

Solution and Crystallographic Structures of the Central Region of the Phosphoprotein from Human Metapneumovirus

Cedric Leyrat^{1*}, Max Renner¹, Karl Harlos¹, Jonathan M. Grimes^{1,2}

¹ Division of Structural Biology, University of Oxford, Oxford, United Kingdom, ² Science Division, Diamond Light Source Ltd., Didcot, United Kingdom

Abstract

Human metapneumovirus (HMPV) of the family *Paramyxoviridae* is a major cause of respiratory illness worldwide. Phosphoproteins (P) from *Paramyxoviridae* are essential co-factors of the viral RNA polymerase that form tetramers and possess long intrinsically disordered regions (IDRs). We located the central region of HMPV P (P_{ced}) which is involved in tetramerization using disorder analysis and modeled its 3D structure *ab initio* using Rosetta fold-and-dock. We characterized the solution-structure of P_{ced} using small angle X-ray scattering (SAXS) and carried out direct fitting to the scattering data to filter out incorrect models. Molecular dynamics simulations (MDS) and ensemble optimization were employed to select correct models and capture the dynamic character of P_{ced}. Our analysis revealed that oligomerization involves a compact central core located between residues 169-194 (P_{core}), that is surrounded by flexible regions with α -helical propensity. We crystallized this fragment and solved its structure at 3.1 Å resolution by molecular replacement, using the folded core from our SAXS-validated *ab initio* model. The RMSD between modeled and experimental tetramers is as low as 0.9 Å, demonstrating the accuracy of the approach. A comparison of the structure of HMPV P to existing *mononegavirales* P_{ced} structures suggests that P_{ced} evolved under weak selective pressure. Finally, we discuss the advantages of using SAXS in combination with *ab initio* modeling and MDS to solve the structure of small, homo-oligomeric protein complexes.

Citation: Leyrat C, Renner M, Harlos K, Grimes JM (2013) Solution and Crystallographic Structures of the Central Region of the Phosphoprotein from Human Metapneumovirus. PLoS ONE 8(11): e80371. doi:10.1371/journal.pone.0080371

Editor: Jamil Saad, University of Alabama at Birmingham, United States of America

Received: July 23, 2013; **Accepted:** October 9, 2013; **Published:** November 4, 2013

Copyright: © 2013 Leyrat et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under SILVER grant agreement n° 260644. Max Renner was supported by a Wellcome trust fellowship (099667/Z/12/Z). This work was also supported by the Wellcome Trust Core award (090532/Z/09/Z). Administrative support was provided by the Wellcome Trust (Grant 075491/Z/04). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have read the journal's policy and have the following conflicts: Dr. Jonathan M. Grimes is employed by Diamond Light Source Ltd. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: cedric@strubi.ox.ac.uk

Introduction

Human metapneumovirus (HMPV) is a major cause of acute respiratory diseases in children, the elderly and immunocompromised patients worldwide [1–5]. HMPV belongs to the *Pneumovirinae* subfamily of the *Paramyxoviridae* and is further classified into the genus *Metapneumovirus*[6]. HMPV is an enveloped virus that forms pleomorphic or filamentous virions. Its genome consists of a ~13-kb single stranded RNA molecule of negative polarity that encodes 9 proteins in the order ^{3'}-N-P-M-F-M2(-1)/(-2)-SH-G-L-5'. HMPV proteins show detectable levels of sequence identity to the respiratory syncytial virus (RSV) (genus *Pneumovirus*); however, the order of the genes is different and HMPV lacks the NS1 and NS2 genes present in RSV. For all paramyxoviruses, the nucleoprotein (N) encapsidates viral RNA, leading to a N-RNA complex which, together with the RNA-dependent RNA

polymerase (L) and the phosphoprotein (P), forms the viral replication complex. The P protein is thought to be responsible for the recruitment of the large polymerase L onto the viral N-RNA template through direct interactions with N and L [7–15]. In addition, P chaperones the nascent N, which is sequestered in the form of an RNA-free NP complex [16,17]. The M2 gene is specific to the *Pneumovirinae* subfamily, and possesses two overlapping open reading frames encoding two proteins, the antitermination/transcription-elongation factor M2-1, which is required for viral transcription [18], and the RNA synthesis regulatory factor M2-2 [19].

For all members of the *Paramyxoviridae* family, the P protein is an intrinsically disordered polypeptide which forms tetramers through a central α -helical coiled-coil region. Available structures of the tetrameric coiled-coil from Sendai virus (SeV) [20] and Measles virus (MeV)[21] show long parallel arrangements of twisted α -helices. However, the structure of

the Mumps virus phosphoprotein strikingly reveals the formation of parallel dimers that further assemble into tetramers by associating in an antiparallel fashion [22]. In contrast, the tetramerization domain of the RSV P protein, which is the closest homologue of HMPV to have been structurally characterized, displays a much shorter coiled-coil region, termed fragment Y*. This fragment has been previously identified via proteolytic digestion and consecutively mapped to residues 119 to 160 by mass-spectrometry and N-terminal sequencing [23,24]. Interestingly, although the length of the HMPV P sequence is greater than that of RSV P by 53 residues and the overall sequence identity is only 28%, conservation is considerably higher in the central structured region of the protein, suggesting similar tetramerization domains.

In this study, we applied bioinformatics approaches to locate the central folded region of HMPV P, and used symmetric homo-oligomeric *ab initio* modeling in combination with small angle X-ray scattering (SAXS) and molecular dynamics simulations (MDS) to determine the structure of the central region of HMPV P (P_{ced}) and capture its flexibility in solution. We used the obtained model to solve the crystal structure of the core region of P_{ced} (residues 168-194) by molecular replacement. We analyze the implications of the structure of P_{ced} for virus function and evolution, and discuss the usefulness of integrative approaches to protein structure determination.

Results

Disorder analysis locates the central structured region of P_{ced}

We used meta-disorder predictions in combination with sequence conservation and secondary structure propensity to locate IDRs and folded regions of HMPV P (Figure 1). The analysis predicts the presence of a central, highly conserved region with α -helical propensity located between residues 158 to 237, which we refer to as P_{ced} . The N-terminal and C-terminal regions flanking P_{ced} are mostly disordered and weakly conserved, with the notable exception of the first 30 residues, which show a narrow peak of conservation and predicted order, suggesting the presence of an α -helical molecular recognition element (MoRE), as has been described for P proteins from other members of the *Paramyxoviridae* [16] and *Rhabdoviridae* families [25].

Structural characterization of P_{ced} by SAXS

P_{ced} was expressed and purified in *E.coli* and its structure was characterized using SAXS (Figure 2 A). The samples were free from aggregates, as evidenced by the linearity of the Guinier region (Figure 2 B). The parameters derived from SAXS data are summarized in Table 1. Radii of gyration (R_g) were independent of protein concentration, and only moderately affected by salt concentration ($R_g = 3.26 \pm 0.02$ nm in 150 mM NaCl vs 3.17 ± 0.04 nm in 800 mM NaCl). However, a significant drop in R_g to a value of 2.98 ± 0.03 nm was observed in the presence of 1M of non-detergent sulfobetaine 201 (NDSB-201), suggesting an induced stabilization of P_{ced} structure. Molecular weights (MW) were estimated based on

calculation of the concentration-independent volume of correlation V_c , as defined in [26], yielding values between 28 and 34 kDa, in agreement with the MW calculated from the amino-acid sequence, assuming a tetramer ($8.8 \times 4 = 35$ kDa).

Modelling of P_{ced} using Rosetta Fold-and-dock and SAXS-based model selection

We employed the Rosetta fold-and-dock application [27] to model the structure of P_{ced} tetramers. We generated $2 \times 30,000$ models using the sequence of residues 155-241 or 156-237. The use of two different sequence lengths leads to an increased structural diversity of sampled models, due to the fragment-based approach implemented in Rosetta. Moreover, the effects of possibly truncating the predicted α -helix ($\alpha 3$) located between residues 207-241 are taken into account (Figure 1). Both ensembles were ranked according to their Rosetta free energy score, and the five best models of each ensemble are shown in Figure 3 A and B. Interestingly, all models apart from a single oblate model, form a tetrameric coiled-coil through the arrangement of α -helices ($\alpha 2$) typically ranging from residues 168 to 198. We refer to this central region as P_{core} . The regions comprising residues 158 to 168, and 200 to 237 display a tendency towards α -helical conformations, but adopt different orientations in each model, resulting in important changes to the overall shape of the predicted models. This lack of convergence suggests that these regions might not assume a single conformation in solution, but rather exist as disordered ensembles.

In a second step, we calculated theoretical SAXS profiles for all of the Rosetta ensembles and fitted them to the experimental SAXS profile, yielding an additional score for each model in the form of a χ_{exp} value, which measures the discrepancy between theoretical and experimental SAXS profiles. The best results were obtained using data measured in the presence of 1M NDSB-201, because of the increase in conformational stability it induces, as evidenced by the significant drop in R_g (Table 1). We filtered out all models displaying χ_{exp} values higher than 1.3, thus eliminating more than 90 % of the models and then ranked the remaining models according to their Rosetta score. The five best models from each ensemble are shown in Figure 3 C & D. Strikingly, we observe that the experimental SAXS profile imposes strong shape constraints on the models, leading to a more homogeneous ensemble (Figure 3C&D opposed to Figure 3A&B). With the exception of an oblate model which ranked third (not shown) and model 4, all models display a relatively similar coiled-coil arrangement of α -helices encompassing residues 168 to 198 (P_{core}), while residues 158-167 and residues 199-237 adopt various, mostly α -helical structures.

MDS confirm the stability of P_{core} and the flexibility of the flanking regions

We tested the stability of the 10 best-fitting *ab initio* models via classical explicit-solvent molecular dynamics simulations (MDS), by performing duplicate runs of approximately 200 ns. The analysis of root mean square fluctuations (RMSF) along the sequence confirms the stability of the coiled coil region, with RMSF values centred around 2 Å for the P_{core} region. The

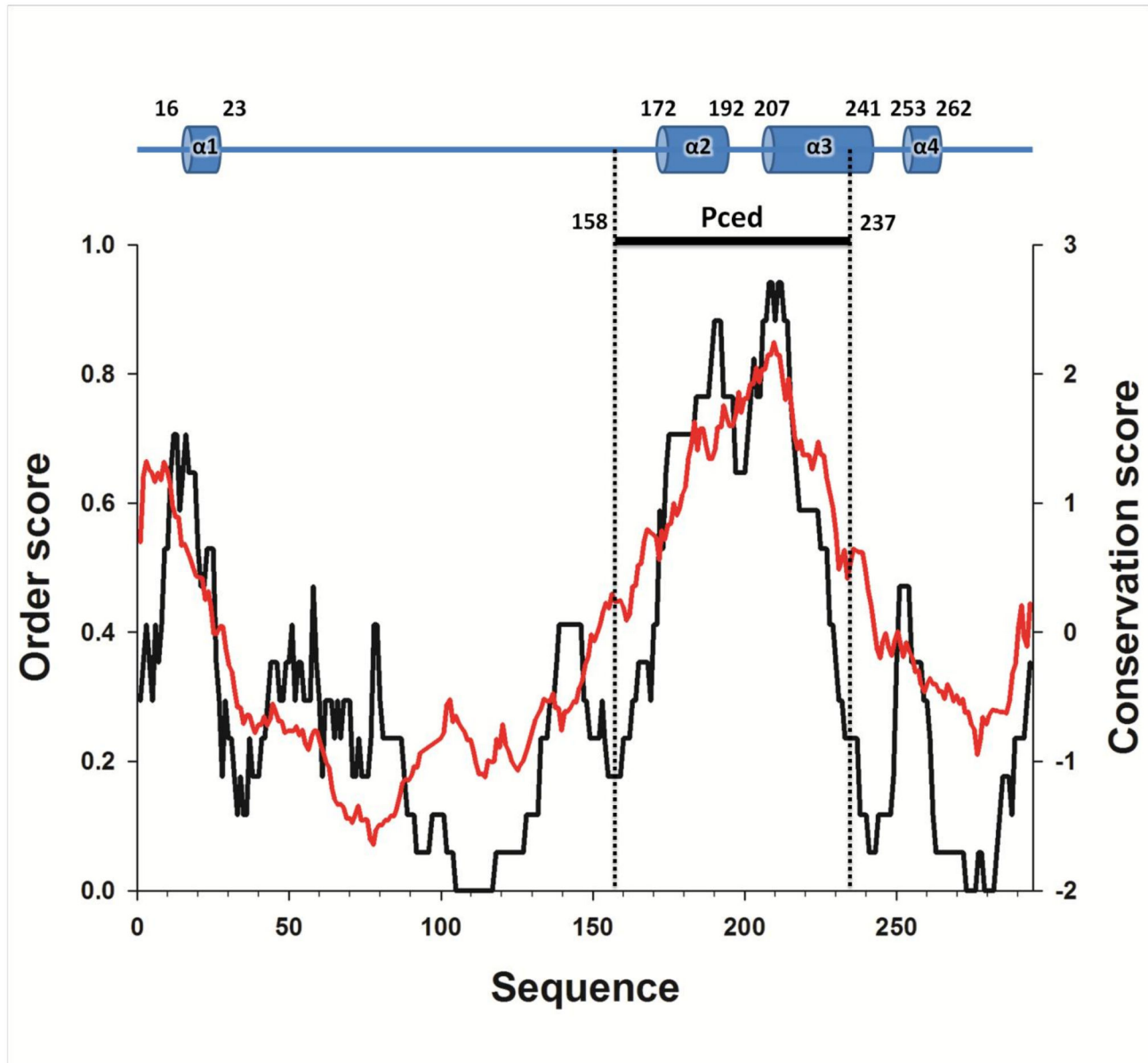


Figure 1. Sequence-based analyses of human metapneumovirus phosphoprotein. The predicted propensity to adopt ordered structures is represented along the amino-acid sequence (black line), together with the conservation score (red line), calculated using AL2CO [76]. The location of the predicted secondary structure elements and the identity of the cloned construct are shown above the graphs.

doi: 10.1371/journal.pone.0080371.g001

flanking regions (residues 158-168 and 200-237) display higher RMSFs, indicating instability of the packing in most models (Figure 4 F), and further supporting the hypothesis that these residues are flexible in solution. Additionally, model 4, which displays a different P_{core} structure, is readily identified as an outlier due to its higher flexibility in the P_{core} region. The RMSD of P_{core} was calculated with respect to the starting structure, showing that model 1L displays the most stable P_{core} structure (Figure 4 A). P_{core} from model 1L was then used as a reference

structure to study the conformational behavior of the other models during MDS (Figure 4 B, C, D and E). Models 1 and 2L fluctuated within about 1.0-1.4 Å RMSD of model 1L P_{core} , showing that they adopt a similar structure (Figure 4 B&C). Interestingly, model 2 was also seen to converge within 1.0 Å RMSD of model 1L P_{core} after 50 ns of MDS, following an initial drop of 0.8 Å (Figure 4 D). Only model 5 P_{core} diverged significantly by more than 2.5 Å in one simulation (Figure 4 E).

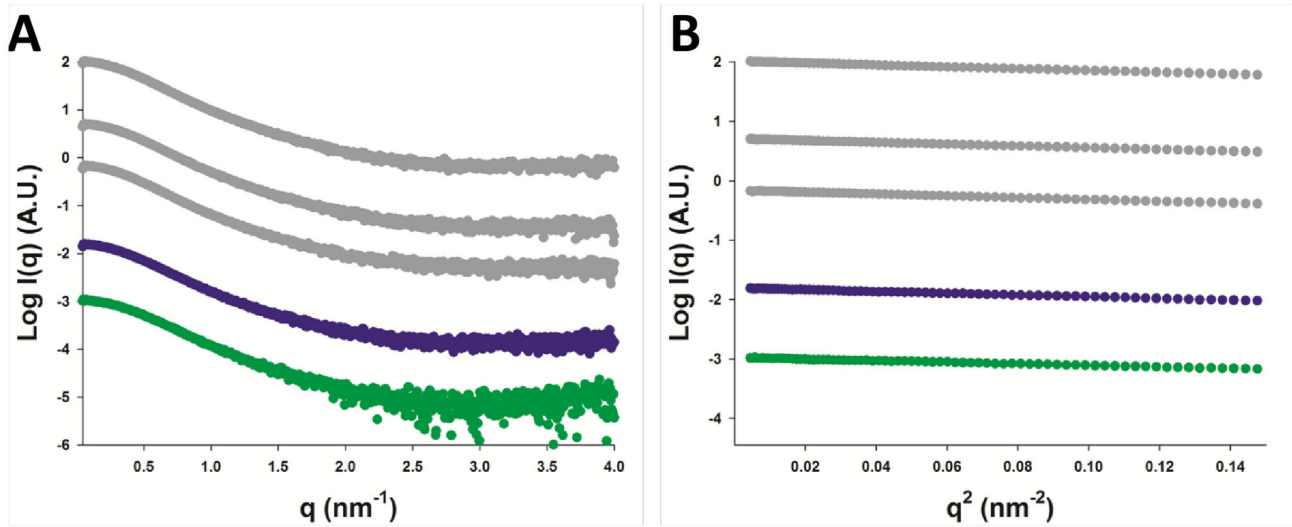


Figure 2. Small Angle X-ray Scattering experiments (SAXS). A. SAXS profiles of P_{ced} measured in the presence of 20 mM Tris pH 7.5, 150 mM NaCl at three different protein concentrations (4, 5 or 6 mg/ml) (grey spheres), in 20 mM Tris pH 7.5, 800 mM NaCl (purple spheres), or 20 mM Tris pH 7.5, 150 mM NaCl and 1M NDSB-201 (green spheres). B. Corresponding Guinier plots, showing linear behavior in the low q range.

doi: 10.1371/journal.pone.0080371.g002

Table 1. SAXS-derived parameters.

Buffer conditions	Concentration (mg/ml)	Molecular weight (kDa)	Radius of gyration (nm)
20 mM Tris pH 7.5 150 mM NaCl	4	33	3.23 ± 0.05
20 mM Tris pH 7.5 150 mM NaCl	5	32	3.25 ± 0.04
20 mM Tris pH 7.5 150 mM NaCl	6	34	3.26 ± 0.02
20 mM Tris pH 7.5 300 mM NaCl 1M NDSB-201	3	30	2.98 ± 0.03
20 mM Tris pH 7.5 800 mM NaCl	3	28	3.17 ± 0.04

doi: 10.1371/journal.pone.0080371.t001

SAXS-based ensemble analysis captures the dynamic character of P_{ced} in solution

The time-averaged R_g of all simulated models ranges from 2.70 to 2.98 nm (Table 2), which is lower than the measured R_g for P_{ced} (3.26 nm) by at least 3 Å, clearly indicating that P_{ced} possesses IDRs that are not adequately modeled in our classical MDS. Interestingly, data measured in the presence of 1M NDSB-201 shows a R_g of 2.98 nm, suggesting a more stable fold of P_{ced} in these conditions, consistently with the lower χ_{exp} values observed for fitting of the Rosetta models (not shown). In order to explicitly model the flexibility of P_{ced} in solution, we employed the ensemble optimization method (EOM) [28]. To take into account the possibility of IDRs outside P_{core} , we used atomistic structure-based models (SBM) [29] as a mean to rapidly sample the conformational space of residues 158-167 and 199-237 and impose extended or α -helical conformations. We then pooled all models from classical MDS and SBM MDS into an ensemble of $\sim 12,300$ models, which we fitted against SAXS data using EOM [28] to yield optimized ensembles (Figure 5). A representative ensemble of ten

conformers is shown in Figure 5 B. SAXS profiles could be adequately fitted using models from the pool ensemble, with χ_{exp} values reaching a plateau at 0.8-0.9 for an ensemble size of 3 to 5 models (Figure 5 C and D), providing direct evidence for the presence of flexible regions in P_{ced} [30]. Interestingly, the drop in χ_{exp} value upon increasing ensemble size was less pronounced for the NDSB-containing sample, confirming the induced increase in protein fold stability, as has been observed previously [31]. The R_g distributions of the optimized ensembles measured in low or high salt conditions, or in the presence of 1M NDSB-201 are shown in Figure 5 A, revealing a dynamic equilibrium between two populations that correspond to the presence or absence of α -helices outside the core region. Interestingly, the addition of NDSB-201 correlated with an increase in the percentage of α -helices in the optimized ensemble (Figure 5 A).

Crystal structure of P_{core}

A single diffracting crystal of HMPV P_{ced} grew after 140 days, suggesting degradation occurred in the drop prior to

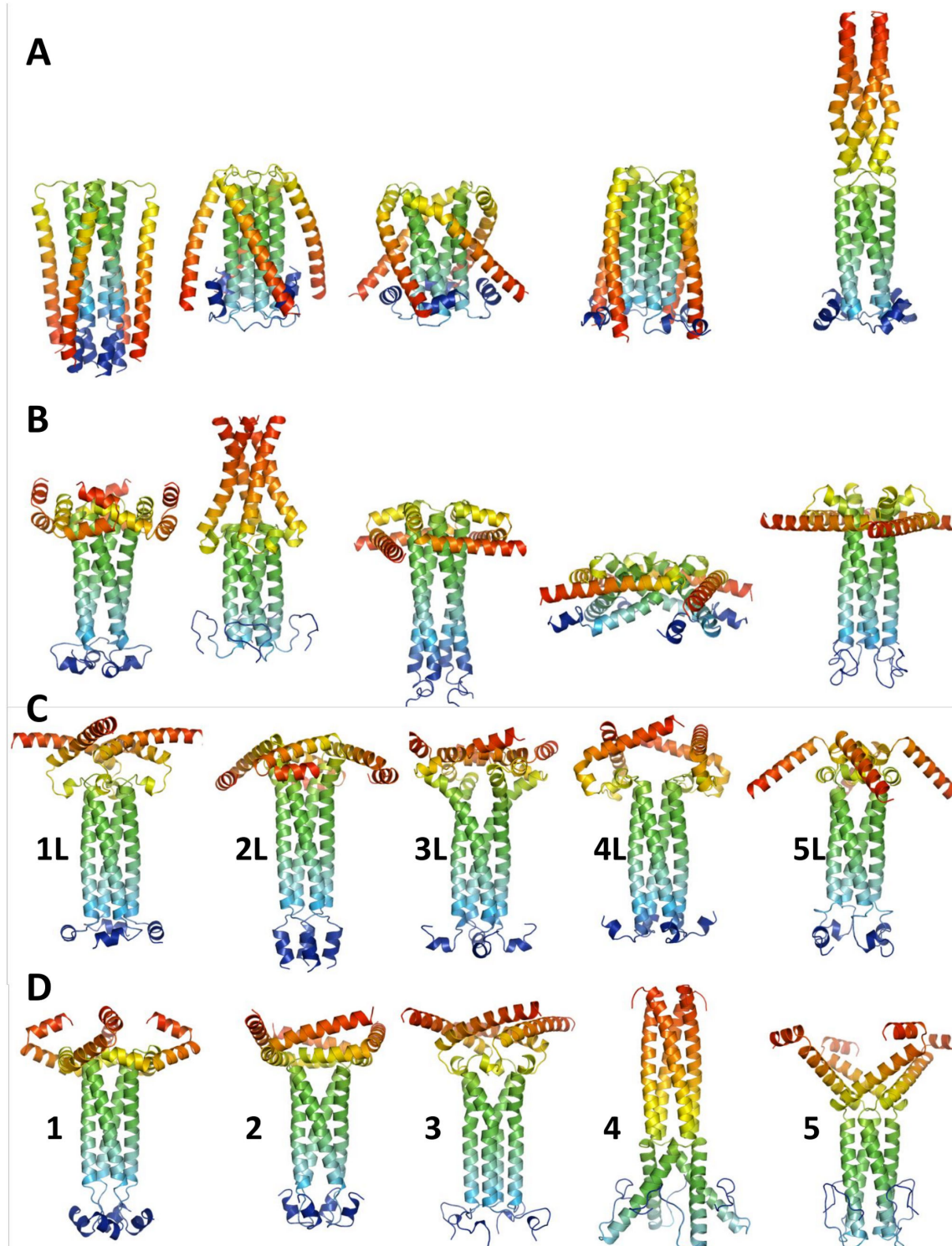


Figure 3. *Ab initio* models of P_{ced} before (A&B) and after (C&D) applying the SAXS filter. A. From left to right, the 5 best-scoring Rosetta fold-and-dock models generated from the sequence of HMPV P residues 155 to 241. Models are shown in cartoon and are coloured from blue (N-terminus) to red (C-terminus). B. 5 best-scoring Rosetta fold-and-dock models from HMPV P residues 156 to 237. C and D. Same as in A and B after filtering out all models with $\chi_{exp} > 1.3$. All models were truncated to residues 156-237 to improve fitting accuracy (residues 156-157 were kept to account for the two extra N-terminal residues resulting from cleavage of the His6 tag by 3C protease).

doi: 10.1371/journal.pone.0080371.g003

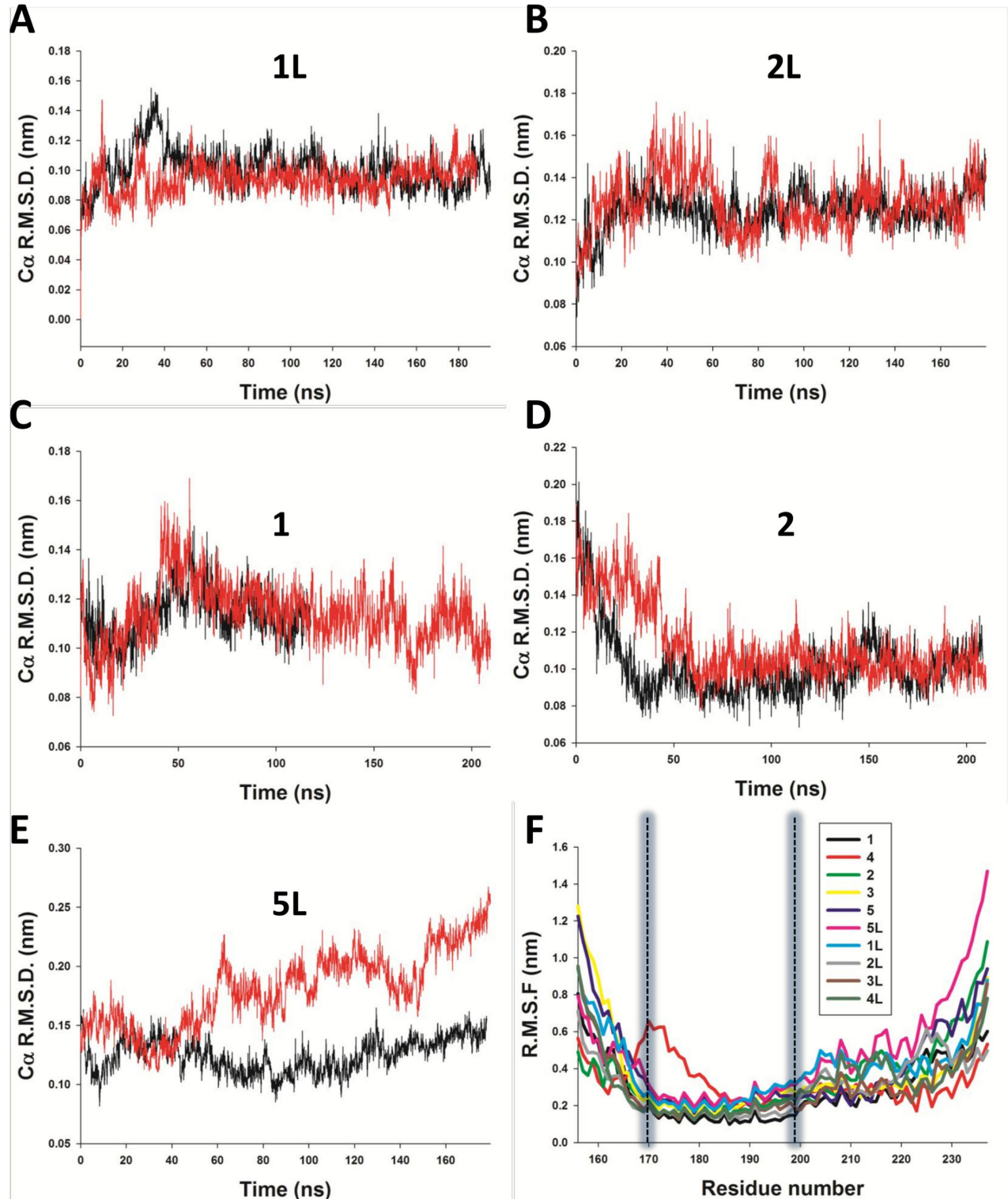


Figure 4. Molecular Dynamics simulations (MDS). A. Plot of root mean square deviation (RMSD) of model 1L P_{core} C α atoms versus simulation time, with respect to the starting structure coordinates. B, C, D and E. RMSD of model 2L, 1, 2 and 5L P_{core} C α atoms using model 1L P_{core} as reference structure. F. Time- and residue-averaged root mean square fluctuations (RMSF) of all simulated models (1L to 5L and 1 to 5). The displayed curves represent average RMSF values over the four chains of a tetramer and two independent simulations.

doi: 10.1371/journal.pone.0080371.g004

Table 2. summary of classical Molecular Dynamics simulations.

Model	Time-averaged χ_{exp}	Time-averaged Rg (nm)	System size (atoms)	Total simulation time (μ s)
1	1.93	2.70	60814	0.32
2	1.37	2.87	58230	0.42
3	1.28	2.95	57043	0.51
4	1.16	2.98	64080	0.41
5	1.55	2.83	64641	0.40
1L	1.31	2.92	71871	0.38
2L	2.34	2.75	67503	0.36
3L	1.72	2.79	57138	0.41
4L	1.92	2.85	56027	0.42
5L	1.28	2.90	81087	0.36

doi: 10.1371/journal.pone.0080371.t002

crystallization. The crystal belonged to space group $P2_12_12_1$ (Table 3). Diffraction data were phased by molecular replacement using residues 169 to 194 from P_{core1} , revealing a tetrameric α -helical arrangement consistent with the modeled structure (Figure 6 A and B). The asymmetric unit contains 2 tetramers, with a solvent content of 40%. Since the asymmetric unit could not physically contain 2 tetramers with residues 158-237 (with 1 tetramer present the solvent content is 16%), packing considerations suggest that degradation necessarily occurred prior to crystallization. Model 1 was selected for molecular replacement because it displayed the lowest RMSF in the core α -helical region during MDS (Figure 4F). The stability of the P_{core} region in *ab initio* models, MDS and SAXS prompted us to use residues 168 to 198 as a search model, however this resulted in rejection of a potential solution with a high translation function Z-score (TFZ) due to steric clashes between tetramers, which were eliminated by using a shorter fragment encompassing residues 169 to 194. The structure was subsequently refined to R_{work} and R_{free} values of 23.5% and 25.2%, respectively, confirming the identity of the modeled residues. The residues 158-168 and 195-237 are absent from the structure, in line with the flexibility observed for these regions by MDS and SAXS-based ensemble optimization. The tetrameric structure is stabilized by a large network of hydrophobic interactions involving Leu176, 183, 187, 189, 190, 193, and Ile172, 179, 186 located on the inner surface of symmetry-related α -helices (Figure 6 C). Additional stability is provided by a solvent-exposed network of ionic interactions created by Glu173/Glu177 and Arg175, or Glu180 and Lys182 side chains from neighbouring protomers. Comparison with representative models from the SAXS optimized ensembles, as well as with the best scoring models from MDS (Figure 3 C & D) demonstrates C α RMSDs ranging from 0.9 to 1.6 Å (Figure 7) over aligned residues, confirming the accuracy of the modeled P_{core} .

Discussion

The structure of P_{ced} reveals the shortest tetrameric coiled-coil among the Paramyxoviridae

The structural data presented here indicates that the α -helical tetramerization domain of HMPV P (residues 171-194) is considerably shorter than the highly conserved central region of the molecule (residues 158-237). Interestingly, the region 195-237 is shown by SAXS-based ensemble optimization to form an IDR with strong α -helical propensity. These transiently folded α -helices can be further stabilized by addition of NDSB-201. Taken together, these features suggest that residues 195 to 237 might constitute a molecular recognition element (MoRE) located directly downstream of the coiled coil region. A sequence alignment of HMPV P with HRSV and BRSV P is shown in Figure 8. The regions that have been mapped in RSV to be required for interaction with the N, L and M2-1 proteins are annotated based on published mutagenesis studies [7,9,11,14,32]. Interestingly, the sequence of P_{core} aligns with a region of RSV P that is necessary for coimmunoprecipitation of the L protein [11], suggesting either direct binding to L or the requirement of a tetrameric P protein for efficient P-L association. Additionally, the α -helical MoRE located between HMPV P residues 195 to 237 shows strong conservation and overlaps with a putative nucleoprotein (N) binding region identified in RSV (residues161-180). Residues 221 to 241 of RSV P are also part of a putative N binding site, which aligns with residues 257 to 277 of HMPV featuring the predicted α 4 helix (Figure 1 and 8). The flexibility of α 3 (residues 195 to 237) relative to α 2 (P_{core}) observed in the SAXS ensembles, together with the requirement of α 3 and α 4 for N binding, suggest that these regions may be part of a nucleoprotein-binding domain (NBD), similar to the C-terminal domain of phosphoproteins from other *Mononegavirales* [33–37]. Alternatively, the C-terminal region of HMPV P may act as a MoRE that folds upon binding to N, and assume an unstable tertiary structure in the absence of binding partner. This hypothesis is supported by the low propensity to form ordered structures observed for the C-terminal region of HMPV P (Figure 1).

P proteins from *Mononegavirales* are large modular proteins that are characterized by extensive IDRs of variable lengths containing multiple MoRE, a central oligomerization domain and, in some but not all viruses, a stable C-terminal domain [13,38–42]. Because P protein sequences vary greatly in length (from 241 residues in RSV P to 709 residues in Nipah virus P), and have diverged beyond remote homology detection, it is difficult to compare P proteins from different families, or sometimes even different genera. However, the available structures of P oligomerization domains allow us to determine phylogenetic relationships from structural alignments. Figure 9 shows a phylogenetic tree of the crystal structures of tetrameric *Paramyxoviridae* P and dimeric *Rhabdoviridae* P oligomerization domains, built using the structure homology program SHP [43]. Interestingly, the tree obtained from P_{ced} structures is similar to trees built based on large numbers of sequences from more conserved proteins such as N, M, F and L [44]. The tree highlights the structural divergence of P protein

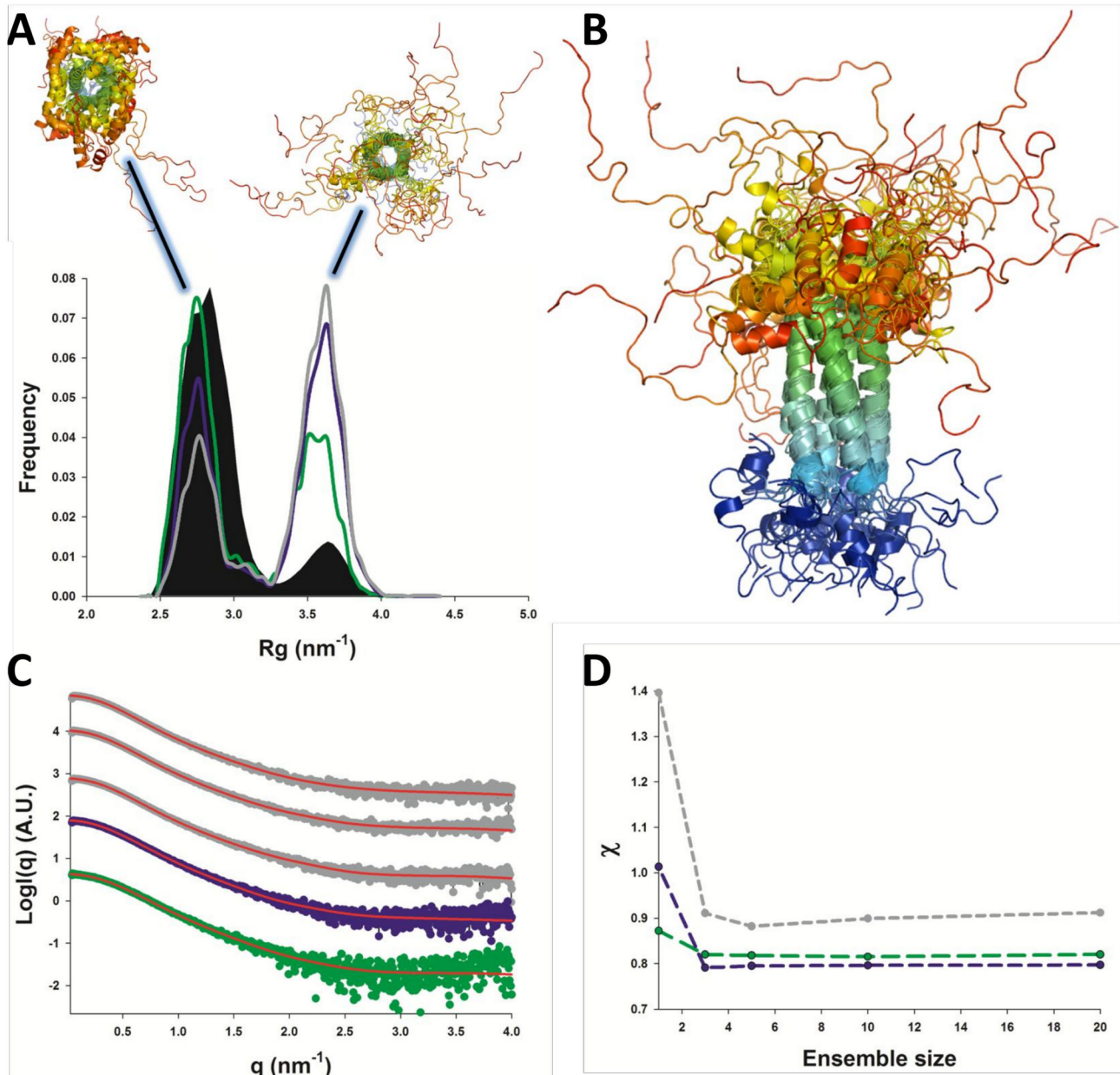


Figure 5. SAXS-based ensemble structure of P_{ced}. A. Radius of gyration (R_g) distributions of the pool (black area) and optimized ensembles selected in the presence of 150 mM NaCl (grey line), 800 mM NaCl (purple line) and 1 M NDSB-201 (green line). Representative conformers populating the two peaks of the R_g distribution are shown above the graph in cartoon representation and coloured from blue (N-terminus) to red (C-terminus). B. An ensemble of 10 conformers selected against the NDSB-containing sample, highlighting the dynamic equilibrium between random coil and α -helical conformations of the C-terminal region. C. EOM-fitted SAXS profiles of P_{ced} are represented by red lines. The colour-coding for the experimental curves is the same as in Figure 2. D. Variation of χ_{exp} with optimized ensemble size for SAXS profiles of P_{ced} measured in the presence of 20 mM Tris pH 7.5, 150 mM NaCl (grey line), in 20 mM Tris pH 7.5, 800 mM NaCl (purple line), or 20 mM Tris pH 7.5, 150 mM NaCl and 1M NDSB-201 (green line).

doi: 10.1371/journal.pone.0080371.g005

oligomerization domains across evolution, with more than three-fold variation in domain length across the *Paramyxoviridae*. The short tetrameric coiled coil from HMPV P

clusters in a separate branch from SeV/MuV/MeV and RV/VSV, emphasizing the pertinence of its classification as a separate subfamily, *Pneumovirinae*.

Table 3. Crystallographic statistics.

Space group	P2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	0.92
Unit cell constants	a=66.9 Å b=48.5 Å c=64.5 Å
Resolution limits ^a	66.9-3.1 Å (3.2-3.1 Å)
Number of measured Reflections	37981
Number of unique Reflections	3602
Completeness of data ^a	90.3% (57.1%)
R _{merge} ^a	35.1% (134.7%)
R _{pim} ^a	15.7% (86.9%)
Multiplicity ^a	10.5 (6.1)
I/σ ^a	6.4(1.4)
R _{work}	23.5%
R _{free}	25.2%
Number of residues in favoured region	177 (96.7%)
Number of residues in allowed region	2 (1.1%)
Number of residues in outlier region	4 (2.2%)
RmsBondLength	0.010
RmsBondAngle	1.290

^a The values for the highest resolution shell are given in parentheses.

doi: 10.1371/journal.pone.0080371.t003

The combined use of *ab initio* modeling with small angle X-ray scattering and molecular dynamics appears to be a promising approach for solving the structure of small homo-oligomeric protein complexes

The work presented in this study combines several computational techniques commonly used in protein structure prediction to yield a correct model for an unknown protein, applying low resolution information about shape and flexibility derived from SAXS as the sole experimental constraint. The advantages of the method can be summarized in four points (1). The recently developed Rosetta fold-and-dock protocol takes advantage of the reduced conformational space available to homo-oligomeric proteins to predict atomistic models [27]. (2) SAXS can be used to successfully filter out a large proportion of incorrect models, as has been shown in *ab initio* protein structure prediction [44–46] and protein-protein docking [47,48]. (3) The usefulness of SAXS data to identify correct models can be increased by extracting information about protein flexibility and disorder through ensemble analysis [28,30], thus tackling the challenges associated with the modeling of partially unstructured proteins. (4) Classical MDS provide an additional mean of selecting and optimizing correct models and detecting flexible regions[49], while its sampling limitations can be overcome by using fast SBM MDS [29,50]. By combining methods (1) to (4), we obtained a detailed, cross-validated picture of HMPV P_{ced} structure and dynamics in solution, showing that it constitutes a promising approach for protein structure determination. The crystal structure of the core region of P_{ced} comes as a validation of the accuracy of the model, and indirectly confirms the flexibility of the degraded flanking regions. The combined use of *ab initio* modeling, MDS and SAXS-based ensemble optimization constitutes a generally applicable method to predict protein structure, both in

the presence of stable, potentially homo-oligomeric domains, transiently structured or completely disordered regions, and should become increasingly useful in the future.

Material & Methods

Sequence-based analyses

Computational meta-disorder predictions and conservation scores based on sequence alignment of sequences from *Pneumovirinae* P were calculated following procedures described in [41]. Consensus secondary structure prediction was obtained from the Dismeta webserver [51].

Protein cloning, expression & purification

The region of the HMPV P gene from strain NL1-00 corresponding to residues 158-237 was amplified by PCR and cloned into pOPINP[52] for expression of P with an N-terminal His6 tag followed by a 3C cleavage site, using a proprietary ligation-independent In-Fusion system (Clontech), following standard procedures. The integrity of the cloned construct was checked by nucleotide sequencing.

The His6-3C-P158-237 construct was expressed in Rosetta2™ *E. coli* cells by overnight incubation under shaking at 17°C following 1 mM IPTG induction of 1 l terrific broth in presence of appropriate antibiotics. Cells were harvested by centrifugation (18°C, 20 min, 4000x g). The resulting cell pellets were resuspended in 20 mM Tris, pH 7.5, 150 mM NaCl, 8 M urea. Cells were lysed by sonication, and the lysate was centrifuged for 45 min at 4°C and 50000x g to remove cell debris. The supernatant was filtered (0.45 µm filter) and loaded on a column containing 2 ml of pre-equilibrated Ni-NTA Agarose (QIAGEN). After extensive washes, the protein was eluted in 20 mM Tris, pH 7.5, 150 mM NaCl, 400mM imidazole. The protein was then subjected to size exclusion chromatography on a S200 column equilibrated in 20 mM Tris, pH 7.5, 1 M NaCl. The His6 tag was removed by addition of 3C protease at 4°C for 72h. The cleaved product was further purified through reverse Ni-NTA purification to remove His-tagged 3C protease followed by an additional gel filtration step in 20 mM Tris, pH 7.5, 150 mM NaCl. The protein was concentrated using a Millipore concentration unit (c/o 10 kDa).

Small angle X-ray scattering experiments

Small angle x-ray scattering measurements of P_{ced} (residues 158 to 237) were performed at the BM29 beamline in the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data was collected at 20°C, a wavelength of 0.0995 nm and a sample-to-detector distance of 1 m. 1D scattering profiles were generated and blank subtraction was performed by the data processing pipeline available at BM29 at the ESRF.

Computational modeling of P_{ced}

The amino-acid sequences of P residues 155-241 or 156-237 were used as input to the Rosetta fold-and-dock protocol with the default recommended parameters [27,53,54]. 2 x 30,000 models were generated and ranked using the Rosetta scoring function. In a second step, models were fitted

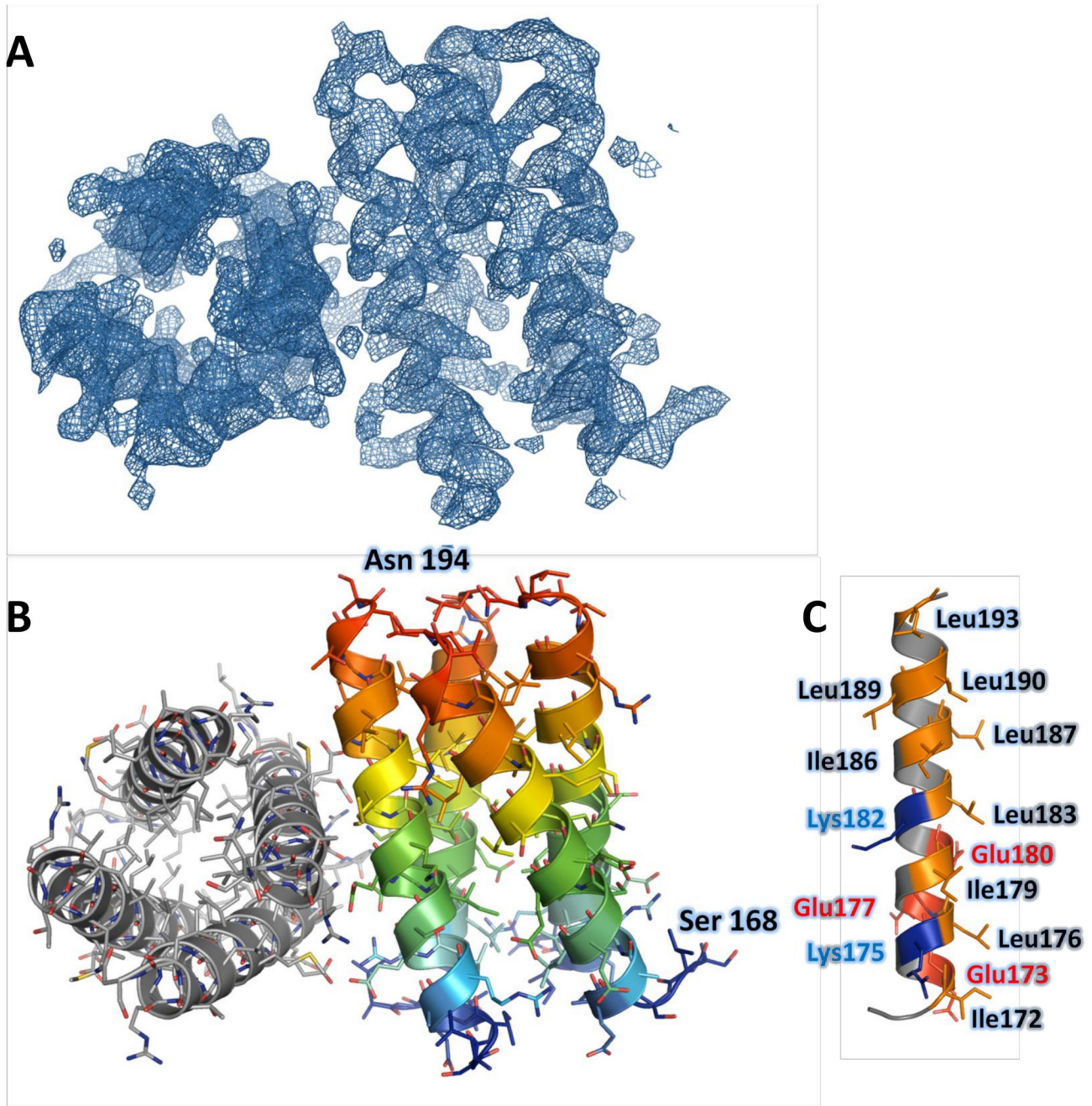


Figure 6. Crystal structure of P_{core} . A. Isomesh surface of the experimental $2F_o - F_c$ electron density contoured at 1.4σ . B. Structure of the asymmetric unit, showing the close packing of two tetrameric molecules. The left tetramer is shown in grey, and the right one is coloured from blue (N-terminus) to red (C-terminus). C. Structure of a single subunit from the crystal, highlighting the residues involved in intermolecular contacts. Hydrophobic residues are coloured in orange, while positively and negatively charged residues are coloured in blue and in red, respectively.

doi: 10.1371/journal.pone.0080371.g006

to experimental SAXS data using CRYSOLOG[55] to yield the agreement between theoretical and experimental profile χ_{exp} . χ_{exp} values were then used to discard incorrect models based on an arbitrary threshold ($\chi_{exp} > 1.3$).

Molecular dynamics simulations and ensemble optimization

All classical MDS were performed using the GROMACS 4 software package [56] and the AMBER99SB-ILDN* force field

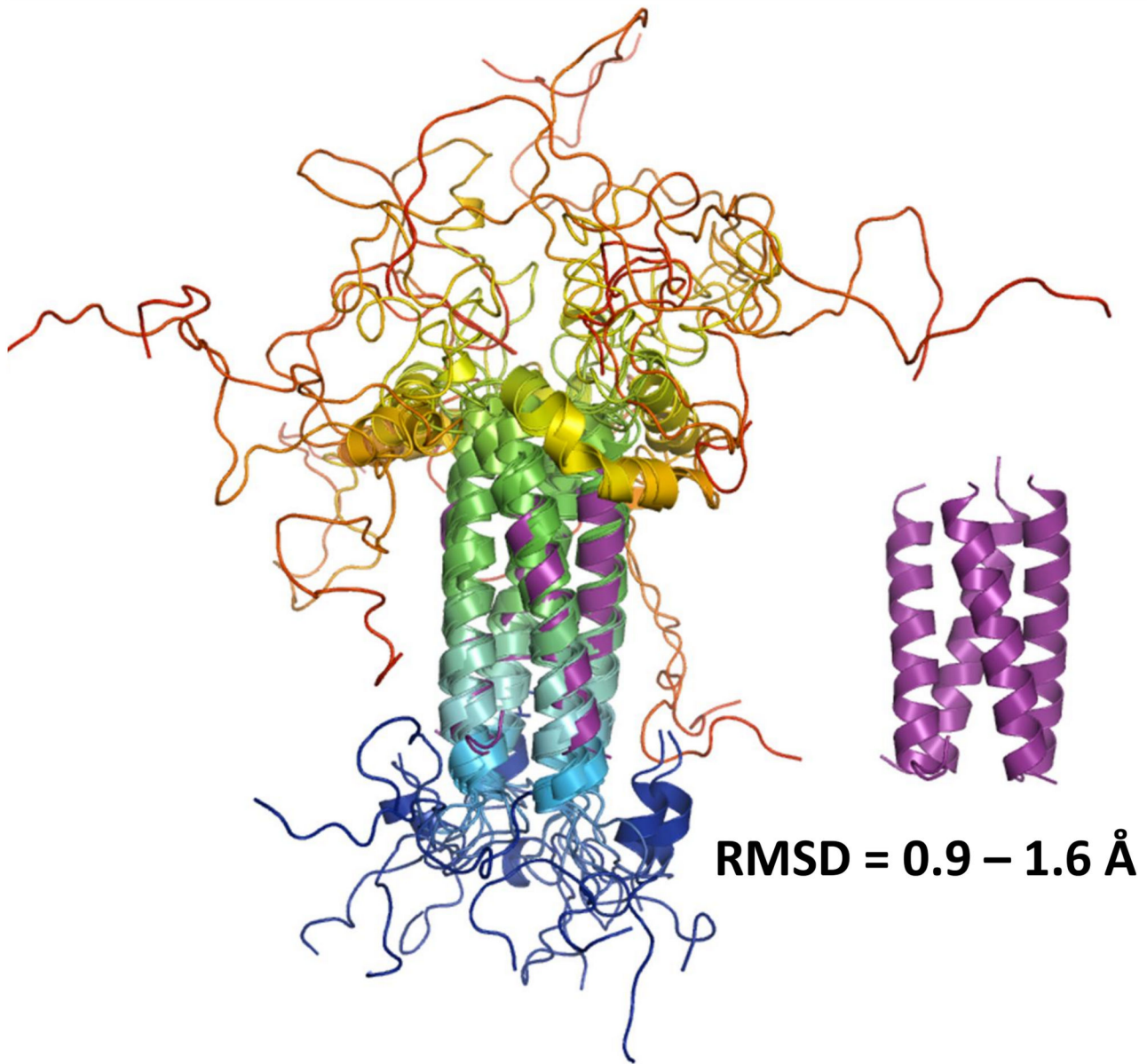


Figure 7. Comparison of the crystal structure of P_{core} with a minimal structural ensemble of P_{ced} . Left panel, the crystal structure from P_{core} (purple cartoon) is overlaid with an optimized ensemble of 5 models of P_{ced} (coloured from blue (N-terminus) to red (C-terminus)) selected from SAXS data measured in 20 mMTris pH 7.5, 150 mMNaCl. Right panel, the crystal structure is additionally shown in purple cartoon. The observed range of C α -RMSD values is indicated (calculated in Pymol).

doi: 10.1371/journal.pone.0080371.g007

[57,58]. At the beginning of each simulation, the protein was immersed in a box of SPC/E water. A minimum distance of 1.0 nm was applied between any protein atom and the edges of the box. Sodium ions were added to reach neutrality. Long range electrostatics were treated with the particle-mesh Ewald summation [59]. Bond lengths were constrained using the P-LINCS algorithm [60]. Hydrogens were treated as virtual sites [61], enabling an integration time step of 5fs. The v -rescale

thermostat [62] and the Parrinello–Rahman barostat [63] were used to maintain a temperature of 300 K and a pressure of 1 atm. Each system was energy minimized using 1,000 steps of steepest descent and equilibrated for 200 ps with restrained protein heavy atoms before the beginning of the production simulation. For each system, two independent production simulations were obtained by using different initial velocities. The aggregated simulation time was $\sim 4.15 \mu\text{s}$ (Table 2).

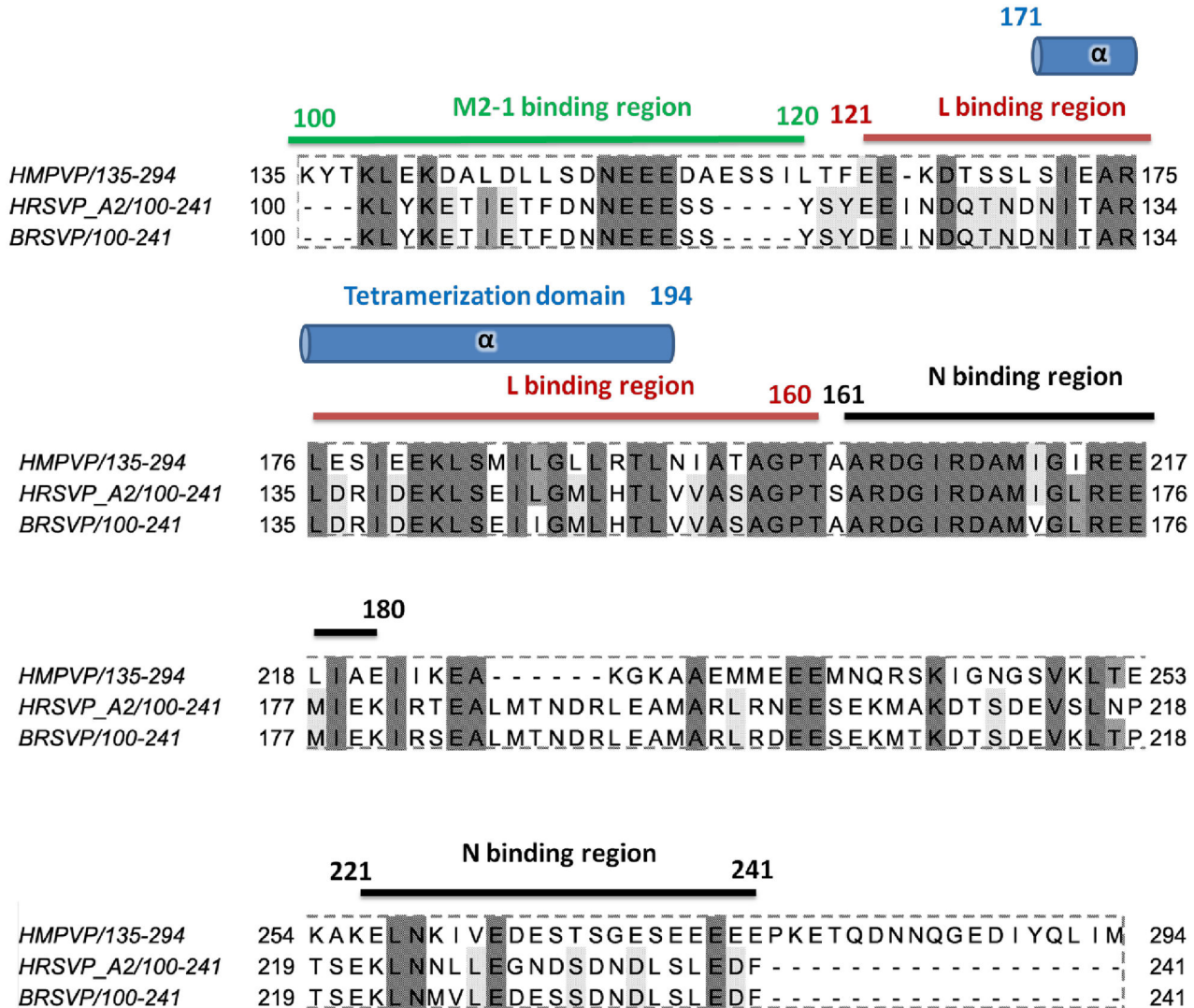


Figure 8. Annotated sequence alignment of the conserved C-terminal region of HMPV, HRSV and BRSV P. The location of the tetramerization domain (P_{core}) and the position of conserved residues are highlighted. Regions that have been associated with N, M2-1 or L interaction in RSV are indicated by a bar based on [7,9,11,14,32].

doi: 10.1371/journal.pone.0080371.g008

Calculation of root mean square deviations (RMSD) and root mean square fluctuations (RMSF) were carried out using GROMACS routines. Snapshots were extracted every 500 ps, resulting in a pool of ~8,300 models.

In order to obtain a more complete sampling of the IDRs, in particular those located between residues 195 and 237, model 1 was simulated using an atomistic coarse-grained structure-based model (SBM) [29,50]. Model 1 was selected because of the higher stability of the α -helical fold adopted by residues 195 to 220, as shown by its low RMSF (Figure 4), and also its higher frequency of selection in optimized ensembles (not shown). Two additional systems were simulated with either residues 220 to 237, or residues 195 to 237 in extended starting conformations, allowing fast sampling of the IDRs

motions. 2000 snapshots were extracted from each simulation, yielding 4,000 additional models to the pool.

For each model from the pool ensemble (~12,300 models), theoretical SAXS patterns were calculated with the program CRYSOLOG [55] and ensemble fitting was performed with GAJOE [28]. The number of models in the selected ensemble was varied from 1 to 20 in order to determine the size of the minimal ensemble required to describe the data.

Crystallization and data collection

Crystallization was carried out via the vapor diffusion method using a Cartesian Technologies pipetting system [64]. The P158-237 construct crystallized after ~142 days in 25 % PEG 3350, 100 mM HEPES pH 7.5 at 20°C. Crystals were frozen in

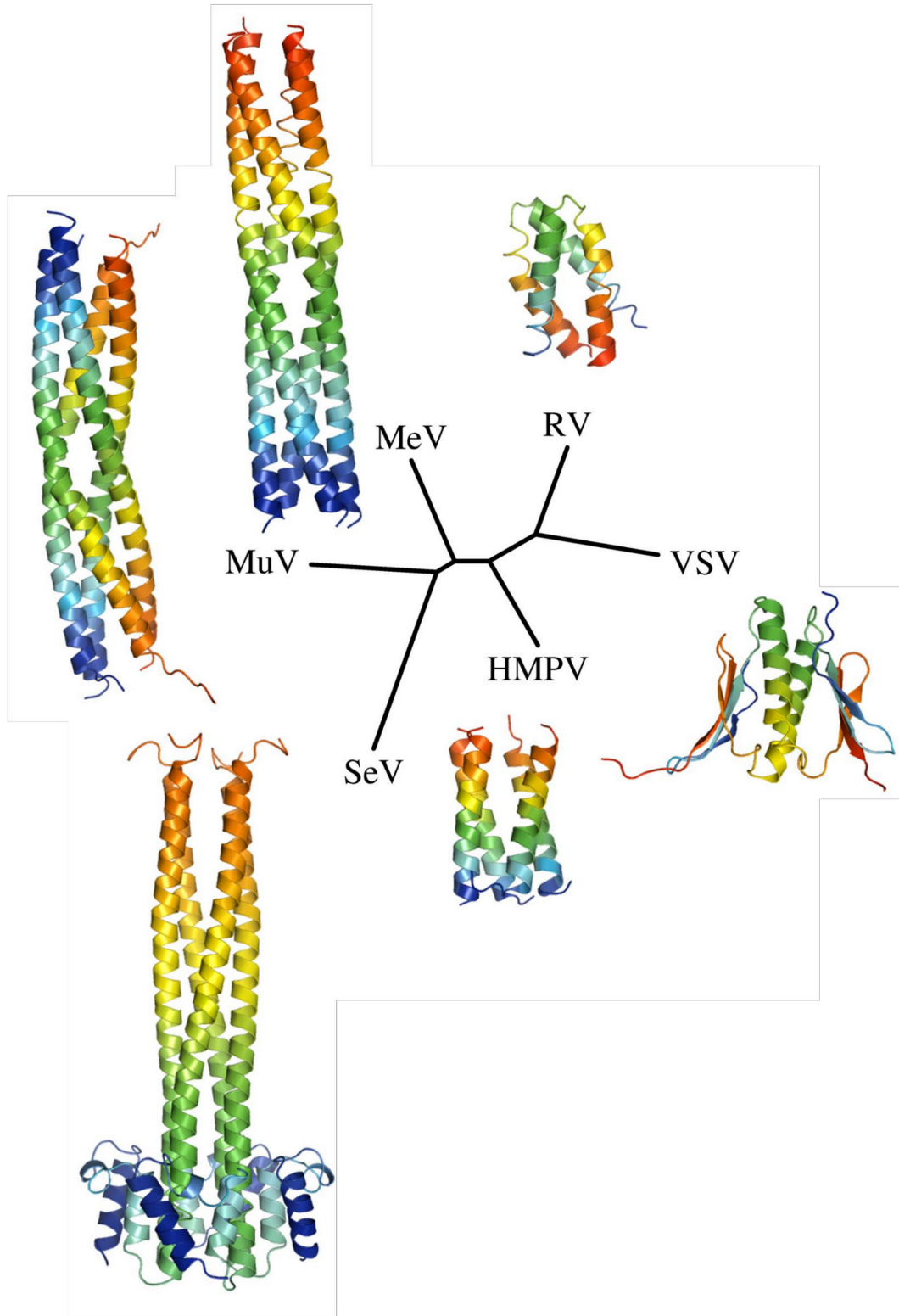


Figure 9. Structure-based phylogenetic tree of *Mononegavirales*P proteins. The tree was built by aligning the experimentally determined structures of Sendai virus (SeV) (PDB ID: 1EZJ), Measles virus (MeV) (PDB ID: 3ZDO), Mumps virus (MuV) (PDB ID: 4EIJ), Rabies Virus (RV) (PDB ID: 3L32), Vesicular stomatitis virus (VSV) (PDB ID: 2FQM) and HMPV P oligomerization domains using SHP [43]. The obtained evolutionary distances were used to draw a tree in PHYLIP [77].

doi: 10.1371/journal.pone.0080371.g009

liquid nitrogen after being soaked in a mother liquor solution supplemented with 25% glycerol. Diffraction data was recorded on the I04 beamline at Diamond Light Source, Didcot, UK.

Structure determination and refinement

Anisotropic diffraction data to a resolution of 3.1Å were indexed and integrated using XDS [65] and scaled with SCALA [66] as implemented in the program xia2 [67]. The structure was determined by molecular replacement using P_{core} residues from model 1 as a search model in PHASER [68]. The solution was subjected to repetitive rounds of restrained refinement in PHENIX [69] and Autobuster [70] and manual building in COOT [71]. Eight-fold non crystallographic local structure similarity restraints [72] were used throughout refinement and TLS parameters were included in the final round of refinement. The CCP4 program suite [73] was used for coordinate manipulations. The structures were validated with Molprobity [74]. Refinement statistics are given in Table 3, and final refined coordinates and structure factors have been deposited in the PDB with accession code 4BXT.

Structure analysis

All the structure-related figures were prepared with the PyMOL Molecular Graphics System (DeLano Scientific LLC). Protein interfaces were analyzed with the PISA webserver [75]. Structural alignments were calculated using PyMOL and SHP [43].

References

- Boivin G, Abed Y, Pelletier G, Ruel L, Moisan D et al. (2002) Virological features and clinical manifestations associated with human metapneumovirus: a new paramyxovirus responsible for acute respiratory-tract infections in all age groups. *J Infect Dis* 186: 1330-1334. doi:10.1086/344319. PubMed: 12402203.
- van den Hoogen BG (2007) Respiratory tract infection due to human metapneumovirus among elderly patients. *Clin Infect Dis* 44: 1159-1160. doi:10.1086/513295. PubMed: 17407032.
- van den Hoogen BG, van Doornum GJ, Fockens JC, Cornelissen JJ, Beyer WE et al. (2003) Prevalence and clinical symptoms of human metapneumovirus infection in hospitalized patients. *J Infect Dis* 188: 1571-1577. doi:10.1086/379200. PubMed: 14624384.
- Williams JV, Harris PA, Tollefson SJ, Halburnt-Rush LL, Pingsterhaus JM et al. (2004) Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. *N Engl J Med* 350: 443-450. doi:10.1056/NEJMoa025472. PubMed: 14749452.
- Xepapadaki P, Psarras S, Bossios A, Tsolia M, Gourgiotis D et al. (2004) Human Metapneumovirus as a causative agent of acute bronchiolitis in infants. *J Clin Virol* 30: 267-270. doi:10.1016/j.jcv.2003.12.012. PubMed: 15135747.
- van den Hoogen BG, Bestebroer TM, Osterhaus AD, Fouchier RA (2002) Analysis of the genomic sequence of a human metapneumovirus. *Virology* 295: 119-132. doi:10.1006/viro.2001.1355. PubMed: 12033771.
- Slack MS, Easton AJ (1998) Characterization of the interaction of the human respiratory syncytial virus phosphoprotein and nucleocapsid protein using the two-hybrid system. *Virus Res* 55: 167-176. doi: 10.1016/S0168-1702(98)00042-2. PubMed: 9725669.
- Marriott AC, Wilson SD, Randhawa JS, Easton AJ (1999) A single amino acid substitution in the phosphoprotein of respiratory syncytial virus confers thermosensitivity in a reconstituted RNA polymerase system. *J Virol* 73: 5162-5165. PubMed: 10233981.
- Mallipeddi SK, Lupiani B, Samal SK (1996) Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P interaction using a two-hybrid system. *J Gen Virol* 77 (5): 1019-1023. doi:10.1099/0022-1317-77-5-1019. PubMed: 8609467.
- Lu B, Brazas R, Ma CH, Kristoff T, Cheng X et al. (2002) Identification of temperature-sensitive mutations in the phosphoprotein of respiratory syncytial virus that are likely involved in its interaction with the nucleoprotein. *J Virol* 76: 2871-2880. doi:10.1128/JVI.76.6.2871-2880.2002. PubMed: 11861854.
- Khattar SK, Yunus AS, Samal SK (2001) Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P and P-L interactions using a minigenome system. *J Gen Virol* 82: 775-779. PubMed: 11257181.
- Khattar SK, Yunus AS, Collins PL, Samal SK (2001) Deletion and substitution analysis defines regions and residues within the phosphoprotein of bovine respiratory syncytial virus that affect transcription, RNA replication, and interaction with the nucleoprotein. *Virology* 285: 253-269. doi:10.1006/viro.2001.0960. PubMed: 11437660.
- Jensen MR, Bernadó P, Houben K, Blanchard L, Marion D et al. (2010) Structural disorder within sendai virus nucleoprotein and phosphoprotein: insight into the structural basis of molecular recognition. *Protein Pept Lett* 17: 952-960. doi: 10.2174/092986610791498885. PubMed: 20450486.
- García-Barreno B, Delgado T, Melero JA (1996) Identification of protein regions involved in the interaction of human respiratory syncytial virus phosphoprotein and nucleoprotein: significance for nucleocapsid assembly and formation of cytoplasmic inclusions. *J Virol* 70: 801-808. PubMed: 8551618.
- Derdowski A, Peters TR, Glover N, Qian R, Utey TJ et al. (2008) Human metapneumovirus nucleoprotein and phosphoprotein interact and provide the minimal requirements for inclusion body formation. *J Gen Virol* 89: 2698-2708. doi:10.1099/vir.0.2008/004051-0. PubMed: 18931065.
- Karlin D, Ferron F, Canard B, Longhi S (2003) Structural disorder and modular organization in Paramyxovirinae N and P. *J Gen Virol* 84: 3239-3252. doi:10.1099/vir.0.19451-0. PubMed: 14645906.
- Leyrat C, Yabukarski F, Tarbouriech N, Ribeiro EA Jr., Jensen MR et al. Structure of the vesicular stomatitis virus N(0)-P complex. *PLOS Pathog* 7: e1002248.
- Fearn R, Collins PL (1999) Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. *J Virol* 73: 5852-5864. PubMed: 10364337.

Acknowledgements

The work presented here made use of the High Performance Computing facility IRIDIS provided by the EPSRC funded Centre for Innovation (EP/K000144/1 and EP/K000136/1) which is owned and operated by the e-Infrastructure South Consortium formed by the universities of Bristol, Oxford, Southampton and UCL in partnership with STFC's Rutherford Appleton Laboratory. The WeNMR project (European FP7 e-Infrastructure grant, contract no. 261572, www.wenmr.eu), supported by the European Grid Initiative (EGI) through the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands (via the Dutch BiG Grid project), Portugal, Spain, UK, South Africa, Malaysia, Taiwan and the Latin America GRID infrastructure via the Gisela project is acknowledged for the use of web portals, computing and storage facilities. Finally the authors would like to thank Diamond Light Source for beamtime (proposal MX8423), and the staff of beamlines I02, I03 I04, I04-1 and I24 for assistance with crystal testing and data collection.

Author Contributions

Conceived and designed the experiments: CL JMG. Performed the experiments: CL MR KH. Analyzed the data: CL. Wrote the manuscript: CL MR JMG.

19. Buchholz UJ, Biacchesi S, Pham QN, Tran KC, Yang L et al. (2005) Deletion of M2 gene open reading frames 1 and 2 of human metapneumovirus: effects on RNA synthesis, attenuation, and immunogenicity. *J Virol* 79: 6588-6597. doi:10.1128/JVI.79.11.6588-6597.2005. PubMed: 15890897.
20. Tarbouriech N, Curran J, Ruigrok RW, Burmeister WP (2000) Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol* 7: 777-781. doi:10.1038/79013. PubMed: 10966649.
21. Communie G, Crépin T, Maurin D, Jensen MR, Blackledge M et al. (2013) Structure of the tetramerization domain of measles virus phosphoprotein. *J Virol* 87: 7166-7169. doi:10.1128/JVI.00487-13. PubMed: 23576502.
22. Cox R, Green TJ, Purushotham S, Deivanayagam C, Bedwell GJ et al. (2013) Structural and functional characterization of the mumps virus phosphoprotein. *J Virol* 87: 7558-7568. doi:10.1128/JVI.00653-13. PubMed: 23637399.
23. Llorente MT, Garcia-Barreno B, Calero M, Camafeita E, López JA et al. (2006) Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. *J Gen Virol* 87: 159-169. doi:10.1099/vir.0.81430-0. PubMed: 16361428.
24. Llorente MT, Taylor IA, López-Viñas E, Gomez-Puertas P, Calder LJ et al. (2008) Structural properties of the human respiratory syncytial virus P protein: evidence for an elongated homotetrameric molecule that is the smallest orthologue within the family of paramyxovirus polymerase cofactors. *Proteins* 72: 946-958. doi:10.1002/prot.21988. PubMed: 18300250.
25. Leyrat C, Jensen MR, Ribeiro EA Jr., Gérard FC, Ruigrok RW et al. (2011) The N(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient alpha-helices. *Protein Sci* 20: 542-556. doi:10.1002/pro.587. PubMed: 21207454.
26. Rambo RP, Tainer JA. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496: 477-481. doi:10.1038/nature12070. PubMed: 23619693.
27. Das R, André I, Shen Y, Wu Y, Lemak A et al. (2009) Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci U S A* 106: 18978-18983. doi:10.1073/pnas.0904407106. PubMed: 19864631.
28. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129: 5656-5664. doi:10.1021/ja069124n. PubMed: 17411046.
29. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY et al. (2009) An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* 75: 430-441. doi:10.1002/prot.22253. PubMed: 18837035.
30. Pelikan M, Hura GL, Hammel M (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* 28: 174-189. doi:10.4149/gpb_2009_02_174. PubMed: 19592714.
31. Expert-Bezançon N, Rabilloud T, Vuillard L, Goldberg ME (2003) Physical-chemical features of non-detergent sulfobetaines active as protein-folding helpers. *Biophys Chem* 100: 469-479. doi:10.1016/S0301-4622(02)00299-5. PubMed: 12646384.
32. Mason SW, Aberg E, Lawetz C, DeLong R, Whitehead P et al. (2003) Interaction between human respiratory syncytial virus (RSV) M2-1 and P proteins is required for reconstitution of M2-1-dependent RSV minigenome activity. *J Virol* 77: 10670-10676. doi:10.1128/JVI.77.19.10670-10676.2003. PubMed: 12970453.
33. Ribeiro EA Jr., Favier A, Gerard FC, Leyrat C, Brutscher B et al. (2008) Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol* 382: 525-538. doi:10.1016/j.jmb.2008.07.028. PubMed: 18657547.
34. Martinez N, Ribeiro EA Jr., Leyrat C, Tarbouriech N, Ruigrok RW et al. (2013) Structure of the C-terminal domain of Lettuce Necrotic Yellow Virus phosphoprotein. *J Virol* 87: 9569-78. PubMed: 23785215.
35. Mavrakīs M, McCarthy AA, Roche S, Blondel D, Ruigrok RW (2004) Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J Mol Biol* 343: 819-831. doi:10.1016/j.jmb.2004.08.071. PubMed: 15476803.
36. Blanchard L, Tarbouriech N, Blackledge M, Timmins P, Burmeister WP et al. (2004) Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology* 319: 201-211. doi:10.1016/j.virol.2003.10.029. PubMed: 14980481.
37. Johansson K, Bourhis JM, Campanacci V, Cambillau C, Canard B et al. (2003) Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem* 278: 44567-44573. doi:10.1074/jbc.M308745200. PubMed: 12944395.
38. Habchi J, Mamelli L, Darbon H, Longhi S. (2010) Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLOS ONE* 5: e11684. doi:10.1371/journal.pone.0011684. PubMed: 20657787.
39. Longhi S, Oglesbee M. (2010) Structural disorder within the measles virus nucleoprotein and phosphoprotein. *Protein Pept Lett* 17: 961-978. doi:10.2174/092986610791498894. PubMed: 20450481.
40. Leyrat C, Gérard FC, de Almeida Ribeiro E Jr., Ivanov I, Ruigrok RW et al. (2010) Structural disorder in proteins of the rhabdoviridae replication complex. *Protein Pept Lett* 17: 979-987. doi:10.2174/092986610791498939. PubMed: 20450482.
41. Gerard FC, Ribeiro Ede A Jr., Leyrat C, Ivanov I, Blondel D et al. (2009) Modular organization of rabies virus phosphoprotein. *J Mol Biol* 388: 978-996. doi:10.1016/j.jmb.2009.03.061. PubMed: 19341745.
42. Leyrat C, Schneider R, Ribeiro EA Jr., Yabukarski F, Yao M et al. (2012) Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein. *J Mol Biol* 423: 182-197. doi:10.1016/j.jmb.2012.07.003. PubMed: 22789567.
43. Abrescia NG, Bamford DH, Grimes JM, Stuart DI. (2012) Structure unifies the viral universe. *Annu Rev Biochem* 81: 795-822. doi:10.1146/annurev-biochem-060910-095130. PubMed: 22482909.
44. Hyndman TH, Marschang RE, Wellehan JF Jr., Nicholls PK. (2012) Isolation and molecular identification of Sunshine virus, a novel paramyxovirus found in Australian snakes. *Infect Genet Evol* 12: 1436-1446. doi:10.1016/j.meegid.2012.04.022. PubMed: 22575820.
45. Zheng W, Doniach S (2002) Protein structure prediction constrained by solution X-ray scattering data and structural homology identification. *J Mol Biol* 316: 173-187. doi:10.1006/jmbi.2001.5324. PubMed: 11829511.
46. Zheng W, Doniach S (2005) Fold recognition aided by constraints from small angle X-ray scattering data. *Protein Eng Des Sel* 18: 209-219. doi:10.1093/protein/gzi026. PubMed: 15845555.
47. Pons C, D'Abramo M, Svergun DI, Orozco M, Bernadó P et al. (2010) Structural characterization of protein-protein complexes by integrating computational docking with small-angle scattering data. *J Mol Biol* 403: 217-230. doi:10.1016/j.jmb.2010.08.029. PubMed: 20804770.
48. Schneidman-Duhovny D, Hammel M, Sali A. (2011) Macromolecular docking restrained by a small angle X-ray scattering profile. *J Struct Biol* 173: 461-471. doi:10.1016/j.jsb.2010.09.023. PubMed: 20920583.
49. Taly JF, Marin A, Gibrat JF (2008) Can molecular dynamics simulations help in discriminating correct from erroneous protein 3D models? *BMC Bioinformatics* 9: 6. doi:10.1186/1471-2105-9-S9-S6. PubMed: 18179702.
50. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. (2010) SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res* 38: W657-W661. doi:10.1093/nar/gkq498. PubMed: 20525782.
51. Wassenaar T, Dijk M, Loureiro-Ferreira N, Schot G, Vries S et al. WeNMR: Structural Biology on the Grid. - 10 - 767.
52. Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R et al. (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* 35: e45. doi:10.1093/nar/gkm047. PubMed: 17317681.
53. Das R, Baker D (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem* 77: 363-382. doi:10.1146/annurev.biochem.77.062906.171838. PubMed: 18410248.
54. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. *Methods Enzymol* 383: 66-93. doi:10.1016/S0076-6879(04)83004-0. PubMed: 15063647.
55. Svergun D, Barberato C, Koch MHJ (1995) CRYSOLO-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28: 768-773. doi:10.1107/S0021889895007047.
56. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Computation* 4: 435-447. doi:10.1021/ct700301q.
57. Best RB, Hummer G (2009) Optimized Molecular Dynamics Force Fields Applied to the Helix α -Coil Transition of Polypeptides. *J Phys Chem B* 113: 9004-9015. doi:10.1021/jp901540t. PubMed: 19514729.
58. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78: 1950-1958. PubMed: 20408171.
59. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103: 8577-8593. doi:10.1063/1.470117.

60. Hess B (2008) P-lincs: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4: 116-122.
61. Berendsen HJC, Gunsteren WF (1984) *Molecular Dynamics Simulations: Techniques and Approaches*. - 135: - 500.
62. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126: 014101. doi:10.1063/1.2408420. PubMed: 17212484.
63. Rühle V (2008) Pressure coupling/barostats.
64. Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG et al. (2005) A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* 61: 651-657. doi:10.1107/S1744309105016933. PubMed: 15930615.
65. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. - 26: - 800.
66. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62: 72-82. doi:10.1107/S0108767306098564. PubMed: 16369096.
67. Winter G, Lobley CM, Prince SM. (2013) Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* 69: 1260-1273. doi:10.1107/S0907444913015308. PubMed: 23793152.
68. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658-674. doi:10.1107/S0021889807021206. PubMed: 19461840.
69. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66: 213-221. doi:10.1107/S0907444909052925. PubMed: 20124702.
70. Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM et al. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* 60: 2210-2221. doi:10.1107/S0907444904016427. PubMed: 15572774.
71. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501. doi:10.1107/S0907444910007493. PubMed: 20383002.
72. Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W et al. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr* 68: 368-380. doi:10.1107/S0907444911056058. PubMed: 22505257.
73. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235-242. doi:10.1107/S0907444910045749. PubMed: 21460441.
74. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12-21. doi:10.1107/S1744309109042018. PubMed: 20057044.
75. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774-797. doi:10.1016/j.jmb.2007.05.022. PubMed: 17681537.
76. Pei J, Grishin NV (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* 17: 700-712. doi:10.1093/bioinformatics/17.8.700. PubMed: 11524371.
77. Retief JD (2000) Phylogenetic analysis using PHYLIP. *Methods Mol Biol* 132: 243-258. PubMed: 10547839.