An Analysis of the Interactions between the Sem−5 SH3 Domain and Its Ligands Using Molecular Dynamics, Free Energy Calculations, and Sequence Analysis

Wei Wang,[†] Wendell A. Lim,[†,‡] Araz Jakalian,[§,#] Jian Wang,[¶,£] Junmei Wang,[£] Ray Luo,[£] Christopher I. Bayly,[§] and Peter A. Kollman*[,†,£]

Contribution from the Graduate Group in Biophysics, Department of Cellular and Molecular Pharmacology, Department of Biochemistry and Biophysics, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, and Merck Frosst Canada & Company, 16711 TransCanada Highway, Kirkland, Quebec H9W 5H7, Canada

Received August 24, 2000

Abstract: The Src-homology-3 (SH3) domain of the *Caenorhabditis elegans* protein Sem−5 binds proline-rich sequences. It is reported that the SH3 domains broadly accept amide N-substituted residues instead of only recognizing prolines on the basis of side chain shape or rigidity. We have studied the interactions between Sem−5 and its ligands using molecular dynamics (MD), free energy calculations, and sequence analysis. Relative binding free energies, estimated by a method called MM/PBSA, between different substitutions at sites −1, 0, and +2 of the peptide are consistent with the experimental data. A new method to calculate atomic partial charges, AM1-BCC method, is also used in the binding free energy calculations for different N-substitutions at site −1. The results are very similar to those obtained from widely used RESP charges in the AMBER force field. AM1-BCC charges can be calculated more rapidly for any organic molecule than can the RESP charges. Therefore, their use can enable a broader and more efficient application of the MM/PBSA method in drug design. Examination of each component of the free energy leads to the construction of van der Waals interaction energy profiles for each ligand as well as for wild-type and mutant Sem−5 proteins. The profiles and free energy calculations indicate that the van der Waals interactions between the ligands and the receptor determine whether an N- or a Cα-substituted residue is favored at each site. A VC value (defined as a product of the conservation percentage of each residue and its van der Waals interaction energy with the ligand) is used to identify several residues on the receptor that are critical for specificity and binding affinity. This VC value may have a potential use in identifying crucial residues for any ligand−protein or protein−protein system. Mutations at two of those crucial residues, N190 and N206, are examined. One mutation, N190I, is predicted to reduce the selectivity of the N-substituted residue at site −1 of the ligand and is shown to bind similarly with N- and Cα-substituted residues at that site.

1. Introduction

Molecular dynamics (MD) has provided dynamic and atomic insights into complicated biological systems. Free energy calculation methods have become powerful tools to provide quantitative measurements of protein−ligand or protein−protein interactions.[1−3] A new method, molecular mechanics/Poisson Boltzmann surface area (MM/PBSA), was recently proposed for use in evaluating solvation and binding free energies of macromolecules and their complexes.[4] When this method is used to calculate binding free energy, the binding free energy is decomposed into contributions from van der Waals and electrostatic energies, nonpolar and electrostatic solvation free energies, and relative solute entropy effects.[5] The van der Waals and electrostatic interactions between the components of the complex are calculated using molecular mechanics (MM) with an empirical force field,[6] the nonpolar part of solvation free energy is estimated by empirical methods on the basis of the solvent accessible (SA) surface, and the electrostatic contribution to solvation is calculated by using a continuum model and solving the Poisson−Boltzmann (PB) equation. The entropy contribution has been estimated using normal-mode analysis.[7] An ensemble of different conformations is extracted from MD trajectories, and each snapshot is analyzed using this MM/PBSA method. The binding free energies are obtained using an ensemble average. This method is able to calculate free energy

* To whom correspondence should be addressed. Tel: (415) 476-4637. Fax: (415) 502-1411. E-mail: pak@cgl.ucsf.edu.

† Graduate Group in Biophysics, UCSF.

‡ Department of Cellular and Molecular Pharmacology, and Department of Biochemistry and Biophysics, UCSF.

§ Merck Frosst Canada & Co.

Current address: Boehringer Ingelheim (Canada) Ltd., Research and Development, 2100 Rue Cunard, Laval, Quebec, H7S 2G6, Canada.

¶ Current address: Ligand Pharmaceutical, 10275 Science Center Dr., San Diego, CA 92121.

£ Department of Pharmaceutical Chemistry, UCSF.

(1) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.

(2) Beveridge, D. L.; Dicapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.

(3) van Gunsteren, W. F. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989; pp 27−59.

(4) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(5) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.

(6) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(7) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285−290.

10.1021/ja003164o CCC: $20.00
Published on Web 04/05/2001

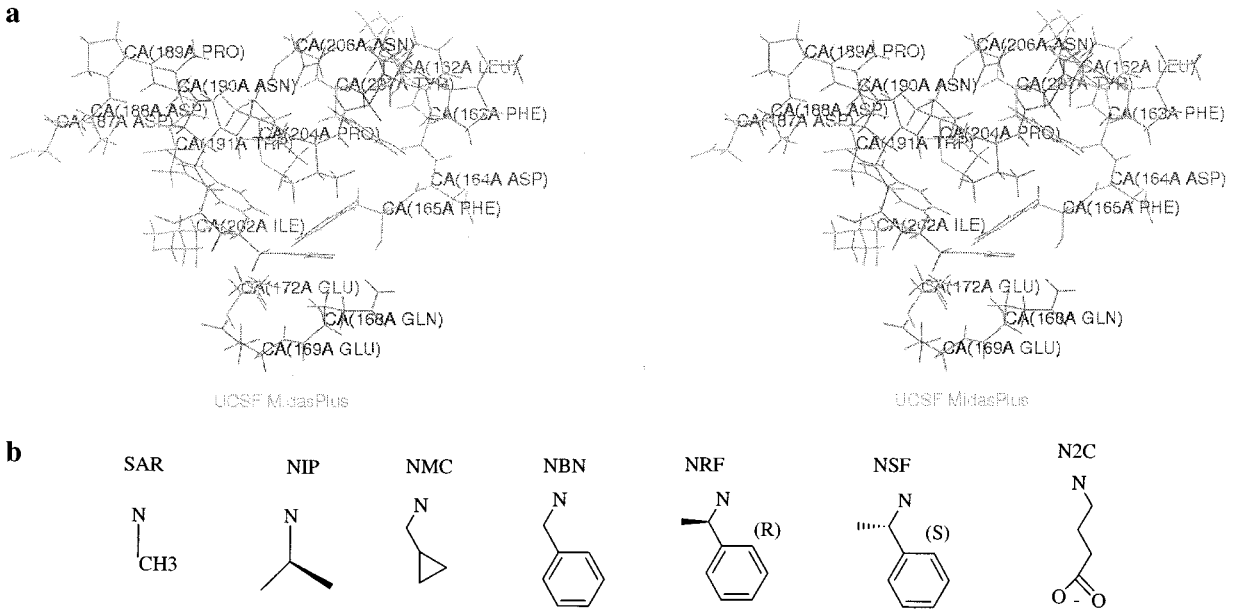

Figure 1. (a) Binding sites of the Sem−5 SH3 domain and its ligands. (b) Side chains of N-substituted peptoids at site −1 of the ligand.

differences between states even when the states are quite dissimilar from each other. It is also significantly more computationally efficient than standard free energy calculations.[1]

With the human genome sequence nearing completion and the advancement of structure genomics, analyzing the amino acid sequence and the structure of a protein can lead to predictions of functions of other proteins. For example, if several critical residues for folding stability or substrate recognition are identified for one sequence whose structure is known, it could be possible to infer which residues are crucial for other sequences whose structures are unknown, which could provide useful guidance for designing new mutagenesis experiments and deducing their functions. An empirical parameter, VC value (see below), is introduced here to serve this purpose. In this paper, we combine molecular dynamics, free energy calculation, and structure and sequence analysis to study interactions between the Sem−5 SH3 domain and its ligands. Better understanding of the SH3 domain can lead to designing potent inhibitors or engineering its specificity.

Protein−protein interactions are essential for transmission of information in cellular signaling pathways. Specific classes of protein−protein interactions are mediated by families of small modular domains. These domains, found in diverse signaling proteins, recognize small peptide motifs in partner proteins. For example, Src-homology-2 (SH2) domains bind to specific phosphotyrosyl motifs, while Src-homology-3 (SH3) domains bind to polyproline motifs. Adaptor proteins that contain both SH2 and SH3 domains can, therefore, assemble multiple proteins around an activated, phosphorylated receptor.[8−10] One example is the *Caenorhabditis elegans* protein, Sem−5, which is composed solely of one SH2 and two SH3 domains. Sem−5 protein couples the receptor, tyrosine kinase, activation to ras signaling.[11−13] The SH3 domains recognize the motif XPXXPXR, where X is any amino acid, found at the C-terminus of the exchange factor protein Sos.[11,14] Recent experimental work has focused on understanding how SH3 domains recognize the core of the PXXP motif. Lim and co-workers found that SH3 domains recognize N-substituted residues instead of only prolines at sites −1 and +2 (Figure 1). Thus, proline is selected at these sites in vivo simply because it is the only natural N-substituted amino acid. In contrast, a Cα-substituted residue is required at site 0.[15] However, little is known about the energetic factors that yield this unusual backbone substitution pattern preference.

In the present study, molecular dynamics simulations are performed on the Sem−5 SH3 domain complexed with ligands. Relative binding free energies between different ligands are calculated using the MM/PBSA method, and the results are consistent with the measured binding affinities. We show that discrimination between the N- and Cα-substituted residues at sites −1, 0, and +2 is primarily due to van der Waals interactions between the SH3 domain and the ligand. N- and Cα-substituted residues are in different conformations, and this conformational heterogeneity is an essential feature of the different binding strengths. We then focus on studying different N-substitutions at site −1 of the ligand. Relative free energies of different ligands estimated by the MM/PBSA method with RESP charges correlate reasonably well with the measured free energies. Free energy calculations have also been performed on these ligands using AM1-BCC charges,[16,17] which can be calculated significantly faster than RESP charges. Results obtained from different charge models are very similar. Since AM1-BCC charges can be easily calculated for any organic molecule, these results suggest a more robust and general application of the MM/PBSA method in drug design. To identify crucial residues for binding, we construct van der Waals interaction energy profiles for the receptor and each ligand. Multiple sequence alignment is also carried out for the Sem−5

(8) Cohen, G. B.; Ren, R.; Baltimore, D. *Cell* **1995**, *80*, 237−248.
(9) Kuriyan, J.; Cowburn, D. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 259−288.
(10) Pawson, T.; Scott, J. D. *Science* **1997**, *278*, 2075−2080.
(11) Clark, S. G.; Stern, M. J.; Horvitz, H. R. *Nature* **1992**, *356*, 340−344.
(12) Simon, M. A.; Dodson, G. S.; Rubin, G. M. *Cell* **1993**, *73*, 169−177.
(13) Olivier, J. P.; et al. *Cell* **1993**, *73*, 179−191.
(14) Lim, W. A.; Fox, R. O.; Richards, F. M. *Protein Sci.* **1994**, *3*, 1261−1266.
(15) Nguyen, J. T.; Turck, C. W.; Cohen, F. E.; Zuckermann, R. N.; Lim, W. A. *Science* **1998**, *282*, 2088−2092.
(16) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2000**, *21*, 132−146.
(17) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. Manuscript in preparation, **2000**.

SH3 domain. An empirical parameter, VC value, is implemented to identify several crucial residues on the receptor. Most of these crucial residues have also been identified in the previous experiments.[18] However, two of them, N190 and N206, were not studied before. Several mutations of these two residues are examined here. Based on the results of our free energy calculations, one mutation, N190I, has a very similar binding affinity with both of the site −1 N- and Cα-substituted ligands. Thus, the selectivity for an N-substituted residue at site −1 should be reduced for this mutant.

2. Methods

A. MD Simulations. All molecular dynamics simulations presented in this work were performed using the AMBER5.0 simulation package[19] and the Cornell et al. force field[6] with the TIP3P water model.[20] The starting structure for the wild-type Sem−5 SH3 domain, which is 58 amino acids long and bound with the PPPVPPR sequence, is taken from the Protein Data Bank. The PDB entry is 1sem. Mutations are made manually using SYBYL6.5 (Tripos Associates Inc., 1998) and MidasPlus.[21] The molecules are solvated in a 60 × 60 × 60 Å^3 box of water. An appropriate number of counterions are added to neutralize the system. Particle Mesh Ewald (PME)[22] is employed to calculate the long-range electrostatic interactions. All structures are minimized first using the SANDER module in AMBER5.0. Molecular dynamics simulations are carried out thereafter. The temperature of the system is raised gradually from 50 to 298 K, and the system is equilibrated at 298 K for 50 ps. Equilibrium is considered to be achieved after the RMSD, compared with the starting structure, reaches a plateau. Such a plateau was found within 50 ps for all the complexes. An additional 120 ps MD simulation is performed for data collection, and 100 snapshots were saved for the subsequent analysis. The average backbone heavy atom RMSDs for all trajectories are around 1 Å. The SHAKE procedure[23] is employed to constrain all bonds. The time step of the simulations is 2 fs. An 8.5 Å cutoff is used for the nonbonded interactions. The nonbonded pairs are updated every 15 steps.

B. The MM/PBSA Method. The binding free energy is calculated as:[24]

$$\Delta G_b = \Delta G_{MM} + \Delta G_{sol}{}^{LP} - \Delta G_{sol}{}^{L} - \Delta G_{sol}{}^{P} - T\Delta S \qquad (1)$$

where ΔG_b is the binding free energy in water; ΔG_{MM} is the interaction energy between the ligand and the protein; $\Delta G_{sol}{}^{L}$, $\Delta G_{sol}{}^{P}$, and $\Delta G_{sol}{}^{LP}$ are solvation free energies for the ligand, protein, and complex, respectively; and $-T\Delta S$ is the conformational entropy contribution to the binding. ΔG_{MM} is calculated from molecular mechanics (MM) interaction energies:

$$\Delta G_{MM} = \Delta G_{int}{}^{ele} + \Delta G_{int}{}^{vdw} \qquad (2)$$

where $\Delta G_{int}{}^{ele}$ and $\Delta G_{int}{}^{vdw}$ are electrostatic and van der Waals interaction energies between the ligand and the receptor, which are calculated using the CARNAL and ANAL modules in the AMBER5.0 software suite.

The solvation energy, ΔG_{sol}, is divided into two parts, the electrostatic contributions, $\Delta G_{sol}{}^{ele}$, and all other contributions, $\Delta G_{sol}{}^{nonpolar}$.

$$\Delta G_{sol} = \Delta G_{sol}{}^{ele} + \Delta G_{sol}{}^{nonpolar} \qquad (3)$$

The electrostatic contribution to the solvation free energy, $\Delta G_{sol}{}^{ele}$, is calculated using the DelPhiII software package,[25] which solves the Poisson−Boltzmann equations numerically and calculates the electrostatic energy according to the electrostatic potential. The grid size used was 0.5 Å. Potentials at the boundaries of the finite-difference lattice are set to the sum of the Debye−Huckel potentials. The value of interior dielectric constant is set to 4. As shown in our previous study,[26] after combining all the terms the binding free energy is calculated as:

$$\Delta G_b = \Delta G_{int}{}^{vdw} + \Delta G_{sol}{}^{nonpolar} + (1/n)\,\Delta G_{1-1}{}^{ele} + (\Delta G_{RFE}{}^{LP}{}_{n-80} - \Delta G_{RFE}{}^{L}{}_{n-80} - \Delta G_{RFE}{}^{P}{}_{n-80}) \qquad (4)$$

where n is the interior dielectric constant, which is 4 in this study. For comparison, free energies are also calculated using an interior dielectric constant of 1 (see Results and Discussion and Supporting Information). $\Delta G_{1-1}{}^{ele}$ is the molecular mechanics electrostatic interaction energy between the ligand and the protein. $\Delta G_{RFE}{}^{LP}{}_{n-80}$, $\Delta G_{RFE}{}^{L}{}_{n-80}$, and $\Delta G_{RFE}{}^{P}{}_{n-80}$ are reaction field energies obtained from DelPhi for the ligand, protein, and complex, respectively, with interior and exterior dielectric constants set to n and 80, respectively.

The exterior dielectric constant is set to that of water (80). The dielectric boundary is taken as the solvent accessible surface defined by a 1.4 Å probe sphere. The radii of atoms are taken from the PARSE parameter set.[27] Partial charges are taken from the Cornell et al. force field for standard amino acids. Partial charges of the nonstandard amino acids were calculated using ab initio and RESP methods.[28] AM1-BCC charges for N-substituted residues at site −1 are calculated by the AM1 semiempirical quantum method, with bond charge corrections.[16,17]

The solvent accessible surfaces (SAS) are calculated using the MSMS program.[29] The nonpolar contribution to the solvation free energy, $\Delta G_{sol}{}^{nonpolar}$, is calculated as 0.00542 × SAS + 0.92 kcal/mol.[27]

Normal-mode analysis is used to estimate conformational entropy $-T\Delta S$. Because this analysis requires extensive computer time, only three snapshots are taken in this study to estimate the order of magnitude of the conformational entropy.

C. Sequence Alignment and Definition of the VC Value. Psi-BLAST[30] with default parameters (BLOSUM62, Expect = 10, E-value threshold for inclusion in Psi-BLAST iteration = 0.002, Descriptions = 500, Alignments = 500, composition based statistics, existence gap penalty = 11, extension gap penalty = 1, λ = 0.319, and gapped λ = 0.270) is used to search the SWISS−PROT database. Multiple sequence alignment is carried out on 207 sequences, with scores >50 and E-value < 5 × 10^{-6}, using the Pileup module in the GCG software package (Version 10.1, Genetics Computer Group, Inc., 2000) with default parameters.

A parameter called the VC value (van der Waals and conservation), defined as the product of the conservation percentage of an amino acid at the Sem−5 SH3 domain and its van der Waals interaction energy with the ligand, is used to identify critical residues for binding. The conservation percentage reflects how conserved the amino acid is and, therefore, it is the sum of the appearance percentage (no gap included) of that specific amino acid and similar ones at a certain position. Appearance percentage reflects how often a specific amino acid appears at a certain position. For example, at position F163 of the Sem−5 SH3 domain, Tyr and Phe have 61% and 37% appearance percentage,

(18) Lim, W. A.; Richards, F. M. *Nat. Struct. Biol.* **1994**, *1*, 221−225.
(19) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(21) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 13−27.
(22) Darden, T. A.; York, D. M.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(23) Rychaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(24) Kollman, P. A.; Massova, I.; Reyes, C. M.; Kuhn, B.; Huo, S.; Chong, L. T.; Lee, M. R.; Lee, T. S.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(25) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327−335.
(26) Wang, W.; Kollman, P. A. *J. Mol. Bol.* **2000**, *303*, 567−582.
(27) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(28) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(29) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.
(30) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J. H.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

Table 1. Binding Free Energies of Sem−5 SH3 Domain with Its Ligands (Mutations at Different Sites)

ligand	exptl ΔG_b (kcal/mol)	ΔG_{int}^{vdw} (kcal/mol)	ΔG_{int}^{ele} (kcal/mol)	ΔG^{nonpol} (kcal/mol)	ΔG_{sol}^{ele} (kcal/mol)	$\Delta G_{int+sol}^{ele}$ [a] (kcal/mol)	ΔG_b [b] (kcal/mol)	$\Delta\Delta G_b$ [c] (kcal/mol)
WT	−5.1	−37.4 ± 0.4	−156.4 ± 2.3	−3.7 ± 0.0	+41.0 ± 0.6	+1.9 ± 0.0	−39.3 ± 0.4	−
SAR−1[d]	−4.4	−35.1 ± 0.5	−152.1 ± 6.7	−3.6 ± 0.0	+39.7 ± 1.1	+1.7 ± 0.6	−37.0 ± 0.0	−1.5
ALA−1[d]	>−2.7	−33.4 ± 0.7	−157.0 ± 3.4	−3.6 ± 0.0	+40.8 ± 1.3	+1.6 ± 0.4	−35.5 ± 0.3	0.0
SAR0[e]	>−2.7	−32.2 ± 0.7	−156.1 ± 2.9	−3.6 ± 0.0	+40.2 ± 0.7	+1.2 ± 0.0	−34.6 ± 0.7	+2.2
ALA0[e]	−4.0	−34.3 ± 0.6	−155.8 ± 0.9	−3.7 ± 0.0	+40.2 ± 0.3	+1.3 ± 0.1	−36.8 ± 0.6	0.0
PRO0[e]	−4.8	−36.3 ± 1.1	−156.4 ± 2.3	−3.7 ± 0.1	+40.7 ± 0.2	+1.6 ± 0.4	−38.3 ± 0.8	−1.5
SAR+2[f]	−4.4	−34.4 ± 0.0	−163.3 ± 1.8	−3.6 ± 0.0	+42.3 ± 0.5	+1.5 ± 0.0	−36.5 ± 0.0	−2.0
ALA+2[f]	>−2.7	−32.2 ± 0.1	−159.1 ± 2.1	−3.4 ± 0.0	+40.9 ± 0.4	+1.1 ± 0.1	−34.5 ± 0.2	0.0

[a] $\Delta G_{int+sol}^{ele} = \Delta G_{int}^{ele} + \Delta G_{sol}^{ele}$. [b] $\Delta G_b = \Delta G_{int}^{vdw} + \Delta G_{int}^{ele}$ (ϵ_{in}=1, ϵ_{out}=1)/4 + ΔG^{nonpol} + ΔG_{sol}^{ele}. [c] $\Delta\Delta G_b$ is calculated for each site. [d] SAR−1 and ALA−1 refer to mutating Pro at site −1 to Sar and Ala, respectively. [e] SAR0, ALA0, and PRO0 refer to mutating Val at site 0 to Sar, Ala, and Pro, respectively. [f] SAR+2 and ALA+2 refer to mutating Pro at site +2 to Sar and Ala, respectively.

respectively, in the multiple sequence alignment. Therefore, the conservation percentage of F163 is 98%.

3. Results and Discussion

A. MM/PBSA Analysis Accounts for Observed SH3 Domain Site Preferences. Lim and co-workers reported that in the SH3 domains, sites −1 and +2 favor N-substituted residues and site 0 favors a Cα-substituted residue. In their study, the representative N- and Cα-substituted residues are sarcosine (Sar) (Figure 1b) and alanine (Ala), respectively.[15] We present here a computer modeling study to provide atomic and dynamic insights of how wild-type and mutant ligands interact with the receptor.

A molecular dynamics (MD) simulation has been performed on the wild-type peptide bound to the Sem−5 SH3 domain. The binding free energy, calculated by the MM/PBSA method (see Methods), is −39.3 kcal/mol (Table 1). Small errors (Table 1) and plateau RMSD compared with the crystal structure (data not shown) suggest convergence of the trajectory. Due to the considerable CPU cost for calculating the entropy contribution to the binding free energy, we only estimate the order of magnitude of the entropy contribution. We assume that the entropy contributions are similar for different ligands because all ligands in our study are just one residue different from the wild-type. The entropy contribution estimated by normal-mode analysis on three conformations is +29.0 ± 1.0 kcal/mol. If this entropy term is included in the calculation, the absolute binding free energy for the wild-type peptide is −10.3 kcal/mol, which is of the same order of magnitude as the measured value of −5.1 kcal/mol.

Binding free energies are also calculated for substitutions at sites −1, 0, and +2 (Table 1). At sites −1 and +2, Sar-substitutions are more favorable than Ala-substitution, and at site 0, Ala is preferred over Sar. Thus, these results are able to reproduce the trend that Ala−1, Sar0, and Ala+2 bind significantly less well than the wild-type sequence, and the remaining mutants (Sar−1, Ala0, Pro0, and Sar+2) bind only slightly less well, which is qualitatively consistent with the experimental data.[15] The correlation coefficient r^2, between relative calculated and experimental binding free energies, is 0.88 (Figure 2). In summary, the MM/PBSA analysis accurately reproduces the ligand site preferences for the Sem−5 SH3 domain.

B. van der Waals Interactions between the Ligand and the Protein Is the Dominant Factor for Site Preferences. As we mentioned above, what energetic factors determine the site preferences are not clear. One opinion is that desolvation is the determinant factor for substituting −NH with $-NCH_3$. To address this problem, we compare correlations between the measured binding free energies and each component of the calculated binding free energies (Table 2). We find that van

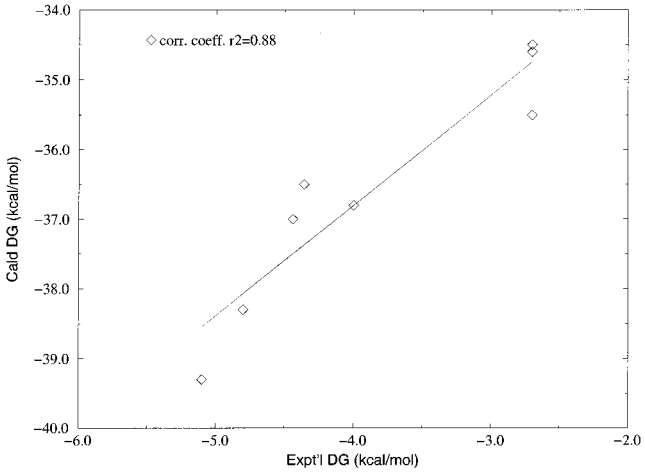

Figure 2. For sites −1, 0, and +2, correlation between measured binding free energy and calculated free energy using RESP charges.

Table 2. Correlation Coefficients r^2 between Measured Binding Free Energies and Different Components of Calculated Ones

component	correlation coefficient, r^2
van der Waals	0.88
electrostatic (Coloumb term)	0.0088
solvation penalty (PB term)	0.026
electrostatic + solvation penalty	0.52
SA	0.45

der Waals energy has the best correlation (r^2 is 0.88). There is no correlation between the measured binding free energies and the electrostatic interaction energy (Coulomb term) ($r^2 = 0.0088$) or the electrostatic solvation energy (PB term) ($r^2 = 0.026$). However, these two terms compensate for each other, and their sum has a better r^2, which is 0.52. The solvent accessible surface term does not correlate well with the measured binding free energies either ($r^2 = 0.45$). It is obvious that van der Waals interactions between the ligand and the receptor are the dominant factor in site preferences.

The average van der Waals interaction energies during the trajectories between the protein and each residue of the ligand are calculated (Supporting Information). Analyzing van der Waals profiles and complex structures, the following pictures of structural changes are suggested for substitutions at sites 0, −1, and +2.

At site 0, discrimination between N- and Cα-substitution is mainly due to the interaction difference between the Trp191 and the site 0 residue. The average distances between the Trp191 CH2 atom and CB in the Ala0 or the CD atom in the Sar0 are 5.5 and 8.2 Å, respectively.

The mutation at site −1 causes a global change of the ligand van der Waals energy profile. In addition to the pair of residues

Table 3. Binding Free Energies of Sem−5 SH3 Domain with Its Ligands Obtained from Computational Mutagenesis Using the Wild-Type Peptide Trajectory

ligand	exptl ΔG_b (kcal/mol)	ΔG_{int}^{vdw} (kcal/mol)	ΔG_{int}^{ele} (kcal/mol)	ΔG^{nonpol} (kcal/mol)	ΔG_{sol}^{ele} (kcal/mol)	$\Delta G_{int+sol}^{ele}$ [a] (kcal/mol)	ΔG_b [b] (kcal/mol)	$\Delta\Delta G_b$ [c] (kcal/mol)
WT	−5.1	−37.4 ± 0.4	−156.4 ± 2.3	−3.7 ± 0.0	+41.0 ± 0.6	+1.9 ± 0.0	−39.3 ± 0.4	−
SAR−1[d]	−4.4	−33.9 ± 0.5	−155.8 ± 2.1	−3.6 ± 0.0	+40.8 ± 0.4	+1.8 ± 0.1	−35.7 ± 0.4	−0.2
ALA−1[d]	>−2.7	−33.9 ± 0.4	−156.4 ± 2.1	−3.7 ± 0.0	+41.1 ± 0.6	+2.0 ± 0.0	−35.5 ± 0.5	0.0
SAR0[e]	>−2.7	−35.1 ± 0.3	−155.5 ± 2.3	−3.7 ± 0.0	+40.5 ± 0.6	+1.6 ± 0.0	−37.2 ± 0.3	+0.3
ALA0[e]	−4.0	−35.6 ± 0.3	−155.9 ± 2.3	−3.7 ± 0.0	+40.8 ± 0.6	+1.8 ± 0.0	−37.5 ± 0.3	0.0
SAR+2[f]	−4.4	−35.1 ± 0.1	−156.6 ± 2.4	−3.6 ± 0.0	+41.1 ± 0.5	+2.0 ± 0.1	−36.7 ± 0.2	−0.8
ALA+2[f]	>−2.7	−34.5 ± 0.1	−156.2 ± 2.4	−3.7 ± 0.0	+41.3 ± 0.6	+2.2 ± 0.0	−35.9 ± 0.1	0.0

[a] $\Delta G_{int+sol}^{ele} = \Delta G_{int}^{ele} + \Delta G_{sol}^{ele}$. [b] $\Delta G_b = \Delta G_{int}^{vdw} + \Delta G_{int}^{ele}$ (ϵ_{in}=1, ϵ_{out}=1)/4+ $\Delta G^{nonpol} + \Delta G_{sol}^{ele}$. [c] $\Delta\Delta G_b$ is calculated for each site. [d] SAR−1 and ALA−1 refer to mutating Pro at site −1 to Sar and Ala, respectively. [e] SAR0, ALA0, and PRO0 refer to mutating Val at site 0 to Sar, Ala, and Pro, respectively. [f] SAR+2 and ALA+2 refer to mutating Pro at site +2 to Sar, and Ala, respectively.

at site −1, the pairs at sites −2 and +2 also have differences of more than 0.5 kcal/mol. Analysis of the profile of the ligand and protein, as well as the φ angle of the ligand residues suggests the following picture of conformational changes due to mutation. For Sar−1, since Sar has a smaller side chain than Pro in the wild-type and thus less attraction to N206, N206 moves toward Pro+2. The distance between N206 CG and Pro+2 CD is 4.7 and 4.3 Å in wild-type and Sar−1, respectively. The φ angle of Pro+2 is larger than that of the wild-type, which means it moves into the pocket formed by F163, N206, and Y207. To maximize the interactions, Sar at site −1 moves toward N206. This makes the peptide more "helical", and Val0 inserts deeper into the pocket formed by F165, W191, P204, and Y207 (more favorable van der Waals interactions for Val0). Since Sar−1 drags Pro−2 and Arg−3 along with it, Pro−2 moves toward N190 and makes more contacts with N190. However, Arg−3 has less favorable van der Waals interactions with Gln168 and Glu172 as it moves a little away from these residues. For the Ala−1 mutant, Pro+2 also intends to move toward N206 and Y207 (larger φ angle). However, since there is no N-substituted group in Ala at site −1, N206 is more flexible. The interactions between Pro+2 and N206/Y207 are similar to those found in the wild-type, but weaker than those found in Sar−1. The side chain of Ala−1 also keeps N190 from moving closer to Pro−2 to compensate for some interactions, as in Sar−1.

If the Pro at site +2 is mutated to Sar (Sar+2) or Ala (Ala+2), the primary difference is from Pro at site +3. The reason is that the new residue (Sar or Ala) has to adjust its conformation to have optimal interactions with both N206/Y207 and F163. Therefore, Sar+2 moves toward N206/Y207 and it brings Pro+3 closer to F163. Pro+3 even has a more favorable van der Waals interaction energy than the wild-type. However, in Ala+2 this adjustment is in the opposite direction, toward F163, which pushes Pro+3 even farther away from the receptor. This introduces the major difference between Sar+2 and Ala+2 (Figure 1).

C. Conformational Changes of Ligands Are Important for Site Preferences. To address the importance of conformational changes of ligands for site preferences, binding free energies for substitutions at sites −1, 0, and +2 with Sar and Ala, respectively, were also calculated using only the trajectory obtained for the wild-type complex (Table 3). The underlying assumption is that the single mutation does not induce significant conformational change of the complex, just like in the computational alanine scanning simulations.[5] This "alanine" scanning approach can only be used if the mutated residue is smaller than the wild-type, which is the case for Pro→Sar or Pro→Ala mutations. From Table 3 we can see that the calculated difference between Sar- and Ala-substitution is small. The van der Waals interaction energies are very similar at site −1 and just slightly different for sites 0 and +2. The calculated ΔG values correlate rather poorly with experimental values (r^2 = 0.34). This suggests that using the wild-type trajectory to estimate the ΔG of mutants is a poor approximation because we have neglected the subtle conformational changes that occur when a residue is substituted. The success of "computational alanine scanning" in the MDM2-p53 protein−protein complex[5] is likely due to the relatively rigid backbone structure; the p53 complex remains α-helical upon Ala mutation. However, in Sem−5 SH3 domain complexes, the ligands are more flexible, and the assumption of backbone rigidity is less accurate.

We also calculated binding free energies using an interior dielectric constant of 1 instead of 4 (see Supporting Information). If we use a single trajectory, the results also correlate poorly with the experimental data, as we found using an interior dielectric constant of 4. However, using separate trajectories, the calculations are consistent with the experimental measurements, as was found with an interior dielectric constant of 4 (r^2 = 0.88).

In summary, substitution dependent conformational flexibility must be taken into account to accurately calculate the observed differences in binding.

D. MM/PBSA Method with Different Charge Models, RESP and AM1-BCC, Can Reasonably Reproduce Relative Binding Free Energies of N-Substituted SH3 Peptoids at Site −1. We next focus on the site −1 and examine different N-substituted peptoids binding to the wild-type receptor. Nguyen et al. used 12-residue peptoids YEVPPPVXPRRR (X is a synthesized, nonnatural residue) in their study of mutations at site −1.[15] Since no crystal structure is available for any entire peptoid, we mutate the residue at site −1 in the shorter ligand, PPPVPPR, whose crystal structure has been solved. We assume that the relative binding free energies of different peptoids do not have significant changes in the longer or shorter peptoid. In Table 4, we present the results of using separate trajectories on different site −1 peptoids, using MM/PBSA to calculate their free energies of binding. Our calculated ΔG values correlate reasonably well with the measured values, with a correlation coefficient (r^2) of 0.78 (N2C excluded, see below).

The largest outlier is N2C. It is worth pointing out that N2C is the only charged residue at site −1 in our calculation. If N2C is included, the correlation coefficient r^2 is 0.60. The van der Waals interaction energy for N2C is not much less favorable than that of the wild-type. The sum of its Coulomb term and the electrostatic contribution to solvation (PB term) is much less favorable compared to other ligands. However, this term is not unfavorable enough.

The stereoisomers of NSF and NRF have similar solvation penalties. The difference between their binding affinities is due to their different protein binding patterns. The phenyl ring of NSF has close contacts with P204, Y207, and F165, while its

Table 4. Binding Free Energies of Sem−5 SH3 Domain with Site −1 Mutant Ligands

ligand	exptl ΔG_b (kcal/mol)	ΔG_{int}^{vdw} (kcal/mol)	ΔG_{int}^{ele} (kcal/mol)	ΔG^{nonpol} (kcal/mol)	ΔG_{sol}^{ele} (kcal/mol)	$\Delta G_{int+sol}^{ele}$ [a] (kcal/mol)	ΔG_b [b] (kcal/mol)	$\Delta\Delta G_b$ [c] (kcal/mol)
ALA	>−2.7	−33.4 ± 0.7	−157.0 ± 3.4	−3.6 ± 0.0	+40.8 ± 1.3	+1.6 ± 0.4	−35.5 ± 0.3	0.0
N2C	−3.48	−38.3 ± 0.7	−115.6 ± 2.5	−4.0 ± 0.0	+32.4 ± 0.4	+3.5 ± 0.4	−38.7 ± 0.5	−3.2
NRF	−4.32	−34.3 ± 1.5	−135.6 ± 5.2	−4.4 ± 0.6	+34.6 ± 1.1	+0.7 ± 0.2	−38.0 ± 1.0	−2.5
SAR	−5.45	−35.1 ± 0.5	−152.1 ± 6.7	−3.6 ± 0.0	+39.7 ± 1.1	+1.7 ± 0.6	−37.0 ± 0.0	−1.5
NSF	−5.82	−37.4 ± 1.2	−162.7 ± 0.4	−3.8 ± 0.1	+41.6 ± 0.5	+0.9 ± 0.4	−40.3 ± 0.8	−4.8
WT	−5.89	−37.4 ± 0.4	−156.4 ± 2.3	−3.7 ± 0.0	+41.0 ± 0.6	+1.9 ± 0.0	−39.2 ± 0.4	−3.7
NMC	−5.97	−37.7 ± 0.7	−156.8 ± 2.5	−3.7 ± 0.0	+40.6 ± 0.5	+1.4 ± 0.1	−40.0 ± 0.6	−4.5
NIP	−6.27	−37.7 ± 0.2	−155.5 ± 3.0	−3.9 ± 0.0	+41.2 ± 0.5	+2.3 ± 0.2	−39.3 ± 0.1	−3.8
NBN	−6.32	−41.1 ± 1.1	−157.0 ± 4.3	−4.0 ± 0.1	+42.2 ± 0.1	+2.9 ± 0.0	−42.2 ± 1.2	−6.7

[a] $\Delta G_{int+sol}^{ele} = \Delta G_{int}^{ele} + \Delta G_{sol}^{ele}$. [b] $\Delta G_b = \Delta G_{int}^{vdw} + \Delta G_{int}^{ele}$ (ϵ_{in}=1, ϵ_{out}=1)/4+ $\Delta G^{nonpol} + \Delta G_{sol}^{ele}$. [c] $\Delta\Delta G_b$ is relative to ALA.

Table 5. Binding Free Energies of Sem−5 SH3 Domain with Site −1 Mutant Ligands Using AM1-BCC Charges

ligand	exptl ΔG_b (kcal/mol)	ΔG_{int}^{vdw} (kcal/mol)	ΔG_{int}^{ele} (kcal/mol)	ΔG^{nonpol} (kcal/mol)	ΔG_{sol}^{ele} (kcal/mol)	$\Delta G_{int+sol}^{ele}$ [a] (kcal/mol)	ΔG_b [b] (kcal/mol)	$\Delta\Delta G_b$ [c] (kcal/mol)
ALA	>−2.7	−33.4 ± 0.7	−157.0 ± 3.4	−3.6 ± 0.0	+40.8 ± 1.3	+1.6 ± 0.4	−35.5 ± 0.3	0.0
N2C	−3.48	−38.3 ± 0.7	−112.3 ± 2.6	−4.0 ± 0.0	+32.1 ± 0.5	+4.0 ± 0.2	−38.2 ± 0.6	−2.7
NRF	−4.32	−34.3 ± 1.5	−135.2 ± 4.5	−4.4 ± 0.6	+35.8 ± 1.0	+2.0 ± 0.1	−36.7 ± 0.9	−1.2
SAR	−5.45	−35.1 ± 0.5	−147.7 ± 7.6	−3.6 ± 0.0	+39.6 ± 1.2	+2.7 ± 0.7	−36.0 ± 0.1	−0.5
NSF	−5.82	−37.4 ± 1.2	−157.2 ± 0.0	−3.8 ± 0.1	+41.8 ± 0.5	+2.5 ± 0.4	−38.7 ± 0.8	−3.2
WT	−5.89	−37.4 ± 0.4	−156.4 ± 2.3	−3.7 ± 0.0	+41.0 ± 0.6	+1.9 ± 0.0	−39.2 ± 0.4	−3.7
NMC	−5.97	−37.7 ± 0.7	−156.3 ± 2.2	−3.7 ± 0.0	+41.0 ± 0.5	+1.9 ± 0.1	−39.5 ± 0.7	−4.0
NIP	−6.27	−37.7 ± 0.2	−147.7 ± 3.4	−3.9 ± 0.0	+39.9 ± 0.6	+3.0 ± 0.3	−38.6 ± 0.0	−3.1
NBN	−6.32	−41.1 ± 1.1	−158.3 ± 4.0	−4.0 ± 0.1	+43.8 ± 1.0	+4.3 ± 0.1	−40.9 ± 1.2	−5.4

[a] $\Delta G_{int+sol}^{ele} = \Delta G_{int}^{ele} + \Delta G_{sol}^{ele}$. [b] $\Delta G_b = \Delta G_{int}^{vdw} + \Delta G_{int}^{ele}$ (ϵ_{in}=1, ϵ_{out}=1)/4+ $\Delta G^{nonpol} + \Delta G_{sol}^{ele}$. [c] $\Delta\Delta G_b$ is relative to ALA.

methyl group points toward N206. However, NRF's phenyl ring packs with N190, and its methyl group points away from the protein. This is reflected in the profiles, as F165, P204, N206, and Y207 have more favorable van der Waals interactions with NSF than with NRF (see Supporting Information).

NIP and NMC are similar. Both of them interact with N206/Y207 through a methyl group, and meanwhile, both have favorable interactions with N190/W191. NMC has a longer side chain. It packs better with W191 than does NIP. However, N190 is pushed a little farther away from the peptide by NMC. The total van der Waals energies of NMC and NIP are similar. Their slightly different binding affinities are due to their different electrostatic contributions. This suggests that the van der Waals interactions dominate in the binding and electrostatic interactions, determine the selectivity, and "fine-tune" the binding strength as well.

NBN packs perfectly with both N190/W191 and N206/Y207. That is why it has such a favorable van der Waals interaction energy. However, its solvation penalty is larger too. This is probably due to the burial of the polar phenyl ring.

In the previous MM/PBSA calculations, an interior dielectric constant of 1 has been used to be consistent with the molecular mechanics force field.[24] Nonetheless, the dielectric constant inside a protein is considered to be in the range of 2−4. To make our model more realistic, we used a value of 4 in this study. Our calculations on site −1 N-substituted peptoids show that the results obtained using a dielectric constant of 4 correlate noticeably better with experimental data (r^2 = 0.78 N2C excluded) than those using a value of 1 (r^2 = 0.21 N2C excluded) (see Supporting Information).

In the above calculations, all atomic charges are calculated using the RESP module in AMBER. This procedure requires a significant effort for each charge determination. Recently, Bayly and co-workers developed a new algorithm to calculate partial charges for atoms, the AM1-BCC charges.[16,17] These are calculated to emulate a HF/6-31G* electrostatic potential around the molecule, as are the RESP charges, only at a fraction of the computational effort. We recalculated the binding free energies for all site −1 peptoids using AM1-BCC charges (Table 5). The results are reasonably well correlated with experimental data (r^2 = 0.64 N2C excluded) and those obtained from RESP charges (r^2 = 0.86 N2C included). This suggests that one can combine the RESP charge for the protein and the AM1-BCC charge for the ligand in applying the MM/PBSA method in drug design.

E. The VC Parameter Allows Identification of SH3 Residues Critical for Binding and Specificity. (i.) Combination of Energetic and Evolutionary Information Can Be Useful in Identifying Critical Residues for Binding. In this section, we focus on studying critical residues in the SH3 domain. As discussed above, the electrostatic solvation penalty compensates for the Coulomb energy, and van der Waals interactions dominate the site preferences. Here, we combine the evolution conservation information with the molecular mechanics energy to evaluate the significance of each residue for binding or stability. The VC value (van der Waals and conservation) is calculated for each residue as the multiple of its van der Waals interaction energy with the ligand and its conservation percentage in the multiple sequence alignment. It is worth noting that we combine the appearance percentages of similar residues, such as Q and N, D and E, and Y and F (see Methods). We observe that critical residues have larger VC values than unimportant residues.

(ii) Critical Residues Identified by VC Value Are Consistent with Findings in the Previous Experiments. First, van der Waals interaction energies between several critical residues of the protein and the ligand are calculated (Supporting Information). The residues with higher than 1 kcal/mol van der Waals interaction energies (absolute value) can be roughly divided into four groups (with some overlaps). The first group includes F165, Q168, E169, E172, and W191, which interact with Arg−3. The second group consists of N190, W191, P204, and N206. They have strong interactions with Pro−1. F165, W191, and P204 also form the third group that interacts with Val0. N206 constitutes the fourth group along with F163 and Y207, they interact with Pro+2.

In the first group, F165, Q168, E169, E172, and W191 have 37, 10, 22, 44, and 95% appearance percentages, respectively, in our 207 sequences obtained from a Psi-BLAST search in the SWISS−PROT database (Supporting Information). Their van der Waals interaction energies are −1.8, −3.0, −1.8, +1.2, and −7.2 kcal/mol, respectively. W191 has the most favorable van der Waals interaction energy and is also well conserved. It is not surprising that no mutant examined experimentally at position W191 can bind with the polyproline peptide.[18] At the F165 position, Phe and Tyr have 37 and 61% appearance percentages, respectively. F165 forms the hydrophobic core of the binding pocket with Pro204, Trp191, and Tyr207. This implies that F165 is more crucial for stabilizing the receptor rather than for binding ligands. This is a possible explanation for the fact that no mutant (e.g., F165V, F165S, F165A, and F165G) but F165L can bind to the peptide. This is presumably because only Leu among those examined residues can still stabilize the hydrophobic core. This tentative explanation will require experimental measurement of the stabilities of F165 mutants to be definitive. Q168 is on the surface and has not been studied experimentally either. Various residues appear at this position in different species, Asn, Gln, and Glu in Crk, Grb2, and Src proteins, respectively. This implies that Q168 is tolerant to mutations. E172 forms crucial hydrogen bonds with Arg−3 to keep Arg−3 in the right position to interact with W191 and Q168. Glu also appears at this position in other species (e.g., Grb2 proteins) to perform the same function. E169 seems to assist E172 in fixing Arg−3, but it is not as crucial as E172 because its 22% appearance percentage is relatively low. In the previous study, double mutations at E169/E172 are shown to have a significant effect on binding.[18]

Each residue in the second group, N190, W191, P204, and N206, has a favorable van der Waals interaction energy of more than 2 kcal/mol. W191, P204, and N206 are well conserved (100 and 73% appearance percentages, respectively, for P204 and N206) while N190 is not well conserved (16% appearance percentage). N190 is part of the binding pocket and interacts strongly with the site −1 residue. However different species have different residues at this position. For example, Crk, Grb2, and Src proteins have Gln, Asn, and Asp residues, respectively, at this position. Our speculation is that this residue may be responsible for substrate specificity. As mentioned above, P204 is part of the hydrophobic core of the binding pocket, and it was shown in the previous experiments to be crucial for the stability of the Sem−5 structure.[18] N206 interacts strongly with residues at sites −1 and +2. It forms a hydrogen bond with the peptide backbone (Supporting Information). It may also be important for keeping the crucial residue Y207 (see below) in the right position. This may explain why this residue is well conserved in different species. No mutations have been studied for Sem−5 N190 or N206.

In the fourth group, F163, N206, and Y207 have strong favorable van der Waals interaction energies, −5.0, −3.9, and −6.1 kcal/mol, respectively. Their appearance percentages are 27, 73, and 84%, respectively. At the F163 position, although Phe is not the dominant residue (27% appearance percentage), Tyr, which also has a phenyl ring, has the highest appearance percentage, 63%. F163 forms one edge of the binding pocket and interacts with Pro at site +2. In the previous study, the F163V mutant shows no binding with the peptide, but F163A does.[18] It is worth pointing out that in the sequence alignment, Ala has a 4% conservation percentage at the F163 position. Thus, position 163 is critical for selectivity of the Sem−5 protein. However, it may be tolerant for a Tyr or Ala mutation with weaker binding. Ala has a smaller side chain, and it allows the peptide to move closer to the pocket. Therefore, the lost interactions between the peptide and the protein due to F163A mutation may be recovered to some extent. However, Val keeps the peptide from approaching closer to the receptor and, thus, the lost interaction cannot be recovered. At the Y207 position, Phe has a 12% appearance percentage, which may explain why only the Y207F mutation does not disrupt the binding with the peptide, as found previously.[18]

Table 6. Critical Residues for Binding Have Larger VC Values than Unimportant Ones

residue	substitution sensitivity (exptl data)	VC value	van der Waals (kcal/mol)	conservation percentage (%)
F163	yes	4.6	−5.0	91
F165	yes	1.8	−1.8	98
E169/E172	yes	0.6/1.0	−1.8/+1.2	33/80
W191	yes	6.8	−7.2	95
P204	yes	2.4	−2.4	100
N206	not studied	2.9	−3.9	73
Y207	yes	5.9	−6.1	96
L162	no	0.3	−0.4	67
D164	no	0.7	−0.9	80
Q168	not studied	0.5	−3.0	16
D187/D188	no	0.0/0.1	−0.02/−0.2	12/54
N190	not studied	0.7	−2.8	24
I202	no	0.1	−0.3	38

It is also worth pointing out that several residues, L162, D164, D187, D188, and I202, shown to be unimportant for binding in ref 18, appear to have much weaker van der Waals interactions with the ligand (<0.5 kcal/mol).

In summary, residues having strong van der Waals interaction energies and being well conserved, such as W191 and Y207, play significant roles in binding affinity and specificity. These are "hot spots" which are not tolerant to mutation. Residues having strong van der Waals interaction energies and being diversified in different species determine specificity. F163 is such an example; if it is mutated to residues appearing in other species, specificity will be reduced. The binding probably will not be completely disrupted, i.e., the substrate probably still can bind, but with a weaker binding affinity. Residues having moderate van der Waals interaction energies but being well conserved are crucial for stabilizing the protein, e.g. F165 and P204. Only those mutations which still can stabilize the protein are tolerated at these positions. Residues having weak van der Waals interaction energies and being varied in different species usually are not important. From our studies, these observations appear in other systems as well (W. Wang and Kollman, unpublished data). One caveat is that charged residues forming strong hydrogen bonds with the ligand should be examined case by case, such as E172.

From Table 6, we can see that the VC value can identify (≥1.0) F163, F165, E172, W191, P204, N206, and Y207 as crucial residues (Table 6). Two residues, Q168 and E169, with strong van der Waals interaction energies but low conservation, and L162 and D164, with high conservation but weak van der Waals interaction energies, have low VC values (Table 6). This suggests the advantage of using the VC value over only using van der Waals energies or conservation information. The significance of E172 may be underestimated, as we point out above. Evaluating the significance of residues for the binding interactions in this way is being tested for other systems (W. Wang and Kollman, unpublished data), and the preliminary results are encouraging. We hope that this VC value can serve as a guide for mutagenesis experiments in the future. We can

Table 7. Sar−1 and Ala−1 Ligands Interact with Mutant Proteins

ligand	exptl ΔG_b (kcal/mol)	ΔG_{int}^{vdw} (kcal/mol)	ΔG_{int}^{ele} (kcal/mol)	ΔG^{nonpol} (kcal/mol)	ΔG_{sol}^{ele} (kcal/mol)	$\Delta G_{int+sol}^{ele}$ [a] (kcal/mol)	ΔG_b [b] (kcal/mol)	$\Delta\Delta G_b$ [c] (kcal/mol)
ALA−1/N190Q	n/a	−36.0 ± 0.4	−143.1 ± 0.7	−3.8 ± 0.0	+38.8 ± 0.3	+3.0 ± 0.1	−36.8 ± 0.5	0.0
SAR−1/N190Q	n/a	−37.1 ± 0.5	−170.8 ± 3.6	−4.0 ± 0.0	+44.9 ± 0.6	+2.2 ± 0.3	−38.9 ± 0.2	−2.1
ALA−1/N190Q/N206Q	n/a	−34.3 ± 0.9	−158.5 ± 3.7	−3.7 ± 0.0	+41.4 ± 0.8	+1.7 ± 0.1	−36.2 ± 1.0	0.0
SAR−1/N190Q/N206Q	n/a	−37.3 ± 0.1	−182.1 ± 3.0	−3.8 ± 0.0	+46.9 ± 0.8	+1.4 ± 0.0	−39.6 ± 0.2	−3.4
ALA−1/N190I	n/a	−37.2 ± 1.0	−140.5 ± 1.8	−3.8 ± 0.1	+37.5 ± 0.7	+2.4 ± 0.2	−38.6 ± 0.8	0.0
SAR−1/N190I	n/a	−36.4 ± 0.1	−147.5 ± 1.6	−3.8 ± 0.0	+38.4 ± 0.3	+1.6 ± 0.1	−38.6 ± 0.1	0.0
ALA−1/N190L	n/a	−32.9 ± 1.0	−175.4 ± 2.3	−3.7 ± 0.1	+45.1 ± 0.5	+1.2 ± 0.0	−35.4 ± 1.0	0.0
SAR−1/N190L	n/a	−39.2 ± 0.6	−161.0 ± 2.2	−4.1 ± 0.0	+43.4 ± 0.2	+3.1 ± 0.7	−40.2 ± 0.1	−4.8

[a] $\Delta G_{int+sol}^{ele} = \Delta G_{int}^{ele} + \Delta G_{sol}^{ele}$. [b] $\Delta G_b = \Delta G_{int}^{vdw} + \Delta G_{int}^{ele}$ (ϵ_{in}=1, ϵ_{out}=1)/4+ $\Delta G^{nonpol} + \Delta G_{sol}^{ele}$. [b] $\Delta\Delta G_b$ is calculated for each type of mutant receptor.

predict that the counterpart residues of F163, F165, E172, W191, P204, N206, and Y207 in other SH3 domains are most crucial for binding.

(iii) N190 Is the Possible Residue Critical for N-Substituted Recognition. The VC value profile leads us to engineer the receptor for tighter binding with site −1 Cα-substituted ligands. Since Ala−1 has more favorable van der Waals interactions with Q168 than does Sar−1 and Y207 is well conserved, the only choices for mutation are N190 and N206, if we want to avoid mutating Y207, which might introduce significant conformational changes. Since N206 forms hydrogen bonds with the peptide backbone, we do not want to disrupt it either. We first tried mutating N190 and N206 to the similar residue Gln. From Table 7, we can see that N190Q and N190Q/N206Q as calculated bind more tightly with Ala-substituted ligands than with the wild-type receptor, which is due to more favorable van der Waals interactions. However, these two mutants are also calculated to bind more tightly with Sar-substituted ligands. Because Gln is similar to Asn, this may suggest that Gln also can maintain this selectivity. As a support of this interpretation, we observe that Gln occupies the N190 position in many sequences, especially in Crk proteins in the multiple sequence alignment. We next calculated two hydrophobic residues, Leu and Ile, at the 190 position. The N190L mutant binds more favorably with Sar−1 than Ala−1 as well. However, the N190I mutant has a very similar binding affinity with Sar−1 and Ala−1. This is because I190 takes a conformation where the short branch of the side chain interacts with Ala−1 and the longer branch can have some favorable interactions with Pro−2 and Arg−3. In all other mutations including N190L, Ala−1 keeps residue 190 away from the peptide. These simulations also suggest that having Asn or Gln at position 190 is crucial for the specificity of N-substituted residues at site −1. If a suitable hydrophobic residue is in position 190 (which can be a nonnatural amino acid), the selectivity for an N-substituted residue of site −1 can be reduced or even reversed. This observation is consistent with the fact that no hydrophobic residue appears at the 190 and 206 positions in our sequence alignment. This suggestion awaits experimental examination.

4. Conclusions

We have presented herein a combination of molecular dynamics, free energy calculations, and sequence alignments to study interactions between the Sem−5 SH3 domain and its ligands. These analyses shed light on understanding SH3 domain−ligand interactions. We have shown that subtle conformational changes of the ligands, due to whether they have N- or Cα-substituted residues, is crucial for reproducing the relative binding free energies since the calculated ΔG's obtained from separate trajectories correlate much better with measured ones than those from a single trajectory. These conformational changes can also be seen from the φ angle profile of different ligands (Supporting Information).

It is also interesting that our results are consistent with experiment by using, in each of the separate trajectories, the ensemble of ligand conformations that exist in the complex. This suggests that these bound conformations are at least representative of the free ligand ensemble, which could not be assumed for small peptides. Perhaps the rigidity of the proline residues enables this to be a good approximation.

In this study, we also test different interior dielectric constants and charge models while applying the MM/PBSA method to estimate the ΔG of binding for flexible ligands. Different dielectric constants have been examined both in simulations where one considers different residues at various sites (sites −1, 0, and +2) and where one considers many residues at the −1 site. Although results at different sites are similar for $\epsilon = 1$ and 4, $\epsilon = 4$ gives noticeably better results at site −1 than $\epsilon = 1$. A new charge model called AM1-BCC,[16,17] which can be calculated for any organic molecules much more efficiently than the RESP[28] charges that are used in the protein force field, has been shown to give comparable results to the RESP charges in the MM/PBSA calculations. It will thus be possible to use RESP charges for proteins and AM1-BCC charges for ligands, which will make the MM/PBSA method more efficient and generally applicable in structure based ligand design.

Discrimination of N- and Cα-substituted residues at different sites in the ligand is shown to be primarily due to van der Waals interactions between the ligand and the receptor. By calculating the van der Waals contribution of each residue to binding and analyzing conservation at each position, we are able to identify several important residues of the receptor, most of which had been shown to be crucial for binding in prior experiments. Our analysis also suggested that mutation at N190 may reduce the selectivity for N- over Cα-substituted residues at site −1, and our free energy calculations further suggest that a specific mutation N190I may bind both types of peptides equally well. This prediction awaits experimental testing.

It was pointed out to us by a reviewer that in protein folding studies, the common folding nucleus for several protein families was identified by looking at the number of contacts that certain amino acids make and how conserved they are.[31,32] This suggests that conserved residues with good intra- or intermolecular packing are crucial for folding or binding. The VC value proposed in this study is the first quantitative parameter to combine energetic and evolutionary information. It thus might be useful for studying protein folding as well.

(31) Ptitsyn, O. B. *J. Mol. Biol.* **1998**, *278*, 655−666.
(32) Ptitsyn, O. B.; Ting, K. L. H. *J. Mol. Biol.* **1999**, *291*, 671−682.

Acknowledgment. This work was supported in part by NIH grant GM-29072 to P.A.K. and GM55040-01 to W.A.L. Part of this investigation was completed using the facilities in the UCSF Computer Graphics Lab, T. Ferrin, director, supported by grant RR-1081 from the NIH. Supercomputer time at the NSF centers (NCSA and SDSC) is gratefully acknowledged.

Supporting Information Available: Tables of binding free energies calculated using different interior dielectric constants, van der Waals interaction energies between single residue of the SH3 domain and different ligands, van der Waals interaction energies between single residue of the ligands and the whole SH3 domain, hydrogen bonds between the ligands and the SH3 domain, and statistics data of the sequence analysis (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA003164O