Pakistan J. Zool., vol. 38(1), pp. 55-63, 2006.

Consensus Sequences as Targets for Phosphorylation of Amino Acids in Phosphoproteins: Statistical Computing Analysis

WAJAHAT M. QAZI, MUNIR AHMED, DANIEL C. HOESSLI, ISHTIAQ AHMAD, IFFAT KHAWAJA, TAYYABA WAJAHAT, AFSHAN KALEEM, EVELYNE WALKER-NASIR, NAZIM RAHMAN, ABDUL R. SHAKOORI AND NASIR-UD-DIN*

Institute of Molecular Sciences and Bioinformatics, Lahore, Pakistan (WMQ, IA, IK, TW, AK, EWN, NR & NUD), National College of Business Administration & Economics (MA), Department of Pathology and Immunology, Centre Médical Universitaire, Geneva, Switzerland (DCH), School of Biological Sciences, University of the Punjab, Lahore, Pakistan (ARS), and HEJ Research Institute of Chemistry, University of Karachi, Karachi, Pakistan (NUD).

Abstract.- Phosphorylation is a post-translational modification (PTM) of a protein on hydroxyl function of Ser/Thr/Tyr (O-linked phosphorylation) or on amino function of His (N-linked phosphorylation) by a phosphate group utilising a specific kinase. Post translational modification has effect on protein conformation and as a consequence results in alteration of protein's function. Present study analyzes the putative target sequence of O-phosphorylation of Ser, Thr and Tyr. The objective is to evaluate the preferential pattern of amino acids on both (right and left) sides of O-phosphorylated amino acid. Calculation of frequency of occurrence of each amino acid around O-linked phosphorylated amino acid residue followed by deviation parameter will serve as the source to define preferential behaviour of each amino acid to develop possible consensus prerequisite pattern for O-linked phosphorylation. These patterns will provide with means to predict phosphorylation potential of proteins and its multifunctional nature based on dynamic and reversible phosphorylation at specific sites.

Key words: Post translational modifications, O-linked phosphorylation, protein conformation.

INTRODUCTION

The comprehensive studies on genome sequence of different organisms define functional implications of the genome. The functional genomics elaborates proteomics, expressed proteins. Proteins are functional molecules in complex organisms. One protein may be involved in different and unrelated functions, and able to mediate coordinated adaptive responses across cellular compartments. Changes in local environments can induce functional switches by altering protein conformation. Protein function depends on the protein's primary structure (amino acid composition) and its conformation in a given environment. The protein functions *in vivo* are regulated by its conformation, and in turn the protein conformation is controlled by its primary structure and post-translational modifications (PTMs) on amino acids. PTMs are amongst multiple biological interactions that make our understanding of the *in vivo* system more complicated as we follow the flow of genetic information.

Protein phosphorylation is the primary means of switching the activity of a cellular protein rapidly from one state to another and is considered as being a key event in many signal transduction pathways of biological systems. Protein phosphorylation can be either N-Linked phosphorylation as on His or O-linked phosphorylation as on Ser, Thr and Tyr. Phosphorylation of substrate sites at Ser, Thr or Tyr residues is performed by members of the protein kinase family, the second largest family in the human genome. Reversible protein phosphorylation is a fundamental regulatory cellular mechanism. Biochemically, this includes a transfer of a phosphate moiety from ATP to the acceptor residue, thereby generating ADP. It is a post translational event which normally occurs in either the cytosol or the nucleus of the cell. Protein kinases catalyze the

*To whom correspondence be addressed. E-mail: prof_nasir@yahoo.com, nasir@super.net.pk
0030-9923/2006/0001-0055 $ 8.00/0

phosphorylation events that are essential for the regulation of cellular processes like metabolism, proliferation, differentiation and apoptosis (Kolibaba *et al.,* 1997; Hunter *et al.,* 1998). This very large family of enzymes share homologous catalytic domains and the mechanism of substrate recognition may be similar despite large variation in sequence. Crystallization studies indicate that a region between seven and twelve residues in size, surrounding the acceptor residue contacts the kinase active site (Songyang *et al.,* 1994). The specificity of protein kinases is dominated by acidic, basic or hydrophobic residues adjacent to the phosphorylated residue, but the large variation makes it difficult to manually inspect protein sequences and predict the position of biologically active sites. A wide range of algorithms have been used to implement prediction strategies. These range from simple motif searches (regular expressions), to more complex methods like neural networks where sequence correlations can be taken into account, when discriminating between potential phosphorylation sites and those which appear never to be modified.

This study analyzes preferential behaviour of amino acid residues around the phosphorylated Ser, Thr and Tyr by calculating the frequency of each amino acid around phosphorylated Ser, Thr and Tyr and will also be helpful in the development of some possible consensus sequence pattern(s). Development of consensus sequence pattern(s) in a protein will be helpful in prediction of protein phosphorylation sites and in explaining its multifunctional nature *in vivo.* The protein phosphorylation is dynamic and is of reversible nature.

MATERIALS AND METHODS

Data collection and pre analysis processing

Phospho-protein database was developed from phosphobase 3.0 (Diella *et al.,* 2004). Phosphobase3.0 is text file formatted database which has 4252 phosphorylation entries for 1663 proteins. In order to perform data processing and to ensure normalization of data in phosphobase3.0, it was cloned as a table in phospho-protein database with table name set to "OrignalPhosphobase3_0". The entity "proteins" contains the information about all phospho-proteins like protein name, sequence and accession number. Information about phosphorylated sites and non phosphorylated sites are managed by "Phosphorylation" and "NonPhosphorylation" entities, respectively.

Phosphobase3.0 contains both N-linked and O-linked phosphorylated proteins. Only one entry for N-linked phosphorylation was found and 4251 entries were found for O-linked phosphorylation. As this study is focused on O-phosphorylation sites so the entry for N-linked phosphorylation was removed from phospho-protein database resulting 1662 O-linked phospho-proteins. The normalization process was carried out with "PhosphoExplorer". PhosphoExplorer was the software written in order to assist the current study. PhosphoExplorer was written using Microsoft Visual C# 2005. Many bioinformatics related softwares have been written using JAVA, PERL and C++ but C# (Visual C#) as a new object oriented language provides greater expressiveness and is more suited to writing performance-critical code than Java, while sharing Java's elegance and simplicity, which makes both much more appealing than C++ (http://genamics.com/developer/csharp_comparative.htm). Bioinformatics as being the scientific and mathematical application requires to be written in a scientific programming language which should be able to produce efficient high-performance code and can be carefully fine-tuned. Gilani (2004) conducted study on C# as scientific computing language and reported that C# is an ideal platform for deploying scientific code and performs well enough to be recognized fully as a competent scientific language.

Out of 1662, 17 proteins were not included in study due to error and misinformation.. Resultantly 1645 phospho-proteins are included in current study.

After normalization process, phospho-protein database was used to generate subsequences/ peptides for phosphorylated sites, which were stored in a table named "Subsequence_Phosphorylated Sites" (for phosphorylated residue at position 0). The length of the sub sequence was 21 such that the subject amino residue (Ser/Thr/Tyr) was at the center *i.e.* position 11 in the subsequence. As the analysis is to be made on the subject amino residue, so in the study, position of the subject amino acid

was considered position 0 and surrounding position were marked as -1,-2, …..-9, -10, and +1, +2, +3 ….. +9, +10. All phosphorylated Ser, Thr and Tyr residues were renamed as PS, PT and PY respectively in order to identify which Ser/Thr/Tyr residue is phosphorylated and which is not, in other words phosphorylated and non phosphorylated residues are treated as separate entities. Similarly for non phosphorylated Ser, Thr and Tyr residues were renamed as NPS, NPT and NPY respectively. This marking will help us to get more generalized information about the surrounding of the amino acids at position 0 of the subsequences and will help us in understanding the role and effect of phosphorylated and non phosphorylated Ser, Thr and Tyr residues in the environment of the position 0 residues (Ser/Thr/Tyr) along with other amino acids.

Data obtained from the above mentioned procedure was divided into three groups which were used for the analysis of Ser, Thr and Tyr phosphorylated residues, respectively.

Statistical analysis

Phosphoprotein database analysis was conducted in order to ascertain the most occurring amino acid residue around the phosphorylated sites (Ser, Thr and Tyr). To accomplish this goal frequency of occurrence for each amino acid of type X on all positions for each group was computed as:

$$F_{observed}(X)_p = \sum \text{Count}_{observed}(X)_p / m$$

where X is the amino acid, p is the position (for this study value of p ranges -10 to +10) and m is the number of peptides in the data set. Significant occurrence of an amino acid residue was identified by marking the corresponding locations in frequency table and deviation parameter table. Deviation Parameter was calculated as:

$$DP(X)_p = ((F_{observed}(X)_p - F_{expected}(X)) / F_{expected}(X)) * 100$$

Amino acids with negative deviation parameter value were considered to be negatively preferred, where as amino acids with positive deviation parameter value was considered to be positively preferred around phosphorylated sites. Deviation Parameters for amino acids were marked to be significant if their corresponding DOE > 2δ, where DOE is the difference between observed and expected counts, *i.e.*

$$DOEC = \text{Count}_{observed}(X)_p \sim \text{Count}_{expected}(X)$$

and δ is given as $\delta = \sqrt{\text{Count}_{observed}(X)_p}$.

RESULTS

Frequency of occurrence

As mentioned in previous section that statistical analysis was performed on three datasets which represent the distribution of amino acid around the amino acid residues (Ser/Thr/Tyr) at position 0, the frequency of occurrence around phosphorylated Ser, Thr and Tyr residues are described below:

Amino acids frequency around phosphorylated serine

Proline has highest frequency at +1 position (33.54%), which is the largest value in whole phosphorylated frequency table. High frequency can also be observed for proline from position -10 to -5 which varies between 7.8% and 8.2% and from position -2 to +3, position number +5 and from position no +7 to +10. Arginine has the second and third highest frequency values at positions -3 and -2 with values 20.97% and 12.049% in the phosphorylated Ser frequency table. Arginine has significant frequencies at positions from -9 to -1, +3, +6 to +9. Phosphorylated Ser has significant but low frequency from position -10 to 10 between 1.9% and 5.6% but on the other hand non - phosphorylated Ser has the highest frequency after proline and arginine at positions -9 with frequency of 9.95%, at -4 with frequency of 10.13%, at -2 with frequency of 10.24% and at +10 with frequency of 9.69%. Phosphorylated Threonine has the lowest but significant frequency from position -10 to 10. Lysine has the significant frequency between 7.67% to 8.265% from position -10 to -6 and Glycine has significant frequency values 8.81% and 7.825% at positions -1 and +3.

Table Ia. Preferred amino acids and their positions around phosphorylated Ser.

Amino acids	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
A																				
C																				
D																				
E																				
F																				
G																				
H																				
I																				
K																				
L																				
M																				
N																				
P																				
Q																				
R																				
PS																				
NPS																				
PT																				
NPT																				
V																				
W																				
PY																				
NPY																				

Positively Preferred amino acid residues

Amino acids frequency around phosphorylated threonine

Proline has the set of highest frequency values in the frequency matrix with following distribution: 40.1% at position +1, 15% at +1 position, 14.240% at position +5, 12.18% at position -2, 11.55% at position -10, 10.12% at position +4 and frequency between 9% to 10% at positions -5, -8 and +6. Arginine is amino acid residue which shares second highest frequency set after proline with following distribution: 15.18% at position -3, 10.4% at position -5 and less then 10% frequency at positions +6, -7, -1, -2, +3 and +9. Phosphorylated Ser has significant but low frequency between positions -10 and +10 but non phosphorylated Ser has significantly high frequency value 11.23% at position -4. Similarly phosphorylated Thr has significant but low frequency between positions -9 to +9 and non phosphorylated Thr has comparatively high significant frequency at position +4. At position +7 non phosphorylated Tyr has 6.48% frequency, Phe has 5.22%, tryptophan has 3.63% and phosphorylated Tyrosine has 0.158% frequency. alanine also has high frequency at position +9 and +10 with values 12.5% and 11.7%. At +2 positions aspartic acid has 8.06%, cystine has 4.74%, and phosphorylated tyrosine has 1.5822% frequency.

Amino acids frequency around phosphorylated tyrosine

As compared to phosphorylated serine and phosphorylated threonine, phosphorylated tyrosine doesn't have observable high frequencies at any

Table Ib. Preferred amino acids and their positions around phosphorylated Thr

Amino acids	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
A																				
C																				
D																				
E																				
F																				
G																				
H																				
I																				
K																				
L																				
M																				
N																				
P																				
Q																				
R																				
PS																				
NPS																				
PT																				
NPT																				
V																				
W																				
PY																				
NPY																				

Positively Preferred amino acid residues

positions. Proline which has the highest occurrence at some positions in phosphorylated serine and phosphorylated threonine has significant frequency range between 5.48% to 6.206% at following positions in order to rising frequencies +10, +9, -10, +5, +8, -2, -8, -3. Similarly Arginine has frequency range between 5.338% and 5.601% at positions +7, -8, +4, -7 but all these positions have negative preference in phosphorylation of tyrosine. Phosphorylated tyrosine has negatively preferred frequency at positions -10, -9, -8 ,-7, -6, -5, -3, -2, +2, +3, +5 and positively preferred frequency at positions -4, -1, +1, +4, +6, +7.

Table Ia, Ib and Ic show the highlighted negatively and positively preferred frequencies at various positions around phosphorylated Ser, Thr and Tyr respectively, whereas Table IIa, IIb and IIc show the frequency values in percentage for each amino acids at various positions around phosphorylated Ser, Thr and Tyr, respectively.

DISCUSSION

The occurrence of amino acids at each position around the *O*- phosphorylated Ser, Thr and Tyr was computed. Presence of phosphorylated amino acid residues as compared to non phosphorylated amino acid residues around those amino acids that are to be phosphorylated has significant effect.

The phosphorylated amino acid residues, designated here as already - phosphorylated, were selected from phosphobase 3.0 as models, to generate phosphorylation data, for phosphorylation sites in the current investigation to facilitate development of prediction methods. Presence of

Table Ic.- Preferred amino acids and their positions around phosphorylated Tyr.

Amino acids	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
A																				
C																				
D																				
E																				
F																				
G																				
H																				
I																				
K																				
L																				
M																				
N																				
P																				
Q																				
R																				
PS																				
NPS																				
PT																				
NPT																				
V																				
W																				
PY																				
NPY																				

Positively Preferred amino acid residues

Negatively Preferred amino acid residues

phophorylated Ser and Thr in the vicinity of phosphorylation sites, designated here as to-be-phosphorylated Ser and Thr, have positively preferred significance. Phosphorylated Ser and Thr residues have no significant effect on to-be-phosphorylated Tyr. Presence of phosphorylated Ser and Thr at positions, -1 to –10 and +1 to +10, and -1 to –9 and +1 to +9 respectively, located continuously forming groups or/and clusters (Table Ia, Ib), around Ser and Thr residues, that are to-be-phosphorylated, is significantly effective. Presence of Pro and Arg around to-be-phosphorylated Ser and Thr is consecutively distributed at various positions (Table Ia, Ib). Phosphorylated Tyr around to-be-phosphorylated Ser has no significantly positive or negative effect (Table Ia, Ib). But phosphorylated Tyr has both negative and positive preference around to-be-phosphorylated Tyr. Proline has significant preference around to-be-phosphorylated Ser at various positions except -3, -4, +4 and +6. Similarly, Arg has no significant effect only on positions –10, +1, +2, +4 and +5 around to-be-phosphorylated Ser.

It is important to note that at position -5 and to some extent at position -2 the four residues including proline, arginine, phosphorylated Ser and phosphorylated Thr are significantly preferred around phosphorylated Ser and Thr sites. Whereas non- phosphorylated Ser is preferred only at position -2 around phosphorylated Ser. On the positions other than -2 and -5 these four residues did not occur simultaneously significantly. In spite of occurring simultaneously, possible combinations with different amino acid residues can also be identified at various positions other than -5 and -2.

It is proposed that preferred amino acid residues for phosphorylation on Ser and Thr is partly similar to sequence preference for Tyr. The results reported by Christlet and Veluraja (2000) has shown that Pro, Ser, Thr and Ala are highly preferred at various positions around glycosylated sites. In this study Pro, Ser, Thr and Arg are highly preffered around phosphorylated sites. Transitory interplay of PTMs is characteristic of multi-functional proteins, and the interplay is often between phosphorylation and glycosylation at hydroxy amino acids in proteins. It is likely that Ser or Thr around which Pro, Ser and Thr are present may be significant sites for the interplay of phosphorylation and glycosylation.

ACKNOWLEDGEMENTS

NUD thanks HEC Pakistan for DNP grant and PAS for partial support to this work. We also acknowledge use of Phohobase 3.0 provided by EBI.

REFERENCES

DIELLA, F., CAMERON, F, GEMÜND, C., LINDING, R., VIA. A., KUSTER, B., SICHERITZ-PONTÉN, T., BLOM. N. AND GIBSON, T., 2004. Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinform.,* **5**: 79.PMID: 15212693.

GILANI, F., 2004. Harness the features of C# to power your scientific computing projects. *MSDN Mag.*, **19**, no 3 (http://msdn.microsoft.com/msdnmag/issues/04/03/ScientificC/)

HUNTER, T., 1998. The Croonian lecture 1997: The phosphorylation of proteins on tyrosin: Its role in cell growth and disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **353:** 583-605.

KOLIBABA, K.S. AND DRUKER, B.J., 1997. Protein tyrosine kinases and cancer. *Biochim. biophys. Act,* **1333**: F217–F248.

SONGYANG, Z., BLECHNER, S., HOAGLAND, N., HOEKSTRA, M.F., PIWNICA-WORMS, H. AND CANTLEY, L.C., 1994. Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.,* **4:** 973–982.

CHRISTLET, T.H.T. AND VELURAJA, K., 2001. Database analysis of O-glycosylation sites in proteins. *Biophys. J.*, **80**: 952 – 960.

(Received 16 February 2006)

Table IIa. Frequency values of amino acids around phosphorylated Ser.

AA	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
A	7.24	6.43	7.09	7.5	7.2	6.5	6.2	5.51	7.7	8.23	3.6	7.09	7.13	6.25	6.98	6.7	7.1	6.36	6.8	6.83
C	1.25	1.25	1.43	0.7	0.6	0.7	0.8	0.92	0.7	0.66	0.51	1.21	0.77	0.84	0.84	0.6	0.8	0.7	0.7	0.84
D	4.78	4.89	4.78	4.5	5.2	5.1	5.3	5.03	5	5.84	7.53	7.24	7.09	5.99	6.76	7.1	5.9	5.14	6	5.29
E	8.01	8.01	7.94	8	7.8	8	8.3	6.1	6.5	6.76	6.28	10.5	14.5	8.63	10.1	9	8.2	8.6	8.5	8.27
F	2.35	2.28	2.2	2.2	2.8	2.8	2.5	1.84	1.8	3.12	4.11	1.43	1.69	2.65	2.17	2.8	1.9	2.65	2.5	1.98
G	6.83	6.8	6.14	7.4	6.2	7.2	6.4	7.35	5.5	8.82	5	6.58	7.83	6.32	6.43	7	6.8	6.87	6.9	6.58
H	1.51	1.84	1.8	1.7	1.9	1.6	1.5	1.25	1.4	1.73	0.77	1.29	1.25	1.4	1.43	1.7	1.7	1.8	1.7	1.36
I	3.27	2.72	3.23	2.9	2.2	3.2	2.7	2.2	2.2	3.09	2.35	2.76	1.58	4.08	2.98	2.6	3.1	3.05	2.2	3.31
K	7.68	7.86	7.94	7.9	8.3	7	7.1	6.72	5.5	4.56	2.24	5.58	6.69	5.51	6.1	7	6	6.76	7.4	6.54
L	6.65	6.21	6.28	5.8	6	7.5	6.4	5.8	5.1	9.07	5.91	5.4	5.03	6.83	6.39	6.4	6.8	6.28	6.1	6.17
M	1.76	2.39	2.5	1.4	1.6	1.9	1.7	0.99	1.6	1.84	1.18	0.88	1.21	2.2	1.58	1.4	1.4	1.32	1.4	1.43
N	2.68	2.94	3.27	3.5	3.1	2.6	2.7	2.46	2.8	4.11	1.73	2.79	3.16	3.01	2.39	2.5	3.3	2.72	2.8	2.68
P	7.79	7.9	8.6	8.9	8.7	8.6	7.1	6.94	9.3	7.9	33.5	8.27	8.89	7.42	9.04	7.2	8.7	8.23	8.1	9.22
Q	4.22	4.81	3.97	4.5	3.9	3.7	4.4	3.38	4.1	2.94	4.81	3.6	3.05	4.67	3.6	4.4	3.7	4.45	4.4	4.19
R	6.47	6.87	7.46	7.2	9.7	9.9	8.4	21	12	6.8	4	6.58	6.72	5.07	6.61	7.1	7	6.83	7.9	6.58
PS	2.28	1.95	2.5	2.5	3.1	3.3	5.6	3.78	5.2	3.71	3.71	5.18	3.78	5.62	3.31	3.1	2.5	2.5	1.9	2.28
NPS	8.89	9.96	8.63	9.3	8	8	10	7.2	10	8.78	2.76	9.15	8.19	8.71	9.18	8.8	8.6	9.18	8.8	9.7
PT	0.29	0.29	0.7	0.4	0.4	0.4	0.8	0.77	1.2	0.7	0.96	0.73	0.81	0.66	0.44	0.6	0.4	0.7	0.6	0.33
NPT	6.06	4.67	4.85	5	5.2	3.9	4.9	3.86	4.9	4.26	2.31	5	3.82	5.88	4.63	4.8	5.1	5.33	5.3	5.25
V	4.26	4.59	3.6	4.4	4.7	4	4	4	4.1	4.67	4.08	5.95	3.67	4.48	4.81	4	4.8	4.08	3.7	4
W	0.88	0.96	0.51	0.6	0.2	0.4	0.5	0.51	0.2	0.4	0.7	0.26	0.37	0.37	0.55	0.6	0.7	0.66	0.3	0.44
PY	0.04	0.04	0.07	0.1	0.1	0.1	0	0.07	0.3	0.07	0.15	0.11	0.18	0.11	0.07	0.1	0.1	0.07	0.1	0.15
NPY	1.73	1.8	2.28	1.7	1.8	2.2	1.7	1.43	2	1.65	1.58	1.73	1.29	1.73	1.69	1.7	2	1.95	1.5	1.73

Table IIb. Frequency values of amino acids around phosphorylated Thr.

AA	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
A	6.8	7.75	7.91	8.54	5.38	6.96	6.65	7.44	7.91	5.85	3.96	4.75	6.5	6.8	7.12	5.7	7.91	8.07	13	11.7
C	1.27	1.27	0.95	1.11	1.42	0.95	0.63	0.47	1.11	1.42	0.63	4.75	1.1	0.95	0.95	0.47	0.95	0.47	0.9	0.47
D	6.65	5.7	4.11	6.01	7.59	5.85	7.44	4.27	3.48	5.54	3.16	8.07	5.2	4.11	4.59	5.54	4.11	4.59	4.1	5.38
E	9.02	7.12	9.18	5.54	8.23	6.65	9.49	6.33	6.33	6.01	5.22	7.75	9.7	7.59	7.28	8.39	8.7	7.75	8.4	7.44
F	2.53	1.11	4.59	1.27	3.16	2.06	2.85	3.16	2.53	2.22	4.59	0.95	1.9	2.37	2.53	2.22	5.22	2.22	2.8	2.69
G	5.7	7.12	6.01	6.8	8.39	7.44	7.91	6.96	6.01	7.75	2.85	5.38	12	6.65	4.59	5.54	6.65	5.22	5.9	6.33
H	2.06	1.74	0.95	2.85	2.69	1.74	1.42	1.74	1.11	1.58	0.95	0.63	2.5	1.42	1.11	1.11	1.42	0.95	1.4	2.06
I	2.53	3.01	5.06	2.53	4.11	4.27	2.37	2.69	3.32	2.69	2.37	1.42	2.2	2.53	2.69	4.43	2.85	3.96	2.5	3.32
K	6.01	6.01	5.06	6.49	6.49	6.17	5.54	7.44	4.59	5.22	3.8	6.17	7.1	5.38	5.7	6.17	6.49	5.85	5.5	7.28
L	7.12	6.96	5.54	9.02	5.22	5.7	4.59	4.75	7.59	9.34	5.85	4.91	6.2	8.23	5.54	7.28	4.75	8.07	6.5	4.75
M	1.27	2.22	3.16	2.22	2.85	2.69	1.11	1.58	2.69	2.37	1.11	0.32	0.8	1.42	2.37	1.58	2.06	2.06	1.6	1.58

Continued

AA	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
P	7.28	7.28	9.49	6.8	6.96	9.02	7.59	8.39	12.2	5.7	40.2	15.5	7.3	10.1	14.2	9.97	7.12	7.28	7.6	11.6
Q	4.11	3.96	5.22	4.75	4.27	3.32	4.27	3.96	3.16	6.65	4.59	2.53	4.1	3.48	4.91	4.27	3.32	4.59	3.6	3.64
R	6.17	6.01	5.06	8.07	5.54	10.4	5.38	15.2	8.54	8.07	3.01	7.59	8.4	5.38	6.65	9.18	5.38	5.06	8.1	5.06
PS	1.42	2.53	3.01	1.9	2.37	1.9	2.85	3.48	3.16	4.11	3.01	5.06	3.3	3.48	1.74	1.58	1.74	3.01	1.3	1.27
NPS	9.18	8.86	5.22	8.23	9.02	6.65	11.2	5.38	8.86	7.12	2.69	4.59	6.3	8.23	6.17	6.17	7.44	9.49	9.3	8.7
PT	0.47	1.27	0.95	0.95	0.79	2.06	1.11	1.74	1.42	1.58	1.58	1.42	1.7	1.11	2.06	0.79	0.95	0.95	1.3	0.47
NPT	5.85	4.91	5.7	4.91	4.59	3.96	4.27	6.17	6.65	5.54	1.11	4.75	3.3	8.54	7.44	6.49	4.91	4.59	3.6	3.16
V	2.69	5.54	3.48	4.11	3.96	4.43	6.01	4.43	5.54	5.06	6.17	6.65	7.1	6.33	4.91	3.96	3.16	3.96	4.4	5.7
W	0.32	0.63	0.95	0.32	0.16	0.32	0.63	0.47	0.32	0.32	0.79	0.47	0.3	0.16	0.79	0	3.64	0.47	0.5	0.16
PY	0	0	0	0	0.16	0	0	0.16	0.16	0.47	0.47	1.58	0.2	0.16	0.32	0.16	0.16	0.16	0	0
NPY	2.69	2.06	1.42	1.27	1.74	2.22	1.58	1.11	1.42	2.69	0.79	1.58	0.9	2.06	1.42	2.85	6.49	4.75	2.2	0.79

Table IIc. Frequency values of amino acids around phosphorylated Tyr.

AA	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10
A	6.39	6.66	6.30	5.64	6.16	6.53	6.52	6.17	6.17	5.66	5.07	6.77	5.96	5.99	6.16	6.63	6.30	6.05	6.81	6.43
C	1.73	2.06	1.87	2.26	1.78	2.08	2.10	1.88	1.71	1.89	2.22	2.63	1.82	2.36	1.98	1.89	1.88	1.93	1.86	1.88
D	4.87	5.65	5.67	5.35	5.22	5.74	5.99	5.26	5.82	6.41	5.57	5.67	5.46	5.09	6.18	6.05	5.44	5.56	5.32	5.51
E	7.06	7.51	7.13	6.86	7.55	7.71	7.39	6.40	7.25	6.94	6.70	7.14	7.33	6.77	7.42	7.63	7.45	7.30	7.10	7.46
F	4.13	3.90	4.10	4.03	3.56	3.80	4.18	3.34	3.30	4.12	3.94	3.85	3.55	3.67	3.58	3.50	3.65	3.70	4.00	3.68
G	6.17	6.48	6.24	6.81	6.79	6.13	6.93	7.74	7.09	7.04	6.89	5.99	5.87	6.29	6.76	6.37	6.31	6.90	6.46	6.91
H	2.76	2.31	2.39	2.78	2.84	2.24	2.33	2.37	2.65	2.92	2.66	2.70	2.48	2.58	2.55	2.43	2.45	2.95	2.47	2.56
I	4.85	4.53	4.82	5.02	4.74	4.59	4.46	5.20	4.43	4.90	5.46	4.89	4.65	4.89	4.41	4.68	4.55	4.45	4.49	4.79
K	6.41	5.94	5.92	5.98	6.02	6.02	5.73	5.77	6.68	6.26	5.52	5.93	6.11	6.29	6.65	6.03	6.30	5.99	6.16	6.19
L	9.33	8.83	8.84	9.37	9.79	8.88	9.39	9.01	8.61	8.43	9.07	8.38	9.13	9.37	8.26	8.73	8.79	8.37	8.66	9.13
M	2.57	2.34	2.29	2.14	2.41	2.14	2.22	2.54	2.31	2.09	2.38	2.27	2.56	2.31	2.00	2.41	2.35	2.18	2.26	2.23
N	3.81	3.97	4.13	3.96	3.97	4.12	4.29	4.13	4.22	4.59	3.79	4.11	3.93	3.92	4.24	4.26	4.17	4.38	3.82	3.62
P	5.63	5.47	6.07	5.68	5.59	5.73	5.72	5.70	5.89	5.39	5.00	5.85	6.21	5.71	5.77	5.76	5.80	5.78	5.58	5.48
Q	4.43	4.70	4.15	4.43	4.66	4.98	4.48	4.56	4.69	4.30	5.16	4.59	4.92	4.43	4.64	4.21	4.19	4.33	4.70	4.55
R	5.49	5.64	5.46	5.60	5.60	5.18	5.29	5.75	5.90	5.34	6.12	5.60	6.10	5.52	5.86	5.50	5.24	5.73	6.09	5.15
PS	0.15	0.13	0.17	0.17	0.14	0.14	0.15	0.11	0.15	0.14	0.14	0.17	0.12	0.14	0.19	0.15	0.14	0.20	0.15	0.15
NPS	7.33	7.52	7.48	7.58	7.35	8.03	7.50	7.50	7.53	7.52	7.76	7.39	7.64	7.77	7.57	7.36	7.58	8.05	7.77	7.37
PT	0.02	0.04	0.09	0.13	0.06	0.03	0.04	0.02	0.03	0.02	0.05	0.03	0.02	0.03	0.04	0.03	0.03	0.03	0.04	0.05
NPT	5.04	5.03	5.37	5.03	5.11	5.38	5.19	5.12	5.32	5.19	5.54	5.05	5.30	5.03	5.08	5.19	5.13	4.88	4.81	5.23
V	6.41	6.23	6.26	5.92	5.81	5.82	5.68	6.74	5.89	6.09	5.98	6.39	5.80	6.90	5.57	5.80	6.04	5.57	5.96	5.56
W	1.23	1.09	1.18	1.25	0.91	1.21	0.88	1.09	1.01	1.27	1.31	1.00	1.11	1.07	1.21	1.10	1.77	1.09	1.03	1.31
PY	0.05	0.06	0.04	0.03	0.11	0.04	0.04	0.03	0.04	0.02	0.02	0.02	0.02	0.03	0.02	0.05	0.07	0.05	0.04	0.03
NPY	3.39	3.20	3.43	3.44	3.40	3.16	3.27	3.46	3.25	3.46	3.46	3.25	3.46	3.27	3.16	3.40	3.44	3.43	3.20	3.39