Molecular architecture of the tyrosine phosphatase HD-PTP and the structural basis for its specific interaction with UBAP1

Deepankar Gahloth[1,*], Colin Levy[1,*], Graham Heaven[2], Flavia Stefani[1], Lydia Wunderly[1], Paul Mold[1], Matthew J. Cliff[1], Thomas Jowitt[1], Clair Baldock[1], Jordi Bella[1], Alistair J. Fielding[2], Philip Woodman[1,$] and Lydia Tabernero[1,$].

[1]*Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK.*
[2]*School of Chemistry and Photon Science Institute, University of Manchester,*

* - equal contribution by these authors

$ - corresponding authors

MS total < 4500 ws (currently 4707)

Abstract

HD-PTP (PTPN23) is a multidomain phosphatase that regulates endosomal trafficking of mitogenic receptors, cell migration and neuronal homeostasis. Essential to HD-PTP function is the ability to coordinate binding of its different domains (Bro1, coiled-coil (CC) and proline-rich regions (PRR)) to the critical endosomal sorting multiprotein complexes ESCRTs. Here, we show by X-ray crystallography, SAXS and biophysical analysis that the CC domain exists in a open extended conformation, in contrast to the closed V shape of related Bro1 proteins. The extended conformation is preserved in solution in the presence of the Bro1 domain and the PRR region, revealing a unique architecture that allows

for simultaneous interaction with the different ESCRTs required to coordinate transfer of ubiquitinated cargo. In addition, the crystal structure of HD-PTP$_{CC}$ in complex with UBAP1, a subunit of ESCRT-I, shows that the open conformation, together with other molecular determinants, is critical to provide a structural basis for selective binding to HD-PTP.

Introduction

Human HD-PTP (PTPN23), is a pleiotropic tumour suppressor phosphatase that sits at the cross roads of three essential processes, endocytosis, down-reguation of mitogenic recpetors and cell migration. The *ptpn23* gene is located in a tumour supressor gene cluster of chromosome 3, 3p21.3, (Toyooka 2000 BBRC), deleted in various human cancers (Kok 1997 Adv Cancer Res) including nasopharyngeal carcinoma (Cheng, PNAS 1998).

Consistent with this, we and others discovered that HD-PTP drives down-regulation of activated mitogenic receptors EGFR, and PDGFRβ (Doyotte 2008, Miura 2008,Ma 2015). This role involves the specific recruitment of endosomal sorting complexes required for transport (ESCRTs) which are crucial for sorting ubiquitinated cargo to the multivesicular body (MVB) and for membrane abcission processes (rev. ref Phil to suggest). HD-PTP is also requried for degradation of ubiquitinated α5β1 integrin (Kharitidi 2015). Furthermore, depletion of HD-PTP causes accumulation of ubiquitinated proteins in aberrant endosomes (Doyotte 2008) and enhances cell proliferation, cell migration and invasion (Mariotti 2009, Lin 2011, Chen 2012, Kharitidi 2015, Ma 2015).

More recenlty, the importance of HD-PTP in neuronal homeostasis and brain disorders has been reported. Myopic/HD-PTP in Drosophila is essential for neurite scission during pruning, (Loncle Sci Rep 2015) and regulation of the survival motor neuron complex (Husedzinovic Mol Cell 2015). HD-PTP is also involved in genetic brain disorders (Alazami et al Cell Reports 10, 148–161, January 13, 2015).

HD-PTP is a multidomain protein that belongs to a family of related Bro1 proteins (Kimura 2015 Euk Cell), that contain a Bro1 domain and a coiled-coil (CC) domain. Human Alix and HD-PTP, also contain a proline-rich region (PRR), and HD-PTP has a C-terminal protein tyrosine phosphatase domain (Fig. 1), which is unique to this protein. Essential to HD-PTP function is its ability to interact simultaneosly with ESCRT-I and -III subunits. CHMP4B/ESCRT-III binds to a conserved site in the Bro1 domain (Ichioka 2007, Doyotte 2008), whereas interactions with ESCRT-I are mediated by the CC domain and the PRR region (Stefani 2011, Ichioka 2007) (Fig. 1). The Bro1 domain also binds to STAM2/ESCRT-0 (Ali 2013). Neither the Bro1 or the CC domain alone are sufficient to rescue RNAi depletion of HD-PTP to restore EGFR trafficking and MVB formation, highlighting that the Bro1-CC region is the minimal functional unit (Doyotte 2008). This suggests that concerted engagement of both domains in binding different ESCRT subunits is necessary for function.

In fact, HD-PTP appears to be an essential link between ESCRT-I and -III in the absence of ESCRT-II, since both down-regulation of ubiquitinated MHC class 1 proteins and neurite scission, require ESCRT-I and –III but not ESCRT-II,

suggesting that an alternative ESCRT pathway exist in which HD-PTP substitutes ESCRT-II (Parkinson 2015, Loncle Sci Rep 2015).

UBAP1, is a ubiquitin-associated protein whose expression is downregulated in nasopharyngeal carcinoma (Quian 2001) and in frontotemporal lobar degeneration (FTLD) (Rollinson 2009). Recent evidence associates UBAP1 with integrin degradation and, together with HD-PTP, appears to be essential to control cell migration and invasion following FN activation of α5β1 integrin (Kharitidi 2015). In agreement with this, we reported that UBAP1 binds to HD-PTP and defines an endosome specific ESCRT-I complex important for ubiquitin-dependent MVB sorting (Stefani 2011, Wunderley 2014). Furthermore, we have identified the CC domain in HD-PTP as the sole responsible for this interaction, and residue F678 in the FYX_2L motif, conserved in all Bro1 proteins, as key for binding (Stefani 2011). However, conservation of a Phe residue in this position is insufficient for binding of UBAP1 to the related human Bro1 protein Alix (Stefani et al. 2011), indicating that other structural determinants may be important to define selectivity.

The molecular details on how HD-PTP is able to coordinate multiple and simultaneous interactions with the ESCRT machinery and other signalling proteins are unknown. To date, structural information on HD-PTP is only available for the Bro1 domain (Sette 2011), which shows a similar fold and shape to the analogous domain in other Bro1 proteins (Fisher 2007, McCullough 2008, Zhai 2001 Plos One, Kim 2005). However, nothing is known about the structure of the other domains of HD-PTP neither on how they are coupled to coordinate

interactions with several partners during cargo trafficking. In particular, the CC domain shows very low homology to the analogous coiled-coil domains, known as V domains, of related Bro 1 proteins Alix and yeast Bro1 (19% and 17% respectively). This low homology, suggests that a significant structural disparity may be key to explain their functional diversity and binding selectivity to biological partners.

Here, we present the first crystal structure of the CC domain of HD-PTP (HD-PTP$_{CC}$) that shows an unexpected open and extended conformation, totally different to that of the closed V shape found in the Alix and yeast Bro1 V domains (Fisher 2007 Cell, Lee 2007, Pashkova 2013 Dev Cell). The SAXS structures of the HD-PTP$_{CC}$, HD-PTP$_{Bro1\text{-}CC}$ and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ proteins reveal that the open architecture is mantained in the presence of the Bro1 domain and the N-terminal region of the PRR, consistent with the role of HD-PTP as a major scaffold for protein-protein interactions. Furthermore, hydrodynamic and biophysical analyses confirm that the open conformation is the only form present in solution showing no evidence of large conformational flexibility.

Finally, the crystal structure of HD-PTP$_{CC}$ in complex with the the central region of UBAP1 (UBAP1$_{C}$), shows that the open conformation is also mantained in the complex and necessary for UBAP1 binding, providing a structural basis for selectivity.

Results

The crystal structure of the coiled-coil domain of HD-PTP reveals an open and extended conformation

The crystallographic structure of HD-PTP$_{CC}$ alone was determined at 2.5 Å resolution (Fig. 1a, Table 1), and shows a unique elongated α-helical architecture, where the dominating feature is a central helix of 104 residues extending the whole length of the molecule (Figure 1b). Two 3-helix bundles are found at both ends of the molecule around this central helix. The overall shape of HD-PTP$_{CC}$ resembles that of an ice hockey stick, in which the N-terminal helical bundle represents the blade (~ 73Å long) and the C-terminal helical bundle forms the shaft (~ 105Å long). The maximal dimension of HD-PTP$_{CC}$ is approximately 155Å from end to end with an average thickness of 20Å. The structure contains seven alpha helices (H1-H7) where H7 is the shared helix forming an N-terminal 3-helix bundle with H1 and H6 and a C-terminal 3-helix bundle with H3 and H4.

Analysis of the coiled-coil motifs in the structure, using the programme SOCKET (Walshaw, J. & Woolfson, D.N. (2001) *J. Mol. Biol.*, **307** (5), 1427-1450), identified two coiled coils regions: one antiparallel 2-stranded coil between H6 and H7 in the blade and a extensive antiparallel 3-stranded coil with tight packing between helices H3, H4 and H7 in the shaft (Fig. 1b).

The topology of HD-PTP$_{CC}$ is such that the polypeptide chain crosses three times over the length of the protein (Fig. 1c). At the base of the shaft, H1 is connected to H2 by an 18-residue long loop (L1) that appears to be highly flexible, as this region is poorly defined in electron density maps. In the same region, a 9-residue long loop (L4) connects H5 to H6. However, L4 sequence is rich in prolines and it is well defined in the density maps.

The extended shape of HD-PTP$_{CC}$ is remarkably different from the analogous V domains in the human Alix and the yeast Bro1 (Fig. 1d, e) (Lee 2007, Fisher

2007, Pahskova 2013). V-domains are built from two distinct arms separated by three unstructured loops that form a flexible hinge (Fig. 1d). In HD-PTP$_{CC}$ the H7 helix forms a continuous backbone along the coiled-coil domain, conferring rigidity to the molecule and maintaining the extended shape observed both in the crystal structure and in solution.

The extended architecture of HD-PTP$_{CC}$ is maintained in solution and in the presence of the Bro1 domain

Small angle X-ray scattering (SAXS) confirmed that the extended conformation of HD-PTP$_{CC}$ is also present in solution. The maximum dimension (D_{max}), obtained from GNOM (Svergun 92), was 152.5 Å, matching almost exactly with the dimension of the crystal structure. The particle shape restored *ab initio* with Gasbor (ref) was consistent with an open conformation for HD-PTP$_{CC}$ (Fig. 2A and Suppl. Fig. 2). Indeed, the crystal structure displayed a good fit to the SAXS profile (χ^2 = 1.6).

We, then, analysed the architecture of two different longer constructs of HD-PTP, one containing the Bro1 domain and the coiled-coil domain (HD-PTP$_{Bro1\text{-}CC}$) and one containing the N-terminal residues of the proline-rich region (PRR) (HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$). This region contains the binding site for STAM2-SH3 (Ali 2013), MONA-SH3 (Karkiolaki 2009) and the "PTAP" motif that binds to TSG101 (Ichioka 2007) and, therefore, is functionally relevant.

The SAXS *ab initio* envelopes generated for HD-PTP$_{Bro1\text{-}CC}$ and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$, showed an extended architecture, in which the Bro1 and the CC domain are in a linear arrngement, to reach a maximum dimension of approximately 193 Å and 204 Å respectively (Fig. 2B, C and Suppl. Fig. 2). There is a visible extension in

the HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ envelope that corresponds to the position of the additional PRR residues (Fig. 2 C) at the C-terminal end of the CC domain.

Molecular models for HD-PTP$_{Bro1\text{-}CC}$ were built from the structures of the HD-PTP$_{CC}$ and HD-PTP$_{Bro1}$ (PDB ID: 3RAU) using torsional angle molecular dynamics (TAMD) with CNS (Brunger 2007). Models were then selected on the basis of their calculated hydrodynamic parameters and compared to the experimental SAXS profile using FoXS (D. Schneidman 2010), as previously described (Tariq 2015). The best model shows a very good agreement to the experimental SAXS data, with χ^2 = 1.15 for HD-PTP$_{Bro1\text{-}CC}$ and χ^2 = 1.36 for HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ (Fig. 2B,C and Suppl. Fig. 2).

The presence of the first residues of the PRR region does not affect the global architecture of the HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ protein, nor the orientation between Bro1 and CC domains. This extended platform is in striking contrast to the crystal structure of Alix$_{Bro1\text{-}V}$, in which both domains are folded in a compact closed conformation (Fisher 2007). In Alix, the C-terminus of the V-domain is in close proximity to the N-terminus of the Bro1 domain, and the presence of the PRR has been proposed to play a regulatory role by locking the closed conformation between these two domains and controlling the activation of Alix (Pires 2009, Zhou 2010, Zhai 2011).

The HD-PTP$_{CC}$, HD-PTP$_{Bro1\text{-}CC}$, and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ proteins were further characterised by biomolecular analyses in solution, including size-exclusion multiangle laser light scattering (SEC-MALLS) and analytical ultracentrifugation (AUC), to provide hydrodynamic and dimensional parameters (Suppl. Table 1).

These data were used to generate estimations of the shape of HD-PTP and to generate suitable models for the structure of the HD-PTP$_{Bro1\text{-}CC}$ proteins as described above.

All three proteins behaved as single species monomers in solution with average molecular weights of 38.2 KDa for the HD-PTP$_{CC}$, and 86 and 88.7 KDa for HD-PTP$_{Bro1\text{-}CC}$ and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ respectively (Suppl Table 1 and Suppl. Figure 1). Hence, in contrast to Alix (Pires 2009), HD-PTP appears to lack an intrinsic ability to dimerise in solution. Furthermore, sedimentation velocity experiments showed that these proteins each migrate as a single particle with sedimentation coeficients of 2.51S for HD-PTP$_{CC}$ and 3.98S for HD-PTP$_{Bro1\text{-}CC}$ (Suppl Table 1 and Suppl. Figure 1). The hydrodynamic radius and frictional ratio (R_h = 3.93, f/f_0 = 1.68) indicate that HD-PTP$_{CC}$ adopts an anisotropic and elongated shape, clearly different from the structure of Alix$_V$ (R_h = 3.17, f/f_0 = 1.41, parameters calculated using SOMO, Rai, 2005, from PDB ID 2OJQ). Likewise, the hydrodynamic studies show that HD-PTP$_{Bro1\text{-}CC}$ adopts an extended conformation with a larger hydrodynamic radius and frictional ratio (R_h = 4.97 and 5.46 and f/f_0 = 1.8), in comparison to the more compact conformation of the Alix$_{Bro1\text{-}V}$ (R_h = 4.28 and f/f_0 = 1.51, form PDB ID 2OEV) (Table S1). Altogether, these data confirms the observations by X-ray crystallography and SAXS that the open, extended architecture of the HD-PTP is completely different to previously characterised Bro1 proteins and that it exists in solution as the sole conformational species.

We believe that this extended open conformation, even in the presence of the PRR extension, is critical for simultaneous binding of Bro1, CC and the PRR to different ESCRT subunits (STAM2, TSG101, UBAP1, CHMP4). Moreover, this conformation provides a structural basis for the coordination of HD-PTP

interactions with ESCRT-I and ESCRT-III during cargo trafficking to MVBs (Doyotte, Stefani, Ali).

HD-PTP$_{CC}$ *shows local conformational dynamics in solution with no large structural rearrengements*

Conformational flexibility between the two arms in the yeast Bro1p and human Alix V domains has been reported, suggesting that in solution these exist both as a closed and open conformations (Pashkova 2013, Pires 2009,), with the loops connecting both arms acting as a hinge region. This flexibility might be important in regulating activation of Alix and ligand binding (Carlton 2008, Pires, 2009, Zhou 2010, Zhai 2011, Sun 2015). Our biomolecular hydrodynamic and SAXS analyses do not show evidence of the presence of a closed conformation in solution for HD-PTP$_{CC}$ or HD-PTP$_{Bro1\text{-}CC}$.

To further investigate the conformational dynamics of HD-PTP$_{CC}$, we used double electron-electron resonance (DEER) spectroscopy. We measured dipolar coupling interactions between unpaired electrons on nitroxide spin-labels (methanethiosulfonate, MTSL) attached to cysteine residues. We conducted DEER experiments on HD-PTP$_{CC}$, which contains three cysteines (C_{425}, C_{628} and C_{697}), and on mutants where one of the cysteines had been substituted for serine. From these measurements, mean interspin distances were calculated.

Significantly, the $C_{425}S$ mutant gave no resolved dipolar coupling indicating that the spin labels on $C_{628,}$ at the N-terminal end of the blade region, and C_{697} at the C-terminal end of the shaft, must be separated by more than 7 nm (maximal mean distance observable in a 6 µs window) (Figure 3). The predicted interspin

distance from the crystal structure is 9.8 nm, thus the lack of dipolar coupling is consistent with the elongated conformation obtained for HD-PTP$_{CC}$.

The distance measured between the C_{425} and C_{697} labels of 6.3 nm, matched very well the predicted value of 6.0 nm based on the crystal structure. However, the experimental distance between the C_{425} and C_{628} labels of 4.2 nm, differed from the predicted value of 5.4 nm, indicating local conformational dynamics. C_{425} is located in the H2 helix, connected by two flanking loops L1 (402-418) and L2 (431-433) to H1 and H3 thus enabling certain flexibility in this region. In fact, the crystallographic data suggests that the L1 loop is indeed highly flexible and that different conformations exist in the crystal. Molecular dynamics simulations (MDS) allowing free movement of the L1 and L2 loops, show that H2 can easily adopt different orientations that are compatible with the shorter interspin distance between C_{425}-C_{628}, while mantaining the same distance for C_{425}-C_{697} (Suppl figure 3).

To confirm the above analyses, an additional Cys residue was engineered at position L521 In H4. DEER experiments provided a mean distance between the C_{628} and C_{521} spins of 4.9 nm (Suppl. Fig. 3), in good agreement with the predicted distance of 5.2nm, and thus confirming that the overall conformation of HD-PTP$_{CC}$ in solution is consistent with the SAXS data and the crystal structure. These data together with the biomolecular and SAXS analysis in solution provide robust proof that the architecture of the HD-PTP provides an inflexible, extended structure, which displays certain local structural dynamics but not global conformational rearrengements.

Identification of the binding region of the ESCRT-I component UBAP1 to HD-PTP

We have previously identified UBAP1 as a binder of HD-PTP using a yeast two-hybrid (Y2H) screen. Binding occurs between the central region of UBAP1(122-309) and the CC domain of HD-PTP (Stefani 2011). We have now confirmed the interaction between full-length UBAP1 and HD-PTP by co-immunoprecipitation of *in vitro* translated UBAP1 with bacterially expressed HD-PTP$_{CC}$ (Suppl fig 6). Binding between UBAP1 and HD-PTP$_{CC}$ was also detected in cells, by co-expressing a mitochondrially-targeted FRB fragment with FKBP-fused to HD-PTP$_{CC}$. FKBP-HD-PTP$_{CC}$ was localised to mitochondria in the presence of rapamycin, and under these conditions it recruited GFP-UBAP1, but not GFP (Fig. 4a). The minimal UBAP1 binding interface with HD-PTP$_{CC}$ was then mapped using Y2H by making nested deletions from both N- and C-termini of UBAP1(122-309) (Figure 4b). This approach identified residues 260-269 in UBAP1 as the minimal region for efective binding. This region contains a FPXL motif that resembles the YPX$_n$L motif within retroviral Gag late-domains and other target proteins that bind to the V domain of Bro1 proteins (Lee 2007, Zhai 2008, Zhai 2011, Kimura 2015). The mutation of the UBAP1 F268S in this motif abolished binding to HD-PTP$_{CC}$, based both on Y2H and mitochondrial targeting assays (Fig 4a, b), suggesting that UBAP1 F268 forms a critical interaction with HD-PTP. However, mutation of P269A or L271A had no obvious effect in the presence of the native F268.

To confirm binding of this region, immobilised HD-PTP$_{CC}$, HD-PTP$_{Bro1\text{-}CC}$, HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ and HD-PTP$_{Bro1}$ were tested in biosensor binding experiments (SPR) with a UBAP1 peptide containing the central region residues 261 to 280 (UBAP1c) as the analyte (Fig. 4 c). Affinity to HD-PTP$_{CC}$, HD-PTP$_{Bro1\text{-}CC}$, and HD-

$PTP_{Bro1\text{-}CC\text{-}PRR}$ was similar, with dissociation constants K_d of 66.3uM, 31.87uM and 57.46uM respectively (Fig. 4 c, d, e, f). However, no binding was observed to HD-PTP_{Bro1} (Fig. 4g), thus confirming that the main binding region of UBAP1 peptide is within the CC domain.

Crystal structure of the HD-PTP_{CC} *in complex with UBAP1c*

The X-ray structure of HD-PTP_{CC} complexed with UBAP1c was determined at 2.5 Å resolution (Table 1). UBAP1c binds to a hydrophobic groove in the shaft region, near the core of the 3-strand coiled-coil (Fig. 5 a). UBAP1 residues 262-271 are visible in the electron density maps and the side chains of I263, L266, F268, P269 and L271, all interact with HD-PTP (Fig. 5 b). No apparent secondary structure is observed for UBAP1c, which buries 562.5 $Å^2$ of solvent-accessible surface area on HD-PTP_{CC}. Interactions between UBAP1c and HD-PTP_{CC} are mainly hydrophobic, with hydrogen bonds present only between HD-PTP K671 and UBAP1 S267 and K270 (Fig. 5c). The binding interface in HD-PTP_{CC} can be described by three hydrophobic pockets (A-C) (Fig. 5b). UBAP1 I263 and L266 bind to pocket A, which is large and shallow, and form hydrophobic interactions with HD-PTP T511, L515, A518 (H4 helix) and A664 and L668 (H7 central helix). UBAP1 F268 binds to pocket B, which is deep and narrow, where it stacks with HD-PTP F678 and forms hydrophobic interactions with HD-PTP A508 and T511 (H4 helix) and G675 and Y679 (H7 helix). UBAP1 P269 accomodates itself near the pocket B where it forms hydrophobic interactions with HD-PTP F678, A148 ans V144. Finally, UBAP1 L271 binds to pocket C at the C-terminal edge of the binding site and forms hydrophobic interactions with HD-PTP V504 (H4 helix),

F678 and L682 (H7 helix). The aryl ring of HD-PTP F678 forms a wall that divides pocket B and C and contributes to hydrophobic interactions in both.

Interestingly, the A-pocket is partially occluded in the structure of HD-PTP$_{CC}$ alone, whereas the B and C pockets are available (Fig. 5 d). Binding of UBAP1c appears to require the re-arrangement of at least three side chains (in K671, D667 and E514) to accommodate the peptide. In the Apo structure, HD-PTP D667 and K671 form a salt bridge, and the carboxyl group of HD-PTP E514 is about 4-5Å closer to D667 and K671 than in the structure of the complex (Fig. 5 e). In the UBAP1 complex, E514 and D667 move away breaking the salt bridge. Instead, NZ of K671 forms hydrogen bonds with the carbonyl oxygens of I263, K264 and L266 in UBAP1 (Fig. 5 c, e). Thus UBAP1c binding results in a widening of the space between H4 and H7 upstream of the 3-stranded coiled-coil.

The extent of the UBAP1 interface with HD-PTP$_{CC}$ was further evaluated using NMR by comparing the ^{1}H ^{13}C-HSQC spectrum of UBAP1c in the absence and presence of HD-PTP$_{CC}$. Significant peak broadening was observed for peaks corresponding to UBAP1 residues 261 to 271 in the presence of HD-PTP$_{CC}$ (Fig. 6 a, b and Suppl. Fig. 4), consistent with this residues binding to HD-PTP as observed in the crystal structure.

To confirm the orientation of the UBAP1c in the HD-PTP$_{CC}$ binding site, we performed paramagnetic relaxation enhancement (PRE) experiments with the MTSL-label at C_{425} and C_{628}. Dipolar interaction with the electron spin-label causes an additional component of the transverse relaxation of the NMR signal, which has strong distance dependence with maximal effect below 40 Å. The

paramagnetic enhancement in this case is dominated by the spin-label attached to C_{425} at the base of the shaft, which is closer to the UBAP1 binding site. Consistent with this, NMR resonances from the N-terminal end of the UBAP1c showed enhancement of their relaxation properties and therefore loss of peak intensity in the presence of the MTSL label (oxidized form), whereas those from the region beyond L271 (> 40 Å from the label), did not (Fig. 6 c). Together these NMR experiments confirmed both the UBAP1c interface with HD-PTP and its orientation at the binding site.

Structural basis for specific interaction of HD-PTP with UBAP1, and comparison with Alix$_V$ complexes with retroviral late-domain peptides.

Several structures of the Alix$_V$ in complex with retroviral late-domain peptides have been reported (Fisher 2007, Lee 2007). These show a common hydrophobic binding site, within the second arm of the V domain that contains the FYX$_2$L motif present in all Bro1 proteins (Kimura 2015). The conserved Phe residue in this motif is essential for binding to the YPX$_n$L motif in viral and cellular target proteins, as demonstrated for Alix (Fisher 2007, Lee 2007), Bro1p and Rim20 (Kimura 2015). In the structures of the Alix complexes, the FY pair stack with the Y present in the late domain motif, thereby serving as the main anchoring point for interaction. Analogously, in HD-PTP the FY pair in the FYX$_2$L motif (F678, Y679 and L682), froms similar stacking interactions to the FP pair in the UBAP1 FPXL motif (F268, P269 and L271) (Fig. 5 c).

The critical importance of HD-PTP F678 and UBAP1 F268 for binding was confirmed by SPR binding studies. HD-PTP$_{CC}$ containing the F678D mutation

displayed no binding to UBAP1c, likewise a UBAP1c peptide with the F268S substitution failed to interact with wt HD-PTP$_{CC}$ (Fig. 7).

Important differences are apparent when comparing the mode of binding and the architecture of the binding sites between the HD-PTP-UBAP1 and Alix-late-domain peptide complexes. First, UBAP1c and late-domain peptides bind in anti-parallel fashion. The N-terminus of the UBAP1c sits in the A pocket extending towards the central region of HD-PTP$_{CC}$ and the C-terminus binds to the B and C pockets (Fig. 8). In contrast, late-domain peptides bind to Alix in the opposite direction, with their N-termini at the pocket containing the conserved Phe residue and extending towards the C-terminal end of the V domain, along a hydrophobic groove (Fig. 8). Second, the hydrophobic groove in Alix is not found in HD-PTP because the core of the 3-stranded coiled-coil in this region results in a very tight packing of the helices, leaving little room for a binding groove. Third, the HD-PTP pocket A is poorly conserved in Alix and does not participate in any binding interactions with the late-domain peptides (Fig. 8 and suppl. Fig. 5). Finally, the architecture of HD-PTP offers an open, extended interface suitable to accommodate UBAP1$_C$. In contrast, in Alix the apex formed between arm1 and arm2 in the V domain precludes binding of any peptides in that region (Fig. 8) because of steric hindrance, thus the binding site is displaced towards the opposite end of arm2 (C-terminal end).

Combined, all the local structural variations together the dramatic architectural differences between HD-PTP and Alix provide a molecular basis to rationalise

the ligand specificity and selectivity reported (Stefani, Ali?, refs CEP55, Gag proteins,…etc).

To support this notion, we analysed the binding of UBAP1c to Alix$_V$ and that of a late-domain peptide (SIV-GAG) to HD-PTP$_{CC}$. As expected, Alix$_V$ did not bind to the UBAP1c, in agreement with our previous report (Stefani et al. 2011). Conversely, the SIV-GAG peptide did not bind to HD-PTP$_{CC}$ but showed affinity to Alix$_V$ (Fig. 8). This indicates that although the presence of the FYX$_2$L motif is necessary for binding but insufficient to determine binding selectivity to biological partners.

Validation of the HD-PTP-UBAP1 interaction

To further explore the HD-PTP-UBAP1 interface identified in the X-ray structure of the complex, we used RNAi rescue experiments using mutations of the binding site residues within HD-PTP. In normal cells, EGFR that has been activated by EGF passes through the endosomal pathway and is degraded within lysosomes. In contrast, cells depleted of HD-PTP are characterised by the accumulation of ligand-activated EGFR in highly clustered early endosomes that label strongly for protein-ubiquitin conjugates (Doyotte et al. 2008). Reintroduction of HD-PTP$_{Bro1\text{-}CC}$ is sufficient to rescue these trafficking defects (Doyotte). Hence, HeLa cells depleted of HD-PTP and pulsed for 3hr with EGF showed intense ubiquitin labelling on cytoplasmic clusters (Fig. 7) which co-labelled with the endosomal marker EEA1 (data not shown). As expected, transfection of these cells with HD-PTP$_{Bro1\text{-}CC}$ restored a wild-type phenotype, in which ubiquitinated proteins were distributed mainly in the cytoplasm and nucleus (Fig. 7). In contrast, HD-PTP$_{Bro1\text{-}}$

$_{CC\text{-}F678D}$ was unable to rescue depletion of HD-PTP, and showed severe endosomal accumulation of ubiquitinated proteins (Fig. 7). The mutation of K671A, failed to completely rescue a wt phenotype, although it showed milder defects than the F678D mutation. K671 forms hydrogen bond interactions with UBAP1c, but clearly the contribution to the binding affinity is minor compared to the more extensive hydrophobic interactions in the binding site, and in particular in the presence of F678.

Discussion

The extended structures of HD-PTP$_{CC}$ and HD-PTP$_{Bro1\text{-}CC}$, present an extensive accessible surface consistent with the proposed role for HD-PTP as a scaffold platform for protein-protein interactions, with the endosomal sorting machinery and other signaling proteins (refs). The long distance between the CHMAP4B binding site in the Bro1 domain and the distal PRR (~200 Å) suggests that this protein may be able to interact simultaneously with subunits in different ESCRTs to coordinate the trafficking of the ubiquitinated cargo. For example, this extended platform would allow for binding of HD-PTP$_{Bro1}$ to STAM2/ESCRT-0 and of HD-PTP$_{CC}$ to UBAP1/ESCRT-I, while CHMP4/ESCRT-III would be competing with or displacing ESCRT-0 for binding to HD-PTP$_{Bro1}$. Particularly relevant to this is the reported essential requirement of HD-PTP, ESCRT-I and ESCRT-III, but not Alix, to promote cargo trafficking in the absence of ESCRT-II. This invokes a model in which HD-PTP may serve as a link between ESCRT-I and

ESCRT-III playing the role of a “clip” or “bracket” to recruit and cluster different ESCRTs together, thus facilitating transfer of ubiquitinated cargo.

The structure of HD-PTP$_{CC}$ in complex with UBAP1$_{C}$ defines three main pockets at the binding interface and identifies the important residues in HD-PTP responsible for interaction. The conserved F678 in pocket B is key for hydrophobic interactions with UBAP1 as mutation to Asp impairs binding to UBAP1. The cognate residue F676 in Alix is responsible for binding to the late-domain of Gag proteins and mutation to Asp abolishes affinity to Gag peptides (Fisher 2007, Lee 2007). Likewise the mutation of the corresponding F687 in Bro1p or F623 in Rim20 to Ala, impairs binding to Rfu1 and Rim101 respectively (Kimura 2015). This confirms that functional relevance of the FYX$_2$L motif, although we have demonstrated that it is not sufficient to elicit binding selectivity.

Mutagenesis of the FPXL motif in UBAP1, show that the Phe residue is key to anchor the UBAP1 peptide into the hydrophobic B pocket in HD-PTP, whereas Pro and Leu play a minor role, in matching with similar results reported for the late-domain YPX$_n$L motif (Lee 2007, Kimura 2015). Furthermore, evidence of the importance of the HD-PTP – UBAP1 interface and the role of the conserved F residues in both proteins was supported by mitochondrial localisation and rescue studies in cells.

The mode of binding of UBAP1 to HD-PTP shows similarities with the binding of the late-domain Gag peptides to Alix, such as the stacking interactions between UBAP1 F268 and HD-PTP F678 in the conserved B pocket. However, significant differences in the orientation of the peptides, in the local structural features of

the binding site and importantly, in the architecture between HD-PTP$_{CC}$ and Alix$_{V}$, established a structural basis for the binding selectivity reported (Stefani). This selectivity was further demonstrated by the inability of a SIV-gag peptide to bind to HD-PTP and of UBAP1 to bind to Alix.

Methods

Methods and any associate references are in the supplementary material

Acknowledgements ALL to complete

This work was support by Medical Research Council (xxxxxx (L.T.) and XXX (add as pertinent)

We thank the staff of beamlines XXXXX at the Diamond Light Source (X-ray diffraction data, BSG mention), the staff of the Biomolecular Analysis Facility of the Faculty of Life Sciences, University of Manchester for assistance; the staff of the beamline XX at ESRF and Hamburg (SAXS diffraction data, proposal?); the staff at the Advanced Photon Science facilities, University of Manchester for assistance. AJF thanks Bruker for sponsorship. All EPR experiments were carried out at the EPSRC National EPR Research Facility & Service. We thank Drs. Melanie Vollmar, Simon Tanley and Efrain Ceh Pavia for advice.

Author Contributions

D.G. cloned HD-PTP constructs, prepared protein samples and crystallised the proteins. C.L. collected data, solved and refined the structures. G.H. prepared labeled proteins, collected data and G.H. and A.J.F. analysed DEER data. G.H. collected CD data and together with J.B. analysed it. M.J.C. collected and analysed

NMR data. P.M. and D.G. collected SPR data and P.M. analysed it. T.J. collected and analysed MALLS and AUC data. C.B. collected and processed SAXS data and analysed it with D.G. JB. Performed MD simulations for EPR and SAXS analyses and generated molecular models. F.S. performed the yeast two-hybrid and co-IP experiments. L.W. performed the mitochondrial localization and rescue experiments. A.J.F initiated and supervised the EPR spectroscopy. P.W. supervised and analysed all cell biology data. L.T designed, supervised the project and analysed the structural and biophysical data. L.T and P.W. wrote the paper.

Competing Financial Interests

The authors declare no competing financial interests.

References (limit 60)

Tables

Table 1. Crystallographic data collection and refinement statistics

	HD-PTP$_{CC}$	HD-PTP$_{CC}$ – UBAP1$_{C}$
Data collection		
Space group	P21	P21 21 21
Cell dimensions		
a, *b*, *c* (Å)	53.5, 47.7, 172.7	48.9, 93.3, 102.2
α, β, γ (°)	90.0, 96.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	2.54(2.63-2.54) *	2.55(2.64-2.55)
R_{merge}	0.11(0.78)	0.087(0.76)
I / σI	19.3(3.84)	13.55(2.51)
Completeness (%)	98(100)	100(100)
Redundancy	13.5(13.8)	6.3(6.6)
Refinement		
Resolution (Å)	2.54	2.55
No. reflections	28609	15835

R_{work} / R_{free}	21.67 / 27.84	20.86 / 25.39
No. atoms	5156	2725
Protein	4991	2639
Ligand/ion	NA	73
Water	165	13
B-factors		
Protein	49.08	73.59
Ligand/ion	NA	89.73
Water	45.55	56.45
R.m.s. deviations		
Bond lengths (Å)	0.009	0.003
Bond angles (°)	1.04	0.51

Each structure was determined from one crystal. *Values in parentheses are for highest-resolution shell.

[AU: Ramachandran statistics should be in Methods section at the end of Refinement subsection.]

Figure captions

Figure 1. The crystalographic structure of HD-PTP$_{CC}$ shows an extended and open conformation. (a) Diagram of HD-PTP domain structure indicating domains boundaries and interacting partners. (b) The structure of HD-PTP$_{CC}$ resembles the shape of an ice hockey stick, where the N-termina region represents the "blade" and the C-terminal region is the "shaft". The seven α-helices are labelled H1 to H7, wth H7 being the longest helix extending the whole leb=ngth of the structure. (c) The structure of HD-PTP$_{CC}$ contains two coiled-coil regions, one two-stranded coiled-coil in the "blade" and one three-stranded coiled coil the "shaft" region. (d) Superimposion of the HD-PTP$_{CC}$ structure (red) with Alix$_V$ (PDB ID: 2OJQ). (e) Structures of Alix$_V$ (PDBID: 2OJQ) and Yeast Bro1$_V$ (PDB ID: XXX). (f) Topology diagram of the structure of HD-PTP$_{CC}$.

Figure 2. SAXS analysis of HD-PTP$_{CC}$, HD-PTP$_{Bro1\text{-}CC}$ and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ confirms the extended architecture of HD-PTP. The open conformation of HD-PTP$_{CC}$ (a) is also present in solution and preserved in the presence of the Bro1 domain (b) and of the N-terminal region of the PRR (c). The experimental SAXS curve is shown in grey open circles, and the fitting of the molecular model in red (produce with FoXS). Damaver envelopes are shown with the molecular model superiposed (produced with Quimera). The agreement between the models and the SAXS profile is shown (χ^2).

Figure 3. DEER spectroscopy of HD-PTP$_{CC}$ confirms the extended structure and rigidity in solution. (a) DEER traces. Inset shows form factors fits after exponential background subtraction (inset). The labelled Cys residues are

indicated. (**b)** DEER derived distance distributions (DEER, black line) using Tikhonov regularisation and crystal structure based predictions (Pred, red line), between labelled Cys residues as indicated. (**c)** Structure of MTSL labelled cysteine. (**d)** Structure of HD-PTP$_{CC}$ with MTSL labels at C_{425}, C_{628} and C_{697} shown as sticks. Mean experimental DEER distances(black) and predicted (red) distances are annotated. (**e)** Structure of HD-PTP$_{CC}$ with MTSL labels at C_{521} and C_{628} shown as sticks. Mean experimental DEER distances (black) and predicted (red) distances are annotated. (**f)** Crystal structure of HD-PTP$_{CC}$ (red) and model obtained with molecular dynamics (green) showing displacement of H2 and movement of flanking loops. Predicted distances between labels at C_{425} and C_{628} show agreement with experimental DEER distances (black).

Figure 4. Identification of the minimal binding region in UBAP1 responsible for interaction with HD-PTP . Phil to fill a, b. labels in second column have Bro1V, should be HD-PTP$_{Bro1\text{-}CC}$. (**c**) Biosensor binding isotherms for binding of the different HD-PTP constructs to UBAP1$_C$. Affinity to HD-PTP$_{CC}$, HD-PTP$_{Bro1\text{-}CC}$ and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ was similar, with dissociation constants K_d, of 66.3±0.35 μM, 31.87±0.85 μM and 57.46±3.10 μM respectively (**d**) Bioensor sensograms for the immobilised HD-PTP constructs binding to UBAP1$_C$, showing that the CC domain is the main responsible for binding since the Bro1 domain alone does not show any binding to UBAP1$_C$.

Figure 5. Binding of UBAP1$_C$ to HD-PTP$_{CC}$ results in an opening on the helix bundle in the shaft. LYT to fill. (**a**) Structue of HD-PTP$_{CC}$ in complex with

UBAP1$_C$ (CPK) . (**b**) Electrostatic surface of HD-PTP$_{CC}$ at the UBAP1$_C$ binding site, the three main pockets A-C are marked as yellow circles. Electron density map for UBAP1$_C$ (sticks) is shown at 1σ level. Redidues in UBAP1$_C$ that interact with HD-PTP$_{CC}$ are labelled: I263, L266 in pocket A; F268 in pocket B and P269 in a shallow pocket next to B; L271 in pocket C. F268, P269 and L271 are in the FPXL motif of UBAP1. (**c**) UBAP1$_C$ binding site showing the HD-PTP$_{CC}$ residues in the conserved FYX$_n$L motif: F678, Y679 and L682 (orange). Residue K671 forms hydrogen bond interactions with three carbonyl oxygens in the UBAP1$_C$ peptide (sticks). (**d**) Electrostatic surface of the structure of HD-PTP$_{CC}$ alone showing that the pocket A is occluded by the side chain of the K671 (yellow arrow). (**e**) Detail of the structure of HD-PTP$_{CC}$ alone (white ribbon left) and in complex with UBAP1$_C$ (green ribbon centre) showing the shift in the positions of the side chains of K671 and D667 that form a salt bridge that is lost in the structure of the complex. Binding of UBAP1$_C$ results in a displacemrnt of the side chains of D667 and E514 (right panel) and results in K671 forming hydrogen bond interactions with the peptide (as in c).

Figure 6. NMR analysis confirms the orientation and binding interface of UBAP1$_C$ to HD-PTP$_{CC}$. (a). Hα region of $^{1}H^{13}C$-PUSH-HSQC of UBAP1$_C$ (natural abundance ^{13}C) showing resonance assignment dissolved in PBS-D_2O. (**b).** Same Hα region of $^{1}H^{13}C$- HSQC of UBAP1$_C$ (natural abundance ^{13}C) in presence of sub-stoichiometric HD-PTP$_{CC}$. D_2O content in these samples was approximately 80% and significant intensity arises from H_2O, compromising interpretation of signals between 4.8 and 4.6 ppm. Residues close in sequence to UBAP1 F268 are

significantly broadened by interaction with HD-PTP$_{CC}$. Residues C-terminal of UBAP1 D272 are less affected by the presence of HD-PTP$_{CC}$ and identified as not being involved in the binding site. (**c).** Per-residue mean peak intensity ratios between UBAP samples containing paramagnetically (I_{ox}) and diamagnetically (I_{red}) labelled HD-PTP$_{CC}$ C697S indicating the extent of paramagnetic relaxation enhancement. For some resonances, peak intensities are low in both samples because of the line-broadening induced by binding, and these residues are marked by red squares in the chart.

Figure 7. Biochemical and functional analysis of the UBAP1-HD-PTP$_{CC}$ interface confirms the role of conserved residues in both proteins. a) Cells depleted of HD-PTP were transiently transfected with HA-tagged HD-PTP and stimulated with EGF for 3 hr before fixing and staining with anti-ubiquitin. Cells transfected with WT HD-PTP display generally even ubiquitin distribution throughout the nucleus and cytoplasm, whilst untransfected cell or those transfected with the F678D mutant display very strong accumulations on cytoplasmic inclusions. These co-label with endosomal markers (data not shown and previous publications)label in a should be HA- HD-PTP$_{Bro1\text{-}cc}$. (b). Scoring of rescue experiments. Cells transfected as indicated were scored for normal ubiquitin distribution. 100 cells from 3 independent experiments were counted, and SDs between these experiments are shown. **(**c) biosensor sensograms of binding of HD-PTP$_{CC}$ containing the F678D mutation to UBAP1 peptide (left) and of binding of HD-PTP$_{CC}$ and HD-PTP$_{Bro1\text{-}CC}$ to UBAP1 peptide containing the F268S mutation (centre and right). Both mutations abolishe interaction.

Figure 8. Comparison of the binding site for UBAP1 to HD-PTP$_{CC}$ with that of Gag peptides to Alix explains the structural basis for selectivity. a. structures of Alix in complex with HIV-Gag (yellow) and SIV-gag (purple) peptides superimposed to the structure of HD-PTP$_{CC}$ with UBAP1c (green). **b.** zoom into the binding site of the Gag peptides in Alix with UBAP1 superimposed. Binding of the Y in the YPXnL motif is similar to the F in the FPXL motif in UBAP1. The Y/F residues fit into the narrow hydrophobic pocket in Alix/HD-PTP containing the conserved F in the FYX$_2$L motif. (**c**) Binding to the FYX$_2$L motif is selective depsite the conservation, UBAP1 does not bind to Alix and the SIV-Gag peptide does not bind to HD-PTP.

Supplementary Material

Supplementary Table 1. Hydrodynamic and dimensional experimental data for HD-PTP$_{CC}$, HD-PTP$_{Bro1\text{-}CC}$ and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ determined from MALLS, AUC and SAXS. Hydrodynamic paramaeters for the crystal structure of HD-PTP$_{CC}$, the SAXS model of HD-PTP$_{Bro1\text{-}CC}$ and the crystals structures of Alix $_V$ and Alix$_{Bro1\text{-}V}$ were calculated with SOMO (Rai) and are shown for comparison. R_h, hydrodynamic radius; $S_{20,w}$, sedimentation coefficient in water at 20 °C; f/f_0, frictional coefficient; R_g, radius of gyration; D_{max}, maximal linear dimension of the particles.

Protein	MALLS		AUC			SAXS	
	R_h (nm)	M_r (Da)	R_h (nm)	f/f_o	$S_{20,w}$	R_g (nm)	D_{max} (nm)

HD-PTP$_{CC}$	3.93 ± 0.12	3.82 10^4	3.80	1.68	2.51 ± 0.07	4.53	15.25
HD-PTP$_{Bro1-CC}$	4.97 ± 0.12	8.59 10^4	5.14	1.80	3.98 ± 0.1	5.55	19.34
HD-PTP$_{Bro1-CC-PRR}$	5.46 ± 0.16	8.87 10^4	-	-	-	5.83	20.36
HD-PTP$_{CC}$ crystal structure	-	-	3.62	1.61	2.64	4.46	16.23
HD-PTP$_{Bro1-CC}$ model	-	-	4.68	1.64	4.03	5.58	20.97
Alix $_V$ (2OJQ)	-	-	3.17	1.41	2.87	2.96	11.17
Alix$_{Bro1-V}$ (2OEV)	-	-	4.28	1.51	4.21	4.51	16.5

Suppl. Figure 1. MALLS (a) and AUC (b) analysis of HD-PTP$_{CC}$ and HD-PTP$_{Bro1-CC}$ shows that both proteins behave as single species monomeric particles and provides hydrodynamic information on their shape in solution (see Suppl. Table 1 for values).

Suppl. Figure 2. P(r)/r plots (a), Kratky plots (b) and Porod-Debye plots (c) from the SAXS analysis of HD-PTP$_{CC}$, HD-PTP$_{Bro1-CC}$ and HD-PTP$_{Bro1-CC-PRR}$ provide dimensional parameters (D_{max}) for the different particles.

Suppl. Figure 3

Circular dichroism of WT HD-PTP$_{CC}$ (unlabelled and MTSL labelled at C_{425}, C_{628} and C_{697}) and mutants C_{425}S (MTSL labelled at C_{628} and C_{697}), C_{628}S (MTSL labelled at C_{425} and C_{697}) and C_{697}S (lablled at C_{425} and C_{628}). No significant changes are seen upon labelling or mutation of HD-PTP$_{CC}$.

Supplementary Figure 4.

a. Aromatic regions of ^{1}H NMR spectra of UBAP1$_{C}$, in the absence (upper panel) and presence (lower panel) of HD-PTP$_{CC}$. Narrow line-width signals arising from the sidechain of Phe268 (Hδ# (7.21ppm), Hε# (7.28ppm) and Hζ(7.24ppm)) are clearly visible in the top panel, but absent from the lower one. Present at approximately 25% intensity are signals (7.18, 7.29 and 7.30 ppm) arising from UBAP1 F268 in the *cis* proline isomer of the peptide, which are unaffected by the presence of HD-PTP$_{CC}$. Effect of electron spin-label at C425: Hα region of 1H-13C HSQC of UBAP1$_{C}$ in the presence of spin-labelled variant of HD-PTP$_{CC}$ (10:1 mixture UBAP1$_{C}$: HD-PTP$_{CC}$). (b) shows spectrum when spin-label is diamagnetic (reduced with excess ascorbate), and (c) shows spectrum when spin-label is paramagnetic.

Suppl. Figure 5.

Sequence alignment of the coiled-coil domains of human HD-PTP and Alix. Alignment done using ESPript 3 online server (Robert, X. and Gouet, P. (2014) "Deciphering key features in protein structures with the new ENDscript server". Nucl. Acids Res. 42(W1), W320-W324).

Suppl. Figure 6. (Phil to fill. Label in HD-PTP should be HD-PTP$_{Bro1\text{-}CC}$

Materials and Methods

Cloning, protein expression and purification. Phil to provide primers info for CC-704

Constructs comprising HD-PTP$_{Bro1}$ (1-361) (sense primer 5'-TACTTCCAATCCATGGAGGCCGTGCCCCGCAT, antisense primer 5'-TATCCACCTTTACTGCTATACCAGTTTGGCAAAGATGTCAGGGCC), HD-PTP$_{CC}$ (363-704?), HD-PTP$_{CC}$ (363-712) (sense primer 5'-TACTTCCAATCCATGGCTGCCCACGAGGCCTCGTCACTGTACAGT antisense primer 5'-TATCCACCTTTACTGCTACAGCTCCCTGTCCAGGAGCTG) and Alix$_V$ (358-702) (sense primer 5'-TACTTCCAATCCATGGTTCCCGTGTCAGTACAGCAGTCTTTG, antisense primer 5'-TATCCACCTTTACTGTCATCTTTCTGTCTTCCGTGCAAAAACTATATC) were generated by PCR amplification and ligation into pNIC28a-Bsa4 vector (Gift from Opher Gileadi (Addgene # 26103)). HD-PTP$_{Bro1\text{-}CC}$ (1-714) (sense primer 5'-GTATCGTATCATATGGAATTCATGGAGGCCGTGC, antisense primer 5'-GGCATACTCGAGTTACTTCTTCAGCTCCCTGTCCAG), and HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ (1-738) (sense primer 5'-GTATCGTATCATATGGAATTCATGGAGGCCGTGC, antisense primer 5'-GGCATACTCGAGTTACACTGCCTCACTCTCCTCCCTGCG) were cloned in pET28a vector with restriction sites Nde1 and Xho1. Point mutants were generated by quick-change primers using Phusion DNA polymerase (New England Biolab). These constructs were transformed in BL-21(DE3) competent *E.coli* cells and protein expression was induced with 0.1 mM IPTG overnight at 20 °C. Overnight grown cultures were harvested by centrifugation at 3,700 x g for 30 mins and cells were re-suspended in 20 mM HEPES 7.0, 500 mM NaCl, 10 mM Imidazole, 2 mM PMSF (Phenylmethylsulphonyl fluoride) lysis buffer. Cells were disrupted by sonication at 4°C and the supernatant clarified by centrifugation at 12,400 x g for 1 h. The His_6-tagged proteins were purified by metal-affinity column chromatography using Nickel-beads (Qiagen) pre-

equilibrated in binding buffer (20 mM HEPES, 500 mM NaCl, 10 mM Imidazole, pH 7.4) and eluted with 250 mM Imidazole in binding buffer. Further purification was performed by anion-exchange chromatography using a MonoQ 5/50 GL column (GE healthcare) in 20 mM HEPES pH 7.4, 2 mM EDTA, 2 mM DTT and eluted with a gradient of NaCl, followed by SEC using a Superdex200 column (GE healthcare) equilibrated with 20 mM HEPES pH 7.0, 0.3 M NaCl, 2 mM EDTA and 2 mM DTT.

Incorporation of L-selenomethionine was done by growing a culture in M9 minimal media (Molecular Dimensions) at 37°C until an OD_{600} of 0.6-0.8 was reached, followed by the addition of phenylalanine, threonine, lysine (100 mg/l each), and isoleucine, leucine and valine (50 mg/l each) and seleno-methionine (80mg/l), prior to induction with 0.1 mM IPTG. Purification was performed as described above.

Multi-angle laser light scattering (MALLS) and analytical ultracentrifugation (AUC).

For MALLS analyses, samples were injected onto a Superdex-200 10/300 GL column (GE healthcare) equilibrated with 20 mM Tris-Cl, pH 8.0, 0.1 M NaCl, 1mM TCEP buffer and the eluted proteins passed through a Wyatt Helios 18-angle laser photometer with Wyatt EOS QELS detector. Concentrations were measured using a Wyatt rEX differential refractive index detector. Light scattering intensities were measured at different angles relative to the incident beam and data analysis was performed with ASTRA 6 software (Wyatt Technology Corp., CA, USA). Protein fractions from MALLS were then used in

sedimentation velocity experiments using either Optima XL-I (HD-PTP$_{CC}$) or XL-A (HD-PTP$_{Bro1\text{-}CC}$) ultracentrifuge (Beckman Instruments) at 50000 ×rpm (18200 g) at 20 ˚C and scanning every 60 or 90 seconds respectively using a wavelength of 280 nm for a total of 200 scans. The sedimentation boundaries were analyzed using the program Sedfit v8.7 (Schuck P., 2000) and hydrodynamic radius (R_h) and frictional ratio (f/f_o) were calculated with Sednterp (Laue, T. M et al., 1992).

Biosensor Binding studies

Biosensor protein binding studies were performed using the multiplex system ProteOn XPR36 surface plasmon resonance instrument (Bio-Rad Laboratories) in 10 mM HePES pH 7.4, 150 mM NaCl, 0.05% Tween 20 as running buffer. His$_6$-tagged proteins were immobilised on a HTE chip (Bio-Rad Laboratories) at a concentration of 50-100 µg/ml. This gave an immobilization level of proteins typically of 5000-8000 response units (RU). All experiments were performed at 25°C. Synthetic peptides (Generon Ltd, UK) were used as analytes in equilibrium binding measurements. Peptide stocks were prepared just prior to the binding experiments and injected (50 µl at 100 µl/min) in the horizontal orientation using five serially diluted peptide concentrations chosen to give a suitable spread of responses below and above half-maximal binding. All the binding sensograms were collected, processed and analyzed using the integrated ProteOn Manager software (Bio-Rad Laboratories), using the equilibrium binding model: Response=[A] * R_{max} / ([A] + K_D) where [A] is the analyte concentration and R_{max} is the maximum response.

EPR-DEER measurements

Protein was purified by affinity chromatography as describe above and then loaded onto a MonoQ 5/50 GL (GE Healthcare) column equilibrated in 20 mM HEPES, 5 mM EDTA, 10 mM DTT pH 7.4 and then eluted using a gradient of 0-500 mM NaCl in the same buffer but with no DTT. A 10-fold molar excess of MTSL (methanethiosulfonate spin label) (Toronto Research Chemicals) was then added from a 100 mM stock solution in methanol. This mixture was incubated at 4°C overnight followed by SEC on a Superdex200 column (GE Healthcare) equilibrated in 20 mM HEPES, 250 mM NaCl, 5 mM EDTA, pH 7.4. Labelling was confirmed by mass spectrometry. All mutants and the wt were checked by circular dichroism to confirm that the addition of MTSL label did not affect the secondary structure of the protein (Suppl. Fig. 3).

For the double electro-electron resonance (DEER), samples were prepared by buffer exchange into deuterated buffer (20 mM HEPES, 250 mM NaCl, pD 7.4) with 30% (v/v) glycerol-d8 and a final protein concentration of 60 µM. 120 µL samples were frozen inside 4 mm quartz tubes (Wilmad) by flash freezing with liquid nitrogen.

DEER experiments were carried out on a pulsed ELEXSYS E580 (9 GHz) spectrometer (Bruker), cooled to 50 K with a continuous-flow helium CF935 cryostat (Oxford Instruments) and a ITC 502 temperature control system (Oxford Instruments). The four-pulse DEER sequence $\pi/2_{\nu obs}-\tau_1-\pi_{\nu obs}-t-\pi_{\nu pump}-(\tau_1+\tau_2-t)-\pi_{\nu obs}-\tau_2-echo$ was applied,[2] with $\pi/2_{\nu obs}$ pulse length of 16 ns, $\pi_{\nu obs}$ pulse length of 32 ns and $\pi_{\nu pump}$ pulse length of 32 ns. Pump pulses were applied at the maximum of the field sweep spectrum with the observe pulses 65

MHz lower. τ_1 was varied by incrementing the first π_{vobs} pulse position over eight steps of 56 ns for averaging of the deuterium nuclear modulation. [3] Phase-cycling was applied. DEERAnalysis2013.2 was used to subtract the exponential background decay due to intermolecular interactions and to calculate the interspin distance distribution by Tikhonov regularization.[4] Distance distribution simulations were calculated by *in silico* spin labelling of the crystal structure (molecule A) using the R1 side chain rotamer library incorporated into program MMM 2013.2 (Multiscale Modelling of Macromolecular systems).[5]

Circular dichroism measurements

Samples for circular dichroism were adjusted to 0.1 mg/mL and buffer exchanged into phosphate buffer (100 mM potassium phosphate, 100 mM potassium fluoride, pH 7.4). Spectra were recorded on a Jasco J-180 spectropolarimeter from 190-260 nm using cuvettes of 0.5 mm path length. A data pitch of 0.2 nm was used with a response time of 8 seconds per point. Units were converted to mean residue ellipticity.

NMR and PRE measurements

For NMR samples, the UBAP1$_C$ peptide was dissolved at a concentration of 1mg/ml (~0.5mM) in phosphate buffered saline pD 7.5 in D_2O. 3mm NMR tubes were used throughout to reduce the Q of the samples. $^1H^{13}C$ gradient selected HSQC spectra were recorded at natural abundance 13C (1%) at 800 MHz, using a Bruker AVANCE III spectrometer equipped with a TCI (1H-13C-15N/2H) cryoprobe with *z*-gradients. To aid assignment, pure-shift (PUSH) HSQC spectra

(Paudel, L, 2013) were recorded for the uncomplexed peptide. Assignment of the H-C correlations was by HSQC-TOCSY and NOESY spectrum, and confirmed by the behavior in the protein complexes. NMR assignments will be deposited at BMRB database. Complexation with HD-PTP$_{CC}$ was detected in a similar sample with the addition of 60μM of the protein.

PRE measurements were recorded by taking MTSL labelled HD-PTP$_{CC}$ C697S mutant under the same conditions, recording $^{1}H^{13}C$ HSQCs, and then reducing the MTSL label (rendering it diamagnetic) with 10-fold excess of sodium ascorbate (added from a 1M stock). HSQC spectra typically used 48 scans with 0.2 s acquisition time in the direct dimension, with 12.5ppm spectral width centred at 4.7ppm, and 256 indirect dimension pairs with 70 ppm spectral width centred at 35ppm. Bi-level adiabatic decoupling was applied during acquisition.

Crystallization and Structure determination

Prior to crystallization, the His$_6$-tag of HD-PTP$_{CC}$ was cleaved overnight at 4 °C with His$_6$-tag TEV protease followed by Nickel-affinity chromatography. Digested HD-PTP$_{CC}$ was further purified using a MonoQ column and Superdex200 size-exclusion column equilibrated with 20 mM HePES pH 7.0, 0.3 M NaCl, 2 mM EDTA and 2 mM DTT. Fractions were pooled and concentrated to 11 mg/ml. Optimised crystals (for the native and Se-derivatised proteins) were obtained by the sitting drop vapour difffusion method. Crystallisation trials were comprised of 180 nl of protein, 20 nl of seed stock mixed with 200 nl of reservoir solution (0.1 M Bis-Tris pH 6.0, 0.1-0.2 M Na-Formate and 13-15% PEG$_{3350}$). Plate-like crystals appeared overnight at 21 °C.

The crystals of the HD-PTP$_{CC}$-UBAP1$_{C}$ complex were obtained by co-crystallisation. A sample of purified HD-PTP$_{CC}$ (1mg/ml) was incubated with a 1mM stock of the peptide solubilised in SEC buffer for a period of 1 hour Following incubation, this sample was concentrated to a final protein concentration of ~ 11 mg/ml and crystalisation trials were performed as described above (seed stock generated from HD-PTP$_{CC}$ crystals). Crystals of the complex were obtained from a reservoir solution of 0.2 M KSCN, 20% PEG$_{3350}$. All crystals were cryo protected in perfluropolyether cryo oil (PFO, Hampton Research) prior to plunge freezing in liquid nitrogen.

Single crystals of both the HD-PTP$_{CC}$ and HD-PTP$_{CC}$ -UBAP1$_{C}$ complex were exposed at i04 of Diamond Light Source (UK) and all data subsequently scaled and merged with XDS. The structure of HD-PTP$_{CC}$ was determined by selenium single-wavelength anomalous diffraction (Se-SAD) with data collected at a wavelength of (0.9795Å). Initial phasing was achieved through the identification of 11 Selenium sites using PHENIX Autosol, and partial automated building into density modified maps using Phenix Autobuild (Adams et al., 2010). The structure was extended and completed by iterative manual model building in COOT (Emsley et al., 2010) and refinement using PHENIX Refine (Adams et al., 2010). Validation with MolProbity and PDB_REDO were integrated into the iterative rebuild and refinement procedure. The data collection and refinement statistics are summarized in Table 1.

[AU: Ramachandran statistics should be in Methods section at the end of Refinement subsection.]

The structure of HD-PTP$_{CC}$ -UBAP1$_{C}$ was determined by MR….etc

[AU: Ramachandran statistics should be in Methods section at the end of Refinement subsection.]

Coordinates will be deposited in the PDB ….etc

Small-angle X-Ray Scattering Analysis

The sample HD-PTP$_{Bro1\text{-}CC}$ was concentrated up to 12 mg/ml in 50 mM Tris-Cl pH 8.0, 100 mM NaCl, 2 mM EDTA and 2 mM DTT. HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$ was prepared at 7.5 mg/ml in 50 mM Tris-Cl pH 8.0, 250 mM NaCl, 2 mM EDTA and 2 mM DTT. SAXS data was collected at beamline X33 Hamburg DESY (HD-PTP$_{Bro1\text{-}CC}$) and beamline BM29 ESRF (HD-PTP$_{Bro1\text{-}CC\text{-}PRR}$). HD-PTP$_{CC}$ was prepared in 20 mM HEPES pH 7.0, 300 mM NaCl, 2 mM EDTA and 2 mM DTT at 2.8 mg/ml and SAXS data was collected at BM29, ESRF. Data processing was performed with the ATSAS suite (Petoukhov et al, 2012). The forward scattering *I(0)* and the radius of gyration *Rg* were estimated with PRIMUS (Konarev et al, 2003) using the Guinier approximation. GNOM (ref), was used to compute the pairwise intraparticle distance distribution function p(r) and D_{max}. Particle shapes were restored ab-initio using GASBOR (Svergun et al., 2001). Twenty simulations were performed and the outputs were averaged and filtered using DAMAVER (Volkov et al., 2003) to produce the final envelopes with a normal spatial discrepancy value of 0.61-0.67 for the DAMMIN models and 1.4-1.8 for the GASBOR models.

The crystallographic structures of the HD-PTP$_{Bro1}$ (PDB ID: 3RAU) and the crystallographic structure of HD-PTP$_{CC}$ (this study) were used to generate molecular models employing torsional angle molecular dynamics (TAMD) with CNS (Brunger 2007). The hydrodynamic and dimensional parameters for all the models generated were calculated with SOMO (Rai 2005), and then used to select suitable conformers that comply with the experimentally determined parameters, as previously described (Tariq 2015). The selected pool of models

was contrasted to the SAXS profile using FoXS (D. Schneidman 2010). The Coral program implemented in ATSAS package was used to add the missing residues in the Bro1-738 protein using the Bro1-714 model.

Yeast two-hybrid analysis

HD-PTP$_{Bro1\text{-}CC}$, HD-PTP$_{Bro1\text{-}CC}$ F678D and Alix $_{Bro1\text{-}V}$(1-xxx), cloned into pGBKT7, were used as described previously (ref). UBAP1 (122-308) were cloned into pGADT7. Further deletion and missense mutations in UBAP1 as indicated were generated by standard PCR-based mutagenesis. Interactions were tested using the Clontech "Matchmaker Gold" system (Clontech) as described previously {Stefani:2011dy} with the following modifications. Test plasmids were co-transformed into the yeast strain Y2HGold and grown on minimal media agar plates (Double Drop-out; DDO) deficient in tryptophan and leucine, to select for the presence of the bait (pGBKT7) and prey (pGADT7) plasmids, respectively. The medium was supplemented with 0.4% D+ Glucose and 1% D+ Galactose to allow optimal yeast growth and expression of all constructs. Transformant colonies were inoculated in triplicate (per condition) into 2.0 ml liquid DDO media containing 2% D+ Glucose and grown for 24-48 hours. A 5µl inoculating loop was used to transfer some of each liquid culture to a square on a DDO and QDO (additionally deficient in histidine and adenine) plates, with 0.4% D+ Glucose and 1% D+ Galactose. Each triplicate experiment was repeated at least 3 times.

UBAP1 *in vitro* translation and binding to HD-PTP

BL21 cells were transformed with His$_6$-Bro1-714 and protein expressed using IPTG induction at 30°C. The culture pellet was resuspended in lysis buffer (50 mM Tris-HCl pH7.5, 0.5M NaCl, 6 mM imidazole) containing EDTA-free protease inhibitor cocktail (Roche). Lysates were generated using a French Press (Thermo

Fisher) and pelleted at 100,000g_{av} for 1 hr. Protein was purified using a NTA Sepharose High Performance column (GE Healthcare) using an imidazole gradient to elute.

UBAP1-strep encoded on a pTriex5 vector (Stefani) was amplified using Pwo Polymerase (Roche). UBAP1 RNA was synthesised from the PCR product using T7 RNA polymerase (Promega, location). Protein was translated in nuclease treated rabbit reticulocyte lysate (Promega, location) containing ^{35}S-methionine (Perkin Elmer, location), and 100 units of RNasin (Promega) for 1 hour at 30°C, followed by 10 min in the presence of 1 mM puromycin. 20 µl translated protein was incubated with 5 µg His_6-HD-PTP(1-714) in 250 µl IP buffer (20 mM Hepes, pH7.4, 100 mM NaCl, 1 mM $MgCl_2$, 1% (w/v) Triton X-100) for 2 hr at 4°C, then overnight with 3µl anti-His antibody (Clone xxx, Sigma). Samples were incubated with 20µl protein A-sepharose beads (Invitrogen) for 2 hr at 4°C, then washed 3 x in IP buffer.

Cell Culture, transfection and siRNA rescue experiments

HeLa cells were grown in DMEM supplemented with 1% NEAA, 10% Foetal Calf Serum (HyClone; Perbio) and 1% Pen-Strep. Transient transfections were performed using Fugene 6 (Roche). Transfection with HD-PTP siRNA was performed using Interferin (QBiogene). using the following nucleotide, as published previously (doyotte):

HDPTP: 5'-GCAAACAGCGGAUGAGCAA-3' 19 check

A non-targeting siRNA (Dharmacon) was used as a control. Knock-down efficiency for HD-PTP was assessed by Western blotting, and efficient depletion

was obtained, as previously reported (Doyotte) . For siRNA rescue experiments, cells were knocked down for 24 hours, then transfected as appropriate with WT or specified mutants of HD-PTP(1-714) (note that the siRNA oligo targets a C-terminal sequence within HD-PTP). Rescues were assessed by visual quantification of phenotypes. For this, cells containing highly clustered and strongly labeled foci of cytoplasmic FK2 staining (previously identified as endosomal; ref) were considered knocked-down. 'Rescued' cells displayed a WT, diffuse cytoplasmic distribution of FK2 staining. Scoring was performed for 3 independent experiments for every construct, with at least 100 transfected cells examined on each occasion. Data was subjected to statistical analysis (x test) using Prism5 software (Graphpad). For graphical representation of rescue data, mean scores +/- SD are provided for the 3 determinations.

References

[2] Pannier, M., Veit, S., Godt, A., Jeschke, G., and Spiess H.W., Dead-Time Free Measurement of Dipole–Dipole Interactions between Electron Spins, 2000*, J. Magn. Reson.*, **142**, 331–40. doi:10.1006/jmre.1999.1944.

[3] Jeschke, G., Bender, A., Paulsen, H., Zimmermann, H., and Godt, A., Sensitivity Enhancement in Pulse EPR Distance Measurements, 2004, *J. Magn. Reson.*, **169**, 1–12. doi:10.1016/j.jmr.2004.03.024.

[4] Jeschke, G., Chechik, V., Ionita, P., Godt, A., Zimmermann, H., Banham, J., Timmel, C.R., Hilger, D., and Jung, H., DeerAnalysis2006 - a Comprehensive Software Package for Analyzing Pulsed ELDOR Data, 2006, *Appl. Magn. Reson.*, **30**, 473–98. doi:10.1007/BF03166213.

[5] Jeschke, G., Polyhach, Y., 2013, http://www.epr.ethz.ch/software/MMM_manual_2013_2.pdf.

[6] Fielding, A.J., Concilio, M.G., Heaven, G., and Hollas, M.A., New Developments in Spin Labels for Pulsed Dipolar EPR, 2014, *Molecules*, **19**,16998–25. doi:10.3390/molecules191016998.

[8] Jeschke, G., and Polyhach, Y., Distance Measurements on Spin-Labelled Biomacromolecules by Pulsed Electron Paramagnetic Resonance, 2007, *Phy. Chem. Chem. Phys.*, **9**, 1895–1910. doi:10.1039/B614920K.

Paudel, L, Adams RW, KirÃ¡ly P, Aguilar JA, Foroozandeh M, Cliff MJ, Nilsson M, SÃ¡ndor P, Waltho JP, Morris GA. 2013. Simultaneously Enhancing Spectral

Resolution and Sensitivity in Heteronuclear Correlation NMR Spectroscopy. Angewandte Chemie International Edition. 52(44):11616-11619.

11]. Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V. and Svergun, D.I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis *J. Appl. Cryst.* 45, 342-350

[12]. V. V. Volkov and D. I. Svergun (2003). Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* 36, 860-864.

[13]. D. I. Svergun (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* 2879-2886.

[14]. Svergun, D.I., Petoukhov, M.V. and Koch, M.H.J. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946-2953.

[15]. P.V.Konarev, V.V.Volkov, A.V.Sokolova, M.H.J.Koch and D. I. Svergun (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J Appl Cryst*. 36, 1277-1282.

[16]. Schuck P (2000) Sizedistribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 78:1606 –1619.

[17]. Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. Analyticall Ultracentrifugation in Biochemistry and Polymer Science, *Royal Society of Chemistry* (Edited by S. Harding & A. Rowe). 90–125 (1992).