Conservation of cis prolyl bonds in proteins during evolution

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · FEBRUARY 2004

Impact Factor: 2.63 · DOI: 10.1002/prot.20342 · Source: PubMed

5 AUTHORS, INCLUDING:

Stephan Lorenzen
Bernhard Nocht Institute for Tropical Medic…
22 PUBLICATIONS **505** CITATIONS
SEE PROFILE

Bjoern Peters
La Jolla Institute for Allergy & Immunology
211 PUBLICATIONS **7,061** CITATIONS
SEE PROFILE

Andrean Goede
Charité Universitätsmedizin Berlin
47 PUBLICATIONS **861** CITATIONS
SEE PROFILE

Robert Preissner
Charité Universitätsmedizin Berlin
122 PUBLICATIONS **2,471** CITATIONS
SEE PROFILE

Conservation of *Cis* Prolyl Bonds in Proteins During Evolution

Stephan Lorenzen,* Björn Peters, Andrean Goede, Robert Preissner, and Cornelius Frömmel
Institute of Biochemistry, Charité Berlin, Berlin, Germany

ABSTRACT **In proteins and peptides, the vast majority of peptide bonds occurs in *trans* conformation, but a considerable fraction (about 5%) of X-Pro bonds adopts the *cis* conformation. Here we study the conservation of *cis* prolyl residues in evolutionary related proteins. We find that overall, in contrast to local, protein sequence similarity is a clear indicator for the conformation of prolyl residues. We observe that *cis* prolyl residues are more often conserved than *trans* prolyl residues, and both are more conserved than the surrounding amino acids, which show the same extent of conservation as the whole protein. The pattern of amino acid exchanges differs between *cis* and *trans* prolyl residues. Also, the *cis* prolyl bond is maintained in proteins with sequence identity as low as 20%. This finding emphasizes the importance of *cis* peptide bonds in protein structure and function. Proteins 2005;58:589–595.**

Key words: protein structure; *cis* proline; homology; evolution; conservation

INTRODUCTION

In the early days of protein structure determination, it was broadly accepted that peptide bonds generally occur in the *trans* conformation.[1] But more and more structures are known that contain *cis* prolyl residues, and it appears that in earlier studies and with poorly resolved X-ray data, *cis* prolyl residues were often assumed as trans.[2]

Cis peptide bonds play a critical role in a variety of processes: The function of the Bowman–Birk protease inhibitor domain has been shown to depend on a critical *cis* prolyl bond.[3] In several proteins (e.g., prothrombin- and mannose-binding protein[4,5]), a *cis–trans* isomerization takes place after binding of ligands. In the case of prothrombin, the isomerization is essential for membrane binding.[6] Prolyl isomerization can also influence ligand–substrate binding of enzymes,[7–9] and isomerization of a single bond can dramatically change the protein architecture and act as a lever arm.[10,11] *Cis–trans* isomerization was also shown to be an important step in regulation: the prolyl isomerase Pin1[12,13] specifically isomerizes pSer/Thr–Pro bonds and thus acts in several steps in mitotic regulation (e.g., by regulating the dephosphorylation of Cdc25C and Tau protein).[14] Several studies on proline mutants showed the impact of *cis* prolyl residues on folding kinetics[15] and stability[16,17] of proteins. The immunosuppressive effect of prolyl isomerase inhibitors (e.g., cyclosporin) is long known (see review[18]).

Hence, the prediction of a *cis* prolyl bond is important for structure prediction and homology modeling. Several groups analyzed different nonredundant sets of protein structures and tried to find local sequence-based rules for the occurrence of a *cis* prolyl residue.[19–22] In general, one observes that aromatic amino acids often precede *cis* prolyl residues. However, these studies are based on just counting the occurrences of different amino acids around the proline without testing the predictive impact and thus relevance of the findings for structure prediction.

Not only the total incidence of *cis* prolyl residues but also the amino acid preferences varied according to different authors: While Stewart et al.[23] found that the Trp–Pro bond never adopts *cis* conformation while Tyr–Pro was in *cis* conformation in 25% of cases, Pal and Chakrabarti[19] later found with an expanded database more than 10% of the *cis* prolines with a Trp proceeding the proline, while Tyr–Pro only adopted *cis* conformation in 9.7% of the cases.

After all, an important question remains whether only the local sequence surrounding the prolyl residue or the general folding pattern controls the formation of *cis* prolyl residues. Studies on peptides strongly indicate that the amino acid preceding proline in a pentamer affects the *cis* prolyl content qualitatively in the same way as in proteins.[20] However, mutations of *cis* prolyl residues to alanyl residues sometimes lead to a *cis* alanyl residue,[24–27] which would indicate that the overall fold of the protein directs the amino acid on this place to a *cis* conformation. An analysis of the conservation of *cis* prolyl residues also provides information about the importance of local versus global sequence environment, which would be relevant for modeling purposes.

In this article, we analyze the conservation of *cis* prolyl residues in related proteins and provide guidelines for the estimation of prolyl bond conformations.

MATERIAL AND METHODS

Data Sets for Propensity Analysis

Subsets of protein structures from the Protein Data Bank[28] (PDB) with different sequence identity thresholds were obtained from the PISCES server.[29]

Grant sponsor: Federal Ministry of Education and Research; Grant numbers: 01K00119 and 0312705B.

*Correspondence to: Stephan Lorenzen, Institute of Biochemistry, Charité Berlin, Monbijoustr. 2, 10117 Berlin, Germany. E-mail: stephan.lorenzen@charite.de

Received 9 July 2004; Accepted 30 August 2004

Published online 17 December 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20342

TABLE I. Occurrence of Nonprolyl *Cis* Amino Acids

Amino acid	No.
Ala	10
Cys	1
Asp	8
Glu	6
Phe	7
Gly	14
His	3
Ile	2
Lys	6
Leu	2
Met	0
Asn	10
Gln	2
Arg	4
Ser	13
Thr	14
Val	6
Trp	0
Tyr	10

For each prolyl residue, the torsion angle of the CO—NH bond (ω) with the preceding amino acid was calculated. Bonds with an angle between −45° and +45° were considered as *cis* prolyl bonds; bonds with an angle between 135° and 225° were considered as *trans*.

To include only structures of good resolution and avoid redundancy in the data set that might bias the analysis, we used a data set with 25% sequence identity threshold, resolutions of ≤ 2.0 Å, and *R*-factors of ≤ 0.25 obtained from the PISCES server. Prolyl residues of 1729 protein structures were examined. Considering only amino acids with at least 10 preceding and 10 succeeding neighbors, we found 742 *cis* and 14,502 *trans* prolyl residues, as well as 118 nonprolyl *cis* amino acids (Table I).

Propensity Analysis

To analyze the predictive impact of the amino acid composition around the prolyl residue, we calculated the propensity of each amino acid at 10 preceding and succeeding positions around the prolyl residue. To obtain a prediction value for each peptide, we multiplied these values for each amino acid of the respective peptide to yield the propensity to contain a *cis* prolyl residue. The propensity of a feature is defined as the fraction of *cis* peptides with this feature, divided by the fraction of *trans* peptides with this feature. We logarithmized the propensities, which are thus normalized to zero. Features with negative values direct a peptide in the *trans* conformation; features with positive values direct to the *cis* conformation.

The accuracy of the prediction was measured as area under receiver operating curve (AUC) value[30]: Peptides were sorted by their prediction score and, starting with the highest scored peptide, successively added to the list of peptides predicted as showing a peptide bond in *cis* conformation. After adding each peptide, the specificity (fraction of correctly identified *trans* peptides, true negatives/negatives) is plotted against the sensitivity (fraction of correctly identified *cis* peptides, true positives/positives). The integral under this curve is a good measure of prediction accuracy and equals the probability that a randomly chosen *cis* peptide has a higher score than a randomly chosen *trans* peptide. It can obtain values from 0.5 for a random assignment to 1.0 for a perfect prediction.

To obtain AUC values, the propensities were calculated based on a training set of variable size, and peptides in the blind set (the remaining data) were predicted. Since the method does not involve any time-consuming learning process, multiple training and blind sets can be evaluated to obtain statistically relevant results. The final AUC value is the mean AUC value of several randomly chosen training and blind sets.

Analysis of Homologous Proteins

Alignments of proteins from our data set with proteins from the PDB were searched for in the HSSP database.[31] To avoid a bias from alignments with very prevalent proteins, we averaged the fractions of amino acids replacing proline or surrounding amino acids, respectively, for each protein in the data set. Standard deviations were calculated by bootstrapping.

Structural alignments for 4 Structural Classification of Proteins (SCOP) superfamilies (Table II; data obtained on April 15, 2004) were done using the combinatorial extension (CE) algorithm.[32]

RESULTS

Like other groups,[2,21,23] we found a strong dependence of *cis* prolyl content on the resolution of the protein structures (Fig. 1). In structures with high resolution, more than 6% of the X-Pro bonds had *cis* conformation, whereas in poorly resolved structures the fraction was below 3%. The fraction of *cis* prolyl residues was not dependent on the length of the proteins (data not shown). For our further analysis, we used only nonhomologous structures of good resolution (see Materials and Methods section).

To simulate predictions of unknown protein sequences, we randomly divided the data set into a training set, which is used to obtain the propensities, and a test set, which is predicted using the propensities generated from the training set. The fraction of the training set is variable: Small training sets use only a little data to "learn" and thus lead to poor results; large training sets result in small test sets whose prediction accuracy thus cannot be calculated precisely. We calculated propensity values for the amino acids around *cis* prolyl residues, which can be used to predict the conformation of prolyl residues. Taking 10 preceding and succeeding positions into account, the AUC value of the prediction rises to about 0.648 when the size of the training set is increased (Fig. 2). Even when taking large fractions of the data as a training set, the curve is still rising—obviously, a larger data set would further increase the prediction quality. We then tested which amino acid positions are relevant for the prediction. By

TABLE II. SCOP Superfamilies Used for Structural Alignment

Name	SCOP Code	No. proteins	No. prolines	No. *cis*-Pro
Restriction endonucleaselike	52980	144	1092	47
Microbial RNAses	53933	242	1007	258
Nucleotide binding	51971	77	1009	52
Alkaline phosphatases	53649	77	1723	49

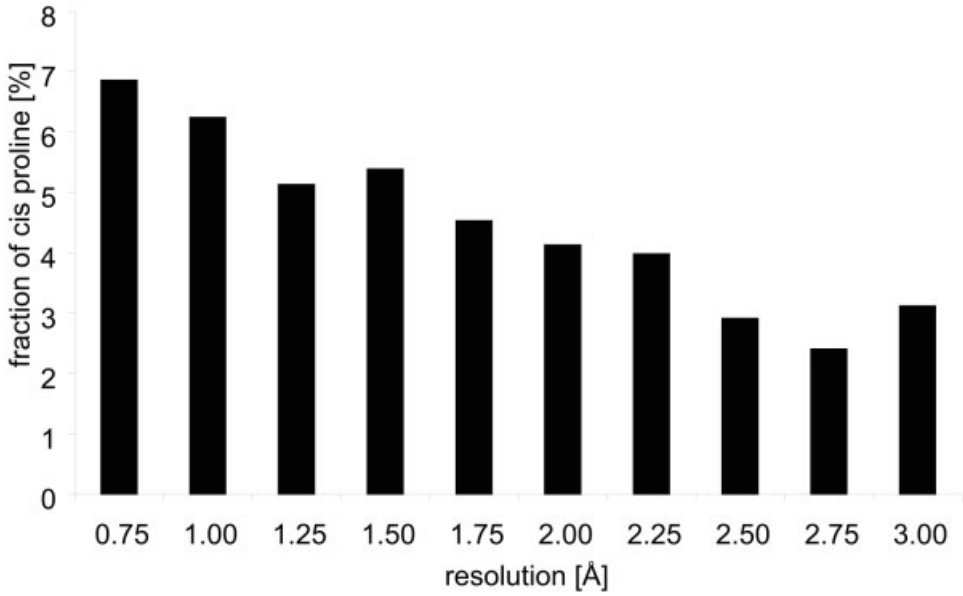

Fig. 1. Percentage of *cis* prolyl residues versus resolution of the data set. Protein structures with less than 25% sequence identity were separated into sets of different resolutions. Every bar shows the percentage of *cis* prolyl residues in proteins with a resolution in a window of width 0.25 Å around the specified value.

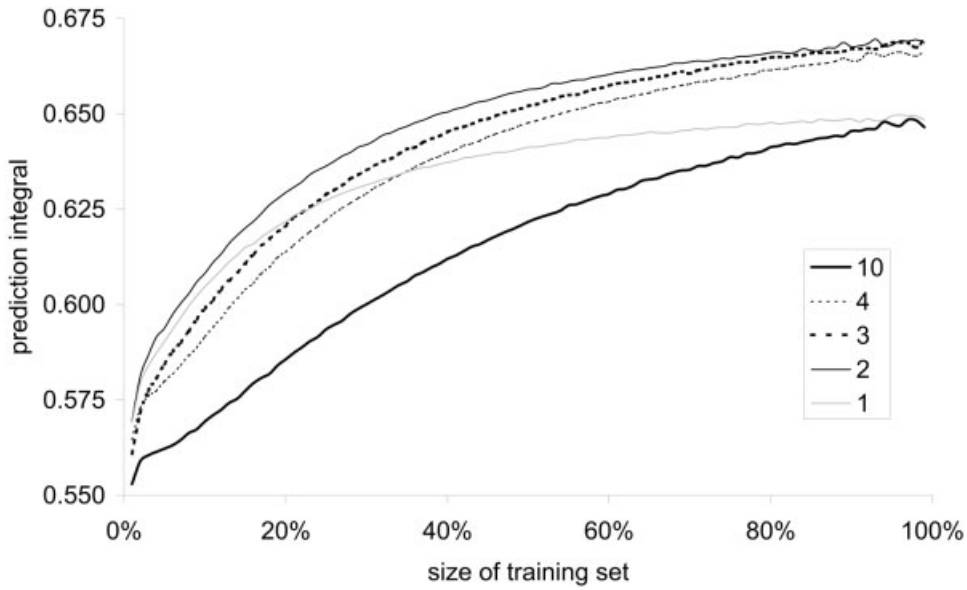

Fig. 2. AUC values for prediction of *cis* prolyl residues versus size of the training set. Starting with 10 preceding and succeeding amino acids around the prolyl residue for the prediction (bold line), the width of the sequence window used for prediction was narrowed to 4 (dashed line), 3 (bold dashed line), 2 (solid line), and 1 (gray line) neighboring amino acid positions. The training set was increased in steps of 1% up to 99%.

TABLE III. Natural Logarithms of Propensities for Amino Acids to Be at Specific Positions Relative to *Cis Versus Trans* Prolines

	−2	−1	+1	+2
A	0.03	0.26	0.00	*−0.52*
C	**0.73**	−0.27	0.44	**0.54**
D	−0.18	*−0.54*	−0.46	0.05
E	−0.22	0.33	*−0.84*	−0.49
F	−0.10	0.25	**0.61**	−0.06
G	−0.06	**0.61**	−0.02	−0.08
H	−0.05	−0.43	0.47	0.06
I	−0.31	*−1.05*	−0.10	−0.11
K	0.12	0.01	−0.25	0.05
L	−0.21	*−0.75*	0.18	−0.38
M	0.22	−0.13	0.02	−0.44
N	0.06	0.01	−0.08	0.37
P	0.27	0.49	0.30	**0.83**
Q	0.39	0.18	−0.15	−0.08
R	−0.38	−0.22	−0.26	0.12
S	0.33	0.07	0.07	−0.30
T	0.13	−0.29	0.12	0.27
V	−0.20	*−0.68*	−0.29	−0.04
W	−0.42	**1.15**	−0.03	−0.01
Y	0.04	**0.84**	**0.70**	0.13

Values > 0.5 and < -0.5 are highlighted by bold and italic, respectively.

narrowing the sequence environment around the prolyl residue that was used for the prediction, we could raise the AUC value to 0.669 when only using 2 preceding and 2 succeeding amino acids (Fig. 2). With the current size of the data set, apparently only information for 4 parameters could be reliably extracted, since more sequence information by a wider sequence window does not lead to a better prediction. As can be seen best at smaller data set sizes, more positions introduce noise and thus decrease the prediction accuracy. Note that even the AUC value with only 1 predictive position on both sides of the prolyl residue is also higher than the value reached with 10 predictive positions on both sides. Fewer predictive positions need less data to converge (note the more asymptotic behavior of the AUC value when fewer positions are used), and the risk of overfitting is considerably smaller.

The resulting propensity matrix is shown in Table III. The amino acids preferred as *cis* proline neighbors show no conspicuous overlap with preferences of prolyl isomerases.[33–35] Grouping of amino acids according to several properties did not lead to an increase in prediction quality (data not shown).

We then determined the degree of conservation of *cis* and *trans* prolyl residues, as well as neighboring amino acids (Fig. 3). While the amino acid positions next to the prolyl residue are as conserved as the whole proteins [Fig. 3(A and C)], prolyl residues themselves are considerably

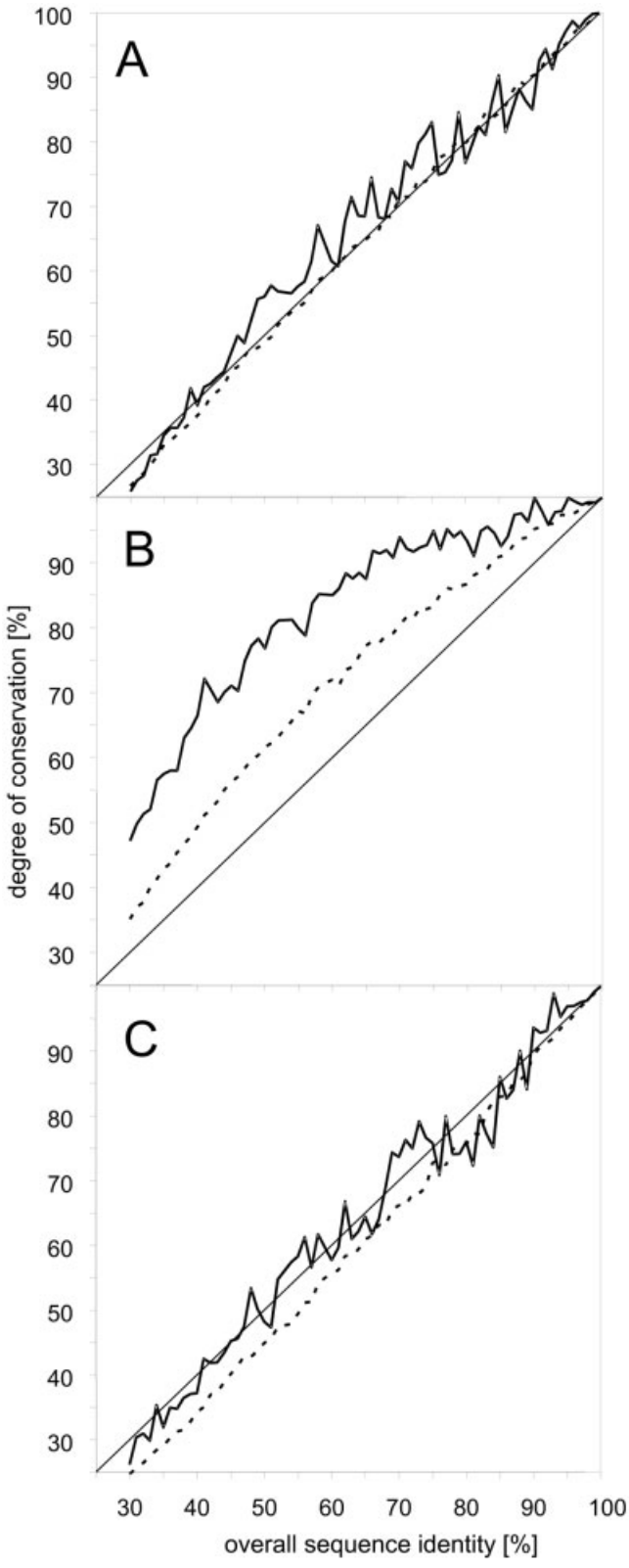

Fig. 3. Conservation of amino acid positions around the prolyl residue. For each prolyl residue in the *cis* and *trans* data set, multiple alignments from the HSSP database were compared. The degree of conservation of prolyl residues or neighboring amino acids is plotted against the percentage of conservation of the whole proteins. Solid line: *cis* prolyl residues; dotted line: *trans* prolyl residues. Thin line: conservation of the whole protein. (**A**) Residues preceding proline. (**B**) Prolyl residues. (**C**) Residues succeeding proline.

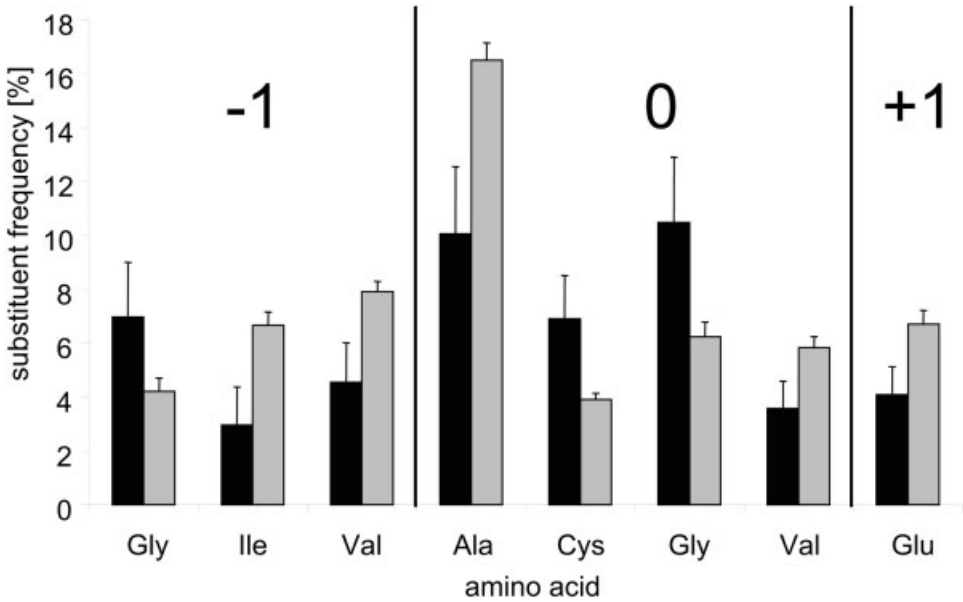

Fig. 4. Fraction of amino acid exchanges for residues preceding proline (−1), prolyl residues themselves (0), and residues succeeding proline (+1). Only amino acids with significant differences ($p < 0.001$) for *cis* (black) and *trans* (gray) prolyl residues are shown.

more conserved [Fig. 3(B)], with *cis* prolyl residues surpassing *trans* prolyl residues.

A comparison of exchange rates is shown in Figure 4. Amino acids preceding *cis* prolyl residues are more frequently replaced by glycine and less frequently replaced by isoleucine and valine than amino acids preceding *trans* prolyl residues ($p < 0.001$). Residues succeeding *cis* prolyl residues are less frequently exchanged to glutamate than residues succeeding *trans* prolyl residues ($p < 0.001$). These amino acid preferences represent the propensities shown in Table III. *Cis* prolyl residues themselves are preferably exchanged to cysteinyl or glycyl residues and less frequently changed to alanyl or valyl residues than *trans* prolyl residues ($p < 0.001$).

As the rather low AUC values have shown, the characteristics of neighboring amino acids alone are not sufficient to decide about the conformation of prolyl residues. Therefore, we studied the impact of local and global sequence homology between proteins from the PDB. For each *cis* prolyl residue, we gathered all peptides in the PDB with rising levels of sequence identity in a given window around the prolyl residue. For these peptides, we computed the fraction of peptides with proline in *cis* conformation (Fig. 5). Considering sequence identity over the whole protein, values of 30% are still sufficient to deduce the presence of a *cis* prolyl residue at a given position. With a declining width of the sequence window, higher degrees of sequence identity are necessary to be indicative of *cis* prolyl residues. Pairs of proteins with overall sequence identity lower than 30% derived from the PALI database[36] also show conserved *cis* prolyl residues (data not shown). To obtain a more comprehensive picture of the lower limit of sequence identity that still preserves *cis* prolyl residues, we structurally superimposed proteins of different superfamilies of the SCOP database[37] with each other (Table II and Fig. 6). All proteins of one superfamily were superimposed pairwise with each other using the CE algorithm,[32] and prolyl residues structurally aligned to *cis* prolyl residues were inspected. Figure 6 shows the fraction of prolyl residues aligned to *cis* prolyl residues where the *cis* conformation is maintained versus the pairwise identity between both proteins. Obviously, a sequence identity as low as 20% still indicates conservation of the proline conformation, whereas smaller values do not have any impact.

DISCUSSION

We were able to show that the sum of characteristics of only a few neighboring amino acids have a direct influence on the probability of forming a *cis* prolyl residue. As reported by other authors,[21,23,38] the amino acid directly preceding the proline has the greatest effect on determining proline conformations. At this position, the aromatic amino acids tryptophan and tyrosine are strongly preferred. Like Pal and Chakrabarti,[19] we found aromatic and small residues, as well as proline itself, on position −1 directing the bond to *cis*, and branched aliphatic amino acids to strongly speak against the formation of a *cis* prolyl residue. In contrast to them, we find no positive influence of polar amino acids. The positive influence of aromatic amino acids might be due to stacking of the 2 rings or CH$\cdots\pi$ interactions.[19,39] The influence of amino acids preceding proline has been shown to be qualitatively the same in peptides and proteins,[20] which is a sign that the local sequence environment plays a role in *cis–trans* isomerization of prolyl residues. Also, the small glycyl residue in this position raises the probability of acquiring a *cis* conformation, probably due to the lack of steric hindrance.

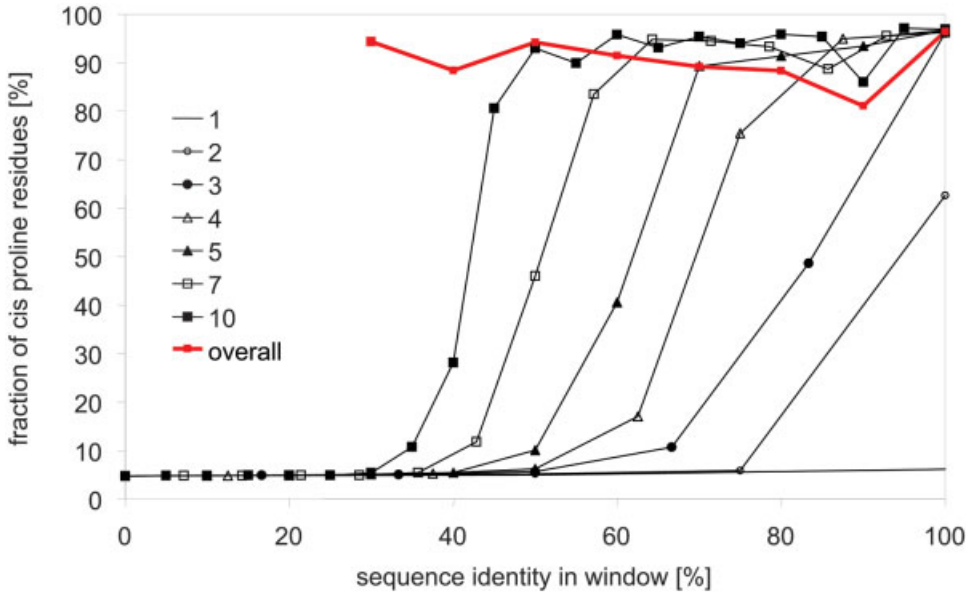

Fig. 5. Fraction of prolyl residues in *cis* conformation for peptides with different sequence identity to *cis* prolyl peptides. For sequence windows of various widths around the prolyl residue (different lines), the fraction is indicated. The line with filled squares shows the dependence on the degree of overall sequence identity.

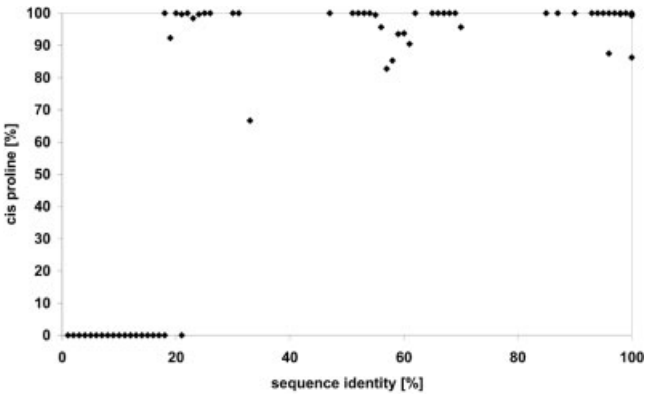

Fig. 6. Degree of conservation of *cis* prolyl residues versus sequence identity of proteins. All proteins belonging to four different SCOP superfamilies (restriction endonucleaselike, microbial ribonucleases, nucleotide-binding, and alkaline phosphataselike) were structurally aligned with each other. The fraction of prolyl residues aligned to *cis* prolyl residues where the *cis* conformation is maintained was plotted against the overall sequence identity of the proteins.

Phenylalanine and tyrosine, but not tryptophan, also facilitate a *cis* bond when succeeding the respective prolyl residue. On position +2, proline and cysteine residues have the highest incidence to occur. While only looking at directly surrounding residues allows no reliable prediction, wider windows around the prolyl residue raise the predictive impact of sequence identity. A more reliable prediction can be made when looking at sequence similarity to other proteins containing *cis* prolyl residues. Since also the consideration of residues more distant from the central proline raises the prediction accuracy, we conclude that overall homology, and not only the local sequence environment, directs the prolyl residue to *cis* or *trans* conformation. This suggestion is supported by the finding that overall sequence identity of only 20% is a solid indicator of *cis* prolyl residues. In homology modeling, the best decision about the presence of a *cis* prolyl residue can thus be made by modeling according to the highest resolution structure with more than 20% sequence identity to the query protein. As early as 1994, Sali and Overington[40] reported that the conformation of prolyl residues can be judged from the conformation of aligned residues.

The importance of *cis* prolyl residues can also be seen in the conservation analysis. While prolyl residues in general are more conserved than other amino acids, *cis* prolyl residues even surpass *trans* prolyl residues and have been described as the second most conserved residue type.[40] In addition to a higher degree of conservation (Fig. 3), the amino acid exchange pattern differs between *cis* and *trans* prolyl residues. We expected that the flexible glycyl residue is favored as substitute for *cis* prolyl residues, but also cysteine is favored as substitute in comparison to *trans* prolyl residues (Fig. 4). Obviously, some amino acids better approximate the critical role of the *cis* prolyl residues than others. The highest difference can be seen with alanine, which substitutes for *trans* prolyl residues in 16% of amino acid exchanges, while it only occurs in 10% of amino acid exchanges for *cis* prolyl residues. Alanine as the most frequent substitute for *trans* prolyl residues can be explained by the genetic exchangeability between proline and alanine (their codons differ by just one nucleotide), whereas the predominant substitute for *cis* prolyl residues, glycine, can be explained by the lack of steric hindrance.

Our studies underline that *cis* prolyl residues fulfill important roles in protein structures, are more conserved, and show different amino acid exchange patterns than

their *trans* counterparts. In contrast to local sequence motifs, overall homology is a much stronger indicator for the occurrence of *cis* prolyl residues.

REFERENCES

1. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. Adv Protein Chem 1968;23:283–438.
2. Weiss MS, Jabs A, Hilgenfeld R. Peptide bonds revisited. Nat Struct Biol 1998;5:676.
3. Brauer AB, Domingo GJ, Cooke RM, Matthews SJ, Leatherbarrow RJ. A conserved cis peptide bond is necessary for the activity of Bowman–Birk inhibitor protein. Biochemistry 2002;41:10608–10615.
4. Ng KK, Park-Snyder S, Weis WI. Ca2+-dependent structural changes in C-type mannose-binding proteins. Biochemistry 1998; 37:17965–17976.
5. Ng KK, Weis WI. Coupling of prolyl peptide bond isomerization and Ca2+ binding in a C-type mannose-binding protein. Biochemistry 1998;37:17977–17989.
6. Evans TC Jr, Nelsestuen GL. Importance of cis-proline 22 in the membrane-binding conformation of bovine prothrombin. Biochemistry 1996;35:8210–8215.
7. Mallis RJ, Brazin KN, Fulton DB, Andreotti AH. Structural characterization of a proline-driven conformational switch within the Itk SH2 domain. Nat Struct Biol 2002;9:900–905.
8. Brandsch M, Thunecke F, Kullertz G, Schutkowski M, Fischer G, Neubert K. Evidence for the absolute conformational specificity of the intestinal H+/peptide symporter, PEPT1. J Biol Chem 1998; 273:3861–3864.
9. Fischer G, Bang H, Berger E, Schellenberger A. Conformational specificity of chymotrypsin toward proline-containing substrates. Biochim Biophys Acta 1984;791:87–97.
10. Reimer U, Fischer G. Local structural changes caused by peptidyl-prolyl cis/trans isomerization in the native state of proteins. Biophys Chem 2002;96:203–212.
11. Andreotti AH. Native state proline isomerization: an intrinsic molecular switch. Biochemistry 2003;42:9515–9524.
12. Lu KP, Liou YC, Zhou XZ. Pinning down proline-directed phosphorylation signaling. Trends Cell Biol 2002;12:164–172.
13. Lu KP. Phosphorylation-dependent prolyl isomerization: a novel cell cycle regulatory mechanism. Prog Cell Cycle Res 2000;4:83–96.
14. Zhou XZ, Kops O, Werner A, Lu PJ, Shen M, Stoller G, Kullertz G, Stark M, Fischer G, Lu KP. Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins. Mol Cell 2000;6:873–883.
15. Wedemeyer WJ, Welker E, Scheraga HA. Proline cis–trans isomerization and protein folding. Biochemistry 2002;41:14637–14644.
16. McHarg J, Kelly SM, Price NC, Cooper A, Littlechild JA. Site-directed mutagenesis of proline 204 in the "hinge" region of yeast phosphoglycerate kinase. Eur J Biochem 1999;259:939–945.
17. Charbonnier JB, Belin P, Moutiez M, Stura EA, Quemeneur E. On the role of the cis-proline residue in the active site of DsbA. Protein Sci 1999;8:96–105.
18. Galat A, Metcalfe SM. Peptidylproline cis/trans isomerases. Prog Biophys Mol Biol 1995;63:67–118.
19. Pal D, Chakrabarti P. Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations. J Mol Biol 1999;294:271–288.
20. Reimer U, Scherer G, Drewello M, Kruber S, Schutkowski M, Fischer G. Side-chain effects on peptidyl-prolyl cis/trans isomerisation. J Mol Biol 1998;279:449–460.
21. Frömmel C, Preissner R. Prediction of prolyl residues in cis-conformation in protein structures on the basis of the amino acid sequence. FEBS Lett 1990;277:159–163.
22. Deane CM, Lummis SC. The role and predicted propensity of conserved proline residues in the 5-HT3 receptor. J Biol Chem 2001;276:37962–37966.
23. Stewart DE, Sarkar A, Wampler JE. Occurrence and role of cis peptide bonds in protein structures. J Mol Biol 1990;214:253–260.
24. Jin L, Stec B, Kantrowitz ER. A cis-proline to alanine mutant of *E. coli* aspartate transcarbamoylase: kinetic studies and three-dimensional crystal structures. Biochemistry 2000;39:8058–8066.
25. Birolo L, Malashkevich VN, Capitani G, De Luca F, Moretta A, Jansonius JN, Marino G. Functional and structural analysis of cis-proline mutants of *Escherichia coli* aspartate aminotransferase. Biochemistry 1999;38:905–913.
26. Mayr LM, Willbold D, Rosch P, Schmid FX. Generation of a non-prolyl cis peptide bond in ribonuclease T1. J Mol Biol 1994;240: 288–293.
27. Tweedy NB, Nair SK, Paterno SA, Fierke CA, Christianson DW. Structure and energetics of a non-proline cis-peptidyl linkage in a proline-202→alanine carbonic anhydrase II variant. Biochemistry 1993;32:10944–10949.
28. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
29. Wang G, Dunbrack RL. PISCES: a protein sequence culling server. Bioinformatics 2003;19:1589–1591.
30. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 1997; 30:1145–1159.
31. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991;9:56–68.
32. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
33. Uchida T, Fujimori F, Tradler T, Fischer G, Rahfeld JU. Identification and characterization of a 14 kDa human protein as a novel parvulin-like peptidyl prolyl cis/trans isomerase. FEBS Lett 1999; 446:278–282.
34. Rahfeld JU, Schierhorn A, Mann K, Fischer G. A novel peptidyl-prolyl cis/trans isomerase from *Escherichia coli*. FEBS Lett 1994; 343:65–69.
35. Harrison RK, Stein RL. Mechanistic studies of enzymatic and nonenzymatic prolyl cis–trans isomerization. J Am Chem Soc 1992;114:3464–3471.
36. Balaji S, Sujatha S, Kumar SS, Srinivasan N. PALI—a database of Phylogeny and ALIgnment of homologous protein structures. Nucleic Acids Res 2001;29:61–65.
37. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
38. MacArthur MW, Thornton JM. Influence of proline residues on protein conformation. J Mol Biol 1991;218:397–412.
39. Weisshoff H, Frost K, Brandt W, Henklein P, Mugge C, Frommel C. Novel disulfide-constrained pentapeptides as models for beta-VIa turns in proteins. FEBS Lett 1995;372:203–209.
40. Sali A, Overington JP. Derivation of rules for comparative protein modeling from a database of protein structure alignments. Protein Sci 1994;3:1582–1596.