SUPPORTING INFORMATION FOR:

A hybrid NMR/SAXS-based approach for discriminating oligomeric protein interfaces using Rosetta

Paolo Rossi[†], Lei Shi[⊥], Gaohua Liu[†], Christopher M. Barbieri[†], Hsiau-Wei Lee[|], Thomas D. Grant[§$], Joseph R. Luft[§$], Rong Xiao[†], Thomas B. Acton[†], Edward H. Snell[§$], Gaetano T. Montelione[†#], David Baker[⊥,≠], Oliver F. Lange[⊥], Nikolaos G. Sgourakis[⊥,&]

[†]Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, and Northeast Structural Genomics Consortium, Rutgers University, Piscataway, NJ 08854, U.S.A.

[⊥]Department of Biochemistry, University of Washington, Seattle, WA 98195, U.S.A.

[|]Departments of Chemistry and Biochemistry & Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, GA, and Northeast Structural Genomics Consortium 30602, U.S.A.

[§]Hauptman–Woodward Medical Research Institute, and Northeast Structural Genomics Consortium Buffalo, NY 14203, U.S.A.

[$]SUNY Buffalo Department of Structural Biology, Buffalo, NY 14203, U.S.A.

[#]Department of Biochemistry, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, U.S.A.

[≠]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, U.S.A

Present addresses: [&]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Table of Contents

Supplementary Methods

Sample preparation

The Aha1 protein from *Colwellia psychrerythraea* including a C-terminal His_6 tag (LEHHHHHH), was cloned, expressed, and purified following standard protocols of the Northeast Structural Genomics Consortium (NESG) in order to prepare [U-^{13}C,^{15}N]- and [U-5%-^{13}C,100%-^{15}N]-Aha1 samples for NMR spectroscopy[1]. Briefly, the 146-residue coding sequence of the gene locus CPS_1688 of *Colwellia psychrerythraea* (UniProtKB/TrEMBL ID, Q484T9_COLP3; NESG ID, CsR251; hereafter referred to as Aha1) was cloned into pET21_NESG vector containing a C-terminal His_6 affinity tag (LEHHHHHH) to yield the plasmid CsR251-21.1 (deposited into the PSI Materials Repository; http://psimr.asu.edu/). The triple-labeled with the methyl groups of Val, Leu, Ile (δ1) selectively protonated, {[U-^{2}H,^{13}C,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Val-[$^{13}CH_3$]}-Aha1 were prepared using $(^{15}NH_4)_2SO_4$ and [U-^{13}C]-D-glucose with addition of [U-$^{13}C_4$, 3,3-$^{2}H_2$]-α-ketobutyrate (50 mg/L), [U-$^{13}C_5$, 3-^{2}H]-α-ketoisovalerate (CIL Inc.) (100mg/L) in D_2O medium[2]. The CsR251-21.1 plasmid was transformed into codon-enhanced BL21(DE3) pMGK *Escherichia coli* cells, and cultured in MJ9 minimal medium[3] containing $(^{15}NH_4)_2SO_4$ and U-^{13}C-glucose as the sole nitrogen and carbon sources. Natural abundance Aha1 was produced in LB media following the identical protocol. Initial cell growth was carried out at 37 °C and protein expression was induced at 17 °C by 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at mid-log phase growth. Expressed proteins were purified using an ÄKTAxpress™ (GE Healthcare) two-step protocol consisting of HisTrap HP affinity chromatography followed directly by HiLoad 26/60 Superdex 75 gel filtration chromatography. The final yield of purified isotopically enriched Aha1 was approximately 20 mg/L of culture. Samples of [U-^{13}C,^{15}N]-, {[U-^{2}H,^{13}C,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Val-[$^{13}CH_3$]}-, and [U-5%-^{13}C,100%-^{15}N]-Aha1 for NMR spectroscopy were concentrated by centrifugation to 1.3, 0.7 and 1.1 mM, respectively, in 10 mM TRIS-HCl, 100 mM NaCl, 5 mM DTT, 50 μM DSS, 10% $^{2}H_2O$ at pH 7.5. Sample purity and molecular mass were confirmed by SDS-PAGE and MALDI-TOF mass spectrometry.

[U-5%-^{13}C,100%-^{15}N]-Aha1 was first aligned in 12.5 mg/mL Pf1 phage medium (ASLA biotech[4]). For the second set of residual dipolar couplings, the protein was aligned in 4.2 % $C_{12}E_5$ polyethylene glycol bicelles (PEG, Sigma Aldrich) using previously published protocols[5].

NMR spectroscopy and resonance assignments

All NMR data were collected at 25 °C on Bruker AVANCE 800 MHz and Varian INOVA 600 MHz spectrometers equipped with 5-mm cryoprobes, processed with NMRPipe[6] and visualized using SPARKY[7];

chemical shift referencing is conducted with 50 μM DSS internal standard and the assigned Cα, Cβ and C' CS were adjusted for ^{2}H shift using TALOS+[8]. First, backbone and selectively protonated methyls were assigned using a {[U-^{2}H,^{13}C,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Val-[$^{13}CH_3$]}-Aha1 sample and the data acquisition and processing strategies follow the previously published work by Lange, Rossi and co-workers[9]. Additional assignments were determined manually starting from the existing triple labeled assignments using two complementary 3D HCCH-TOCSY experiments, with the chemical shift of ^{1}H and ^{13}C indirectly recorded respectively, and 3D ^{13}C- and ^{15}N-edited NOESY spectra acquired on a [U-^{13}C,^{15}N]-Aha1 with initially simulated peak lists following a previously published strategy of Liu and coworkers[10]. The complete initial NOESY peak list were simulated from the existing triple labeled assignments for assigned atoms and averaged chemical shifts from the BioMagResBank (BMRB) for all remaining assignable side-chain resonances. Stereospecific isopropyl methyl assignments for all Val and Leu residues were deduced from characteristic cross-peak fine structures in high resolution 2D ^{1}H-^{13}C HSQC spectra of [U-5%-^{13}C,100%-^{15}N]-Aha1[11]. Complete datasets that include ^{1}H, ^{13}C, and ^{15}N resonance assignments, raw FIDs, and peak lists for *C. psychrerythraea* Aha1 were deposited in BMRB (BMRB ID 19235). Dynamics and oligomerization states of Aha1 were assessed by gradient and sensitivity-enhanced 2D {^{1}H} -^{15}N heteronuclear NOE and 1D ^{15}N T_1 and T_2 relaxation experiments[12]. 3D ^{13}C- and ^{15}N-edited NOESY spectra on [U-^{13}C,^{15}N]-Aha1 sample were acquired with a mixing time of 120 ms. A 300 ms mixing time was chosen for NOESY acquisitions on {[U-^{2}H,^{13}C,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Val-[$^{13}CH_3$]}-Aha1 sample.

Automatic NOESY cross-peak assignment and initial structure calculations of monomeric Aha1 units were performed with an automated CYANA 3.0 run[13] as previously described by Lange, Rossi and co-workers[9] using 23 manually determined restraints and carrying two sets of peak lists and chemical shifts assignments from double and triple labeled Aha1 samples. The initial run yielded 3463 upper limit distance restraints per protomer. The initial NOE set was trimmed back to the 1493 long-range ($|i - j| \geq 5$) restraints set and used for the RASREC-Rosetta run. No experimental intermolecular contact information was present prior to Rosetta structure calculations.

For further structure validation, intermolecular protein-protein NOE contacts were obtained using a double half-filtered (3D F1-^{13}C/^{15}N-filtered, F3-^{13}C-edited) NOESY experiment[14] on a 1:1 mixed sample of unlabeled and [U-^{13}C,^{15}N]-Aha1 acquired with 120 ms mixing time (Table S3). Assignment of the 3D double half-filtered NOESY was straightforward from the existing sidechain assignments that were highly resolved in the ^{1}H-^{13}C HSQC for the majority of the interchain peaks. All spectra were acquired on the same spectrometer to insure consistent peak positions. Amide backbone one bond ^{1}H-^{15}N residual dipolar coupling (RDC) values ($^{1}D_{NH}$) for Aha1 were measured using $^{1}J_{NH}$-modulated HSQC experiments[15] on isotropic and two partially aligned samples of [U-5%-^{13}C,100%-^{15}N]-Aha1 in Pf1 phage (12.5 mg/mL) and polyethylene glycol bicelles (4.2%) as previously described[4,5]. A high degree of linear dependence was

obtained for the two RDC datasets, manifested in a correlation coefficient of 0.8 in 5-dimensional tensor space[16]. Dihedral angle predictions were computed using TALOS+ [8].

Rosetta structure calculation and analysis

The ensemble of Aha1 monomeric structures were modeled using the RASREC CS-Rosetta method[17]. In summary, CYANA 3.0 upper distance restraints were first separated into the restraints with highest reliability (SUP = 1), and those with lower reliability (SUP < 1) and then converted into ROSETTA flat-bottom restraints using an exponential penalty function with a variable upper limit[18]. RASREC combines the lower reliability distance restraints into random pairs of ambiguous restraints at the start of each individual structure calculation trajectory. All RASREC stages were terminated as soon as the acceptance rate for new conformers into the structural pool drops below 10%, as described previously[9].

We applied a penalty term that is proportional to the root-mean square deviation between experimental and calculated RDCs, during Rosetta scoring and gradient-based minimization. For a given structural model and input RDC data, the five alignment tensor parameters were fitted using the non-linear Levenberg-Marquardt algorithm. The three Euler angles that define the orientation of the alignment tensor in the molecular frame were optimized while keeping the axial (D_a) and Rhombic component (R) of the alignment tensor fixed. The D_a and R-values used for peg were 7.4 Hz, 0.33 and for phage 12.7 Hz, 0.18, respectively. These values were estimated from the powder pattern distribution of the RDC data[19]. The contribution of each RDC dataset was weighted according to the inverse of the magnitude of the alignment tensor, D_a.

The dimeric structure of Aha1 was modeled using the RosettaOligomers[20] protocol that involves Monte-Carlo-based search of the rigid body degrees of freedom using a coarse-grained representation of the system, followed by symmetric, all-atom optimization of the rigid-body, backbone and sidechain degrees of freedom. Briefly, starting from an ensemble of 10 monomeric structures provided by RASREC, we generated 40,000 full-atom conformations using a symmetric docking protocol that samples the 4 rigid-body degrees of freedom that uniquely define a dimer orientation in the molecular frame globally using a low-resolution docking stage followed by symmetric all-atom refinement (phase I). In a first step, we selected the phase I dimer models having the lowest 10% Rosetta energy and 20% interface interaction free energy ($\Delta\Delta G$ or DDG) (800 models). From this set, we further selected 28 models showing good fits (below the average score values in the full set) to both the RDC and SAXS data that were used as starting structures for additional localized docking calculations, that involve random perturbations of the four rigid-body degrees of freedom, using Gaussian displacements and rotations between the two subunits with a standard deviation of 3 Å and 5^0, respectively (phase II). From these runs, conformations having below

75% Rosetta all-atom energy, RDC and SAXS penalty and with solvent exposed surface area (SASA) greater than 800 $Å^2$ were kept for further analysis. Finally, we selected with the top 10 structures according to DDG as the final ensemble. The model with the lowest interface interaction score was used as a reference structure to compute RMS (excluding the flexible C-terminal loops at residues 132-148). A high degree of convergence to the reference structure was obtained, suggesting that the restrained search of conformational space leads to a global minimum in Rosetta's energy landscape (Fig. S8 and Fig. 5). All generated dimer models were further rescored according to an independent dataset of 23 inter-subunit NOEs (Table S3) not applied as calculation restraints, using a flat-bottom potential with a 5.5 Å upper limit and an exponential penalty function. The final 10 Rosetta structures are highly consistent with the intermolecular NOEs as well as all other available NMR and SAXS data (Table S2).

We used a reduced SAXS dataset, consisting of 10% of the original points (this was done by only including every 10th data point from the low-noise region of the experimental profile up to Q values of 0.35 $Å^{-1}$) during Rosetta structure calculations. This did not affect model discrimination, but led to a much faster computation of the SAXS score, performed using a coarse-grained representation of the system as described previously. The CRYSOL fits to the experimental SAXS profile in Figures 3, 5, and S8 were carried out using the full experimental dataset.

Additionally, to evaluate the agreement of the dimer orientations observed in the available Xray structures of the bet-V1 superfamily with the NMR and SAXS data, we performed local perturbation calculations (same as phase II above), where conformational sampling is carried out in the vicinity of the rigid-body degrees of freedom extracted from the homologous dimeric structures SSP2350 (PDB ID 3Q6A), MM0500 (PDB ID 1XUV), and MLL2253 (PDB ID 3Q63). The models generated from these calculations were found to be largely inconsistent with the RDC, SAXS and intermolecular NOE data.

Structural statistics and global structure quality scores for Aha1, presented in Table S2, were computed using the PSVS 1.4 software package[21]. The RDC statistics were computed using PALES[22]. The final coordinates (excluding the C-terminal 6-His polypeptide segment) for the ensemble of 10 structures and NMR-derived restraints for Aha1 were deposited to the Protein Data Bank (PDB ID 2M89). All structure diagrams were made using PyMOL ver. 1.4 (www.pymol.org).

Small angle X-ray Scattering (SAXS)

SAXS data were collected on Beamline 4-2 of the Stanford Synchrotron Radiation Lightsource (SSRL)[23]. The data were integrated with SSRL and in-house software Sastools[23] and analyzed using the Primus package[24]. Three protein concentrations were used that were ×0.5, ×0.375 and ×0.25 the starting 1.4

mM Aha1 (CsR251) stock solution with eight 1s exposures of each. There was no evidence of radiation damage within each concentration series and no aggregation or concentration dependent effects seen. Twenty independent *ab initio* molecular envelope reconstructions were carried out yielding a mean normalized spatial discrepancy of 0.540 with a variation of 0.023 (Fig. 2). CRYSOL[25] was used to examine the fit to the lowest energy Rosetta structure, a dimer, which gave a χ^2 of 1.71.

Analytical ultracentrifugation

Analytical ultracentrifugation experiments were carried out in an Optima XL-I ultracentrifuge using an An-50 Ti rotor (Beckman Coulter, Fullerton, CA). Aha1 was equilibrated in a pH 7.5 buffer containing 10 mM Tris, 100 mM NaCl and 2 mM TCEP. Buffer density, buffer viscosity and the partial specific volumes of the proteins were calculated using the program SEDNTERP[26]. All analytical ultracentrifugation experiments were conducted in epon charcoal-filled double-sector centerpieces and quartz windows. Sedimentation velocity (SV) experiments were conducted at 25 ºC and the samples were spun at 50,000 rpm using seven different protein concentrations (40, 20, 10, 5, 3, 1.5, and 0.75 μM). Protein gradients were monitored using either the interference optical system (40 μM Aha1) and at either 280 nm (20 and 10 μM Aha1) or 230 nm (5, 3, 1.5, and 0.75 μM Aha1) using the optical absorbance system. Approximately 50 SV scans for each protein concentration were fit simultaneously to calculate a *c*(s) distribution that was generated using the program SEDPHAT[27]. During the *c*(s) analysis, the frictional ratio and meniscus position were treated as floating parameters. After optimization of these parameters, the final distribution was calculated using a resolution setting of 200, in a range from 0-10 Svedberg (resolution 0.05 S, 10/200) a confidence interval of 0.8. For the sedimentation equilibrium (SE) experiments, 4 different protein concentrations (5.0, 2.5, 1.5, and 0.25 μM) at three different speeds (15,000, 23,000 and 29,000 rpm) were analyzed at 25 °C, with protein gradients being monitored at both 280 and 230 nm. Profiles of Aha1 were globally fit to a monomer-dimer self-association model using the program SEDPHAT (v 4.0)[27] with the monomer molecular weight fixed at 17550 Da. Error limits, which represent the 95% confidence interval, were determined using F-statistics[28].

Software use (command lines)

These command lines are compatible with ROSETTA3 SVN version 51540 https://svn.rosettacommons.org/source/trunk/rosetta/rosetta_source. A full description of the use of the program can be found at www.csrosetta.org

Step 1: Generate monomer models from ILV NOE data using RASREC CS-Rosetta protocol.

Please refer to the supporting information in Lange, Rossi et. al.[9] for the exact command lines used to generate models of the monomeric subunit.

Step 2: Global symmetric docking (Phase I)

```
minirosetta.static.linuxgccrelease \
      -run:protocol symdock \
      -database rosetta_database \
      -symmetry:symmetry_definition C2.symm \
      -symmetry:initialize_rigid_body_dofs \
      -packing:ex1 \
      -packing:ex2aro \
      -use_input_sc \
      -ignore_unrecognized_res \
      -out:nstruct 200 \
      -out:file:silent phaseI.silent \
      -out:file:silent_struct_type binary \
      -out:file:fullatom \
      -use_incorrect_hbond_deriv false \
      -docking:dock_lowres_filter 15.0 20.0 1500.0 \
      -docking:high_min_patch patch_high_min \
      -score:weights score12_full \
       -restore_pre_talaris_2013_behavior \
     -in:file:s monomer_input_rasrec.pdb
```

Step 3: Local symmetric perturbation with rdcs and saxs

This step starts from selected monomer models from step 2 based on their RDC and SAXS scores (generating sel_monomer_from_global_docking.pdb)

```
minirosetta.static.linuxgccrelease \
        -run:protocol symdock \
        -database rosetta_database \
        -symmetry:symmetry_definition C2.sym \
        -symmetry:perturb_rigid_body_dofs 3 5 \
        -packing:ex1 \
        -packing:ex2aro \
        -use_input_sc \
        -ignore_unrecognized_res \
        -out:nstruct $1 \
        -out:file:silent phaseII.silent \
        -out:file:silent_struct_type binary \
        -out:file:fullatom \
        -use_incorrect_hbond_deriv false \
        -docking:low_patch patch_rdc_saxs \
        -docking:high_patch patch_rdc_saxs \
        -docking:high_min_patch patch_high_min_rdc_saxs \
        -docking:pack_patch patch_rdc_saxs \
        -docking:dock_lowres_filter 15.0 20.0 1500.0 \
        -docking:kick_relax \
        -default_max_cycles 200 \
        -relax:default_repeats 2 \
        -jump_move true \
        -score:weights score12_full \
         -restore_pre_talaris_2013_behavior \
        -score::patch patch_relax_rdc_saxs \
        -rdc:fit_method nls \
        -in::file::rdc rdc_ medium1.txt rdc_ medium2.txt  \
        -rdc::fix_normAzz 0.001 0.001 \
        -residues:patch_selectors CENTROID_HA \
        -score:saxs:ref_spectrum saxs_sparse.dat \
        -in:file:s sel_monomer_from_phaseI.pdb
```

Inspection of the DDG, Rosetta energy and SAXS and RDC score terms in the phase I/phase II output models can be performed by extracting the score lines from the silent output file: eg. grep SCORE: phaseI.out > scores.txt

Patch files

Patch files are used to set weights of the score function. Four different patch

files are used to control sampling (low resolution, all atom) and selection into the pool of structures.

==> patch_high_min<==

fa_rep *= 4.22

==> patch_rdc_saxs <==

rdc = 20.0
fastsaxs = 1.0

==> patch_high_min_rdc_saxs <==

rdc = 20.0
fastsaxs = 1.0
fa_rep *= 4.22

==> patch_relax_rdc_saxs <==

rdc = 10.0
fastsaxs = 0.05

Symmetry definition files

==> C2.symm <==

```
symmetry_name paolo_dimer__2
E = 2*VRT0_base + 1*(VRT0_base:VRT1_base)
anchor_residue 93
virtual_coordinates_start
xyz VRT0  -0.9971370,0.0568482,0.0498606  -0.0569190,-0.9983788,0.0000000  14.2107039,-4.3937104,-0.3663004
xyz VRT0_base  -0.9971370,0.0568482,0.0498606  -0.0569190,-0.9983788,0.0000000  26.1198784,-5.0726689,-0.9618041
xyz VRT1  0.9971370,-0.0568482,-0.0498606  -0.0756160,-0.7488722,-0.6583864  14.2107039,-4.3937104,-0.3663004
xyz VRT1_base  0.9971370,-0.0568482,-0.0498606  -0.0756160,-0.7488722,-0.6583864  2.3015295,-3.7147518,0.2292032
xyz VRT  1.0000000,0.0000000,0.0000000  0.0000000,1.0000000,0.0000000  15.2107039,-4.3937104,-0.3663004
virtual_coordinates_stop
connect_virtual JUMP0_to_com VRT0 VRT0_base
connect_virtual JUMP0_to_subunit VRT0_base SUBUNIT
connect_virtual JUMP1_to_com VRT1 VRT1_base
connect_virtual JUMP1_to_subunit VRT1_base SUBUNIT
connect_virtual JUMP0 VRT VRT0
connect_virtual JUMP1 VRT0 VRT1
set_dof JUMP0_to_com x(11.9433682565431) angle_x
set_dof JUMP0_to_subunit angle_x(0:360) angle_y(0:360) angle_z(0:360)
set_jump_group JUMPGROUP2 JUMP0_to_com JUMP1_to_com
set_jump_group JUMPGROUP3 JUMP1_to_subunit JUMP0_to_subunit
```

Data files

For each input data, we only print the first 10 lines each to familiarize readers with the file-format.

==> rdc_medium1.txt <==

```
3 H 3 N 9.279
151 H 151 N 9.279
7 H 7 N 0.396
155 H 155 N 0.396
15 H 15 N 8.252
163 H 163 N 8.252
16 H 16 N 15.043
164 H 164 N 15.043
17 H 17 N 12.097
165 H 165 N 12.097
```

==> rdc_medium2.txt <==

```
3 H 3 N 10.360
151 H 151 N 10.360
7 H 7 N -16.609
155 H 155 N -16.609
15 H 15 N 18.553
163 H 163 N 18.553
16 H 16 N 24.677
164 H 164 N 24.677
17 H 17 N 23.681
165 H 165 N 23.681
```

==> saxs_sparse.txt <==

```
0.00771096 7554.24 70.6635
0.017006 7253.15 9.33698
0.0263011 6830.58 7.28595
0.0355961 6285.76 6.17379
0.0448912 5670.72 5.27
0.0541862 4985.43 4.56433
0.0634813 4285.38 3.83601
0.0727763 3587.02 3.27892
0.0820714 2926.51 2.97787
```

Supplementary Tables

Table S1. Aha1 SAXS data collection summary.

Data Collection parameters	
Instrument	SSRL beamline 4-2
Beam geometry	Slit
Wavelength	1.13
q range	0.008-0.531
Exposure time (s)	1 (x8) for 3 concs
Concentration range (mM)	0.25-0.50 × 1.4
Temperature (°C)	18
Structural parameters	
R_g from P(r)	20.2
R_g from Guinier	20.5
D_{max}	62.65
Porod volume estimate	56,560
Dry volume estimated from sequence	21,231
Molecular-mass determination	
Molecular mass Mr from Porod (kDa)	35.35
Calculated monomeric Mr from sequence (kDa)	17.55
Software employed	
Primary data reduction	Sastool
Data processing	Primus
Ab initio analysis	Dammif
Validation and averaging	Damaver
Computation of model intensities	Crysol

Table S2. Summary of NMR Structural Statistics for Aha1 ensemble.[a]

Completeness of resonance assignments [b]	**Aha1**
Backbone (%)	92
Side chain (%)	77
Aromatic (%)	48
Stereospecific methyl (%)	100
Conformationally-restricting restraints [c]	
NOE restraints	
Total	1483
Intra-residue ($i = j$)	0
Sequential ($\|i - j\| = 1$)	0
Medium range ($1 < \|i - j\| < 5$)	0
Long range ($\|i - j\| \geq 5$)	1483
NOE restraints / residue	10
Interchain protein/protein NOEs [d]	23
NH RDC restraints (PEG/Phage)	78/91
Number of restraints / residue (total / long range)	1662/1483
Residual constraint violations [c]	
Average distance restraint violations / structure	
0.1 - 0.2 Å	20.0
0.2 - 0.5 Å	5.1
> 0.5 Å	0.0
Average RMS of distance violation / restraint (Å)	0.14
Maximum distance violation (Å)	0.49
Model Quality [c]	
RMSD from average coordinates (Å)	
All Backbone atoms (ordered/all)	0.6/0.6
All Heavy atoms (ordered/all)	0.7/0.7
RMSD Bond lengths (Å)	0.018
RMSD Bond angles (°)	1.3
Molprobity Ramachandran plot [e]	
Most favored regions (%)	96.1
Additionally allowed regions (%)	3.5
Disallowed regions (%)	0.3
Global quality scores (Raw / Z-score) [c]	
Verify3D	0.47/0.16
ProsaII	0.55/-0.22
Procheck G-factor (ϕ,ψ) [e]	-0.23/-0.59
Procheck G-factor (all dihedrals) [e]	0.08/0.47
MolProbity clashscore	1.48/1.27
Residual Dipolar Couplings (RDC) Scores [f]	
Q-factor (PEG/Phage)	0.23/0.13
R (PEG/Phage)	0.16/0.09
Small Angle X-ray scattering (SAXS) Validation [g]	
χ^2	1.71

[a] Structural statistics were computed for the ensemble of 10 deposited structures (PDB ID 2M89)

[b] Computed for residues 1-148, using the AVS software[29]. Resonances that were not included were exchangeable protons (N-terminal NH_3^+, Lys NH_3^+, Arg NH_2, Cys SH, Ser/Thr/Tyr OH) and Pro N, C-terminal carbonyl, side chain carbonyl and non-protonated aromatic carbons.

[c] Calculated for protein using the PSVS 1.4 program[21]. Average distance constraints were calculated using the sum of r^{-6}.

[d] Not used during Rosetta structure calculation. Used only for validation of the models.

[e] Ordered residue ranges [$S(\phi) + S(\psi) > 1.8$] : 1-134 (chain A), 1-134 (chain B). Secondary structure elements: 2-10 (β1, β1′), 34-36 (β2, β2′), 44-48 (β3, β3′), 55-62 (β4, β4′), 66-72 (β5, β5′), 83-90 (β6, β6′), 95-104 (β7, β7′), 14-22 (α1, α1′), 24-27 (α2, α2′), 109-130 (α3, α3′).

[f] RDC quality scores[30] averaged over 10 lowest energy Aha1 Rosetta structures.

[g] SAXS validation score for lowest energy Aha1 Rosetta structure computed with Crysol[25].

Table S3. Aha1 Assigned Experimental Interchain NOEs. [a]

Assigned Interchain atoms						upl [Å]
Chain A			Chain B			
Number	Residue	Atom	Number	Residue	Atom	
6	His	HA	116	Thr	HG2	5.50
8	Ile	HD1	116	Thr	HG2	5.50
116	Thr	HG2	6	His	HA	5.50
116	Thr	HG2	8	Ile	HD1	5.50
116	Thr	HA	119	Ala	HB	5.50
116	Thr	HG2	122	Leu	HG	5.50
116	Thr	HG2	123	Leu	HG	5.50
116	Thr	HG2	123	Leu	HD2	5.50
119	Ala	HB	119	Ala	HB	5.50
119	Ala	HB	116	Thr	HA	5.50
119	Ala	HB	120	Val	HG2	5.50
120	Val	HG1	123	Leu	HD2	5.50
120	Val	HG2	123	Leu	HD2	5.50
120	Val	HG1	123	Leu	HD1	5.50
120	Val	HG2	123	Leu	HD1	5.50
120	Val	HG2	119	Ala	HB	5.50
122	Leu	HG	116	Thr	HG2	5.50
123	Leu	HD1	120	Val	HG1	5.50
123	Leu	HD1	120	Val	HG2	5.50
123	Leu	HD2	120	Val	HG1	5.50
123	Leu	HD2	120	Val	HG2	5.50
123	Leu	HD2	116	Thr	HG2	5.50
123	Leu	HG	116	Thr	HG2	5.50

[a]The intermolecular distances listed are based on manually assigned peaks (chain A→B and symmetric B→A). A qualitative 5.50 Å upper limit distance was used to determine the restraint scores shown in Figure 5. These NOE derived distance restraints were not yet available when the structure was solved, and thus were not used during the structure calculation.

Supplementary Figures

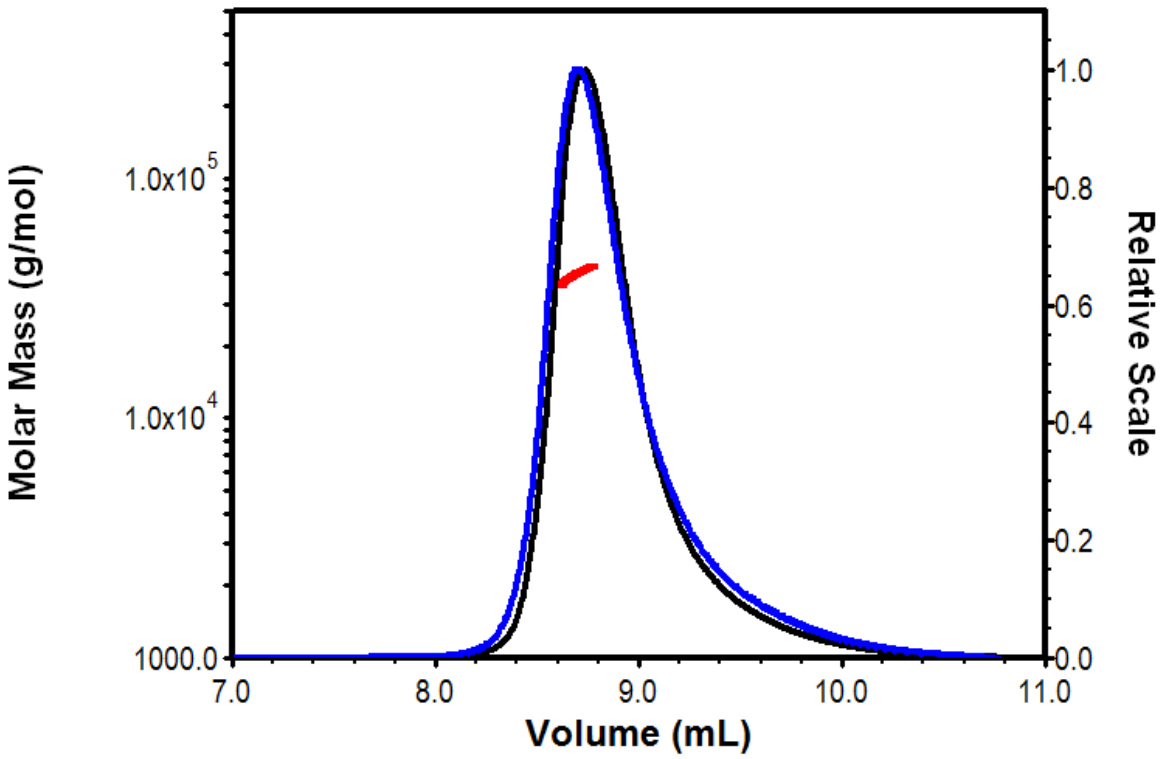


Figure S1. Static light scattering results for Aha1 (CsR251).

The NMR sample (30 ml) of [U-5%^{13}C-100%^{15}N]-Aha1 at 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM DTT, 1X Proteinase Inhibitors, 5% D_2O was injected onto an analytical gel-filtration column (Protein KW-802.5, Shodex, Japan) with the effluent monitored by refractive index (black trace, Optilab rEX) and 90° static light-scattering (blue trace; miniDAWN TREOS, Wyatt Technology) detectors. The resulting experimental molecular weight of isotopically labeled Aha1 is 40.3 kDa (red), the expected Aha1 molecular weight including affinity tag is 17.8 kDa.

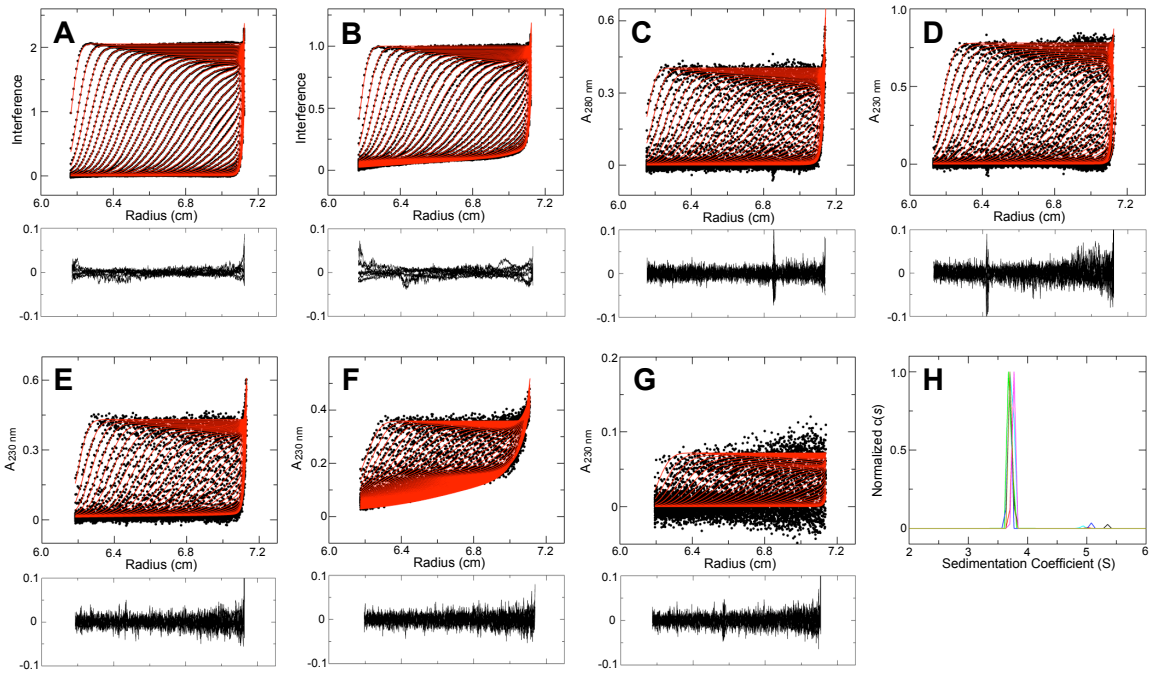


Figure S2. Sedimentation velocity analysis of Aha1.

A-G) Interference (A and B), absorbance at 280 nm (C), and absorbance at 230 nm (D-G) across the centripetal field for solutions of 40, 20, 10, 5, 3, 1.5, and 0.75 μM of Aha1, respectively. Data recorded in 1 min intervals. For clarity of presentation only every fifth point is plotted. Residuals are shown in bottom panels. **H)** Normalized c(*s*) distribution plots each depicting a single sedimenting species of 39 ± 3 kDa generated from the plots A-G, respectively. Data generated from 40, 20, 10, 5, 3, 1.5, and 0.75 μM of Aha1 are indicated by black, red, green, blue, cyan, magenta, and dark yellow, respectively.

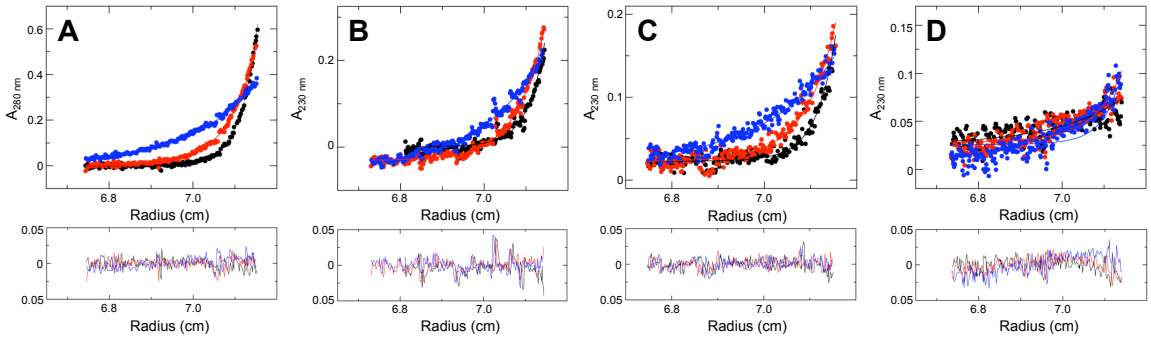


Figure S3. Concentration dependence of sedimentation equilibrium for Aha1.

A-D) Absorbance at 280 nm (A) or 230 nm (B-D) at equilibrium across the centripetal field for solutions of 4, 2, 1 and 0.25 μM Aha1 (A-D respectively). The black, red, blue circles represent data collected at 15,000, 23,000 and 29,000 rpm, respectively. For clarity only every third measured point is plotted. Fitted lines represent the global best fit of the data to a model for a monomer to dimer equilibrium with the upper limit for the dimerization constant (K_D) being 80 nM. Residuals are shown in bottom panels.

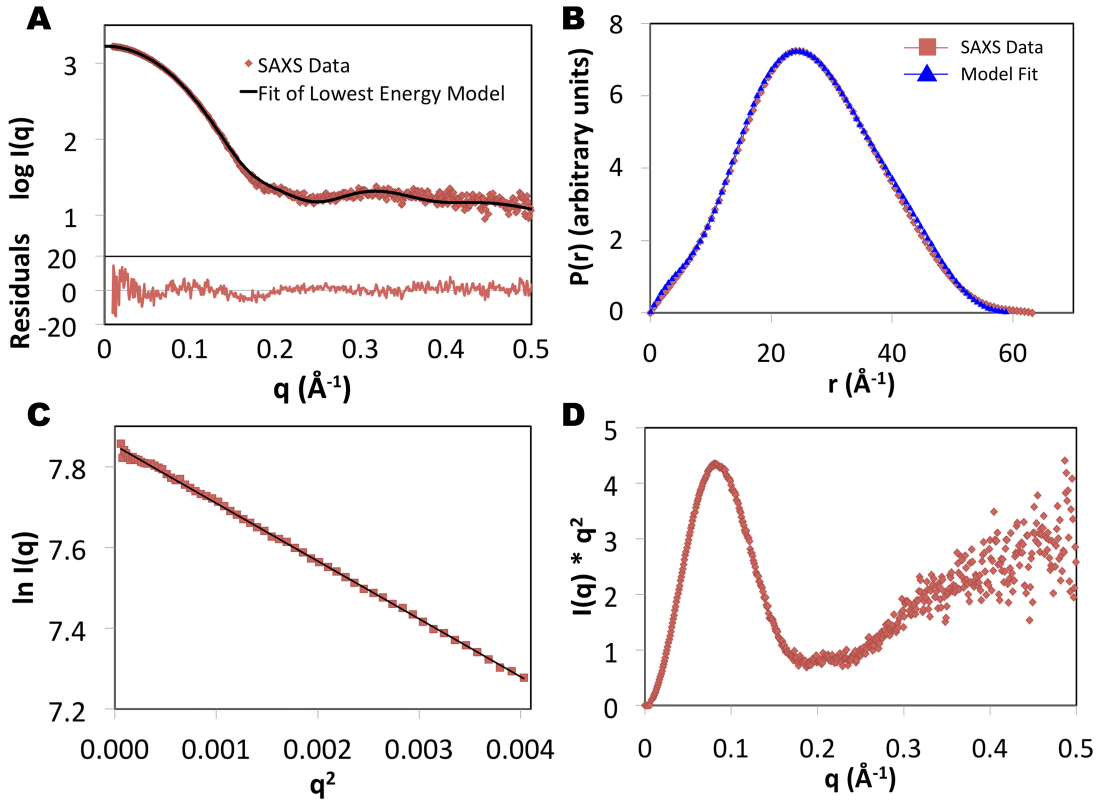


Figure S4. Analysis of experimental SAXS data and model fits.

A) The calculated fit to the lowest energy Rosetta structure (black line, χ^2 =1.71) overlaid onto the experimental SAXS data (red squares). The residuals of the fit are shown in the bottom panel. **B)** P(r) distributions (histogram of interatomic scattering vectors) derived from the raw data (red) versus model fitted values (blue). **C)** Guinier plot (showing no aggregation). **D)** Kratky plot demonstrating that Aha1 is folded in solution.

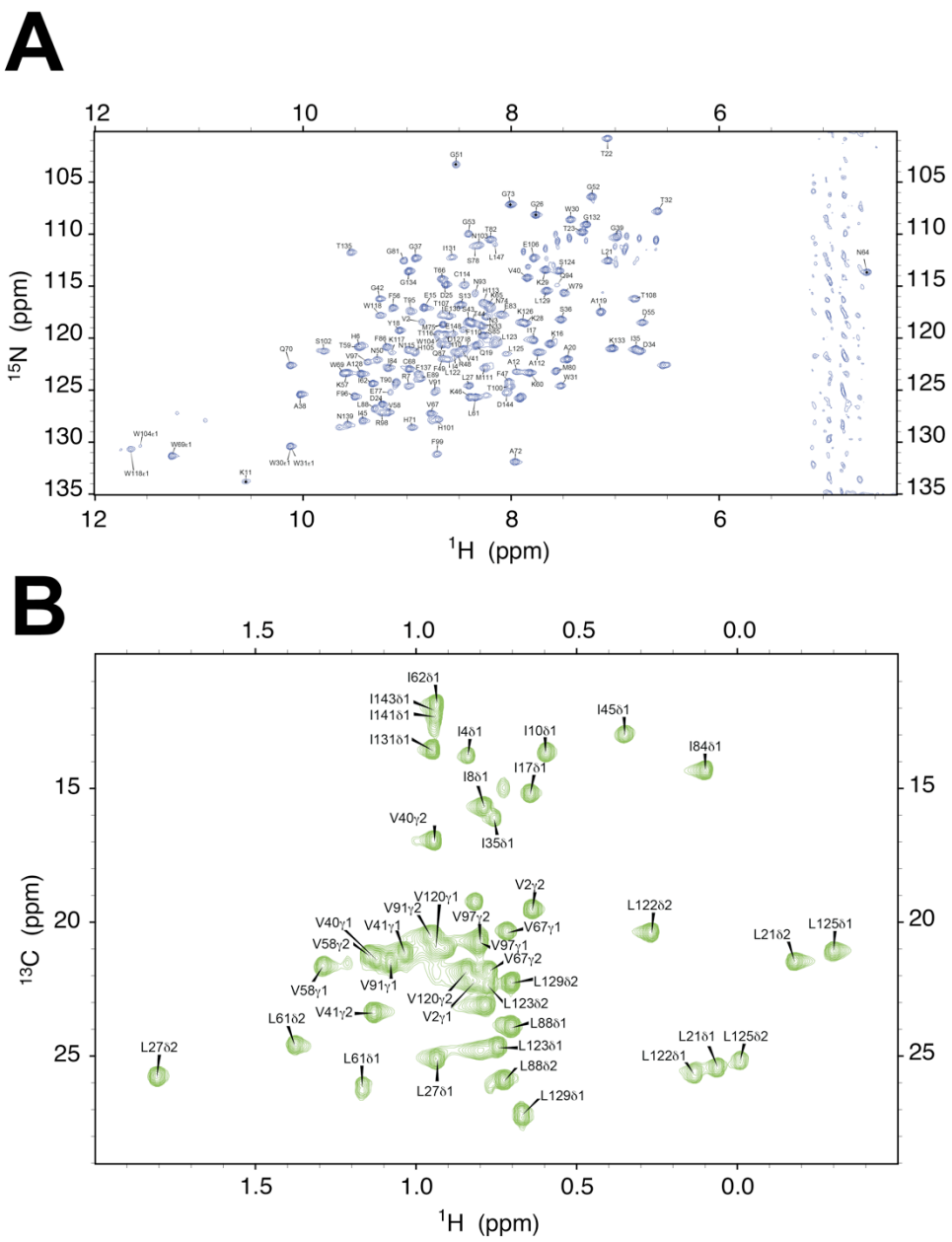


Figure S5. Two-dimensional HSQC spectra.

A) ^{1}H-^{15}N and **B)** ^{1}H-^{13}C HSQC spectra of *C. psychrerythraea* {[U-^{2}H,^{13}C,^{15}N]; Ileδ1-[$^{13}CH_3$]; Leu,Val-[$^{13}CH_3$]}-Aha1 in 10 mM Tris, 100 mM NaCl, 5 mM DTT, 50 mM DSS, 90% H_2O / 10% $^{2}H_2O$, pH 7.5 buffer collected at 25 °C on a Bruker AVANCE 800 MHz spectrometer. Backbone amide, side chain resonances of Trp as well as selectively protonated Ile, Leu and Val methyls are labeled with one-letter amino acid codes followed by their sequence numbers.

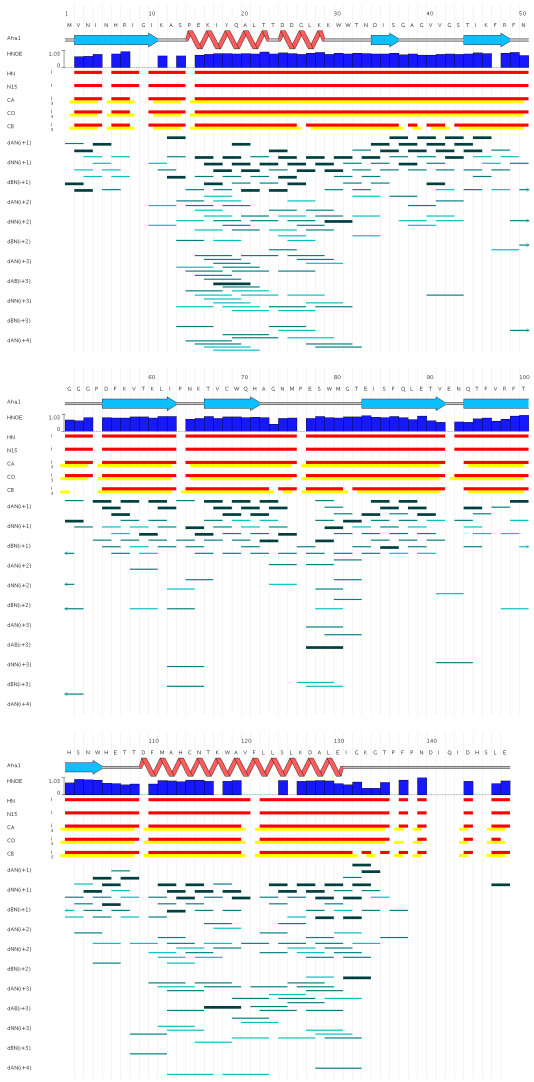
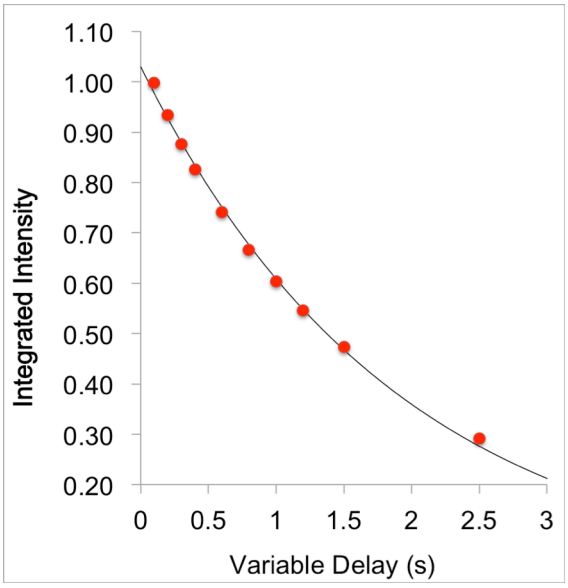
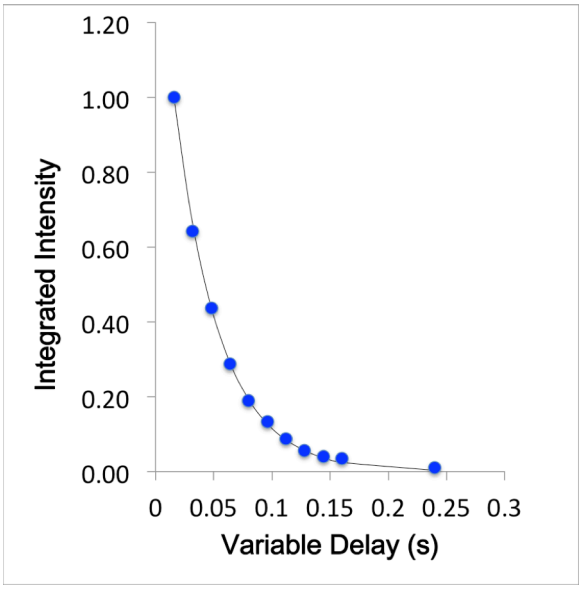


Figure S6. NMR chemical shifts connectivity map.

NMR Chemical shift are used to determine resonance assignments and secondary structure for Aha1 (BMRB ID 19235). The final six unassigned histidines in the C-terminal tag have been omitted. Intraresidue (i) and sequential (s) connectivities and sequential C′, C^{α}, and C^{β} resonances are shown as horizontal red and yellow lines, respectively. The complete inter-residue CYANA-derived NOE connectivities are shown as thin, medium, and thick black lines, corresponding to weak, medium, and strong NOE interactions. Bar graphs of $\{^{1}H\}$-^{15}N heteronuclear NOE data is shown in blue. The secondary structural elements in the final Aha1 NMR structure (PDB ID 2M89) are also shown.


Integrated Intensity
1.10
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0
0.5
1
1.5
2
2.5
3
Variable Delay (s)


Integrated Intensity
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0
0.05
0.1
0.15
0.2
0.25
0.3
Variable Delay (s)

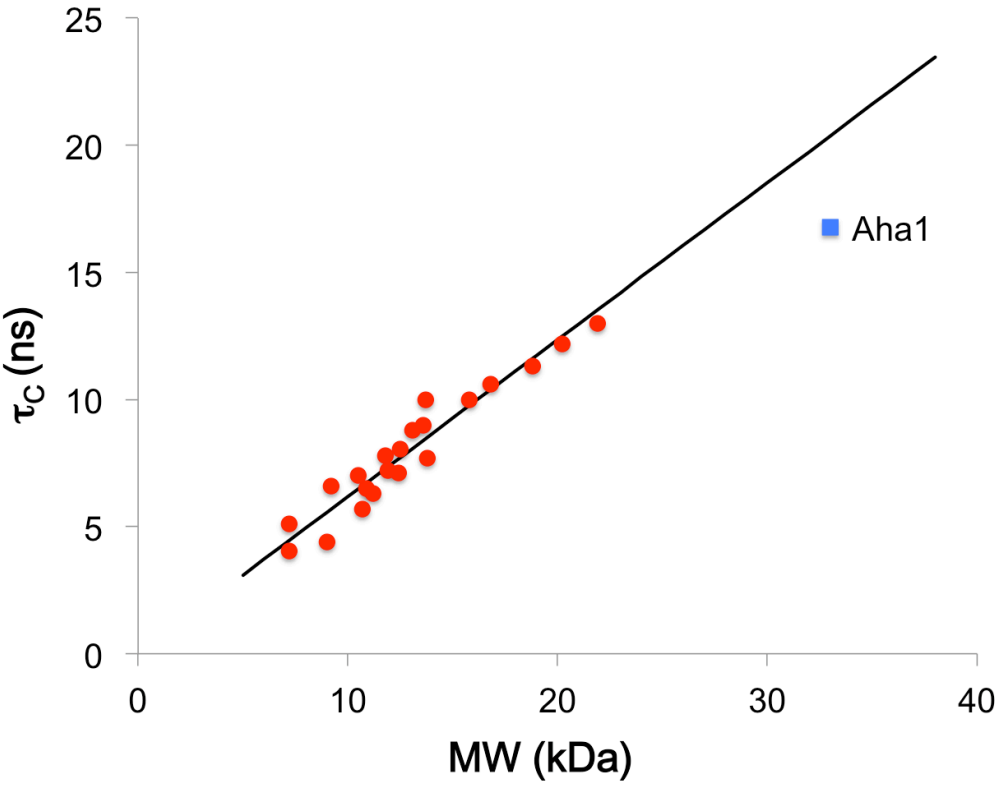

25
20
15
10
5
0
τC (ns)
0
10
20
30
40
MW (kDa)
Aha1

Figure S7. Backbone amide ^{15}N T_1 and T_2 relaxation data.

1D ^{15}N T_1 and T_2 relaxation data for *C. psychrerythraea* [U-5%^{13}C,^{15}N]-Aha1. Data were acquired on a Bruker AVANCE 800 MHz spectrometer at 25 °C using pseudo-2D ^{15}N T_1 and T_2 gradient experiments[12]. T_1 spectra were acquired with variable delays ranging from 0.05 to 2.5 s and a relaxation delay of 5 s. T_2 spectra were acquired with CPMG time ranging from 16 to 240 ms and a relaxation delay of 1.5 s. (Top): ^{15}N T_1 and T_2 values were extracted by plotting the decay of integrated $^1H^N$ intensity between $\delta \approx 7$ to 11 ppm and fitting the curves with standard exponential equations using the program 't1guide' within TopSpin 2.1 (Bruker BioSpin). (Bottom): Plot of rotational correlation time, τ_c (ns), versus protein molecular weight (kDa) for known monomeric NESG targets (red) of ranging size (taking into account isotope enrichment as well as affinity tags in the sequence). ^{15}N T_1/T_2 data for all monomeric proteins used for the τ_c vs. MW plot were obtained at 25 °C, and analyzed as described above. For each protein, the τ_c was calculated from the ^{15}N T_1/T_2 ratio using the following approximation of literature relaxation equations[31,32]:

Eq. 1 $$\tau_c \approx \left(\sqrt{\frac{6T_1}{T_2} - 7} \right) / 4\pi\nu_N$$

where ν_N is the frequency of ^{15}N in Hz. The Aha1 τ_c value point matches the approximate value for dimeric MW, confirming the dimer association in solution. The ^{15}N T_1, ^{15}N T_2, and τ_c for Aha1 are 1.9 sec, 39 ms and 17.0 ns, respectively.

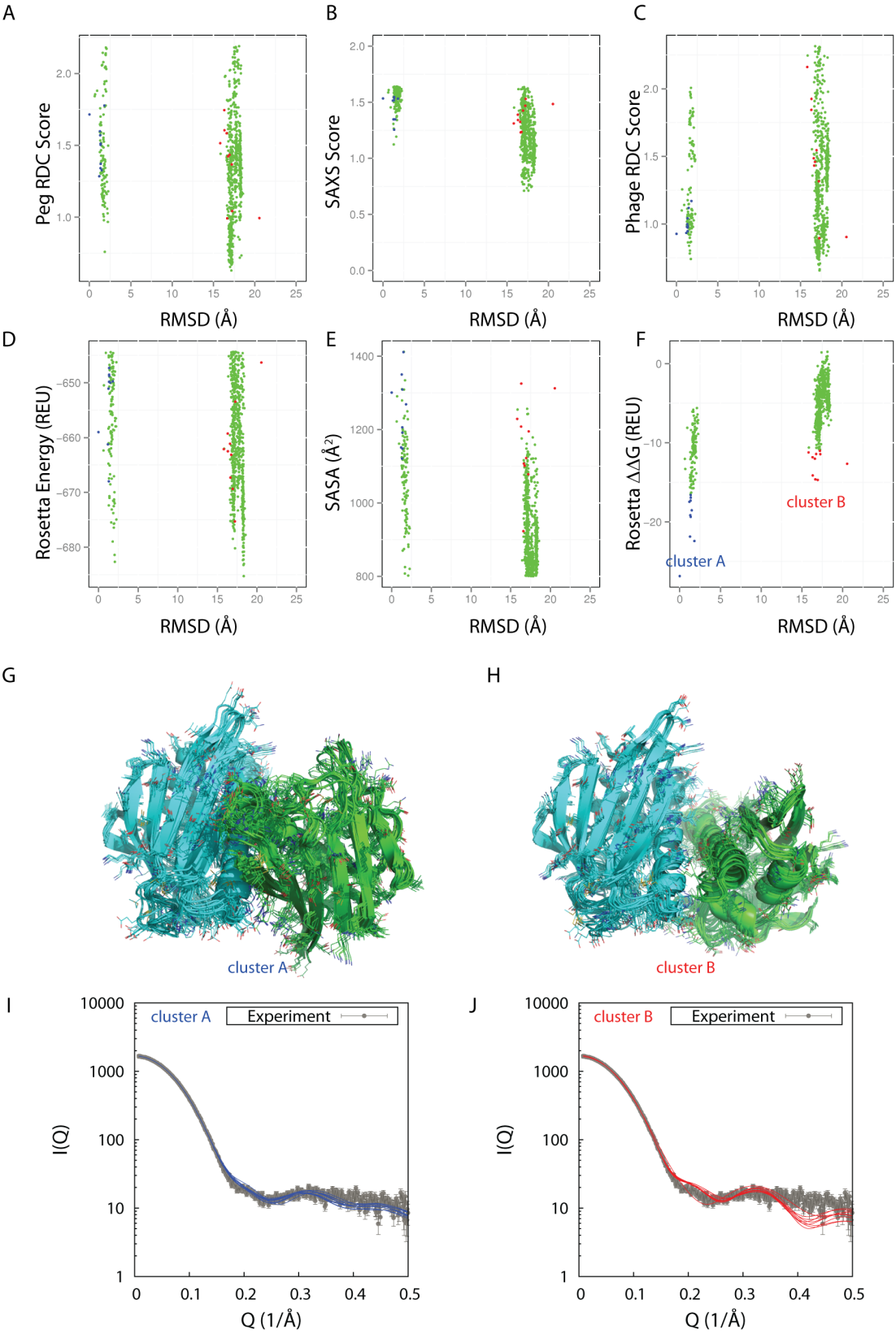

A
B
C
Peg RDC Score
SAXS Score
Phage RDC Score
RMSD (Å)
D
E
F
Rosetta Energy (REU)
SASA (Å²)
Rosetta ΔΔG (REU)
cluster A
cluster B
G
H
cluster A
cluster B
I
J
cluster A
cluster B
Experiment
I(Q)
Q (1/Å)

Figure S8. Comparison of alternative Phase II dimer conformations.

Experimental data score terms, structure quality terms and Rosetta energies vs. RMSD plot for all models generated in the Phase II/local refinement calculations after filtering (900 conformations – green points). As a reference structure, we used the model with the lowest interface interaction score (1st model in the submitted PDB ensemble, residues 1-131). The 10 conformations with lowest predicted ΔΔG values were selected as the final ensemble ("cluster A"- blue points), while an alternative cluster of conformations is further highlighted in the plots ("cluster B"). In detail: **A)** Peg RDC score computed as RMS(DEXP-DCALC) /DA, where DA is the alignment tensor magnitude. **B)** SAXS score: RMS(IEXP-ICALC). **C)** Phage RDC scores. **D)** Rosetta score12 all-atom energy (in Rosetta Energy Units) **E)** SASA (Å^2). **F)** Interface free energy (ddg or ΔΔG) defined as: ΔG_{dimer} - 2 * $\Delta G_{monomer}$. **G)** Structural superposition of the 10-dimer conformations in cluster A, also the final ensemble. **H)** Structure superposition of the 10-lowest ΔΔG structures in cluster B. Overlay of SAXS profiles computed independently for each member in cluster A (**I)** and cluster B **(J)** alongside the experimental profile (grey).

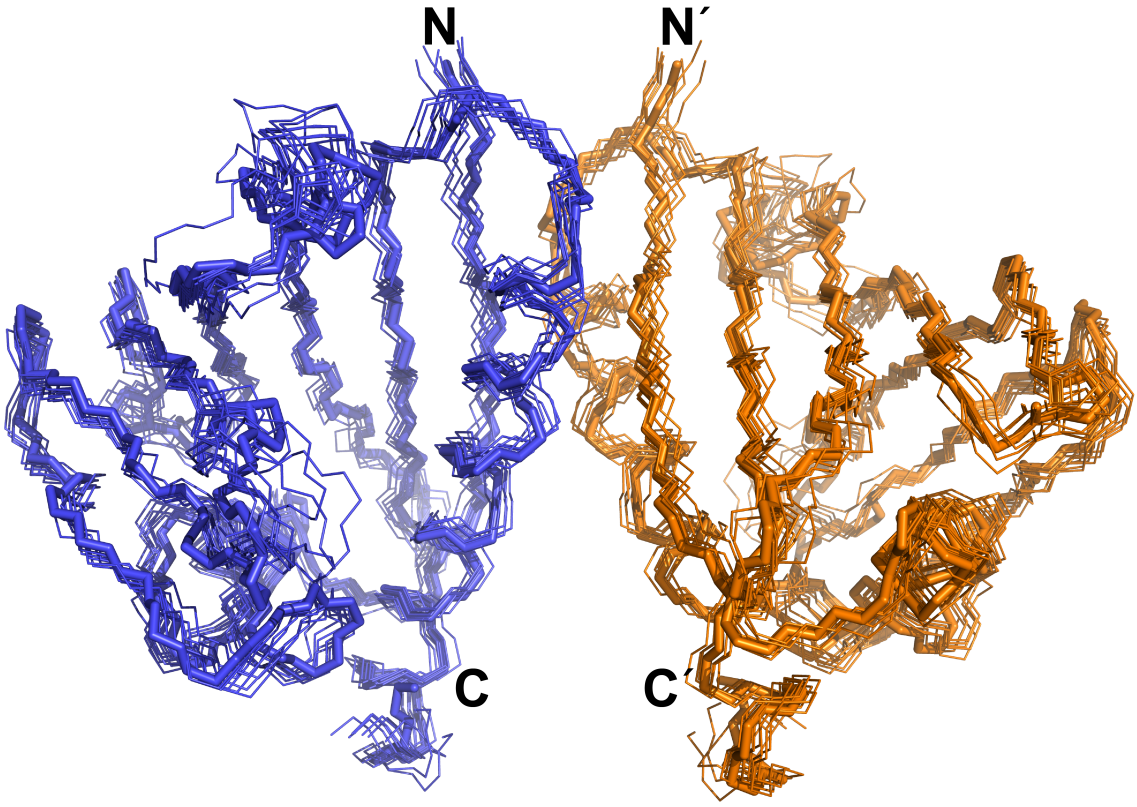


Figure S9. Final hybrid structural ensemble.

Final Rosetta ensemble of the *C. psychrerythraea* Aha1 dimer (PDB ID 2M89), computed using the sparse intra-residue NOEs, RDCs and SAXS data. The two subunits are colored blue and orange and the location of C-termini are labeled. An average value of 0.6 Å backbone RMSD was computed relative to the 1st structure for the ordered region, res. 1-134 (chain A and B), indicating a highly converged ensemble.

References

1. Acton TB, Xiao R, Anderson S, Aramini J, Buchwald WA, Ciccosanti C, Conover K, Everett J, Hamilton K, Huang YJ, Janjua H, Kornhaber G, Lau J, Lee DY, Liu G, Maglaqui M, Ma L, Mao L, Patel D, Rossi P, Sahdev S, Shastry R, Swapna GV, Tang Y, Tong S, Wang D, Wang H, Zhao L, Montelione GT. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. Methods Enzymol 2011;493:21-60.

2. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. J Biomol NMR 1999;13(4):369-374.

3. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione BT, Nilsson B. High-level production of uniformly 15N- and 13C-enriched fusion proteins in Escherichia coli. J Biomol NMR 1996;7(2):131-141.

4. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 1998;5(12):1065-1074.

5. Ruckert M, Otting G. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. J Am Chem Soc 2000;122(32):7793-7797.

6. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6(3):277-293.

7. Goddard TD, Kneller DG. SPARKY 3: University of California, San Francisco; 2006.

8. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 2009;44(4):213-223.

9. Lange OF, Baker D. Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation. Proteins 2012;80(3):884-895.

10. Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T. NMR data collection and analysis protocol for high-throughput

protein structure determination. Proc Natl Acad Sci U S A 2005;102(30):10487-10492.

11. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry 1989;28(19):7510-7516.

12. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 1994;33(19):5984-6003.

13. Guntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol 2004;278:353-378.

14. Zwahlen C, Legault P, Vincent SJ, Greenblatt J, Konrat R, Kay LE. Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage λ N-Peptide/boxB RNA Complex. J Am Chem Soc 1997;119:6711-6721.

15. Tjandra N, Grzesiek S, Bax A. Magnetic Field Dependence of Nitrogen,àíProton J Splittings in 15N-Enriched Human Ubiquitin Resulting from Relaxation Interference and Residual Dipolar Coupling. J Am Chem Soc 1996;118(26):6264-6272.

16. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S. An easy way to include weak alignment constraints into NMR structure calculations. J Biomol NMR 2001;21(3):275-280.

17. Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. Atomic model of the type III secretion system needle (vol 486, pg 276, 2012). Nature 2012;488(7413).

18. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 2011;487:545-574.

19. Clore GM, Gronenborn AM, Bax A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. J Magn Reson 1998;133(1):216-221.

20. Sgourakis NG, Lange OF, DiMaio F, Andre I, Fitzkee NC, Rossi P, Montelione GT, Bax A, Baker D. Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings. J Am Chem Soc 2011;133(16):6288-6298.

21. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66(4):778-795.

22. Zweckstetter M. NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 2008;3(4):679-690.

23. Smolsky IL, Liu P, Niebuhr M, Ito K, Weiss TM, Tsuruta H. Biological small-angle x-ray scattering facility at the Stanford synchrotron radiation laboratory. J Appl Crystallogr 2007;40:S453-S458.

24. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr 2003;36:1277-1282.

25. Svergun DI, Barberato C, Koch MH. CRYSOL - a program to evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J Appl Crystallogr 1995;28:768-773.

26. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Computer-aided interpretation of analytical sedimentation data for proteins. In: S.E. Harding AJR, J.C. Horton, editor. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Cambridge, United kingdom: J R Soc Chem; 1992. p 90–125.

27. Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, Schuck P. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem 2004;326(2):234-256.

28. Johnson ML, Straume M. Comments on the analysis of sedimentation equilibrium experiments. In: Schuster TM, Laue TM, editors. Modern Analytical Ultracentrifugation. Boston, MA: Birkhäuser; 1994. p 37-65.

29. Moseley HN, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 2004;28(4):341-355.

30. Cornilescu GM, J. L.; Ottiger, M.; Bax, A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. J Am Chem Soc 1998;120:6836-6837.

31. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 1989;28(23):8972-8979.

32. Fushman D, Ohlenschlager O, Ruterjans H. Determination of the backbone mobility of ribonuclease T1 and its 2'GMP complex using molecular dynamics simulations and NMR relaxation data. J Biomol Struct Dyn 1994;11(6):1377-1402.