research articles Journal of proteome research

Prediction of Binding Affinities between the Human Amphiphysin-1 SH3 Domain and Its Peptide Ligands Using Homology Modeling, Molecular Dynamics and Molecular Field Analysis

Tingjun Hou, William McLaughlin, Benzhuo Lu, Ken Chen, and Wei Wang*

Department of Chemistry and Biochemistry and Center for Theoretical Biological Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0319

Received July 19, 2005

The SH3 domain of the human protein amphiphysin-1, which plays important roles in clathrin-mediated endocytosis, actin function and signaling transduction, can recognize peptide motif PXRPXR (X is any amino acid) with high affinity and specificity. We have constructed a complex structure of the amphiphysin-1 SH3 domain and a high-affinity peptide ligand PLPRRPPRA using homology modeling and molecular docking, which was optimized by molecular dynamics (MD). Three-dimensional quantitative structure-affinity relationship (3D-QSAR) analyses on the 200 peptides with known binding affinities to the amphiphysin-1 SH3 domain was then performed using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). The best CoMSIA model showed promising predictive power, giving good predictions for about 95% of the peptides in the test set (absolute prediction errors less than 1.0). It was used to validate peptide-SH3 binding structure and provide insight into the structural requirements for binding of peptides to SH3 domains. Finally, MD simulations were performed to analyze the interaction between the SH3 domain and another peptide GFPRRPPPRG that contains with the PXRPXsR (s represents residues with small side chains) motif. MD simulations demonstrated that the binding conformation of GFPRRPPPRG is quite different from that of PLPRRPPRAA especially the four residues at the C terminal, which may explain why the CoMSIA model cannot give good predictions on the peptides of the PXRPXsR motif. Because of its efficiency and predictive power, the 3D-QSAR model can be used as a scoring filter for predicting peptide sequences bound to SH3 domains.

Keywords: amphiphysin-1 • SH3 domain • protein−protein interaction • molecular dynamics • CoMFA • CoMSIA • quantitative structure-affinity relationship (QSAR) • molecular field analysis

Introduction

Protein−protein interactions are often mediated by small modular domains, among which Src Homology 3 (SH3) domain is the most abundant in the human genome and has been well characterized. SH3 domains are 50−70 amino acids long and share a highly conserved fold.[1] They are present in proteins with critical roles in signal transduction, cytoskeleton organization and other important biological processes.[2−5] The initial understanding of the nature of SH3 domain ligands was derived from two proteins, 3BP1 and 3BP2, which bind to the Abl SH3 domain. Mutagenesis studies on these proteins has identified the consensus sequence XPXXPPPψXP (X represents any residue and ψ represents hydrophobic residues) to which the Abl SH3 domain binds.[6] It was shown that individual peptides with these sequences could bind to the Abl- and Fyn-SH3 domains with similar affinities (5−40 μM).[7]

Experimental results using combinatorial peptide libraries show that most SH3 domain binding peptides possess the sequence motif PXXP (X represents any residue) and bind to SH3 domains in a left-handed polyproline helix type II conformation.[8] Isolated SH3 domains can bind to contiguous proline-rich ligands, characterized by a core region of 7−9 amino acid long.[1] The two proline residues in the PXXP motif occupy two hydrophobic pockets, formed by aromatic residues that are conserved in most SH3 domains. The third binding pocket is lined by negative-charged residues and thus accommodates a positively charged side-chain. The orientation of peptide ligands depends on the position of the positive-charged residue in the peptides. SH3 domain binding peptides can be divided into two classes, which contain +XXPXXP and PXXPX+ (where + refers to a positively charged residue) motifs, respectively. Figure 1 shows the binding position and the nomenclature of the two classes of peptides. Pro is preferred at position P_0 and P_3 in class I ligands and position P_{-1} and P_2 in class II ligands.[8]

Identification of the sequence motifs recognized by different SH3 domains is a crucial step in understanding the protein-peptide interactions and peptide library experiments are often used to serve this purpose.[9−11] Although the peptide library has

* To whom correspondence should be addressed. Tel: (858) 822-4240. Fax: (858) 822-4236. E-mail: wei-wang@ucsd.edu.

10.1021/pr0502267 CCC: $33.50
Published on Web 12/15/2005

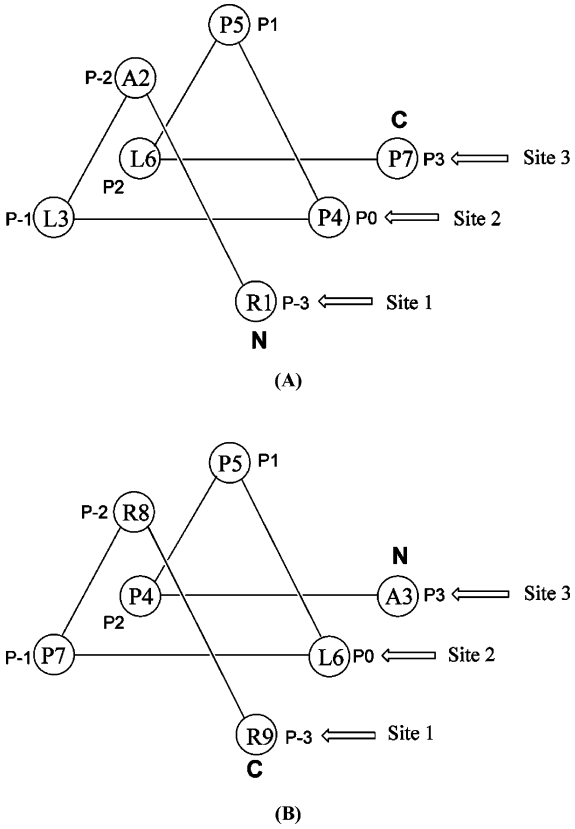

Figure 1. Diagram of the binding position and nomenclature of class I peptides (A) and class II peptides (B) following the nomenclature proposed by Lim et al.[8] The sequences in Figure A and B are the c-Src RLP2 ligand and c-Src PLR1 ligand, respectively.[12] The RLP2 ligand has the sequence RAL**P**PL**P**RY and the PLR1 ligand has the sequence AFA**P**PL**P**RR (sequence shown in bold are two proline residues in PXXP motif).

a limited coverage of all possible peptides of, for example, 10 residues (about 10^{-6} to 10^{-3}), it is too time-consuming and expensive to synthesize all 10-residue long peptides found in the human genome and perform the peptide-SH3 domain binding assay to identify the SH3 binding partners. Therefore, it is important to develop computational methods to derive such information. A rigorous computational approach would be developed to calculate the binding free energy between the SH3 domains and peptides to identify binding sequences.[13−15] However, accurate computation of binding free energy is far from trivial. Besides, it is quite time-consuming on protein-peptide systems. In the field of computer-aided drug design, the correlation between the structures and the biological activities of small organic molecules can be described by 2D- or 3D-QSAR methods. We hypothesized that these methods could also be used to characterize the structural properties of peptide ligands that bind to specific protein modular domains such as the SH3 domain. The previous study by Doytchinova et al. investigated the interactions between the MHC protein and its peptide ligands using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), and obtained models with good predictive power.[16] Here we have applied 3D-QSAR to study the interactions between the human amphiphysin-1 (hAmph1) SH3 domain and its peptide ligands.

Amphiphysin represents two very similar proteins, amphiphysins 1 and 2, which may be involved in clathrin-mediated endocytosis, actin function and signaling pathways.[17−19] Amphiphysin 1[20] has an SH3 domain at its C-terminus[21] that interacts with high affinity and specificity to a class II SH3 domain binding site PSRPNR in the proline-rich C terminus of dynamin.[22] The specific interaction between the SH3 domain of amphiphysin 1 and dynamin plays a critical role in endocytic function.[22] Since the structure of the human amphiphysin 1 (hAmph1) SH3 domain has not been solved, we built a model of the hAmph1 SH3 domain in complex with its peptide ligand using homology modeling and molecular dynamics (MD) simulations. We then aimed to obtain a predictive model of peptide binding affinity, which may be used to identify potential binding proteins of the hAmph1 SH3 domain. Two types of 3D-QSAR techniques, CoMFA[23] and CoMSIA,[24] were used in the study. CoMFA and CoMSIA are well-documented and validated techniques for the study of QSAR at the 3-D level.[25] However, CoMFA or CoMSIA is unable to appropriately describe all of the binding forces based only on steric and electrostatic molecular fields to model receptor−ligand interactions. The direct and complete representations of the binding forces between the peptide and the SH3 domain can only be fully described by the atomic level peptide/protein interactions derived from the complex structures. Therefore, we expect that combining the information obtained from receptor- and ligand-based approaches can give a full description of peptide-SH3 domain interactions.

		10	20	30	40
hAmph1 SH3	626	KVETLHDFEAANSDELTLQRGDVVLVVPSDSEADQDAGWL			
1bb9	46	KVQAQHDYTATDTDELQLKAGDVVLVIPFQNPEEQDEGWL			
hAmph1 SH3		VGVKESDWLQYRDLATYKGLFPENFTRR			
1bb9		MGVKESDWNQHKELEKCRGVFPENFTERVQ			

Figure 2. Sequence alignment between hAmph1 SH3 and 1bb9.

Materials and Methods

Homology Modeling of hAmph1 SH3 and Construction of the SH3/PLPRRPPRAA Complex. The structure of the hAmph1 SH3 domain was modeled based on the crystal structure of the rat amphiphysin-2 (rAmph2) SH3 domain (PDB entry 1bb9).[26] A sequence alignment using FASTA program[27] indicated that the percentage of sequence identity between the two domains was 55.9% (Figure 2). The prediction of the hAmph1 SH3 model was accomplished by using the Homology module in INSIGHTII.[28] The modeled structure was minimized using the CFF91 force field[29] implemented in the DISCOVER module in INSIGHTII. Then, the final model was quality-verified using the Profile-3D module in INSIGHTII[30] and PROCHECK.[31]

The 10 core residues (PLPRRPPRAA) of peptide ESPLPRRP-PRAARS reported by Landgraf et al.[32] was selected as the effective binder to construct the SH3/peptide complex. As described in the work reported by Cestra et al.,[33] the pXRPXR amphiphysin consensus (where lowercase p indicates that Pro is not absolutely conserved at that position) could be portrayed as a novel class II peptide ligand with an unusual positive charge at position P_0 and a tolerance for hydrophobic residue at P_2. Therefore, peptide ESPLPRRPPRAARS should adopt the class II binding orientation. The structure of the hAmph1 SH3 domain complexed with peptide PLPRRPPRAA was modeled based on the crystal structure of C−Crk N-terminal SH3 domain complexed with C3G peptide (PPPALPPKKR) (PDB entry: 1cka) as the template[34] using INSIGHTII and the *scap* program.[34] More detailed descriptions of the homology modeling of

hAmph1 SH3 and the construction of the SH3/PLPRRPPRAA complex can be found in Part1 and Part2 in the Supporting Information.

Molecular Dynamics Simulations of the SH3/Peptide Complex. MD simulations were employed to optimize the modeled complex structure as well as to investigate the binding mode of the peptide PLPRRPPRAA and the dynamic characteristics of the hAmph1 SH3 structure. All MD simulations presented in this work were performed using the AMBER8.0 software package[36] and the AMBER03 force field.[37] Counterions of Na+ were placed near the SH3 domain using a Coulombic potential on a grid to keep the whole simulated system neutral, and then the system was solvated in a rectangular box of about 3900 TIP3P water molecules.[38] The water box extended 9 Å away from any solute atom. Particle Mesh Ewald (PME) was employed to calculate the long-range electrostatic interactions.[39] Prior to MD simulations, a series of minimizations were preformed. All hydrogen atoms were first minimized, followed by all water molecules, and then water molecules with counterions. Next, a cycle of minimization was done to relax all atoms without constraints. The maximum number of minimization steps was 20 000 steps and the convergence criterion for the energy gradient was 0.05 kcal/mol/Å. In MD simulations, the SHAKE procedure was employed to constrain all bonds involving hydrogen atom,[40] and the time step was 2.0 fs. The MD simulations consisted of a gradual temperature increase from 10K to 300K over 30 ps and a 2900 ps simulation for equilibration and data collection, during which a snapshot was saved every 500 fs. The resulting trajectories were analyzed using the *ptraj* and *mm_pbsa* modules in AMBER8.0.

Residue−Residue Interaction Analysis. The interactions between each residue in peptide and each residue in the SH3 domain were analyzed using the generalized Born model/solvent area (GB/SA) method in the *mm_pbsa* module in AMBER8.0. The interaction between residue−residue pairs includes three terms: van der Waals contribution (ΔG_{vdw}), electrostatic contribution (ΔG_{ele}) and solvation contribution ($\Delta G_{solvation}$).

$$\Delta G_{residue\_pair} = \Delta G_{vdw} + \Delta G_{ele} + \Delta G_{solvation} = \Delta G_{vdw} + \Delta G_{ele} + \Delta G_{GB} + \Delta G_{SA} \quad (1)$$

where ΔG_{vdw} and ΔG_{ele} are nonbonded van der Waal interaction and electrostatic interaction between two residues, respectively. The solvation contribution $\Delta G_{solvation}$ was calculated using the GB/SA method. The polar contribution (ΔG_{GB}) to solvation was computed using GB with radii parametrized by Onufriev et al.[41] and AMBER charges. The nonpolar contribution (ΔG_{SA}) to solvation was estimated based on the accessible surface area ($G_{SA} = 0.0072 \times$ Area). All energy components in eq 1 were calculated from 50 snapshots taken from 0.8 to 3 ns of the MD simulation.

Data Set. In total, 884 peptides reported by Landraf et al. were selected for 3D-QSAR analyses.[32] In Landraf et al. study, 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. For each domain, peptides matching the defined patterns were synthesized at high density on cellulose membrances by SPOT synthesis technology, and the membranes were probed with the corresponding SH3 domain fused to glutathione S-transferase (GST). The intensity of each SPOT was measured quantitatively in Boehringer light unites (BLUs). BLU, in contrast with the quantities using other high-throughput approaches, has obvious correlation with the dissociation constant.[32] So here, the BLUs values were used as the measure of binding in 3D-QSAR analyses. For the hAmph1 SH3 domain, the authors totally reported 2033 peptides with experimental binding values. Considering the noise in a single experiment, we thus only selected the 884 peptides that had at least 2 experimental results for our analyses. The 884 peptides were randomly divided into a training set of 200 peptides (Table S1) and a test set of 684 peptides (Table S2). The selection of training set and test set molecules was made such that both sets contained peptides covering the entire spectrum of the binding affinities.

Binding Modes of the Peptides and the Structural Alignment. The predicted complexed structure from homology modeling and MD simulations was used as the template to determine the binding conformation and molecular alignment of all peptides. The average structure was fully minimized using the *sander* program in AMBER8.0. And then the peptide PLPRRPPRAA in the template molecule was mutated to a new peptide from the peptide list. The peptide ligand extracted from the minimized complexed structure was considered as the active conformation of the peptide. We thus obtained an ensemble of active conformation and natural orientation for all peptides, which were used as alignment for the CoMFA and CoMSIA calculations. The detailed descriptions of the determination of the binding mode of peptides can be found in Part3 in the Supporting Information.

CoMFA and CoMSIA Studies. (a) CoMFA Setup. CoMFA calculations were performed using the QSAR method in SYBYL.[42] The overlapped peptides in the training set were surrounded by a 3D grid of points in the three dimensions extending at least 4 Å beyond the union volume occupied by the superimposed molecules. The default sp^3 carbon atom with $+1|e|$ charge was selected as probe atom for the calculations of the steric and electrostatic fields around the aligned peptides. In the CoMFA study, besides the usually used steric and electrostatic fields, the H-bonding fields were also included.[43]

To choose the optimum number of components (ONC) and check the statistical significance of the models, leave-one-out (LOO) cross-validation was used by the SAMPLS method,[44] an enhanced version of Partial Least Square (PLS). The final 3D-QSAR models were obtained using ONC on the entire training set. The models were assessed by the explained variance r^2, standard error of estimate (SD), and F ratio. Column filtering (minimum σ) was used at the default value of 2 kcal/mol in the cross-validation part.

(b) CoMSIA Setup. CoMSIA calculations were performed using the QSAR option of SYBYL with five types of physicochemical properties in CoMSIA implemented in SYBYL, steric contributions by the third power of the atomic radii, electrostatics by partial charges, hydrophobicities by atom-based hydrophobic parameters and H-bond properties by suitably placed pseudoatoms, using a common probe with 1Å radius, +1 charge, +1 hydrophobicity and H-bond property of +1. The extent and orientation of the grids surrounding the tested molecules were the same as those in the CoMFA study. The attenuation factor α, which is the coefficient of the squared mutual distance in the Gaussian-type function in the calculation of similarity indices, was set to 0.3. The statistical evaluation for the CoMSIA analyses was performed in the same manner as described for CoMFA.

(c) Calculations of Partial Charges. To consider the influence of different charge models to CoMFA and CoMSIA models, three types of partial atomic charges other than the AMBER charges were considered: Gasteiger−Marsili charges,[45] Gasteiger−Hückel charge[46] and MMFF94 charges.[47] Both of the Gasteiger−Marsili and the Gasteiger−Hückel methods calculate atomic charges based on information about the atoms and their connectivity within the molecule. The MMFF94 atomic charges are computed based on the bond increment parameters in the MMFF94 force field. The calculations of these charges in SYBYL were automated using the SYBYL programming language (SPL) scripts.

(d) q^2-Guided Region Selection (q^2-GRS). The q^2-GRS process has been described in detail elsewhere.[48] In CoMFA, a significant number of lattice points are excluded from the analysis because CoMFA standard scaling applies equal weight to data from each lattice point in a given field. The q^2-GRS technique refines a model by increasing the weight for those lattice points that are most pertinent to the model. Thus, the resultant q^2 exhibits better statistics compared to the conventional q^2. q^2-GRS was executed using the region focusing command in SYBYL.

Results and Discussions

Homology Modeling the Complex Structure of the hAmph1 SH3 Domain and its Peptide Ligand. The modeled structure of the hAmph1 SH3 domain was checked by Profile-3D and PROCHECK in InsightII.[28] The self-compatibility score of Profile-3D was 21.6, which was higher than the low score 13.8. The Ramachandran plot given by PROCHECK is shown in Figure S2 in the Supporting Information. Fifty-four (94.7%) of the 57 nonglycine and nonproline residues were located in the most favored regions, and the remaining 3 residues were located in the additional allowed regions of the Ramachandran plot. Compared with most other SH3 domains, the hAmph1 SH3 domain contains an insertion of 4−5 residues in the n-Src region. Figure S3 in the Supporting Information shows the structural alignment of eight SH3 domains. Apparently the n-Src regions of hAmph1 SH3 domain (pink line) and rAmph2 SH3 domain (blue line) are much longer than those of the other SH3 domains. Of these 10 residues, there are four acidic residues (Asp661, Asp659, Glu657, and Asp655) among which Asp661 and Asp659 are close to the peptide binding pocket and may have significant contributions to binding. Also, as in other structures, the side-chains of the residues His631, Phe633, Trp664, Pro687, and Phe714 in the hAmph1 SH3 domain form hydrophobic surface patches against the conserved proline residues of the binding peptide. Along with the internal packing positions of the peptides, the binding specificity is usually determined by the interactions between residues in the n-Src and RT loops of the SH3 domain.

Several experiments show that the amphiphysin SH3 domains prefer two Arg's in the internal packing positions of a polyproline helix,[32−33] which may be due to the favorable electrostatic interactions between the domain and the Arg's in the peptide. We thus calculated the electrostatic potentials of the hAmph1 SH3 domain (Figure 3). Electrostatic potentials were calculated using the finite different Poisson−Boltzmann method (Delphi program[52] in INSIGHTII with AMBER charges). Compared with the other four SH3 domains, Abl tyrosine kinase SH3 domain,[5] c-Crk N-terminal SH3 domain,[34] Grb2 N-terminal SH3 domain,[51] and rat amphiphysin-2 SH3 domain,[26] whose electrostatic potentials were calculated in our previous work,[50]

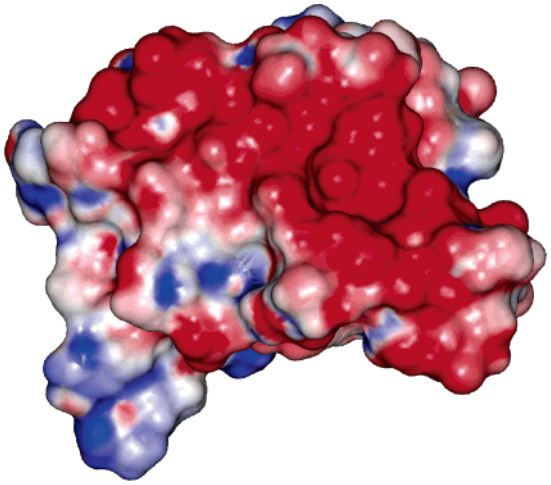

Figure 3. Electrostatic potentials of the peptide binding interfaces for the hAmp1 SH3 domain. Negative areas are shown in red, and positive areas are shown in blue. These figures are produced using INSIGHTII.

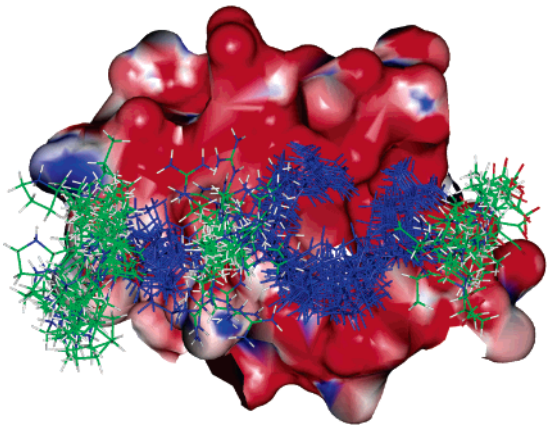

Figure 4. Superposition of 21 snapshots extracted from the MD trajectory from 1.0 to 3.0 ns. The solvent-accessible molecular surface of hAmp1 SH3 is colored according to the electrostatic potential calculated using the Delphi program in INSIGHTII. The residues Pro at P2, Arg at P0, Pro at P-1, Pro at P-2, and Arg at P-3 are colored by blue, and other residue are colored by atom types.

the hAmph1 SH3 domain, similar to the rAmph2 SH3 domain, has large areas of negative electrostatic potentials, which explains why the hAmph1 and rAmph2 SH3 domains prefer PXRPXR sequences containing two positive charges.

Interactions between the hAmph1 SH3 Domain and PLPRRPPRAA from MD Simulations. Time course of the total energy relative to the initial snapshot is shown in Figure S4(a) in the Supporting Information. The main chain root-mean-square deviation (RMSD) compared with the starting structure is shown in Figure S4(b), which indicates that the system reaches equilibrium after about 800 ps. Analyses of the trajectories taken from the MD simulations showed that there was a significant motion of the peptide complexed with the SH3 domain, especially the N- and C-terminal residues. The 21 snapshots extracted from the MD trajectory from 1.0 to 3.0 ns are superposed in Figure 4. We found that Pro at P_2 and residues from position P_0 (Arg) to position P_{-3} (Arg) were restrained by the binding to the SH3 domain. In contrast, the N- and C-terminal residues as well as Arg at P_1 exhibited great

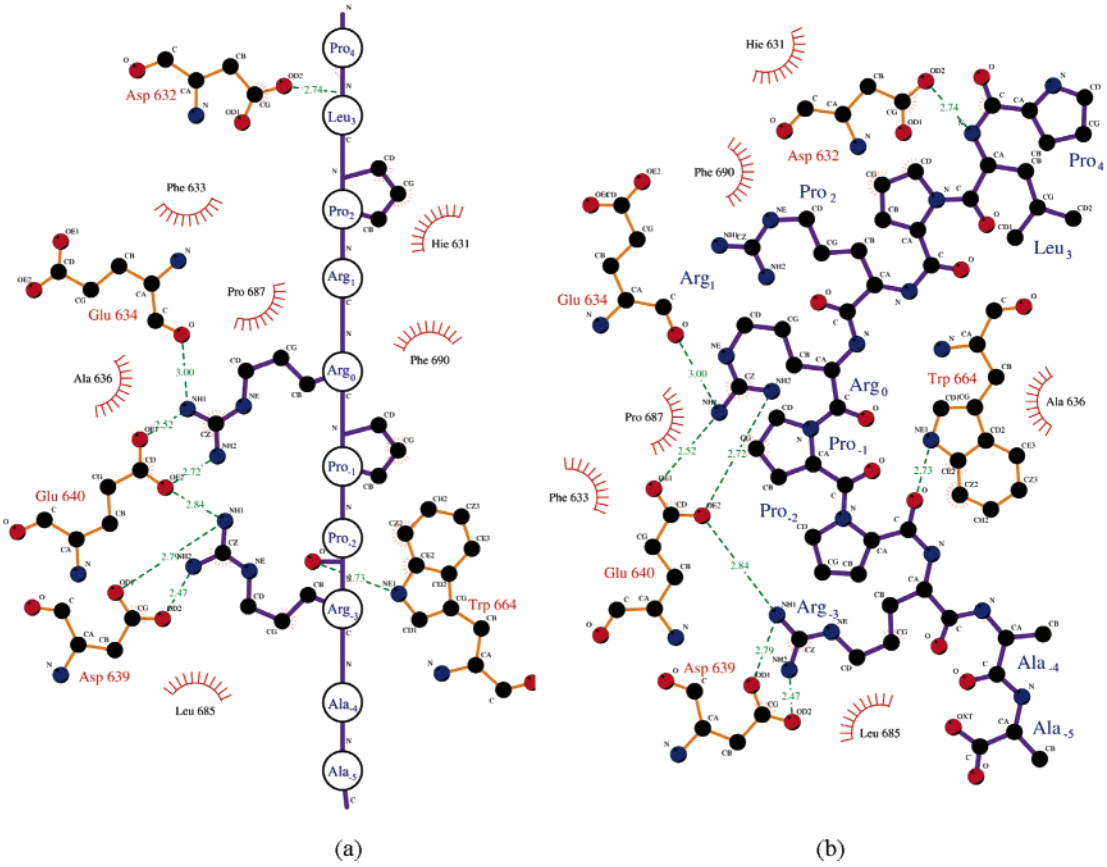

Figure 5. Schematic representation of interactions between peptide PLPRRPPRAA and hAmph1 SH3 produced using the Ligplot program developed by Wallace et al.[52] The average structure after minimization was used in the analysis. (a) only the side chains which have strong interaction with SH3 are shown, (b) all ligand atoms are shown.

conformational flexibility as they were relatively unrestrained by the binding.

To evaluate the contribution of each peptide residue to binding, the total binding free energy of was decomposed to residue−residue pairs (Figure 5, Figure S5, and Table S3). Apparently the middle six residues of the ligand make significant contacts with the hAmph1 SH3 domain, while the C-terminal Ala and the N-terminal Pro have little contribution to binding. Moreover, the interactions are mainly between the side chains of the peptide and the SH3 domain as there are only two backbone/backbone H-bonds: one between the backbone N atom of Leu at P_3 and Asp632 and the other between the backbone carbonyl oxygen atom of Pro at P_{-2} and Trp 664 (Figure 5). These two hydrogen bonds are not very stable during MD simulations.

Both prolines at P_2 and P_{-1} positions have strong interactions with the protein. Three protein residues His631, Asn689 and Phe690 form effective interactions with Pro at P_2 and the van der Waals interaction is the dominant component (Table S3 in the Supporting Information). The interaction between Pro at P_{-1} and the protein is highly favorable (−7.09 kcal/mol). Trp664 and Pro687 are very important to the strong binding of this residue. Pro687 can form strong van der Waals interaction with Pro at P_{-1}, whereas both van der Waals and electrostatic interactions (including the electrostatic contribution from solvation) are important to the binding between Trp664 and Pro at P_{-1}. In fact, the interactions between these two prolines in peptide and the four protein residues, Phe690, His631, Pro687, and Trp664, have been shown to be conserved in SH3 domains.

The interaction between Arg at P_1 and SH3 is relatively weak (Table S3). This residue can only form effective electrostatic interaction with Asn689. In the modeled complex structure, the side chain of the arginine points away from the protein and therefore has less contribution to binding. Arg at P_0 can form strong interaction with the SH3 domain (−22.88 kcal/mol in Table S3). This arginine residue has effective interactions with five SH3 residues: Phe633, Glu634, Glu640, Trp664, and Phe690. As shown in Figure 5, arginine at P_0 can form two stable hydrogen bonds with two carboxyl oxygen atoms in Glu640, which was validated by the strong interaction between Arg at P_0 and Glu640 (−14.97 kcal/mol). Glu634 can form strong electrostatic interaction with arginine at P_0 (−19.91 kcal/mol), but the strong electrostatic interaction was largely compensated by the unfavorable electrostatic component of desolvation upon binding. There exists a relatively weak hydrogen bond between the carbonyl oxygen atom in Glu634 and arginine at P_0 (Figure 5). The hydrophobic side chains of Phe633, Trp664 and Phe690 can make favorable van der Waals interactions with the side chain of arginine at P_0.

Even the backbone carbonyl oxygen atom of the proline at P_{-2} of the peptide can form a hydrogen bond with Trp 664, this residue is exposed to solvent and is not as important as the two prolines at P_{-1} and P_2 to binding.

The arginine at P_{-3} of the peptide may contribute most to the peptide/SH3 interactions. It can form effective interactions with five SH3 domain residues, Asp639, Glu640, Asp659, Gln660, and Trp664. Asp639 and Glu640 are more important than the other three residues (−11.66 kcal/mol for Asp639 and −8.10 kcal/mol for Glu640) because of the stable hydrogen bonds

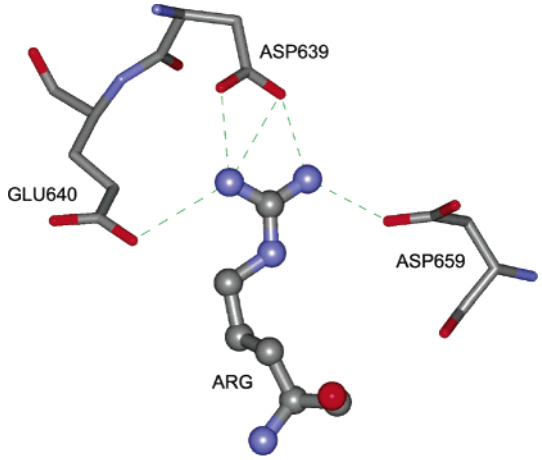

Figure 6. H-bond network formed by Arg at P-3 in peptide and three residues in Ahmp1 SH3. The protein residues are shown in stick, and the ligand residue is shown in ball-and-stick. The hydrogen bonds between protein and peptide are shown as green dotted lines.

formed between these two resides and the peptide. In the average complex structure obtained from the MD simulation, three hydrogen bonds can be found between Arg at P_{-3} and Asp639/Glu640 (Figure 5). Besides Asp639 and Glu640, the interaction between Asp659 and Arg at P_{-3} is also very significant (−4.56 kcal/mol). From the analysis of the MD trajectories, it can be found that the OD1 atom in the carboxyl group of the Asp659 is very close to the NH2 atom of the Arg at P_{-3}. The residue Asp639, Glu640, and Asp 659 form a nearly isosceles triangle, with the Arg nitrogen at the plane of the triangle formed by the Asp carboxyl carbons (Figure 6). The binding site is thus remarkably well suited for interactions with Arg and it is not surprising that five hydrogen bonds formed between Arg at P_{-3} and four carboxylate oxygens in Asp639, Glu640 and Asp659.

CoMFA and CoMSIA Models. The CoMFA and CoMSIA models (2.0 Å grid) with different fields based on the AMBER charge model are shown in Table 1. It shows that using steric and electrostatic fields only, the quality of the CoMSIA model ($q^2 = 0.580$) was quite similar to the CoMFA model ($q^2 = 0.587$). When the hydrophobic interaction is included, the predictive power of the CoMSIA model ($q^2 = 0.593$) is slightly increased, which indicates a contribution of the hydrophobic field to peptide binding. Also, after adding the H-bond donor field, the cross-validation coefficient of the 3D-QSAR was increased significantly to 0.606, which indicates that H-bond donor properties are important for peptide binding. However, inclusion of the H-bond acceptor field does not improve the CoMSIA model. Moreover, the inclusion of H-bond field does not improve the predictive power of the CoMFA model. The best 3D-QSAR model was built using all four fields in CoMSIA ($q^2 = 0.606$).

For QSARs developed with CoMFA or CoMSIA, a shift in the q^2 values may be often observed using different charge models.[54−55] The CoMFA and CoMSIA models using three different sets of partial charges are shown in Table 2. The CoMSIA models show better regression results. For the three CoMSIA models, all of them exhibit good statistical quality between the predicted and experimentally determined values of log(BLU). Considering that the internal predictive ability of CoMSIA3 model in Table 2 is slightly higher than those of the other models, Gasteiger−Marsili partial charges were used in the subsequent studies.

The region focusing technique in the 'Advanced CoMFA' module in SYBYL was used to refine the model by increasing the weights for those lattice points most pertinent to the model and all 3D-QSAR models show obvious improvement. The improvement of the CoMFA model with the steric, electrostatic

Table 1. CoMFA and CoMSIA Models Using the Different Combinations of Fields based on the AMBER Charge Model

	CoMFA1	CoMFA2	CoMSIA1	CoMSIA2	CoMSIA3	CoMSIA4
q^2	0.587	0.593	0.580	0.593	0.606	0.604
r^2	0.827	0.793	0.794	0.768	0.766	0.775
standard error	0.375	0.406	0.407	0.431	0.432	0.424
F	153.577	250.723	187.722	216.245	214.255	225.415
n	6.000	3.000	4.000	3.000	3.000	3.000
grid spacing	2.000		2.000	2.000	2.000	2.000
fraction						
steric	0.473	0.165	0.242	0.155	0.096	0.086
electrostatic	0.527	0.193	0.758	0.496	0.326	0.281
hydrophobic				0.349	0.216	0.191
HB donor		0.274			0.361	0.324
HB acceptor		0.368				0.117

Table 2. Influence of Different Charge Models to CoMFA and CoMSIA Model (grid spacing: 2.0)

	CoMFA1	CoMFA2	CoMFA3	CoMSIA1	CoMSIA2	CoMSIA3
charge model	Gasteiger− Marsili	Gast−Huck	MMFF94	Gasteiger− Marsili	Gast−Huck	MMFF94
q^2	0.600	0.599	0.598	0.618	0.612	0.614
r^2	0.804	0.805	0.805	0.766	0.762	0.767
standard error	0.396	0.395	0.395	0.433	0.436	0.432
F	268.765	270.108	269.460	213.692	209.090	215.044
n	3.000	3.000	3.000	3.000	3.000	3.000
grid spacing	2.000	2.000	2.000	2.000	2.000	2.000
fraction						
steric	0.177	0.177	0.176	0.102	0.107	0.105
electrostatic	0.132	0.132	0.136	0.275	0.257	0.263
hydrophobic				0.234	0.239	0.237
HB donor	0.293	0.294	0.292	0.389	0.397	0.395
HB acceptor	0.397	0.398	0.396			

Table 3. CoMFA and CoMSIA Models after Region Focusing

	CoMFA1	CoMFA2	CoMSIA1	CoMSIA2	CoMSIA3	CoMSIA4
q^2	0.620	0.633	0.606	0.625	0.636	0.631
r^2	0.750	0.746	0.796	0.756	0.767	0.729
standard error	0.449	0.450	0.407	0.443	0.425	0.466
F	145.900	191.827	125.762	150.865	160.302	175.777
n	4.000	3.000	6.000	4.000	4.000	3.000
grid spacing						
fraction						
steric	0.566	0.227	0.412	0.239	0.142	0.128
electrostatic	0.434	0.171	0.588	0.370	0.283	0.193
hydrophobic				0.392	0.246	0.208
HB donor		0.257			0.329	0.350
HB acceptor		0.345				0.122

and H-bond fields appeared to be higher than that with the steric and electrostatic fields. After using region focusing, the CoMSIA model was the best according to several statistical parameters. Leave-one-out cross-validation of the PLS analysis of the model resulted in a model with q^2 of about 0.64 using four principle components. The non-cross-validated PLS analysis yields a model with a higher r^2 of 0.767 with a low standard error of estimate (SD) of 0.43 (Table 3). Figure 7 shows the linear correlation between the experimental affinities and the predicted values of the peptides in the training set.

The predictability of the models derived by CoMFA and CoMSIA is the most important criteria to assess the two methods. The best CoMFA and CoMSIA models were applied to predict the binding of the 684 peptides in the test set (Table S2 in the Supporting Information). By comparing the results of the best CoMFA and CoMSIA models, one can see that the best CoMSIA models also possesses good predictive ability (r_{pred} = 0.79) with the average absolute error of 0.41 log units across a range of 3.87 log units, which is better than the prediction of the best CoMFA to the test set (r_{pred} = 0.76 and the average absolute error of 0.42 log units). The binding values (logBLU), the predicted values, and the prediction error using the best CoMSIA model for the test set are given in Table S2 the Supporting Information. For the 684 peptides in the test set, 454 have absolute prediction errors smaller than 0.5 (66.4%), 193 absolute prediction errors between 0.5 and 1.0 (28.2%), 28 absolute prediction errors between 1.0 and 1.5 (4.1%) and 9 peptides with absolute prediction errors larger than 1.5 (1.3%). Totally, about 95% peptide can be well predicted with absolute prediction errors smaller than 1.0. The correlation between the predicted and measured log(BLU) values is presented in Figure 8.

CoMSIA Contour Maps. The CoMSIA steric, electrostatic, hydrophobic, and H-bond donor fields were plotted as three-dimension colored contour maps in Figure 9. The steric contour map for the best CoMSIA model depicts regions around the molecules where enhanced SH3 binding affinity is associated with increasing (green) and decreasing (yellow) steric bulk (see Figure 9a). The steric contours show that there are two sterically favorable regions: one is around the side chains of Lys at P_1, and the other is located near the terminal of the side chain of Arg at P_{-3}. In fact, the presence of these two regions is supported by the MD simulations. The side chain of Lys at P_1 is external and thus large group is allowed here and does not impair the SH3/peptide interaction. The hAmph1 SH3 domain has relatively large space at site 3, which can accommodate large residues, such as Arg, Lys, and Phe. Moreover, in Figure 9a, there are two unfavorable regions: the large one is close to Pro at P_2 and the small one is close to the region between Pro at P_{-2} and Lys at P_{-4}. These large yellow regions indicate that additional steric interaction in these regions would lead to a decreased binding. For example, Pro at P_2 forms close contact

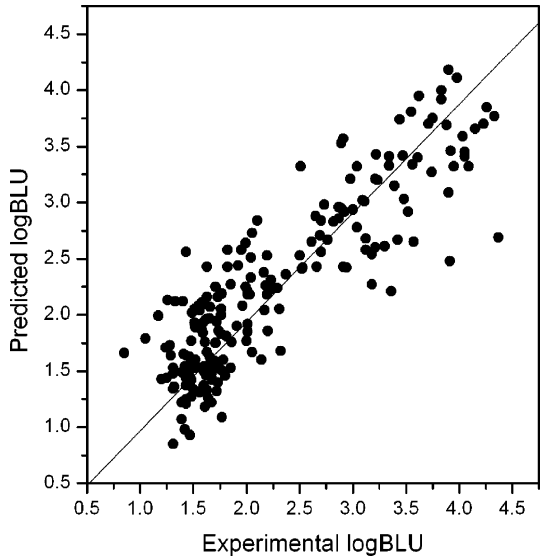

Figure 7. Comparison of experimental logBLU with predicted logBLU using the best CoMSIA model for the peptides in the training set.

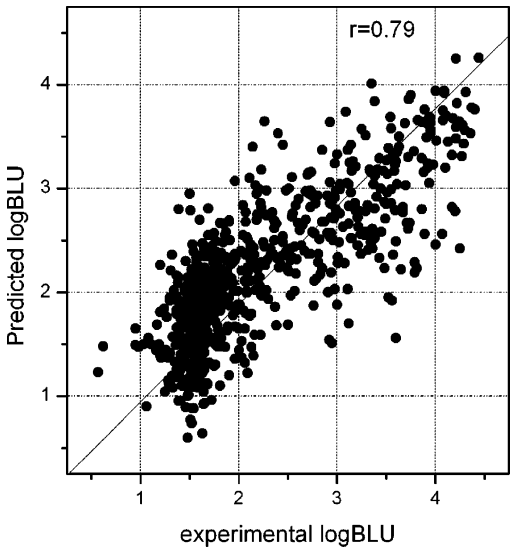

Figure 8. Comparison of experimental logBLU with predicted logBLU using the best CoMSIA model for the peptides in the test set.

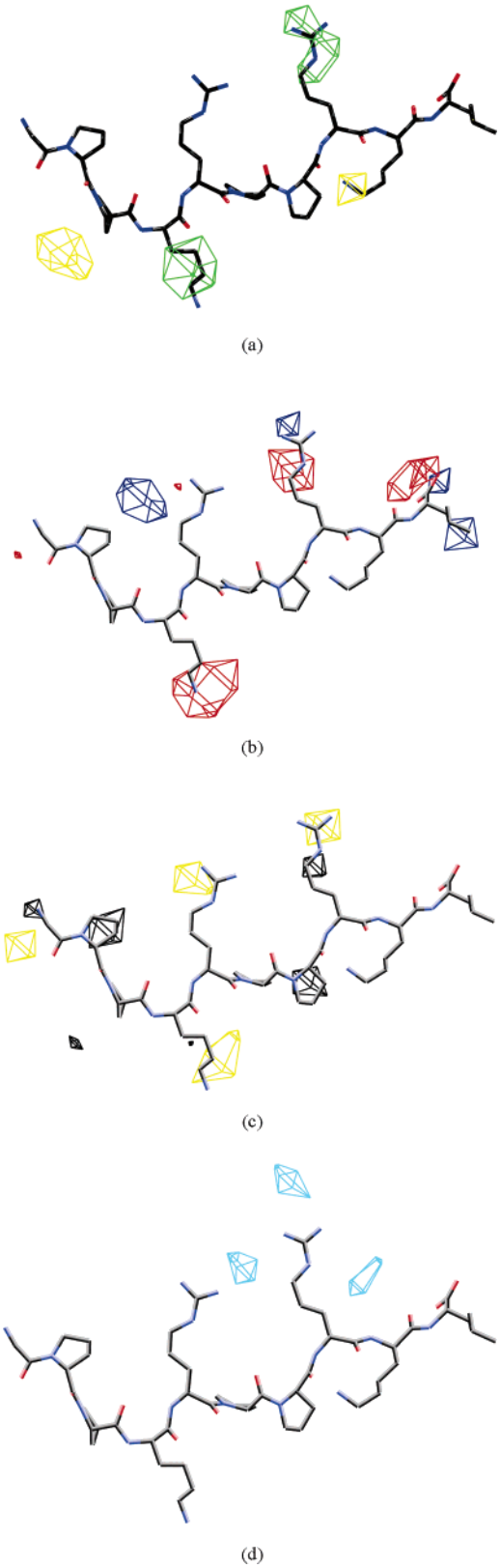

Figure 9. CoMSIA contour maps for (a) steric field, (b) electrostatic field, (c) hydrophobic field and (d) H-bond donor field.

with His631 and Phe690, and large group at this position should destroy the surface complementarity between peptide and SH3.

In the electrostatic contour maps (Figure 9b), peptides orienting side chains with positive electrostatic potentials into areasin red will enhance peptide binding, as well as the side chains with negative electrostatic potentials into areas in blue. In Figure 9b, there are three blue areas: one is at the end of Arg at P_{-3}, one is close to area between reside at P_3 and Arg at P_0, and the other one is close to the C terminal of the peptide. The blue areas are consistent with the distribution of electrostatic potentials of the SH3 binding potential. The residues at P_0, P_{-3}, or P_{-5} prefer to form strong electrostatic interactions with SH3. It is interesting to find that there are two red areas near two blue areas. In fact, in many cases, the blue and red regions are adjacent to each other.[56] It is not very surprising because a positive and negative charge group often are adjacent to each other. There is large red area around the side chain of Arg at P_1, which is not straightforward to be interpreted.

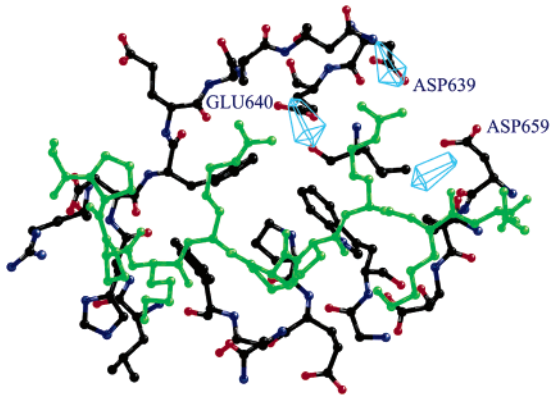

Figure 10. Alignment between the CoMSIA H-bond donor contour maps and the average structure of SH3 after minimizations. Peptide ligand was shown as green stick.

The maps of hydrophobic properties are shown in Figure 9c. Yellow and black meshes indicate regions where hydrophobic or hydrophilic groups are favorable to peptide binding. In Figure 9c, there are three obvious hydrophobic favorable areas. These three yellow areas are located near the side chains of Arg's at P_1, P_0, and P_{-3}, which appears to contradict the MD results. From MD, we found that two residues at P_0 and P_{-3} can form strong electrostatic interaction with the SH3 domain. The improper output of the CoMSIA model may be caused by the CoMSIA algorithm. In our data set, the residues at P_1, P_0, and P_{-3} are usually Arg or Lys. The terminal carbon atom of the side chain of Arg or Lys has significant positive charges, but CoMSIA may identify these atoms as hydrophobic centers. Although the hydrophobic distribution in Figure 9c is unreasonable, we believe that the hydrophobic property of the peptide should still be effectively considered by CoMSIA.

H-bond donor field is shown in Figure 9d, where cyan mesh indicates regions where hydrogen bond acceptor groups on SH3 will enhance the binding. In Figure 9d, there are three cyan areas near Arg at P_0 and Arg at P_{-3}. To directly compare the CoMSIA contour map with the structure of the hAmph1 SH3 domain, residues in the SH3 domain within 5 Å of peptide ligand were merged to the CoMSIA contour map (Figure 10). In Figure 10, it is very clear that two cyan areas are superimposed with the side chain of Glu640 and the side chain of Asp639, and the other small one is close to the side chain of Asp659, which is well consistent with the results from MD simulations. The MD simulations have shown these two Arg's can form strong hydrogen bonds with these residues in SH3.

When using the CoMSIA model to predict affinities on the test set, we found that 37 of the peptides had poor predictions.

Table 4. Peptides with Absolute Prediction Errors Larger than 1.5 log Units

no.	peptide sequence	protein name	logBLU (exp)	logBLU (pred)[a]	residue[a]	logBLU (pred)[b]	residue[b]
A37	RPPSRPPSRP	Q99700	4.37	2.54	1.83	2.69	1.68
B30	RPPRRPPWGR	Q99785	4.00	1.86	2.14	2.46	1.54
B32	GFPRRPPPRG	Q9UER3	3.81	2.28	1.53	2.23	1.58
B42	LLPRRPPPRV	O15056	3.65	2.25	1.40	2.12	1.53
B52	EGPRRPPPRP	Q9BVH8	3.52	2.08	1.44	1.95	1.57
B53	AFPRRPPPRG	Q9Y5N2	3.79	2.27	1.53	2.19	1.60
B54	VVPRRPPAHI	Q9NTY8	3.79	2.39	1.41	2.29	1.50
B211	GPPPRPPPRQ	Q9HCU4	4.25	2.16	2.09	2.42	1.83
B448	VAPRRPRDAV	O43541	3.56	1.81	1.75	1.92	1.64
B460	RPFRRPNANF	Q9NVW8	3.60	1.40	2.20	1.56	2.05

[a] Prediction based on CoMFA2 model in Table 5. [b] Prediction based on CoMSIA3 model in Table 5.

Among these 37 peptides, 9 peptides have absolute prediction errors larger than 1.5 log units. There are many reasons that could have caused these poor predictions. First, since the energy estimate can be noisy, there may have been an inaccurate binding energy estimate by experiment.[32] Second, the binding characteristics of a peptide in the binding site cannot be well represented by a single conformation of a peptide. The MD simulation results have shown that there was significant conformational flexibility of the peptide in the binding pocket. Therefore, a single conformation of the peptide cannot describe the dynamic property of the peptide upon binding. Third, the side chain installation and molecular mechanics minimization may not produce the correct binding conformation of a peptide in the binding site. The basic assumption in CoMFA or CoMSIA is that the modeled conformation for molecular alignment is the biologically active one. For flexible ligands, especially peptides, this assumption may not be the case.

Table 4 shows the peptides with prediction errors larger than 1.5. One can see that the binding affinities of all these peptides are underestimated by the CoMSIA model. It implies that the CoMSIA model cannot fully represent some of the chemical features that contribute to peptide binding. Among these 10 peptides, six share following the common feature: an arginine at P_{-4} and a small residue (proline or serine) at P_{-3}. These common features can be defined as the motif PXRPXsR, where s represents small residues. The peptide motif is not fully consistent with the results from MD simulations. Our previous MD simulations show that Arg at P_{-3} is the most important contributor to the binding. However, all of these six peptides do not have Arg at P_{-3}. It seems that these six peptides may have different binding conformations in the binding pocket compared with the peptide PLPRRPPRAA. To investigate the interaction between the SH3 domain and the peptides with the motif PXRPXsR, we performed a 3 ns MD simulation for the SH3/GFPRRPPPRG complex. The initial structure for MD simulation was also generated by side chain installation and molecular mechanics minimization. We used the same simulation condition in our previous MD simulation for the SH3/PLPRRPPRAA complex.

The MD simulations of an example peptide (GFPRRPPPRG) from these six poorly predicted peptides suggested that they adopt a different conformation from the rest of the peptides in the test set. According to Figure S6, the total energy becomes stable after about 400 ps but RMSD does not reach plateau until about 700 ps. The contribution of each residue in the peptide GFPRRPPPRG to binding is summarized in Table S4 in the Supporting Information. Upon examining the differences of the binding conformations of the peptides PLPRRPPRAA and GFPRRPPPRG (Tables S3 and S4 in the Supporting Information), it was found that the binding modes for these six residues from P_4 to P_{-2} in GFPRRPPPRG are quite similar to those in PLPRRPPRAA. However, for the three residues from P_{-3} to P_{-5}, these two peptides show some differences. The proline at P_{-3} in GFPRRPPPRG cannot form any effective interaction with protein, whereas Arg at P_{-4} can form strong interactions with three residues in SH3, including Ala658, Asp659 and Asp661. In the average structure of SH3/GFPRRPPPRG complex, we find that Arg at P_{-4} can form three hydrogen bonds with Ala658 and Asp659 (Figure 11). Figure 12 shows the alignment of average structures of SH3/PLPRRPPPAA and SH3/GFPRRPPPRG. As shown in Figure 12, the backbone structures of these four residues of PLPRRPPPAA (red peptide shown in Figure 12) and GFPRRPPPRG (green peptide shown in Figure 12) at the C terminal have some differences. Compared with PLPRRPPPAA, three residues from P_{-1} to P_{-3} in PLPRRPPPAA move away from the binding surface of the hAmph1 SH3 domain (the direction is labeled by an arrow). Moreover, the conformation of several

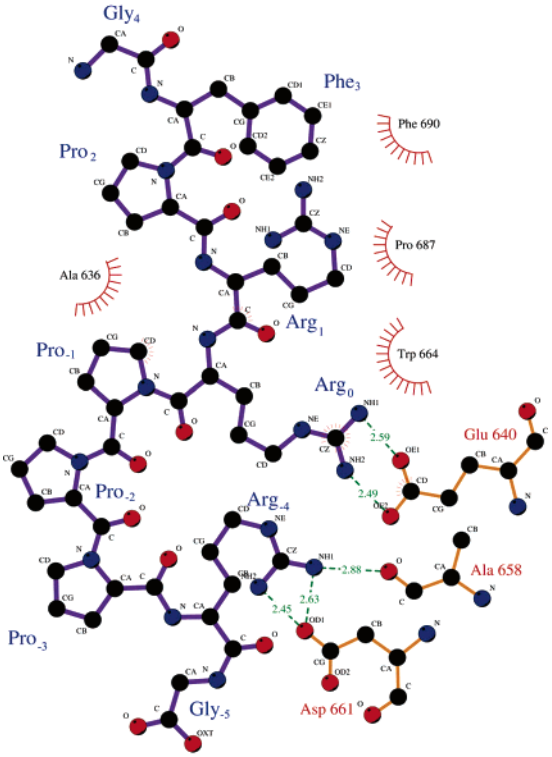

Figure 11. Schematic representation of interactions between peptide GFPRRPPPRG and hAmph1 SH3 produced using the Ligplot program developed by Wallace et al.[53] The average structure after minimization was used in analysis.

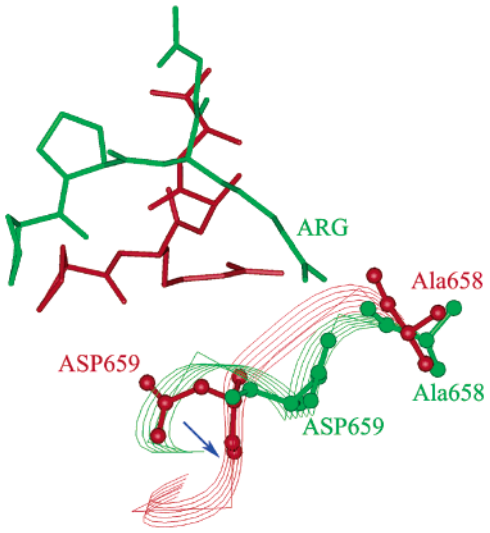

Figure 12. Alignment of average structures of SH3/PLPRRPPRAA and SH3/GFPRRPPPRG. SH3/PLPRRPPRAA is shown in red and SH3/GFPRRPPPRG in green. The residues in SH3 are represented in ball-and-stick, and those in peptides in stick.

residues in the n-Src loop of the SH3 domain shows significant difference when interacting with PLPRRPPPAA and GFPRRPPPRG. To form an interaction with Arg at P_{-4} in GFPRRPPPRG, the side chain of Asp659 rotated about 180° and he overall structure near the Asp of SH3 is quite different between interacting with PLPRRPPPAA and with GFPRRPPPRG. From MD simulations, it is clear why the GFPRRPPPRG have relatively high binding interaction with SH3 although there is no interaction with Arg at P_{-3}: the interaction between SH3 and Arg at P_{-4} can compensate the loss of binding energy by the substitution of Arg or Lys at P_{-3}.

The underlying reason for the poor prediction of peptides with PXRPXsR motif is therefore that the PXRPXsR motif has different binding conformation compared with the PX(R/K)-PX(R/K) motif. We believe that for the hAmph1 SH3 domain, there may be a number of binding modes. The hAmph1 SH3 domain has long and flexible n-Src domain and it may adjust the conformation easily to adopt peptides with other motifs, especially these peptides with positively charged C-terminal residues because the hAmph1 SH3 domain has strong negative potentials near the RT loop and n-Src loop.

Conclusion

Homology modeling and MD simulations were applied to determine the structure of a strong binding peptide complexed with the human amphiphysin-1 SH3 domain. The interaction between the peptide PLPRRPPRAA and the SH3 domain were systematically analyzed and the key binding forces were identified. On the basis of the molecular alignment and the modeled structure from homology modeling and MD simulations, the 3D-QSAR techniques CoMFA and CoMSIA were applied to study the relationship between structure and binding to 884 10-residue peptides with experimental binding data for the human amphiphysin-1 SH3 domain. The preferred and not-preferred areas of steric bulk, electron density, local hydrophobicity and hydrogen-bond forming properties for the binding peptides were defined. Most CoMSIA contour features can be well explained by the interacting information discovered by the MD simulations.

To investigate the reason the CoMSIA model cannot predict well for some peptides with the PXRPXsR motif, an MD simulation was performed on the GFPRRPPPRG/SH3 complex. The simulations showed that the peptide GFPRRPPPRG could adopt a binding conformation different from the PLPRRPPRAA. Since the calculated results from MD simulations explain why the CoMSIA model gives poor predictions for several peptides, the combination of ligand-based and receptor-based modeling is proposed to be a powerful approach to predict the binding affinities of peptides to SH3 domain and to analyze the binding features of peptides in the SH3 binding pocket. In the future, the ability to predict peptide binding using a CoMSIA model may enable us to analyze human genomes, identifying the most potential binding partner of the human amphiphysin-1 SH3 domain. Moreover, although we presently focus on the binding between peptides and the hAmph1 SH3 domain, the method we have described may be useful for other peptide/protein systems, such as the other SH3 domains as well as other modular signaling domains, e.g., SH2, WW, and PDZ domains.

Acknowledgment. The MD simulations were performed on the Linux cluster in the Center for Theoretical Biological Physics (CTBP) at UCSD. This work has been supported by the NSF PFC-sponsored Center for Theoretical Biological Physics (Grant Nos. PHY-0216576 and PHY-0225630). T.H. is supported by a CTBP postdoctoral scholarship. B.L. is supported by an NIH grant to the McCammon's group (Grant No. GM31749). We thank Prof. J. Andrew McCammon for providing access to computer software such as InsightII and SYBYL.

Supporting Information Available: The peptide sequence, the experimental and predicted binding affinities for peptides in the training set and test set are shown in Table S1 and S2; the peptide-residues interaction based on residue−residue decomposition analysis for the PLPRRPPRAA/SH3 and GFPRRPPPRG/SH3 complexes are shown in Table S3 and S4. The molecular alignment for the 200 peptides in the training set is shown in Figure S1; Ramachandran plot for the homology model after MD simulations is shown in Figure S2; the structural alignments of eight SH3 domains is shown in Figure S3; the MD time course of (a) the total energy and (b) the RMSD of the SH3 backbone in the PLPRRPPRAA/SH3 complex is shown in Figure S4; the contribution of each residue in peptide PLPRRPPRAA to the SH3/peptide interactions is shown in Figure S5; the MD time course of (a) the total energy and (b) the RMSD of the SH3 backbone in the GFPRRPPPRG/SH3 complex is shown in Figure S6. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dalgarno, D. C.; Botfield, M. C.; Rickles, R. J. SH3 domains and drug design: ligands, structure, and biological function. *Biopolymers* **1997**, *43*, 383−400.
(2) Stahl, M. L.; Ferenz, C. R.; Kelleher, K. L.; Kriz, R. W.; Knopf, J. L. Sequence similarity of phospholipase C with the noncatalytic region of Src. *Nature* **1988**, *332*, 269−272.
(3) Musacchio, A.; Wilmanns, M.; Saraste, M. Structure and function of the SH3 domain. *Prog. Biophys. Mol. Biol.* **1994**, *61*, 283−297.
(4) Mayer, B. J. SH3 domains: complexity in moderation. *J. Cell Sci.* **2001**, *114*, 1253−1263.
(5) Musacchio, A.; Saraste, M.; Wilmanns, M. High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nature Struct. Biol.* **1994**, *1*, 546−551.
(6) Ren, R. B.; Mayer, B. J.; Cicchetti, P.; Baltimore, D. Identification of a 10-amino acid proline-rich SH3 binding-site. *Science* **1993**, *259*, 1157−1161.

(7) Viguera, A. R.; Arrondo, J. L. R.; Musacchio, A.; Saraste, M.; Serrano, L. Characterization of the interaction of natural proline-rich peptides with 5 different SH3 domains. *Biochemistry* **1994**, *33*, 10925−10933.
(8) Lim, W. A.; Richards, F. M.; Fox, R. O. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **1997**, *372*, 375−379.
(9) Rickles, R. J.; Botfield, M. C.; Weng, Z.; Taylor, J. A.; Green, O. M.; Brugge, J. S.; Zoller, M. J. Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligangds by screening a random phage display library. *EMBO J.* **1994**, *13*, 5598−5604.
(10) Rickles, R. J.; Botfield, M. C.; Zhou, X. M.; Henry, P. A.; Brugge, J. S.; Zoller, M. J. Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10909−10913.
(11) Sparks, A. B.; Rider, J. E.; Hoffman, N. G.; Fowlkes, D. M.; Quillam, L. A.; Kay, B. K. Distinct ligand preferences of Src homology 3 domain from Src, Yes, Abl, Cortactin, p53bp2, Plc-g, Crk, and Grb2. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1540−1544.
(12) Feng, S. B.; Chen, J. K.; Yu, H. T.; Simon, J. A.; Schreiber, S. L. 2 binding orientations for peptides to the Src SH3 domain − development of a general-model for SH3-ligand interactions. *Science* **1994**, *266*, 1241−1247.
(13) Froloff, N.; Windemuth, A.; Honig, B. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Protein Sci.* **1997**, *6*, 1293−1301.
(14) Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J. M.; Luo, R.; Bayly, C. T.; Kollman, P. A. 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.
(15) Donnini, A.; Donnini, S. Calculation of affinities of peptides for proteins. *J. Comput. Chem.* **2004**, *25*, 393−411.
(16) Doytchinova, I. A.; Flower, D. R. 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.
(17) David, C.; McPherson, P. S.; Mundigl, O.; De Camilli, P. A role of amphiphysin in synaptic vesicle endocytosis suggested by its binding to dynamin in nerve terminals. *Proc. Natl Acad. Sci. U.S.A.* **1996**, *93*, 331−335.
(18) Slepnev, V. I.; Ochoa, G. C.; Butler, M. H.; Grabs, D.; Camilli, P. D. Role of phosphorylation in regulation of the assembly of endocytic coat complexes. *Science* **1998**, *281*, 821−824.
(19) Wigge, P.; Kohler, K.; Vallis, Y.; Doyle, C. A.; Owen, D.; Hunt, S. P.; McMahon, H. T. Amphiphysin heterodimers: potential role in clathrin-mediated endocytosis. *Mol. Biol. Cell* **1997**, *8*, 2003−2015.
(20) Lichte, B.; Veh, R. W.; Meyer, H. E.; Kilimann, M. W. Amphiphysin, a novel protein associated with synaptic vesicles. *EMBO J.* **1992**, *11*, 2521−2530.
(21) David, C.; Solimena, M.; De Camilli, P. Autoimmunity in stiff-man syndrome with breast-cancer is targeted to the c-terminal region of human amphiphysin, a protein similar to the yeast proteins, rvs167 and rvs161. *FEBS Lett.* **1994**, *351*, 73−79.
(22) McPherson, P. S. Regulatory role of SH3 domain-mediated protein−protein interactions in synaptic vesicle endocytosis. *Cell Signal.* **1999**, *11*, 229−238.
(23) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular-field analysis (CoMFA). 1. effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1989**, *110*, 5959−5967.
(24) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indexes in a comparative-analysis (CoMSIA) of drug molecules to correlate and predict their biological-activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.
(25) Hou, T. J.; Xu, X. J. Recent progress in comparative molecular field analysis. *Prog. Chem.* **2001**, *13*, 436−440.
(26) Owen, D. J.; Wigge, P.; Vallis, Y.; Moore, J. D.; Evans, P. R.; McMahon, H. T. Crystal structure of the amphiphysin-2 SH3 domain and its role in the prevention of dynamin ring formation. *EMBO J.* **1998**, *17*, 5273−5285.
(27) Pearson, W. R.; Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2444−2448.
(28) Insight II 2000 User Guide, MSI Inc., San Diego, USA, **2000**.
(29) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. Derivation of Class II force fields. 1. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comput. Chem.* **1994**, *15*, 162−182.
(30) Bowie, J. U.; Lüthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **1991**, *253*, 164−170.
(31) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(32) Landgraf, C.; Panni, S.; Montecchi-Palazzi, L.; Castagnoli, L.; Schneider-Mergener, J.; Volkmer-Engert, R.; Cesareni, G. Protein interaction networks by proteome peptide scanning. *PLOS Biol.* **2004**, *2*, 94−103.
(33) Cestra, G.; Castagnoli, L.; Dente, L.; Minenkova, O.; Petrelli, A.; Migone, N.; Hoffmuller, U.; Schneider-Mergener, J.; Cesareni, G. The SH3 domains of endophilin and amphiphysin bind to the proline-rich region of synaptojanin 1 at distinct sites that display an unconventional binding specificity. *J. Biol. Chem.* **1999**, *274*, 32001−32007.
(34) Wu, X. D.; Knudsen, B.; Feller, S. M.; Zheng, J.; Sali, A.; Cowburn, D.; Hanafusa, H.; Kuriyan, J. Structural basis for the specific interaction of lysine-containing proline-rich peptides with the n-terminal SH3 domain of c-Crk. *Structure* **1995**, *3*, 215−226.
(35) Xiang, Z.; Honig, B. Extending the accuracy limits of prediction of side-chain conformations. *J. Mol. Biol.* **2001**, *311*, 421−430.
(36) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. AMBER 8, University of California, San Francisco, 2004.
(37) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A. Point-charge force field for molecular mechanics simulations of proteins. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Phys. Chem.* **1983**, *79*, 926−935.
(39) Darden, T.; York, D.; Pedersen, L. Particale mesh Ewald: and N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(40) Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C. Numerical-integration of Cartesian equations of motion of a system with constraints − molecular-dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(41) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.
(42) SYBYL User Guide, Tripos Inc. St. Louis, U.S.A. **2003**. http://www.tripos.com.
(43) Bohacek, R. S.; McMartin, C. Definition and display of steric, hydrophobic, and hydrogen-bonding properties of ligand-binding sites in proteins using Lee and Richards accessible surface − validation of a high-resolution graphical tool for drug design. *J. Med. Chem.* **1992**, *35*, 1671−1684.
(44) Bush, B. L.; Nachbar, R. B. Sample-distance partial least-squares − PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587−619.
(45) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity − a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(46) Streitwieser, A. Molecular Orbital Theory for Organic Chemists; Wiley: New York, 1961.
(47) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parametrization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(48) Cho, S. J.; Tropsha, A. Cross-validated r(2)-guided region selection for comparative molecular-field analysis − a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060−1066.
(49) Pisabarro, M. T.; Serrano, L.; Wilmanns, M. Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J. Mol. Biol.* **1998**, *281*, 513−521.
(50) Hou, T. J.; Chen, K.; McLaughlin, W.; Lu, B. Z.; McCammon, J. A.; Wang, W. Analysis and prediction of peptide binding partners of the Abl SH3 domain, in press.

(51) Wittekind, M.; Mapelli, C.; Lee, V.; Goldfarb, V.; Friedrichs, M. S.; Meyers, C. A.; Mueller, L. Solution structure of the Grb2 N-terminal SH3 domain complexed with a ten-residue peptide derived from SOS: direct refinement against NOEs, J-couplings and 1H and 13C chemical shifts. *J. Mol. Biol.* **1997**, *267*, 933−952.
(52) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144−1149.
(53) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: A program to generate schematic diagrams of protein−ligand interactions. *Prot. Eng.* **1995**, *8*, 127−134.
(54) Kroemer, R. T.; Hecht, P.; Liedl, K. R. Different electrostatic descriptors in comparative molecular field analysis: A comparison of molecular electrostatic and Coulomb potentials. *J. Comput. Chem.* **1996**, *17*, 1296−1308.
(55) Jung, D.; Floyd, J.; Gund, T. M. A Comparative Molecular Field Analysis (CoMFA) study using semiempirical, density functional, ab initio methods and pharmacophore derivation using DISCOtech on sigma 1 ligands. *J. Comput. Chem.* **2004**, *25*, 1385−1399.
(56) Hou, T. J.; Li, Z. M.; Li, Z.; Liu, J.; Xu, X. J. Three-dimensional quantitative structure−activity relationship analysis of the new potent sulfonylureas using comparative molecular similarity indices analysis. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1002−1009.

PR0502267