Supporting Information for: The Homodesmotic Reference: A Two-Regime Geometric Model for Cycloalkane Ring Strain

Yiming Ma

S1 Supplementary Computational Details

S1.1 CREST Conformational Search Parameters

All initial CREST calculations used the command:

```
crest <input>.xyz --gfn2 --chrg 0 --uhf 0 --T 8 --mdlen 6.0 --metadyn --ewin 10.0
```

The metadynamics simulation length was 6.0 ps; the energy window was set to 10.0 kcal mol^{-1} to capture a broad ensemble of conformers. Subsequently, a refined RT (Room Temperature) filtering was applied (see Section S1.2) to retain only conformers within an energy window of 0.59 kcal mol^{-1} above the global minimum before proceeding to high-level DFT optimizations.

S1.2 Boltzmann Weighting

After final composite energy calculations, for molecules possessing multiple conformers within a Gibbs free energy window $\Delta G \leq 0.59$ kcal mol^{-1} (which corresponds to the thermal energy kT at 298.15 K), Boltzmann-weighted averaging was performed:

$$G_{\text{final}} = \sum_i w_i G_i, \quad w_i = \frac{e^{-G_i/kT}}{\sum_j e^{-G_j/kT}},$$

where $kT = 0.59$ kcal mol^{-1} $= 0.000944$ Hartree. This window is based on the approximate thermal accessibility at room temperature.

S1.3 Composite Energy Scheme

$$G_{\text{final}} = E_{\text{CC}} + (G_{\text{DFT}} - E_{\text{DFT}})$$

Similarly, the composite enthalpy H_{final} is obtained by replacing G_{DFT} with H_{DFT}:

$$H_{\text{final}} = E_{\text{CC}} + (H_{\text{DFT}} - E_{\text{DFT}})$$

Here, E_{CC} refers to the single-point energy computed at the DLPNO-CCSD(T)/def2-TZVPP level, while E_{DFT}, G_{DFT}, and H_{DFT} are obtained at the B3LYP-D3(BJ)/def2-QZVPP level.

S1.4 Descriptor Definitions

See Table S1.

All DFT calculations used RI-J and COSX approximations, the def2/J auxiliary basis set, and Grimme's D3(BJ) dispersion correction.[**Grimme2011**, **Neese2020**, **Weigend2005**]

Table 1: Descriptor definitions.

Descriptor	Formula	Description
$\Delta\theta^2$	$(\theta - \theta_0)^2$	Bond angle deviation squared ($^{\circ 2}$)
ΔR^2	$(R - R_0)^2$	Bond length deviation squared (Å^2)
Θ	—	Traceless quadrupole moment (a.u.)
μ	—	Dipole moment (a.u.)
$\lvert Q_3 \rvert$	—	Octupole moment (a.u.)
$\langle r^2 \rangle$	—	Electronic spatial extent (a.u.)
q_C	—	ADCH charge averaged over carbon atoms (e)

S2 Carbon Rings (C3–C14): Energies, Weights, Descriptors

S2.1 Composite Gibbs Free Energies and Enthalpies with Boltzmann Weights

Tables S2 and S3 list composite energies and weights for cycloalkanes and linear alkanes, respectively. The G_{CC} and H_{CC} values are computed using the high-precision raw data (see Section S1.3).

Table 2: Composite energies and weights for cycloalkanes (C3–C14).

Molecule	Conformer	E_{CC} (Eh)	G_{DFT} (Eh)	E_{DFT} (Eh)	G_{CC} (Eh)	H_{CC} (Eh)
C3	C3_1	−117.671774	−117.817322	−117.874074	−117.61502158	−117.58642311
C4	C4_1	−156.914939	−157.098385	−157.182571	−156.83075249	−156.79950317
C5	C5_1	−196.186648	−196.409400	−196.520885	−196.07516320	−196.04033163
C6	C6_1	−235.436817	−235.697112	−235.837995	−235.29593373	−235.26048769
C7	C7_1	−274.666941	−274.968011	−275.135805	−274.49914344	−274.46055135
	C7_2	−274.661996	−274.963679	−275.130541	−274.4941[†]	−274.4555[†]
	Weighted avg				−313.70581902	−313.66492011
C8	C8_1	−313.901871	−314.221448	−314.417366	−313.70595264	−313.66505374
	C8_2	−313.898504	−314.241817	−314.435633	−313.70468763	−313.66378873
	Weighted avg				−313.70581902	−313.66492011
C9	C9_1	−353.138495	−353.517593	−353.741052	−352.91503571	−352.87199543
	C9_2	−353.134934	−353.515819	−353.738210	−352.91254349	−352.86950321
	C9_3	−353.136809	−353.517531	−353.740349	−352.91399136	−352.87095108
	Weighted avg				−352.91485152	−352.87181207
C10	C10_1	−392.378553	−392.797833	−393.047110	−392.12927594	−392.08379514
C11	C11_1	−431.619349	−432.079323	−432.356702	−431.34197021	−431.29340300
	C11_2	−431.619253	−432.079043	−432.356414	−431.34188241	−431.29331520
	C11_3	−431.617807	−432.077812	−432.355244	−431.34037459	−431.29180738
	Weighted avg				−431.34187462	−431.29330741
C12	C12_1	−470.867422	−471.366453	−471.671698	−470.56217658	−470.51175673
C13	C13_1	−510.107325	−510.646524	−510.975753	−509.77809652	−509.72406932
C14	C14_1	−549.352210	−549.931748	−550.288085	−548.99587266	−548.93920852

[†]Individual G_{CC} and H_{CC} for C7_2 are not listed; its contribution to the weighted average is negligible due to the very small weight (0.001).

S2.2 Descriptors for Carbon Rings

Table S4 lists all descriptors for carbon rings. The last column of the original table (ring-opening energy with $m = 1$) has been removed to avoid confusion with the strain energies computed using the homodesmotic or fixed-reference reactions discussed in Section S4.

S3 Full-Subset Regression Results for the Ring-Opening Reaction Energy

The ring-opening reaction energies were computed using both composite Gibbs free energies (G_{CC}) and composite enthalpies (H_{CC}). The same set of 8 descriptors was used for full-subset OLS regression (1–3 descriptors). Models were retained if $R^2 > 0.95$, RMSE < 0.5 kcal mol^{-1}, and maximum VIF < 5. The top models are listed in Tables S5a and S5b. A direct comparison of the best models is given in Table S6. Note that the regression coefficients for bond-length stretching (ΔR^2) are in units of kcal mol^{-1} Å^{-2}, and those for angle bending ($\Delta\theta^2$) are in kcal mol^{-1} deg^{-2}.

As shown in Table S6, the enthalpy-based composite (H_{CC}) yields a substantially smaller intercept (0.131 vs. 0.718) while maintaining essentially identical geometric coefficients. This near-zero intercept under H_{CC} is a recurring theme that culminates in the identification of the homodesmotic reaction as the physically most meaningful scheme (Section S4).

Reference Molecules

Table 9 lists the reference molecules used to obtain unstrained geometric parameters for carbon rings.

S4 Reaction Schemes: Forced Geometric Regression vs. Experiment

To systematically evaluate which reaction scheme yields strain energies that are both statistically accurate and physically interpretable, we performed two complementary scans.

S4.1 Fair Arena: Forced Two-Descriptor Regression

All reaction schemes were evaluated using only two geometric descriptors, $\Delta\theta^2$ and ΔR^2. Each scheme's calculated strain energies were (i) compared directly against experimental values (MAE, RMSE, R^2_{exp}), and (ii) regressed against the two geometric descriptors to obtain the intercept and coefficients (R^2_{geom}). To assess physical consistency, we define three criteria (the "physical standard"):

- Intercept |Intercept| < 0.1 kcal mol^{-1} (ideally near zero),
- $\Delta\theta^2$ coefficient c_θ in the range 0.0009–0.0010 kcal mol^{-1} deg^{-2},
- ΔR^2 coefficient c_R in the range 10000–11000 kcal mol^{-1} Å^{-2}.

These ranges are derived from the robust correlation observed in CarbonQSAR and reflect the intrinsic deformation–energy relationship where bond-length stretching dominates over angle bending.

Key observations from Table 10:

- **Homo_1 ($n-1 \rightarrow n$) based on $H_{\mathbf{CC}}$** is the only scheme that simultaneously satisfies all three physical criteria: intercept -0.018 (essentially zero), $c_\theta = 0.000905$ and $c_R = 10697$. It also covers the full set of 12 rings (C3–C14). This establishes it as the *physical gold standard*.
- **Fixed_m=3 (H)** shows deceptively excellent statistics (RMSE = 0.096, $R^2_{\text{exp}} = 0.997$). However, this is achieved by (i) excluding C3, the most strained ring, and (ii) severely distorting the physical coefficients: c_θ is inflated by a factor of ~ 7 (0.00775) and c_R is halved (5512) relative to the intrinsic deformation–energy relationship. This scheme is a statistical artifact and not physically meaningful.
- The comparison between G_{CC} and H_{CC} reveals that enthalpy-based strain energies are systematically more physically interpretable, as evidenced by near-zero intercepts and stable coefficients across multiple schemes.

S4.1.1 Leave-One-Out Cross-Validation of the Homodesmotic Scheme

To assess the predictive stability of the geometric model against overfitting, leave-one-out cross-validation (LOOCV) was performed on the homodesmotic scheme (Homo_1, $n-1 \rightarrow n$) using both the two-descriptor geometric model

($\Delta\theta^2 + \Delta R^2$) and the best three-descriptor extensions. The cross-validated coefficient Q^2 and the associated LOOCV RMSE are reported in Table 11.

The LOOCV analysis reveals that the two-descriptor geometric model has comparable or better predictive stability than any three-descriptor extension. The negative Q^2 values, while seemingly unfavorable, arise from the intrinsic statistical instability of leave-one-out validation with only 12 samples, particularly when the data contain non-linear electronic residuals that are not linearly separable by the descriptors. Notably, the addition of a third descriptor increases the LOOCV error without meaningfully improving R^2, confirming that the third descriptor is unnecessary and that the two-descriptor model is not overfitted.

S4.2 Cut-off Scanning: Locating the Valid Range of the Homodesmotic Reaction

To determine the maximum ring size up to which the homodesmotic reaction remains valid without introducing artificial corrections, we performed a cut-off scan. For a given cut-off point t, rings with $n \leq t$ were computed using the homodesmotic reaction ($m = n - 1$), while rings with $n > t$ were computed using the fixed reference reaction $m = t$. The scan covers $t = 5$ to 14, where $t = 14$ corresponds to pure homodesmotic treatment for all rings.

Key observations from Table 12:

The cut-off scan reveals three distinct regimes:

1. **Early collapse regime ($t = 5$–7)**: The intercept oscillates wildly ($+0.115 \rightarrow +0.764 \rightarrow -0.601$) as the fixed reference chain is applied to increasingly larger rings. Although $t = 5$ shows a low RMSE (0.404), its intercept ($+0.115$) already deviates from the physical zero, and the scheme is not continuous with the stable plateau; the subsequent dramatic jump at $t = 6$ confirms that the apparent accuracy of $t = 5$ is coincidental. This demonstrates that mixing the homodesmotic reaction with arbitrary fixed references outside the intrinsic validity window introduces uncontrolled systematic offsets.

2. **Golden plateau ($t = 8$–11)**: This is the most significant finding. For cut-off points between C8 and C11, the intercept is locked near zero (-0.086 to $+0.035$), R^2_{geom} remains stable at $\sim$0.96, and the geometric coefficients are highly consistent ($c_\theta \approx 0.00094$, $c_R \approx 10400$). Notably, the full-subset regression in this region is dominated exclusively by geometric descriptors ($\Delta\theta^2$, ΔR^2, Θ), with no need for charge or electronic terms. The RMSE vs. experiment reaches its minimum ($\sim$0.52) here, proving that for rings up to C11, the homodesmotic reaction captures the physical essence of ring strain with purely geometric descriptors.

3. **Tail collapse regime ($t = 12$–14)**: Applying the homodesmotic reaction to C12 and beyond causes R^2_{geom} to drop below 0.91 and RMSE to rise to $\sim$0.83. However, crucially, the intercept returns to near zero at $t =$

13 and $t = 14$ (-0.018), confirming that the deviation originates from conformational flexibility and entropic effects in large rings, not from a failure of the reaction scheme. The pure homodesmotic reaction ($t = 14$) is the only scheme that maintains a zero intercept across the entire C3–C14 range, establishing it as the *physical gold standard*, albeit with larger RMSE due to limitations of the current DFT conformational sampling for large rings.

S4.3 Summary: Practical Recommendation

Based on the combined evidence from Tables 10 and 12, we recommend:

- **For physical interpretation**: Use the pure homodesmotic reaction ($n - 1 \rightarrow n$) with H_{CC}-based composite energies. This scheme uniquely yields a near-zero intercept across all ring sizes and preserves the intrinsic deformation–energy relationship ($c_\theta \approx 0.001$, $c_R \approx 10000$).
- **For optimal prediction accuracy**: Apply the homodesmotic reaction to rings up to C11 (the golden plateau), where RMSE ≈ 0.52 kcal mol^{-1} is achievable with purely geometric descriptors. The residual errors for C12–C14 should be addressed through improved conformational sampling or inclusion of dispersion corrections, rather than by altering the reaction scheme.

Cartesian Coordinates

The Cartesian coordinates (in Å) of the optimized geometries for the representative (global minimum) conformers of each carbon ring discussed in the main text are provided in the accompanying file (`coordinates.txt`). Coordinates of additional low-weight conformers and all linear chain structures are not included here but are fully reproducible from the computational workflow described in Section S1 and are available upon reasonable request.

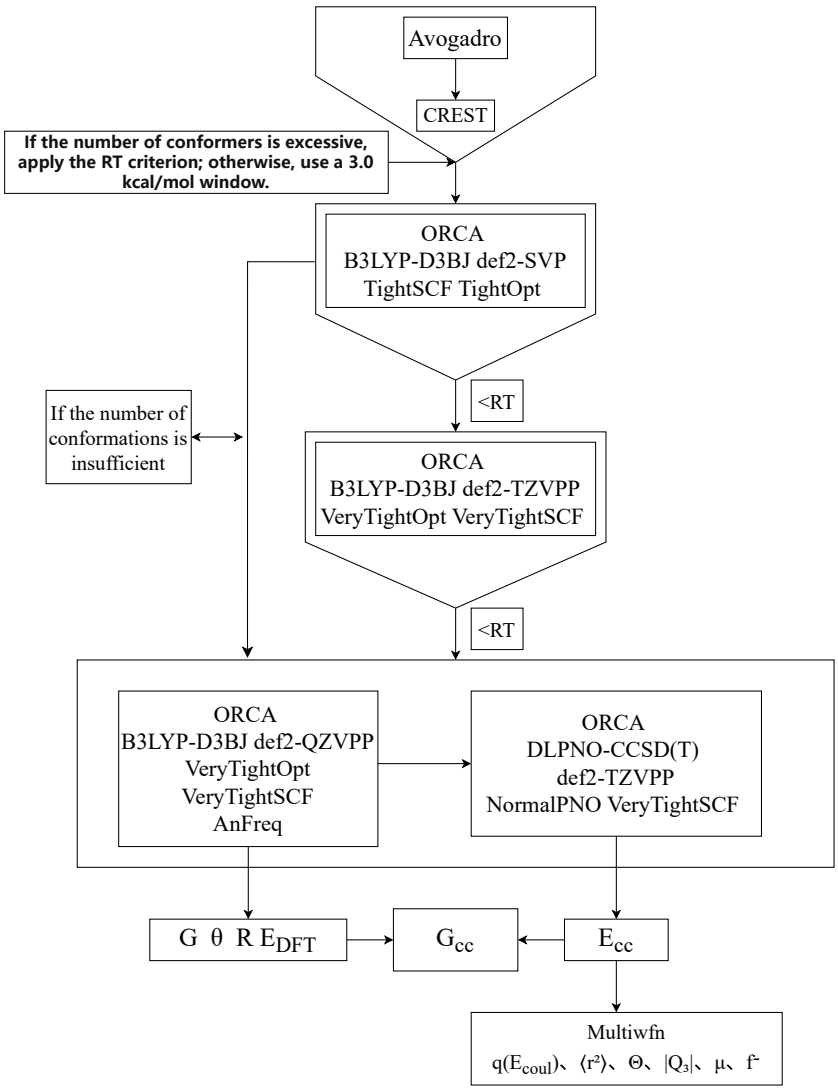

Figure 1: Overall computational workflow: CREST conformational search → multi-tier DFT refinement → composite energy scheme → QSAR and cut-off scan analysis, Gibbs free energy can be replaced by enthalpy.

Table 3: Composite energies and weights for linear alkanes (LC1–LC12), Part 1.

Molecule	Conformer	E_{CC} (Eh)	G_{DFT} (Eh)	E_{DFT} (Eh)	G_{CC} (Eh)	H_{CC} (Eh)
LC1	LC1_1	−40.438647	−40.460017	−40.508357	−40.39030626	−40.39030626
LC2	LC2_1	−79.675302	−79.733590	−79.810136	−79.59875620	−79.59875620
LC3	LC3_1	−118.915161	−119.041343	−119.119275	−118.83722932	−118.80673892
LC4	LC4_1	−158.155076	−158.322160	−158.426229	−158.05085597	−158.01681130
	LC4_2	−158.154210	−158.321055	−158.425306	−158.05085121	−158.01680654
	Weighted avg				−158.05085498	−158.01681031
LC5	LC5_1	−197.394909	−197.602883	−197.733195	−197.26446015	−197.22676007
	LC5_2	−197.394080	−197.602339	−197.732235	−197.26418427	−197.22648419
	Weighted avg				−197.26446015	−197.22676007
LC6	LC6_1	−236.634788	−236.884066	−237.040135	−236.47871907	−236.43728854
	LC6_2	−236.633987	−236.882245	−237.039217	−236.47773511	−236.43630458
	LC6_3	−236.633983	−236.882955	−237.039226	−236.47764631	−236.43621578
	Weighted avg				−236.47849283	−236.43706230
LC7	LC7_1	−275.874644	−276.164541	−276.345794	−275.69339215	−275.64873318
	LC7_2	−275.873909	−276.163368	−276.344834	−275.69286297	−275.64820400
	LC7_3	−275.873853	−276.163399	−276.344805	−275.69273004	−275.64807107
	Weighted avg				−275.69311931	−275.64846034
LC8	LC8_1	−315.114505	−315.444940	−315.654073	−314.90537183	−314.85753991
	LC8_2	−315.113759	−315.443892	−315.653158	−314.90481716	−314.85698524
	LC8_3	−315.113810	−315.443943	−315.653198	−314.90483523	−314.85700331
	LC8_4	−315.113715	−315.443910	−315.653160	−314.90478628	−314.85695436
	Weighted avg				−314.90501745	−314.85718553
LC9	LC9_1	−354.354343	−354.724696	−354.961052	−354.11798617	−354.06721577
	LC9_2	−354.353596	−354.724261	−354.960099	−354.11775825	−354.06698785
	LC9_3	−354.353682	−354.724385	−354.960186	−354.11788078	−354.06711038
	LC9_4	−354.353569	−354.724286	−354.960119	−354.11773563	−354.06696523
	Weighted avg				−354.11785795	−354.06708755
LC10	LC10_1	−393.594203	−394.005719	−394.268012	−393.33191001	−393.27779800
	LC10_2	−393.593524	−394.004883	−394.267142	−393.33136957	−393.27725756
	LC10_3	−393.593485	−394.004475	−394.267004	−393.33116283	−393.27705082
	LC10_4	−393.593537	−394.004723	−394.267116	−393.33134819	−393.27723618
	LC10_5	−393.593399	−394.004654	−394.267082	−393.33113066	−393.27701865
	Weighted avg				−393.33150834	−393.27739633
LC11	LC11_1	−432.834032	−433.285438	−433.574985	−432.54448470	−432.48735542
	LC11_2	−432.833399	−433.285150	−433.574098	−432.54445126	−432.48732198
	LC11_3	−432.833315	−433.285217	−433.574072	−432.54446037	−432.48733109
	LC11_4	−432.833394	−433.285138	−433.574070	−432.54446213	−432.48733285
	LC11_5	−432.833259	−433.285067	−433.574066	−432.54425927	−432.48712999
	Weighted avg				−432.54443403	−432.48730475
LC12	LC12_1	−472.073888	−472.566426	−472.881959	−471.75835520	−471.69802704
	LC12_2	−472.073242	−472.565341	−472.881023	−471.75740136	−471.69707320
	LC12_3	−472.073120	−472.565511	−472.880972	−471.75758912	−471.69726096
	LC12_4	−472.073215	−472.565330	−472.881023	−471.75745468	−471.69712652
	LC12_5	−472.073117	−472.565474	−472.880960	−471.75748873	−471.69716057
	LC12_6	−472.073075	−472.565239	−472.880931	−471.75734187	−471.69701371
	Weighted avg				−471.75790894	−471.69758078

Table 4: Composite energies and weights for linear alkanes (LC13–LC14), Part 2.

Molecule	Conformer	E_{CC} (Eh)	G_{DFT} (Eh)	E_{DFT} (Eh)	G_{CC} (Eh)	H_{CC} (Eh)
LC13	LC13_1	−511.313717	−511.846748	−512.186618	−510.97358461	−510.91016130
	LC13_2	−511.313100	−511.845897	−512.185730	−510.97324706	−510.90982375
	LC13_3	−511.313113	−511.845762	−512.185627	−510.97307087	−510.90964756
	LC13_4	−511.312995	−511.846015	−512.187997	−510.97543693	−510.91201362
	Weighted avg				−510.97358461	−510.91016130
LC14	LC14_1	−550.553569	−551.127061	−551.495883	−550.18448416	−550.11773842
	LC14_2	−550.552954	−551.126408	−551.494984	−550.18420578	−550.11746004
	LC14_3	−550.552975	−551.126127	−551.494932	−550.18401844	−550.11727270
	LC14_4	−550.552950	−551.126498	−551.494986	−550.18412490	−550.11737916
	LC14_5	−550.552813	−551.126760	−551.494940	−550.18400882	−550.11726308
	LC14_6	−550.552900	−551.126434	−551.495023	−550.18406176	−550.11731602
	LC14_7	−550.552887	−551.126302	−551.495005	−550.18414938	−550.11740364
	Weighted avg				−550.18448393	−550.11773819

Table 5: Descriptors for carbon rings (C3–C14).

Ring	n	θ (°)	R (Å)	Θ (a.u.)	μ (a.u.)	$\lvert Q_3\rvert$ (a.u.)	$\langle r^2\rangle$ (a.u.)	q_C (e)
C3	3	60.00	1.5038	0.7337	0.00015	11.8224	235.89	−0.2109
C4	4	88.53	1.5500	0.1244	0.00026	9.8343	562.52	−0.1695
C5	5	104.65	1.5411	0.4081	0.00910	6.4140	413.57	−0.1601
C6	6	111.43	1.5307	0.3886	0.00030	3.3819	944.40	−0.1544
C7	7	115.42	1.5337	0.3013	0.00436	2.5940	815.36	−0.1497
C8	8	117.85	1.5336	0.3334	0.01560	3.8650	1099.24	−0.1466
C9	9	115.38	1.5378	0.3516	0.00489	8.8575	1348.58	−0.1482
C10	10	116.77	1.5365	0.1920	0.00001	0.0008	1691.40	−0.1461
C11	11	115.38	1.5363	0.2093	0.00902	6.5907	2101.22	−0.1465
C12	12	114.13	1.5338	0.3284	0.00336	0.1898	2561.88	−0.1478
C13	13	114.08	1.5333	0.4620	0.02220	8.4397	3112.89	−0.1455
C14	14	114.35	1.5315	0.2866	0.00001	0.0018	3691.20	−0.1457

Table 6: Top regression models for the ring-opening reaction energy based on G_{CC}.

Rank	Descriptors	R^2	RMSE	max VIF	AIC	BIC
1	$\Delta\theta^2$, ΔR^2, $\langle r^2\rangle$	0.9822	0.3609	2.98	17.59	19.53
2	$\Delta\theta^2$, ΔR^2, $\lvert Q_3\rvert$	0.9819	0.3634	3.24	17.76	19.70
3	$\Delta\theta^2$, ΔR^2, μ	0.9818	0.3642	2.44	17.81	19.75
4	$\Delta\theta^2$, ΔR^2	0.9818	0.3643	2.37	15.82	17.28

Table 7: Top regression models for the ring-opening reaction energy based on H_{CC}.

Rank	Descriptors	R^2	RMSE	max VIF	AIC	BIC
1	$\Delta\theta^2$, ΔR^2, $\langle r^2 \rangle$	0.9860	0.3241	2.98	15.02	16.96
2	$\Delta\theta^2$, ΔR^2, $\lvert Q_3 \rvert$	0.9859	0.3246	3.24	15.05	16.99
3	$\Delta\theta^2$, ΔR^2, μ	0.9859	0.3249	2.44	15.07	17.01
4	$\Delta\theta^2$, ΔR^2	0.9859	0.3252	2.37	13.09	14.55
5	ΔR^2, q_C, $\langle r^2 \rangle$	0.9684	0.4867	3.90	24.77	26.71

Table 8: Comparison of the best models (descriptor set: $\Delta\theta^2$, ΔR^2, $\langle r^2 \rangle$) obtained using G_{CC} and H_{CC}.

Parameter	Based on G_{CC}	Based on H_{CC}
Intercept	0.717663	0.130981
$\Delta\theta^2$ coefficient	0.001571	0.001567
ΔR^2 coefficient	9877.732	10222.868
$\langle r^2 \rangle$ coefficient	-0.000058	-0.000030
R^2	0.9822	0.9860
RMSE (kcal mol^{-1})	0.3609	0.3241
max VIF	2.98	2.98
AIC	17.59	15.02
BIC	19.53	16.96

Table 9: Reference molecules for carbon rings.

Reference molecule	Bond type	Parameter	Value
Ethane (CH_3–CH_3)	C–C bond length	R_0	1.527 Å
Neopentane ($C(CH_3)_4$)	C–C–C bond angle	θ_0	109.47°

Table 10: Results for all reaction schemes using forced geometric regression ($\Delta\theta^2 + \Delta R^2$). Schemes are ranked by RMSE vs. experiment within each thermodynamic quantity. Units: MAE, RMSE in kcal mol^{-1}; c_θ in kcal mol^{-1} deg^{-2}; c_R in kcal mol^{-1} Å^{-2}.

Scheme	n	Based on G_{CC}				Based on H_{CC}			
		MAE	RMSE	R^2_{exp}	R^2_{geom}	MAE	RMSE	R^2_{exp}	R^2_{geom}
Fixed_m=3	11	1.0888	1.3295	0.4083	0.8480	0.0726	0.0959	0.9969	0.9847
Fixed_m=4	10	0.9251	1.0740	-3.8335	0.5811	0.1152	0.1294	0.9298	0.9021
Fixed_m=5	9	0.5223	0.6501	-0.7598	0.7402	0.1203	0.1387	0.9199	0.9104
Fixed_m=6	8	0.2747	0.3132	0.5029	0.8332	0.7998	0.8043	-2.2794	0.8781
Fixed_m=7	7	1.6939	1.7116	-12.0281	0.9670	0.8824	0.8870	-2.4988	0.9335
Fixed_m=8	6	1.0442	1.0663	-3.8797	0.9669	0.1630	0.1799	0.8611	0.9959
Fixed_m=9	5	0.4747	0.5088	-0.4360	0.9968	0.1219	0.1498	0.8755	0.9977
Fixed_m=10	4	0.8516	0.8583	-5.8032	0.9955	0.1230	0.1345	0.8329	0.9968
Homo_1 ($n-1 \to n$)	12	2.2613	3.7346	-0.8721	0.9827	0.5903	0.8270	0.9082	0.8978
Homo_2 ($n-2 \to n-1$)	12	2.5535	4.3458	-1.5350	0.8015	0.5490	0.7802	0.9183	0.9347
Fixed_m=1	12	4.5593	4.6762	-1.9351	0.8914	1.0560	1.0593	0.8494	0.9867
Fixed_m=2	12	14.2805	14.3182	-26.5176	0.8914	1.3492	1.3517	0.7547	0.9867

Table 11: LOOCV results for the homodesmotic scheme based on H_{CC}: two-descriptor vs. three-descriptor models.

Descriptor set	R^2	Q^2 (LOOCV)	$\Delta = R^2 - Q^2$	LOOCV RMSE
$\Delta\theta^2$, ΔR^2	0.898	-1.367	2.265	1.54
$\Delta\theta^2$, ΔR^2, Θ	0.898	-5.389	6.287	2.43
$\Delta\theta^2$, ΔR^2, q_C	0.901	-0.997	1.899	1.41
$\Delta\theta^2$, ΔR^2, μ	0.911	-2.698	3.609	2.00

RMSE in kcal mol^{-1} atom^{-1}. The negative Q^2 values arise from the small sample size ($N = 12$); however, the two-descriptor model gives the smallest absolute LOOCV error, confirming that the geometric model is not improved by adding a third descriptor.

Table 12: Cut-off scan results: C3–Ct using homodesmotic reaction, C($t+1$)–C14 using fixed reference $m = t$. All values based on H_{CC}. Units: MAE, RMSE in kcal mol^{-1}; coefficients c_θ in kcal mol^{-1} deg^{-2}, c_R in kcal mol^{-1} Å^{-2}.

t	Description	n	MAE	RMSE	R^2_{exp}	Intercept	R^2_{geom}	c_θ
5	C3–C5 homo, C6–C14 fixed $m = 5$	12	0.2203	0.4036	0.9781	0.1152	0.9752	0.000925
6	C3–C6 homo, C7–C14 fixed $m = 6$	12	0.6732	0.7622	0.9220	0.7640	0.9381	0.000953
7	C3–C7 homo, C8–C14 fixed $m = 7$	12	0.7242	0.8164	0.9105	-0.6011	0.9447	0.000934
8	C3–C8 homo, C9–C14 fixed $m = 8$	12	0.3599	0.5299	0.9623	-0.0864	0.9617	0.000936
9	C3–C9 homo, C10–C14 fixed $m = 9$	12	0.3453	0.5264	0.9628	0.0695	0.9608	0.000953
10	C3–C10 homo, C11–C14 fixed $m = 10$	12	0.3380	0.5233	0.9632	-0.0581	0.9607	0.000931
11	C3–C11 homo, C12–C14 fixed $m = 11$	12	0.3401	0.5250	0.9630	0.0351	0.9612	0.000957
12	C3–C12 homo, C13–C14 fixed $m = 12$	12	0.5812	0.8094	0.9121	0.4340	0.9137	0.001082
13	C3–C13 homo, C14 fixed $m = 13$	12	0.5903	0.8270	0.9082	-0.0182	0.8978	0.000905
14	Pure homodesmotic (all homo)	12	0.5903	0.8270	0.9082	-0.0182	0.8978	0.000905