Contribution of Proline to the Pre-Structuring Tendency of Transient Helical Secondary Structure Elements in Intrinsically Disordered Proteins

Chewook Lee[1], Lajos Kalmar[2], Bin Xue[3], Peter Tompa[2], Gary W. Daughdrill[4], Vladimir N. Uversky[3,5,6], and Kyou-Hoon Han[1,7]*

[1]*Division of Convergent Biomedical Research, Biomedical Translational Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro, Yuseong-gu, Daejeon, 305-806, Korea*

[2]*Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, 1518 Budapest, P.O. Box 7, Hungary*

[3]*Department of Molecular Medicine, University of South Florida, Tampa, Florida 33612, United States*

[4]*Department of Cell Biology, Microbiology, and Molecular Biology and Center for Drug Discovery and Innovation, University of South Florida, Tampa, Florida 33612, United States*

[5] *USF Health Byrd Alzheimer's Research Institute, Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612, United States.*

[6]*Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia*

[7] *Department of Bioinformatics, University of Science and Technology, 113 Gwahak-ro, Yuseong-gu, Daejeon, 305-333, Korea.*

* To whom correspondence should be addressed.

Tel : 82-42-860-4250

Fax : 82-42-860-4259

Email : khhan600@kribb.re.kr

ABSTRACT

Background: IDPs function without relying on three-dimensional structures. No clear rationale for such a behavior is available yet. PreSMos are transient secondary structures observed in the target-free IDPs and serve as the target-binding "active" motifs in IDPs. Prolines are frequently found in the flanking regions of PreSMos. Contribution of prolines to the conformational stability of the helical PreSMos in IDPs is investigated.

Methods: MD simulations are performed for several IDP segments containing a helical PreSMo and the flanking prolines. To measure the influence of flanking-prolines on the structural content of a helical PreSMo calculations were done for wild type as well as for mutant segments with Pro→Asp, His, Lys, or Ala. The change in the helicity due to removal of a proline was measured both for the PreSMo region and for the flanking regions.

Results: The α-helical content in ~70% of the helical PreSMos at the early stage of simulation decreases due to replacement of an N-terminal flanking proline by other residues whereas the helix content in nearly all PreSMos increases when the same replacements occur at the C-terminal flanking region. The helix destabilizing/terminating role of the C-terminal flanking prolines is more pronounced than the helix promoting effect of the N-terminal flanking prolines.

General Significance: This work represents a novel example demonstrating that a proline is encoded in an IDP with a defined purpose. The helical PreSMos presage their target-bound conformations. As they most likely mediate IDP-target binding via conformational selection their helical content can be an important feature for IDP function.

Keywords: Intrinsically Disordered Protein (IDP), PreSMo (Pre-Structured Motif), Flanking Proline, Molecular Dynamics Simulation

1. INTRODUCTION

Intrinsically disordered proteins (IDPs) [1-4] do not form uniquely folded structures under physiological conditions but they are still able to perform their specific functions, such as transcriptional activity [5], translational activity [6], and binding to target proteins [4]. IDPs exist as a dynamic conformational ensemble and have no fixed equilibrium values for backbone dihedral angles (ψ, ϕ). Various reasons including uncompensated charge groups and the low percentage of hydrophobic amino acid residues have been hypothesized as the reason that IDPs do not form well-ordered three-dimensional structures [7,8]. The conformational states of both globular proteins and IDPs can be affected by temperature, pH, counter ions, and other factors [9]. However, the target-bound structures of IDPs are more significantly governed by their binding partners than globular proteins because they fold into relatively ordered structures upon binding [10]. Furthermore, the structural flexibility of IDPs makes them more promiscuous in target interactions [2,11]. For example, the tumor suppressor p53, which is 50% disordered in its monomeric form, can bind to hundreds of target proteins [2]. More specifically, the N-terminal transactivation domain (TAD) of p53, containing three pre-structured motifs [5,12], can bind to three different target proteins [12-14]. In addition, a short disordered segment in the C-terminal region of p53 binds with four different partners, adopting different conformations depending on the context provided by the binding partners [2,15-18].

Although IDPs are topologically disordered, they often contain transient secondary structural elements that were originally named ***lo***cal ***st***ructural (lost) elements which have recently been redefined as ***pre-s***tructured ***mo***tif***s*** (**PreSMos) [**5,12]. These results contradicted the simple notion which prevailed in the early days of IDP research that IDPs with extended disorder were completely unstructured down to the level of secondary structures. It turned out that PreSMos are present in over two dozens of target-unbound IDPs which represent approximately two-thirds of IDPs whose structural states have been characterized in detail by NMR [4]. In fact, IDPs can be divided into two classes, a mostly unstructured (MU, "μ",) subtype and a completely unstructured (CU, "Q") subtype based upon whether they contain PreSMos or not [4]. PreSMos are critically involved in the target binding of many IDPs, such as p53 TAD [13], VP16 TAD [19,20], preS1 of HBV [21], and 4EBP1 [6], and represent the "active sites" in these IDPs [4]. Unlike the spatially disposed active pockets found in globular

proteins the PreSMos are linear contiguous motifs most of which are a short stretch (2~3 turns) of a helix. Presence of PreSMos in target-unbound IDPs has an important bearing to the question of whether the IDP-target binding mechanism involves conformational selection or not.

Proline is one of two natural amino acids that do not have the typical backbone torsion angles. The ring formation in prolines involving the peptide backbone restricts the allowed backbone torsion ψ and φ angles to a smaller region than other amino acids [22]. Theoretical studies on the proline-containing α-helix model peptides revealed an interesting correlation between the position of proline and the stability of a helix [23]. The highest positional propensities of proline were found in the helical peptides when a proline is situated at the N-terminus of a helix [23,24]. Without much surprise one finds that IDPs have many disorder-promoting proline residues [25,26]. Furthermore an interesting point was revealed during a recent examination of the amino acid sequences for the μ type IDPs that prolines are more frequently observed in the flanking regions of PreSMos than other regions of IDPs. Since many PreSMos are α-helical [4] it is likely then that the conformational stability of PreSMos may be affected by the proline residues in the flanking regions of PreSMos. We have used bioinformatics analysis tools and molecular dynamics (MD) simulation to explore the effect of prolines on the structural integrity of the helical PreSMos. We have replaced the prolines both in the N- and the C-terminal flanking regions of PreSMos with 4 different amino acid residues and have monitored the incurring conformational changes of PreSMos in several μ type IDPs [4] such as p53 TAD, securin, I-2, ENSA, CREB KID, and DARPP-32.

2. MATERIALS AND METHODS

2.1 PreSMo Selection

The helical PreSMos for MD simulation were selected manually from the recently published list of PreSMos [4]. These PreSMo peptides shown in Table 1 were selected because they are the ones defined most clearly by NMR experiments. The degree of pre-population or pre-structuring for most PreSMos in IDPs is 20~50% [4]. All PreSMo peptides under investigation contained at least one proline both at the N-terminal flanking region (NFR), and the C-terminal flanking region (CFR) within 5 amino acid residues from a PreSMo except for the case of the PreSMo **VII** in ENSA and the PreSMo **IX** in the CFR of DARPP-32 (see below). We have limited the number of flanking residues to 5 since beyond this number of residues the capping effect is likely to be insignificant. For convenience we have divided the PreSMo peptides into three regions; an NFR, a PreSMo region (PR), and a CFR (Figure 1). In Table 1 we summarized the sequences of 9 PreSMo peptides, i.e., p53 TAD (12-28, 35-48, and 45-59, denoted as **I , II**, and **III**) [5,12], securin (144-164 and 172-183, denoted as **IV** and **V**) [27], I-2 (90-107, denoted as **VI**) [28], ENSA (27-43, denoted as **VII**) [29,30], CREB KID (131-147, denoted as **VIII**) [31,32], and DARPP-32 (27-46, denoted as **IX**) [28].

2.1.1 p53 TAD

p53 TAD has three PreSMos, one is helix (residue 18-26), and the other two are short turns (residue 40-44 and 48-53) [5]. The populations of each PreSMo obtained from SSP calculation were approximately 25, 5, and 15% [4], respectively. Pro residues were located in both flanking regions of all three PreSMos in p53 TAD.

2.1.2 Securin

Securin was shown to have three PreSMos, one is β-strand (residues113-127) and the others (residue 150-159 and 174-178) are helices [27]. SSP calculations show that the populations of the helices are 45 and 20%, respectively [4]. All three PreSMos have Pro resides in their both terminal flanking regions. We selected two helical PreSMos since all the other PreSMos examined in this study are helical.

2.1.3 I-2

Dancheck *et al*. showed that I-2 had three PreSMos (residues 36-42, 96-106, and 132-138) [28] with the helical populations of 30, 48, and 98%, respectively. The second helix PreSMo was used for the current study.

2.1.4 ENSA

The predicted populations of three PreSMos of ENSA were 40, 10, and 30% for residue 32-36, 48-50, and 65-70, respectively [4,29,30]. Only the first PreSMo has Pro residues in both flanking regions.

2.1.5 CREB KID

Originally CREB KID was viewed as a CU type requiring an induced-fit for IDP-target binding [32]. However, upon examination of NMR parameters we concluded that it in fact has two helical PreSMos in its unbound state [4,31]; the populations of two PreSMo helices are >50% (A-Helix: residue 119-129) and ~10% (B-Helix; 134-143), respectively [1,31]. In the case of A-Helix, Pro was found only in its CFR, but the B-Helix has Pro residues both in NFR and CFR. Thus, we have chosen the B-Helix for our study.

2.1.6 DARPP-32

DARPP-32 has two helical PreSMos (residues 22-29 and 103-114) [28] and their predicted populations were 50 and 25% [4]. Only the first PreSMo selected in this study has prolines in both flanking regions.

2.2 Amino Acid Composition in PreSMos/α-MoRFs and Their Flanking Sequences

In Table 2, we show the amino acid frequencies of all reported PreSMo peptides [4] were calculated not only for the PR themselves but also for CFRs and NFRs. A previous bioinformatics analysis identified so called α-helical molecular recognition elements (α-MoRFs) that are short segments of IDPs to form an α-helix upon target binding [33-35]. PreSMos and α-MoRFs are similar in their helical structural nature even though α-MoRFs are different by definition from the helical PreSMos since the latter are found to be helical in

the target-unbound state whereas the former are helical segments of IDPs only observed in the target-bound x-ray structures. We have analyzed the amino acid frequencies in the α-MoRFs and their flanking regions using the same logic used for analyzing the amino acid sequence of PreSMos and the results are summarized in Table 3. The second column was calculated by taking the residue specific averages of five contiguous residues preceding the first residue in an α-MoRF. These averages were normalized to the residue specific frequencies in all disordered proteins. The third column shows the frequency of prolines in the C-terminal flanking region of α-MoRFs. The values in this column were calculated by taking the residue specific averages of five contiguous residues following the last residue in an α-helical conformation. These averages were also normalized to the residue specific frequencies in all disordered proteins.

2.3 Molecular Dynamics (MD) Simulations

2.3.1 MD Procedure

All MD simulations were performed with SANDER module of AMBER10 program package [36] using the ff99SB force field [37]. Nine wild-type PreSMo peptides and four mutants of each wild-type PreSMo were generated by Swiss PDB Viewer [38]. Since we wanted to observe how the structural content of a helical PreSMo is affected by presence or removal of a proline in the flanking regions we started MD calculations with the PreSMo peptides which are initially in an ideal alpha-helical conformation and followed the loss of helicity (unfolding) due to mutations during the course of MD simulations. The helical structure contents of the wild-type PreSMos were compared with those of the mutants containing Asp, His, Lys, or Ala; Asp and Lys were chosen for their negative or positive charge effects. Ala was selected for its small hydrophobic sidechain whereas His was selected for its somewhat larger sidechain than a proline.

Simulations were performed for 10ns at 300K under a neutral pH condition (total 81 trajectories) with the peptides capped by ACE (acetyl) and NME (N-methylamine) at the N- and C-terminus, respectively. Short simulations were performed since the main purpose of our study is to compare the effect of different mutations to the loss (unfolding) of the helicity. On the other hand, for a few cases a longer simulation time (40~50 ns) was used in order to

check if short simulation results may be extrapolated to longer simulations. The starting PreSMo peptides were explicitly solvated with 1,600~4,100 TIP3P water molecules in the rectangular box where the distance to the edge of solvent box from PreSMo peptide was chosen to be 10~14 Å. To neutralize the systems, Na^+ or Cl^- counter ions were added. The particle mesh Ewald method [39] was applied to account for long-range electrostatic interactions, and a 10.0 Å cutoff was used for nonbonded interactions. The hydrogen atoms were constrained to the equilibrium bond length using the SHAKE algorithm [40]. In order to remove unfavorable van der Waals contacts, 500~4,000 steps of steepest minimization and 500~6,000 steps of conjugate gradient minimization were performed with 500 kcal $(mol\ Å^2)^{-1}$ harmonic potential. Then, the whole system (peptide and solvent) was minimized using 4,000~12,000 steps of steepest descent minimization without any restraints. These systems were subsequently subjected to 50~100 ps equilibration process where the temperature of the system was gradually raised to 300K under atomic positional restraint of peptide. After equilibration, the production runs were preformed for 10 ns with 2fs time steps and NPT ensemble and 5,000 structures were collected for every PreSMo peptide.

2.3.2 Trajectory Analysis

The trajectories were analyzed using several parameters. The root-mean-square deviation (RMSD) value for Cα atoms is a useful parameter for quantifying the conformational changes between two structures of the same peptide or protein. In this study, Cα RMSDs between structures in the trajectories and the initial α-helical state were calculated. To analyze the fluctuation and statistics of the secondary structure of the wild-type and its four mutants, the dictionary of secondary structure of proteins (DSSP) program developed by Kabsch and Sander was used [41]. All secondary structures were characterized by DSSP and were classified into three helix types (alpha-helix, 3_{10}-helix, and π-helix), two sheet types (antiparallel and parallel β-sheet), and others (turn and coil). Here we considered only alpha-helical structure to keep the consistency in comparing several PreSMo peptides. The clustering analysis was employed to generate the pools of conformation for each MD trajectory. The clustering method used the pairwise Cα RMSDs between different frames. Using this clustering method, 40 groups were produced for each trajectory and the closest structure to the centroid of the largest cluster was selected (See Figure 2). The helicity of each stored structure was calculated from secondary structure analysis. Also various distance

analyses between two specific atoms were directly measured from MD trajectories. The *ptraj* toolset available in the AMBER program as well as in-house analysis programs were used to analyze the resulting MD trajectories. The VMD [42] and PyMol [43] software were used for the preparation of the structural figures and the visual analyses in this study.

3. RESULTS

3.1 Amino Acid Frequency in PreSMos/α-MoRFs and Their Flanking Regions

The amino acid composition within PRs and their flanking regions are summarized in Table 2. For several hydrophobic residues we observed reduced frequency inside the PRs where an overrepresentation of polar residues is also observed. The number of experimentally verified PreSMos is ~ 50 [4]. Nonetheless, the observed trend is remarkably clear for prolines even with this small sample size. The frequency of prolines is ~ 3 times higher in NFRs and CFRs than that from UniProt is observed whereas it is only half of the UniProt value inside PRs.

A similar but weaker trend is observed in the case of α-MoRFs. The second column in Table 3 shows the relative frequency of prolines in the N-terminal flanking region with respect to those of α-MoRFs. In concurrence with the data on PreSMos presented in Table 2, the frequency of the C-terminal flanking prolines in α-MoRFs is higher than the frequency of 0.0718 observed in all IDPs [44]. The frequency of N-terminal flanking prolines of α-MoRFs is also higher in CFRs when compared with the average frequency in all IDPs. Interestingly, if the average is taken over 6 N-terminal flanking residues instead of 5 the proline frequency is even higher (1.23). Overall in both PreSMos and α-MoRF data sets, the frequency of prolines flanking the alpha helix-forming regions in IDPs is higher than the overall frequency of prolines in IDPs. These results strongly suggest presence of positive selection to maintain proline residues in flanking regions of transient helical segments.

3.2 MD Simulations Suggest a Helix-Destabilizing Role for Flanking Prolines

Figure 2 shows the relative average Cα RMSDs of the mutant PreSMo peptides with respect to the wild-types containing prolines can be calculated by subtracting the average values of the wild-types from those of mutants. The Cα RMSDs for the NFR mutants were on the average increased by 0.302, but those for the CFR mutants decreased in the opposite direction by 0.439, indicating that the NFR mutants, i.e., removing a proline in the N-terminal flanking region of a helix, destabilized the helical nature when compared with the CFR mutants where a proline is replaced by other amino acids in the C-terminal flanking region of a helix. Figure 3 shows the most populated structures of each PreSMo peptide that are close to the centroid of the largest cluster obtained from MD simulations. The size of the largest

clusters increased if a proline was removed in CFR whereas the size decreased when we a proline is replaced by other amino acids in NFR. In case of Pro→Ala mutation in NFR, the population of the largest cluster decreased by 0.5% on the average, but for Pro→Asp mutation in NFR the population remained same as that in the wild-type. On the contrary all the mutations in CFR increased the population of the largest clusters more noticeably by 1.4~2.6%. In almost all cases the helical content of the largest cluster in the CFR mutants was higher than that in the wild-type. Such a tendency was not observed for the NFR mutants. These results indicate that the overall effect of the prolines present in CFR of the wild type PreSMo peptides is larger than that in NFR.

It should be noted that the structures shown in Figure 3 are not necessarily the same as the experimentally determined structures of PreSMos since we focused on only the most populated cluster of each PreSMo peptide. Figure 4 summarizes the residual alpha helical content for each residue. It can be seen that ~90% (36 out of 40) of cases the helical content in the CFR mutants is higher than that in the wild type and that ~70% (32 out of 40) of the NFR mutants show a lower helix content than the wild type. The overall summary of variation in the alpha helical content of the mutant PreSMo peptides with respect to the wild-type PreSMo is summarized in Figure 5.

3.2.1 p53 TAD

p53 TAD 1 (**I**) is a relatively well-defined helix. This PreSMo is the critical site in p53 that triggers many factors involved in tumor suppressing function [5,12,45]. When mdm2 blocks this site of p53 the tumor suppressive activity of p53 disappears [46]. Figure 3 shows that the helicity for the entire PreSMo-forming residues (residues 7-15) significantly decreases if a proline in its NFR is replaced by other amino acid residues. The helicity of NFR itself also decreases drastically except for the case of Lys mutant. On the contrary when a proline in CFR is replaced by Asp, His, Lys or Ala the helicity of this PreSMo increases by a large amount.

p53 TAD 2 (**II**) was originally found to form one turn of a helix by residues 40-44. The p53 TAD 2 PreSMo (**II**) peptide we studied here has 14 residues and hence forms a slightly longer PreSMo helix with two turns. With a proline mutation in its CFR the helicity of this PreSMo increases slightly. But, the CFR mutations involving Asp, His, or Ala do not influence the helicity of the NFR-forming residues. Interestingly, a lower helicity for the

NFR-forming residues is observed in the case of CFR Lys mutation. The helicity of the CFR-forming residues in CFR mutants is always higher than that of the wild type. In contrast, NFR proline mutations reduce the helicity of the PreSMo noticeably (40~60%), in particular, in the case of His and Lys mutants. The reduction is less pronounced (~10%) in Ala and Asp mutants. The helicity of the NFR residues also decreases significantly with N-terminal mutations, i.e., by ~60%, 20%, 40% and 10% respectively, for Asp, His, Lys and Ala mutation. On the other hand minimal reduction in the helicity of the CFR-forming residues is observed with N-terminal mutations.

p53 TAD 3 (**III**) is also a turn PreSMo, which by itself can bind to mdm2 or p62 [12,13]. When RPA binds to p53 TAD, p53 TAD 2 (**II**) and p53 TAD 3 (**III**) help the binding by forming a two-turn helix [14]. The effect of removing a proline from its NFR is not pronounced; the helicity of this PreSMo remains the same before and after mutations. On the other hand, the helicity of the entire PreSMo peptide including NFR, PR and CFR dramatically is enhanced due to the proline removal at its CFR.

3.2.2 Securin

Securin PreSMo 1 (**IV**) is 10 residues long (residues 150-159) with 4 glutamic acids. The wild type PreSMo peptide used for MD simulations is a 21-residue long peptide. This PreSMo peptide shows an interesting dip in its helicity within the PreSMo-forming residues (Figure 4I**V**). In three NFR mutants (Asp, His, and Ala) this dip disappears with a concomitant increase in the helicity of the PreSMo-forming residues. Removal of a proline in NFR decreases the helicity of the NFR-residues in two (Asp and Lys) mutants following a similar trend observed in p53 TAD. Yet, it is remarkable that the overall helicity increases when a proline is mutated to a histidine (the second panel from the left in Figure 4I**V**). This represents the only case showing such a behavior in the current study. As in p53 TAD the helicity of all three regions, NFR, PR and CFR, with His and Lys mutations in CFR is higher than that of the wild type. The helicity remains similar to that of the wild type in the case of Asp and Ala mutants in CFR. The helicity of the NFR-forming residues in Asp mutant is lower than the wild type.

Securin PreSMo 2 (**V**) is a short 5-residue PreSMo. A significant reduction in the overall helicity due to NFR mutations was noted in the case of His mutant. The other mutants have a similar degree of helicity as the wild type. In the case of CFR mutants a slight increase

in the helicity was observed except for the case of Lys mutant where the helicity decreased due to removal of a proline.

3.2.3 I-2

I-2 PreSMo (**VI**) is an 11-residue PreSMo having a ~50% degree of helix pre-population [28]. We have used an 18-residue long PreSMo peptide for calculations. It is notable that all the CFR mutations drastically increased the helicity of the PreSMo especially at the C-terminus of PR. For three mutants the NFR-forming residues did not experience any noticeable change in the helicity except for the case of Asp mutant which showed ~15% decrease in the helicity. The Asp mutation in NFR decreased the helicity of the NFR-forming residues, but not that of the PreSMo-forming residues. The other NFR mutations did not influence the helicity of the NFR-forming residues. The Lys mutation in NFR increased the helicity of the N-terminal half of the PreSMo-forming residues to a significant degree (~20%). For Ala mutant removal of a proline in NFR did not affect the structural stability of PreSMo whereas His mutation changed the helicity of the N-terminal 4 residues of the PreSMo by 10~50%.

3.2.4 ENSA

The ENSA PreSMo (**VII**) formed by residue 32-36 is ~ 40% pre-populated [4], and the length of the ENSA PreSMo peptide used for MD simulations is 17. The Asp and Ala mutants in NFR drastically reduced the helicity of NFR-, PR- and CFR-forming residues. But it did not change to any significant degree by His and Lys mutations in NFR. In the case of CFR mutants none of the mutations seems to affect the helicity noticeably.

3.2.5 CREB KID

The NFR mutations all seem to increase the helicity at the C-terminal half of the PreSMo-forming residues except for the case of Ala mutant. As for the CFR mutations two (Asp and Lys) mutations increased the helicity by a large amount (20~70%). But, His and Ala mutations seem to decrease the helicity by 10~50%.

3.2.6 DARPP-32

The DARPP-32 PreSMo (residues 22-29, **IX**) is the only PreSMo containing a

proline (residue 24) in the middle [28]. Although this PreSMo is ~50% populated by SSP calculations [4] it contains a kink within a helix. It is difficult to find any general trend with NFR mutations. In three CFR mutants (Asp, Lys, and Ala) the helicity seems to increase. However in His mutant in CFR the effect seems to be in the opposite direction at least for ~5 residues which show a lower helicity.

4. DISCUSSION

One of the characteristics observed in IDPs is a higher composition of disorder-promoting residues including proline than in globular proteins [26]. An apparent role of prolines is to render proteins flexible. But, prolines may play a specific structural role by stabilizing helical structures when they are placed at the N-terminus of a helix, i.e., act preferentially as a helix initiator at an N-terminus rather than a helix breaker [47]. For example, several computational studies on the helical pentadecapeptide, Ac-$(Ala)_n$-Pro-$(Ala)_m$-NHME, demonstrated that the highest positional propensity of a proline is found for the Ncap+1 position in a helical peptide [22,48,49]. Would such a positional propensity of prolines hold for the prolines flanking the helix-forming segments such as PreSMos in IDPs? Our study was designed to investigate if prolines in IDPs, in addition to such a general disorder-promoting role, may have a further specific role of influencing the length and the stability of the helix-forming PreSMos.

Figure 5 summarizes the variations in the alpha-helical contents associated with the proline mutations in the helix PreSMo peptides studied. When a proline in the NFR is replaced by other amino acids, the alpha helical contents of NFR, PR, and CFR of PreSMo peptides were all noticeably decreased (Figure 5(A)). Among all the mutants, the alpha helical contents in NFR of **I**-Ala, **II**-Ala, **VII**-Lys, **VII**-Ala, **II**-His, and **I**-Asp mutant peptides significantly decreased by -65.8%, -63.9%, -48.4%, -47.6%, -45.0%, and -42.9%, respectively. For the PRs, the alpha helical contents of **II**-His and **VII**-Ala changed by a similar degree, i.e., -58.7% and -51.7%, respectively. In case of CFR, the variations in the alpha helical contents for most mutants were below 25%. Also the average variation in NFR (-12.9%) is much higher than that in the other two regions; PR by ~ -3.1% and CFR by ~ -3.6% on the average. These results show that the prolines in NFR tend to serve as an alpha helix initiator/promoter in the two-thirds of the cases we have studied similar to the prolines preceding α-helices in globular proteins.

In contrast to the prolines in NFR the prolines in CFR of PreSMos take up an opposite role. As shown in Figure 5 we clearly observed that the alpha helical contents for the most part of the CFR and PR increased when a proline was removed in CFR. Among all the C-terminal mutants, the alpha helical contents in NFR of **V**-Ala, **II**-Asp, **VIII**-His mutants increased over +30% and for the PRs, those of **III**-Lys, **III**-Ala and **VI**-His were increased over +53.2%, +45.7%, and +41.0%, respectively. In case of CFR, the variations of alpha

helical contents for the most of mutants were significantly increased; more than 50% and 40% increase in **I**-Asp, **III**-Lys, **III**-Ala and in **VIII**-Ala, **II**-His, **IX**-Ala, **I**-Ala, and **II**-Asp, respectively. The average variations of alpha helical contents by the C-terminal mutations were -2.9%, +12.0% and +26.8% for NFR, PR, and CFR, respectively. Therefore, the prolines in CFR of PreSMo are definite α-helix breakers. Also the proline removal from CFR more effectively increased the helical contents of PreSMos than that in NFR. Placing two prolines in any terminus of a PreSMo appears to be a highly efficient means of forming short (1~3turns) helix PreSMos in IDPs. A proline residue in middle of an α-helix may introduce a bend or kink into a helix [49-52]. Among the PreSMo peptides we studied, only **IX** (DARPP-32) came under this category. But the helix kink or bend was not observed due to the short (3 residues) N-terminal region of **IX** (DARPP-32).

Undoubtedly, the helix is the most studied and understood secondary structure in proteins. Nonetheless, our knowledge on this important structural element still seems incomplete especially in terms of how its stability may be controlled by the capping residues at its termini (Ncap and Ccap) and by surrounding residues (N', N'' or C', C'' etc) which not only form hydrogen bonds but exert hydrophobic effects [24]. Even though several studies addressing this issue were carried out with the model peptides consisting of a simple sequence, e.g., all Ala, it is difficult to extend the results of such studies to our PreSMo peptides which have widely varying sequences. In addition, the flanking prolines we have studied the effects of are not the capping residues, being a few residues apart from the termini. Hence, the observed roles of PreSMo-flanking prolines, probably via more hydrophobic effects than hydrogen bonding, we provided in this study is more like an overall *comparative* tendency (wild type vs. mutants) especially at the early stage of peptide structure evolution noting the peptide conformation may fluctuate much further during a long simulation [53,54]. What seems interesting is that even with such a variation in the sequence the role of the CFR prolines as helix terminators seems to be rather clearly manifesting even at some long simulation times even though it is less pronounced than at short simulation times (Figure 6 and Table 4). On the other hand, the less clear tendency (~50% of the cases at a longer simulation time instead of ~70% observed at a short simulation) observed for the NFR prolines is not surprising if we appreciate the fact that the results are obtained for PreSMo peptides, not for a simple all-Ala model peptide. In addition, one must note that there is a certain uncertainty for the boundaries of a helix PreSMo determined by NMR.

IDP-target binding can be described by induced fit or conformational selection. Recent studies show that the binding would be better explained by a combination of an induced fit and a conformational selection mechanism in the case of the μ type IDPs which constitute ~70% structurally characterized IDPs [55-58]. PreSMos presage target-bound conformations of the binding segment and are specificity determinants for IDP recognition of the μ type IDPs by target proteins [4]. As our results show that the flanking prolines control the conformational stability of PreSMos the prolines that flank PreSMo- or α-MoRFs must have been designed to act as the "switches" that control the level of helical content, which in turn govern the degree of conformational selection by target proteins. Conformational selection mechanism is more energetically favorable than an induced fit. In the scenario of conformational selection, only a minor structural adjustment of PreSMos or local structures will be needed in binding process rather than a full induced fit/folding, where structural induction from random coil into a helix would have to occur after the binding segment is already seated into the binding pocket of target protein.

5. CONCLUSIONS

In summary, we have examined the effect of flanking prolines on the structural stability of helix-forming PreSMos in several μ type IDPs using molecular dynamics simulations. We observed that in the early stage of conformation evolution the alpha-helical contents were significantly decreased by replacement of proline residues with other amino acid residues in NFR, but were increased by the same mutations when they occur in CFR. Even though the role of prolines in the N-terminal region of a peptide as a helix initiator/promoter was examined by the various methods [59-64], this work is the first to our knowledge that examined the effect of proline mutations in the C-terminal region of peptides by MD simulations, in particular, in IDPs. We conclude that the proline residue in NFR may act as an alpha-helix initiator whereas that in CFR is a strong α-helix breaker in IDPs; the proline in CFR of a PreSMo was most likely selected during evolution as a "terminator" of helical PreSMos in IDPs. A helically restrained PreSMo by a proline residue in an IDP which would become a stable helix upon target binding may favorably follow the conformational selection mechanism during target binding. While the number of investigated PreSMos in this study is small the conclusion drawn from this study has provided a useful insight into the role of proline residues around PreSMos or similar helix-forming segments such as α-MoRFs. Further studies with a larger data set of PreSMos, with longer PreSMo peptides entailing longer regions of IDPs and with mutation of prolines into more amino acid residues examined for a longer simulation time than used in this study are expected to provide a more concrete picture regarding the role of prolines in IDPs.

6. ACKNOWLEDGEMENTS

This work was supported by Korea Research Council of Fundamental Science and Technology (KRCF) with a Korean-Hungarian Joint Laboratory Program [2010-88343 to P. T. & K.H.] & NRF [2010-0022224 to K.H]. The computing resources were supported by the strategic support program (KSC-2011-C2-15, KSC-2011-C3-13) of Korea Institute of Science and Technology Information (KISTI).

7. REFERENCES

[1] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. 293 (1999) 321-331.

[2] V.N. Uversky, A.K. Dunker, Understanding protein non-folding, Biochim. Biophys. Acta 1804 (2010) 1231-1264.

[3] P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions, Trends Biochem. Sci. 33 (2008) 2-8.

[4] S.-H. Lee, D.-H. Kim, J.J. Han, E.-J. Cha, J.-E. Lim, Y.-J. Cho, C. Lee, K.-H. Han, Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins, Curr. Prot. Pept. Sci. 13 (2012) 34-54.

[5] H. Lee, K.H. Mok, R. Muhandiram, K.-H. Park, J.-E. Suk, D.-H. Kim, J. Chang, , Y.C. Sung, K.Y. Choi, K.-H. Han, Local structural elements in the mostly unstructured transcriptional activation domain of human p53, J. Biol. Chem. 275 (2000) 29426-29432.

[6] C.M. Fletcher, G. Wagner, The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein, Protein Sci. 7 (1998) 1639-1642.

[7] H. Hemmings, A. Nairn, D. Aswad, P. Greengard, DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Purification and characterization of the phosphoprotein from bovine caudate nucleus, J. Neurosci. 4 (1984) 99-110.

[8] K. Gast, H. Damaschun, K. Eckert, K. Schulze-Forster, H.R. Maurer, M. Mueller-Frohne, D. Zirwer, J. Czarnecki, G. Damaschun, Prothymosin α: A biologically active protein with random coil conformation, Biochemistry 34 (1995) 13211-13218.

[9] V. Uversky, Intrinsically disordered proteins and their environment: Effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding, Protein J. 28 (2009) 305-325.

[10] V.N. Uversky, Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes, Chem. Soc. Rev. 40 (2011) 1623-1634.

[11] V.N. Uversky, Intrinsic disorder-based protein interactions and their modulators, Curr. Pharm. Design 23 (2013) 4191-4213.

[12] S.-W. Chi, S.-H. Lee, D.-H. Kim, M.-J. Ahn, J.-S. Kim, J.-Y. Woo, T. Torizawa, M. Kainosho, K.-H. Han, Structural details on mdm2-p53 interaction, J. Biol. Chem. 280 (2005) 38795-38802.

[13] P. Di Lello, L.M.M. Jenkins, T.N. Jones, B.D. Nguyen, T. Hara, H. Yamaguchi, J.D. Dikeakos, E. Appella, P. Legault, J.G. Omichinski, Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53, Mol. Cell 22 (2006) 731-740.

[14] E. Bochkareva, L. Kaustov, A. Ayed, G.-S. Yi, Y. Lu, A. Pineda-Lucena, J.C.C. Liao, A.L. Okorokov, J. Milner, C.H. Arrowsmith, A. Bochkarev, Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A, Proc. Natl. Acad. Sci. USA 102 (2005) 15412-15417.

[15] S. Mujtaba, Y. He, L. Zeng, S. Yan, O. Plotnikova, Sachchidanand, R. Sanchez, N.J. Zeleznik-Le, Z. Ronai, M.-M. Zhou, Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation, Mol. Cell 13 (2004) 251-263.

[16] S. Chuikov, J.K. Kurash, J.R. Wilson, B. Xiao, N. Justin, G.S. Ivanov, K. McKinney, P. Tempst, C. Prives, S.J. Gamblin, N.A. Barlev, D. Reinberg, Regulation of p53 activity through lysine methylation, Nature 432 (2004) 353-360.

[17] J.L. Avalos, I. Celic, S. Muhammad, M.S. Cosgrove, J.D. Boeke, C. Wolberger, Structure of a Sir2 enzyme bound to an acetylated p53 peptide, Mol. Cell 10 (2002) 523-535.

[18] E.D. Lowe, I. Tews, K.Y. Cheng, N.R. Brown, S. Gul, M.E.M. Noble, S.J. Gamblin, L.N. Johnson, Specificity determinants of recruitment peptides bound to phospho-CDK2/Cyclin A, Biochemistry 41 (2002) 15625-15634.

[19] D.-H. Kim, S.-H. Lee, K.H. Nam, S.-W. Chi, I. Chang, K.-H. Han, Multiple hTAFII31-binding motifs in the intrinsically unfolded transcriptional activation domain of VP16, BMB Rep. 42 (2009) 411-417.

[20] H.R.A. Jonker, R.W. Wechselberger, R. Boelens, G.E. Folkers, R. Kaptein, Structural properties of the promiscuous VP16 activation domain, Biochemistry 44 (2005) 827-839.

[21] S.-W. Chi, D.-H. Kim, S.-H. Lee, I. Chang, K.-H. Han, Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus, Protein Sci. 16 (2007) 2108-2117.

[22] A.A. Morgan, E. Rubenstein, Proline: The distribution, frequency, positioning, and

common functional roles of proline and poly proline sequences in the human proteome, PLoS ONE, 8 (2013) e53785.

[23] M.K. Kim, Y.K. Kang, Positional preference of proline in α-helices, Protein Sci. 8 (1999) 1492-1499.

[24] R. Aurora, G.D. Rose, Helix capping, Protein Sci. 7 (1998) 21-38.

[25] J.A. Marsh, J.D. Forman-Kay, Sequence determinants of compaction in intrinsically disordered proteins, Biophys. J. 98 (2010) 2383-2390.

[26] F.-X. Theillet, L. Kalmar, P. Tompa, K.-H. Han, P. Selenko, A.K. Dunker, G.W. Daughdrill, V.N. Uversky, The alphabet of intrinsic disorder I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins, Intrinsically Disordered Proteins 1 (2013) e24360.

[27] V. Csizmok, I.C. Felli, P. Tompa, L. Banci, I. Bertini, Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy, J. Am. Chem. Soc. 130 (2008) 16873-16879.

[28] B. Dancheck, A.C. Nairn, W. Peti, Detailed structural characterization of unbound protein phosphatase 1 inhibitors, Biochemistry 47 (2008) 12346-12356.

[29] J. Boettcher, K. Hartman, D. Ladror, Z. Qi, W. Woods, J. George, C. Rienstra, 1H, 13C, and 15N resonance assignment of the cAMP-regulated phosphoprotein endosulfine-alpha in free and micelle-bound states, Biomol. NMR Assign. 1 (2007) 167-169.

[30] J.M. Boettcher, K.L. Hartman, D.T. Ladror, Z. Qi, W.S. Woods, J.M. George, C.M. Rienstra, Membrane-induced folding of the cAMP-regulated phosphoprotein endosulfine-α, Biochemistry 47 (2008) 12357-12364.

[31] I. Radhakrishnan, G.C. Pérez-Alvarado, H.J. Dyson, P.E. Wright, Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB, FEBS Lett. 430 (1998) 317-322.

[32] I. Radhakrishnan, G.C. Pérez-Alvarado, D. Parker, H.J. Dyson, M.R. Montminy, P.E. Wright, Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions, Cell 91 (1997) 741-752.

[33] V. Vacic, C.J. Oldfield, A. Mohan, P. Radivojac, M.S. Cortese, V.N. Uversky, A.K. Dunker, Characterization of molecular recognition features, MoRFs, and their binding partners, J. Proteome Res. 6 (2007) 2351-2366.

[34] C.J. Oldfield, Y. Cheng, M.S. Cortese, P. Romero, V.N. Uversky, A.K. Dunker, Coupled folding and binding with α-helix-forming molecular recognition elements, Biochemistry 44 (2005) 12454-12470.

[35] A. Mohan, C.J. Oldfield, P. Radivojac, V. Vacic, M.S. Cortese, A.K. Dunker, V.N. Uversky, Analysis of molecular recognition features (MoRFs), J. Mol. Biol. 362 (2006) 1043-1059.

[36] D.A. Case, T. Darden, T.E. Cheatham III, C. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, et al., AMBER10, (2008) University of California, San Francisco.

[37] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712-725.

[38] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling, Electrophoresis 18 (1997) 2714-2723.

[39] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089.

[40] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327-341.

[41] W. Kabsch, C. Sander, How good are predictions of protein secondary structure? FEBS Lett. 155 (1983) 179-182.

[42] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graphics 14 (1996) 33-38.

[43] W.L. DeLano, PyMol. (2002) DeLano Scientific, San Carlos, CA.

[44] V. Vacic, V.N. Uversky, A.K. Dunker, S. Lonardi, Composition Profiler: a tool for discovery and visualization of amino acid composition differences, BMC Bioinformatics 8 (2007) 211.

[45] S. Shangary, S. Wang, Small-molecule inhibitors of the MDM2-p53 Protein-protein interaction to reactivate p53 function: A novel approach for cancer therapy, Annu. Rev. Pharmacol. Toxicol. 49 (2009) 223-241.

[46] P.H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A.J. Levine, N.P. Pavletich, Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor

transactivation domain, Science 274 (1996) 948-953.

[47] J. Richardson, D. Richardson, Amino acid preferences for specific locations at the ends of alpha helices, Science 240 (1988) 1648-1652.

[48] A.R. Viguera, L. Serrano, Stable proline box motif at the N-terminal end of α-helices, Protein Sci. 8 (1999) 1733-1742.

[49] R.H. Yun, A. Anderson, J. Hermans, Proline in α-helix: Stability and conformation studied by dynamics simulation, Proteins 10 (1991) 219-228.

[50] R. Sankararamakrishnan, S. Vishveshwara, Characterization of proline-containing α-helix (helix F model of bacteriorhodopsin) by molecular dynamics studies, Proteins 15 (1993) 26-41.

[51] R. Sankararamakrishnan, S. Vishveshwara, Conformational studies on peptides with proline in the right-handed α-helical region, Biopolymers 30 (1990) 287-298.

[52] R. Sankararamakrishnan, S. Vishveshwara, Geometry of proline-containing alpha-helices in proteins, Int. J. Pept. Prot. Res. 39 (1992) 356-363.

[53] M.A.M Ahmed, M. De Avila, E. Polverini, K. Bessonov, V.V. Bamm, G. Harauz, Solution NMR structure and molecular dynamics simulations of murine 18.5-kDa myelin basic protein segment (S72-S107) in association with dodecylphosphocholine micelles, Biochemistry 51 (2012) 7475-7487.

[54] E. Polverini, E.P. Coll, D.P. Tieleman, G. Harauz, Conformational choreography of a molecular switch region in myelin basic protein - Molecular dynamics shows induced folding and secondary structure type conversion upon threonyl phosphorylation in both aqueous and membrane-associated environments. Biochim. Biophys. Acta 1808 (2011) 674-683.

[55] J.-P. Changeux, S. Edelstein, Conformational selection or induced fit? 50 years of debate resolved, F1000 Biol. Rep. 3 (2011) 19.

[56] C.-J. Tsai, B. Ma, Y.Y. Sham, S. Kumar, R. Nussinov, Structured disorder and conformational selection, Proteins 44 (2001) 418-427.

[57] D.D. Boehr, R. Nussinov, P.E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, Nat. Chem. Biol. 5 (2009) 789-796.

[58] G.G. Hammes, Y.-C. Chang, T.G. Oas, Conformational selection or induced fit: A flux description of reaction mechanism, Proc. Natl. Acad. Sci. USA 106 (2009) 13737-13741.

[59] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, Nat.

Rev. Mol. Cell. Biol. 6 (2005) 197-208.

[60] G. von Heijne, Proline kinks in transmembrane α-helices, J.Mol. Biol. 218 (1991) 499-503.

[61] P. Tompa, Intrinsically unstructured proteins, Trends Biochem. Sci. 27 (2002) 527-533.

[62] D.N. Woolfson, R.J. Mortishire-Smith, D.H. Williams, Conserved positioning of proline residues in membrane-spanning helices of ion-channel proteins, Biochem. Biophys. Res. Commun. 175 (1991) 733-737.

[63] C.J. Brandl, C.M. Deber, Hypothesis about the function of membrane-buried proline residues in transport proteins, Proc. Natl. Acad. Sci. USA 83 (1986) 917-921.

[64] R.M. Kini, H.J. Evans, A hypothetical structural role for proline residues in the flanking segments of protein-protein interaction sites, Biochem. Biophys. Res. Commun. 212 (1995) 1115-1124.

TABLES

Table 1. The locations, amino acid sequences, and populations of PreSMos.

	IDP	Location including PreSMos	Pre-Population (%)[a]	Sequence (N-term. – *PreSMo* – C-term.)	Target protein	Ref
I	p53-TAD	12-28	20	P**P**LSQE–*TFSDLWKLL*–**P**E	Mdm2, RPA, TFEIIH	[5,12]
II		35-48	5	L**P**SQA–*MDDLM*–LS**P**D		
III		45-59	15	S**P**–*DDIEQW*–FTED**P**G		
IV	Securin	144-164	45	V**P**LMIL–*DEERELEKLF*–QLG**P**P	-	[27]
V		172-183	20	P**P**–*WESNL*–LQS**P**S		
VI	I-2	90-107	**48**(70)	A**P**DILA–*RKLAAAEGLE**P***–K	PP1	[28]
VII	ENSA	27-43	40	L**P**ER–*AEEAK*–LKAKY**P**S	-	[29,30]
VIII	CREB KID	131-147	**10**	R**P**S–*YRKILNDLSS*–DA**P**G	KIX	[31,32]
IX	DARPP-32	17-36	50	P**P**SQ–*LDPRQVEM*–IRRRR**P**T	PP1	[28]

a. Most of the populations were obtained from the SSP score calculation in ref 4. PreSMo populations obtained from the original papers were written in bold.

Table 2. The amino acid frequencies in the PreSMo and the flanking regions compared to the average frequencies of each amino acid in UniProt and the average amino acid frequencies in different parts of PreSMos, normalized to the average frequency of amino acids in disordered regions (in parenthesis).

Amino acid[a]	N-terminal flanking	PreSMo	C-terminal flanking	UniProt
Trp	0.0161 (-)[b]	0.0194 (3.076)	0.0111 (-)	0.0130
Phe	0.0323 (1.366)	0.0360 (1.523)	0.0056 (-)	0.0403
Tyr	0.0215 (1.092)	0.0194 (0.985)	0.0222 (1.127)	0.0304
Ile	0.0323 (1.016)	0.0277 (0.872)	0.0278 (0.875)	0.0601
Met	0.0215 (1.187)	0.0360 (1.988)	0.0222 (1.226)	0.0248
Leu	0.0860 (1.378)	0.1274 (2.041)	0.0889 (1.424)	0.0997
Val	0.0484 (0.852)	0.0388 (0.683)	0.0389 (0.685)	0.0679
Asn	0.0269 (0.754)	0.0526 (1.474)	0.0611 (1.713)	0.0411
Cys	0.0000 (-)	0.0000 (-)	0.0000 (-)	0.0123
Thr	0.0323 (0.582)	0.0388 (0.699)	0.0333 (0.600)	0.0557
Ala	0.0753 (0.928)	0.0776 (0.957)	0.0556 (0.685)	0.0710
Gly	0.0753 (1.042)	0.0332 (0.460)	0.0556 (0.770)	0.0710
Arg	0.0376 (0.819)	0.0859 (1.871)	0.0389 (0.847)	0.0544
Asp	0.0860 (1.373)	0.0886 (1.414)	0.0611 (0.975)	0.0532
His	0.0161 (-)	0.0111 (-)	0.0111 (-)	0.0220
Gln	0.0591 (1.138)	0.0360 (0.693)	0.0778 (1.498)	0.0400
Lys	0.0323 (0.387)	0.0582 (0.697)	0.0500 (0.599)	0.0526
Ser	0.1075 (1.226)	0.0637 (0.727)	0.0944 (1.077)	0.0662
Glu	0.0484 (0.480)	0.1302 (1.292)	0.0944 (0.937)	0.0618
Pro	0.1452 (1.886)	0.0194 (0.252)	0.1500 (1.949)	0.0467

[a] Residues are arranged according to their disorder propensity (from order promoting to disorder promoting).

[b] Due to the absence, or very low frequency of the residue (caused by the low sample number), some data were excluded

Table 3. Averaged amino acid frequency in different parts of α-MoRF.

Residue[a]	N-terminal region[b]	C-terminal region[c]
Trp	1.343	0.000
Phe	0.685	0.943
Tyr	0.730	0.914
Ile	0.777	0.917
Met	1.610	1.153
Leu	1.127	0.888
Val	0.814	0.815
Asn	1.427	1.441
Cys	0.721	1.192
Thr	1.102	0.759
Ala	0.669	0.745
Gly	0.662	0.821
Arg	1.212	1.577
Asp	1.373	1.267
His	1.050	0.842
Gln	0.615	1.043
Lys	0.812	0.911
Ser	1.594	1.291
Glu	0.982	0.748
Pro	1.016	1.328

[a] Residues are arranged according to their disorder propensity (from order to disorder promoting).

[b] Calculated as the residue specific average of five residues that flank the N-terminus of α-MoRF normalized to the amino acid frequency in all disordered proteins.

[c] Calculated as the residue specific average of five residues that flank the C-terminus of α-MoRF normalized to the amino acid frequency in all disordered proteins.

Table 4. The averaged variations of residual helical contents for short (10ns) and long (40~50ns) MD simulations. All values were obtained by subtracting residual helical contents of wild-type from those of mutants.

	NFR		PR		CFR	
	short	long	short	long	short	long
A. N-terminal mutation						
P53 TAD 1	-42	-5	-22	-24	0	0
P53 TAD 2	-36	-39	-32	-25	-8	-5
Securin 2	-7	-10	-13	-13	-13	-6
CREB KID	+6	+11	+14	+10	+8	+4
B. C-terminal mutation						
P53 TAD 1	-18	+1	+26	+4	+35	+42
P53 TAD 2	-6	-16	+7	-5	+37	+24
Securin 2	+2	+5	-2	+6	+20	+17
CREB KID	+22	+32	+9	+15	+28	+26

FIGURE LEGENDS

Figure 1. A schematic diagram for a PreSMo peptide composed of an N-terminal flanking region (NFR) prior to a PreSMo, a PreSMo region (PR), and a C-terminal flanking region (CFR). Both NFR and CFR are composed of 2~6 residues, one of which corresponds to a proline.

Figure 2. Relative average Cα RMSDs of mutant PreSMo peptides with respect to the wild-type PreSMo peptides. Each bar was obtained by subtracting the average Cα RMSD values of the wild-type PreSMo peptides from the average Cα RMSDs (in Å).

Figure 3. The representative structures of the largest clusters obtained by MD simulations for each PreSMo peptide. The residues constituting PreSMos are shown in green. The N- and C-termini are colored blue and red, respectively. The three letter code of amino acids under each structure indicates the residue inserted into the position of a proline in the flanking regions. The capital letters N and C in the parentheses denote the N- and C- terminal flanking regions where mutation of a proline was made. "WT" under each PreSMo peptide (**I**~**IX**) means the wild-type.

Figure 4. The residual helical content of each residue in the wild-type (WT, black), and the mutant (NFR mutants, blue; CFR mutants; red) PreSMo peptides. (from left to right: Asp, His, Lys, and Ala residues)

Figure 5. A summary of variations in the helical content of the residues belonging to NFR (left), PR (middle) and CFR (right) for NFR mutants (A) and CFR mutants (B). The range of variation is from -70% (decrease in helicity, blue) to 70% (increase in helicity, red).

Figure 6. Residual secondary structure contents during 40ns MD simulations of a few selected PreSMo peptides; (a) p53TAD 1, (b) p53TAD 2, (c) Securin 2, and (d) CREB KID. Black, blue, and red lines represent the wild-type (WT) polypeptide including PreSMo, NFR mutant, and CFR mutant, respectively. (From left to right: Asp, His, Lys, and Ala mutant).

FIGURES

Figure 1.

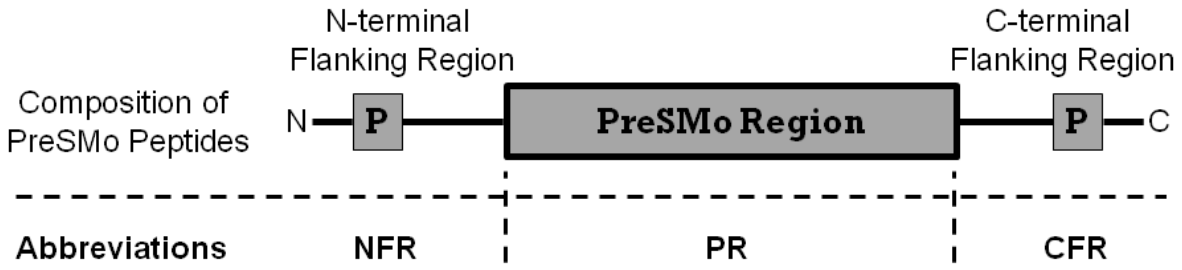

Figure 2.

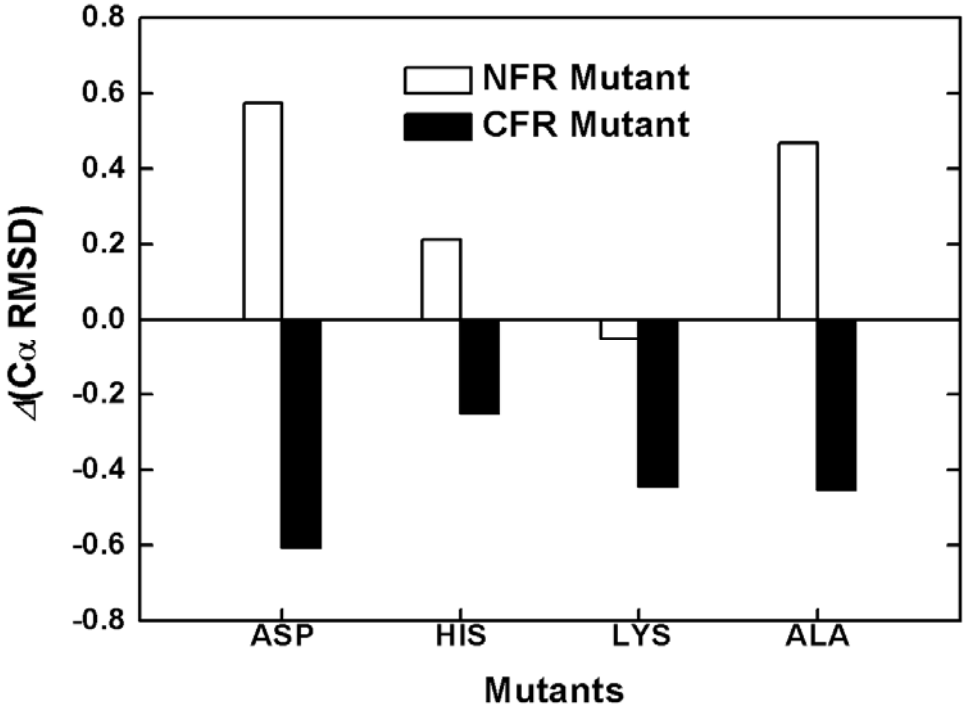
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
Δ(Cα RMSD)
NFR Mutant
CFR Mutant
ASP
HIS
LYS
ALA
Mutants

Figure 3.

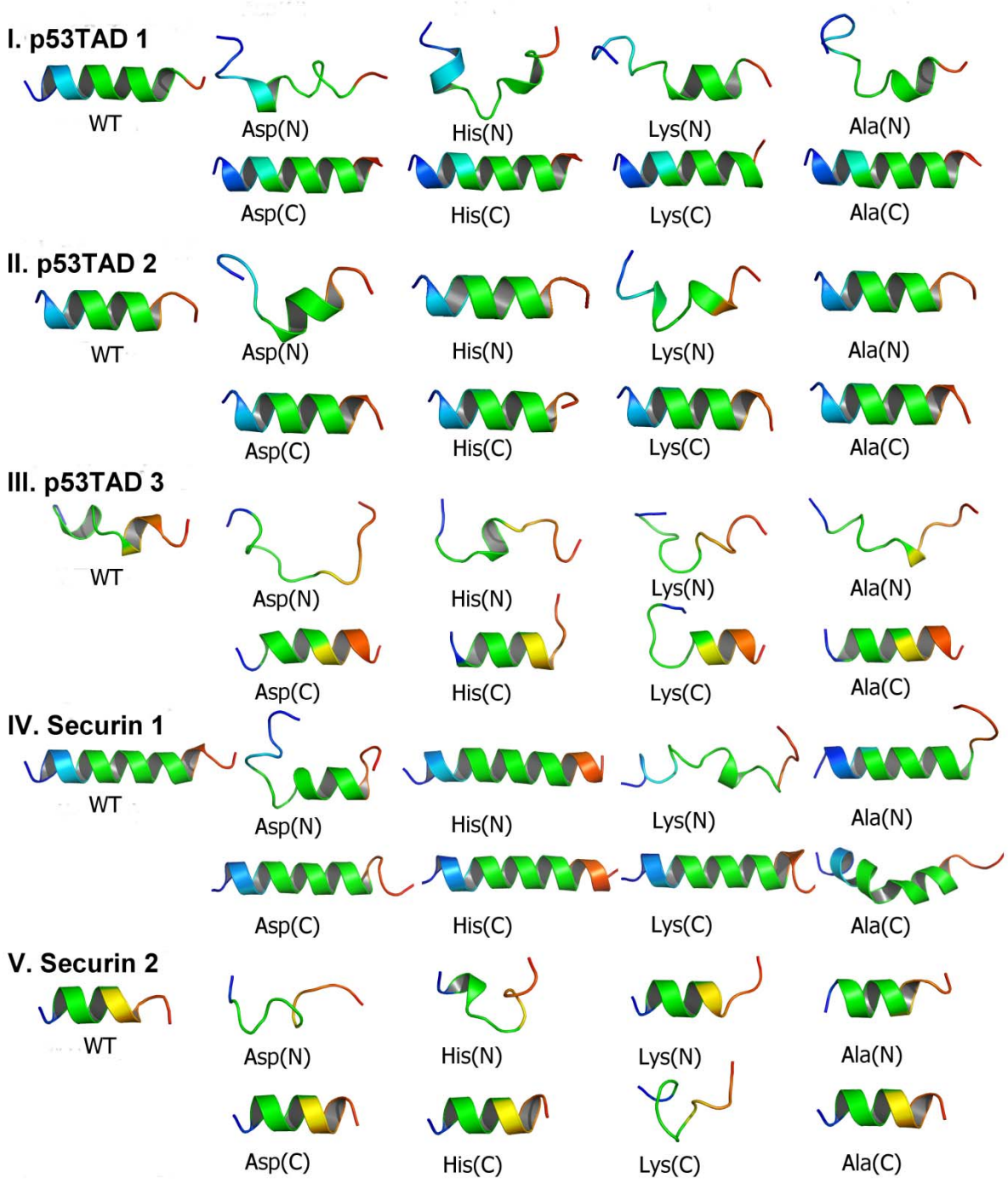
I. p53TAD 1
WT
Asp(N)
His(N)
Lys(N)
Ala(N)
Asp(C)
His(C)
Lys(C)
Ala(C)
II. p53TAD 2
WT
Asp(N)
His(N)
Lys(N)
Ala(N)
Asp(C)
His(C)
Lys(C)
Ala(C)
III. p53TAD 3
WT
Asp(N)
His(N)
Lys(N)
Ala(N)
Asp(C)
His(C)
Lys(C)
Ala(C)
IV. Securin 1
WT
Asp(N)
His(N)
Lys(N)
Ala(N)
Asp(C)
His(C)
Lys(C)
Ala(C)
V. Securin 2
WT
Asp(N)
His(N)
Lys(N)
Ala(N)
Asp(C)
His(C)
Lys(C)
Ala(C)

Figure 3. (Continued.)

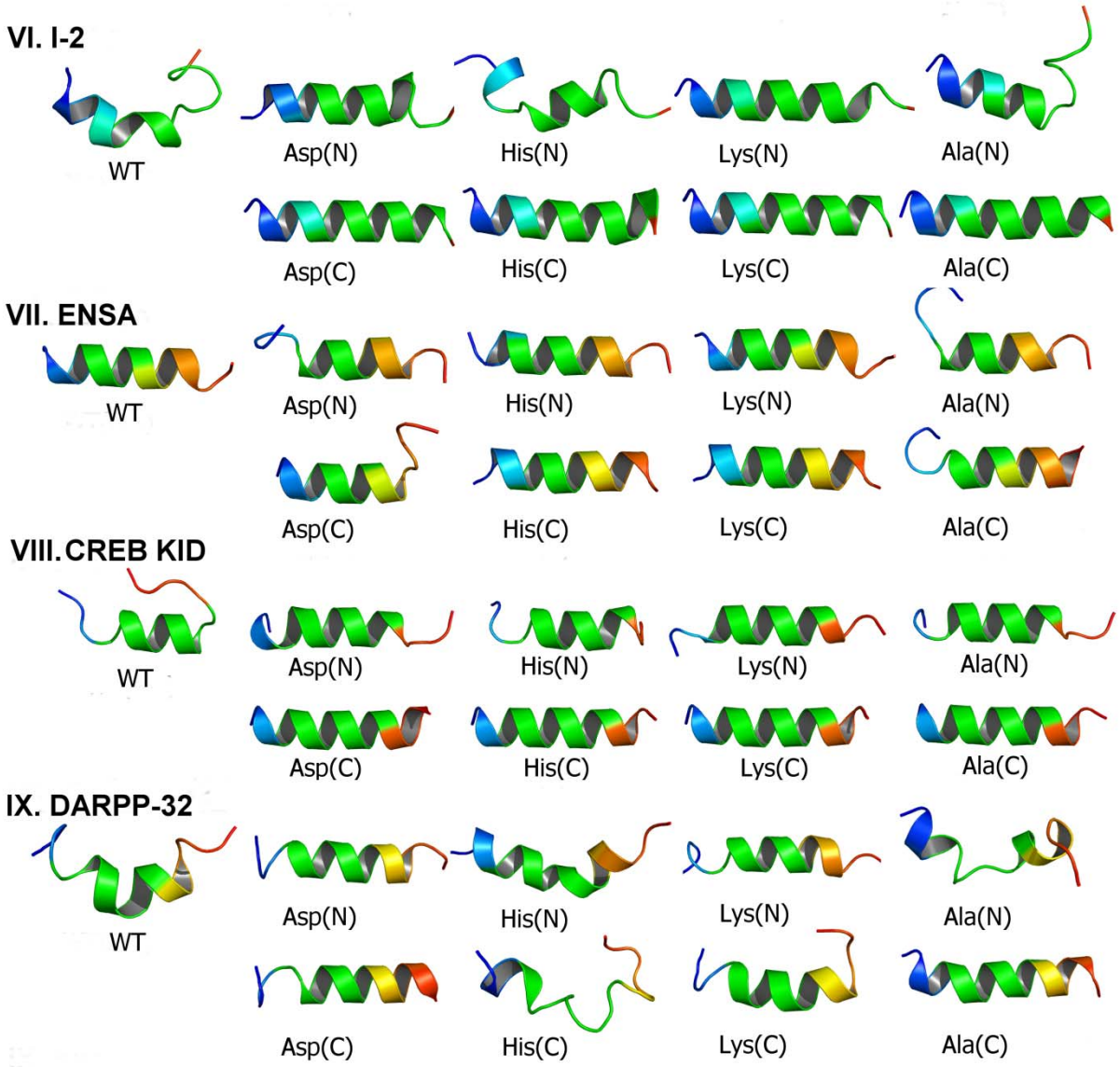

Figure 4.

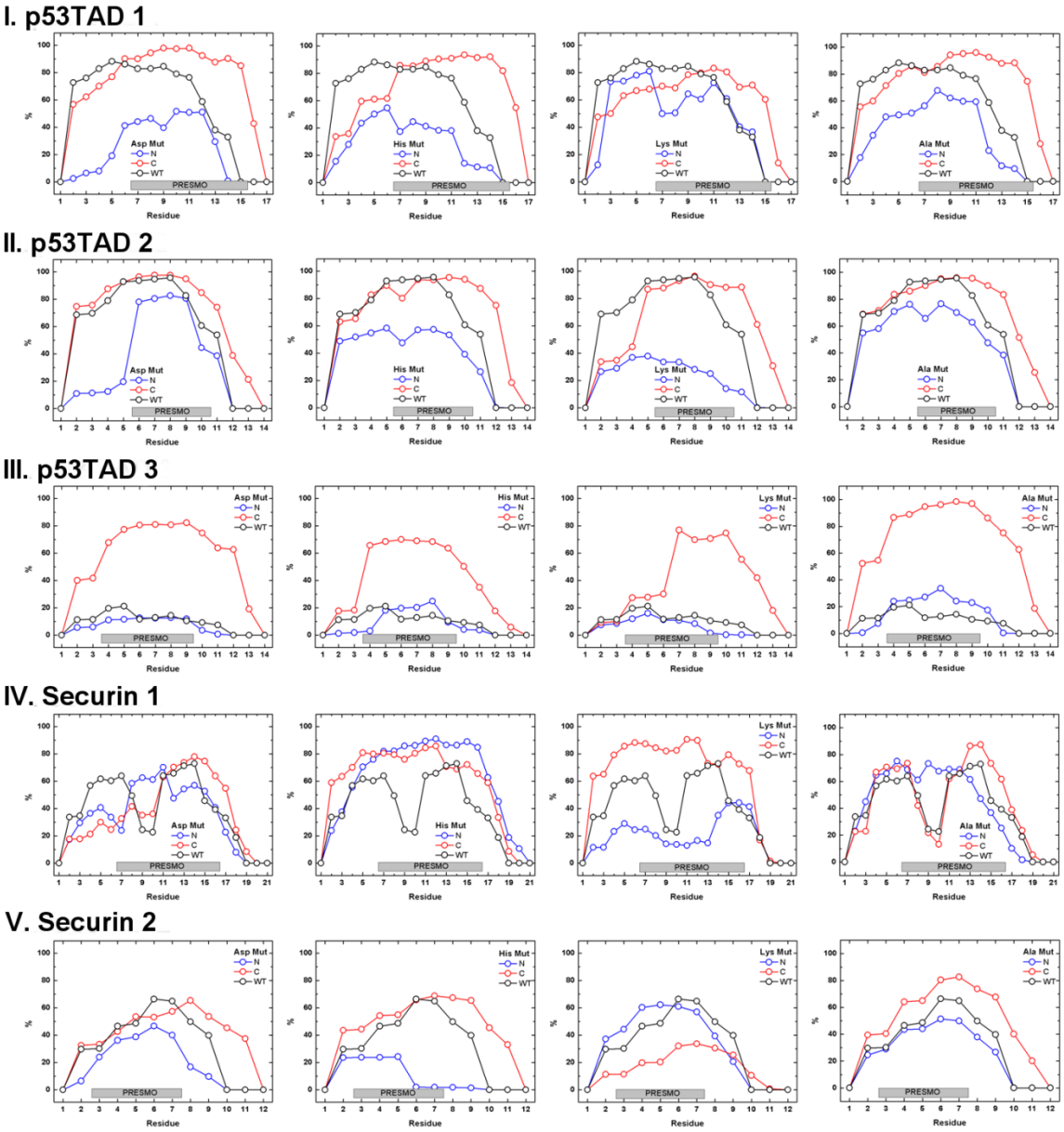
I. p53TAD 1
II. p53TAD 2
III. p53TAD 3
IV. Securin 1
V. Securin 2
Asp Mut
His Mut
Lys Mut
Ala Mut
N
C
WT
PRESMO
%
Residue

Figure 4. (Continued)

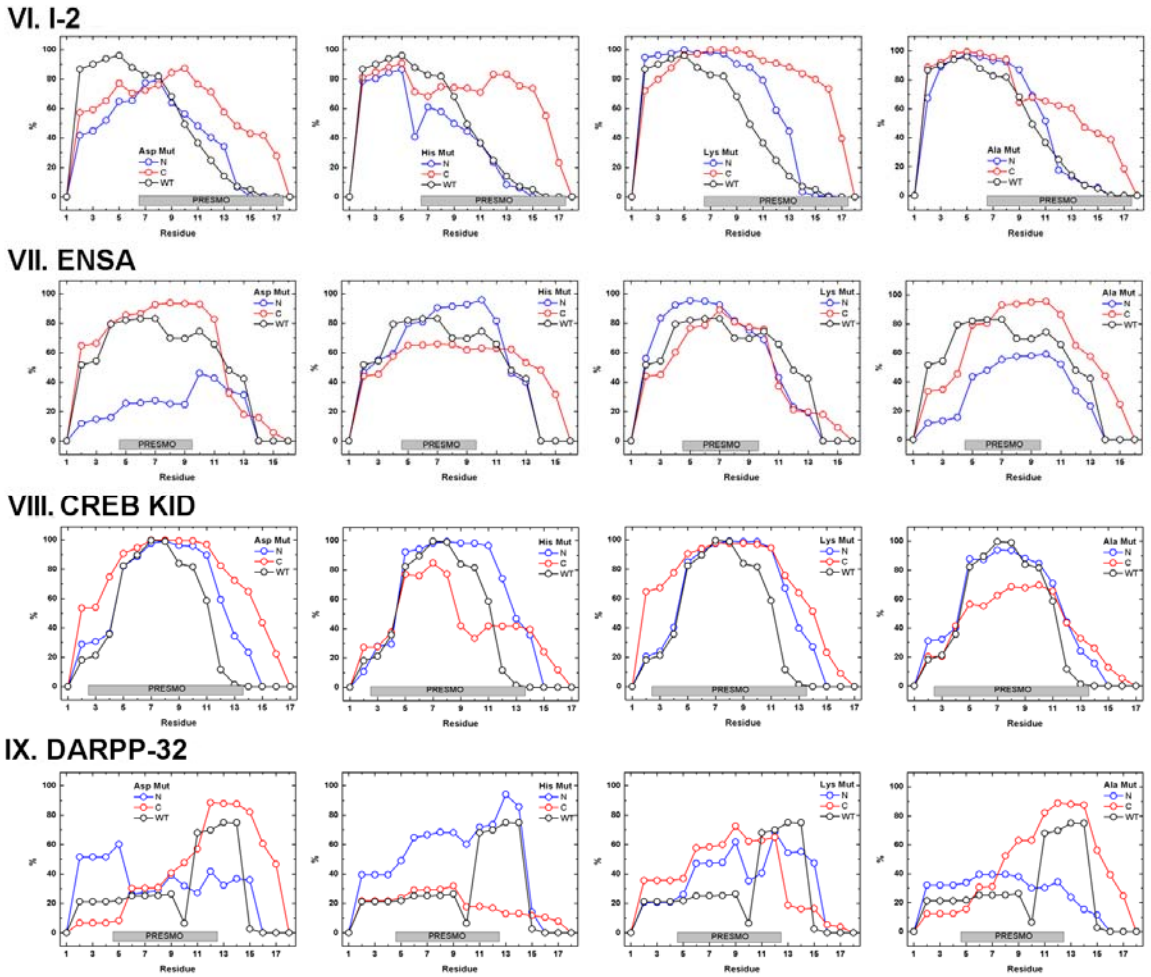

Figure 5.

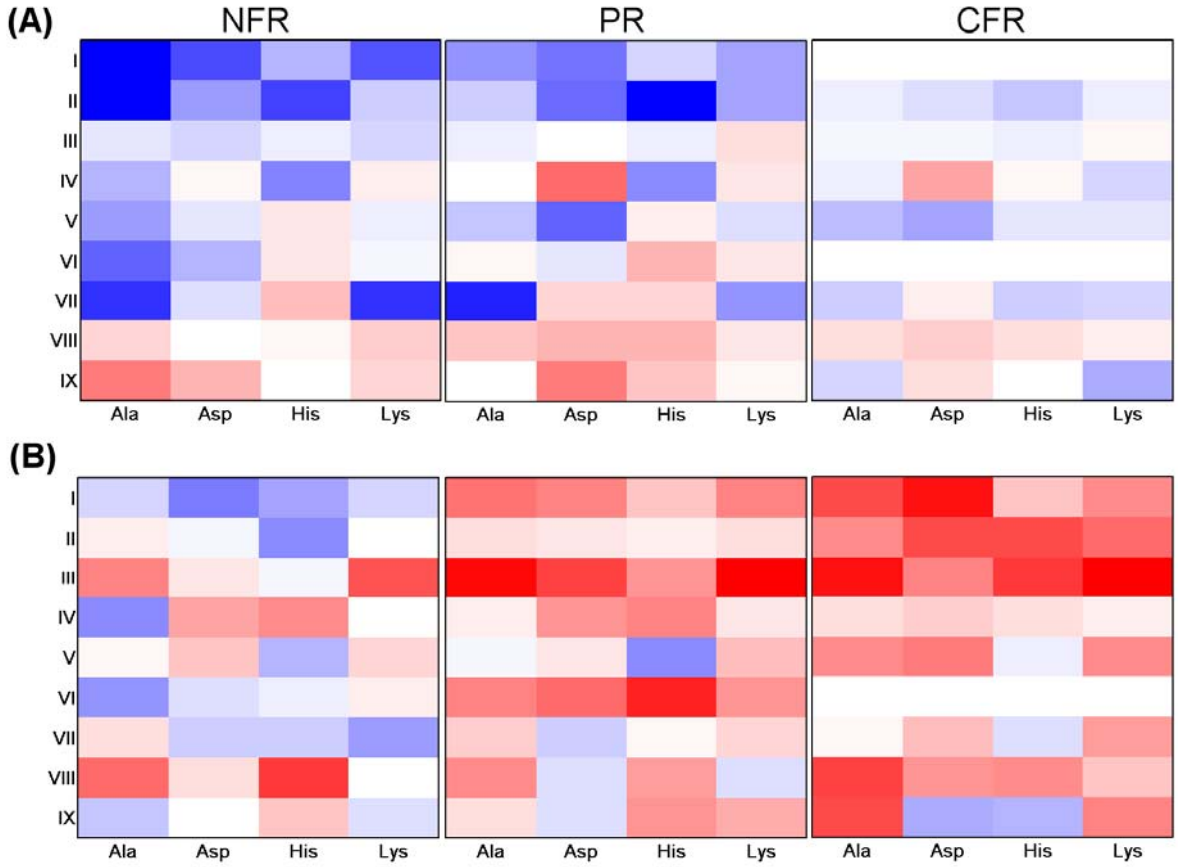

Figure 6.