Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules

Jonathan P. McNamara and Ian H. Hillier*

Table S1 Interaction energies (kcal mol^{-1}) for geometry optimised *S22* complexes

No.	Complex (symmetry)	AM1	PM3	$PM3_{BP}$[a]	AM1-D	PM3-D	Database[b]
Hydrogen bonded complexes (7)							
1	$(NH_3)_2$ (C_{2h})	–1.39	–0.71	–0.87	–3.03	–1.99	–3.17
2	$(H_2O)_2$ (C_s)	–3.30	–3.55	–4.19	–7.22	–6.53	–5.02
3	Formic acid dimer (C_{2h})	–6.62	–9.58	–12.60	–12.45	–16.16	–18.61
4	Formamide dimer (C_{2h})	–2.06	–6.99	–14.48	–14.64	–14.42	–15.96
5	Uracil dimer (C_{2h})	10.48	–10.70	–16.96	–17.80	–18.83	–20.65
6	2-pyridoxine·2-aminopyridine (C_1)	–6.15	–7.06	–14.90	–13.06	–18.32	–16.71
7	Adenine·thymine WC (C_1)	–5.06	–6.90	–15.36	–12.66	–18.66	–16.37
MUE		8.78	7.28	2.45	2.86	1.77	
Complexes with predominant dispersion contribution (8)							
8	$(CH_4)_2$ (D_{3d})	–0.21	–0.32	–0.18	–4.10	–2.38	–0.53
9	$(C_2H_4)_2$ (D_{2d})	–0.13	–1.08	–0.82	–4.85	–4.11	–1.51
10	Benzene·CH_4 (C_3)	0.35	–0.20	–0.23	–2.93	–2.88	–1.50
11	Benzene dimer (C_{2h})	0.01	–0.02	–0.54	–3.10	–4.59	–2.73
12	Pyrazine dimer (C_s)	–0.34	–0.26	–0.38	–4.87	–4.45	–4.42
13	Uracil dimer (C_2)	–6.05	–4.26	–4.79	–11.25	–7.59	–10.12
14	Indole·benzne (C_1)	–1.33	–1.65	–2.01	–8.16	–6.26	–5.22
15	Adenine·thymine stack (C_1)	–5.15	–6.50	–16.57	–15.13	–11.70	–12.23
MUE		3.23	3.00	2.68	2.01	1.48	
Mixed complexes (7)							
16	Ethene.ethine (C_{2v})	–0.57	–1.23	–1.23	–2.47	–2.58	–1.53
17	Benzene·H_2O (C_s)	–1.03	–1.63	–1.77	–3.90	–4.46	–3.28
18	Benzene·NH_3 (C_s)	–0.80	–0.93	–1.12	–4.04	–3.99	–2.35
19	Benzene·HCN (C_s)	–0.92	–1.85	–2.24	–4.28	–4.40	–4.46
20	Beznene dimer (C_{2v})	–0.09	–0.52	–0.60	–4.22	–4.39	–2.74
21	Indole·benzene T-shape (C_1)	–1.24	–1.67	–1.85	–7.74	–7.20	–5.73
22	Phenol dimer (C_1)	–3.39	–4.33	–3.41	–11.55	–8.95	–7.05
MUE		2.73	2.14	2.13	1.63	1.28	
Statistics							
MSE[c]		–4.83	–4.09	–2.04	0.52	0.59	
MD[d]		3.27	2.74	1.48	1.99	1.34	
MUE[e]		4.83	4.09	2.43	2.16	1.51	
RMSE[f]		6.25	5.22	2.90	2.65	1.65	
MAXE–MINE[g]		13.16	9.74	10.35	10.65	5.13	

[a] Giese *et al*.[20] [b] *S22* database.[1] [c] Mean signed error. [d] Mean deviation. [e] Mean unsigned error. [f] Root mean square error. [g] Error spread (largest positive minus largest negative error).

Table S2 Interaction distances (Angstroms) for geometry optimised *S22* complexes[a]

No.	Complex (symmetry)	AM1	PM3	$PM3_{BP}$[b]	AM1-D	PM3-D	Database[c]
Hydrogen bonded complexes (7)							
1	$(NH_3)_2$ (C_{2h})	2.784	3.241	3.129	2.646	2.726	2.504
2	$(H_2O)_2$ (C_s)	2.094	1.809	1.761	1.911	1.769	1.952
3	Formic acid dimer (C_{2h})	2.101	1.776	1.729	1.925	1.737	1.670
4	Formamide dimer (C_{2h})	2.072	1.807	1.730	1.981	1.763	1.841
5	Uracil dimer (C_{2h})	2.044	1.787	1.722	1.946	1.744	1.775
6	2-pyridoxine·2-aminopyridine (C_1)	2.511, 2.107	1.798, 1.815	1.717, 1.737	1.980, 1.981	1.722, 1.768	1.859, 1.874
7	Adenine·thymine WC (C_1)	2.476, 2.101	1.780, 1.821	1.696, 1.741	1.807, 2.018	1.708, 1.768	1.819, 1.929
MUE		0.34	0.14	0.18	0.12	0.12	
Complexes with predominant dispersion contribution (8)							
8	$(CH_4)_2$ (D_{3d})	3.721	3.447	3.717	2.881	3.160	3.718
9	$(C_2H_4)_2$ (D_{2d})	3.714	3.706	3.759	3.305	3.469	3.718
10	Benzene·CH_4 (C_3)	3.746	3.718	3.729	3.315	3.450	3.716
11	Benzene dimer (C_{2h})	6.952	6.096	6.202	3.643	3.499	3.765
12	Pyrazine dimer (C_s)	4.848	4.760	4.633	3.695	3.437	3.479
13	Uracil dimer (C_2)	5.805	6.732	6.671	3.097	3.406	3.166
14	Indole·benzne (C_1)	5.572	5.520	4.781	4.448[i]	3.415	3.444
15	Adenine·thymine stack (C_1)	6.202	5.788	5.599	4.320[i]	3.280	3.172
MUE		1.55	1.52	1.36	0.53	0.22	
Mixed complexes (7)							
16	Ethene.ethine (C_{2v})	2.468	2.429	2.440	2.319	2.366	2.752
17	Benzene·H_2O (C_s)	4.020	3.746	3.656	2.986	2.982	2.531
18	Benzene·NH_3 (C_s)	4.092	4.025	3.988	2.995	3.069	3.592
19	Benzene·HCN (C_s)	3.472	3.694	3.441	3.228	3.343	3.387
20	Beznene dimer (C_{2v})	5.225	3.606	3.646	3.253	3.370	3.513
21	Indole·benzene T-shape (C_1)	3.811	3.807	3.427	3.010	3.233	3.210
22	Phenol dimer (C_1)	2.174, 5.925	1.829, 5.712	1.775, 5.674	2.001, 5.040	1.778, 5.265	1.937, 4.921
MUE		0.74	0.48	0.39	0.29	0.26	
Statistics							
MSE[d]		–0.83	–0.60	–0.51	–0.02	0.08	
MD[e]		0.79	0.79	0.74	0.30	0.18	
MUE[f]		0.85	0.69	0.63	0.30	0.20	
RMSE[g]		1.28	1.17	1.09	0.42	0.25	
MAXE–MINE[h]		3.47	3.89	3.82	1.99	1.01	

[a] Refer to Figure S1 for structures and (average) interaction distances. [b] Giese *et al*.[20] [c] *S22* database.[1] [d] Mean signed error. [e] Mean deviation. [f] Mean unsigned error. [g] Root mean square error. [h] Error spread (largest positive minus largest negative error). [i] Indole·benzene (C_1) collapses to the T-shape arrangement; adenine·thymine stack (C_1) is distorted by hydrogen-bonding interactions.

Table S3 Interaction energies (kcal mol^{-1}) for *JSCH-2005* complexes[a]

No.	Complex	AM1	PM3	$PM3_{BP}$[b]	AM1-D	PM3-D	Dispersion[c]	Database[d]
Hydrogen bonded DNA base pairs (31)								
23	G···C WC	–11.82 (–14.71)	–16.48 (–14.90)	–21.21 (–26.97)	–30.00 (–25.34)	–29.94 (–30.79)	–4.57	–32.06
24	mG···mC WC	–10.71 (–14.68)	–17.12 (–14.84)	–22.45 (–26.93)	–30.46 (–25.34)	–31.61 (–30.87)	–4.78	–31.59
25	A···T WC	–4.28 (–5.06)	–6.79 (–6.14)	–9.66 (–15.36)	–16.58 (–12.66)	–17.33 (–18.66)	–4.05	–16.86
26	mA···mT H	–3.39 (–5.09)	–8.13 (–7.71)	–11.85 (–16.55)	–16.81 (–12.78)	–19.51 (–19.86)	–4.21	–18.16
27	8oG···C WC pl	–11.76 (–15.54)	–19.03 (–16.31)	–24.70 (–28.77)	–31.79 (–26.01)	–33.67 (–32.06)	–4.69	–33.30
28	I···C WC pl	– 8.36 (–10.41)	–13.87 (–12.87)	–18.50 (–21.14)	–24.11 (–18.79)	–26.15 (–24.09)	–4.20	–24.90
29	G···U wobble	–2.46 (–8.39)	–8.23 (–7.35)	–11.64 (–13.05)	–18.39 (–16.65)	–18.60 (–15.80)	–3.63	–19.10
30[e]	CCH+	–24.60 (–33.54)	–35.96 (–36.18)	–43.20 (–48.31)	–48.96 (–40.04)	–53.21 (–51.64)	–4.82	–51.40
31	U···U Calcutta pl	–2.98 (–5.13)	–3.15 (–3.75)	–3.93 (–6.24)	–11.45 (–11.10)	–8.73 (–11.69)	–2.52	–10.30
32	U···U pl	–1.34 (–6.11)	–5.34 (–4.93)	–7.35 (–10.27)	–13.75 (–11.84)	–13.57 (–12.89)	–2.92	–13.70
33	2-aminoA···T	–5.51 (–7.74)	–7.85 (–7.81)	–10.84 (–17.53)	–23.21 (–17.81)	–22.15 (–23.10)	–5.11	–19.50
34	2-aminoA···T pl	–5.80 (–7.82)	–8.29 (–9.40)	–11.06 (–18.76)	–22.58 (–17.80)	–21.72 (–23.60)	–4.86	–19.70
35	A···F	–1.04 (–2.15)	–0.62 (–2.14)	–1.34 (–4.04)	–6.85 (–7.18)	–4.95 (–11.64)	–2.66	–5.20
36	A···C pl	–3.96 (–4.78)	–8.98 (–10.43)	–12.91 (–19.70)	–15.14 (–10.50)	–20.30 (–23.53)	–3.91	–17.60
37	G···G pl	–8.42 (–9.66)	–9.22 (–8.76)	–12.51 (–14.38)	–19.54 (–16.81)	–18.75 (–20.46)	–3.97	–21.30
38	G···A 1	–4.19 (–7.43)	–8.64 (–8.36)	–12.30 (–16.67)	–19.43 (–15.60)	–21.43 (–21.87)	–4.94	–19.40
39	G···A 1 pl	–4.26 (–7.07)	–9.23 (–10.15)	–12.47 (–17.78)	–18.55 (–15.60)	–20.60 (–22.21)	–4.44	–18.90
40	G···A 2	–1.17 (–3.70)	–4.12 (–4.84)	–5.94 (–12.92)	–10.70 (–7.90)	–14.00 (–19.89)	–3.89	–14.40
41	G···A 2 pl	–0.83 (–2.78)	–4.74 (–7.11)	–6.76 (–14.92)	–10.63 (–7.89)	–14.92 (–20.69)	–3.89	–12.80
42	G···A 3	–3.90 (–6.12)	–8.47 (–7.98)	–11.85 (–16.12)	–18.04 (–14.35)	–20.58 (–20.70)	–4.76	–18.80
43	G···A 4	–1.31 (–4.19)	–4.39 (–5.17)	–6.63 (–13.68)	–11.89 (–8.78)	–15.24 (–19.69)	–3.96	–13.50
44	A···A 1 pl	–1.87 (–3.33)	–6.08 (–7.58)	–9.28 (–16.28)	–12.06 (–8.42)	–16.61 (–20.11)	–3.70	–14.50
45	A···A 2 pl	–0.88 (–3.05)	–5.46 (–5.53)	–8.28 (–15.96)	–10.54 (–7.75)	–15.51 (–20.24)	–3.64	–13.70
46	A···A 3 pl	0.31 (–3.11)	–3.70 (–4.62)	–5.93 (–13.87)	–8.63 (–8.99)	–13.16 (–19.03)	–3.58	–12.20
47	8oG···G	–3.37 (–10.04)	–9.61 (–8.27)	–13.74 (–15.46)	–21.10 (–18.58)	–21.10 (–17.79)	–3.78	–22.80
48	A···T WC	–4.53	–6.38	–9.27	–17.13	–17.35	–4.30	–16.40
49	G···C WC	–12.79	–20.81	–26.56	–33.32	–35.95	–4.89	–35.80
50	A···T WC	–6.05	–6.80	–9.79	–18.00	–17.37	–4.27	–18.40
51	G···A HB	0.27	–5.40	–8.41	–12.01	–17.96	–4.54	–11.30
52	C···G WC	–9.32	–18.79	–25.37	–31.18	–34.60	–4.81	–30.70
53	G···C WC	–9.11	–18.59	–25.14	–30.60	–34.12	–4.68	–31.40
MUE[f]		15.17 (11.76)	10.63 (10.50)	7.06 (2.97)	1.58 (4.67)	1.66 (3.41)		
Interstrand base pairs (32)								
54	GG0/3.36 CGis036	–1.22	–1.82	–2.26	–3.35	–3.96	–1.86	–3.68
55	GG0/3.36 GCis036	7.36	8.67	9.71	–3.41	–4.59	–11.23	–4.82
56[g]	AA20/3.05 ATis2005	–1.37	–0.97	–1.48	–3.58	–3.33	–1.72	–2.16
57[g]	AA20/3.05 TAis2005	4.90	6.08	6.45	–4.82	–5.21	–9.76	–2.34
58	GC0/3.25 C//Cis	2.92	2.86	3.60	2.97	2.50	–0.39	3.09
59	GC0/3.25 G//Gis	4.39	5.34	6.43	2.05	2.21	–2.75	1.93

60[g]	CG0/3.19 G//Gis	2.10	2.52	3.06	1.57	1.65	–0.77	1.24
61[g]	CG0/3.19 C//Cis	1.08	4.46	4.10	–6.11	–4.69	–7.04	–3.91
62	GA10/3.15 A//Cis	1.40	1.37	2.01	–0.31	–0.42	–1.61	–0.31
63	GA10/3.15 T//Gis	1.48	1.48	1.85	0.64	0.36	–1.05	0.58
64	AG08/3.19 T//Gis	1.01	2.01	2.25	–1.82	–1.40	–2.64	–0.47
65	AG08/3.19 A//Cis	1.32	1.99	2.55	–0.32	0.06	–1.53	–0.18
66	TG03.19 A//Gis	0.63	1.84	1.45	–5.21	–5.12	–5.54	–4.22
67	TG03.19 T//Cis	–0.52	–0.71	–0.94	–1.10	–1.29	–0.54	–1.15
68	GT10/3.15 T//Cis	0.55	0.53	0.57	0.30	0.12	–0.42	0.30
69	GT10/3.15 A//Gis	–1.46	–1.83	–2.79	–4.70	–5.56	–2.96	–4.06
70	AT10/3.26 T//Tis	1.13	1.15	1.44	0.60	0.40	–0.74	0.88
71	AT10/3.26 A//Ais	0.85	1.70	2.30	–1.59	–1.10	–2.16	–0.92
72	TA08/3.16 A//Ais	2.28	3.58	4.18	–2.73	–2.37	–4.67	–1.55
73	TA08/3.16 T//Tis	0.96	0.90	1.13	0.67	0.43	–0.47	0.70
74	AA0/3.24 A//Tis	–0.08	–0.33	–0.46	–1.48	–1.78	–1.33	–1.71
75	AA0/3.24 T//Ais	–0.20	–0.54	–0.79	–1.39	–1.82	–1.16	–1.30
76	A···A IS	0.78	1.87	2.43	–2.19	–1.73	–2.66	–0.70
77	T···T IS	1.32	1.35	1.66	0.77	0.53	–0.80	1.00
78	G···G IS	1.33	3.77	3.55	–6.29	–5.64	–7.68	–4.50
79	C···C IS	2.58	3.59	4.26	1.30	1.82	–1.34	1.40
80	A···G IS	–1.85	–2.49	–3.57	–4.79	–5.82	–2.75	–4.80
81	T···C IS	0.34	0.23	0.20	–0.01	–0.29	–0.52	–0.10
82	C···A IS	–0.21	0.68	0.21	–3.09	–2.70	–2.28	–3.00
83	G···G IS	–3.60	–3.36	–4.31	–6.00	–5.68	–1.91	–5.22
84	G···G IS	4.00	6.12	6.93	–0.03	0.86	–4.18	0.80
85	C···C IS	3.08	3.02	3.75	3.16	2.68	–0.37	3.10
MUE		2.30	2.87	2.99	0.61	0.58		
Stacked base pairs (54)								
86	G···C S	–5.87	2.33	0.62	–21.10 (–22.20)	–14.50 (–15.75)	–11.98	–19.02
87	mG···mC S	–3.20	6.49	4.97	–22.51 (–23.59)	–15.75 (–18.91)	–15.98	–20.35
88	A···T S	2.91	7.37	6.56	–12.20 (–15.13)	–10.63 (–11.70)	–14.18	–12.30
89	mA···mT S	4.77	11.54	10.74	–15.80 (–18.45)	–12.46 (–15.18)	–18.32	–14.57
90	CC1	8.58	10.10	11.50	0.72	–0.14	–8.76	2.45
91	CC2	5.33	6.31	6.61	–3.72	–4.81	–9.40	–3.85
92	CC3	1.66	2.29	1.79	–8.27	–9.07	–9.47	–8.88
93	CC4	1.12	2.74	2.67	–9.19	–8.65	–9.42	–9.92
94	CC5	8.31	9.99	11.31	–0.06	–0.78	–9.10	0.32
95	CC6	8.60	10.60	11.87	0.19	–0.37	–9.23	0.64
96	CC7	4.79	4.64	5.22	–0.47	–2.01	–5.73	–0.98
97	CC8	0.72	0.80	0.83	–8.24	–9.05	–8.14	–9.10
98	CC9	1.48	3.41	3.23	–8.68	–7.76	–9.36	–9.11
99	CC10	1.24	3.16	3.17	–8.82	–8.00	–9.12	–8.27
100	CC11	0.64	1.99	1.66	–9.00	–8.56	–8.80	–9.43
101	CC12	–0.79	–1.73	–1.91	–6.36	–7.70	–4.99	–7.43

102	CC13	0.69	2.36	2.15	–8.03	–7.23	–7.82	–8.80
103	CC14	1.34	3.39	3.49	–9.12	–8.29	–9.53	–9.11
104	AAst	3.51	4.50	4.26	–7.88	–9.03	–11.38	–8.58
105	GGst	0.98	2.05	1.20	–12.12	–13.18	–12.71	–12.67
106	ACst	1.61	2.72	2.15	–9.08	–9.62	–10.40	–10.22
107	GAst	2.46	3.21	2.72	–9.87	–11.20	–12.13	–11.38
108	CCst	0.83	2.55	2.13	–9.00	–8.15	–9.00	–10.02
109	AUst	2.15	3.52	2.72	–8.81	–9.21	–10.73	–9.79
110	GCst	–0.31	1.10	–0.17	–10.91	–10.92	–10.03	–10.60
111	CUst	0.23	0.97	0.02	–9.34	–9.69	–9.00	–10.42
112	UUst	1.90	2.45	1.71	–6.45	–6.87	–8.15	–7.46
113	GUst	0.22	1.54	0.24	–11.10	–11.53	–10.93	–12.09
114[g]	GG0/3.36 GGs036	5.64	6.00	6.73	–0.87	–2.24	–7.04	–1.62
115[g]	GG0/3.36 CCs036	–3.10	–3.69	–4.61	–4.46	–4.76	–0.96	–3.54
116	AA20/3.05 AAs2005	5.29	7.58	8.18	–6.43	–6.88	–11.81	–6.06
117	AA20/3.05 TTs2005	–0.34	0.29	0.21	–3.18	–2.46	–2.16	–4.18
118	GC0/3.25 G//Cs	1.04	1.45	0.41	–10.94	–12.68	–11.66	–10.80
119	CG0/3.19 G//Cs	1.18	0.54	0.56	–6.43	–8.25	–7.22	–7.88
120	GA10/3.15 A//Gs	3.35	5.08	4.21	–9.84	–11.18	–13.24	–9.14
121	GA10/3.15 T//Cs	3.38	4.52	4.27	–4.47	–5.06	–8.08	–4.69
122	AG08/3.19 A//Gs	3.44	2.89	2.78	–6.43	–8.75	–10.01	–7.58
123	AG08/3.19 T//Cs	1.87	0.88	0.76	–6.15	–7.89	–7.68	–6.07
124	TG03.19 T//Gs	2.73	2.95	3.01	–4.68	–5.98	–7.39	–5.67
125	TG03.19 A//Cs	3.78	4.93	5.57	–4.58	–5.37	–8.23	–4.96
126	GT10/3.15 T//Gs	5.57	7.01	7.40	–6.46	–7.25	–11.91	–4.96
127	GT10/3.15 A//Cs	4.56	5.97	6.44	–5.48	–6.01	–10.04	–5.44
128	AT10/3.26 A//Ts	3.85	4.29	4.05	–6.87	–7.97	–10.45	–6.64
129	TA08/3.16 A//Ts	7.89	12.21	12.78	–5.80	–4.94	–13.52	–6.07
130	AA0/3.24 A//As	4.93	6.61	7.06	–6.21	–7.03	–11.23	–6.25
131	AA0/3.24 T//Ts	5.07	6.00	6.37	–5.27	–5.62	–9.85	–3.86
132	A···T S	4.71	5.23	5.10	–7.95	–9.20	–12.48	–8.10
133	G···C S	0.04	–2.11	–2.13	–6.38	–8.71	–5.85	–7.90
134	A···C S	4.54	5.20	6.02	–6.47	–7.34	–10.82	–6.70
135	T···G S	4.51	4.91	5.17	–6.79	–7.95	–11.23	–6.20
136	G···C S	2.76	4.93	5.13	–7.86	–6.92	–9.85	–7.70
137	A···G S	3.34	7.40	8.31	–8.92	–7.43	–11.78	–6.50
138	C···G S	0.61	0.14	–0.63	–9.98	–12.08	–10.27	–12.40
139	G···C S	–0.56	–0.31	–1.27	–10.74	–12.08	–9.69	–11.60
MUE[f]		10.36	11.80	11.68	0.81 (3.28)	1.11 (1.48)		
Amino acid base pairs								
140	F30-K46	–0.34	–1.08	–1.00	–4.27	–4.50	–3.28	–3.10
141	F30-L33	–0.47	0.02	0.25	–9.69	–8.29	–7.23	–5.00
142	F30-Y13	–0.20	–1.27	–1.14	–5.75	–6.37	–4.78	–3.90
143	F30-F49	0.30	–0.54	–0.47	–4.03	–4.44	–3.80	–3.30

144	F30-Y4	2.38	0.97	1.14	–6.59	–7.81	–8.60	–7.00
145	F49-K46	–1.53	0.81	1.16	–9.28	–5.70	–4.93	–4.80
146	F49-V5	–1.19	0.09	0.15	–12.28	–9.31	–7.79	–6.70
147[h]	F49-W37	0.28	–0.54	–0.45	–2.83	–3.37	–2.76	–2.50
148	F49-Y4	0.28	–2.07	–0.90	–8.22	–8.85	–6.54	–3.10
149	F49-PB (Y4-V5)	0.46	–0.04	0.21	–3.64	–3.97	–3.60	–2.80
150	F49-PB (V5-C6)	–0.66	0.14	–0.42	–10.17	–8.49	–7.30	–8.20
151	E47-K6 (1IU5)	–69.83	–68.09	–67.23	–79.87	–76.11	–5.58	–80.73
152	E49-K6 (1BQ9)	–99.84	–101.37	–100.15	–110.05	–109.40	–5.54	–113.35
153	E54-K2 (1SMM)	–86.71	–87.04	–87.46	–90.75	–89.95	–2.10	–88.29
154	E50-K30 (1BRF)	–58.22	–58.78	–59.16	–58.85	–59.14	–0.30	–60.36
155	E50-K52 (1BRF)	–84.55	–83.42	–83.97	–93.65	–91.40	–6.03	–97.14
156	E49-K6 (1BRF)	–63.83	–67.09	–66.20	–73.65	–73.41	–5.47	–74.24
MUE		5.93	5.60	5.82	2.32	2.28		
Statistics								
MSE[i]		–8.98	–8.60	–7.78	–0.01	0.41		
MD[j]		4.74	4.29	3.97	1.13	1.17		
MUE[k]		8.99	8.61	7.79	1.13	1.26		
RMSE[l]		10.69	10.11	9.24	2.52	3.04		
MAXE–MINE[m]		26.97	27.07	26.39	9.29	12.40		

[a] (In parenthesis) interaction energies calculated using semi-empirical optimised geometries. [b] Giese *et al*.[20] [c] AM1-D and PM3-D dispersion contributions for *JSCH-2005* geometries only. [d] *JSCH-2005* database.[1] Reference complexes 23-47 and 86-89 are *ab initio* optimised geometries whereas remainder are experimental geometries. Complexes omitted containing sulfur; hydrogen bonded complexes 6tG···C WC pl, A···4tU WC, G···4tU, G···2tU, G···6tG pl, 6tG···G pl, 2tU···2tU pl and amino acid pairs F49-C39 and F49-C6. [e] Overall charge +1. [f] (In parenthesis) MUEs for optimised complexes only. For comparison, for the stacked complexes the MUEs at the reference geometries are 1.39 (AM1-D) and 3.23 (PM3-D) kcal mol^{-1}. [g] Interaction energies of complexes 56 and 57, 60 and 61, 114 and 115 are interchanged compared to the *JSCH-2005* database. [h] Residue 37 of 1RB9 is tryptophan (W), not tyrosine (Y). [i] Mean signed error. [j] Mean deviation. [k] Mean unsigned error. [l] Root mean square error. [m] Error spread (largest positive minus largest negative error).

Table S4 Molecular mechanics and semi-empirical interaction energies (kcal mol^{-1}) of F30 cluster in rubredoxin[a]

	F30						
Amino Acid Base Pair	**F49**	**K46**	**L33**	**Y13**	**Y4**	**Sum**	**MUE**
Reference[b]	–3.30	–3.10	–5.00	–3.90	–7.00	–22.30	
Semi-empirical							
AM1	0.30	–0.34	–0.47	–0.20	2.38	1.67	4.79
PM3	–0.54	–1.08	0.02	–1.27	0.97	–1.90	4.08
$PM3_{BP}$[c]	–0.47	–1.00	0.25	–1.14	1.14	–1.22	3.98
AM1-D	–4.03	–4.27	–9.69	–5.75	–6.59	–30.33	1.77
PM3-D	–4.44	–4.50	–8.29	–6.37	–7.81	–31.41	1.82
DFT-D[d]							
BLYP-D/TZV(2d,2p)	–2.60	–3.05	–5.14	–4.14	–5.10	–20.03	0.61
Molecular Mechanics[e]							
AMBER gaff	–18.90	–6.40	–6.70	0.60	–5.30	–36.70	5.36
AMBER parm94, parm99	–19.30	–13.00	–6.60	–3.10	–5.60	–47.60	5.94
AMBER ff02	–24.10	–15.90	–7.80	–4.80	–5.40	–58.00	7.78
AMBER ff03	–16.20	–8.20	–5.20	–3.50	–5.40	–38.50	4.04
CHARM22	–2.10	–1.30	–2.70	–2.40	–4.70	–13.20	1.82
MMFF94	–1.50	–4.10	–2.20	–2.10	–5.00	–14.90	1.88
Engh-Huber	–3.10	–3.10	–4.10	–2.90	–8.20	–21.40	0.66
OPLS-AA	0.00	–0.30	–0.10	0.40	–1.20	–1.20	4.22
TAFF	–2.20	–1.80	–3.60	–2.60	–5.30	–15.50	1.36
Rule	–10.00	–10.60	–14.60	–13.90	–26.00	–75.00	10.54

[a] PDB Code: 1RB9. [b] *JSCH-2005* database.[1] [c] Giese *et al.*[20] [d] Morgado *et al.*[31] [e] Vondrášek *et al.*[30]

Table S5 Interaction distances (Å) for semi-empirically optimised *JSCH-2005* complexes

No.	Complex	AM1	PM3	$PM3_{BP}$[a]	AM1-D	PM3-D	Database[b]
Hydrogen bonded DNA base pairs (25)							
23	G···C WC	2.087	1.803	1.721	1.948	1.739	1.761
		2.349	1.781	1.717	1.964	1.745	1.906
		2.115	1.840	1.761	1.989	1.764	1.922
24	mG···mC WC	2.085	1.802	1.720	1.948	1.739	1.730
		2.366	1.784	1.720	1.972	1.744	1.874
		2.111	1.837	1.759	1.985	1.765	1.880
25	A···T WC	2.101	1.821	1.741	2.018	1.768	1.929
		2.476	1.780	1.696	1.870	1.708	1.819
26	mA···mT H	2.491	1.766	1.679	1.896	1.692	1.746
		2.130	1.827	1.748	2.051	1.770	1.930
27	8oG···C WC pl	2.072	1.805	1.723	1.952	1.741	1.729
		2.089	1.768	1.707	1.941	1.732	1.853
		2.084	1.838	1.759	1.984	1.765	1.868
28	I···C WC pl	2.088	1.795	1.729	1.994	1.754	1.790
		1.988	1.779	1.710	1.848	1.718	1.775
29	G···U wobble	2.052	1.794	1.728	1.929	1.747	1.744
		2.147	1.802	1.732	2.036	1.759	1.754
30[c]	CCH+	2.168	1.858	1.771	2.014	1.780	1.892
		1.886	1.697	1.644	1.809	1.668	1.714
		1.995	1.767	1.714	1.880	1.714	1.610
31	U Calcutta pl	2.106	1.808	1.745	1.988	1.759	1.847
		2.207	2.481	2.461	1.998	1.781	2.177
32	U···U pl	2.109	1.805	1.735	1.973	1.758	1.810
		2.109	1.799	1.733	1.971	1.752	1.811
33	2-aminoA···T	2.120	1.850	1.771	1.981	1.778	1.942
		2.036	1.758	1.676	1.932	1.714	1.799
		2.110	1.844	1.760	1.976	1.770	1.870
34	2-aminoA···T pl	2.111	1.832	1.761	1.981	1.770	1.905
		2.055	1.758	1.677	1.933	1.719	1.868
		2.106	1.828	1.756	1.976	1.765	1.873
35	A···F	2.160	1.801	1.776	1.997	1.787	2.105
		2.777	2.598	2.530	2.477	1.746	2.272
36	A···C pl	2.571	1.792	1.715	2.004	1.726	1.903
		2.468	1.794	1.721	1.980	1.727	1.889
37	G···G pl	2.585	1.810	1.734	1.881	1.721	1.846
		2.244	2.485	2.459	2.071	1.800	2.162
38	G···A 1	2.431	1.805	1.722	2.029	1.735	1.823
		2.100	1.805	1.728	1.974	1.752	1.806
39	G···A 1 pl	2.385	1.778	1.712	2.028	1.739	1.939

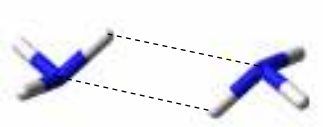
		2.088	1.807	1.730	1.974	1.749	1.774
40	G···A 2	2.636	1.835	1.744	2.279	1.735	2.005
		2.756	1.811	1.719	2.172	1.728	1.961
41	G···A 2 pl	2.569	1.786	1.704	2.176	1.720	1.928
		2.635	1.817	1.739	2.280	1.732	2.006
42	G···A 3	2.539	1.790	1.706	2.073	1.720	1.821
		2.123	1.815	1.739	2.016	1.762	1.851
43	G···A 4	2.693	1.815	1.725	2.023	1.731	1.921
		2.653	1.820	1.730	2.048	1.729	1.947
44	A···A 1 pl	2.577	1.802	1.722	2.046	1.733	1.923
		2.577	1.802	1.722	2.046	1.733	1.925
45	A···A 2 pl	2.562	1.834	1.736	2.298	1.743	1.951
		2.629	1.810	1.711	2.194	1.726	1.938
46	A···A 3 pl	2.607	1.836	1.736	2.725	1.744	1.987
		2.608	1.836	1.736	2.725	1.744	1.981
47	8oG···G	2.080	1.789	1.716	2.032	1.741	1.717
		2.136	1.815	1.742	1.969	1.765	1.712
MUE		0.41	0.09	0.15	0.17	0.14	
Stacked Base Pairs (4)							
86[d]	G···C S	5.587	5.707	5.774	3.693	3.641	3.191
87[d]	mG···mC S	6.012	6.140	6.219	3.159	3.183	3.056
88[d]	A···T S	6.202	5.788	5.599	4.320[j]	3.281	3.172
89[d]	mA···mT S	6.536	6.821	5.984	3.092	3.238	3.054
MUE		2.97	3.00	2.78	0.45	0.22	
Statistics							
MSE[e]		–0.58	–0.17	–0.08	–0.18	0.11	
MD[f]		0.38	0.39	0.38	0.12	0.10	
MUE[g]		0.58	0.29	0.32	0.19	0.15	
RMSE[h]		0.89	0.79	0.55	0.27	0.18	
MAXE-MINE[i]		3.45	4.07	3.49	1.33	0.98	

[a] Giese *et al*.[20] [b] *JSCH-2005* benchmark database. Sulfur containing complexes 6tG···C WC pl, A···4tU WC, G···4tU, G···2tU, G···6tG pl, 6tG···G pl and 2tU···2tU pl omitted. [c] Overall charge +1. [d] Distance between centre of mass of monomer units. [e] Mean signed error. [f] Mean deviation. [g] Mean unsigned error. [h] Root mean square error. [i] Error spread (largest positive minus largest negative error). [j] Hydrogen-bonding interactions distort structure.

Figure S1 Intermolecular interaction distances for *S22* complexes.

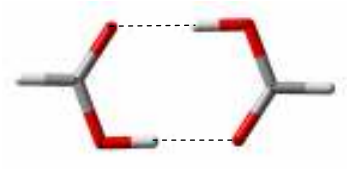
Note: ● refers to the centre of mass (COM) of the monomer unit

Hydrogen bonded complexes (7)

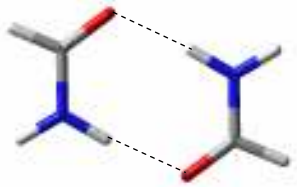


$(NH_3)_2$ (C_{2v})

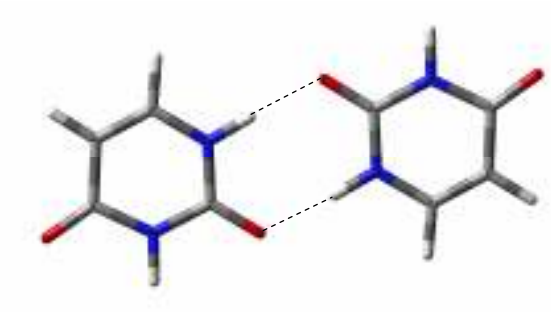
$(H_2O)_2$ (C_s)



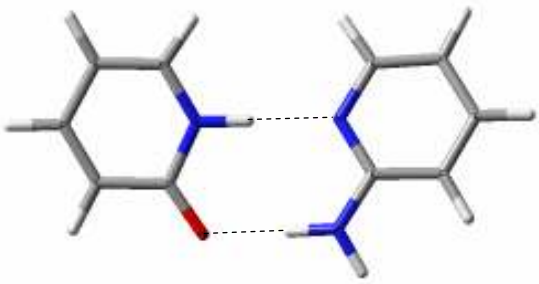
Formic acid dimer (C_{2h})



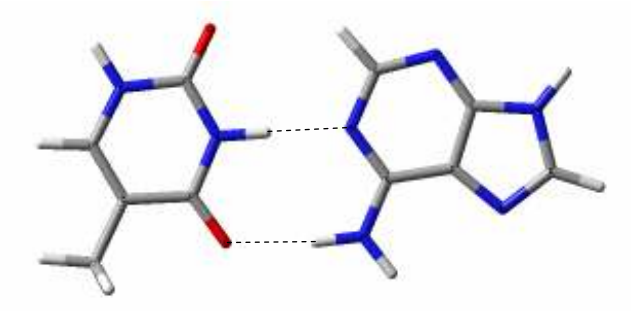
Formamide dimer (C_{2h})



Uracil dimer (C_{2h})



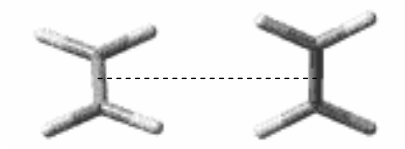
2-pyridoxine·2-aminopyridine (C_1)



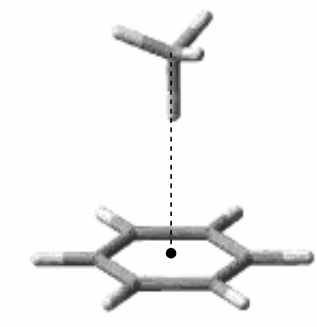
Adenine·thymine WC (C_1)

Complexes with predominant dispersion contribution (8)

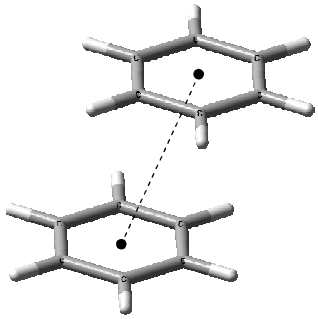
$(CH_4)_2$ (D_{3d})



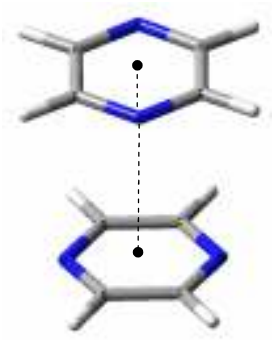
$(C_2H_4)_2$ (D_{2d})



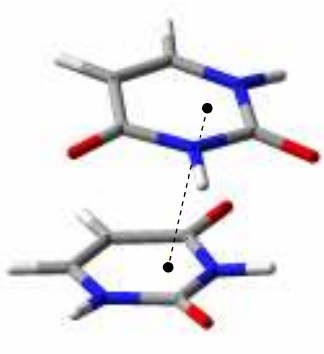
Benzene·CH_4 (C_3)



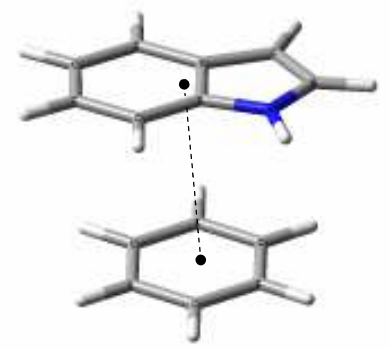
Benzene dimer (C_{2h})



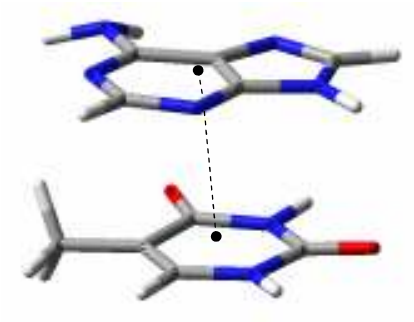
Pyrazine dimer (C_s)



Uracil dimer (C_2)

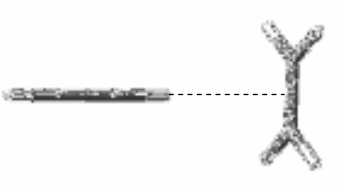


Indole·benzene (C_1)

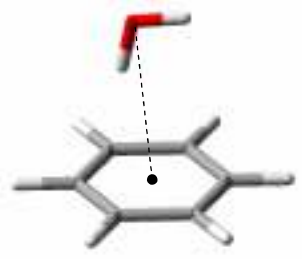


Adenine·thymine stack (C_1)

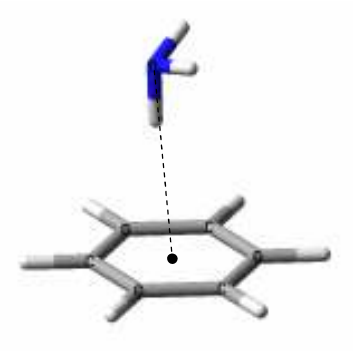
Mixed Complexes (7)



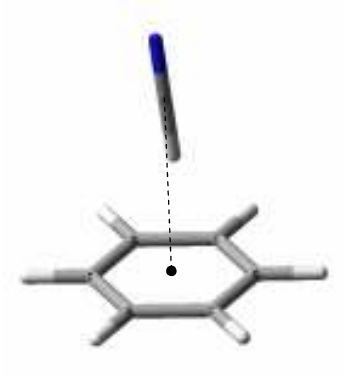
Ethene·ethine (C_{2v})



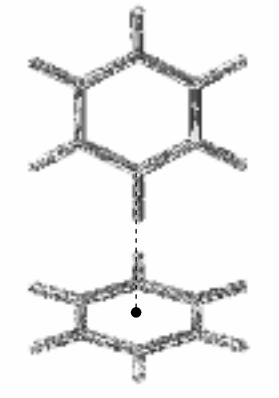
Benzene·H_2O (C_s)



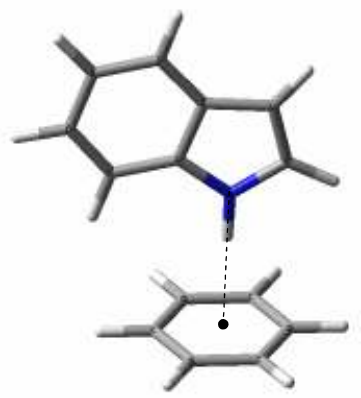
Benzene·NH_3 (C_s)



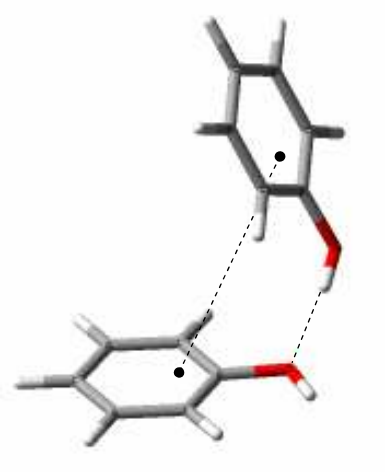
Benzene·HCN (C_s)



Benzene dimer (C_{2v})



Indole·benzene T-shape (C_1)



Phenol dimer (C_1)