ELSEVIER

Biophysical Chemistry 96 (2002) 203–212

Biophysical Chemistry

www.elsevier.com/locate/bpc

# Local structural changes caused by peptidyl–prolyl *cis/trans* isomerization in the native state of proteins

Ulf Reimer[a,1], Gunter Fischer[b,*]

[a]*Institut für Organische Chemie, AK Griesinger, Universität Frankfurt, Marie-Curie Str. 11, D-60439 Frankfurt/Main, Germany*
[b]*Max Planck Unit for Enzymology of Protein Folding, Weinbergweg 22, D-06120 Halle/Saale, Germany*

Received 7 December 2000; received in revised form 8 October 2001; accepted 12 October 2001

**Abstract**

Peptidyl–prolyl *cis/trans* isomerization, observed in the native state of an increasing number of proteins, is of considerable biological significance. The first evidence for an asymmetric transmission along the polypeptide chain of the structural effects of prolyl isomerization is now derived from the statistics of the $C^{\alpha}/C^{\alpha}$-atom distance distributions in the crystal structures of 848 non-homologous proteins. More detailed information on how isomerization affects segments adjacent to proline is obtained from crystal structures of proteins, that are more than 95% homologous, and that exhibit two different states of isomerization at a particular prolyl bond. The resulting 64 cases, which represent 3.8% of the database used, form pairs of coordinates which were analyzed for the existence of isomer-specific intramolecular nonbonded $C^{\alpha}/C^{\alpha}$-atom distances around the critical proline, and for the positional preferences for particular amino acids in the isomeric sequence segment. The probability that a native protein exhibits both prolyl isomers in the crystalline state increases in particular with a Pro at the third position N-terminal to the isomeric bond ($-3$ position), and with Ser, Gly and Asp at the position preceding the isomeric bond ($-1$ position). Structural alignment of matched pairs of isomeric proteins generates three classes with respect to position-specific distribution of $C^{\alpha}$-atom displacements around an isomeric proline imide bond. In the majority of cases the distribution of these intermolecular isomer-specific $C^{\alpha}$-atom distances shows a symmetric behavior for the N-terminal and C-terminal segment flanking the proline residue, and the magnitude did not exceed $1.3 \pm 0.6$ Å including the $C^{\alpha}$ atoms in proximity to the prolyl bond. However, in the remaining 12 protein pairs the structural changes are unidirectional relative to the isomerizing bond whereby the magnitude of the isomer-specific effect exceeds $3.0 \pm 2.0$ Å even at positions remote to proline. Interestingly, the magnitude of the intramolecular isomer-specific $C^{\alpha}$ atom displacements reveals a lever–arm amplification of the isomerization-mediated structural changes in a protein backbone. The observed backbone effects provide a structural basis for isomer-specific reactions of proline-containing polypeptides, and thus may play a role in biological recognition and regulation. 

*Keywords:* *Cis/trans* isomerization; Peptide bond; Protein structure; Native state isomerization; Proline; Isomer-specific reactions

*Corresponding author. Tel.: +49-345-552-2801; fax: +49-345-551-1972.
*E-mail address:* fischer@enzyme-halle.mpg.de (G. Fischer).
[1] Present address: Jerini AG, Rudower Chaussee 29, D-12489 Berlin, Germany.

0301-4622/02/$ - see front matter 
PII: S0301-4622(02)00013-3

## 1. Introduction

Proteins exhibit considerable conformational dynamics, undergoing remarkable structural fluctuations and inner motions during folding and during biological function. Backbone flexibility in the polypeptides arises mainly from the rotational freedom of the single bonds that are characterized by the torsion angles $\phi$ and $\psi$. The properties of the rigid peptide bond itself (angle $\omega$) substantially restricts the available conformational space, allowing molecules to populate only two ground state conformations, *cis* and *trans*, for rotation about the C–N bond. The general preference of amidic peptide bonds for the *trans* conformation [1] is somewhat reduced for the imidic prolyl peptide bond which commonly exhibits an isomer ratio [*cis*]/[*trans*] of 0.1 to 1 in unstructured peptide chains. However, all peptide bond isomers are separated by rotational barriers of 60–100 kJ $mol^{-1}$, allowing them to react individually in many biochemical processes. In fact, isomer-specificity has been observed directly for protein–ligand and enzyme–substrate interactions [2–7] and for protein folding [8], and indirectly through the catalytic effects of peptidyl–prolyl *cis*/*trans* isomerases [9–11]. The *cis* and *trans* isomers coexist in peptide chains that are in various folding states that have various chain length as well as in native proteins [5,12–21]. There is increasing evidence that coexisting isomers of native proteins, which differ by the *cis*/*trans* disposition of a particular prolyl bond, may occur frequently in solution but in greatly different relative amounts. Zhou et al. already showed native-state isomerization in two arbitrarily chosen proteins to arise from one particular prolyl bond, relying on isomer-specific dephosphorylation by protein phosphatase 2a and catalysis of this dephosphorylation by the peptidyl–prolyl *cis*/*trans* isomerases Pin1 [11].

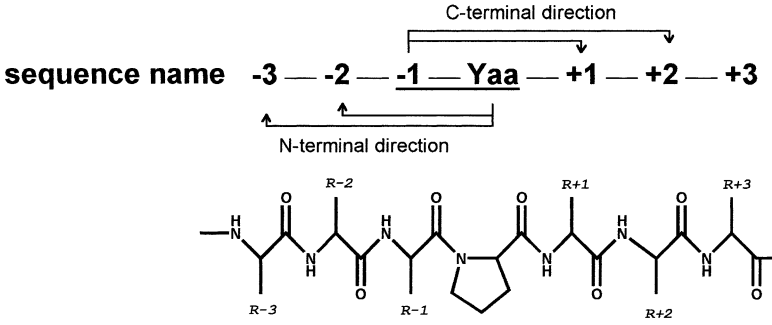
The conformational-selectivity discrimination factor for the association constants of isomer-specific protein–ligand interactions or the rate constants of isomer-specific enzyme reactions is often high ($>10^3$ fold) but the origin of the isomer-specific discrimination has a poorly defined structural basis. The general impact of proline residues on protein structure is familiar and reasonable on the base that (a) hydration is favored at proline residues; (b) proline frequently occurs at corner positions in $\beta$-turns and at the ends of strands and helices; and (c) the $\alpha$-helix breaking properties of proline [22–25]. An additional feature of proline is related to the uniquely high rotational barrier of prolyl isomerization. This rotation does not couple dynamically to the framework movement of the backbone. Due to this fact, functional properties of polypeptides could be monitored in dependence on the folding state of a particular segment of a peptide chain. Obviously, therefore, knowledge about the extent of structural reorganization during prolyl bond *cis*/*trans* isomerization either in the proximity of the prolyl bond or at remote positions of the protein backbone is critical and particularly so for the relationship between folding state and protein activity.

Direct comparison of the three-dimensional structure of individual proteins that exhibit alternative isomeric states for a particular prolyl bond should provide insight into such isomer-specific structural changes. However, multiple isomerization sites, low propensity for the *cis* isomer, the presence of the isomerization site in flexible protein segments, and poor dispersion of the isomer-specific NMR chemical shifts, usually prevent simultaneous structure determination in solution. In the present report, we have used the coordinates in the protein structure database of homologous proteins that have different native conformations of single prolyl bonds. These data were analyzed with respect to the consequences of isomerization for the backbone structure around the critical proline and for the positional preferences of particular amino acids. In addition, a statistical analysis of distances between $C^{\alpha}$ atoms for a number of non-homologous proteins has been performed in an attempt to predict the maximal distances across which alterations will occur as a result of prolyl isomerization.

## 2. Methods

### *2.1. Statistical distribution of $C^{\alpha}/C^{\alpha}$ distances in non-homologous proteins*

For a set of non-homologous protein structures from the pdb [38] the distances between $C^{\alpha}$ atoms




Scheme 1. Positional characterization of a polypeptide segment adjacent to a 'central' peptide bond. In the upper, schematic drawing of the 'central' peptide bond is underlined. In the lower panel, the corresponding formula of the peptide segment is shown. In this case, Yaa is a prolyl residue.

in the vicinity of one certain proteinogenic amino acid Yaa was computed. For all 20 amino acids Yaa a data set was created containing the following distance information: $C^{\alpha}(-3)$ to $C^{\alpha}$(Yaa), $C^{\alpha}(-1)$ to $C^{\alpha}(+2)$, $C^{\alpha}(-2)$ to $C^{\alpha}$(Yaa) and $C^{\alpha}(-1)$ to $C^{\alpha}(+1)$ (for numbering see Scheme 1).

### 2.2. Evaluation of the isomeric state of prolyl bonds in homologous proteins and its structural consequences

An all against all FASTA sequence alignment was performed for 12606 protein chains with a length of at least 35 amino acids. Only structures determined by X-ray crystallography with a resolution of 2.6 Å or better were used. Sets of coordinates were created, each containing all homologous protein chains with a sequence identity of at least 95%. For the resulting 1699 sets of protein chains that contain more than one protein chain, dihedral angles ω were compared for all Xaa–Pro moieties with a minimal distance of 9 amino acids from the termini and sequence identity for all residues at positions $\pm 9$ from the proline. Sets containing structures with a particular prolyl-bond in the *cis* ($-20^{\circ} \geqslant \omega \geqslant 20^{\circ}$) conformation in one structure and in the *trans* ($-180^{\circ} \geqslant \omega \geqslant -160^{\circ}$ or $160^{\circ} \geqslant \omega \geqslant 180^{\circ}$) conformation in another structure were then selected for further evaluation.

For sets with multiple entries, the structures with the best resolution were chosen. For all sets of two homologous proteins, a structural alignment was performed using the Swiss-PdbViewer v3.6 [26]. In this procedure, the coordinates of all corresponding $C^{\alpha}$ atoms in both structures were aligned with each other. The average RMSD of all aligned protein pairs is then 0.52 Å for the $C^{\alpha}$ atoms. All aligned protein pairs were saved in separate PDB files for distance evaluation using locally developed software. First, we calculated the intermolecular distances between corresponding $C^{\alpha}$ atoms from the two aligned structures. In a separate step, we calculated intramolecular distances between the $C^{\alpha}$ atom at position (i) and a $C^{\alpha}$ atom separated by a variable number of residue ($n$). These distances were compared for the respective *cis* and *trans* isomeric protein chain using Eq. (1).

$$\Delta = \left(\vec{C}^{\alpha}_{trans}(i) - \vec{C}^{\alpha}_{trans}(i+n)\right) - \left(\vec{C}^{\alpha}_{cis}(i) - \vec{C}^{\alpha}_{cis}(i+n)\right) \quad (1)$$

The following PDB data sets were evaluated in this way (pdb-code *cis*, pdb-code *trans*, sequence number for the respective prolyl residue;…): 1a4m_A, 2ada, 114; 1a8e, 1a8f, 142, 145; 1ado_A, 1ald, 158; 1ao6_B, 1bj5, 96; 1aox_A, 1aox_B, 307; 1asc, 1art, 200; 1b1i_A, 2ang_A, 38; 1bem, 1cpg, 190; 1btc, 1byb, 201; 1btk_A, 1btk_B, 110; 1c4a_A, 1feh_A, 100; 1cak, 1avn, 30; 1cbz_A, 1cbz_B, 354; 1cci, 1cca, 190; 1chd, 1a2o_B, 258; 1cmv_A, 1lay, 114; 1djx_B, 1djx_A, 510; 1dpg_A, 1dpg_B, 149; 1eai_B, 1eai_A, 198; 1fdy_A, 1fdy_D, 273; 1gfl_B, 1c4f_A, 89;

1ggu_A, 1ggy_A, 411; 1gym, 2ptd, 282; 1hcl, 1b38_A, 254; 1hfp, 1drf, 66; 1hmy, 6mht_A, 113; 1hse, 1lcf, 141; 1iow, 1iov, 186; 1iso, 1sjs, 262; 1jsu_A, 1hcl, 155; 1kif_A, 1an9_A, 41; 1kxu, 1jkw, 73; 1lya_B, 1lyw_B, 177; 1pcz_A, 1d3u_A, 53; 1qa7_A, 1hav_A, 134; 1qf5_A, 1hop_B, 236; 1qnw_C, 1qnw_B, 42; 1qpg, 3pgk, 204; 1qq5_A, 1qq5_B, 107; 1qr2_B, 1qr2_A, 129; 1ryc, 1aes, 190; 1ubs_A, 2wsy_A, 28; 1vin, 1fin_B, 346; 1wbf_A, 1wbl_A, 83; 1yaa_A, 1yaa_B, 138; 1ytb_A, 1ytf_A, 200; 2bam_A, 2bam_B, 39; 2bmh_A, 1bu7_A, 196; 2msb_B, 1buu_A, 186; 2oat_A, 2can_A, 199; 2pf2, 2pf1, 54; 2prk, 1bjr_E, 171; 2vub_H, 3vub, 28; 3dap_B, 3dap_A, 121; 3por, 2por, 117; 3rn3, 1c0b_A, 93; 3thi_A, 4thi_A, 191; 4enl, 1ebh_A, 265; 4icb, 3icb, 43; 4ktq_A, 2ktq_A, 579; 4nul, 1fln, 58. The three sets 1tlf_C, 1tlf_A, 332; 1trh, 1lpp, 92 and 1mst_B, 1mst_A, 78 exhibit massive structural changes in the region of the respective prolyl bonds, and were not included in the evaluation of $C^{\alpha}/C^{\alpha}$-distances.

A control set of 24 protein structural pairs with no isomeric peptide bond showed RMSD values $<0.4$ Å in calculations corresponding to Fig. 3, $<0.1$ Å in calculations corresponding to Figs. 4 and 5.

### *2.3. Positional preferences in protein data sets with different isomeric structures*

The sequences of the segments of the proteins containing an isomeric peptide bond in different structures were evaluated for preferences of certain amino acids in such areas.

The frequencies were normalized by dividing the observed percentage of each amino acid by the statistical percentage of each amino acid in the PDB [27].

## 3. Results and discussion

### *3.1. Statistical distribution of $C^{\alpha}/C^{\alpha}$ distances in non-homologous proteins*

Obviously, peptide bond isomerism should particularly affect atom distances around the isomeric bond between residues $-1$ and Yaa (Scheme 1). Four intramolecular nonbonded $C^{\alpha}/C^{\alpha}$ distances were computed for each 6 residue sequence in the entire group of 848 non-homologous proteins that had one particular residue of the normal 20 at position Yaa. The distances $C^{\alpha}(-3)$ to $C^{\alpha}$(Yaa) and $C^{\alpha}(-2)$ to $C^{\alpha}$(Yaa) reflect the influence of isomerism across $C^{\alpha}(-1)/C^{\alpha}$(Yaa) on the $C^{\alpha}$ positions N-terminal to the potentially isomerizing peptide bond. The distances $C^{\alpha}(-1)$ to $C^{\alpha}(+1)$ and $C^{\alpha}(-1)$ to $C^{\alpha}(+2)$ similarly reflect the influence of isomerism across $C^{\alpha}(-1)/C^{\alpha}$(Yaa) on the $C^{\alpha}$ positions C-terminal to the potentially isomeric bond.

For 19 of the 20 natural amino acids at position Yaa the statistical distribution pattern of $C^{\alpha}/C^{\alpha}$ distances in the N-terminal direction was identical to that in the C-terminal direction (data not shown). However, with proline at position Yaa the statistical distribution of the $C^{\alpha}/C^{\alpha}$ distances in the two directions were different and fell into two categories depending on whether the $C^{\alpha}(-1)/C^{\alpha}$(Yaa) peptide bond was in the *cis* or *trans* conformation (Fig. 1). The isomer-specific differences are most dramatically illustrated by the $C^{\alpha}(-1)$ to $C^{\alpha}(+2)$ data. The *trans* data show a bimodal distribution centered approximately 6 Å, where the *cis* data show a unimodal distribution centered approximately 5 Å. These data suggest that there is a high probability for a protein to undergo a considerable backbone expansion during *cis*-to-*trans* isomerization. The effect appears in fact to be present also in $C^{\alpha}(-1)$ to $C^{\alpha}(+2)$ data but the broadened distribution render less visually apparent the distinction.

In contrast the distribution patterns for $C^{\alpha}(-2)$ to $C^{\alpha}$(Yaa) and $C^{\alpha}(-3)$ to $C^{\alpha}$(Yaa) distances are essentially similar for the *cis* and the *trans* isomers. Thus, the *cis*/*trans* interconversion results in almost no propagation of conformational effects along the backbone in the N-terminal direction. Clearly, conformational changes upon prolyl isomerization affect remote peptide segments asymmetrically.

### *3.2. The isomeric state of prolyl bonds in homologous proteins*

For non-homologous proteins the statistical characterization of chain displacements caused by

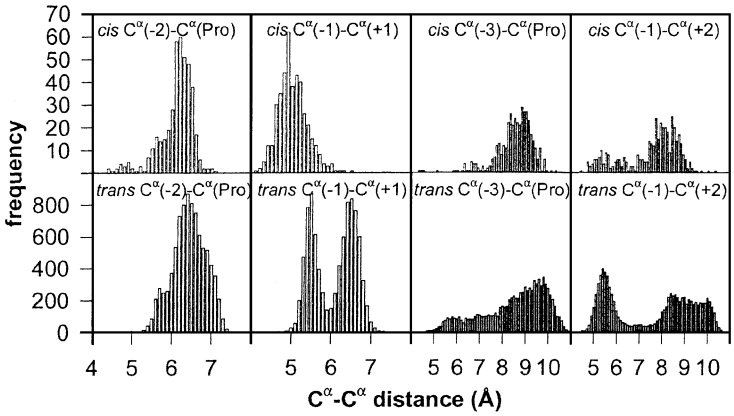



Fig. 1. Distribution of $C^{\alpha}/C^{\alpha}$ distances around Xaa–Pro sequences in 848 non-homologous protein structures [38]. The numbering is according to Scheme 1. The red columns are derived from structures with a *cis* peptide bond between amino acid −1 and Yaa (identical to proline), while the black columns have this bond in *trans* conformation.

prolyl isomerization can only be approximate because no clear separation of isomer-specific effects from the effects of primary-structure differences can be made.

However, protein structure database does contain crystallographic coordinates of proteins which are identical, or nearly identical in sequence and which differ in the isomeric state of a single prolyl bond. As a result, a direct comparison of isomer-specific atom coordinates is possible for these proteins, and such comparison should reveal isomer-specific atom displacements that arise from *cis/trans* isomerization. Out of 12 606 protein structures compiled, as many as 1699 form subsets of at least two members with >95% sequence identity. The conformational state of prolyl bonds was then compared within these 1699 subsets. A *cis/trans* distinction at a single prolyl bond was found in 64 cases. They were collected in a sample of 64 pairs of coordinates. Possible reasons for the structural differences between the proteins in a given subset include crystallization in different crystal forms (2 cases), subunit structures in oligomers (14 cases), isolation from different organisms (1 case), binding of different ligands (34 cases), mutations at sites distant from the proline (10 cases) and crystallization under different conditions such as pH (1 case), temperature (1 case) or solvent composition (1 case).

The prolyl bonds involved are located in rather different regions of the proteins. Alternative conformations are often expected in flexible, solvent-exposed loops devoid of well-defined secondary structure but prolyl bond isomers are actually seen in other regions as well. The isomerization sites were found in segments characterized by a wide range of flexibility as indicated by the B-factors. Thus, isomeric prolyl bonds are located in buried environment as well as in solvent-accessible regions of the proteins. Linker regions, turns, or bends are frequent locations.

### 3.3. Positional preferences

Amino acid preferences in positions +3 to +1 and −3 to −1, flanking the isomeric bond (Scheme 1), were extracted from the database of the 64 protein pairs. Certain amino acids occur with a frequency higher by at least twofold the expected value: for position −3: Phe, Pro, Trp; position −2: Trp; position −1: Asp, Gly, Ser; position +1: Cys, Phe, Trp; position +2: Met; position +3: Pro (Fig. 2). Interestingly, there are 12 cases of coexisting isomers in the native state of proteins observed by NMR techniques in solution in which glycine preceding proline (position −1) is present in three cases and serine in one case [13,15,18,20]. Therefore, the probability of

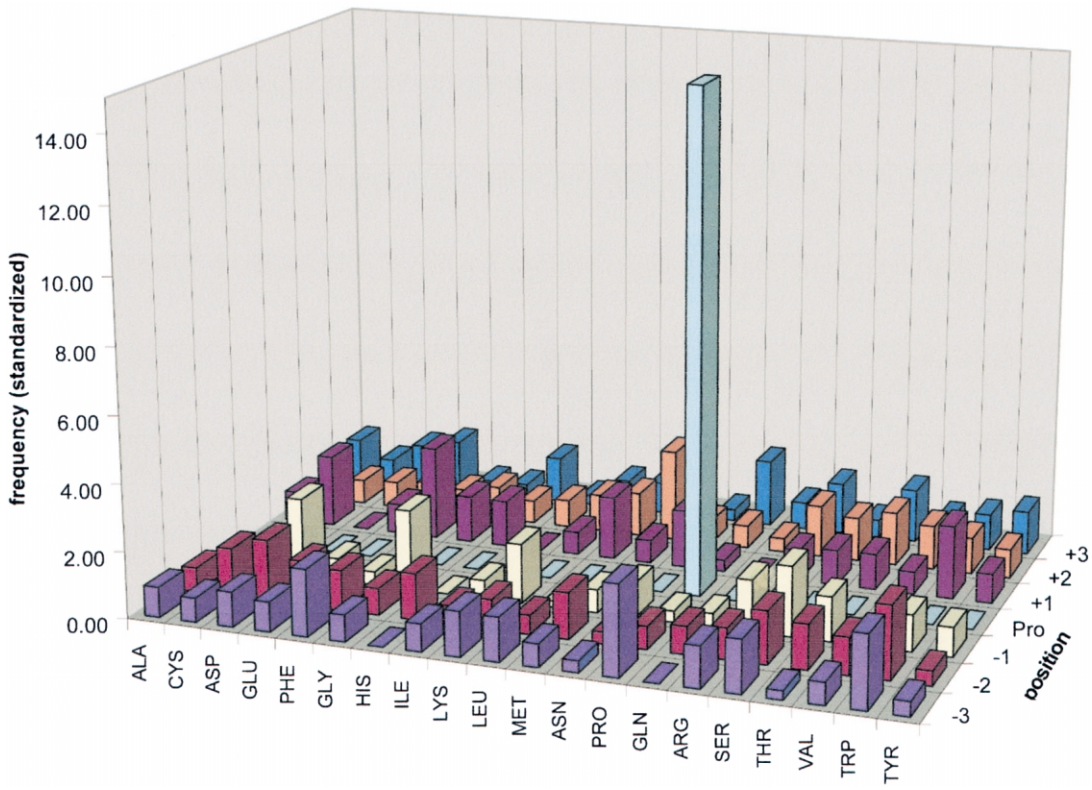



Fig. 2. Three-dimensional plot of the positional propensity of amino acids adjacent to the isomeric proline for a database of 64 homologous protein structures. The numbers of the frequency of occurrence of the amino acids were normalized by dividing the percentage found in the 64 structures divided by the statistical percentage of the respective amino acid in the PDB [27]. The value for proline is 22.22 since this position is in each case a proline (percentage 100%) and the statistical percentage for proline to occur in a pdb structure is 4.5%.

the occurrence of the two isomers of a particular Xaa–Pro in the crystalline state may be related to the tendency of a protein to acquire native state isomers in solution. This tendency may increase in sequence segments containing several favorable amino acid residues in sensitive positions. In this context, the preference scale permits the designation of the proline-rich segment Pro–$(Xaa)_2$–Pro–$(Yaa)_2$–Pro as a signature for enhanced native state conformational diversity in proteins. Proline-rich signature sequences are among the most frequently occurring motifs in *D. melanogaster* and *C. elegans* [28]. In addition, the region of the human p53 protein between amino acids 61 and 94 contains five repeats of the sequence Pro–$(Xaa)_2$–Pro) [29]. Deletion of this segment impairs the ability of p53 to suppress tumor cell growth. Using the MCF10AT model for early human breast cancer, it was shown that a relatively stable cell-specific conformation of wt p53 protein is responsible for alterations in growth and morphology that accompany preneoplastic stages of breast tumor progression. It should be noted that the uniquely slow interconversion rates of multiple prolyl bonds [30] might lead to long-lived native state protein conformers in which conformational information about different folding conditions in cells is stored.

### 3.4. Evaluation of structural changes

For the characterization of isomer-specific structural changes, we constructed 64 separate data sets

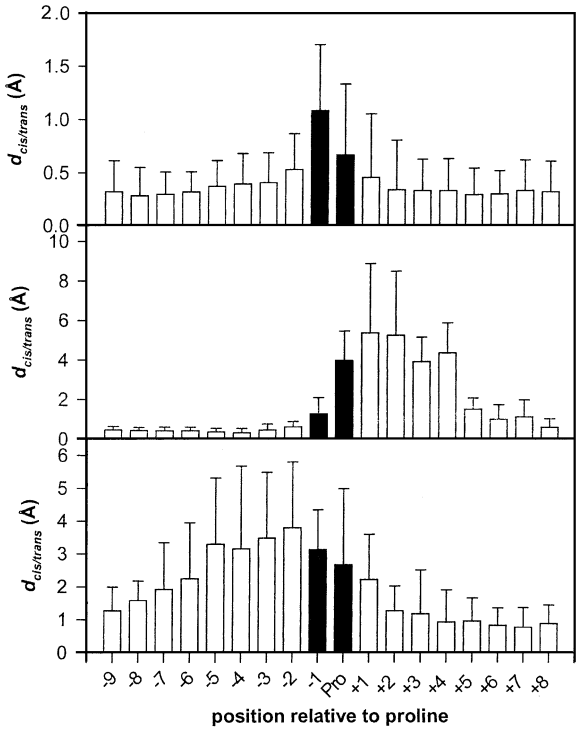



Fig. 3. Isomer-specific intermolecular distances of the respective $C^{\alpha}$ atoms in aligned pairs of homologous protein structures. The $d_{cis/trans}$-values represent average distances for (a) 49, (b) 2 and (c) 8 different protein sets. Black columns designate both $C^{\alpha}$ atoms of the Xaa–Pro moiety. The numbering of the residues is relative to the $C^{\alpha}$ atom of the isomeric proline. The average overall RMS deviation between the $C^{\alpha}$ atoms of all 61 aligned protein pairs is 0.52 Å.

from the pairwise three-dimensional alignment of the matched isomeric protein pairs. The average RMSD value for the structural alignment of the 64 pairs was 0.52 Å. Atomic displacements between isomeric pairs were analyzed for $C^{\alpha}$ atoms of the residues at positions +9 to −9 about the isomeric bond (numbered as in Scheme 1). For 49 protein pairs we found very small isomer-specific atomic displacements, averaging $<1$ Å (mean value 0.4 Å) at all positions from −9 to +9 (Fig. 3a). Only the $C^{\alpha}$ atoms of the residues that form the isomerizing bond show isomer-specific displacements greater than the average RMSD value of 0.52 Å. Thus, the backbone contraction observed at the prolyl bond itself during a *trans* to *cis* isomerization is not propagated noticeably through the chains for this group of proteins.

However, in the remaining 15 protein pairs the $C^{\alpha}$ atoms suffered a massive isomer-specific displacement even at positions remote from the proline residue. The effects suggest classification into three distinct subsets of proteins.

One subset with three members shows massive structural alterations with RMSD values of $>6$ Å in the entire peptide segment from positions −9 to +9. These three proteins were (1) *Candida rugosa* lipase (1trh, 1lpp) where the prolyl isomerization occurs at a hinge point for a flap movement important for the activity of this enzyme [31]; (2) the bacteriophage MS2 coat protein (1mst_B, 1mst_A) where a large loop has a Leu-Pro moiety in the *cis* and *trans* state [32]; (3) the dimeric *E. coli* Lac repressor where each of the constituent monomers shows a different conformation (1tlf_C, 1tlf_A). Since the origin of these effects remains unclear, these data have been omitted from the data analyses.

In the remaining 12 cases, the average RMSD values of the $C^{\alpha}$ atoms for the 18-residue segment about the isomeric peptide bond are between 1 and 3 Å. However, $C^{\alpha}$ atom displacements for these proteins do not exhibit a Gaussian distribution with peak maxima occurring at the proline position. The structures fall into two subsets according to whether displacements are more pronounced in the C-terminal or in the N-terminal direction. The propagation of the isomer-specific $C^{\alpha}$ atom displacement favors the C-terminal direction in 4 cases (Fig. 3b). Three out of these 4 proteins are cytochrome C oxidases having mutations in the position preceding the proline residue, with an overall sequence identity of $>98\%$. They show very similar structural effects of prolyl isomerization that include a displacement of $>4$ Å for the $C^{\alpha}$ atoms of the 4 amino acid residues following proline. Amplification by lever–arm action should be considered because this value is greater than the displacement at the prolyl bond itself. The plot in Fig. 3b arose from these three structures and an additional protein with similar isomer-specific structural properties.

For 8 proteins propagation of the isomer-specific $C^{\alpha}$ atom displacements favors the N-terminal

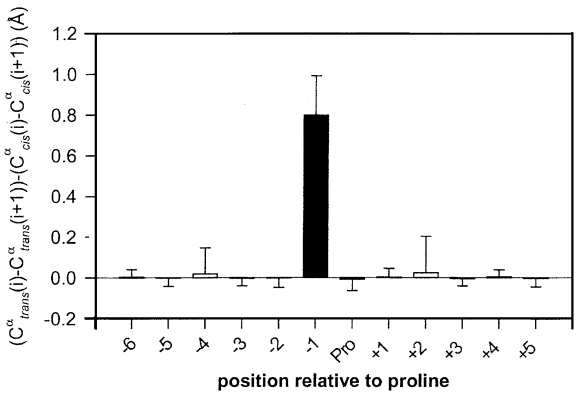



Fig. 4. Differences Δ of the $C_i^{\alpha}/C_{i+1}^{\alpha}$ intramolecular distances calculated for the isomeric proteins of a homologous protein pair in the two isomeric states. Averaged Δ values result from 61 protein pairs. The $C^{\alpha}$ of the respective proline is named Pro and the numbering of the $C^{\alpha}$ atoms of the adjacent amino acids refers to it. The $C_i^{\alpha}/C_{i+1}^{\alpha}$ distances from the respective *cis* conformer of the protein were subtracted from these of the homologous *trans* conformer.

region ($>3$ Å) over the C-terminal region ($<1.1$ Å) (Fig. 3c). The effect is still visible at the position $-5$ preceding proline whereas the amino acid segment following proline remains fixed. This suggests that unidirectional segment rearrangements can occur during the *cis*/*trans* isomerization.

However, calculation of intermolecular $C^{\alpha}$ atom distances cannot serve to differ between backbone contraction and backbone expansion during the isomerization reaction. Instead, isomer-specific intramolecular distance calculations provide insight into directional movements.

The isomer-specific quantity Δ [Eq. (1)] measures the magnitude and the sign of the intramolecular displacement of $C^{\alpha}$ atoms with a distance of *n* amino acid residues. These calculations should be performed for a number of constant step widths *n* scanning through the amino acid segment on both sides of proline. Small numbers of Δ indicate that the position *i* is relatively inert to the adjacent prolyl isomerization whereas large numbers indicate considerable reorganization. A positive sign of Δ indicates a chain contraction during a *trans* to *cis* isomerization whereas a negative Δ value describes chain expansion.

$$\Delta = \left(\vec{C}^{\alpha}_{trans}(i) - \vec{C}^{\alpha}_{trans}(i+1)\right) - \left(\vec{C}^{\alpha}_{cis}(i) - \vec{C}^{\alpha}_{cis}(i+1)\right) \quad (2)$$

Eq. (2) was used to analyze the intramolecular displacement of neighboring $C^{\alpha}$ atoms ($n=1$) in the whole set of 61 protein pairs (Fig. 4). No isomer-specific distance variations between two adjacent $C^{\alpha}$ atoms exist at all positions remote to proline with the exception of the distance between the $C^{\alpha}$ atoms attached to the isomeric bond itself, which lead to chain contraction in the *cis* relative to the *trans* isomer (Fig. 4).

Unfortunately, the intrinsic uncertainties in the three-dimensional protein coordinates cause a large level of error when isomer-specific data were analyzed with the whole set of proteins for larger distances *n*. Since larger distances *n* are expected to be more sensitive for structural alterations a subset of structures was used to determine the isomer-specific distance Δ for $n=2$. The error level could be tolerated using the subset of proteins already presented in Fig. 3b. The distance pattern of Fig. 5 indicates that the unidirectional propa-

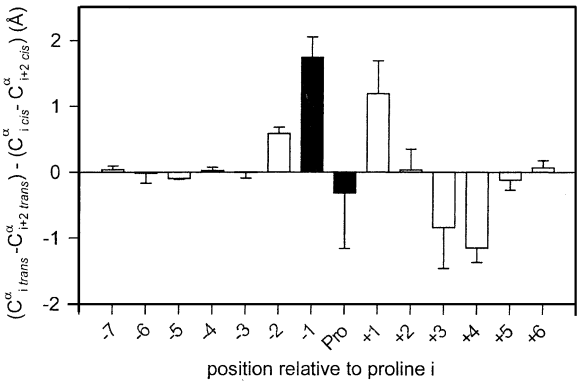



Fig. 5. Average differences Δ of the $C_i^{\alpha}/C_{i+2}^{\alpha}$ distances calculated for three isomeric protein pairs of Cytochrome C peroxidase. The isomerizing peptide bond is between Gly189 and Pro 190. Each of the three protein pairs has a different amino acid in the position following the proline ($i+1$). The $C^{\alpha}$ of the respective proline is named Pro and the numbering of the $C^{\alpha}$ atoms of the adjacent amino acids refers to it. The $C_i^{\alpha}/C_{i+2}^{\alpha}$ distances from the respective *cis* conformer of the protein were subtracted from these of the homologous *trans* conformer.

gation of the intermolecular $C^{\alpha}$ atoms distances in C-terminal direction (Fig. 3b) results from different intramolecular effects during the *cis*/*trans* isomerization. The clustering of negative $\Delta$ values at positions $+3$ and $+4$ indicates a more relaxed backbone structure in the *cis* conformation whereas compactness dominates in the proximity of proline. However, $\Delta$ values of about zero show fixed atomic positions in N-terminal direction. The results can be discussed in terms of a backbone region with immobile structure utilized to mediate a slow unidirectional chain movement during prolyl isomerization.

## 4. Conclusions

The isomerization of prolyl bonds in the native state of proteins has attracted little attention previously [33], doubtless because suitable probes for native-state isomerization are still lacking. In the present work, we have shown that three-dimensional structures with two distinct prolyl isomers exist for 3.8% of the homologous proteins in the database. It is clear that our analysis of these structural data correlates to the existence of two distinct isomeric structures in the crystalline state. Even though there is evidence for an equilibrium between these two isomeric states from NMR-measurements [13,15,18,20], this is not shown for our examples.

Proline-containing motifs have been shown to be important in the binding of ligands (substrates) to proteins including proteases [34], kinases [39], phosphatases [11] and peptidyl–prolyl *cis*/*trans* isomerases [35], and protein domains like the SH3- and WW-domains [36,37], involved in cellular regulation. Interactions involving these molecules show isomer-specificity with a complete absence of biological activity for one of the prolyl isomers. The structural basis for the discrimination of the inactive isomer remained unknown.

The present work may offer insight on how a very local bond rearrangement propagates through the backbone to remote locations in a protein. The range of effects of prolyl isomerization includes small structural changes restricted to the residues in proximity to the isomerizing peptide bonds, in some cases, and large changes that extend along the protein backbone asymmetrically in other cases. Such effects may enable a protein–ligand complex to utilize extended isomer-specific binding sites for functional interactions. Moreover, the unique dynamic properties of prolyl bond isomerization need enzyme catalysis in order for conformational control of protein function to occur.

## Acknowledgments

We thank Prof. Richard L. Schowen for fruitful discussions and correcting the manuscript.

## References

[1] G. Scherer, M.L. Kramer, M. Schutkowski, U. Reimer, G. Fischer, Barriers to rotation of secondary amide peptide bonds, J. Am. Chem. Soc. 120 (1998) 5568–5574.

[2] L.N. Lin, J.F. Brandts, Evidence suggesting that some proteolytic enzymes may cleave only the trans form of peptide bonds, Biochemistry 18 (1979) 43–47.

[3] G. Fischer, H. Bang, E. Berger, A. Schellenberger, Conformational specificity of chymotrypsin toward proline-containing substrates, Biochim. Biophys. Acta 791 (1984) 87–97.

[4] M. Brandsch, F. Thunecke, G. Küllertz, M. Schutkowski, G. Fischer, K. Neubert, Evidence for the absolute conformational specificity of the intestinal $H^+$/peptide symporter, J. Biol. Chem. 273 (1998) 3861–3864.

[5] K.K.-S. Ng, W.I. Weis, Coupling of prolyl peptide bond isomerization and $Ca^{2+}$ binding in a C-type mannose-binding protein, Biochemistry 37 (1998) 17977–17989.

[6] B.L. Stoddard, S. Pietrovski, Breaking up is hard to do, Nat. Struct. Biol. 5 (1998) 3–5.

[7] J.-B. Charbonnier, P. Belin, M. Moutiez, E.A. Stura, E. Quéméneur, On the role of the cis-proline residue in the active site of DsbA, Protein Sci. 8 (1999) 96–105.

[8] F.X. Schmid, Prolyl isomerase: enzymatic catalysis of slow protein-folding reactions, Annu. Rev. Bioph. Biom. 22 (1993) 123–142.

[9] Y.G. Chen, F. Liu, J. Massague, Mechanism of TGF-Beta receptor inhibition by FKBP12, EMBO J. 16 (1997) 3866–3876.

[10] M. Lopez-Ilasaca, C. Schiene, G. Küllertz, T. Tradler, G. Fischer, R. Wetzker, Effects of FK506-binding protein 12 and FK506 on autophosphorylation of epidermal growth factor receptor, J. Biol. Chem. 273 (1998) 9430–9434.

[11] X.Z. Zhou, O. Kops, A. Werner, M. Shen, P.-J. Lu, G. Stoller, G. Küllertz, M. Stark, G. Fischer, K.P. Lu, Essential role of phosphorylation-specific prolyl isomerization in phosphorylation signaling and cell survival', Mol. Cell. 6 (2000) 873–883.

[12] P.A. Evans, C.M. Dobson, R.A. Kautz, G. Hatfull, R.O. Fox, Proline isomerism in staphylococcal nuclease characterized by NMR and site-directed mutagenesis, Nature 329 (1987) 266–268.

[13] W.J. Chazin, J. Kordel, T. Drakenberg, E. Thulin, P. Brodin, T. Grundstrom, S. Forsen, Proline isomerism leads to multiple folded conformations of calbindin D9k: direct evidence from two-dimensional $^{1}$H-NMR spectroscopy, Proc. Natl. Acad. Sci. USA 86 (1989) 2195–2198.

[14] J. Bouckaert, R. Loris, F. Poortmans, L. Wyns, Crystallographic structure of metal-free concanavalin A at 2.5 Å resolution, Proteins 23 (1995) 510–524.

[15] R.K. Gitti, B.M. Lee, J. Walker, M.F. Summers, S. Yoo, W.I. Sundquist, Structure of the amino-terminal core domain of the HIV-1 capsid protein, Science 273 (1996) 231–235.

[16] M.P. O'Keefe, D. Nedelkov, A.L. Bieber, R.A. Nieman, Evidence for isomerization in myotoxin A rom the prairie rattlesnake (crotalus viridis viridis), Toxicon 34 (1996) 417–434.

[17] Y.Q. Feng, W.F. Hood, R.W. Forgey, A.L. Abegg, M.H. Caparon, B.R. Thiele, R.M. Leimgruber, C.A. McWherter, Multiple conformations of a human interleukin-3 variant, Protein Sci. 6 (1997) 1777–1782.

[18] X.R. Sheng, H.J. Zhang, X.M. Pan, X.F. Li, J.M. Zhou, Domain movement in rabbit muscle adenylate kinase might involve proline isomerization, FEBS Lett. 413 (1997) 429–432.

[19] E. Adjadj, V. Naudat, E. Quiniou, D. Wouters, P. Sautière, C.T. Craescu, Solution structure of the Lqh-8/6, a toxin like peptide from a scorpion venom. Structural heterogeneity induced by proline cis/trans isomerization, Eur. J. Biochem. 246 (1997) 218–227.

[20] X. Yuan, J.M. Werner, V. Knott, P.A. Handford, J.D. Campbell, A.K. Downing, Effects of proline cis-trans isomerization on TB domain secondary structure, Protein Sci. 7 (1998) 2127–2135.

[21] J.C. Cheetham, D.M. Smith, K.H. Aoki, J.L. Stevenson, T.J. Hoeffel, R.S. Syed, J. Egrie, T.S. Harvey, NMR structure of human erythropoietin and a comparison with its receptor bound conformation, Nature Struct. Biol. 5 (1998) 861–866.

[22] P.R. Gibbs, A. Radzicka, R. Wolfenden, The anomalous hydrophilic character of proline, J. Am. Chem. Soc. 113 (1991) 4714–4715.

[23] D.E. Stewart, A. Sarkar, J.E. Wampler, Occurrence and role of cis peptide bonds in protein structures, J. Mol. Biol. 214 (1990) 253–260.

[24] M.W. MacArthur, J.M. Thornton, Influence of proline residues on protein conformation, J. Mol. Biol. 218 (1991) 397–412.

[25] D. Pal, P. Chakrabarti, Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations, J. Mol. Biol. 294 (1999) 271–288.

[26] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.

[27] J. Wojcik, J.-P. Mornon, J. Chomilier, New efficient statistical sequence-dependent structure prediction of short to medium-sized protein loops based on an exhaustive loop classification, J. Mol. Biol. 289 (1999) 1469–1490.

[28] G.M. Rubin, et al., Comparative genomics of the eucaryotes, Science 287 (2000) 2204–2215.

[29] K.K. Walker, A.L. Levine, Identification of a novel p53 functional domain that is necessary for efficient growth suppression, Proc. Natl. Acad. Sci. USA 93 (1996) 15335–15340.

[30] U. Reimer, G. Scherer, M. Drewello, S. Kruber, M. Schutkowski, G. Fischer, Side chain effects on peptidyl-prolyl cis/trans isomerisation, J. Mol. Biol. 279 (1998) 449–460.

[31] P. Grochulski, Y. Li, J.D. Schrag, M. Cygler, Two conformational states of *Candida rugosa* lipase, Protein Sci. 3 (1994) 82–91.

[32] N.J. Stonehouse, K. Valegard, R. Golmohammadi, S. Vandenworm, C. Walton, P.G. Stockley, L. Liljas, Crystal structure of MS2 capsids with mutations in the subunit FG loop, J. Mol. Biol. 256 (1996) 330–339.

[33] W.-F. Yu, C.-S. Tung, H. Wang, M.L. Tasayco, NMR analysis of cleaved *Escherichia Coli* thioredoxin (1–73/74–108) and its P76A variant: cis/trans peptide isomerization, Protein Sci. 9 (2000) 20–28.

[34] D.F. Cuningham, B. O'Connor, Proline specific peptidases, Biochim. Biophys. Acta 1343 (1997) 160–186.

[35] G. Fischer, Peptidyl-Prolyl cis/trans isomerases and their effectors, Angew. Chem. Int. Ed. Engl. 33 (1994) 1415–1436.

[36] S. Feng, J.K. Chen, H. Yu, J.A. Simon, S.L. Schreiber, Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions, Science 266 (1994) 1241–1247.

[37] O. Staub, D. Rotin, WW domains, Structure 4 (1996) 495–499.

[38] U. Hobohm, C. Sander, Enlarged representative set of protein structures, Protein Sci. 3 (1994) 522–524.

[39] M. Weiwad, G. Küllertz, M. Schutkowski, G. Fischer, Evidence that the substrate backbone conformation is critical to phosphorylation by p42 MAP kinase, FEBS Lett. 478 (2000) 39–42.