Supplementary Information

Hexamer ID	GFN2 (A^3)	ORCA (A^3)	Gaussian (A^3)
B:1073-1156-000011	205.010	1281.299	1278.897
C:520-515-111111	184.176	1339.674	1339.809

Table S1 Outlier Polarizability Comparison

Hexamer ID	Tanimoto Coefficient
520-515-111111	0.535
154-515-001100	0.384
71-515-101010	0.355
222-364-001000	0.261
337-515-101010	0.241

Table S2 Top 5 Chemically Similar Wide Range Set Molecules to Outlier A (222-515-111100)

Hexamer ID	Tanimoto Coefficient
1175-67-010101	0.324
496-979-000011	0.263
1130-1192-010111	0.262
104-680-100010	0.242
930-485-000100	0.239

Table S3 Top 5 Chemically Similar Wide Range Set Molecules to Outlier B (1073-1156-000011)

Hexamer ID	Tanimoto Coefficient
222-515-111100	0.535
71-515-101010	0.427
154-515-001100	0.375
337-515-101010	0.257
411-525-000000	0.228

Table S4 Top 5 Chemically Similar Wide Range Set Molecules to Outlier C (520-515-111111)

PubChemQC ID	Gaussian ωB97XD ($Å^3$)	ORCA ωB97X ($Å^3$)
10844636	15.8972868	15.78617472
19800353	31.9664682	31.60996769
21545508	23.31098235	23.09620301
65163876	26.31024675	26.07089537

Table S5 Polarizability Method Comparison for PubChemQC Sample

The MAE comparing ORCA ωB97X to Gaussian ωB97XD for this sample of the PubChemQC subset is 0.184$Å^3$, which equates to approximately 0.7% error on the scale of this sample set.

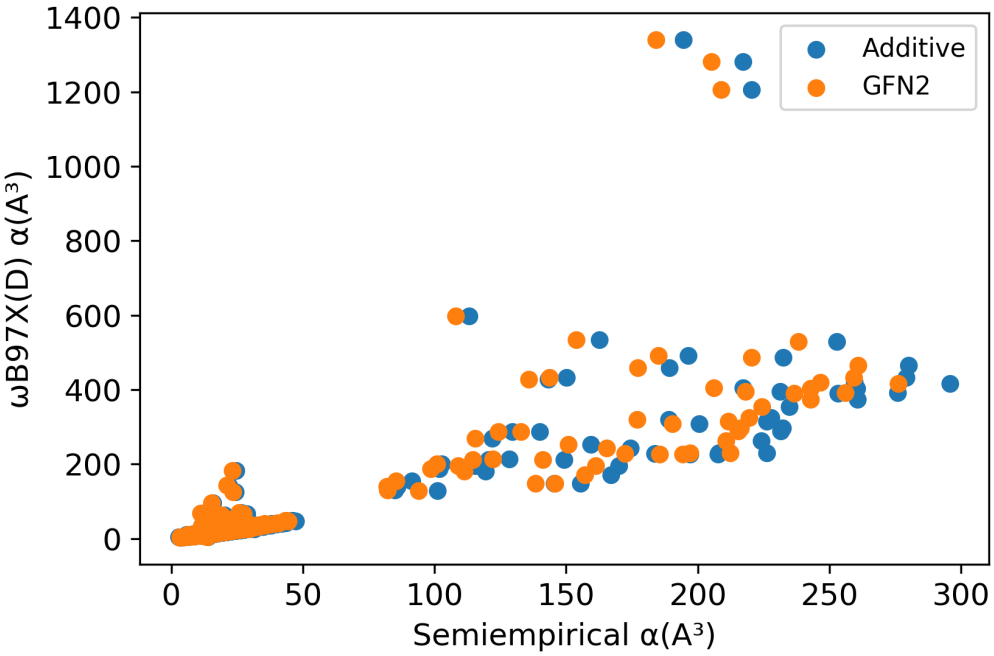

Fig. S1 Comparison of additive model and GFN2 polarizabilities to DFT results.

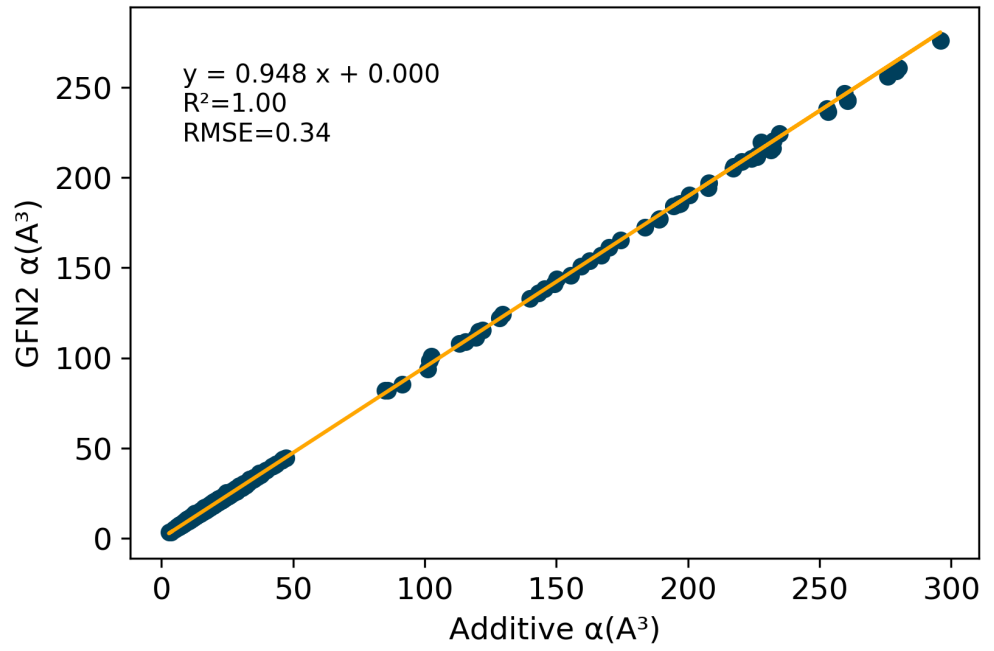

Fig. S2 Strong correlation is shown between GFN2 and additive model polarizabilites.

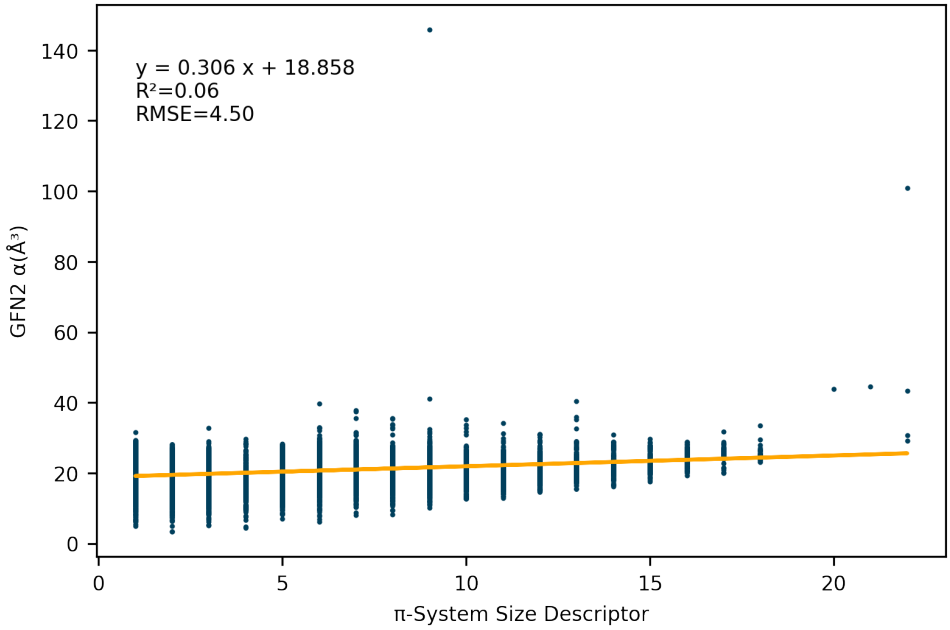

Fig. S3 Almost no correlation is shown between GFN2/D4 calculated polarizabilities and a π-system size descriptor for molecules with 25 or fewer π-systems.

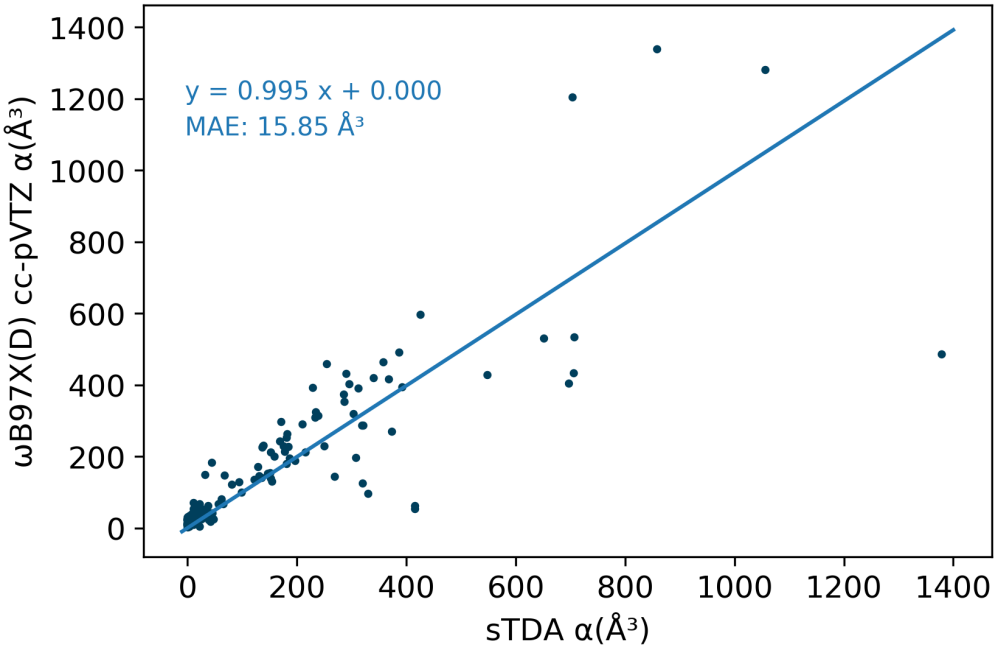

Fig. S4 sTDA-xTB shows worse correlation with DFT than GFN2.

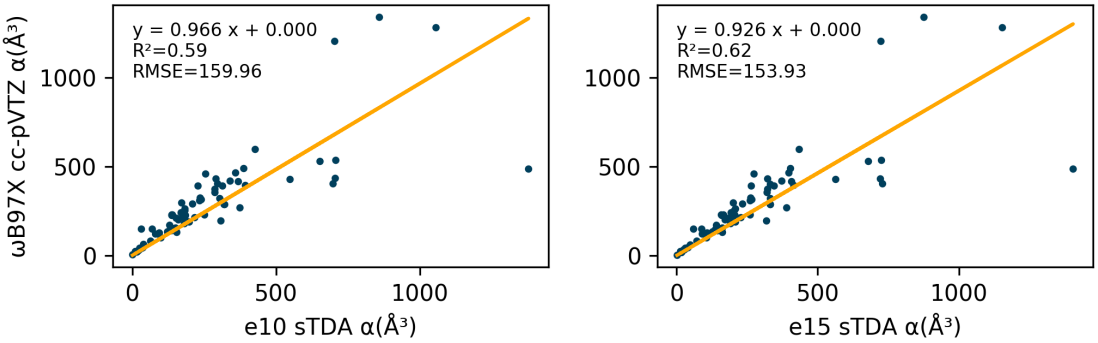

Fig. S5 Comparison of DFT vs sTDA isotropic polarizability results for wide range molecule set. Energy thresholds used for sTDA calculations are noted as their keyword arguments: e10 for 10 eV and e15 for 15 eV.

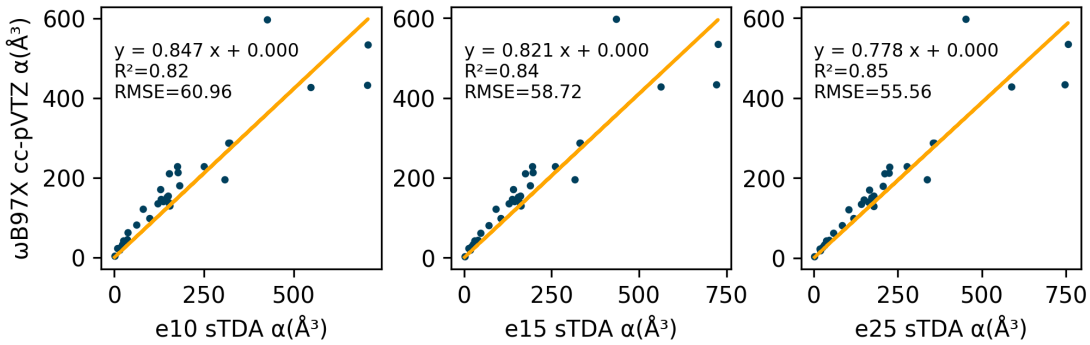

Fig. S6 Comparison of DFT vs sTDA isotropic polarizability results for the 34 members of the wide range molecule set we were able to successfully run sTDA with a 25 eV energy threshold. Energy thresholds used for sTDA calculations are noted as their keyword arguments: e10 for 10 eV, e15 for 15 eV, and e25 for 25 eV.

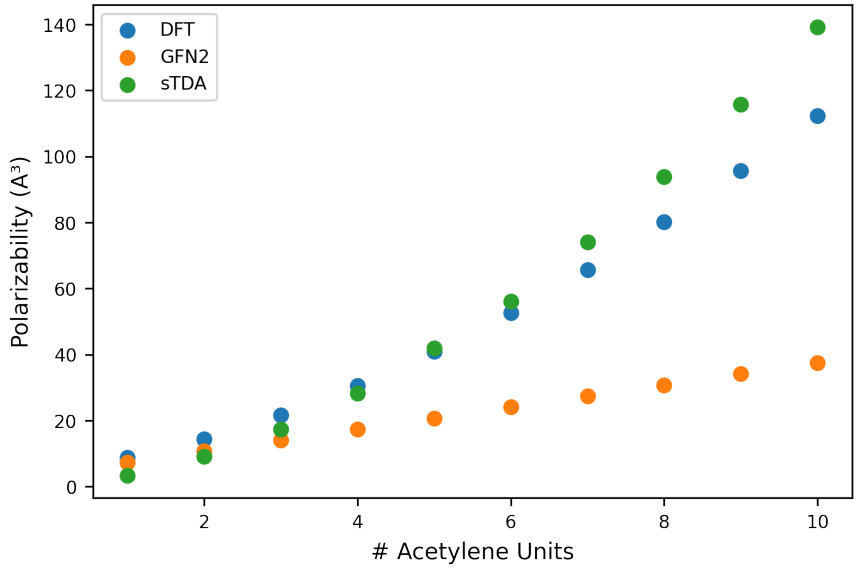

Fig. S7 Comparison of isotropic polarizabilities calculated for polyacetylene oligomers of increasing length using DFT with the ωB97X functional and the cc-pVTZ basis set corrected to aug-cc-pVTZ accuracy using our linear correction from Figure S13, GFN2, and sTDA.

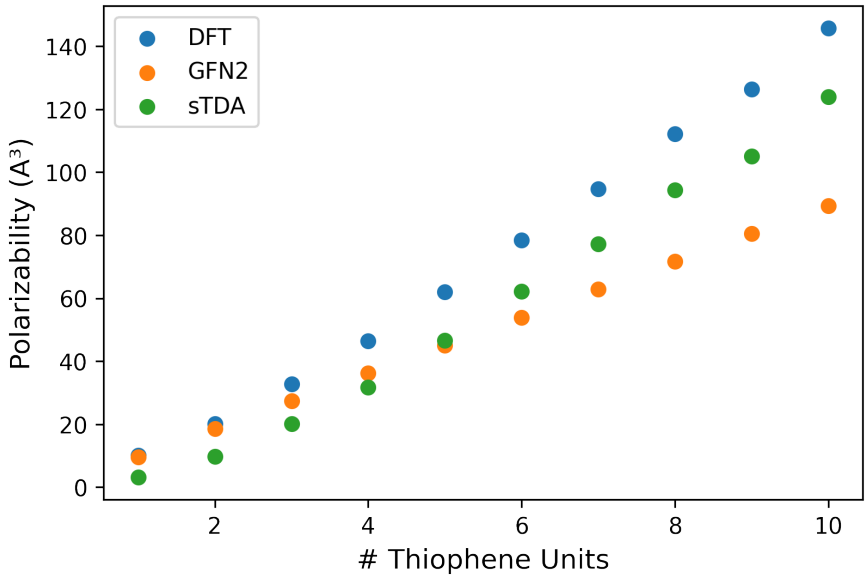

Fig. S8 Comparison of isotropic polarizabilities calculated for polythiophene oligomers of increasing length using DFT with the ωB97X functional and the cc-pVTZ basis set corrected to aug-cc-pVTZ accuracy using our linear correction from Figure S13, GFN2, and sTDA.

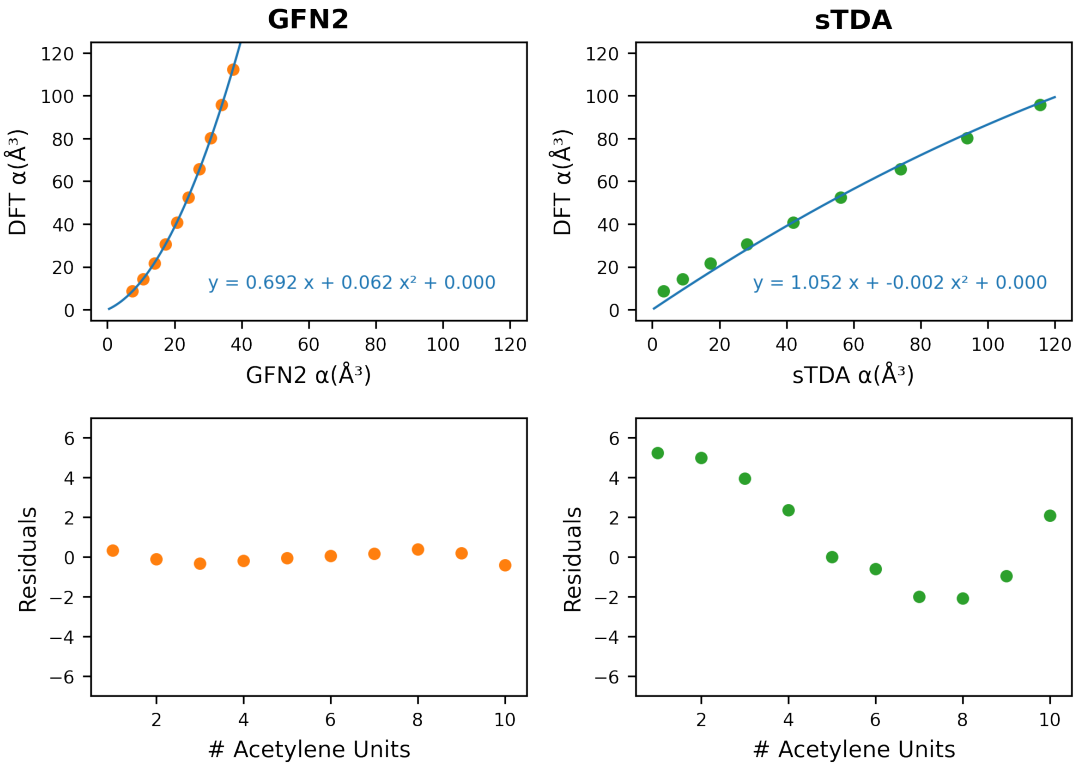

Fig. S9 Quadratic regression with forced zero intercept of isotropic polarizabilities calculated for polyacetylene oligomers of increasing length using DFT with the ωB97X functional and the cc-pVTZ basis set corrected to aug-cc-pVTZ accuracy using our linear correction from Figure S13 and GFN2 and sTDA. The residuals from the DFT vs. GFN2 regression are notably smaller than those from the DFT vs. sTDA regression, indicating the latter's larger random error.

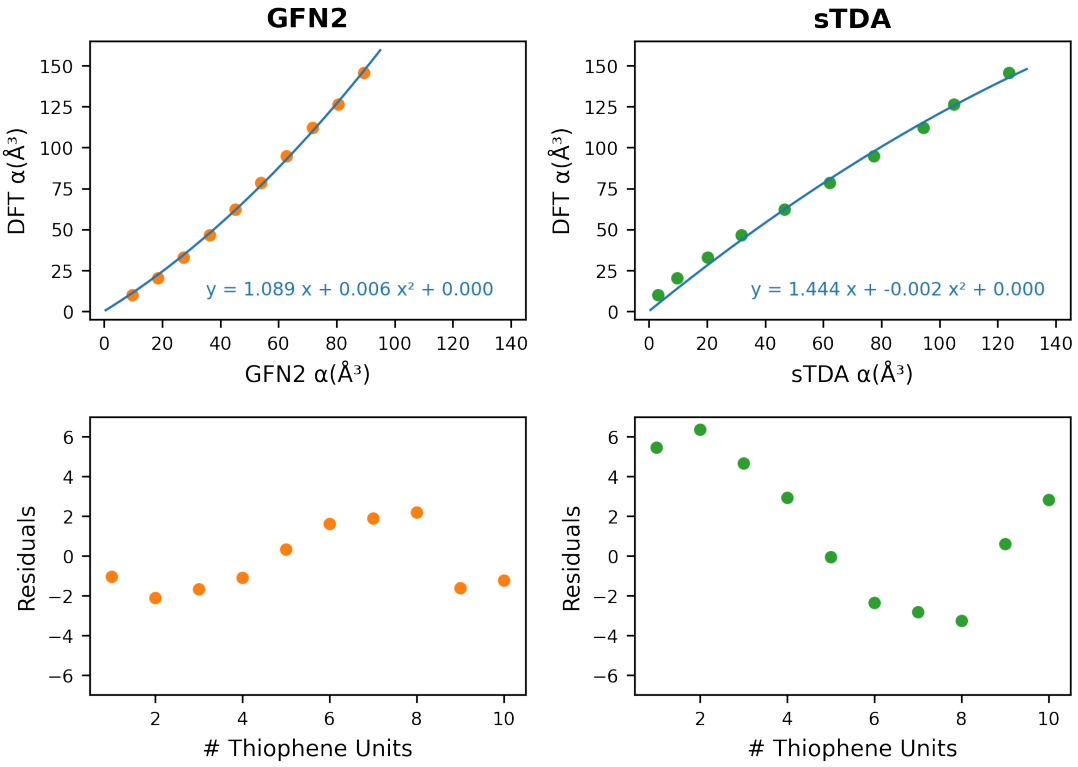

Fig. S10 Quadratic regression with forced zero intercept of isotropic polarizabilities calculated for polythiophene oligomers of increasing length using DFT with the ωB97X functional and the cc-pVTZ basis set corrected to aug-cc-pVTZ accuracy using our linear correction from Figure S13 and GFN2 and sTDA. The residuals from the DFT vs. GFN2 regression are notably smaller than those from the DFT vs. sTDA regression, indicating the latter's larger random error.

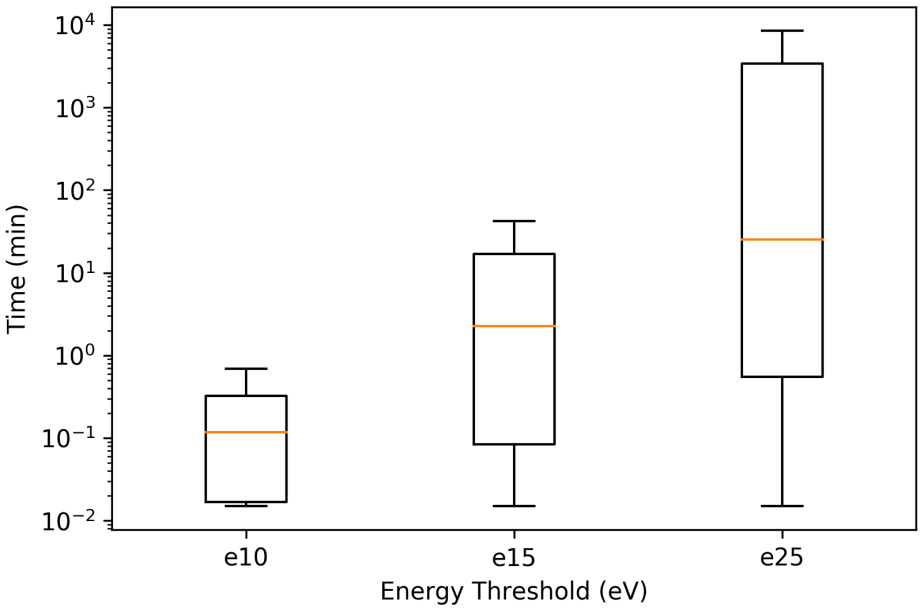

Fig. S11 Timing data for sTDA calculations for the 34 members of the wide range molecule set we were able to successfully run sTDA with a 25 eV energy threshold. Note that the 15 hexamers in this set were run on four cores instead of one, so their raw calculation times were quadrupled and should be considered an estimate. Energy thresholds used for sTDA calculations are noted as their keyword arguments: e10 for 10 eV and e15 for 15 eV.

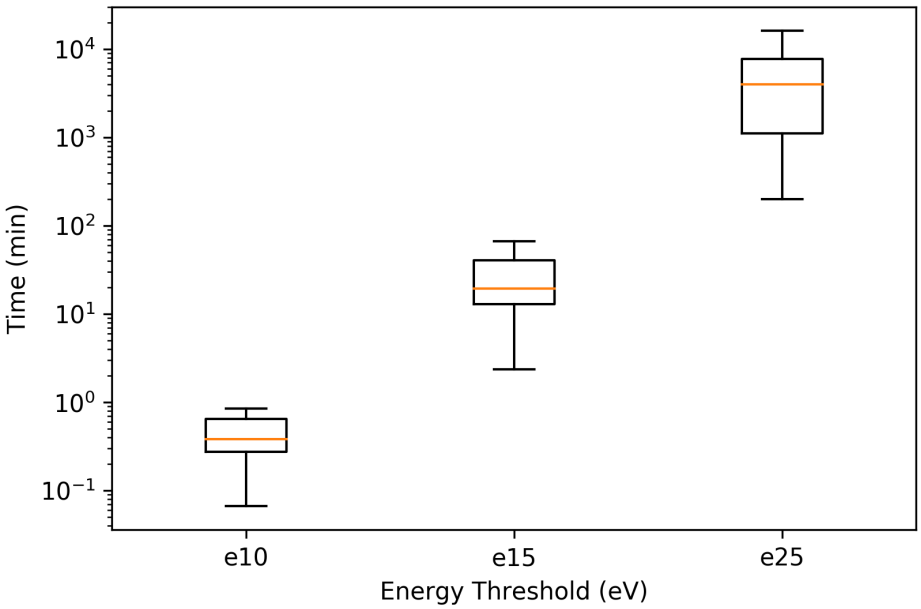

Fig. S12 Timing data for sTDA calculations on subset of 15 hexamers. Note that the 15 hexamers in this set were run on four cores instead of one, so their raw calculation times were quadrupled and should be considered an estimate. Energy thresholds used for sTDA calculations are noted as their keyword arguments: e10 for 10 eV and e15 for 15 eV.

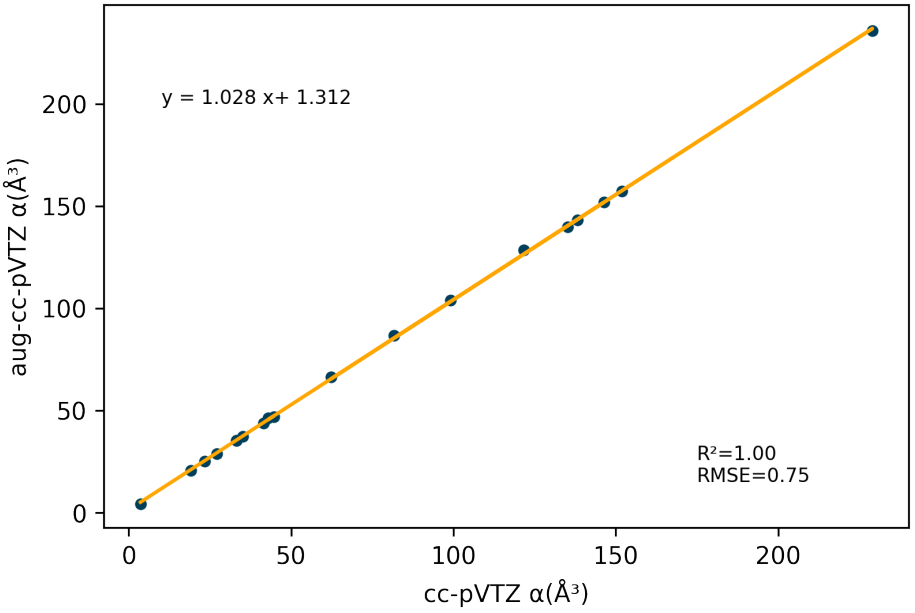

Fig. S13 Linear regression performed on isotropic polarizabilities calculated with cc-pVTZ and aug-pVTZ basis sets shows a suggested linear correction to improve results calculated with the former basis set to the accuracy of the latter.