5.1 Introduction

The experimental results presented in the previous chapter establish Xaa–Pro–Aro as a cis-Pro-stabilizing motif. The aromatic residue takes part in a CH\text{...}\pi with the C^\alpha–H of N-terminal Pro. The importance of the cis peptidyl-prolyl motif often leads to the conservation of cis-Pro residue among homologous proteins over during evolution. It was shown that a cis-Pro residue could be conserved upto 20% sequence identity (compared to the original protein sequence). This is much stronger than the conservation of trans proline in similar cases.\textsuperscript{17, 79} Cis proline residues are also often known to play important functional and structural role in proteins.\textsuperscript{22}

The work on short peptides clearly showed that Pro–Pro–Aro is a sequence motif that stabilizes the Pro–cisPro bond by local interaction, operative between the N-terminal proline residue and the aromatic ring. Does this translate into a similar picture in proteins? A survey of Pro-Pro-Aro motifs in known protein structures, where the Pro–Pro bond is in cis-conformation, shows that a CH\text{...}\pi interaction, between the N-cap Pro and the aromatic residue is present in some cases but absent in many. This was reported by Pal \textit{et. al.}\textsuperscript{45} and Dasgupta \textit{et. al.}\textsuperscript{44} For example, the actin interacting protein 1 (pdb code: 1nr0) shows a CH\text{...}\pi interaction while glutathione synthetase (pdb code: 1gsa) does not. Both contain a Pro–Pro–Phe sequence where
Pro–Pro peptide bond is in cis conformation. However, the mere absence of the interaction does not mean that it plays no role in forming the cisPro conformation. It must be understood that the structure of a protein is dynamic and in this context it is quite possible that after stabilization of the cis Pro–Pro bond, the aromatic group following it could take part in some other interaction, with a different purpose – the experimental crystal structure being a mere snapshot of the many possible structures and the final conformation in which the protein crystallized might not contain the transient CH···π interaction. The detection and identification of CH···π interaction is often a tricky matter in protein structures, especially those solved using X-Ray crystallography. Since CH···π interactions are not included in the restraint libraries of the structure-refinement programs, it may well be possible that some of the interactions are ignored by design during refinement. Similarly, it is also possible that a CH···π interaction may show up as an artifact due to crystal packing forces and the real biological molecule may not have the interaction.

Therefore, given only a protein structure with a Pro–cisPro–Aro motif, it is difficult to judge whether the Pro–cisPro appears in the structure because local Pro–Aro CH···π interaction or it appears due to global folding compulsions. If the CH···π interaction is responsible for the observed Pro–cisPro, then one should expect a co-conservation of Pro–Aro residues among homologous proteins. In this chapter we investigate the sequence conservation of individual residues in the Pro–Pro–Aro motif in proteins that display a Pro–cisPro–Aro motif. In addition, we experimentally investigate the effect of mutating a cisPro motif (Gly–Pro–Tyr) in the protein glutamyl-tRNA synthetase (GluRS) in E. coli.
5.2 Materials and Methods

5.2.1 Sequence conservation

The non-redundant dataset described in Chapter 3 (Section 3.2.3) contain a total of 18 structures with the Pro–cisPro–Aro motif (Table 5.1). A database of homologous sequences, corresponding to each of these proteins, was constructed by collecting sequences from the NCBI sequence database.\textsuperscript{81, 82} The protein name (as in the pdb file) was used as a query in the NCBI server (four hypothetical proteins were excluded, see Table 5.1). The downloaded sequences were manually curated as: i) all putative proteins and hypothetical proteins were rejected, ii) all sequences bearing \( \geq \) 95\% identity (using the PISCES server\textsuperscript{58}) with any other sequence were removed, iii) sequences were rejected unless they belonged to the same kingdom (see Table 5.1) as the query sequence. This resulted in seven proteins with more than 40 associated homologous sequences (see Table 5.1). The curated sequences were then analyzed by multiple sequence alignment (MSA) using CLUSTALW\textsuperscript{83} and pair-wise alignment EMBOSS NEEDLE\textsuperscript{84} (with default parameters). The results from the alignment file and sequence identity output was further analyzed using in-house computer program written in perl programming language. The structures were viewed using PyMOL\textsuperscript{85} and CHIMERA\textsuperscript{86} molecular viewer software.

5.2.2 Protein expression and purification

The \textit{E. coli} GluRS gene (plasmid PLQ7612) was obtained from Prof. Jacques Lapointe (University of Laval, Quebec, Canada), which was cloned in pET-28c vector
with 6X-His tag. The following primers were designed and purchased from Sigma for the construction of the P72A mutant of *E. coli* GluRS.

Forward primer: 5’ GAGTGGGATGAAGGTGCGTACTACCAGACC 3’

Reverse primer: 5’ GGTCTGGTAGTACGCACCTTCATCCCACTC 3’

Using *E. coli* GluRS gene as the template, the primers were used for the PCR-based *in vitro* site directed mutagenesis (Stratagene®) for the construction of the mutant. The resultant plasmid DNAs were purified by QIAGEN®-plasmid purification protocol (Qiagen). Mutations were confirmed by sequencing the corresponding genes, with appropriate primers in DNA sequencer (Applied Biosystems).

Plasmid DNAs corresponding to the wild type *E. coli* GluRS and the P72A mutant were separately transformed into *E. coli* BL21 (DE3) competent cells. Single colonies from the plates were inoculated in Luria Broth (1 lit) and were expressed at 37 °C at 200 r.p.m (with 50 µg/ml of kanamycin). At an O.D$_{595}$ nm $= 0.5–0.6$, cells were induced by 1 mM IPTG and shaken at 37 °C for six hours. The cells were harvested by centrifugation at 5000 r.p.m. for 15 min at 4°C. The harvested cells were re-suspended in lysis buffer (50 mM phosphate buffer, pH 7.5, 100 mM KCl, 10mM β-mercaptoethanol and 5 mM imidazole). The cells were lysed by sonication, followed by centrifugation at 12,000 r.p.m. for 45 mins at 4°C. The supernatant was loaded to 5ml His-TRAP® column (GE Life sciences), which was pre-equilibrated with the lysis buffer. The column was then washed with 25ml of wash-buffer (50 mM phosphate buffer, pH 7.5, 100 mM KCl, 10mM β-mercaptoethanol and 15 mM
imidazole). Finally, the column was charged with elution buffer (50 mM phosphate buffer, pH 7.5, 100 mM KCl, 10 mM β-mercaptoethanol and 400 mM imidazole) and the eluted fractions were collected. The fractions were then analyzed by 10% Bis-tris–PAGE with coomassie staining (HIMEDIA) (Fig. 5.12). Purified fractions were dialyzed in appropriate buffer and were stored at −80°C in small aliquots.

5.2.3 Functional Assay

In vitro glutamylation assay of *E. coli* GluRS and the P72A mutant were carried out in 50 mM HEPES pH 7.5, 0.1 mM unlabeled l-glu, 16 mM MgCl₂, 2 mM ATP, 0.8 mM β-mercaptoethanol and [³⁻H] l-glu (purchased from Perkin Elmer) with 5 µM *E. coli* tRNA^{Glu} (purchased from Sigma) at 37°C as described before. The assay buffer (100 µl) contained 1 µl of [³⁻H] l-glu. The resultant radioactivity of the products was measured by liquid scintillation counter (Beckman-Coulter). Protein concentrations used per assay point were 0.9 nM (*E. coli* GluRS) and 1.4 nM (the P72A mutant). The acceptor capacity of *E. coli* tRNA^{Glu} used in the experiment was ~1.2 nmol/OD₂₆₀nm. The tRNA^{Glu} concentration was determined assuming 1.6 nmol/ml/OD₂₆₀nm for 100% aminoacylation. The specific radioactivity of [³⁻H] L-glu was 50 Ci/mmol.

5.2.4 Equilibrium and kinetic studies of protein unfolding studies

For urea-induced equilibrium unfolding studies, purified *E. coli* GluRS and P72A mutants were dialyzed in 50 mM potassium phosphate buffer (pH 7.5) and 100 mM KCl. Ultrapure urea was purchased from Sigma. Urea stock solutions
(0–9M) were prepared in the above buffer. In each of these urea stock solutions, desired volume of protein was added (final concentration 3 µM) and were incubated for ∼16 h before spectroscopic measurements at 25 °C (Fluorescence in HITACHI Spectro-fluorimeter and far-UV Circular Dichroism in Jasco J-815 spectropolarimeter). For fluorescence measurements, the excitation wavelength was 295nm. After correcting for buffer background signals, the center of gravity of fluorescence spectra normalized with respect to the area, \( \lambda_{cg} = \Sigma(I_\lambda) \), and ellipticity at 222 nm (\( \Theta_{222} \)) were plotted against urea concentrations and were fitted to a two-state model:\(^88\):

\[
S = \frac{(S^0_U + m_U[X]) + (S^0_F + m_F[X])\exp\{(\Delta G^0 - m[X])/RT\}}{1 + \exp\{(\Delta G^0 - m[X])/RT\}}
\]

(Eq. 5.1)

where \( S \) is the observed signal, \( S^0 \) is the signal from pure folded (F) and unfolded (U) states, \([X]\) is the denaturant concentration, \( m_U \) and \( m_F \) represent linear dependencies of \( S^0 \) on \([X]\), \( \Delta G^0 \) is the free energy of unfolding and \( m \) is the associated cooperativity parameter (\([X]_{1/2} \) is given by \( \Delta G^0/m \)). Temperature induced equilibrium unfolding studies of \( E. coli \) GluRS and P72A was monitored by CD spectroscopy (\( \Theta_{222} \)) in a Jasco J-815 spectropolarimeter attached with a temperature-controlled peltier.

5.2.5 ATP binding

Dissociation constants associated with of ATP-binding to wt-GluRS and the P72A mutant were measured by monitoring the intrinsic protein fluorescence
5.3 Results

5.3.1 Sequence conservation of PPY motif in tyrosinase (pdb code: 1wxc)

The crystal structure of tyrosinase (pdb code: 1wxc) contains a Pro–cisPro peptide bond (168PPY). The tryrosine residue is involved in a CH–π interaction with Cα–H of N-terminal proline residue in PPY, and its side chain maintains g(−) rotameric state Fig. 5.1. Homologous sequences of 1wxc (≥ 39% identical) were downloaded using BLAST and analyzed by multiple sequence alignment (MSA). The choice for the cutoff for sequences with 39% or more sequence identity was just to make sure that the global fold of the protein is not altered to a dramatic level. The sequence Logo Plot (for 168PPY) shows a complete conservation of the cis proline residue. The Tyr also only changes to other aromatic amino acids. This strongly supports the case for co-conservation of the Pro–Aro motif. When the conservation pattern for the position 168 is observed, it is seen that 168P is dispensable, changing to Ala (for most of the cases), Ser and Asp. The observed occurrence of various amino acid combinations for these three positions is shown in the Fig. 5.1.

This result led to the design of Ac–Ala–Pro–Trp–NH2 (APW) peptide and subsequent study of its cis-content with respect to Ac–Ala–Pro–Ala–NH2 (APA). It was found that the cis-content of APW (21%) was four times higher than APA (5%) (Fig. 5.1). Compatible cis-content for similar, but not identical, Ala–Pro–Trp sequence containing peptide have been obtained by Nardi et. al. It was also seen that the Cα–H of Ala was upfield shifted (3.5 ppm) for the cis conformer of APW. The sequence analysis shows that 170Tyr in tyrosinase stabilizes the cis peptidyl-prolyl
bond just as it does in the tripeptide APY (see Chapter 4) or just as Trp does in the tripeptide APW (this Chapter).

5.3.2 Sequence conservation of Pro–cisPro–Aro motif in known protein structures

Analysis on 1wxc showed how the Pro–Aro is co-conserved in cisPro–Aro motifs in proteins. This calls for a detailed analysis of the Pro–cisPro–Aro sequences in a larger set of non-redundant protein structures. A summary of similar analyses, performed on seven representative proteins (see first seven entries of Table 5.1), is given below.

5.3.2.1 PDB code: 1GSA

We analyzed (154 homologous sequences) glutathione synthetase, complexed with ADP and glutathione (1gsa), from Escherichia coli. In the crystal structure the Phe ring is in the t-rotameric state and does not participate in CH–π interaction with Pro (Fig. 5.2). However Pro is close to the active site of the protein and is known to interact with the sulfhydryl group, which makes the residue very important. The diproline motif is completely conserved when sequence identity was ≥ 30%. The Aro (Phe) is also strongly, but not exclusively, conserved till 30% sequence identity. The aromatic residue is conserved exclusively till about 60% sequence identity and majority of the sequences (~82%) till 30% sequence identity (Fig. 5.2b). Pro, on the other hand, is exclusively conserved till ≥ 30% identity and 97% of the sequences with ≥ 20% sequence identity have a conserved Pro. The Pro and the Pro start to disappear between 20-30% sequence identity bracket (Fig. 5.2b).
5.3.2.2 PDB code: 1V5V

Aminomethyltransferase (1v5v)\textsuperscript{91} from \textit{Pyrococcus horikoshii} (archea) show a Pro–cisPro–Phe motif at position 387 – 389 where \textsuperscript{389}Phe participates in CH–\(\pi\) interaction with the C\(\alpha\)–H of the \textsuperscript{387}Pro. This motif occurs almost at the C-terminal end of the protein, which ends in the sequence number 401. The aminomethyltransferase has \(~40\%\) sequence identity with many bacterial sequences of the same protein. Some of these bacterial proteins were also crystallized. In those crystal structures it was seen that when the \textit{cis}Pro was conserved the aromatic residue was co-conserved with it (\textit{Bacillus subtilis}, pdb code 1yx2). On the other hand, in case of \textit{Escherichia coli}, the central Pro is not conserved and as a result, the \textit{cis} peptide bond is not observed despite the presence of the aromatic group (Fig. 5.3). A total of 42 archeal aminomethyltransferase sequences were analyzed. In majority of cases, the central Pro is conserved with the aromatic residue being co-conserved (Fig. 5.3d). Bacterial aminomethyltransferase sequences (162) were also analyzed where 132 out of 162 (~81\%) sequences have the central Pro conserved (Fig. 5.3c). Of these, 129 sequences (with sequence identity as low as 20\% with respect to 1v5v) have the aromatic amino acid co-conserved with the central Pro. In the 129 sequences in only 12 sequences the N-cap Pro proline is conserved.

5.3.2.3 PDB code: 1VKY
In the crystal structure of bacterial (Thermotoga maritima) S–adenosylmethionine: tRNA ribosyltransferase-isomerase (pdb: 1vky), the 277Pro–cisPro–Phe does not exhibit a Cα–H (277Pro) to 279Phe CH···π interaction. The 279Phe is seen to be in the t-rotameric state and it participates in a CH···π interaction with the 275Ile (Fig. 5.4a). The conservation profile of the 277Pro–cisPro–Phe motif is also not strong. First of all, no bacterial sequence falls within 60% sequence identity of T. maritima S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Among the 183 sequences explored, only in 24 sequences the 277Pro residue is conserved. Of these, 22 have the Pro–Aro motif (Fig. 5.4c). The disappearance of the central proline residue hints at the unimportance of the prolyl-prolyl motif in this protein. It is worthwhile to note that the crystal structure of the protein from Bacillus subtilis also lacks the central proline residue and the cis peptide bond is absent in that position (Fig. 5.4b).

5.3.2.4 PDB code: 1X54

In the crystal structure (1x54) of asparaginyl-tRNA synthetase (AsnRS) from Pyrococcus horikoshii (archea), in the 277Pro–cisPro–Phe motif (Fig. 5.5), the 279Phe ring is in g(‘)-rotameric state but due to a non-canonical $\chi^2$ value (≠ ±90°) the ring does not face the Cα–H of 277Pro. This cis prolyl peptide bond is very special and is maintained for the entire aspartyl-tRNA synthetase (AspRS) and AsnRS system, spread over bacterial, archeal, as well as eukaryotal species (Fig. 5.5a). Even higher organisms like nematodes have the cis proline maintained along with the aromatic amino acid following the cis proline. Among the 98 bacterial species (with 40–50% sequence identity with 1x54) 83 have the Pro–Aro motif conserved. In
almost all the archaean sequences (except for one) the Pro−Pro−Phe is conserved as 
Pro−Pro−Aro or Xaa−Pro−Aro (Fig. 5.6). Of 275 bacterial sequences (with 30−40% 
sequence identity with 1x54), the central proline is conserved in 36 sequences. 35 
sequences (of these 36) have the aromatic residue co-conserved with the proline 
residue[Fig. 5.5b].

5.3.2.5 PDB code: 1MPG

The crystal structure with pdb ID 1mpg is of the protein 3-Methyladenine 
DNA Glycosylase II from the species Escherichia coli, with a cis prolyl-prolyl bond 
occurring at the very start of its sequence and C-capped with a Tyr residue 
(8Pro−cis9Pro−10 Tyr). The PPY motif lies in between a strand and helix. The Tyr 
residue does not take part in the CH−π interaction with the 8Pro (Fig. 5.7). Its side 
chain shows a \( \chi^1 \) of \( t (\pm 180^\circ) \) that places the aromatic ring far away from the proline 
residue. When conservation pattern of the PPY motif for different bacterial species 3-
Methyladenine DNA Glycosylase II were analyzed and it was seen that the central 
proline residue and aromatic residue C-capping the proline is completely conserved 
till 40% sequence identity as Pro−Pro−Aro (Fig. 5.7). Till 30% sequence identity 15 
out of 17 (~88%) sequences show conservation of the central proline as well as the 
aromatic residue. The bacterial 3-Methyladenine DNA Glycosylase II bearing 30-
40% sequence identity with the E. coli has 8 Xaa−Pro−Aro motif. Even below 30% 
sequence identity till 20% the proline residue and the aromatic residue is conserved, 
11 out of 33 sequences has Xaa−Pro−Aro motif (Fig 5.7). It is striking to note that 
there is no sequence with the central proline conserved in the absence of the aromatic 
residue following it up. This imply that although the crystal structure gives the
impression that the aromatic residue do not take part in the CH–π, its importance can be understood from its strong co-conservation with the central proline residue. This suggests that the aromatic residue might play a decisive role in guiding the proline residue to a cis conformation via a local interaction.

5.3.2.6 PDB code: 1WXC

The protein corresponding to crystal structure with pdb code 1wxc is tyrosinase from the bacterial species Streptomyces castaneoglobisporus. The Tyr residue of the \(^{168}\text{Pro} \rightleftharpoons \text{cisPro} \rightarrow \text{Tyr}\) motif takes part in a CH–π interaction with the Cα–H of \(^{168}\text{Pro}\). When the sequences of the protein tyrosinase were analyzed it was seen that 23 sequences bear ≥40% identity with Streptomyces castaneoglobisporus. All 23 sequences have the central proline residue conserved along with the aromatic residue following it, except for one case where a glycine occupies in that position (Fig. 5.8). With sequences occurring in between 20–30% identity of Streptomyces castaneoglobisporus the proline residue starts to disappear. Of the 7 sequences occurring between 20–30% sequence identity with the original pdb sequence only 3 sequences have the proline conserved and the aromatic residue is co-conserved with this proline. Looking at all sequences bearing ≥20% there are 32 sequences containing the proline residue and of which 31 sequences has the aromatic residue co-conserved with it (Fig. 5.8).

5.3.2.7 PDB code: 2VWS
The crystal structure of YFAU or 2-Keto-3-deoxy sugar aldolase (2-keto-3-deoxy-L-rhamnonate aldolase; pdb ID: 2vws) from the bacteria *Escherichia coli* contains the \textsuperscript{115}Pro–cisPro–Tyr motif with no CH–π interaction between \textsuperscript{115}Pro and the Tyr residue. Alignment of homologous sequences showed that \textsuperscript{115}PP segment is absolutely conserved but not \textsuperscript{117}Y (Fig. 5.9a). Interestingly, a number of residues surrounding the \textsuperscript{115}PP segment, including \textsuperscript{114}Y are conserved. We could identify a number of crystal structures of homologs without a conserved \textsuperscript{114}Y. In all cases \textsuperscript{116}P is in cis-conformation (Fig. 5.9b). Superposition of four structures (Fig. Fig. 5.9c) showed that \textsuperscript{114}YPpX is part of a turn with two conserved interaction pairs (see Fig. Fig. 5.9a and Fig. Fig. 5.9c): 1) between \textsuperscript{115}P and \textsuperscript{122}I, and, 2) between \textsuperscript{114}Y and \textsuperscript{120}R. Thus, it seems that for this case two residues just preceding cisPro, Pro (i-1) and Tyr (i-2) interact with two residues at the C-terminal side of cisPro, Val/Ile/Met (i+5) and Arg (i+3). This may be an example of cation–π interaction (Arg-Tyr) stabilized cisPro motif. Only experiments with short designed peptides would be able to validate this.

5.3.3 The effect of perturbing of a conserved cisPro residue in *E. coli* GluRS

Bacterial glutamyl-tRNA synthetase (GluRS) is a two-domain protein, consisting of a N-terminal domain that binds ATP and Glu, and a C-terminal domain that recognizes the anticodon nucleotides of tRNA\textsuperscript{Glu}. The N-terminal domain (see Fig. 5.10a) is also the catalytic domain since it is directly involved in transferring Glu to the acceptor stem of tRNA\textsuperscript{Glu}. As shown in Fig. 5.10, ATP binds in a cleft in the N-terminal catalytic domain in the crystal structure of GluRS from *Thermus thermophilus* (pdb ID: 1j09). On the opposite side of the cleft, a hinge is formed
between a conserved Tyr (residue position 80 in 1j09) and a sequentially distant but co-conserved Met (residue position 198 in 1j09). The Tyr residue is part of a conserved Gly-Pro-Tyr motif (see Fig. 5.11) and in all bacterial GluRS structures solved so far, the cisPro conformation is present in the motif (of the five crystal structures of T. thermophilus GluRS, 1G59, 1GLN, 1J09, 1N77 and 1N78, cisPro is present only in 1J09, 1N77 and 1N78, however, local sequence around GPY in these, VGGPHGPY, is different from that in the other two, VAAPTGPY). In order to understand the structural and functional role of the conserved Pro in GluRS, we mutated the corresponding Pro (Pro72 in E. coli GluRS) in E. coli GluRS to Ala and undertook some preliminary structural and functional studies.

5.3.3.1 Structural studies of P72A mutant of E. coli GluRS.

Although the crystal structure of E. coli GluRS has not been solved, its structure can be deduced from crystal structure of homologous proteins. In addition, the solution structure of GluRS has been studied using CD and fluorescence spectroscopy.\textsuperscript{87} We studied solution conformation of wt-GluRS and P72A mutant (see Fig. 5.12a for Bis-tris gel pictures of the purified proteins). In Figs. 5.12b and 5.12c the CD and the fluorescence spectra of wild type GluRS (wt-GluRS) and the P72A mutant (P72A) are shown. The CD spectra show that, over all, GluRS maintains its secondary structure upon P72A mutation, although a slight increase in ellipticity indicate a little gain of helical secondary structure. This is not unexpected since a Pro to Ala mutation directly affects the local backbone. The fluorescence spectra also show that the degree of Trp exposure in the two proteins is very similar (the $\lambda_{\text{max}}$ of fluorescence of the two proteins are identical). The decrease in intensity in the P72A
mutant indicates that non-radiative pathways of de-excitation of Trp in the two proteins are not identical, which may arise due to perturbation of the local environment of one or more Trp residues close to Pro72 (there are eight Trp residues in *E. coli* GluRS, of which Trp68 is within 6 Å of Pro72 in homologous structures).

5.3.3.2 *Equilibrium unfolding studies of P72A mutant of E. coli GluRS.*

Having shown that the P72A mutation induces very slight change in the native structure of wt-GluRS, we undertook urea-induced and temperature-induced equilibrium unfolding studies of the P72A mutant of *E. coli* GluRS. As shown in Figs. 5.13a and 5.13b, the fluorescence-detected and the CD-detected unfolding profiles of both, the wt-GluRS and the P72A mutant, are almost overlapped. In other words, as a first approximation it can be assumed that both the proteins unfold via a two-state model. The fluorescence-detected and the CD-detected data for wt-GluRS and P72A mutant were combined and are plotted in Fig. 5.13c for comparison. The comparison highlights two important features: i) P72A mutant unfolds at a lower urea concentration than the wt-GluRS, ii) the unfolding isotherm for the P72A mutant is much sharper at the transition than wt-GluRS. A quantitative analysis using a two-state model (Eq. 5.1) yielded the following parameters: $\Delta G_{\text{unfold}}^0 = 7.6 \text{ kcal/mol}$ and $m_{\text{unfold}} = 1.7 \text{ kcal/mol/M}$ (wt-GluRS) and $\Delta G_{\text{unfold}}^0 = 10.4 \text{ kcal/mol}$ and $m_{\text{unfold}} = 3.1 \text{ kcal/mol/M}$ (P72A mutant). It should be pointed out that these experiments were performed only once and therefore the data should be considered only preliminary (no error bars are given with fit parameters). Nonetheless, even if we ignore the extra-stability of the P72A mutant (about 3 kcal/mol), the sharpness of
transition of the P72A mutant (larger m value) indicates that same amount of urea brings about a larger perturbation to the native structure of the P72A mutant (than the wt-GluRS). As discussed above, the mutated residue, Pro72, is part of a hinge to the cleft where ATP binds. Our experiments did not include any ATP and therefore it is expected that urea could directly interact with residues that form the cleft (no ATP means an empty ATP-binding site). Changing the hinge residue, from Pro to Ala, may have brought about a change in the dynamics of the cleft that allowed urea to access it more than that in the wt-GluRS. However, this model needs to be verified with more complementary experiments like following the urea-induced unfolding in presence of ATP.

In the temperature-induced unfolding experiments (Fig. 5.13c), P72A mutant melted at a slightly lower temperature than the wt-GluRS. However, unlike the case of urea-induced unfolding, the sharpness of P72A mutant transition was almost identical, if not a little less sharp, than the wt-GluRS. Unfortunately, because both the proteins precipitated upon temperature induced unfolding (as visible from turbidity of the solution), the unfolding profile is not strictly reversible and therefore no attempt was made to quantitatively analyze the temperature-induced unfolding data.

5.3.3.3 Kinetic studies of unfolding of the P72A mutant of E. coli GluRS.

Going beyond equilibrium unfolding studies, we attempted to follow the kinetics of unfolding and refolding of wt-GluRS and the P72A mutant. Note that due to the unavailability of a Stopped-Flow machine, kinetic studies reported here only pertain to hand-mixing experiments (any relaxation with a time scale shorter than a second would appear as burst phase in these experiments). Time-resolved unfolding
experiments were performed by mixing a concentrated solution of urea with a small volume of concentrated solution of the protein (in appropriate buffer) that yielded the desired urea concentration where the protein is expected to unfold. Immediately after this, ellipticity (CD spectroscopy) at 222 nm was measured as a function of time. Similarly, refolding studies involved dilution of a protein solution in concentrated urea (unfolded protein) with aqueous buffer to reach a desired final urea concentration where the protein is expected to refold. This was again followed by monitoring the ellipticity at 222 nm as a function of time.

Typical unfolding and refolding time-dependent traces are shown in Figs. 5.14a and 5.14b. The CD signal is lost during the unfolding process (signal approaches towards 0 from an initial negative value) and is gained during the refolding process (signal goes to a negative value starting from 0). The time-dependent signals were fitted to a single exponential (with a life time $\tau$). Such experiments were done for different urea concentration in order to generate the Chevron plot where $\ln(1/\tau)$, or $\ln(k_{obs})$ where $k_{obs} = k_{fold} + k_{unfold}$, was plotted against the final urea concentration, as shown in Fig. 5.14c.

As expected, both wt-GluRS and the P72A mutant show a characteristic variation of $\ln(k_{obs})$, where $\ln(k_{obs})$ reaches a minimum at a urea concentration corresponding to the midpoint of equilibrium unfolding and linearly increases on both sides. The kinetic data can only be discussed with a plausible model (Fig. 5.15) of folding and unfolding of wt-GluRS and its P72A mutant. GluRS is a two domain protein, and earlier studies have shown that the C-terminal domain (anticodon binding domain) folds very cooperatively but the N-terminal domain (or the catalytic domain that contains Pro72) doesn’t. In order to analyze the kinetic data, the model also must
include a slow (seconds or more) kinetic step corresponding to the data presented in Fig. 5.14c. The unfolding/refolding experiments performed by manual mixing is associated with a dead time, which is around 15 seconds. The burst phase for unfolding kinetic studies (dotted horizontal lines in blue and in red), corresponding to the signal lost during the dead time, for both for wt-GluRS and for the P72A mutant, was about 60%. This means that unfolding data of Fig. 5.14c represents roughly about 40% of the total signal. On the other hand, the burst phase for the re-folding kinetic studies was ~ 80% for wt-GluRS and ~ 95% for the P72A mutant. In other words, the re-folding data of Fig. 5.14c represents roughly about 5-20% of the total signal.

Previous CD studies on wt-GluRS and the isolated C-terminal domain suggests that the isolated C-terminal domain contributes about 40% of the total ellipticity (at 222 nm) of wt-GluRS. Based on this and the fact that only 40% of the total CD signal could be observed during hand-mixing unfolding of wt-GluRS, the model (Fig. 5.15) assumes that the N-terminal domain unfolds fast (F → IC) followed by a slower unfolding of the C-terminal domain (IC → U). For the folding process, the model (Fig. 5.15) assumes that the C-terminal domain and most of the N-terminal domain fold fast (U → IC → ICN', about 80% of total signal lost in burst phase). Only small structural adjustments at the N-terminal domain (ICN' → F), corresponding to about 33% of the N-terminal domain CD signal, happens in the slow time scale (the slow phase corresponds to 20% of the total signal; if the N-terminal domain contributes 60% of the total signal and the slow phase appears exclusively due to the N-terminal then the slow phase (ICN' → F) can be considered as the final folding of about 33% of the N-terminal domain). Interestingly, for the P72A mutant this (ICN' → F) reduces to 5% of the total signal or only about 8% of the N-terminal folds with a slow time scale.
An unexpected feature of the Chevron plot is the leveling of \( \ln(k_{\text{obs}}) \) for wt-GluRS, rather than linearly increasing, at low urea concentrations. The value at which the leveling occurs, \( \ln(k_{\text{obs}}) \approx -5.5 \), corresponds to a rate constant of \( \sim 40 \times 10^{-4} \text{ sec}^{-1} \), compatible with the \( \text{cis/trans} \) isomerization rate constant of Gly-Pro peptides.\(^{35}\) Therefore, the kinetic data strongly suggest that the native structure of wt-GluRS from \( E. \text{ coli} \), whose crystal structure is not known, exhibits a \( ^{71}\text{Gly-cisPro} \) motif, just like GluRS from other species. For the P72A mutant, no such leveling was seen. However, because the Chevron curve also shifts towards low urea for the P72A mutant, any leveling due to a rate limiting process arising from \( \text{cis/trans} \) isomerization would not be observable.

Unfolding rate constants for the C-terminal domain can be estimated from the intercept of the unfolding arm of the Chevron plot of Fig. 5.14c. This value, \( k_{\text{H}_2\text{O}}^{\text{unfold}} \), is \( \sim 6 \times 10^{-6} \text{ sec}^{-1} \) for wt-GluRS and \( \sim 2 \times 10^{-5} \text{ sec}^{-1} \) for the P72A mutant. In other words, the P72A mutation destabilizes the folded C-terminal domain (compared to the transition state) by about \(-0.592 \ln(10) \approx 0.7 \text{ kcal/mol} \).

The re-folding data for wt-GluRS (refolding of a small portion of the N-terminal domain) only show an initial linear increase with decreasing urea. The intercept of this linear increase at zero urea is \( \sim \ln(k_{\text{H}_2\text{O}}^{\text{refold}}) = 0 \), indicating that the corresponding re-folding rate constant is \( \sim 1 \text{ sec}^{-1} \). However, due to some rate-limiting process (probably \( \text{cis/trans} \) isomerization), the actual rate at zero urea levels off at around \( \ln(k_{\text{H}_2\text{O}}^{\text{refold}}) \approx -5.5 \) (\( k_{\text{obs}} \approx 40 \times 10^{-4} \text{ sec}^{-1} \)). For the P72A mutant, the re-folding rate decreases linearly with decreasing urea with an intercept of about \( \ln(k_{\text{H}_2\text{O}}^{\text{refold}}) \approx -5.5 \) (\( k \approx 40 \times 10^{-4} \text{ sec}^{-1} \)), with no clear indication that it involves a \( \text{cis/trans} \) isomerization step (for P72A mutant). Therefore, P72A mutation has two effects on
the final structural adjustment of the N-terminal domain during refolding: i) only a minor fraction of the N-terminal shows a slow adjustment, ii) the rate of the adjustment is slower by two order of magnitude compared to that of wt-GluRS, meaning the penultimate unfolded state of the P72A mutant, corresponding the final adjustment ($I^{CN} \rightarrow F$), is about 3 kcal/mol ($0.592 \ln 240$) more stable than the transition state when compared to similar values for the wt-GluRS.

5.3.3.4 Aminoacylation kinetics and ATP binding studies on the P72A mutant.

What are the functional effects of replacing Pro72 by Ala? To address this issue we performed two preliminary experiments on the P72A mutant and compared it with that of wt-GluRS. The first is the measurement of ATP-affinities for P72A mutant and the wild type protein. Since Pro72 is part of a hinge to the ATP binding pocket, we also determined the dissociation constant of ATP, with wt-GluRS and the P72A mutant, to check whether mutating Pro72 affects the ATP binding properties of GluRS. As shown in Fig. 5.15a, the $K_d$ value for ATP-binding to the P72A mutant (10 µM) is about four times smaller than that of wt-GluRS (40 µM), meaning that the P72A mutation facilitates a slightly stronger ATP binding. These results are very preliminary and the experiments were performed only once. To draw a more robust conclusion, the glutamylation kinetics needs to be followed by measuring $K_m$ and $k_{cat}$ values. Also, it would be worthwhile to study the kinetics of ATP-binding because there is a strong possibility that perturbing the hinge residue perturbs the dynamics of ATP binding.
We also performed glutamylation assays where wt-GluRS and the P72A mutant was used to aminoacylate tRNA$^{\text{Glu}}$ with radioactive L-Glu. The concentration of the final product, Glu-tRNA$^{\text{Glu}}$, was measured as a function of time, as shown in Fig. 5.15b. The initial slopes associated with the rise of concentration of products are directly correlated with glutamylation efficiencies. These were 3 pmol/min and 2 pmol/min for the P72A mutant and wt-GluRS. After normalizing with the concentrations of the two enzymes used in the assay studies (0.9 and 1.4 nM for wt-GluRS and the P72A mutant respectively), it was observed that the glutamylation efficiency of the P72A mutant and wt-GluRS are almost identical.
4.4 Conclusions and Implications

Using a small dataset of proteins with known high-resolution crystal structures and containing a Pro-cisPro-Aro motif we showed that in homologous sequences for which crystal structures are known, the central proline remains in the cis conformation. When the central proline residue disappears in the sequence, the crystal structure of the homologous proteins (if available) showed the disappearance of the cis conformer. This shows that the experimental results obtained for short peptides containing the Pro-cisPro-Aro motif are also valid in proteins and may be structurally or functionally important since these motifs are evolutionarily conserved. However, more detailed studies with a much larger dataset need to be conducted before assessing the robustness of the general conclusions drawn.

We also experimentally studied a protein, GluRS from *E. coli*, probing its structure-function relationship with the presence and absence (replacement by Ala) of a conserved Pro residue that is found adopt the cis-conformation in homologous proteins from other bacterial species. We showed that The Pro to Ala mutation abolishes a cis/trans isomerization rate-limiting step in the refolding of GluRS. We also showed that the Pro to Ala mutation affects the stability of the protein. Finally we showed that although there is no apparent change in the glutamylation efficiency of the Pro to Ala mutant, the ATP-binding affinity of the mutant is about four-fold higher than the wild type protein. It should be noted that the conserved Pro residue is part of a hinge region that might have a role in modulating the opening and closing of the ATP-binding cleft. The data presented is only preliminary. Nevertheless, our experiments set the stage for a more elaborate experimental work where not only more studies could be performed on the Pro72Ala mutant but also other mutants,
especially the Tyr73Ala mutant needs to be studied. That would directly focus on disrupting the CH⋯π interaction. Preliminary experiments have already been initiated to realize this.