2. Spinks, J. W. T. and Woods, R. J., In *An Introduction to Radiation Chemistry*, John Wiley, New York, 1990, pp. 257–258.
3. Dimitrijevic, N. M., Henglein, A. and Meisel, D., Charge separation across the silica/water interface. *J. Phys. Chem. B*, 1999, **103**, 7073–7076.
4. Seino, S., Yamamoto, T. A., Fujimoto, R., Hashimoto, K., Katsura, M., Okuda, S. and Okitsu, K., Enhancement of hydrogen yield form water dispersing nanoparticles irradiated with gamma ray. *J. Nucl. Sci. Technol.*, 2001, **38**(8), 633–636.
5. Wada, Y., Kawaguchi, K. and Myouchin, M., Decomposition of water and production of H_2 using semiconductor–photocatalytic effect induced by gamma ray from high radioactive waste. *Prog. Nucl. Energy*, 1995, **29**, 251–256.
6. Yoshida, T., Sawasaki, T., Chen, A. Y. K. and Tanabe, T., Enhanced gamma-ray energy conversion in water vessel – Part-II: hydrogen production by water radiolysis. *Nucl. Sci. Eng.*, 2005, **150**, 357–361.
7. Fujishima, A. and Honda, K., Electrochemical photolysis of water at semiconductor electrode. *Nature*, 1972, **238**, 37–38.
8. Allen, A. O., In *The Radiation Chemistry of Water and Aqueous Solutions*, 32, D. Van. Nostrand Company, Inc., New York, 1961, pp. 32–48.
9. Kumbhar, A. G., Belapurkar, A. D., Venkateswaran, G. and Kishore, K., Impact of different metal turbidities on radiolytic hydrogen generation in nuclear power plant. *Power Plant Chem.*, 2005, **7**(11), 674–679.
10. Buxton, G. V., Greenstock, C. L., Helman, W. P. and Ross, A. B. J., Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ($^{\bullet}OH/O^{-}$) in aqueous solutions. *J. Phys. Chem. Ref. Data*, 1988, **17**, 513–886.
11. Hanglien, A., Reactions of organic free radicals at colloidal silver in aqueous solution. Electron pool effect and water decomposition. *J. Phys. Chem.*, 1979, **83**(17), 2209–2216.
12. Sowden, R. G., Radiolytic problems in nuclear reactors. *J. Nucl. Mater.*, 1963, **8**(1), 81–101.
13. Feradini, C. and Jay-grein, J. P., Quelqes aspects actuels de la radiolyse du methanol liquide: une revue. *Radiat. Phys. Chem.*, 1996, **48**(4), 473–480.
14. Jung, J., Jeong, H. S., Chung, H. H., Lee, M. J., Jin, J. H. and Park, K. B., Radiolytic H_2 production with gamma-irradiation and TiO_2 catalyst. *J. Radioanal. Nucl. Chem.*, 2003, **258**(3), 543–546.
15. Maeda, Y., Kawana, Y., Kawamura, K., Hayami, S., Sugihara, S. and Okai, T., Hydrogen gas evolution from water included in silica gel cavity and on metal oxides with $\dot{\gamma}$-ray irradiation. *J. Nucl. Radiochem.*, 2005, **6**(2), 131–134.

ACKNOWLEDGEMENTS. We thank to Dr T. Mukherjee, Director, Chemistry Group, BARC and Dr S. V. Narasimhan, Head, Water and Steam Chemistry Division, BARC for their interest and encouragement in the study. We also thank Shri B. K. Gokhale for preparing the schematic of test loop for radiolytic H_2 production.

Received 15 September 2010; revised accepted 3 January 2011

MetPred: a web server for classification, identification and prediction of metalloproteases

Kumud Pant[1,*], Neeru Adlakha[2] and Alok Mittal[3]

[1]Department of Bioinformatics, and [3]Department of Chemistry, Maulana Azad National Institute of Technology, Bhopal 462 051, India
[2]Department of Applied Mathematics, Sardar Vallabhbhai National Institute of Technology, Surat 395 007, India

Metalloproteinases (MP) are a family of proteases which play a major role not only in normal physiological processes like normal turnover of extracellular matrix macromolecules but also in diseased conditions such as arthritis, cancer, brain disorders, skeletal dysplasias, coronary artery and heart disease. Members of the matrix metalloproteinases (MMP) family include MMPs, MT-MMP, ADAMs and ADAMTS. Majority of these proteins are drug targets with many of the MP drugs currently in trial phase. In the past, machine learning has been used to classify other proteins but no attempt has been made for classification of MPs. Realizing their importance, an attempt has been made to develop a support vector machine model to predict, classify and correlate major subclasses of MPs with their amino acid composition. The method was trained and tested on 229 proteins of MPs. The method discriminated MPs from other enzymes with Matthew's correlation coefficient of 1.00 and 100% accuracy. In classifying different subclasses of MPs with amino acid composition, an overall average accuracy of 98% was achieved. The performance of the method was evaluated using five-fold cross-validation. For understanding them in a better way, a web server MetPred has been developed for predicting MPs from its amino acid sequence at www.bifmanit.org/MetPred/.

Keywords: Metalloproteinases, support vector machine, web server.

METALLOPROTEINASES or metalloproteases (MPs) constitute a family of enzymes from the group of proteases, classified by the nature of the most prominent functional group in their active site. These are proteolytic enzymes whose catalytic mechanism involves a metal. Most MPs are zinc-dependent, some use cobalt. The metal ion is coordinated to the protein via three histadine imidazole ligands. The fourth coordination position is taken up by a labile water molecule[1]. There are two subgroups of MPs: (i) exopeptidases: metalloexopeptidases (EC number: 3.4.17) and (ii) endopeptidases: metalloendopeptidases (3.4.24). Well-known metalloendopeptidases include: (i) matrix metalloproteinase (MMPs), (ii) a disintegrin and metalloproteinase (ADAM) proteins and (iii) a disintegrin and metalloproteinase with thrombospondin motifs

*For correspondence. (e-mail: pant.kumud@gmail.com)

(ADAMTS). All the three endopeptidases, viz. MMPs, ADAMs and ADAMTS have very important roles to play in the normal physiological activity of the body. MMPs belong to a larger family of proteases known as the metzincin superfamily. Collectively they are capable of degrading all kinds of extracellular matrix proteins, but can also process a number of bioactive molecules. They are known to be involved in the cleavage of cell surface receptors, the release of apoptotic ligands (such as the FAS ligand), and chemokine/cytokine in/activation[2]. MMPs are also thought to play a major role on cell behaviours such as cell proliferation, migration (adhesion/dispersion), differentiation, angiogenesis, apoptosis and host defence. MMPs share a common domain structure. The three common domains are the pro-peptide, the catalytic domain and the haemopexin-like C-terminal domain which is linked to the catalytic domain by a flexible hinge region. On the basis domain information and the catalytic activity involved in the working of MMPs, they have been classified into various subclasses such as collagenases, gelatinases, stromelysin and membrane-type MMPs[3–7]. ADAM proteins are members of the same superfamily as MMPs, viz. the metzincins, named for their zinc-binding domains and their structurally important C-terminal conserved methionine residue. As the name suggests, ADAM proteins are cell surface proteins that possess both an adhesion domain and a protease domain[8]. ADAM are a fascinating family of transmembrane and secreted proteins with important roles in regulating the cell phenotype via their effects on cell adhesion, migration, proteolysis and signalling. They are involved in diverse processes such as development, cell–cell interactions and protein ectodomain shedding[9]. There are more than 35 members of the ADAM family of proteins; the precise function of many of the ADAM family members is unknown, but some, such as ADAM17 have known biological functions. An additional class of ADAM-related proteins is known as the ADAMTS proteins. ADAMTS proteins are structurally homologous to ADAM proteins, but they contain at least one C-terminal thrombospondin type 1 (TSP1) repeat and are secreted rather than membrane bound[10,11]. ADAMTS1 and ADAMTS8 are inhibitors of angiogenesis, and others, such as ADAMTS5, cleave extracellular proteoglycans such as aggrecan. ADAM proteins are involved in diverse processes including sperm–egg binding and fusion, myoblast fusion, protein ectodomain processing, shedding cytokine receptors and adhesion proteins, whereas ADAMTS have a role in post-translational modification of procollagen as well as in the pathologic destruction of cartilage. Recent study showed that activation of mitogen-activated protein kinase/extracellular signal-regulated kinase (MEK/ERK) and phosphatidylinositol 3-inase (PI3K)/Akt pathways by fibronectin requires integrin V-mediated ADAM activity in hepatocellular carcinoma, indicating ADAM protein may be a target for anti-cancer drug development[12]. In addition, ADAM protease families have been shown to degrade aggrecan, one of the most important structural components of joint cartilage *in vivo*[13]. ADAMs harbouring a sheddase function recently became candidates for causing several diseases, like distinct forms of the Alzheimer disease. Thus, ADAMs can be a target for preclinical drug development[14].

Similarly, the study on MMPs which began with the discovery of collagenase in tadpole tail in 1962 has now become a vast research field. This family of enzymes plays a central role in the pathological degradation and physiological turnover of the extracellular matrix (ECM) and other proteins. MMPs, ADAMs and ADAMTS share the substrate specificity and their expression pattern under certain conditions, but their mutual roles are not well understood[15].

ADAMs recently have been found to play an important role in various types of cancer[16]. Similarly, MMPs are found to be a key player in skin ageing and many physiological disorders. In animal body they have a role in cancer, neurological disorders, rheumatoid, osteoarthritis, etc.[17]. Synthetic inhibitors have recently been developed for MMPs, so as to control cartilage matrix degradation by them in arthritis and osteoporosis[18]. Similarly, ADAMs have recently also been implicated in diseases such as various types of cancers and ADAM-dependent proteolytic attach could be a valuable therapeutic target for Alzheimer's disease and other neurodegenerative disorders characterized by the deposition of insoluble proteinaceous aggregates[16,19,20].

Despite emerging implications for MMPs, ADAMs and ADAMTS in disease progression, the mechanisms that lead to activation of specific MMPs, ADAMs and ADAMTS and their actions in various diseases are still incompletely understood. Some attempts for the understanding of structure and function of MMPs, ADAMs and ADAMTS family of proteins have been made in the past five years. These results have demonstrated the importance of these proteins in diverse biological processes. Studies have also raised many interesting questions that remain to be answered. Questions concerning substrate specificities of these, the physiological regulators that activate or inhibit these proteases the regulation of the protease, adhesion and signalling activities of these in response to developmental, physiological and pathological stimuli. etc. remain unanswered[21].

Identification of novel type of cell surface proteases and their cognate substrate is the major focus of pharmaceutical companies. Hence, highly accurate identification of protease types will solve the problem of efficacy and side effects of various drugs. Currently, efforts are underway to develop new therapeutic agents and elucidation of the metabolic pathway associated with diseases. Moreover, the mere understanding of different types of ADAM, ADAMTS and MMPs and their substrate-binding properties will assist in finding novel drug target

with minimum side effects. The experimental attempts are reported in the literature for functional classification of MMPs, ADAMs and ADAMTS using the structure of their catalytic sites[22]. But no computational technique is available in the literature for classification based on other parameters like amino acid composition, dipeptide composition and physiochemical properties. As their experimental identifications are labour- and cost-intensive, computational biology can provide a better alternative to develop a method for classifying the different enzymes.

An attempt has been made in this communication to develop a computational approach for predicting and classifying three types of cell surface proteases: ADAMs, ADAMTS and MMPs. In the first step, a binary classification method has been adopted where MPs can be discriminated from the rest of the proteins. In the second step a multiclass classification approach has been used where the three MPs can be distinguished from each other.

It has been shown in the past that the support vector machine (SVM) is an elegant technique for the classification of biological data[16,19,20,23–25]. Here, a SVM model has been developed for amino acid composition-based prediction, identification and classification of cell surface proteases ADAMs, ADAMTS and MMPs.

This communication is a step in the direction where machine learning and computational biology techniques can be used to complement existing wet-lab techniques. To the best of our knowledge, there is no web server that allows recognition and classification of MPs. On the basis of our study, an online web tool 'MetPred' has been made available at http://www.bifmanit.org/MetPred/.

To achieve our goal and develop our methodology, we obtained the dataset from Swissprot/Uniprot databank of Expasy server[26]. The following three datasets were used. The dataset were obtained after selecting all the instances present in the databse and removing the fragments and redundant sequences.

Dataset 1 consisted of all MMP proteins, ADAM family members and ADAMTS. The total instances were 97 for MMPs, 90 for ADAM subclass, 42 for ADAMTS. The final dataset consisted of 229 sequences belonging to MMP and ADAM, and ADAMTS subclass of MPs family.

Dataset 2 was used to validate our methodology; 229 globular proteins belonging to various enzyme classes other than hydrolases to which MPs belong were taken into consideration. They were treated as negative instances.

For the training dataset, we consider 458 sequences belonging to both MPs and non-MPs whereas for testing, three datasets were prepared. First test dataset comprised 97 MMPs obtained from dataset 1 and an equal number of non-MMPs obtained from dataset 2. The second test dataset comprised 90 ADAMs belonging to dataset 1 and an equal number of non-ADAMs obtained from dataset 2 and the third dataset consisted of 42 ADAMTS and an equal number of non-ADAMTS taken from dataset 2.

SVM is a supervised machine learning method which is based on the statistical learning theory[27,28]. When used as a binary classifier, a SVM will construct a hyperplane, which acts as the decision surface between the two classes. This is achieved by maximizing the margin of separation between the hyperplane and those points nearest to it. SVMs were implemented using a freely downloadable software, libSVM[29]. In this software, there is a facility to define parameters and choose among various inbuilt kernels. They can be radial basis function (RBF) or a polynomial kernel (of given degree), linear or sigmoid.

Simulations were preformed using libSVM version 2.89 (a freely available software package)[29]. For our study RBF kernel was found to be the best. SVM training was carried out by the optimization of the value of the regularization parameter and the value of RBF kernel parameter.

Previously, the amino acid composition parameter has been used for predicting the subcellular localization of proteins[9]. The amino acid composition is the fraction of each amino acid type within a protein. The fractions of all 20 natural amino acids were calculated by using eq. (1).

$$\text{Amino acid composition} = \frac{\begin{pmatrix}\text{Total number of}\\ \text{amino acid } (i)\end{pmatrix}}{\begin{pmatrix}\text{Total number of amino}\\ \text{acids in a protein}\end{pmatrix}} \quad (1)$$

where i can be any amino acid.

The performance of our classifier was judged by a 10-fold cross-validation. The libSVM provides a parameter

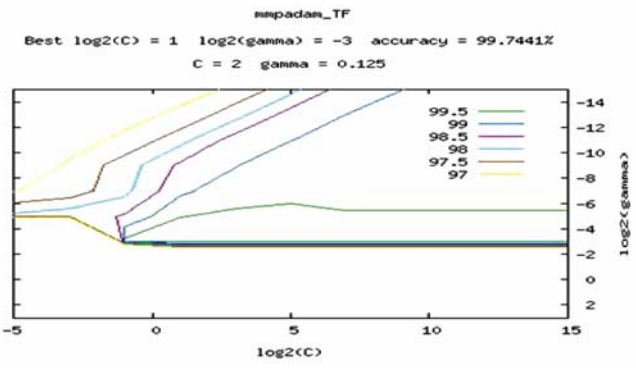

Figure 1. Coarse grid search on $C = 2^{-5}, 2^{-4}, \ldots, 2^{10}$ and gamma = $2^{5}, 2^{4}, \ldots, 2^{-10}$.

Table 1. Classification details for the three subclasses

Protein sub-family	MCC	Accuracy (%)	Sensitivity	Recall (%)	Precision (%)
MMPs	0.97	98.98	0.99	98.96	98.96
ADAMs	0.97	98.88	0.98	98.87	98.87
ADAMTS	0.97	98.88	0.98	98.87	98.87

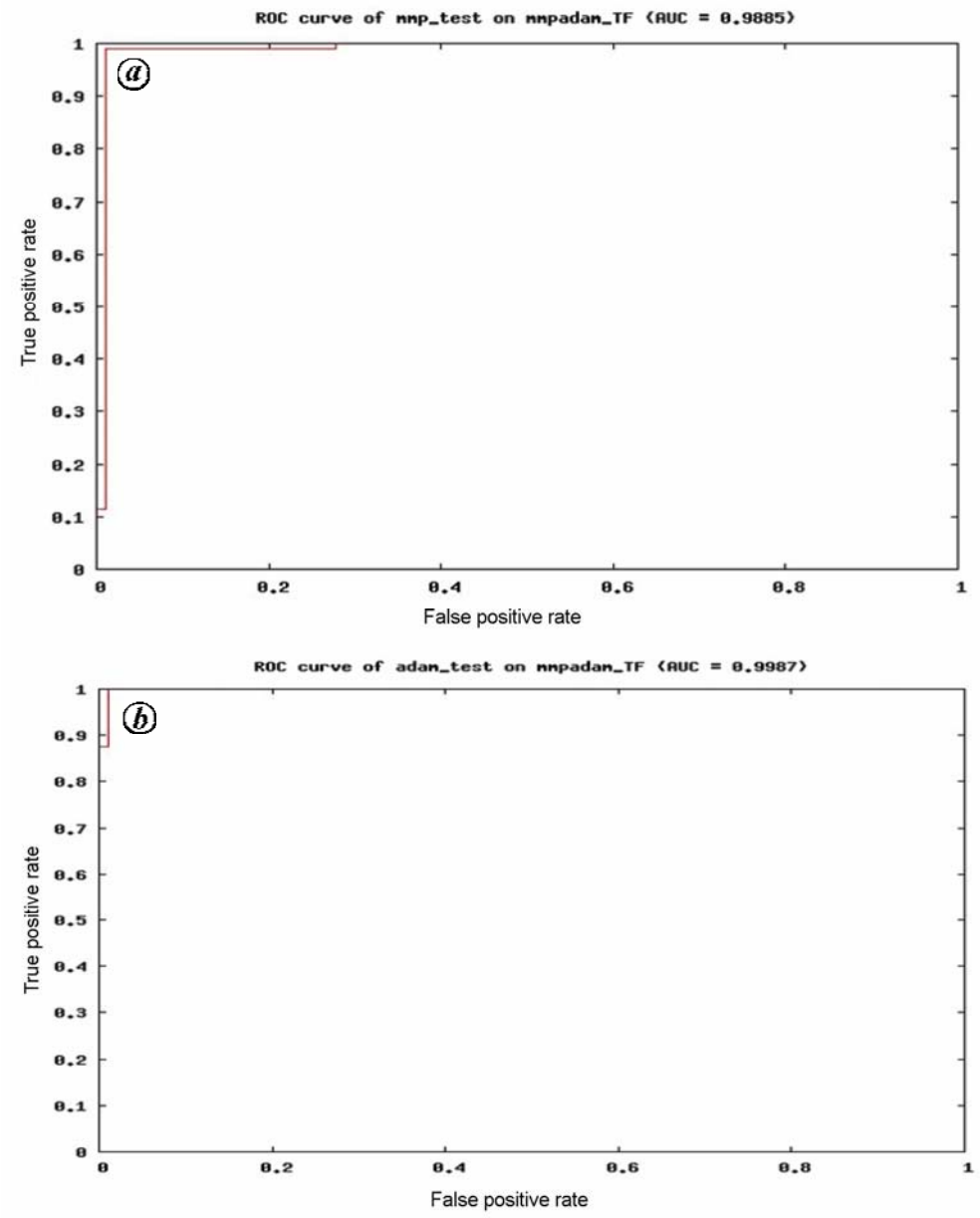

Figure 2. ROC curve for both MMP (***a***) and ADAM (***b***) proteases depicting FP and TP.

selection tool using the RBF kernel: cross-validation via grid search. A grid search was performed on *C* and gamma using an inbuilt module of libSVM tools as shown in Figure 1. Here pairs of *C* and gamma are tried and the one with the best cross-validation accuracy is picked. On using the values of $C = 2$ and gamma = 0.125 obtained through grid search, an accuracy of 98.87% was obtained.

True positives (TP) and true negatives (TN) were identified as the positive and negative samples respectively. False positives (FP) were negative samples identified as positive. False negatives (FN) were positive samples identified as negative. The prediction performance was tested with sensitivity (TP/(TP + FN)), specificity (TN/(TN + FP)), overall accuracy (Q2), and the Matthews correlation coefficient (MCC). The accuracy and MCC for each subfamily of MPs were calculated as described by Hua and Sun[30] and are shown below in eqs (2) and (3).

$$\text{Accuracy}\,(x) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{MCC} = \frac{(\text{TP})(\text{TN}) - (\text{FP})(\text{FN})}{\sqrt{(\text{CP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}. \tag{3}$$

All the three formulae of MPs, viz. MMP, ADAMs and ADAMTS have been implicated in various diseases and one key players in many protein degradation processes.

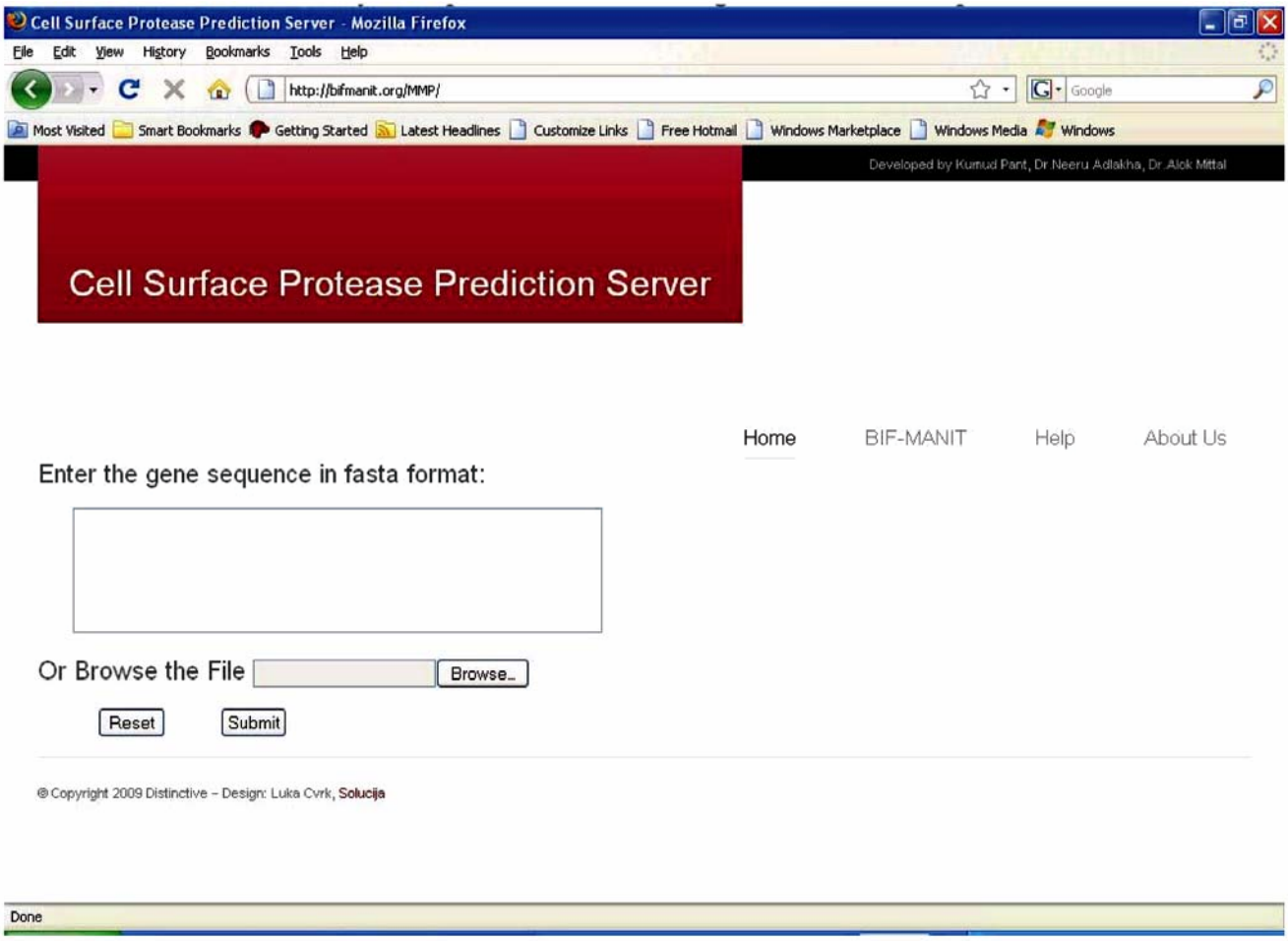

Figure 3. Snapshot of MetPred (cell surface protease prediction server).

Realizing their implications, we have chosen these three families of MPs for our study.

The results of classifying MMPs, ADAMs and ADAMTS using amino acid composition are given in Table 1. The results obtained here will be helpful in differentiating among the three MPs. A new protein discovered can be shown to either belong to ADAMs, ADAMTS or MMP sub-family of proteins.

Our results clearly highlight the importance of amino acid composition in differentiating between these families. This model can also be an important tool to understand the differences between these and hence a step towards assisting various wet-lab techniques in devising novel drugs and therapeutic agents against these two. The correlation of MMPs, ADAMs and ADAMTS with their amino acid composition explored here can be useful to obtain better insight into these proteins. Their molecular and physiological roles along with the substrate affinity can also be correlated with amino acid composition.

The overall accuracy and MCC of the amino acid composition-based classifier for classifying the three subfamilies of MPs was 98% and 0.97 respectively. It proved that MPs can be correlated with amino acid composition and can be easily distinguished on this basis.

The receiver operating characteristics (ROC) score was usually used as the primary measure of the machine learning method performance and provided an overview of the possible cut-off levels in the test performance. The ROC curves for two of the subfamilies, i.e. MMPs and ADAMs are depicted in Figure 2 *a* and *b* which shows that majority of instances fall in the TP range.

MetPred is freely available at www.bifmanit.org/MetPred/. MetPred server is installed on a Windows Server environment. The user can provide the input sequence by cut–paste or directly uploading sequence file from disk. The server accepts the sequence in standard FASTA format. A snapshot sequence submission page of server is shown in Figure 3. The user can predict the type of MP based on amino acid composition. On submission the server will give results in user-friendly format.

With amino acid composition as evaluation parameter of MPs, an overall average accuracy of 98% was obtained in classifying various subclasses. MetPred developed at www.bifmanit.org/MetPred/ can be an efficient and time-

saving tool. These kinds of web servers can be an economical and time-saving approach for annotation of piled-up genomic data. They can be used to effectively complement the existing wet-lab techniques.

This model can be used to analyse other enzymes such as entire proteomics data. Such prediction systems can be useful in understanding these proteases in a better way, i.e. a novel method for classifying MMPs, ADAMs and ADAMTS is presented. This method will nicely complement the existing wet-lab methods. It will assist in assigning a correct class to these proteins or classify them as any of the three subclasses. The prediction method presented here may be useful for the annotation of the piled-up proteomic data.

We await the discovery of more of these proteins in the future so that accuracy of the prediction model can be increased further and a server developed for public use.

1. Remacle, A. G., Rozanov, D. V., Fugere, M., Day, R. and Strongin, A Y., Furin regulates the intracellular activation and the uptake rate of cell surface-associated MT1-MMP. *Oncogene*, 2006, **25**(41), 5648–5655.
2. Van Lint, P. and Libert, C., Chemokine and cytokine processing by matrix metalloproteinases and its effect on leukocyte migration and inflammation. *J. Leukoc. Biol.*, 2007, **82**(6), 1375–1381.
3. Pei, D., Kang, T. and Qi, H., Cysteine array matrix metalloproteinase (CA-MMP)/MMP-23 is a type II transmembrane matrix metalloproteinase regulated by a single cleavage for both secretion and activation. *J. Biol. Chem*., 2000, **275**(43), 33988–33997.
4. Trexler, M., Briknarová, K., Gehrmann, M., Llinás, M. and Patthy, L., Peptide ligands for the fibronectin type II modules of matrix metalloproteinase 2 (MMP-2). *J. Biol. Chem*., 2003, **278**(14), 12241–12246.
5. Browner, M. F., Smith, W. W. and Castelhano, A. L., Matrilysin-inhibitor complexes: common themes among metalloproteases. *Biochemistry*, 1995, **34**(20), 6602–6610.
6. Kester, W. R. and Matthews, B. W., Crystallographic study of the binding of dipeptide inhibitors to thermolysin: implications for the mechanism of catalysis. *Biochemistry*, 1977, **16**(11), 2506–2516.
7. Manzetti, S., McCulloch, D. R., Herington, A. C. and van der Spoel, D., Modeling of enzyme-substrate complexes for the metalloproteases MMP-3, ADAM-9 and ADAM-10. *J. Comput. Aided Mol. Des*., 2003, **17**(9), 551–565.
8. Wolfsberg, T. G. *et al.*, ADAM, a novel family of membrane proteins containing A disintegrin and metalloprotease domain: multipotential functions in cell–cell and cell–matrix interactions. *J. Cell Biol*., 1995, **131**, 275–278.
9. Chou, K. C. and Cai, Y. D., Using functional domain composition and support vector machines for prediction of protein subcellular location. *J. Biol. Chem.*, 2002, **277**, 45765–45769.
10. Minireviews 'ADAMs', R&D Systems' 2001 Catalog.
11. Stocker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F. X., McKay, D. B. and Bode, W., The metzincins – topological and sequential relations between the astacins, ADAMalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.*, 1995, **4**, 823–840.
12. Matsuo, M., Sakurai, H., Ueno, Y., Ohtani, O. and Saiki, I., Activation of MEK/ERK and PI3K/Akt pathways by fibronectin requires integrin α-mediated ADAM activity in hepatocellular carcinoma: A novel functional target for gefitinib. *Cancer Sci*., 2006, **97**, 155–162.
13. Peppard, J., Glickman, F., He, Y., Hu, S. I., Doughty, J. and Goldberg, R., Development of a high-throughput screening assay for inhibitor of aggrecan cleavage using luminescent oxygen channeling (AlphaScreen). *J. Biochem. Screen*, 2003, **8**, 149–156.
14. Lu, X., Lu, D., Scully, M. F. and Kakkar, V. V., Structure-activity relationship studies on ADAM protein–integrin interactions, cardiovascular and hematological agents in medicinal chemistry. 2007, **5**, 29–42.
15. Vincent, B., ADAM proteases: protective role in Alzheimer's and prion diseases. *Curr. Alzheimer Res.*, 2006, **10**, 165–174.
16. Bhasin, M. and Raghava, G. P., SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. *Bioinformatics*, 2004, **20**, 421–423.
17. Chang, C. and Werb, Z., The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. *Am. J. Pathol*., 2001, S37–S43.
18. Elliott, S. L. and Cawston, T., The clinical potential of matrix metalloproteinase inhibitors in the rheumatic disorders authors. *Drugs & Aging*, 2001, **18**(13), 87–99.
19. Blobel, C. P., Metalloprotease-disintegrins: links to cell adhesion and cleavage of TNF alpha and Notch. *Cell*, 1997, **90**, 589.
20. Cai, Y. D., Zhou, G. P., Jen, C. H., Lin, S. L. and Chou, K. C., Identify catalytic triads of serine hydrolases by support vector machines. *J. Theor. Biol*., 2004, **228**, 551–557.
21. Tenth Keio University International Symposium for Life Sciences and Medicine, 17–19 October 2001.
22. Hooper, N. M., Families of zinc metalloproteases. *FEBS Lett*., 1994, **354**, 1–6.
23. Cai, Y. D., Liu, X. J., Xu, X. B. and Chou, K. C., Support vector machines for prediction of protein subcellular location by incorporating quasi-sequence-order effect. *J. Cell. Biochem.*, 2002, **84**, 343–348.
24. Cai, Y. D., Pong-Wong, R., Feng, K., Jen, J. C. H. and Chou, K. C., Application of SVM to predict membrane protein types. *J. Theor. Biol*., 2004, **226**, 373–376.
25. Cai, Y. D., Zhou, G. P. and Chou, K. C., Support vector machines for predicting membrane protein types by using functional domain composition. *Biophys. J.*, 2003, **84**, 3257–3263.
26. ExPASy Proteomics Server, www.expasy.org/
27. Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, New York, 1995.
28. Wang, M., Yang, J., Liu, G. P., Xu, Z. J. and Chou, K. C., Weighted-support vector machines for predicting membrane protein types based on pseudo amino acid composition. *Protein Eng. Des*., 2004, **17**, 509–516.
29. Chang, C.-C. and Lin, C.-J., LIBSVM: a library for support vector machines, 2001; http://www.csie.ntu.edu.tw/~cjlin/libSVM
30. Hua, S. and Sun, Z., Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, 2001, **17**, 721–728.

ACKNOWLEDGEMENTS. We thank the Department of Biotechnology, Delhi, and MP Council of Science and Technology, Bhopal, for providing support in the form of bioinformatics infrastructure facility to carry out this work.

Received 4 June 2010; revised accepted 20 December 2010