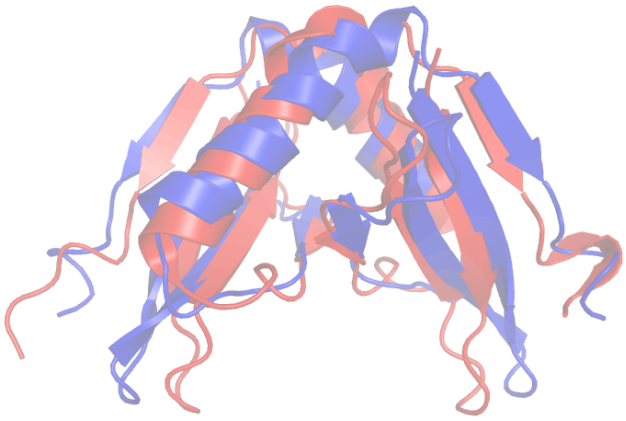
Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings

*Nikolaos G. Sgourakis[1], Oliver F. Lange[1], Frank DiMaio[1], Ingemar André[3], Nicholas C. Fitzkee[4], Paolo Rossi[5], Gaetano T. Montelione[5], Ad Bax[4] and David Baker[1,2]**

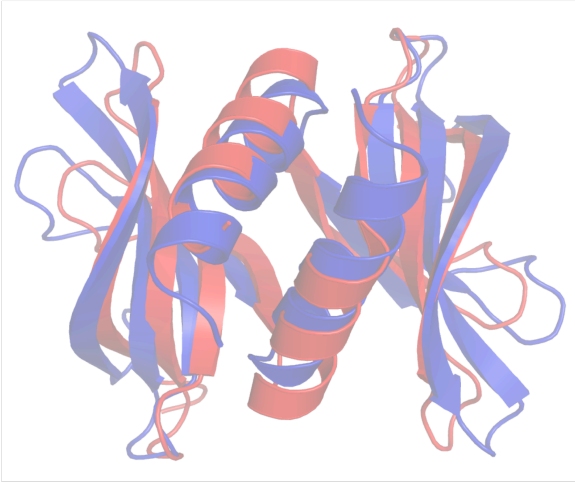
[1] Department of Biochemistry, University of Washington, Seattle, WA 98195-7350, USA; [2] Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195-7350 USA; [3] Department of Biochemistry and Structural Biology, Centre for Molecular Protein Science, Chemical Centre, Lund University, PO Box 124, SE-22100 Lund, Sweden; [4] Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA; [5] Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, and Northeast Structural Genomics Consortium, Rutgers University, Piscataway, NJ 08854, USA

Supporting information

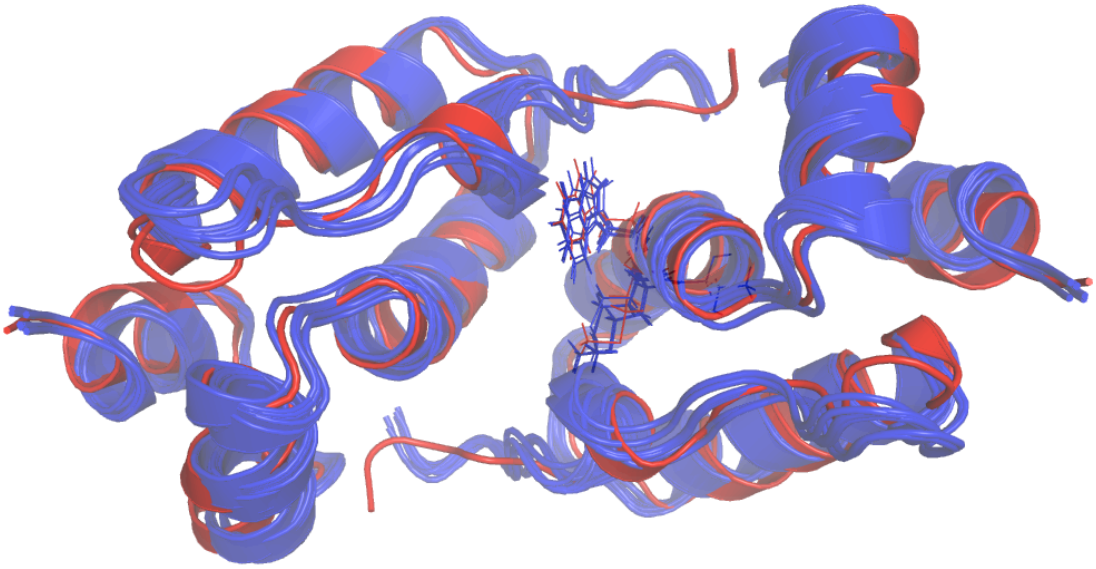
*David Baker: Department of Biochemistry, University of Washington, J wing, Health Sciences Building, Box 357350, Seattle, WA, 98185; Phone: 206-543-1295; Fax:206-685-1792; email: dabaker@u.washington.edu;



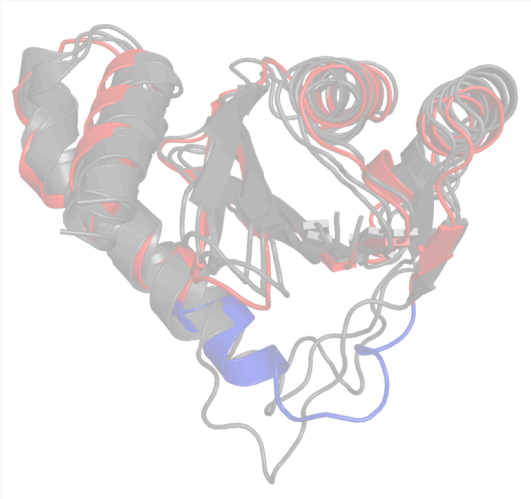
Supplementary figure 1. Structure of the protein ATU0232 from *Agrobacterium tumefaciens* (PDB ID 2K7I) using the fold-and dock branch of our method supplemented with RDC data. The structure determined here falls very close (2.5Å) to the previously reported NMR structure.



Supplementary figure 2. Application of the fold-and-dock branch of the method to structural genomics target KR150. We obtain a lowest-energy structure (blue) that falls very close to the available crystal structure from the homologous protein SP_0782 from *Streptococcus pneumoniae* shown in red (PDB ID 3OBH).

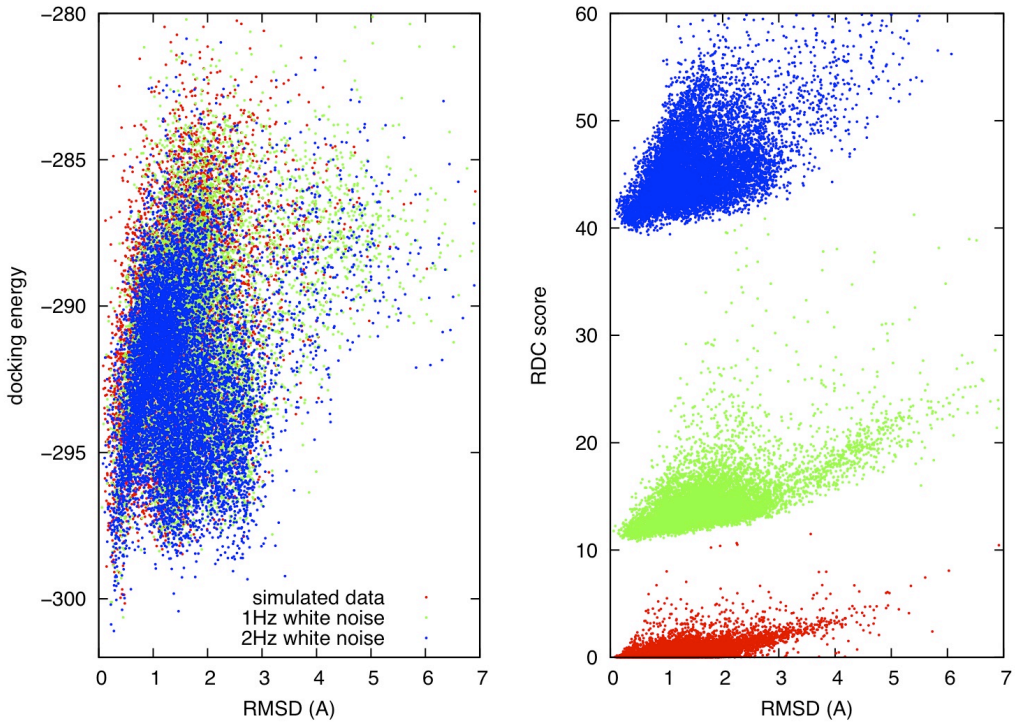


Supplementary figure 3. Low-energy ensemble of the HIV-1 capsid protein C-terminal domain determined here (blue) superimposed on the previously determined NMR structure (PDB ID 2KOD) (red). The side chains of the interface residues Trp 184 and Met 185 are shown demonstrating a high degree of convergence to the rotameric states observed in the native structure.

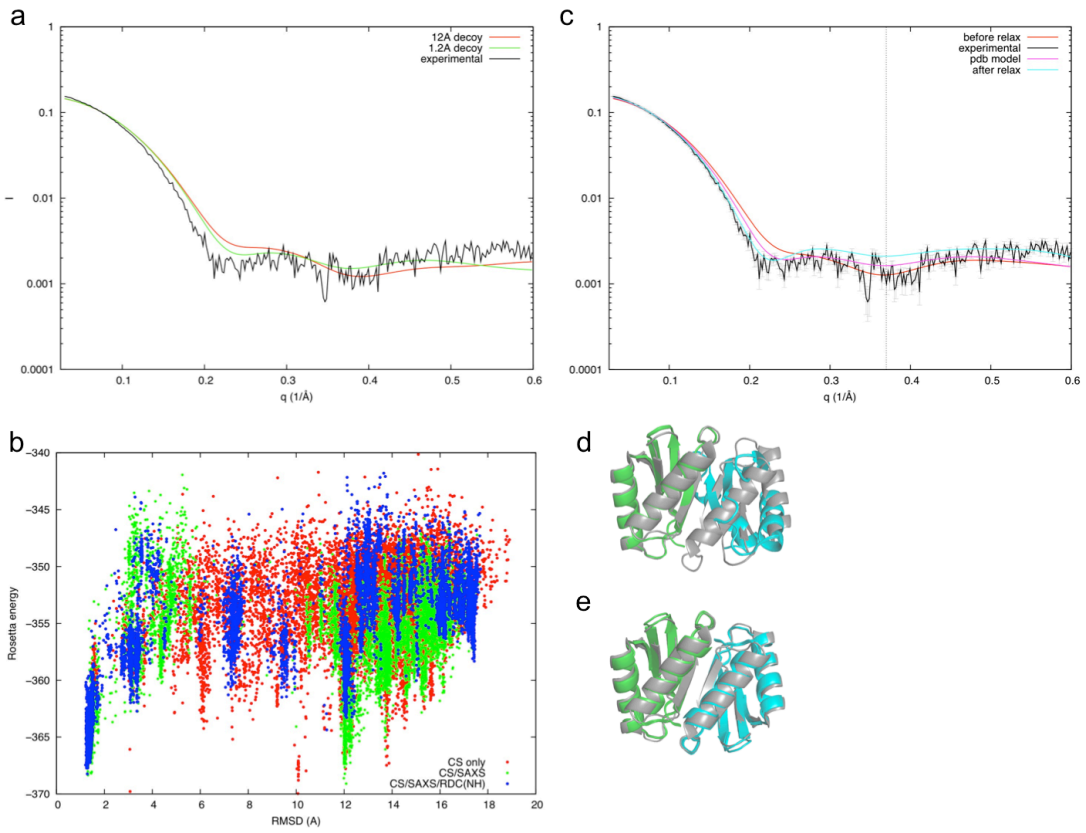


Supplementary figure 4. Lowest-energy structures of the monomer of the HIV Integrase core domain (residues 50-185 show here) produced using CS-Rosetta (gray) overlaid on the monomeric subunit from the crystal structure (red) (PDB ID 1BIS). We

obtain a high degree of convergence of the core of the structure and a large structural variability in the catalytic loop spanning residues 140-153 (highlighted in blue in the crystal structure). The region spanning residues 185-212 adopts a loop-helix motif that is found in various orientations due to structural variability in the loop region at residues 185-195.

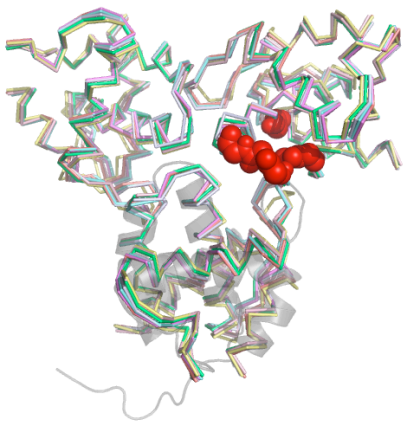


Supplementary figure 5. Symmetric docking results using synthetic RDC data with increasing levels of added noise are shown for the HIV Integrase catalytic core domain. Convergence of the calculations to the native structure is observed using synthetic RDCs with up to 2Hz of added noise (left). The RDC score shows a strong funnel towards the native structure (right). At higher noise levels, non-native local minima are beginning to emerge in the docking energy landscape.



Supplementary figure 6. Effects of including Small Angle X-ray data on symmetric docking calculations starting from a CS-Rosetta monomer for the TolR dimer. **(a)** Back-calculated SAXS spectra are shown for two low-energy conformations (shown in the structure diagrams of d, e respectively) derived from the symmetric docking calculations of (b). **(b)** Using the SAXS score to bias the search, symmetric docking converges to two local minima in Rosetta's energy. Here the use of SAXS data (green points) biases sampling away from non-native local minima otherwise observed in the naïve docking energy landscape (red points). By using limited RDC data (NH vectors only) in addition to SAXS we achieve convergence at a single, native minimum (blue points). Representative structures from the native-like minimum and the minimum at ~12 are shown in **(d, e).** The two subunits are colored as green and cyan, respectively. **(c)**

Detailed comparison with experimental data at different points of the symmetric docking calculation. Experimental data points before the dotted line are used for both fitting and scoring of a computed structure, while after for scoring only. The use of a gradient-based minimization of the full-atom coordinates to the SAXS score results in improved agreement with the experimental data.



Supplementary figure 7. Structural model of the equine anemia matrix virus homotrimer determined using NH RDCs and backbone chemical shifts using CS-Rosetta followed by symmetric docking. A high degree of convergence is observed in the low-energy Rosetta ensemble. In red spheres are residues that show significant chemical shifts perturbations upon increasing protein concentration, indicating their location on the trimer interface.

Supplementary table 1. Table with backbone amide RDC values measured at 900 MHz for the HIV integrase catalytic core domain in PEG and Pf1 liquid crystalline media respectively.

Residue	RDC(Hz)	RDC(Hz)
2	10.826	-3.879
3	6.739	28.861
6	N/A	21.227
8	-17.73	-16.997
9	-7.723	-18.827
11	-6.735	-23.331
12	-7.043	-12.294
13	6.788	6.372
14	-5.728	-20.634
15	-14.836	-22.574
16	N/A	-15.658
18	N/A	34.009
20	0.798	44.327
21	8.874	21.429
22	6.673	-0.723
23	11.436	-9.505
24	4.661	38.625
25	7.299	10.244
27	-14.807	8.662
29	-13.978	8.946
31	-12.072	-26.431

33	6.404	-13.78
34	3.446	-16.235
35	-10.581	12.243
37	5.349	-10.257
38	-6.61	-15.57
39	-0.443	-11.311
40	5.858	-6.955
41	4.337	-14.489
42	-2.009	-11.608
43	3.032	-1.301
45	4.239	-20.128
46	0.696	-9.362
47	6.669	-7.544
48	9.509	-11.98
49	-3.301	-16.079
50	N/A	-5.133
52	-4.658	-19.784
53	5.555	4.452
54	3.437	29.521
55	4.946	31.253
56	-0.804	29.717
57	0.714	32.804
60	8.027	8.216
61	0.965	-5.887
62	4.53	0.134
63	-6.639	N/A

65	6.625	23.119
68	-6.21	-14.697
69	1.856	-14.683
70	6.312	-13.097
71	0.702	-17.91
72	-0.587	-14.549
73	7.72	-8.605
74	5.247	-5.272
75	-0.915	-15.706
76	6.151	-0.954
77	8.349	-9.907
78	7.077	9.588
79	7.349	29.975
80	-3.503	34.791
81	2.878	31.326
96	2.755	15.751
97	-2.78	21.281
99	4.632	N/A
100	-2.126	29.917
101	-7.941	16.756
102	5.685	17.799
103	3.396	17.683
104	-5.366	N/A
105	-5.717	10.941
106	4.857	20.338
107	-4.256	N/A

108	-11.909	-9.604
109	0.606	-17.42
110	-5.476	6.029
112	N/A	-12.271
113	-2.694	-14.472
116	11.253	-9.179
117	N/A	-6.495
118	8.807	-12.589
119	7.052	-7.149
120	N/A	-5.917
121	8.415	-10.155
122	8.111	-26.305
123	8.819	-9.335
124	9.102	-11.674
125	9.236	-12.843
126	7.157	-13.622
128	5.063	5.098
137	-9.486	-21.564
138	6.253	3.959
139	8.357	24.567
140	3.932	-4.737
141	6.467	6.771
142	8.58	13.914
143	8.424	25.771
144	7.976	9.899
145	7.367	7.978

146	8.169	22.671
147	8.395	20.981
149	6.286	6.914
150	3.985	15.165

Command-line options for running the different stages of the method

A. Fragment selection (application: *picker.default.linuxgccrelease*)

-database [minirosetta database directory]

-in:file:vall [fragment structure database]

-frags:n_frags 200

-frags:frag_sizes 3 9

-frags:describe_fragments frags.fsc.score

-out:file:frag_prefix frags.score

-frags:scoring:config [weights parameter file]

-in:file:checkpoint [blast checkpoint file]

-in:file:talos_cs [chemical shift input file in TALOS format]

-frags:ss_pred [TALOS+ disorder prediction file (pred.ss.tab)]

-in:file:talos_phi_psi [TALOS+ prediction file (pred.tab)]

-frags:sigmoid_cs_A 2

-frags:sigmoid_cs_B 4

B. CS-Rosetta (application: *minirosetta.default.linuxgccrelease*).

-run:protocol broker

-in:file:fasta [fasta sequence file]

-file:frag3 [3mer fragments file]

-file:frag9 [9mer fragments file]

-abinitio:stage1_patch [RDC scoring weight file]

-abinitio:stage2_patch [RDC scoring weight file]

-abinitio:stage3a_patch [RDC scoring weight file]

-abinitio:stage3b_patch [RDC scoring weight file]

-cst_file [NOE restraints file mapped to the low-resolution representation]

-cst_fa_file [NOE restraints file]

-in:file:rdc [experimental RDCs file]

-score:patch [RDC scoring weight file]

-score:weights score12_full

-nstruct [number of docking calculations for each input monomer]

-out:file:silent [output structure file in silent format]

A more detailed description of obtaining and running CS-Rosetta can be found at: http://spin.niddk.nih.gov/bax/software/CSROSETTA/

C. Symmetry definition file generation.

To generate a default symmetry definition file for use in fold and dock or symmetric docking:

Cyclic symmetry:

```
make_symmdef_file_denovo.py -symm_type cn -nsub [number of subunits]
```

Dihedral symmetry:

```
make_symmdef_file_denovo.py -symm_type dn -nsub [number of subunits]
```

To extract the symmetric orientation from the coordinates of a symmetric oligomer, for use in symmetric docking only (to be used in perturbation studies).

Cyclic symmetry (this program will infer the number of subunits from the relative orientation of the two chains A,B in the pdb file) :

make_symmdef_file.pl -a A -i B -p [pdb file] > [C_n symmetry definition file]

Dihedral symmetry (assuming subunits A, B belong to the first oligomer and C to the second):

make_symmdef_file.pl -a A -i B C -p [pdb file] > [D_n symmetry definition file]

D. Symmetric docking (application: *SymDock.default.linuxgccrelease*)

-database [minirosetta database directory]

-in:file:s [series of pdb files]

-symmetry:symmetry_definition [symmetry definition file]

-smmetry:initialize_rigid_body_dofs

-packing:ex1

-packing:ex2aro

-use_input_sc

-ignore_unrecognized_res

-out:file:silent output.out

-in:file:rdc [experimental RDCs file]

-docking:low_patch [RDC scoring weight file]

-docking:high_patch [RDC scoring weight file]

-docking:high_min_patch [RDC scoring weight file]

-docking:pack_patch [RDC scoring weight file]

-docking:full_relax

-relax:fast

-relax:jump_move true

-score:patch [RDC scoring weight file]

-score:weights score12_full

-nstruct [number of docking calculations for each input monomer]

-out:file:silent [output structure file in silent format]

E. Fold and dock (application: *minirosetta.default.linuxgccrelease*).

-run:protocol broker

-broker:setup [topology file]

-file:frag3 [3mer fragments file]

-file:frag9 [9mer fragments file]

-in:file:fasta [fasta sequence file]

-symmetry:symmetry_definition [symmetry definition file]

-relax:fast

-relax:jump_move

-symmetry:initialize_rigid_body_dofs

-fold_and_dock::rotate_anchor_to_x

-rg_reweight 0.001

-abinitio:increase_cycles 0.2

-rigid_body_cycles 1

-abinitio:recover_low_in_stages 0

-rigid_body_frequency 0.2

-rigid_body_disable_mc

-run:reinitialize_mover_for_each_job

-abinitio:stage1_patch [RDC scoring weight file]

-abinitio:stage2_patch [RDC scoring weight file]

-abinitio:stage3a_patch [RDC scoring weight file]

-abinitio:stage3b_patch [RDC scoring weight file]

-in:file:rdc [experimental RDCs file]

-score:patch [RDC scoring weight file]

-score:weights score12_full

-nstruct [number of docking calculations for each input monomer]

-out:file:silent [output structure file in silent format]

For RDC refinement, the magnitude of the alignment tensor can then be set constant to this value using the flag:

-rdc:fix_normAzz [value]

used in both the CS-Rosetta, SymDock and fold-and-dock stages.

To include RDC data that require an Hα atom (such as Cα-Hα RDCs) at the low-resolution stage of all protocols used here, we use the flag:

-residues:patch_selectors CENTROID_HA

To include SAXS data in the low-resolution and full-atom stages of all protocols, we apply the flag:

-score:saxs:ref_spectrum [SAXS data file]

and add the line:

fastsaxs = 2.5

in the corresponding score weights files to engage the SAXS scoring term.

A standard file format with the values: (q I(q) err) separated by whitespace is used to input the experimental SAXS data.

Non-standard input files used in the different steps of the method

[fragment selection weights file]

score name priority wght min_allowed extras

CSScore 400 1 -

ProfileScoreL1 300 1.0 -

TalosSSSimilarity 200 0.25 - talos

RamaScore 100 1 - talos

PhiPsiSquareWell 50 0.15 -

FragmentCrmsd 30 0.0 -

[RDC scoring weight file]

rdc = 10.0

atom_pair_constraint = 5.0

experimental RDCs file format:

residue_number_i atom_name_i residue_number_j atom_name_j RDC_value(Hz)

e.g.

2 N 2 H -13.580

3 N 3 H -8.490

4 N 4 H -21.560

5 N 5 H -9.640

6 N 6 H -0.010

7 N 7 H -10.370

NOE restraints file format:

atom_name_i residue_number_i atom_name_j residue_number_j dist_min dist_max curvature

e.g

```
AmbiguousNMRDistance    H    2    H    20  BOUNDED  1.5  5.140  0.3
AmbiguousNMRDistance    H    2    H    21  BOUNDED  1.5  6.150  0.3
AmbiguousNMRDistance    H    3    H    82  BOUNDED  1.5  6.150  0.3
```

```
[symmetry definition file]

symmetry_name c2

subunits 2

recenter

number_of_interfaces  1

E = 2*VRT0001 + 1*(VRT0001:VRT0002)

anchor_residue COM

virtual_transforms_start

start      -1,0,0      0,1,0      0,0,0

rot Rz 2

virtual_transforms_stop

connect_virtual JUMP1 VRT0001 VRT0002

set_dof BASEJUMP x(50) angle_x(0:360) angle_y(0:360) angle_z(0:360)


[topology file]

CLAIMER FoldandDockClaimer

END_CLAIMER
```

Complete Reference 41

Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y. E.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popovic, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. *Methods Enzymol* **2011**, *487*, 545.