Journal of Biomolecular NMR **27:** 143–149, 2003.

Structural investigations of a GYF domain covalently linked to a proline-rich peptide

Christian Freund[a,*], Ronald Kühne[b], Sunghyouk Park[d], Katharina Thiemke[a], Ellis L. Reinherz[c] & Gerhard Wagner[d]

[a]*Protein Engineering Group and* [b]*Molecular Modeling Group, Forschungsinstitut für Molekulare Pharmakologie and Freie Universität Berlin, Robert-Rössle-Str. 10, 13125 Berlin, Germany;* [c]*Laboratory of Immunobiology, Dana-Farber Cancer Institute and Department of Medicine and* [d]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, U.S.A.*

Received 25 February 2003; Accepted 30 April 2003

Abstract

Protein structure determination of low affinity complexes of interacting macromolecules is often hampered by a lack of observable NOEs between the binding partners. Covalent linkage offers a way to shift the equilibrium of the interaction partners to the bound state. Here we show that a single-chain protein containing the GYF domain of CD2BP2 and the target peptide SHRPPPPGHRV from CD2 allows for the intramolecular association of the binding partners. We obtained NOEs between the GYF domain and the peptide that could define the principal orientation of the peptide in the complex. In conjunction with general recognition rules for proline-rich sequence recognition these NOEs allowed the accurate modeling of the protein-peptide complex.

Introduction

Many important biological interactions are characterized by high off rates of the macromolecules involved (Kirschner et al., 2000). In signal transduction high on and off-rates of the part-taking proteins guarantee a dynamic response to changing environmental stimuli. Adapter domains that recognize proline-rich sequences offer an example of widely used and important protein modules that bind the respective target sequences with measured K_D's in the μM to mM range (Kay et al., 2000). Five different folds that bind proline-rich sequences have been identified to date, namely the SH3 domain (Stahl et al., 1988; Mayer et al., 1988), the WW domain (Bork and Sudol, 1994), the EVH1 domain (Niebuhr et al., 1997), profilin (Carlsson et al., 1977) and the GYF domain (Nishizawa et al., 1998; Freund et al., 1999). Structures of complexes of representative members of these fold families have been solved (Musacchio et al., 1994; Yu et al., 1994; Macias et al., 1996; Mahoney et al., 1997; Prehoda et al., 1999; Freund et al., 2002) and show striking similarities across domain borders: (i) the proline-rich ligand binds in an extended conformation and a stretch of at least four amino acids forms a left-handed polyproline-typeII helix (ii) the interaction surface of the protein domains contain aromatic amino acids that are highly conserved within the respective family (iii) the presence of 1–3 hydrogen bonds between the side-chains of the protein domains and the backbone carbonyl groups of the peptide ligand is observed and probably helps to orient and stabilize the primarily hydrophobic interactions between the prolines of the peptide and the aromatic side-chains of the protein.

Given the importance of proline-rich sequence recognition in the assembly of intracellular protein complexes it is desirable to create a large database of structural and functional information regarding these domains. Despite the small size of the adapter domains and peptides, the high K_D values of the binding partners often hamper a straightforward NOE based structural investigation of the interface by NMR spectrocsopy. On the other hand, the knowledge of the

*To whom correspondence should be addressed. E-mail: freund@fmp-berlin.de

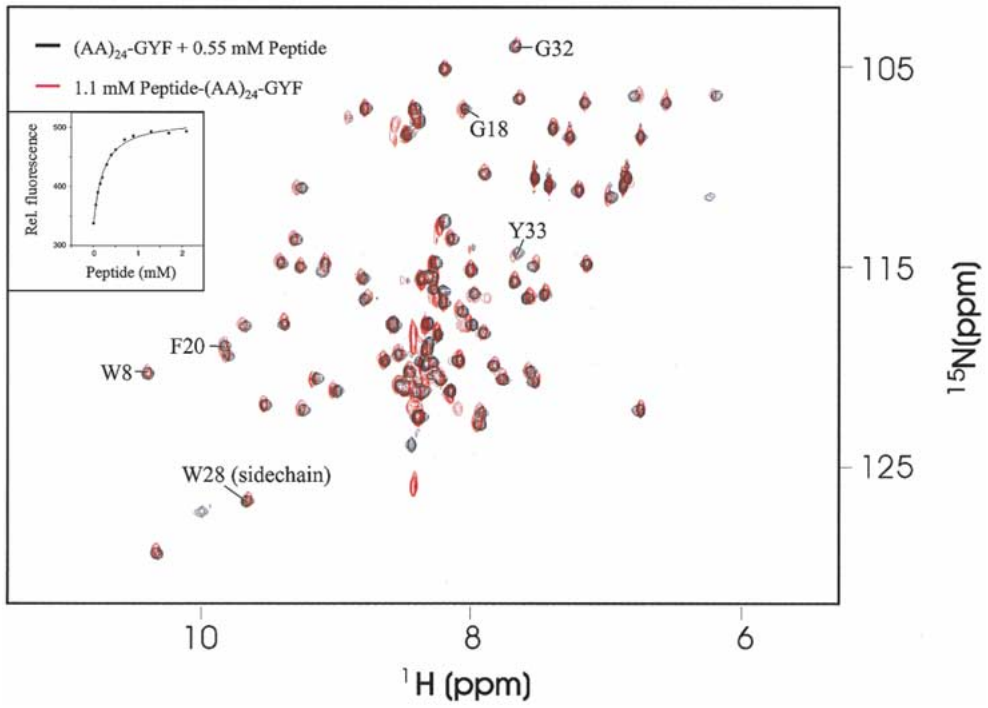

Figure 1. Overlap of the ^{15}N-^{1}H HSQC spectra of the $(AA)_{24}$-GYF domain after addition of 0.55 mM SHRPPPPGHRV peptide (black) and a 1.1 mM sample of the Peptide-$(AA)_{24}$-GYF protein (red). The resonances of binding-site residues that are shifted upon addition of the peptide are labeled with residues type and number (the numbering starting with the Asp1 of the GYF domain). The inset shows the fluorescence titration curve of the GYF domain upon addition of increasing amounts of the peptide.

general rules for proline-rich sequence recognition and the rigid structure of the polyproline type II helix sets the stage for molecular modeling of homologous interactions on the basis of experimentally determined structures. Here, the major obstacle is the prediction of the correct orientation of a given peptide relative to the shallow surface of the respective adapter domains. Furthermore, it has been shown that proline-rich peptides can bind in two inverse directions to adapter domains as a result of the pseudosymmetry of the polyproline helix (Feng et al., 1994).

We therefore attempted to combine a knowledge-based approach with an experimental approach to obtain a minimal number of NOE constraints. We show that the covalent linkage of the CD2 peptide SHRPPPPGHRV to the N-terminus of the CD2BP2 GYF domain allows for the intramolecular binding of the peptide to the GYF domain. In addition, it was possible to obtain intra- and intermolecular NOEs from a single ^{15}N-labeled sample of the single-chain protein. When combined with five constraints derived from the general recognition rules, we obtain ensembles of structures that are very similar to the recently determined structure of the GYF domain in complex with the SHRPPPPGHRV peptide (Freund et al., 2002). This approach should therefore be useful in the accurate modeling of unknown complexes of adapter domains and prolin-rich sequences.

Materials and methods

Protein expression and purification

The single chain proteins were expressed in the vector pTFT74 in *E. coli* BL21(DE3). The large amount of overexpressed, soluble protein in E.coli allowed the purifiaction of the single-chain protein without tags. Purification of the protein was achieved by the application of a MonoQ ion exchange column and a subsequent gel filtration (Superdex-75).

Sedimentation equilibrium experiments

Analytical ultracentrifugation was done with Beckmann Optim XL-A analytical ultracentrifuge (Palo Alto, CA) equipped with a 4-hole 60-Ti rotor. Different detection wavelengths were used for different

concentration of the protein to avoid nonlinear response at higher concentrations: (20 μM: 280 nm), (50 μM: 290 nm). Two different rotor speeds (30,000, 45,000 rpm) were tried to check the consistency of the data and the representative data is shown. All the samples were prepared in 20 mM phosphate pH 6.3 and 100 mM NaCl buffer. Data were analyzed using a customized version of Microcal Origin software supplied by the ultracentrifuge vendor. The partial specific volume of the protein was calculated from the amino acid sequence (0.715 cm^3/g).

Fluorescence spectroscopy

The K_D of the unlinked GYF domain was determined by measuring the change in the intrinsic fluorescence of aromatic residues of the GYF domain binding site upon addition of increasing amounts of the CD2 derived peptide SHRPPPPGHRV. The excitation wavelength in the experiments was 280 nm and emission spectra between 300 and 400 nm were recorded after each addition. For data analysis, the intensity increase at 325 nm of the emission spectra was evaluated by fitting the data to the equation:

$$[\mathrm{GYF}]_b/[\mathrm{GYF}]_t = 0.5 * (1 + K_D/[\mathrm{GYF}]_t + [\mathrm{Pep}]_t/[\mathrm{GYF}]_t) - 0.5 * \mathrm{sqrt}((1 - K_D/[\mathrm{GYF}]_t - [\mathrm{Pep}]_t/[\mathrm{GYF}]_t)^2 + 4 * K_D/[\mathrm{GYF}]_t). \quad (1)$$

To account for the difference in the protein concentration in the fluorescence ($[\mathrm{GYF}]_t = 46$ μM) and the NMR titration experiment ($[\mathrm{GYF}]_t = 250$ μM), the fraction bound in the NMR experiment was calculated from (1) with the K_D determined by the fluorescence experiments $K_D = 203$ μm) and a $[\mathrm{GYF}]_t$ of 250 μM.

NMR spectroscopy

All NMR experiments were performed at a Bruker 600 MHz DMX spectrometer at 298 K. Assignments of the single-chain proteins were based on the assignments of the unlinked GYF domain (Freund et al., 1999) and performed by the combined use of 15N-edited NOESY and 15N-TOCSY spectra and 2D NOESY and 2D TOCSY spectra in D_2O. Intra-Peptide and Peptide-GYF NOEs were derived from either a 15N-edited NOESY spectrum with a mixing time of 150 ms or from a 2D NOESY spectrum in D_2O with a mixing time of 150 ms. The data were processed and analyzed with the programs PROSA (Güntert et al., 1992) and XEASY (Eccles et al. 1991).

Structure calculation

200 structures were calculated with CNS1.0 (Brünger et al., 1998) using a torsion angle dynamics simulated annealing protocol. The temperature of the hot phase was 50,000 °K and the force constants for NOE and dihedral angle restraints were 50 $kcalmol^{-1}$ $Å^{-2}$ and 200 $kcalmol^{-1}$ rad^{-2}, respectively. Finally, all structures were minimized 200 steps. The twenty lowest energy structures displayed no distance violations greater than 0.3 Å and they were taken for further analysis. The implemented pseudo-constraint for the hydrogen bond between NHε of W28 and CO of Pro5 was set to an upper limit of 2.0 Å, while the proline ψ angle of residues 5–8 was restrained to 146 ± 20°, based on the ideal PII helical conformation.

Results and discussion

The covalently linked peptide binds to the major interaction surface of the GYF domain

The sequence of the single-chain GYF domain construct (scGYF) used in this study is:

Peptide Linker
GSHRPPPPGHRV-*AEEELETPTPTQRGEAESRGDGLV*-
DVMWEYKWENTGDAELYGPFTSAQMQTWVSEGY
FPDGVYCRKLDPPGGQFYNSKRIDFDLYT
GYF domain

The 24 amino acid linker in the scGYF protein was derived from the CD2BP2 sequence adjacent to the C-terminal GYF domain. This 'natural' linker sequence was shown to be unstructured and not to interact with the GYF domain (Freund et al., 1999). The black resonances in Figure 1 show the HSQC spectrum of 0.25 mM Linker-GYF mixed with 0.55 mM CD2 peptide. At these concentrations, approximately 66% of the protein is in the bound state as derived from the fluorescence titration curve in the inset of Figure 1. Superposition of this spectrum with the spectra of the scGYF at 1.1 mM concentration (red in Figure 1) shows that the GYF domain peaks within the scGYF overlap well with the Linker-GYF domain/peptide mixture (black resonances) except for the additional peaks that arise from the Peptide-NH resonances of the scGYF protein. More specifically, the residues W8, Y17, G18, F20, W28, Y33, G34 are shifted as compared to the uncomplexed, unlinked GYF domain

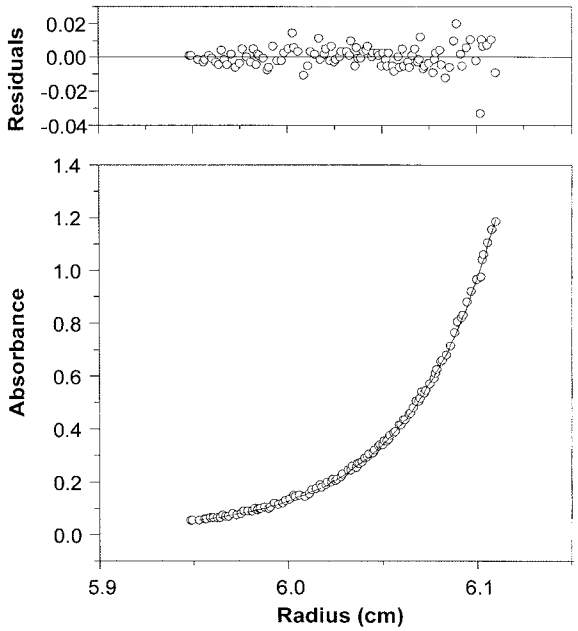

Figure 2. Results of the equilibrium ultracentrifugation of the Peptide-$(AA)_{24}$-GYF protein at 20 mM concentration, 298 K, 45.000 rpm as measured at 280 nm. Data were fitted using the single ideal species model with the vendor provided software. The lower panel shows the experimental data (open circles) and the fitted line (red lines). The upper panel shows the residuals as the difference between the experimental and the fitted data. Random distribution of the residuals indicates a good fit to the single ideal species model.

(Nishizawa et al., 1998). These residues form the binding site of the GYF domain creating a hydrophobic patch at the surface of the protein. This demonstrates that the GYF domain in the scGYF adopts a very similar conformation compared to the unlinked domain and that the binding surface of the domain is the same for the unlinked versus the peptide linked protein.

Intramolecular interaction of the peptide and the GYF domain

To test whether the linked proline-rich peptide binds to the GYF domain intramolecularly (cis) or intermolecularly (trans) we varied the protein concentration and found that there was no concentration dependence for the fraction bound in the NMR experiments. This observation indicates that the binding is unimolecular rather than bimolecular. To test this hypothesis further, analytical ultracentrifugation experiments were performed with either 20 or 50 μM solutions of the scGYF. The results are shown in Figure 2 for the 20 μM sample. The experimentally determined curve resulted in a molecular weight of 12,353 D at a protein concentration of 20 μM protein. These values are close to the theoretical mass of the monomer (11,407 D) and indicate that the scGYF is primarily monomeric. Since the major fraction of the protein at 20 μM is in the bound state, we conclude that the protein-peptide interaction occurs in cis at low protein concentrations and that the linker does not interfer with the intramolecular association of the peptide with the GYF domain. Very similar results (experimentally determined molecular weight = 12,520 D) were obtained for the scGYF at 50 μM.

Table 1. NOEs between the GYF domain and peptide residues

Atom in GYF domain	Atom in peptide
W8/HD1	Gly8/HA1
W8/HD1	Gly8/HA2
W8/HE1	Gly8/HA∗
F17/QE	Gly8/HA1
F17/QE	Gly8/HA2
F20/QE	Pro7/HD1
F20/QE	Pro7/HD2
F20/QE	Pro7/HB∗
Y33/QE	Arg3/HG∗

Structure calculations of the scPeptide-Linker-GYF protein

To confirm that the conformation of the GYF domain is unchanged upon covalent linkage with the CD2 peptide, the NOESY-spectra of the linked proteins were assigned on the basis of the unlinked GYF domain. The observed NOE cross peak pattern was almost identical to the NOE pattern of the unlinked GYF domain, confirming the notion that the presence of the linker does not alter the protein conformation. This is not surprising given the fact that the binding site comprises mainly tightly packed aromatic amino acids. Therefore, in the structure calculations, we used the NOE constraints of the refined unlinked GYF domain in combination with intraligand and intermolecular NOEs of the single-chain protein. The latter were derived either from a ^{15}N-NOESY-HSQC or a 2D NOESY spectrum in D_2O. A total of 22 intra-peptide and 9 peptide-GYF NOEs was used in the calculations. The 9 peptide-GYF NOEs are observed between the side-chains of aromatic residues of the GYF domain and protons of peptide residues 4, 8, and 9 (Table1).

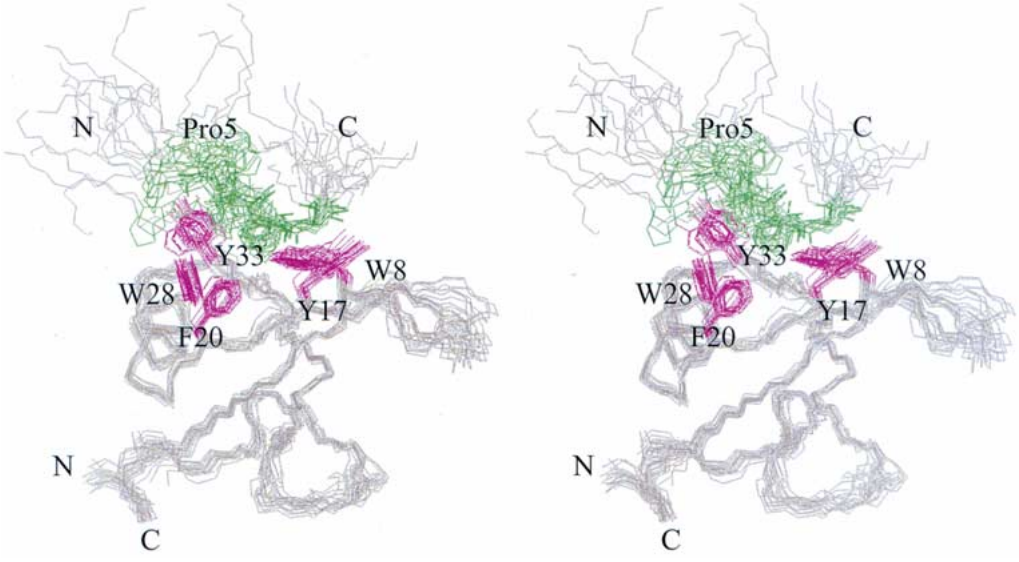

Figure 3. Ensemble of the 20 lowest energy structures of the GYF domain in complex with the peptide SHRPPPPGHRV as derived from CNS calculations of the Peptide-$(AA)_{24}$-GYF protein that contained experimental NOEs of the single-chain protein. The linker has been omitted for clarity and residue numbers of the GYF domain are depicted in the amino acid one-letter code while the peptide residues are marked by the amino acid three-letter code. The numbering of the GYF domain and the peptide is the same as described previously (Freund et al., 2002). The N- and C-termini of the GYF domain and the peptide are marked by the letters N and C, respectively.

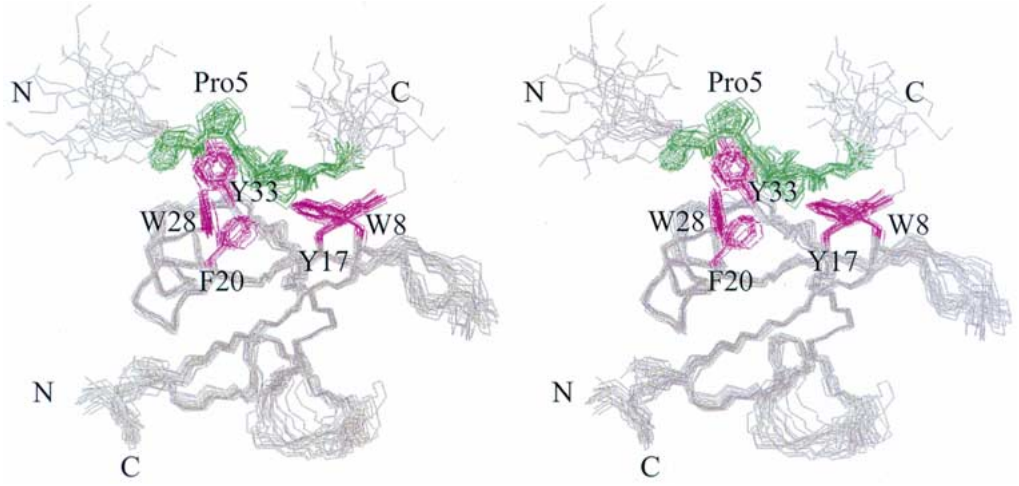

Figure 4. Ensemble of the 20 lowest energy structures of the GYF domain in complex with the peptide SHRPPPPGHRV using NOE data from the scGYF and 5 pseudo-constraints derived from the general recognition rules for proline-rich sequence recognition. Residue marking is the same as in Figure 3.

The result of the initial structure calculation is shown in Figure 3. The linker is omitted for clarity, since no NOE constraints were used for this part of the molecule. The structural ensemble (Figure 3) shows that the peptide adopts an extended structure with a preferred orientation that places Pro6 and Pro7 (shown in green in Figure 3) into a hydrophobic pocket of the domain that is constituted predominantly by aromatic side chains. Gly8 is also contacting the binding site and allows the ligand to kink in order to prevent a steric clash with the side chains of W8, E15 and Y17 of the GYF domain. The linker peptide of 24 amino acids was of sufficient length to geometrically allow the CD2 peptide to assume the two inverse orientations that have been observed for proline-rich peptide binding to adapter domains (Feng et al., 1994). However, we observed a prefered orientation of the peptide, where the C-terminal residues of the peptide are in

proximity to the GYF domain W8 side-chain (Figure 3). This observed orientation requires the linker to wrap around the domain in order to reach the N-terminus of the peptide, while the inverse orientation would allow the linker to adopt a more relaxed conformation. We therefore conclude that the linker does not influence the orientation of the peptide bound to the GYF domain.

Structure refinements implementing pseudo-constraints

In some of the structures the diagnostic hydrogen bond between the side-chain of Trp28 and either Pro4 or Pro5 of the ligand is observed and justifies the use of a pseudo constraint that fixes either of the two hydrogen bonds in separate calculations. Initial structure calculations indicated that the implementation of a hydrogen bond between NHε of Trp28 and CO of Pro4 of the ligand results in lower energies than the implementation of the hydrogen bond between NHε of Tpr28 and CO of Pro5. We only show the result for the lower energy ensemble. In addition, in some of the structures the prolines adopt a PPII helix in agreement with the conformation of the proline-rich core motif of all known structures of complexes of proline-rich peptides with protein recognition domains. We therefore set the ψ angle of the four prolines to 146°. Figure 4 shows a stereoview of the ensemble of the 20 lowest energy structures of the single-chain protein by using 9 experimental Peptide-GYF NOEs in addition to the pseudo-constraints. Compared to the calculation without the addition of the pseudo-constraints (Figure 3), the convergence of the central part of the peptide is much improved and clearly defines the binding site in atomic detail. Figure 5 shows a comparison of the lowest energy structure of the Peptide-Linker-GYF protein with the previously determined structure of the GYF domain in complex with the SHRPPPPGHRV peptide. The good agreement between the experimentally determined structure and the structure obtained by the combined use of experimental and knowledge-based constraints indicates the validity of our approach for the accurate, semi-empirical modeling of homologous protein-proline-rich-peptide interactions.

Another advantage of the covalent linkage of proline-rich peptides to adapter domains is the intramolecular binding of the peptide at low protein concentrations that allows studying the bound conformation of the peptide under these conditions. With the advent of cryo-probe technology, structural investigations by NMR spectroscopy at protein concentrations as low as 50–100 μM become amenable. For the unlinked GYF domain at 50 μM protein and equimolar peptide, only 10% is in the complex as compared to the approximately 66% of the scGYF. For proteins and peptides with limited solubility and low affinity, it is therefore impossible to obtain NOE data for structure calculations. The covalent linkage of proline-rich peptides to their respective adapter domains offers a way to obtain experimental constraints in such unfavorable cases.

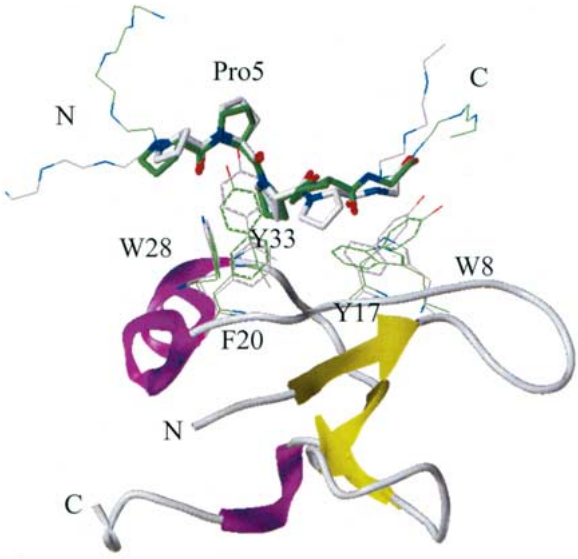

Figure 5. Comparison of the structure of the complex of the GYF domain with the peptide SHRPPPPGHRV as determined previously (Freund et al., 2002) (grey) with the lowest energy structure of the ensemble calculated in Figure 4 of this study (green). Residue numbering is as described for Figure 4.

Acknowledgements

This work was supported by the BMBF (grant 0311879 to C.F.) and by grants from the NIH (AI 37581 to G.W. and AI 21226 to E.L.R.).

References

Bork, P. and Sudol, M. (1994) *Trends Biochem. Sci.*, **19**, 531–533.

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. and Warren, G.L. (1998) *Acta Crystallogr. D. Biol. Crystallogr.*, **54 (Pt 5)**, 905–921.

Carlsson, L., Nystrom, L.E., Sundkvist, I., Markey, F. and Lindberg, U. (1977) *J. Mol. Biol.*, **115**, 465–483.

Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 111–130.

Feng, S., Chen, J.K., Yu, H., Simon, J.A. and Schreiber, S.L. (1994) *Science*, **266**, 1241–1247.

Freund, C., Dötsch, V., Nishizawa, K., Reinherz, E.L. and Wagner, G. (1999) *Nat. Struct. Biol.*, **6**, 656–660.

Freund, C., Kühne, R., Yang, H., Park, S., Reinherz, E.L. and Wagner, G. (2002) *EMBO J.*, **21**, 5985–5995.

Güntert, P., Dötsch, V., Wider, G. and Wüthrich, K. (1992) *J. Biomol. NMR*, **2**, 619–630.

Kay, B.K., Williamson, M.P. and Sudol, M. (2000) *FASEB J.*, **14**, 231–241.

Kirschner, M., Gerhart, J. and Mitchison, T. (2000) *Cell*, **100**, 79–88.

Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M. and Oschkinat, H. (1996) *Nature*, **382**, 646–649.

Mahoney, N.M., Janmey, P.A. and Almo, S.C. (1997) *Nat. Struct. Biol.*, **4**, 953–960.

Mayer, B.J., Hamaguchi, M. and Hanafusa, H. (1988) *Nature*, **332**, 272–275.

Musacchio, A., Saraste, M. and Wilmanns, M. (1994) *Nat. Struct. Biol.*, **1**, 546–551.

Niebuhr, K., Ebel, F., Frank, R., Reinhard, M., Domann, E., Carl, U.D., Walter, U., Gertler, F.B., Wehland, J. and Chakraborty, T. (1997) *EMBO J.*, **16**, 5433–5444.

Nishizawa, K., Freund, C., Li, J., Wagner, G. and Reinherz, E.L. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 14897–14902.

Prehoda, K.E., Lee, D.J. and Lim, W.A. (1999) *Cell*, **97**, 471–480.

Stahl, M.L., Ferenz, C.R., Kelleher, K.L., Kriz, R.W. and Knopf, J.L. (1988) *Nature*, **332**, 269–272.

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. and Schreiber, S.L. (1994) *Cell*, **76**, 933–945.