ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理
Chinese Physics B

The CALYPSO methodology for structure prediction

Qunchao Tong(童群超), Jian Lv(吕健), Pengyue Gao(高朋越), Yanchao Wang(王彦超)

Citation: Chin. Phys. B . 2019, 28(10): 106105 . **doi:** 10.1088/1674-1056/ab4174
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Geoscience material structures prediction via CALYPSO methodology

Andreas Hermann

Chin. Phys. B . 2019, 28(10): 106107 . **doi:** 10.1088/1674-1056/ab43bc

Predicted novel insulating electride compound between alkali metals lithium and sodium under high pressure

Yang-Mei Chen(陈杨梅), Hua-Yun Geng(耿华运), Xiao-Zhen Yan(颜小珍), Zi-Wei Wang(王紫薇), Xiang-Rong Chen(陈向荣), Qiang Wu(吴强)

Chin. Phys. B . 2017, 26(5): 056102 . **doi:** 10.1088/1674-1056/26/5/056102

New crystal structure and physical properties of TcB from first-principles calculations

Zhang Gang-Tai, Bai Ting-Ting, Yan Hai-Yan, Zhao Ya-Ru

Chin. Phys. B . 2015, 24(10): 106104 . **doi:** 10.1088/1674-1056/24/10/106104

The ground-state structure and physical properties of RuC: first-principles calculations

Zhang Mei-Guang, Yan Hai-Yan, Zhang Gang-Tai, Wang Hui

Chin. Phys. B . 2012, 21(7): 076103 . **doi:** 10.1088/1674-1056/21/7/076103

High-pressure structures of InBi predicted by particle swarm optimization algorithm

Liu Huan-Huan,Liu Yan-Hui

Chin. Phys. B . 2012, 21(2): 026102 . **doi:** 10.1088/1674-1056/21/2/026102

TOPICAL REVIEW — CALYPSO structure prediction methodology and its applications to materials discovery

The CALYPSO methodology for structure prediction*

Qunchao Tong(童群超), Jian Lv(吕健)†, Pengyue Gao(高朋越), and Yanchao Wang(王彦超)‡

Innovation Center of Computational Physics Methods and Software, State Key Laboratory of Superhard Materials, College of Physics, Jilin University, Changchun 130012, *China*

(Received 30 July 2019; revised manuscript received 28 August 2019; published online 30 September 2019)

Structure prediction methods have been widely used as a state-of-the-art tool for structure searches and materials discovery, leading to many theory-driven breakthroughs on discoveries of new materials. These methods generally involve the exploration of the potential energy surfaces of materials through various structure sampling techniques and optimization algorithms in conjunction with quantum mechanical calculations. By taking advantage of the general feature of materials potential energy surface and swarm-intelligence-based global optimization algorithms, we have developed the CALYPSO method for structure prediction, which has been widely used in fields as diverse as computational physics, chemistry, and materials science. In this review, we provide the basic theory of the CALYPSO method, placing particular emphasis on the principles of its various structure dealing methods. We also survey the current challenges faced by structure prediction methods and include an outlook on the future developments of CALYPSO in the conclusions.

Keywords: structure prediction, CALYPSO method, crystal structure, potential energy surface

PACS: 61.50.Ah, 31.50.–x, 02.60.Pn **DOI:** 10.1088/1674-1056/ab4174

1. Introduction

The atomic structure is the basis for a deep understanding of the properties or functionalities of materials and is closely relevant to many areas of science as diverse as physics, chemistry, biology, pharmaceutics, etc. Predicting structures utilizing modern theory and supercomputers is highly desirable for many reasons. On the one hand, experimental measurements of structures are usually indirect or under extreme conditions, leading to poor or incomplete data from which the complete structure determination is impossible. For example, photoelectron spectroscopy or collision cross section measurements are usually used to study the structures of atomic clusters at the nanoscale, but only indirect data, such as the electron binding energy and cross section, can be extracted.[1,2] Powder x-ray diffraction combined with a diamond anvil cell apparatus is a major tool for solving the crystal structures of materials at high pressures, but the small sample size generally induces poor diffraction data.[3] In these cases, theoretical structure predictions can provide putative structure models to supplement experiments for a complete structure determination. On the other hand, traditional methods of functional material discovery require expensive and laborious experimental attempts involving long-period trial-and-error processes. In this context, structure prediction combined with quantum mechanical simulation can be much faster and cheaper than experiments to search a wide range of systems for discovering promising new materials, which will provide guidance for purposive experimental synthesis.

Structure prediction with given information of chemical composition can be mathematically formulated as a global optimization problem, since the most stable structure generally corresponds to the global free energy (reduced to enthalpy at 0 K) minimum of the potential energy surface (PES). In principle, the exact solution can be obtained by visiting all local minima on the PES to determine the global one. However, the number of local minima on the PES is expected to grow exponentially with the number of atoms in the system. This growth renders the exhaustively ergodic searching strategy unfeasible, even for systems with a few atoms, and the issues can be infinitely exacerbated with an increasing system size. Therefore, structure prediction has been long thought to be a formidable problem.[4]

Over the years, much effort has been devoted to solving the problem of structure prediction.[5,6] Many algorithms have been proposed either to reduce the searching space or to enhance the sampling efficiency of the PES; these approaches have led to steady progress in the structure prediction community. A number of powerful structure prediction methods have been proposed, where typical examples are simulated annealing,[7] basin hopping,[8] minima hopping,[9] metadynamics,[10] random sampling,[11] genetic algorithm,[12–17] and our developed CALYPSO (Crystal struc-

*Project supported by the National Natural Science Foundation of China (Grant Nos. 11534003 and 11604117), the National Key Research and Development Program of China (Grant No. 2016YFB0201201), the Program for JLU Science and Technology Innovative Research Team (JLUSTIRT) of China, and the Science Challenge Project of China (Grant No. TZ2016001).

†Corresponding author. E-mail: lvjian@calypso.cn

‡Corresponding author. E-mail: wyc@calypso.cn

© 2019 Chinese Physical Society and IOP Publishing Ltd

ture AnaLYsis by Particle Swarm Optimization) method.[18,19] These methods significantly push the upper limit of the system size that is tractable for structure prediction and endow computational scientists with the predictive power to guide experiments. More and increasingly more theory-driven discoveries of new materials have been made,[20–24] such as the sulfur hydrides[25,26] and lanthanum superhydrides[27,28] with near room-temperature superconductivity,[29–31] the sodium chlorides with unexpected stoichiometries[32] and the Cu_2Si monolayer with planar hexacoordinate bonding.[33] The structure prediction method has become a standard tool in essentially all fields concerning the arrangement of atoms. There are considerable studies in the literature that review the methodology of different structure prediction methods and their applications in different areas.[5,6,11,34,35] In this work, we only focus on our CALYPSO method, paying particular attention to its basic theory and design principles. We refer the readers to Refs. [36] and [37] and review papers in the same issue for the method's wide applications.

This review is organized as follows. In Section 2, we provide a brief introduction on the general feature of the PES and the problem of structure prediction. In Section 3, we review the basic theory and design principles of the CALYPSO method, as well as the various functional modules of the CALYPSO software. Finally, the current challenges and future developments of the structure prediction method are discussed.

2. Potential energy surfaces and theoretical structure prediction

In the context of theoretical structure prediction, the PES represents the energy as a function of geometric structures of assemblies of atoms (Fig. 1). The dimension of PES increases linearly with the number of atoms N within the system, which are $3N+6$ and $3N-3$ for periodic crystal and isolated molecules, respectively. Theoretical and numerical analyses have demonstrated that the number of local minima on the PES grows exponentially as the number of atoms increases.[38,39] For example, the number of local minima is approximately 10^3 for the 13-atom Lennard–Jones (LJ) cluster, and it increases to at least 10^{12} for the 55-atom cluster.[40] There is some consensus on the general features of the PES, as described in Refs. [5], [9], [11], and [41], which we briefly restate as follows:

(i) The PES is composed of basins of attraction on which local optimization leads to an energy minimum that corresponds to a dynamically stable structure.

(ii) The substantial fraction of the PES corresponds to structures in which some atoms are very close to each other. These structures generally have energies even much higher than that of a set of isolated atoms, and the corresponding high-energy regions of the PES contain almost no minima.

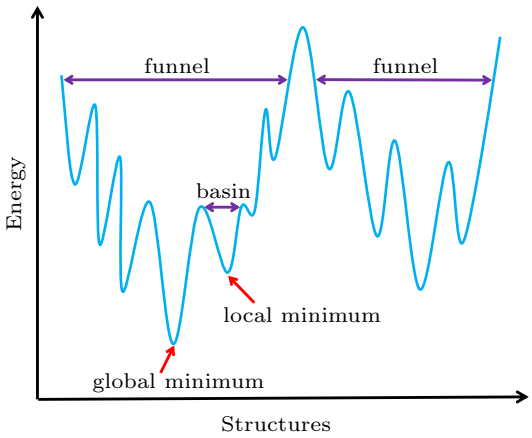


Fig. 1. Schematic illustration of a one-dimensional potential energy surface.

(iii) According to the Bell–Evans–Polanyi principle,[42,43] it is more likely to find a low energy local minimum if one moves from the current basin over a low barrier into a new basin than if one is required to overcome a high barrier. As a consequence, the low-energy basins on the PES tend to gather together, forming a superbasin, which is normally called a funnel. The structures that reside in the same funnel are generally similar to each other and can be regarded as corresponding to the same structural motif.

(iv) The probability distribution of the energies of the local minima of a PES is close to Gaussian when the system is sufficiently large. Symmetric structures tend to correspond to very low or very high-energy minima.[44]

The goal of structure prediction is to find the global minimum of a PES as well as the corresponding structure. The exponential increase in the number of local minima indicates that it belongs to a notorious NP-hard (non-deterministic polynomial-time hard) problem. Therefore, any algorithm cannot guarantee working without failing in polynomial time. All effective algorithms would be bound to have an upper limit in the system size that can be tractable, and they may eventually converge when averaged over all the systems.[45] Although the prospect for completely solving the problem of structure prediction seems bleak, significant progress has been made in the community during the last few decades, which make structure prediction a popular tool routinely used in modern computational physics, chemistry, and material science. This is partially because knowledge about the general features of the PES allows one to disregard the most irrelevant part or bias the desired part of the PES when developing a structure prediction method. This significantly simplifies the problem and makes structure prediction methods applicable to systems with moderate sizes. For example, the CALYPSO method can reliably address systems with several tens

of atoms when the energy calculation is taken at the quantum mechanical level, and it can address more than one hundred atoms when empirical potentials are used. A large number of structures have been predicted and experimentally confirmed, validating the method. Some of them possess a large system size, e.g. the 64-atom BC_3[46] and 76-atom $Bi_4Si_3O_{12}$.[47] Note that, the majority of inorganic compounds tend to have small system sizes. Approximately 90% of crystals in the Inorganic Crystal Structure Database have less than 60 atoms in the primitive cell.[48] This number indicates that state-of-the-art structure prediction methods are already applicable to most ordered inorganic compounds.

3. The CALYPSO method

CALYPSO is a numerical method for structure prediction with given information with respect to the chemical composition alone for a material.[18,19] The essence of the method is a heuristic global minimization strategy for the PES, which is composed of several structure dealing methods, such as, structure generation with constraints of symmetry, structure characterization through the bond characterization matrix (BCM), and structure evolution via the swarm-intelligence based particle swarm optimization (PSO) algorithm.

Structure predictions through CALYPSO consist of mainly four steps, as depicted in the flow chart in Fig. 2. First, the initial structures are randomly generated with the constraints of symmetry to allow a diverse sampling of the PES. Then, BCM will be used to characterize the newly generated structures and examine their distances/similarities with all the previous ones. Structures with distances less than a threshold will be eliminated. After a user-specified number of structures (a population or generation) have been generated, local structural optimizations are performed to eliminate the noise of the energy surface and drive the structures to the local minima. Eventually, the swarm-intelligence PSO algorithm is applied to produce new structures for the next generation. This process continues iteratively until convergence criteria are reached.

The various structure dealing methods within CALYPSO were designed by taking into account the following three aspects for solving the problem of structure prediction: (i) simplifying the problem using physical constraints, (ii) extracting subproblems through structure characterization, and (iii) solving subproblems via a heuristic global optimization algorithm. Below we introduce the various structure dealing methods within CALYPSO from the above three aspects.

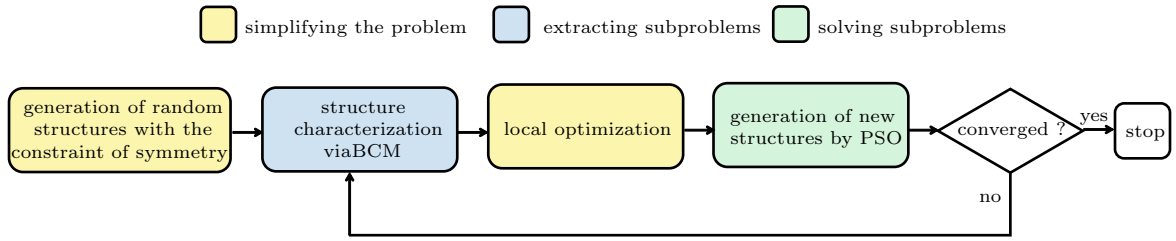


Fig. 2. The flow char of CALYPSO. Various structure dealing methods are colored according to their role in solving the problem of structure prediction.

3.1. Simplifying the problem using physical constraints

Knowledge regarding the PES, as stated in Section 2, allows us to impose various physical constraints on structures to simplify the problem of structure prediction. The most straightforward one is the constraint of the minimum interatomic distance, which has been adopted by most structure prediction methods. From feature ii of the PES, we known that a substantial fraction of the PES possesses high energy and contains almost no minima, which corresponds to structures with some atoms very close proximity to each other. Therefore, imposing the constraint of the minimum interatomic distance on structures during the structure prediction can disregard a large fraction of high-energy regions of the PES. This constraint will significantly reduce the search space and simplify the problem. In the CALYPSO method, we used the constraint of the minimum interatomic distance when generating new structures. Note that there is no risk at losing the global minimum for imposing this constraint, as we know that all real structures show interatomic distances within a reasonable range (even distances between H atoms under high compression are not less than ~ 0.74 Å).

The second constraint on structures is symmetry. The CALYPSO structure search involves the generation of a large number of random structures to sample the PES. Feature iv of the PES tells us that fully random structures correspond to the local minima with a Gaussian-like energetic distribution. This has been numerically demonstrated using isolated LJ systems in our previous work,[49] where we randomly generated 10000 structures for LJ_{38} and LJ_{100} clusters, receptively, and then locally optimized them. As shown in Fig. 3, both clusters demonstrate a Gaussian-like distribution with a peak centered at a high energy. As the number of atoms increases from 38 to 100, the peaks shift to higher energies and become apparently narrower. A careful inspection of the structures around the peak revealed a large number of disordered or liquid-like struc-

tures without symmetry. Therefore, although intuitively a fully random generation of structures is expected to give a diverse sampling of the PES, it in fact frequently produces nonsymmetric liquid-like structures with high energies, which is unfavorable for structure prediction. Given that symmetric structures tend to correspond to a very low- or a very high-energy minimum (Feature iv), it is beneficial to impose symmetry during the generation of random structures, which would achieve a more diverse sampling of the PES in terms of energies and symmetries. For example, when we randomly impose C_1 to C_6 point group symmetry for the generation of the LJ_{38} and LJ_{100} clusters (Fig. 3), the energetic distributions are more spread, and more low-energy local minima have been detected. In this respect, biasing symmetric structures during random structure generation is more sensible and "random".

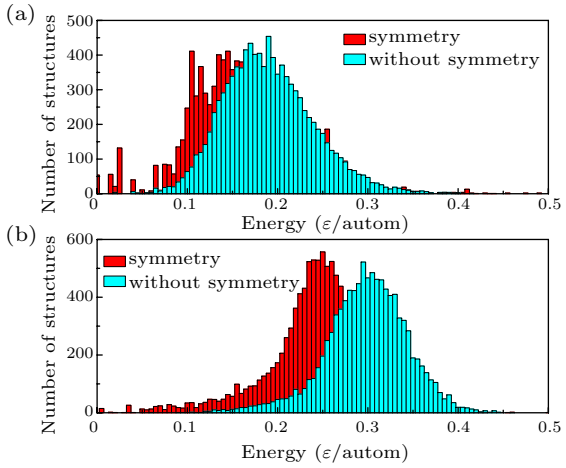


Fig. 3. Energetic distributions of randomly generated structures with and without symmetry constraints for the (a) LJ_{38} and (b) LJ_{100} clusters. Energies are shown relative to the global minimum. The LJ potential for a pair of atoms is given by $U_{ij} = \varepsilon\,[(r/r_{ij})^{12} - 2(r/r_{ij})^{6}]$, where ε and r are the pair equilibrium well depth and separation, respectively. The reduced units, i.e., $\varepsilon = r = 1$ are employed throughout. Reprinted from Ref. [49], with the permission of AIP Publishing.

In CALYPSO, different symmetry groups are adopted for generating structures with different dimensions. In particular, 230 space group symmetries are used to constrain bulk crystals, and 17 in-plane space group symmetries are used to constrain two-dimensional (2D) systems, such as layered structures and surfaces. In principle, an infinite number of point groups exist for zero-dimensional (0D) isolated systems (molecules or nanoclusters). Here, we adopt 48 point group symmetries that frequently appear in such systems. For example, when we generate a crystal structure, one of the 230 space groups will be randomly selected, then the lattice parameters are generated within the chosen symmetry according to a confined volume, and the atomic positions are obtained by combining several sets of symmetry-equivalent coordinates (Wyckoff Positions) in accordance with the symmetry and the number of atoms in the simulation cell. We have demonstrated that the inclusion of symmetry constraints during structure generation significantly increases the efficiency and success rate of structure prediction.[19,49]

The third constraint on the structures is the dynamical stability, which is a prerequisite for the real structures and can be readily achieved by local optimizing a structure to its corresponding local minimum (Feature i of the PES). Each newly generated structure is subject to local optimization during the CALYPSO structure search. Although this step is the most time-consuming part (since it involves a large number of energy calculations) during structure prediction, it plays a role as transforming the searching space from an entire continuous PES to a series of discrete local minima, which significantly simplifies the problem. Local structural optimization has been an indispensable step in modern structure prediction methods and is largely responsible for their success.

3.2. Extracting subproblems through structure characterization

If a structure prediction problem can be split into several subproblems, then each subproblem should be easier to solve than the whole problem. Feature iii of the PES tell us that the basins of attraction arrange themselves into funnels, each of which represents a class of structures with a certain structural motif. Therefore, it is reasonable to split the problem on the basis of the funnels. This approach requires a method to characterize structures and provide a metric for measuring the distance (similarity/dissimilarity) between structures in the configuration space. Structure characterization is widely involved in many areas, such as structure prediction,[18,50–52] machine learning for constructing interatomic potentials or materials discovery,[53,54] and chemical informatics.[55] There are different requirements for structure characterization in different applications. In the context of structure prediction, a real valued function is required to uniquely characterize a structure solely from the geometry and chemical identities of the constituent atoms. The function should be invariant to the translation and rotation of a structure, as well as the permutation of atoms of the same type within the structure. Moreover, a reasonable metric for measuring distances between structures should be defined based on the function, i.e., the distances should be zero between the identical structures and gradually increase as structures depart from each other.

Based on the above requirements, we designed a BCM to characterize structures in the CALYPSO method.[19,49] This matrix is realized by extending the bond-orientational order parameters, which was previously introduced by Steinhardt *et al.*[56] The BCM is constructed based on all bond information of a structure. This method utilizes spherical harmonics and exponential functions to describe the orientations and lengths of bonds. A bond vector $\boldsymbol{r}_{ij}$ between atoms i and j is defined

if the interatomic distance between them is less than a cutoff distance. The vector $\boldsymbol{r}_{ij}$ is associated with the spherical harmonics $Y_{lm}(\theta_{ij}, \phi_{ij})$, where θ_{ij} and ϕ_{ij} are the polar angles. A weighted average over all bonds formed by types A and B atoms is then performed by the equation:

$$\bar{Q}_{lm}^{\delta_{\mathrm{AB}}} = \frac{1}{N_{\delta_{\mathrm{AB}}}} \sum_{i\in A, j\in B} \mathrm{e}^{-\alpha\left(r_{ij}-r_{\mathrm{AB}}\right)} Y_{lm}(\theta_{ij}, \phi_{ij}),$$

where δ_{AB} and $N_{\delta_{\mathrm{AB}}}$ denote the type and number of bonds, respectively. Only even-l spherical harmonics are used in the above equation to guarantee the invariant bond information with respect to the direction of the bonds. To avoid dependence on the choice of the reference frame, it is important to consider the rotationally invariant combinations,

$$Q_l^{\delta_{\mathrm{AB}}} = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} \left|\bar{Q}_{lm}^{\delta_{\mathrm{AB}}}\right|},$$

where each series of $Q_l^{\delta_{\mathrm{AB}}}$ for $l = 0$, 2, 4, 6, 8, and 10 can be used to represent a type of bond, thus being an element of the BCM of structure. As a result, the similarity of the two structures can be quantitatively represented by the Euclidean distance between their BCMs,

$$D_{\mathrm{uv}} = \sqrt{\frac{1}{N_{\mathrm{type}}} \sum_{\delta_{\mathrm{AB}}} \sum_{l} \left(Q_l^{\delta_{\mathrm{AB,u}}} - Q_l^{\delta_{\mathrm{AB,v}}}\right)^2},$$

where u and v denote two individual structures and N_{type} is the number of bond types. Since all Q_l except for Q are zero for isotropic systems, e.g., infinitely large liquid, it is therefore plausible to define a distance from a structure to the isotropic system as

$$D_u = \sqrt{\frac{1}{N_{\mathrm{type}}} \sum_{\delta_{\mathrm{AB}}} \sum_{l} \left(Q_l^{\delta_{\mathrm{AB,u}}}\right)^2},$$

which can be used to estimate the degree of order for a given structure. A numerical demonstration of the validity of the BCM for structure characterization has been shown for isolated clusters in Ref. [49] and for crystals in Ref. [19].

Although we are able to uniquely quantify structures through the BCM, the unambiguous definition of the funnels of a PES is still difficult since it requires a detailed understanding of the PES. Therefore, we designed a scheme to dynamically decide the center of funnels during the CALYPSO structure search. Given a set of structures that have been detected during a structure search, the lowest-energy structure is always selected as the center of a funnel. Then, the second one is defined as the lowest-energy structure among the structures that have BCM distances (relative to previous defined centers) larger than the threshold value. This procedure repeats until a predefined number of centers have been chosen. As the structure search proceeds, new structures are continually included at each generation, and the centers of the funnels will be updated when a new structure with lower energy is found. The threshold value of distance will also be adjusted to allow a sufficient number of centers to be defined. Based on the defined centers, CALYPSO can readily classify a structure into a funnel which it is nearest to (without the need of any prior information regarding the PES). Therefore, the problem of structure prediction can be divided into several funnels, each of which can be solved through a heuristic global optimization.

3.3. Solving subproblems via a heuristic global optimization algorithm

The CALYPSO method is within the evolutionary scheme where structures evolve according to the PSO algorithm. The PSO algorithm is a typical swarm-intelligence scheme for global optimization inspired by natural biological systems (e.g., ants, bees, or birds),[57,58] and has been applied to a variety of fields in engineering and chemical science. In practice, a candidate structure in the configurational space is regarded as a particle, and a set of individual particles is called a population or a generation. Each particle explores the search space via a velocity vector, which is affected by both its personal best experience and the best position found by the population so far (Fig. 4(a)).

Within the PSO scheme, the position of ith particle at the jth dimension is updated according to the following equation:

$$x_{i,j}^{t+1} = x_{i,j}^{t} + v_{i,j}^{t+1},$$

where t denotes the generation index and $j \in \{1, 2, 3\}$. The new velocity ($v_{i,j}^{t+1}$) is calculated on the basis of its previous location, previous velocity ($v_{i,j}^{t}$), the personal best location ($pbest_{i,j}^{t}$) with an achieved best fitness of this individual and the population best location ($gbest_{i,j}^{t}$) with the best fitness for the entire population,

$$v_{i,j}^{t+1} = \omega v_{i,j}^{t} + c_1 r_1 \left(pbest_{i,j}^{t} - x_{i,j}^{t}\right) + c_2 r_2 \left(gbest_{i,j}^{t} - x_{i,j}^{t}\right),$$

where ω (in the range of 0.9–0.4) denotes the inertia weight, $c_1 = 2$, $c_2 = 2$, r_1 and r_2 are two random numbers that are uniformly distributed in the range [0,1].

We have implemented two versions of the PSO algorithm (global and local version) in CALYPSO, which are illustrated in Figs. 4(b) and 4(c), respectively. In the global version, the problem is solved as a whole though one PSO evolution. All particles seek new positions according to one best location ($gbest$) for the entire population and their personal best locations ($pbest$). It has been demonstrated to be efficient and powerful for global structural convergence, especially for small systems. However, it is noteworthy that for large systems with much more complex PES, finer structural searches are desirable. Therefore, we introduced a local version of PSO in which the whole problem is solved though several PSO evolutions.[49,59] As stated in the above section, we assume

that each particle belongs to a funnel based on the BCM distances. Then, its velocity is adjusted according to both its personal best position and the best position (*lbest*) achieved so far within the funnel,

$$v_{i,j}^{t+1} = \omega v_{i,j}^{t} + c_1 r_1 \left(pbest_{i,j}^{t} - x_{i,j}^{t}\right) + c_2 r_2 \left(lbest_{i,j}^{t} - x_{i,j}^{t}\right).$$

Definitely, *lbest* denotes the center of the funnel to which a particle belongs. Therefore, the entire swam diffuses into several funnels, in which several PSO evolutions are performed simultaneously. The *local* PSO algorithm can be seen as a combination of several information-shared *global* PSOs. This method apparently avoids premature structures by maintaining multiple attractors and allows the exploration of the larger space of PESs, although it is computationally more demanding.

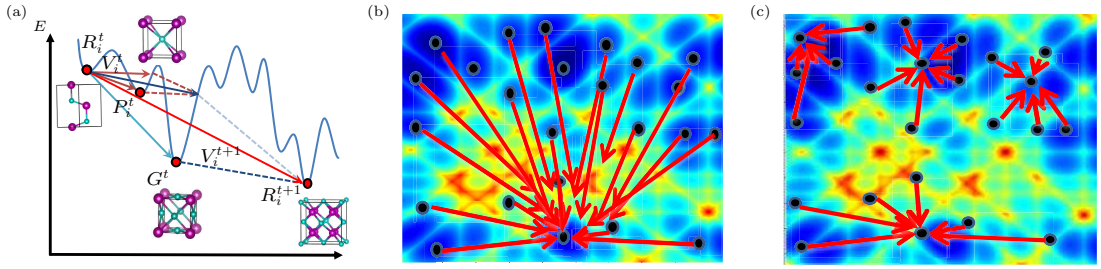


Fig. 4. Schematics of (a) the velocity and position updates in PSO, (b) global PSO, and (c) local PSO. Reprinted from Ref. [37], with the permission of IOP Publishing.

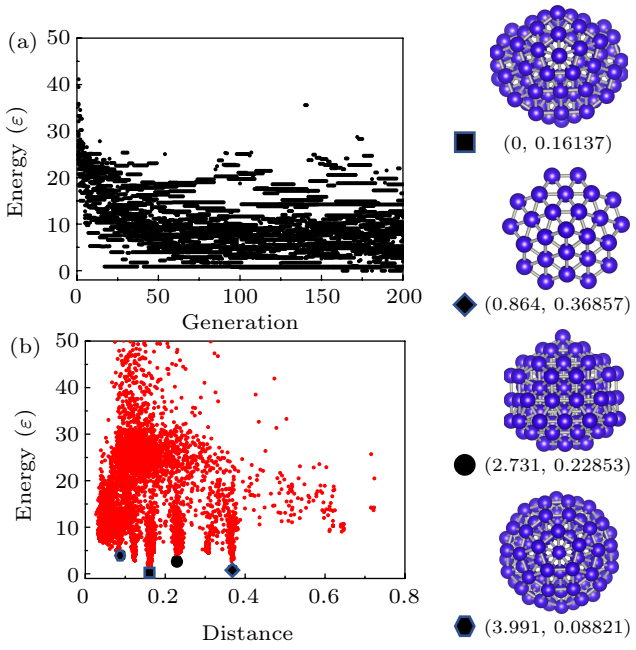


Fig. 5. (a) Energy as a function of the evolution step of a CALYPSO structure search performed on the LJ_{100} cluster. (b) Energy–distance distribution for the minima searched during the CALYPSO run. The energies are shown relative to the global minimum and the BCM distances were calculated with respect to the isotropic system. Symbols denote the four *lbest*s at the 200th generation, which are shown in the right panel. The numbers below each structure are the energies in ε relative to the global minimum and BCM distances with respect to the isotropic system. See text for detailed structural descriptions. Reprinted from Ref. [49], with the permission of AIP Publishing.

We illustrate the *local* PSO through a CALYPSO structure search of the LJ_{100} cluster. Figure 5 (a) depicts the energies as a function of the evolution step of the structure search, where four *lbest*s were used. In the first 50 evolution steps, the overall energies decreased rapidly. Then, structural searches were mainly conducted on the low-energy regions, and the global stable structure was successfully produced at the 163rd generation. We plot the energy–distance distribution for the minima detected during this structure search in Fig. 5(b). Structures possessing similar structural motifs show similar BCM distances to the isotropic system, and these structures are expected to reside in the same energy funnel. Different funnels can be clearly identified by the calculated BCM distances in Fig. 5(b). The symbols denote the four *lbest* structures determined at the 200th generation (the corresponding structures are shown in the right panel in Fig. 5). These structures reside in different funnels and possess structural motifs that are very different from each other. The first *lbest* structure is the global stable structure with an icosahedral structural motif and a BCM distance of 0.16137. The second *lbest* structure contains a Marks decahedral motif and has a BCM distance of 0.36857. The third and fourth *lbest*s consist of Marks decahedral and Mackay icosahedral cores with incomplete anti-Mackay overlayers, respectively. This example clearly shows that *local* PSO enables a simultaneous search in different energy funnels of a potential energy surface.

3.4. Other strategies for biasing low-energy structures and increasing the structure diversity

We have introduced the basic idea of the CALYPSO structure search. However, several other efficient techniques have also been implemented in CALYPSO to further increase the search efficiency. To maintain the structural diversity as well as further biasing the structure search toward low-energy regions, we have included a penalty function during the structure search. In each generation, a certain percentage of low-energy structures are used to produce the next generation by PSO, while the remaining high-energy structures are rejected and replead by newly generated random structures. This technique has been demonstrated to be important for enriching the structure diversity and is crucial for increasing the search ef-

ficiency. Moreover, inspired by the Monte Carlo simulation, the Metropolis criterion can also be included in CALYPSO for acceptance or rejection of new structures. A structure will be accepted if it has a lower energy than its parent structure. Otherwise, a selection probability is imposed based on the relative energies according to the Boltzmann distribution. This implementation further improves the possibility of generating low-energy structures during structural evolution.

3.5. Features of the CALYPSO software

The CALYPSO method as stated above has been implemented into the same-name software package. Currently, it contains ten functional modules, which allows to perform unbiased search for the energetically stable/metastable structures of isolated clusters/nanoparticles,[49] 2D layered materials,[60] surfaces,[61] interfaces,[62] adsorbate systems,[63] proteins,[64] and three-dimensional (3D) crystals.[18] It can also be used to design novel functional materials with desirable functionalities when we change the global fitness function from the total energy to a specific property (e.g., hardness, bandgaps, and even experimental x-ray diffraction patterns).[65–67] Moreover, we have recently proposed a simple and unambiguous definition for crystal structure prototypes based on hierarchical clustering and constructed a crystal structure prototype database (CSPD) by filtering existing crystallographic structure databases.[68] The CSPD also accumulates a low-energy structure during CALYPSO structure prediction to construct a separate theoretical structure database. This database can be used for generating initial structures for structure prediction or determining the prototype of a structure, as well as high-throughput calculation.

4. Conclusion and perspectives

We have presented a short review on the basic theory of our developed CALYPSO method, paying particular attention to the design principles of the various structure dealing methods. The validity of the CALYPSO method has been extensively demonstrated through its application to various materials ranging from 0D to 3D systems. We refer the readers to review papers in the same issue for their latest applications on 0D nanoclusters, 2D layered materials, and 3D bulk crystals for superconductors and superhard materials, as well as exotic materials at high pressure.

Long thought to be impossible, modern theoretical structure prediction has achieved substantial progress and can routinely deal with systems with several tens of atoms. However, large systems containing hundreds of atoms is still a great challenge for structure prediction, where a large number of structure-related problems exist concerning realistic materials (e.g., alloys, minerals, and drugs). Both searching and ranking problems should be further addressed. Quantum mechanical methods for energy calculations can provide a reasonable ranking of structures, but the computational cost is very demanding for large systems, while empirical potential suffers from poor transferability. One promising solution is the utilization of machine learning potentials,[69] which show accuracy comparable to those of quantum mechanical methods, but with less computational cost by several orders of magnitude. The utilization of machine learning potentials to accelerate structure predictions is now in its early stages but has already generated several encouraging results.[70–72] Recently, we have developed an acceleration scheme for the CALYPSO structure prediction of large systems, in which a machine-learning potential (Gaussian approximation potential[73]) is trained in an on-the-fly manner from scratch during the structure searches and is used to accelerate the structure searches.[74] The scheme substantially reduced the computational cost by at least ∼1–2 orders of magnitude compared with full DFT-based structure searches. For the search problem, it is still difficult to give a definitive answer as to whether we can cross the "exponential wall" of the number of local minima. Further simplifying the PES for a specific problem is a practical way. For example, the partial or complete fixation of the geometry of molecules is a wise choice for predicting molecular crystals.[75] In addition to large systems, several other important issues are also faced by structure prediction, such as predicting materials at finite temperature and the synthesizability of a material. With these challenges addressed, we imagine more wide applications of the structure prediction method in all fields relevant to atomic structures.

References

[1] Wang L S 2016 *Int. Rev. Phys. Chem.* **35** 69
[2] Oger E, Crawford N R M, Kelting R, Weis P, Kappes M M and Ahlrichs R 2007 *Angew. Chem. Int. Ed.* **46** 8503
[3] Zhang L, Wang Y, Lv J and Ma Y 2017 *Nat. Rev. Mater.* **2** 17005
[4] Maddox J 1988 *Nature* **335** 201
[5] Wang Y and Ma Y 2014 *J. Chem. Phys.* **140** 040901
[6] Oganov A R, Pickard C J, Zhu Q and Needs R J 2019 *Nat. Rev. Mater.* **4** 331
[7] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671
[8] Wales D J and Doye J P K 1997 *J. Phys. Chem. A* **101** 5111
[9] Goedecker S 2004 *J. Chem. Phys.* **120** 9911
[10] Martoňák R, Laio A and Parrinello M 2003 *Phys. Rev. Lett.* **90** 075503
[11] Pickard C J and Needs R J 2011 *J. Phys. Condens. Matter* **23** 053201
[12] Oganov A R and Glass C W 2006 *J. Chem. Phys.* **124** 244704
[13] Lonie D C and Zurek E 2011 *Comput. Phys. Commun.* **182** 372
[14] Kolmogorov A N, Shah S, Margine E R, Bialon A F, Hammerschmidt T and Drautz R 2010 *Phys. Rev. Lett.* **105** 217003
[15] Trimarchi G and Zunger A 2007 *Phys. Rev. B* **75** 104113
[16] Bahmann S and Kortus J 2013 *Comput. Phys. Commun.* **184** 1618
[17] Bi W, Meng Y, Kumar R S, Cornelius A L, Tipton W W, Hennig R G, Zhang Y, Chen C and Schilling J S 2011 *Phys. Rev. B* **83** 104106
[18] Wang Y, Lv J, Zhu L and Ma Y 2010 *Phys. Rev. B* **82** 094116
[19] Wang Y, Lv J, Zhu L and Ma Y 2012 *Comput. Phys. Commun.* **183** 2063

[20] Li Y, Wang L, Liu H, Zhang Y, Hao J, Pickard C J, Nelson J R, Needs R J, Li W, Huang Y, Errea I, Calandra M, Mauri F and Ma Y 2016 *Phys. Rev. B* **93** 020103
[21] Li Y, Hao J, Liu H, Lu S and Tse J S 2015 *Phys. Rev. Lett.* **115** 105502
[22] Li Y, Hao J, Liu H, Tse J S, Wang Y and Ma Y 2015 *Sci. Rep.* **5** 9948
[23] Li Y, Wang Y, Pickard C J, Needs R J, Wang Y and Ma Y 2015 *Phys. Rev. Lett.* **114** 125501
[24] Li Y, Feng X, Liu H, Hao J, Redfern S A T, Lei W, Liu D and Ma Y 2018 *Nat. Commun.* **9** 722
[25] Li Y, Hao J, Liu H, Li Y and Ma Y 2014 *J. Chem. Phys.* **140** 174712
[26] Duan D, Liu Y, Tian F, Li D, Huang X, Zhao Z, Yu H, Liu B, Tian W and Cui T 2015 *Sci. Rep.* **4** 6968
[27] Peng F, Sun Y, Pickard C J, Needs R J, Wu Q and Ma Y 2017 *Phys. Rev. Lett.* **119** 107001
[28] Liu H, Naumov I I, Hoffmann R, Ashcroft N W and Hemley R J 2017 *Proc. Natl. Acad. Sci.* **114** 6990
[29] Drozdov A P, Eremets M I, Troyan I A, Ksenofontov V and Shylin S I 2015 *Nature* **525** 73
[30] Ahart M, Somayazulu M, Meng Y, Struzhkin V V, Baldini M, Mishra A K, Geballe Z M and Hemley R J 2019 *Phys. Rev. Lett.* **122** 27001
[31] Drozdov A P, Kong P P, Minkov V S, Besedin S P, Kuzovnikov M A, Mozaffari S, Balicas L, Balakirev F F, Graf D E, Prakapenka V B, Greenberg E, Knyazev D A, Tkacz M and Eremets M I 2019 *Nature* **569** 528
[32] Zhang W, Oganov A R, Goncharov A F, Zhu Q, Boulfelfel S E, Lyakhov A O, Stavrou E, Somayazulu M, Prakapenka V B and Konôpková Z 2013 *Science* **342** 1502
[33] Yang L M, Bačić V, Popov I A, Boldyrev A I, Heine T, Frauenheim T and Ganz E 2015 *J. Am. Chem. Soc.* **137** 2757
[34] Woodley S M and Catlow R 2008 *Nat. Mater.* **7** 937
[35] Zhao J, Shi R, Sai L, Huang X and Su Y 2016 *Mol. Simul.* **42** 809
[36] Wang H, Wang Y, Lv J, Li Q, Zhang L and Ma Y 2016 *Comput. Mater. Sci.* **112** 406
[37] Wang Y, Lv J, Zhu L, Lu S, Yin K, Li Q, Wang H, Zhang L and Ma Y 2015 *J. Phys. Condens. Matter* **27** 203203
[38] Stillinger F H 1999 *Phys. Rev. E* **59** 48
[39] Tsai C J and Jordan K D 1993 *J. Phys. Chem.* **97** 11227
[40] Doye J P K and Wales D J 1995 *J. Chem. Phys.* **102** 9659
[41] Wales D 2004 *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge: Cambridge University Press)
[42] Jensen F 2007 *Introduction to Computational Chemistry* (Wiley)
[43] Roy S, Goedecker S and Hellmann V 2008 *Phys. Rev. E* **77** 56707
[44] Wales D J 1998 *Chem. Phys. Lett.* **285** 330
[45] Wolpert D H and Macready W G 1997 *IEEE Trans. Evol. Comput.* **1** 67
[46] Zhang M, Liu H, Li Q, Gao B, Wang Y, Li H, Chen C and Ma Y 2015 *Phys. Rev. Lett.* **114** 15502
[47] Chen F, Ju M, Kuang X and Yeung Y 2018 *Inorg. Chem.* **57** 4563
[48] Xie T and Grossman J C 2018 *Phys. Rev. Lett.* **120** 145301
[49] Lv J, Wang Y, Zhu L and Ma Y 2012 *J. Chem. Phys.* **137** 084104
[50] Oganov A R and Valle M 2009 *J. Chem. Phys.* **130** 104504
[51] Zhu L, Amsler M, Fuhrer T, Schaefer B, Faraji S, Rostami S, Ghasemi S A, Sadeghi A, Grauzinyte M, Wolverton C and Goedecker S 2016 *J. Chem. Phys.* **144** 034203
[52] Sadeghi A, Ghasemi S A, Schaefer B, Mohr S, Lill M A and Goedecker S 2013 *J. Chem. Phys.* **139** 184118
[53] Behler J 2011 *J. Chem. Phys.* **134** 074106
[54] Bartók A P, Kondor R and Csányi G 2013 *Phys. Rev. B* **87** 184115
[55] Todeschini R and Consonni V 2009 *Mol. Descriptors For Chemoinformatics: Volume I: Alphabetical Listing/volume II: Appendices References* Vol 41 (John Wiley & Sons)
[56] Steinhardt P J, Nelson D R and Ronchetti M 1983 *Phys. Rev. B* **28** 784
[57] Kennedy J and Eberhart R 1995 *Proc. ICNN'95 - Int. Conf. Neural Netw.* **4** 1942
[58] Eberhart R and Kennedy J 1995 *Proc. Sixth Int. Symp. Micro Mach. Hum. Sci.* p. 39
[59] Wang Y, Liu H, Lv J, Zhu L, Wang H and Ma Y 2011 *Nat. Commun.* **2** 563
[60] Wang Y, Miao M, Lv J, Zhu L, Yin K, Liu H and Ma Y 2012 *J. Chem. Phys.* **137** 224108
[61] Lu S, Wang Y, Liu H, Miao M and Ma Y 2014 *Nat. Commun.* **5** 3666
[62] Gao B, Gao P, Lu S, Lv J, Wang Y and Ma Y 2019 *Sci. Bull.* **64** 301
[63] Gao B, Shao X, Lv J, Wang Y and Ma Y 2015 *J. Phys. Chem. C* **119** 20111
[64] Gao P, Wang S, Lv J, Wang Y and Ma Y 2017 *RSC Adv.* **7** 39869
[65] Zhang X, Wang Y, Lv J, Zhu C, Li Q, Zhang M, Li Q and Ma Y 2013 *J. Chem. Phys.* **138** 114101
[66] Gao P, Tong Q, Lv J, Wang Y and Ma Y 2017 *Comput. Phys. Commun.* **213** 40
[67] Zhang Y, Wang H, Wang Y, Zhang L and Ma Y 2017 *Phys. Rev. X* **7** 011017
[68] Su C, Lv J, Li Q, Wang H, Zhang L, Wang Y and Ma Y 2017 *J. Phys. Condens. Matter* **29** 165901
[69] Behler J 2016 *J. Chem. Phys.* **145** 170901
[70] Jacobsen T L, Jørgensen M S and Hammer B 2018 *Phys. Rev. Lett.* **120** 026102
[71] Deringer V L, Csányi G and Proserpio D M 2017 *ChemPhysChem.* **18** 873
[72] Deringer V L, Pickard C J and Csányi G 2018 *Phys. Rev. Lett.* **120** 156001
[73] Bartók A P, Payne M C, Kondor R and Csányi G 2010 *Phys. Rev. Lett.* **104** 136403
[74] Tong Q, Xue L, Lv J, Wang Y and Ma Y 2018 *Faraday Discuss.* **211** 31
[75] Reilly A M, Cooper R I, Adjiman C S, *et al.* 2016 *Acta Crystallogr. Sect. B* **72** 439