Epothilones: Quantitative Structure Activity Relations Studied by Support Vector Machines and Artificial Neural Networks

Annalen Bleckmann and Jens Meiler*

University of Washington, Department of Biochemistry, BOX 357350, Seattle, WA 98195, USA

Full Paper

In this paper the relation between the structure of epothilones (a new class of anti-tumour agents) and their potential to influence the tubulin-microtubule equilibrium is investigated. Insights into the character of the tubulin-epothilone interactions are derived as the accuracy and reliability of support vector machines and artificial neural networks to model such relations quantitatively is compared. Both methods are well qualified to model relationships between the structure of epothilone derivatives and their anti-tumour activities. Artificial neural networks achieve lower residual standard deviations (22%) compared to support vector machines (25%) and better classification results (89% compared to 75%). However, the reproducibility of the results is greater for support vector machines, which suggests a stronger convergence. The mapping of the influence of individual structural descriptors on the three-dimensional epothilone structure suggests one side of the rather flat molecule to be more important for its activity. The "LIBSVM" software which is used for simulating the support vector machines is freely available from www.csie.ntu.edu.tw/~cjlin/libsvm. The Program "Smart" which is used for simulating artificial neural networks is free for academic use and can be obtained together with the database of epothilones and their activities from www.jens-meiler.de.

1 Introduction

Epothilones A (Figure 1) and B can be isolated from the myxobacterium *Sorangium cellulosum* strain 90 [1, 2]. The recognition of their cytotoxic action against tumor cells led to intense research activities in chemistry and biology. Bollag et al. [3] discovered the induction of the tubulin-polymerization (TP) similar to agents like Taxol [4]. The effect of microtubule stabilization even in taxol-resistant tumor cell lines [5] increased their potential in cancer chemotherapy further [3, 6, 7].

The complete elucidation of the structure including stereochemistry was published by Höfle [8]. Soon after, the synthesis of epothilone, analogues and precursors was described in a large number of reports [9–16]. The biological activity of these analogues was investigated. These data served as basis of qualitative structure activity relations [10, 17–19]. Moreover, Wang et al. [20] introduced a unified and quantitative receptor model for microtubule binding of paclitaxel and epothilone by docking 26 derivatives into a "minireceptor".

After publishing tubulin structures at 6.5 Å and 3.7 Å resolution, the structure of αβ-tubulin stabilized with taxol was refined to 3.5 Å resolution utilizing electron crystallography [21–23]. The tubulin-bound conformation of epothilone A [24] which was solved by NMR-spectroscopy shows significant differences in comparison to the x-ray structure of free epothilone [8]. The macrocycle undergoes conforma-

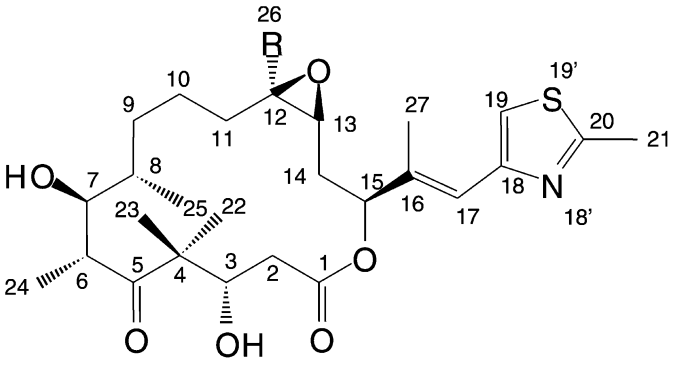


Figure 1. Two dimensional structure of epothilone A.

* To receive all correspondence: Jens Meiler, University of Washington, Department of Biochemistry, Box 357350, Seattle, Washington 98195-7350, USA, e-mail: jens@jens-meiler.de, Phone: +1 206 543 7134, Fax: +1 206 685 1792

Key words: support vector machines; artificial neural networks; quantitative structure activity relationship; epothilones; tubulin

 DOI: 10.1002/qsar.200330837

tional changes which become most obvious in the change of the C13-C14-C15-C16 (Figure 1, Figure 5) torsion angle by ~100°. More importantly, the side chain dihedral angle C16-C17-C18-C19 (Figure 1, Figure 5) changes from *trans* to *cis* conformation, which results in having the electron pair on the nitrogen freely accessible for interactions.

In recent years the potential of support vector machines (SVM) [25, 26] for distinguishing biologically active from non-active substances has been investigated [27]. SVMs use distinct mathematical functions to transform a given set of descriptors into a different hyperspace. The transformation is optimized to separate a set of data points according to a given and usually binary property by introducing a hyperplane in this new hyperspace. Slightly modified versions are also able to handle classifications in more than two groups or the fit of analog functions. The latter case is a specialization of three-layer artificial neural networks (ANN) with a restricted set of transformations possible.

ANNs are used for several years in chemistry and biochemistry to describe structure activity relations quantitatively [28]. Their critical advantage is the flexibility of the model. It can adapt to complex interrelations and is capable to detect even small signals at a large noise level. Hence ANNs are applied when no simple mathematical model can be assumed, many potential parameters interact and the experimental errors are high.

The major difference between ANNs and SVMs is the training procedure. ANNs are trained by minimizing the prediction error for the training set of data. The training stops in a local minimum. This is not critical, if all local minima are close to the global minimum as usually for big large sets of training data. However, if the set of data becomes small as it frequently happens with biological data, the local minima tend to represent substantially different models although their overall accuracy is close. In contrast to ANNs, SVMs offer a well-defined decision which of these local minima is to be chosen – the one that ensures the greatest separation between positives and negatives, e.g. active and inactive drugs. So the somewhat random decision which of the minima is adopted depending on the starting values of the weights when training an ANN is replaced by a well defined algorithm. Thus, while for ANNs often dissimilar models are obtained after retraining the same set of data several times from different randomized starting points, the SVM obtained for one and the same set of data will be the same.

In this paper, ANNs and SVMs are applied to the same set of epothilone structures and biological activities to compare the two methods in terms of accuracy and reliability. General conclusions about the epothilone-tubulin-interactions were obtained.

2 Materials and Methods

All biological activities used in the following analysis are taken from Nicolaou et al. and Altmann et al. [10, 16, 29]. Epothilone derivatives that did not match the template structure of epothilone A (Figure 1) were excluded (e.g. open macrocycle, missing side chain at C15). These structures were exclusively inactive. Hence only 223 structures out of ~250 initial structures were used for the analysis. The activity of these substances [10, 16, 29] was assessed by incubating purified tubulin for 30 min at 37 °C in the presence of the compound. The mixture is filtered and the collected polymerized tubulin is stained with amido black solution. Quantification is yielded by measuring the absorbance of the dyed solution. The given%-polymerization (TP) is calculated relative to the presumed absorbance of 100% polymerized tubulin. The values lay in a range between 1% and 98%.

In addition biological activities were assessed for 37 derivatives by measuring the inhibition of carcinoma cell growth as IC_{50}-values for four cell lines[10]. The cell lines include a parental ovarian cell line (1A9), two of its mutants that are taxol resistant (1A9PTX10: Phe270→Val and 1A9PTX22: Ala364→Thr), and a mutated MCF7 breast-cell line. The cell growth is evaluated by measuring the increase in cellular protein. The resistant cell lines are gained by treating the cells with increasing concentrations of taxol. However, their taxol resistance does not cause a resistance to epothilone, which further increases the potential of epothilones in cancer chemotherapy. Finally one more biological parameter, an EC_{50}-value obtained from a quantitative glutamate assay, is defined as the concentration of the drug that reduces the concentration of the protein in the solution to 50% [10, 16, 29].

In order to analyze these data with ANNs or SVMs, a numerical representation for the ligand structure is needed. 25 parameters (compare Figure 2) were derived that code the 223 structures uniquely. P01 describes the stereochemistry on atom C3. P02 distinguishes whether the two carbon atoms at C4 are methyl groups or connected by a single bond to form a three membered ring. The stereochemistry at C6 and C7 is always alternate and can therefore be coded with a single parameter P03. P04 codes the four possible substituent configurations on C8 (Me;H), (H;H), (H;Me), (Me;Me). The number of CH_2-groups between C8 and C12 (between 1 and 5) is described by P05. P06 codes the bond type between C12 and C13 (single or double). The stereochemistry of C12 and C13 is described by P07 and P08. The large variety of substituents found on C12 is described by P09-P12 giving their volume, average electronegativity, valence electrons and π-electrons. The substituent on C13 (H, OH, F) is described by P13. The stereochemistry at C15 is coded by P14. The structure of the substituent on C15 was varied over a wide range. To describe this structural space with a small number of parameters, predicted chemical shift values [30] have proven to be efficient. They include information about the covalent as well as the electronical structure in only one number per atom. Chemical shift values for C16-C18 were used as P15-P17. The atom types at positions 18', 19, 19', 20 are described by P18-P21 (C, N, O, S), while P22-P25

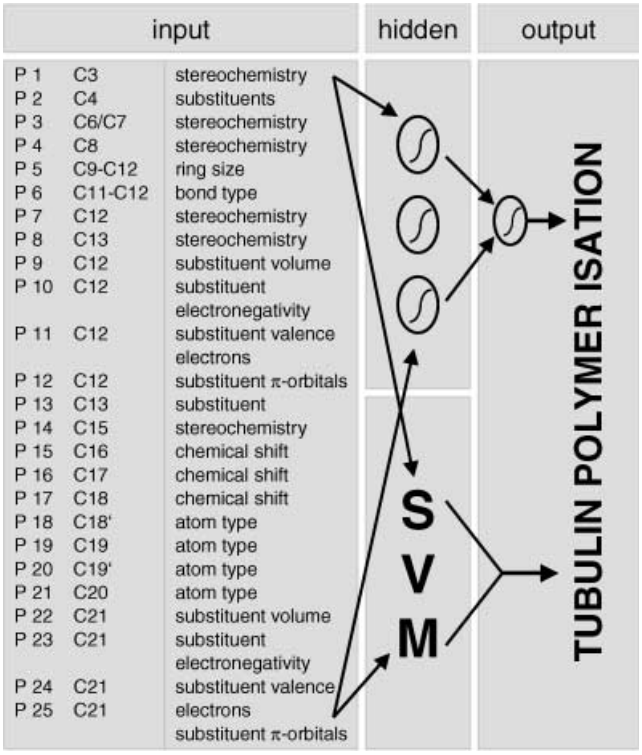


Figure 2. Scheme of data processing. A chemical structure is translated by 25 parameters into a numerical code. This code is applied to ANNs and to SVMs. Both are trained to predict a certain biological activity.

describe the substituent of C21 in the same manner as applied for the substituent on C12 (volume, average electronegativity, valence and π-electrons). Eleven of these 25 parameters (P01, P02, P03, P06, P07, P08, P14, P18, P19, P20, and P21) are binary, P04, P05, and P13 are multi-state and the remaining 11 parameters (P09-P12, P15-P17, and P22-P25) are real numbers. To train ANNs with published inhibition values for carcinoma cell growth constants the number of epothilone structures with known values reduces to 37, which limits the number of necessary parameters to 8.

The SVMs were established with the LIBSVM software (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). The SVMs were trained by setting the type of SVM to epsilon-SVM-regression. The type of kernel function was chosen to be a radial basis function. The gamma in the kernel function (radial basis function) was set to 0.0005. The parameter c of the epsilon-SVM-regression was set to 10. To get a binary output, the used type of SVM was C-SVM-classification.

All ANNs were generated with the program "Smart" [30]. The three-layer networks contain one bias in every layer and are trained with the back propagation algorithm. The learning rate η is decreased during the training procedure from 0.1 to 0.001, the momentum α is constant with 0.5. The input sensitivity is defined to be the first derivative of the output with respect to every single input $s_{ij} = \frac{d_{out_i}}{d_{inp_j}}$. For obtaining classification results ANN were trained to predict a binary output value. Both setups were optimized parameter-wise to give the best results for the monitoring data (see below).

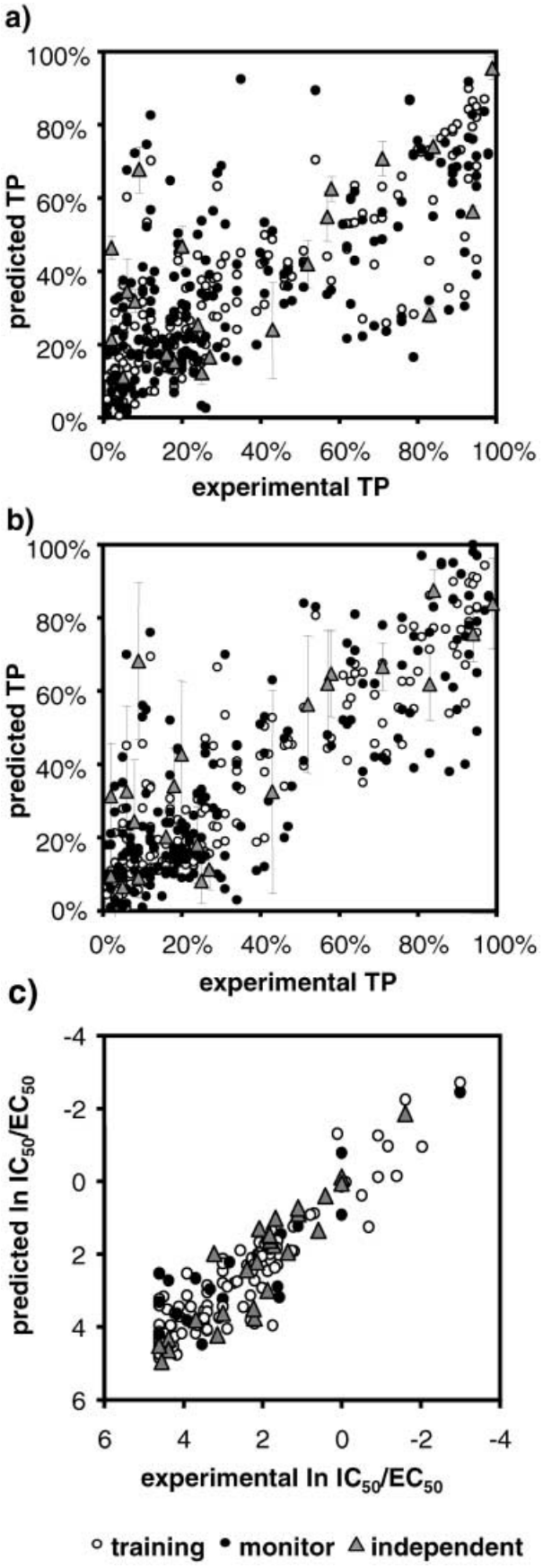
3 Results and Discussion

200 epothilone structures were used to create 10 sets of data for cross-validation, each consisting of 180 structures for training und 20 structures for monitoring. The 23 remaining structures form the independent set of data.

The following three experiments were performed: 1) Prediction of tubulin polymerization (TP): ANNs and SVMs were trained for each of the ten sets of data. The net architecture was 25 inputs, four hidden neurons and one neuron in the output layer. The number of weights amounts to 109. The results for all trained ANNs and SVMs are given in table 1. The residual standard deviations for predicting TP for the independent set of data are 25.1% ± 0.7% when using the ten SVMs and 22.2% ± 2.2% for the ten ANNs. While the correlation is only slightly better for the ANNs when looking at the independent set of data, the remarkable convergence of the SVMs becomes clear from the low standard deviation. In Figure 3 a) and b) the standard deviations are marked for the ten predictions of the 23 data points in the independent set of data. It is striking that SVMs suggest in all ten cases rather similar good models while ANNs suggest in average slightly better but always different models. 2) Classification: In a second experiment all epothilone structures with a TP constant that exceeds 50% were counted as active the remaining as inactive. The training of ANNs and SVMs was repeated under those binary conditions. The relation between active and non active structures was 61 : 162. While the SVMs have a success

Table 1. Standard deviations and classification results of predicted TP activities

		Training data	Monitoring data	Independent data
SVM	std. dev.	18.0% ± 3.9%	22.6% ± 2.6%	25.1% ± 0.7%
	classification	86.1% ± 1.2%	79.5% ± 8.5%	74.8% ± 4.0%
ANN	std. dev.	14.6% ± 1.5%	17.3% ± 2.7%	22.2% ± 2.2%
	classification	91.7% ± 1.8%	87.5% ± 5.4%	89.1% ± 3.1%
NN6	TP	12.7%	11.3%	13.9%
	$\ln(IC_{50}/EC_{50})$	7.1%	6.2%	6.5%



◀ **Figure 3.** Results of experiments I, II and III. The correlation diagrams for SVMs (a) and ANNs (b) with the experimental induction of tubulin-polymerization data on the x-axis and the predicted value on the y-axis for all training, monitoring and independent sets of data. Diagram (c) shows the experimental LN of IC_{50}-values and EC_{50}-values on the y-axis and the predicted LN of IC_{50}-values and EC_{50}-values the x-axis. The line of same experimental data results from substances, whose activities are all greater than 100.

rate of 74.8% ± 4.0%, ANNs achieve values of 89.1 ± 3.1% (table 1). While for both models around 10% of only the active substances are classified as inactive, the rate of false positive is much higher for SVMs (close to 50%) compared to ANN's (around 12%). 3) Prediction of carcinoma cell growth inhibition: For the last experiment four values for carcinoma cell growth inhibition and two values for induction of the TP are predicted by a single ANN. Because of the small available set of data (only 37 structures having at least two of these values determined) only one training, one monitoring, and one independent set of data were defined. The training set of data consists of 28, the monitoring set of data of four and the independent set of data of five epothilone structures. However, since several experimental data points per structure are available, the overall number of data points amounts to 137 for the training set of data, 20 for the monitoring set of data and 32 for the independent set of data. The number of weights in the ANN is 66 having 8 inputs, 4 hidden neurons and 6 output neurons to predict the TP, EC_{50} (TP) and four IC_{50} values for carcinoma cell growth inhibition. ANNs have the advantage, that they are able to predict several output values in parallel. The result profits from this parallel prediction, especially if dependent output values for small data sets are available. The usual drawback in applying ANNs on small set of data (a hidden layer too small to describe a complex non-linear dependence because of the limited number of degrees of freedoms allowed) can be avoided: in our case, the number of training data points is multiplied by six, which allows a sufficient broad hidden layer to train the network. While the number of available data points was too small to establish ANNs predicting individual IC_{50}, a combined ANN that predicts six values for every structure enhances the model immediately. The standard deviation values for the TP are 13.9% for the independent dataset, whereas the EC_{50}- and the IC_{50}-values achieve standard deviations of 6.5%.

To assess the importance of individual input parameters for predicting the activity of a structure, the so-called input sensitivity was computed. The input-sensitivities of each parameter as obtained from the ten ANNs from experiment I are shown in Figure 4. Aside from the stereochemistry at C12, the ANN seems to utilize all provided parameters for computing the activity. The high influence of the ring size between C9 and C11 likely results from a change in the conformation of the macrocycle rather than a direct

Table 2. Detailed classification results including the fraction of false and true positives and negatives

		Training data		Monitoring data		Independent data	
		exp. active	exp. inactive	exp. active	exp. inactive	exp. active	exp. inactive
SVM	pred. active	23.2%	10.8%	20.5%	14.5%	30.9%	21.3%
	pred. inactive	3.1%	62.8%	6.0%	59.0%	3.9%	43.9%
ANN	pred. active	20.9%	2.9%	18.5%	4.5%	30.9%	7.0%
	pred. inactive	5.4%	70.8%	8.0%	69.0%	3.9%	58.2%

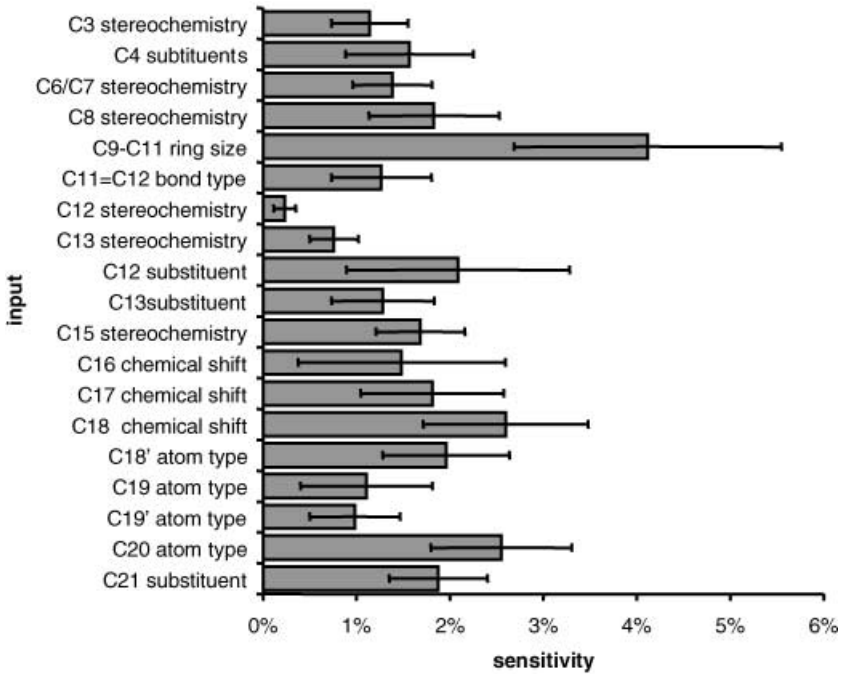


Figure 4. Diagram of input-sensitivity for all parameters. The input sensitivity is given on the x-axis. The parameters as introduced in figure 2 are given on the y-axis. The error bars indicate the obtained standard deviation over the ten cross validation set of data.

interaction with the protein. The presence of the two hetero atoms in the correct position in the side chain is critical for the activity of epothilone. The strong interaction of this part of the molecule with the protein is also shown by the high sensitivity values obtained for the chemical shift values and the substituent at C21. As already seen by the increased activity of epothilone B compared to epothilone A a methyl-group at C12 increases the activity. This observation agrees with the high sensitivity of the relevant parameters. Also critical are the OH-group at C3 and the methyl-group on C8 that have to point into the plane of Figure 1 in order to increase activity.

These sensitivities can now be mapped onto the volume contour of epothilone in its tubulin-bound conformation to visualize the potential binding site. In Figure 5 a high sensitivity is indicated by red regions and a low sensitivity by blue regions. Comparing the two sides of the molecule it is striking that descriptors with high influence on the activity of epothilone tend to cluster on the b-side of the molecule. The substituents at C21 and C12 as well as the nitrogen in position 22, the methyl groups at C4, the OH at C7 point in this direction. The 180° turn obtained in the side chain when going from the unbound to bound conformation of epothilone does rotate the nitrogen atom as well as the substituent at C21 on this side of the molecule. Since the absence of the nitrogen causes epothilone to loose its activity, it can be suggested that an interaction between this nitrogen and the protein (e.g. via a hydrogen bond) exists. Hence, the b-side of the molecule is suggested to play the major role in the epothilone-tubulin interaction.

4 Conclusions

SVMs as well as ANNs are well qualified to establish quantitative relations between the chemical structure of epothilone derivatives and their potential to induce the tubulin polymerization and in turn their ability to inhibit the

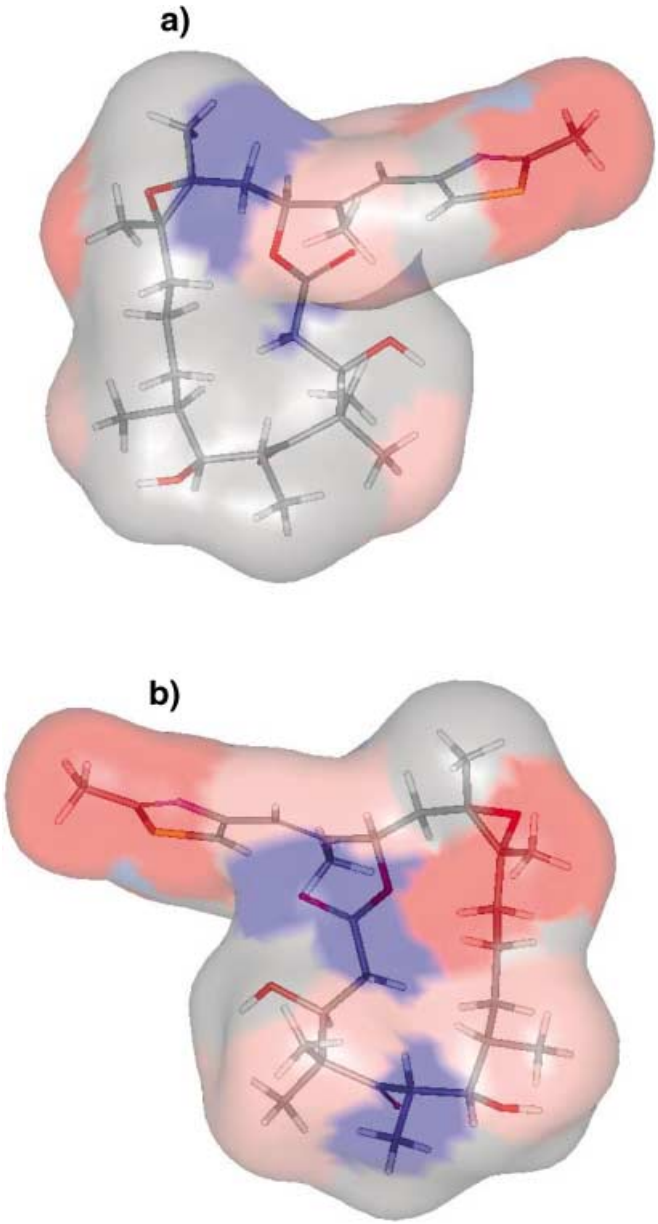


Figure 5. Plot of input-sensitivities on a three dimensional structure of epothilone. The surface colours are related to the sensitivity of this region. Most important regions for activity are shown in red and the least important in blue. The most active regions cluster on one side of the molecule (b). Interaction with tubulin will most likely occur on this side while the second side (a) seems to be less involved in interactions.

growth of cancer cells. In predicting the induction of tubulin polymerization for an independent set of data SVMs achieve standard deviations of 25%, while in using ANNs values of 22% are obtained. For classification, SVMs achieve a success rate of 75%, whereas ANNs are successful in 89% of the cases. Hence ANNs get lower standard deviations and significantly better classification results. In contrast to ANNs, a much higher convergence is observed for SVMs. Their high reproducibility is a clear advantage. In predicting the induction of tubulin polymerization, EC_{50}-values for the tubulin polymerization and four IC_{50}-values for carcinoma cell growth inhibition parallel, ANNs achieve standard deviations of 14% for the induction of tubulin polymerization, whereas the EC_{50}-value and the four IC_{50}-values yield standard deviations of 7%. An analysis of the input sensitivity of the trained ANNs allows mapping the influence of every structural feature of epothilone on its three-dimensional structure. The network finds the substituents at C12 and C21, stereochemistry at C3, C6, C7 and C8, methyl groups at C4 and most importantly the presence of a nitrogen atom in the aromatic ring to be critical for the activity of epothilone derivatives. Interestingly, those features cluster on one side of the three dimensional epothilone structure which suggests this side to be in contact with tubulin.

Acknowledgement

The authors would like to thank Christian Griesinger for useful discussion. We thank Bill Wedemeyer and Phil Bradley for careful reading the manuscript. J.M. is supported by a Human Frontier Science Program fellowship.

References

[1] G. Höfle, N. Bedorf, K. Gerth, Epothilone derivatives, *Chemical Abstracts* **1994**, *120*, 836.
[2] K. Gerth, N. Bedorf, G. Höfle, H. Irschik, H. Reichenbach, Epothilons A and B: Antifungal and Cytotoxic Compounds from Sorangium cellulosum (Myxobacteria) – Production, Physico-chemical and Biological Properties, *J. Antibiot.* **1997**, *49*, 560 – 563.
[3] D. M. Bollag, P. A. McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides, C. M. Woods, Epothilones, a New Class of Microtubulue-stabilizing Agents with a Taxol-like Mechanism of Action, *Cancer Res.* **1995**, *55*, 2325 – 2333.
[4] P. B. Schiff, J. Fant, S. B. Horwitz, Promotion of microtubule assembly in vitro by taxol, *Nature* **1979**, *277*, 665 – 668.
[5] R. J. Kowalski, P. Giannakakous, E. Hamel, Activities of the Microtubule-stabilizing Agents Epothilone A and B with Purified and in Cells Resistant to Paclitaxel (Taxol), *J. Biol. Chem.* **1997**, *272*, 2534 – 2541.
[6] P. Giannakakou, R. Gussio, E. Nogales, K. H. Downing, D. Zaharevitz, B. Bollbuck, G. Poy, D. Sackett, K. C. Nicolaou, T. Fojo, A common pharmacophore for epothilone and taxanes: molecular basis for drug resistance conferred by tubulin mutations in human cancer cells, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2904 – 2909.
[7] L. A. Martello, H. M. McDaid, D. L. Regl, C.-P. H. Yang, D. Meng, T. R. R. Pettus, M. D. Kaufman, H. Arimoto, S. J. Danishefsky, A. B. Smith, III, S. B. Horwitz, Taxol and discodermolide represent a synergistic drug combination in human carcinoma cell lines, *Clin. Cancer Res.* **2000**, *6*, 1978 – 1987.
[8] G. Höfle, N. Bedorf, H. Steinmetz, H. Reichenbach, K. Gerth, Epthilon A und B – neuartige, 16gliedrige Makrolide mit

cytotoxischer Wirkung: Isolierung, Struktur im Kristall und Konformation in Lösung, *Angew. Chem.* **1996**, *108*, 1671 – 1673.
[9] D. Schinzer, A. Limberg, O. M. Böhm, Studies Towards the Total Synthesis of Epothilones: Asymmetric Synthesis of the Key Fragments, *Chem. Eur. J.* **1996**, *2*, 1477 – 1482.
[10] K. C. Nicolaou, F. Roschangar, D. Vourloumis, Chemie und Biologie der Epothilone, *Angew. Chem.* **1998**, *110*, 2120 – 2153.
[11] E. Von Angerer, Tubulin as a target for anticancer drugs, *Curr. Opin. Drug Discovery Dev.* **2000**, *3*, 575 – 584.
[12] K. C. Nicolaou, R. Scarpelli, B. Bollbuck, B. Werschkun, M. M. A. Pereira, M. Wartmann, K. H. Altmann, D. Zaharevitz, R. Gussio, P. Giannakakou, Chemical synthesis and biological properties of pyridine epothilones, *Chem. Biol.* **2000**, *7*, 593 – 599.
[13] J. Mulzer, Epothilone B and its derivatives as novel antitumor drugs: total and partial synthesis and biological evaluation, *Monatsh. Chem.* **2000**, *131*, 205 – 238.
[14] C. B. Lee, T.-C. Chou, X.-G. Zhang, Z.-G. Wang, S. D. Kuduk, M. D. Chappell, S. J. Stachel, S. J. Danishefsky, Total Synthesis and Antitumor Activity of 12,13-Desoxyepothilone F: An Unexpected Solvolysis Problem at C15, Mediated by Remote Substitution at C21, *J. Org. Chem.* **2000**, *65*, 6525 – 6533.
[15] J. Johnson, S.-H. Kim, M. Bifano, J. DiMarco, C. Fairchild, J. Gougoutas, F. Lee, B. Long, J. Tokarski, G. Vite, Synthesis, Structure Proof, and Biological Activity of Epothilone Cyclopropanes, *Org. Lett.* **2000**, *2*, 1537 – 1540.
[16] K.-H. Altmann, G. Bold, G. Caravatti, N. End, A. Florsheimer, V. Guagnano, T. O'Reilly, M. Wartmann, Epothilones and their analogs – potential new weapons in the fight against cancer, *Chimia* **2000**, *54*, 612 – 621.
[17] J. D. Winkler, P. H. Axelsen, A Model for the Taxol (Paclitaxel)/Epothilone Pharmacophore, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2963 – 2966.
[18] D.-S. Su, A. Balog, D. Meng, P. Bertinato, S. J. Danishefsky, Y.-H. Zheng, T.-C. Chou, L. He, S. B. Horwitz, Structure-activity relationships of the epothilones and the first in vivo comparison with paclitaxel, *Angew. Chem., Int. Ed.* **1997**, *36*, 2093 – 2096.
[19] L. He, P. G. Jagtap, D. G. I. Kingston, H.-J. Shen, G. A. Orr, S. B. Horwitz, A Common Pharmacophore for Taxol and the Epothilones Based on the Biological Activity of a Taxane Molecule Lacking a C-13 Side Chain, *Biochemistry* **2000**, *39*, 3972 – 3978.
[20] M. Wang, X. Xia, Y. Kim, D. Hwang, J. M. Jansen, M. Botta, D. C. Liotta, J. P. Snyder, A Unified and Quantitative Receptor Model for the Microtubule Binding of Paclitaxel and Epothilone, *Org. Lett.* **1999**, *1*, 43 – 46.
[21] E. Nogales, S. G. Wolf, I. A. Khan, R. F. Luduena, K. H. Downing, Structure of tubulin at 6.5A ad location of the taxol-binding site, *Nature* **1995**, *375*, 424 – 427.
[22] E. Nogales, S. G. Wolf, K. H. Downing, Structure of the α,β tubulin dimer by electron crystallography, *Nature* **1998**, *39*, 199 – 203.
[23] J. Lowe, H. Li, K. H. Downing, E. Nogales, Refined structure of alpha-beta-tubulin at 3.5 A resolution, *J. Mol. Biol.* **2001**, *313*, 1045.
[24] T. Carlomagno, M. J. J. Blommers, J. Meiler, W. Jahnke, T. Schupp, F. Petersen, D. Schinzer, K.-H. Altmann, C. Griesinger, The High-Resolution Solution Structure of Epothilone A Bound to Tubulin: An Understanding of the Structure-Activity Relationships for a Powerful Class of Antitumor Agents, *Angew. Chem.* **2003**, *42*, 2511 – 2515.
[25] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, **1998.**
[26] B. Schoelkopf, A. J. Smola, *Learning with Kernels*, The MIT Press, Cambridge, Massachusetts, **2002.**
[27] J. Weston, F. Perez-Cruz, O. Bousquet, O. Chapelle, A. Elisseeff, B. Schoelkopf, Feature selection and transduction for prediction of molecular bioactivity for drug design, *Bioinformatics* **2003**, *19*, 764 – 771.
[28] J. Zupan, J. Gasteiger, *Neural Networks for Chemists*, VCH Verlagsgesellschaft mbH, Weinheim, **1993.**
[29] K. H. Altmann, Epothilone B and its analogs – a new family of anticancer agents, *Mini Rev. Med. Chem.* **2003**, *3*, 149 – 158.
[30] J. Meiler, M. Will, R. Meusinger, Fast Determination of 13C-NMR Chemical Shifts Using Artificial Neural Networks, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1169 – 1176.

Received on June 11, 2003; Accepted on July 16, 2003