J Comput Aided Mol Des (2012) 26:1267–1275
DOI 10.1007/s10822-012-9614-6

Identification of tissue-specific targeting peptide

Eunkyoung Jung · Nam Kyung Lee · Sang-Kee Kang · Seung-Hoon Choi · Daejin Kim · Kisoo Park · Kihang Choi · Yun-Jaie Choi · Dong Hyun Jung

Received: 14 May 2012 / Accepted: 19 October 2012 / Published online: 27 October 2012

Abstract Using phage display technique, we identified tissue-targeting peptide sets that recognize specific tissues (bone-marrow dendritic cell, kidney, liver, lung, spleen and visceral adipose tissue). In order to rapidly evaluate tissue-specific targeting peptides, we performed machine learning studies for predicting the tissue-specific targeting activity of peptides on the basis of peptide sequence information using four machine learning models and isolated the groups of peptides capable of mediating selective targeting to specific tissues. As a representative liver-specific targeting sequence, the peptide "DKNLQLH" was selected by the sequence similarity analysis. This peptide has a high degree of homology with protein ligands which can interact with corresponding membrane counterparts. We anticipate that our models will be applicable to the prediction of tissue-specific targeting peptides which can recognize the endothelial markers of target tissues.

Keywords Machine learning · Partial least squares · Artificial neural network · Bayesian · Support vector machine · Tissue-specific targeting peptide · ROC score

Electronic supplementary material The online version of this article (doi:10.1007/s10822-012-9614-6) contains supplementary material, which is available to authorized users.

E. Jung · S.-H. Choi · D. Kim · K. Park · D. H. Jung (✉)
Insilicotech Co. Ltd., C-602 Korea Bio Park, 694-1, Sampyeong, Bundang-Gu, Seongnam-Shi 463-400, Korea
e-mail: dhjung@insilicotech.co.kr

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

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

D. Kim
e-mail: djkim@insilicotech.co.kr

K. Park
e-mail: kspark@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

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

In pharmaceutical research and development, targeted drug delivery (TDD) is as important as the optimization of pharmacological specificity and potency of the drug because TDD in specific disease sites can achieve a therapeutic effect with a lower dose and fewer undesirable side effects. To date, the TDD system has been exploited as a therapeutic tool in various treatments using appropriate homing ligands, including antibodies, peptides, aptamers and polymeric particles [1–4]. One approach is based on a unique set of marker molecules on endothelial surfaces within different tissues [5–7]. Because endothelial cells lining blood vessels are heterogeneous and express tissue-specific markers [8, 9], some of the most specific can be used as prospective targets in directed therapy, as also applies to other tissues [10–14].

Peptide ligands that home to specific cells or tissues provide a suitable transport mechanism that is structural simplicity itself, with ease of synthesis and a low probability of undesirable immunogenicity [15]. To date, various homing peptides and their modified platforms have successfully intensified the therapeutic availability of drugs against several tumors [16–18] or autoimmune diseases [19, 20] via peptide-mediated TDD. In addition, some peptides have been identified that enhance the molecular trafficking efficiency across tight tissue barriers in the body [21–23]. To obtain these particular peptides, the phage display technique (which massively selects convergent peptide sequences specific to targets of interest using random peptide-displaying phage libraries) is widely used as a rational screening method both in vitro and in vivo heterogenic molecular populations. As a result, a number of peptides that home to the vasculature of normal tissues, including brain, kidney [5], lung, skin, pancreas, intestine, adrenal gland, retina [6], breast [12], prostate [13], heart [24], and pathological tissues such as tumor [10, 25–27] have been reported. Recently, a report by Arap et al. [28] of phage library screening of a patient described peptides capable of selectively targeting to vascular receptors of bone marrow, skin, fat, muscle and prostate. Homing peptides identified by this screening technique have been successfully and promise to give good delivery of drug molecules and other agents to designated sites. These strategies should increase therapeutic efficacy and site specificity of drugs whilst reducing side effects. For example, the RGD and NGR peptide motifs that are the first generation tumor-homing peptides were coupled to doxorubicin [10], pro-apoptotic peptides [29, 30], cytotoxic agents [31], and cytokines [32, 33], which resulted in enhanced therapeutic efficiency compared to the untargeted drug. The pro-apoptotic peptide $(KLAKLAK)_2$ linked with the prostate-homing peptide (SMSIARL) led to tissue destruction, reduction in prostate size, and delayed cancer progression [12].

To analyze effectively the vast amounts of biological data on peptides now that thousands of different peptides have been identified by screening procedures and biological assays such as phage display experiment, we have carried out the machine learning studies to help predict the diverse properties of peptides [34–36]. Machine learning models based on sequence information that predict and rank peptides might help find suitable targeted delivery vehicles that increase the efficacy of therapeutic agents. Using sequence sets of phage-displayed peptides selected from phage display experiments, we constructed machine learning models to screen tissue-targeting peptides, followed by the isolation of peptides from tissue-targeting peptide sets capable of mediating selective targeting to specific tissues.

Materials and methods

Phage-peptide library and animals

To screen cell- or tissue-targeting peptide sequences, we employed combinatorial phage display peptide library (New England Biolabs, MD, USA) of random peptides fused to a pIII coat protein of the M13 phage. The peptides were displayed separately on the N-termini of pIII, which was followed by a Gly–Gly-Gly-Ser linker to the wildtype pIII sequence. Phage-peptide libraries displaying flexible linear heptapeptides (Ph.D.-7^{TM}) or disulfide-constrained cyclic heptapeptides with 2 flanking cysteines (Ph.D.-$C7C^{TM}$) were used for in vitro or in vivo screening.

All animal studies were conducted in accordance with the Institutional Animal Care and Use Committee guidelines of the Seoul National University. BALB/c mice, Wistar rats, or Sprague–Dawley (SD) rats (Samtako, Osan, Korea) were kept under standard laboratory conditions with a 12 h light/dark cycle, at constant temperature (20 °C), and a humidity of 48 %, to be used for the following phage display screenings.

Preparation of peptide pools

In vitro cell-based screening for identification of peptide sequences targeting bone marrow-derived dendritic cells (BMDCs)

To obtain immature DCs from bone marrow (BM), BM cells were isolated from the femurs and tibias of mice, and BMDCs were prepared as described in Ref. [37]. After phenotypic analysis by detecting BMDCs-specific markers using flow cytometry [38], the BMDCs were used for in vitro phage display screening as follows. After the final biopanning, a total of 151 peptide sequences were identified from randomly selected individual phage recombinants as candidates of BMDCs-targeting peptides. The in vitro screening procedure is depicted in Fig. 1a.

In vivo tissue-based screening I for identification of peptide sequences targeting visceral adipose tissue through transdermal route

To identify peptide moieties targeting visceral adipose tissue through the transdermal route, an in vivo phage display screening was conducted as previously described [22], with slight modifications. After the final biopanning, a total of 269 peptide sequences were identified from randomly selected individual phage recombinants eluted from the visceral adipose tissue as candidates of skin-to-visceral

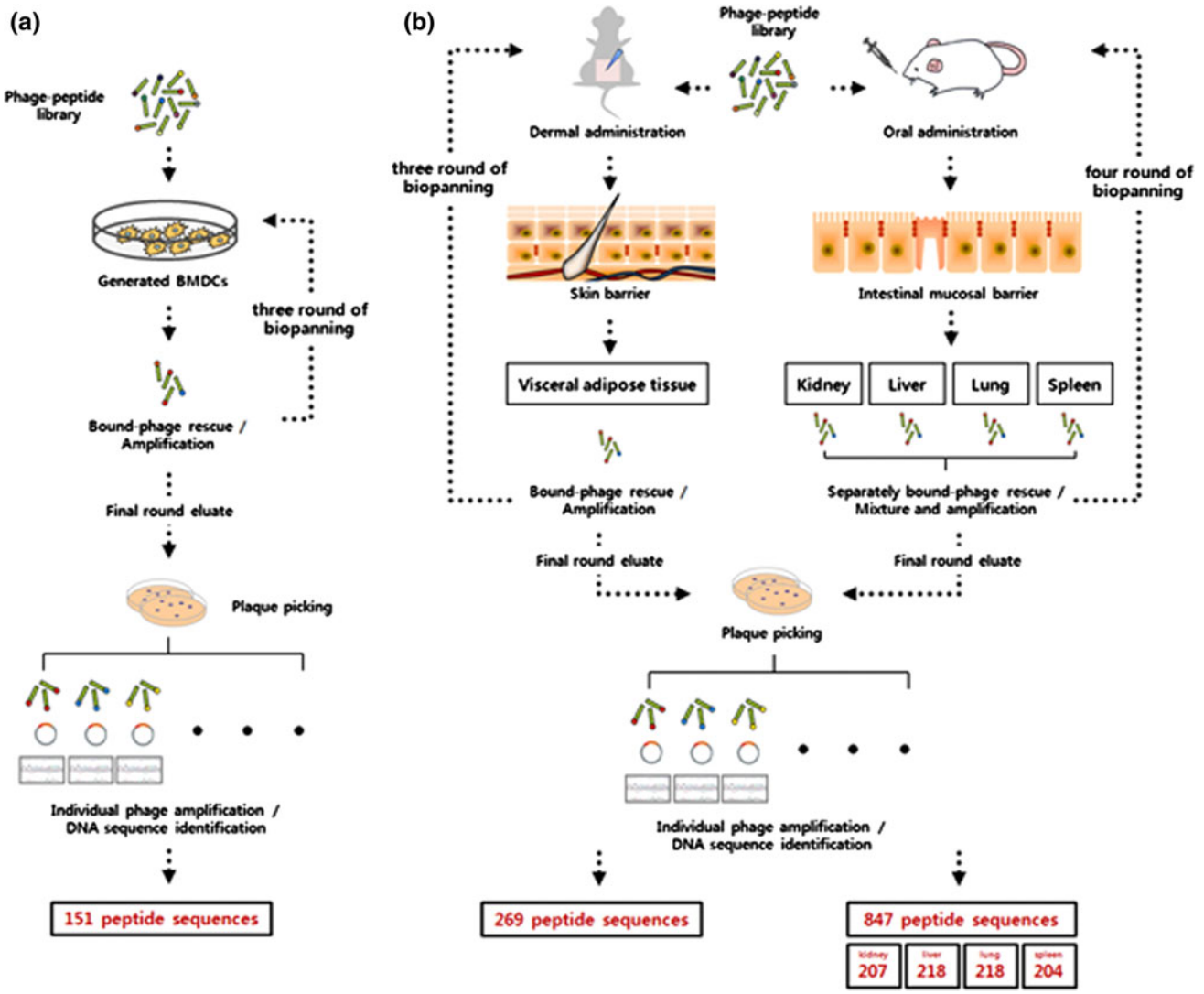

Fig. 1 Flow charts of the in vitro and in vivo phage display screening procedures for the preparation of peptide pools

adipose tissue-targeting peptide moieties. The procedure of this in vivo screening is outlined in Fig. 1b.

In vivo tissue-based screening II for identification of peptide sequences permeable to intestinal barrier

To identify peptide moieties that penetrate the intestinal mucosal barrier, an in vivo phage display screening was conducted as previously described [21]. After the final biopanning, individual phage recombinants were randomly selected from each organ eluate (kidney (207), liver (218), lung (218), and spleen (204)), and separately amplified for identification of their peptide sequences. The outlines procedure of this in vivo screening is given in Fig. 1b.

Detailed methodologies of in vitro and in vivo phage display screenings indicated above are described in the Supporting Information.

Data sets

The positive control data set of peptides homing to each tissue was obtained from sequences identified by phage display experiments. The negative control data set was generated from random sequences that had the same frequencies of occurrence of each amino acid residue in the primary structures of unrelated proteins of known sequence [39]. 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 (detailed methodologies are described in the Supporting Information).

Descriptors

The VHSE descriptor was used to encode important features of individual peptide sequences. It is composed of 8

variables for each amino acid, describing its hydrophobic, steric and electronic properties [40].

Machine learning models

The partial least squares (PLS) models with 10 latent variables were carried out using the pls package implementing PLS regression (PLSR) and principal component regression (PCR) [41]. The Artificial Neural Network (ANN) models were used the nnet of the VR 7.2 package [42] for feed-forward neural networks and multinomial log-linear models. The neural network architecture of the ANN models was set as previously described [36]. The Bayesian models with 10 bins were carried out using a two-class Bayesian categorization component of the Pipeline Pilot 8.0 [41]. The SVM models with radial basis function were implemented using the function svm of the e1071 package [42]. The input value of the positive control was 0.9 and the negative control was 0.1 before the learning network was applied. Workflow for all calculations was automated through Pipeline Pilot 8.0 [41].

Evaluation and validation

Predictive performance, assessed using the receiver operating characteristics (ROC) score [43], was used for each training and test 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.

More information on machine learning and the validation methods is given in the Supporting Information.

Sequence similarity calculation and candidate motif search

The pair-wise similarity between tissue-specific targeting peptides was calculated using the Align Sequences protocol in Discovery Studio 3.1 [44]. The candidate motifs for the tissue-specific targeting peptide were searched in the SWISSPROT databases at the NCBI, using the BLAST Search (NCBI Server) protocol in Discovery Studio 3.1 [44], and proteins originated from humans were selected from the BLAST hits. The BLOSUM 62 [45] scoring matrix was used for searching homologs and calculating pair-wise similarity of residues.

Results

Screening tissue-targeting peptides

Using the phage display screening, heptapeptide sequences homing to target tissues have been identified. The targeting peptide sequences for 7 tissues are listed in the Supporting Information (Table S1). They were used as the positive control set in further analyses. Table 1 shows the number of tissue-targeting sequences and peptides in the training and test sets.

Building the machine learning models

To develop and optimize the method for screening of tissue-targeting peptides, 4 machine learning algorithms, including partial lease squares (PLS), artificial neural network (ANN), Bayesian, and support vector machine (SVM) were examined. The model parameters were optimized to find the best prediction model. Because the VHSE descriptor showed significant predictive power for permeability [34, 36] and target delivery [35] in previous studies, it was used for screening of tissue-targeting peptide.

The receiver operating characteristics (ROC) scores of the models for each tissue are listed in Table 1. The bone-marrow dendritic cell (BMDC)-targeting models that have a relatively high sequence redundancy in the positive control set had relatively higher prediction scores than the models for the other tissues. Figure 2 shows the ROC curves of PLS, ANN, Bayesian and SVM models for the test sets. The left panel in Fig. 2 shows the curves for models trained with the BMDC-targeting peptide set that has the highest sequence redundancy in the positive control set. The right panel shows those for models trained with the small intestine-targeting peptide set that has the greatest number of sequences.

Validation of machine learning models

To evaluate the stability of machine learning models in predicting tissue-targeting peptides, we performed leave-group-out cross-validation; the results are listed in the Supporting Information (Table S2 and S3). For 4 machine learning methods, the standard deviation of the ROC scores was small for the training set, but relatively large for the test set. Although not markedly different, the SVM models performed better than average in others.

To test the reliability of the peptide sequences defined as the positive control in the machine learning methods, and validate the strength of our models in predicting tissue-targeting peptides, separate Y-randomization sets were generated by a random permutation of activity values to change true order of the activity data. Supplementary models trained with these Y-randomization sets were then used for comparison with the models trained with the actual data set identified by the phage display experiment. The results suggested that the predictive power of the models constructed with the actual positive set (Table 1) was considerably greater than that of the models trained with the

Table 1 The number of data sets and prediction accuracy for machine-learning models[a]

Tissues	Data set			PLS[b]		ANN[c]		Bayesian		SVM[d]	
	No. of targeting sequences	Training set	Test set	Training	Test	Training	Test	Training	Test	Training	Test
BMDC	151	244	58	0.96	0.95	0.96	0.97	0.97	0.96	1.00	0.99
Kidney	207	330	84	0.87	0.75	*0.88*	*0.75*	0.90	0.78	0.99	0.84
Liver	218	348	88	0.86	0.75	*0.86*	*0.75*	0.89	0.71	0.99	0.70
Lung	218	348	88	0.86	0.77	*0.86*	*0.76*	0.88	0.75	0.99	0.85
Spleen	204	326	82	0.88	0.76	*0.88*	*0.76*	0.90	0.81	0.99	0.79
Small intestine	847	1,348	346	0.81	0.70	0.82	0.70	0.82	0.73	0.97	0.76
Visceral adipose tissue	269	434	104	*0.91*	*0.85*	*0.91*	*0.86*	0.92	0.88	*1.00*	*0.91*

[a] The prediction scores written in italic type were obtained from our previous studies [35, 36]

[b] The PLS models trained with 10 latent variables

[c] The ANN models trained with network architecture with zero neuron in hidden layer and one in output layer

[d] The SVM models trained with a radial basis function

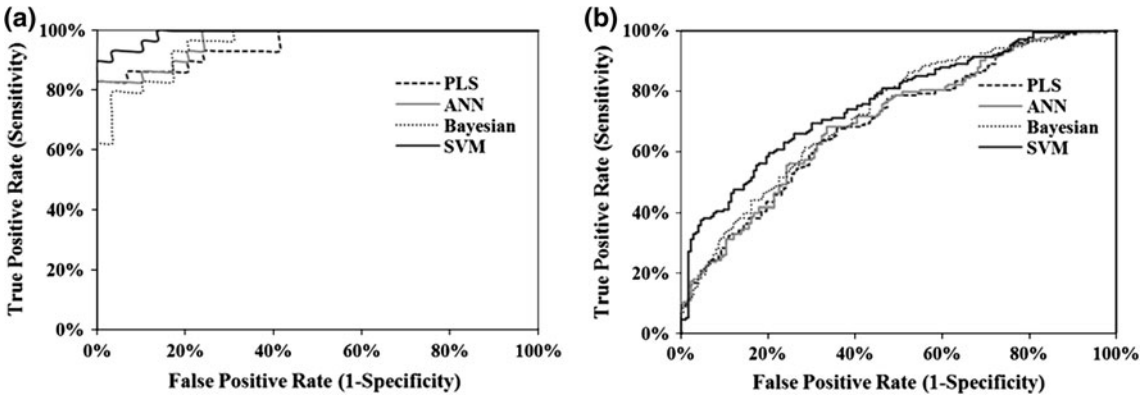

Fig. 2 Receiver operating characteristics (ROC) curve of the models for test sets. **a** The models for the predictions of BMDC-targeting peptides and **b** the model for the prediction of intestinal-barrier permeable peptides

Y-randomization set (Table S4 in the Supporting Information). This result confirms that the peptides identified from phage display screening have meaningful sequence patterns related to their tissue-targeting activities, although the mechanism for the tissue-targeting is not fully understood.

More detailed statistics of the predictive capacities of our models are listed in the Supporting Information (Table S5), which shows the truth table analysis of the binary outcome based on tissue-targeting activity. Comparing sensitivity and specificity, PLS, ANN and SVM models seem to be more sensitive in predicting tissue-targeting peptides rather than specific in screening out non-tissue-targeting peptides, but there was no such tendency for Bayesian models.

Prediction of tissue-specific targeting peptides

We tried to predict tissue-specific targeting peptides using SVM models that showed higher overall predictive capability than the others. The tissue-specific targeting peptides were selected by filtering peptides with the prediction score satisfying the specific conditions; peptides with a prediction score >0.7 (or 0.8) for target tissue and those <0.5 for the other 5 tissues. For example, BMDC targeting peptides with the prediction score >0.7 for BMDC targeting model and those <0.5 for kidney, liver, lung, spleen and visceral adipose tissue targeting models were selected as the BMDC-specific targeting peptides. The tissue-specific targeting peptides filtered by the condition are listed in Table 2, and their sequences in Table 3. The tissue-specific targeting sequences in Table 3 were not included in targeting peptide sets for other than the target tissue. For example, the "TTWNPLD" peptide predicted as a BMDC-specific targeting peptide was not identified as a kidney, liver, lung, spleen or visceral adipose tissue-targeting peptide in our phage display experiment. As a representative specific targeting sequence, the peptide "DKNLQLH" was selected from sequence similarity

Table 2 Prediction of tissue-specific targeting peptides using the SVM models with radial basis function

Target tissue	No. of sequences[a]	No. of sequences[b]
BMDC	6	4
Kidney	2	1
Liver	8	3
Lung	5	3
Spleen	3	2
Visceral adipose tissue	10	5

[a] Peptides with the prediction score higher than 0.7 for the target tissue and those lower than 0.5 for the others

[b] Peptides with the prediction score higher than 0.8 for the target tissue and those lower than 0.5 for the others

analysis conducted with liver-specific targeting peptides (Table S6 in the Supporting Information). To evaluate possible ligands corresponding to the "DKNLQLH" peptide, candidate motifs were searched in the SWISSPROT databases at the NCBI. Proteins with an identity in the aligned region of >80 % are listed in the Supporting Information (Table S7), and some proteins in the results were seen as specific proteins associated with molecular adhesion or transport, with varying degrees of homology (Table 4).

Discussion

There has been increasing interest in using peptide moieties to deliver therapeutic and diagnostic molecules to specific select targets. Although peptide moieties suffer from short biological half-life and low target-binding affinities in the clinical application, much effort has been made to overcome these defects by combining peptides with polymeric structures to enhance their stability [46], and by introducing modified amino acids to optimize the affinity [47]. But there is also another consideration for more effective analysis of peptide sequences identified from phage display biopanning. The phage populations filtered by binding affinities to the targets are in the thousands to millions, and among these numerous candidates, the number of peptide sequences that can be detected from individual phage recombinants is severely limited due to the labor and cost-intensive sequencing steps using phage-DNAs. Thus, the prediction of the most effective sequences based on the limited experimental data would greatly facilitate phage display screening and subsequent peptide analysis. For this purpose, we have developed machine learning models to predict peptide enhancers for use in targeted molecular delivery.

Our QSAR models are based on the PLS, ANN, Bayesian and SVM methods. The model parameters in each were set to the same conditions shown to give the best performance in predicting properties in our previous studies [35, 36]. These 4 machine learning methods performed well in predicting tissue-targeting peptides (Table 1). There was no significant difference between machine learning models in their accuracy for the prediction of tissue-targeting peptides, but that of SVM models proved slightly superior to the others. r^2 of SVM model for BMDC, kidney, liver, lung, spleen, small intestine and visceral adipose tissue was 0.94, 0.81, 0.79, 0.83, 0.83, 0.72 and 0.87, respectively.

We analyzed the regression coefficients of the PLS model to identify major property for tissue-targeting property of peptide. The coefficients for variables listed in Supporting Information (Table S8) indicated that the steric property (v3 or v4) at 4, 5 or 6 position of targeting heptapeptide is important to have tissue-targeting property of heptapeptides and amino acids with negative scales for steric property are generally preferred. 4 and 7 positions in liver-targeting peptides are worth paying special attention to. Referring to the descriptor scales, hydrophobic character at those positions in liver-targeting model is preferred and indeed liver-specific targeting peptides listed in the Table 3 have hydrophobic amino acids at the positions.

Although we tested diverse methods and optimized model parameters, one possible problem that might cause prediction errors is the reliability of the data set. Unlike positive control peptides identified by phage display experiments, the sequences used as negative controls may not be true negatives since the negative control peptides were randomly chosen. We believe, however, that the heptapeptides in the negative data set are probably non-tissue-targeting because the tissue-targeting peptides identified by the experiment covered only a fraction of the entire 'heptapeptide space', as discussed in our previous studies [34–36]. More reliable models might be developed if we were able to confirm the negative data set by experiments.

Regarding the mechanism of tissue targeting peptides, several studies have reported the existence of receptors for homing peptides. For example, Zhang et al. [24] identified protein targets for the heart-homing peptides (CRPPR, CKRAVR, CPKTRRVPS, CRSTRANPC, and CARPAR). By using peptide affinity chromatography, Rajotte et al. [48] identified membrane dipeptidase [MDP] as the receptor for lung-homing peptide GFE-1 (CGFECVRQCPERC) and Kolonin et al. [49] identified prohibitin as binding to fat-homing peptide (CKCCRAKDC). As an attempt to understand the mechanism of tissue-specific targeting peptides, we have performed a test analysis to search ligand proteins containing the sequence motif, D–LQL, derived from a liver-specific targeting peptide (Table 4). One of the ligands is von Willbrand Factor A domain-containing protein 5B2 (coch-5B2), which circulates in bloodstream and binds to

Table 3 Tissue-specific peptide sequences

Target tissue	Peptide sequence	Prediction scores[a]					
		BMDC	Kidney	Liver	Lung	Spleen	Visceral adipose tissue
BMDC	TTWNPLD	**0.860**	0.108	0.390	0.282	0.125	0.383
	RVPTWPS	**0.860**	0.485	0.419	0.373	0.325	0.475
	GPHYINY	**0.860**	0.331	0.407	0.489	0.388	0.345
	FNVVALH	**0.841**	0.285	0.371	0.410	0.416	0.280
	YIHAPAP	**0.784**	0.446	0.230	0.303	0.356	0.456
	HIPYDPP	**0.774**	0.250	0.062	0.027	-0.006	0.473
Kidney	PKNGSDP	0.114	**0.860**	0.290	0.451	0.163	0.339
	DSHKDLK	0.217	**0.744**	0.477	0.369	0.348	0.259
Liver	GVRTNLL	0.359	0.383	**0.860**	0.343	0.420	0.226
	ENRSDKV	0.085	0.449	**0.860**	0.441	0.327	0.362
	NSELNMQ	0.268	0.329	**0.855**	0.329	0.264	0.348
	DSLTYMH	0.236	0.180	**0.747**	0.317	0.190	0.413
	GVLFTQD	0.168	0.242	**0.739**	0.082	0.091	0.462
	PNEKEPD	0.238	0.372	**0.724**	0.379	0.474	0.167
	NNNSNVE	0.217	0.295	**0.722**	0.178	0.316	0.453
	DKNLQLH	0.099	0.385	**0.714**	0.478	0.451	0.383
Lung	LPYDKRI	0.295	0.402	0.348	**0.860**	0.305	0.307
	TQMGYTM	0.217	0.431	0.416	**0.860**	0.261	0.432
	LHGFMDR	0.074	0.388	0.258	**0.800**	0.342	0.229
	LHTRTPL	0.397	0.476	0.347	**0.800**	0.458	0.367
	MDRYSPR	0.298	0.251	0.293	**0.788**	0.457	0.422
Spleen	RTLPINV	0.143	0.329	0.298	0.326	**0.860**	0.334
	EPTTAQW	0.484	0.465	0.252	0.457	**0.834**	0.462
	RDPVLST	0.330	0.323	0.158	0.225	**0.731**	0.474
Visceral adipose tissue	NWSQQST	0.285	0.457	0.201	0.328	0.304	**0.860**
	HQVFGAY	0.270	0.299	0.240	0.362	0.227	**0.860**
	SISYNSF	0.399	0.476	0.218	0.294	0.443	**0.860**
	QWSVYNN	0.222	0.379	0.102	0.213	0.399	**0.860**
	DRQPPQP	0.150	0.390	0.302	0.379	0.219	**0.831**
	IESMDSH	0.190	0.290	0.313	0.419	0.252	**0.792**
	ATYSPDR	0.451	0.236	0.324	0.340	0.340	**0.787**
	QLNYPHD	0.477	0.387	0.344	0.125	0.414	**0.738**
	QWSVYTN	0.244	0.223	0.053	0.224	0.291	**0.718**
	TSTWSFR	0.338	0.212	0.343	0.412	0.363	**0.717**

[a] The prediction scores obtained from SVM models trained with a radial basis function for each target tissue and the scores in bold represents the prediction scores for target tissue of tissue-specific targeting peptide

Table 4 Examples of candidate human proteins mimicked by DKNLQLH peptide

Protein[a]	Protein description	Identity[b]	Similarity[c]	Accession number	Homology sequence
von Willebrand Factor A domain-containing protein 5B2 [Coch-5B2]	Cell/ECM/integrin adhesion domain, cell aggregation effector	85	100	Q8N398	DQNLQLH
Myosin-Vb	Rab11 interactor, transcytosis regulation	83	100	Q9ULV0	DKSLQL-
Neuropilin-2	VEGF/VEGFR interactor	83	100	O60462	DKSLQL-

[a] Selection from searched proteins with the identity greater than 80 % against DKNLQLH sequence in aligned region

[b] Sequence identity against DKNLQLH sequence in the aligned region

[c] Sequence similarity against DKNLQLH sequence in the aligned region

endothelial or extracellular matrix components such as collagens or integrins [50]. Another protein bearing the same motif, Myosin-Vb, is a key molecule interacting with Rab11, which localizes to the apical recycling endosome and regulates transcytosis [51]. In addition, Neuropilin-2 is possible protein interacting with VEGFR-2 and -3 localized on the membrane of endothelial cells [52]. The results revealed that the sequence motif might represent sequences present in circulation, targeting cell-surface proteins that participate in molecular adhesion, transport machinery or home to vascular endothelium. Although the way that predicted peptides bind to their corresponding tissue is not clear, we consider that homological sequences such as D–LQL may be important in targeting specificity of the peptide to liver tissue. Further studies are underway to investigate biological functions of the predicted peptides as tissue-homing ligands and to expand the scope of target tissues.

We have tried to isolate tissue-specific targeting peptides by filtering those with the prediction score that satisfied the specific cutoff (listed in Table 4). Although peptides given in Table 4 bind selectively to their target tissue set out in this study, they may also bind to other tissues that had not been used as targets in phage display experiment. To use the peptides in Table 4 as targeted delivery vehicles, we need to validate the tissue-targeting properties of peptides by in vivo experiments for diverse tissues. To predict peptides capable of binding selectively to target tissue through machine learning study for efficient target delivery of therapeutic agents, we would need to carry out a systematic search for more diverse tissues.

Conclusions

We developed computational models to rapidly screen tissue-specific targeting peptides for BMDC, kidney, liver, lung, spleen and small intestine on the basis of their sequence information directly obtained by phage display experiments. These models trained by using the PLS, ANN, Bayesian and SVM methods are capable of providing a reasonable prediction of tissue-targeting peptides and the combination of prediction models turns out to be successful in discriminating candidate peptide sequences specifically targeting to one of the six tissues. This approach is expected to find applications in the prediction of tissue-specific targeting peptides which can recognize the endothelial markers of target tissues.

Acknowledgments This research is supported by National Research Foundation of Korea (NRF), Korea government (MEST) (Project No. 2011-0029416). We thank Accelrys Korea for the support of SciTegic Pipeline Pilot and Discovery Studio software, and acknowledge the assistance of BioMedES (http://www.biomedes.co.uk/home).

References

1. Alley SC, Okeley NM, Senter PD (2010) Antibody-drug conjugates: targeted drug delivery for cancer. Curr Opin Chem Biol 14:529–537
2. Vives E, Schmidt J, Pelegrin A (2008) Cell-penetrating and cell-targeting peptides in drug delivery. BBA Rev Cancer 1786: 126–138
3. Ray P, White RR (2010) Aptamers for targeted drug delivery. Pharmaceuticals 3:1761–1778
4. Singha R, Lillard JWM (2009) Nanoparticle-based targeted drug delivery. Exp Mol Pathol 86:215–223
5. Pasqualini R, Ruoslahti E (1996) Organ targeting in vivo using phage display peptide libraries. Nature 380:364–366
6. Rajotte D, Arap W, Hagedorn M, Koivunen E, Pasqualini R, Ruoslahti E (1998) Molecular heterogeneity of the vascular endothelium revealed by in vivo phage display. J Clin Invest 102:430–437
7. Trepel M, Arap W, Pasqualini R (2000) Exploring vascular heterogeneity for gene therapy targeting. Gene Ther 7:2059–2060
8. William CA (2007) Phenotypic heterogeneity of the endothelium : I. Structure, function, and mechanisms. Circ Res 100:158–173
9. William CA (2007) Phenotypic heterogeneity of the endothelium. Circ Res 100:174–190
10. Arap W, Pasqualini R, Ruoslahti E (1998) Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. Science 279:377–380
11. Kolonin MG, Pasqualini R, Arap W (2001) Molecular addresses in blood vessels as targets for therapy. Curr Opin Chem Biol 5:308–313
12. Arap W, Haedicke W, Bernasconi M, Kain R, Rajotte D, Krajewski S, Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti E (2002) Targeting the prostate for destruction through a vascular address. Proc Natl Acad Sci USA 99:1527–1531
13. Durr E, Yu J, Krasinska KM, Carver LA, Yates JR, Testa JE, Oh P, Schnitzer JE (2004) Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. Nat Biotechnol 22:985–992
14. Oh P, Li Y, Yu J, Durr E, Krasinska K, Carver LA, Testa JE, Schnitzer JE (2004) Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. Nature 429:629–635
15. Majumdar S, Siahaan TJ (2010) Peptide-mediated targeted drug delivery. Med Res Rev. doi:10.1002/med.20225
16. Binétruy-Tournaire R, Demangel C, Malavaud B, Vassy R, Rouyre S, Kraemer M, Plouët J, Derbin C, Perret G, Mazié JC (2000) Identification of a peptide blocking vascular endothelial growth factor(VEGF)-mediated angiogenesis. EMBO J 19:1525–1533
17. Askoxylakis V, Zitzmann S, Mier W, Graham K, Krämer S, von Wegner F, Fink RH, Schwab M, Eisenhut M, Haberkorn U (2005) Preclinical evaluation of the breast cancer cell-binding peptide, p160. Clin Cancer Res 11:6705–6712
18. Zang L, Shi L, Guo J, Pan Q, Wu W, Pan X, Wang J (2009) Screening and identification of a peptide specifically targeted to NCI-H1299 from a phage display peptide library. Cancer Lett 281:64–70
19. Yanofsky SD, Baldwin DN, Butler JH, Holden FR, Jacobs JW, Balasubramanian P, Chinn JP, Cwirla SE, Peters-Bhatt E, Whitehorn EA, Tate EH, Akeson A, Bowlin TL, Dower WJ, Barrett RW (1996) High affinity type I interleukin 1 receptor antagonists discovered by screening recombinant peptide libraries. Proc Natl Acad Sci USA 93:7381–7386
20. Tian W, Bai G, Li ZH, Yang WB (2006) Antagonist peptides of human interferon-alpha2b isolated from phage display library

inhibit interferon induced antiviral activity. Acta Pharmacol Sin 27:1044–1050
21. 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
22. 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 Biothechnol 24:455–460
23. 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
24. Zhang L, Hoffman JA, Ruoslahti E (2005) Molecular profiling of heart endothelial cells. Circulation 112:1601–1611
25. Pasqualini R, Koivunen E, Ruoslahti E (1997) Alpha v integrins as receptors for tumor targeting by circulating ligands. Nat Biotechnol 15:542–546
26. Burg MA, Pasqualini R, Arap W, Ruoslahti E, Stallcup WB (1999) NG2 proteoglycan-binding peptides target tumor neovasculature. Cancer Res 59:2869–2874
27. Laakkonen P, Akerman ME, Biliran H, Yang M, Ferrer F, Karpanen T, Hoffman RM, Ruoslahti E (2004) Antitumor activity of a homing peptide that targets tumor lymphatics and tumor cells. Proc Natl Acad Sci USA 101:9381–9386
28. Arap W, Kolonin MG, Trepel M, Lahdenranta J, Cardó-Vila M, Giordano RJ, Mintz PJ, Ardelt PU, Yao VJ, Vidal CI, Chen L, Flamm A, Valtanen H, Weavind LM, Hicks ME, Pollock RE, Botz GH, Bucana CD, Koivunen E, Cahill D, Troncoso P, Baggerly KA, Pentz RD, Do KA, Logothetis CJ, Pasqualini R (2002) Steps toward mapping the human vasculature by phage display. Nat Med 8:121–127
29. Ellerby HM, Arap W, Ellerby LM, Kain R, Andrusiak R, Rio GD, Krajewski S, Lombardo CR, Rao R, Ruoslahti E. Bredesen DE, Pasqualini R (1999) Anti-cancer activity of targeted pro-apoptotic peptides. Nat Med 5:1032–1038
30. Chen Y, Xu X, Hong S, Chen J, Liu N, Underhill CB, Creswell K, Zhang L (2001) RGD-tachyplesin inhibits tumor growth. Cancer Res 61:2434–2438
31. Yoneda Y, Steiniger SC, Capková K, Mee JM, Liu Y, Kaufmann GF, Janda KD (2008) A cell-penetrating peptidic GRP78 ligand for tumor cell-specific prodrug therapy. Bioorg Med Chem Lett 18:1632–1636
32. Curnis F, Sacchi A, Borgna L, Magni F, Gasparri A, Corti A (2000) Enhancement of tumor necrosis factor α antitumor immunotherapeutic properties by targeted delivery to aminopeptidase N (CD13). Nat Biotechnol 18:1185–1190
33. Curnis F, Arrigoni G, Sacchi A, Fischetti L, Arap W, Pasqualini R, Corti A (2002) Differential binding of drugs containing the NGR motif to CD13 isoforms in tumor vessels, epithelia, and myeloid cells. Cancer Res 62:867–874
34. 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 Bioinformat 8:245
35. 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
36. Jung E, Choi SH, Lee NK, Kang SK, Choi YJ, Shin JM, Choi K, Jung DH (2011) Machine learning study for the prediction of transdermal peptide. J Comput Aided Mol Des 25:339–347
37. Inaba K, Inaba M, Romani N, Aya H, Deguchi M, Ikehara S, Muramatsu S, Steinman RM (1992) Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. J Exp Med 176:1693–1702
38. Shepherd DM, Steppan LB, Hedstrom OR, Kerkvliet NI (2001) Anti-CD40 treatment of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)-exposed C57BL/6 mice induces activation of antigen presenting cells yet fails to overcome TCDD-induced suppression of allograft immunity. Toxicol Appl Pharmacol 170:10–22
39. McCaldon P, Argos P (1988) Proteins 4:99–122
40. 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
41. Pipeline Pilot 8.0 (http://accelrys.com/products/pipeline-pilot/)
42. The R Project for Statistical Computing (http://www.r-project.org/)
43. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36
44. Discovery Studio 3.1 (http://accelrys.com/products/discovery-studio/)
45. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89:10915–10919
46. Gauthier MA, Klok HA (2008) Peptide/protein–polymer conjugates: synthetic strategies and design concepts. Chem Commun 23:2591–2611
47. Molek P, Strukelj B, Bratkovic T (2011) Peptide phage display as a tool for drug discovery: targeting membrane receptors. Molecules 16:857–887
48. 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
49. 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
50. Bhattacharya SK, Peachey NS, Crabb JW (2005) Cochlin and glaucoma: a mini-review. Vis Neurosci 22:605–613
51. Tuma PL, Hubbard AL (2003) Transcytosis: crossing cellular barriers. Physiol Rev 83:871–932
52. Favier B, Alam A, Barron P, Bonnin J, Laboudie P, Fons P, Mandron M, Herault JP, Neufeld G, Savi P, Herbert JM, Bono F (2006) Neuropilin-2 interacts with VEGFR-2 and VEGFR-3 and promotes human endothelial cell survival and migration. Blood 108:1243–1250