PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Very broad distribution of β sheet registries of the HIV gp41 fusion peptide supports mutational robustness for fusion and infection

Yijin Zhang[a,1,2], Scott D. Schmick[a,1,3], Li Xie[a], Ujjayini Ghosh[a,4], and David P. Weliky[a,5]

Affiliations are included on p. 10.

Edited by Ann McDermott, Columbia University, New York, NY; received February 14, 2024; accepted October 16, 2024

HIV, like other membrane-enveloped viruses, has protein spikes that include a fusion peptide (Fp) segment that binds the host cell membrane and plays a critical role in fusion (joining) viral and cell membranes. The HIV Fp is the ~23 N-terminal residues of the gp41 spike protein. Fp adopts intermolecular antiparallel β sheet structure when lipid fraction cholesterol ≈0.3, which is comparable to host cells. Rotational-echo double-resonance NMR was applied to probe the registries (alignments) of adjacent Fp molecules in membrane-bound sheets. The data were fitted to determine quantitative populations, f(t)'s, of individual antiparallel registries indexed by t, the number of residues in the registry. Both wild-type (WT) and fusion-defective V2E Fp sheets have broad but very different registry distributions, each with at least eight populated registries with f(t) > 0.02, and $\langle t \rangle^{WT} = 16.1$ and $\langle t \rangle^{V2E} = 18.5$. The broad WT distribution likely improves mutational robustness for HIV, as Fp is a neutralization epitope of the immune system, and Fp mutations are required for immune evasion during chronic HIV infection. V2E fusion is reduced because longer Fp sheets increase separation between initially apposed membranes. The $f(t)^{WT}$ were well-fitted to free energies that were sums of contributions from sheet length, aligned leucines, and sidechain membrane insertion. The $f(t)^{V2E}$'s were similarly well-fitted except there wasn't the insertion contribution. Relative to V2E, WT fusion is enhanced by deeper membrane insertion of Fp with accompanying greater dislocation of neighboring lipids. This study provides a rare quantitative determination of broad molecular structural distributions by experiment.

HIV | NMR | beta sheet | fusion peptide | distribution

Significance

This study provides the experimentally-derived quantitative populations of antiparallel β sheet registries of both wild-type (WT) and dominant fusion-defective Valine-2 to Glutamate-2 (V2E) mutant of the membrane-bound HIV gp41 fusion peptide. The peptide plays a critical role in HIV fusion and infection and is a neutralization epitope of the immune system. Each distribution contains at least eight registries, with quite different distributions for WT vs. V2E and weighting towards shorter vs. longer registries. The breadth of the WT distribution likely improves mutational robustness for HIV, a chronic infection. Fusion loss for V2E is likely due to larger initial membrane apposition distance and shallower membrane location. This study provides a rare quantitative determination of broad molecular structural distributions by experiment.

Membrane-enveloped viruses are a large group that includes many families including HIV, influenza, and coronaviruses (1–5). Cellular infection for these viruses requires fusion (joining) the viral and cellular membranes, and depending on family, the latter is the plasma and/or an endosomal membrane. The fusion rate is typically negligible in the absence of catalyst, so each virus family has protein spikes that protrude from the viral membrane and catalyze fusion. There is homology in the spike sequence within a virus family but not between families. For "class I" viruses like HIV, each spike has three glycoproteins and each glycoprotein is a receptor-binding subunit and a fusion subunit. For HIV, the glycoprotein 160 kD (gp160) is cleaved into the gp120 receptor-binding and gp41 fusion subunits, with ~510 and ~350 residues, respectively. Gp41 has a ~170-residue ectodomain outside the virus followed by a transmembrane domain (Tmd) and a ~150-residue endodomain in the virus interior, Fig. 1 (6–12). The spike protruding from the virus has a core formed by the three gp41 ectodomains and the three gp120 subunits bound noncovalently to this core (13, 14). Target T and macrophage cells are identified by gp120 binding to CD4 and chemokine receptors followed by gp120 movement away from gp41 (2). The gp41 residues ~25 to 160 then spontaneously transform to a different and thermostable trimer-of-hairpins structure, Fig. 1 (i–iv) (7–10). Each hairpin has ~60-residue N-helix and C-helix segments separated by a loop. The N-helices from three gp41's form an interior parallel coiled-coil and the C-helices are antiparallel and bound to the exterior grooves of the N-helix coil.

The ~23 N-terminal residues of gp41 are not part of the final hairpin and are named the "fusion peptide" (Fp) (6). The Fp sequence is fairly well-conserved among HIV isolates, with some variability (15). The Fp is the epitope of some broadly neutralizing antibodies in infected individuals and also a HIV vaccine target (16, 17). The Fp in the absence of the rest of gp160 binds membrane and has been commonly proposed to bind the target membrane during fusion, Fig. 1 (1, 18). Such binding could be important in overcoming activation energy barriers between different membrane structures during fusion. For

[1]Y.Z. and S.D.S. contributed equally to this work.

[2]Present address: Interdisciplinary Institute of Nuclear Magnetic Resonance (NMR) and Molecular Sciences, School of Chemistry and Chemical Engineering, The State Key Laboratory of Refractories and Metallurgy, Wuhan University of Science and Technology, Wuhan 430081, People's Republic of China.

[3]Present address: Specialty Materials Development Laboratory, Corporate Research and Development, 3M Company, Saint Paul, MN 55144.

[4]Present address: Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105.

[5]To whom correspondence may be addressed. Email: weliky@chemistry.msu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2402953121/-/DCSupplemental.

Published December 2, 2024.

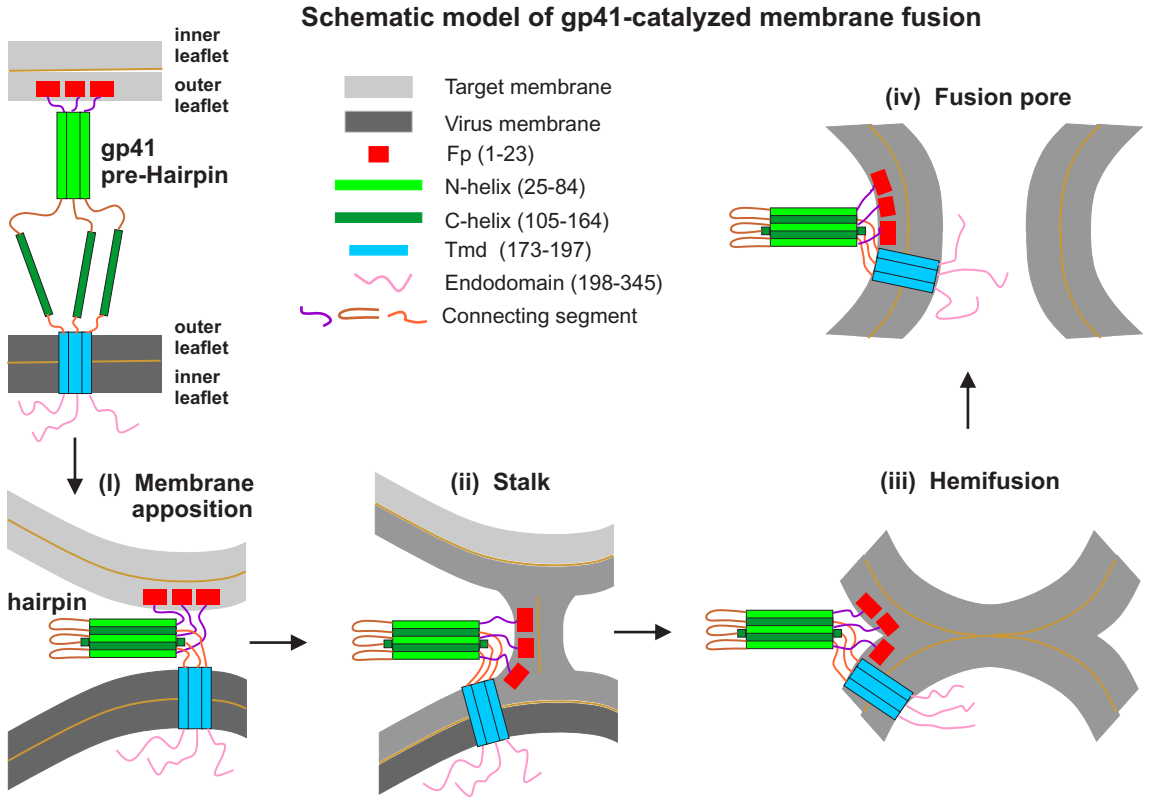

Fig. 1. Schematic model for changes in gp41 and membrane structures during fusion. The *Top-Left* prehairpin structure forms after the gp120 subunits move away from the gp41 trimer. The endodomain forms a well-defined structure that is not displayed in this figure.

example, close apposition between viral and target membranes is likely the initial step in fusion and requires ~25 kcal/mole of free energy, Fig. 1 (i) (2, 4, 19). Some of this free energy could be provided through the combination of Fp in target membrane, Tmd in viral membrane, and intervening thermostable trimer-of-hairpins. There are ~10 kcal/mole barriers to form subsequent membrane structures during fusion, Fig. 1 (ii–iv), in part because these membrane transformations require large-amplitude motions of acyl chains of lipids (4).

In detergent-rich media, Fp is a monomer and residues 2 to 22 are a continuous single helix (20). In membrane without cholesterol, there are two Fp populations with distinct structures. One structure is the monomer helix observed in detergent and the other is an intermolecular β sheet oligomer with each Fp as a strand in the sheet (18, 21–23). There is a positive correlation between the mole fraction cholesterol in the membrane and the β sheet oligomer population, with >90% sheet population when the cholesterol fraction is ~0.3, which is typical for the plasma membrane of host cells of HIV (23–25). NMR data have shown the β sheets to be >90% antiparallel, with a relatively small number (~10) of molecules in the sheet (26, 27). NMR spectra of a large membrane-bound gp41 construct with Fp and hairpin regions supported Fp with predominant sheet rather than helix structure (28). Other NMR data evidenced predominant antiparallel vs. parallel Fp sheet (29).

The present study addresses the distribution of registries (residue alignments) of adjacent antiparallel Fp molecules in the membrane-bound β sheets, as well as the important role of this distribution in HIV fusion and infection. A convenient index for a registry is t, the total number of residues in one strand that could be hydrogen-bonded to residues in the neighboring strand, starting from the N termini. Fig. 2*A* shows a schematic of the t = 16 registry for a Fp with ^{13}CO label at L12 and ^{15}N label at G5. There have been limited experimental data about the Fp registry distribution. NMR studies of a few Fp samples with selective ^{13}C,^{15}N labelings supported some population of the t = 16 and t = 17 registries, with large populations in other unknown registries (22, 27). The present study provides a complete and accurate determination of all the registry populations using global analysis of NMR data from many differently labeled samples. Our study also provides the determination of populations for the V2E Fp mutant. Earlier work showed that V2E caused complete loss of gp160-mediated fusion and HIV infection (30). V2E is phenotypically dominant for mixed WT:V2E gp160 trimers and functional loss vs. WT:V2E ratio has been the main data to understand intertrimer cooperativity in fusion (31). The wild-type (WT) vs. V2E registry distributions determined in the present study are important information to understand the bases of the fusion -active vs. -defective functions. Samples for the present study contain Fp without the rest of gp160 but the relevance of our study for understanding the large functional difference in V2E vs. WT gp160 is supported by earlier observation of ~10× less vesicle fusion after exogenous addition of V2E vs. WT Fp (32). Mixtures of WT and V2E Fp's do not exhibit V2E-dominant loss of vesicle fusion, but V2E-dominant loss was observed for a large Fp+hairpin construct that included most of the gp41 ectodomain and adopted trimer-of-hairpins structure (33, 34). The V2E mutant also exhibited ~15% loss in helicity. There were very similar dependences of V2E-dominant losses vs. WT:V2E ratio for HIV gp160 fusion and infection, and Fp+hairpin-induced fusion and helicity. Global fitting of all the data supported a requirement of at least two WT trimers for efficient fusion and infection.

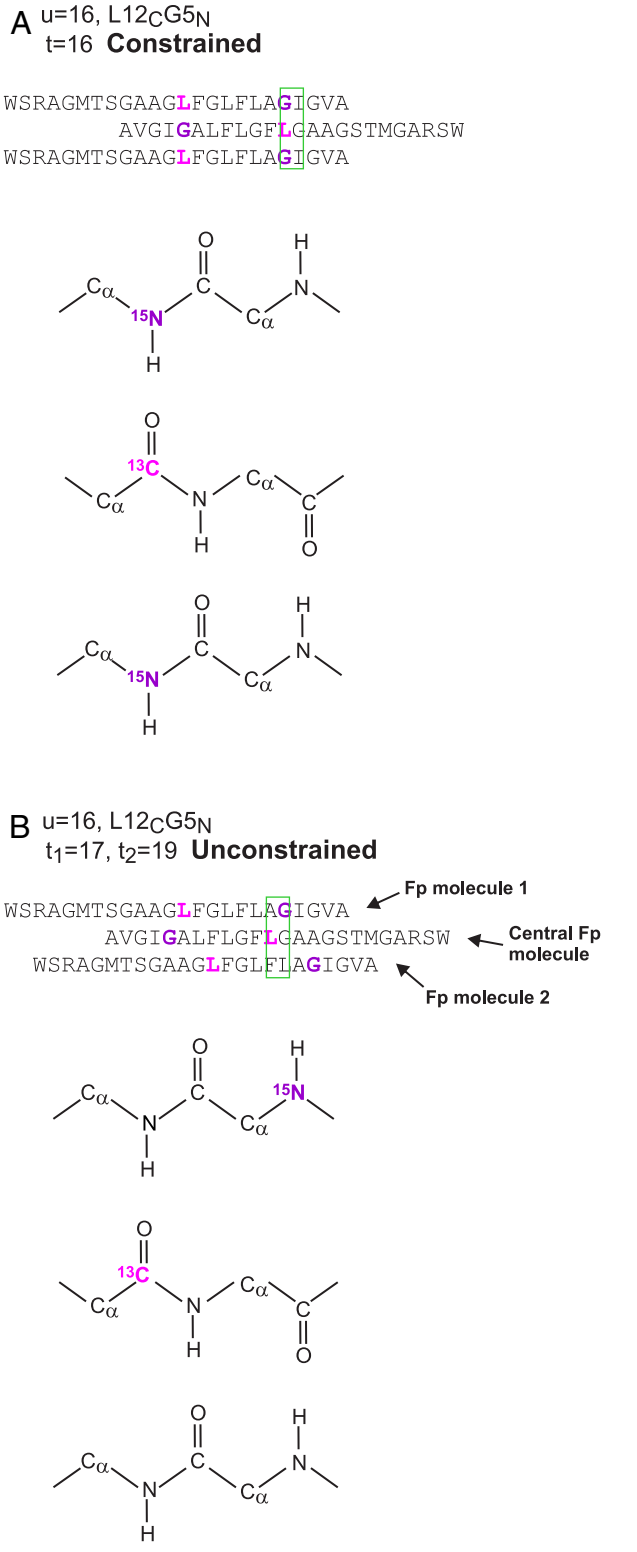

Fig. 2. Schematic representations of antiparallel β sheet registries and $^{13}CO/^{15}N$ spin systems of the WT Fp. The Fp labeling is u = 16 ($L12_CG5_N$), the L12/^{13}CO are magenta, and the G5/^{15}N are purple. Panel *A* displays a constrained sheet with t = 16 and panel *B* displays an unconstrained sheet with t_1 = 17 and t_2 = 19. The backbone geometry in each panel is for the green rectangle region in the schematic sheet. The t parameter is the total number of residues in one strand that could be hydrogen-bonded to residues in the neighboring strand, starting from the N termini. The u parameter is the value of t that aligns the ^{13}CO-labeled residue on one strand and the ^{15}N-labeled residue on the neighboring strand, and is therefore labeling-dependent. The t_1 and t_2 in the unconstrained sheet are the total number of residues of molecule 1 or 2, respectively, that could be hydrogen-bonded to the central molecule. There are more than three molecules in a Fp β sheet.

Results

Samples and NMR Spectra. WT Fp peptides had sequence AVGIGALFLGFLGAAGSTMGARSWKKKKKKA, with underlining for the 23 N-terminal residues of HIV gp41, HXB2 laboratory strain, followed by a nonnative W as a A280 chromophore, and a polylysine tag so that Fp is monomeric in aqueous solution prior to membrane binding (35). Comparison of Fp-induced vesicle fusion with vs. without tag shows both similar extents of vesicle fusion for WT sequence and similar reductions for V2E mutant (36–38). The membrane composition was DTPC:DTPG:cholesterol with 8:2:5 molar ratio where DTPC is 1,2-di-*O*-tetradecyl-*sn*-glycero-3-phosphocholine and DTPG is 1,2-di-*O*-tetradecyl-*sn*-glycero-3-[phospo-*rac*-(1-glycerol)]. The plasma membranes of HIV host cells contain similar fractions of negatively charged lipid and cholesterol and significant phosphatidylcholine lipid (25). The bound Fp:lipid mole ratio was ~1:60 and the lipids were in bilayer phase (39, 40). This molar ratio is substantially larger than the bulk gp160:lipid molar ratio calculated from the number of gp160 spikes in a virion (41). However, each spike contains three gp160 molecules and spikes cluster together in the mature virion which leads to much higher local gp160:lipid molar ratio at the fusion site (42, 43). A Fp had one residue with a backbone ^{13}CO label and a different residue with a backbone ^{15}N label. Each labeled Fp sample was indexed by an integer u which is the registry t value that aligns the ^{13}CO-labeled residue on one strand and the ^{15}N-labeled residue on the neighboring strand. The u parameter is labeling-dependent. Fig. 2*A* displays a schematic antiparallel sheet with t = 16 and $L12_CG5_N$ labeling which corresponds to u = 16. There were individual samples with WT and V2E Fp's for all values of u in the 8 to 24 range as well as WT Fp with u = 28, i.e., 18 total WT and 17 total V2E samples.

There was acquisition of S_0 and S_1 data from ^{13}C-^{15}N rotational-echo double-resonance (REDOR) NMR experiments, which yielded ^{13}C spectra that respectively were not or were attenuated by dipolar coupling to nearby ^{15}N nuclei (44–46). Data were acquired for dephasing times τ_k = 2.2, 8.2, 16.2, 24.2, 32.2, 40.2, and 48.2 ms with k = 1, 2, 3, 4, 5, 6, and 7, respectively. *SI Appendix*, Fig. S1 displays plots of the ^{13}CO regions of the S_0 and S_1 spectra at τ_k = 40.2 ms and *SI Appendix*, Fig. S2 displays S_0 and $\Delta S = S_0 - S_1$ spectra for u = 17, 20 samples. The S_0 spectra include both labeled (lb) and natural abundance (na) ^{13}CO contributions whereas ΔS spectra are predominantly lb ^{13}CO signals. Each S_0 and ΔS lineshape is well-fitted by a single Gaussian function and *SI Appendix*, Table S1 lists the fitted peak chemical shifts, δ_{peak}'s, and full-width at half-maximum linewidths, δ_{FWHM}'s, with labeled ^{13}CO sites at A6, L7, F8, L9, and L12. For both WT and V2E Fp's, the δ_{peak} values correlate with β sheet structure, and the typical δ_{FWHM} is between 3 and 4 ppm (47, 48). For a particular sample, the $\delta_{peak,S0}$ and $\delta_{peak,\Delta S}$ typically agree within 0.3 ppm and the $\delta_{FWHM,S0}$ is usually larger than $\delta_{FWHM,\Delta S}$ by 0.2 to 0.5 ppm. The F8, L9, and L12 sites are ^{13}CO labeled in multiple samples and for each site, the average values of $\delta_{peak,S0}$, $\delta_{peak,\Delta S}$, $\delta_{FWHM,S0}$, and $\delta_{FWHM,\Delta S}$ are very similar among samples with typical rmsd <0.3 ppm. This similarity demonstrates reproducibility of sample preparation. The ^{13}CO spectral integrals are denoted $S_0(u,\tau_k)$ and $S_1(u,\tau_k)$ with accompanying uncertainties, $\sigma(u,\tau_k)$. Fig. 3 and *SI Appendix*, Table S2 show the buildups of experimental dephasing vs. τ_k where dephasing is $\Delta S/S_0 = 1 - S_1(u,\tau_k)/S_0(u,\tau_k)$. Dephasing is only observed for the isotropic ^{13}CO signals, as shown in representative full ^{13}C S_0 and S_1 spectra of the V2E u = 16 sample, *SI Appendix*, Fig. S3. *SI Appendix*, Table S3 shows very similar $\Delta S/S_0$ for replicate samples, which demonstrates experimental reproducibility.

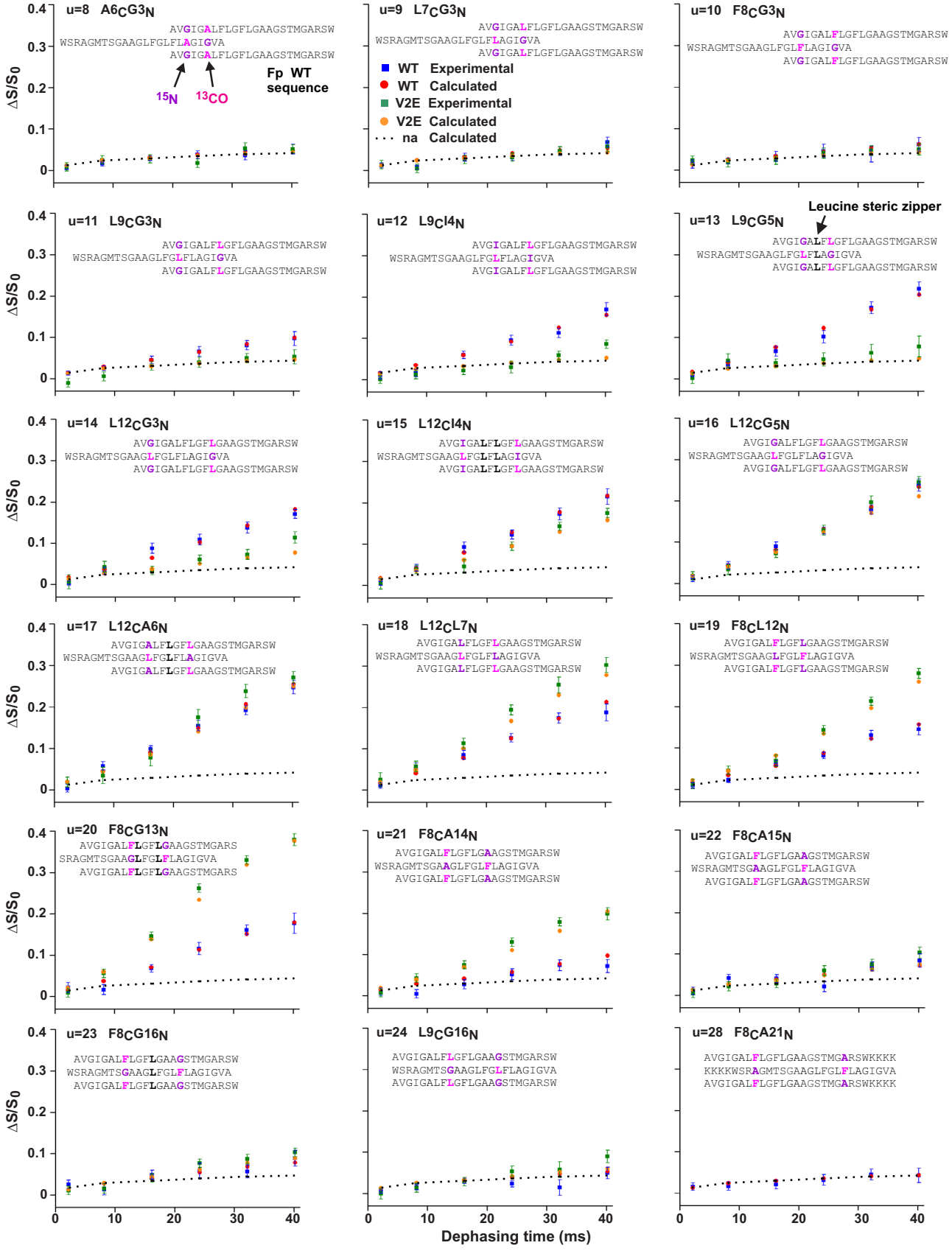

Fig. 3. Plots of ^{13}CO-^{15}N REDOR NMR $\Delta S/S_0$ vs. dephasing time (τ_k) for membrane-bound Fp, WT, and V2E, u = 8 to 24, 28, and τ_k=2.2 to 40.2 ms (k = 1 to 6). Experimental and calculated $\Delta S/S_0$ are displayed with symbols and the dotted line connects the natural abundance $\Delta S/S_0$ calculated with Eq. **1**. The $\Delta S/S_0$ in the u = 8 to 24 samples larger than calculated na $\Delta S/S_0$ values have a contribution from dipolar dephasing of labeled ^{13}CO signals by nearby labeled ^{15}N spins. Each panel displays schematic constrained WT sheets with t = u and ^{13}CO- and ^{15}N- labeled residues bolded in magenta and purple, respectively. Aligned leucines in the β sheet are also bolded. For u = 8 to 24, calculated $\Delta S/S_0$ are based on unconstrained fitting with b = 0.98 scaling factor for $\gamma_{t1,t2}^{lb,lb}(\tau_k)$, Eq. **3** and *SI Appendix*, Table S4, and for u = 28, natural abundance dephasing, Eq. **1**. The t, t_1, t_2, and u parameters are defined in the main text and Fig. 2 legend. The experimental and calculated $\Delta S/S_0$ numerical values are presented in *SI Appendix*, Tables S6 and S7.

Determination of f(t) Populations by Constrained and Unconstrained Fittings. The WT u = 28 data are representative of na dephasing, e.g. $\Delta S/S_0 = 0.043$ for $\tau_k = 40.2$ ms. Relative to u = 28, Fig. 3 shows that $\Delta S/S_0$ are larger in many samples with other values of u, and these larger $\Delta S/S_0$ are due to dipolar dephasing of lb ^{13}CO signals by nearby lb ^{15}N spins, where "nearby" means r_{CN} is smaller than ~8 Å. This lb ^{13}CO/lb ^{15}N proximity only holds for t = u, u ± 1, or u ± 2 registries, so large $\Delta S/S_0$ for many WT samples with different u supports a broad distribution of antiparallel registries. This finding also holds for V2E. In addition, for the same u, i.e. labeling, there is often a large difference between $\Delta S/S_0$ for WT vs. V2E Fp, which supports very different registry distributions for the two Fp's. These distributions are important information to understand the reasons for their fusion -active vs. -defective functions. These WT and V2E registry distributions are quantitatively described by a set of registry populations, $f(t)^{WT}$ and $f(t)^{V2E}$, with $\sum_t f(t)^{WT} = \sum_t f(t)^{V2E} = 1$. A significant point is that each $\Delta S/S_0(u)$ value has contributions from $t \neq u$ as well as $t = u$ registries so $f(t = u)$ is not proportional to $\Delta S/S_0(u)$. The interconnected dependences of the $\Delta S/S_0(u)$ data on a single set of f(t) populations means the f(t) must be determined by simultaneous global fitting of all 17 u = 8 to 24 sample data, with $f(t)^{WT}$ and $f(t)^{V2E}$ determined from WT and V2E data, respectively. Fittings are based on χ^2 calculated with the $[S_1/S_0](u,\tau_k)^{exp}$ experimental intensity ratios and $[S_1/S_0](u,\tau_k)$ calculated from mathematical models. The SI provides a detailed description of the models and analyses. In brief, the total $S_0^{tot}(\tau_k)$ and total $S_1^{tot}(\tau_k)$ signals are each considered to be sums of lb and na ^{13}CO signals. The $S_0^{lb}(\tau_k)$ is assigned to be 1.0 so that $S_0^{na}(\tau_k) = 0.33$, based on the 30 other backbone carbonyl spins. Both the $S_1^{lb}(u,\tau_k)$, which is u and therefore labeling-dependent, and $S_1^{na}(\tau_k)$ signals are sums of contributions from different ^{13}CO populations that experience different ^{13}CO-^{15}N dipolar dephasings. The $S_1^{lb}(u,\tau_k)$ signal includes an attenuated $S_1^{lb,na}(\tau_k)$ contribution from lb ^{13}CO spins close to na ^{15}N spins. The remaining $S_1^{lb}(u,\tau_k)$ signal is divided into two categories that are registry-dependent: 1) $S_1^{lb,lb}(u,\tau_k)$ is signal from lb ^{13}CO spins in registries with t values close to u, and is attenuated by dephasing from lb ^{15}N spins; and 2) $S_1^{lb,X}(u,\tau_k)$ is S_1 signal of lb ^{13}CO in all other registries, which are collectively designated as "X," with $S_1^{lb,X}(u,\tau_k) = S_0^{lb,X}(u,\tau_k)$. The specific values of t designated as X are given later in this section. The $S_1^{na}(\tau_k)$ signal is similarly separated and includes the attenuated $S_1^{na,lb}(\tau_k)$ from na ^{13}CO near lb ^{15}N spins. The S_1 signal of other na ^{13}CO spins is not attenuated with $S_1(\tau_k) = S_0(\tau_k)$.

When Fp labeling results in β sheets without proximal lb ^{13}CO/lb ^{15}N pairs, there is only na dephasing so that:

$$\frac{\Delta S}{S_0}(\tau_k) = 0.0442 - \left[0.01105 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m, \tau_k)\right], \quad [1]$$

where the sum is over lb/na pairs with shortest ^{13}CO-^{15}N distances (r_m), the ^{13}CO-^{15}N dipolar coupling, $d_m(\text{Hz}) = 3{,}080/r_m(\text{Å})^3$, and the $\gamma^{lb,na}(d_m,\tau_k)$ are calculated using a mathematical expression for REDOR (49). For the present study, the parameter γ always refers to a S_1/S_0 intensity ratio calculated for a specific spin geometry. The Eq. **1** model for na dephasing is validated by quantitative agreement between the calculated $\Delta S/S_0$ and WT u = 28 $\Delta S/S_0$ data, Fig. 3. Similar na dephasing models have been validated experimentally in earlier REDOR NMR studies of Fp's (27, 50, 51). The other u = 8 to 24 $\Delta S/S_0$ data are fitted with contributions to dephasing from na spins (using the Eq. **1** model) and from proximal lb ^{13}CO/lb ^{15}N pairs. The latter contribution depends on how Fp strands are arranged in the β sheets, and "constrained" and "unconstrained" models are considered. For constrained fitting, all neighboring strands within a single sheet have the same registry and value of t, and f(t) is the fractional population of sheets with a particular t. Fig. 2*A* displays schematic representations of a constrained t = 16 sheet and the relevant spin geometry for this sheet for $L12_CG5_N$ labeling, i.e. u = 16. The $S_1^{lb,lb}$ contributions are from sheets with t = u, u + 1, u–1, u + 2, or u–2 registries and *SI Appendix*, Table S4 displays the $\gamma_{t=u}^{lb,lb}(\tau_k)$, $\gamma_{t=u\pm1}^{lb,lb}(\tau_k)$, and $\gamma_{t=u\pm2}^{lb,lb}(\tau_k)$ values calculated using the SIMPSON simulation program (52, 53). For each t value, the intermolecular β sheet structure means that the simulation spin system is one lb ^{13}CO and two lb ^{15}N spins in the 3D geometry of a β sheet. The other $t \neq u$, u ± 1, u ± 2 are designated as X with $\gamma_X^{lb,lb}(\tau_k) = 1$. This approximation is supported by $\gamma_{t=u\pm2} = 0.9334$ at $\tau_k = 40.2$ ms. The constrained f(t) populations are described:

$$\sum_{t=u-2}^{u+2} \left\{ f(t) \times \left[1 - b \times \gamma_t^{lb,lb}(\tau_k)\right]\right\} = 1.350 \times \left\{ \left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) - \left[0.01105 \times \left(\sum_{m=1}^{4} 1 - \gamma^{lb,na}(d_m,\tau_k)\right)\right]\right\}, \quad [2]$$

where exp≡experimental and b is a user-selected number between 0.95 and 1. Having b < 1 accounts for dephasing due to more distant lb ^{15}N spins not included in the spin geometries used to calculate $\gamma_t^{lb,lb}(\tau_k)$. Eq. **2** shows that the $(1 - b \times \gamma_t^{lb,lb})$ weighted f(t) sum is proportional to $\Delta S/S_0(u)$ after subtracting the na contribution between the right-hand square brackets.

For unconstrained fitting, there can be multiple registries within a single sheet, and f(t) is the fractional population of neighboring strands with a particular t. Each "central" Fp molecule has neighboring Fp molecules denoted 1 and 2, with assignment of molecule 1 vs. 2 based on having vs. not having a NH hydrogen-bonded to the lb ^{13}CO of the central molecule. Registries are indexed by t_1 and t_2, the total number of residues of molecule 1 or 2, respectively, that could be hydrogen-bonded to the central molecule. Fig. 2*A* displays $u = t_1 = t_2 = 16$ and Fig. 2*B* displays $u = 16$, $t_1 = 17$, and $t_2 = 19$. *SI Appendix*, Table S4 presents the SIMPSON-calculated $\gamma_{t1,t2}^{lb,lb}(\tau_k)$, when t_1 and/or t_2 = u, u + 1, or u–1. The t_1 = X means $t_1 \neq u$, u ± 1 and t_2 = X means $t_2 \neq u$, u + 1 and $\gamma_{X,X}^{lb,lb}(\tau_k) = 1$. The intermolecular β sheet structure means the simulation spin system is one lb ^{13}CO and two lb ^{15}N spins when $t_1 \neq X$ and $t_2 \neq X$, Fig. 2*A*, and one lb ^{13}CO and one lb ^{15}N when $t_1 \neq X$ or $t_2 \neq X$, Fig. 2*B*. The unconstrained f(t) populations are described:

$$\sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{u+1,X} \left[f(t_1) \times f(t_2) \times b_{t_1,t_2} \times \gamma_{t_1,t_2}^{lb,lb}(\tau_k)\right] = 1 - 1.350 \times \left\{ \left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) - 0.01105 \times \left[\sum_{m=1}^{4} 1 - \gamma^{lb,na}(d_m,\tau_k)\right]\right\}. \quad [3]$$

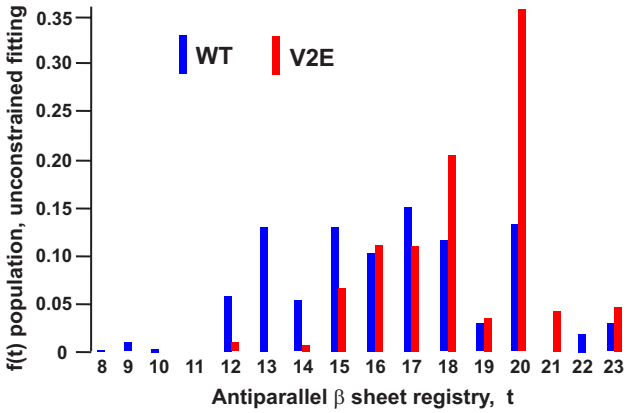

Fig. 4. WT (blue bar) and V2E (red bar) populations, f(t), vs. antiparallel β sheet registry t, which is the total number of residues in one strand that could be hydrogen-bonded to residues in the neighboring strand. The $f(t)^{WT}$ values were determined using Eq. **3** unconstrained fitting of WT $(S_1/S_0)^{exp}$ data from u = 8 to 24 samples with τ_k = 2.2 to 40.2 ms and b = 0.98. The $f(t)^{V2E}$ were similarly determined using the V2E data. The $f(t)^{WT}$ and $f(t)^{V2E}$ numerical values are presented in *SI Appendix*, Table S5.

There is a single f(t) population for each t, i.e. when $t_1 = t_2$, $f(t_1) = f(t_2)$. The $b_{t1,t2}$ in Eq. **3** are user-selected numbers when t_1 and/or t_2 = u or u ± 1 and are between 0.98 and 1. The $b_{t1,t2} < 1$ account for dephasing due to more distant lb ^{15}N spins not included in the spin geometries used to calculate $\gamma_{t1,t2}^{lb,lb}(\tau_k)$ in *SI Appendix*, Table S4. Eq. **3** shows that the $b_{t1,t2} \times \gamma_{t1,t2}^{lb,lb}$ weighted sum of $f(t_1) \times f(t_2)$ products is equal to one minus a term proportional to $\Delta S/S_0(u)$ minus the natural abundance contribution.

The f(t), t = 8 to 24, are determined by global χ^2 fitting of the u = 8 to 24 data using Eqs. **2** or **3** and Python code presented in the *SI Appendix*. Fig. 3 displays plots of experimental $\Delta S/S_0$ vs. τ_k for WT and V2E samples, u = 8 to 24 and τ_k = 2.2 to 40.2 ms, as well as best-fit $\Delta S/S_0$ from unconstrained fittings using "b = 0.98", i.e. $b_{t1=u,u\pm1,t2=u,u\pm1} = 0.98$, $b_{t1=u,u\pm1,t2=X} = b_{t1=X,t2=u,u\pm1} = 0.99$, and $b_{t1=X,t2=X} = 1$. Fig. 4 and *SI Appendix*, Table S5 present best-fit $f(t)^{WT}$ and $f(t)^{V2E}$ from these fittings, and *SI Appendix*, Table S6 (WT) and *SI Appendix*, Table S7 (V2E) numerically present the experimental and calculated $\Delta S/S_0$. The fitting quality is evidenced by best-fit χ^2 of 107 for WT and 145 for V2E, which are close to the number of data, 102. *SI Appendix*, Table S5 also presents best-fit $f(t)^{WT}$ and $f(t)^{V2E}$ from constrained fittings with b = 0.98, with calculated $\Delta S/S_0$ in *SI Appendix*, Tables S6 and S7. *SI Appendix*, Table S8 presents f(t) and χ^2 values from all fittings. Differences among fittings included the value of the scaling factor b and/or whether τ_k = 48.2 ms data were included. For a particular t, the f(t) value is usually similar among all fittings with typical variation of ~0.015, which supports the accuracy of the f(t) determination. The $f(t)^{WT}$ and $f(t)^{V2E}$ determinations are the core result of this study and show that both WT and fusion-defective V2E have very broad but different registry distributions, with each distribution having at least eight registries with f(t) ≥ 0.02. Relative to WT, the V2E registry distribution is weighted to larger t/longer registries. For each fitting, *SI Appendix*, Table S8 provides $\langle t \rangle$, the average value of t. The $\langle t \rangle^{WT}$ is highly conserved among different WT fittings with $\left\langle \langle t \rangle^{WT} \right\rangle = 16.132 \pm 0.048$, and $\langle t \rangle^{V2E}$ is similarly conserved among V2E fittings with $\left\langle \langle t \rangle^{V2E} \right\rangle = 18.475 \pm 0.028$. Although constrained fitting includes $|t-u| = 2$ contributions to dephasing, Eq. **2**, the equivalent t_1, $t_2 = u \pm 2$ contributions are not included in unconstrained fitting, Eq. **3**. One reason for this approximation is that relative to constrained, the unconstrained model has a larger number of distinct lb ^{13}CO/lb ^{15}N spin geometries, with nine t_1, t_2 combinations when t_1 and t_2 independently adopt values of u, u ± 1, and X, *SI Appendix*, Table S4. The validity of this unconstrained approximation is supported by χ^2 values from these fittings that are comparable to the number of data and typically smaller than χ^2 from constrained fittings.

Unlike most earlier REDOR NMR studies for which data were analyzed to determine interspin distances and/or angles, the molecular structural information determined in the present study is structural populations, specifically the f(t) antiparallel registry populations. The importance of simultaneous global fitting of data of the 17 different samples rather than separate fittings of data of individual samples is highlighted by lack of proportionality between $\Delta S/S_0(u)$ and f(t = u). For example, the V2E experimental $\Delta S/S_0(u)$ at τ_k = 40.2 ms are very similar for u = 16, 17, 18, 19, with values of 0.25, 0.27, 0.30, 0.28, respectively, *SI Appendix*, Table S2. By contrast, there are very different f(t = u), with unconstrained values of 0.11, 0.11, 0.21, 0.04, *SI Appendix*, Table S5. Another example is very different $\Delta S/S_0$ for WT u = 11 and V2E u = 19 samples, 0.10 and 0.28, respectively, vs. very similar $f(t = 11)^{WT} = f(t = 19)^{V2E} = 0.04$ in unconstrained fittings, i.e. $\Delta S/S_0(u = 19)$ values are dominated by t ≠ 19 registries.

Free Energy Contributions. The accurate and quantitative determinations of $f(t)^{WT}$ and $f(t)^{V2E}$, Fig. 4 and *SI Appendix*, Table S5, and accompanying finding of very broad but very different registry distributions for WT vs. V2E, are significant information about Fp structure and provide a basis for understanding Fp role(s) in fusion. The f(t) are thermodynamic equilibrium values, based on reproducible $\Delta S/S_0$ for replicate samples, *SI Appendix*, Table S3. The free energy function, G(t), likely has registry length- and sequence-dependent contributions, e.g. specific sidechain/sidechain and sidechain/lipid contacts. Elucidating these contributions should provide insight into the broad but different WT vs. V2E registry distributions and their fusion -catalytic vs. -defective functions. The SI includes detailed description of fittings of $f(t)^{WT}$ to different G(t) contributions, as well as similar fittings of $f(t)^{V2E}$. The G(t) functions must have contribution(s) that are nonmonotonic in t to match the nonmonotonic f(t), e.g., for t = 13, 14, and 15, $f(t)^{WT}$ are 0.13, 0.05, and 0.13. In addition, contributions must be different for WT vs. V2E to reflect differences in $f(t)^{WT}$ vs. $f(t)^{V2E}$. The $f(t)^{WT}$ distribution from unconstrained fitting with b = 0.98 was successfully fitted using:

$$f(t)^{WT} = C^{WT} \times exp\left\{\frac{-\left[\left(G_\beta^{WT} \times t\right) + \left(G_{Leu}^{WT} \times L(t)\right) + \left(G_{sc}^{WT}(t) \times g^{WT}\right)\right]}{RT}\right\}, \quad [4]$$

and the $f(t)^{V2E}$ distribution was fitted using

$$f(t)^{V2E} = C^{V2E} \times exp\left\{\frac{-\left[\left(G_\beta^{V2E} \times t\right) + \left(G_{Leu}^{V2E} \times L(t)\right)\right]}{RT}\right\}, \quad [5]$$

with RT = 0.6 kcal/mole. The G_β^{WT} is the free energy-per-residue of β sheet formation, with sheet extension from A1 to t evidenced by β sheet ^{13}C chemical shifts (18, 22). The G_{Leu}^{WT} is the free energy when leucines are aligned in adjacent strands in a β sheet, with L(t) = 1 when at least one residue position is aligned, and L(t) = 0 in absence of such alignment. The aligned leucines are bolded in the schematic registries of Fig. 3. The $G(t)_{sc}^{WT}$ is the sum of free energies of membrane insertion of sidechains for residues V2 to t−1, with sidechain energy relative to Ala (54). Earlier studies support membrane insertion of WT Fp starting near V2 (40, 55). The g^{WT} is a scaling factor that accounts for the positive free energy of membrane insertion of the Fp backbone and C^{WT} is an overall scaling factor. Fitting of $f(t)^{WT}$, t = 11 to 20 with Eq. **4** yielded $R^2 = 0.88$ with typical residual magnitude of ~0.01, $G_\beta^{WT} = -0.113 \pm 0.038$ kcal/mole, $G_{Leu}^{WT} = -0.350 \pm 0.079$ kcal/mole, and $g^{WT} = 0.129 \pm 0.040$. Fig. 5*A* and *SI Appendix*, Table S9 display $f(t)^{WT}$ and best-fit values, and the three $G(t)^{WT}$ contributions. Models different than Eq. **4** were tried but resulted in poorer fitting, with examples described in the *SI Appendix*.

The $f(t)^{V2E}$, t = 15 to 21, were first fitted using Eq. **4**, with separate fittings done for insertion starting at I4, A6, or L7, and in all cases, the best-fit $g^{V2E} \approx 0$. This correlates with shallower membrane insertion for V2E vs. WT Fp (55). The $f(t)^{V2E}$ fitting was then done using Eq. **5**, and $R^2 = 0.78$, $G_\beta^{V2E} = -0.184 \pm 0.056$ kcal/mole, and $G_{Leu}^{V2E} = -1.21 \pm 0.44$ kcal/mole. The typical residual was ~0.01 except for t = 16 which fitted poorly with residual of ~0.1. When $f(16)^{V2E}$ was not included, $R^2 = 0.98$, $G_\beta^{V2E} = -0.195 \pm 0.021$ kcal/mole, and $G_{Leu}^{V2E} = -1.40 \pm 0.22$ kcal/mole, Fig. 5*B* and *SI Appendix*, Table S10. T = 16 is the longest registry with only apolar residues, other than E2, and $G(16)^{V2E}$ may have a membrane insertion contribution not included in Eq. **5**. There isn't a large contribution to $G(23)^{V2E}$ from E2-R22 salt bridges, based on $f(23)^{V2E} < 0.05$, and only ~0.015 larger than $f(23)^{WT}$. Further analysis of G(t) contributions in context of earlier studies is in the *Discussion* section.

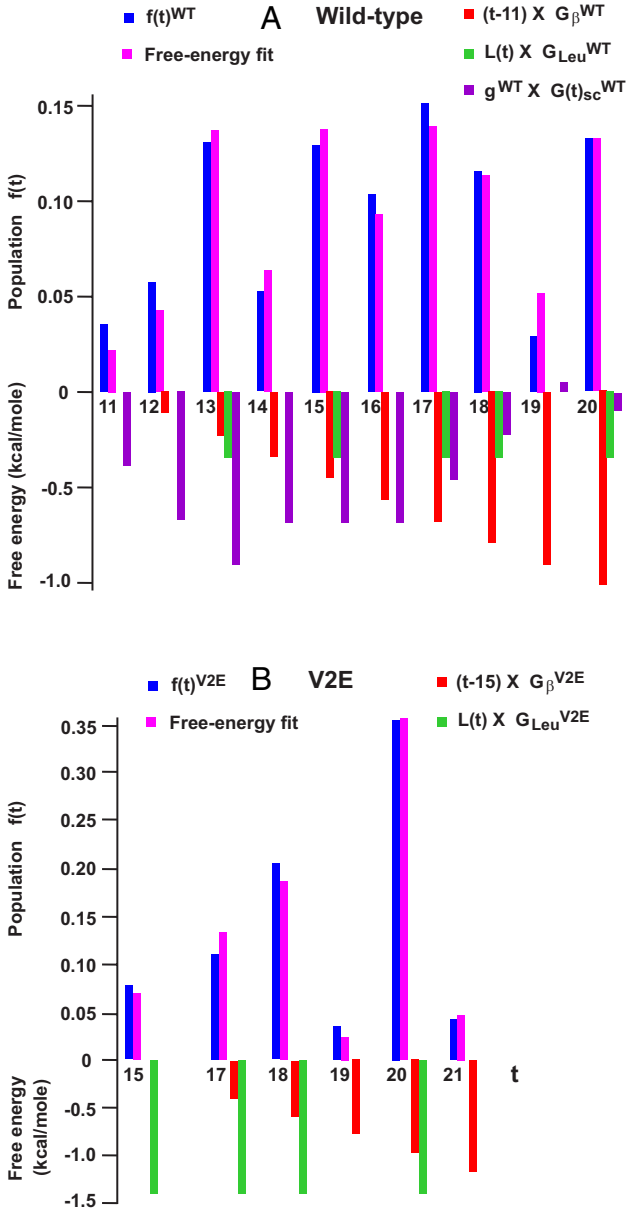

Fig. 5. WT (*A*) and V2E (*B*) bar plots of f(t) determined from unconstrained fitting of REDOR data (blue) and f(t) calculated from free energy fitting (magenta). There are also bar plots of free energy contributions from β sheet length (red), leucine alignment (green), and membrane insertion (purple). Free energy fitting for WT was done using Eq. **4** and $f(t)^{WT}$, t = 11 to 20, and for V2E, Eq. **5** and $f(t)^{V2E}$, t = 15, 17 to 21. These t ranges respectively include ~95% and ~85% of the total registry population. The f(t) and free energy numerical values are presented in *SI Appendix*, Table S9 (WT) and *SI Appendix*, Table S10 (V2E). The f(t) values are determined by relative rather than absolute values of G(t), i.e., adding or subtracting a constant energy value to all G(t) does not change the f(t). The energy offsets of $G(t)_\beta^{WT} = (t - 11) \times G_\beta^{WT}$ in panel a and $G(t)_\beta^{V2E} = (t - 15) \times G_\beta^{V2E}$ in panel b are respectively chosen so that $G(11)_\beta^{WT} = 0$ and $G(15)_\beta^{V2E} = 0$.

Discussion

Broad Linewidths Correlate with Broad Registry Distributions. Each spectral line profile is typically well-fitted to a single Gaussian function with 3 to 4 ppm linewidth, *SI Appendix*, Fig. S2 and Table S1. These profiles are likely due to superposition of unresolved signals with different peak shifts from the individual β sheet registries within the distribution, based on the ~1 ppm typical linewidth for a single-site backbone ^{13}C signal in a membrane protein with a unique structure (21, 56, 57). The multiple-registry explanation for the broad Fp linewidths is supported by $\delta_{FWHM,WT} > \delta_{FWHM,V2E}$ for most u, with typical difference ≈ 0.5 ppm. This correlates with the larger number of populated registries for WT vs. V2E. For example, unconstrained b = 0.98 fitting results in 12 values of t with $f(t)^{WT} > 0.015$ vs. only 8 values with $f(t)^{V2E}$, *SI Appendix*, Table S5.

Registry Distributions Are Very Similar with Unconstrained and Constrained Models and for Fp with and without C-Terminal Hairpin. The robustness and accuracy of the f(t) determinations are supported by very similar f(t) values for many types of fittings, including unconstrained vs. constrained, without vs. with $\tau_k = 48.2$ ms data, and b < 1 vs. b = 1, *SI Appendix*, Table S8. Comparison among all WT or all V2E fittings shows typical variation <0.015. There is some breadth of χ^2 values with largest contributions typically from data with larger τ_k. The average value of t, ⟨t⟩, is highly conserved. For all WT fittings, the $\left\langle \langle t \rangle^{WT} \right\rangle = 16.132 \pm 0.048$, and for all V2E fittings, $\left\langle \langle t \rangle^{V2E} \right\rangle = 18.475 \pm 0.028$. Earlier NMR studies of membrane-bound WT Fp+hairpin protein evidenced Fp's with antiparallel β sheet structure and multiple populated registries (28, 29). This result supported interleaved strands from two hairpin trimers, Fig. 6*A*. A requirement of two hairpin trimers is also supported by analyses of V2E-dominant losses vs. WT:V2E ratio for HIV gp160 fusion and infection as well as Fp+hairpin-induced vesicle fusion and helicity (30, 31, 34). The parallel coiled-coil alignment of the three N-helices in each trimer could favor constrained Fp

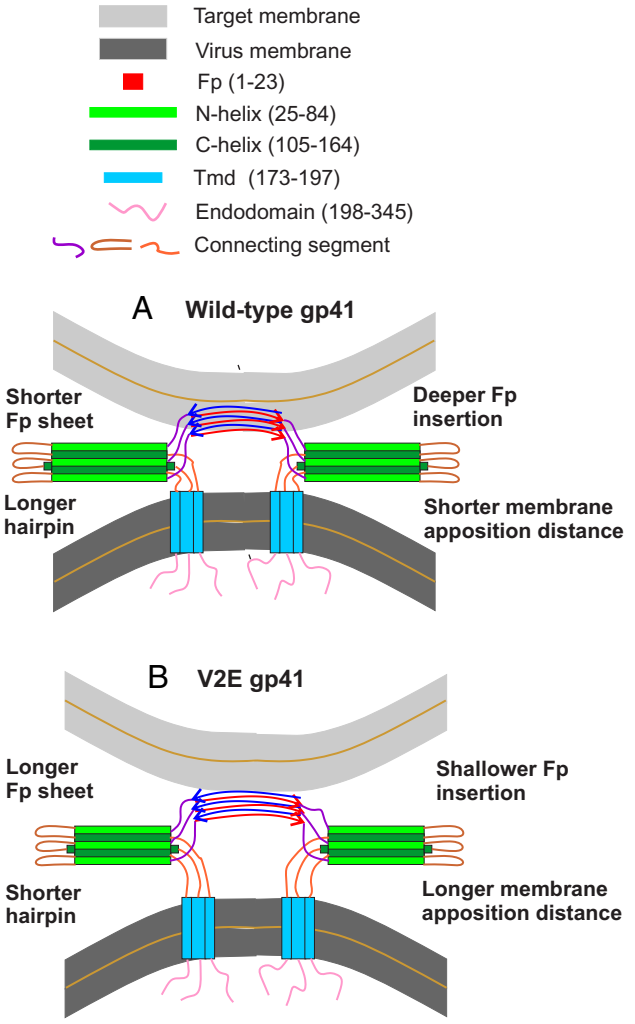

Fig. 6. Structural models for (*A*) WT and (*B*) V2E gp41. Slower fusion rate for V2E is ascribed to longer membrane apposition distance and shallower membrane location of the Fp β sheet. The longer distance is correlated with longer V2E Fp sheets detected in the present study and shorter V2E hairpins detected in a previous study. The longer distance means that stalk formation, see Fig. 1 (ii), requires larger-amplitude lipid chain motions and therefore proceeds at a slower rate. Shallower Fp location also reduces the probability of lipid chain protrusion into the aqueous region.

registries within a single β sheet. The populations of a few specific Fp registries in Fp+hairpin were probed using NMR detection of dipolar dephasing due to proximity between lb ^{13}CO spins on adjacent Fp strands (29). The experiment was constant-time double-quantum buildup with finite-pulses (fpCTDQBU), which results in $\Delta S/S_0$ vs. τ_k data with appearance similar to the REDOR data in Fig. 3 (58). The fpCTDQBU experiment was first tested using crystalline GFF peptide with lb ^{13}CO at G1 and F3 and intramolecular ^{13}CO–^{13}CO distance (r_{CC}) of 5.4 Å. After accounting for na contributions, the GFF lb ^{13}CO $\Delta S/S_0$ vs. τ_k data matched quantitatively with $\Delta S/S_0$ values calculated using r_{CC} = 5.4 Å. For longer dephasing times, i.e. $\tau_k \approx 40$ ms, $\Delta S/S_0$ leveled off at ~0.95 which correlates with GFF labeling in which nearly all lb ^{13}CO spins are in proximal lb G1/lb F3 ^{13}CO spin pairs. The fpCTDQBU experiment was then applied to Fp+hairpin samples with each sample having a single ^{13}CO label at a specific residue v in Fp. The t = 2v – 1 and t = 2v registries have the shortest interstrand lb ^{13}CO - lb ^{13}CO distances (5 to 6 Å) and therefore largest dipolar couplings that contribute to $\Delta S/S_0$, e.g., *SI Appendix*, Fig. S4 displays schematic geometries of F8 ^{13}CO label in t = 15 and t = 16 constrained registries. After accounting for na contributions, the lb ^{13}CO $\Delta S/S_0$ vs. τ_k buildups are approximately proportional to the GFF buildup but with smaller $\Delta S/S_0$ for Fp+hairpin. These results are consistent with Fp+hairpin $\Delta S/S_0$ being due to interstrand proximity of lb ^{13}CO spins and with their long-time $\Delta S/S_0$ being an estimate of the sum of populations of the 2v – 1 and 2v registries. For v = 4, 7, 8, 11, and 12 labelings, the long-time $\Delta S/S_0$ were ~0.04, 0.3, 0.3, 0.06, and 0.02, respectively, which agree semiquantitatively with the $f(2v-1)^{WT} + f(2v)^{WT}$ sums of the present study of ~0.001, 0.18, 0.23, 0.02, and 0.03 for unconstrained fitting and ~0, 0.19, 0.23, 0.003, and 0.02 for constrained fitting, *SI Appendix*, Table S5. This agreement supports the $f(t)^{WT}$ distribution of HIV gp41 in its final hairpin state to be similar to the $f(t)^{WT}$ distribution of the present study.

Broad Registry Distribution May be Advantageous for Chronic Infection by HIV. Broad registry distributions are unusual in peptides and proteins and the underlying reasons and functional advantages of them for Fp are considered. The unconstrained $f(t)^{WT}$, t = 11 to 20, were fitted with Eq. **4**, see Fig. 5*A* and *SI Appendix*, Table S9. The total free energies, $G(t)^{WT}$, were sums of contributions: 1) $t \times G_\beta{}^{WT}$, β sheet length; 2) $L(t) \times G_{Leu}{}^{WT}$, with L(t) = 1 or 0 for presence vs. absence of aligned leucines in adjacent strands; and 3) $g^{WT} \times G_{sc}(t)^{WT}$, sidechain membrane insertion. The typically negative values of all three contributions are ascribed to the hydrophobic effect, specifically release of water solvating the Fp because there is 1) β sheet hydrogen bonding; 2) packing of leucine sidechains; and 3) Fp solvation by lipid acyl chains. The relative fusion activities of different WT registries are considered. The Fp β sheet is likely inserted in a single leaflet rather than traversing the bilayer, Fig. 6, as is reasonable for interleaved Fp's from different hairpin trimers and also consistent with earlier observation that multiple residues within the G5-L12 region contact the lipid chain termini (40). Sheets with more negative $G_{sc}{}^{WT}$ are likely more deeply inserted and will therefore induce larger displacements of neighboring lipids which will catalyze fusion. For the unconstrained model, $G_{sc}{}^{WT}$ might be an average over the different registries, and for the constrained model, each sheet has a single specific t and therefore $G_{sc}(t)^{WT}$ and membrane insertion depth. This registry-dependent depth hypothesis is supported by an earlier NMR study that probed proximity between ^{13}C spins in Fp and ^{2}H spins in lipid acyl chains (59). The NMR data were only reasonably understood with two Fp β sheet populations, one inserted close to the bilayer center and the other with shallower insertion. Fusion activity also correlates with the number of sheet-neighboring lipids, $N_{lipid,nb}(t)$, with $N_{lipid,nb}(t) \propto$ sheet area $\propto$ t. Most registries are expected to have similar fusion activities, based on t = 12 to 16 having more negative $g^{WT} \times G_{sc}(t)^{WT}$, –0.91 to –0.67 kcal/mole, and smaller $N_{lipid,nb}(t)$, whereas t = 17 to 20 have less negative $g^{WT} \times G_{sc}(t)^{WT}$, –0.46 to 0.01 kcal/mole, but larger $N_{lipid,nb}(t)$. Similar fusion activities among most populated registries also confer fusion activity to unconstrained sheets with mixed registries.

HIV is a chronic infection that relies on constant mutation to escape neutralization by the immune system, with Fp being one of the neutralization epitopes (16, 17). Relative to a narrow registry distribution, the observed broad distribution for Fp likely enables mutants that can both evade the immune system and also remain fusion-competent. Many mutations likely cause only moderate changes in the f(t) distribution so fusion competence can be retained because most registries are fusion-active. This explanation for the broad distribution is supported by comparison between

the nonhomologous Fp's of HIV gp41 and influenza virus Ha2. Influenza is also membrane-enveloped, and fusion is likely catalyzed by a mechanism similar to Fig. 1, with Ha2 replacing gp41 (1). However, influenza is an acute rather than chronic infection and does not experience long-time immune pressure within a single person. The long- vs. short- time immune pressures of HIV vs. influenza are manifested in the contrasts between: 1) HIV Fp which has sequence variety among patient isolate strains and an intermolecular β sheet structure with broad registry distribution; vs. 2) influenza Fp which exhibits very high sequence conservation among viral isolates and adopts two very similar and monomeric helical hairpin structures (15, 60–62).

Longer V2E Registries with Shallower Membrane Insertion Explain V2E-Dominant Loss of Fusion, Infection, and Hairpin Helicity. Comparison of $\Delta S/S_0$ data among WT samples shows largest values for u = 17 whereas V2E samples show largest values for u = 20, Fig. 3. This correlates with the registry distribution weighted toward larger t for V2E, Fig. 4. The $f(t)^{V2E}$ distribution, t = 15, 17 to 21 was fitted with a free energy function that does not include sidechain membrane insertion contribution which correlates with earlier observation of deeper membrane insertion for WT vs. V2E Fp (55). There would therefore be greater lipid chain displacement for WT vs. V2E which correlates with ~10× greater vesicle fusion induced by WT vs. V2E Fp (38). The G_{Leu} and G_β in Eqs. **4** and **5** are due in part to less water contact for leucine and other hydrophobic amino acid sidechains when they are more tightly packed in β sheets than in monomeric peptides. The ratios $G_{Leu}{}^{V2E}/G_{Leu}{}^{WT} \approx 4$ and $G_\beta{}^{V2E}/G_\beta{}^{WT} \approx 1.7$, both >1, evidence this hydrophobic effect, as shallower membrane location of V2E Fp correlates with higher water content.

V2E mutation results in complete loss of HIV gp160-mediated fusion and infection with accompanying V2E dominance in mixed WT/V2E gp160 trimers (30). There are also V2E-dominant losses of helicity and vesicle fusion for mixed WT/V2E Fp+hairpin trimers which are quantitatively similar to losses in fusion and infection for mixed gp160 trimers. The mole fraction V2E dependences of losses in fusion, infection, and helicity of gp160 and Fp+hairpin were globally fitted and supported a requirement of cooperativity between at least 6 WT molecules for efficient fusion and infection, i.e. 2 WT trimers, Fig. 6 (34). One conundrum is how V2E, which is ~20 residues N-terminal of the hairpin, dominantly changes the hyperthermostable and autonomously folding structure of the hairpin trimer. This question is addressed by the finding of the present study that the V2E registry distribution is weighted to larger t. Several unstructured residues are likely required between the C-terminus of the Fp β sheet and the hairpin N-helix, so the longer V2E β sheets likely result in unfolding of the N-helix region closest to the Fp, Fig. 6*B*. The C-helices pack in the exterior grooves of the N-helix trimeric bundle, so loss of N-helix residues likely results in unfolding of the C-helix region closest to the Tmd. There are ~12 fewer helical residues in the hairpin for V2E vs. WT which would correspond to ~6 fewer N-helix and ~6 fewer C-helix residues. The N- and C-helices have heptad repeat sequences, so these losses correspond to about one repeat in both the N- and C-helices. Relative to WT, the unfolding of N- and C-helix segments will result in a longer distance between the apposed membranes, Fig. 6*B*. Stalk formation, Fig. 1 (ii), will therefore require larger-amplitude lipid chain protrusion into the aqueous phase and will happen at a slower rate (63). The protrusion probability will also be smaller because of shallower membrane location of V2E Fp sheets. V2E-dominant phenotypes in mixed WT/V2E trimers are understood by trimeric bundle formation by the N-helices and the bundle N terminus being determined by the longest registry in the Fp β sheet. There is consequently V2E-dominance in shortening the two hairpins and in lengthening the apposition distance. This mechanism is supported by V2E-dominant loss of hairpin helicity and vesicle fusion induced by Fp+hairpin trimers whereas V2E is not dominant for vesicle fusion induced by Fp-only (33, 34).

Quantitative Determination of Broad Structural Distributions Using REDOR NMR of Multiple Differently Labeled Samples. To our knowledge, this study is one of only a few reports of experimentally based quantitative determination of populations of >10 different structures of a molecule in a sample. Structural populations can sometimes be obtained in computational simulations, although it can be difficult to ascertain whether the populations represent a thermodynamic equilibrium distribution, and there is dependence of the energies that underlie the distribution on the force field parameters of the simulation (64). The present study highlights an underutilized strength of solid-state NMR to experimentally determine broad population distributions, particularly with a pulse sequence like REDOR whose data can be analyzed quantitatively and are insensitive to: 1) small variations in instrument parameters during acquisition over several days; and 2) similar small variations in parameters among acquisitions for different samples. Determination of the full registry distribution required REDOR NMR data for eighteen differently labeled samples, both because the distribution is broad and because the ^{13}C NMR signals from different registries are not resolved. The >10 structural populations determined in this study is larger than the few structures typically distinguished with other experimental approaches such as crystal diffraction or cryo-EM (65). There are typically only a few structures in earlier NMR studies based on spectrally resolved signals among the structures (56, 57).

The extent to which broad, but defined, structural distributions are important in biological processes is an open question. For processes like membrane fusion which are molecular movement rather than chemical reactions, the present study shows that a broad structural distribution can confer catalytic function that is mutationally robust. This is evolutionarily advantageous for a pathogen like HIV which requires constant mutation to escape neutralization by the immune system.

Materials and Methods

See *SI Appendix* for complete descriptions of Fp synthesis and purifications, sample preparation, NMR data acquisition and processing, constrained and unconstrained fittings, and free energy fittings. In brief, WT and V2E Fp's were synthesized manually and typical purity was >95% (23). NMR samples were made by suspending lipids in *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (HEPES) buffer at pH 7 and forming unilamellar vesicles with freeze-thaw cycles and extrusion. Fp solution was then added dropwise to vesicle solution, followed by centrifugation to pellet vesicles with bound Fp. The pellet was lyophilized, packed in a 4 mm NMR rotor, and rehydrated. REDOR NMR S_0 and S_1 spectra were acquired on a 9.4 T spectrometer with dephasing times τ_k in the 2.2 to 48.2 ms range. Samples were at −30 °C and pulse parameters were optimized for isotropic ^{13}CO signals without consideration of other ^{13}C signals (27). $S_0(u,\tau_k)$ and $S_1(u,\tau_k)$ ^{13}CO peak intensities were determined by integration. For either WT or V2E, f(t) registry populations were determined by χ^2 fitting of experimental $[S_1/S_0](u,\tau_k)$ ratios. Fittings were done with Python code provided in *SI Appendix* and included $\gamma(\tau_k) = S_1/S_0$ ratios calculated with the SIMPSON program (52).

Data, Materials, and Software Availability. All study data are included in *SI Appendix*, Tables S1-S10 and Software S1.

ACKNOWLEDGMENTS. We acknowledge assistance from the Mass Spectrometry, NMR, and Institute for Cyber-Enabled Research Facilities at Michigan State University and from facility staff including Dr. Daniel Holmes, Dr. Xiaoge Wang, and Dr. Dirk Colbry. HP from Cyber Kunlun Lab, China helped with the Python coding. Portions of the paper were developed from the thesis of Y.Z. This work was supported by the United States NIH grant number R01 AI047153.

Author affiliations: [a]Department of Chemistry, Michigan State University, East Lansing, MI 48824

Author contributions: Y.Z., S.D.S., L.X., U.G., and D.P.W. designed research; Y.Z., S.D.S., L.X., and U.G. performed research; Y.Z., S.D.S., L.X., U.G., and D.P.W. analyzed data; and D.P.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

1. J. M. White, S. E. Delos, M. Brecher, K. Schornberg, Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme. *Crit. Rev. Biochem. Mol. Biol.* **43**, 189–219 (2008).
2. R. Blumenthal, S. Durell, M. Viard, HIV entry and envelope glycoprotein-mediated fusion. *J. Biol. Chem.* **287**, 40841–40849 (2012).
3. S. C. Harrison, Viral membrane fusion. *Virology* **479**, 498–507 (2015).
4. S. Boonstra *et al.*, Hemagglutinin-mediated membrane fusion: A biophysical perspective. *Ann. Revs. Biophys.* **47**, 153–173 (2018).
5. T. Tang, M. Bidon, J. A. Jaimes, G. R. Whittaker, S. Daniel, Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res.* **178**, 104792 (2020).
6. W. R. Gallaher, Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. *Cell* **50**, 327–328 (1987).
7. D. C. Chan, D. Fass, J. M. Berger, P. S. Kim, Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997).
8. M. Caffrey *et al.*, Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J.* **17**, 4572–4584 (1998).
9. Z. N. Yang *et al.*, The crystal structure of the SIV gp41 ectodomain at 1.47 A resolution. *J. Struct. Biol.* **126**, 131–144 (1999).
10. V. Buzon *et al.*, Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathog.* **6**, e1000880 (2010).
11. B. Kwon, M. Lee, A. J. Waring, M. Hong, Oligomeric structure and three-dimensional fold of the HIV gp41 membrane-proximal external region and transmembrane domain in phospholipid bilayers. *J. Am. Chem. Soc.* **140**, 8246–8259 (2018).
12. A. Piai *et al.*, NMR model of the entire membrane-interacting region of the HIV-1 fusion protein and its perturbation of membrane morphology. *J. Am. Chem. Soc.* **143**, 6609–6615 (2021).
13. M. Pancera *et al.*, Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014).
14. A. B. Ward, I. A. Wilson, The HIV-1 envelope glycoprotein structure: Nailing down a moving target. *Immunol. Revs.* **275**, 21–32 (2017).
15. N. F. Parrish *et al.*, Phenotypic properties of transmitted founder HIV-1. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6626–6633 (2013).
16. R. Kong *et al.*, Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. *Science* **352**, 828–833 (2016).
17. K. Xu *et al.*, Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1. *Nat. Med.* **24**, 857–867 (2018).
18. J. Yang, C. M. Gabrys, D. P. Weliky, Solid-state nuclear magnetic resonance evidence for an extended beta strand conformation of the membrane-bound HIV-1 fusion peptide. *Biochemistry* **40**, 8126–8137 (2001).
19. L. V. Chernomordik, M. M. Kozlov, Mechanics of membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 675–683 (2008).
20. C. P. Jaroniec *et al.*, Structure and dynamics of micelle-associated human immunodeficiency virus gp41 fusion domain. *Biochemistry* **44**, 16167–16180 (2005).
21. C. M. Wasniewski, P. D. Parkanzky, M. L. Bodner, D. P. Weliky, Solid-state nuclear magnetic resonance studies of HIV and influenza fusion peptide orientations in membrane bilayers using stacked glass plate samples. *Chem. Phys. Lipids* **132**, 89–100 (2004).
22. W. Qiang, M. L. Bodner, D. P. Weliky, Solid-state NMR spectroscopy of human immunodeficiency virus fusion peptides associated with host-cell-like membranes: 2D correlation spectra and distance measurements support a fully extended conformation and models for specific antiparallel strand registries. *J. Am. Chem. Soc.* **130**, 5459–5471 (2008).
23. W. Qiang, D. P. Weliky, HIV fusion peptide and its cross-linked oligomers: Efficient syntheses, significance of the trimer in fusion activity, correlation of β strand conformation with membrane cholesterol, and proximity to lipid headgroups. *Biochemistry* **48**, 289–301 (2009).
24. A. L. Lai, A. E. Moorthy, Y. L. Li, L. K. Tamm, Fusion activity of HIV gp41 fusion domain is related to its secondary structure and depth of membrane insertion in a cholesterol-dependent fashion. *J. Mol. Biol.* **418**, 3–15 (2012).
25. M. Lorizate *et al.*, Comparative lipidomics analysis of HIV-1 particles and their producer cell membrane in different cell lines. *Cell. Microbiol.* **15**, 292–304 (2013).
26. M. L. Bodner *et al.*, Temperature dependence and resonance assignment of ^{13}C NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes. *Magn. Reson. Chem.* **42**, 187–194 (2004).
27. S. D. Schmick, D. P. Weliky, Major antiparallel and minor parallel beta sheet populations detected in the membrane-associated Human Immunodeficiency Virus fusion peptide. *Biochemistry* **49**, 10623–10635 (2010).
28. P. U. Ratnayake, K. Sackett, M. J. Nethercott, D. P. Weliky, pH-dependent vesicle fusion induced by the ectodomain of the human immunodeficiency virus membrane fusion protein gp41: Two kinetically distinct processes and fully-membrane-associated gp41 with predominant beta sheet fusion peptide conformation. *Biochim. Biophys. Acta* **1848**, 289–298 (2015).
29. K. Sackett, M. J. Nethercott, Z. X. Zheng, D. P. Weliky, Solid-state NMR spectroscopy of the HIV gp41 membrane fusion protein supports intermolecular antiparallel beta sheet fusion peptide structure in the final six-helix bundle state. *J. Mol. Biol.* **426**, 1077–1094 (2014).
30. E. O. Freed, E. L. Delwart, G. L. Buchschacher Jr., A. T. Panganiban, A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 70–74 (1992).
31. O. F. Brandenberg, C. Magnus, R. R. Regoes, A. Trkola, The HIV-1 entry process: A stoichiometric view. *Trends Microbiol.* **23**, 763–774 (2015).
32. F. B. Pereira, F. M. Goni, J. L. Nieva, Liposome destabilization induced by the HIV-1 fusion peptide effect of a single amino acid substitution. *FEBS Lett.* **362**, 243–246 (1995).
33. Y. Kliger *et al.*, Fusion peptides derived from the HIV type 1 glycoprotein 41 associate within phospholipid membranes and inhibit cell-cell fusion, structure-function study. *J. Biol. Chem.* **272**, 13496–13505 (1997).
34. M. Rokonujjaman *et al.*, A large HIV gp41 construct with trimer-of-hairpins structure exhibits V2E mutation-dominant attenuation of vesicle fusion and helicity very similar to V2E attenuation of HIV fusion and infection and supports: (1) hairpin stabilization of membrane apposition with larger distance for V2E; and (2) V2E dominance by an antiparallel beta sheet with interleaved fusion peptide strands from two gp41 trimers. *Biophys. Chem.* **293**, 106933 (2023).
35. R. Yang, M. Prorok, F. J. Castellino, D. P. Weliky, A trimeric HIV-1 fusion peptide construct which does not self-associate in aqueous solution and which has 15-fold higher membrane fusion rate. *J. Am. Chem. Soc.* **126**, 14722–14723 (2004).
36. F. B. Pereira, F. M. Goni, A. Muga, J. L. Nieva, Permeabilization and fusion of uncharged lipid vesicles induced by the HIV-1 fusion peptide adopting an extended conformation: Dose and sequence effects. *Biophys. J.* **73**, 1977–1986 (1997).
37. J. Yang, M. Prorok, F. J. Castellino, D. P. Weliky, Oligomeric β-structure of the membrane-bound HIV-1 fusion peptide formed from soluble monomers. *Biophys. J.* **87**, 1951–1963 (2004).
38. C. M. Gabrys *et al.*, Solid-state nuclear magnetic resonance measurements of HIV fusion peptide ^{13}CO to lipid ^{31}P proximities support similar partially inserted membrane locations of the α Helical and β sheet peptide structures. *J. Phys. Chem. A* **117**, 9848–9859 (2013).
39. C. M. Gabrys *et al.*, Nuclear magnetic resonance evidence for retention of a lamellar membrane phase with curvature in the presence of large quantities of the HIV fusion peptide. *Biochim. Biophys. Acta* **1798**, 194–201 (2010).
40. L. H. Jia *et al.*, REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins. *J. Magn. Reson.* **253**, 154–165 (2015).
41. P. Zhu *et al.*, Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**, 847–852 (2006).
42. R. Sougrat *et al.*, Electron tomography of the contact between T cells and SIV/HIV-1: Implications for viral entry. *PLoS Pathog.* **3**, e63 (2007).
43. J. Chojnacki *et al.*, Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* **338**, 524–528 (2012).
44. T. Gullion, Introduction to rotational-echo, double-resonance NMR. *Concepts Magn. Reson.* **10**, 277–289 (1998).
45. H. Yang *et al.*, REDOR NMR reveals multiple conformers for a Protein Kinase C ligand in a membrane environment. *ACS Cent. Sci.* **4**, 89–96 (2018).
46. M. Kashefi, N. Malik, J. O. Struppe, L. K. Thompson, Carbon-nitrogen REDOR to identify ms-timescale mobility in proteins. *J. Magn. Reson.* **305**, 5–15 (2019).
47. A. T. Petkova *et al.*, A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16742–16747 (2002).
48. H. Y. Zhang, S. Neal, D. S. Wishart, RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25**, 173–195 (2003).
49. K. T. Mueller, Analytical solutions for the time evolution of dipolar-dephasing NMR signals. *J. Magn. Reson. Ser. A* **113**, 81–93 (1995).
50. J. Yang, D. P. Weliky, Solid state nuclear magnetic resonance evidence for parallel and antiparallel strand arrangements in the membrane-associated HIV-1 fusion peptide. *Biochemistry* **42**, 11879–11890 (2003).
51. Z. Zheng, R. Yang, M. L. Bodner, D. P. Weliky, Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. *Biochemistry* **45**, 12960–12975 (2006).
52. M. Bak, J. T. Rasmussen, N. C. Nielsen, SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.* **147**, 296–330 (2000).
53. M. Bak, R. Schultz, T. Vosegaard, N. C. Nielsen, Specification and visualization of anisotropic interaction tensors in polypeptides and numerical simulations in biological solid-state NMR. *J. Magn. Reson.* **154**, 28–45 (2002).
54. C. P. Moon, K. G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10174–10177 (2011).
55. W. Qiang, Y. Sun, D. P. Weliky, A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15314–15319 (2009).
56. M. P. Bhate, B. J. Wylie, L. Tian, A. E. McDermott, Conformational dynamics in the selectivity filter of KcsA in esponse to potassium ion concentration. *J. Mol. Biol.* **401**, 155–166 (2010).
57. L. B. Andreas *et al.*, Structure and mechanism of the influenza A $M2_{18-60}$ dimer of dimers. *J. Am. Chem. Soc.* **137**, 14877–14886 (2015).
58. Z. Zheng, W. Qiang, D. P. Weliky, Investigation of finite-pulse radiofrequency-driven recoupling methods for measurement of intercarbonyl distances in polycrystalline and membrane-associated HIV fusion peptide samples. *Magn. Reson. Chem.* **245**, S247–S260 (2007).
59. L. Xie, L. H. Jia, S. Liang, D. P. Weliky, Multiple locations of peptides in the hydrocarbon core of gel-phase membranes revealed by peptide ^{13}C to lipid ^{2}H rotational-echo double-resonance solid-state nuclear magnetic resonance. *Biochemistry* **54**, 677–684 (2015).
60. E. Nobusawa *et al.*, Comparison of complete amino acid sequences and receptor binding properties among 13 serotypes of hemagglutinins of influenza A viruses. *Virol.* **182**, 475–485 (1991).

61. J. L. Lorieau, J. M. Louis, A. Bax, The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid:Water interface. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11341–11346 (2010).
62. U. Ghosh, L. Xie, L. H. Jia, S. Liang, D. P. Weliky, Closed and semiclosed interhelical structures in membrane vs closed and open structures in detergent for the Influenza Virus hemagglutinin fusion peptide and correlation of hydrophobic surface area with fusion catalysis. *J. Am. Chem. Soc.* **137**, 7548–7551 (2015).
63. P. Larsson, P. M. Kasson, Lipid tail protrusion in simulations predicts fusogenic activity of influenza fusion peptide mutants and conformational models. *PLoS Comp. Biol.* **9**, e1002950 (2013).
64. A. Panahi, M. Feig, Conformational sampling of Influenza fusion peptide in membrane bilayers as a function of termini and protonation states. *J. Phys. Chem. B* **114**, 1407–1416 (2010).
65. Y. F. Cai *et al.*, Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586–1592 (2020).

Supporting Information for

Very broad distribution of β sheet registries of the HIV gp41 fusion peptide supports mutational robustness for fusion and infection

Yijin Zhang[1,2], Scott D. Schmick[1,2], Li Xie, Ujjayini Ghosh[2], and David P. Weliky*

Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA

[1] These authors contributed equally to this work.
[2] Present Addresses: Y. Zhang, Interdisciplinary Institute of NMR and Molecular Sciences, School of Chemistry and Chemical Engineering, The State Key Laboratory of Refractories and Metallurgy, Wuhan University of Science and Technology, Wuhan 430081, People's Republic of China; S. D. Schmick, Specialty Materials Development Laboratory, Corporate Research and Development, 3M Company, Saint Paul, MN 55144; U. Ghosh, Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105

* Corresponding author, Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA; weliky@chemistry.msu.edu

This PDF file includes:
Supporting text with detailed descriptions of Materials and Methods

Supporting Information Text of Detailed Descriptions of Materials and Methods

Sample preparation. WT Fp peptides had sequence AVGIGALFLGFLGAAGSTMGARSWKKKKKKA, with underlining for the 23 N-terminal residues of HIV gp41, HXB2 laboratory strain, followed by a non-native W which served as a A280 chromophore, and a polylysine tag which resulted in Fp monomers in aqueous solution prior to membrane binding (1, 2). The peptides were synthesized manually by solid-phase peptide synthesis using 9-Fluorenylmethoxycarbonyl (Fmoc) chemistry, followed by cleavage from the resin with a trifluoroacetic acid solution, and then purification with reverse-phase high-performance liquid chromatography with a semi-preparative C4 column. The synthesis and purification followed published methods, with typical final Fp purity >95% as assessed by MALDI-TOF mass spectrometry (3). Each Fp had one residue with a backbone ^{13}CO label and a different residue with a backbone ^{15}N label. Labeled amino acids were purchased from Cambridge Isotopes (Andover, MA) and were Fmoc-protected using published methods (4). Each labeled Fp was indexed by an integer u. The u value is defined using the parameter t which is the registry length of two neighboring strands, i.e. the total number of hydrogen-bonded residues in either strand. The u is the value of t for which the ^{13}CO-labeled residue on one strand is aligned with the ^{15}N-labeled residue on the adjacent antiparallel strand. The integer value of u is also the ^{13}CO residue number plus ^{15}N residue number minus 1, e.g. Fp with $L12_CG5_N$ labeling has u=16, **Fig**. **2a** in the main text. Both WT and V2E Fp's were produced with 17 different labelings that correspond to all values of u in the 8-24 range. WT Fp with u = 28 was also produced.

The lipid composition of the samples was 1,2-di-*O*-tetradecyl-*sn*-glycero-3-phosphocholine (DTPC), 1,2-di-*O*-tetradecyl-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (DTPG), and cholesterol in an 8:2:5 molar ratio. Ether- rather than the ester-linked phospholipids were used so that there wouldn't be lipid natural abundance (na) ^{13}CO NMR signal. This study relies on analysis of the labeled (lb) Fp ^{13}CO NMR signals and this analysis would likely be less accurate if there were lipid contributions to the ^{13}CO signals.

The DTPC, DTPG, and cholesterol, ~32, 8, and 20 μmole, were dissolved in chloroform followed by chloroform removal with nitrogen gas and vacuum. The solid was suspended in 2 mL of 5 mM HEPES buffer (pH 7.0) with 0.01% NaN_3 preservative and large unilamellar vesicles were formed with 10 freeze-thaw cycles followed by extrusion through 100 nm diameter pores of a polycarbonate filter (Avestin, Ottawa, ON). A solution containing ~5 mg Fp in ~30 mL HEPES buffer was added dropwise into the vesicle solution followed by gentle stirring overnight. Earlier analytical ultracentrifugation data showed that Fp is a monomer in this buffer (1). Vesicles with bound Fp were pelleted by centrifugation at ~150000*g* for 4 h and the bound Fp:total lipid mole ratio is estimated to be ~1:60, based on an earlier study, with unbound Fp in the supernatant (5). The pellet was lyophilized and transferred to a NMR magic angle spinning rotor with 4 mm outer diameter, followed by sample hydration with ~20 μL water.

There are many experimental data that support very similar structures of membrane-bound Fp: (1) before lyophilization; and (2) after lyophilization and rehydration. In several earlier studies, the sample preparation was very similar to the present study except there weren't the final lyophilization/rehydration steps (2, 6, 7). As in the present study, these earlier studies had peak ^{13}C NMR chemical shifts that correlated with β sheet structure. In addition, the ^{13}CO linewidths in samples without lyophilization/hydration were typically 3-4 ppm, similar to the linewidths of the present study, **Table S1**. The β sheet ^{13}CO shifts and 3-4 ppm linewidths were also observed when Fp was incorporated in membrane by organic co-solubilization followed by removal of solvent, rehydration, and centrifugation (2, 5, 8). For this preparation approach, β sheet was predominant both when there wasn't and when there was subsequent lyophilization/rehydration. As noted in the Discussion in the main text, ^{13}CO NMR signals with β sheet shifts and 3-4 ppm linewidths were also observed for Fp in

membrane-bound Fp+hairpin protein; these samples were prepared without lyophilization/rehydration (9). As described in the Discussion, sums of registry populations estimated from long-time $\Delta S/S_0$ of Fp+hairpin samples matched semi-quantitatively the sums of registry populations of the present study (10). Other earlier studies provided some ^{13}CO-^{15}N REDOR data for Fp samples that were prepared without lyophilization/rehydration. Although labelings were different than in the present study, the $\Delta S/S_0$ values between studies can sometimes be correlated. For example, a sample that covered u=15-19 labelings had $\Delta S/S_0 = 0.13$ at $\tau_k = 24$ ms which is similar to the average $\langle \Delta S/S_0 \rangle^{u=15\text{-}19} = 0.12$ of the present study (11). Another sample in the earlier study covered u=18-22 labelings and its $\Delta S/S_0 = 0.09$ is similar to $\langle \Delta S/S_0 \rangle^{u=18\text{-}22} = 0.08$ of the present study. In another study with samples prepared without lyophilization/rehydration, the $\langle \Delta S/S_0 \rangle$ for a u=17-type labeled sample at $\tau_k = 24$, 32 ms was 0.16, 0.19 which are the same as $\langle \Delta S/S_0 \rangle^{u=17} = 0.16, 0.19$ in the present study, and the u=16 comparison is 0.16, 0.21 vs. 0.13, 0.18 in the present study (6).

REDOR NMR spectroscopy. Spectra were acquired with a 9.4 T spectrometer with Varian Infinity Plus console. The samples were maintained at ~ -30 ºC by cooling with nitrogen gas at -50 °C. This cooling helped to maintain sample stability and hydration, and reduced molecular motion so that internuclear dipolar couplings were close to the rigid values, with consequent larger signals from $^1H \rightarrow {}^{13}C$ cross-polarization (CP) and more accurate analysis of the ^{13}CO-^{15}N REDOR data (5, 12, 13). Cooling did not modify Fp structure, as evidenced by earlier NMR spectra showing very similar ^{13}C shifts of samples near ambient and cooled temperatures, with typical difference ≤ 0.5 ppm (12). The rotor with sample was in a probe tuned to ^1H, ^{13}C, and ^{15}N frequencies. The ^{13}C transmitter shift was typically at 153 ppm with external ^{13}C shift referencing done using the adamantane -$^{13}CH_2$ signal at 40.5 ppm (14). The REDOR experiment was done with 10 kHz magic angle spinning frequency, 2 s recycle delay between scans, and temporal sequence: (1) 50 kHz ^1H $\pi/2$ pulse; (2) 2.2 ms $^1H \rightarrow {}^{13}C$ CP with a 60 kHz ^1H field and 63-68 kHz ramped ^{13}C field; (3) time period τ_k which alternated between S_0 reference scans with refocusing 54 kHz ^{13}C π pulses at the end of each rotor cycle except the last cycle, and S_1 scans with ^{13}C-^{15}N dipolar recoupling because of ^{13}C π and 45 kHz ^{15}N π pulses at the end and in the middle of each rotor cycle, respectively; and (4) ^{13}C detection (15). The rf pulses were set using a lyophilized helical peptide containing a single labeled ^{13}CO-^{15}N spin pair with r = 4.1 Å (16). There was XY-8 phase cycling for the ^{13}C and ^{15}N π pulses, and 80 kHz ^1H TPPM decoupling during periods 3 and 4 (17, 18). The phase cycle of S_0 and S_1 acquisitions was: ^1H $\pi/2$, 0, 180, 0, 180; ^1H CP, 90, 90, 90, 90; ^{13}C CP, 270, 270, 180, 180; final ^{13}C π, 270, 270, 180, 180; receiver, 180, 0, 90, 270. Typically ~20,000 S_0 or S_1 acquisitions were summed for each τ_k = 2.2, 8.2, 16.2, 24.2, 32.2, 40.2, and 48.2 ms with k = 1, 2, 3, 4, 5, 6, and 7, respectively.

S_0 and S_1 data were separately processed with 200 Hz Gaussian line broadening, Fourier transformation, and baseline correction followed by integration about the ^{13}CO peak with 3 ppm window that was the same for all spectra of a single sample and resulted in $S_0(u,\tau_k)$ and $S_1(u,\tau_k)$, with u=8-24,28 and τ_k=2.2-48.2 ms, k=1-7. The uncertainty, $\sigma(u,\tau_k)$, was the root mean squared deviation (RMSD) of 24 different 3 ppm integrations of noise regions of both the S_0 and S_1 spectra. For either WT or V2E samples, the experimental $[S_1/S_0](u,\tau_k)$ ratios are the basis for determination of registry populations, f(t)'s, using χ^2 fitting that includes the $\sigma_{S1/S0}(u,\tau_k)$ calculated with error propagation. The data are typically presented as a dephasing buildup, i.e. $\Delta S/S_0 = 1 - S_1/S_0$.

f(t) fitting. **Fig**. **S2** and **Table S1** show that S_0 and ΔS ^{13}CO NMR peaks are very well-fitted by single Gaussian lineshapes with peak chemical shifts (δ_{peak}'s) characteristic of β sheet structure and FWHM linewidths (δ_{FWHM}'s) typically in the 3-4 ppm range. The f(t) fitting approach that follows relies on the

approximation that the $\Delta S/S_0$ data in **Table S2** are only due to β sheet Fp, i.e. no α helix contribution. The validity of this approximation is supported by the following calculation. Earlier Fp studies showed that ^{13}CO $\delta_{peak,\alpha\ helix} - \delta_{peak,\beta\ sheet} \geq 4$ ppm (16, 19). The α helix peaks are either not apparent or barely apparent in the **Fig. S1** spectra and the α helix:β sheet population ratio is estimated to be <1:9, i.e. <10% α helix population. We consider α helix and β sheet peaks with Gaussian lineshapes with δ_{FWHM} = 3.6 ppm and $\delta_{peak,\alpha\ helix} - \delta_{peak,\beta\ sheet}$ = 4 ppm. The 3 ppm integration window is centered on the β sheet peak and results in 0.0509 fraction of the total α helix intensity and 0.6938 fraction of the total β sheet intensity, with calculations done using the Gaussian Error Function. For α helix:β sheet population ratio = 1:9, the α helix:β sheet ratio of contributions to the S_0 data is 0.0084 which supports the approximation of $\Delta S/S_0$ data being only due to β sheet Fp. The absence of α helix contribution to $\Delta S/S_0$ is also supported by smaller $\Delta S/S_0$ for the u=19 vs. 20 sample, **Table S2**, despite shorter $^{13}CO\cdots H^{15}N$ distance in an α helix for u=19 $F8_CL12_N$ vs. u=20 $F8_CG13_N$ labeling, **Fig. 3** in the main text.

The total $S_0^{tot}(\tau_k)$ and total $S_1^{tot}(\tau_k)$ signals are each a sum of labeled (lb) and natural abundance (na) ^{13}CO signals. The $S_0^{lb}(\tau_k)$ is assigned to be 1.0 so that $S_0^{na}(\tau_k)$ = 0.33, based on the 30 other backbone carbonyl spins. Both the $S_1^{lb}(u,\tau_k)$, which is u-dependent, and $S_1^{na}(\tau_k)$ signals are sums of contributions from different ^{13}CO populations that experience different ^{13}CO-^{15}N dipolar dephasings. The $S_1^{lb}(u,\tau_k)$ signal includes an attenuated $S_1^{lb,na}(\tau_k)$ contribution from the lb ^{13}CO spins that experience dephasing from nearby na ^{15}N spins. The other $S_1^{lb}(u,\tau_k)$ signal is from lb ^{13}CO spins not near na ^{15}N spins and is divided into two categories that are registry-dependent: (1) $S_1^{lb,lb}(u,\tau_k)$ is the signal from lb ^{13}CO spins in registries with t values close to u, and is attenuated because of dipolar couplings to the nearby lb ^{15}N spins; and (2) $S_1^{lb,X}(u,\tau_k)$ is the S_1 signal of the lb ^{13}CO in all other registries and is not attenuated so that $S_1^{lb,X}(u,\tau_k) = S_0^{lb,X}(u,\tau_k)$. These other registries are collectively-designated as "X", and the specific t values in X are provided later in this section.

The $S_1^{na}(\tau_k)$ signal of the na ^{13}CO spins is similarly separated and includes one attenuated contribution that is denoted $S_1^{na,lb}(\tau_k)$ and is from the na ^{13}CO spins near lb ^{15}N spins. The S_1 signal of the other na ^{13}CO spins is not attenuated so that $S_1(\tau_k) = S_0(\tau_k)$. The lb ^{13}CO/na ^{15}N and na ^{13}CO/lb ^{15}N populations are calculated using the ^{13}C and ^{15}N na probabilities of 0.011 and 0.0037, respectively. These probabilities are small and only isolated spin pairs are considered. Selection of a specific pair in the β sheet as lb ^{13}CO/na ^{15}N or na ^{13}CO/lb ^{15}N is based on the magnitude of the ^{13}CO-^{15}N dipolar coupling, d, with d(Hz) = 3080/r(Å)^3, where r is the ^{13}CO-^{15}N internuclear distance. Among all samples, the smallest dephasing is for the WT u=28 ($F8_CA21_N$) sample, and this dephasing is used to validate a model of dephasing only due to lb ^{13}CO/na ^{15}N and na ^{13}CO/lb ^{15}N spin pairs and not lb ^{13}CO/lb ^{15}N pairs. For this model, the largest dephasing contributions are from the 4 pairs with r < 5 Å in a model β sheet. Three of the pairs are intra-strand with r = 1.3, 2.4, and 4.6 Å and one pair is inter-strand with r = 4.1 Å. For example, the labeled F8 ^{13}CO has natural abundance intra-strand ^{15}N at L9, F8, and G10, respectively, and natural abundance inter-strand $^{15}NH\cdots O^{13}C$ (F8) hydrogen bond. The ratio $\gamma = S_1/S_0$ is calculated for each pair, with a pair indexed by m = 1, 2, 3, or 4. More specifically, the $\gamma^{lb,na}(d_m,\tau_k)$ is calculated using an expression with n^{th}-order Bessel functions J_n of the first kind and the dimensionless parameter $\lambda_{m,k} = d_m \times \tau_k$ (20):

$$\gamma^{lb,na}(d_m,\tau_k) = \left[J_0\left(\sqrt{2}\lambda_{m,k}\right)\right]^2 - \left\{2 \times \sum_{n=1}^{5} \frac{\left[J_n\left(\sqrt{2}\lambda_{m,k}\right)\right]^2}{16n^2-1}\right\} \quad \text{(S1)}$$

The $S_0^{lb,na} = 4 \times 0.0037 = 0.0148$ and the calculated $S_1^{lb,na}$:

$$S_1^{lb,na}(\tau_k) = 0.0037 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m, \tau_k) \tag{S2}$$

The four CO-N distances that are <5Å are the same for lb ^{13}CO-na ^{15}N and na ^{13}CO-lb ^{15}N spin pairs, so the two types of spin pairs also share the same four d_m and same four $\gamma(d_m,\tau_k)$, i.e. $\gamma^{lb,na}(d_m,\tau_k) = \gamma^{na,lb}(d_m,\tau_k)$. The $S_0^{na,lb} = 4 \times 0.011 = 0.044$ and the calculated $S_1^{na,lb}$:

$$S_1^{na,lb}(\tau_k) = 0.011 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m, \tau_k) \tag{S3}$$

For any sample, the non-dephased S_1^{na} signal is $0.33 - (4 \times 0.011) = 0.286$. For a sample for which $S_0^{lb,lb} = 0$, i.e. no populated registries with t close to u, the non-dephased $S_1^{lb,X} = S_0^{lb,X} = 1 - (4 \times 0.0037) = 0.9852$ so that:

$$S_1^{tot}(\tau_k) = 1.2712 + S_1^{lb,na}(\tau_k) + S_1^{na,lb}(\tau_k) = 1.2712 + \left[0.0147 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m, \tau_k)\right] \tag{S4}$$

Eq. **S4** is incorporated into the dephasing expression:

$$\frac{\Delta S}{S_0}(\tau_k) = \frac{[1.33 - S_1^{tot}(\tau_k)]}{1.33} \tag{S5a}$$

and the calculated natural abundance dephasing is:

$$\frac{\Delta S}{S_0}(\tau_k) = 0.0442 - \left[0.01105 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m, \tau_k)\right] \tag{S5b}$$

which is **Eq**. **1** in the main text. The $\Delta S/S_0$ calculated with **Eq**. **S5b** agree quantitatively with the experimental $\left(\frac{\Delta S}{S_0}\right)^{exp}(u = 28, \tau_k)$ in **Fig**. **3** in the main text.

For the other samples, the $S_1^{na}(\tau_k)$ is the sum of the dephased signal from the 4 na ^{13}CO spins sites close to a lb ^{15}N spin and the remaining undephased signal = 0.286 from the other na ^{13}CO spins. The combined $S_1^{na} + S_1^{lb,na}$:

$$S_1^{na}(\tau_k) + S_1^{lb,na}(\tau_k) = 0.286 + \left[0.0147 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m, \tau_k)\right] \tag{S6}$$

The population of lb ^{13}CO that have not been dephased by na ^{15}N is 1 – [4 × 0.0037] = 0.9852. Calculation of the S_1^{lb} signal is done with two different models, referred to as constrained and unconstrained. For constrained fitting, all neighboring strands within a single sheet have the same registry and value of t, and f(t) is the fractional population of sheets with a particular t. The determination of the f(t) populations is done by fitting the $S_1^{lb,lb} + S_1^{lb,X}$ signal contributions from the u=8-24 samples. The $S_1^{lb,lb}$ contributions are from lb ^{13}CO spins in sheets with the t = u, u+1, u–1, u+2, or u–2 registries, i.e. registries with substantial lb ^{13}CO-lb ^{15}N dipolar coupling. The $\gamma_{t=u}^{lb,lb}(\tau_k)$, $\gamma_{t=u\pm1}^{lb,lb}(\tau_k)$, and $\gamma_{t=u\pm2}^{lb,lb}(\tau_k)$ values (where $\gamma=S_1/S_0$) were calculated using the SIMPSON simulation program and the relevant geometry with one ^{13}CO and two ^{15}N spins (21). As one example, **Fig**. **2a** in the main text displays schematic representations of the constrained t=16 registry for the u=16 sample, and the geometry of the three spins. In general, the spin geometries were based on atomic coordinates of the crystal structure of β barrel outer membrane protein G (OMPG, PDB file 2IWW). Simulation inputs were determined using these coordinates and the SIMMOL program and included the dipolar couplings and the Euler angles for each coupling vector and for the principal axis system of the ^{13}CO chemical shift anisotropy (CSA), as described in an earlier study (22, 23). The ^{13}CO CSA principal values were 247, 176, and 99 ppm. Other SIMPSON simulation inputs were the spinning frequency and pulses in the dephasing time of the REDOR pulse program including ^{13}C and ^{15}N transmitter shifts, pulse lengths, rf fields, and phases. Each $\gamma(\tau_k)$ was an average from ~10 SIMPSON simulations that were each based on coordinates of different atoms in OMPG. Neither ^{1}H's nor relaxation were considered in the simulations. **Table S4** presents the $\gamma_{t=u}^{lb,lb}(\tau_k)$, $\gamma_{t=u\pm1}^{lb,lb}(\tau_k)$, and $\gamma_{t=u\pm2}^{lb,lb}(\tau_k)$ values determined from these simulations. The approximations $\gamma_{t=u+1}^{lb,lb}(\tau_k) = \gamma_{t=u-1}^{lb,lb}(\tau_k)$ and $\gamma_{t=u+2}^{lb,lb}(\tau_k) = \gamma_{t=u-2}^{lb,lb}(\tau_k)$ are based on the differences in γ values between the two spin geometries being smaller than the differences due to variations in β sheet structure.

The lb ^{15}N spins in the other (X) registries, i.e. $t \neq u, u\pm1, u\pm2$, are considered too distant to dephase the lb ^{13}CO spins so that $S_1^{lb,X}(u,\tau_k) = S_0^{lb,X}(u,\tau_k) = f_X(u)$ and:

$$f_X(u) = 1 - \sum_{t=u-2}^{t=u+2} f(t) \tag{S7}$$

$$S_1^{lb,X}(u,\tau_k) + S_1^{lb,lb}(u,\tau_k) = 0.9852 \times \left\{ f_X(u) + \sum_{t=u-2}^{t=u+2} \left[f(t) \times \gamma_t^{lb,lb}(\tau_k) \right] \right\} \tag{S8}$$

$$S_1^{tot}(u,\tau_k) = S_1^{na}(\tau_k) + S_1^{lb,na}(\tau_k) + S_1^{lb,lb}(u,\tau_k) + S_1^{lb,X}(u,\tau_k) \tag{S9}$$

The $\frac{S_1^{tot}(u,\tau_k)}{1.33} = \left(\frac{S_1}{S_0}\right)^{exp}(u,\tau_k)$ so that:

$$1.33 \times \left(\frac{S_1}{S_0}\right)^{exp}(u,\tau_k) = 0.286 + \left[0.0147 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m,\tau_k) \right] + (0.9852) \times \left\{ 1 + \sum_{t=u-2}^{t=u+2} \left[f(t) \times \left\{ \gamma_t^{lb,lb}(\tau_k) - 1 \right\} \right] \right\}$$

(S10a)

which is algebraically equivalent to:

$$\left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) = \left\{0.7408 \times \sum_{t=u-2}^{t=u+2}\left[f(t) \times \left\{1-\gamma_t^{lb,lb}(\tau_k)\right\}\right]\right\} + \left\{0.01105 \times \left[\sum_{m=1}^{4} 1-\gamma^{lb,na}(d_m,\tau_k)\right]\right\}$$

(S10b)

Algebra is used to place the f(t) terms on the left-side and the other terms on the right-side:

$$\sum_{t=u-2}^{t=u+2}\left\{f(t) \times \left[1-\gamma_t^{lb,lb}(\tau_k)\right]\right\} = 1.350 \times \left\{\left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) - \left[0.01105 \times \left(\sum_{m=1}^{4} 1-\gamma^{lb,na}(d_m,\tau_k)\right)\right]\right\}$$

(S11)

Eq. **S10b** shows that $\Delta S/S_0(u)$ is proportional to the $(1 - \gamma_t^{lb,lb})$ weighted f(t) sum (first term in parentheses on right-side) plus a natural abundance contribution (second term). Relative to the t = u ± 1, u ± 2 terms in the f(t) sum, t = u has the smallest $\gamma_t^{lb,lb}$ and is therefore the most heavily-weighted term in the sum, **Table S4**. **Eq**. **S11** similarly shows that the $(1 - \gamma_t^{lb,lb})$ weighted f(t) sum on the left-side is proportional to $\Delta S/S_0(u)$ after subtracting the natural abundance contribution. The f(t), t=8-24, are determined by χ^2 fitting using Python code and the $\left(\frac{S_1}{S_0}\right)^{exp}(u,\tau_k)$ data with $\sigma_{\frac{S_1}{S_0}}^{exp}(u,\tau_k)$ uncertainties. The f(6), f(7), f(25), and f(26) are set to 0 in the fittings. The f(6), f(7), f(25), and f(26) are set to 0 in the fittings and the basis for this approximation is described in the following text. Because each f(t) contributes to $\Delta S/S_0(u)$ for five values of u, the f(t) fitting must be done simultaneously for all $\Delta S/S_0(u)$. Also, as noted in the main text and in **Table S4**, the largest contributions to $\Delta S/S_0(u)$ are from the population of Fp strands with registry t=u. There aren't $\Delta S/S_0$ data for u = 6, 7, 25, or 26 but registries with corresponding t values contribute to $\Delta S/S_0$ for u = 8, 9, 23, 24. **Table S2** shows that the $\Delta S/S_0$ generally decrease as u approaches 8 or 24, i.e. at τ_k = 40.2 ms for u = 12, 11, 10, 9, 8, the $\Delta S/S_0^{WT}$ = 0.170, 0.097, 0.062, 0.068, 0.052, and $\Delta S/S_0^{V2E}$ = 0.086, 0.053, 0.050, 0.058, 0.052. For u = 20, 21, 22, 23, 24, the $\Delta S/S_0^{WT}$ = 0.177, 0.072, 0.084, 0.089, 0.052, and $\Delta S/S_0^{V2E}$ = 0.379, 0.198, 0.103, 0.103, 0.090. The $\Delta S/S_0$ for u=8, 24 are similar to $\Delta S/S_0^{WT}(u=28) = \Delta S/S_0^{na}$ = 0.043, **Table S6**. Given these trends and the absence of u=t data for t = 6, 7, 25, 26, the approximation is made that f(t)=0 for these t values. With this approximation, WT and V2E fittings result in f(8) < 0.003 and f(24) = 0, **Table S8**. The $\Delta S/S_0(u=8)$ values only depend on f(8), f(9), f(10), and $\Delta S/S_0(u=24)$ values only depend on f(22), f(23), f(24). The f(8) and f(24) from fittings with the approximation are likely larger than those from fittings without the approximation, based on the possibility in the latter fittings that f(6), f(7), f(25), and/or f(26) are non-zero. The finding that f(8) and f(24) are always negligible with the approximation further supports the validity of the approximation.

Somewhat smaller χ^2 are sometimes obtained when the SIMPSON-calculated $\gamma_t^{lb,lb}(\tau_k)$ are all multiplied by a scaling parameter, b, with 0.95 < b < 1.

$$\sum_{t=u-2}^{t=u+2}\left\{f(t) \times \left[1-b \times \gamma_t^{lb,lb}(\tau_k)\right]\right\} = 1.350 \times \left\{\left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) - \left[0.01105 \times \left(\sum_{m=1}^{4} 1-\gamma^{lb,na}(d_m,\tau_k)\right)\right]\right\}$$

(S12)

This is the same as **Eq**. **2** in the main text. Better fitting with smaller $\gamma_t^{lb,lb}(\tau_k)$ could reflect inclusion of contributions from couplings to more distant lb ^{15}N spins.

The second approach is unconstrained fitting for which there can be multiple registries within a single sheet, and f(t) is the fractional population of adjacent strands with a particular t. The spin system is a single ^{13}CO spin in a central Fp molecule and the ^{15}N spins in the two neighboring Fp molecules in the β sheet. These neighboring molecules are denoted 1 and 2, with assignment of molecule 1 vs. 2 based on having vs. not having a backbone NH hydrogen-bonded to the lb ^{13}CO of the central molecule. For **Fig**. **2** in the main text, molecule 1 is above the central molecule and molecule 2 is below the central molecule. Registries are indexed by t_1 and t_2, the total number of residues of molecule 1 or 2, respectively, that could be hydrogen-bonded to the central molecule. **Fig**. **2a** displays the schematic β sheet with $t_1 = t_2 = u = 16$ and **Fig**. **2b** displays the schematic sheet with three adjacent Fp molecules in an unconstrained sheet with u=16, t_1=17, and t_2=19. Relative to the constrained model, there are a larger number of distinct lb ^{13}CO/lb ^{15}N spin geometries in the unconstrained model and the unconstrained analysis is done based on $\gamma_{t1,t2}(\tau_k) < 1$ only when t_1 = u, u+1, or u-1 and/or t_2 = u, u+1, or u-1. The t_1=X means $t_1 \neq u, u\pm1$ and t_2=X means $t_2 \neq u, u\pm1$, and $\gamma_{X,X}^{lb,lb}(\tau_k) = 1$. The SIMPSON-calculated $\gamma_{t1,t2}^{lb,lb}(\tau_k)$ are presented in **Table S4**. The intermolecular β sheet structure means the simulation spin system is one lb ^{13}CO and two lb ^{15}N spins when $t_1 \neq$X and $t_2 \neq$X, like in **Fig**. **2a**, and one lb ^{13}CO and one lb ^{15}N when $t_1 \neq$X or $t_2 \neq$X, like in **Fig**. **2b**. **Eqs**. **S1**-**S6** are valid for the unconstrained analysis and **Eqs**. **S7**, **S8** become:

$$f_X(u) = 1 - \sum_{t=u-1}^{t=u+1} f(t) \tag{S13}$$

$$S_1^{lb,X}(u,\tau_k) + S_1^{lb,lb}(u,\tau_k) = 0.9852 \times \left\{ \sum_{t_1=u-1}^{u+1,X} \quad \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times \gamma_{t_1,t_2}^{lb,lb}(\tau_k) \right] \right\} \tag{S14}$$

And:

$$S_1^{tot}(u,\tau_k) = S_1^{na}(\tau_k) + S_1^{lb,na}(\tau_k) + S_1^{lb,lb}(u,\tau_k) + S_1^{lb,X}(u,\tau_k) \tag{S15}$$

There is a single f(t) population for each t, i.e. when t_1=t_2, f(t_1)=f(t_2). The $S_1^{na}(\tau_k) + S_1^{lb,na}(\tau_k)$ are described by **Eq**. **S6** and $\frac{S_1^{tot}(u,\tau_k)}{1.33} = \left(\frac{S_1}{S_0}\right)^{exp}(u,\tau_k)$ so that:

$$1.33 \times \left(\frac{S_1}{S_0}\right)^{exp}(u,\tau_k) = 0.286 + \left[0.0147 \times \sum_{m=1}^{4} \gamma^{lb,na}(d_m,\tau_k)\right] + (0.9852) \times \left\{\sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times \gamma_{t_1,t_2}^{lb,lb}(\tau_k)\right]\right\} \quad \text{(S16a)}$$

which is algebraically equivalent to:

$$\left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) = 0.7408 \times \left\{1 - \sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times \gamma_{t_1,t_2}^{lb,lb}(\tau_k)\right]\right\} + \left\{0.01105 \times \left[\sum_{m=1}^{4} 1 - \gamma^{lb,na}(d_m,\tau_k)\right]\right\} \quad \text{(S16b)}$$

Algebra is used to place the f(t) terms on the left-side and the other terms on the right-side:

$$\sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times \gamma_{t_1,t_2}^{lb,lb}(\tau_k)\right] = 1 - 1.350 \times \left\{\left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) - 0.01105 \times \left[\sum_{m=1}^{4} 1 - \gamma^{lb,na}(d_m,\tau_k)\right]\right\} \quad \text{(S17)}$$

Eq. **S16b** shows that $\Delta S/S_0(u)$ is proportional to one minus the $\gamma_{t1,t2}^{lb,lb}$ weighted sum of $f(t_1) \times f(t_2)$ products (first term in parentheses on right-side) plus a natural abundance contribution (second term). **Eq**. **S17** similarly shows that the $\gamma_{t1,t2}^{lb,lb}$ weighted sum of $f(t_1) \times f(t_2)$ products is equal to one minus a term proportional to $\Delta S/S_0(u)$ minus the natural abundance contribution.

The f(t), t=8-24, are determined by fitting to the $\left(\frac{S_1}{S_0}\right)^{exp}(u,\tau_k)$ data, with f(7) = f(25) = 0. Similar to constrained fittings, unconstrained fittings are sometimes done with scaled $\gamma_{t1,t2}(\tau_k)$:

$$\sum_{t_1=u-1}^{u+1,X} \sum_{t_2=u-1}^{t_2=u+1,X} \left[f(t_1) \times f(t_2) \times b_{t_1,t_2} \times \gamma_{t_1,t_2}^{lb,lb}(\tau_k)\right] = 1 - 1.350 \times \left\{\left(\frac{\Delta S}{S_0}\right)^{exp}(u,\tau_k) - 0.01105 \times \left[\sum_{m=1}^{4} 1 - \gamma^{lb,na}(d_m,\tau_k)\right]\right\} \quad \text{(S18)}$$

Eq. **S18** is the same as **Eq**. **3** in the main text. The "b=0.98" scaling factors are: $b_{t1=u,u\pm1,t2=u,u\pm1}$ = 0.98, $b_{t1=u,u\pm1,t2=X}$ = $b_{t1=X,t2=u,u\pm1}$ = 0.99, and $b_{t1=X,t2=X}$ = 1.

Free energy fittings of f(t). For either WT or V2E, the substantial REDOR $\Delta S/S_0$ data for many differently-labeled samples are the basis for the broad distributions of populated registries and for the population weighting towards longer registries for V2E vs. WT, **Figs**. **4**, **5** in the main text and **Tables S5**, **S8**. The reproducibility of the sample preparation and REDOR NMR approaches is supported by typical agreement within uncertainties between $\Delta S/S_0$ values from replicate samples, **Table S3**. The $f(t)^{WT}$ and $f(t)^{V2E}$ distributions from unconstrained fittings with b=0.98 were quantitatively-analyzed with thermodynamic equilibrium models with RT=0.6 kcal/mole:

$$f(t)^{WT} = C^{WT} \times exp\left\{\frac{-\left[\left(G_{\beta}^{WT} \times t\right) + \left(G_{Leu}^{WT} \times L(t)\right) + \left(G_{sc}^{WT}(t) \times g^{WT}\right)\right]}{RT}\right\} \quad \text{(S19)}$$

$$f(t)^{V2E} = C^{V2E} \times exp\left\{\frac{-\left[\left(G_{\beta}^{V2E} \times t\right) + \left(G_{Leu}^{V2E} \times L(t)\right)\right]}{RT}\right\} \quad \text{(S20)}$$

These are the same as **Eqs**. **4**, **5** in the main text. The G_β is the free energy-per-residue of β sheet formation. The G_{Leu} is the free energy when leucines are aligned in adjacent strands in a β sheet, with L(t)=1 when at least one residue position is aligned, and L(t)=0 in the absence of such alignment. These leucines are bolded in the schematic registries of **Fig**. **3** in the main text. The $G_{sc}(t)^{WT}$ is the sum of free energies of membrane insertion of sidechains for residues V2 to t-1, with sidechain energy relative to Ala, and g^{WT} is a scaling factor that is <1 and may help to account for the positive free energy of membrane insertion of the Fp backbone (24). There are earlier studies that support membrane insertion of WT Fp starting near V2 (5, 7). The $f(t)^{WT}$, t=11-20, were fitted with **Eq**. **S19** and encompass ~95% of the total WT Fp population. The L(t)=1 for t=13, 15, 17, 18, and 20, and L(t)=0 for t=11, 12, 14, 16, and 19. Fitting was done using RT = 0.6 kcal/mole and variation of the parameters C^{WT}, G_β^{WT}, G_{Leu}^{WT}, and g^{WT}. **Fig**. **5a** in the main text displays bar plot comparison between $f(t)^{WT}$ and values from **Eq**. **S19** best-fitting, along with a bar plot of the three contributions to G(t). **Table S9** presents the numerical values. The fitting R^2 = 0.88, the typical magnitude of a residual is ~0.01, G_β^{WT} = -0.113 ± 0.038 kcal/mole, G_{Leu}^{WT} = -0.350 ± 0.079 kcal/mole, and g^{WT} = 0.129 ± 0.040. The negative-signed G_{Leu}^{WT} contribution is illustrated by larger $f(t)^{WT}$ for t=15 vs. 14 or 16. These three registries all have the same value of $g^{WT} \times G_{sc}(t)^{WT}$, but differ in the presence (t=15) vs. absence (t=14, 16) of aligned leucines. The extension of the β sheet for registry t over the region from A1 to t is supported by significant intensity assigned to β sheet chemical shifts in ^{13}C NMR spectra of membrane-bound Fp, with ^{13}C-labeled sites between A1 and A21 (6, 25). Models different than **Eq**. **S19** resulted in poorer fitting, i.e. smaller R^2. Examples of these alternate models were: (1) not including G_β^{WT} and/or not including G_{Leu}^{WT} contributions to free energy; (2) setting L(t) as the number of leucine and/or phenylalanine residue positions that are aligned in the registry; (3) not including the g^{WT} parameter; and (4) calculating $G_{sc}^{WT}(t)$ as the sum starting at a specific residue between G3 and L7 and ending at the corresponding residue between t-2 and t-6.

The $f(t)^{V2E}$, t=15-21, were also fitted and encompass ~95% of the total V2E Fp population. Fittings were first done using **Eq**. **S19**, with $G_{sc}(t)^{V2E}$ calculated as the sum of sidechain free energies. Separate fittings were done for insertion of the I4→t-3, A6→t-5, and L7→t-6 regions, and in all cases, the best-fit $g^{V2E} \approx 0$. The $f(t)^{V2E}$ fitting was then done using **Eq**. **S20**, i.e. without a contribution to free energy from membrane insertion, and $R^2 = 0.78$, $G_{\beta}^{WT} = -0.184 \pm 0.056$ kcal/mole, and $G_{Leu}^{WT} = -1.21 \pm 0.44$ kcal/mole. The typical residual had magnitude ≈ 0.01 except for t=16 which fitted poorly with residual ≈ 0.1. When $f(16)^{V2E}$ wasn't included, $R^2 = 0.98$, $G_{\beta}^{V2E} = -0.195 \pm 0.021$ kcal/mole, and $G_{Leu}^{V2E} = -1.40 \pm 0.22$ kcal/mole, **Fig**. **6b** in the main text and **Table S10**.

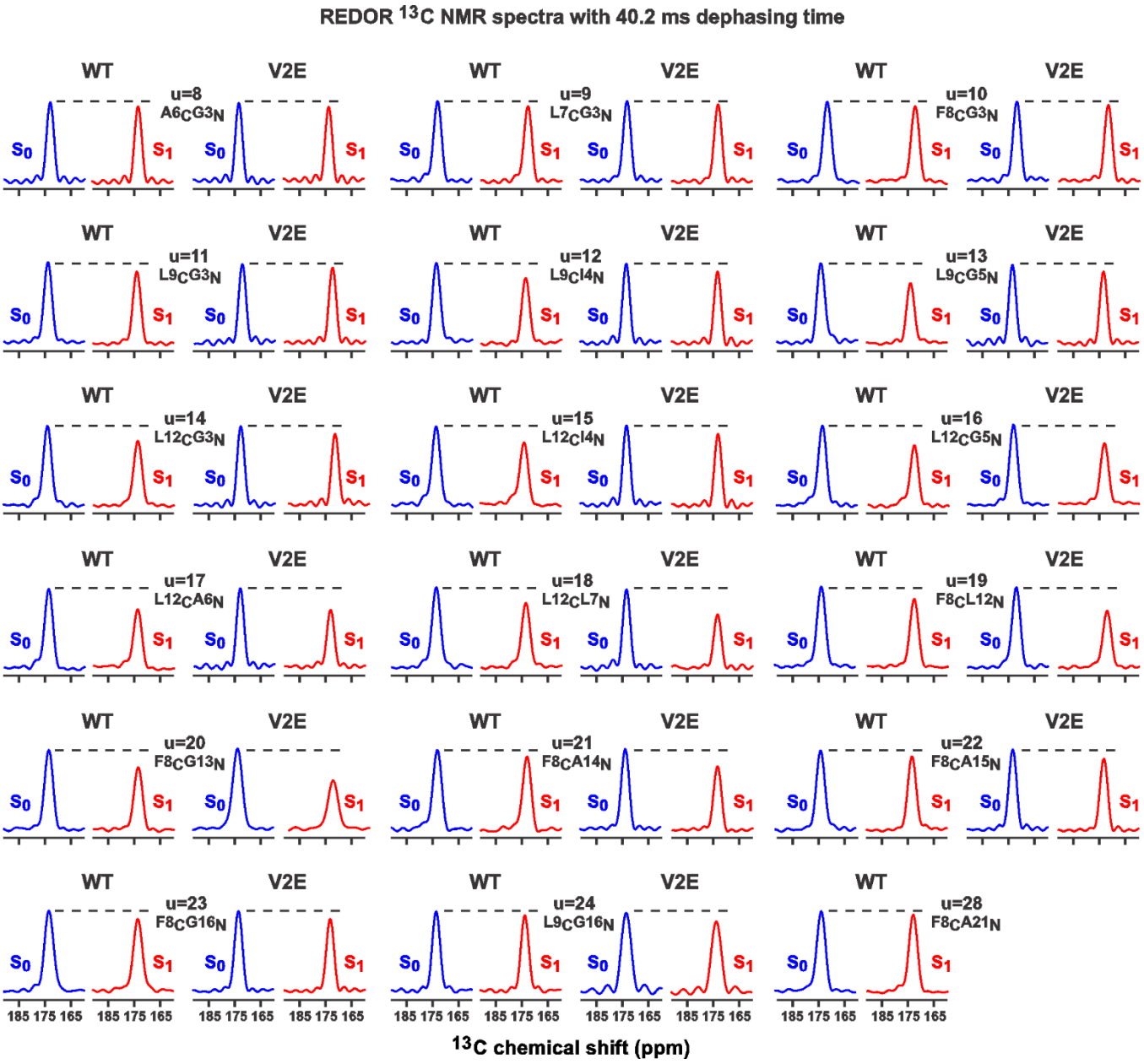

Figure S1. Plots of the ^{13}CO region of REDOR NMR S_0 (blue) and S_1 (red) spectra with τ_k = 40.2 ms dephasing time. The samples are membrane-bound Fp with u=8-24 labeling, WT or V2E, and u=28, WT, i.e 18 different WT and 17 different V2E samples. Each column of spectra is for either WT or V2E samples. The ^{13}CO and ^{15}N labelings are shown for each Fp, e.g. $L12_{C}G5_{N}$ for u=16. The peak intensities are scaled to be the same height for all the S_0 spectra and are marked with dashed lines. The WT u=28 sample has S_1/S_0 intensity ratio = 0.957, which matches the ratio calculated using only natural abundance dephasing, cf. **Eq**. **1**, **Fig**. **3**, and **Table S6**. Spectra were processed with 100 Hz Gaussian line broadening and 5th order polynomial baseline correction.

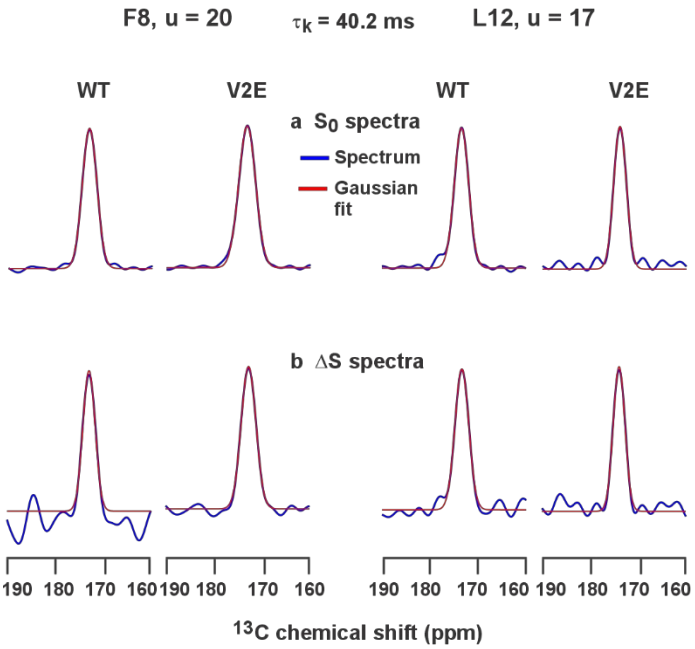

Figure S2. Plots of **a** S_0 and **b** ΔS ^{13}CO REDOR NMR spectra, blue traces, and Gaussian fits, red traces. Spectra were scaled to have similar vertical heights. **Table S1** presents the peak shifts and linewidths from fittings for all samples.

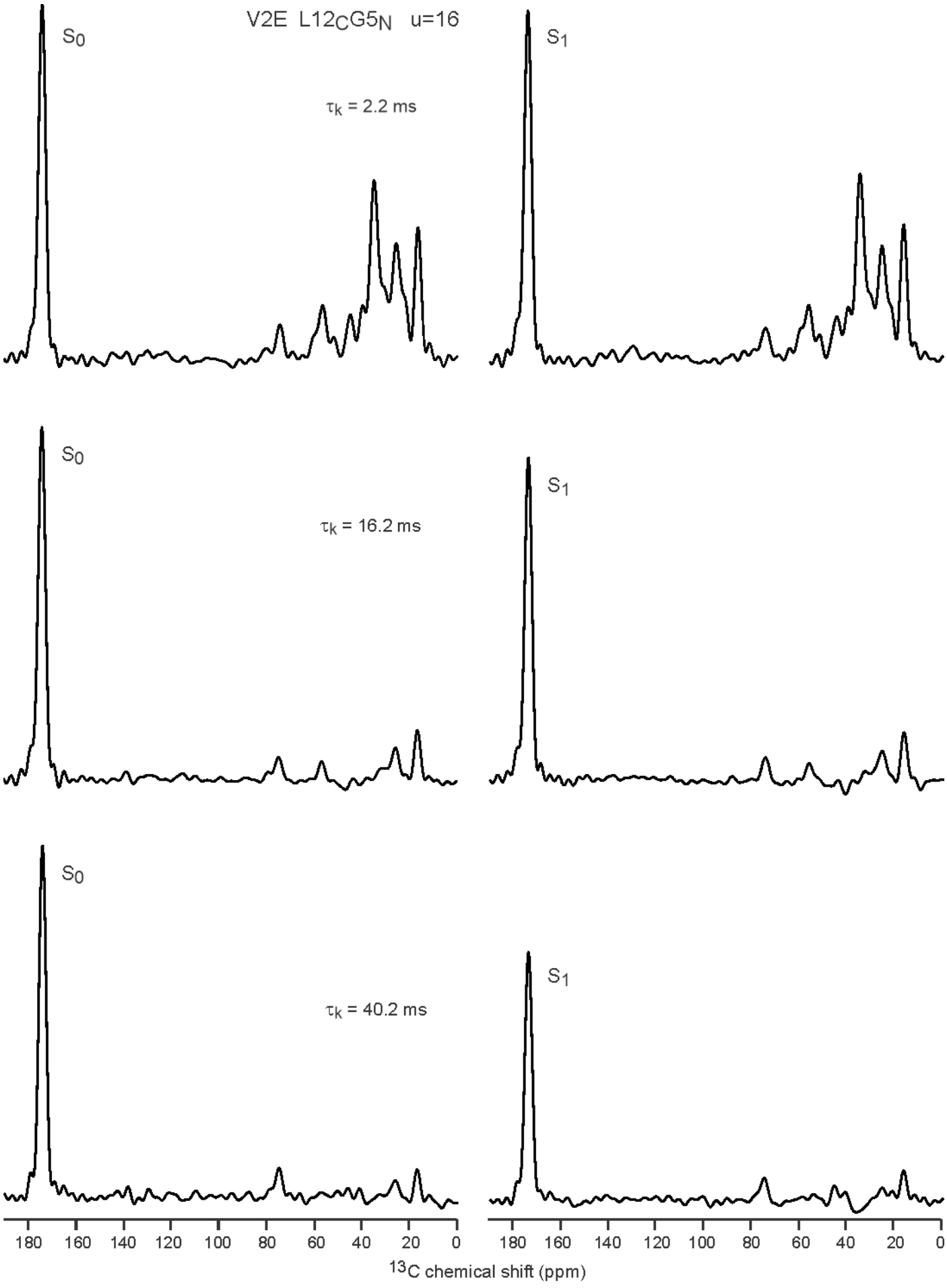
V2E L12CG5N u=16
S0
S1
τk = 2.2 ms
S0
S1
τk = 16.2 ms
S0
S1
τk = 40.2 ms
180 160 140 120 100 80 60 40 20 0
180 160 140 120 100 80 60 40 20 0
13C chemical shift (ppm)

Figure S3. ^{13}C REDOR NMR spectra of the V2E $L12_CG5_N$ (u=16) sample. The S_0 spectra are left panels and S_1 spectra are right panels, and dephasing time τ_k = 2.2, 16.2, and 40.2 ms for top, middle, and bottom rows, respectively. For all τ_k values, the spectra have been scaled to have the same S_0 ^{13}CO peak height. Spectra were processed with 200 Hz Gaussian line broadening and 5th order polynomial baseline correction. The transmitter was at 153 ppm and pulse parameters including cross-polarization were optimized specifically for the isotropic ^{13}CO signal at 174 ppm without consideration of ^{13}C signals in any other spectral region. The MAS frequency was 10 kHz and the m = -1 sideband at 74 ppm is observed in all spectra. Dephasing is only observed for the isotropic ^{13}CO signal. Relative to ^{13}CO, the other ^{13}C signals are both less intense and decay much faster with τ_k. These observations for the non-^{13}CO signals are likely due to lack of optimization, large off-resonance effects for ^{13}C pulses, and poorer ^{1}H decoupling because of directly-bonded ^{1}H spins.

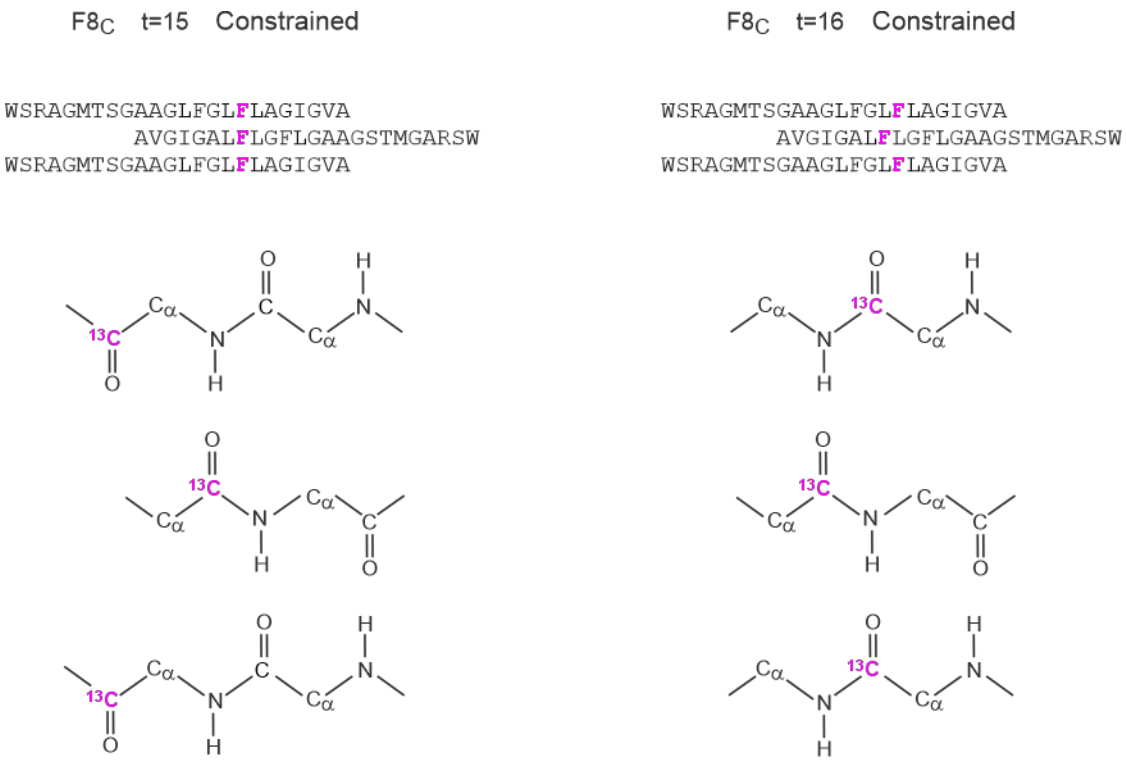

Figure S4. Schematic representations of constrained Fp sheets with F8 ^{13}CO labeling and t=15 (left) and t=16 (right) registries. In the earlier NMR study of Ib ^{13}CO-Ib ^{13}CO intermolecular dipolar dephasing of membrane-bound Fp+hairpin, one of the samples had a single ^{13}CO label at F8, i.e. v=8, and showed long-time dephasing of ~0.3 (10). The largest contributions to the dephasing were from the Fp+hairpin molecules with t=15 and t=16 registries, i.e. t=2v-1 and v, which have the smallest ^{13}CO-^{13}CO distances, as shown schematically in the Figure. The long-time dephasing of the Fp+hairpin sample is similar to the f(15) + f(16) = 0.23 sum for constrained or unconstrained fittings in the present study, **Table S5**. This agreement supports the $f(t)^{WT}$ distribution of HIV gp41 in its final hairpin state to be similar to the $f(t)^{WT}$ distribution of the present study.

Table S1. ^{13}CO peak chemical shifts and linewidths determined from REDOR spectra with τ_k = 40.2 ms [a]

Residue	U	WT				V2E			
		S_0		ΔS		S_0		ΔS	
		δ_{peak}	δ_{FWHM}	δ_{peak}	δ_{FWHM}	δ_{peak}	δ_{FWHM}	δ_{peak}	δ_{FWHM}
		ppm		ppm		ppm		ppm	
A6	8	173.87	2.95		.	173.94	2.81		
L7	9	173.59	3.34			173.78	3.21		
	10	172.85	3.44			172.75	3.18		
	19	172.58	3.63	172.35	3.57	173.00	3.55	172.66	3.39
F8	20	173.07	3.58	173.13	3.12	173.12	4.15	172.79	3.59
	21	172.78	3.85			173.05	3.24	173.40	3.14
	22	173.17	3.63			173.02	3.27		
	23	173.14	3.88			172.94	3.21		
Average(RMSD)		172.93(23)	3.67(17)	172.74(55)	3.35(32)	172.98(13)	3.43(38)	172.95(40)	3.37(23)
	11	173.50	3.38			173.22	2.96		
L9	12	173.39	3.29	173.62	3.12	173.74	2.87		
	13	173.63	3.38	173.79	3.02	173.88	2.87		
	24	173.33	3.26			173.10	2.86		
Average(RMSD)		173.46(13)	3.33(6)	173.71(12)	3.07(7)	173.49(38)	2.89(5)		
	14	173.49	3.51	173.24	3.08	173.69	2.98		
	15	173.54	3.66	173.77	3.29	174.06	3.59		
L12	16	173.75	3.53	173.61	3.45	173.92	3.02	174.04	3.16
	17	173.32	3.54	173.33	3.38	173.83	2.98	173.96	2.86
	18	173.31	3.70			173.95	3.04	173.93	2.94
Average(RMSD)		173.48(18)	3.59(9)	173.49(25)	3.30(16)	173.89(14)	3.12(27)	173.98(6)	2.99(16)

[a] ^{13}CO peak shifts and full-width at half-maximum linewidths from Gaussian lineshape fitting of REDOR S_0 and $\Delta S \equiv S_0 - S_1$ spectra acquired with τ_k = 40.2 ms, see **Fig**. **S2** for example fittings. A ΔS spectrum was fitted when the spectral signal-to-noise appeared by eye to be sufficient, which in practice typically meant that $\Delta S/S_0 > 0.15$. Averages for a particular ^{13}CO site with RMSD's in parentheses are represented using the convention that the RMSD corresponds to the right-most digits in the average, e.g. 173.48(18) means 173.48 ± 0.18.

Table S2. Experimental REDOR $\Delta S/S_0$ values [a]

	WT							V2E						
	Dephasing time (ms)							Dephasing time (ms)						
u	2.2	8.2	16.2	24.2	32.2	40.2	48.2	2.2	8.2	16.2	24.2	32.2	40.2	48.2
8	0.006(7)	0.017(6)	0.030(6)	0.038(9)	0.037(11)	0.052(12)	0.057(16)	0.009(10)	0.026(10)	0.028(10)	0.018(10)	0.053(14)	0.052(10)	0.055(12)
9	0.012(6)	0.009(6)	0.032(9)	0.033(9)	0.047(11)	0.068(12)	0.063(18)	0.014(10)	0.005(10)	0.027(10)	0.032(10)	0.047(10)	0.058(12)	0.062(17)
10	0.015(10)	0.022(7)	0.033(12)	0.046(16)	0.039(19)	0.062(17)	0.111(21)	0.024(10)	0.018(10)	0.024(10)	0.041(12)	0.047(12)	0.050(13)	0.068(16)
11	0.014(5)	0.026(6)	0.046(9)	0.066(12)	0.081(11)	0.097(17)	0.151(22)	-0.010(10)	0.006(11)	0.031(10)	0.040(12)	0.050(11)	0.053(17)	0.062(13)
12	0.011(9)	0.016(9)	0.060(9)	0.095(12)	0.113(11)	0.170(16)	0.215(23)	0.001(10)	0.012(10)	0.022(10)	0.030(14)	0.059(10)	0.086(10)	0.085(16)
13	0.010(5)	0.034(8)	0.067(12)	0.102(15)	0.172(14)	0.218(16)	0.256(25)	0.002(13)	0.044(17)	0.038(10)	0.047(16)	0.063(20)	0.078(25)	0.129(25)
14	0.003(6)	0.033(9)	0.088(12)	0.109(13)	0.138(14)	0.171(11)	0.235(21)	0.007(17)	0.043(13)	0.034(10)	0.061(10)	0.073(12)	0.114(14)	0.154(12)
15	0.008(8)	0.043(8)	0.093(12)	0.123(11)	0.173(14)	0.215(19)	0.244(19)	0.004(13)	0.039(10)	0.046(16)	0.095(10)	0.143(10)	0.175(11)	0.195(10)
16	0.012(7)	0.044(9)	0.090(10)	0.128(8)	0.179(11)	0.238(15)	0.253(15)	0.019(10)	0.034(12)	0.073(10)	0.129(12)	0.197(15)	0.245(14)	0.287(23)
17	0.004(9)	0.058(10)	0.099(7)	0.155(13)	0.192(11)	0.247(16)	0.275(21)	0.019(12)	0.035(20)	0.078(20)	0.175(19)	0.238(16)	0.271(14)	0.310(20)
18	0.011(7)	0.055(11)	0.085(11)	0.126(10)	0.174(12)	0.188(20)	0.201(21)	0.025(17)	0.057(13)	0.113(12)	0.194(12)	0.254(19)	0.302(18)	0.303(22)
19	0.010(7)	0.022(5)	0.064(6)	0.082(8)	0.131(11)	0.145(13)	0.157(13)	0.013(11)	0.046(10)	0.069(14)	0.144(10)	0.213(10)	0.280(12)	0.346(12)
20	0.022(12)	0.017(12)	0.068(9)	0.116(15)	0.161(12)	0.177(24)	0.175(15)	0.009(10)	0.056(10)	0.146(10)	0.262(10)	0.330(10)	0.379(14)	0.398(17)
21	0.010(9)	0.005(10)	0.028(11)	0.052(13)	0.074(13)	0.072(16)	0.112(16)	0.007(10)	0.043(10)	0.075(10)	0.130(10)	0.179(10)	0.198(15)	0.257(11)
22	0.011(9)	0.042(9)	0.041(9)	0.021(12)	0.070(11)	0.084(16)	0.096(13)	0.005(10)	0.022(11)	0.029(10)	0.060(12)	0.075(12)	0.103(14)	0.155(15)
23	0.026(10)	0.014(14)	0.049(11)	0.059(18)	0.057(16)	0.089(19)	0.113(19)	0.011(10)	0.016(11)	0.041(10)	0.077(10)	0.087(11)	0.103(10)	0.113(11)
24	0.006(5)	0.016(7)	0.031(6)	0.024(8)	0.015(19)	0.050(14)	0.046(14)	0.001(13)	0.014(11)	0.029(10)	0.054(13)	0.058(19)	0.090(15)	0.101(18)
28	0.016(9)	0.017(10)	0.021(11)	0.032(13)	0.045(13)	0.043(17)	0.044(21)							

[a] The uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$, e.g. 0.015(10) means 0.015 ± 0.010. The uncertainties were calculated based on estimates of spectral noise, with the estimation approach provided in the REDOR NMR spectroscopy section of Detailed Descriptions of Materials and Methods.

Table S3. Experimental REDOR $\Delta S/S_0$ values for replicate samples with WT Fp's [a]

τ_k	u = 13		u = 16		u = 17		u = 20		
(ms)	Fp	Fp-dimer	Fp	Fp-dimer	Fp	Fp-dimer	Fp	Fp	Fp-dimer
2.2	0.010(5)	0.012(5)	0.012(7)	0.016(8)	0.004(9)	0.005(6)	0.022(12)	0.018(10)	0.024(8)
8.2	0.034(8)	0.039(7)	0.044(9)	0.049(7)	0.058(10)	0.047(7)	0.017(12)	0.033(10)	0.040(8)
16.2	0.067(12)	0.086(8)	0.090(10)	0.085(8)	0.099(7)	0.097(8)	0.068(9)	0.066(10)	0.066(11)
24.2	0.102(15)	0.125(8)	0.128(8)	0.138(8)	0.155(13)	0.149(11)	0.116(15)	0.108(10)	0.113(18)
32.2	0.172(14)	0.170(9)	0.179(11)	0.178(14)	0.192(11)	0.211(18)	0.161(12)	0.162(11)	0.151(13)
40.2	0.218(16)	0.212(15)	0.238(15)	0.232(13)	0.247(16)	0.216(13)	0.177(24)	0.170(12)	0.157(15)
48.2	0.256(25)	0.236(15)	0.253(15)	0.277(18)	0.275(21)	0.253(13)	0.175(15)	0.206(11)	0.201(20)

[a] For each u value, the left-most column is the $\Delta S/S_0$ data listed in **Table S2**. The right column (or for u=20 center and right columns) is $\Delta S/S_0$ data of replicate samples in which the Fp was separately -synthesized and -purified but had the same ^{13}CO-labeled residue and the same ^{15}N-labeled residue, see **Fig**. **3** in the main text. The experimental uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$ value, e.g. an entry of 0.058(10) means 0.058 ± 0.010. The uncertainties were calculated based on estimates of spectral noise, with the estimation approach provided in the REDOR NMR spectroscopy section of Detailed Descriptions of Materials and Methods. The Fp-dimer was synthesized by cross-linking in air AVGIGALFLGFLGAAGSTMGARSWKKKKKCA, with underlining for the 23 N-terminal residues of HIV gp41. The synthesis and purification of Fp-dimer followed published procedures (3). For the present study, Fp-dimer is only significant as a sample replicate. The original motivation of the Fp-dimer experiments was that close proximity of the cross-linked cysteines is similar to the proximity C-terminal of Fp's for Gp41 trimer-of-hairpins, see **Fig**. **6**. We thought that the cysteine proximity might change the registry distribution and therefore the REDOR $\Delta S/S_0$. Instead, Fp-dimer has $\Delta S/S_0$ values that within error are typically the same as non-cross-linked Fp, the main subject of the present study. The very similar $\Delta S/S_0$ for the replicate samples supports a thermodynamic equilibrium distribution of antiparallel registries in the Fp β sheet. This distribution is also consistent with previously-reported NMR data for Fp+hairpin that are described in the Discussion. There is variability in the numerical difference between: (1) the magnitude of the difference in $\Delta S/S_0$ between sample replicates; and (2) the uncertainty of an individual sample. The magnitude difference is sometimes smaller than the uncertainty which supports the reproducibility of sample preparation. However, sometimes the magnitude difference is larger than the uncertainty, but the magnitude difference is typically smaller than the magnitude difference uncertainty. For example, the u=16 $\Delta S/S_0$ at τ_k = 24.2 ms are 0.128(8) and 0.138(8) and the magnitude difference and associated uncertainty is 0.10(11).

Table S4. SIMPSON-calculated values of $\gamma^{lb,lb}(\tau_k) = [S_1/S_0]^{lb,lb}(\tau_k)$ [a]

	Constrained Fitting									
	t = u	**t = u ± 1**	**t = u ± 2**							
	Unconstrained Fitting									
	t_1=u **t_2=u**	t_1 = u **± 1** t_2 = u **± 1**		**t_1=u** **t_2= u ± 1**	**t_1=u** **t_2=X**	**t_1= u ± 1** **t_2=u**	**t_1=X** **t_2=u**	t_1 = u **± 1** t_2 = u **-+1**	t_1 = u **± 1** t_2 = X	t_1 = X t_2 = u **± 1**
	Typical ^{13}CO-^{15}N Distances (Å)									
	4.0 5.5	5.7 5.9	7.8 8.9	4.0 5.9	4.0	5.7 5.5	5.5	5.7 5.9	5.7	5.9
τ_k (ms)										
2.2	0.9917	0.9984	0.9998	0.9921	0.9928	0.9980	0.9989	0.9984	0.9991	0.9992
8.2	0.8938	0.9785	0.9971	0.8974	0.9064	0.9737	0.9850	0.9785	0.9885	0.9898
16.2	0.6453	0.9186	0.9890	0.6476	0.6710	0.9012	0.9427	0.9191	0.9560	0.9608
24.2	0.3786	0.8263	0.9755	0.3565	0.3778	0.7914	0.8754	0.8293	0.9039	0.9143
32.2	0.1964	0.7103	0.9569	0.1288	0.1236	0.6571	0.7870	0.7200	0.8348	0.8521
40.2	0.1186	0.5814	0.9334	0.0156	-0.0230	0.5132	0.6824	0.6037	0.7520	0.7769
48.2	0.0939	0.4501	0.9052	0.0011	-0.0470	0.3739	0.5675	0.4926	0.6592	0.6917

[a] This table provides the $\gamma=S_1/S_0$ intensity ratios calculated using the SIMPSON program at different dephasing times (τ_k). The SI section on f_t fitting describes the SIMPSON calculations. The first row in this Table is the t values for constrained fitting, **Eq**. **2** in the main text, and has three columns of values. The second row is the t_1 and t_2 values for unconstrained fitting, **Eq**. **3** in the main text, and shares the first two columns with constrained fitting, doesn't include the third column, and then has seven additional columns that are only for unconstrained fitting. The third row lists typical ^{13}CO-^{15}N inter-spin distances. The top line in this row is the central molecule ^{13}CO-molecule 1 ^{15}N distance and the bottom line is the central molecule ^{13}CO-molecule 2 ^{15}N distance. For constrained fitting, all neighboring strands within a single sheet have the same registry and value of t, **Fig**. **2a**. The t parameter describes the registry alignment of two neighboring strands in the antiparallel β sheet and is the total number of residues in one strand that could be hydrogen-bonded to residues in the neighboring strand. The u parameter is a labeling- and therefore sample- dependent parameter, and is the value of t that aligns the ^{13}CO-labeled residue on one strand and the ^{15}N-labeled residue on the neighboring strand. For unconstrained fitting, there can be multiple registries within a single sheet, **Fig**. **2b**. Each "central" Fp molecule in an unconstrained sheet has neighboring Fp molecules denoted 1 and 2, with assignment of molecule 1 vs. 2 based on having vs. not having a NH hydrogen-bonded to the labeled ^{13}CO of the central molecule. Registries are indexed by t_1 and t_2, the total number of residues of molecule 1 or 2, respectively, that could be hydrogen-bonded to the central molecule. For constrained sheets, t=X means $t\neq u,u\pm1,u\pm2$. For unconstrained sheets, t_1=X means $t_1\neq u,u\pm1$ and t_2=X means $t_2\neq u,u\pm1$. When t=X or $t_1=t_2$=X, the ^{13}CO-^{15}N distances are approximated to be large enough so that $\gamma^{lb,lb}(\tau_k)=1$. The other $\gamma^{lb,lb}(\tau_k)$ values are calculated by the SIMPSON simulation program and are typically for one ^{13}CO and two ^{15}N spins. For unconstrained registries with either t_1=X or t_2=X, the simulation system is one ^{13}CO and one ^{15}N spin. Each $\gamma^{lb,lb}(\tau_k)$ is an average from ~10 SIMPSON simulations that are each based on coordinates of different atoms in OMPG β barrel outer membrane protein (OMPG, PDB file 2IWW).

Table S5. f(t) registry populations for fittings with b = 0.98 and 2.2-40.2 ms data [a]

	WT; Unconstrained	WT; Constrained	V2E; Unconstrained	V2E; Constrained
t	χ^2=107; ⟨t⟩=16.1843	χ^2=131; ⟨t⟩=16.1734	χ^2=145; ⟨t⟩=18.4585	χ^2=231; ⟨t⟩=18.4957
8	0.0015	0	0	0
9	0.0090	0.0027	0	0
10	0.0032	0	0	0
11	0.0355	0.0247	0	0
12	0.0579	0.0672	0.0092	0
13	0.1306	0.1384	0	0
14	0.0524	0.0545	0.0065	0
15	0.1297	0.1266	0.0769	0.0806
16	0.1035	0.1069	0.1113	0.1114
17	0.1514	0.1678	0.1106	0.1254
18	0.1159	0.1206	0.2054	0.1993
19	0.0290	0.0138	0.0351	0.0058
20	0.1325	0.1490	0.3564	0.4275
21	0	0	0.0425	0.0137
22	0.0193	0.0033	0	0
23	0.0285	0.0244	0.0460	0.0364
24	0	0	0	0

[a] Unconstrained and constrained fittings were done using u = 8-24, k = 1-6, τ_k = 2.2-40.2 ms data, and b=0.98 scaling factor for $\gamma^{lb,lb}(\tau_k)$ in **Eq**. **2** or **3**. For unconstrained fittings, "b=0.98" means $b_{t1=u,u\pm1,t2=u,u\pm1}$ = 0.98, $b_{t1=u,u\pm1,t2=X}$ = $b_{t1=X,t2=u,u\pm1}$ = 0.99, and $b_{t1=X,t2=X}$ = 1.

Table S6. WT $\Delta S/S_0$ values from experiment and from unconstrained and constrained fittings with b = 0.98 [a]

τ_k	u = 8			u = 9			u = 10			u = 11			u = 12		
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98
2.2	0.006(7)	0.0127	0.0126	0.012(6)	0.0128	0.013	0.015(10)	0.0134	0.014	0.014(5)	0.0143	0.0163	0.011(9)	0.0164	0.0174
8.2	0.017(6)	0.0247	0.0244	0.009(6)	0.0253	0.025	0.022(7)	0.026	0.0263	0.026(6)	0.0295	0.031	0.016(9)	0.0349	0.0364
16.2	0.030(6)	0.0309	0.0299	0.032(9)	0.0328	0.031	0.033(12)	0.0341	0.0333	0.046(9)	0.0449	0.0447	0.060(9)	0.0593	0.0613
24.2	0.038(9)	0.0379	0.0361	0.033(9)	0.0414	0.0378	0.046(16)	0.0441	0.0418	0.066(12)	0.0645	0.0613	0.095(12)	0.0917	0.0918
32.2	0.037(11)	0.0441	0.0413	0.047(11)	0.0491	0.0435	0.039(19)	0.0541	0.05	0.081(11)	0.0839	0.0771	0.113(11)	0.1257	0.1202
40.2	0.052(12)	0.0482	0.0443	0.068(12)	0.0541	0.0468	0.062(17)	0.0626	0.0565	0.097(17)	0.0998	0.09	0.17(16)	0.1562	0.1429

τ_k	u = 13			u = 14			u = 15			u = 16			u = 17		
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98
2.2	0.010(5)	0.0171	0.0197	0.003(6)	0.0178	0.0205	0.008(8)	0.0178	0.0224	0.012(7)	0.0192	0.0221	0.004(9)	0.0193	0.0218
8.2	0.034(8)	0.04	0.0433	0.033(9)	0.0372	0.0403	0.043(8)	0.0413	0.046	0.044(9)	0.0426	0.0461	0.058(10)	0.0452	0.049
16.2	0.067(12)	0.0763	0.0799	0.088(12)	0.0646	0.0681	0.093(12)	0.0794	0.0831	0.090(10)	0.081	0.0846	0.099(7)	0.0903	0.0955
24.2	0.102(15)	0.1228	0.1223	0.109(13)	0.102	0.1041	0.123(11)	0.1287	0.1274	0.128(8)	0.1321	0.1327	0.155(13)	0.1487	0.1505
32.2	0.172(14)	0.1678	0.1579	0.138(14)	0.1428	0.141	0.173(14)	0.1772	0.1669	0.179(11)	0.1853	0.179	0.192(11)	0.2066	0.1988
40.2	0.218(16)	0.2033	0.1824	0.171(11)	0.1817	0.1746	0.215(19)	0.2169	0.1971	0.238(15)	0.2329	0.2184	0.247(16)	0.2547	0.2348

τ_k	u = 18			u = 19			u = 20			u = 21			u = 22		
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98
2.2	0.011(7)	0.0179	0.0218	0.010(7)	0.0172	0.0197	0.022(12)	0.0158	0.0178	0.010(9)	0.015	0.0156	0.011(9)	0.0134	0.0153
8.2	0.055(11)	0.0407	0.0453	0.022(5)	0.0347	0.0366	0.017(12)	0.0377	0.0405	0.005(10)	0.029	0.0296	0.042(9)	0.027	0.0279
16.2	0.085(11)	0.0772	0.0819	0.064(6)	0.057	0.0573	0.068(9)	0.0709	0.0742	0.028(11)	0.0412	0.0419	0.041(9)	0.0376	0.0358
24.2	0.126(10)	0.1248	0.1259	0.082(8)	0.0878	0.0856	0.116(15)	0.1126	0.1111	0.052(13)	0.058	0.0584	0.021(12)	0.0507	0.0455
32.2	0.174(12)	0.1724	0.1656	0.131(11)	0.1224	0.1174	0.161(12)	0.1513	0.1387	0.074(13)	0.0774	0.0768	0.070(11)	0.0632	0.055
40.2	0.188(20)	0.2126	0.1965	0.145(13)	0.1567	0.1491	0.177(24)	0.1791	0.1533	0.072(16)	0.0972	0.0944	0.084(16)	0.0727	0.0628

τ_k	u = 23			u = 24			u = 28	
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Eqs. S4,S5
2.2	0.026(10)	0.0135	0.0131	0.006(5)	0.013	0.013	0.016(9)	0.01257
8.2	0.014(14)	0.0276	0.0267	0.016(7)	0.0252	0.0251	0.017(10)	0.02433
16.2	0.049(11)	0.0397	0.0366	0.031(6)	0.0319	0.0316	0.021(11)	0.02974
24.2	0.059(18)	0.0545	0.0475	0.024(8)	0.0399	0.0393	0.032(13)	0.03571
32.2	0.057(16)	0.0684	0.056	0.015(19)	0.0477	0.0464	0.045(13)	0.04071
40.2	0.089(19)	0.0784	0.0605	0.050(14)	0.0537	0.0515	0.043(17)	0.04349

[a] For u=28, the values are from experiment and from calculation with **Eqs**. **S4**, **S5**. The experimental uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$ value, e.g. 0.015(10) means 0.015 ± 0.010.

Table S7. V2E $\Delta S/S_0$ values from experiment and from unconstrained and constrained fittings with b = 0.98 [a]

τ_k	u = 8			u = 9			u = 10			u = 11			u = 12		
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98
2.2	0.009(10)	0.0126	0.0126	0.014(10)	0.0126	0.0126	0.024(10)	0.0126	0.0126	-0.01(1)	0.0127	0.0126	0.001(10)	0.0128	0.0126
8.2	0.026(10)	0.0243	0.0243	0.005(10)	0.0243	0.0243	0.018(10)	0.0243	0.0243	0.006(11)	0.0246	0.0243	0.012(10)	0.0252	0.0243
16.2	0.028(10)	0.0297	0.0297	0.027(10)	0.0297	0.0297	0.024(10)	0.0297	0.0297	0.031(10)	0.0304	0.0297	0.022(10)	0.0325	0.0297
24.2	0.018(10)	0.0357	0.0357	0.032(10)	0.0357	0.0357	0.041(12)	0.0357	0.0357	0.040(12)	0.0371	0.0357	0.030(14)	0.0409	0.0357
32.2	0.053(14)	0.0407	0.0407	0.047(10)	0.0407	0.0407	0.047(12)	0.0407	0.0407	0.050(11)	0.043	0.0407	0.059(10)	0.0482	0.0407
40.2	0.052(10)	0.0435	0.0435	0.058(12)	0.0435	0.0435	0.050(13)	0.0435	0.0435	0.053(17)	0.0468	0.0435	0.086(10)	0.0527	0.0435

τ_k	u = 13			u = 14			u = 15			u = 16			u = 17		
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98
2.2	0.002(13)	0.0128	0.0138	0.007(17)	0.0139	0.0155	0.004(13)	0.0161	0.0179	0.019(10)	0.0179	0.0212	0.019(12)	0.02	0.0214
8.2	0.044(17)	0.0248	0.0257	0.043(13)	0.0273	0.0287	0.039(10)	0.0352	0.0372	0.034(12)	0.0405	0.0442	0.035(20)	0.0444	0.0468
16.2	0.038(10)	0.0309	0.0316	0.034(10)	0.0375	0.0382	0.046(16)	0.0613	0.0628	0.073(10)	0.0764	0.0798	0.078(20)	0.0856	0.0888
24.2	0.047(16)	0.038	0.0383	0.061(10)	0.0507	0.0507	0.095(10)	0.0952	0.0931	0.129(12)	0.1233	0.1231	0.175(19)	0.1406	0.1407
32.2	0.063(20)	0.0445	0.0444	0.073(12)	0.0646	0.064	0.143(10)	0.1292	0.1198	0.197(15)	0.1706	0.1629	0.238(16)	0.1981	0.1896
40.2	0.078(25)	0.0491	0.0486	0.114(14)	0.0773	0.0762	0.175(11)	0.1577	0.1397	0.245(14)	0.2107	0.1946	0.271(14)	0.2497	0.23

τ_k	u = 18			u = 19			u = 20			u = 21			u = 22		
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98
2.2	0.025(17)	0.0192	0.0269	0.013(11)	0.0223	0.0248	0.009(10)	0.0213	0.0248	0.007(10)	0.0192	0.0203	0.005(10)	0.014	0.0198
8.2	0.057(13)	0.048	0.0557	0.046(10)	0.0448	0.0463	0.056(10)	0.06	0.0676	0.043(10)	0.0392	0.0393	0.022(11)	0.027	0.0331
16.2	0.113(12)	0.1004	0.106	0.069(14)	0.0818	0.0808	0.146(10)	0.1381	0.1521	0.075(10)	0.0689	0.066	0.029(10)	0.0364	0.0432
24.2	0.194(12)	0.1667	0.1646	0.144(10)	0.1347	0.1313	0.262(10)	0.2337	0.2441	0.130(10)	0.1104	0.1037	0.060(12)	0.0488	0.0567
32.2	0.254(19)	0.2295	0.2143	0.213(10)	0.1963	0.1917	0.33(10)	0.3181	0.31	0.179(10)	0.1574	0.1471	0.075(12)	0.0622	0.0717
40.2	0.302(18)	0.2777	0.2498	0.280(12)	0.2599	0.2558	0.379(14)	0.376	0.3422	0.198(15)	0.2045	0.1914	0.103(14)	0.075	0.0864

τ_k	u = 23			u = 24		
(ms)	Expt.	Uncons. b=0.98	Cons. b=0.98	Expt.	Uncons. b=0.98	Cons. b=0.98
2.2	0.011(10)	0.0135	0.0135	0.001(13)	0.0133	0.0132
8.2	0.016(11)	0.0287	0.0279	0.014(11)	0.0257	0.0254
16.2	0.041(10)	0.0434	0.04	0.029(10)	0.0332	0.0324
24.2	0.077(10)	0.0614	0.0531	0.054(13)	0.0425	0.0408
32.2	0.087(11)	0.0777	0.0631	0.058(19)	0.0519	0.0489
40.2	0.103(10)	0.0887	0.0682	0.090(15)	0.06	0.0551

[a] The experimental uncertainties are in parentheses using the convention that the uncertainty corresponds to the right-most digits in the $\Delta S/S_0$ value, e.g. 0.015(10) means 0.015 ± 0.010.

Table S8. f(t), χ^2, and ⟨t⟩ values for all fittings

	WT	WT	WT	WT	WT	WT	WT	V2E	V2E	V2E	V2E	V2E	V2E	V2E	V2E	V2E
	Uncons	Uncons	Uncons	Uncons	Cons.	Cons	Cons	Uncons	Uncons	Uncons	Uncons	Cons.	Cons.	Cons.	Cons.	Cons.
	b=0.98	b=0.98	b=1	b=1	b=0.98	b=1	b=1	b=0.98	b=0.98	b=1	b=1	b=0.98	b=1	b=1	b=0.9641	b=0.9554
	k=1-6	k=1-7	k=1-6	k=1-7	k=1-6	k=1-6	k=1-7	k=1-6	k=1-7	k=1-6	k=1-7	k=1-6	k=1-6	k=1-7	k=1-6	k=1-7
	χ^2=107	χ^2=130	χ^2=97	χ^2=123	χ^2=131	χ^2=117	χ^2=163	χ^2=145	χ^2=221	χ^2=168	χ^2=250	χ^2=231	χ^2=277	χ^2=422	χ^2=220	χ^2=333
	⟨t⟩=16.18	⟨t⟩=16.08	⟨t⟩=16.19	⟨t⟩=16.09	⟨t⟩=16.17	⟨t⟩=16.11	⟨t⟩=16.09	⟨t⟩=18.46	⟨t⟩=18.42	⟨t⟩=18.49	⟨t⟩=18.45	⟨t⟩=18.50	⟨t⟩=18.49	⟨t⟩=18.49	⟨t⟩=18.50	⟨t⟩=18.49
t																
8	0.0015	0.0029	0	0.0009	0	0	0	0	0	0	0	0	0	0	0	0
9	0.0090	0.0066	0.0090	0.0073	0.0027	0	0	0	0	0	0	0	0	0	0	0
10	0.0032	0.0130	0	0.0100	0	0	0	0	0	0	0	0	0	0	0	0
11	0.0355	0.0379	0.0346	0.0375	0.0247	0.0230	0.0293	0	0	0	0	0	0	0	0	0
12	0.0579	0.0608	0.0598	0.0626	0.0672	0.0681	0.0688	0.0092	0.0095	0.0067	0.0073	0	0	0	0	0
13	0.1306	0.1335	0.1294	0.1321	0.1384	0.1369	0.1400	0	0	0	0	0	0	0	0	0
14	0.0524	0.0543	0.0555	0.0575	0.0545	0.0577	0.0586	0.0065	0.0242	0.0038	0.0223	0	0	0	0	0.0017
15	0.1297	0.1285	0.1285	0.1268	0.1266	0.1301	0.1282	0.0769	0.0743	0.0776	0.0756	0.0806	0.0762	0.0950	0.0843	0.1013
16	0.1035	0.0990	0.1073	0.1025	0.1069	0.1113	0.1095	0.1113	0.1098	0.1118	0.1096	0.1114	0.1199	0.1140	0.1047	0.0976
17	0.1514	0.1509	0.1536	0.1524	0.1678	0.1720	0.1712	0.1106	0.1061	0.1104	0.1063	0.1254	0.1083	0.0972	0.1389	0.1322
18	0.1159	0.1071	0.1152	0.1060	0.1206	0.1182	0.1064	0.2054	0.1903	0.2014	0.1861	0.1993	0.2108	0.2020	0.1895	0.1781
19	0.0290	0.0260	0.0342	0.0302	0.0138	0.0280	0.0255	0.0351	0.0444	0.0434	0.0523	0.0058	0.0104	0.0123	0.0020	0.0034
20	0.1325	0.1240	0.1302	0.1218	0.1490	0.1384	0.1341	0.3564	0.3437	0.3549	0.3409	0.4275	0.4263	0.4225	0.4291	0.4286
21	0	0	0	0	0	0	0	0.0425	0.0413	0.0463	0.0465	0.0137	0.0183	0.0203	0.0091	0.00076
22	0.0193	0.0245	0.0184	0.0242	0.0033	0.0070	0.0152	0	0.0107	0	0.0081	0	0	0	0	0
23	0.0285	0.0309	0.0244	0.0282	0.0244	0.0094	0.0131	0.0460	0.0455	0.0436	0.0450	0.0364	0.0298	0.0367	0.0424	0.0495
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[a] The k=1-6 corresponds to τ_k=2.2-40.2 ms and k=1-7 corresponds to τ_k=2.2-48.2 ms. For all WT fittings, the average value of ⟨t⟩ and RMSD is $\langle \langle t \rangle^{WT} \rangle$ = 16.132 ± 0.048. For all V2E fittings, $\langle \langle t \rangle^{V2E} \rangle$ = 18.475 ± 0.028.

Table S9. Values of $f(t)^{WT}$ and fitted f(t) and contributions to G(t) [a]

	$f(t)^{WT}$	Fitted f(t)	$(t-11) \times G_{\beta}^{WT}$	$L(t) \times G_{Leu}^{WT}$	$g^{WT} \times G(t)_{sc}^{WT}$
t				kcal/mole	
11	0.0355	0.0221	0	0	-0.387
12	0.0579	0.0429	-0.113	0	-0.671
13	0.1306	0.1371	-0.226	-0.350	-0.905
14	0.0524	0.0638	-0.339	0	-0.683
15	0.1297	0.1380	-0.452	-0.350	-0.683
16	0.1035	0.0930	-0.565	0	-0.683
17	0.1514	0.1390	-0.679	-0.350	-0.461
18	0.1159	0.1132	-0.792	-0.350	-0.225
19	0.0290	0.0520	-0.905	0	0.005
20	0.1325	0.1325	-1.018	-0.350	-0.093

[a] The $f(t)^{WT}$ are from the unconstrained model with b=0.98 and the k=1-6, τ_k=2.2-40.2 ms data, see **Table S5**. Fitting was done using **Eq**. **4** in the main text and the best-fit G_{β}^{WT} = -0.113 kcal/mole, G_{Leu}^{WT} = -0.350 kcal/mole, and g^{WT} = 0.129. Each $G(t)_{sc}^{WT}$ is the sum of free energies of membrane insertion of sidechains for residues between V2 and t-1 with sidechain energy relative to Ala (24). The f(t) values are determined by relative rather than absolute values of G(t), i.e. adding or subtracting a constant energy value to all G(t) doesn't change the f(t). The energy offset of $G(t)_{\beta}^{WT} = (t-11) \times G_{\beta}^{WT}$ is chosen so that $G(11)_{\beta}^{WT} = 0$. The $f(t)^{WT}$ and the free energy contributions are displayed as a bar plot in **Fig**. **5a** in the main text.

Table S10. Values of $f(t)^{V2E}$ and fitted f(t) and contributions to G(t) [a]

	$f(t)^{WT}$	Fitted f(t)	$(t-15) \times G_{\beta}^{V2E}$	$L(t) \times G_{Leu}^{V2E}$
t			kcal/mole	
15	0.0769	0.0703	0	-1.404
17	0.1106	0.1348	-0.391	-1.404
18	0.2054	0.1868	-0.586	-1.404
19	0.0351	0.0249	-0.782	0
20	0.3564	0.3583	-0.977	-1.404
21	0.0425	0.0478	-1.173	0

[a] The $f(t)^{V2E}$ are from the unconstrained model with b=0.98 and the k=1-6, τ_k=2.2-40.2 ms data, see **Table S5**. Fitting was done using **Eq**. **5** in the main text and the best-fit G_{β}^{V2E} = -0.195 kcal/mole and G_{Leu}^{V2E} = -1.40 kcal/mole. The f(t) values are determined by relative rather than absolute values of G(t), i.e. adding or subtracting a constant energy value to all G(t) doesn't change the f(t). The energy offset of $G(t)_{\beta}^{V2E} = (t-15) \times G_{\beta}^{V2E}$ is chosen so that $G(15)_{\beta}^{V2E} = 0$. The $f(t)^{V2E}$ and the free energy contributions are displayed as a bar plot in **Fig**. **5b** in the main text.

References

1. R. Yang, M. Prorok, F. J. Castellino, D. P. Weliky, A trimeric HIV-1 fusion peptide construct which does not self-associate in aqueous solution and which has 15-fold higher membrane fusion rate. *J. Am. Chem. Soc.* **126**, 14722-14723 (2004).
2. J. Yang, M. Prorok, F. J. Castellino, D. P. Weliky, Oligomeric β-structure of the membrane-bound HIV-1 fusion peptide formed from soluble monomers. *Biophys. J.* **87**, 1951-1963 (2004).
3. W. Qiang, D. P. Weliky, HIV fusion peptide and its cross-linked oligomers: efficient syntheses, significance of the trimer in fusion activity, correlation of β strand conformation with membrane cholesterol, and proximity to lipid headgroups. *Biochemistry* **48**, 289-301 (2009).
4. L. Lapatsanis, G. Milias, K. Froussios, M. Kolovos, Synthesis of N-2,2,2-(trichloroethoxycarbonyl)-L-amino acids and N-(9-fluorenylmethoxycarbonyl)-L-amino acids involving succinimidoxy anion as a leaving group in amino-acid protection. *Synthesis* **8**, 671-673 (1983).
5. L. H. Jia *et al.*, REDOR solid-state NMR as a probe of the membrane locations of membrane-associated peptides and proteins. *J. Magn. Reson.* **253**, 154-165 (2015).
6. W. Qiang, M. L. Bodner, D. P. Weliky, Solid-state NMR spectroscopy of human immunodeficiency virus fusion peptides associated with host-cell-like membranes: 2D correlation spectra and distance measurements support a fully extended conformation and models for specific antiparallel strand registries. *J. Am. Chem. Soc.* **130**, 5459-5471 (2008).
7. W. Qiang, Y. Sun, D. P. Weliky, A strong correlation between fusogenicity and membrane insertion depth of the HIV fusion peptide. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15314-15319 (2009).
8. L. Xie, L. H. Jia, S. Liang, D. P. Weliky, Multiple locations of peptides in the hydrocarbon core of gel-phase membranes revealed by peptide ^{13}C to lipid ^{2}H rotational-echo double-resonance solid-state nuclear magnetic resonance. *Biochemistry* **54**, 677-684 (2015).
9. P. U. Ratnayake, K. Sackett, M. J. Nethercott, D. P. Weliky, pH-dependent vesicle fusion induced by the ectodomain of the human immunodeficiency virus membrane fusion protein gp41: Two kinetically distinct processes and fully-membrane-associated gp41 with predominant beta sheet fusion peptide conformation. *Biochim. Biophys. Acta* **1848**, 289-298 (2015).
10. K. Sackett, M. J. Nethercott, Z. X. Zheng, D. P. Weliky, Solid-state NMR spectroscopy of the HIV gp41 membrane fusion protein supports intermolecular antiparallel beta sheet fusion peptide structure in the final six-helix bundle state. *J. Mol. Biol.* **426**, 1077-1094 (2014).
11. J. Yang, D. P. Weliky, Solid state nuclear magnetic resonance evidence for parallel and antiparallel strand arrangements in the membrane-associated HIV-1 fusion peptide. *Biochemistry* **42**, 11879-11890 (2003).
12. M. L. Bodner *et al.*, Temperature dependence and resonance assignment of ^{13}C NMR spectra of selectively and uniformly labeled fusion peptides associated with membranes. *Magn. Reson. Chem.* **42**, 187-194 (2004).
13. M. Kashefi, N. Malik, J. O. Struppe, L. K. Thompson, Carbon-nitrogen REDOR to identify ms-timescale mobility in proteins. *J. Magn. Reson.* **305**, 5-15 (2019).
14. C. R. Morcombe, K. W. Zilm, Chemical shift referencing in MAS solid state NMR. *J. Magn. Reson.* **162**, 479-486 (2003).
15. T. Gullion, Introduction to rotational-echo, double-resonance NMR. *Concepts Magn. Reson.* **10**, 277-289 (1998).
16. Z. Zheng, R. Yang, M. L. Bodner, D. P. Weliky, Conformational flexibility and strand arrangements of the membrane-associated HIV fusion peptide trimer probed by solid-state NMR spectroscopy. *Biochemistry* **45**, 12960-12975 (2006).
17. T. Gullion, D. B. Baker, M. S. Conradi, New, compensated Carr-Purcell sequences. *J. Magn. Reson.* **89**, 479-484 (1990).
18. A. E. Bennett, C. M. Rienstra, M. Auger, K. V. Lakshmi, R. G. Griffin, Heteronuclear decoupling in rotating solids. *J. Chem. Phys.* **103**, 6951-6958 (1995).
19. C. M. Gabrys *et al.*, Solid-state nuclear magnetic resonance measurements of HIV fusion peptide ^{13}CO to lipid ^{31}P proximities support similar partially inserted membrane locations of the α Helical and β sheet peptide structures. *J. Phys. Chem. A* **117**, 9848-9859 (2013).
20. K. T. Mueller, Analytical solutions for the time evolution of dipolar-dephasing NMR signals. *J. Magn. Reson. Ser. A* **113**, 81-93 (1995).

21. M. Bak, J. T. Rasmussen, N. C. Nielsen, SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.* **147**, 296-330 (2000).
22. M. Bak, R. Schultz, T. Vosegaard, N. C. Nielsen, Specification and visualization of anisotropic interaction tensors in polypeptides and numerical simulations in biological solid-state NMR. *J. Magn. Reson.* **154**, 28-45 (2002).
23. S. D. Schmick, D. P. Weliky, Major antiparallel and minor parallel beta sheet populations detected in the membrane-associated Human Immunodeficiency Virus fusion peptide. *Biochemistry* **49**, 10623-10635 (2010).
24. C. P. Moon, K. G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10174-10177 (2011).
25. J. Yang, C. M. Gabrys, D. P. Weliky, Solid-state nuclear magnetic resonance evidence for an extended beta strand conformation of the membrane-bound HIV-1 fusion peptide. *Biochemistry* **40**, 8126-8137 (2001).

Software S1: Example Python code for data fitting

S1. Experimental REDOR and sigma data of WT-HFP

```
exp8 = 0.006, 0.017, 0.030, 0.038, 0.037, 0.052, 0.057
exp9 = 0.012, 0.009, 0.032, 0.033, 0.047, 0.068, 0.063
exp10 = 0.015, 0.022, 0.033, 0.046, 0.039, 0.062, 0.111
exp11 = 0.014, 0.026, 0.046, 0.066, 0.081, 0.097, 0.151
exp12 = 0.011, 0.016, 0.060, 0.095, 0.113, 0.170, 0.215
exp13 = 0.010, 0.034, 0.067, 0.102, 0.172, 0.218, 0.256
exp14 = 0.003, 0.033, 0.088, 0.109, 0.138, 0.171, 0.235
exp15 = 0.008, 0.043, 0.093, 0.123, 0.173, 0.215, 0.244
exp16 = 0.012, 0.044, 0.090, 0.128, 0.179, 0.238, 0.253
exp17 = 0.004, 0.058, 0.099, 0.155, 0.192, 0.247, 0.275
exp18 = 0.011, 0.055, 0.085, 0.126, 0.174, 0.188, 0.201
exp19 = 0.010, 0.022, 0.064, 0.082, 0.131, 0.145, 0.157
exp20 = 0.022, 0.017, 0.068, 0.116, 0.161, 0.177, 0.175
exp21 = 0.010, 0.005, 0.028, 0.052, 0.074, 0.072, 0.112
exp22 = 0.011, 0.042, 0.041, 0.021, 0.070, 0.084, 0.096
exp23 = 0.026, 0.014, 0.049, 0.059, 0.057, 0.089, 0.113
exp24 = 0.006, 0.016, 0.031, 0.024, 0.015, 0.050, 0.046

sigma8 = 0.007, 0.006, 0.006, 0.009, 0.011, 0.012, 0.016
sigma9 = 0.006, 0.006, 0.009, 0.009, 0.011, 0.012, 0.018
sigma10 = 0.010, 0.007, 0.012, 0.016, 0.019, 0.017, 0.021
sigma11 = 0.005, 0.006, 0.009, 0.012, 0.011, 0.017, 0.022
sigma12 = 0.009, 0.009, 0.009, 0.012, 0.011, 0.016, 0.023
sigma13 = 0.005, 0.008, 0.012, 0.015, 0.014, 0.016, 0.025
sigma14 = 0.006, 0.009, 0.012, 0.013, 0.014, 0.011, 0.021
sigma15 = 0.008, 0.008, 0.012, 0.011, 0.014, 0.019, 0.019
sigma16 = 0.007, 0.009, 0.010, 0.008, 0.011, 0.015, 0.015
sigma17 = 0.009, 0.010, 0.007, 0.013, 0.011, 0.016, 0.021
sigma18 = 0.007, 0.011, 0.011, 0.010, 0.012, 0.020, 0.021
sigma19 = 0.007, 0.005, 0.006, 0.008, 0.011, 0.013, 0.013
sigma20 = 0.012, 0.012, 0.009, 0.015, 0.012, 0.024, 0.015
sigma21 = 0.009, 0.010, 0.011, 0.013, 0.013, 0.016, 0.016
sigma22 = 0.009, 0.009, 0.009, 0.012, 0.011, 0.016, 0.013
sigma23 = 0.010, 0.014, 0.011, 0.018, 0.016, 0.019, 0.019
sigma24 = 0.005, 0.007, 0.006, 0.008, 0.019, 0.014, 0.014
```

1. WT_DualAnnealing_5registries_6data_B0.98

```
import numpy as np
from scipy.optimize import dual_annealing

# Input data from files
fname_exp = '/Users/yijinzhang/Desktop/REDOR/python_WT/exp.txt'
fh_exp = open(fname_exp)

fname_sig = '/Users/yijinzhang/Desktop/REDOR/python_WT/sigma.txt'
fh_sig = open(fname_sig)

lst_exp = []
for line in fh_exp:
    line = line.rstrip()
```

```
    lst_exp.append(line)

exp_0 = {}
for i in range(0,len(lst_exp)):
    line = lst_exp[i].split('=')[1]
    exp0 = line.split(',')
    new_lst=[]
    for j in range(0,len(exp0)-1):
        k= float(exp0[j])
        new_lst.append(k)
    exp_0['exp'+str(i+8)] = np.array(new_lst)

dephasing = []
for i in range (8,25):

    item = sum([exp_0['exp'+str(i)][j] for j in range(6)])
    dephasing.append(item)

def dephasing_time(time):

    if time == 2.2:
        x = 0
    if time == 8.2:
        x = 1
    if time == 16.2:
        x = 2
    if time == 24.2:
        x = 3
    if time == 32.2:
        x = 4
    if time == 40.2:
        x = 5

    return [exp_0['exp'+str(i+8)][x] for i in range(17)]

lst_sig = []
for line in fh_sig:
    line = line.rstrip()
    lst_sig.append(line)

sig = {}
for i in range(0,len(lst_sig)):
    line = lst_sig[i].split('=')[1]
    sig0 = line.split(',')
    new_lst=[]
    for j in range(0,len(sig0)-1):
        k= float(sig0[j])
        new_lst.append(k)
    sig['sigma'+str(i+8)] = np.array(new_lst)

gamma_of_na = np.array((0.7156, 0.44982500000000003, 0.32737499999999997,
                    0.192225, 0.07915000000000001, 0.016399999999999998))
gamma_nad = gamma_of_na * 0.0588 + 0.286
gon = np.array((0.991709633,0.893821899,0.645290442,0.378629871,
                    0.196398598,0.118554369))
```

```
goff = np.array((0.998385086,0.978470252,0.918588976,0.826254485,
            0.710329562,0.581357776))
goff_5registry= np.array((0.9998,0.9971,0.9890,0.9755,0.9569,0.9334))
S0 = 1.33
Slab = 0.9852

#### objective function
def chi_square_all(data):

    def ind_chi(i):

        j = data[i-2]
        x = data[i-1]
        y = data[i]
        z = data[i+1]
        k = data[i+2]
        t = i + 6

        gon_def = gon
        goff_def = goff
        goff_def_2 = goff_5registry
        gamma_naddef = gamma_nad
        exp_def = exp_0['exp'+str(t)]
        sig_def = sig['sigma'+str(t)]

        chi_square_ind = np.array(np.sum(pow((S0-Slab/sum(data)*(0.98*gon_def*y+(z+x)*0.98*goff_def+(j+k) *
0.98*goff_def_2
                                            + sum(data) -x-z-y-j-k)-gamma_naddef)/S0
                                    -exp_def,2)/pow(sig_def,2)))
        return chi_square_ind

    chi_square_sum = np.sum([ind_chi(i) for i in range (2, len(data)-2)])

    return chi_square_sum
another_bounds = [[0,1.0]]*21
result_1 = dual_annealing(chi_square_all, another_bounds, maxiter=1000, accept=-5)
```

2. WT_DA_3reg_b_Uncon_6data (b1 = 0.98 b2 = 0.99 b3 = 1.0)

```
import numpy as np
from scipy.optimize import dual_annealing

# Input data from files
fname_exp = '/Users/yijinzhang/Desktop/REDOR/python_WT/exp.txt'
fh_exp = open(fname_exp)
fname_sig = '/Users/yijinzhang/Desktop/REDOR/python_WT/sigma.txt'
fh_sig = open(fname_sig)

lst_exp = []
for line in fh_exp:
    line = line.rstrip()
    lst_exp.append(line)
exp_0 = {}

for i in range(len(lst_exp)):
```

```
    line = lst_exp[i].split('=')[1]
    exp0 = line.split(',')
    new_lst=[]
    for j in range(len(exp0)-1):
        k= float(exp0[j])
        new_lst.append(k)
    exp_0['exp'+str(i+8)] = np.array(new_lst)

lst_sig = []
for line in fh_sig:
    line = line.rstrip()
    lst_sig.append(line)

sig = {}
for i in range(0,len(lst_sig)):
    line = lst_sig[i].split('=')[1]
    sig0 = line.split(',')
    new_lst=[]
    for j in range(0,len(sig0)-1):
        k= float(sig0[j])
        new_lst.append(k)
    sig['sigma'+str(i+8)] = np.array(new_lst)

gamma_of_na = np.array((0.7156, 0.44982500000000003, 0.32737499999999997,
                    0.192225, 0.07915000000000001, 0.016399999999999998))
gamma_nad = gamma_of_na * 0.0588 + 0.286
guu = np.array((0.9917,0.8938,0.6453,0.3786,
                    0.1964,0.1186))
# gamma for t_top=u-1 t_bottom=u
gum1u = np.array((0.9980, 0.9737, 0.9012, 0.7914, 0.6571, 0.5132))
# gamma for t_top=u+1 t_bottom=u
gup1u = np.array((0.9980, 0.9737, 0.9012, 0.7914, 0.6571, 0.5132))
# gamma for t_top=x t_bottom=u
gxu = np.array((0.9989, 0.985, 0.9427, 0.8754, 0.787, 0.6824)
# gamma for t_top=u t_bottom=u-1
guum1 = np.array((0.9921, 0.8974, 0.6476, 0.3565, 0.1288, 0.0156))
# gamma for t_top=u-1 t_bottom=u-1
gum1um1 = np.array((0.9984,0.9785,0.9186,0.8263,
            0.7103,0.5814))
# gamma for t_top=u+1 t_bottom=u-1
gup1um1 = np.array((0.9984, 0.9785, 0.9191, 0.8293, 0.7200, 0.6037))
# gamma for t_top=x, t_bottom=u-1
gxum1 = np.array((0.9992, 0.9898, 0.9608, 0.9143, 0.8521, 0.7769))
# gamma for t_top=u t_bottom=u+1
guup1 = np.array((0.9921, 0.8974, 0.6476, 0.3565, 0.1288, 0.0156))
# gamma for t_top=u-1 t_bottom=u+1
gum1up1 = np.array((0.9984, 0.9785, 0.9191, 0.8293, 0.7200, 0.6037))
# gamma for t_top=u+1 t_bottom=u+1
gup1up1 = np.array((0.9984,0.9785,0.9186,0.8263,
            0.7103,0.5814))
# gamma for t_top=x t_bottom=u+1
gxup1 = np.array((0.9992, 0.9898, 0.9608, 0.9143, 0.8521, 0.7769))
# gamma for t_top=u t_bottom=x
gux = np.array((0.9928, 0.9064, 0.671, 0.3778, 0.1236, -0.023))
# gamma for t_top=u-1 t_bottom=x
```

```
gum1x = np.array((0.9991, 0.9885, 0.956, 0.9039, 0.8348, 0.752))
# gamma for t_top=u+1 t_bottom=x
gup1x = np.array((0.9991, 0.9885, 0.956, 0.9039, 0.8348, 0.752))
S0 = 1.33
Slab = 0.9852

#### objective function
def chi_square_all(data):

    def ind_chi(i):

        x = data[i-1]
        y = data[i]
        z = data[i+1]
        rest = sum(data)-x-y-z
        t = i + 7

        # gamma_naddef = gamma_nad
        exp_def = exp_0['exp'+str(t)]
        sig_def = sig['sigma'+str(t)]

        # scaling factor b1(t_top=u,u±1;t_bottom=u,u±1) = 0.98
        # b2(t_top=u,u±1; t_bottom=x) = b3 (t_top=x;t_bottom=u,u±1) = 0.99
        # b3 (t_top=x;t_bottom=x)=1.0
        b1 = 0.98
        b2 = 0.99
        b3 = 1.0

        cal = (b1*guu*y*y+b1*gum1u*x*y+b1*gup1u*z*y+b2*gxu*rest*y+
         b1*guum1*y*x+b1*gum1um1*x*x+b1*gup1um1*z*x+b2*gxum1*rest*x+
         b1*guup1*y*z+b1*gum1up1*x*z+b1*gup1up1*z*z+b2*gxup1*rest*z+
         b2*gux*y*rest+b2*gum1x*x*rest+b2*gup1x*z*rest+
         b3*rest*rest)

        chi_square_ind = np.array(np.sum(pow(((S0-Slab/pow(sum(data),2)*cal
                                    -gamma_nad)/S0
                                    -exp_def),2)/pow(sig_def,2)))
        return chi_square_ind

    chi_square_sum = np.sum([ind_chi(i) for i in range (1, len(data)-1)])

    return chi_square_sum

another_bounds =[[0,0.00001]]+ [[0,1.0]]*17+[[0,0.00001]
result = dual_annealing(chi_square_all, another_bounds, maxiter=1000, accept=-5)
```