HHS Public Access

Author manuscript

Pac Symp Biocomput. Author manuscript; available in PMC 2016 November 23.

Published in final edited form as:

Pac Symp Biocomput. 2012 ; : 94–103.

CORRELATION BETWEEN POSTTRANSLATIONAL MODIFICATION AND INTRINSIC DISORDER IN PROTEIN

JianJiong Gao[*] and **Dong Xu**[†]

Department of Computer Science, C.S. Bond Life Sciences Center, University of Missouri Columbia, Missouri 65211, USA

Abstract

Protein intrinsic disorder has been shown to play an important role in some posttranslational modifications (PTM). In this paper, we systematically investigated the correlation between protein disorder and dozens of PTMs using data from UniProt/Swiss-Prot and 3-D structures solved by NMR from Protein Data Bank. We observed that many PTMs have a preference for occurrence in disordered regions, including phospho-serine/-threonine/-tyrosine, hydroxylation, sulfotyrosine, S-geranylgeranyl cysteine, deamidated glutamine, 4-carboxyglutamate, 6'-bromotryptophan and most of methylation; while a few PTMs have a preference for occurrence in ordered regions, including 4-aspartylphosphate, S-nitrosocysteine, tele-methylhistidine, FMN conjugation, 4,5-dihydroxylysine, 3-methylthioaspartic acid, most of ADP-ribosylation, and most of FAD attachment. It is also noted that acetyllysine does not show any significant preference for occurrence in either disordered or ordered regions. Further analysis of NMR structures suggested disorder-to-order transitions might be introduced by modifications of phospho-serine/-threonine, mono-/di-/tri-methyllysine, sulfotyrosine, 4-carboxyglutamate, and potentially 4-hydroxyproline. This study sheds light on the functions and mechanisms of various PTMs.

1. Background

Almost all proteins undergo certain chemical modifications on their side chains, called posttranslational modifications (PTM) at some cellular state. Many PTM sites have been shown to occur in disordered regions. For example, it has been reported that phosphorylation was overrepresented in disordered regions;[1,2] the regions containing acetylated and methylated lysines in histone proteins was shown to be disordered;[3] methylated arginine was observed to be enriched in disordered regions;[4] various aspects of ubiquitination process were reported to occur predominately in disordered regions,[5] and protein disorder was suggested to facilitate hydroxylation of proline residues.[6] Large-scale studies on the relationship between protein disorder and PTM were also carried out previously. Pang et al. investigated the correlation between 44 types of PTMs and surface accessibility/disorder.[7] Xie et al. studied the correlations between predicted disorder and PTMs annotated by Swiss-Prot functional keywords, and they reported significant associations between PTMs and predicted protein disorder.[8] We noted that these two large-scale studies were both based on

[†]Corresponding author: xudong@missouri.edu.

[*]Current address: Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

computationally predicted disorder. While most disorder prediction tools achieve high accuracies when predicting long intrinsic disordered regions, they typically do not target prediction of short disordered regions. Besides, the current way to define disorder regions (residues with atoms missing in X-ray structures) for training these tools may not be reliable. Therefore, correlations between PTM and disorder implied from these prediction tools may have some bias.

In this study, we systematically investigated the correlations between various PTMs and protein disorder/flexibility. We utilized large-scale PTM annotations from UniProt/Swiss-Prot too but with more specific types and on a finer level than noted in previous studies (e.g. Xie et al). We excluded N/C-terminal modifications (e.g. N-linked acetylation and N-linked methylation, and C-linked amidation) for analysis, since terminal residues in general are inherently more flexible/disordered. In addition, to overcome the limitations of predicted disorder, we explored for the first time NMR 3-D structures in Protein Data Bank (PDB)[9] to study the relationship between PTM and disorder/flexibility. In this regard, we do not treat protein regions with binary states, i.e. order and intrinsic disorder. Instead, we characterize protein order/disorder using a continuous and quantitative measure, i.e. flexibility defined by the distribution of multiple structural models in the same PDB file. By combining results of predicted disorder and NMR structures, PTMs were categorized according to their correlations with protein disorder more reliably.

We also compared the disorder/order state before and after modifications occur. Previous studies suggested disorder-to-order transitions after modifications such as phosphorylation.[10] Here, we further studied the NMR data innovatively approaching this issue in a more systematical way thereby observing modification-induced disorder-to-order transitions for several PTMs.

2. Results

2.1. Correlation of PTM sites and their predicted disorder scores

Protein sequences and annotations of known PTMs were retrieved from UniProt/Swiss-Prot (release 2010_09)[11], using a bioinformatics tool, Musite.[2] Disordered regions were predicted for the retrieved proteins by applying a widely used protein disorder prediction tool VSL2B.[12] The disorder prediction scores of PTM sites were then extracted and compared with those of non-PTM sites, as shown in Table 1 with major findings summarized below:

- For phosphorylation, the average predicted disorder scores of phosphoserines, phosphothreonines and phosphotyrosines are significantly greater than those of their unmodified counterparts. However, phosphohistidine and 4-aspartylphosphate have significant lower mean disorder scores than their unmodified counterparts.
- We investigated 17 subtypes of methylation/dimethylation/trimethylation. Ten of them (including all 4 subtypes of dimethylation) have significantly greater mean predicted disorder scores, including cysteine methyl ester, asymmetric dimethylarginine, N6-methyllysine, symmetric

dimethylarginine, N6,N6-dimethyllysine, glutamate methyl ester (Glu), omega-N-methylarginine, Omega-N-methylated arginine, leucine methyl ester and N6-methylated lysine. Tele-methylhistine and S-methylcysteine have significantly lower mean disorder scores. N6,N6,N6-trimethyllysine does not have any significant difference in mean disorder score between PTM and non-PTM lysine residues.

- All 3 subtypes of hydroxylation (4-hydroxyproline, 3-hydroxyproline and 5-hydroxylysine) have significantly greater mean disorder scores.
- For acetylation, N2-acetylarginines have much greater mean scores than non-PTM arginine residues. N6-acetyllysine has lower mean scores than non-PTM lysine residues. Although the difference is statistically significant, the absolute value of difference is very small. Therefore, we assume that N6-acetyllysine does not have any significant preference for occurrence in either disordered or ordered regions.
- The residues with all 4 subtypes of ADP-ribosylation have lower mean disorder scores than their unmodified counterparts, but the difference for ADP-ribosylcysteine is not significant.
- For deamidation, deamidated glutamine has a significantly greater mean disorder score, while diamidated asparagine has a significantly lower one.
- For FAD attachment, tele-8alpha-FAD histidine and S-8alpha-FAD cysteine have significantly lower disorder scores.
- For FMN conjugation, both FMN phosphoryl threonine and S-4a-FMN cysteine have significantly lower mean disorder scores.
- 4-carboxyglutamate, S-geranylgeranyl cysteine, 6'-bromotryptophan and sulfotyrosine all have significantly greater mean disorder scores than their unmodified counterparts.
- 4,5-dihydroxylysine and S-nitrosocysteine both have significantly lower mean disorder scores than their unmodified counterparts.
- S-palmitoyl cysteine has a lower mean disorder score although the difference is not large.
- S-diacylglycerol cysteine does not have a significantly different mean disorder score.
- 3-methylthioaspartic acid has a significantly lower mean disorder score.

2.2. Correlation of PTM sites and their spatial fluctuations in NMR 3-D structures

We also investigated the flexibility of spatial fluctuations of PTM sites in protein 3-D structures determined by NMR spectroscopy. 7,714 NMR-based protein structures were retrieved from PDB (as of May 4th, 2011). Since the number of modified residues in 3-D structures is limited (as shown in Section 2.3), PTM annotations from UniProt/Swiss-Prot were mapped and aligned onto known protein structures, based on the mapping downloaded

from European Bioinformatics Institute (ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/text/pdb_chain_uniprot.lst). We excluded PTM sites that had any mismatch (i.e. with any different types of amino acids) between UniProt and PDB.

The spatial fluctuation score of a residue among multiple NMR models of the same protein was calculated based on the following equation:

$$F = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^T (X_i - \overline{X})}{n}} \quad (1)$$

where n is the number of models in the PDB file, X_i is a 3-element vector representing the 3-D coordinates of the C-alpha atom for the i^{th} model, and $\bar{X}$ is the mean vector of X_i's. Structures with less than 10 models (n<10) were excluded from the analysis. A residue with a larger spatial fluctuation typically results from sparse spatial restraints derived from nuclear Overhauser effects (NOEs) in NMR spectroscopy. Thus, the residue is more flexible and has higher tendency to be disordered. Conventionally, short mobile regions of proteins may not be considered as intrinsic disorder, which often refers to long protein fragments (>40 residues) that cannot be observed in X-ray crystallography. However, we believe the protein flexibility has a continuous spectrum, from highly rigid, to mobile and then to completely disordered. Therefore, we use the parameter F to characterize protein order/disorder, which can quantify the relationship between PTM and protein disorder continuously. The spatial fluctuations of PTM sites and non-PTM sites were then compared using student's t-test as shown in Table 2. Details are explained below:

- For phosphorylation, phospho-serine, -threonine and -tyrosine have significantly greater mean fluctuation scores than corresponding non-PTM residues; and 4-aspartylphosphate has a significantly lower one. These results are consistent with Table 1.
- The result for phosphohistinine is not consistent with the corresponding comparison in Table 1. The mean fluctuation score of phosphohistidine is greater than non-PTM histidine, mainly because the residue H243 in the structure with PDB accession of 1JOY (corresponding to the residue H243 of Swiss-Prot entry P0AEJ4) has very high fluctuation scores of 4.85 Å (chain A) and 5.11 Å (chain B). The fluctuation scores of residues corresponding to the other two phosphohistidines (H842 in PDB:1SR2 and H58 in PDB:1Y6D) are actually low (0.58 Å and 0.96 Å). Since the two high fluctuation scores may be outliers (which could explain the insignificant t-test p-value), one cannot make any inference until more data are available.
- The results for the subtypes of methylation are consistent with Table 1, except that asymmetric dimethylarginine does not have a significant p-value. Omega-N-methylarginine, N6-methyllysine, symmetric dimethylarginine and N6,N6-dimethyllysine have significantly greater

mean fluctuation scores than corresponding non-PTM residues. Similar to Table 1, N6,N6,N6-trimethyllysine has a mean fluctuation score very close to non-PTM lysine.

- Again, N6-acetyllysine has a mean fluctuation score that is almost the same as non-PTM lysine, providing more evidence that N6-acetyllysine may not have preference on either disordered or ordered regions.
- 6'-bromotryptophan and 4-carboxyglutamate have significantly greater mean fluctuation scores, which are consistent with Table 1.
- S-nitrosocysteine has a significantly lower mean fluctuation score, which is consistent with Table 1.
- Sulfotysine has a greater mean fluctuation score but not statistically significant. More data are needed for a significance test.
- S-palmitoyl cysteine has a significantly higher mean fluctuation score than non-PTM cysteine, which is inconsistent with Table 1. We noted that the disorder scores for both S-palmitoyl cysteine and corresponding non-PTM sites are relatively low in Table 1 (0.27, and 0.35, respectively). We observed from the structures (e.g. C422 in PDB:1Q68 and C5 in PDB: 1SPF) that S-palmitoyl cysteine tends to be in short, highly mobile regions (less than 15 amino acids), which are not considered as disordered regions by protein disorder prediction tools.
- The inconstancy of results for 4-hydroxyproline between Table 1 and Table 2 will be explained in Section 2.3.

2.3. Spatial fluctuation changes in 3-D structure due to PTM

Although sparse, there are some modified residues characterized in the NMR-based 3-D structures. To study the possibility of conformational changes after modifications, we separated PTM sites used in Section 2.2 into two groups: one group containing those PTM sites that are actually modified in structures; and the other group containing PTM sites (mapped from UniProt/Swiss-Prot) that are in pre-modified apo-forms in structures. For each type of PTM, we then compared the spatial fluctuations between the two groups, if there were cases available in both groups, as shown in Table 3. Interestingly, for all groups except 4-hydroxyproline, the modified residues have lower mean spatial fluctuations than unmodified residues. The differences for phosphoserine, phosphothreonine, N6-methyllysine, N6,N6-dimethyllysine, N6,N6,N6-trimethyllysine, sulfotyrosine and 4-carboxyglutamate are significant. This finding could indicate disorder-to-order transitions triggered by those PTMs. For 4-hydroxyproline, most of the PTM sites (65 out of 68) are modified in the structures (Table 3). The mean spatial fluctuation of the three residues in apo-form is lower than the 65 residues in modified form mainly because one of the three unmodified residues, i.e. P6 in PDB:2H8S, has a very low fluctuation score (0.32 Å). If we assume this residue is an outlier, 4-hydroxyproline could also follow the same disorder-to-order transition, which is also supported by the data in Tables 1 and 2. The disorder scores in Table 1 for pre-modified residues suggest 4-hydroxyprolines are very likely to occur on

disordered proline residues, and the spatial fluctuations in Table 2 for post-modified residues suggest that 4hydroxyprolines' transition are more ordered.

As an example, coagulation factor IX (UniProt accession of P00740) undergoes disorder-to-order transition in 3-D structure after gamma-carboxylation on glutamic acid, as shown in Figure 1. PDB entries 1CFI[13] and 1CFH[14] are two NMR solved structures for residues 1–47 of coagulaiton factor IX. 1CFI is heavily carboxylated, containing 12 4-carboxyglutamates (Figure 1(A)), while none of glutamic acid residues in 1CFH are carboxylated (Figure 1(D)). From the secondary structures (Figures 1(B) and 1(E)), it is obvious that 1CFI is substantially more ordered with increased helical content than 1CFH. From Figures 1(C) and 1(F), the structure fluctuation among NMR models for 1CFI is much lower than that of 1CFH.

3. Discussion

Based on observed correlations between PTMs and predicted disorder and spatial fluctuation in 3-D structures, we can divide PTMs into three categories: (1) PTMs that have preferences for occurrence in disordered regions; (2) PTMs that have preferences for occurrence in ordered regions; and (3) PTMs that have no obvious preferences for occurrence in either disordered or ordered regions. Some PTMs have positive correlations to both predicted disorder and high spatial fluctuation in NMR structures. Therefore, they have strong preferences for occurrence in disordered regions including phosphoserine, phosphothreonine, phosphotyrosine, omega-N-methylarginine, N6-methyllysine, symmetric dimethylarginine, N6,N6-dimethylarginine, 6'-bromotryptophan, and 4-carboxyglutamate. Some PTMs were observed to be positively correlated to both predicted order and low spatial fluctuation, including 4-aspartylphosphate and S-nitrosocysteine, and thus they are highly overrepresented in ordered regions. Many PTMs have significant positive correlations to predicted disorder only, but with no corresponding NMR structures available including cysteine methyl ester, glutamate methyl easter, N6-methylated lysine, leucine methyl easter, asymmetric dimethylarginine, omega-N-methylated arginine, 5-hydroxylysine, 3-hydroxyproline, N2-acetylarginine, deamidated glutamine, S-geranylgeranyl cysteine, and sulfotyrosine. Some PTMs have positive correlations to predicted order only but are also with no corresponding NMR structures available including tele-methylhistidine, ADP-ribosylasparagine, ADP-ribosylserine, ADP-ribosylarginine, tele-8alpha-FAD histidine, S-8alpha-FAD cysteine, 4,5-dihydroxylysine and 3-methylthioaspartic acid. N6-acetyllysine has no significant correlation to either predicted disorder or spatial fluctuation in 3-D structures. Further analysis of NMR structures also provided evidences of disorder-to-order transitions after modifications of phospho-serine/-threonine, mono-/di-/tri-methyllysine, sulfotyrosine, 4-carboxyglutamate, and potentially 4-hydroxyproline. Disorder-to-order transition could be a general mechanism that many PTMs use to control the functions of proteins. The 4-hydroxyproline residues have high mean predicted disorder but low mean spatial fluctuation. This could be due to disorder-to-order transition after hydroxylation and, therefore, 4-hydroxyproline may still target proline residues predominately in disordered regions.

It is noted that both data of disorder prediction and NMR structures have limitations: predicted disorder may have a certain level of inaccuracy depending on the training data and algorithm, and NMR structures may have bias since the models of some NMR structures may have been selected in an *ad hoc* way by the experimentalists. By combining both data, we hope to better assess the results and hence gain more credibility with consistent results.

Most of the results in this paper are novel findings. It is worth mentioning the differences between this study and previous ones by Pang et al.[7] and Xie et al.[8] These two studies correlated PTMs with predicted disorder/order, while this study departed from tradition to take advantage of NMR structures to verify the correlations and to investigate PTM-induced disorder-to-order transitions. Xie et al. only reported results for general types of PTMs (e.g. methylation and phosphorylation). In contrast, we investigated PTMs with many specific types and subtypes. We found that subtypes of PTMs (e.g. N6-methyllysine and tele-methylhistidine) in the same general type (e.g. methylation) could have different correlations to disorder regions. All these findings provided useful insight into the mechanisms of various PTMs and may facilitate further investigations into the structural and functional implications of these PTMs.

Acknowledgments

We like to thank Dr. A. Keith Dunker for helpful discussions. This work was supported by the funding from the National Institute of Health (grant number R21/R33 GM078601) and National Science Foundation (grant number DBI-0604439).

References

1. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. Nucleic Acids Res. 2004; 32(3):1037. [PubMed: 14960716]
2. Gao J, Thelen JJ, Dunker AK, Xu D. Mol. Cell. Proteomics. 2004; 9(12):2586.
3. Hansen JC, Lu X, Ross ED, Woody RW, Biol J. Chem. 2006; 281(4):1853.
4. Daily KM, Radivojac P, Dunker AK. IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. 2005:475.
5. Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebl MG, Iakoucheva LM. Proteins. 2010; 78(2):365. [PubMed: 19722269]
6. Hon WC, Wilson MI, Harlos K, Claridge TDW, Schofiel CJ, Pug CW, Maxwell PH, Ratcliffe PJ, Stuart DI, Jones EY. Nature. 2002; 417(6892):975. [PubMed: 12050673]
7. Pang CNI, Hayen A, Wilkins MR. J Proteome Res. 2007; 6(5):1833. [PubMed: 17428077]
8. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. J Proteome Res. 2007; 6(5):1917. [PubMed: 17391016]
9. Rose PW, Beran B, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlic A, Quesada M, Qulinn GB, WestBrook JD, Young J, Yukich B, Zardecki C, Berman HM, Bourne PE. Nucleic Acids Res. 2011; 39:D392. [PubMed: 21036868]
10. Lilia MI, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. Nucleic Acids Res. 2004; 32(3):1037. [PubMed: 14960716]
11. Farriol-Mathis N, Garavelli JS, Boeckmann B, Duvaud S, Gasteiger E, Gateau A, Veuthey AL, Bairoch A. Proteomics. 2004; 4(6):1537. [PubMed: 15174124]
12. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Proteins. 2005; 61(S7):176.
13. Freedman SJ, Furie BC, Furie B, Baleja JD, Biol J. Chem. 1995; 270:7980.
14. Freedman SJ, Furie BC, Furie B, Baleja JD. Biochemistry. 1995; 34:12126. [PubMed: 7547952]

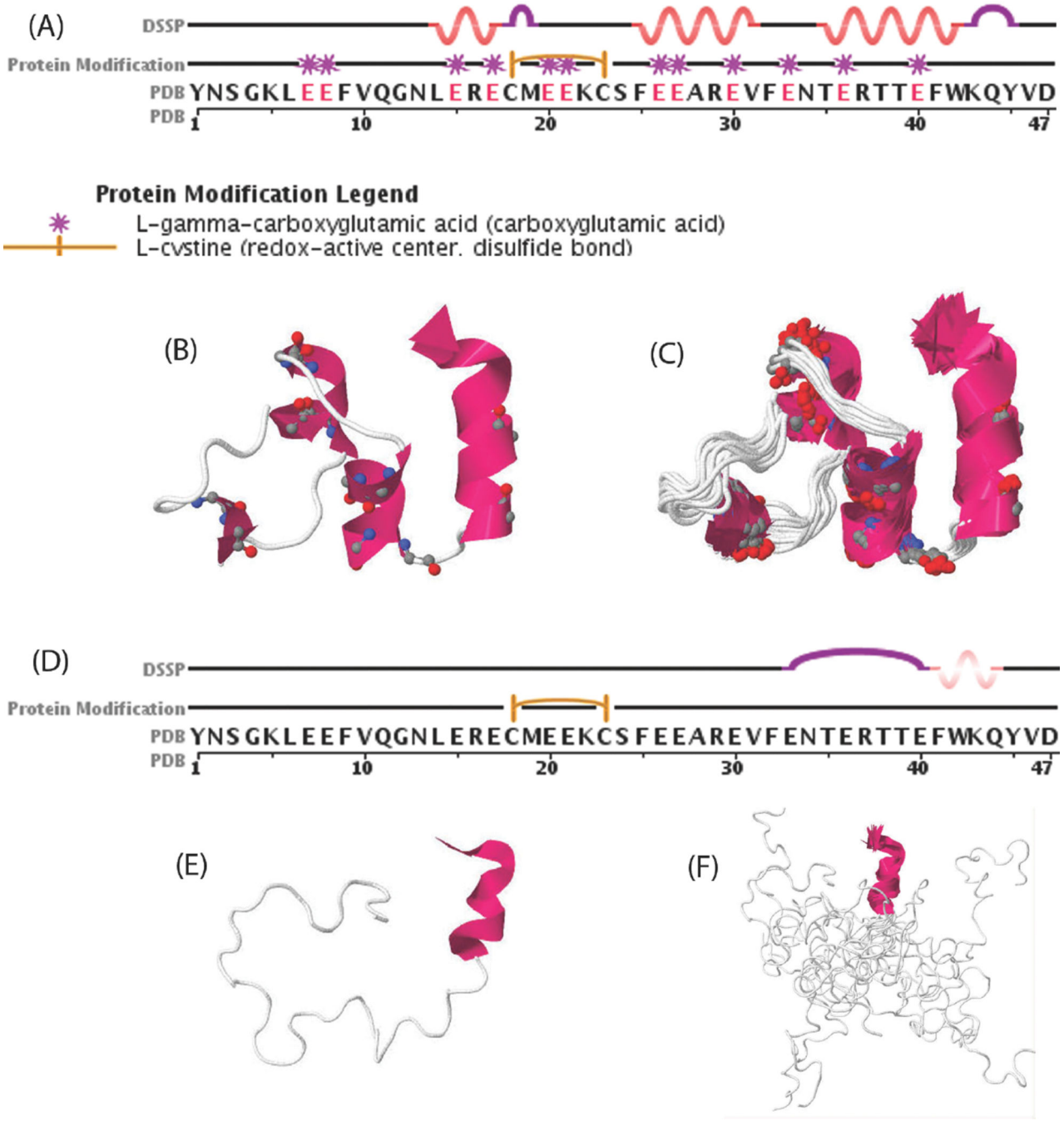

Fig. 1.

Disorder-to-order transition after gamma-carboxylation on coagulation factor IX (UniProt accession P00740). (A) Sequence and secondary structures of PDB 1CFI (heavily carboxylated); (B) 3-D visualization of the first model in 1CFI; (C) 3-D visualization of all 17 models overlaid in 1CFI; (D) Sequence and secondary structures of PDB 1CFH (with no carboxylation); (E) 3-D visualization of the first model in 1CFH; (F) 3-D visualization of all 16 models overlaid in 1CFH.

Pac Symp Biocomput. Author manuscript; available in PMC 2016 November 23.

Table 1

Comparison of predicted disorder scores between PTM sites and non-PTM sites. Predicted disorder score rages from 0 to 1. Higher score mean larger possibility of being in a disordered region. The p-value for the significance of the disorder score difference is based on two-tailed, two-sample student t-test.

				Mean fluctuation (Å)			
PTM	**UniProt keyword**	**# of PTM sites**	**Amino acid**	**PTM sites**		**Non-PTM sites**	**p-value**
Phosphorylation	Phosphoserine	53080	S	0.81	>	0.59	0
Phosphorylation	Phosphothreonine	13477	T	0.69	>	0.49	0
Phosphorylation	Phosphotyrosine	8075	Y	0.46	>	0.35	7.58E-287
Phosphorylation	Phosphohistidine	357	H	0.40	<	0.47	2.28E-14
Phosphorylation	4-aspartylphosphate	529	D	0.20	<	0.51	1.74E-240
Methylation	Cysteine methyl ester	491	C	0.78	>	0.35	6.82E-199
Methylation	N6-methyllysine	499	K	0.78	>	0.56	2.22E-64
Methylation	Glutamate methyl ester	40	E	0.84	>	0.58	8.63E-21
Methylation	Omega-N-methylarginine	98	R	0.81	>	0.52	1.01E-20
Methylation	Leucine methyl ester	14	L	0.69	>	0.41	5.59E-11
Methylation	N5-methylglutamine	10	Q	0.57	>	0.55	9.22E-02
Methylation	Pros-methylhistidine	10	H	0.37	<	0.47	1.07E-01
Methylation	N4-methylasparagine	65	N	0.46	<	0.49	6.14E-02
Methylation	5-methylarginine	11	R	0.46	<	0.53	5.30E-02
Methylation	S-methylcysteine	12	C	0.26	<	0.35	3.14E-02
Methylation	Tele-methylhistidine	47	H	0.12	<	0.47	7.77E-167
Dimethylation	Asymmetric dimethylarginine	385	R	0.90	>	0.52	3.67E-171
Dimethylation	Symmetric dimethylarginine	57	R	0.91	>	0.52	3.76E-25
Dimethylation	N6,N6-dimethyllysine	340	K	0.72	>	0.56	8.13E-25
Dimethylation	Omega-N-methylated arginine	247	R	0.72	>	0.52	1.04E-18
Trimethylation	N6-methylated lysine	374	K	0.69	>	0.56	1.36E-18
Trimethylation	N6,N6,N6-trimethyllysine	696	K	0.57	>	0.56	1.92E-01
Hydroxylation	4-hydroxyproline	900	P	0.89	>	0.60	2.27E-236
Hydroxylation	5-hydroxylysine	179	K	0.97	>	0.56	8.65E-136

PTM	UniProt keyword	# of PTM sites	Amino acid	Mean fluctuation (Å)			p-value
				PTM sites		Non-PTM sites	
Hydroxylation	3-hydroxyproline	11	P	1.00	>	0.60	6.28E-128
Acetylation	N2-acetylarginine	32	R	0.67	>	0.52	3.09E-05
Acetylation	N6-acetyllysine	11436	K	0.54	<	0.56	6.83E-09
ADP-ribosylation	ADP-ribosylcysteine	12	C	0.30	<	0.35	3.79E-01
ADP-ribosylation	ADP-ribosylasparagine	8	N	0.15	<	0.49	1.71E-06
ADP-ribosylation	ADP-ribosylserine	7	S	0.46	<	0.59	4.80E-10
ADP-ribosylation	ADP-ribosylarginine	87	R	0.18	<	0.52	3.26E-78
Deamidation	Deamidated glutamine	43	Q	0.77	>	0.55	4.86E-20
Deamidation	Deamidated asparagine	74	N	0.42	<	0.49	3.80E-02
FAD attachment	Pros-8alpha-FAD histidine	21	H	0.54	>	0.47	4.39E-02
FAD attachment	Tele-8alpha-FAD histidine	36	H	0.33	<	0.47	2.04E-11
FAD attachment	S-8alpha-FAD cysteine	52	C	0.21	<	0.35	2.40E-23
FMN conjugation	FMN phosphoryl threonine	33	T	0.37	<	0.49	3.01E-10
FMN conjugation	S-4a-FMN cysteine	21	C	0.20	<	0.35	6.46E-13
Carboxylation	4-carboxyglutamate	647	E	0.65	>	0.58	1.66E-19
Geranyl-geranylation	S-geranylgeranyl cysteine	29	C	0.81	>	0.35	3.20E-12
Palmitoylation	S-palmitoyl cysteine	155	C	0.29	<	0.35	1.03E-02
S-diacylglycerol cysteine	S-diacylglycerol cysteine	37	C	0.39	>	0.35	2.36E-01
S-Nitrosylation	S-nitrosocysteine	56	C	0.27	<	0.35	8.63E-06
Sulfation	Sulfotyrosine	561	Y	0.58	>	0.35	2.87E-77
Dihydroxylation	4,5-dihydroxylysine	8	K	0.32	<	0.56	3.93E-05
Bromination	6′-bromotryptophan	24	W	0.46	>	0.30	1.21E-03
Beta-methylthiolation	3-methylthioaspartic acid	503	D	0.38	<	0.51	0

Pac Symp Biocomput. Author manuscript; available in PMC 2016 November 23.

Table 2

Comparison of spatial fluctuations in 3-D structure between PTM sites and non-PTM sites. PTM annotations in UniProt/Swiss-Prot were mapped onto NMR-based 3-D structures. PTMs with one or more sites are included. The p-value for the significance of the fluctuation difference is based on two-tailed, two-sample student t-test.

				Mean fluctuation (Å)			
PTM	**UniProt keyword**	**# of PTM sites**	**Amino acid**	**PTM sites**		**Non-PTM sites**	**p-value**
Phosphorylation	Phosphoserine	585	S	1.88	>	1.16	1.3E-21
Phosphorylation	Phosphothreonine	131	T	1.62	>	1.05	7.0E-5
Phosphorylation	Phosphotyrosine	164	Y	1.10	>	0.91	0.037
Phosphorylation	Phosphohistidine	4	H	2.87	>	1.01	0.224
Phosphorylation	4-aspartylphosphate	6	D	0.57	<	1.10	4.0E-5
Methylation	Omega-N-methylarginine	4	R	3.20	>	1.07	0.021
Methylation	N6-methyllysine	28	K	1.51	>	1.07	0.034
Dimethylation	Symmetric dimethylarginine	3	R	2.30	>	1.07	0.008
Dimethylation	N6,N6-dimethyllysine	13	K	2.11	>	1.07	0.011
Dimethylation	Asymmetric dimethylarginine	4	R	2.16	>	1.07	0.163
Trimethyllysine	N6,N6,N6-trimethyllysine	32	K	1.00	<	1.07	0.689
Hydroxylation	4-hydroxyproline	68	P	0.73	<	1.20	6.4E-13
Acetylation	N6-acetyllysine	395	K	1.02	<	1.07	0.348
Bromination	6′-bromotryptophan	5	W	1.50	>	0.86	0.150
Carboxylation	4-carboxyglutamate	62	E	2.13	>	1.07	7.3E-8
Palmitoylation	S-palmitoyl cysteine	9	C	2.24	>	0.87	0.004
S-Nitrosylation	S-nitrosocysteine	10	C	0.56	<	0.87	1.4E-4
Sulfation	Sulfotyrosine	6	Y	1.02	>	0.91	0.102

Table 3

Comparison of spatial fluctuations of PTM sites in 3-D structure before and after modifications. PTMs with one or more modified residue in NMR-based 3-D structure are included. The p-value for the significance of the fluctuation difference is based on two-tailed, two-sample student t-test.

			Modified in structure		**Pre-modified in structure**		
PTM	UniProt keyword	Amino acid	# of sites	Mean fluctuation (Å)	# of sites	Mean fluctuation (Å)	p-value
Phosphorylation	Phosphoserine	S	9	0.98	576	1.90	8.2E-4
Phosphorylation	Phosphothreonine	T	15	1.07	116	1.69	0.022
Phosphorylation	Phosphotyrosine	Y	3	0.93	161	1.10	0.447
Methylation	N6-methyllysine	K	10	0.94	18	1.84	0.006
Dimethylation	N6,N6-dimethyllysine	K	5	0.84	8	2.90	2.6E-04
Trimethylation	N6,N6,N6-trimethyllysine	K	12	0.71	20	1.18	0.073
Hydroxylation	4-hydroxyproline	P	65	0.74	3	0.71	0.917
Acetylation	N6-acetyllysine	K	7	0.95	388	1.02	0.442
Sulfation	Sulfotyrosine	Y	4	0.93	2	1.19	0.001
Carboxylation	4-carboxyglutamate	E	49	1.76	13	3.52	0.004

Pac Symp Biocomput. Author manuscript; available in PMC 2016 November 23.