Supporting Information

Das et al. 10.1073/pnas.0904407106

SI Materials and Methods

Secondary Structure Definition. Only residues in the regular secondary structures are used for rmsd calculation for the NMR cases:

Gb1 secondary elements: 18, 12–20, 23–33, 44–46, 51–55, 58–64, 69–75, 79–89, 98–102, 107-

Gb1 secondary elements: 1–8, 12–20, 23–33, 44–46, 51–55, 58–64, 69–75, 79–89, 98–102, 107–111

NS1 secondary elements: 3–24, 31–49, 54–70, 203–224, 231–249, 254–270

Gr83 secondary elements: 3–16, 26–32, 41–54, 64–70

McR1 secondary elements: 5–31, 37–62, 68–94, 100–125

HR2106 secondary elements: 2–11, 17–24, 29–31, 36–60, 66–73, 77–83, 86–92, 98–107, 113–120, 125–127, 132–156, 162–169, 173–179, 182–188

Sft1 secondary elements: 10–31, 35–50, 210–231, 235–250

Atc0223 secondary elements: 3–9, 13–20, 24–28, 35–48, 53–56, 65–71, 75–82, 86–90, 97–110, 115–118

Availability. The code is freely available to academic users at www.rosettacommons.org in release Rosetta 2.3.0. The following command line carried out the protocol described above, here exemplifying the simulation of a C4 and D2 tetramer:

rosetta.exe xx pdb1 _ -output_silent_gz -read_all_chains -randomize1 -symm_type cn -pose_fold_and_dock -docking_pose_symmetry -nstruct 1 -pose1 -slide_contact_frequency 0.01 -rigid_body_frequency 0.01 -rigid_body_cycles 100 -new_centroid_packing -rg_reweight 0.0001 -move_anchor_points -monomer_input -pose_relax -pose_relax_fragment_moves -increase_cycles 1 -output_chi_silent -stringent_relax -ex1 -ex2 -vary_omega -omega_weight 0.5 -docking_pose_symm_full -combinatorial_rms_sym -pose_symm_jump_moves_during_relax -pose_symm_n_monomers 4 -n crystal_pdb1_native.pdb

rosetta.exe xx 1fe6 _ -output_silent_gz -read_all_chains -randomize1 -symm_type dn -pose_fold_and_dock -docking_pose_symmetry -nstruct 30 -pose1 -slide_contact_frequency 0.01 -rigid_body_frequency 0.01 -rigid_body_cycles 100 -new_centroid_packing -rg_reweight 0.0001 -move_anchor_points -monomer_input -pose_relax -pose_relax_fragment_moves -increase_cycles 1 -output_chi_silent -stringent_relax -ex1 -ex2 -vary_omega -omega_weight 0.5 -docking_pose_symm_full -combinatorial_rms_sym -pose_symm_jump_moves_during_relax -dn_twist_angle 5.0 -dn_twist_rings -pose_symm_n_monomers 4 –n pdb1_native.pdb

In Rosetta 3.1 the following command line exemplifies the simulation of a C3 system:

minirosetta.exe -run:protocol broker -broker:setup setup_init.tpb -nstruct 1 -out:file:scorefile score.fsc -database ./minirosetta_database -file:frag3 3mer_frgment_file -file:frag9 9mer_fragment_file -in:file:fasta pdb.fasta -symmetry:symmetry_definition ./sym_def_trimer.dat -out:file:silent pdb_name.out -out:file:silent_struct_type binary -relax:fast -relax:jump_move -symmetry:initialze_rigid_body_dofs -fold_and_dock::rotate_anchor_to_x -rg_reweight 0.01 -run:reinitialize_mover_for_each_job

where setup_init.tpb contains:

```
CLAIMER FoldandDockClaimer
END_CLAIMER
```

And sym_def_trimer.dat contains:

```
symmetry_name c3
subunits 3
recenter
number_of_interfaces 1
E = 3*VRT0001 + 3*(VRT0001:VRT0002)
anchor_residue 17
virtual_transforms_start
start -1,0,0 0,1,0 0,0,0
rot Rz 3
virtual_transforms_stop
connect_virtual JUMP1 VRT0001 VRT0002
connect_virtual JUMP2 VRT0002 VRT0003
set_dof BASEJUMP x(50) angle_x(360) angle_y(360) angle_z(360)
```

1. Bradley P, Misura KM, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* 309:1868–1871.

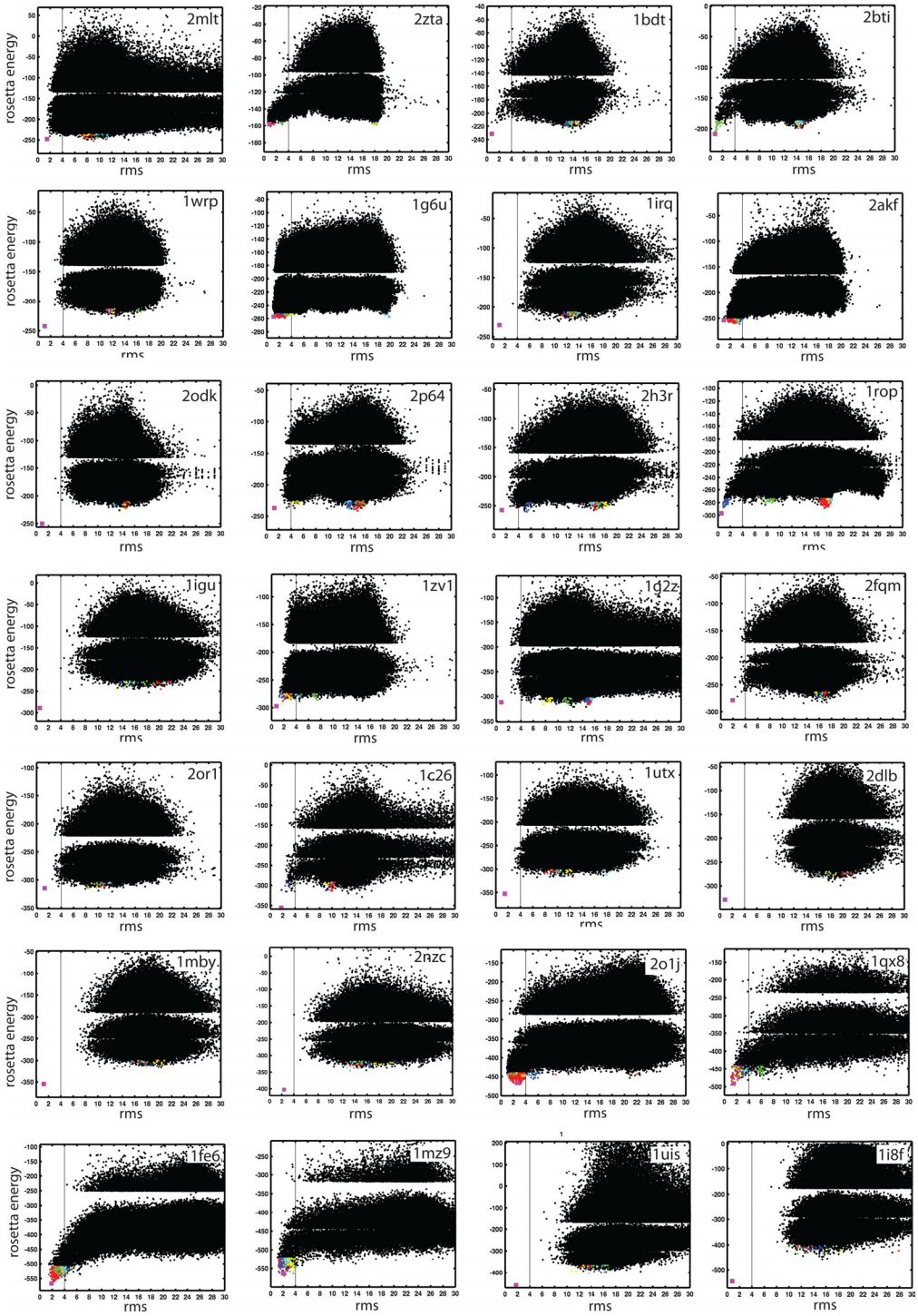

Fig. S1. Energy landscapes for proteins in the crystal structure benchmark. Rosetta all-atom energy (*y* axis) is plotted against C^{α} rmsd (*x* axis) for models generated with the fold-and-dock protocol. Colored points represent the five most populous clusters among the lowest energy models (first, red; second, green; third, blue; fourth, yellow; and fifth, cyan). Violet square, Rosetta all-atom energy of native structure after all-atom refinement. Two horizontal striations evident in each plot are due to early termination of runs that failed to pass 50th percentile cutoffs at intermediate stages of full-atom refinement, a strategy that reduces computational expense (1).

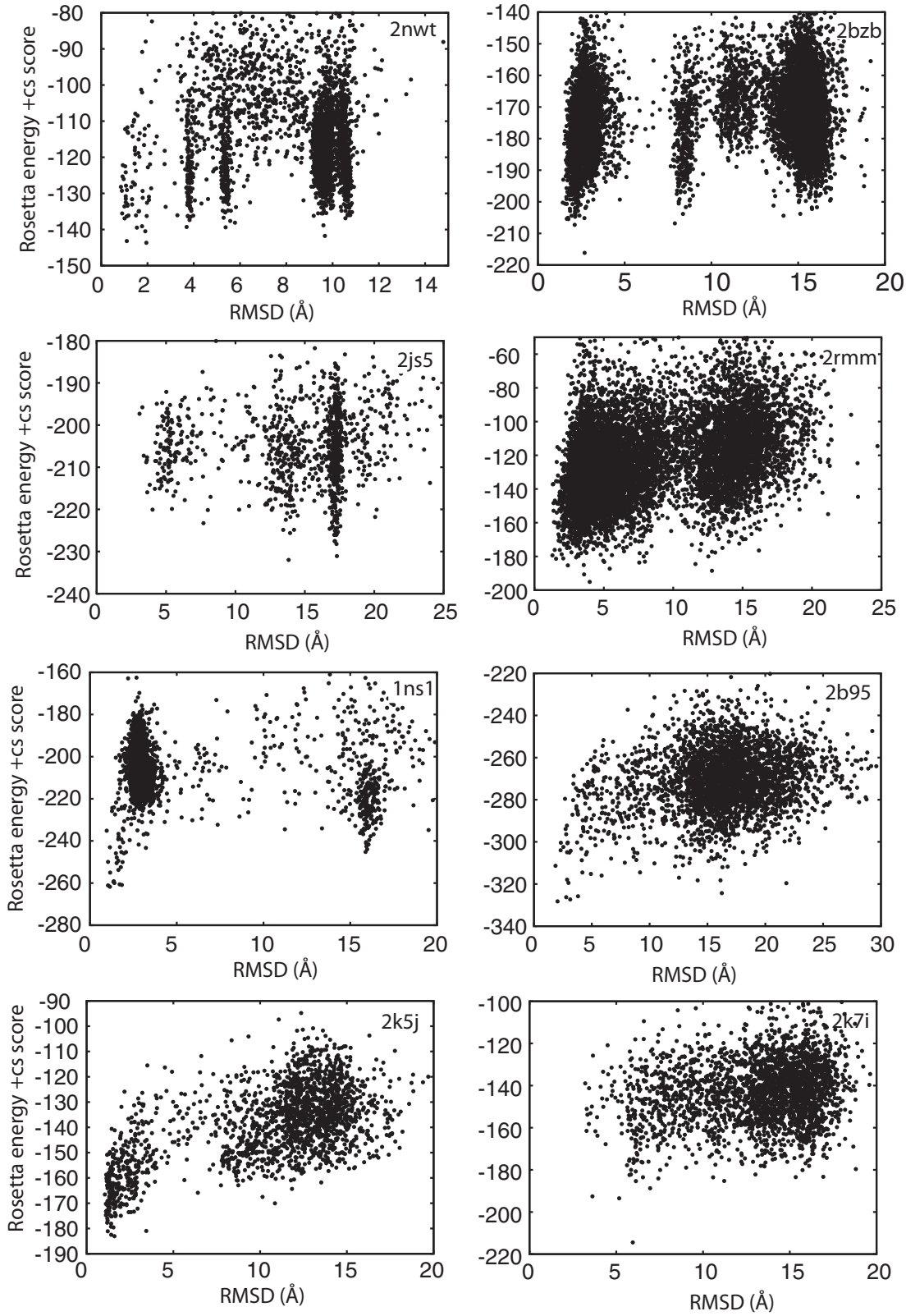

Fig. S2. Energy landscapes for proteins in the NMR structure benchmark. Rosetta all-atom energy + chemical shift score (*y* axis) is plotted against C^{α} rmsd (*x* axis) for models generated with the fold-and-dock protocol using backbone fragments generated with chemical shifts.

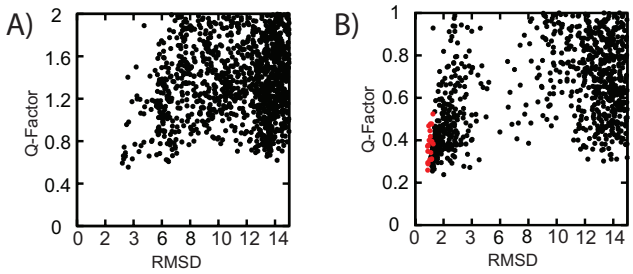

Fig. S3. Correlation between quality (*Q*)-factor (agreement with residual dipolar coupling data) and C^{α} rmsd from the native structure for low-energy Rosetta models. (*A*) 2k7i. (*B*) 2k5j. *Q*-factors for the 20 conformers in the NMR ensemble are plotted in red.

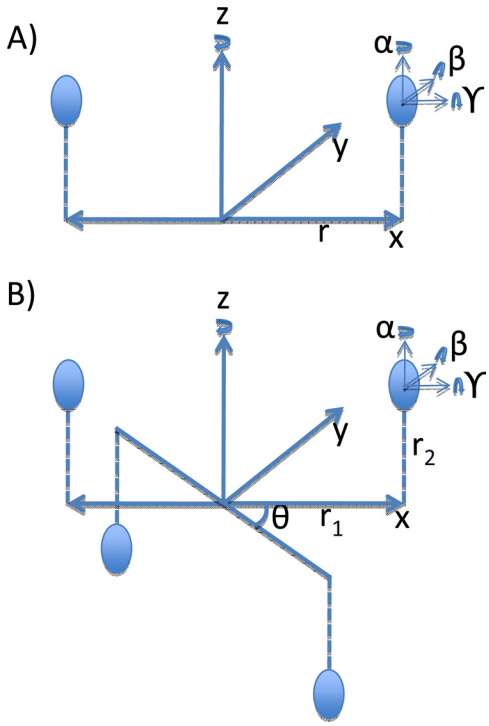

Fig. S4. Coordinate system used in the fold-and-dock simulation. (*A*) In a simulation of cyclical symmetry, there are four symmetrical degrees of freedom to sample: three rotational (α, β, and γ) and one translational (r). (*B*) In a simulation of dihedral symmetry (represented here by a D2 system), there are six degrees of freedom to sample: four rotational (α, β, γ, together with θ) and two translational (r_1 and r_2).

Table S1. Phasing results using models from traditional Rosetta modeling of single chains and from the new fold-and-dock protocol

					De novo models of single chains		Fold-and-dock models, monomers		Fold-and-dock models, complexes	
Modeled protein PDB ID	Structure factors*	Space group	No. molecules in asymm. unit	No. residues in asymm. unit	Max. TFZ score†	Frac. Residues built and sequence assigned‡	Max. TFZ score†	Frac. Residues built and sequence assigned‡	Max. TFZ score†	Frac. Residues built and sequence assigned‡
1c26	1aie	$P422$	1	31	5.0	—	5.6	0.84	0.0	—
1fe6	1ybk	$P3_121$	4	208	6.5	—	6.1	—	10.8	0.86
1g2z	1g2z	$P2_12_12$	2	64	5.0	—	5.5	—	—	—
1g6u	1g6u	$P2_13$	2	96	5.8	—	6.1	—	8.2	0.88
1i8f	1i8f	$C2$	7	567	5.3	—	5.6	—	5.9	—
1igu	1igu	$P65$	2	124	5.9	—	5.9	—	5.2	—
1mby	1mby	$P3_212$	2	176	6.0	—	5.8	—	5.4	—
1qx8	1qx8	$C2$	2	116	5.6	—	11.5	0.75	—	—
1rop	2iji	$P3_121$	1	63	8.0	0.62	10.6	0.89	—	—
1uis	1uis	$P6_522$	2	462	7.0	—	6.8	—	—	—
1utx	1utx	$P4_1$	2	132	6.3	—	6.1	—	5.1	—
1wrp	1p6z	$P2_1$	2	214	6.4	—	6.2	—	4.9	—
1zv1	1zv1	$P2_12_12_1$	2	130	6.1	—	6.2	—	21.6	0.75
2akf	2akf	$P1$	3	96	5.7	—	5.2	—	4.7	—
2bti	2bti	$P4_322$	2	126	6.2	—	6.3	—	12.7	0.79
2dlb	2dlb	$P2_12_1\ 2_1$	2	160	6.4	—	6.0	—	6.0	—
2fqm	2fqm	$P4_12_12$	6	450	5.9	—	5.2	—	5.0	—
2h3r	2h3r	$P2$	4	440	8.2	—	7.3	—	8.8	—
2mlt	2mlt	$C222_1$	2	52	8.9	—	7.3	—	—	—
2nzc	2nzc	$C222_1$	4	344	—	—	4.6	—	4.1	—
2o1j	2o1j	$C222_1$	4	208	—	—	11.6	0.25	11.9	0.52
2odk	2odk	$C2$	4	356	6.0	—	6.0	—	5.5	—
2p64	2p64	$P4_32_12$	2	104	6.7	—	7.5	0.62	10.7	0.88
2zta	2ahp	$C2$	2	66	4.8	—	5.8	0.42	5.8	0.82

*In several cases, structure factors were not deposited for the modeled sequence, but for a variant with slightly different termini or a small number of substitutions.

†Maximum Phaser Translation Function Z score observed with 400 lowest energy models. For 2nzc and 2o1j, single-chain models all gave sterically disallowed solutions in Phaser searches. For several cases, the asymmetric unit spanned only a subset of the biological complex; phasing was only carried out with monomer structures extracted from the modeled complex structures.

‡Maximum fraction of residues automatically built and sequence-assigned with the Flex-Warp version of ARP/wARP; trials were carried out with the molecular replacement models with the three highest Phaser TFZ scores.