Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Engineering the catalytic domain of human protein tyrosine phosphatase β for structure-based drug discovery

Artem G. Evdokimov,[a]* Matthew Pokross,[a] Richard Walter,[b] Marlene Mekel,[c] Brooke Cox,[c] Chuiying Li,[c] Randy Bechard,[c] Frank Genbauffe,[c] Ryan Andrews,[c] Conrad Diven,[a] Brian Howard,[d] Vinit Rastogi,[a] Jeffrey Gray,[d] Matthew Maier[d] and Kevin G. Peters[d]

[a]Structural Biology, Procter and Gamble Pharmaceuticals, 8700 Mason-Montgomery Road, Mason, OH 45140, USA, [b]Corporate Research Division, The Procter and Company, Miami Valley Innovation Center, 11810 East Miami River Road, Cincinnati, OH 45252, USA, [c]Protein Production, Procter and Gamble Pharmaceuticals, 8700 Mason-Montgomery Road, Mason, OH 45140, USA, and [d]Cardiac Focus Area, Procter and Gamble Pharmaceuticals, 8700 Mason-Montgomery Road, Mason, OH 45140, USA

Correspondence e-mail: artem@xtals.org

Received 20 June 2006
Accepted 16 September 2006

PDB References: engineered catalytic domains of HPTPβ and complexes with inhibitors, 2i3r, r2i3rsf; 2i3u, r2i3usf; 2hc2, r2hc2sf; 2hc1, r2hc1sf; 2i4e, r2i4esf; 2i5x, r2i5xsf; 2i4g, r2i4gsf; 2i4h, r2i4hsf.

Protein tyrosine phosphatases (PTPs) play roles in many biological processes and are considered to be important targets for drug discovery. As inhibitor development has proven challenging, crystal structure-based design will be very helpful to advance inhibitor potency and selectivity. Successful application of protein crystallography to drug discovery heavily relies on high-quality crystal structures of the protein of interest complexed with pharmaceutically interesting ligands. It is very important to be able to produce protein–ligand crystals rapidly and reproducibly for as many ligands as necessary. This study details our efforts to engineer the catalytic domain of human protein tyrosine phosphatase β (HPTPβ-CD) with properties suitable for rapid-turnaround crystallography. Structures of apo HPTPβ-CD and its complexes with several novel small-molecule inhibitors are presented here for the first time.

1. Introduction

Protein tyrosine phosphorylation is a post-translational modification that regulates many important biological processes, including embryonic patterning, tissue growth and repair, metabolism and angiogenesis (Andersen *et al.*, 2004; Dube & Tremblay, 2005; Kappert *et al.*, 2005; Ostman & Bohmer, 2001; Stoker, 2005; Zhang, 2002). The balance of protein tyrosine phosphorylation is maintained in part by the opposing actions of protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs). Importantly, disruption of this balance has been implicated in a variety of disease processes such as cancer, atherosclerosis and diabetes, making both PTKs and PTPs prime candidates for drug discovery. Like the PTKs, the PTPs comprise a large diverse family of proteins represented by over 90 verified genes in humans (Alonso *et al.*, 2004; Andersen *et al.*, 2004). Also like the PTKs, the PTP family includes receptor-type and cytoplasmic proteins with a variety of associated domains that are likely to play roles in ligand binding, cellular localization or regulation of enzymatic activity (Alonso *et al.*, 2004; Andersen *et al.*, 2004; Stoker, 2005). Currently, the most intensely studied member of this family has been PTP1B, a cytoplasmic PTP that negatively regulates insulin-receptor action (Elchebly *et al.*, 1999; Klaman *et al.*, 2000; Zinker *et al.*, 2002). Unfortunately, elucidation of the functional roles of additional PTPs has lagged behind progress on the PTKs. Future advances towards determining the biological function of additional members of the PTP family will no doubt increase interest in developing therapies designed to modulate PTP activity.

Although selective and potent PTK inhibitors have been developed and approved for clinical use (Ventura & Nebreda,

Table 1
Phosphatase construct details.

Construct	N-terminus	Loop 1746–1757	Loop 1849–1857	C-terminus	Crystallized
CD-1	1640 LICRQ...	LSNVDDDPCSDY	CGEEQLDAH	...VYSRH 1997	No
CD-2	1680 SCPIK...	LSNVDDDPCSDY	CGEEQLDAH	...VYSRH 1997	No
CD-3	1640 LICRQ...	LSNVDDDPCSDY	CGEEQLDAH	...RSEGE 1975	Yes
CD-4	1676 NRKTS...	LSNVDDDPCSDY	CGEEQLDAH	...ARKLR 1970	Yes
CD-4-Δ1	1676 NRKTS...	LSGGSDY	CGEEQLDAH	...ARKLR 1970	Yes
CD-4-Δ2	1676 NRKTS...	LSNVDDDPCSDY	CGAGGAH	...ARKLR 1970	Yes

2006), the development of pharmaceutically suitable PTP inhibitors has proven difficult. Structure-based drug design was critical for the development of clinically useful PTK inhibitors and has also been a useful tool for understanding the enzymatic mechanism of PTPs, as well as for early attempts at inhibitor development. For example, elucidation of the molecular mechanism for the dephosphorylation of the insulin receptor by PTP1B was facilitated by studying the interaction of PTP1B with the diphosphorylated activation loop of the insulin receptor (Salmeen *et al.*, 2000). Moreover, X-ray crystallography has been utilized by a number of industrial and academic researchers (Andersen *et al.*, 2000, 2002; Iversen *et al.*, 2000; Burke & Zhang, 1998; Rye & Baell, 2005) to understand and optimize the potency and selectivity of PTP1B inhibitors. Clearly, additional structural information will help to advance our understanding of this critically important family of enzymes.

HPTPβ is a receptor-type phosphatase that is expressed primarily in endothelial cells that form the protective lining of blood vessels (Fachinger *et al.*, 1999; Krueger *et al.*, 1990; Wright *et al.*, 2000). In cultured endothelial cells, HPTPβ expression increases markedly, suggesting a role for this phosphatase in density-dependent growth arrest and/or maintenance of the quiescent endothelial phenotype (Gaits *et al.*, 1995). Further supporting an important role in vascular biology, HPTPβ associates with and regulates the phosphorylation of Tie2, a receptor PTK for the angiopoietin family of polypeptide growth factors (Fachinger *et al.*, 1999). Targeted disruption of the HPTPβ catalytic domain (CD) in transgenic mice results in embryonic lethality owing to abnormalities in the development of the cardiovascular system (Baumer *et al.*, 2006). Together, these data demonstrate that HPTPβ plays a crucial role in blood-vessel growth and maintenance and suggest that it may be a target for therapy of cardiovascular disease.

Phosphatases can be considered challenging targets (Tautz *et al.*, 2006): the hydrophilic and charged phosphate group of their natural substrates is difficult to mimic without compromising the membrane permeability of the inhibitor. Nevertheless, based on our recent experiences, we believe that membrane-permeable phosphatase inhibitors can be produced, although their oral availability is difficult, maybe even impossible, to attain.

To facilitate the development of potent and selective inhibitors, we undertook an effort to solve the crystal structure of HPTPβ. The ability to solve multiple structures of the protein of interest with a number of inhibitor candidates is a great advantage to the early drug-discovery process because it enables the designers to develop a comprehensive structure–activity relationship. The more rapid the turnaround between cycles of compound design, activity evaluation and structure determination, the more successful and efficient is the structure-based inhibitor-discovery process. Therefore, for pharmaceutical structural studies two factors are of paramount importance: (i) that individual protein–ligand complex structures be obtained rapidly and (ii) that the protein–ligand structures be obtained routinely with any given high-affinity ligand. In this paper, we illustrate stepwise protein engineering that enabled the development of an HPTPβ-CD construct with greatly improved crystallization properties that fulfill the two criteria for optimal structure-based drug design. We also describe for the first time the structure of HPTPβ-CD and provide a brief overview of the fundamental interactions of this phosphatase with some representative ligands, including a novel class of HPTPβ inhibtors.

2. Materials and methods

2.1. Initial structure of HPTPβ-CD

2.1.1. Cloning and expression. Catalytic domain residues of HPTPβ were identified by analyzing multiple sequence alignments with phosphatases of known structure (*BLAST*; Altschul *et al.*, 1990). DNA encoding the catalytic domain region was PCR-amplified from a human kidney cDNA library using standard procedures. The sequence of the clone was verified and PCR artefacts were corrected by mutagenesis (QuikChange, Stratagene) as needed.

Based on the work of Wang & Pallen (1992) as well as on sequence alignments, the first-generation HPTPβ-CD constructs contained protein residues 1640–1997 (CD-1), 1680–1997 (CD-2) and 1640–1975 (CD-3) (see Table 1 for a summary of all constructs). Relevant portions of the gene were amplified by PCR and inserted into pET-15b (N-terminal His tag) or pET-28a (C-terminal His tag) vector backbones using standard restriction-enzyme-based cloning. Constructs were verified by sequencing of the insert regions. When these constructs were expressed in BL21(DE3) (Stratagene) *Escherichia coli* a significant increase in phosphatase activity of the clarified cell lysates was observed, but the amount of soluble protein was not sufficient to warrant large-scale purification.

For improved functional phosphatase domain expression the same portions of the gene were inserted into a pMAL-c2

vector backbone. The resulting fusion proteins contained a thrombin-cleavage site separating the maltose-binding protein (MBP) and the HPTPβ-CD. Expression of these constructs in BL21 *E. coli* grown at 310 K in Luria Broth (LB) and induced at 295 K for 16 h afforded soluble fusion proteins with high levels of tyrosine phosphatase activity (based on PNPP assay; Bessey *et al.*, 1946) for MBP-thrombin-CD-1 and MBP-thrombin-CD-3, whereas MBP-thrombin-CD-2 had virtually no activity and was only marginally soluble.

Starter cultures of MBP-thrombin-CD-1 or MBP-thrombin-CD-3 in BL21 *E. coli* were grown from glycerol stocks in Luria Broth supplemented with 0.4% glucose overnight. These cultures were used to inoculate 8 l Biostat-C fermenters loaded with the same medium. Fermentation was carried out at 310 K with 700 rev min^{-1} agitation and two vessel volumes per minute (vvm) air in pH-stat mode (pH 7.2) until the optical density of the cultures reached $\sim$3 A_{600} units, at which point IPTG was added to 1 m*M* and the temperature was lowered to 295 K overnight. Bacterial cells were harvested by centrifugation and stored at 193 K.

2.1.2. Purification and crystallization. Both recombinant proteins, MBP-thrombin-CD-1 and MBP-thrombin-CD-3, were purified in the same manner: the bacterial cell pellet (100 g) was resuspended in 300 ml 250 m*M* NaCl, 25 m*M* Tris–HCl pH 8.0, 0.05% β-mercaptoethanol (BME) supplemented with EDTA-free protease-inhibitor cocktail (Boehringer-Mannheim). Cells were lysed *via* two passages through a French press (69 MPa) and the resulting lysate was clarified by centrifugation (25 000*g* for 30 min). The supernatant was loaded onto an amylose affinity column (Qiagen) which was washed with the same buffer until A_{280} stabilized. The protein was eluted with 20 m*M* maltose in the same buffer. Relevant fractions were pooled and supplemented with β-octyl glucoside (BOG) to a final concentration of 0.1%. Fusion protein was cleaved with thrombin (GE-Pharmacia) overnight at 277 K. It is important to note that in the absence of BOG the cleaved CD-1/CD-3 proteins precipitate. The cleavage reaction was stopped by addition of 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF) to 0.2 m*M* and diluted with 25 m*M* Tris pH 8.0 containing 0.1% BOG and 1 m*M* DTT until the conductivity of the solution was below 8 mS cm^{-1}. Cleavage products were separated on a Q-15 ion-exchange resin (GE-Pharmacia) using a gradient of NaCl from 20 m*M* to 1 *M* in the same buffer. Fractions were analyzed by SDS–PAGE and LC–MS. Fractions containing CD-1/CD-3 were pooled and concentrated in a stirred cell to 10 mg ml^{-1} (using a calculated specific A_{280}). Concentrated proteins were supplemented with DTT to 5 m*M*, flash-frozen in 100 µl aliquots in liquid nitrogen and stored at 193 K.

Crystallization of CD-1 and CD-3 was accomplished using commercial screens set up as 24-well hanging-drop experiments. Crystallization of CD-3 was observed with reservoir conditions containing PEG 2500–8000, varying concentrations of Mg salts and pH 6.5–8.5. Crystallization conditions were refined until suitably sized crystals (Fig. 1*a*) were obtained using crystallization drops of 3 + 3 µl (protein + reservoir) equilibrated against 1 ml reservoir containing 18–22% PEG 8000, 200 m*M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.0, 1% BME. Construct CD-1 mostly precipitated in crystallization and further studies were therefore not pursued.

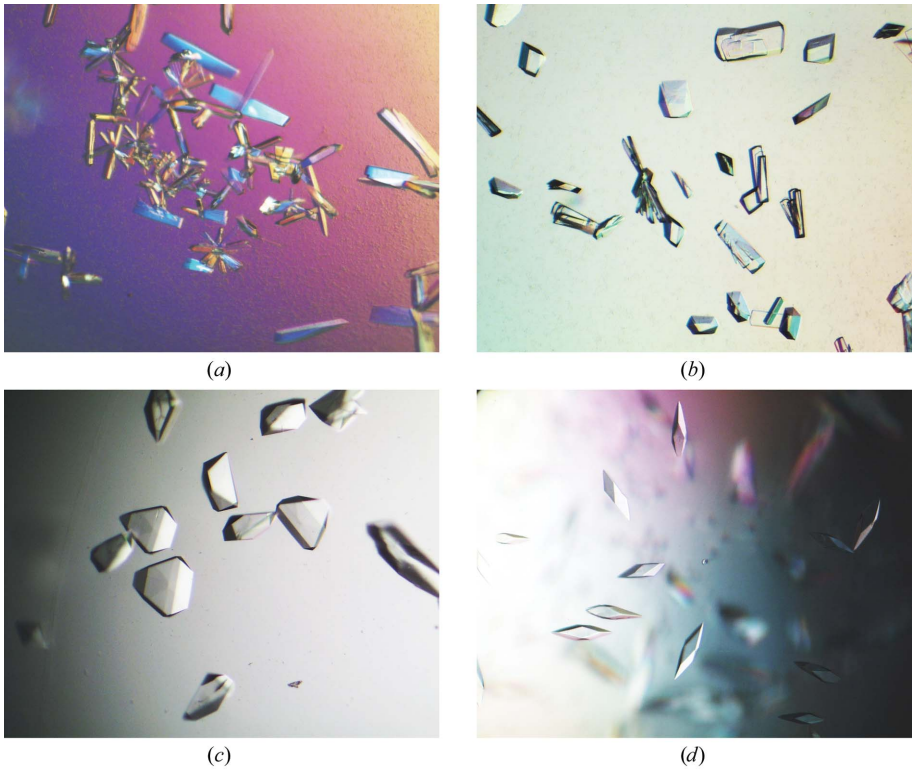

Figure 1
Various crystal forms of PTPβ-CD. (*a*) $P2_1$ crystals, construct CD-3. (*b*) $P2_12_12_1$ crystals, construct CD-4. (*c*) $P2_12_12_1$ crystals, construct CD-4-Δ1. (*d*) $C2$ crystals, construct CD-4-Δ1.

2.1.3. *Apo* CD-3 data collection, structure determination and refinement. The first CD-3 data set was obtained using a home source (Rigaku RU-300 generator equipped with MAR165 CCD). Crystals were mounted in Paratone-N using a monofilament loop and flash-frozen in a nitrogen stream at 100 K. Considerable twinning of crystals was observed: their X-ray diffraction patterns were characteristic of those produced by multiple independently oriented clusters. Changing the cryoprotection conditions did not result in any improvement in the diffraction; however, we were able to collect a non-twinned set of data to 2.4 Å in space group $P2_1$ by testing multiple crystals and carefully indexing the diffraction pattern (*HKL*-2000; Otwinowski & Minor, 1997). The unit cell contained two protein molecules. Another data set was subsequently obtained in space group $P2_12_12_1$ with one protein molecule in the unit cell.

Table 2
Essential data-collection and structure-refinement parameters for the representative structures.

Values in parentheses are for the highest resolution shell. Compound (1), vanadate (Figs. 2, 3*a* and 3*b*); compound (2), *p*-ethyl phenylsulfamic acid (Figs. 3*c* and 4); compound (3), *N*-(*t*-butoxycarbonyl)-L-tyrosyl-*N*-methyl-4-(sulfoamino)-L-phenylalanylamide (Fig. 5).

PDB code	2i3r	2i3u	2hc2	2hc1	2i4e	2i5x	2i4g	2i4h
Construct	CD-3	CD-3	CD-4-Δ1	CD-4-Δ1†	CD-3	CD-3	CD-3†	CD-3
Ligand	—	—	—	—	(1)‡	(2)‡	(3)‡	(3)§
Beamline	17-ID	17-ID	22-ID	22-ID	22-ID	22-ID	22-ID	22-ID
Space group	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$	$C2$
Unit-cell parameters								
a (Å)	61.86	38.75	38.31	39.06	61.97	62.06	38.55	113.31.98
b (Å)	71.64	69.78	70.78	70.71	71.30	71.72	69.60	38.79
c (Å)	70.53	118.52	49.85	118.45	69.69	69.85	117.78	66.98
α (°)	90	90	900	90	900	90	90	900
β (°)	93.58	90	92.52	90	93.40	93.31	90	104.94
γ (°)	90	90	90	90	90	90	90	90
Resolution (Å)	30–1.90 (2.00–1.90)	30–1.85 (1.95–1.85)	30–1.40 (1.50–1.40)	30–1.35 (1.45–1.35)	30–1.70 (1.80–1.70)	30–1.75 (1.85–1.75)	30–1.90 (2.00–1.90)	30–2.15 (2.25–2.15)
No. of reflections	47833 (6862)	77971 (3939)	48397 (6761)	53852 (10720)	66068 (9811)	59776 (8987)	20756 (3222)	15799 (2150)
Redundancy	2.8 (2.2)	3.5 (2.6)	4.2 (2.1)	2.6 (1.3)	3.7 (2.1)	3.2 (2.0)	2.1(1.8)	3.2 (2.3)
$\langle I/\sigma(I)\rangle$	20.2 (7.6)	17.7 (3.2)	20.2 (4.87)	16.8 (5.8)	16.9 (3.6)	24.8 (7.1)	19.4 (8.7)	12.5 (5.3)
Completeness (%)	98.4 (99.9)	98.2 (93.3)	92.5 (69.4)	73.7 (77.3)	99.6 (97.9)	97.1 (97.7)	80.8 (88.2)	99.5 (97.2)
$R_{\rm merge}$	0.03 (0.13)	0.03 (0.20)	0.02 (0.22)	0.03 (0.30)	0.04 (0.27)	0.02 (0.13)	0.03 (0.11)	0.06 (0.19)
$R_{\rm cryst}$	0.19	0.18	0.13	0.13	0.17	0.18	0.17	0.17
$R_{\rm free}$	0.23	0.23	0.18	0.17	0.22	0.22	0.22	0.24
Protein residues	564	276	283	290	584	561	289	280
Ligand residues	0	0	0	0	2	2	1	1
Ions	2	0	1	3	0	0	0	2
Solvent residues	195	204	298	371	465	276	249	182
R.m.s.d., bonds (Å)	0.012	0.015	0.009	0.008	0.013	0.012	0.009	0.010
R.m.s.d., angles (°)	1.4	1.7	1.4	1.6	1.5	1.6	1.2	1.4
Ramachandran plot regions								
Preferred (%)	89.4	90.1	90.4	91.5	86.7	90.3	88.3	88.7
Allowed (%)	9.8	8.2	8.4	7.4	12.0	9.1	11.3	10.9
Generously allowed (%)	0.8	1.6	1.2	1.2	1.3	0.6	0.4	0.4

† Data set incomplete owing to diffraction anisotropy. ‡ Soak. § Cocrystal.

The structure was solved by molecular replacement (*AMoRe*; Navaza, 2001) using the catalytic domain of protein tyrosine phosphatase-μ (PDB code 1rpm; Hoffmann *et al.*, 1997) as the search model. Extensive rebuilding [*O* (Jones *et al.*, 1991) and *QUANTA* (Accelrys Inc.)] and refinement (*REFMAC*5; Murshudov *et al.*, 1997) were conducted until the parameters of the structure were within acceptable ranges. High-resolution native data sets were subsequently collected using synchrotron radiation as described in the following section. Relevant data-collection, refinement and structure parameters are summarized in Table 2. All the structures mentioned in this article were deposited with the RCSB (see Table 2 for PDB codes).

2.1.4. PTPβ-CD-3 in complex with inhibitors: crystallization, data collection, structure determination and refinement. Our attempts to cocrystallize CD-3 with various prototype inhibitors were only successful for a few compounds. To complement the cocrystallization experiments, we also soaked various inhibitors into both the orthorhombic and the monoclinic crystals at a concentration of 1–10 m*M* in the crystallization buffer, using soak times of between 2 and 24 h. Soaking was only partially successful, since most of the compounds cracked or dissolved the protein crystals.

X-ray diffraction data were collected using cryogenically preserved (Paratone-N) native, cocrystallized and soaked crystals at Advanced Photon Source beamlines 17-ID and 17-BM (IMCA-CAT). Beamlines were tuned to the wavelength of 1.0 Å; data collection was conducted in 0.25 or 0.5° oscillations using 1 s deg^{-1} exposure, typically over a span of 220–300° in order to achieve good completeness and redundancy. Data were indexed and processed using the *HKL*-2000 (Otwinowski & Minor, 1997) suite of programs. The structures of complexes were solved *via* molecular replacement using the corresponding native structures for each crystal form as search models. Several rounds of manual rebuilding (*Coot*; Emsley & Cowtan, 2004) and refinement (*REFMAC*5; Murshudov *et al.*, 1997) were employed, after which the geometry of the structures was analyzed using *SFCHECK* (Vaguine *et al.*, 1999). The relevant data-collection and structure-solution parameters are presented in Table 2.

2.2. HPTPβ-CD variants with improved properties

2.2.1. Cloning and expression. The crystal structure of CD-3 was used to further optimize a new construct for crystallization. Since the termini of the CD-3 protein were found to be disordered, a second-generation amplicon containing HPTPβ residues 1676–1973 was obtained by PCR from CD-3 using primers that encoded the N-terminal TVMV protease-cleavage sequence (ETVRFQ|S; Nallamsetty *et al.*, 2004; Tozser *et al.*, 2005) flanked by the *att*B recombination sites of the Gateway cloning system (Invitrogen). This amplicon was

recombined with the donor vector pDONR-221 to generate the entry clone pENTR-TVMV-HPTPβ-1676-1970 (CD-4). Two additional entry clones harbouring substitutions and deletions in disordered loops were produced *via* QuikChange (Stratagene) as follows: the pENTR-TVMV-HPTPβ-1676-1970-Δ1 (CD-4-Δ1) clone has the 1746–1757 loop residues **LS**NVDDDPCS**DY** mutated to **LS**GGS**DY** (intact residues are in bold), whereas the pENTR-TVMV-PTPβ-1676-1970-Δ2 (CD-4-Δ2) clone has the 1849–1857 loop residues **CG**EEQLD**AH** mutated to **CG**AGG**AH**. These sequence substitutions were intended to convert flexible loops into compact turns. All entry clones were verified *via* sequencing. Relevant details of all the constructs are summarized in Table 1.

In order to generate the N-terminally His-tagged MBP fusions of the three new catalytic domain variants, we recombined the respective entry clones with an in-house destination vector pDEST-HisMBP which has a His_6 tag inserted after the first three residues of the mature form of *E. coli* MBP, followed by the *att*R1-CmR-ccdB-*att*R2 recombination cassette (the same as the recently described dual-tag destination vector developed by Waugh, 2005). The LR reaction (the final stage of Gateway recombinational cloning) was performed according to the manufacturer's recommendations and the recombination products were transformed into TOP10 (Invitrogen) and BL21 (Stratagene) *E. coli* cells for expression.

Small-scale testing for active protein expression was conducted as follows: 5 ml cultures were grown in Terrific Broth II (QBiogene) supplemented with 1% glycerol, 0.4% glucose and 100 mg l^{-1} carbenicillin at 310 K until the optical density of the cultures reached $\sim$1 A_{600} unit. Induction of protein expression was studied using a matrix of conditions and the cultures were induced with either IPTG (0.1, 0.5 or 1.5 m*M*) or lactose (0.1, 0.3 or 1.0%) using induction temperatures of 289, 295 and 310 K for either 4, 8 or 16 h. Soluble protein extracts were prepared using detergent lysis with B-PER (Pierce) supplemented with bovine DNAse-I, hen egg-white lysozyme and proteinase-inhibitor cocktail (Sigma). His-tagged proteins were purified from these extracts *via* IMAC using 1 ml Zeba columns (Pierce) containing 50 µl His-Select HF resin (Sigma). The columns were washed twice with 1 ml of 250 m*M* NaCl, 25 m*M* Tris pH 7.8, 10 m*M* imidazole and eluted with 60 µl of the same buffer supplemented with 200 m*M* imidazole. The protein content of the eluates was visualized by SDS–PAGE stained with SimplyBlue SafeStain (Invitrogen). Estimates of protein abundance were obtained *via* scanning densitometry. Optimum expression conditions for all three constructs were found to be TOP10 cells induced for 16 h at 289 K using 0.5–1.0 m*M* IPTG.

2.2.2. Fermentation, purification and crystallization. All second-generation constructs were fermented, purified and crystallized using the same protocol, described below.

Starter cultures of the His-MBP-TVMV-CD-4, His-MBP-TVMV-CD-4-Δ1 or His-MBP-TVMV-CD-4-Δ2 constructs in TOP10 *E. coli* were grown from glycerol stocks in Terrific Broth II supplemented with 1% glycerol, 0.4% glucose and 100 mg l^{-1} carbenicillin overnight. These cultures were used to inoculate 2 l Biostat-B or Biostat-MD fermenters loaded with the following medium: 1.2$\times$ Terrific Broth II, 2.5% glycerol, 0.8% glucose and 120 mg l^{-1} carbenicillin. Fermentation was carried out at 310 K with 700 rev min^{-1} agitation and 2 vvm air in pH-stat mode (pH 7.2) until the optical density of the cultures reached 8–11 A_{600} units (typically 6–7 h), at which point IPTG was added to 1 m*M* and the temperature was lowered to 291 K overnight. Bacterial cells were harvested by centrifugation and stored at 193 K.

Protein purification was accomplished as follows. 100 g frozen cell paste was suspended in 200 ml B-PER (Pierce) supplemented with 1 mg ml^{-1} DNAse-I and 0.1 mg ml^{-1} hen egg-white lysozyme (Sigma), as well as an EDTA-free proteinase-inhibitor cocktail (Sigma). After complete breakdown of the frozen cell pellet, the suspension was stirred for an additional 20 min at 277 K and then clarified by centrifugation (30 min at 25 000*g*). Clarified supernatant was supplemented with 10 m*M* imidazole and passed through a 40 ml His-Select HF column. The column was washed with 250 m*M* NaCl, 25 m*M* Tris pH 7.8, 10 m*M* imidazole until the OD_{280} stabilized and was then eluted with an imidazole concentration gradient from 10 to 200 m*M* in the same buffer. Relevant fractions were pooled and supplemented with BOG and DTT to final concentrations of 0.1% and 2 m*M*, respectively. CD-4, CD-4-Δ1 or CD-4-Δ2 were cleaved off the His-MBP fusion protein by means of overnight treatment with recombinant TVMV protease used at 1 mg per 50 mg of fusion protein at 277 K. The vector (pRK1035, developed by Dr Waugh, NCI-FCRF) encoding this protease was purchased from http://www.ScienceReagents.com. The cleavage reaction was diluted with 25 m*M* Tris pH 8.0 containing 0.1% BOG and 1 m*M* DTT until the conductivity of the solution was below 8 mS cm^{-1}. Cleavage products were separated on a Q-15 ion-exchange resin using a gradient of NaCl from 20 m*M* to 1 *M* in the same buffer. Fractions were analyzed by SDS–PAGE and LC–MS. Fractions containing the proteins of interest were pooled and concentrated to 10 mg ml^{-1} (by theoretical specific A_{280}). Concentrated proteins were supplemented with DTT to 5 m*M*, flash-frozen in 100 µl aliquots in liquid nitrogen and stored at 193 K. The final protein yield averaged $\sim$120 mg per 100 g of frozen cell paste. The proteins were monomeric according to the results of gel-filtration and dynamic light-scattering experiments (data not shown).

Crystallization conditions were established in the same way as for the first-generation constructs and were refined until suitably sized crystals (Fig. 1*c*) were obtained using crystallization drops of 3 + 3 µl (protein + reservoir) equilibrated against 1 ml reservoir containing 21% PEG 8000, 220 m*M* $MgCl_2$, 100 m*M* Tris–HCl pH 8.0, 1% BME. Depending on the inhibitor used for cocrystallization, additional optimization of crystallization conditions was often necessary. To this end, we have developed a subscreening 3 $\times$ 3 matrix of conditions as follows: each condition was comprised of a 1 ml aliquot of the solution described above, supplemented in one direction with 0, 100 or 200 µl 1 *M* sodium citrate buffered to pH 5 with citric acid and in the other direction with 0, 100 or 300 µl 7 *M*

ammonium acetate buffered to pH 8.0 with ammonium hydroxide. Hanging drops containing 3 µl of protein–inhibitor complex and 3 µl of the reservoir solution were set up against each of the matrix conditions. Almost invariably, one or more of these conditions produced crystals of sufficient size and quality.

2.2.3. Data collection, structure determination and refinement. Crystals of CD-4, CD-4-Δ1 or CD-4-Δ2 were mounted in a monofilament loop using Paratone-N and flash-frozen in a nitrogen stream at 100 K. Data collection was conducted both at home (MacScience generator equipped with a Bruker SMART-6000 CCD) and at the synchrotron (beamlines 22-ID and 22-BM, SER-CAT, APS as described above for the CD-3 construct). Of the three new constructs, CD-4 and CD-4-Δ2 showed twinning behavior similar to that of the original construct CD-3, although the severity and frequency of twinning was considerably reduced. The CD-4-Δ1 construct produced mostly non-twinned cocrystals with essentially every compound attempted. Crystals typically diffracted to 1.5–1.7 Å resolution, with some of the compounds producing complex crystals diffracting to better than 1.2 Å resolution. Therefore, this construct was used for the remainder of the study. The relevant data-collection and structure-solution parameters are presented in Table 2.

2.2.4. NMR studies. We obtained ^{15}N-labeled CD-3 protein by growing the respective construct in TOP10 *E. coli* using an M9 minimal medium with $^{15}NH_4Cl$ as the sole source of nitrogen. The concentrations of glycerol and glucose in the medium were 1 and 0.4%, respectively. Fermentation and purification procedures did not significantly differ from those described above, except that recombinant protein induction was initiated at an OD_{600} of 1–2 and the final yield of *E. coli* biomass was only 25–30 g per litre of medium owing to nutrient limitation. ^{15}N–^{1}H heteronuclear correlation spectra were collected on a Varian-500 instrument using standard pulse programs. Protein concentration was kept at 10 mg ml^{-1} in perdeuterated Tris buffer pH 7.5, 100 m*M* NaCl, 0.1% BOG and 10% percent D_2O for internal lock standard.

3. Results and discussion

3.1. Design and crystallization of the initial constructs

Successful crystallization of a specific domain of a multi-domain protein can be greatly facilitated by leveraging pre-existing information about domain boundaries in the target or its close relatives. Unfortunately, experience shows that theoretical knowledge of domain boundaries can be insufficient to design and express soluble protein owing to the fact that protein folding is unpredictable, especially in the case of heterologous expression. Fortunately, the PTPs are a well characterized protein family with numerous examples of crystal structures reported (see supplementary material[1]). The high degree of sequence and structure conservation between the PTPs makes it fairly easy to determine catalytic domain boundaries with a reasonable expectation of success. By combining the information published by Wang & Pallen (1992) with our own sequence alignments, we designed three first-generation HPTPβ-CD constructs, two of which expressed active protein, and one of the two active variants (CD-3) was crystallized (Table 1). High levels of soluble protein were obtained by expressing HPTPβ-CDs as fusion proteins with the mature form of *E. coli* MBP, whereas expression with only a His tag at either the C- or N-terminus of the protein was insufficient to generate useful amounts of soluble protein. The use of MBP fusions has been previously demonstrated to be a particularly successful tool for obtaining soluble properly folded recombinant proteins in *E. coli* (Fox & Waugh, 2003; Nallamsetty *et al.*, 2005), although the underlying reasons for such improvement remain to be elucidated (Nallamsetty & Waugh, 2006).

The expression, purification and crystallization of CD-3 were fairly straightforward: no unusual reagents or conditions were required and the protein was well behaved and stable provided that a reducing environment was maintained. The first crystal structures were produced within approximately one month from the cloning of the three first-generation constructs and individual crystallization experiments were optimized to yield useful crystals within a few days, thus satisfying the requirement for rapid turnaround between compound availability and cocomplex crystals as formulated in §1 (Fig. 1*a*). The CD-3 crystals diffracted very well at the synchrotron (1.8–1.5 Å) and reasonably well at home (2.2–2.4 Å).

3.2. Crystal structures of HPTPβ-CD

PTPs from eukaryotes as well as bacterial pathogens employ the same mechanism for recognition and cleavage of

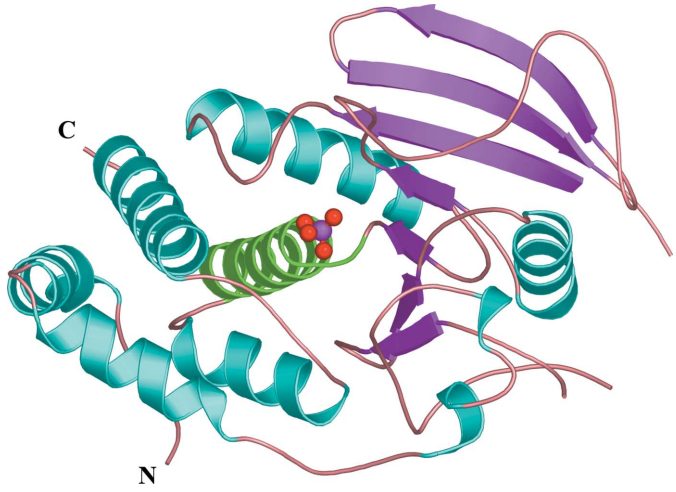

Figure 2
Overall view of the HPTPβ-CD structure. α-Helices and β-sheets are colored blue and magenta, respectively, with the exception of the helix (1909–1926) proposed to be involved in ligand binding, which is colored green. The position of the active site is indicated by the ligand (vanadate ion), which is shown as a ball-and-stick model.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: SX5059). Services for accessing this material are described at the back of the journal.

tyrosine phosphate (pTyr; see, for example, Zhang, 2002). Therefore, it is not surprising that the crystal structure of HPTPβ-CD closely resembles those of other tyrosine phosphatases; in fact, the C^{α} r.m.s.d. between different receptor-type PTPs are in the same range (~1–1.5 Å) as the C^{α} r.m.s.d. between different structures of HPTPβ-CD. The structure of apo HPTPβ-CD recently released by the Structural Genomics Consortium (PDB code 2ahs; http://www.sgc.utoronto.ca/SGC-WebPages/StructureDescription/2AHS.php) almost exactly overlaps one of the apo forms solved by us (C^{α} r.m.s.d. of 0.7 Å; see supplementary material).

The overall structure of HPTPβ-CD is a typical representative of the higher molecular-weight PTP fold, comprising a curled eight-stranded β-sheet that wraps around a short α-helix and is flanked by six additional α-helices on the other side (Fig. 2). The active site is positioned in the cleft between the β-rich and the α-rich portions of the protein. Like other metal-independent protein phosphatases (both tyrosine and dual-specificity), the HPTPβ active site is shallow, with the phosphate-binding element formed by the horseshoe-shaped backbone of residues 1904–1910 and by the side chain of Arg1910 (Fig. 3*a*). A long α-helix (1909–1926) is positioned directly under the bound phosphate, suggesting a role in neutralization of the negative charge(s) by the partial positive charge associated with the N-terminal end of this α-helix (Dillet *et al.*, 2000).

The side chain of the catalytic Cys1904 is located directly under the centre of the horseshoe; this position puts it on the line connecting the ester O atom and the phosphorus of pTyr and enables the cysteine residue to act as a nucleophile (Fig. 3*b*). The phosphate-binding site is flanked by residues Tyr1733, Asn1735, Ile1736 and Gln1948 (Fig. 3*c*), which form hydrophobic and hydrophilic interactions with the remainder of pTyr. Similar to other PTPs, the active site of HPTPβ is capped by a flexible loop (commonly referred to as the WPD-loop owing to its signature sequence) that carries the catalytic aspartic acid residue Asp1870. We have solved both the open and the closed forms of HPTPβ-CD and were able to observe the transition of the WPD-loop as shown in Fig. 4.

In general, the structural rearrangement of the enzyme between the two forms closely resembles that of PTP1B: upon binding of pTyr or a pTyr-mimetic inhibitor the WPD-loop undergoes conformational change that allows His1871 to form a π-stack with the pTyr phenyl ring. The change is accompanied by conformational and positional changes extending several residues into the core of the protein on each side of the WPD-loop. As a result, the catalytic Asp1870 is delivered into the active site, where it can play a general acid role: protonation of the leaving tyrosyl group and activation of a water molecule responsible for the cleavage of the phospho-cysteine covalent intermediate (Pannifer *et al.*, 1998). Unlike PTP1B, the side chain of the tryptophan residue of the WPD-loop (Trp179 in PTP1B and Trp1868 in HPTPβ) does not flip upon transition from open to closed form; instead, it moves ~1.5 Å towards the core of the protein (Fig. 4). Also unlike in PTP1B, we did not observe ordering of the C-terminal segment of HPTPβ-CD upon ligand binding.

Similar to PTPμ (Hoffmann *et al.*, 1997), the structure and solution behavior of HPTPβ-CD do not indicate that it may be transiently inhibited *via* dimerization of catalytic domains, which further supports the opinion that catalytic domain dimerization is not a common regulatory mechanism in receptor PTPs.

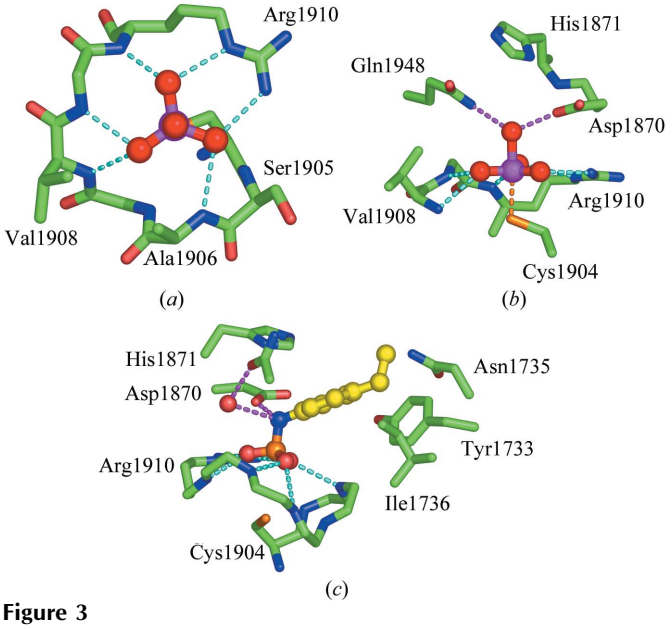

Figure 3
Detailed view of the PTPβ-CD active site. (*a*) Hydrogen-bonding network (blue dashes) bridging the ligand (vanadate ion, balls-and-stick representation) and the protein (sticks). (*b*) Side view depicting additional interactions (magenta dashes) between the ligand and polar residues in the active site. The partial covalent bond between Cys1904 and the V atom is shown as an orange dashed line. (*c*) Side view of a pTyr mimetic (*p*-ethyl phenylsulfamic acid) bound in the active site of the enzyme. Relevant hydrogen bonds are shown as colored dashes.

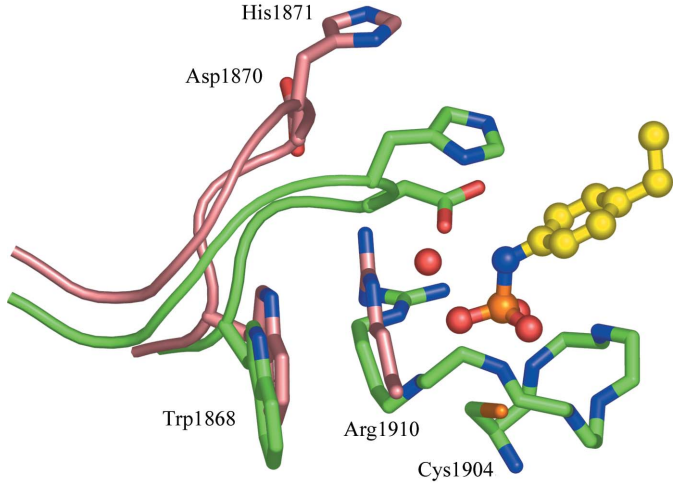

Figure 4
Transition of the WPD-loop between open (light brown) and closed (green) states. The ligand (sulfamic acid, yellow) is shown in ball-and-stick representation and relevant protein residues are shown as sticks colored according to the loop state. The catalytically important water molecule is shown as a red sphere.

3.3. Optimizing HPTPβ-CD crystallography as a tool for structure-based drug discovery

3.3.1. Drawbacks of the CD-3 construct. In the course of structure-based inhibitor-discovery process it is not uncommon to produce crystal structures of complexes by the dozen and in many cases by the hundred. It is always tempting to rely solely on a soaking procedure in order to minimize the time and effort required per structure. Our experience indicates that soaking, while undoubtedly the quickest route to cocrystals, may not be adequate for proteins with shallow active sites and for large ligands where crystal packing has a high likelihood of generating artefacts. As we found out, the CD-3 protein was not suitable for either soaking or cocrystallization procedures.

While the initial crystallization of CD-3 was relatively straightforward, we soon discovered that this construct had two significant limitations. Firstly, the crystals grow almost exclusively as multiple twins (Fig. 1*a*). Nevertheless, we were able to determine several structures of CD-3 in the apo form as well as in complex with several small-molecule ligands. However, a second and more serious drawback of CD-3 was discovered: this construct could be cocrystallized with only three inhibitors. Soaking of CD-3 crystals with small molecules allowed us to solve four additional structures, but this was not a robust solution because the inhibitors often destroyed the crystals and with larger compounds we became increasingly concerned that crystal-packing effects produce artefactual binding modes. Consideration of potential structural differences between soaking and cocrystallization of the same complex is important for structure-driven drug discovery (Danley, 2006) and indeed such effects have been observed previously (Zhu *et al.*, 1999; Hiller *et al.*, 2006). We report perhaps the most dramatic example of this behavior: one of our early compounds binds in a radically different way depending on whether it is soaked into apo CD-3 crystals or is cocrystallized with the protein (Fig. 5). Not only do the ligand conformation and position change, but the protein residues involved in ligand binding are also affected. We have observed two other cases of significant differences between soaked and cocrystallized structures of the HPTPβ-CD–ligand complexes, which we cannot describe here because of the proprietary nature of the compounds involved.

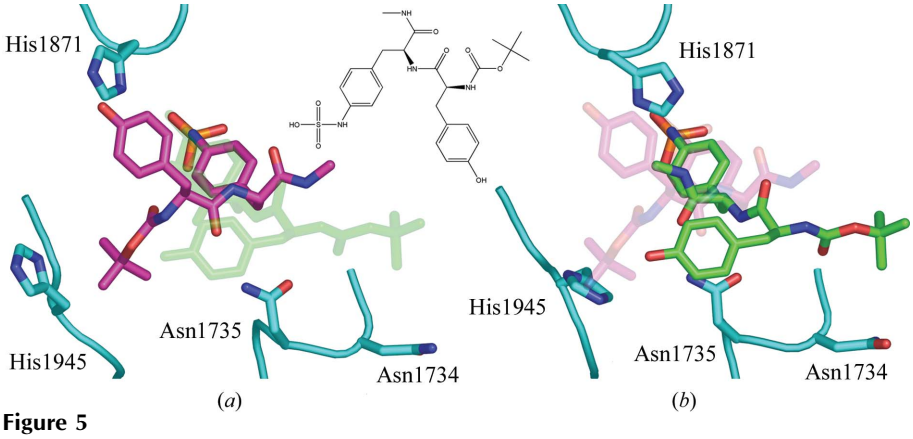

Figure 5
Radical differences in binding of the ligand depending on the method used for cocrystal production. Protein residues are shown as blue sticks. (*a*) Soaking: ligand is shown as magenta-colored sticks. (*b*) Cocrystallization: ligand is shown as green-colored sticks. In both cases the alternative position of the ligand is shown as transparent sticks.

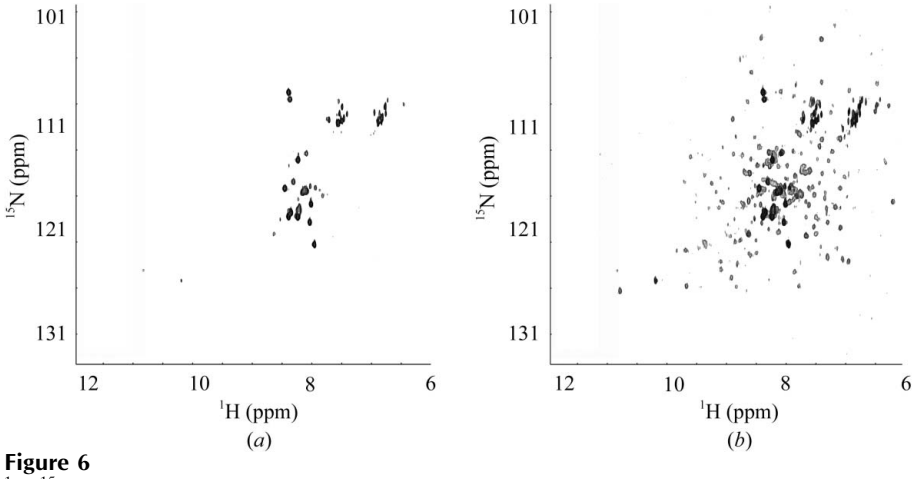

Figure 6
^{1}H–^{15}N correlation NMR spectra of PTPβ-CD (*a*) in the absence of an inhibitor and (*b*) in the presence of tightly binding sulfamic acid inhibitor (the same as in Fig. 5).

All in all, the requirement for routine availability of numerous crystalline CD-3–ligand complexes was clearly not satisfied and therefore CD-3 crystals were suboptimal for efficient structure-based drug design.

3.3.2. Engineering of PTPβ-CD for efficient structure-based drug design. To alleviate the crystallization issues inherent to the CD-3 construct, we decided to pursue a protein-engineering approach because repeated attempts to change crystallization conditions and other protein-independent parameters did not result in a consistent improvement. Several structures of apo and ligand-bound CD-3 were available for analysis. We also evaluated the protein by NMR: the quality of the ^{15}N–^{1}H spectrum of the protein in the apo state was exceptionally poor, but the spectrum improved greatly upon the addition of any potent ligand (Fig. 6). Finally, we performed a limited proteolysis experiment with CD-3 in the presence and the absence of a potent inhibitor. In the presence of the inhibitor the protein was protected, whereas in the absence of ligand CD-3 was rapidly degraded (data not shown).

Given the information above, we focused our attention on disorder: in all of the structures we observed that the electron density was missing for a

number of residues at the N- and the C-termini (proteolysis was ruled out by mass spectroscopy). There were also two disordered stretches of polypeptide: the 1746–1757 loop (LSNVDDDPCSDY) and the 1849–1857 loop (CGEEQLDAH). The situation with the first loop was particularly interesting because in the predominant CD-3 crystal form ($P2_1$, two protein molecules in asymmetric unit) this disordered loop is near the interface between symmetry-related protein molecules (Fig. 7*a*). We speculated that such disorder in one of the major crystal contacts was a key contributor to the persistent twinning observed for CD-3 crystals and that eliminating this disorder may be the key to changing the crystal form.

New constructs were generated by PCR and PCR mutagenesis (Table 1). Construct CD-4 had the disordered termini truncated; it contained HPTPβ residues 1676–1970. The two mutants of this construct (CD-4-Δ1 and CD-4-Δ2) contained replacements of the mobile loops with shorter turn-like sequences (Table 1). Expression, purification and crystallization of the three new constructs were straightforward. Catalytic activity and compound selectivity were measured for the new constructs and found not to vary significantly compared with those of CD-3 (data not shown). The two crystal forms of the CD-4 construct were identical to the $P2_1$ and $P2_12_12_1$ forms of CD-3. The tendency to produce multiple twins was slightly reduced, but this construct was still difficult to crystallize with more than only a few useful compounds. The CD-4-Δ2 construct likewise did not constitute a major improvement over CD-3.

The real breakthrough was the CD-4-Δ1 construct: its tendency towards twinning was essentially abolished and it was readily cocrystallized with nearly all HPTPβ inhibitors to

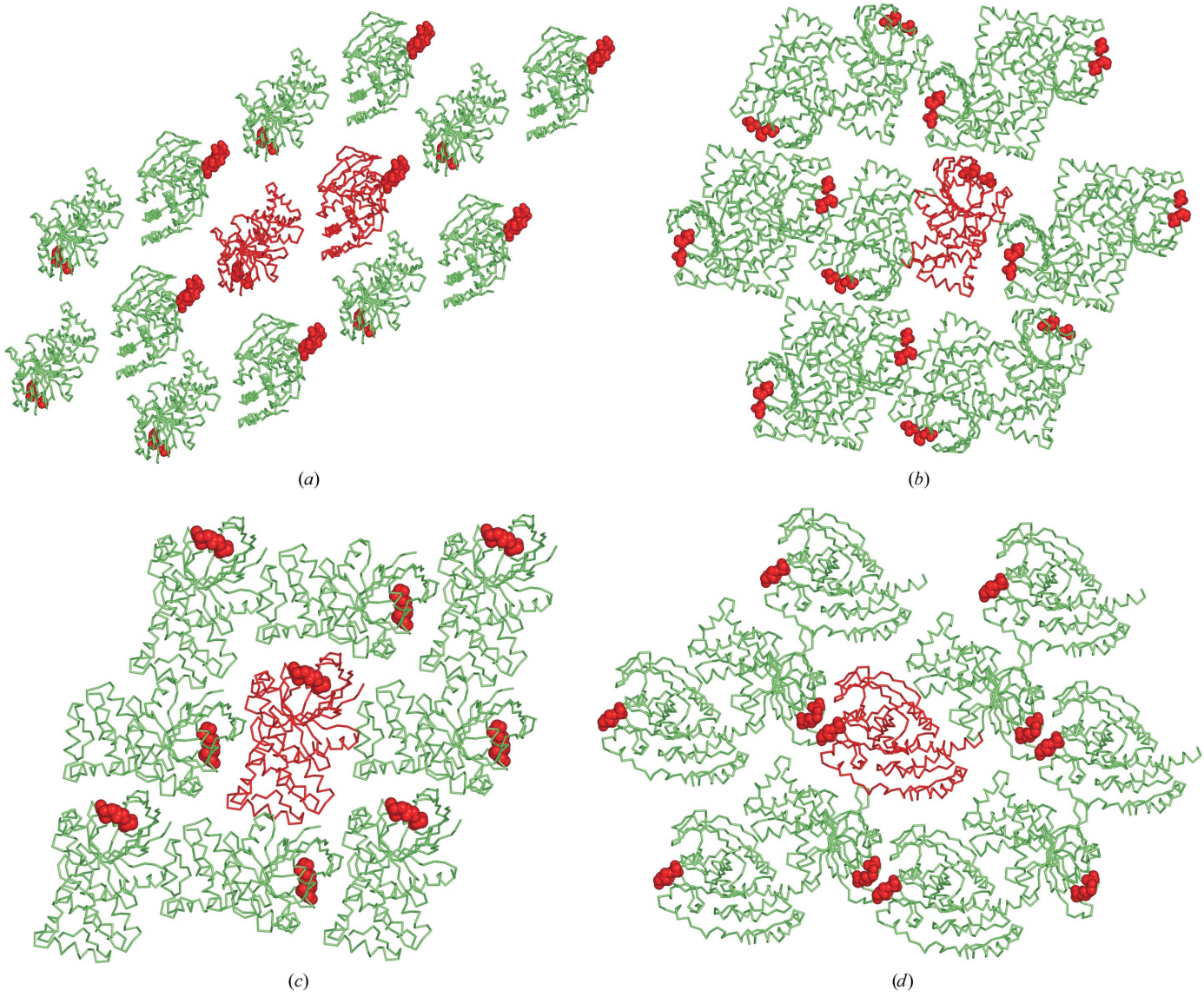

Figure 7
Packing modes of various HPTPβ-CD crystal forms: (*a*) $P2_1$ crystals, CD-3 construct. (*b*) $P2_12_12_1$ crystals, CD-3 construct. (*c*) $P2_1$ crystals, CD-4-Δ1 construct. (*d*) $C2$ crystals, CD-4-Δ1 construct. The asymmetric unit contents and symmetry mates are shown as red and green ribbons, respectively. Residues of the 1746–1757 loop are shown as red spheres.

which it was exposed. Well diffracting crystal forms in several space groups ($P2_1$, $P2_12_12_1$, $C2$; Table 2; Figs. 1 and 7) were readily obtained. The GGS residues replacing the disordered part of loop 1746–1757 adopted a stable and ordered turn conformation and were often located near the crystallographic contact interfaces (Figs. 7*c* and 7*d*). The structure of the rest of the protein was essentially unperturbed. The CD-4-Δ1 construct was therefore used for cocrystallization of numerous inhibitors, which resulted in over 120 crystal structures.

Modification of the primary sequence of an enzyme can have a profound effect on its kinetic properties and on the affinity of ligands. Since the ultimate target of our drug-discovery efforts is the native human enzyme, we had to be careful to consider the implications of the design and use of truncated and altered constructs. The disordered loop that is replaced in the CD-4-Δ1 construct is positioned far from the active site; therefore, the impact of this substitution on enzyme activity and on ligand binding is likely to be minimal. Indeed, the structure and enzyme kinetic properties of this mutant are virtually identical to those of the original construct. Most importantly, as multiple rounds of compound design, chemical synthesis and cocrystallization were carried out, every successful generation of compounds had improved activity in the *in vitro*, cellular and *in vivo* assays (Amarasinghe *et al.*, 2006, and manuscripts in preparation). This indicates that our interpretation of data obtained using truncated and modified HPTPβ catalytic domain expressed in *E. coli* is relevant to the ultimate target of this work: the full-length HPTPβ expressed in living human tissue.

To summarize, the method exemplified above consists of finding disordered regions in the early protein structure and eliminating them *via* mutagenesis. In the example above a particular replacement sequence (GGS) was used, but we anticipate that other short turn-like sequences should be just as applicable as long as they are long enough to bridge the gap while being short enough to not be disordered themselves.

4. Conclusions

Elimination of localized protein disorder *via* primary sequence manipulation has allowed us to greatly improve the quality and quantity of HPTPβ-CD crystals. This optimization resulted in the rapid and efficient solution of over 100 complex structures that were used to drive structure-based design of potent and selective HPTPβ inhibitors.

During the course of pharmaceutically oriented structural studies, we frequently find the first structure of the protein target of interest to be inadequate for rapid inhibitor design, which optimally requires rapid production of multiple complex structures. Localized disorder is common in protein structure and the disordered elements are often found in the solvent-exposed regions, which at the same time are very likely to form crystal contacts. Therefore, as shown above, elimination of local disorder can result in a dramatic change in crystallization behavior. Armed with the early structure and with the modern arsenal of rapid DNA-mutagenesis and protein-expression techniques, we can quickly design and obtain multiple protein variants. Based on the example of HPTPβ-CD presented in this article, we propose that this form of rational protein design should be useful for optimizing the ability to rapidly generate multiple ligand-bound structures of pharmaceutically relevant proteins, enhancing the efficiency of structure-guided inhibitor optimization in this important new area of drug discovery. In the future, we intend to expand the scope and usefulness of this method by reducing the need for the initial crystal structure through the use of sequence-based disorder prediction and related techniques.

Data were collected at the SER-CAT 22-ID and IMCA-CAT 17-ID beamlines at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at http://www.ser-cat.org and http://www.imca.aps.anl.gov. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38.

References

Alonso, A., Sasin, J., Bottini, N., Friedberg, I., Friedberg, I., Osterman, A., Godzik, A., Hunter, T., Dixon, J. & Mustelin, T. (2004). *Cell*, **117**, 699–711.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.

Amarasinghe, K. K., Evdokimov, A. G., Xu, K., Clark, C. M., Maier, M. B., Srivastava, A., Colson, A. O., Gerwe, G. S., Stake, G. E., Howard, B. W., Pokross, M. E., Gray, J. L. & Peters, K. G. (2006). *Bioorg. Med. Chem. Lett.* **16**, 4252–4256.

Andersen, H. S., Iversen, L. F., Jeppesen, C. B., Branner, S., Norris, K., Rasmussen, H. B., Moller, K. B. & Moller, N. P. (2000). *J. Biol. Chem.* **275**, 7101–7108.

Andersen, H. S. *et al.* (2002). *J. Med. Chem.* **45**, 4443–4459.

Andersen, J. N., Jansen, P. G., Echwald, S. M., Mortensen, O. H., Fukada, T., Del, V. R., Tonks, N. K. & Moller, N. P. (2004). *FASEB J.* **18**, 8–30.

Baumer, S., Keller, L., Holtmann, A., Funke, R., August, B., Gamp, A., Wolburg, H., Wolburg-Buchholz, K., Deutsch, U. & Vestweber, D. (2006). *Blood*, **107**, 4754–4762.

Bessey, O. A., Lowry, O. H. & Brock, M. J. (1946). *J. Biol. Chem.* **1**, 321–329.

Burke, T. R. Jr & Zhang, Z. Y. (1998). *Biopolymers*, **47**, 225–241.

Danley, D. E. (2006). *Acta Cryst.* D**62**, 569–575.

Dillet, V., Van Etten, R. L. & Bashford, D. (2000). *J. Phys. Chem. B*, **104**, 11321–11333.

Dube, N. & Tremblay, M. L. (2005). *Biochim. Biophys. Acta*, **1754**, 108–117.

Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A. L., Normandin, D., Cheng, A., Himms-Hagen, J., Chan, C. C., Ramachandran, C., Gresser, M. J., Tremblay, M. L. & Kennedy, B. P. (1999). *Science*, **283**, 1544–1548.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Fachinger, G., Deutsch, U. & Risau, W. (1999). *Oncogene*, **18**, 5948–5953.

Fox, J. D. & Waugh, D. S. (2003). *Methods Mol. Biol.* **205**, 99–117.

Gaits, F., Li, R. Y., Ragab, A., Ragab-Thomas, J. M. & Chap, H. (1995). *Biochem. J.* **311**, 97–103.

Hiller, N., Fritz-Wolf, K., Deponte, M., Wende, W., Zimmermann, H. & Becker, K. (2006). *Protein Sci.* **15**, 281–289.

Hoffmann, K. M., Tonks, N. K. & Barford, D. (1997). *J. Biol. Chem.* **272**, 27505–27508.

Iversen, L. F., Andersen, H. S., Branner, S., Mortensen, S. B., Peters, G. H., Norris, K., Olsen, O. H., Jeppesen, C. B., Lundt, B. F., Ripka,

W., Moller, K. B. & Moller, N. P. (2000). *J. Biol. Chem.* **275**, 10300–10307.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kappert, K., Peters, K. G., Bohmer, F. D. & Ostman, A. (2005). *Cardiovasc. Res.* **65**, 587–598.
Klaman, L. D., Boss, O., Peroni, O. D., Kim, J. K., Martino, J. L., Zabolotny, J. M., Moghal, N., Lubkin, M., Kim, Y. B., Sharpe, A. H., Stricker-Krongrad, A., Shulman, G. I., Neel, B. G. & Kahn, B. B. (2000). *Mol. Cell. Biol.* **20**, 5479–5489.
Krueger, N. X., Streuli, M. & Saito, H. (1990). *EMBO J.* **9**, 3241–3252.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nallamsetty, S., Austin, B. P., Penrose, K. J. & Waugh, D. S. (2005). *Protein Sci.* **14**, 2964–2971.
Nallamsetty, S., Kapust, R. B., Tozser, J., Cherry, S., Tropea, J. E., Copeland, T. D. & Waugh, D. S. (2004). *Protein Expr. Purif.* **38**, 108–115.
Nallamsetty, S. & Waugh, D. S. (2006). *Protein Expr. Purif.* **45**, 175–182.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Ostman, A. & Bohmer, F. D. (2001). *Trends Cell Biol.* **11**, 258–266.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannifer, A. D., Flint, A. J., Tonks, N. K. & Barford, D. (1998). *J. Biol. Chem.* **273**, 10454–10462.
Rye, C. S. & Baell, J. B. (2005). *Curr. Med. Chem.* **12**, 3127–3141.
Salmeen, A., Andersen, J. N., Myers, M. P., Tonks, N. K. & Barford, D. (2000). *Mol. Cell*, **6**, 1401–1412.
Stoker, A. W. (2005). *J. Endocrinol.* **185**, 19–33.
Tautz, L., Pellecchia, M. & Mustelin, T. (2006). *Expert Opin. Ther. Targets*, **10**, 157–177.
Tozser, J., Tropea, J. E., Cherry, S., Bagossi, P., Copeland, T. D., Wlodawer, A. & Waugh, D. S. (2005). *FEBS J.* **272**, 514–523.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Ventura, J. J. & Nebreda, A. R. (2006). *Clin. Transl. Oncol.* **8**, 153–160.
Wang, Y. & Pallen, C. J. (1992). *J. Biol. Chem.* **267**, 16696–16702.
Waugh, D. S. (2005). *Trends Biotechnol.* **23**, 316–320.
Wright, M. B., Seifert, R. A. & Bowen-Pope, D. F. (2000). *Arterioscler. Thromb. Vasc. Biol.* **20**, 1189–1198.
Zhang, Z. Y. (2002). *Annu. Rev. Pharmacol. Toxicol.* **42**, 209–234.
Zhu, D. W., Campbell, R., Labrie, F. & Lin, S. X. (1999). *J Steroid Biochem. Mol. Biol.* **70**, 229–235.
Zinker, B. A. *et al.* (2002). *Proc. Natl Acad. Sci. USA*, **99**, 11357–11362.