Pakistan J. Zool., vol. 46(2), pp. 409-416, 2014.

GEARS: A Genetic Algorithm Based Machine Learning Technique to Develop Prediction Models

Wajahat M. Qazi[1], Zeeshan Iqbal[2], M. Saleem Khan[1], Muhammad Rehan[1], Jawaria Munir[2], Abdul Rauf Shakoori[3] and Nasir-ud-Din[2,4*]

[1]*Cybernetic Intelligence Research Lab (Bioinformatics Group), Department of Computer Science, GC University Lahore*

[2]*Institute of Molecular Sciences & Bioinformatics, Lahore, Pakistan*

[3]*School of Biological Sciences, University of the Punjab, Lahore, Pakistan*

[4]*HEJ Research Institute of Chemistry, University of Karachi, Karachi, Pakistan*

Abstract.- The development of new prediction models to identify potential modified residues are based on different machine learning methods. Primary sequences, biochemical properties of the amino acids and 3D structural information of proteins are used to evolve prediction models. The information about the significant residues to govern different biological processes has not been considered yet to develop a prediction model. MAPRes is an efficient tool which has been utilized to mine significant residues and association patterns for surrounding amino acids of some specific modifications on hydroxyl and amino group such as phosphorylation and acetylation. The primary sequences of the proteins and association patterns of surrounding amino acids of modified residues may use to train new dataset for the development of an efficient and reliable prediction model. Biophysical and biochemical properties of the amino acids are also important parameters for the prediction of the modified residues. This study proposes, GEARS (Genetic Evolution of ClAssifers by Learning Residue Rules and Sequences), a classifier rule learning model, which considered different machine learning techniques such as ANNs, HMM and MAPRes were considered for the development of GEARS model. The GEARS, by combining these models, will have the capacity to reduce the false negative and positive predictions.

Key words: Machine learning, bioinformatics, post translational modification, artificial neural networks, backpropagation, genetic algorithm, association rules learning and mining, hidden markov models.

INTRODUCTION

Proteins are multi-functional molecules which often perform diverse functions (Jeffery, 1999). The 3D structural changes are one of the significant factor behind the functional switch of proteins (Bork *et al.*, 1998; Attwood, 2000; Li *et al*., 2004). These changes are regulated by a number of biochemical processes, particularly posttranslational modifications (PTMs) such as phosphorylation, sulfation, acetylation, methylation, glycosylation etc. (Varki and Kornfeld, 1980; Boynton *et al.*, 2001; Berlot *et al*., 2002; Yuan *et al*., 2003). These PTMs are usually take place at amino group, hydroxyl group, carboxyl group and side chain of the amino acids. The amino group of the amino acids modified by acetyl, methyl, formyl groups and etc. while modifications on hydroxyl group are *O*-Linked glycosylation and phosphorylation (Ahmad *et al.*, 2006; Butt *et al.*, 2011; Schiza *et al.*, 2013). These modifications are vitally important to regulate various biological processes and the development of the prediction models to identify the potential modification sites for different types of modifications are equally important. To analyze the environment and potential modified sites for different PTMs several machine learning techniques and statistical models have been developed that are playing vital role to understand the structure-function relationship of proteins (Christlet and Veluraja, 2001; Blom *et al*., 2004; Fujii *et al*., 2004 Lee *et al*., 2006; Qazi *et al*., 2006; Wong *et al*., 2007). The general classifications of these statistical and computational models are prediction, classification and clustering techniques. Indeed these models have assisted biologists to provide information about the complex biological processes that are extremely difficult to find out utilizing

* Corresponding author: prof_nasir@yahoo.com, professor_nasir@yahoo.com, chairman@imsb.edu.pk, Web: www.imsb.edu.pk.

0030-9923/2014/0002-0409 $ 8.00/0

existing experimental and/or theoretical studies. MAPRes (Mining Association Patterns among preferred amino acid Residues in the vicinity of amino acids targeted for post-translational modifications) is a useful computational model that can be utilize to mine association rules/patterns for significantly preferred amino acids in the vicinity of specific PTM sites (Ahmad *et al.*, 2008a). MAPRes is one of the example of clustering type protein sequence analyses.

MAPRes has been utilized to mine association rules/patterns for the surrounding amino acids that are biologically important in the vicinity of phosphorylation, glycosylation and acetylation sites (Ahmad e*t al*., 2008a,b, 2009; Khawaja *et al*., 2008; Iqbal *et al.*, 2013)**.** The upgraded version of MAPRes has also been released (unpublished). This version of MAPRes provides mechanism to mine association rules by utilizing the properties of neighboring amino acids. This version of MAPRes has been applied to study the acetylation sites based on the properties of surrounding amino acids such as polarity and charge (Iqbal *et al.*, 2013). These studies have encouraged that association rules can be used to build a better classifier by combining it with sequence learning.

Indeed existing predictive and descriptive analyses methods are either based on data learning method (Blom *et al*., 2004) or data mining for association analyses (Ahmad *et al*., 2008a). This suggests, in either case, these methods lack some significant aspects associated with data. Hence new methods are required to develop prediction models, utilizing the data derived from previous analyses in form of certain rules/patterns coupled with protein sequence data, by developing an assembly of learning schemes for different types of data. The concept of combinatorial classification for the development of prediction models has already been practiced such as DictyOGlyc (Gupta *et al*., 1999). In this prediction method, the jury of 6 ANNs with different configurations and architectures were utilized but the learning data was in form of protein sequence windows only.

This study reports the GEARS model which is a novel machine leaning simulator (under development). The development phases are divided into various functional and reportable deliverable. This study will report following functional deliverables, conceptual overview of the simulator and its utilization scheme, bio-data preprocessing modules, MAPRes and neural network subsystems. Functional modules (MAPRes and Neural Network) reported here can be used to train a classifier based on the induction of data as well as significant biological rules capable to perform prediction. These reported modules will use to develop a novel prediction model with probably highest rate of accuracy and efficiency. In this study, a test dataset of phosphorylated-tyrosine (P-Tyr) modified by EGFR kinase were brought into account to validate the function and hypothesis of GEARS model. Not only the GEARS model but this test module too have the capacity to train data for the prediction of other modified residues. The availability of the data is the key factor for the training of these suggested modules.

MATERIALS AND METHODS

The GEARS model

GEARS, is a classifier evolution and learning manager that work on the bases of parallel and distributed computing. The structure of the GEARS is designed on layered and plug-in architecture. The layered architecture will have the capacity to handle the computational complexities of genetic algorithm to evolve respective models while plug-in architecture will provide the flexibility to handle different types of data servers. Software artifacts of GEARS system are grouped into three layers (data repository layer, windows application module layer and windows service module layer) (Fig. 1). On the bases of these layers, the GEARS model can be integrated into three basic modules which are further divided into some subsystems. The development of GEARS model is executing in C# (Fig. 2.). The overview of the main modules and subsystems of GEARS model are described below.

Windows application modules layer

In this module many software artifacts are composed in a group. These software artifacts are Classifier Evolution and Learning Manager (CEL Manager), Job Monitoring Console (JMC) and Project Registry Explorer (PRE).

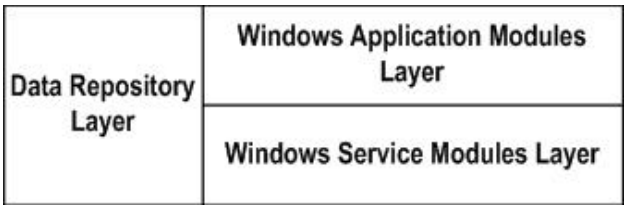

Fig. 1. A block diagram of GEARS System containing data repository layer, windows application module layer and windows service module layer.

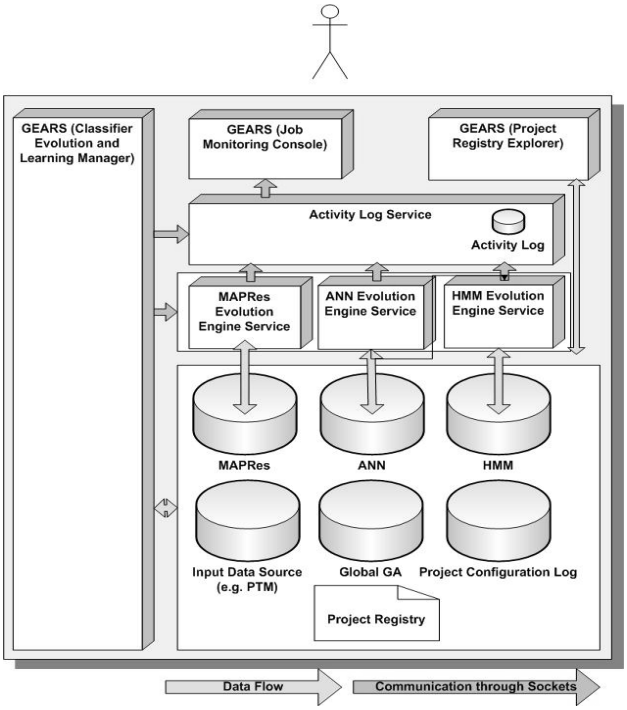

Fig. 2. System view and architecture of GEARS showing the data flow and training modules of protein sequence learning by ANNs and HMMs and rules learning mined by MAPRes and finally the evolution of classifiers and combining the three models into a single model.

The CEL Manager will provide an interface to human user to create and configure GEARS project. In this module, after the defining the project, rest of the configurations such project registry and repositories will be created by GEARS itself. This application will also provide the interface to run the learning and evolution simulation. CEL Manager will communicate with configuration parameters like the software artifacts (Services) in Windows Service Modules Layer (WSML). Services in WSML will be responsible to configure and run the respective jobs. Data management of input and output data sources will be mange by Data Repository Layer (DRL) that will be based on plug-in architecture. The user required plug-ins will be select at the time of project creation and application will load the installed plug-ins. This will make the GEARS' framework flexible enough to handle any type of data server.

In next step, Job Monitoring Console application will provide the interface to illustrate the job status, progress of dot net and grid threads, learning curves etc. This application receives data streams from Activity Log Service (ALS) (Fig. 2).

In last step, the Project Repository Explorer (PRE) will be utilize to browse the reports, configurations and architecture of evolved classifier candidate. This explorer will also provide the support to generate and compile the selected candidate as a dot net (DLL) component.

Windows service modules layer

This layer will consist of different Windows services which are the part of GEARS subsystems such as MAPRes Evolution Engine Service, ANN Evolution Engine Service, HMM Evolution Engine Service and Activity Log Service (Fig. 2). The idea behind splitting of GEARS application is to balance the computational load in distributed fashion. Each distribution is a grid application (except for Activity Log Service) and will be responsible to perform their specific tasks. Application Log Service which is an interface between other GEARS services, CEL Manager and Job Monitoring Console. This service will have the responsibility to entertain incoming data streams from other GEARS services and CEL Manager, writes in a log database and stream them to Job Monitoring Console application. The data streams will contain the information about the status and progress of each dot net and grid thread executing in other services.

The MAPRes, ANN and HMM Evolution Engine Services are the examples of grid services and the development of these models are based on genetic algorithm. These Windows evolution engine services are responsible to evolve their respective classifier models *e.g.*, MAPRes Evolution Engine

Services will evolve MAPRes classification model. Each model will be executed as grid thread and to enhance the further computational efficiency the dot net threads will be executing within each candidate model.

Data repository layer (DRL)

All Windows applications and services which are the members of GEARS system, will communicate with DRL with exception for ALS and JMC. Activity log database is not the part of DRL because log database can be viewed as temporary storage. On every execution of CEL Manager, activity log database will be cleaned. Whereas, data objects in DRL will persist in their states as long as project registry is not deleted by Project Registry Manager. In other words, each project will have its own unique DRL profile and will be treated as internal storage. As mentioned earlier that access to data objects in DRL will be performed through plug-in interface, means the GEARS implementation is not bound to a specific arrangement of DRL data objects and their database servers.

DRL management plug-ins is the implementation of DRL management interface library. MYSQL server based DRL management (MySQLDRL) plug-in will be implemented as default. Each data object in MYSQLDRL will be treated as a separate database, so that each data object can utilize maximum data storage. If required in any study to have different management strategy and/or database server, one can implement his/her own DRL management plug-in by extending interfaces defined in DRL management Interface Library.

Data object in DRL are following: MAPRes Data Object, ANN Data Object, HMM Data object, Input Source Data Object, Global Genetic Algorithm (Global GA) Data Object, Project Configuration Log and Project Registry (Fig. 2).

Data Objects associated to MAPRes, ANN and HMM are responsible to hold information on the respective model architecture and configurations, learning curves and error rates, training, validation and testing data sets, testing/validation reports for each candidate against its ModelID, Generation No. and initial conditions. Whereas, Global GA Data Object will be responsible to persist information on overall learning profile including learning curves and error rates of candidate models in a combinatorial way (ensemble mode), accommodating the new configurations of the data including size of peptides, training/validation/testing datasets.

Several components of GEARS have been developed and tested such as Project Creation Module, Main Console Module and Ensemble Learning Module. In current study, the justifications of relations behind the GEARS is performed by considering the assembled learning using ANN and MAPRes subsystems to evolve a model to predict the modification potential of P-Tyr data modified by EGFR kinase.

Test model to predict the modification potential of P-Tyr data

A part of the proposed model was tested by developing a prediction-model for the prediction of the P-Tyr modified by EGFR kinase. The P-Tyr data based on EGFR kinase was obtained from Phospho.ELM. Data preprocessing and cleaning was performed by using Data Inconsistency modules of MAPRes in order to remove the inconsistencies in downloaded data from Phospho.ELM.

After cleaning of data, the MAPRes model was utilized to mine the association patterns from the resultant dataset. MAPRes mine these association patterns at different support levels (5%, 10%, 15%, 20% and 25%). The resultant dataset of P-Tyr were also take into account to make a peptide dataset of modified residues, the length of each peptide were consist of 21 amino acids (modified residue at 0 position and 10 amino acids at each side). This peptide dataset were utilized to train and validate the neural networks. The association patterns mined by MAPRes were compared with each peptide of the peptide dataset to find the suitable match. The association patterns and peptides which were matched, combine as a single record. To generate the binary form of the amino acids, sparse encoding was applied on the peptide dataset. For each amino acid a binary code of 21 digit was generated hence a 420 digit code represent a single peptide. The position “0” was not considered for sparse encoding as it is populated

only with Tyr. The flow diagram for the current study are shown in Figure 3.

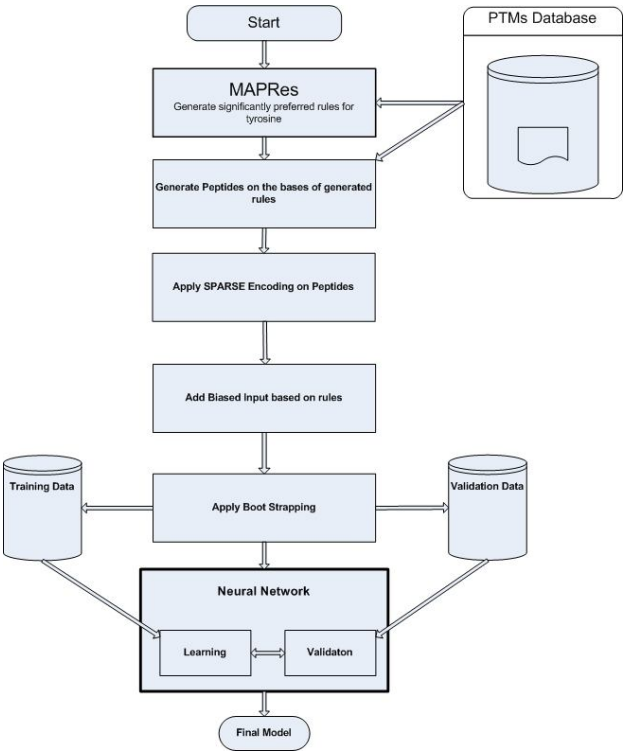

Fig. 3. Flow diagram of the data trained by utilizing ANN. The MAPRes rules/patterns and peptide dataset were generated from assembled dataset of P-Tyr modified by EGFR kinase. The sparse coding apply on the peptide and rules dataset to generate the binary code of the amino acids. On the bases of MAPRes rules the biasness factor introduce against each peptide for better learning of the ANNs. Boot strapping were performed for the validation and training of the dataset.

The biasness factor was added against each encoded peptide on the bases of MAPRes rules for effective learning of neural network. After the calculation of the biasness factor, the bootstrap sampling process was introduced to train and validate the peptide dataset. The training of the neural network was performed by varying the values of the sigmoidal lambda and root mean square (RMS) error. There were 21 experiments in total was performed by changing these values to identify the neural network configuration with optimized values, maximizing Mathew's Correlation Coefficient (MCC), sensitivity, specificity, accuracy and minimizing RMS error (Table I).

Table I.- Parameter alterations for training of ANNs.

Sr.	Sigmoidal Lambda	Acceptable RMS Error
1.	0.01	0.07
2.		0.18
3.	0.02	0.07
4.		0.075
5.		0.09
6.		0.1
7.		0.1
8.		0.12
9.		0.13
10.	0.03	0.05
11.		0.06
12.		0.07
13.		0.08
14.		0.1
15.		0.11
16.		0.12
17.	0.04	0.05
18.		0.05
19.		0.08
20.		0.1
21.	0.05	0.04

Table II.- Association patterns mined by MAPRes.

Sr.	Association rule	Support level
1.	<E,-3><V,1>=>PP[1]	5
2.	<G,-6><E,-3>=>PP	5
3.	<G,-6><V,1>=>PP	5
4.	<E,-3>=>PP	20
5.	<G,-6>=>PP	20
6.	<V,1>=>PP	24
7.	<F,-4>=>NP[2]	5
8.	<F,-9>=>NP	5
9.	<I,2>=>NP	5
10.	<I,-7>=>NP	5
11.	<K,-3>=>NP	5
12.	<Q,8>=>NP	5
13.	<R,5>=>NP	5

[1]PP represents positively P-Tyr
[2]NP represents non P-Tyr

RESULTS AND DISCUSSION

The GEARS simulator tends to evolve amino acid sequence and association rule. The classifiers

Table III.- Input and output parameters for training of ANNs.

Sr. No	Sigmoidal Lambda	Acceptable RMS Error	MCC	Sensitivity	Specificity	Accuracy
1.	0.03	0.11	0.538703	1	0.719697	0.764331
2.	0.04	0.1	0.538703	1	0.719697	0.764331
3.	0.04	0.08	0.545951	1	0.727273	0.770701
4.	0.03	0.12	0.545951	1	0.727273	0.770701
5.	0.03	0.08	0.553342	1	0.734849	0.77707
6.	0.03	0.1	0.553342	1	0.734849	0.77707
7.	0.03	0.07	0.592655	1	0.772727	0.808917
8.	0.02	0.13	0.627409	1	0.80303	0.834395
9.	0.02	0.1	0.646096	1	0.818182	0.847134
10.	0.04	0.05	0.655806	1	0.825758	0.853503
11.	0.04	0.05	0.655806	1	0.825758	0.853503
12.	0.03	0.06	0.67603	1	0.840909	0.866242
13.	0.02	0.12	0.732015	1	0.878788	0.898089
14.	0.02	0.09	0.770131	1	0.901515	0.917198
15.	0.02	0.07	0.770131	1	0.901515	0.917198
16.	0.02	0.1	0.783741	1	0.909091	0.923567
17.	0.05	0.04	0.797856	1	0.916667	0.929936
18.	0.03	0.05	0.797856	1	0.916667	0.929936
19.	0.02	0.075	0.797856	1	0.916667	0.929936
20.	0.01	0.18	0.800714	0.68	1	0.949045
21.	0.01	0.07	0.8436007	1	0.939394	0.9490446

are evolve in this method by utilizing genetic algorithm. The classifiers which are evolved as a jury of machine learning models are MAPRes, ANNs and HMMs. GEARS uses the concept of heterogeneous jury of aforementioned models. Every model have its own inherited strengths and different nature of learning *e.g.*, ANNs are empirical and connectionist system, HMMs have statistical grounds whereas MAPRes offers a rule learning mechanism from unsupervised to supervised learning. This combinatorial scheme of three machine learning techniques in GEARS provides it the capability to accommodate complexities in protein data.

In this specific case of study, MAPRes and neural networks which are the two basic proposed modules of GEARS methodology were utilized to perform sequence learning with rule inductions. Association rules which were used for rule induction in the form of biasness factor were generated by MAPRes and the sequence learning was performed by using neural networks.

In first step of this method, the association rules were generated at various support levels by using MAPRes (Table II). At different support levels, some similar association rules were found that were removed from the list and kept only one rule that have the maximum support level. After removing the similar rules from the list, the remaining association rules were have only three support levels 5, 20 and 24. There were only 13 association rules were left in the list and out of these total association rules, 6 were found for positively P-Tyr (PP) and 7 for non-phosphorylated Tyr (NP). Among the 6 association rules for PP, 3 were found at 5% support level, 2 were found at 20% support level and 1 at 24% support level. Whereas, in case of NP all association rules were found at 5% support level. The association patterns indicated that the Val at +1 position were found at maximum support level (24%), while, Glu at -3 position and Gly at -6 position have considerably high support level. It is interestingly noticed that all these three amino acids, found in surrounding of PP with maximum support level are non-polar.

There were 21 experiments were performed to train the neural networks by changing the values of the sigmoidal lambda and root mean square (RMS) error (Table III). These trainings of the data are critical for the prediction of potential P-Tyr. As

mentioned earlier sigmoidal Lambda and acceptable RMS were the training configurations, whereas the training performance of the neural networks was measured on the bases of following performance parameters such as Mathew's correlation coefficient (MCC), sensitivity, specificity and accuracy. These experiments were conducted with 5 different values of sigmoidal lambda. Variation in root mean square (RMS) error was recorded during the learning phase. The value of sigmoid lambda ranges from 0.01 to 0.04 and RMS error from 0.04 to 0.18 correspondingly set the value of accuracy. The experiment 17 to 21 have shown the highest values of the accuracy than the previous experiments (Table III). It is indicating that the training of the data was good in these experiments. The maximum learning of the data was done in last experiment with maximum value of accuracy. The finally selected experiment for prediction which has MCC = 0.8436007, Sensitivity = 1, Specificity = 0.939394 and Accuracy = 0.9490446. This finally selected trained dataset will have the capacity to develop a reliable prediction model of potential P-Tyr residues.

In conclusion, the development of new computational prediction models by considering the combinatorial approaches is the need of the hour to improve the accuracy of the prediction models. This study proposes the architecture of the GEARS model which is designed to combine the different data learning and data mining techniques. The GEARS model will have the capacity to predict the modified residues with better accuracy. A part of the GEARS model have been developed and results of this module have shown a good values of data learning or accuracy. This test module were utilized to train the data of P-Tyr but it also have the capacity to execute on other modified residues such as Ser, Thr, Lys and Gln. The extensive data available for phosphorylation of OH- group compels to investigate this particular study. The modification on NH_3- group such as acetylation and methylation are equally important and control the various function of the proteins. In future, the data from other modification will also be considered for the training of developed module of the GEARS.

ACKNOWLEDGEMENTS

Nasir-ud-din acknowledges financial support from Pakistan Academy of Sciences for amino group studies.

REFERENCES

AHMAD, I., HOESSLI, D.C., WALKER-NASIR, E., CHOUDHARY, M.I., RAFIK, S.M., SHAKOORI, A.R. AND NASIR-UD-DIN, 2006. Phosphorylation and glycosylation interplay: protein modifications at hydroxy amino acids and prediction of signaling functions of the human beta3 integrin family. *J. cell. Biochem.*, **99**: 706-718.

AHMAD, I., HOESSLI, D.C., QAZI, W.M., KHURSHID, A., MAHMOOD, A., WALKER-NASIR, E., AHMAD, M., SHAKOORI, A.R. AND NASIR-UD-DIN, 2008b. MAPRes: An efficient method to analyze protein sequence around post translational modification sites. *J. cell. Biochem.*, **104**: 1220–1231.

AHMAD, I., MEHMOOD, A., KHURSHID, A., QAZI, W.M., HOESSLI, D.C., WALKER-NASIR, E., SHAKOORI, A.R. AND NASIR-UD-DIN, 2009. Phosphoproteome sequence analysis and significance: Mining association patterns around phosphorylation sites utilizing MAPRes. *J. cell. Biochem.*, **108**: 64-74.

AHMAD, I., QAZI, W.M., KHURSHID, A., AHMAD, M., HOESSLI, D.C., WALKER-NASIR, E., CHOUDHARY, M.I., KHWAJA, I., SHAKOORI, A.R. AND NASIR-UD-DIN, 2008a. MAPRes: Mining association rules among preferred amino acid residues in the vicinity of target amino acids in proteins. *Proteomics*, **8**: 1954–1958.

ATTWOOD, T., 2000. The quest to deduce protein function from sequence: the role of pattern databases. *Int. J. Biochem. Cell Biol*., **32**:139-155.

BERLOT, S., AISSAOUI, Z., PAVON-DJAVID, G., BELLENEY, J., JOZEFOWICZ, M., HELARY, G. AND MIGONNEY, V., 2002. Biomimetic poly(methyl methacrylate)-based terpolymers: modulation of bacterial adhesion effect. *Biomacromolecules*, **3**: 63-68.

BLOM, N., SICHERITZ-PONTEN, T., GUPTA, R., GAMMELTOFT, S. AND BRUNAK, S., 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, **4**:1633-1649.

BORK, P., DANSEKAR, T., DIAZ-LAZCOZ, Y., EISENHABER, F., HUYNEN, M. AND YUAN, Y., 1998. Predicting function: from genes to genome and back. *J. mol. Biol*., **283**:707-725.

BOYNTON, J.C.R., HEINEGARD, D. AND BARRY, F., 2001. Post translational modifications to human bone sialoprotein determined by mass spectrometry. *Biochemistry*, **40**:12983-12991.

BUTT, A., KALEEM, A., IQBAL, Z., WALKER-NASIR, E., SALEEM, M., SHAKOORI, A.R. AND NASIR-UD-DIN., 2011. Functional regulation of DNA binding of FOXO1 by post translational modifications: *In silico* study. *Pakistan J. Zool.*, **43**: 1167-1175.

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

FUJII, K., ZHU, G., LIU, Y., HALLAM, J., CHEN, L., HERRERO, J. AND SHAW, S., 2004. Kinase peptide specificity: improved determination and relevance to protein phosphorylation. *Proc. natl. Acad. Sci., USA*, **101**: 13744-13749.

GUPTA, R., JUNG, E., GOOLEY, A.A., WILLIAMS, K.L., BRUNAK, S. AND HANSEN, J., 1999. Scanning the available *Dictyostelium discoideum* proteome for O-linked GlcNAc glycosylation sites using neural networks. *Glycobiology*, **9**: 1009-1022.

IQBAL, Z., HOESSLI, D.C., KALEEM, A., MUNIR, J., SALEEM, M., AFZAL, I., SHAKOORI, A.R. AND NASIR-UD-DIN, 2013. Influence of the sequence environment and properties of neighboring amino acids on amino-acetylation: relevance for structure-function analysis. *J. cell. Biochem*., **114**: 874-887.

JEFFERY, C.J., 1999. Moonlighting proteins. *Trends biochem.. Sci*., **24**:8-11.

KHWAJA, T.A., WAJAHAT, T., AHMAD, I., HOESSLI, D.C., WALKER-NASIR, E., KALEEM, A., QAZI, W.M., SHAKOORI, A.R. AND NASIR-UD-DIN, 2008. *In silico* modulation of apoptotic Bcl-2 proteins by mistletoe lectin-1: functional consequences of protein modifications. *J. cell. Biochem.*, **103**: 479–491.

LEE, T., HUANG, H., HUNG, J., HUANG, H., YANG, Y. AND WANG, T., 2006 dbPTM: An information repository of protein post translational modification. *Nucl. Acids Res.*, **34**: D622--D627.

LI, J., BIGELOW, D.J. AND SQUIER, T.C., 2004. Conformational changes within the cytosolic portion of phospholamban upon release of Ca-ATPase inhibition. *Biochemistry*, **43**:3870-3879.

QAZI, W.M., SHAKOORI, A.R. AND NASIR-UD-DIN., 2006. Rules required for phosphorylating tyrosine using as a model kinase EGFR utilizing AME. In: *Abstracts of International Symposium on Nano Chemistry: Chemistry, Biochemistry, Molecular Biology & Bioinformatics of Enzymes*, pp. 10.

SCHIZA, V., MOLINA-SERRANO, D., KYRIAKOU, D., HADJIANTONIOU, A. AND KIRMIZIS, A., 2013. N-alpha-terminal acetylation of histone H4 regulates arginine methylation and ribosomal DNA silencing. *PLoS. Genet.*, **9**: e1003805.

VARKI, A. AND KORNFELD, S., 1980. Structural studies of phosphorylated high mannose-type oligosaccharides. *J biol. Chem*., **255**:10847-10858.

WONG, Y., LEE, T., LIANG, H., HUANG, C., WANG, T., YANG. Y., CHU, C., HUANG, H., KO, M. AND HWANG, J., 2007. KinasePhos2.0: A web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucl. Acids Res.*, **35:**W588-594.

YUAN, Z.Q., FELDMAN, R.I., BUSMAN, G.E., COPPOLA, D., NICOSIA, S.V. AND CHENG, J.Q., 2003. AKT2 inhibition of capsulation-induced JNK/p38 and box activation by phosphorylation of ASK1: implication of AKT2 in chemo resistance. *J. biol. Chem*., **278**:23432-23440.

(Received 13 March 2009, revised 15 February 2014)