Biochemistry

© Copyright 2003 by the American Chemical Society Volume 42, Number 32 August 19, 2003

Current Topics

Native State Proline Isomerization: An Intrinsic Molecular Switch†

Amy H. Andreotti*

Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011

Received June 20, 2003

Exquisite control of biological function is achieved via tight regulation of the catalytic and binding activities of cellular proteins. The mechanistic details of protein regulation vary from targeted chemical modification of amino acid side chains (*1*) to the quite drastic global unfolding of an entire polypeptide chain (*2*). Peptidyl prolyl cis/trans isomerization is emerging as a potentially general mechanism for the control of protein function (*3*). While most structures of native, folded proteins reveal peptidyl−prolyl imide bonds that adopt *either* the cis or trans conformation, there are a growing number of folded proteins that exhibit conformational heterogeneity about one or more peptidyl−prolyl bonds. Unlike covalent modification or global unfolding, proline isomerization is an intrinsic conformational exchange process that has the potential to direct ligand recognition and to control protein activity within the confines of the native state.

Proline is the only naturally occurring amino acid for which both the cis and trans peptide bond conformations are thermodynamically feasible (*4*, *5*) lending the proline-containing polypeptide backbone the remarkable ability to populate discrete conformational and dynamic states separated by a modest interconversion activation energy. Available data for short proline containing peptides indicate that prolyl isomerization proceeds via a one-step mechanism involving a planar transition state (*6*, *7*) (Figure 1A). The barrier to interconversion between the cis and trans imide bond conformers is 14−24 kcal/mol, and the slow rate of exchange between conformers arising from this energy barrier (*8*) manifests itself as a rate-limiting step in protein folding (*9*−*11*). A subject that has received significantly less attention in the literature to date is the occurrence of peptidyl−prolyl cis/trans isomerization in the context of the native state of a folded protein. This review will discuss the role of proline isomerization-induced conformational heterogeneity and its significance in controlling the binding and function of globular proteins. The role of proline in signaling via transmembrane α helices has recently been reviewed elsewhere (*12*).

Experimental Detection of Proline Isomerization. Proline cis/trans isomerization likely occurs in many folded proteins, and it is possible that proline-mediated conformational exchange events control important functions of numerous well-studied systems, but have thus far eluded detection. Direct identification of proline switches is challenging, as the isomerization itself is invisible to most biochemical methodologies. Nuclear magnetic resonance (NMR) spectroscopy is the best available method for direct detection of cis/trans prolyl isomerization in a folded protein (*13*). With respect to the NMR chemical shift time scale, peptidyl−prolyl cis/trans isomerization is slow, and therefore two resonance frequencies are observed in NMR spectra per nuclear spin in the proximity of the isomerizing bond (Figure 1B). Characteristic NOE patterns readily discriminate the resonances arising from the cis and trans conformers allowing specific assignment of each substate (Figure 1C). The relative populations of the cis and trans forms can be determined by integration of the separate peak volumes and under the appropriate conditions the rate of exchange between conformers can be measured (*14*). In addition to NMR spectroscopy, direct detection of peptidyl−prolyl cis/trans isomerization within short peptides can be achieved by monitoring changes in the ultraviolet (UV) absorbance following a rapid

† This work was supported by grants from the Roy J. Carver Charitable Trust and the National Institutes of Health (AI 43957).
* E-mail: amyand@iastate.edu. Phone: (515) 294-4953.

10.1021/bi0350710 CCC: $25.00 © 2003 American Chemical Society
Published on Web 07/24/2003

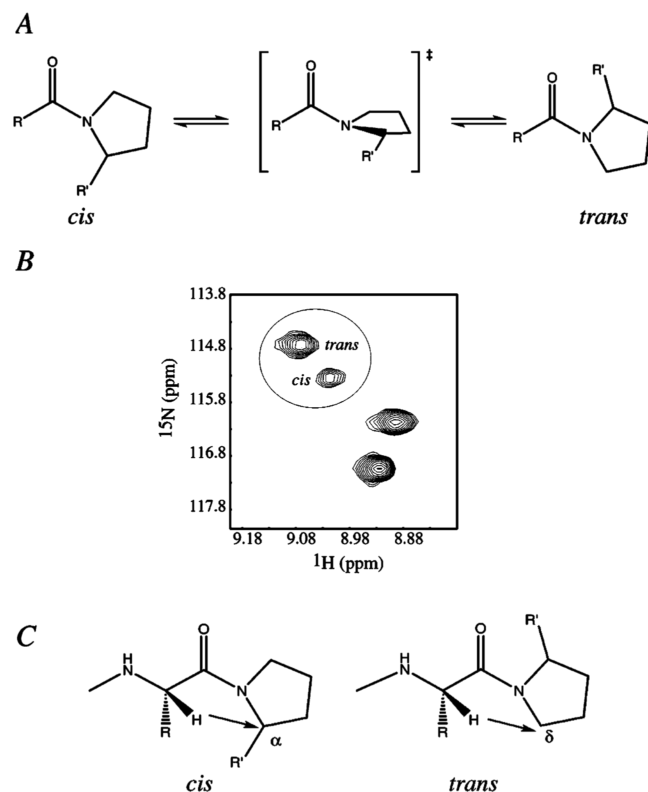

FIGURE 1: (A) Cis/trans isomerization of an Xaa-Pro imide bond (where Xaa is any amino acid) progresses through a planar transition state. (B) Selected region of a ^{1}H-^{15}N heteronuclear single quantum coherence spectrum of a folded protein that contains a conformationally heterogeneous proline. The circled resonances correspond to a single backbone amide NH group and are doubled due to the slow rate of cis/trans isomerization. Integration of cross-peak volumes indicates that the trans/cis ratio is 60/40 for this protein. Residues that give rise to only a single resonance (not circled) are unaffected by the isomerization. (C) Short ^{1}H-^{1}H distances (NOEs) between the CαH proton of the residue preceding proline and the Pro CαH proton are diagnostic of the cis conformation, while the NOE between the CαH proton of the residue preceding proline and the Pro CδH proton are diagnostic of the trans conformation.

change in pH (*15*). The range of pH change required to alter the cis/trans ratio is large and therefore impractical for application of this approach to native folded proteins. As well, reversed-phase liquid chromatography has been used as a diagnostic for cis/trans isomerization within dipeptides (*16*, *17*). The applicability of this technique to native proteins that harbor prolyl isomerization has not been demonstrated.

While NMR is the most direct method for detection of prolyl isomerization in native proteins and peptides, protein structures solved by X-ray crystallography can also reveal structural heterogeneity (*18*). Often, a cis conformation is apparent under certain crystallization conditions, while the same imide bond may adopt the trans conformation when the protein is crystallized under different conditions. The observation of two prolyl imide bond conformations in a crystal structure is therefore somewhat fortuitous. There are likely many protein structures that are biased by the crystallization process favoring a single imide bond conformer, thereby obscuring potentially important conformational heterogeneity. Moreover, the frequent occurrence of proline residues in flexible loops means that, for the region of interest, many crystal structures may not be of sufficient resolution to provide evidence for proline isomerization.

Proline Isomerization Is Catalyzed by Several Conserved Families of Isomerases. The slow interconversion between cis and trans prolyl imide bond conformers is catalyzed by the peptidyl−prolyl cis/trans isomerases (PPIases). The PPIases (namely, the cyclophilins (Cyp), FK506 binding proteins (FKBPs) (*19*−*21*), and the parvulins (*22*)) bind and stabilize a transition state that is characterized by partial rotation around the C−N imide bond (*23*). The energetic cost of distortion is compensated by favorable interactions between enzyme and transition state. The ubiquitous nature of these enzymes and the existence of three distinct families of PPIases underscores the importance of proline isomerization in protein structure and function. Cyclophilin A and FKBP-12 have attracted significant attention as the protein targets of the immunosuppressive agents cyclosporine A (CsA) and FK-506, respectively (*24*). These small molecules inhibit PPIase activity and block T cell activation. Unexpectedly, the immunosuppressant activity of these small molecules derives not from inhibition of isomerase activity, but rather from interruption of signaling events by the inhibitor−PPIase complexes themselves (*25*−*27*). Thus, while CsA and FK506 have shed considerable light on the signal transduction pathways involved in T cell signaling, the normal cellular roles of the PPIases remain unclear.

Functional Roles of the PPIases. The PPIases accelerate protein folding in vitro by catalyzing the rate-limiting peptidyl−prolyl isomerization step, implying a role for these enzymes as folding catalysts (*28*−*31*). There is also accumulating evidence that the PPIases modulate intracellular signaling events (*32*, *33*). The most recently described family of PPIases, the parvulins (*22*), are in fact the only members of the PPIase superfamily that are essential for cell growth. The prototypical parvulin, Pin1, interacts with target proteins in a phosphorylation-dependent manner (*34*, *35*) and is required for normal progression through mitosis (*22*). Despite the picture that is forming for the biological importance of Pin1 isomerase activity (*36*), an understanding of the physiological role of the cyclophilins and FKBPs and the specific nature of their respective protein targets is still lacking.

In contrast to studies involving short peptide model substrates, the catalytic action of PPIases on *folded* protein substrates has not yet been extensively studied at the molecular level. An exception is the amino-terminal core domain of the human immunodeficiency virus-type 1 (HIV-1) capsid protein. The interaction between HIV-1 capsid and cyclophilin A is essential for the subsequent packaging of multiple copies of CypA into each HIV-1 virion (*37*−*40*). While the precise function of CypA in HIV virulence is not yet known and the isomerase activity of CypA per se may not be required for HIV replication (*41*), virions lacking CypA (either by interruption of the interaction with cyclosporin or by mutations in the amino-terminal core domain) are less infectious (*39*, *42*).

The molecular details of the HIV-1 capsid/cyclophilin A interaction emerged in 1996 when the three-dimensional structure of the HIV-1 capsid protein (Figure 2A) was solved by multidimensional heteronuclear magnetic resonance spectroscopy (*43*). Shortly thereafter, the crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid (Figure 2B) was reported (*44*). Together, these structures provide valuable insight into the nature of a

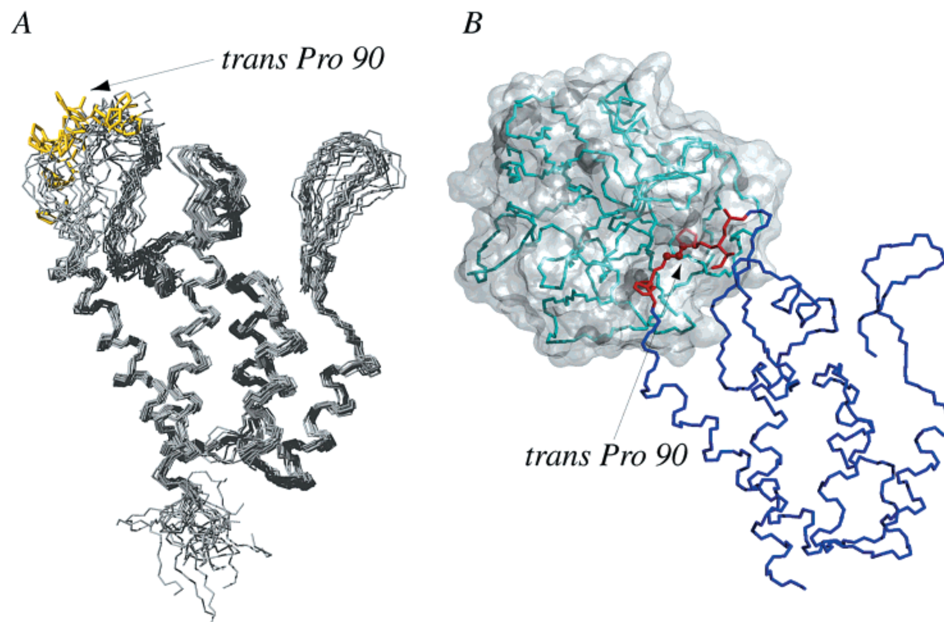

FIGURE 2: (A) Overlay of 17 low energy structures of the trans conformation of the HIV-1 capsid protein as determined by NMR spectroscopy (pdb: 1GDS). Pro 90 (yellow) is conformationally heterogeneous within the context of the folded HIV-1 capsid protein and is located at the apex of a disordered loop. (B) The first crystal structure of the HIV-1 capsid protein bound to the active site of cyclophilin A (pdb: 1AK4). Cyclophilin A is shown as a transparent surface over the backbone trace (light blue), while only the backbone trace of the HIV-1 capsid protein (dark blue) is represented. In this structure, the imide bond preceding Pro 90 (ball-and-stick representation) exclusively adopts the trans conformation upon binding to cyclophilin A. The residues adjacent to Pro 90 and Pro 90 itself within HIV-1 capsid are shown in red. All structures were created in MolMol (*87*).

physiological cyclophilin A substrate as well as the mode of recognition between cyclophilin A and the folded capsid protein. Two features are especially noteworthy. First, the NMR structure of the capsid protein alone reveals that the exposed loop that serves as the binding site of cyclophilin A includes a single conformationally heterogeneous proline residue (Pro 90, Figure 2A). The less populated conformer (14%) corresponds to the protein containing the cis prolyl imide bond at position 90, while the major species in solution (86%) is the trans imide bond-containing conformer. The low population of the cis imide bond-containing conformer prohibited detailed structural analysis of that substate, and the reported NMR structure consequently corresponds to the trans form of the domain. Second, the structure of the cyclophilin A−capsid complex reveals that Pro 90 adopts exclusively the trans conformation upon binding to cyclophilin A (Figure 2B) in contrast to the previously solved structures of cyclophilin A bound to short peptides where the prolyl imide bond adopts the cis conformation (*45*−*48*).

Very recently, a series of crystal structures of CypA complexed to several sequence variants of the HIV-1 capsid protein were solved and reveal that the cis conformation can also be accommodated in the CypA active site (*49*). In conjunction with NMR dynamics data (*50*), these structures are providing critical insight into the mechanism of prolyl isomerization by cyclophilin A. However, it is unclear whether the primary role of cyclophilin A is an enzyme in vivo or instead as a protein binding module that mediates specific protein−protein interactions (*51*). While the static structures of the HIV-1 capsid−cyclophilin A complex do not resolve this issue, more recent work by Kern and co-workers using NMR exchange spectroscopy indicates that cyclophilin A has the capacity to accelerate isomerization of the Gly 89−Pro 90 imide bond within the HIV-1 capsid protein substrate in vitro (*52*). Nevertheless, the extent to which catalysis is important for the biological functions of the cyclophilins remains to be determined. As a step toward this goal, Kern and co-workers describe mutations within cyclophilin A that retain binding but abolish catalysis of the Gly 89−Pro 90 imide bond of HIV-1 capsid (*52*). These and other such mutants will be invaluable in future experiments aimed at dissecting the respective contribution of binding and catalysis to cyclophilin A function in vivo.

Proline Isomerization Leads to a Range of Structural Consequences in Folded Proteins. Notably, proline isomerization has a minimal effect on the overall structure of the amino-terminal domain of HIV-1 capsid. The NMR data for the free HIV-1 capsid protein reveal doubled resonances for only 4 of the 151 residues that comprise the domain: Ala 88, Gly89, Pro 90, and Ile 91 (*43*). This shows that the structural perturbations associated with proline isomerization are very localized. There are in fact numerous examples of local conformational heterogeneity within folded proteins induced by proline isomerization (*53*−*64*). One of the first such proteins to be extensively characterized is staphylococcal nuclease. NMR studies show that two interconverting folded forms of staphylococcal nuclease coexist in solution (*65*, *66*) and that isomerization about the prolyl imide bond at position 117 is the source of the observed conformational heterogeneity (*67*). A more recent example of the coexistence of alternative prolyl conformers in similar folded structures is reported for Tendamistat (*68*). For this protein, kinetic characterization of refolding reveals a nativelike state that interconverts to native Tendamistat via proline isomerization. In this case, cyclophilin accelerates the rate of proline isomerization and therefore favorably affects the interconversion rate between the nativelike and native states of Tendamistat. These results provide indirect evidence that Tendamistat serves as a folded protein substrate of cyclophilin. For the majority of conformationally heterogeneous

native proteins, evidence for interactions with the PPIases is currently lacking (*69*). Thus, until additional folded protein substrates for the PPIases are identified and characterized it is not yet clear whether the presence of a conformationally heterogeneous proline in a folded protein is a recognition motif for cyclophilin and/or related enzymes. In fact, it is likely that proline isomerization itself plays an important functional role that is independent of the action of the peptidyl−prolyl isomerases.

Several specific proteins will now be considered to explore the notion that a drastic structural rearrangement induced by peptidyl−prolyl isomerization can serve as an intrinsic molecular switch that controls protein function. In each case, three-dimensional structures of both imide bond-containing conformers are available, the structural consequences of the isomerization are pronounced, and in some cases the cis and trans conformers have been ascribed distinct functional roles. Undoubtedly, as additional protein structures are solved and as new methods are developed to detect native state proline isomerization, the list of proteins known to exist in multiple native-state conformations that differ with respect to the isomerization state around one or more prolyl imide bonds will grow. Ultimately, characterization of such conformationally heterogeneous proteins will provide the information that is needed to understand both the molecular determinants that favor dual imide bond conformers within a folded protein, as well as the role of proline isomerization as a molecular switch that has the potential to mediate conformer-specific ligand binding and/or function.

Native-State Proline Isomerization Causes Extensive Structural Rearrangements: Itk SH2 Domain. The nonreceptor tyrosine kinase, interleukin-2 tyrosine kinase (Itk), participates in the intracellular signaling events leading to T cell activation (*70*, *71*). In an effort to elucidate the mechanism by which the catalytic activity of Itk and related tyrosine kinases is controlled during signaling, detailed structural analyses of the noncatalytic regulatory domains have been carried out (*72*−*74*). Of particular importance in the context of this review is the structure determination of the Src homology 2 (SH2) domain of Itk (*75*). The Itk SH2 domain exhibits a hallmark indicator of multiple conformations in slow exchange: the appearance of "extra" resonances in the NMR spectra. Cis/trans isomerization about the imide bond between Asn 286 and Pro 287 within the SH2 domain is the source of this conformational heterogeneity as mutation of Pro 287 to Gly leads to greatly simplified spectra consistent with a single species in solution (*33*).

Itk is also a physiological substrate of cyclophilin A as the interaction between these two proteins has been detected in T cells and cyclophilin A activity regulates Itk catalytic function both in vitro and in vivo (*33*). However, the detailed structural insights that have been reported for the HIV-1 capsid−cyclophilin A complex (*52*) are not yet available for the Itk−cyclophilin A complex, and so the molecular details of this interaction are not known. Nevertheless, proline isomerization within the Itk SH2 domain has pronounced structural and functional consequences that are independent of the association with cyclophilin A. Interestingly, the conformational heterogeneity arising from proline isomerization within the Itk SH2 domain differs from the conformationally heterogeneous proteins cited above. First, proline isomerization causes *long-range* structural perturbations within the Itk SH2 domain rather than the limited structural heterogeneity observed, for example, in structural studies of the HIV-1 capsid protein. Second, the population of the cis and trans imide bond-containing conformers is nearly equal (40% cis, 60% trans) as determined by integration of NMR peak volumes permitting detailed structural investigations of *both* species in solution.

NMR derived structural models for the cis and trans imide bond-containing SH2 conformers are shown in Figure 3A. The overall fold of the domain matches those of previously solved SH2 domains. The conformationally heterogeneous Pro 287 is located at the carboxy-terminus of the CD loop. In the trans form of the domain, the CD loop extends away from the body of the domain. Upon isomerization to the cis form, the CD loop bends down at two hinge points located at Pro 287 itself and at Lys 280 located directly across the loop structure. Moreover, proline isomerization from trans to cis affects the dynamics of the loop. Backbone dynamics measurements (^{15}N T_1, T_2, and heteronuclear ($\{^1H\}$-^{15}N) NOE) indicate significant mobility in the trans form of the CD loop (*75*) as expected for a large surface loop. In contrast to the trans form, backbone dynamics measurements indicate significantly less mobility for the CD loop residues in the cis form (*75*). Consistent with the pronounced change in the conformational preference and dynamics of the CD loop, the resonance frequency differences between the cis and trans forms are largest for the residues in the CD loop. Nondegenerate cis and trans chemical shifts are, however, not confined to the proline-containing loop and in fact extend across one-third of the domain surface (Figure 3B). In addition to the large displacement of the CD loop, statistically significant structural differences are apparent between the refined structures of the cis and trans forms of the Itk SH2 domain in the AB and BG loops, the top of the central β sheet and the top turn of β helix (Figure 3C) (*75*).

The structural and dynamic differences that arise from isomerization about the Asn 286−Pro 287 imide bond confer conformer-specific properties to this domain. The Itk SH2 domain binds to two distinct ligands. In the canonical fashion, the domain binds a phosphotyrosine containing peptide (Figure 4A). Phosphopeptide binding occurs preferentially with the trans form of the domain (*33*). The Itk SH2 domain also forms a novel and specific intermolecular complex with the Src homology 3 (SH3) domain of Itk (*73*). This interaction favors the cis form of the SH2 domain (Figure 4B) (*33*). Thus, proline cis/trans isomerization and the concomitant structural rearrangements within the SH2 domain serve as a molecular switch governing ligand recognition. Elucidation of the mechanism by which this occurs must await further characterization of the structures and dynamics of the respective complexes. Nevertheless, this example makes clear that proline isomerization is sufficient to alter the ligand-binding surface of a protein to an extent that can be exploited by incoming ligands.

MS2 Capsid. A comparison of the cis and trans Itk structures with other proteins in which a proline residue adopts both the cis and trans imide bond conformations suggests that the conformer-specific structural rearrangements that have been characterized for Itk may be a general phenomenon in controlling protein−protein interactions. A striking example is found in the subunits of the bacteriophage MS2 coat protein (*76*). The icosahedral virus bacteriophage

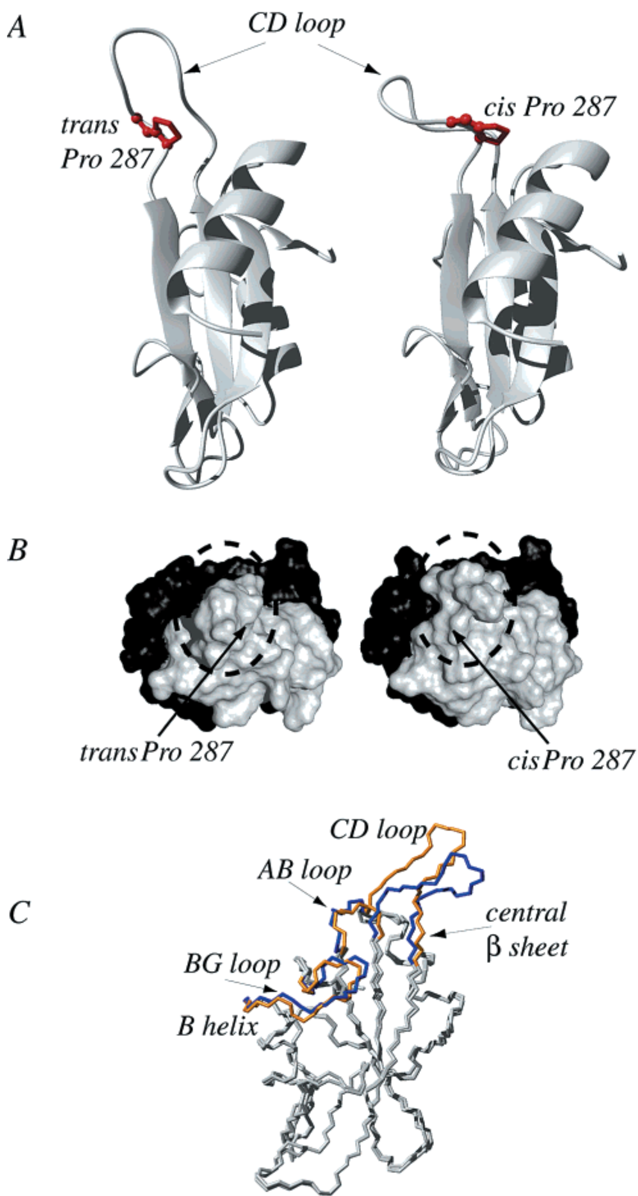

FIGURE 3: (A) Ribbon diagrams of the energy minimized average structures of the trans (left) and cis (right) Itk SH2 conformers (pdb: 1LUN, 1LUK). Pro 287 in the CD loop and the preceding imide bond is highlighted in red in each case. (B) Solvent-accessible surface plot of the trans (left) and cis (right) SH2 domain. For each structure, the residues that give rise to nondegenerate chemical shifts due to proline isomerization are shown in light gray and black shading represents residues that are unaffected by proline isomerization. The position of Pro 287 is indicated and the CD loop is circled on each structure. The conformational change in the CD loop that accompanies proline isomerization causes changes in the solvent accessible surface area. (C) Superposition of the minimized average cis and trans structures. (Backbone superposition was carried for all residues excluding the CD loop (residues 278−290)). Regions of conformational heterogeneity (CD, BG, and AB loops and the tops of the central β-sheet and B helix) are highlighted in orange (trans) and blue (cis). Regions of the protein which are unaffected by proline isomerization are gray. The view of the domain in panel C is rotated approximately 90° from that shown in B.

MS2 is comprised of 180 copies of a coat protein surrounding a single-stranded RNA. The structure of bacteriophage MS2 has been solved at 2.8 Å resolution (Figure 5) (*77*). The fold of the MS2 coat protein differs from that of other viral coat proteins and contains a five-stranded β-sheet, two nearly perpendicular shorter strands and a bent helix. Furthermore, MS2 lacks the terminal polypeptide segment that in other viruses mediates assembly of the viral coat by switching between ordered and disordered forms (*78*). Instead, the MS2 coat protein contains a conformationally heterogeneous loop (the FG loop) that appears to fulfill a similarly critical role in regulating assembly (*77*, *79*). Interestingly, cis/trans isomerization around the imide bond preceding Pro 78 is responsible for conferring two distinct conformations to this region of the protein (*77*).

Three structurally independent coat protein subunits (A−C) are shown in Figure 5A. Each monomeric subunit is a chemically equivalent 13.7 kDa protein. An overlay of the three subunits reveals large conformational differences between monomer B and monomers A and C in the region of the proline-containing FG loop (Figure 5B) (*77*). The cis imide bond in the B subunit leads to a bend in the FG loop toward the short β-strands of the protein (Figure 5C). A number of hydrogen bonds stabilize this conformation (*77*), and it is interesting to note that in a manner similar to Itk, a kink at the cross strand residue (Thr 69) accompanies isomerization to the cis proline. Indeed, the structural differences between the cis and trans prolyl bond-containing proteins are strikingly similar for MS2 and Itk. Moreover, as observed for Itk, dynamic differences between the cis and trans MS2 conformers are apparent in a comparison of the temperature factors for the residues of the FG loop. As observed for Itk, it is the cis imide bond containing structure that is less flexible than the corresponding trans form (*77*).

Candida rugosa Lipase. In a manner similar to Itk and MS2, a functionally relevant loop rearrangement driven by proline isomerization occurs in *Candida rugosa* lipase. Crystal structures of the "open" and "closed" forms of this enzyme have been solved to 2.06 and 2.1 Å resolution, respectively (*80*, *81*). The two crystal forms differ with respect to the conformation of a single loop or "flap" region that in the "open" form (*81*) protrudes away from the surface of the enzyme allowing access to the active site and in the "closed" form (*80*) makes extensive hydrophobic contacts to the rest of the protein and shields the active site from solvent (Figure 6A). The rearrangement between "open" and "closed" forms involves trans to cis prolyl isomerization (Pro 92) and a concomitant bending at a hinge point across the loop (Glu 66). The *C. rugosa* lipase loop (26 residues) is significantly longer than that of Itk and MS2 and adopts distinct secondary structure in both forms. The energetically more stable trans prolyl imide bond is adopted in the "open" form, while the "closed" form contains the cis conformation (Figure 6B,C).

Is there a Motif Common to Proline Switches? It is interesting to note that in each of the examples of native state proline isomerization (Itk, MS2, and lipase), the trans form corresponds to the more extended, solvent-exposed loop conformation, while the cis form bends down. In each case, more extensive contacts are present between the loops in the cis prolyl imide bond conformation and the body of the protein than in the corresponding trans conformation. Whether hydrophobic or hydrophilic in nature, such contacts likely provide the necessary stabilization energy to favor the inherently less stable cis conformer. In fact, studies that have directly compared the cis/trans ratio of a proline residue in the context of the folded protein with that of a corresponding

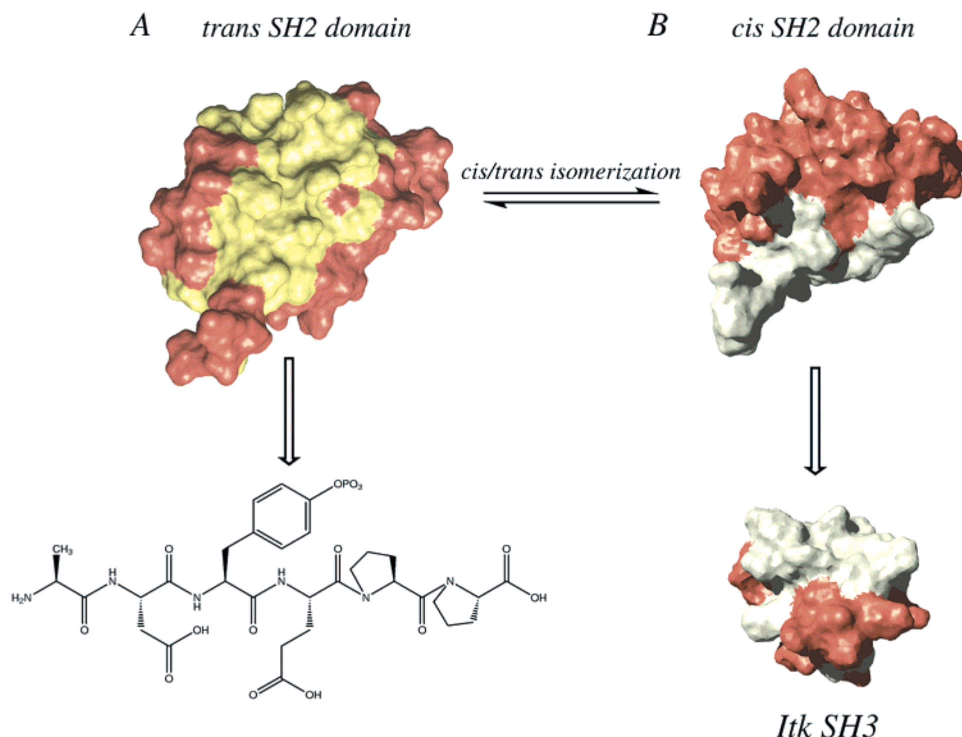

FIGURE 4: (A) Solvent accessible surface of the Itk SH2 domain containing the trans imide bond at position 287. The residues for which significant resonance frequency changes (>0.1 ppm ^{1}H, >0.5 ppm ^{15}N) are observed upon titration of the phosphopeptide (ADpYEPP) are highlighted in yellow. Those residues that do not exhibit changes in resonance frequency upon peptide binding are orange. (B) The alternative (cis) form of the SH2 domain binds in an intermolecular fashion to the conserved binding pocket of the Itk SH3 domain. For each protein (cis SH2 and SH3), residues for which chemical shift perturbations are observed upon complex formation are shown in light yellow and residues unaffected by the interaction are orange.

short peptide designed to approximate the unfolded state have revealed significant differences (*82*). Specifically, the linear peptide derived from the proline-containing loop in staphylococcal nuclease is predominantly trans in solution, whereas the same sequence adopts the cis form in the folded protein (*83*). It is argued that noncovalent intramolecular forces between the loop and the protein alter the cis/trans ratio and serve to stabilize the cis form in that case (*82*). The examples highlighted here all appear to conform to the same general mechanism to achieve stabilization of dual imide bond conformers.

There are particularly notable similarities between Itk and MS2 at both the primary and secondary structural levels. First, the proline residue in each case is located at the carboxy-terminus of a loop that extends from two strands of an antiparallel β-sheet. The position of proline at the base of a loop and the physiochemical properties of the cross-strand residue may be important components of the hinge movement that results from isomerization. The Itk and MS2 loops themselves are close in size and share some sequence similarity (Figure 7). In particular, a cluster of hydrophobic residues occupy the central region of each loop (Val-Gly-Gly-Val in MS2 and Ala-Ile-Ile in Itk). Itk and MS2 also share a glutamate residue in the position directly preceding the conformationally heterogeneous proline. Additionally, the large conformational changes that occur upon proline isomerization in both Itk (CD loop) and MS2 (FG loop) cause smaller structural changes in the neighboring loops (AB and BG loops of Itk SH2 (Figure 3C) and the AB and DE loops of the MS2 protein (Figure 5C)). Of particular note is the presence of multiple glycine residues in one of these proximal loops in both proteins. For Itk, the BG loop contains three consecutive glycines (Gly 326, Gly 327, and Gly 328) and the AB loop of MS2 contains three nearly consecutive glycine residues (Gly 13, Gly 14, and Gly 16). The flexibility inherent to glycine containing sequences in close proximity to a conformationally heterogeneous proline may be required to accommodate the conformational change and maintain the overall structure of the protein. The similarities between Itk SH2 and the MS2 coat protein are striking and suggest there may be characteristic structural and sequence motifs that could serve as a signature for proline switches in native proteins.

In support of this hypothesis, the zinc finger protein, Zpr1, has recently been identified in a screen for cyclophilin A dependent yeast strains (*84*). Amino acid mutations in Zpr1 were identified that confer cyclophilin A dependence on yeast cells lacking cyclophilin A. Interestingly, two of the three mutant alleles are identical and encode an amino change directly preceding a conserved proline residue near the carboxy-terminus of Zpr1 (*84*). The primary sequence in this region of Zpr1 contains a similar pattern of amino acids as already highlighted for Itk and MS2 (Figure 7). Namely, Zpr1, like Itk and MS2, contains a cluster of hydrophobic residues and a glutamate adjacent to the proline. High-resolution structural data for Zpr1 is not yet available, and so it is not possible to determine whether this region of Zpr1 may serve as a "proline switch" motif in the context of the native protein. Nevertheless, the sequence similarities with already characterized proline-containing sequences along with the genetic evidence for a functional role for cyclophilin A in regulating Zpr1 provide tantalizing hints that this region of Zpr1 is functionally important and may contain a conformationally heterogeneous proline residue.

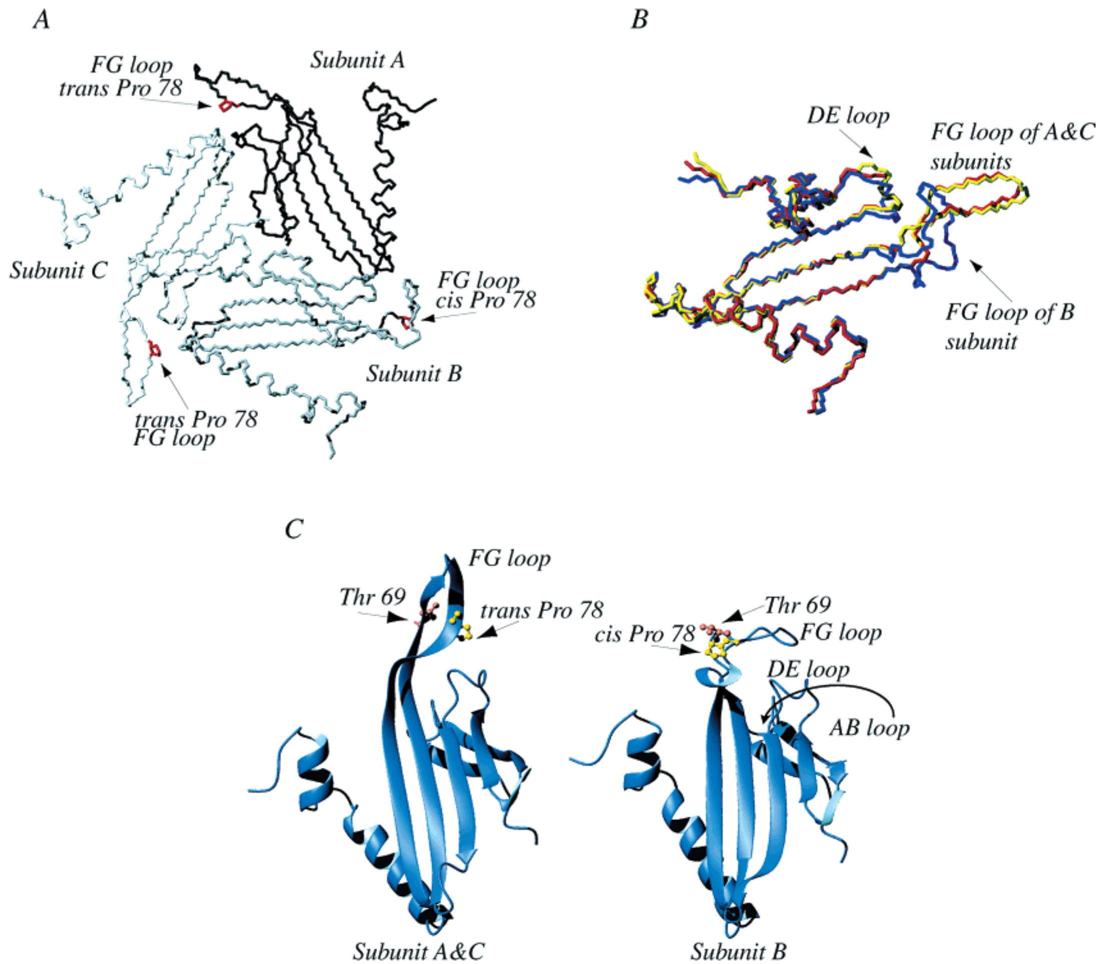

FIGURE 5: (A) The three monomeric subunits of the bacteriophage MS2 viral coat (pdb: 1ZDI) are chemically identical but differ with respect to the imide bond conformation at position 78. (Pro 78 is highlighted in red for each monomer.) (B) The conformational differences in the FG loop of subunit B (blue) are apparent in the superposition of the three subunits. (Backbone superposition was carried out for the all residues excluding those of the FG loop.) Smaller structural differences are also apparent in this view for the DE loop. (C) Ribbon representations of the subunits A and C (left) and subunit B (right). The FG loop containing the cis proline is bent ~90° compared to the same loop in the trans proline containing subunits A and C. Proline 78 and the preceding imide bond are shown in yellow. Thr 69 is labeled and indicated in orange.

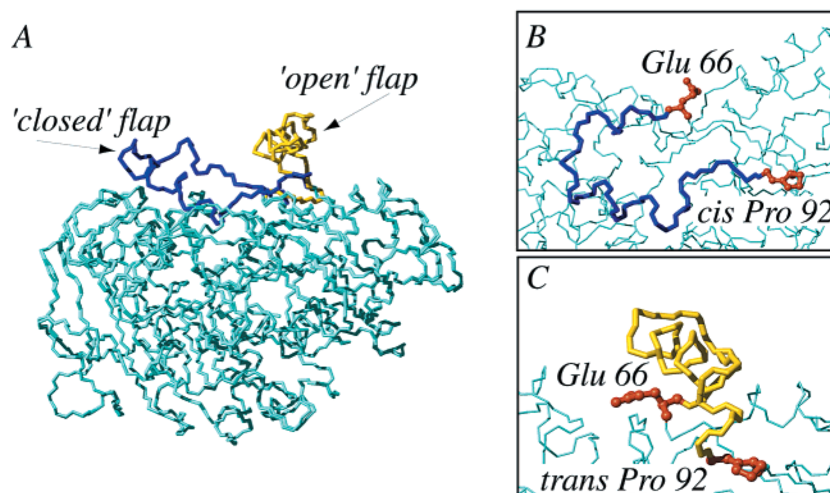

FIGURE 6: (A) Superposition of the "closed" (blue, pdb: 1TRH) and "open" (yellow, pdb: 1CRL) forms of *C. rugosa* lipase. The two forms of the protein are identical outside of the conformationally heterogeneous loop region. (Regions of structural identity are shown in cyan.) Backbone superposition was carried out for all residues excluding those of the loop (residues 66−92). (B) View of the "closed" loop showing, in red, the cis prolyl residue at position 92 (red) and the corresponding hinge point at glutamate 66 (red). (C) Similar view of the "open" form of the loop where Pro 92 (red) adopts the trans conformation. Glu 66 at the base of the loop is also shown in red.

Native State Proline Isomerization Is Not Limited to Loop Regions. TB Domain. One might expect that a conformationally heterogeneous proline would be located in a region of relatively high flexibility in the context of a native fold

$Itk_{278-287}$ ...FTKAIISENP...
$MS2_{70-78}$ ...QTVGGV-ELP...
$Zpr1_{437-445}$ ...KFTVIM-EDP...

FIGURE 7: Limited sequence comparison of the "proline-switch" region of Itk with that of MS2 and a putative "proline-switch" of Zpr1. The proline residue and a conserved glutamate (E) are highlighted in red. A hydrophobic stretch (boxed) is present in each sequence.

as opposed to the more stable regions of secondary structure (*56*). While the three examples already discussed follow this general rule, proline isomerization has been observed at least in one case within the confines of stable secondary structure. Structural analysis of the transforming growth factor β binding protein-like (TB) domain from human fibrillin-1 reveals two stable conformers that differ with respect to the isomerization state of two proline residues (Pro 22 and Pro 52) (*85*). Two structures of this protein were solved by NMR spectroscopy corresponding to the major and minor forms of the domain (Figure 8). In contrast to the examples discussed above, the conformationally heterogeneous Pro 22 of the TB domain is embedded in a region of stable secondary structure rather than a flexible loop region (Figure 8). In conjunction with the three-dimensional structures, dynamics data provide further evidence that the polypeptide region surrounding Pro 22 is well ordered for both conformers. Despite the particular location of Pro 22, interconversion between imide bond conformers occurs and causes measurable changes in the secondary structure topology of the domain in the β-strands surrounding Pro 22 (Figure 8) (*85*). While this example provides an alternative view into the structural consequences of native state proline isomerization, the functional significance of the major and minor forms of the TB domain have yet to be elucidated. Certainly, the extent of the observed structural perturbations that arise from proline isomerization are on the order of those observed for the Itk SH2 domain, and it is therefore conceivable that ligand recognition by the TB domain could be directed by proline isomerization in a similar manner.

The examples of native state proline isomerization discussed above point to the potentially general role of proline as a molecular switch that can direct protein interactions by altering a receptor binding surface. The unique properties of proline, the energy barrier for interconversion between the cis and trans forms and the potential for conformational and dynamical differences between the two stable conformations, make this intrinsic switch ideal for conferring two distinct, low energy structures to the overall native fold of a protein. Continued structural analyses will illuminate the mechanisms by which structural perturbations arising from isomerization about a single imide bond are propagated throughout a significant portion of a molecule. Such knowledge should lead to an understanding of why some conformationally heterogeneous proline residues cause very minor local perturbations while others significantly alter the surface binding properties of a protein. The elusive nature of proline isomerization is a problem that will need to be addressed before dramatic increases in the number of proteins exhibiting native state isomerization will be realized. New experimental approaches that are less demanding than NMR spectroscopy and X-ray crystallography in terms of sample quantities and instrumentation must be developed to facilitate identification and analysis of native state proline isomerization. Along these lines, the development of efficient, high throughput screens for native state proline isomerization will be necessary to fully characterize the extent to which this molecular switch is at work within the ensemble of proteins that comprise the cell.

Of the naturally occurring amino acids, proline is unique and appears to fulfill multiple roles in the context of native, folded proteins. In some proteins, proline isomerization may confer conformer-specific properties to a native protein fold by modulating the features of a protein surface. Alternatively, a conformationally heterogeneous proline residue may cause minimal structural perturbations but could instead serve as a recognition site for a PPIase. Proline residues that exhibit dual functionality with respect to these functions are also likely to exist. In much the same way that mutation of tyrosine to phenylalanine disrupts important phosphorylation sites and abrogates signaling (*86*), it is possible that disease states will be identified in the future that arise from mutation of a functionally significant proline residue to an amino acid that cannot access the cis conformation at a reasonable energy

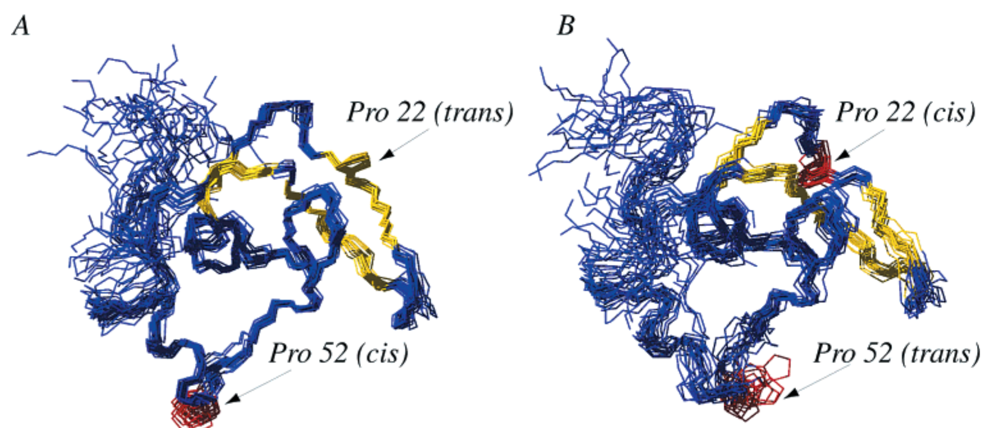

FIGURE 8: (A) Overlay of the 21 lowest energy solution structures of the major form of the TB domain (pdb: 1APJ). Pro 22 adopts the trans conformation and Pro 52 adopts the cis conformation in this structure. The region of stable secondary structure (β-sheet) containing Pro 22 is shown in yellow. Pro 52 (red) is located in a loop region outside of the defined secondary structure of the domain. (B) Same representation as in panel A for the minor form of the TB domain where the proline conformers are inverted (Pro 22 (red) is cis and Pro 52 is trans). Isomerization of Pro 22 to the cis form disrupts the regular secondary structure in this region (remaining secondary structure shown in yellow).

cost. Thus, future efforts should strive to ascertain the extent to which native-state proline isomerization is a general control mechanism and simultaneously advance our fundamental understanding of the molecular basis for this intrinsic conformational exchange process.

ACKNOWLEDGMENT

The author would like to thank Drs. D. Bruce Fulton and Mark Hargrove for careful reading of the manuscript and insightful comments. The coordinates for the minor form of the TB domain were kindly provided by K. Downing (University of Oxford).

REFERENCES

1. Buck, M., and Rosen, M. K. (2001) *Science 291*, 2329−30.
2. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol. 293*, 321−31.
3. Yaron, A., and Naider, F. (1993) *Crit. Rev. Biochem. Mol. Biol. 28*, 31−81.
4. Grathwohl, C., and Wuthrich, K. (1976) *Biopolymers 15*, 2043−57.
5. Cheng, H. N., and Bovey, F. A. (1977) *Biopolymers 16*, 1465−72.
6. Stein, R. L. (1993) *Adv. Protein Chem. 44*, 1−24.
7. Fischer, G. (2000) *Chem. Soc. Rev. 29*, 119−127.
8. Grathwohl, C., and Wüthrich, K. (1981) *Biopolymers 20*, 2623−2633.
9. Schmid, F. X., and Baldwin, R. L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4764−8.
10. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) *Biochemistry 14*, 4953−63.
11. Wedemeyer, W. J., Welker, E., and Scheraga, H. A. (2002) *Biochemistry 41*, 14637−44.
12. Sansom, M. S., and Weinstein, H. (2000) *Trends Pharmacol. Sci. 21*, 445−51.
13. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
14. Hsu, V. L., Handschumacher, R. E., and Armitage, I. M. (1990) *J. Am. Chem. Soc. 112*, 6745−6747.
15. Zhang, Y., Fussel, S., Reimer, U., Schutkowski, M., and Fischer, G. (2002) *Biochemistry 41*, 11868−77.
16. Bouabdallah, S., Trabelsi, H., Bouzouita, K., and Sabbah, S. (2002) *J. Biochem. Biophys. Methods 54*, 391−405.
17. Trabelsi, H., Bouabdallah, S., Sabbah, S., Raouafi, F., and Bouzouita, K. (2000) *J. Chromatogr. A 871*, 189−99.
18. Smith, J. L., Hendrickson, W. A., Honzatko, R. B., and Sheriff, S. (1986) *Biochemistry 25*, 5018−27.
19. Harding, M. W., Galat, A., Uehling, D. E., and Schreiber, S. L. (1989) *Nature 341*, 758−60.
20. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T., and Schmid, F. X. (1989) *Nature 337*, 476−8.
21. Takahashi, N., Hayano, T., and Suzuki, M. (1989) *Nature 337*, 473−5.
22. Lu, K. P., Hanes, S. D., and Hunter, T. (1996) *Nature 380*, 544−7.
23. Harrison, R. K., and Stein, R. L. (1990) *Biochemistry 29*, 1684−9.
24. Rosen, M. K., and Schreiber, S. L. (1992) *Angew. Chem. Int. Ed. Engl. 31*, 384−400.
25. O'Keefe, S. J., Tamura, J., Kincaid, R. L., Tocci, M. J., and O'Neill, E. A. (1992) *Nature 357*, 692−4.
26. Liu, J., Farmer, J. D., Jr., Lane, W. S., Friedman, J., Weissman, I., and Schreiber, S. L. (1991) *Cell 66*, 807−15.
27. Walsh, C. T., Zydowsky, L. D., and McKeon, F. D. (1992) *J. Biol. Chem. 267*, 13115−8.
28. Kern, G., Kern, D., Schmid, F. X., and Fischer, G. (1995) *J. Biol. Chem. 270*, 740−5.
29. Kern, D., Kern, G., Scherer, G., Fischer, G., and Drakenberg, T. (1995) *Biochemistry 34*, 13594−602.
30. Schmid, F. X., Mayr, L. M., Mucke, M., and Schonbrunner, E. R. (1993) *Adv. Protein Chem. 44*, 25−66.
31. Gothel, S. F., and Marahiel, M. A. (1999) *Cell Mol. Life Sci. 55*, 423−36.
32. Arevalo-Rodriguez, M., Cardenas, M. E., Wu, X., Hanes, S. D., and Heitman, J. (2000) *EMBO J. 19*, 3739−49.
33. Brazin, K. N., Mallis, R. J., Fulton, D. B., and Andreotti, A. H. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 1899−904.
34. Shen, M., Stukenberg, P. T., Kirschner, M. W., and Lu, K. P. (1998) *Genes Dev 12*, 706−20.
35. Yaffe, M. B., Schutkowski, M., Shen, M., Zhou, X. Z., Stukenberg, P. T., Rahfeld, J. U., Xu, J., Kuang, J., Kirschner, M. W., Fischer, G., Cantley, L. C., and Lu, K. P. (1997) *Science 278*, 1957−60.
36. Lu, K. P., Liou, Y. C., and Zhou, X. Z. (2002) *Trends Cell Biol. 12*, 164−72.
37. Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C. T., Sodroski, J., and Gottlinger, H. G. (1994) *Nature 372*, 363−5.
38. Luban, J., Bossolt, K. L., Franke, E. K., Kalpana, G. V., and Goff, S. P. (1993) *Cell 73*, 1067−78.
39. Franke, E. K., Yuan, H. E., and Luban, J. (1994) *Nature 372*, 359−62.
40. Braaten, D., Franke, E. K., and Luban, J. (1996) *J. Virol. 70*, 3551−60.
41. Saphire, A. C., Bobardt, M. D., and Gallay, P. A. (2002) *J. Virol. 76*, 2255−62.
42. Braaten, D., Franke, E. K., and Luban, J. (1996) *J. Virol. 70*, 4220−7.
43. Gitti, R. K., Lee, B. M., Walker, J., Summers, M. F., Yoo, S., and Sundquist, W. I. (1996) *Science 273*, 231−5.
44. Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I., and Hill, C. P. (1996) *Cell 87*, 1285−94.
45. Zhao, Y., and Ke, H. (1996) *Biochemistry 35*, 7356−61.
46. Zhao, Y., and Ke, H. (1996) *Biochemistry 35*, 7362−8.
47. Ke, H., Mayrose, D., and Cao, W. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3324−8.
48. Kallen, J., Spitzfaden, C., Zurini, M. G., Wider, G., Widmer, H., Wuthrich, K., and Walkinshaw, M. D. (1991) *Nature 353*, 276−9.
49. Howard, B. R., Vajdos, F. F., Li, S., Sundquist, W. I., and Hill, C. P. (2003) *Nat. Struct. Biol. 10*, 475−81.
50. Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002) *Science 295*, 1520−3.
51. Schreiber, S. L., and Crabtree, G. R. (1992) *Immunol. Today 13*, 136−42.
52. Bosco, D. A., Eisenmesser, E. Z., Pochapsky, S., Sundquist, W. I., and Kern, D. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 5247−52.
53. Amodeo, P., Morelli, M. A., and Castiglione Motta, A. (1994) *Biochemistry 33*, 10754−62.
54. Langsetmo, K., Fuchs, J., and Woodward, C. (1989) *Biochemistry 28*, 3211−20.
55. Chazin, W. J., Kordel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundstrom, T., and Forsen, S. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2195−8.
56. Feng, Y., Hood, W. F., Forgey, R. W., Abegg, A. L., Caparon, M. H., Thiele, B. R., Leimgruber, R. M., and McWherter, C. A. (1997) *Protein Sci. 6*, 1777−82.
57. Higgins, K. A., Craik, D. J., Hall, J. G., and Andrews, P. R. (1988) *Drug Des. Delivery 3*, 159−70.
58. Marsh, H. C., Scott, M. E., Hiskey, R. G., and Koehler, K. A. (1979) *Biochem. J. 183*, 513−7.
59. Perera, L., Darden, T. A., and Pedersen, L. G. (1998) *Biochemistry 37*, 10920−7.
60. Scanlon, M. J., and Norton, R. S. (1994) *Protein Sci. 3*, 1121−4.
61. Ogura, K., Nagata, K., Horiuchi, M., Ebisui, E., Hasuda, T., Yuzawa, S., Nishida, M., Hatanaka, H., and Inagaki, F. (2002) *J. Biomol. NMR 22*, 37−46.
62. Brown, R. D., 3rd, Brewer, C. F., and Koenig, S. H. (1977) *Biochemistry 16*, 3883−96.
63. Yang, Y., Sanchez, J. F., Strub, M. P., Brutscher, B., and Aumelas, A. (2003) *Biochemistry 42*, 4669−80.
64. Reimer, U., and Fischer, G. (2002) *Biophys. Chem. 96*, 203−12.
65. Markley, J. L., Williams, M. N., and Jardetzky, O. (1970) *Proc. Natl. Acad. Sci. U.S.A. 65*, 645−51.
66. Fox, R. O., Evans, P. A., and Dobson, C. M. (1986) *Nature 320*, 192−4.
67. Evans, P. A., Dobson, C. M., Kautz, R. A., Hatfull, G., and Fox, R. O. (1987) *Nature 329*, 266−8.

68. Pappenberger, G., Bachmann, A., Muller, R., Aygun, H., Engels, J. W., and Kiefhaber, T. (2003) *J. Mol. Biol. 326*, 235−46.
69. Kordel, J., Drakenberg, T., Forsen, S., and Thulin, E. (1990) *FEBS Lett. 263*, 27−30.
70. Heyeck, S. D., and Berg, L. J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 669−73.
71. Siliciano, J. D., Morrow, T. A., and Desiderio, S. V. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 11194−8.
72. Laederach, A., Cradic, K. W., Brazin, K. N., Zamoon, J., Fulton, D. B., Huang, X. Y., and Andreotti, A. H. (2002) *Protein Sci. 11*, 36−45.
73. Brazin, K. N., Fulton, D. B., and Andreotti, A. H. (2000) *J. Mol. Biol. 302*, 607−23.
74. Andreotti, A. H., Bunnell, S. C., Feng, S., Berg, L. J., and Schreiber, S. L. (1997) *Nature 385*, 93−7.
75. Mallis, R. J., Brazin, K. N., Fulton, D. B., and Andreotti, A. H. (2002) *Nat. Struct. Biol. 9*, 900−5.
76. Valegard, K., Liljas, L., Fridborg, K., and Unge, T. (1990) *Nature 345*, 36−41.
77. Golmohammadi, R., Valegard, K., Fridborg, K., and Liljas, L. (1993) *J. Mol. Biol. 234*, 620−39.
78. Stehle, T., and Harrison, S. C. (1997) *EMBO J. 16*, 5139−48.
79. Stonehouse, N. J., Valegard, K., Golmohammadi, R., van den Worm, S., Walton, C., Stockley, P. G., and Liljas, L. (1996) *J. Mol. Biol. 256*, 330−9.
80. Grochulski, P., Li, Y., Schrag, J. D., and Cygler, M. (1994) *Protein Sci. 3*, 82−91.
81. Grochulski, P., Li, Y., Schrag, J. D., Bouthillier, F., Smith, P., Harrison, D., Rubin, B., and Cygler, M. (1993) *J. Biol. Chem. 268*, 12843−7.
82. Truckses, D. M., Somoza, J. R., Prehoda, K. E., Miller, S. C., and Markley, J. L. (1996) *Protein Sci. 5*, 1907−16.
83. Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E., Jr., Richardson, D. C., Yonath, A., and Richardson, J. S. (1971) *J. Biol. Chem. 246*, 2302−16.
84. Ansari, H., Greco, G., and Luban, J. (2002) *Mol. Cell. Biol. 22*, 6993−7003.
85. Yuan, X., Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D., and Downing, K. (1998) *Protein Sci. 7*, 2127−35.
86. Cooper, J. A., and Howell, B. (1993) *Cell 73*, 1051−4.
87. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph. 14*, 51−5, 29−32.

BI0350710