Supporting Information

Protein Structure Determination from Pseudocontact Shifts Using ROSETTA

Christophe Schmitz, Robert Vernon, Gottfried Otting, David Baker and Thomas Huber

Table S0. Biological Magnetic Resonance Data Bank (BMRB) accession codes of chemical shift data used for target proteins.

Protein name	diamagnetic chemical shift [a]	pseudocontact shifts [a]
protein G (A)	BMRB7280	Ref 34
calbindin (B)	BMRM6699	Ref 4
θ subunit (C)	BRMB6571	Ref 37
ArgN (D)	Ref 21	Ref 21
ArgN (E)	Ref 21	Ref 38
N-calmodulin (F)	Ref 39	Ref 39
thioredoxin (G)	BRMB1813 [b]	Ref 42
parvalbumin (H)	BRMB6049	Ref 43
calmodulin (I)	BRMB15852	BRMB7423, BRMB7424, BRMB7425 and ref 44
ε186 (J)	BRMB6184	Ref 46

[a] Reference numbers from the list of references in main text are given when data was not available in BMRB
[b] only H^N and ^{15}N chemical shifts

Table S1. PCS data information and grid search parameters used.

Protein name	Residues[a]	Metal ions used	Atom types	cs corr[b]	*w(c)*	*cg*[c]	*sg*[d]	*co*[d]	*ci*[d]
protein G (A)	1-56	Tb^{3+}, Tm^{3+}, Er^{3+}	H^N	0.53	15.5	E19 CA	6	17	7
calbindin (B)	2-75	Ce^{3+}, Dy^{3+}, Er^{3+}, Eu^{3+}, Ho^{3+}, Nd^{3+}, Pr^{3+}, Sm^{3+}, Tb^{3+}, Tm^{3+}, Yb^{3+}	H^N, N, C'	2.72	1.98	D54 CA	6	8	4
θ subunit (C)	10-64	Dy^{3+}, Er^{3+}	H^N	-0.16	7.1	D14 CA	6	25	15
ArgN (D)	8-70	Tb^{3+}, Tm^{3+}, Yb^{3+}	H^N, N	2.09	13.5	C68 CB	6	10	4
ArgN (E)	8-70	Tb^{3+}, Tm^{3+}	H^N	2.09	48.9	K12 CB	6	15	0
N-calmodulin (F)	3-79	Tb^{3+}, Tm^{3+}	H^N, CA, CB	0.00	4.7	D60 CA	6	8	4
thioredoxin (G)	2-108	Ni^{2+}	H^N	1.23	106.3	S1 N	3.8	4	0
parvalbumin (H)	2-109	Dy^{3+}	H^N, N	2.65	2.86	D93 CA	6	8	4
calmodulin (I)	3-146	Tb^{3+}, Tm^{3+}, Yb^{3+}, Dy^{3+}	H^N	0.59	5.1	D60 CA	6	8	4
ε186 (J)	7-180	Tb^{3+}, Dy^{3+}, Er^{3+}	H^N, N, C'	0.53	8.2	D12 CA	6	8	4

[a] Ordered residues

[b] Uniform offset used for ^{13}C chemical shifts (in ppm) compared to published values. In the case of thioredoxin, the offset was applied to ^{15}N chemical shifts

[c] Residue and atom name defining the center of the grid search to position the paramagnetic center.

[d] In Ångström

Table S2. Comparison of PCS-ROSETTA and CS-ROSETTA, evaluating their performance only for the structured core residues defined in Table S1.

Targets	PCS-ROSETTA run[a]		CS-ROSETTA run[b]	
	rmsd[c]	convergence[d]	rmsd[c]	convergence[d]
protein G (A)	0.61	0.92	0.80	0.88
calbindin (B)	1.46	2.09	4.96	4.72
θ subunit (C)	1.30	0.55	1.56	2.25
ArgN[e] (D)	1.00	0.77	1.31	2.21
ArgN[f] (E)	0.83	0.94	1.65	5.43
N-calmodulin (F)	1.74	1.49	4.69	4.49
thioredoxin (G)	2.58	2.44	4.61	5.55
parvalbumin (H)	11.26	10.25	11.80	11.30
calmodulin (I)	2.80	2.12	6.35	2.94
ε186 (J)	20.57	18.03	17.07	17.74

[a] The structures used to calculate the rmsds were identified using the combined PCS-score and ROSETTA full atom energy across the core residues.

[b] Using the ROSETTA full-atom energy across the core residues.

[c] C^{α} rmsd (with respect to the native structure) of the structure of lowest score, in Å. All C^{α} rmsd values were calculated using the core residues.

[d] Average C^{α} rmsd calculated between the lowest score structure and the next four lowest scoring structures, in Å.

[e] PCSs measured with a covalent tag attached to the N-terminal domain of the *E. coli* arginine repressor (ArgN).

[f] PCSs measured with a non-covalent tag bound to ArgN.

Text S1. Fragment Assembly Using PCSs Only. In order to gain a better understanding of the merit of PCS data, we generated 10000 decoys per protein with all ROSETTA force field components turned off except for the PCS score. In seven of the ten protein structure calculations, the PCS score alone produced decoys with a C^{α} rmsd of less than 2.5 Å to the target structure (Figure S2, solid blue line). Control calculations without any scoring function produced not a single useful decoy. This highlights the power of PCS data to define the overall topology of a protein at the fragment assembly stage. The effect was particularly pronounced for the target proteins θ and ArgN (Figure S2 C and D).

The second set of PCS data of ArgN (Table 1; structure E) yielded worse decoys in the PCS-only computations with PCS-ROSETTA than CS-ROSETTA. Remarkably, however, using the PCS score in combination with the ROSETTA force field yielded much better structures than when used separately (Figure S3 E). This shows that the PCS score adds information that is not captured by the ROSETTA energy score alone.

Text S2. Scoring over Core Residues. Disordered residues can add noise to the ROSETTA energy, and this noise can prevent identification of low rmsd structures. Notably, three of the targets that succeeded under the PCS-ROSETTA protocol and failed under the CS-ROSETTA protocol (targets C, D, and E in Table 1) have disordered termini accounting for ten or more residues each. In practice, the disordered character of N- and C-terminal polypeptide segments can readily be identified by NMR spectroscopy. Therefore, we produced an additional set of structures by removing disordered N- and C-terminal peptide segments before the final rescoring step and retaining only the core residues defined in Table S1. Knowledge of the target structures allowed perfect identification of the core residues. Selection of the core residues only improved the capability of the CS-ROSETTA protocol to identify low rmsd structures in four of the ten cases (including targets C, D, and E), and produced convergence to a low rmsd structure in three of the ten cases (targets C, D, and E in Table S2). In contrast, removing the disordered residues had little effect on the rmsd values achieved with PCS-ROSETTA, indicating that the combined PCS and ROSETTA score greatly alleviates the sensitivity to disordered polypeptide segments. The remaining targets had few or no disordered residues and removal of disordered terminal residues had little effect on the results.

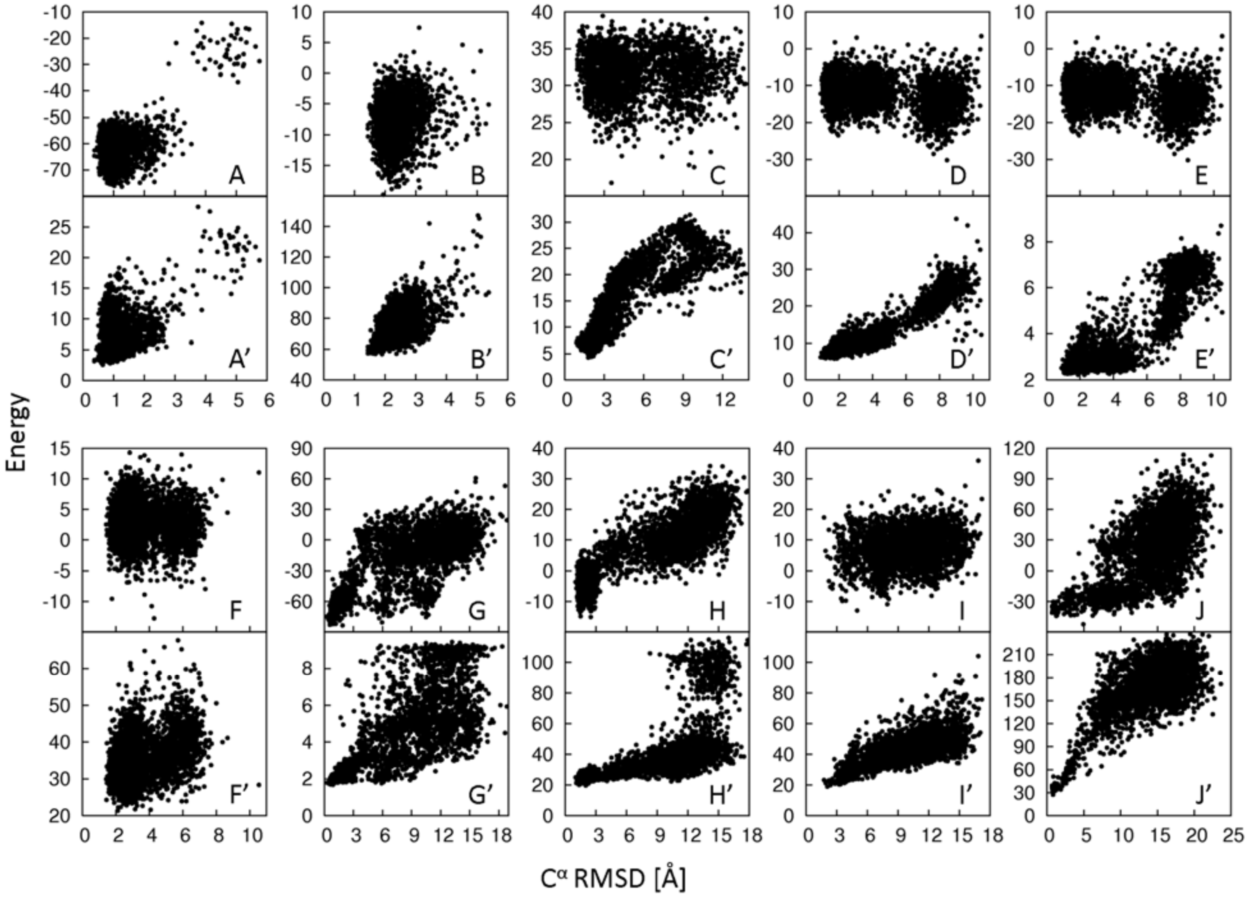


Figure S1. Fold identification by pseudocontact shift score and ROSETTA energy. 3000 decoys were generated using CS-ROSETTA. In order to ensure that some decoys with small rmsd to the target structure were obtained, the starting set of peptide fragments was reduced and included the fragments from the known target structures. A to J: ROSETTA energies plotted versus the C^{α} rmsd to the target structure. A' to J': PCS scores plotted versus the C^{α} rmsd to the target structure. The targets are labeled A-J as in Table 1.

Table S3. Correlation coefficients between rmsd and score in fold identification calculations (Figure 1 and S1). The targets are labeled A-J as in Table 1.

	total score		ROSETTA score		PCS score	
	r[a]	ρ[b]	r[a]	ρ[b]	r[a]	ρ[b]
protein G (A)	0.64	0.07	0.50	-0.06	0.59	0.34
calbindin (B)	0.62	0.52	0.17	0.17	0.63	0.52
θ subunit (C)	0.76	0.83	0.03	0.02	0.81	0.88
ArgN[c] (D)	0.72	0.69	-0.26	-0.23	0.93	0.91
ArgN[d] (E)	0.07	0.06	-0.26	-0.23	0.88	0.77
N-calmodulin (F)	0.32	0.32	-0.03	-0.01	0.36	0.35
thioredoxin (G)	0.81	0.80	0.80	0.79	0.78	0.83
parvalbumin (H)	0.79	0.90	0.84	0.84	0.65	0.86
calmodulin (I)	0.65	0.65	0.21	0.20	0.68	0.69
ε186 (J)	0.77	0.64	0.65	0.60	0.70	0.52

[a] Pearson correlation coefficient.

[b] Spearman’s rank correlation coefficient.

[c] PCSs measured with a covalent tag attached to the N-terminal domain of the *E. coli* arginine repressor (ArgN).

[d] PCSs measured with a non-covalent tag bound to ArgN.

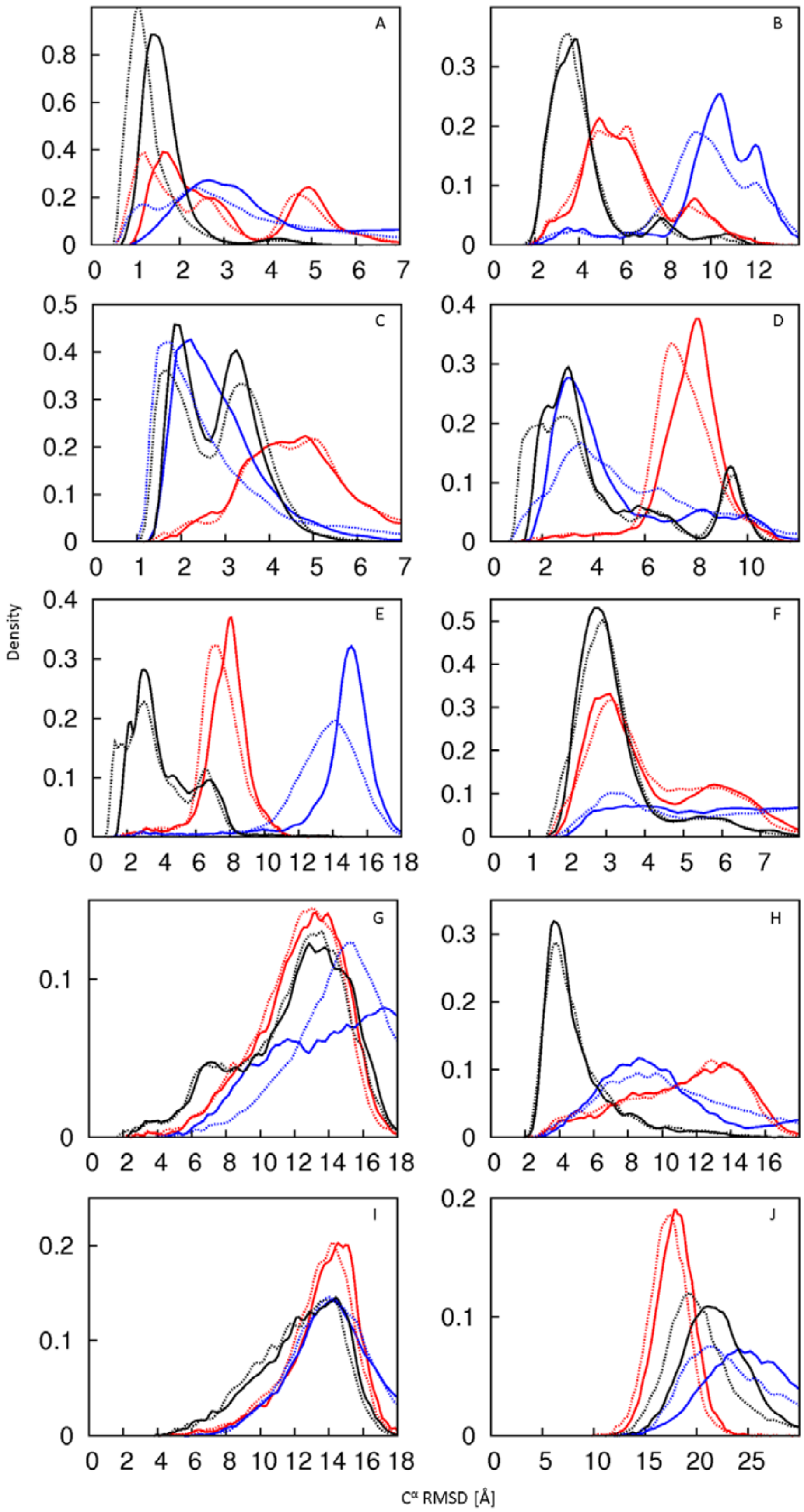

A
0.8
0.6
0.4
0.2
0
0 1 2 3 4 5 6 7
B
0.3
0.2
0.1
0
0 2 4 6 8 10 12
C
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7
D
0.4
0.3
0.2
0.1
0
0 2 4 6 8 10
E
0.4
0.3
0.2
0.1
0
0 2 4 6 8 10 12 14 16 18
F
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7
G
0.1
0
0 2 4 6 8 10 12 14 16 18
H
0.3
0.2
0.1
0
0 2 4 6 8 10 12 14 16
I
0.2
0.1
0
0 2 4 6 8 10 12 14 16 18
J
0.2
0.1
0
0 5 10 15 20 25
Density
Cα RMSD [Å]

Figure S2. Improved fragment assembly by PCS-ROSETTA. Fragments were assembled in 10000 different runs of CS-ROSETTA (red), 10000 different runs of PCS-ROSETTA (black), and 10000 different runs using exclusively the PCS score of PCS-ROSETTA (blue). The plots show the frequency with which structures of different C^{α} rmsd values to the target structure were found. The red and black solid lines reproduce the data of Figure 2. The dashed lines show the corresponding data obtained in independent calculations that included the full atom refinement step. The same colors were used for calculations with and without the full atom refinement step. The full atom refinement step does not significantly change the C^{α} rmsd of the structures produced in the fragment assembly step with respect to the target structure. The targets are labeled A-J as in Table 1.

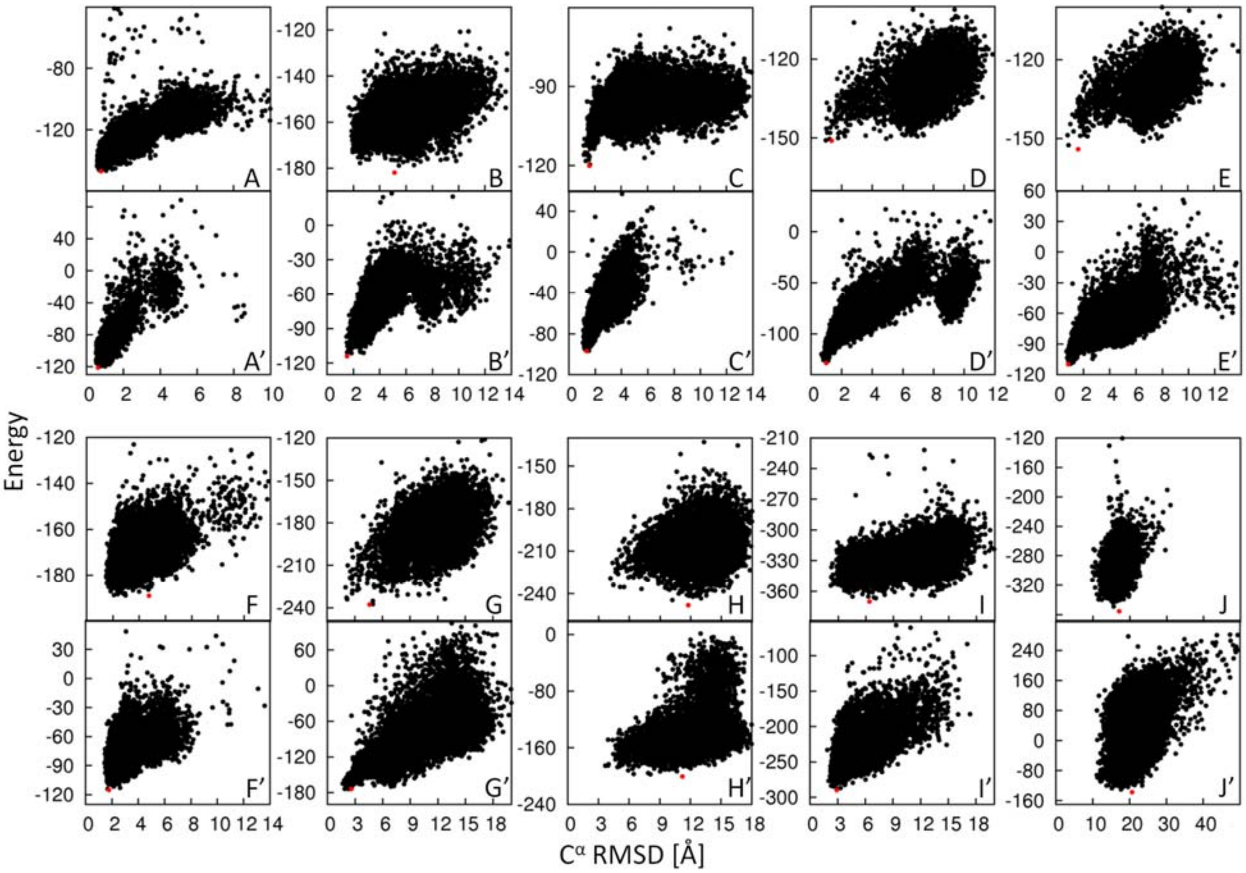


Figure S3. Energy landscape generated by CS-ROSETTA and PCS-ROSETTA, with full atom ROSETTA energies and C$^{\alpha}$ rmsd values calculated using only the core residues as defined in Table S1. A to J: full atom ROSETTA energies plotted versus the C$^{\alpha}$ rmsd to the target structure for structures calculated using CS-ROSETTA. A' to J': Combined ROSETTA energy and PCS score plotted versus the C$^{\alpha}$ rmsd to the target structure for structures calculated using PCS-ROSETTA. The lowest energy structures are indicated in red. The targets are labeled A-J as in Table 1.

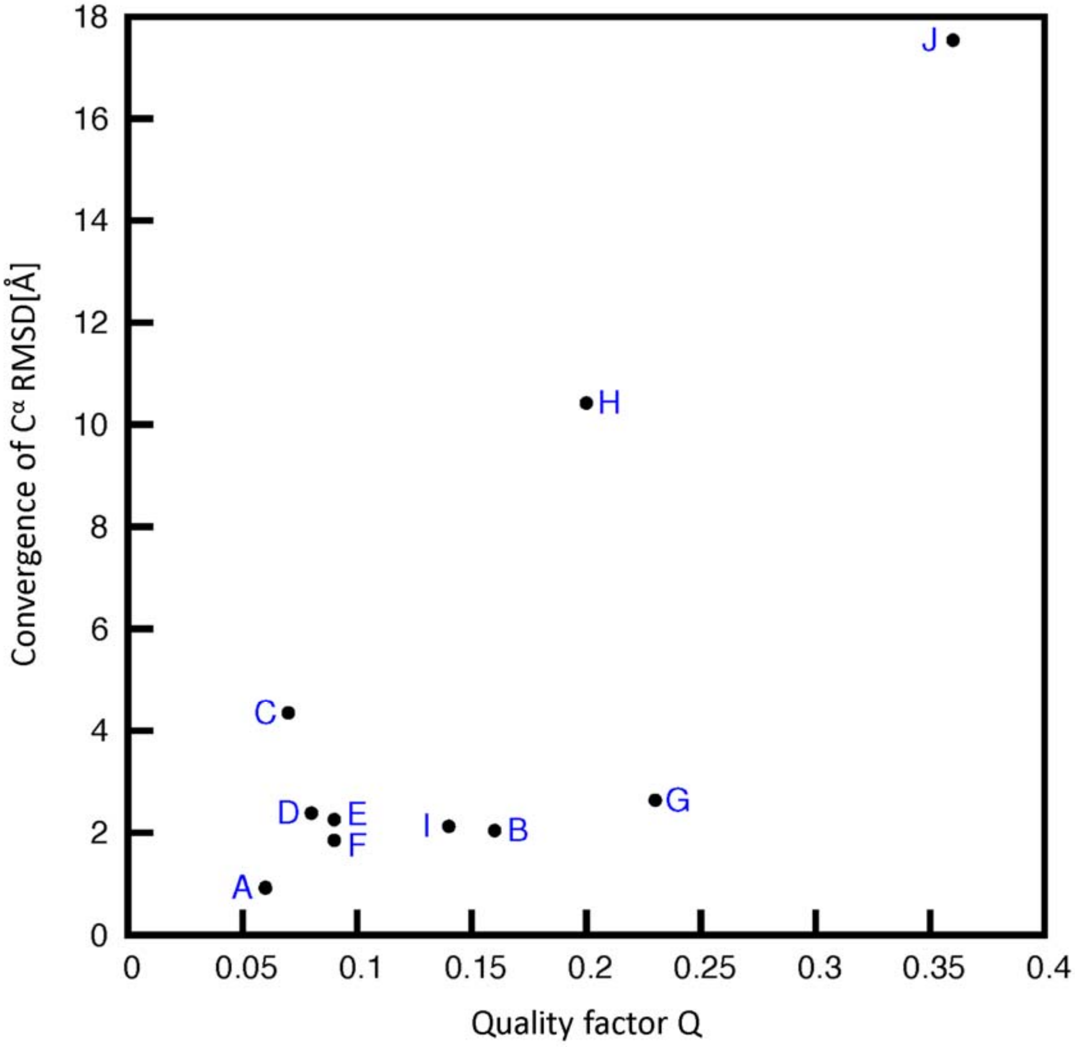


Figure S4. Identification of successful calculations with PCS-ROSETTA. The quality factor Q reports on the agreement between the experimental and calculated PCSs. A value below 20% indicates that the calculated structure satisfies the PCS restraints well. Above 25%, the quality of the structure is poor. The y axis displays the average C^α rmsd value calculated between the structure with the lowest score and the next four lowest scoring structures. Rmsd values below 3 Å indicate convergence of the protocol. Convergence criterion and quality factor can be combined to further ascertain the success of the calculations for the targets A, B, C, D, E, F, G, and I, and reject targets H and J. The targets are labeled A-J as in Table 1. The plot displays the figures of columns 7 and 8 of Table 1 on the y and x axis, respectively.

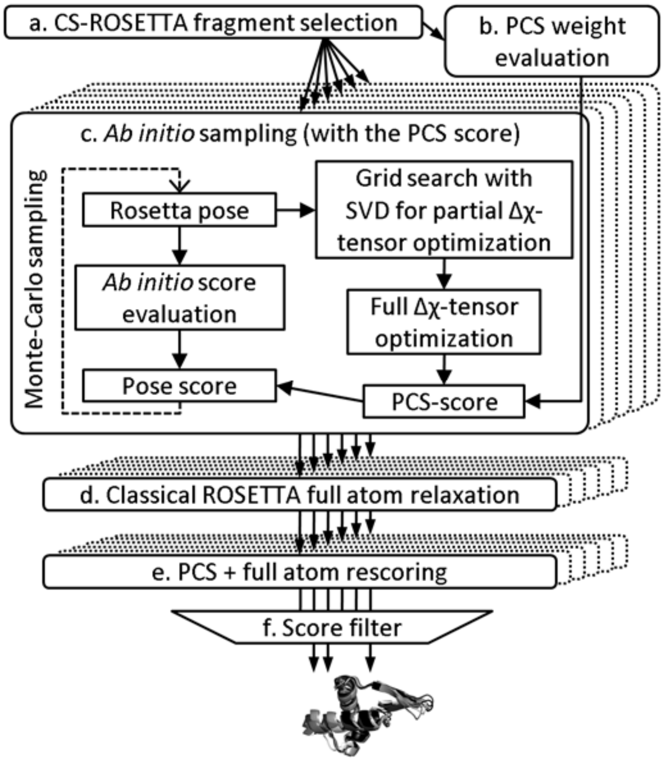

a. CS-ROSETTA fragment selection
b. PCS weight evaluation
c. *Ab initio* sampling (with the PCS score)
Monte-Carlo sampling
Rosetta pose
Grid search with SVD for partial Δχ-tensor optimization
Ab initio score evaluation
Full Δχ-tensor optimization
Pose score
PCS-score
d. Classical ROSETTA full atom relaxation
e. PCS + full atom rescoring
f. Score filter

Figure S5. Flow diagram of PCS-ROSETTA. (a) Fragments are selected by their chemical shifts using CS-ROSETTA. (b) The PCS weight is calculated using Eq. 4 on 1000 decoys generated with CS-ROSETTA. (c) Structures are produced by the classical fragment assembly protocol of ROSETTA with addition of the PCS-score. (d) Side chains are added to the structures and subjected to a full atom minimization. (e) Resulting structures are rescored using a combination of the ROSETTA full atom energy score and the PCS score. (f) Best structures are selected by their lowest score.

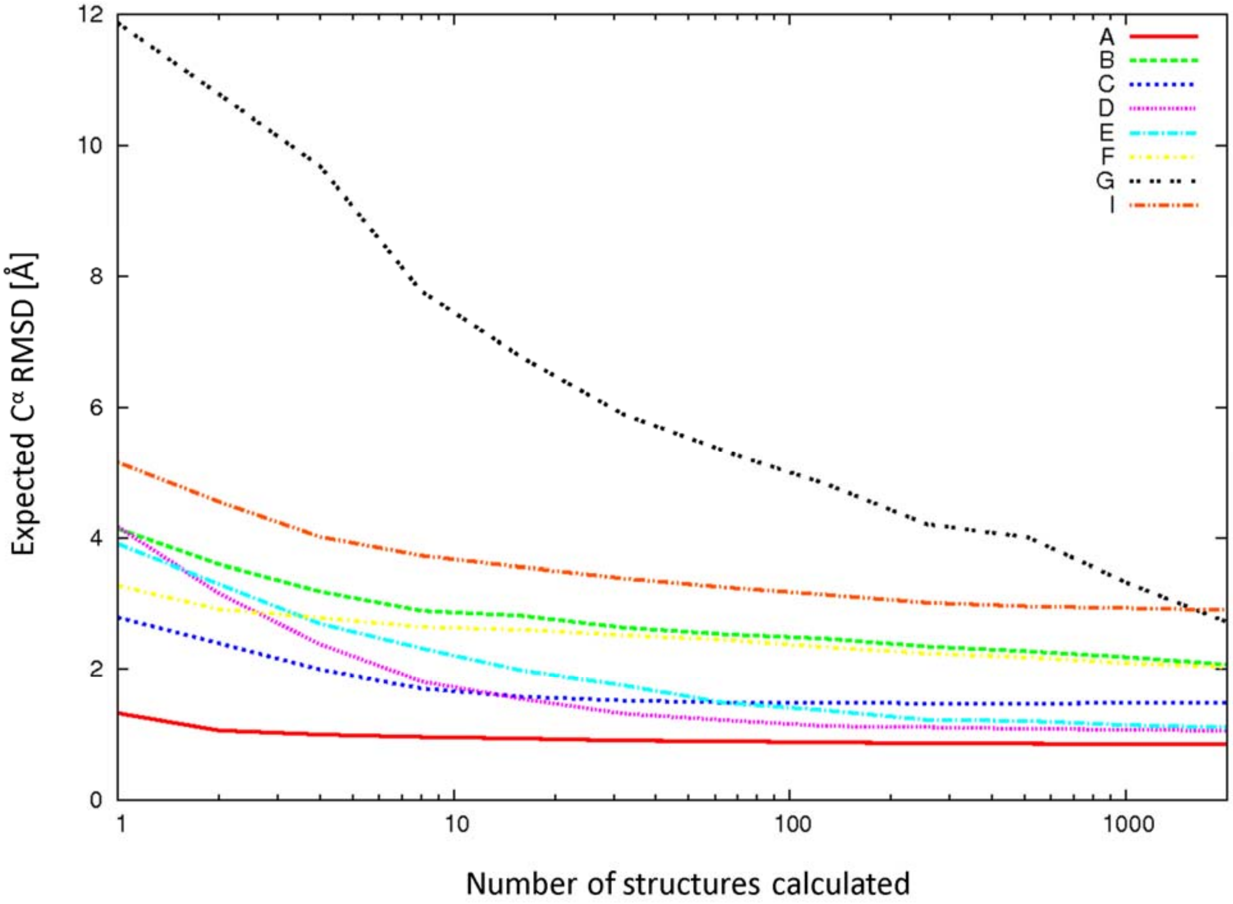


Figure S6. Expected C^{α} rmsd of the lowest energy structure calculated with PCS-ROSETTA. A given number n of structures (x axis) was randomly chosen 5000 times from the total of 10 000 generated structures and the average of the C^{α} rmsd of the lowest energy structure found in each of the 5000 trials is graphed. The curves show a posteriori that 1000 structures calculated for all the targets would have been sufficient to ensure convergence of the PCS-ROSETTA calculations. The targets are labeled A-J as in Table 1. The curves for the targets parvalbumin (H) and ε186 (J) are not shown since they didn't converge.

Table S4. Comparison of $\Delta\chi$–tensor parameters from reference target structure and lowest energy structure calculated with PCS-ROSETTA. The targets are labeled A-J as in Table 1.

		target structure		lowest energy structure					
	metal ion	$\Delta\chi_{ax}$ / 10^{-32} m^3	$\Delta\chi_{rh}$ / 10^{-32} m^3	$\Delta\chi_{ax}$ / 10^{-32} m^3	$\Delta\chi_{rh}$ / 10^{-32} m^3	α-x[a] / degree	α-y[b] / degree	α-z[c] / degree	d_{MM}[d] / Å
protein G (A)	Er^{3+}	11.7	3.5	10.7	4.0	11.9	14.9	17.5	1.39
	Tb^{3+}	41.6	22.5	30.9	25.4	13.9	13.3	18.8	
	Tm^{3+}	28.2	7.4	27.0	6.1	11.7	26.0	28.6	
calbindin (B)	1	2.1	0.8	2.6	0.8	5.6	7.0	7.1	1.45
	2	2.8	1.4	4.4	1.9	7.3	21.0	22.1	
	3	1.6	0.4	2.2	0.6	8.1	11.0	13.6	
	4	-1.8	-0.6	-1.7	-0.9	10.6	9.3	13.6	
	5	41.8	9.4	39.9	18.6	4.0	1.8	4.1	
	6	31.6	19.9	36.5	18.1	9.4	20.3	19.7	
	7	17.8	4.4	22.1	4.6	14.7	28.3	26.6	
	8	-11.7	-7.3	-15.5	-8.1	7.9	12.2	10.2	
	9	25.9	13.7	31.3	21.3	9.9	8.5	10.4	
	10	7.1	4.1	7.9	5.2	5.7	1.2	5.5	
	11	0.3	0.0	0.4	0.2	67.5	79.5	45.2	
θ subunit (C)	Dy^{3+}	65.9	25.6	27.4	14.9	25.0	69.3	74.6	7.25
	Er^{3+}	-17.7	-9.2	-6.7	-1.5	41.2	77.9	86.2	
ArgN[e] (D)	Tb^{3+}	-11.6	-7.7	-9.9	-7.3	4.4	6.6	5.7	1.53
	Tm^{3+}	12.5	7.7	11.3	6.4	4.0	5.6	4.8	
	Yb^{3+}	-6.5	-4.1	-4.5	-4.2	18.3	11.4	15.5	
ArgN[f] (E)	Tb^{3+}	-7.5	-1.6	-7.2	-0.6	10.4	8.8	5.7	2.29
	Tm^{3+}	4.1	0.7	3.7	0.5	40.8	6.0	41.0	
N-calmodulin (F)	Tb^{3+}	35.5	16.7	18.9	15.9	28.4	14.9	24.2	1.37
	Tm^{3+}	28.1	10.6	20.2	6.6	11.2	4.6	12.0	
thioredoxin (G)	Ni^{2+}	-1.1	-0.6	-0.9	-0.9	16.4	22.2	22.6	3.19
parvalbumin (H)	Dy^{3+}	31.1	12.2	26.3	7.0	38.6	35.3	44.3	3.24
calmodulin (I)	Tb^{3+}	36.7	19.2	38.1	17.0	21.7	8.8	22.1	1.03
	Tm^{3+}	26.1	12.2	23.1	14.8	6.1	3.8	7.0	
	Yb^{3+}	10.1	1.7	9.6	3.8	3.8	4.4	3.5	
ε186 (J)	Dy^{3+}	39.8	4.4	133.3	193.3	67.3	40.9	63.8	23.21
	Er^{3+}	-10.2	-4.4	-37.3	-44.6	29.1	3.5	29.1	
	Tb^{3+}	27.3	5.5	89.4	102.4	47.6	39.1	42.8	

[a] Angle between the x-axes of the $\Delta\chi$–tensors of the target and the calculated structure.

[b] Angle between the y-axes of the $\Delta\chi$–tensors of the target and the calculated structure.

[c] Angle between the z-axes of the $\Delta\chi$–tensors of the target and the calculated structure.

[d] Distance between the metal ion position of the target and the calculated structure.

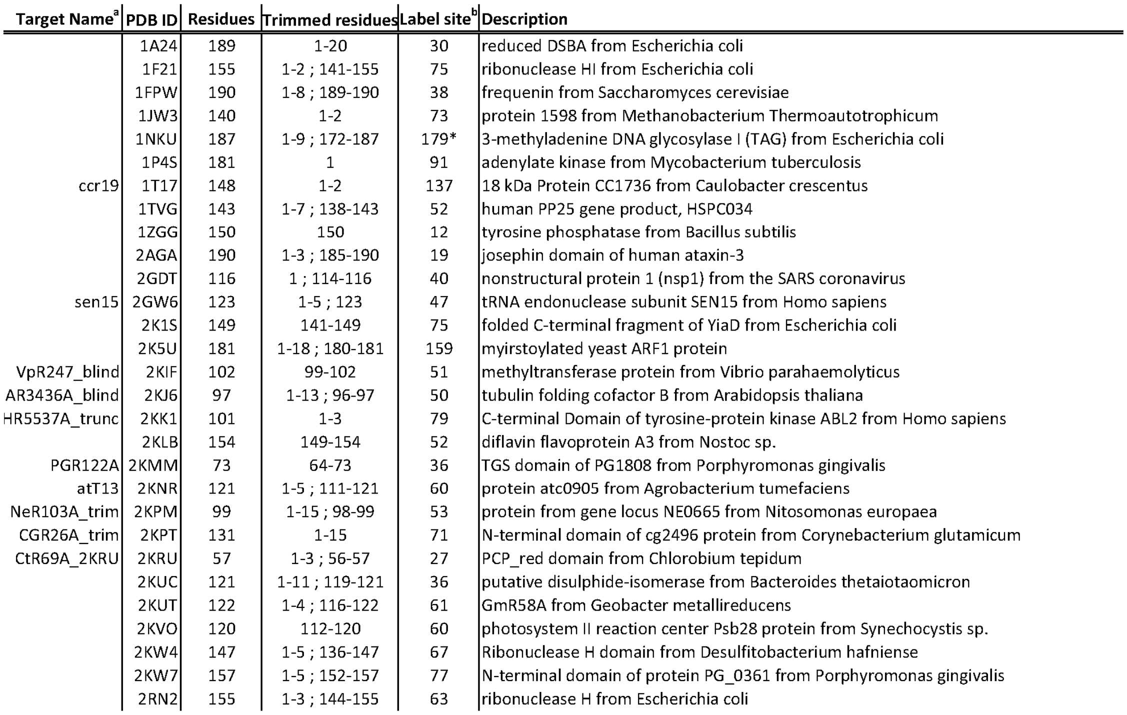
[e] PCSs measured with a covalent tag attached to the N-terminal domain of the *E. coli* arginine repressor (ArgN).

[f] PCSs measured with a non-covalent tag bound to ArgN.

Text S3. PCS-ROSETTA on large proteins. Due to the long-range nature of PCS data, PCS-ROSETTA could potentially be suitable for 3D structure determinations of much larger proteins than CS-ROSETTA. To test this hypothesis, we performed extensive PCS-ROSETTA calculations with simulated PCS data, using 29 proteins (Table S5) that had either failed previously to converge with CS-ROSETTA and/or are larger in size. For each protein, a lanthanide ion was positioned at a single site and H^N and ^{15}N PCS data were generated using the three lanthanide $\Delta\chi$-tensors (Tb^{3+}, Tm^{3+}, Yb^{3+}) of target D (ArgN), allowing for experimental errors of ±0.05 ppm and excluding residues closer than 12 Å to the paramagnetic center to account for line broadening beyond detection arising from paramagnetic relaxation enhancements. Using the same number and type of lanthanide labels allowed a stringent comparison of PCS-ROSETTA with CS-ROSETTA. The structure calculations followed the protocol described in the main text. The calculations took on average 200 CPU days per target on a local cluster.

The results from the PCS-ROSETTA calculation on this test set of challenging proteins confirmed the trends observed in our calculations on proteins with experimentally determined PCS data. While the inclusion of PCS data did not always produce low rmsd values to the target structure where CS-ROSETTA had failed, sampling of more native-like conformations nonetheless improved consistently for all protein sizes. Figure S7 compares the C^α rmsd density distributions of structures generated with CS-ROSETTA and PCS-ROSETTA. In all but two cases low rmsd structures were more often generated with PCS-ROSETTA, indicating that the availability of long range PCS restraints extended the radius within which elements of natively formed (sub-)structures were recognized, even for structures with rmsd values of 5 Å or greater to the target structure. This result is remarkable, as structures that are very different from the native structure are generally associated with low quality $\Delta\chi$-tensors. Clearly, however, even the restraints from poorly determined $\Delta\chi$-tensors improved the quality of the structures sampled, as well as helping to discriminate wrong folds from structures with native-like elements. This effect is illustrated in Figure S8 and S9. Interestingly, the biggest improvement is in the remote similarity range (Global distance test (Zemla et al. 1999) GDT 0.4-0.7; RMSD 10-5 Å) where partially correct topologies are present in the generated structure but it is notoriously difficult to recognize these elements and improve the fold. Reliable, accurate 3D structure determinations of large proteins is likely to require the combination of improved sampling convergence and recognition of native-like sub-structures in protein models as demonstrated here with new computational approaches such as broken chain sampling or iterative refinement.

Table S5. Protein structures with simulated PCS data used to evaluate the performance of PCS-ROSETTA.



Target Name[a]	PDB ID	Residues	Trimmed residues	Label site[b]	Description
	1A24	189	1-20	30	reduced DSBA from Escherichia coli
	1F21	155	1-2 ; 141-155	75	ribonuclease HI from Escherichia coli
	1FPW	190	1-8 ; 189-190	38	frequenin from Saccharomyces cerevisiae
	1JW3	140	1-2	73	protein 1598 from Methanobacterium Thermoautotrophicum
	1NKU	187	1-9 ; 172-187	179*	3-methyladenine DNA glycosylase I (TAG) from Escherichia coli
	1P4S	181	1	91	adenylate kinase from Mycobacterium tuberculosis
ccr19	1T17	148	1-2	137	18 kDa Protein CC1736 from Caulobacter crescentus
	1TVG	143	1-7 ; 138-143	52	human PP25 gene product, HSPC034
	1ZGG	150	150	12	tyrosine phosphatase from Bacillus subtilis
	2AGA	190	1-3 ; 185-190	19	josephin domain of human ataxin-3
	2GDT	116	1 ; 114-116	40	nonstructural protein 1 (nsp1) from the SARS coronavirus
sen15	2GW6	123	1-5 ; 123	47	tRNA endonuclease subunit SEN15 from Homo sapiens
	2K1S	149	141-149	75	folded C-terminal fragment of YiaD from Escherichia coli
	2K5U	181	1-18 ; 180-181	159	myirstoylated yeast ARF1 protein
VpR247_blind	2KIF	102	99-102	51	methyltransferase protein from Vibrio parahaemolyticus
AR3436A_blind	2KJ6	97	1-13 ; 96-97	50	tubulin folding cofactor B from Arabidopsis thaliana
HR5537A_trunc	2KK1	101	1-3	79	C-terminal Domain of tyrosine-protein kinase ABL2 from Homo sapiens
	2KLB	154	149-154	52	diflavin flavoprotein A3 from Nostoc sp.
PGR122A	2KMM	73	64-73	36	TGS domain of PG1808 from Porphyromonas gingivalis
atT13	2KNR	121	1-5 ; 111-121	60	protein atc0905 from Agrobacterium tumefaciens
NeR103A_trim	2KPM	99	1-15 ; 98-99	53	protein from gene locus NE0665 from Nitosomonas europaea
CGR26A_trim	2KPT	131	1-15	71	N-terminal domain of cg2496 protein from Corynebacterium glutamicum
CtR69A_2KRU	2KRU	57	1-3 ; 56-57	27	PCP_red domain from Chlorobium tepidum
	2KUC	121	1-11 ; 119-121	36	putative disulphide-isomerase from Bacteroides thetaiotaomicron
	2KUT	122	1-4 ; 116-122	61	GmR58A from Geobacter metallireducens
	2KVO	120	112-120	60	photosystem II reaction center Psb28 protein from Synechocystis sp.
	2KW4	147	1-5 ; 136-147	67	Ribonuclease H domain from Desulfitobacterium hafniense
	2KW7	157	1-5 ; 152-157	77	N-terminal domain of protein PG_0361 from Porphyromonas gingivalis
	2RN2	155	1-3 ; 144-155	63	ribonuclease H from Escherichia coli

[a] CS-ROSETTA targets in the CASD experiment (Rosato et al. 2009) or difficult targets (ccr19 and sen15) in the CS-ROSETTA benchmark (Shen et al. 2008).

[b] The selection of lanthanide labeling sites was guided by native cysteine residues and the lanthanide ion was placed 4 Ångstrom from the C^{β}–atom along the C^{α}- C^{β} bond, consistent with experimental results for small lanthanide binding tags (Su et al. 2008, Man et al 2010, Jia et al. 2011). For proteins without cysteine residue, a solvent exposed residue located approximately in the middle of the amino acid sequence was chosen. For proteins with multiple cysteines, a solvent exposed cysteine residue was used. In the case of 1NKU the natural metal binding site of the protein was used as the paramagnetic center and the coordinating residue 179 listed in the table identifies its position in the sequence (marked by an asterisk).

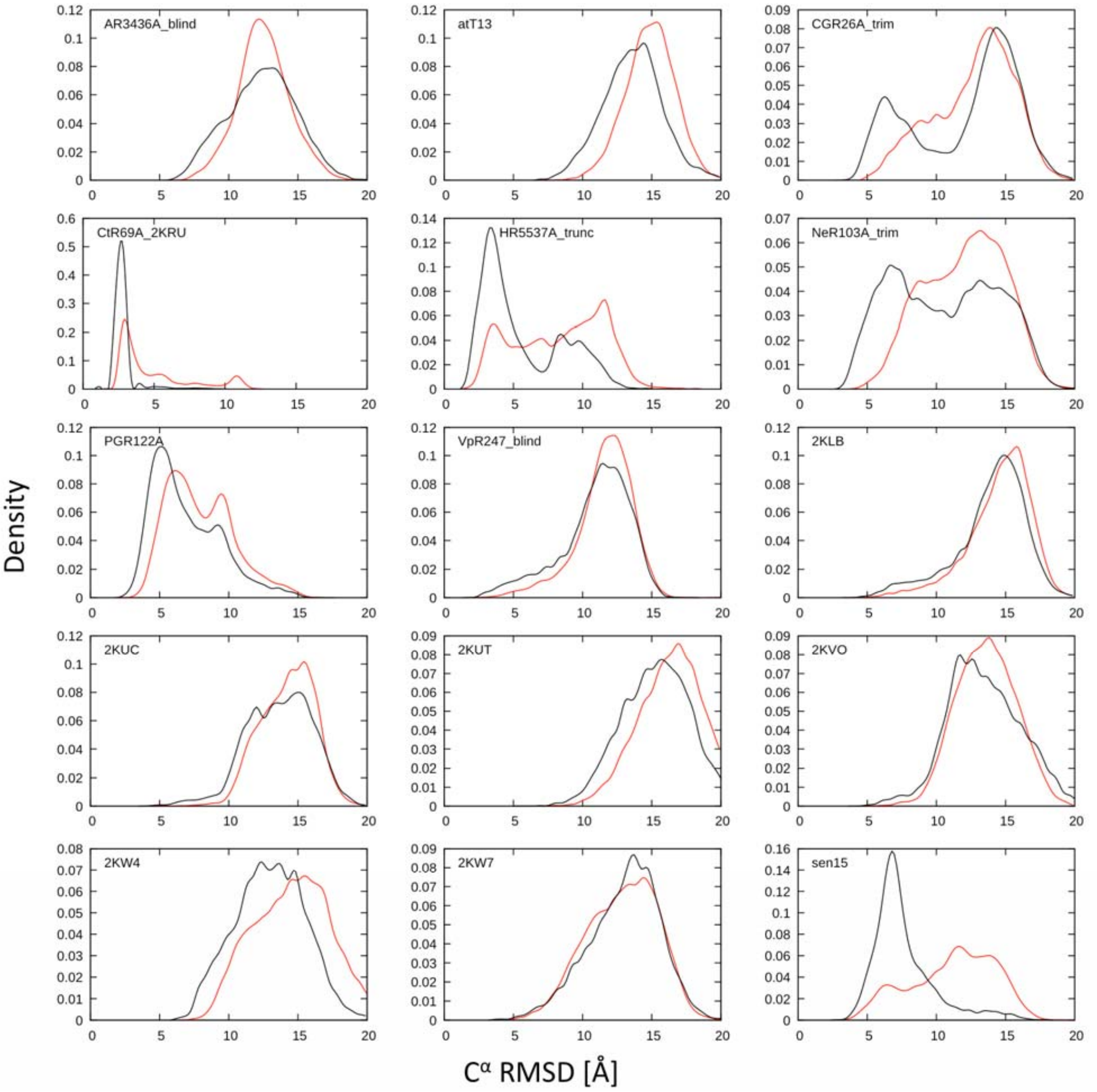


Figure S7. C^{α} rmsd density distributions of structures generated by PCS-ROSETTA (black) and CS-ROSETTA (red) for 29 test proteins with simulated PCS data.

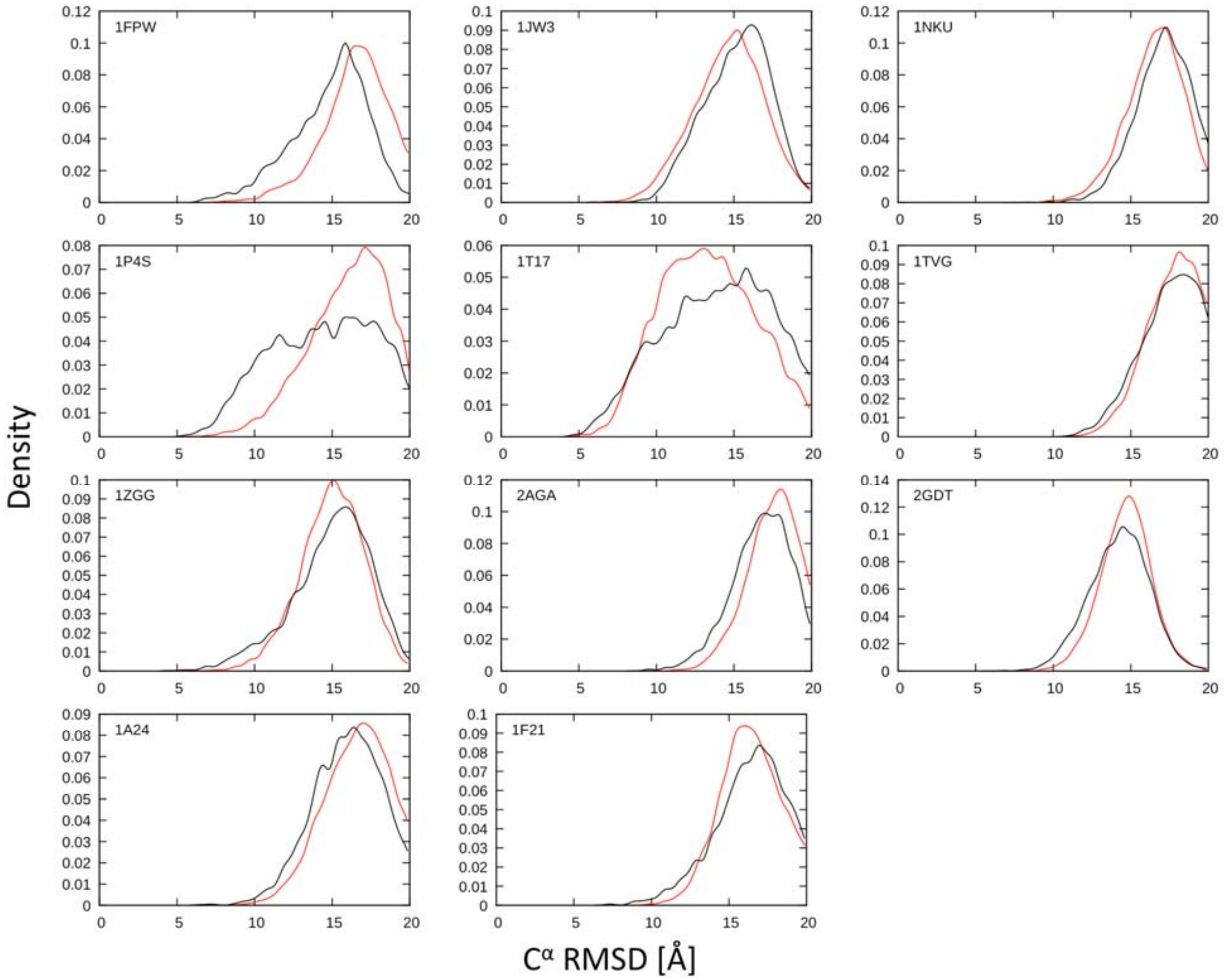


Figure S7 (continued).

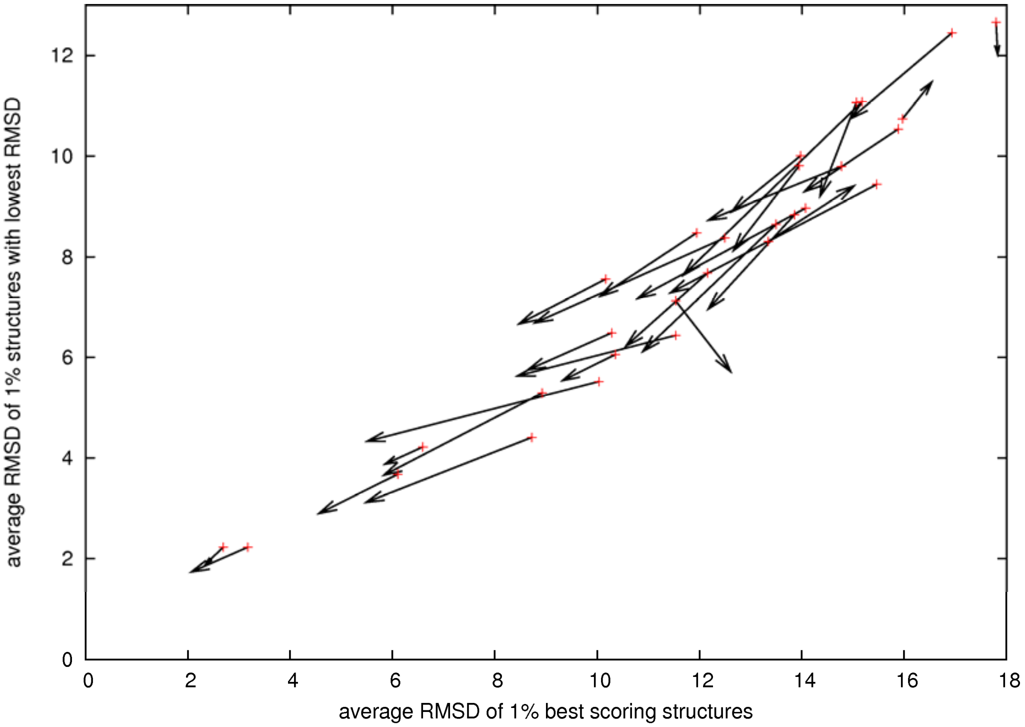


Figure S8. Consistent improvement of sampling and recognition of structures with lower C^{α} rmsd to the target structure. The x-axis of the plot shows the average C^{α} rmsd value of the best scoring 1% of the structures and the y-axis of the plot shows the average C^{α} rmsd value of the 1% of structures with lowest C^{α} rmsd. Lower y-values indicate that better structures were generated, whereas lower x-values indicate that better structures were also discriminated by the score function used. Arrows show the change between the CS-ROSETTA control (red cross) and the PCS-ROSETTA calculation (arrowhead).

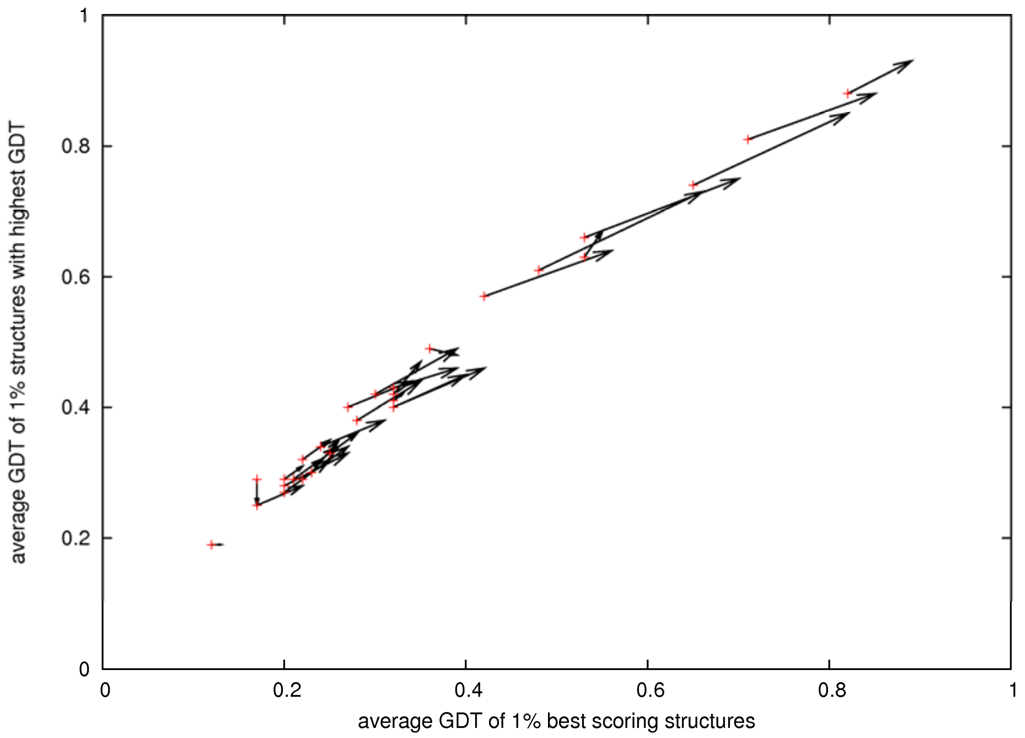


Figure S9. Consistent improvement of sampling and recognition of structures with higher GDT values. The x-axis of the plot shows the average GDT value of the best scoring 1% of the structures and the y-axis of the plot shows the average GDT value of the 1% of structures with highest GDT. Higher y-values indicate that better structures were generated, whereas higher x-values indicate that better structures were also discriminated by the score function used. Arrows show the change between the CS-ROSETTA control (red cross) and the PCS-ROSETTA calculation (arrowhead).

References

Jia, X. et al. (2011). 4,4'-Dithiobis-dipicolinic acid: a small and convenient lanthanide binding-tag for protein NMR spectroscopy. *Chemistry – A European Journal* **17,** 6830-6836.

Man, B. et al. (2010). 3-Mercapto-2,6-pyridinedicarboxylic acid: a small lanthanide-binding tag for protein studies by NMR spectroscopy *Chemistry – A European Journal* **16,** 3827-3832.

Shen, Y. et al. (2008). Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4685-4690.

Rosato, A. et al. (2009). CASD-NMR: critical assessment of automated structure determination by NMR. *Nature Methods* **6,** 625-626.

Su, X.-C. et al. (2008). A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy. *Journal of the American Chemical Society* **130**, 10486-10487.

Zemla, A. et al. (1999). Processing and analysis of CASP3 protein structure predictions. *Proteins* S3, 22–29.