Cis-trans isomerization of omega dihedrals in Proteins

Pierrick Craveur[1,2,3,4,#], Agnel Praveen Joseph[5,#], Pierre Poulain[1,2,3,4,†], Alexandre G. de Brevern[1,2,3,4,+,*]& Joseph Rebehmed[1,2,3,4,+,*]

[1] INSERM, U665, DSIMB, F-75739 Paris, France
[2]Univ. Paris Diderot, Sorbonne Paris Cité, UMR_S 665, F-75739 Paris, France
[3] Institut National de la Transfusion Sanguine (INTS), F-75739 Paris, France
[4] Laboratoire d'Excellence GR-Ex, F-75739 Paris, France
[5]National Centre for Biological Sciences, UAS-GKVK Campus, Bellary Road, Bangalore 560 065, Karnataka, India.

[#]: both first authors contribute equally
[+]: both last authors contribute equally
†: present address: Ets Poulain, Pointe-Noire, Congo.

Short title: cis-trans conformations

* Both corresponding authors:
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

Mailing address: Rebehmed J., 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: joseph.rebehmed@univ-paris-diderot.fr
Tel: +33(1) 44 49 30 91
Fax: +33(1) 47 34 74 31

Key words : protein structures; amino acids; proline; protein folds; steric constraints; secondary structures; protein flexibility; structural alphabet; Protein Blocks; Protein Data Bank, backbone angle.

Abstract

Peptide bonds in protein structures are mainly found in *trans* conformation with a torsion angle ω close to 180°. Only a very low proportion is observed in *cis* conformation with ω angle around 0°. *Cis-trans* isomerization leads to local conformation changes which play an important role in many biological processes. In this paper, we reviewed the recent discoveries and research achievements in this field.

First, we presented some interesting cases of biological processes in which *cis-trans* isomerization is directly implicated. It is involved in protein folding and various aspect of protein function like dimerization interfaces, autoinhibition control, channel gating, membrane binding. Then we reviewed conservation studies of *cis* peptide bonds which emphasized evolution constraints in term of sequence and local conformation. Finally we made an overview of the numerous molecular dynamics studies and prediction methodologies already developed to take into account this structural feature in the research area of protein modeling.

Many *cis* peptide bonds have not been recognized as such due to the limited resolution of the data and to the refinement protocol used. *Cis-trans* proline isomerization reaction represents a vast and promising research area that still needs to be further explored for a better understanding of isomerization mechanism and improvement of *cis* peptide bond predictions.

Introduction

Proteins are major components of all cells. Composed of a sequence of amino acids, they are the support of critical biological functions. Amino acids are composed of a succession of N, C_α, C and O atoms defining the protein backbone while more versatile side-chains give them specific physiochemical specificities. The backbone chain of proteins is constituted of three repeating dihedral angles: the phi-angle (φ) described the torsion angle around the N-Cα bond, the psi-angle (ψ) around the Cα-C bond, and the omega-angle (ω) around the C-N bond – peptide bond. The φ and ψ-angle define the structure of a protein backbone. Due to the delocalization of the carbonyl π electrons and the nitrogen lone pair, the ω angle tends to be planar and therefore is very rigid. From a theoretical point of view, the planar peptide bonds, between the carbonyl carbon atom and the amide nitrogen, in protein structures are more favorable in *trans* conformation as the amide hydrogen of trans peptide bonds offers less steric repulsion to the preceding Cα atom than it does in the *cis* isomers (see Figure 1).

The first protein structures obtained by X-ray crystallography clearly agree with this point (Ramachandran, Sasisekharan 1968). The occurrence of a peptide bond in the *cis* conformation was occasional. Only a specific case had been underlined with the Xaa-Pro peptide bond (Xaa represents any amino acids), where both *cis* and *trans* isomers experience similar steric clashes with the neighboring and are nearly equivalent energetically. Therefore, it was considered that the vast majority of *cis* peptides are observed with a proline. The *cis* conformation for such bond becomes feasible due to a steric interaction between the Cδ-atom of the proline side chain and the backbone atoms destabilizing the *trans* conformer (Stewart et al. 1990,Weiss et al. 1998a). Nonetheless non-proline *cis* bonds are also found in proteins, although they occur much less frequently (Jabs et al. 1999). In a non-redundant set of 571 proteins, 5.2% of Xaa-Pro and 0.03% of Xaa-nonPro peptide bonds were found in *cis* conformation. In random coil polypeptides, the proportion of Xaa-Pro *cis*

bonds varies between 5% and 30%. The size of this population is influenced mainly by the identity of neighboring residues (Brandts et al. 1975,Grathwohl, Wuthrich 1976,Raleigh et al. 1992). Figure 2 presents the *cis* and *trans* conformational isomers of a cysteine-proline peptide bond in the Bovine NPC2 (Nieman-Pick C2) protein (PDB code 1NEP) (Friedland et al. 2003).

Even if they represent a very limited number of residues, the *cis* conformers have biological implications. It was shown for the first time (35 years ago) by Schmid et al. that the *cis-trans* isomerization of peptide bonds on the N-terminal side of a proline plays an important role in the folding process of the protein (Schmid, Baldwin 1978). With the increasing of three-dimensional structures of proteins available today, the importance of *cis* peptide bonds started to emerge. Later on, the importance of the *cis-trans* isomerization process of the peptide bond in protein folding has grown bigger (Levitt 1981,Kiefhaber et al. 1990,Wedemeyer et al. 2002). Additionally, the isomerization reaction is also used as a molecular timer in a number of biological processes, including cell signaling (Sarkar et al. 2007,Wulf et al. 2005), ion channel gating (Lummis et al. 2005) and gene expression (Nelson et al. 2006). Its deregulation is related to pathological conditions, such as cancer (Suizu et al. 2006), amyloid formation (Eakin et al. 2006,Pastorino et al. 2006) and Alzheimer disease(Lu et al. 2007).

In this review, we will underline some characteristic biological examples of *cis-trans* isomerization of omega dihedrals in proteins. Then, we will highlight the interest of structural analyses to conclude with the different *in silico* approaches applied on this subject.

I. IMPORTANCE IN BIOLOGICAL PROCESSES

The *cis-trans* isomerization of omega angles consists of a local conformation change that is often compensated by local variations of backbone angles of the residues flanking the *cis* peptide, thereby avoiding an important global change. Nonetheless, these small changes are sometimes

biologically crucial. Some interesting cases are presented below.

Active sites and dimerization interface. Many *cis* peptide bonds were found to be located at the active sites of proteins and dimerization interfaces. For instance, the recombinant human cellular factor XIII zymogen (Weiss et al. 1998b), which catalyzes the enzymatic cross-linking of fibrin monomers into stable polymers and protects polymers from plasmatic and nonspecific degradation, has two non-proline *cis*-peptide bonds in its crystal structure. One is between Arg310 and Tyr311 close to the active site and the other is between Gln425 and Phe426 at the dimerization interface (see Figure 3). In the same way, the structure of Chitinase B, an hydrolytic enzyme that breaks down glycosidic bonds in chitin, holds a non-proline *cis* bond between Glu144 and Tyr145 in the active site; these two residues are involved in direct contact with the substrate and are also in the vicinity of the dimer interface (van Aalten et al. 2001).

Modulation of signaling - Autoinhibition control. Autoinhibition is a natural control of protein activities and is typically regulated by extrinsic factors. For instance, the Kit kinase is autoinhibited through an intramolecular interaction with the juxtamembrane (JM) domain. Kit JM behaves as an autonomously folding domain that binds directly to the amino-terminal lobe of the Kit kinase and prolongs its induction time to activation (Chan et al. 2003).

Another example is the adaptor molecule Crk that binds to several tyrosine-phosphorylated proteins. It is involved in several signaling pathways, recruiting cytoplasmic proteins in the vicinity of tyrosine kinase through src-homology domains 2 -phosphotyrosine interaction. Recently, Sarkar and collaborators (Sarkar et al. 2007,Sarkar et al. 2010) showed that autoinhibition of Crk can be also controlled by an intrinsic intramolecular switch afforded by a *cis-trans* isomerization. A proline of the linker tethering the two SH3 domains of the Crk adaptor protein interconverts between the *cis*

and *trans* conformation (see Figure 4). In the *cis* conformation, the two domains interact intramoleculary leading to autoinhibition. The isomerization facilitates the interactions between domains of the same protein. Conversely, in the *trans* conformation Crk exists in an extended, uninhibited conformation that serves to activate the protein upon ligand binding. This interconversion between the *cis* and *trans* conformations can be accelerated by the action of peptidylisomerases. Authors concluded that proline isomerization make an ideal switch that can regulate the kinetics of activation, thereby modulating the dynamics of signal response.

Channel gating. The 5-Hydroxytryptamine type 3 receptor (5-HT_3R) is a transmembrane protein, a G-protein-coupled one. It belongs to the superfamily of ligand-gated ion channels and is closely related to the nicotinic acetylcholine receptor. Composed of 5 subunits arranged around a central ion conducting pore, it is permeable to potassium and calcium ions and is well known to bind serotonin leading to an excitatory response in neurons.

Lumnis and co-workers (Lummis et al. 2005,Melis et al. 2009) suggest that *cis-trans* isomerization of a single proline provides the switch that interconverts the open and closed states of the channel. A model has been proposed for 5-HT_3R gating whereby *trans−cis* isomerization of a specific proline (Pro308) at the apex of the M2−M3 loop would induce a conformational change in M2, resulting in a displacement of the hydrophobic zone that occludes the pore and opening the channel. A series of proline analogues was incorporated at the position 308 by mutagenesis. Proline analogues that strongly favor the *trans* conformation produced non-functional channels. Molecular rearrangement, after ligand binding, of the protein backbone at Pro308 from the *trans* conformation to the *cis* conformation can be the structural mechanism that opens the receptor pore. Figure 5 underlines this particular position.

Proline has unique structural and conformational properties, and it is found with anomalously high frequency in the transmembrane regions of ion channels and transporters, suggesting a key role

in structural changes associated with transmembrane signaling (Brandl, Deber 1986,Sansom, Weinstein 2000).

Membrane binding.The winged helix gene Brain factor-1 (BF1 or Forkhead box protein G1, FOXG1) has a pleiotropic role in the development of the cerebral hemispheres of the brain. Mice lacking BF1 have defects in the morphogenesis of the structures of the dorsal and the ventral telencephalon with important consequences, *e.g*. for development of the vertebrate olfactory system. This 489 residues protein length binds membrane. In the process of phospholipid binding, Pro22 of BF1 may undergo, in the presence of calcium ions, isomerization from the *trans* to *cis* conformation in a step essential for membrane binding(Evans, Nelsestuen 1996).Perera and co-workers used molecular dynamics (MD) simulations to investigate the relative importance of the two conformations of Pro22 to the structural and dynamical properties of BF1(Perera et al. 1998).They underlined that the overall structural changes associated with the *trans-cis* isomerization is negligible; the calculated electrostatic potential energy surfaces is very similar and cannot explain the binding specificities. Only minor modifications to the hydrogen bond network take place.

Protein stability. It was also observed that *cis-trans* isomerization of the peptide bond can impact proteins stability. Truckses and collaborators encounter an example with the *Staphylococcus aureus*nuclease. This enzyme is an endo-exonuclease that preferentially digests single-strandednucleic acids. They observed a mixture of *cis* and *trans* configuration of the Lys116-Pro117 peptide bond. In terms of free energy, they measured that approximately 17% of the increase in protein stability manifests itself as stabilization of the *cis* configuration at Lys116-Pro117(Truckses et al. 1996).

II. Analyses and conservation of cis-trans peptide bonds in proteins

Analyses. As *cis* conformations are occasional, they are expected to occur for specific biological and physical reasons; they would also be more conserved. MacArthur and Thornton calculated the ω peptide angle distribution in the Cambridge Structural Database of small molecules (Allen et al. 1979).They found only 54 on 1712 peptides with *cis* conformation. They extended their approach to a non-redundant set of proteins from the Brookhaven protein database (Bernstein et al. 1977) and they found out that peptide bonds in the *cis* conformation represented less than 0.1% of the peptide bonds in their dataset(MacArthur, Thornton 1996).

Two years later Weiss and co-workers analyzed specifically the *cis*-peptide bonds (Weiss et al. 1998a). They used a non-redundant set of 571 X-ray protein structures taken from the Brookhaven protein database (Bernstein et al. 1977) with a resolution of 3.5 Å or better. *Cis* conformation represented 0.28% of all omega angles. They underlined that most of the peptide bonds in *cis*conformation occur where the peptide bond is an imide (Xaa-Pro) rather than an amide bond (Xaa-nonPro). Authors also discussed the discrepancy between the fraction of *cis* peptide bonds observed and what can be predicted from free enthalpy values. One of the main results of this study was the significant correlation between the resolution of the structure solved and the number of *cis* peptides detected. High resolution structures (resolution < 2.0 Å) contain almost twice the number of Xaa-Pro bonds than medium and low resolution structures (resolution ≥ 2.5 Å) and almost four times the number of Xaa-nonPro bonds.

This conclusion brought our attention to the fact that many *cis* peptide bonds may have passed unnoticed due to an automatic usage of refinement programs (Huber, Steigemann 1974). Moreover, Gunasekaran and co-workers showed that half of disallowed regions residues (often due to *cis*-conformation) in the Ramachandran plot are linked to biological functions (Gunasekaran et al.

1996). These unnoticed *cis* peptide bonds could impact the understanding of many protein functions mechanisms.

Evolution.Due to their structural and functional importance,cis peptide bonds in proteins are expected to be conserved during evolution. The first conservation study was made on a set of non-homologous and high resolution structures (sequence identity <25% and resolution ≤2.0Å) (Lorenzen et al. 2005). The set contained 1729 proteins corresponding to 742 *cis* and 14,502 *trans* prolyl residues. They show, as previously reported (Pal, Chakrabarti 1999), that position preceding cis-proline is strongly occupied by the aromatic amino acids tryptophan and tyrosine, and by small residue glycine, as well as proline itself. Phenylalanine and tyrosine, but not tryptophan, are the most present in the succeeding position of the prolyl residue. This predominance for aromatic and small residues at neighboring positions of *cis*-proline could be respectively explained by π interactions between the aromatic and the proline rings, and by the lack of steric hindrance (Wu, Raleigh 1998).Authors analyzed the impact of local and global sequence homology and showed that (i) *cis*prolyl residues are more often conserved than *trans* prolyl residues, (ii) *cis* and *trans* prolyl residues are more conserved than the surrounding amino acids, and (iii) neighboring amino acids have a direct influence on the probability of forming a *cis* prolyl residue.

The residue conservation in term of structural homology is also an important feature. It was firstly analyzed to illustrate the usefulness of structure alignments database to improve comparative protein modeling(Sali, Overington 1994). The underlying idea is that restraints of *cis/trans* isomerism in a modeled sequence could be provided by homologous structures. Despite the small number of example in the used dataset (238 *cis*-prolines), the authors estimated that the knowledge of equivalent conformations increases the prediction success from 0% to 82.9% for *cis*-proline and from 93.3% to 96.2% for *trans*-proline, compared with the only overall stereochemical preference of proline taken into account. The same conclusion was also obtained by structurally aligning

proteins from same SCOP (Murzin et al. 1995) superfamily (Lorenzen et al. 2005). The best decision about the presence of a *cis* prolyl residue can be made by modeling according to the highest resolution structure with more than 20% sequence identity to the query protein.

A very recent study consisted of analyzing the preferences of amino acid types and local backbone conformations associated with *cis* peptide bonds. In this study, a larger set of high quality protein structures solved by x-ray crystallography with a resolution better than 1.6 Å was used(Joseph et al. 2012). Authors emphasized the variations between *cis* and *trans* conformers in structurally similar proteins. They described the local conformations associated with these peptides using a Structural alphabet named Protein Blocks (PBs) (de Brevern et al. 2000,Etchebest et al. 2005,Joseph et al. 2010). PBs are local approximation of protein structures and are useful to analyze local protein conformations(see for instance(Jallu et al. 2012)). Considering the Xaa-Proline *cis* bonds, 34% are not conserved in structural homologous proteins; and proline has a high tendency to get replaced by another amino acid in the *trans* conformer (87%). Authors showed 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-Proline *cis* peptides are more prone to such changes than the Xaa-Pro peptides(Joseph et al. 2012). This study also underline the fact that aligned structural regions can easily have *cis* and *trans* conformation in front of each other despite the huge change of omega angles; in most cases, the other angles compensate the local change in the backbone directions.

III. In silico

Prediction of cis-trans isomerization in proteins. In regards to other related research areas, the number of studies aiming to predict the peptide bond conformation between amino acids is limited. The main limitation of these approaches is that the fraction of peptide bonds adopting a *trans*

conformation is 350 times more important than the *cis* conformation, and even 20 times more for the prolyl peptide bond. Therefore, the prediction tools of *cis-trans* isomerization focus exclusively on Xaa-Pro *cis* peptide bonds.

The beginning of the 90's witnessed the first attempt to predict peptide bond conformation of prolines in proteins. The methodology consisted of determining sequence patterns flanking a centered proline(Frommel, Preissner 1990). Authors considered windows up to the sixth residues around the proline, and their physicochemical properties. They identified six different patterns to discriminate between the *cis* and *trans* conformations. They correctly predicted about 73% of *cis* prolyl residues and did not predict any proline in *trans*-conformation as *cis* conformation. Nonetheless, the dataset used was too restricted with only 242 Xaa-Pro cases to make reliable prediction on larger dataset.

COPS algorithm (Pahlke et al. 2005a,Pahlke et al. 2005b) was the first to predict the peptide bond also for Xaa-nonPro bonds. The algorithm is based on an extension of the Chou-Fasman parameters (Chou, Fasman 1974). It predicts conformation of the peptide bond by taking into account only the secondary structure of amino acid triplets. Despite a large dataset (8,584 proteins yielded 25,663 amino acids in *cis* conformation and ~ 11 million amino acids in *trans* conformation), a sensitivity of 35% and an accuracy of 66% under lied the difficulty of the approach.

Many methods are based on Support Vector Machine (SVM) (Vapnik 1999). The first study of prolyl *cis*/*trans* isomerization took into account amino acid properties such as secondary structure prediction, solvent accessibility prediction and the physicochemical properties of the surrounding amino acids (Wang et al. 2004). They used larger dataset than previous studies (2,193 structures) and reported a better success rate. However, no software or webserver had been available making it difficult to assess its real performance as the validation steps seem biased.

More interestingly, Song and co-workers proposed another method based on SVM in 2006

(Song et al. 2006). It used the amino acid compositions of local sequence flanking centered proline residues (different window sizes were tested), the position specific scoring matrix (PSSMs) extracted from PSI-BLAST(Altschul et al. 1997) and the predicted secondary structure extracted from PSIPRED (Jones 1999). During this study, different kernel functions and parameters were tested for the SVM training and testing to unbias the prediction. Their method yielded a prediction accuracy of 71.5% and a Matthews Correlation Coefficient (MCC)(Matthews 1975) of 0.43. Authors also compared their newly developed method with those published (Frommel, Preissner 1990,Pahlke et al. 2005b) (see Table 6 of (Song et al. 2006)). Authors pointed the fact that the method accuracy depends on the training set size, and mentioned that they have also developed a webserver named CISPEPpred (http://sunflower.kuicr.kyoto-u.ac.jp/~sjn/cispep/).

Exarchos and co-workers presented another method to predict all types of *cis* peptide bonds. It is related to previous studies using evolutionary profiles, secondary structure information, solvent accessibility predictions for each amino acid and the physicochemical properties of the surrounding residues with SVMs. The methodology was more complex than the previous ones for the selection of features. They also used a large dataset of high quality proteins (3,050 structures) with resolution < 2.0Å and sequence identity < 25%. They made distinction of the four peptide bonds *cis*-Pro, *cis*-nonPro, trans-Pro and *trans*-nonPro and also developed the PBOND web server (that it currently no more available) (Exarchos et al. 2009a). They reported a sensitivity up to 77% and an accuracy up 74%.The extracted patterns were also compared against the PROSITE database, in order to gain insight into the implications of *cis* prolyl bonds in protein functions (Exarchos et al. 2009b).

Molecular dynamics simulations. The usefulness of molecular dynamics simulation techniques for the study of *cis-trans* isomerization of proline has been demonstrated in many studies. A first interest in computational approach was given by Levitt in the beginning of the 80's (Levitt 1981).He performed conformational energy calculations to study the role of the proline residues in

the folding of the bovine pancreatic trypsin inhibitor (BPTI) by forcing the *trans-cis* isomerization of 4 prolines. It underlined the existence of a potential alternative folded native-like conformation for either isomeric form when the *cis* form is calculated to destabilize slightly the folded structure.

Another example was presented later by Ikura and co-workers(Ikura et al. 1997). Using the *staphylococcus aureus* nuclease (SNase), they studied the urea-induced unfolding transition and five proline mutants, coupling experimental methods (circular dichroism, absorption spectroscopy, refolding−unfolding kinetics of the proteins by stopped-flow circular dichroism and stopped-flow absorption techniques) and *in silico* approaches (molecular dynamics). Two phases were observed in the unfolding of the wild-type and mutant proteins (that contained a specific Pro117); a fast phase corresponding to the unfolding of the *trans* isomer and a slow phase corresponding to that of the *cis* isomer. On the basis of these results, they also discussed the folding scheme of SNase.

Hodel and co-workers obtained thermodynamic and structural description of the conformational equilibrium of a particular protein loop. Authors used an exhaustive conformational search that identified several substrates followed by free energy simulations between the substrates (Hodel et al. 1995). These simulations correctly predicted a small free energy difference between the *cis* and *trans* forms composed of larger, compensating differences in enthalpy and entropy. The structural predictions of these simulations were qualitatively consistent with known X-ray structures of nuclease variants.

Another interesting approach was developed by Darve and co-workers with an adaptive biasing force molecular dynamics simulation (Darve et al. 2008).This approach was used to force *cis-trans* isomerization at Pro32 of β2-microglobulin and to calculate the relative free energy in the folded and unfolded state (Fogolari et al. 2011). Authors carried clockwise and anticlockwise rotations to sample the whole -180:180 interval of the ω angle between His31 and Pro32 to generate the energy profile. They underlined clearly that *cis* state is favored in the native structure.

In our laboratory, we have also observed the effect of *cis* and *trans* conformation on protein

dynamics of the platelet integrin αIIbβ3. Composed of two sub-units, it is implicated into fibrillar aggregation and responsible of alloimmune thrombocytopenia or Glanzmann's thrombasthenia. Through *in silico* approaches, we showed the impact of punctual mutation (Jallu et al. 2012,Jallu et al. 2010). We also studied the mutation S163P of the β3 subunit (see Figure 6). As no observation is known about the local conformation of the Pro163 residue, we tested both *cis* and *trans* conformations with long MD simulations (see (Jallu et al. 2012)) for technical methodology and results). We observed that both isomers are highly rigid; they do not locally move during the different simulations, but due to the steric constraints they also keep distinct local conformations, *i.e.* distinct PBs series (Poulain, Jallu and de Brevern, *unpublished*).

QM/MM studies.QM/MM methods were applied to study the electronic forces that facilitate the peptidyl *cis* to *trans* interconversion, because conventional force fields based on classical mechanics do not take into account the effect of electronic state changes. The general mechanism of peptide bond isomerization involves a change of the bond order from a partial double bond to a single bond, followed by a return to partial double bond character. This pyrrolidine N changing from a planar sp^2 state to a pyramidal sp^3 state allows the rotation (Yonezawa et al. 2009,Kang 2006). By cons, QM cannot be applied to whole proteins systems because the computational cost exceeds current resources.

Umbrella sampling methods were used to explore the free energy landscape during the isomerization of a dipeptide (Yonezawa et al. 2009). It showed that the *trans* state is more stable than the *cis* state by 4 kcal/mol and the high energy barrier separating the two states is about 20 kcal/mol.

Force-clamp molecular dynamics (FCMD) simulations were used to accelerate the interconversion from *cis* to *trans* conformation of a tetrameric peptide (AAPA) (Chen et al. 2012). Its termini Cα atoms were subjected to a constant force in opposite directions. Authors observed the

variation of different order parameters related to the isomerization of the peptide bond in respect to the simulation time: torsion angle of the peptide bond, distance between the Cα atoms, and length of the peptide bond. Force releases the partial double bond and converts it to a single bond. Rotation around the peptide bond can then occur and afterward the partial double bond reforms in the *trans* conformation. Their study demonstrated that *cis* to *trans* isomerization can be triggered by mechanical force.

QM/MM MD simulations in combination with mean reaction force (MRF) (Vohringer-Martinez, Toro-Labbe 2011) were applied to Pin1, an enzyme that specifically catalyzes the isomerization of peptide bonds between phosphorylated threonine or serine residues and proline. Authors wanted to discern the catalysis influence of Pin1 enzyme in the isomerization (Vohringer-Martinez et al. 2012). They observed the influence of the enzyme and identified structural and electronic contributions to activation barriers and reaction free energies for the isomerization. They carried out simulations of the peptide in two different environments: in solution and in the enzyme. The enzyme induces a rotation of the ψ angle of the preceding residue to avoid unfavorable interactions with the rotating part of the peptide.

CONCLUSION

The *cis-trans* isomerization and conformations are well known from researchers involved in protein structures analyses. Nonetheless, as they are quite non frequent, they are not often considered. This review emphasizes two points: first, the biological importance of the *cis* conformation of peptide bonds in protein structures and functions; and second, the possibility that many *cis* peptide bonds have not been recognized as such due to the limited resolution of the data and to the refinement protocol used.

Even if the number of Xaa-Pro peptide bonds in the *cis* and *trans* conformations are quite

unbalanced, it is possible to observe some general preferences concerning the preceding amino acid types, with aromatic and small residues favoring the *cis* isomer slightly for example. In the same way, pertinent methodology for cis peptide bonds prediction seems to give quite correct prediction rate in regards to the difficulty. Accurate prediction of *cis-trans* isomerization in proteins would have many important applications towards the understanding of protein structure and function. Nonetheless, an effort is still for an accurate prediction of the Xaa-nonPro *cis* peptide bonds.

An interesting observation was done on structural homologues that are well superimposed, but share different cis / trans conformation. In an important number of cases, no significant difference can be observed in terms of structural local conformations (Joseph et al. 2012). The difference of ω angles is compensated by the neighboring dihedrals.

It has been shown, that this local similarity could be associated with biological specificity. For instance, Glycinamideribonucleotide (GAR)transformylases and GARsynthetase have the same fold and bind similar substrates. Both enzymes bind ATP (or its derivatives) and GAR, but the conformation of bound ATP (or ADP) is different in the two crystal structures with *cis* and *trans* isomers. They also have similar binding sites, but the catalysis mechanisms are different. Hence, the *cis–trans* peptide changes can be implicated in the emergence of new functionamong similar protein folds during evolution (see (Joseph et al. 2012)).

Structure-based *in silico* approaches provide further details for the explanation of protein functions and specific mechanisms. Due to the computation time, most of the QM/MM studies were made on small peptides and proposed explanations may not be correct on a real protein where extra constraints can be applied on the peptide bond. *Cis-trans* proline isomerization reaction represents a vast and promising research area that still needs to be further studied.

ACKNOWLEDGEMENTS

This work was supported by grants from the Ministry of Research (France), University Paris Diderot, Sorbonne Paris Cité (France), National Institute for Blood Transfusion (INTS, France), National Institute for Health and Medical Research (INSERM, France) and "Investissements d'avenir", Laboratories of Excellence GR-Exto APJ, PC, JR and AdB, (France);HFSP to APJ; and ANR NaturaDyRe (France, ANR-2010-CD2I-014-04) to JR.

FIGURES

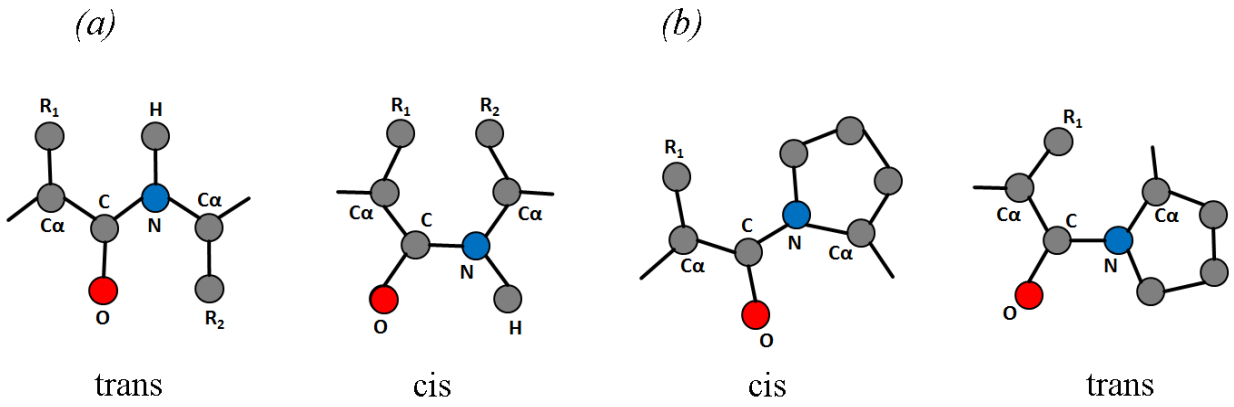

Figure 1: (a) Peptide bonds can adopt two different conformations, *trans* and *cis*. In the *trans* isomer, the C=O and N-H groups point in the opposite directions, whereas in the *cis* isomer, they point in the same direction. (b) When the second residue of a peptide bond is a proline, the icis-form is more frequent. [warning: an error was found].

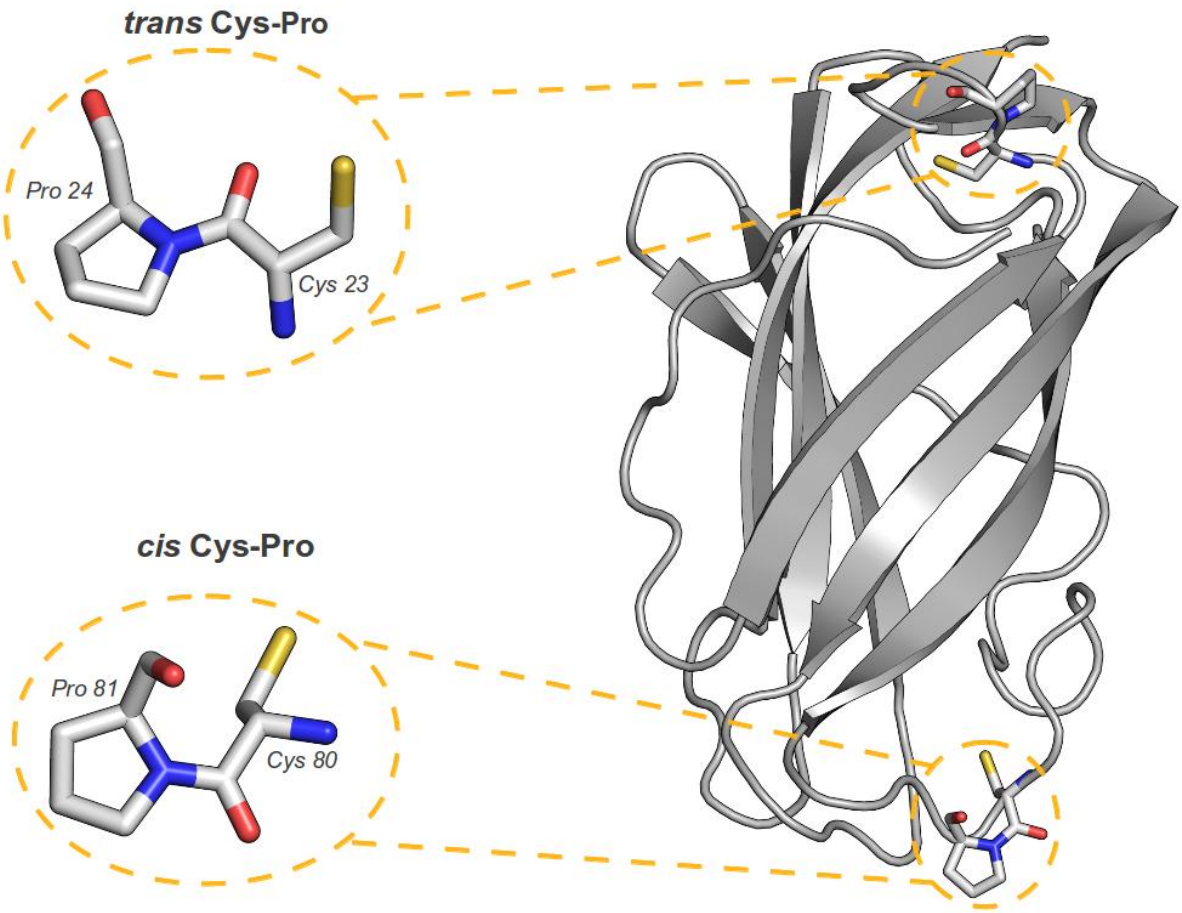

Figure 2: Representation of the *cis* and *trans* conformations in a cysteine-proline (Cys-Pro) peptide bond. The displayed structure is the Bovine NPC2 (Nieman-Pick C2) protein (PDB code 1NEP) (Friedland et al. 2003) obtained from the Protein Data Bank. In the *cis* conformation, the alpha carbons are locked on the same side of the peptide bond, whereas in the *trans* conformation, they lie on opposite sides of the peptide bond.

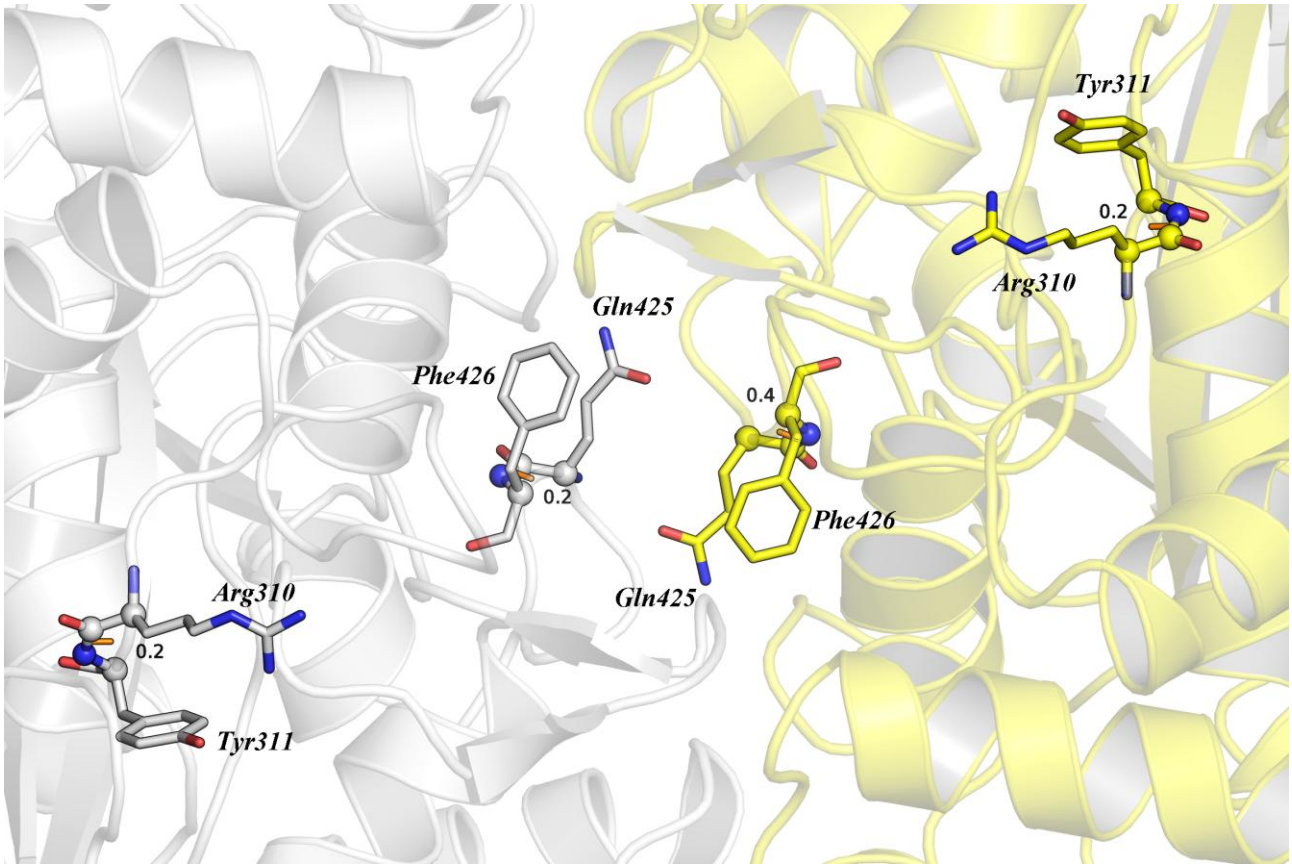

Figure 3: Structure of the factor XIII dimer (PDB code 1F13, (Weiss et al. 1998b)).One subunit is colored in silver and the second in yellow. Two non-proline *cis* peptide bonds are shown in each subunit with partner residues in sticks. One is between Arg310 and Tyr311 close to the active site and the other is between Gln425 and Phe426 at the dimerization interface. The value of the ω angle for the different *cis* peptide bonds is labeled on the figure and marked by an orange bar. The atoms composing the ω angle are represented in spheres.

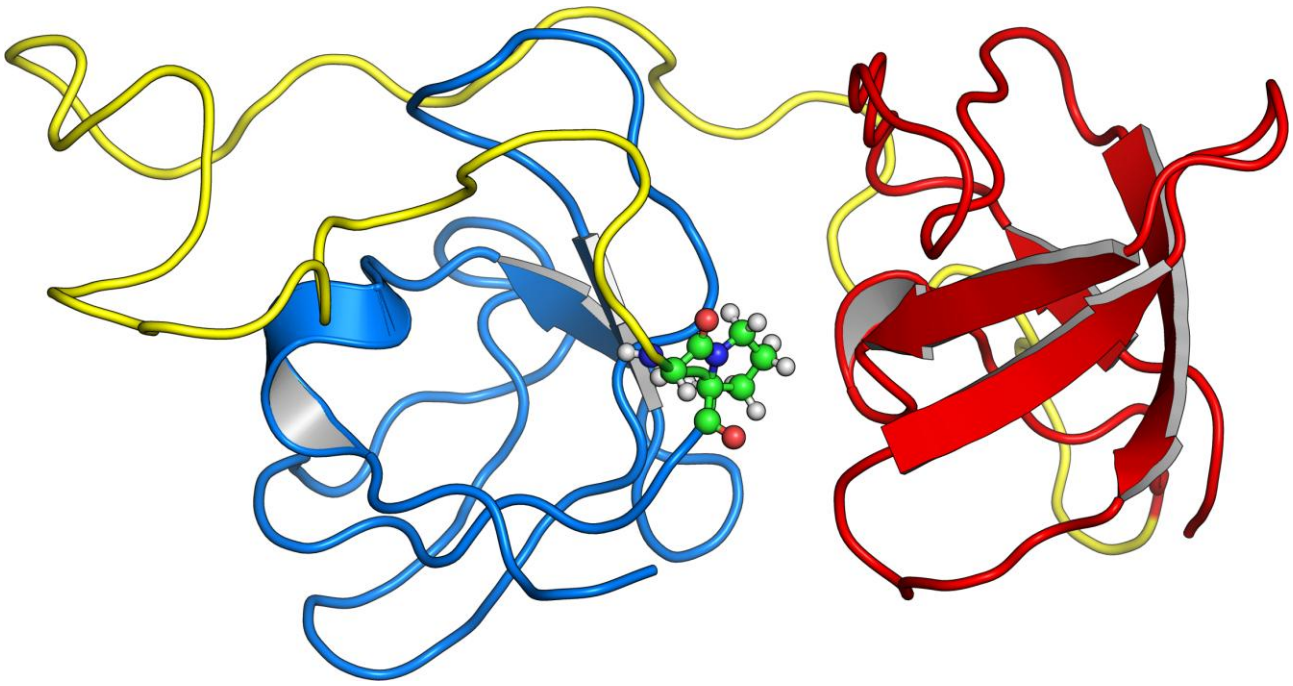

Figure 4:A NMR model of Crk adaptor (PDB code 2L3S, (Sarkar et al. 2010)). The presence of a heterogeneous proline residue undergoing *cis*-trans isomerization provides a potential binding site for PPIase enzymes. The Gly237-Pro238 bond (in green) is located on the linker (colored in yellow) between the two SH3 domains (colored in red and blue).

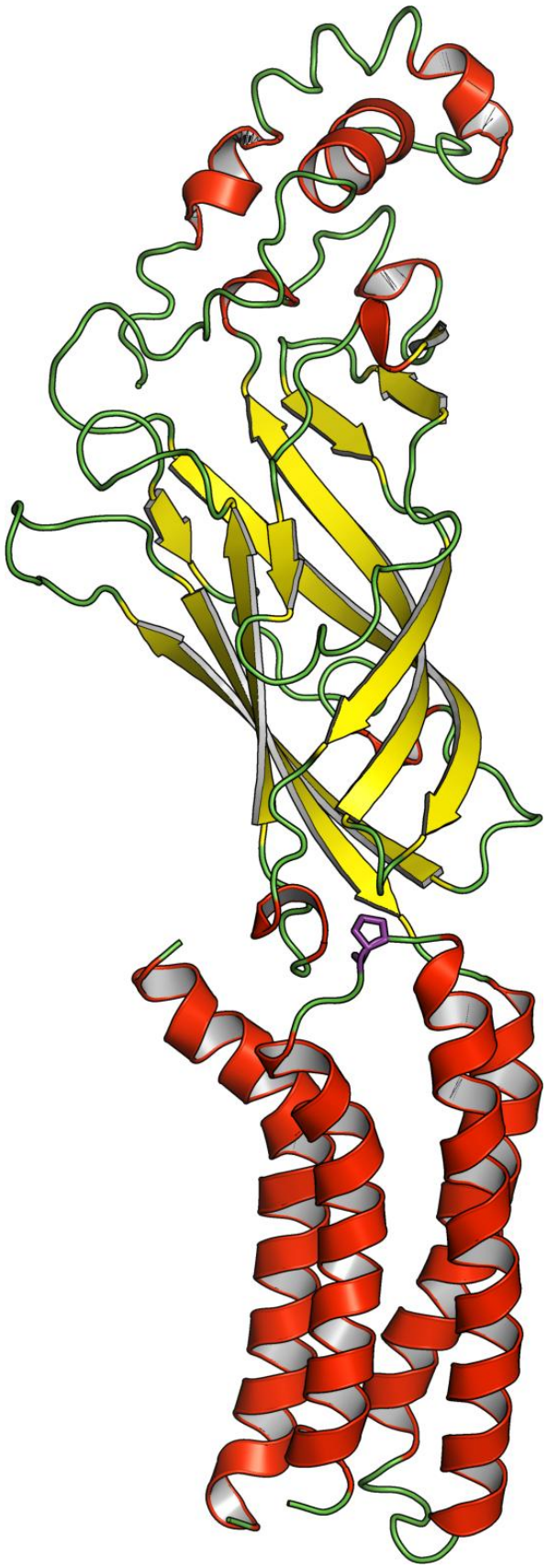

Figure 5: Model of The 5-Hydroxytryptamine type 3 receptor.Pro308, in its classical *trans* form, is shown as sticks and colored in purple. This model was generated by SPARKS (Yang et al. 2011).

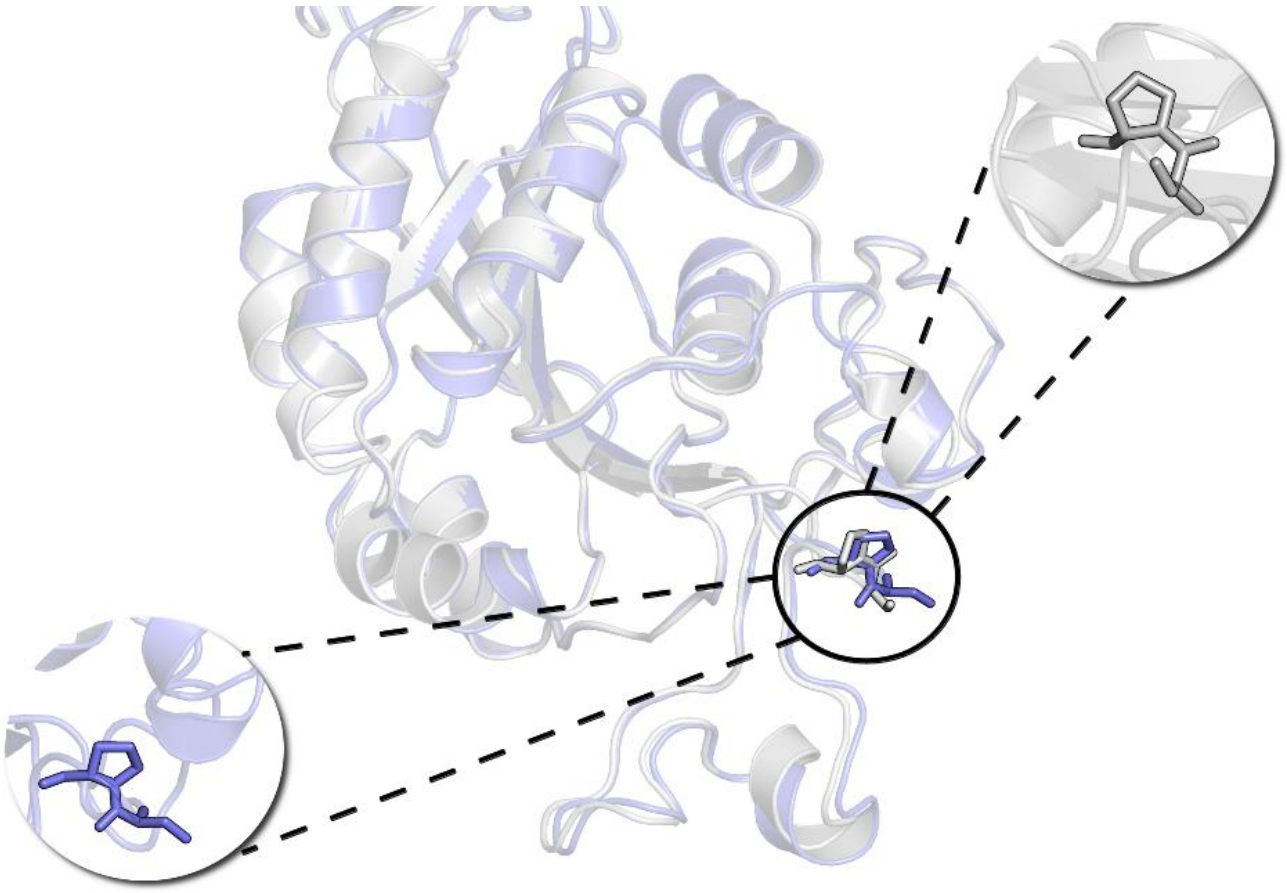

Figure 6: βI domain of the β3 subunit of the platelet integrin αIIbβ3. Structure colored in purple and silver contained the *trans* and *cis* Ser162-Pro163 peptide bond respectively.

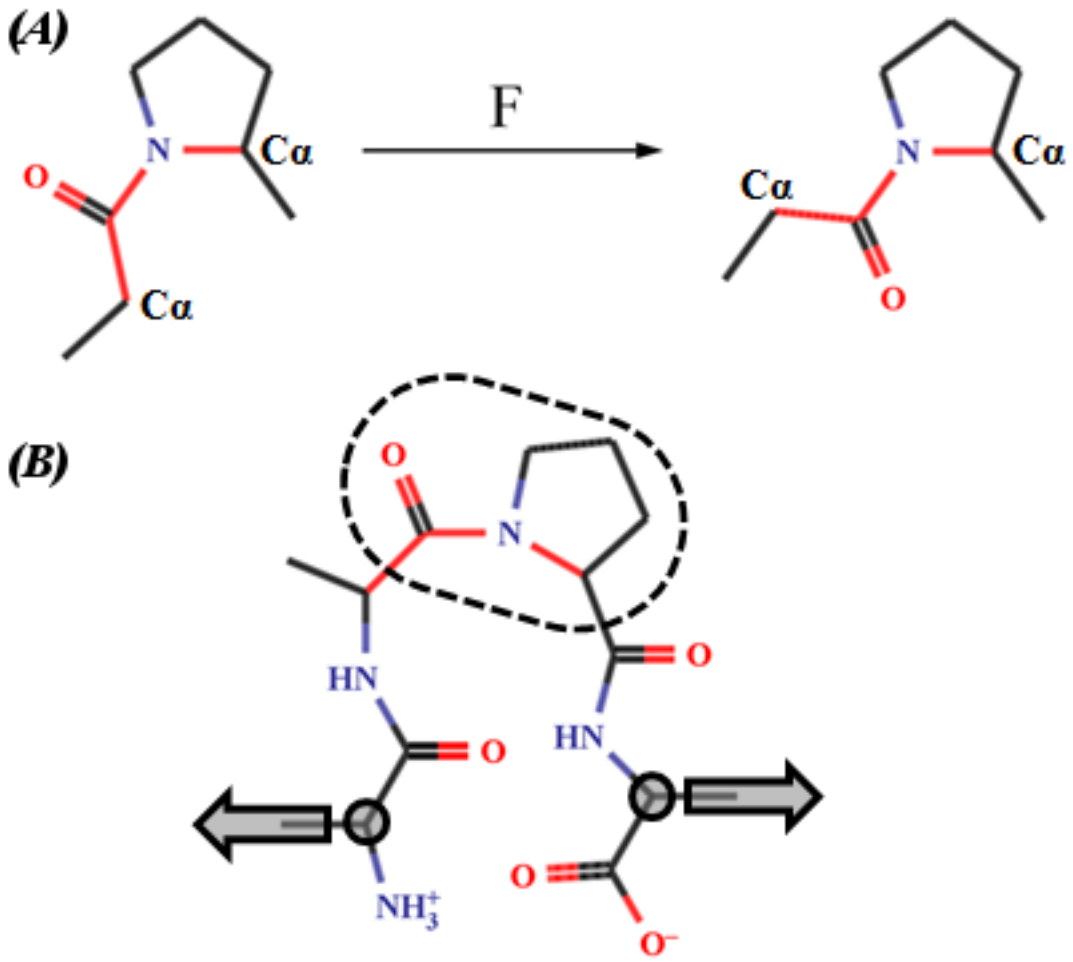

Figure 7: (A) Schematic presentation of the *cis* to *trans* isomerization. Peptide bonds are highlighted in red. (B) The simulated AAPA peptide with QM region separated from the MM region by a dashed line. In the Force-clamp molecular dynamics, the Cα atoms (gray spheres) were subjected to a constant force in opposite directions.

REFERENCES

Allen FH, Bellard S, Brice MD, Cartwright BA, Doubleday A, Higgs H, Hummelink T, Hummelink-Peters BG, Kennard O, Motherwell WDS, Rodgers JR, Watson DG (1979) The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. Acta Crystallographica Section B 35:2331-2339. doi:10.1107/S0567740879009249

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25 (17):3389-3402. doi:gka562 [pii]

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Jr., Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112 (3):535-542

Brandl CJ, Deber CM (1986) Hypothesis about the function of membrane-buried proline residues in transport proteins. Proc Natl Acad Sci U S A 83 (4):917-921

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

Chan PM, Ilangumaran S, La Rose J, Chakrabartty A, Rottapel R (2003) Autoinhibition of the kit receptor tyrosine kinase by the cytosolic juxtamembrane region. Mol Cell Biol 23 (9):3067-3078

Chen J, Edwards SA, Grater F, Baldauf C (2012) On the cis to trans isomerization of prolyl-peptide bonds under tension. J Phys Chem B 116 (31):9346-9351. doi:10.1021/jp3042846

Chou PY, Fasman GD (1974) Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry 13 (2):211-222

Darve E, Rodriguez-Gomez D, Pohorille A (2008) Adaptive biasing force method for scalar and vector free energy calculations. J Chem Phys 128 (14):144120. doi:10.1063/1.2829861

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. doi:10.1002/1097-0134(20001115)41:3<271::AID-PROT10>3.0.CO;2-Z [pii]

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

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

Evans TC, Jr., Nelsestuen GL (1996) Importance of cis-proline 22 in the membrane-binding conformation of bovine prothrombin. Biochemistry 35 (25):8210-8215. doi:10.1021/bi9606354

bi9606354 [pii]

Exarchos KP, Exarchos TP, Papaloukas C, Troganis AN, Fotiadis DI (2009a) PBOND: web server for the prediction of proline and non-proline cis/trans isomerization. Genomics Proteomics Bioinformatics 7 (3):138-142. doi:S1672-0229(08)60042-X [pii]

10.1016/S1672-0229(08)60042-X

Exarchos KP, Papaloukas C, Exarchos TP, Troganis AN, Fotiadis DI (2009b) Prediction of cis/trans isomerization using feature selection and support vector machines. J Biomed Inform 42 (1):140-149. doi:10.1016/j.jbi.2008.05.006

S1532-0464(08)00064-6 [pii]

Fogolari F, Corazza A, Varini N, Rotter M, Gumral D, Codutti L, Rennella E, Viglino P, Bellotti V, Esposito G (2011) Molecular dynamics simulation of beta(2)-microglobulin in denaturing and stabilizing conditions. Proteins 79 (3):986-1001. doi:10.1002/prot.22940

Friedland N, Liou HL, Lobel P, Stock AM (2003) Structure of a cholesterol-binding protein deficient in Niemann-Pick type C2 disease. Proc Natl Acad Sci U S A 100 (5):2512-2517. doi:10.1073/pnas.0437840100

0437840100 [pii]

Frommel C, Preissner R (1990) Prediction of prolyl residues in cis-conformation in protein structures on the basis of the amino acid sequence. FEBS Lett 277 (1-2):159-163. doi:0014-5793(90)80833-5 [pii]

Grathwohl C, Wuthrich K (1976) The X-Pro peptide bond as an nmr probe for conformational studies of flexible linear peptides. Biopolymers 15 (10):2025-2041. doi:10.1002/bip.1976.360151012

Gunasekaran K, Ramakrishnan C, Balaram P (1996) Disallowed Ramachandran conformations of amino acid residues in protein structures. J Mol Biol 264 (1):191-198. doi:S0022-2836(96)90633-6 [pii]

10.1006/jmbi.1996.0633

Hodel A, Rice LM, Simonson T, Fox RO, Brunger AT (1995) Proline cis-trans isomerization in staphylococcal nuclease: multi-substrate free energy perturbation calculations. Protein Sci 4 (4):636-654. doi:10.1002/pro.5560040405

Huber R, Steigemann W (1974) Two cis-prolines in the Bence-Jones protein Rei and the cis-pro-bend. FEBS Lett 48 (2):235-237. doi:0014-5793(74)80475-8 [pii]

Ikura T, Tsurupa GP, Kuwajima K (1997) Kinetic folding and cis/trans prolyl isomerization of staphylococcal nuclease. A study by stopped-flow absorption, stopped-flow circular dichroism, and molecular dynamics simulations. Biochemistry 36 (21):6529-6538. doi:10.1021/bi963174v

bi963174v [pii]

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

Jallu V, Dusseaux M, Panzer S, Torchet MF, Hezard N, Goudemand J, de Brevern AG, Kaplan C (2010) AlphaIIbbeta3 integrin: new allelic variants in Glanzmann thrombasthenia, effects on ITGA2B and ITGB3 mRNA splicing, expression, and structure-function. Hum Mutat 31 (3):237-246. doi:10.1002/humu.21179

Jallu V, Poulain P, Fuchs PF, Kaplan C, de Brevern AG (2012) Modeling and molecular dynamics of HPA-1a and -1b polymorphisms: effects on the structure of the beta3 subunit of the alphaIIbbeta3 integrin. PLoS One 7 (11):e47304. doi:10.1371/journal.pone.0047304

PONE-D-12-10352 [pii]

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292 (2):195-202. doi:10.1006/jmbi.1999.3091

S0022-2836(99)93091-7 [pii]

Joseph AP, Agarwal G, Mahajan S, Gelly JC, Swapna LS, Offmann B, Cadet F, Bornot A, Tyagi M, Valadie H, Schneider B, Etchebest C, Srinivasan N, De Brevern AG (2010) A short survey on protein blocks. Biophys Rev 2 (3):137-147. doi:10.1007/s12551-010-0036-1

Joseph AP, Srinivasan N, de Brevern AG (2012) Cis-trans peptide variations in structurally similar proteins. Amino Acids 43 (3):1369-1381. doi:10.1007/s00726-011-1211-9

Kang YK (2006) Conformational preferences of non-prolyl and prolyl residues. J Phys Chem B 110 (42):21338-21348. doi:10.1021/jp0647481

Kiefhaber T, Grunert HP, Hahn U, Schmid FX (1990) Replacement of a cis proline simplifies the mechanism of ribonuclease T1 folding. Biochemistry 29 (27):6475-6480

Levitt M (1981) Effect of proline residues on protein folding. J Mol Biol 145 (1):251-263. doi:0022-2836(81)90342-9 [pii]

Lorenzen S, Peters B, Goede A, Preissner R, Frommel C (2005) Conservation of cis prolyl bonds in

proteins during evolution. Proteins 58 (3):589-595. doi:10.1002/prot.20342

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

MacArthur MW, Thornton JM (1996) Deviations from planarity of the peptide bond in peptides and proteins. J Mol Biol 264 (5):1180-1195. doi:S0022-2836(96)90705-6 [pii]

10.1006/jmbi.1996.0705

Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta 405 (2):442-451

Melis C, Bussi G, Lummis SC, Molteni C (2009) Trans-cis switching mechanisms in proline analogues and their relevance for the gating of the 5-HT3 receptor. J Phys Chem B 113 (35):12148-12153. doi:10.1021/jp9046962

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]

Nelson CJ, Santos-Rosa H, Kouzarides T (2006) Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell 126 (5):905-916. doi:S0092-8674(06)01019-1 [pii]

10.1016/j.cell.2006.07.026

Pahlke D, Freund C, Leitner D, Labudde D (2005a) Statistically significant dependence of the Xaa-Pro peptide bond conformation on secondary structure and amino acid sequence. BMC Struct Biol 5:8. doi:1472-6807-5-8 [pii]

10.1186/1472-6807-5-8

Pahlke D, Leitner D, Wiedemann U, Labudde D (2005b) COPS--cis/trans peptide bond conformation prediction of amino acids on the basis of secondary structure information. Bioinformatics 21 (5):685-686. doi:bti089 [pii]

10.1093/bioinformatics/bti089

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

Perera L, Darden TA, Pedersen LG (1998) Trans-cis isomerization of proline 22 in bovine prothrombin fragment 1: a surprising result of structural characterization. Biochemistry 37 (31):10920-10927. doi:10.1021/bi980263u

bi980263u [pii]

Raleigh DP, Evans PA, Pitkeathly M, Dobson CM (1992) A peptide model for proline isomerism in the unfolded state of staphylococcal nuclease. J Mol Biol 228 (2):338-342. doi:0022-2836(92)90822-2 [pii]

Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. Adv Protein Chem 23:283-438

Sali A, Overington JP (1994) Derivation of rules for comparative protein modeling from a database of protein structure alignments. Protein Sci 3 (9):1582-1596. doi:10.1002/pro.5560030923

Sansom MS, Weinstein H (2000) Hinges, swivels and switches: the role of prolines in signalling via transmembrane alpha-helices. Trends Pharmacol Sci 21 (11):445-451. doi:S0165614700015534 [pii]

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

Sarkar P, Saleh T, Tzeng SR, Birge RB, Kalodimos CG (2010) Structural basis for regulation of the Crk signaling protein by a proline switch. Nat Chem Biol 7 (1):51-57. doi:nchembio.494 [pii]
10.1038/nchembio.494

Schmid FX, Baldwin RL (1978) Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. Proc Natl Acad Sci U S A 75 (10):4764-4768

Song J, Burrage K, Yuan Z, Huber T (2006) Prediction of cis/trans isomerization in proteins using PSI-BLAST profiles and secondary structure information. BMC Bioinformatics 7:124. doi:1471-2105-7-124 [pii]
10.1186/1471-2105-7-124

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:10.1016/0022-2836(90)90159-J

Suizu F, Ryo A, Wulf G, Lim J, Lu KP (2006) Pin1 regulates centrosome duplication, and its overexpression induces centrosome amplification, chromosome instability, and oncogenesis. Mol Cell Biol 26 (4):1463-1479. doi:26/4/1463 [pii]
10.1128/MCB.26.4.1463-1479.2006

Truckses DM, Somoza JR, Prehoda KE, Miller SC, Markley JL (1996) Coupling between trans/cis proline isomerization and protein stability in staphylococcal nuclease. Protein Sci 5 (9):1907-1916. doi:10.1002/pro.5560050917

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. doi:10.1073/pnas.151103798
98/16/8979 [pii]

Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10 (5):988-999. doi:10.1109/72.788640

Vohringer-Martinez E, Duarte F, Toro-Labbe A (2012) How does Pin1 catalyze the cis-trans prolyl peptide bond isomerization? A QM/MM and mean reaction force study. J Phys Chem B 116 (43):12972-12979. doi:10.1021/jp307946h

Vohringer-Martinez E, Toro-Labbe A (2011) The mean reaction force: a method to study the influence of the environment on reaction mechanisms. J Chem Phys 135 (6):064505. doi:10.1063/1.3624388

Wang ML, Li WJ, Xu WB (2004) Support vector machines for prediction of peptidyl prolyl cis/trans isomerization. J Pept Res 63 (1):23-28. doi:100 [pii]

Wedemeyer WJ, Welker E, Scheraga HA (2002) Proline cis-trans isomerization and protein folding. Biochemistry 41 (50):14637-14644. doi:bi020574b [pii]

Weiss MS, Jabs A, Hilgenfeld R (1998a) Peptide bonds revisited. Nat Struct Biol 5 (8):676. doi:10.1038/1368

Weiss MS, Metzner HJ, Hilgenfeld R (1998b) Two non-proline cis peptide bonds may be important for factor XIII function. FEBS Lett 423 (3):291-296. doi:S0014-5793(98)00098-2 [pii]

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. doi:10.1002/(SICI)1097-0282(19980415)45:5<381::AID-BIP6>3.0.CO;2-H [pii]

10.1002/(SICI)1097-0282(19980415)45:5<381::AID-BIP6>3.0.CO;2-H

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

Yang Y, Faraggi E, Zhao H, Zhou Y (2011) Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. Bioinformatics 27 (15):2076-2082. doi:btr350 [pii]

10.1093/bioinformatics/btr350

Yonezawa Y, Nakata K, Sakakura K, Takada T, Nakamura H (2009) Intra- and intermolecular interaction inducing pyramidalization on both sides of a proline dipeptide during isomerization: an ab initio QM/MM molecular dynamics simulation study in explicit water. J Am Chem Soc 131 (12):4535-4540. doi:10.1021/ja807814x