BIOINFORMATICS

Vol. 19 no. 13 2003, pages 1650–1655
DOI: 10.1093/bioinformatics/btg223

Secondary structure prediction with support vector machines

*J. J. Ward, L. J. McGuffin, B. F. Buxton and D. T. Jones**

Bioinformatics Group, Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK

Received on November 25, 2002; revised on February 6, 2003; accepted on February 10, 2003

ABSTRACT

Motivation: A new method that uses support vector machines (SVMs) to predict protein secondary structure is described and evaluated. The study is designed to develop a reliable prediction method using an alternative technique and to investigate the applicability of SVMs to this type of bioinformatics problem.

Methods: Binary SVMs are trained to discriminate between two structural classes. The binary classifiers are combined in several ways to predict multi-class secondary structure.

Results: The average three-state prediction accuracy per protein (Q_3) is estimated by cross-validation to be 77.07 ± 0.26% with a segment overlap (Sov) score of 73.32 ± 0.39%. The SVM performs similarly to the 'state-of-the-art' PSIPRED prediction method on a non-homologous test set of 121 proteins despite being trained on substantially fewer examples. A simple consensus of the SVM, PSIPRED and PROFsec achieves significantly higher prediction accuracy than the individual methods.

Availability: The SVM classifier is available from the authors. Work is in progress to make the method available on-line and to integrate the SVM predictions into the PSIPRED server.

Contact: dtj@cs.ucl.ac.uk

INTRODUCTION

The computational complexity of predicting the full 3D conformation of proteins has stimulated the development of knowledge-based approaches that solve simple intermediate problems such as the prediction of solvent accessibility and secondary structure. These predictions are often core elements of methods that recognize the folds of proteins that have negligible sequence identity to existing structures (McGuffin and Jones, 2003). Secondary structure prediction has been tackled using numerous learning algorithms including multi-layer perceptrons and recurrent neural networks, and therefore represents a useful problem for testing the effectiveness of new techniques. Support vector machines (SVMs) have shown promising results on several biological pattern classification problems and are becoming established as a standard tool in bioinformatics. SVMs have been successfully applied to recognition of protein translation–initiation sites in DNA sequences (Zien *et al*., 2000) and functional annotation of genes from expression profiles (Brown *et al*., 2000).

SVMs perform well compared with other learning algorithms because they are effective in controlling the classifier's capacity and the associated potential for overfitting. This is achieved by ensuring that the decision boundary separating two classes does so with a large margin (Christianini and Shaw-Taylor, 2000; Vapnik, 1998). SVMs have other desirable properties such as efficient solutions, relatively few adjustable parameters and the interchangeable use of kernel functions, which define a mapping of the data to a higher-dimensional feature space.

Since the publication of the PHD prediction method in the early 1990s (Rost and Sander, 1993), three-state accuracies (Q_3) have continued to rise to their current peak of around 77%. The main reason for the increased accuracy is that modern methods incorporate evolutionary information from a larger number of homologues to the target protein. One source of this improvement is the continual expansion of the sequence databases (Rost, 2001b) and the other is the use of scoring matrices from iterated PSI-BLAST searches (Altschul *et al*., 1997) or hidden Markov model homology searches (Karplus *et al*., 1998), which recover information from more remote homologous sequences (Rost, 2001a).

SVMs have previously been shown to predict secondary structure (Hua and Sun, 2001) from multiple alignments with slightly greater accuracy than the early PHD prediction method. The highest quoted Q_3 score of 73.5% and the fairly modest improvement over PHD suggests that the SVM classifier, described in (Hua and Sun, 2001), is not competitive with the best modern methods. The main reason for the improved Q_3 score of the new method is that the SVM is trained on PSI-BLAST profiles from searches carried out over the latest sequence databases. The SVM also uses a quadratic kernel function, which maps the data to a feature space that can represent pair interactions between features, increasing the accuracy of helix prediction. The other sources of improvement are that probability estimates are used to combine binary classifiers and to successfully train a second structure-to-structure classifier.

*To whom correspondence should be addressed.

Bioinformatics 19(13) © Oxford University Press 2003; all rights reserved.

The SVM is benchmarked against the PSIPRED secondary stucture prediction method (Jones, 1999), which is an expertly tuned feed-forward neural network that is consistently ranked amongst the best methods available (Rost and Eyrich, 2001).

SYSTEM AND METHODS

The training set was constructed by clustering a set of proteins with solved crystal structures, such that no two proteins had >25% sequence identity between clusters. A representative structure with the highest resolution was then selected from each cluster and placed in the training set totalling 1460 non-redundant proteins. This set is a comprehensive sampling of structure space and is also suitable for estimating the accuracy of the prediction system using cross-validation (Rost and Eyrich, 2001).

The inputs from each sequence appear in the form of a $20 \times M$ position-specific scoring matrix from three iterations of a PSI-BLAST search, where M is the length of the target sequence. The scoring matrix for a window of 15 positions, centred on the target residue, is used as the input to the SVM. In cases where the window extends beyond the protein termini, 'empty' attributes are filled with zeros. The inclusion of extra inputs to indicate the ends did not alter the performance of the SVM and were therefore omitted.

The structural class of each residue was obtained by performing a standard reduction of the eight states outputted by the DSSP program (Kabsch and Sander, 1983) to the helix (H), strand (E) and the default coil (C) states. This study uses a reduction scheme whereby α-helix and 3_{10}-helix are converted to H, sheet and isolated β-bridge to E and the rest (π-helix, turn, bend and other) to C. This reduction allows comparison with PSIPRED and most other prediction methods listed on the EVA server (Rost and Eyrich, 2001).

Three-state predictions are made by combining the outputs of several binary SVMs. The 'one-versus-rest' and the 'one-versus-all' methods are two standard approaches for combining outputs from binary classifiers into a multiple-class prediction (Hsu and Lin, 2002). The 'one-versus-rest' method constructs a binary classifier for each class. Each SVM is trained on all the examples with those in class i given positive labels and examples belonging to all other classes given a negative labelling. An example is then assigned to the class with the largest functional output. The 'one-versus-one' approaches construct classifiers for the $\binom{k}{2} = k(k-1)/2$ possible class pairings with each classifier trained on the subset of the examples belonging to the two classes. In testing, the outputs are combined by casting a vote for the 'winner' of each pair-wise comparison and assigning the example to the class with the most votes.

This study also investigates other methods for combining outputs from binary classifiers that use estimates of the posterior probability rather than the functional output. An acknowledged deficiency of SVMs is that they do not provide estimates of the posterior probability of class membership, in contrast to many neural networks. The heuristic probability estimates developed by Platt (1999) provide a reasonable approximation in this case. The outputs of the SVM $f(\mathbf{x})$ are mapped to posterior probabilities using a logistic sigmoid function

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(Af(\mathbf{x}) + B)} \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ is the d-dimensional feature vector representing each example and y is a binary class label. The parameters A and B were found by maximum likelihood estimation. This function is monotone in the magnitude of the functional output and maps the uncalibrated SVM outputs to the range $[0, 1]$. These estimates can be used to assign an example to the class with the highest posterior probability for the 'one-versus-rest' classifier or using the pair-wise coupling technique for a 'one-versus-one' classifier (Hastie and Tibshirani, 1998). The pair-wise coupling method models the output of the classifier trained on classes i and j by

$$P(\text{class } i|\text{class } i \text{ or class } j) = \frac{p_i}{p_i + p_j} \tag{2}$$

where p_i and p_j are the probabilities of the example being in classes i and j, respectively. The three-state probabilities are estimated by minimizing the cross-entropy between the model and the binary estimates of posterior probability from each classifier.

DESIGN

The publicly available *SVMlight* software was used to train several binary SVMs (Joachims, 1999). The LOQO quadratic program solver was used because of its faster convergence for problems with a high error rate (Burges, 1998). The performance of several kernels was investigated using a 300 protein training set and a 75 protein validation set. The order 2 polynomial kernel was found to have the highest validation accuracy.

$$K(\mathbf{x}, \mathbf{z}) = \left(\frac{\mathbf{x} \cdot \mathbf{z} + 1}{50}\right)^2 \tag{3}$$

where $\mathbf{x}$ and $\mathbf{z}$ are the input vectors for two examples. This kernel function maps the data to a feature space that includes the d input attributes $(x_1, x_2, \ldots, x_d)$ and the $\binom{d+1}{2}$ degree 2 monomials $(x_1^2, x_1x_2, \ldots, x_d^2)$. This allows the SVM to represent periodic patterns that characterize helices, such as high scores for alanine in the central residue and glutamic acid at the $n+4$ position. The constant constrains $K(\mathbf{x}, \mathbf{z})$ to the range $[-1, 1]$. The optimum value for the parameter, C, controlling the trade-off between low training error and large margin was found to be 0.5 for the quadratic kernel. These parameters appear to be optimal for all six binary classifiers shown in Table 1.

Table 1. The percentage of the training set that form support vectors and accuracy on the test set (the above random column shows the SVM's improvement over the trivial prediction)

Classifier	SVs (at upper bound)	Accuracy	Above random
C/¬C	55.0 (48.8)	77.7	20.9
H/¬H	40.9 (34.9)	86.4	19.8
E/¬E	36.5 (30.4)	85.6	9.8
C/H	46.1 (39.5)	84.2	30.1
C/E	48.5 (40.7)	81.3	20.3
H/E	36.0 (29.6)	88.0	34.3

Table 2. Table shows the three-state accuracy Q_3 (Rost and Sander, 1993) and segment overlap (Sov) scores (Zemla *et al*., 1999) for a test set of 75 proteins

	Q_3	Sov	C_C	C_H	C_E
(a) Sequence–structure					
One-versus-rest SVM	74.54	68.24	0.55	0.68	0.59
Maxprob SVM	74.52	68.45	0.55	0.68	0.59
One-versus-one SVM	74.04	67.11	0.54	0.68	0.58
Pair-wise couple SVM	74.24	68.34	0.54	0.68	0.59
Neural network	73.28	64.65	0.54	0.66	0.58
(b) Structure–structure					
Pair-wise couple SVM	75.44	70.74	0.55	0.70	0.60
Neural network	74.72	68.92	0.55	0.69	0.59

C_X are the Matthew's correlation co-efficients for coil, helix and strand. (a) Shows results of binary sequence-to-structure SVMs and comparison with neural network. (b) Shows the results for the SVM and neural network structure-to-structure classifiers. In both (a) and (b), the Q_3 and Sov-scores for the SVMs are significantly higher than the comparable neural network (at the 95% level) according to a paired t-test.

The number of support vectors is extremely high for all six binary classifiers. This is undesirable because the fraction of the training examples that become support vectors places an upper bound on the generalization error rate (Christianini and Shaw-Taylor, 2000). It also affects the time complexity, in testing, which scales linearly with the number of support vectors. The high empirical error is a characteristic of secondary structure prediction caused by noise in the evolutionary profiles and some ambiguity in the structure assignments. This high error leads to a large fraction of the data set becoming bounded support vectors. The examples are also not independently and identically distributed (i.i.d), with the feature vector for a particular residue closely resembling a rotation of the adjacent positions in the protein chain. Correlations between examples have been shown to increase the number of support vectors (Burges and Schölkopf, 1997), as transformations of an initial set of support vectors generates more.

The accuracy of 'one-versus-rest' binary classifiers suggest that coil and helix states are more easily discriminated than sheet, contradicting results from balanced training of neural networks (Rost and Sander, 2000). This may be a consequence of the helix and coil formation being influenced by predominantly local interactions whereas strands are formed between two strings of complementary residues that may be distantly separated in the protein sequence. Longer range interactions are not fully encoded by local windows although the profiles can represent the evolutionary constraints on each residue, which may indicate structural constraints in the globular protein.

The results of several methods for combining binary classifiers are shown in Table 2. These classifiers are compared to a feed-forward neural network with a similar architecture to PSIPRED. The network has two layers of adaptive weights, a hidden layer with 70 units and an output node for each of the three structural classes (Bishop, 1995). This network was trained using batch resilient back-propagation (Riedmiller and Braun, 1993) with early stopping.

The relatively low segment overlap score of all the methods is a characteristic of prediction methods that treat each residue as an i.i.d example and is caused by occasional inconsistencies in the middle of structural elements. This study follows others (Rost and Sander, 1993; Jones, 1999) in training a second classifier on the outputs to filter the first set of predictions, as shown below for a fragment of protein 1qkr(A).

classifier 1: `HHCECEHECHHHCC`

classifier 2: `HHHHEEEEEHHHCC`

The structure-to-structure SVMs were trained on the outputs of another window of 15 residues with additional binary features to indicate the protein termini. The pair-wise coupled probability estimates were used as they give better prediction accuracy than training on raw SVM outputs. The isotropic Gaussian kernel was found to have lowest error on the validation set.

$$K(\mathbf{x}, \mathbf{z}) = \exp(-\gamma||\mathbf{x} - \mathbf{z}||)^2 \quad (4)$$

The kernel and regularization parameters ($\gamma = 1.28$ and $C = 0.2$) were found on the training set using the leave-one-out model selection (looms) program (Lee and Lin, 2000). The results for two cascaded pair-wise coupled SVMs, in Table 2, show that the SVM continues to outperform a comparable neural network. The filtering neural net is trained on the outputs of the first network and has 60 hidden units. Both filters have improved accuracy and segment overlap, with the correlation co-efficients also indicating improved prediction of helices and strands.

RESULTS

The final prediction method includes three 'one-versus-one' binary SVMs with the quadratic kernel for sequence-to-structure classification. This allows training to be carried out on the full set of 1460 proteins because the memory requirements are lower than the 'one-versus-rest' class assignments.

Table 3. Results from 3-fold cross-validation of the final SVM prediction method on a data set of 1095 proteins

	H	E	C	
(a)				
obs(helix)	80.40	3.31	16.29	
obs(sheet)	4.76	68.75	26.50	
obs(coil)	10.63	10.15	79.22	
(b)				
pred(helix)	83.93	4.97	11.10	
pred(sheet)	4.03	83.62	12.34	
pred(coil)	13.35	21.71	64.93	
(c) Q_3	Sov	C_H	C_E	C_C
$77.07 \pm 0.26\%$	$73.32 \pm 0.39\%$	0.725	0.634	0.585

(a) Shows the SVM's assignment of the observed structural classes with diagonal entries representing the per residue Q_X^{obs} scores for each structure type. (b) Shows the true class assignments of the predictions with diagonal entries indicating the Q_X^{pred} scores. (c) Shows the mean Q_3 and Sov scores per protein. The confidence interval is given by $\sigma/\sqrt{n}$, where n is the number of protein sequences. C_X represents Matthew's correlation co-efficients for helix, sheet and coil.

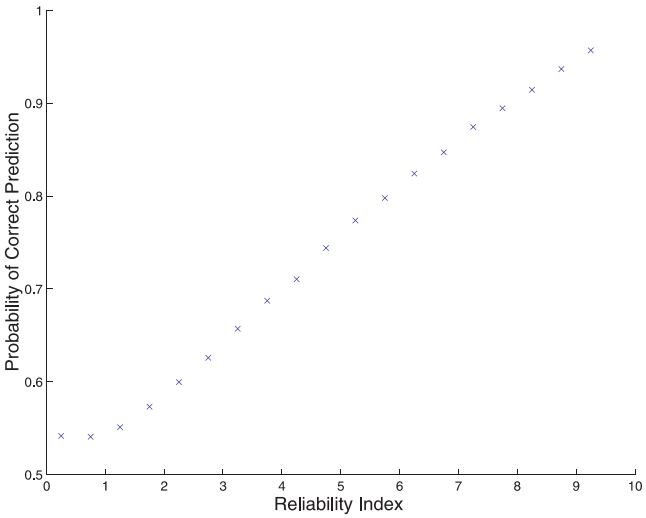

Fig. 1. Reliability index for cross-validation set of 1095 proteins against posterior probability for bins of width 0.5.

```
sequence:    -----EEEE----E--EE--EE--------EEEE-HHHEEE-
prediction:  -E---EEEEE------EE--EEEEEE----EEEE--EE-EE-
errors:       O       L   U        LLLL         LWWL
```

Fig. 2. First two rows show target and prediction, with dashes representing coil structures. The four types of errors are shown in the lower row. These are length errors, which occur at the ends of correctly predicted helix or sheet elements; over-predictions, where a coil segment is predicted as helix or strand; under-predictions, where helix or strand segments are predicted as coil; and wrong predictions, where strand is misclassified as helix and *vice versa*.

The 'one-versus-one' approach is also roughly an order of magnitude faster to train and the frequencies of each class are more evenly balanced. The outputs were mapped to probabilities and then inputted to a second classifier comprising binary structure-to-structure SVMs with the Gaussian kernel. The error matrices of the cross-validated predictions are shown in Table 3. The results are shown for the 1095 structures without unresolved chain breaks out of the full set of 1460 proteins.

The SVM's Q_X^{obs} and Q_X^{pred} scores are quite similar to those of several other methods with comparable accuracies that are listed on the EVA server (Rost and Eyrich, 2001). The main differences appear to be that the SVM is more accurate in predicting helices but is slightly more conservative in predicting sheet residues. The outputs from the first set of binary classifiers give better estimates of the probability of a correct prediction than the outputs of the structure-to-structure classifiers. These estimates can be used to calculate a reliability index, as shown in Figure 1, which can be used to indicate the regions of a protein where the classifier has high confidence.

Benchmarking of secondary structure methods is complicated by the great variation in accuracy between test proteins and the continual improvement in profiles resulting from the expansion of the sequence databases. As a result, an objective comparison of two methods can only be made on the same test set and with both methods having access to the same sequence database. This was achieved by benchmarking the SVM against PSIPRED on a set of 121 proteins released between January and June 2002. This set was filtered to remove any sequences that are homologous to any in PSIPRED's training set (the SVM is trained on a subset) and each other. The cut-off was set at a PSI-BLAST expectation score of 0.1 with sequences below that threshold considered to have significant similarity. The inputs were constructed using PSI-BLAST profiles from the same databases using identical search parameters. The results are shown in Table 4 along with those from PROFsec (Rost and Eyrich, 2001) and a consensus of the three methods using a simple voting scheme similar to that used by JPRED (Cuff and Barton, 1999). Table 4 also shows a breakdown of the type of errors made by each prediction method, as shown in Figure 2. This breakdown was found to

Table 4. Accuracy scores for the SVM method compared to PSIPRED, PROFsec and a consensus of these methods on a test set of 121 proteins

	Q_3	Sov	Length	Over	Under	Wrong
SVM	74.92	70.48	16.63	1.74	4.43	2.22
PSIPRED	74.97	71.81	16.37	1.85	4.30	2.50
PROFsec	74.90	71.15	16.27	1.63	4.88	2.28
Consensus	76.17	72.37	15.75	1.74	4.08	2.23

None of the differences in the scores between the SVM, PSIPRED and PROFsec are statistically significant (at the 95% level) according to a paired sample t-test. The differences in Q_3, Sov and length error scores between the consensus and the other three methods are statistically significant according to the same criterion.

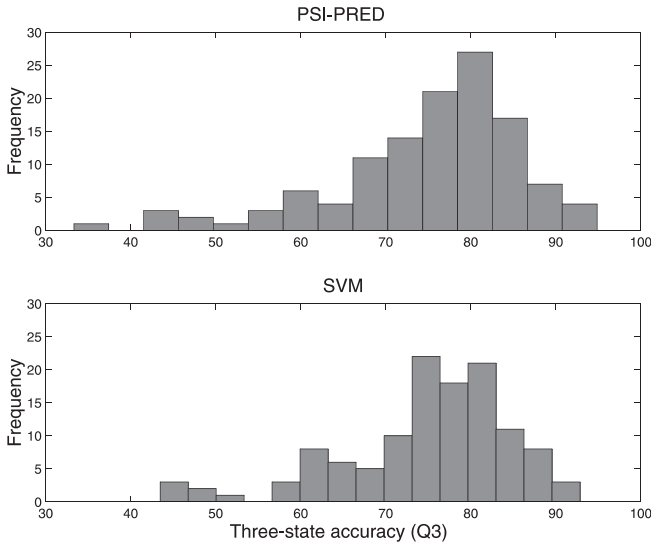

Fig. 3. Histogram of Q_3 scores per protein chain for PSIPRED and SVM for set of 121 test proteins.

be useful for benchmarking, as under-predictions and wrong predictions appear to be more detrimental to fold recognition using alignments of predicted secondary structure than other types of errors (McGuffin and Jones, 2003).

The SVM's error rates are significantly higher for this benchmark set of proteins than those estimated by cross-validation. However, both PSIPRED and PROFsec also achieve accuracies that are far lower than their overall averages and the difference appears to be due to the nature of this particular test set. The similarity of the scores and the Q_3 distributions for these methods (Fig. 3) suggest that the SVM has similar prediction accuracy to PSIPRED and the other highest ranking methods on the EVA server such as PROFsec and SSpro, since common sets of approximately 100 proteins have been sufficient to demonstrate a significant improvement over older methods such as PHD (Rost and Sander, 1993). The SVM appears to have achieved similar prediction accuracy to PSIPRED despite being trained on 1460 rather than 5000 structures.

The consensus achieves a Q_3 score that is significantly higher than each individual method according to a paired t-test. This demonstrates that a consensus of the highest-accuracy methods continues to improve on individual predictions, even in the current era of structure prediction based on PSI-BLAST profiles. Most of the improvement occurs in correctly predicting the ends of structural elements, although under-predictions may also be reduced.

DISCUSSION

The support vector machine has demonstrated a similar accuracy to PSIPRED on the benchmark set of proteins, with a smaller training set and without use of ensemble averaging from several networks. The SVM also attains greater accuracy than a neural network when both classifiers are trained on small data sets but this advantage is lost as the number of training examples is increased. The linear growth in the number of support vectors with the size of the training set coupled with the high empirical error of secondary structure prediction means that the final method has a very large number of support vectors, which occupy around 680 MB of memory. A single prediction therefore requires 1.6×10^8 multiply-adds from the sequence-to-structure classifier alone and the SVM is not yet a viable alternative to PSIPRED for implementation as a web service. Although, there is significant redundancy in the set of support vectors and exact simplification of the decision surface (Downs *et al.*, 2001) would be likely to yield a classifier with acceptable classification rates. Alternatively, the SVM could be re-trained on a smaller training set without greatly compromising generalization.

Other experiments have indicated that there is no advantage in using SVMs for the filtering structure-to-structure classifier. A hybrid method incorporating a neural network would reduce computational complexity and may also improve prediction accuracies, since the multiple-state classification can be made in a single step. Further improvements of the SVM classifier may be achieved through design of a kernel function that incorporates domain knowledge of how structure is conserved between homologous proteins or that makes effective use of information from outside a relatively narrow input window. Consensus predictions have also been shown to improve upon individual methods (Cuff and Barton, 1999) and a system that makes use of the most accurate modern methods may provide further gains in accuracy.

An acknowledged deficiency of SVMs is that the uncalibrated outputs do not provide estimates of posterior probability of class membership. This study has shown that a combination of the probability measures of Platt (1999) with the pair-wise coupling technique (Hastie and Tibshirani, 1998) can be used to indicate confidence successfully and may be applicable to other areas of biological pattern recognition.

Recent efforts have been made to predict the eight structural states assigned by DSSP using recurrent neural networks (Pollastri *et al.*, 2002) and this may be the direction of future developments in secondary structure prediction. Investigations, using the SVM, suggest that α- and 3_{10}-helices can be distinguished using evolutionary profiles, although this is complicated by the very unequal frequencies of the sub-divided structure classes. In addition, the SVM has some success in recognizing parallel and anti-parallel strands, and these predictions could benefit fold recognition methods that incorporate secondary structure predictions.

ACKNOWLEDGEMENTS

This work has been supported by the MRC (JJW) and the BBSRC (LJM).

REFERENCES

Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Bishop,C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.

Brown,M., Grundy,W., Lin,D., Christianini,N., Sugnet,C., Furey,T., Ares,J. and Haussler,D. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl Acad. Sci. USA*, **97**, 262–267.

Burges,C.J.C. (1998) A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**, 121–167.

Burges,C.J.C. and Schölkopf,B. (1997) Improving the accuracy and speed of support vector machines. *Adv. Neural Inform. Process. Syst.*, **9**.

Christianini,N. and Shaw-Taylor,J. (2000) *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge.

Cuff,J.A. and Barton,G.J. (1999) Evalution and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins*, **35**, 508–519.

Downs,T., Gates,K. and Masters,A. (2001) Exact simplification of support vector solutions. *J. Mach. Learn. Res.*, **2**, 293–297.

Hastie,T. and Tibshirani,R. (1998) Classification by pairwise coupling. In Jordan,M.I., Kearns,M.J. and Solla,S.A. (eds.), *Advances in Neural Information Processing Systems*, The MIT Press, Cambridge, MA, Vol. 10.

Hsu,C.-W. and Lin,C.-J. (2002) A comparison on methods for multi-class support vector machines. *IEEE Trans. Neural Networks*, **13**, 415–425.

Hua,S. and Sun,Z. (2001) A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. *J. Mol. Biol.*, **308**, 397–407.

Joachims,T. (1999) Making large-scale SVM learning practical. In Schölkopf,B., Burges,C. and Smola,A. (eds.), *Advances in Kernel Methods—Support Vector Learning*.

Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 196–202.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Karplus,K., Barrett,C. and Hughey,R. (1998) Hidden markov models for detecting remote protein homologies. *Bioinformatics*, **14**, 846–856.

Lee,J.-H. and Lin,C.-J. (2000) Automatic model selection for support vector machines. Technical report, Department of Computer Science and Information Engineering, National Taiwan University.

McGuffin,L.J. and Jones,D.T. (2003) Benchworking secondary structure prediction for fold recognition. *Proteins*, **52**, 166–175.

Platt,J.C. (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Smola,A., Bartlett,P., Schölkopf,B. and Shuurmans,D. (eds.), *Advances in Large Margin Classifiers*. MIT Press, Cambridge, MA.

Pollastri,G., Przybylski,D., Rost,B. and Baldi,P. (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins*, **47**, 228–235.

Riedmiller,M. and Braun,H. (1993) A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proc. of the IEEE Intl. Conf. on Neural Networks*, pp. 586–591.

Rost,B. (2001a). Protein structure prediction continues to rise. *J. Struct. Biol.*, **134**, 204–218.

Rost,B. (2001b). Twilight zone of protein sequence alignments. *Protein Eng.*, **134**, 204–218.

Rost,B. and Eyrich,V.A. (2001) EVA: large-scale analysis of secondary structure prediction. *Proteins*, **5**, 192–199.

Rost,B. and Sander,C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.*, **232**, 584–599.

Rost,B. and Sander,C. (2000) Third generation prediction of secondary structures. In Webster,D. (ed.), *Methods in Molecular Biology*, Humana Press in., Totowa, NJ, pp. 71–96.

Vapnik,V. (1998) *Statistical Learning Theory*. John Wiley and Sons, New York.

Zemla,A., Venclovas,C., Fidelis,K. and Rost,B. (1999) A modified definition of sov, a segment-based measure for protein secondary structure prediction assessment. *Proteins*, **34**, 220–223.

Zien,A., Rätsch,G., Mika,S., Schölkopf,B., Lengauer,T. and Müller,K.-R. (2000) Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, **16**, 799–807.