Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 306 (2003) 225–230
BBRC
www.elsevier.com/locate/ybbrc

An investigation of phosphopeptide binding to SH2 domain

Jin Kak Lee,[a] Taesung Moon,[a] Myung Whan Chi,[a] Jin-Soo Song,[a] Young-Sang Choi,[b] and Chang No Yoon[a,*]

[a] *Bioanalysis and Biotransformation Research Center, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, South Korea*
[b] *Department of Chemistry, Korea University, Seoul 136-761, South Korea*

Received 22 April 2003

Abstract

A comparative molecular field analysis (CoMFA) was carried out to investigate quantitative structure–activity relationships for SH2-binding phosphopeptides. Two alignment rules were applied in our CoMFA model. The phosphopeptide backbone atoms were used for superposition in alignment I and the backbone atoms of peptide-binding residues of SH2–phosphopeptide complexes were used in alignment II to consider the position of phosphopeptides in SH2-binding sites. The higher correlation and predictivity in alignment II (r^2 value of 0.961 and cross-validated r^2 value of 0.682) suggest that the consideration of peptide-binding position at the binding sites gives rise to better results when the ligand–receptor complex structure is considered. In addition, CoMFA contour and electrostatic maps were well accorded with the experimental results in which the replacement of N-terminal residues with an acetyl group reduced the binding affinity. Therefore, the modification of molecular size and charge of phosphopeptides can be carried out based on these contour maps in order to increase binding affinities.

Keywords: QSAR; CoMFA; SH2; Phosphopeptides; Binding affinity; Alignment rule

The src homology 2 (SH2) domains mediate specific protein–protein interactions through binding to specific, phosphotyrosine-containing peptide sequences in their targets. Tyrosine phosphorylation of cellular target proteins mediates normal signal transduction and plays a significant role in regulation of cell growth, differentiation, and oncogenesis [1–3]. The cellular level of tyrosine phosphorylation is maintained by kinase and phosphatase. Tyrosine kinase activity is implicated in the activation of growth factor receptor, such as EGF-R (epidermal growth factor receptor) and PDGF-R (platelet-derived growth factor receptor). The binding of growth factor to its surface receptor leads to activation of intrinsic protein kinase activity which is followed by tyrosine autophosphorylation of target substrate proteins [3]. In addition to the kinase catalytic domain (SH1; src homology 1), tyrosine kinases of the src family have highly conserved amino acid motifs named SH2 (src homology 2) and SH3 (src homology 3), which are non-catalytic domains [4]. The SH3 domain is a distinct motif which may modulate interactions with the cytoskeleton and membrane together with SH2 domain. The transmission of growth factor-mediated signals, for example, depends on the sequence-specific recognition of phosphorylated tyrosines by SH2 domains. Many studies report that they are involved in mediating protein–protein interactions in the downstream of growth factor receptors, including GAP (Ras GTPase-activating protein), PIK (phosphatidylinositol 3′-kinase), and phospholipase C-γ [5]. The function of the SH2 domain is less clear. The SH2 may play a role in the regulation of kinetic activity itself. When Tyr-527 at the C-terminus of src is phosphorylated, it suppresses the kinase activity [6,7]. Substitution of Tyr-527 with phenylalanine increases kinase activity [8]. Many crystal and solution structures of SH2 have been reported [9–16] and crystal structures of the SH2 from src and lck with a phosphopeptide-containing pYEEI motif have been determined [9,16]. These studies show the presence of two pockets. One pocket binds the pY (phosphotyrosine) with ion-paring, hydrogen-bonding, and amino-aromatic interactions. The other pocket

* Corresponding author. Fax: +82-2-958-5059.
E-mail address: cody@kist.re.kr (C.N. Yoon).

0006-291X/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0006-291X(03)00932-X

binds the pY+3 (isoleucine) with hydrophobic interactions. The peptide is anchored by insertion of pY and pY+3 side chains into their pockets. In the low affinity phosphopeptide/v-src complex, the pY+3 do not insert to the binding pocket and the peptide main chain is displaced from the surface of the domain [16]. A series of quantitative studies of the relative affinities of synthetic phosphopeptides for SH2 domain have been carried out [17–19]. The computer-aided molecular design has been greatly enhanced in recent years by the development and application of three-dimensional quantitative structure–activity relationships (3D-QSAR). The comparative molecular field analysis (CoMFA) [20] is one of the most widely used 3D-QSAR methods. This method is based on the idea that the ligand–receptor interaction is primarily noncovalent and shape dependent. To obtain further insight into the relationships between the structure and activity of SH2-binding phosphopeptides, the QSAR study using CoMFA method was carried out.

Methods

Data set for analysis. Published data on 30 pY324 phosphopeptides for GST-Lck SH2 [17] were used for this study (Table 1). Of the 30 phosphopeptides, 20 were used as training set and 10 as test set. The activity of phosphopeptides is expressed by IC_{50} values of multiplied relative affinities in original paper by 2.87 μM [17].

Conformation sampling. For the construction of the phosphopeptide-binding structures, the crystal structure of pY324 SH2–phosphopeptide complex [9] was used as a template for other 29 structures. Each complex structure was built by the replacement of corresponding residues and the orientation of side chains was adjusted manually. All 30 structures were then subjected to molecular dynamics simulations to get optimum conformations. Prior to molecular dynamics, these complexes were immersed in water box and energy-minimized using steepest descents for 200 steps and conjugate gradients for 300 steps with periodic boundary conditions. By this procedure short contacts in model protein were relieved and the water molecules were adjusted. The minimized coordinates were then used as starting point for 25-ps molecular dynamics simulations at 300 K with periodic boundary conditions. The initial velocities were taken from a Maxwell–Boltzmann distribution for target temperature. The leap-frog algorithm was used to integrate the equations of motion with an integration time-step of 1 fs (femtosecond). The minimum energy conformations during molecular dynamics were sampled and optimized for QSAR analysis. All the molecular mechanics and dynamics calculations were carried out by using Discover2.98 program (Accelrys Inc.) with CVFF (Consistent Valence Force Field) with nonbonded cutoff of 9.0 Å running on a CRAY Y-MP at KISTI Supercomputing Center and graphic display was achieved by InsightII97.0 (Accelrys Inc.) and SYBYL6.3 (Tripos Inc.) implemented on an Silicon Graphics O2 workstation.

Structure alignment. Positioning of the molecular structure within the fixed lattice is very important in CoMFA, since the relative interaction energies depend strongly on relative molecular positions. Two alignment rules were defined in this study. In alignment I, the

Table 1
The sequences and relative affinities of altered pY324 phosphopeptides for GST-LckSH2 [17]

No.	Phosphopeptide	Phosphopeptide sequence	Relative affinity (μM)	IC_{50}	$-\log IC_{50}$
1	ac-py324+1a	Ac-pYAEIPI	42	120.540	3.9189
2	ac-py324+1d	Ac-pYDEIPI	13	37.310	4.4282
3	ac-py324+1e	Ac-pYEEIPI	1.7	4.879	5.3117
4	ac-py324+1f	Ac-pYFEIPI	18	51.660	4.2868
5	ac-py324+1g	Ac-pYGEIPI	>47	>134.890	~3.8700
6	ac-py324+1i	Ac-pYIEIPI	18	51.660	4.2868
7	ac-py324+1k	Ac-pYKEIPI	≫47	≫134.890	~3.8700
8	ac-py324+1l	Ac-pYLEIPI	47	134.890	3.8700
9	ac-py324+1m	Ac-pYMEIPI	26	74.620	4.1271
10	ac-py324+1q	Ac-pYQEIPI	5.9	16.933	4.7713
11	ac-py324+1s	Ac-pYSEIPI	17	48.790	4.3117
12	ac-py324+4a-NH2	Ac-pYEEIA-NH2	3.2	9.184	5.0370
13	py324	EPQpYEEIPIYL	1.0	2.870	5.5421
14	py324+2i	EPQpYEIIPIYL	1.6	4.592	5.3380
15	py324+3a	EPQpYEEAPIYL	10	28.700	4.5421
16	py324+3e	EPQpYEEEPIYL	6.1	17.507	4.7568
17	py324+3l	EPQpYEELPIYL	4.8	13.776	4.8609
18	py324+3m	EPQpYEEMPIYL	2.7	7.749	5.1108
19	py324+3q	EPQpYEEQPIYL	24	68.880	4.1619
20	py324+3v	EPQpYEEVPIYL	2.2	6.314	5.1997
21	py324+4n	EPQpYEEINIYL	3.5	10.045	4.9981
22	py355	ASQVpYFTYDPYSE	260	746.200	3.1271
23	py697	GGVDpYKNIHLE	430	1234.100	2.9086
24	py708	LEKKpYVRRDSG	≫760	≫2181.200	~2.6613
25	py723	VDTpYVEMRPVS	>760	>2181.200	~2.6613
26	py730	CVSpYVVPTKAD	270	774.900	3.1108
27	py745	IGSpYIERDVTG	760	2181.200	2.6613
28	py809	NDSNpYIVKGNA	>760	>2181.200	~2.6613
29	py936	NHIpYSNLANSS	120	344.400	3.4629
30	py969	QPNNpYQFS	140	401.800	3.3960

phosphopeptides were aligned by using root mean square (rms) fitting of backbone atoms of pY, pY+1, pY+2, and pY+3 of phosphopeptides. In alignment II, the phosphopeptides were aligned by using rms fitting of backbone atoms of peptide-binding residues in SH2–phosphopeptide complexes in order to consider the position of phosphopeptides at the binding sites in SH2 domain.

CoMFA interaction energy calculation. For alignment I and alignment II, the steric (van der waals 6–12) and electrostatic (Columbic with a $1/r$ dielectric) interaction energies were separately calculated at each lattice interaction on a regularly spaced (2.0 Å) grid of dimensions $29.64 \times 46.13 \times 32.12$ Å and $26.77 \times 47.06 \times 30.07$ Å, respectively. An sp^3 carbon atom with a van der Waals radius and +1.0 charge was selected as a probe to calculate the steric and electrostatic interactions. The cutoff values were +30 kcal/mol for steric energy and +15 kcal/mol for electrostatic energy. Whenever the probe atom experiences a steric repulsion greater than cutoff value, the steric interaction is set to the cutoff value and the electrostatic interaction is set to the mean value of the other molecules' electrostatic interactions at the same locations. The standard CVFF atomic charges were used for the calculation of electrostatic interactions.

PLS analysis. Partial least square (PLS) fitting was used to obtain a 3D-QSAR. PLS is a regression method for solving linear models. It has been applied in lots of QSAR studies to rationalize the structural features affecting biological activity. The predictive activity of a model is quantitated in terms of r^2 which is defined as [21]

$$r^2 = 1.0 - \frac{\sum (y_{\text{pred}} - y_{\text{actual}})^2}{\sum (y_{\text{actual}} - y_{\text{mean}})^2}, \quad (1)$$

where y_{pred}, y_{actual}, and y_{mean} are predicted, actual, and mean values of the target property, respectively. The predictive r^2 can assume a negative value reflecting a complete lack of predictive ability. The optimum number of components was then determined by leave-one-out cross-validation as that which yields either the smallest rms error or the largest cross-validated r^2 value. In the leave-one-out cross-validation each molecule is systematically excluded from the set and its activity is predicted by a model which is derived from the rest of the molecules. A final PLS analysis was carried out on the entire data set using the reported optimum number of molecules without cross-validation. The steric and electrostatic fields were scaled using CoMFA standard scaling. In most cases the use of CoMFA standard scaling gave better r^2 values compared with those obtained using no scaling. All CoMFA and PLS analyses were performed using SYBYL6.3 (Tripos Inc.) implemented on Silicon Graphics O2 workstation.

Table 2
Summary of CoMFA results for training set of SH2-binding phosphopeptides

	Alignment I	Alignment II
No. of training set	20	20
No. of components	2	3
$r^2_{\text{cross-validated}}$	0.739	0.682
$r^2_{\text{conventional}}$	0.948	0.961
F value	154.347	130.841
Standard error	0.200	0.179
Steric contributions	0.395	0.402
Electric contributions	0.605	0.598

Results and discussion

CoMFA of SH2-binding phosphopeptides

The results of CoMFA studies for 20 training set members are summarized in Table 2. The values of r^2 suggest that each subset compound provides adequate structure–activity information to construct a quality 3D-QSAR (0.948 for alignment I and 0.961 for alignment II). The r^2 value of alignment II is higher than that of alignment I, which suggests that the information about the binding mode of ligands is important in the prediction of biological activities. Two and three components were used in PLS analyses for alignments I and II in order to give optimal fitting, respectively. The relative contributions of steric (van der waals) and electrostatic (Coulombic) interactions in the CoMFA models were 0.395 and 0.605 for alignment I and 0.402 and 0.598 for alignment II, which proposes that the binding activities of the phosphopeptides in this model system are more dependent on the electrostatic interactions. The experimental and calculated activities of alignments I and II in training set are shown in Fig. 1, which shows good correlations between experimental and calculated values. The analysis based on steric (van der waals) and electrostatic (Columbic) fields gives rise to the predictive r^2 values of 0.531 and 0.717 for alignments I and II, respectively. If model number 25 is not considered in this analysis, the predictive r^2 value becomes higher (0.681 for alignment I and 0.792 for alignment II). It is reasonable that model number 25 is not considered because it contains the experimental error [17]. The experimental and corresponding predicted activities in test set are

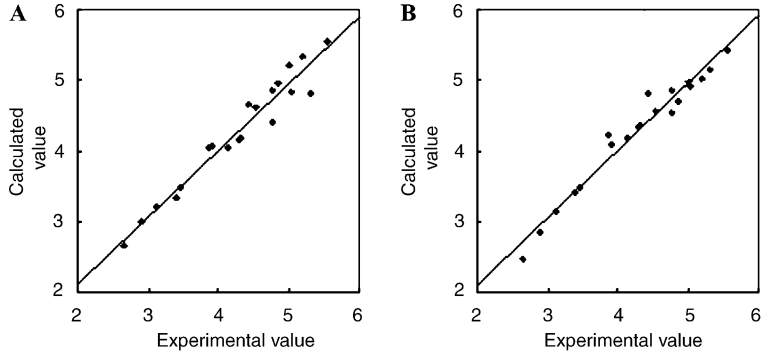

Fig. 1. Plots of experimental versus calculated $-\log IC_{50}$ values for training set: (A) Alignment I and (B) Alignment II.

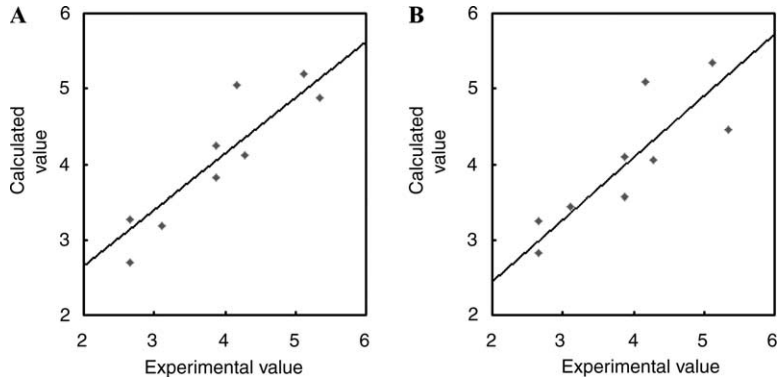

Fig. 2. Plots of experimental versus calculated $-\log IC_{50}$ values for test set: (A) Alignment I and (B) Alignment II. No. 25 is not considered because No. 25 contains experimental error.

shown in Fig. 2. Comparison of these calculated values with experimental (IC_{50}) values shows that CoMFA can generally predict the binding activities in test set within a log unit. It is noticeable that the difference of r^2 values between alignments I and II is greater in test set than in training set. Although alignment II shows a little higher r^2 values (order of 0.01) in training set, it shows much higher values in test set (order of 0.1). The predictive r^2 values in test set were 0.681 and 0.792 for alignments I and II, respectively, while the r^2 values in training set were 0.948 and 0.961. The higher correlations in alignment II may be caused from the consideration of binding modes in this model system. The alignment I constructed in the absence of the knowledge of binding modes may not explain the actual binding conformation, since the different ligands (phosphopeptides) may not bind the receptor (SH2 domain) in the same way. Therefore, it is important to consider the binding mode between ligand and receptor in CoMFA analysis when the information about the binding modes is available.

CoMFA fields

In order to visualize the derived 3D-QSAR model, CoMFA contour maps were obtained using an sp^3 car-

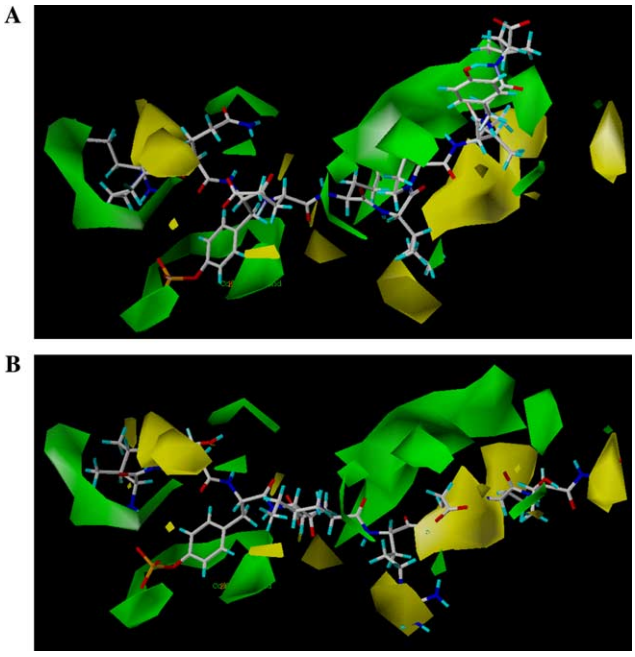

Fig. 3. The CoMFA steric contour plots based on alignment II: (A) No. 13 and (B) No. 27. Green regions represent sterically favored areas and yellow regions represent sterically unfavored areas.

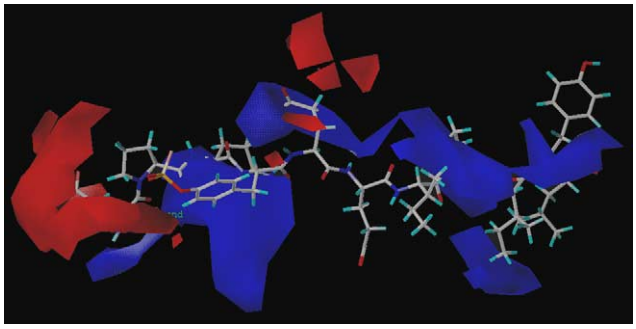

Fig. 4. The CoMFA electrostatic contour plot based on alignment II. Blue regions represent positive charge favored areas and red regions represent negative charge favored areas.

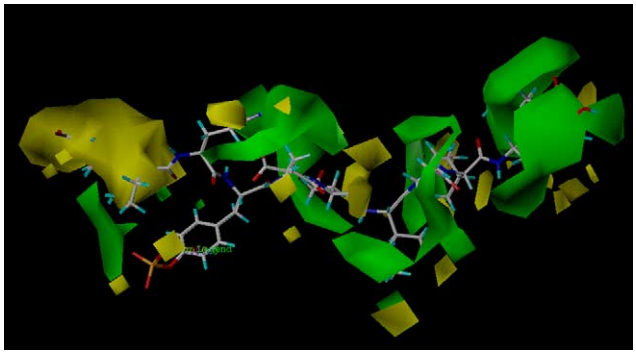

Fig. 5. The CoMFA steric contour plot based on energy-minimized structure (without molecular dynamics). Green regions represent sterically favored areas and yellow regions represent sterically unfavored areas.

bon atom with +1 charge. The fields were calculated by multiplying the β coefficient and the standard deviation of columns in the table generated by CoMFA. The steric contour maps for alignment II were shown in Fig. 3 representing 80% (green) and 20% (yellow) level contribution. The green and yellow contours describe regions of space around the molecules where increases in steric bulk enhance and diminish the affinity, respectively. Most of the phosphopeptides with low binding affinity generally occupy the yellow (sterically unfavorable) region, whereas the highly active phophopeptides tend to occupy the green (sterically favorable) region. CoMFA contour maps were well accorded with experimental results, for example, the replacement of N-terminal residues with an acetyl group reduced the binding affinity (Table 1). Therefore, the modification of phosphopeptides can be carried out based on this steric contour map in order to increase binding affinities in two ways; the increase of molecular size in green regions and the decrease of molecular size in yellow regions.

Electrostatic contour map is shown in Fig. 4 representing 90% (blue) and 30% (red) level contributions. The blue and red contours describe the positive charge favored and the negative charge favored regions, respectively. The red region near pY+1 suggests that the negative charge of pY+1 plays an important role in binding affinity. It is well accorded with the experiment that the substitution of Glu with hydrophobic side chain had an effect on binding affinity ranging from 18- to 47-fold lower [17]. Based on this electrostatic contour map, the modification of phosphopeptides can be carried out as follows; the increases of positive charge in blue regions and negative charge in red regions.

The energy-minimized (EM) structures (without molecular dynamics) gave higher r^2 value in training set and predictive r^2 value in test set than the molecular-dynamics (MD) structures did. The r^2 values were 0.973 and 0.723 (0.794 if No. 25 was not considered) in training and test sets, respectively. In contrast to the steric contour map based on the MD structures in which the N-terminal region is sterically favorable, the steric contour map (60% green and 40% yellow) based on the EM structures in Fig. 5 shows that the N-terminal region is sterically unfavorable. This result based on EM structures is not accorded with the experiment [17] in which the replacement of N-terminal residues with an acetyl group reduced the binding affinity. It is difficult to say which structure of the EM or MD structure is closer to the real one because SH2-altered phosphopeptide complexes used in this study are not the structures based on X-ray or NMR structure. But it can be suggested that the MD structure is closer to real one than EM structure. The MD structures of phosphopeptides had large structural variation, whereas the EM structures yield little conformational change after point mutation (replacement of residue). The conformational changes induced by these point mutations were suggested by Payne et al. [17].

Acknowledgments

This work was supported by National Research Laboratory program of the Ministry of Science and Technology and advanced backbone IT technology development program of the Ministry of Information and Communication of Korea.

References

[1] S.K. Hanks, A.M. Quinn, T. Hunter, The protein kinase family: conserved features and deduced phylogeny of the catalytic domains, Science 241 (1988) 42–52.
[2] A. Ullrich, J. Schlessinger, Signal transduction by receptors with tyrosine kinase activity, Cell 61 (1990) 203–212.
[3] L.T. Williams, Signal transduction by the platelet-derived growth factor receptor, Science 243 (1989) 1564–1570.
[4] C.A. Koch, D. Anderson, M.F. Morgan, C. Ellis, T. Pawson, SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins, Science 252 (1991) 668–674.
[5] L.C. Cantley, K.R. Auger, C. Carpenter, B. Duckworth, A. Granziani, R. Kapeller, S. Soltoff, Oncogenes and signal transduction, Cell 64 (1991) 281–302.
[6] K.E. Amrein, B.M. Sefton, Mutation of a site of tyrosine phosphorylation in the lymphocyte-specific tyrosine protein kinase, $p56^{lck}$, reveals its oncogenic potential in fibroblasts, Proc. Natl. Acad. Sci. USA 185 (1988) 4247–4251.
[7] J.A. Cooper, K.L. Gould, C.A. Cartwright, T. Hunter, Tyr^{527} is phosphorylated in $pp60^{c\text{-}src}$: implications for regulation, Science 231 (1986) 1431–1434.
[8] C.A. Cartwright, W. Eckhart, S. Simon, P.L. Kaplan, Cell transformation by $pp60^{c\text{-}src}$ mutated in the carboxy-terminal regulatory domain, Cell 49 (1987) 83–91.
[9] M.J. Eck, S.E. Shoelson, S.C. Harrison, Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of $p56^{lck}$, Nature 362 (1993) 87–91.
[10] M.J. Eck, S.K. Atwell, S.E. Shoelson, S.C. Harrison, Structure of the regulatory domains of the src-family tyrosine kinase Lck, Nature 368 (1994) 764–769.
[11] V. Makol, G. Baumann, T.H. Keller, U. Manning, M.G.M. Zurini, The crytal structures of the SH2 domain of $p56^{lck}$ complexed with two phosphopeptides suggest a gated peptide binding site, J. Mol. Biol. 246 (1995) 344–355.
[12] S.S. Narula, R.W. Yuan, S.E. Adams, O.M. Green, J. Green, T.B. Philips, L.D. Zydowsky, M.C. Botfield, M. Hatada, E.R. Laird, M.J. Zoller, J.L. Karas, D.C. Dalgarno, Solution structure of C-terminal SH2 domain of the human tyrosine kinase syk complexed with a phosphotyrosine pentapeptide, Structure 3 (1995) 1061–1073.
[13] S.M. Pascal, A.U. Singer, G. Gish, T. Yamazaki, S.E. Shoelson, T. Pawson, L.E. Kay, J.D. Forman-kay, Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-γ complexed with a high affinity binding peptide, Cell 77 (1994) 461–472.
[14] L. Tong, T.C. Warren, J. King, R.B.J. Rose, S. Jakes, Crytal structures of the human $p56^{lck}$ SH2 domain in complex with two short phosphotyrosyl peptides at 1.0 and 1.8 Å resolution, J. Mol. Biol. 256 (1996) 601–610.
[15] G. Waksman, D. Kominos, S.C. Robertson, N. Pant, D. Baltimore, R.B. Birge, D. Cowburn, H. Hanafusa, B.J. Mayer, M. Overduin, M.D. Resh, C.B. Rios, L. Silverman, J. Kuriyan,

Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides, Nature 358 (1992) 646–653.
[16] G. Waksman, S.E. Shoelson, N. Pant, D. Cowburn, J. Kuriyan, Binding of high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms, Cell 72 (1993) 779–790.
[17] G. Payne, L.A. Stolz, D. Pei, H. Band, S.E. Shoelson, C.T. Walsh, The phosphopeptide-binding specificity of Src family SH2 domains, Chem. Biol. 1 (1991) 99–105.
[18] E. Piccione, R.D. Case, S.M. Domchek, P. Hu, M. Chaudhuri, J.M. Backer, J. Schlessinger, S.E. Shoelson, Phosphatidylinocitol 3-kinase p85 SH2 domain specificity defined by direct phosphopeptide/SH2 domain binding, Biochemistry 32 (1993) 3197–3292.
[19] Z. Songyang, S.E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W.G. Haser, F. King, T. Roberts, S. Ratnofsky, R.J. Lechleider, B.G. Neel, R.B. Birge, J.E. Fajardo, M.M. Chou, H. Hanafusa, B. Schaffhausen, L.C. Cantley, SH2 domains recognize specific phosphopeptide sequences, Cell 72 (1993) 767–778.
[20] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier protein, J. Am. Chem. Soc. 110 (1988) 5059–5067.
[21] Sybyl Molecular Modeling Software, version 6.3. Sybyl theory manual. Tripos, Inc., St. Louis, 1996.