispen, W.; Espinosa, J. R.; Sanz, E.; Vega, C.; Dijkstra, M. Variational umbrella seeding for calculating nucleation barriers. *J. Chem. Phys.*, **2024**, *160*, 17. .

(53) Farkas, L. Keimbildungsgeschwindigkeit in übersättigten Dämpfen. *J. Phys. Chem.* **1927**, *125*, 236−242.

(54) Volmer, M.; Weber, A. Keimbildung in übersättigten Gebilden. *J. Phys. Chem.* **1926**, *119*, 277−301.

(55) Becker, R.; Doring, W. Kinetische Behandlung der Keimbildung in übersättigten Dämpfen. *Annalen der Physik* **1935**, *416*, 719.

(56) Sharma, A. K.; Escobedo, F. A. Nucleus-Size Pinning for Determination of Nucleation Free-Energy Barriers and Nucleus Geometry. *J. Chem. Phys.* **2018**, *148* (18), 184104.

(57) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. Bond-Orientational Order in Liquids and Glasses. *Phys. Rev. B* **1983**, *28*, 784−805.

(58) https://www.iza-structure.org/databases/.

(59) https://www.ccdc.cam.ac.uk/.

(60) https://github.com/PabloPiaggi/EnvironmentFinder/tree/master.

(61) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.

(62) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.

(63) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.

(64) Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO−the Open Visualization Tool. *Modelling Simul. Mater. Sci. Eng.* **2010**, *18* (1), 015012.

NOTE ADDED AFTER ASAP PUBLICATION

Changes were made to Table 5 after this paper was published ASAP on May 1, 2025. The corrected version was reposted on May 13, 2025.