Crystal Structure of *Cydia Pomonella* granulovirus PTP-2

Authors

Guangmei Huang[a], Michael R. Oliver[a], Jeremy R. Keown[ab], David C. Goldstone[a] and Peter Metcalf[a]*

[a]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand
[b]Division of Structural Biology, University of Oxford, Oxford, United Kingdom

Correspondence email: Peter.Metcalf@auckland.ac.nz

Abstract Many viral genomes encode kinase and phosphatase enzymes to manipulate pathways controlled by phosphorylation events. The majority of viral phosphatase genes occur in the *Baculoviridae* and *Poxviridae* families of large DNA viruses. The corresponding protein sequences belong to four major homology groups and structures are currently available for only two of these. Here we describe the first structure from the third group, the protein tyrosine phosphatase-2 class of viral phosphatases. We show that *Cydia pomonella* granulovirus PTP-2 has the same general fold and active site architecture as described previously for other phosphatases, despite lacking significant sequence homology. It has in addition a novel projecting C-terminal extension in an area corresponding to the interface of dimeric poxvirus phosphatases belonging to the Tyr-Ser protein phosphatase homology group.

Keywords: *Cydia pomonella* granulovirus; PTP-2; phosphatase

Synopsis The crystal structure of the betabaculovirus *Cydia pomonella* granulovirus protein phosphatase PTP-2 has been determined at 1.65Å resolution. The structure shows that the molecule has a typical protein phosphatase fold with a unique β-sheet extension at the C-terminus.

1. Introduction

Protein kinases and phosphatases regulate cellular processes by adding or removing phosphates at specific sites on target proteins. These two enzyme classes together constitute about 3% of eukaryotic proteomes (Chen *et al.*, 2017). Approximately 30% of the human proteome is phosphorylated (Vlastaridis *et al.*, 2017) illustrating the extent and complexity of eukaryotic phosphorylation networks. Phosphorylation events control the cell cycle and signalling pathways and are important in the control of cell growth, proliferation, differentiation and transformation (Olsen *et al.*, 2006). Some large (> 100Kbp) DNA virus genomes also encode protein kinases and

phosphatases. Fig. 1 represents all the predicted protein kinases and phosphatases derived from the complete virus genomes currently available at the NCBI proteinclusters database (http://www.ncbi.nlm.nih.gov/proteinclusters). Protein kinases and phosphatases occur mainly in the *Chordopoxvirinae* and *Baculoviridae* virus families (i.e. poxviruses and baculoviruses). Poxviruses are predominantly vertebrate viruses and include significant human pathogens (e.g. the variola virus that causes smallpox). Baculoviruses are ubiquitous insect viruses and are utilized for the biocontrol of specific insect pests. They are divided into two main subfamilies, the *Alphabaculoviridae* and the *Betabaculoviridae*. The prototypical alphabaculovirus *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV) is the best characterized baculovirus and is well known because it is widely used to express recombinant proteins in cultured insect cells. The AcMNPV genome encodes two predicted kinases and three predicted protein phosphatases (Table 1). The genome of the prototypical betabaculovirus *Cydia pomonella* granulovirus (CpGV) encodes one predicted protein kinase PK-1 and three predicted protein phosphatases. Both viruses have representatives of the most common kinase and phosphatase proteinclusters (2744809 and 274817) but the other predicted protein kinases and phosphatases in the viral genomes are not shared between the two.

Structures have been determined for the vaccinia VH1 (Koksal *et al.*, 2009) and variola H1 (Phan *et al.*, 2007) poxvirus protein phosphatases and for the dimeric orfvirus phosphatase OH1 (Segovia *et al.*, 2017) which all belong to proteincluster 2509785. There is a single structure in proteincluster 2509865, that for the AcMNPV NP_054030 BVP phosphatase which can hydrolyze the 5'- phosphate from RNA and free nucleotide di- and tri-phosphates (Changela *et al.*, 2005). No structures have yet been reported for the major class of viral phosphatases in proteincluster 2744817, that comprise over a third of all predicted viral phosphatases.

Here we report the crystal structure of the phosphatase PTP-2 from *Cydia pomonella* granulovirus, the first structure of a member of the protein tyrosine phosphatase-2 proteincluster 294255. A BLAST search with the *C. pomonella* PTP-2 amino acid sequence revealed that most homologues occur in insect baculoviruses and entomopoxviruses (Supplementary Fig. S3). The best functionally characterised viral PTP-2 phosphatase is that of the alphabaculovirus *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV), where expression of the ptp2 gene induces apoptosis and increases yields of infectious occlusion bodies in infected larvae, compared to recombinant viruses without the ptp2 gene (Han *et al.*, 2018).

2. Materials and methods

2.1. Macromolecule production

DNA was extracted from purified CpGV and the ptp2 gene amplified by PCR using the forward primer (5'-CCGGAATTCGATGTACGACGCCACGCGAATCAAC-3') and the reverse primer (5'-CCGCTCGAGCTATTCAACTCTAACCACCAACTTATTATC-3'). The PCR product was digested using EcoRI and XhoI, and ligated into a pETDuet plasmid including a sequence encoding a His_6 tag directly upstream from the gene sequence. The recombinant plasmid was transformed into *Escherichia coli* BL21 (DE3) competent cells. The transformed cells were cultured at 291 K in 2 L Erlenmeyer flasks with 100 µg mL^{-1} ampicilin at 310 K and grown to an $OD_{600\ nm}$ of 0.6-0.8. Gene expression was

induced with isopropyl β-D-1-thiogalactopyranoside at a final concentration of 0.25 mM and the cells grown a further 18 hours.

The cells were harvested by centrifugation and the pellet resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole). Lysozyme was added to a final concentration of 1 mg mL^{-1} and the resuspended cells incubated at 277 K for 15 min. The resuspended cells were then sonicated at 277 K and the resulting homogenate centrifuged at 17,000 RCF at 277 K for 30 min. The supernatant was then applied to a 5 mL Ni-IMAC column equilibrated with lysis buffer. The column was washed with twenty column volumes of lysis buffer before the His_6 tagged PTP-2 protein was eluted using three column volumes of elution buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 250 mM imidazole). SDS-PAGE analysis of the eluted fractions revealed a single band. Pooled fractions containing this band were concentrated and applied to a Superdex S200 gel-filtration column that had been previously equilibrated in 20 mM Tris-HCl pH 8.0, 200 mM NaCl. Gel filtration was monitored by measuring absorbance at 280 nm. The resultant absorption profile showed a single peak and SDS-PAGE analysis of this peak revealed a single band similar to that found in the samples eluted from the Ni-IMAC column. Fractions containing this protein band were pooled and concentrated to
22 mg mL^{-1} for crystallization. Selenomethionine protein was purified from bacteria grown in SeMet medium(Park *et al.*, 2017) using similar methods, and concentrated to 40 mg mL^{-1} for crystallization.

2.2. Crystallization, data collection, and structure refinement

Initial crystallization conditions were found using the sitting-drop vapour-diffusion method and solutions from a commercial crystallisation screening kit (Molecular Dimensions). Refined drops were prepared by mixing 1 µL protein solution and 1 µL reservoir solution and were equilibrated against 80 µL reservoir solution. Crystals grew within 24 hours at 291K using the reservoir solution HEPES-NaOH pH 7.5, 200mM $CaCl_2$, and 14% - 20% PEG 6000. Selenomethionine PTP-2 was crystallized using the same conditions as the native protein. Crystals were briefly soaked in a cryo-protectant solution made by substituting water in the reservoir solution with 15 % glycerol prior to flash-cooling in liquid nitrogen. Diffraction data were measured at the Australian Synchrotron using beam line MX1 equipped with an ADSC QUANTUM 315r CCD detector. The MX1 nominal beam size at the crystal was 130 x 90 µm and the flux 1.5 x 10^{11} photons/second (http://archive.synchrotron.org.au/34-australian-synchrotron/mx/254-technical-information-pc). The crystal used for the native dataset had approximate dimensions 100 x 100 x 20 µm. The data was processed using IMOSFLM (Battye *et al.*, 2011) and AIMLESS (Evans & Murshudov, 2013). The structure was solved by the SAD method using AutoSol (Adams *et al.*, 2010) which led to an initial model for the protein. This model was then used for molecular replacement with PHASER to solve the structure of the native protein structure. The model was then improved with iterative rounds of manual building and refining with Coot (Cowtan, 2010) and subsequently refined with PHENIX (Adams *et al.*, 2010). Data and refinement statistics are listed in Table 2. Atomic coordinates and structure factors have been deposited into the Protein Data Bank (www.pdb.org) with the PDB accession code 6I28.

3. Results and discussion

The electron density revealed a single molecule in the asymmetric unit and the complete polypeptide was clearly interpretable allowing an unambiguous model to be fitted for all 161 amino acids (Fig. 2). The refinement statistics and model parameters indicate that the model is well defined. No significant potential oligomerisation interfaces were found using the PISA server (www.ccp4.ac.uk/pisa).

A search for structural homologues using PDBefold (www.ebi.ac.uk/msd-srv/ssm) revealed that the N-terminal ~130 amino acids of PTP-2 form a CC1 protein phosphatase fold, the most commonly occurring phosphatase fold (Chen *et al.*, 2017). This fold corresponds to the DSPc (dual specificity protein phosphatase) architecture in the SMART catalogue (http://smart.embl.de/smart/do_annotation.pl?DOMAIN=SM00195), and is also evident from the results of a NCBI BLAST sequence homology search (Fig S3). The DSPc family is widely represented in major phyla and currently includes 125 structures and 5580 sequences.

The closest DSPc structure to PTP-2 is a human dual specificity phosphatase DUSP13b (PDB ID 2PQ5) that matches with an RMS deviation of 1.4 Å over 135 Cα positions with a sequence identity of 24%. The structures of viral protein phosphatases in NCBI proteinclusters 2509785 and 2509865 (Fig. 1) fit to PTP-2 with an RMS deviation of Cα positions of ~2.0 Å over ~130 residues when compared to the PTP-2 structure using GESAMT (Krissinel, 2012). The corresponding sequence identities are 15% and 17%.

The DsPc fold has a core structure consisting of a five stranded parallel β-sheet with two α-helices on one side and five α-helices on the other. In addition to the common fold DSPc structures share a conserved active site architecture and catalytic mechanism (Andersen *et al.*, 2001). The cysteine in the active site motif CX5R 'P-loop' (C97 in PTP-2) is located at the base of a positively charged pocket at the N-terminal end of helix H5 and forms a phosphor-cysteine intermediate with the substrate phosphate. This intermediate is resolved by the neighbouring aspartate in the 'D-loop' (D64 in PTP-2).

The fold of PTP-2 is shown in Fig. 2 together with side-chains of the five P-loop residues and the D-loop aspartate that have atoms within 5Å of the active site. The active site residues are also shown in the alignment of NCBI proteincluster 294255 homologues (Fig. 3) and the wider alignment of BLAST hits to PTP-2 shown in Supplementary Fig. S3. The regions of sequence homology extend well beyond the active site. PTP-2 shares the larger HCxxGxxR motif and other similarities to dual specificity protein phosphatases and the active site pocket topography is similar to that of DUSP14 (Jeong *et al.*, 2014). The closest full-length BLAST matches to PTP-2 are from alphabaculoviruses and entomopoxviruses, but not from other betabaculoviruses. The shorter matches shown in Supplementary Fig. S3 are mainly to fragments of insect proteins and these matches cover only the C-terminal part of the core of the molecule extending from strand S5 to H6 and including the active site P-loop but excluding the D-loop. The *Cydia pomonella* granulovirus genome encodes one of these shorter homologues, which we call here PTP-3. The sequence alignment of PTP-2 and PTP-3 is shown in Supplementary Fig. S4. PTP-3 is 60% identical to PTP-2 suggesting it is functional despite the lack of a D-loop, having presumably been maintained in the genome for sufficient time to

diverge from PTP-2. Interestingly, similar truncated versions of PTP-2 are not found in other baculoviruses.

PTP-2 and its viral homologues and PTP-3 all have C-terminal extensions beyond the conserved DSPc core structure (Supplementary Fig. S3). In PTP-2 this extension forms an anti-parallel beta sheet containing the two strands S6-S7 which extends from the molecule in a different configuration to the C-terminal extensions previously described for DUSP14 and DUSP18 (Jeong *et al.*, 2014). The proteincluster 294255 sequence alignment (Fig. 3) shows that other alphabaculovirus PTP-2 homologues have a either a similar C-terminal extension or an N-terminal extension, but not both. The only other betabaculovirus PTP-2 (*Crytophlebia leucotreta* granulovirus) in the proteincluster sequence alignment does not have extensions at either end of the molecule.

Given the complexity of phosphorylation networks in cells, it is likely that PTP-2 has a number of phosphorylated protein and non-protein substrates. It may also function in cells as an oligomer or in a complex with other proteins. Homologous viral phosphatases have already been described that are dimeric (e.g. OH1 (Segovia *et al.*, 2017)) or that have non-protein substrates (Changela *et al.*, 2005) (Segovia *et al.*, 2017). In this context we note that while there is no evidence that PTP-2 forms a dimer in solution, when the two PTP-2 models are separately superimposed onto the two subunits of the Orf virus OH1 dimer, the PTP-2 C-terminal beta sheet extensions are then in a position to form an extended beta-sheet dimer interface in the same region as the OH1 dimer interface (Supplementary Fig. S2). Clearly much remains to be learned about the role of PTP-2 in virally infected cells. Future proteomics investigations can be expected to identify substrates and interacting partners. This structure, together with biochemical characterization of its activity on defined substrates will lead to a much better understanding of the role of phosphorylation networks in granulovirus infected cells.

Acknowledgements We thank Johannes Jehle for providing purified granulovirus. Funding for the project was provided by the New Zealand Royal Society Marsden grant UOA221. X-ray data collection was carried out at the MX1 beamline at the Australian Synchrotron, Victoria, Australia, with support from the New Zealand Synchrotron Group.

Author Contributions

GH cloned, expressed, purified and crystallized proteins

JK collected carried out the X-ray crystallographic structure determination

DG helped with the structure determination

MO and GH planned and designed the research

PM and GH analysed data and wrote the paper

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.

Andersen, J. N., Mortensen, O. H., Peters, G. H., Drake, P. G., Iversen, L. F., Olsen, O. H., Jansen, P. G., Andersen, H. S., Tonks, N. K. & Møller, N. P. H. (2001). *Mol. Cell Biol*. **21**, 7117–7136.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281.

Changela, A., Martins, A., Shuman, S. & Mondragón, A. (2005). *J. Biol. Chem.* **280**, 17848–17856.

Chen, M. J., Dixon, J. E. & Manning, G. (2017). *Sci. Signal.* **10**, eaag1796.

Cowtan, K. (2010). *Acta Crystallogr. D Biol. Crystallogr*. **66**, 470–478.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr. D Biol. Crystallogr*. **69**, 1204–1214.

Han, Y., van Houte, S., van Oers, M., Ros, V., Han, Y., van Houte, S., van Oers, M. M. & Ros, V. I. D. (2018). *Viruses*. **10**, 181.

Jeong, D. G., Wei, C. H., Ku, B., Jeon, T. J., Chien, P. N., Kim, J. K., Park, S. Y., Hwang, H. S., Ryu, S. Y., Park, H., Kim, D.-S., Kim, S. J. & Ryu, S. E. (2014). *Acta Crystallogr. D Biol. Crystallogr.* **70**, 421–435.

Koksal, A. C., Nardozzi, J. D. & Cingolani, G. (2009). *J. Biol. Chem.* **284**, 10129–10137.

Krissinel, E. (2012). *J. Mol. Biochem*. **1**, 76–85.

Notredame, C., Higgins, D. G. & Heringa, J. (2000). *J. Mol. Biol.* **302**, 205–217.

Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P. & Mann, M. (2006). *Cell*. **127**, 635–648.

Park, J., Kim, H., Kim, S., Lee, D. & Shin, D. H. (2017). *Acta Crystallogr. F Struct. Biol. Commun*. **73**, 90–94.

Phan, J., Tropea, J. E. & Waugh, D. S. (2007). *Acta Crystallogr. D. Biol. Crystallogr.* **63**, 698–704.

Schrödinger, LLC (2015). The PyMOL Molecular Graphics System, Version 2.1.1.

Segovia, D., Haouz, A., Porley, D., Olivero, N., Martínez, M., Mariadassou, M., Berois, M., André-Leroux, G. & Villarino, A. (2017). *J. Mol. Biol*. **429**, 2816–2824.

Vlastaridis, P., Kyriakidou, P., Chaliotis, A., Van de Peer, Y., Oliver, S. G. & Amoutzias, G. D. (2017). *Gigascience*. **6**,.

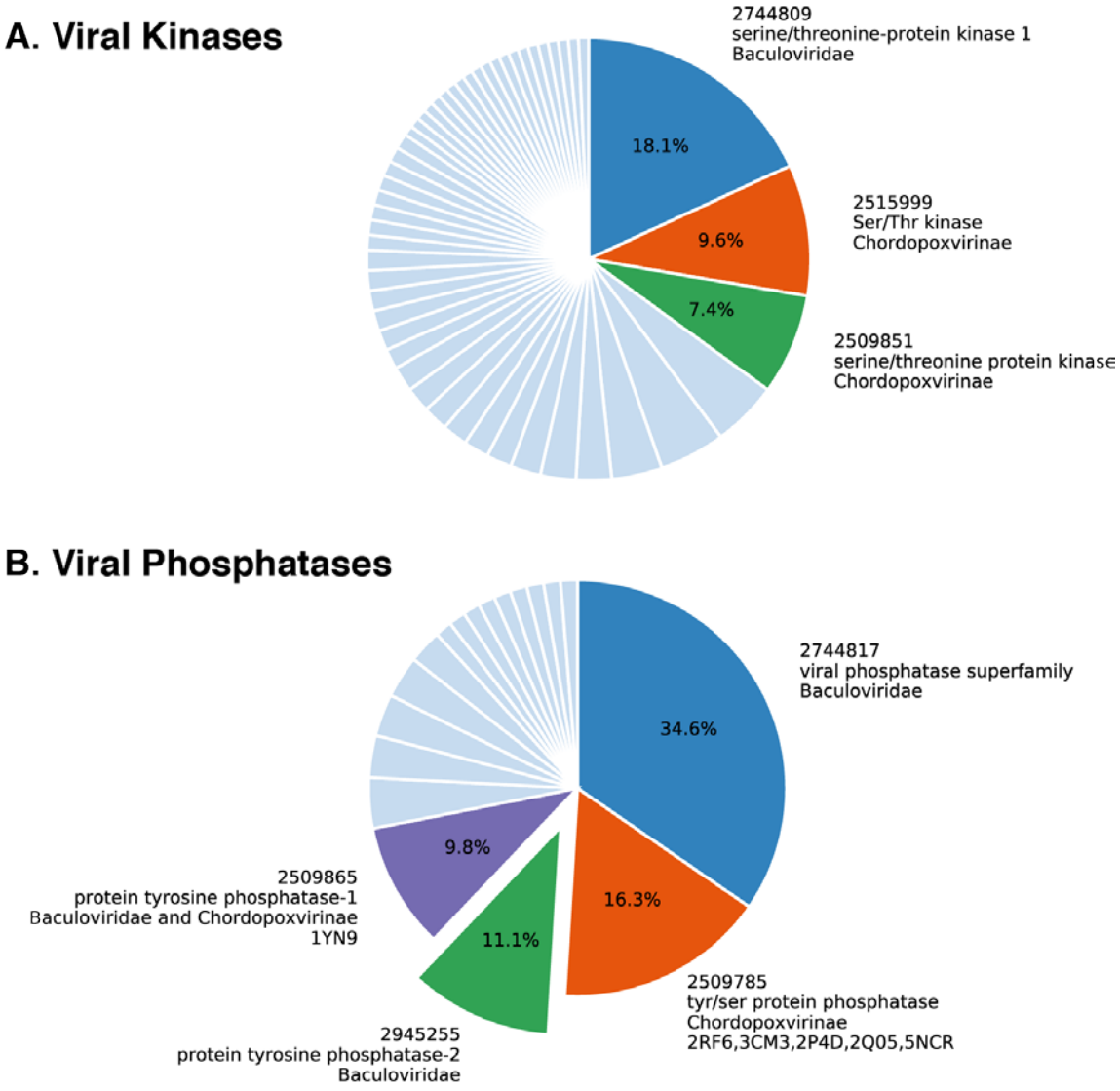

Figure 1 Predicted protein kinases(A) and phosphatases(B) derived from virus genomes. Homology groups were obtained from the NCBI proteinclusters website (www.ncbi.nlm.nih/proteinclusters) using the search string "viruses"[organism] together with either of the terms "kinase" or "phosphatase". Kinases and phosphatases with non-protein targets were omitted. Each segment of the pie charts is labelled with the NCBI proteincluster number and name together with the included virus families and PDB codes for atomic structures. Clusters representing less than 5% of the total are indicated by the light blue unlabelled segments. *Cydia pomonella* PTP-2 phosphatase belongs to the protein tyrosine phosphatase-2 proteincluster shown by the expanded green segment in B.

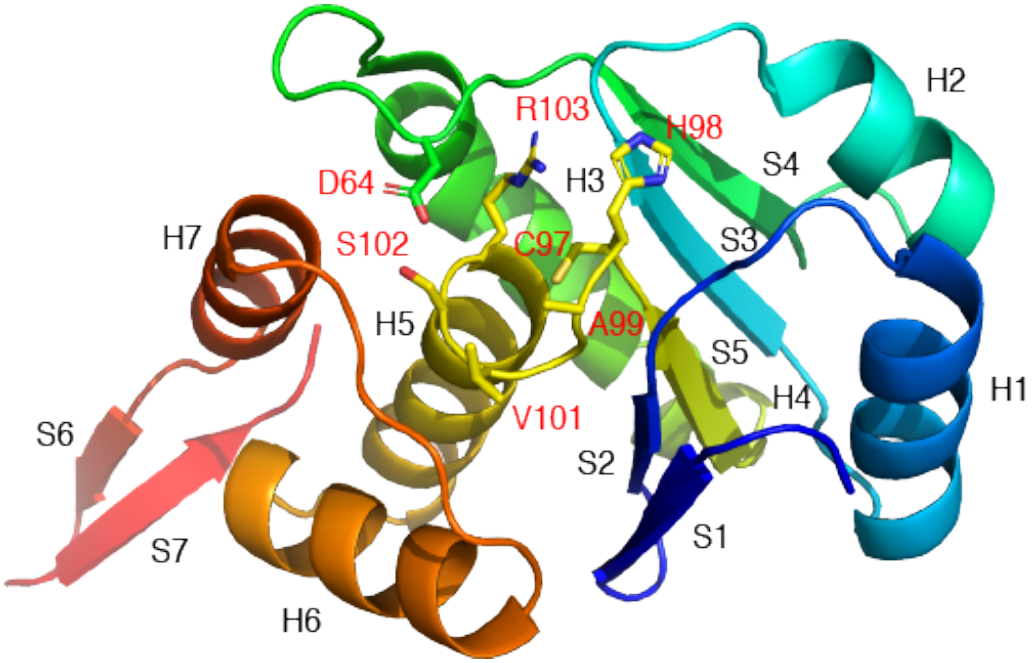

Figure 2 Ribbon diagram of PTP-2 showing secondary structure elements (black labels) and conserved active site residues (red labels). Figure prepared using PyMol(Schrödinger, LLC, 2015)

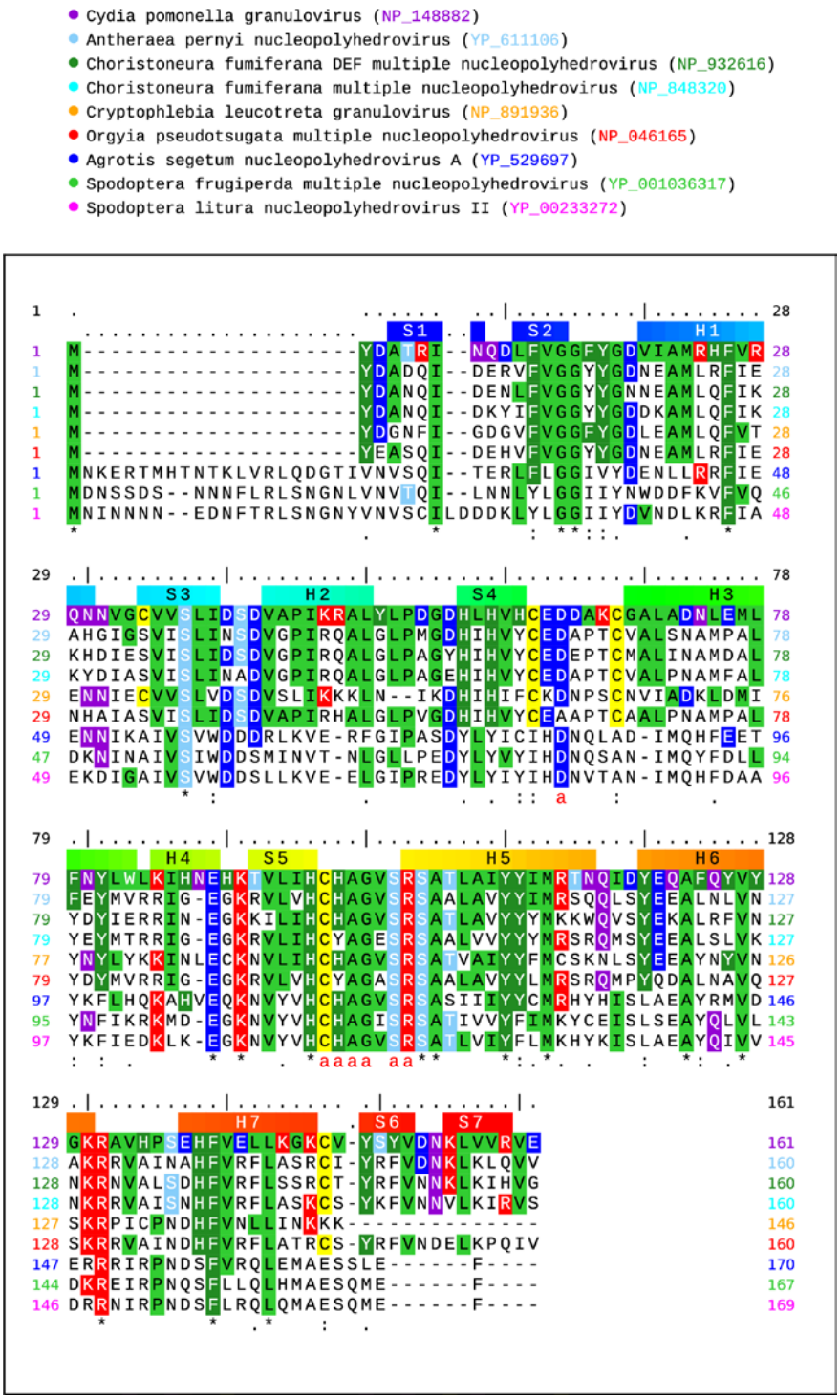

Figure 3 PTP-2 multiple sequence alignment made using the PTP-2 atomic structure. The header lists species names and NCBI codes for representative sequences selected from the NCBI proteincluster 294255 (the whole cluster is shown in Supplementary Fig. S1). Coloured dots correspond to the amino acid numbers in the alignment below. The top row above the alignment is the index for *C. pomonella* PTP-2. The next row shows secondary structure elements from Fig. 2. The row below the alignment indicates homology. Active site residues shown in Fig. 2 are marked in red

with the letter ‘a’. The structure based alignment was prepared using T-Coffee (Notredame *et al.*, 2000).

Table 1 Predicted protein kinases and phosphatases in the genomes of AcMNPV and CpGV

Virus	Predicted function	Protein name	NCBI protein ID	NCBI proteincluster ID
AcMNPV	protein kinase	PK-1	NP_054039	2744809[1]
		PK2	NP_054153	2509980
	protein phosphatase	PTP	NP_054030	2509865
		38K	NP_054128	2744817[2]
			NP_054062	2509896
CpGV	protein kinase	PK-1	NP_148787	2744809[1]
	protein phosphatase	ORF88	NP_148872	2744817[2]
		PTP-2	NP_148882	2945255
		PTP-3*	NP_148850	

Data from https://www.ncbi.nlm.nih.gov/proteinclusters. Superscripts indicate homologues in proteinclusters which occur in both AcMNPV and CpGV.

* Renamed here as PTP-3 because the NCBI names for NP_148882 and NP_148850 are both ‘PTP-2’

Table 2 Data collection and processing

	SeMet_Ptp_2	Native_Ptp_2
Wavelength (Å)	0.9537	0.9537
Resolution range (Å)	54.7 - 2.20 (2.27 -2.20)	33.7 - 1.65 (1.68 - 1.65)
Space group	I 2 2 2	I 2 2 2
Unit cell (Å)	64.4 65.97 97.56 90 90 90	64.7 66.7 98.0 90 90 90
Total reflections	154,401 (13,430)	365,717 (17,715)
Unique reflections	10,898 (937)	25886 (1250)
Multiplicity	14.2 (14.3)	14.1 (14.2)
Completeness (%)	100 (100)	100 (100)
Mean I/sigma(I)	10.4 (1.8)	11.2 (0.7)
Wilson B-factor	22.1	20.7
R-merge	0.25 (1.7)	0.17 (3.9)
R-meas	0.27 (1.84)	0.18 (4.1)
R-pim	0.01 (0.67)	0.05 (1.1)
CC1/2	0.99 (0.68)	0.90 (0.40)
Refinement		
Reflections used in refinement	-	25883 (2526)
Reflections used for R-free		1307 (142)
R-work		0.19 (0.37)
R-free		0.22 (0.40)
CC(work)		0.94 (0.24)
CC(free)		0.87 (0.13)
Number of non-hydrogen atoms		1445
protein chain A		1328
glycerol (1)		6
Ca (1)		1
water		110
Protein residues		161
RMS(bonds)		0.016

RMS(angles)		1.29
Ramachandran favored (%)		96.9
Ramachandran allowed (%)		3.1
Ramachandran outliers (%)		0
Rotamer outliers (%)		2.1
Clashscore		3.4
Average B-factor		34.6
protein chain A		33.7
glycerol (1)		63.4
Ca (1)		44.5
water		43.5
PDB ID	-	6I28

Supporting information

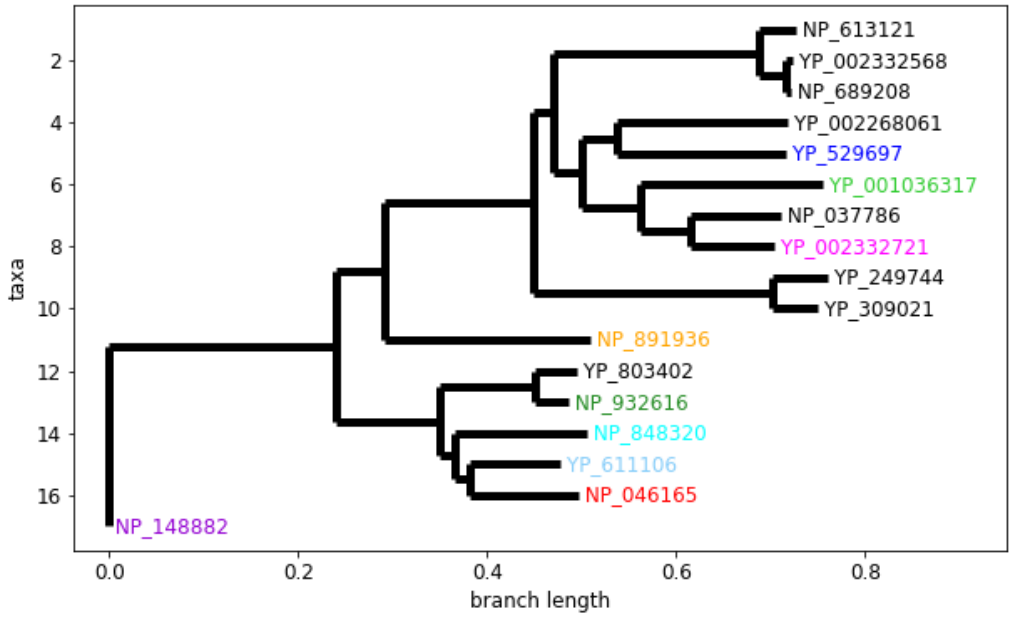

Figure S1 Rooted phylogenetic tree for the PTP-2 NCBI proteincluster 2945255. The root sequence on the left side of the figure is *Cydia pomonella* PTP-2 (NP_148882). Coloured sequences were included in the multiple sequence alignment shown in Fig. 3.

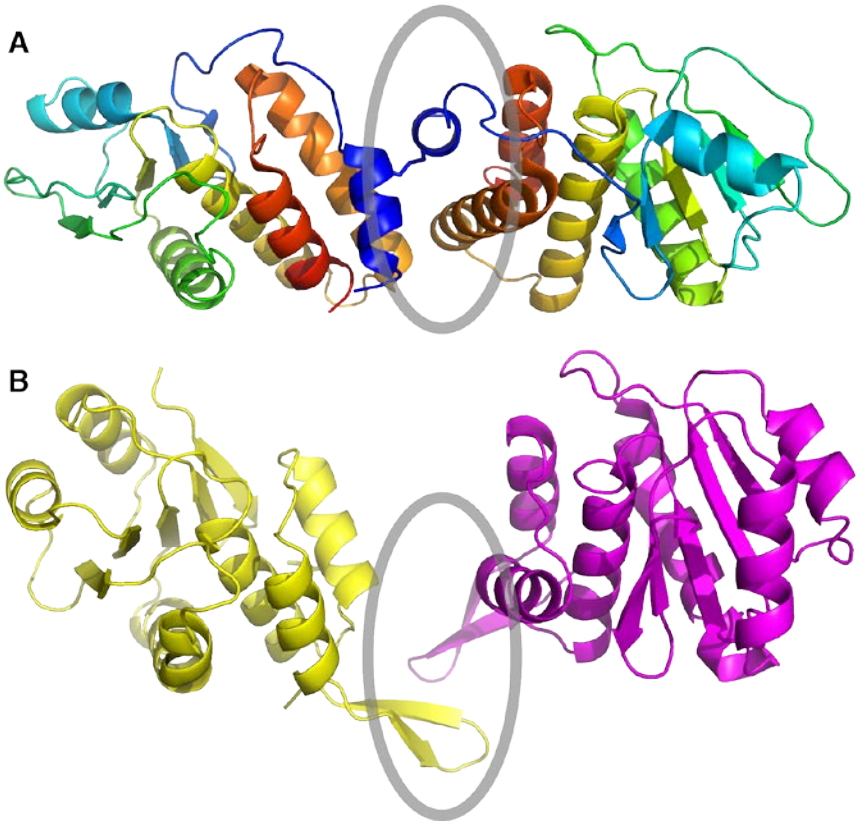

Figure S2 A. Cartoon representation of the dimeric Orf Virus OH1 phosphatase(PDB 5NCR) (Segovia *et al.*, 2017) showing monomers A and B which are linked by a disulphide bond joining the N-terminal α-helices, shown in blue. The circle indicates the dimeric interface.
B. PTP-2 molecules separately aligned onto the OH1 A and B monomers. The N-terminal α-helices are absent in PTP-2. The C-terminal β-strands of PTP-2 are located in the same region as the OH1 AB interface (circled).

Figure S3 Results summary from a *Cydia pomonella* PTP-2 protein sequence BLAST search against the NCBI refseq_protein database. Horizontal coloured lines represent the 20 best matching pairwise sequence alignments listed in order below the target *C. pomonella* PTP-2, shown by the top red line. Except for *C. pomonella* PTP-3 the magenta lines all represent alphabaculovirus PTP-2 homologues. The shorter green lines are protein fragments from a variety of insects. The two longer green lines are full-length entomopox insect viral PTP-2 homologues. The vertical red lines represent the active site residues shown in Figures 2 and 3. The green box shows the region of PTP-2 that corresponds to the DSPc superfamily fold-type identified by the NCBI routines. All matches have E values < 1e-6.

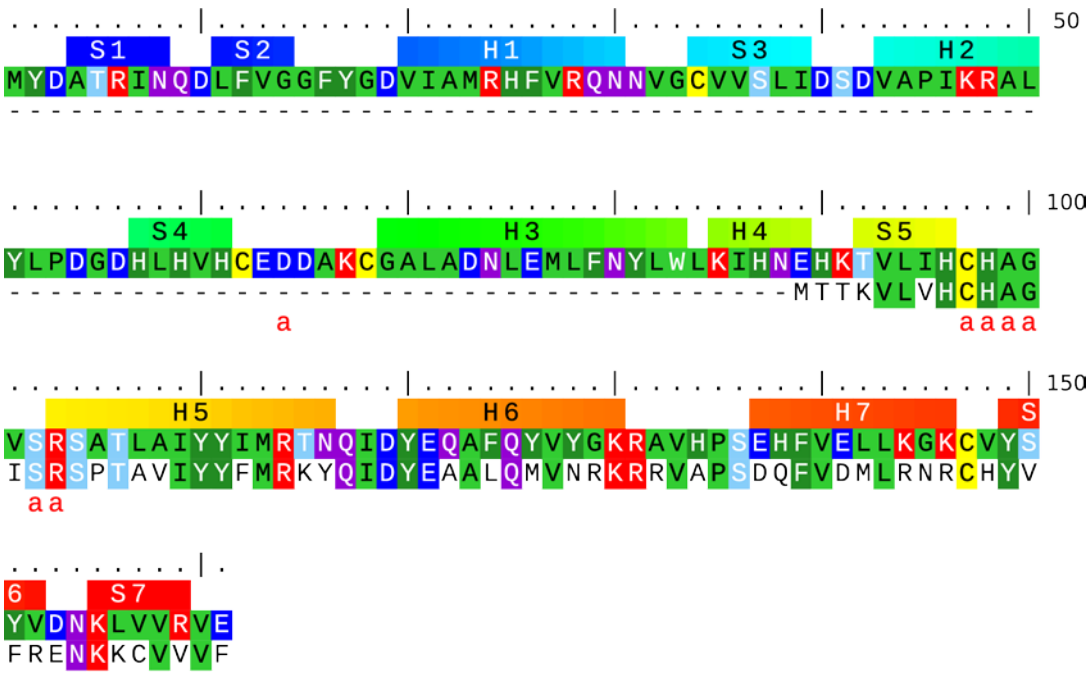

Figure S4 Sequence alignment of *C. pomonella* PTP-2 and PTP-3. The figure shows the two amino acid sequences together with an index line for PTP-2 and secondary structure elements corresponding to the Fig. 2 ribbon diagram.