Supporting Information: Permanent Electrostatic Moments Through the Lens of Atoms: Assessing Variational Hirshfeld Methods

Maximilian van Zyl,[†] Carlos Castillo-Orellana,[‡] Leila Pujal,[†] and Farnaz Heidar-Zadeh*,[†]

†*Department of Chemistry, Queen's University, 90 Bader Lane, Kingston, Ontario, K7L-3N6, Canada*

‡*Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, Concepción, Chile*

E-mail: farnaz.heidarzadeh@queensu.ca

Table of Contents

1 Datasets

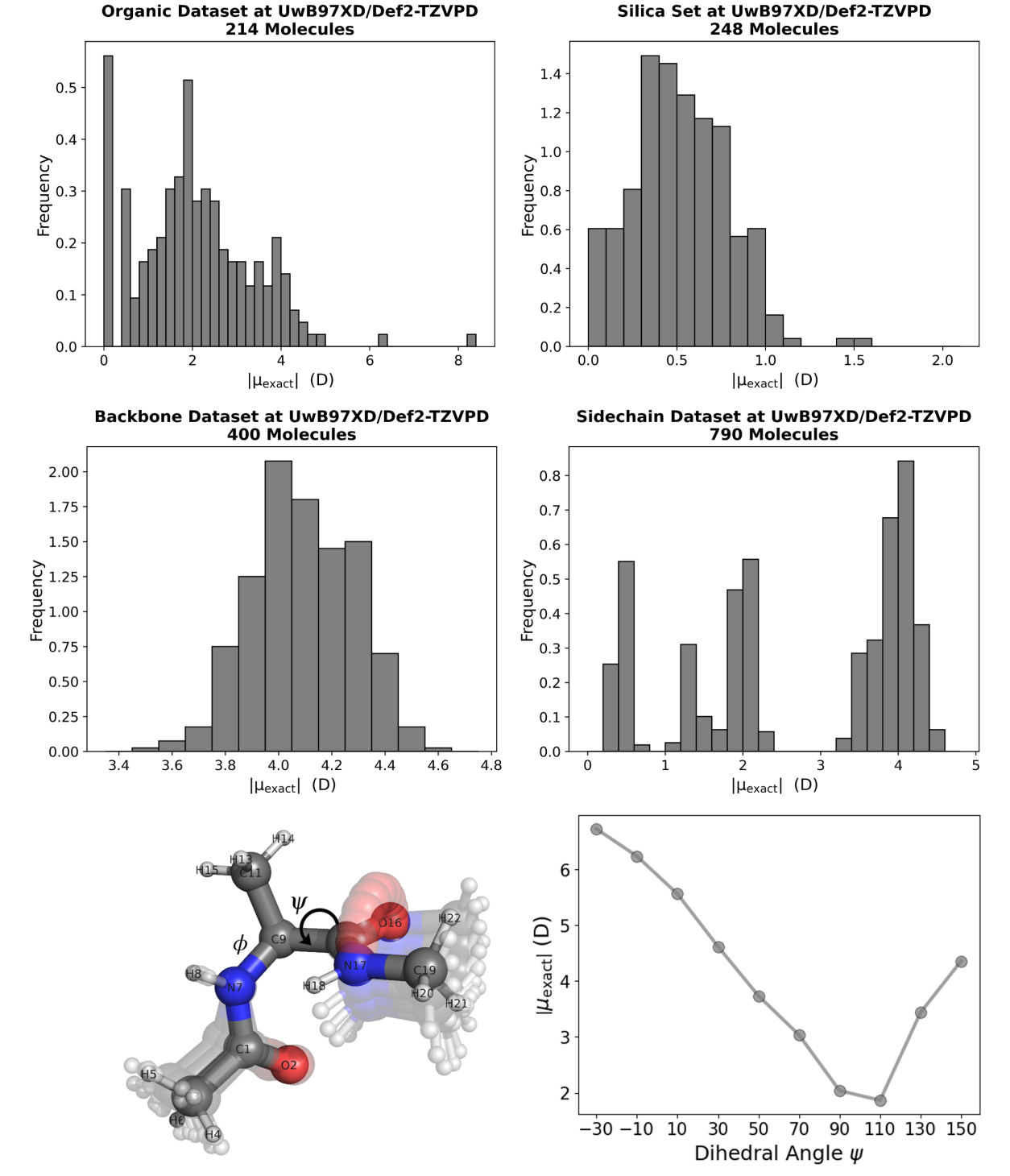

Figure S1: Histogram of molecular dipole norms (in Debye) computed at uωB97XD/def2-TZVPD level of theory for each dataset.

2 Sensitivity to Numerical Integration Grids

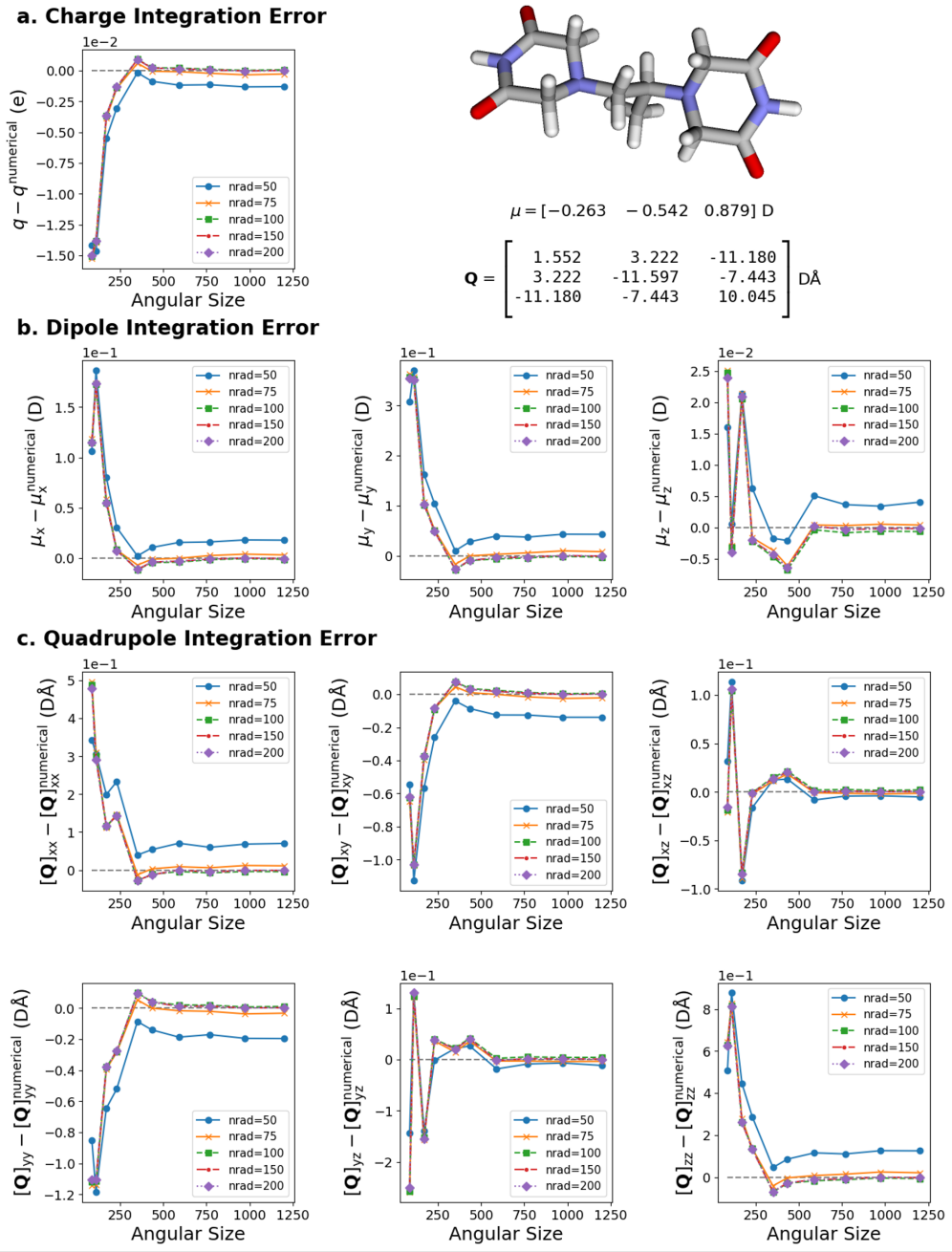

Figure S2: Numerical integration accuracy in reproducing each component of molecular multipole moments, against the exact moments from the ωB97X-D/Def2-TZVPD level of theory, for the largest molecule in the Organic dataset. Molecular integrations were performed with the exponential radial grid and the Becke–Lebedev angular grid of `HORTON 2`, using different numbers of radial (nrad label) and angular (x-axis) grid points.

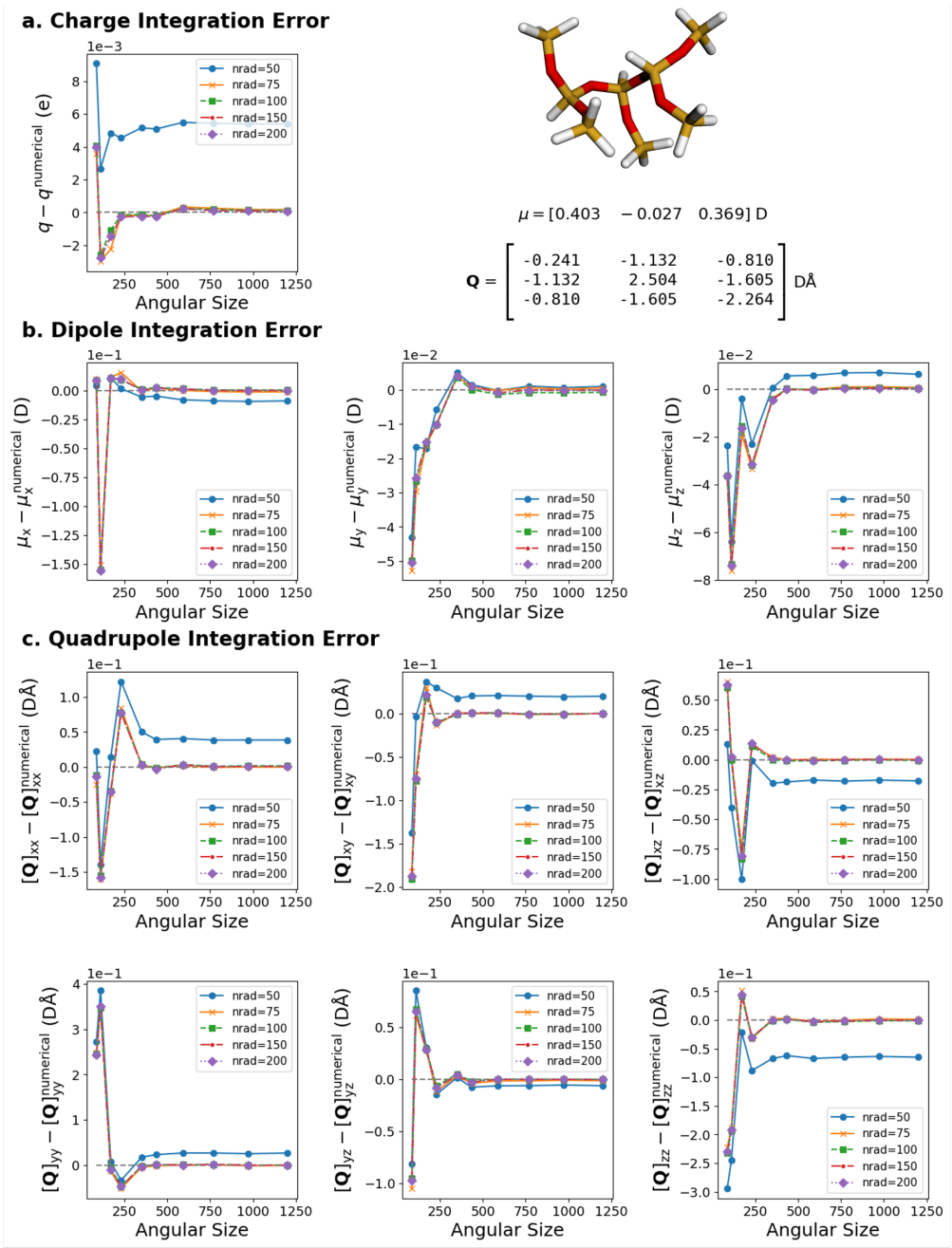

Figure S3: Numerical integration accuracy in reproducing each component of molecular multipole moments, against the exact moments from the ωB97X-D/Def2-TZVPD level of theory, for a molecule in the Silica dataset. Molecular integrations were performed with the exponential radial grid and the Becke–Lebedev angular grid of `HORTON 2`, using different numbers of radial (nrad label) and angular (x-axis) grid points.

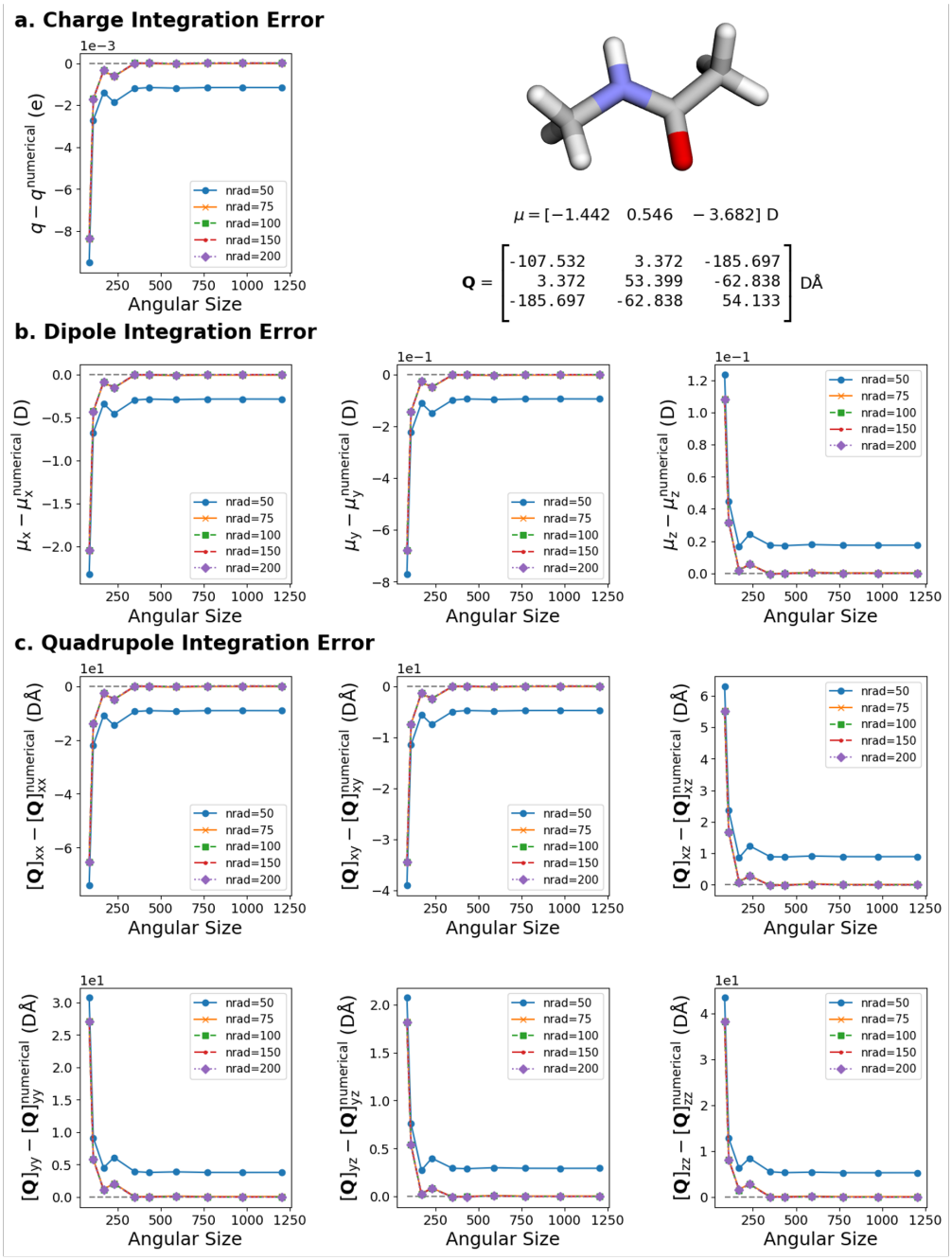

Figure S4: Numerical integration accuracy in reproducing each component of molecular multipole moments, against the exact moments from the ωB97X-D/Def2-TZVPD level of theory, for an N-Methylacetamide conformer in the Backbone dataset. Molecular integrations were performed with the exponential radial grid and the Becke–Lebedev angular grid of `HORTON 2`, using different numbers of radial (nrad label) and angular (x-axis) grid points.

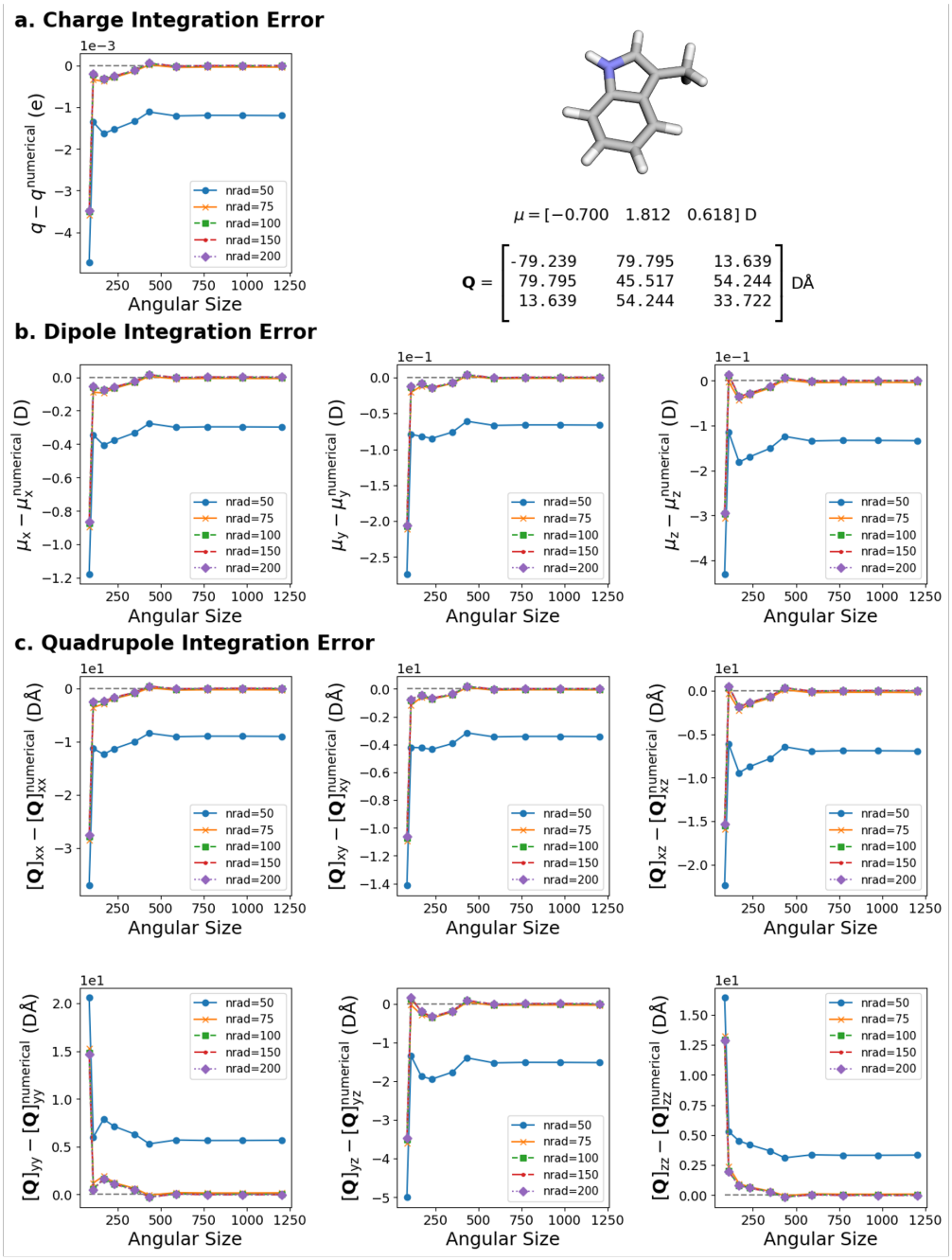

Figure S5: Numerical integration accuracy in reproducing each component of molecular multipole moments, against the exact moments from the ωB97X-D/Def2-TZVPD level of theory, for a TRP sidechain conformer in the Sidechain dataset. Molecular integrations were performed with the exponential radial grid and the Becke–Lebedev angular grid of `HORTON 2`, using different numbers of radial (nrad label) and angular (x-axis) grid points.

3 Approximating Molecular Dipole

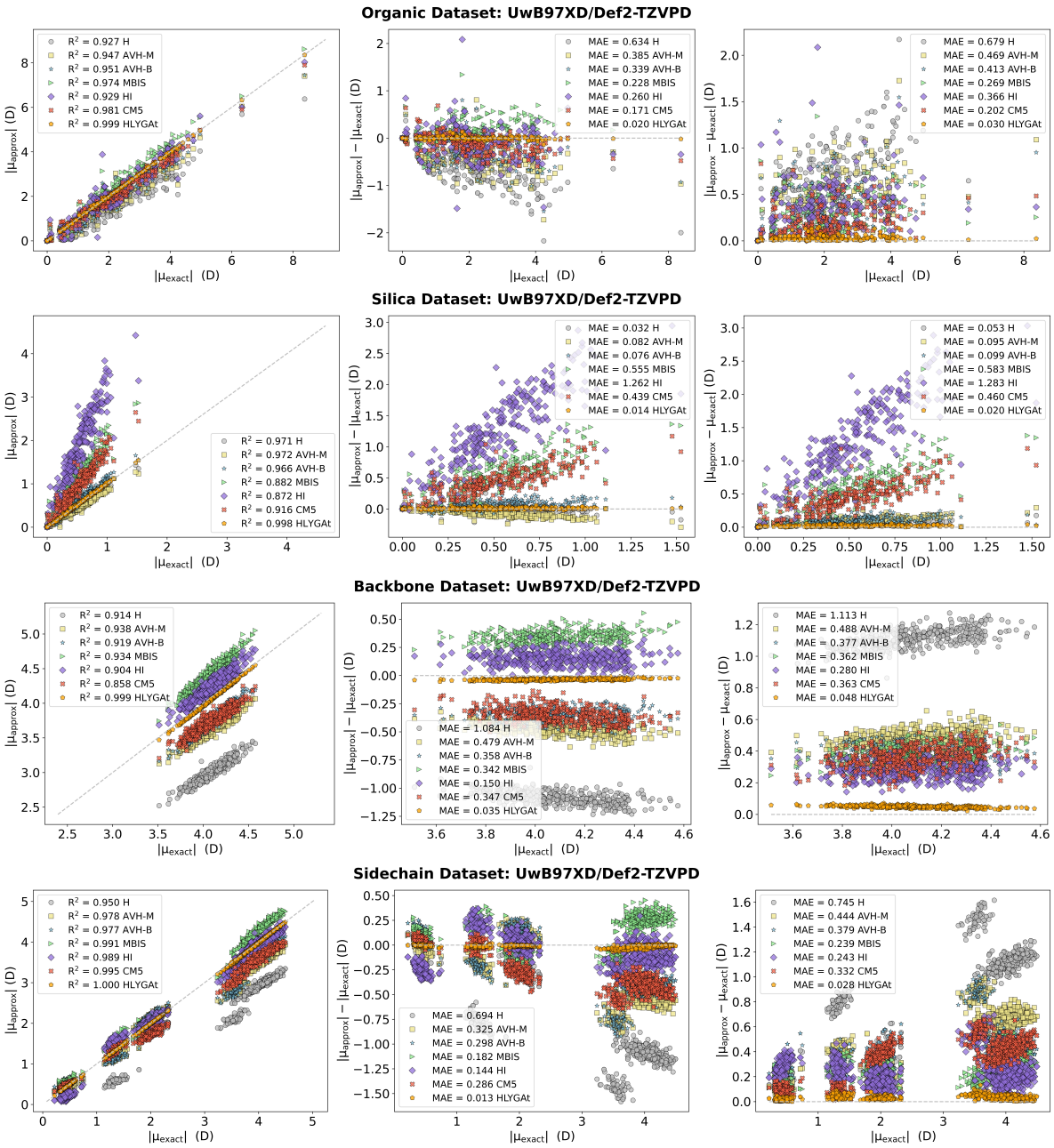

Figure S6: Accuracy of the point-charge approximation of the molecular dipole, $\mu_{\text{approx}} = \sum_A q_A \mathbf{R}_A$ for various charge schemes across different datasets at the uωB97XD/def2-TZVPD level of theory. The first column plots the norm of the approximate dipole versus the norm of the exact dipole, while the second and third columns display the error in the dipole norm and the error in the dipole vector, respectively. The partitioning schemes included are Hirshfeld (H), iterative Hirshfeld (HI), Minimal Basis Iterative Stockholder (MBIS), and the Additive Variational Hirshfeld (AVH) with bound proatoms (AVH-B) and minimal proatoms (AVH-M), Charge Model 5 (CM5), and electrostatic potential-fitted charges based on Hu, Lu, and Yang's procedure (HLYGAt).

3.1 Computational Robustness of Approximated Molecular Dipole

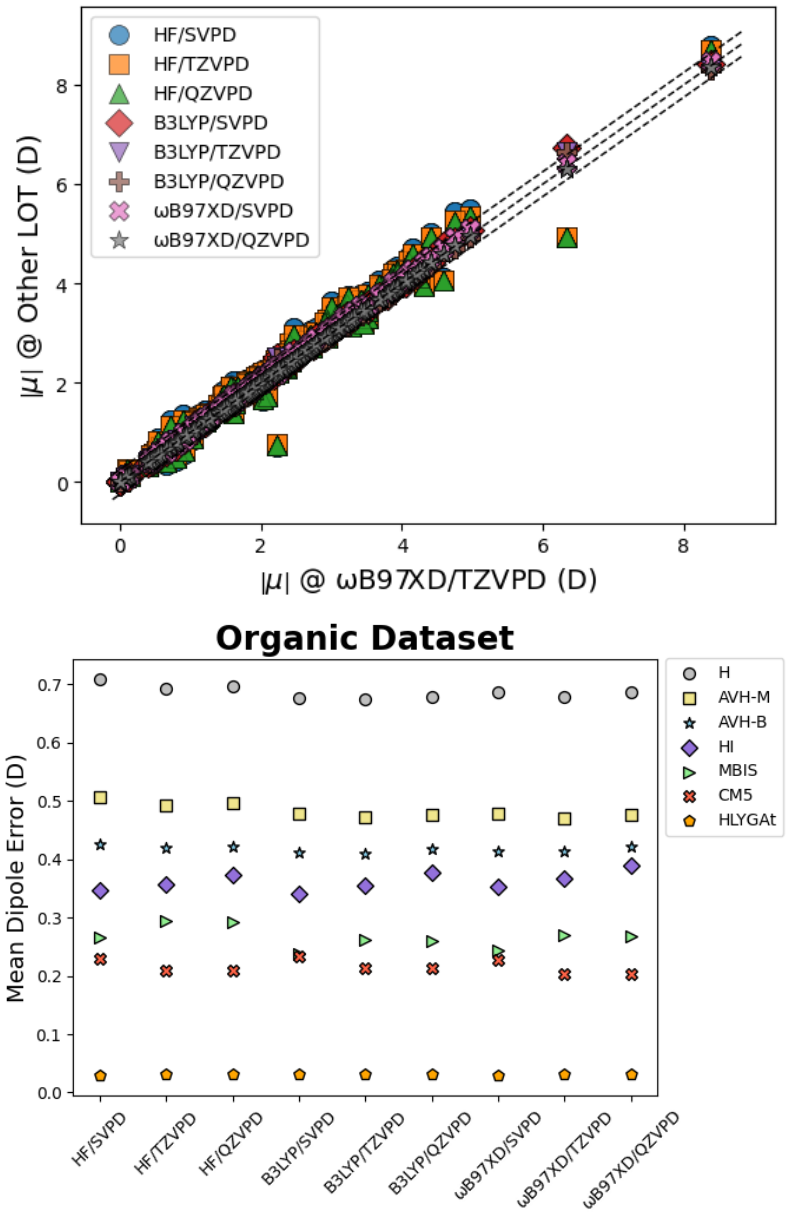

Figure S7: Robustness of molecular dipole (top plot) and its point-charge approximation error (bottom plot) defined in Equation (9), with respect to the level of theory for the Organic dataset (in Debye). The basis set corresponds to the Def2 (Karlsruhe) basis sets. For each level of theory, the error is calculated as the norm of the difference vector $|\mu_{\text{exact}} - \mu_{\text{approx}}|$ for each molecule, and then averaged over all molecules in each dataset. The values corresponding to ωB97X-D/Def2-TZVPD level of theory are presented in Table 1. For each molecule, the same geometry is used for all levels of theory (i.e., single-point calculations were performed).

4 Approximating Molecular Quadrupole

Note that the `Symmetry=None` keyword was used in the Gaussian input files, which sets the origin for molecular quadrupole moment calculations at $(0,0,0)$. Since this origin is not aligned with the molecular center of mass, the magnitude of the atomic coordinates $\{\mathbf{R}_A\}$ directly influences the calculated molecular quadrupole moments, and consequently, the associated errors reported when approximating molecular quadrupole with only atomic charges. This effect is particularly relevant for the Backbone and Sidechain datasets, as their geometries were extracted from molecular dynamics trajectories of proteins, resulting in relatively large atomic coordinate components.

Table S1: Mean quadrupole error (in Debye $\cdot$ Å) for approximating the molecular quadrupole using atomic charges ($\mathbf{Q}^{(0)}_{\text{approx}}$), plus atomic dipole ($\mathbf{Q}^{(0)}_{\text{approx}} + \mathbf{Q}^{(1)}_{\text{approx}}$), plus atomic quadrupole ($\mathbf{Q}^{(0)}_{\text{approx}} + \mathbf{Q}^{(1)}_{\text{approx}} + \mathbf{Q}^{(2)}_{\text{approx}}$) as defined in Equation (10), computed at the ωB97X-D/Def2-TZVPD level of theory. For each level of approximation, the $|\mathbf{Q}_{\text{exact}} - \mathbf{Q}_{\text{approx}}|$ error defined in Equation (11) is computed for each molecule, and then averaged over all molecules in each dataset. CM5 and HLYGAt are purely charge-based models, so higher-order approximations are not applicable for these methods. For brevity, the subscript approx is omitted in the column labels, with each column indicating only the additional atomic quadrupole term included in the approximation. The $+\mathbf{Q}^{(1)}$ column corresponds to results in Table 2, and Figure 4 illustrates their error distribution for each scheme across the different datasets. Figure S8 shows the numerical error distribution of $+\mathbf{Q}^{(2)}$ for each scheme and dataset.

Schemes	**Organic**			**Silica**			**Backbone**			**Sidechain**		
	$\mathbf{Q}^{(0)}$	$+\mathbf{Q}^{(1)}$	$+\mathbf{Q}^{(2)}$	$\mathbf{Q}^{(0)}$	$+\mathbf{Q}^{(1)}$	$+\mathbf{Q}^{(2)}$	$\mathbf{Q}^{(0)}$	$+\mathbf{Q}^{(1)}$	$+\mathbf{Q}^{(2)}$	$\mathbf{Q}^{(0)}$	$+\mathbf{Q}^{(1)}$	$+\mathbf{Q}^{(2)}$
H	3.247	0.983	0.000	0.668	0.573	0.002	81.377	0.774	0.040	64.527	0.934	0.048
AVH-M	2.180	1.001	0.000	0.841	0.468	0.002	35.269	0.795	0.040	38.595	1.044	0.048
AVH-B	2.720	0.667	0.000	1.068	0.326	0.002	27.433	0.236	0.041	32.588	0.382	0.048
HI	1.492	1.407	0.000	10.223	3.040	0.007	20.282	1.296	0.040	20.881	2.032	0.048
MBIS	1.211	0.879	0.003	4.389	0.398	0.046	26.645	2.100	1.983	20.391	2.992	2.705
CM5	1.670	-	-	3.565	-	-	26.645	-	-	28.595	-	-
HLYGAt	0.467	-	-	0.257	-	-	3.618	-	-	2.500	-	-

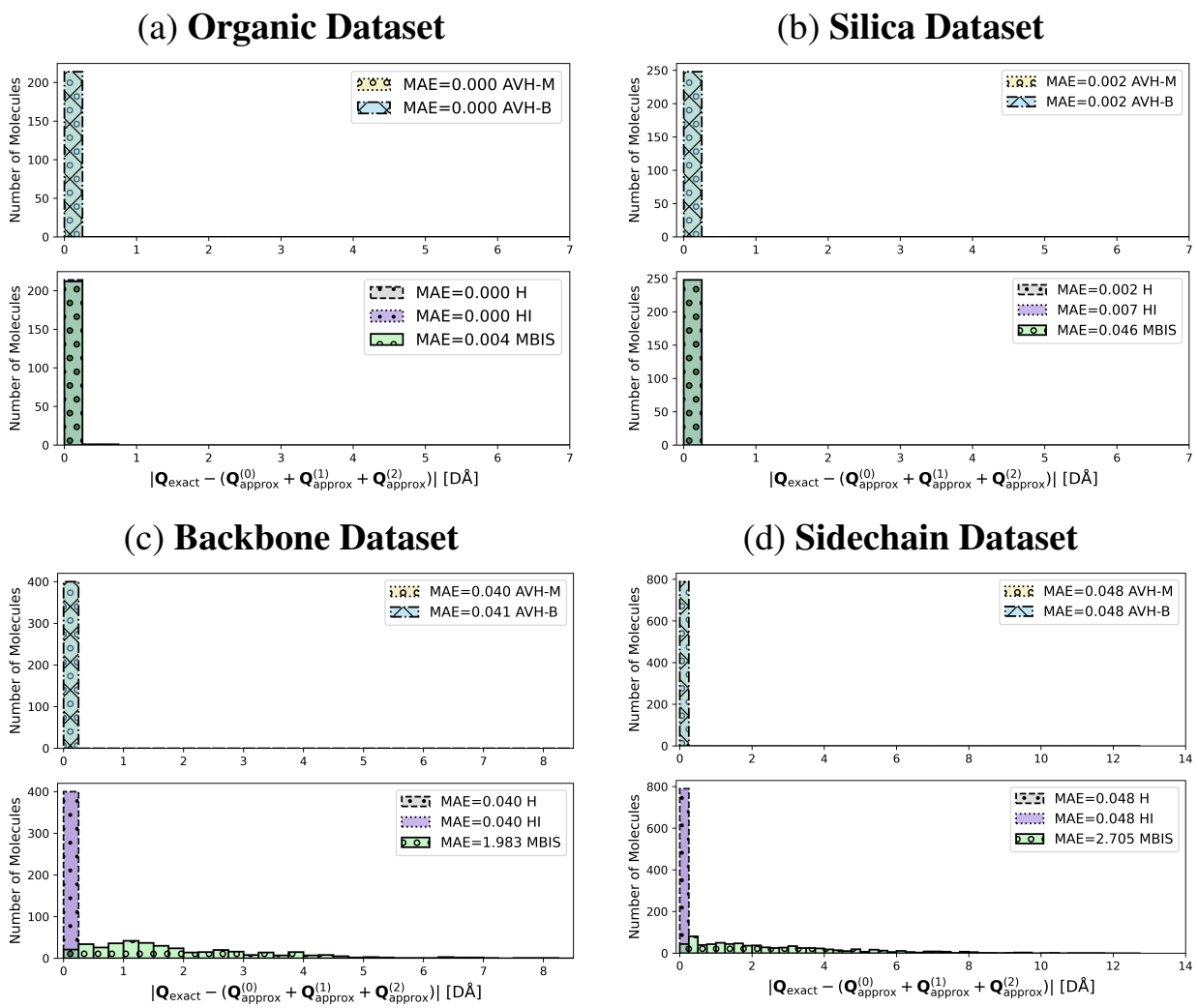

Figure S8: Distribution of errors in approximating the molecular quadrupole moment (in Debye $\cdot$ Å) using atomic charges and dipoles, i.e., $\mathbf{Q}^{(0)}_{\text{approx}} + \mathbf{Q}^{(1)}_{\text{approx}} + \mathbf{Q}^{(2)}_{\text{approx}}$, as defined in Equation (10), computed at the ωB97X-D/Def2-TZVPD level of theory. For each molecule, the error is calculated as the Frobenius norm of the full difference matrix $|\mathbf{Q}_{\text{exact}} - \mathbf{Q}_{\text{approx}}|$ defined in Equation (11). The mean absolute error (MAE) values indicated in the legend match the corresponding column in Table S1.

Table S2: The percentage of molecules in each dataset for which adding the higher-order atomic moments degrades approximating the molecular quadrupole moment at the ωB97X-D/Def2-TZVPD level of theory. $\%P_1$ denotes the percentage of molecules in each dataset for which $|\mathbf{Q}_{\text{exact}} - (\mathbf{Q}^{(0)}_{\text{approx}} + \mathbf{Q}^{(1)}_{\text{approx}})| - |\mathbf{Q}_{\text{exact}} - (\mathbf{Q}^{(0)}_{\text{approx}}| > \varepsilon$ and $\%P_2$ denotes the percentage of molecules in each dataset for which $|\mathbf{Q}_{\text{exact}} - (\mathbf{Q}^{(0)}_{\text{approx}} + \mathbf{Q}^{(1)}_{\text{approx}})| - |\mathbf{Q}_{\text{exact}} - (\mathbf{Q}^{(0)}_{\text{approx}} + \mathbf{Q}^{(1)}_{\text{approx}} + \mathbf{Q}^{(2)}_{\text{approx}})| > \varepsilon$. The norms are defined in Equation (11). The value of $\varepsilon = 0.05$ to make sure we are not comapring numerical error.

Schemes	**Organic**		**Silica**		**Backbone**		**Sidechain**	
	$\%P_1$	$\%P_2$	$\%P_1$	$\%P_2$	$\%P_1$	$\%P_2$	$\%P_1$	$\%P_2$
H	4.7	0.0	36.3	0.0	0.0	0.0	0.0	0.0
AVH-M	5.1	0.0	15.3	0.0	0.0	0.0	0.0	0.0
AVH-B	4.7	0.0	2.0	0.0	0.0	0.0	0.0	0.1
HI	43.9	0.0	0.4	0.0	0.0	0.0	0.6	0.0
MBIS	29.4	0.0	0.0	0.0	0.0	30.0	0.5	31.4

4.1 Computational Robustness of Approximated Molecular Quadrupole

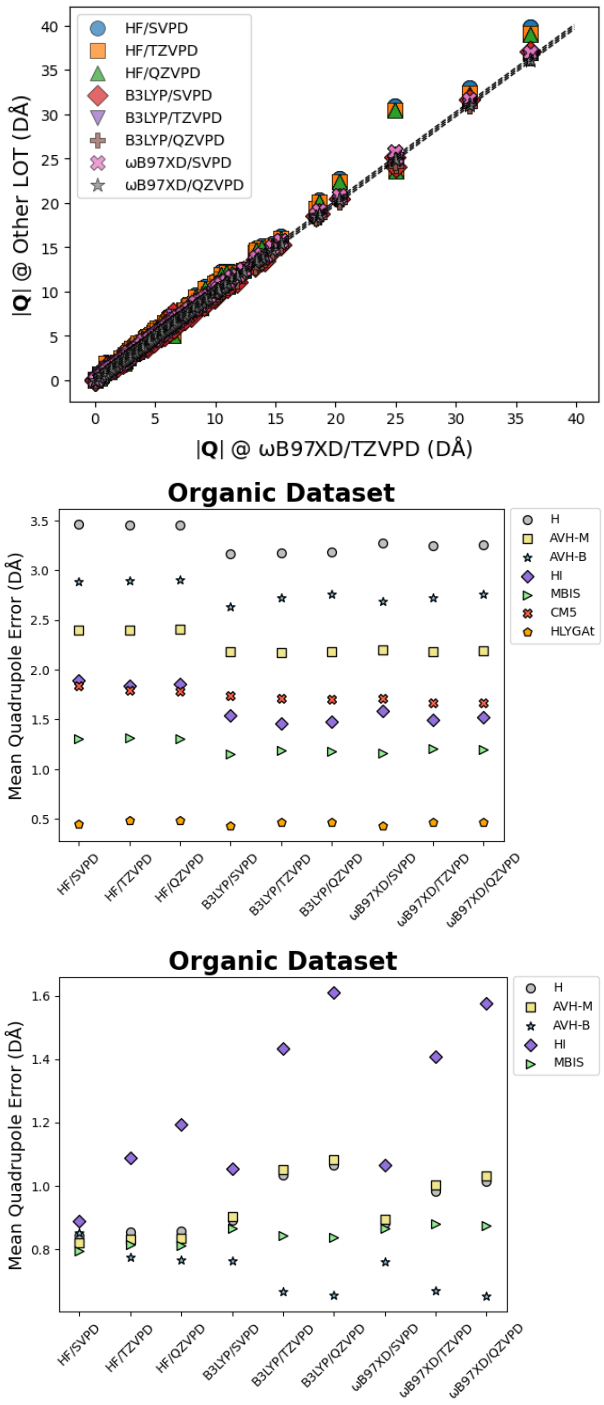

Figure S9: Robustness of mean quadrupole error (in DÅ) for approximating the molecular quadrupole with respect to the level of theory for the Organic dataset. The basis set corresponds to the Def2 (Karlsruhe) basis sets. The error is calculated as the Frobenius norm of the difference matrix $|\mathbf{Q}_{\text{exact}} - \mathbf{Q}_{\text{approx}}|$ for each molecule, and then averaged over all molecules in each dataset. The values corresponding to ωB97X-D/Def2-TZVPD level of theory are presented in Table 2. For each molecule, the same geometry is used for all levels of theory (i.e., single-point calculations were performed).

5 Conformational Stability

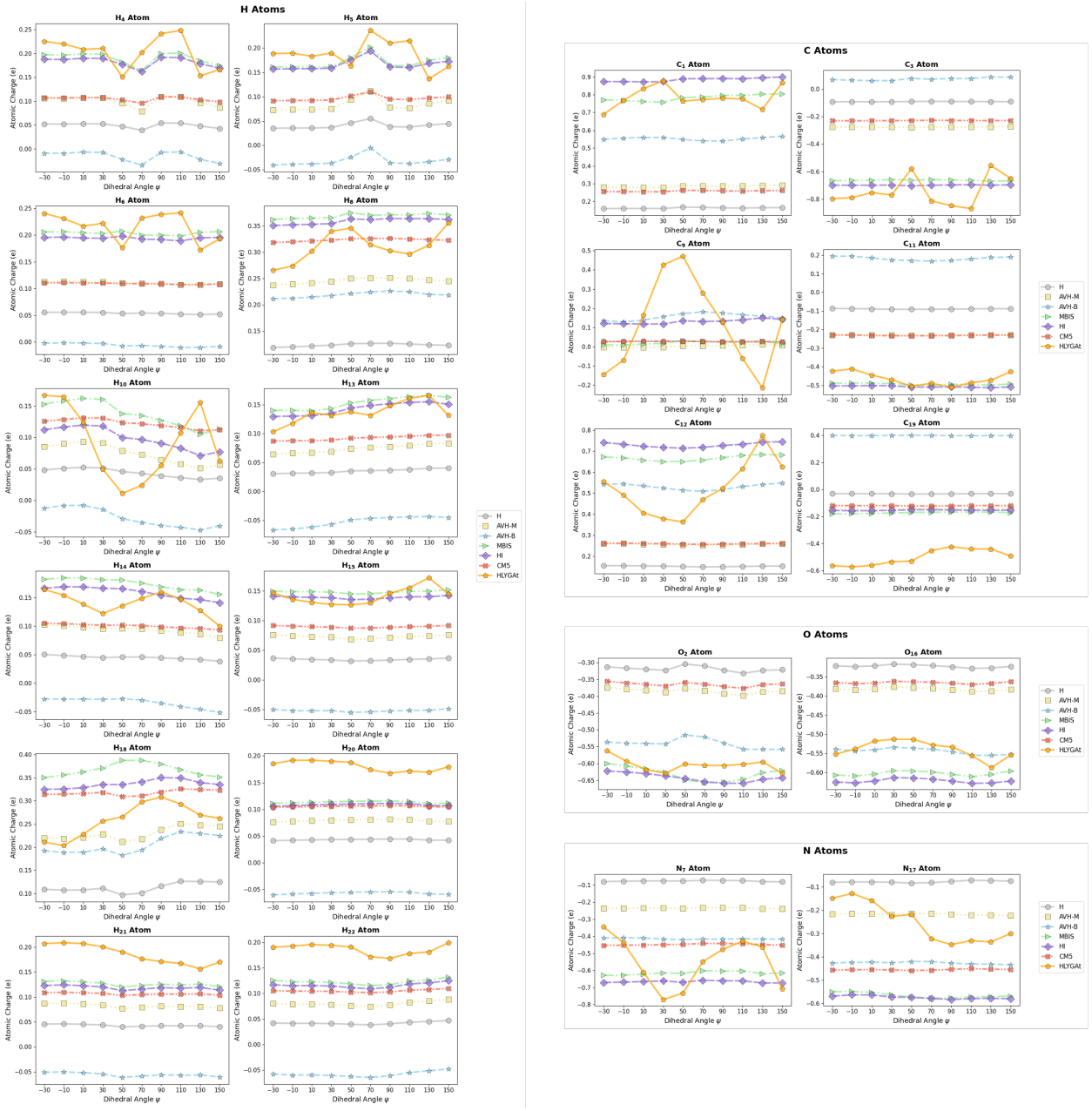

Figure S10: Variation in atomic charges for atoms in alanine dipeptide conformers across different partitioning schemes as a function of the dihedral angle. Atom numbering corresponds to that in Figure S1. Atomic charges were obtained at the ωB97X-D/Def2-TZVPD level of theory. The partitioning schemes included are Hirshfeld (H), iterative Hirshfeld (HI), Minimal Basis Iterative Stockholder (MBIS), and the Additive Variational Hirshfeld (AVH) with bound proatoms (AVH-B) and minimal proatoms (AVH-M).

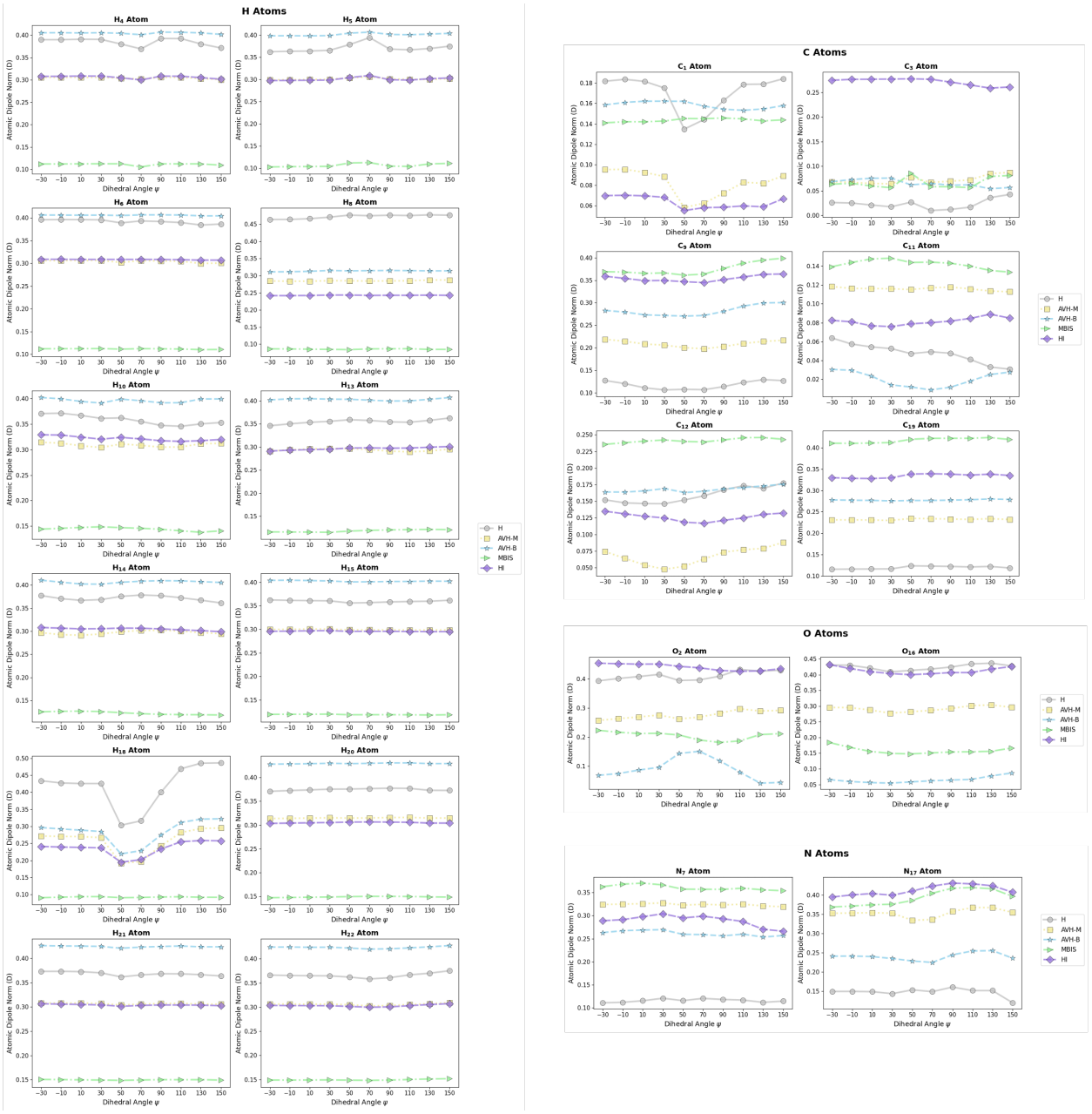

Figure S11: Variation in atomic dipole norms for atoms in alanine dipeptide conformers across different partitioning schemes as a function of the dihedral angle. Atom numbering corresponds to that in Figure S1. Atomic dipoles were obtained at the ωB97X-D/Def2-TZVPD level of theory. The partitioning schemes included are Hirshfeld (H), iterative Hirshfeld (HI), Minimal Basis Iterative Stockholder (MBIS), and the Additive Variational Hirshfeld (AVH) with bound proatoms (AVH-B) and minimal proatoms (AVH-M).

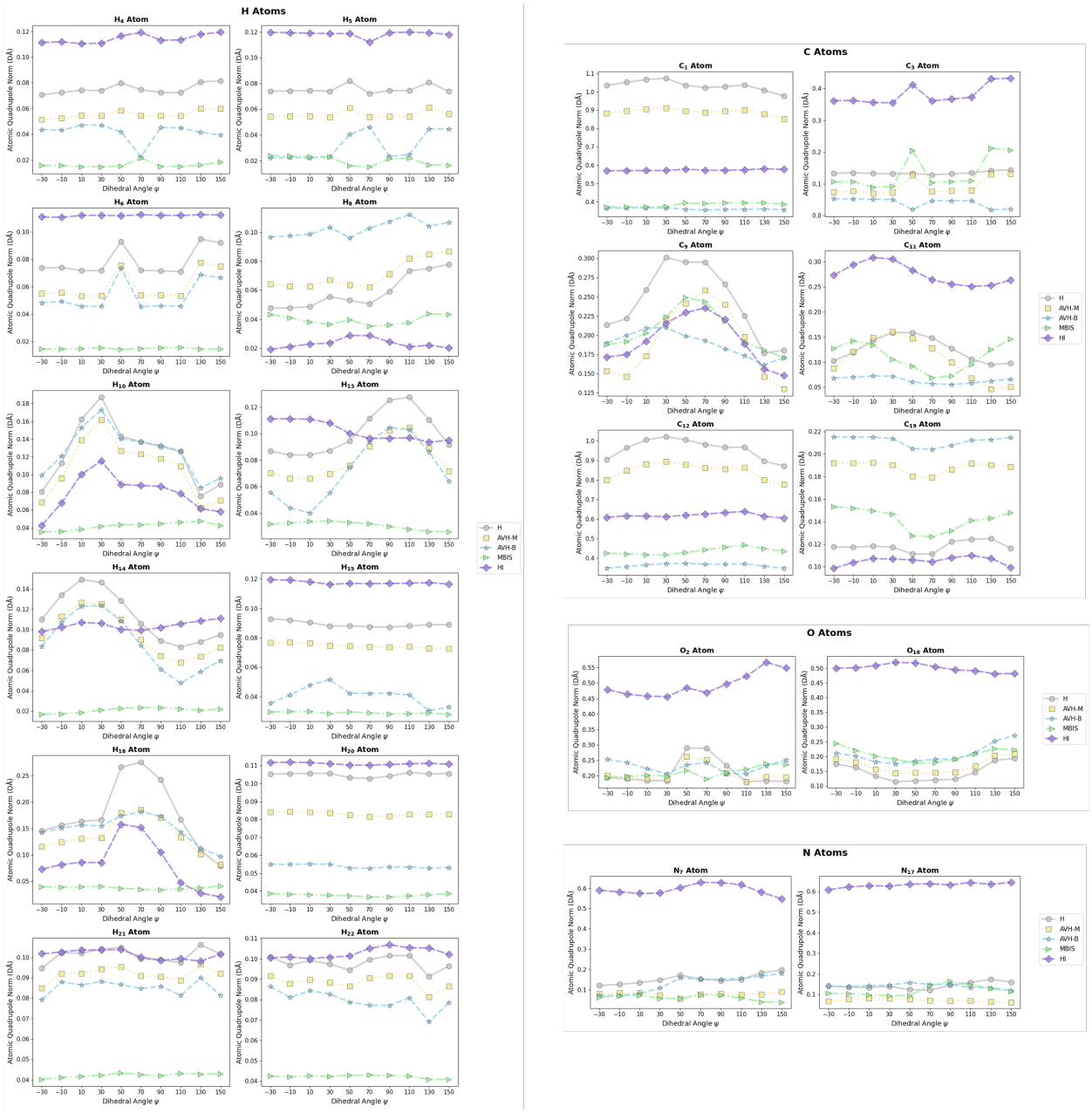

Figure S12: Variation in atomic quadrupole norms for atoms in alanine dipeptide conformers across different partitioning schemes as a function of the dihedral angle. Atom numbering corresponds to that in Figure S1. Atomic quadrupoles were obtained at the ωB97X-D/Def2-TZVPD level of theory. The partitioning schemes included are Hirshfeld (H), iterative Hirshfeld (HI), Minimal Basis Iterative Stockholder (MBIS), and the Additive Variational Hirshfeld (AVH) with bound proatoms (AVH-B) and minimal proatoms (AVH-M).