Electronic Supplementary Material (ESI) for Chemical Science.

Supporting Information: Predicting Phosphorescence Energies and Inferring Wavefunction Localization with Machine Learning

Andrew E. Sifain,[ß1,2] Levi Lystrom,[ß1,2] Richard A. Messerly,[1] Justin S. Smith,[1,2] Benjamin Nebgen,[1,3] Kipton Barros,[1,2] Sergei Tretiak,[1,2,3] Nicholas Lubbers,*[4] and Brendan J. Gifford*[1,2,3]

[1]Theoretical Division, [2]Center for Nonlinear Studies, [3]Center for Integrated Nanotechnologies, and [4]Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA, 87545

Email: nlubbers@lanl.gov, giff@lanl.gov

[ß]These authors contributed equally to this work

Neural Network Architecture and Training Procedure

The neural network architecture used in our work is HIPNN[1] with tensor sensitivities.[2] In particular, the network has 2 interaction blocks, each consisting of 1 interaction layer, followed by 3 on-site layers, and a linear layer to form a hierarchical contribution to target energy. In the HIP-loc model, localization weights are determined using linear layers in the same way as how atomic energy contributions are constructed. The target energy at each linear layer is computed as a sum over all atomic contributions.

The training procedure closely follows that described in Ref. 1. In addition to training the model to molecular energies, we also trained to atomic forces, as this has been shown to improve energy predictions.[3,4] The cost function reflects this additional information and consists of the root-mean-square error (RMSE) and mean-absolute error (MAE) of the singlet (S_0) and triplet (T_1) state energies as well as the RMSEs and MAEs of the atomic forces:

$$\mathcal{L} = \mathcal{L}_{L2} + \mathcal{L}_R + \sum_{i=S_0,T_1} \lambda_E \left(\sqrt{\langle (E_i - E_i')^2 \rangle} + \langle |E_i - E_i'| \rangle \right) + \lambda_{\vec{F}} \left(\sqrt{\frac{1}{3} \langle \left(\vec{F}_i - \vec{F}_i' \right)^2 \rangle} + \frac{1}{3} \langle \left| \vec{F}_i - \vec{F}_i' \right| \rangle \right) \quad (S1)$$

where angle brackets $\langle \cdots \rangle$ denote an average over the training batch, which was taken to be 512 structures. Primed versus unprimed represents ML versus reference quantum mechanical quantities and the scaling factors were set to $\lambda_E = 1/10$ and $\lambda_{\vec{F}} = 1/30$. We

also included $\mathcal{L}_{L2}$ regularization ($\lambda_{L2} = 10^{-6}$) and $\mathcal{L}_R$ hierarchicality regularization ($\lambda_R = 10^{-1}$) to prevent overfitting.[1] The factor of $1/3$ is a normalization that reflects the three degrees of freedom in an atomic force.

The dataset used to build and test the HIPNN and HIP-loc models was split using an 80-10-10 scheme, where 80% of the dataset was used for training, 10% for validation, and 10% for testing. Additional training parameters include minimum and maximum number of epochs of 10 and 500, $t_{\text{patience}} = 10$ epochs, and initial learning rate of $\eta = 10^{-3}$. See Ref. 1 for more details of the training procedure including the gradient-based optimization with the Adam optimizer[5] and annealing/early stopping algorithm.

The cost function shown above was used to train both the HIPNN and HIP-loc models described in the main text. The only difference between the models is that HIPNN trains directly to the S_0 and T_1 state energies and atomic forces, whereas for HIP-loc, the triplet state energy is represented as the energy of the singlet plus the singlet-triplet energy gap, $E_T = E_S + \Delta E$, where now ΔE is computed as a weighted sum over atomic contributions. This feature of HIP-loc encompasses the novelty of the approach because unlike total energy, which is an extensive property that increases with system size, the energy gap is a non-extensive quantity that does not vary with molecule size in a clear way (Fig. S2). Furthermore, the singlet-triplet transition may be attributed to only a handful of atoms or localized regions of the molecule which, by way of the HIP-loc localization weights, can be accounted for by weighting atomic contributions disproportionately.

Localization Metric

To quantify the agreement in localization computed using HIP-loc and DFT, we introduce a localization metric, η, defined as the distance between the centers of localization computed using HIP-loc weights ($\vec{r}_{\text{HIP-loc}}$) and DFT spin density ($\vec{r}_{\text{DFT}}$) divided by the relative size of the molecule (R):

$$\eta = \frac{|\vec{r}_{\text{HIP-loc}} - \vec{r}_{\text{DFT}}|}{R} \qquad (S2)$$

The center of localization is defined as

$$\vec{r} = \frac{\sum_{i=1}^{N} \alpha_i \vec{r}_i}{\sum_i^N \alpha_i} \qquad (S3)$$

where the coefficient α_i is the "density" contribution on the i-th atom and $\vec{r}_i$ is the position of the atom. For computing $\vec{r}_{\text{DFT}}$, these coefficients are taken to be atom-centered

densities (q_i) approximated from Hirshfeld charge partitioning,[6] while for $\vec{r}_{\mathrm{HIP-loc}}$, they are the localization weights (w_i). Finally, we quantify the relative size of the molecule (R) with the radius of gyration (R_g), assuming atoms are of equal mass,

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\vec{r}_i - \vec{r}_{cm})^2} \qquad (S4)$$

where the center of mass or, more accurately, the center of atomic distribution is defined as

$$\vec{r}_{cm} = \frac{1}{N}\sum_{i=1}^{N}\vec{r}_i \qquad (S5)$$

For $\eta \ll 1$, the centers of localizations are in close proximity to one another, whereas $\eta \sim 1$ or greater signifies that the centers of localization differ by approximately the radius of the molecule and therefore there is very little or no agreement in the predicted localization.

References

(1) Lubbers, N.; Smith, J. S.; Barros, K. Hierarchical Modeling of Molecular Energies Using a Deep Neural Network. *J. Chem. Phys.* **2018**, *148* (24), 241715. https://doi.org/10.1063/1.5011181.

(2) Smith, J. S.; Chigaev, M.; Barros, K.; Lubbers, N. Incorporating Angular and Higher Order Tensor Information in Atomistic Neural Networks. *Manuscript in Preparation*

(3) Cooper, A. M.; Kästner, J.; Urban, A.; Artrith, N. Efficient Training of ANN Potentials by Including Atomic Forces via Taylor Expansion and Application to Water and a Transition-Metal Oxide. *Npj Comput. Mater.* **2020**, *6* (1), 54. https://doi.org/10.1038/s41524-020-0323-8.

(4) Smith, J. S.; Lubbers, N.; Thompson, A. P.; Barros, K. Simple and Efficient Algorithms for Training Machine Learning Potentials to Force Data. *ArXiv200605475 Cond-Mat Physicsphysics Stat* **2020**.

(5) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* **2017**.

(6) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44* (2), 129–138. https://doi.org/10.1007/BF00549096.

GDB	# of Compounds
2	2
3	6
4	24
5	108
6	727
7	4701
8	3625
9	2241
10	1770
11	900
12	449

Table S1: Number of molecules sampled from each GDB dataset. Here, GDB# means the sampled compounds are made up of # heavy atoms.

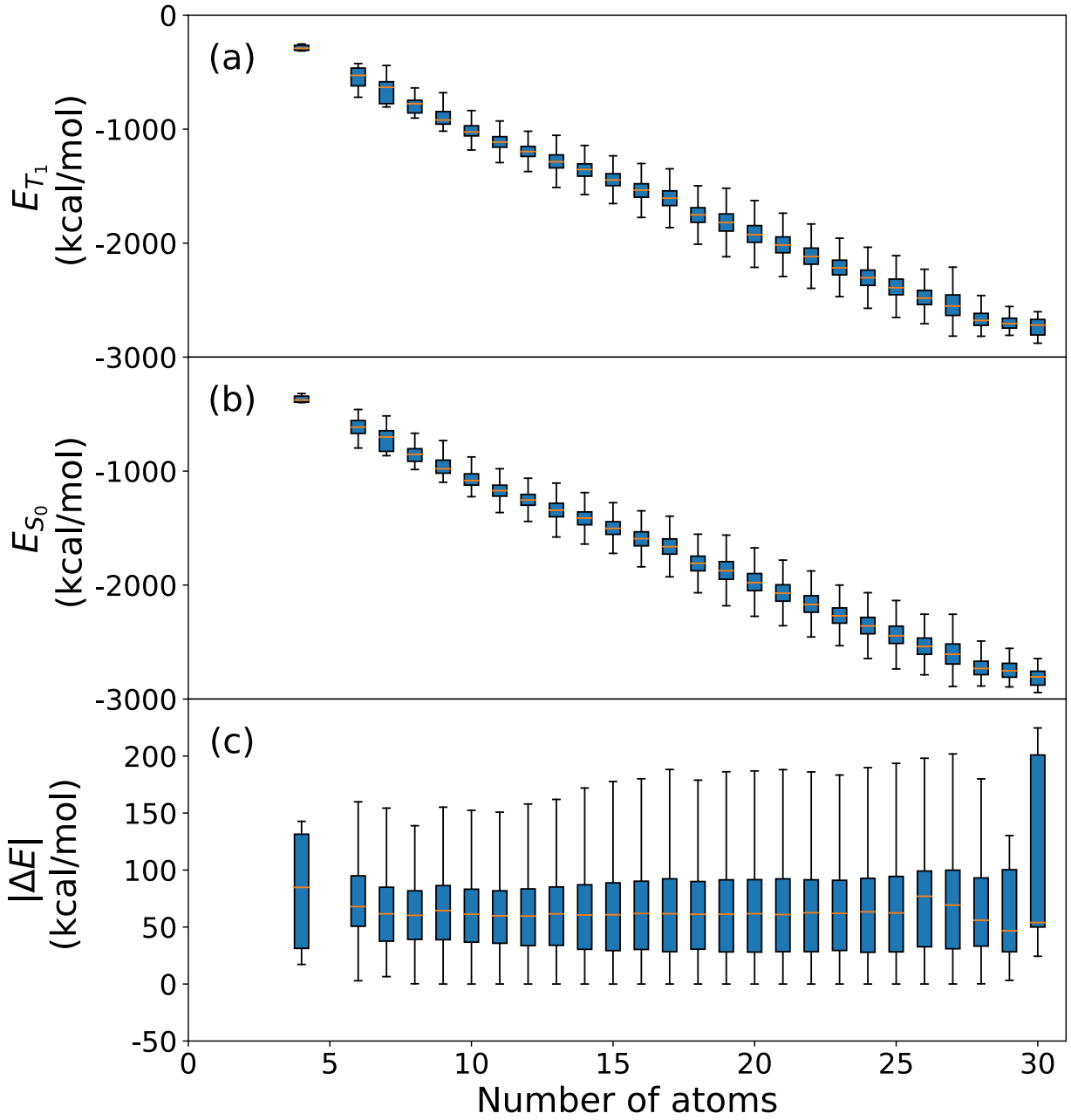

Figure S1: Energies as a function of number of atoms (including H) for all molecules in the dataset. (a) Triplet (E_{T1}), (b) singlet (E_{S0}), and (c) absolute $|\Delta E|$ energies are shown. Each box plot shows the median (orange line), first (Q_1) to third (Q_3) quartiles (blue box), and whiskers (black lines), capped at $Q_{1/3} \mp 1.5(Q_3 - Q_1)$. Unlike E_{T1} and E_{S0} that decrease with molecule size, ΔE does not. The extensivity assumption of HIPNN does not apply for ΔE, prompting development of HIP-loc.

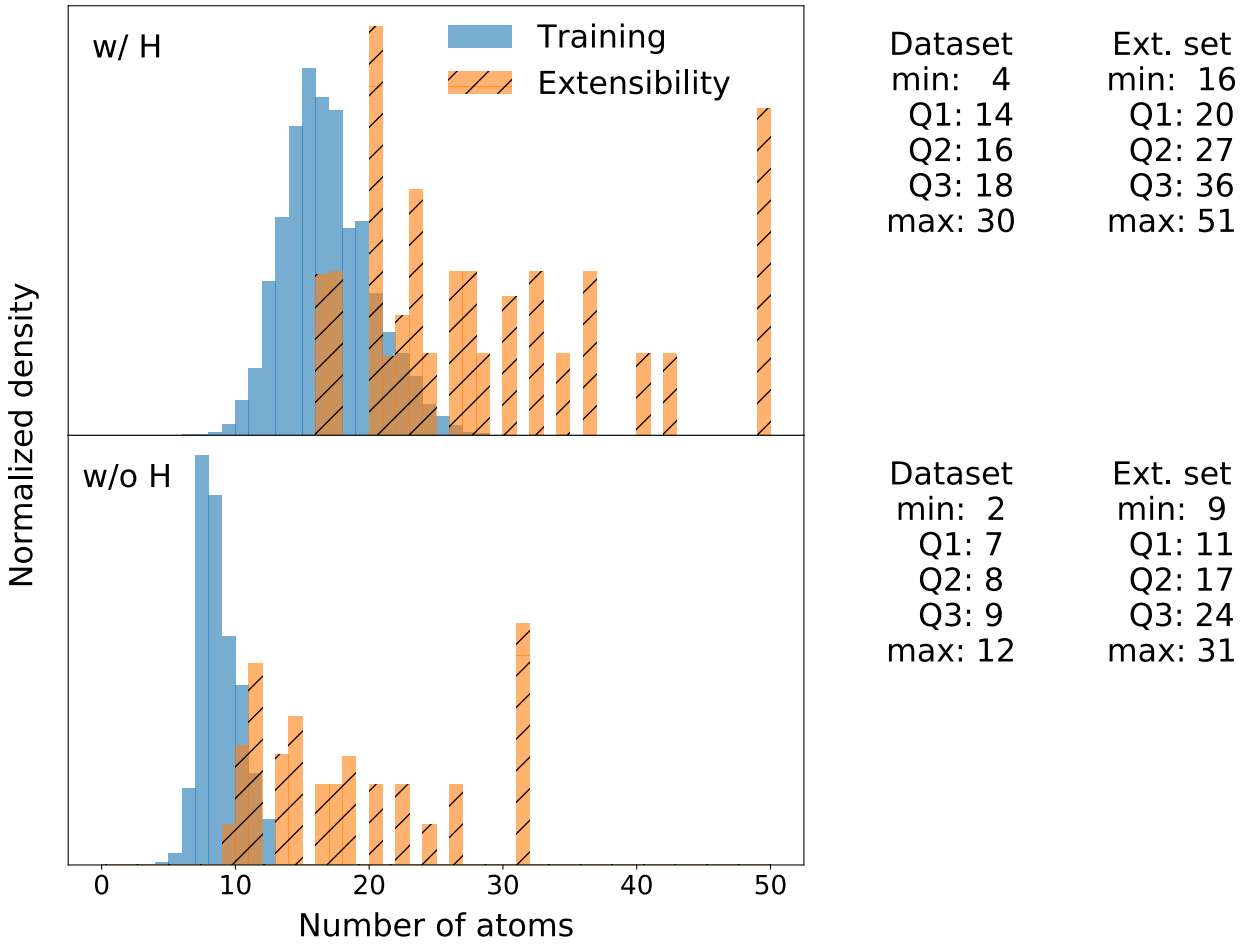

Figure S2: Distributions of molecule size for the dataset used to train the ML models as well as the extensibility test set with (top panel) and without (bottom panel) hydrogen included. Statistical values of the distributions including the minimum value, 25th percentile (Q_1), 50th percentile (Q_2), 75th percentile (Q_3), and maximum value are shown to the right of the corresponding panel.

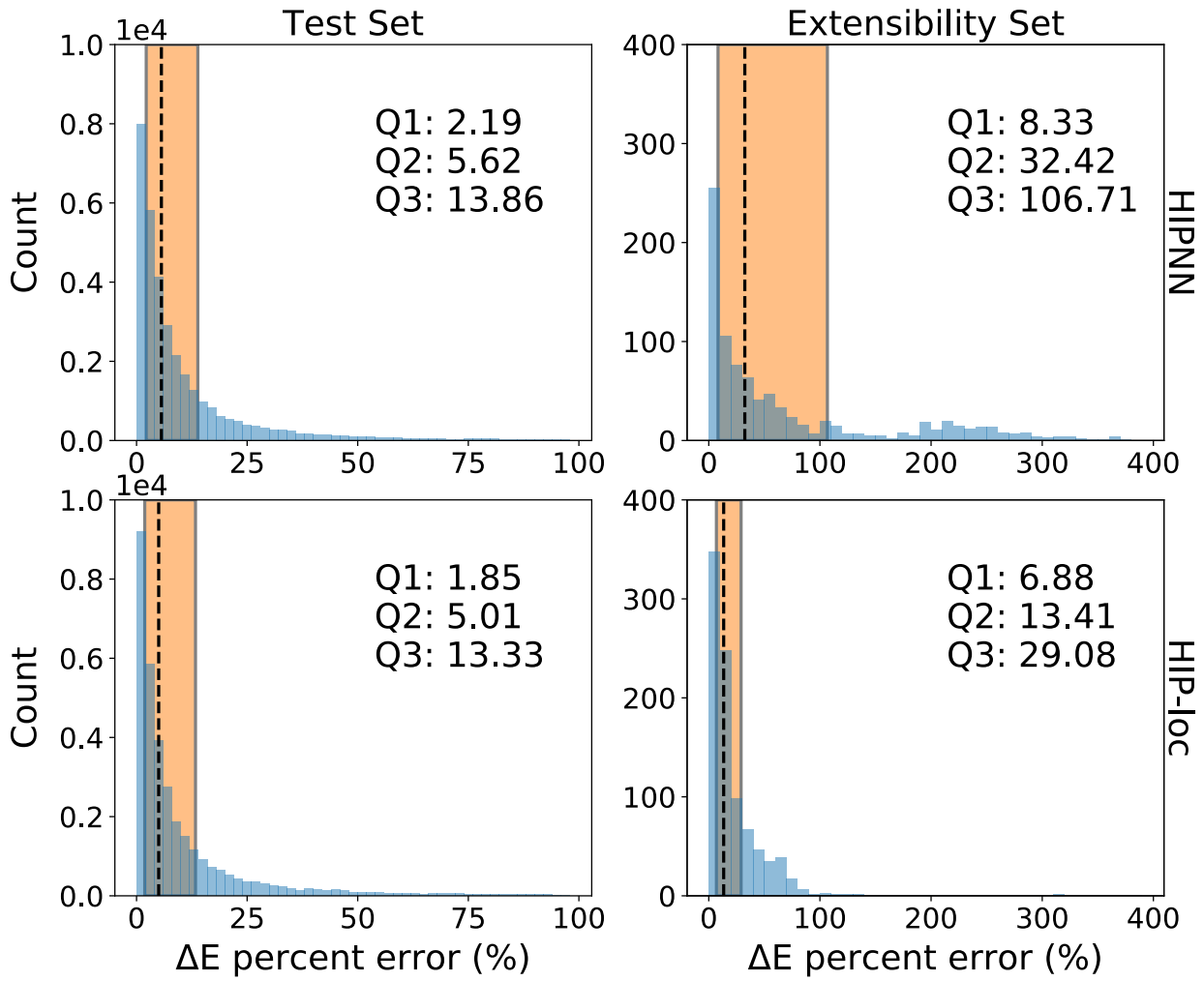

Figure S3: Histograms of percentage error in *ΔE* energy. Data are of thermal conformers sampled around the triplet equilibria of molecules in the held-out test set (left panels) and the extensibility set (right panels) using HIPNN (top panels) and HIP-loc (bottom panels). In each panel, the first (Q_1) to third (Q_3) quartiles (orange shaded regime) and median (Q_2) (black dashed line) are labeled and whose values are explicitly shown. HIP-loc outperforms HIPNN, especially on the extensibility set.

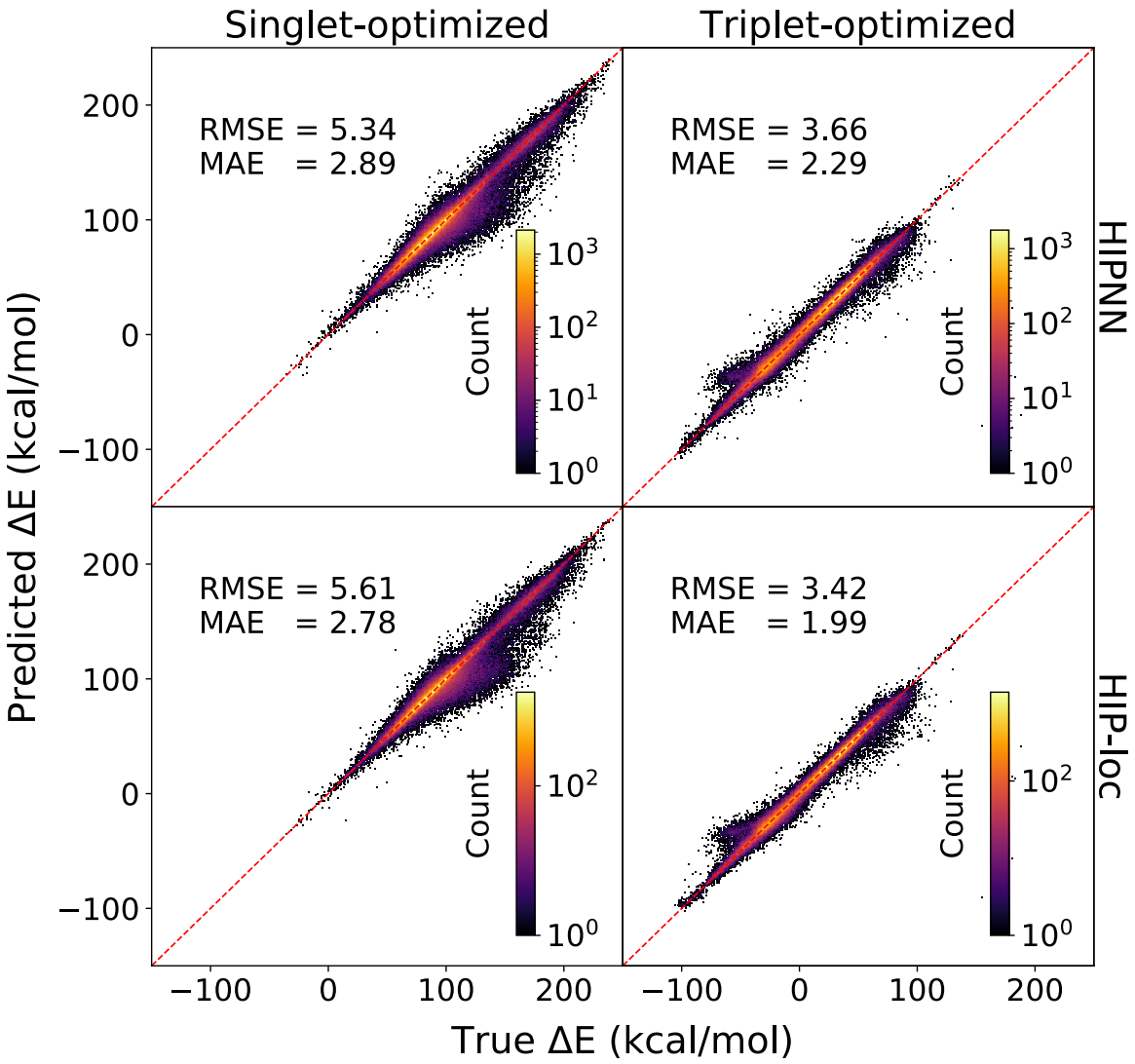

Figure S4: Parity plots of predicted versus true *ΔE* energy on thermal conformers sampled around the singlet-optimized (left panels) and triplet-optimized (right panels) geometries of the full dataset using HIPNN (top panels) and HIP-loc (bottom panels). Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

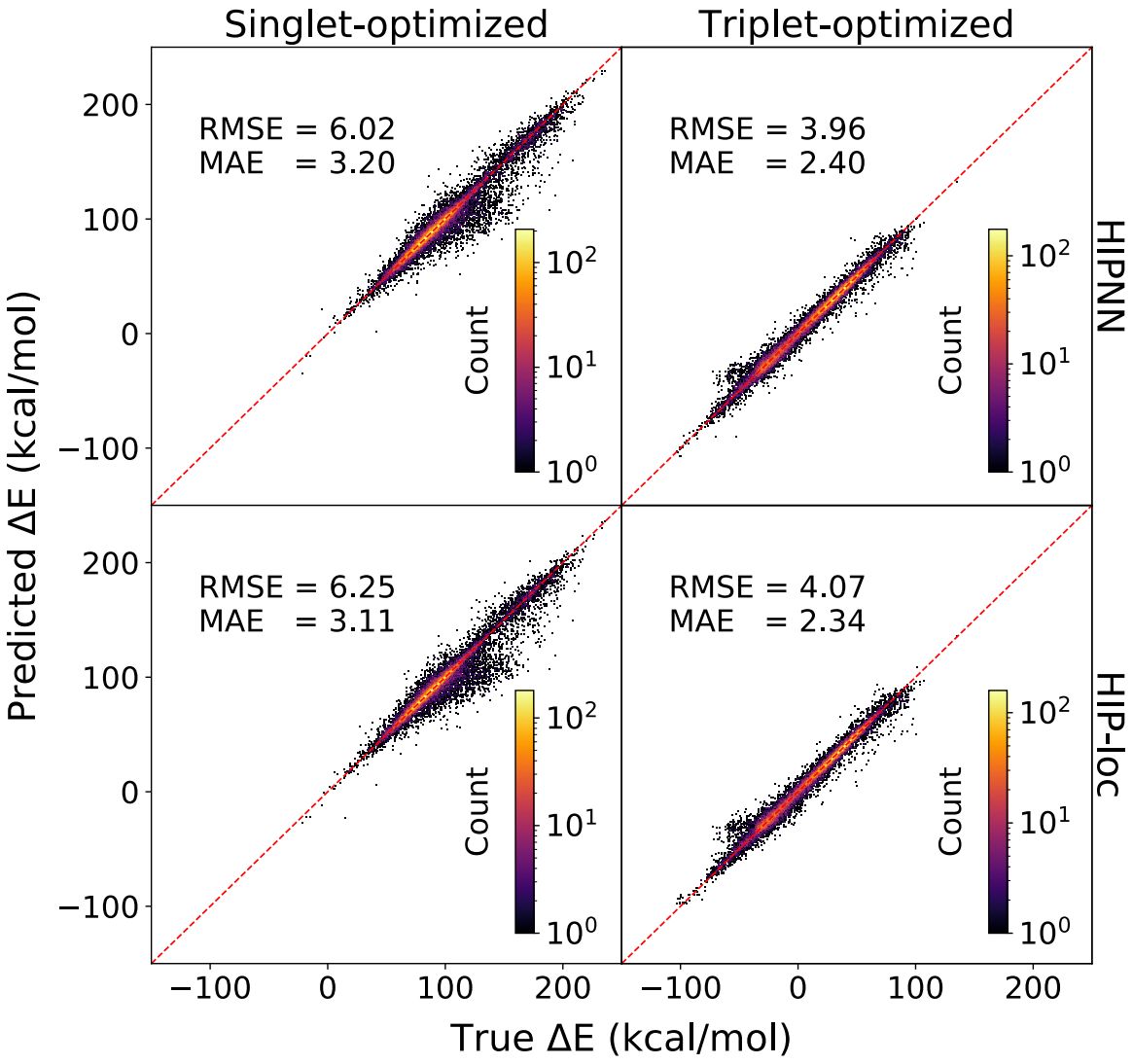

Figure S5: Parity plots of predicted versus true *ΔE* energy on thermal conformers sampled around the singlet-optimized (left panels) and triplet-optimized (right panels) geometries of the held-out test set using HIPNN (top panels) and HIP-loc (bottom panels). Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

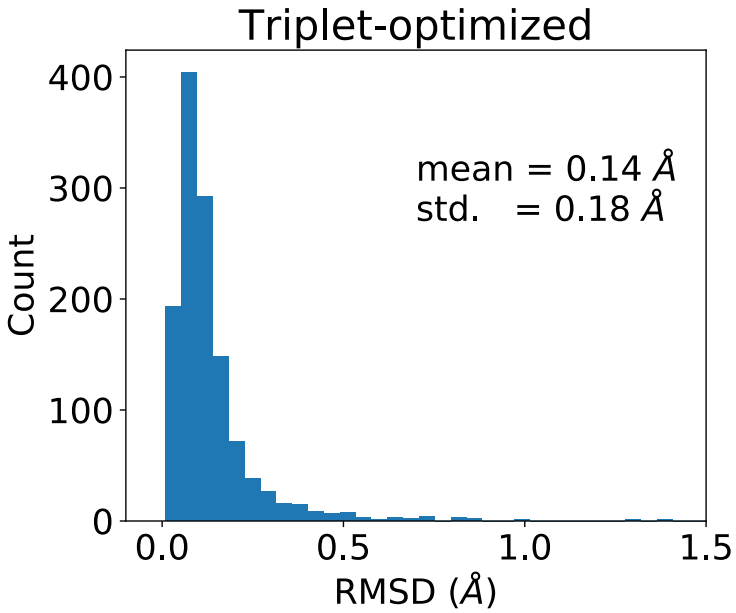

Figure S6: Root-mean-square deviation (RMSD) between DFT and HIP-loc for triplet-optimized geometries on a subset of thermal conformers in the held-out test set.

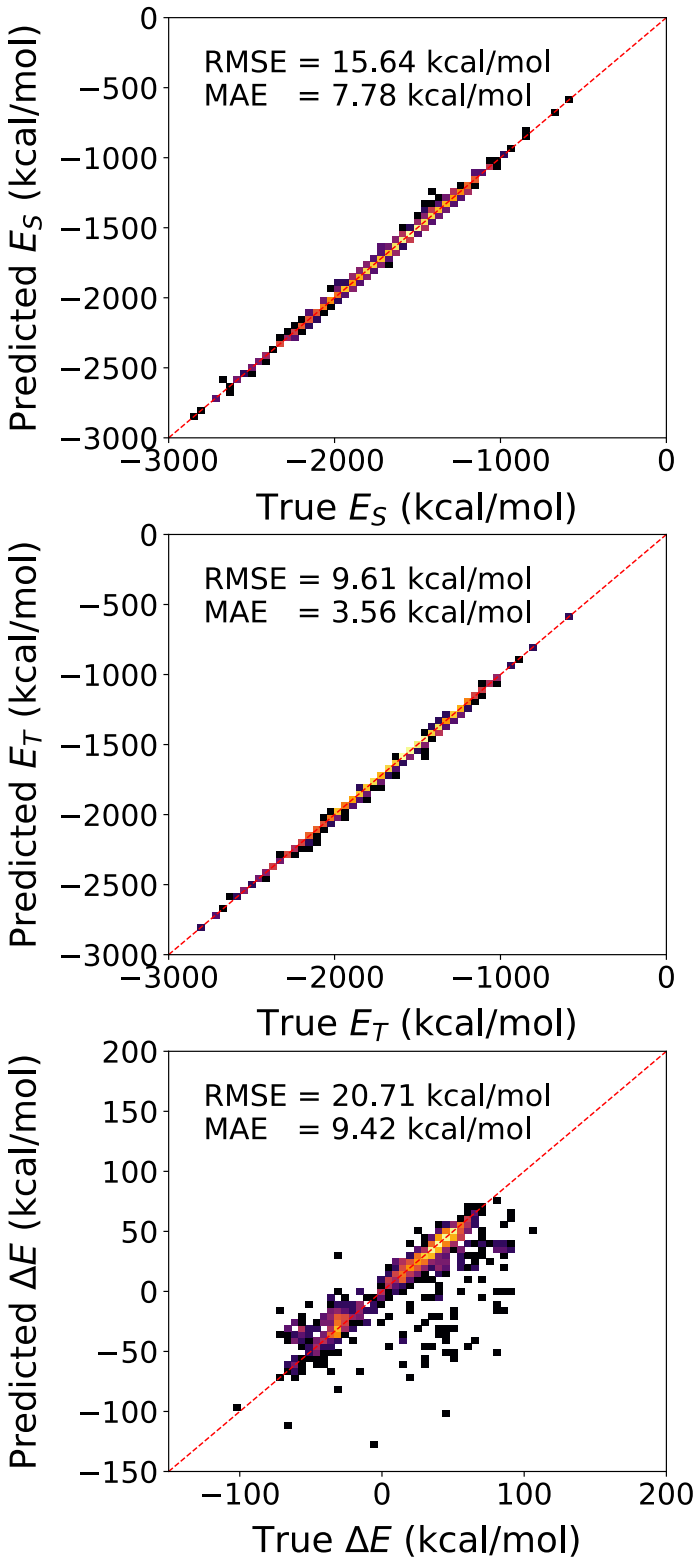

Figure S7: S_0 energies (top panel), T_1 energies (middle panel), and singlet-triplet energy gaps (bottom panel) on a subset of molecules from the held-out test set optimized using DFT (x-axes) and HIP-loc (y-axes). All molecules were optimized on the T_1 PES.

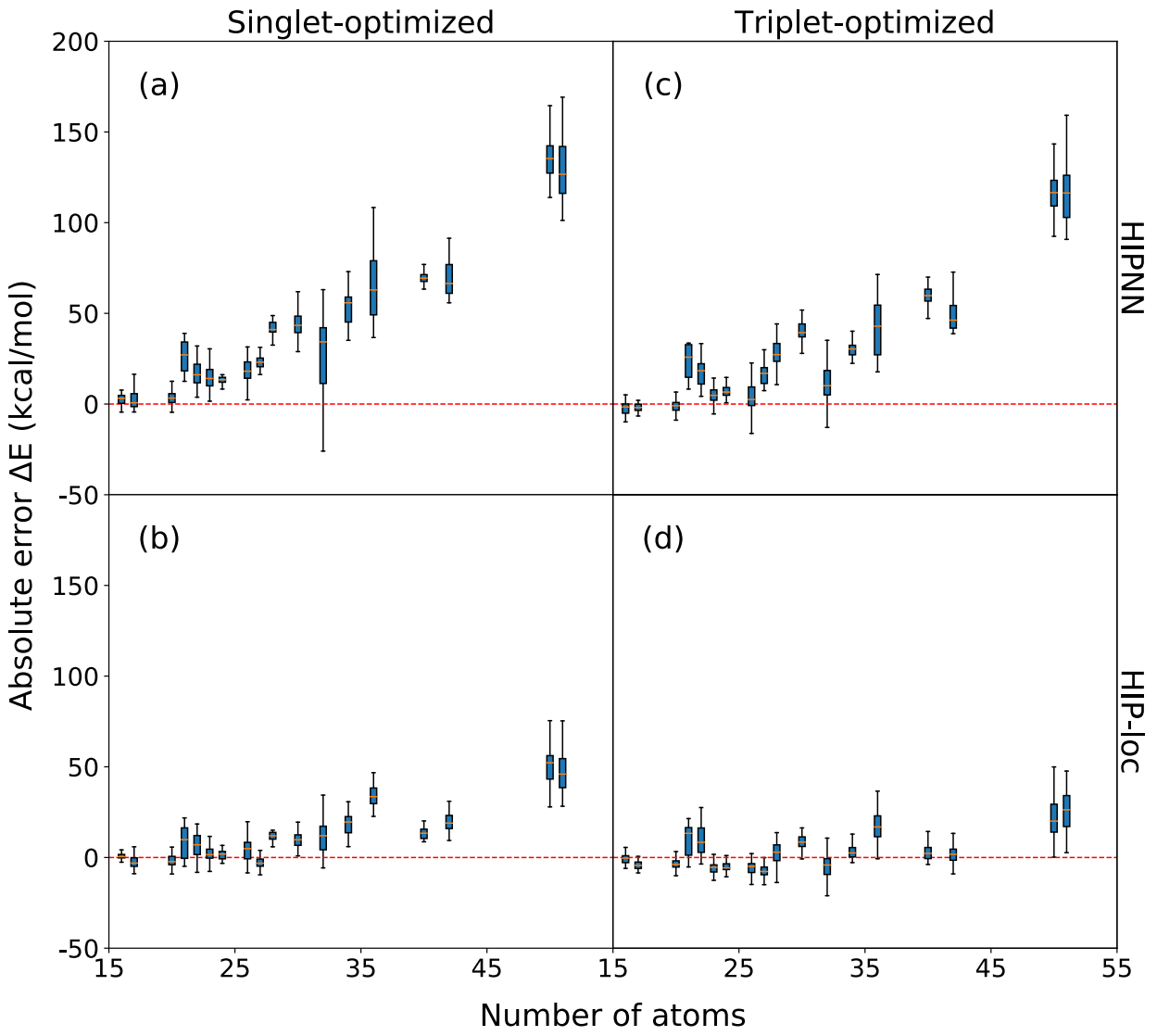

Figure S8: Absolute error in *ΔE* energy as a function of molecule size (including H). Data are of thermal conformers sampled around the singlet-optimized (left panels) and triplet-optimized (right panels) geometries of the extensibility set using HIPNN (top panels) and HIP-loc (bottom panels). Each box plot shows the median (orange line), first (Q_1) to third (Q_3) quartiles (blue box), and whiskers (black lines), capped at $Q_{1/3} \mp 1.5(Q_3 - Q_1)$. Perfect agreement between ML and reference DFT is labeled with a red dashed line. HIP-loc outperforms HIPNN, especially as molecule size increases.

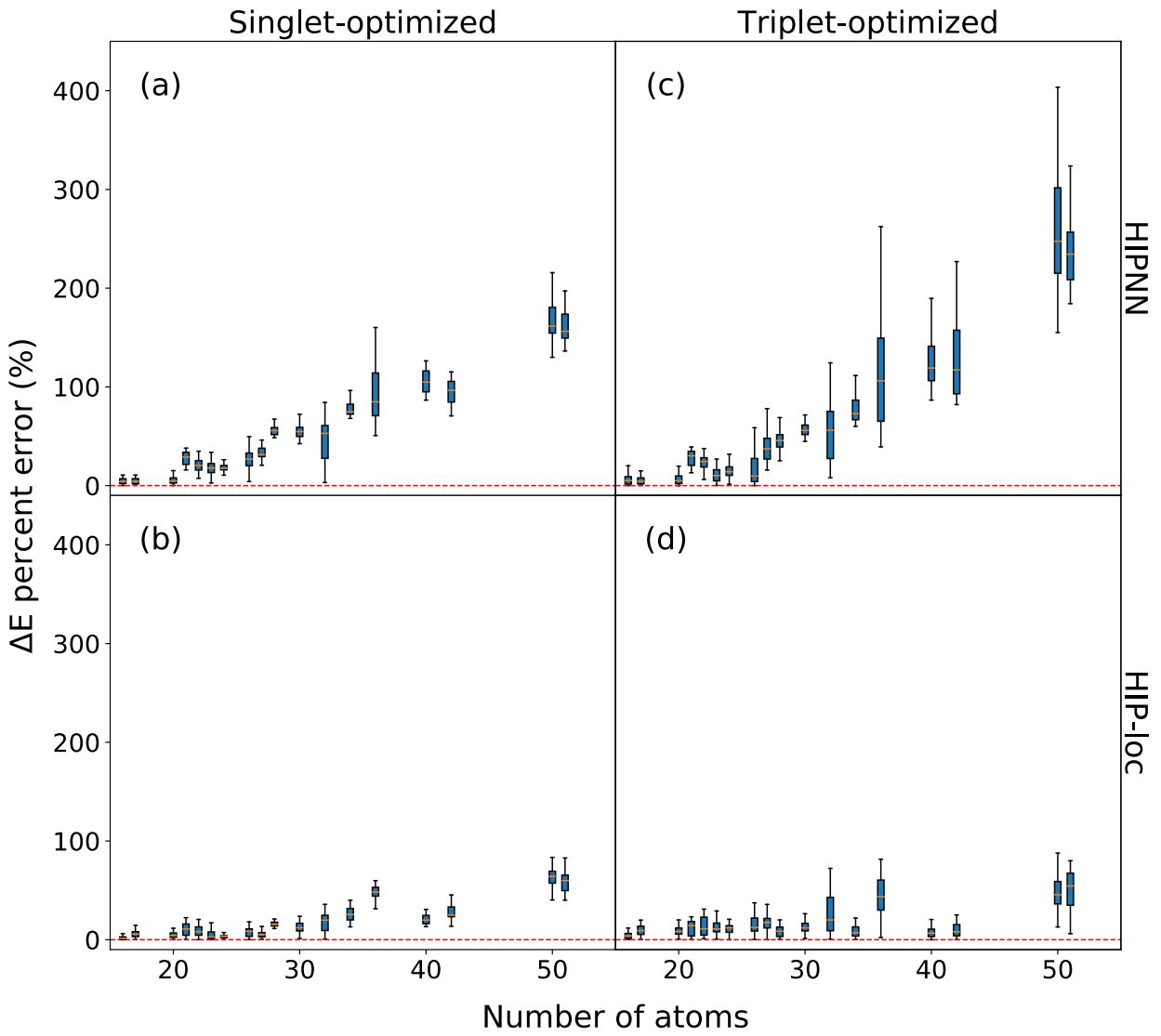

Figure S9: Percentage error in *ΔE* energy as a function of molecule size (including H). Data are of thermal conformers sampled around the singlet-optimized (left panels) and triplet-optimized (right panels) geometries of the extensibility set using HIPNN (top panels) and HIP-loc (bottom panels). Each box plot shows the median (orange line), first (Q_1) to third (Q_3) quartiles (blue box), and whiskers (black lines), capped at $Q_{1/3} \mp 1.5(Q_3 - Q_1)$. Perfect agreement between ML and reference DFT is labeled with a red dashed line. HIP-loc outperforms HIPNN, especially as molecule size increases.

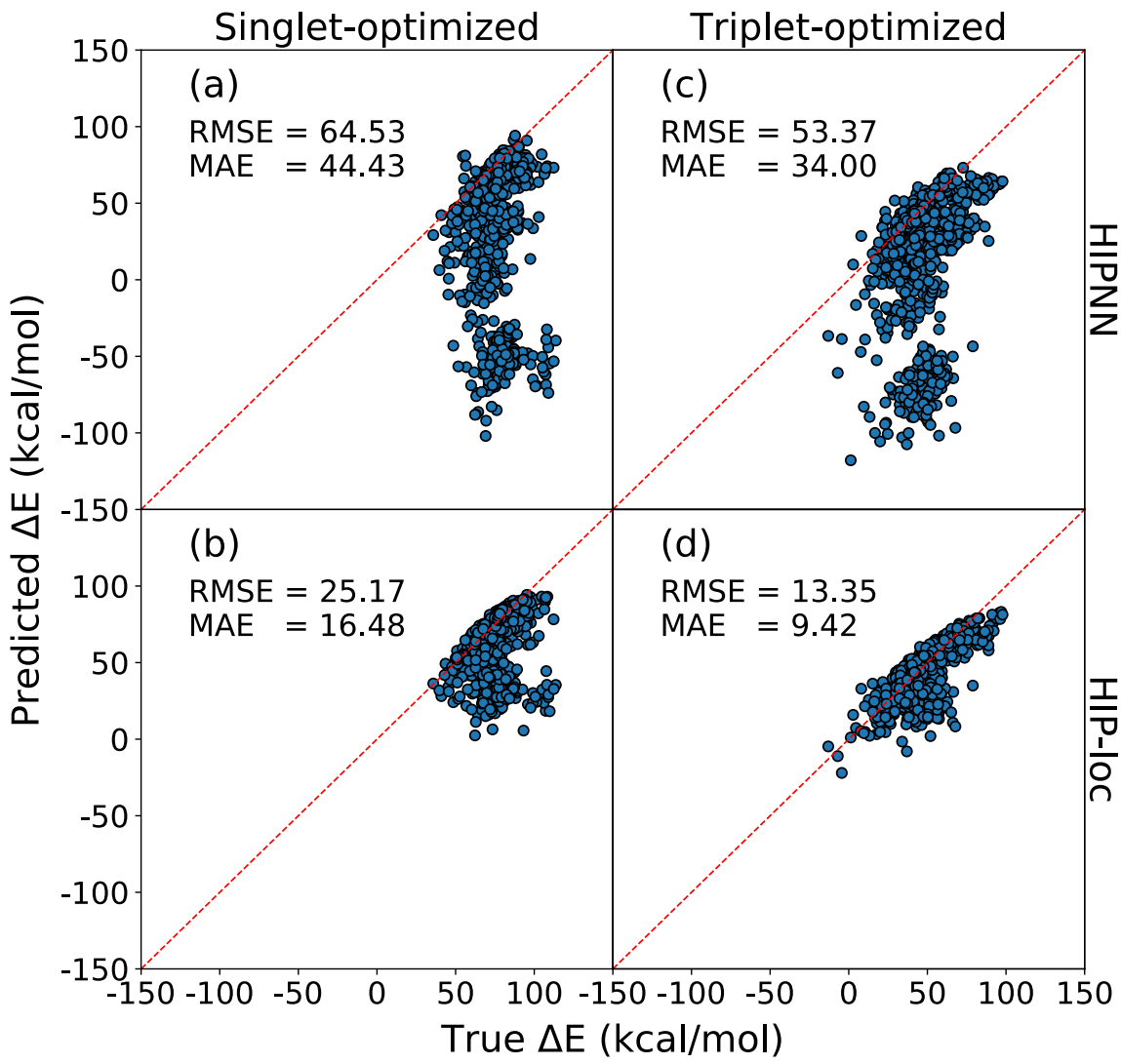

Figure S10: Parity plots of predicted versus true *ΔE* energy on thermal conformers sampled around the singlet-optimized (left panels) and triplet-optimized (right panels) geometries of the extensibility set using HIPNN (top panels) and HIP-loc (bottom panels). Predicted *ΔE* energies are significantly improved with HIP-loc. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

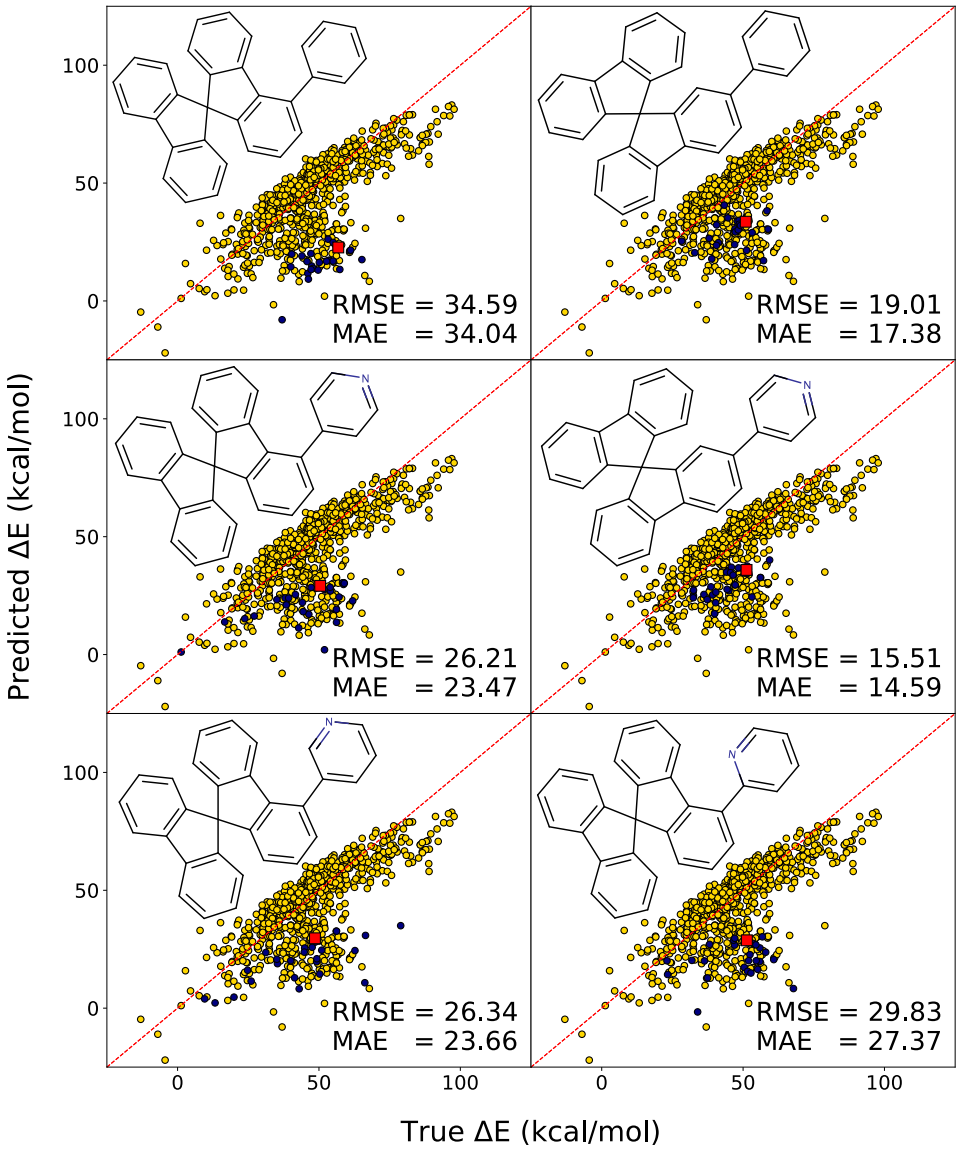

Figure S11: Parity plots of predicted versus true *ΔE* energy for a group of chemically similar molecules of the extensibility set, each overlaid with the results of all molecules in the extensibility set. The optimized structure is labeled with a red square and thermal conformers are labeled with blue circles. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

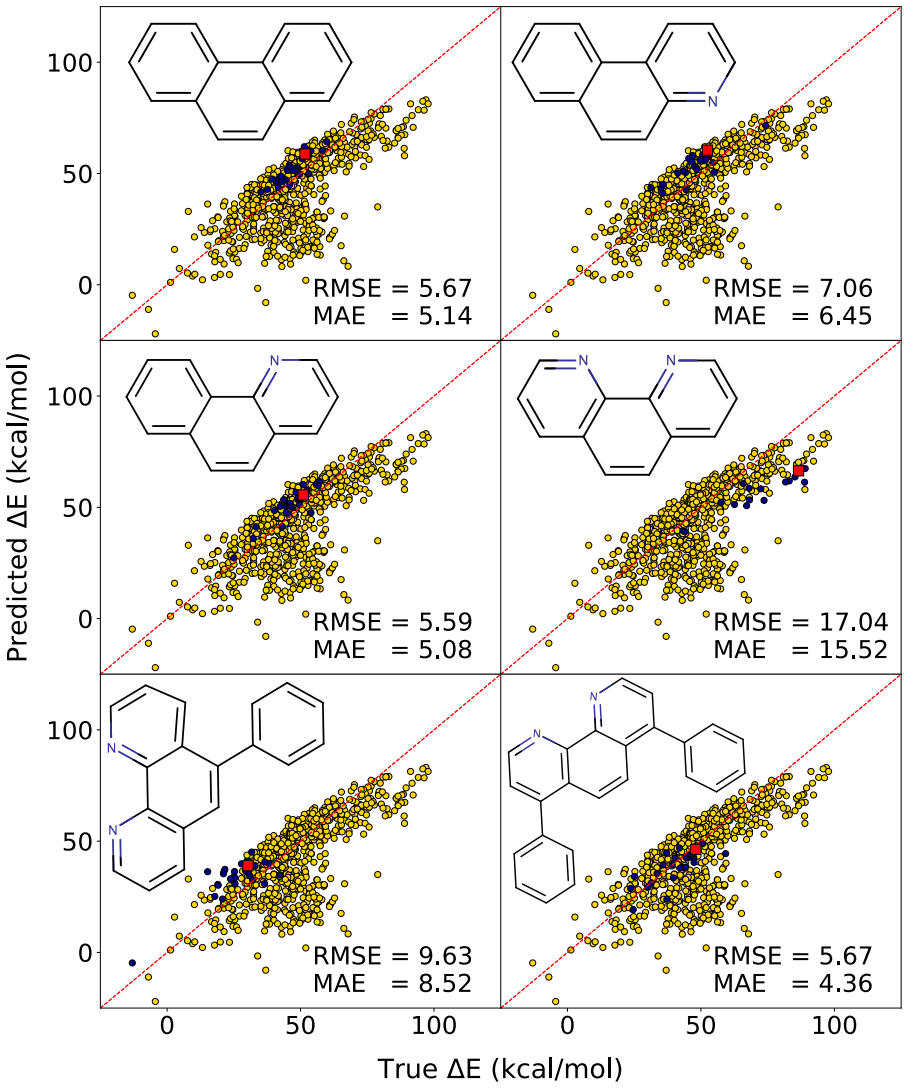

Figure S12: Parity plots of predicted versus true *ΔE* energy for a group of chemically similar molecules of the extensibility set, each overlaid with the results of all molecules in the extensibility set. The optimized structure is labeled with a red square and thermal conformers are labeled with blue circles. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

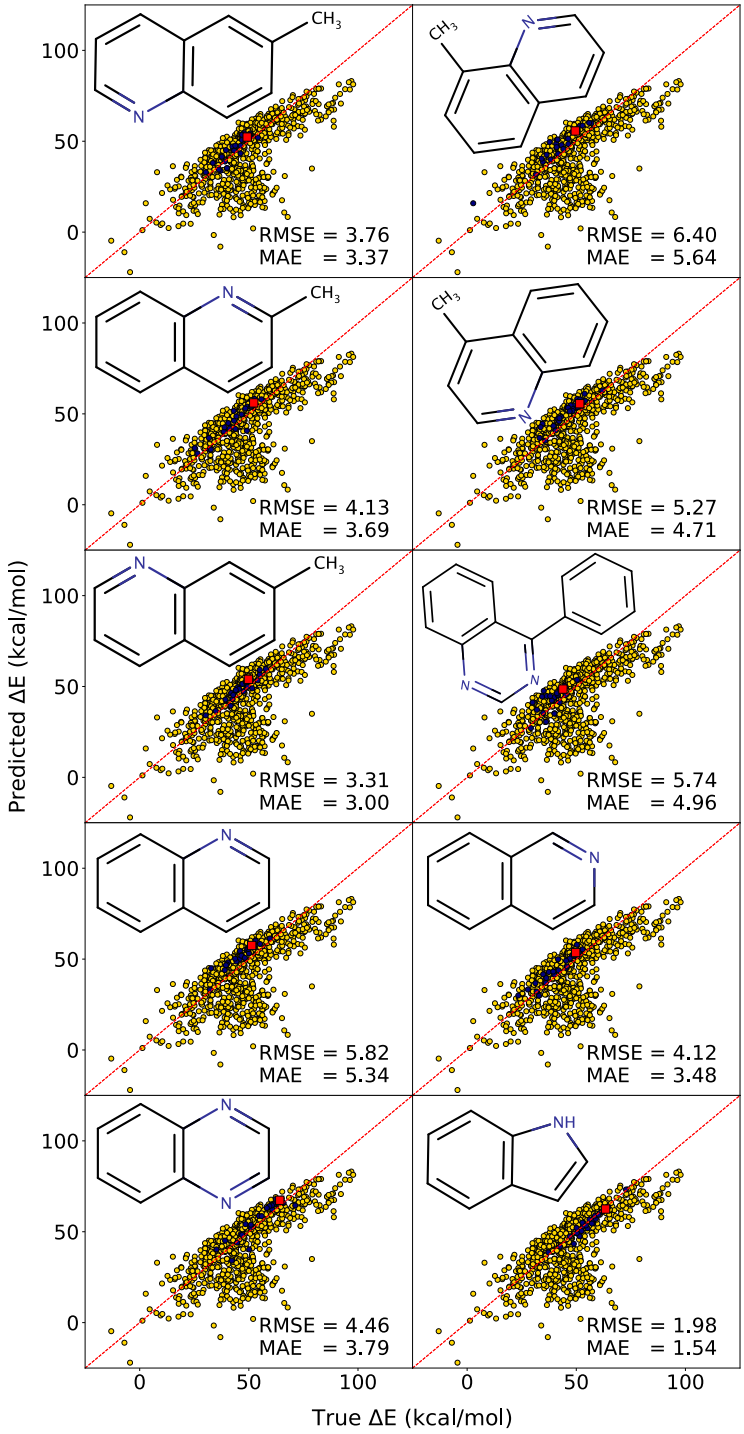

Figure S13: Parity plots of predicted versus true *ΔE* energy for a group of chemically similar molecules of the extensibility set, each overlaid with the results of all molecules in the extensibility set. The optimized structure is labeled with a red square and thermal conformers are labeled with blue circles. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

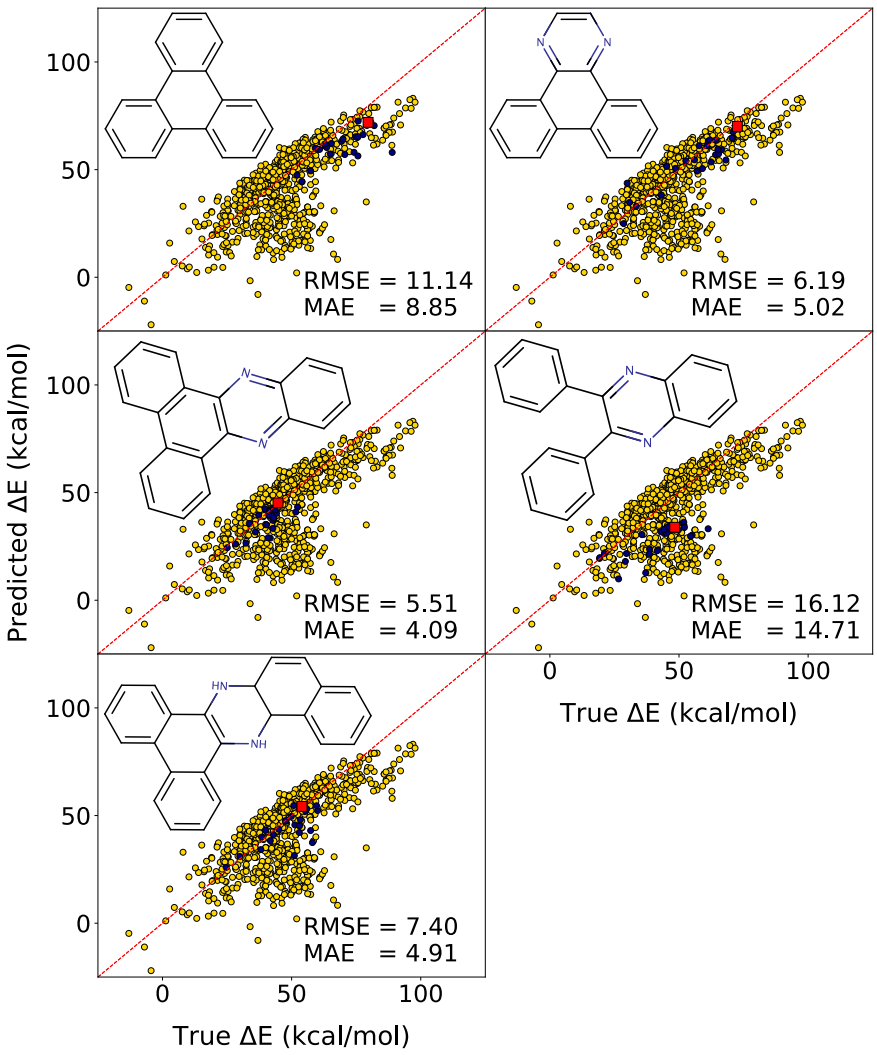

Figure S14: Parity plots of predicted versus true *ΔE* energy for a group of chemically similar molecules of the extensibility set, each overlaid with the results of all molecules in the extensibility set. The optimized structure is labeled with a red square and thermal conformers are labeled with blue circles. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

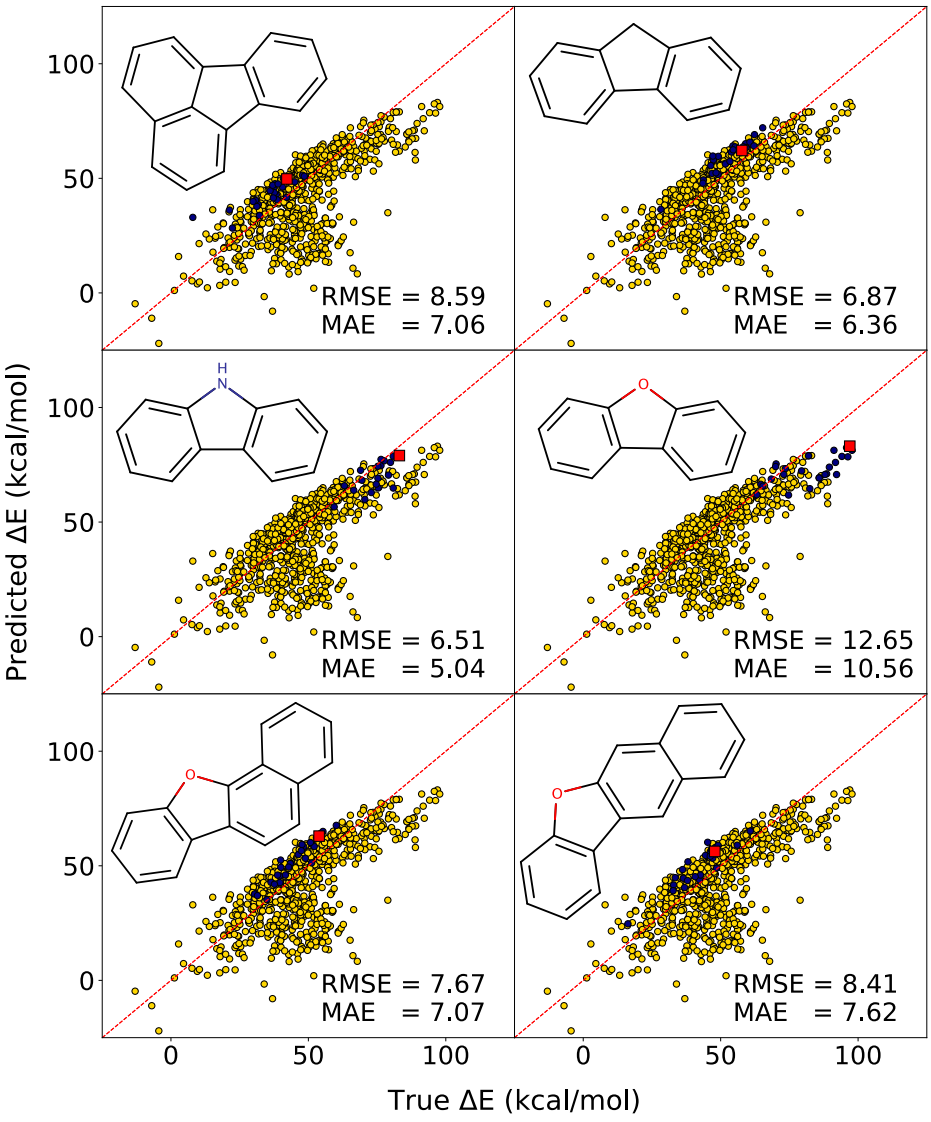

Figure S15: Parity plots of predicted versus true *ΔE* energy for a group of chemically similar molecules of the extensibility set, each overlaid with the results of all molecules in the extensibility set. The optimized structure is labeled with a red square and thermal conformers are labeled with blue circles. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

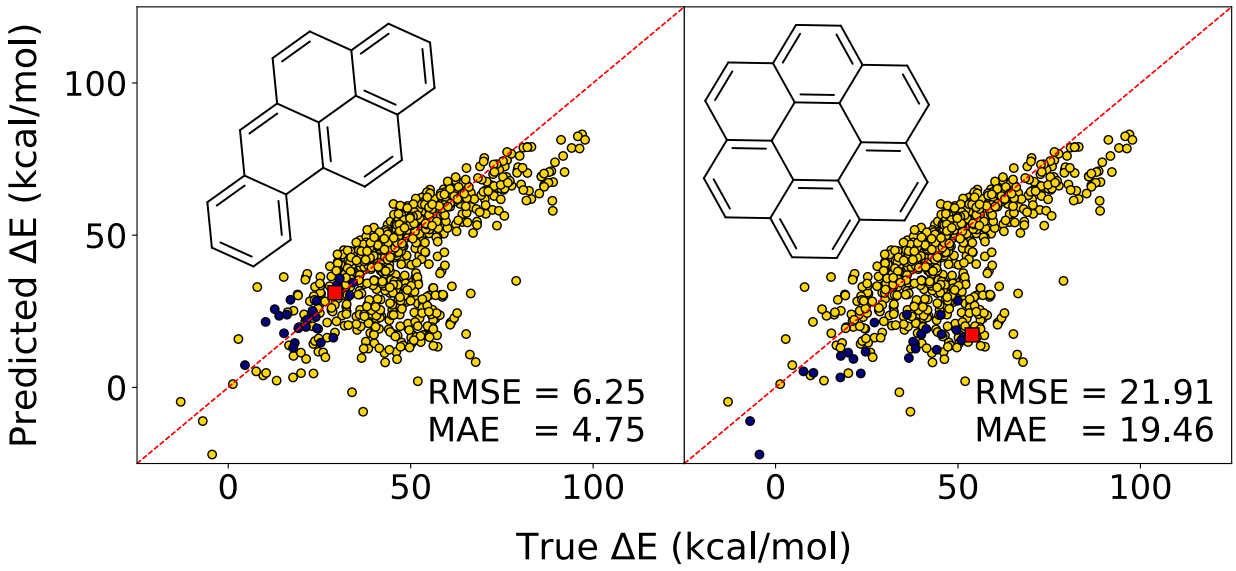

Figure S16: Parity plots of predicted versus true *ΔE* energy for a group of chemically similar molecules of the extensibility set, each overlaid with the results of all molecules in the extensibility set. The optimized structure is labeled with a red square and thermal conformers are labeled with blue circles. Prediction errors are expressed in root-mean-square error (RMSE) and mean-absolute error (MAE).

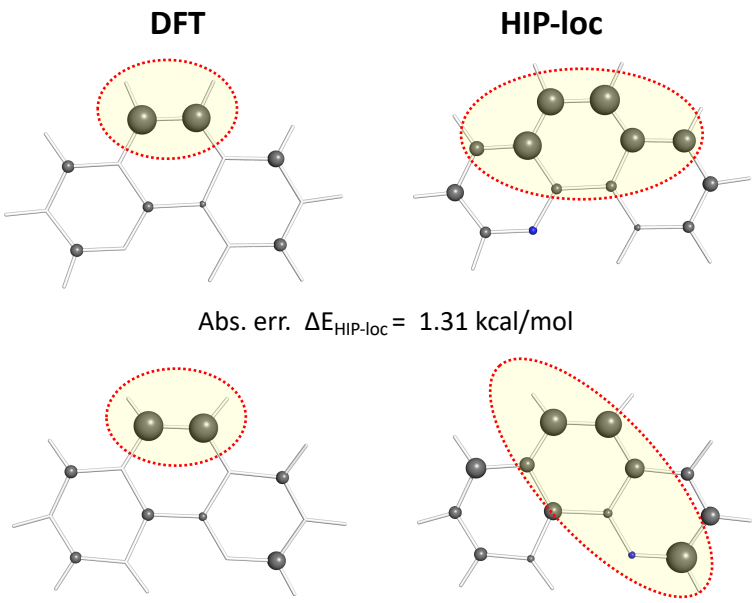

Figure S17: A comparison of DFT spin density and HIP-loc localization weights for a select molecule of the extensibility set. Conformers with the best and worst predicted *ΔE* energy, measured in absolute error (Abs. err.), are shown. Predominant localities associated with the spin transitions are circled. Localities inferred by HIP-loc weights are in qualitative agreement with reference DFT spin densities for the best performing conformer. The root-mean-square error (RMSE) and mean-absolute error (MAE) over all 27 conformers are also shown.

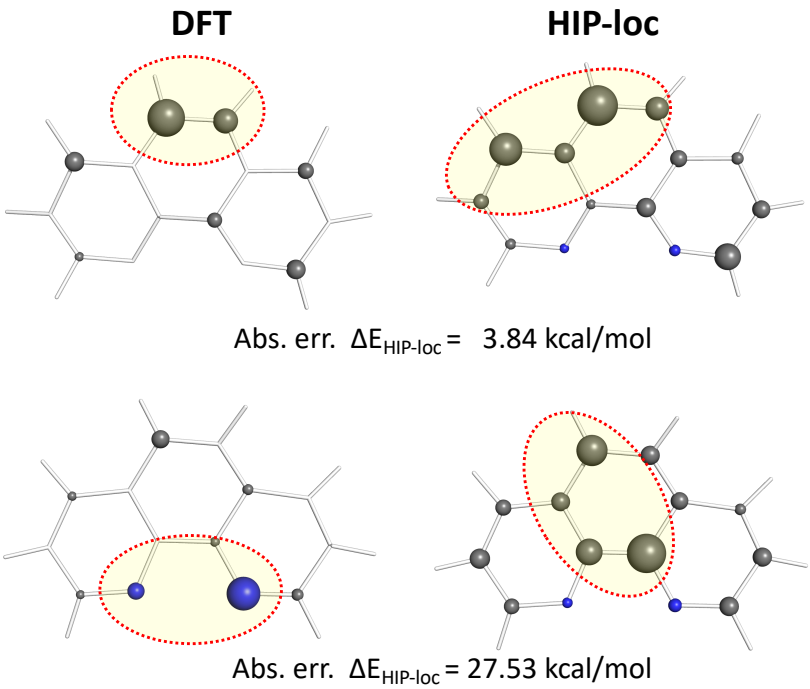

Figure S18: A comparison of DFT spin density and HIP-loc localization weights for a select molecule of the extensibility set. Conformers with the best and worst predicted *ΔE* energy, measured in absolute error (Abs. err.), are shown. Predominant localities associated with the spin transitions are circled. Localities inferred by HIP-loc weights are in qualitative agreement with reference DFT spin densities for the best performing conformer. The root-mean-square error (RMSE) and mean-absolute error (MAE) over all 27 conformers are also shown.

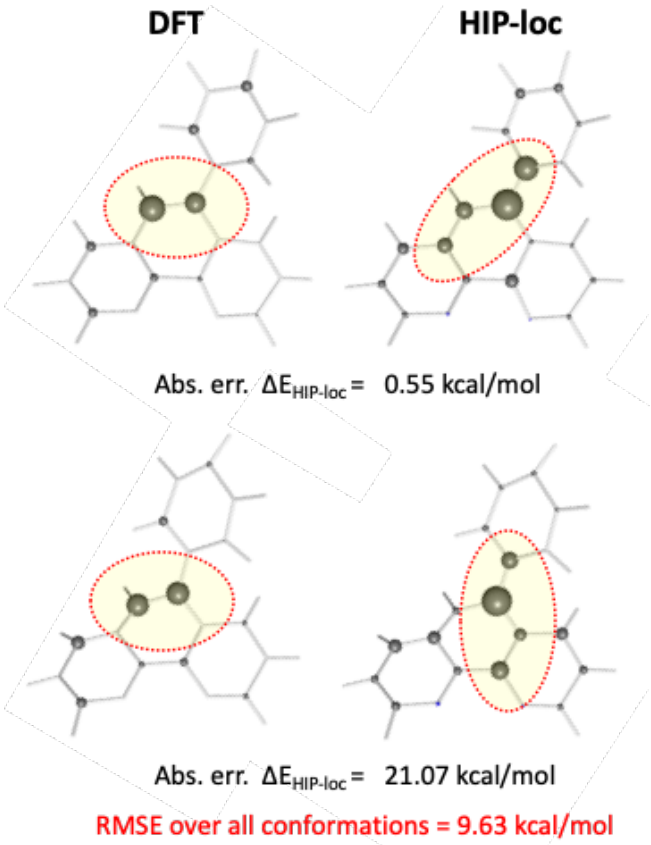

Figure S19: A comparison of DFT spin density and HIP-loc localization weights for a select molecule of the extensibility set. Conformers with the best and worst predicted *ΔE* energy, measured in absolute error (Abs. err.), are shown. Predominant localities associated with the spin transitions are circled. Localities inferred by HIP-loc weights are in qualitative agreement with reference DFT spin densities for the best performing conformer. The root-mean-square error (RMSE) and mean-absolute error (MAE) over all 27 conformers are also shown.

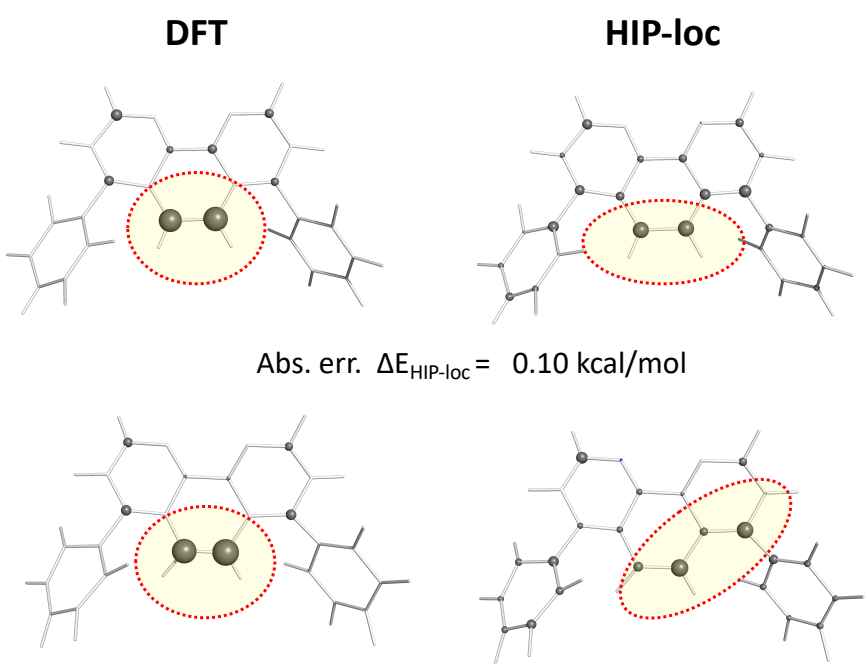

Figure S20: A comparison of DFT spin density and HIP-loc localization weights for a select molecule of the extensibility set. Conformers with the best and worst predicted *ΔE* energy, measured in absolute error (Abs. err.), are shown. Predominant localities associated with the spin transitions are circled. Localities inferred by HIP-loc weights are in qualitative agreement with reference DFT spin densities for the best performing conformer. The root-mean-square error (RMSE) and mean-absolute error (MAE) over all 27 conformers are also shown.

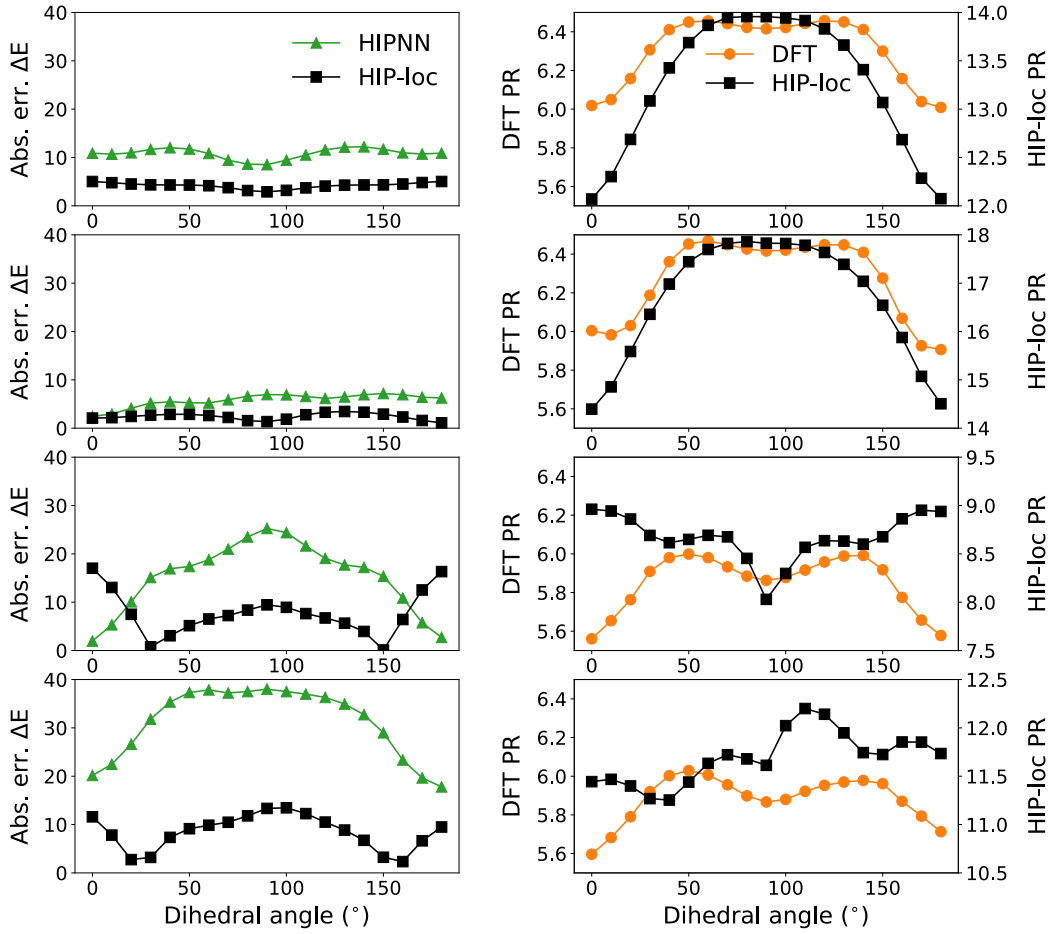

Figure S21: Dihedral scans of molecules with a single torsional angle. Absolute error (Abs. err.) in *ΔE* computed with HIPNN and HIP-loc (left panels). Participation ratio (PR) estimated using atom-centered DFT spin density (right panels, left axis) and HIP-loc localization weights (right panels, right axis) as a function of dihedral angle. HIP-loc outperforms HIPNN in predicting *ΔE* for all molecules. For the top two molecules with relatively low *ΔE* errors, the trend in the PRs computed with DFT and HIP-loc qualitatively agree with one another. By contrast, the trends in PR differ significantly between DFT and HIP-loc for the bottom two molecules and coincidently the errors in *ΔE* are relatively large for HIP-loc, albeit considerably lower than the errors without accounting for localization (HIPNN).

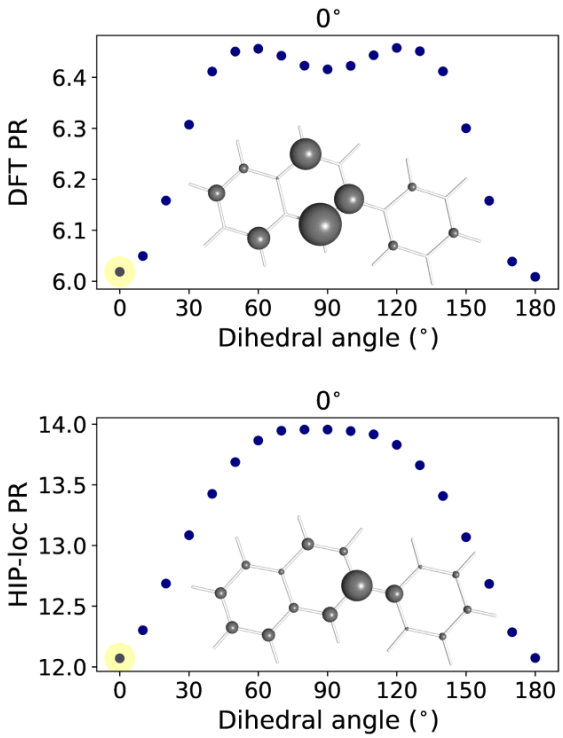

Figure S22: Scan of the central dihedral angle in a representative molecule in Fig. 4 of the main text. Participation ratios (PRs) are computed using atom-centered DFT density (top panel) and HIP-loc localization weights (bottom panel) as a function of dihedral angle. For both DFT and HIP-loc, the PR in the planar structure (dihedral of 0° and 180°) is delocalized across atoms on each ring, whereas the PR in the non-planar structure (dihedral of 90°) is localized to more atoms exclusively on the larger ring. The net effect is an increase in PR for the non-planar conformation. Compared to DFT, HIP-loc infers more delocalized transitions, but qualitative agreement between the methods' PRs is observed. An animated figure covering all molecular conformations can be found in the ESI.

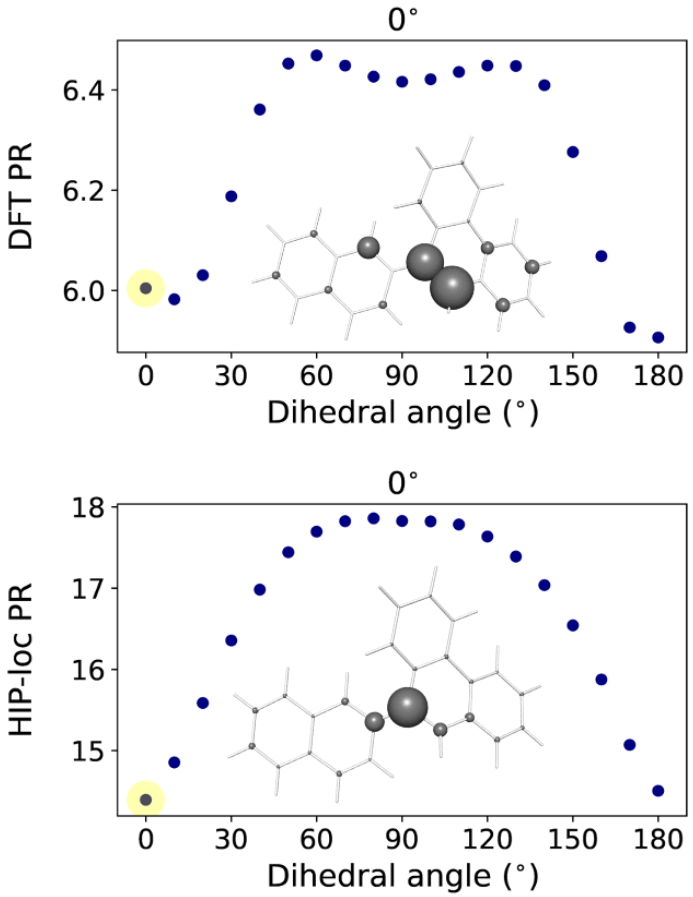

Figure S23: Scan of the central dihedral angle in a representative molecule in Fig. S19 of the main text. Participation ratios (PRs) are computed using atom-centered DFT density (top panel) and HIP-loc localization weights (bottom panel) as a function of dihedral angle. An animated figure covering all molecular conformations can be found in the ESI.

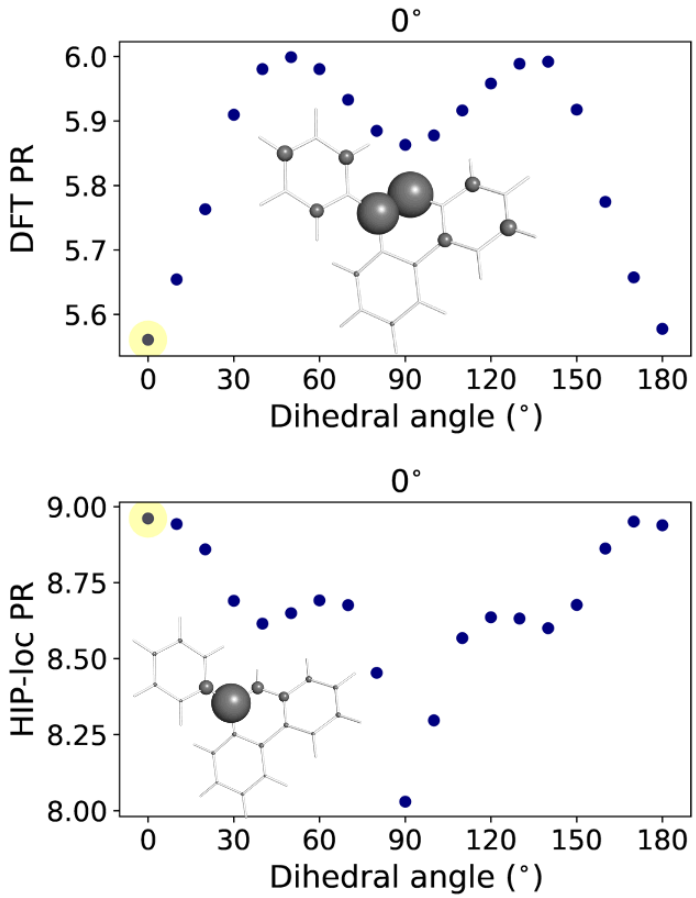

Figure S24: Scan of the central dihedral angle in a representative molecule in Fig. S19 of the main text. Participation ratios (PRs) are computed using atom-centered DFT density (top panel) and HIP-loc localization weights (bottom panel) as a function of dihedral angle. An animated figure covering all molecular conformations can be found in the ESI.

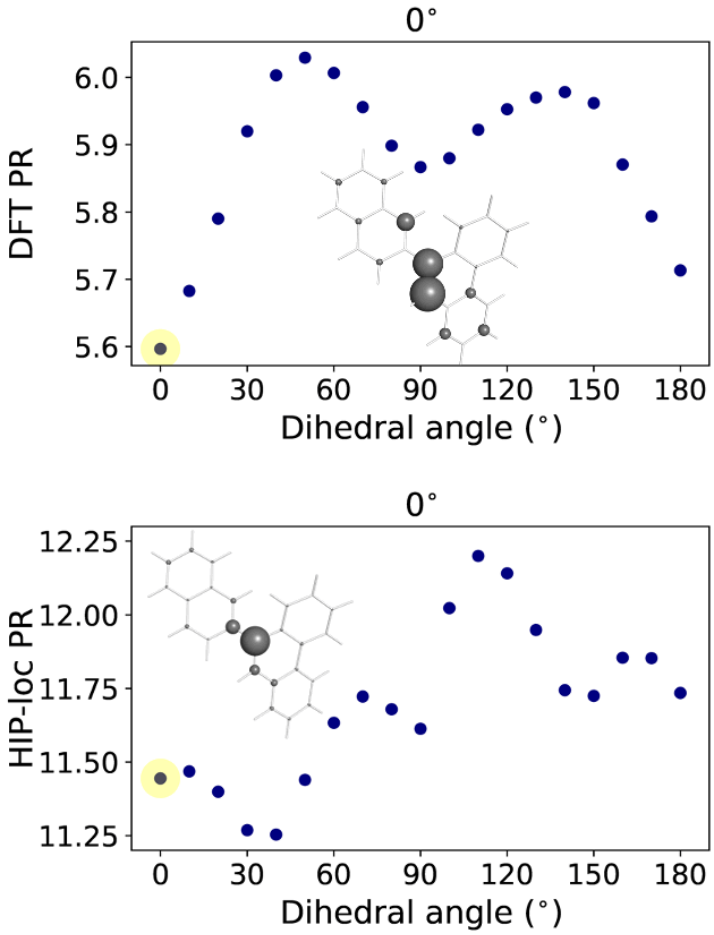

Figure S25: Scan of the central dihedral angle in a representative molecule in Fig. S19 of the main text. Participation ratios (PRs) are computed using atom-centered DFT density (top panel) and HIP-loc localization weights (bottom panel) as a function of dihedral angle. An animated figure covering all molecular conformations can be found in the ESI.