Accepted for publication in Amino Acids 2012

Cis-trans peptide variations in structurally similar proteins

Agnel Praveen Joseph[1,2,3,+], Narayanaswamy Srinivasan[4] & Alexandre G. de Brevern[1,2,3*]

[1]INSERM, UMR_S 665, DSIMB, , F-75739 Paris, France.
[2]Univ Paris Diderot, Sorbonne Paris Cité, UMR-S665, Paris, F-75739, France.
[3]INTS, F-75739 Paris, , France.
[4]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India.

Short title: Conservation of ω dihedral angle.

* Corresponding author:
mailing address: de Brevern A.G., INSERM UMR-S 665, Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), Université Denis Diderot - Paris 7, INTS, 6, rue Alexandre Cabanel, 75739 Paris cedex 15, France
E-mail: alexandre.debrevern@univ-paris-diderot.fr
Tel: +33(1) 44 49 31 14
Fax: +33(1) 47 34 74 31

+ Current address: National Centre for Biological Sciences, TIFR, Bellary Road, Bangalore 560065, India

Key words : protein folds; *cis* peptides ; ω dihedral; *cis-trans* isomerization; structural alignment; structural alphabet; Protein Blocks; Protein Data Bank.

Abstract

The presence of energetically less favourable *cis* peptides in protein structures has been observed to be strongly associated with its structural integrity and function. Inter-conversion between the *cis* and *trans* conformations also has an important role in the folding process. In this study, we analyse the extent of conservation of *cis* peptides among similar folds. We look at both the amino acid preferences and local structural changes associated with such variations.

Nearly 34% of the Xaa-Proline *cis* bonds are not conserved in structural relatives; Proline also has a high tendency to get replaced by another amino acid in the trans conformer. At both positions bounding the peptide bond, Glycine has a higher tendency to lose the *cis* conformation. The *cis* conformation of more than 30% of β turns of type VIb and IV are not found to be conserved in similar structures. A different view using Protein Block based description of backbone conformation, suggests that many of the local conformational changes are highly different from the general local structural variations observed among structurally similar proteins.

Changes between *cis* and *trans* conformations are found to be associated with the evolution of new functions facilitated by local structural changes. This is most frequent in enzymes where new calalytic activity emerges with local changes in the active site. *Cis-trans* changes are also seen to facilitate inter-domain and inter-protein interactions. As in the case of folding, *cis-trans* conversions have been used as an important driving factor in evolution.

Introduction

The large majority of peptide bonds in protein structures are constrained to a planar conformation (Ramachandran and Sasisekharan 1968; Pauling et al. 1951) (*trans*) due to the partial double bond nature of the peptide linkage. This has been used as a strict dihedral angle restraint (ω~180°) for protein structure refinement in X-ray crystallography (Priestle 2003). However, several studies have identified deformations in planarity that has been tolerated in protein structures (MacArthur and Thornton 1991; Weiss et al. 1998).

A very small percent of all peptides are observed to occur in *cis* conformation (Ramachandran and Mitra 1976). Significant portion of about 5% of all Xaa-Pro peptides (Xaa being the amino acid preceding the Proline residue) were found in this conformation with the ω dihedral close to 0^o (Stewart et al. 1990; MacArthur and Thornton 1991; Weiss et al. 1998). On the other hand, only about 0.03% of the Xaa-nonPro peptides are reported to be in the *cis* form (Weiss et al. 1998). This behaviour is due to the imide bond character of the peptide joining a proline residue, which limits the steric cost to have ω close to 0°. Due to the strict constraints used during structure determination and refinement, many of the *cis* peptide bonds may have gone unreported, especially in low resolution structures (Weiss et al. 1998). The $C\alpha_i$-$C\alpha_{i+1}$ distance in *cis* conformation is nearly 1Å less than that seen in *trans* and hence there can be a strong influence of structure resolution on the *cis* bond content (Stewart et al. 1990; Weiss et al. 1998).

Cis bonds are preserved in protein structures for both structural and functional reasons (Brauer et al. 2002; Lummis et al. 2005; Wu and Matthews 2002). Xaa-nonPro peptides in *cis* conformation are observed near the active sites and are shown to be important for the protein activity (Weiss et al. 1998; Stoddard and Pietrokovski 1998; Jabs et al. 1999; Guan et al. 2004). Amino acid preferences associated with the *cis* conformation have also been studied with the help of high resolution structures. Aromatic residues such as Tyr, Trp and other amino acids like Gly, Ala and Pro are more frequent at the position Xaa, before the Pro (Pal and Chakrabarti 1999; Weiss et al. 1998).

The inter-conversion between the *cis* and *trans* isomers is a slow reaction and is catalysed in cell by peptidyl *cis-trans* isomerases (Fischer and Aumuller 2003). An activation barrier of about 20 kcal/mol has been proposed for this change (Grathwohl and Wuthrick 1981; Scherer et al. 1998). Unlike the Xaa-Pro *cis* bond, the non-prolyl peptides require a considerable additional cost (~2 kcal/mol) to be stabilized in this conformation. Several

studies on *cis-trans* isomerisation underline its importance in protein folding kinetics and function (Andreotti 2003). Proline isomerisation is shown to guide the unfolding and refolding paths of proteins (Schmid 1986; Jakob and Schmid 2008). Mutations destabilizing the *cis* conformation are also shown to reduce the folding rates (Odefey et al. 1995; Guan et al. 2004; Brandts et al. 1975).

Cis-trans conversion is implicated in cell signalling (Sarkar et al. 2007; Wulf et al. 2005), neuro-degeneration (Pastorino et al. 2006), amyloid formation (Eakin et al. 2006), channel gating (Lummis et al. 2005), phage infection (Eckert et al. 2005), structural dynamics (Grochulski et al. 1994; Guan et al. 2004) etc. This isomerisation has been recognized as a molecular timer that modulates the amplitude and duration of different cellular processes (Lu et al. 2007; Nicholson and Lu 2007). Hence, it has been also considered as a possible target for therapeutic studies (Lu et al. 2007).

The values of ω are associated with the other backbone dihedrals and a strict correlation was observed with the adjacent φ angle (Esposito et al. 2005). A strong association between the side chain and backbone conformation has also been noted for the *cis* conformation (Pal and Chakrabarti 1999). Local structural changes associated with *cis-trans* isomerisation have been studied based on the Cα distances (Reimer and Fischer 2002). Except for the position X (Xaa-Pro) the distances between consecutive Cαs did not show much variations. However significant changes were observed in the distances between $C\alpha_i$ and $C\alpha_{i+2,}$ where i stands for the residue at position X. β-turns of type VI has the characteristic backbone φ,ψ angles for the *cis* conformation, at positions *i+1* and *i+2*. β-turns of types VIa-1 ($\varphi_{i+1},\psi_{i+1},\varphi_{i+2},\psi_{i+2}$: -60°,120°,-90°,0°) and VIb-1 ($\varphi_{i+1},\psi_{i+1},\varphi_{i+2},\psi_{i+2}$: -135°,135°(& >100°),-75°,160°) are predominant among the different sub-classes of type VI turns (Pal and Chakrabarti 1999).

The preferences for both amino acid types and backbone conformations associated with *cis* peptides have been analysed but specific and comprehensive studies on the conservation of *cis* conformation have not been carried out. In this study, we analyse the deformations in the planarity of peptide bonds in protein structures with special emphasis on the variations between *cis* and *trans* conformers in similar folds. Using protein family alignments involving high resolution structures, we look at the amino acid preferences and the local conformations associated with such *cis-trans* conversion. A Structural Alphabet namely Protein Blocks (PBs) was used for local structure description (de Brevern et al. 2000; Etchebest et al. 2005; Joseph et al. 2010). This helps to obtain a precise categorization of

pentapeptide backbone conformations into 16 frequent and regular conformations. Only the conserved regions of the alignment are considered for the study.

Methods

Dataset. A set of high quality protein structures solved by X-ray crystallography, with resolution better than 1.6Å and R-factor less than 0.25 is extracted from the PDB. The SCOP domains (version 1.75) corresponding to these structures were identified and all those domains belonging to the same SCOP family were aligned. This resulted in multiple structural alignments of 775 families. The conservation of ω dihedral angles was studied by analysing well-aligned (less than 30% gaps) columns in the alignment.

Protein Blocks. Protein Blocks (PBs) correspond to a set of 16 prototypes of pentapeptide conformations, described based on the ϕ and φ dihedral angles (Supplementary data 1). They were obtained by grouping different pentapeptide conformations observed in protein structures using an unsupervised classifier similar to Kohonen Maps (Kohonen 2001) and hidden Markov models (Rabiner 1989). This structural alphabet allows a reasonable approximation of local backbone conformations (de Brevern et al. 2000) with an average root mean square deviation (*rmsd*) evaluated at 0.42 Å (de Brevern 2005; Joseph et al. 2010). PBs (de Brevern 2005) have been assigned using a in-house Python software as in the previous studies (Gelly et al. 2011; Joseph et al. 2011).

Secondary Structure Assignment. The secondary structure types associated with the PBs were identified with the help of assignments made by PROMOTIF (Hutchinson and Thornton 1996).

Multiple Structural Alignment. A structural alignment approach had been developed based on the representation of protein 3D structures as PB sequences (Gelly et al. 2011; Joseph et al. 2010; Joseph et al. 2011; Tyagi et al. 2008). The protein structural backbone was transformed into PB sequence and these sequences were aligned to obtain a comparison of the 3D structures. Dynamic programming algorithm was used to generate PB sequence alignments (Joseph et al. 2011). Pairwise structure comparison based on PB representation (iPBA) outperformed other popular structural alignments tools in several benchmark tests

(Joseph et al. 2011; Tyagi et al. 2008; Gelly et al. 2011).

The PB based pairwise alignment was extended to perform multiple structure comparisons. A progressive alignment strategy similar to that used in CLUSTALW (Thompson et al. 1994), was adopted for multiple PB sequence alignments. The various different parameters and gap penalties were optimized based on the alignment quality. This method performed better than established tools like MUSTANG (Konagurthu et al. 2006), MULTIPROT (Shatsky et al. 2004), SALIGN (Madhusudhan et al. 2009), MATT (Menke et al. 2008) and MAMMOTH (Lupyan et al. 2005) (Joseph et al, submitted).

Results

The amino acid and conformational preferences associated with *cis-trans* peptide variations were studied using a dataset of alignment of families of high quality structures. Initially, the preferences associated with the *cis* conformation, were studied. Then the sequence and structural features associated with *cis-trans* changes were analysed. PB representation was used to have a reasonable categorization of the backbone conformations associated with such changes.

Figure 1 gives the distribution of the values of ω angles observed in the databank. In the alignment databank, 94% of the peptides have ω angles that varies at most 10^{o} from planarity (180^{o}). Only about 0.31% of all peptides are found to have *cis* conformation with ω close to 0^{o}. The distribution of ω angles was also studied in a larger dataset of structures solved at resolutions under 2.5Å. Only 0.23% of the peptide conformations are in *cis* form.

Cis peptide preferences

Throughout this work, we indicate the two amino acids that form the peptide, as Xaa1 -Xaa2. The preference for each amino acid to occur at these two positions of a *cis* peptide was calculated. At the position Xaa1, aromatic amino acids have a higher preference compared to others. About 12.6% of the position Xaa1 is occupied by Gly (Figure 2A), while 87.6% of the position Xaa2 is occupied by Pro (Figure 2B). About 0.62% and 0.67% of all Tryptophans and Tyrosines are in the first position Xaa1 of a *cis*-bond (Figure 2C). Apart from the aromatic amino acids (W,Y, F), Gly, Ala, Asn, Pro and Cys also prefer to occupy the position Xaa1 (> 0.3%). Aromatic amino acids are reported to stabilize Pro in the *cis* conformation by forming *pi* interactions (Wu and Raleigh 1998).

5.4% of all prolines occur as Xaa2 in the *cis*-conformation (Figure 2D). All other amino acids are under-represented at this position, Tyr has a relatively higher occurance of about 0.12%. For 61% of the *cis* peptides, the solvent accessibility at the second position was higher than the first reflecting a greater exposure at this position (Supplementary data 2).

About 26% of the *cis* peptides occur in β turns of type IV (Figure 3). Other β turns of types VIa1, VIa2 and VIb (that are associated with a *cis*-Pro (Hutchinson and Thornton 1996)) also dominate the *cis* conformation. The conformation at the first position (Figure 3A) is represented by β turns VIa1 and VIb with an occurrences of 13.1% and 16.2% respectively. The coil and extended (β strand) conformations share about 14% each, at this position, which is quite low compared to the background frequencies. At the second postion (Figure 3B), β turns VIa1 and VIb contribute to 12.2% and 21.1% of the *cis* conformations. The strand conformation is less frequent (4.4%) at this position. The preferred conformations were also studied in terms of Protein Blocks (PBs). PBs generally associated with β strands, *a – f*, are preferred in both positions (Figure 3C & D), PB *b*, a specific PB often located after long loop at the entrance of β-strand, has a higher preference for the 2^{nd} position. Loop associated PBs, *g* and *j* are preferred in the first position while PBs *h* and *k* are preferred in the 2^{nd}. Note that the PBs *g* and *h* are frequently found in succession, similarly *j* and *k* share common dihedrals. Supplementary data 3 gives the dihedral angles associated with these PBs.

Cis-Trans Variations

The multiple structural alignments involving high resolution structures were used to study the conservation of *cis* conformation in structurally related proteins. The conserved regions in the alignment (< 30% gaps) were only considered for analysis. The frequency of changes in the ω dihedral was also computed from the alignments (Figure 4). 65% of the ω angles are highly conserved with a maximum variation of only 10^{o}. On the other hand, the *cis* conformation is not well conserved; about 36% of the *cis* peptides prefer exchange with the *trans* conformer.

The change from *cis* to *trans* is indicated by the change Xaa1-Xaa2 to Xaa1'-Xaa2', where Xaa1 and Xaa2 are amino acids linked by the peptide in *cis* conformation and Xaa1' and Xaa2' represents amino acids linked by the peptide in *trans* conformation after the change. It was quite striking to note that nearly 65% of the positions that were associated with changes in the *cis* conformation, are marked with gaps in the alignment.

Based on the residue occupancies at the two positions, *cis* peptides can have different tendencies to go to the *trans* state. To study the relationship of amino acid types with the stability of the *cis* conformation, we calculated the percentage of residues undergoing *cis-trans* variations. Figure 5 gives the percentage propensity of each amino acid (in *cis* peptides) to go to the *trans* conformation in a related protein. At positions Xaa1 and Xaa2, proline has about 31.8% and 33.8% chance of undergoing the change (Figure 5A & B). At position Xaa1, isoleucine and glutamine have the highest tendency for isomerisation (64.3% and 61.6%). Other amino acids like glycine also show high preference of about 47.5%, for conversion (Figure 5A). Trp in the position Xaa1 prefers to conserve the *cis* conformation (only 11.5% change). At position Xaa2, the very low occurrences of non-Pro residues restrict a reliable quantification of the preferences (Figure 5B). Only the changes with at least 10 instances are considered for analysis. Lys and Gly shows higher preferences (47.5% and 55.5% respectively) for the change, at the second position Xaa2. As mentioned above, only 33.8% of Pro undergoes *cis-trans* changes.

Figure 5C & D gives the preferences in amino acid substitutions associated with *cis-trans* variation. Only the conserved alignment positions (no gaps at both positions) are considered for studying the substitution preferences. At the first position (Xaa1 to Xaa1') many of the amino acids prefer replacement with non-polar amino acids like Gly, Ala, Leu and Ile, for undergoing the change. Substitutions with Glu and Pro are also preferred. On the other hand, at the second position (Xaa2 to Xaa2'), most of the amino acids prefer to be conserved (Figure 5C). Pro, which is dominant at this position, is not highly conserved (only 13.3%) upon the *cis-trans* conversion. Some amino acids like Thr, Asp, Gln and His prefer substitutions with Gly upon conversion. Leu and Ile are preferably replaced with Asp or Phe for the *cis-trans* change.

The local conformational changes associated with *cis-trans* changes was also studied (Figure 6). At the first position (Figure 6A), β turns of types VIII and IV are more prone to *cis-trans* change (61% and 41%). However it has to be noted that only a few *cis* peptides occur in type VIII β turns (Figure 3A). β turn VIb and the coil states also have conversion propensities above 30% (38.5% and 34.9% respectively). At the second position (Figure 6B), the coil and strand conformations have conversion frequencies close to 40%. The β turn VIb also show a relatively higher preference (38.3%).

The propensities were also studied in terms of PB changes. At the first position (Figure 6C), PBs *l* and *p* have higher propensities for *cis-trans* change (75.9% and 69.7%

respectively). PB *l* is largely found in the N-cap of helices while *p* is frequent at the C terminus (and at the N terminus of strands). PB *i* mainly associated with coils and has a high dominance of type IV β turns, has a propensity of about 50% for conversion. For the second position on the other hand (Figure 6D), PBs *j* and *a* have a relatively higher preferences with frequencies 54.4% and 46.0% respectively.

On the conserved regions of alignment (no gaps at both positions), the PB changes associated with *cis-trans* change was analysed (Figure 7). At both positions a preference for substitution with a strand associated conformation (PBs *b*, *c* and *d*), was observed. At the first position (Figure 7A), loop associated PBs *i* and *j*, prefer replacement with PB *c* which is frequent at the N terminus of strands. PB *g*, that represents a significant portion of 3_{10} helices, prefers substitution with central helix conformation (PB *m*). This preference is observed at the second position also. At the second position (Figure 7B), PBs *i* and *j*, have greater preferences to get replaced by strand associated PBs, *b* and *d*. Apart from *g*, PBs *h*, *k* and *p* also prefer change with PB *m*, upon *cis-trans* change.

These substitution preferences were compared with the favourable PB changes (substitution with different PBs) observed in all the aligned regions in the dataset. Figure 7C & D gives the distribution of differences in the substitution preferences. It can be seen that many of the preferred changes associated with *cis-trans* variation differ highly from the general preferences for PB substitutions. Also, the largely expected changes involving the replacement of central helix conformation (PB *m*) with the capping regions (*l* and *n*) are under-represented.

Discussion

The *cis* peptides are maintained in protein structures for both structural and functional reasons. Hence these are expected to be well-conserved across the proteins sharing the same structure and function. In this study we have looked into the amino acid and conformational preferences associated with the *cis* state and how these preferences influence the replacement with a *trans* conformer in the structurally equivalent regions. Multiple structural alignments of different SCOP (Murzin et al. 1995) families involving high resolution crystal structures were used for analysis.

SCOP domain families where at least one instance of *cis-trans* change, were extracted (181 families). Nearly 80% of the domain families correspond to enzymes and the rest are also dominated by ligand binding activity (Supplementary data 4). A few cases among these

were looked into.

The structure of Chitinase B holds a non-Pro *cis* bond between Glu 144 and Tyr 145 (numbering taken from the PDB file) in the active site at the centre of the TIM barrel (Figure 8A) (Kolstad et al. 2002). These two residues are involved in direct contact with the substrate (van Aalten et al. 2001) and are also in the vicinity of the dimer interface (Figure 8A). This *cis* conformation is not conserved in Imaginal Disc Growth Factor – 2 (IDGF2) (Varela et al. 2002) which shares a similar TIM barrel fold with chitinase B. However the absence of many active site residues preclude the chitinase activity, the protein is reported to hold the ability to bind carbohydrates but in a different conformation (Figure 8B). Comparison of the two structures (Figure 7A & B) showed that the local PB change at the site of *cis* bond is from PB series *ce* to *ae,* involving variation in β turn type IV conformation. By comparing IDGF2 protein with related folds, Valera and co-workers suggested that the protein may have evolved from chitinases to acquire the new function as growth factors (Varela et al. 2002).

Another set of proteins that have the same fold and bind similar substrates, are the Glycinamide ribonucleotide transformylases and Glycinamide ribonucleotide synthetase enzymes. The transformylase catalyses the formylation of glycinamide ribonucleotide (GAR) using Mg(2+)ATP and formate while the synthetase uses phosphoribosylamine, glycine, and ATP to form glycinamide ribonucleotide. Both enzymes are important in the purine biosynthesis pathway. A *cis* bond between Ala and Pro is found in the vicinity of the active site of the synthetase (Wang et al. 1998) (Figure 8C). This region is at the C terminus of a β strand and is associated with the PBs *eb*. However the *cis* peptide is not conserved in the transformylase enzyme (Thoden et al. 2002) (Figure 8D). The *cis* conformation is replaced by *trans* with a change in the strand conformation (PB change *eb* to *df*). This part is in direct contact with the substrate glycinamide ribonucleotide (GAR). Though both enzymes bind ATP or its derivatives and GAR, and they also have similar binding sites, the catalysis mechanisms are different. The conformation of bound ATP (or ADP) is also different in the two crystal structures (Figure 8C and D). Hence the *cis*-*trans* peptide changes can be implicated in the evolution of new function.

Variation in *cis*-conformation is also observed while comparing the structures of beta keto acyl carrier protein synthase III (Qiu et al. 2001) and HMG (3-hydroxy-3-methylglutaryl) - CoA synthase (Theisen et al. 2004), which have similar folds. Changes in CoA binding sites are also observed (Supplementary data 5) between the two structures. Hence the changes in the *cis* conformation are generally associated with the acquiring of a different function with

variations in substrate binding modes.

Apart from ligand binding and catalysis, *cis* peptides are also involved in intra and inter protein interactions. The structure of Hypoxic Response Protein I (HRP1) (Sharpe et al. 2008) has a cystathionine beta synthase (CBS) domain that is reported to lack AMP binding activity. A novel protein with similar fold from *Thermoplasma acidophilum* (Proudfoot et al. 2008) has an additional N terminal Zn^{++} ribbon domain (Figure 9A). A *cis* proline is observed at the interface between this domain and CBS and this directly takes part in the interaction. The absence of this *cis* peptide loop in HRP1 reflects its functional relevance. It is interesting to analyse the structural or functional importance of a *cis-trans* variation with conserved amino acids. In the structure of hydroxynitrile lyase (from Almond), a *cis* Proline is observed at a Glycosylation site (Figure 9B), when bound to sugar moiety. Though the Proline is conserved at the second position, the structure of related fold, pyranose 2 oxidase (Bannwarth et al. 2006), has a *trans* conformation at this position which is likely to be due to the absence of the glycosylation site.

Non Pro *cis-trans* isomerisation was observed in the structures of myrosinase bound to different substrate analogs (Figure 10A) (Bourderioux et al. 2005; Burmeister et al. 2000). The B-factors associated with this region were not higher compared to the rest of the protein, reflecting a reliable conformational change dependant on the nature of ligand. A similar case was also observed in the structures of quinone reductase 2 where a *cis* conformation was adopted upon substrate binding (Figure 10B) (Calamini et al. 2008). In the absence of the substrate (melatonin), the loop conformation changes from β turn VIb to type I in the *trans* state.

Conclusion

Specific studies have shown that *cis-trans* isomerisation has an important role in the protein folding process. Due to the high energy cost, *cis* peptides accounts for only a negligible percent of the total population. But the ones that are observed in protein structures are seen to be essential for the structural integrity and function. Here we show that the change between *cis* and *trans* conformations are efficiently utilized for the emergence of new function among similar protein folds. Many of the non-Pro *cis* peptides are more prone to such changes than the Xaa-Pro peptides. The local conformational changes associated with the *cis-trans* variations are significantly different from those generally observed in similar structures.

Acknowledgments

These works were supported by grants from the French Ministry of Research, University of Paris Diderot – Paris 7, French National Institute for Blood Transfusion (INTS), French Institute for Health and Medical Research (INSERM) and Indian Department of Biotechnology. APJ is supported by CEFIPRA/IFCPAR number 3903-E. NS and AdB also acknowledge to CEFIPRA/IFCPAR for collaborative grant (number 3903-E).

Figure Legends

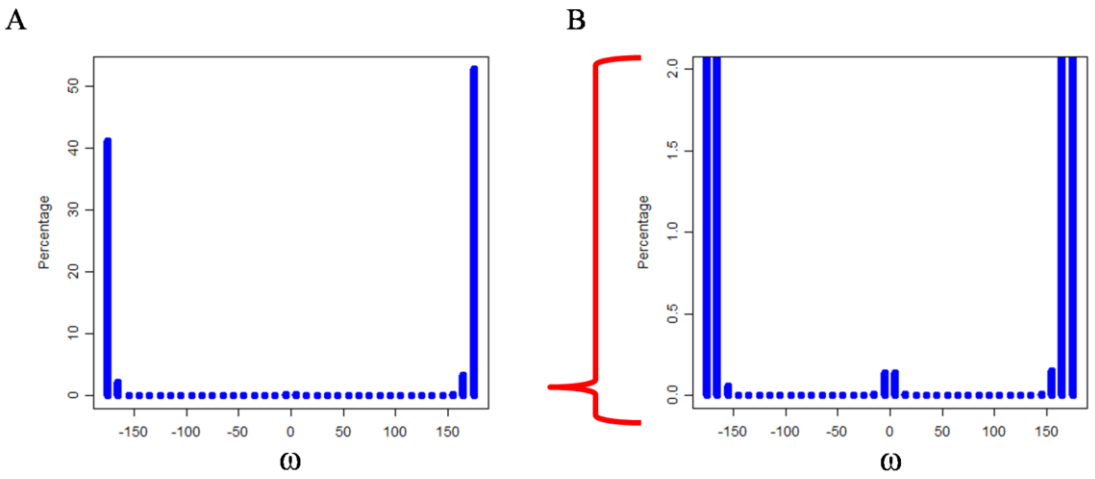

Figure 1. *Distribution of ω angles*. (A) The percentage distribution of ω angles is plotted. The plot on the right (B) zooms the distribution with occurrence less than 2%.

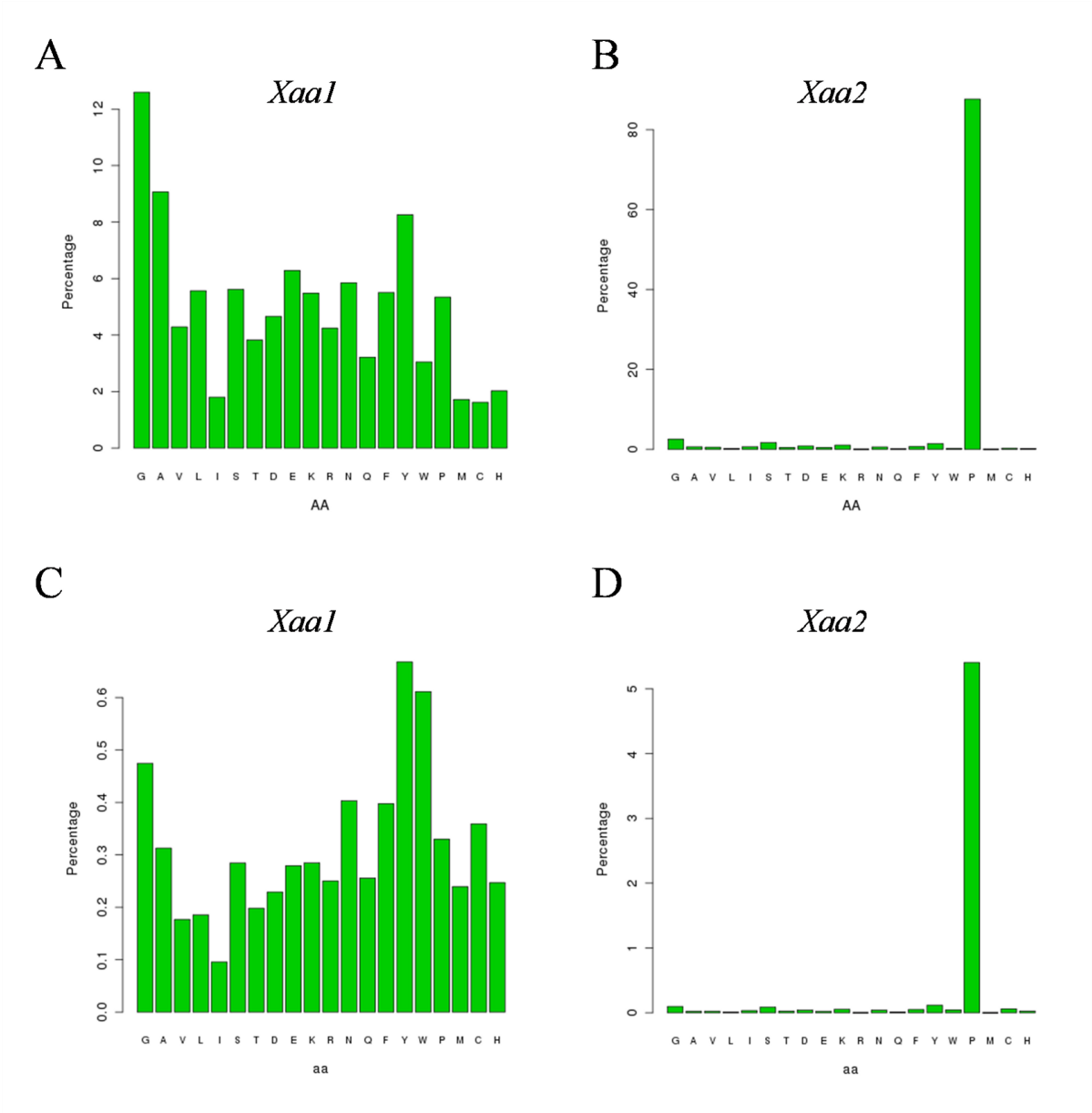

Figure 2. *Amino acid propensity for cis - conformation*. Percentage of each amino acid in the *cis*-conformation at the positions Xaa1 (A) and Xaa2 (B) that forms the peptide bond. For (C) and (D), the percentage is calculated with respect to individual amino acid occurrence frequency.

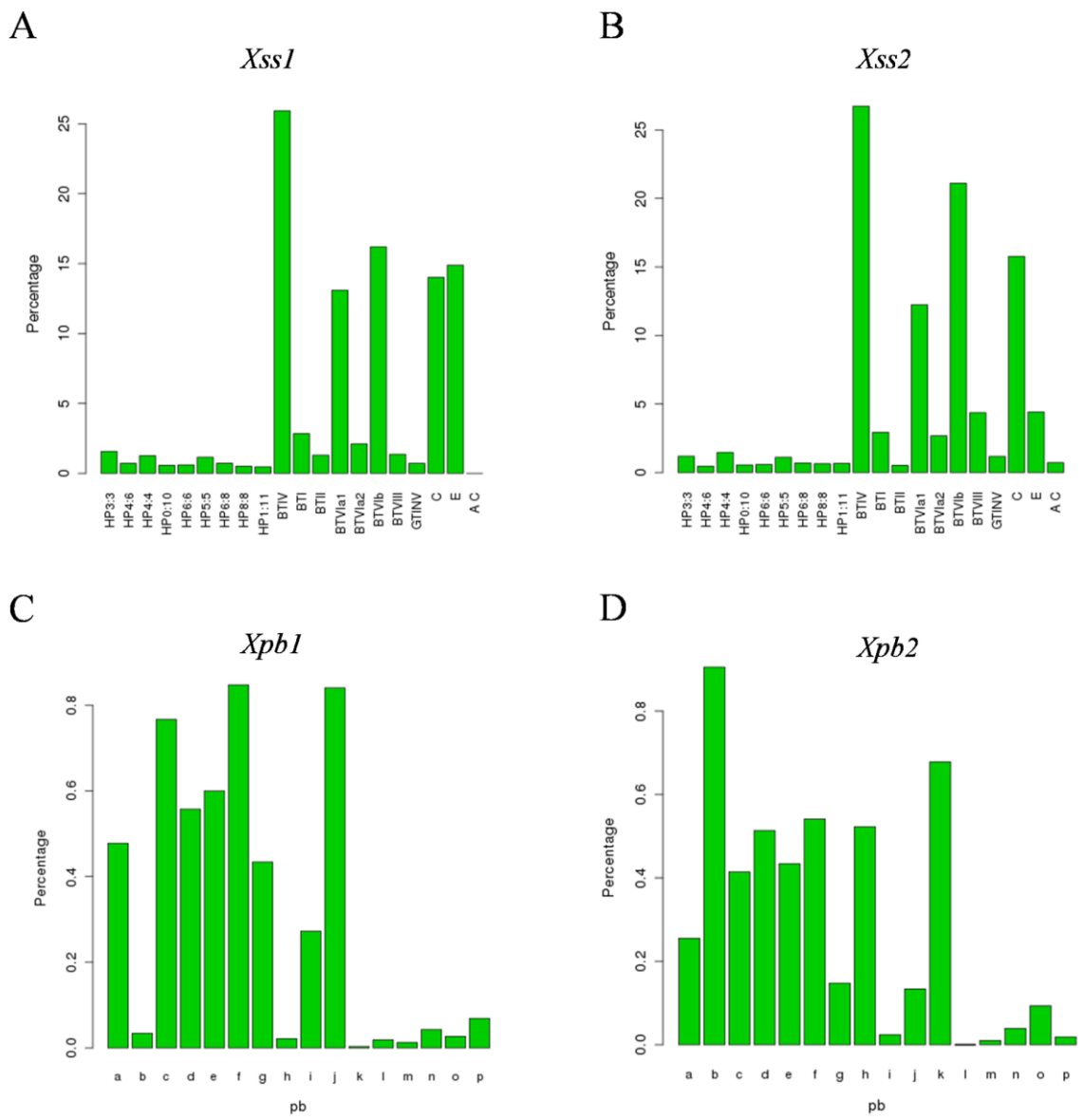

Figure 3. *Local conformational preferences for cis - conformation*. The local backbone conformation was analysed in terms of secondary structures and PBs. PROMOTIF (Hutchinson and Thornton 1996) was used for assignment of the local fold corresponding to the *cis* peptides. Percentage of each type of local fold associated with the first (A) and second (B) positions, are given. Abbreviation of PROMOTIF assignments: BT*X* – β-turns, X is the type of β-turn, AG – Antiparallel strands, G1 type β-bulge, where the first residue is in the left handed helical conformation (usually Glycine), AC – Antiparallel strands, Classic type beta bulge, one extra residue forms the bulge, GTINV – Inverse γ-turns, HP*X:Y* – β-hairpins, X and Y indicate the number of residues in loop, based on two different rules (Hutchinson and Thornton 1996). (C) and (D) indicates the percentage of each PB in the *cis*-conformation at the positions corresponding to Xaa1 and Xaa2. For (C) and (D), the percentage is calculated with respect to individual PB occurrence frequency.

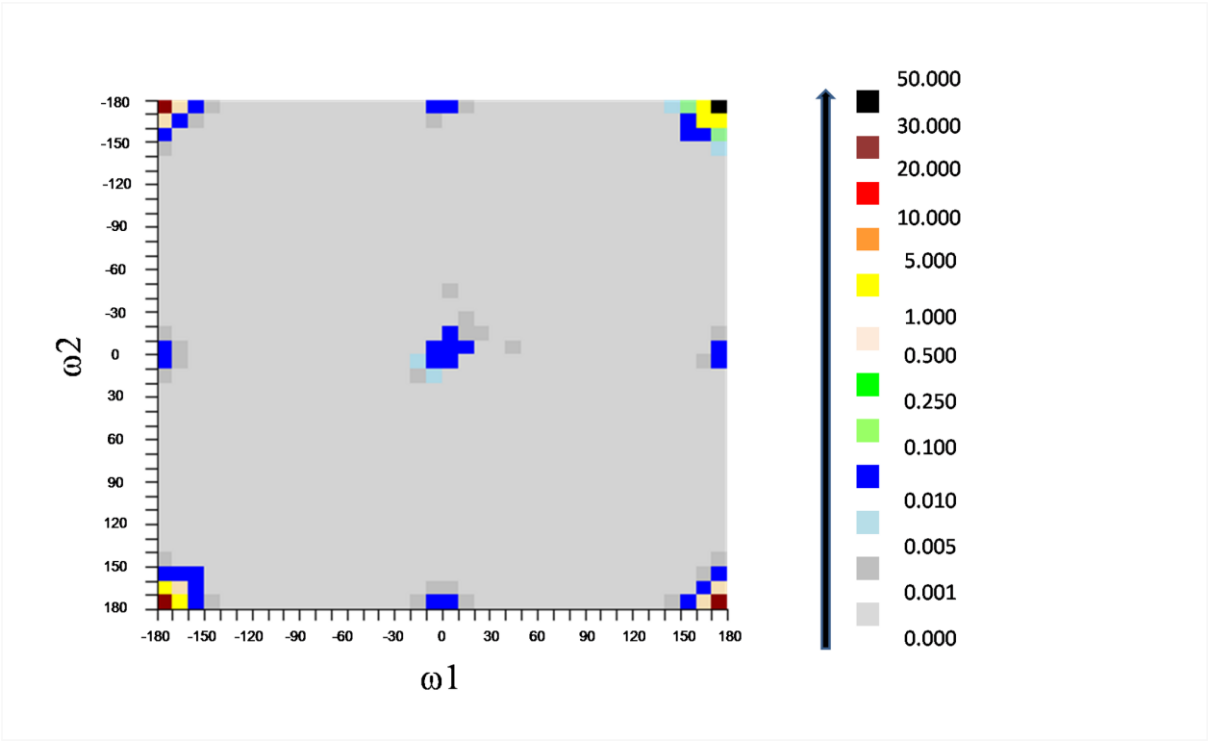

Figure 4. Conservation of ω angles represented by plotting the substitutions observed in ω angles (divided into 10^{o} bins).

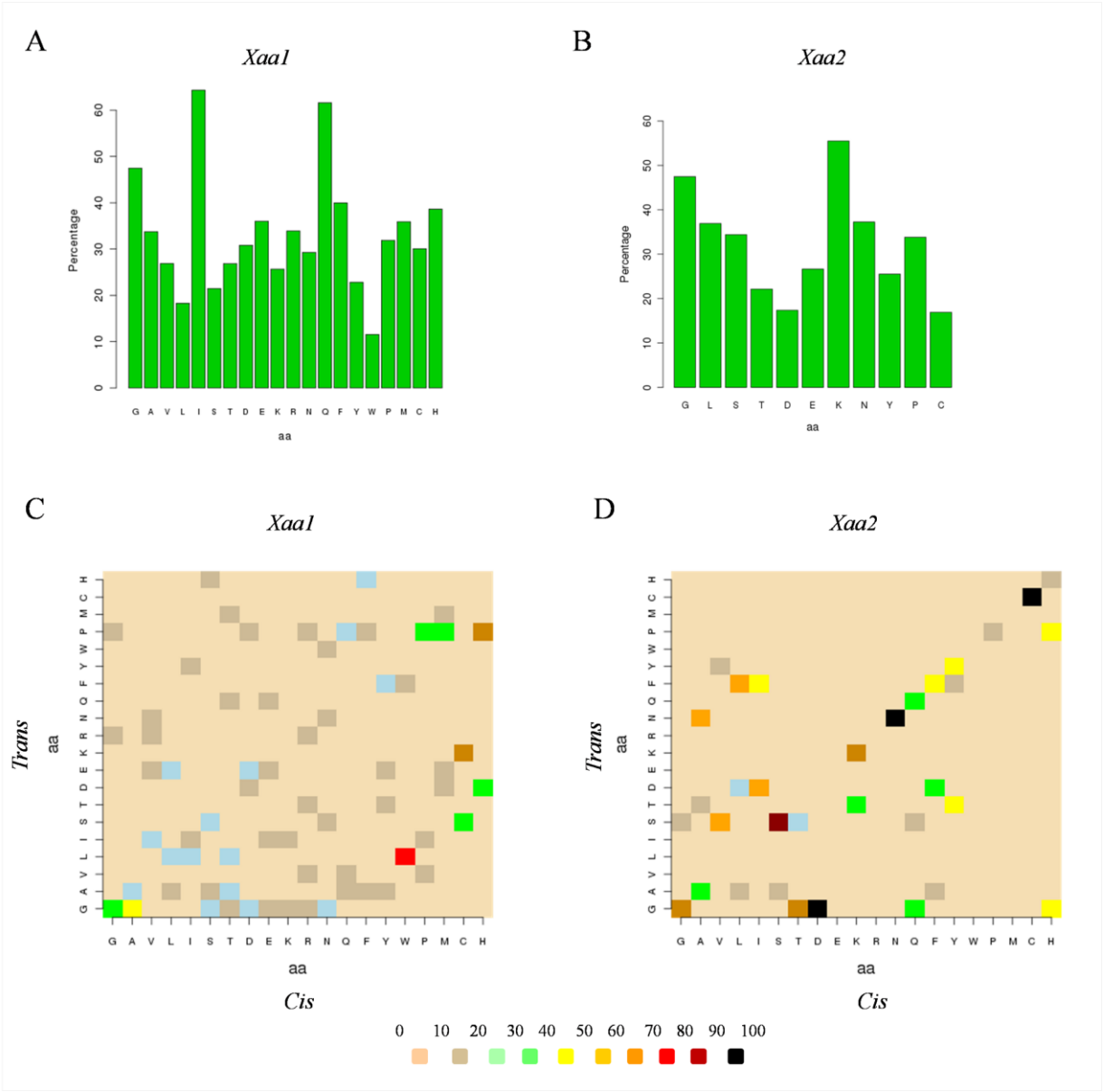

Figure 5. *Conservation of cis state: Amino acid propensity for cis-trans change*. Percentage of each amino acid that participate in *cis-trans* change at the positions Xaa1 (A) and Xaa2 (B). The percentage is calculated with respect to individual amino acid occurrence frequency in cis conformation. (C) and (D) gives the percentage substitution of each amino acid (x axis) (*cis*) by all the 20 amino acids (*trans*), associated with *cis-trans* conversion. The legend gives the substitution percentages; each column adds up to 100%.

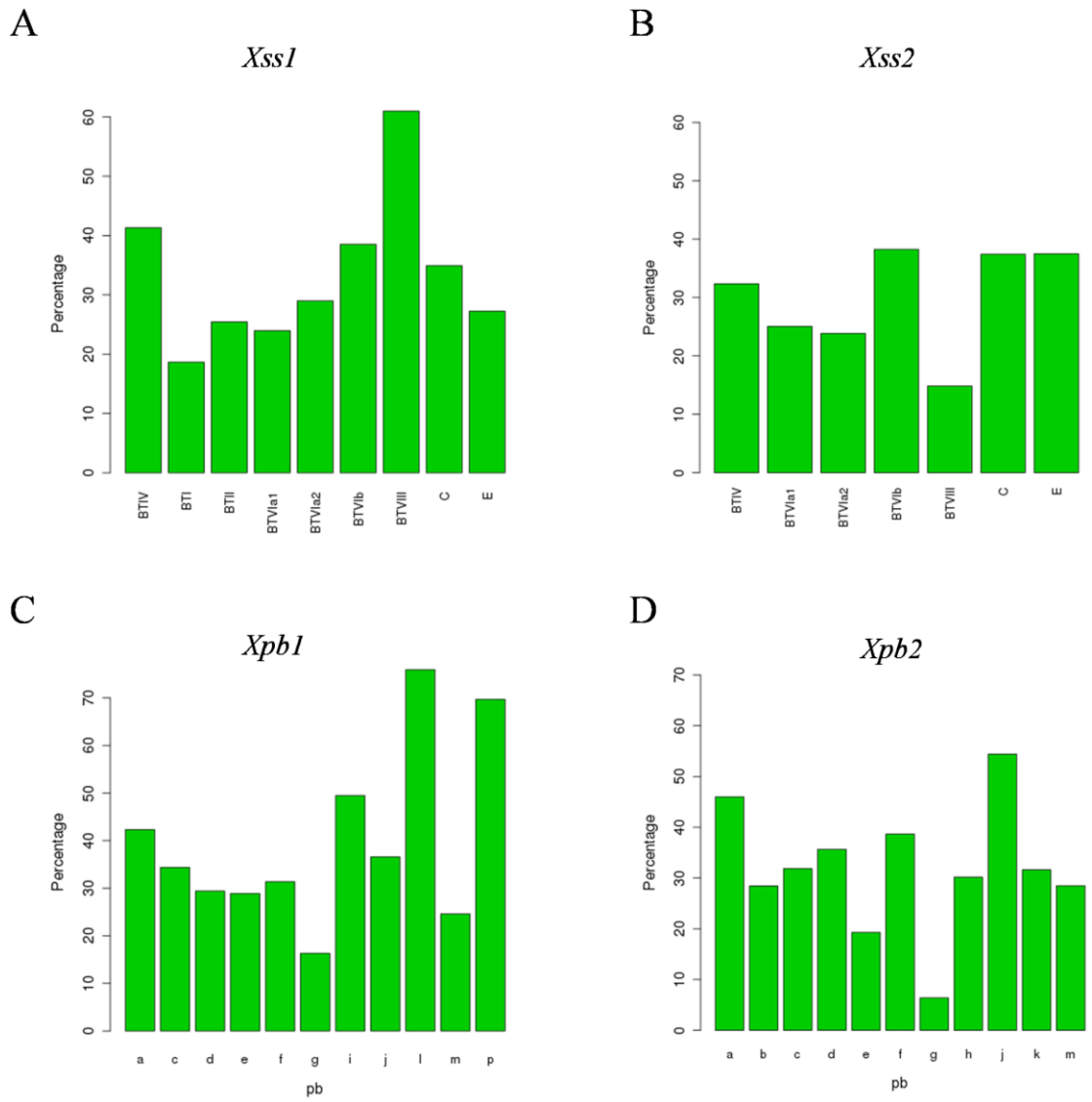

Figure 6. *Local conformational changes associated with cis – trans conversion*. The local backbone conformation was analysed in terms of secondary structures and PBs. PROMOTIF (Hutchinson and Thornton 1996) was used for assignment of the local fold. Percentage of each type of local fold involved in *cis-trans* conversion at the first (A) and second (B) positions, are given. Abbreviation of PROMOTIF assignments: BT*X* – β-turns, X is the type of β-turn, AG – Antiparallel strands, G1 type β-bulge, where the first residue is in the left handed helical conformation (usually Glycine), AC – Antiparallel strands, Classic type beta bulge, one extra residue forms the bulge, GTINV – Inverse γ-turns, HP*X:Y* – β-hairpins, X and Y indicate the number of residues in loop, based on two different rules (Hutchinson and Thornton 1996). (C) and (D) indicates the percentage of each PB involved in *cis-trans* change at the positions corresponding to Xaa1 and Xaa2. The percentage is calculated with respect to individual occurrence frequency in *cis* conformation.

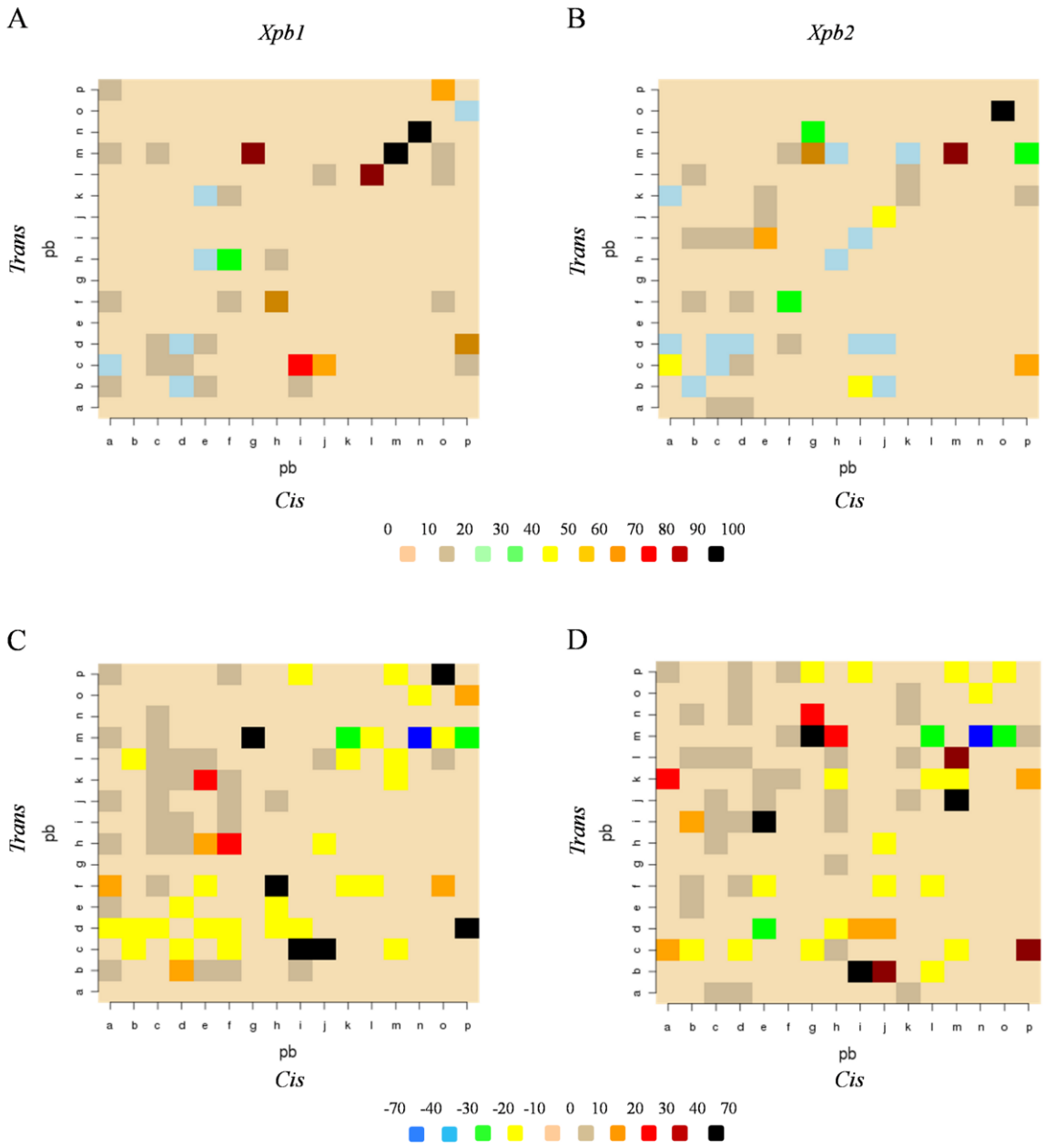

Figure 7. *Preferences for PBs at different positions for cis-trans change*. (A) and (B) gives the percentage substitution of each PB (x axis) (*cis*) with all the 16 PBs (*trans*), associated with *cis-trans* conversion. The legend gives the substitution percentages, each column adds up to 100%. The percentages of non-similar (except the conserved ones) PB changes associated with *cis-trans* change were compared with the general preferences observed in the databank. The differences in these preferences (*cis-trans* - general) are plotted in (C) and (D).

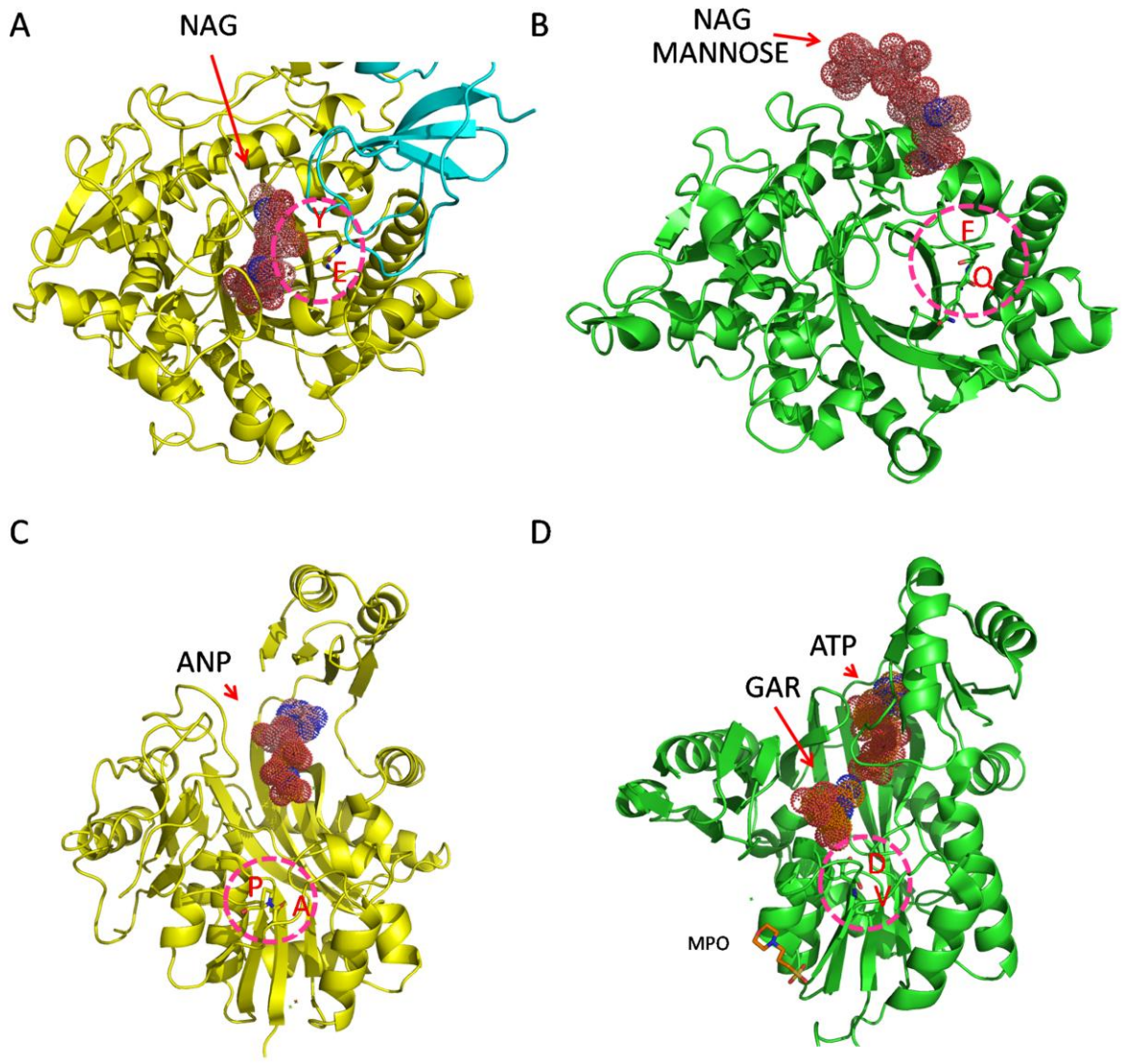

Figure 8. *Evolution of new functions with cis-trans conversions*. (A) The structure of Chitinase B (PDB ID: 1GOI) (Kolstad et al. 2002) (yellow) highlighting the binding site of N-acetyl D-glucosamine (NAG) (orange) (van Aalten et al. 2001). The residues involved in *cis* peptide (EY) are also indicated (red). The other monomer is shown in cyan. (B) The structure of Imaginal Disc Growth factor – 2 (IDGF2) (green) has a similar fold but lack the *cis* peptide (Varela et al. 2002). The sugar moiety (NAG and mannose) bound to the structure is also shown. The residues structurally equivalent to the *cis* peptide are also labelled. (C) The structure of Glycinamide Ribonucleotide Synthetase from *E-coli* (Wang et al. 1998) (yellow) that holds a *cis* peptide (AP) at the active site. The ADP binding site is also highlighted, ANP – Phosphoaminophosponic acid Adenylate ester. (D) The structure of Glycinamide Ribonucleotide Transformylase (Thoden et al. 2002) (green) bound to ATP and Glycinamide Ribonucleotide (GAR).

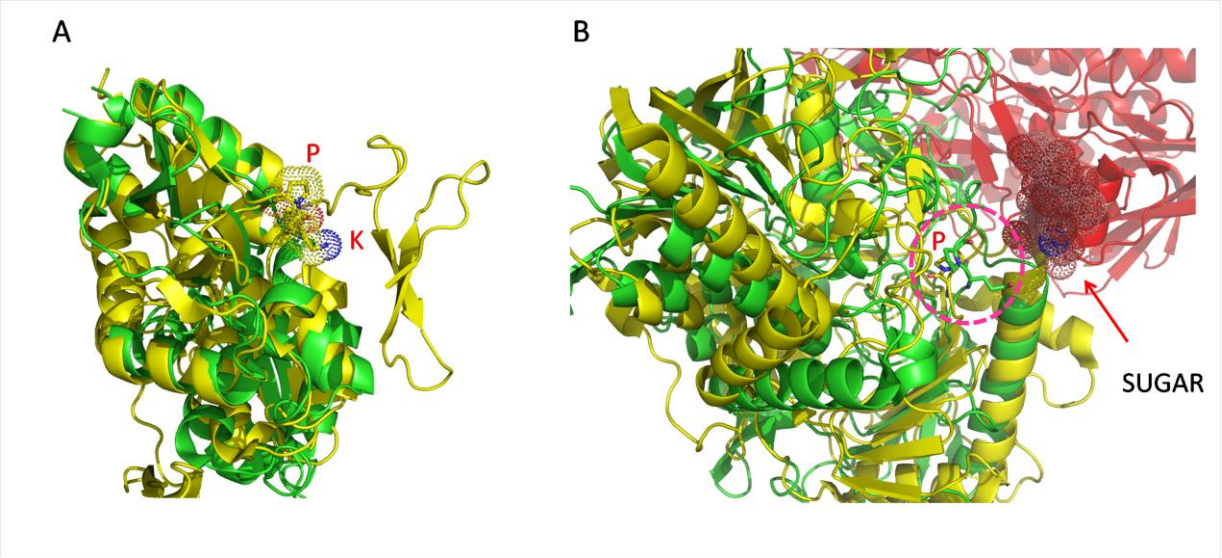

Figure 9. *Structural modifications associated with cis-trans change*. (A) Comparison of structures of Cystathione beta synthase (CBS) domain containing protein (Proudfoot et al. 2008) (yellow, PDB ID: 1PVM) and Hypoxic Response Protein 1 (HRP1) (Sharpe et al. 2008) (green). The location of the *cis* peptide (KP) in 1PVM (yellow), is highlighted. (B) Comparison of structures of Hydroxynitrile Lyase (Dreveny et al. 2001) (yellow) and Pyranose 2 oxidase (Bannwarth et al. 2006) (green). The sugar moiety bound to HRP1 is also highlighted.

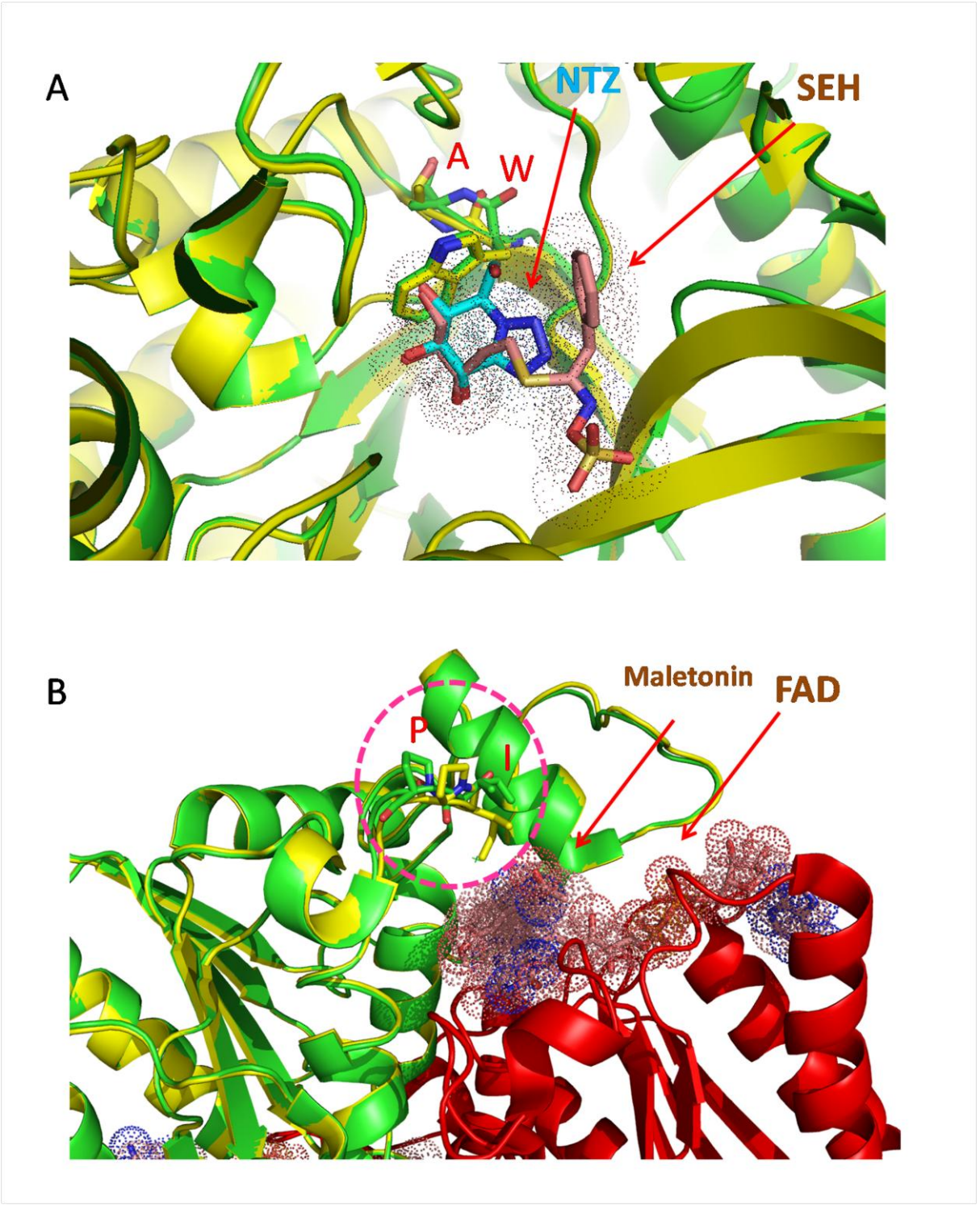

Figure 10. *Role of cis-trans isomerisation in ligand binding.* (A) Comparison of structures of Myrisinases bound to substrate analogs (inhibitors) S-benzyl-phenylacetothiohydroximate-O-sulphate (SEH) (Bourderioux et al. 2005) (yellow) and Nojirimycine Tetrazole (NTZ) (Burmeister et al. 2000). The *cis* peptide (WA) is highlighted. (B) Comparison of Quinone reductase structures (Calamini et al. 2008), one with the *cis* peptide (IP) (yellow) bound to Melatonin and the other (green) unbound at this site.

REFERENCES

Andreotti AH (2003) Native state proline isomerization: an intrinsic molecular switch. Biochemistry 42 (32):9515-9524

Bannwarth M, Heckmann-Pohl D, Bastian S, Giffhorn F, Schulz GE (2006) Reaction geometry and thermostable variant of pyranose 2-oxidase from the white-rot fungus Peniophora sp. Biochemistry 45 (21):6587-6595

Bourderioux A, Lefoix M, Gueyrard D, Tatibouet A, Cottaz S, Arzt S, Burmeister WP, Rollin P (2005) The glucosinolate-myrosinase system. New insights into enzyme-substrate interactions by use of simplified inhibitors. Org Biomol Chem 3 (10):1872-1879

Brandts JF, Halvorson HR, Brennan M (1975) Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. Biochemistry 14 (22):4953-4963

Brauer AB, Domingo GJ, Cooke RM, Matthews SJ, Leatherbarrow RJ (2002) A conserved cis peptide bond is necessary for the activity of Bowman-Birk inhibitor protein. Biochemistry 41 (34):10608-10615. doi:bi026050t [pii]

Burmeister WP, Cottaz S, Rollin P, Vasella A, Henrissat B (2000) High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. J Biol Chem 275 (50):39385-39393

Calamini B, Santarsiero BD, Boutin JA, Mesecar AD (2008) Kinetic, thermodynamic and X-ray structural insights into the interaction of melatonin and analogues with quinone reductase 2. Biochem J 413 (1):81-91

de Brevern AG (2005) New assessment of a structural alphabet. In Silico Biol 5 (3):283-289

de Brevern AG, Etchebest C, Hazout S (2000) Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins 41 (3):271-287

DeLano WLT (2002) The PyMol Molecular Graphics System. Schrodinger, Delano Scientific LLC,

Dreveny I, Gruber K, Glieder A, Thompson A, Kratky C (2001) The hydroxynitrile lyase from almond: a lyase that looks like an oxidoreductase. Structure 9 (9):803-815

Eakin CM, Berman AJ, Miranker AD (2006) A native to amyloidogenic transition regulated by a backbone trigger. Nat Struct Mol Biol 13 (3):202-208. doi:nsmb1068 [pii]
10.1038/nsmb1068

Eckert B, Martin A, Balbach J, Schmid FX (2005) Prolyl isomerization as a molecular timer in phage infection. Nat Struct Mol Biol 12 (7):619-623. doi:nsmb946 [pii]
10.1038/nsmb946

Esposito L, De Simone A, Zagari A, Vitagliano L (2005) Correlation between omega and psi dihedral angles in protein structures. J Mol Biol 347 (3):483-487. doi:S0022-2836(05)00123-3 [pii]
10.1016/j.jmb.2005.01.065

Etchebest C, Benros C, Hazout S, de Brevern AG (2005) A structural alphabet for local protein structures: improved prediction methods. Proteins 59 (4):810-827. doi:10.1002/prot.20458

Fischer G, Aumuller T (2003) Regulation of peptide bond cis/trans isomerization by enzyme catalysis and its implication in physiological processes. Rev Physiol Biochem Pharmacol 148:105-150. doi:10.1007/s10254-003-0011-3

Gelly JC, Joseph AP, Srinivasan N, de Brevern AG (2011) iPBA: a tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Res. doi:gkr333 [pii]
10.1093/nar/gkr333

Grathwohl C, Wuthrick K (1981) NMR studies of the rates of proline *cis-trans* isomerization

in oligopeptides. Biopolymers 20:2623-2633
Grochulski P, Li Y, Schrag JD, Cygler M (1994) Two conformational states of Candida rugosa lipase. Protein Sci 3 (1):82-91. doi:10.1002/pro.5560030111
Guan RJ, Xiang Y, He XL, Wang CG, Wang M, Zhang Y, Sundberg EJ, Wang DC (2004) Structural mechanism governing cis and trans isomeric states and an intramolecular switch for cis/trans isomerization of a non-proline peptide bond observed in crystal structures of scorpion toxins. J Mol Biol 341 (5):1189-1204. doi:10.1016/j.jmb.2004.06.067
S0022-2836(04)00772-7 [pii]
Hutchinson EG, Thornton JM (1996) PROMOTIF--a program to identify and analyze structural motifs in proteins. Protein Sci 5 (2):212-220
Jabs A, Weiss MS, Hilgenfeld R (1999) Non-proline cis peptide bonds in proteins. J Mol Biol 286 (1):291-304. doi:S0022-2836(98)92459-7 [pii]
10.1006/jmbi.1998.2459
Jakob RP, Schmid FX (2008) Energetic coupling between native-state prolyl isomerization and conformational protein folding. J Mol Biol 377 (5):1560-1575
Joseph AP, Agarwal G, Mahajan S, Gelly J-C, Swapna LS, Offmann B, Cadet F, Bornot A, Tyagi M, Valadié H, Schneider B, Cadet F, Srinivasan N, de Brevern AG (2010) A short survey on Protein Blocks. Biophysical Reviews 2:137-145
Joseph AP, Srinivasan N, de Brevern AG (2011) Improvement of protein structure comparison using a structural alphabet. Biochimie 93 (9):1434-1445
Kohonen T (2001) Self-Organizing Maps (3rd edition). Springer,
Kolstad G, Synstad B, Eijsink VG, van Aalten DM (2002) Structure of the D140N mutant of chitinase B from Serratia marcescens at 1.45 A resolution. Acta Crystallogr D Biol Crystallogr 58 (Pt 2):377-379
Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: a multiple structural alignment algorithm. Proteins 64 (3):559-574. doi:10.1002/prot.20921
Lu KP, Finn G, Lee TH, Nicholson LK (2007) Prolyl cis-trans isomerization as a molecular timer. Nat Chem Biol 3 (10):619-629. doi:nchembio.2007.35 [pii]
10.1038/nchembio.2007.35
Lummis SC, Beene DL, Lee LW, Lester HA, Broadhurst RW, Dougherty DA (2005) Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. Nature 438 (7065):248-252. doi:nature04130 [pii]
10.1038/nature04130
Lupyan D, Leo-Macias A, Ortiz AR (2005) A new progressive-iterative algorithm for multiple structure alignment. Bioinformatics 21 (15):3255-3263. doi:bti527 [pii]
10.1093/bioinformatics/bti527
MacArthur MW, Thornton JM (1991) Influence of proline residues on protein conformation. J Mol Biol 218 (2):397-412
Madhusudhan MS, Webb BM, Marti-Renom MA, Eswar N, Sali A (2009) Alignment of multiple protein structures based on sequence and structure features. Protein Eng Des Sel 22 (9):569-574
Menke M, Berger B, Cowen L (2008) Matt: local flexibility aids protein multiple structure alignment. PLoS Comput Biol 4 (1):e10. doi:07-PLCB-RA-0329 [pii]
10.1371/journal.pcbi.0040010
Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247 (4):536-540. doi:10.1006/jmbi.1995.0159
S0022283685701593 [pii]
Nicholson LK, Lu KP (2007) Prolyl cis-trans Isomerization as a molecular timer in Crk

signaling. Mol Cell 25 (4):483-485. doi:S1097-2765(07)00082-2 [pii]
10.1016/j.molcel.2007.02.005
Odefey C, Mayr LM, Schmid FX (1995) Non-prolyl cis-trans peptide bond isomerization as a rate-determining step in protein unfolding and refolding. J Mol Biol 245 (1):69-78
Pal D, Chakrabarti P (1999) Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations. J Mol Biol 294 (1):271-288. doi:10.1006/jmbi.1999.3217
S0022-2836(99)93217-5 [pii]
Pastorino L, Sun A, Lu PJ, Zhou XZ, Balastik M, Finn G, Wulf G, Lim J, Li SH, Li X, Xia W, Nicholson LK, Lu KP (2006) The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-beta production. Nature 440 (7083):528-534. doi:nature04543 [pii]
10.1038/nature04543
Pauling L, Corey RB, Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci U S A 37 (4):205-211
Priestle JP (2003) Improved dihedral-angle restraints for protein structure refinement. JACS 36:34-42
Proudfoot M, Sanders SA, Singer A, Zhang R, Brown G, Binkowski A, Xu L, Lukin JA, Murzin AG, Joachimiak A, Arrowsmith CH, Edwards AM, Savchenko AV, Yakunin AF (2008) Biochemical and structural characterization of a novel family of cystathionine beta-synthase domain proteins fused to a Zn ribbon-like domain. J Mol Biol 375 (1):301-315
Qiu X, Janson CA, Smith WW, Head M, Lonsdale J, Konstantinidis AK (2001) Refined structures of beta-ketoacyl-acyl carrier protein synthase III. J Mol Biol 307 (1):341-356
Rabiner LR (1989) A tutorial on hidden Markov models and selected application in speech recognition. Proceedings of the IEEE 77 (2):257-286
Ramachandran GN, Mitra AK (1976) An explanation for the rare occurrence of cis peptide units in proteins and polypeptides. J Mol Biol 107 (1):85-92
Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. Adv Protein Chem 23:283-438
Reimer U, Fischer G (2002) Local structural changes caused by peptidyl-prolyl cis/trans isomerization in the native state of proteins. Biophys Chem 96 (2-3):203-212
Sarkar P, Reichman C, Saleh T, Birge RB, Kalodimos CG (2007) Proline cis-trans isomerization controls autoinhibition of a signaling protein. Mol Cell 25 (3):413-426. doi:S1097-2765(07)00008-1 [pii]
10.1016/j.molcel.2007.01.004
Scherer G, Kramer M, Schutkowski M, Reimer U, Fischer G (1998) Barriers to rotation of secondary amide peptide bonds. J Am Chem Soc 120:5568-5574
Schmid FX (1986) Proline isomerization during refolding of ribonuclease A is accelerated by the presence of folding intermediates. FEBS Lett 198 (2):217-220
Sharpe ML, Gao C, Kendall SL, Baker EN, Lott JS (2008) The structure and unusual protein chemistry of hypoxic response protein 1, a latency antigen and highly expressed member of the DosR regulon in Mycobacterium tuberculosis. J Mol Biol 383 (4):822-836
Shatsky M, Nussinov R, Wolfson HJ (2004) A method for simultaneous alignment of multiple protein structures. Proteins 56 (1):143-156. doi:10.1002/prot.10628
Stewart DE, Sarkar A, Wampler JE (1990) Occurrence and role of cis peptide bonds in protein structures. J Mol Biol 214 (1):253-260. doi:0022-2836(90)90159-J [pii]

10.1016/0022-2836(90)90159-J
Stoddard BL, Pietrokovski S (1998) Breaking up is hard to do. Nat Struct Biol 5 (1):3-5
Theisen MJ, Misra I, Saadat D, Campobasso N, Miziorko HM, Harrison DH (2004) 3-hydroxy-3-methylglutaryl-CoA synthase intermediate complex observed in "real-time". Proc Natl Acad Sci U S A 101 (47):16442-16447
Thoden JB, Firestine SM, Benkovic SJ, Holden HM (2002) PurT-encoded glycinamide ribonucleotide transformylase. Accommodation of adenosine nucleotide analogs within the active site. J Biol Chem 277 (26):23898-23908
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22 (22):4673-4680
Tyagi M, de Brevern AG, Srinivasan N, Offmann B (2008) Protein structure mining using a structural alphabet. Proteins 71 (2):920-937. doi:10.1002/prot.21776
van Aalten DM, Komander D, Synstad B, Gaseidnes S, Peter MG, Eijsink VG (2001) Structural insights into the catalytic mechanism of a family 18 exo-chitinase. Proc Natl Acad Sci U S A 98 (16):8979-8984
Varela PF, Llera AS, Mariuzza RA, Tormo J (2002) Crystal structure of imaginal disc growth factor-2. A member of a new family of growth-promoting glycoproteins from Drosophila melanogaster. J Biol Chem 277 (15):13229-13236
Wang W, Kappock TJ, Stubbe J, Ealick SE (1998) X-ray crystal structure of glycinamide ribonucleotide synthetase from Escherichia coli. Biochemistry 37 (45):15647-15662
Weiss MS, Jabs A, Hilgenfeld R (1998) Peptide bonds revisited. Nat Struct Biol 5 (8):676. doi:10.1038/1368
Wu WJ, Raleigh DP (1998) Local control of peptide conformation: stabilization of cis proline peptide bonds by aromatic proline interactions. Biopolymers 45 (5):381-394
Wu Y, Matthews CR (2002) A cis-prolyl peptide bond isomerization dominates the folding of the alpha subunit of Trp synthase, a TIM barrel protein. J Mol Biol 322 (1):7-13. doi:S0022283602007374 [pii]
Wulf G, Finn G, Suizu F, Lu KP (2005) Phosphorylation-specific prolyl isomerization: is there an underlying theme? Nat Cell Biol 7 (5):435-441. doi:ncb0505-435 [pii]
10.1038/ncb0505-435