

# Native proteins trap high-energy transit conformations

Andrew E. Brereton and P. Andrew Karplus\*

During protein folding and as part of some conformational changes that regulate protein function, the polypeptide chain must traverse high-energy barriers that separate the commonly adopted low-energy conformations. How distortions in peptide geometry allow these barrier-crossing transitions is a fundamental open question. One such important transition involves the movement of a non-glycine residue between the left side of the Ramachandran plot (that is,  $\varphi < 0^\circ$ ) and the right side (that is,  $\varphi > 0^\circ$ ). We report that high-energy conformations with  $\varphi \sim 0^\circ$ , normally expected to occur only as fleeting transition states, are stably trapped in certain highly resolved native protein structures and that an analysis of these residues provides a detailed, experimentally derived map of the bond angle distortions taking place along the transition path. This unanticipated information lays to rest any uncertainty about whether such transitions are possible and how they occur, and in doing so lays a firm foundation for theoretical studies to better understand the transitions between basins that have been little studied but are integrally involved in protein folding and function. Also, the context of one such residue shows that even a designed highly stable protein can harbor substantial unfavorable interactions.

## INTRODUCTION

Proteins carry out a myriad of functions that are enabled by their three-dimensional structures, and decades of research have led to more than 100,000 structures in the Protein Data Bank [PDB (1)] and substantial understanding of protein folding and dynamics [for example, (2, 3)]. In pioneering work, Ramachandran and co-workers (4) introduced the  $\varphi$  and  $\psi$  torsion angles to describe protein backbone conformations (see Fig. 1, A and B), defining some conformations as “allowed” and others as “disallowed” due to collisions between atoms. Now, state-of-the-art energetics calculations (5) and the distributions of  $\varphi, \psi$  angles seen in high-resolution protein structures (6, 7) recapitulate the main features of the original  $\varphi, \psi$  plots remarkably well. For alanine-like residues (Fig. 1D), these include two well-populated low-energy regions—typically called the  $\alpha$  and  $\beta$  basins—on the left-hand side of the plot (having  $\varphi < 0^\circ$ ) and a single, smaller, reasonably populated low-energy basin—called  $\alpha_L$ —on the right-hand side (having  $\varphi \sim +60^\circ$ ).

Although much study has been devoted to the geometries and relative energetics of the well-populated basins [for example, (8)], how alanine-like residues cross the high-energy barriers near  $\varphi = 0^\circ$  or  $+135^\circ$  (Fig. 1D) that match classically-disallowed regions and separate the common conformations having  $\varphi < 0^\circ$  from those having  $\varphi \sim +60^\circ$  proves to be much more difficult to study [for example, (5, 9)] and remains poorly understood. As estimated by Faller *et al.* (5), the heights of the barriers between the basins are about 5 to 7 kcal/mol (Fig. 1D). These barriers are much lower than the barrier ( $\sim 20$  kcal/mol) associated with *cis-trans* isomerization of proline that can be rate-limiting for folding (10), and thus, the transitions should not be rate-limiting but rather common occurrences during protein folding. Such transitions have also been seen to be important for regulatory conformational switches that govern the function of certain proteins, such as modulating peptide binding by an Src homology 2 domain (SH2 domain) (11) or switching between the low- and high-affinity states of the cell adhesion mediator CD44 (12).

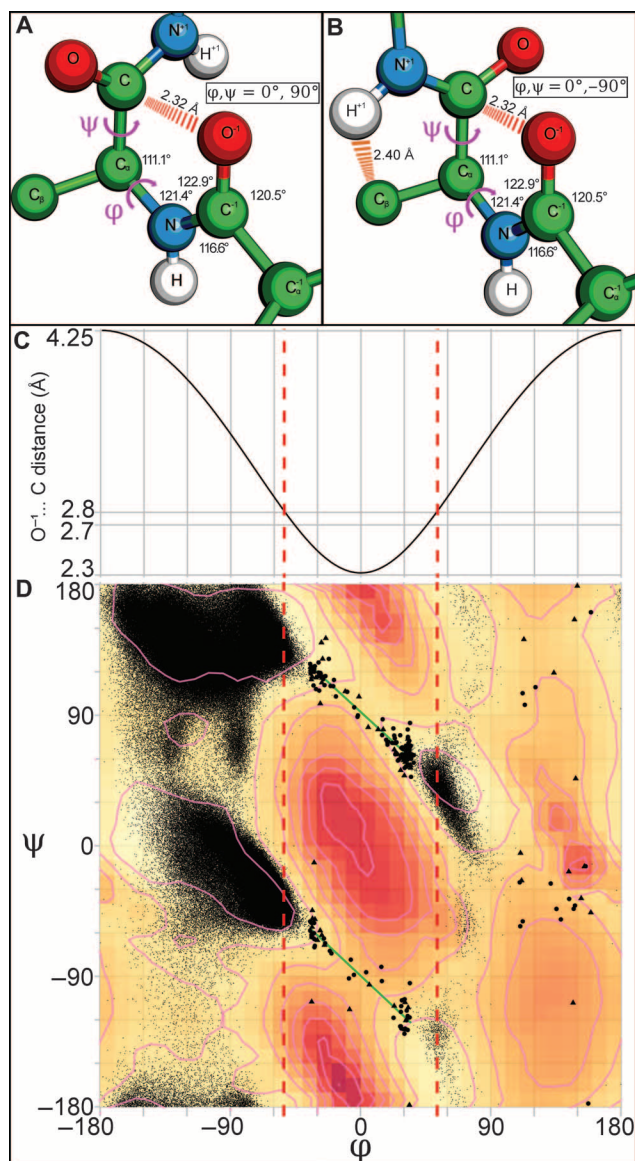
Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331, USA.  
\*Corresponding author. E-mail: karplusp@science.oregonstate.edu

As noted above, a residue must cross one of the two high-energy swaths near  $\varphi = 0^\circ$  or  $+135^\circ$  to transition between the populated conformations having  $\varphi < 0^\circ$  or  $\varphi \sim +60^\circ$ . These regions were classically described as disallowed because of collisions between the carbonyl carbon (C) or the  $C_\beta$  carbon, respectively, and the peptide oxygen of the previous residue ( $O^{-1}$ ). For example, with standard peptide geometry (13), the  $O^{-1}\dots C$  approach at  $\varphi = 0^\circ$  is 2.32 Å (Fig. 1, A to C), which is much closer than the expected extreme contact limit of 2.7 Å (14). Like all transition states, these high-energy transit conformations are expected to be only fleetingly populated and inaccessible to direct experimental characterization, so that there cannot be certainty about what the transition structures really look like. Contrary to this expectation, we have discovered and describe here high-resolution observations of a series of conformations that have been trapped in native protein structures deposited in the PDB and that cover the full range of the  $\varphi \sim 0^\circ$  transitions. The analysis of these observations provides an experimentally derived detailed map of the geometric distortions that take place during these conformational transitions.

## RESULTS

### Reliably modeled residues exist in the two high-energy passes near $\varphi = 0^\circ$

While surveying the conformations of residues in high-resolution ( $\leq 1.5$  Å) protein structures, we were surprised to discover two narrow strings of observations that span completely across the classically disallowed transition regions near  $\varphi = 0^\circ$  (an upper one with  $\psi \sim +90^\circ$  and a lower one with  $\psi \sim -90^\circ$ ) as well as a few sporadic observations in the regions near  $\varphi = +135^\circ$  (Fig. 1D). The existence of residues adopting conformations in the two “mountain passes” through the  $\varphi \sim 0^\circ$  high-energy landscape can be seen in some previously published Ramachandran plots [for example, (15, 16)], but, to our knowledge, the reliability and potential importance of these residues have not been investigated. Even a recent paper that explicitly focused on describing residues in sparsely populated regions of the Ramachandran plot made no mention of these residues, which is consistent with them



**Fig. 1. Populated high-energy passes for transitions between  $\phi < 0^\circ$  and  $\phi > 0^\circ$  conformations.** (A) A standard geometry (13) alanine dipeptide with  $\phi, \psi = 0^\circ, +90^\circ$ . The positive rotation direction for the  $\phi$  and  $\psi$  torsion angles (magenta), the standard values for the five backbone bond angles not involving  $C_\beta$  (black), and the  $O^{-1}\dots C$  clash (red dashes with distance) are indicated. (B) An alanine dipeptide, as in (A), but with  $\phi, \psi = 0^\circ, -90^\circ$ . The  $H^{+1}\dots C_\beta$  approach (orange dashes with distance) also shown, matches the “normal” close approach limit of 2.4 Å for these atoms (14) and causes the  $\psi \sim -90^\circ$  transition track to be somewhat more unfavorable than the  $\psi \sim +90^\circ$  track [ $\sim 7$  versus  $\sim 5$  kcal/mol as seen in (D)]. (C)  $O^{-1}\dots C$  distances as a function of  $\phi$  for standard geometry alanine dipeptides. The expected normal (2.8 Å) and extreme (2.7 Å) approach limits (14) are indicated; red dashed lines at  $\phi = \pm 53^\circ$  mark where the normal approach limit is crossed. (D) A Ramachandran plot with energy contours for the alanine dipeptide calculated using an adaptive force biasing algorithm (5) displayed in steps of 2 kcal/mol (pink). Also shown (small black dots) are 616,212 non-glycine residues from representative  $\leq 1.5$  Å resolution structures; of these, 16,613 (or  $\sim 3\%$ ) have  $\phi > 0^\circ$ . Reliable (large circles) and unreliable (large triangles) observations between  $-35^\circ < \phi < +35^\circ$  and  $+110^\circ < \phi < +160^\circ$  are highlighted. The best-fit lines for reliable residues between  $-35^\circ < \phi < +35^\circ$  are shown for both the  $\psi = +90^\circ$  and  $\psi = -90^\circ$  passes (green).

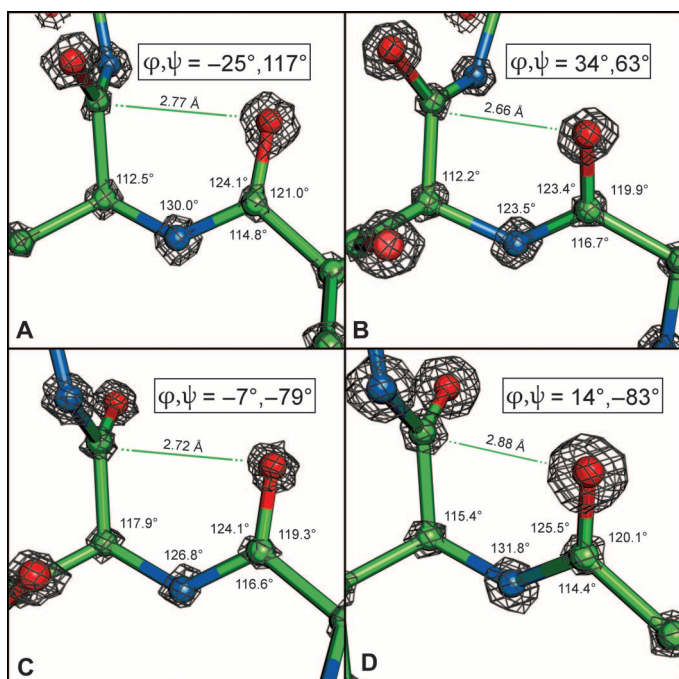
not being considered as reliably observable (15). We carried out visual checks of each of the putative transition residues against its electron density (for example, Fig. 2 and fig. S1) and found that most are reliably defined (Fig. 1D, circles). All the reliably defined residues have  $\phi, \psi$  angles roughly falling within the predicted lowest-energy passes through the high-energy terrain (Fig. 1D). Furthermore, as might be anticipated, the observed residues having  $\phi \sim 0^\circ$  that are not in the low-energy passes were found to be the result of incorrect or unreliable modeling (Fig. 1D, triangles). Because the reliably determined residues with high-energy conformations near  $\phi = 0^\circ$  are real and relatively abundant (146 observations in the  $-35^\circ < \phi < 35^\circ$  transit zone; see table S1), they represent fortuitous “natural experiments” that provide an unprecedented ability to experimentally define at high resolution exactly how the standard peptide geometry becomes distorted as a residue passes through these highly strained conformational transition states. Although 15 residues in the passes near  $\phi \sim +135^\circ$  are also well defined (Fig. 1D), those populations are not yet sufficiently large enough to enable an accurate description of the pathways they represent.

### The $\phi \sim 0^\circ$ transition residues exist in diverse contexts

The  $\phi \sim 0^\circ$  transition residues trapped in native proteins exist in a variety of conformational contexts (fig. S2) and are distributed among 17 of the 20 standard residue types (table S2), implying that they are not special cases but represent realistic snapshots along a transition pathway. Many of these residues are present in or near active sites, but others are not (for example, fig. S3). The cases occurring in two proteins are particularly instructional. In one case, the occurrence proves that even a small, highly stable, designed helical bundle with a melting temperature of  $105^\circ\text{C}$  can accommodate a residue with such high local strain energy (fig. S3A). In the second case, it has been shown that a simple Cys-to-Ala mutation that removes a single hydrogen bond in the active site of an isocyanide hydratase (fig. S3B) leads to the rearrangement of a short backbone segment and the loss of the high-energy conformation (17). Furthermore, it was also shown that a Cys-to-Ser mutation that strengthened the hydrogen bond actually enhanced the stability of the segment in the native conformation (17). This example implies that the energy cost for a residue adopting a high-energy transition conformation can apparently be offset by the formation of a single hydrogen bond and the rearrangement of a few residues.

### Systematic $\phi$ -dependent bond-angle distortions allow passage through the transition region

On a Ramachandran plot, the strip of observations near  $\psi = -90^\circ$  nearly perfectly matches through inversion symmetry that near  $\psi = +90^\circ$  (see Fig. 1D, green lines, and table S3), making it reasonable to treat the two passes as a single phenomenon, roughly doubling the density of observations available for mapping the barrier crossing. To define the patterns of distortion that allow peptides to traverse this barrier, we calculated  $\phi$ -dependent average values for the  $O^{-1}\dots C$  distance and all backbone bond angles. Given the diverse contexts of the residues, treating them as an aggregate should average out specific features due to each particular context and provide a view of the generic transition properties that are solely due to local factors and are generally relevant. This is supported by previous studies showing that the average conformation dependence of backbone bond angles and planarity, found in ultrahigh-resolution protein structures, agrees well with those from quantum mechanics calculations of simple model compounds and those from structures of small peptides (18–21).

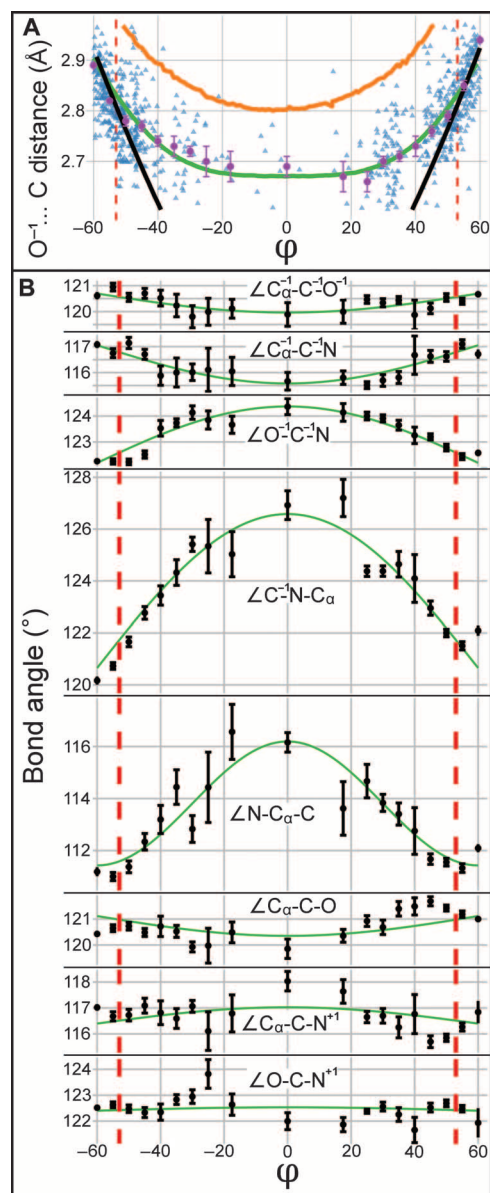


**Fig. 2. Electron density evidence for four residues adopting conformations in the  $-35^\circ < \phi < +35^\circ$  range.** Each panel shows a residue with its  $2F_{\text{O}} - F_{\text{C}}$  electron density, its backbone bond angle values (black), its  $\phi, \psi$  angles (inset box), and its  $\text{O}^{-1} \dots \text{C}$  approach (green line with distance). (A) His<sup>261</sup> from PDB entry 4N1I [1.0 Å resolution; contoured at  $6.2 \times \rho_{\text{rms}}$ ]. (B) Ser<sup>115</sup> from PDB entry 2DDX (0.86 Å resolution; contoured at  $7.0 \times \rho_{\text{rms}}$ ). (C) Asp<sup>249</sup> from PDB entry 4AYO (0.85 Å resolution; contoured at  $7.0 \times \rho_{\text{rms}}$ ). (D) Ile<sup>152</sup> from PDB entry 3NOQ (1.0 Å resolution; contoured at  $5.5 \times \rho_{\text{rms}}$ ).

The behavior of the  $\text{O}^{-1} \dots \text{C}$  distance is striking (Fig. 3A). The average values near  $\phi = \pm 60^\circ$  track with the distance expected for standard geometry, until the distance reaches 2.8 Å (near  $\phi \sim \pm 50^\circ$ ), and then distortion begins and the average distance decreases much less rapidly than predicted by standard geometry, until it reaches  $\sim 2.7$  Å (near  $\phi = \pm 25^\circ$ ). Then, between  $\phi$  of  $-25^\circ$  and  $+25^\circ$ , the observed distance is remarkably flat, with the average distance of  $2.68 \pm 0.02$  Å over that range matching remarkably well with the 2.7 Å “extreme approach limit” for these atom types defined nearly 50 years ago (14).

The  $\phi$ -dependent variations of the backbone bond angles are also systematic, with each angle roughly matching its standard value at  $\phi = \pm 60^\circ$  and varying smoothly to its maximal deformation at  $\phi = 0^\circ$ . Only three bond angles— $\angle \text{O}^{-1} - \text{C}^{-1} - \text{N}$ ,  $\angle \text{C}^{-1} - \text{N} - \text{C}_\alpha$ , and  $\angle \text{N} - \text{C}_\alpha - \text{C}$ —expand substantially, with expansions of roughly  $2^\circ$ ,  $6^\circ$ , and  $4^\circ$ , respectively (Fig. 3B). The lesser expansion of  $\angle \text{O}^{-1} - \text{C}^{-1} - \text{N}$  is consistent with the expectation that as a purely  $\text{sp}^2$ -hybridized center, it would have a higher force constant for resisting distortion. Given the expanding  $\angle \text{O}^{-1} - \text{C}^{-1} - \text{N}$  angle, the  $\angle \text{C}_\alpha^{-1} - \text{C}^{-1} - \text{O}^{-1}$  and  $\angle \text{C}_\alpha^{-1} - \text{C}^{-1} - \text{N}$  angles decrease in a coordinated fashion by  $\sim 1^\circ$  and  $1.5^\circ$ , respectively, to keep the  $\text{C}^{-1}$  carbonyl group largely planar.

To check the validity of treating the  $\psi \sim +90^\circ$  and  $\psi \sim -90^\circ$  passes as equivalent, we analyzed the data from the two passes separately and found that all angles behaved similarly, except that for the  $\psi \sim -90^\circ$  transition,  $\angle \text{C}_\alpha - \text{C} - \text{N}^{+1}$  also expands  $\sim 2^\circ$  (fig. S4), as makes sense to minimize the clash between the  $\text{N}^{+1}$  hydrogen and  $\text{C}_\beta$  (Fig. 1B). We



**Fig. 3. Systematic deformations of geometry associated with transition through the high-energy  $\phi \sim 0^\circ$  passes.** (A) Observed average  $\text{O}^{-1} \dots \text{C}$  distance (large purple dots and error bars) plotted as a function of  $\phi$  (see table S4 for details), along with each data point (blue triangles), and the  $\text{O}^{-1} \dots \text{C}$  distance predicted by standard geometry (black curve), by the empirically defined  $\phi$ -dependent geometry functions (green curve), and by the AMBER FF99SB force field (orange curve). Red dashed lines at  $\phi = \pm 53^\circ$  are as in Fig. 1B. (B) Average backbone bond angles (black dots with error bars) as a function of  $\phi$  (see table S4 for details) along with cosine functions fit to the data (green curves; see table S3 for the equations). All error bars are SEM.

note that, for two reasons, these empirical bond angle distortions may slightly underestimate the actual average distortions: first, structures in the 1.0 to 1.5 Å resolution range are still somewhat influenced by refinement restraints tethering them to the standard values (18, 22), and second, at  $\phi = 0^\circ$ —the point of expected maximal distortion—because of limited data, the empirical value is an average over the broad  $\phi$  range of  $-12.5^\circ$  to  $+12.5^\circ$  (table S4).



## An analytical model for the transition is not matched by predictions from molecular mechanics

These observed  $\phi$  dependencies of the backbone bond angles were modeled as a set of smooth conformation-dependent functions (Fig. 3B, green curves; table S3) that could be used to generate prototype models for the conformational transition. That these yield  $O^{-1}\dots C$  distances (Fig. 3A, green line) matching reasonably well with the empirical averages supports the validity of these functions as capturing a realistic general model for how the  $\phi \sim 0^\circ$  transition is traversed. As noted above, the variations of the individual observations from the average behavior (such as in the examples shown in Fig. 2) are not primarily due to experimental uncertainty but are real variations reflecting the forces caused by the unique tertiary environments that stabilize the transit conformations.

To assess how accurately a state-of-the-art molecular mechanics force field handles these high-energy transition conformations, we used AMBER and the FF99SB force field [recently demonstrated (23) to perform best in a protein modeling test] to minimize conformational energy while restraining  $\phi$  and  $\psi$  to the values along the upper narrow transit path. The energy-minimized  $O^{-1}\dots C$  separation distances (Fig. 3A, orange line) and backbone bond angles (fig. S5) showed qualitative similarity to the empirical variations but were not in good quantitative agreement: the limiting  $O^{-1}\dots C$  approach was  $\sim 0.15$  Å too high and four bond angles had notable systematic displacements from the empirical values, with the largest difference of  $\sim 4^\circ$  occurring for  $\angle C^{-1}-N-C_\alpha$  (fig. S5). These discrepancies imply that the empirical conformational details defined here for the  $\phi \sim 0^\circ$  high-energy conformations provide a real advance in our understanding of this transition, and represent a rare resource for enhancing force field parameterizations.

## DISCUSSION

The observation of these conformations and their conformation-dependent bond angles represent a remarkably detailed experimental characterization of two important conformational transition states that had not been thought to be accessible to direct observation. Importantly, the residues adopting these conformations are not transition-state analogs, artificially held in place by a covalent modification that might alter the pathway; rather, they are authentic residues that are free to transition through the barrier, yet are stabilized partway through by noncovalent interactions with their environment. The fact that the  $\phi, \psi$  angles of the observed transition residues match so well with the low-energy pathway calculated for an isolated dipeptide (Fig. 1D) supports the conclusion that neither the specific protein environments nor the cryogenic temperatures at which most of the structures were determined have changed the nature of the pathway.

In one sense, these images contribute to our understanding of how this transition occurs in the same way that Muybridge's striking "series of instantaneous photographs" of horses provided information previously considered unobservable and showed "with absolute accuracy the motions of horses when walking, trotting, and running" (24). These proved that all four legs of the horse are off the ground roughly half of the time even during a trot. Similarly, the observations presented here provide indisputable evidence that proteins can truly adopt these unfavorable  $\phi \sim 0$  conformations and, on the basis of direct observation, can reveal, in high-resolution detail, the nature of the bond angle deformations that are involved. Just as Muybridge's photographs strung together could provide an observation-based movie of a horse in motion,

our empirically derived analytical functions allow us to generate such a movie of a peptide traversing the mountain pass (movie S1).

In contrast, although molecular simulations are powerful, if simulations were the only source of information, many uncertainties would remain. One illustration of this is the discrepancies between the approach distances and distortions observed here and those predicted by the AMBER force field (Fig. 3A and fig. S5). Another illustration is a molecular dynamics study of the conformational switch in the SH2 domain for which a residue goes from the  $\alpha_L$  basin ( $\phi, \psi \sim +60^\circ, +60^\circ$ ) to the  $\beta$  basin ( $\phi, \psi \sim -60^\circ, +120^\circ$ ). Acknowledging they could not be certain which was the preferred path, and on the basis of a lower predicted energy in their molecular mechanics calculations, the authors proposed that the residue traversed the longer path through the high-energy pass near  $\phi \sim +135^\circ$  (11). Our results suggest that the shorter path through the mountain pass at  $\phi, \psi \sim +0^\circ, +90^\circ$  should be reconsidered as an a priori more likely path.

In terms of the larger picture of protein folding and function, these analyses bring new clarity on how this fundamentally important transition occurs and the level of distortions that peptides are subject to. As such, they provide a foundation for future investigations of the important but little studied area of high-energy barrier crossings and open the door for a richer understanding of folding routes and conformational transitions. On a practical level, this work also provides conformation-dependent restraints, similar to those previously developed for well-populated areas of the Ramachandran plot (18, 22) that can both guide force field development and enhance the accuracy that can be achieved in experimental [for example, (25)] and predictive [for example, (26, 27)] modeling of proteins having residues adopting these rare but important conformations in the  $\phi \sim 0^\circ$  transition region. Finally, this study holds the promise that other high-energy transition conformations can be similarly characterized as the size of the PDB increases and more such observations accumulate.

This work also provides some insight into the thermodynamics of native proteins. It is well known that naturally occurring proteins are not optimized for stability, and this has recently been dramatically illustrated by the creation of a set of designed proteins adopting five different folds and having melting temperatures above  $95^\circ\text{C}$  (28) and of a similarly designed set of superstable helical bundles, one of which had a stability of  $\sim 60$  kcal/mol and a melting temperature above  $135^\circ\text{C}$  (29). In the latter report, it was concluded that "low-energy structures must have unstrained backbone conformations..." (29), but this is not the case given that one of the proteins in our sample was a highly stable designed protein (fig. S3A). That example and the other example noted above in which the high-energy conformation was apparently stabilized by the folding of just a small segment of the protein (fig. S3B) emphasize two things: first, the potential stability achievable by a folded protein is so high that even highly stable proteins may still contain many suboptimal and even some highly unfavorable interactions, and second, suboptimal interactions (that is, "frustration") present in native proteins need not only be present in the form of many slightly unfavorable interactions but can also include individual interactions that are even as high as 5 to 7 kcal/mol destabilizing.

## MATERIALS AND METHODS

### Protein Geometry Database searches

The data set plotted as small dots in Fig. 1D was created using the Protein Geometry Database (PGD) (30), and it includes 616,212 non-

glycine residues. Each of these residues is at the center of a three-residue segment that has backbone, average side-chain, and  $\gamma$ -atom  $B$ -factors  $\leq 25 \text{ \AA}^2$  and is present in a protein crystal structure refined to  $R_{\text{work}}/R_{\text{free}} \leq 0.2/0.25$  at a resolution of  $1.5 \text{ \AA}$  resolution or better and from a protein having  $\leq 90\%$  sequence identity to any other structure in the set. To obtain amino acid frequencies that are representative of diverse sets of proteins (table S2), another smaller data set was generated using a  $\leq 25\%$  sequence identity cutoff.

### Manual curating of the observations in the high-energy passes

On the basis of the above search, all observations having their  $\phi$  torsion angle in one of the high-energy pass regions, either  $-35^\circ < \phi < +35^\circ$  or  $110^\circ < \phi < 160^\circ$ , were manually curated as to the reliability of their conformation on the basis of a visual assessment of the fit to their electron density map. Using conservative criteria, each residue was designated as either reliable (shown as large black dots in Fig. 1D) or unreliable (shown as triangles in Fig. 1D). Residues designated as reliable had to have a strong and well-defined electron density that is not highly anisotropic and a model that was well fit in that density. Observations that lead residues to be deemed unreliable also included the presence of alternate conformations or a close association with uninterpreted density that might indicate alternate conformations. These criteria erred on the side of possibly excluding residues that may have been accurately modeled, rather than including any residues that might not be accurately modeled.

### Generation of modeled peptide structures

All peptides were generated using the PeptideBuilder Python program and library (31), which was slightly modified to be able to handle  $\phi$ -dependent equations instead of single-value standard geometries.

### Calculations of the protein geometries

The set of curated observed residues in the  $-35^\circ < \phi < +35^\circ$  range output by the PGD was used as input for a custom script written in R, which made use of the Bio3D (32) package to read PDB files and then calculate specific geometric details for all residues of interest for each protein, excluding any residue not having at least two residues on both sides of it without a chain break. The quantities calculated included all of the relevant backbone torsion angles and bond angles and the  $\text{O}^{-1} \dots \text{C}$  distances. Bond lengths were not analyzed because it has been shown that their variations are too small to be reliably determined in crystal structures at these resolutions, and also because the conformation-dependent variations are too small to substantially affect modeling accuracy (18, 25). Even those quantities available from the PGD search were recalculated so that the information could also be obtained for the noncrystallographic symmetry (ncs) mates of the PGD hits (which are not present in the PGD). This allowed the 100 unique and curated residues with  $\phi$  in the  $-35^\circ$  to  $+35^\circ$  range to be expanded by the addition of 46 ncs chains (which were also manually curated and deemed as reliable), making for 146 total observations.

### Statistical analyses and least-squares modeling of the data

All averages and standard errors of the mean (SEM) were calculated using conventional formulae written in R. Best-fit lines in Fig. 1D were generated using principal components analysis to fit an

orthogonal linear regression, due to experimental uncertainty in both  $x$  and  $y$  values. The function `prcomp()` in R was used, where

```
x = phi
y = psi
r = prcomp(~x + y)
slope = r$rotation[2,1]/r$rotation[1,1]
intercept = r$center[2] - slope * r$center[1]
```

Independent best-fit lines were calculated for each transit region separately, and there was negligible difference between the two (table S3 and Fig. 1D). The dependence of bond angles on  $\phi$  was fit using the `geom_smooth()` function from the R package “ggplot2,” while specifying “formula = ( $y \sim I(\cos(x * \text{pi}/120))$ ),” where  $x$  is the central value in the phibin and  $y$  is the mean value of the bond angle.

### AMBER minimizations

AMBER calculations were performed using AMBER12 (33) and the FFSB99 force field. Peptides were capped with N-terminal acetyl and C-terminal  $N$ -methyl amide groups. SANDER minimizations were done every  $1^\circ$  in  $\phi$  over the range of  $-60^\circ$  to  $+60^\circ$  with the  $\phi$  and  $\psi$  torsion angles restrained using “NMR restraints” of  $300.0 \text{ kcal/mol} \cdot \text{rad}$  and the  $\psi$  target value set according to the best-fit line given in table S3 for the  $\psi > 0$  pass. Minimizations were carried out two ways, once starting from standard backbone bond angles and another time starting from the  $\phi$ -dependent backbone bond angles as defined in the equations in table S3. For all calculations, the dielectric constant was set to 80 and minimizations were run for 2000 cycles. The results based on both starting points were equivalent; hence, only one is shown in Fig. 3A and fig. S5.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/1/9/e1501188/DC1>

Fig. S1. Electron density evidence for a reliable residue adopting a conformation in the  $+110^\circ < \phi < +160^\circ$  range.

Fig. S2.  $\phi, \psi$  angles describing the local conformational context of the mountain pass residues.

Fig. S3. Four diverse examples showing the contexts of residues adopting a  $\phi \sim 0^\circ$  conformation.

Fig. S4. How the average bond angle variations obtained by treating the  $\psi \leq 0^\circ$  and  $\psi \geq 0^\circ$  transitions separately compare with each other and with those based on the combined data.

Fig. S5. AMBER minimizations of alanine dipeptides distort bond angles to alleviate the  $\text{O}^{-1} \dots \text{C}$  steric clash in  $\phi \sim 0$  conformations.

Table S1. Complete list of analyzed  $\phi \sim 0$  mountain pass residues.

Table S2. Frequency of amino acid types in the mountain pass transition region.

Table S3. Equations governing  $\phi$ -dependent changes in geometry during transition through the mountain pass.

Table S4. Further details of data plotted in Fig. 3 including the ranges for and numbers of observations in each  $\phi$  bin and the average distances and angles.

Movie S1. An alanine dipeptide animation generated according to the “general” model of the  $\psi \sim +90^\circ$  conformational transition described in this paper.

References (34, 35)

### REFERENCES AND NOTES

- P. W. Rose, A. Prlic, C. Bi, W. F. Bluhm, C. H. Christie, S. Dutta, R. K. Green, D. S. Goodsell, J. D. Westbrook, J. Woo, J. Young, C. Zardecki, H. M. Berman, P. E. Bourne, S. K. Burley, The RCSB Protein Data Bank: Views of structural biology for basic and applied research and education. *Nucleic Acids Res.* **43**, D345–D356 (2015).
- K. A. Dill, J. L. MacCallum, The protein-folding problem, 50 years on. *Science* **338**, 1042–1046 (2012).
- C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).

4. G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
5. C. E. Faller, K. A. Reilly, R. D. Hills Jr., O. Guvench, Peptide backbone sampling convergence with the adaptive biasing force algorithm. *J. Phys. Chem. B* **117**, 518–526 (2013).
6. K. Gunasekaran, C. Ramakrishnan, P. Balaram, Disallowed Ramachandran conformations of amino acid residues in protein structures. *J. Mol. Biol.* **264**, 191–198 (1996).
7. V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *MolProbity*: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
8. J. Grdadolnik, V. Mohacek-Grosec, R. L. Baldwin, F. Avbelj, Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1794–1798 (2011).
9. Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker, J. A. McCammon, Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J. Chem. Theory Comput.* **10**, 2677–2689 (2014).
10. P. A. M. Schmidpeter, J. R. Koch, F. X. Schmid, Control of protein function by prolyl isomerization. *Biochim. Biophys. Acta* **1850**, 1973–1982 (2015).
11. O. Guvench, C.-K. Qu, A. D. MacKerell, Tyr66 acts as a conformational switch in the closed-to-open transition of the SHP-2 N-SH2-domain phosphotyrosine-peptide binding cleft. *BMC Struct. Biol.* **7**, 14 (2007).
12. F. W. Jamison II, T. J. Foster, J. A. Barker, R. D. Hills Jr., O. Guvench, Mechanism of binding site conformational switching in the CD44-hyaluronan protein-carbohydrate binding interaction. *J. Mol. Biol.* **406**, 631–647 (2011).
13. R. A. Engh, R. Huber, in *International Tables for Crystallography*, M. G. Rossmann, E. Arnold, Eds. (Kluwer Academic Publishers, Dordrecht, Netherlands, 2001), pp. 382–392.
14. G. N. Ramachandran, V. Sasisekharan, Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–438 (1968).
15. N. V. Kalmankar, C. Ramakrishnan, P. Balaram, Sparsely populated residue conformations in protein structures: Revisiting “experimental” Ramachandran maps. *Proteins Struct. Funct. Bioinf.* **82**, 1101–1112 (2014).
16. J. S. Richardson, D. C. Richardson, D. A. Keedy, in *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map* (World Scientific Publishing Company, Singapore, Singapore, 2013), pp. 46–61.
17. M. Lakshminarasimhan, P. Madzellan, R. Nan, N. M. Milkovic, M. A. Wilson, Evolution of new enzymatic function by structural modulation of cysteine reactivity in *Pseudomonas fluorescens* isocyanide hydratase. *J. Biol. Chem.* **285**, 29651–29661 (2010).
18. D. S. Berkholz, M. V. Shapovalov, R. L. Dunbrack Jr., P. A. Karplus, Conformation dependence of backbone geometry in proteins. *Structure* **17**, 1316–1325 (2009).
19. P. A. Karplus, Experimentally observed conformation-dependent geometry and hidden strain in proteins. *Protein Sci.* **5**, 1406–1420 (1996).
20. L. Esposito, N. Balasco, A. De Simone, R. Berisio, L. Vitagliano, Interplay between peptide bond geometrical parameters in nonglobular structural contexts. *BioMed. Res. Int.* **2013**, e326914 (2013).
21. R. Improta, L. Vitagliano, L. Esposito, Peptide bond distortions from planarity: New insights from quantum mechanical calculations and peptide/protein crystal structures. *PLOS One* **6**, e24533 (2011).
22. D. S. Berkholz, C. M. Driggers, M. V. Shapovalov, R. L. Dunbrack Jr., P. A. Karplus, Nonplanar peptide bonds in proteins are common and conserved but not biased toward active sites. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 449–453 (2012).
23. D. S. Cerutti, P. L. Freddolino, R. E. Duke Jr., D. A. Case, Simulations of a protein crystal with a high resolution x-ray structure: Evaluation of force fields and water models. *J. Phys. Chem. B* **114**, 12811–12824 (2010).
24. E. Muybridge, The science of the horse's motions. *Sci. Am.* **39**, 241 (1878).
25. N. W. Moriarty, D. E. Tronrud, P. D. Adams, P. A. Karplus, Conformation-dependent backbone geometry restraints set a new standard for protein crystallographic refinement. *FEBS J.* **281**, 4061–4071 (2014).
26. A. Stein, T. Kortemme, Improvements to robotics-inspired conformational sampling in Rosetta. *PLOS One* **8**, e63090 (2013).
27. P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding, D. Baker, Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* **23**, 47–55 (2014).
28. N. Koga, N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, D. Baker, Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
29. P.-S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
30. D. S. Berkholz, P. B. Krenesky, J. R. Davidson, P. A. Karplus, Protein Geometry Database: A flexible engine to explore backbone conformations and their relationships to covalent geometry. *Nucleic Acids Res.* **38**, D320–D325 (2009).
31. M. Z. Tien, D. K. Sydykova, A. G. Meyer, C. O. Wilke, PeptideBuilder: A simple Python library to generate model peptides. *PeerJ* **1**, e80 (2013).
32. B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. D. Caves, Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
33. D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, *AMBER 12* (2012); <http://ambermd.org/>.
34. N. L. Ogihara, G. Ghirlands, J. W. Brysons, M. Gingery, W. F. DeGrado, D. Eisenberg, Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1404–1409 (2001).
35. J. Herrou, S. Crosson, *myo*-Inositol and *D*-ribose ligand discrimination in an ABC periplasmic binding protein. *J. Bacteriol.* **195**, 2379–2388 (2013).

**Acknowledgments:** We thank O. Guvench for providing the data for the energy contours in Fig. 1D. **Funding:** This work was supported by NIH grant R01-GM083136 (to P.A.K.). **Author contributions:** The project was conceived by P.A.K.; experiments, analyses, and figure preparation were carried out by A.E.B.; and writing was done by A.E.B. and P.A.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the PDB, with specific details reported in table S1.

Submitted 28 August 2015  
Accepted 14 September 2015  
Published 16 October 2015  
10.1126/sciadv.1501188

**Citation:** A. E. Brereton, P. A. Karplus, Native proteins trap high-energy transit conformations. *Sci. Adv.* **1**, e1501188 (2015).