

Origin of Conformational Dynamics in a Globular Protein

Adam M. Damry, Marc M. Mayer, Aron Broom, Natalie K. Goto & Roberto A. Chica*

Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie-Curie,
Ottawa, Ontario, Canada K1N 6N5

* To whom correspondence should be addressed:

Roberto A. Chica, Email: rchica@uottawa.ca

Abstract

Protein structures are dynamic, undergoing specific motions that can play a vital role in function. However, the link between primary sequence and conformational dynamics remains poorly understood. Here, we studied how conformational dynamics can arise in a globular protein by evaluating the impact of individual substitutions of core residues in DANCER-3, a streptococcal protein G domain $\beta 1$ ($G\beta 1$) variant that we previously designed to undergo a specific mode of conformational exchange that has never been observed in the wild-type protein. Using a combination of solution NMR experiments and molecular dynamics simulations, we demonstrate that only two mutations are necessary to create this conformational exchange, and that these mutations work synergistically, with one destabilizing the native $G\beta 1$ structure and the other allowing two new conformational states to be accessed on the energy landscape. Overall, our results show how conformational dynamics can appear in a stable globular fold, a critical step in the molecular evolution of new dynamics-linked functions.

Introduction

Proteins carry out most of the intricate biological processes required to sustain life. These functions often involve motions,¹⁻³ with conformational dynamics playing a vital role in enzyme catalysis,^{4,5} allosteric regulation,⁶ and molecular recognition.⁷ Despite the demonstrated importance of dynamics for protein function, the link between protein sequence and dynamics remains poorly understood,⁸ and epistasis resulting from the ruggedness of the protein energy landscape complicates attempts to study how sequence elements contribute to dynamics and thereby function in natural proteins.⁹ Progress towards this goal has been made by evolution-based studies,¹⁰⁻¹² with key findings having shown that novel protein functions can arise from new dynamic regimes that reorganize functional sites.^{13,14} However, these studies have focused on function-expanding dynamics that develop in metastable regions of proteins where preexisting motions responsible for a promiscuous activity can be amplified and modulated with small changes to protein sequence and structure.^{10,15} Therefore, a fundamental understanding of how sequence changes can lead to novel modes of conformational exchange in globular proteins remains elusive.

Previously, we used computational design to create a series of Streptococcal protein G domain $\beta 1$ (G $\beta 1$) variants termed DANCERS¹⁶ that undergo a specific mode of conformational exchange that has never been observed in the wild-type protein.¹⁷⁻²⁶ This exchange involves spontaneous switching of the Trp43 side chain between core-buried and solvent-exposed states,¹⁶ which is accompanied by rotameric changes to residues in the protein core and by displacement of the helix relative to the β -sheet that nevertheless preserves the native G $\beta 1$ fold (Fig. 1a). In contrast with other proteins previously used to study the evolution of dynamics,^{10,15,27-30} DANCERS are unique in that the designed Trp43 conformational exchange did not result from alteration of exchange rates and amplitudes between existing low-occupancy conformational states observed in

the parent. WT G β 1 and DANCERs therefore comprise an ideal system to study how conformational exchange can arise in a globular protein, helping us to gain insights on the link between protein sequence and dynamics.

Here, we evaluated the impact on Trp43 conformational exchange of individual core-residue substitutions in a representative DANCER in order to understand the role of each residue in the designed dynamic exchange. Solution NMR experiments showed that only two mutations were necessary to convert the static WT G β 1 into a DANCER capable of undergoing Trp43 conformational exchange. The molecular mechanism by which these mutations induce Trp43 conformational exchange was probed using molecular dynamics simulations, showing that these mutations work synergistically, with one destabilizing the native G β 1 structure and the other allowing two new conformational states to be accessed on the energy landscape. Our results thus provide a model for the molecular evolution of conformational dynamics in globular proteins.

Results

Contribution of individual mutations to exchange

To investigate how sequence changes can lead to novel modes of conformational exchange in a globular protein, we individually reverted each of the five mutations present in DANCER-3 to the corresponding amino acid found in WT G β 1 (Fig. 1b). We chose DANCER-3 for this analysis because of its slow exchange rate and distinct ^1H and ^{15}N chemical shifts for both Trp43 conformational states, facilitating analysis of exchange using nuclear magnetic resonance (NMR) spectroscopy. The resulting five DANCER-3 point mutants (D3-F3Y, D3-I7L, D3-F34A, D3-L39V, and D3-I54V, each named for the mutation they possess relative to DANCER-3) all adopted the native G β 1 fold (Supplementary Fig. 1) and were stably folded (Table 1, Supplementary Fig.

2 and 3). ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectra for all variants with the exception of D3-F34A showed a population of additional peaks not observed in the WT G β 1 spectrum (Supplementary Fig. 4 and 5), indicating that these variants adopt more than one distinct structural state, as observed for DANCER-3.¹⁶ Given the difference in NMR spectra between D3-F34A and all other DANCER mutants studied, we investigated the possibility that DANCER variants form dimers in solution, as the A34F mutation common to these proteins has been shown to induce dimerization when introduced on its own into WT G β 1.^{25,31} Analytical size-exclusion chromatography using WT G β 1 and its A34F point mutant (WT-A34F) as controls for monomeric and dimeric proteins, respectively, revealed that all DANCER-3 variants containing the A34F mutation form weak dimers with K_d values measured by NMR between 31 and 240 μM (Supplementary Fig. 6 and Supplementary Table 1). Although monomer-dimer equilibrium could give rise to the observed populations of peaks in HSQC spectra, we previously confirmed that the Trp43 side chain of DANCERs undergoes conformational exchange by demonstrating that the nuclear Overhauser effect (NOE) contacts made by this residue cannot arise from its adoption of a single conformation in the context of the native G β 1 fold, and by solving solution structures of the major and minor states of the DANCER-2 variant.¹⁶ As the Trp43 conformation in the WT-A34F dimer structure is identical to that of the WT G β 1 monomer (Fig. 1b and Supplementary Fig. 7), these contacts cannot be explained by monomer-dimer equilibrium. Furthermore, the dimer interface lies on the opposite face of the protein from Trp43 (Supplementary Fig. 7), which would only give rise to small changes in local chemical environment around Trp43 upon dimerization, whereas conformational exchange between buried and solvent-exposed states would be expected to give rise to larger changes.

To verify whether the second population of Trp43 peaks in our DANCER variants resulted from monomer-dimer equilibrium or Trp43 conformational exchange, we measured Trp43 ϵ chemical shift displacements ($\Delta\delta$) for these variants as well as WT-A34F and DANCER-3. We observed a smaller Trp43 ϵ $\Delta\delta$ in the WT-A34F spectrum than that of DANCER-3 (0.34 ppm vs. 1.18 ppm, Fig. 1c). HSQC spectra of D3-F3Y, D3-I7L, and D3-I54V showed a $\Delta\delta$ close to that of DANCER-3 (0.94–1.32 ppm). Conversely, the spectrum of D3-L39V instead exhibited a $\Delta\delta$ similar to the one observed for WT-A34F (0.32 ppm), suggesting that D3-L39V does not undergo the Trp43 conformational exchange characteristic of DANCERs, with its Trp43 ϵ $\Delta\delta$ arising from the monomer-dimer equilibrium as in WT-A34F.

To determine if these two peak populations interconvert on a millisecond timescale, as was observed for the Trp43 conformational exchange in DANCER-3, ^1H - ^{15}N HSQC ZZ-exchange experiments were performed on all variants (Supplementary Fig. 8). We observed exchange for Trp43 ϵ as well as Thr17/Thr18, which can be used as G β 1 dimerization reporters due to their localization at the dimer interface (Supplementary Fig. 7). To determine whether Trp43 ϵ exchange arises from monomer-dimer equilibrium, we compared activation energies for Trp43 ϵ exchange with that of Thr17/Thr18. In both WT-A34F and D3-L39V, activation energies calculated from both Thr17 and Trp43 ϵ were the same within error (Table 2, Supplementary Fig. 9), strongly suggesting that dynamics probed at both positions are representative of monomer-dimer exchange. In contrast, DANCER-3 rates for Trp 43 ϵ minor to major state exchange decreased as a function of temperature (Fig. 2a), while Thr17 instead demonstrated the expected increase of exchange rate with temperature. This non-Arrhenius behavior indicates a complex exchange trajectory that may be explained by competing effects between increasing rates of Trp43 side-chain dynamics and decreasing populations of monomeric G β 1 at the higher temperatures (Fig. 2b). Therefore, the net

reduction in measured Trp43 ϵ exchange rates at higher temperatures suggests that this process is inhibited in the dimer. Our prior results with the L39I mutant of DANCER-3 (previously reported as NERD-S, here renamed D3-L39I)¹⁶ corroborate this hypothesis, as D3-L39I exhibits no detectable Trp43 conformational exchange at the concentration range used in our NMR experiments despite significant sequence and spectral similarity to DANCER-3.¹⁶ At these concentrations, the monomer form of DANCER-3 is observable while that of D3-L39I is not. Indeed, we observed only the dimer of D3-L39I whose K_d of $16 \pm 2 \mu\text{M}$ is the lowest among all variants tested here (Supplementary Table 1). Non-Arrhenius behavior for Trp43 ϵ similar to that observed for DANCER-3 was also observed in D3-F3Y, D3-I7L, and D3-I54V (Table 2, Supplementary Fig. 10), confirming that these three mutants exhibit Trp43 conformational exchange similar to that of DANCER-3. In contrast, D3-F34A, which exhibits only a single population of peaks in its HSQC, does not undergo Trp43 conformational or monomer-dimer exchange, and adopts a structure highly similar to that of WT G β 1 (Supplementary Fig. 11, Supplementary Table 2). Altogether, these results confirm that DANCER-3 mutants D3-F3Y, D3-I7L, and D3-I54V undergo Trp43 conformational exchange, but that this mode of molecular motion does not occur in WT G β 1, WT-A34F, D3-F34A, or D3-L39V.

Phe34 and Leu39 synergistically create conformational exchange

Having identified Phe34 and Leu39 as two key residues required for Trp43 conformational exchange in DANCER-3, we next investigated whether these mutations could create this mode of conformational exchange when introduced together into WT G β 1. The resulting WT-A34F/V39L double mutant adopted the native G β 1 fold, was stable, and formed a dimer with a K_d of $116 \pm 5 \mu\text{M}$ (Table 1, Supplementary Fig. 1–3 and 6, Supplementary Table 1). Its HSQC spectrum

demonstrated the presence of two peak populations (Supplementary Fig. 5) with a $\Delta\delta$ of 0.94 ppm (Fig. 1c) for the Trp43 side-chain peak. Furthermore, temperature-dependent exchange of the minor state back to the major state for W43ε, as monitored by ZZ-exchange spectra, showed non-Arrhenius behaviour (Supplementary Fig. 8–10), indicating that Trp43 conformational exchange does occur when both mutations are introduced simultaneously. In contrast, the WT-V39L point mutant, which adopted the native Gβ1 fold and was stable (Table 1, Supplementary Fig. 1–3), was instead monomeric and gave rise to an HSQC spectrum with high similarity to that of WT Gβ1 (Supplementary Fig. 6 and 12), demonstrating that its Trp43 conformation is the same as in WT Gβ1. Taken together, these results show that neither V39L nor A34F alone is sufficient to alter the conformation of Trp43 or create dynamic exchange of this residue when introduced into WT Gβ1. There is therefore a synergistic effect between Leu39 and Phe34 that leads to the specific mode of Trp43 conformational exchange that we designed into DANCERs.

Leu39 stabilizes loop connecting helix with strand β3

To gain insights on the role of the conservative V39L mutation in Trp43 conformational exchange and evaluate its underlying synergistic effect with A34F, we performed μ s-timescale molecular dynamics simulations on DANCER-3, D3-F34A, and D3-L39V. We also included D3-L39I in this analysis because it contains an isosteric mutation at position 39 that might further help to demonstrate the role of this residue in DANCER dynamics. Over the course of each 10- μ s trajectory, the Trp43 side chain in these proteins samples a number of different conformations, defined by their χ_1 and χ_2 dihedrals (Supplementary Fig. 13). The number of transitions between these Trp43 conformations over the course of the simulation is highest in DANCER-3, and close to zero in D3-F34A (Fig. 3a), in agreement with the observed presence or absence of Trp43

conformational exchange measured by NMR (Fig. 1c). D3-L39V also undergoes very few transitions during the course of the molecular dynamics simulations, which is consistent with the fact that Trp43 conformational exchange in this variant was not detected by NMR (Fig. 1c, Table 2). The D3-L39I monomer undergoes the second highest number of transitions even though no measurable Trp43 conformational exchange was observed at the concentrations used for NMR at which D3-L39I is predominantly in the dimeric form. This result, along with a lack of NOEs observed from the W43 H_{ε1} to the protein core,¹⁶ supports our hypothesis that only the monomeric form of DANCERs is capable of undergoing Trp43 conformational exchange.

Molecular dynamics also allowed us to observe other motions in close proximity to Trp43 that may be coupled with its conformational exchange (defined here as a transition between *g*(+), *g*(−), or *t* conformations for the χ_1 dihedral). These simulations revealed one mode of Phe34 motion that appeared to be correlated with Trp43 conformational exchange in DANCER-3 and D3-L39I (Fig. 3b, Supplementary Fig. 13c). Specifically, the introduction of a large aromatic side chain at this position stabilizes a state where the Trp43 side chain has rotated out of the protein core by adopting the *g*(+) conformation, since the Phe34 side-chain can occupy the cavity that would otherwise be created by the Trp43 rotation.¹⁶ In D3-F34A, the smaller Ala34 side chain cannot serve this purpose, and therefore the state with Trp43 in the *g*(+) conformation is not energetically accessible, which prevents this conformational exchange. The role of Leu39 is less clear, as it exhibits no concerted transitions with Trp43. However, we observed that D3-L39V displays enhanced motion in the loop connecting the helix with strand β_3 , which contains position 39, compared to DANCER-3, D3-F34A, and D3-L39I (Fig. 3c). Moreover, this loop in D3-L39V frequently occupies an alternate conformation created by partial unfolding of the helix at its C-terminus (Fig. 3d), which positions residue 39 away from Phe34 and thereby creates a state that

was seldom observed in simulations run for the other variants. A similar loss of helicity in this region of the protein has also been seen in the WT-A34F solution structure (Supplementary Fig. 7). The increase in conformational flexibility in this region may decouple Phe34 side-chain reorientation from that of Trp43 (Supplementary Fig. 13c), since conformational dynamics in the backbone could accommodate the different Phe34 orientations in a way that prevents steric clashes with the native Trp43 side-chain conformation (Fig. 3e). On the other hand, in DANCER-3 the loop containing Leu39 predominantly adopts the native conformation, positioning Leu39 adjacent to Phe34 and maintaining a helical state in the backbone of residue 34 (Supplementary Fig. 13). This more stable backbone conformation is required in order for the Phe34 side chain to move in concert with that of Trp43, enabling Trp43 conformational exchange.

Discussion

Our results demonstrate that two mutations synergistically create a novel mode of conformational exchange in DANCERs. The first mutation, A34F, positions a bulky side chain directly next to Trp43, likely leading to steric clashes that destabilize the protein by 1.6 kcal/mol (Table 1). These clashes could be alleviated by movement of the Trp43 side chain outside of the core. However, when A34F is introduced into WT G β 1, it is instead the helix and the loop connecting it to strand β 3 that undergo a conformational change, which minimizes steric clashes between Phe34 and Trp43. Meanwhile, the V39L mutation does not significantly impact the free energy of unfolding when introduced into WT G β 1. In the context of DANCERs, however, this permissive mutation prevents local unfolding of the G β 1 α -helix, shifting the equilibrium of the loop towards the native conformation. This in turn forces Trp43 to move out of the core, leading to conformational exchange due to the equilibrium between the solvent-exposed and core-buried

Trp43 side-chain conformations. Interestingly, the addition of a single methylene group to residue 39 is sufficient to prevent local unfolding of the helix, demonstrating that a subtle structural change to a key region of a protein can give rise to a large effect on dynamics. Yet, it is not sufficient to simply overpack the core to create dynamics, since mutations such as V54I do not lead to Trp43 conformational exchange in G β 1, and DANCERs are at least as stable as WT-A34F, which has not been shown to undergo this mode of exchange. Instead, it is the introduction of a specific combination of mutations that remodels the energy landscape in a way that favors the emergence of dynamics.

Epistasis caused by ruggedness in the protein energy landscape complicates analysis of the evolution of new modes of conformational exchange in natural proteins. However, the G β 1 variants described here provide a unique opportunity to start addressing this question since there are only two potential pathways by which the conformational exchange might be introduced (Fig. 4). As neither A34F nor V39L alone are sufficient to introduce Trp43 conformational exchange in the context of the WT G β 1 sequence, we expect that the effect of these individual mutations on protein stability will be the primary determinant of their fitness in an evolutionary setting since sufficient stability of the native state is required for function.³² Thus, as the A34F mutation is considerably more destabilizing to the protein than V39L, we would expect the latter to occur earlier in this hypothetical evolutionary pathway for Trp43 conformational exchange. However, the V39L mutation alone cannot displace the mutant protein from its native energy well. Given the differential effects of the A34F mutation when introduced into the context of either WT G β 1 or WT-V39L, we postulate that V39L reshapes the protein's energy landscape to disfavor the alternate loop conformation and thereby favor Trp43 conformational exchange without significantly altering the energy well depth of the native fold. Conversely, the A34F mutation is

highly destabilizing to the native fold and promotes the alternate loop conformation, facilitating access to other energy wells in the surrounding landscape. This explains the synergistic effects of these two mutations, where one destabilizes the native energy well and the other makes neighboring wells more accessible in a way that allows novel dynamic motions to develop.

Based on these results, we propose a hypothetical model for the molecular evolution of conformational dynamics in globular proteins. First, a permissive mutation such as V39L that silently reshapes unexplored regions of the protein energy landscape could spontaneously emerge. This mutation would not result in any phenotypic change to the protein but would remodel the energy landscape in a way that would increase accessibility to alternate states, for example by lowering barriers between these states, and stabilize them such that they could be significantly populated. Next, a destabilizing mutation such as A34F could appear and cause the emergence of a new dynamic process by making the native state less stable than the alternate conformations now made accessible by the permissive mutation, allowing these states to be significantly populated and interconvert. The combination of these two mutations would then form the basis for the evolution of a new function, as it has been shown that plasticity enhances evolvability.³³ Once these function-expanding mutations have been introduced, additional mutations could afterwards emerge to recover stability, such as the stabilizing Y3F, L7I, and V54I mutations in DANCER-3, or, with the right selective pressure, to fine-tune this new dynamic regime in order to create a novel function such as allosteric regulation or molecular recognition. This process parallels previously studied evolutionary processes wherein the development of new dynamics-linked function occurs via the introduction of mutations that amplify dynamics, followed by selection against undesirable competing motions that negatively impact activity or stability.¹⁰ Overall, our results demonstrate how novel dynamic motions arise in proteins and suggest how they could form the basis for the

molecular evolution of new dynamics-linked function, shedding light on the link between protein sequence and dynamics.

Methods

Mutagenesis. Codon-optimized and His-tagged (N-terminus) genes for WT Gβ1 (GenBank entry M13825.1) and DANCER-3 cloned into the pJ414 vector were obtained from ATUM (formerly DNA2.0). All other variants tested here were generated through splicing by overlap extension mutagenesis³⁴ using either WT Gβ1 or DANCER-3 as starting templates, and cloned into the pET-11a vector (Novagen) via the *NdeI/BamHI* restrictions sites. The resulting plasmids were then transformed into chemically competent *E. coli* XL-1 Blue cells (Agilent). The entire reading frame of each plasmid was verified by DNA sequencing, after which the plasmids were transformed into chemically competent *E. coli* BL21-Gold (DE3) cells (Agilent) for protein expression.

Protein expression and purification. Proteins for chemical denaturation assays and circular dichroism measurements were expressed using Luria-Bertani (LB) broth supplemented with 100 µg/mL ampicillin. Proteins for NMR spectroscopy were instead expressed using M9 minimal expression medium supplemented with 1 g/L ¹⁵N-ammonium chloride and/or 3 g/L ¹³C-D-glucose for isotopic enrichment. Cultures were grown at 37 °C with shaking to an OD600 of approximately 0.6 after which protein expression was initiated with 1 mM isopropyl β-D-1-thiogalactopyranoside. Following overnight incubation at either 15 °C or 37 °C with shaking (250 rpm) for cultures grown in LB or M9 medium, respectively, cells were harvested by centrifugation and lysed with an EmulsiFlex-B15 cell disruptor (Avestin). Proteins were purified by immobilized metal affinity chromatography according to the manufacturer's protocol (Qiagen), followed by gel filtration in

10 mM sodium phosphate buffer (pH 7.4) using an ENrich SEC 650 size-exclusion chromatography column (BioRad). Purified samples were concentrated using Amicon Ultracel-3K centrifugal filter units (EMD Millipore).

Analytical size-exclusion chromatography. Analytical size-exclusion chromatography measurements were performed with a Superdex 200 Increase column (ÄKTA). Protein samples of 1 mL at a concentration of 1 mM were loaded onto the column and eluted with 10 mM sodium phosphate buffer (pH 7.4) at a flow rate of 1 mL/min. Peak UV absorbance was used to determine the elution time of each sample.

Circular dichroism and thermal denaturation assays. Circular dichroism measurements were performed with a Jasco J-815 spectrometer using 650- μ L aliquots of each G β 1 sample at a concentration of 40 μ M in 10 mM sodium phosphate buffer (pH 7.4) in a 1-mm path-length quartz cuvette (Jasco). For structural characterization of protein folds, circular dichroism spectra were acquired from 185 to 250 nm, sampled every 1 nm at a rate of 10 nm/min. Three scans were acquired and averaged for each sample. For thermal denaturation assays, samples were heated at a rate of 1 $^{\circ}$ C per minute, and ellipticity at 208 nm was measured every 2 $^{\circ}$ C. T_m values were determined by fitting a 2-term sigmoid function with baseline correction³⁵ using nonlinear least-squares regression. Reversibility was confirmed by comparing circular dichroism spectra acquired before and after thermal denaturation experiments from 185–250 nm at 25 $^{\circ}$ C.

Chemical denaturation assays. Chemical denaturation assays were performed in triplicate using protein samples at a 1 mg/mL concentration. Protein aliquots of 25 μ L in individual wells of UV-

Star 96-well plates (Greiner Bio-One) were mixed with 175 μ L of 0–5 M guanidium chloride solutions (12 points, evenly spaced) and incubated at room temperature for an hour. Fluorescence emission spectra were measured for each sample from 300 nm to 450 nm (excitation at 295 nm and step size of 2 nm) using an Infinite M1000 plate reader (Tecan). Fluorescence was integrated and converted into fraction of unfolded protein values. Error on these values is reported as standard deviation from three replicates at each denaturant concentration for each G β 1 variant. C_m values (concentration of denaturant at midpoint of denaturation) were determined by fitting a 2-term sigmoid function using nonlinear least-squares regression.

NMR spectroscopy. 15 N- and 13 C-labelled G β 1 samples for NMR consisted of 0.1–2.0 mM uniformly labelled protein in 10 mM sodium phosphate buffer (pH 7.4), 10 μ M EDTA, 0.02% sodium azide, 1 \times cOmplete EDTA-free Protease Inhibitor Cocktail (Roche), and 10% D₂O for experiments requiring detection of amide protons or 99% D₂O otherwise. All NMR experiments were performed on a Bruker AVANCEIII HD 600 MHz spectrometer equipped with a triple resonance cryoprobe. HSQC, chemical shift assignment, and NOESY experiments were performed at 25 $^{\circ}$ C, and ZZ-exchange experiments were performed at temperatures varying from 5 $^{\circ}$ C to 25 $^{\circ}$ C. NMR data sets were processed with the NMRPipe software package³⁶ and spectra were analyzed with NMRViewJ (One Moon Scientific).³⁷ Backbone and side-chain chemical shift assignments were obtained from the standard suite of 3D triple resonance experiments, including HSQC, HNCO, HNCACB, and CBCA(CO)NH spectra for backbone assignments, and NH-TOCSY (with mixing time (τ_{mix}) = 75 ms), CCH-TOCSY (τ_{mix} = 32.5 ms), and HCCH-TOCSY (τ_{mix} = 27.5 ms) spectra for side-chain assignments of D3-F34A. Dimer K_d values were calculated using peak volumes for linked major and minor state peaks for Thr17 taken at four different

concentrations per protein sample tested. Average backbone amide chemical shift differences ($\Delta\delta$) between W43 ϵ major and minor states were calculated using $\Delta\delta = ((\Delta\delta_{\text{HN}})^2 + (\Delta\delta_{\text{N}}/5)^2)^{0.5}$ where $\Delta\delta_{\text{HN}}$ and $\Delta\delta_{\text{N}}$ are chemical shift differences calculated for protons and nitrogen atoms, respectively. To obtain D3-F34A distance restraints, ^{15}N -edited and ^{13}C -edited HSQC-NOESY ($\tau_{\text{mix}} = 120$ ms) spectra were acquired. ZZ-exchange spectra were acquired at temperatures between 9 °C and 30 °C for which Trp43 ϵ and Thr17 peaks could be deconvolved from neighbouring peaks (or Trp43 ϵ and Thr18 peaks for WT-A34F/V39L and D3-I7L as Thr17 crosspeaks could not be deconvolved at any temperature), and analyzed by fitting a Gaussian model to peaks in distinct clusters and integrating using NMRDraw.³⁶ Rates of exchange were determined by fitting 4-term relaxation and exchange curves³⁸ using nonlinear least-squares regression, and thermodynamic parameters were determined by fitting exchange rates and temperatures to the Eyring equation.

Structure determination. TALOS+³⁹ was used to determine secondary structure propensities and backbone dihedral restraints for D3-F34A on the basis of measured chemical shifts for $^1\text{H}_\alpha$, ^{15}N , $^{13}\text{C}'$, $^{13}\text{C}_\alpha$, and $^{13}\text{C}_\beta$ chemical shifts. Simultaneous NOE assignment and structure calculation was performed using CYANA 2.1.⁴⁰ Chemical shifts from cross peaks in 3D ^{15}N -edited and ^{13}C -edited NOESY-HSQC spectra were used as input, in addition to TALOS+ derived dihedral angle restraints. A total of 618 unique and non-redundant distance restraints were used to calculate 100 conformers, and NMR ensembles were represented by the 10 lowest energy conformers. No distance violations >0.5 Å or torsion angle violations $>5^\circ$ were observed in the resulting structures.

Molecular dynamics simulations. Structures of DANCER-3, D3-L39I, D3-L39V, and D3-F34A were generated from the NMR structure of D3-L39I by introducing mutations using the TRIAD protein design software (Protabit LLC, Pasadena, CA). Prior to molecular dynamics, the structures were parameterized with the AMBER FF14SB forcefield using the LEaP program from the amber suite (<http://ambermd.org/>). The protein was surrounded in a cubic box of TIP3P water with 6 Å as the shortest clearance on a single side, and the electrostatic charge neutralized by addition of sodium ions.

Molecular dynamics simulations were performed using in-house code written with the OpenMM API (version 7, <http://openmm.org/>). A Langevin thermostat was used with a temperature of 298.15 K and a timestep of 2 fs. A Monte Carlo barostat was employed with a pressure of 1 atm, periodic boundary conditions and particle mesh Ewald summation with a long-range cutoff of 12 Å. The lengths of all bonds to hydrogen atoms were constrained. Energy minimization was performed until a convergence of 10 kJ/mol, followed by 100 ns of equilibration. Following equilibration, 5 production replicas were run for 3 μs each, for a total of 15 μs of simulation time per protein. All analysis of the molecular dynamics trajectories was performed using VMD (www.ks.uiuc.edu/Research/vmd/) on the final 2 μs of each replica trajectory for each protein. Thus, a total of 10 μs simulation time per protein was used for analysis.

Accession codes

Structure coordinates for D3-F34A have been deposited in the Protein Data Bank with the accession code 6NJF. NMR data for D3-F34A has been deposited in the Biological Magnetic Resonance Data Bank with the accession code 30532.

Acknowledgments

R.A.C. acknowledges an Early Researcher Award from the Ontario Ministry of Economic Development & Innovation (ER14-10-139), and grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-04831) and the Canada Foundation for Innovation (26503). N.K.G. acknowledges a grant from NSERC (RGPIN-2011-20298378). A.M.D. is the recipient of a NSERC postgraduate scholarship. A.B. is the recipient of a postdoctoral fellowship from NSERC. M.M.M. is the recipient of an Ontario Graduate Scholarship. We acknowledge Dr. Glenn Facey for assistance with NMR experiments.

Author Contributions

A.M.D. and R.A.C. conceived the project. A.M.D. and M.M.M. performed biophysical characterization experiments. A.B. performed all computational experiments. N.K.G. and A.M.D. designed NMR experiments and analyzed data. A.B. and R.A.C. designed computational experiments. All authors wrote the manuscript.

Competing Interests

The authors declare no competing financial interests.

References

- 1 Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: signalling and other functions. *Curr Opin Struct Biol* **20**, 142-147 (2010).
- 2 Bhabha, G. *et al.* Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol* **20**, 1243-1249 (2013).
- 3 Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669-673, doi:10.1038/nature08615 (2009).
- 4 Bhabha, G. *et al.* A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**, 234-238, doi:10.1126/science.1198542 (2011).

- 5 Kerns, S. J. *et al.* The energy landscape of adenylate kinase during catalysis. *Nat Struct Mol Biol* **22**, 124-131, doi:10.1038/nsmb.2941 (2015).
- 6 Tzeng, S. R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462**, 368-372, doi:10.1038/nature08560 (2009).
- 7 Tuinstra, R. L. *et al.* Interconversion between two unrelated protein folds in the lysozyme native state. *Proceedings of the National Academy of Sciences* **105**, 5057-5062, doi:10.1073/pnas.0709518105 (2008).
- 8 Hensen, U. *et al.* Exploring Protein Dynamics Space: The Dynasome as the Missing Link between Protein Structure and Function. *PLoS One* **7**, e33931 (2012).
- 9 Miton, C. M. & Tokuriki, N. How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci* **25**, 1260-1272, doi:10.1002/pro.2876 (2016).
- 10 Campbell, E. *et al.* The role of protein dynamics in the evolution of new enzyme function. *Nature chemical biology* **12**, 944-950, doi:10.1038/nchembio.2175 (2016).
- 11 Obexer, R. *et al.* Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat Chem* **9**, 50-56 (2017).
- 12 Risso, V. A. *et al.* *De novo* active sites for resurrected Precambrian enzymes. *Nat Commun* **8** (2017).
- 13 Campbell, E. C. *et al.* Laboratory evolution of protein conformational dynamics. *Curr Opin Struct Biol* **50**, 49-57 (2018).
- 14 Tokuriki, N. & S., T. D. Protein dynamism and evolvability. *Science* **324**, 203-207 (2009).
- 15 Clifton, B. E. & Jackson, C. J. Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins. *Cell Chem Biol* **23**, 236-245 (2016).
- 16 Davey, J. A., Damry, A. M., Goto, N. K. & Chica, R. A. Rational design of proteins that exchange on functional timescales. *Nature chemical biology* **13**, 1280-1285, doi:10.1038/nchembio.2503 (2017).
- 17 Gronenborn, A. M. *et al.* A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657-661 (1991).
- 18 Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **33**, 4721-4729 (1994).
- 19 Derrick, J. P. & Wigley, D. B. The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab. *J Mol Biol* **243**, 906-918, doi:10.1006/jmbi.1994.1691 (1994).
- 20 Wylie, B. J. *et al.* Ultrahigh resolution protein structures using NMR chemical shift tensors. *Proc Natl Acad Sci U S A* **108**, 16974-16979, doi:10.1073/pnas.1103728108 (2011).
- 21 Tomlinson, J. H., Green, V. L., Baker, P. J. & Williamson, M. P. Structural origins of pH-dependent chemical shifts in the B1 domain of protein G. *Proteins* **78**, 3000-3016, doi:10.1002/prot.22825 (2010).
- 22 Wilton, D. J., Tunnicliffe, R. B., Kamatari, Y. O., Akasaka, K. & Williamson, M. P. Pressure-induced changes in the solution structure of the GB1 domain of protein G. *Proteins* **71**, 1432-1440, doi:10.1002/prot.21832 (2008).
- 23 Strop, P., Marinescu, A. M. & Mayo, S. L. Structure of a protein G helix variant suggests the importance of helix propensity and helix dipole interactions in protein design. *Protein Sci* **9**, 1391-1394, doi:10.1110/ps.9.7.1391 (2000).
- 24 Saio, T., Ogura, K., Yokochi, M., Kobashigawa, Y. & Inagaki, F. Two-point anchoring of a lanthanide-binding peptide to a target protein enhances the paramagnetic anisotropic effect. *Journal of biomolecular NMR* **44**, 157-166, doi:10.1007/s10858-009-9325-z (2009).

- 25 Jee, J., Ishima, R. & Gronenborn, A. M. Characterization of specific protein association by 15N CPMG relaxation dispersion NMR: the GB1(A34F) monomer-dimer equilibrium. *J Phys Chem B* **112**, 6008-6012, doi:10.1021/jp076094h (2008).
- 26 Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *Journal of the American Chemical Society* **121**, 2337-2338 (1999).
- 27 Tokuriki, N. *et al.* Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun* **3**, doi:10.1038/ncomms2246 (2012).
- 28 Kaltenbach, M., Jackson, C. J., Campbell, E. C., Hollfelder, F. & Tokuriki, N. Reverse evolution leads to genotypic incompatibility despite functional and active site convergence. *eLife* **4**, doi:10.7554/eLife.06492 (2015).
- 29 Biel, J. T., Thompson, M. C., Cunningham, C. N., Corn, J. E. & Fraser, J. S. Flexibility and Design: Conformational Heterogeneity along the Evolutionary Trajectory of a Redesigned Ubiquitin. *Structure* **25**, 739-749, doi:10.1016/j.str.2017.03.009 (2017).
- 30 Phillips, A. H. *et al.* Conformational dynamics control ubiquitin-deubiquitinase interactions and influence in vivo signaling. *Proc Natl Acad Sci U S A* **110**, 11379-11384, doi:10.1073/pnas.1302407110 (2013).
- 31 Jee, J., Byeon, I. J., Louis, J. M. & Gronenborn, A. M. The point mutation A34F causes dimerization of GB1. *Proteins* **71**, 1420-1431, doi:10.1002/prot.21831 (2008).
- 32 DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* **6**, 678-687, doi:10.1038/nrg1672 (2005).
- 33 Sikosek, T. & Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* **11**, doi:10.1098/rsif.2014.0419 (2014).
- 34 Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59, doi:0378-1119(89)90358-2 [pii] (1989).
- 35 Koepf, E. K., Petrassi, H. M., Sudol, M. & Kelly, J. W. WW: An isolated three-stranded antiparallel beta-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Science : A Publication of the Protein Society* **8**, 841-853 (1999).
- 36 Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **6**, 277-293, doi:10.1007/bf00197809 (1995).
- 37 Johnson, B. & Blevins, R. NMR View: A computer program for the visualization and analysis of NMR data. *Journal of biomolecular NMR* **4**, 603-614, doi:10.1007/BF00404272 (1994).
- 38 Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *Journal of biomolecular NMR* **4**, 727-734, doi:10.1007/bf00404280 (1994).
- 39 Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* **44**, 213-223, doi:10.1007/s10858-009-9333-z (2009).
- 40 Güntert, P. in *Protein NMR Techniques* (ed A. Kristina Downing) 353-378 (Humana Press, 2004).

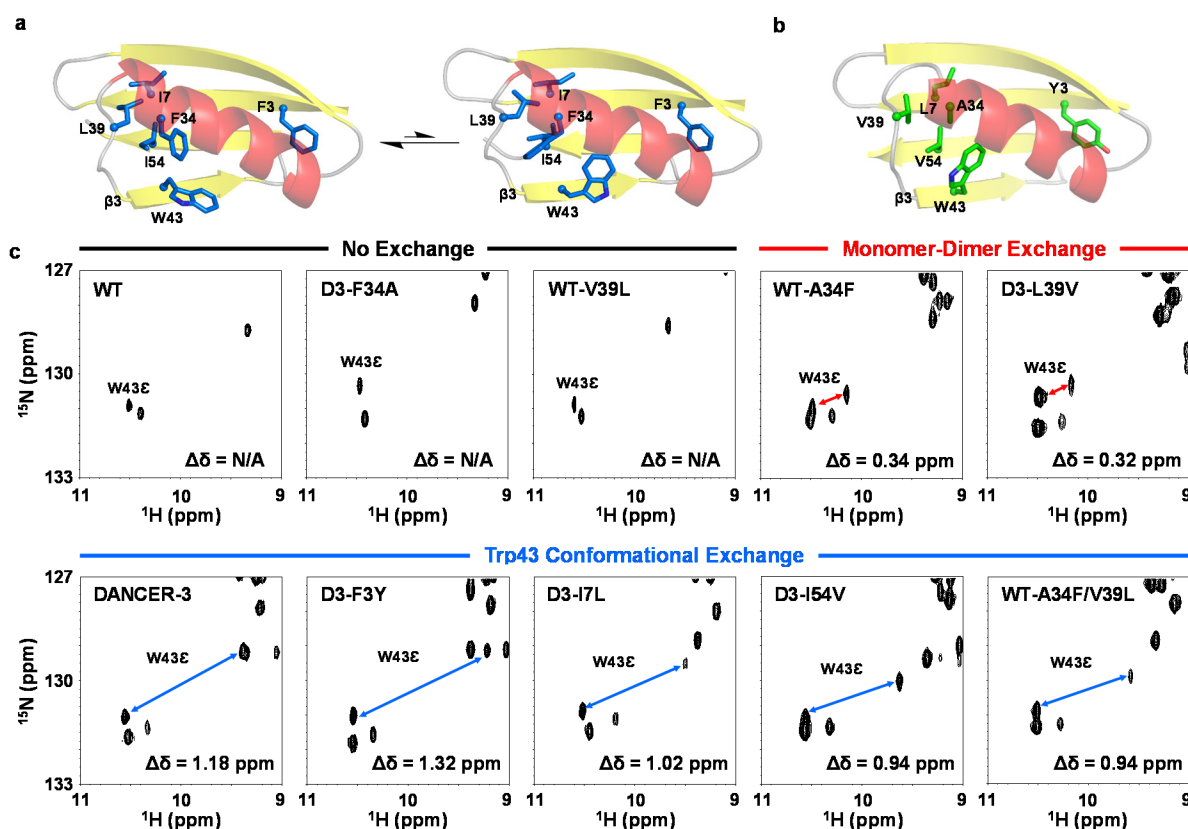


Figure 1. Contribution of individual mutations to exchange. **a**, Structures of the DANCER-3 major and minor states (modeled on PDB ID: 5UBS) and **b**, WT G β 1 (PDB ID: 1PGA) show the side chains of Trp43 and core residues at mutated positions. α -helices, β -strands, and loops are colored red, yellow, and grey, respectively, and strand β 3 is highlighted. **c**, ^1H - ^{15}N HSQC spectra of selected G β 1 variants. W43 ϵ minor state peaks identified using ZZ-exchange experiments are shown linked to their corresponding major state peak with a double-sided arrow. Chemical-shift displacements ($\Delta\delta$) between major and minor state peaks are indicated (N/A = not applicable). Variants were grouped as displaying no exchange, monomer-dimer exchange, or Trp43 conformational exchange, according to the magnitude of their calculated $\Delta\delta$.

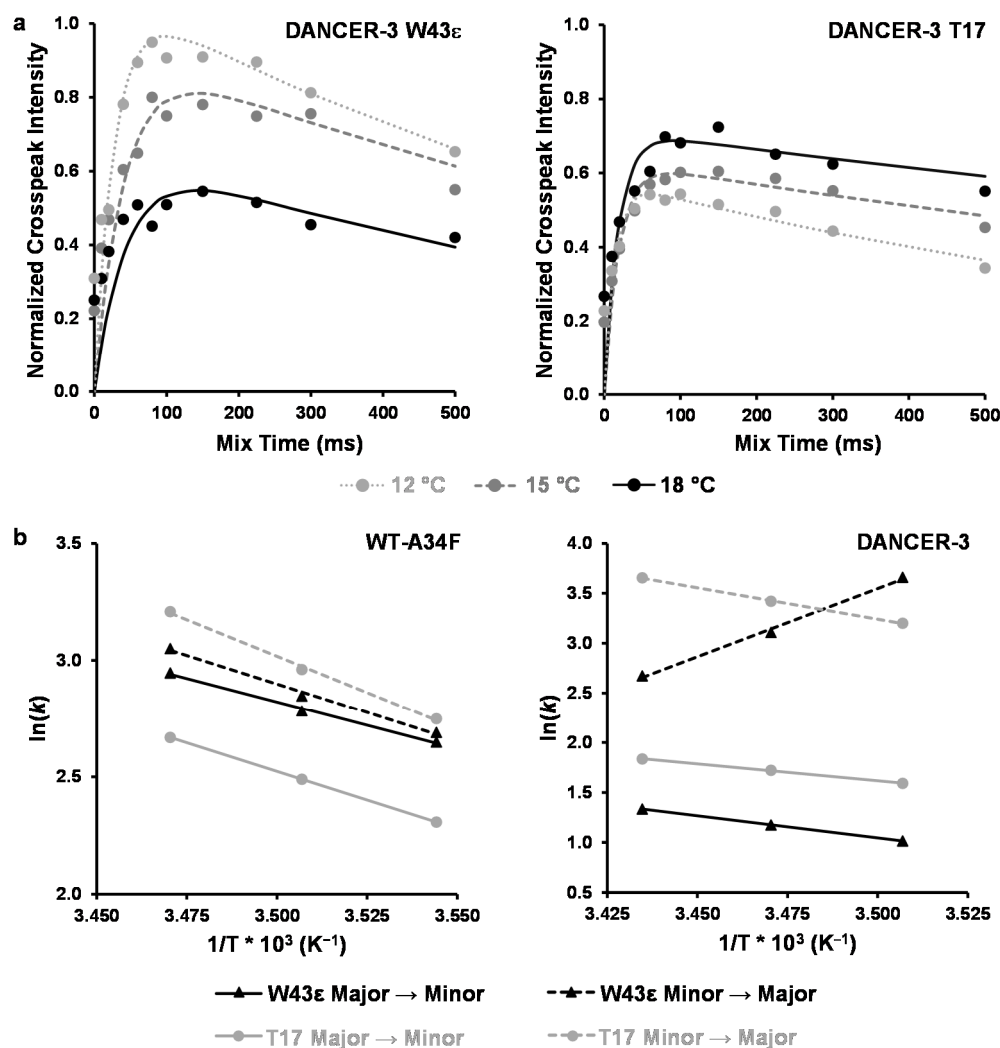


Figure 2. Trp43 conformational exchange does not follow Arrhenius behavior in DANCERs. **a**, ^1H - ^{15}N HSQC ZZ-exchange crosspeak intensity curves for the W43ε minor to major state transition show that this transition slows in DANCER-3 as temperature increases. This unusual behavior is not observed with T17, demonstrating a different mode of exchange for W43ε than monomer-dimer exchange. **b**, Arrhenius plots demonstrate Arrhenius behavior for both W43ε and T17 exchange profiles in WT-A34F, as demonstrated by similar slopes (i.e. activation energies) for all four transitions. DANCER-3 however exhibits non-Arrhenius behavior, as shown by a W43ε minor to major state transition with a non-physical negative activation energy, which indicates that this residue does not follow a two-state Arrhenius model. In contrast, a normal Arrhenius behavior is observed for DANCER-3 T17.

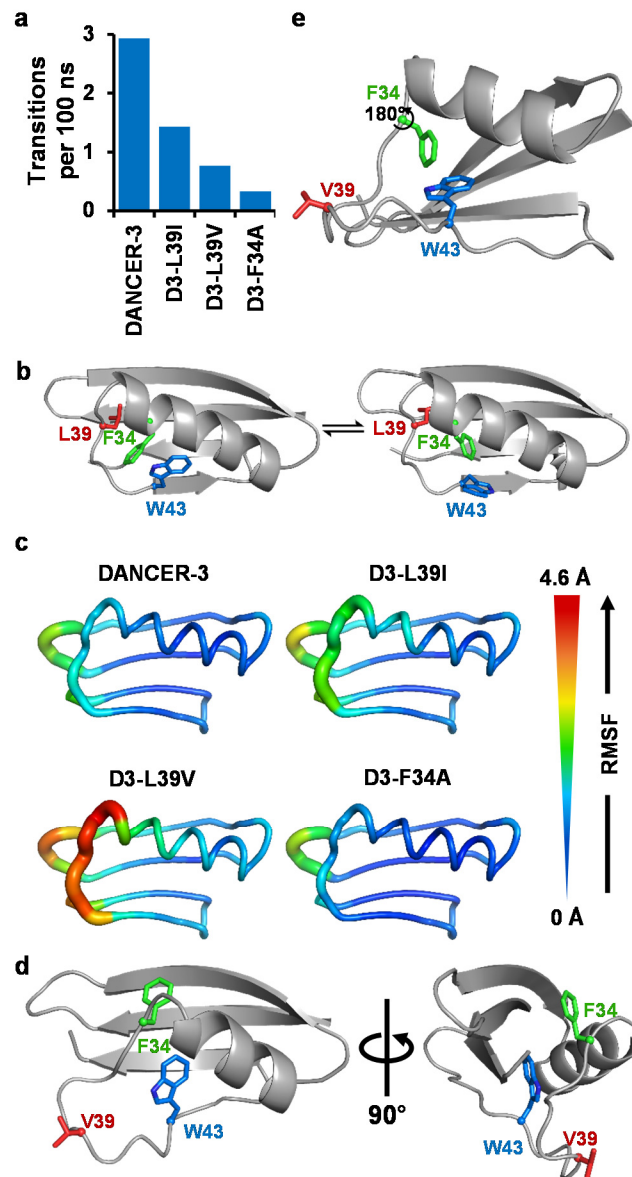


Figure 3. Molecular dynamics results. **a**, The number of Trp43 conformational transitions, defined as χ_1 and χ_2 dihedral changes between bins centered at -60° , 180° , and $+60^\circ$, or -90° and $+90^\circ$, respectively, decreases as a function of experimentally-validated dynamicity. **b**, Representative snapshots sampled from the DANCER-3 simulation demonstrating the concerted motion between Phe34 and Trp43. **c**, Backbone root mean square fluctuation (RMSF) over each simulation plotted on a model backbone for each respective protein. Increased RMSF seen in D3-L39V at the C-terminus of the α -helix and in the loop containing position 39 results from increased mobility and ability to adopt alternate backbone conformations. **d**, D3-L39V populates an alternate loop conformation comprising a partially unfolded α -helix. **e**, Phe34 is incapable of displacing Trp43 while satisfying the backbone conformation described in panel **d** as shown by a model demonstrating that a 180° χ_1 rotation of the Phe34 side-chain could be allowed without significant displacement of the Trp43 side-chain.

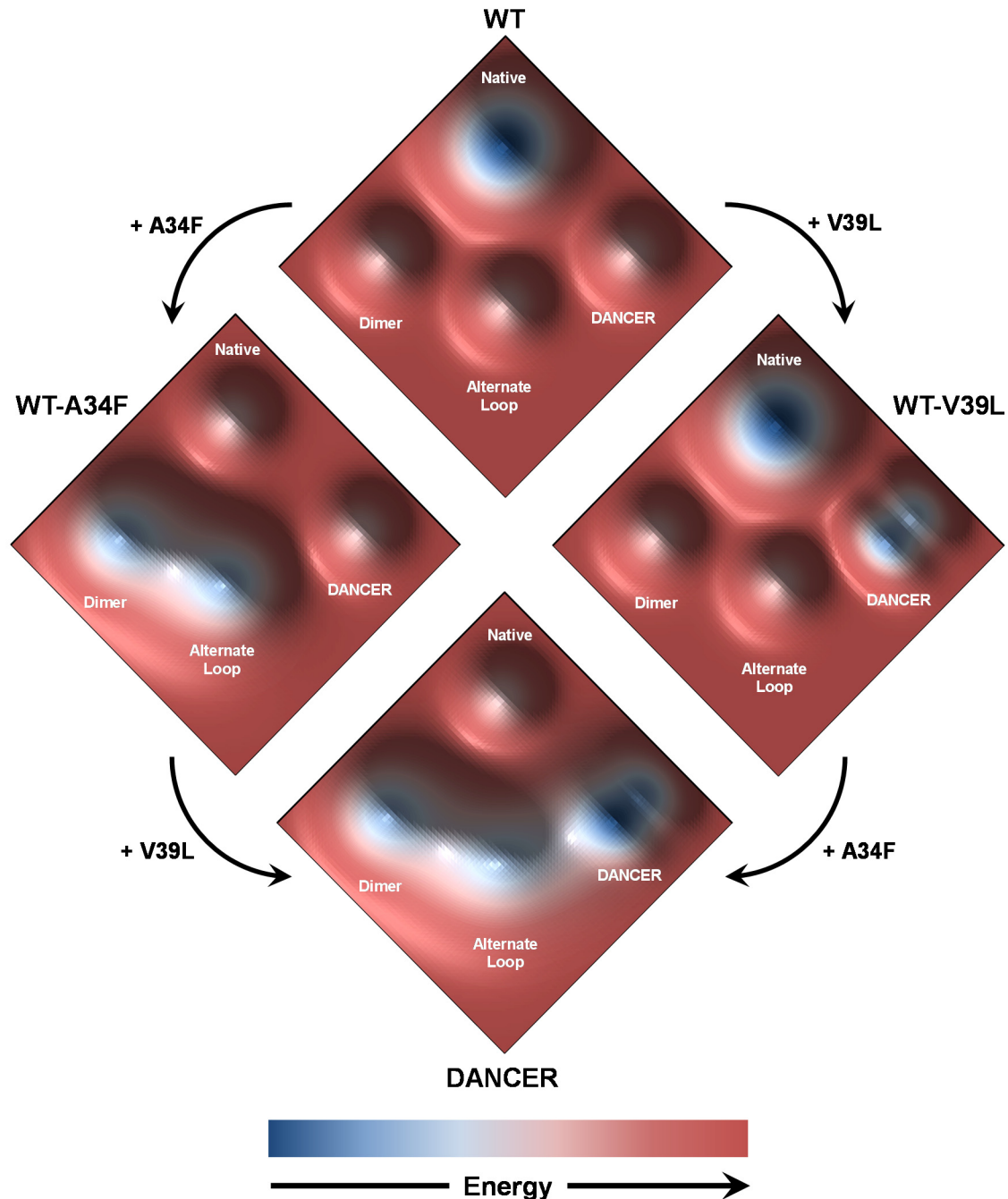


Figure 4. Proposed model for the origin of conformational exchange in DANCERS. Major conformational states sampled by Gβ1 variants are shown on a hypothetical energy landscape. Hypothetical energy well depths are based on NMR and molecular dynamics data suggesting which states can be sampled by each variant, and on measured equilibrium constants between these states. Though the DANCER energy well contains two distinct structural states (e.g., core-buried and solvent-exposed states), the energy barrier between these states is much lower than those separating them from either the native or alternate loop states. Therefore, a single DANCER state is indicated for simplicity.

Table 1. Stability of Gβ1 variants

Protein	Mutations from WT	T _m ^a (° C)	C _m ^b (M)	ΔG _U ^c (kcal/mol)
Wild Type (WT)	–	81	2.66 ± 0.04	4.3 ± 0.3
WT-A34F	A34F	57	1.02 ± 0.04	2.7 ± 0.2
WT-V39L	V39L	74	2.32 ± 0.02	4.3 ± 0.4
WT-A34F/V39L	A34F/V39L	61	1.34 ± 0.01	3.0 ± 0.1
DANCER-3	Y3F/L7I/A34F/V39L/V54I	71	1.72 ± 0.05	3.4 ± 0.4
D3-F3Y	L7I/A34F/V39L/V54I	66	1.66 ± 0.06	3.3 ± 0.3
D3-I7L	Y3F/A34F/V39L/V54I	60	1.30 ± 0.03	2.8 ± 0.4
D3-F34A	Y3F/L7I/V39L/V54I	84	2.74 ± 0.07	4.3 ± 0.1
D3-L39V	Y3F/L7I/A34F/V54I	68	1.75 ± 0.05	2.9 ± 0.1
D3-I54V	Y3F/L7I/A34F/V39L	65	1.82 ± 0.06	3.2 ± 0.4

^a Thermal denaturation midpoint temperature determined through loss of circular dichroism signal at 208 nm

^b Concentration of guanidinium chloride at denaturation midpoint (25 °C, n = 3, mean ± s.d.)

^c Free energy of unfolding determined by chemical denaturation with guanidinium chloride at 25 °C (n = 3, mean ± s.d.)

Table 2. Monomer-dimer and Trp43 conformational exchange activation energies.

Sequence	Peaks	$E_a^{\text{app}_{1 \rightarrow 2}}$ ^a (kcal/mol)	$E_a^{\text{app}_{2 \rightarrow 1}}$ ^b (kcal/mol)	Arrhenius Behavior ^c
WT-A34F	W43ε	8 ± 1	10 ± 2	Yes
	T17	10 ± 1	12 ± 1	Yes
D3-L39V	W43ε	19 ± 2	16 ± 2	Yes
	T17	18 ± 2	17 ± 2	Yes
DANCER-3	W43ε	9 ± 1	-27 ± 3	No
	T17	7 ± 1	12 ± 1	Yes
D3-F3Y	W43ε	3 ± 1	1.6 ± 0.3	Yes ^d
	T17	13 ± 1	16 ± 2	Yes
D3-I7L	W43ε ^e	N/A	N/A	N/A
	T18	31 ± 3	25 ± 2	Yes
D3-I54V	W43ε	0 ^f	0 ^f	No ^d
	T17	8 ± 2	12 ± 1	Yes
WT-A34F/V39L	W43ε	1.6 ± 0.2	-4.7 ± 0.6	No
	T18	9 ± 1	13 ± 1	Yes

^a Apparent activation energy for exchange from the major state (the most populated state at the lowest temperature tested) to the minor state (the least populated state at the lowest temperature tested) (n = 2 analytical replicates, mean ± s.d.)

^b Apparent activation energy for exchange from the minor state to the major state (n = 2 analytical replicates, mean ± s.d.)

^c Arrhenius behaviour is defined here as a positive activation energy for both transitions

^d Although D3-F3Y activation energies are both positive, they are weak compared to T17 (Supplementary Figure 10). It is expected that for one process, they would be similar. This is not the case, thus different dynamic processes are being sensed by D3-F3Y Trp43ε and T17

^e W43ε minor state peaks and crosspeaks could not be quantified for D3-I7L but are present

^f Exchange rates did not vary as a function of temperature over the range measured