34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

SVM-based Prediction of the Calpain Degradome using Bayes Feature Extraction

L.J.K. Wee and H.M. Low

Abstract— **Calpains belong to a family of calcium-dependent cysteine proteases which are implicated in a myriad of pathologies such as cancer and neurodegeneration. Despite extensive experimental studies on these proteases, our knowledge of the calpain degradome is still limited. Using a dataset of 341 unique, experimentally verified calpain cleavage sites, we conducted extensive sequence analyses and discovered novel residue propensities in the region flanking the cleavage site which could be modeled for prediction using machine learning algorithms. We have developed a series of computational models incorporating support vector machines and Bayes Feature Extraction for the prediction of calpain cleavage sites. The best models achieved A_{ROC} and accuracy scores ranging from 0.79 to 0.93 and 71% to 86% respectively when tested on independent test sets. We predicted calpain cleavage sites on proteins from the receptor tyrosine kinase family and discovered potential sites of cleavage at critical regulatory domains. The results suggest a novel role of calpains as a direct regulator of receptor tyrosine kinase activity in cell survival and cell death pathways.**

I. Introduction

Proteolysis - the specific and limited cleavage of proteins by enzymes called proteases - represents an important mechanism for post-translational control in all living organisms [1]. Calpains constitute an important family of intracellular, calcium-dependent, non-lysosomal cysteine proteases which exhibit specific proteolytic activities at neutral pH [2]. While the calpain system comprises of several protease members, much of the current work has been centered on two calpains; μ-calpain (or commonly called calpain 1) and *m*-calpain (calpain 2). Disruption to normal calpain-mediated proteolysis have been shown to affect several critical biological processes such as cytoskeleton organization and protein secretion, leading to abnormalities in cell structure, shape and cellular interactions. Significantly, calpain involvement in cytoskeletal protein proteolysis has been associated to neuronal diseases such as Huntington's and Alzheimer's disease [2,3]. Experimental and computational studies on the specificity of calpain cleavage of substrates and its intricacies would be highly valuable for deepening our biochemical knowledge of this important class of proteases. In this work, we constructed a comprehensive dataset of experimentally verified calpain cleavage sites for analysis and development of prediction models. We discovered that flanking sequences of cleavage sites possess distinctive residue composition and position-specific propensity patterns which could be helpful in discriminating the cleavage sites from non-cleavage sites *in silico*. We rigorously tested support vector machines (SVM) classifiers - employing simple binary encoding and the Bayes Feature Extraction schemes - for their effectiveness in predicting calpain cleavage sites. Results also showed that our best classifiers are comparable with, if not better than existing algorithms. We further predicted calpain cleavage sites on proteins from the receptor tyrosine kinase family and discovered potential sites of cleavage at critical regulatory domains, suggesting a novel role of calpains as a direct regulator of receptor tyrosine kinase activity in cell survival and cell death pathways.

L.J.K. Wee is with the Institute for Infocomm Research, Singapore, Singapore 138632 (phone: +65 6408 2182; fax: +65 6776 1378; e-mail: lawrence@bic.nus.edu.sg).

H.M. Low is with the Genome Institute of Singapore, Singapore, Singapore 138672 (e-mail: lowhm@gis.a-star.edu.sg).

II. Materials and Methods

A. Datasets

We extracted 341 unique, experimentally verified calpain cleavage site sequences from three publicly available sources (schematically shown in Figure 1; dataset is available from the authors upon request): the general protease substrates database CutDB [4], calpain prediction servers CaMPDB [5] and GPS-CCD [6]. A comprehensive search was also conducted on PUBMED for recently reported calpain cleavage events between 1st Jan 2009 through 31st Dec 2010, resulting in the identification of 20 previously unreported calpain cleavage site sequences. Based on the cleavage site sequences, we constructed five datasets containing sequences with the cleavage site flanked by four, eight, twelve, sixteen and twenty residues on both sides of the cleavage site (designated as P_4P_4', P_8P_8', $P_{12}P_{12}$' $P_{16}P_{16}$'and $P_{20}P_{20}$' datasets respectively; and described schematically in Figure 2). In addition, we further constructed two asymmetrical datasets to encapsulate the scissile bond and extension of four and twelve amino acids on either sides (P_4P_{12}' and $P_{12}P_4$'). An equal number of "non-cleavage sites" or negative examples were obtained by randomly extracting P_1 residues on the substrates while ensuring that no reported cleavage sites were inadvertently selected. A corresponding set of sequence segments of the aforementioned lengths and compositions were obtained as detailed earlier. The final datasets each comprised of 682 sequences (341 positives and 341 negatives; the complete dataset of sequences is available from authors upon request). For analysis, all 341 positives and 341 negatives from the $P_{12}P_{12}$' and P_8P_8' datasets were used. For SVM model development, all datasets were partitioned into training and test sets consisting of 291 positives/291 negatives and 50 positives/50 negatives respectively.

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

B. *Sequence analysis*

The relative position-specific residue propensity P_x was computed as the ratio of the frequency of occurrence of a particular amino acid in the cleavage sites pool to its frequency of occurrence in the non-cleavage sites pool at a specific position on the sequence. Using the $P_{20}P_{20}$' dataset, P_x scores were calculated for every amino acid at each of the twenty residue positions and visualized on heat maps. Additionally, we constructed a sequence logo representation of the positive sequences from the P_8P_8' dataset using WebLogo [7].

C. *Feature representation*

To encapsulate sequence information for SVM training and testing, input vectors were constructed using simple binary or bi-profile Bayes Features encoding. For simple binary encoding, each amino acid is represented by a 20-dimensional vector. For example, alanine was represented as [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1] and cysteine as [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0]. Hence, in this case, a 20-mer sequence will be represented by a vector of 400 dimensions (20 x 20). Detailed description on bi-profile vector encoding using Bayes Features is available in Shao *et al*. [8]. In short, feature vectors contain information from both positive position-specific and negative position-specific profiles. These profiles were generated by accounting for the frequency of occurrence of each amino acid at each position of the sequences in the positives pool (cleavage site sequences) and negatives pool (non-cleavage site sequences) respectively. Therefore, a 24-mer sequence (from the $P_{12}P_{12}$' dataset) would be represented by a feature vector of 48 dimensions (24 x 2), containing information of the residues in both positive (cleavage site sequences) and negative (non-cleavage site sequences) spaces.

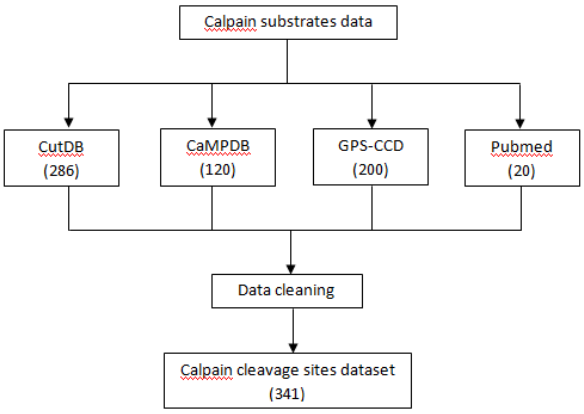

Figure 1. Calpain cleavage sites dataset construction. *(Numbers in parentheses indicate the number of unique calpain cleavage sites from the respective sources.)*

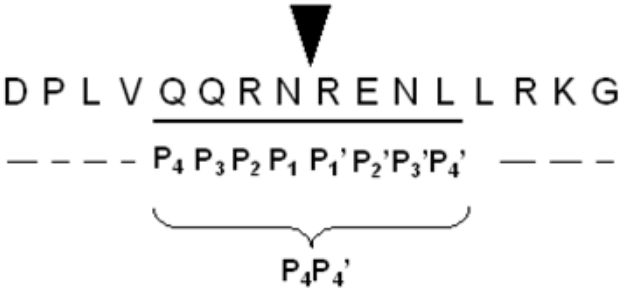

Figure 2. Example of a sequence segment for P_4P_4' dataset. (*Example shows a subsequence of human CDK5R2 (Uniprot: Q13319): The P_4P_4' dataset is comprised of octapeptides - in this case, "QQRNRENL" (underlined). Amino acids to the left of the scissile bond (indicated by the inverted triangle) are labelled P_1 (N) to P_4 (Q). Amino acids to the right of the scissile bond are labelled P_1' (R) to P_4' (L).*

D. *SVM model development*

To train and test the SVM models, we used the LIBSVM package provided by Chang and Lin [9]. For details on the SVM method, readers are advised to consult the article by Burges [10]. We used the radial basis function (RBF) kernel and performed grid-based optimization for γ and C using 10-fold cross-validation. In 10-fold cross-validation, the training set was randomly partitioned into ten subsets where one of the subsets was used as the test set while the other subsets were used for training the classifier. The trained classifier was evaluated using the test set. This procedure was repeated ten times using different subsets for testing, hence making sure that all subsets were utilized for both training and testing. The optimized γ and C values were applied towards training the entire training sets to generate the SVM classifiers for independent testing on the test sets.

E. *Evaluation of model performance*

A set of statistical variables were established to evaluate the performance of the SVM classifier for the prediction of calpain cleavage sites:

- True Positives (TP), for the number of correctly classified cleavage sites.
- False Positives (FP), for the number of incorrectly classified non-cleavage sites.
- True Negatives (TN), for the number of correctly classified non-cleavage sites.
- False Negatives (FN), for the number of incorrectly classified cleavage sites.

Sensitivity (Sn) and Specificity (Sp), which measures the capability of the model to correctly classify the cleavage sites and non-cleavage sites respectively, were computed as well. To measure the overall model performance, we computed Accuracy (A_{CC}).

$$S_n = \frac{TP}{TP + FN} \quad (1)$$

$$S_p = \frac{TN}{TN + FP} \quad (2)$$

$$A_{cc} = \frac{TP + TN}{TP + FN + TN + FP} \quad (3)$$

In addition, we plotted the receiver operating characteristic curve and computed the area under the curve (A_{ROC}) for threshold independent evaluation.

III. Results and Discussion

A. Sequence analysis of calpain cleavage sites

Experimental studies on calpain substrates have unveiled unique sequence specificities at the cleavage site and the flanking upstream and downstream regions [11]. It was shown that, unlike caspases and granzymes, calpains recognize a more diverse range of cleavage site sequence signatures, particularly at the P_1 position. As earlier protease-substrate functional studies were carried on a restricted range of bona fide substrates or through the use of *in vitro* peptide-based tools, it was difficult to accurately define the sequence specificities governing substrate recognition and cleavage. Hence, we have compiled the largest available dataset of experimentally verified cleavage sites of calpain substrates, and conducted an extensive analysis on the cleavage sites and flanking sequences to identify residue specificities which could assist hypothesis generation and the development of computational methods for predicting its degradome. Here, a total of 341 unique calpain cleavage sites were extracted from various publicly available databases and recent literature. We extracted sequence segments of varying lengths and compositions centered on the P_1 cleavage site and derived seven corresponding datasets, labeled P_4P_4', P_8P_8', $P_{12}P_{12}$', $P_{16}P_{16}$', $P_{20}P_{20}$' P_4P_{12}' and $P_{12}P_4$'. As a control, we further extracted an equal number of "non-cleavage sites" (for negative examples) by randomly selecting P_1 residues on substrates which were not present in the cleavage sites pool. On the $P_{20}P_{20}$' dataset, we computed the relative position-specific propensities of each amino acid (termed as P_x) at the different residue positions on the 40-mer segment. In this case, P_x quantifies the relative frequency of a particular amino acid in the cleavage site sequences over the frequency of the same amino acid in the non-cleavage site sequences at the same position. As shown in Table 1, measurements of average P_x in the $P_{20}P_{20}$' sequences indicate a decidedly greater enrichment of some PEST (for Pro, Glu/Asp, Ser and Thr) residues in the cleavage site sequences. Pro, Ser and Thr residues were found to have average P_x scores of 1.40, 1.23 and 1.18 respectively, while Glu and Asp were both found to have average P_x slightly under 1.0. While enrichment of PEST residues in the vicinity was shown to correlate with proteolytic cleavage in a range of protease-substrate systems, its biochemical significance in calpain substrate cleavage is still uncertain. It was previously demonstrated that calpain cleavage often occurred at regions adjacent to PEST regions [12] - however, mutations of PEST residues in a calmodulin-binding domain on Ca^{2+}-ATPase were shown to have little or no inhibitory effect on calpain cleavage in a separate study. Interestingly, our results suggest that presence of unusually high concentration of PEST residues, especially proline, could at least serve as one of the factors for the recognition of potential cleavage sites. Next, to quantify position-specific residue propensities on the cleavage sites, we plotted a sequence logo using the P_8P_8' cleavage site sequences and a heat map of P_x scores based on the $P_{20}P_{20}$' dataset (in Figures 3 and 4 respectively). It was observed that the P_1 position showed a strong, but comparatively balanced, conservation patterns for Ser, Gly, Arg and Thr residues. Interestingly, Leu, and to lesser extent Thr and Leu, were found to be dominant at the P_2 position. These results contrasted somewhat with Tompa *et al*. [11], who examined the amino acids preferences in the cleavage site vicinity of a small dataset of 49 calpain substrates and found that both tyrosine and arginine were preferred at the P_1 position, while leucine, threonine and valine are more prevalent at P_2. In addition, we observed that Ser, Ala, Leu were dominant at the P_1' position, and an unusually high propensity for proline residue in the immediate downstream region of the cleavage site (from P_2' to $P1_{10}$'), with a relative peak at P_3. Our results highlight several unique sequential determinants of calpain cleavage in the vicinity of the cleavage site – and we hypothesize that these complex patterns, as represented by the unique position-specific residue propensities, could be exploited using machine learning algorithms for the development of accurate computational prediction models.

TABLE I. AVERAGE P_X OF AMINO ACIDS

Amino acid	Average P_x
A	1.00
C	0.38
D	0.93
E	1.02
F	0.95
G	1.12
H	0.87
I	0.91
K	1.08
L	0.90
M	1.23
N	1.10
P	1.40
Q	1.13
R	1.20
S	1.30
T	1.19
V	1.09
W	1.30
Y	1.00

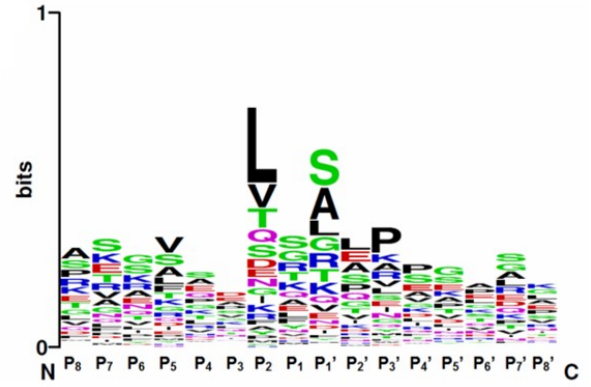

Figure 3. Sequence logo of P_8P_8' dataset

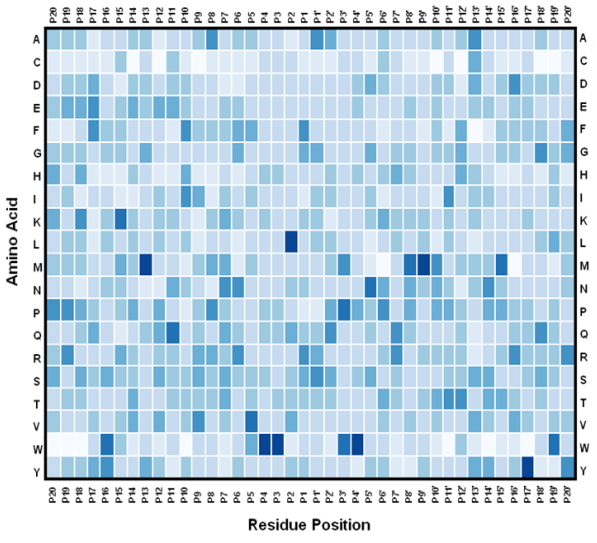

Figure 4. Heat map of P_x values of $P_{20}P_{20}'$ dataset

B. SVM-based prediction of calpain cleavage sites

To encapsulate these complex patterns of residue conservation and position-specific propensities for computational prediction, we constructed SVM prediction classifiers employing simple binary encoding and Bayes Feature Extraction (BFE) for vector representation. Vector representation using the BFE approach was shown to significantly improve performance in several bio-computational problems - such as the prediction of protein methylation sites [8], caspase cleavage [13] and linear B-cell epitopes [14] - over simple binary encoding schemes. Essentially, in the application of BFE, feature vectors are encoded in a bi-profile manner containing attributes from positive position-specific and negative position-specific profiles. These bi-profile vectors are generated by accounting for the frequency of occurrence of each amino acid at each position of the sequence in the positives (cleavage sites) pool and negatives (non-cleavage sites) pool respectively. In this problem, we trained a series of SVM classifiers on sequence windows of diverse lengths and compositions using both simple binary encoding and BFE schemes. Sequences from all datasets were partitioned into training and independent test sets. 10-fold cross-validation was conducted within a grid-search optimization process using the training sets to obtain the optimal set of SVM parameter values. The final SVM classifiers were trained on the entire training set using the optimized parameters and evaluated on the independent test sets. As shown in Table 2, the P_4P_4' classifier utilizing simple binary encoding (P_4P_4'-SVM) achieved an accuracy of 77.00% and A_{ROC} of 0.83 on independent testing. As the sequence window extends further upstream and downstream residues from the cleavage site, a slight reduction in prediction performance was observed. These results were contrary to those obtained with sequence-based prediction of other protease cleavage sites, such as caspases. One possibility for this disparity could be in the limited ability of the learning algorithm to model more complex sequence patterns which is likely to be inherent in calpain substrate cleavage sites. When Bayes Feature Encoding scheme was utilized, the P_4P_4' classifier (P_4P_4'-Bayes) achieved an accuracy of 78.00% and A_{ROC} of 0.84 (Table 3), while classifiers trained using longer sequence windows showed gradual improvements in the prediction performance, with accuracy and A_{ROC} scores peaking at 86.00% and 0.93 using the $P_{16}P_{16}$-Bayes and $P_{20}P_{20}'$-Bayes classifiers respectively. Interestingly, in both feature representation schemes, prediction performances did not significantly improve with sequences longer than P_8P_8'. This could be due to that fact that much of the information specific for differentiating cleavage sites from non-cleavage sites are encoded within the sequences situated closer to the cleavage sites, as evidenced by the unique residue propensities discussed earlier. In addition, accuracy and A_{ROC} scores across most sequence lengths and compositions were generally higher for classifiers trained using the BFE scheme, with the greatest improvements observed when longer sequences (P_6P_6', P_8P_8', $P_{10}P_{10}'$ and $P_{14}P_{10}'$) were employed. GPS-CCD was demonstrated recently to produce one of the most accurate models for calpain cleavage site prediction [6]. Using a unique amino acid substitution matrix derived from 368 experimentally-verified calpain cleavage sites, it achieved a cross-validated A_{ROC} of 0.838 and accuracy of 89.98% using the best performing classifier. The GPS-CCD algorithm was shown also to out-perform other methods, such as PoPS [15] and SitePrediction [16], when tested using common independent datasets. These results suggest that our best models are comparable with the state of the art algorithms.

TABLE II. PERFORMANCE OF SVM CLASSIFIERS WITH SIMPLE BINARY ENCODING

SVM Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)	A_{ROC}
P_4P_4'-SVM	70.00	84.00	77.00	0.83
P_8P_8'-SVM	74.00	76.00	75.00	0.80
$P_{12}P_{12}'$-SVM	66.00	76.00	71.00	0.79
$P_{16}P_{16}'$-SVM	64.00	82.00	73.00	0.83
$P_{20}P_{20}'$-SVM	62.00	82.00	72.00	0.83

TABLE III. PERFORMANCE OF SVM CLASSIFIERS WITH BAYES FEATURE EXTRACTION

SVM Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)	A_{ROC}
P_4P_4'-Bayes	76.00	80.00	78.00	0.84
P_8P_8'-Bayes	80.00	80.00	80.00	0.89
$P_{12}P_{12}'$-Bayes	82.00	82.00	82.00	0.91
$P_{16}P_{16}'$-Bayes	88.00	84.00	86.00	0.92
$P_{20}P_{20}'$-Bayes	86.00	84.00	85.00	0.93

C. *Calpain cleavage of receptor tyrosine kinases*

Receptor tyrosine kinases (RTKs) belong to a sub-class of the protein kinase superfamily which function as plasma membrane-bound receptors transducing extracellular signals mediating cell survival, proliferation, embryonic development, adult homeostasis and many other critical processes [17]. As RTK activity in resting, normal cells is tightly controlled, mutations or structural aberrations in RTKs were shown to convert them to potent oncoproteins, contributing to the development and progression of many cancers. Despite its widespread influences in several cellular pathways, there are limited studies on the role of calpain cleavage in RTK signalling and related cell survival pathways. The epidermal growth factor receptor (EGFR) and its relative Erbb2 were reported to be cleaved by calpain at multiple sites along their cytoplasmic domain *in vitro* [18,19]. In both cases, calpain-mediated cleavage was found to abrogate intracellular signalling downstream of receptor activation, leading to the perturbation of cell survival responses. Most notably, calpain cleavage of Erbb2 was reported to mediate drug sensitivity and pathway of survival Erbb2-positive breast cancer cells, suggesting the potential of this protease as a therapeutic target. Interestingly, recent studies have also implicated caspases as upstream modulators of RTK function. In particular, caspase cleavage of EGFR [20,21], Erbb2 [22], MET [23-25], RET [26] and ALK [27], was shown to obstruct RTK signalling cascade at various stages through specific cleavage of the receptor intracellular domains. Given that calpains are intimately involved in a number of regulatory cross-talks with caspases and target many similar substrates, it is tempting for us to speculate if an unexpectedly broader range of RTKs could be cleaved and regulated by calpains in a similar fashion. Accordingly, we applied the SVM classifier (SVM-$P_{20}P_{20}'$ using BFE) on a well-characterized group of caspase-cleaved RTKs - namely EGFR, Erbb2, RET, MET and LTK - and predicted for potential cleavage sites (results in Figure 5). All RTKs were predicted to harbor cleavage sites which are distributed throughout the extracellular and intracellular regions, with a number of them co-localizing with established caspase cleavage sites. In particular, calpain cleavage sites were found at the juxtamembrane regions of EGFR (Gly-696), Erbb2 (Gly-704 and Met-709) and Met (Asp-981 and Ala-1005). It was previously demonstrated that caspase cleavage of MET receptor at the juxtamembrane region (Asp-1000) resulted in the truncation of the full-length receptor into a membrane-bound portion and an intracellular fragment, and consequently the termination of MET signaling [23]. In addition, cleavage of the ALK protein by caspases at a similar juxtamembrane position (Asp-1160) was shown to expose an intracellular region which induces pro-apoptotic signals downstream [27]. The co-localization of calpain cleavage sites suggest that calpains could potentially disrupt receptor structure and function in the same manner as the caspases. We noted that several calpain cleavage sites were found within the kinase domains of RET (Gly-828 and Ser-835), Erbb2 (Gly-727), EGFR (Gly-719) and MET(Ala-1196). As intact kinase domains are required for auto-phosphorylation of the receptors, it is plausible that calpain cleavage could directly disrupt kinase activity and terminate downstream signaling. Interestingly, earlier studies have reported that proteolytic fragments bearing the motif "RLLGI" derived from the tyrosine kinase domains of EGFR and Errb2 were able to induce apoptosis in cells - further suggesting that calpain cleavage at these sites could generate downstream pro-apoptotic signals in addition to signaling inhibition [29]. We found a notable proportion of calpain cleavage sites at the C-terminus of EGFR and Errb2 proteins. This suggests another point of functional intervention since it was earlier reported that the caspase-mediated cleavage of the cytoplasmic tail of Errb2 releases an intracellular carboxy terminal which was pro-apoptotic, and cleavage of EGRF at a similar location was shown to terminate downstream signaling through the removal of the binding sites for Cbl and Grb2 on the receptor [29,30]. Calpains have been shown to cleave a myriad of substrates, where most if not all, are localized in the cytoplasm. However, they are also known to be released from cells into the extracellular environment through leakage from injured and dying cells or via regulated secretory pathways. Here, we noted that a sizeable proportion of calpain cleavage sites were located on the extracellular domains of EGFR, RET, MET and ALK. It is quite conceivable that cleavage sites localized within ligand-binding regions of these proteins could be targeted by calpains in the extracellular milieu for cleavage, and as a result, inhibiting transmembrane receptor signaling. Taken together, the presence and distribution of these predicted cleavage sites across the RTK family highlights potentially novel roles of caspase cleavage in the direct regulation of RTK activity.

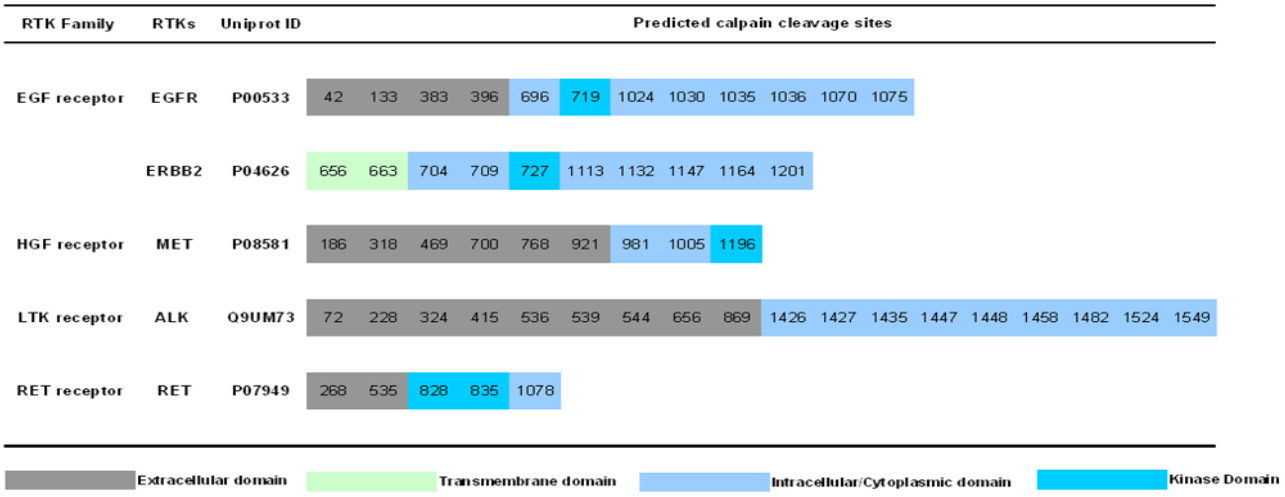

Figure 5. Schematic map of predicted calpain cleavage sites on selected receptor tyrosine kinase (RTK) family members *(P_1 positions of predicted cleavage sites on each RTK are listed. Grey sections indicates location of cleavage site within the extracellular domain, green indicates location within transmembrane domain, light blue indicates location within intracellular domain and darker blue indicates location within kinase domain.)*

As in the case for caspases, it is tempting to speculate a general phenomenon whereby calpains terminates the function of the pro-apoptotic receptor protein, and in some cases, converts it into a pro-apoptotic, through the cleavage of functional domains or release of pro-apoptotic cytoplasmic domains. Indeed, such dramatic reversal of protein function has been well documented in studies on caspase cleavage of serine/threonine protein kinases such as MEKK1 and MEKK4, as well as the anti-apoptotic Bcl-2 and Bcl-xl proteins [13]. Clearly, further experimental investigations will be necessary to elucidate the deeper intricacies of calpain-mediated RTK cleavage and its broader biochemical consequences.

IV. Conclusion

In this paper, we constructed a comprehensive database of experimentally verified calpain cleavage sites for analysis and development of prediction methods. We discovered that flanking sequences of cleavage sites possess distinctive residue composition and position-specific propensity patterns which could be helpful in discriminating the cleavage sites from non-cleavage sites *in silico*. We have rigorously tested SVM classifiers employing simple binary encoding and the Bayes Feature Extraction schemes to predict calpain cleavage sites. Results also show that our best classifiers are comparable with, if not better than existing algorithms. We predicted calpain cleavage sites on proteins from the receptor tyrosine kinase family and discovered potential sites of cleavage at critical regulatory domains, suggesting a novel role of calpains as a direct regulator of receptor tyrosine kinase activity in cell survival and cell death pathways. In the immediate future, we will be exploring the influence of cleavage site secondary structures, solvent accessibilities and other physicochemical properties on protease-substrate cleavage specificities, as well as their potential for enhancing the performance of our SVM prediction models. Computational prediction of calpain substrates will complement on-going experimental efforts and refine our understanding of the biochemistry of this fascinating family of proteases.

Acknowledgment

This study was sponsored by a research grant from the Joint Council Office (JCO) of A*STAR Singapore.

References

[1] C. Lopez-Otin and C. M. Overall, "Protease degradomics: a new challenge for proteomics," Nat Rev Mol Cell Biol, vol. 3, pp. 509-19, Jul 2002.

[2] J. S. Evans and M. D. Turner, "Emerging functions of the calpain superfamily of cysteine proteases in neuroendocrine secretory pathways," J Neurochem, vol. 103, pp. 849-59, Nov 2007.

[3] S. L. Chan and M. P. Mattson, "Caspase and calpain substrates: roles in synaptic plasticity and cell death," J Neurosci Res, vol. 58, pp. 167-90, Oct 1 1999.

[4] Y. Igarashi, A. Eroshkin, S. Gramatikova, K. Gramatikoff, Y. Zhang, J. W. Smith, A. L. Osterman, and A. Godzik, "CutDB: a proteolytic event database," Nucleic Acids Res, vol. 35, pp. D546-9, Jan 2007.

[5] D. duVerle, I. Takigawa, Y. Ono, H. Sorimachi, and H. Mamitsuka, "CaMPDB: a resource for calpain and modulatory proteolysis," Genome Inform, vol. 22, pp. 202-13, Jan 2010.

[6] Z. Liu, J. Cao, X. Gao, Q. Ma, J. Ren, and Y. Xue, "GPS-CCD: a novel computational program for the prediction of calpain cleavage sites," PLoS One, vol. 6, p. e19001, 2011.

[7] G. E. Crooks, G. Hon, J. M. Chandonia, and S. E. Brenner, "WebLogo: a sequence logo generator," Genome Res, vol. 14, pp. 1188-90, Jun 2004.

[8] J. Shao, D. Xu, S. N. Tsai, Y. Wang, and S. M. Ngai, "Computational identification of protein methylation sites

through bi-profile Bayes feature extraction," PLoS One, vol. 4, p. e4920, 2009.
[9] C. Chang and C. Lin, "LIBSVM : a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 1-27, 2011.
[10] C. J. Burges, "A tutorial on support vector machines for pattern recognition. ," Data Mining and Knowledge Discovery, vol. 2, pp. 121-167, 1998.
[11] P. Tompa, P. Buzder-Lantos, A. Tantos, A. Farkas, A. Szilagyi, Z. Banoczi, F. Hudecz, and P. Friedrich, "On the sequential determinants of calpain cleavage," J Biol Chem, vol. 279, pp. 20775-85, May 14 2004.
[12] M. Molinari, J. Anagli, and E. Carafoli, "PEST sequences do not influence substrate susceptibility to calpain proteolysis," J Biol Chem, vol. 270, pp. 2032-5, Feb 3 1995.
[13] L. J. Wee, T. W. Tan, and S. Ranganathan, "SVM-based prediction of caspase substrate cleavage sites," BMC Bioinformatics, vol. 7 Suppl 5, p. S14, 2006.
[14] L. J. Wee, D. Simarmata, Y. W. Kam, L. F. Ng, and J. C. Tong, "SVM-based prediction of linear B-cell epitopes using Bayes Feature Extraction," BMC Genomics, vol. 11 Suppl 4, p. S21, 2010.
[15] S. E. Boyd, R. N. Pike, G. B. Rudy, J. C. Whisstock, and M. Garcia de la Banda, "PoPS: a computational tool for modeling and predicting protease specificity," J Bioinform Comput Biol, vol. 3, pp. 551-85, Jun 2005.
[16] J. Verspurten, K. Gevaert, W. Declercq, and P. Vandenabeele, "SitePredicting the cleavage of proteinase substrates," Trends Biochem Sci, vol. 34, pp. 319-23, Jul 2009.
[17] S. R. Hubbard and W. T. Miller, "Receptor tyrosine kinases: mechanisms of activation and signaling," Curr Opin Cell Biol, vol. 19, pp. 117-23, Apr 2007.
[18] M. Gregoriou, A. C. Willis, M. A. Pearson, and C. Crawford, "The calpain cleavage sites in the epidermal growth factor receptor kinase domain," Eur J Biochem, vol. 223, pp. 455-64, Jul 15 1994.
[19] S. Kulkarni, K. B. Reddy, F. J. Esteva, H. C. Moore, G. T. Budd, and R. R. Tubbs, "Calpain regulates sensitivity to trastuzumab and survival in HER2-positive breast cancer," Oncogene, vol. 29, pp. 1339-50, Mar 4 2010.
[20] S. S. Bae, J. H. Choi, Y. S. Oh, D. K. Perry, S. H. Ryu, and P. G. Suh, "Proteolytic cleavage of epidermal growth factor receptor by caspases," FEBS Lett, vol. 491, pp. 16-20, Feb 23 2001.
[21] Y. Y. He, J. L. Huang, and C. F. Chignell, "Cleavage of epidermal growth factor receptor by caspase during apoptosis is independent of its internalization," Oncogene, vol. 25, pp. 1521-31, Mar 9 2006.
[22] O. Tikhomirov and G. Carpenter, "Caspase-dependent cleavage of ErbB-2 by geldanamycin and staurosporin," J Biol Chem, vol. 276, pp. 33675-80, Sep 7 2001.
[23] D. Tulasne, J. Deheuninck, F. C. Lourenco, F. Lamballe, Z. Ji, C. Leroy, E. Puchois, A. Moumen, F. Maina, P. Mehlen, and V. Fafeur, "Proapoptotic function of the MET tyrosine kinase receptor through caspase cleavage," Mol Cell Biol, vol. 24, pp. 10328-39, Dec 2004.
[24] B. Foveau, C. Leroy, F. Ancot, J. Deheuninck, Z. Ji, V. Fafeur, and D. Tulasne, "Amplification of apoptosis through sequential caspase cleavage of the MET tyrosine kinase receptor," Cell Death Differ, vol. 14, pp. 752-64, Apr 2007.
[25] J. Deheuninck, B. Foveau, G. Goormachtigh, C. Leroy, Z. Ji, D. Tulasne, and V. Fafeur, "Caspase cleavage of the MET receptor generates an HGF interfering fragment," Biochem Biophys Res Commun, vol. 367, pp. 573-7, Mar 14 2008.
[26] M. C. Bordeaux, C. Forcet, L. Granger, V. Corset, C. Bidaud, M. Billaud, D. E. Bredesen, P. Edery, and P. Mehlen, "The RET proto-oncogene induces apoptosis: a novel mechanism for Hirschsprung disease," EMBO J, vol. 19, pp. 4056-63, Aug 1 2000.
[27] J. Mourali, A. Benard, F. C. Lourenco, C. Monnet, C. Greenland, C. Moog-Lutz, C. Racaud-Sultan, D. Gonzalez-Dunia, M. Vigny, P. Mehlen, G. Delsol, and M. Allouche, "Anaplastic lymphoma kinase is a dependence receptor whose proapoptotic functions are activated by caspase cleavage," Mol Cell Biol, vol. 26, pp. 6209-22, Aug 2006.
[28] T. Nakagawa and J. Yuan, "Cross-talk between two cysteine protease families. Activation of caspase-12 by calpain in apoptosis," J Cell Biol, vol. 150, pp. 887-94, Aug 21 2000.
[29] O. Tikhomirov and G. Carpenter, "Identification of ErbB-2 kinase domain motifs required for geldanamycin-induced degradation," Cancer Res, vol. 63, pp. 39-43, Jan 1 2003.
[30] A. M. Strohecker, F. Yehiely, F. Chen, and V. L. Cryns, "Caspase cleavage of HER-2 releases a Bad-like cell death effector," J Biol Chem, vol. 283, pp. 18269-82, Jun 27 2008.