

Supplementary Information for: Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape

Hugh P. G. Thompson and Graeme M. Day

1 Full details of crystal structure and molecular structure optimisations

Prior to lattice energy minimisation, any missing hydrogen atoms in the reported crystal structures were placed at standardised positions and, in a few cases where their positions were poorly determined in the reported crystal structure, hydrogen atoms were removed and replaced at standard positions.

All DFT-D calculations were performed using the CRYSTAL09^{1,2} software, with the B3LYP hybrid functional^{3,4}, 6-31G* basis set, and the rescaled empirical dispersion correction suggested by Civalleri *et al.*⁵ A fine sampling of reciprocal space (shrinking factors of 6 for the Monkhorst Pack and Gilat nets) was used to ensure convergence of the total energy. CRYSTAL09 uses Lebedev and Gauss-Legendre quadrature for the angular and radial numerical integration in calculating the exchange-correlation. We used the pruned (75,434) grid, containing 75 radial points and a maximum of 434 angular points. All atomic positions were relaxed during lattice energy minimisation, with unit cell dimensions constrained at experimentally determined values. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-5} , 1.2×10^{-4} , 4.5×10^{-5} and 1.8×10^{-4} a.u., respectively.

All single molecule calculations were also performed in CRYSTAL09, using the same functional, basis set, DFT integration grid and optimisation thresholds as the periodic lattice energy minimisations. We estimate that the maximum errors in relative molecular energies due to optimisation convergence tolerances and our choice of DFT integration grids are approximately 0.7 kJ mol^{-1} (see below).

2 DFT Calculation settings

Several tests were performed to examine the influence of the energy minimisation method and DFT integration grid on the results of molecular optimisations. As a set of test structures, we chose the independent molecular geometries of two of the packing polymorphic systems studied: the 9 geometries of MABZNA and the 2 geometries of GALCAX. Geometry optimisation starting from each of these starting points should result in the same final conformer of the isolated molecule. However, numerical differences will result from differences in the parameters used in evaluating energies and forces, as well as convergence thresholds for geometry optimisation. Therefore, differences in the resulting energies provide a measure of the inherent errors involved with a chosen set of computational parameters.

The following combinations of computational parameters were assessed:

Standard: Optimisation performed in Cartesian coordinates. The default pruned (55,434) DFT integration grid, containing 55 radial points and a maximum of 434 angular points. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-4} , 1.2×10^{-3} , 4.5×10^{-4} and 1.8×10^{-3} a.u., respectively.

TightOpt: Optimisation performed in Cartesian coordinates. The default pruned (55,434) DFT integration grid, containing 55 radial points and a maximum of 434 angular points. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-5} , 1.2×10^{-4} , 4.5×10^{-5} and 1.8×10^{-4} a.u., respectively.

INTREDUN: The same as Standard, but optimising in redundant internal coordinates in place of Cartesian coordinates. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-5} , 1.2×10^{-4} , 4.5×10^{-5} and 1.8×10^{-4} a.u., respectively.

LGRID: Optimisation performed in Cartesian coordinates. Pruned (75,434) DFT integration grid, containing 75 radial points and a maximum of 434 angular points. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-5} , 1.2×10^{-4} , 4.5×10^{-5} and 1.8×10^{-4} a.u., respectively.

INT_LGRID: Optimisation performed in redundant internal coordinates. Pruned (75,434) DFT integration grid, containing 75 radial points and a maximum of 434 angular points. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-5} , 1.2×10^{-4} , 4.5×10^{-5} and 1.8×10^{-4} a.u., respectively.

XLGRID: Optimisation performed in Cartesian coordinates. Pruned (75,974) DFT integration grid, containing 75 radial points and a maximum of 974 angular points. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-5} , 1.2×10^{-4} , 4.5×10^{-5} and 1.8×10^{-4} a.u., respectively.

INT_XLGRID: Optimisation performed in redundant internal coordinates. Pruned (75,974) DFT integration grid, containing 75 radial points and a maximum of 974 angular points. Thresholds for convergence on RMS gradient, RMS displacement, maximum gradient and maximum displacement were set to 3×10^{-5} , 1.2×10^{-4} , 4.5×10^{-5} and 1.8×10^{-4} a.u., respectively.

Figure S1 summarises the errors in calculated molecular energies using each set of computational parameters. Each coloured slice of the total bar represents the difference in calculated energy between the optimised geometry resulting from that starting structure, calculated with respect to the lowest calculated energy achieved from any of the starting points for the same molecule, with the same set of computational parameters.

The results show the importance of the greater number of radial points in the integration grid - the average errors decrease by almost an order of magnitude when the grid is increased from the pruned (55,434) to the pruned (75,434) grid. This improved accuracy comes at a minor increase in computational cost (Figure S2). The use of an even larger grid (more angular points) has no effect on the average errors.

The choice of optimising in Cartesian or redundant internal coordinates has little impact on the resulting errors in energy (Figure S1), but the use of redundant internal coordinates approximately halves the total cost of molecular geometry optimisations (Figure S2).

Based on these data the decision was taken to use internal redundant coordinates during optimisation and the large (not extra large) integration grid (INT_LGRID settings).

Using these settings, the mean and maximum difference in energies from optimisations that should result in the same molecular geometry are 0.22 kJ mol^{-1} and 0.36 kJ mol^{-1} , respectively. We estimate that maximum errors in relative energies may be up to double this value, approximately 0.7 kJ mol^{-1} , given that the error in a particular structure may have either sign.

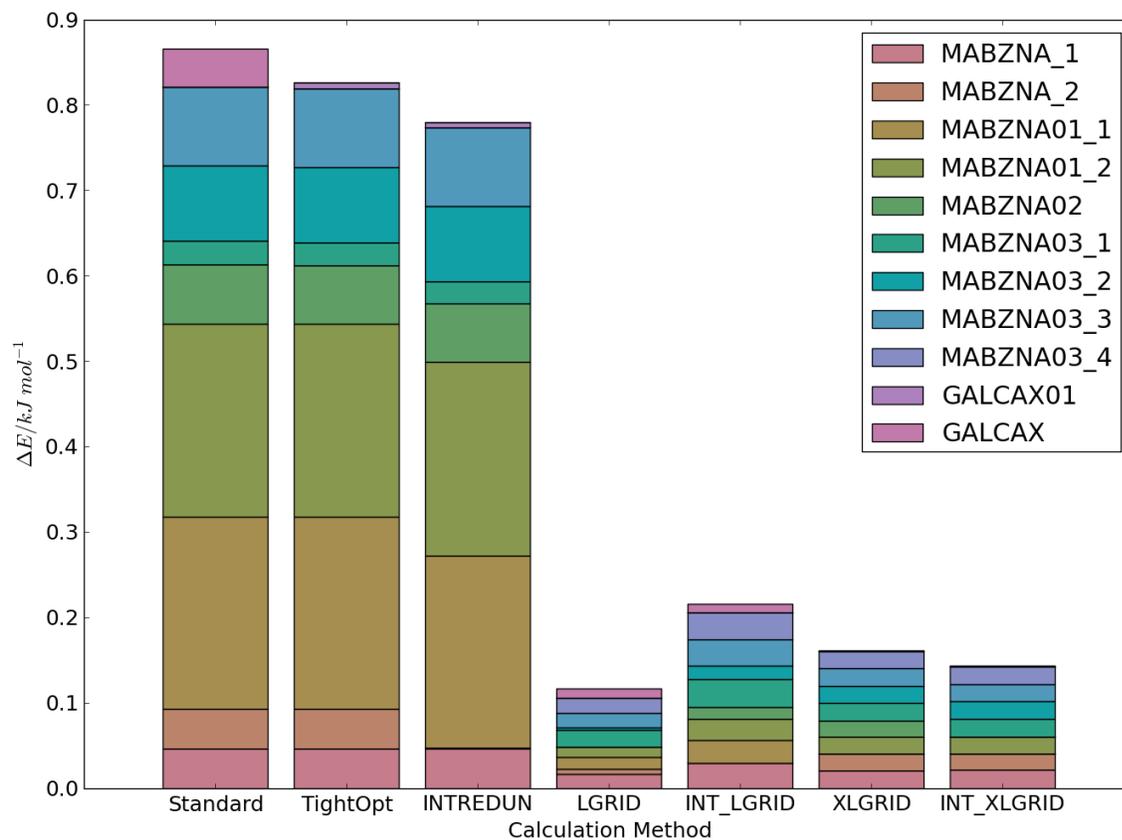


Fig. S1 Bar graph showing the error in the calculated energies depending on the computational parameters used in the molecular optimisation. Errors were calculated as the difference in energy between structures that are expected to optimise to the same conformer. The total bar height is the mean error for that calculation method; the coloured stripes represent the proportion of that mean error due to a given structure.

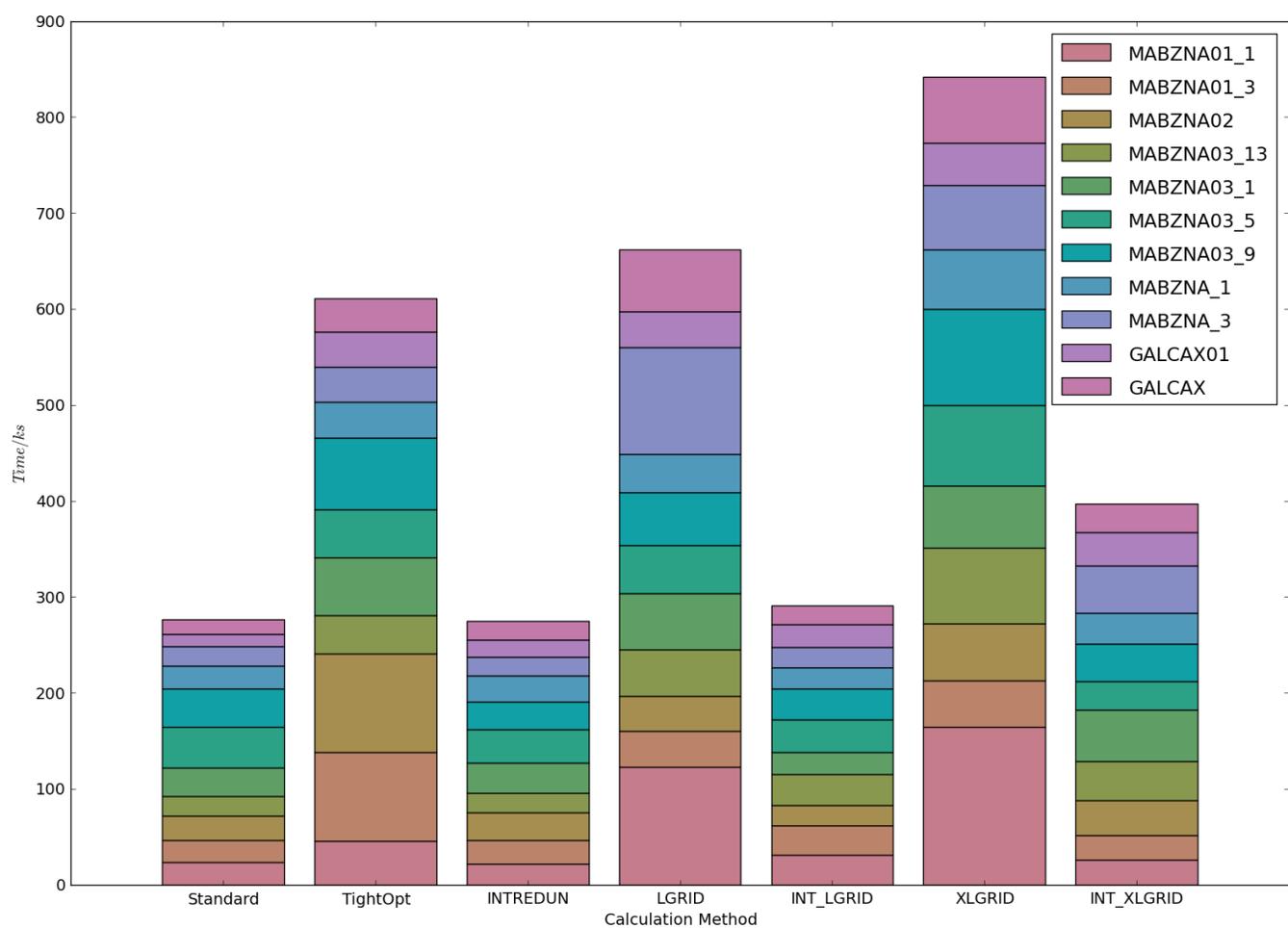


Fig. S2 Bar graph showing the effect of computational parameters on CPU time taken for a molecular optimisation. The bar height is the total calculation time. Each coloured stripe is the calculation time for a given conformation.

3 Full details of conformational searches

The LMCS method for conformer generation was chosen based on a recent review of conformer generators.⁶ LMCS is a mode-following algorithm - a starting molecular geometry is perturbed along one or a combination its calculated normal modes before re-minimising. Due to the number of energy minimisations required for an exhaustive search, this initial search is based on a molecular mechanics description of the intramolecular energy; OPLS2005⁷ was chosen based on its performance in a recent review of available force fields.⁸ Minimum and maximum move distances of 3 and 6 Å were applied and 2,000 search steps were performed per flexible dihedral angle present in the molecule. A gradient of $< 0.05 \text{ kJ mol}^{-1} \text{Å}^{-1}$ was set as a criterion for convergence of geometry optimisations. All conformations within a 50 kJ mol^{-1} window of the global minimum were saved, both to keep all conformers that might be relevant to crystal packing and to allow for significant inaccuracies of the force field. Duplicate molecular geometries were identified and removed first using an all-atom RMS deviation of atomic positions (within Macromodel), with a 0.05 Å tolerance, followed by clustering based on selected dihedral angles (performed using in-house software), with tolerances of 5° RMS and 10° maximum dihedral angle difference to identify duplicate conformers.

4 Strain energy of all observed geometries

REFCODE family	Crystal Structure	Z'	Molecule number	$\Delta E_{strain}/kJmol^{-1}$		
HIBGUV	{	HIBGUV	2	{	1	6.88
		HIBGUV01	1		2	8.00
		HIBGUV02	1		1	14.36
					1	13.45
MABZNA	{	MABZNA	2	{	1	3.18
					2	3.14
		MABZNA01	2	{	1	3.96
					2	3.98
		MABZNA02	1		1	3.68
		MABZNA03	4	{	1	6.14
					2	6.91
3	4.11					
			4	7.85		
SIKRIN	SIKRIN	1	1	14.65		
FAHNOR	{	FAHNOR	1	1	5.83	
		FAHNOR05	1	1	9.20	
ODNPDS	{	ODNPDS02	1	1	5.46	
		ODNPDS11	1	1	7.10	
COCAIN	COCAIN10	1	1	4.64		
VEMTOW	{	VEMTOW	1	1	6.18	
		VEMTOW01	1	1	3.52	
FIBKUW	{	FIBKUW01	1	1	18.34	
		FIBKUW02	1	1	21.58	
NEQNIG	NEQNIG	1	1	2.77		
HAJYUN	{	HAJYUN01	1	1	6.73	
		HAJYUN02	1	1	4.02	
GALCAX	{	GALCAX	1	1	3.77	
		GALCAX01	1	1	2.42	
SEVJAF	SEVJAF	1	1	7.11		
DANQEP	{	DANQEP	1	1	13.14	
		DANQEP01	1	1	11.63	
		DANQEP02	2	{	1	7.86
					2	6.15
CELHIL	{	CELHIL01	1	1	18.99	
		CELHIL	1	1	14.41	
DADNUR	DADNUR	1	1	16.63		

Table S1 Calculated strain energies of all 36 observed molecular geometries.

5 Correlation between strain energy and geometry change

Figures S3 and S4 show the correlation between strain energy and the change in molecular geometry between the crystalline molecular geometry and the associated optimised isolated conformer. The change in molecular geometry is measured as the root mean squared deviation (RMSD) in atomic positions (Fig S3) and as the RMSD in the flexible torsion angles (Fig S4). RMSDs were calculated using a command line program based on the molecular overlay function in CCDC's Mercury software. The calculated value excludes hydrogen atom positions. Both show a similar trend, demonstrating that the molecular strain energy tends to be higher when the geometrical distortion is larger. However, there is considerable scatter in both graphs.

Error bars are estimated for both the molecular strain energy and the geometrical distortion. The error in the molecular energy is based on the value obtained from tests reported in the see previous section.

The errors in the RMSD atomic position and RMSD torsion values are calculated by comparing the molecular geometry from the optimised crystal structure after isolated molecule optimisation with its best match from the sets of predicted conformers. The geometrical match should be perfect if the structural optimisation is fully converged. However, the convergence criteria used mean there may be small differences between the optimised geometries. The error in the geometry of the conformers is approximated as the mean difference in RMSD between the conformer from the crystal and the that from the conformational search.

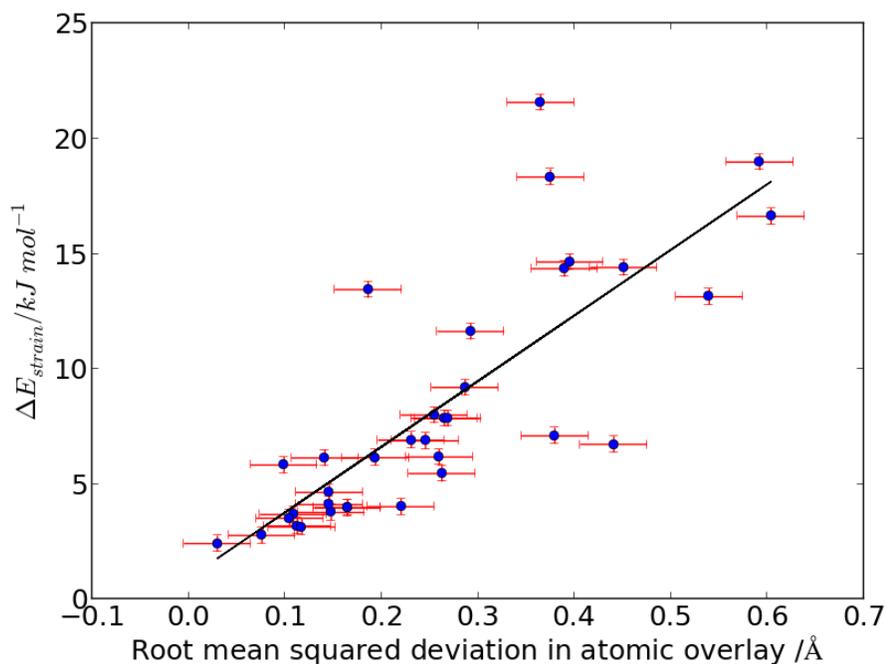


Fig. S3 Plot of ΔE_{strain} against RMSD in atom positions between the crystalline molecular geometry and the isolated molecule conformer.

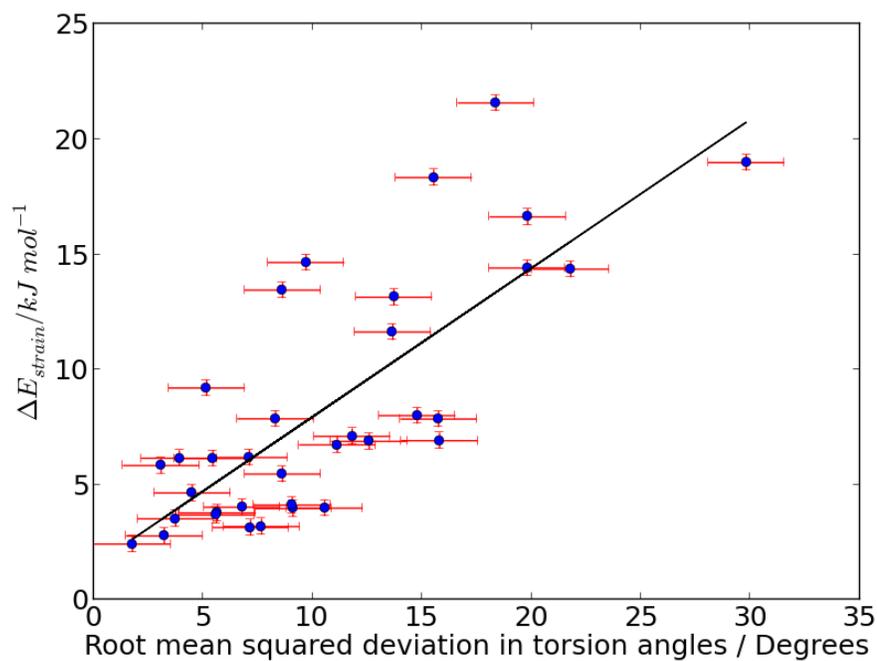


Fig. S4 Plot of ΔE_{strain} against RMSD in flexible torsion angles between the crystalline molecular geometry and the isolated molecule conformer.

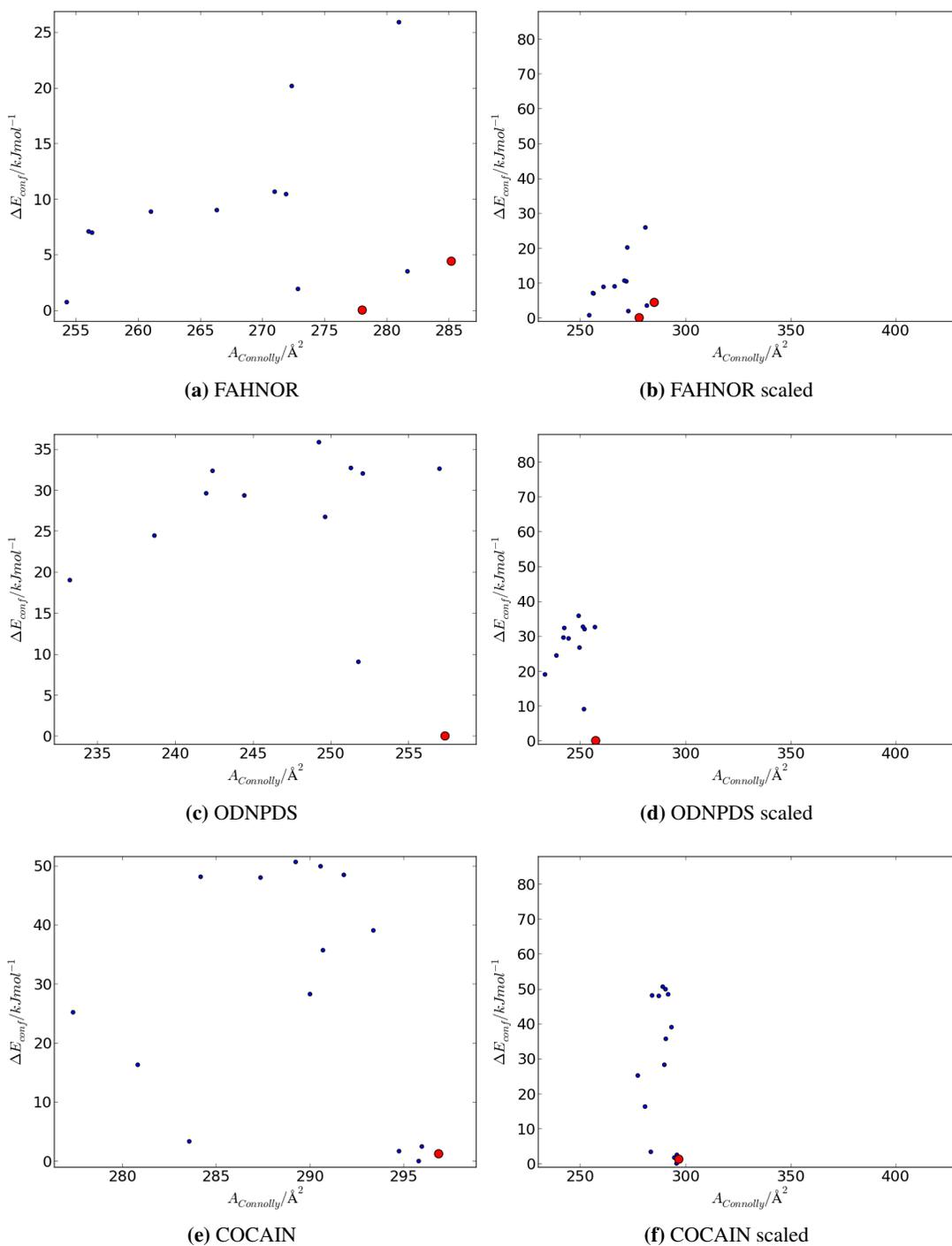


Fig. S6 ΔE_{conf} vs $A_{connolly}$ for all predicted conformers (blue) of (a) FAHNOR, (c) ODNPDS and (e) COCAIN. Plots (b), (d) and (f) show the same data using a common range for all molecules. The conformers corresponding to the observed crystal structures are highlighted in red.

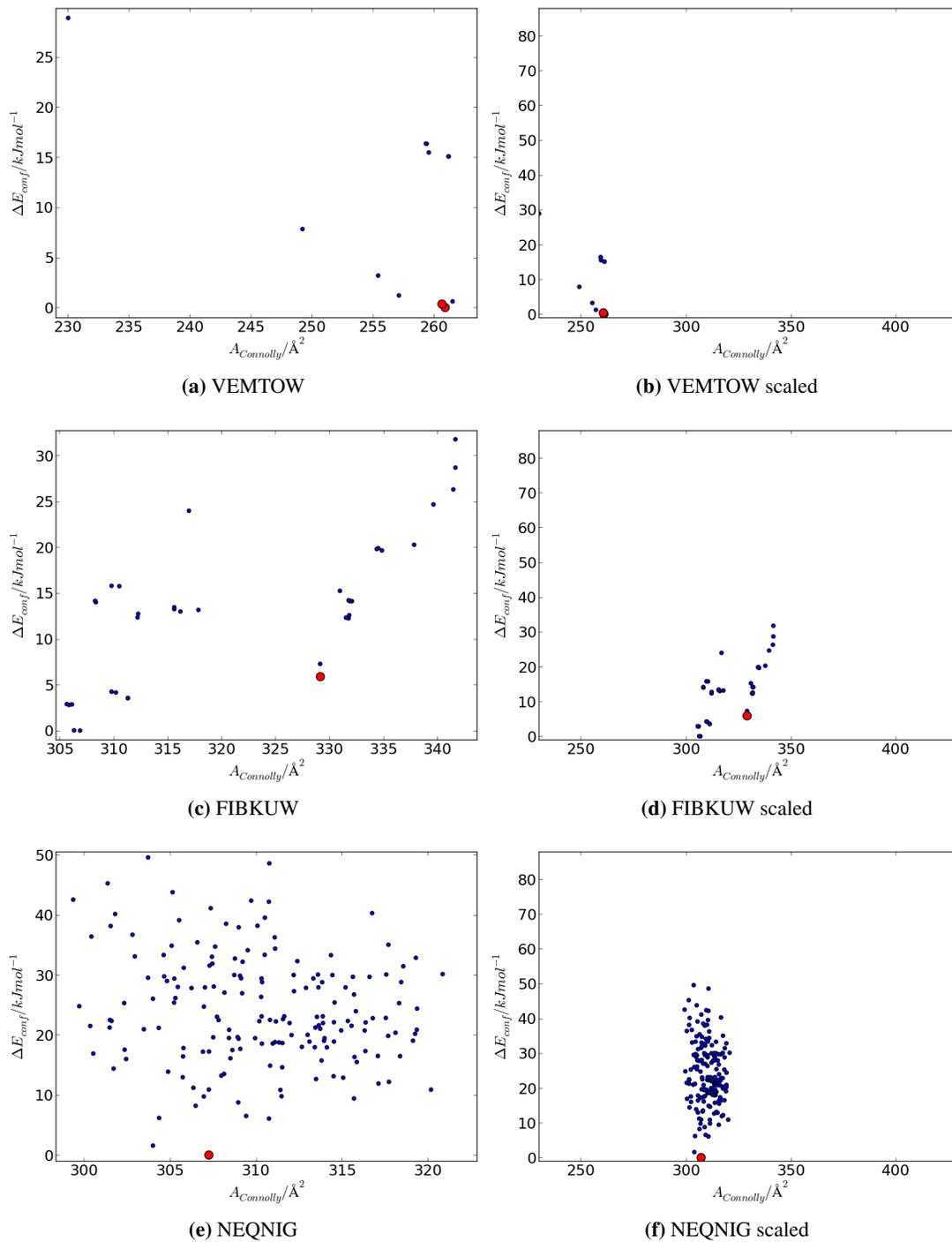


Fig. S7 ΔE_{conf} vs $A_{connolly}$ for all predicted conformers (blue) of (a) VEMTOW, (c) FIBKUW and (e) NEQNIG. Plots (b), (d) and (f) show the same data using a common range for all molecules. The conformers corresponding to the observed crystal structures are highlighted in red.

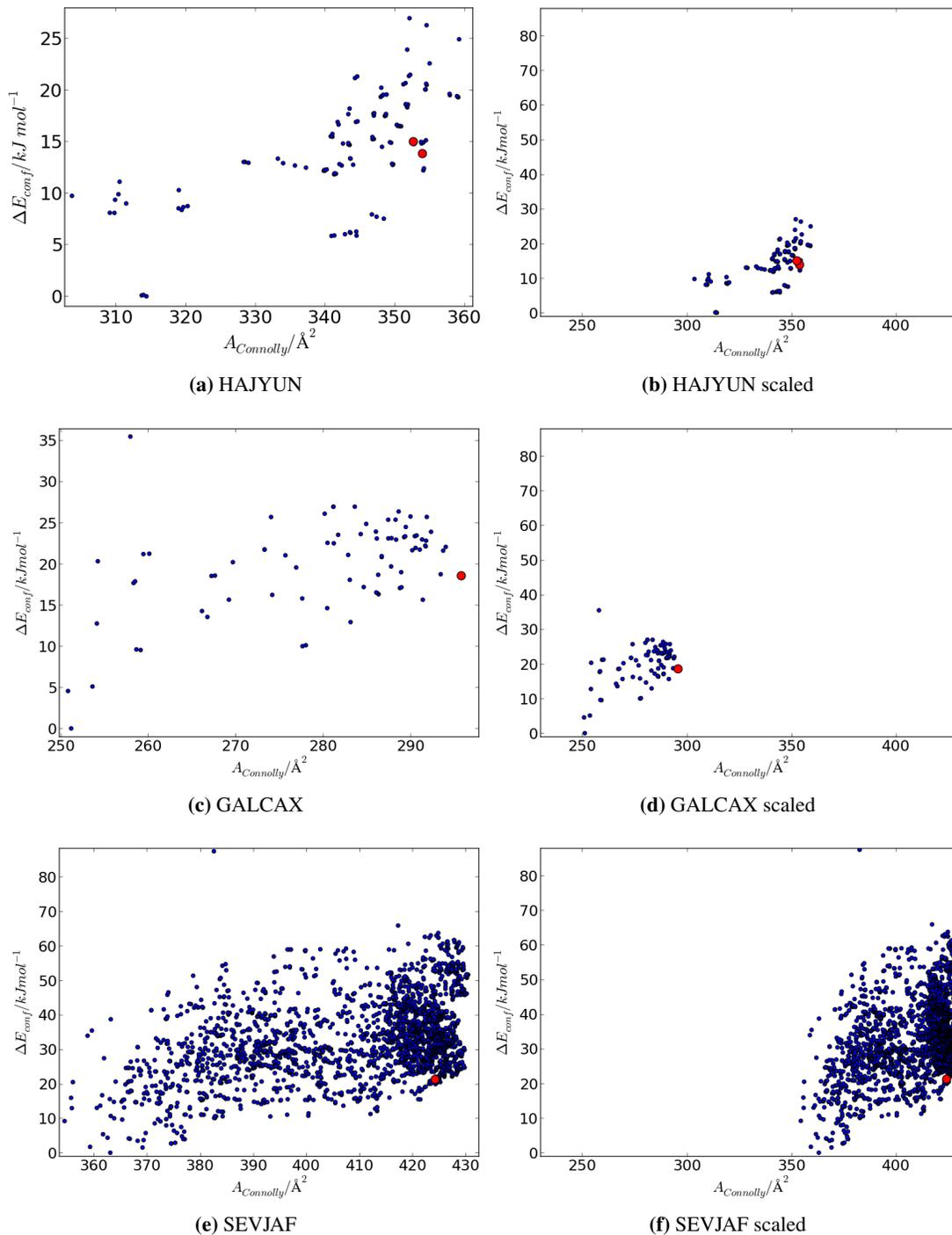


Fig. S8 ΔE_{conf} vs $A_{connolly}$ for all predicted conformers (blue) of (a) HAJYUN, (c) GALCAX and (e) SEVJAF. Plots (b), (d) and (f) show the same data using a common range for all molecules. The conformers corresponding to the observed crystal structures are highlighted in red.

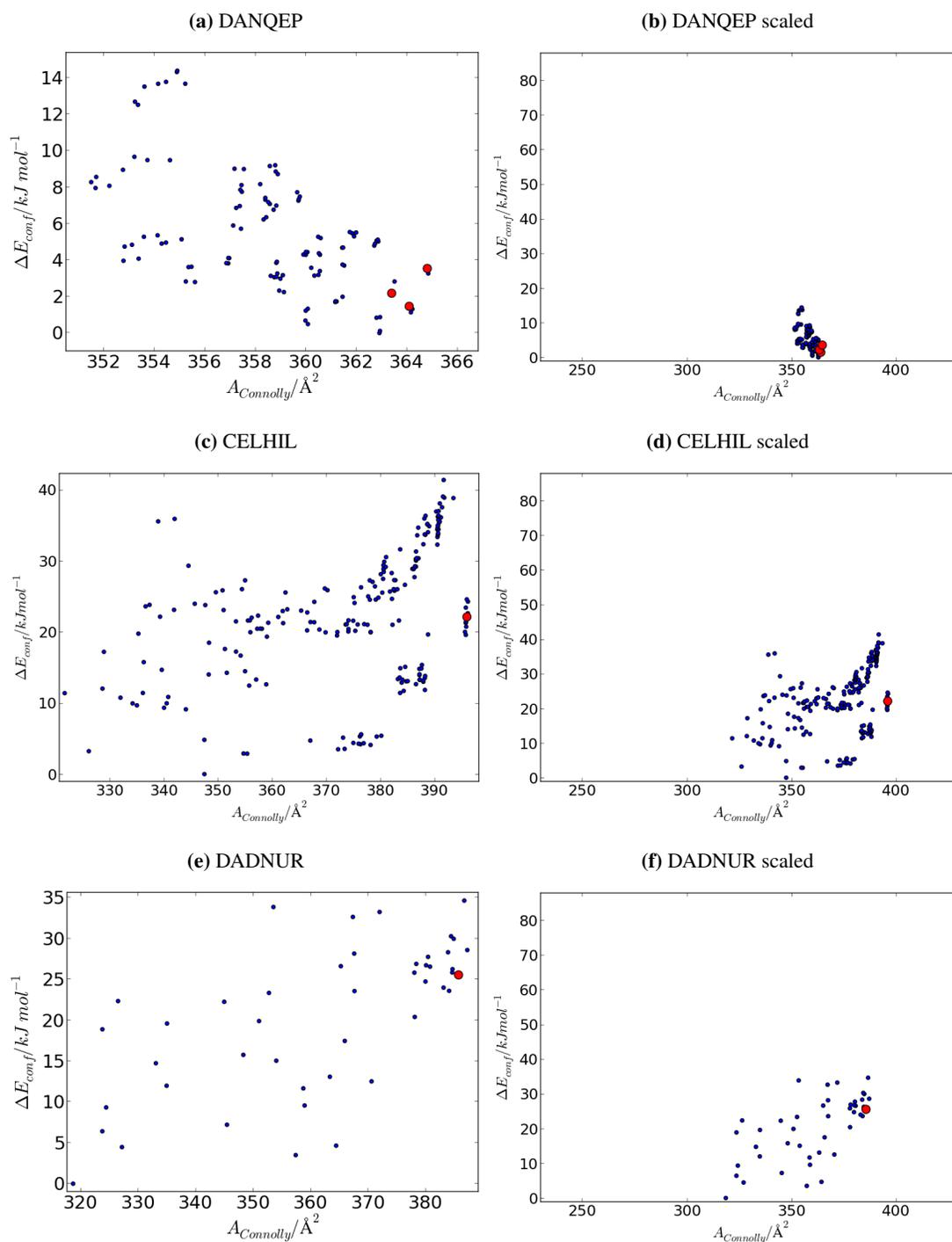


Fig. S9 ΔE_{conf} vs $A_{connolly}$ for all predicted conformers (blue) of (a) DANQEP, (c) CELHIL and (e) DADNUR. Plots (b), (d) and (f) show the same data using a common range for all molecules. The conformers corresponding to the observed crystal structures are highlighted in red.

7 Rigid molecules and experimental sublimation enthalpies

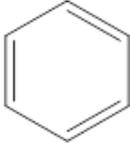
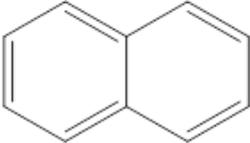
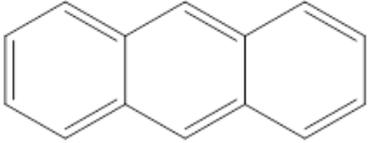
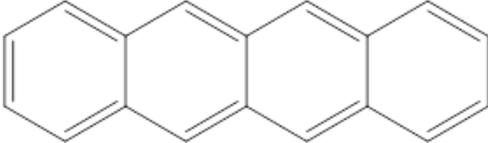
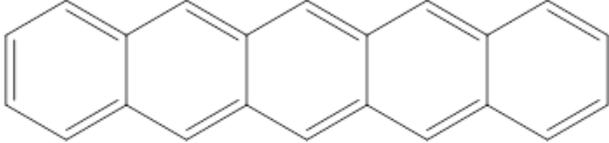
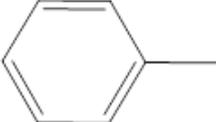
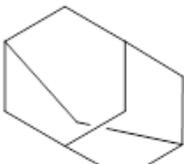
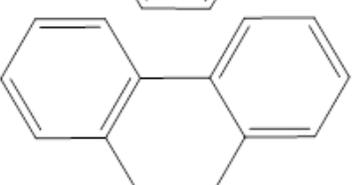
Molecule	$\Delta H_{\text{sublimation}}/\text{kJ mol}^{-1}$	$A_{\text{connolly}}/\text{\AA}^2$	2D structure
Benzene (BENZEN15)	45.5 ⁹⁻¹³	100.0	
Napthalene (NAPHTA36)	71.3 ^{10,14-26}	149.4	
Anthracene (ANTCEN16)	97.8 ^{16,23,26-34}	194.9	
Tetracene (TETCEN01)	126 ^{19,35-38}	227.6	
Pentacene (PENCEN02)	171 ^{19,38}	275.7	
Toluene (TOLUEN02)	43.1 ³⁹	125.1	
Adamantane (ADAMAN08)	57.2 ⁴⁰⁻⁴⁶	145.8	
para-Xylene (ZZZITY02)	60.1 ⁴⁷	143.7	
Phenanthrene PHENAN13	88.8 ^{14,16,26,48}	190.5	

Table S2 Details of the rigid molecules used to relate Connolly surface to lattice energy. The Cambridge Structural Database REFCODE for the crystal structure used for the molecular surface area calculation is given following the molecular name. The sublimation enthalpy is taken as a mean of reported values given in the references for each molecule.

References

- 1 R. Dovesi, R. Orlando, B. Civalleri, C. Roetti, V. R. Saunders and C. M. Zicovich-Wilson, *Zeitschrift für Kristallographie*, 2005, **220**, 571–573.
- 2 R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, M. Llunell, C. Science and A. Technologies, *Crystal09 2.0.1*, 2013.
- 3 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623–11627.
- 4 A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648.
- 5 B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano and P. Ugliengo, *CrystEngComm*, 2008, 1–6.
- 6 M. J. Loferer, I. Kolossváry and A. Aszódi, *J. Mol. Graph Model.*, 2007, **25**, 700–10.
- 7 W. L. Jorgensen and J. Tirado-Rives, *Journal of the American Chemical Society*, 1988, **110**, 1657–1666.
- 8 R. S. Paton and J. M. Goodman, *J. Chem. Inf. Model.*, 2009, **49**, 944–55.
- 9 D. R. Stull, *Industrial & Engineering Chemistry*, 1947, **39**, 517–540.
- 10 C. De Kruif and C. Van Ginkel, *The Journal of Chemical Thermodynamics*, 1977, **9**, 725–730.
- 11 A. H. Jones, *Journal of Chemical & Engineering Data*, 1960, **5**, 196–200.
- 12 S. M. Richard M. Stephenson, *AIChE Journal*, 1989, **35**, 877–877.
- 13 D. Brown, J. A. Connor, C. P. Demain, M. L. Leung, J. A. Martinho-Simoes, H. A. Skinner and M. T. Zafarani Moattar, *Journal of Organometallic Chemistry*, 1977, **142**, 321–335.
- 14 L. A. Torres-Gómez, G. Barreiro-Rodríguez and A. Galarza-Mondragón, *Thermochimica Acta*, 1988, **124**, 229–233.
- 15 F. Wania, W.-Y. Shiu and D. Mackay, *Journal of Chemical & Engineering Data*, 1994, **39**, 572–577.
- 16 W. J. Sunnefeld, W. H. Zoller and W. E. May, *Analytical Chemistry*, 1983, **55**, 275–280.
- 17 M. Colomina, P. Jimenez and C. Turrion, *The Journal of Chemical Thermodynamics*, 1982, **14**, 779–784.
- 18 C. de Kruif, T. Kuipers, J. van Miltenburg, R. Schaake and G. Stevens, *The Journal of Chemical Thermodynamics*, 1981, **13**, 1081–1086.
- 19 C. De Kruif, *The Journal of Chemical Thermodynamics*, 1980, **12**, 243–248.
- 20 D. Ambrose, I. Lawrenson and C. Sprake, *The Journal of Chemical Thermodynamics*, 1975, **7**, 1173–1176.
- 21 J. S. Chickos, *Journal of Chemical Education*, 1975, **52**, 134.
- 22 D. McEachern, O. Sandoval and J. C. Iñiguez, *The Journal of Chemical Thermodynamics*, 1975, **7**, 299–306.
- 23 D. McEachern and O. Sandoval, *Journal of Physics E: Scientific Instruments*, 1973, **6**, 155–161.
- 24 G. A. Miller, *Journal of Chemical & Engineering Data*, 1963, **8**, 69–72.
- 25 A. Aihara, *Bulletin of the Chemical Society of Japan*, 1959, **32**, 1242–1248.
- 26 R. S. Bradley and T. G. Cleasby, *Journal of the Chemical Society (Resumed)*, 1953, 1690.
- 27 P. C. Hansen and C. A. Eckert, *Journal of Chemical & Engineering Data*, 1986, **31**, 1–3.
- 28 B. F. Rordorf, *Chemosphere*, 1986, **15**, 1325–1332.
- 29 R. Bender, V. Bieling and G. Maurer, *The Journal of Chemical Thermodynamics*, 1983, **15**, 585–594.
- 30 J. W. Taylor and R. J. Crookes, *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 1976, **72**, 723.

- 31 L. Malaspina, *The Journal of Chemical Physics*, 1973, **59**, 387.
- 32 G. Beech and R. M. Lintonbon, *Thermochimica Acta*, 1971, **2**, 86–88.
- 33 J. D. Kelley and F. O. Rice, *The Journal of Physical Chemistry*, 1964, **68**, 3794–3796.
- 34 G. W. Sears and E. R. Hopke, *Journal of the American Chemical Society*, 1949, **71**, 1632–1634.
- 35 M. V. Roux, M. Temprado, J. S. Chickos and Y. Nagano, *Journal of Physical and Chemical Reference Data*, 2008, **37**, 1855.
- 36 K. Nass, D. Lenoir and A. Kettrup, *Angewandte Chemie International Edition in English*, 1995, **34**, 1735–1736.
- 37 H. Inokuchi, S. Shiba, T. Handa and H. Akamatu, *Bulletin of the Chemical Society of Japan*, 1952, **25**, 299–302.
- 38 N. Wakayama and H. Inokuchi, *Bulletin of the Chemical Society of Japan*, 1967, **40**, 2267–2271.
- 39 C. Lenchitz and R. W. Velicky, *Journal of Chemical & Engineering Data*, 1970, **15**, 401–403.
- 40 R. Jochems, H. Dekker, C. Mosselman and G. Somsen, *The Journal of Chemical Thermodynamics*, 1982, **14**, 395–398.
- 41 T. Clark, T. M. Knox, M. A. McKervey, H. Mackle and J. J. Rooney, *Journal of the American Chemical Society*, 1979, **101**, 2404–2410.
- 42 T. Clark, T. Knox, H. Mackle, M. A. McKervey and J. J. Rooney, *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 1975, **71**, 2107.
- 43 R. H. Boyd, S. N. Sanwal, S. Shary-Tehrany and D. McNally, *The Journal of Physical Chemistry*, 1971, **75**, 1264–1271.
- 44 R. Butler, A. Carson, P. Laye and W. Steele, *The Journal of Chemical Thermodynamics*, 1971, **3**, 277–280.
- 45 P.-J. Wu, L. Hsu and D. A. Dows, *The Journal of Chemical Physics*, 1971, **54**, 2714.
- 46 W. K. Bratton, I. Szilard and C. A. Cupas, *The Journal of Organic Chemistry*, 1967, **32**, 2019–2021.
- 47 P. M. Burkinshaw and C. T. Mortimer, *Journal of the Chemical Society, Dalton Transactions*, 1984, 75.
- 48 A. G. Osborn and D. R. Douslin, *Journal of Chemical & Engineering Data*, 1975, **20**, 229–231.