Electronic Supplementary Material (ESI) for Chemical Society Reviews.
This journal is © The Royal Society of Chemistry 2020

Electronic Supplementary Information for

Guest-occupiable space in the crystalline solid state: a simple rule-of-thumb for predicting occupancy

Dewald P. van Heerden and Leonard J. Barbour*

Department of Chemistry and Polymer Science, University of Stellenbosch, Matieland, 7600, South Africa. Email: ljb@sun.ac.za

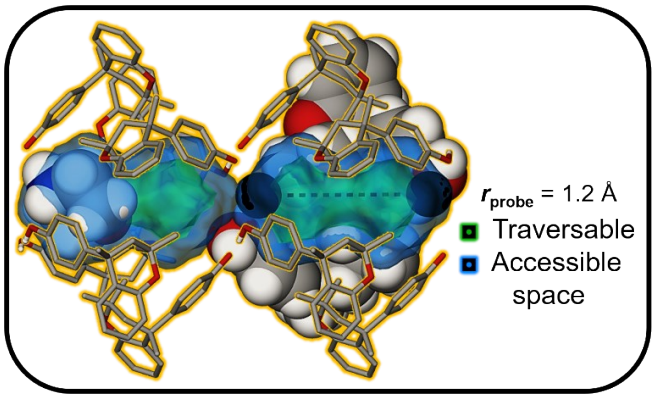

Table of Contents

1 Group Additivity sets

Table S1 Comparison of functional group volumes. Estimates from Kitaigorodsky[1] and Cady[2] are spheres-and-caps derived, while those from other authors are derived from formula unit volumes; i.e., assuming no void space. Dashes indicate values that were not parameterized.

Group	V_{FU} (Å³)[a]				$V_{s\&c}$ (Å³)[b]	
	Hofmann[3]	Immirzi & Perini[4]	Stein[5]	Ammon[6]	Kitaigorodsky	Cady
–H	5.1	6.9	6.0	5.4	2.0	3.1
–F	11.2	12.8	12.9	13.5	9.6	9.3
–Cl	25.8	26.7	–	27.8	19.9	19.4
–Br	32.7	33.0	–	33.3	26.0	–
⩾CH	19.0	17.9	19.4	21.5	11.1	11.2
>CH_2	24.0	24.8	25.4	24.4	17.1	16.9
–CH_3	29.1	31.7	31.3	32.2	23.5	22.6
–OH	16.5	–	12.5	12.9	–	–
–NH_2	22.0	21.0	19.6	19.7	19.7	16.6
–NO_2	34.6	35.2[c]	35.9	35.6	23.0	28.3

[a] Derived using equation 1 of the main text.

[b] Derived using equation 7 of the main text.

[c] The erroneous original value of 25.6 Å³ was amended as per the recommendation of Cady.

Table S2 Comparison of functional group displacement volumes for groups bonded to and including an aromatic carbon. All formula unit-derived values were scaled by $PC_K = {}^{7}/_{10}$ to compare with spheres-and-caps derived volumes by Kitaigorodsky[1] and Cady.[2] Dashes indicate values that were not parameterized.

Group	$PC_K \times V_{FU}$ (Å³)					$V_{s\&c}$ (Å³)	
	Hofmann[3]	Tarver[a]	Immirzi & Perini[4]	Stein[5]	Ammon[6]	Kitaigorodsky	Cady
C_a–H	13.27	13.81	14.42	13.71	12.49	14.7	13.54
C_a–CH_3	30.09	30.19	31.78	30.4	31.2	31.9	31.64
C_a–OH	21.24	18.63	–	17.25	17.74	–	–
C_a–NH_2	25.08	18.21	24.29	22.21	22.48	28.1	25.71
C_a–NO_2	33.92	34.53	34.23	33.58	33.62	31.4	37.34

[a] Values from Tarver *et al.*[7] incorporate the carbon atom that the functional group is bonded to, preventing a direct comparison as in Table S1.

2 Guest Molecular volume

In this work, the vdW surface was used to estimate V_{mol}. However, as Meyer pointed out, contours of electron density do not have the sharp clefts of the 'vdW surface' that result from the clockwork-like union of spheres, but are more rounded out or 'peanut necked.'[17] The shape of a molecule ultimately depends on its electron density distribution, which can be determined from diffraction experiments or taken as the complex square of its quantum-mechanical wavefunction representation, ψ^2.[18] Bader proposed the isodensity value of 0.014 e Å^{-3} as a theoretical measure of molecular size, but later recommended the 0.007 e Å^{-3} contour that encompasses 98% of the total electron density.[19,20] The $\rho_{0.007}$ isosurface of the formula unit approximates its volume V_{FU};[21,22] it extends beyond the molecular surface and is a suitable contour for mapping the molecular surface electrostatic potential (MESP) relevant to the onset of noncovalent interactions.[23,24] In a preliminary study relating free space and impact sensitivity of energetic compounds, Politzer *et al.* used a contour value of 0.014 e Å^{-3}, but later recommended the $\rho_{0.020}$ isosurface as an estimate of V_{mol}.[25,26] Bouhmaida and Ghermani investigated the suitability of electronic isosurfaces for estimating ρ_{cryst} and determined PC_K values of 0.99, 0.85, 0.71 and 0.58 for isodensity thresholds of 0.007, 0.014, 0.027 and 0.054 e Å^{-3}, respectively.[27] This underscores that a $\rho_{0.007}$ isosurface is an estimate of V_{FU}, while 0.027 e Å^{-3} is a more appropriate threshold for estimating V_{mol}. For an assessment of MESP methods for organic energetic materials, see for example Ghule *et al.*[28]

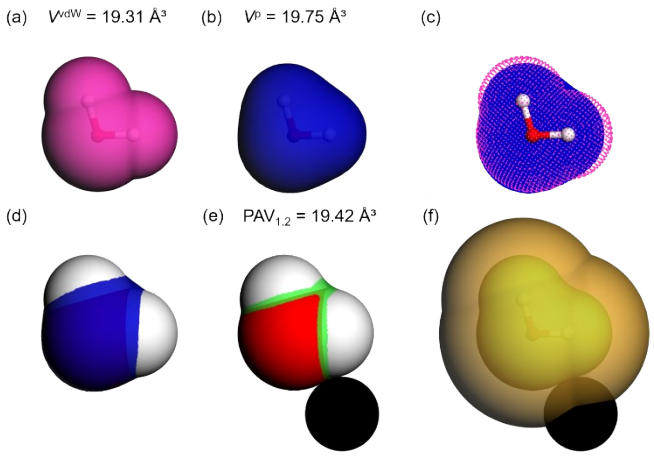

Fig. S1 Different surface typologies of water. The semi-transparent (a, pink) vdW and (b, blue) 0.017 e Å^{-3} isodensity surfaces are overlaid in (c) with the former shown as dots while a section through the molecular plane is shown for the latter. Water is shown in the ball-and-stick metaphor with oxygen and hydrogen atoms coloured red and white, respectively. In (d) the isodensity surface is overlaid with the space-fill representation of water. The PAS (green) is overlaid in (e) and shown along with the PTS (orange) in (f); the probe with radius 1.2 Å is shown as a black sphere. Smoothing of clefts present in the hard-sphere model is apparent in (d) and (e) as saddle areas.

The molecular volume of a solvent was estimated as the mean V_g determined for each unique guest molecule. Values obtained by MStudio, CCDC and Zeo++ using r_p = 0.0 Å for a guest molecule located in the empty unit cell (from which the host and other guest molecules were omitted) are listed under "periodic" in Table S3. Molecules were kept in their crystallographic geometries. MStudio *in vacuo* V^{vdW} evaluations of force-field geometry optimised solvent molecules appear under "non-periodic" along with the volume enclosed by a 0.017 e Å^{-3} contour of the electron density calculated at the DFT/Dζ level of theory (see §3). Values listed under "bulk" were calculated using the molecular weight of the solvent and its liquid density at 25 °C:[29,30]

$$V_{\mathrm{mol}}^{\mathrm{bulk}} = 0.9586\frac{MW}{\rho_l} \qquad \text{S1}$$

A comparison of V_{mol} determined using the additivity methods of Hofmann and Ammon reflects the increased accuracy of the latter when the chemical environment is accounted for. However, in the interest of a robust application, an immutable set of element-specific r_{vdW} were employed. MStudio non-periodic $V_{\mathrm{g}}^{\mathrm{vdW}}$ were employed as estimations of solvent V_{mol} as they are intermediate between the crystallographic-geometry spheres-and-caps values and the smoother electronic isodensity values.

The geometry of the water molecule was optimised and the electron density derived using DMol3[31,32] implemented in MStudio with the B3LYP density functional[33] and the DNP basis set (double numeric plus polarizing d- and p-functions on non-hydrogen and hydrogen atom, respectively) to a 3.3 Å cutoff. Threshold values for SCF convergence and change in energy, maximum force and maximum displacement were 6.28 × 10^{-4} and 6.28 × 10^{-3} kcal mol^{-1}, 1.255 kcal mol^{-1} Å^{-1} and 0.005 Å, respectively. A DIIS (direct inversion in an iterative subspace, Pulay scheme) size of 6 was applied to speed up convergence. The MStudio default electronic isodensity value of 0.017 e Å^{-3} (0.0025 e bohr^{-3}) was employed.

Table S3 Comparison of solvent $V^{\text{bulk}}_{\text{mol}}$ (Å³). Bulk values were calculated using Eqn S1, while formula unit-derived values were scaled by PC_K = $^7/_{10}$ to account for V_{void}. Periodic $V^{\nu\text{dW}}_{\text{g}}$ values are for solvent molecules in their crystallographic geometries, whereas non-periodic $V^{\nu\text{dW}}_{\text{g}}$ and 0.017 e Å^{-3} V^{ρ}_{g} values are for *in vacuo* force-field geometry optimised molecules.

Solvent	$V^{\text{bulk}}_{\text{mol}}$	$PC_K \times V_{FU}$		Periodic $V^{\nu\text{dW}}_{\text{g}}$			Non-periodic	
		Hofmann	Ammon	MStudio	CCDC	Zeo++	$V^{\nu\text{dW}}_{\text{g}}$	V^{ρ}_{g}
acetic acid	55.108	49.588	51.422	53.233	53.115	53.219	55.889	58.588
acetone	70.968	58.436	61.727	60.191	60.181	60.636	64.640	68.403
acetonitrile	50.086	38.346	43.948	43.510	43.481	43.809	45.855	48.357
benzene	85.429	79.59	82.908	78.662	79.276	78.998	84.024	89.007
chloroform	77.384	67.445	73.296	69.106	68.936	68.606	70.461	79.389
cyclohexane	104.244	100.926	97.789	91.865	93.078	92.331	101.764	109.053
dcm	61.373	52.941	55.89	54.855	54.738	54.979	56.625	62.926
diethyl ether	99.541	82.369	88.038	81.096	80.938	81.116	88.466	94.512
dma	89.109	87.073	93.204	86.352	86.157	87.572	93.438	99.265
dmf	80.275	70.252	75.228	71.042	70.967	72.079	76.755	81.394
dmso	68.028	66.367	72.687*	66.350	66.231	67.669	71.892	77.947
dioxane	81.705	83.23	82.879	79.478	80.254	80.345	85.134	90.681
ethanol	55.949	48.727	48.626	48.031	47.943	49.792	53.979	57.108
ethyl acetate	93.811	83.23	90.033	84.142	84.090	84.387	90.492	95.957
hexane	125.049	108.038	113.284	103.625	104.690	103.110	113.067	121.342
iso-propanol	73.77	65.548	69.133	63.609	63.570	65.408	71.041	75.774
methanol	38.812	31.906	31.562	32.748	32.710	34.520	36.814	38.399
pentane	110.447	91.217	96.22	87.416	87.534	87.207	95.948	103.098
thf	78.253	83.23	74.075	71.372	71.366	71.831	78.114	82.934
toluene	102.429	96.411	100.292	94.377	94.180	94.431	100.591	106.793

* The value for dmso was calculated as: 2 × "CH_3 of CH_3–S" + "SO_2 of sulfone" – "C=O not in a ring" + "Csp not in a ring".

Given the small deviation of $V^{\nu\text{dW}}_{\text{g}}$ from V^{ρ}_{g}, we opted to estimate the molecular volume of a guest as the volume of its readily attainable vdW surface.

3 Molecular isodensity volumes

Geometries were optimised using the COMPASSII force field[34] implemented in the Forcite module of MStudio. Force field-assigned charges and atom-based summation for non-bond interactions with cubic spline truncation was employed with threshold values of 0.001 kcal mol^{-1}, 0.5 kcal mol^{-1} Å^{-1} and 0.015 Å for the change in energy, maximum force and maximum displacement criteria, respectively. Single point energy evaluations were carried out with DMol3 implemented in MStudio using the PBE density functional[35] and the DND basis set (double numeric plus polarization d-functions on non-hydrogen atoms) with a SCF tolerance of 6.28 × 10^{-3} kcal mol^{-1}and an orbital cut-off of 3.3 Å.

4 Van der Waals radii

The conceptually simplest boundary of a molecule is that of the molecular surface resulting from a set of fused spheres of suitable radii centred at the position of atomic nuclei. It is the steepness of the repulsive term in the potential function for noncovalent interactions that is responsible for the success of the "hardness" of the hard-sphere model.[8] Following Bragg's[9] *ansatz* of considering crystals as 'an assemblage of spheres' – each with element-characteristic diameters – packed tightly together, and that of 'crystal radii' by Mack,[10] Pauling introduced the term 'van der Waals radius' (r_{vdW}) to describe the distance midway between atoms in van der Waals contact.[11] For a comparison of sets of r_{vdW}, see for example Hu *et al.*[12] and Batsanov.[13]

See www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/Elemental_Radii.xlsx for r_{vdW} values implemented by the CCDC. Although $r_{vdW}^{hydrogen}$ is listed as 1.09 Å,[14] the actual value is 1.20 Å. Radii not available in either Bondi[15,16] or Rowland and Taylor[14] have r_{vdW} = 2.00 Å. The ***–r*** argument was invoked for Zeo++ analyses to read a list of van der Waals radii consistent with that of CCDC. There was no convenient manner in which to universally implement a set of r_{vdW}; however, discrepancies only occur for metallic elements usually concealed by a coordination sphere.

Table S4 List of element-specific van der Waals radii employed by each software suite.

Element		vdW radius (Å)			
Name	Symbol	CCDC	Zeo++	MStudio	PLATON
Aluminium	Al	2	1.43		2.15
Americium	Am	2	1.73		2.31
Antimony	Sb	2	1.61		2.26
Argon	Ar		1.88		
Arsenic	As		1.85		
Barium	Ba	2	2.24		2.14
Beryllium	Be	2	1.12		1.15
Bismuth	Bi	2	1.82		2.34
Boron	B	2	1.75		1.63
Bromine	Br		1.85		
Cadmium	Cd		1.58		
Caesium	Cs	2	2.72		2.47
Calcium	Ca	2	1.97		1.79
Carbon	C		1.7		
Cerium	Ce	2	1.81		2.63
Chlorine	Cl		1.75		
Chromium	Cr	2	1.29		2.15
Cobalt	Co	2	1.25		2.03
Copper	Cu		1.4		
Dysprosium	Dy	2	1.78		2.55
Erbium	Er	2	1.76		2.53
Europium	Eu	2	2.08		2.79
Fluorine	F		1.47		
Gadolinium	Gd	2	1.8		2.59
Gallium	Ga		1.87		
Germanium	Ge	2	1.39		1.97

Element		vdW radius (Å)			
Name	Symbol	CCDC	Zeo++	MStudio	PLATON
Gold	Au		1.66		
Hafnium	Hf	2		1.59	2.37
Helium	He		1.4		
Holmium	Ho	2		1.76	2.54
Hydrogen	H		1.2		
Indium	In		1.93		
Iodine	I		1.98		
Iridium	Ir	2		1.36	2.12
Iron	Fe	2		1.26	2.14
Krypton	Kr		2.02		
Lanthanum	La	2		1.88	2.67
Lead	Pb		2.02		
Lithium	Li		1.82		
Lutetium	Lu	2		1.73	2.52
Magnesium	Mg		1.73		
Manganese	Mn	2		1.37	2.15
Mercury	Hg		1.55		
Molybdenum	Mo	2		1.4	2.27
Neodymium	Nd	2		1.81	2.61
Neon	Ne		1.54		
Neptunium	Np	2		1.55	2.35
Nickel	Ni		1.63		
Niobium	Nb	2		1.47	2.28
Nitrogen	N		1.55		
Osmium	Os	2		1.35	2.17
Oxygen	O		1.52		

Element		vdW radius (Å)			
Name	Symbol	CCDC	Zeo++	MStudio	PLATON
Palladium	Pd		1.63		
Phosphorus	P		1.8		
Platinum	Pt	1.72		1.75	1.72
Plutonium	Pu	2		1.59	2.33
Polonium	Po	2		1.64	2.48
Potassium	K		2.75		
Praseodymium	Pr	2		1.82	2.62
Promethium	Pm	2		1.83	2.6
Protactinium	Pa	2		1.63	2.41
Radium	Ra	2		2.35	2.7
Rhenium	Re	2		1.37	2.15
Rhodium	Rh	2		1.34	2.25
Rubidium	Rb	2		2.5	2.27
Ruthenium	Ru	2		1.34	2.3
Samarium	Sm	2		1.8	2.6
Scandium	Sc	2		1.64	2.24
Selenium	Se		1.9		
Silicon	Si		2.1		
Silver	Ag		1.72		
Sodium	Na		2.27		

Element		vdW radius (Å)			
Name	Symbol	CCDC	Zeo++	MStudio	PLATON
Strontium	Sr	2		2.15	1.92
Sulphur	S		1.8		
Tantalum	Ta	2		1.47	2.23
Technetium	Tc	2		1.35	2.15
Tellurium	Te		2.06		
Terbium	Tb	2		1.77	2.56
Thallium	Tl		1.96		
Thorium	Th	2		1.96	2.59
Thulium	Tm	2		1.75	2.52
Tin	Sn		2.17		
Titanium	Ti	2		1.47	2.27
Tungsten	W	2		1.41	2.17
Uranium	U		1.86		
Vanadium	V	2		1.35	2.13
Xenon	Xe		2.16		
Ytterbium	Yb	2		1.93	2.74
Yttrium	Y	2		1.82	2.58
Zinc	Zn	1.39		1.39	2.25
Zirconium	Zr	2		1.6	2.36

5 Software investigated parameter comparison

Table S5 Influence of choice of grid spacing for MStudio and CCDC APIs and the number of Monte Carlo samplings employed by Zeo++ for the integration of surface volume. Mean surface volumes and 99% confidence-level intervals were derived for 323 virtually porous pentane host structures.

Grid spacing (Å)	MStudio		CCDC		Zeo++		
	$^{MS}PTV_{1.2}$ (Å³)	$^{MS}PAV_{1.2}$ (Å³)	$^{C}PTV_{1.2}$ (Å³)	$^{C}PAV_{1.2}$ (Å³)	Monte Carlo samples	$^{Z}PTV_{1.2}$ (Å³)	$^{Z}PAV_{1.2}$ (Å³)
0.1	48.36 ± 2.04	258.83 ± 7.46	48.53 ± 2.06	219.73 ± 5.19	5×10^3	48.25 ± 2.28	153.90 ± 4.63
0.2	48.36 ± 2.04	258.83 ± 7.46	48.72 ± 2.09	211.30 ± 4.94			
0.3	48.36 ± 2.04	258.84 ± 7.47	48.92 ± 2.13	202.77 ± 4.79	5×10^4	48.33 ± 2.21	196.90 ± 4.72
0.4	48.32 ± 2.04	258.80 ± 7.46	49.10 ± 2.17	195.36 ± 4.68			
0.5	48.37 ± 2.04	258.81 ± 7.47	49.33 ± 2.21	186.74 ± 4.61	5×10^5	48.32 ± 2.19	213.23 ± 5.35
0.6	48.46 ± 2.04	258.85 ± 7.46	49.50 ± 2.26	179.57 ± 4.60			
0.7	48.47 ± 2.05	258.85 ± 7.49	49.69 ± 2.32	166.53 ± 4.50			
0.8	48.43 ± 2.05	258.78 ± 7.46	49.89 ± 2.36	162.86 ± 4.49			

6 Refcode references

Table S6 References to CSD refcodes used in this work not cited in the main text.

CSD refcode	Reference
NOPSOA	A. L. Llamas-Saiz, C. Foces-Foces, J. Elguero and W. Meutermans, *Supramol. Chem.*, 1994, **4**, 53–62
PASHIB/GOG	K. Kato, M. Sugahara, N. Tohnai, K. Sada and M. Miyata, *Cryst. Growth Des.*, 2004, **4**, 263–272
TOTFAJ	P. Fonte, F. H. Kohnke, M. F. Parisi, S. Menzer and D. J. Williams, *Tetrahedron Lett.*, 1996, **37**, 6205–6208
NOJZUI	D. Das, L. J. Barbour, *Chem. Commun.*, 2008, 5110–5112

7 Outliers summary

Table S7 Collection of $PAV_{1.2}$. SD = standard deviation. Refined data, after omission of outliers (< Q1 – 1.5IQR or > Q3 + 1.5IQR), are shown in boldface.

Solvent	MStudio				CCDC				Zeo++				PLATON			
	Count	Mean	Median	SD	Count	Mean	Median	SD	Count	Mean	Median	SD	Count	Mean	Median	SD
acetic acid	266	111.63	108.07	27.22	266	95.68	92.90	22.08	266	94.46	92.00	18.67	266	96.53	93.75	21.60
	256	**109.07**	**107.79**	**18.09**	**256**	**93.65**	**92.70**	**13.37**	**255**	**92.68**	**91.48**	**11.99**	**257**	**94.67**	**93.00**	**13.32**
acetone	1 516	162.59	150.54	49.80	1 516	135.29	127.19	37.44	1 516	128.64	122.35	31.62	1 516	135.88	128.00	36.75
	1 358	**149.95**	**146.79**	**24.57**	**1 388**	**126.77**	**124.87**	**18.22**	**1 382**	**121.74**	**120.73**	**15.82**	**1 388**	**127.62**	**126.00**	**18.00**
acetonitrile	3 727	123.73	115.12	41.06	3 727	97.86	92.65	30.80	3 727	92.23	88.19	26.53	3 727	98.55	93.50	30.40
	3 366	**115.70**	**112.15**	**22.89**	**3 396**	**92.27**	**91.18**	**15.10**	**3 410**	**87.86**	**86.85**	**13.41**	**3 385**	**92.89**	**91.50**	**14.61**
benzene	3 478	207.98	195.4	55.53	3 478	175.18	167.31	40.38	3 478	165.57	159.16	35.15	3 478	174.16	166.37	39.51
	3 128	**196.03**	**191.46**	**31.22**	**3 143**	**166.99**	**164.98**	**20.57**	**3 159**	**159.32**	**157.44**	**17.99**	**3 147**	**166.20**	**164.25**	**20.14**
chloroform	3 863	175.58	165.36	50.86	3 863	150.89	143.95	41.63	3 863	145.10	139.31	36.92	3 863	151.43	144.50	41.20
	3 475	**165.48**	**162.39**	**22.82**	**3 515**	**143.51**	**142.06**	**16.47**	**3 493**	**138.79**	**137.65**	**14.28**	**3 493**	**143.91**	**142.50**	**15.86**
cyclohexane	158	267.61	242.91	91.33	158	227.25	205.94	80.27	158	212.89	192.89	73.70	158	224.73	202.75	78.80
	139	**240.28**	**231.75**	**43.26**	**137**	**200.81**	**201.91**	**29.97**	**140**	**192.41**	**190.32**	**30.25**	**137**	**199.86**	**198.00**	**29.55**
dcm	7 647	153.57	143.44	51.44	7 647	124.57	117.85	42.60	7 647	117.39	111.77	39.16	7 647	125.19	118.38	42.25
	6 970	**144.19**	**140.55**	**24.49**	**7 008**	**117.54**	**115.96**	**16.61**	**7 033**	**111.66**	**110.41**	**14.74**	**6 995**	**118.18**	**116.50**	**16.18**
diethyl ether	1 266	241.31	226.08	77.30	1 266	199.05	187.52	62.77	1 266	185.15	176.46	55.26	1 266	199.43	188.50	62.00
	1 105	**223.52**	**219.96**	**36.98**	**1 109**	**185.39**	**183.24**	**25.66**	**1 113**	**174.58**	**174.15**	**22.46**	**1 108**	**186.06**	**185.00**	**25.30**
dma	168	197.13	187.28	41.64	168	174.85	170.58	33.39	168	170.44	167.58	30.95	168	174.89	171.25	32.37
	158	**192.82**	**184.6**	**32.86**	**157**	**170.43**	**167.1**	**21.76**	**157**	**165.89**	**165.39**	**18.28**	**157**	**170.32**	**169.50**	**20.58**
dmf	2 440	161.48	153.18	42.46	2 440	142.70	136.89	35.89	2 440	139.08	133.88	32.31	2 440	143.04	137.00	35.37
	2 241	**153.46**	**151.3**	**20.90**	**2 261**	**136.48**	**135.26**	**16.13**	**2 262**	**133.71**	**132.57**	**14.69**	**2 258**	**136.85**	**135.50**	**15.90**
dmso	1 189	138.48	134.42	26.03	1 189	122.52	119.79	20.56	1 189	121.22	119.09	18.89	1 189	123.20	120.25	20.28
	1 127	**134.66**	**133.09**	**17.28**	**1 143**	**120.19**	**119.39**	**14.04**	**1 145**	**119.31**	**118.71**	**12.8**	**1 139**	**120.65**	**119.50**	**13.36**
dioxane	430	172.38	161.12	61.18	430	151.11	145.92	42.20	430	146.76	142.71	34.58	430	149.51	143.58	42.19
	404	**162.84**	**160.21**	**25.85**	**412**	**145.86**	**145.29**	**19.97**	**414**	**142.85**	**141.88**	**17.82**	**412**	**144.25**	**143.00**	**19.42**
ethanol	1 214	131.52	119.97	62.00	1 214	105.93	97.89	55.25	1 214	101.04	94.24	53.48	1 214	106.67	99.00	55.44
	1 105	**120.97**	**117.5**	**23.15**	**1 118**	**98.10**	**96.44**	**16.84**	**1 116**	**94.07**	**93.08**	**14.75**	**1 117**	**98.97**	**97.50**	**16.53**
ethyl acetate	482	211.68	204.54	47.83	482	179.69	176.11	35.31	482	172.18	168.10	30.48	482	180.33	176.00	34.37
	445	**204.27**	**202.57**	**29.24**	**449**	**174.97**	**174.33**	**21.57**	**448**	**168.3**	**166.63**	**18.47**	**451**	**176.16**	**175.00**	**21.68**
hexane	1 096	334.31	308.26	92.98	1 096	266.93	251.81	62.02	1 096	245.85	233.63	52.24	1 096	264.00	249.00	61.29
	954	**310.94**	**301.05**	**59.54**	**954**	**249.73**	**246.35**	**36.49**	**959**	**231.9**	**230.34**	**30.87**	**954**	**246.87**	**243.00**	**35.63**
iso-propanol	175	164.25	155.82	37.95	175	136.55	131.48	23.39	175	131.98	128.33	21.15	175	137.24	133.00	22.89
	160	**155.98**	**154.61**	**23.09**	**163**	**132.64**	**130.38**	**17.61**	**164**	**128.8**	**127.43**	**16.02**	**164**	**133.64**	**131.50**	**17.16**
methanol	4 499	89.32	80.65	37.41	4 499	66.66	61.80	29.01	4 499	62.73	58.82	25.42	4 499	67.44	62.50	28.43
	2 725	**86.71**	**83.52**	**18.34**	**2 746**	**67.39**	**65.61**	**11.24**	**2 746**	**64.02**	**62.90**	**9.55**	**2 743**	**68.03**	**66.25**	**11.16**
pentane	323	278.14	260.49	96.17	323	229.07	214.47	85.19	323	211.77	199.16	77.66	323	228.80	214.42	84.41
	266	**253.33**	**246.69**	**38.45**	**269**	**208.7**	**207.52**	**27.17**	**269**	**194.65**	**191.76**	**24.33**	**269**	**208.65**	**207.50**	**26.52**
thf	2 410	184.55	168.83	66.92	2 410	157.57	146.38	56.77	2 410	149.67	140.15	51.03	2 410	157.65	146.50	56.05
	2 126	**169.64**	**165.79**	**25.50**	**2 135**	**145.54**	**143.66**	**17.12**	**2 145**	**139.33**	**138.13**	**14.95**	**2 134**	**145.64**	**144.00**	**16.66**
toluene	2 446	244.68	230.48	67.89	2 446	209.92	200.30	57.37	2 446	199.48	190.71	52.05	2 446	210.24	200.50	56.89
	2 168	**230.51**	**226.28**	**31.29**	**2 166**	**198.36**	**197.15**	**20.65**	**2 167**	**189.55**	**188.46**	**18.18**	**2 169**	**198.89**	**197.50**	**20.73**

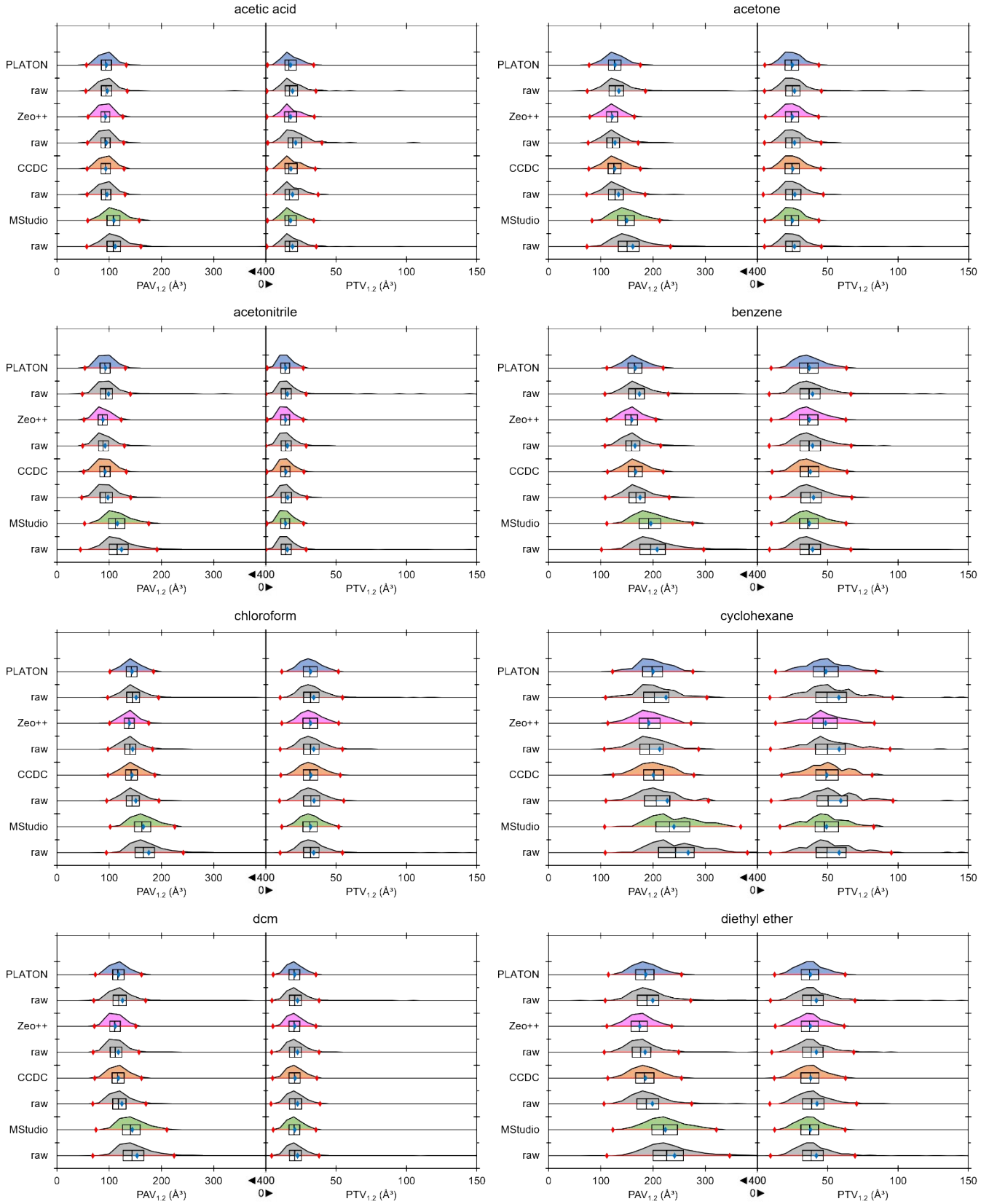

Fig. S2 Comparison of probe-surface volumes obtained using various software packages employing a 1.2 Å probe radius for 20 common solvents.

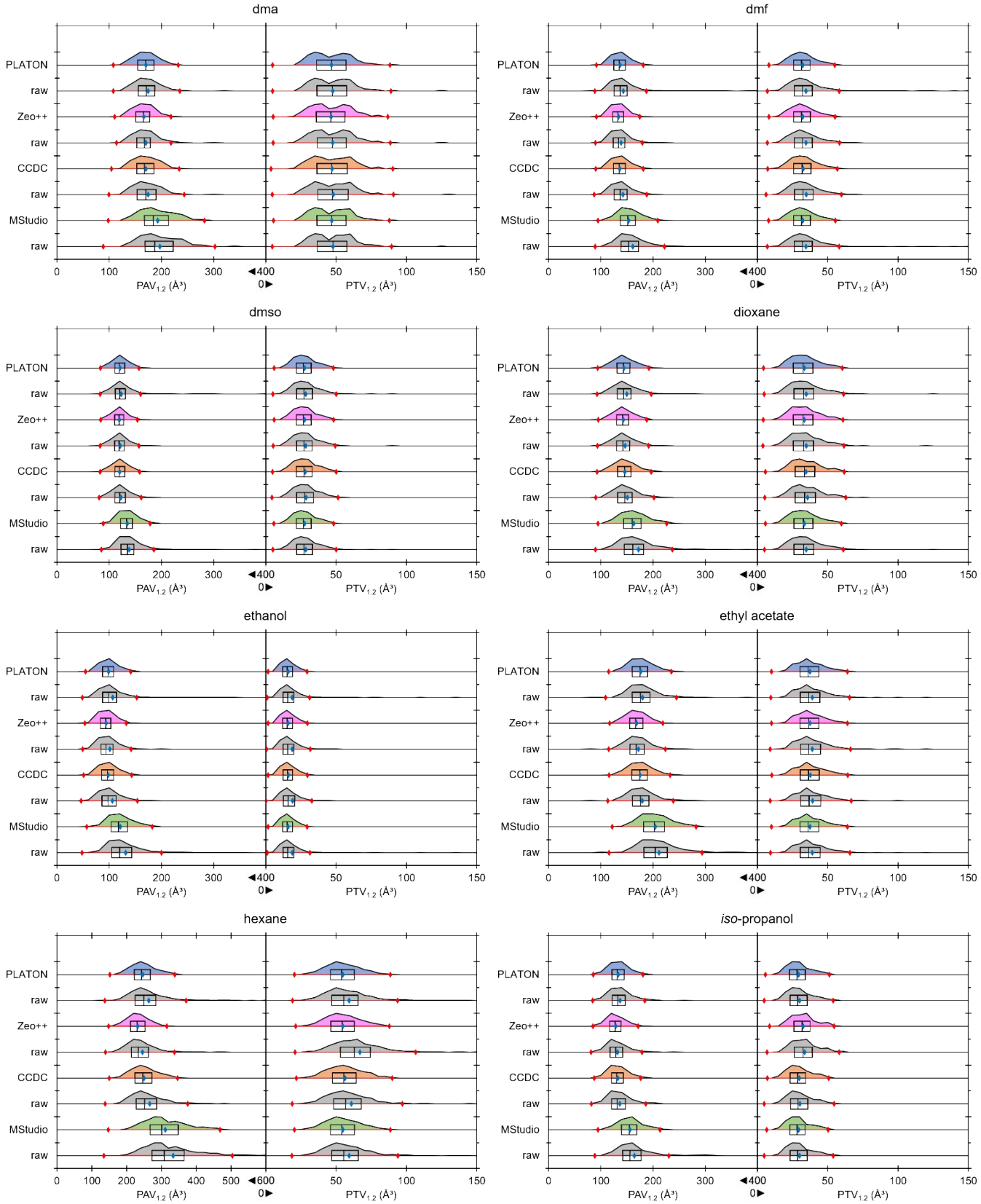

Fig. S2 continued.

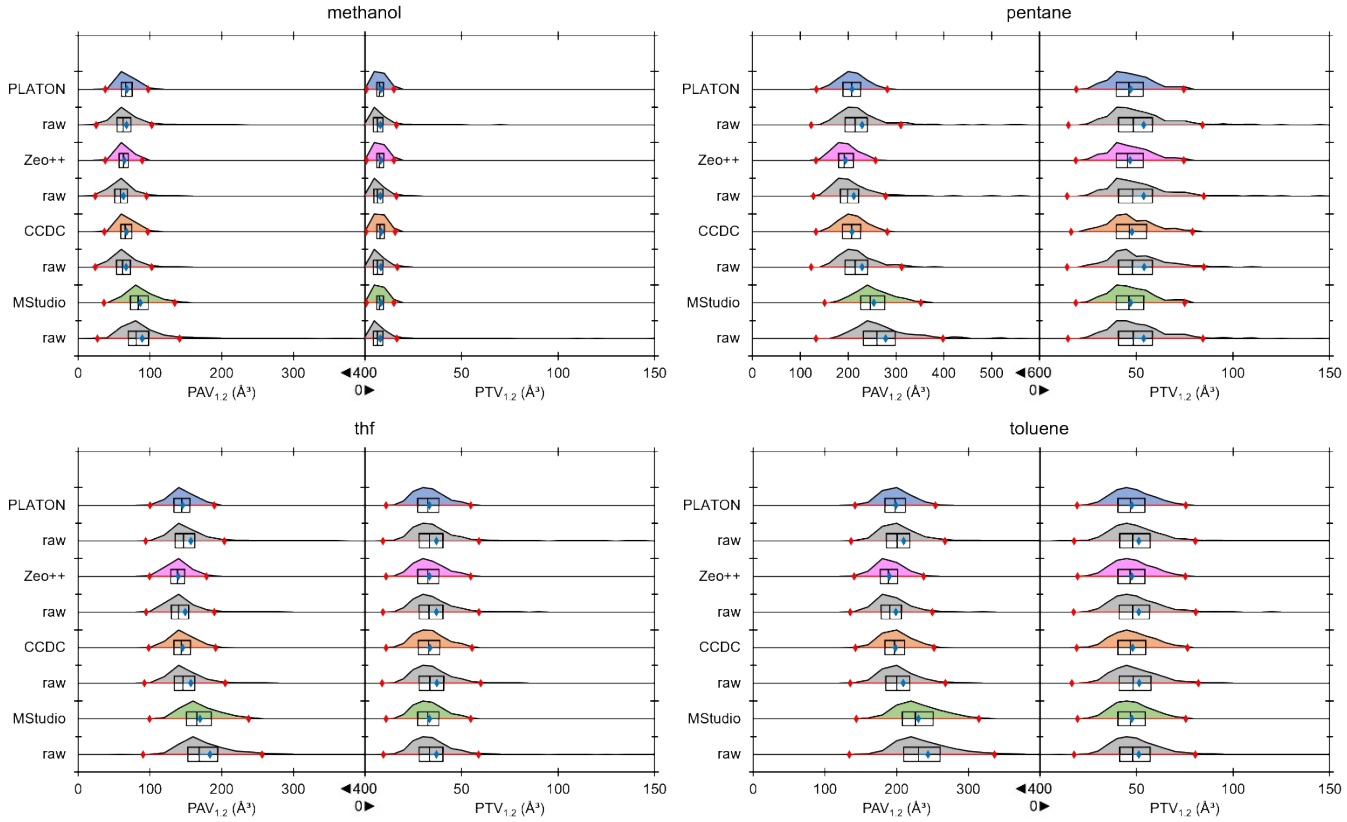

Fig. S2 continued.

8 Fractions summary

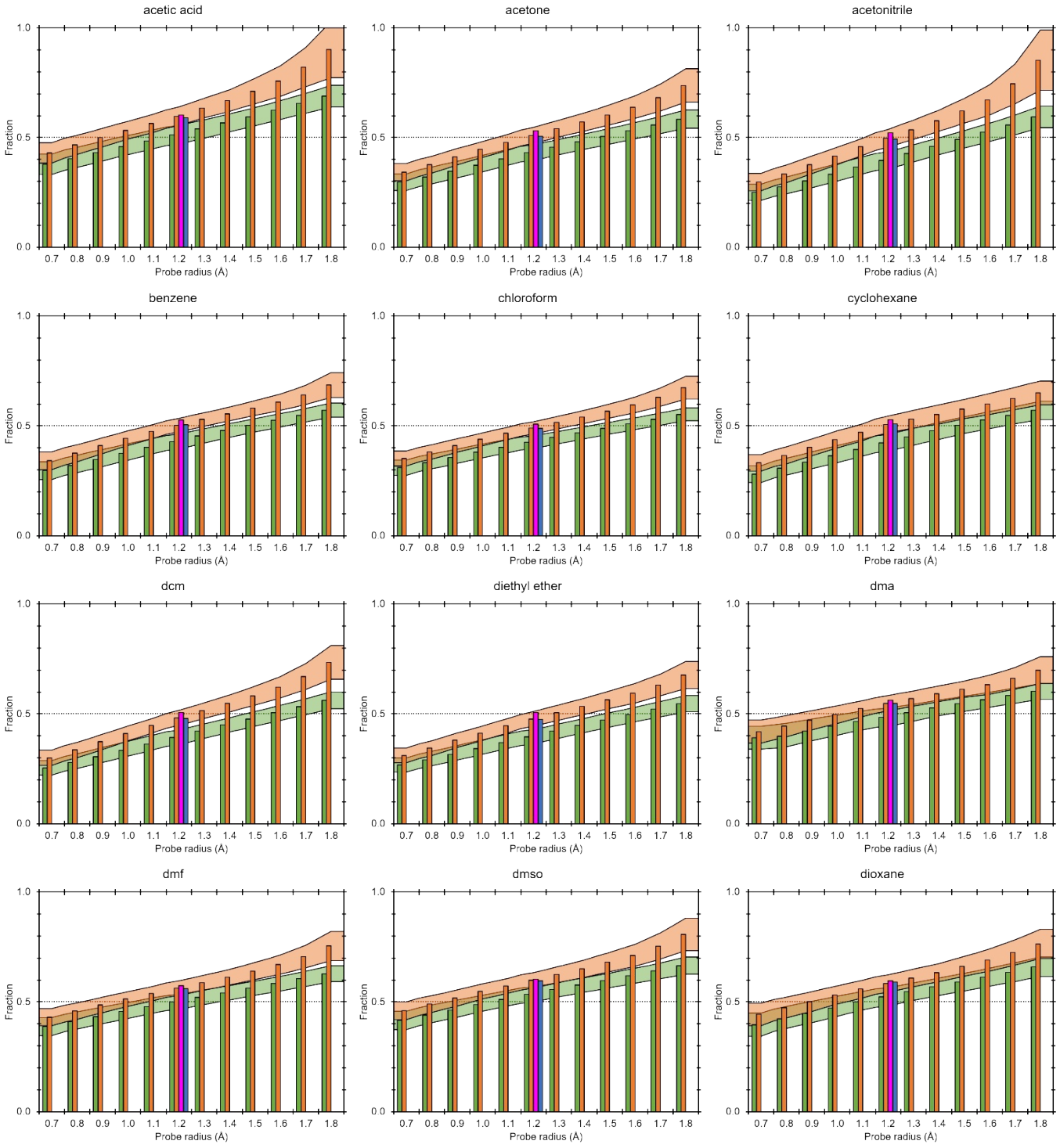

Fig. S3 Comparison of fraction of PAV occupied for 20 common solvents using MStudio (green), CCDC (orange), Zeo++ (pink) and PLATON (blue).

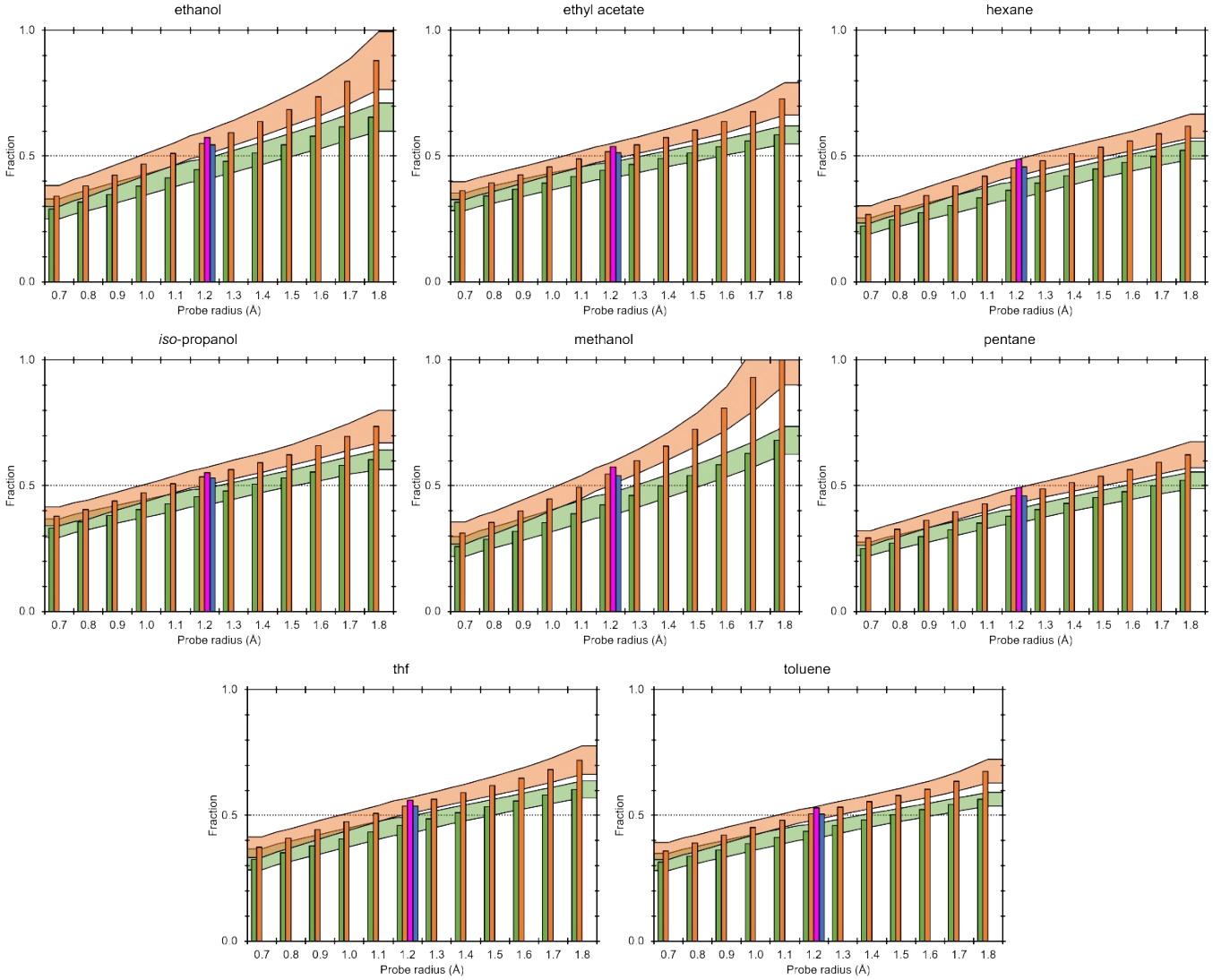

Fig. S3 continued.

9 Predictions

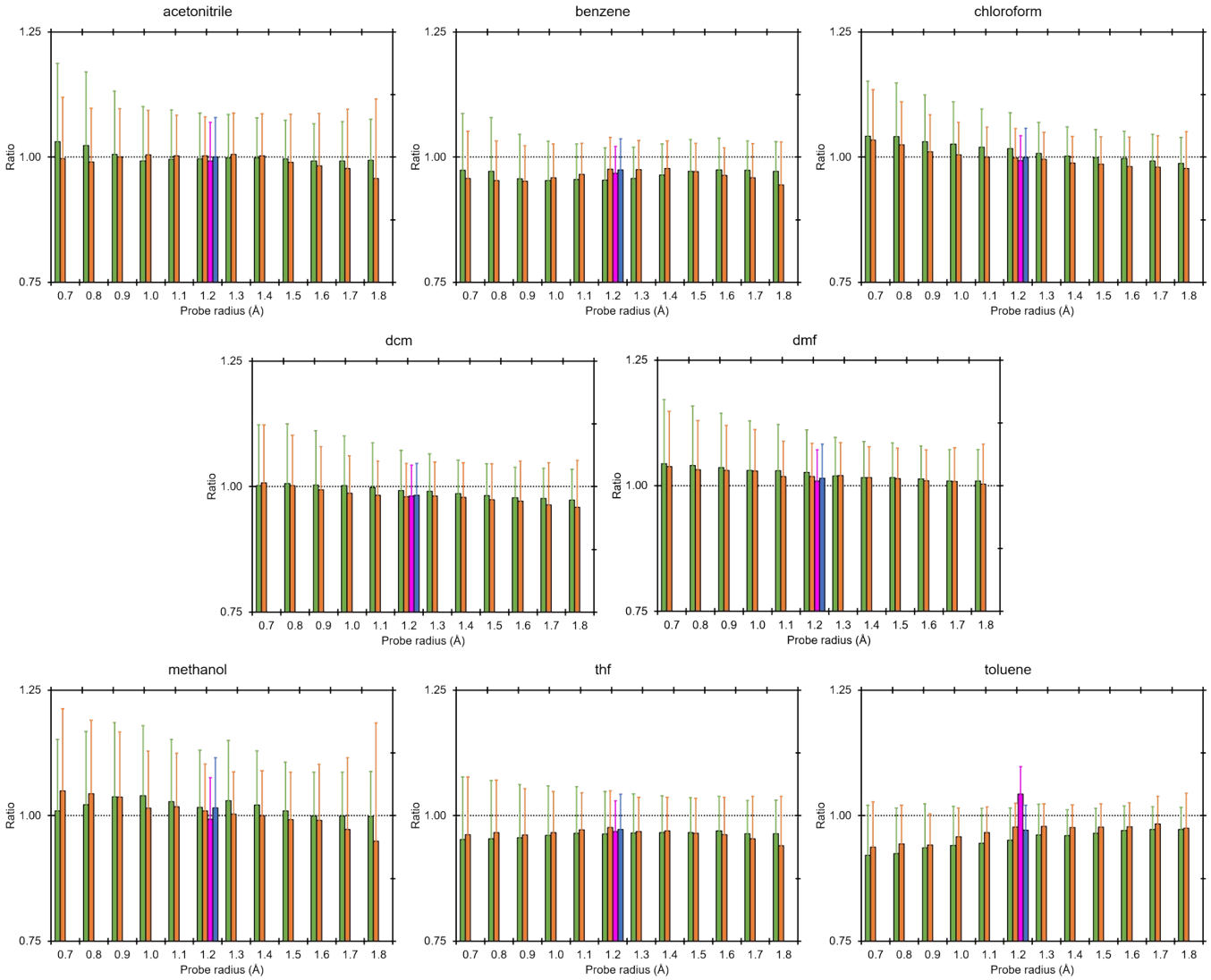

Fig. S4 Ratio of the estimated and observed number of guest molecules per unit cell employing various r_p in MStudio (green), CCDC (orange), Zeo++ (pink) and PLATON (blue). Error bars indicate the RMSE.

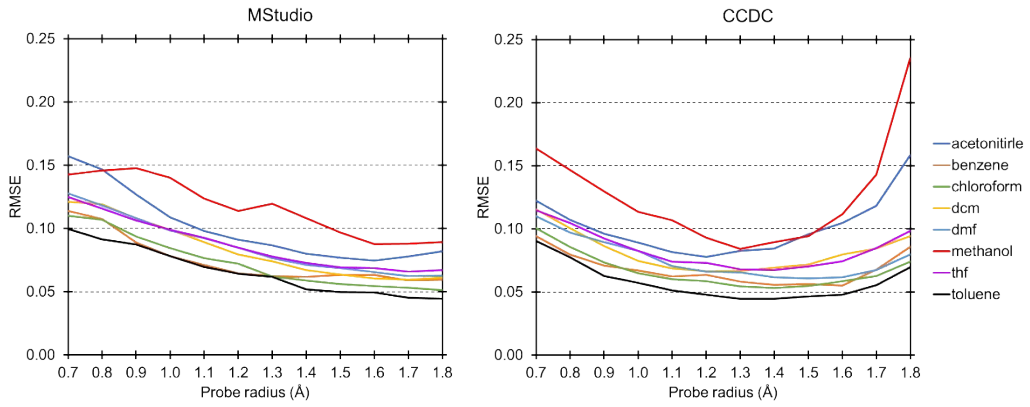

Fig. S5 RMSE on the $\frac{N_g^{est}}{N^{obs}_g}$ ratio determined using the MStudio and CCDC APIs for different probe radii.

10 Fractions for other solvents

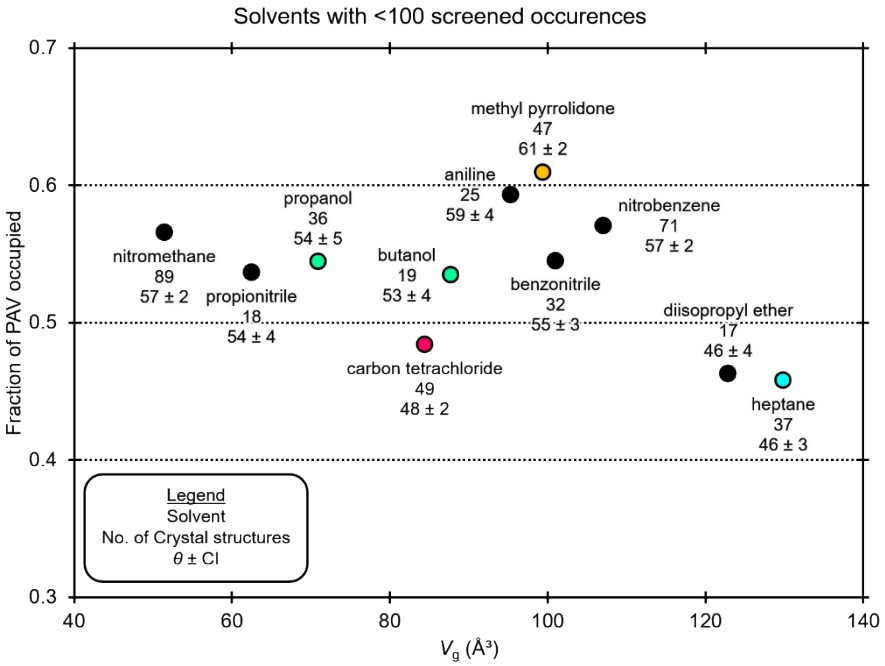

Fig. S6 CCDC-derived θ values with 99% confidence-level intervals as a function of V_g for eleven solvents with less than 100 statistically significant crystal structures after application of the filters mentioned in the main text. Markers are coloured loosely according to class of compound: hydrocarbons, cyan; alcohols, lime; chlorinated, magenta; carbonyl-containing, yellow.

11 Interaction energies

Hydrogen atom positions were optimised using the COMPASSII force field employing force field-assigned charges and Ewald summation for non-bond interactions. Only crystal structures with COMPASSII atom types for each atom were evaluated. Threshold values were the same as before (see §3). Single point energy evaluations were carried out with CASTEP[36] implemented in MStudio using the PBE density functional in combination with Grimme's DFT-D dispersion correction.[37] On-the-fly generated Vanderbilt-type ultrasoft pseudopotentials in combination with the Koelling-Harmon scalar-relativistic approach and a plane-wave expansion to an energy cut-off of 1.32×10^4 kcal mol^{-1}. Integration in the reciprocal lattice was performed using a Monkhorst-Pack grid with a 0.04 Å^{-1} k-point separation and self-consistent field convergence was set to 2.3×10^{-5} kcal mol^{-1}. A 50% admixture of the charge density was applied in conjunction with a DIIS (direct inversion in an iterative subspace, Pulay scheme) size of 20 to speed up convergence.

12 Disorder test

Table S8 Number of crystal structures from a survey of the CSD for disordered solvates. Useable structures contained no unbound atoms, while only structures with two unique chemical units and "solvate" or "clathrate" in the chemical name were evaluated. RMSE are given for crystal structures in which solvent molecules were identified.

Solvent	No. of crystal structures	Useable	Evaluated	Guest identified	RMSE^{raw} (molc UC^{-1})			
					MStudio	CCDC	Zeo++	PLATON
acetone	3 583	644	200	73	1.11	1.19	1.07	1.25
acetonitrile	13 195	1 630	357	140	1.68	1.99	1.97	1.96
benzene	4 592	706	311	191	1.00	1.07	0.97	1.12
chloroform	7 175	995	435	190	1.18	1.26	1.14	1.31
dcm	18 880	3078	1 240	389	0.88	0.97	0.85	1.06
diethyl ether	5 903	945	310	72	0.90	0.97	0.89	0.95
dmf	7 955	2 242	570	105	1.83	2.00	1.89	2.03
ethanol	4 970	729	250	68	0.67	0.70	0.64	0.76
hexane	4 070	791	437	113	0.65	0.70	0.62	0.77
methanol	12 767	1 909	557	199	0.99	1.13	1.17	1.11
thf	12 479	1 826	618	183	1.41	1.53	1.39	1.61
toluene	9 621	1 874	1 017	230	1.23	1.39	1.42	1.38
	105 190	17 369	6 302	1 953	⟨1.13⟩	⟨1.24⟩	⟨1.17⟩	⟨1.28⟩

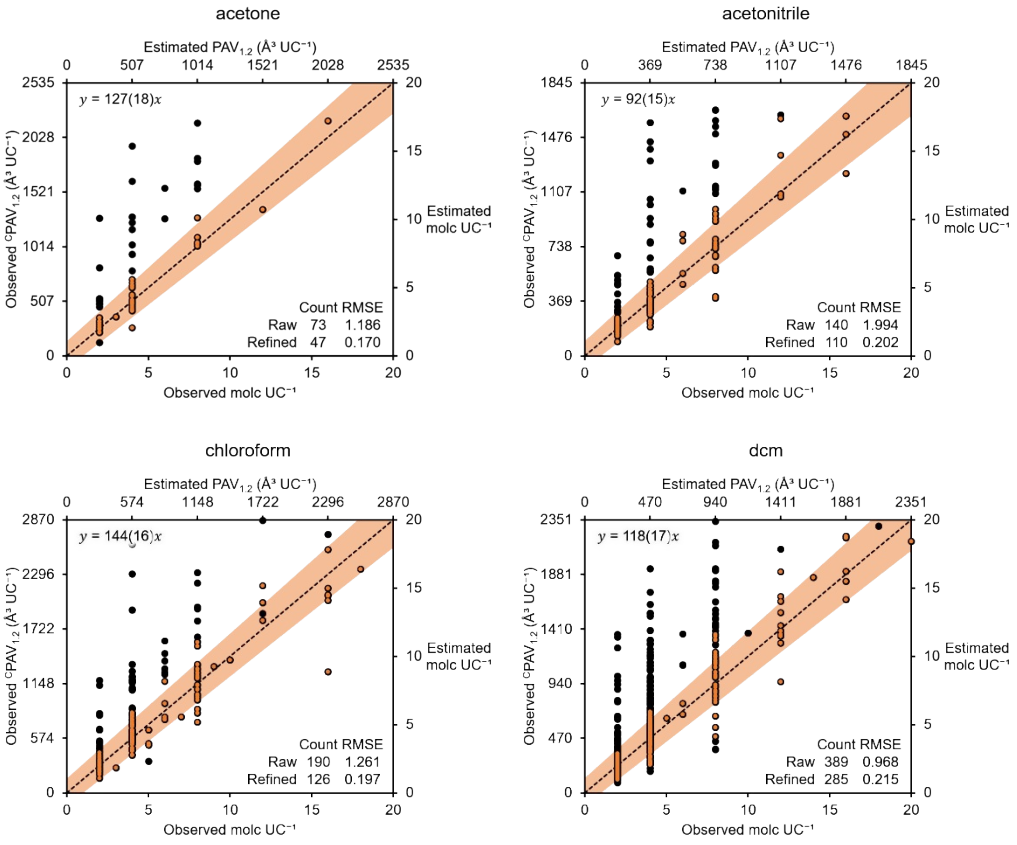

Fig. S7 Prediction of the number of guest molecules per unit cell in disorder-flagged crystal structures employing CCDC-derived fractions of $PAV_{1.2}$ occupied. Surfaces indicate one standard deviation from the dashed line-of-estimation and the number of structures before (Raw, black markers) and after (Refined, coloured markers) omission of crystal structures with $PTV_{1.2,complex}$ >50 Å³ UC^{-1}.

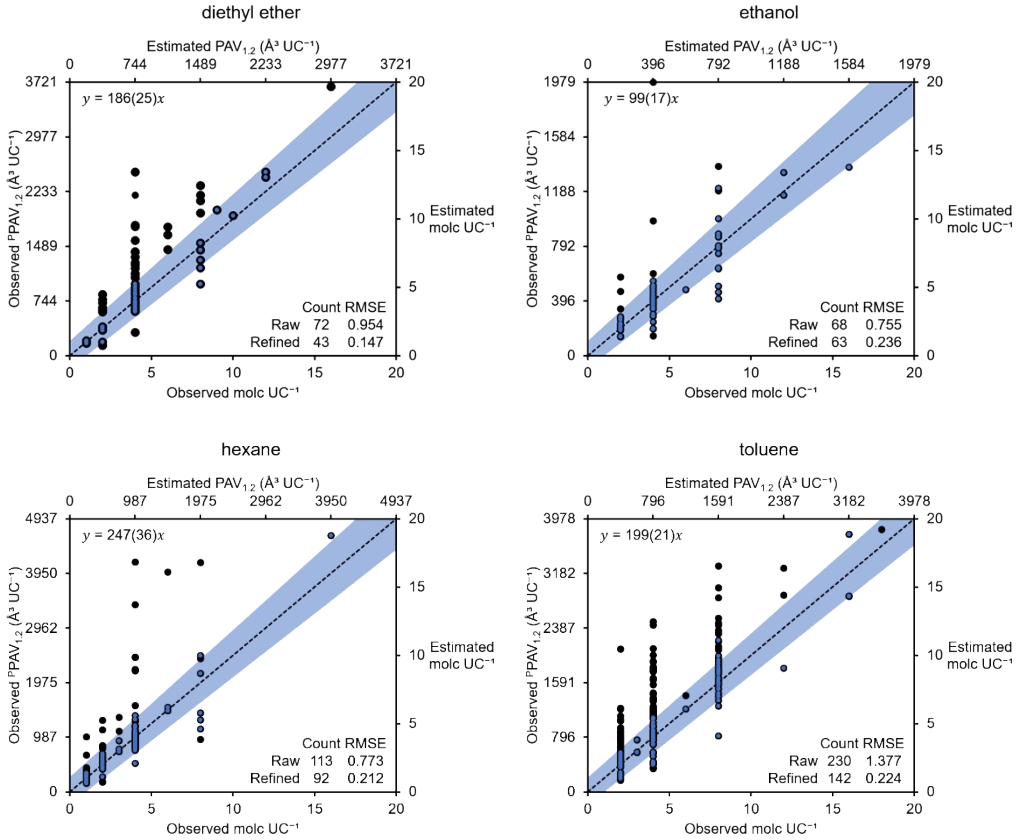

Fig. S8 Prediction of the number of guest molecules per unit cell in disorder-flagged crystal structures employing PLATON-derived fractions of $PAV_{1.2}$ occupied. Surfaces indicate one standard deviation from the dashed line-of-estimation and the number of structures before (Raw, black markers) and after (Refined, coloured markers) omission of crystal structures with $PTV_{1.2,complex}$ >50 Å³ UC⁻¹.

13 Software comparison

Table S9 Comparison of $PAV_{1.2}$ (Å³ UC⁻¹) generated employing different software suites.

Refcode	Guest	V_{mol}	N_{guest}	$PAV^{est}_{1.2}$[a]	MStudio	CCDC	Zeo++[b]	PLATON	X-Seed/MSRoll[c]	Olex2
NOJZOC	methanol	36.814	2	134.78	212.7	155.41	157.56	152	157.04	164.4
NOJZUI	acetonitrile	45.855	8	738.11	905.88	723.12	629.21	688	756.38	696
COZWAQ	dcm	56.625	4	470.21	735.69	506.7	494.49	514	–	528.2
COZWEU	def[d]	110.659	2	–	421.46	394.44	**410.94**	399	–	411.8
COZWIY	dma	93.438	4	681.66	869.13	721.06	**584.05**	728	–	789
COZWOE	dmf	76.755	2	272.96	352.47	325.44	**339.12**	329	–	343.2
COZWUK	dmso	71.892	2	240.36	327.32	280.41	299.69	289	314.83	304.9
COZXAR	thf	78.114	4	582.18	804.98	607.62	563.35	628	–	660.6

[a] Calculated as: $V_{mol} \div \theta \times N_g$ employing CCDC-derived θ values for methanol (0.55), acetonitrile (0.50), dcm (0.48), dma (0.55), dmf (0.56), dmso (0.60) and thf (0.54).

[b] Values in boldface and lightface indicate accessible and non-accessible probe-occupiable volume, respectively.

[c] Only cases that successfully trap the probe yield sensible surface volumes.

[d] *N*,*N*–diethylformamide

The over- and underestimation of PAV by MStudio and Zeo++, respectively, is apparent for most cases in Table S9.

Table S10 Comparison of $PTV_{1.2}$ (Å³ UC⁻¹) generated employing different software suites.

Refcode	Guest	V_{mol}	N_g	MStudio	CCDC	Zeo++[a]	PLATON	X-Seed/MSRoll[b]
NOJZOC	methanol	36.814	2	18.05	18.06	20.08	18.1	18.04
NOJZUI	acetonitrile	45.855	8	134.51	145.17	142.35	134	134.47
COZWAQ	dcm	56.625	4	87.37	88.03	92.81	87.6	87.41
COZWEU	def[c]	110.659	2	110.11	111.96	**115.28**	110.4	–
COZWIY	dma	93.438	4	117.94	114.31	**119.38**	117.8	–
COZWOE	dmf	76.755	2	70.24	68.29	**74.478**	70.2	–
COZWUK	dmso	71.892	2	53.89	51.98	57.29	53.9	53.93
COZXAR	thf	78.114	4	108.23	107.82	118.83	84.4	108.04

[a] Values in boldface and lightface indicate accessible (channel) and non-accessible (pocket) "probe-occupiable" volume, respectively.

[b] Only cases that successfully trap the probe yield sensible surface volumes; i.e., PTS delineated as pockets by Zeo++.

[c] *N*,*N*–diethylformamide

Table S11 Comparison of surface volumes (Å³ UC⁻¹) generated by different software suites employing r_p = 1.2 Å.

Refcode	Guest	V_{mol}	N_g	RPluto	ATOMS	CrystalExplorer	V^p
NOJZOC	methanol	36.814	2	52.50	19.88439	363.49	34.42
NOJZUI	acetonitrile	45.855	8	340.25	246.5558	1650.79	193.84
COZWAQ	dcm	56.625	4	160.29	129.2669	1069.02	135.38
COZWEU	def[c]	110.659	2	175.92	125.6539	440.23	159.50
COZWIY	dma	93.438	4	178.55	175.0481	1112.66	182.92
COZWOE	dmf	76.755	2	97.50	83.22811	374.88	111.72
COZWUK	dmso	71.892	2	75.19	70.49202	359.21	82.83
COZXAR	thf	78.114	4	202.26	174.3585	1117.39	161.70

X-Seed allows the user to interactively generate a non-periodic *.pdb file that can be parsed to MSRoll for surface analysis. If there are no pockets capable of confining the probe then only the exterior surface of the host structure will be mapped. RPluto and ATOMS employ an algorithm first proposed by Gavezzotti; to avoid the cumbersome geometric spheres-and-caps method of calculating the total volume of spheres less the volume of intersecting gaps,[1] he proposed enclosing the system in a box of known volume and the molecular volume is then readily determined using the ratio of grid points that lies within the constituting atoms' r_{vdW}:[38]

$$V_{\mathrm{mol}} = V_{\mathrm{box}} \frac{N_{\mathrm{occupied}}}{\mathrm{N_{box}}} \qquad \text{S2}$$

Whereas the "Crystal Voids" calculation of CrystalExplorer uses a 0.0134 e Å⁻³ promolecule isodensity value, *ab initio* determined V^p values are volumes circumscribed by the 0.000675 e Å⁻³ isosurface. Only for the determination of V^p were hydrogen atom positions optimised using CASTEP implemented in MStudio using the PBE density functional in combination with Grimme's DFT-D dispersion correction. On-the-fly generated Vanderbilt-type ultrasoft pseudopotentials in combination with the Koelling-Harmon scalar-relativistic approach and a plane-wave expansion

to an energy cut-off of 1.13×10^4 kcal mol^{-1}. Integration in the reciprocal lattice was performed using a Monkhorst-Pack grid with a 0.05 Å^{-1} k-point separation and self-consistent field convergence was set to 4.61×10^{-5} kcal mol^{-1}. Convergence tolerances for geometry optimization using the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm[39] were set to 4.61×10^{-4} kcal mol^{-1} atom^{-1}, 1.15 kcal mol^{-1} Å^{-1} and 2.0×10^{-3} Å on energy, maximum force and maximum displacement, respectively.

14 References

1 A. I. Kitaigorodsky, *Molecular Crystals and Molecules*, Academic Press, New York, NY, 1973.
2 H. H. Cady, *Estimation of the Density of Organic Explosives from their Structural Formulas*, Los Alamos, New Mexico, 1975.
3 D. W. M. Hofmann, *Acta Crystallogr.*, 2002, **B58**, 489–493.
4 A. Immirzi and B. Perini, *Acta Crystallogr.*, 1977, **A33**, 216–218.
5 J. R. Stein, *Prediction of Crystal Densities of Organic Explosives by Group Additivity*, Los Alamos, New Mexico, 1981.
6 H. L. Ammon, *Propellants, Explos. Pyrotech.*, 2008, **33**, 92–102.
7 C. M. Tarver, C. L. Coon and J. M. Guimont, *Density estimation for new solid and liquid explosives*, 1977.
8 F. M. Richards, *Methods Enzymol.*, 1985, **115**, 440–464.
9 W. L. Bragg, *Philos. Mag.*, 1920, **40**, 169–189.
10 E. Mack Jr., *J. Am. Chem. Soc.*, 1932, **54**, 2141–2165.
11 L. Pauling, *Nature of the Chemical Bond*, Cornell University Press, Ithaca, NY, Second., 1948.
12 S.-Z. Hu, Z.-H. Zhou, Z.-X. Xie and B. E. Robertson, *Zeitschrift fuer Krist. - Cryst. Mater.*, 2014, **229**, 517–523.
13 S. S. Batsanov, *Inorg. Mater. (Translation Neorg. Mater.*, 2001, **37**, 871–885.
14 R. S. Rowland and R. Taylor, *J. Phys. Chem.*, 1996, **100**, 7384–7391.
15 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.
16 A. Bondi, *J. Phys. Chem.*, 1966, **70**, 3006–3007.
17 A. Y. Meyer, *J. Comput. Chem.*, 1988, **9**, 18–24.
18 P. L. Warburton, J. L. Wang and P. G. Mezey, *J. Chem. Theory Comput.*, 2008, **4**, 1627–1636.
19 R. F. W. Bader, W. H. Henneker and P. E. Cade, *J. Chem. Phys.*, 1967, **46**, 3341–3363.
20 R. F. W. Bader, M. T. Carroll, J. R. Cheeseman and C. Chang, *J. Am. Chem. Soc.*, 1987, **109**, 7968–7979.
21 J. S. Murray and P. Politzer, *J. Mol. Struct. THEOCHEM*, 1998, **425**, 107–114.
22 P. Politzer and J. S. Murray, *Struct. Chem.*, 2016, **27**, 401–408.
23 J. S. Murray and P. Politzer, *Croat. Chem. Acta*, 2009, **82**, 267–275.
24 P. Politzer and J. S. Murray, *Theor. Chem. Acc.*, 2002, **108**, 134–142.
25 M. Pospisil, P. Vavra, M. C. Concha, J. S. Murray and P. Politzer, *J. Mol. Model.*, 2010, **16**, 895–901.
26 M. Pospisil, P. Vavra, M. C. Concha, J. S. Murray and P. Politzer, *J. Mol. Model.*, 2011, **17**, 2569–2574.
27 N. Bouhmaida and N. E. Ghermani, *J. Chem. Phys.*, 2005, **122**, 114101.
28 A. Nirwan, A. Devi and V. D. Ghule, *J. Mol. Model.*, 2018, **24**, 166.
29 C. E. Webster, R. S. Drago and M. C. Zerner, *J. Am. Chem. Soc.*, 1998, **120**, 5509–5516.
30 D. R. Lide, Ed., *Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 88th edn., 2008.
31 B. Delley, *J. Chem. Phys.*, 1990, **92**, 508–517.
32 B. Delley, *J. Chem. Phys.*, 2000, **113**, 7756–7764.
33 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
34 H. Sun, Z. Jin, C. Yang, R. L. C. Akkermans, S. H. Robertson, N. A. Spenley, S. Miller and S. M. Todd, *J. Mol. Model.*, 2016, **22**, 47.
35 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
36 S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson and M. C. Payne, *Zeitschrift fuer Krist.*, 2005, **220**, 567–570.
37 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.
38 J. D. Dunitz, G. Filippini and A. Gavezzotti, *Tetrahedron*, 2000, **56**, 6595–6601.
39 B. G. Pfrommer, M. Cote, S. G. Louie and M. L. Cohen, *J. Comput. Phys.*, 1997, **131**, 233–240.