Supporting Information for: A Unified Geometric Model of Ring Strain: From Carbon to Nitrogen and the Role of Lone-Pair Coulomb Repulsion

Yiming Ma

S1 Supplementary Computational Details

S1.1 CREST Conformational Search Parameters

All 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 10.0 kcal mol^{-1}.

S1.2 Boltzmann Weighting

For multiple conformers within $\Delta G \leq 0.59$ kcal mol^{-1} (298.15 K), Boltzmann-weighted averaging was used:

$$G_{\text{final}} = \sum_i w_i G_i, \quad w_i = \frac{e^{-G_i/kT}}{\sum_j e^{-G_j/kT}}, \quad kT = 0.000592 \text{ Hartree}$$

S1.3 Composite Energy Scheme

$G_{\text{final}} = E_{\text{DLPNO-CCSD(T)/def2-TZVPP}} + (G_{\text{B3LYP-D3(BJ)/def2-QZVPP}} - E_{\text{B3LYP-D3(BJ)/def2-QZVPP}})$

S1.4 Descriptor Definitions

See Table S1.

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_X	—	ADCH charge averaged over heavy atoms (e)
E_{coul}	$\frac{q_X^2 (n-3)}{2\bar{R}} \times 332.06$	Per-atom Coulomb repulsion (kcal mol^{-1})

S1.5 Linear Chains LN13 and LN14

For LN13 and LN14, CREST failed to converge. The two most stable LN12 conformers were used as templates to construct extended chains, followed by the same DFT refinement.

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

S2.1 Composite Gibbs Free Energies and Boltzmann Weights

Tables S2 and S3 list composite energies and weights for cycloalkanes and linear alkanes.

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

Molecule	Conformer	E_{CC} (Eh)	G_{DFT} (Eh)	E_{DFT} (Eh)	G_{CC} (Eh)	Weight
C3	C3	−117.671774	−117.817322	−117.874074	−117.615022	1.000
C4	C4	−156.914939	−157.098385	−157.182571	−156.830753	1.000
C5	C5	−196.186648	−196.409400	−196.520885	−196.075163	1.000
C6	C61	−235.436817	−235.697112	−235.837995	−235.295934	1.000
C7	C71	−274.666941	−274.968011	−275.135805	−274.499147	1.000
C8	C81	−313.901871	−314.221448	−314.417366	−313.705953	0.894
	C82	−313.898504	−314.241818	−314.435633	−313.704688	0.106
C9	C91	−353.138495	−353.517593	−353.741052	−352.915036	0.843
	C92	−353.134934	−353.515819	−353.738210	−352.912543	0.013
	C93	−353.136809	−353.517531	−353.740349	−352.913992	0.144
C10	C10	−392.378553	−392.797833	−393.047110	−392.129276	1.000
C11	C111	−431.619349	−432.079323	−432.356702	−431.341970	0.518
	C112	−431.619253	−432.079043	−432.356414	−431.341882	0.447
	C113	−431.617807	−432.077812	−432.355244	−431.340375	0.035
C12	C12	−470.867422	−471.366453	−471.671698	−470.562177	1.000
C13	C13	−510.107325	−510.646524	−510.975753	−509.778097	1.000
C14	C14	−549.352210	−549.931748	−550.288085	−548.995873	1.000

Table 3: Composite energies and weights for linear alkanes (LC3–LC14).

Molecule	Conformer	E_{CC} (Eh)	G_{DFT} (Eh)	E_{DFT} (Eh)	G_{CC} (Eh)	Weight
LC3	LC3	−118.915161	−119.041343	−119.119275	−118.837229	1.000
LC4	LC41	−158.155076	−158.322160	−158.426229	−158.051008	1.000
LC5	LC51	−197.394909	−197.602883	−197.733195	−197.264598	0.668
	LC52	−197.394080	−197.602339	−197.732235	−197.264184	0.332
LC6	LC61	−236.634788	−236.884066	−237.040135	−236.478719	1.000
LC7	LC71	−275.874644	−276.164541	−276.345794	−275.693392	1.000
LC8	LC81	−315.114505	−315.444940	−315.654073	−314.905372	1.000
LC9	LC91	−354.354343	−354.724696	−354.961052	−354.117986	0.315
	LC92	−354.353596	−354.724261	−354.960099	−354.117758	0.214
	LC93	−354.353682	−354.724385	−354.960186	−354.117881	0.264
	LC94	−354.353569	−354.724286	−354.960119	−354.117736	0.207
LC10	LC101	−393.594203	−394.005719	−394.268012	−393.331910	1.000
LC11	LC111	−432.834032	−433.285438	−433.574985	−432.544485	0.220
	LC112	−432.833399	−433.285150	−433.574098	−432.544451	0.208
	LC113	−432.833315	−433.285217	−433.574072	−432.544460	0.211
	LC114	−432.833394	−433.285138	−433.574070	−432.544462	0.211
	LC115	−432.833259	−433.285067	−433.574066	−432.544259	0.150
LC12	LC121	−472.073888	−472.566426	−472.881959	−471.758355	1.000
LC13	LC131	−511.313717	−511.846748	−512.186618	−510.973585	1.000
LC14	LC141	−550.553569	−551.127061	−551.495883	−550.184484	1.000

S2.2 Descriptors for Carbon Rings

Table S4 lists all descriptors for carbon rings.

Table 4: 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)	E_{strain}
C3	3	60.00	1.5038	0.7337	0.00015	11.8224	235.89	−0.2109	9.752
C4	4	88.53	1.5500	0.1244	0.00026	9.8343	562.52	−0.1695	6.984
C5	5	104.65	1.5411	0.4081	0.00910	6.4140	413.57	−0.1601	1.721
C6	6	111.43	1.5307	0.3886	0.00030	3.3819	944.40	−0.1544	0.729
C7	7	115.42	1.5337	0.3013	0.00436	2.5940	815.36	−0.1497	1.649
C8	8	117.85	1.5336	0.3334	0.01560	3.8650	1099.24	−0.1466	1.852
C9	9	115.38	1.5378	0.3516	0.00489	8.8575	1348.58	−0.1482	1.912
C10	10	116.77	1.5365	0.1920	0.00001	0.0008	1691.40	−0.1461	1.672
C11	11	115.38	1.5363	0.2093	0.00902	6.5907	2101.22	−0.1465	1.539
C12	12	114.13	1.5338	0.3284	0.00336	0.1898	2561.88	−0.1478	1.054
C13	13	114.08	1.5333	0.4620	0.02220	8.4397	3112.89	−0.1455	0.961
C14	14	114.35	1.5315	0.2866	0.00001	0.0018	3691.20	−0.1457	0.584

S3 Nitrogen Rings (N3–N14): Energies, Weights, Descriptors

S3.1 Composite Gibbs Free Energies and Boltzmann Weights

Tables S5 and S6 list energies and weights for cyclo-nitrogen rings and linear nitrogen chains.

Table 5: Composite energies and weights for cyclo-nitrogen rings (N3–N14).

Molecule	Conformer	E_{CC} (Eh)	G_{DFT} (Eh)	E_{DFT} (Eh)	G_{CC} (Eh)	Weight
N3	N31	−165.682574	−165.900014	−165.922901	−165.659686	1.000
N4	N41	−220.918378	−221.200798	−221.242089	−220.877087	1.000
N5	N51	−276.200791	−276.552540	−276.608930	−276.144401	1.000
N6	N61	−331.442486	−331.864613	−331.937470	−331.369630	1.000
N7	N73	−386.677874	−387.165483	−387.253766	−386.589591	1.000
N8	N81	−441.912981	−442.473480	−442.576525	−441.809936	1.000
N9	N91	−497.157204	−497.779238	−497.899046	−497.037396	1.000
N10	N101	−552.394217	−553.085775	−553.220274	−552.259718	1.000
N11	N113	−607.638265	−608.396772	−608.547592	−607.487486	1.000
N12	N121	−662.875100	−663.701030	−663.866786	−662.709358	1.000
N13	N131	−718.122484	−719.012769	−719.194939	−717.940304	1.000
N14	N142	−773.356651	−774.316694	−774.516962	−773.156369	1.000

Table 6: Composite energies and weights for linear nitrogen chains (LN3–LN14).

Molecule	Conformer	E_{CC} (Eh)	G_{DFT} (Eh)	E_{DFT} (Eh)	G_{CC} (Eh)	Weight
LN3	LN31	−166.945178	−167.148588	−167.193679	−166.900087	1.000
LN4	LN41	−222.186889	−222.457359	−222.517338	−222.126910	0.172
	LN42	−222.187348	−222.457754	−222.518181	−222.126921	0.175
	LN43	−222.187349	−222.457762	−222.518187	−222.126923	0.176
	LN44	−222.187812	−222.457933	−222.518234	−222.127511	0.477
LN5	LN51	−277.428474	−277.766376	−277.841620	−277.353229	0.128
	LN52	−277.429572	−277.766839	−277.842118	−277.354294	0.771
	LN53	−277.428644	−277.765912	−277.841463	−277.353092	0.101
LN6	LN61	−332.672306	−333.076304	−333.166958	−332.581652	1.000
LN7	LN71	−387.915273	−388.385847	−388.491998	−387.809123	1.000
LN8	LN81	−443.155060	−443.692678	−443.815178	−443.032560	0.527
	LN82	−443.155060	−443.692624	−443.815188	−443.032496	0.473
LN9	LN91	−498.401429	−499.004933	−499.142232	−498.264130	1.000
LN10	LN101	−553.644624	−554.314532	−554.467463	−553.491692	1.000
LN11	LN111	−608.887894	−609.624163	−609.792727	−608.719330	1.000
LN12	LN121	−664.131154	−664.933783	−665.119306	−663.946219	1.000
LN13	LN131	−719.374476	−720.243438	−720.443368	−719.174545	1.000
LN14	LN141	−774.615281	−775.551644	−775.767064	−774.400619	1.000

S3.2 Descriptors for Nitrogen Rings

Table 7: Descriptors for nitrogen rings (N3–N14).

Ring	n	θ (°)	R (Å)	Θ (a.u.)	μ (a.u.)	$\lvert Q_3 \rvert$ (a.u.)	$\langle r^2 \rangle$ (a.u.)	q_N (e)	E_{coul}	E_{strain}
N3	3	60.00	1.4575	0.6185	0.58108	17.4131	117.14	−0.3092	0.000	13.558
N4	4	89.77	1.4698	0.4475	0.00000	38.9610	201.74	−0.2724	8.385	11.683
N5	5	106.01	1.4450	0.6014	0.51893	29.8362	312.03	−0.2633	15.930	4.274
N6	6	115.65	1.4196	0.3686	0.00000	0.0004	456.58	−0.2465	21.415	3.809
N7	7	113.43	1.4238	0.0496	0.55752	62.7266	616.98	−0.2624	32.062	3.929
N8	8	118.76	1.4138	1.0032	0.47238	73.4818	856.06	−0.2524	37.322	3.687
N9	9	115.33	1.4283	0.5495	0.72361	94.6515	1013.52	−0.2444	41.726	3.566
N10	10	115.53	1.4223	0.2497	1.11130	108.1678	1300.64	−0.2474	49.771	3.539
N11	11	115.89	1.4201	1.0190	1.77461	97.7601	1556.98	−0.2355	51.812	3.212
N12	12	115.39	1.4231	0.3063	1.16671	127.7812	1856.76	−0.2380	59.315	3.174
N13	13	114.13	1.4250	0.6288	0.61095	52.8430	2319.75	−0.2438	69.188	2.831
N14	14	116.75	1.4230	0.8065	0.60317	108.1290	2470.09	−0.2273	66.013	3.043

S4 Full-Subset Regression Results

S4.1 Carbon Rings (C3–C14)

Table S8 lists top models satisfying $R^2 > 0.95$, RMSE < 0.5, VIF < 5.

Table 8: Regression models for carbon rings.

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

S4.2 Nitrogen Rings (N3–N14)

Table S9 lists top models satisfying $R^2 > 0.9$, RMSE < 1.0, VIF < 10.

Table 9: Regression models for nitrogen rings.

Rank	Descriptors	R^2	RMSE	max VIF	AIC	BIC
1	$\Delta\theta^2$, ΔR^2, E_{coul}	0.9683	0.6387	2.20	29.35	30.94
2	$\Delta\theta^2$, ΔR^2, $\langle r^2 \rangle$	0.9614	0.7042	2.01	31.50	33.09
3	ΔR^2, Θ, E_{coul}	0.9545	0.7642	1.76	33.30	34.89
4	ΔR^2, Θ, $\langle r^2 \rangle$	0.9381	0.8916	1.38	36.69	38.28
5	$\Delta\theta^2$, ΔR^2, $\lvert Q_3 \rvert$	0.9379	0.8932	2.06	36.73	38.32

S5 Detailed Analysis of the N5 Outlier

S5.1 CASSCF(6,5) Natural Orbital Occupations

To characterize the electronic origin of the N5 anomaly, CASSCF(6,5) calculations were performed on the N5H5 ring (6 electrons in 5 orbitals, corresponding to the five in-plane nitrogen lone pairs and their interactions with adjacent N–N σ^* orbitals). The natural orbital occupation numbers are listed in Table S10.

Table 10: CASSCF(6,5) natural orbital occupations for N5H5 (selected orbitals).

Orbital	Occupation	Ideal closed-shell	Deviation
HOMO (orbital 19)	1.34	2.00	-0.66
LUMO (orbital 20)	0.34	0.00	$+0.34$

The HOMO is significantly depleted (deficit -0.66) and the LUMO is occupied (0.34), indicating strong multireference character. For a typical closed-shell molecule, HOMO occupations are > 1.98 and LUMO occupations < 0.02. This confirms that single-reference methods (DFT, MP2) cannot adequately describe N5H5.

S5.2 Dynamic Correlation Energy

The dynamic correlation energy was estimated as the difference between DLPNO-CCSD(T) and CASSCF total energies at the same CASSCF(6,5)/def2-TZVP geometry:

This 14.85 kcal mol^{-1} gap represents the dynamic correlation energy arising from lone-pair delocalization, directly associated with σ-aromatic stabilization.

S5.3 NICS Details

NICS calculations were performed at the B3LYP-D3(BJ)/def2-QZVPP level using the GIAO method. The ghost atom coordinates were placed at the geometric center of the heavy atoms. For C5H10, NICS(0) = +5.86 ppm; for N5H5,

Table 11: Energy comparison at the CASSCF-optimized geometry (def2-TZVP).

Method	Total energy (Hartree)
CASSCF(6,5)	−276.159373
DLPNO-CCSD(T)	−276.183045
Difference	−0.023672 Hartree = 14.85 kcal mol^{-1}

NICS(0) = −15.18 ppm. The NICS(1) values (1 Å above the plane) are +3.70 and −5.65 ppm, respectively. The rapid attenuation in N5H5 (from −15.18 to −5.65 ppm, compared to benzene from −9.7 to −11.5 ppm) indicates that the ring current is confined to the molecular plane by lone-pair repulsion.

S5.4 Statistical Diagnosis of N5 as an Outlier

Table S11 summarizes the influence of N5 on the nitrogen regression model.

Table 12: Effect of including/excluding N5.

Dataset	R^2	RMSE	AIC
Including N5	0.899	1.096	44.25
Excluding N5	0.968	0.639	29.35

N5 has a studentized residual of −4.00 and Cook's distance > 1.0, confirming it as a strong outlier.

S5.5 Energy Decomposition

The additive geometric + Coulomb model predicts for N5: - $\Delta\theta^2$ = 19.15 - ΔR^2 = 0.00117 - E_{coul} = 15.93 kcal mol^{-1} - Predicted strain: 8.28 kcal mol^{-1} - Actual strain: 4.27 kcal mol^{-1} - Residual: −4.00 kcal mol^{-1}

This residual is attributed to σ-aromatic stabilization (∼14–18 kcal mol^{-1} from CASSCF/CCSD(T) and NICS) competing with lone-pair repulsion (15.93 kcal mol^{-1}).

S5.6 IGMH Method Details

S5.6 IGMH Method Details

IGMH (Independent Gradient Model based on Hirshfeld partition) analysis was performed using Multiwfn . Each N atom in N5H5 and each C atom in C5H10 was defined as an independent fragment. The δg_{inter} isosurface was plotted at isovalue 0.006 a.u. and colored by sign(λ_2)ρ ranging from −0.04 to +0.04 a.u. The resulting images (Figure 5 in the main text) visualize the competition

between stabilizing σ-delocalization (blue) and destabilizing lone-pair repulsion (red).

S6 Reference Molecules

Table S12 lists reference molecules used for unstrained geometric parameters.

Table 13: Reference molecules.

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°
Hydrazine (N_2H_4)	N–N bond length	R_0	1.479 Å
Triaminonitrogen ($N(NH_2)_3$)	N–N–N bond angle	θ_0	110.39°

All calculations used RI-J and COSX approximations, the def2/J auxiliary basis set, and Grimme's D3(BJ) dispersion correction.

S7 Cartesian Coordinates

S7 Cartesian Coordinates

The Cartesian coordinates (in Å) of the optimized geometries for the representative (global minimum) conformers of each carbon and nitrogen ring discussed in the main text are provided in the accompanying file (`coordinates.xyz`). 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.