Predicting polymeric crystal structures by evolutionary algorithms

Qiang Zhu,[1,a),b)] Vinit Sharma,[2,a),c)] Artem R. Oganov,[1,3,4] and Ramamurthy Ramprasad[2]

[1]*Department of Geosciences, Stony Brook University, Center for Materials by Design, Institute for Advanced Computational Science, Stony Brook, New York 11794, USA*
[2]*Materials Science and Engineering, Institute of Materials Science, University of Connecticut, Storrs, Connecticut 06626, USA*
[3]*Department of Problems of Physics and Energetics, Moscow Institute of Physics and Technology, 9 Institutskiy Lane, Dolgoprudny City, Moscow Region 141700, Russia*
[4]*Department of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China*

The recently developed evolutionary algorithm USPEX proved to be a tool that enables accurate and reliable prediction of structures. Here we extend this method to predict the crystal structure of polymers by constrained evolutionary search, where each monomeric unit is treated as a building block with fixed connectivity. This greatly reduces the search space and allows the initial structure generation with different sequences and packings of these blocks. The new constrained evolutionary algorithm is successfully tested and validated on a diverse range of experimentally known polymers, namely, polyethylene, polyacetylene, poly(glycolic acid), poly(vinyl chloride), poly(oxymethylene), poly(phenylene oxide), and poly (p-phenylene sulfide). By fixing the orientation of polymeric chains, this method can be further extended to predict the structures of complex linear polymers, such as all polymorphs of poly(vinylidene fluoride), nylon-6 and cellulose. The excellent agreement between predicted crystal structures and experimentally known structures assures a major role of this approach in the efficient design of the future polymeric materials. © *2014 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4897337]

I. INTRODUCTION

Recent methodological developments have made it possible to predict crystal structures from the knowledge of chemical species, compositions, or the molecular geometries.[1–5] Especially for organic crystal structure predictions (CSP), the recent blind tests organized by the Cambridge Crystallographic Data Centre show a steady progress in this field.[6–10] There are many global energy minimization methods which have been successfully applied.[3,11] However, extension and application of such methods to polymers is complicated and challenging, as one has to look for a possible solution within the constraint of a given polymeric topology, while unconstrained crystal structure prediction is difficult in large cells since the number of free variables increases rapidly with the number of atoms that are included.

The ability to predict the physical and chemical properties of polymers from their molecular structure can give structure prediction great value for the design of polymers for numerous technological applications such as capacitive energy storage, transistors and photovoltaic devices.[12–17] Polymer molecules can adopt multiple chain conformations and also multiple packings of same conformations, for which a method to predict the crystal structures is the key to understand and rationalize the fundamental properties of polymeric materials.

The most popular way is to conjecture some initial configurations based on the packing of polymer chains, followed by energy minimization.[18] Alternatively, some automated algorithms were used to assemble the polymeric chains by proper mathematical treatments, without information about the crystal class, cell parameters, and space group.[19–21] Most of the previous attempts were based on either classical molecular mechanics or molecular dynamics, requiring accurate force fields. However, such classical force fields are usually insufficient to describe the potential energy surface. If one has to evaluate total energy calculation at *ab initio* level when sampling the configuration space, an efficient structural searching method is crucial - here we focus on formulating such efficient approaches.

In our recent work, we proved that evolutionary algorithms (EA) can efficiently predict the crystal packing from only a knowledge of molecular geometry.[22] To the best of our knowledge, so far no attempt has been made to predict the structure of crystalline polymers that possess maximal stability or optimal physical properties, at an *ab initio* level. Here, we describe a specifically designed constrained evolutionary algorithm, implemented in the USPEX code,[23] and employing well-defined molecular units or motifs. The high efficiency of this approach enables prediction of polymeric crystal structures using quantum-mechanical computations. The power of our scheme, combining this efficient EA-based global optimization technique with accurate energy calculations to search for stable polymeric crystals, is demonstrated by the successful identification of various experimentally known polymers.

a)Q. Zhu and V. Sharma contributed equally to this work.
b)Electronic mail: qiang.zhu@stonybrook.edu
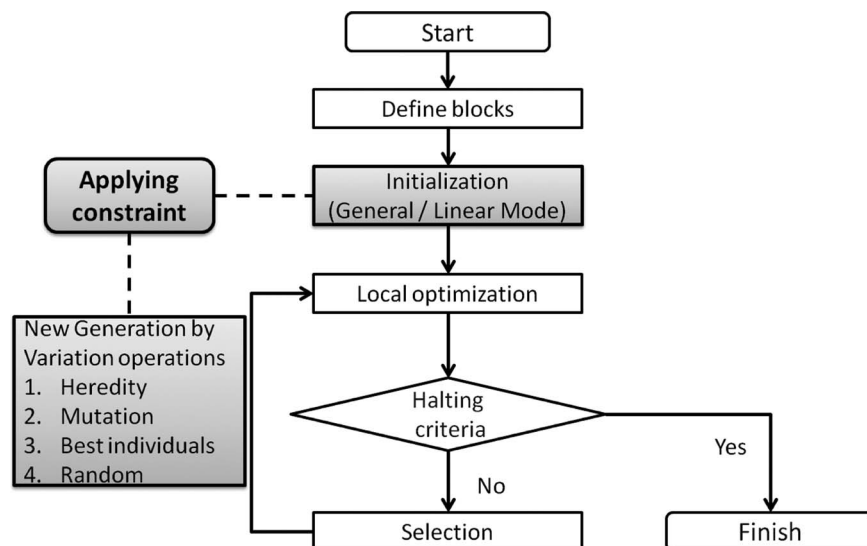c)Electronic mail: vinit.sharma.mlsu@gmail.com

FIG. 1. Flowchart of the constrained evolutionary algorithm. The key feature is that all the generated structures before local optimization are constructed based on the pre-specified building blocks (highlighted in gray box).

II. EVOLUTIONARY ALGORITHMS

The USPEX code has been successfully applied to various classes of systems (bulk crystals,[2,23] nanoclusters,[24] 2D crystals,[25] and surfaces[26]). Extending the range of its applicability, we proposed a new *constrained global optimization* method to predict the packing of molecular crystals.[22] A similar concept can be applied to polymers as well, with further developments that we describe below.

If we start to search for the global energy minimum with randomly generated structures (according to the given chemical formula), it is very likely that most of the time will be spent on exploring many disordered structures characterized by irrelevant structural motifs. More importantly, the desired polymeric crystals are usually not the thermodynamic equilibrium in the given chemical system. The truly interesting searching target is actually the optimum sequence of monomers, and 3D-packing of the pre-formed polymeric chains. This problem can be solved by *constrained global optimization*—finding the most stable packing of monomers with fixed bond connectivity. It requires whole motifs rather than individual atoms to be considered as the minimum building blocks in our search. This strategy does not only make the global optimization meaningful, but at the same time simplifies it, leading to a drastic reduction of the number of variables in the search space.

In the context of EA, our procedure is as shown in Fig. 1.

(1) **Define blocks**. Since the monomeric units can be made of one or multiple types of blocks, we represent them by using Z-matrix, which has been widely used to describe the molecular structure in organic chemistry. For the atom in a given molecule, its bond connectivity can be defined by internal coordinates (i.e., the bond length, bond angle, and torsional angle). As shown in Fig. 2, Cartesian coordinates can be transformed to Z-matrix representation according to the constraints by those internal coordinates. In Z-matrix, the top three atoms lack six constraints, since there are no reference atoms to define their internal coordinates. The 6 missing components in the Z-matrix correspond the 3 translational and 3 rotational variables in 3D space. From now on, we treat each block as a rigid body, and construct the crystal structures by varying only those 6 variables in each Z-matrix. It is essential here that the monomeric units used here are chemically unsaturated, i.e., prone to spontaneously forming polymers.

(2) **Initialization**. At the beginning of EA, the random symmetric structures for the first generation are randomly produced. Note that different techniques of generating symmetric crystal structures have been widely used in organic CSP community[4–10] since Dzyabchenko's early work on benzene,[27] in the hope of obtaining the target structures directly from random generation or after relaxation. But here we choose to use symmetry also for other purposes. A fully random initialization is a poor choice for large systems, as it always leads to nearly identical glassy structures that have similar (high) energies and low degree of order.[2,24] From such
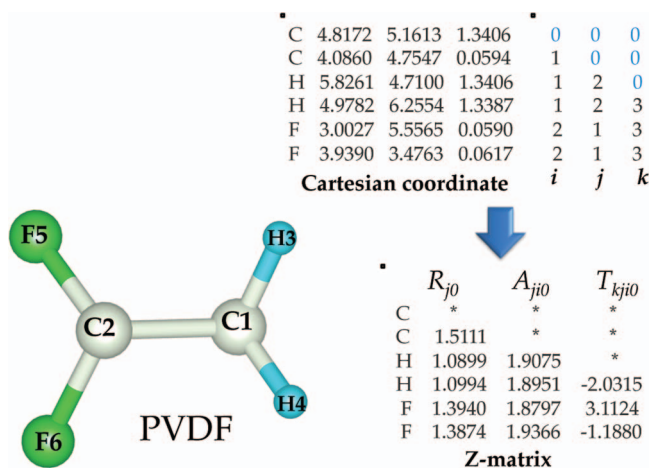

FIG. 2. Transformation from Cartesian coordinates to Z-matrix representation. The $i - j - k$ table specifies the topological relations in the Z-matrix style. R is distance from atom 0 to atom number i, A is the angle made by the present atom with atoms j and i, while T is the torsion angle made by the present atom with atoms k, j, and i.
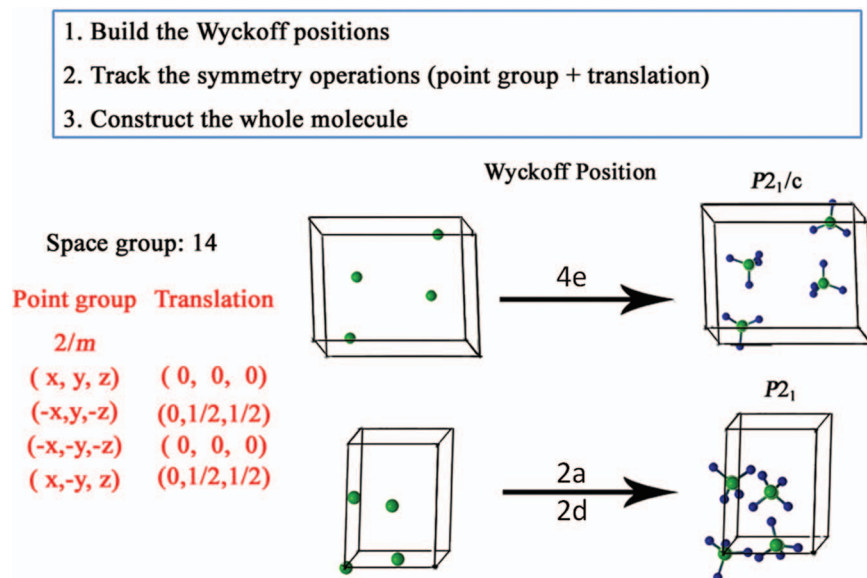
FIG. 3. Illustration of generating a random symmetric structure with 4 molecules per cell. For a given space group randomly assigned by the program (in this case, $P2_1/c$), the Bravais cell is generated, and molecular center is placed onto a random position (in this case, the general position 4e or 2a+2d). Molecules are then built at the Wyckoff sites according to their connectivity and with their orientations obeying space group symmetry operations. If special Wyckoff sites are occupied, molecular geometry often breaks space group symmetry, leading to a subgroup. For clarity of the figure, molecules occupying positions at the corners and faces of the unit cell are shown only once.

starting conditions, it is difficult to obtain ordered crystalline states. To achieve both high structural diversity and quality, a better way is to create symmetric structures for the initial population.[22, 24]

If the polymer consists of multiple types of blocks and their arrangement is unclear, we need to navigate all the possibilities. We treat each block independently and generate structures with each block randomly located in the unit cell, just as we did for the prediction of molecular crystals.[22] We first generate a random symmetric structure with the geometric centers of each block being located at general or special Wyckoff positions in a 3D primitive lattice, by using a special random symmetric algorithm.[24] For each Wyckoff site, the first molecular block (R) is built around it with random orientation, and the replica blocks (R') can be obtained by symmetry operations, which are a combination of point group (P) and translations (T) operations,

$$R' = R \cdot P + T. \tag{1}$$

This whole scheme is illustrated in Fig. 3, which exactly works for molecules occupying the general Wyckoff site. If special Wyckoff sites are involved, the generated structures are likely to have lower symmetry: if the molecular block itself has low symmetry, there will be symmetry breaking, leading to a subgroup symmetry, and we allow this.

(3) **Local optimization.** Structural relaxation is done stepwise from low to high precision, as described in Ref. 23, to achieve greater efficiency.

(4) **Selection.** At the end of each generation, all relaxed structures in the generation are compared using their fingerprints[22] and all non-identical structures are ranked by their energies.[28]

(5) **Variation operators.** Child structures (new generation) are produced from parent structures (old generation)

using two general types of variation operators: heredity and several kinds of mutations.[23] Heredity, mating two different parent structures, establishes communications between good candidate solutions. Mutations are aimed at introducing new features into the population and probing the neighborhood of the already found low-energy structures by strongly perturbing them. Different from those in atomic crystals, these variation operators act on the geometric centers of the molecules and their orientations. It is very important to note that symmetric initialization scheme might favor only symmetric structures. Yet, our variation operators can break symmetry, and this ensures the even asymmetric ground states will not be missed.

Heredity: This is a basic variation operator in EA. It cuts planar slices from two selected individuals and combines them to produce a child structure. During this process, each monomer is represented by its geometric center and orientation (Fig. 4(a)), and the entirety of all monomers is retained during the operation of heredity (as well as of all the other variation operators).

Rotational mutation: A certain number of randomly selected monomers are rotated by random angles. Different from the previous implementation of purely random rotation,[22] the rotational axes are determined by the eigendecomposition of the inertia tensor matrix of the given molecule (Fig. 4(b)).

$$[I] = [Q][\Delta][Q^T]. \tag{2}$$

The inertia tensor can be computed by

$$I = \sum_{i=1}^{N} r_i^2, \tag{3}$$

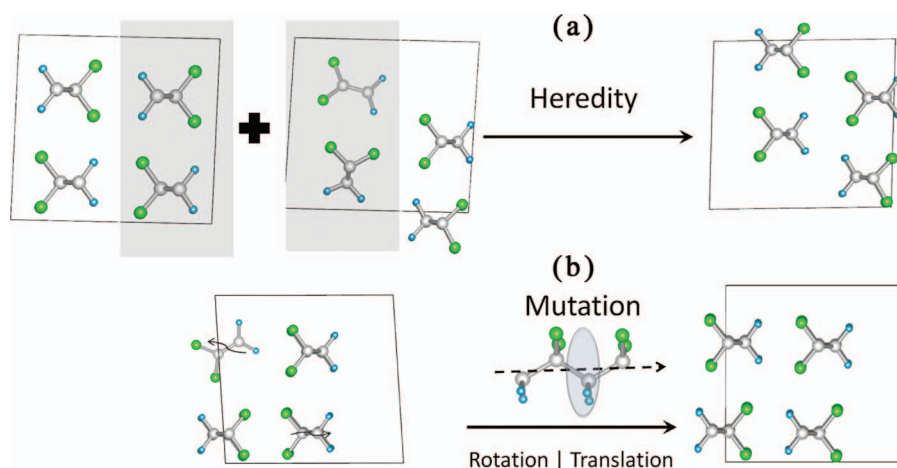where r is the corresponding distance from each atom to the geometric center.

FIG. 4. Variation operators in EA. (a) Heredity; (b) mutation.

The columns of the rotation matrix [Q] define the directions of the principal axes of the body, and [Δ] is a diagonal matrix,

$$[\Delta] = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix}, \tag{4}$$

where I_1, I_2, and I_3 are called principal moments of inertia, determining which direction in [Q] is easier to rotate. For the polymeric chain, I_1 is usually significantly larger than I_2 and I_3. And the first direction in [Q] nearly coincides with the direction of the polymers chain.

Translational mutation: All the centers of monomers are displaced in random directions, the displacement magnitude for molecule i being picked from a zero-mean Gaussian distribution with σ defined as

$$\sigma_i = \sigma_{\max} \frac{\Pi_{\max} - \Pi_i}{\Pi_{\max} - \Pi_{\min}}, \tag{5}$$

where Π is the local degree of order of the molecule[22] and $\sigma_{\max}$ is of the order of a typical intermolecular distance. We calculate the Π of each molecule's geometric center from its fingerprint.[29] Thus molecules with more ordered environment are perturbed less than molecules with less ordered environment.

Softmutation: This operator involves atomic displacements along the softest mode eigenvectors, or a random linear combination of softest eigenvectors.[30] For molecular crystals, it becomes a hybrid operator, combining rotational and coordinate mutations. In this case, the eigenvectors are calculated first, and then projected onto translational and rotational degrees of freedom of each molecule and the resulting changes of molecular positions and orientations are applied preserving rigidity of the fixed intra-molecular degrees of freedom. For the rapid calculation of vibrational modes, here we use the approach of Ref. 30.

Addition of random structures: Although the search space has been effectively decreased by applying geometric constraints, we are still facing a high dimensional configuration space. A general challenge for EA (and many other global optimization methods) is how to avoid getting stuck in a local minimum when dealing with multidimensional search spaces—in other words, avoiding decrease of population diversity during evolution. A key to maintain the population diversity is to add "new blood." Therefore, we produce some fraction of each generation using the random symmetric algorithm described above. Fingerprint niching also helps to retain the diversity.[30]

The percentages of structures obtained by different variation operators shall be pre-specified by the user, and can be adjusted during the calculation. From our experience, a good setting should be 50%–70% for heredity, 10%–20% for translational or softmutation, 10%–20% for rotational mutation, and 15%–30% for random structures, respectively. In all the tests demonstrated below, we choose a population size of 30, 50% for heredity, 10% for translational/softmutation, 10% for rotational mutation, and 30% for random structures.

(6) **Halting and post-processing**. When the lowest-energy structure remains unchanged for a certain number of generations, the calculation automatically stops and the lowest-energy structures found in USPEX are then carefully relaxed with higher precision: the all-electron projector-augmented wave (PAW) method,[31] as implemented in the VASP code,[32] at the level of generalized gradient approximation (GGA-PBE functional).[33] van der Waals (vdW) dispersion and hydrogen bond are crucial factors in determining the crystal packing of the crystalline polymers. There have been major efforts in improving the accuracy of vdW functional, for example, see Ref. 34. Our previous work proved that Tkatchenko-Scheffler method[35] gives results in satisfactory agreement with experimental data.[36] Therefore, we employed PBE-TS functional as implemented in VASP[37] for all calculations. We used the plane wave kinetic energy cutoff of 550 eV and the Brillouin zone was sampled with a resolution of $2\pi \times 0.07$ Å^{-1}, which showed excellent convergence of the energy differences, stress tensors, and structural parameters.

III. PREDICTION OF THE CRYSTAL STRUCTURES OF SIMPLE POLYMERS

The new constrained EA in combination of state-of-the-art quantum-mechanical computational methods
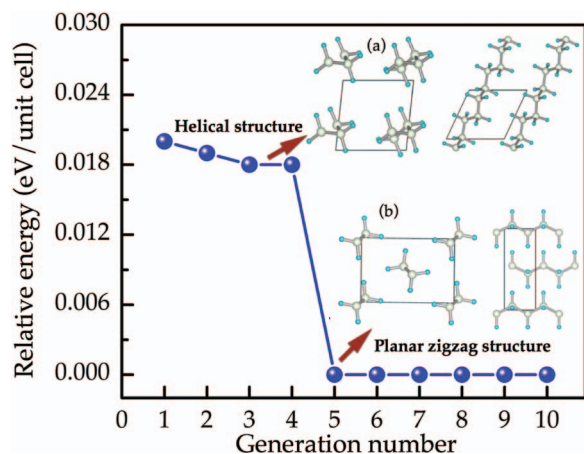
FIG. 5. The evolution of lowest-energy structure as a function of generation number for PE. The predicted structures (top and side view) of (a) the helical structure identified in a distinct local minimum for PE and (b) planar zigzag structure are also shown. Gray and cyan spheres represent the carbon and hydrogen atoms, respectively.
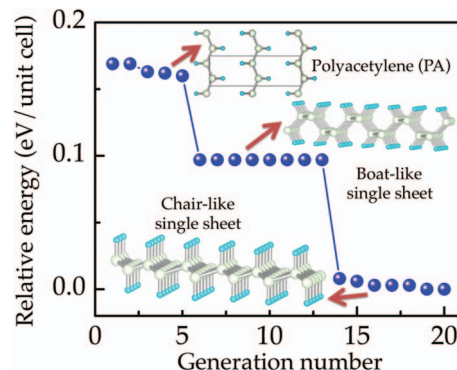

FIG. 6. The evolution of lowest-energy structure as a function of generation number for PA with repeating unit $[-C_2H_2-]_n$. Along with experimentally known structure, boat-like and chair-like structures of graphane are also shown. Gray and cyan spheres represent the carbon and hydrogen atoms, respectively.

successfully predict the crystal structures, lattice parameters and densities for several polymers composed of simple monomeric units, namely polyethylene (PE), polyacetylene (PA), poly(glutamic acid) (PGA), poly(vinyl chloride) (PVC), poly(oxymethylene) (POM), poly(phenylene oxide) (PPO), and poly (p-phenylene sulfide) (PPS). We note that while past computational work assumed experimentally known structural parameters and space group, no such assumptions were made in the present study. Symmetry and other structural parameters are the outcome of the search process. As the new constrained EA significantly reduces the search space, the correct structures of these known polymers are found in the first few generations. It is noteworthy to point out that the number of formula units (f.u.) is a pre-specified parameter in the current implementation. Since the aim of this paper is to benchmark our method by comparing with the experimental data, we simply take the number of formula units from the known results. Ideally, one needs to try many different numbers, which is computationally more expensive. Below we discuss these systems and their structures.

PE: We start with the simplest polymer, polyethylene ($[-CH_2-CH_2-]_n$). In Fig. 5, the evolution of lowest-energy structure as a function of generation number for PE is plotted. The carbon backbone of the equilibrium structure has a planar all-trans zigzag structure. Its space group (*Pnma*), equilibrium geometry and density were found to be in agreement with available experimental measurements and previous *ab initio* results.[36,38,39] It is evident from Fig. 5 that the new constrained evolutionary search scheme explored the configurational space in such an efficient manner that our algorithm is capable of finding the meta-stable phases in first few generations (<5). Our EA based structural search also identifies that in addition to the global minima, the total energy also has a distinct local minimum for the helical structure as shown in Fig. 5, which is 0.005 eV/f.u. higher in energy than the ground state.

PA: The 3D geometry of the crystalline PA polymer with the repeat unit $[-C_2H_2-]_n$ has been a matter of debate. To resolve this issue, attempts have been made by various experiments[40] as well as computations.[41,42] The two proposed structures with space groups $P2_1/a$ and $P2_1/n$ have a slight difference in the orientation of double bonds of adjacent chains. In the case of $P2_1/a$, the double bonds of adjacent chains are in-phase while in the case of $P2_1/n$, they are out-of-phase. Along the chain axis, translating alternate chains of the $P2_1/a$ structure by $c/2$ results in the $P2_1/n$ structure. Our calculations predict that the structure where double bonds of adjacent chains are in-phase ($P2_1/a$) is more stable. The evolution of lowest-energy structure as a function of generation number for PA is shown in Fig. 6. Calculated lattice parameters are in agreement with experimental values (Table I). Moreover, it is worth mentioning that in this search we also obtained isomeric two-dimensional sheets of graphane with CH stoichiometry (shown in Fig. 6), which are even more stable than both benzene (C_6H_6) and PA ($C_2H_2)_n$.[40] The boat-like and chair-like structures were also identified with a previous version of the USPEX method.[42]

We also studied a set of other polymers (including PGA, PVC, POM, PPO, PPS). The predicted crystal structures, lattice parameters and densities for all polymers considered are shown in Table I. It can clearly be seen that PBE-TS significantly improves the agreement with available experimental data, compared with PBE without any vdW correction. The PBE-TS results obtained in our search, are slightly different from our previous study,[36] indicating a flat energy landscape of polymers. The deviations are mainly from the lattice vectors in non-fibre axis which are sensitive to numerical noise and description of vdW dispersions. In general, the comparison from Table I proves that (1) TS-vdW correction allows one to reproduce the experiment lattice parameters very well, and thus is sufficient to describe inter-chain interactions; (2) our EA approach is very efficient for predicting polymeric structures and their crystal packing. Moreover, such searches from small molecular building blocks can yield a comprehensive picture of the energy landscape (as shown in the examples of PA). Furthermore, the search not only identifies the ground state configuration, as low energy metastable phases can also be observed in this type of search.[44] With this encouragement, we proceed to the cases of more complex polymers.

TABLE I. The predicted lattice parameters and density for the considered polymers, namely, PE, PA, PPO, POM, PPS, PGA, and PVC. For comparison available experimental and computational results at the same level of theory are also listed. All considered polymers have four f.u. per unit cell ($Z = 4$), except PA has $Z = 2$.

Polymer	Method	a(Å)	b(Å)	c(Å)	β(°)	Density (g/cm^3)
PE	Expt.[39]	7.12	4.85	2.55		0.997
$[-CH_2-]_n$	PBE[36]	8.20	5.60	2.55		0.796
(Pnma)	PBE-TS[36]	7.01	4.76	2.56		1.091
	PBE-TS[a]	7.02	4.76	2.56		1.090
PA	Expt.[39]	4.24	7.32	2.46	91-94	1.130
$[-C_2H_2-]_n$	PBE[36]	5.00	7.74	2.46	90.3	0.908
($P2_1/n$)	PBE-TS[36]	4.01	7.19	2.46	90.3	1.219
	PBE-TS[a]	4.00	7.22	2.46	90.6	1.217
PGA	Expt.[39]	5.22	6.19	7.02		1.700
$[-C_2H_2O_2-]_n$	PBE[36]	5.07	5.58	6.96		1.958
(Pcmn)	PBE-TS[36]	5.09	6.11	7.03		1.763
	PBE-TS[a]	5.13	6.09	7.01		1.761
PVC	Expt.[39]	10.24	5.24	5.08		1.523
$[-CH_2-CHCl-]_n$	PBE[36]	10.45	5.50	5.05		1.430
(Pbcm)	PBE-TS[36]	10.11	5.15	5.08		1.540
	PBE-TS[a]	10.14	5.16	5.08		1.562
POM	Expt.[39,43]	4.77	7.65	3.56		1.536
$[-CH_2-O-]_n$	PBE[36]	5.40	8.37	3.63		1.216
($P2_12_12_1$)	PBE-TS[36]	4.59	7.72	3.57		1.577
	PBE-TS[a]	4.55	7.75	3.59		1.576
PPO	Expt.[39]	8.07	5.54	9.72		1.408
$[-C_6H_4O-]_n$	PBE[36]	8.42	5.88	9.85		1.254
(Pbcn)	PBE-TS[36]	8.04	5.37	9.75		1.453
	PBE-TS[a]	8.02	5.36	9.75		1.460
PPS	Expt.[39]	8.67	5.61	10.26		1.440
$[-C_6H_4S-]_n$	PBE[36]	8.85	5.73	10.26		1.381
(Pbcn)	PBE-TS[36]	8.48	5.54	10.25		1.492
	PBE-TS[a]	8.48	5.53	10.26		1.494

[a]This work.

IV. PREDICTION OF THE CRYSTAL STRUCTURES OF COMPLEX LINEAR POLYMERS

So far, we have demonstrated a general framework to predict crystal structures of polymers, from only the building blocks. The prediction still needs to sample a considerably large configurational space, including the connectivity sequence, the conformational diversity, and the manner in which the chains pack together. This method, although very powerful in predicting polymers made of simple monomers, is very likely to face hurdles for complex polymeric systems. Considering that most of the existing crystalline polymers are composed of neatly packed straight chains, let us focus on linear polymers. The conformation of the chain is the primary interest in polymer chemistry. Therefore, we can simplify the searching problem by starting from the conformation of an individual chain, and predicting the optimal packing of such chains. Note that similar techniques have been used in the early work of polymeric crystal structure predictions.[19–21]

Provided that the chain conformation is known, the factors of defining their packing in the crystal are (1) the rela-

(a) Anti-parallel (twisted chain, γ-nylon 6)

(b) Parallel (twisted chain)

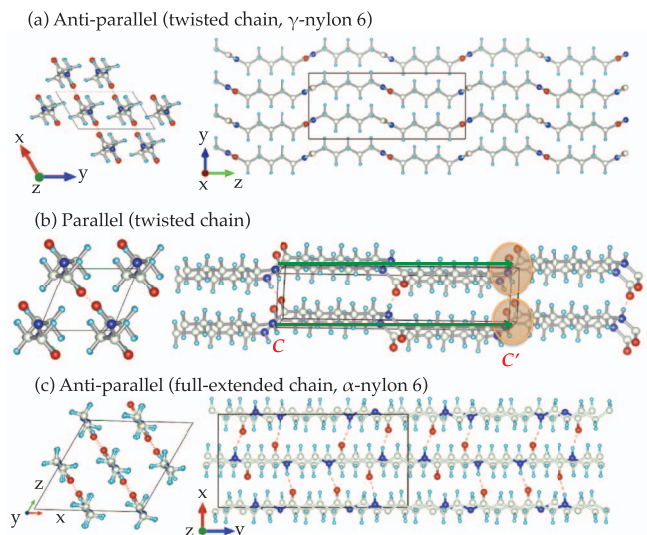(c) Anti-parallel (full-extended chain, α-nylon 6)

FIG. 7. Crystal structures of nylon 6: (a) γ phase with the twisted chains; (b) another low-energy configuration with the twisted chain; (c) α phase with the full extended chain. Note that (a) and (b) differ in the direction of two adjacent H-bonded sheets: anti-parallel in (a), and parallel in (b). Structure initialization in the context of linear chain mode is also represented. C and C' are the geometric centers of monomers. The monomers are assembled in such a way that the $C-C'$ connections are parallel (or anti-parallel) to the c-axis of the cell. The unsaturated connecting groups (CO− and −NH) are marked. The degrees of freedom for each monomer include the position of its geometric center (C) and its rotation along c-axis in ab-plane. Gray, cyan, red, and blue spheres represent the carbon, hydrogen, oxygen, and nitrogen atoms, respectively.

tive positions of the chains in the crystal; (2) the rotational degrees of freedom associated with the lateral groups: (3) the orientation of the chains. In this case, another constraint can be imposed: all the infinite chains in a crystalline structure must be parallel or antiparallel to each other. Therefore, we propose a new structure prediction scheme of *Linear chain mode*, where we assume that the polymeric chain runs parallel to the crystallographic c-axis. Here, we assemble the monomers by ensuring the neighboring contacts of these bridging atoms are close to the real situation (in terms of bond length and bond angle). Mathematically, the propagation orientation can be determined by the vector between the geometric centers of two connected monomers, $C − C'$, as shown in Fig. 7, where structure initialization of nylon-6 in the context of linear chain mode is demonstrated. Thus we can reorient the linear chain in the (001) or (00$\bar{1}$) direction. To predict the crystal structures, initially we create a 2D primitive cell in the $a − b$ plane for the geometric centers, according to the randomly assigned plane group symmetry. Then the c-axis is defined by the length of the chain, and the reoriented monomers are constructed around the geometric centers in the 3D unit cell. During the course of new structure generation the chain orientations are randomly assigned, allowing the freedom of parallel and anti-parallel packings. To enrich structural diversity, a certain degree of variation from the rotation and translation of polymers along the c-axis is permitted. Accordingly, the rotational axis is fixed to the c direction when rotational mutations are performed.

By imposing the above constraints, the linear chain mode significantly speeds up the searching process. Here we

illustrate its power by the prediction of all poly(vinylidene fluoride) (PVDF) polymorphs, and two other well-known complex polymers, nylon-6 and cellulose.

PVDF: It is composed of polar $[-CH_2-CF_2-]_n$ repeating units. The molecular chain can be assembled in different conformations, depending on the trans (T) or gauche (G) linkages. The variations of chain conformation and arrangements of dipole moments lead to polymorphism. So far, four different known phases have been well characterized experimentally; α (TGTG'), β (TTTT), γ (TTTGTTTG'), and δ (TGTG').[47,48] The *general mode*, efficiently predicts the β phase with all-T chains, but fails to obtain other known polymorphs with complex chain conformations. For instance, the γ phase is has a more complex chain conformation and relatively high energy, although it has been synthesized and exists as a metastable state. The chance of finding this phase in a "predictive EA" is very low, while a wise idea is to confine the search by starting from a well-defined chain, instead of the basic monomeric unit. Here, we perform several searches using the new scheme of *linear chain mode*, by starting from a well-defined chain in different conformations, namely, TT, TGTG', and TTTGTTTG'. For the TT chain, we found the β phase, which is well known for its piezoelectric properties and is a prototype of a family of piezoelectric materials. Interestingly, two other low-energy phases (S1 and S2 as shown in Fig. 8) are also observed, which differ from the β-phase in the orientation of the dipole moments. In S1 the orientations of the dipoles moments are antiparallel, while a noncollinear orientation of neighboring chain dipoles is observed in S2. The monoclinic α-PVDF and orthorhombic δ-PVDF are successfully identified as well in the search starting from TGTG' chain. In α-PVDF, the dipole moments are antiparallel and mutually cancelled, while all dipole moments are oriented in the same direction in δ-PVDF. Therefore, α is a non-polar phase, while δ is polar. In both phases, the adjacent chains are antiparallel. Interestingly, we also noticed an energetically competitive configuration in TGTG' conformation with two adjacent chains being parallel (denoted as S3 in Fig. 8). Starting from the TTTGTTTG' chain, we observed that the most stable configuration (S4) consists of two parallel adjacent chains with dipoles arranged in the same direction. The experimentally known γ-phase was also identified as a metastable phase in this search. Among the discovered structures, δ-PVDF has the lowest energy. Although the energy differences between different structures are small, only a few phases have been observed experimentally. This indicates either entropic stabilization of the observed phases, or their kinetic preference. The predicted lattice parameters for all known PVDF polymorphs are summarized in Table II.[49] Clearly, our predictions based on PBE-TS functional are in satisfactory agreement with experiment.

Nylon 6: Two crystalline forms of nylon 6 have been experimentally characterized, α and γ. There is a substantial confusion regarding the structure of the α phase.[50] The earliest reported crystal structure[51] had some incorrect atomic coordinates in Cambridge structural database (CSD entry: LILSUU). Here we used the model suggested in the previous theoretical studies, which is described by the packing of the full-extended chains, possessing eight monomeric units of $[-(CH_2)_5-CO-NH-]$ per unit cell. The γ-phase has a simpler structure ($Z = 4$) based on twisted chains.

We have performed a search with $Z = 8$ starting from the full-extended chain, and $Z = 4$ starting from the twisted chain, in the hope of finding the α and γ-phase.[52] Indeed, we found that the most stable configuration has a monoclinic symmetry for $Z = 4$ (space group $P2_1/c$, $a = 4.77$ Å, $b = 8.35$ Å, $c = 16.88$ Å (fiber axis), $\gamma = 121.2°$, in good agreement with experimental results,[53] except that there is a considerable deviation in cell vector b (the direction between the
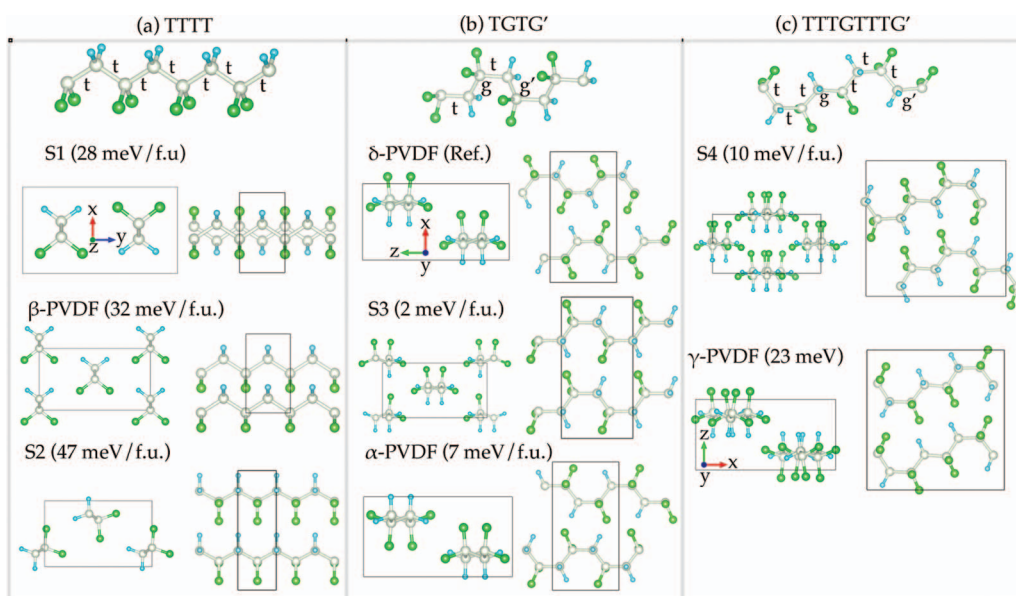

FIG. 8. Top and side views of low-energy crystal structures of PVDF found in USPEX searches starting from (a) TTTT chain; (b) TGTG' chain; (c)TTTGTTTG' chain. The energetics relative to the ground state (δ-PVDF) are also shown. Gray, cyan, and green spheres represent the carbon, hydrogen and fluorine atoms, respectively.

TABLE II. The predicted lattice parameters and density for all polymorphs of PVDF. For comparison available experimental and computations results at the same level of theory are also listed.

Polymer	Method	a(Å)	b(Å)	c(Å)	β(°)	Density (g/cm^3)
α	Expt.[39]	9.64	4.96	4.62	90.0	1.92
$Z = 4$	PBE[a]	9.83	5.07	4.68	90.0	1.82
($P2_1/c$)	PBE-TS[a]	9.32	4.83	4.67	90.0	2.02
β	Expt.[39]	8.58	4.91	2.56		1.97
$Z = 2$	PBE[a]	8.75	4.91	2.58		1.91
($Cm2m$)	PBE-TS[a]	8.67	4.81	2.59		2.12
γ	Expt.[45]	4.97	9.18	9.66		1.93
$Z = 8$	PBE[a]	5.01	9.32	9.81		1.85
($Pca2_1$)	PBE-TS[a]	4.81	9.29	9.55		1.99
δ	Expt.[46]	4.96	9.64	4.62		1.93
$Z = 4$	PBE[a]	5.04	10.01	4.68		1.80
($Pna2_1$)	PBE-TS[a]	4.83	9.21	4.63		2.04

[a]This work.

alternating sheets with purely vdW bonding) as shown in Fig. 7. We also note that the discrepancy in the cell parameters can be significantly reduced when the optPBE-vdW functional is used, as shown in Table III. Compared with the previous study based on a classical force field,[54] our results show a significant improvement. While different functionals show somewhat different cell parameters, the structural topology does correspond to the experimental γ-phase. In the γ phase, the antiparallel twisted chains form pleated sheets via hydrogen bonds, and the chain directions are opposite in alternating sheets. Interestingly, we also observed another extremely low-energy configuration in which the corresponding chains in adjacent sheets are parallel; this phase is only 3 meV/f.u. higher in energy than the γ-phase.[49]

In our $Z = 8$ search, we found the ground state which is very similar to what has been described in the literature.[54]

Similarly, this structure features nylon sheets joined by H bonds in the antiparallel way. Moreover, the predicted lattice parameters are close to the experimental results.[51] Again, the optPBE-vdW functional appears to give better lattice parameters than PBE-TS, but the energetics are similar. The α phase is calculated to be -28.8 meV/f.u. lower in energy than γ, which is also consistent with the fact that α is the thermodynamically most stable crystalline form.[54]

Cellulose: Cellulose is a polymer with repeating D-glucose units $[-C_6H_{10}O_5-]_n$. Microfibrils of naturally occurring cellulose correspond to two crystalline forms, I_α and I_β. I_α has a triclinic unit cell and crystallizes in $P1$ space group. Since it contains only one disaccharide unit with two D-glucose units in the unit cell, the prediction is much simpler from the mathematical viewpoint (here one just needs to consider the lattice parameters). Indeed, it can be trivially found by random sampling together with lattice mutation described in our previous work.[23] The optimized cell parameters are listed in Table III.

Let us focus on the more challenging case of I_β. It was found that I_β is the thermodynamically more stable phase. Although its crystal structure has been intensively studied,[55,56] the crystallographic coordinates were only recently reliably determined by low-temperature neutron diffraction techniques.[57] A computational study[58] showed that PBE-TS yields a remarkable agreement with the experimental reports on cellulose-I_β, and hence we use the PBE-TS approach in our search. Starting from the D-glucopyranosyl chains ($Z = 4$), we indeed identified I_β as the ground state configuration, and the calculated unit cell parameters ($a = 7.36$ Å, $b = 8.16$ Å, $c = 10.44$ Å, $\gamma = 96.4°$) agree well with previous reports[57,58] (see Table III). As shown in Fig. 9(a), cellulose chains are arranged parallel-up and edge to edge, making flat sheets that are held together by H-bonds. Sheets formed by H-bonded D-glucopyranosyl chains are in the bc-plane, while there are no strong H-bonds perpendicular to the sheets. Most importantly, the complex

TABLE III. The comparison of experimental and theoretical cell parameters for nylon-6 and cellulose.

|	Method	a (Å)	b (Å)	c (Å)	α(°)	β(°)	γ(°)	Density(g/cm^3)
α nylon 6	Expt.295 K[51]	9.56	8.01	17.24			67.5	1.23
$[-(CH_2)_5-CO-NH-]_n$	optPBE-vdW[a,59]	9.47	7.34	17.46			68.5	1.34
$Z = 8$	PBE[a]	9.76	8.27	17.49			69.1	1.15
($P2_1$)	PBE-TS[a]	9.80	8.49	17.48			67.5	1.12
γ nylon 6	Expt.295 K[53]	4.78	9.33	16.88			121.0	1.17
$[-(CH_2)_5-CO-NH-]_n$	optPBE-vdW[a,59]	4.76	8.96	16.91			123.5	1.22
$Z = 4$	PBE[a]	4.90	9.98	16.81			120.0	1.06
($P2_1/a$)	PBE-TS[a]	4.77	8.35	16.98			121.2	1.29
Cellulose-I_α	Expt.293 K[60]	6.72	5.96	10.40	118.1	114.8	80.4	1.64
$[-C_6H_{10}O_5-]_n$	PBE-D2(A)[61]	6.56	5.83	10.39	117.5	114.2	81.8	1.70
$Z = 2$	PBE-D2(B)[61]	6.77	5.81	10.30	117.6	113.7	85.0	1.66
($P1$)	PBE-TS[a]	6.84	6.07	10.45	116.8	113.7	78.8	1.54
Cellulose-I_β	Expt.15 K[57]	7.64	8.18	10.37			96.5	1.61
$[-C_6H_{10}O_5-]_n$	PBE-TS[58]	7.63	8.14	10.41			96.4	1.64
$Z = 4$	PBE[37]	8.70	8.23	10.46			95.5	1.41
($P2_1$)	PBE-TS[a]	7.49	8.13	10.42			96.4	1.66

[a]This work.
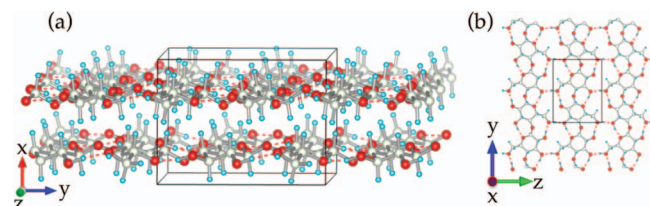
FIG. 9. Crystal structure of cellulose-I$_\beta$ (a) side view (b) top view. Gray, cyan, and red spheres represent the carbon, hydrogen and oxygen atoms, respectively.

hydrogen bond network in the flat sheets is also correctly predicted (Fig. 9(b)).

V. CONCLUSIONS

In summary, we demonstrated and successfully tested a specifically designed constrained evolutionary algorithm, implemented in the USPEX code,[23] employing well-defined molecular units or motifs, successfully predict polymeric crystal structures from first-principles quantum-mechanical computations. This strategy makes the problem well-defined, significantly reduces the search space and improves the efficiency of search. The diversity of the sampling is enhanced by using space group symmetry combined with random cell parameters, random Wycoff positions and orientations of molecular units. There are two schemes in our approaches. In the general mode, the structures are assembled from the unsaturated monomeric units, while the remaining variables (i.e., the connectivity of the monomeric units, the conformation of chain, and the packing in 3D spaces) are optimized by the global EA approach in conjunction with *ab initio* structure relaxation. This method has been successfully tested and validated on a diverse range of experimentally known polymers. To predict complex linear polymers, we have proposed a *linear chain* scheme. Unlike previous attempts of polymeric crystal structure prediction within the framework of classical force fields,[19–21] all the systems were investigated by the accurate (and more expensive) energy minimization at the *ab initio* level. Very recently, this method has been successfully applied to predict crystal structures of polymer dielectrics.[17] It should be emphasized that this strategy avoids problems due to inadequacies of the empirical force fields, but also becomes notably demanding at the same time (for instance, the prediction of α nylon-6 is already extremely expensive in this scheme). If one performs *ab initio* calculations, structure perdition will currently be limited at 0 K. While we only discussed global optimization of the energy here, global optimizations of other properties (e.g., density,[62] hardness,[63] dielectric constants,[64] etc.) are readily available as well.

ACKNOWLEDGMENTS

[1]*Modern Methods of Crystal Structure Prediction*, edited by A. R. Oganov (Wiley-VCH, Weinheim, 2010).

[2]A. R. Oganov, A. O. Lyakhov, and M. Valle, Acc. Chem. Res. **44**, 227–237 (2011).

[3]M. A. Neumann, F. J. J. Leusen, and J. Kendrick, Angew. Chem., Int. Ed. **47**, 2427–2430 (2008).

[4]M. U. Schmidt and U. Englert, J. Chem. Soc., Dalton Trans. **1996**, 2077–2082.

[5]C. J. Pickard and R. J. Needs, Nat. Mater. **7**, 775–779 (2008).

[6]J. P. M. Lommerse *et al.*, Acta Cryst. B **56**, 697–714 (2000).

[7]W. D. S. Motherwell *et al.*, Acta Cryst. B **58**, 647–661 (2002).

[8]G. M. Day *et al.*, Acta Cryst. B **61**, 511–527 (2005).

[9]G. M. Day *et al.*, Acta Cryst. B **65**, 107–125 (2009).

[10]D. A. Bardwell *et al.*, Acta Cryst. B **67**, 535–551 (2011).

[11]S. L. Price, Chem. Soc. Rev. **43**, 2098–2111 (2014).

[12]L. Zhu and Q. Wang, Macromolecules **45**, 2937–2954 (2012).

[13]B. Chu, X. Zhou, K. Ren, B. Neese, M. Lin, Q. Wang, F. Bauer, and Q. M. Zhang, Science **313**, 334–336 (2006).

[14]A. Facchetti, Chem. Mater. **23**, 733–758 (2011).

[15]Y.-g. Ha, J. D. Emery, M. J. Bedzyk, H. Usta, A. Facchetti, and T. J. Marks, J. Am. Chem. Soc. **133**, 10239–10250 (2011).

[16]C. Wang, G. Pilania, S. Boggs, S. Kumar, C. Breneman, and R. Ramprasad, Polymer **55**, 979–988 (2014).

[17]V. Sharma *et al.*, Nat. Commun. **5**, 4845 (2014).

[18]N. Karasawa and W. A. I. Goddard, Macromolecules **25**, 7268–7281 (1992).

[19]R. A. Sorensen, W. B. Liau, and R. H. Boyd, Macromolecules **21**, 194–199 (1988).

[20]J. Aerts, Polym. Bull. **36**, 645–652 (1996).

[21]S. León, J. Navas, and C. Alemán, Polymer **40**, 7351–7358 (1999).

[22]Q. Zhu, A. R. Oganov, C. W. Glass, and H. T. Stokes, Acta Cryst. B **68**, 215–226 (2012).

[23]A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).

[24]A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, Comput. Phys. Commun. **184**, 1172–1182 (2013).

[25]X.-F. Zhou, X. Dong, A. R. Oganov, Q. Zhu, Y. Tian, and H.-T. Wang, Phys. Rev. Lett. **112**, 085502 (2014).

[26]Q. Zhu, L. Li, A. R. Oganov, and P. B. Allen, Phys. Rev. B **87**, 195317 (2013).

[27]A. Dzyabchenko, J. Struct. Chem. **25**, 416–420 (1984).

[28]If one can do free energy minimization for each structural relaxation (i.e., local optimization), the global crystal structure prediction at finite temperature is also possible. This strategy can be used when an accurate classical force field is available. However, free energy minimization based on DFT methods is still impractical at the moment. Therefore, we limit our prediction to zero Kelvin.

[29]A. R. Oganov and M. Valle, J. Chem. Phys. **130**, 104504 (2009).

[30]A. O. Lyakhov, A. R. Oganov, and M. Valle, Comput. Phys. Commun. **181**, 1623–1632 (2010).

[31]P. E. Blöchl, Phys. Rev. B **50**, 17953–17979 (1994).

[32]G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169–11186 (1996).

[33]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865–3868 (1996).

[34]J. Klimes and A. Michaelides, J. Chem. Phys. **137**, 120901 (2012).

[35]A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[36]C.-S. Liu, G. Pilania, C.-C. Wang, and R. Ramprasad, J. Phys. Chem. A **116**, 9347–9352 (2012).

[37]T. Bucko, J. Hafner, S. Lebegue, and J. G. Angyan, J. Phys. Chem. A **114**, 11814–11824 (2010).

[38]G. Avitabile, R. Napolitano, B. Pirozzi, K. D. Rouse, M. W. Thomas, and B. T. M. Willis, J. Polym. Sci., Polym. Lett. Ed. **13**, 351–355 (1975).

[39]R. L. Miller and L. E. Nielsen, J. Poly. Sci. **44**, 391–395 (1960).

[40]Q. Zhu, J. E. Fischer, R. Zusok, and S. Roth, Solid State Commun. **83**, 179–183 (1992).

[41]C. M. Zicovich-Wilson, B. Kirtman, B. Civalleri, and A. Ramirez-Solis, Phys. Chem. Chem. Phys. **12**, 3289–3293 (2010).

[42]X.-D. Wen, L. Hand, V. Labet, T. Yang, R. Hoffmann, N. W. Ashcroft, A. R. Oganov, and A. O. Lyakhov, Proc. Natl. Acad. Sci. U.S.A. **108**, 6833–6837 (2011).

[43]G. Carazzolo and M. Mammi, J. Polym. Sci., Part A: Gen. Pap. **1**, 965–983 (1963).

[44]Due to the property of EA to zoom in on the most promising regions of the search space—though there is no guarantee that all low-energy metastable states will be found by the EA.

[45]A. J. Lovinger, Macromolecules **14**, 322–325 (1981).

[46]M. Li, H. J. Wondergem, M.-J. Spijkman, K. Asadi, I. Katsouras, P. W. Blom, and D. M. de Leeuw, Nat. Mater. **12**, 433–438 (2013).

[47]N. J. Ramer and K. A. Stiso, Polymer **46**, 10431–10436 (2005).

[48]N. J. Ramer, T. Marrone, and K. A. Stiso, Polymer **47**, 7160–7165 (2006).

[49]See supplementary material at http://dx.doi.org/10.1063/1.4897337 for all other metastable structures of PVDF and nylon-6 in cif format.

[50]S. Dasgupta, W. B. Hammond, and W. A. Goddard, J. Am. Chem. Soc. **118**, 12291–12301 (1996).

[51]D. R. Holmes, C. W. Bunn, and D. J. Smith, J. Polym. Sci. **17**, 159–177 (1955).

[52]Since α-nylon-6 has a large unit cell, we used soft PAW potential in our search for $Z = 8$ to reduce the computational cost. The top five low-energy structures were selected and further relaxed with normal PAW potential and higher precision, as for the other systems.

[53]H. Arimoto, M. Ishibashi, M. Hirai, and Y. Chatani, J. Polym. Sci., Part A: Gen. Pap. **3**, 317–326 (1965).

[54]Y. Li and W. A. Goddard, Macromolecules **35**, 8440–8455 (2002).

[55]J. Sugiyama, R. Vuong, and H. Chanzy, Macromolecules **24**, 4168–4175 (1991).

[56]R. J. Viëtor, K. Mazeau, M. Lakin, and S. Pérez, Biopolymers **54**, 342–354 (2000).

[57]Y. Nishiyama, G. P. Johnson, A. D. French, V. T. Forsyth, and P. Langan, Biomacromolecules **9**, 3133–3140 (2008).

[58]T. Bučko, S. Lebègue, J. Hafner, and J. G. Ángyán, Phys. Rev. B **87**, 064110 (2013).

[59]J. Klimes, D. R. Bowler, and A. Michaelides, J. Phys: Condens. Matter **22**, 022201 (2010).

[60]Y. Nishiyama, J. Sugiyama, H. Chanzy, and P. Langan, J. Am. Chem. Soc. **125**, 14300–14306 (2003).

[61]T. Bucko, D. Tunega, J. G. Angyan, and J. Hafner, J. Phys. Chem. A **115**, 10097–10105 (2011).

[62]Q. Zhu, A. R. Oganov, M. A. Salvadó, P. Pertierra, and A. O. Lyakhov, Phys. Rev. B **83**, 193410 (2011).

[63]A. O. Lyakhov and A. R. Oganov, Phys. Rev. B **84**, 092103 (2011).

[64]Q. Zeng, A. R. Oganov, A. O. Lyakhov, C. Xie, X. Zhang, J. Zhang, Q. Zhu, B. Wei, I. Grigorenko, L. Zhang *et al.*, Acta Cryst. C **70**, 76–84 (2014).