

The pattern of co-existed posttranslational modifications-A case study

Zheng-Rong Yang

School of Biosciences, University of Exeter, Correspondence should be addressed to Zheng- Rong Yang (Z.R.Yang@exeter.ac.uk)

Received August 20th, 2008; revised December 1st, 2008; accepted December 5th, 2008

ABSTRACT

Posttranslational modifications are a class of important cellular activities in various biochemical processes including signalling transduction, gene/metabolite networks, and disease development. It has been found that multiple posttranslational modifications with the same or different modification residues can co-exist in the same protein and this co-occurrence is critical to signalling networks in cells. Although some biological studies have spotted this phenomenon, little bioinformatics study has been carried out for understanding its mechanism. Four data sets were downloaded from NCBI for the study. The joint probabilities of any two neighbouring posttranslational modification sites of different modification residues were analyzed. The Bayesian probabilistic network was derived for visualizing the relationship between a target modification and the contributing modifications as the predictive factors.

Keywords: posttranslational modification, bioinformatics, pattern analysis, posttranslational modification, pattern analysis, probability model

1. INTRODUCTION

Posttranslational modifications (PTMs) are a chemical process of modifying a protein's chemical or structural properties after the translation of the protein has been completed. Posttranslational modifications are closely related with signalling networks of molecules in cells. The modifications include attaching a chemical, chemical structural change of amino acids, or protein structural change. The modification will also alter the functions of a protein in both ways, adding functions or removing functions, for instance, phosphorylation and dephosphorylation, carboxylation and decarboxylation. Because of these changes, proteins will carry different *signals* for functioning in cells. Posttranslational modifications are then the focus of many signalling transduction network studies. For instance, properly folded and posttransla-

tional modified endoplasmic reticulum is related with stress and will lead to different pathological states [1]. Poly (ADP-ribosylation) is related with DNA repair and cell cycle checkpoint pathways as the unique signal for protein function modulations [2]. Posttranslational modifications are the mediators for the transporters for multiple functions of human copper-transporting ATPases [3]. In the study of various chronic diseases, it is found that protein 3-nitrotyrosine (nitration) plays an important role in pathological conditions [4]. Together mutations and aberrant mRNA splicing, hyperphosphorylation will lead to a number of neurodegenerative disorders [5]. S-Nitrosation has been recently found to have similar function as phosphorylation and acetylation because of its association with various pathological cell reactions in signalling networks [6]. In cell-cycle control, differentiation, metabolism, stress response and programmed cell-death, the FOXO subgroup of forkhead transcription factors have been found being tightly controlled by phosphorylation, acetylation and ubiquitination [7].

In biological experiments, it has been found that the co-occurrence of posttranslational modifications is critical for many cellular functions in recent a few years. For instance, it has been found that the necessary condition for stable transcriptional activity of p53 is the cooperation of multiple posttranslational modifications such as phosphorylation, acetylation, and ubiquitination [8]. In studying the complex pattern of posttranslational modifications and its impact on cellular processes, it has been found that lysine acetylation, arginine/lysine methylation and serine/threonine phosphorylation will work together cooperatively for regulating the high mobility group proteins [9]. In the experiments with human cancer specimens, it has been found that the extent of acetylation, formylation and methylation is higher in cultured cells [10]. It has also been found that proteins with multiple posttranslational modifications may make contribution to similar signalling functions [11]. In studying DNA repair, apoptosis and senescence, it has been found that the interplay between multiple protein modifications, including phosphorylation, ubiquitylation, acetylation and sumoylation is critical for properly propagating DNA damage signals [12] and the interplay between methylation and acetylation has been found important for activating p53 by responding to DNA damage signals

[13]. It is even found that there is crosstalk between different posttranslational modifications [14]. In glycogen synthases kinase-3, it has been found that O-GlcNAcylation O-phosphate is interplaying for cellular regulation [15]. The interplay has been also found in the steroid receptor coactivators [16]. In the study of transcriptional programming, it is found that interplay between posttranslational modifications exists in H3 termini [17]. In histone, it has been found that there are multiple arginine posttranslational modifications which are critical for some disease development. Also in histone, the interplay between sumoylation and either acetylation or ubiquitylation has been observed contributing to complex functions of proteins [18]. A recent study has used laboratory method to identify co-occurrence posttranslational modifications [19]. A computational method was proposed to predict the interplay between phosphorylation sites and O-GlcNAc sites based on peptides around modification residues [20,21]. The analysis was based on the prediction results from various PTM prediction tools and was based on peptide information only. Moreover, the method only focused on the competition mechanism between phosphorylation sites and O-GlcNAc modifications at the same residues.

The patterns of co-occurrence of posttranslational modifications are so far unclear or have not yet emerged through large scale studies. Bioinformatics study will help revealing those patterns and will benefit many desirable cellular engineering processes, i.e. disease control and prevention based on handling signalling pathways subjectively. This study is aimed to analyse the patterns of co-occurrence of multiple posttranslational modifications and visualizing their relationship through a probabilistic analysis.

Computational approaches, such as structural bioinformatics [22], molecular docking [23], molecular packing [24,25], pharmacophore modelling [26], Mote Carlo simulated annealing approach [27], diffusion-controlled reaction simulation [28], bio-macromolecular internal collective motion simulation [29], QSAR [30,31], protein subcellular location prediction [32,33], identification of membrane proteins and their types [34], identification of enzymes and their functional classes [35], identification of GPCR and their types [36], identification of proteases and their types [37], protein cleavage site prediction [38,39,40], and signal peptide prediction [41,42], can timely provide very useful information and insights for both basic research and drug design and are widely welcome by science community. The present study is devoted to develop a computational approach for studying co-occurrence of posttranslational modifications.

2. APPROACH

The first thing we need to do is to scan all the sequences in a data set to find all neighbouring modifications. We use frequency as the joint probability to measure the quantitative

property of co-occurrence of modifications first. Through analyzing the frequency of two modifications, the likelihood that a pair of modifications occurs can be quantified.

However, the frequency analysis only shows how likely two modifications can occur simultaneously in the same sequence. For instance, we may observe that the probability for hydroxyproline and hydroxylysine to occur simultaneously in the same sequence is 18.6%. But this does not indicate how likely hydroxylysine depends on hydroxyproline. In other words, if we have observed a hydroxyproline in a sequence, how likely can we find a hydroxylysine in the same sequence as the neighbouring modification? We first define the joint probability of two different modification residues as the frequency defined as

$$P(X, Y) = \frac{\text{the number that } X \text{ and } Y \text{ occur simultaneously}}{\text{the number of neighbouring modifications}} \quad (1)$$

below Here, X and Y are two different modification residues. We then define the marginal probability as the frequency for one modification residue to occur in a data set

$$P(X) = \frac{\text{the number of } X \text{ modification}}{\text{the number of all modifications}} \quad (2)$$

Using the product theory in probability, we have Here

$$P(X, Y) = P(X | Y)P(Y) = P(Y | X)P(X) \quad (3)$$

$P(X | Y)$ reads out as the conditional probability for X to occur given that Y has happened. Based on the above calculation, we will have two conditional probabilities, either the probability of observing X if Y has been observed or the probability of observing Y if X has been observed. Based on the conditional probabilities and the marginal probabilities, we can use the Bayes rule to determine the posterior probabilities which are commonly used for decision-making. The Bayes rule is defined as

$$P(X_i | Y) = \frac{P(Y | X_i)P(X_i)}{\sum_m P(Y | X_m)P(X_m)} \quad (4)$$

Here we treat Y as a target modification, for instance a hydroxyproline residue. X_i is the i th potential contributing modification for the target modification Y . The posterior probability indicates if Y has been observed, what is the probability that it has a neighbour X_i . It quantifies quantitatively how possible X_i is most probable neighbouring modification of Y . Note that $\sum_i P(X_i | Y) = 1$. The posterior probabilities are then used to draw networks to illustrate the relationship between a target modification residue and other modification residues. This probabilistic network here is a simple

type of Bayesian networks [43].

3. DATA AND EXPERIMENTAL DESIGN

Two classes of posttranslational modifications are used for the study, *i.e.* hydroxylation and methylation. Both have two most common modification residues. The hydroxylation mainly functions at a lysine residue or a proline residue while methylation mainly functions at a lysine residue or an arginine residue. Both have ample experimentally verified data for the study. Four keywords, hydroxyproline, hydroxylysine, methyllysine and methylarginine were used to scan NCBI database to download sequences. All the identical sequences were removed from the study. All three types of phosphorylations were grouped together named as phosphorylation. All the amidation activities were also grouped into one type of modification. Various acetylation modifications are grouped together. Because there are only two poly (methylaminopropyl) lysine sites, they are treated as methyllysine. Two methyl-hydroxylysine residues are treated separated as one methyllysine and one hydroxylysine.

Table 1 shows the statistics of these four data sets. There are 10, 17, 10, and 8 different modification residues in the hydroxylysine, hydroxyproline, methylarginine, and methyllysine data sets, respectively.

Here modification residue means a specific posttranslational modification activity at residues in proteins, for instance a hydroxyproline residue means a proline which can be hydroxylated and has been confirmed in experiments. The percentages of sites per sequence are 15.3, 7.1, 7.3, and 5.8 for the hydroxylysine, hydroxyproline, methylarginine, and methyllysine data sets, respectively. The hydroxyproline data set has the double number of modification residues compared with other three data sets. The details of multiple modifications are listed in **Table 2**. The abbreviations are seen in **Table 3**.

Table 4 shows the distribution of neighbouring PTMs of different modification residues. It can be seen that at least 25% (and up to 35%) of neighbouring PTMs are of different modification residues.

Table 1. The statistics of four data sets

Data sets	Sequences	Sites	PTM types	Sites per seq
Hydroxylysine	38	581	10	15.3
Hydroxyproline	199	1405	17	7.1
Methylarginine	23	167	10	7.3
Methyllysine	65	376	8	5.8

Table 2. The details of modifications in four data sets

Hydroxylysine	Hydroxyproline	Methylarginine	Methyllysine
hydroxyproline	hydroxyproline	phosphorylation	phosphorylation
hydroxylysine	hydroxylysine	methylarginine	acetylation
Amidation	hydroxyphenylalanine	methylhistidine	citrulline
Allylsine	allylsine	methylglutamine	methyllysine
phosphorylation	amidation	thioglycine	methylarginine
methyllysine	acetylation	acetylation	amidation
bromotryptophan	carboxylation	methyllysine	methylhistidine
hydroxyphenylalanine	hydroxyasparagine	citrulline	methylalanine
hydroxyarginine	bromotryptophan	methylcysteine	
proteolytic	methylcysteine	deamidation	
	phosphorylation		
	methyllysine		
	sulfoation		
	didehydrobutyrine		
	chlorotryptophan		
	didehydroalanine		
	hydroxyvaline		

Table 3. The abbreviations

AC	acetylation	AK	allylsine
AM	amidation	BR	bromotryptophan
CA	carboxylation	CH	chlorotryptophan
CT	citrullination	DA	didehydroalanine
DM	deamidation	DP	D-phenylalanine
DT	D-tryptophan	DY	didehydrobutyrine
HK	hydroxylysine	HP	hydroxyproline
HR	hydroxyarginine	HN	hydroxyasparagine
HV	hydroxyvaline	HF	hydroxyphenylalanine
MA	methylalanine	MC	methylcysteine
MQ	methylglutamine	MH	methylhistidine
MK	methyllysine	MR	methylarginine
NT	nitration	PH	phosphorylation
PR	proteolytic	SU	sulfoation
TH	thioglycine		

Among these neighbouring PTMs of different modification residues, 33%, 68%, 78%, 84% have the distance less than 10 residues for the methylarginine, the methyllysine, the hydroxylysine, and the hydroxyproline data sets, respectively as seen in **Table 5**. Because of this, two PTMs of different modification residues may likely share similar structure (at least an overlapped local structure) for binding. This suggests that the cooperative activities of PTMs of different modification residues are critical to cellular signalling/functioning.

Based on the collected sequences, we produce a program to search for all the posttranslational modification residues in four data sets. The sites must have the notation as `</site_type="modified">`, `</experiment="experimental evidence, ...">`, and `</note_type=X>`. Here X can be various types, for instance, 4-hydroxyproline, 3-hydroxyproline, 5-hydroxyproline, etc. For each of four posttranslational modifications, we find all the involved posttranslational modification sites. The sequences in different data sets are analyzed separately although there are some overlaps among them.

4. RESULTS

Table 6 shows the frequencies as the joint probabilities of nine types of modifications for the hydroxylysine data set. Proteolytic is removed as there is only one such site in the data set. The highest joint probability is 60.2% for two hydroxyproline residues to be neighbours. However, the joint probability for two hydroxylysine residues to be neighbours is only 6.81%. The co-occurrence probability for these two types of hydroxylation is 18.8%. These two probabilities indicate that for every hydroxylysine residue, the probability for it to have a hydroxyproline as the neighbour is three times higher than the probability for it to have the same hydroxylysine residue as the neighbour. The joint probabilities for a hydroxylysine to have an amidation, allysine, phosphorylation residue, bromotryptophan, hydroxyphenylalanine, and hydroxyarginine residue as the neighbour are 0.55%, 0.55%, 0.18%, 0.74%, and 0.18%, respectively. This means that except for the same type of modification,

Table 4. The distribution of neighbouring PTMs of different modification residues

	sites	Percentage over total
Hydroxylysine	147	25%
Hydroxyproline	434	31%
Methylarginine	40	24%
Methyllysine	132	35%

Table 5. The times for different modifications to occur in one sequence

	1-2	3-4	5-6	7-8	9-10	>11	Sum
Hydroxylysine	32	32	33	3	14	33	147
Hydroxyproline	175	97	46	18	29	69	434
Methylarginine	3	2	2	2	4	27	40
Methyllysine	48	24	5	11	2	42	132

Table 6. The joint probability as frequency of co-occurred modifications for the hydroxylysine data set

	HP	HK	AM	AK	PH	MK	BR	HF	HR
HP	60.2	18.8	0	0	0	0	0.18	0	0
HK	18.8	6.81	0.55	0.55	0.18	0	0.18	0.74	0.18
AM	0	0.55	0	0	0	0	0	0	0
AK	0	0.55	0	0	0	0	0	0	0
PH	0	0.18	0	0	4.05	5.52	0	0	0
MK	0	0	0	0	5.52	1.47	0	0	0
BR	0.18	0.18	0	0	0	0	0	0	0.18
HF	0	0.74	0	0	0	0	0	0.37	0
HR	0	0.18	0	0	0	0	0.18	0	0

hydroxylysine has a high correlation with allysine, amidation, and hydroxyphenylalanine modifications.

Figure 1 visualises the probabilistic relationship among different types of modifications in the hydroxylysine data set using the posterior probabilities, where all the posterior probabilities less than 10% are omitted for simplicity. The network demonstrates that hydroxylysine only depends on hydroxyproline (24%). However, hydroxylysine has great impacts on five modification residues, allysine (100%), amidation (100%), hydroxyphenylalanine (67%), hydroxyarginine (50%), and bromotryptophan (33%). Phosphorylation and methyllysine modification residues are independent from the hydroxylysine block. They are mutually correlated to each other.

Table 7 shows the joint probabilities for the hydroxyproline data set. It can be seen that the probability for two hydroxyproline residues to be neighbours is 57.7%. The likelihood for a hydroxyproline to have a hydroxylysine

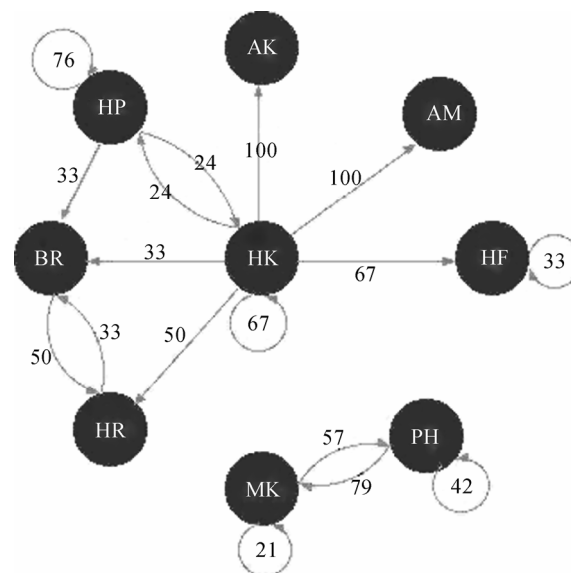


Figure 1. The probabilistic network as the relationship among modification residues in the hydroxylysine data set. The values represent the posterior probabilities. The arcs mean the directions. For instance, the arc from HK to AM with a number 100 means $P(\text{HK}|\text{AM})=100\%$. In other words, the probability of observing a hydroxylysine residue nearby an observed amidation residue is almost certain

as the neighbour is 8.46%. However, the likelihood for a hydroxyproline to have a hydroxyphenylalanine as the neighbour is 18.2%. This means that a hydroxyproline is more likely to have a hydroxyphenylalanine to co-occur rather than hydroxylysine. The other two important modifications for hydroxyproline are carboxylation and amidation. The probability for a hydroxyproline residue and an amidation residue to be neighbours is 4.06% and the probability for a hydroxyproline residue and a carboxylation residue to be neighbours is 1.91%. The other co-occurred modifications with joint probabilities larger than 0.2% are acetylation and bromotryptophan.

The probabilistic relationship among different modifications shown in **Figure 2** is built for the hydroxyproline data set using the posterior probabilities. All the posterior probabilities less than 10% are not shown. In the probabilistic network, it can be seen that the most contributing modifications for hydroxyproline is hydroxyphenylalanine (20%). However, hydroxyproline has contributed to 11 other modification residues. For instance, the posterior probability $P(HP|AC)$ is 100% meaning that whenever we have found an acetylation residue, it is certain that there is a hydroxyproline residue nearby. The posterior probability $P(HP|HP)$ is 63% while $P(HK|HP)=9\%$, $P(AK|HP)=1\%$, $P(AM|HP)=4\%$, $P(AC|HP)=1\%$, $P(CA|HP)=2\%$, $P(BR|HP)=1\%$.

Table 7. The joint probability (larger than 0.2%) as frequency of co-occurred modifications for the hydroxyproline data set

	HP	HK	HF	AM	AC	CA	BR
HP	57.7	8.46	18.2	4.06	0.41	1.91	0.5
HK	8.46	1.58	0	0.08	0	0	0.08
HF	18.2	0	0.25	0	0	0	0
AM	4.06	0.08	0	0.17	0	0.17	0.41
AC	0.41	0	0	0	0	0	0
CA	1.91	0	0	0.17	0	2.24	0.33
BR	0.5	0.08	0	0.41	0	0.33	0

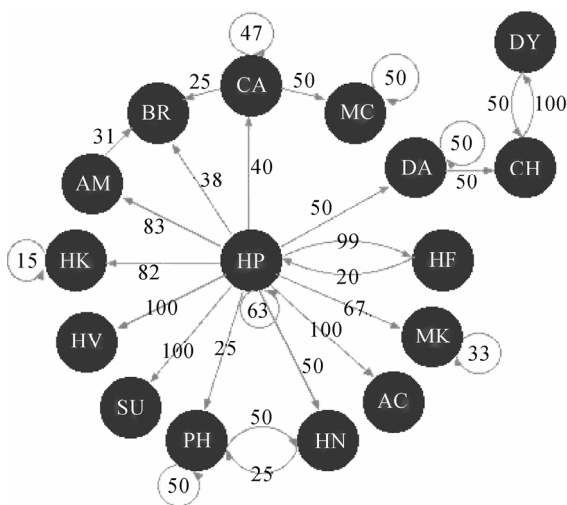


Figure 2. The probabilistic network as the relationship among modifications in the hydroxyproline data set. The values represent the posterior probabilities

Table 8 shows the frequencies of modifications in the methylarginine data set. The likelihood for two methylarginine residues to co-occur as neighbours is 40.3%. The interesting phenomenon is that the co-occurrence probability between methylarginine and methyllysine residues (as neighbours) is 0%. The contributing modifications to methylarginine residues are phosphorylation (11.1%), acetylation (2.78%), methylhistidine (1.39%), and citrulline (1.39%). This is not as expected as it is thought that both two dominant methylation modifications should be highly correlated.

Shown in **Figure 3** is the probabilistic relationship among the modifications derived from the methylarginine data set. The network shows that the most contributing modifications to methylarginine modification is phosphorylation (19%), i.e. $P(PH|MR)=19\%$. For any observed methylarginine residue in a protein, the probability of observing a phosphorylation residue is 19%. Meanwhile, methylarginine residue can be an important pre-request for other modification residues. For instance, the probability of observing a methylarginine residue if a methylhistidine residue has been observed is 100% and the probability that a methylarginine residue is standing near an observed acetylation residue is 33%.100%.

Table 8. The joint probability as frequency of co-occurred modifications for the methylarginine data set

	PH	MR	MH	MG	TH	AC	MK	CT
PH	26.4	11.1	0	0	0	2.78	1.39	0.69
MR	11.1	40.3	1.39	0.69	0.69	2.78	0	1.39
MH	0	1.39	0	0	0	0	0	0
MG	0	0.69	0	0	0.69	0	0	0
TH	0	0.69	0	0.69	0	0	0	0
AC	2.78	2.78	0	0	0	2.08	0.69	0
MK	1.39	0	0	0	0	0.69	0.69	0
CT	0.69	1.39	0	0	0	0	0	2.78

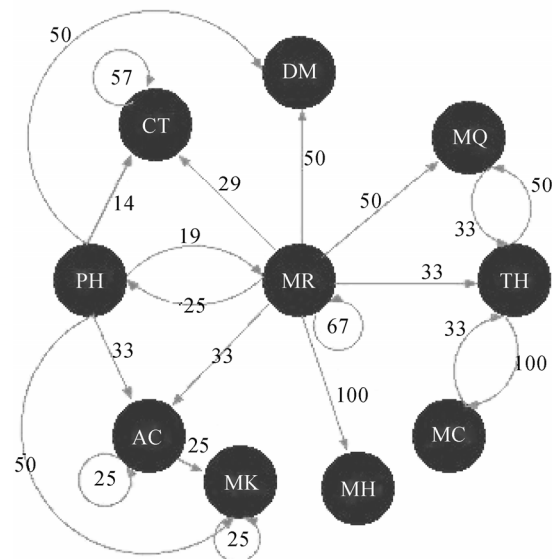


Figure 3. The probabilistic network as the relationship among modifications in the methylarginine data set. The values represent the posterior probabilities

Table 9 shows the frequencies for the methyllysine data set. The probability for two methyllysine residues to be neighbours is 28.3%, which is not dominantly high. Interestingly, we have found the contributing modifications for methyllysine are acetylation (15.11%) and phosphorylation (10.61%). It is again difficult to find the evidence that two methylation modifications are highly correlated.

Figure 4 shows the probabilistic network as the relationship among the modifications in the methyllysine data set. Here, the most contributing modifications to methyllysine are phosphorylation (22%) and acetylation (31%). The methyllysine residue has a high correlation with methylarginine, methyhistine, and methylalanine. All have the posterior probabilities as 100%.

5. CONCLUSION

This paper has studied the co-occurrence pattern of two types of posttranslational modifications with four modification residues. The study aims to reveal how posttranslational modifications are correlated to each other, i.e. how one posttranslational modification contributes to the others. It has been found that the hydroxylysine and hydroxyproline residues are not the most mutually dependent modification residues, so are the methylarginine

Table 9. The joint probability as frequency of co-occurred modifications for the methyllysine data set

	PH	AC	CT	MK	MR	AM	MH	MA
PH	8.04	11.58	0	10.61	0	0.32	0	0
AC	11.58	28.3	0.96	15.11	1.93	0	0	0
CT	0	0.96	0	0.96	0	0	0	0
MK	10.61	15.11	0.96	21.22	0	0	0.32	0.64
MR	0	1.93	0	0	0	0	0	0
AM	0.32	0	0	0	0	0	0	0
MH	0	0	0	0.32	0	0	0	0
MA	0	0	0	0.64	0	0	0	0

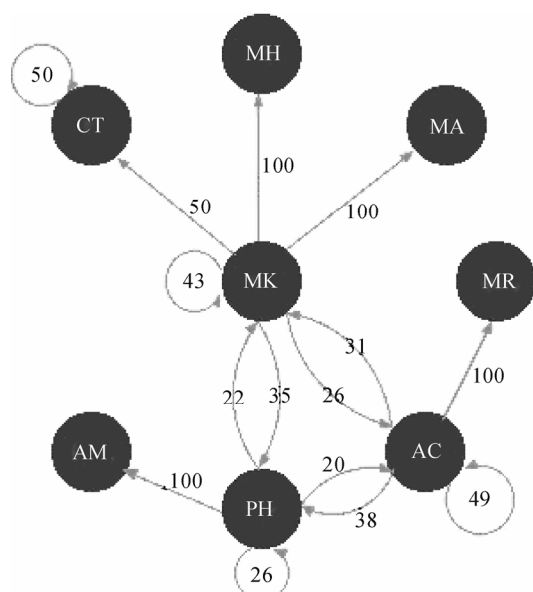


Figure 4. The probabilistic network as the relationship among modifications in the methyllysine data set. The values represent the posterior probabilities

and methyllysine residues. We have found that the hydroxylysine residues depend on the hydroxyproline residues with a posterior probability 24% and the hydroxyproline residues are the unique major contributing modification residue for the hydroxylysine residues. However, we have found the hydroxyproline residues nearly do not depend on the hydroxylysine residues. Instead, the hydroxyphenylalanine residues are the contributing modification residue to the hydroxyproline residues with a posterior probability 20%. Among the methylarginine residues and the methyllysine residues, we have found that the phosphorylation residues are the main player for both of these two modification residues. In addition, the acetylation residues are needed for the methyllysine residues as well. Surprisingly, two different methylation residues also do not rely on each other. Although the study is limited to two modification classes with four modification residues, it is expected that this method can be generalized to a wide range of multiple posttranslational modification pattern discovery.

REFERENCES

- [1] E. Lai, T. Teodoro, and A. Volchuk, (2007) Endoplasmic reticulum stress: signaling the unfolded protein response. *Physiology (Bethesda)*, 22, 193–201.
- [2] J. F. Haince, S. Kozlov, V. L. Dawson, T. M. Dawson, M. J. Hendzel, M.F. Lavin, and G. G. Poirier, (2007) Ataxia telangiectasia mutated (ATM) signaling network is modulated by a novel poly (ADP-ribose)-dependent pathway in the early response to DNA-damaging agents. *J Biol Chem.*, 282, 16441–16453.
- [3] S. Lutsenko, A. Gupta, J. L. Burkhead, and V. Zuzel, (2008) Cellular multitasking: The dual role of human Cu-ATPases in cofactor delivery and intracellular copper balance. *Arch Biochem Biophys.*, (in press).
- [4] J. M. Souza, G. Peluffo, and R. Radi, (2008) Protein tyrosine nitration-Functional alteration or just a biomarker? *Free Radic Biol Med.*, (in press).
- [5] J. Z. Wang, and F. Liu, (2008) Microtubule-associated protein tau in development, degeneration and protection of neurons. *Prog Neurobiol.*, 85, 148–175.
- [6] V. V. Sumbayev and I. M. Yasinska, (2008) Protein S-nitrosation in signal transduction: assays for specific qualitative and quantitative analysis. *Methods Enzymol.*, 440, 209–219.
- [7] T. Obsil, and V. Obsilova, (2008) Structure/function relationships underlying regulation of FOXO transcription factors. *Oncogene*, 27, 2263–2275.
- [8] A. Scoumanne and X. Chen, (2008) Protein methylation: a new mechanism of p53 tumor suppressor regulation. *Histol Histopathol.*, 23, 1143–1149.
- [9] Q. Zhang, and Y. Wang, (2008) High mobility group proteins and their post-translational modifications. *Biochim Biophys Acta.*, (in press).
- [10] J. R. Wisniewski, A. Zougman, S. Kruger, P. Ziołkowski, M. Pudielko, M. Bębenek and M. Mann, (2008) Constitutive and dynamic phosphorylation and acetylation sites on NUCKS, a hypermodified nuclear protein, studied by quantitative proteomics. *Proteins*, (in press).
- [11] C. A. Galea, A. Nourse, Y. Wang, S. G. Sivakolundu, W. T. Heller and R. W. Kriwacki, (2008) Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *J Mol Biol.* 376, 827–838.
- [12] M. S. Huen and J. Chen, (2008) The DNA damage response pathways: at the crossroad of protein modifications. *Cell Res.*, 18, 8–16.
- [13] M. S. Huen and J. Chen, (2008) The DNA damage response pathways: at the crossroad of protein modifications. *Cell Res.*,

- 18, 8–16.
- [14] T. Hunter, (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell.*, 28, 730–738.
- [15] Z. Wang, A. Pandey and G. W. Hart, (2007) Dynamic interplay between O-linked N-acetylglucosamylation and glycogen synthase kinase-3-dependent phosphorylation. *Mol Cell Proteomics*, 6, 1365–1379.
- [16] S. Li and Y. Shang (2007) Regulation of SRC family coactivators by post-translational modifications. *Cell Signal*, 19, 1101–1112.
- [17] P. R. Thompson and W. Fast, (2006) Histone citrullination by protein arginine deiminase: is arginine methylation a green light or a roadblock? *ACS Chem Biol.*, 21, 433–41.
- [18] D. Nathan, K. Ingvarsdottir, D. E. Sterner, G. R. Bylebyl, M. Dokmanovic, J. A. Dorsey, K. A. Whelan, M. Krsmanovic, W. S. Lane, P. B. Meluh, E. S. Johnson, and S. L. Berger, (2006) Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes Dev.*, 20, 966–976.
- [19] J. Seo, J. Jeong, Y. M. Kim, N. Hwang, E. Paek and K. J. Lee, (2008) Strategy for comprehensive identification of post-translational modifications in cellular proteins, including low abundant modifications: application to glyceraldehyde-3-phosphate dehydrogenase. *J Proteome Res.*, 7, 587–602.
- [20] A. Ahmad, T. S. Khan, D. C. Hoessli, E. Walker-Nasir, A. Kaleem, A. R. Shakoori and S. Nasir-ud-Din, (2008) In silico modulation of HMGN-1 binding to histones and gene expression by interplay of phosphorylation and O-GlcNAc modification. *Protein Pept Lett.*, 15, 193–199.
- [21] A. Kaleem, D. C. Hoessli, I. Ahmad, E. Walker-Nasir, A. Nasim, A. R. Shakoori and S. Nasir-ud-Din, (2008) Immediate-early gene regulation by interplay between different post-translational modifications on human histone H3. *J Cell Biochem.*, 103, 835–851.
- [22] K. C. Chou, (2004) Review: Structural bioinformatics and its impact to biomedical science. *Current Medicinal Chemistry*, 11, 2105–2134.
- [23] K. C. Chou, D. Q. Wei and W. Z. Zhong (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. (Erratum: *ibid.*, 2003, Vol.310, 675). *Biochem Biophys Res Comm*, 308, 148–151.
- [24] K. C. Chou, G. Nemethy and H. A. Scheraga, (1984) Energetic approach to packing of α -helices: 2. General treatment of non-equivalent and nonregular helices. *Journal of American Chemical Society*, 106, 3161–3170.
- [25] K. C. Chou, G. M. Maggiora, G. Nemethy and H. A. Scheraga, (1988) Energetics of the structure of the four- α -helix bundle in proteins. *Proceedings of National Academy of Sciences, USA*, 85, 4295–4299.
- [26] S. Sirois, D. Q. Wei, Q. S. Du and K. C. Chou, (2004) Virtual-Screening for SARS-CoV Protease Based on KZ7088 Pharmacophore Points. *J Chem Inf Comput Sci*, 44, 1111–1122.
- [27] K. C. Chou, (1992) Energy-optimized structure of antifreeze protein and its binding mechanism. *Journal of Molecular Biology*, 223, 509–517.
- [28] K. C. Chou and G. P. Zhou, (1982) Role of the protein outside active site on the diffusion-controlled reaction of enzyme. *Journal of American Chemical Society*, 104, 1409–1413.
- [29] K. C. Chou, (1988) Review: Low-frequency collective motion in biomacromolecules and its biological functions. *Biophysical Chemistry*, 30, 3–48.
- [30] Q. S. Du, R. B. Huang, Y. T. Wei, L. Q. Du and K. C. Chou, (2008) Multiple Field Three Dimensional Quantitative Structure-Activity Relationship (MF-3D-QSAR). *Journal of Computational Chemistry*, 29, 211–219.
- [31] H. Gonzalez-Diaz, Y. Gonzalez-Diaz, L. Santana, F. M. Ubeira and E. Uriarte, (2008) Proteomics, networks, and connectivity indices. *Proteomics*, 8, 750–778.
- [32] K. C. Chou and H. B. Shen, (2007) Review: Recent progresses in protein subcellular location prediction. *Analytical Biochemistry*, 370, 1–16.
- [33] K. C. Chou and H. B. Shen, (2008) Cell-PLoc: A package of web-servers for predicting subcellular localization of proteins in various organisms. *Nature Protocols*, 3, 153–162.
- [34] K. C. Chou and H. B. Shen, (2007) MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. *Biochem Biophys Res Comm*, 360, 339–345.
- [35] H. B. Shen, and K. C. Chou, (2007) EzyPred: A top-down approach for predicting enzyme functional classes and subclasses. *Biochem Biophys Res Comm*, 364, 53–59.
- [36] K. C. Chou, (2005) Prediction of G-protein-coupled receptor classes. *Journal of Proteome Research*, 4, 1413–1418.
- [37] K. C. Chou and H. B. Shen, (2008) ProfIdent: A web server for identifying proteases and their types by fusing functional domain and sequential evolution information. *Biochem Biophys Res Comm*, 376, 321–325.
- [38] K. C. Chou, (1993) A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. *J Biol Chem*, 268, 16938–16948.
- [39] K. C. Chou, (1996) Review: Prediction of HIV protease cleavage sites in proteins. *Analytical Biochemistry*, 233, 1–14.
- [40] H. B. Shen and K. C. Chou, (2008) HIVcleave: a web-server for predicting HIV protease cleavage sites in proteins. *Analytical Biochemistry*, 375, 388–390.
- [41] K. C. Chou and H. B. Shen, (2007) Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. *Biochem Biophys Res Comm*, 357, 633–640.
- [42] H. B. Shen and K. C. Chou, (2007) Signal-3L: a 3-layer approach for predicting signal peptide. *Biochem Biophys Res Comm*, 363, 297–303.
- [43] P. Pourret and B. Naim, (2008) *Marcot, Bayesian Networks: A Practical Guide to Applications*. Chichester, UK: Wiley.