Supporting Information

Table S1, X-ray data collection, refinement and model statistics

	Native protein	SeMet-labelled protein	$VFT2_{F375E+Q461E}$
A. *Data collection*			
	Native	Peak	
Wavelength (Å)	0.979	0.97910	1.5418
ESRF beam-line	Id23-1	Id23-1	Cu
Resolution range (Å)	20-2.0	20-2.0	5,6-2.1
Space group		$P2_12_12$	$P2_12_12$
Cell parameters (Å)		48.95 95.41 64.02	48.80 101.48 53.83
Completeness (%)	97.1	99.6	94.1
$I/\sigma I$	27.47	12.3	20.74
R_{sym}[a] (%)	5.3	10.5	8.4
No. reflections	97935	147040	107317
No. unique reflections	17616	39129	15270
B. *Refinement statistics*			
Resolution (Å)	19.56-2.04		5.58-2.1
No. reflections	18142		14491
R-factor[b] (%)	17.81		19.08
R_{free}[c] (%)	21.86		24.57
C. *Model statistics*			
B-factor analysis			
All (Å^2)	19.113		25.45
RMSD from ideal			
Bond angles (deg.)	1.934		1.844
Bond lengths (Å)	0.027		0.023

Values in parentheses are for the highest resolution shell.
[a] $R_{sym}=\Sigma\Sigma|I_{hkl} - I_{hkl}(j)|/\Sigma I_{hkl}$ where $I_{hkl}(j)$ is observed intensity and I_{hkl} is the final average value of intensity.
[b] R-factor$=\Sigma||F_{obs}| - |F_{calc}||/|F_{obs}|$.
[c] $R_{free}=\Sigma||F_{obs}| - |F_{calc}||/|F_{obs}|$ where all reflections belong to a set of 5% data randomly selected in Refmac omitted from the refinement

Table S2. Structural parameters of structural homologs of VFT2

Structures	Related structures	Description	RMSD	Opening angle
1HSL		Complex with histidine of a PBP from E. coli		27.8
	1HPB	Complex with histidine	0.41	29.1
1IIT		Complex with Serine of GluRO from Synechocystis		39.4
	1II5	Complex with glutamate	0.2	39.7
	1IIW	non-liganded closed form	0.2	39.4
1LAF		Complex with arginine of a PBP from Salmonella		28.9
	1LST	Complex with lysine	0.1	28.9
	1LAG	Complex with histidine	0.2	28.7
	1LAH	Complex with ornithine	0.2	28.5
	2LAO	non-liganded open form	**4.7**	**59.8**
1S9T		Complex with quisqualate (agonist) of GluR6 from Rat		36.7
	1S50	Complex with glutamate (agonist)	0.5	38.8
	2IOC	SS-bridge Mutant	0.6	37.9
	1YAE	Complex with domoic acid (agonist)	1.5	43.7
1WDN		Complex with glutamine of a PBP from E. coli		33.3
	1GGG	non-liganded open form	**5.3**	**64.9**
1XT8		Complex with cysteine of a PBP from Campylobacter		26.8
1YCJ		Complex with glutamate of GluR5 from Rat		36.4
	1VSO	Complex with S(ATPO) (antagonist)	**2.9**	**52.2**
	2WKY	Complex with 4-AHCP (agonist)	1.2	42.2
	2F34	Complex with Ubp310 (antagonist)	**3.2**	**53.9**
2PYY		Complex with glutamate of GluRO from Nostoc		36.6
2Q2C		Complex with histidine of a PBP from Geobacillus		32.5
	2Q2A	Complex with arginine	0.3	31.0
	2PVU	Complex with lysine	0.3	30.8
2V25		Complex with aspartate of a PBP from Campylobacter		29.9
3H7M		PBP of a histidine kinase from Geobacter		36.6
3I6V		Complex with lysine of a PBP from Silicibacter		32.1
BvgS		PBP(VFT2) of a histidine kinase from Bordetella		39.8
	$VFT2_{F375E+Q461E}$	Mutant F375E/Q461E of VFT2	1.1	40.8
2Q88		Complexe with ectoine of a PBP from Sinorhizobium		33.1
	2Q89	Complex with hydroxyectoine	0.3	33.4

The RMSD values correspond to distinct structures of the same protein in each case.
See supporting materials and methods (below) for the calculation of opening angles.
Values in bold denote open forms either non liganded or complexed with an antagonist.

Figure S3. Electrostatic surface potentials.
The electrostatic potentials of the VFT2 and VFT2F375E+Q461E cavities are shown at the left and right, respectively.

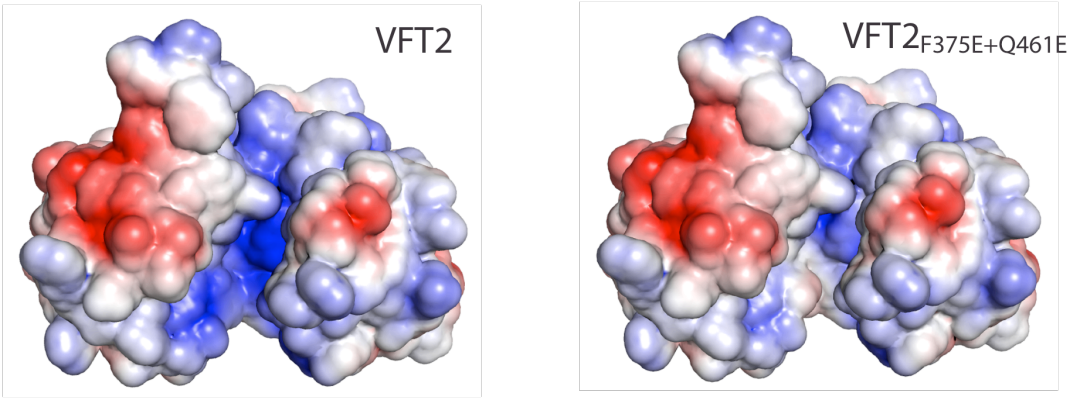

Figure S4. Superposition of the VFT2 and VFT2$_{F375E+Q461E}$ structures. VFT2 and VFT2$_{F375E+Q461E}$ are shown in blue and magenta ribbons, respectively. The lobes I of the two proteins have been superposed.

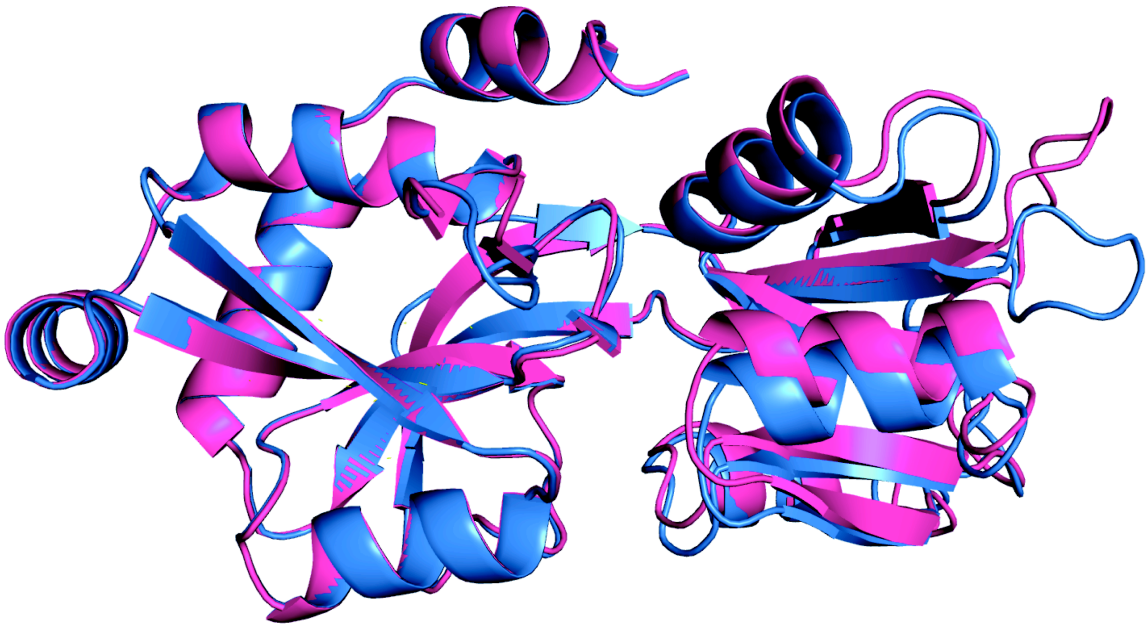

Figure S5. Close-up view of the cavity of $VFT2_{F375E+Q461E}$. Lobes I and II are represented in pale green and orange ribbons, respectively. $VFT2_{F375E+Q461E}$ is viewed from the top, with the two strands of the hinge shown in grey. The H bonds are shown as dotted yellow lines and the water molecules as red spheres. The major interactions between the newly introduced glutamates (in cyan) and cavity residues are depicted. Side-chain –side chain interactions that could contribute to stabilization of the closed conformation include a salt bridge between Arg 421 and Glu 357 and a hydrogen bond between the amide group of Ala 424 and the introduced Glu 375.

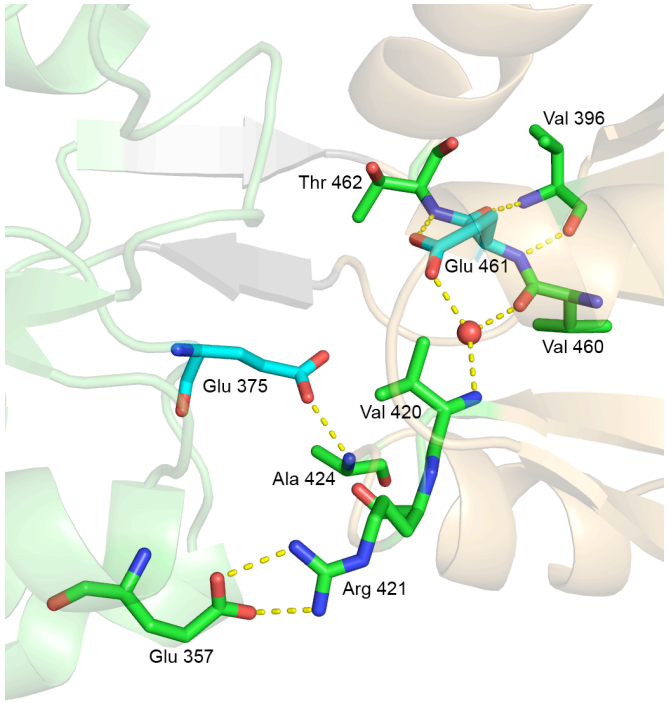

Figure S6. Structural alignment of VFT2 with its closest homologues. For each protein only the residues involved in ligand binding as based on the liganded structure are shown. Lobes I and II of VFT2 are delineated by dark red and green lines, respectively, with strands in red and helices in blue. The residues of the VFT2 cavity targeted for substitution are showed by yellow boxes. The PBPs used in the alignment are *E. coli* GlnBP (PDB code: **1WDN**), *Synechocystis* GluR0 (**1II5**), *Nostoc punctiforme* GluR0 ligand binding core (**2PYY**), *Campylobacter jejuni* Asp/Glu receptor (**2V25**), *Geobacillus stearothermophilus* Arg/Lys/HisBP (**2Q2A**), *E. coli* HisBP (**2HSL**), *C. jejuni* CysBP (**1XT8**), *Rhizobium meliloti* Ectoine BP (**2Q88**).

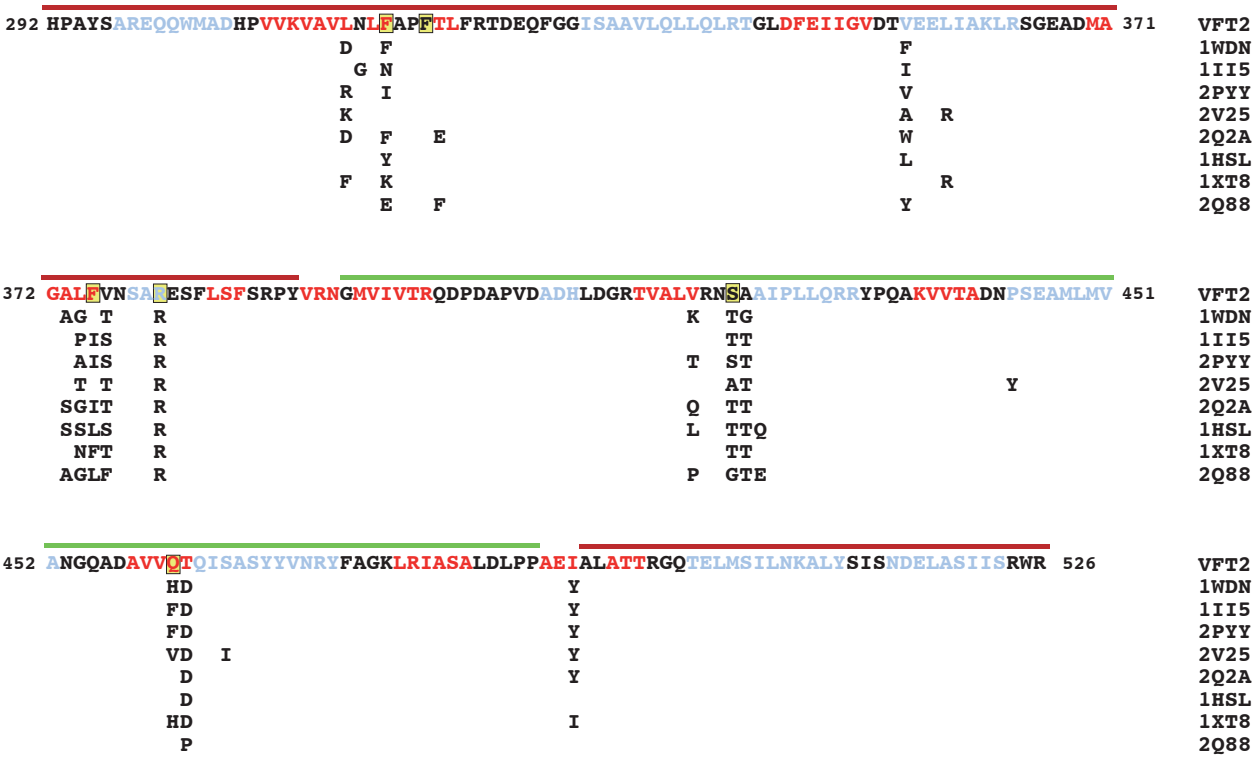

Table S7. Oligonucleotides used in this study

Oligonucleotides	Sequences	Fragment amplified
VFT1-Up	5'-AT**GGATCC**GCGCAGGCAAGCCAGGA-3'	Sequence coding the VFT1 recombinant protein
VFT1-Lo	5'-AT**CTCGAG**CGAAATGCTGCTGCCCAGTCC-3'	
VFT2-Up	5'-AT**GGATCC**CTCGATTTCGCGCACC-3'	Sequence coding the VFT2 recombinant protein
VFT2-Lo	5'-AT**AAGCTT**CTAGATCTCGTTGCGGTAGGC-3'	
VFT2zip-Up	5'-ACAGCATTTCGAACGACGAGC-3'	Sequence coding a part of the VFT2 domain for the VFT1+2zip construction
VFT2zip-Lo	5'-TCTTCCAGCTGTTTCATACGTTGCTCGTTGCGGTAGGCGTACCA-3'	
VFT1+2-Up	5'-TA**GGATCC**GCGCAGGCAAGCCAG-3'	Sequence coding the VFT1+2 recombinant protein
VFT1+2-Lo	5'-TA**CTCGAG**GTAGATCTCGTTGCGGTAGGC-3'	
ZIP-Up	5'-CAACGTATGAAACAGCTGGAAGA-3'	Sequence coding a leucine zipper
ZIP-Lo	5'-TACTCGAGCCCACGTTCACCCACCAGTTT-3'	
F317A-Up	5'-GCGGTCCTGAATCTGGCAGCGCCCTTCACCCTGTTCCGC-3'	Replacement of the phenylalaline 317 by an alanine in the lobe 1 of VFT2
F317A-Lo	5'-GCTGCCAGATTCAGGACCGCCAC-3'	
F320A-Up	5'-TCTGTTCGCGCCCGCAACCCTGTTCC-3'	Replacement of phenylalanine 320 by alanine in the lobe 1 of VFT2
F320A-Lo	5'-GGAACAGGGTTGCGGGCGCGAACAGA-3'	
F375A-Up	5'-GACATGGCCGGCGCCCTGGCAGTCAACAGCGCGCGGGAGTCC-3'	Replacement of phenylalaline 375 by alanine in the lobe 1 of VFT2
F375A-Lo	5'-TGCCAGGGCGCCGGCCATGTC-3'	
R380A-Up	5'-CTGTTCGTCAACAGCGCGGCGGAGTCCTTCCTCAGTTTCAG-3'	Replacement of arginine 380 by alanine in the lobe 1 of VFT2
R380A-Lo	5'-CGCCGCGCTGTTGACGAACAG-3'	
F375E-Up	5'-GACATGGCCGGCGCCCTGGAGGTCAACAGCGCGCGGGAGTCC-3'	Replacement of phenylalanine 375 by glutamate in the lobe 1 of VFT2
F375E-Lo	5'-CTCCAGGGCGCCGGCCATGTC-3'	
S423A-Up	5'-TTGGTGCGCAACGCTGCCGCCATTCCCCTGCTGCAG-3'	Replacement of serine 423 by alanine in the lobe 2 of VFT2
S423A-Lo	5'-GCGGCAGCGTTGCGCACCAACGCGAC-3'	
Q461A-Up	5'-CAGGCCGACGCCGTCGTGGCAACGCAGATCAGCGCCAGCTAT-3'	Replacement of glutamine 461 by alanine in the lobe 2 of VFT2
Q461A-Lo	5'-TGCCACGACGGCGTCGGCCTG-3'	
Q461E-Up	5'-CAGGCCGACGCCGTCGTGGAGACGCAGATCAGCGCCAGCTAT-3'	Replacement of glutamine 461 by glutamate in the lobe 2 of VFT2
Q461E-Lo	5'-CTCCACGACGGCGTCGGCCTG-3'	

Supplementary Material and methods

Construction of expression plasmids

VFT1-Up and VFT2-Up were used as the 5′ primer with VFT1-Lo and VFT2-Lo as the 3' primers to amplify the sequences of VFT1 and VFT2. The sequences of the oligonucleotides are given in Table S7. The amplicons were inserted into pCR®II-TOPO (Invitrogen) and sequenced. They were then introduced as a *Bam*HI-*Hin*dIII (for VFT2) and *Bam*HI-*Xho*I (for VFT1 and VFT1+VFT2) fragments into the corresponding sites of pQE-30 (QIAGEN) and pGEV2 [1], respectively. The resulting plasmids, pQE-VFT2, pGEV-VFT1 and pGEV-VFT1+2 code for VFT2 with an N-terminal 6-His tag, and for VFT1 and VFT1+VFT2 with N-terminal GB1 domain and C-terminal 6-His tags, respectively. The GB1 domain increases the solubility and stability of the latter proteins.

To produce a dimeric form of the BvgS PBP domains (VFT1+2_{zip}), a recombinant protein with a C-terminal leucine zipper domain was constructed by overlapping PCR. The sequence for a leucine zipper was amplified from pT18-zip [2] by using the ZIP-Up and ZIP-Lo primers, and the 3' part of the VFT2 sequence was amplified by using the VFT2zip-Up and VFT2zip-Lo primers. The amplicon resulting from the overlapping PCR was used to replace the *Xho*I-*Sfu*I fragment of pGEV-VFT1+2.

Protein purification

VFT1, VFT1+VFT2 and VFT1+$VFT2_{zip}$ were produced in *E. coli* BL21(DE3), and VFT2 was produced in *E. coli* SG13009(pREP4). The cultures were incubated at 37 °C under rotary shaking (220 rpm). Expression of the recombinant genes was induced at an

OD_{600nm} of 0.8 by adding IPTG to 1 mM. After 3 h of incubation, the cells were harvested by centrifugation.

All purification steps were carried out at room temperature. Typically, the cell pellets were resuspended in 10 mM Tris–HCl (pH 7.4), 150 mM NaCl, 10 mM imidazole with 5 µg/ml of DNase I and a tablet of Complete EDTA-free Protease Inhibitor Cocktail (Roche). Because the recombinantVFT1+VFT2 proteins precipitated in these conditions, a 10 mM Tris–HCl (pH 8.8), 500 mM NaCl, 10 mM imidazole buffer was used instead.

Cells were disrupted by two passages in a French pressure cell, and the cell debris were removed by centrifugation for 20 min at 20,000x*g*. The supernatant was loaded onto a nickel-Sepharose affinity column (HiTrap Chelating HP 5 ml, GE Healthcare) pre-equilibrated with the respective lysis buffers. Two washing steps were performed using 10 mM and 75 mM followed by the elution step with 200 mM imidazole in the same buffers. The protein solutions were then dialysed against the same buffers as above.

Gel filtration experiments using a HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) were performed with VFT1+2 and VFT1+2_{zip} at 70 mg/ml. The column was calibrated with molecular masses markers, yielding estimated molecular masses of 69 ± 13 kDa for VFT1+2 (calculated mass, 64 kDa) and 153 ± 31 kDa for VFT1+2_{zip} (calculated mass, 68 kDa), indicating that they are monomeric and dimeric, respectively.

$VFT2_{F320A}$ tended to precipitate upon dialysis, and thus TSA was performed with imidazole using a VFT2 control in the same conditions. These data are thus not included in table 1.

Ligand-binding studies by fluorescence spectroscopy

The purified proteins were dialysed against 50 mM sodium phosphate (pH 7.4), 50 mM NaCl at 4°C. The final concentrations of the protein solutions were determined by using a NanoVue spectrophotometer (GE Healthcare). For the titration experiments, 2.5 ml-protein solutions (0.1 µM for VFT1, 0.3 µM for VFT2 and 0.06 µM for VFT1+VFT2) were mixed with small volumes of concentrated solutions of ligand (typically 2.5 µl of a 10-30 mM solution). The fluorescence intensity of the recombinant proteins was measured using a Perking Elmer LS 55 spectrofluorimeter with an excitation wavelength of 295 nm, and the emission spectrum was scanned from 290 nm to 450 nm. Fluorescence was maximal at 348 nm. The sample cell was maintained at 20 °C with a circulating water bath, and the sample solution was mixed by using an included magnetic stirrer.

Crystallization of VFT2

Purified VFT2 was concentrated to 40 mg/ml using a stirred Ultrafiltration Cell (Amicon-Millipore). Protein purity was estimated to be 95%. The crystallization screening was carried out with the sitting-drop, vapour-diffusion technique in 96-wells microplates, a Cybi-workstation robot (Cybio) and commercial crystallisation kits (Qiagen and JBScreen kits). The drops were set up by mixing equal volumes of the protein and the precipitant solution equilibrated against 75 µl of the precipitant solution. The best crystals were obtained at 20°C by hand refinement with 0.1 M sodium acetate (pH 5), 20% MPD (w/v) based on the MPDSuite kit, crystallisation condition #50 (Qiagen). Crystals grew to their final size in seven to ten days. All crystallization

attempts by hand were carried out with the hanging-drop, vapour-diffusion technique using 24-wells plates. The drops were set up by mixing equal volumes of the protein and the precipitant solution equilibrated against 500 µl of the precipitant solution. Attempts to obtain crystal in the absence of acetate or in presence of a potential antagonist ligand for VFT2 were unsuccessful.

For the phasing step in structure determination, the selenomethionine-substituted (SeMet) VFT2 protein was produced as described [3]. Full incorporation of SeMet in place of the six Met residues was confirmed by tryptic digestion and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. The dialysis buffer was supplemented with 0.2 mM EDTA and 10 mM DTT to prevent SeMet oxidation. Crystals of SeMet-VFT2 grew under similar conditions to those used to produce native crystals. Soaking the crystals in the crystallization solution containing 25% glycerol allowed their flash-freezing directly at 100° K in liquid nitrogen.

To crystallize $VFT2_{F375E+Q461E}$, a new round of screening was performed as above. The best crystals were obtained at 20°C by hand refinement with 0.085 M HEPES (pH 7.5), 10% PEG 40000 (w/v), 6.8% ethylene glycol, 5 % glycerol, based on the CryoSuite kit, crystallization condition #95 (Qiagen). Crystals grew to their final size in ten to fifteen days.

X-ray data collections and processing

The structure was solved by the SAD method using the anomalous signal of selenium. A single SeMet-labelled protein crystal was used to collect a SAD dataset at beamline Id23-1 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The

datasets were collected at a resolution between 1.8 Å and 2 Å on an ADSC Q315R CCD detector. A single dataset of the native protein crystal was collected later at beamline Id23-1 at ESRF to 2 Å resolution. All diffraction data were processed and scaled with the XDS/ XSCALE program package. The estimated solvent content was 53% (v/v) assuming one molecule in the asymmetric unit. The final model was refined with 235 residues, 146 water, 2 acetate and 2 glycerol molecules.

For $VFT2_{F375E+Q461E}$, diffraction datasets were collected by using an in-house Mar345dtb detector at a resolution of 2.10 Å. Molecular replacement was performed by using the Amore and CCP4i softwares.

Structure determination and refinement

The positions of 6 Se atoms were determined with ShelxD [4]. The SAD parameters were refined, phases were calculated and the density modification procedure was carried out in SHARP [5]. The automated model building was accomplished by the WarpNTrace procedure, implemented with the CCP4 package [6], which allowed the building of one molecule of the asymmetric unit. A randomly selected 5 % of the observations were set aside for cross-validation analysis, and the behaviour of R_{free} and the likelihood free figure of merit were used to monitor refinement. Atomic models were refined using a maximum likelihood-based refinement strategy employing the REFMAC5 program [7], followed by manual fitting into σA-weighted electron density maps with COOT. The acetate and glycerol molecules present in the crystal were identified from Fo – Fc and 2Fo – Fc electron-density maps, and parameters defining it, present in REFMAC5 dictionaries were used without modification. An omit map was used to reduce model

bias. The coordinates of the final refined model were randomly shifted up to 0.3 Å using PDBSET, and the ligand molecules were omitted. The resulting model was subjected to 10 cycles of restrained refinement with REFMAC5, and maps were calculated from the refined omit model.

The stereochemical quality of structures was evaluated with PROCHECK [8]. The secondary structures were assigned by using the DSSP program, and interactions between the protein and the molecules present in the cavity were determined with Ligplot. Illustrations of the protein structure were produced with Pymol (www.schrodinger.com).

Structural models

Electrostatic surface potentials were calculated with the PBEQ Solver module of CHARMM. Three residues were defined –one on the lip of lobe I, one in the hinge between the lobes and on the lip of lobe II – that are structurally equivalent between all proteins. For GlnBP and VFT2, the sequence positions of these residues are 10, 86 and 139, and 314, 390 and 444, respectively. The angle formed between the vectors defined by the α carbons of residues 2 and 1 and the α carbons of residues 2 and 3, respectively, was thus determined.

Bibliography

1. Huth, J.R., et al., *Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR.* Protein Sci, 1997. **6**: p. 2359-2364.
2. Karimova, G., et al., *A bacterial two-hybrid system based on a reconstituted signal transduction pathway.* Proc Natl Acad Sci USA, 1998. **95**: p. 5752-5756.

3. Doublie, S., *Preparation of selenomethionyl proteins for phase determination.* Methods Enzymol., 1997. **276**: p. 523-530.
4. Schneider, T. and G. Sheldrick, *Substructure solution with SHELXD.* Acta Crystallogr D Biol Crystallogr., 2002. **58**(Pt 10 Pt 2): p. 1772-1779.
5. Bricogne, G., et al., *Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0.* Acta Crystallogr D Biol Crystallogr., 2003. **59**(Pt 11): p. 2023-2030.
6. Bailey, S., *Collaborative computational project: The CCP4 Suite: Programs for Protein Crystallography.* Acta Crystallogr., 1994. **D50**: p. 760-763.
7. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
8. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures.* J Appl Cryst, 1993. **26**: p. 283-291.