Research Article

Journal of PeptideScience

Received: 8 February 2010 Revised: 6 May 2010 Accepted: 10 June 2010 Published online in Wiley Online Library: 16 September 2010

(wileyonlinelibrary.com) DOI 10.1002/psc.1274

Modeling and predicting interactions between the human amphiphysin SH3 domains and their peptide ligands based on amino acid information

Jianfeng Cai,[a‡] Rongying Ou,[a‡] Yun-Sheng Xu,[a*] Li Yang,[b] Zhihua Lin[c] and Mao Shu[b,c*‡]

In this paper, VHESH, which was a novel set of amino acid descriptors including hydrophobic, electronic, steric, and hydrogen bond contribution properties, were proposed to characterize the structures of the decapeptides binding the human amphiphysin-1 Src homology 3 (SH3) domains, and QSAR model was constructed by partial least square (PLS) with genetic algorithm-variable selection. It was found that diversified properties of the residues between P_2 and P_{-3} (including P_2 and P_{-3}) of the decapeptide ($P_4P_3P_2P_1P_0P_{-1}P_{-2}P_{-3}P_{-4}P_{-5}$) may contribute remarkable effect to the interactions between the SH3 domain and decapeptides. Particularly, hydrogen bond and steric properties of P_2 and electronic properties, steric properties of P_{-3} may provide relatively large positive contributions to the interactions. Based on the GA-PLS model, a series of decapeptides, with relatively high binding affinities were designed. These results showed that VHESH descriptors can well represent the decapeptides. Furthermore, the model obtained, which showed low computational complexity, correlated VHESH descriptors with the binding affinities as well as that VHESH may also be applied in QSAR studies of peptides. Copyright © 2010 European Peptide Society and John Wiley & Sons, Ltd.

Supporting information may be found in the online version of this article

Keywords: human amphiphysin SH3 domain; peptide; VHESH descriptor; genetic algorithm; partial least square; QSAR

Introduction

Protein–protein interactions play critical roles in regulating many key biological processes. Many important protein–protein interactions are mediated by peptide recognition modular domains, among which Src homology 3 (SH3) domain is the most abundant modular domain in the human proteome and presents in a wide variety of proteins, such as kinases, GTPases, lipases, and adaptor proteins, to lead diverse cellular processes [1–3]. SH3 domains are small (55–70 amino acids) noncatalytic protein modules that mediate protein–protein interactions by binding to Pro-rich peptide sequences [4–6]. They recognize the proline-rich peptides with the consensus motif PXXP (where P is proline and X is any amino acid). Thereinafter, many other peptide recognition domains have also been characterized for their roles in signal transduction by mediating weak and transient protein–protein interactions [7,8]. Human amphiphysin exists as two similar proteins, amphiphysin-1 and amphiphysin-2, which may be concerned with clathrin-mediated endocytosis, actin function, and signaling pathways [9]. The specific interaction between the amphiphysin-1 SH3 domain and dynamin plays a crucial role in endocytic function.

Identification of the sequence motifs recognized by different SH3 domains is an important step in understanding protein–peptides interactions, and peptide library experiments are often used to accomplish this purpose [10], it is time-consuming and costly to synthesize all ten-residue-long peptides appearing in the human genome. Therefore, it is very important to develop effective computation method and acquire some binding information. A rigorous computational method concerning the peptide–SH3 domain binding free energy for future recognition binding sequences has been reported [11]. However, accurate calculation of binding free energy is very complicated. In addition, constructing protein–peptide systems are also time-consuming. QSAR provides a practical tool for exploring the peptide–SH3 relationship. Previously, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) are used to explore the interactions between MHC proteins and their peptide ligands coordination, and to acquire models

* *Correspondence to: Yun-Sheng Xu, Department of Dermatovenereology – the First Affiliated Hospital of Wenzhou Medical College, Institute of Dermatovenereology, Wenzhou Medical College, Wenzhou 325000, P.R. China. E-mail: xuyunsh@sohu.com*

Mao Shu, College of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing 400050, P.R. China. E-mail: sm7507@126.com

a *Department of Dermatovenereology – the First Affiliated Hospital of Wenzhou Medical College, Institute of Dermatovenereology, Wenzhou Medical College, Wenzhou 325000, P.R. China*

b *College of Bioengineering, Chongqing University, Chongqing 400030, P.R. China*

c *College of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing 400050, P.R. China*

‡ *These authors contributed equally to this study.*

Table 1. VHESH scales for amino acids

AAs	$VHESH_1$	$VHESH_2$	$VHESH_3$	$VHESH_4$	$VHESH_5$	$VHESH_6$	$VHESH_7$	$VHESH_8$	$VHESH_9$	$VHESH_{10}$
Ala (A)	0.18	−1.35	−0.37	−0.9	−0.42	0.28	−0.9	−1.19	−1.13	−0.01
Arg (R)	−1.34	1.28	0.44	2.09	−1.81	−0.2	0.47	1.83	2.56	0.07
Asn (N)	−1.04	0.07	0.99	0.35	0.34	0.06	−0.6	0.67	1.01	−0.18
Asp (D)	−1.06	−0.15	2.73	−0.47	1.15	0.13	−0.8	0.76	0.72	−0.14
Cys (C)	0.81	−1.07	−0.23	0.68	1.95	−0.14	−0.17	−1.26	−0.99	1.5
Gln (Q)	−1.06	0.21	0.36	0.3	−0.27	0.49	−0.26	0.99	1.12	0.81
Glu (E)	−1.04	0.33	2	−1.28	−0.35	0.49	−0.39	0.99	0.82	1.22
Gly (G)	−0.16	−1.69	0.01	−1.7	−1.32	0.36	−2.37	−1.01	−1.25	−0.73
His (H)	−0.36	0.41	0.38	1.34	−0.13	−0.92	0.34	0.23	0.25	0.14
Ile (I)	1.39	−0.14	−1.2	−0.95	−0.39	0.5	0.92	−1.19	−0.64	−0.72
Leu (L)	1.42	−0.37	−1.03	−0.79	−0.23	0.66	0.82	−0.9	−0.6	−1.72
Lys (K)	−1.27	0.65	0.07	1.32	−2.11	−0.09	−0.07	1.81	1.28	−1.58
Met (M)	0.85	−0.27	−0.61	0.24	0.64	0.6	0.82	−0.63	−0.58	−1.04
Phe (F)	1.49	0.68	−0.73	0.59	1.18	0.34	1.31	−0.73	−0.81	−0.27
Pro (P)	−0.35	−0.68	−0.57	−0.88	0.22	−3.9	−0.68	0.08	−0.97	1.71
Ser (S)	−0.72	−0.82	0.29	−0.23	0.36	−0.18	−1.2	0.01	−0.08	0.66
Thr (T)	−0.37	−0.6	−0.09	−0.17	−0.05	−0.08	−0.62	0.01	0.1	0.07
Trp (W)	1.18	2.39	−0.94	0.86	1	0.77	1.8	0.15	−0.13	−1.36
Tyr (Y)	0.37	1.62	−0.51	0.59	0.7	0.38	1.17	0.68	0.14	1.27
Val (V)	1.06	−0.48	−0.99	−0.98	−0.46	0.46	0.42	−1.29	−0.83	0.29

with good prediction ability [12]. Hou *et al*. [13] investigated the interactions between the SH3 domain and its peptide ligands, and constructed satisfying QSAR models using homology modeling, molecular dynamics, and molecular field analysis. Liang *et al*. [14] employed amino acid information to characterize the primary sequence of peptide, and built satisfying QSAR models using partial least square (PLS). In this study, a novel amino acid descriptors, VHESH (involving hydrophobic properties, steric properties, electronic properties, and hydrogen bond contribution), was acquired and optimized based on previous work [15,16], and was then employed to represent the structures of a set of decapeptides, and some information concerning the interactions between SH3 domains and decapeptides was acquired based on QSAR model which was constructed by PLS with genetic algorithm-variable selection.

Principles and Methods

VHESH Descriptors and Structural Characterization

Considering that the nonbonding factors are influential to protein–peptide binding, 50 hydrophobic properties, 23 electronic properties, 35 steric properties, and 5 hydrogen bond contribution properties of 20 coded amino acids properties were selected from AA index data (supporting information Table S1). Then, the four categories of properties matrices were processed by principal components analysis (PCA), respectively. For the matrices of hydrophobic, electronic, steric, and hydrogen bond contribution properties, the top two, four, two, and two significant principal components accounting for 71.40, 77.74, 72.95, and 77.76% variance of original data matrices, respectively. These ten principal components can replace the corresponding original data matrices with less losing information. Here, tentatively called the ten score vectors as VHESH (principal component score Vector of Hydrophobic, Electronic, Steric, and Hydrogen bond properties) for short, VHESH1–VHESH2 were relate to hydrophobic properties, VHESH3–VHESH6 to electronic properties, VHESH7–VHESH8 to steric properties, and VHESH9–VHESH10 to hydrogen bonds contribution properties (Table 1). PCA is done with statistics software package SPSS 13.0.

For a set of peptide analogs, the peptide sequence can be now quantified by ten VHESH variables. Here, a dipeptide is exemplified: each of ten VHESH specially corresponded to each amino acid site and there would be 20 VHESH scales ($2 \times 10 = 20$) represent the dipeptide molecule. Accordingly, peptide sequence with n amino acids will generate $n \times 10$ variables.

Variables Selection and PLS Modeling

In a QSAR data set, not all the structural descriptors were relevant to biological activity; therefore, those redundant descriptors should be deleted from the model in order to promote its predictive capability. Here, variables selection method is carried out by GA–PLS as a popular variables selection tool, which is a sophisticated hybrid approach combining GA as a powerful optimization method with PLS as a robust statistical method for variables selection [17,18]. GA–PLS procedure includes five steps: (i) the initial population of chromosomes is created by a random value; (ii) the fitness of each chromosome is evaluated by the internal prediction of the PLS model; (iii) the chromosome with the least number of variables and the highest fitness is marked as an informative chromosome; (iv) GA manipulation including crossover, mutation, and replication is carried out; (v) return to Step 2 and repeat Steps 2–4 unless the optimal chromosomes are achieved.

PLS regression is a widely used modeling method, which can avoid harmful effects in modeling due to multicollinearity, and is particularly fit for regression when the number of observations is less than the number of variables. PLS regression combines basic functions of regression model, PCA, and canonical correlation analysis [19,20]. And an excellent model should have both favorable estimation ability for any internal sample and

Journal of PeptideScience

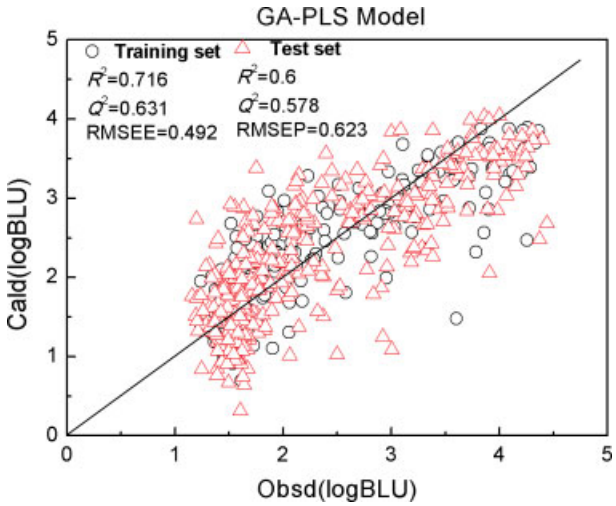

Figure 1. Calculated *versus* observed logarithm BLU values for the SH3 domain binding peptide. Training set samples are depicted by circles, test set samples by squares.

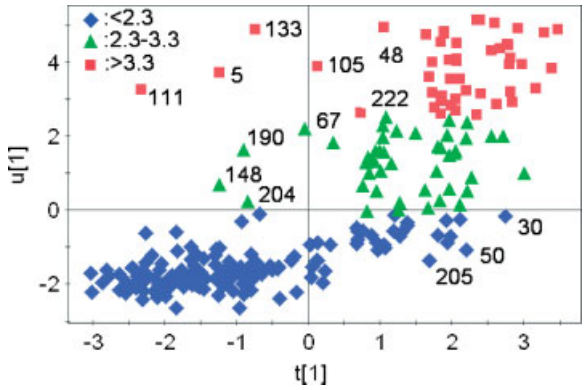

Figure 2. Plots of the GA–PLS scores *t*1 *versus* *u*1 for the SH3 binding decapeptides (different biological activities with different marks).

outstanding predictive ability for any external sample [21]. PLS is done with statistics software package SIMCA-P 10.0.

Dataset and Sample Selection

A set of decapeptides dataset was obtained from the report by Landgraf *et al*. [22]. A combination of phage display and SPOT synthesis was used to find all peptides in the yeast proteome binding to eight yeast SH3 domains, and peptides in the human proteome binding to two human SH3 domains by Landgraf *et al*. [22]. For each domain, peptides matching the defined patterns were synthesized at high density on cellulose membranes by SPOT synthesis technology, and the membranes were probed with the corresponding SH3 domain fused to glutathione-*s*-transferase (GST). To enhance the data quality, we set two criteria to ensure the reliability of selected samples: (i) peptides should have at least two experimental results, and (ii) the SD of repeated experiments should be less than the half of average value. As a result, 592 decapeptides were selected for QSAR modeling. Among 592 peptides data, 243 peptides with at least three experimental binding values were regarded as training set to construct QSAR model, and remaining 349 peptides were treated as test set to validate the external predictive power of the QSAR model.

Results and Discussion

The optimal PLS model of training set is obtained by GA–PLS techniques. Parameters were set as follows: the number of population was 200, the maximum number of generations was 200, the generation gap was 0.8, the crossover frequency was 0.5, the mutation rate was 0.005, and the fitting function was Q^2_{cv}. An optimal model including 39 variables was obtained from all trained 10 models. As a result, a QSAR model was constructed by extracting two significant principal components, cumulatively explaining the 71.2% variance of *Y* variable with cross-validation achieving 63.1%. The relative statistics are listed as $R^2 = 0.712$, $Q^2 = 0.631$, $Q_{ext}{}^2 = 0.578$, RMSEE = 0.492, and RMSEP = 0.623. Figure 1 is the plots of GA–PLS-calculated predicted *versus* observed logarithm BLU values for the SH3 domain binding peptide (training set samples are depicted by circles, test set samples by squares), wherein most samples are uniformly dispersed along a line passing through the origin and forming an angle of 45°. It was shown a robust model.

Figure 2 is the plots of *t*1 *versus* *u*1 in the PLS model (*t*1 and *u*1 indicate the first principal component in the *X* and *Y* scoring space, respectively) with different biological activities for three marking symbols: square marks samples with the potency was greater than 3.3, while triangle marks samples with the potency between 2.3 and 3.3 and rhombus marks samples with the potency was less than 2.3. It can be seen that distribution of decapeptide sequences exhibit an increasing trend from the root left corner to the top right corner in terms of their activities, and except for 13 samples (Nos. 5, 30, 48, 50, 67, 105, 111, 133, 148, 190, 204, 205, and 222), most compounds with similar bioactivities are close together. By further analysis, it is found that high binding decapeptides (including Nos. 5, 48, 105, 111, 133, and 222) possessing arginine residue in fifth residue of the sequences, the middle binding decapeptides (including Nos. 67, 148, 190, and 204) having arginine or lysine residue in the first or fifth residue of the sequence, and the low ones (including Nos. 30, 50, 199, and 205) are arginine or lysine in the fifth or eighth residue of decapeptide. These positive charge amino acid residues may take key roles in the SH3 domain–peptide interaction, hence, all samples were carefully reserved in the model.

To validate normal hypothesis, we then implement the normal probability of the standardized residual for the regression model. Figure 3 shows that most of the residue errors follow a normal distribution, with the only exceptions of samples 5, 105, 111, and 133, for which the standardized residues (SD) are beyond the range of 2.5 [23]. So, the normal hypotheses are confirmed to be true. It can be seen from the 20-random-permutation validation of the PLS model (Figure 4) that intercepts of R^2_{cum} and Q^2_{cv} are 0.047 and −0.225, respectively. Therefore, it is considered that relatively high values of R^2_{cum} and Q^2_{cv} are not resulted from accidental factors.

According to the report by Hou *et al*. [13], the positions of the decapeptide are defined as: $P_4P_3P_2P_1P_0P_{-1}P_{-2}P_{-3}P_{-4}P_{-5}$. Description, coefficient, and variable importance in the projection (VIP) of the variables of the PLS model are summarized in Table 2. VIP is the sum of the variable influence over all model dimensions and is a measure of variable importance. Higher VIP values indicate good correlation between the variable and the model. It can be seen that ten variables, including electronic properties of P_{-3}, hydrogen bond properties of P_{-3}, steric properties of P_{-3}, steric properties of P_0, steric properties of P_1, steric properties of P_2,

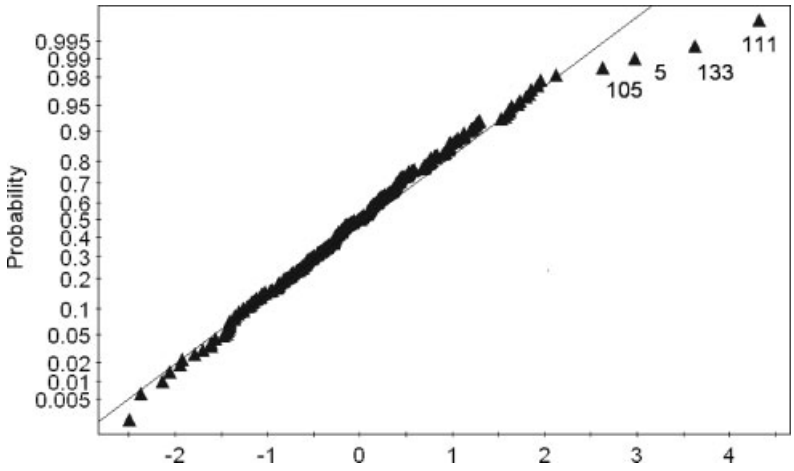

Figure 3. Plots of cumulative probability of standardized residuals.

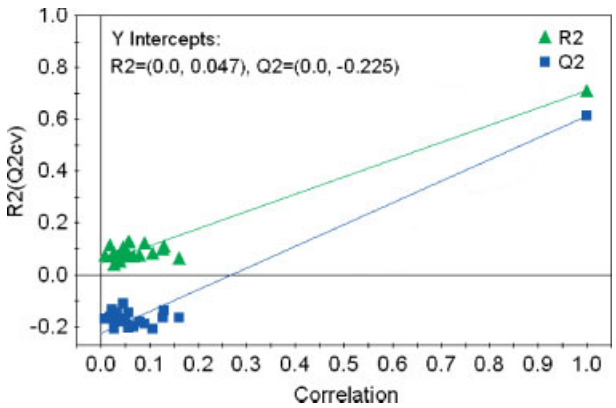

Figure 4. *Y* random permutations test in the GA–PLS model.

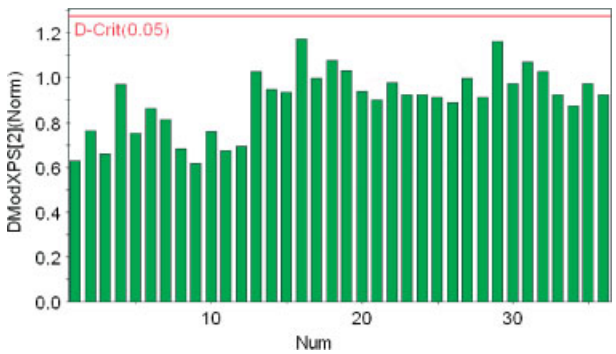

Figure 5. Distance of *X* space for predicted samples.

Table 2. Description, coefficient, and VIP of the variables for GA–PLS model

No.	Variable	Site	Variables property	Coefficient	VIP
1	v_{10}	P_4	Hydrogen bond ($VHESH_{10}$)	0.033	0.413
2	v_{11}	P_3	Hydrophobic ($VHESH_1$)	0.014	0.158
3	v_{13}	P_3	Electronic ($VHESH_3$)	−0.065	0.424
4	v_{14}	P_3	Electronic ($VHESH_4$)	0.043	0.421
5	v_{15}	P_3	Electronic ($VHESH_5$)	−0.072	0.408
6	v_{17}	P_3	Steric ($VHESH_7$)	0.040	0.286
7	v_{19}	P_3	Hydrogen bond ($VHESH_9$)	0.032	0.410
8	v_{28}	P_2	Steric ($VHESH_8$)	0.179	1.185
9	v_{30}	P_2	Hydrogen bond ($VHESH_{10}$)	0.162	1.133
10	v_{37}	P_1	Steric ($VHESH_7$)	0.064	0.984
11	v_{38}	P_1	Steric ($VHESH_8$)	−0.043	1.442
12	v_{40}	P_1	Hydrogen bond ($VHESH_{10}$)	0.081	0.781
13	v_{47}	P_0	Steric ($VHESH_7$)	0.135	0.760
14	v_{64}	P_{-2}	Electronic ($VHESH_4$)	0.004	0.278
15	v_{65}	P_{-2}	Electronic ($VHESH_5$)	0.026	0.191
16	v_{66}	P_{-2}	Electronic ($VHESH_6$)	−0.087	0.550
17	v_{68}	P_{-2}	Steric ($VHESH_8$)	0.035	0.274
18	v_{74}	P_{-3}	Electronic ($VHESH_4$)	0.258	2.421
19	v_{75}	P_{-3}	Electronic ($VHESH_5$)	−0.178	2.122
20	v_{77}	P_{-3}	Steric ($VHESH_7$)	0.128	1.400
21	v_{79}	P_{-3}	Hydrogen bond ($VHESH_9$)	0.281	2.472
22	v_{80}	P_{-3}	Hydrogen bond ($VHESH_{10}$)	0.159	1.169
23	v_{81}	P_{-4}	Hydrophobic ($VHESH_1$)	−0.063	0.361
24	v_{83}	P_{-4}	Electronic ($VHESH_3$)	−0.016	0.171
25	v_{84}	P_{-4}	Electronic ($VHESH_4$)	0.093	0.544
26	v_{90}	P_{-4}	Hydrogen bond ($VHESH_{10}$)	0.063	0.383
27	v_{93}	P_{-5}	Electronic ($VHESH_3$)	−0.081	0.452
28	v_{96}	P_{-5}	Electronic ($VHESH_6$)	−0.005	0.067
29	v_{100}	P_{-5}	Hydrogen bond ($VHESH_{10}$)	−0.003	0.057
30	Constant	–	–	2.522	–

and hydrogen bond properties of P_2, have the relatively larger VIP values. These ten variables concern with the sites between P_2 and P_{-3} (including P_2 and P_{-3}), indicating that middle six residues of decapeptide are closely correlated with the SH3 domains.

According to the coefficients of the PLS model equation (Table 2), we can evaluate the contributions of various properties of each peptide residue to the binding affinities. At the P_{-3}, five VHESH descriptors were selected, representing electronic, steric property, and hydrogen bond contribution properties(v_{74}, v_{75}, v_{77}, v_{79}, and v_{80}). Among the electronic and hydrogen bond properties (v_{74}, v_{79}, and v_{80}) are remarkable with interactions between the peptides and the SH3 domains. Cesareni *et al*. [24] suggest that P_{-3} is essential to the binding specificity.

The VIP values of hydrogen bond and steric properties of P_2 are more than 1, and their corresponding coefficient are positive in correlation with the GA–PLS model. Through the above analysis, we can speculate that the properties of P_2 may provide the

Journal of PeptideScience

Table 3. Preferred amino acids at positions which are important for affinities of decapeptide–SH3 domain

Sequence position	Property	Preferred amino acids	Contribution to activity
P_2	Steric	R, K, E, Q, L	+
P_2	Hydrogen bond	C, E, P, Y	+
P_1	Steric	A, C, G, I, V	−
P_0	Steric	Y, W, F, I	+
P_{-3}	Electronic	R, K, H, W	+
P_{-3}	Electronic	R, K, G	−
P_{-3}	Steric	Y, W, F, I	+
P_{-3}	Hydrogen bond	R, N, Q, K	+
P_{-3}	Hydrogen bond	C, E, P, Y	+

+, positive; −, negative.

most contributions to the interactions between the peptides and SH3 domains. In addition, Hou *et al*. [13] have reported that the proline at P_2 position has strong interactions with the SH3 domain. In addition, coefficients of hydrogen bond and steric properties at P_1, and coefficient steric properties at P_1 are also positive contribution to the model. Preferred amino acids at positions that are important for binding decapeptides containing five residues are shown in Table 3. We can obtain peptides with demanding decapeptide derivates by alteration of these important amino acid residues.

Loading contributions, coefficients, and VIPs of PLS model show that variables including steric, hydrogen bond, and electronic properties of P_2 and P_{-3} of decapeptide play important roles for interactions between peptide and SH domain, to a certain extent which supports the conclusion that fundamental action mode of P_2 and P_{-3} are anchoring to the SH3 domain [13,24].

Thus, the properties of P_2 and P_{-3} of peptide sequence must be considered when designing a new peptide sequence. For example, derivatives of the SH3 binding peptide sequence (corresponding to P81408, O94885, Q9NNY5, and Q16632) were obtained by modifying the P_2 and P_{-3} residue of the sequence. Table 4 presents the new peptide sequences and their prediction binding affinities. The results showed that most derivatives possess relatively high antimicrobial activity, which may be due to the no charge amino acid residue proline (P) replaced by positive charge ones (arginine (R), lysine (K)), which increase electronic properties, hydrogen bond properties, and binding affinity.

Plots of the distance to the PLS model in the *X* space are depicted in Figure 5 to investigate the efficiency on recombination for new sequences. It can be seen from it that the only two samples overflow normalized distance to *X* of GA–PLS model critical value of 1.276 at 5% significance level [25]. It shows that new design decapeptide sequences are adapted to application domain of the GA–PLS model, so some predicted sequences with high binding affinity are considered to synthesis and test their binding affinities.

Conclusion

The VHESH descriptors were constructed by multivariate statistical analysis and applied in the QSAR study of the decapeptides binding the human amphiphysin-1 SH3 domains. By analysis, steric, hydrogen bond, and electronic properties of decapeptide play important roles on binding affinity. Especially, these properties of P_2 and P_{-3} of decapeptide make high contribution to binding affinity. New derivatives of decapeptide (corresponding to P81408, O94885, Q9NNY5, and Q16632) were obtained by modifying the P_2 and P_{-3} residues of the sequence and display that some derivatives possess relatively high binding affinity. The results showed that VHESH descriptors have strong characterization capability, easy

Table 4. The new sequences of SH3 domain binding decapeptide with predicted activities

Identity	Peptide	Obsd	Pred	Identity	Peptide	Obsd	Pred
P81408	PLPRRPPRAA	4.37	3.77	Q9NNY5	LPPKRPIKEV	2.57	2.55
P81408-1	PLRRRPPRAA	ND	4.17	Q9NNY5-1	LPPKRPIREV	ND	3.23
P81408-2	PLRRRPPKAA	ND	3.49	Q9NNY5-2	LPRKRPIKEV	ND	2.94
P81408-3	PLRRRPPHAA	ND	3.33	Q9NNY5-3	LPRKRPIREV	ND	3.62
P81408-4	PLRRRPPYAA	ND	3.36	Q9NNY5-4	LPRKRPIYEV	ND	2.81
P81408-5	PLKRRPPRAA	ND	4.00	Q9NNY5-5	LPKKRPIREV	ND	3.45
P81408-6	PLERRPPRAA	ND	4.01	Q9NNY5-6	LPEKRPIKEV	ND	2.79
P81408-7	PLERRPPKAA	ND	3.34	Q9NNY5-7	LPEKRPIREV	ND	3.46
P81408-8	PLQRRPPRAA	ND	3.97	Q9NNY5-8	LPQKRPIREV	ND	3.42
P81408-9	PLYRRPPRAA	ND	3.92	Q9NNY5-9	LPYKRPIREV	ND	3.37
O94885	EKPKRPTRRR	4.36	3.85	Q16632	QSPKRPPEDF	1.62	2.35
O94885-1	EKRKRPTRRR	ND	4.24	Q16632-1	QSPKRPPRDF	ND	3.48
O94885-2	EKRKRPTKRR	ND	3.56	Q16632-2	QSRKRPPKDF	ND	3.20
O94885-3	EKRKRPTHRR	ND	3.41	Q16632-3	QSRKRPPRDF	ND	3.87
O94885-4	EKRKRPTYRR	ND	3.43	Q16632-4	QSRKRPPYDF	ND	3.06
O94885-5	EKKKRPTRRR	ND	4.07	Q16632-5	QSKKRPPRDF	ND	3.70
O94885-6	EKKKRPTKRR	ND	4.07	Q16632-6	QSEKRPPRDF	ND	3.04
O94885-7	EKQKRPTRRR	ND	4.04	Q16632-7	QSEKRPPHDF	ND	3.72
O94885-8	EKCKRPTRRR	ND	3.40	Q16632-8	QSQKRPPRDF	ND	3.68
O94885-9	EKYKRPTRRR	ND	3.99	Q16632-9	QSQKRPPFDF	ND	3.62

ND, no data.

operation, and clear physicochemical significance, and further suggest broad application prospects in the QSAR field of peptides.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Nos. 60873103, 30500537, and 30972800), and supported by the Key Project of Natural Science Foundation of China (No. 30830090), and supported by Program for New Century Excellent Talents in University (No. NCET-06-0780).

References

1 Lim WA, Richards FM, Fox RO. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* 1994; **372**: 375–379.
2 Feng SB, Chen JK, Yu H, Simon JA, Schreiber SL. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3–ligand interactions. *Science* 1994; **266**: 1241–1247.
3 Kuriyan J, Cowburn D. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* 1997; **26**: 259–288.
4 Hou TJ, Zhang W, Case DA, Wang W. Characterization of domain–peptide interaction interface: a case study on the amphiphysin-1SH3 domain. *J. Mol. Biol.* 2008; **376**: 1201–1214.
5 Cohen GB, Ren R, Baltimore D. Modular binding domains in signal transduction proteins. *Cell* 1995; **80**: 237–248.
6 Dalgarno DC, Botfield MC, Rickles RJ. SH3 domains and drug design: ligands, structure, and biological function. *Biopolymers* 1997; **43**: 383–400.
7 Castagnoli L, Costantini A, Dall'armi C, Gonfloni S, Montecchi-Palazzi L, Panni S. Selectivity and promiscuity in the interaction network mediated by protein recognition modules. *FEBS Lett.* 2004; **567**: 74–79.
8 Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. *Science* 2003; **300**: 445–452.
9 Slepnev VI, Ochoa GC, Butler MH, Grabs D, De Camilli P. Role of phosphorylation in regulation of the assembly of endocytic coat complexes. *Science* 1998; **281**: 821–824.
10 Rickles RJ, Botfield MC, Zhou X, Henry PA, Brugge JS. Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc. Natl. Acad. Sci. USA* 1995; **92**: 10909–10913.
11 Wang W, Lim WA, Jakalian A, Wang J, Wang JM, Luo R, Bayly CI, Kollman PA. An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. *J. Am. Chem. Soc.* 2001; **123**: 3986–3994.
12 Doytchinova IA, Flower DR. Toward the quantitative prediction of T-cell epitopes: CoMFA and CoMSIA studies of peptides with affinity for the class I MHC molecule HLA-A*0201. *J. Med. Chem.* 2001; **44**: 3572–3581.
13 Hou TJ, McLaughlin W, Lu BZ, Chen K, Wang W. Prediction of binding affinities between the human amphiphysin-1 SH3 domain and its peptide ligands using homology modeling, molecular dynamics and molecular field analysis. *J. Proteome Res.* 2006; **5**: 32–43.
14 Liang G, Chen G, Niu W, Li Z. Factor analysis scales of generalized amino acid information as applied in predicting interactions between the human amphiphysin-1 SH3 domains and their peptide ligands. *Chem. Biol. Drug Des.* 2008; **71**: 345–351.
15 Mei H, Liao Z, Zhou Y, Li Z. A new set of amino acid descriptors and its application in peptide QSARs. *Biopolymers (Pept. Sci.)* 2005; **80**: 775–786.
16 Shu M, Mei H, Yang S, Liao L, Li Z. Structural parameter characterization and bioactivity simulation based on peptide sequence. *QSAR Comb. Sci.* 2009; **28**: 27–35.
17 Hasegawa K, Miyashita Y, Funatsu K. GA strategy for variable selection in QSAR studies: GA based PLS analysis of calcium channel antagonists. *J. Chem. Inf. Comput. Sci.* 1997; **37**: 306–310.
18 Hasegawa K, Kimura T, Funatsu K. GA strategy for variable selection in QSAR studies: enhancement of comparative molecular binding energy analysis by GA-based PLS method. *Quant. Struct. Act. Relat.* 1999; **18**: 262–272.
19 Sjöström M, Rännar S, Wieslander Å. Polypeptide sequence property relationships in *Escherichia coli* based on auto cross covariances. *Chemomet. Intell. Lab. Syst.* 1995; **29**: 295–305.
20 Wold S, Trygg J, Berglund A, Antti H. Some recent developments in PLS modeling. *Chemomet. Intell. Lab. Syst.* 2001; **58**: 131–150.
21 Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* 2003; **22**: 69–77.
22 Landgraf C, Panni S, Montecchi-Palazzi L. Protein interaction networks by proteome peptide scanning. *PLoS Biol.* 2004; **2**: 94–103.
23 Wold S, Sjöström M, Eriksson L. PLS regression: a basic tool of chemometrics. *Chemomet. Intell. Lab. Syst.* 2001; **58**: 109–130.
24 Cesareni G, Panni S, Nardelli G, Castagnoli L. Can we infer peptide recognition specificity mediated by SH3 domains? *FEBS Lett.* 2002; **513**: 38–44.
25 Cho SJ, Zheng W, Tropsha A. Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches. *J. Chem. Inf. Comput. Sci.* 1998; **38**: 259–268.

