Supplementary Material

Implicit solvation parameters derived from explicit water forces in large-scale Molecular Dynamics simulations

Jens Kleinjung,[†] Walter R.P. Scott,[‡] Jane R. Allison,[¶] Wilfred F. van Gunsteren,[¶] and Franca Fraternali[*,§]

Division of Mathematical Biology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom, Chemistry Department, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia, Canada V6T 1Z1, Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH Hönggerberg, CH-8093 Zürich, Switzerland, and Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, London SE1 1UL, United Kingdom

E-mail: franca.fraternali@kcl.ac.uk

[*]To whom correspondence should be addressed
[†]NIMR London
[‡]University of British Columbia
[¶]ETH Zürich
[§]King's College London

Figure Legends

Figure S1 Distribution of atomic SASA values A_i for all GROMOS atom types included in this study.

Figure S2 Distribution of the size of the explicit solvent (water) forces $\mathbf{f}_i^{expl}$ for the GROMOS atom types and for different data sets. The grey distributions show the forces on <ALL> atoms. The overlaid coloured distributions show subsets: the <SA> forces in orange, the <SA & θ+> forces in blue. Insets show the <SA & θ+> data in a uniform scale of 10^4 data points for a quantitative comparison of the atom type frequency. The selected atom types are the same as in Figure S1.

Figure S3 Distribution of the angle θ_i between the explicit and implicit force vectors for the GROMOS atom types. The atom type selection, data subsets and colour scheme are the same as in Figure S2.

Figure S4 Distribution of the subset $\sigma^{SASA}_{<SA>}$ values as function of the atomic SASA value A_i for the GROMOS atom types.

Figure S5 Distribution of the subset $\sigma^{SASA}_{<SA\ \&\ \theta+>}$ values as function of the atomic SASA value A_i for the GROMOS atom types.

Figure S6 Q–Q plots showing the log-normal behaviour of <ALL> water forces. Sampled water forces $\mathbf{f}^{smp}_{<ALL>}$ are plotted over a theoretical log–normal distribution $\mathbf{f}^{theo}_{<ALL>}$. $\mathbf{f}^{theo}_{<ALL>}$ are 1000 points from a distribution with 'mean' and 'sd' as estimated from a fit of the $\mathbf{f}^{expl}_{<ALL>}$ distribution. $\mathbf{f}^{smp}_{<ALL>}$ are 1000 points sampled from the $\mathbf{f}^{expl}_{<ALL>}$ distribution. An ideal log–normal distribution would follow the diagonal (solid line). Plots are scaled to the range of frequently occurring forces.

Figure S7 Q–Q plots of water forces of the <SA> subset. The selection on SASA does not change the log–normal shape of the distribution.

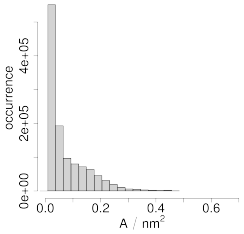
Figure S8 Q–Q plots of the <SA & θ+> water forces. The selection on SASA and angle range does not change the log–normal shape of the distribution.

Figure S9 Q–Q plots of re–sampled <SA & θ+> forces over <SA> forces. The distribution of the <SA & θ+> forces were re–sampled to fit the 'mean' and 'sd' values of the <SA> force distribution. The re–sampled distributions are both log–normal.

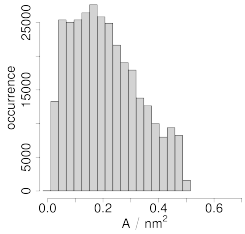
Figure S10 Ratio I_{norm} of the Mutual Information and the Joint Entropy between the GROMOS atom type classification and the atom groups derived by partitioning. The partitions were computed on the basis of σ_i^{SASA} value ranges (bins) of $1\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. The partitioning shows a relative information maximum at 3 groups.

Figure S11 Local structural properties of the test proteins trp (a,b), drk (c,d), ubq (e,f) and lys (g,h). in implicit solvent (a,c,e,g) and water (b,d,f,h). Colour scheme: red, helical (including α–helix); blue: extended (including β–strand); green-yellow: turns and loops. Conformations were plotted every 25 ps.

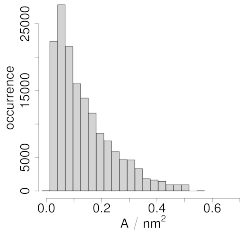
Figure S1



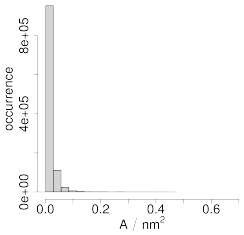
(a) atom type 1 O



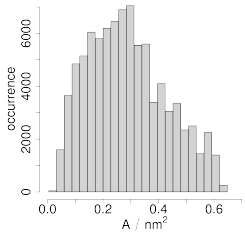
(b) atom type 2 OM



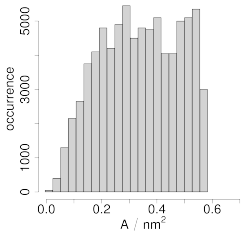
(c) atom type 3 OA



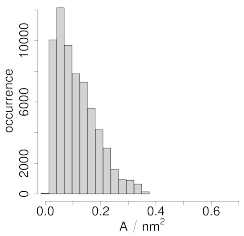
(d) atom type 5 N



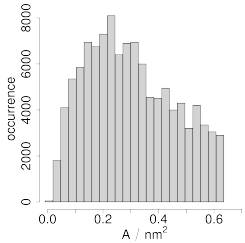
(e) atom type 6 NT



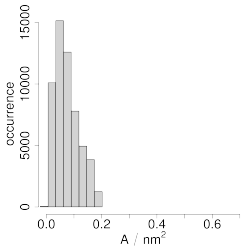
(f) atom type 7 NL



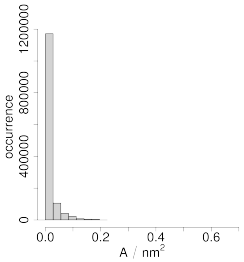
(g) atom type 8 NR



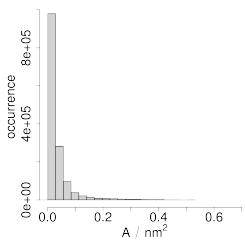
(h) atom type 9 NZ



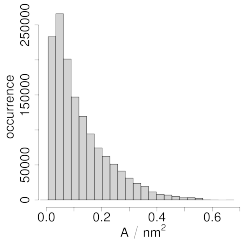
(i) atom type 10 NE



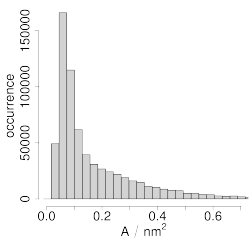
(j) atom type 11 C



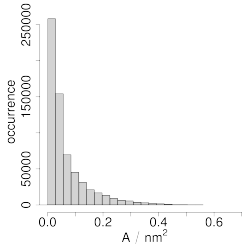
(k) atom type 12 CH



(l) atom type 13 CH2

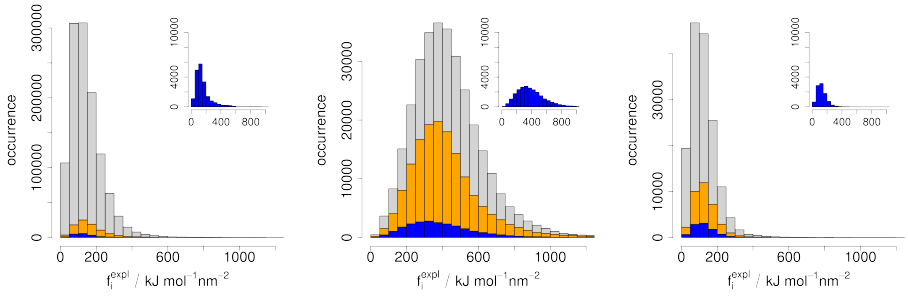


(m) atom type 14 CH3

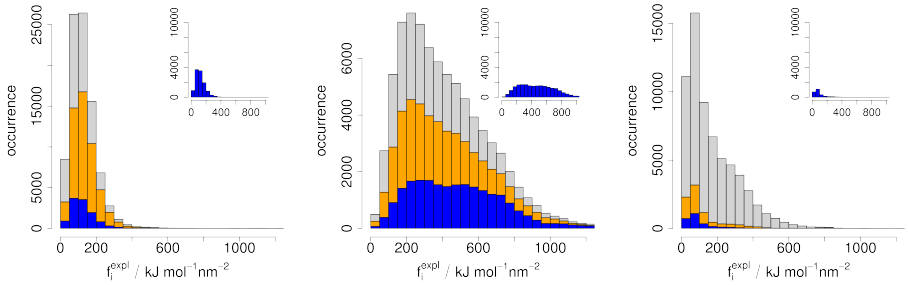


(n) atom type 16 CR1

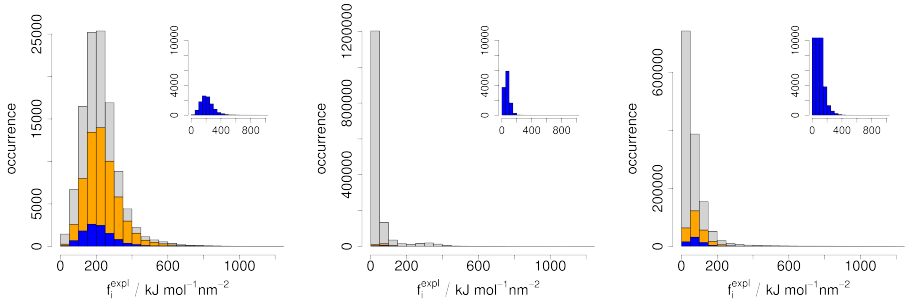
Figure S2



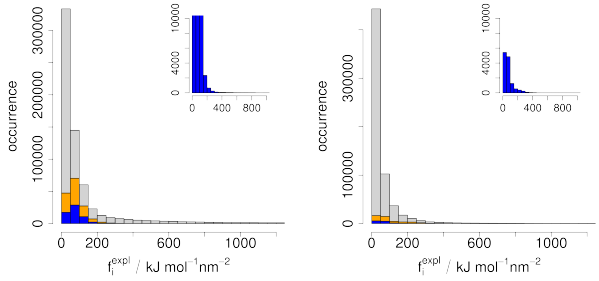
(a) atom type 1 O (b) atom type 2 OM (c) atom type 3 OA



(d) atom type 6 NT (e) atom type 7 NL (f) atom type 8 NR

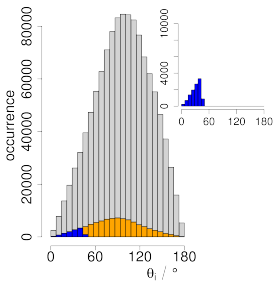


(g) atom type 9 NZ (h) atom type 12 CH (i) atom type 13 CH2

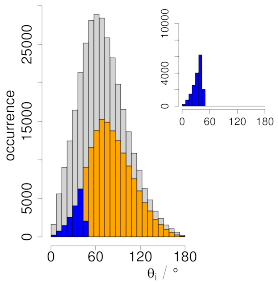


(j) atom type 14 CH3 (k) atom type 16 CR1

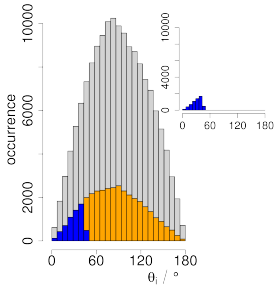
Figure S3



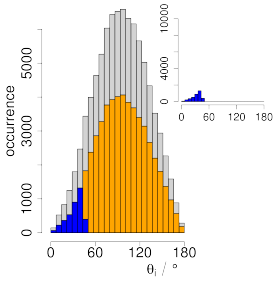
(a) atom type 1 O



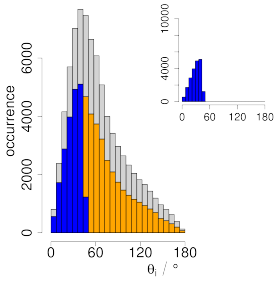
(b) atom type 2 OM



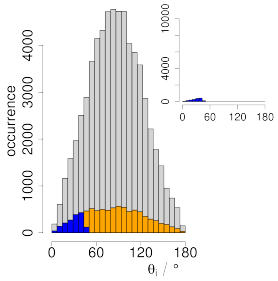
(c) atom type 3 OA



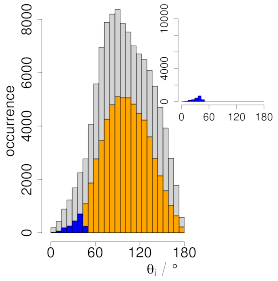
(d) atom type 6 NT



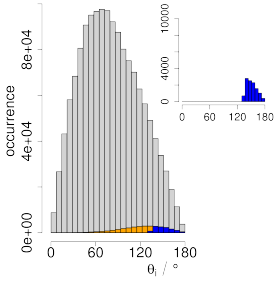
(e) atom type 7 NL



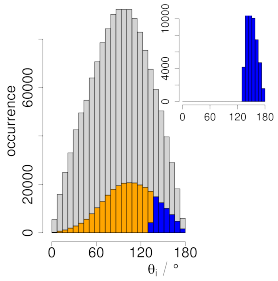
(f) atom type 8 NR



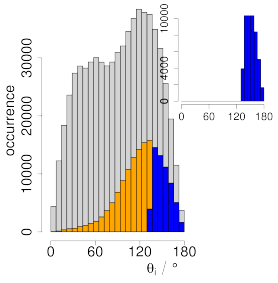
(g) atom type 9 NZ



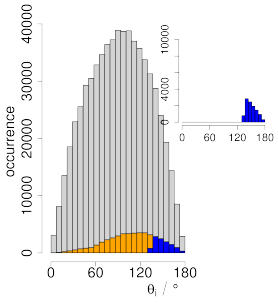
(h) atom type 12 CH



(i) atom type 13 CH2

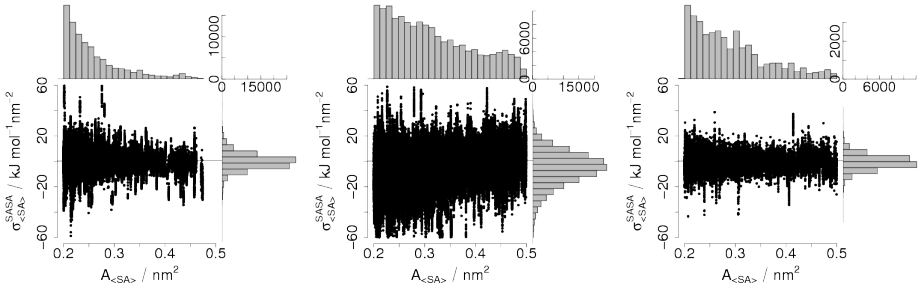


(j) atom type 14 CH3



(k) atom type 16 CR1

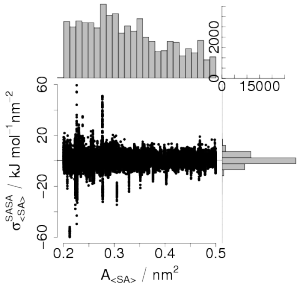
Figure S4



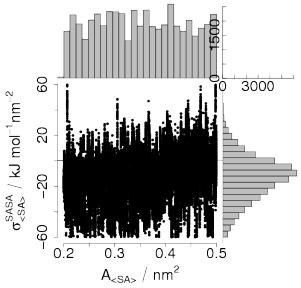
(a) atom type 1 O

(b) atom type 2 OM

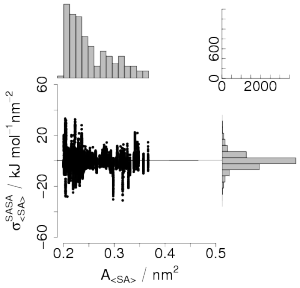
(c) atom type 3 OA



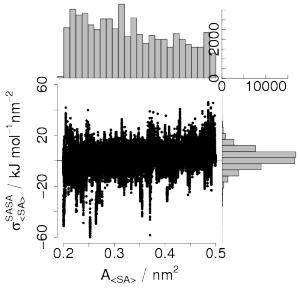
(d) atom type 6 NT



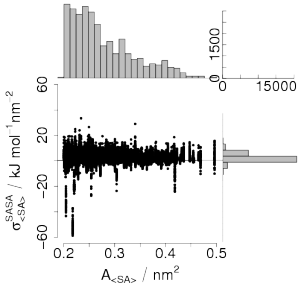
(e) atom type 7 NL



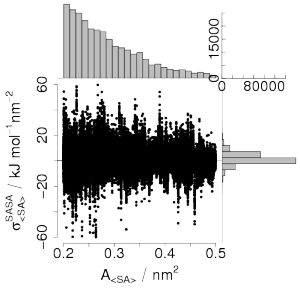
(f) atom type 8 NR



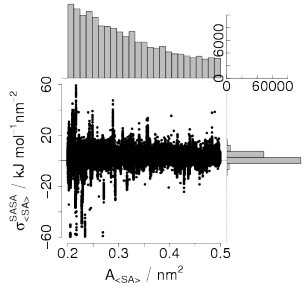
(g) atom type 9 NZ



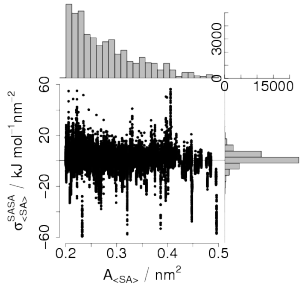
(h) atom type 12 CH



(i) atom type 13 CH2

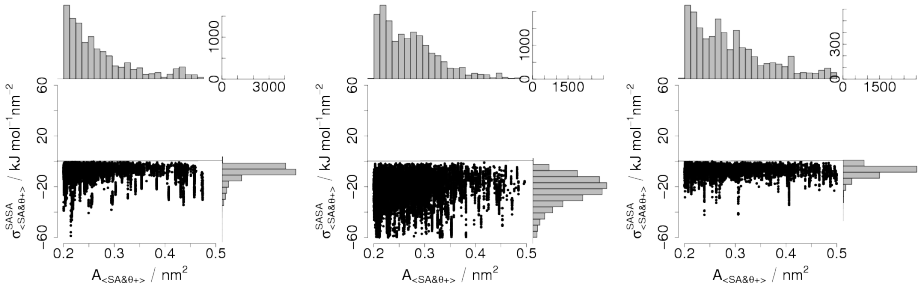


(j) atom type 14 CH3



(k) atom type 16 CR1

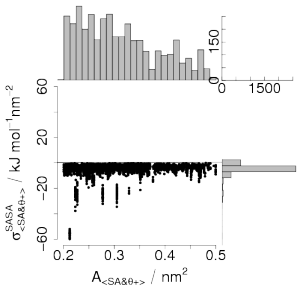
Figure S5



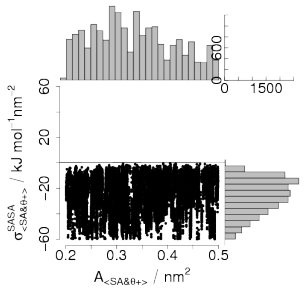
(a) atom type 1 O

(b) atom type 2 OM

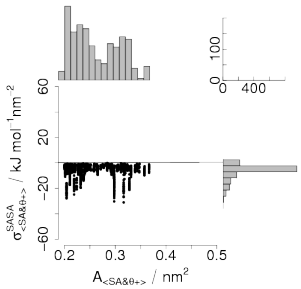
(c) atom type 3 OA



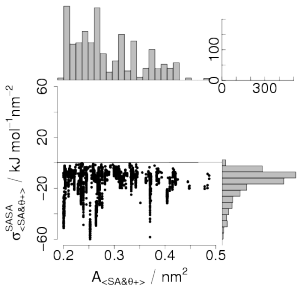
(d) atom type 6 NT



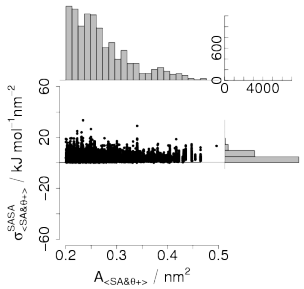
(e) atom type 7 NL



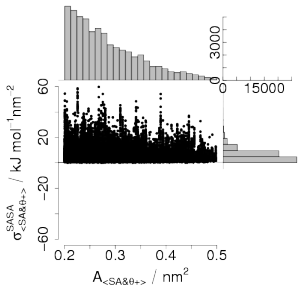
(f) atom type 8 NR



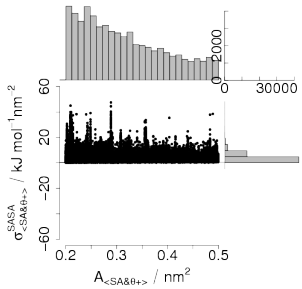
(g) atom type 9 NZ



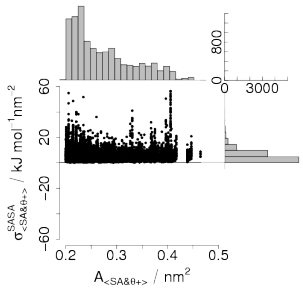
(h) atom type 12 CH



(i) atom type 13 CH2

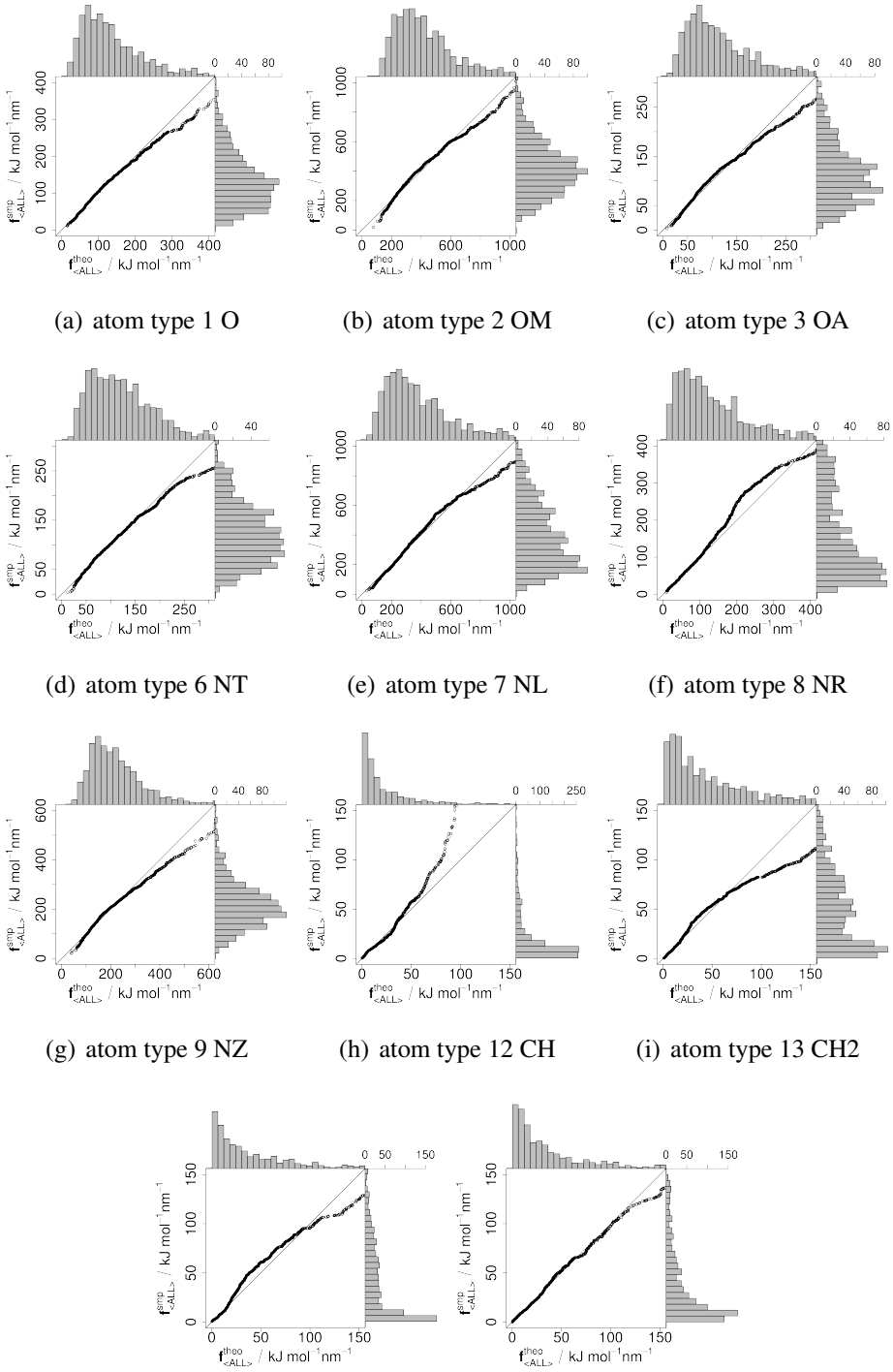


(j) atom type 14 CH3



(k) atom type 16 CR1

Figure S6



(a) atom type 1 O (b) atom type 2 OM (c) atom type 3 OA

(d) atom type 6 NT (e) atom type 7 NL (f) atom type 8 NR

(g) atom type 9 NZ (h) atom type 12 CH (i) atom type 13 CH2

(j) atom type 14 CH3 (k) atom type 16 CR1

Figure S7

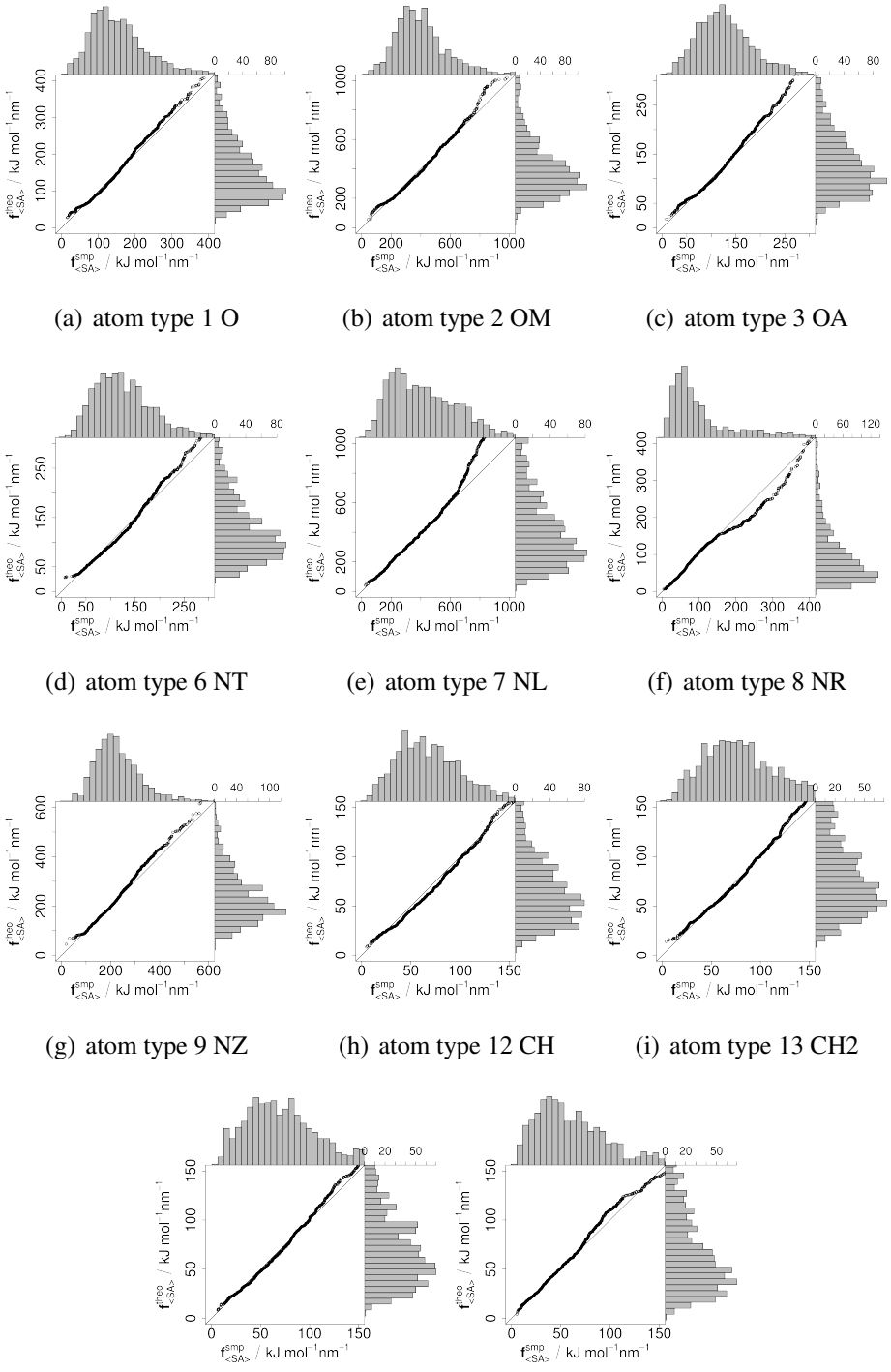
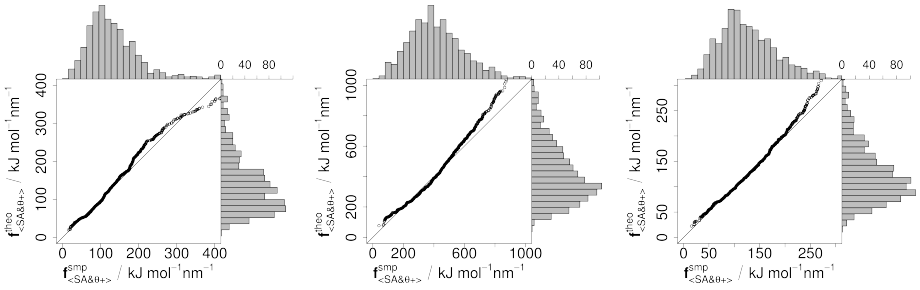
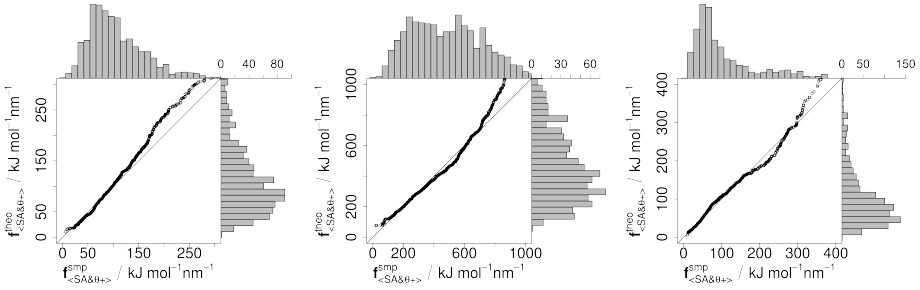

(a) atom type 1 O
(b) atom type 2 OM
(c) atom type 3 OA
(d) atom type 6 NT
(e) atom type 7 NL
(f) atom type 8 NR
(g) atom type 9 NZ
(h) atom type 12 CH
(i) atom type 13 CH2
(j) atom type 14 CH3
(k) atom type 16 CR1

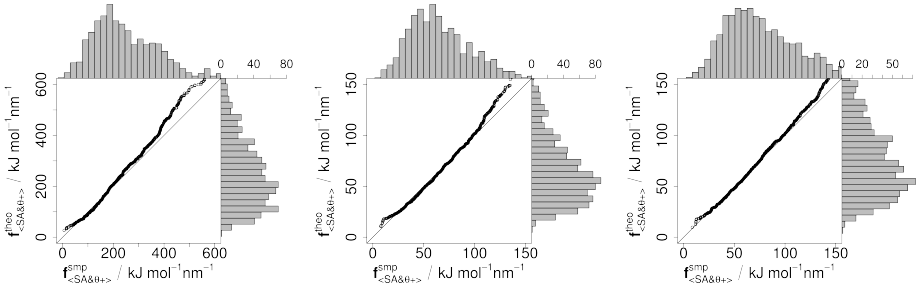
Figure S8



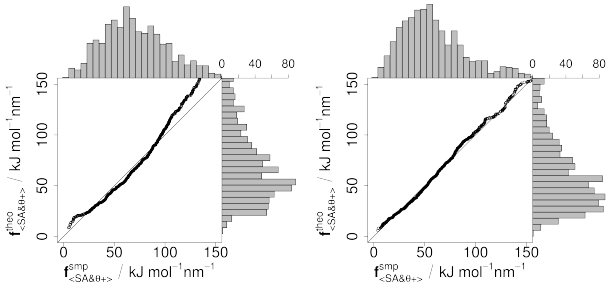
(a) atom type 1 O (b) atom type 2 OM (c) atom type 3 OA



(d) atom type 6 NT (e) atom type 7 NL (f) atom type 8 NR

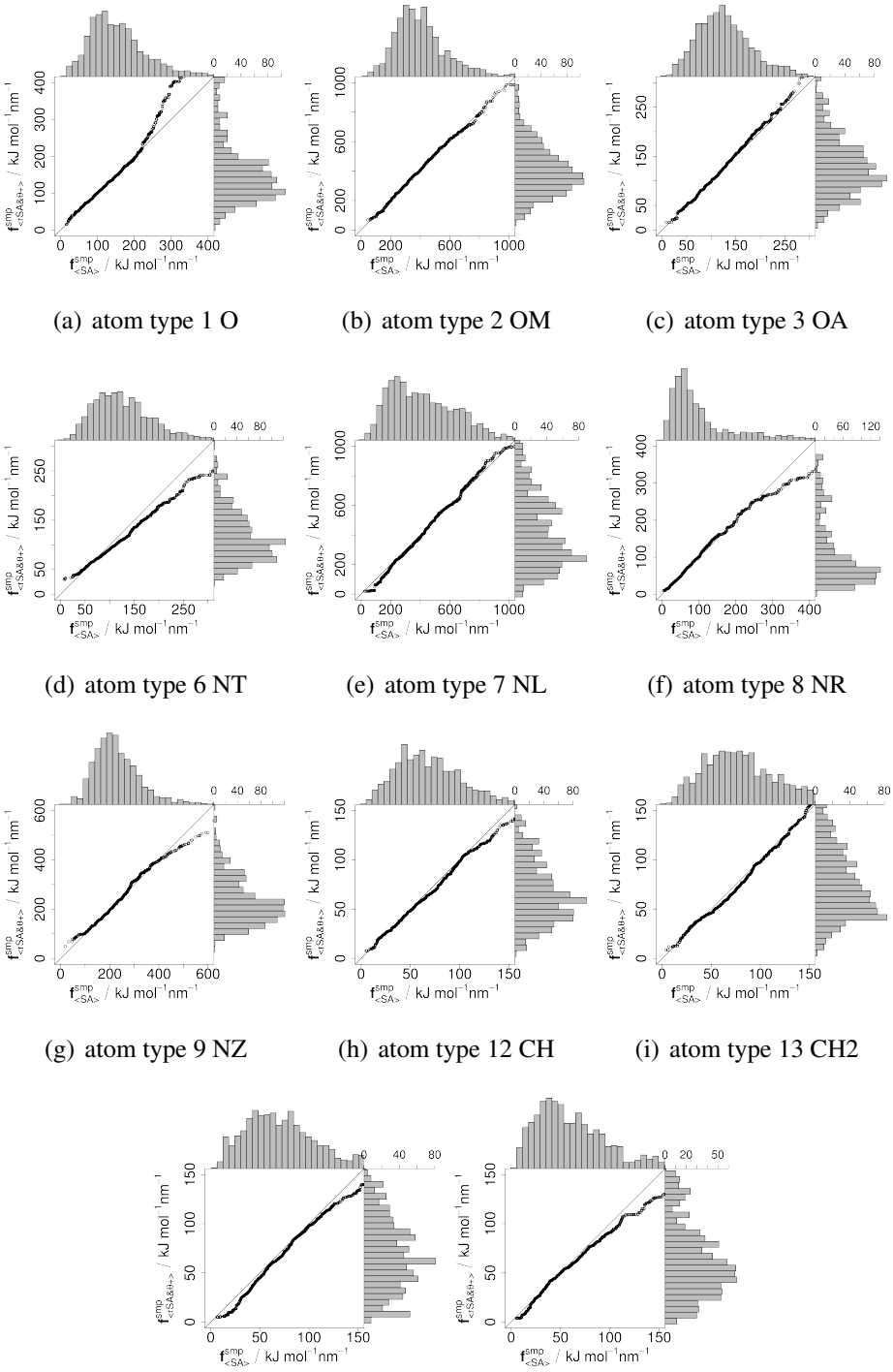


(g) atom type 9 NZ (h) atom type 12 CH (i) atom type 13 CH2



(j) atom type 14 CH3 (k) atom type 16 CR1

Figure S9



(a) atom type 1 O (b) atom type 2 OM (c) atom type 3 OA

(d) atom type 6 NT (e) atom type 7 NL (f) atom type 8 NR

(g) atom type 9 NZ (h) atom type 12 CH (i) atom type 13 CH2

(j) atom type 14 CH3 (k) atom type 16 CR1

Figure S10

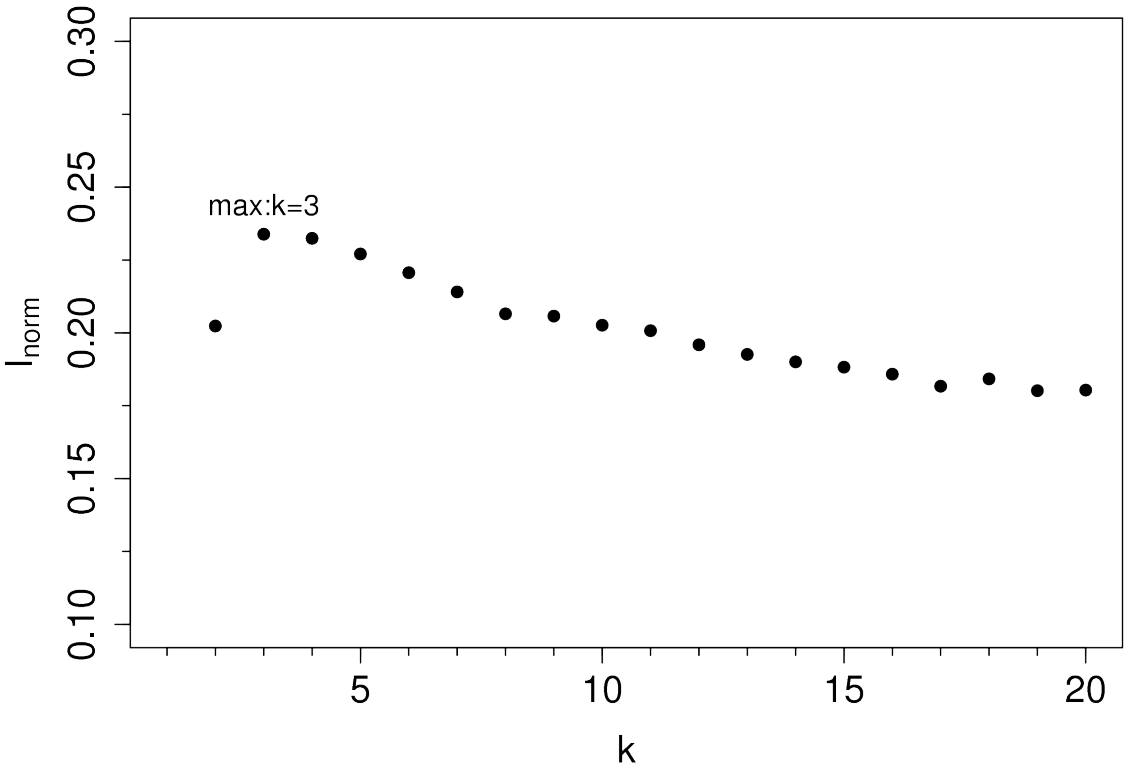
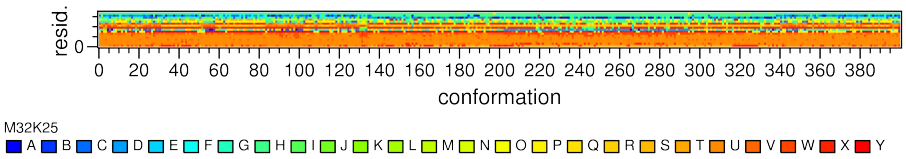
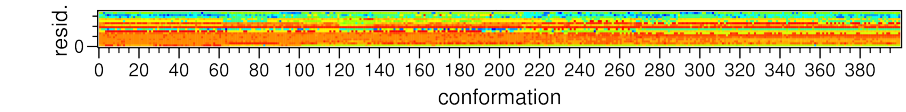


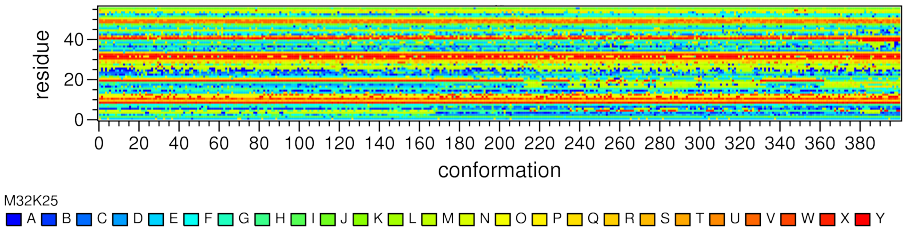
Figure S11



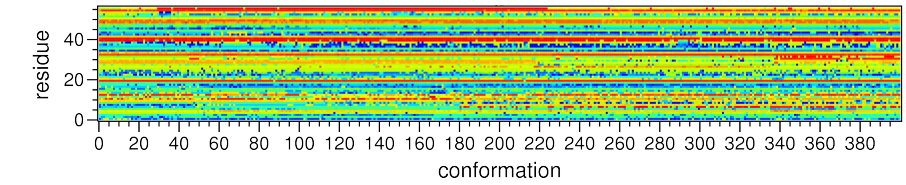
(a) trp (1l2y) in implicit solvent



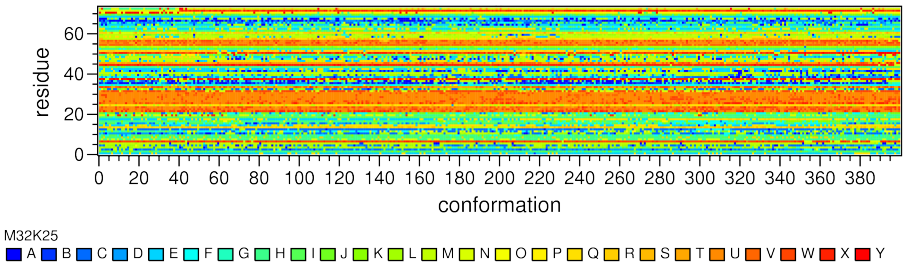
(b) trp (1l2y) in explicit solvent



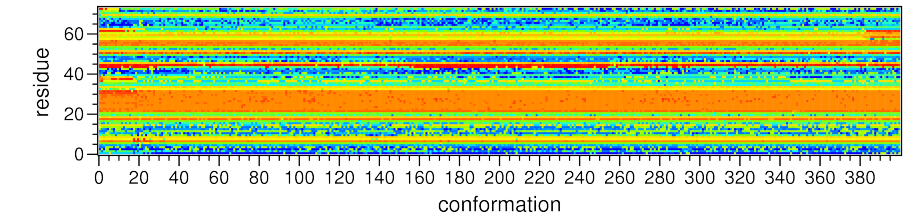
(c) drk (2a36) in implicit solvent



(d) drk (2a36) in explicit solvent

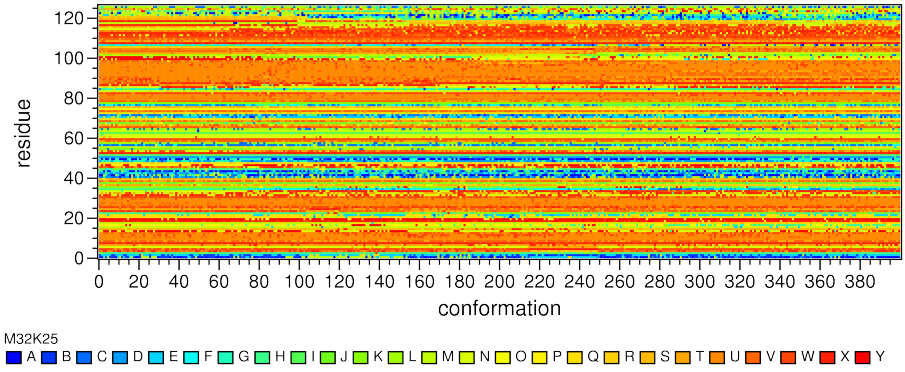


(e) ubq (1ubq) in implicit solvent

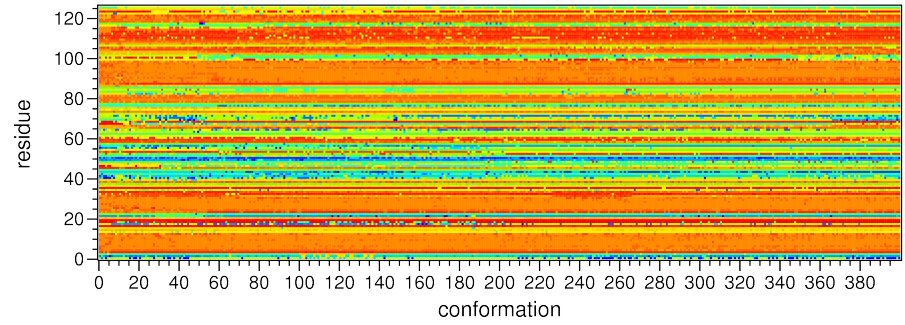


(f) ubq (1ubq) in explicit solvent

Figure S11



(a) lys (1aki) in implicit solvent



(b) lys (1aki) in explicit solvent

Table S1: Topological alphabet defined by the super–secondary structure of two successive α or β elements and the angle between their central axis. The combination of super–secondary structure and angle range defines a 'state', which is denoted by the letter given in the table.

angle range/ $^\circ$	β-β	β-α	α-β	α-α
0–60	a	d	g	j
60–120	b	e	h	k
120–180	c	f	i	l

Table S2: Characteristics of the protein domains used in this study for the implicit solvent parametrisation. The last six proteins (t1 to t6) belong to the performance test set. Columns: domain, domain identifier of the SCOP ASTRAL40 database;[1] description, SCOP protein name and domain/chain specification; topology, topological string defining the sequence and relative orientation of secondary structure elements (see Methods); class, SCOP domain classification; N_{res}, number of residues; N_{cha}, absolute net charge; SPACI, the SPACI structure quality score of the SCOP ASTRAL database (a SPACI score ≥ 0.4 indicates a 'good' structure).

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
1	d1a2pa_	Barnase	jjkgbaab	d.1.1.2	108	2	0.80
2	d1a8l_2	Protein disulfide oxidoreductase	ggdgdha	c.47.1.2	107	4	0.53
3	d1aa7a_	Influenza virus matrix protein M1	jjjjgfjjjg	a.95.1.1	158	2	0.49
4	d1ash__	Hemoglobin (domain I)	jjjljjkg	a.1.1.2	147	4	0.48
5	d1ay7b_	Barstar (barnase inhibitor)	ddlgdjg	c.9.1.1	89	6	0.69
6	d1b37a2	Polyamine oxidase	bbaadjga	d.16.1.5	112	1	0.58
7	d1b66a_	6-pyruvoyl tetrahydropterin synthase	ddhdjjga	d.96.1.2	138	1	0.53
8	d1b8da_	Phycoerythrin alpha subunit	jjjjjkj	a.1.1.3	164	2	0.55
9	d1bd8__	Cell cycle inhibitor p19ink4D	jjljjjj	d.211.1.1	156	4	0.63
10	d1bea__	Hageman factor/amylase inhibitor	jjjgbdg	a.52.1.2	116	1	0.40
11	d1bhta1	Hepatocyte growth factor	aaadgaaa	g.10.1.1	92	12	0.48
12	d1bi5a2	Chalcone synthase	aaadgdjjg	c.95.1.2	154	6	0.82
13	d1bm8__	DNA-binding domain of Mlu1-box binding protein MBP1	aaadjge	d.34.1.1	99	6	0.63

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
14	d1c02a_	Phosphorelay protein ypd1	jjjjgdgd	a.24.10.2	166	0	0.60
15	d1c0pa2	D-aminoacid oxidase	aaaaadh	d.16.1.3	95	1	1.01
16	d1c52__	Cytochrome c552	kkjgaadkj	a.3.1.1	131	7	0.93
17	d1c8ua1	Thioesterase II	ggdgaaba	d.38.1.3	114	0	0.54
18	d1cg5b_	Hemoglobin, beta-chain	jjjkjjj	a.1.1.2	141	3	0.71
19	d1coja2	Mn superoxide dismutase	kkgacdj	d.44.1.1	122	2	0.61
20	d1cot__	Cytochrome c6 (cytochrome c553)	ggadgejl	a.3.1.1	121	3	0.62
21	d1cqxa1	Hemoglobin I	jjhdjjj	a.1.1.2	150	7	0.61
22	d1cs6a1	Axonin-1	aaabbaa	b.1.1.4	97	4	0.61
23	d1cxya_	Cytochrome b558	ggadjjge	d.120.1.1	81	5	0.65
24	d1cyo__	Cytochrome b5	ddgadjjjj	d.120.1.1	88	5	0.74
25	d1d0qa_	Zinc-binding domain of DNA primase	jjgbafk	g.41.3.2	102	3	0.59
26	d1d3ba_	D3 core SNRNP protein	ggacaaa	b.38.1.1	72	2	0.53
27	d1d3bb_	B core SNRNP protein	aacaaaa	b.38.1.1	81	4	0.53
28	d1dp7p_	Class II MHC transcription factor RFX1	ggadgdga	a.4.5.20	76	6	0.71
29	d1dq3a3	PI-Pfui intein	ddidgad	d.95.2.2	98	3	0.54
30	d1dq3a4	PI-Pfui intein	jjgadgadj	d.95.2.2	109	2	0.54
31	d1dqea_	Moth pheromone-binding protein	jjlkjjg	a.39.2.1	137	8	0.60

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
32	d1duga1	Class pi Glutathione S-transferase	jjjjjjg	a.45.1.1	140	2	0.66
33	d1duga2	Class pi Glutathione S-transferase	ddgdhad	c.47.1.5	80	2	0.66
34	d1dvoa_	Repressor of bacterial conjugation FinO	kkhdgdjg	a.136.1.1	152	7	0.57
35	d1e42a2	Beta2-adaptin AP2, C-terminal subdomain	ggegaaa	d.105.1.1	113	1	0.62
36	d1e6ba2	Class pi Glutathione S-transferase	ddgdiad	c.47.1.5	80	1	0.65
37	d1eara2	Urease metallochaperone UreE, C-terminal domain	ddgadga	d.58.38.1	68	3	0.61
38	d1ed1a_	SIV matrix antigen	ggejjjkg	a.61.1.1	114	2	0.50
39	d1eexg_	Diol dehydratase, gamma subunit	ggejjjj	a.23.2.1	137	1	0.59
40	d1ej8a_	Copper chaperone for superoxide dismutase, C-terminal domain	aaaaaaa	b.1.8.1	140	1	0.67
41	d1ekga_	C-terminal domain of frataxin	ggaaaadg	d.82.2.1	119	7	0.59
42	d1enfa1	Staphylococcal enterotoxin H	ggbadgc	b.40.2.2	100	4	0.70
43	d1enfa2	Staphylococcal enterotoxin H	aadgbdg	d.15.6.1	112	5	0.70
44	d1eqfa1	TAFII250 double bromodomain module	ddjlgdjjj	a.29.2.1	139	5	0.50
45	d1et9a2	Streptococcal superantigen Spe-H	bbdgadg	d.15.6.1	109	4	0.54
46	d1evha_	Enabled/VASP homology 1 domain	bbaaaaa	b.55.1.4	111	4	0.54
47	d1exra_	Troponin C	kkgdjkgd	a.39.1.5	146	25	1.19
48	d1eyva_	Antitermination factor NusB	jjgdjkj	a.79.1.1	131	4	0.73
49	d1f00i1	Intimin	aabaaaa	b.1.14.1	95	0	0.52

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
50	d1f7ta_	Holo-(acyl carrier protein) synthase ACPS	ddjjjgaa	d.150.1.2	119	6	0.62
51	d1fc3a_	Spo0A	jjjjkjl	a.4.6.3	119	4	0.46
52	d1fp2a1	Isoflavone O-methyltransferase	jjgdkgae	a.4.5.29	101	3	0.76
53	d1g4ia_	Phospholipase A2	kkkkgadg	a.133.1.2	123	1	1.18
54	d1g73d_	BIR domains of XIAP	jjiadjj	g.52.1.1	102	3	0.46
55	d1gaka_	SP18	ddjjjge	a.19.1.1	137	15	0.54
56	d1glqa1	Glutathione S-transferase, C-terminal domain	jjjjjjk	a.45.1.1	131	2	0.64
57	d1glqa2	Glutathione S-transferase, N-terminal domain	ddgdjhad	c.47.1.5	78	0	0.64
58	d1go3f_	RNA polymerase II subunit RBP7	ddjkjjk	a.60.8.2	107	7	0.63
59	d1gpqa_	Inhibitor of vertebrate lysozyme	jjgdhaaaad	d.233.1.1	127	4	0.69
60	d1gsma1	Mucosal addressin cell adhesion molecule-1	ccaaaaa	b.1.1.4	90	4	0.48
61	d1gtfa_	Trp RNA-binding attenuation protein	aaabaaa	b.82.5.1	69	0	0.63
62	d1h4ra2	Moesin	aaacaad	b.55.1.5	99	8	0.65
63	d1h7wa5	Dihydropyrimidine dehydrogenase, C-terminal domain	kkgdgadg	d.58.1.5	173	1	0.59
64	d1h9ka1	Cytoplasmic molybdate-binding protein ModG	aaaadga	b.40.6.2	76	1	0.60
65	d1hfoa_	MIF-related	ddgaadg	d.80.1.3	113	0	0.60
66	d1hqz1_	Cofilin-like domain of actin-binding protein abp1p	ggfgadgd	d.109.1.2	139	10	0.49
67	d1hufa_	YopH tyrosine phosphatase N-terminal domain	ggaegdgb	d.195.1.1	123	0	0.46

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
68	d1hyua3	Alkyl hydroperoxide reductase subunit F, N-terminal domain	ggdgabd	c.47.1.2	102	2	0.55
69	d1hyua4	Alkyl hydroperoxide reductase subunit F, N-terminal domain	ggdgdhf	c.47.1.2	96	5	0.55
70	d1i4ga2	Staphylococcal enterotoxin A	aaegadg	d.15.6.1	113	0	0.50
71	d1i4ma_	Prion protein domain	aaadgel	d.6.1.1	108	4	0.54
72	d1i7na1	Synapsin I	ddgaaaad	c.30.1.5	102	1	0.50
73	d1in4a1	Holliday junction helicase RuvB	kkkgae	a.4.5.11	75	1	0.68
74	d1iow_1	D-Ala-D-Ala ligase	ddgdgdg	c.30.1.2	96	5	0.52
75	d1iq6a_	(R)-specific enoyl-CoA hydratase	ddkjgaab	d.38.1.4	132	2	0.76
76	d1it2a_	Hagfish hemoglobin	ddjjkjj	a.1.1.2	146	2	0.60
77	d1iuja_	Hypothetical protein TT1380	ddgadkg	d.58.4.5	102	2	0.61
78	d1iyha1	Glutathione S-transferase, C-terminal domain	jjjjjjh	a.45.1.1	124	5	0.65
79	d1j3wa_	Giding protein MglB	ggaegaad	d.110.7.1	134	5	0.78
80	d1jd1b_	Highdosage growth inhibitor YER057cp	aaadgdg	d.79.1.1	125	4	0.62
81	d1jfla1	Aspartate racemase	ddgdjgdg	c.78.2.1	115	4	0.60
82	d1k0ia2	p-Hydroxybenzoate hydroxylase	aaaaaej	d.16.1.2	102	1	0.58
83	1k8kd1	ARPC2	ggaadji	d.198.2.1	120	3	0.48
84	d1klxa_	Cysteine rich protein B	jjjjjjid	a.118.18.1	133	9	0.59
85	d1knca_	Antioxidant defense protein AhpD	jjjjjlj	a.152.1.1	174	1	0.52

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
86	d1krha2	Benzoate dioxygenase reductase	ddgdgcd	c.25.1.2	133	6	0.66
87	d1ks9a1	Ketopantoate reductase PanE	kkjgdkk	a.100.1.7	124	4	0.56
88	d1kwma2	Pancreatic carboxypeptidase, activation domain	ddgbdgf	d.58.3.1	95	4	0.73
89	d1kyfa2	Alpa-adaptin AP2, C-terminal subdomain	ddgdgaaa	d.105.1.1	114	1	0.94
90	d1le6a_	Snake phospholipase A2	jjkgbdg	a.133.1.2	123	5	0.54
91	d1lj9a_	Transcriptional regulator SlyA	jjkkgad	a.4.5.28	144	3	0.66
92	d1luqa_	Streptavidin	aaabaad	b.61.1.1	119	0	1.15
93	d1m2da_	Thioredoxin-like 2Fe-2S ferredoxin	ddjgaad	c.47.1.11	101	1	1.06
94	d1m4ra_	Interleukin-22	ffjkjjj	a.26.1.3	142	0	0.55
95	d1mbya_	Serine/threonine-protein kinase Sak C-terminal domain	aaaaaad	d.223.1.1	75	0	0.47
96	d1mg4a_	Doublecortin-like kinase Dclk	aadhaad	d.15.11.1	101	6	0.78
97	d1mgta2	O6-alkylguanine-DNA alkyltransferase	aaadgadj	c.55.7.1	88	0	0.66
98	d1mola_	Monellin	ddgaaab	d.17.1.1	94	2	0.66
99	d1mu5a2	Topoisomerase VI-B subunit	ggaadgae	d.14.1.3	164	0	0.50
100	d1mwxa1	Penicillin binding protein 2a, N-terminal domain	jjjkgaaab	d.17.4.5	112	4	0.54
101	d1n5ua2	Serum albumin	jjjjkjkkjj	a.126.1.1	192	9	0.52
102	d1n7ea_	Discs large protein homolog	aaacdkg	b.36.1.1	95	1	0.71
103	d1n8va_	Chemosensory protein Csp2	jjjhdkj	a.118.21.1	101	1	0.84

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
104	d1nbua_	7,8-dihydroneopterin aldolase	aafgdjjg	d.96.1.3	118	5	0.63
105	d1ng6a_	Hypothetical protein YqeY	jjjjlkj	a.182.1.1	148	1	0.81
106	d1ngka_	Protozoan/bacterial hemoglobin	jjjkjkjj	a.1.1.1	126	5	0.50
107	d1nh2a2	TATA-box binding protein	aadgaabd	d.129.1.1	85	3	0.58
108	d1ntya1	Triple functional domain protein TRIO	kkjkjjjl	a.87.1.1	184	8	0.66
109	d1oiub1	Cyclin A	kkjjgdjk	a.74.1.1	133	4	0.48
110	d1oiub2	Cyclin A	jjjjkkg	a.74.1.1	123	1	0.48
111	d1ok7a3	DNA polymerase III, beta subunit	ddgaaadga	d.131.1.1	122	5	0.63
112	d1opc__	OmpR	aaadkkg	a.4.6.1	99	1	0.47
113	d1opd__	Histidine-containing phosphocarrier protein	ddgadgd	d.94.1.1	85	4	0.66
114	d1osya_	Fungal immunomodulatory protein	ggcaaca	b.1.21.1	114	0	0.65
115	d1ou8a_	Stringent starvation protein B	ggabbaa	b.136.1.1	106	5	0.71
116	d1oxja2	RNA-binding protein Smaug	jjgdjjk	a.118.1.13	108	6	0.61
117	d1oyja2	Class pi GST	ddgdhbd	c.47.1.5	84	4	0.57
118	d1p0za_	Sensor kinase CitA	jjjgfgaa	d.110.6.1	131	2	0.75
119	d1p5dx2	Phosphomannomutase/phosphoglucomutase	ggdgaega	c.84.1.1	104	2	0.73
120	d1p8xa1	Gelsolin	aaadgeg	d.109.1.1	121	2	0.49
121	d1p8xa2	Gelsolin	aacadgd	d.109.1.1	96	1	0.49

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
122	d1p9ya_	Trigger factor ribosome-binding domain	aadkgab	d.241.2.1	117	1	0.50
123	d1pama1	Galactose oxidase, C-terminal domain	aaababa	b.1.18.2	86	2	0.62
124	d1pbwa_	p50 RhoGAP domain	jjkjjjj	a.116.1.1	184	6	0.49
125	d1pbya1	Quinohemoprotein amine dehydrogenase A chain, domains 1 and 2	ggbdgdj	a.3.1.7	85	5	0.63
126	d1pgva_	Ribonuclease inhibitor	llllgd	c.10.1.1	167	2	0.63
127	d1poc__	Phospholipase A2	ggbdjgc	a.133.1.1	134	2	0.48
128	d1q1ca2	FKBP52, N-terminal domain	aaaaega	d.26.1.1	117	6	0.57
129	d1q3oa_	Discs large protein homolog	aaabdkg	b.36.1.1	104	2	0.51
130	d1q5za_	Invasion protein A	ddkjjkjjj	a.196.1.1	145	1	0.67
131	d1qba_1	Bacterial chitobiase	jjgbabaa	b.1.18.2	105	4	0.63
132	d1qksa1	N-terminal (heme c) domain of cytochrome cd1-nitrite reductase	jjgadkjj	a.3.1.2	127	9	0.86
133	d1qzma_	gamma subunit of DNA polymerase III, N-domain	jjgdjgd	c.37.1.20	94	4	0.49
134	d1r1ta_	SmtB repressor	ddjgdkgad	a.4.5.5	98	1	0.66
135	d1r26a_	Thioredoxin	ddgdgdh	c.47.1.1	113	5	0.83
136	d1r5ra_	Pheromone-binding protein asp1	jjgdkjgd	a.39.2.1	117	15	0.67
137	d1r5ta_	mono-domain cytidine deaminase	ggadgbdga	c.97.1.1	141	0	0.53
138	d1rewc_	BMP receptor Ia ectodomain	aaaadha	g.7.1.3	86	4	0.49
139	d1rgxa_	Resistin	ggaaaaa	g.77.1.1	89	2	0.51

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
140	d1rlja_	Flavoprotein NrdI	ggdgcdgdg	c.23.5.7	135	0	0.49
141	d1rlka_	Hypothetical protein TA0108	ddjgdgad	c.131.1.1	116	0	0.53
142	d1rq2a2	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase IspF	ggdgdga	d.79.2.1	107	0	0.57
143	d1sdsa_	Eukaryotic ribosomal protein L30	ggdgdli	d.79.3.1	115	2	0.63
144	d1sh5a1	Actin binding domain of plectin	ggdjgdgej	a.40.1.1	120	5	0.56
145	d1siha3	Copper amine oxidase, domains 1 and 2	jjgaaaa	d.17.2.1	115	11	0.69
146	d1sra__	C-terminal (EC) domain of BM-40/SPARC/osteonectin	kkgdjjjk	a.39.1.3	151	10	0.54
147	d1sumb_	PhoU homolog TM1734	jjjjjjg	a.7.12.1	224	14	0.51
148	d1szha_	Her-1	jjkkjjj	a.226.1.1	147	3	0.73
149	d1t15a2	Breast cancer associated protein	ddgbdgdg	c.15.1.3	102	7	0.54
150	d1t1ja_	Hypothetical protein PA1492	ddgejgdg	c.137.1.1	119	6	0.63
151	d1t7ra_	Nuclear receptor ligand-binding domain	ggdjhdjjjjgdg	a.123.1.1	250	5	0.82
152	d1tfe__	Elongation factor Ts, dimerisation domain	aadjljjga	d.43.1.1	142	3	0.64
153	d1tlua_	S-adenosylmethionine decarboxylase	ddjgaaadg	d.156.1.2	117	6	0.69
154	d1tuaa1	Hypothetical protein APE0754	ddjgadj	d.51.1.1	84	1	0.74
155	d1tuwa_	Tetracenomycin polyketide synthesis protein TcmI	ddjhadjg	d.58.4.8	106	9	0.42
156	d1ty0a2	Streptococcal pyrogenic exotoxin Spe-J	aadgadg	d.15.6.1	107	2	0.61
157	d1u0sy_	CheY protein	ddlgdgdg	c.23.1.1	118	0	0.54

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
158	d1udxa1	Obg GTP-binding protein N-terminal domain	aaaaaca	b.117.1.1	156	1	0.52
159	d1udxa3	Obg GTP-binding protein C-terminal domain	aadjjga	d.242.1.1	76	3	0.52
160	d1ugia_	Uracil-DNA glycosylase inhibitor protein	ggadiaa	d.17.5.1	83	12	0.63
161	d1ugna1	Vascular cell adhesion molecule-1	aaadjga	b.1.1.4	96	2	0.56
162	d1unnc_	DNA polymerase IV	ddgadha	d.240.1.1	111	1	0.53
163	d1ut9a2	Cellulose 1,4-beta-cellobiosidase CbhA, precatalytic domain	aaaabaa	b.1.18.2	98	3	0.52
164	d1uvqa2	Class II MHC alpha chain, N-terminal domain	aaaaadj	d.19.1.1	83	3	0.64
165	d1uw4a_	RNA processing protein UPF3x, RRM domain	eegaadg	d.58.7.4	90	3	0.56
166	d1v74a_	Colicin D nuclease domain	kkgaaae	d.243.1.1	107	2	0.60
167	d1v97a4	Xanthine oxidase, domain 4	aaaafgd	d.87.2.1	114	3	0.58
168	d1vc1a_	Anti-sigma factor antagonist SpoIIaa	aadgdgdg	c.13.2.1	110	1	0.50
169	d1vgya2	Carboxypeptidase G2	ddgbadg	d.58.19.1	113	3	0.53
170	d1vjja2	Transglutaminase, two C-terminal domains	aabaaadg	b.1.5.1	115	5	0.56
171	d1vk5a_	Hypothetical protein At3g22680	llgdjjj	a.220.1.1	121	2	0.81
172	d1vlba3	Aldehyde oxidoreductase, domain 3	ggadidg	d.41.1.1	117	3	0.88
173	d1vlga_	Ferritin	jjgadjj	a.25.1.1	164	12	0.60
174	d1vsra_	Very short patch repair endonuclease	ggbbdgd	c.52.1.15	134	2	0.65
175	d1wlfa1	Peroxisome biogenesis factor 1, N-terminal domain	ddjhaab	b.52.2.3	80	6	0.52

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
176	d1wmub_	Hemoglobin, beta-chain	jjjljjj	a.1.1.2	146	2	0.68
177	d2cbla2	N-terminal domain of cbl	ddgejjj	a.48.1.1	131	6	0.45
178	d2cbp__	Pseudoazurin	hhaadga	b.6.1.1	96	6	0.65
179	d2cpl__	Cyclophilin (eukaryotic)	aadgaaeg	b.62.1.1	164	2	0.69
180	d2cy3__	Cytochrome c3	aaaegdk	a.138.1.1	118	3	0.59
181	d2end__	T4 endonuclease V	ggejgad	a.18.1.1	137	5	0.83
182	d2gdm__	Leghemoglobin	jjjjgdjj	a.1.1.2	153	5	0.62
183	d2sak__	Staphylokinase	aaadhaa	d.15.5.1	121	2	0.61
184	d3chbd_	Cholera toxin	ggaaadg	b.40.2.1	103	2	0.92
185	d3eipa_	Colicin E3 immunity protein	aadidga	d.26.2.1	84	13	0.57
186	d3sdha_	Hemoglobin I	jjjjjjj	a.1.1.2	145	4	0.81
187	d3seb_2	Staphylococcal enterotoxin A	aadgaadg	d.15.6.1	116	1	0.74
188	d3vub__	CcdB	aabbaad	b.34.6.1	101	1	0.76
t1	d1l2ya_	Trp-Cage miniprotein construct TC5b (trp)	-	k.32.1.1	20	1	-
t2	2a36	N-terminal SH3 domain of the Drosophila adapter protein (drk)	aaBaCA	-	59	6	-
t3	d1ubqa_	Human erythrocytic ubiquitin (ubq)	aADhAb	d.15.1.1	76	0	0.58
t4	d1tiga_	Translational initiation factor (if3c)	dDGDI	d.68.1.1	88	5	0.53
t5	d1akia_	Hen egg-white lysozyme (lys)	jJhAFJkJ	d.2.1.2	129	8	0.65

id.	domain	description	topology	class	N_{res}	N_{cha}	SPACI
t6	2jsw	C-terminal actin-binding domain of talin (talin)	gGaeJJJ	-	189	2	-

Table S3: Parameters of central tendency of log-likelihood log-normal fits of various explicit solvent force distributions. Columns: <ALL>, all water forces; <SA>, SASA selected forces; <SA & θ>, SASA and angle selected forces; <SA & θ+>, one-sided SASA and angle selected distribution; r<SA & θ+>, the <SA & θ+> distribution re–sampled using the 'mean' and 'sd' of the <SA> distribution; mean, mean value in $\mathrm{kJ\,mol^{-1}\,nm^{-1}}$; sd, standard deviation in $\mathrm{kJ\,mol^{-1}\,nm^{-1}}$.

atom			<ALL>		<SA>		<SA & θ>		<SA & θ+>		r<SA & θ+>	
id.	type	description	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
1	O	carbonyl oxygen (C=O)	4.83	0.67	4.97	0.56	4.80	0.54	4.90	0.61	4.99	0.59
2	OM	carboxyl oxygen (CO^-)	5.95	0.49	5.91	0.47	5.94	0.51	5.84	0.54	5.91	0.46
3	OA	hydroxyl oxygen (OH)	4.68	0.65	4.76	0.52	4.74	0.51	4.74	0.55	4.73	0.50
5	N	peptide nitrogen (NH)	–	–	–	–	–	–	–	–	–	–
6	NT	terminal nitrogen (NH_2)	4.69	0.66	4.80	0.52	4.66	0.68	4.71	0.59	4.70	0.46
7	NL	terminal nitrogen (NH_3^+)	5.89	0.65	5.89	0.65	6.10	0.59	6.10	0.63	5.86	0.67
8	NR	aromatic nitrogen (-N=)	4.85	0.91	4.33	0.83	4.39	0.69	4.30	0.72	4.40	0.80
9	NZ	Arg amino nitrogen (NH_2^+)	5.28	0.50	5.39	0.42	5.53	0.72	5.31	0.51	5.40	0.40
10	NE	Arg imino nitrogen (NH)	–	–	–	–	–	–	–	–	–	–
11	C	bare carbon (C)	–	–	–	–	–	–	–	–	–	–
12	CH1	aliphatic methine carbon (CH)	2.64	1.40	4.14	0.61	4.16	0.50	4.09	0.56	3.97	0.77
13	CH2	methylene carbon (CH_2)	3.70	1.22	4.28	0.62	4.27	0.57	4.28	0.63	4.24	0.66
14	CH3	methyl carbon (CH_3)	3.71	1.73	4.16	0.69	4.21	0.59	4.21	0.68	3.99	0.89
16	CR1	aromatic methine carbon (-CH=)	3.13	1.44	4.20	0.81	3.97	0.64	4.05	0.76	4.14	0.80

Table S4: Solvation parameters σ_g^{SASA} of atom groups derived by partitioning *via* dynamic programming. $\widetilde{\sigma_g^{SASA}}$: median value; (iqr) : inter-quartile range; sd_{bs}, standard deviation of the property in 1000 bootstrap (with replacement) samples; n : number of data points. Data are derived for the angle ranges 0–30° and 150–180° (0–45° and 135–180° in the main text). See Table S3 for a description of the atom types.

group		atom		solvation parameter		
id.	description	id.	type	$\widetilde{\sigma_g^{SASA}}$ (iqr) $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$	sd_{bs}	n(σ_g^{SASA})
1	charged	2, 7, 9	OM, NL, NZ	-26.0 (22.2)	0.2	15487 [507350]
2	polar	1, 3, 6, 8	O, OA, NT, NR	-7.8 (6.2)	0.08	8533 [1493950]
3	hydrophobic	12, 13, 14, 16	CH1, CH2, CH3, CR1	4.5 (3.9)	0.01	68272 [4191050]

Table S5: Mutual Information (I), Joint Entropy (H) and the normalised Mutual Information I_{norm} = I/H of the partitioning *via* dynamic programming into k groups for $1\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ binning of the σ_i value range. The maximal I_{norm} value is emphasised.

k	I / nat	H / nat	I_{norm}
2	0.73	3.63	0.202
3	0.95	4.08	*0.234*
4	1.02	4.41	0.232
5	1.05	4.62	0.227
6	1.06	4.82	0.220
7	1.07	5.02	0.214
8	1.08	5.24	0.207
9	1.09	5.29	0.206
10	1.09	5.40	0.203
11	1.10	5.47	0.201
12	1.10	5.63	0.196
13	1.11	5.74	0.193
14	1.11	5.84	0.190
15	1.11	5.90	0.188
16	1.11	5.99	0.186
17	1.12	6.14	0.182
18	1.12	6.06	0.184
19	1.12	6.21	0.180
20	1.12	6.22	0.180

Table S6: Original implicit solvation parameters[2] for GROMOS atom types.[3]

atom id.	atom type	atom description	solvation parameter σ_i/kJ mol^{-1} nm^{-2}
1	O	carbonyl oxygen (C=O)	-25
2	OM	carboxyl oxygen (CO^-)	-25
3	OA	hydroxyl oxygen (OH)	-25
5	N	peptide nitrogen (NH)	-25
6	NT	terminal nitrogen (NH_2)	-25
7	NL	terminal nitrogen (NH_3^+)	-25
8	NR	aromatic nitrogen (-N=)	-25
9	NZ	Arg amino nitrogen (NH_2^+)	-25
10	NE	Arg imino nitrogen (NH)	-25
11	C	bare carbon (C)	5
12	CH1	aliphatic methine carbon (CH)	5
13	CH2	methylene carbon (CH_2)	5
14	CH3	methyl carbon (CH_3)	5
16	CR1	aromatic methine carbon (-CH=)	5

References

(1) Brenner, S. E.; Koehl, P.; Levitt, M. *Nucleic Acids Res* **2000**, *28*, 254–256.

(2) Fraternali, F.; van Gunsteren, W. F. *J Mol Biol* **1996**, *256*, 939–948.

(3) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v., 1996.