J Comput Aided Mol Des (2011) 25:339–347
DOI 10.1007/s10822-011-9424-2

Machine learning study for the prediction of transdermal peptide

Eunkyoung Jung · Seung-Hoon Choi · Nam Kyung Lee · Sang-Kee Kang · Yun-Jaie Choi · Jae-Min Shin · Kihang Choi · Dong Hyun Jung

Received: 26 October 2010 / Accepted: 21 March 2011 / Published online: 30 March 2011

Abstract In order to develop a computational method to rapidly evaluate transdermal peptides, we report approaches for predicting the transdermal activity of peptides on the basis of peptide sequence information using Artificial Neural Network (ANN), Partial Least Squares (PLS) and Support Vector Machine (SVM). We identified 269 transdermal peptides by the phage display technique and use them as the positive controls to develop and test machine learning models. Combinations of three descriptors with neural network architectures, the number of latent variables and the kernel functions are tried in training to make appropriate predictions. The capacity of models is evaluated by means of statistical indicators including sensitivity, specificity, and the area under the receiver operating characteristic curve (ROC score). In the ROC score-based comparison, three methods proved capable of providing a reasonable prediction of transdermal peptide. The best result is obtained by SVM model with a radial basis function and VHSE descriptors. The results indicate that it is possible to discriminate between transdermal peptides and random sequences using our models. We anticipate that our models will be applicable to prediction of transdermal peptide for large peptide database for facilitating efficient transdermal drug delivery through intact skin.

Keywords Machine learning · Artificial neural network · Partial least squares · Support vector machine · Transdermal peptide · ROC score · VHSE descriptor

E. Jung · S.-H. Choi · D. H. Jung (✉)
Insilicotech Co. Ltd., A-1101 Kolontripolis, 210, Geumgok-Dong, Bundang-Gu, Seongnam-Shi 463-943, Korea
e-mail: dhjung@insilicotech.co.kr

E. Jung
e-mail: jungek@insilicotech.co.kr

S.-H. Choi
e-mail: shchoi@insilicotech.co.kr

N. K. Lee · S.-K. Kang · Y.-J. Choi
School of Agriculture Biotechnology, Seoul National University, San 56-1, Shilim-Dong, Kwanak-gu, Seoul 151-742, Korea
e-mail: lnk025@snu.ac.kr

S.-K. Kang
e-mail: haman@unitel.co.kr

Y.-J. Choi
e-mail: cyjcow@snu.ac.kr

J.-M. Shin
SBScience Co. Ltd., Sung-Ok BD, Sunae-Dong, Bundang-Gu, Seongnam-Shi 463-825, Korea
e-mail: sbscience@gmail.com

K. Choi
Department of Chemistry, Korea University, 1, Anam-dong 5-Ga, Seongbuk-Gu, Seoul 136-701, Korea
e-mail: kchoi@korea.ac.kr

Introduction

Skin, the largest and most easily accessible organ of the body, provides a painless and compliant interface for systemic drug administration [1–3]. Therefore, transdermal delivery of drugs has been considered as a convenient route of administration for therapeutic drugs [4]. However, the delivery of large chemical drugs or macromolecules such as insulin across the skin barrier is difficult because transdermal delivery is largely controlled by the stratum corneum in the outer skin surface that only permits penetration of small, lipophilic drugs [5, 6]. There have been many studies on chemical enhancers (such as surfactants [7, 8] and fatty acid/esters [9–11]) and techniques

(electroporation [12], iontophoresis [13, 14], ultrasound [15] and microneedles [16, 17]) to increase the permeability and/or absorption rate of molecules with high molecular weight and low lipophilicity. Although various strategies have been proposed to overcome the impermeability of intact skin, transdermal delivery of macromolecules in therapeutics still remains a formidable task. This is partly due to the problems associated with chemical enhancers (such as skin irritation), and techniques (inconvenience or high cost of electrical apparatus), and partly due to the difficulties associated with macromolecules [6, 10].

Since phage display was known as an experimental method for the study of protein–protein and protein-peptide interactions, there have been many studies with phage display peptide libraries [18–20] to identify peptides that home selectively to the specific tissues or organs. In addition, several studies used in vivo phage display to identify peptides that transport across the intestinal mucosal barrier [21, 22]. Recently, Chen et al. [3] reported a study for transdermal protein delivery using a coadministered peptide which had been identified via phage display. In the study, the peptide coadministrated with insulin increased plasma insulin levels and, as a result, reduced blood glucose levels. These experimental methods are rather labor-intensive and not easily applicable to high-throughput screening and, therefore, it is important to develop more effective methods of screening large peptide libraries for their transdermal activities.

In this work, we propose machine learning models for predicting transdermal properties of peptides based on sequence information. Using a sequence set of phage-displayed peptides selected from in vivo experiments, we constructed machine learning models to screen transdermal peptides using various descriptors of peptide physicochemical properties.

Materials and methods

Preparation of skin-to-abdominal adipose tissue-targeting peptides

To identify peptide moieties targeting abdominal adipose tissue after transdermal penetration, an in vivo phage display screening was conducted as described in previous report [22]. Briefly, 2×10^{11} pfu of random phage-peptide library (Ph.D.-C7C: New England BioLabs, Beverly, MD, USA) was spread onto the abdominal skin surface (4 cm^2 of area) of anesthetized Wistar rats (10-week-old male: Samtako, Osan, Korea) after hair-trimming without skin damage. The rats were sacrificed by abdominal incision after 1 h of retention, then phage recombinants crossing the skin barrier were recovered from the abdominal adipose tissue and amplified by infection into *E. coli* ER 2738 (New England BioLabs, Beverly, MD, USA). The next rounds of phage display biopanning were conducted with newly amplified phage recombinants at each round consecutively to enhance the specificity of the selected peptide ligands to abdominal adipose tissue after skin penetration. After the third round of biopanning, total 269 peptide sequences were identified from randomly selected individual phage recombinants rescued from the abdominal adipose tissue as candidates of skin-to-abdominal adipose tissue-targeting peptide ligands. Overall procedure of in vivo transdermal phage display screening in this study was depicted in Fig. 1. All applicable national laws and Seoul National University policies regarding the care and use of laboratory animals were observed.

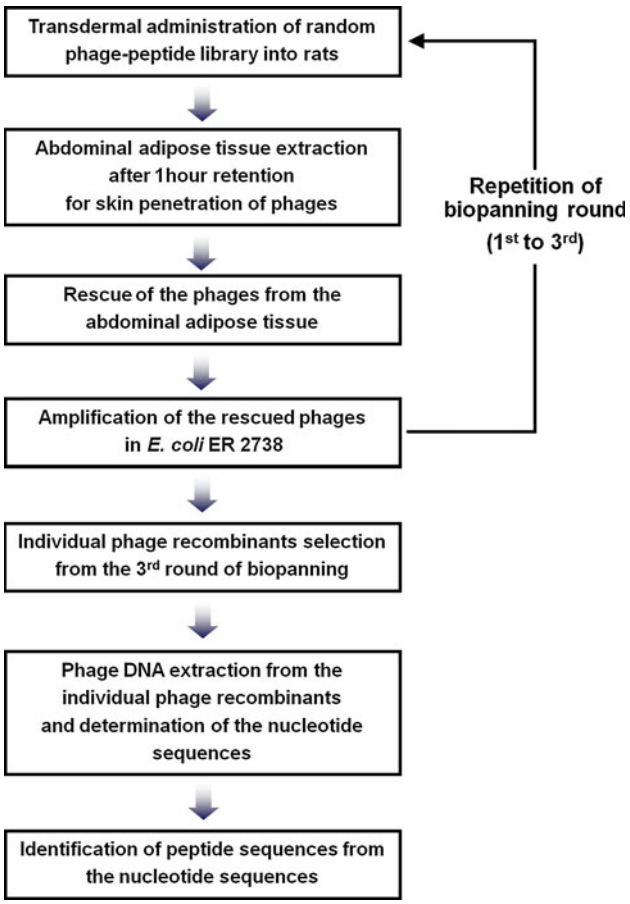

Fig. 1 Flow chart of the in vivo transdermal phage display screening procedure

Data sets

The positive control data set of peptides that can be delivered across skin barrier was obtained from transdermal heptapeptide sequences identified by in vivo phage display experiments as already described. Fig. 2 shows the

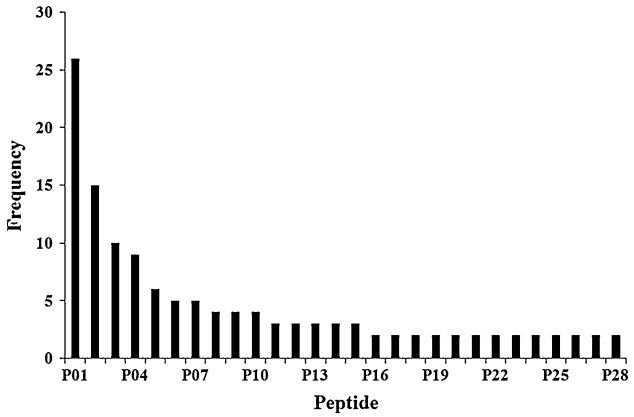

Fig. 2 The number of occurrences of redundant peptides found in the positive controls

number of occurrences of peptides that were found more than once in the positive control set. For example, peptide sequence P01 was found 26 times in the positive control set. The decoy negative control data set was generated from random sequences that had the same frequencies of occurrence of each amino acid residue in primary structures of unrelated proteins of known sequence [23]. The decoy sequences were then compared with the positive control data and any common sequences were removed from the negative control data. Sequences that are much similar to the positive control sequences were also removed because some sequences that have high sequence similarity with transdermal peptides might also cross the skin barrier. About 80% of each data set was used for network training and the remaining data were used for the test set to validate the trained network.

Descriptors

Three types of amino acid descriptors, VHSE, Z3 and Z5, were used to encode important features of individual peptide sequences. The VHSE descriptor is composed of eight variables for each amino acid describing hydrophobic, steric and electronic properties [24] and, therefore, each heptapeptide was encoded with $7 \times 8 = 56$ variables. The Z3 descriptor is composed of three z-scales describing hydrophobic, steric and electronic properties [25]. The Z5 descriptor is composed of five variables as the extended z-scales describing lipophilicity, size/polarizability, and electronic properties of each amino acid [26].

Artificial neural network (ANN) model

The calculations were carried out using the nnet of the VR 7.2 package [27] for feed-forward neural networks and for multinomial log-linear models. We used a three-layer neural network architecture containing a single hidden layer in which the number of neurons was increased from 0 to 2. The neural network architecture of the ANN models were set as described in our previous study [28]. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [29] was used as the optimization function. To help the optimization process and to avoid overfitting, the weight decay was set at 0.001 and five-fold cross-validations were performed [28]. The maximum number of iterations for network training was 50,000 and the other parameters were given as the default values set by the nnet of the VR 7.2 package. Before the learning network was applied, the input value of the positive control was 0.9 and that of the negative control was 0.1. All calculations were performed through Pipeline Pilot 7.5 [30].

Partial least squares (PLS) model

The calculations were carried out using the pls package implementing partial least squares regression (PLSR) and principal component regression (PCR) [27]. The number of latent variables to be used for building the model was set at 3, 5, 10 and 15. To improve the performance of the learned model on 'new' data, five-fold cross-validations were performed during the training. All calculations were performed through Pipeline Pilot 7.5 and data were preprocessed identically to the ANN models.

Support vector machine (SVM) model

The SVM models were implemented using the function svm of the e1071 package [27] and the workflow for all calculations was automated through Pipeline Pilot 7.5. The ε-regression technique, which ensures that training set residuals do not exceed ε by increasing the number of support as needed, was used for training with ε set to 0.1. The linear, polynomial and radial basis functions were examined as a kernel for mapping input space to feature space. For the polynomial kernel function, the degree (D) of polynomial was set to 2 or 3, and the constant term was set to 0. For all kernel functions, the kernel parameter γ was set to 1 over the number of variables (data dimension) and the other parameters were given as the default value set by the Learn R Support Vector Machine Model component of the Pipeline Pilot. To help the optimization process and to avoid over-fitting, five-fold cross-validations were performed. Data were preprocessed identically to the ANN models.

Evaluation and validation

To assess the predictive performance, the receiver operating characteristics (ROC) score, which is the area under the ROC curve [31], was used for each training and test set.

All the ROC scores reported here were generated from a leave-group-out cross-validation for real and decoy set. The sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV) and accuracy (Acc) were assessed for the models showing the best performance.

Finally, we performed validation tests on our models using a decoy set [28, 32]. The validation using a decoy set has turned out to be very informative to prove the statistical importance of the machine learning models [28].

Results

Using the phage display screening, we identified 269 transdermal heptapeptide sequences targeting abdominal adipose tissue after transdermal penetration. These transdermal peptides were used as the positive control set for further analysis. Nine ANN models, twelve PLS models and twelve SVM models were derived for the training data set by varying the peptide descriptor, the number of neurons in the single hidden layer, the number of latent variables and the kernel function. Table 1 shows the prediction accuracy of our models using VHSE, Z3 and Z5 descriptors. The ROC score was used as the primary yardstick of performance providing an overview of the possible cut-off levels for performance test. For the ANN model, several network architectures have been tested to find the best structure which provides high prediction accuracy of the output. As for the neural network architecture, the difference in the ROC scores between the training and test sets increased with the number of neurons in the hidden layer. This is presumably due to overfitting of the networks, as mentioned in our previous studies [28, 33]. Because higher complexity models use a large number of parameters, these models generally require greater amounts of data to minimize overfitting. Therefore, models with simple architecture would be more appropriate for predicting short peptides that can cross the skin barrier.

We employed PLS methods to build a linear model to avoid overfitting and a various number of latent variables have been tested to find the best prediction model. For each descriptor, the best performance was obtained when the number of latent variables was set at 10 and ROC score of the model with 10 latent variables was similar to ANN model with network architecture with zero neuron in hidden layer and one in output layer.

Several kernels have been examined to improve the predictive power of the SVM model. The models with a linear kernel showed small differences in the ROC scores between the training and test sets. In case of the SVM models with a polynomial kernel, the degree of polynomial was increased up to three but the quality of the models was not markedly improved. In terms of the ROC score and overfitting, the best result was obtained from the SVM model with a radial basis function. VHSE descriptors tend to produce slightly better training models than the others for the equal training method. The predictive features of the SVM model with a radial basis function are presented

Table 1 Prediction accuracy for machine-learning models

Descriptor	ANN[a]			PLS			SVM		
	No. of HN[b]	Training	Test	No. of latent variables	Training	Test	Kernel[c]	Training	Test
VHSE	0	0.914	0.861	3	0.891	0.838	Linear	0.890	0.839
	1	0.982	0.770	5	0.906	0.851	Polynomial (D = 2)	0.994	0.817
	2	0.998	0.804	10	0.910	0.851	Polynomial (D = 3)	1.000	0.897
				15	0.913	0.858	RBF	0.997	0.909
Z3	0	0.758	0.751	3	0.757	0.733	Linear	0.753	0.776
	1	0.879	0.770	5	0.758	0.749	Polynomial (D = 2)	0.947	0.666
	2	0.910	0.722	10	0.758	0.751	Polynomial (D = 3)	0.997	0.800
				15	0.758	0.751	RBF	0.977	0.838
Z5	0	0.806	0.784	3	0.789	0.794	Linear	0.798	0.780
	1	0.909	0.752	5	0.804	0.796	Polynomial (D = 2)	1.000	0.750
	2	0.979	0.750	10	0.806	0.783	Polynomial (D = 3)	0.982	0.854
				15	0.806	0.785	RBF	0.991	0.855

[a] The network architecture A–B–C indicates the total number of descriptors in an input layer, where A is (7, the sequence length of a peptide) × (the number of descriptors for each amino acid), B and C are the numbers of neurons in hidden and output layers, respectively. For instance, the network architecture (7 × 8)-0-1 specifies a model constructed with zero neuron in hidden layer and one in output layer using the VHSE descriptor. All the models have one neuron in output layer

[b] The number (B) of neurons in a hidden layer

[c] D indicates the degree of polynomial when the kernel parameter is set to polynomial and RBF indicates the Radial Basis Function

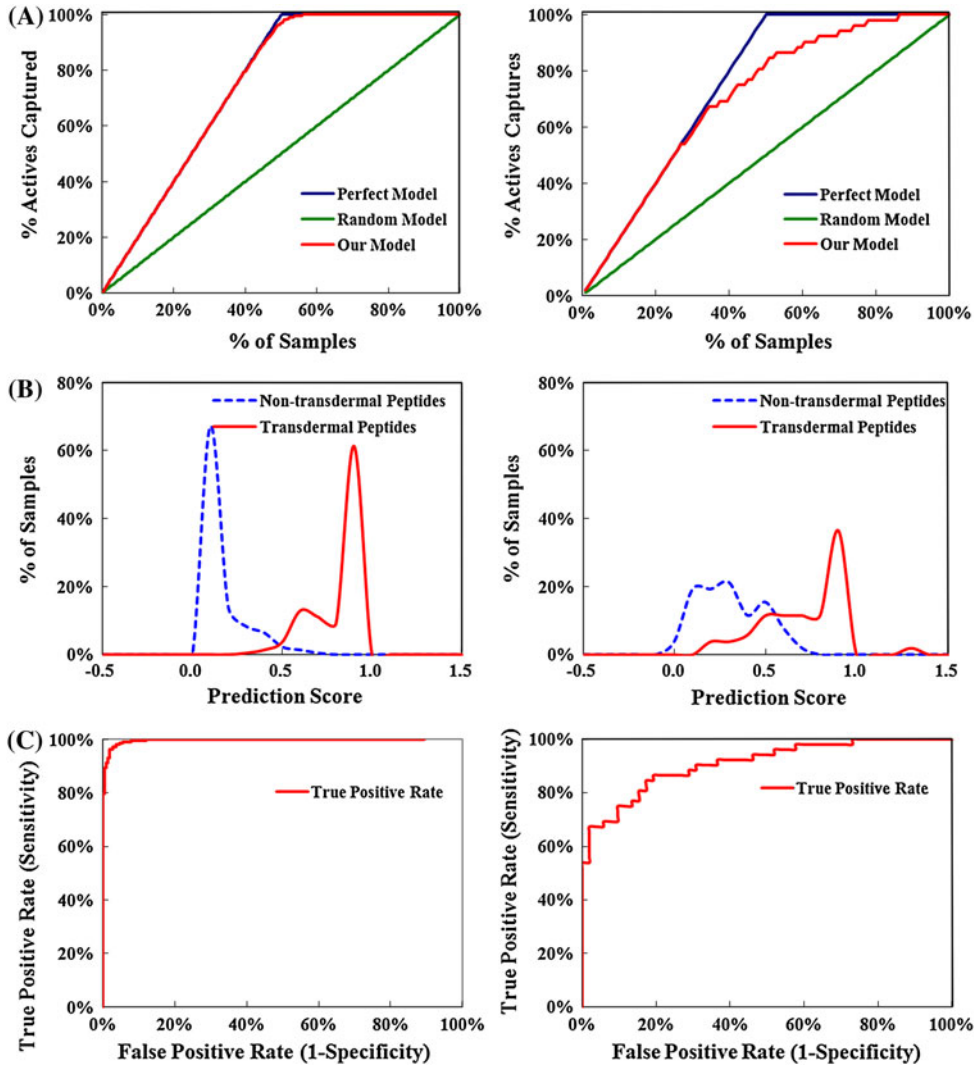

Fig. 3 Predictive features of the SVM model. The SVM model was constructed with a radial basis function using VHSE descriptor. **a** Enrichment curve, **b** histogram actives versus model values, and **c** receiver operating characteristic (ROC) curve. The features for the training and test set were plotted in the *left* and *right* panels, respectively

in Fig. 3 showing that the model can distinguish between transdermal and non-transdermal peptides effectively.

We performed leave-5%-out cross-validation to evaluate the stability of machine learning models in predicting transdermal peptides and the results are listed in Table 2. The SVM models, although not markedly different, were performed better than the others in average. For three machine learning methods, the standard deviation of the ROC scores is small for the training set but relatively large for the test set. The models using the VHSE descriptor showed considerably greater predictive power than the others for the test set.

To test the reliability of the peptide sequences defined as the positive control in the machine learning methods and to validate the strength of our models in predicting transdermal peptides, a separate decoy set was generated and then supplementary models trained with this decoy set were compared with the models trained with the real data set identified by the phage display experiment. The results suggest that the predictive power of the models constructed with the real set (Table 1) is considerably greater than that of the models trained with the decoy sequence (Table 3). In contrast to the model trained with the real data set, the SVM model constructed with the decoy set cannot discriminate between transdermal and non-transdermal peptides (Fig. 4). This result confirms that the peptides obtained from in vivo phage display have meaningful sequence patterns related with their transdermal activities, although the exact mechanism for the transdermal delivery of peptides is not fully understood yet.

The models with the VHSE descriptor showed the best performance for three machine learning methods and,

Table 2 The results (ROC scores) of validation using leave-5%-out method for machine-learning models

Descriptor	ANN[a]		PLS[b]		SVM[c]	
	Training	Test	Training	Test	Training	Test
VHSE	0.895 ± 0.004	0.780 ± 0.075	0.893 ± 0.004	0.777 ± 0.082	0.990 ± 0.001	0.819 ± 0.061
Z3	0.766 ± 0.004	0.685 ± 0.065	0.766 ± 0.004	0.685 ± 0.065	0.971 ± 0.002	0.776 ± 0.078
Z5	0.809 ± 0.005	0.702 ± 0.083	0.809 ± 0.005	0.701 ± 0.084	0.984 ± 0.001	0.767 ± 0.079

The results of rigorous test using leave-5%-out method. The results of 20 rigorous tests are averaged and expressed as mean ± standard deviation

[a] ANN models with network architecture with zero neuron in hidden layer and one in output layer

[b] PLS models with 10 latent variables

[c] SVM models with a radial basis function

Table 3 The results (ROC scores) of decoy analysis for machine-learning models

Descriptor	ANN[a]		PLS[b]		SVM[c]	
	Training	Test	Training	Test	Training	Test
VHSE	0.704	0.619	0.702	0.614	0.969	0.523
Z3	0.642	0.527	0.642	0.527	0.945	0.550
Z5	0.679	0.508	0.679	0.506	0.974	0.559

[a] ANN models with network architecture with zero neuron in hidden layer and one in output layer

[b] PLS models with 10 latent variables

[c] SVM models with a radial basis function

therefore, were selected for the truth table analysis of the binary outcome based on transdermal activity (Table 4). In comparison of sensitivity and specificity, our models seem to be sensitive in predicting transdermal peptides rather than specific in screening out non-transdermal peptides. Also, the analysis results show that the SVM model has higher overall prediction accuracy (76%) than the others (70% for ANN model and 71% for PLS model). To improve the prediction performance, we combined the three models and tried a consensus prediction in which the peptides predicted to be transdermal by three models are considered as transdermal and the other peptides are considered as non-transdermal. The positive predictive value (84%) of this combined model is much higher than those of the individual models and the accuracy of the prediction is also increased (Table 4).

Discussion

In a continuing effort to develop QSAR models for the analysis of peptide sequence information studies on [28, 33], we performed a machine learning study on transdermal peptides to predict and rank skin-permeability on the basis of peptide sequences.

We tried to develop QSAR models using the ANN, PLS and SVM methods. Like in our previous study [28], the VHSE descriptor tended to produce slightly better training models than the z scale descriptor. The models with the VHSE descriptor have also showed significant predictive power for intestinal barrier-permeability [33] and target delivery [28], and it seems that this descriptor is suitable for developing reliable sequence-based activity prediction models.

Three machine learning models using the VHSE descriptor showed good performance in predicting transdermal peptides although the standard deviation of ROC scores over the different test runs is slightly large. In order for a QSAR model to be used for the virtual screening of transdermal peptides, it must have high positive predictive value (PPV). The ANN, PLS and SVM models showing the best performance have high PPV of 68, 70 and 73%,

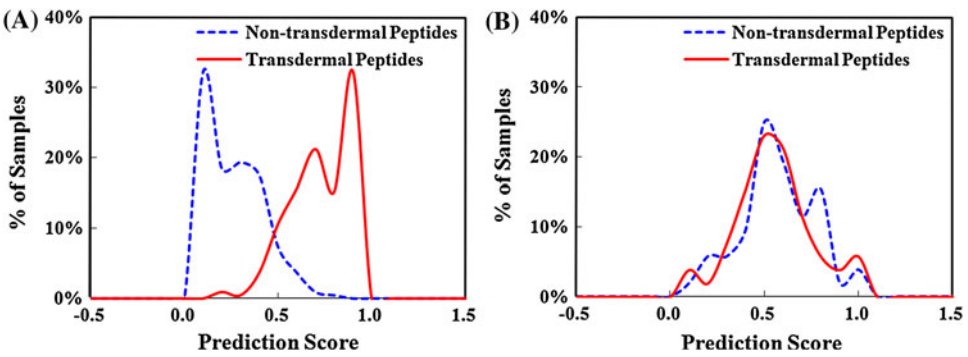

Fig. 4 The features of the SVM model constructed with the decoy set. The SVM model was constructed with a radial basis function using VHSE descriptor. **a** Training set and **b** test set

Table 4 Comparison of truth table statistics for the test set

Model	SE[a]	SP[b]	PPV[c]	NPV[d]	Acc[e]
ANN[f]	0.75	0.65	0.68	0.72	0.70
PLS[g]	0.75	0.67	0.70	0.73	0.71
SVM[h]	0.83	0.69	0.73	0.80	0.76
Consensus prediction[i]	0.71	0.87	0.84	0.75	0.79

[a] *SE* sensitivity: the proportion of all transdermal peptides correctly predicted, SE = TP/(TP + FN) where TP is the number of transdermal peptides correctly predicted and FN is the number of transdermal peptides incorrectly predicted as non-transdermal peptides

[b] *SP* specificity: the proportion of non-transdermal peptides correctly predicted, SP = TN/(TN + FP) where TN is the number of non-transdermal peptides correctly predicted and FP is the number of non-transdermal peptides incorrectly predicted as transdermal peptides

[c] *PPV* positive predictive value: the probability that a predicted transdermal peptide is in fact a transdermal peptide, PPV = TP/(TP + FP)

[d] *NPV* negative predictive value: the probability that a predicted non-transdermal peptide is in fact non-transdermal peptide, NPV = TN/(TN + FN)

[e] *Acc* accuracy: the percentage of all predictions that are correct, Acc = (TP + TN)/Total

[f] ANN models with network architecture with zero neuron in hidden layer and one in output layer; trained using the VHSE descriptor

[g] PLS models with 10 latent variables; trained using the VHSE descriptor

[h] SVM models with a radial basis function kernel; trained using the VHSE descriptor

[i] In the consensus prediction, transdermal peptides predicted by three models were considered as transdermal peptides, otherwise they were considered as non-transdermal peptides

respectively (Table 4). Recently, several studies have been reported combining different types of models to improve the accuracy of prediction [34–36]. We also tried to combine the three models to improve the performance of QSAR models and the PPV of the combined model is much higher than those of the individual models because the number of false positive is decreased.

Several studies have been reported QSAR models on the basis of peptide sequence information and examined diverse methods to predict the activity of peptides [37–41]. Burden et al. [38] developed Bayesian regularized neural network models using nonapeptide (nine amino acids) sequences as training set for the prediction of peptides binding to the major histocompatibility complex molecules and compared the results with previous studies. In the study, the authors restricted the number of neurons in a hidden layer to one or two to ensure that the number of weights is less than the number of training set, and obtained robust models with reliable performance. Because heptapeptides were used as training set and the number of peptides in training set is 434 in this study, we also restricted the number of neurons up to 2 to avoid overfitting. As shown in Table 1, the ANN models with zero neuron showed the smallest difference in the ROC scores between the training and test set, and the models with 2 neurons showed non-ignorable overfitting. Like ANN models, the SVM models with more complex kernel were overfitted relative to a simpler one. Although a part of models were overfitted, three models shown in Table 4 showed reliable performance in terms of the accuracy and PPV. To enhance predictive ability and robustness of model for prediction of transdermal peptide, the development of QSAR models using more robust methods or conditions would be a fruitful approach of future work.

One possible problem of our models which might cause prediction errors is the reliability of the data set. In this study, the positive control heptapeptides were identified by in vivo experiments through the three rounds of biopanning but the negative control peptides were randomly generated. This means that the sequences used as the negative control may not be true negatives. We believe, however, that the heptapeptides in the negative data set are very likely to be non-transdermal because the transdermal peptides identified by the in vivo experiment covered only a very small portion of the entire 'heptapeptide space' [28, 33], and because random sequences that are similar to the positive control sequences were removed from the negative control set. We could produce more reliable models if we were able to obtain more experimental data for model training.

Since peptides are in general relatively large and hydrophilic, the peptide-mediated transport uses an active transport strategy involving receptor-mediated endocytosis mechanism. In the active transport, the peptide interacts with a specific receptor that undergoes energy-dependent transport process such as endocytosis. Several studies have been shown that the peptides identified by phage display are transported across epithelial or endothelial barriers by binding to a specific receptor [19, 42–44]. To understand the mechanism underlying active transport of peptide at the skin, we will identify the receptor of transdermal peptide identified repetitively through an in vivo phage display technique.

The peptides have been shown to be promising materials for the improvement of the transport efficiency of macromolecules administrated via the convenient routes of drug delivery such as transdermal [3], peporal [22] and nasal route [45]. For example, peptide ligands crossing intact skin have been linked to small drug compounds or coadministrated with proteins [3], thus enhancing their therapeutic efficacy and convenience. We expect that our QSAR study on the prediction of transdermal peptides will be applicable to the development of peptide-based enhancers for the transdermal delivery of therapeutic agents.

In the recent years, nanoparticle have been studied as an effective delivery system of the skin because nanoparticles

such as liposomes [46, 47], nanostructure lipid carriers [48, 49] and polymeric nanoparticles [50] could release sustainedly the drug for a prolonged period of time and protect the therapeutic agent from chemical degradation. An alternative delivery strategy is the use of cell penetrating peptides (CPP) that have ability to translocate micro and macromolecules across the cell membrane. Rothbard et al. [51] firstly reported the application of CPP for delivery of peptides in the skin. Recently, Patlolla et al. [52] reported a study for transdermal delivery system using TAT peptide as CPP. The study showed that TAT peptide can translocate lipid crystal nanoparticle into skin layer. Like the CPP, the peptide predicted as transdermal peptide by our QSAR study could derive active transdermal transport of macromolecules and be available as a leading peptide for the carrier-drug conjugate strategy with macromolecular therapeutics such as the polymeric nanoparticle encapsulated peptide-drug conjugate or peptide-polymer conjugate encapsulated a drug.

Conclusions

We developed computational models to rapidly evaluate transdermal properties of peptides on the basis of their sequence information. These models trained using the ANN, PLS and SNM methods are capable of providing a reasonable prediction of transdermal peptide and expected to find applications in the selection of transdermal peptides from large peptide libraries. The selected peptides might be used as peptide-based enhancers to increase the efficacy and to reduce the toxicity associated with systemic administration of therapeutic agents.

Acknowledgments This work was supported by the Korea Science and Engineering Foundation (KOSEF) NRL Program grant funded by the Korea government (MEST) (No. R0A-2008-000-20024-1). We thank Accelrys Korea for the support of SciTegic Pipeline Pilot software.

References

1. Prausnitz MR, Mitragotri S, Langer R (2004) Current status and future potential of transdermal drug delivery. Nat Rev Drug Discov 3:115–124
2. Thomas BJ, Finnin BC (2004) The transdermal revolution. Drug Discov Today 9:697–703
3. Chen Y, Shen Y, Guo X, Zhang C, Yang W, Ma M, Liu S, Zhang M, Wen LP (2006) Transdermal protein delivery by a coadministered peptide identified via phage display. Nat Biotechnol 24:455–460
4. Benson HA (2005) Transdermal drug delivery: penetration enhancement techniques. Curr Drug Deliv 2:23–33
5. Madison KC (2003) Barrier function of the skin: "la raison d'etre" of the epidermis. J Invest Dermatol 121:231–241
6. Lopes LB, Brophy CM, Furnish E, Flynn CR, Sparks O, Komalavilas P, Joshi L, Panitch A, Bentley MV (2005) Comparative study of the skin penetration of protein transduction domains and a conjugated peptide. Pharm Res 22:750–757
7. Goodman M, Barry BW (1989) Lipid-protein-partitioning (LPP) theory of skin enhancer activity: finite dose technique. Int J Pharm 57:29–40
8. Rigg PC, Barry BW (1990) Shed snake skin and hairless mouse skin as model membranes for human skin during permeation studies. J Invest Dermatol 94:235–240
9. Takahashi K, Sakano H, Numata N, Kuroda S, Mizuno N (2002) Effect of fatty acid diesters on permeation of anti-inflammatory drugs through rat skin. Drug Dev Ind Pharm 2:1285–1294
10. Barry BW (2004) Breaching the skin's barrier to drugs. Nat Biotechnol 22:165–167
11. Karande P, Jain A, Mitragotri S (2004) Discovery of transdermal penetration enhancers by high-throughput screening. Nat Biotechnol 22:192–197
12. Prausnitz MR, Bose VG, Langer R, Weaver JC (1993) Electroporation of mammalian skin: a mechanism to enhance transdermal drug delivery. Proc Natl Acad Sci USA 90:10504–10508
13. Pillai O, Nair V, Panchagnula R (2004) Transdermal iontophoresis of insulin: IV. Influence of chemical enhancers. Int J Pharm 269:109–120
14. Kalia YN, Naik A, Garrison J, Guy RH (2004) Iontophoretic drug delivery. Adv Drug Deliv Re 56:619–658
15. Lavon I, Kost J (2004) Ultrasound and transdermal drug delivery. Drug Discov Today 9:670–676
16. McAllister DV, Allen MG, Prausnitz MR (2000) Microfabricated microneedles for gene and drug delivery. Annu Rev Biomed Eng 2:289–313
17. Prausnitz MR (2004) Microneedles for transdermal drug delivery. Adv Drug Deliv Re 56:581–587
18. Rajotte D, Ruoslahti E (1999) Membrane dipeptidase is the receptor for a lung-targeting peptide identified by in vivo phage display. J Biol Chem 274:11593–11598
19. Pasqualini R, Koivunen E, Kain R, Lahdenranta J, Sakamoto M, Stryhn A, Ashmun RA, Shapiro LH, Arap W, Ruoslahti E (2000) Aminopeptidase N is a receptor for tumor-homing peptides and a target for inhibiting angiogenesis. Cancer Res 60:722–727
20. Trepel M, Arap W, Pasqualini R (2002) In vivo phage display and vascular heterogeneity: implications for targeted medicine. Curr Opin Chem Biol 6:399–404
21. Duerr DM, White SJ, Schluesener HJ (2004) Identification of peptide sequences that induce the transport of phage across the gastrointestinal mucosal barrier. J Virol Methods 116:177–180
22. Kang SK, Woo JH, Kim MK, Woo SS, Choi JH, Lee HG, Lee NK, Choi YJ (2008) Identification of a peptide sequence that improves transport of macromolecules across the intestinal mucosal barrier targeting goblet cells. J Biotechnol 135:210–216
23. McCaldon P, Argos P (1988) Proteins 4:99–122
24. Mei H, Lian ZH, Zhou Y, Li SZ (2005) A new set of amino acid descriptors and its application in peptide QSARs. Biopolymer (Peptide Science) 80:775–786
25. Hellberg S, Sjöström M, Skagerberg B, Wold S (1987) Peptide quantitative structure-activity relationships, a multivariate approach. J Med Chem 30:1126–1135
26. Sandberg M, Eriksson L, Jonsson J, Sjöströ M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 41:2481–2491
27. The R Project for Statistical Computing (2009) http://www.r-project.org/. Version R 2.10.1
28. Jung E, Kim J, Choi SH, Kim M, Rhee H, Shin JM, Choi K, Kang SK, Lee NK, Choi YJ, Jung DH (2010) Artificial neural network study on organ-targeting peptides. J Comput Aided Mol Des 24:49–56

29. Venables WN, Ripley BD (2002) Modern applied statistics with S, Fourth Edition edn. Springer, Berlin
30. Pipeline Pilot 7.5. http://accelrys.com/products/pipeline-pilot/
31. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36
32. Springer C, Adalsteinsson H, Young MM, Kegelmeyer PW, Roe DC (2005) PostDock: a structural, empirical approach to scoring protein ligand complexes. J Med Chem 48:6821–6831
33. Jung E, Kim J, Kim M, Jung DH, Rhee H, Shin JM, Choi K, Kang SK, Kim MK, Yun CH, Choi YJ, Choi SH (2007) Artificial neural network models for prediction of intestinal permeability of oligopeptides. BMC Bioinformatics 8:245
34. Natt NK, Kaur H, Raghava GP (2004) Prediction of transmembrane regions of beta-barrel proteins using ANN- and SVM-based methods. Proteins: Str, Funct, Bioinf 56:11–18
35. Bhasin M, Raghava GP (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine 22:3195–3204
36. Wang D, Larder B, Revell A, Montaner J, Harrigan R, De Wolf F, Lange J, Wegner S, Ruiz L, Pérez-Elías MJ, Emery S, Gatell J, D'Arminio Monforte A, Torti C, Zazzi M, Lane C (2009) A comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy. Artif Intell Med 47:63–74
37. Bhasin M, Raghava GP (2004) SVM-based method for predicting HLADRB1*0401 binding peptides in an antigen sequence. Bioinformation 20:421–423
38. Burden FR, Winkler DA (2005) Predictive Bayesian neural network models of MHC class II peptide binding. J Mol Graph Model 23:481–489
39. Ivanciuc O, Braun W (2007) Robust quantitative modeling of peptide binding affinities for MHC molecules using physical-chemical descriptors. Protein Pept Lett 14:903–916
40. Zhang GL, Bozic I, Kwoh CK, August JT, Brusic V (2007) Prediction of supertype-specific HLA class I binding peptides using support vector machines. J Immunol Methods 320:143–154
41. Fjell CD, Jenssen H, Hilpert K, Cheung WA, Panté N, Hancock RE, Cherkasov A (2009) Identification of novel antibacterial peptides by chemoinformatics and machine learning. J Med Chem 52:2006–2015
42. Essler M, Ruoslahti E (2002) Molecular specialization of breast vasculature: a breast-homing phage-displayed peptide binds to aminopeptidase P in breast vasculature. Proc Natl Acad Sci USA 99:2252–2257
43. Kolonin MG, Saha PK, Chan L, Pasqualini R, Arap W (2004) Reversal of obesity by targeted ablation of adipose tissue. Nat Med 10:625–632
44. Lee NK, Kim M, Choi JH, Kim EB, Lee HG, Kang SK, Choi YJ (2010) Identification of a peptide sequence targeting mammary vasculature via RPLP0 during lactation. Peptides 31:2247–2254
45. Wan XM, Chen YP, Xu WR, Yang WJ, Wen LP (2009) Identification of nose-to-brain homing peptide through phage display. Peptides 30:343–350
46. Loan Honeywell-Nguyen P, Bouwstra JA (2005) Vesicles as a tool for transdermal and dermal delivery. Drug Discovery Today: Technol 2:67–74
47. Elsayed MM, Abdallah OY, Naggar VF, Khalafallah NM (2007) Lipid vesicles for skin delivery of drugs: reviewing three decades of research. Int J Pharm 332:1–16
48. Mehnert W, Mäder K (2001) Solid lipid nanoparticles: production, characterization and applications. Adv Drug Deliv Rev 47:165–196
49. Patlolla RR, Chougule M, Patel AR, Jackson T, Tata PN, Singh M (2010) Formulation, characterization and pulmonary deposition of nebulized celecoxib encapsulated nanostructured lipid carriers. J Control Release 144:233–241
50. Paños I, Acosta N, Heras A (2008) New drug delivery systems based on chitosan. Curr Drug Discov Technol 5:333–341
51. Rothbard JB, Garlington S, Lin Q, Kirschberg T, Kreider E, McGrane PL, Wender PA, Khavari PA (2000) Conjugation of arginine oligomers to cyclosporin A facilitates topical delivery and inhibition of inflammation. Nat Med 6:1253–1257
52. Patlolla RR, Desai PR, Belay K, Singh MS (2010) Translocation of cell penetrating peptide engrafted nanoparticles across skin layers. Biomaterials 31:5598–5607