Prediction of Specificity and Cross-Reactivity of Kinase Inhibitors

Nitish Kumar Mishra and Gajendra P.S. Raghava*

Institute of Microbial Technology, Sector 39A, Chandigarh, India

Received July 13, 2010: Revised October 21, 2010: Accepted October 22, 2010

Abstract: Designing kinase inhibitors is always an area of interest because kinases are involved in many diseases. In the last one decade a large number of kinase inhibitors have been launched successfully; six inhibitors have been approved by FDA and more are under clinical trials. Cross-reactivity or off-target is one of the major problems in designing inhibitors against protein kinases; as human, have more than 500 kinases with high sequence similarity. In this study an attempt has been made to develop a model for predicting specificity and cross-reactivity of kinase inhibitors. The dataset used for testing and training consists of binding affinities of 20 chemical kinase inhibitors with protein kinases.

We developed QSAR based SVM models for predicting binding affinity of an inhibitor against protein kinases using most relevant 5,10 and 15 structure descriptors and achieving average correlation of 0.64, 0.488 and 0.442 respectively. In order to predict specificity and cross-reactivity of an inhibitor, we developed 16 QSAR based SVM models for 16 protein kinases; one model for each kinase. We achieved average correlation 0.719 between actual and predicted binding affinity using kinase specific models. Based on the above study a web server DMKPred has been developed for predicting binding affinity of a drug molecule with 16 kinases. The SVM based model used in this study can be used to predict kinase specific inhibitors. This study will be useful for designing kinase specific inhibitors.

Keywords: Support vector machine (SVM), Molecular descriptors, kinase inhibitors, QSAR, dissociation constant, prediction.

INTRODUCTION

Protein kinases are critical components of cellular signal transduction, directly involved in many diseases including cancer and inflammation. Thus kinases are one of most important drug targets [1-5]. This is the reason that most of the pharmaceutical companies are concentrating on designing of kinases inhibitors [6]. Fortunately researchers achieved success in developing six drug molecules approved by FDA (Gleevec, Iressa, Tarceva, Erbitux, Herceptin, Nexavar) against kinases and 40 are under clinical trial [7]. One of the major challenges in designing kinases inhibitors is cross-reactivity or specificity [8]. There is need to develop a method for estimating cross-reactivity in order to discover inhibitors against specific kinase. It is not practically possible to examine cross-reactivity of an inhibitor using experimental technique. Thus there is a need to develop *in-silico* high throughput techniques for screening chemical libraries against kinases.

Researchers are trying to understand the specificity and mechanism of action of kinase inhibitors [9]. Recently Fabin *et al*. [10] studies the cross reactivity of kinase inhibitors at a large scale where they calculate K_d of 20 kinase inhibitors against 119 protein kinases using a high throughput phase display method. They have several observations that include- i) most of the protein kinase inhibitor targets of the ATP binding site, and because more than 500 protein kinases are identified in human genome have an ATP site with very high similarity [11], there is great potential of cross-reactivity, ii) Specificity varies widely and is not strongly correlated with the chemical structure or identity of the intended target and off-targets, iii) compounds which may bind with one sub-family of protein kinase also bind to other sub-family of protein kinase with good affinity and iv) some allosteric site binding inhibitors also bind with other kinases with good K_d.

In this study a systematic attempt has been made to understand cross-reactivity and specificity of kinase inhibitors. First, we developed common models for predicting inhibition capability of an inhibitor against any protein kinase. Then we developed kinase specific models for predicting inhibition potential of a molecule against a desired kinase. These models will be useful for predicting specificity and cross-reactivity of kinase inhibitors.

METHODS

Dataset of Inhibitors

We extracted 20 kinase inhibitors and their experimentally validated dissociation constant with 119 kinases [10]. It has been observed that for a number of protein kinases dissociation constant was 10 (no significant binding affinity). We create a clean data set where we take only those kinases which satisfy the following two criterias; i) Kinases for which six or more than six chemicals have significant inhibition ($K_d < 10$) and ii) Kinases for which SVM models were developed in reasonable time.

Molecular Descriptors

In order to understand the property of a chemical molecule, it is important to calculate its molecular descriptors. In chemoinformatics the structural feature of the individual molecules are derived from the molecular structure, so called descriptors [12]. Molecular descriptors include constitutional, topological, geometrical, physiochemical, electrostatic descriptors.

*Address correspondence to this author at the Institute of Microbial Technology, Sector 39A, Chandigarh, India; Tel: +91-172-2690557, 2695225; Fax: +91-172-2690632, 2690585; E-mail: raghava@imtech.res.in

1570-1808/11 $58.00+.00

In this study our aim was to develop models for academic use. Thus, we calculated molecular descriptors using two freely available software packages Molinspiration and PreADMET.

Molinspiration

Molinspiration [13] is a freely available online software for calculating molecular descriptors. This software calculates total nine molecular descriptors which are commonly used in QSAR studies. We use eight molecular descriptors for all kinases in our generalized models; these descriptors are milogP (octanol / water partition coefficient), TPSA (topological polar surface area), nAtoms (number of atoms), MW (molecular weight), nON (Number of Hydrogen bond acceptor), nOHNH (Number of Hydrogen bond donar), nrotab (number of rotable bonds) and volume.

PreADMET

This program allows to calculate more than 900 molecular descriptors for a molecule [14]. It is not practically possible to use all these descriptors for developing any model because these descriptors include a number of redundant descriptors; dimensionality of feature space is also very high. In order to reduce the descriptors dimensionality and selecting the most relevant descriptors, we used the following criterias. We removed all those descriptors, which had no value for any kinase. We also removed all those descriptors which had value zero for more than 10. We Selected only those molecular descriptors, that were having Pearson's correlation less than 0.8. In this way we finally got 60 molecular descriptors. In addition to 60 descriptors, two molecular descriptors were also calculated using Molinspiration, they were also used for developing models.

The molecular descriptors were further reduced from 62 to 15 descriptors based on correlation between K_d & descriptors. We used top 5,10 and 15 descriptors for developing models (Table **1**). For developing general QSAR model we used only the descriptors having highest average correlation. In order to develop, kinase specific model we used the descriptors having highest correlation with K_d of a particular kinase.

SVM Algorithm

An excellent machine learning technique SVM has been used for the prediction of K_d. The theory of SVM has been extensively described in literature [15]. In the present study, a freely downloadable package of SVM, SVM_light has been used to predict the K_d of kinase inhibitors with specific protein kinases. The SVM_light software is downloaded from http://www.cs.cornell.edu/People/tj/svm_light/. This software enables users to define a number of parameters as well as to select from some inbuilt kernel functions, including a Radial Basis Function (RBF), polynomial and linear kernel. We used different parameters and different kernels for our study.

Construction of SVM Models

In this study regression models have been developed, where descriptors are used as dependent or input variables and inhibition constant as output or independent variable. We use different kernels and optimizing parameters to generate SVM model. These models are called QSAR based SVM models, because they compute inhibition constant from descriptors of kinase inhibitors.

Evaluation of Models

In the present study we used Jack-Knife test/leave one out cross-validation (LOOCV) technique for evaluating our models. In LOOCV one chemical kinase inhibitor has been

Table 1. List of Top 15 Molecular Descriptors of PreADMET, these Descriptors are Ranked Based on Correlation Between Descriptor and Inhibition Constant (K_d)

Descriptor	Correlation	Descriptor category
Negative charge SA	0.24	Electrostatic descriptor
2D VSA H bond acceptor	0.18	Geometrical descriptor
SK log P_{VP}	0.15	Physiochemical descriptor
SK log P value	0.14	Physiochemical descriptor
2D VSA H bond all	0.13	Geometrical descriptor
Maximum positive charge	0.13	Electrostatic descriptor
Charge polarization	0.12	Electrostatic descriptor
Fraction of 2D VSA polar	0.12	Geometrical descriptor
Relative positive charge	0.11	Electrostatic descriptor
Molecular weight	0.10	Constitutional descriptor
SK BP	0.10	Physiochemical descriptor
Polarity parameter	0.09	Electrostatic descriptor
SK log D value	0.09	Physiochemical descriptor
Aromatic bond	0.09	Constitutional descriptor
SK log S value	0.07	Physiochemical descriptor

used for testing and remaining chemicals for testing, this process repeats *n* times in such a way that each chemical kinase inhibitor is used for testing, where *n* is the total number of chemicals. In order to assess performance we compute Pearsons's correlation coefficient R, between predicted and actual K_d value using the following formula.

$$R = \frac{n\sum K_d^{act} K_d^{pred} - \sum K_d^{act} \sum K_d^{pred}}{\sqrt{n\sum \left(K_d^{act}\right)^2 - \left(\sum K_d^{act}\right)^2}\sqrt{n\sum \left(K_d^{pred}\right)^2 - \left(\sum K_d^{pred}\right)}}$$

Where n is the size of test set, K_d^{pred} is the predicted dissociation constant and K_d^{act} is the actual dissociation constant. Value of R always ranges from -1 to +1 negative.

RESULTS

General Models

QSAR based SVM models have been developed for predicting kinase inhibitors as potential drug molecules. These models were developed for whole of kinase family rather than a specific member of kinase family. The objective of this model is to examine the inhibition potential of a chemical molecule to inhibit the proteins of a kinase family. We achieved correlation (R) 0.647, 0.488 and 0.442 respectively for 5, 10 and 15 for best molecular descriptor models. We achieved R more than 0.5 for four protein kinases using Molinspiration model. In case of 15 descriptors model we achieved correlation more than 0.5 for four kinases out of these one has more than 0.65 correlation. In case of 10 descriptor models 9 kinases have correlation more than 0.5 including two, that have more than 0.65 and no model generated in case of AAK1. In case of 5 descriptors based models 13 kinases have more correlation than 0.5 including eight which have more than 0.65. We used 8 Molinspiration descriptors for developing model and achieved correlation 0.398 (Table **2**).

Kinase Specific Models

Our above models predict whether a chemical will inhibit protein kinases or not but they did not provide any information about specificity and cross-reactivity because we used a same set of molecular descriptors for all kinases which have better correlation with dissociation constant. Here an attemp has been made to develop kinase specific models where the separate models have been developed for each member of kinase family. These models will be helpful in predicting specificity and cross-reactivity of kinase inhibitors. Our above result shows that both ten and five descriptor models performed very well for many kinases and both have some advantage on each other. So in this study we used both models and computed their performance, we also computed to a performance of models that were developed, using top negatively correlated descriptors. We achieved overall correlation 0.719 using five descriptor models with 16 protein kinases which have correlation more than 0.5 including 9 protein kinases which have correlation more than 0.65. We also used top ten descriptor models to check their performance and achieved an overall correlation 0.518 and only six protein kinases had correlation more than 0.5 and only six had more than 0.65 but no model generated in case of six protein kinases.

We also wanted to check the role of negatively correlated descriptors on dissociation constant and achieved an overall correlation 0.491 but no model developed for five kinase proteins and six had correlation more than 0.5 including

Table 2. Performance of SVM Based QSAR Models Using Top 5, 10 and 15 Descriptors

Protein	Top 5	Top 10	Top 15	Molinspiration
AAK1	0.514	NM	0.45	0.42
ABL1	0.449	0.714	0.43	0.472
ABL1E255K	0.851	0.623	0.473	0.48
ABL1H296P	0.851	0.623	0.473	0.48
ABL2	0.737	0.621	0.473	0.44
EPHA5	0.428	0.445	0.349	0.464
EPHB1	0.462	0.195	0.578	0.577
JNK2	0.669	0.601	0.627	0.064
LCK	0.653	0.771	0.727	0.621
MAP4K5	0.658	0.515	0.41	0.129
PDGFR	0.583	0.268	0.363	0.036
RIPK2	0.736	0.249	0.25	0.557
SLK	0.643	0.363	0.293	0.345
SRC	0.77	0.541	0.54	0.395
STK10	0.792	0.511	0.396	0.373
TNIK	0.563	0.274	0.235	0.514
Average	**0.647**	**0.488**	**0.442**	**0.398**

Table 3. Performance of Kinase Specific Models Developed Using Top Positive and Negative Correlated Descriptors

Protein	Top 5 +ve	Top 5 -ve	Top 10 +ve	Top 10 mixed
AAK1	0.55	0.8	NM	0.22
ABL1	0.52	0.38	0.16	NM
ABL1E255K	0.530	0.480	0.435	0.585
ABL1M351T	0.72	0.56	0.66	0.48
BIKE	0.96	NM	0.76	NM
EGFR	0.81	0.29	0.91	0.91
EPHA5	0.91	0.95	-0.02	0.95
EPHA6	0.79	NM	0.68	0.22
EPHB1	0.89	NM	0.38	NM
JNK3	0.69	0.52	0.63	0.52
KIT	0.57	NM	NM	0.29
LCK	0.62	0.56	NM	0.51
P38Alpha	0.57	NM	NM	NM
RIPK2	0.63	0.65	0.8	0.57
SLK	0.59	0.3	0.22	NM
TNIK	0.88	0.07	NM	0.76
Average	**0.719**	**0.491**	**0.518**	**0.543**

three that had more than 0.65. Similarly in case of the model developed by using top five positively and top five negatively correlated descriptors we achieved overall correlation 0.543 and 6 kinase had no model (Table **3**).

DISCUSSION

First genomes sequenced in 1996, at present full genomes of human and several pathogenic organisms have been sequenced. Attempts have been made to combine genomics, proteomics, bioinformatics, combinatorial chemistry, QSAR to screen new drug compounds in less time and money which may developed as an effective drug [16, 17]. In the last decades a large number of bioinformatic tools have been developed for the annotation of protein and its function prediction in order to identify the novel potential drug targets. According to literature main drug targets belong to GPCRs, Kinase protein, ion-channels, nuclear receptors, hormones and DNA [18].

Kinase proteins are one of the important proteins in drug designing so designing kinase inhibitors is always an area of interest for pharmaceutical companies [19]. Despite tremendous growth in the area of kinase inhibitors research from all over the world, cross-reactivity and specificity still remains a major challenge. Because most of the kinase inhibitor molecules bind to the active site of kinase molecules and this site is highly conserved in kinase family**.** There is a need to address the issue of cross-reactivity in order to develop the drug free from side effects.

The general artificial intelligence (AI) based techniques such as SVM and neural network are elegant approaches for the extraction of complex pattern from chemical structure descriptors. These techniques are highly successful in protein annotation [20, 21], subcellular localization [22, 23], structure prediction [24, 25], interacting residue prediction [26-29], antibacterial peptide [30], QSAR studies [31, 32], microarray data analysis [33,34], cancer classification [35], drug designing [36,37] and chemical toxicity studies [38]. Several studies shows that SVM over perform on other AI techniques in drug designing [39]. We used SVM for QSAR model development for K_d prediction and achieved overall correlation 0.719. Our SVM based QSAR model worked very well for 9 protein kinases. Result also suggests that the performance of model decreases when negatively correlated descriptors are included or feature dimension increases. SVM parameters for all these models are available in tables which are freely available at DMKPred web server.

CONCLUSIONS

In the present study, for the first time an attempt has been made to predict specificity and cross-reactivity of kinase inhibitors using *in-silico* approach. We used the experimental data of Fabin *et al*. for developing our models. First we developed a general method for predicting kinase inhibitors against kinase family; these models became helpful for screening of kinase inhibitors. As shown in Table **2** we achieved reasonable correlation for kinase inhibitor prediction. Later on we use kinase specific model which may be helpful in prediction of cross-reactivity and specificity of kinase inhibitors. These models also gave reasonable correlation (Table **3**) between actual and predicted K_d value.

In conclusion, our QSAR based SVM models directly predict the dissociation constant of chemical kinase inhibitors against protein kinases. This approach should provide a valuable result in determining dissociation constant of

chemical molecules against kinase proteins without taking time and using an experimental setup.

DESCRIPTION OF WEB SERVER

Based on our study, we have developed a web server, DMKPred, which allow the users to predict the dissociation constant of chemical molecules with 16 protein kinases. DMKPred is freely available at http://www.imtech.res.in/raghava/dmkpred/. The Common Gateway Interface (CGI) script for DMKPred is written using PERL. This server is installed on a Sun Server under a Solaris environment and launched using Apache. Required molecular descriptors name for each protein kinases are given in the submission page (Fig. **1**). User can enter the molecular descriptors of the chemical molecules for the prediction of dissociation constant with each protein kinase. Results after prediction, will be displayed on result in a tabular form (Fig. **2**).

AUTHOR'S CONTRIBUTIONS

This study was designed by G. P. S. Raghava, performed by Nitish Kumar Mishra and analyzed and final manuscript approved by both authors.

ACKNOWLEDGEMENTS

The authors are gratefully acknowledging the financial support provided by the Council of Science and Industrial (CSIR) and Department of Biotechnology (DBT), Government of India.

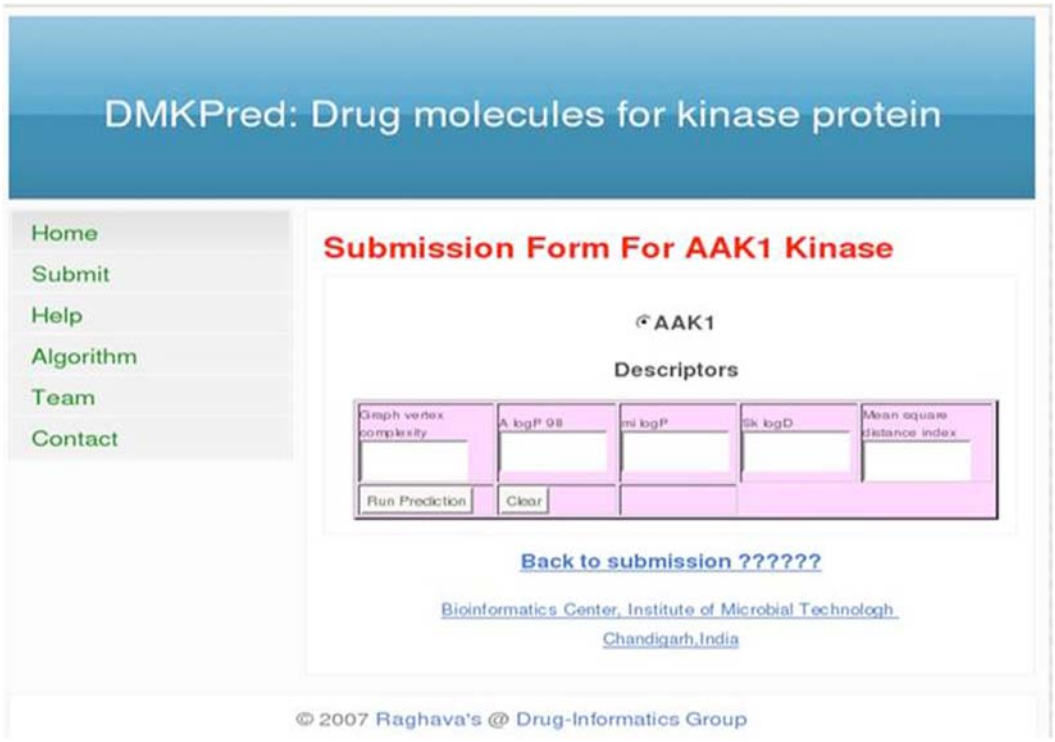

Fig. (1). A snapshot of submission page of DMKPred.

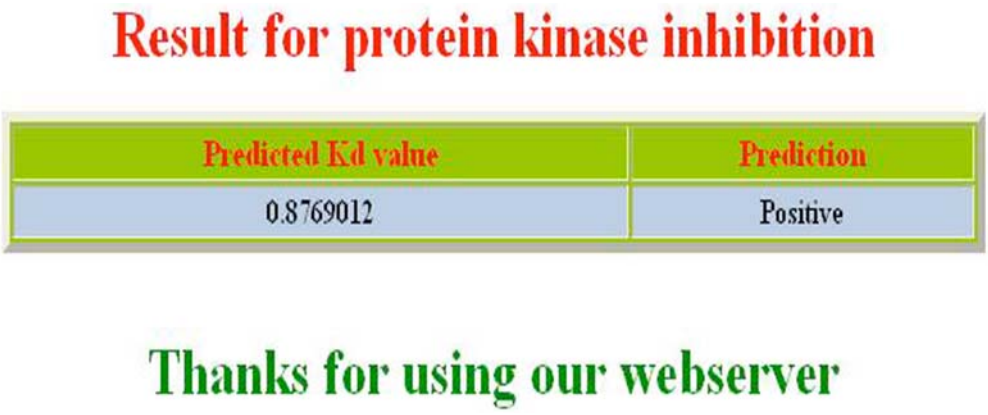

Fig. (2). Result of prediction for given chemical compound.

REFERENCES

[1] Cohen, P. Protein kinases the major drug targets of the twenty first century? *Nat. Rev. Drug. Discov.*, **2002**, *2*, 296-313.

[2] Huse, M.; Kuriyan, J. The conformational plasticity of protein kinases. *Cell*, **2002**, *109*, 275-282.

[3] Hopkins AL Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.*, **2006**, 16, 127-136.

[4] Veith, M. Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochem. Biophys. Acta*, **2004**, *1697*, 243-257.

[5] Kumar, S.; Lee, J.C. P38 MAP kinase: key signaling molecules as therapeutic targets for inflammatory diseases. *Nat. Rev. Drug. Discov.*, **2003**, *2*, 717-726.

[6] Dancey, J.; Sausville, E.A. Issue and progress with protein kinase inhibitors for cancer treatment. *Nat. Rev. Drug. Discov.*, **2003**, *2*, 296-313.

[7] Garberken The second wave in kinase cancer drugs. *Nat. Biotechnol.*, **2006**, *24*, 127-130.

[8] Bain, J.; Cohen, P. The specificities of protein kinase inhibitors: an update. *Biochemistry* **2003**, *371*, 199-204.

[9] Davies, S.P.; Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem. J.*, **2000**, *351*, 95-105.

[10] Fabin, M.A. A small molecule kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.*, **2005**, *23*, 329-336.

[11] Manning, G.; Whyte, D.B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science*, **2002**, *298*, 1912-1934.

[12] Gasteiger, J. Calculation of structure descriptors. In *Chemoinformatics A text Book*. Wiley-VCH; 2003: 401-431.

[13] Molinspiration [http://www.molinspiration.org]

[14] PreADMET [http://www.preadmet.bmdrc.org/preadmet/index.html]

[15] Joachims, T. Making large scale SVM learning practical. In *Advanced in kernel methods- support vector learning (eds. Scholkopf, B. et al.)*. MIT Press Cambridge USA; 1999: 42-56.

[16] Perveen, S.; Yasmeen, A.; Khan, M.A.; Dar, A.; Jafri, R.; Ahmed, A. Structure Activity Relationship of Organic Alcohol and Esters for Antidepressant Like Activity. *Letters in Drug Design & Discovery*, **2010**, *7*, 14-17.

[17] Kallech-Ziri, O.; Luis, J.; Faljoun, Z.; Sabatier, J.M.; Lehmann, M.; Ayeb, M.E1.; Marrakchi. N.; Loret, E. Structure Function Relationships of KTS Disintegrins and Design of Antiangiogenic Drugs. *Letters in Drug Design & Discovery*, **2010**, *7*, 36-40.

[18] Imming, P.; Sinning, C.; Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug. Discov.*, **2006**, *5*, 821-834.

[19] Gray, N.S. Exploiting chemical libraries, structure and genomics in the search of kinase inhibitors. *Science*, **2003**, *281*, 533-538.

[20] Mishra, N.K.; Kumar, M.; Raghava, G.P.S. Support vector machine based prediction of glutathione s-transferase proteins. *Protein Pept. Lett.*, **2007**, *14*, 575-80.

[21] Kumar, M.; Verma, R.; Raghava, G.P.S. Prediction of mitochondrial proteins using support vector machine and hidden markov model. *J. Biol. Chem.*, **2006**, *281*, 5357-5363.

[22] Garg, A.; Raghava, G.P.S. ESLpred2: improved method for predicting subcellular localization of eukaryotic proteins. *BMC Bioinformatics*, **2008**, *9*, 503.

[23] Bhasin, M.; Raghava, G.P.S. PSLPred: prediction of subcellular localization of bacterial proteins. *Bioinformatics*, **2005**, *21*, 2522-2524.

[24] Kaur, H.; Raghava, G.P.S. Prediction of alpha-turns in proteins using PSI-BLAST profiles and secondary structure information. *Proteins*, **2004**, *55*, 83-90.

[25] Kaur, H.; Raghava, G.P.S. A neural network method for prediction of beta-turn types in proteins using evolutionary information. *Bioinformatics*, **2004**, *20*, 2751-2758.

[26] Mishra, N.K.; Raghava, G.P.S. Prediction of FAD interacting residues in a protein from its primary sequence using evolutionary information. *BMC Bioinformatics*, **2010**, *11*, S48.

[27] Chauhan, J.S.; Mishra, N.K.; Raghava, G.P.S. Identification of ATP binding residues of a protein from its primary sequence. *BMC Bioinformatics*, **2009**, *10*, 434.

[28] Chauhan, J.S.; Mishra, N.K.; Raghava, G.P.S. Prediction of GTP interacting residues, dipeptides and tripeptides in a protein from its evolutionary information. *BMC Bioinformatics*, **2010**, *11*, 301.

[29] Ansari, H.R.; Raghava, G.P.S. Identification of NAD interacting residues in proteins. *BMC Bioinformatics*, **2010**, *11*, 160.

[30] Lata. S.; Mishra, N.K.; Raghava, G.P.S. AntiBP2: improved version of antibacterial peptide prediction. *BMC Bioinformatics*, **2010**, *11*, S19.

[31] Mishra, N.K.; Agarwal, S.; Raghava, G.P.S. Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule. *BMC Pharmavol.*, **2010**, *10*, 8.

[32] Garg, A.; Tewari, R.; Raghava, G.P.S. KiDoQ: using docking based energy to develop ligand based model for predicting antibacterials. *BMC Bioinformatics*, **2010**, *11*, 125.

[33] Furey, T.S.; Haussler, D. Support vector machine classification and validation of cancer tissue sample using microarray express data. *Bioinformatics*, **2002**, *16*, 906-914.

[34] Brown, M.P.S.; Lin, D. Knowledge-based analysis of microarray gene expression data by using support vector machine. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 262-267.

[35] Trotter, M.W.B.; Holden, S.B. Support vector machines for ADME property classification. *QSAR Comb. Chem.*, **2003**, *22*, 533-548.

[36] Burdige, R. Drug design by machine learning: support vector machine for pharmaceutical data analysis. *Comput. Chem.*, **2001**, *26*, 5-14.

[37] Totter, M.B.W.; Halden, S.B. Support vector machine in combinatorial chemistry. *Measurement and Control*, **2003**, *34*, 235-239.

[38] Zernov, V.V. Drug discovery using support vector machine. The case studies of drug-likeness, agrochemical-likeness and enzyme inhibition prediction. *J. Chem. Info. Comp. Model.*, **2003**, *43*, 2048-2056.

[39] Yap, C.W.; Chen, Y.Z. Prediction of tarsade-causing potential of drugs by support vector machine approach. *Toxicol. Sci.*, **2004**, *79*, 170-177.