

## Electronic Supplementary Information for

### Polycyclic Aromatic Hydrocarbons:

### From Small Molecules Through Nano-Sized Species Towards Bulk Graphene

Bun Chan<sup>a</sup> and Amir Karton<sup>b</sup>

<sup>a</sup> Graduate School of Engineering, Nagasaki University, Bunkyo 1-14, Nagasaki-shi, Nagasaki 852-8521, Japan.

<sup>b</sup> School of Molecular Sciences, University of Western Australia, Crawley, Perth, WA 6009, Australia

#### Table of Contents

Full set of results corresponding to Table 2 of the main text	S2
Reference and highest-level calculated heats of formation	S3
Working example for constructing isodesmic-type scheme	S4–S7
Results of assessments presented in the unit of kcal mol <sup>-1</sup>	S8–S9

**Table S1** Mean absolute deviation (MAD, kJ mol<sup>-1</sup>) for the bond-separation reactions (set A, Fig. 1 in the main text) and reactions constructed for better preservation of chemical characteristics (set B)

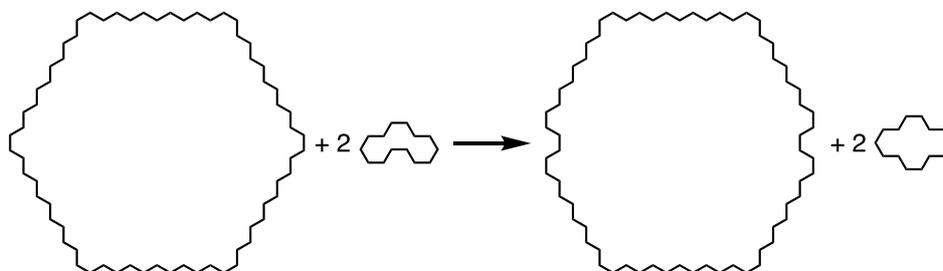
Method	Set A		Set B	
W1X-1	0.2		0.1	
WG	5.1		1.0	
G4	16.9		3.9	
CBS-QB3	17.2		3.6	
G4(MP2)	9.1		2.5	
G4(MP2)-6X	8.1		1.9	
G4(MP2)-XK	4.8		1.1	
DSD-PBEP86	15.0		1.3	
MPW2PLYP	5.2		1.6	
B2PLYP	8.3		1.5	
LMPW2PLYP	4.7		2.0	
	def2-TZVP	def2-SVP	def2-TZVP	def2-SVP
B3LYP	16.1	21.1	3.9	6.6
PBE0	6.4	9.6	2.6	4.5
TPSSh	8.5	12.9	2.6	4.0
SCAN0	7.8	9.5	2.6	4.2
$\omega$ B97M-V	7.7	12.5	2.8	5.7
MN15	5.2	7.1	2.3	3.4
BLYP	16.5	24.3	3.1	6.2
PBE	5.5	10.5	2.9	4.1
TPSS	8.4	13.5	3.1	4.0
SCAN	11.2	11.3	2.2	3.3
B97M-V	23.5	13.5	3.1	3.4
MN15-L	32.1	19.4	6.5	3.7
PM7	43.5		8.8	
DC-DFTB3	7.9		4.4	

**Table S2** Reference 298 K heats of formation (kJ mol<sup>-1</sup>)

species	method	calc	expt
methane	ATcT		-74.53 ± 0.055
ethene	ATcT		52.35 ± 0.12
ethane	ATcT		-83.98 ± 0.13
butadiene	ATcT		110.82 ± 0.36
benzene	W1-F12	83.1	82.9 ± 0.9
toluene	W1-F12	50.8	50.1 ± 1.1
indene	W1-F12	159.8	161.2 ± 2.3
indane	W1-F12	58.2	60.9 ± 2.1
naphthalene	W1-F12	148.9	150.6 ± 1.5
acenaphthylene	W1-F12	262.9	263.2 ± 3.7
biphenylene	W1-F12	419.4	417.2 ± 1.9
acenaphthene	W1-F12	154.0	156.8 ± 3.1
biphenyl	W1-F12	179.8	180.3 ± 3.3
fluorene	W1-F12	185.9	176.7 ± 3.1
diphenylmethane	W1-F12	164.9	163.7 ± 2.3
pyracylene	W1-F12	433.0	408.6
anthracene	W1-F12	229.7	229.4 ± 2.9
phenanthrene	W1-F12	205.7	202.2 ± 2.3
pyracene	W1-F12	190.7	174.3 ± 5.3
pyrene	W1-F12	225.9	225.5 ± 2.5
11H-Benzo-b-fluorene	W1-F12	249.6	
chrysene	W1-F12	268.6	268.7 ± 4.7
<i>m</i> -terphenyl	W1-F12	275.5	280.0 ± 3.9
<i>p</i> -terphenyl	W1-F12	275.5	284.4 ± 3.8
perylene	MPW2PLYP	322.3	318.3 ± 3.7
anthanthrene	MPW2PLYP	323.9	
coronene	MPW2PLYP	295.8	295 ± 11
ovalene	MPW2PLYP	400.2	
circumanthracene	MPW2PLYP	520.5	
circumcoronene	MPW2PLYP	625.6	
C <sub>96</sub> H <sub>24</sub>	MN15	1114.1	
C <sub>150</sub> H <sub>30</sub>	MN15	1677.0	
C <sub>216</sub> H <sub>36</sub>	MN15	2335.6	

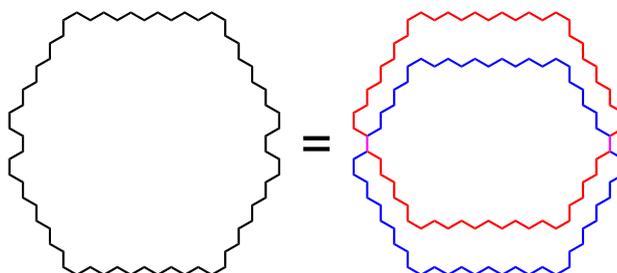
### Working Example for Constructing Isodesmic-Type Scheme

In this case study with  $C_{216}H_{36}$ , let us begin by considering the general approach for the deconstruction. We first chop off two “vertices” of  $C_{216}H_{36}$  with the equation  $C_{216}H_{36} + 2 C_{14}H_{10} \rightarrow C_{212}H_{36} + 2 C_{16}H_{10}$ :

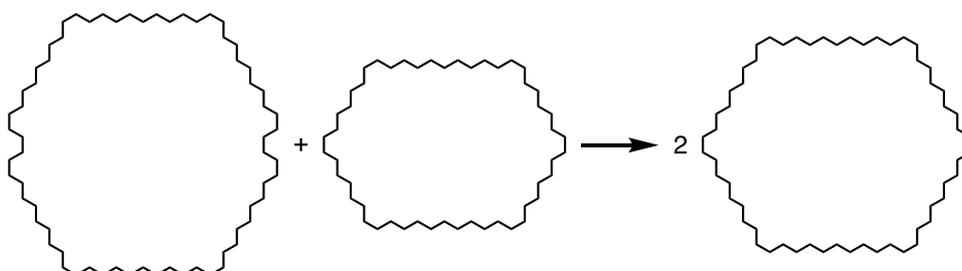


Importantly, the four carbon atoms in these two vertices are represented as differences in two similar aromatic molecules phenanthrene ( $C_{14}H_{10}$ ) and pyrene ( $C_{16}H_{10}$ ). If one instead employs the bond separation approach (reaction 24' in the main text), the ethane molecules would not adequately account for the aromatic characters for the connections between the four vertex carbons and the main body of the resulting  $C_{212}H_{36}$  molecule.

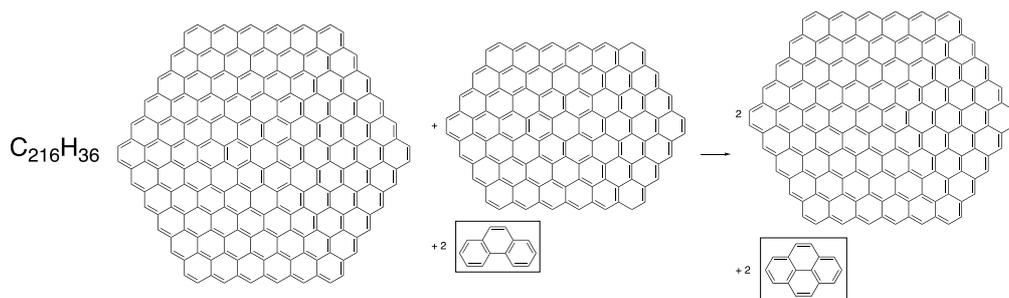
Next, we treat this  $C_{212}H_{36}$  molecule as an overlap of two identical halves:



This enables us to straightforwardly construct a reaction with these halves and the overlapping region:

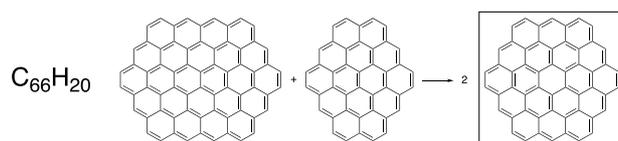
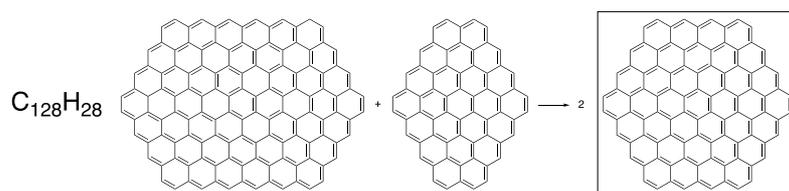
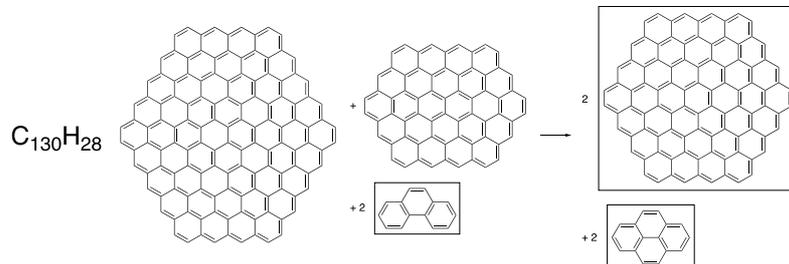
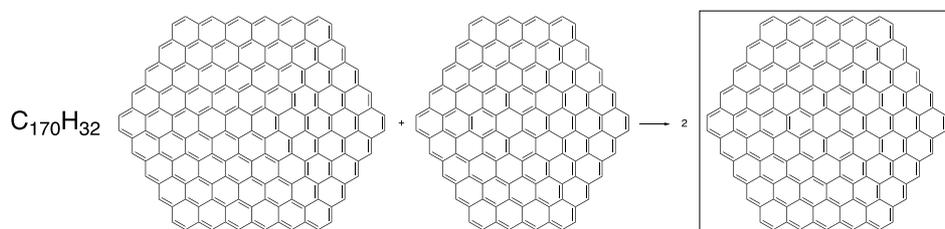


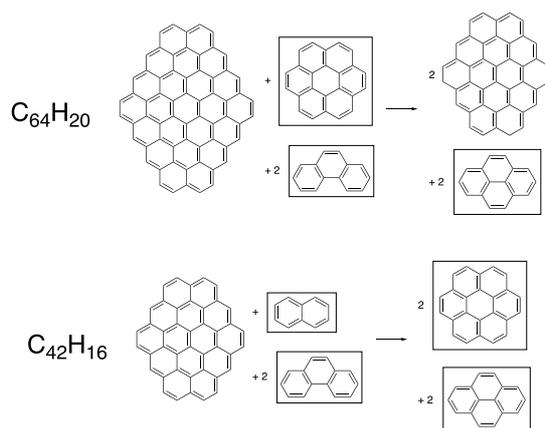
Putting all of above together, we arrive at our first reaction for the deconstruction of  $C_{216}H_{36}$ :



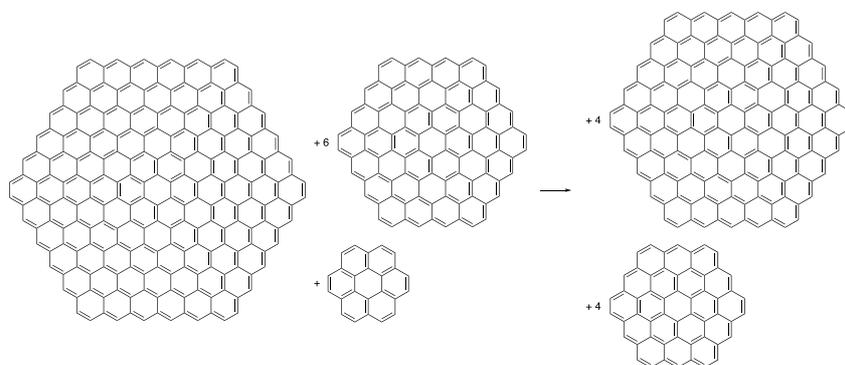
Note that  $C_{14}H_{10}$  and  $C_{16}H_{10}$ , shown in the boxes, are among the molecules that we use in the present study.

We now further deconstruct the two intermediate species that are not in our molecule set using the same general approach, and then repeat the process for all intermediate species. This leads to a series of reactions:

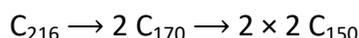




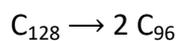
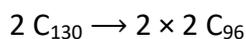
Eventually, the last reaction in the series contains only the target molecule plus spectator species that are all in our molecule set. Combining all steps together and eliminate identical species on the two sides of the equation leads to the final formula:



The four  $C_{150}H_{30}$  molecules come from the sequence:

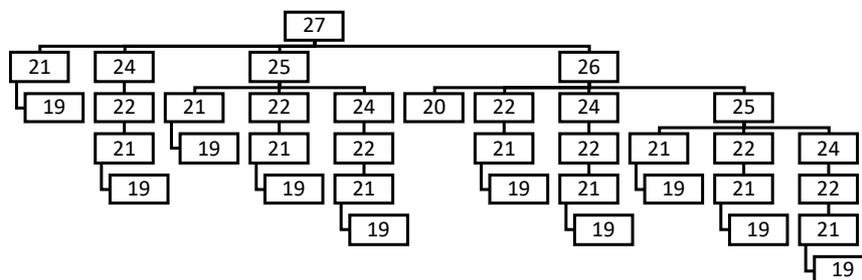


They represent the main body of the target molecule. The six  $C_{96}H_{24}$  molecules come from the intermediate steps:

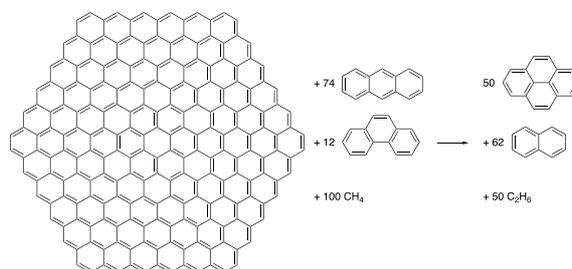


The molecules  $C_{130}H_{28}$  and  $C_{128}H_{28}$  are the overlapping regions for, respectively, the  $C_{170}H_{32}$  and  $C_{216}H_{26}$  equations. Thus, the  $C_{96}H_{24}$  molecules are the main representations for the overlapping regions. Finally, the  $C_{54}H_{18}$  and  $C_{24}H_{12}$  species are results of deconstructing smaller intermediates  $C_{66}H_{20}$  and  $C_{64}H_{20}$ , respectively.

We can also deconstruct the final isodesmic-type reaction into smaller fragments for which W1-F12 reference values are available, by combining the reactions 19–27 in Figs. 2 and 3 of the main text:



This then leads to the following formula:



It involves a substantial number of species. In addition, the aliphatic  $\text{CH}_4$  and  $\text{C}_2\text{H}_6$  molecules also enter the equation. These factors would lead to a large uncertainty in the calculated reaction energy.

Let us consider this from another perspective. An advantage of applying the chemistry-conserving reaction  $\text{C}_{216}\text{H}_{36} + 6 \text{C}_{96}\text{H}_{24} + \text{C}_{24}\text{H}_{12} \rightarrow 4 \text{C}_{150}\text{H}_{30} + 4 \text{C}_{54}\text{H}_{18}$  is that the atomization energies for the spectator species ( $\text{C}_{24}\text{H}_{12}$ ,  $\text{C}_{54}\text{H}_{18}$ ,  $\text{C}_{96}\text{H}_{24}$  and  $\text{C}_{150}\text{H}_{30}$ ) are obtained at the MPW2PLYP level. In comparison, the use of just the small molecules is equivalent to applying the lower-level MN15 method to the intermediate steps, thus leading to less accurate results.

**Table S3** Results in kcal mol<sup>-1</sup> corresponding to Table 1 of the main text

Method	MAD	MD	SD	LD
Composite methods				
(1) W1X-1	0.1	0.0	0.1	0.1
(2) WG	1.0	1.0	0.5	1.7
(3) G4	3.5	3.5	1.3	5.8
(4) CBS-QB3	3.4	3.4	1.5	6.1
(5) G4(MP2)	2.0	2.0	0.7	3.0
(6) G4(MP2)-XK	1.0	1.0	0.4	2.2
Double-hybrid DFT				
(7) DSD-PBEP86	2.8	2.8	1.7	5.8
(8) MPW2PLYP	1.1	0.1	1.4	2.8
(9) B2PLYP	1.7	0.8	1.8	3.8
(10) LMPW2PLYP	1.1	-0.4	1.3	-2.7
Hybrid DFT				
(11) B3LYP	3.4	-3.4	2.0	-7.4
(12) PBE0	1.5	-1.1	1.7	-4.9
(13) TPSSh	2.1	-1.5	2.4	-6.6
(14) SCAN0	1.7	0.3	2.0	4.1
(15) $\omega$ B97M-V	1.5	-1.5	1.2	-4.2
(16) MN15	1.2	0.8	1.3	3.0
Non-hybrid DFT				
(17) BLYP	3.6	-3.3	2.5	-7.9
(18) PBE	1.6	-0.6	2.0	-4.3
(19) TPSS	2.2	-1.4	2.6	-6.5
(20) SCAN	2.2	1.7	2.3	6.4
(21) B97M-V	4.5	4.4	3.4	10.9
(22) MN15-L	6.4	6.4	4.0	12.2

**Table S4** Results in kcal mol<sup>-1</sup> corresponding to Table S2 of this document

Method	Set A		Set B	
W1X-1	0.1		0.0	
WG	1.2		0.2	
G4	4.0		0.9	
CBS-QB3	4.1		0.9	
G4(MP2)	2.2		0.6	
G4(MP2)-6X	1.9		0.5	
G4(MP2)-XK	1.1		0.3	
DSD-PBEP86	3.6		0.3	
MPW2PLYP	1.3		0.4	
B2PLYP	2.0		0.4	
LMPW2PLYP	1.1		0.5	
	def2-TZVP	def2-SVP	def2-TZVP	def2-SVP
B3LYP	3.9	5.0	0.9	1.6
PBE0	1.5	2.3	0.6	1.1
TPSSh	2.0	3.1	0.6	1.0
SCAN0	1.9	2.3	0.6	1.0
$\omega$ B97M-V	1.8	3.0	0.7	1.4
MN15	1.2	1.7	0.6	0.8
BLYP	3.9	5.8	0.7	1.5
PBE	1.3	2.5	0.7	1.0
TPSS	2.0	3.2	0.7	1.0
SCAN	2.7	2.7	0.5	0.8
B97M-V	5.6	3.2	0.7	0.8
MN15-L	7.7	4.6	1.5	0.9
PM7	10.4		2.1	
DC-DFTB3	1.9		1.1	