Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Binding of flexible and constrained ligands to the Grb2 SH2 domain: structural effects of ligand preorganization

John H. Clements, John E. DeLorbe, Aaron P. Benfield and Stephen F. Martin*

Department of Chemistry and Biochemistry, The Institute of Cellular and Molecular Biology, Texas Institute of Drug and Diagnostic Development, The University of Texas, Austin, Texas 78712, USA

Correspondence e-mail: sfmartin@mail.utexas.edu

Received 2 April 2010
Accepted 4 September 2010

PDB References:
Grb2 SH2 domain, complex with **2**, 3c7i; complex with **3**, 3in8; complex with **4**, 3imd; complex with **5**, 3kfj; complex with **6**, 2huw; complex with **7**, 3imj; complex with **8**, 3in7.

Structures of the Grb2 SH2 domain complexed with a series of pseudopeptides containing flexible (benzyl succinate) and constrained (aryl cyclopropanedicarboxylate) replacements of the phosphotyrosine (pY) residue in tripeptides derived from Ac-pY*X*N-NH_2 (where X = V, I, E and Q) were elucidated by X-ray crystallography. Complexes of flexible/constrained pairs having the same pY + 1 amino acid were analyzed in order to ascertain what structural differences might be attributed to constraining the phosphotyrosine replacement. In this context, a given structural dissimilarity between complexes was considered to be significant if it was greater than the corresponding difference in complexes coexisting within the same asymmetric unit. The backbone atoms of the domain generally adopt a similar conformation and orientation relative to the ligands in the complexes of each flexible/constrained pair, although there are some significant differences in the relative orientations of several loop regions, most notably in the BC loop that forms part of the binding pocket for the phosphate group in the tyrosine replacements. These variations are greater in the set of complexes of constrained ligands than in the set of complexes of flexible ligands. The constrained ligands make more direct polar contacts to the domain than their flexible counterparts, whereas the more flexible ligand of each pair makes more single-water-mediated contacts to the domain; there was no correlation between the total number of protein–ligand contacts and whether the phosphotyrosine replacement of the ligand was preorganized. The observed differences in hydrophobic interactions between the complexes of each flexible/constrained ligand pair were generally similar to those observed upon comparing such contacts in coexisting complexes. The average adjusted B factors of the backbone atoms of the domain and loop regions are significantly greater in the complexes of constrained ligands than in the complexes of the corresponding flexible ligands, suggesting greater thermal motion in the crystalline state in the former complexes. There was no apparent correlation between variations in crystal packing and observed structural differences or similarities in the complexes of flexible and constrained ligands, but the possibility that crystal packing might result in structural variations cannot be rigorously excluded. Overall, it appears that there are more variations in the three-dimensional structure of the protein and the ligand in complexes of the constrained ligands than in those of their more flexible counterparts.

1. Introduction

One of the challenges in molecular recognition in biological systems is correlating how a change in the structure of a small

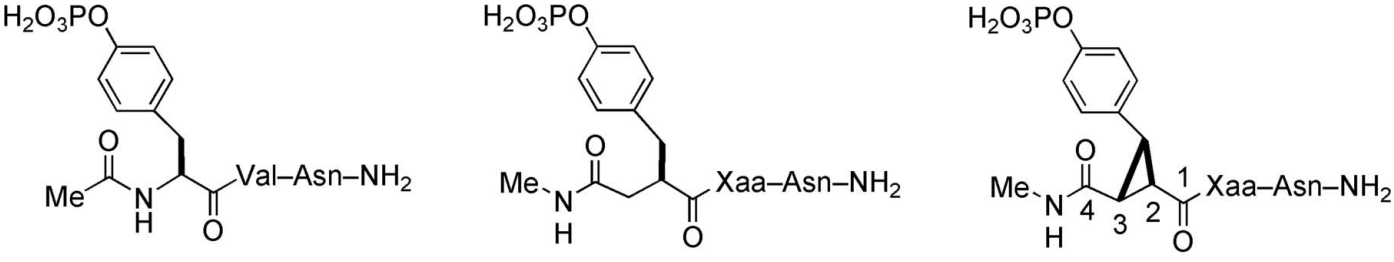
molecule affects the energetics and structure of the resultant protein–ligand complex. In this context, we initiated a study several years ago that was directed towards assessing how changing the structure of phosphotyrosine-derived ligands affected their affinities for the SH2 domain of the human growth receptor-bound protein 2 (Grb2; Bradshaw & Waksman, 2002; Machida & Mayer, 2005). Grb2 is a 25 kDa cytosolic protein consisting of 217 residues that comprises an SH2 domain and two flanking SH3 domains (Chardin *et al.*, 1995; Rahuel *et al.*, 1996, 1998). Grb2 serves as an adapter protein, with the SH2 domain binding pY residues on membrane-bound receptor tyrosine kinases (RTKs) and the proline-rich SH3 domains binding to the Ras exchange factor Son of Sevenless (Sos), thereby leading to activation of Ras, an important regulator of cell growth (Vojtek & Ders, 1998). Because the role of Grb2 in over-activation of the Ras pathway has been linked to hyperproliferative diseases such as leukemia, and breast and ovarian cancers (Chardin *et al.*, 1995; Rahuel *et al.*, 1996, 1998), the identification of potent antagonists that selectively prevent binding of the Grb2 SH2 domain to RTKs has been pursued as a potential strategy that may lead to novel treatments for such diseases.

The SH2 domain of Grb2 recognizes peptides containing the tripeptide consensus sequence pY*X*N, where *X* is typically a hydrophobic amino acid, although hydrophilic amino acids such as Gln, Glu and Lys are also found in potent Grb2 SH2-binding ligands (Kessels *et al.*, 2002). Ligand preorganization is a common strategy for enhancing affinity (Mann, 2003; Nakanishi & Kahn, 2003), so we queried how preorganizing pY*X*N-derived ligands in their biologically active conformations might affect the binding energetics and structure of their complexes with the SH2 domain of Grb2. Although the increased affinity arising from ligand preorganization has been commonly associated with a more favorable entropy of binding (Gerhard *et al.*, 1993; Kahn *et al.*, 1998; Davidson *et al.*, 2002), we and others have recently demonstrated that introducing a conformational constraint can result in enhanced affinity that arises from a more favorable enthalpy of binding (Benfield *et al.*, 2006; Delorbe *et al.*, 2009; Udugamasooriya & Spaller, 2008). Indeed, it is now apparent that ligand preorganization is not necessarily accompanied by a more favorable binding entropy.

In order to correlate how specific changes in ligand flexibility affect binding energetics, a detailed structural analysis of the complexes formed between the target protein and the ligands being compared is required. Accordingly, we prepared the pseudopeptide ligands **2**–**8**, which are related to the native tripeptide Ac-pYVN-NH_2 (**1**; Fig. 1), and determined the thermodynamic parameters for their binding to the Grb2 SH2 domain (Benfield *et al.*, 2006; Delorbe *et al.*, 2009). The structures of the flexible and constrained replacements of the phosphotyrosine moiety in **2**–**5** and **6**–**8**, respectively, were inspired by our previous studies of Src SH2-binding ligands (Davidson *et al.*, 2002). Namely, the benzyl succinyl moiety in **2**–**5** is a flexible replacement of the pTyr amino acid in **1** and the substituted cyclopropane ring in **6**–**8** serves as a rigid mimic of the pTyr amino acid in **1**. In order to minimize differences in desolvation effects associated with binding, the corresponding flexible and constrained ligand pairs (*e.g.* **2** and **6**, **3** and **7**, and **4** and **8**) each possess the same number and type of non-H atoms, the same functional groups and the same number of hydrogen-bond donors and acceptors. The structures of the complexes of **2**–**8** with the Grb2 SH2 domain were determined by X-ray diffraction and details relevant to the observed differences in their binding affinities have been published (Benfield *et al.*, 2006; Delorbe *et al.*, 2009). We now provide a more detailed analysis of the structures of the complexes of **2**–**8** with the Grb2 SH2 domain.

2. Materials and methods

The syntheses of **2**–**8** and the expression and purification of the Grb2 SH2 domain have been described in detail elsewhere (Benfield *et al.*, 2006; Delorbe *et al.*, 2009). Briefly, the DNA construct containing the QE60 plasmid and residues 53–163 of the Grb2 SH2 domain was obtained from the Schering–Plough Research Institute and expressed in *Escherichia coli* (SG13009, Qiagen; McNemar *et al.*, 1997). Cultures were grown at 303 K in LB medium containing 0.1 g l^{-1} ampicillin (Acros Organics) and 0.035 g l^{-1} kanamycin (Sigma–Aldrich) to an $OD_{600\,nm}$ of 0.5–0.8, at which point expression was induced by 1 m*M* isopropyl *β*-D-1-thiogalactopyranoside (IPTG; Acros Organics). After 15 h incubation followed by sedimentation of the suspension by centrifugation, the cells



1: Ac-pYVN

2: Xaa = Val (fpYVN)
3: Xaa = Ile (fpYIN)
4: Xaa = Gln (fpYQN)
5: Xaa = Glu (fpYEN)

6: Xaa = Val (cpYVN)
7: Xaa = Ile (cpYIN)
8: Xaa = Gln (cpYQN)

Figure 1
Structures of the native tripeptide Ac-pYVN (**1**), flexible ligands **2**–**5** and analogous constrained ligands **6**–**8** in which the C atoms belonging to the succinyl moiety are labeled 1–4.

Table 1
X-ray diffraction data and refinement statistics of phosphotyrosine-derived flexible ligands **2**–**5** complexed with the Grb2 SH2 domain.

Values in parentheses are for the outer shell.

Data set	Grb2 SH2–**2**	Grb2 SH2–**3**	Grb2 SH2–**4**	Grb2 SH2–**5**
Data collection				
Total reflections†	32135/30927	58613/58278	108054/107744	59636/59525
Unique reflections†	11034/10073	11224/10789	18279/18133	6814/6728
Resolution range (Å)	20.00–1.70 (1.76–1.70)	30.00–1.70 (1.76–1.70)	50.00–1.79 (1.85–1.79)	50.00–2.02 (2.09–2.02)
Completeness (%)	95.6 (70.9)	97.2 (79.7)	94.7 (96.1)	98.4 (100.0)
Multiplicity	2.9 (1.6)	5.2 (2.6)	5.9 (5.9)	8.8 (9.0)
$R_{\rm merge}$‡	0.045 (0.158)	0.054 (0.174)	0.049 (0.084)	0.050 (0.102)
$I/\sigma(I)$	27.7 (4.5)	33.4 (6.8)	40.0 (27.6)	42.0 (27.4)
Crystal				
Space group	$P4_32_12$	$P4_32_12$	$P2_12_12_1$	$P4_32_12$
Unit-cell parameters				
$a = b$ (Å)	42.05	42.03	42.24	41.87
c (Å)	109.43	109.76	110.38	108.81
$\alpha = \beta = \gamma$ (°)	90.0	90.0	90.0	90.0
No. of complexes in asymmetric unit	1	1	2	1
Solvent content (%)	30.4	34.0	35.0	32.9
Matthews coefficient $V_{\rm M}$ (Å^3 Da^{-1})	1.77	1.86	1.89	1.83
Bulk-solvent B factor (Å^2)	44.3	48.9	11.4	16.3
Refinement				
No. of reflections§	10430/555	10605/560	12970/641	6373/350
$R_{\rm cryst}$¶/$R_{\rm free}$ (%)	20.4/23.7	18.9/21.1	20.0/23.7	19.8/23.0
R.m.s. deviation from ideal values				
Bond lengths (Å)	0.0048	0.0046	0.0121	0.0117
Bond angles (°)	1.32	1.33	1.54	1.62
B-factor restraints (Å^2)				
Backbone bonds (r.m.s./σ)	1.29/1.5	1.21/1.5	1.11/1.5	1.11/1.5
Side-chain bonds (r.m.s./σ)	1.79/2.0	1.91/2.0	1.78/2.0	1.77/2.0
Backbone angles (r.m.s./σ)	2.08/2.0	1.93/2.0	1.66/2.0	1.62/2.0
Side-chain angles (r.m.s./σ)	2.65/2.5	2.86/2.5	2.53/2.5	2.48/2.5
Final model				
No. of protein residues	101	100	200	100
No. of protein atoms	832	832	1666	825
No. of ligand atoms	36	37	76	38
No. of water molecules	122	117	117	92
No. of solvent molecules	1	1	5	3
No. of residues in alternate conformations	0	0	1	2

† No. of reflections/No. for which $I/\sigma(I) \geq 1.0$. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the scaled intensity of the ith observation and $\langle I(hkl)\rangle$ is the mean intensity for that reflection. § No. of reflections used in refinement: working set/free R set. ¶ $R_{\rm cryst} = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$, where $F_{\rm calc}$ and $F_{\rm obs}$ are the calculated and observed structure-factor amplitudes, respectively.

were resuspended in 25 m*M* Tris, 1 m*M* EDTA pH 7.5 and lysed using a French press at a pressure of 3.4 MPa. The lysate was centrifuged and the supernatant was purified on a phosphotyrosine affinity column, followed by dialysis of the eluent containing the protein in 25 m*M* Tris, 1 m*M* EDTA pH 7.5. The dialysed protein was further purified on a Q-Sepharose FF column (GE Healthcare).

2.1. Preparation of protein–ligand solutions

Prior to the cultivation of crystals of the domain complexed with each ligand, purified protein was dialyzed twice in 3.5 l distilled water that had been passed through a Nanopure water-purification system (Barnstead) to give a resistivity within the range 17.0–18.2 MΩ cm. The resulting protein solution was placed in 15 ml Centriplus concentrators (Millipore; molecular-weight cutoff 2000) and centrifuged at 3000*g* and 277 K until the protein concentration was within the range 5–15 mg ml^{-1}. Ligands **2**–**8** were each dissolved in this solution to give protein:ligand molar ratios within the range 1:1–1:2. These solutions were each heated to 323 K for 10 min to convert any Grb2 SH2 domain-swapped dimer to the monomeric domain (Benfield *et al.*, 2007), filtered through a 0.45 µm PVDF filter disk and cooled and stored at 277 K. Crystal Screen, Crystal Screen 2 and Crystal Screen Lite (Hampton Research) were used to identify initial crystallization conditions; additional screening was performed if necessary. Crystals were cultivated by the hanging-drop vapor-diffusion method, employing 7 µl drops and 350–400 µl well solution in standard 24-well flat-bottom polystyrene plates. Specific details pertaining to the growth and diffraction of crystals of Grb2 SH2 complexed with each of the ligands are given below.

2.1.1. Grb2 SH2 complexed with 2. An aqueous solution containing 15.0 mg ml^{-1} Grb2 SH2 and 0.92 mg ml^{-1} **2** (protein:ligand molar ratio of 1:1.5) was prepared. This solution (3.5 µl) was mixed with a precipitant solution containing 4.0 *M* sodium formate (3.5 µl) and allowed to equilibrate against the aforementioned precipitant well solution (400 µl) at 298 K. Useable crystals grew within four weeks.

2.1.2. Grb2 SH2 complexed with 3. An aqueous solution containing 15.0 mg ml^{-1} Grb2 SH2 and 0.61 mg ml^{-1} **3** (protein:ligand molar ratio of 1:1) was prepared. This solution

Table 2
X-ray diffraction data and refinement statistics of phosphotyrosine-derived constrained ligands **6–8** complexed with the Grb2 SH2 domain.

Values in parentheses are for the outer shell.

Data set	Grb2 SH2–**6**	Grb2 SH2–**7**	Grb2 SH2–**8**
Data collection			
Total reflections†	39543/36919	29405/26449	89346/87412
Unique reflections†	16569/15394	13832/12774	19967/16579
Resolution range (Å)	20.00–1.90 (1.97–1.90)	50.00–2.02 (2.09–2.02)	50.00–1.83 (1.91–1.83)
Completeness (%)	94.7 (95.5)	96.5 (84.7)	90.1 (50.8)
Multiplicity	2.3	2.1 (1.7)	4.5 (1.7)
$R_{\rm merge}$‡	0.060 (0.270)	0.047 (0.265)	0.054 (0.302)
$I/\sigma(I)$	13.8 (2.7)	21.2 (3.1)	25.2 (2.4)
Crystal			
Space group	$P2_1$	$P2_1$	$P2_12_12_1$
Unit-cell parameters			
a (Å)	31.82	31.60	40.38
b (Å)	85.49	85.51	63.96
c (Å)	41.44	41.64	92.73
$\alpha = \gamma$ (°)	90.0	90.0	90.0
β (°)	92.6	98.5	90.0
No. of complexes in asymmetric unit	2	2	2
Solvent content (%)	40.2	42.5	46.6
Matthews coefficient $V_{\rm M}$ (Å^3 Da^{-1})	2.06	2.14	2.30
Bulk-solvent B factor (Å^2)	62.6	31.6	20.7
Refinement			
No. of reflections§	16538/833	13813/687	14090/709
$R_{\rm cryst}$¶/$R_{\rm free}$ (%)	19.5/23.1	20.0/22.9	22.9/25.0
R.m.s. deviation from ideal values			
Bond lengths (Å)	0.0060	0.0120	0.0122
Bond angles (°)	1.39	1.65	1.66
B-factor restraints (Å^2)			
Backbone bonds (r.m.s./σ)	1.46/1.5	1.34/1.5	1.41/1.5
Side-chain bonds (r.m.s./σ)	1.96/2.0	1.85/2.0	1.88/2.0
Backbone angles (r.m.s./σ)	2.32/2.0	2.07/2.0	2.27/2.0
Side-chain angles (r.m.s./σ)	2.95/2.5	2.64/2.5	2.77/2.5
Final model			
No. of protein residues	199	199	197
No. of protein atoms	1642	1658	1645
No. of ligand atoms	72	74	76
No. of water molecules	151	182	119
No. of solvent molecules	1	0	0
No. of residues in alternate conformations	0	1	2

† No. of reflections/No. for which $I/\sigma(I) \geq 1.0$. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the scaled intensity of the ith observation and $\langle I(hkl)\rangle$ is the mean intensity for that reflection. § No. of reflections used in refinement: working set/free R set. ¶ $R_{\rm cryst} = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm calc}$ and $F_{\rm obs}$ are the calculated and observed structure-factor amplitudes, respectively.

(3.5 µl) was mixed with a precipitant solution containing 0.5 *M* sodium formate (3.5 µl) and allowed to equilibrate against the aforementioned precipitant well solution (400 µl) at 298 K. Useable crystals grew within three weeks.

2.1.3. Grb2 SH2 complexed with 4. An aqueous solution containing 9.1 mg ml^{-1} Grb2 SH2 and 0.87 mg ml^{-1} **4** (protein:ligand molar ratio of 1:2) was prepared. This solution (3.0 µl) was mixed with a precipitant solution containing 0.2 *M* magnesium chloride hexahydrate, 0.1 *M* Tris pH 8.5 and 30%(*w*/*v*) PEG 4000 (4.0 µl; Hampton Research Crystal Screen condition No. 6) and allowed to equilibrate against the aforementioned precipitant solution (350 µl) at 298 K. Useable crystals grew within four weeks.

2.1.4. Grb2 SH2 complexed with 5. An aqueous solution containing 9.5 mg ml^{-1} Grb2 SH2 and 0.82 mg ml^{-1} **5** (protein:ligand molar ratio of 2:1) was prepared. This solution (4.0 µl) was mixed with a precipitant solution containing 0.1 *M* magnesium chloride hexahydrate, 0.1 *M* Tris pH 8.5 and 30%(*w*/*v*) PEG 4000 (3.0 µl) and allowed to equilibrate against the aforementioned precipitant solution (350 µl) at 298 K. Useable crystals grew within four weeks.

2.1.5. Grb2 SH2 complexed with 6. An aqueous solution containing 50.0 mg ml^{-1} Grb2 SH2 and 3.05 mg ml^{-1} **6** (protein:ligand molar ratio of 1:1.5) was prepared. This solution (3.5 µl) was mixed with a precipitant solution containing 50 m*M* sodium cacodylate, 2.0 *M* ammonium phosphate pH 6.0 and 4%(*v*/*v*) PEG 400 (3.5 µl) and allowed to equilibrate against the aforementioned precipitant well solution (400 µl) at 298 K. Useable crystals grew within four weeks.

2.1.6. Grb2 SH2 complexed with 7. An aqueous solution containing 11.2 mg ml^{-1} Grb2 SH2 and 0.90 mg ml^{-1} **7** (ligand:protein molar ratio of 2.0) was prepared. This solution (4.0 µl) was mixed with a precipitant solution containing 0.2 *M* ammonium acetate, 0.1 *M* sodium acetate trihydrate pH 4.6 and 30%(*w*/*v*) PEG 4000 (3.0 µl; Hampton Research Crystal Screen condition No. 10) and allowed to equilibrate against the aforementioned precipitant well solution (350 µl) at 298 K. Useable crystals grew within four weeks.

2.1.7. Grb2 SH2 complexed with 8. An aqueous solution containing 7.6 mg ml^{-1} Grb2 SH2 and 0.53 mg ml^{-1} **8** (protein:ligand molar ratio of 1:1.7) was prepared. Undissolved ligand was removed by filtration (0.45 µm PVDF filter disk) to obtain a homogeneous solution. This solution (3.5 µl) was mixed with a precipitant solution containing 0.1 *M* HEPES pH 7.5 and 20%(*w*/*v*) PEG 10 000 (3.5 µl; Hampton Research Crystal Screen 2 condition No. 38) and allowed to equilibrate against the aforementioned precipitant solution (400 µl) at 298 K. Useable crystals grew within two weeks.

2.2. Collection of X-ray diffraction data

Prior to data collection, crystals were cryoprotected by transferring them to solutions containing salt, buffer and precipitant concentrations equal to their theoretical initial concentrations in the hanging-drop experiment, yet containing sequentially increasing concentrations of glycerol up to the concentration range 20–30%(*v*/*v*). Crystals were allowed to equilibrate in each successive solution for 1–3 min. Once equilibrated in solution containing 20–30%(*v*/*v*) glycerol, the crystals were removed and flash-frozen in liquid nitrogen prior to affixing them to the goniometer. X-ray diffraction data were collected at 100 K using a Rigaku R-AXIS IV area detector positioned on a Rigaku RU-200 rotating-angle X-ray generator operated at 40 kV and 70 mA producing Cu $K\alpha$ graphite-monochromatic radiation (1.5418 nm). Data frames were collected in 0.5° intervals with exposure times of 240–300 s at crystal-to-detector distances of 125–175 mm.

2.3. Data processing and structure refinement

Data frames were processed and scaled using *HKL*-2000 (*XDISPLAY*, *DENZO* and *SCALEPACK*; Otwinowski & Minor, 1997); *CCP*4 (*MOLREP* and *Phaser*; Collaborative Computational Project, Number 4, 1994) was used to identify a molecular-replacement solution from a published Grb2 SH2–peptide crystal structure (Nioche *et al.*, 2002). Structures were refined using the *CNS* suite of programs (Brünger *et al.*, 1998) and structural manipulation was performed using the program *O* (Jones *et al.*, 1991). Ligands containing amino-acid replacements were docked into the protein model and the necessary topology and parameter files were built using the online program *PRODRG* (Schüttelkopf & van Aalten, 2004).

During refinement of the domain complexed with cyclopropane-constrained ligands **6**–**8**, difficulties were encountered in maintaining planarity about the C4—N bond of the N-terminal methylamide moiety (Fig. 1). In extreme cases, the O—C4—N—Me dihedral angle describing rotation about the C4—N bond deviated from planarity by 57.7°, suggesting an almost complete loss of amide resonance. In order to better enforce planarity, two additional parameters were added to the ligand *CNS* topology and parameter files: an additional O—C4—N—Me dihedral angle and an additional C4—O—Me—N improper angle. The inclusion of these additional parameters resulted in O—C4—N—Me dihedral angles that deviated from planarity by no more than 11.9°.

B-factor data were obtained from the output of the *CNS* file `refine.inp` and the resulting values were compared with those obtained from the output of the *CNS* file `bindividual.inp`. Values obtained *via* the latter method yielded lower values than the former method for all atoms of all complexes.

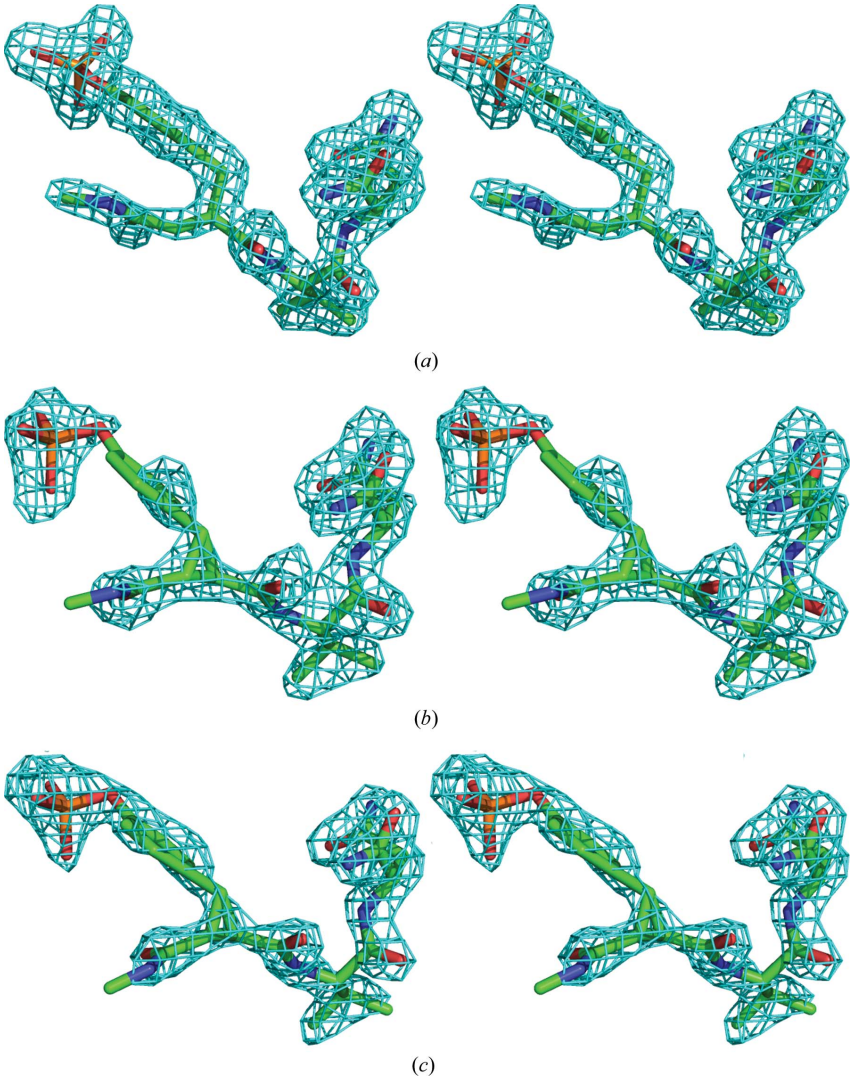


Figure 2
Electron-density difference OMIT maps showing the presence of the bound flexible ligand **2** and the cyclopropane-derived ligand **6** in stereo. The maps indicated by the cyan wire mesh are unweighted $F_o - F_c$ OMIT maps contoured at $+3\sigma$, showing only the portion within 1.0–1.5 Å of each ligand atom in the complexes for clarity. (*a*) Complex of the domain with **2**. (*b*) Complex *a* of the domain with **6**. (*c*) Complex *b* of the domain with **6**.

However, the values obtained from the two methods differed by only 1.0, 0.9, 2.2 and 1.7 Å^2 for all domain atoms, all ligand atoms, all water molecules and those water molecules equivalent to W1–W10, respectively, averaged over all complexes. As such, the original values obtained from `refine.inp` were used in further analysis.

3. Results and discussion

The amino acids comprising the Grb2 SH2 domain are identified according to the nomenclature used by Rahuel *et al.* (1998). X-ray diffraction data and refinement statistics for the Grb2 SH2 domain complexed with the flexible ligands **2**–**5** and the cyclopropane-derived ligands **6**–**8** are presented in Tables 1 and 2, respectively. Data for the complexes of the domain with **2** and **6** were collected to 1.7 and 1.9 Å resolution, respectively, and the structures were solved by molecular replacement using the structure of the complex of the domain with PSpYVNVQN reported by Nioche and coworkers as an initial solution (Nioche *et al.*, 2002; Delorbe *et al.*, 2009). Data for the complexes of **3**–**5** were collected to 1.7, 1.8 and 2.0 Å resolution, respectively, and the structures were solved by molecular replacement using a structure of the complex of the domain with **2** (PDB code 3c7i) as an initial solution. Similarly, data for the complexes of **7** and **8** were collected to 2.0 and 1.7 Å resolution, respectively, and the structures were solved by molecular replacement using the structure of the domain complexed with **6** (PDB code 2huw) as an initial solution. There is one complex in the asymmetric unit for the complexes of **2**, **3** and **5**, whereas there are two complexes in the asymmetric units for the complexes of **4**, **6**, **7** and **8**. At least 90.1% of the available diffraction data was collected for all of the complexes.

Following structure refinement, the ligands were removed and the molecular model was refined to produce $F_o - F_c$ density difference OMIT maps revealing the presence of bound flexible and cyclopropane-derived ligands. Representative maps showing the presence of the bound ligands **2** and **6** are shown in Fig. 2. Ideally, such maps yield wire-mesh surfaces that completely envelope the positions of all atoms of the ligand in the refined molecular models. Although all of the atoms of **2** and **3** in their respective complexes are fully enveloped within this mesh, some atoms in the complexes of **4**–**8** are not. For example, the N-terminal methyl groups are not enveloped in any of the complexes of **4**–**7** or in one of the two coexisting complexes of **8**. The N atoms of the pY replacements in the complex of **5** and in one of the two coexisting complexes for each of **6** and **7** are also not fully within this mesh. Other atoms of the phosphotyrosine (pY) replacement that are not fully enveloped in the complexes of **4**–**8** include the C^{β}–C^{δ} atoms of the phenyl rings and one or more atoms belonging to the pY + 1 amino acid. These are typically the C^{γ} or C^{δ} atoms of the side chain and/or the backbone C^{α} or carbonyl C atoms. Given the high binding affinities of **4**–**8** for the Grb2 SH2 domain (Table 3; Benfield *et al.*, 2006; Delorbe *et al.*, 2009), it seems unlikely that incomplete envelopment of various atoms of the ligands arises from low occupancy in the binding pocket. Rather, it seems more likely that there is greater molecular motion associated with these atoms, thereby rendering their positions and observed electron densities more diffuse.

$2F_o - F_c$ density difference maps of the refined structures were also examined.

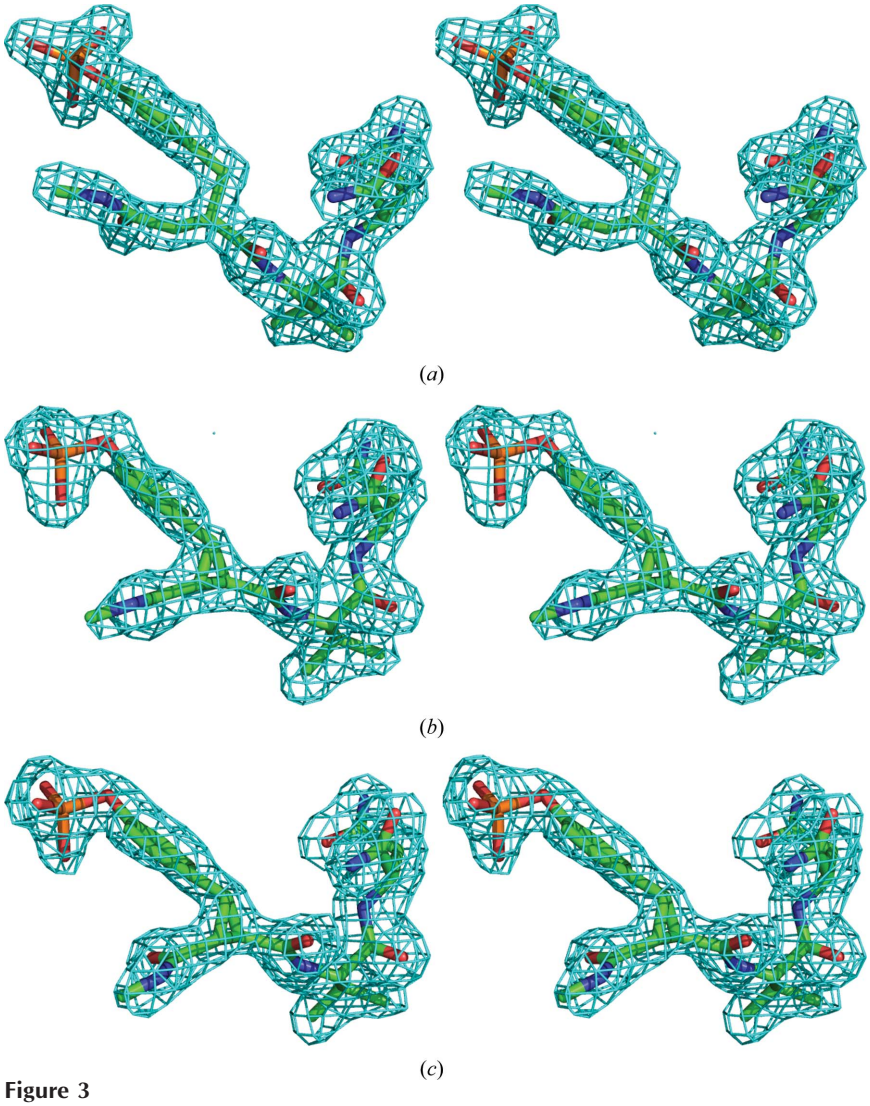


Figure 3
Electron-density difference maps showing the presence of the bound flexible ligand **2** and the cyclopropane-derived ligand **6** in stereo. The maps indicated by the cyan wire mesh are unweighted $2F_o - F_c$ maps contoured at $+1\sigma$, showing only the portion within 1.0–1.5 Å of each ligand atom in the complexes for clarity. (*a*) Complex of the domain with **2**. (*b*) Complex *a* of the domain with **6**. (*c*) Complex *b* of the domain with **6**.

Representative maps showing the presence of the bound ligands **2** and **6** are depicted in Fig. 3. The density difference maps gave wire-mesh surfaces that completely enveloped all atoms of all ligands, with the exception of the N-terminal methyl group in the complex of **5** and in one of the two complexes for each of **4**, **6** and **7**.

3.1. Structural analysis

The structures of complexes of the Grb2 SH2 domain with pairs of flexible and constrained ligands were compared in order to identify any differences that might be associated with changes in ligand preorganization. In particular, we wanted to evaluate whether there are differences in (i) the conformation and orientation of the protein backbones, (ii) protein–ligand pairwise interactions, (iii) networks involving interfacial water molecules, (iv) the *B* factors of the protein backbone and (v) crystal packing that might be correlated with differences in structure.

As a prelude to comparing structures, it is necessary to establish what constitutes a significant difference. For the purpose of this study, we posit that structural differences observed between multiple complexes within the same asymmetric unit provide a useful benchmark. Since the potential energies of coexisting complexes are likely to be within a factor of RT, where R is the gas constant ($kN_{\rm A}$) and T is the temperature (K), they represent nearly isoenergetic structures that vary by no more than $\sim$2.5 kJ mol^{-1} at 298 K. Accordingly, the maximal difference observed for a given structural feature in any set of coexisting complexes will be taken to represent a reasonable lower boundary as to what constitutes a significant dissimilarity.

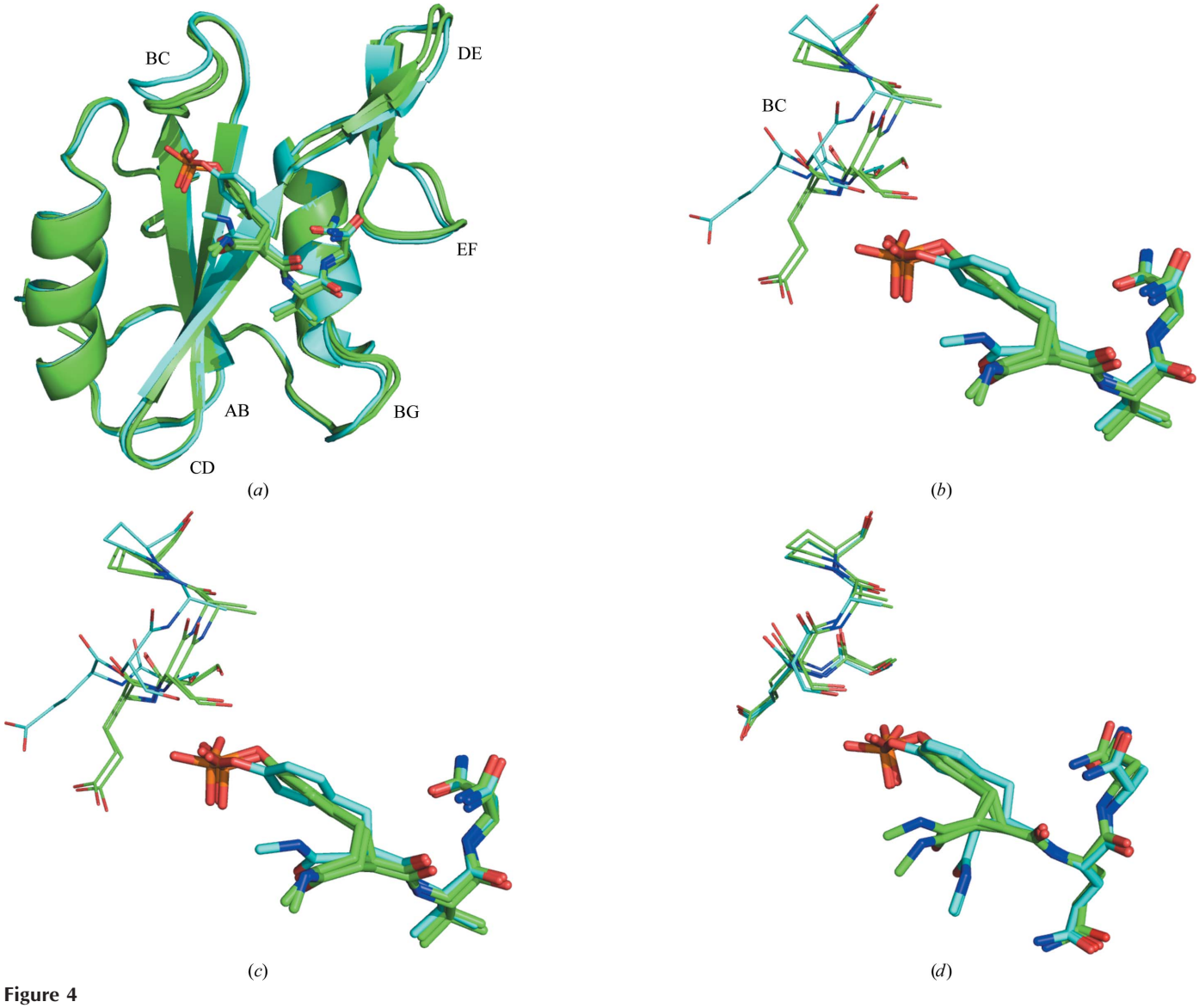


Figure 4
Domain (ProA1–IleβG2) backbone-atom alignment of Grb2 SH2 (lines) complexed with flexible and constrained ligands (sticks). Only residues SerβB7–ProBC4 of the BC loop are shown in (*b*)–(*d*). Atoms are colored red (oxygen), blue (nitrogen), orange (phosphorus) and cyan or green (carbon). (*a*, *b*) Domain alignment of the complex of **2** (cyan) and the two complexes of **6** (green). (*c*) Domain alignment of the complex of **3** (cyan) and the two complexes of **7** (green). (*d*) Domain alignment of the two complexes of **4** (cyan) and the two complexes of **8** (green).

Table 3
Dissociation constants for complexes of the Grb2 SH2 domain with the flexible ligands **2**–**5** and the constrained ligands **6**–**8**.

Ligand	K_d (μ*M*)
2	2.20 ± 0.08
3	2.48 ± 0.13
4	1.78 ± 0.07
5	3.34 ± 0.13
6	0.36 ± 0.02
7	0.47 ± 0.02
8	0.86 ± 0.06

3.1.1. Comparing complexes of flexible/constrained ligand pairs. The backbone atoms of the Grb2 SH2 domain, defined here as residues ProA1–IleβG2, in its complexes with the pairs of flexible and constrained ligands **2/6**, **3/7** and **4/8** were aligned and these alignments will be referred to as domain alignments. Additional expressed residues N- and C-terminal to the ProA1–IleβG2 sequence were omitted from the calculations because the inherent flexibility in these terminal regions results in either the partial or complete absence of electron density for these residues. Inasmuch as there are two complexes in the asymmetric unit for **4**, **6**, **7** and **8**, each ligand pair yielded multiple alignments. The domain alignments of the complexes of flexible and constrained ligand pairs **2/6**, **3/7** and **4/8** yielded root-mean-square (r.m.s.) deviations in the range 0.4–0.6 Å. Since domain alignments of coexisting complexes yielded a maximal r.m.s. deviation of 0.4 Å, the backbone atoms of the domain are displaced significantly in most alignments of flexible and constrained ligands, with the most notable dissimilarities residing in the flexible BC, DE, EF and BG loop regions (Fig. 4*a*). The BC loop region, which comprises residues SerβB7–ProBC4, forms part of the phosphotyrosine-binding pocket. Although the loop formally consists of residues GluBC1–GlyBC5, SerβB7 was included in observations regarding the loop given the large deviations relative to the domain of backbone atoms belonging to this residue in many of the complexes. GlyBC5 was excluded owing to the tendency of this residue to adopt variable ψ angles from one complex to another.

Following domain alignment, the backbone atoms belonging to the BC loop in the two coexisting complexes of **6** are displaced with r.m.s. deviations of 1.3 and 1.0 Å relative to the same loop region in the complex of **2** (Fig. 4*b*). Similarly, the BC loop regions in the two coexisting complexes of **7** are displaced relative to that in **3** with r.m.s. deviations of 0.4 and 0.9 Å (Fig. 4*c*), whereas the relative positions of these loop regions in the complexes of **4** and **8**, each of which have two coexisting complexes in the asymmetric unit, vary with r.m.s. deviations of 0.4, 0.3, 0.4 and 0.3 Å (Fig. 4*d*). Because the maximal r.m.s. deviation of the BC loop region in coexisting complexes is 0.6 Å, the relative orientations of the BC loop in both flexible/constrained ligand pairs **2/6** and in one of the two pairs **3/7** are considered to be significant, whereas the variations in the relative positions of the BC loop regions in the complexes of the flexible/constrained ligand pair **4/8** are not. Alignment of backbone atoms belonging only to the BC loop region, referred to here as local alignment, of the complexes of flexible/constrained ligand pairs **2/6**, **3/7** and **4/8** yielded r.m.s. deviations of $\leq$0.3 Å. Because these values are within the experimental error limit of the crystallographic data, there are no significant variations in the local conformations of the backbone atoms of the BC loop region in any of the complexes. Hence, although the BC loop regions in the complexes of flexible/constrained ligand pairs may be displaced significantly from one another relative to the domain, their backbone conformations are similar.

Table 4
R.m.s. deviations of loop regions and full domain following alignment of all backbone atoms of the protein in complexes of flexible/constrained ligand pairs **2/6**, **3/7** and **4/8**.

	Displacements (Å)			
Loop region†	**2/6**	**3/7**	**4/8**	MDCC‡
AB	0.2–0.4	0.2–0.7	0.4–0.6	0.6
BC	1.0–1.3	0.4–0.9	0.3–0.4	0.6
CD	0.2–0.4	0.3–1.0§	0.3–0.5	1.1§
DE	0.9–1.1	0.8–0.9	0.7–0.9	0.7
EF	0.7	0.5–0.7	0.4–0.6	0.3
BG	0.7–0.9	0.6–0.9	0.7–1.3¶	1.3¶
Domain	0.5–0.6	0.5	0.4–0.5	0.4

† Loop regions were defined as follows: AB, ArgAB3–AspAB5; BC, SerβB7–ProBC4; CD, GlyCD1–AspβD1; DE, AspβD′3–GlyDE3; EF, LeuβE4–LysβF2; BG, SerBG3–GlnBG7. ‡ Maximal displacement in coexisting complexes (MDCC). § These displacements are the result of >135° differences in the ψ angle of GlyCD1 and are not a consequence of differences in the overall position of the backbones. ¶ The large maximal displacement of the BG loop is a consequence of the relative position of this loop in one of the coexisting complexes of **8**.

Inasmuch as there are some significant variations in the relative positions of the backbone atoms in the BC loop region, one might anticipate that there would be corresponding differences in the orientations of the amino-acid side chains in this loop in order to maintain polar contacts between the domain and the phosphate groups of the ligands. Following local alignment of the BC loop region in the complexes of ligand pairs **2/6**, **3/7** and **4/8**, all side chains belonging to the BC loop residues align closely, with the exception of the GluBC1 side chain in the complexes with ligand pairs **3/7** and **4/8** (Fig. 5). These displacements appear to result from φ, ψ and χ_1 angles that differ maximally by 21°, 14° and 9°, respectively.

Displacements of other loop regions following domain alignment of the complexes of the flexible/constrained ligand pairs **2/6**, **3/7** and **4/8** together with the maximal displacements of these loops in coexisting complexes are given in Table 4. Examination of these data reveal that variations in the DE and EF loop regions exceed the maximal displacements for the corresponding loops in coexisting complexes in most cases, so these variations are considered to be significant. Although the DE loop region does not contact the ligand in any of the complexes, the LeuβB4 backbone carbonyl O atom of the EF loop region makes a direct contact to the pY + 2 Gln side-chain amide N atom of the ligand in all complexes. Moreover, residues LeuβB4 and TrpEF1 of the loop make hydrophobic interactions with the pY + 1 and pY + 2 amino acids of the ligand in all complexes.

Following domain alignment of the complexes of ligand pairs **2/6**, **3/7** and **4/8**, backbone atoms belonging to the EF loop region are displaced by 0.4–0.7 Å, which exceeds the

maximal displacement of 0.3 Å in coexisting complexes. In contrast, local alignment of the loop regions leads to deviations of ≤0.3 Å. As such, the EF loop regions in the complexes of flexible/constrained ligand pairs are displaced significantly from one another relative to the domain in all cases, yet their backbone conformations are nearly identical. The LeuβB4 and TrpEF1 side chains, which make hydrophobic interactions with the bound ligand, are displaced 0.3–0.9 Å in the complexes with ligand pairs **2**/**6**, **3**/**7** and **4**/**8** following local alignment of the EF loop region. However, since the side chains are displaced maximally by 0.6 Å in the coexisting complexes with **7**, these differences are considered to be significant in only one of the two complexes with each of **6** and **7**. Hence, side-chain variations in the complexes of flexible/constrained ligand pairs are generally less meaningful in the EF loop region than in GluBC1 of the BC loop region.

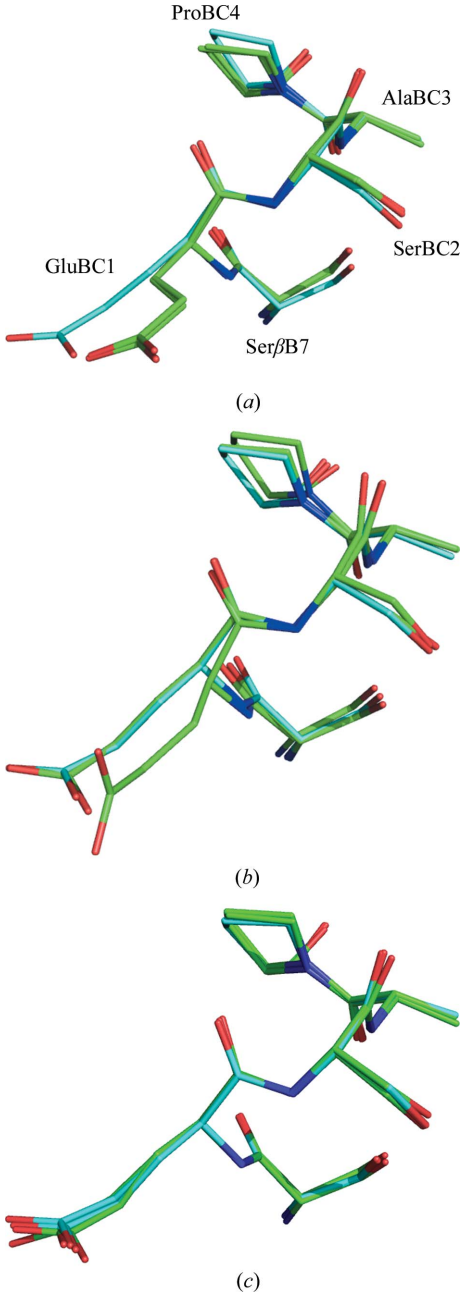


Figure 5
Backbone and side-chain atoms (lines) of the BC loop region following local alignment. Atoms are colored red (oxygen), blue (nitrogen), orange (phosphorus) and cyan or green (carbon). Alignment of the complexes with flexible/constrained ligand pairs (*a*) **2** (cyan) and **6** (green), (*b*) **3** (cyan) and **7** (green) and (*c*) **4** (cyan) and **8** (green).

A comparison of the three-dimensional structures of the bound ligands themselves is also important. All of the ligands **2**–**8** bind to the Grb2 SH2 domain in a β-turn configuration that is stabilized by an intramolecular hydrogen bond between the C-terminal amide N atom and the backbone carbonyl O atom of the pY replacement. All-atom alignments of the bound ligands in the complexes of the flexible/constrained pairs **2**/**6**, **3**/**7** and **4**/**8** give deviations in the range 0.6–1.1 Å. However, alignment of those atoms belonging only to the pY + 1 and pY + 2 amino acids and the C-terminal amide, referred to here as the *X*–N region, yields r.m.s. deviations of ≤0.2 Å. Moreover, the P atoms in the phosphate groups of **2**–**8** align with r.m.s. deviations of ≤0.2 Å. Given that structures that align with r.m.s. deviations <0.25 Å are typically considered to be identical by conformational searching and molecular-dynamics programs (*MacroModel* v.9.1; Schrödinger, New York, USA), these amino acids are considered to adopt identical conformations in all complexes. The significant differences in the bound conformations of flexible and constrained ligands are in the positions of atoms belonging to the *N*-methylamide group and the pY replacement, referred to here as the Ac-pY region (see Fig. 4). All-atom alignments of the Ac-pY regions of the pairs of flexible and constrained ligands in the various complexes give r.m.s. deviations in the range 0.8–1.2 Å. These larger deviations are the result of different atom types in the flexible and constrained pY replacements and variations in the orientation of the N-terminal methylamide moieties, particularly in the complexes of **4** and **8**. Since the Ac-pY region differs maximally in the coexisting complexes by only 0.3 Å, these deviations between flexible and constrained ligands are significant but not unexpected owing to the specific nature of the pY replacements. The large differences in the orientation of the *N*-methylamide moieties between the complexes of **4** and **8** arise because the O atom of the *N*-methylamide group of **4** is oriented so that it can make a single water-mediated contact with the side chain of the pY + 1 Gln amino acid; a similar conformation is not possible for **8** because the *N*-methylamide moiety is directly attached to the cyclopropane ring and the relevant rotor is completely restricted.

3.1.2. Comparing sets of complexes of flexible and constrained ligands. Structural features specific to the set of complexes of flexible ligands **2**–**5** and the set of complexes of constrained ligands **6**–**8** were then considered. Differences in the set of complexes with flexible ligands **2**–**5** were evaluated by domain alignment of each with the complex of **2**, thereby giving four pairwise alignments. Differences in the set of complexes with the cyclopropane-derived ligands **6**–**8** were similarly evaluated by domain alignment of each with the *a* complex of the domain with **6**, thereby giving five pairwise alignments. Following these alignments, backbone atoms belonging to the BC loop region are displaced ≤0.3 Å in the

set of complexes of **2**–**5** (Fig. 6*a*) and 0.4–1.3 Å in the set of complexes with **6**–**8** (Fig. 6*b*). Since the BC loop region is maximally displaced by 0.6 Å following domain alignment of the coexisting complex with **7**, displacements of the loop region in the set of flexible ligands are not significant following any of the four pairwise alignments. However, displacements of the BC loop region in the set of constrained ligands **6**–**8** are meaningful for three of the five pairwise alignments, these being with one of the two complexes with **7** and both of the complexes with **8**. Hence, there is greater variation in the position of the BC loop region relative to the domain from one complex to another in the set of constrained ligands relative to the set of flexible ligands.

Differences in the EF and BG loop regions were evaluated because these regions are involved in interactions with the ligand. The EF loop region is displaced by $\leq$0.3 Å following all pairwise domain alignments in the set of complexes with flexible and constrained ligands (Fig. 6). Hence, based upon the established metric, variations in this loop region are not judged to be significant in either set of complexes. The BG loop region is displaced by $\leq$0.3 Å following the four pairwise domain alignments in the set of complexes with flexible ligands **2**–**5** and by 0.2–1.2 Å following the five pairwise domain alignments in the set of complexes with constrained ligands **6**–**8**. While the latter variations are larger than the former, they do not appear to be meaningful because the BG loop is displaced maximally by 1.3 Å following domain alignment of the coexisting complexes with **8**. Therefore, while differences in the EF and BG loop regions in the set of complexes with constrained ligands **6**–**8** were larger than those in the set of flexible ligands **2**–**5**, none of these displacements are regarded as significant.

3.2. Direct polar contacts

There are ten direct polar contacts that are conserved in all complexes. Of these, five involve the phosphate moiety of the Ac-pY replacement, one involves the N-terminal amide O atom of the Ac-pY replacement and four involve the *X*–N region of the ligand. Four additional direct contacts are partially conserved and appear in the majority of complexes and these are (i) a contact between the side chain of ArgαA2 and the N-terminal amide O atom of the Ac-pY replacement that is observed in all complexes except those of **4**; (ii) a contact between the side chain of ArgβB5 and a nonbridging phosphate O atom that is observed in all complexes except those of **4**, **5** and one of the two coexisting complexes of **7**; (iii) a contact between the side chain of ArgβB5 and a nonbridging phosphate O atom that is observed in both complexes of **4** and all complexes of constrained ligands **6**–**8** and (iv) a contact between the side chain of SerβC3 and the bridging O atom of the phosphate moiety that is observed in all complexes with constrained ligands **6**–**8**, yet is absent in all complexes with flexible ligands **2**–**5**.

The cyclopropane-derived ligands **6**–**8** make 1–3 more direct contacts to the Grb2 SH2 domain on average than their more flexible counterparts **2**–**5**. These additional contacts arise because there are more interactions between the domain and the Ac-pY replacement in **6**–**8** than between the corresponding replacement in **2**–**5** and the domain. However, because the Ac-pY replacement of **7** makes three more direct contacts with the domain in one coexisting complex than in the other, one cannot make a compelling case that the Ac-pY replacement itself leads to a significant difference in the number of direct contacts with the domain. It is perhaps noteworthy that direct contacts in some complexes are replaced with single water-mediated interactions in others. For example, direct contacts between differing nonbridging phosphate O atoms and the backbone amide N atoms of GluBC1 and SerBC2 present in one of the complexes of **7** are replaced with a single water-mediated interaction in the other complex.

3.3. Water-mediated contacts

Interfacial water molecules play important roles in protein–ligand interactions. In our analysis of single-water-mediated

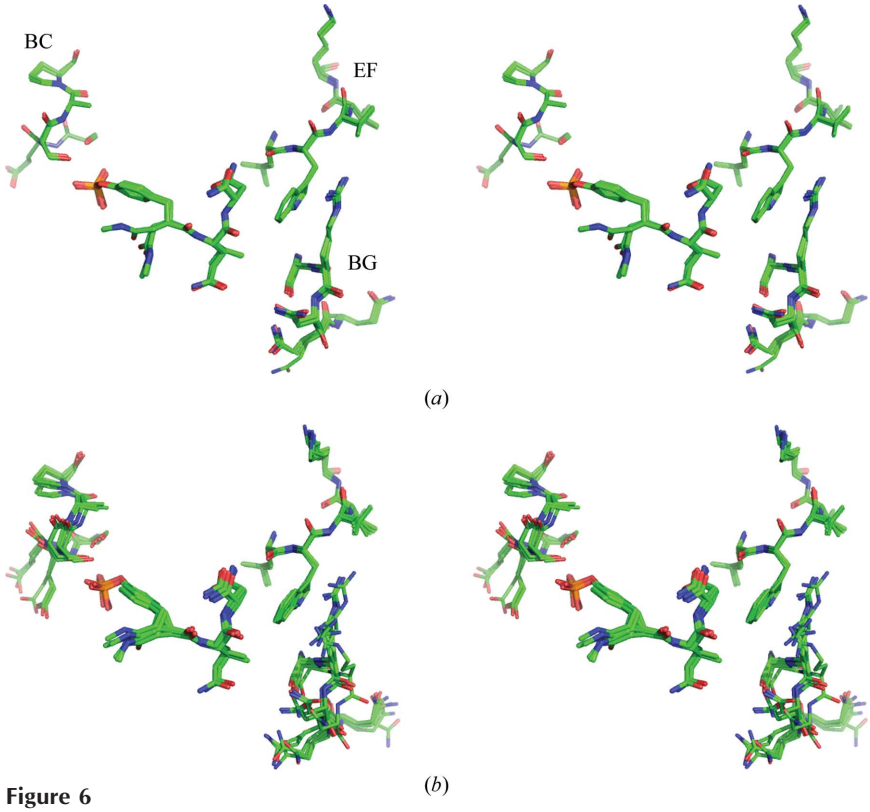


Figure 6
BC, EF and BG loop regions following domain alignment of the set of complexes with flexible ligands **2**–**5** (*a*) and constrained ligands **6**–**8** (*b*) in stereo. Ligands are shown as sticks and loop residues as lines. Atoms are colored red (oxygen), blue (nitrogen), orange (phosphorus) and green (carbon).

contacts, only those water molecules with sufficient electron density ($2F_o - F_c$ density difference mesh contoured at 1σ) were considered. The ordered water molecules depicted in Fig. 7 are labeled W1–W10 to indicate those that are equivalent across the complexes of **2**–**8**. We adopted the criteria used by Boström *et al.* (2006) to define equivalent ordered water molecules. Briefly, ordered water molecules in different complexes must participate in at least one identical protein–ligand interaction and must reside <2.0 Å from one another following domain alignment.

The interfacial ordered water molecules W1, W2 and W4 mediate contacts between domain residues GluBC1, SerBC2 and SerβC3, respectively, and the phosphate moiety of the pY replacement. W1 is conserved in all complexes of **2**–**8** except for one of the two complexes with **6** and **7**. In these two complexes the BC loop is ∼2.0 Å closer to the phosphate moiety than in the other complexes, so there is insufficient space to accommodate an interfacial water molecule in these complexes. W2 is conserved in all but one of the complexes of the flexible ligands **2**–**5**, whereas W4 is conserved in all but one of the complexes of the constrained ligands **6**–**8**. Therefore, introduction of the constraint appears to result in the loss of a protein–ligand interaction mediated by W2 that is balanced by the concomitant gain of a contact that is mediated by W4. W5 mediates contacts between HisβD4 and the N-terminal carbonyl O atom of the Ac-pY replacement and with the polar moiety of the pY + 1 side chain in both complexes of **4** and the complex with **5**; W5 also mediates the latter contact in both complexes with **8**. W8 mediates a contact between SerBG3 and the pY + 1 side chain in both complexes of **4**, the complex of **5** and one of the two complexes of **8**. W9 mediates a contact between AsnBG5 and the pY + 1 side chain in both complexes of **4** and the complex of **5**. It is noteworthy that although ligands **2**, **3**, **6** and **7** are incapable of making contacts mediated by W5, W8 and W9, ordered water molecules equivalent to W5, W8 and W9 appear in most of the complexes of **2**–**8**, suggesting that the identity of the pY + 1 amino acid has little effect on the local water network. Similarly, W10 mediates a contact between AsnBG5 and the pY + 1 carbonyl O atom in only one of the complexes of **2**–**8**, yet an ordered water molecule equivalent to W10 is present in all complexes with flexible ligands **2**–**5**.

In summary, the succinyl-derived Ac-pY replacement of **2**–**5** makes one or two more water-mediated contacts with the domain than the cyclopropane-derived Ac-pY replacement of **6**–**8**; there are also one or two more water-mediated contacts between the domain and the *X*–N region of **2**–**5** than with the corresponding region of **6**–**8**. The flexible ligands thus make 2–4 more water-mediated contacts to the domain than the constrained ligands. Because the maximal difference in the number of single-water-mediated contacts in the coexisting complexes is two, the difference in the number of water-mediated contacts between the complexes of flexible and constrained ligands appears to be significant in all instances except perhaps that between the complex of **2** and one of the complexes of **6**. Thus, while the Ac-pY replacement of the constrained ligands typically make more direct contacts to the domain, the flexible ligands make more single-water-mediated contacts.

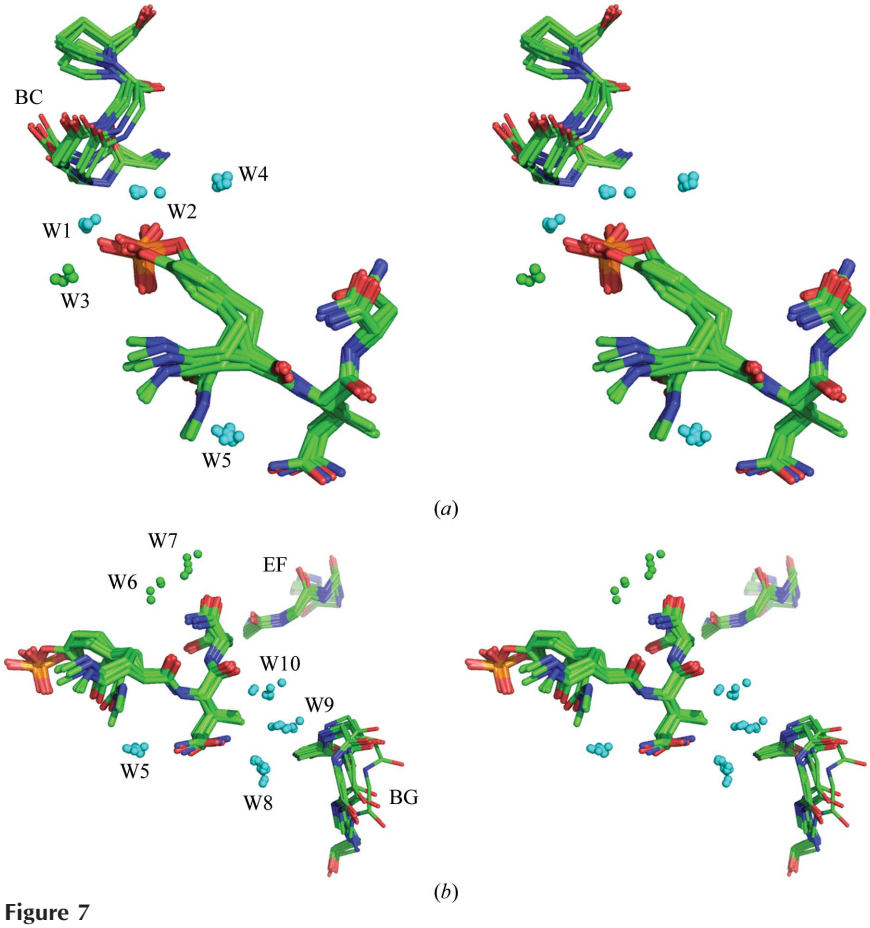


Figure 7
Protein–ligand contacts mediated by W1–W10 in the complexes with **2**–**8**, showing the BC, EF and BG loop regions (lines), the ligands (sticks) and the ordered interfacial water molecules (spheres) in stereo. Only the backbone atoms of residues in the loop regions are shown for clarity. Protein and ligand atoms are colored red (oxygen), blue (nitrogen), orange (phosphorus) and green (carbon). Ordered water molecules colored cyan are either conserved in at least eight of the 11 complexes of **2**–**8**, four of the five complexes with flexible ligands **2**–**5** or five of the six complexes with constrained ligands **6**–**8**. Ordered water molecules that do not meet the aforementioned criteria are colored green. (*a*) W1–W5 mediate contacts between the BC loop region and the Ac-pY replacement. (*b*) W5–W10 mediate contacts between the EF and BG loop regions and the *X*–N region of the ligand.

3.4. Analysis of protein–ligand contacts

Polar interactions between the Grb2 SH2 domain and the bound ligands **2**–**8** were analyzed by constructing contact diagrams such as those depicted in Fig. 8 for the complex of **5**, in which contacts between polar non-H atoms of the protein and ligand

are illustrated (Benfield *et al.*, 2006). Only contacts with non-H atom–atom distances in the range 2.5–3.4 Å are included in this analysis. Contacts mediated by a single ordered water molecule conforming to the same distance criteria are also included. Owing to the differences in the resolution of the crystallographic data for the various complexes, polar interactions were not further characterized based upon their measured contact distances or angles.

3.5. Hydrophobic interactions

Residues PheβD5, LysβD6, LeuβD′1, LeuβE4 and TrpEF1 of the Grb2 SH2 domain are involved in hydrophobic interactions with **2**–**8** that vary between 3.3 and 4.5 Å. PheβD5, LeuβD′1, LeuβE4 and TrpEF1 make contacts with the *X*–N region of the ligand that vary by less than 0.3 Å from one complex to another; these minor differences are not regarded as meaningful. LysβD6 makes contacts to both the phenyl ring of the Ac-pY region and the pY + 2 side chain of the ligands in all complexes that vary maximally by 0.8 and 0.6 Å, respectively, but these variations are insignificant relative to the differences observed in coexisting complexes. LysβD6 also makes hydrophobic interactions with the C-terminal amide N atom and carbonyl O atoms in all complexes with the flexible ligands **2**–**5**, but with the sole exception of one of the two coexisting complexes of **8** it does not engage in such contacts with the constrained ligands **6**–**8**. However, LysβD6 makes additional and/or closer contacts to the pY replacement in these complexes such that differences in hydrophobic interactions between the complexes of flexible and constrained ligands do not appear to be significant.

3.6. Analysis of atomic *B* factors

Comparisons of the atomic *B* factors, or Debye–Waller factors, arising from structure refinements of X-ray crystallographic data have been used as an indicator of the relative molecular motion in the crystalline state (Frauenfelder, 1989). The extraction of meaningful information from such analyses has historically been met with some skepticism because *B* factors vary with crystal lattice disorder as well as a number of other parameters that cannot be controlled (Frauenfelder & Petsko, 1980). Comparison of *B* factors from independently determined structures presents an even greater challenge. Nevertheless, it is possible to place two different sets of mean-square displacements belonging to different structures on an approximately identical scale by adjusting the data set possessing the highest overall displacement such that its lowest value corresponds to the lowest value in the second structure (Frauenfelder & Petsko, 1980). The adjusted *B*-factor data for all complexes are presented in Table 5.

The maximal difference in the average adjusted *B* factors of the backbone atoms of the domains in the coexisting complexes of **8** is 3.9 Å^2, so differences that exceed this value are considered to be meaningful by our criterion. With the sole exception of one of the two complexes of **6**, the average adjusted *B* factors of the backbone atoms of the domains are thus significantly greater in the complexes of the constrained ligands **6**–**8** than in the corresponding complexes of the flexible ligands **2**–**5**. The average adjusted *B* factors of the backbone atoms in the AB, BC, EF and BG loop regions are also typically greater in the complexes of **6**–**8** than in the complexes of **2**–**5** and the average adjusted *B* factors of these loop regions in the complexes of **2**–**5** are not significantly greater than their domain averages. Although the average adjusted *B* factors of the CD and DE loop regions are significantly greater than the domain averages in most of the complexes, these *B* factors do

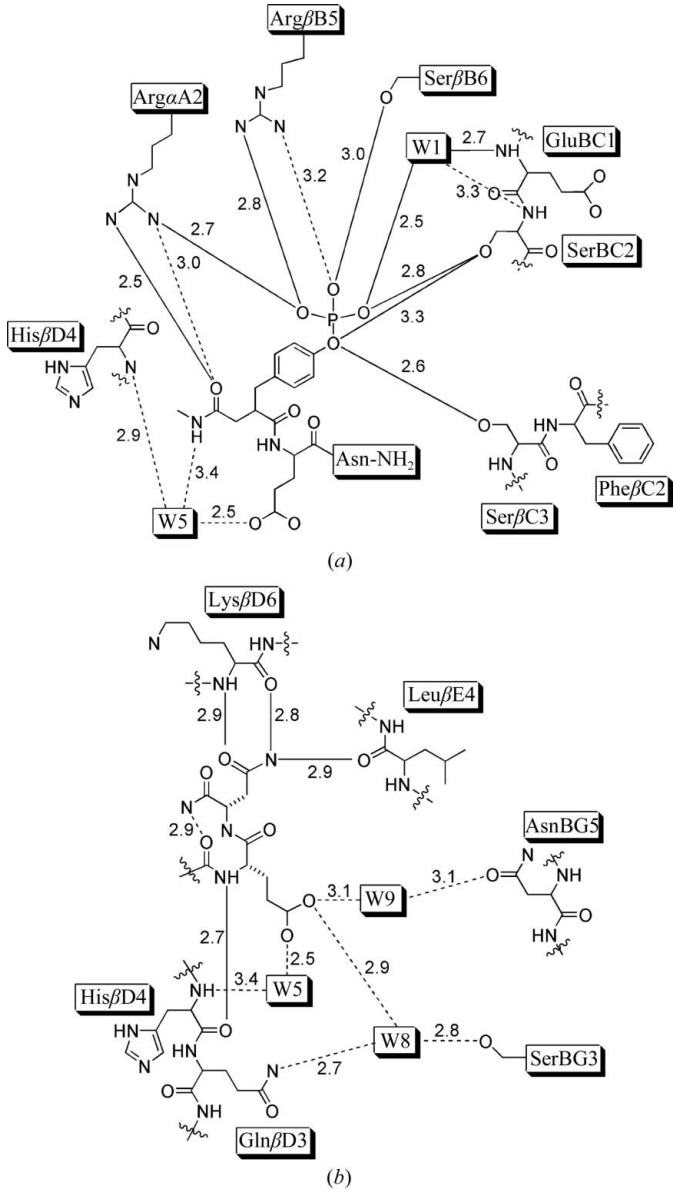


Figure 8
Polar interactions in the complex of **5** with the Grb2 SH2 domain. All labile H atoms have been omitted for clarity except those on protein backbone N atoms. Only those ordered water molecules that mediate a single protein–ligand interaction are shown and these are numbered so that water molecules that are conserved in at least two complexes have the same number. Solid lines in (*a*) and (*b*) indicate those polar contacts that are conserved for all complexes. (*a*) Interactions between the domain and the Ac-pY replacement. (*b*) Interactions between the domain and the *X*–N region of the ligand.

Table 5
Average adjusted *B* factors of the backbone atoms of the domain and loop regions of all complexes.

Values in bold refer to average adjusted *B* factors of loop regions that are judged to be significantly greater in the complexes of constrained ligands than in the complexes of the corresponding flexible ligands. Values in parentheses refer to average adjusted *B* factors of loop regions that are considered to be significantly greater relative to the domain.

	Average adjusted *B* factors (Å^2)						
		Loop regions					
Complex	Domain	AB	BC	CD	DE	EF	BG
2	22.9	24.6	23.3	(35.3)	26.3	20.9	22.4
6a	**27.3**	**28.8**	**31.0**	(33.4)	**(34.1)**	22.6	**(37.3)**
6b	25.7	**(34.7)**	**29.5**	(32.0)	26.2	**25.1**	**26.5**
3	21.8	23.0	21.9	(32.2)	(25.8)	20.0	21.9
7a	**26.2**	26.2	**(33.3)**	28.6	**(32.9)**	**25.2**	**(32.6)**
7b	**25.9**	**(32.3)**	**(31.8)**	**(36.7)**	24.9	**25.4**	**27.5**
4a	22.9	23.8	22.2	(30.2)	(27.3)	21.3	20.8
4b	23.1	23.6	22.9	(30.6)	(28.3)	21.1	20.5
8a	**28.4**	**30.4**	**(39.1)**	(33.4)	24.5	**25.6**	**31.3**
8b	**32.3**	**31.3**	**33.5**	27.9	**(37.5)**	**(37.6)**	**(43.8)**
5	22.9	24.3	24.2	(29.2)	(28.9)	21.4	22.3

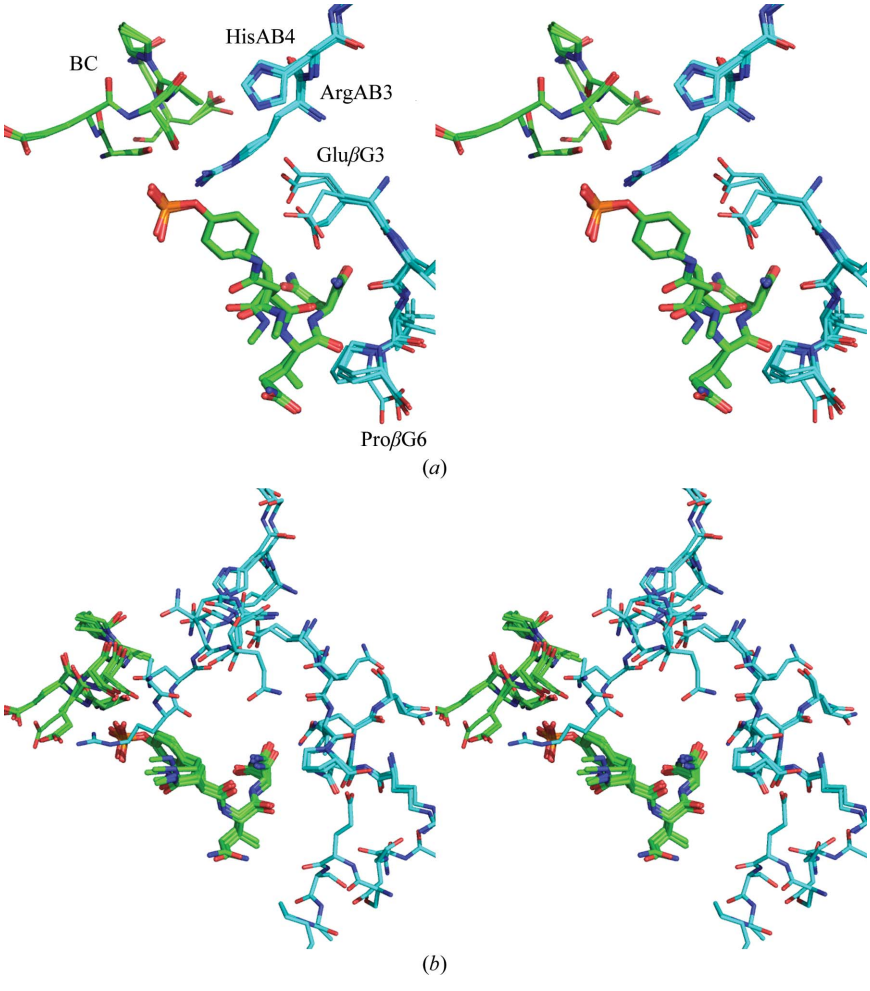


Figure 9
Domain alignments of complexes showing the BC loop region (lines) and the ligands (sticks) of each complex in the asymmetric unit and residues of symmetry-related complexes (lines). Atoms are colored red (oxygen), blue (nitrogen), orange (phosphorus) and cyan or green (carbon). C atoms of symmetry-related residues are colored cyan (*a*) The complexes of **2**–**5**. (*b*) The complexes of **6**–**8**.

not vary meaningfully between the complexes of flexible/constrained ligand pairs **2**/**6**, **3**/**7** and **4**/**8**.

3.7. Crystal packing

The orientation and position of neighboring symmetry-related complexes can influence the structure of complexes within the asymmetric unit, so we queried whether any of the observed structural differences in these complexes might be correlated with dissimilarities in crystal packing. Following pairwise domain alignment of the complex of **2** with each of the complexes of the flexible ligands **3**–**5**, crystal packing in the complexes is nearly identical (Fig. 9*a*). In particular, the side chains of ArgAB3 and HisAB4 in a symmetry-related complex make polar contacts to the phosphate moiety of the ligand and the BC loop region, respectively, in all of the complexes of **2**–**5**. The side chains of GluβG3 and ProβG6 in a different symmetry-related complex also make hydrophobic interaction contacts to the pY + 1 and the Ac-pY replacement of the ligand, respectively, in all of the complexes of **2**–**5**. Thus, for the series of flexible ligands both the crystal packing and the structures of the complexes are similar. The situation is markedly different for the complexes of the constrained ligands **6**–**8**. Pairwise domain alignment of the *a* complex of **6** with each of the remaining complexes of **6**–**8** reveal few similarities in crystal packing and none of the aforementioned contacts between side chains in symmetry-related complexes and the BC loop or the ligand are present (Fig. 9*b*). Furthermore, the structures of the two complexes of **8** in the asymmetric unit are similar, but the crystal packing in these complexes is significantly different. On the other hand, the relative positions of the BC loop regions in the *a* complexes of **6** and **7** are significantly different, even though the crystal packing is similar. We thus find that variations in the structures of complexes of constrained ligands cannot be directly correlated with crystal packing. Comparing the complexes of flexible ligands with those of their constrained

counterparts reveals that there are invariably differences in crystal packing. Accordingly, it is possible that some of the structural dissimilarities observed upon comparing flexible and constrained ligands in this study arise from variations in crystal packing.

4. Summary and conclusions

Structures of the Grb2 SH2 domain complexed with a series of pseudopeptides containing flexible (benzyl succinate) and constrained (aryl cyclopropanedicarboxylate) replacements of the phosphotyrosine (pY) in tripeptides derived from Ac-pYXN-NH_2 (X = V, I, E and Q) were elucidated by X-ray crystallography. Structural variations were judged to be significant based on whether or not they exceed maximal variations between complexes coexisting within the same asymmetric unit. Excepting the loop regions, the backbone of the Grb2 SH2 domain adopts a similar conformation and orientation relative to the bound ligands in all complexes. Although the backbones of the loop regions also adopt similar conformations, they do vary in their relative positions and consistently greater variations are observed in the set of complexes of constrained ligands **6**–**8** than in the set of complexes of flexible ligands **2**–**5**. Only the BC loop region, which contacts the Ac-pY replacement of the ligands, differs significantly relative to the domain and bound ligands following alignment of the complexes of flexible/constrained ligand pairs **2**/**6**, **3**/**7** and **4**/**8**. Accordingly, the effects of introducing a cyclopropyl constraint into the flexible ligands **2**–**4** are most notable in the BC loop region of the resulting complexes.

The constrained ligands **6**–**8** typically make more direct contacts with the Grb2 SH2 domain than do their more flexible counterparts **2**–**4**; however, the observed differences are comparable to those seen in the coexisting complexes of **7**, so they may not be meaningful. On the other hand, the succinate-derived ligands make 2–4 additional single water-mediated contacts to the domain relative to their constrained counterparts. The nature of the pY + 1 side chain has little effect on the surrounding water network in the complexes studied, even though compounds **2**, **3**, **6** and **7** lack polar functionality that can participate in water-mediated contacts. The differences in hydrophobic interactions between the complexes of flexible/constrained ligand pairs **2**/**6**, **3**/**7** and **4**/**8** were generally similar to those observed upon comparing such contacts in coexisting complexes, so significant variations in these contacts do not appear to arise from constraining the phosphotyrosine replacement in **2**–**4**. The average adjusted B factors of the backbone atoms of the domain and loop regions are significantly greater in the complexes of constrained ligands than in the complexes of the analogous flexible ligands, suggesting greater thermal motion in the crystalline state in the former complexes. Although there is no direct correlation between crystal packing and the structures of the complexes of flexible and constrained ligands, the possibility that some structural dissimilarities in these complexes arise from variations in crystal packing cannot be excluded.

5. PDB codes and supplementary data

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 3c7i, 3in8, 3imd, and 3kfj for complexes of the Grb2 SH2 domain with flexible ligands **2**, **3**, **4** and **5**, respectively, and 2huw, 3imj and 3in7 for complexes of the Grb2 SH2 domain with the cyclopropyl-constrained ligands **6**, **7** and **8**, respectively. $F_o - F_c$ density difference OMIT maps and $2F_o - F_c$ density difference maps for **3**–**5**, **7** and **8**, protein–ligand contact diagrams for the complexes of **2**–**4** and **6**–**8** and Ramachandran plots for the complexes of **2**–**8** have been deposited as supplementary material[1].

We thank the National Institutes of Health (GM 84965), the National Science Foundation (CHE 0750329), the Robert A. Welch Foundation (F-652), the Norman Hackerman Advanced Research Program, the Texas Institute for Drug and Diagnostic Development and Welch Foundation Grant #H-F-0032 for their generous support of this research. We are also thankful to Dr Stuart Black (Schering–Plough Research Institute) for the Grb2 SH2 expression vector and helpful discussions. We are also grateful to undergraduates Naveed Nosrati and Nicole Pulliam for their help in expressing and purifying protein and setting up drops for crystallization studies.

References

Benfield, A. P., Teresk, M. G., Plake, H. R., DeLorbe, J. E., Millspaugh, L. E. & Martin, S. F. (2006). *Angew. Chem. Int. Ed.* **45**, 6830–6835.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M. (1977). *J. Mol. Biol.* **122**, 535–542.

Benfield, A., Widdon, B. & Clements, J. H. (2007). *Arch. Biochem. Biophys.* **462**, 47–53.

Boström, J., Hogner, A. & Schmitt, S. (2006). *J. Med. Chem.* **49**, 6716–6725.

Bradshaw, J. M. & Waksman, G. (2002). *Adv. Protein Chem.* **61**, 161–210.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chardin, P., Cussac, D., Maignan, S. & Ducruix, A. (1995). *FEBS Lett.* **369**, 47–51.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Davidson, J. P., Lubman, O., Rose, T., Waksman, G. & Martin, S. F. (2002). *J. Am. Chem. Soc.* **124**, 205–215.

Delorbe, J. E., Clements, J. H., Teresk, M. G., Benfield, A. P., Plake, H. R., Millspaugh, L. E. & Martin, S. F. (2009). *J. Am. Chem. Soc.* **131**, 16758–16770.

Frauenfelder, H. (1989). *Int. J. Quantum Chem.* **35**, 711–715.

Frauenfelder, H. & Petsko, G. A. (1980). *Biophys. J.* **32**, 465–478.

Gerhard, U., Searle, M. S. & Williams, D. H. (1993). *Bioorg. Med. Chem. Lett.* **3**, 803–808.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: XB5018). Services for accessing this material are described at the back of the journal.

Kahn, A. R., Parrish, J. C., Fraser, M. E., Smith, W. W., Barlett, P. A. & James, M. N. G. (1998). *Biochemistry*, **37**, 16839–16845.
Kessels, H. W., Ward, A. C. & Schumacher, T. N. M. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 8524–8529.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Machida, K. & Mayer, B. J. (2005). *Biochim. Biophys. Acta*, **1747**, 1–25.
Mann, A. (2003). *The Practice of Medicinal Chemistry*, edited by C. G. Wermuth, pp. 233–250. London: Academic Press.
McNemar, C., Snow, M. E., Windsor, W. T., Prongay, A., Mui, P., Zhang, R., Durkin, J., Le, H. V. & Weber, P. C. (1997). *Biochemistry*, **36**, 10006–10014.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.
Nakanishi, H. & Kahn, M. (2003). *The Practice of Medicinal Chemistry*, edited by C. G. Wermuth, pp. 477–500. London: Academic Press.
Nioche, P., Liu, W.-Q., Broutin, I., Charbonnier, F., Latreille, M.-T., Vidal, M., Roques, B., Garbay, C. & Ducruix, A. (2002). *J. Mol. Biol.* **315**, 1167–1177.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rahuel, J., Gay, B., Erdmann, D., Strauss, A., Garcia-Echeverria, C., Furet, P., Caravatti, G., Fretz, H., Schoepfer, J. & Grütter, M. G. (1996). *Nature Struct. Biol.* **3**, 586–589.
Rahuel, J., Garcia-Echeverria, C., Furet, P., Strauss, A., Caravatti, G., Fretz, H., Schoepfer, J. & Gay, B. (1998). *J. Mol. Biol.* **279**, 1013–1022.
Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* D**60**, 1355–1363.
Udugamasooriya, D. G. & Spaller, M. R. (2008). *Biopolymers*, **89**, 653–667.
Vojtek, A. & Ders, C. (1998). *J. Biol. Chem.* **273**, 19925–19928.