

## Evolutionary construction of a formation-energy convex hull: Practical scheme and application to a carbon-hydrogen binary system

Takahiro Ishikawa<sup>1,\*</sup> and Takashi Miyake<sup>1,2</sup><sup>1</sup>*ESICMM, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan*<sup>2</sup>*CD-FMat, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan*

(Received 12 March 2020; revised manuscript received 19 May 2020; accepted 28 May 2020; published 8 June 2020)

We present an evolutionary construction technique of formation-energy convex hull to search for thermodynamically stable compounds. In this technique, candidates with a wide variety of chemical compositions and crystal structures are created by systematically applying evolutionary operators, “mating,” “mutation,” and “adaptive mutation,” to two target compounds, and the convex hull is directly updated through the evolution. We applied the technique to carbon-hydrogen binary system at 10 GPa and obtained 15 hydrocarbons within the convex hull distance less than 0.5 mRy/atom: graphane, polybutadiene, polyethylene, butane, ethane, methane, three molecular compounds of ethane and methane, and six molecular compounds of methane and hydrogen. These results suggest that our evolutionary construction technique is useful for the exploration of stable phases under extreme conditions and the synthesis of new compounds.

DOI: [10.1103/PhysRevB.101.214106](https://doi.org/10.1103/PhysRevB.101.214106)

### I. INTRODUCTION

Search for thermodynamically stable compounds is crucial for the design and synthesis of novel functional materials and the understanding of the behavior of materials under extreme conditions. For example, an impressive achievement with respect to it is the discovery of high-temperature superconductivity in hydrides stabilized under high pressure conditions [1–8]. In first-principles calculations, thermodynamically stable phases of materials are predicted by constructing a convex hull with respect to formation energy of the compounds. If a binary compound,  $\alpha_{1-x}\beta_x$ , is considered as an example, thermodynamically stable chemical compositions are typically explored through the following four steps: (1) determination of the most stable crystal structure at a fixed composition rate  $x$  using structure search techniques, (2) calculation of formation energy  $E$  of the compound with respect to  $\alpha$  and  $\beta$ , (3) repeat of the same calculations changing  $x$ , and (4) construction of the convex hull curve by plotting  $E$  for  $x$ . The compositions on the convex hull correspond to thermodynamically stable ones. This approach has been successful in the discovery of new materials including superconducting hydrides mentioned above, whereas it requires greater computational resources, i.e., a large number of structural optimizations, to construct the convex hull accurately. Moreover, many calculations end up in a wasted effort because the appearance of the compounds on the convex hull is a rare event. Therefore, it is important for the increase of the search efficiency to use a direct construction approach of the convex hull without the procedure through the four steps, and various improvements have been carried out on the prediction codes for

multicomponent systems, e.g., USPEX [9–11], CALYPSO [12,13], AIRSS [14], and so on. For the USPEX code, thermodynamically stable compositions are explored on the line connecting between the selected two compounds by applying evolutionary operators called “heredity” and “transmutation.” For the CALYPSO code, different compositions are created by giving the variation range of the number of atoms for each atomic species included in a calculation cell as the input parameter, and stable compounds are explored by a particle-swarm optimization. For the AIRSS code, stable compounds are explored by randomly creating compositions and structures.

In this study, we propose an approach for the direct search, which we call “evolutionary construction technique of formation-energy convex hull.” In this technique, candidates with a wide variety of chemical compositions and crystal structures are created by systematically applying evolutionary operators, “mating,” “mutation” consisting of permutation, distortion, reflection, modulation, addition, elimination, and substitution, and “adaptive mutation,” to stable and metastable compounds on or near the convex hull. After the structural optimizations, the convex hull is updated. By repeatedly performing this process, the convex hull is evolutionarily constructed, which achieves a more efficient search for thermodynamically stable compounds compared with the typical approach based on the exhaustive search. In this paper, we show the details of the method and its application to the carbon-hydrogen binary system.

### II. DETAILS OF EVOLUTIONARY CONSTRUCTION TECHNIQUE

Evolutionary algorithm requires setting and tuning many parameters such as “heredity rate,” “permutation rate,”

\*ISHIKAWA.Takahiro@nims.go.jp

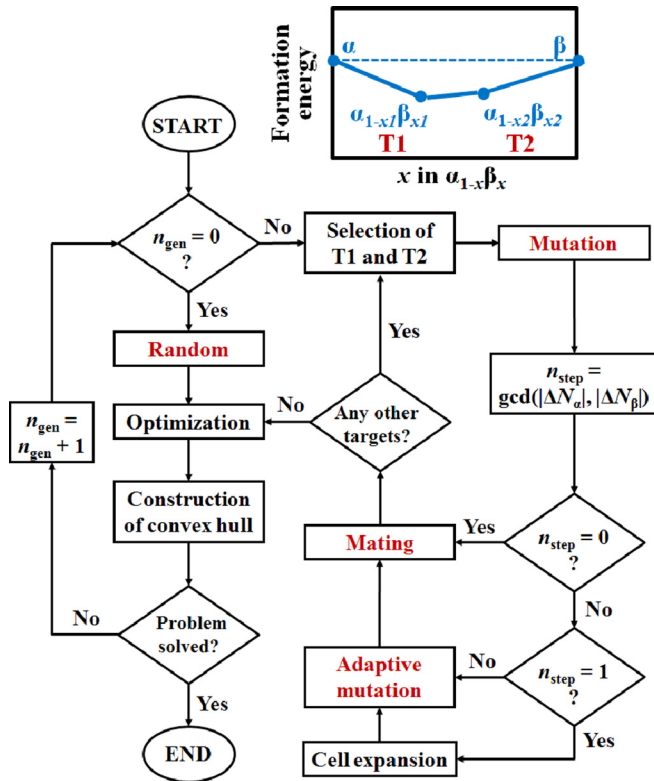


FIG. 1. Flowchart for the evolutionary construction of formation-energy convex hull with respect to binary systems.

“mutation rate,” “selection rate,” “population size,” and so on. On the other hand, in our technique the parameter setting and tuning are reduced because potential candidates are systematically created according to rules described below. Here, for simplicity, let us consider the case of a binary system,  $\alpha$  ( $x = 0$ ) and  $\beta$  ( $x = 1$ ). Figure 1 shows the flowchart of the evolutionary construction technique. At the preliminary generation ( $n_{\text{gen}} = 0$ ), chemical compositions and crystal structures are generated by randomly mixing the two elements, and the structural optimizations are performed. Then, the formation energies from the simple substances are calculated, and the preliminary convex hull is constructed.

From the first generation ( $n_{\text{gen}} = 1$ ), two target compounds, T1 and T2, are selected from the compounds which emerge on or near the convex hull. This means that our method needs no setting of the population size. In Ref. [15], the authors reported that more than 80% of experimentally synthesized compounds included in the Inorganic Crystal Structure Database (ICSD) have the distance from the convex hull ( $\Delta H$ ) less than 36 meV/atom ( $= 2.65$  mRy/atom). This error is associated with the approximations used in first-principles calculations and the omission of temperature effects. Therefore, T1 and T2 should be selected from the compounds with  $\Delta H < 2.65$  mRy/atom. Then, the evolutionary operator, mutation, is applied to each of T1 and T2, as shown in Fig. 2. The mutation consists of seven operators as follows: permutation, distortion, reflection, modulation, addition, elimination, and substitution. The permutation is the operator that the positions are exchanged between different atomic species. The distortion is the operator that the lattice

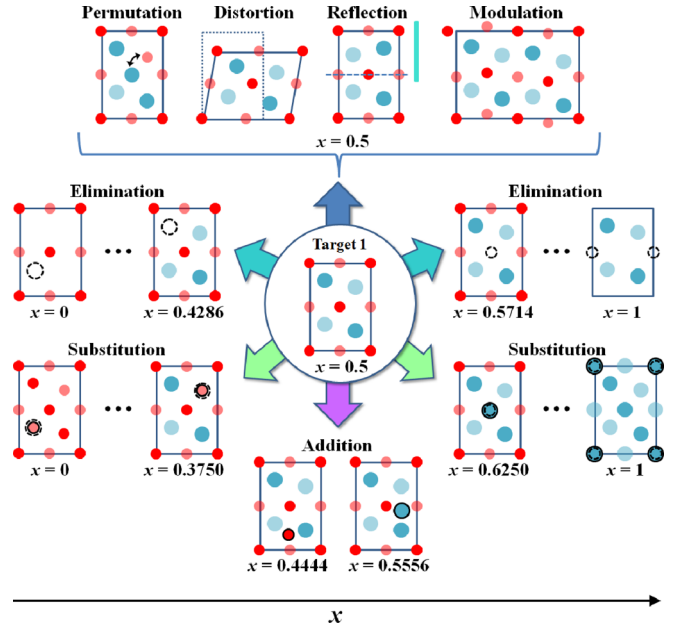


FIG. 2. Application of the mutation to a target compound, which consists of seven operators, permutation, distortion, reflection, modulation, addition, elimination, and substitution. The composition  $x$  is defined as  $N_l/(N_s + N_l)$ , where  $N_s$  ( $N_l$ ) is the number of the small (large) balls included in the cell.

parameters, i.e., three lengths and three angles, are randomly changed. The reflection is the operator that atomic positions are partially converted according to a reflection about a mirror on the  $ab$ ,  $bc$ , or  $ca$  plane. The modulation is the operator that a zigzag modulation is randomly provided to the calculation cell doubled along the axis with the smallest length of the three, which is effective for the discovery of long-period structures such as modulated structures observed in elemental metals under high pressure [16–18]. In contrast to the above four operators, the addition, elimination, and substitution are the operators varying the chemical composition  $x$ . The addition is the operator that the atom is randomly added in the calculation cell. The elimination is the operator that crystal structures with a wide variety of  $x$  are created by step-by-step eliminating the atoms in the cell. Similarly, the substitution is the operator that the structures are created by step-by-step substituting the atomic species. For example, in the case of Fig. 2,  $x$  is gradually decreased (increased) to 0 (1) by eliminating the large (small) balls or substituting the large (small) ones for the small (large) ones. Especially, these operators are effective for the exploration with respect to the case that the structure is basically similar, although the chemical composition is different, e.g., low-, intermediate-, and high- $T_c$  superconducting phases predicted in compressed sulfur hydrides [19,20].

Next, the adaptive mutation and the mating are performed, in which the candidates are created using the structural information about both T1 and T2 (Fig. 3). The adaptive mutation is a controversial theory of biological evolution [21,22] and a hypothesis that the mutations are much less random and more purposeful than those considered in traditional evolutionary theory. We adopt this idea in our composition search algorithm, in which the adaptive mutation is defined

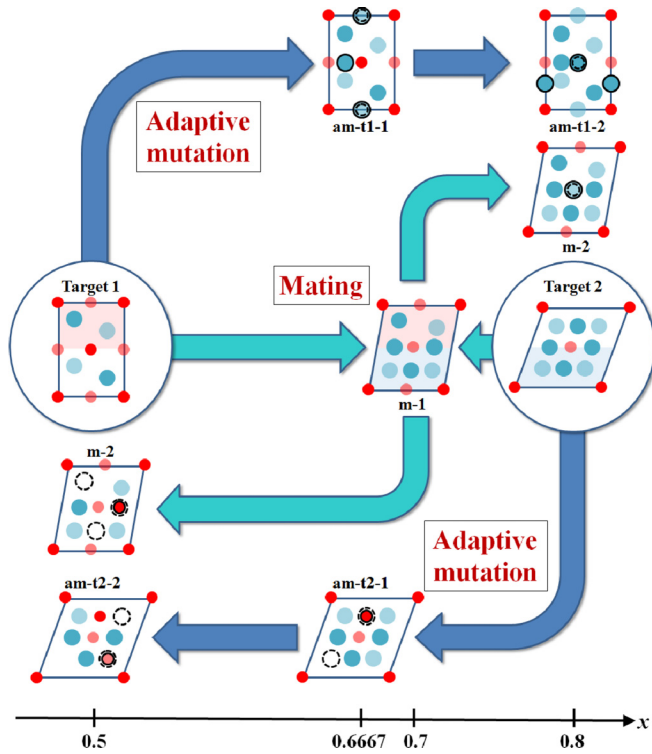


FIG. 3. Application of the adaptive mutation and the mating to two target compounds. The composition  $x$  is defined as  $N_l/(N_s + N_l)$ , where  $N_s$  ( $N_l$ ) is the number of the small (large) balls included in the cell. The added, eliminated, and substituted atoms are represented as the solid, broken, and solid-broken circles, respectively. The structures created by the adaptive mutation from T1 to T2, that from T2 to T1, and the mating are labeled by “am-t1-,” “am-t2-,” and “m-,” respectively.

as the operator that compositions are created not randomly but sequentially based on the structural information with respect to two stable substances. This is considered as the constraint for the choice of probable compositions in significantly large search space. The structures are created by step-by-step transforming T1 (T2) until the composition  $x$  and the number of the formula unit included in the calculation cell,  $n_{f.u.}$ , coincide with those of T2 (T1). Here, we represent variations of the number of the  $\alpha$  and  $\beta$  atoms according to the transformation as  $\Delta N_\alpha$  and  $\Delta N_\beta$ , respectively. The number of the adaptive mutation steps,  $n_{step}$ , is defined as the greatest common divisor (gcd) of the absolute values of  $\Delta N_\alpha$  and  $\Delta N_\beta$ , i.e.,  $n_{step} = \text{gcd}(|\Delta N_\alpha|, |\Delta N_\beta|)$ . The number of the atoms in the calculation cell is varied by  $\delta N_\alpha = \Delta N_\alpha/n_{step}$  for  $\alpha$  and  $\delta N_\beta = \Delta N_\beta/n_{step}$  for  $\beta$  at each step. The atoms in the unit cell are operated according to five cases determined by the condition of  $\delta N_\alpha$  and  $\delta N_\beta$ , as shown in Table I. For example, in Fig. 3,  $\delta N_\alpha$  (small ball) and  $\delta N_\beta$  (large ball) are  $-1$  and  $2$  for the transformation from T1 to T2, respectively. The value of  $\delta N_\alpha + \delta N_\beta$  is equal to  $1$ , which is classified into the case (iii)-(b); first a large ball is substituted for a small ball selected randomly, and then a large ball is randomly added. At this stage, the composition is increased to  $x = 0.6667$ , which is the intermediate one between T1 and T2. Performing the similar procedure again, a candidate structure

TABLE I. Operations for the atoms in the calculation cell with respect to the adaptive mutation.

Case	Operation
(i) $\delta N_\alpha \geq 0 \cap \delta N_\beta \geq 0$	addition
(ii) $\delta N_\alpha \leq 0 \cap \delta N_\beta \leq 0$	elimination
(iii) others	(a) $\delta N_\alpha + \delta N_\beta = 0$ substitution
	(b) $\delta N_\alpha + \delta N_\beta > 0$ sub. + add.
	(c) $\delta N_\alpha + \delta N_\beta < 0$ sub. + elm.

with  $x = 0.8$  is created, based on the structural information of T1. In contrast, when the transformation starts from T2, corresponding to the case (iii)-(c), candidate structures with  $x = 0.6667$  and  $0.5$  are created by the random elimination of the large balls in addition to the substitution, based on the structural information of T2.

The mating is the operator that a structure is created by combining T1 and T2, which is similar to the “heredity” operator proposed by Oganov *et al.* [9]. First, each structure is cut into two at the plane, which passes through a point  $s$  ( $0 \leq s < 1$ ) on the axis randomly selected and is parallel to the plane formed by the other two axes. Then, the  $[0, s)$  region for T1 and the  $[s, 1)$  region for T2 are combined. The calculation cell of the combined structure is obtained by mixing the lattice parameters of T1 and those of T2 with a rate of  $s : 1 - s$ . In this way, a candidate structure is created with a composition different from T1 and T2, and further two structures are created from the structure, as with the case of the adaptive mutation (Fig. 3).

Note that there are three cases with respect to  $n_{step}$ :  $n_{step} \geq 2$ ,  $n_{step} = 1$ , and  $n_{step} = 0$  (Fig. 1). For the first case, both the adaptive mutation and the mating are applied as shown in Fig. 3. For the second case, cell expansion is performed before the application of the adaptive mutation and mating because no intermediate chemical compositions between T1 and T2 are created by the procedure mentioned above. First the calculation cell is doubled along the axis with the smallest length of the three for both T1 and T2, and then the adaptive mutation and mating are applied to the expanded cells. This procedure achieves the increase of  $n_{step}$  to 2 and the creation of the intermediate compounds between T1 and T2. For the third case, only the mating is applied because T1 has the same  $x$  and  $n_{f.u.}$  as T2 and the adaptive mutation is unnecessary.

For the case of T1 and T2 shown in Figs. 2 and 3, 55 structures are created by the evolutionary operations: 22 by the mutation for T1, 26 by the mutation for T2, 4 by the adaptive mutation, and 3 by the mating. If computational resources are still available, further structures can be created using other T1 and T2. Then, the structural optimizations are performed for all the created structures, and the convex hull is updated to determine whether the created compounds are favorable. We define the update of the convex hull as “generation” in this method.

Our evolutionary construction technique can be applied to not only a binary but also a ternary or multinary system. Although the division into the cases shown in Table I is more complicated, the same operators are available. For a ternary or multinary system, the search space is significantly

enlarged due to the existence of large number of possible compositions. USPEX and CALYPSO show a marked tendency to a local search (exploitation) with respect to the composition space because the search for stable compounds is intensively carried out on the line connecting between the selected substances. On the other hand, AIRSS shows the tendency to a global search (exploration) because different compositions are randomly created. Therefore, in all three cases, a huge number of the structural optimizations is required for the discovery of thermodynamically stable compounds with respect to the ternary or multinary system. The most notable feature of our evolutionary method is that potential stable compounds are searched for with a balance of the exploitation and the exploration using the structural information of stable substances; the exploitation is achieved by the permutation, distortion, reflection, modulation, addition, adaptive mutation, and mating, and the exploration is achieved by the elimination and substitution. This enables us to more efficiently obtain an overview of the convex hull.

### III. APPLICATION TO CARBON-HYDROGEN BINARY SYSTEM

We applied the evolutionary construction technique to the carbon-hydrogen binary system ( $C_{1-x}H_x$ ) under high pressure. Carbon, hydrogen, and hydrocarbon have profoundly affected humankind, and the knowledge of thermodynamically stable phases of  $C_{1-x}H_x$  system is important for materials science, earth and planetary science, life science, and so on. Gao *et al.* predicted that methane,  $CH_4$  ( $x = 0.8$ ), dissociates into ethane,  $C_2H_6$  ( $x = 0.75$ ), at 95 GPa, butane,  $C_4H_{10}$  ( $x = 0.7143$ ), at 158 GPa, and diamond ( $x = 0$ ) at 287 GPa at zero temperature [23]. Liu *et al.* searched for stable phases of the binary system in the pressure range of 100–300 GPa and predicted that ethylene,  $C_2H_4$  ( $x = 0.6667$ ),  $C_2H_6$ , and  $CH_4$  emerge on the convex hull at 100 GPa, and  $CH_4$  and  $C_2H_6$  become unstable at 100 GPa and 200 GPa, respectively [24]. These results indicate that hydrogen-rich hydrocarbons get to be unstable with the increase of pressure. In contrast, there is a possibility that novel hydrocarbons are stabilized in the low pressure region, and we explored them by performing the evolutionary construction technique at the pressure of 10 GPa.

First, we developed the calculation code of the evolutionary construction technique following the flowchart shown in Fig. 1 and combined it with the QUANTUM ESPRESSO (QE) code [25] to perform the optimization for the structures created by the evolutionary operators. In this study, we intentionally used no experimental and theoretical results on the stable compositions and structures reported earlier in order to verify the prediction ability of our search technique. For the preliminary generation ( $n_{gen} = 0$ ), we prepared for the calculation cell including 10 carbon atoms and step-by-step substituted a hydrogen atom for a carbon one. In this way,  $x$  was increased from 0 to 1 with an interval of 0.1, and 11 compositions were created. The lattice parameters and atomic positions were randomly generated at each  $x$ . The number of the atoms in the calculation cell, which is increased by the addition, the mating, the modulation, and the cell expansion, was limited to less than 80. The atoms were displaced by  $\pm 0.05$  in fractional coordinates with respect to the modulation. The

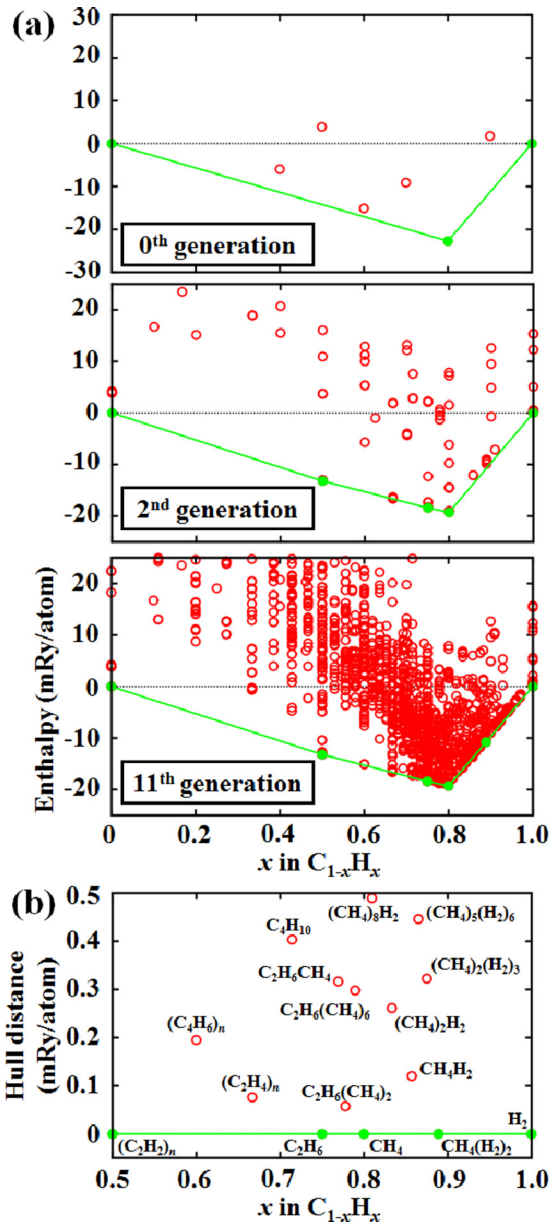


FIG. 4. (a) Evolution of formation-energy convex hull with respect to  $C_{1-x}H_x$  at 10 GPa. (b) Compounds with the convex hull distance less than 0.5 mRy/atom.

pressure was set at 10 GPa. We used Perdew, Burke, and Ernzerhof [27] for the exchange-correlation functional, and the Rabe-Rappe-Kaxiras-Joannopoulos ultrasoft pseudopotential [28]. The  $k$ -space integration over the Brillouin zone (BZ) was performed on a  $16 \times 16 \times 16$  grid for the compounds with 1–4 atoms in the calculation cell,  $12 \times 12 \times 12$  for 5–8 atoms, and  $8 \times 8 \times 8$  for more than 8 atoms. The energy cutoffs were set at 100 Ry for wave function and 800 Ry for charge density, respectively.

Figure 4(a) shows the evolution of the formation enthalpy convex hull with respect to the composition  $x$  in  $C_{1-x}H_x$ . The convex hull is roughly converged at the second generation by performing 169 structural optimizations, in which the compounds with  $x = 0.5$ ,  $0.75$ , and  $0.8$  appear on the convex hull. Then, the compound with  $x = 0.8889$  emerges

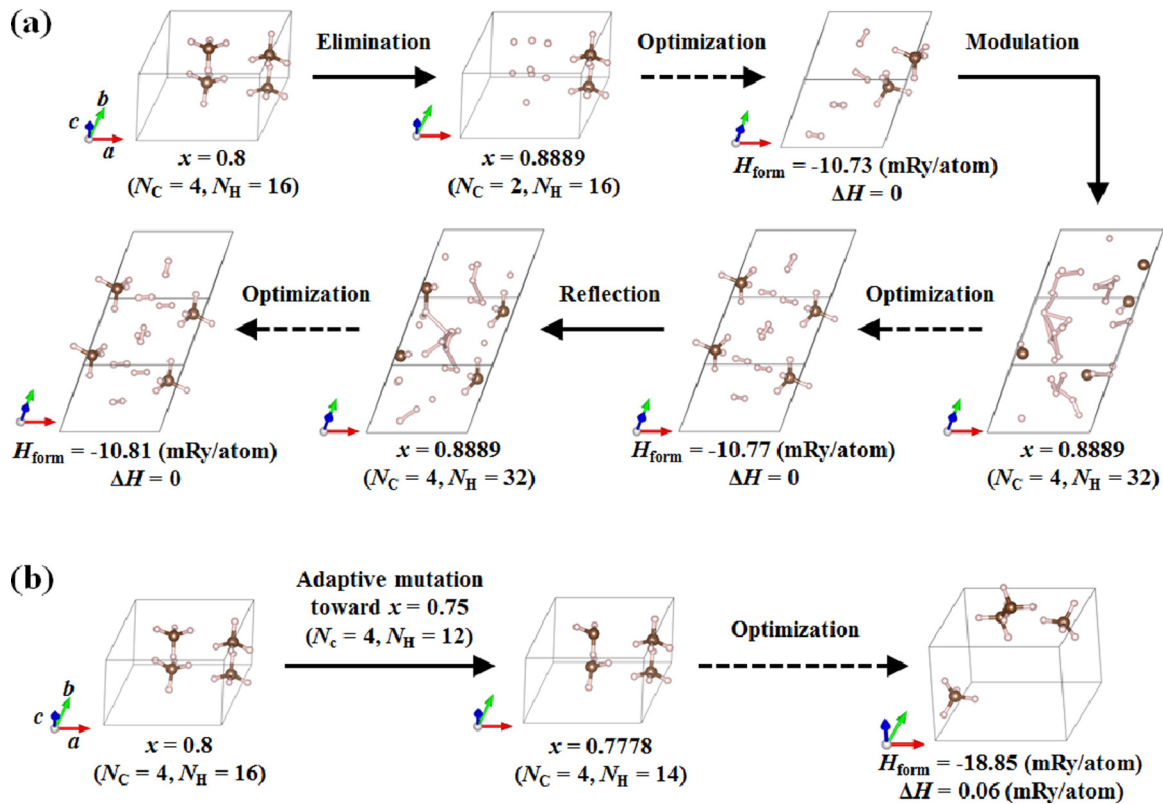


FIG. 5. (a) Formation process of the compound with  $x = 0.8889$  obtained through the evolutionary construction technique.  $N_C$  ( $N_H$ ),  $H_{\text{form}}$ , and  $\Delta H$  represent the number of carbon (hydrogen) atoms in the calculation cell, the formation enthalpy, and the convex hull distance, respectively. At the sixth generation, the two carbon atoms in the compound with  $x = 0.8$  are eliminated from the cell by the evolutionary operator “elimination,” and four  $\text{H}_2$  molecules are formed by the structural optimization. At the seventh generation,  $N_C$  and  $N_H$  are doubled and the atoms are displaced by the “modulation” along the  $b$  axis, and four  $\text{CH}_4$  molecules and eight  $\text{H}_2$  molecules are moved to more stable positions by the optimization. At the tenth generation, the positions of the atoms included in the region of  $0 \leq a < 0.5$  are converted by “reflection” about a mirror on the  $ab$  plane, and then the most stable structure is obtained by the optimization, in which the hydrogen atoms are slightly displaced from the positions before the reflection. (b) Formation process of the compound with  $x = 0.7778$ . At the fifth generation, two hydrogen atoms in the compound with  $x = 0.8$  are eliminated by the adaptive mutation toward another stable compound with  $x = 0.75$  ( $N_C = 4$  and  $N_H = 12$ ), and two  $\text{CH}_4$  molecules and a  $\text{C}_2\text{H}_6$  molecule are formed by the optimization. Crystal structures were drawn with VESTA [26].

on the convex hull at the sixth generation (1047 optimizations). The compound is first created by applying the operator “elimination” to the compound with  $x = 0.8$  including four  $\text{CH}_4$  molecules in the calculation cell and then is refined by applying “modulation” at the seventh generation and “reflection” at the tenth generation [Fig. 5(a)]. As a metastable compound with the small hull distance, we obtained the compound with  $x = 0.7778$  by “adaptive mutation” from the compound with  $x = 0.8$  ( $N_C = 4$  and  $N_H = 16$ ) toward that with  $x = 0.75$  ( $N_C = 4$  and  $N_H = 12$ ), where  $N_C$  ( $N_H$ ) represents the number of carbon (hydrogen) atoms in the calculation cell [Fig. 5(b)]. The result at the 11th generation (2528 optimizations) shows that the hull distance  $\Delta H$  is gradually decreased with the increase of  $x$  and many hydrogen-rich compounds are close to the convex hull, which implies that hydrogen-rich compounds are more stable than carbon-rich ones at 10 GPa.

Figure 4(b) shows the compounds included in the region that the hull distance  $\Delta H$  is less than 0.5 mRy/atom, which is less than one fifth of the tolerance stated above, 2.65 mRy/atom [15]. All the compounds with  $\Delta H <$

0.5 mRy/atom, i.e., 15 hydrocarbons, are listed in Table II. The compounds with  $x = 0.5, 0.6,$  and  $0.6667$  are formed by polymerization of  $\text{C}_2\text{H}_2$ ,  $\text{C}_4\text{H}_6$ , and  $\text{C}_2\text{H}_4$  molecules, respectively. Here, we indicate the polymerization with the subscript  $n$ .  $(\text{C}_2\text{H}_2)_n$  takes a trigonal  $P-3m1$  structure polymerized two dimensionally in the  $ab$  plane [Fig. 6(a)], which has been known as the chair-type graphane [29].  $(\text{C}_4\text{H}_6)_n$  (polybutadiene) and  $(\text{C}_2\text{H}_4)_n$  (polyethylene) take a monoclinic  $C2/m$  polymerized along the  $a$  axis [Fig. 6(b)] and a monoclinic  $P2_1/m$  polymerized along the  $b$  axis [Fig. 6(c)], respectively. The structural parameters are listed in Table III. The compounds with  $0.7143 \leq x \leq 1$  are formed by the condensation of isolated  $\text{C}_4\text{H}_{10}$ ,  $\text{C}_2\text{H}_6$ ,  $\text{CH}_4$ , and  $\text{H}_2$  molecules. The structural parameters are listed in Table IV.  $\text{C}_4\text{H}_{10}$  and  $\text{C}_2\text{H}_6$  take a triclinic  $P-1$  [Fig. 7(a)] and a monoclinic  $C2/c$  [Fig. 7(b)], respectively. In the region of  $0.75 < x < 0.8$ , molecular compounds of  $\text{C}_2\text{H}_6$  and  $\text{CH}_4$  are thermodynamically stable, and especially the compound with  $\text{C}_2\text{H}_6 : \text{CH}_4 = 1 : 2$ , i.e.,  $\text{C}_2\text{H}_6(\text{CH}_4)_2$ , with a triclinic  $P-1$  [Fig. 7(c)] is energetically competing with the dissociated state,  $\text{C}_2\text{H}_6 + \text{CH}_4$ :  $\Delta H = 0.06$  mRy/atom.  $\text{CH}_4$  with  $x = 0.8$  takes a monoclinic  $C2/c$

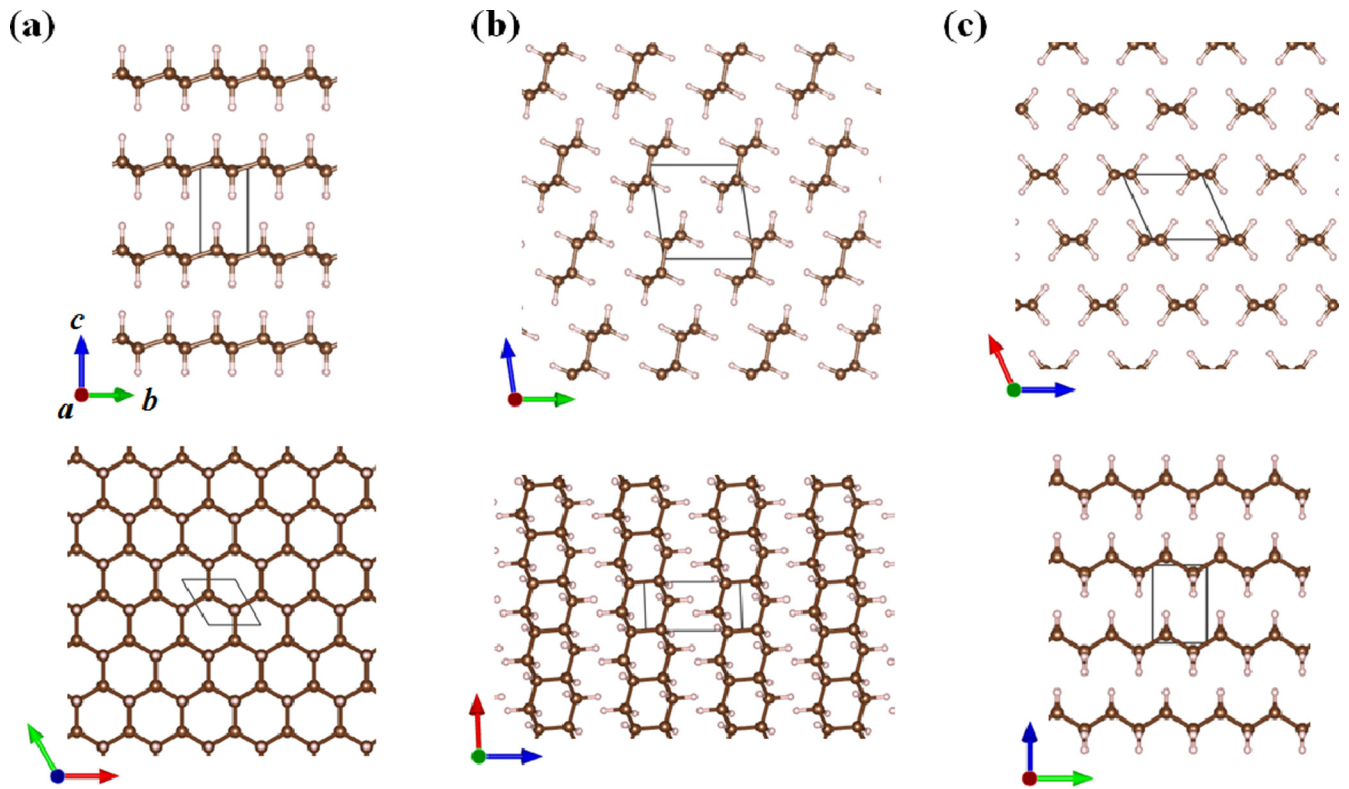


FIG. 6. Crystal structures at 10 GPa: (a)  $(\text{C}_2\text{H}_2)_n$ , (b)  $(\text{C}_4\text{H}_6)_n$ , and (c)  $(\text{C}_2\text{H}_4)_n$ . Large and small balls represent carbon and hydrogen atoms, respectively. Crystal structures were drawn with VESTA [26].

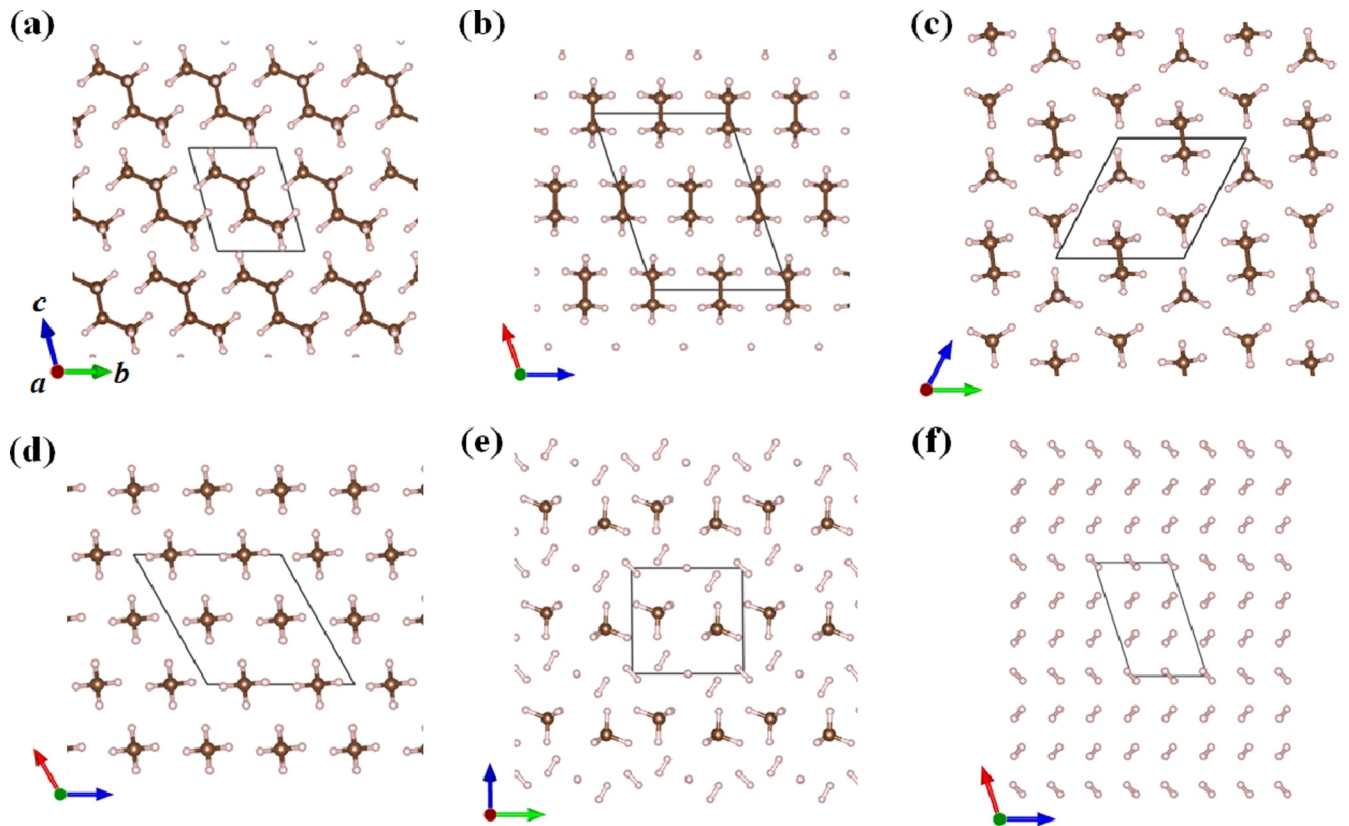


FIG. 7. Crystal structures at 10 GPa: (a)  $\text{C}_4\text{H}_{10}$ , (b)  $\text{C}_2\text{H}_6$ , (c)  $\text{C}_2\text{H}_6(\text{CH}_4)_2$ , (d)  $\text{CH}_4$ , (e)  $\text{CH}_4(\text{H}_2)_2$ , and (f)  $\text{H}_2$ . Large and small balls represent carbon and hydrogen atoms, respectively. Crystal structures were drawn with VESTA [26].

TABLE II. Compounds with the hull distance  $\Delta H$  less than 0.5 mRy/atom.

$x$			$\Delta H$ (mRy/atom)	Band gap (eV)
0	C	diamond	0	4.3
0.5	(C <sub>2</sub> H <sub>2</sub> ) <sub>n</sub>	graphane	0	4.7
0.6	(C <sub>4</sub> H <sub>6</sub> ) <sub>n</sub>	polybutadiene	0.19	6.1
0.6667	(C <sub>2</sub> H <sub>4</sub> ) <sub>n</sub>	polyethylene	0.08	6.2
0.7143	C <sub>4</sub> H <sub>10</sub>	butane	0.40	6.7
0.75	C <sub>2</sub> H <sub>6</sub>	ethane	0	7.7
0.7692	C <sub>2</sub> H <sub>6</sub> CH <sub>4</sub>	eth. + met.	0.32	8.0
0.7778	C <sub>2</sub> H <sub>6</sub> (CH <sub>4</sub> ) <sub>2</sub>	eth. + met.	0.06	7.9
0.7895	C <sub>2</sub> H <sub>6</sub> (CH <sub>4</sub> ) <sub>6</sub>	eth. + met.	0.30	8.1
0.8	CH <sub>4</sub>	methane	0	8.3
0.8095	(CH <sub>4</sub> ) <sub>8</sub> H <sub>2</sub>	met. + hyd.	0.49	8.0
0.8333	(CH <sub>4</sub> ) <sub>2</sub> H <sub>2</sub>	met. + hyd.	0.26	8.3
0.8571	CH <sub>4</sub> H <sub>2</sub>	met. + hyd.	0.12	8.2
0.8649	(CH <sub>4</sub> ) <sub>5</sub> (H <sub>2</sub> ) <sub>6</sub>	met. + hyd.	0.45	8.4
0.875	(CH <sub>4</sub> ) <sub>2</sub> (H <sub>2</sub> ) <sub>3</sub>	met. + hyd.	0.32	7.8
0.8889	CH <sub>4</sub> (H <sub>2</sub> ) <sub>2</sub>	met. + hyd.	0	7.9
1	H <sub>2</sub>	hydrogen	0	7.3

structure [Fig. 7(d)]. Similarly, in the region of  $0.8 < x < 1$ , molecular compounds of CH<sub>4</sub> and H<sub>2</sub> (hydrogen) are stabilized, and especially CH<sub>4</sub>(H<sub>2</sub>)<sub>2</sub> with CH<sub>4</sub> : H<sub>2</sub> = 1 : 2, which takes a triclinic *P*-1 [Fig. 7(e)], emerges on the convex hull. For H<sub>2</sub> with  $x = 1$ , we obtained a structure with a monoclinic *P*2<sub>1</sub>/*c* [Fig. 7(f)], in which the orientation of the H<sub>2</sub> molecules is different from those of the structures predicted earlier [30,31].

Our results suggest that, in the region of  $\Delta H < 0.50$  mRy/atom, the most hydrogen-rich hydrocarbon is CH<sub>4</sub>(H<sub>2</sub>)<sub>2</sub> ( $x = 0.8889$ ), which is qualitatively consistent with the results of earlier experiments [32] and recent first-principles calculations [33]. The compounds with  $x > 0.8889$  is thermodynamically unstable, and  $\Delta H$  is increased with increase of  $x$ . The most and second most hydrogen-rich compounds created in this study are CH<sub>4</sub>(H<sub>2</sub>)<sub>15</sub> with CH<sub>4</sub> : H<sub>2</sub> = 1 : 15 ( $x = 0.9714$ ) and CH<sub>4</sub>(H<sub>2</sub>)<sub>14</sub> with CH<sub>4</sub> : H<sub>2</sub> = 1 : 14 ( $x = 0.9697$ ), and they show  $\Delta H$  of 1.32 mRy/atom and 0.78 mRy/atom, respectively. We investigated the band gaps of 15 hydrocarbons (see Table II). The band gap increases from 4.3 to 7.7 eV with increase of  $x$  from 0 to 0.75 and varies in the range of 7.9–8.1 eV for the molecular compounds

TABLE III. Structural parameters of (C<sub>4</sub>H<sub>6</sub>)<sub>n</sub> and (C<sub>2</sub>H<sub>4</sub>)<sub>n</sub>.

		Cell (Å, °)	Atomic position
(C <sub>4</sub> H <sub>6</sub> ) <sub>n</sub>	<i>C</i> 2/ <i>m</i>	<i>a</i> 9.2041	C 4i 0.5266 0 0.1569
		<i>b</i> 2.5128	C 4i 0.8869 0 0.7557
		<i>c</i> 4.8316	H 4i 0.4284 0 0.2612
		$\beta$ 98.70	H 4i 0.1485 0 0.4707
			H 4i 0.7858 0 0.8510
(C <sub>2</sub> H <sub>4</sub> ) <sub>n</sub>	<i>P</i> 2 <sub>1</sub> / <i>m</i>	<i>a</i> 3.6115	C 2e 0.4983 0.25 0.3946
		<i>b</i> 2.5210	H 2e 0.7594 0.25 0.3224
		<i>c</i> 3.9636	H 2e 0.2296 0.25 0.1272
		$\beta$ 114.82	

TABLE IV. Structural parameters of C<sub>4</sub>H<sub>10</sub>, C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>6</sub>(CH<sub>4</sub>)<sub>2</sub>, CH<sub>4</sub>, CH<sub>4</sub>(H<sub>2</sub>)<sub>2</sub>, and H<sub>2</sub>.

		Cell (Å, °)	Atomic position
C <sub>4</sub> H <sub>10</sub>	<i>P</i> -1	<i>a</i> 3.7519	C 2i 0.9532 0.2335 0.7636
		<i>b</i> 4.5375	C 2i 0.1387 0.5177 0.6456
		<i>c</i> 5.3520	H 2i 0.2062 0.7489 0.7741
		$\alpha$ 94.70	H 2i 0.6742 0.2247 0.7914
		$\beta$ 109.37	H 2i 0.1777 0.2493 0.9618
		$\gamma$ 112.77	H 2i 0.1432 0.0011 0.3711
			H 2i 0.5560 0.4574 0.3470
C <sub>2</sub> H <sub>6</sub>	<i>C</i> 2/ <i>c</i>	<i>a</i> 8.6551	C 8f 0.5835 0.0919 0.0364
		<i>b</i> 3.3924	H 8f 0.6038 0.2681 0.9023
		<i>c</i> 6.2579	H 8f 0.1788 0.3679 0.0847
		$\beta$ 108.70	H 8f 0.8990 0.2138 0.8189
C <sub>2</sub> H <sub>6</sub> (CH <sub>4</sub> ) <sub>2</sub>	<i>P</i> -1	<i>a</i> 3.4521	C 2i 0.0795 0.4212 0.1286
		<i>b</i> 5.6046	C 2i 0.3595 0.1591 0.6933
		<i>c</i> 5.8933	H 2i 0.2635 0.5498 0.1716
		$\alpha$ 61.74	H 2i 0.7505 0.6406 0.3646
		$\beta$ 85.06	H 2i 0.3245 0.8289 0.3165
		$\gamma$ 89.97	H 2i 0.8449 0.3304 0.2843
			H 2i 0.7569 0.9805 0.1095
			H 2i 0.2560 0.2577 0.1337
			H 2i 0.2733 0.0858 0.5636
CH <sub>4</sub>	<i>C</i> 2/ <i>c</i>	<i>a</i> 6.2809	C 4e 0 0.0964 0.25
		<i>b</i> 3.3144	H 8f 0.8386 0.0937 0.6830
		<i>c</i> 6.1895	H 8f 0.4831 0.2144 0.5992
		$\beta$ 119.59	
CH <sub>4</sub> (H <sub>2</sub> ) <sub>2</sub>	<i>P</i> -1	<i>a</i> 3.7932	C 2i 0.7066 0.7632 0.4235
		<i>b</i> 5.1936	H 2i 0.5864 0.3091 0.1891
		<i>c</i> 5.2688	H 2i 0.6763 0.9573 0.3438
		$\alpha$ 90.46	H 2i 0.1761 0.2273 0.3559
		$\beta$ 111.02	H 2i 0.1048 0.3459 0.6505
		$\gamma$ 90.09	H 2i 0.9818 0.0495 0.9461
			H 2i 0.5690 0.3301 0.6422
	H 2i 0.9006 0.5007 0.9977		
		H 2i 0.4611 0.7652 0.9402	
H <sub>2</sub>	<i>P</i> 2 <sub>1</sub> / <i>c</i>	<i>a</i> 5.4733	H 4e 0.9589 0.0649 0.5394
		<i>b</i> 3.4470	H 4e 0.7078 0.5718 0.4044
		<i>c</i> 3.4760	H 4e 0.3733 0.5759 0.7347
		$\beta$ 108.39	

of C<sub>2</sub>H<sub>6</sub> and CH<sub>4</sub> and in the range of 7.9–8.4 eV for the molecular compounds of CH<sub>4</sub> and H<sub>2</sub>.

Finally, we compare the present results with the previous ones. Figure 8 shows the convex hull of the C-H binary system at 10 GPa, which was developed by optimizing the earlier-reported substances [23,29,31,33–36]. C, (C<sub>2</sub>H<sub>2</sub>)<sub>n</sub>, C<sub>2</sub>H<sub>6</sub>, CH<sub>4</sub>, (CH<sub>4</sub>)<sub>2</sub>H<sub>2</sub>, and H<sub>2</sub> take the vertices of the convex hull, and (C<sub>2</sub>H<sub>4</sub>)<sub>n</sub>, (CH<sub>4</sub>)<sub>2</sub>(H<sub>2</sub>)<sub>3</sub>, and CH<sub>4</sub>(H<sub>2</sub>)<sub>2</sub> are metastable compounds. We plotted the data of all the compounds shown in Fig. 4(b) and confirmed that all the earlier-reported compounds are well reproduced by our evolutionary search, except for crystal structures. Table V shows the comparison of the structures between our results and previous ones. The structures of C and (C<sub>2</sub>H<sub>2</sub>)<sub>n</sub> are completely consistent with the previous ones. Besides, we found structures being more stable by 40  $\mu$ Ry/atom for (C<sub>2</sub>H<sub>4</sub>)<sub>n</sub> and by 68  $\mu$ Ry/atom for CH<sub>4</sub>(H<sub>2</sub>)<sub>2</sub> than previous ones. For CH<sub>4</sub>, our predicted *C*2/*c* and the earlier-reported *P*2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> [23] are energetically

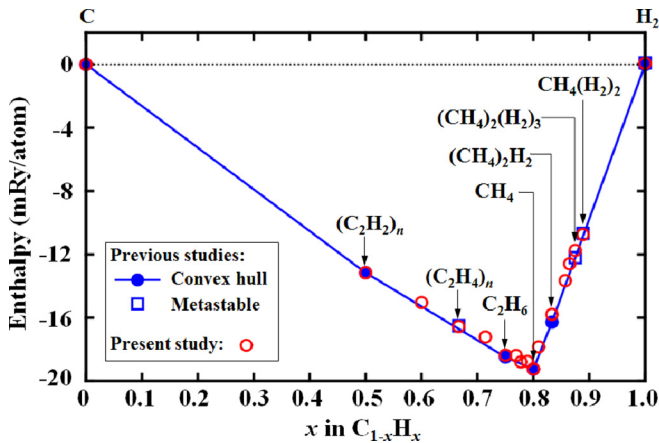


FIG. 8. Formation enthalpy convex hull of C-H binary system at 10 GPa.

competing, and  $C2/c$  is very slightly unstable only by 12  $\mu\text{Ry/atom}$ . For  $\text{H}_2$ , our predicted  $P2_1/c$ , which includes 12 hydrogen atoms in the primitive cell, is slightly more stable by 29  $\mu\text{Ry/atom}$  than the earlier-predicted  $P2_1/c$  including 24 atoms [31], whereas it is slightly more unstable by 72  $\mu\text{Ry/atom}$  than  $Pa-3$  observed by the experiments [36]. For  $\text{C}_2\text{H}_6$ ,  $(\text{CH}_4)_2\text{H}_2$ , and  $(\text{CH}_4)_2(\text{H}_2)_3$ , the most stable structures or energetically competing structures were failed to be obtained by the search up to the 11th generation in the present study. The favorable structures, however, are expected to be found by further continuing the evolutionary search. Experiments have revealed that methane and hydrogen form four molecular compounds,  $(\text{CH}_4)_2\text{H}_2$  with  $x = 0.8333$ ,  $\text{CH}_4\text{H}_2$  with  $x = 0.8571$ ,  $\text{CH}_4(\text{H}_2)_2$  with  $x = 0.8889$ , and  $\text{CH}_4(\text{H}_2)_4$  with  $x = 0.9231$ , at pressure up to 10 GPa [32].  $(\text{CH}_4)_2\text{H}_2$ ,  $\text{CH}_4\text{H}_2$ , and  $\text{CH}_4(\text{H}_2)_2$  were obtained as stable compounds with  $\Delta H < 0.30$  mRy/atom at 10 GPa, whereas  $\text{CH}_4(\text{H}_2)_4$  was a more unstable one with  $\Delta H = 0.69$  mRy/atom. Another important result is that  $\text{C}_2\text{H}_6(\text{CH}_4)_2$  with  $x = 0.7778$ , found in the present study [Fig. 7(c)], is predicted to be a potential metastable compound with  $\Delta H = 0.12$  mRy/atom, which suggests that molecular compounds of ethane and methane can be synthesized at around 10 GPa, similar to the

TABLE V. Comparison of crystal structures between the present and previous studies.  $\Delta H_{\text{str}}$  shows the enthalpy of the present structure, relative to the previous one.

	Present	Previous	$\Delta H_{\text{str}}$ ( $\mu\text{Ry/atom}$ )
C	$Fd-3m$	$Fd-3m$	0
$(\text{C}_2\text{H}_2)_n$	$P-3m1$	$P-3m1$ [29]	0
$(\text{C}_2\text{H}_4)_n$	$P2_1/m$	$Cmcm$ [23]	-40
$\text{C}_2\text{H}_6$	$C2/c$	$P2_1/c$ [23]	+118
$\text{CH}_4$	$C2/c$	$P2_12_1$ [23]	+12
$(\text{CH}_4)_2\text{H}_2$	$P-1$	$P-1$ [34]	+489
$(\text{CH}_4)_2(\text{H}_2)_3$	$P-1$	$P-1$ [35]	+457
$\text{CH}_4(\text{H}_2)_2$	$P-1$	$P2_1/c$ [33]	-68
$\text{H}_2$	$P2_1/c$	$P2_1/c-24$ [31]	-29
		$Pa-3$ [36]	+72

molecular compounds of methane and hydrogen experimentally observed.

#### IV. CONCLUSION

We proposed the evolutionary construction technique of the formation-energy convex hull to search for thermodynamically stable phases in compounds. The potential candidates for the stable compounds are created by applying the three evolutionary operators, “mating,” “mutation” (permutation, distortion, reflection, modulation, addition, elimination, and substitution), and “adaptive mutation,” to two compounds on and near the convex hull. In other words, the compositions and the structures for the next generation are created using the structural information of the stable and metastable compounds at the present generation. Evolutionary algorithm usually requires setting many parameters such as heredity rate, permutation rate, mutation rate, population size, etc., to perform the structure search and tuning them to find the structures efficiently. Our technique reduces the burden on the parameter setting and tuning because the candidates are systematically created based on the rules shown in Figs. 2 and 3. It is important for search algorithms to achieve a balance between a global search (exploration) and a local search (exploitation). In our search technique, the elimination and the substitution achieve the exploration with respect to the composition search, and the others achieve the exploitation. We emphasize that this technique can be applied to not only binary but also ternary or multinary systems.

We applied the evolutionary construction technique to the C-H binary system at 10 GPa and searched for thermodynamically stable compounds. As a result, in addition to diamond and  $P2_1/c$   $\text{H}_2$ , we obtained 15 hydrocarbons with the convex hull distance less than 0.5 mRy/atom:  $(\text{C}_2\text{H}_2)_n$ ,  $(\text{C}_4\text{H}_6)_n$ ,  $(\text{C}_2\text{H}_4)_n$ ,  $\text{C}_4\text{H}_{10}$ ,  $\text{C}_2\text{H}_6$ ,  $\text{C}_2\text{H}_6\text{CH}_4$ ,  $\text{C}_2\text{H}_6(\text{CH}_4)_2$ ,  $\text{C}_2\text{H}_6(\text{CH}_4)_6$ ,  $\text{CH}_4$ ,  $(\text{CH}_4)_8\text{H}_2$ ,  $(\text{CH}_4)_2\text{H}_2$ ,  $\text{CH}_4\text{H}_2$ ,  $(\text{CH}_4)_5(\text{H}_2)_6$ ,  $(\text{CH}_4)_2(\text{H}_2)_3$ , and  $\text{CH}_4(\text{H}_2)_2$ , in which the earlier-reported compounds were well reproduced. These results indicate that hydrocarbons with  $x \leq 0.6667$  are polymerized, those with  $0.75 < x < 0.8$  form the molecular compounds of  $\text{C}_2\text{H}_6$  and  $\text{CH}_4$ , and those with  $0.8 < x < 1$  form the molecular compounds of  $\text{CH}_4$  and  $\text{H}_2$ . Our calculations show that  $\text{CH}_4(\text{H}_2)_2$  ( $x = 0.8889$ ) is the most hydrogen-rich compound at 10 GPa, which is qualitatively consistent with the previous experimental and first-principles results. The compounds with  $x > 0.8889$  are thermodynamically unstable and the hull distance increases with the increase of  $x$ . These results suggest that the evolutionary construction technique is useful to obtain the knowledge on thermodynamically stable compositions and structures.

#### ACKNOWLEDGMENTS

This study was supported by JSPS KAKENHI Scientific Research (C) (Grant Nos. 17K05541 and 20K03868) and Scientific Research (S) (No. 16H06345), and “Exploratory Challenge on Post-K Computer” (Frontiers of Basic Science: Challenging the Limits) and the Elements Strategy Initiative Center for Magnetic Materials (ESICMM) (No. JPMXP0112101004), through the Ministry of Education, Culture, Sports, Science and Technology (MEXT).



- [1] A. P. Drozdov, M. I. Erements, I. A. Troyan, V. Ksenofontov, and S. I. Shylin, *Nature (London)* **525**, 73 (2015).
- [2] M. Somayazulu, M. Ahart, A. K. Mishra, Z. M. Geballe, M. Baldini, Y. Meng, V. V. Struzhkin, and R. J. Hemley, *Phys. Rev. Lett.* **122**, 027001 (2019).
- [3] A. P. Drozdov, P. P. Kong, V. S. Minkov, S. P. Besedin, M. A. Kuzovnikov, S. Mozaffari, L. Balicas, F. F. Balakirev, D. E. Graf, V. B. Prakapenka *et al.*, *Nature (London)* **569**, 528 (2019).
- [4] B. Guigue, A. Marizy, and P. Loubeyre, *Phys. Rev. B* **95**, 020104(R) (2017).
- [5] A. F. Goncharov, S. S. Lobanov, V. B. Prakapenka, and E. Greenberg, *Phys. Rev. B* **95**, 140101(R) (2017).
- [6] Z. M. Geballe, H. Liu, A. K. Mishra, M. Ahart, M. Somayazulu, Y. Meng, M. Baldini, and R. J. Hemley, *Angew. Chem. Int. Ed.* **57**, 688 (2018).
- [7] Y. Sun, J. Lv, Y. Xie, H. Liu, and Y. Ma, *Phys. Rev. Lett.* **123**, 097001 (2019).
- [8] T. Ishikawa, T. Miyake, and K. Shimizu, *Phys. Rev. B* **100**, 174506 (2019).
- [9] A. R. Oganov and C. W. Glass, *J. Chem. Phys.* **124**, 244704 (2006).
- [10] A. R. Oganov (ed.), *Modern Methods of Crystal Structure Prediction* (Wiley-VCH, Weinheim, Germany, 2010).
- [11] A. R. Oganov, Y. Ma, A. O. Lyakhov, M. Valle, and C. Gatti, *Rev. Mineral. Geochem.* **71**, 271 (2010).
- [12] Y. Wang, J. Lv, L. Zhu, and Y. Ma, *Phys. Rev. B* **82**, 094116 (2010).
- [13] Y. Wang, J. Lv, L. Zhu, and Y. Ma, *Comput. Phys. Commun.* **183**, 2063 (2012).
- [14] C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter* **23**, 053201 (2011).
- [15] Y. Wu, P. Lazic, G. Hautier, K. Persson, and G. Ceder, *Energy Environ. Sci.* **6**, 157 (2013).
- [16] M. I. McMahon and R. J. Nelmes, *Chem. Soc. Rev.* **35**, 943 (2006).
- [17] T. Ishikawa, H. Nagara, K. Kusakabe, and N. Suzuki, *Phys. Rev. Lett.* **96**, 095502 (2006).
- [18] T. Ishikawa, A. Ichikawa, H. Nagara, M. Geshi, K. Kusakabe, and N. Suzuki, *Phys. Rev. B* **77**, 020101(R) (2008); **78**, 029901(E) (2008).
- [19] T. Ishikawa, A. Nakanishi, K. Shimizu, H. Katayama-Yoshida, T. Oda, and N. Suzuki, *Sci. Rep.* **6**, 23160 (2016).
- [20] R. Akashi, W. Sano, R. Arita, and S. Tsuneyuki, *Phys. Rev. Lett.* **117**, 075503 (2016).
- [21] J. Cairns, J. Overbaugh, and S. Miller, *Nature (London)* **335**, 142 (1988).
- [22] H. Hendrickson, E. S. Slechta, U. Bergthorsson, D. I. Andersson, and L. R. Roth, *Proc. Natl. Acad. Sci. USA* **99**, 2164 (2002).
- [23] G. Gao, A. R. Oganov, Y. Ma, H. Wang, P. Li, Y. Li, T. Iitaka, and G. Zou, *J. Chem. Phys.* **133**, 144508 (2010).
- [24] H. Liu, I. I. Naumov, and R. J. Hemley, *J. Phys. Chem. Lett.* **7**, 4218 (2016).
- [25] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Cereso, G. L. Chiarott, M. Cococcioni, I. Dabo *et al.*, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [26] K. Momma and F. Izumi, *J. Appl. Crystallogr.* **44**, 1272 (2011).
- [27] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [28] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, *Phys. Rev. B* **41**, 1227 (1990).
- [29] J. O. Sofo, A. S. Chaudhari, and G. D. Barber, *Phys. Rev. B* **75**, 153401 (2007).
- [30] C. J. Pickard and R. J. Needs, *Nat. Phys.* **3**, 473 (2007).
- [31] C. J. Pickard and R. J. Needs, *Phys. Status Solidi B* **246**, 536 (2009).
- [32] M. S. Somayazulu, L. W. Finger, R. J. Hemley, and H. K. Mao, *Science* **271**, 1400 (1996).
- [33] L. J. Conway and A. Hermann, *Geosciences* **9**, 227 (2019).
- [34] A. S. Naumova, S. V. Lepeshkin, and A. R. Oganov, *J. Phys. Chem. C* **123**, 20497 (2019).
- [35] G. Saleh and A. R. Oganov, *Sci. Rep.* **6**, 32486 (2016).
- [36] I. F. Silvera, *Rev. Mod. Phys.* **52**, 393 (1980).