# Supporting Information

## Lorieau et al. 10.1073/pnas.1213801109

### SI Materials and Methods.

**NMR Sample Preparation.** Serotype H1 HAfp23 and HAfp23-G8A fusion peptides, with the sequence GLFGAIA[G/A]FI EGGWTGMIDG WYGSGKKKKD, and the HAfp14 peptide, with the sequence of GLFGAIAGFIEGGWKKKKD, were expressed with a high-solubility tag (1) (underlined sequence) and purified as previously described (2). NMR samples with $^{15}$N; $^{13}$C,$^{15}$N; and $^{2}$H,$^{13}$C,$^{15}$N isotopic labeling scheme were prepared to final concentrations of 0.3–0.6 mM peptide in 130–150 mM perdeuterated dodecylphosphocholine (DPC; Anatrace), 93% $H_2O$/7% (vol/vol) $D_2O$ or 99.9% $D_2O$ [for measurement of 2D nuclear Overhauser effect spectroscopy (NOESY) spectra] and 25 mM $^{2}$H·Tris at pH 7.3 ± 0.1 (Cambridge Isotopes), and a final volume of 280 μL. For the pH 4 wild-type sample, the sample pH was adjusted using 50 mM $^{2}$H-citric acid.

For extracting the populations of closed and open states from chemical shifts, a scaled chemical shift, $\delta^*$, was used to account for the differences in range between different nuclei. Scaling factors of 1.00, 0.181, and 0.315 for $^{1}H^{N}$, $^{15}$N, and $^{13}C^{\alpha}$ nuclei were used, which are proportional to the inverse of the widths of their distributions in the protein chemical shift database.

Residual dipolar coupling (RDC) measurements were conducted using stretched acrylamide gel (SAG) as previously described (2–4) using 2-(acrylamido)-2-methyl-1-propanesulfonic acid doped acrylamide, and a cast diameter of 5.4 mm that was stretched axially to a final diameter of 4.2 mm.

**Lipid-Mixing Fusion Assay.** The lipid-mixing assay was conducted as previously described in the supporting information of ref. 5, with the following changes: the total lipid concentration used in the assay was 3 mM and the final protein concentration was 7 μM.

**NMR Data Collection and Analysis.** All experiments were conducted at 32 °C unless noted otherwise. Assignments of chemical shifts were made using Rance–Kay-detected $^{15}$N-heteronuclear single-quantum coherence (HSQC), HNCO, HNCA, and constant-time $^{13}$C-HSQC spectra, as well as 3D HNH-NOESY-HMQC and 2D NOESY experiments. The HNH-NOESY-HMQC experiments (100-ms NOE mixing time) were conducted at 600 MHz, and the $^{1}$H-$^{1}$H 2D NOESY experiments (70-ms NOE mixing time) were carried out at 900 MHz on $^{15}$N-labeled peptide in 99.8% $^{2}H_2O$ using a 10-Hz HDO presaturating radiofrequency field between transients.

NOE intensity to distance conversion was calibrated using fixed-geometry backbone and side-chain interproton pairs from the 12 N-terminal residues, which form a well-defined ideal α-helix based on $^{1}$H-$^{15}$N and $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ RDCs and secondary chemical shifts. NOEs involving only $AX/AX_2$ spin systems were grouped separately from NOEs with $AX_3$ spin systems. To account for the effect of spin diffusion, best-fit distance exponent values were derived (6), with values that ranged between −2.8 and −4.2 for the NOESY and NOESY-HMQC spectra.

The $^{1}D_{NH}$ and $^{1}D_{C\alpha H\alpha}$ RDCs in SAG-aligned samples were measured with 2D $^{1}$H-$^{15}$N IPAP-HSQC (7) and 3D $^{1}H^{\alpha}$-coupled HNCOCA spectra, respectively. An open-only HAfp23-G8A RDC dataset was generated by subtracting the contribution from the 15% closed helical hairpin conformation using the RDCs of the HAfp23 structure (2). This direct subtraction is possible because the HAfp23 and HAfp23-G8A RDCs were collected under the same sample and alignment conditions, the timescale of motion is much slower than the overall rotational correlation time, and the $^{2}H_2O$ residual quadrupole coupling was 2.5 ± 0.1 Hz for both HAfp23 and HAfp23-G8A datasets. These two sets of RDCs, for HAfp23 and the open-only dataset, were used as input to the structure calculations (see below).

The backbone $^{15}$N $R_1$, $R_2$, and {$^{1}$H}-$^{15}$N NOE relaxation rates were measured on a $^{13}$C,$^{15}$N-labeled peptide at 600 MHz and 900 MHz using a Rance–Kay detection scheme, as previously described (8, 9). A reduced spectral density analysis of the J(0) term (10), as determined from the 600-MHz data, is presented in Fig. S2.

The $T_{1\rho}$($^{1}$H)-unlike relaxation dispersion experiments (11) were conducted at 12 °C, and the relaxation dispersion curves were fit against Eq. 1, including an $R_1$ term for small $^{1}$H offsets and using a common value of $\tau_{ex}$ for all residues.

Hydrogen exchange rates were measured using a WEX-III pulse sequence with a transverse relaxation-optimized spectroscopy detection (12) scheme, and exchange delay periods of 50, 200, 400, 800, 1,100, and 1,400 ms. Hydrogen exchange rates were obtained from $K_{obs}$ rates (12) at pH 7.14 and 7.96 at 37 °C. The protection factors were derived relative to the intrinsic exchange rates calculated by the program SPHERE (13).

**Structure Calculations.** Structure calculations were performed with XPLOR-NIH version 2.31 using the variable-step internal variable module (IVM) algorithm (14–16). The helical hairpin structure (PDB ID code 2KXA) was used as the starting point for all structure calculations with Gly8 replaced by Ala. Structure calculations were conducted with a three-member ensemble. The initial temperature of 800 K was used, and the simulation was cooled down linearly to 30 K in 10-K temperature steps. Experimental restraints included Talos+ dihedral angles from chemical shifts (17), $^{1}$H-$^{1}$H NOE distance restraints as well as $^{1}$H-$^{15}$N and $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ RDCs. The nonexperimental energy terms were applied equally to all three conformers, and a quartic repulsive nonbonded potential was used with the atomic radii scaled from the van der Waals values multiplicatively, from 0.90 to 0.81.

The contribution of each conformer to the NOE and RDC observables must be factored into the refinement. Although the population of the closed conformer and the total population of the two open conformers together are known, the relative population of one open conformer to the other is not. This uncertainly does not impact the RDC term in the refinement because the alignment tensors, $D_a$, are floating variables during the simulated annealing procedure (18), and are proportional to the population-weighted alignment strength of the corresponding conformers (Table S1). Comparable $D_a$ values were obtained for the two open conformers (Table S1), suggesting populations for the two open conformers that are also comparable. Even though the treatment of NOEs in principle is affected by the assumed relative population of the two open states, due to the $r^{-6}$ dependence of the NOE on internuclear distance, even a fivefold deviation from equal population of the two open conformers would not impact the NOE distance-derived restraints beyond their intrinsic experimental uncertainty. An equal population of the two open conformers was therefore used for refinement.

Amide hydrogen bonds were included using the hydrogen-bonding database (HBDB) potential (19) with a default energy constant and a fixed list of α-helical hydrogen bonds for residues 1–11 and 15–23. Aliphatic hydrogen bonds between Gly1 and Trp21, Ala5 and Met17, Met17 and Ala5, and Phe9 and Gly13 were included for the closed conformer only, using a H—O upper-distance bound of 2.8 Å and a harmonic potential ramped multiplicatively from 0.2 to 20 kcal/Å$^2$.

The Talos+ dihedral restraints were applied with a 300 kcal/(mol·rad$^2$) energy constant. NOE restraint force constants were ramped multiplicatively from 0.2 to 20 kcal/Å$^2$ using a soft-square potential. NOEs were refined against a population-weighted average of the NOE rate ($r^{-6}$ averaging) for each conformer in the refinement, and a population weighting of 15%, 42.5%, and 42.5% was used for the closed and two open conformations, respectively. No consistent NOE violations greater than 0.35 Å were found. A total of 316 $^1$H-$^1$H NOE restraints were used, including 202 short-range, 92 medium-range, and 22 long-range restraints.

The published wild-type RDCs (2) were used to refine the closed helical hairpin conformer, and the open-only RDC dataset was used to refine the two open conformers together. Refinement against the RDCs was accomplished using a harmonic $^1D_{NH}$ potential with a force constant ramped multiplicatively from $1.5{\cdot}10^{-5}$ to 0.045 kcal/Hz$^2$ for the closed conformer and from $4.5{\cdot}10^{-5}$ to 0.1275 kcal/Hz$^2$ for each of the two open conformers. Force constants for $^1D_{C\alpha H\alpha}$ were scaled relative to $^1D_{NH}$ by a factor of 0.2. All of the alignment tensor's parameters ($D_a$, rhombicity, and orientation) were allowed to float during refinement (16).

The $Q$-factor was estimated using the following equation for multiple conformations, which accounts for the powder average of the dipolar coupling second-moment for multiple conformations when the alignment tensor orientations are not correlated between conformers:

$$Q = \sqrt{\frac{\sum_{i}^{RDC}\left(\left(\sum_{j}^{states} D_{(i,j),pred}\right) - D_{i,obs}\right)^2}{N_{RDC}\sum_{j}^{states}\left(D_{a,j}^2\left(4+3R_j^2\right)/5\right)}}. \qquad [S1]$$

The $D_{a,j}$, $R_j$, and $D_{(i,j),pred}$ are determined from a multiconformer singular-value decomposition fit (20), and the fitted $D_{a,j}$ parameter encompasses the strength of the alignment scaled by the population of the $j$th conformer. Free $Q$-factors were estimated with a jack-knifing procedure that iteratively discards a single dipolar coupling from a refinement, followed by back-calculating this coupling's value from the refined structures and comparing it to the observed value.

A total of 480 sets of three structures were calculated, and the 33 sets with lowest energy were retained for analysis, with structural statistics reported in Table S1.


1. Han X, Tamm LK (2000) A host-guest system to study structure-function relationships of membrane fusion peptides. *Proc Natl Acad Sci USA* 97(24):13097–13102.
2. Lorieau JL, Louis JM, Bax A (2010) The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface. *Proc Natl Acad Sci USA* 107(25):11341–11346.
3. Tycko R, Blanco FJ, Ishii Y (2000) Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *J Am Chem Soc* 122:9340–9341.
4. Meier S, Häussinger D, Grzesiek S (2002) Charged acrylamide copolymer gels as media for weak alignment. *J Biomol NMR* 24(4):351–356.
5. Lorieau JL, Louis JM, Bax A (2011) Whole-body rocking motion of a fusion peptide in lipid bilayers from size-dispersed 15N NMR relaxation. *J Am Chem Soc* 133(36): 14184–14187.
6. Güntert P, Braun W, Wüthrich K (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol* 217(3):517–530.
7. Yao LS, Ying JF, Bax A (2009) Improved accuracy of 15N-1H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J Biomol NMR* 43(3):161–170.
8. Mandel AM, Akke M, Palmer AG, 3rd (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme. *J Mol Biol* 246(1):144–163.
9. Cole R, Loria JP (2003) FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data. *J Biomol NMR* 26(3):203–213.
10. Farrow NA, Zhang OW, Szabo A, Torchia DA, Kay LE (1995) Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR* 6(2):153–162.
11. Ishima R, Louis JM, Torchia DA (1999) Transverse 1H cross relaxation in 1H-15N correlated 1H CPMG experiments. *J Magn Reson* 137(1):289–292.
12. Fitzkee NC, Torchia DA, Bax A (2011) Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain. *Protein Sci* 20(3):500–512.
13. Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17(1):75–86.
14. Schwieters CD, Clore GM (2001) Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J Magn Reson* 152(2):288–302.
15. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73.
16. Clore GM, Schwieters CD (2004) How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *J Am Chem Soc* 126(9):2923–2938.
17. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4):213–223.
18. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2001) An easy way to include weak alignment constraints into NMR structure calculations. *J Biomol NMR* 21(3):275–280.
19. Grishaev A, Bax A (2004) An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *J Am Chem Soc* 126(23):7281–7292.
20. Losonczi JA, Andrec M, Fischer MWF, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* 138(2): 334–342.

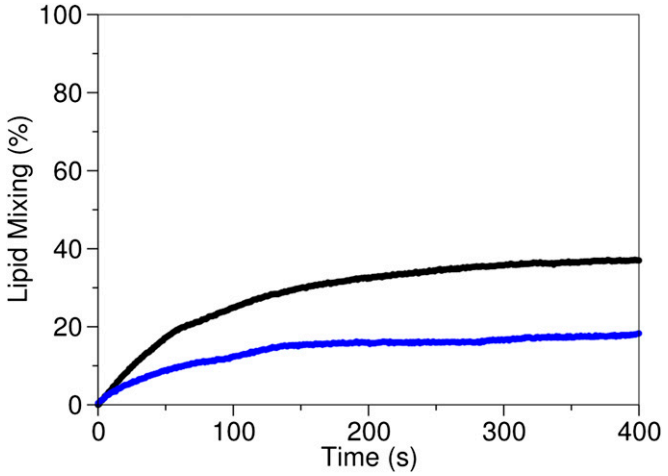


**Fig. S1.** Lipid-mixing fusion activity of the wild-type HAfp23 (black) and mutant G8A (blue) fusion peptides. The fluorescence scale has been calibrated such that the zero level corresponds to the initial residual fluorescence of the labeled vesicles and the 100% value to complete mixing of all of the lipids in the system (obtained by adding reduced Triton X-100 to a final concentration of 50 mM).

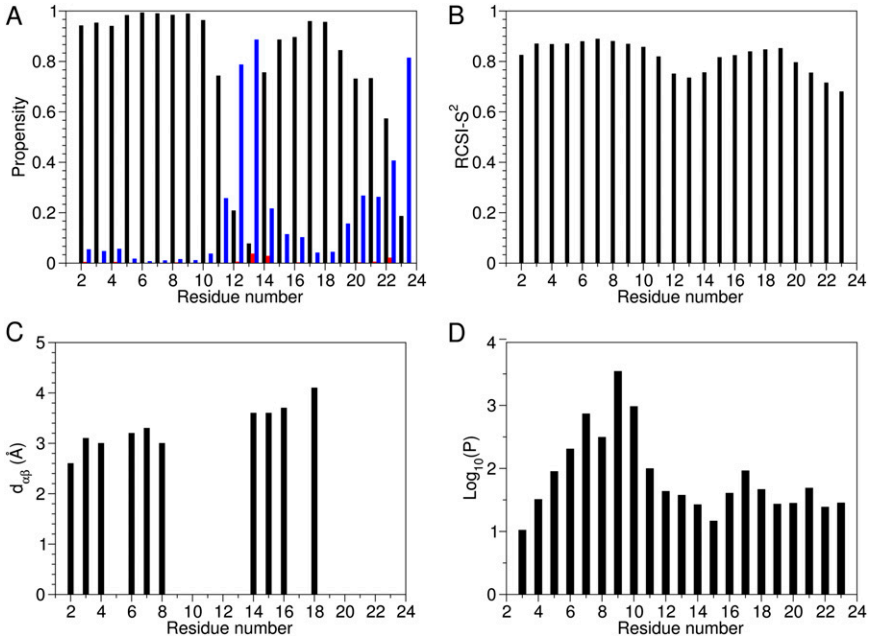


**Fig. S2.** The HAfp23-G8A peptide has a helical secondary structure from residues 2–11 and 14–22. (*A*) The TALOS+ secondary structure predictions (1) are shown for α-helical (black), β-sheet (red), and coil (blue). These are calculated based on the $^{1}H^{N}$, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$ $^{13}C^{\beta}$, and $^{15}N$ secondary chemical shifts. (*B*) RCSI-$S^2$ predictions (2) derived from the secondary chemical shifts. A high RCI-$S^2$ (>0.75) represents a well-ordered backbone for residues in the N-terminal (L2–E11) and C-terminal (Trp14–Trp21) helices. (*C*) The α(i)-β(i + 3) NOE patterns are consistent with an α-helical structure. Some α(i)-β(i + 3) NOEs are missing because they either are not observable (Gly1–Gly4, Phe9–Gly12, Met17–Gly20, and Gly20–Gly23 do not have an $H^{\beta}$ proton on the i + 3th residue), are in a congested region of the spectrum (Ala5–Ala8), or are masked by $t_1$ noise (Ile18–Trp21). (*D*) HAfp23-G8A $H_2O$ protection factors for backbone amide protons.

1. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4):213–223.
2. Berjanskii M, Wishart DS (2006) NMR: Prediction of protein flexibility. *Nat Protoc* 1(2):683–688.

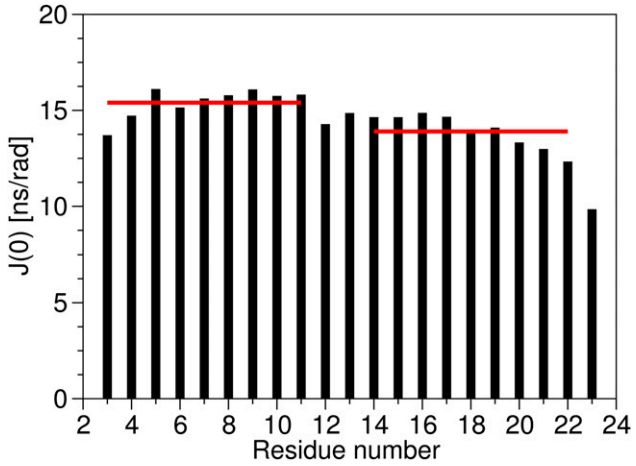


**Fig. S3.** The J(0) spectral density term derived from a reduced spectral density analysis (1) of the HAfp23-G8A peptide using $^{15}$N relaxation rates collected at 600 MHz. The J(0) spectral densities (black bars) are shown together with the N-terminal and C-terminal averages (red bars). The N-terminal residues (Phe3–Glu11) have a J(0) spectral density of 15.4 ns/rad, on average, and the C-terminal residues (Trp14–Tyr22) have a J(0) spectral density of 13.9 ns/rad, on average, corresponding to an additional motion on the ~100-ns time scale. This analysis makes use of a motional order parameter $S^2$ of 0.137, estimated from the RDC $D_a$ ratio of the N- and C-terminal helices in Fig. 3. A $^{15}$N chemical shift anisotropy of −173 ppm and a $^1$H–$^{15}$N bond length of 1.04 Å, to account for sub-picosecond librational motion, were used in the reduced spectral density analysis.

1. Farrow NA, Zhang OW, Szabo A, Torchia DA, Kay LE (1995) Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR* 6(2):153–162.

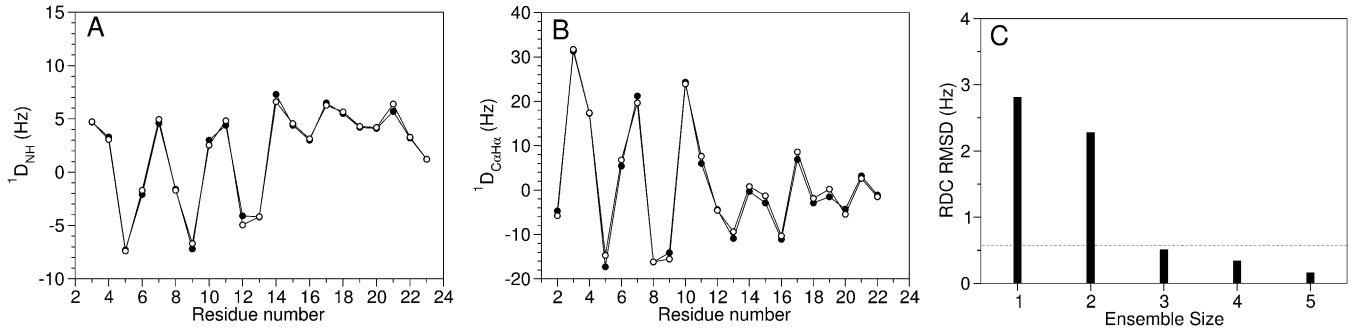


**Fig. S4.** Fits of the open-only RDC dataset for the mutant HAfp23-G8A peptide to the lowest-energy, RDC-refined open structures. (*A*) $^1D_{NH}$ and (*B*) $^1D_{C\alpha H\alpha}$ experimental RDCs (●) from the open-only G8A dataset (see text) are plotted together with the best-fitted RDCs for the open-only conformations of a three-conformer refinement (○). The singular value decomposition (SVD) fit was carried out for the L-shaped and extended conformers, producing a *Q*-factor of 3.6%. Experimental errors in the RDC measurements are 0.4 Hz for $^1D_{NH}$ and 1.6 Hz for $^1D_{C\alpha H\alpha}$, i.e., about the size of the symbols. (*C*) Normalized (relative to $^1D_{NH}$) residual rmsd of the RDC SVD fits for the multiconformer ensembles of varying size. The one-member "ensemble" did not include a closed helical hairpin structure or aliphatic hydrogen-bonding distance restraints, and the structures were refined directly against the HAfp23-G8A RDCs. Other ensembles included a fixed, 15% population of hairpin conformers, and equal populations for the remaining conformers. The dashed line corresponds to the uncertainty in the RDC measurement.

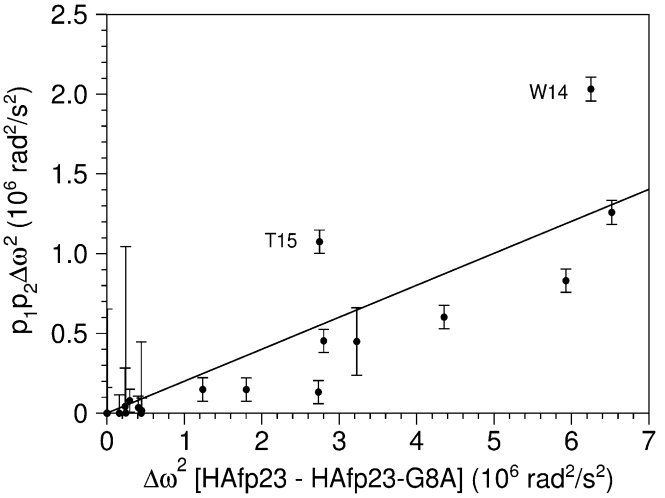


**Fig. S5.** Estimation of the population for open conformers from the fit parameters of the $^1$H relaxation dispersion curves for wild-type HAfp23 at pH 4 and 12 °C. The values for $\Delta\omega^2$ are taken from the differences in chemical shift between HAfp23 and HAfp23-G8A (at pH 7 and 12 °C), and are used to model the chemical shifts of the closed and open states, respectively. However, because the HAfp23-G8A peptide is only 85% open (see text), this difference has been scaled by $0.85^{-1}$. The best-fit slope, $p_1p_2 = 0.20$, yields a population of 28 ± 8% for the open conformers; the best-fit slope from the fit without W14 and T15, $p_1p_2 = 0.16$, yields a population of 20 ± 5% for the open conformers. An alternative method for calculating this population is presented in Fig. S5.

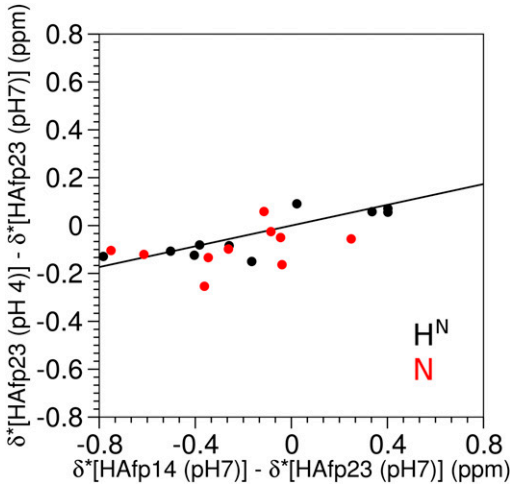


**Fig. S6.** Populations of the open and closed conformations for HAfp23 at pH 4 and 12 °C from chemical shifts. The changes in backbone chemical shifts between HAfp23 (closed) and HAfp14 (open) at pH 7 and HAfp23 at pH 4 (conformational mixtures in the fast exchange limit) show a linear correlation with a slope of 0.21, corresponding to a population of 21 ± 5% for the open conformation(s) of HAfp23 at pH 4 and 12 °C. The chemical shift differences have been scaled by the following factors derived from database SDs in secondary chemical shifts: N = 0.181, $H^N$ = 1.373. Chemical shifts from all residues between Phe3 to Gly13 have been included in the fit, excluding residue 8, the site of mutation, and Glu11, which becomes protonated at pH 5.3 (1).

1. Lorieau JL, Louis JM, Bax A (2011) Helical hairpin structure of influenza hemagglutinin fusion peptide stabilized by charge-dipole interactions between the N-terminal amino group and the second helix. *J Am Chem Soc* 133(9):2824–2827.

**Table S1. Structural statistics for the 33 lowest-energy three-conformer sets of structures of HAfp23-G8A at pH 7**

| Simulation property | Value |
|---|---|
| Population of the closed conformer | 15% |
| Population of the open extended conformer | 42.5% |
| Population of the open L-shaped conformer | 42.5% |
| $Q_{work}$, closed[†] | 9.7% |
| $Q_{work}$, open[‡] | 3.6% |
| $Q_{free}$, open | 11.8% |
| RDC tensor parameters | |
| $D_a$(NH)/R, closed | 8.2 Hz/0.09 |
| $D_a$(NH)/R, open, L-shaped conformer | 8.1 Hz/0.62 |
| $D_a$(NH)/R, open, extended conformer | −12.0 Hz/0.36 |
| Rmsd from experimental restraints (no. of restraints) | |
| $^1D^{NH}$ (21), closed | 0.42 ± 0.06 Hz |
| $^1D^{C\alpha H\alpha}$ (25), closed[§] | 1.9 ± 0.2 Hz |
| $^1D^{NH}$ (21), open | 0.42 ± 0.11 Hz |
| $^1D^{C\alpha H\alpha}$ (21), open | 1.5 ± 0.3 Hz |
| Dihedral angles (40) | 0.29 ± 0.16° |
| Short-range NOE (202) | 0.044 ± 0.005 Å |
| Medium-range NOE (92) | 0.049 ± 0.003 Å |
| Long-range NOE (22) | 0.048 ± 0.001 Å |
| Deviations from idealized geometry | |
| Bonds | 0.0050 ± 0.0007 Å |
| Angles | 0.363 ± 0.007° |
| Impropers | 0.196 ± 0.01° |
| Energy, kcal/mol | |
| Total | −78 ± 6 |
| Bond | 9 ± 2 |
| Angle | 61 ± 2 |
| Improper | 6.3 ± 1.0 |
| van der Waals | 30 ± 5 |
| Ramachandran | −189 ± 4 |
| HBDB[¶] | −39 ± 4 |
| Dipolar | 35 ± 4 |
| NOE | 4.4 ± 4.0 |
| Dihedral | 0.62 ± 0.89 |
| Ramachandran statistics[‖] | |
| $\phi/\psi$ in most favored regions, closed conformer | 93.8% |
| $\phi/\psi$ in most favored regions, open L-shaped conformer | 100% |
| $\phi/\psi$ in most favored regions, closed extended conformer | 100% |
| $\phi/\psi$ in additionally allowed regions, closed conformer | 6.2% |
| $\phi/\psi$ in additionally allowed regions, open extended conformer | 0% |
| $\phi/\psi$ in additionally allowed regions, open L-shaped conformer | 0% |
| Atomic rmsd** | |
| Backbone heavy atom (residues 3–22), closed conformer | 0.16 ± 0.05 Å |
| Backbone heavy atom (residues 3–22), L-shaped conformer | 0.52 ± 0.17 Å |
| Backbone heavy atom (residues 3–22), extended conformer | 0.66 ± 0.29 Å |
| Structure quality factors[††] | |
| MolProbity Z-score, closed conformer | 0.50 |
| MolProbity Z-score, open L-shaped conformer | −0.52 |
| MolProbity Z-score, open extended conformer | 0.50 |

[†]RDC dataset for closed conformer, taken from the wild-type SAG-aligned sample. *Q*-factor calculated for residues 2–23, not including residue 8. See Lorieau et al. (1).
[‡]RDC dataset for open conformers, generated from the HAfp23-G8A dataset (see text for details.) *Q*-factor calculated for residues 2–23 using Eq. **S1**.
[§]Includes stereospecific RDCs $^1H^{\alpha}$(pro-S)$^{13}C$ and $^1H^{\alpha}$(pro-R)$^{13}C$ for Gly1, Gly12, and Gly13.
[¶]HBDB, the empirical hydrogen bond potential for XPLOR-NIH (2).
[‖]Evaluated using PROCHECK (3).
**Rmsd calculated from the average structure for each conformer.
[††]Evaluated using PSVS (4).


1. Lorieau JL, Louis JM, Bax A (2010) The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:water interface. *Proc Natl Acad Sci USA* 107(25):11341–11346.
2. Grishaev A, Bax A (2004) An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *J Am Chem Soc* 126 (23):7281–7292.
3. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17(4):355–362.
4. Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* 66(4):778–795.