Supplementary information: Systematic solution to homo-oligomeric structures determined by NMR

Jeffrey W. Martin Pei Zhou Bruce R. Donald

November 4, 2014

1 Mathematical structure of the operators

The folds and operators for DAGK have a mathematical structure that provides a systematic way to sample topologically distinct folds and also precisely models the symmetry inherent to many homo-oligomers. The roll and swap operators (R and S respectively) form the basis set of an Abelian group G where the group operation is binary composition of operators. G has order 36 and the presentation

$$\langle \mathtt{R}, \mathtt{S} | 6\mathtt{R} = 0, 6\mathtt{S} = 0 \rangle \,, \tag{1}$$

so the canonical form of G due to the structure theorem for finitely presented Abelian groups [1, 2] is $\mathbb{Z}_6 \oplus \mathbb{Z}_6$. Here, $\mathbb{Z}_p$ denotes the ring of integers modulo p, also written $\mathbb{Z}/p\mathbb{Z}$. There are 36 operators in the group, yet there are 48 distinct folds for DAGK due to the multiple possibilities for linker routes after the helices have been placed.

Interestingly, R and S each generate a cyclic sub-group of order 6 (i.e., $\mathbb{Z}_6$) which is decomposed into $\mathbb{Z}_3 \oplus \mathbb{Z}_2$. The order-3 torsion subgroup reflects the trimeric quaternary structure of DAGK, while the order-2 torsion subgroup reflects the two remaining positions (modulo symmetry) for the H3 helix after the H1 helix has been placed (i.e., $3 - 1 = 2$). Therefore, the factorization of G into torsion subgroups of prime order is

$$G \;\cong\; \mathbb{Z}_3 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}_2. \tag{2}$$

The group action of G on F, the set of folds, is the function

$$G \times F \;\rightarrow\; F \tag{3}$$

$$(g, f) \;\mapsto\; g \cdot f \tag{4}$$

where $g \in G$, $f \in F$, and $g \cdot f$ denotes operator g applied to fold f.

2 Refinement using Xplor-NIH

Xplor-NIH [3] v2.33 was used to perform the subunit and the trimeric structural refinements. We annealed subunit structures using 2000 steps of dynamics at 3000° K followed by a cooling phase where the temperature dropped to 25° K over 20,000 steps of additional dynamics. Then, models were minimized using 1000 steps of torsion angle minimization followed by 1000 steps of Cartesian minimization. Throughout the simulations, the models were restrained by the usual chemical potentials: bond, angl, impr, and the non-bonded atom repel potential. Also, the simulation used potentials for experimental intra-subunit restraints: NOEs, hydrogen bonds, dihedral restraints from TALOS, and RDCs. The potential used for the NOEs and the hydrogen bonds was the "hard" type NOE potential with the default averaging exponent of 6. The alignment tensor used for the RDC potential was restricted to have zero rhombicity, but its magnitude was allowed to vary. The orientation of the tensor was also fixed so its z axis aligned with the frame of the template structure, and hence the symmetry axis for the DAGK trimer. The C^{α} atoms of the subunit template structure were used to restrain corresponding C^{α} atoms in the models using an RMSD potential. Since these subunit structures were destined for further refinement in the trimeric state, only a single structure was calculated for each template. The single structures were found to have adequate satisfaction statistics and therefore computing large ensembles was deemed unnecessary.

For the subunit simulations, weights for the potentials were generally set to low values for the high-temp dynamics, and raised linearly during the cooling phase to final values, or simply held constant throughout the simulations. Table S1 shows the weights used for the subunit refinements.

Table S1: Weights for the subunit refinements using Xplor-NIH.

Potential	**High-temp weight**	**Low-temp weight**
repel (radii scale)	0.4	0.8
repel (weight)	0.2	1
bond	1	1
angl	0.4	1
impr	0.1	1
NOE	100	100
hbond	100	100
dihedral	500	500
RDC	0.01	1
template RMSD	1	10

Multiplicative ramps are often preferred in Xplor-NIH refinements. However, we observed during our refinements that potential weights were rising too quickly at low temperatures to allow the dynamics simulations to find relaxed conformations. Using a ramp with a slower rate of increase at low temperatures should have solved the problem, but the linear ramp gave poor results. Further investigation revealed numerical round-off errors in the implementation of linear ramps. After implementing a numerically-stable linear ramp, cooling phases using the new linear ramps gave greatly improved results.

For trimeric refinements, we used a slightly different approach than the one used for subunit refinements. Instead of annealing an extended chain once, we refined the "seed" structures many times to compute a traditional NMR ensemble. We refined each of the 26 trimeric seed structures 64 times. From the resulting ensemble for a chosen fold, we chose the single lowest-energy structure to represent the fold. We refined trimer structures using 4000 steps of dynamics at 3000° K followed by a cooling phase where the temperature dropped to 25 K over 40,000 steps of additional dynamics. Finally, models were minimized using 4000 steps of Cartesian minimization. The energy function was composed of the usual chemical potentials: bond, angl, impr, the non-bonded atom repel potential, and an additional RMSD potential between subunits to enforce the trimeric symmetry. The energy function also incorporated potentials for experimental restraints: NOEs, hydrogen bonds, dihedral restraints from TALOS, RDCs, disulfide bond restraints, and restraints from PRE. The potential for NOEs and hydrogen bonds was the "hard" type NOE potential with the default averaging exponent of 6. For the disulfide bond restraints and restraints from PRE, which are restraints with larger upper distances, we obtained better results by using an exponent of 12 with the "hard" type NOE potential. For the alignment tensor used in the RDC potential, the rhombicity was fixed at zero, but the magnitude and orientation were allowed to vary. Table S2 gives the weights for the potentials used during the trimeric refinement.

Table S2: Weights for the trimer refinements using Xplor-NIH.

Potential	**High-temp weight**	**Low-temp weight**
repel (radii scale)	0.6	0.8
repel (weight)	1	1
bond	0.6	1
angl	0.6	1
impr	0.6	1
symmetry	5	5
NOE	20	40
hbond	20	80
dihedral	200	400
RDC	0.1	1
disulfide bond	1	40
PRE	0.01	30

Empirically, we found that a weight on the order of hundreds was needed for the dihedral potential to have any noticeable effect. We also found that the weight for the PRE potential needed to be initialized with a very low value to prevent the simulation from becoming trapped in the many local energy minima defined by these very ambiguous restraints.

In a final step, the lowest-energy structure from each trimeric ensemble was subjected to an additional 4000 steps of Cartesian minimization. The conditions for this minimization were the same as for the previous low-temperature minimization with one notable exception. The van der Waals potential was used (with a weight of 1) instead of the non-bonded atom repel potential to drive molecular packing. We found that for structures already in a low-energy conformation, switching to the van der Waals potential gave even better results.

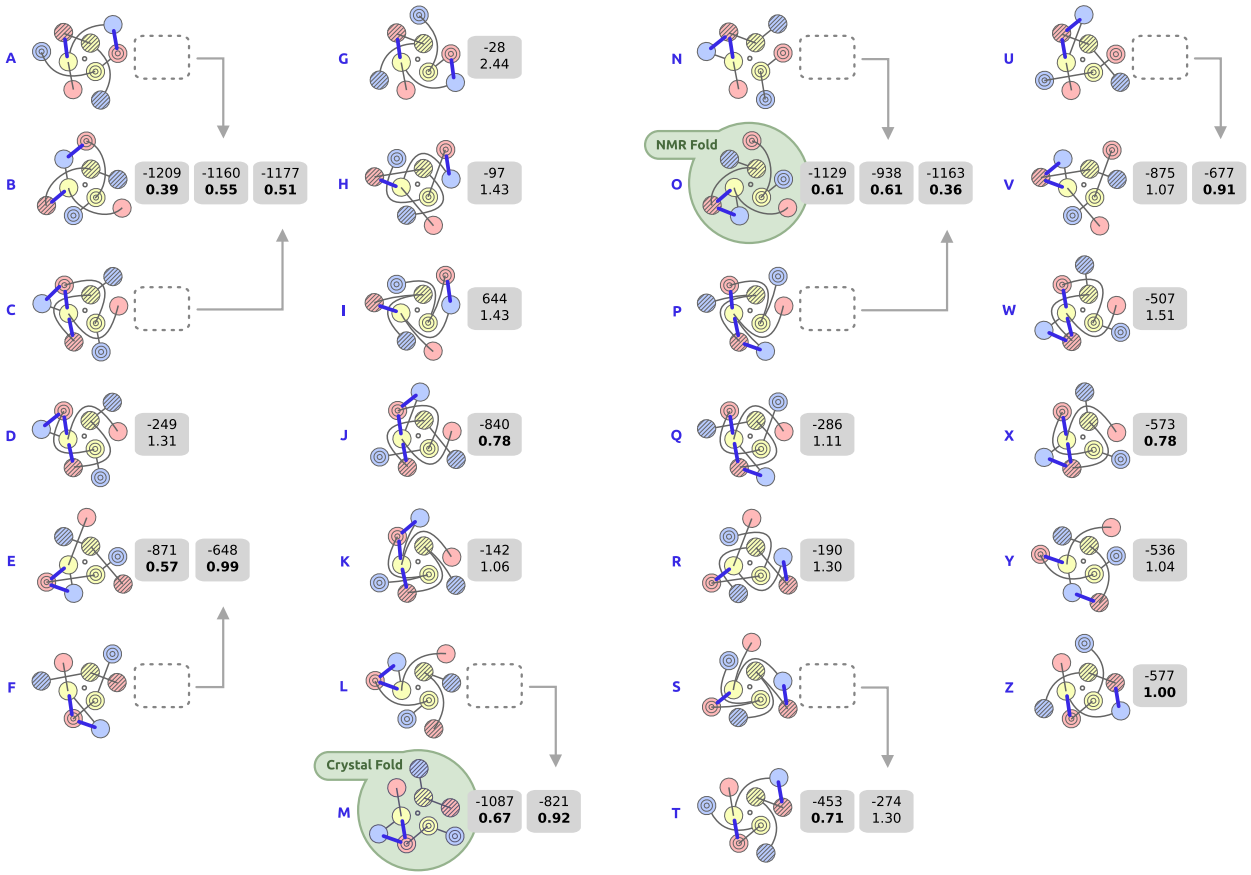

Figure S1: All 26 DAGK structures grouped by post-refinement fold. Folds are labeled with their names, A-Z, in blue and structures are represented by their statistics in grey boxes. The structure statistics are the same as in Figure 5 in the main text. RMS violation indices of 1 or less are highlighted in bold. For structures that switched folds during refinement, an empty box indicates the pre-refinement fold.

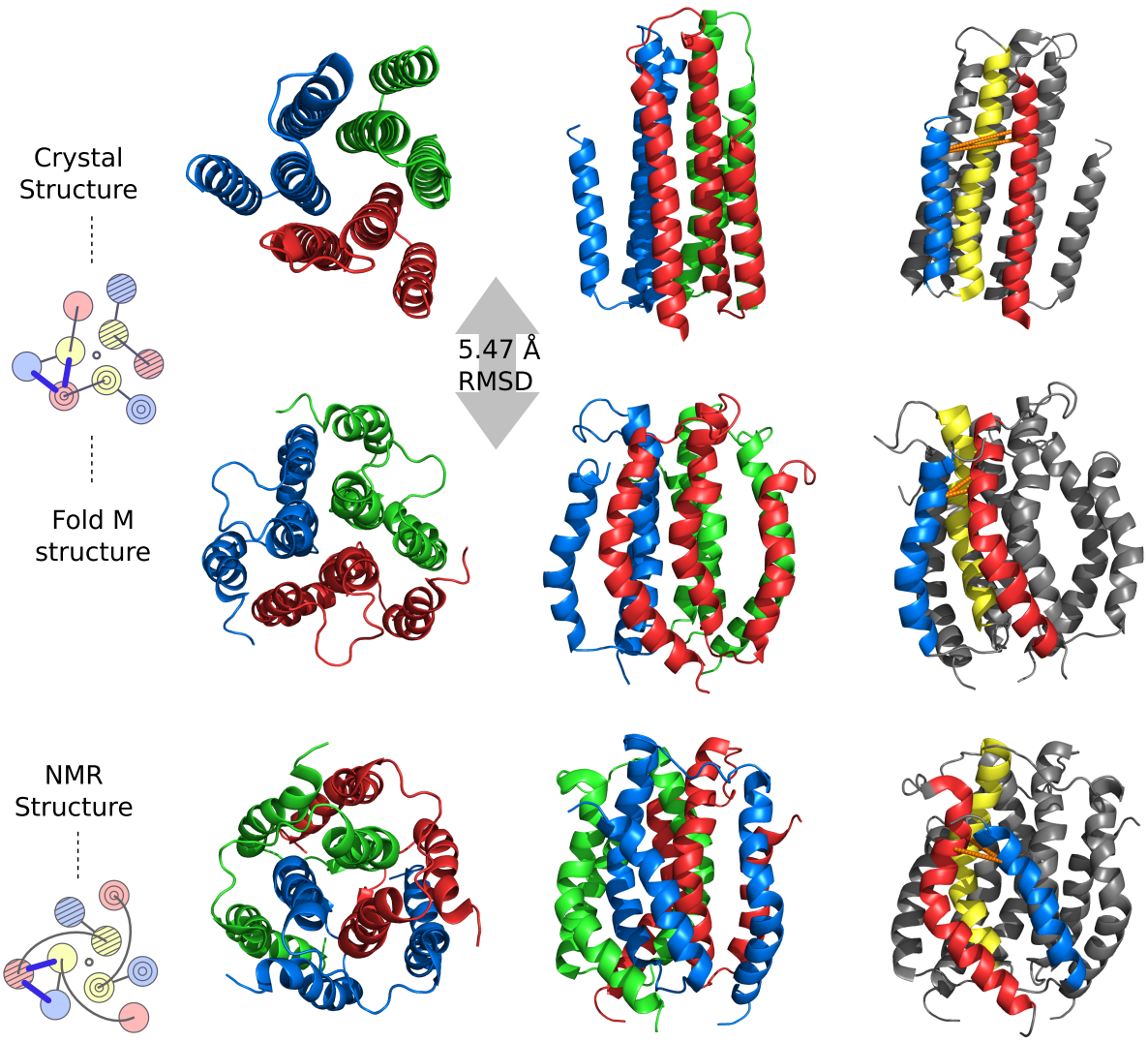

Figure S2: Left, Middle: The crystal structure (top, PDB: 3ZE4) and the structure of Fold M (middle) share the same fold, but have noticeable differences in helix shape and position. The curved helices in the NMR structure (bottom, PDB: 2KDC, model 1) and the Fold M structure stand in contrast to the straight helices in the crystal structure. Right: Coloring the helices with the same colors as the fold schematics (next to the structure names) shows how satisfying two disulfide bond restraints in particular ($33C^{\alpha}$–$96C^{\alpha}$ and $33C^{\alpha}$–$97C^{\alpha}$, shown using orange lines) causes the structure of Fold M to differ from the crystal structure. The depicted disulfide bond distance restraints are satisfied in the refined structure from fold M and the NMR structure, but in the crystal structure, they have violations of 7.4–7.6 Å and 9.7–10.2 Å respectively.

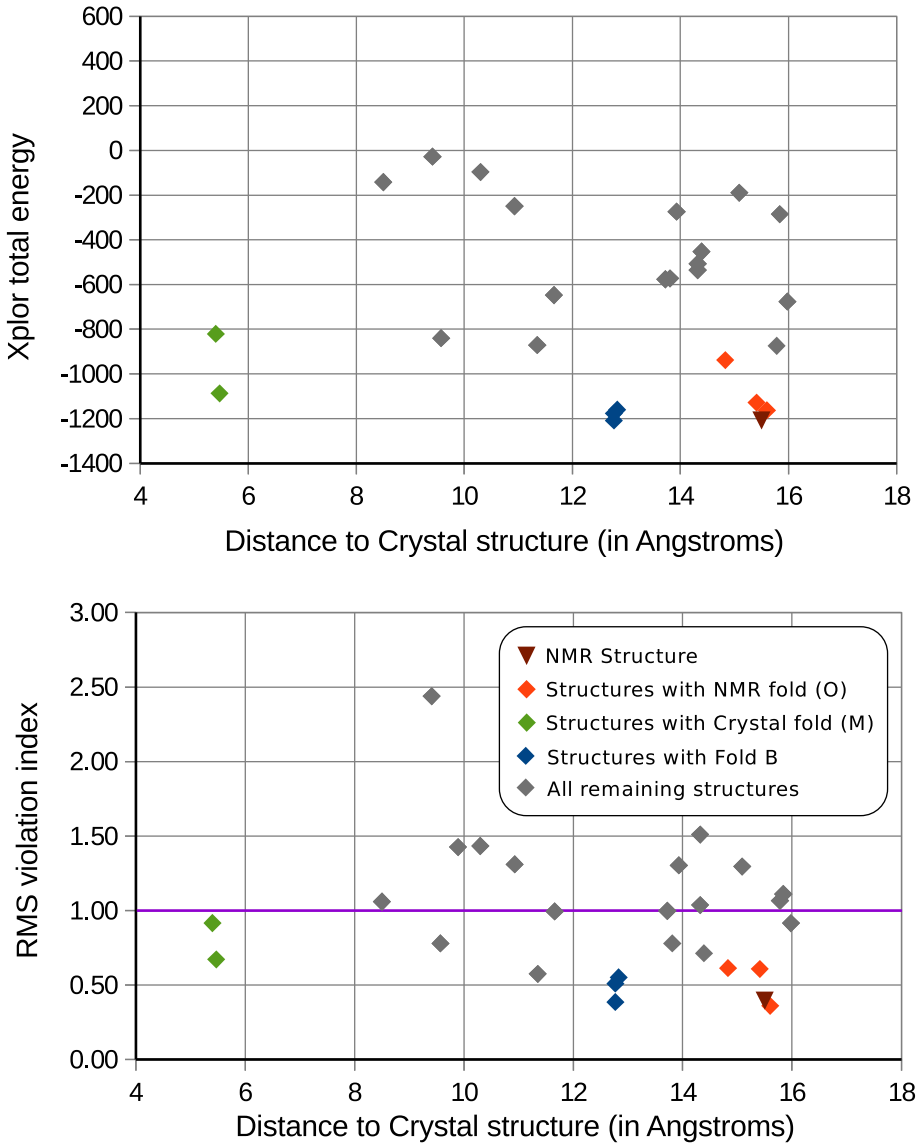

Figure S3: Xplor total energies (top) and RMS violation indices (bottom) for structures organized by distance to the crystal structure, PDB: 3ZE4. All structural distances (x-axis) are backbone atom (N,C$^{\alpha}$,C$'$) RMSD values in Å computed for the helical residues 30-48, 51-83, and 90-119 only. Variations in the loop regions were not considered in this score.

Table S3: Violation indices for all DAGK structures

Fold[1]	RMS	NOEs 0.5 Å	HBonds 0.5 Å	RDCs 1.0 Hz	Dihedrals 5°	DBonds 2.0 Å	PREs 2.0 Å	Xplor Energy	MolProbity Score[5]
P → O	0.36	0.40	0.16	0.33	0.41	0.28	0.5	-1163	3.45
B	0.39	0.35	0.14	0.43	0.34	0.23	0.63	-1209	3.37
2KDC[2]	0.40	0.35	0.12	0.70	0.45	0.24	0.23	-1209	3.12
C → B	0.51	0.53	0.18	0.63	0.49	0.26	0.73	-1177	3.56
A → B	0.55	0.40	0.13	1.08	0.36	0.24	0.54	-1160	3.46
E	0.57	0.47	0.20	0.94	0.76	0.30	0.41	-871	3.49
O	0.61	0.33	0.20	1.28	0.40	0.25	0.46	-1127	3.19
N → O	0.61	0.49	0.20	1.04	0.54	0.27	0.73	-938	3.60
M	0.67	0.56	0.17	1.21	0.77	0.27	0.49	-1087	3.44
T	0.71	0.64	0.27	1.33	0.61	0.40	0.52	-452	3.77
J	0.78	1.15	0.20	1.08	0.78	0.29	0.65	-840	3.62
X	0.78	0.71	0.25	1.42	0.71	0.27	0.70	-573	3.83
U → V	0.91	0.54	0.22	1.78	0.91	0.33	0.76	-677	3.75
L → M	0.92	0.60	0.16	1.70	1.08	0.28	0.72	-821	3.55
F → E	0.99	0.51	0.29	2.13	0.77	0.26	0.62	-648	3.77
Z	1.00	0.62	0.29	2.11	0.89	0.26	0.44	-577	3.80
Y	1.04	0.59	0.34	2.10	1.05	0.32	0.61	-536	3.66
K	1.06	0.44	0.37	2.32	0.80	0.27	0.51	-142	4.01
V	1.07	0.39	0.28	2.22	1.11	0.26	0.59	-875	3.43
Q	1.11	1.00	0.44	2.09	0.96	0.42	0.87	-286	3.93
R	1.30	0.85	0.31	2.85	0.69	0.36	0.73	-190	3.84
S → T	1.30	0.59	0.41	2.80	1.10	0.33	0.72	-274	3.68
D	1.31	0.71	0.30	2.89	0.88	0.32	0.65	-249	3.81
I	1.43	0.89	0.53	2.91	1.36	0.42	0.80	644	3.85
H	1.43	1.21	0.28	2.99	1.19	0.47	0.46	-99	4.13
W	1.51	0.51	0.32	3.29	1.31	0.24	0.81	-507	3.26
G	2.44	0.91	0.38	5.66	1.40	0.44	0.72	-28	3.83
3ZE4[3]	14.02	4.10	7.30	10.92	30.24	6.05	6.25	ND[4]	1.36

[1] $\alpha \rightarrow \beta$ indicates the structure switched from fold α to fold β during refinement. These statistics describe fold β in this case. The lack of an arrow means the structure remained in the same fold at the start and end of refinement. [2]The published NMR structure, PDB ID: 2KDC, model 1 after energy minimization in Xplor. [3]The published Crystal structure, PDB ID: 3ZE4 without any refinement in Xplor. [4]Not determined. Xplor energies were not determined for the crystal structure since refinement in Xplor was not performed. [5]Overall score from MolProbity [4]. Violation index values are all unitless. A value of 1 or less indicates the structure meets expectations for restraint satisfaction. Normalization constants for the violation indices are shown under each column heading. Violation indices are further described in the main text.

Table S4: Restraint satisfaction for all DAGK structures

Fold[1]	NOEs[2]		HBonds[3]		RDCs[4]		Dihedrals[5]		DBonds[6]		PREs[7]	
	Num	RMS	Num	RMS	RMSD	Q	Num	RMS	Num	RMS	Num	RMS
P → O	0	0.03	0	0.01	0.49	0.06	0	0.3	0	0.18	0	0.11
B	0	0.03	0	0.01	0.50	0.06	0	0.3	0	0.16	6	0.19
2KDC[8]	0	0.03	0	0.01	0.51	0.06	0	0.4	0	0.16	0	0.04
C → B	0	0.03	0	0.01	0.61	0.08	0	0.4	0	0.15	0	0.21
A → B	0	0.03	0	0.01	0.64	0.08	0	0.4	0	0.18	3	0.13
E	0	0.04	0	0.02	0.79	0.10	15	0.7	0	0.21	0	0.11
O	0	0.03	0	0.02	0.53	0.07	0	0.4	0	0.19	0	0.12
N → O	0	0.03	0	0.02	0.68	0.08	3	0.4	0	0.20	9	0.20
M	3	0.03	0	0.02	0.64	0.08	3	0.5	0	0.17	0	0.10
T	6	0.04	0	0.02	0.88	0.11	10	0.6	0	0.25	3	0.15
J	2	0.05	0	0.02	0.82	0.10	0	0.6	0	0.19	0	0.15
X	0	0.04	0	0.02	0.99	0.12	0	0.7	0	0.19	0	0.21
U → V	3	0.03	0	0.02	0.83	0.10	9	0.7	0	0.21	9	0.21
L → M	3	0.03	0	0.02	0.76	0.09	14	0.8	0	0.19	6	0.17
F → E	3	0.04	0	0.02	0.88	0.11	12	0.7	0	0.19	6	0.16
Z	4	0.03	0	0.02	1.03	0.13	12	0.7	0	0.18	0	0.13
Y	9	0.04	0	0.02	1.05	0.13	18	0.8	0	0.24	3	0.15
K	0	0.03	0	0.03	1.35	0.17	0	0.9	0	0.22	0	0.17
V	0	0.02	0	0.02	0.85	0.11	6	0.7	0	0.18	3	0.13
Q	0	0.05	0	0.03	1.02	0.13	0	0.8	0	0.24	0	0.21
R	5	0.05	0	0.03	1.02	0.13	12	0.8	0	0.28	3	0.16
S → T	2	0.04	0	0.02	1.15	0.14	19	0.9	0	0.19	5	0.14
D	0	0.05	0	0.03	1.17	0.14	0	0.9	0	0.24	0	0.17
I	3	0.04	6	0.06	1.32	0.16	27	1.1	0	0.28	15	0.23
H	12	0.06	0	0.03	1.20	0.15	26	0.9	0	0.29	0	0.13
W	0	0.03	0	0.02	1.37	0.17	1	0.7	0	0.17	0	0.16
G	9	0.05	0	0.03	1.25	0.16	19	1.0	0	0.30	12	0.21
3ZE4[9]	19	0.17	6	0.27	3.46	0.42	92	17.1	24	4.55	97	2.13

[1]$\alpha \rightarrow \beta$ indicates the structure switched from fold α to fold β during refinement. These statistics describe fold β in this case. The lack of an arrow means the structure remained in the same fold at the start and end of refinement. [2]NOEs: number of violations of at least 0.25 Å, RMS violation shown in Å. [3]Hydrogen bonds: number of violations of at least 0.25 Å, RMS violation shown in Å. [4]RDCs: RMSD units are Hz, Q-factor values are unitless. [5]Dihedrals angles: number of violations of at least 2.5°, RMS violation shown in °. [6]Disulfide bonds: number of violations of at least 1 Å, RMS violation shown in Å. [7]PREs: number of violations of at least 1 Å, RMS violation shown in Å. [8]The published NMR structure, PDB ID: 2KDC, model 1 after energy minimization in Xplor. [9]The published Crystal structure, PDB ID: 3ZE4 without any refinement in Xplor. In general, the violation threshold for the number of violations shown here is half the violation threshold used for the violation index.

References

[1] Hungerford, T. (1980) *Algebra.* (Springer, New York).

[2] Donald, B. R. (2011) *Algorithms in Structural Molecular Biology.* (MIT Press, Cambridge, MA). 464 pages.

[3] Schwieters, C. D, Kuszewski, J. J, Tjandra, N, & Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **160**, 65–73.

[4] Chen, V. B, Arendall, III, W. B, Headd, J. J, Keedy, D. A, Immormino, R. M, Kapral, G. J, Murray, L. W, Richardson, J. S, & Richardson, D. C. (2010) *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**, 12–21.