Structural Characterization of the Self-Association Domain of Swallow Supplementary Material

Nikolaus M. Loening[1,*] and Elisar Barbar[2]

[1]Department of Chemistry, Lewis & Clark College, Portland, Oregon 97219, USA
[2]Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331, USA

Materials and Methods:

NMR Sample Preparation

Isotopically-labeled (^{15}N and ^{15}N/^{13}C) samples of Swa$_{DIMER}$ were prepared as previously described[1]. Briefly, the sequence for Swallow residues 205-275 was cloned into a Champion pET SUMO vector (Invitrogen, Carlsbad, CA), which was then used to transform *Escherichia coli* BL21(DE3) cells. Bacterial cultures were grown in isotopically-labeled (^{15}N or ^{15}N/^{13}C) MJ9 minimal media[2] at 37 ºC, and protein synthesis was induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG). Cells were lysed by sonication and then, after clarification by ultracentrifugation, purified by immobilized metal affinity (IMAC) chromatography. Following cleavage of the His$_6$-SUMO tag by SUMO protease, a subsequent IMAC step was used to remove the tag and protease, and the resulting flow-through was further purified by size-exclusion chromatography (SEC). SEC fractions containing Swa$_{DIMER}$ were pooled, exchanged into NMR buffer [20 mM MES (pH 5.6), 10 mM NaCl, 1 mM sodium azide, 0.2 mM 2,2-dimethylsilapentane-5-sulfonic acid (DSS), and 1× cOmplete™ protease inhibitor cocktail (Roche, Mannheim, Germany)], and concentrated to 0.5 mM. Protein concentrations were determined by measuring the absorbance at 280 nm.

The pH of the NMR buffer was chosen based on previous work using multiangle laser light scattering (MALLS) that showed that, at this pH, Swa$_{DIMER}$ exclusively forms homodimers, whereas at a higher pH (8 and above) higher-order oligomers are formed[3]. The oligomeric state of Swa$_{DIMER}$ is not only pH-dependent but also temperature-dependent; we observed that when stored at 4 ºC, the sample would form a white aggregate that would rapidly disappear after a few minutes of incubation at 40 ºC.

NMR Spectroscopy

All NMR spectra were acquired using a sample temperature of 40 ºC. NMR spectra for assignment and for determining structural constraints were acquired on Bruker (Billerica, MA) spectrometers at 600, 800, 900, and 950 MHz for ^{1}H. Backbone resonances were assigned using band-selective excitation short transient (BEST) variants of transverse relaxation-optimized spectroscopy (TROSY)-based triple resonance sequences[4] and sidechains were assigned using TROSY-H(CCO)NH, TROSY-(H)CC(CO)NH, and TROSY-HBHA(CO)NH experiments. ^{1}H chemical shifts were referenced to the ^{1}H signal for DSS, and ^{13}C and ^{15}N chemical shifts were referenced indirectly using chemical shift ratios[5]. For analyzing the periodicity of secondary chemical shifts,

random coil chemical shift values for Swa_{DIMER} at 40 ºC and pH 5.6 were calculated using a web-based tool[6] that uses published reference values[7,8].

To collect $^{1}D_{HN}$ residual dipolar coupling (RDC) values, a ^{15}N-labeled sample of Swa_{DIMER} was aligned with Pf1 phage[9] (ASLA Biotech). At a phage concentration of 10 mg/mL and a temperature of 40 ºC, the quadrupole splitting of the D_2O signal was 3.1 Hz. $^{1}D_{HN}$ RDC values were measured using a variation of the ARTSY[10] experiment suitable for protonated samples. For resonances with sufficient chemical shift resolution, the RDC values were confirmed by comparing the $^{1}D_{HN}$ values from the ARTSY experiment to those calculated from measuring the change in peak position between a BEST-HSQC spectrum and a BEST-TROSY spectrum. The alignment tensor was determined using REDCAT[11] to have a magnitude of 32.0 Hz and a rhombicity of 0.337. We also attempted to align Swa_{DIMER} by soaking it into a compressed polyacrylamide gel, but the protein interacted too strongly with the gel matrix, resulting in poor-quality spectra.

$\{^{1}H\}$-^{15}N NOE (HetNOE) values were determined on a Bruker spectrometer operating at 600 MHz for ^{1}H by collecting, in an interleaved fashion, a spectrum in which ^{1}H spins were saturated by 90º radiofrequency pulses every 5 ms during the 3 s relaxation delay prior to the initial ^{15}N pulse and a control spectrum without these saturation pulses. Transverse (R_2) and longitudinal (R_1) ^{15}N relaxation rates were measured at 800 MHz for ^{1}H using HSQC-based experiments. The longitudinal relaxation measurement used delay times of 0, 100, 200, 400, 600, 800, 1100, 1200, and 1500 ms, whereas the transverse relaxation measurement used CPMG periods of 10, 30, 50, 70, 90, 130, 170, 210, and 250 ms.

Further Analyses:

Secondary Chemical Shifts:

Plots of the secondary chemical shifts for the backbone residues, with coloring to highlight the positions of the amino acids in the coiled-coil heptad repeat, are shown in Figure S1. Systematic positive deviations in the secondary chemical shifts are observed for $^{13}C_{\alpha}$ and systematic negative deviations are observed for $^{13}C_{\beta}$, which is indicative of α-helical secondary structure and consistent with a coiled-coil structure. A correlation between secondary chemical shift and heptad position is seen for $^{1}H_{N}$ shifts, with residues in the **a**, **d**, and **e** positions tending to have more positive secondary shifts, and those in the **c** and **f** positions having more negative secondary shifts, a trend that has been previously noted for the leucine-zipper region of GCN4[12]. Interestingly, the pattern for the $^{1}H_{\alpha}$ secondary chemical shifts appears to be the opposite of that for the $^{1}H_{N}$ chemical shifts, with α-protons at the **a**, **d**, and **e** positions tending to have more negative secondary shifts than α-protons located at the other heptad positions. The carbonyl ($^{13}C_{o}$) secondary chemical shifts show a pattern that has previously been observed for coiled-coils[13] in which the magnitudes of these secondary shifts are diminished for the **a** and **d** positions relative to the other heptad positions. For the other carbon atoms, no obvious pattern correlating secondary chemical shifts and heptad position was observed, which stands in contrast to recently reported results for the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ secondary shifts for the leucine-zipper region of GCN4[14]. This discrepancy may be due to the more regular sequence of GCN4, which always has leucine in the **d** position, or because the GCN4 sequence only has three heptad repeats, which could result in patterns arising by chance.

Coiled-Coil Propensity for Swa$_{DIMER}$:

Coiled-coil prediction algorithms were used to analyze the sequence for Swa$_{DIMER}$ as well as the wild-type sequence for residues 205-275 of Swallow (Swa$_{WT}$), which lacks the R224E, K244I, C253D, and C265A mutations. As shown in Figure S3A, the COILS algorithm[15] predicts a lower probability for coiled-coil structure in the center of Swa$_{WT}$. The size of this region of low coiled-coil propensity becomes smaller, but does not disappear, when the sequence for Swa$_{DIMER}$ is used for the COILS prediction. In contrast, the more recently developed Multicoil2 algorithm[16] predicts a relatively uniform coiled-coil propensity for residues 210-270 for both Swa$_{WT}$ and Swa$_{DIMER}$. Analysis of the sequences with the Marcoil algorithm[17] produced results that were almost identical to those from the Multicoil2 algorithm (data not shown).

The secondary structure prediction (SSP) score[18] uses experimentally-determined backbone chemical shifts to provide residue-level information about the expected fraction of α-helical and β-sheet structure. Stretches of sequence with SSP scores close to +1 indicate fully-formed α-helices, whereas stretches with scores close to –1 indicate fully-formed β-sheets. Values closer to 0 reflect lower secondary structural propensities, which typically indicates greater disorder. Our SSP analysis using all available backbone chemical shifts (Figure S3B) indicates that Swa$_{DIMER}$ is uniformly α-helical for most of its length, with only a small amount of disorder at the C-terminus. As the SSP score provides insight on secondary structural propensity, and not coiled-coil propensity, the absence of a dip in SSP values in the middle of the sequence is not inconsistent with the COILS analysis results shown in Figure S3A.

Residual Dipolar Coupling Measurements and NMR Relaxation

In most proteins, a distribution of positive and negative $^{1}D_{HN}$ residual dipolar coupling (RDC) values is observed due to the variation in the orientation of the N-H bond vectors for the backbone amides. In the case of the backbone amides of Swa$_{DIMER}$, however, the measured $^{1}D_{HN}$ RDC values (Figure S3C) are predominantly in the range of –35 to –25 Hz. This preponderance of similar $^{1}D_{HN}$ RDC values is consistent with a coiled-coil structure, as in coiled-coils the N-H bond vectors of all the backbone amides are oriented nearly parallel to the long axis of the coiled coil. Some correlations between RDC values and heptad position are observable, with amino acids in the **a**, **d**, and **e** positions tending to have more negative RDC values than amino acids in the other heptad positions.

Although the ^{15}N relaxation rates for Swa$_{DIMER}$ were previously published[1], the graphs in Figures S3D-G are from a reanalysis of the original data that corrects some errors. Within the error of the measurements, the ^{15}N R_2, ^{15}N R_1, and $\{^{1}H\}$-^{15}N HetNOE values are relatively uniform for the central part of Swa$_{DIMER}$, indicating that residues 215-270 form a rigid structure without notable variations in dynamics on the pico- and nanosecond time scales. The dynamics indicate higher flexibility at the termini. This is especially apparent for the C-terminus, an observation that matches the fraying of this terminus evident in the NMR structure.

CCCP Analysis of the Swa$_{DIMER}$ Structure

We analyzed the Swa$_{DIMER}$ structure using the CCCP (Coiled-Coil Crick Parameterization) tool[19], which fits α-helices to the Crick parameterization for ideal coiled-coils[20]. Analysis of residues 205-272 of Swa$_{DIMER}$ (Table S1) shows a structure with nearly average values for the superhelix radius (R_0 = 5.21 Å) and helix radius (R_1 = 2.178 Å), but with a –6.28º/residue superhelical (ω_0) frequency, which is considerably more negative than the median value of –3.5º/residue found in

an analysis of 868 coiled-coil structures[19]. The superhelical frequency of –6.28º/residue indicates that Swa_{DIMER} forms a relatively tightly wound left-handed coiled-coil in which approximately 57 residues complete one period of the superhelix. The helix frequency (ω_1) represents the turn of the helix per residue; the 103.44º value found for Swa_{DIMER} is very close to the ideal value of 102.86º (=720º/7) for a coiled-coil formed from heptad repeats. Likewise, the rise per residue (d) of 1.508 Å for Swa_{DIMER} is very close to the ideal value of 1.5 Å.

These parameters represent average values for the entire length of the coiled-coil region of Swa_{DIMER} and disguise variations in the structure. Closer examination of the structure showed that it is not actually uniform, but rather has a pronounced bulge in the middle, a feature which is readily apparent when Swa_{DIMER} is compared to another recently determined NMR coiled-coil structure (Figure 2D). This bulge has a location that matches the region of lower coiled-coil propensity predicted by COILS (Figure S3A), and its presence is reflected in the variation in the coiled-coil parameter values calculated for subsets of the structure (see rightmost columns in Table S1). For example, the superhelix radius (R_0), which is the distance between the center of each α-helix to the coiled-coil axis, is ~5.0 Å for the first 30 residues but increases to ~5.7 Å for residues 235-255 before decreasing to ~5.3 Å for residues 255-272. The variation in the coiled-coil parameters for different regions of the structure is reflected in the RMSD (root-mean-square deviation) values provided in Table S1, which compare the position of the C_α atoms in the Swa_{DIMER} structure to where they would be in an ideal coiled-coil structure. This RMSD for residues 207-272 is relatively high as the entire length of the Swa_{DIMER} structure is not well-described by a single set of coiled-coil parameters, whereas smaller RMSD values are seen when subsets of the structure are analyzed as these subsets can be better described by single sets of parameters.

References

1. Kidane AI, Song Y, Nyarko A, Hall J, Hare M, Löhr F, Barbar E (2013) Structural features of LC8-induced self-association of swallow. Biochemistry 52:6011–6020.
2. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, Nilsson B (1996) High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in Escherichia coli. J. Biomol. NMR 7:131–141.
3. Barbar E, Nyarko A (2014) NMR Characterization of Self-Association Domains Promoted by Interactions with LC8 Hub Protein. Comput. Struct. Biotechnol. J. [Internet] 9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3995210/
4. Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J. Biomol. NMR 55:311–321.
5. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J. Biomol. NMR 6:135–140.
6. Maltsev AS (2019) Poulsen IDP/IUP random coil chemical shifts. Available from: https://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/
7. Kjaergaard M, Brander S, Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. J. Biomol. NMR 49:139–149.

8. Kjaergaard M, Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. J. Biomol. NMR 50:157–165.
9. Hansen MR, Hanson P, Pardi A [15] Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. In: Methods in Enzymology. Vol. 317. RNA - Ligand Interactions, Part A. Academic Press; 2000. pp. 220–240.
10. Fitzkee NC, Bax A (2010) Facile measurement of ^{1}H-^{15}N residual dipolar couplings in larger perdeuterated proteins. J. Biomol. NMR 48:65–70.
11. Valafar H, Prestegard JH (2004) REDCAT: a residual dipolar coupling analysis tool. J. Magn. Reson. San Diego Calif 1997 167:228–241.
12. Goodman EM, Kim PS (1991) Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide. Biochemistry 30:11615–11620.
13. Greenfield NJ, Huang YJ, Palm T, Swapna GV, Monleon D, Montelione GT, Hitchcock-DeGregori SE (2001) Solution NMR structure and folding dynamics of the N terminus of a rat non-muscle alpha-tropomyosin in an engineered chimeric protein. J. Mol. Biol. 312:833–847.
14. Kaplan AR, Brady MR, Maciejewski MW, Kammerer RA, Alexandrescu AT (2017) Nuclear Magnetic Resonance Structures of GCN4p Are Largely Conserved When Ion Pairs Are Disrupted at Acidic pH but Show a Relaxation of the Coiled Coil Superhelix. Biochemistry 56:1604–1619.
15. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164.
16. Trigg J, Gutwin K, Keating AE, Berger B (2011) Multicoil2: Predicting Coiled Coils and Their Oligomerization States from Sequence in the Twilight Zone. PLOS ONE 6:e23519.
17. Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18:617–625.
18. Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. Protein Sci. Publ. Protein Soc. 15:2795–2804.
19. Grigoryan G, DeGrado WF (2011) Probing Designability via a Generalized Model of Helical Bundle Geometry. J. Mol. Biol. 405:1079–1100.
20. Crick FHC (1953) The packing of α-helices: simple coiled-coils. Acta Crystallogr. 6:689–697.

Table S1: Coiled-coil parameters calculated for the Swa_{DIMER} ensemble using CCCP[a].

Coiled-Coil Parameter	Residues 207-272	Residues 207-235	Residues 235-255	Residues 255-272
R_0 (superhelix radius, Å)	5.21 (±0.07)	4.98 (±0.16)	5.71 (±0.06)	5.26 (±0.13)
R_1 (helix radius, Å)	2.178 (±0.012)	2.205 (±0.018)	2.310 (±0.012)	2.12 (±0.04)
ω_0 (superhelix frequency, °/residue)	–6.28 (±0.15)	–6.3 (±0.3)	–5.9 (±0.2)	–8.9 (±0.4)
ω_1 (helix frequency, °/residue)	103.44 (±0.10)	104.3 (±0.3)	102.0 (±0.2)	106.0 (±0.6)
α (pitch angle, °)	–22.2 (±0.6)	–21.0 (±1.4)	–22.8 (±1.0)	–32.1 (±1.6)
d (rise per residue, Å)	1.508 (±0.007)	1.519 (±0.012)	1.527 (±0.009)	1.537 (±0.017)
RMSD (Å)	1.35 (±0.10)	0.88 (±0.07)	0.54 (±0.04)	1.00 (±0.14)

[a]Values are averages calculated using the 20 structures in the NMR ensemble, and errors represent the standard deviations for these values.

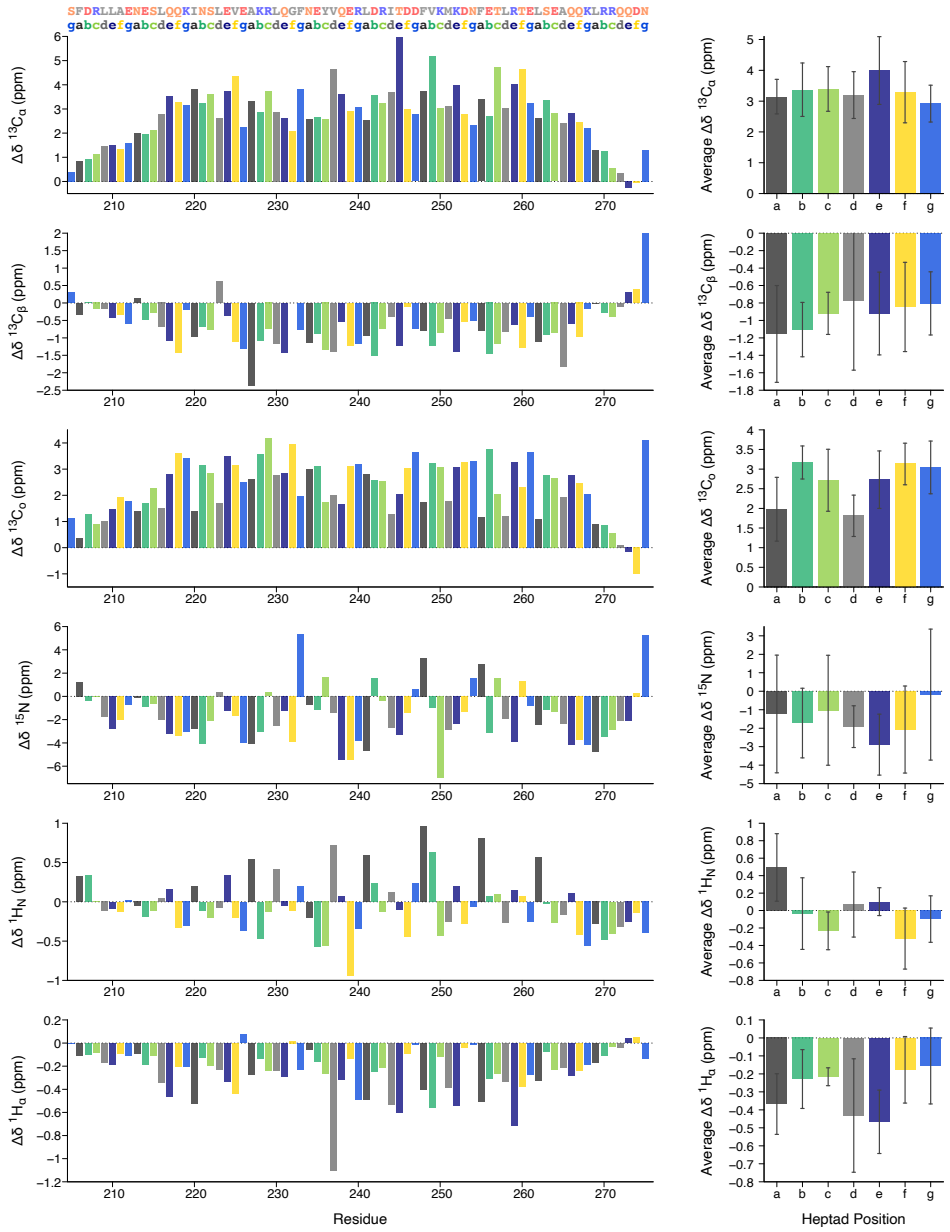

Figure S1: Secondary chemical shifts for Swa$_{DIMER}$ backbone resonances.

On the left are the secondary chemical shifts (Δδ) for the Swa$_{DIMER}$ backbone $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C_o$, ^{15}N, $^{1}H_N$, and $^{1}H_\alpha$ resonances. The average values for the secondary shifts by heptad position for residues 215-265 are shown on the right; the error bars are set to ± the standard deviation. The sequence for Swa$_{DIMER}$ at the top is colored by amino acid type using the same scheme [hydrophobic (grey), positive (red), negative (blue), neutral (orange)] as in Figure 1C. Below the sequence are letters that indicate the position of the individual amino acids in the coiled-coil heptad repeat (**a** and **d** in greys, **e** and **g** in blues, and **b** and **c** in greens, and **f** in yellow). This coloring scheme is retained in the plots to highlight any correlations between the plotted values and heptad position.

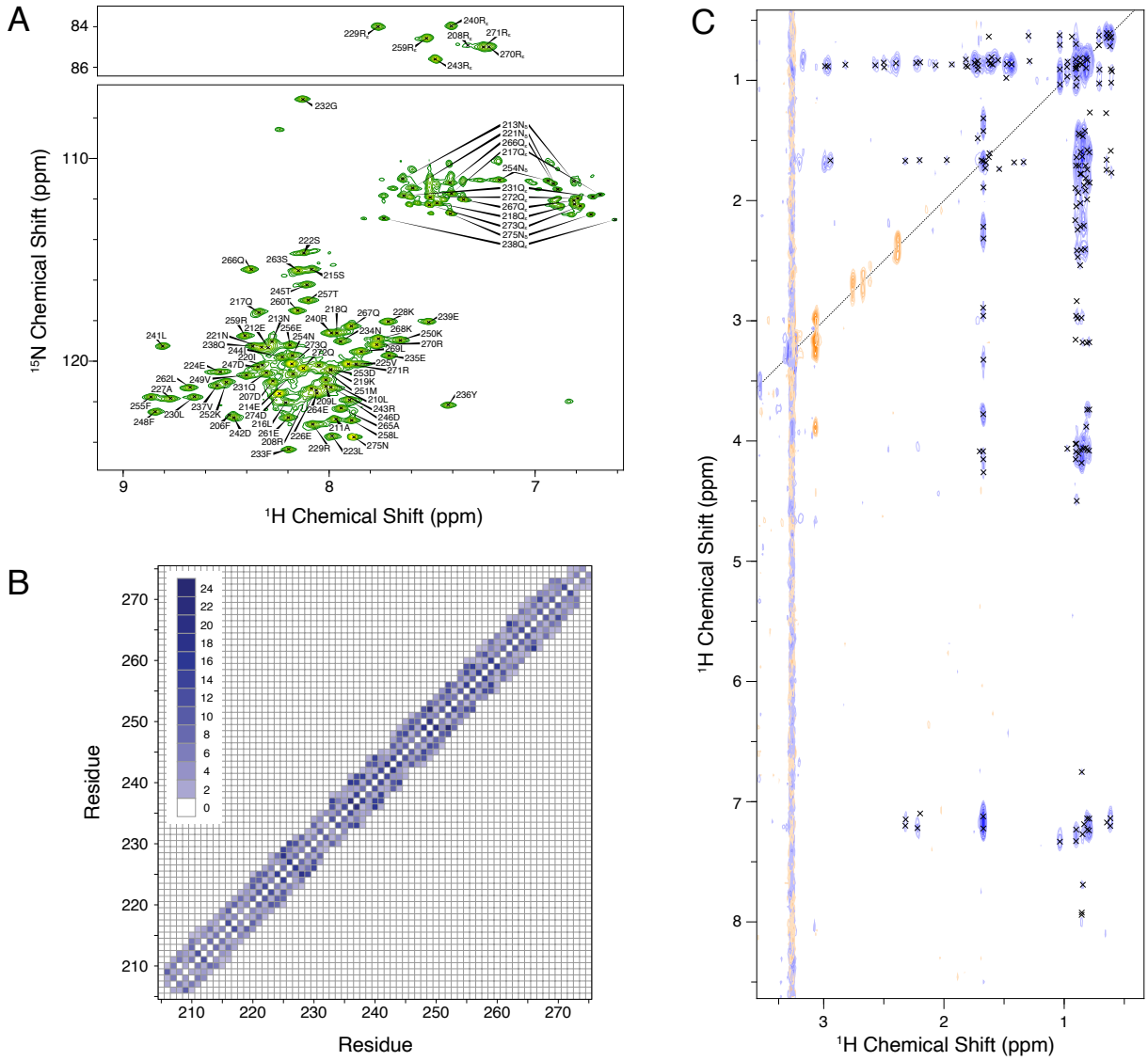

Figure S2: ^{1}H-^{15}N correlation spectrum and NOE contact map.

(A) ^{1}H-^{15}N TROSY-HSQC spectrum of Swa$_{DIMER}$ acquired at 950 MHz for ^{1}H with a sample temperature of 40 ºC. Peaks in this TROSY spectrum were shifted by 46 Hz ($^{1}J_{NH}/2$) in each dimension so that the chemical shift assignments for the backbone amides match those from conventional heteronuclear correlation experiments. For NH_2 groups [the asparagine (N_δ) and glutamine (Q_ε) sidechains] this shift only partially corrects the ^{15}N dimension, so these peaks are located 46 Hz (0.48 ppm) downfield of where they would appear in the ^{15}N dimension of conventional heteronuclear correlation experiments.

(B) Interresidue NOE contact map for Swa$_{DIMER}$ after using ARIA to assign NOE peaks. The pattern of medium-range contacts (i±3 and i±4) observed for all residues is from amino acids

separated by one complete turn of the α-helical secondary structure. The coloring represents the number of NOE restraints contributing to each contact.

(C) Two-dimensional projection of a 3D $^{13}C/^{15}N$-filtered, ^{13}C-separated NOESY-HSQC spectrum acquired at 950 MHz for ^{1}H with a 120 ms mixing time and a sample temperature of 40 ºC. The × symbols denote peaks due to intermolecular NOEs that were used in the structure calculation. The sample used for this experiment contained equimolar amounts of $^{13}C/^{15}N$-labeled and unlabeled Swa_{DIMER}.

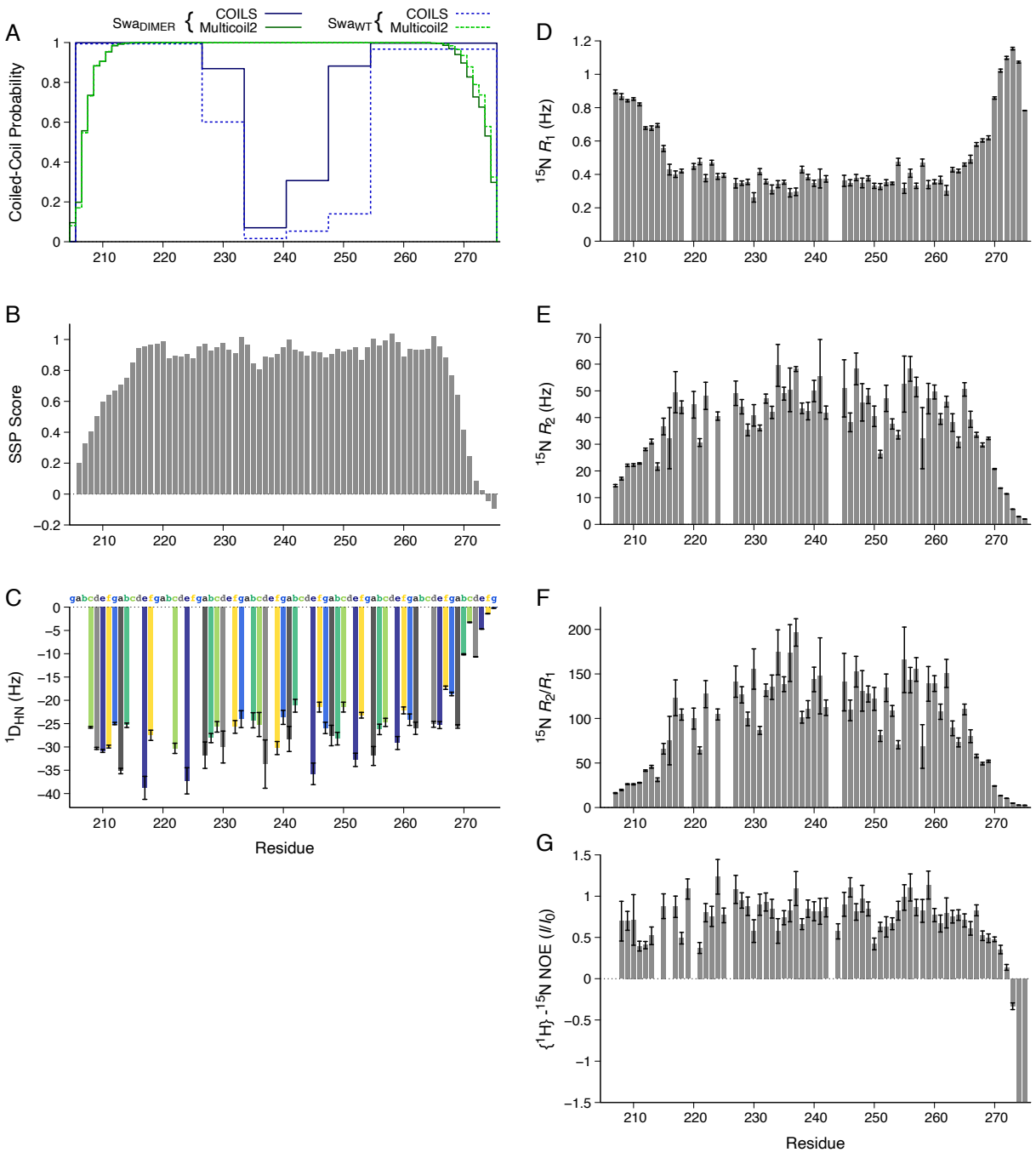

Figure S3: Coiled-coil predictions, SSP score, $^{1}D_{HN}$ RDC values, ^{15}N relaxation rates, and {^{1}H}-^{15}N NOE for Swa$_{DIMER}$.

(A) The COILS and Multicoil2 coiled-coil prediction algorithms were used to analyze the Swa$_{DIMER}$ sequence, as well as the same sequence without the four mutations (Swa$_{WT}$). For the COILS analysis a window size of 21 amino acids was used. (B) The secondary structure prediction (SSP) score was calculated using the $^{13}C_o$, $^{13}C_\alpha$, $^{13}C_\beta$, ^{15}N, $^{1}H_N$, and $^{1}H_\alpha$ chemical shifts as inputs. SSP values close to 1 reflect fully-formed α-helical regions, whereas values of zero indicate more disordered regions. (C) The residual dipolar couplings for the backbone amides ($^{1}D_{HN}$) of

Swa_{DIMER} were measured at 40 ºC using a variation of the ARTSY[10] experiment and a ^{15}N-labeled sample aligned with 10 mg/mL Pf1 phage. The coloring scheme indicates the position of the individual amino acids in the coiled-coil heptad repeat (**a** and **d** in greys, **e** and **g** in blues, **b** and **c** in greens, and **f** in yellow). Measurements of the ^{15}N longitudinal (D, R_1) and transverse (E, R_2) relaxation rates, as well as the ratio of these rates (F, R_2/R_1) and the $\{^1H\}$-^{15}N NOE (G), show that residues 215-270 form a single rigid structure without any flexible regions, indicating that the region where the coiled-coil bulges does not differ in dynamics from the rest of the coiled-coil. The R_2 values are much larger than what would typically be expected for a protein with the molecular weight of Swa_{DIMER} (17.1 kDa for the dimer) due to the combination of anisotropic tumbling and the nearly parallel orientation of the N-H bond vectors relative to the long axis of the structure. Values below –1.5 are truncated in the plot for the $\{^1H\}$-^{15}N NOE. Gaps in the charts are due to residues where spectral overlap and/or low signal intensity prevented accurate measurements.