Insights from the energetics of water binding at the domain-ligand interface of the Src SH2 domain

Gianni De Fabritiis,[1][*][†] Sebastien Geroult,[2,3][*][†] Peter V. Coveney,[4] and Gabriel Waksman[2,3,5][*]

[1] Computational Biochemistry and Biophysics Laboratory (GRIB), Universitat Pompeu Fabra,
Barcelona Biomedical Research Park (PRBB), C/Dr Aiguader 88, 08003 Barcelona, Spain

[2] Institute of Structural Molecular Biology, UCL and Birkbeck, London WC1E 7HX, United Kingdom

[3] School of Crystallography, Birkbeck College, London WC1E 7HX, United Kingdom

[4] Department of Chemistry, Centre for Computational Science, University College London, London WC1H 0AJ,
United Kingdom

[5] Department of Biochemistry and Molecular Biology, University College London, London WC1E 6BT, United Kingdom

ABSTRACT

SH2 domains play important roles in signal transduction by binding phosphorylated tyrosine residues on cell surface receptors. In an effort to understand the mechanism of ligand binding and more specifically the role of water, we have designed a general computational protocol based on the potential of mean force to compute the thermodynamics of water molecules at the protein–ligand interface for two SH2 domain complexes of the Src kinase, those bound to the two peptides Ac-PQpYEpYI-NH2 and Ac-PQpYIpYV-NH2 where pY indicates a phosphotyrosine. These two peptides were chosen because they have similar binding affinities but very different entropic/enthalpic thermodynamic binding signatures, indicating different interactions with solvent. We find that the isoleucine to valine mutation at position +3 (the third amino acid C-terminal to pY) in the ligand has only limited impact on the water structure. By contrast, the glutamic acid to isoleucine mutation at position +1 has a significant impact by not only abrogating a local hydrophilic binding site but, more importantly and surprisingly, inducing a favorable nonlocal entropic contribution from the water molecules around the phophorylated tyrosine at the +2 position. Our study demonstrates the validity of the method reported here for exploring the thermodynamic solvation landscape of protein–protein interactions.

INTRODUCTION

SH2 domains are small protein domains present in a large number of signal transducing proteins. They are comprised of ~100 amino acids and bind specifically to tyrosine-phosphorylated sequences within proteins involved in signal transducing pathways. The SH2 domain structure consists of a large (central β-sheet flanked by two α-helices. The tyrosine phosphorylated peptide binds the SH2 domain perpendicular to the central β-sheet. The central β-sheet divides the structure in two regions. The first region is located on the N-terminal side of the central (β-sheet). In this region, the phosphotyrosine in the phosphopeptide inserts into a positively-charged binding-pocket. The second region is located on the C-terminal side of the central β-sheet and interacts with the C-terminal residues of the peptide ligand.[1–3]

The binding specificity of SH2 domains was first investigated using a degenerate phosphopeptide library in which the three first residues C-terminal to the phosphotyrosine were randomized.[4–6] These studies resulted in the identification of classes of SH2 domains differing in the types of sequence motifs they preferentially bind to. In the case of the SH2 domains of the Src kinase family, it was shown that a sequence EEI C-terminal to the phosphotyrosine appeared to generate substantial binding affinity. This was confirmed by subsequent studies investigating the details of binding specificity in the Src SH2 domain[7,8] where it was shown that, in the EEI motif, the +3 Ileu appears to be the more important of the three motif's residues.

Table I

Thermodynamic Signatures of pYEpYI- and pYIpYV-Binding to the Src SH2 Domain

	ΔH (Kcal/Mol)	$T\Delta S$ (Kcal/Mol)	ΔG (Kcal/Mol)
pYEpYI	−7.0	2.7	−9.7
pYIpYV	−1.5	7.2	−8.7
$\Delta\Delta$ (Kcal/Mol)	−5.5	−4.5	−1.0

Reprinted from J Mol Biol, 328, Lubman OY, Waksman G, Structural and thermodynamic basis for the interaction of the Src SH2 domain with the activated form of the PDGF β-receptor; pp 655–668, © 2003, with permission from Elsevier.

In the platelet-derived growth factor (PDGF) signaling pathway, the Src kinase binds preferentially to a region of the PDGF β-receptor with the sequence pYIpYV (where pY indicates a phosphotyrosine) i.e. an Ile, phosphotyrosine, and Val at the +1, +2, and +3 positions C-terminal to the first phosphotyrosine. This region is located in the juxta-membrane sequence of the receptor. The structure and thermodynamic signature of this ligand have been investigated to gain a better understanding of the role played by the phosphotyrosine at the +2 position of the pYIpYV ligand in the binding process.[9] A mutation to pY at the +2 position of a peptide containing the pYEEI sequence (pYEpYI) resulted in a more favorable affinity. It was also observed that, while the binding of a pYEpYI-containing phosphopeptide is enthalpically driven (i.e., the contribution of the enthalpy change to the binding free energy dominates the binding thermodynamics), the binding of a pYIpYV-containing peptide is entropically driven (i.e., the contribution of the entropy change to the binding free energy dominates the binding thermodynamics; see Table I).

The structures of the pYEpYI- and pYIpYV-containing phosphopeptides bound to the Src SH2 domain were determined using X-ray crystallography at a resolution of 1.9 and 2.1 Å, respectively (1NZL and 1NZV in the protein data bank, respectively), and unveiled 13 and 9 crystallographic water molecules at their binding interface respectively. This difference in water structure appeared to be the only significant one as all other aspects of the structures are very similar (the crystallization conditions of the two complexes were identical). Four subsequent studies using very different computational methods attempted to make sense of the contrasted binding thermodynamic signatures of the pYEpYI- and pYIpYV-containing ligands.[10–13] Using molecular dynamics (MD) and thermodynamic integration (TI) on a high performance computing grid, Fowler et al.[10,13] investigated the $\Delta\Delta G$ between two alchemical mutations at +1 and +3 positions of the ligand, mutating from pYEpYI to pYIpYV. This study attempted *inter alia* to reproduce the experimental ΔG value of −1.0 ± 0.2 kcal/mol between the two dually phosphorylated peptides. Increasing the duration of simulations to enhance the sampling of phase space improved the quality of

results; however the computed value underestimated the experimental measurements by 3 kcal/mol, suggesting that the approach is not accurate enough to be used for double alchemical mutations, in contrast to the situation pertaining for single point mutations. Two other studies by Geroult *et al.* (2006, 2007) used the surface area method pioneered by Freire and Amzel (reviewed in Baker and Murphy[14]) to compute binding thermodynamics.[11,12] This method, simple in its principles and implementation, failed to provide predicted binding thermodynamic parameters anywhere close to the observed values, even in a rigorous and systematic treatment of interfacial water molecules. However, the failure was at least in part ascribed to a flawed ΔC_p (heat capacity) equation and it was shown that, provided the ΔC_p is known, one can expect to derive a reasonable estimate of the binding entropy change.

The present work seeks to investigate the interactions of the Src SH2 domain with the ligands Ac-PQpYEpYI-NH2 and Ac-PQpYIpYV-NH2 in terms of the thermodynamics of solvation, that is the free energy cost $W(\mathbf{x})$ associated with the removal of a water molecule from the surface of the SH2 domain-ligand complex at any position $\mathbf{x}$. The free energy $W(\mathbf{x})$ is computed via the potential of mean force (PMF) method of displacing reversibly a water molecule to the bulk from any small volume localized around the protein. By subtracting the different water binding free energies around the two protein–ligand systems, we can derive the free energy difference of water binding between the two systems at any position $\mathbf{x}$ (see Fig. 1). The information on the thermodynamics of water molecules mediating binding is then exploited to gain insights into the overall mechanism of phosphopeptide-binding to the Src SH2 domain. We demonstrate here that water binding energetics is a useful "informant" to describe parts of the mechanism of protein–ligand interaction mediated by SH2 domains.

The use of PMF calculations by direct unbiased measurement over a MD trajectory is a simple and useful method to derive the thermodynamics of small molecules when the states are sufficiently well populated.[15,16] In the work reported here, we use the PMF method to probe the water energetic landscape at the binding interface and thereby *indirectly* derive information related to the overall free energy landscape of the protein-ligand complex. We are able to compute the work from the PMF formula without any special sample biasing protocol due to the abundance of water at the protein-ligand interface, which makes the statistical sampling of states from the MD trajectories entirely reliable. This significantly reduces the computational cost of the approach as well as the difficulty of implementing the protocol. This protocol furnishes rapid qualitative insight into protein-ligand binding and is a useful complement to more involved calculations of the binding affinity using standard free energy perturbation (FEP)[17] and steered molecular dynamics (SMD) methods,[18] which, by contrast,
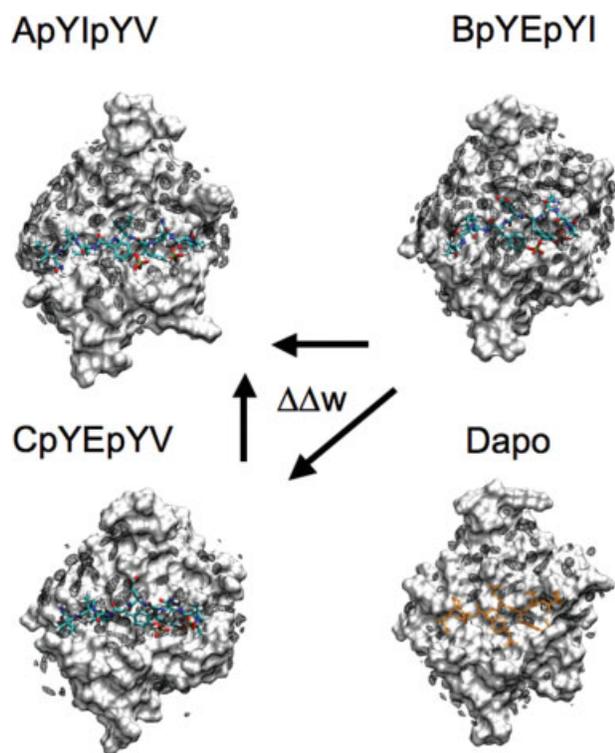
Figure 1

Thermodynamic cycles used to compute the free energy differences of water molecules across a different set of protein–ligand interfaces. The Src SH2 domain is represented by its water accessible surface area colored in white and the isosurfaces in grey chicken-wire represent $w = W(r)$ with $w = -0.5$ Kcal/ mol highlighting water binding sites. The identity of the bound peptides in panels A, B, and C is indicated next to the A, B, and C labels. The peptides in those panels are shown in stick representation color-coded by atom types (C, N, O, and H atoms in cyan, blue, red, and white, respectively). In panel D where the apo form of the SH2 domain is represented, a pYEpYI peptide (in stick representation color-coded in orange) is shown to help locate the binding site.

require a larger computational and human effort in terms of preparation of the calculations and computer time. It also compares favorably with algorithms for computing binding sites for water which probe the energy landscape with respect to topology, electrostatic and Lennard–Jones potentials.[19,20] However, those approaches neglect entropic effects and therefore fail to recover the free energy surface associated with the binding site. This is the major advantage of our new protocol which allows us to reconstruct the water thermodynamics and thereby to compare directly different structures by computing free energy differences of water binding at protein–ligand interfaces.

THEORY AND METHODS

The PMF $W(\xi)$ along a reaction coordinate $\xi = \xi(R^N)$ is defined up to an arbitrary constant by Kirkwood[21] as:

$$W(\xi) = -K_b T \log(\rho(\xi)) \qquad (1)$$

in which the probability density $\rho(\xi)$ of encountering ξ, is given by the statistical mechanical ensemble average:

$$\rho(\xi) = \frac{\int dR^N dP_R^N \delta(\xi(R^N) - \xi) \exp(-\beta H)}{\int dR^N dP_R^N \exp(-\beta H)}$$
$$= \langle \delta(\xi(R^N) - \xi) \rangle, \qquad (2)$$

where H is the Hamiltonian of the molecular system (based here on the CHARMM force field[22,23]), R^N and P^N the atom positions and momenta, $\beta = 1/k_B T$, and δ is the Dirac delta function.

We define the reaction coordinate $\xi_{x,a}$ to be the occupancy of water molecules within a sphere $S_{x,a}$ of radius a centered at position x in three-dimensional (3D) space. More formally, given M water molecules and the positions $r_1,...,r_M$ of their oxygen atoms, the reaction coordinate is defined as $\xi_{x,a} = \sum_i^M \chi_{x,a}(r_i)$, where $\chi_{x,a}$ is the characteristic function of the sphere $S_{x,a}$ centered in x of radius a, that is $\chi(x) = 1$ if $x \in S$ and 0 otherwise. The probability density $\rho(\xi_{x,a})$ can be directly measured from a standard molecular dynamics trajectory by counting the presence of water molecules within the sphere $S_{x,a}$ for a given probe radius a and over a set of spatial positions x. For instance, we have used a regular lattice embedding the protein-ligand complex with a lattice spacing of 0.5 Å and radius $a = 1$ Å. The density is then directly used to determine the PMF in Eq. (1). The PMF difference $\Delta W = W(x_2) - W(x_1)$ between two reference positions x_2 and x_1 is the free energy difference required to reversibly move a water molecule from ξ_{x1} to ξ_{x2} (see Fig. 1).

Within this description, a water molecule is said to be in the bulk if its free energy does not change on moving it further away from the protein–ligand complex. The reference bulk free energy $W(x_1)$ when x_1 is in bulk water can be computed by the most probable value of the free energy histogram of the W values in the lattice. This reference value W_{bulk} is identical between different simulations on different ligands; it can then be subtracted from the free energy to give the reference delta free energy between x_2 and bulk water (x_1). Computing the occupancy density at different neighboring sites also gives us an indication of the degree of convergence of our occupancy averages which are better than 0.1 kcal/mol in the bulk. This protocol allows us to compute the reversible work required to move a water molecule from the bulk water environment to any volume $S_{x,a}$ close to the SH2 domain, and to build a 3D map and isosurface of the free energy profile obtained.

Computational protocol for pmf calculations

We used the bound crystallographic structures of the Src SH2 domain in the PDB files 1NZL and 1NZV to set up the MD simulations of the SH2 domain/pYEpYI ligand and the SH2 domain/pYIpYV ligand complexes, respectively. The starting structure for the SH2 domain/

pYEpYV ligand complex was generated from the crystal structure of the SH2 domain/pYEpYI ligand complex by mutating the isoleucine to a valine residue. The SH2 domain in the apo form is also taken from the PDB file 1NZL (this structure contains indeed two molecules of SH2 domain in the asymmetric unit, one bound to the doubly phosphorylated peptide and the other unbound). All MD simulations were carried out using the NAMD program[24] with the CHARMM27 forcefield[23] and TIP3P water. During equilibration the position of the protein heavy atoms were kept fixed with the hydrogen atoms and water molecules allowed to evolve for 200 ps at ambient temperature (298 K). The backbone of the protein, and the phospho-peptide, were then restrained by a small harmonic potential with a spring constant of 2 kcal/mol/ Å^2 and thermalised from 108 to 298 K in steps of 10 K over 200 ps. The magnitude of the restraining spring constant was then reduced to 0.5 kcal/mol/Å^2 and finally to zero. A Berendsen barostat was applied to maintain the pressure at 1 atm. The simulation cell was finally equilibrated for at least 2 ns in the isothermal-isobaric (NPT) ensemble.

The PMF calculation uses trajectories generated during a 10 ns simulation run in the NPT ensemble at 298 K and 1 atm for the three equilibrated SH2 domain complex structures (SH2 domain/pYEpYI ligand, SH2 domain/pYIpYV ligand, SH2 domain/pYEpYV ligand complexes) and also for the apo form (see Fig. 1). The 10 ns production MD trajectories were postprocessed in order to ensure perfect alignment. First, the proteins were re-centered in the simulation box and the water molecule positions transformed within the periodic domain. Then every trajectory was aligned on a chosen reference initial position of the C_α atoms comprising the SH2 domain. This procedure allows us to compare the relative positions of the water molecules around the SH2 domain/ligand complex system. For every trajectory, each configuration is partitioned into small cubic volumes of side 0.5 Å and the average number of water molecules counted within the spherical volume $S_{x,l}$ centered at the grid point $\mathbf{x}$ with radius 1 Å. We obtain the density $\rho(\xi_{x,a})$ after normalizing for the number of frames and the PMF from the definition in Eq. (1), resulting in $W_{\text{SH2+pYEpYI}}(\mathbf{x})$, $W_{\text{SH2+pYIpYV}}(\mathbf{x})$, $W_{\text{SH2+pYEpYV}}(\mathbf{x})$, and $W_{\text{apo}}(\mathbf{x})$. The value of the PMF in the bulk W_{bulk} is then subtracted from the volumetric free energy data to obtain the free energy landscapes of water binding at the interface of the protein-ligand systems $\Delta W_{\text{SH2+pYEpYI}}(\mathbf{x})$, $\Delta W_{\text{SH2+pYIpYV}}(\mathbf{x})$, $\Delta W_{\text{SH2+pYEpYV}}(\mathbf{x})$, and $\Delta W_{\text{apo}}(\mathbf{x})$. The four isosurfaces at $w = -0.5$ Kcal/mol are shown in Figure 1.

From the three data sets of the ligand-bound structures we can further compute the "ΔΔ" free energies between different SH2 complexes, which highlight the differences in binding sites of water molecules between different ligands. We compute the free energy changes for the dual mutation $\Delta\Delta W_{\text{IV-EI}}(\mathbf{x}) = \Delta W_{\text{SH2+pYIpYV}}(\mathbf{x}) - \Delta W_{\text{SH2+pYEpYI}}(\mathbf{x})$, and for the single mutations $\Delta\Delta W_{\text{EV-EI}}(\mathbf{x}) = \Delta W_{\text{SH2+pYEpYV}}(\mathbf{x}) - \Delta W_{\text{SH2+pYEpYI}}(\mathbf{x})$, and $\Delta\Delta W_{\text{IV-EV}}(\mathbf{x}) = \Delta W_{\text{SH2+pYIpYV}}(\mathbf{x}) - \Delta W_{\text{SH2+pYEpYV}}(\mathbf{x})$. Using the same protocol, we have also computed the free energy difference for the doubly mutated ligands pYEpYI and pYIpYV in bulk water $\Delta\Delta W_{\text{IV-EI, bulk}}(\mathbf{x}) = \Delta W_{\text{pYI-pYV}}(\mathbf{x}) - \Delta W_{\text{pYEpYI}}(\mathbf{x})$ with water coordinates restrained to their crystallographically determined bound positions, in order to compare the "bare" effect of the mutation in bulk with that when bound to the SH2 domain. All volumetric data are visualized by isosurfaces using the volume rendering capabilities of VMD.[25] All necessary files to reproduce the results (psf, pdb, cube files) can be found as supplementary files in supplementary materials.

RESULTS

We use the PMF method to provide an explanation of the differences observed between the binding thermodynamics of the pYEpYI and pYIpYV peptides to the Src SH2 domain (shown in Table I) in terms of a quantitative definition of hydrophobicity. In the following, we look at the results on the apo form of SH2 domain and the ΔW change ($\Delta\Delta W$) between these two ligands by using the singly mutated intermediate pYEpYV peptide to help differentiate the contribution from each single amino acid mutation.

A quantitive definition of hydrophobicity/hydrophilicity

The definition of the PMF given in Eq. (1) allows us to give a clear, quantitative definition of hydrophobicity. A small volume centered in $\mathbf{x}$ is hydrophobic if $\Delta W(\mathbf{x}) = W(\mathbf{x}) - W_{\text{bulk}} > 0$, that is it requires positive work to reversibly displace a water molecule from the bulk to the volume centered in $\mathbf{x}$. In contrast, hydrophilic sites are characterized by negative work $\Delta W < 0$. The isosurface at a given value of the PMF $w = \Delta W(\mathbf{x})$ defines a solvent iso-energetic surface. Water molecules experience a zero average thermodynamic force when moving on this surface (i.e., they undergo free diffusion there), while the direction normal to this surface at each point represents the least likely permeation direction for water molecule diffusion because it corresponds to the greatest energetic barrier. Volumes characterized by a negative free energy value are therefore binding sites for water molecules. The information obtainable by this methodology is displayed in Figure 1 where the isosurfaces for -0.5 Kcal/Mol are shown highlighting the different binding sites for water at the protein–ligand interface.

ΔW of the Src SH2 domain apo form

In all panels of Figure 2, the pYEpYI ligand is shown in a semitransparent stick representation superimposed
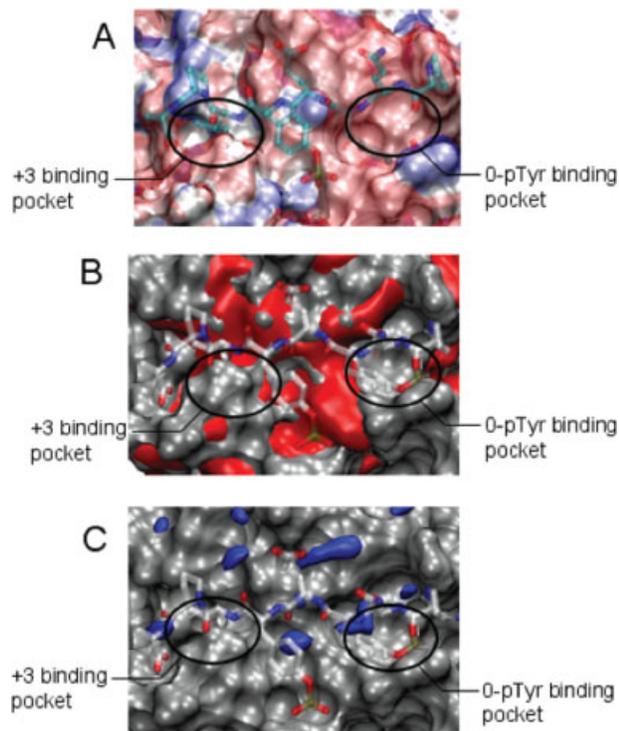
Figure 2

ΔW of the Src SH2 apo form. In **A**, the Src SH2 apo form is shown with the $+3k_bT/e$ (red) and $-3k_bT/e$ electrostatic potential (blue) isosurface. A large positive electrostatic region characterizes the binding pocket. In **B** and **C**, the free energy ΔW of the Src SH2 apo form computed with the PMF method is represented with 3D isosurfaces; ΔW with values higher than $+3k_bT$ and lower than $-1.5k_bT$ are illustrated by red and blue isosurfaces, respectively. In each panel, the crystallographic structure of the pYEpYI ligand has been added at the domain interface for illustrative purposes and is shown in semi-transparent stick representation color-coded by atom types red, blue, and gray for oxygen, nitrogen, and carbon atoms, respectively. The Src SH2 domain water solvation accessible surface area is represented in gray.

ligand. The binding pocket is expected to be hydrophobic because it would require a lower free energy to displace water molecules in order for the ligand to bind. In Figure 2(C), ΔW of the Src SH2 domain apo form is represented by free energy 3D isosurfaces with values lower than $-1.5k_BT$ (blue). The negative ΔW domains highlight the hydrophilic sites of the apo form.

$\Delta\Delta W$ of Src SH2 domain bound to pYEpYI and pYIpYV

The free energy change $\Delta\Delta W_{\text{IV-EI}}(\mathbf{x}) = \Delta W_{\text{SH2}+\text{pYIpYV}}(\mathbf{x}) - \Delta W_{\text{SH2}+\text{pYEpYI}}(\mathbf{x})$ provides the difference in solvation energy between the SH2 domain/pYEpYI ligand and SH2 domain/pYIpYV ligand complexes. In Figure 3(A), the 3D isosurfaces of $\Delta\Delta W_{\text{IV-EI}}$ are drawn for potentials of
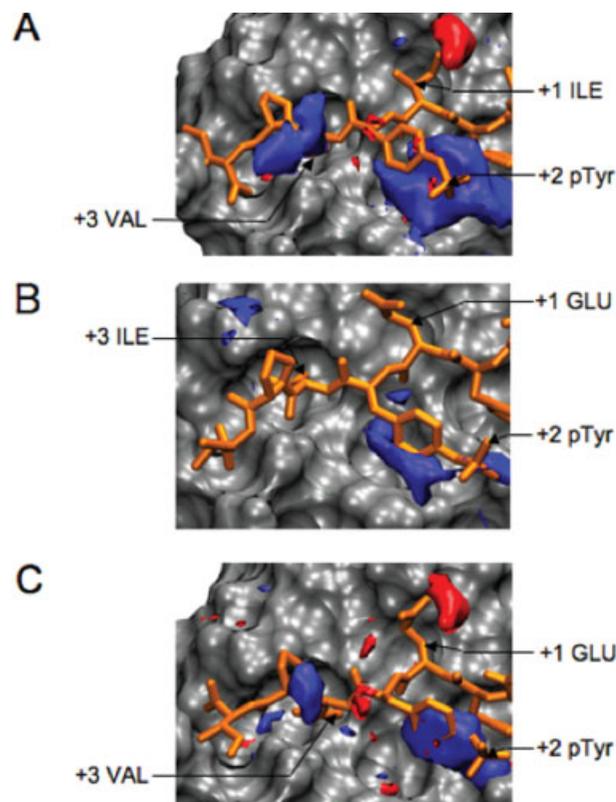
Figure 3

$\Delta\Delta W(\mathbf{r})$ between pYEpYI-, pYEpYV-, and pYIpYV-SH2 domain complexes.(**A**) $\Delta\Delta W_{\text{IV-EI}}(\mathbf{x}) = \Delta W_{\text{SH2}+\text{pYIpYV}}(\mathbf{x}) - \Delta W_{\text{SH2}+\text{pYEpYI}}(\mathbf{x})$ between the pYEpYI- and pYIpYV-SH2 domain complexes. The 3D isosurfaces of $\Delta\Delta W$ with a value higher than $+4k_bT$ and lower than $-4k_bT$ are represented in red and blue color, respectively. The time-averaged structure of the pYIpYV ligand is shown at the domain interface in orange stick representation. The solvation accessible surface area of the Src SH2 domain interface is represented in a grey isosurface. (**B**) $\Delta\Delta W(\mathbf{r})$ between pYEpYI- and pYEpYV-SH2 domain complexes. The color coding is the same as in A. The time-averaged structure of the pYEpYI ligand is shown at the domain interface in orange stick representation. (**C**) $\Delta\Delta W(\mathbf{r})$ between the pYEpYV- and pYIpYV-SH2 domain complexes. The color coding is the same as in A. The time-averaged structure of the pYEpYV ligand is shown at the domain interface in orange stick representation.

on the apo SH2 domain to help locate the binding interface. First, the electrostatic potential (U) of the Src SH2 domain in its apo form is computed using APBS (adaptive Poisson–Boltzmann solver, http://apbs.sourceforge.net)[26] and shown as an isosurface plot in Figure 2(A) at the potential values of $U = +3k_BT/e$ (red color) and $U = -3k_BT/e$ (blue color). Interestingly, the binding interface shows an electrostatic potential manifestly positive, a feature that attracts negatively-charged ligands such as the pYEEI, pYIpYV or pYEpYI peptides. The static information on the electrostatic potential can be complemented with van der Waals interactions and entropic effects which contribute to the full thermodynamic landscape. In our PMF method, these effects are explicitly taken into account. This is illustrated by the hydrophobic free energy difference $\Delta W_{\text{SH2-apo}}(\mathbf{x}) = +3k_BT$ isosurface (red) shown in Figure 2(B). The hydrophobic region coincides reasonably well with the binding site of the

Figure 4

$\Delta\Delta W(\mathbf{r})$ between the apo form of the pYEpYI (orange) and pYIpYV (green) ligands restrained in their crystallographic conformations. The three-dimensional isosurfaces of $\Delta\Delta W$ with values higher than $+3k_bT$ and lower than $-3K_bT$ are represented in red and blue, respectively.

$+4k_BT$ (red) and $-4k_BT$ (blue) representing the positive and negative free energy differences in solvation from the SH2 domain/pYEpYI complex to the SH2 domain/pYIpYV complex. The free energy isosurface $\Delta\Delta W_{\text{IV-EI}}$ between the two complex structures shows a strongly reduced water affinity at the +1 position. This is to be expected from the nature of the mutated residues. Indeed, the same mutations for the ligands in bulk water (without the SH2 domain) reproduce a similar energetic profile (see Fig. 4) due to removal of the polar groups of the glutamic acid, which makes the region more hydrophobic. The I–V mutation at the +3 position also reduces steric interactions, creating more available volume for water binding but it has no direct solvation effect within the protein–ligand complex because it is not directly accessible by water. However, the $\Delta\Delta W_{\text{IV-EI}}$ isosurface between the bound structures also exhibits a strong higher hydrophilic component at the +2 position. This is unexpected as it is not present in $\Delta\Delta W_{\text{IV-EI}}$ computed from the peptides alone (see Fig. 4). Thus, the negative $\Delta\Delta W_{\text{IV-EI}}$ observed at the +2 position in the bound structures is due to the interaction with the protein. We next investigated which of the two residues mutated at the +1 and +3 positions is responsible for this effect.

$\Delta\Delta W$ between Src SH2 bound to pYEpYI and pYEpYV

We applied the PMF protocol to the Src SH2 domain/pYEpYV peptide complex and computed the free energy change $\Delta\Delta W_{\text{EV-EI}}(\mathbf{x}) = \Delta W_{\text{SH2+pYEpYV}}(\mathbf{x}) - \Delta W_{\text{SH2+pYEpYI}}(\mathbf{x})$. In Figure 3(B), the isosurfaces $\Delta\Delta W_{\text{EV-EI}}$ with a value higher than $+4k_BT$ (red) and lower than $-4k_BT$ (blue) are shown. We did not observe significant changes of solvation close to the +3 position in the solvation isosurface because, as explained earlier, the region is buried inside the protein–ligand interface. A small change in the

solvation $\Delta\Delta W_{\text{EV-EI}}$ isosurface is observed close to the phosphotyrosine at the +2 position. It appears that the presence of an isoleucine compared to a valine at +3 position produces a gain in solvation around the +2 position, leading to increased water affinity under the ring of the phosphotyrosine in the SH2 domain/pYEpYV peptide complex.

$\Delta\Delta W$ between Src SH2 bound to pYEpYV and pYIpYV

Finally, we investigated $\Delta\Delta W_{\text{IV-EV}}(\mathbf{x}) = \Delta W_{\text{SH2+pYIpYV}}(\mathbf{x}) - \Delta W_{\text{SH2+pYEpYV}}(\mathbf{x})$ between the Src SH2 domain/pYEpYV peptide and/pYIpYV peptide complexes. The $\Delta\Delta W_{\text{IV-EV}}$ isosurface is shown in Figure 3(C), where we observe a loss of solvation at the +1 position (as expected, see Fig. 4) due the replacement of a hydrophilic by a hydrophobic residue. However, we observe a net gain in the solvation energetics at the +2 phosphotyrosine and +4 positions.

Monitoring the mobility of the +2 pY side chain

As both mutations at the +1 and +3 position impact on the hydrophilicity under the +2-pY position, we investigated a potential impact of this increased hydrophilicity on the mobility of the +2-pY side chain along the MD trajectory. We monitored two distances: D1 between the +2-pY C_ζ atom and the C_α atom of the N-terminal pY (the least mobile atom in the ligand) and D2 between +2-pY C_ζ atom of the ligand and the C_α atom of LysβD6 in the SH2 domain [Fig. 5(A)]. The distances between the atoms were computed for each frame, smoothed with moving means ($\pm$50 ps) and plotted as a function of frame number. Figures 5(B,C) show the variation of the distances D1 and D2, respectively, during the simulation. We observed a reduced displacement fluctuation of the +2-pY side chain in the SH2 domain/pYEpYI peptide complex (blue trace) compared to the SH2 domain/pYEpYV peptide complex (red trace) or SH2 domain/pYIpYV peptide complex (green trace). Fluctuations of this side chain in the SH2 domain/pYIpYV peptide complex were much larger than in the SH2 domain/pYEpYV peptide complex. Also, the +2 side chain in the ligand appears to be more anchored to the SH2 domain interface in the SH2 domain/pYEpYI peptide complex than in the SH2 domain/pYIpYV peptide complex [Fig. 5(C)]. Thus, there appears to be a correlation between solvation and greater mobility of the +2-pY side chain in the SH2 domain/pYIpYV peptide complex.

DISCUSSION

In our study, the role of solvation in influencing binding thermodynamics of the Src SH2 domain to doubly phos-
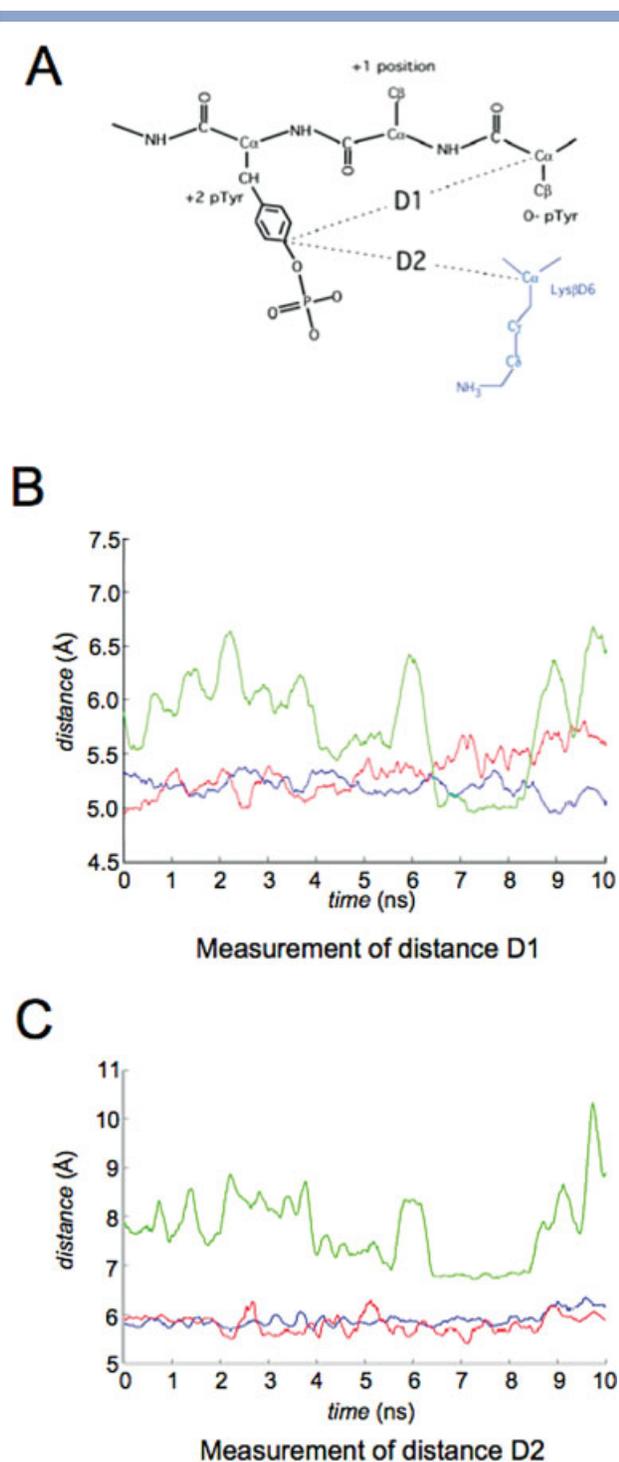
Figure 5

Variation in the distance D1 between the +2-pTyr C$_\zeta$ atom and the C$_\alpha$ atom of the 0-pTyr and distance D2 between the +2-pTyr C$_\zeta$ atom and the C$_\alpha$ atom of LysβD6 in the SH2 domain. (A) Schematic diagram defining D1 and D2. (B) D1 is plotted for all three complexes examined with the pYEpYI-SH2 domain complex in blue, the pYEpYV-SH2 domain complex in red, and pYIpYV-SH2 domain complex in green. The distances between the atoms are computed for each molecular dynamics snapshot, smoothed with a moving mean (±50 ps) and plotted as a function of frame number. (C) D2 is plotted for all three complexes with the same color coding than B.

phorylated peptides has been investigated. If, in general terms, the release of bound water to the bulk is believed to be entropically favorable,[27] on the other hand, recruitment of water to a binding interface may be enthalpically favorable.[28–30] Overall, although there is a fine balance between the enthalpic and entropic contributions, binding of water molecules through formation of several hydrogen-bonds is believed to be energetically favorable,[29,30] if these water molecules are placed such that they extend the binding pocket by forming several hydrogen bonds with the ligand and the protein at the same time. Although, these rules must be used with great care, it is clear that whatever the impact water molecules make on binding, they are key players in the thermodynamic signature of Src SH2 domain/phosphopeptide complexes.

Our results provide a detailed, quantitative, thermodynamic map of all the binding sites of water at the protein–ligand interface. We also showed that a gain in solvation energy is observed in the region of the $\Delta\Delta W$ isosurface located on the +2-pY and that the +2-pY side chain in the pYIpYV peptide-bound ligand appears to be more mobile and to assume a different conformation compared to the pYEpYI peptide-bound ligand. Overall, the double mutation introduced in the pYIpYV peptide at position +1 and +3 compared to the pYEpYI peptide results in two indirect events around the +2-pY. It is thus remarkable to see that these mutations do not just produce local effects, that is ones that occur close to the mutated residues, but combine to affect more distant regions of the binding interface.

We suggest here that the region around the +2-pY therefore plays a central role in the entropic contribution to the free energy changes shown in Table I. By decomposing the double E to I and I to V mutation into single mutations, we gain a better understanding of the cause of the progressively better solvation around the +2 pY in going from pYEpYI to pYEpYV to pYIpYV. Indeed, we observe that this effect is not entirely due to one residue but rather to the combination of two single mutations, albeit with different magnitudes (+1 position being the dominant one). The favorable enthalpic difference from pYIpYV to pYEpYI (−4.7 Kcal/mol from Table I) is likely due to the binding of water molecules at the interface between the glutamic acid and the SH2 domain with a possibly much lower contribution from the V to I mutation at the +3 position. The high entropic loss (−4.5 kcal/mol) is partially due to the bound water at +1 but more importantly to a secondary event at +2-pY with a change in mobility and conformation of the phosphotyrosine ring producing a significantly reduced water accessible volume at the ligand–protein interface.

In conclusion, the PMF protocol described in this work, in combination with MD simulations, provides an effective means of analyzing the thermodynamic signature of protein–ligand interactions on a qualitative basis and useful insights into the molecular mechanism of

binding. However, it has the obvious limitation that it does not provide a quantitative answer in terms of the thermodynamics of protein–ligand binding. In the future, however, the same protocol may lead to more quantitative results for binding affinities. In fact, the "$\Delta\Delta$" free energy isosurfaces $\Delta\Delta W$ computed here provide additional information on the shape and depth of the free energy wells of the water binding sites which are of interest in estimating, at least approximately, the relative enthalpic/entropic contributions of solvation in a more quantitative manner.

ACKNOWLEDGMENTS

REFERENCES

1. Kuriyan J, Cowburn D. Modular peptide recognition domains in eukaryotic signaling. Annu Rev Biophys Biomol Struct 1997;26: 259–288.

2. Pawson T. Organization of cell-regulatory systems through modular-protein-interaction domains. Philos Transact A Math Phys Eng Sci 2003;361:1251–1262.

3. Waksman G, Kumaran S, Lubman O. SH2 domains: role, structure and implications for molecular medicine. Expert Rev Mol Med 2004;6:1–18.

4. Songyang Z, Gish G, Mbamalu G, Pawson T, Cantley LC. A single point mutation switches the specificity of group III Src homology (SH) 2 domains to that of group I SH2 domains. J Biol Chem 1995;270:26029–26032.

5. Songyang Z, Shoelson SE, McGlade J, Olivier P, Pawson T, Bustelo XR, Barbacid M, Sabe H, Hanafusa H, Yi T, Ren R, Baltimore D, Ratnofsky S, Feldman RA, Cantley LC. Specific motifs recognized by the SH2 domains of Csk3BP2fps/fes. GRB-2, HCP, SHC, Syk, and Vav. Mol Cell Biol 1994;14:2777–2785.

6. Songyang Z, Shoelson SE, Chaudhuri M, Gish G, Pawson T, Haser WG, King F, Roberts T, Ratnofski S, Lechleider RJ, Neel B, Birge RB, Fajardo JE, Chou MM, Hanafusa H, Schaffhausen B, Cantley LC. SH2 domains recognize specific phosphopeptide sequences. Cell 1993;72:767–778.

7. Bradshaw JM, Waksman G. Calorimetric examination of high-affinity Src SH2 domain-tyrosyl phosphopeptide binding: dissection of the phosphopeptide sequence specificity and coupling energetics. Biochemistry 1999;38:5147–5154.

8. Lubman OY, Waksman G. Dissection of the energetic coupling across the Src SH2 domain-tyrosyl phosphopeptide interface. J Mol Biol 2002;316:291–304.

9. Lubman OY, Waksman G. Structural and thermodynamic basis for the interaction of the Src SH2 domain with the activated form of the PDGF β-receptor. J Mol Biol 2003;328:655–668.

10. Fowler PW, Jha S, Coveney PV. Grid-based steered thermodynamic integration accelerates the calculation of binding free energies. Philos Trans A Math Phys Eng Sci 2005;363:1999–2015.

11. Geroult S, Virdee S, Waksman G. The role of water in computational and experimental derivation of binding thermodynamics in SH2 domains. Chem Biol Drug Des 2006;67:38–45.

12. Geroult S, Hooda M, Virdee S, Waksman G. Prediction of solvation sites at the interface of Src SH2 domain complexes using molecular dynamics simulations. Chem Biol Drug Des 2007;70:87–99.

13. Fowler PW, Geroult S, Jha S, Waksman G, Coverney PV. Rapid accurate, and precise calculation of relative binding affinities for the SH2 domain using a computational grid. J Chem Theory Comput 2007;3:1193–1202.

14. Baker BM, Murphy KP. Prediction of binding energetics from structure using empirical parameterization. Methods Enzymol 1998;295:294–315.

15. De Fabritiis G, Villa-Freixa J, Coveney PV. Multiple scale modeling of permeation through membrane channels using pregenerated molecular dynamics trajectories. Int J Mod Phys 2007;18:511.

16. Choen J, Arkhipov A, Braun R, Schulten K. Imaging the migration pathways for O2, CO, NO, and Xe inside myoglobin. Biophys J 2006;91:1844–1857.

17. Frenkel D, Smit B. Understanding molecular simulations. San Diego: Academic Press; 1996.

18. Jensen M, Park S, Tajkhorshid E, Schulten K. Energetics of glycerol conduction through aquaglyceroporin GlpF. Proc Natl Acad Sci USA 2002;99:6731–6736.

19. Blundell TL. Structure-based drug design. Nature 1996;384:23–26.

20. Verlinde CLMJ, Hol WGJ. Structure-based drug design: progress, results and challenges. Structure 1994;2:577–587.

21. Kirkwood JG. Statistical mechanics of fluid mixtures. J Chem Phys 1935;3:300–313.

22. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.

23. MacKerell AD, Jr, Bashford D, Bellott M, Dunbrack RL, Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.

24. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Laxmikant K, Schulten K. Scalable molecular dynamics with NAMD. J Comp Chem 2005;26:1781–1802.

25. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:27–38.

26. Baker NA, Sept D, Joseph D, Holst MJ, McGammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001;98:10037–10041.

27. Dunitz J. The entropic cost of bound water in crystals and biomolecules. Science 1994;264:670–674.

28. Clarke C, Woods RJ, Gluska J, Cooper A, Nutley MA, Boons G-J. Involvement of water in carbohydrate-protein binding. J Am Chem Soc 2001;123:12238–12247.

29. Ladbury JE. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. Chem Biol 1996;3:973–980.

30. Ladbury JE, Williams MA. The extended interface: measuring nonlocal effects in biomolecular interactions. Curr Opin Struct Biol 2004;14:562–569.