STRUCTURAL SCIENCE
CRYSTAL ENGINEERING
MATERIALS

Acta Cryst

B

Supporting information for article:

Crystal structure prediction of rigid molecules

Dennis M. Elking, Laszlo Fusti-Molnar and Anthony Nichols

Additional results are given below as Supporting Information. In the first sub-section, results for locally optimizing unit cells are presented. Next, the CPU times for unit cell optimization are given for both rigid and flexible molecule crystal optimization. In sub-section three, results for gas phase dimers are presented. In sub-section four, the three high ranking problem structures mentioned in section 3.5 of the main text are discussed in more detail. A comparison of space group representations P 1 21/c 1 and P 1 21/n 1 is given as a check of the overall method for generating crystal structures in sub-section five. Lastly, various benchmark code tests are given in the last sub-section.

S1. Locally optimized unit cells

The two component crystal with CSD tag 'ACOYOG' is given as a second example. A picture of the experimental structure overlaid with the rigid molecule optimized structure using DMA atomic quadrupoles ($l_{max} = 2$) is given in Figure S1. In Table S1, the experimental cell parameters are compared to the rigid and flexible optimized cell parameters. The unit cell coordinate rmsd's are 0.145 and 0.203 Å for rigid and flexible molecule optimization, respectively. The average CPU times to optimize the unit cell are 7.5 and 15.8 secs for rigid and flexible molecule optimization, respectively.

S2. Unit cell optimization CPU time

The average CPU time to perform a single local optimization of a unit cell is plotted in Figure S2 as a function of average number of atoms in the unit cell for both rigid and flexible molecule optimization using atomic quadrupoles ($l_{max} = 2$). For unit cells with a small number of atoms N, the average CPU times for rigid molecule optimization are 1.5 sec for $N < 40$, 5.7 sec for $40 \leq N < 80$, and 13 sec for $80 \leq N < 120$. In comparison, the average CPU times for flexible molecule optimization are 2.9 sec for $N < 40$, 9.9 sec for $40 \leq N < 80$, and 22 sec for $80 \leq N < 120$. For larger unit cells, the average CPU times for rigid molecule optimization are 142 sec for $280 \leq N < 320$ and 173 sec for $320 \leq N < 360$. In comparison, the average CPU times for flexible molecule optimization are 170 sec for $280 \leq N < 320$ and 248 sec for $320 \leq N < 360$. Thus, the CPU times for flexible molecule optimization are generally larger than rigid molecule optimization by a factor of ~2*x* for small crystals and ~1.5*x* for large crystals.

S3. Dimer energy

The force field is tested further by comparing the geometries and intermolecular energies of 110 gas phase dimers. As an example, the intermolecular energy is plotted as a function of the closest atom-atom distance in Figure S3 for the dimethyl sulfide - water dimer. The force field or MM energy is calculated with DMA quadrupoles ($l_{max} = 2$) while the reference qm or *ab initio* energy is calculated at the CCSD(T)/CBS level. As shown in Figure S3, the force field can approximately

reproduce the location and well depth of the energy minima. When restricted to a one dimensional distance scan (along the qm energy minima), the qm and mm minima are located at 2.4 and 2.5 Å, respectively, with qm and mm dimer energies of -21.2 and -19.4 kJ/mol, respectively. The one dimensional distance scan gives a qualitative illustration of the dimer potential surface.

A more quantitative comparison between the force field and reference *ab initio* dimers may be made by optimizing the dimer geometry while keeping the monomer geometries rigid using the force field energy function. At the optimized equilibrium dimethyl sulfide - water dimer, the force field dimer energy and the closest atom-atom distance are -21.6 kJ/mol and 2.66 Å, respectively. The force field optimized dimer energies and distances can be compared to the *ab initio* reference values of -21.2 kJ/mol and 2.40 Å. Note the optimized force field dimer energy of -21.6 kJ/mol is lower than the dimer energy of -19.4 kJ/mol when restricted to a one dimensional scan. The coordinate rmsd between the optimized force field dimer with the reference *ab initio* dimer structure is 0.212 Å.

As a second example, the intermolecular energy is plotted as a function of the closest atom-atom distance in Figure S4 for the N-methyl methanesulfonamide ammonia dimer. The force field or MM energy is calculated with DMA quadrupoles (l_{max} = 2) while the reference qm or *ab initio* energy is calculated at the CCSD(T)/CBS level. When restricted to a one dimensional distance scan, the qm and mm minima are located at 2.4 and 2.5 Å, respectively, with qm and mm dimer energies of -20.9 and -18.2 kJ/mol, respectively. At the fully relaxed equilibrium dimer geometry, the force field dimer energy and the closest atom-atom distance are -22.6 kJ/mol and 2.0 Å, respectively. The coordinate rmsd of the optimized force field dimer with the reference qm dimer structure is 0.318 Å. Thus, there may be a significant difference between the dimer structure when it is fully optimized as compared to a restricted one dimensional distance scan.

A more extensive comparison between force field and *ab initio* optimized dimers is made on a set of 110 gas phase dimers. The dimer geometries are optimized with force field energy while keeping the monomer geometries rigid. Results for average rmsd in energy (*E*), average rmsd in closest atom-atom distance (*R*), and averaged dimer coordinate rmsd are given in Table S2 for ESP atomic charges, DMA atomic quadrupoles, and DMA atomic octupoles. The dimer coordinate rmsd is greater than 1.0 Å in some cases, which may indicate the force field optimized dimer is in a different local minimum from the *ab initio* optimized dimer. Thus, only dimers in which the dimer coordinate rmsd is less than 1.0 Å are used in the rmsd averages. Out of the 110 dimers, the number of dimers with a coordinate rmsd less than 1.0 Å are 95, 102, and 100 for ESP atomic charges, DMA atomic quadrupoles, and DMA atomic octupoles.

The average rmsd errors in optimized dimer energy are 7.32, 4.35, and 4.68 kJ/mol for ESP atomic charges, DMA atomic quadrupoles, and DMA atomic octupoles. Thus, the errors in DMA atomic quadrupoles and octupoles are substantially smaller than the errors in ESP atomic charges. However, there is an unexpected small increase in average error of 0.3 kJ/mol in going from DMA

atomic quadrupoles to DMA atomic octupoles. This may be due to the fact that a single set of vdW repulsion parameters A_i fit with DMA atomic quadrupoles are also used in both ESP atomic charges and DMA atomic octupoles calculations. The average rmsd errors in optimized dimer closest atom-atom distances and average dimer coordinate rmsd errors follow the expected decreasing trend in order of ESP atomic charges, DMA atomic quadrupoles, and DMA atomic octupoles.

In Figures S5 and S6, histograms of optimized dimer energy and closest atom-atom distance errors are given for DMA atomic quadrupoles. For dimer energy, most of the dimer errors lie between -4 and 4 kJ/mol. For closest atom-atom distance, most of the dimer errors like between -0.25 and 0.25 Å.

S4. High Ranking Problem Structures

Recall the results for ranking the rigid molecule optimized crystals using the correct space group from Figure 4 of the main text. Out of the 2440 crystals, the correct experimental structure has rank 1 for 1780 structures. The correct structure is in the top 10 ranks for 2331 crystals and in the top 20 ranks for 2388 crystals. The three worst ranking structures have force field ranks of 152, 567, and 148 corresponding to CSD tags 'HNAPAC01', 'HPBTAZ01', and 'WOMXUR'. Recall the hydrogen atomic positions in the original experimental crystals structures are optimized in the crystal lattice using the IEFF-MMFF with MMFF atomic charges. The hydrogen atom optimized structures are then taken as the reference crystal. Atomic multipoles and charges are calculated for the hydrogen atom optimized geometries. The crystal local and global optimization calculations are then performed. The original experimental and hydrogen optimized structures are overlaid and shown in Figures S7 – S9.

Note that all three structures involve an –O–H group in which the dihedral angle involving O-H had changed. The calculations are repeated for these three structures using the original experimental structures as the reference crystal, i.e. the hydrogen atomic positions are not optimized. The new rigid molecule ranks for 'HNAPAC01', 'HPBTAZ01', and 'WOMXUR' are 10, 653, and 87. Thus, not optimizing hydrogen positions improved one case significantly but not the other two. It is not yet clear what is causing the large errors in these three cases.

S5. Comparison of Equivalent Space Group Representations

Crystal structures are generated for equivalent representations P 1 21/c 1 and P 1 21/n 1 of space group number 14 for the crystal used as XXII in the recent polymorph blind challenge. The unit cell energies and volumes for the 6 lowest energy structures are given in Tables S3 and S4, respectively. As expected, the optimized energies and volumes in both lists are identical to 8 and 7 significant figures, respectively. The purpose of this calculation is to check the crystal sampling, local optimization, and space group representation for internal consistency.

S6. Methodological Tests

In order to test the basic methodology used in the local optimization of crystal cells, various numerical tests have been performed. In Table S5, analytical and numerical atomic forces on the -O-H oxygen atom are given for the 2 component crystal with CSD tag 'ACOYOG' illustrated in Figure S1. As shown, the analytical and numerical atomic forces agree to 7-8 significant figures. Similarly, the analytical and numerical gradients of energy with respect to cell parameters are given in Table S6 for the crystal with CSD tag 'ACOYOG' illustrated in Figure S1. The analytical and numerical cell parameter gradients agree to 7-8 significant figures. The electrostatic energy of the unit cell with CSD tag 'ACANIL06' is calculated with DMA atomic quadrupoles for various values of the Ewald parameter β and presented in Table S7. As shown, the total Ewald electrostatic energies agree for all values of β and with the result calculated by the TINKER (Shi et al. 2013) software program.

Table S1   Experimental and optimized cell parameters for the two component crystal with CSD tag 'ACOYOG'. DMA atomic quadrupoles ($l_{max} = 2$) are used as the force field.

	a	b	c	α	β	γ
Experiment	7.707	9.146	11.80	87.94	86.55	71.33
Rigid Opt.	7.857	9.135	11.80	91.11	87.29	71.36
Flexible Opt.	7.882	9.145	11.71	91.24	87.28	71.83

Table S2   Average rmsd errors in dimer energy (E), closest atom-atom distance (R), and dimer coordinate rmsd for rigid monomer optimized dimers as compared to *ab initio* reference values.

	ESP ($l_{max} = 0$)	Quadrupoles ($l_{max} = 2$)	Octupoles ($l_{max} = 3$)
E_{rmsd} (kJ/mol)	7.32	4.35	4.68
R_{rmsd} (Å)	0.419	0.319	0.270
average dimer rmsd (Å)	0.329	0.324	0.296
N^* with rmsd < 1.0 Å	95	102	100

[*]Out of 110.

Table S3   Unit cell energies (kJ/mol) of top 6 predicted structures of CSD blind challenge crystal XXII in equivalent space group representations P 1 21/c 1 and P 1 21/n 1 of space group number 14.

Rank (sg = 14)	P 1 21/c 1	P 1 21/n 1
1	-107.36970	-107.36969
2	-106.13814	-106.13813
3	-104.07758	-104.07758
4	-104.06944	-104.06943
5	-102.45716	-102.45716
6	-102.15771	-102.15771

Table S4   Unit cell volumes ($\mathring{A}^3$) of top 6 predicted structures of CSD blind challenge crystal XXII in equivalent space group representations P 1 21/c 1 and P 1 21/n 1 of space group number 14.

Rank (sg = 14)	P 1 21/c 1	P 1 21/n 1
1	245.119882	245.119881
2	247.604410	247.604381
3	250.361805	250.361796
4	250.496683	250.496680
5	250.413976	250.413966
6	245.584252	245.584241

Table S5   Analytical and numerical atomic forces (kJ/mol/Å) on the -O-H oxygen atom for the crystal with CSD tag 'ACOYOG' in Figure S1.

$\mathbf{F}_x$	$\mathbf{F}_y$	$\mathbf{F}_z$	
-16.9188391	8.5437325	3.3707826	Analytical
-16.9188394	8.5437324	3.3707834	Numerical

Table S6   Analytical and numerical gradients of cell parameters for the crystal with CSD tag 'ACOYOG' in Figure S1.

	a	b	c	α	β	γ
Analytical	-72.172730	19.254542	-18.1914560	-4.2089414	-6.740732	0.558974
Numerical	-72.172725	19.254535	-18.1914554	-4.2089410	-6.740729	0.558969

Table S7   Electrostatic energy of the unit cell with CSD tag 'ACANIL06' calculated with DMA atomic quadrupoles for various Ewald parameters β. The reference energy calculated by TINKER is -46.5327 kJ/mol.

β (Å^{-1})	E (kcal/mol)
1.0	-46.533018
0.7	-46.532910
0.5	-46.532903
0.4	-46.532903
0.3	-46.532903
0.2	-46.532902
0.1	-46.532902
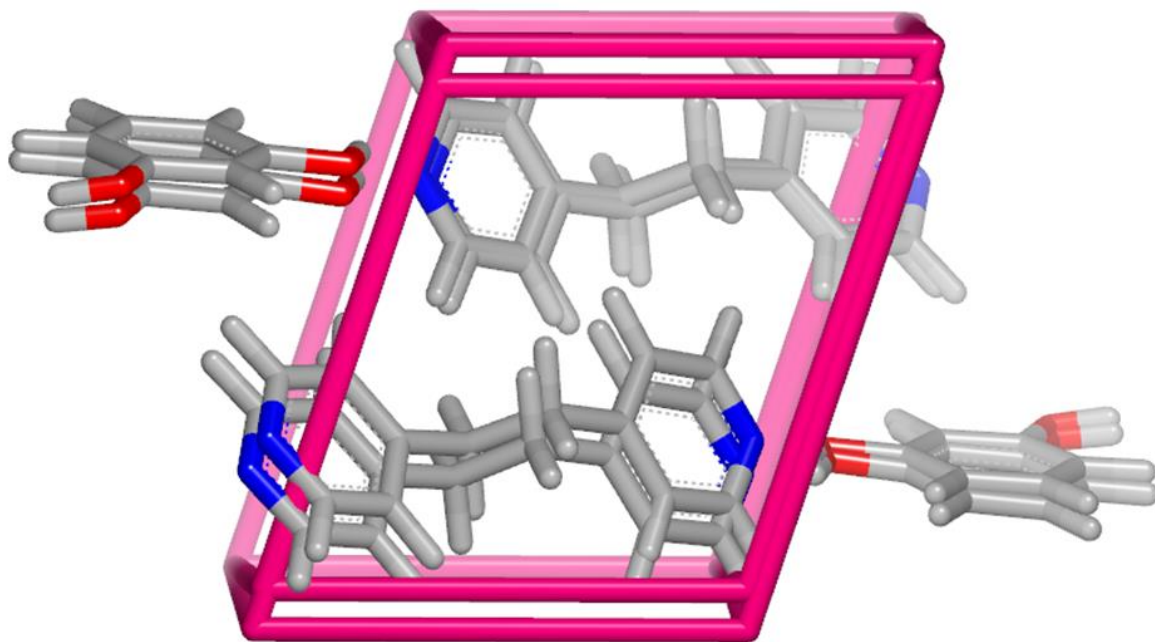
Figure S1 Experimental and optimized two component crystal structure for ACOYOG with space group P-1.
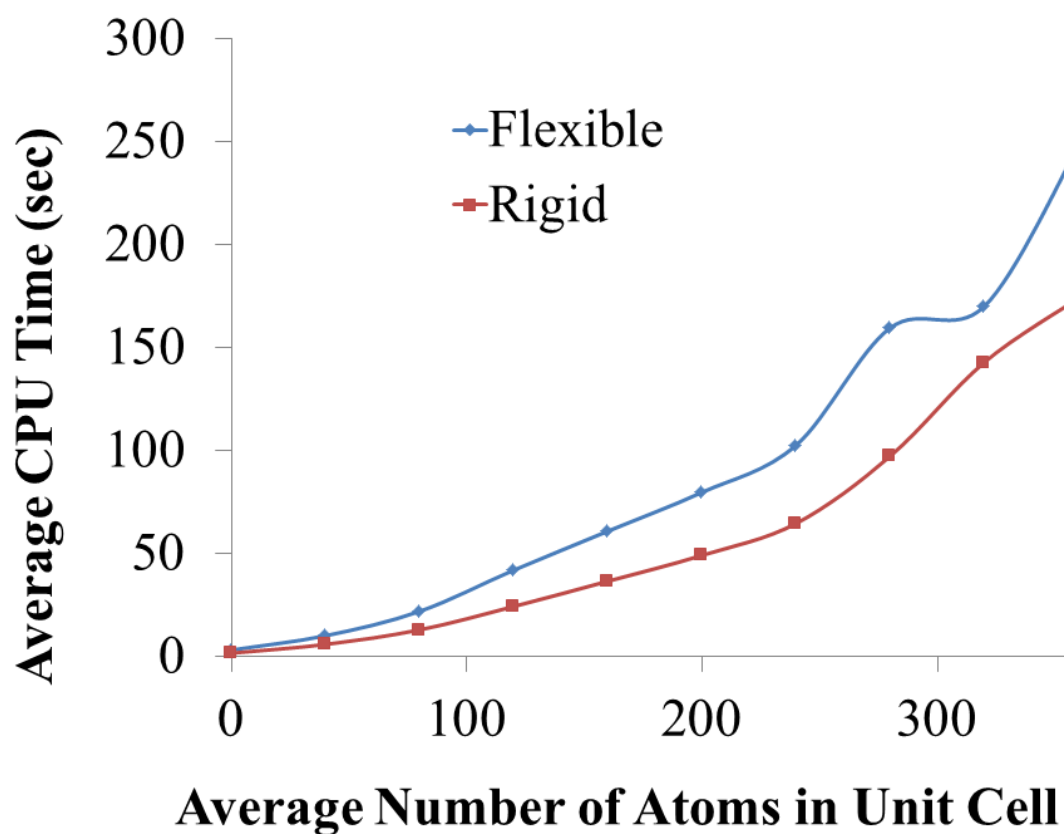
Figure S2 Plot of average CPU time (seconds) to perform a single local optimization of a unit cell as a function of number of atoms in the unit cell. Both flexible and rigid molecule optimization times are plotted for atomic quadrupoles ($l_{max} = 2$).
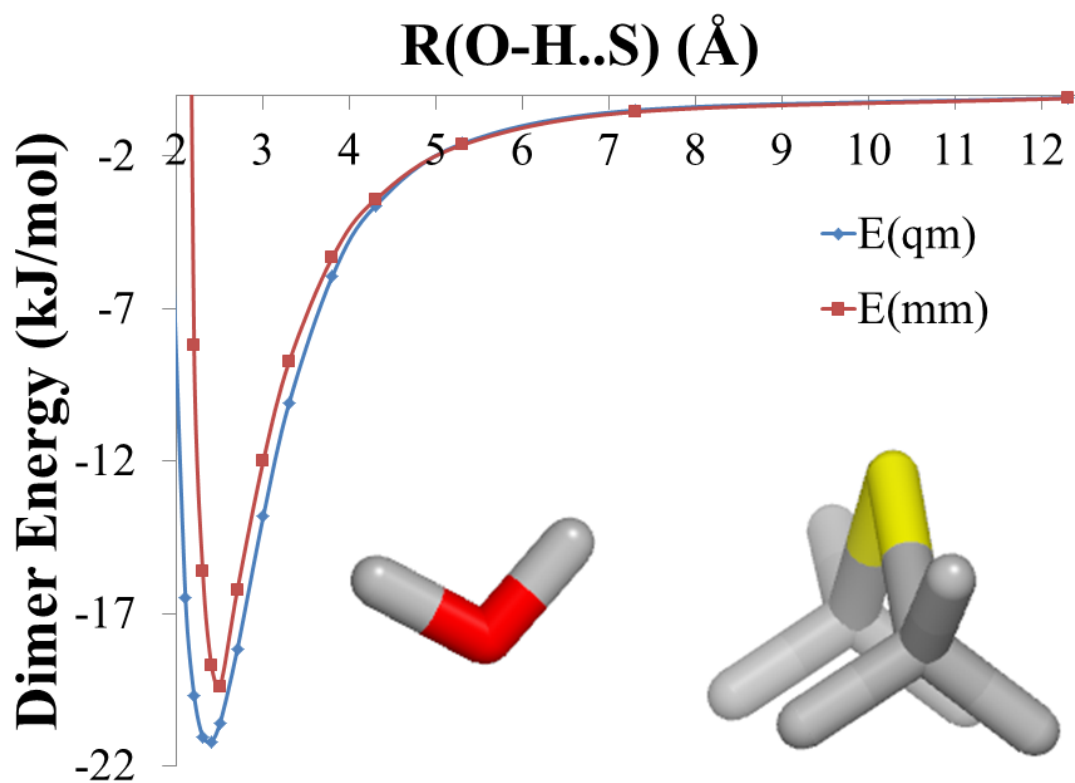
Figure S3 Plot of the intermolecular dimer energy as a function of O-H..S distance for the dimethyl sulfide - water dimer. The MM energy is calculated with DMA quadrupoles (l_{max} = 2), while the QM energy is calculated at the CCSD(T)/CBS level.
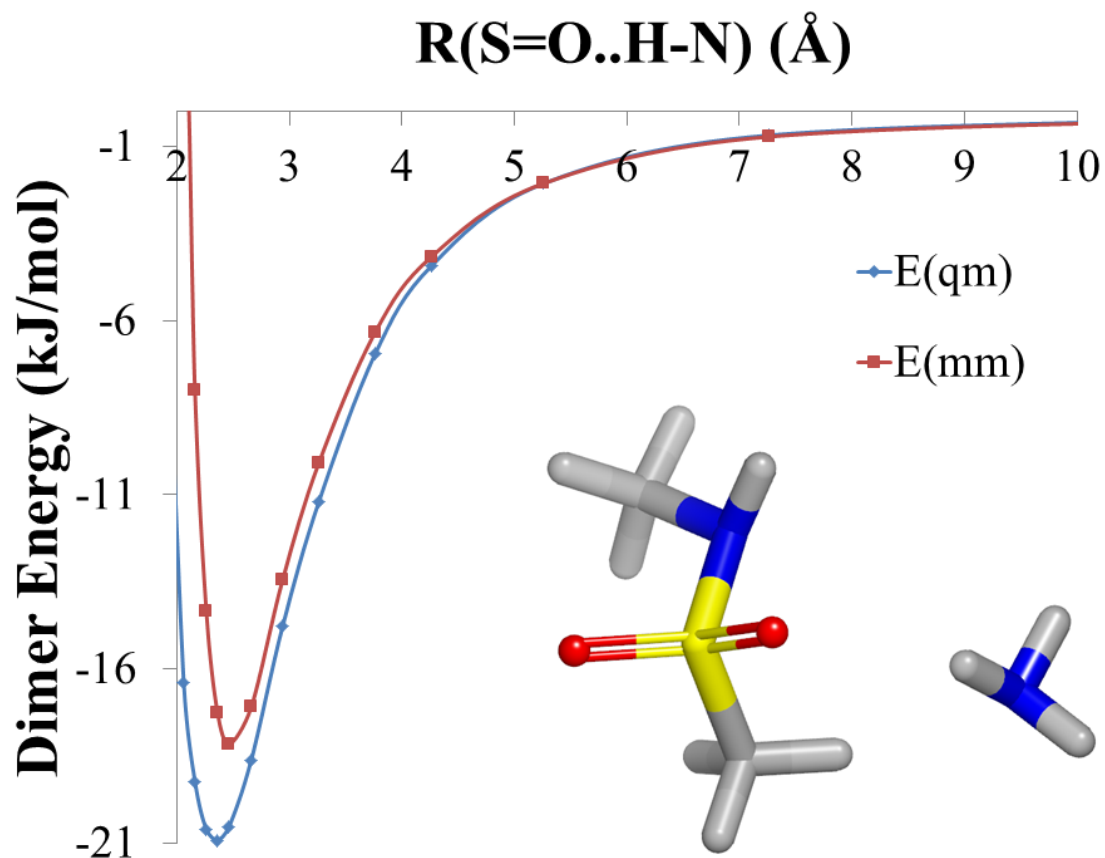
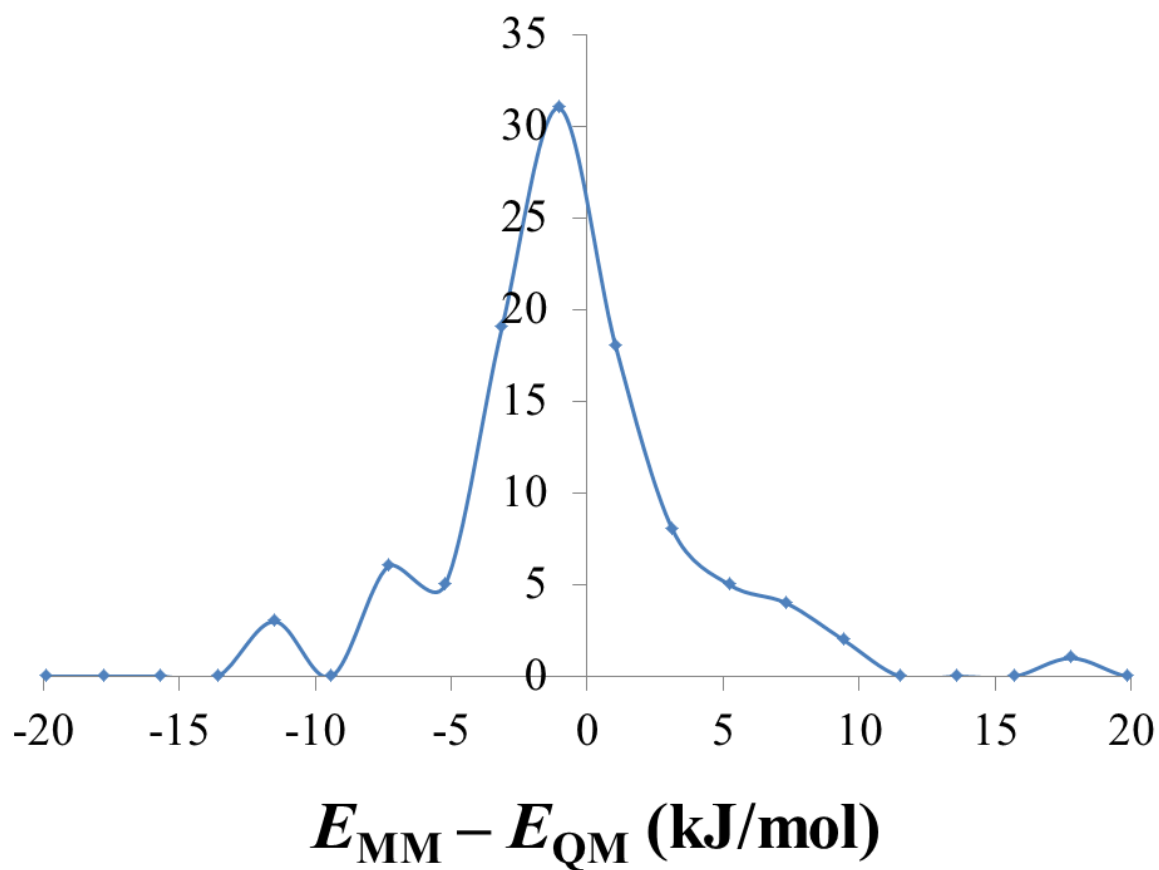**Figure S4** Plot of the intermolecular dimer energy as a function of S=O..H-N distance for the N-methyl methanesulfonamide ammonia dimer. The MM energy is calculated with DMA quadrupoles ($l_{max} = 2$), while the QM energy is calculated at the CCSD(T)/CBS level.

Figure S5 Histogram of the difference between optimized force field (MM) and optimized *ab initio* (QM) dimer energies. The MM energy is calculated with DMA quadrupoles ($l_{max} = 2$), while the QM energy is calculated at the CCSD(T)/CBS level.
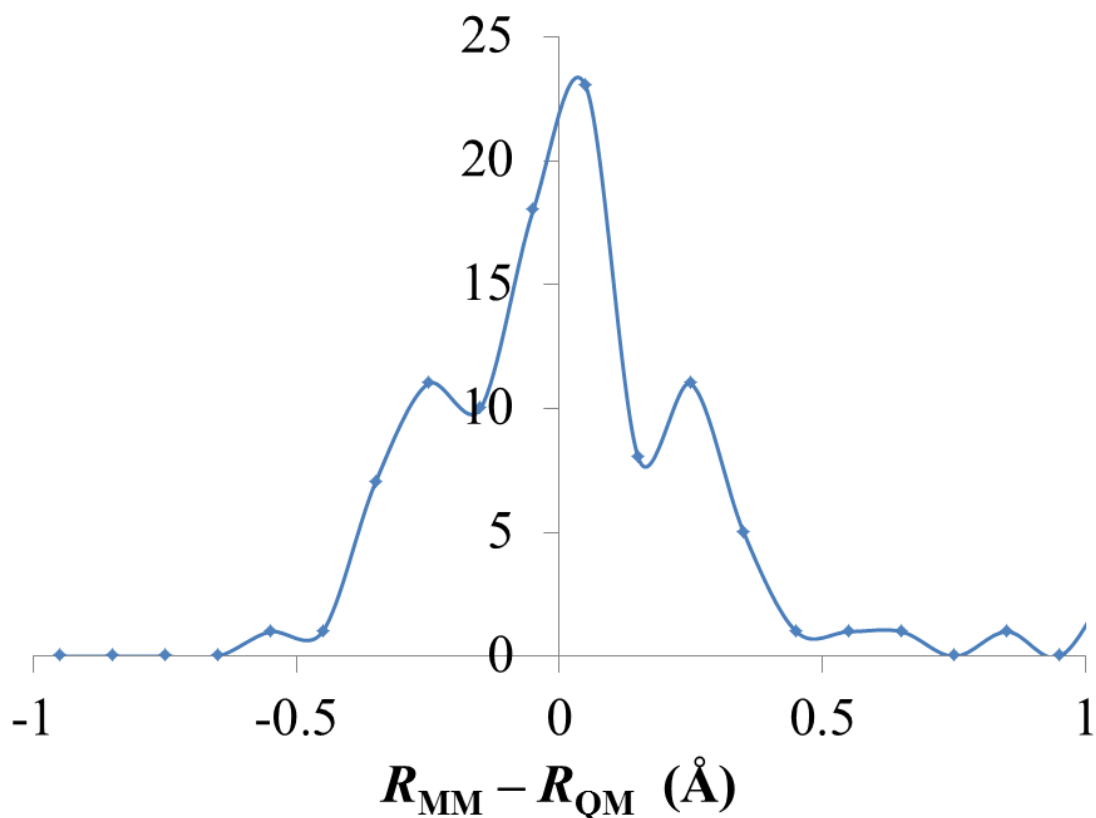
Figure S6 Histogram of the difference between optimized force field (MM) and optimized *ab initio* (QM) closest atom-atom distances. The MM energy is calculated with DMA quadrupoles ($l_{max} = 2$), while the QM energy is calculated at the CCSD(T)/CBS level.
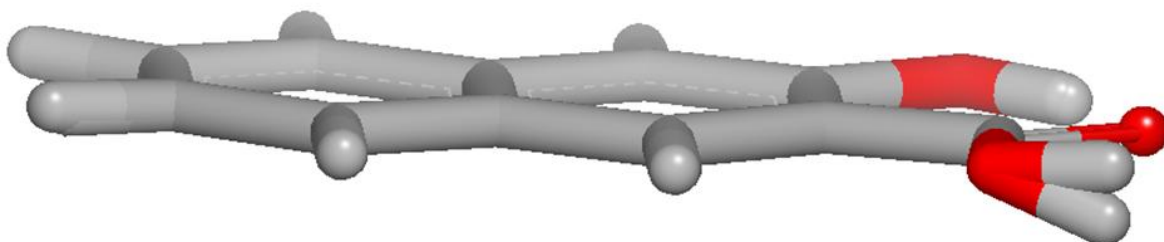

Figure S7 Original experimental and hydrogen optimized structure for HNAPAC01. The O-H is in the plane for the experimental structure.
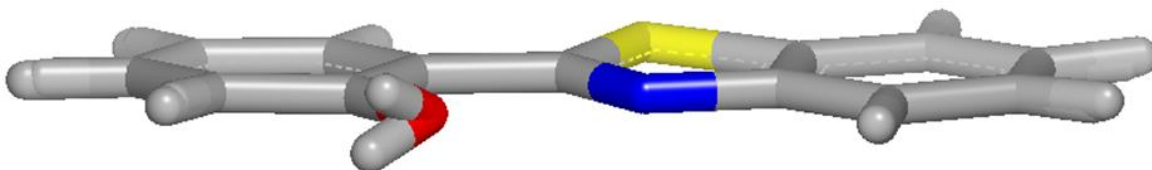

Figure S8 Original experimental and hydrogen optimized structure for HPBTAZ01. The O-H is out of the plane for the experimental structure.
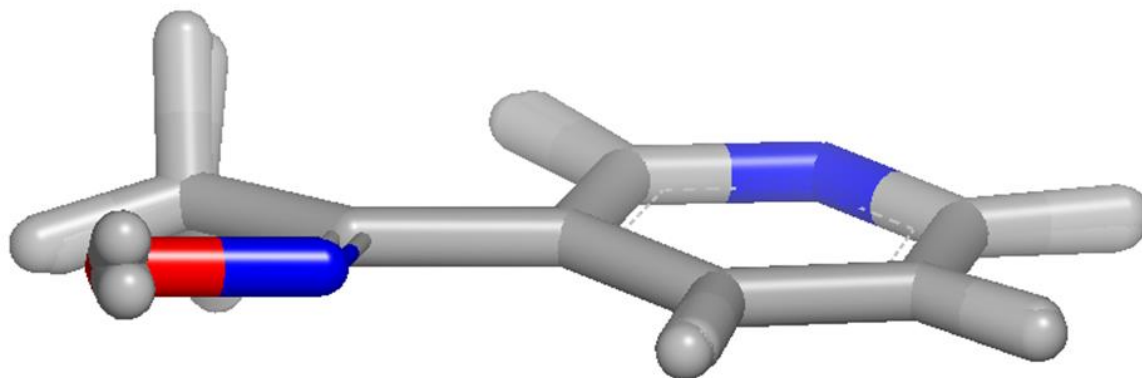
Figure S9   Original experimental and hydrogen optimized structure for WOMXUR. The O-H is in the plane for the experimental structure.