Supporting Information

Dobbins *et al.* 10.1073/pnas.0802496105

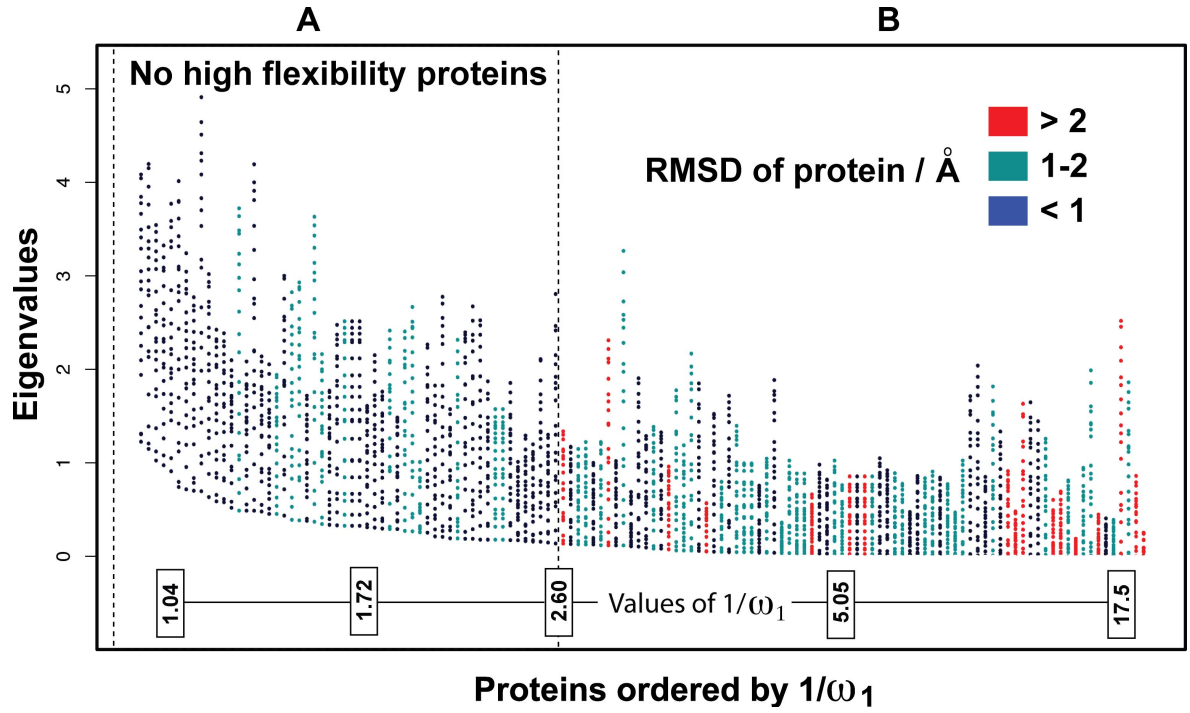

Fig. S1. The lowest 20 mode frequencies are shown for each protein in the main dataset. Proteins are ranked in descending order of lowest mode frequency. The spectrum for each protein is colored according to extent of conformational change. The plot is split into two sections A and B, where section A contains no proteins with large conformational change, and B contains the remainder.

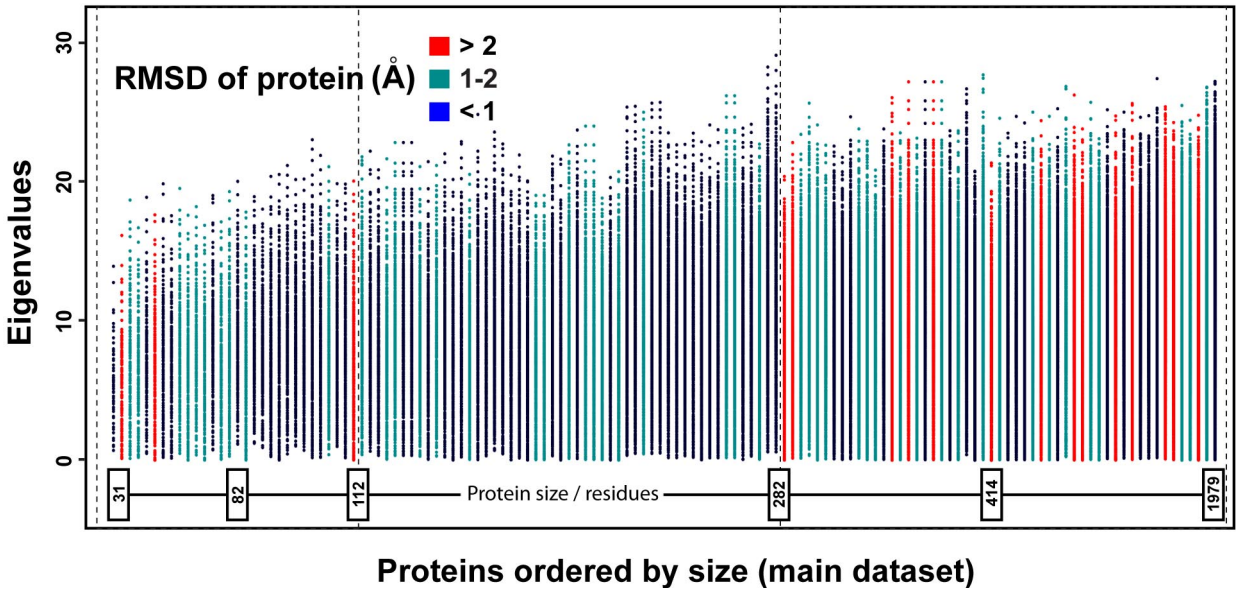

Fig. S2. All mode frequencies are shown for each protein in the main dataset. Proteins are ranked by size (number of residues). The spectrum for each protein is colored according to extent of conformational change.

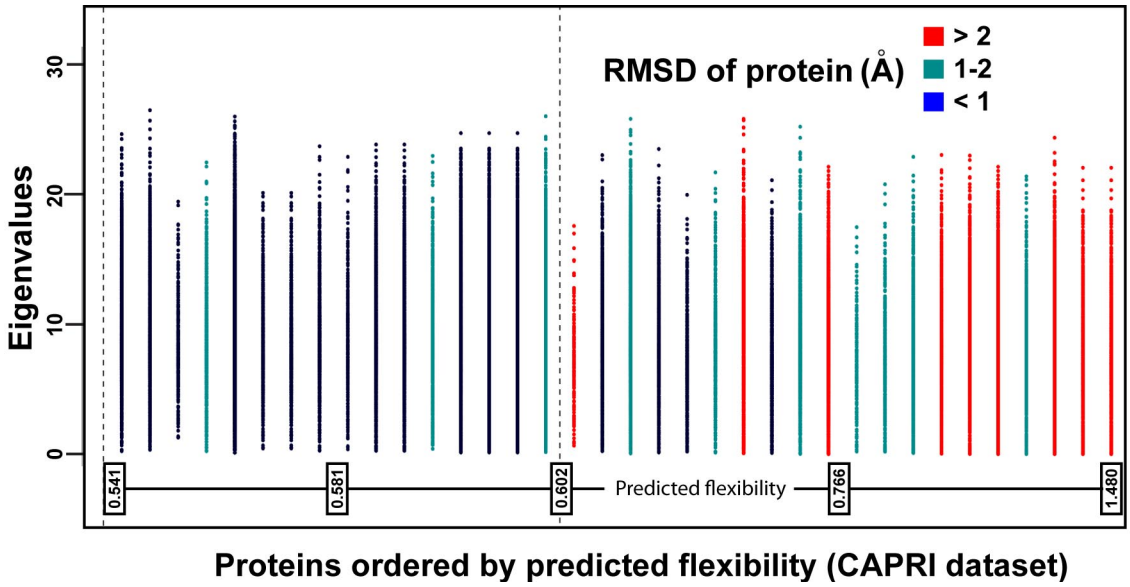

Fig. S3. All mode frequencies are shown for each protein in the CAPRI dataset. Proteins are ranked by predicted flexibility (see Eq. **2**). The spectrum for each protein is colored according to extent of conformational change.

Table S1. Main dataset

	PDB ID code	Size	RMSD, Å	$1/\omega_1$	Predicted flexibility
1	1 sup	282	0.249	1.305	0.4978
2	1ubn_A	274	0.28	1.295	0.507
3	7nn9	388	0.247	1.605	0.5285
4	1c3d	294	1.1	1.68	0.5366
5	2cpl	164	0.524	1.187	0.5448
6	1trm_A	223	0.354	1.431	0.5451
7	2hpr	88	0.491	0.9013	0.5454
8	1btp	223	0.433	1.377	0.5456
9	1poh	85	0.709	0.9301	0.5467
10	1qqu_A	223	0.49	1.419	0.5566
11	2tgt	223	1.28	1.471	0.5615
12	1 hrc	104	0.366	0.9844	0.5631
13	1ycc	108	0.498	1.01	0.5643
14	1srr_C	121	0.803	1.247	0.5707
15	3min_ABCD	1,979	0.464	5.598	0.5709
16	1j06_B	533	0.745	2.338	0.5752
17	1a19_B	89	0.618	1.045	0.5754
18	1hoe	74	0.474	1.156	0.577
19	1jae	472	0.396	2.644	0.5779
20	2rac_A	105	0.626	1.176	0.5784
21	2pka_XY	222	0.781	1.694	0.5848
22	2nip_AB	575	4.1	3.391	0.5896
23	1czp_A	98	0.813	1.5	0.5911
24	1aly_ABC	439	1.65	1.927	0.5924
25	1nob_F	185	0.51	1.777	0.5929
26	1cje_D	107	1.03	1.591	0.598
27	1bvl_BA	224	0.822	2.3	0.6009
28	1b1u_A	118	1.12	1.579	0.6017
29	3dni	258	0.407	1.819	0.6021
30	1pig	495	0.646	2.798	0.6038
31	2cga_B	245	1.48	2.303	0.6049
32	2cga_B	245	1.75	2.303	0.6049
33	1vfa_AB	224	0.534	2.33	0.6071
34	1ppi	496	0.571	2.702	0.6074
35	3lzt	129	1.23	1.722	0.6079
36	3lzt	129	0.847	1.722	0.6079
37	3lzt	129	0.751	1.722	0.6079
38	821p	166	0.643	1.471	0.6091
39	1udh	228	0.477	1.766	0.6147
40	8lyz	129	1.07	1.878	0.6195
41	9pti	58	0.554	1.419	0.6218
42	1f5w_B	121	0.964	1.711	0.6218
43	1m08_B	131	1.73	1.817	0.6264
44	1yrg_B	343	0.794	4.042	0.6274
45	1cd8_AB	229	0.987	2.209	0.6285
46	1gcp_B	67	1.12	1.417	0.6296
47	1he9_A	131	0.916	2.134	0.64
48	1a43	72	1.68	1.623	0.6405
49	1tnd_C	316	0.54	3.253	0.6435
50	1cx8_AB	1,278	1.68	6.135	0.6435
51	1ias_A	330	1.92	2.661	0.6437
52	1tbg_DH	405	1.96	4.663	0.6447
53	4pep	326	0.686	3.074	0.6449
54	1 gia	310	1.73	3.249	0.6481
55	1bv1	159	0.655	1.803	0.651
56	1b39_A	290	3.29	2.608	0.6541
57	1hh8_A	192	0.633	2.924	0.6581
58	1lu0_A	31	0.457	1.192	0.6616
59	1 se4	239	0.537	2.398	0.6647
60	1d0n_B	729	14	4.928	0.6686
61	1hcl	294	1.09	2.689	0.6691
62	1d6o_A	107	0.382	2.239	0.6738
63	1m9z_A	105	0.628	2.294	0.6741
64	1kdc	136	0.878	2.114	0.6787

	PDB ID code	Size	RMSD, Å	$1/\omega_1$	Predicted flexibility
65	1ste	238	0.95	2.427	0.6799
66	1fgn_LH	428	0.71	4.929	0.6807
67	1qbl_HL	433	0.913	5.336	0.6815
68	2pab_ABCD	457	0.814	2.895	0.6834
69	1e6o_HL	429	1.15	5.053	0.6859
70	1jpt_HL	425	0.701	5.347	0.6884
71	2ugi_B	83	1.09	1.892	0.6913
72	1e1n_A	455	1.07	4.138	0.6917
73	1rrp_AB	338	1.6	4.484	0.7027
74	1m0z_B	266	1.7	3.672	0.7063
75	1rgh_B	96	0.525	1.944	0.7084
76	1e8z_A	839	1.55	4.652	0.713
77	1jb1_ABC	484	2.76	3.98	0.7173
78	1hbp	175	0.541	2.515	0.7177
79	3ssi	108	0.667	2.45	0.7178
80	1fvu_AB	254	0.967	4.154	0.7206
81	1mkf_A	372	0.617	4.971	0.7228
82	1mlb_AB	432	1.03	6.237	0.7243
83	1a4r_A	190	1.75	2.754	0.731
84	1ay1_HL	423	1.55	5.985	0.7372
85	1dqq_CD	424	0.796	6.008	0.7375
86	1ijj_B	371	2.72	5.313	0.7396
87	1ijj_B	371	0.888	5.313	0.7396
88	1ijj_B	371	2.72	5.313	0.7396
89	1ijj_B	371	1.6	5.313	0.7396
90	2clr_DE	376	1.12	5.252	0.7617
91	1gig_LH	431	5.44	7.107	0.7652
92	2bnh	456	1.51	8.29	0.7742
93	1pne	139	0.751	4.551	0.7792
94	1a6z_AB	371	1.4	6.133	0.7825
95	1h15_AB	369	1.36	5.948	0.7838
96	2ci2_I	65	0.471	2.605	0.7868
97	1wer	324	0.975	4.401	0.7869
98	1fsc	61	0.77	2.243	0.799
99	1 hpt	56	1.8	2.805	0.801
100	1ly2_A	130	0.752	3.773	0.8164
101	1f32_A	127	1.05	3.647	0.8198
102	1ixm_AB	343	1.3	5.547	0.8365
103	1e6j_P	137	1.46	4.181	0.855
104	1k9b_A	58	1.03	2.226	0.8734
105	1hrp_AB	196	1.83	4.19	0.8811
106	1hrp_AB	196	1.83	4.19	0.8811
107	1f59_A	440	2.96	8.535	0.8844
108	1tfh_A	202	1.39	5.365	0.91
109	1s6p_AB	979	3.63	10.84	0.9101
110	1gjr_A	295	1.07	8.609	0.9111
111	1bec	238	0.891	5.79	0.9164
112	1kw2_B	453	2.12	7.308	0.9167
113	1tfh_B	182	1.08	4.875	0.9363
114	1tfh_B	182	1	4.875	0.9363
115	1fc1_A	206	0.813	7.106	0.9369
116	1bx8	50	2.13	2.78	0.9579
117	2vpf_GH	189	0.776	6.914	0.9857
118	1dks_A	76	1.05	3.702	1.017
119	1ccz_A	171	0.764	6.306	1.019
120	1cmw_A	818	1.42	16.23	1.026
121	1qjb_AB	460	2.34	11.99	1.074
122	1hnf	179	0.851	7.981	1.115
123	1jvm_A	97	0.701	6.878	1.131
124	1ecz_AB	284	2.29	8.472	1.135
125	1fqi_A	133	0.914	7.862	1.138
126	1unk_D	87	1.18	6.998	1.154
127	1tgk	112	2.22	7.547	1.171
128	1gri_B	211	1.14	11.16	1.284
129	1i49_AB	402	0.949	15.07	1.39

	PDB ID code	Size	RMSD, Å	$1/\omega_1$	Predicted flexibility
130	1qfk_HL	357	6.21	18.52	1.419
131	1dol	72	1.24	11.26	1.838
132	1egl	70	1.51	17.61	2.434
133	1bdd	60	3.13	17.49	2.657
134	1ern_AB	414	2.46	3.16E+04	2693

Proteins from the main dataset in order of predicted flexibility (see Eq. **2**). The following information is given: PDB ID code, number of residues (used for ranking proteins in Fig. S2), observed RMSD (in Å), inverse of the lowest-frequency mode (used for ranking of proteins in Fig. S1), and predicted flexibility (used for ranking of proteins in Fig. 4).

Table S2. Features of the high-flexibility dataset

	Protein name	PDB ID code	In complex with	RMSD, Å	RMSD BS, Å	Dynamic domain	Predicted flexibility rank	Max overlap
1	Staphylococcus A	1bdd	1fc1_AB	3.13	2.73	N	2	0.4
2	Ran GTPase	1qg4_A	1a12_A	2.63	3.65	N	–	0.26
3	14-3-3	1qjb_AB	1kuy_A	2.34	2.73	N	6	0.49
4	Actin	1ijj_B	3dni	2.72	9.47	N	12	0.56
5	Erythropoietin	1buy_A	1ern_AB	4.10	2.18	N	–	0.41
6	FAB	1gig_LH	2viu_A	5.44	7.57	Y	11	0.91
7	TGF-β	1tgk	1m9z_A	2.22	0.79	N	4	0.28
8	Actin	1ijj_B	1pne	2.72	1.51	Y	13	0.57
9	Coagulation factor VIIa	1qfk_HL	1tfh_B	6.21	12.52	Y	3	0.71
10	Ran GTPase	1qg4_A	1f59_A	3.91	2.31	N	–	0.41
11	HPr kinase C-ter domain	1jb1_ABC	2hpr	2.79	3.40	N	14	0.63
12	HIV1 reverse transcriptase	1s6p_AB	2hmi_CD	3.63	5.75	Y	9	0.53
13	Ecotin	1ecz_AB	1trm_A	2.29	2.95	N	5	0.62
14	EPO receptor	1ern_AB	1buy_A	2.46	2.65	Y	1	0.62
15	Vitamin D binding	1kw2_B	1ijj_B	2.12	1.74	Y	8	0.68
16	Nitrogenase Fe	2nip_AB	3min_ABCD	4.10	3.56	N	17	0.83
17	CDK2 kinase	1b39_A	1fpz_F	3.29	5.27	Y	16	0.33
18	Gelsolin	1d0n_B	1ijj_B	14.05	12.52	Y	15	0.30
19	Importin-β	1f59_A	1qg4_A	2.96	3.35	Y	10	0.75
20	Hirustatin	1bx8	2pka_XY	2.13	2.10	Y	7	0.47

Proteins from the high-flexibility dataset. The following information is given: PDB ID code, the PDB ID code of the complexation partner, RMSD (in Å), binding site (BS) RMSD (in Å), presence of dynamic domain (defined by software DynDom, see *Materials and Methods*), rank of predicted flexibility relative to the other high-flexibility proteins (ranging from 1 to 17 because proteins with $K_{lowest} < 0.03$ are excluded), and the maximum overlap of the lowest 20 modes.

Table S3. More features of the high-flexibility dataset

	PDB ID code	Collectivity observed	Collectivity NMs	Overlap NM7 (O_7)	Overlap NM8 (O_8)	Overlap NM9 (O_9)	Periphery/core >2
1	1bdd	0.04	0.48	0.02	0.14	0.17	Y
2	1qg4_A	0.01	0.35	0.03	0.07	0.06	Y
3	1qjb_AB	0.54	0.12	0.33	0.31	0.07	Y
4	1ijj_B	0.08	0.08	0.56	0.13	0.20	N
5	1buy_A	0.02	0.33	0.03	0.36	0.06	N
6	1 gig_LH	0.76	0.75	0.91	0.28	0.08	N
7	1tgk	0.54	0.23	0.04	0.10	0.05	N
8	1ijj_B	0.08	0.07	0.57	0.27	0.03	N
9	1qfk_HL	0.37	0.22	0.04	0.71	0.09	N
10	1qg4_A	0.01	0.19	0.01	0.08	0.05	N
11	1jb1_ABC	0.10	0.21	0.00	0.01	0.63	N
12	1 s6p_AB	0.45	0.63	0.53	0.32	0.49	N
13	1ecz_AB	0.51	0.56	0.16	0.62	0.51	N
14	1ern_AB	0.61	0.59	0.01	0.01	0.01	N
15	1kw2_B	0.56	0.54	0.68	0.17	0.12	N
16	2nip_AB	0.72	0.73	0.07	0.83	0.02	N
17	1b39_A	0.37	0.19	0.25	0.33	0.04	Y
18	1d0n_B	0.54	0.50	0.14	0.08	0.11	N
19	1f59_A	0.60	0.55	0.42	0.75	0.26	Y
20	1bx8	0.60	0.67	0.12	0.47	0.23	N

Proteins from the high-flexibility dataset. The following information is given: PDB ID code, collectivity (K) of observed conformational change, and lowest-frequency normal mode (see Eq. **5**), overlap (O_j) *of the three lowest-frequency (nontrivial) normal modes (see Eq.* **4**), and whether the ratio of observed motion of peripheral to core residues in BS is >2.

Table S4. CAPRI dataset

	CAPRI target	Size	RMSD, Å	$1/\omega_1$	Predicted flexibility
1	capri_016_B	301	0.97	0.229	0.541
2	capri_014_A_swissmodel	309	0.83	0.337	0.542
3	capri_001_A	86	0.53	1.263	0.550
4	capri_017_B_swissmodel	190	1.65	0.226	0.573
5	capri_002_ABC	1,194	0.63	0.099	0.577
6	capri_011_A	140	0.44	0.439	0.580
7	capri_012_A	140	0.44	0.439	0.580
8	capri_018_A	182	0.65	0.270	0.581
9	capri_026_B	111	0.60	0.364	0.584
10	capri_016_A	274	0.45	0.253	0.590
11	capri_017_A	274	0.40	0.253	0.590
12	capri_025_A	165	1.10	0.405	0.593
13	capri_004_A	496	0.41	0.129	0.596
14	capri_005_A	496	0.40	0.129	0.596
15	capri_006_A	496	0.37	0.129	0.596
16	capri_018_B	370	1.13	0.156	0.602
17	capri_011_B_swissmodel	56	4.54	0.651	0.620
18	capri_007_A	220	0.49	0.230	0.631
19	capri_026_A	417	1.71	0.058	0.649
20	capri_027_A	200	0.92	0.127	0.661
21	capri_021_B	125	0.93	0.187	0.700
22	capri_022_A	135	1.48	0.102	0.700
23	capri_001_CEF	506	2.76	0.063	0.701
24	capri_027_B	158	0.78	0.083	0.743
25	capri_020_A	276	1.65	0.027	0.766
26	capri_010_B	764	11.46	0.009	0.766
27	capri_022_B	62	1.29	0.199	0.769
28	capri_019_A_swissmodel	102	1.61	0.183	0.774
29	capri_021_A	195	1.27	0.079	0.805
30	capri_007_R	236	2.22	0.030	0.911
31	capri_013_A	246	10.01	0.019	0.922
32	capri_010_A	382	11.33	0.009	1.080
33	capri_008_A	164	1.88	0.015	1.240
34	capri_020_B_swissmodel	299	8.31	0.007	1.270
35	capri_009_A	206	12.75	0.006	1.480
36	capri_009_B	206	12.75	0.006	1.480

Proteins from the CAPRI dataset in order of predicted RMSD (see Eq. **2**). The following information is given: Capri target name, number of residues, observed RMSD (in Å), inverse of the lowest-frequency mode, and predicted flexibility (used for ranking of proteins in Fig. S3). A number of the protein structures were produced by using swiss-model (Schwede, *et al.,* 2003), which will impact on the expected accuracy of the calculated RMSD. Indeed, protein 17, which is built by using a homology model with only 50% sequence identity, is the only high-flexibility protein in the lower 60% of Fig. S3.

Table S5. Input parameters for calculating normal modes

Input parameters for PDBMAT (for use in pdbmat.dat)	
Coordinate FILENAME	pdbmat.structure
INTERACtion DISTance CUTOF	12.000
INTERACtion FORCE CONStant	1.000
Origin of MASS values	COOR
Output PRINTing level	0
Bond DEFINITION	NONE
Maximum bond LENGTH	0.000
BOND FORCE CONStant	0.000
ANGLE FORCE CONStant	0.000
LevelSHIFT	1.0E-09
Matrix FORMAT	FREE
Input parameters for DIAGRTB (for use in diagrtb.dat)	
MATRix filename	pdbmat.sdijf
COORdinates filename	structure.pdb
Eigenvector OUTPut filename	diagrtb.eigenfacs
Nb of VECTors required	10000
EigeNVALues chosen	LOWE
Type of SUBStructuring	NONE
Nb of residues per BLOck	1
Origin of MASS values	CONS
Temporary files cleaning	ALL
MATRix FORMat	FREE
Output PRINting level	0

Parameter files for calculation of the normal modes. The normal modes were calculated with the programs PDBMAT and DIAGRTB (each downloadable at www.igs.cnrs-mrs.fr/elnemo). PDBMAT requires only a PDB file as input. When PDBMAT is run, a file called pdbmat.dat_run is created. This can be edited and saved as pdbmat.dat and sets the values of parameters for future runs. DIAGRTB requires the PDB file and the output from PDBMAT as input. The file diagrtb.dat_run is created as above and can be altered in the same way to set the parameters. Further details can be found on the elnemo website.

Table S6. Benchmark proteins

Protein 1	Protein 2	Complex
1qqu_A	1ba7_B	1avx_A:B
1rgh_B	1a19_B	1ay7_A:B
1pig	1 hoe	1bvn_P:T
2cga_B	1 hpt	1cgi_E:I
2tgt	1k9b_A	1d6r_A:I
9rsa_B	2bnh	1dfj_E:I
1e1n_A	1cje_D	1e6e_A:B
1eax_A	9pti	1eaw_A:B
1gjr_A	1czp_A	1ewy_A:C
1trm_A	1ecz_AB	1ezu_C:AB
1fgn_LH	1tfh_A	1ahw_AB:C
1bvl_BA	3lzt	1bvk_DE:F
1dqq_CD	3lzt	1dqj_AB:C
1e6o_HL	1a43	1e6j_HL:P
1qg4_A	1oun_AB	1a2k_C:AB
2cpl	1e6j_P	1ak4_A:D
1d6o_A	1ias_A	1b6c_A:B
1hcl	1dks_A	1buh_A:B
1mh1	1hh8_A	1e96_A:B
1ixm_AB	1srr_C	1f51_AB:E
1bdd	1fc1_A	1fc2_C:D
1tnd_C	1fqi_A	1fqj_A:B
1gri_B	1gcp_B	1gcq_B:C
1bj1_HL	2vpf_GH	1bj1_HL:VW
1fsk_BC	1bv1	1fsk_BC:A
1iqd_AB	1d7p_M	1iqd_AB:C
1k4c_AB	1jvm_A	1k4c_AB:C
1kxq_H	1ppi	1kxq_H:A
2cga_B	1egl	1acb_E:I
1ay1_HL	1cmw_A	1bgx_HL:T
1gia	1tbg_DH	1gp2_A:BG
1a4r_A	1rgp	1grn_A:B
821p	1e8z_A	1he8_B:A
1qg4_A	1a12_A	1i2m_A:B
1qjb_AB	1kuy_A	1ib1_AB:E
1ijj_B	3dni	1atn_A:D
1a6z_AB	1cx8_AB	1de4_AB:CF
1buy_A	1ern_AB	1eer_A:BC
2hmi_CD	1s6p_AB	2hmi_CD:AB
4pep	1f32_A	1f34_A:B
1j06_B	1fsc	1mah_A:F
1btp	1lu0_A	1ppe_E:I
Protein 1	Protein 2	Complex
1udh	2ugi_B	1udi_E:I
2bbk_JM	2rac_A	2mta_HL:A
1ccp	1ycc	2pcc_A:B
1sup	3ssi	2sic_E:I
1ubn_A	2ci2_I	2sni_E:I
1unk_D	1m08_B	7cei_A:B
1jpt_HL	1tfh_B	1jps_HL:T
1mlb_AB	3lzt	1mlc_AB:E
1vfa_AB	8lyz	1vfb_AB:C
1qbl_HL	1hrc	1wej_HL:F
1gig_LH	2viu_A	2vis_AB:C
1c3d	1ly2_A	1ghq_A:B
1mh1	1 he9_A	1 he1_C:A
1mh1	1i49_AB	1i4d_D:AB
1nob_F	1f5w_B	1kac_A:B
1h15_AB	1ste	1klu_AB:D
1tgk	1m9z_A	1ktz_A:B
1ijj_B	1kw2_B	1kxp_A:D
1hnf	1ccz_A	1qa9_A:B
1bec	1se4	1sbb_A:B
1ijj_B	1pne	2btf_A:P
1nca_HL	7nn9	1nca_HL:N
1nsn_HL	1kdc	1nsn_HL:S
1qfw_HL	1hrp_AB	1qfw_HL:AB
1qfw_IM	1hrp_AB	1qfw_IM:AB
2jel_HL	1poh	2jel_HL:P
1fvu_AB	1auq	1ijk_BC:A
1rrp_AB	1yrg_B	1k5d_AB:C
1auq	1m0z_B	1m10_A:B
3min_ABCD	2nip_AB	1n2c_ABCD:EF
6q21_D	1wer	1wq1_R:G
1qfk_HL	1tfh_B	1fak_HL:T
1fpz_F	1b39_A	1fq1_A:B
1ijj_B	1d0n_B	1h1v_A:G
1qg4_A	1f59_A	1ibr_A:B
2pka_XY	1bx8	1hia_AB:I
1jae	1b1u_A	1tmq_A:B
2clr_DE	1cd8_AB	1akj_AB:DE
2pab_ABCD	1hbp	1rlb_ABCD:E
1i9r_HL	1aly_ABC	1i9r_HL:ABC
1jb1_ABC	2hpr	1kkl_ABC:H
1mkf_A	1dol	1ml0_A:D

All proteins in docking benchmark set. The PDB ID codes are shown for each of the components and also the complex along with relevant chain information.