Electronic Supplementary Information

Deep Learning-Driven Prediction of Chemical Addition Patterns for Carboncones and Fullerenes

Zhengda Li, Xuyang Chen, and Yang Wang*

School of Chemistry and Chemical Engineering, Yangzhou University, Yangzhou, Jiangsu 225002, China

E-mail: yangwang@yzu.edu.cn

1 Cutoff energies of $\mathrm{RE_{xTB}}$ and $\mathrm{RE_{DNN}}$

Table S1. Cutoff energy of $\mathrm{RE_{xTB}}$ for screening seed regioisomers of $\mathbf{C_{70}H_{20+(n-2)}}$ to generate structures of $\mathbf{C_{70}H_{20+n}}$, and cutoff energy of $\mathrm{RE_{DNN}}$ for selecting lower-energy regioisomers of $\mathbf{C_{70}H_{20+n}}$ for subsequent xTB geometry optimizations.

n	Cutoff of $\mathrm{RE_{xTB}}$ (kcal/mol)	Cutoff of $\mathrm{RE_{DNN}}$ (kcal/mol)
6	30.0	50.0
8	30.0	45.0
10	30.0	35.0
12	30.0	30.0
14	20.0	45.0
16	20.0	40.0
18	20.0	60.0
20	15.0	40.0
22	15.0	50.0
24	20.0	40.0
26	15.0	50.0
28	15.0	55.0
30	20.0	40.0
32	20.0	40.0
34	15.0	50.0
36	20.0	50.0
38	25.0	50.0
40	25.0	60.0

Table S2. Cutoff energy of $\mathrm{RE_{xTB}}$ for screening seed regioisomers of $\mathbf{C_{62}H_{16+(n-2)}}$ to generate structures of $\mathbf{C_{62}H_{16+n}}$.

n	Cutoff of $\mathrm{RE_{xTB}}$ (kcal/mol)
6	25.0
8	15.0
10	15.0
12	15.0
14	20.0
16	15.0
18	15.0
20	20.0
22	15.0
24	20.0
26	25.0
28	15.0
30	25.0

2 PCA feature dimensionality reduction

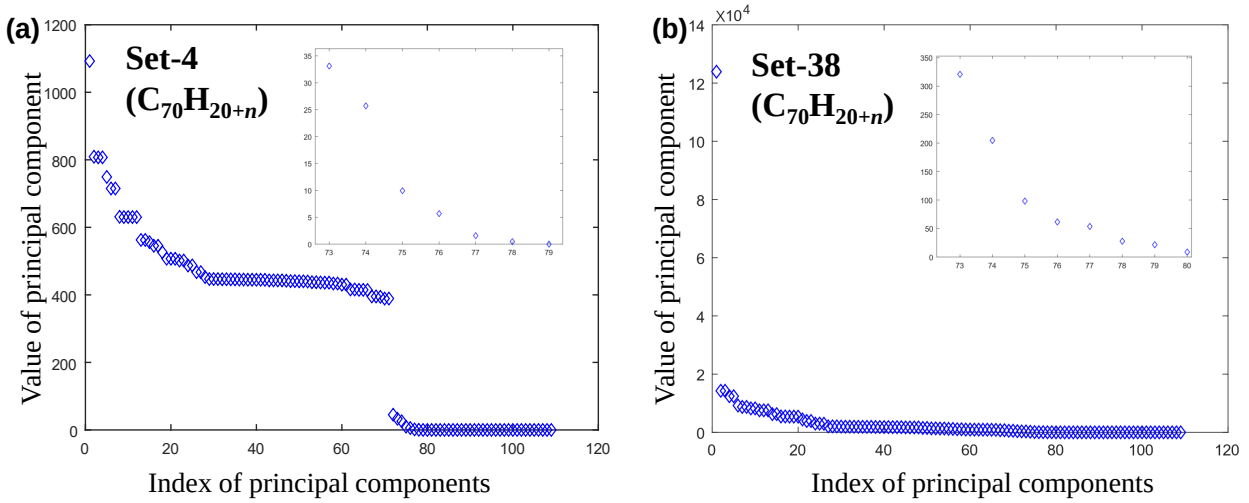

Figure S1. PCA results of the 109 features for regioisomers of $C_{70}H_{20+n}$ in (a) Set-4 and (b) Set-38.

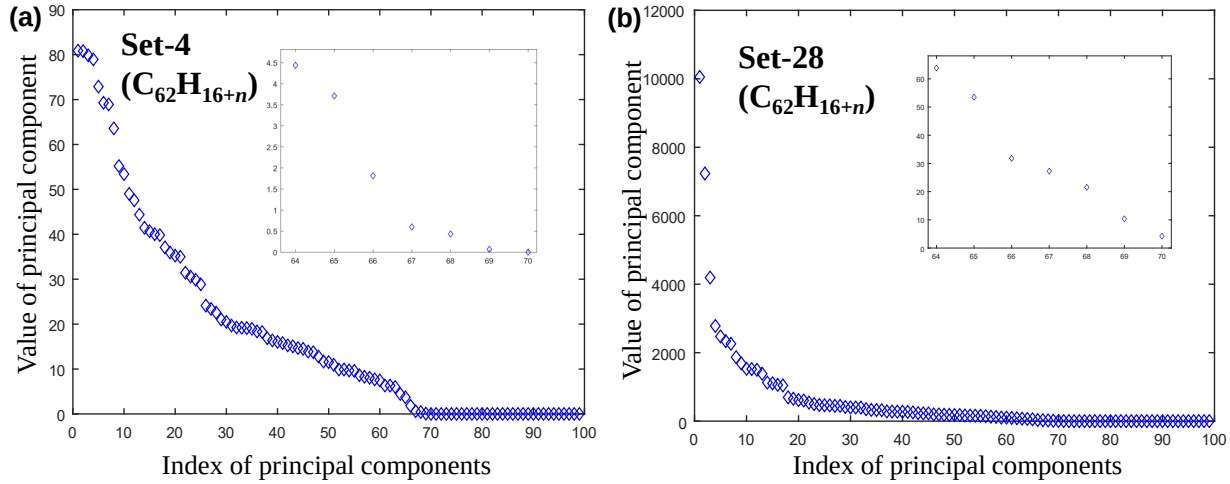

Figure S2. PCA results of the 99 features for regioisomers of $C_{62}H_{16+n}$ in (a) Set-4 and (b) Set-28.

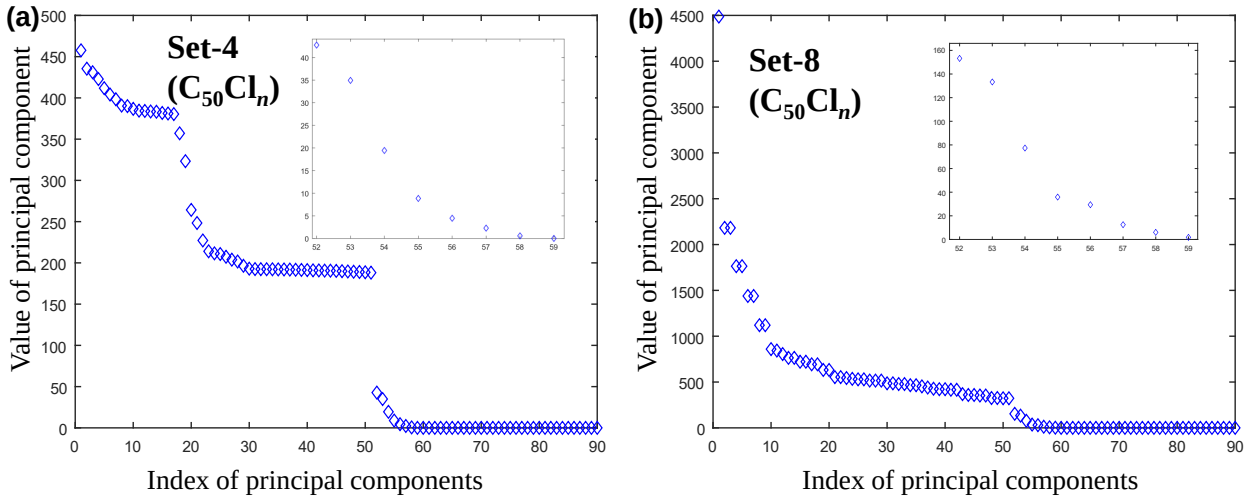

Figure S3. PCA results of the 90 features for regioisomers of $C_{50}Cl_n$ in (a) Set-4 and (b) Set-8.

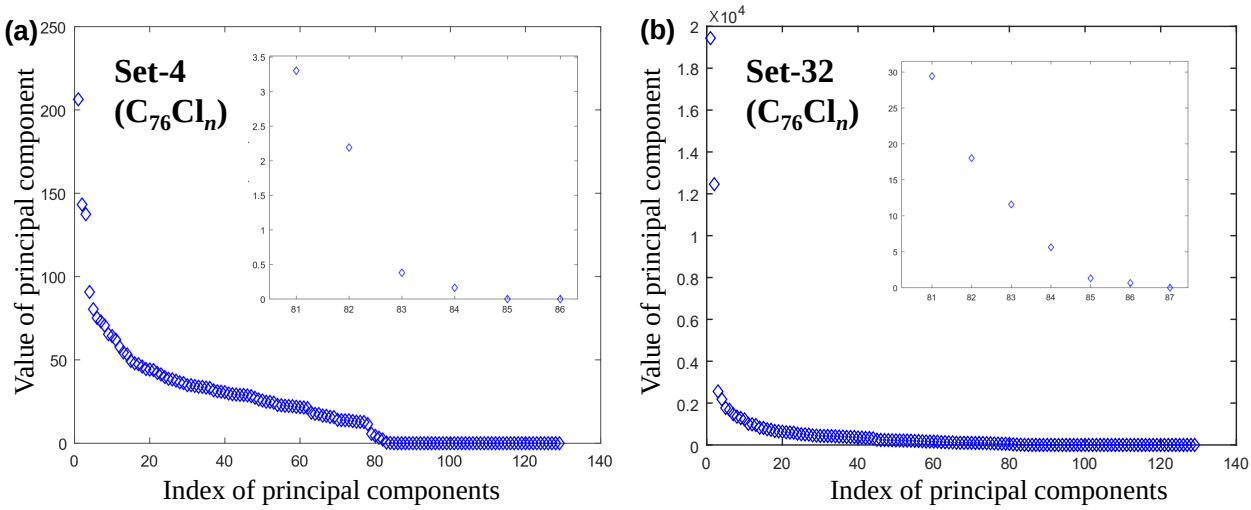

Figure S4. PCA results of the 129 features for regioisomers of $C_{76}Cl_n$ in (a) Set-4 and (b) Set-32.

The PCA is applied prior to feeding the features to the neural network. For each incremental model, DNN-n, PCA is consistently used to reduce feature dimensionality as the dataset (Set-n) grows with the increasing number of instances. Dimensionality reduction is based on the following criterion: only the first N principal components are retained if the $(N+1)$-th component is sufficiently smaller than the N-th component. In this work, we chose a typical threshold value of 10_{-10} for the ratio between the $(N+1)$-th and N-th components. Table R1 S3 lists the number of original features and the number of retained features after PCA for training each DNN-n model on the four systems studied.

Table S3. Number of original features (N) and number of retained features after PCA (N_{PCA}) for training each DNN-n model on hydrogenated carboncones ($C_{62}H_{16+n}$ and $C_{70}H_{20+n}$) and chlorinated fullerenes ($C_{50}Cl_n$ and $C_{76}Cl_n$).

System	n	N	N_{PCA}	System	n	N	N_{PCA}	System	n	N	N_{PCA}
	4	99	68		4	109	76		4	90	59
	6	99	68		6	109	77	$C_{50}Cl_n$	6	90	59
	8	99	69		8	109	77		8	90	59
	10	99	69		10	109	77		4	129	84
	12	99	70		12	109	77		6	129	85
	14	99	70		14	109	78		8	129	86
$C_{62}H_{16+n}$	16	99	70		16	109	78		10	129	86
	18	99	70		18	109	78		12	129	86
	20	99	70		20	109	78		14	129	86
	22	99	70	$C_{70}H_{20+n}$	22	109	78		16	129	86
	24	99	70		24	109	78		18	129	86
	26	99	70		26	109	79	$C_{76}Cl_n$	20	129	86
	28	99	70		28	109	79		22	129	86
					30	109	79		24	129	86
					32	109	79		26	129	87
					34	109	79		28	129	87
					36	109	79		30	129	87
					38	109	79		32	129	88

3 Comparison of relative isomer energies between xTB and DFT calculations

To assess the reliability of the xTB method for chlorinated fullerenes, we evaluate two representative cases, $C_{50}Cl_{10}$ and $C_{76}Cl_{28}$, both of which include experimentally determined structures.[1–3] In each case, we have randomly chosen additional regioisomers beyond the lowest-energies ones, so that the data points span a wide range of relative energies.

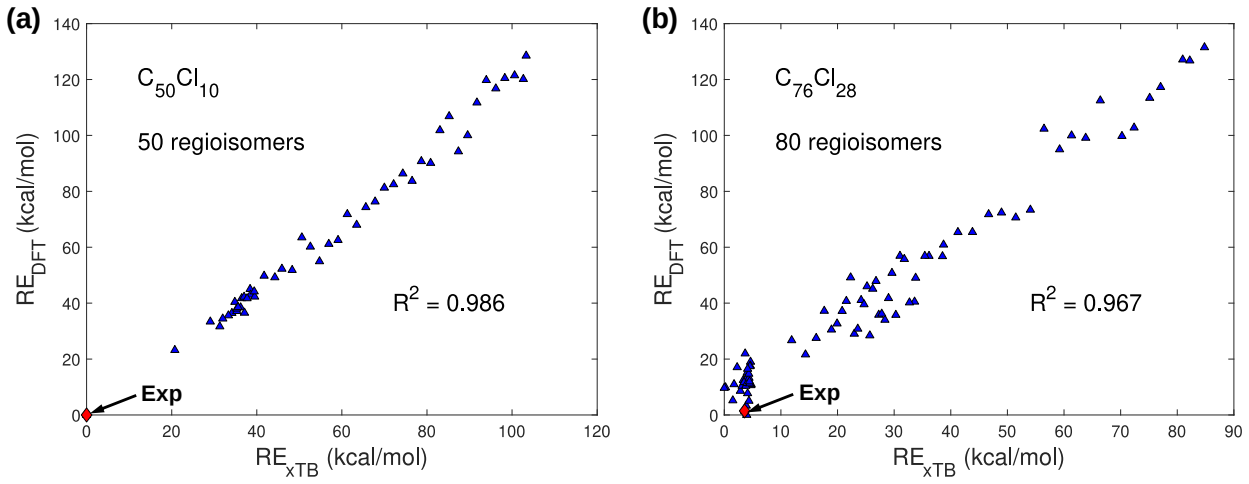

Figure S5. Comparison of relative isomer energies between the xTB and DFT calculations for (a) $C_{50}Cl_{10}$[1,2] and (b) $C_{76}Cl_{28}$.[3] The experimentally identified regioisomers[1–3] are represented by red diamonds. The squared correlation coefficient, R^2, is indicated in each plot.

As shown in Figure S5, the xTB relative energies correlate well with the DFT relative energies for both $C_{50}Cl_{10}$ and $C_{76}Cl_{28}$, suggesting that the xTB method is a reasonable choice for prescreening the relatively stable regioisomers.

4 Distortion of carbon framework upon addition

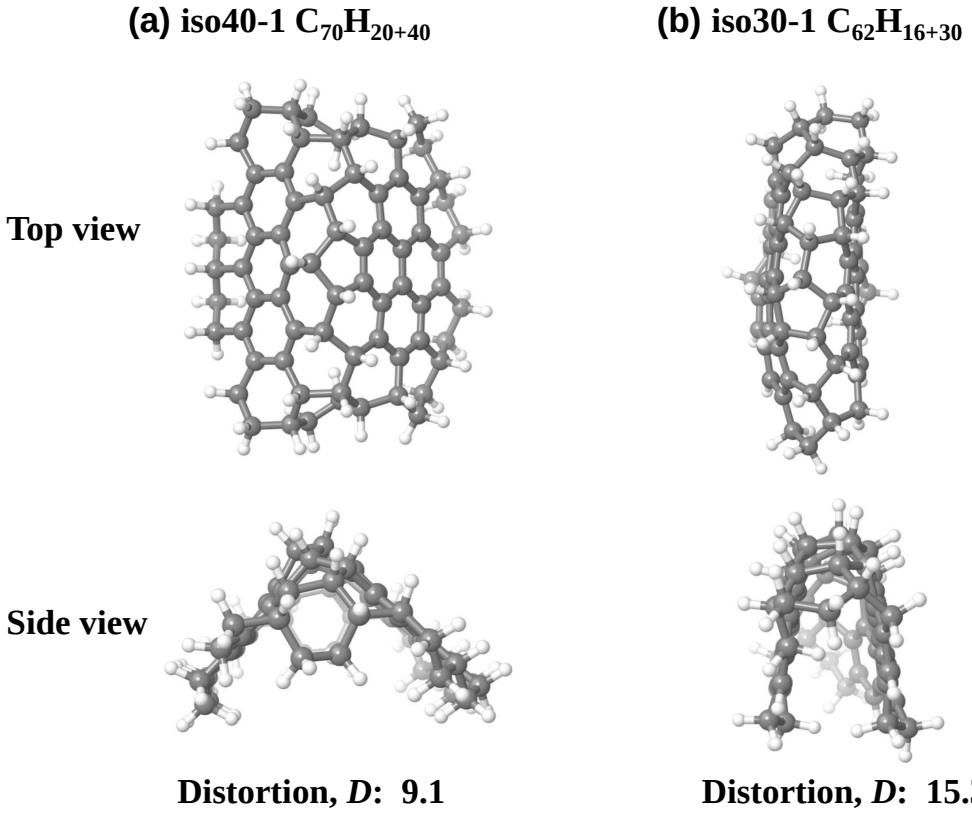

Figure S6. Top view and side view of the DFT optimized structures of the lowest-energy regioisomer of (a) $C_{70}H_{20+40}$ and (b) $C_{62}H_{16+30}$. The distortion of carbon framework, D (defined in Equation S1), is provided in units of Å.

To evaluate the distortion of carbon framework of a given carboncone or fullerene molecule upon exohedral addition of atoms, we compare the difference in equilibrium geometry between the pristine molecule and the adduct molecule. To this end, we first reorient the adduct molecule using the Kabsch–Umeyama algorithm[4–6] so that the RMSD between the positions of C atoms in the adduct and those in the pristine molecule is minimized. Then, the distortion of carbon framework upon addition is defined as the above-obtained RMSD:

$$D = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\Big[(x_i - x_i^0)^2 + (y_i - y_i^0)^2 + (z_i - z_i^0)^2\Big]}, \tag{S1}$$

where $\{x_i, y_i, z_i\}$ are the Cartesian coordinates of C atoms in the adduct and $\{x_i^0, y_i^0, z_i^0\}$ are those in the pristine molecule.

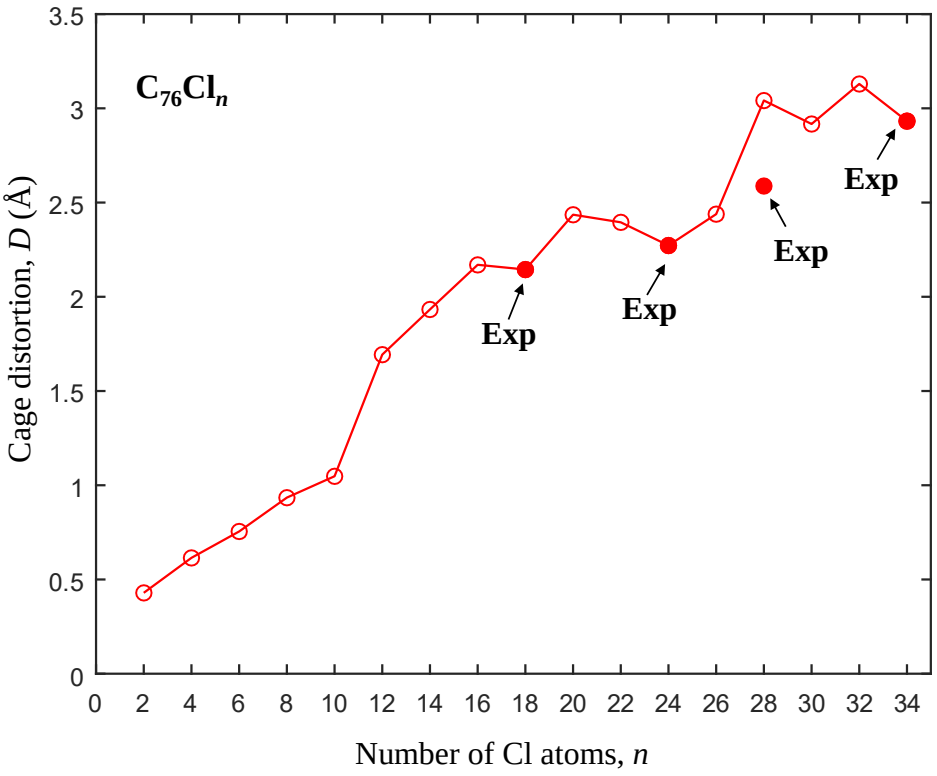

Figure S7. Cage distortion, D (defined in Equation 4 in the main text), as a function of the number of added Cl atoms, n, for the lowest-energy structures of chlorinated fullerene $C_{76}Cl_n$. The experimentally identified structures are indicated by arrows.

5 Performance of DNN models on test set

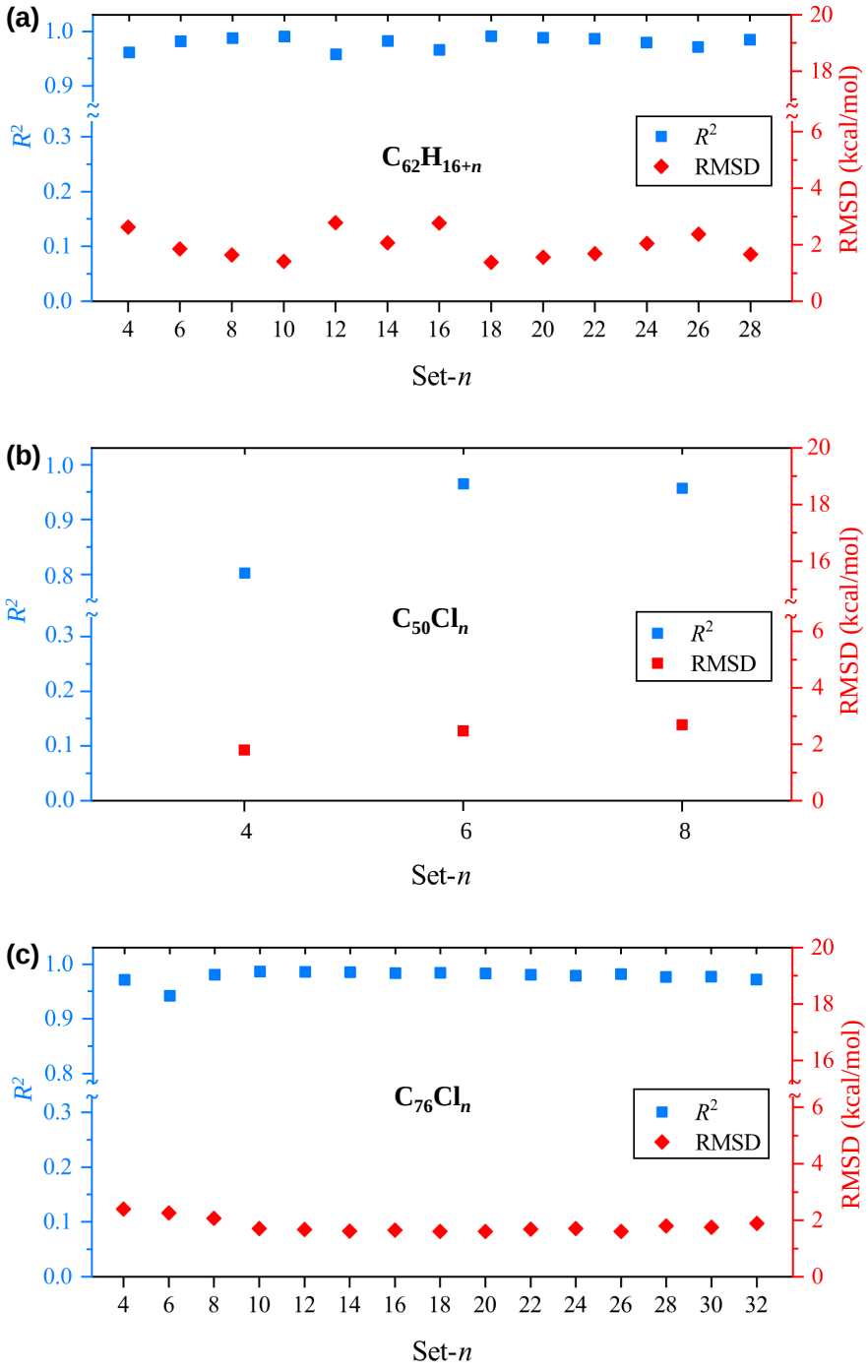

Figure S8. Predictive performance of the DNN-n models in predicting relative energies of regioisomers of (a) $C_{62}H_{16+n}$, (b) $C_{50}Cl_n$, and (c) $C_{76}Cl_n$. The results are evaluated on the test set from Set-n. Red diamonds (right y-axis) denote the RMSD of the model-predicted relative isomer energies from the xTB computed ones, while blue squares (left y-axis) represent the corresponding squared correlation coefficient, R^2.

6 Comparison of performance between DNN and other models

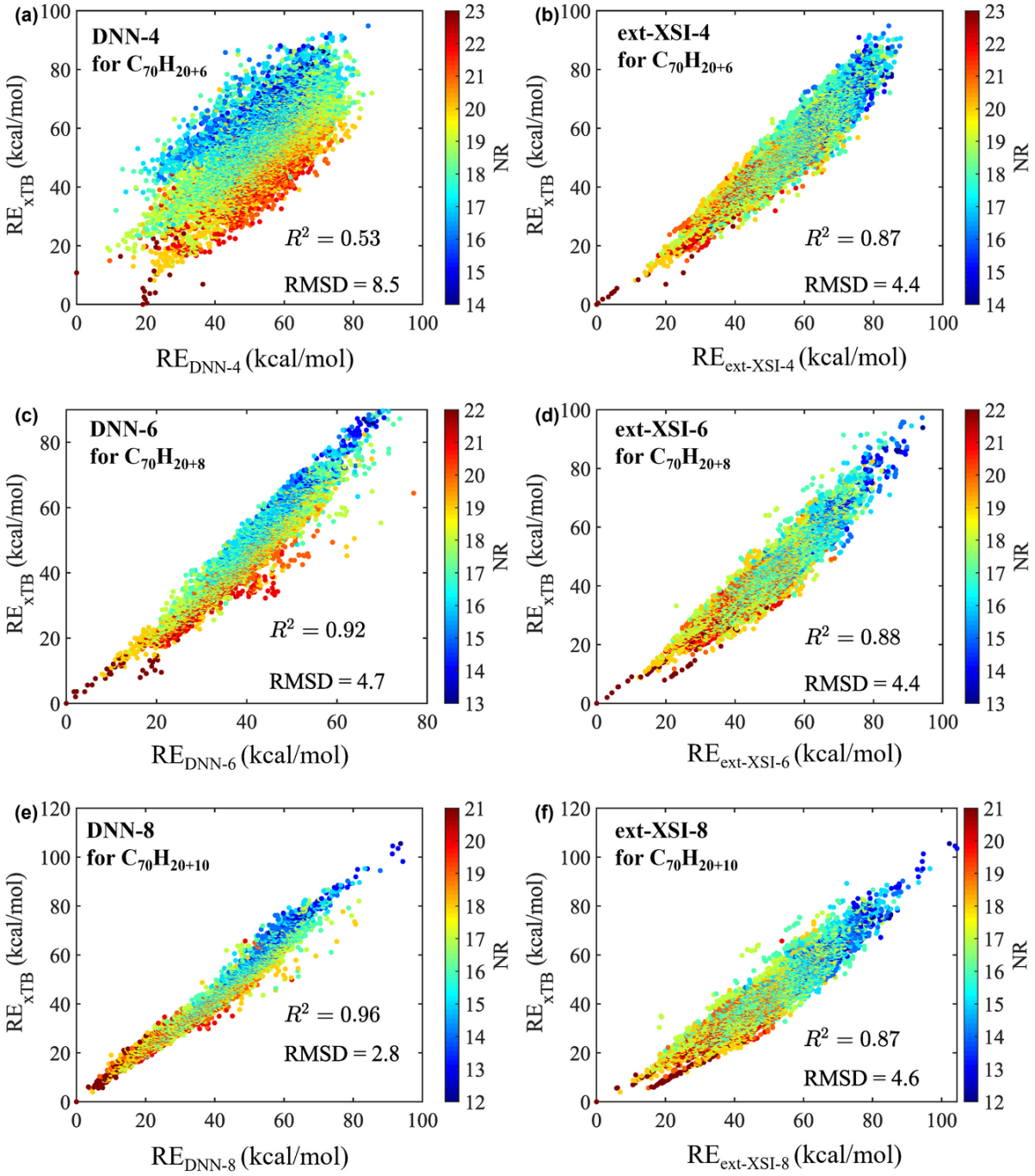

Figure S9. Performances of the DNN-n and ext-XSI-n models in predicting the xTB relative energies, RE_{xTB}, of $C_{70}H_{20+(n+2)}$ regioisomers, with n = 4–8. RE_{xTB} is compared with the relative energies predicted by (a) DNN-4, (b) ext-XSI-4, (c) DNN-6, (d) ext-XSI-6, (e) DNN-8, and (f) ext-XSI-8 models.

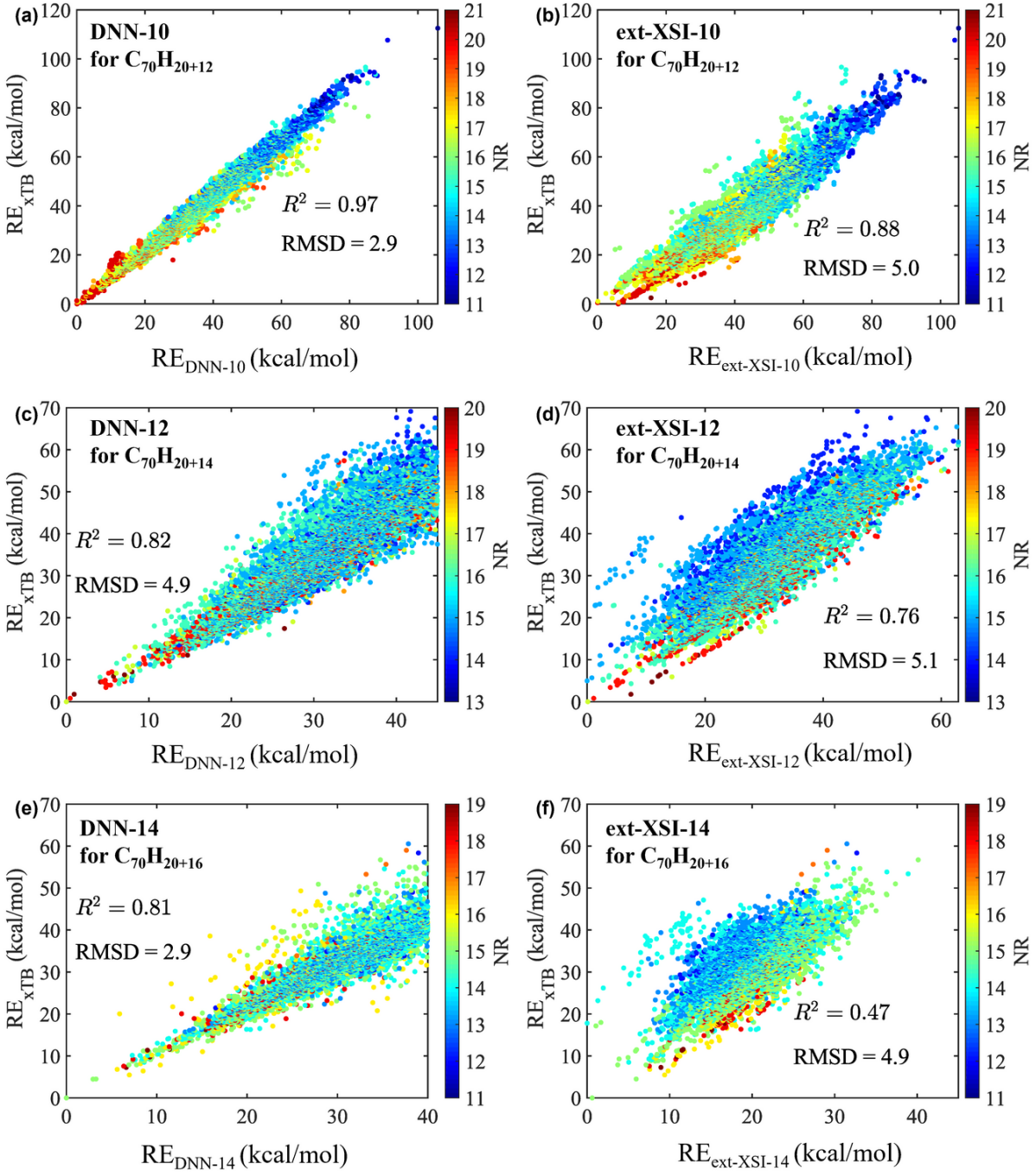

Figure S10. Idem Figure S9 for n = 10–14. The results are presented for (a) DNN-10, (b) ext-XSI-10, (c) DNN-12, (d) ext-XSI-12, (e) DNN-14, and (f) ext-XSI-14 models.

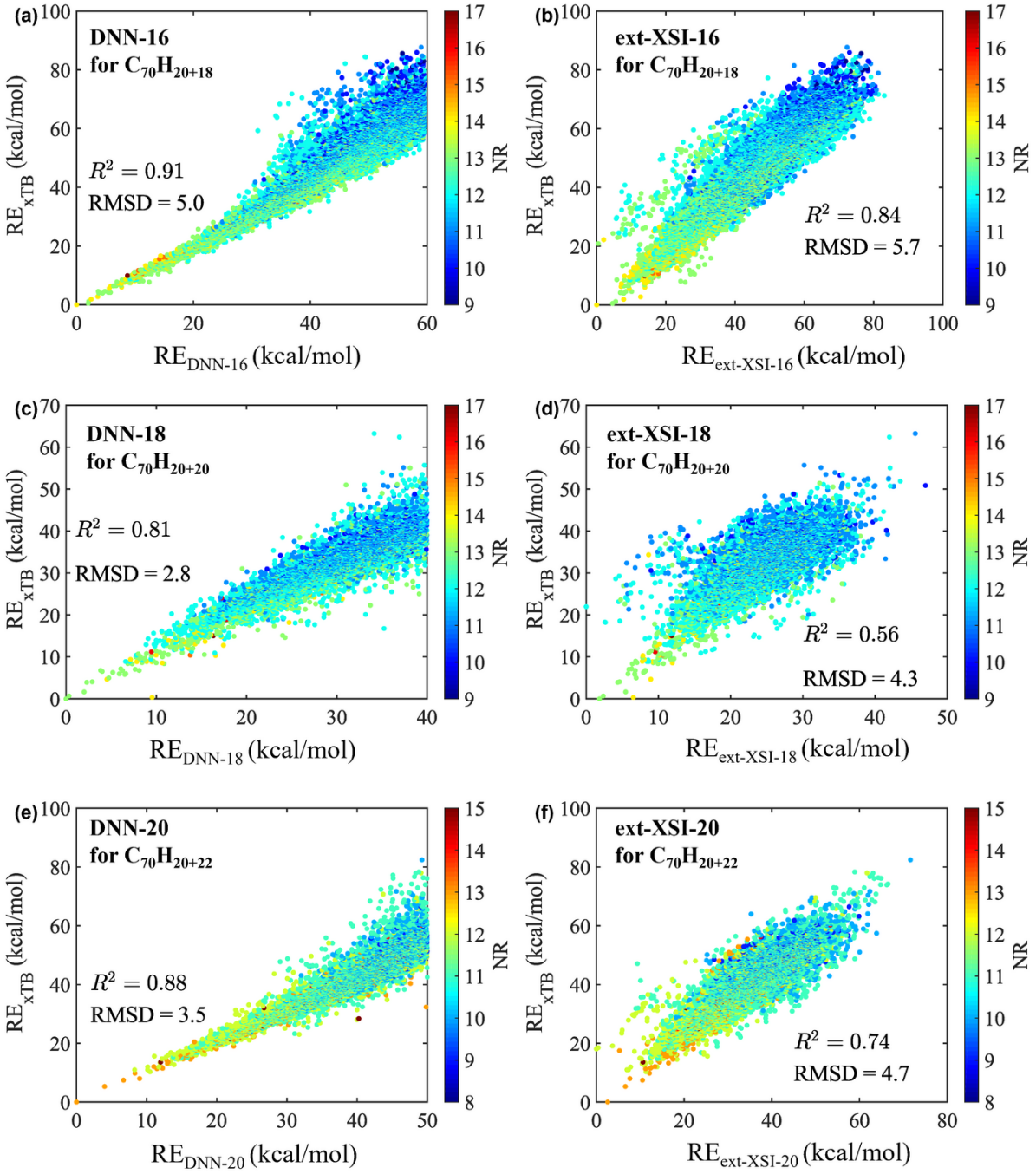

Figure S11. Idem Figure S9 for n = 16–20. The results are presented for (a) DNN-16, (b) ext-XSI-16, (c) DNN-18, (d) ext-XSI-18, (e) DNN-20, and (f) ext-XSI-20 models.

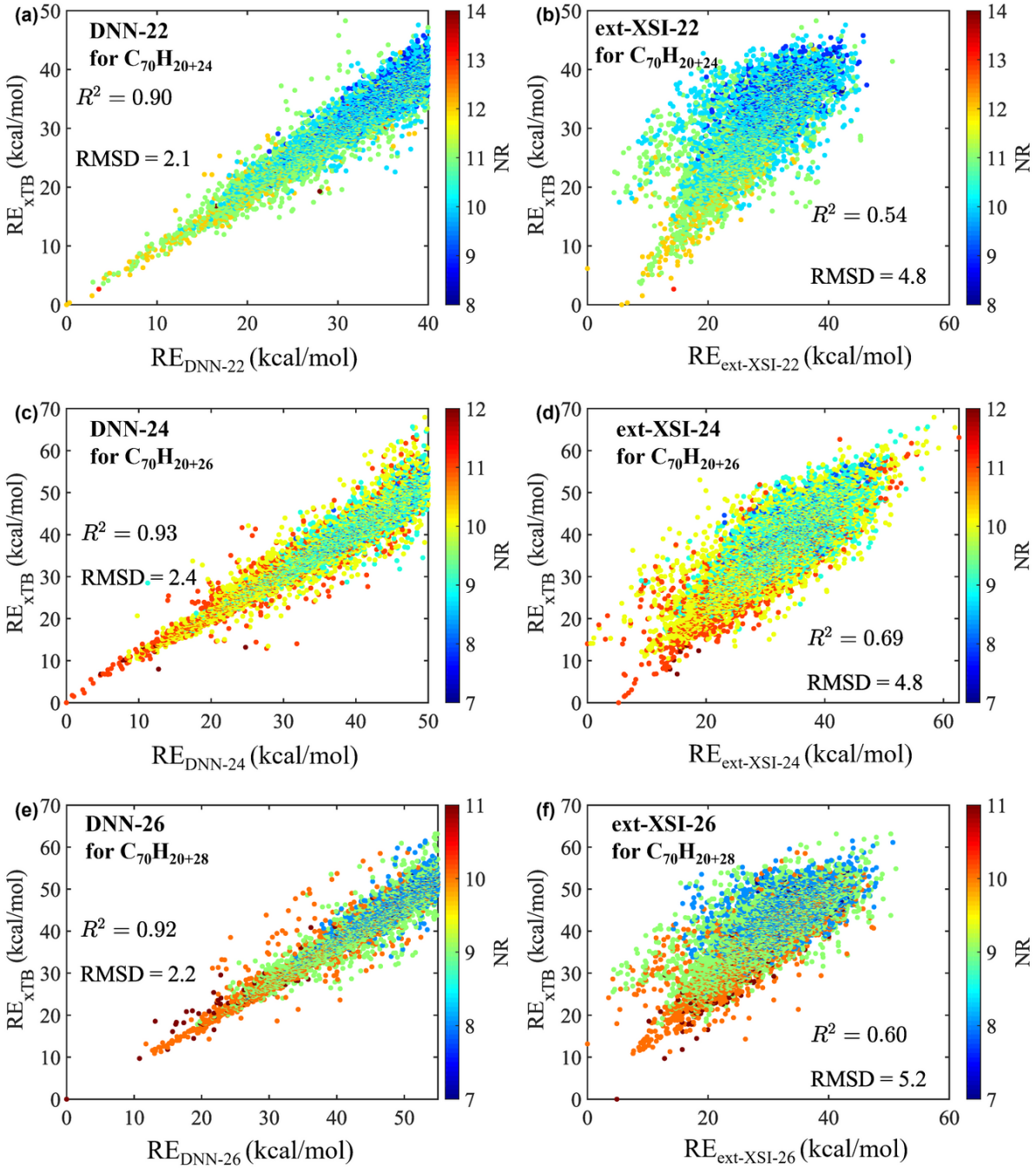

Figure S12. Idem Figure S9 for n = 22–26. The results are presented for (a) DNN-22, (b) ext-XSI-22, (c) DNN-24, (d) ext-XSI-24, (e) DNN-26, and (f) ext-XSI-26 models.

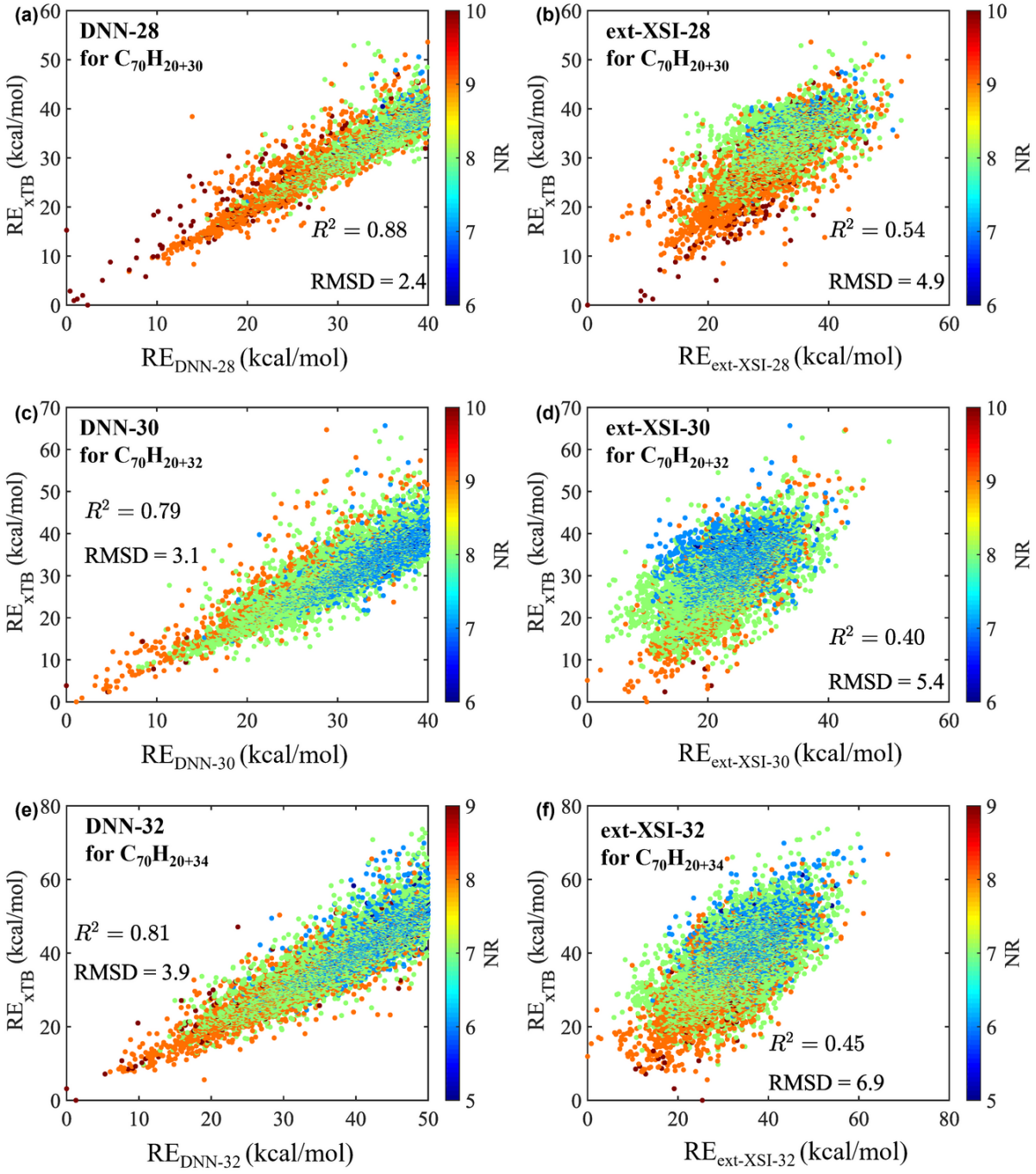

Figure S13. Idem Figure S9 for n = 28–32. The results are presented for (a) DNN-28, (b) ext-XSI-28, (c) DNN-30, (d) ext-XSI-30, (e) DNN-32, and (f) ext-XSI-32 models.

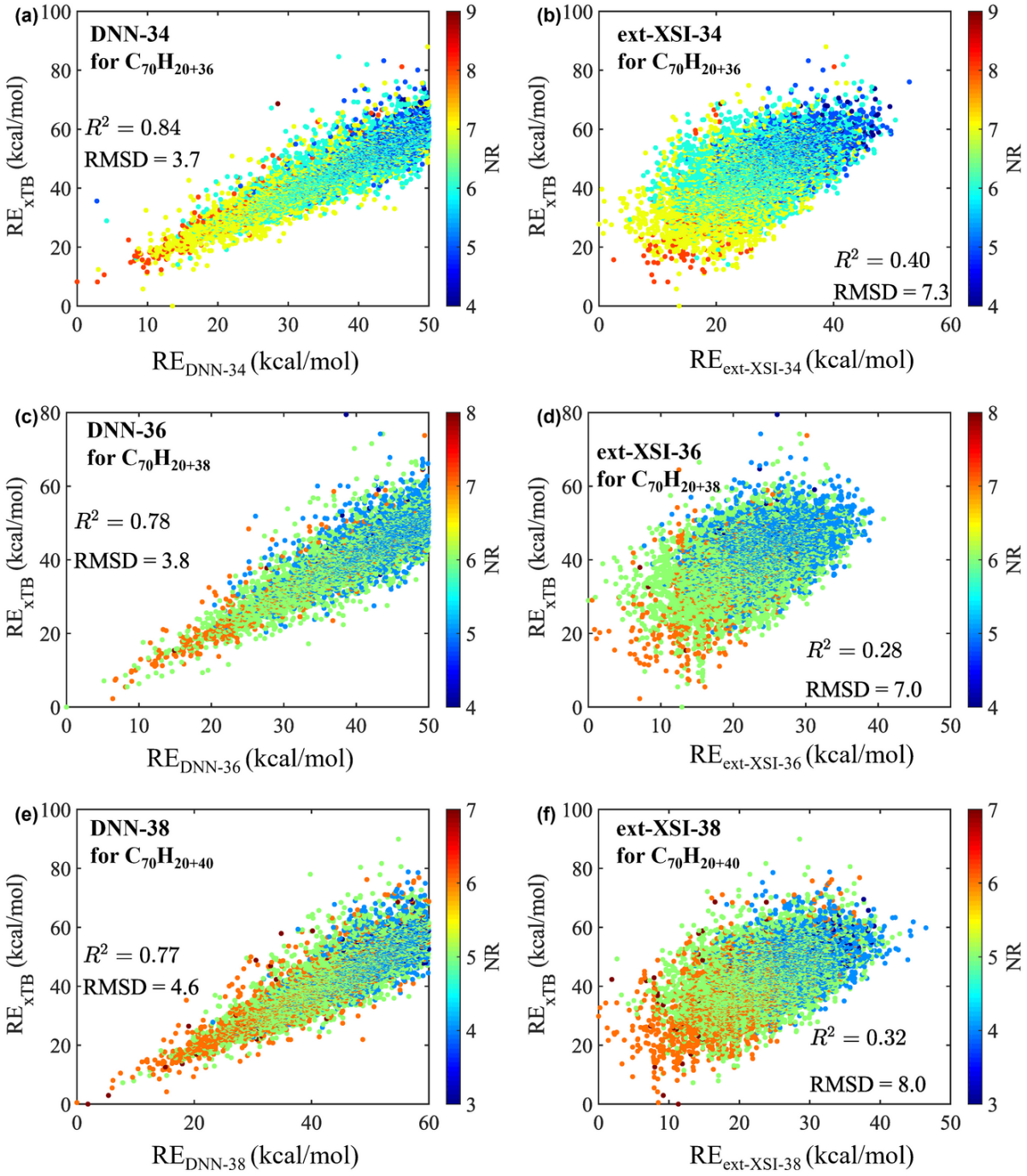

Figure S14. Idem Figure S9 for n = 34–38. The results are presented for (a) DNN-34, (b) ext-XSI-34, (c) DNN-36, (d) ext-XSI-36, (e) DNN-38, and (f) ext-XSI-38 models.

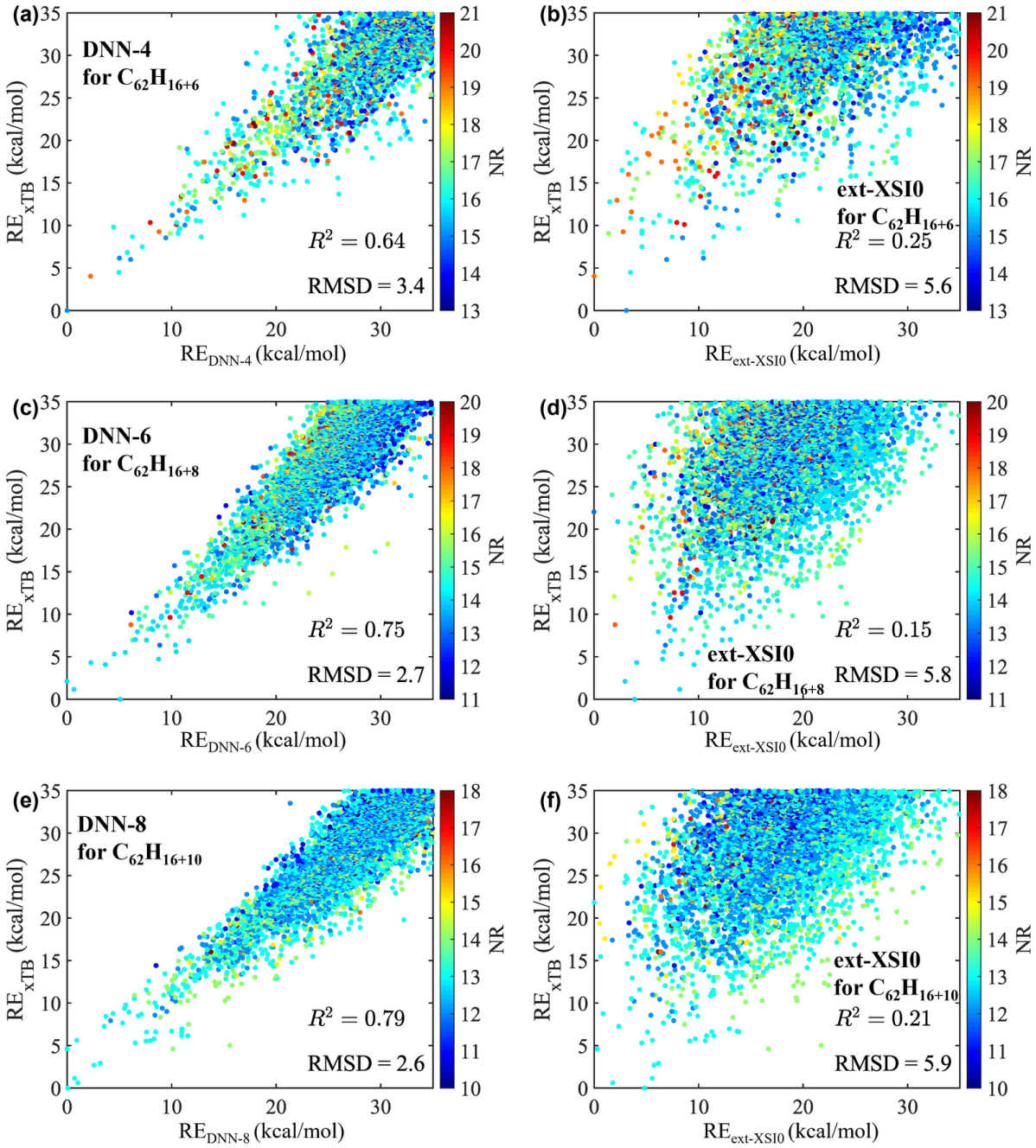

Figure S15. Performances of the DNN-n and ext-XSI0 models in predicting the xTB relative energies, RE_{xTB}, of $C_{62}H_{16+(n+2)}$ regioisomers, with n = 4–8. RE_{xTB} is compared with the relative energies predicted by (a) DNN-4, (b) ext-XSI0 for $C_{62}H_{16+6}$, (c) DNN-6, (d) ext-XSI0 for $C_{62}H_{16+8}$, (e) DNN-8, and (f) ext-XSI0 for $C_{62}H_{16+10}$. R^2 and RMSD (in kcal/mol) between the xTB and the model-predicted values are indicated in each plot. The data points are colorized according to the NR values for the corresponding addition patterns.

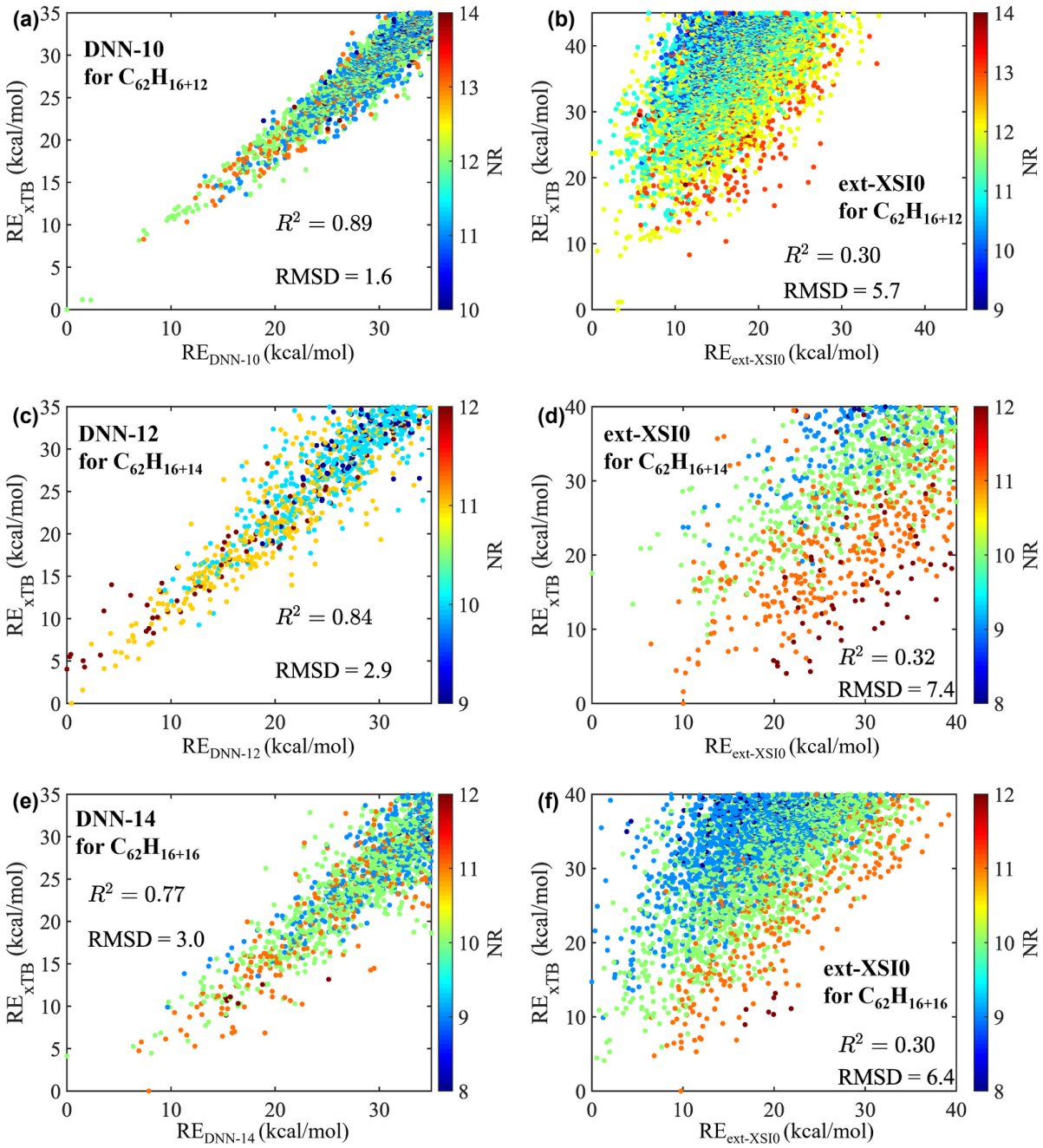

Figure S16. Idem Figure S15 for n = 10–14. The results are presented for (a) DNN-10, (b) ext-XSI0 for $C_{62}H_{16+12}$, (c) DNN-12, (d) ext-XSI0 for $C_{62}H_{16+14}$, (e) DNN-14, and (f) ext-XSI0 for $C_{62}H_{16+16}$.

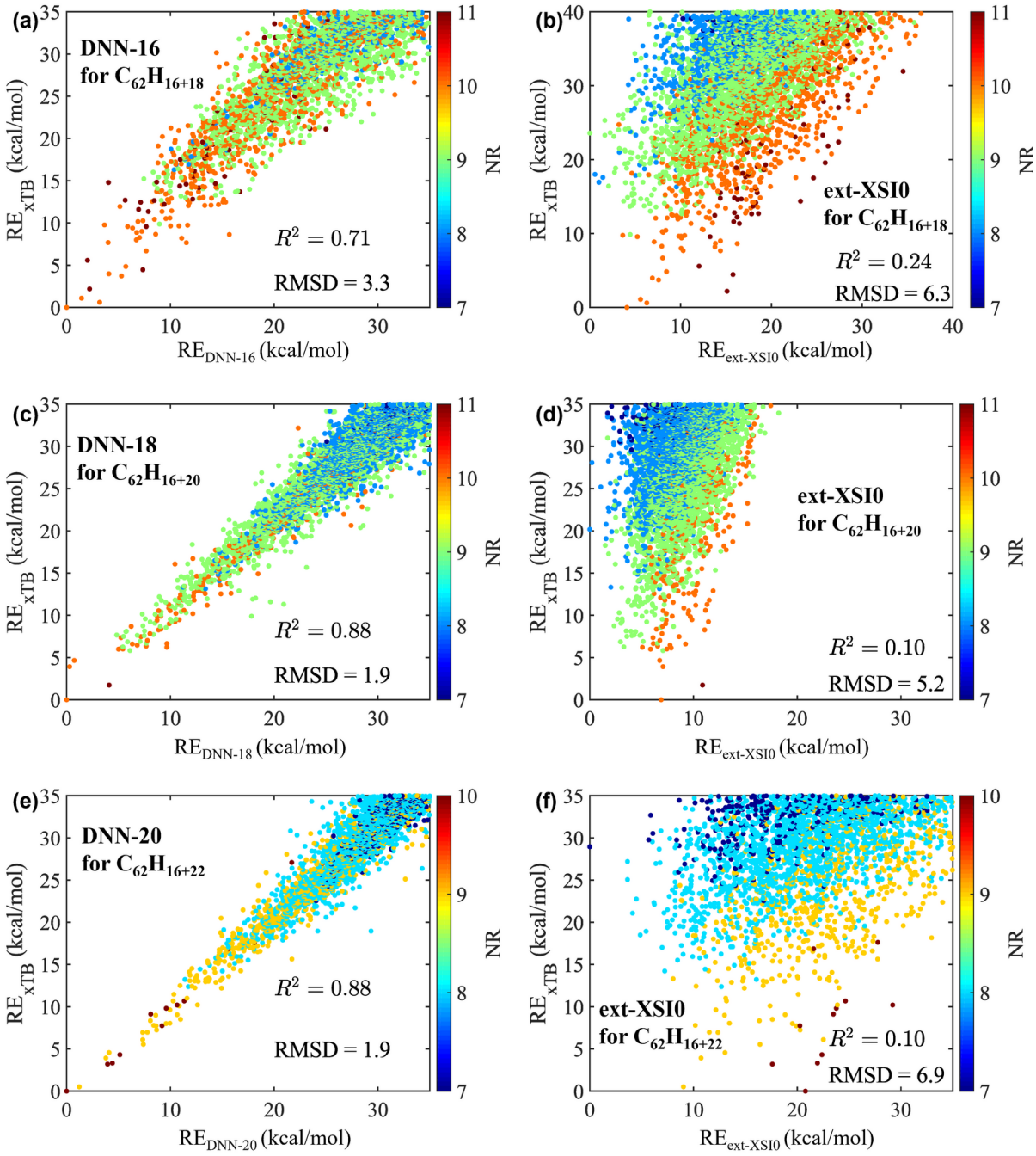

Figure S17. Idem Figure S15 for n = 16–20. The results are presented for (a) DNN-16, (b) ext-XSI0 for $C_{62}H_{16+18}$, (c) DNN-18, (d) ext-XSI0 for $C_{62}H_{16+20}$, (e) DNN-20, and (f) ext-XSI0 for $C_{62}H_{16+22}$.

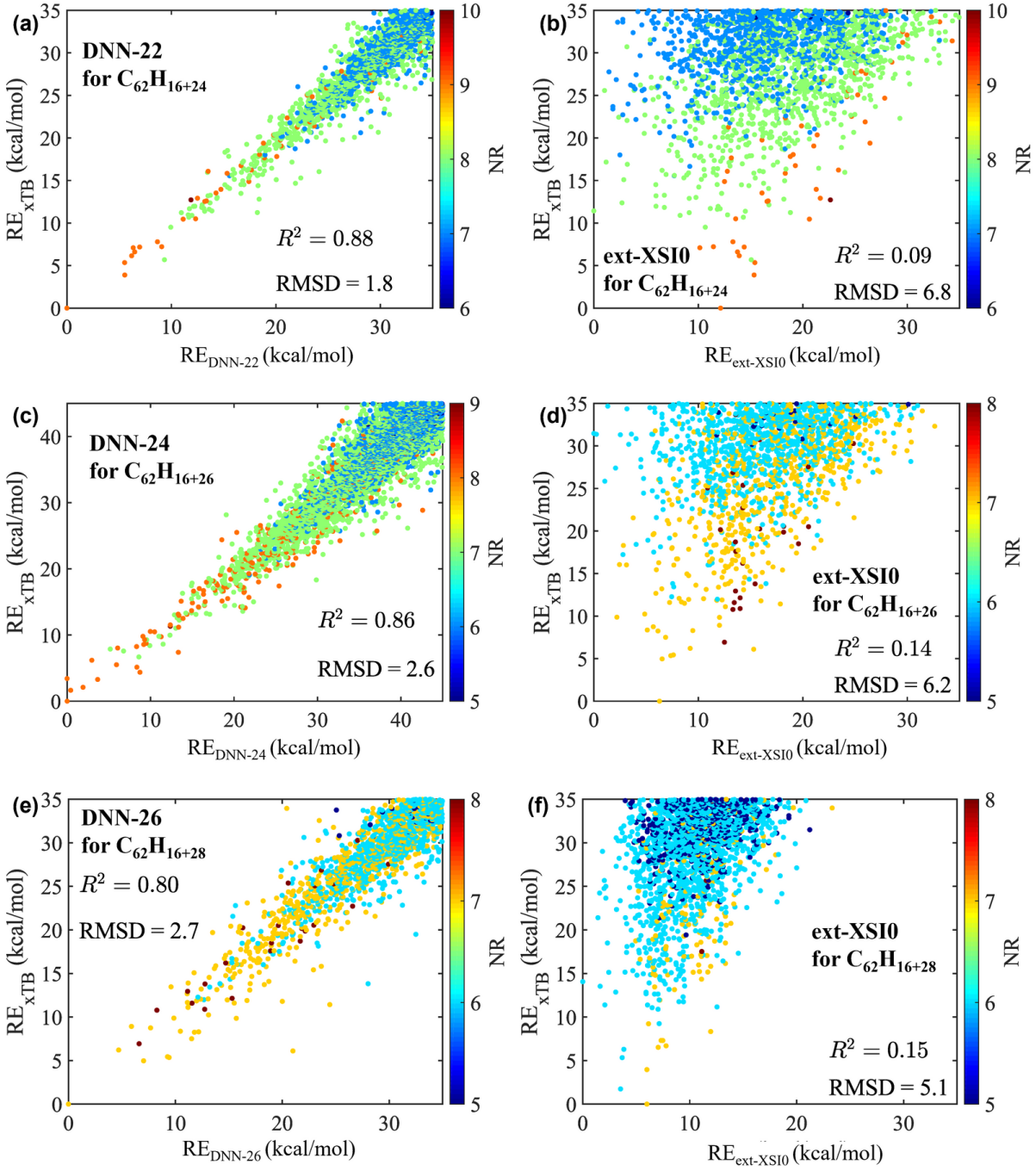

Figure S18. Idem Figure S15 for n = 22–26. The results are presented for (a) DNN-22, (b) ext-XSI0 for $C_{62}H_{16+24}$, (c) DNN-24, (d) ext-XSI0 for $C_{62}H_{16+26}$, (e) DNN-26, and (f) ext-XSI0 for $C_{62}H_{16+28}$.

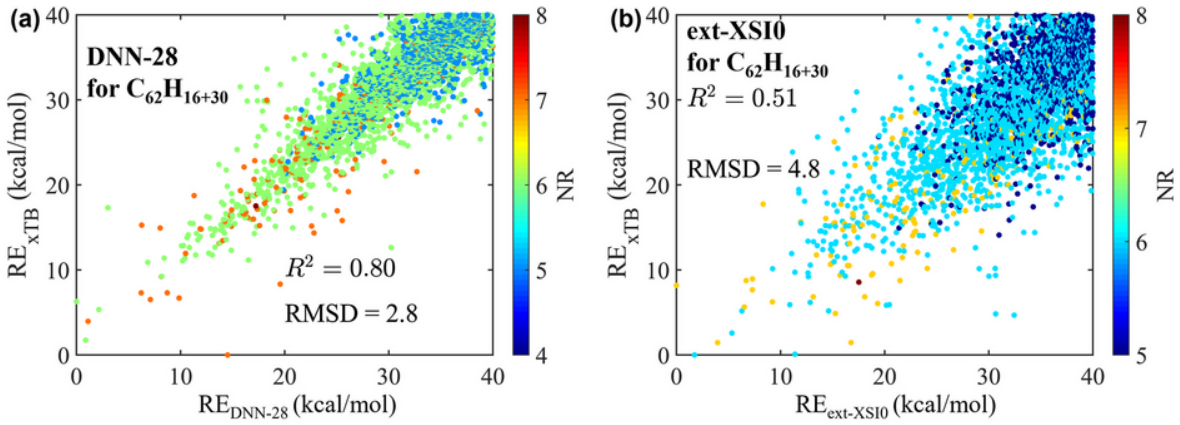

Figure S19. Idem Figure S15 for $n = 28$. The results are presented for (a) DNN-28 and (b) ext-XSI0 models.

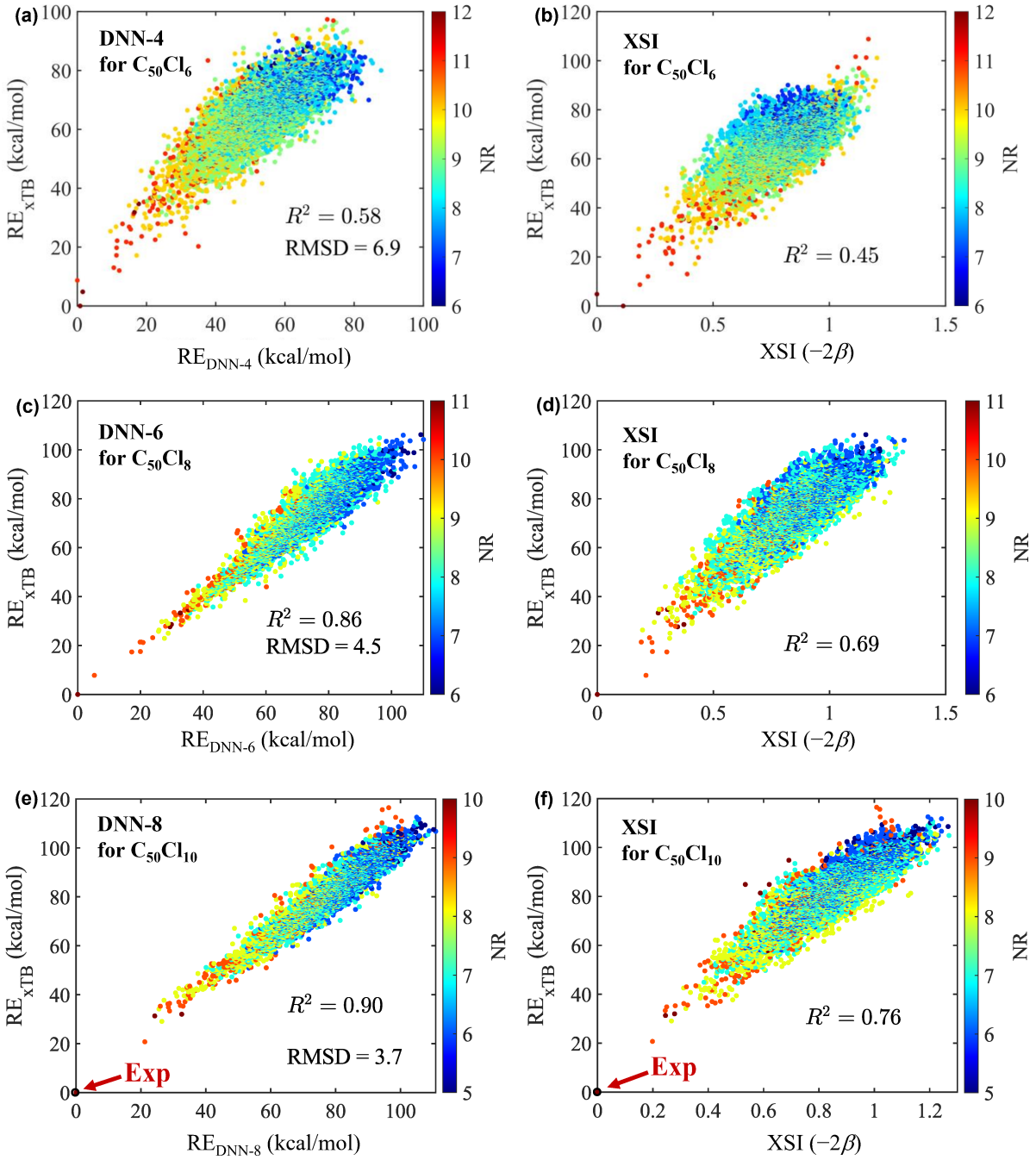

Figure S20. Performances of the DNN-n and XSI models in predicting the xTB relative energies, $\mathrm{RE_{xTB}}$, of $C_{50}Cl_{n+2}$ regioisomers, with n = 4–8. $\mathrm{RE_{xTB}}$ is compared with the relative energies predicted by (a) DNN-4, (b) XSI for $C_{50}Cl_6$, (c) DNN-6, (d) XSI for $C_{50}Cl_8$, (e) DNN-8, and (f) XSI for $C_{50}Cl_{10}$ models. R^2 and RMSD (in kcal/mol) between the xTB and the model-predicted values are indicated in each plot. The data points are colorized according to the NR values for the corresponding addition patterns.

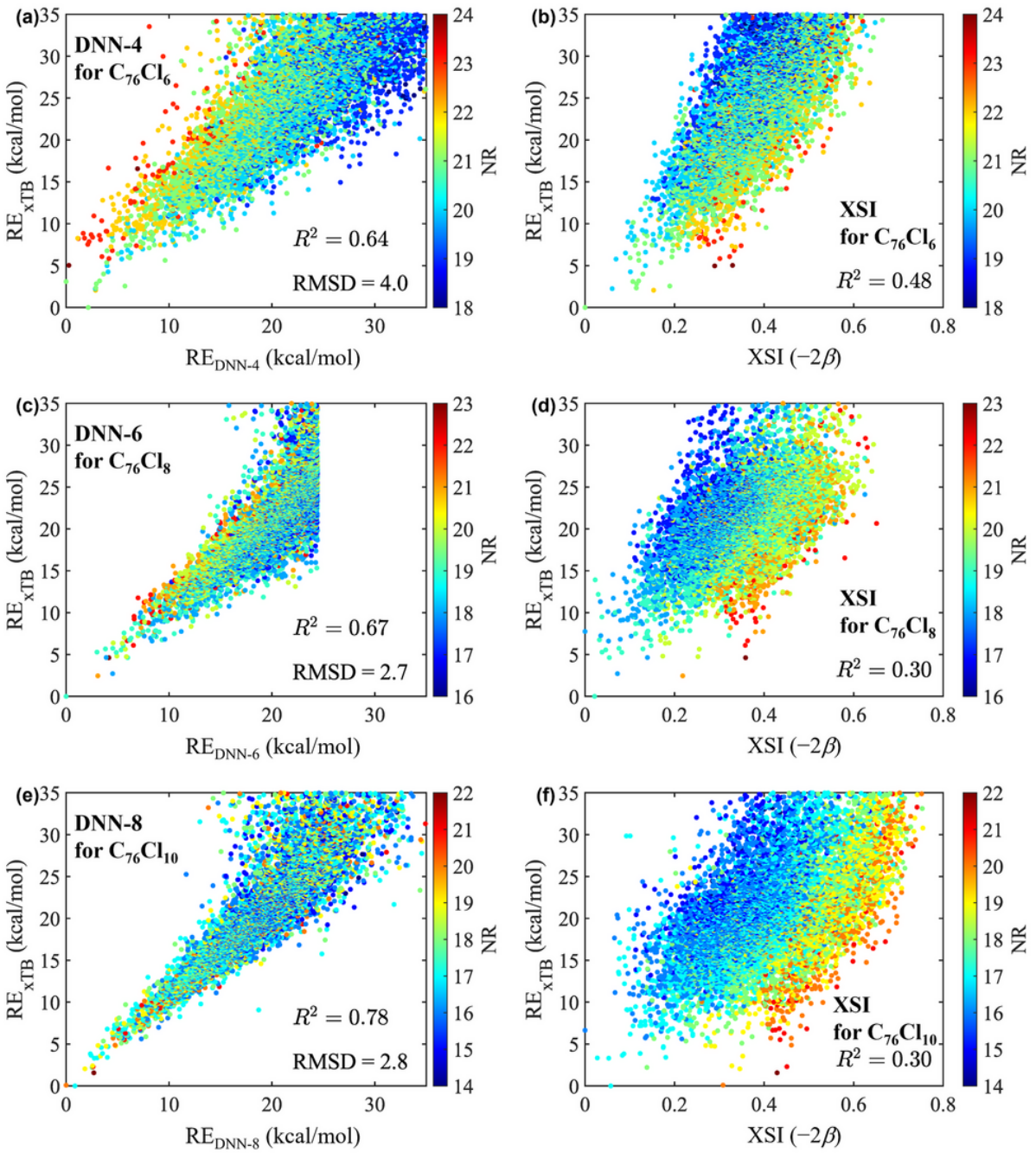

Figure S21. Performances of the DNN-n and XSI models in predicting the xTB relative energies, RE_{xTB}, of $C_{76}Cl_{n+2}$ regioisomers, with n = 4–8. RE_{xTB} is compared with the relative energies predicted by (a) DNN-4, (b) XSI for $C_{76}Cl_6$, (c) DNN-6, (d) XSI for $C_{76}Cl_8$, (e) DNN-8, and (f) XSI for $C_{76}Cl_{10}$ models. R^2 and RMSD (in kcal/mol) between the xTB and the model-predicted values are indicated in each plot. The data points are colorized according to the NR values for the corresponding addition patterns.

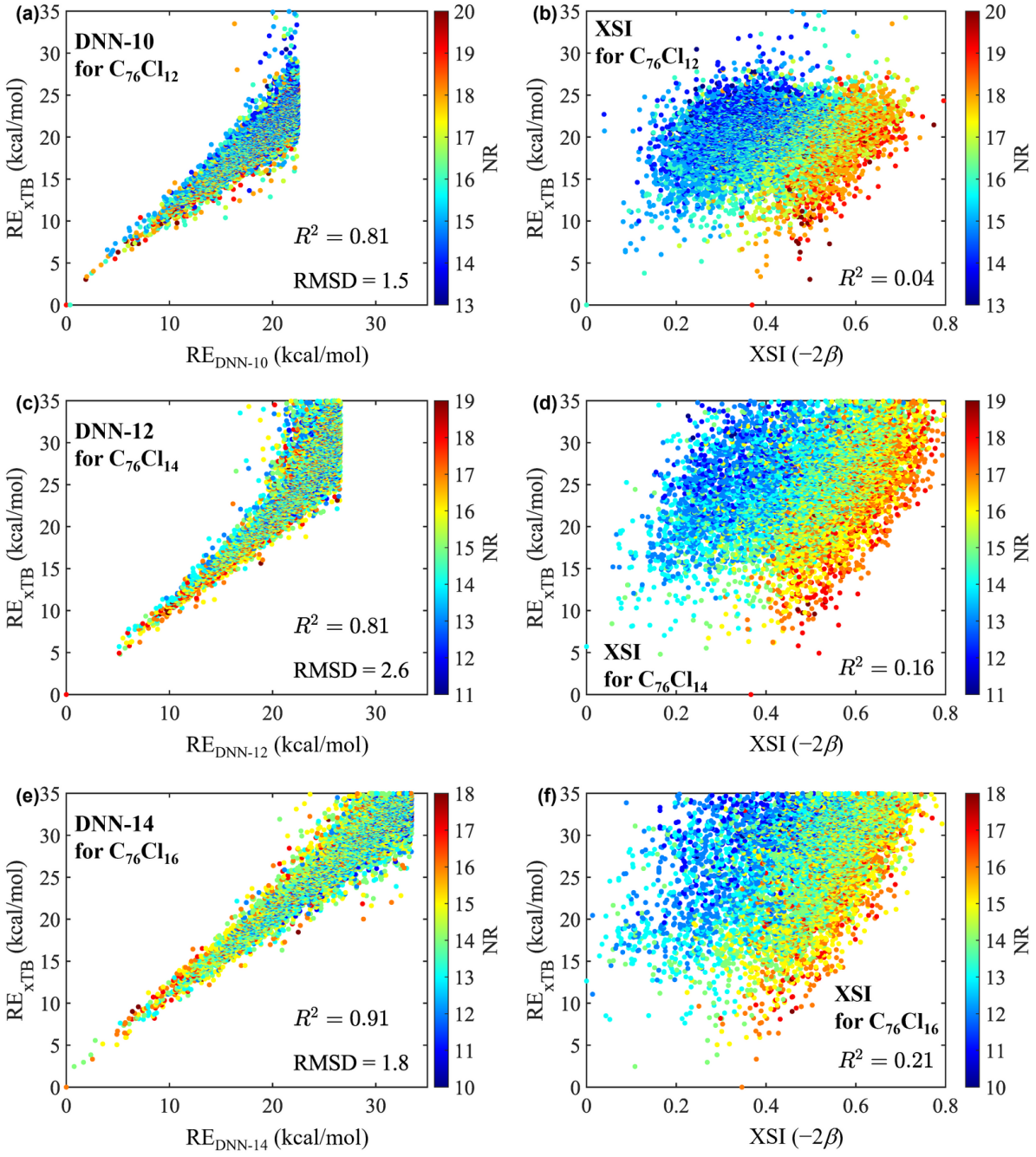

Figure S22. Idem Figure S21 for n = 10–14. The results are presented for (a) DNN-10, (b) XSI for $C_{76}Cl_{12}$, (c) DNN-12, (d) XSI for $C_{76}Cl_{14}$, (e) DNN-14, and (f) XSI for $C_{76}Cl_{16}$.

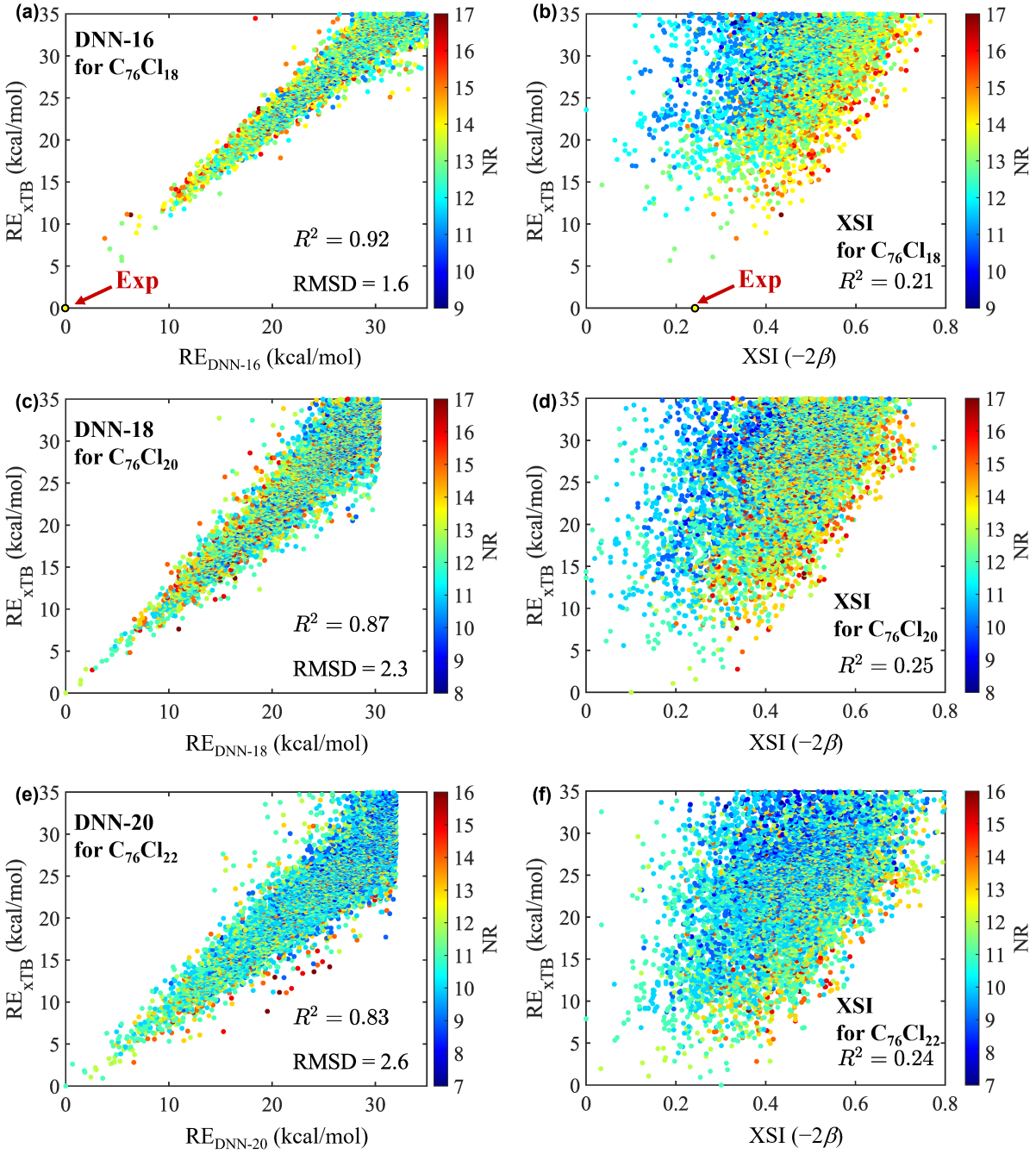

Figure S23. Idem Figure S21 for n = 16–20. The results are presented for (a) DNN-16, (b) XSI for $C_{76}Cl_{18}$, (c) DNN-18, (d) XSI for $C_{76}Cl_{20}$, (e) DNN-20, and (f) XSI for $C_{76}Cl_{22}$.

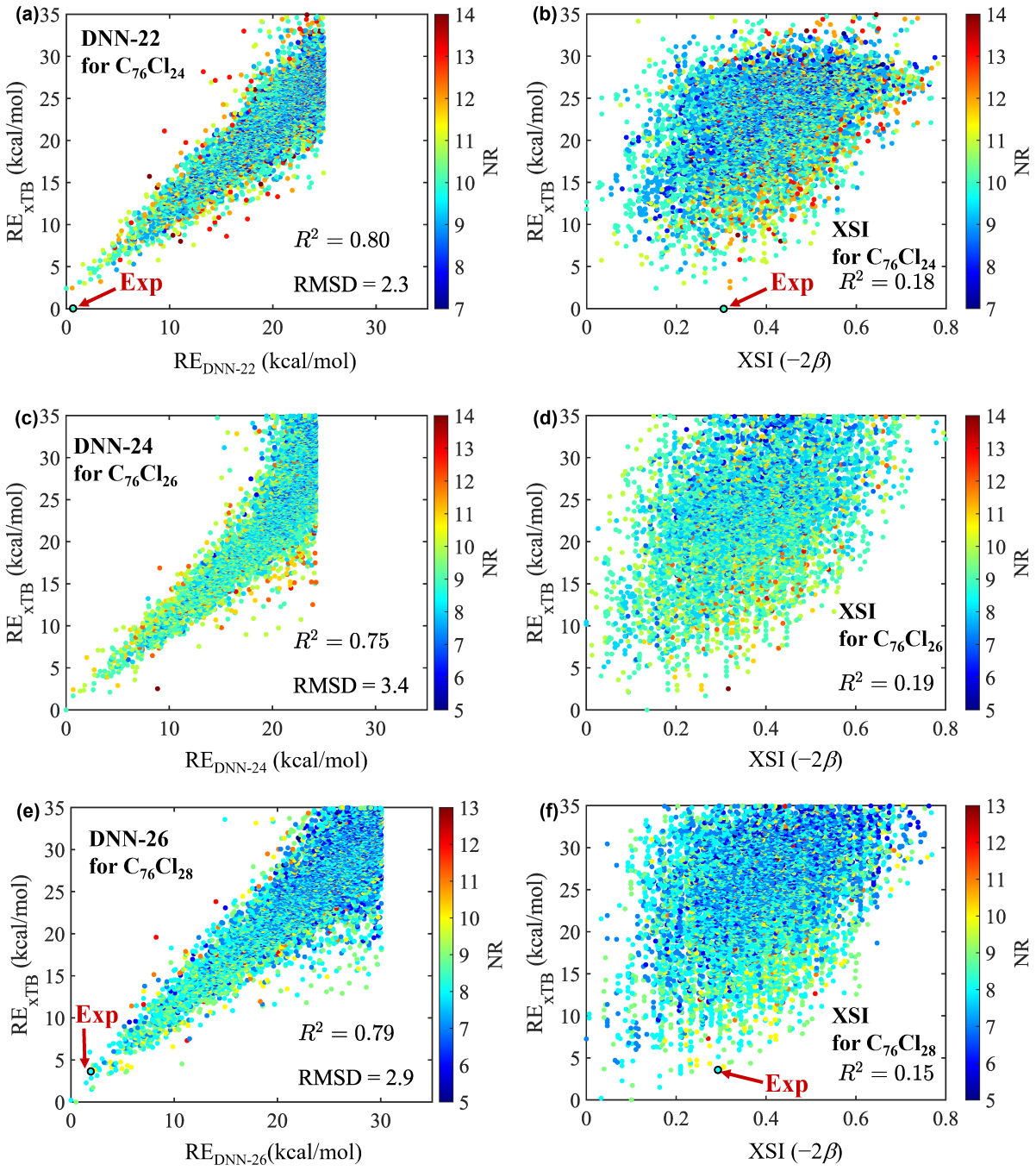

Figure S24. Idem Figure S21 for n = 22–26. The results are presented for (a) DNN-22, (b) XSI for $C_{76}Cl_{24}$, (c) DNN-24, (d) XSI for $C_{76}Cl_{26}$, (e) DNN-26, and (f) XSI for $C_{76}Cl_{28}$.

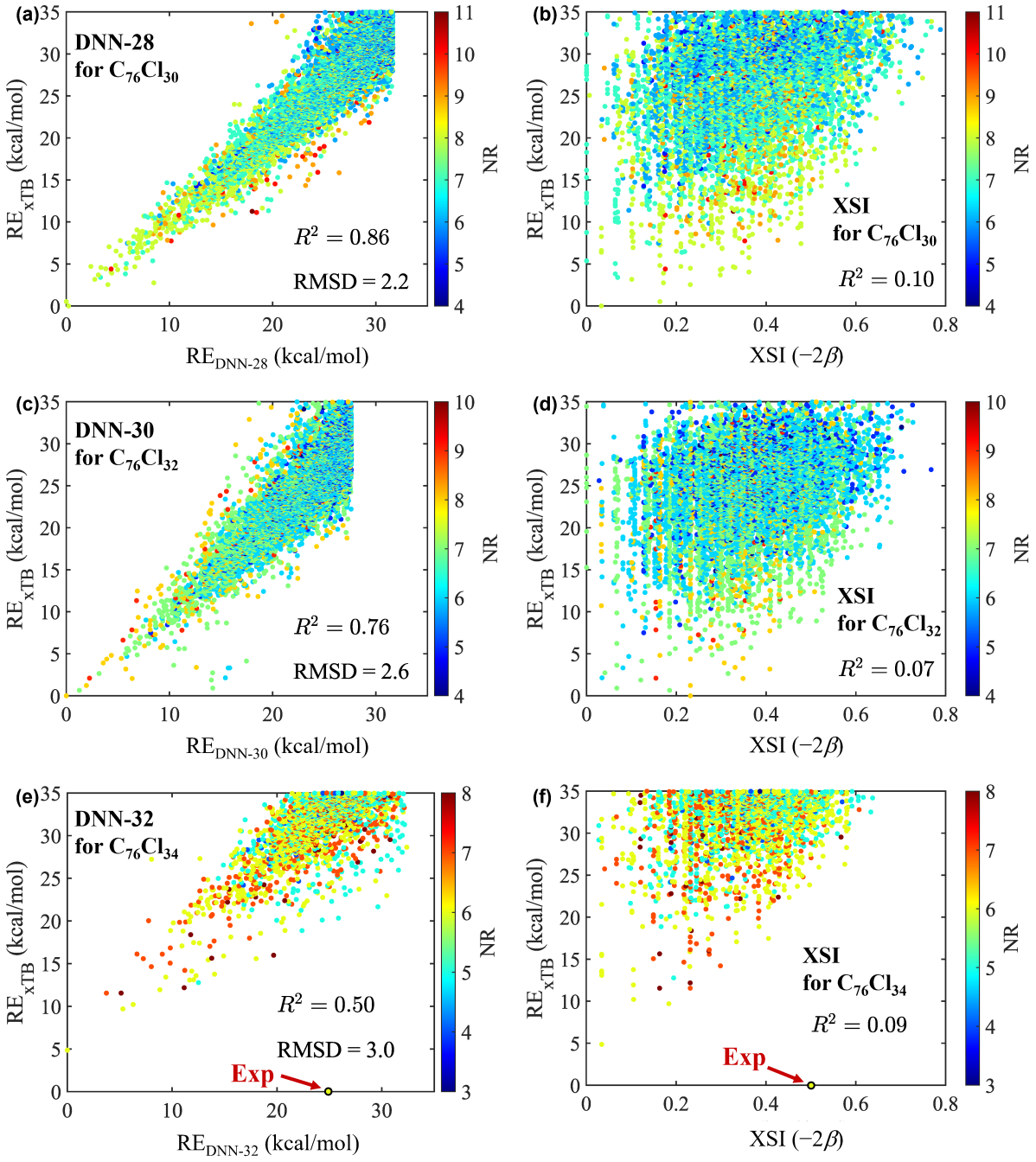

Figure S25. Idem Figure S21 for n = 28–32. The results are presented for (a) DNN-28, (b) XSI for $C_{76}Cl_{30}$, (c) DNN-30, (d) XSI for $C_{76}Cl_{32}$, (e) DNN-32, and (f) XSI for $C_{76}Cl_{34}$.

7 Lowest-energy addition patterns of hydrogenated carboncones

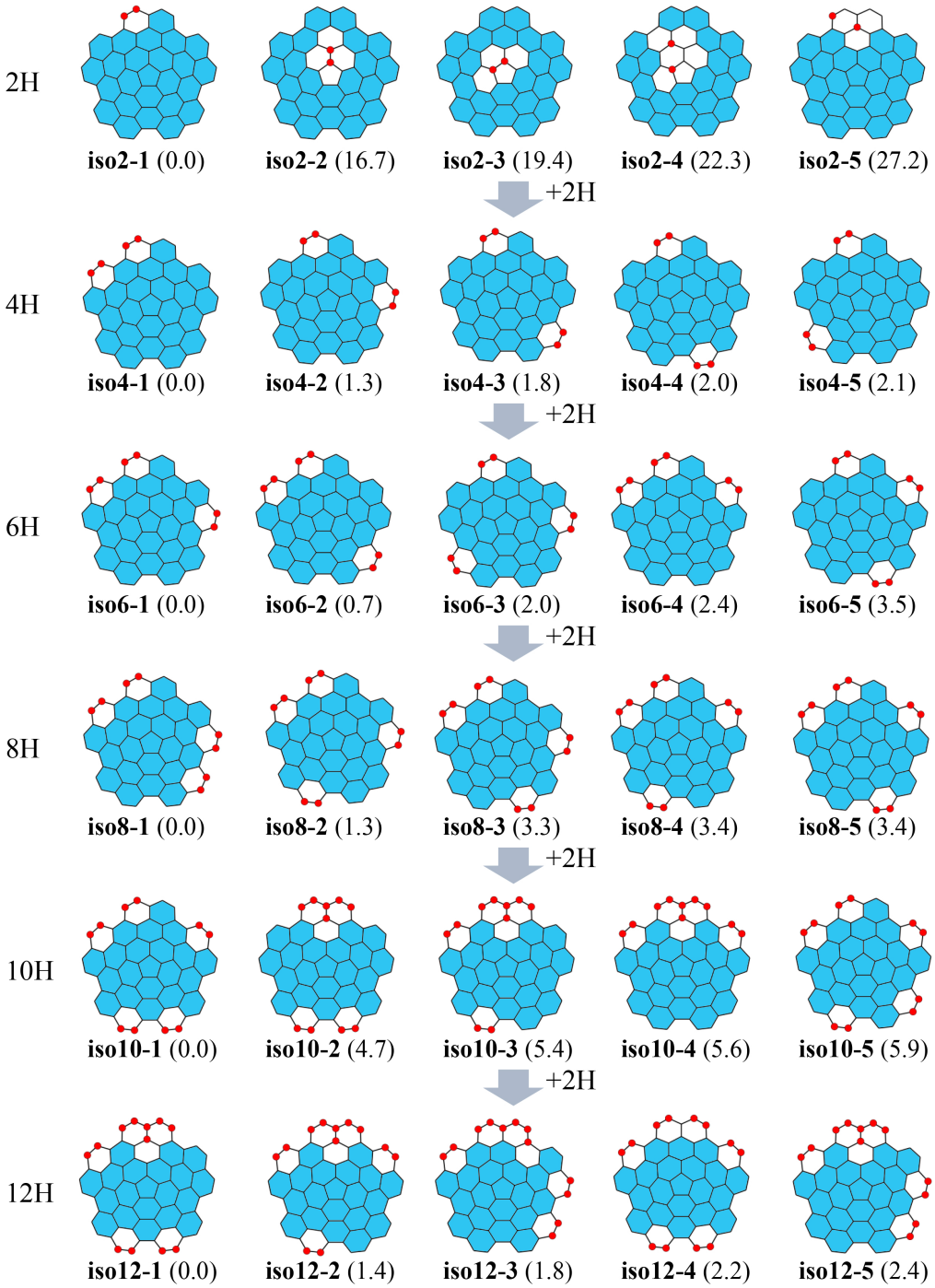

Figure S26. The five lowest-energy regioisomers $C_{70}H_{20+n}$ for n = 2–12. Below each of the addition pattern illustrations the DFT relative energy including ZPE are indicated in parentheses.

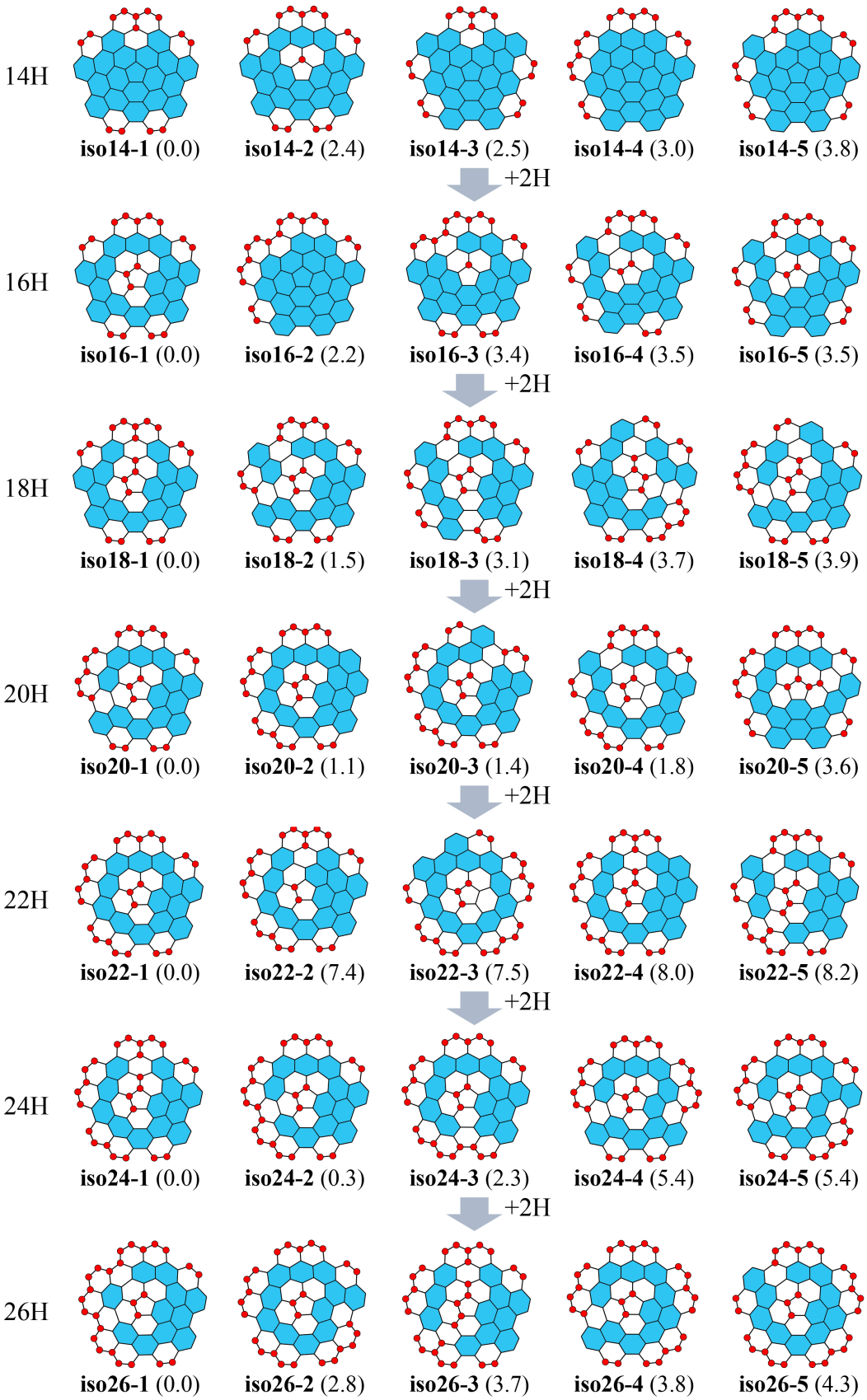

Figure S27. Idem Figure S26 for $C_{70}H_{20+n}$ with n = 14–26.

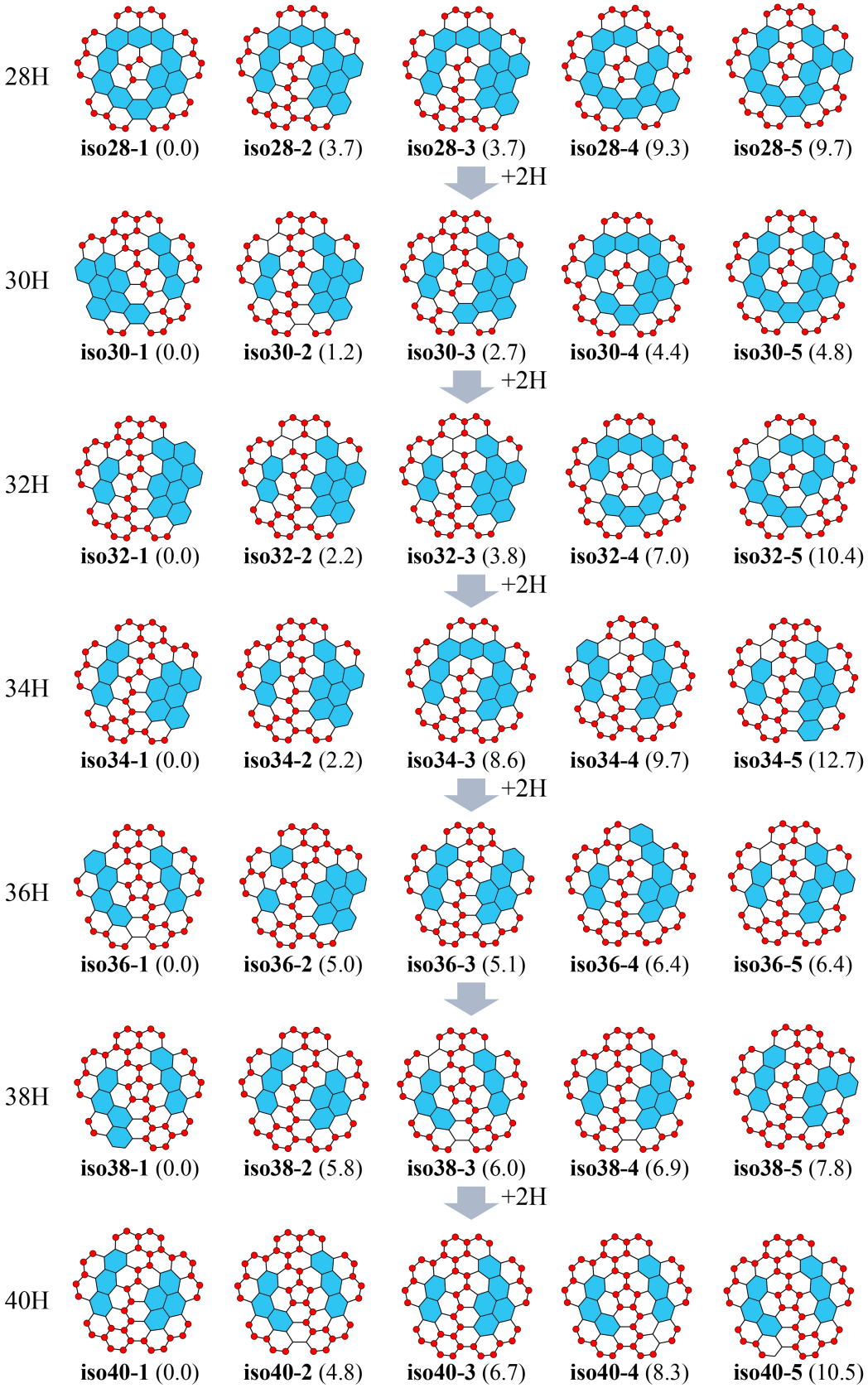

Figure S28. Idem Figure S26 for $C_{70}H_{20+n}$ with n = 28–40.

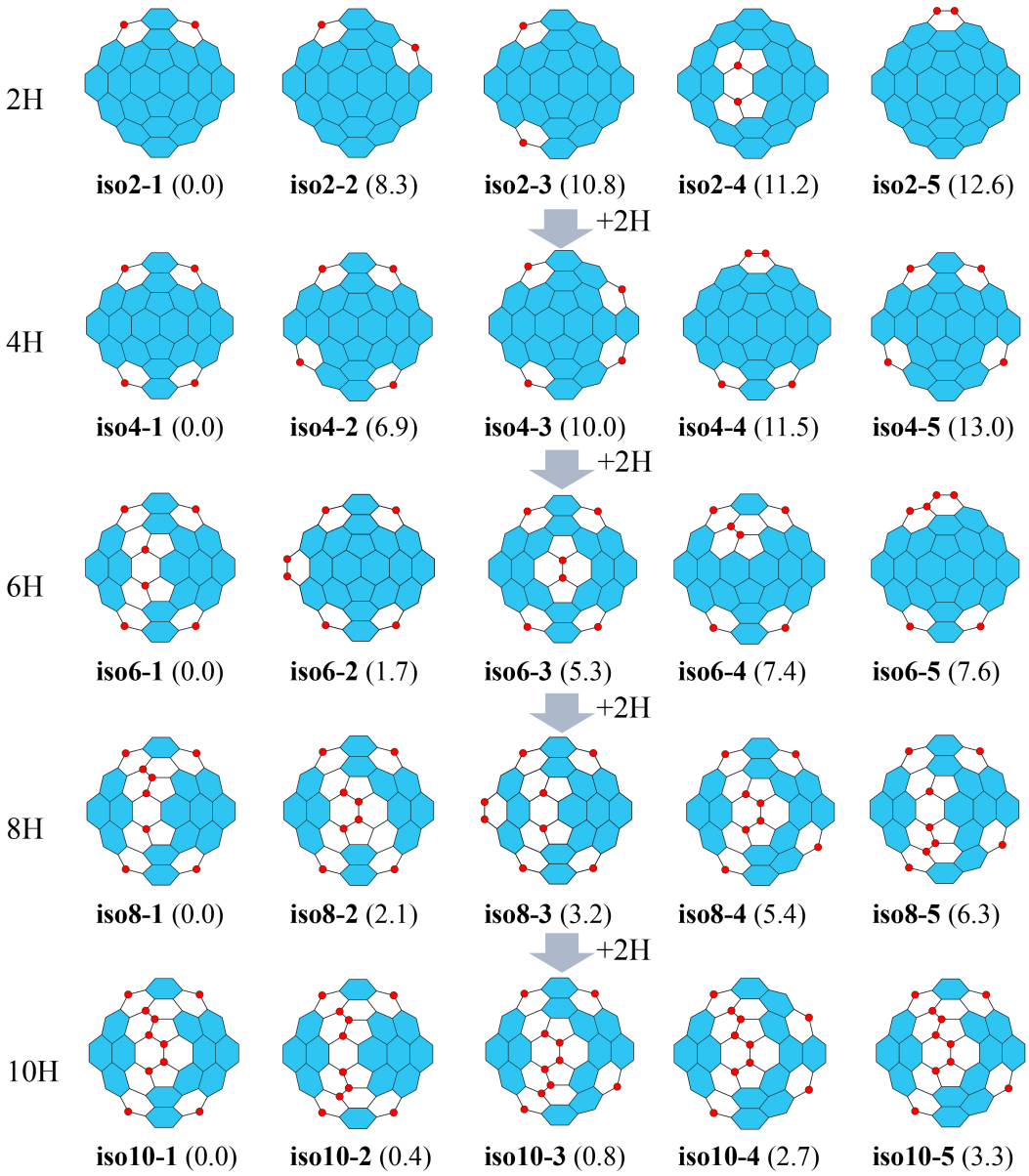

Figure S29. Idem Figure S26 for $C_{62}H_{16+n}$ with n = 2–10.

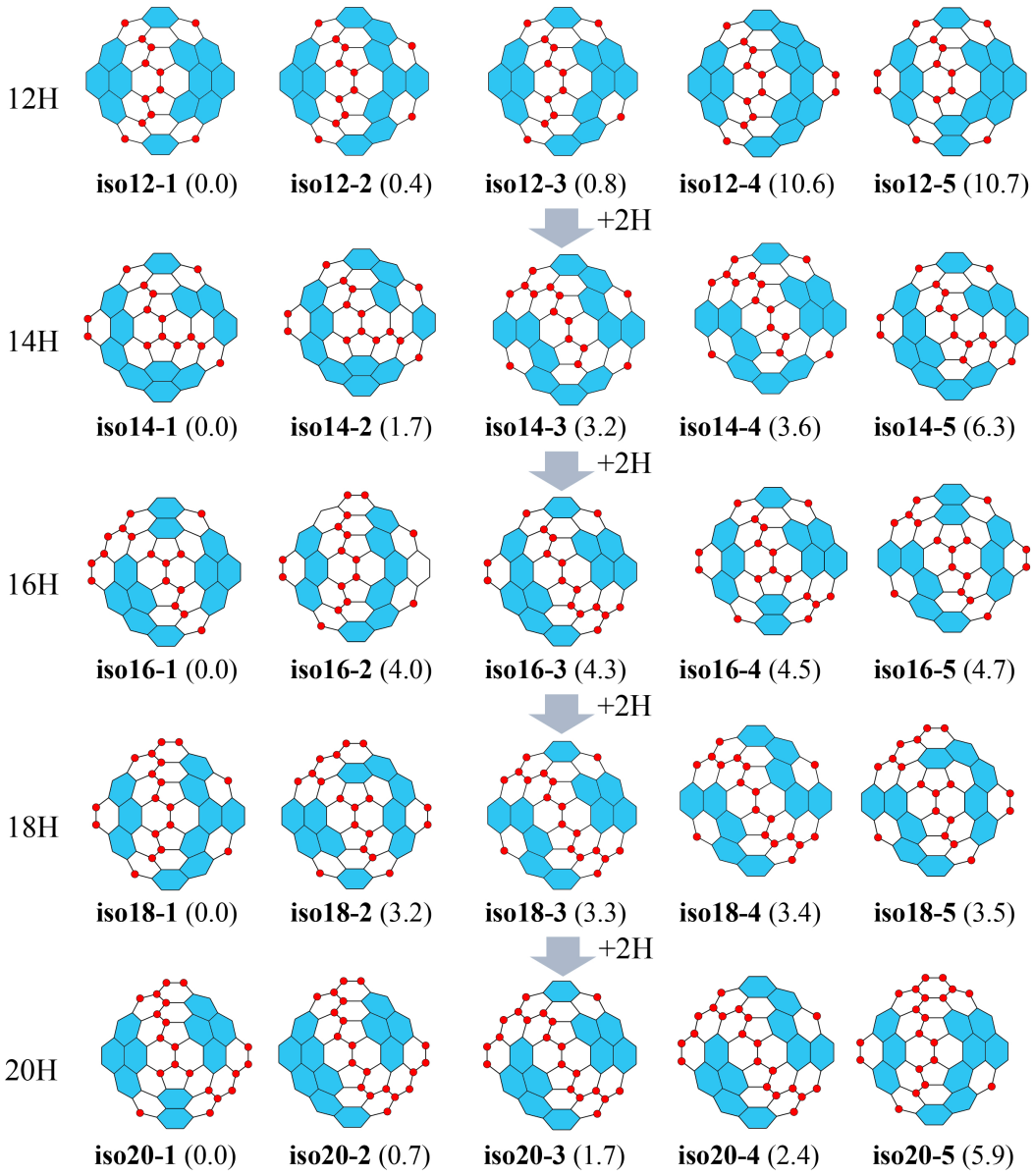

Figure S30. Idem Figure S26 for $C_{62}H_{16+n}$ with n = 12–20.

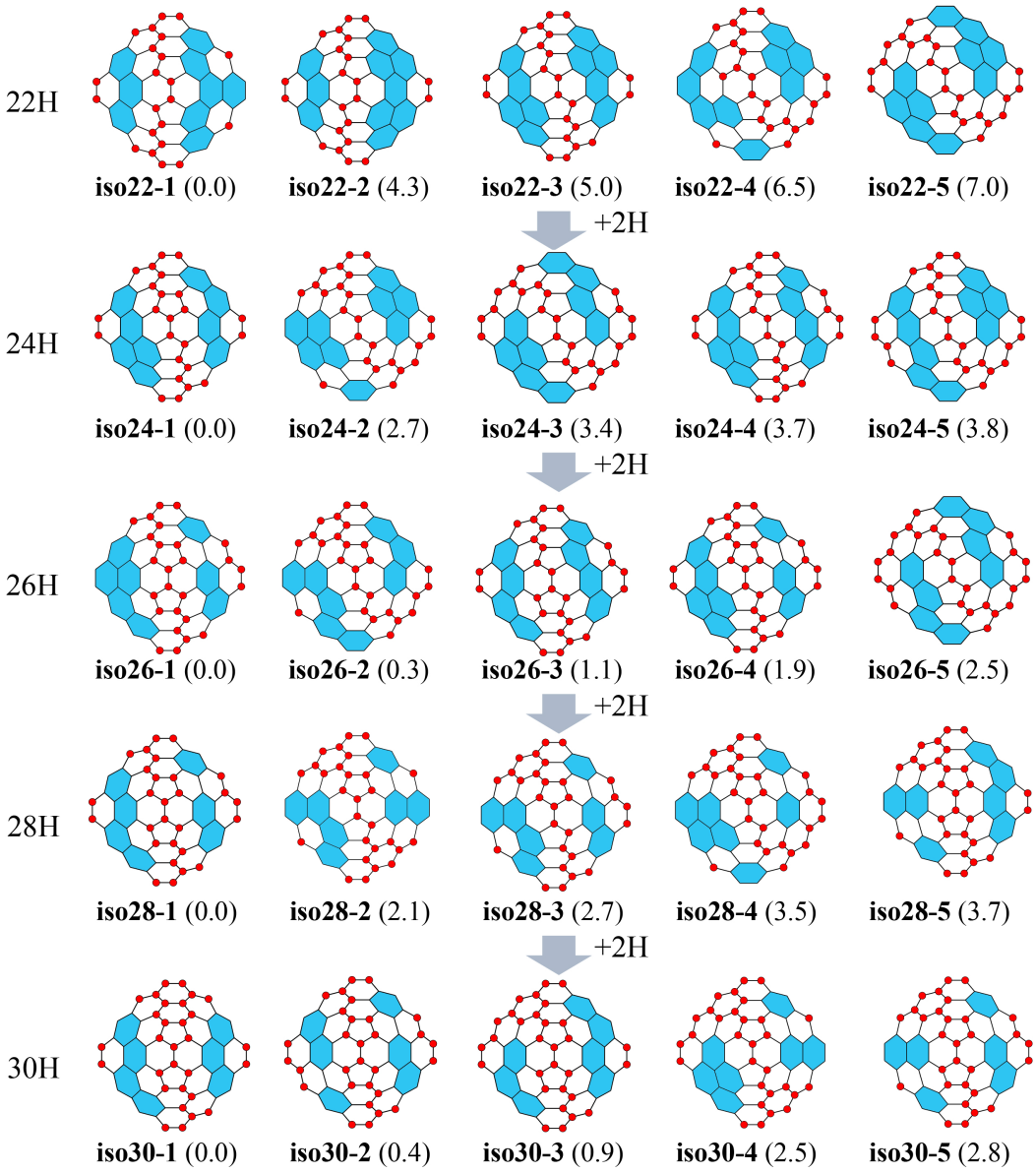

Figure S31. Idem Figure S26 for $C_{62}H_{16+n}$ with n = 22–30.

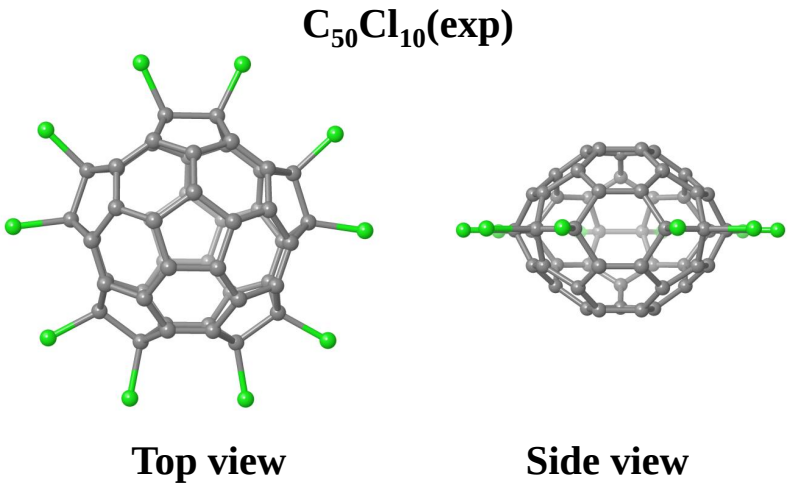

Figure S32. Top and side views of the DFT optimized structures of the experimentally synthesized regioisomer of $C_{50}Cl_{10}$.[1,2]

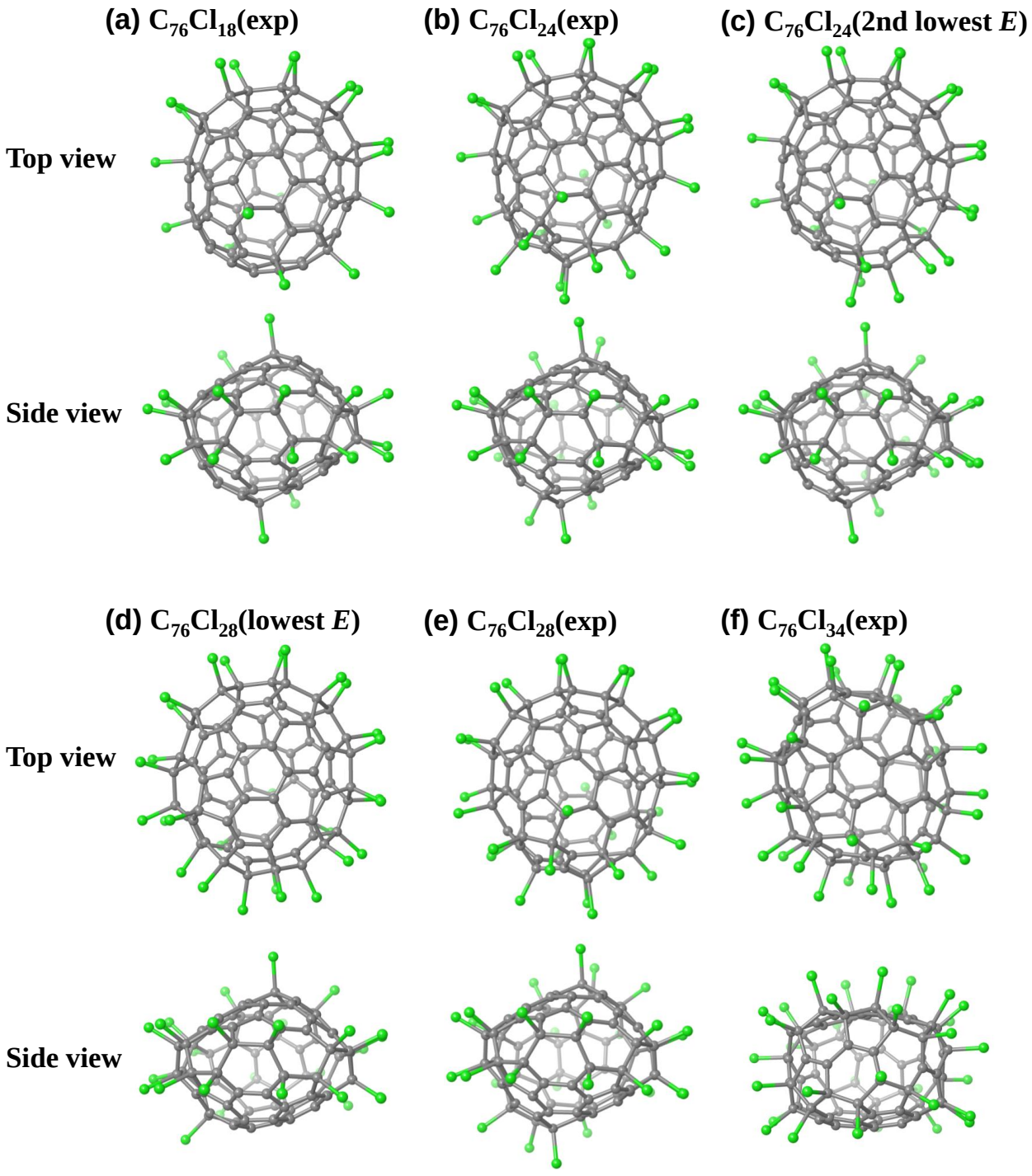

Figure S33. Top and side views of the DFT optimized structures of (a) experimental isomer of $C_{76}Cl_{18}$,[7] (b) experimental isomer of $C_{76}Cl_{24}$,[3] (c) second lowest-energy isomer of $C_{76}Cl_{24}$, (d) lowest-energy isomer of $C_{76}Cl_{28}$, (e) experimental isomer of $C_{76}Cl_{28}$,[3] and (f) experimental isomer of $C_{76}Cl_{34}$.[3]

8 Details on feature importance calculations

We have estimated the importances of the original, topology-based features using the strategy as follows. For all instances (regioisomers) in the test set, we replace the values of the feature in question with a universal constant, set as the mean value across all instances. This operation neutralizes the feature's impact by ensuring no variation in its value. Next, we apply the same PCA transformation as used in the original training phase to obtain reduced-dimensional features as inputs for the pre-tranined DDN model. The feature's importance is defined as the difference between the baseline score (R^2 between the predicted and actual relative energies) and the score after neutralization. As a result, if the difference is large, the corresponding feature has a significant impact on the model's performance. Conversely, if the difference small, then the feature plays a minor role in the model prediction.

References

(1) Xie, S.-Y.; Gao, F.; Lu, X.; Huang, R.-B.; Wang, C.-R.; Zhang, X.; Liu, M.-L.; Deng, S.-L.; Zheng, L.-S. Capturing the Labile Fullerene[50] as $C_{50}Cl_{10}$. *Science* **2004**, *304*, 699–699.

(2) Han, X.; Zhou, S.-J.; Tan, Y.-Z.; Wu, X.; Gao, F.; Liao, Z.-J.; Huang, R.-B.; Feng, Y.-Q.; Lu, X.; Xie, S.-Y.; Zheng, L.-S. Crystal Structures of Saturn-Like $C_{50}Cl_{10}$ and Pineapple-Shaped $C_{64}Cl_4$: Geometric Implications of Double- and Triple-Pentagon-Fused Chlorofullerenes. *Angew. Chem. Int. Ed.* **2008**, *47*, 5340–5343.

(3) Ioffe, I. N.; Mazaleva, O. N.; Chen, C.; Yang, S.; Kemnitz, E.; Troyanov, S. I. C_{76} fullerene chlorides and cage transformations. Structural and theoretical study. *Dalton Trans.* **2011**, *40*, 11005–11011.

(4) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **1976**, *32*, 922–923.

(5) Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **1978**, *34*, 827–828.

(6) Umeyama, S. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 376–380.

(7) Simeonov, K. S.; Amsharov, K. Y.; Jansen, M. Connectivity of the Chiral D2-Symmetric Isomer of C_{76} through a Crystal-Structure Determination of $C_{76}Cl_{18}{\cdot}TiCl_4$. *Angew. Chem. Int. Ed.* **2007**, *46*, 8419–8421.