Electronic Supplementary Information for Accelerating CALYPSO Structure Prediction by Data-driven Learning of Potential Energy Surface

Qunchao Tong,[1] Lantian Xue,[1] Jian Lv,[2,*] Yanchao Wang[1] and Yanming Ma[1,3*]

[1]*State Key Laboratory of Superhard Materials, College of Physics, Jilin University, Changchun 130012, China. E-mail*: mym@calypso.cn

[2]*College of Materials Science and Engineering, Jilin University, Changchun 130012, China. E-mail:* lvjian@calypso.cn

[3]*International Center of Future Science, Jilin University, Changchun 130012, China.*

Atom centered symmetry functions (ACSF)

The ACSF descriptor[1] is a set of radial and angular functions which describes the coordination environment of an atom depending on the positions of the neighboring atoms up to a cutoff radius R_{cut}. We here adopt radial functions

$$G_i^1 = \sum_j f_c\left(\boldsymbol{R}_{ij}\right),$$

$$G_i^2 = \sum_j e^{-\eta\left(\boldsymbol{R}_{ij}-\boldsymbol{R}_s\right)^2}.f_c\left(\boldsymbol{R}_{ij}\right),$$

and the angular function

$$G_i^3 = 2^{1-\zeta}\sum_{j,k\neq i}^{all}\left(1+\lambda\cos\theta_{ijk}\right)^{\zeta}.e^{-\eta\left(R_{ij}^2+R_{ik}^2+R_{jk}^2\right)}.f_c\left(R_{ij}\right).f_c\left(R_{ik}\right).f_c\left(R_{jk}\right),$$

where R_{ij} is the distance between atoms i and j, $\theta_{ijk} = a\cos\left(\boldsymbol{R}_{ij}\cdot\boldsymbol{R}_{ik}/R_{ij}\cdot R_{ik}\right)$ is the angle centered at atom i with neighboring atoms j and k, $\eta, R_s, \zeta, \lambda$ are adjustable parameters, the $f_c(R_{ij})$ is the cutoff function defined as

$$f_c\left(R\right) = \begin{cases}\left(cos\left(\dfrac{\pi R}{R_{cut}}\right)+1\right)\Big/2 & for\ R \le R_{cut}\\ 0 & for\ R>R_{cut}\end{cases}.$$

All G^1, G^2 and G^3 are many-body functions depending explicitly on the positions of

all atoms in the local environment and invariant to the choice of reference frame. An arbitrary number (typically 50-100) of symmetry function values can be obtained by adjusting $\eta, R_s, \zeta, \lambda, R_{cut}$ for atom *i*, which can be used to represent the environment of atom *i* in the sturcture and as input vectors $(\boldsymbol{d}_i)$ for the ML model.

Computational Details

The underlying energy calculations and local structure optimizations in the current work were performed at the DFT level using the plane wave basis set, projected augmented wave potentials,[2] and PBE exchange–correlation functional[3] through the Vienna ab initio simulation package.[4] The cut-off energy of 430 eV for wave-function expansion and Γ point for Brillouin zone sampling are chosen to ensure that the total energy converges to better than 1 meV per atom. A vacuum region of 13 Å is used to isolate B clusters to avoid their interactions with adjacent periodic images. The predicted low-lying isomers for B_{84} cluster were further optimized at the DFT level using the hybrid PBE0 and TPSSh functionals through the Gaussian 09 package.[5] An all-electron 6-31G(d) basis set was used.

Reference

1 J. Behler, *J. Chem. Phys.*, 2011, **134**, 74106.

2 P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953.

3 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

4 G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169.

5 M. J. Frisch, et al., Gaussian 09, revision C.01, Gaussian, Inc.,Wallingford, CT, 2009.

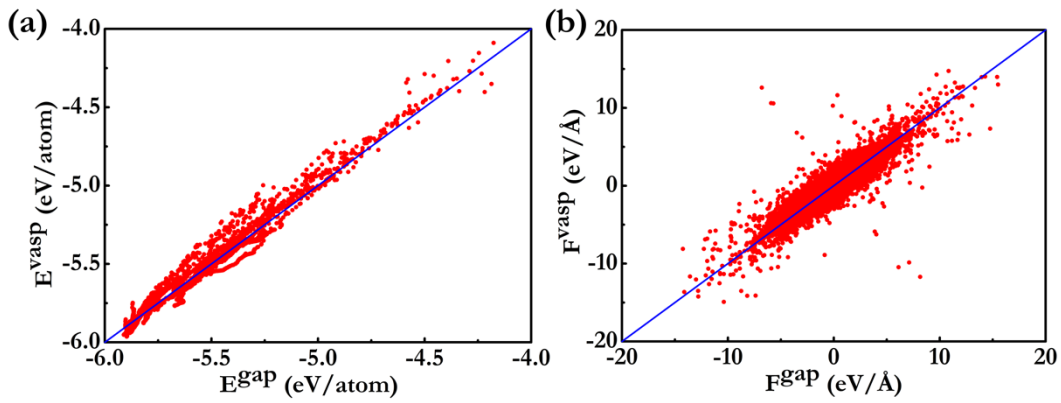

Fig. S1 Comparisons between GAP and DFT results of (a) energies and (b) forces for the testing set. The GAP is constructed from 12,000 training data (see main text).

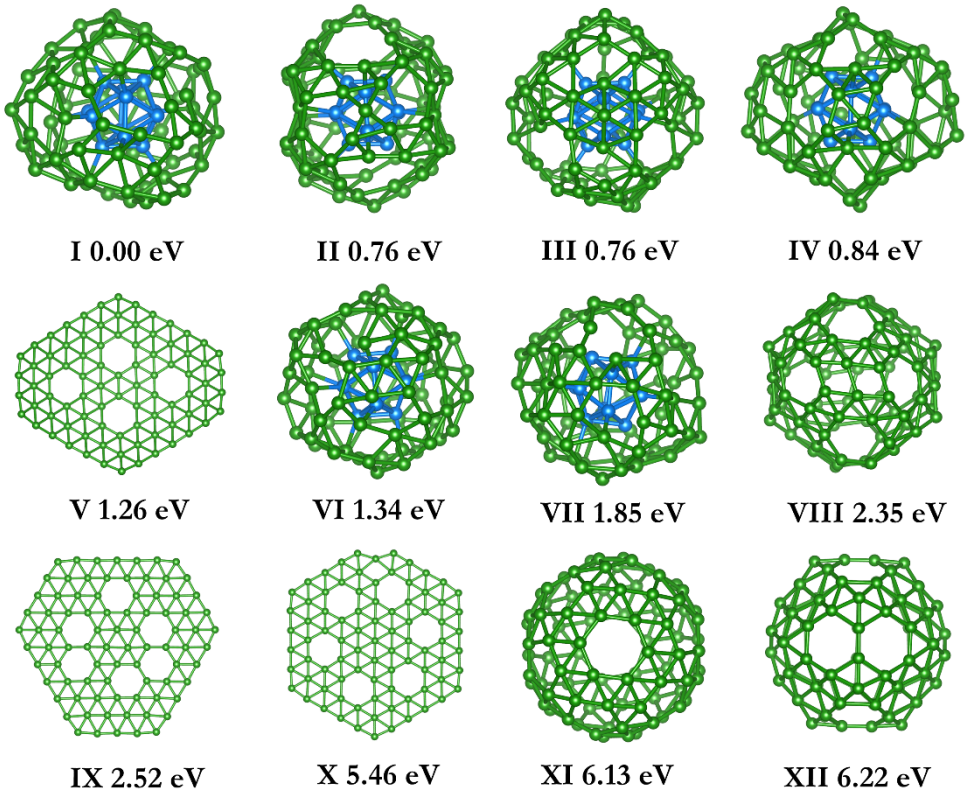

Fig. S2 Low-lying isomers of B_{84} cluster form CALYPSO structure search. Energies calculated at TPSSh/6-31G(d) level of theory are relative to that of the lowest-energy core-shell structure.

Table SI. The parameters used in the ACSF descriptor with 89 function values.

No.	type	ζ	λ	η	R_s	R_{cut}	No.	type	ζ	λ	η	R_s	R_{cut}
1	G^1	**0.0**	**0.0**	**1.0**	**0.0**	**6.0**	**46**	G^3	**2.0**	**1.0**	**0.01**	**0.0**	**2.0**
2	G^1	**0.0**	**0.0**	**0.5**	**0.0**	**6.0**	**47**	G^3	**2.0**	**1.0**	**0.01**	**0.0**	**1.0**
3	G^1	**0.0**	**0.0**	**0.2**	**0.0**	**6.0**	**48**	G^3	**2.0**	**1.0**	**0.001**	**0.0**	**6.0**
4	G^1	**0.0**	**0.0**	**0.1**	**0.0**	**6.0**	**49**	G^3	**2.0**	**1.0**	**0.001**	**0.0**	**5.0**
5	G^1	**0.0**	**0.0**	**0.04**	**0.0**	**6.0**	**50**	G^3	**2.0**	**1.0**	**0.001**	**0.0**	**4.0**
6	G^1	**0.0**	**0.0**	**0.001**	**0.0**	**6.0**	**51**	G^3	**2.0**	**1.0**	**0.001**	**0.0**	**3.0**
7	G^2	**0.0**	**0.0**	**1.0**	**5.5**	**6.0**	**52**	G^3	**2.0**	**1.0**	**0.001**	**0.0**	**2.0**
8	G^2	**0.0**	**0.0**	**1.0**	**5.0**	**6.0**	**53**	G^3	**2.0**	**1.0**	**0.001**	**0.0**	**1.0**
9	G^2	**0.0**	**0.0**	**1.0**	**4.5**	**6.0**	**54**	G^3	**2.0**	**-1.0**	**0.01**	**0.0**	**6.0**
10	G^2	**0.0**	**0.0**	**1.0**	**4.0**	**6.0**	**55**	G^3	**2.0**	**-1.0**	**0.01**	**0.0**	**5.0**
11	G^2	**0.0**	**0.0**	**1.0**	**3.5**	**6.0**	**56**	G^3	**2.0**	**-1.0**	**0.01**	**0.0**	**4.0**
12	G^2	**0.0**	**0.0**	**1.0**	**3.0**	**6.0**	**57**	G^3	**2.0**	**-1.0**	**0.01**	**0.0**	**3.0**
13	G^2	**0.0**	**0.0**	**1.0**	**2.5**	**6.0**	**58**	G^3	**2.0**	**-1.0**	**0.01**	**0.0**	**2.0**
14	G^2	**0.0**	**0.0**	**1.0**	**2.0**	**6.0**	**59**	G^3	**2.0**	**-1.0**	**0.01**	**0.0**	**1.0**
15	G^2	**0.0**	**0.0**	**1.0**	**1.5**	**6.0**	**60**	G^3	**2.0**	**-1.0**	**0.001**	**0.0**	**6.0**
16	G^2	**0.0**	**0.0**	**1.0**	**1.0**	**6.0**	**61**	G^3	**2.0**	**-1.0**	**0.001**	**0.0**	**5.0**
17	G^3	**0.0**	**1.0**	**0.01**	**0.0**	**6.0**	**62**	G^3	**2.0**	**-1.0**	**0.001**	**0.0**	**4.0**
18	G^3	**0.0**	**1.0**	**0.01**	**0.0**	**5.0**	**63**	G^3	**2.0**	**-1.0**	**0.001**	**0.0**	**3.0**
19	G^3	**0.0**	**1.0**	**0.01**	**0.0**	**4.0**	**64**	G^3	**2.0**	**-1.0**	**0.001**	**0.0**	**2.0**
20	G^3	**0.0**	**1.0**	**0.01**	**0.0**	**3.5**	**65**	G^3	**2.0**	**-1.0**	**0.001**	**0.0**	**1.0**
21	G^3	**0.0**	**1.0**	**0.01**	**0.0**	**2.0**	**66**	G^3	**3.0**	**1.0**	**0.01**	**0.0**	**6.0**
22	G^3	**0.0**	**1.0**	**0.01**	**0.0**	**1.0**	**67**	G^3	**3.0**	**1.0**	**0.01**	**0.0**	**5.0**
23	G^3	**0.0**	**1.0**	**0.001**	**0.0**	**6.0**	**68**	G^3	**3.0**	**1.0**	**0.01**	**0.0**	**4.0**
24	G^3	**0.0**	**1.0**	**0.001**	**0.0**	**5.0**	**69**	G^3	**3.0**	**1.0**	**0.01**	**0.0**	**3.0**
25	G^3	**0.0**	**1.0**	**0.001**	**0.0**	**4.0**	**70**	G^3	**3.0**	**1.0**	**0.01**	**0.0**	**2.0**
26	G^3	**0.0**	**1.0**	**0.001**	**0.0**	**3.5**	**71**	G^3	**3.0**	**1.0**	**0.01**	**0.0**	**1.0**
27	G^3	**0.0**	**1.0**	**0.001**	**0.0**	**3.0**	**72**	G^3	**3.0**	**1.0**	**0.001**	**0.0**	**6.0**
28	G^3	**0.0**	**1.0**	**0.001**	**0.0**	**2.0**	**73**	G^3	**3.0**	**1.0**	**0.001**	**0.0**	**5.0**
29	G^3	**0.0**	**1.0**	**0.001**	**0.0**	**1.0**	**74**	G^3	**3.0**	**1.0**	**0.001**	**0.0**	**4.0**
30	G^3	**0.0**	**-1.0**	**0.01**	**0.0**	**6.0**	**75**	G^3	**3.0**	**1.0**	**0.001**	**0.0**	**3.0**
31	G^3	**0.0**	**-1.0**	**0.01**	**0.0**	**5.0**	**76**	G^3	**3.0**	**1.0**	**0.001**	**0.0**	**2.0**
32	G^3	**0.0**	**-1.0**	**0.01**	**0.0**	**4.0**	**77**	G^3	**3.0**	**1.0**	**0.001**	**0.0**	**1.0**
33	G^3	**0.0**	**-1.0**	**0.01**	**0.0**	**3.0**	**78**	G^3	**3.0**	**-1.0**	**0.01**	**0.0**	**6.0**
34	G^3	**0.0**	**-1.0**	**0.01**	**0.0**	**2.0**	**79**	G^3	**3.0**	**-1.0**	**0.01**	**0.0**	**5.0**
35	G^3	**0.0**	**-1.0**	**0.01**	**0.0**	**1.0**	**80**	G^3	**3.0**	**-1.0**	**0.01**	**0.0**	**4.0**
36	G^3	**0.0**	**-1.0**	**0.001**	**0.0**	**6.0**	**81**	G^3	**3.0**	**-1.0**	**0.01**	**0.0**	**3.0**
37	G^3	**0.0**	**-1.0**	**0.001**	**0.0**	**5.0**	**82**	G^3	**3.0**	**-1.0**	**0.01**	**0.0**	**2.0**
38	G^3	**0.0**	**-1.0**	**0.001**	**0.0**	**4.0**	**83**	G^3	**3.0**	**-1.0**	**0.01**	**0.0**	**1.0**

39	G^3	**0.0**	**-1.0**	**0.001**	**0.0**	**3.0**	**84**	G^3	**3.0**	**-1.0**	**0.001**	**0.0**	**6.0**
40	G^3	**0.0**	**-1.0**	**0.001**	**0.0**	**2.0**	**85**	G^3	**3.0**	**-1.0**	**0.001**	**0.0**	**5.0**
41	G^3	**0.0**	**-1.0**	**0.001**	**0.0**	**1.0**	**86**	G^3	**3.0**	**-1.0**	**0.001**	**0.0**	**4.0**
42	G^3	**2.0**	**1.0**	**0.01**	**0.0**	**6.0**	**87**	G^3	**3.0**	**-1.0**	**0.001**	**0.0**	**3.0**
43	G^3	**2.0**	**1.0**	**0.01**	**0.0**	**5.0**	**88**	G^3	**3.0**	**-1.0**	**0.001**	**0.0**	**2.0**
44	G^3	**2.0**	**1.0**	**0.01**	**0.0**	**4.0**	**89**	G^3	**3.0**	**-1.0**	**0.001**	**0.0**	**1.0**
45	G^3	**2.0**	**1.0**	**0.01**	**0.0**	**3.0**							