Supplementary Information for: Accurate force fields and methods for modelling organic molecular crystals at finite temperatures

Jonas Nyman, Orla Sheehan Pundyke and Graeme M. Day

1 Intramolecular relaxation energies

The intramolecular relaxation energy is the difference in energy between the in-crystal molecular geometry and charge density, and the *in vacuo* ground state of the molecule. All calculations were performed with Gaussian 09, at the B3LYP/6-311G(d,p) level of theory, with or without a polarizable continuum model (PCM). When using PCM, this energy includes the induction energy. Two calculated relaxation energies are (marginally) negative. This is caused by minute SCF convergence differences and are of no consequence to the results.

Structure	No PCM	PCM3	PCM5
1,4-cyclohexanedione	1.2229	2.6651	3.6287
Acetic acid	0.1601	0.7341	1.0885
Adamantane	0.3710	0.3790	0.3842
Ammonia	0.2647	0.7355	1.0532
Anthracene	0.1353	0.6417	1.0513
Benzene	0.0161	0.2216	0.3765
Cyanamide	0.4108	2.5704	3.9487
Cytosine	0.5588	4.2822	6.8184
Ethylcarbamine	0.9387	1.8617	2.4551
Formamide	0.5746	2.1986	3.2462
Imidazole	0.1975	1.8754	3.0516
Naphtalene	-0.0262	0.3374	0.6262
Oxalic acid α	14.8293	15.8083	16.4410
Oxalic acid β	14.7199	15.6928	16.3203
Pyrazine	-0.0457	0.4768	0.8312
Pyrazole	0.2722	1.2054	1.8329
Triazine	0.0009	0.3298	0.5398
Trioxane	0.3476	0.9903	1.4734
Uracil	0.7730	2.9385	4.3615
Urea	3.1316	4.9236	6.0972
Hexamine	0.1051	0.4757	0.7356
Succinic acid	3.1846	3.9777	4.4818
Carbon dioxide	0.0	0.5790	0.6603

Table S1: Relaxation energies in kJ/mol, calculated without PCM and with PCM using relative permittivities of 3.0 and 5.0, respectively.

2 Unit cell dimensions and atomic RMSD

Otero-de-la-Roza & Johnson[5] includes a table in their supplementary information that compares calculated lattice parameters with the experimental reference values. We reproduce our corresponding results in Table S2, in units of Å. Urea is excluded from comparison since the force fields describe the experimental structure as a thermally averaged structure over two lower symmetry minima.

The columns are as follows: experimental reference values (Ref.), results obtained with the FIT and the W99rev6311P5 force fields. The two rightmost columns (FIT_th and W99rev6311P5_th) are results for explicitly thermally expanded structures with the FIT and W99rev6311P5 force fields.

Structure		Ref.	FIT	W99_pcm5	FIT_th	W99rev6311P5_th
1,4-cyclohexanedione	a	6.650	6.650	6.723	6.718	6.804
	b	6.210	6.390	6.304	6.448	6.384
	c	6.870	6.798	6.857	6.828	6.884
Acetic acid	a	13.151	13.392	13.501	13.43	13.607
	b	3.923	3.874	3.907	3.901	3.942
	c	5.762	5.699	5.749	5.736	5.778
Adamantane	a	6.639	6.578	6.665	6.657	6.787
	c	8.918	8.953	9.018	9.074	9.186
Ammonia	a	5.130	4.523	5.231	4.647	5.495
Anthracene	a	8.414	8.418	8.594	8.519	8.700
	b	5.990	5.936	5.905	5.920	5.905
	c	11.095	11.042	11.119	11.098	11.182
Benzene	a	7.390	6.937	7.653	7.098	7.985
	b	9.420	9.522	8.178	9.522	8.311
	c	6.810	7.152	7.887	7.160	8.154
Cyanamide	a	6.856	6.894	6.430	6.6461	6.564
	b	6.628	6.665	6.836	6.7324	6.971
	c	9.147	9.050	9.656	9.0557	9.704
Cytosine	a	13.044	12.940	12.909	13.040	12.886
	b	9.496	9.369	9.622	9.363	9.686
	c	3.814	3.963	3.942	4.014	4.085
Ethylcarbamate	a	5.051	4.928	5.050	4.971	5.103
	b	7.011	7.040	7.220	7.034	7.259
	c	7.543	7.497	7.532	7.646	7.695
Formamide	a	3.604	3.679	3.794	3.749	3.872
	b	9.041	9.254	9.499	9.306	9.564
	c	6.994	6.509	6.652	6.535	6.694

Structure		Ref.	FIT	W99rev6311P5	FIT_th	W99rev6311P5_th
Imidazole	a	7.582	7.555	7.751	7.650	7.913
	b	5.371	5.374	5.282	5.442	5.313
	c	9.790	9.620	9.837	9.638	9.924
Naphtalene	a	8.085	8.085	8.279	8.150	8.308
	b	5.938	5.837	5.682	5.841	5.716
	c	8.633	8.462	8.238	8.508	8.310
Oxalic acid α	a	6.548	6.664	6.662	6.844	6.839
	b	7.844	7.926	8.098	8.028	8.249
	c	6.086	6.178	6.190	6.232	6.279
Oxalic acid β	a	5.330	5.391	5.447	5.408	5.479
	b	6.015	6.578	6.391	6.665	6.513
	c	5.436	4.671	5.008	4.854	5.285
Pyrazine	a	9.325	9.477	9.455	9.529	9.524
	b	5.850	5.658	5.589	5.662	5.564
	c	3.733	3.753	3.831	3.898	4.034
Pyrazole	a	8.190	7.708	7.862	7.828	8.067
	b	12.588	13.063	13.234	13.234	13.346
	c	6.773	7.164	7.171	7.446	7.516
Triazin	a	9.647	9.603	9.605	9.828	9.841
	c	7.281	6.912	6.912	7.369	9.841
Trioxane	a	9.320	9.373	9.388	9.444	9.468
	c	8.196	8.085	8.181	8.162	8.264
Uracil	a	11.938	11.849	11.957	11.976	12.104
	b	12.376	12.449	12.587	12.478	12.649
	c	3.655	3.809	3.792	3.916	3.908
Hexamine	a	7.028	6.939	6.946	6.982	6.991
Succinic acid	a	5.464	5.574	5.622	5.637	5.719
	b	8.766	8.342	8.490	8.439	8.590
	c	5.004	5.093	5.146	5.176	5.249
Carbon dioxide	a	5.624	5.706	5.445	-	-
MAE lengths		-	0.170	0.224	0.188	0.337
MA%E lengths		-	2.497	3.060	2.819	3.700

Table S2: Calculated lattice parameters and experimental reference values in units of Å.

To compare molecular geometries and crystal packing, we use COMPACK to calculate a root mean square deviation in atomic coordinates for a cluster of 20 molecules,[2] as implemented in CCDC's program MERCURY. In Table S3 we present $RMSD_{20}$ values for the FIT, W99rev6311P3 and W99rev6311P5 force fields as well as thermally expanded structures calculated with the thermal pressure method in the W99rev6311P5 force field.

Structure	FIT	W99rev6311P3	W99rev6311P5	W99rev6311P5_th
1,4-cyclohexanedione	0.144	0.132	0.128	0.179
Acetic acid	0.186	0.096	0.092	0.098
Adamantane	0.071	0.065	0.065	0.188
Ammonia	0.499	0.105	0.103	0.325
Anthracene	0.210	0.272	0.280	0.340
Benzene	0.393	0.847*	0.862*	0.954*
Cyanamide	0.174	0.311	0.312	0.313
Cytosine	0.272	0.177	0.189	0.329
Ethylcarbamate	0.335	0.375	0.370	0.318
Formamide	0.447	0.447	0.449	0.487
Imidazole	0.107	0.179	0.179	0.265
Naphtalene	0.404	0.842	0.865	0.843
Oxalic acid α	0.094	0.137	0.142	0.235
Oxalic acid β	0.527	0.347	0.338	0.284
Pyrazine	0.137	0.201	0.207	0.387
Pyrazole	0.315	0.291	0.296	0.414
Triazin	0.178	0.178	0.178	0.120
Trioxane	0.057	0.048	0.049	0.085
Uracil	0.314	0.243	0.231	0.264
Succinic acid	0.198	0.189	0.191	0.250
Hexamine	0.013	0.006	0.008	0.036
Carbon dioxide	0.070	0.146	0.148	0.197
Average $RMSD_{20}$	0.234	0.256	0.258	0.314

Table S3: Calculated root mean square deviations in atomic coordinates in a cluster of 20 overlayed molecules. *20 molecules could not be overlayed, the reported numbers are for 19 overlayed molecules.

3 Co-prime splitting of linear supercells.

In order to sample **k**-points in the Brillouin zone, we construct supercells of the unit cell. The crystal's primitive unit cell is expanded into linear supercells expanded along **a**, **b** or **c** only. The expansion along each direction is such that the distance between sampled **k**-points is strictly smaller than some target value. In our experience 0.3 – 0.1 Å^{-1} is required for convergence.

The long supercells are then split into several shorter cells with mutually co-prime expansion coefficients according to the scheme in Table S4.

The co-prime splitting used in this article differs slightly from that used in our previous paper[4]. We have changed the co-prime splitting scheme to make the convergence less erratic and more monotonic. This is achieved by making sure that as the target **k**-point distance decreases, the number of sampled **k**-points increase strictly.

Lattice dynamics calculations are then performed with each supercell using DMACRYS.

The co-prime splitting used in this work is as follows. If the supercell expansion coefficient $s < 6$, the lattice dynamics calculation is performed on the supercell as is. For $s \geqslant 6$ the supercell (1x1xs) is split into n smaller supercells (1x1xk, 1x1xℓ, 1x1xm ...) such that $k, \ell, m...$ are all mutually co-prime and $k+\ell+m-n \geqslant s$. This ensures that at least s **unique** **k**-points are sampled. The phonons for $\mathbf{k} = \mathbf{0}$ will be calculated in each supercell, and we only include these phonons from one of the split supercells. The long linear supercells are split into 2, 3 or 4 co-prime supercells according to the scheme in Table S4.

Table S4: Co-prime supercell expansion coefficients.

2	→	2
3	→	3
4	→	4
5	→	5
6	→	3, 4
7	→	3, 5
8	→	4, 5
9	→	3, 4, 5
10	→	3, 4, 5
11	→	3, 4, 7
12	→	3, 4, 7
13	→	3, 5, 7
14	→	4, 5, 7
15	→	5, 6, 7
16	→	3, 4, 5, 7
17	→	5, 7, 8
18	→	5, 7, 7
19	→	3, 5, 7, 8
20	→	3, 5, 7, 8

4 Calculating the Debye frequency

The Debye frequency is obtained from the velocity of sound through the crystal, since the group velocity corresponds to $\partial\omega/\partial\mathbf{k}$. In the Debye approximation, where dispersion is linear near the Brillouin zone centre, the group and phase velocities are equal. Since the velocity of sound in organic crystals is anisotropic, we calculate the velocity in 13 different directions and average over these. These directions are vectors pointing from Γ to the Brillouin zone boundary in the direction towards the centre of 13 neighbouring Brilluoin zones. The first Brillouin zone has 26 adjacent zones, but since the velocity of sound is equal for directions $\mathbf{k}$ and $-\mathbf{k}$, only 13 need to be sampled. The 13 principal component vectors pointing to the neighbouring cells are:

$$
\begin{gathered}
(1/2, 0, 0) \\
(0, 1/2, 0) \\
(0, 0, 1/2) \\
(1/2, 1/2, 0) \\
(1/2, 0, 1/2) \\
(0, 1/2, 1/2) \\
(1/2, 1/2, 1/2) \\
(1/2, -1/2, 0) \\
(1/2, 0, -1/2) \\
(0, 1/2, -1/2) \\
(1/2, 1/2, -1/2) \\
(1/2, -1/2, 1/2) \\
(-1/2, 1/2, 1/2)
\end{gathered}
$$

We then calculate the direction cosines $\hat{\mathbf{k}}_m$ of these vectors with respect to the reciprocal axis system. They are calculated with respect to the reciprocal axes of a supercell with the expansion coefficents being those of the longest linear supercell in each direction. These 13 vectors span the Debye ellipsoid. The first three will coincide with the nearest sampled k-vectors.

5 Derivation of thermodynamic expressions

For molecules with six degrees of freedom, there are $n = 6ZN_{\mathrm{k}} - 3$ non-zero phonon frequencies ω_i, where Z is the number of molecules per unit cell and N_{k} the number of sampled unique **k**-points.

The phonon density of states is approximated with a kernel density estimate. The phonon density of states is normalised to $6Z$. The Boltzmann constant is denoted k_B and $\hbar$ is the reduced Planck's constant. The equations below have been derived from the works of Bonadeo[1] and Day[3].

5.1 Lattice vibrational energy

The total lattice vibrational energy is a sum of the zero-point energy and optical contributions. To these are added Debye contributions to each of these terms.

Per unit cell, we have:

$$F_{\mathrm{vib}}(T) = ZPE + F_{\mathrm{optic}}(T) + F_{\mathrm{acoustic}}(T) + ZPE_{\mathrm{Debye}} \tag{1}$$

The zero point energy is:

$$ZPE = \frac{1}{2}\Big(\frac{n}{n+3}\Big) \int_0^\infty \hbar\omega g(\omega)\ d\omega \tag{2}$$

The optical lattice vibrational energy is:

$$F_{\mathrm{optic}}(T) = k_B T\Big(\frac{n}{n+3}\Big) \int_0^\infty g(\omega) \ln\left(1 - \exp\left(\frac{-\hbar\omega}{k_B T}\right)\right)\ d\omega \tag{3}$$

The Debye contribution is:

$$F_{\mathrm{acoustic}}(T) = \frac{3k_B T}{N_k} \ln\left(1 - \exp\left(\frac{-\hbar\omega_D}{k_B T}\right)\right) - \frac{k_B T}{N_{\mathrm{k}}} \mathrm{D}\Big(\frac{\hbar\omega_D}{k_B T}\Big) \tag{4}$$

Here $\mathrm{D}(x)$ is the Debye function

$$\mathrm{D(x)} = \frac{3}{\mathrm{x}^3} \int_0^{\mathrm{x}} \frac{\mathrm{t}^3}{\exp(\mathrm{t}) - 1}\ \mathrm{dt}. \tag{5}$$

The Debye contribution to the zero point energy is:

$$ZPE_{\mathrm{Debye}} = \frac{9\hbar\omega_D}{8N_{\mathrm{k}}} \tag{6}$$

The Debye integral has no known analytical solution. We calculate all integrals numerically using SciPy.

5.2 Entropy

The lattice vibrational entropy is $S_{\text{vib}}(T) = -\partial F_{\text{vib}}(T)/\partial T$. The zero point energy is temperature independent so:

$$S_{\text{vib}}(T) = S_{\text{optic}}(T) + S_{\text{acoustic}}(T) = -\frac{\partial F_{\text{optic}}(T)}{\partial T} - \frac{\partial F_{\text{acoustic}}(T)}{\partial T} \quad (7)$$

$$S_{\text{optic}}(T) = -k_B\Big(\frac{n}{n+3}\Big)\int_0^\infty g(\omega)\ln\Big(1-\exp\Big(\frac{-\hbar\omega}{k_BT}\Big)\Big)\,d\omega \quad (8)$$

$$-\frac{1}{T}\Big(\frac{n}{n+3}\Big)\int_0^\infty \hbar\omega g(\omega)\left(\frac{\exp\left(\frac{-\hbar\omega}{k_BT}\right)}{1-\exp\left(\frac{-\hbar\omega}{k_BT}\right)}\right)d\omega \quad (9)$$

$$S_{\text{acoustic}}(T) = -\frac{3k_B}{N_{\text{k}}}\ln\Big(1-\exp\Big(\frac{-\hbar\omega_D}{k_BT}\Big)\Big) \quad (10)$$

$$+\frac{3}{N_{\text{k}}T}\hbar\omega_D\left(\frac{\exp\left(\frac{-\hbar\omega_D}{k_BT}\right)}{1-\exp\left(\frac{-\hbar\omega_D}{k_BT}\right)}\right) \quad (11)$$

$$+\frac{k_B}{N_{\text{k}}}\text{D}\Big(\frac{\hbar\omega_D}{k_BT}\Big) \quad (12)$$

$$-\frac{\hbar\omega_D}{N_{\text{k}}T}\text{D}'\Big(\frac{\hbar\omega_D}{k_BT}\Big) \quad (13)$$

Where $\text{D}'(x)$ is

$$\text{D}' = \frac{3}{\exp(x)-1} - \frac{1}{x^4}\int_0^x \frac{t^3}{\exp(t)-1}\,dt \quad (14)$$

5.3 Specific heat capacity

The heat capacity C is a sum of optic contributions and a Debye contribution from acoustic vibrations. The heat capacity for constant volume C_v is obtained in the harmonic approximation. The quasi-harmonic approximation allows us to calculate C_p with the same expression.

$$C(T) = C_{\text{optic}}(T) + C_{\text{acoustic}}(T) \quad (15)$$

$$C_{\text{optic}}(T) = k_B\Big(\frac{n}{n+3}\Big)\int_0^\infty \Big(\frac{\hbar\omega}{k_BT}\Big)^2 \frac{\exp\Big(\frac{\hbar\omega}{k_BT}\Big)g(\omega)}{\Big(\exp\Big(\frac{\hbar\omega}{k_BT}\Big)-1\Big)^2}\,d\omega \qquad (16)$$

$$C_{\text{acoustic}}(T) = \frac{12k_B}{N_{\text{k}}}\text{D}\Big(\frac{\hbar\omega_D}{k_BT}\Big) - \frac{9\hbar\omega_D}{N_{\text{k}}T}\Big(\exp\Big(\frac{\hbar\omega_D}{k_BT}\Big)-1\Big)^{-1} \qquad (17)$$

6 Convergence tests

The convergence of the vibrational energy with respect to target **k**-point distance was studied, as this is the main difficulty in the free energy calculations. The entropy, being the derivative of the vibrational energy, converges at the same rate. The zero point energy converges more easily. Here we present data for the acetic acid, cyanamide, benzene and urea crystal structures at 300 K.

The expression for the vibrational energy is;

$$F_{\text{vib}}(T) = \frac{1}{2}\sum_{i,\mathbf{k}}\hbar\omega_{i,\mathbf{k}} + k_BT\sum_{i,\mathbf{k}}\ln\Big(1-\exp\Big(\frac{-\hbar\omega_{i,\mathbf{k}}}{k_BT}\Big)\Big). \qquad (18)$$

but this 'Neat' expression converges poorly with respect to the number of sampled **k**-points, see Fig S1 – S2.

To improve the convergence and reduce the dependence of F_{vib} on the choice of **k**-points, we replace the discrete phonons with the kernel density 'KDE'. Adding the acoustic contribution 'Debye' is used to account for low-frequency phonons which are difficult to sample with the supercell method. When both the KDE and the Debye methods are used, we obtain the expression for the vibrational energy given in the paper. The convergence for this expression is much smoother and almost independent of the choice of supercells/**k**-points.

The W99rev6311 force field was used in these convergence calculations. A kernel bandwidth of $0.05\text{std}(\omega_{\text{i}})$ was used in the kernel density estimate of the phonon density of states, where $\text{std}(\omega_{\text{i}})$ denotes the standard deviation of the phonon frequencies.

The kernel bandwidth has a negligible influence on the calculated vibrational energy in the range 0.01-0.12 standard deviations, see Fig S3. There is however a smooth trend for all four tested structures; increasing bandwidth leads to larger absolute vibrational energies. This trend is shown for benzene in Fig S4.

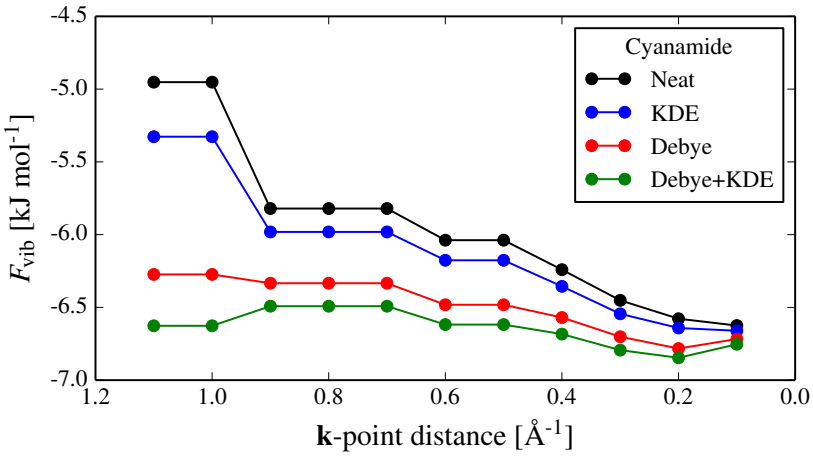

Figure S1: Convergence of the lattice vibrational energy for crystalline cyanamide. The minimum and maximum number of sampled **k**-points are 1 and 25.

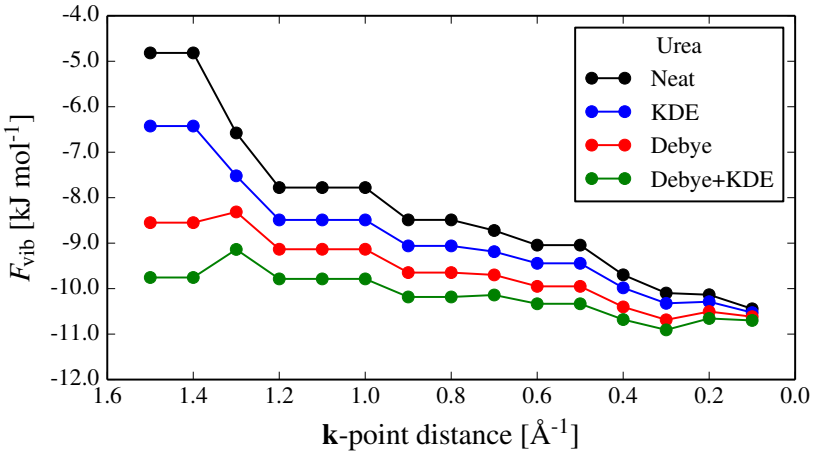

Figure S2: Convergence of the lattice vibrational energy for crystalline urea. The minimum and maximum number of sampled **k**-points are 1 and 26.

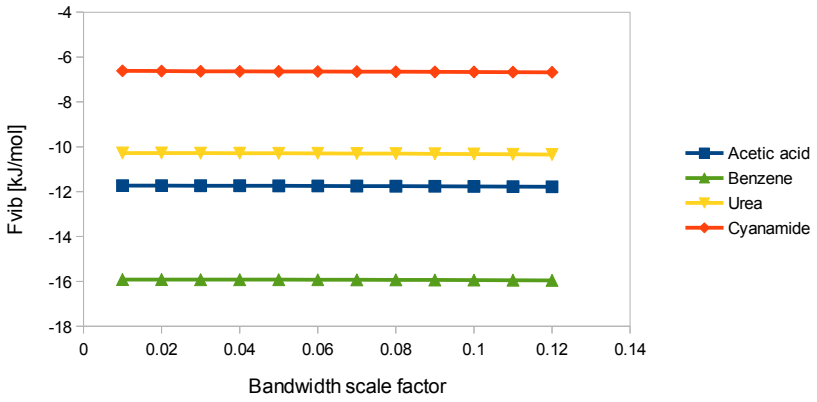

Figure S3: Vibrational energy as a function of the kernel bandwidth factor. A **k**-point target distance of 0.3 Å^{-1} was used.

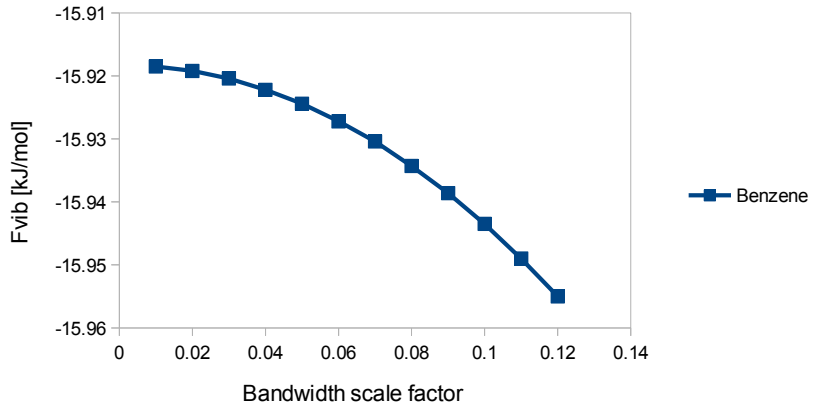

Figure S4: Vibrational energy as a function of the kernel bandwidth factor for crystalline benzene, shown on a narrow energy scale. A **k**-point target distance of 0.3 Å^{-1} was used.

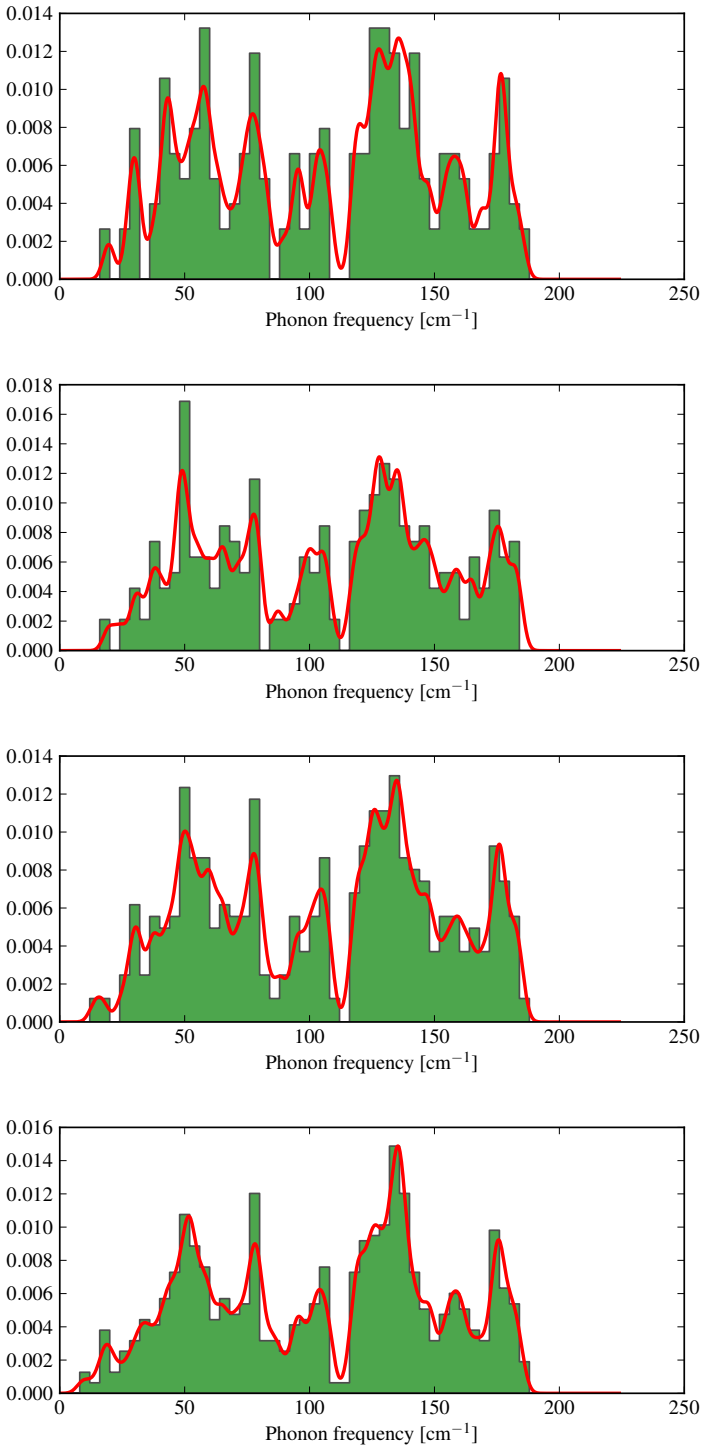

Figure S5: Phonon density of states of the acetic acid crystal plotted as histograms (green) and kernel density of states (red curve) for different **k**-point samplings. The **k**-point distances are 0.4, 0.3, 0.2 and 0.1 Å^{-1}. The kernel density bandwidth is std(ω_i)/20.

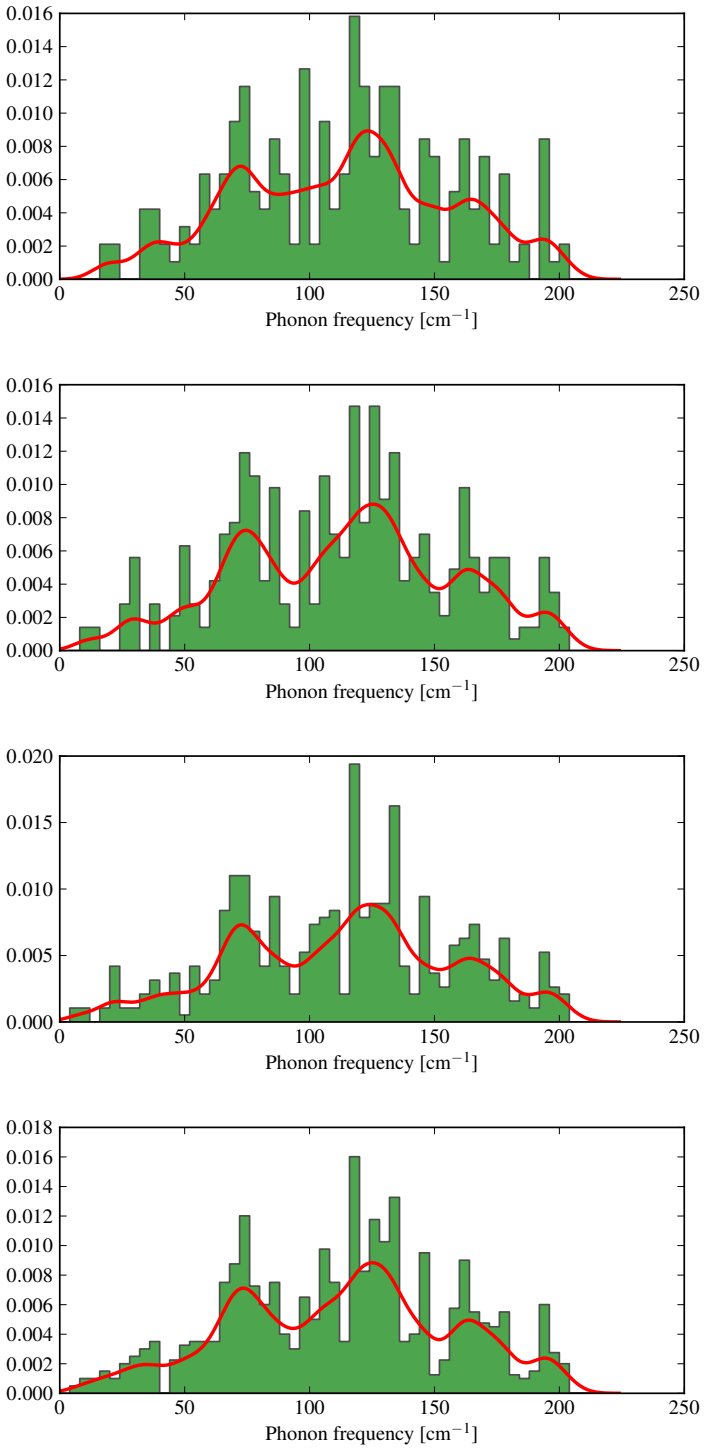

Figure S6: Phonon density of states of cyanamide plotted as histograms (green) and kernel density of states (red curve) for different **k**-point samplings. The **k**-point distances are 0.4, 0.3, 0.2 and 0.1 Å^{-1}. The kernel density bandwidth is std(ω_i)/20.

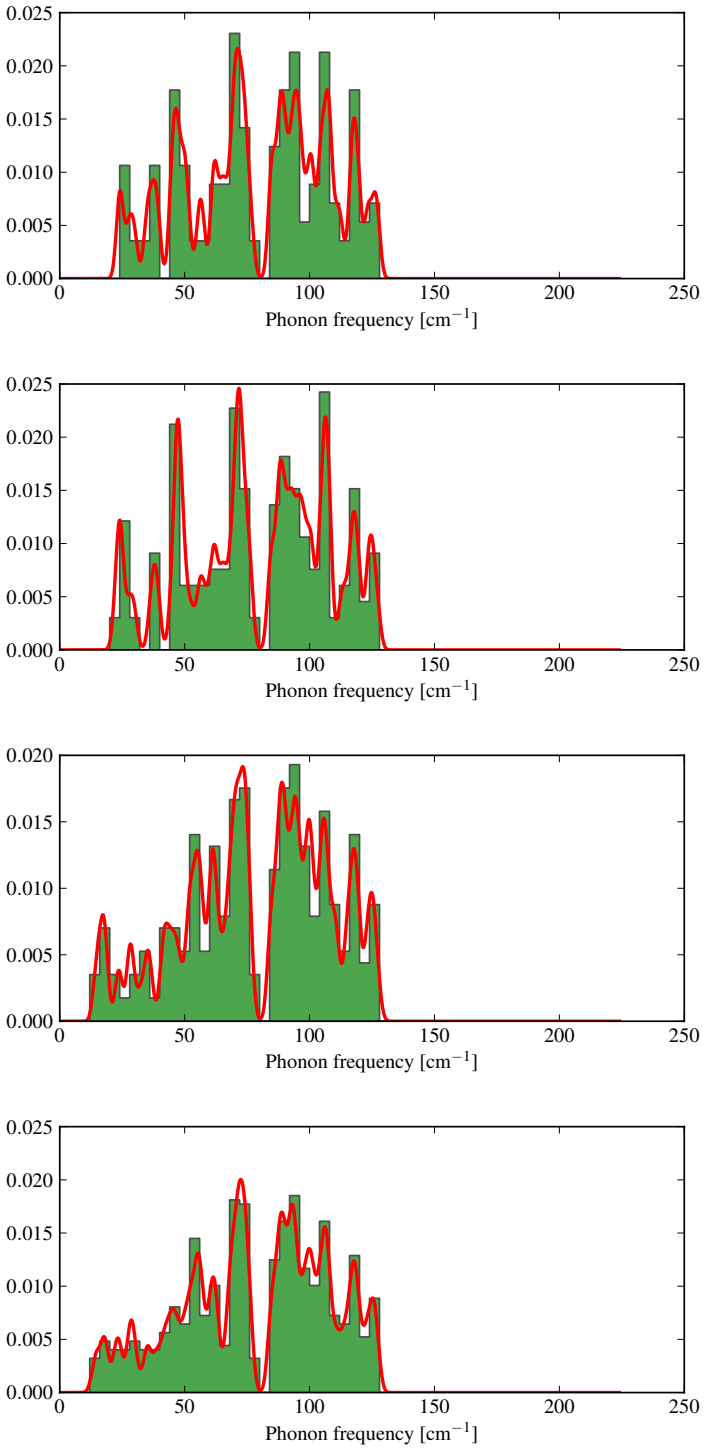

Figure S7: Phonon density of states of benzene plotted as histograms (green) and kernel density of states (red curve) for different **k**-point samplings. The **k**-point distances are 0.4, 0.3, 0.2 and 0.1 Å^{-1}. The kernel density bandwidth is std(ω_i)/20.

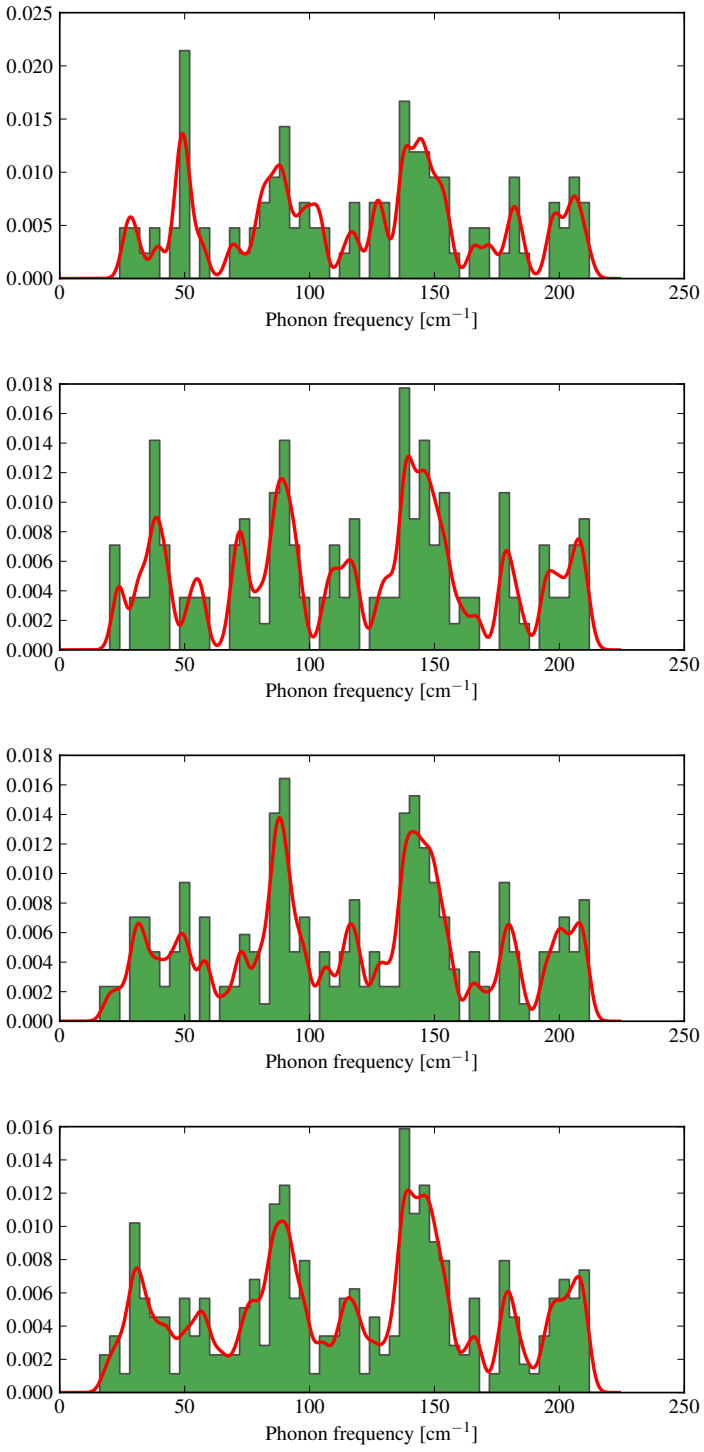

Figure S8: Phonon density of states of urea plotted as histograms (green) and kernel density of states (red curve) for different **k**-point samplings. The **k**-point distances are 0.4, 0.3, 0.2 and 0.1 Å^{-1}. The kernel density bandwidth is std(ω_i)/20.

References

[1] H Bonadeo, E D'alessio, E Halac, and E Burgos. Lattice dynamics, thermodynamic functions, and phase transitions of p-dichloro-and 1, 2, 4, 5-tetrachlorobenzene. *The Journal of Chemical Physics*, 68(10):4714–4721, 1978.

[2] James Alexander Chisholm and Sam Motherwell. COMPACK: a program for identifying crystal structure similarity using distances. *Journal of applied crystallography*, 38(1):228–231, 2005.

[3] Graeme M. Day. *Lattice Dynamical Studies of Molecular Crystals with Application to Polymorphism and Structure Prediction.* PhD thesis, University College London, 2003.

[4] Jonas Nyman and Graeme M. Day. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm*, 17:5154–5165, 2015.

[5] A Otero-de-la-Roza and Erin R Johnson. A benchmark for non-covalent interactions in solids. *The Journal of chemical physics*, 137(5):054103, 2012.