Published in final edited form as:

Sens Actuators B Chem. 2007 July 16; 125(1): 149–157.

Using Support Vector Machine Regression to Model the Retention of Peptides in Immobilized Metal-affinity Chromatography

B. G. Kermani, I. Kozlov, P. Melnyk, C. Zhao, J. Hachmann, D. Barker, and M. Lebl

Illumina, Inc. 9885 Towne Centre Drive, San Diego, CA 92122

Abstract

Retention of histidine-containing peptides in immobilized metal-affinity chromatography (IMAC) has been studied using several hundred model peptides. Retention in a Nickel column is primarily driven by the number of histidine residues; however, the amino acid composition of the peptide also plays a significant role. A regression model based on support vector machines was used to learn and subsequently predict the relationship between the amino acid composition and the retention time on a Nickel column. The model was predominantly governed by the count of the histidine residues, and the isoelectric point of the peptide.

Keywords

Support Vector Machines; Regression; Peptide; Metal-affinity Chromatography; Retention Time

Introduction

Immobilized metal-affinity chromatography (IMAC) is an important tool for purification of proteins containing residues that form metal complexes, e.g., histidine, cysteine and tryptophan [1–4]. The poly-histidine tag is extremely useful in molecular biology where it serves to facilitate isolation of genetically engineered proteins from complex mixtures and can be used for targeted immobilization of these proteins [4,5].

We are developing a high-throughput method for the purification of peptide-oligonucleotide conjugates. One of our strategies is to place three histidines (His3) at the amino-terminus of the peptide, and three histidines at the 5' end of the oligonucleotide. When joined together, the six histidines (His6) should form a tag that can be bound to a Nickel Sepharose affinity column. After washing away unreacted components, the purified peptide-oligonucleotide conjugate can be eluted with a gradient of increasing imidazole concentration. For this strategy to be successful, the concentration of imidazole that elutes His3 must be significantly less than the concentration that elutes His6. In addition, the elution concentration for His6 should be relatively insensitive to the identity of amino acids surrounding the His6 tag. For brevity, throughout this manuscript, we use the term “imidazole concentration” in lieu of “the concentration of imidazole at which elution occurs.”

Surprisingly, there is little information available in the literature on the relative affinity for Nickel-Sepharose, in the presence of imidazole gradients, of polyhistidine-containing peptides and the influence of surrounding amino acids. Consequently, we synthesized an array of model compounds containing different numbers of histidines in various sequential arrangements and

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

in combination with various other amino acids. Here we present comprehensive information that should facilitate the design and purification of engineered peptides and proteins.

By training a model using a subset of the peptides (with measured imidazole concentration required for elution from a Nickel Sepharose affinity column) we were able to predict the imidazole concentration needed for elution from the column for a large group of peptides. Table 1 shows the amino acid sequence for the peptides of this experiment, along with their measured imidazole concentrations. Figure 1 shows the dependency of imidazole concentration on several parameters, including the number of histidines and the isoelectric point. As can be seen from Figure 1a, there is a relationship between the number of histidines (nHis, nH) and the imidazole concentration; however, this relationship is not simple. Figure 1b shows that imidazole concentration is also a function of the composition of the peptide (in this case as measured by the isoelectric point). Figure 1c shows how imidazole concentration varies as a function of the number of histidines and the number of histidine pairs (nHisHis, nHH).

Methods

The objective of this study is to generate a regression model that takes several parameters from the peptide sequence, and generates an estimate for the imidazole concentration. The necessary (optimal) number of the inputs is unknown (Figure 2). The goal is, thus, to find the best model and the minimal number of informative inputs that are capable of predicting the imidazole concentration. A regression (as opposed to a classification) model is needed for this predictor, since the output, i.e., imidazole concentration, takes continuous values.

To address the problem, we started with the simplest models, i.e., a linear model. However, a simple linear model was found to be insufficient in capturing all the dependencies of the system. Therefore, we adopted various nonlinear regression models for trial. The most simplistic (nonlinear) model is a model with one input –number of histidines. The performance of this model was deemed insufficient in view of the alternatives. The next model that we tried was a model with two inputs –nHis and isoelectric point. A thin-plate spline was used to model this two-input system. While the performance of this system was better than the previous one, it was found that in this model, a linear combination of nHis and nHisHis is more informative than nHis by itself. Therefore, a third model with the following parameters was made: 1) isoelectric point, 2) $nHis + \text{Beta} * nHisHis$ (Figure 3). The parameter Beta was found to have an optimal value of approximately 2 (Figure 4). The Beta value of 2 has a physical interpretation, i.e., it matches the hypothesis that in a Nickel column, each Nickel atom interacts with two histidine residues.

While the performance of the thin-plate spline is rather satisfactory, it has the following shortcomings:

1. It is only capable of handling 2 inputs. Note that in order to handle the nHisHis situation, we had to sum it with the nHis, in a linear fashion, in order to avoid having an extra input.
2. The surface of the thin-plate spline had local modulations that are aggravated by local noise or data sparsity. These modulations are expected and are due to the “local” (as opposed to global) nature of a spline-based model.
3. Finding the optimal smoothing factor for the thin-plate spline is non-trivial, and requires human supervision.

Due to the above limitations, we opted to use support vector machines. By doing so, we were able to easily extend the model to larger number of inputs (than 2). Support vector machines

have been in the forefront of the machine learning algorithms and have gained great popularity in the last decade. This popularity is mostly stemmed from the following [6,7]:

1. SVMs are based on the structural risk minimization, and thus have a builtin mechanism for regularization, i.e., they are robust to over-training.
2. Soft-margin SVMs can combine the empirical risk minimization with the structural risk minimization, in order to find an adequate trade-off between the model complexity and the prediction error.
3. By using nonlinear kernels, nonlinear systems can be modeled using a linear system in the feature space, and yet without having a need to project the solution into the feature space.
4. The solution to the underlying optimization problem in SVMs is amenable to classical optimization techniques, namely Quadratic Programming and Linear Programming.

The most popular application of SVM is in binary classification [8,9]. However, it has been shown that regression problems can also be modeled using SVM [6,7,10]. The SVM regression (SVM-R) models are based on the epsilon-insensitive error models [6,7]. In an epsilon-insensitive model, the (absolute) error values of epsilon or lower are mapped to zero, and the other error values are mapped either linearly or quadratically. The linear error models are generally more resilient to outliers (with high values). Therefore, in this study, we considered only linear epsilon-sensitive regression SVMs.

Parameter Selection

Table 2 shows the specifications of the SVM-R that was used for this study. The SVM-R models (similar to most binary classifier SVMs) use a regularization parameter C [6] in order to provide a balance between the prediction error and model complexity. The optimal value of C is problem-dependent. Various methods (including cross-validation) are usually used to set this value. In this study, based on several trial-and-errors, the value of 10 was found to be appropriate for the parameter C, although the results were not too sensitive to the specific choice of this parameter.

In order to be able to define the parameter epsilon, we defined an alternate error model. In this model, first for each experiment, the correlation coefficient of the target values and the output values was found. Then, the error was defined as one minus this correlation coefficient. An error which was based on correlation coefficient had a more tangible physical meaning for our scientists, and thus was preferred over the alternatives, e.g., mean-squared error, mean absolute deviation. With this definition, the error values of 0.01 or lower were deemed clinically insignificant, and thus the epsilon value was set to 0.01.

Equation 1

$$\begin{aligned}
 a) \text{ TrainingError} &= 1 - \rho(\text{TrainingOutput}, \text{TrainingTarget}) \\
 b) \text{ TestError} &= 1 - \rho(\text{TestOutput}, \text{TestTarget}) \\
 c) \text{ BootstrapError} &= 0.632 * \text{TestError} + 0.368 * \text{TrainingError} \\
 d) \text{ BootstrapCorrelation} &= 1 - \text{BootstrapError} \\
 e) \text{ BootstrapRsquared} &= (\text{BootstrapCorrelation})^2
 \end{aligned}$$

A polynomial SVM [11] was chosen for this study. An SVM with order 1 was unable to perform as well as an SVM with order 2. However, beyond order 2, the SVM did not significantly improve the performance, and only resulted in longer training sessions. Therefore, the order was fixed at 2.

Model Validation

In order to validate the model, a bootstrapping technique [12] was used. In this technique, a total of 100 runs were executed for each scenario. For each run, a subset of the total data was chosen at random (with substitution). This set was called the training set. The difference between this set and the original set was defined as the test set. For each run, the training error and the test error were defined as given in Equation 1a,b. Subsequently, a composite 0.632 bootstrap error [13] was computed (Equation 1c); and this error was used to find a bootstrap R-squared value (Equation 1e). This R-squared value was used as a representation for the appropriateness of the model.

Results

Synthesized model peptides and the concentration of imidazole needed for their elution (extrapolated from the retention time in gradient elution), are given in Table 1. All sequences were synthesized on tyrosine-modified resin to simplify UV detection (280 nm) of peptides eluted with increasing imidazole concentrations. Figure 5 shows the results of running SVM-R on systems with the following inputs:

1. Number of arginines (R) (one input);
2. Number of histidines (H) (one input);
3. Number of histidine pairs (one input);
4. Compound histidine count (Hc), i.e., number of histidines plus twice the number of histidine pairs (one input);
5. nHis and nHisHis (two inputs);
6. isoelectric point (one input);
7. Hc and isoelectric point (two inputs);
8. nHis, nHisHis and isoelectric point (three inputs);
9. Hc, isoelectric point, and number of arginines (three inputs);
10. nHis, nHisHis, isoelectric point, and the number of arginines (four inputs).

By comparing the bootstrap R-squared values from the tested systems, one would make the following observations:

1. The Nickel column is also responsive to arginine, although it is more sensitive to histidine.
2. The Nickel column is more sensitive to pairs of histidines, as opposed to single histidines.
3. Isoelectric point by itself is informative in the predictive system.
4. The performance of the system improves significantly if the isoelectric point and histidine (or histidine pair) information are supplied simultaneously.
5. The best performance is achieved in a system with four inputs –number of histidines, number of histidine pairs, isoelectric point, and the number of arginines.
6. The performance of a three-input system with inputs –Hc, isoelectric point and number of arginines is not too different from the optimal system.

Visualizing systems with more than 2 inputs is difficult. Figure 6 shows the performance of a 2-input (Hc and isoelectric point) system. The quadratic shape of the surface is due to the

specific form of the SVM-R kernel, i.e., second-order polynomial. The proximity of the actual and predicted points is a testimony to the appropriateness of the fit.

Figure 7 illustrates the differences between the target and predicted values for one of the 100 bootstrap runs. The residuals have patterns, which indicate that the fit has not been optimal. This is partially due to the small sample size, and partially due to the low order (i.e., 2) of the nonlinear kernel. However, due to the small sample size, the order of the system was not increased (beyond 2) in order to keep the model parsimonious, and thus more resilient to over-training.

Conclusions

We performed an exhaustive study of the affinity of histidine-rich peptide sequences for a Nickel Sepharose column, using elution by a gradient of increasing imidazole concentration as a measure of affinity. Retention of histidine-containing peptides depends on the arrangement of histidines as well as the composition of other amino acids in the sequence. The highest performing regression model was obtained with a 4-input system, with the following inputs: 1) number of histidines, 2) number of histidine pairs, 3) number of arginines, and 4) isoelectric point of the peptide sequence. We used the R-squared metric for gauging the performance of the learning algorithm. This metric was a function of the agreement between the predicted and the observed values of imidazole concentration in our model. The system rendered a bootstrap R-squared value of approximately 0.85, while trained with a second-degree polynomial soft-kernel epsilon-insensitive regression SVM.

Acknowledgements

The project was partially supported by NIH SBIR grant R44 AI056869-02.

References

1. Porath J, Carlsson J, Olsson I, Belfrage G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 1975;258:598–599. [PubMed: 1678]
2. Hochuli E, Dobeli H, Schacher A. New metal chelate adsorbent selective for proteins and peptides containing neighboring histidine residues. *J Chromatogr* 1987;411:177–184. [PubMed: 3443622]
3. Sulkowski E. Purification of proteins by IMAC. *Trends Biotechnol* 1985;3:1–7.
4. Smith MC, Furman TC, Ingolia TD, Pidgeon C. Chelating peptide-immobilized metal ion affinity chromatography. *J Biol Chem* 1988;263:7211–7215. [PubMed: 3284883]
5. Hochuli E, Bannwarth W, Dobeli H, Gentz R, Stuber D. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate absorbent. *Bio/Technology* 1988;6:1321–1325.
6. Cristianini, N.; Shawe-Taylor, J. Support vector machines and other kernel-based learning methods. Cambridge University Press; Cambridge, CB2 2RU, UK: 2000.
7. Scholkopf, B.; Smola, AJ. Learning with Kernels. The MIT Press; 2002.
8. Brown M, Grundy WN, Lin D, Cristianini N, Sugnet CW, Ares M, Haussler D. Genetics knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci* 2000;97:262–267. [PubMed: 10618406]
9. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci* 2001;98(26):15149–54. [PubMed: 11742071]
10. Bennett KP, Campbell C. Support Vector Machines: Hype of Hallelujah? *SIGKDD Explorations* 2000;2:1–13.
11. SVM and Kernel Methods MATLAB Toolbox, Insa de Rouen. <http://asi.insa-rouen.fr/~arakotom/toolbox/index.html>
12. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. Chapman & Hall/CRC; Boca Raton, FL, U.S.A: 1994.

13. Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J of American Statistical Association* 1983;78:316–331.
14. Lebl M. Centrifugation Based Automated Synthesis Technologies. *J Assoc Lab Autom* 2003;8:30–36.
15. Lebl M. New Technique for High-Throughput Synthesis. *Bioorg Med Chem Lett* 1999;9:1305–1310. [PubMed: 10340619]
16. Hachmann JP, Lebl M. Alternative to piperidine in Fmoc solid phase synthesis. *Journal of Combinatorial Chemistry* 2006;8:149. [PubMed: 16529506]
17. King DS, Fields CG, Fields GB. A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. *Int J Peptide Prot Res* 1990;36:255–266.

Appendix

Experimental Setup

Fmoc amino acids, BOP reagent and Rink resin (0.42 mmol/g) were purchased from Novabiochem (EMD Biosciences, Inc. San Diego, CA, USA). Solvents were from VWR International, Inc. (West Chester, PA, USA). 4-Methylpiperidine was from Sigma-Aldrich (Milwaukee, WI, USA).

Rink resin (300 mg) was added into a mixture of DMF and DCM (10 ml total) to form a non-sedimenting suspension which was distributed into the wells of flat bottom polypropylene microtiterplates (Evergreen Scientific, Los Angeles, CA, USA). The plates were placed into a centrifugal synthesizer [14,15]. An additional 100 μ l of DMF was added into the plate wells (beads sedimented) and the plate was centrifuged with a tilt of 6 degrees. A standard protocol was used for the synthesis to remove the Fmoc protecting group; 4-Methylpiperidine was used instead piperidine [16]. Individual Fmoc-protected amino acids (0.3M solution in 0.3M HOBt in DMF) were pipetted to the wells, and a solution of BOP (0.6M in DMF) and 1.2 M DIEA in DMF was delivered to each well. Plates were oscillated five times and allowed to rest for 50 seconds. (During oscillation, the plates are rotated at a speed at which the liquid does not overflow the wall of the well and solid support moves towards the outer side of the well. When the rotation is stopped, liquid returns to the horizontal position and beads distribute at the well bottom, thus mixing the well content.) This procedure was repeated 30 times. The plate was centrifuged and the addition of amino acids and reagents was repeated. After another 30 cycles of oscillation and pausing, the reagents were removed by centrifugation and washing and de-protection was repeated to prepare the plate for the next cycle of synthesis.

At the end of the synthesis the plate was dried in vacuo and 150ul of mixture K [17] (TFA/thioanisole/water/phenol/EDT: 82.5:5:5:5:2.5 v/v) was added. The plate was capped and shaken on the plate shaker for 3 h. The suspension was transferred by multi-channel pipettor to a filter plate (Orochem Technologies, Lombard, IL, USA). The filtrate was collected in a deep well plate (VWR) and precipitated with ether (600 μ l), and after standing in the refrigerator for 2 h, a pellet was formed by centrifugation. The supernatant was removed by a surface suction device and the pellet was re-suspended in ether (600 μ l) and centrifuged again. The process of supernatant removal and re-suspension was repeated three times. The product was dried in a Speedvac (ThermoSavant, Waltham, MA, USA), dissolved in 200ul of H₂O, or 50% dimethylsulfoxide (DMSO)-50% H₂O and samples of 20 μ l were taken into 180 μ l of water. Twenty microliters were injected onto an HPLC column (Waters, Milford MA, USA, μ Bondapak, C18, 10 μ particle, 125 Å pore, 3.9 $\times$ 150 mm, gradient 0.05% TFA in H₂O to 70% acetonitrile, 0.05% TFA in 15 minutes, flow rate, 1.5ml/min, detection by UV at 217 nm). MS was performed at HT-Labs (San Diego, CA, USA).

The peptides were analyzed using HPLC containing a 1 ml volume HisTrap column (Amersham Biosciences, Piscataway, NJ, USA) with the detection at 260 nm. The peptides

were injected in 0.02M sodium phosphate buffer pH 7.4 containing 0.5 M NaCl. The concentration of imidazole was increased linearly from 0 to 0.5M during 20 min.

Figure 1.

(a: top-left) relationship between imidazole concentration and nHis; (b: top-right) relationship between imidazole concentration and isoelectric point; (c: bottom-left) the number of pairs of histidine versus the number of single histidines in the set; (d: bottom-right) relationship between imidazole concentration as a function of nHis and isoelectric point.

Figure 2.
The general diagram of the regression problem

Figure 3.

The prediction results of the thin-plate spline model with the following inputs: 1) isoelectric point, 2) a linear combination of nHis and nHisHis. The points that are on the surface are the predicted. The separation between the actual and the predicted points shows the error in prediction.

Figure 4.

The error in the prediction of imidazole concentration while using the above thin-plate spline model, as a function of the parameter Beta. The optimal value of Beta appears to be close to 2.

Figure 5.

The Box Plot of the Bootstrap R-squared values for the described SVM-R system, with the following inputs: 1) Number of argenines (R) (one input); 2) Number of histidines (H) (one input); 3) Number of HisHis pairs; 4) Compound histidine count (Hc), i.e., number of histidines plus twice the number of HisHis pairs (one input); 5) nHis and nHisHis (two inputs); 6) isoelectric point (one input); 7) Hc and isoelectric point (two inputs); 8) nHis, nHisHis and isoelectric point (three inputs); 9) Hc, isoelectric point, and number of argenines (three inputs); 10) nHis, nHisHis, isoelectric point, and the number of argenines.

Figure 6. Surface plot of the imidazole concentration, for a two input systems, with Isoelectric point and Hc as the two inputs. The blue points are the expected (experimental) measurements, and the red points are the SVM-predicted points.

Figure 7.

The results of a single training event. The left plot shows the correlation of the predicted values for imidazole concentration versus the target (experimentally measured) values.

Table 1

Peptides and Imidazole concentration

Peptide	Im concentration, M
GAGAHGAGAGY	0.08
GAGHHDHHGAY	0.09
GAGHHEHHGAY	0.09
HGAGAGAGAGY	0.09
HHAGAGAGAGY	0.09
HRIFLAGDEDY	0.09
GAGAHHGAGA	0.1
GAGAHHGAGA	0.1
GAGAHHGAGAY	0.1
GAGAHHGAGAY	0.1
GAGAHWGAGAY	0.1
GAGRWWGAGAY	0.1
GAHGAGAHGAY	0.1
GAHGAGHAGAY	0.1
GAHGAHAGAY	0.1
HGAGAGAGARY	0.1
HRIFLAGDKDY	0.1
GAGHAHAGAY	0.11
HRAGAGAGAGY	0.11
GAGAWWGAGAY	0.12
GAHHEEHHGAY	0.12
GAHHGAGAHGY	0.12
GAHHGAGHAGY	0.12
GAHHGAGHAGY	0.12
HHHGAGAGAGY	0.12
AHHSASGASGASGHHY	0.13
ASGAHHSASGASGHHY	0.13
ASGASGAHHSASGHHY	0.13
ASGASGASGAHHSGHHY	0.13
ASGASGASHHGASGHHY	0.13
ASGASGHHSASGASGHHY	0.13
ASGASHHGASGASGHHY	0.13
ASGHHSASGASGASGHHY	0.13
ASHHGASGASGASGHHY	0.13
EEEEHHHEEY	0.13
EEEEHHHEEY	0.13
EEHHHEEY	0.13
ESEHHHESEY	0.13
GAHHEGHHGAY	0.13
GAHHGHAGAGY	0.13
HHGAGAGAGRY	0.13
RRHHGGHHEEY	0.13
ASGASGASGAHHGHY	0.14
ASGASGASGHASGHHY	0.14
EEHHHEEKY	0.14
ESHHRSHESY	0.14
GAGAHHHGAG	0.14
GAGAHHHGAG	0.14
GAGAHHHGAGY	0.14
GAGAHHHGAGY	0.14
GAGARWWGAGY	0.14
GAGHHWGAGAY	0.14
GAHHHHHGAY	0.14
GAHLLHHGAY	0.14
GAQHHAHHQAY	0.14
HHAGAGAGRRY	0.14
HHASGASGASGASGHHY	0.14
GAHHERHHGAY	0.15
GAHHGAGHHAY	0.15
GAHHGAHHGAY	0.15
GAHHGAHHGAY	0.15
GAHHGGHHGAY	0.15
GAHHGGHHGAY	0.15
GAHLLGHGAY	0.15
GAHHSGHGAY	0.15
HASGAHSGASHGASGHY	0.15
HASGHASGHASGHASGY	0.15
HASGHASGHASGHASGY	0.15
HHASGASGASGASGHHY	0.15
HRHGAGAGAGY	0.15
KEHHHEEY	0.15

Peptide	Im concentration, M
REEHHHHEEY	0.15
EKKHHHKEEY	0.16
GAGHHAHHGAY	0.16
GAGHHHHGAY	0.16
GAGHHLHHGAY	0.16
GAGHHMHHGAY	0.16
GAGHHPHHGAY	0.16
GAGHHQHHGAY	0.16
GAHHFGHHGAY	0.16
GAHHGHGAGAY	0.16
GANHHAHHNAY	0.16
GHHAGAGHHAY	0.16
KDHHHHDDDY	0.16
AAHHHHAADY	0.17
DDHHHHDDDY	0.17
DDHHHHDDDY	0.17
FFLHHHHESEY	0.17
GAGHHFHHGAY	0.17
GAGHHGHHGAY	0.17
GAGHHKHHGAY	0.17
GAGHHNHHGAY	0.17
GAHHAHHGAGY	0.17
GAHHFFHHGAY	0.17
HAHSHGHASGASGASY	0.17
HASGASGASGASGHHY	0.17
HHHHAGAGAGY	0.17
HRHRAGAGAGY	0.17
KKHHHHDDDY	0.17
KKKHHHHEEY	0.17
KSKHHHHESEY	0.17
REEHHHHEERY	0.17
RSHHESHHSY	0.17
AAHHHHAAAY	0.18
AAKHHHHAADY	0.18
AAKHHHHAAY	0.18
AAKHHHHADAY	0.18
AAKHHHHAAY	0.18
AAKHHHDAAY	0.18
AAKHHHHAAY	0.18
AKAHHHHAADY	0.18
DDHHHHDDKY	0.18
DDKHHHKKDY	0.18
EEGASGASGASGHHHY	0.18
EKKHHHKEKY	0.18
FFLHHHHSKY	0.18
GAGHRRHHGAY	0.18
HASGASGASGASGHHY	0.18
KAHHHHHAAY	0.18
KKHHHHEEY	0.18
PPHHHHPPPY	0.18
RDDHHHHDDDY	0.18
RREESGASGASGHHHY	0.18
RREHHHHEEY	0.18
SSHHHHSSSY	0.18
AAHHHHAAY	0.19
AARHHHHADAY	0.19
AARHHHDAAY	0.19
AKAHHHHAAY	0.19
ASGASGASGASGHHHY	0.19
EERHHHHEEY	0.19
GAGAHHHGA	0.19
GAGAHHHHGAY	0.19
GAGAHHHGAY	0.19
GAGAHHHHGAY	0.19
GAGAHHHHGAY	0.19
GAHRRHHGAY	0.19
HHHGAGAGRRY	0.19
HHHHHSGESPSYFTA	0.19
ILVHHHHVLIY	0.19
KEHHHHEKKY	0.19
KKHHHHDDKY	0.19
KKHHHDKDY	0.19
KKHHHKKDY	0.19
RAAHHHHAADY	0.19
RDDHHHDDRY	0.19

Sens Actuators B Chem. Author manuscript; available in PMC 2007 October 16.

Peptide	Im concentration, M
HHHHHSGRSPRYLTA	0.23
HHHHHSGRSPSNFTA	0.23
KKHHHHEKKY	0.23
KSKHHHHSKY	0.23
RRRHHHHEEY	0.23
AARHHHHARAY	0.24
AARHHHHRAAY	0.24
ARAHHHHAARY	0.24
GAGHHHHGA	0.24
GAGHHHHGAY	0.24
HHHHHHAGAGY	0.24
HHHHHSGEALWNFTA	0.24
HHHHHSGEAPRRLTA	0.24
HHHHHSGRALRYLTA	0.24
HHHHHSGRALSYFTA	0.24
HHHHHSGRAPRYFTA	0.24
HHHHHSGRAPWNFTA	0.24
HHHHHSGRSLRYLTA	0.24
HHHHHSGRSPSRLTA	0.24
HHHHHSGRSPWNFTA	0.24
HRHRHRAGAGY	0.24
RAAHHHHAARY	0.24
AARHHHHHAARY	0.25
AARHHHHARAY	0.25
AARHHHHRAAY	0.25
HHHHHSGEALRNFTA	0.25
HHHHHSGEALRYFTA	0.25
HHHHHSGEALWRLTA	0.25
HHHHHSGEAPWYFTA	0.25
HHHHHSGRAPRNFTA	0.25
HHHHHSGRAPSRFTA	0.25
HHHHHSGRAPWRLTA	0.25
HHHHHSGRSLRNFTA	0.25
HHHHHSGRSLRYFTA	0.25
HHHHHSGRSLWNFTA	0.25
HHHHHSGRSPRNFTA	0.25
HHHHHSGRSPSRFTA	0.25
HHHHHSGRSPWRLTA	0.25
KKKHHHKKKY	0.25
RAAHHHHAARY	0.25
RERHHHHRERY	0.25
RRRHHHDDR	0.25
RRRHHHRRDY	0.25
ARAHHHHAARY	0.26
FFLHHHHSRY	0.26
HHHAGAGHHY	0.26
HHHAGAGHHGY	0.26
HHHAGAGHHGY	0.26
HHHAGAGHHGY	0.26
HHHAGAGHHGY	0.26
HHHHHSGESLWYFTA	0.26
HHHHHSGRALRYFTA	0.26
HHHHHSGRSLRNFTA	0.26
HHHHHSGRSLSRFTA	0.26
HHHHHSGRSLWRLTA	0.26
HHHHHSGRSPRRLTA	0.26
HHHHHSGRSPWYFTA	0.26
RREHHHHERRY	0.26
RRRHHHDDR	0.26
FFLHHHHLFFY	0.27
HHHGHGAGY	0.27
HHHHHSGRALWRLTA	0.27
RRRHHHHEERY	0.27
RRRHHHHEERY	0.27
RRRHHHHEERY	0.27
GAHHHHHHGA	0.28
GAHHHHHHGAY	0.28
HHHHHSGEALWYFTA	0.28
HHHHHSGRALSRFTA	0.28
HHHHHSGRAPWYFTA	0.28
HHHHHSGRSLWRLTA	0.28
HHHHHSGRSLWYFTA	0.28
HHHHHSGRSPRRFTA	0.28
HHHHHSGRSPWYFTA	0.28
RDRHHHHRDY	0.28
RRDHHHDDR	0.28

Peptide	Im concentration, M
DRRHHHHRDY	0.29
HHHHHHHGAGY	0.29
KKKHHHKKKY	0.29
HHHHHSGRALRRFTA	0.3
RSRHHHRSRY	0.3
ERRHHHRREY	0.31
HHHHHSGRALWYFTA	0.31
RRRHHHDDRY	0.32
GHHHHHHGAY	0.33
HHHHHHHGAGY	0.33
RRRHHHHERRY	0.34
HHHHHSGRSPWRFTA	0.35
GHHHHHHHAY	0.37
RRRHHHKKKY	0.37
RRRHHHRRRY	0.39
HHHHHHHHAY	0.41
HHHHHHHHHY	0.46
RRRHHHRRRY	0.46

Table 2

Specifications of the SVM

Type	Polynomial
Order	2
Margin	Soft
C	10
Epsilon	0.01