

Vadim Rimsa, Thomas Eadsforth  
and William N. Hunter\*

Division of Biological Chemistry and Drug  
Discovery, College of Life Sciences, University  
of Dundee, Dundee DD1 5EH, Scotland

Correspondence e-mail:  
w.n.hunter@dundee.ac.uk

Received 1 November 2010  
Accepted 16 February 2011

PDB References: SENP1, 2xph; 2xre.

## The role of $\text{Co}^{2+}$ in the crystallization of human SENP1 and comments on the limitations of automated refinement protocols

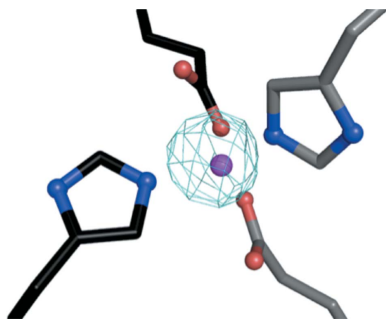
Metal ions often stabilize intermolecular contacts between macromolecules, thereby promoting crystallization. When interpreting a medium-resolution electron-density map of the catalytic domain of human sentrin-specific protease 1 (SENP1), a strong feature indicative of an ordered divalent cation was noted. This was assigned as  $\text{Co}^{2+}$ , an essential component of the crystallization mixture. The ion displays tetrahedral coordination by Glu430 and His640 from one molecule and the corresponding residues from a symmetry-related molecule. Analysis of the data derived from a previous structure of SENP1 suggested that  $\text{Co}^{2+}$  had been overlooked and re-refinement supported this conclusion. High-throughput automated re-refinement protocols also failed to mark the  $\text{Co}^{2+}$  position, supporting the requirement for the incorporation of as much information as possible to enhance the value of such protocols.

### 1. Introduction

The conjugation of a small ubiquitin-like modifier (SUMO) onto target proteins is a reversible post-translational modification that regulates many processes including gene expression, the cell cycle and stress responses (Cheng *et al.*, 2006). Sentrin-specific protease 1 (SENP1), one of six SUMO-deconjugating enzymes in humans, is overexpressed in prostatic intraepithelial neoplasia and prostate cancer lesions and promotes androgen receptor (AR) dependent transcription and cell proliferation. The siRNA-mediated ablation of SENP1 expression in prostate cancer cells significantly decreases AR-dependent proliferation, while transgenic mice with targeted overexpression of SENP1 in the prostate develop prostatic intraepithelial neoplasia (Bawa-Khalfe *et al.*, 2007; Kaikkonen *et al.*, 2009). These data indicate that SENP1 overexpression is intimately associated with the development of prostate cancer and that inhibition of SENP1 may offer a therapeutic approach. Small-molecule inhibitors of SENP1 are sought to further investigate this potential drug target.

The structures of the catalytic domain of SENP1 and its complex with SUMO have been determined previously (see, for example, Shen *et al.*, 2006). The enzyme contains a catalytic triad (histidine, aspartate and cysteine) and a conserved glutamine residue required for the formation of an oxyanion hole in the active site (Yeh, 2008). SENP1 does not appear to require metal binding for either structural stability or catalytic activity.

We plan structural studies of human SENP1–ligand interactions and required a supply of protein for cocrystallization experiments and crystals suitable for soaking in of ligands. Here, we used similar crystallization conditions to those previously determined (Shen *et al.*, 2006) and elucidated the structure of the SENP1 catalytic domain at 2.4 Å resolution. The presence of  $\text{CoCl}_2$  is essential for crystal formation and in the analysis we observed an ordered  $\text{Co}^{2+}$  ion that bridges symmetry-related molecules. Our reinterpretation of the published data revealed that this ion has previously been overlooked in the initial study and also by automated re-refinement protocols (Joosten *et al.*, 2009).



© 2011 International Union of Crystallography  
All rights reserved

**Table 1**  
Crystallographic statistics.

Values in parentheses are for the highest resolution shell.

	2xph	2xre (2iyc re-refined)	2iyc
Space group	<i>P</i> 3 <sub>1</sub> 21	<i>P</i> 3 <sub>1</sub> 21	<i>P</i> 3 <sub>1</sub> 21
Unit-cell parameters (Å)	<i>a</i> = <i>b</i> = 71.17, <i>c</i> = 199.99	<i>a</i> = <i>b</i> = 71.98, <i>c</i> = 200.64	<i>a</i> = <i>b</i> = 71.98, <i>c</i> = 200.64
Resolution (Å)	45–2.40 (2.53–2.40)	45–2.45 (2.51–2.45)	54–2.45 (2.51–2.45)
No. of reflections recorded	120621 (17892)	NR†	NR†
Unique reflections	23808 (3415)	21829 (1555)	21832 (1557)
Completeness (%)	99.8 (100.0)	99.9 (100.0)	100.0 (100.0)
Average multiplicity	5.1	8.8	8.8
$\langle I/\sigma(I) \rangle$	8.5 (2.5)	19.0 (3.0)	19.0 (3.0)
Wilson <i>B</i> factor (Å <sup>2</sup> )	59.5	65.5	65.5
Radiation source and beamline	ESRF ID29	ESRF ID14-4	ESRF ID14-4
Wavelength (Å)	0.977	0.979	0.979
Residues	418–644	418–644	419–644
Water/glycerol/Co <sup>2+</sup>	70/6/1	67/10/1	90/0/0
$R_{\text{merge}}^{\ddagger}$ (%)	9.0 (42.8)	10.0 (43.0)	10.0 (43.0)
$R_{\text{work}}^{\S}$ (%)	23.1	24.3	21.9
$R_{\text{free}}^{\P}$ (%)	31.3	32.5	26.7
Average <i>B</i> factor for all atoms (Å <sup>2</sup> )	58.7	51.7	28.9
Cruickshank DPI†† (Å)	0.38	0.45	0.41
Real-space <i>R</i> value‡‡	0.172 (0.071)	0.183 (0.083)	0.119 (0.036)
Real-space correlation coefficient‡‡	0.927 (0.063)	0.912 (0.090)	0.936 (0.042)
Significant regions‡‡			
Chain <i>A</i>	18 outliers [8%]	6 outliers [3%]	0 outliers [0%]
Chain <i>B</i>	20 outliers [9%]	30 outliers [13%]	2 outliers [1%]
Ramachandran plot§§			
Most favoured (%)	91.6	92.5	96.9
Additional allowed (%)	8.0	7.5	2.2
Outliers (%)	0.4	0.0	0.9
R.m.s.d. on ideal values¶¶			
Bond lengths (Å)	0.02	0.01	0.02
Bond angles (°)	1.62	1.49	1.48

† Not reported. ‡  $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the intensity of the *i*th measurement of reflection *hkl* and  $\langle I(hkl) \rangle$  is the mean value of  $I_i(hkl)$  for all *i* measurements. §  $R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  is the observed structure factor and  $F_{\text{calc}}$  is the calculated structure factor. ¶  $R_{\text{free}}$  is the same as  $R_{\text{work}}$  except calculated with a subset (5%) of data that were excluded from refinement calculations. †† Diffraction-component Precision Indicator (Cruickshank, 1999). ‡‡ Calculated using the Electron Density Server (Kleywegt *et al.*, 2004). §§ Chen *et al.* (2010). ¶¶ Engh & Huber (1991).

## 2. Methods

### 2.1. Protein expression and purification

The gene fragment encoding the catalytic domain of human SENP1 (amino acids 415–644; UniProt ID Q9P0U3; EC 3.4.22.68) was provided by Ron Hay (University of Dundee) in a pHISTEV30a vector and expressed an N-terminally His-tagged protein. The construct was verified by DNA sequencing (DNA Sequencing Unit, University of Dundee). The recombinant protein was expressed in *Escherichia coli* BL21 (DE3) Gold cells (Novagen), purified using an Ni<sup>2+</sup>-charged HisTrap HP column (GE Healthcare) and dialyzed in buffer (50 mM Tris–HCl, 250 mM NaCl pH 7.5) in the presence of tobacco etch virus (TEV) protease to remove the His tag. After TEV protease cleavage, the protein was further purified by a second nickel-affinity step, which removed uncleaved protein, followed by a gel-filtration step (Superdex 200 column; GE Healthcare). Protein concentration was determined using a theoretical molar extinction coefficient of 37 930 M<sup>−1</sup> cm<sup>−1</sup> at 280 nm and the purity of the sample, which was estimated at >95%, was checked by SDS–PAGE and mass spectrometry (Fingerprint Proteomics Facility, University of Dundee). The single protonated species as observed by mass spectrometry had a mass of 28 096 kDa, which is in close agreement with the theoretical mass of 28 041 kDa.

### 2.2. Crystallization, data collection and structure determination

A number of commercially available screens were used in initial attempts at crystallization. Some of these solutions contained di-

valent cations, including Ca<sup>2+</sup>, Cd<sup>2+</sup>, Co<sup>2+</sup>, Mg<sup>2+</sup>, Ni<sup>2+</sup> and Zn<sup>2+</sup>. The screening did not identify any conditions that we judged suitable for optimization. Crystals of the human SENP1 catalytic domain were then obtained at room temperature by hanging-drop vapour diffusion based on previously published conditions (Shen *et al.*, 2006). The reservoir solution consisted of 1.8 M ammonium sulfate, 50 mM CoCl<sub>2</sub> and 100 mM MES pH 6.5. Single crystals appeared after 2 d from equal volumes of protein solution (20 mg ml<sup>−1</sup> in 20 mM Tris–HCl pH 8.0 and 50 mM NaCl) and reservoir solution. Crystals were cryoprotected in reservoir buffer containing 20% glycerol. Diffraction data were collected on European Synchrotron Radiation Facility beamline ID29. The data were indexed with *MOSFLM* (Leslie, 2006) and scaled using *SCALA* (Evans, 2006) from the *CCP4* program suite (Collaborative Computational Project, Number 4, 1994). The structure was solved by molecular replacement with *Phaser* (Read, 2001) using the coordinates of a monomer (chain *A*) of human SENP1 (PDB code 2iyc; Shen *et al.*, 2006) as a search model. The model was manipulated using *Coot* (Emsley & Cowtan, 2004) and refinement was carried out using *REFMAC5* (Murshudov *et al.*, 1997). Water and glycerol molecules were included in the model and a positive peak (>10σ) in the  $F_o - F_c$  difference map was assigned as a Co<sup>2+</sup> ion. The refinement continued until no significant changes in  $R_{\text{work}}$  and  $R_{\text{free}}$  were observed and until inspection of the difference density map suggested that no further corrections or additions were required. The stereochemistry of the structure was checked using *MolProbity* (Chen *et al.*, 2010). Data-collection and structure-refinement statistics are shown in Table 1.

### 3. Results and discussion

#### 3.1. Structure determination

The crystal structure of human SENP1 was redetermined at 2.4 Å resolution with  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 23.1% and 31.3%, respectively. Whilst these  $R$ -factor values fall within the range observed for structures at comparable resolution, the difference of 8.2% would be considered to be on the high side. Our general experience is that the discrepancy between  $R_{\text{work}}$  and  $R_{\text{free}}$  is often greater when derived from diffraction data displaying larger  $R_{\text{merge}}$  values. In this case  $R_{\text{merge}}$  is 9.0%. However, a further comment on this point will be made below.

The asymmetric unit contains two polypeptides, referred to as *A* and *B*, comprising residues 418–644 and 419–644, respectively, 70 waters, six glycerol molecules and one  $\text{Co}^{2+}$  ion. Most of the amino acids had well defined electron density, apart from several poorly ordered residues on the surface of the protein. These atoms were included in the model but their occupancy was set at zero. In this respect, therefore, our refinement models consist of a reduced number of protein atoms.

#### 3.2. Overall structure

The crystal structure of the human SENP1 catalytic domain was previously determined at 2.45 Å resolution (PDB code 2iyc; Shen *et al.*, 2006). We used similar crystallization conditions to obtain isomorphous crystals, although we found it beneficial to reduce the concentration of  $\text{CoCl}_2$  in the reservoir from 100 to 50 mM. Least-squares superposition of the asymmetric unit of our SENP1 structure on that of PDB entry 2iyc gives an r.m.s.d. (root-mean-square deviation) of 0.78 Å for 452  $\text{C}^\alpha$  atoms. The SENP1 catalytic domain displays the cysteine protease superfamily fold. The catalytic domain consists of a five-stranded  $\beta$ -sheet, in which the middle strand is antiparallel to the other four, positioned between several helices. The residues Cys603, His533 and Asp550 are in close proximity and form a catalytic triad (data not shown).

#### 3.3. The role of $\text{Co}^{2+}$ in stabilizing the crystal lattice

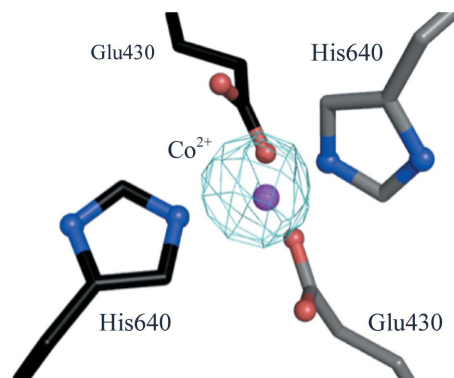
The interactions that promote crystallization are primarily non-covalent and are dominated by hydrogen-bonding or salt-bridge associations with polar residues and generally a minor contribution from van der Waals forces (Dasgupta *et al.*, 1997). Often, water molecules and cations can bridge and stabilize intermolecular contacts (Durbin *et al.*, 1996). Here, the  $F_o - F_c$  OMIT electron-density map of our SENP1 structure (PDB code 2xph) revealed the presence of an ordered  $\text{Co}^{2+}$  ion (Fig. 1). The ion binds at the surface of the protein and mediates interactions between neighbouring asymmetric units. The residues that most often coordinate  $\text{Co}^{2+}$  ions in proteins are histidine, followed by aspartate and glutamate (Dokmanić *et al.*, 2008). In this case, the residues that provide tetrahedral coordination are Glu430 and His640 from chain *B* and the corresponding residues from chain *A* of a symmetry-related molecule (symmetry operation  $-y, x - y, z + 1/3$ ; Fig. 1).

The coordination number displayed by  $\text{Co}^{2+}$  is variable since there is no particular stabilization of the  $d^7$  configuration. The most common coordination number assigned to  $\text{Co}^{2+}$  is six, followed closely by five, but in 10% of protein structures investigated tetrahedral coordination is observed (Dokmanić *et al.*, 2008). In that study the authors considered any distance of less than 3.0 Å to represent coordination, a value that in our opinion is too long for metal–oxygen and metal–nitrogen coordinate bonds since it exceeds the sum of the ionic/crystal radii of the atoms and ions involved. The distances

between cobalt and electron donors vary depending on the functional groups of the electron donors. The mean distances, derived from analysis of high-resolution structures, between cobalt and imidazole nitrogen and between cobalt and carboxyl oxygen are  $2.07 \pm 0.12$  and  $2.22 \pm 0.19$  Å (Dokmanić *et al.*, 2008), respectively. The distances observed in our SENP1 structure were approximately 2.0 Å for the Co–NE2 and 1.8 Å for the Co–OE1 separation. We did not impose any restraints on these distances, but they reflect the medium resolution of the analysis. The ability of imidazole and carboxyl groups to coordinate cations depends upon their protonation state; at near-neutral pH, such as in the crystallization conditions, glutamate and histidine residues can readily coordinate  $\text{Co}^{2+}$ . The carboxylate groups of the two Glu430 residues coordinate  $\text{Co}^{2+}$  in a *syn*-monodentate fashion using OE1, whilst the OE2 atoms accept hydrogen bonds donated from Trp636 NE1. This *syn*-monodentate mode of binding for carboxylate groups is the most prevalent form of carboxylate–metal-ion coordination in both the Protein Data Bank and the Cambridge Crystallographic Database (Dudev & Lim, 2004). The distances of OE2 from the cation are 2.7 and 2.6 Å for Glu430 from molecules *A* and *B*, respectively. This, together with the acute angle values of less than  $60^\circ$  subtended at  $\text{Co}^{2+}$  by OE1 and OE2, suggests to us that only the contact with OE1 should be considered as a coordinate bond.

Other residues also contribute to the inter-asymmetric unit contacts in the crystals (data not shown); for example, the hydrogen bonds donated from Arg641 of molecule *A* to Glu425 and Glu426 from a symmetry-related molecule *B* (symmetry operation  $-x + y, -x, z + 2/3$ ). Lys64 of molecule *B* makes side-chain and main-chain interactions with the Glu422 main-chain and Thr424 side-chain atoms, respectively, of the symmetry-related molecule *A* (symmetry operation  $-y, x - y, z + 1/3$ ). Thus, coordination of  $\text{Co}^{2+}$  together with other intermolecular interactions contributes to crystal lattice formation. We note that one molecule in the asymmetric unit, molecule *B*, is less well defined by the electron density. This is likely to be a reflection of the smaller surface area and fewer interactions that are involved in intermolecular contacts in the crystal lattice for this molecule (data not shown).

We wondered whether the  $\text{Co}^{2+}$  ion had been inadvertently omitted in the previous structure of this enzyme (PDB code 2iyc) and so re-refined that structure. The  $\text{Co}^{2+}$  was indeed evident, again at a level of  $>10\sigma$  positive density in a difference map (data not shown).



**Figure 1**

The  $F_o - F_c$  OMIT electron density for  $\text{Co}^{2+}$ .  $F_o$  are the observed and  $F_c$  are the calculated structure factors. The map is contoured at  $5\sigma$  (cyan chicken wire). Glu430 and His640 from two asymmetric units are represented as ball-and-stick models. Carbon positions are coloured grey for molecule *B* and black for the symmetry-related molecule *A*. Oxygen and nitrogen positions are coloured red and blue, respectively; the purple sphere represents  $\text{Co}^{2+}$ .

We also included several glycerol molecules in the model on the basis of strong features in the electron-density and difference density maps. The new coordinates have been deposited in the PDB with accession code 2xre. The crystallographic statistics for the re-refined model and (for comparison) statistics relevant to 2iyc are given in Table 1. The re-refinement resulted in  $R_{\text{work}}$  and  $R_{\text{free}}$  values greater than in the original structure determination. This may be the result of fewer atoms and a different  $B$ -factor model.

The  $\text{Co}^{2+}$  ion and glycerol molecules are not of physiological relevance, but their omission means that the model was incomplete and the new data inform our experimental strategies as we seek to characterize SENP1–ligand complexes.

### 3.4. Comments on automated refinement approaches

PDB\_REDO is an automated high-throughput data bank that incorporates re-refined PDB structures using the most up-to-date methods (Joosten *et al.*, 2009). This is an excellent initiative and one that we use, and would encourage others to use, to obtain optimized structural models. The protocols applied in PDB\_REDO work over a wide resolution range and employ methods such as translation/libration/screw (TLS) refinement and different weighting schemes and in many cases clearly improve the models deposited in the PDB. However, as yet there is limited input and the assignment of previously unrecognized ligands that bind to protein structures has not been incorporated into PDB\_REDO protocols. PDB\_REDO failed to detect or correct errors around the  $\text{Co}^{2+}$  ion in PDB entry 2iyc, but inspection of the electron-density maps clearly indicated that the divalent cation was present in the same position as in the 2xph structure. In the 2iyc structure the side chain of Glu430 had been positioned in density that corresponds to the  $\text{Co}^{2+}$  ion and this error remained in the PDB\_REDO entry (data not shown). This reiterates an earlier observation (Joosten *et al.*, 2009) that although automated re-refinement protocols are a useful tool and are constantly being improved, a need still remains for manual intervention in some instances.

The submission of PDB entry 2xre (2iyc re-refined) to PDB\_REDO, now with the assigned metal ion and glycerol molecules, produced improved statistics, with the  $R_{\text{work}}$  and  $R_{\text{free}}$  values reduced to 22.2% and 26.0%, respectively. Strikingly, the difference between  $R_{\text{work}}$  and  $R_{\text{free}}$  has been reduced from 8.2% to 3.8%. The major difference that we observe between the two models is that PDB\_REDO has selected weaker  $B$ -factor restraints and produced lower  $B$  factors. The lesson for us is that it might be useful to test more  $B$ -factor weighting schemes during refinement. Note, however, that PDB\_REDO derives lower  $R_{\text{work}}$  and  $R_{\text{free}}$  values (20.3% and 24.3%) for 2iyc, *i.e.* the model without the  $\text{Co}^{2+}$  ion or the glycerol molecules!

We anticipate that the inclusion of information concerning the chemical entities present in the crystallization mixtures together with

improved automated refinement and electron-density interpretation protocols will lead to significant improvements in facilities such as PDB\_REDO. Work to address the issue of metal-ion identification in protein crystal structures using valence calculations was carried out more than a decade ago (Nayal & Di Cera, 1996) and such approaches, together with the options available in *Coot* (Emsley & Cowtan, 2004), offer a means to progress in this aspect of macromolecular crystallography. It remains as important as ever that investigators accurately report crystallization protocols that might inform decision making during any refinement process and, as a reviewer of our manuscript pointed out, storage of such information in a convenient machine-readable format might be a useful advance in this respect.

VR is funded by the Medical Research Council and AstraZeneca. The WNH laboratory is supported by The Wellcome Trust (WT082596 and WT083481). We acknowledge the European Synchrotron Radiation Facility for beam time and excellent staff support and are grateful to Professor Ron Hay for provision of the plasmid.

### References

- Bawa-Khalife, T., Cheng, J., Wang, Z. & Yeh, E. T. (2007). *J. Biol. Chem.* **282**, 37341–37349.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Cheng, J., Bawa, T., Lee, P., Gong, L. & Yeh, E. T. (2006). *Neoplasia*, **8**, 667–676.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cruickshank, D. W. J. (1999). *Acta Cryst.* **D55**, 583–601.
- Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. (1997). *Proteins*, **28**, 494–514.
- Dokmanić, I., Šikić, M. & Tomić, S. (2008). *Acta Cryst.* **D64**, 257–263.
- Dudev, T. & Lim, C. (2004). *J. Phys. Chem. B*, **108**, 4546–4557.
- Durbin, S. D. & Feher, G. (1996). *Annu. Rev. Phys. Chem.* **47**, 171–204.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Joosten, R. P. *et al.* (2009). *J. Appl. Cryst.* **42**, 376–384.
- Kaikkonen, S., Jääskeläinen, T., Karvonen, U., Rytinki, M. M., Makkonen, H., Gioeli, D., Paschal, B. M. & Palvimo, J. J. (2009). *Mol. Endocrinol.* **23**, 292–307.
- Kleywegt, G. J., Harris, M. R., Zou, J., Taylor, T. C., Wählby, A. & Jones, T. A. (2004). *Acta Cryst.* **D60**, 2240–2249.
- Leslie, A. G. W. (2006). *Acta Cryst.* **D62**, 48–57.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Nayal, M. & Di Cera, E. (1996). *J. Mol. Biol.* **256**, 228–234.
- Read, R. J. (2001). *Acta Cryst.* **D57**, 1373–1382.
- Shen, L. N., Dong, C., Liu, H., Naismith, J. H. & Hay, R. T. (2006). *Biochem. J.* **397**, 279–288.
- Yeh, E. T. H. (2008). *J. Biol. Chem.* **284**, 8223–8227.