How and why phosphotyrosine-containing peptides bind to the SH2 and PTB domains

Yingyao Zhou and Ruben Abagyan

Background: Specific recognition of phosphotyrosine-containing protein segments by Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domains plays an important role in intracellular signal transduction. Although many SH2/PTB-domain-containing receptor–peptide complex structures have been solved, little has been done to study the problem computationally. Prediction of the binding geometry and the binding constant of any peptide–protein pair is an extremely important problem.

Results: A procedure to predict binding energies of phosphotyrosine-containing peptides with SH2/PTB domains was developed. The average deviation between experimentally measured binding energies and theoretical evaluations was 1.8 kcal/mol. Binding states of unphosphorylated peptides were also predicted reasonably well. *Ab initio* predictions of binding geometry of fully flexible peptides correctly identified conformations of two pentapeptides and a hexapeptide complexed with a v-Src SH2 domain receptor with root mean square deviations (rmsds) of 0.3 Å, 1.2 Å and 1.5 Å, respectively.

Conclusions: The binding energies of phosphotyrosine-containing complexes can be effectively predicted using the procedure developed here. It was also possible to predict the bound conformations of flexible short peptides correctly from random starting conformations.

Address: Skirball Institute of Biomolecular Medicine, Structural Biology, New York University Medical Center, 540 1st Avenue, New York, NY 10016, USA.

Correspondence: Ruben Abagyan
E-mail: abagyan@earth.med.nyu.edu

Key words: binding energy, protein docking, signal transduction, structure prediction, tyrosine phosphorylation

Received: **18 August 1998**
Revisions requested: **07 October 1998**
Revisions received: **19 October 1998**
Accepted: **03 November 1998**

Published: **11 December 1998**
http://biomednet.com/elecref/1359027800300513

Folding & Design 11 December 1998, **3**:513–522

Introduction

Specific recognition of phosphotyrosine-containing protein segments by Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domains (also known as phosphotyrosine interaction, PI, domains) plays an important role in intracellular signal transduction. This type of interaction leads to a broad range of cellular processes, such as growth, proliferation, differentiation, mitogenesis and transformation [1–3]. SH2-domain-containing proteins can either modulate enzymatic activities of downstream signal-transduction events or act as adaptor molecules in protein–protein interaction. PTB domains serve as alternative binding domains and recruit additional signaling proteins to the vicinity of an activated receptor [4–6]. SH2 and PTB domains have attracted considerable interest in recent years; more than 20 structures of peptide–receptor complexes have been published since the first structure by Waksman *et al.* [7].

The key property of the SH2-modulation mechanism is that phosphorylation/dephosphorylation of a specific tyrosine residue of the peptide ligand acts as an on/off switch [1]. It was experimentally demonstrated that phosphorylation plays a critical role in association of an SH2–EGFR complex (epidermal growth factor receptor) [8], an SH2–FAK complex (focal adhesion kinase) [8]; a p85α SH2-containing complex [9,10], a Cγ1 Syk SH2-containing complex [11], a Syk C-terminal SH2-containing complex [12], $p56^{lck}$ SH2 containing complexes [13–15], a Src Hck–quercetin complex [16], as well as other complexes [17,18]. The recently discovered PTB domains do not always require phosphorylation for binding [1,4], but phosphorylation usually increases the binding affinity. For the Shc adaptor protein the increase was estimated as at least 7.5 kcal/mol [19]; for the insulin receptor substrate-1 (IRS-1) specific peptide, the increase was estimated to be at least 4.1 kcal/mol (a factor of 1000, Zhou *et al.* [20]; MM Zhou, personal communication). The X11 PTB domain, however, binds both phosphorylated and unphosphorylated peptides rather strongly with K_ds of 8.3 μM and 4.6 μM, respectively, and phosphorylation decreases binding by 0.4 kcal/mol [21].

These complex behaviors require a quantitative explanation. Little has been done, however, to study the problem computationally. The computational analysis is complicated by the magnitude of conformational rearrangement of the binding proteins, lack of adequate description of the unbound states and the inaccuracies of the force-field parameters and partial charges. Two properties of these complexes, however, make theoretical investigation feasible. First, it was experimentally demonstrated that short peptides from the binding segments could mimic specific complexes with the SH2 domains [11,15]. Secondly, it has been shown that SH2 domains do not undergo large conformational changes upon association [7,22,23].

In this paper we address two computational problems: prediction of the binding energy of both phosphorylated and unphosphorylated peptides given the geometry of the complex, and prediction of the association geometry of a fully flexible peptide to its receptor.

Theoretical binding energy evaluation is a complex and, as yet, unsolved problem [24–26]. Some methods use sophisticated and computationally costly approaches, such as molecular dynamics (MD) [27], free energy perturbation (FEP) and thermodynamic integration (TI) [28], the protein dipoles Langevin dipoles (PDLD) model and linear-response approximation (LRA) [29]. Other methods use simplified binding functions that attempt to capture the major contributions to the binding energy [24,30,31], such as polar and apolar contributions [32], entropy loss of the ligand [33,34] and surface complementarity [35]. In general, fast and simple evaluations of binding energy are essential for a large scale screening of compound databanks [36]. The evaluation scheme used in this work was similar to the one used by Novotny and coworkers [33,34,37] with the exception of two modifications. First, we use a rigorous boundary-element evaluation of the electrostatic components of the interaction. Second, we took advantage of a rigorous sidechain-refinement procedure using optimal-bias Monte Carlo minimization (OBMCM, also known as biased probability Monte Carlo [38]), that improves the accuracy of the electrostatic contribution (see the Results and discussion section).

Flexible docking of peptides to their protein receptors remains an unsolved problem [39–41]. Even a single peptide in isolation has too many degrees of freedom for a reliable prediction. To avoid this difficulty, ligands are usually considered to be rigid or are constructed incrementally from rigid parts [30,42]. The OBMCM algorithm [38] for efficient global optimization of a peptide was shown to predict the conformation of isolated peptides [43]. Here we demonstrate that this algorithm can be combined with pseudo-Brownian random moves [44] to allow efficient sampling of both the energy landscape of a continuously flexible peptide, as well as peptide interaction with its receptor represented by a set of grid of potentials [45,46].

In this paper we describe the binding energy evaluation of 12 phosphopeptide–receptor complexes. Comparison of the results between unrefined and refined complex structures suggests the sidechain-refinement procedure is important in the binding affinity evaluation. The predicted energies for phosphorylated and unphosphorylated peptides explain most of the experimental measurements and suggest that similar calculations can be used to evaluate the binding and role of phosphorylation in uncharacterized peptide–receptor complexes. Furthermore, the bound conformation of three short peptides was predicted with reasonable accuracy.

Results and discussion

Binding energy function

Models of 12 selected phosphorylated complexes (Table 1) were built using the regularization procedure followed by the sidechain-refinement procedure (see the Materials and methods section). The sidechain-refinement procedure resolves clashes after hydrogen atoms have been added to a structure and corrects geometrical inaccuracies and ambiguities. For example, some interface residues in the $p56^{lck}$ complex (PDB code 1LKK) have dual sidechain conformations [47]; the peptide ends in a Cγ1 Syk complex (2PLD) are undefined [48]; part of the ligand-binding loop in the c-Src complex 1SHD is disordered [8]; and backbone peptide bonds deviate significantly from both *cis* and *trans* conformations in the Shc-adaptor complex 1TCE [49]. The refinement procedure is expected to eliminate clashes, thus optimizing electrostatic contacts between the two interacting domains [50].

We applied the following empirical binding energy function (see the Materials and methods section) to the refined flexible peptide–receptor complexes in Table 1:

$$\Delta G_{calc} = \Delta G_{BEP}(\varepsilon) + s\Delta A - T\Delta S + C \qquad (1)$$

Where $\Delta G_{BEP}(\varepsilon)$ is the electrostatic free energy change, ε is the dielectric constant, s is the surface tension, ΔA is the change of solvent accessible area, T is room temperature, ΔS is the entropy term evaluated via the number of ligand free backbone torsions and the relative accessibility change of the interface sidechains (see the Materials and methods section), and C is a constant. The dielectric constant ε, surface tension s, and the constant C can be adjusted to partially absorb other missing contributions.

For various combinations of ε and C, the optimal surface-tension constants s were determined by minimizing the average deviation between experimental measurements and theoretical predictions, $\langle|\Delta G_{calc} - \Delta G_{exp}|\rangle$, on the set of phosphorylated complexes as discussed in the next section. During the above minimization process, ε values above 4 and C values ranging from 0–30 kcal/mol were examined. The best combination was found to be $\varepsilon = 24$, $s = 19.0$ cal/mol/Å^2, and $C = 5.0$ kcal/mol.

The experimental binding data have considerable inaccuracies: about 1.1 kcal/mol for a $p56^{lck}$ complex (1LCK) [14], 1.0 kcal/mol for the Cγ1 Syk complex [48], and possibly even larger for another $p56^{lck}$ complex (1LCJ) [13] (Table 1). Furthermore, the binding data might also be sensitive to solution conditions, such as pH and salt concentration [51]. We assumed 1.0 kcal/mol to be the experimental error, and found that many combinations of the parameters with $\varepsilon \geq 9$, $C \leq 14$ kcal/mol, and $14.6 \leq s \leq 27.9$ cal/mol/Å^2 can all explain the available data accurately within 1 kcal/mol. The dependence of the performance of

Table 1

Protein–phosphopeptide complexes and their experimentally determined binding energies.

Complex (PDB code)	Receptor	Ligand sequence	K_d (μM)	ΔG_{exp}* (kcal/mol)
1IRS	Insulin receptor substrate-1 PTB domain	LVIAGNPApYRS	6 [20]	−7.17
1LCJ	p56^{lck} tyrosine kinase SH2 domain	EPQpYEEIPIYL	0.001 [13], 0.0035–0.0063†	−7.93‡
1LCK	p56^{lck} tyrosine kinase SH2 domain	EGQpYQPQPA	2.92§, 20†	−7.02#
1LKK	p56^{lck} tyrosine kinase SH2 domain	Ac-pYEEI	0.14¶	−9.41
1LKL	p56^{lck} tyrosine kinase SH2 domain	Ac-pYEEG	1.54¶	−7.98
1PIC	Phosphatidylinositol 3-kinase P85α SH2 domain	Ac-pYVPM	0.058 [10]	−9.93
2PLD	Phospholipase Cγ1 Syk SH2 domain	DNDpYIIPLPDPK	0.1-0.5 [48]	−9.13#
1SHA	v-Src tyrosine kinase SH2 domain	pYVPML	5.9 [23]	−7.11
1SHC	Shc adaptor protein PTB domain	HIIENPQpYFSDA	0.053 [19,56]	−9.99
1SHD	c-Src tyrosine kinase SH2 domain	Ac-pYEEI	~0.1 [8]	~−9.61
1SPS	v-Src tyrosine kinase SH2 domain	EPQpYEEIPIYL	0.003-0.006 [57], 0.2 [23]	−9.11¥
1TCE	Shc adaptor protein SH2 domain	GHDGLpYQGLSTATK	50 [49]	−5.90
Complex (PDB code)	**Receptor**	**Ligand sequence**	**Relative binding affinity**	**$\Delta\Delta G_{exp}$ (kcal/mol)**
1AYA	Tyrosine phosphatase Syp SH2 domain	SVLpYTAVQPNE	0.08**	0.55**
1AYB	Tyrosine phosphatase Syp SH2 domain	SPGEpYVNIEF	0.2**	0.0**
1AYC	Tyrosine phosphatase Syp SH2 domain	DGGpYMDMSKDE	< 0.003**	> 2.50**

*ΔG_{exp} is calculated from K_d value using $\Delta G_{exp} = RT\ln K_d$, where R is the gas constant and T is room temperature. †From Table 2 in [15]; K_ds were measured using surface plasmon resonance (SPR). ‡The binding energy was derived from measurements for 1LCK [14], as well as from the ID_{50} values determined for these two ligands (Table 1 in [15]). The data for 1LCK were determined using isothermal titration calorimetry (ITC), which directly measures the heat of reaction associated with binding [1,69]. ITC avoids the artifact that can arise when the target peptides are immobilized in the SPR method, which can lead to higher determined affinity. ID_{50} values used for 1LCJ and 1LCK were1.8 μM and 8.2 μM, respectively. §From Table 1 in [17]; K_d was measured using the ITC method. #The average value of the measurements was used. ¶K_d for 1LKL was derived from 1LKK, because its binding was 11-fold weaker [47]. ¥K_d value determined in [23] was used as suggested by Waksman *et al.* [23]. **From Figure 1 in [22]. The binding energy was compiled from data presented by Case *et al.* [70]. 1AYB was used as the reference ligand, whose $\Delta\Delta G_{exp}$ is defined to be 0 kcal/mol.

binding prediction to the binding parameters is also discussed in a separate work (M Schapira and R.A., unpublished observations).

Binding energy calculation for phosphorylated complexes

Binding energies for phosphorylated complexes calculated using the above optimal parameters and the corresponding experimental data are listed in Table 2 and plotted in Figure 1 (filled circles) for comparison. The average deviation between the two data sets is 1.8 kcal/mol with a high correlation coefficient of 0.87. Underestimation of the binding affinity in an Shc SH2-domain-containing model (1TCE) might be attributable to the problematic ligand backbone coordinates [49].

Evaluation of binding energies on initially regularized but unrefined models resulted in a lower correlation coefficient of 0.71 (open circles in Figure 1), even when the three parameters were readjusted using the unrefined models themselves. On average, the electrostatic contribution to the binding energies of the studied complexes was modified by 2.3 kcal/mol after the sidechain refinement, which mainly led to the improved fit.

We analyzed the effect of several alternative forms of the binding energy terms on the prediction accuracy. After each modification the parameters were re-optimized to maximize the correlation. When the entropy contribution to binding was calculated as N_{torsion}*0.6 kcal/mol, the correlation dropped from 0.87 to 0.74. When the molecular surface area [52] (the contact area of the probe sphere) was used instead of the solvent-accessible area (the center of the probe sphere) to calculate the surface contribution, as recommended by Tunon *et al.* [53] and Novotny *et al.* [34], the optimal correlation remained unchanged. The surface-based atomic solvation parameter method [43,54] combined with the Coulomb evaluation of the intermolecular interactions slightly reduced the correlation coefficient to 0.82. The current set is too small to make definite conclusions about the optimal choice of terms. A more systematic analysis of binding energy terms, based on a larger set of complexes, will be published elsewhere (M Schapira and R.A., unpublished observations).

Binding energy prediction for unphosphorylated complexes

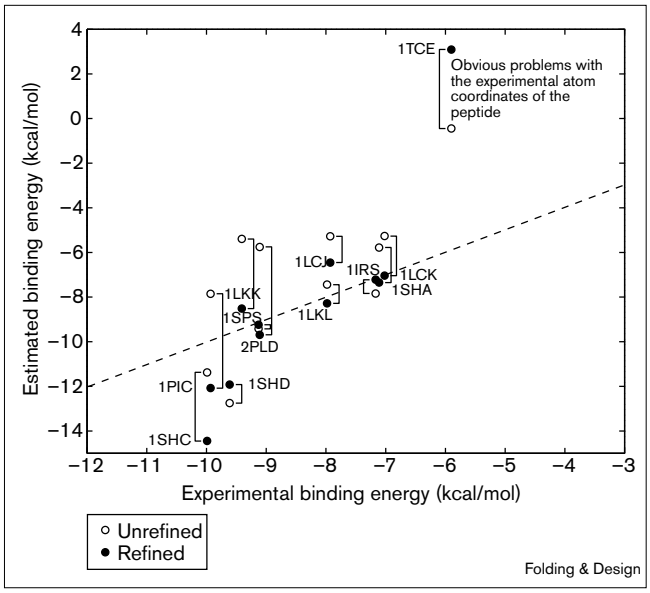
Is the accuracy of the above algorithm for binding energy evaluation sufficient to predict whether a given peptide will stay bound or disassociate after dephosphorylation? To answer this question, we performed a series of calculations on unphosphorylated complexes. Because structural information for the unphosphorylated complexes is generally not available, their initial models were derived from

Table 2

Theoretical estimations of binding energies of the selected phosphorylated complex set.

Complex (PDB code)	ΔG_{el} (kcal/mol)	$s\Delta A$ (kcal/mol)	$-T\Delta S$ (kcal/mol)	ΔG_{calc} (kcal/mol)	ΔG_{exp} (kcal/mol)
1IRS	−4.37	−23.85	15.99	−7.22	−7.17
1LCJ	−4.16	−22.98	15.69	−6.45	−7.93
1LCK	−6.26	−19.47	13.68	−7.04	−7.02
1LKK	−5.55	−16.03	8.07	−8.51	−9.41
1LKL	−6.91	−15.31	8.94	−8.28	−7.98
1PIC	−4.78	−21.70	9.40	−12.08	−9.93
2PLD	−4.43	−25.93	16.12	−9.24	−7.11
1SHA	−4.70	−16.30	8.65	−7.35	−9.99
1SHC	−6.19	−33.69	20.43	−14.45	−9.99
1SHD	−9.88	−16.02	8.98	−11.92	−9.61
1SPS	−10.93	−21.24	17.48	−9.69	−9.11
1TCE	1.92	−23.46	19.64	3.09	−5.90
Complex*	$\Delta\Delta G_{el}$ (kcal/mol)	$s\Delta\Delta A$ (kcal/mol)	$-T\Delta\Delta S$ (kcal/mol)	$\Delta\Delta G_{calc}$ (kcal/mol)	$\Delta\Delta G_{exp}$ (kcal/mol)
1AYA	3.84	1.91	−1.85	3.90	0.55
1AYC	5.24	8.65	−1.57	12.32	> 2.50

Figure 1



Comparison of theoretical predictions with experimental measurements using the models before (open circles) and after (filled circles) the interface sidechain-refinement procedure.

the corresponding PDB entries by omitting the phosphate groups. The initial models were then refined using the same procedure as for the phosphorylated proteins.

The predictions for the unphosphorylated ligands using the previous optimal binding function are shown in Figure 2. For those SH2-domain-containing complexes whose experimental binding data are not available, it is generally believed that they do not bind or bind only weakly. For two $p56^{lck}$ complexes (1LCJ [13] and 1LCK [14]), a Shc SH2 domain-containing complex (1TCE [49]), a Grb SH2 domain-containing complex (1TZE [55]), two v-Src complexes (1SHA and 1SHB [7]) and two Syp complexes (1AYA and 1AYC [22]), the unphosphorylated peptides were predicted to be in the free form. Other SH2-domain-containing complexes were predicted to be weakly bound. (A 50 μM dissociation value was used for distinguishing weakly bound from bound states.) The affinities of the unphosphorylated complexes were correctly predicted to be much weaker than the phosphorylated ones, with an average gap of 6.6 kcal/mol.

The binding energy differences are more variable for the PTB domain-containing complexes and these variations were correctly captured using our method. The method correctly predicted a 6 kcal/mol increase of the 11-residue peptide to the IRS-1 PTB domain [20] and the 13-residue peptide to the Shc PTB domain [56] upon phosphorylation. In contrast, the evaluation suggested only 0.1 kcal/mol difference between the phosphorylated and unphosphorylated models in the X11 PTB domain-containing complex (1AQC [21]), which agrees well with the 0.4 kcal/mol experimental measurement.

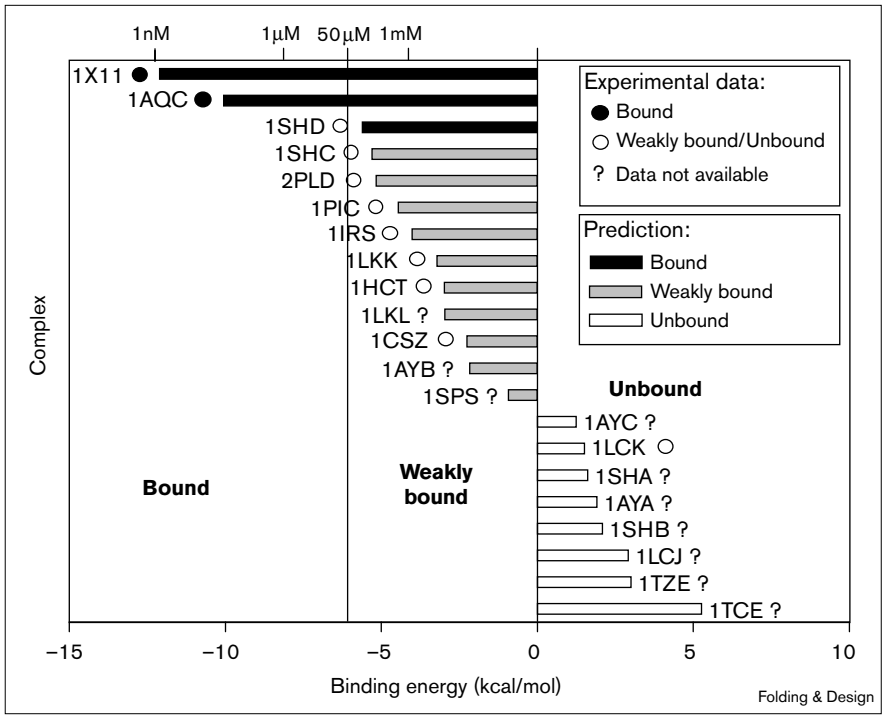
These results are encouraging as all the unphosphorylated binding energies were evaluated using parameters that were derived purely on the basis of the phosphorylated structures listed in Table 1, without prior knowledge of any binding energies on unphosphorylated peptides. The fact that on average the electrostatic contribution accounts for 5.5 kcal/mol out of the 6.6 kcal/mol (83%) binding energy enhancement after phosphorylation agrees with our qualitative understanding. Eight complexes, however, were predicted to have binding energy lower than –2 kcal/mol, which is the error of the empirical binding function. These numbers indicate that the binding energy for the unphosphorylated models can be overestimated using this method.

Ab initio ligand docking simulations

As individual receptor structures and complex structures are both available for certain targets in our complex set, we also made an attempt to predict not only the binding energies but also the ligand conformation *ab initio* given only minimal information about the approximate location of the receptor-binding site. The candidates for our docking simulations were three v-Src complexes (1SHA, 1SHB [7] and 1SPS [57]) and three Syp complexes (1AYA, 1AYB and 1AYC [22]). These were the only structures of the SH2 family that had both isolated SH2 domain structures and complex structures deposited in the Brookhaven PDB. Simulations were performed for the six structures and the 1IRS complex [20] model (Table 3), where the receptor structures were taken from the refined models described

Figure 2

Binding energy predictions of unphosphorylated complexes. The proteins 1TCE [49], 1TZE [55], 1LCJ [13], 1SHB [7], 1AYA [22], 1SHA [7], 1LCK [14] and 1AYC [22] were predicted to be in the unbound form, and 1X11 and 1AQC [21] were correctly predicted to be strongly bound (the experimentally measured binding energies are −8.87 kcal/mol for 1X11 and −6.98 kcal/mol for 1AQC). Other complexes were predicted to be weakly bound, if a dissociation constant of 50 μM was used as the boundary between weakly bound complexes and the more strongly bound ones.



before. Due to the fact that the C terminus of the 1SPS ligand interacts with another receptor domain (which complicates the calculation and interpretation), a shorter peptide Ac-PQpYEEI-NH_2 was used in the docking simulation of 1SPS as suggested by one of our referees.

The prediction was performed in two major steps as recommended previously [46,58] (see the Materials and methods section). The first step involves docking of fully flexible, full atom model of the peptide into the set of receptor grid potentials. The grid potentials are precalculated potential fields generated by the rigid receptor, and enabled us to dramatically improve the computational efficiency of the interaction energy [45,46]. A combination of OBMCM [38] and pseudo-Brownian moves was used to sample the conformational space. The first step resulted in a set of about 40 low-energy conformations. In the second step, all of the low-energy solutions were refined with the explicit receptor using the OBMCM interface sidechain-refinement procedure (see the Materials and methods section) and sorted according to the final energy values.

In a 1SHA model simulation, the lowest energy conformation found by the grid potential folding simulation was already very close to the crystallographic conformation (rmsd ~1.7 Å). After the global OBMCM refinement of the 40 low-energy conformations, the near-native solution remained the most energetically favoured. Furthermore, the backbone rmsd from the crystallographic structure dropped to the even lower value of only 0.3 Å after the refinement. The 1SHB model simulation illustrated the power and the importance of the OBMCM sidechain-refinement step. After the initial grid docking, two near-native conformations with rmsd less than 2 Å were ranked 11th and 30th, respectively. The lowest energy conformation had an rmsd of 4.5 Å, and solution 24 had an rmsd of 2.9 Å. The OBMCM refinement of all the solutions resulted in the rank 24 solution being improved, however,

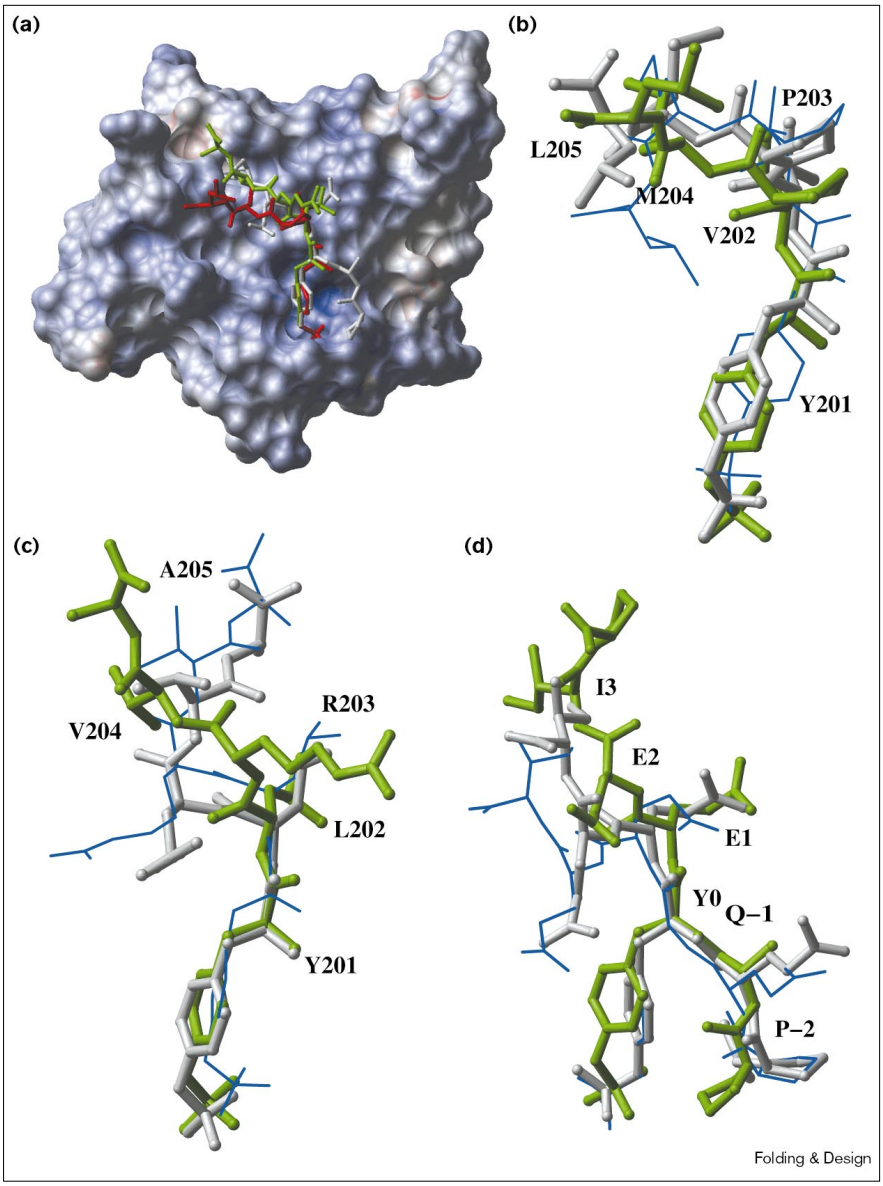
Table 3

Results of docking of six peptides to their receptors taken from the complex models.

Complex (PDB code)	Peptide length	E_{min} (kcal/mol)	$rmsd_{min}$ (Å)	$E_{closest}$ (kcal/mol)	$rmsd_{closest}$ (Å)
1SHA*	5	−1231	0.3	−1231	0.3
1SHB*	5	−1347	1.2	−1347	1.2
1SPS*	6	−1263	1.5	−1253	1.2
1IRS	11	−1358	3.9	−1317	3.1
1AYA	11	−1240	21.1	−1236	6.1
1AYB	12	−1222	15.9	−1202	13.3
1AYC	11	−1349	7.0	−1296	3.3

*The pentapeptide conformation could be predicted without a tether being attached to the phosphorus atom in the phosphate group. E_{min} is the energy of the lowest conformation found by the docking simulation, and $rmsd_{min}$ is the rmsd between the lowest energy ligand backbone conformation and its native conformation. $E_{closest}$ is the energy of the closest-to-native conformation found, and $rmsd_{closest}$ is the rmsd between this conformation and the native conformation.

Figure 3



(a) Superposition of three crystal structures of the Src SH2 domain with different peptides (1SHA, 1SHB and 1SPS). The molecular surface of SH2 domain is colored by electrostatic potentials calculated with REBEL algorithm [46]. The peptide ligands in complexes 1SHA, 1SHB and 1SPS are red, green and white, respectively. **(b)** *Ab initio* prediction of ligand conformation for 1SHA [7]. **(c)** *Ab initio* prediction of ligand conformation for 1SHB [7]. **(d)** *Ab initio* prediction of ligand conformation for 1SPS [57]. The green sticks present the experimentally determined ligand structures, the blue wires represent the ligand conformations identified by the *ab initio* folding within the grid potentials, which was later refined in the ECEPP/3 [64–66] forcefield to produce the final conformations (the white sticks).

to the most energetically favored conformation, with a backbone rmsd of 1.2 Å to the crystallographic structure.

A similar result was obtained for the high-affinity hexapeptide (1SPS entry) simulation. The initial grid docking found a conformation with an rmsd of 1.0 as rank seven. The lowest energy conformation had an rmsd of 9.1 Å, and the rank six conformation had an rmsd of 2.4 Å. The OBMCM refinement of all the retained conformations brought the rank six solution to the best energy conformation with a backbone rmsd of 1.5 Å. The rank seven conformation was moved to rank four with only a slightly increased rmsd of 1.2 Å (compared to the initial 1.0 Å). The crystal structures of the complexes and the predicted ligand conformations of 1SHA, 1SHB, and 1SPS are shown in Figure 3.

We also applied the same procedure to longer peptides (around 11 residues) in the three Syk complexes and the IRS-1 complex. These peptides have twice the number of free variables and we could not find sampling conditions that led to the correct predictions. After a tether was set between a single phosphorus atom of the phosphate group and the corresponding location in a binding pocket, however, the sampling space was reduced. The simulations found the best near-native conformations with rmsd about 3 Å for the complex 1IRS and 1AYC, although these conformations were energetically too high to be recognized

Table 4

Results of restrained peptide docking to the peptide-free SH2 domains.

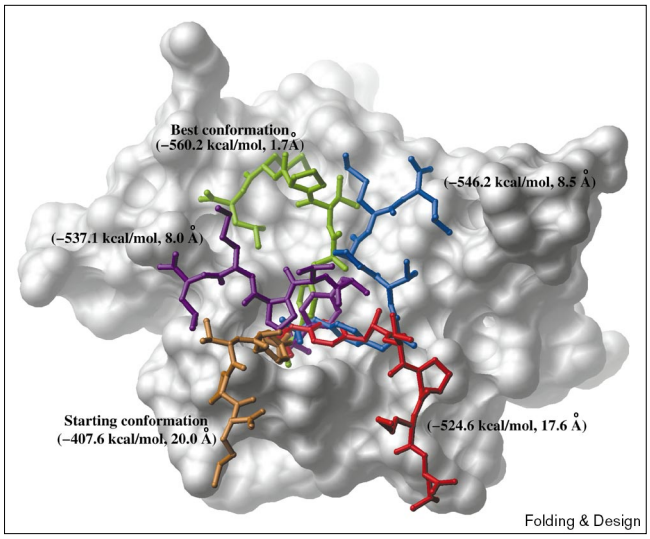
Complex	Peptide length	E_{min} (kcal/mol)	$rmsd_{min}$ (Å)	$E_{closest}$ (kcal/mol)	$rmsd_{closest}$ (Å)
1SHA	5	−1197	2.0	−1164	1.4
1SHB	5	−1311	1.3	−1311	1.3
1SPS	6	−1172	1.2	−1156	1.1
1AYA	11	−1260	18.9	−1230	13.9
1AYB	12	−1264	12.2	−1260	11.6
1AYC	11	−1336	20.8	−1291	14.7

by the prediction. It should be noted that a single tether to the phosphate atom is a biologically sensible restraint and still leaves a lot of room for backbone rearrangements — for example, backbone rmsd values up to 17 Å were found in the simulation stack.

Docking a peptide into an uncomplexed receptor structure is a difficult task because of the conformational changes that are induced upon complex formation. We found that tethering only the phosphate group to its binding pocket was sufficient for successful docking of the rest of the peptides into experimentally solved uncomplexed SH2 structures (1SPR receptor for peptides from 1SHA, 1SHB [7] and 1SPS [57]; and 1AYD receptor for peptides from 1AYA, 1AYB and 1AYC [22]). It appeared that the correct conformation of the short peptides was not found primarily because of insufficient sampling, rather than higher interaction energy with the deformed receptor. After a single tether was imposed on the phosphorus atom and the binding pocket, the correct folding of the rest of a peptide in the grid potentials occurred rather quickly (Table 4). The peptides from the 1SHA complex, 1SHB complex, and the 1SPS complex folded into the expected conformations (the backbone rmsd was around 2.0 Å) in the field of the isolated SH2 domain from the 1SPR entry. Docking of the long peptides into the isolated receptor structure (1AYD) resulted in larger deviations than using the receptor structures of the refined complexes.

The *ab initio* docking simulation of the three short peptides performed remarkably well, given the giant size of the conformational space and the lack of single clear binding groove at the receptor. This work supports the previous finding that the interface sidechain refinement significantly improves the recognition of a correct conformation among false positives [46,58]. A realistic ligand simulation with the free receptor conformation 1SPR was more challenging but was still successfully carried out with the aid of a single tether imposed between the phosphate group and its binding site. The difficulties of simulating docking for the long peptides were a result of both insufficient conformational sampling and an inaccurate force field, as the near-native conformations sometimes had lower energy than the lowest-energy solutions found by grid sampling and sometimes had higher energy than those of other refined solutions.

Figure 4



Snapshots of an *ab initio* folding simulation of the 1SHA [7] pentapeptide ligand using grid potentials. The receptor SH2 domain is shown as a molecular surface, and five representative ligand conformations sampled by the simulation are drawn as sticks. The parameters in the parentheses show both the energy and the rmsd of the corresponding conformations. The best conformation, with an rmsd of 1.7 Å, was also the lowest energy conformation in the simulation stack.

Unlike any previous docking simulation (e.g. [59]), here we considered fully flexible peptide ligand models and completely random starting peptide conformations with rmsd values up to 20 Å. Figure 4 shows some snapshots of the 1SHA pentapeptide simulation and illustrates a wide sampling range of the procedure. Quicktime movies of 1SHA, 1SHB, and 1SPS docking simulations are available at the following URL: http://saturn.med.nyu.edu/groups/AbagyanLab/sh2/.

Conclusions

In this work, an empirical binding energy that considers accurate electrostatic contributions, hydrophobic effects and configurational entropy changes has been applied to analyze the association of SH2/PTB domains with their specific phosphotyrosyl peptide ligands. The experimentally measured binding energies of phosphopeptides were accurately reproduced, within 1.8 kcal/mol (Figure 1), and the complex pattern of the association states of unphosphorylated peptide ligands was predicted correctly in most cases (Figure 2). Even such a subtle phenomenon as different magnitudes of binding enhancement upon phosphorylation in different complexes can be predicted reasonably well using the proposed algorithm.

The native binding geometries of some protein–peptide complexes were correctly predicted *ab initio* from random initial conformations, that had rmsd values from the correct position as large as 20 Å and their internal geometry fully randomized (Figure 4)! It was demonstrated that a recently developed two-step energy optimization procedure, consisting of grid potential sampling followed by interface refinement, could predict the binding modes of two pentapeptides and a hexapeptide with 1 Å accuracy (Figure 3).

Predictions using longer peptides and/or the uncomplexed conformations of the receptors proved to be more difficult, but a minimal experimental restraint (such as the location of the phosphate group) was sufficient to make a near-correct prediction. This algorithm could be useful for predicting the geometry and energetics of binding for new, as yet uncharacterized interactions between protein domains involved in signal transduction.

Materials and methods

Peptide–receptor complexes studied

In total, 21 structures were selected for this study. They are X11 PTB domain-containing complexes (1AQC and 1X11 [21]), Syp SH2 domain-containing complexes (1AYA, 1AYB, and 1AYC [22]), a Syk C-terminal SH2 domain-containing complex (1CSZ [12]), Src SH2 domain-containing complexes (1HCT [60], 1SHA, 1SHB [7], 1SHD [8], and 1SPS [57]), IRS-1 domain-containing complex (1IRS [20]), p56^{lck} domain-containing complexes (1LCJ [13], 1LCK [14], 1LKK and 1LKL [48]), a p85α SH2 domain-containing complex (1PIC [9]), a Cγ1 Syk SH2 domain-containing complex (2PLD [48]), a Shc PTB domain-containing complex (1SHC [56]), a Shc SH2 domain-containing complex (1TCE [49]), and a Grb2 SH2 domain-containing complex (1TZE [55]). A subset of these structures with available binding energies (or dissociation constants) was used to adjust the binding parameters and is shown in Table 1. X11 protein [21] was excluded because of its incomplete structure in the PDB. The last three complexes were used for control purposes only, because only the relative binding energies determined through competitive assays are available. Where IC_{50} or similar parameters are used to derive the binding energies, the conversion to relative binding energies $\Delta\Delta G_{ij}$ for two ligands i and j bound to an identical receptor was performed as follows (derived from [61]):

$$\Delta\Delta G_{ij} = RT\log(IC_{50i}/IC_{50j}) \quad (2)$$

Where R is the gas constant and T is room temperature.

Model building and refinement

Models of all complexes were regularized and energy minimized. The regularization procedure builds a full-atom model with idealized covalent geometry as close to the initial Cartesian coordinates of the corresponding PDB entries as possible and has been described elsewhere [44]. The C_α rmsd for the models were less than 0.5 Å with respect to their PDB structures. Internal coordinates (bond length, bond angle, torsion angle and phase angle) were used to describe both the receptors and the ligands, and six virtual variables were associated with the orientation of each ligand with respect to its receptor. Partial atomic charges of the phosphate atoms were assigned according to the widely parameterized Merck molecular force field 94 (MMFF94) [62]; the phosphorous atom carries partial charge of +1.37, whereas each of the three oxygen atoms carries a partial charge of −1.03.

The refinement procedure, as described in previous flexible docking studies [44,46], was applied to the initial models. The optimal-bias Monte Carlo minimization procedure [38] was applied to sample the sidechain χ angles at the ligand–receptor interface, defined as the atoms within 10 Å from the shared boundary. Torsion angles of the ligand (φ, ψ) were not globally sampled, but relaxed by local minimization after each conformational change. A stack of 40 conformations was used to keep track of the lowest conformations and to guide the simulation process [63]. A constant temperature of 700K was used in the Metropolis criterion. Loose distance restraints (defined in [44]) between the selected ligand–receptor C_α atom pairs were imposed to prevent the ligand from moving away from the binding site during refinement. Additional restraints were also assigned to C_β atoms in the models built for a Cγ1 Syk domain containing complex [45], a c-Src/YEEI complex [8], a Shc SH2 domain containing complex [49], and a Shc PTB domain containing complex [56] in order to keep tyrosine aromatic rings in the binding pockets. The ECEPP/3 [64–66] force field was used for the refinement, and the calculation for each model took about 7 CPU hours on a 195 MHz MIPS R10000 chip (~200,000 functional calls).

The above sidechain refinement procedure has previously been successfully applied to the simulation of two GCN4 helices [44], as well as the uncomplexed lysozyme and HyHel5 antibody [58] and β-lactamase and its inhibitor [41,67].

Binding free energy evaluation

The binding energy ΔG is the change in free energies when two molecules associate. Here the free energy of the complex was calculated with its lowest energy conformation instead of its conformational ensemble, a technique known as the 'predominant-states approximation' [68]. Free energies of the isolated receptor and ligand were evaluated using the bound conformations of the two molecules, but the entropy term was added according to the estimated number of free variables. It was also assumed that the van der Waals interactions at the complex interface could be estimated as proportional to the surface buried upon interaction, because of the low accuracy of its explicit evaluation. Therefore, the van der Waals contribution was taken into account by changing the surface energy density due to the solvation effect. This approximated binding energy can be presented as:

$$\Delta G_{calc} = \Delta G_{BEP}(\varepsilon) + s\Delta A - T\Delta S + C \quad (3)$$

where ΔG_{BEP} is the electrostatic contribution, s is the surface tension, ΔA denotes the buried surface area, $-T\Delta S$ is an entropy contribution associated with ligand sidechains and backbone, ε is the dielectric constant of the solute and C is a constant compensating the lost of translational and rotational entropies of a ligand after association, as well as other relatively constant contributions unaccounted for in (3). ΔG_{BEP} was calculated by solving the Poisson equation with a rapid-exact-boundary element (REBEL) method [50]. Hydrophobic effects, van der Waals interactions and other interface related contributions were modeled by multiplication of a surface tension constant s to the change of solvent-accessible surface area. The surface was constructed by rolling a probe of 1.4 Å in radius over the all-atom model. We used the following van der Waals radii: H, 1 Å; C, 1.6 Å; O, 1.35 Å; N, 1.45 Å. The sidechain entropy calculation was described previously [38], and the backbone entropy was calculated by multiplying the number of ligand backbone torsion angles by 0.5 kcal/mol.

Ab initio *ligand docking simulation by grid potentials*

The grid potentials were calculated in a box placed around the receptor-binding pocket. The box dimensions were approximately 40 Å × 40 Å × 40 Å and the grid size was 0.5 Å. Four different potentials, including electrostatic grid potential, van der Waals grid potentials for hydrogen and heavy atom probes, and hydrogen-bonding potential, were calculated for each point. The partial charges were taken from the ECEPP/3 force field [64–66]. All free torsion angles of docked peptides were randomized with the amplitude of 180° to generate a starting conformation. A tether with the strength of 1 kcal/Å$^{1/2}$ was assigned between the phosphorus atom of the phosphate ring and the binding pocket to reduce sampling space. Peptide conformations

were sampled with the OBMCM protocol [38] for 3×10^6 functional calls (~10 CPU hours on a 195 MHz R10000 processor). The lowest 40 conformations accumulated were then subjected to the refinement process described above. After 20,000 functional calls for each solution, the 40 candidates were reordered by the more realistic ECEPP/3 force field. A successful prediction was expected to be able to identify the native conformation among the few lowest energy candidates. The rmsd, defined as the root mean square deviation of two ligand backbone coordinates without optimal superimposition, was used as the criterion for the quality of predicted ligand conformations.

Acknowledgements

We thank the NIH (Grant R01 GM55418-01) and DOE (Grant DE-FG02-96ER62268) for financial support (this does not constitute an endorsement by either agency of the views expressed in the article). We are grateful to Mingming Zhou and Gabriel Waksman for their helpful suggestions. We also thank Alex Morrill for careful reading of the manuscript.

References

1. Kuriyan, J. & Cowburn, D. (1997). Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259-88.
2. Pawson, T. & Schlessinger, J. (1993). SH2 and SH3 domains. *Curr. Biol.* **3**, 434-442.
3. Cantley, L.C., *et al*., & Soltoff, S. (1991). Oncogenes and signal transduction. *Cell* **64**, 281-302.
4. Margolis, B. (1996). The PI/PTB domain: a new protein interaction domain involved in growth factor receptor signaling. *J. Lab. Clin. Med.* **128**, 235-241.
5. Zhou, M.M. & Fesik, S.W. (1995). Structure and function of the phosphotyrosine binding (PTB) domain. *Prog. Biophys. Mol. Biol.* **64**, 221-235.
6. Kavanaugh, W.M. & Williams, L.T. (1994). An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science* **266**,1862-1865.
7. Waksman, G., *et al*., & Kuriyan, J. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **358**, 646-653.
8. Gilmer, T., *et al*., & Berman, J. (1994). Peptide inhibitors of src SH3–SH2–phosphoprotein interactions, *J. Biol. Chem.* **269**, 31711-31719.
9. Breeze, L., *et al*., & Cartlidge, S.A. (1996). Structure of a specific peptide complex of the carboxy-terminal SH2 domain from the p85α subunit of phosphatidynositol 3-kinase. *EMBO. J.* **15**, 3579-3589.
10. Panayotou, G., *et al*., & Waterfield, M.D. (1993). Interactions between SH2 domains and tyrosine-phosphorylated platelet-derived growth factor β-receptor sequences: analysis of kinetic parameters by a novel biosensor-based approach. *Mol. Cell. Biol.* **13**, 3567-3576.
11. Piccione, E., *et al*., & Shoelson, S.E. (1993). Phosphatidylinositol 3-kinase p85 SH2 domain specificity defined by direct phosphopeptide/SH2 domain binding. *Biochem.* **13**, 3197-3202.
12. Narula, S.S., *et al*., & Dalgarno, D.C. (1995). Solution structure of the C-terminal SH2 domain of the human tyrosine kinase Syk complexed with a phosphotyrosine peptapeptide. *Structure* **3**, 1061-1073.
13. Eck, M.J., Shoelson, S.E. & Harrison, S.C. (1993). Recognition of a high-affinity phophotyrosyl peptide by the Src homology-2 domain of $p56^{lck}$. *Nature* **362**, 87-91.
14. Eck, M.J., Atwell, S.K., Shoelson, S.E. & Harrison, S.C. (1994). Structure of the regulatory domains of the Src-family tyrosine kinase Lck. *Nature* **368**, 764-769.
15. Payne, G., Shoelson, S.E., Gish, G.D., Pawson, T. & Walsh, C.T. (1993). Kinetics of $p56^{lck}$ and $p60^{src}$ homology 2 domain binding to tyrosine-phosphorylated peptides determined by a competition assay or surface plasmon resonance. *Proc. Natl Acad. Sci. USA* **90**, 4902-4906.
16. Sicheri, F., Moarefi, I. & Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature* **385**, 602-609.
17. Ladbury, J.E., Lemmon, M.A., Zhou, M., Creen, J., Botfield, M.C. & Schlessinger, J. (1995). Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. *Proc. Natl Acad. Sci. USA* **92**, 3199-3203.
18. Felder, S., *et al*., & Schlessinger, J. (1993). SH2 domain exhibit high-affinity binding to tyrosine phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Mol. Cell Biol*. **13**, 1449-1455.
19. Zhou, M.M., *et al*., & Fesik, S.W. (1995). Binding affinities of tyrosine-phosphorylated peptides to the COOH-terminal SH2 and NH2-terminal phosphotyrosine binding domains of Shc. *J. Biol. Chem.* **270**, 31119-31123.
20. Zhou, M.M., *et al*., & Fesik, S.W. (1996). Structural basis for IL-4 receptor phosphopeptide recognition by the IRS-1 PTB domain. *Nat. Struct. Biol*. **3**, 388-393.
21. Zhang, Z., *et al*., & Kuriyan., J. (1997). Sequence-specific recognition of the interalizaiton motif of the Alzheimer's amyloid precursor protein by the X11 PTB domain. *EMBO J.* **16**, 6141-6150.
22. Lee, C.H., *et al*., & Kuriyan, J. (1994). Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase. *Structure* **2**, 423-438.
23. Bradshaw, J.M., Grucza, R.A., Ladbury, J.E. & Waksman, G. (1998). Probing the 'two-pronged plug two-holed socket' model for the mechanism of binding of the Src SH2 domain to phosphotyrosyl peptides: a thermodynamic study. *Biochemistry* **37**, 9083-9090.
24. Timms, D. & Wilkinson, A.J. (1997). Estimation of binding affinity in structure-based design. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Volume 3, (van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., ed.), pp. 466-493, Kluwer Academic Publishers, Boston.
25. Ajay & Murcko, M.A. (1995). Computational methods to predict binding energy in ligand-receptor complexes. *J. Med. Chem.* **38**, 4953-4967.
26. Kollman, P.A. (1994). Theory of macromolecule–ligand interactions, *Curr. Opin. Struct. Biol*. **4**, 240-245.
27. Lesyng, B. & McCammon, J.A. (1993). Molecular modeling methods:basic techniques and challenging problems. *Pharmacol. Ther.* **60**:149-167.
28. Bash, P.A., Singh, U.C., Langridge, R. & Kollman, P.A. (1987). Free energy calculations by computer simulation. *Science* **236**, 564-568.
29. Muegge, I., Schweins, T. & Warshel, A. (1998). Electrostatic contributions to protein-protein binding affinities: application to Rap/Raf interaction. *Proteins* **30**, 407-423.
30. Vajda, S., Sippl, M. & Novotny, J. (1997). Empirical potentials and functions for protein folding and binding. *Curr. Opin. Struct. Biol.* **7**, 222-228.
31. Rosenfeld, R., Vajda, S. & DeLisi, C. (1995). Flexible docking and design. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 677-700.
32. Horton, N. & Lewis, M. (1992). Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**, 169-181.
33. Novotny, J., Bruccoleri, R.E. & Saul, F.A. (1989). The attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, HyHEL-5. *Biochemistry* **28**, 4735-4749.
34. Novotny, J., Bruccoleri, R.E., Davis, M. & Sharp, K.A. (1997). Empirical free energy calculations: a blind test and futher improvements to the method. *J. Mol. Biol*. **268**, 401-411.
35. Chothia, C. & Janin, J. (1975). Principles of protein–protein recognition. *Science* **256**, 705-708.
36. Kuntz, I.D. (1992). Structure-based strategies for drug design and discovery. *Science* **257**, 1078-1082.
37. Krystek, S., Stouch, T. & Novotny, J. (1993). Affinity and specificity of serine endopeptidase-protein inhibitor interactions: empirical free energy calculations based on X-ray crystallographic structures. *J. Mol. Biol*. **234**, 661-679.
38. Abagyan, R.A. & Totrov, M.M. (1994). Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol*. **235**, 983-1002.
39. Lengauer. T. & Rarey, M. (1996). Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **6**, 402-406.
40. Shoichet, B. & Kuntz, I.D. (1996). Predicting the structure of protein complexes: a step in the right direction. *Chem. Biol.* **3**, 151-156.
41. Janin, J. (1995). Protein–protein recognition. *Prog. Biophys. Molec. Biol*. **64**, 145-166.
42. Gulukota, K., Vajda, S. & DeLisi, C. (1996). Peptide docking using dynamic programming. *J. Comp. Chem.* **17**, 418-428.
43. Abagyan, R.A. (1997). Protein structure prediction by global energy optimization. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Volume 3, (van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., ed.), pp.363-394, Kluwer Academic Publishers, Boston.
44. Abagyan, R.A., Totrov, M.M. & Kuznetsov, D.N. (1994). ICM – a new method for protein modeling and design. Applications to docking and structure prediction from the distorted native conformation. *J. Comp. Chem.* **15**, 488-506.
45. Luty, B.A., Wasserman, Z.R., Stouten, P.F.W. & Hodge, D.N. (1995). A molecular mechanics/grid method for evaluation of ligand–receptor interactions. *J. Comp. Chem.* **16**, 454-464.
46. Totrov, M.M. & Abagyan, R.A. (1997). Flexible protein–ligand docking by global energy optimization in internal coordinates. *Proteins* **(suppl 1),** 215-220.

47. Tong, L., Warren, T.C., King, J., Betageri, R., Rose, J. & Jakes, S. (1996). Crystal structures of the human p56^{lck} SH2 domain in complex with two short phophotyrosyl peptides at 1.0 Å and 1.8 Å resolution. *J. Mol. Biol.* **256**, 601-610.
48. Pascal, S.M., *et al*., & Forman-Kay J.D. (1994). Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-γ1 complexed with a high affinity binding peptide. *Cell* **77**, 461-472.
49. Zhou, M., *et al*., & Fesik, S.W. (1995). Solution structure of the shc SH2 domain complexed with a tyrosine-phosphorylated peptide from the T-cell receptor. *Proc. Natl Acad. Sci. USA* **92**, 7784-7788.
50. Gilson, M.K. (1997). Modeling protonation equilibria in biomolecules. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Volume 3. (van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., ed.), pp.199-222, Kluwer Academic Publishers, Boston.
51. Bradshaw, J. & Waksman, G. (1998). A calorimetric investigation of proton linkage by monitoring both the enthalpy and association constant of binding: application to the interaction of the Src SH2 domain with a high affinity tyrosyl phosphopeptide. *Biochemistry* **37**, 15400-15407.
52. Totrov, M.M. & Abagyan, R.A. (1996). The contour-buildup algorithm to calculate the analytical molecular surface. *J. Struct. Biol.* **116**, 138-143.
53. Tunon, I., Silla, E., Pascual-Ahuir, J.L. (1992). Molecular surface area and hydrophobic effect. *Protein Eng.* **5**, 715-716.
54. Wesson, L. & Eisenberg, D. (1992). Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1**, 227-235.
55. Rahuel, J., *et al*., & Grutter, M.G. (1996). Structural basis for specificity of Grb2-SH2 revealed by a novel ligand binding mode. *Nat. Struct. Biol.* **3**, 586-589.
56. Zhou, M., *et al*., & Feslk, S.W. (1995). Structure and ligand recognition of the phosphotyrosine binding domain of shc. *Nature* **378**, 584-592.
57. Waksman, G., Shoelson, S.E., Pant, N., Cowburn, D. & Kuriyan J. (1993). Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **72**, 770-790.
58. Totrov, M.M. & Abagyan, R.A. (1994). Detailed ab initio prediction of lysozyme-antibody complex with 1.6 Å accuracy. *Nat. Struct. Biol.* **1**, 259 -263.
59. Caflisch, A., Niederer, P. & Anliker, M. (1992). Monte Carlo docking of oligopeptides to proteins. *Proteins* **13**, 223-230.
60. Xu, R.X., Word, J.M., Davis, D.G., Rink, M.J., Willard, D.H., Jr. & Gampe, R.T., Jr. (1995). Solution structure of the human pp60 c-src sh2 domain complexed with a phosphorylated tyrosine pentapeptide. *Biochemistry* **34**, 2107-2121.
61. Pratt, W.B. Taylor, P. (1990). *Principle of drug action: the basis of pharmacology,* (3rd edn), Churchill Livingstone, New York.
62. Halgren, T.A. (1992). Merck Molecular Force Field. I-V. *J. Comp. Chem.* **17**, 1995, 490-641.
63. Abagyan, R.A. & Argos, P. (1992). Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J. Mol. Biol.* **225**, 519-532.
64. Momany, F.A., McGuire, R.F., Burgess, A.W. & Scheraga, H.A. (1975). Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* **79**, 2361-2381.
65. Nemethy, G., Pottle, M.S. & Scheraga, H.A. (1983). Energy parameters in polypeptides. IX. Updating of geometric parameters, nonbonded interactions and hydrogen bond interactions for the naturally occurring amino acid. *J. Phys. Chem.* **87**, 1883-1887.
66. Nemethy, G., *et al*., & Scheraga, H.A. (1992). Energy parameters in polypeptides. 10. Improved geometric parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* **96**, 6472-6484.
67. Strynadka, N.C.J., *et al*., & James, (1996). M.N.G. Molecular docking programs successfully determine the binding of a β-lactamase inhibitory protein to term-1 β-lactamase. *Nat. Struct. Biol.* **3**, 233-239.
68. Gilson, M.K., Given, J.A. & Head, M.S. (1997). A new class of models for computing receptor–ligand binding affinities. *Chem. Biol.* **4**, 87-92.
69. Lemmon, M.A. & Ladbury, J.E. (1994). Thermodynamic studies of tyrosyl-phosphopeptide binding to the SH2 domain of p56^{lck}. *Biochemistry* **33**, 5070-5076.
70. Case, R.D., *et al*., & Shoelson, S.E. (1994). SH-PTP2 SH2 domain binding specificity is defined by direct interactions with platelet-growth factor beta-receptor, epidermal growth factor receptor, and insulin receptor substrate-1-derived phosphopeptide. *J. Biol. Chem.* **269**, 10467-10474.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad – for further information, see the explanation on the contents pages.