



OPEN

Crystal structures of FNIP/FGxxFN motif-containing leucine-rich repeat proteins

Trevor Huyton¹✉, Mamta Jaiswal¹, Waltraud Taxer¹, Matthias Fischer² & Dirk Görlich¹✉

The *Cafeteria roenbergensis* virus (Crov), *Dictyostelium*, and other species encode a large family of leucine-rich repeat (LRR) proteins with FGxxFN motifs. We determined the structures of two of them and observed several unique structural features that set them aside from previously characterized LRR family members. Crov588 comprises 25 regular repeats with a LxxLxFGxxFNQxlxENVLPxx consensus, forming a unique closed circular repeat structure. Novel features include a repositioning of a conserved asparagine at the middle of the repeat, a double phenylalanine spine that generates an alternate core packing arrangement, and a histidine/tyrosine ladder on the concave surface. Crov539 is smaller, comprising 12 repeats of a similar LxxLxFGxxFNQPIExVxW/LPxx consensus and forming an unusual cap-swapped dimer structure. The phenylalanine spine of Crov539 is supplemented with a tryptophan spine, while a hydrophobic isoleucine-rich patch is found on the central concave surface. We present a detailed analysis of the structures of Crov588 and Crov539 and compare them to related repeat proteins and other LRR classes.

Abbreviations

LRR Leucine-rich repeat
Crov *Cafeteria roenbergensis* Virus

Cafeteria roenbergensis virus (Crov) is a giant virus in the family *Mimiviridae* that infects marine microplankton such as the bicosoecid flagellate *Cafeteria* sp.^{1–3}. The virus was originally isolated during the 1980s from seawater samples taken in the Gulf of Mexico³. Since its isolation, the genomics and structure of the virus have been well described. Crov has an exceptionally large genome with 692 kilobase pairs of double-stranded DNA and > 500 predicted genes^{1,2,4–8}. Its capsid has a diameter of 3000 Å and represents one of the largest structures reconstructed by cryo-electron microscopy⁹. Crov is one of several giant viruses that are parasitized by smaller DNA viruses called virophages⁵.

Our interest in Crov arose after discovering that it contains multiple proteins with long repetitive motifs^{1,10}, which we will refer to as FGxxFN repeat proteins. Specifically, these motifs contain FG dipeptides that are best known from intrinsically disordered FG nucleoporin repeat domains^{11,12}, which can cohesively interact with each other and condense into an FG phase¹³. This FG phase functions as a sieve-like permeability barrier of nuclear pore complexes (NPCs). It allows for controlled transport between the nuclear and cytoplasmic compartments via shuttling nuclear transport receptors that interact with FG motifs (for a detailed review, see¹⁴). We have also shown that isolated nucleoporin FG repeats form hydrogels that reproduce the permeability properties of NPCs remarkably well^{13–17}. However, structural information on FG-FG interactions is extremely limited due to the technical problems of measuring transient hydrophobic interactions. We therefore reasoned that the structure of a soluble FG-containing protein might give some further insights into the interactions and stacking of phenylalanine residues within FG-containing peptides.

The FGxxFN repeat proteins belong to a larger family termed FNIP proteins (Pfam PF05725), which in turn appear related to the large and well-studied Leucine-rich repeat (LRR) family of proteins. LRRs are present in over 60 000 proteins identified in viruses, bacteria, archaea, and eukaryotes. The structures of many LRR proteins are known. The arrangement of tandem LRR repeats gives rise to a typical curved α/β -solenoid structure. Many LRR-containing proteins function as scaffolds that participate in protein–protein or protein–ligand interactions and help to regulate processes like signal transduction, cell adhesion, DNA repair, recombination, transcription, RNA processing, disease resistance, apoptosis, and innate immune response in mammals, fish,

¹Department of Cellular Logistics, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany. ²Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Heidelberg, Germany. ✉email: trevor.huyton@mpinat.mpg.de; goerlich@mpinat.mpg.de

and plants¹⁸. LRR-containing proteins include RNase inhibitors, toll-like receptors (TLRs), hormone receptors, tyrosine kinase receptors, cell-adhesion molecules, bacterial virulence factors, and many others^{18,19}. LRRs vary in consensus sequence and repeat length, ranging typically between 20 and 30 residues.

LRRs are divided into a highly conserved and a variable segment. The conserved segment of the prototypic LRR proteins consists of an 11-residue LxxLxLxxNxL stretch or a 12-residue LxxLxLxxCxxL stretch, (where L = leucine/isoleucine/valine, N = asparagine/cysteine/threonine/serine, C = cysteine/ and x = any amino acid). Traditionally, LRRs are divided into eight classes. These include (1) typical, (2) ribonuclease inhibitor (RI)-like, (3) cysteine-containing (CC), (4) plant-specific (PS), (5) SDS22-like, (6) bacterial, (7) TpLRR (*Treponema pallidum* LRR), and (8) IRREKO^{20–22}.

The FNIP family (Pfam PF05725) so far comprises > 2000 family members, including many repeat proteins from Crov^{1,10}, Megavirus baoshan²³, and the slime mold genera *Polysphondylium* and *Dictyostelium*^{24,25}. The FNIP repeats were named after their 22 residues long hxxhxxhxxxFNxxhxxxIPxx consensus (with h being hydrophobic residues). The here described FGxxFN repeats (including the Crov repeat proteins) are a major subfamily of them. However, to our knowledge, no experimental structures of FGxxFN or FNIP leucine-rich repeats have been elucidated so far. We now close this gap and report the crystal structures of Crov588 (with an LxxLxLFGxxFNQxIxENVLPxx consensus) and Crov539 (with an LxxLxLFGxxFNQPIExVxWPxx consensus) that are quite representative of the two subfamilies of Crov FGxxFN repeat proteins (supplementary sequence file 1). The structures reveal a so far unique hydrophobic packing with a buried phenylalanine spine and thus define a new sub-class of LRR proteins.

Results

Recombinant expression of Crov LRR proteins and their thermostability. Initial analysis of the Crov genome identified ~ 30 genes encoding FGxxFN/ FNIP repeat-containing proteins¹. Recent re-sequencing using long-read technology¹⁰ resolved additional repeat clusters and increased this number to ~ 70 (for repeat annotations, see Supplementary sequence file 1). Using codon-optimized gene constructs, we expressed and purified five of these proteins in *E. coli*: Crov527, Crov539, Crov563, Crov564, and Crov588 (Fig. 1a). An alignment of the repeat sequences for all five overexpressed Crov repeat proteins and their corresponding Logo plots (<https://weblogo.berkeley.edu/logo.cgi>) for their regular repeats (22 residues) within these sequences is presented in Fig. 1b.

Initial heat denaturation experiments at 97 °C for 5 min using purified protein indicated high thermal stability for all the Crov FGxxFN LRR proteins with ~ 25% of Crov527, 539, and 563 remaining in the supernatant after centrifugation. Crov564 and Crov588 remained soluble to even ~ 50% after heat treatment. Thermofluor assays, using SYPRO orange as an unfolding indicator, revealed, however, lower melting temperatures of 30–48 °C (Fig. 1c). We interpret this as efficient renaturation following melting, with refolding of Crov564 and 588 being more efficient than that of Crov527, 539, or 563. The relatively low initial T_m values for Crov539 and 563 were however surprising despite the standard growth temperature of 20–25 °C for *Cafeteria* sp. Many repeat proteins contain capping motifs which serve to shield the hydrophobic core from solvent and maintain structural integrity²⁶. Other chemical denaturation studies on the LRR protein pp32 have shown dependence on the capping motifs^{27,28}. The capping motifs from the selected Crov FGxxFN LRR proteins show little homology and this could be a reason for their distinct differences in their thermal stability (Fig. 1c). This becomes even more apparent when looking at the full Crov FGxxFN repeat proteins list (supplementary sequence file 1). The capping regions often (but not always) contain duplicate FN motifs features. The repeat patterns also show some variability, with Crov588 and Crov539 defining two sub-classes, differing foremost in the C-terminal half of the repeat.

Structure determination of Crov588. We were able to obtain well-diffracting crystals of the purified Crov588 protein. We determined its structure using a combination of W-SAD phasing on crystals soaked with an Anderson-Evans polyoxotungstate [TeW₆O₂₄]⁶⁻ (TEW) cluster (Supplementary Figure 1), which is known as an effective phasing tool²⁹. Soaking with tungstate appeared to induce a change in spacegroup (P12₁1 > P2₁2₁2₁) and reduced diffraction to 3.0 Å resolution. Initially, the tungstate structure was solved in spacegroup P2₁2₁2₁. However, it became clear at later stages of building and refinement that P12₁1 is the true spacegroup and that the spacegroup ambiguity was caused by a disorder in the termini due to the binding of clusters at the N/C-terminal interface of the circular structure. The structure of the native data was subsequently solved by molecular replacement using a partial model, and capping motifs at the termini were rebuilt (see methods; Table 1).

The refined structure from native data at 2.4 Å resolution contains two molecules in the asymmetric unit (see Table 1). The two copies superpose poorly with a root-mean-square deviation (RMSD) of 2.06 Å between the equivalent Ca positions of residues 2–680. This is predominantly due to a more twisted reorientation of the terminal capping regions within molecule B. Chain A from the native structure shows the best order. It was therefore chosen as the reference molecule for detailed structural analyses.

Overall structure of Crov588, a novel LRR fold. The Crov588 structure reveals a compact core of 25 regular, 22 residues long LRRs flanked by N- and C-terminal capping modules that build a circular structure. At the N-terminus, four irregular 21–24 residue repeat sequences cap the structure, while a flanking strand extends to contact the C-terminus. The C-terminus is capped by a single 24 residue repeat and flanking strand. To our knowledge, this represents a unique LRR structure as it is the first “fully closed/circular” structure of a native LRR protein, where the N and C-terminal caps interact (Fig. 2). The circular LRR structure is ~ 90 Å in diameter with a central hole of ~ 45 Å.

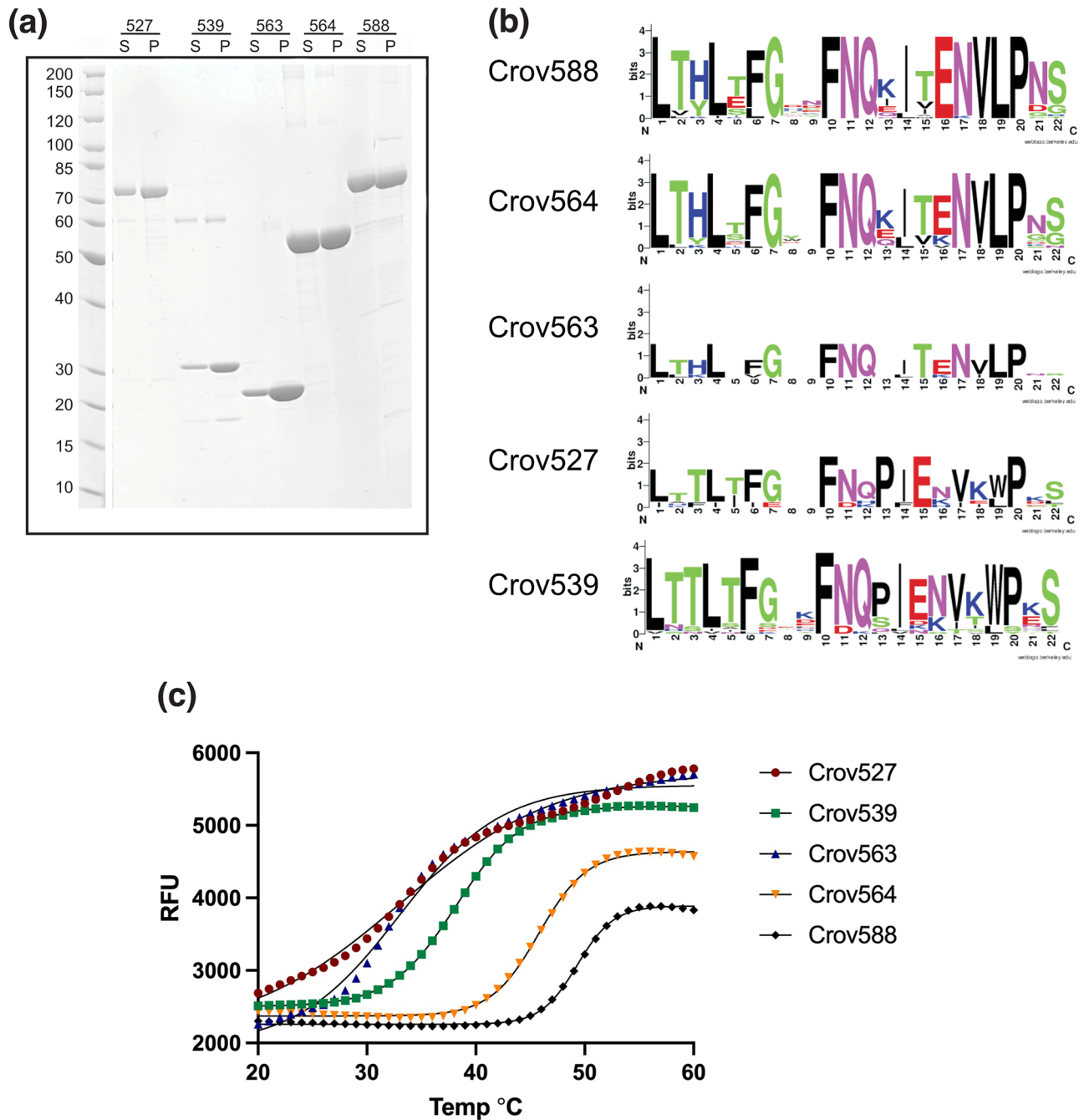


Figure 1. Properties of selected FGxxFN LRR proteins from *Cafeteria roenbergensis* virus. **(a)** SDS gel depicting soluble (S) and pellet (P) fractions in a heat stability assay, following heat denaturation at 97 °C for 5 min. **(b)** Logo plots of the corresponding 22-residue FGxxFN repeats. See supplementary sequence file for a comprehensive repeat annotation for all ~70 Crov FGxxNN repeat proteins. **(c)** Thermofluor measurements of unfolding. Calculated V50 melting temperature values are: Crov527 T_m 30.1 °C, Crov539 T_m 38.2 °C, Crov563 T_m 32.9 °C, Crov564 T_m 45.8 °C, Crov588 T_m 49.3 °C.

Sequence conservation within the Crov588 FGxxFN repeats. The 25 core LRRs are, in comparison to other LRR proteins, almost identical throughout the whole protein and contain a very highly conserved sequence motif LxxLxFGxxFNQxIxENVLPxx (where x is a variable residue) and this appears representative of the first sub-class of Crov FGxxFN repeat proteins (supplementary sequence file 1).

Structural homology between Crov588 and other LRR-containing proteins. In almost all LRR structures studied, the highly conserved segments (LxxLxLxxNxL or LxxLxLxxCxxL) form a short β-strand, which is completely, or almost completely, conserved. This feature positions an asparagine ladder, forming a

Structure PDB code	Crov588 Native (6NYR)	Crov588 Tungstate (6NYS)	Crov539 Native (7ZAL)
Data collection			
Wavelength (Å)	1.0	1.2108	1.0
Space group	P12 ₁ 1	P12 ₁ 1	P22 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	65.4, 172.6, 92.3	65.2, 171.7, 92.0	101.0, 113.4, 236.2
α , β , γ (°)	90, 94.7, 90	90, 90.1, 90	90, 90, 90
Resolution (Å)	49.09–2.43 (2.48–2.43)	46.02–3.1 (3.31–3.1)	49.46–2.73 (2.79–2.73)
<i>R</i> _{merge}	0.1253 (1.14)	0.2419 (1.79)	0.123 (2.85)
<i>R</i> _{p.i.m}	0.0525 (0.48)	0.067 (0.49)	0.034 (0.78)
CC _{1/2}	0.99 (0.67)	0.98 (0.61)	0.99 (0.44)
Mean <i>I</i> / σ <i>I</i>	11.59 (1.44)	15.11 (1.15)	14.4 (1.0)
Completeness (%)	99.56 (99.95)	99.8 (99.92)	99.7 (95.0)
Multiplicity	6.8 (6.5)	13.8 (13.6)	13.8 (13.6)
Refinement			
Resolution (Å)	49.09–2.43	46.02–3.1	49.44–2.73
No. of reflections	76,129 (7384)	36,693 (3654)	68,959 (4858)
<i>R</i> _{work} (%)	0.2204 (0.306)	0.2400 (0.351)	0.2197 (0.360)
<i>R</i> _{free} (%)	0.2568 (0.306)	0.2975 (0.415)	0.2666 (0.377)
No. of atoms			
Protein	11,110	11,086	14,496
Ligand/ion	30	465	0
Water	814	0	304
Mean <i>B</i> values (Å ²)	56.84	116.5	46.59
Protein	57.29	109.49	45.87
Ligand/ion	75.95	284.11	N/A
Water	49.99	N/A	80.85
R.m.s. deviations			
Bond lengths (Å)	0.004	0.005	0.005
Bond angles (°)	1.04	1.03	1.11
Ramachandran analysis (%)			
Preferred	90.56	88.7	94.66
Allowed	9.22	10.78	5.34
Outliers	0.22	0.52	0.00

Table 1. Data collection and refinement statistics.

continuous network of hydrogen bonds between consensus asparagines and neighboring backbone oxygens³⁰. In FGxxFN repeat proteins, the conservation of this motif has deteriorated to only the first six residues (**LxxLx**F-GxxFNxxLxxxxLPxx) before accommodation of the restricted peptide sequence that allows the double phenylalanine spine. The two residues between the FG and FN motifs are the most variable in all repeat proteins from Crov, its virophage mavirus, and *Dictyostelium*²⁴. The asparagine ladder and subsequent continuous network of hydrogen bonds between repeats are conserved but have been shifted from position 9 in the regular LRR to position 12 in FGxxFN repeats. This stacking arrangement appears highly optimized and likely reflects an important protein folding constraint. In the archetypical LRR structure of ribonuclease inhibitor (pdb 1DFJ), the convex region of the repeat forms an α -helix. In other structures, such as the InlB structure (pdb 1D0B), loops or 3_{10} -helices combined with highly ordered water molecules form a complex hydrogen bonded spine along the convex face between the adjacent LRR repeats³¹. Similarly, in the Crov588 structure, tightly bound water molecules participate in hydrogen bonding and stabilize the irregular loop structure along the convex surface of the structure at residue position 14 between the repeats.

An unusual double phenylalanine spine. A single spine formed by consecutive consensus phenylalanine residues has been observed in several LRR structures, such as the Nogo receptor and TLR3^{32,33} (Fig. 3). However, the spine formed by the FGxxFN repeat is subtly different in its positioning and the planar stacking arrangement of the highly conserved phenylalanine residues that make up the repetitive hydrophobic core. This constrained arrangement is possible only because the conserved glycine in the FG motif can adopt appropriate PHI/PSI angles in the bottom left quadrant of the Ramachandran plot. To our knowledge, the only other structure containing repeats with a double phenylalanine spine is the recently published structure of LGI1³⁴. In this case, the phenylalanine spine is packed in a less restricted manner, and the repeat follows the more conventional

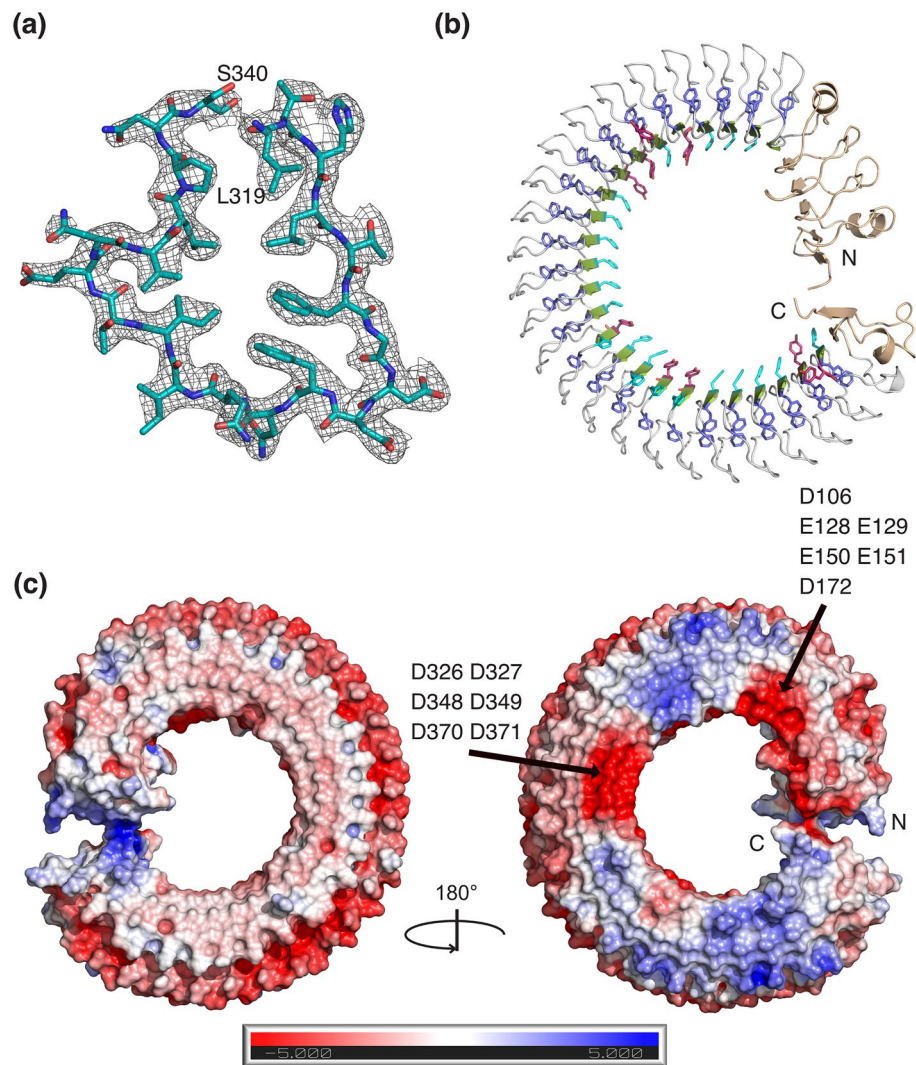


Figure 2. Structure of the Crov588 protein. **(a)** A zoomed view of an individual repeat is presented with 2Fo-Fc density contoured at 1.0 σ . **(b)** Ribbon representation of the Crov588 structure. Highlighted are N and C-terminal capping repeats (wheat), the double phenylalanine spine (slate) histidine repeats (cyan) and tyrosine repeats (magenta). **(c)** Charged surface features of the Crov588 structure calculated with APBS (Pymol), highlighting the pronounced negatively charged patches in the N-terminal and middle of the structure.

packing arrangement with the asparagine ladder being located at the end of the initial β -strand (Fig. 2a, b). Although the FG-containing peptide in the Crov588 structure is restricted by the packing of neighboring repeats within the protein core, the planar stacking arrangement within the FGxxFN motif is in itself interesting. The intermolecular Potential Energy Surface (PES) of a Phe–Phe interaction calculated by MD simulations shows two minima for antiparallel and parallel stacking arrangements³⁵. The parallel stacking FF2 minimum shows a distance between the two ring centroids of 4.03 Å and a 5.1° angle between the rings³⁵. The internal F-F stacking within the Crov588 FGxxFN repeat is also a parallel stacking mode, albeit with a centroid distance of 3.8 Å and a stacking angle of 34°. The centroid distance between Phe residues in subsequent repeats is 5.6 Å with a stacking angle of 62°.

Crov564 and Crov563 have repeat sequences that are very similar to Crov588 (Fig. 1b). Therefore, we assume that the Crov588 is also representative of these other two family members.

Surface features of Crov588. The Crov588 repeat shows a preference for histidine or tyrosine at position three, and this feature is conserved in Crov564 and Crov563, but not in Crov527 and Crov539, where this position is replaced with threonine or serine (Fig. 1b). These surface histidine patches proved to be perfect binding sites for the Anderson-Evans polyoxotungstate [TeW₆O₂₄]⁶⁻ (TEW) cluster used to solve the structure; and indeed, 8 of the 14 cluster sites were found on these histidine patches (Supplementary Figure 1). The concave surface of LRR family proteins is commonly used as a site for protein–protein interactions. In this repeat, however, variable surface features are restricted due to the conserved Histidine ladder and the FG motif. Only residue 5 on the concave surface of each repeat comprises sequence variability and shows a preference for small

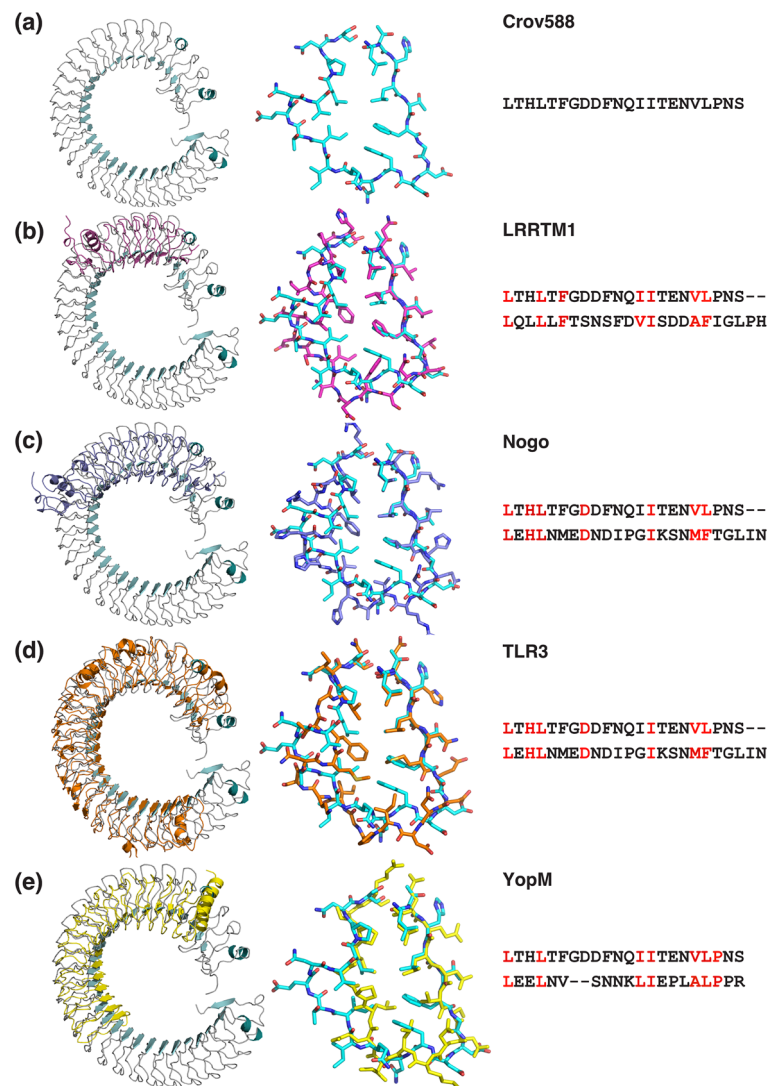


Figure 3. Overlay of Crov588 with structurally similar LRR-containing proteins. Side-by-side view of structures and aligned LRR motifs for (a) Crov588, (b) LRRTM1 (5Y30), (c) Nogo Receptor (1OZN), (d) TLR3 (1ZIW) and (e) YopM (1JL5).

polar residues S/T or negatively charged E. The most variable residues in the repeat are at positions 8–9 in between the FG and FN elements along the edge of the concave surface. Here, the charged residues EE in repeats 2–3 (Glu¹²⁸, Glu¹²⁹, Glu¹⁵⁰, Glu¹⁵¹) and DD in repeats 11–13 (Asp³²⁶, Asp³²⁷, Asp³⁴⁸, Asp³⁴⁹, Asp³⁷⁰, Asp³⁷¹) form two clusters of negative charges along the edge of the concave surface of the Crov588 structure.

Overall structure of Crov539. Crov539 and Crov527 represent the second sub-class of the here analyzed repeat proteins with an LxxLxFGxxFNQPIExVxW/LPxx consensus (Fig. 1b, supplementary sequence file). Differences to Crov588 are the lack of histidine at the N-termini of the repeats and the occurrence of a proline in the middle and of tryptophan near the C-terminus of the repeat. We crystallized Crov539 and solved its structure by molecular replacement using a homology model based upon the Crov588 structure. Crov539 has a short and curved core comprising 12 regularly shaped repeats (Fig. 4). This core is flanked by a compact 15-residue N-terminal capping module. At the C-terminus, a unique domain swapping arrangement occurs where the last repeat(s) (261–290) of the first molecule extends across and caps another molecule to form a dimer. Dimerization is also evident in gel filtration experiments. To our knowledge, such an arrangement has not been reported for other LRR structures. For both the Crov527 and Crov539 repeats, the packing within the hydrophobic core is altered through a substitution of the LP motif for WP at positions 19–20 (Fig. 1b). This appears to be restricted to those repeat proteins that also contain a restraining proline at position 13. The restriction between prolines 13 and 20 and the inclusion of a bulky tryptophan residue gives rise to a reversal of the hydrophobic core packing at residues 17 and 18. A search of the DALI database³⁶ highlighted our Crov588 structure (RMSD 2.1 Å) and the structure of human fibromodulin (5mx0, RMSD 2.9 Å)³⁷ as the closest structural homologs. While fibromodu-

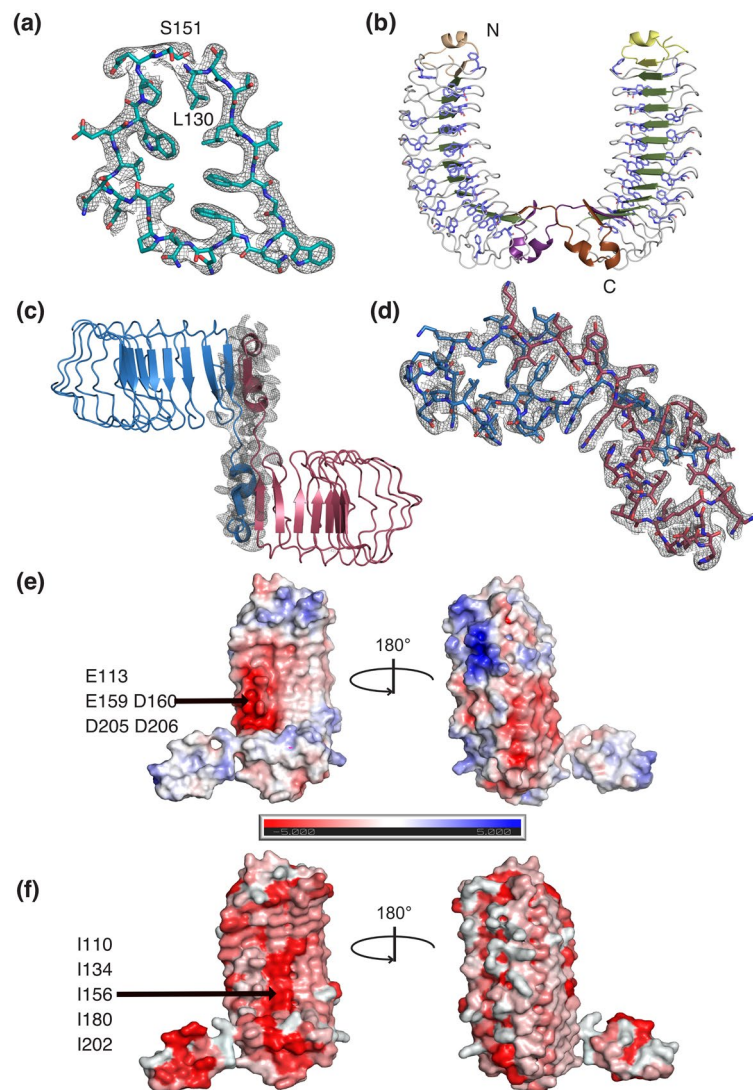


Figure 4. Structure of the Crov539 FGxxFN LRR protein. **(a)** Zoomed view of an individual repeat is presented with 2Fo-Fc density contoured at 1.0σ . **(b)** Ribbon representation of the Crov539 structure. Highlighted are N and C-terminal capping repeats (wheat-purple, yellow-brown), the phenylalanine/tryptophan spine (Slate). **(c)** Representation of the Crov563 dimer with individual chains colored blue and raspberry, 2Fo-Fc density at the dimer interface is contoured at 1.0σ . **(d)** A close-up of the motif swapped dimer region illustrated in sticks with 2Fo-Fc density at the dimer interface is contoured at 1.0σ . **(e)** Charged surface features of the Crov539 structure calculated with APBS, highlighting the pronounced negatively charged patch in the central curvature of the structure. **(f)** Hydrophobic surface features of the Crov539 structure colored according to the Eisenberg hydrophobicity scale with red being most hydrophobic.

lin is of similar size and curvature to Crov539, its repeat sequence and structure show significant differences in hydrophobic core packing (Fig. 5).

Surface features of Crov539. Similarly, to the Crov588 structure, tightly bound water molecules participate in hydrogen bonding and stabilize the irregular loop structure along the convex surface of the structure.

Unlike Crov588, Crov539 lacks the histidine ladder derived from position 3 of the repeat as this residue is almost exclusively threonine. Residue 5 on the concave surface of the repeat is uncharged and either threonine at the N-terminus of the structure or isoleucine (Ile¹¹⁰, Ile¹³⁴, Ile¹⁵⁶, Ile¹⁸⁰, Ile²⁰²) towards the C-terminal region, creating a hydrophobic patch at the center of the lower concave surface (Fig. 4f). Adjacent to this hydrophobic patch at the edge of the concave surface is a negatively charged surface generated from the variable residues (Glu¹¹³, Glu¹⁵⁹, Asp¹⁶⁰, Asp²⁰⁵, Asp²⁰⁶) at positions 8 and 9 in the repeat, directly following the FG motif (Fig. 4e).

Analysis of the Crov588 and Crov539 repeat geometry. Given the completely circular architecture and the highly repetitive repeat structure of Crov588, we were intrigued to compare its geometry to other clas-

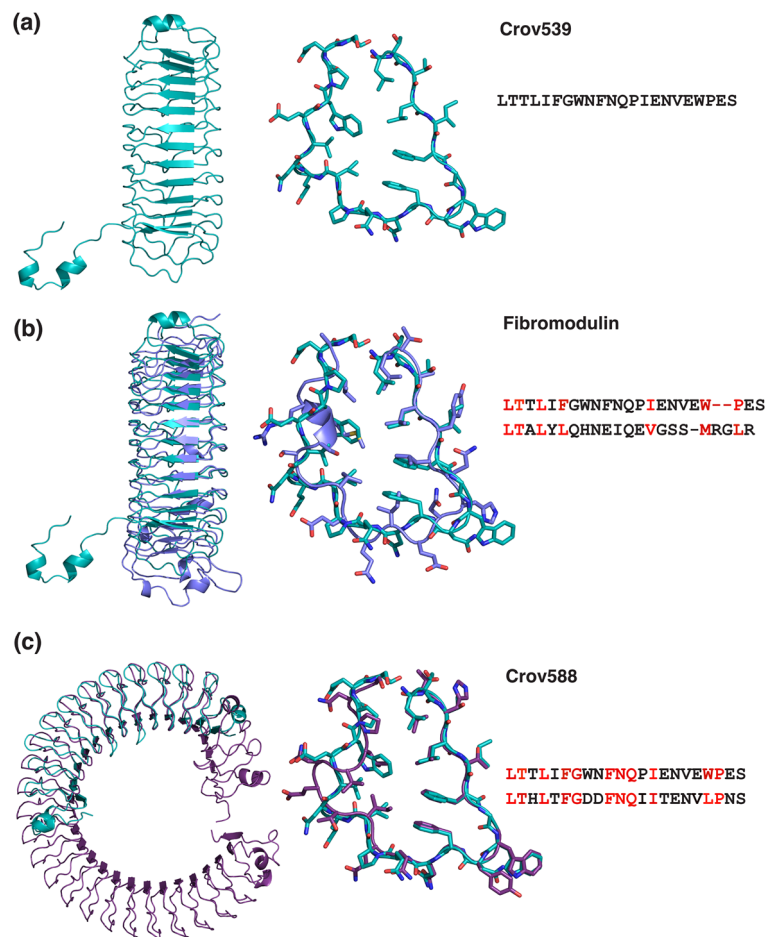


Figure 5. Overlay of structures and aligned LRR motifs for (a) Crov539, (b) Fibromodulin (5MX0) and (c) Crov588 (6NYR).

sical LRR proteins. The curvature, twist, and lateral bending angles for each repeat relative to its predecessor can be calculated using the ANGULATOR webserver (<http://bragi2.helmholtz-hzi.de/Angulator/>)³⁸ (Supplementary Figure 2). We used this to compare the Crov588 structure with its structural homologs from the PDB; TLR3 (1ZIW) and YopM (1JL5)^{32,39} as determined from searching the DALI database³⁶ (Fig. 3). Additionally, we included the archetypical ribonuclease inhibitor structure (2BNH)⁴⁰. The central diameter of ribonuclease inhibitor is roughly half that of Crov588 at only 20 Å. Therefore, despite the circular structure of Crov588, its cumulative curvature of 343° (12.2° per repeat) is not as pronounced as that of ribonuclease inhibitor with a cumulative curvature of 289° (19.3° per repeat) (Supplementary Figure 2). In comparison, Crov539 illustrated a similar curvature to Crov588 with a cumulative value of 129° (11.7° per repeat).

Discussion

FNIP repeat proteins are a large and still growing class of leucine-rich repeat (LRR) proteins for which, so far, no experimental structures were available. We now report the first structures, namely those of Crov588 and Crov537. Their perhaps most striking structural feature is a fully buried double phenylalanine spine that originates from their diagnostic FGxxFN motifs. The two repeat proteins are related but show some differences in their repeat consensus (see Fig. 1b) and the packing of their hydrophobic core. The *Cafeteria roenbergensis* virus encodes ~70 FNIP repeat proteins (see Supplementary sequence file). The majority of them follow either the Crov588 repeat consensus (LxxLxFGxxFNQxIxENVLPxx; bold residues forming the hydrophobic core) or the Crov539 repeat consensus (LxxLxFGxxFNxPxIENxWPxx). The two structures are, therefore, ideal templates for modeling structures of the remaining Crov repeat proteins.

The FNIP LRR protein class was named after the occurrence of the FNx₇IP motif. The C-terminal IP sub-motif is, however, not conserved but substituted by LP in Crov588 and by either WP or LP in the Crov539 class. The preceding FG dipeptide shows, in fact, better conservation. ‘FGxxFN LRR proteins’ might therefore be a more accurate denomination also because it considers the structural relevance of the FGxxFN motif.

We searched for further FGxxFN LRR proteins using hidden Markov modeling tools (<https://www.ebi.ac.uk/Tools/hmmer/>) and the Crov588 sequence as an input. This returned, amongst others, 557 sequences from *Dictyostelium* (mostly with C-terminal “LP” sub-motifs), 104 sequences from the brown alga *Endocarpus* (mostly with

C-terminal WP sub-motifs), and 51 from Megavirus chiliensis (predominantly with C-terminal IP sub-motifs). Recently, 82 Megavirus baoshan proteins have been reported that feature a similar but shorter (21 instead of 22 residues) VxxLxFGxxFNQxIxxxIPxx repeat consensus²³.

Obviously, these species have exploited the variability within a single FGxxFN repeat to evolve a family of proteins that may recognize a large number of potential ligand molecules. They are not quite as diverse as the Variable Lymphocyte Receptors (VLRs) of jawless fish, which are generated by somatic gene rearrangements and have an antibody-like adaptive immune function⁴¹. Nevertheless, it is astounding how large the FGxxFN LRR repertoire of individual species can be.

So far, little is known about the functions of FGxxFN LRR proteins—though their diversity suggests also functional diversity and possibly cross-species differences. For example, many of the Megavirus baoshan FGxxFN LRR domains are fused to an N-terminal F-box domain, indicating that they function in the ubiquitin–proteasome pathway²³ and that the repeat domain might select proteins for degradation. In contrast, numerous FGxxFN LRR domains from *Polysphondylium pallidum* are linked to kinase domains and thus might confer specificity in protein phosphorylation. The Cafeteria repeat proteins, on the other hand, appear not associated with any enzymatic activities. Instead, they might have structural roles in the cytoplasmic virus replication factories⁴², or bind cellular factors as stoichiometric agonists or antagonists to subvert the host's anti-viral response.

The Crov588 structure is, to our knowledge, the first structure of a naturally occurring closed circle of LRR motifs. This was previously only shown for a designed LRR structure with 20 LRRs created to explore engineered binding scaffolds⁴³. The concave (inner) surface of LRR family proteins is commonly used as a site for protein–protein interactions³⁰; although in some cases, exterior surfaces are used as the predominant ligand binding site, as in the case for TLR3 and CD14^{44,45}. FGxxFN LRR proteins probably also follow a concave surface-binding mode since positions 8–9 (at the FGxxFN motif) on the edge of the repeat are the most variable ones.

Our initial reason for studying the Crov repeat proteins was to investigate the packing arrangement of phenylalanine residues with respect to the conserved FG dipeptide motif within the repeat sequence. Similar motifs are commonly found in the intrinsically disordered repeat regions of nucleoporins. The interactions of the different types of FG motifs (FxFG, PxFG, SAFG, GLFG, and FG)^{13,46} within the nuclear pore are dynamic and likely to include mixtures of stacking arrangements such as staggered stacking, parallel in plane, tilted, edge-ring-face, and cogwheel, which have all been described in bioinformatic analyses of protein structures^{47–50}. More recently, the structure of a GFGNFGTS peptide trapped in a cross- β arrangement has been determined⁵¹. Here the peptide stacks as segments of kinked β -sheets that pair into protofilaments with layers of phenylalanine interactions. The layering within this structure restrains the distance between the Phe residues because of the β -sheet formation, while the shortest van der Waals contact is 3.5 Å. The phenylalanine residues stack here in a parallel staggered arrangement with an angle of 48° between the rings and a centroid distance of 4.8 Å. In the same way, we can consider the double phenylalanine spine of consecutive FGxxFN repeats simply as another parallel stacking of FG motifs, albeit within the hydrophobic core of a soluble protein and not an assembled fiber. At present, it is unclear how representative this arrangement for cohesive interactions within intrinsically disordered FG repeat domains is; however, the Crov structures illustrate one possible mode of interaction.

Methods

Protein expression and purification. We expressed and purified codon-optimized (for *E. coli*) versions of Crov527, Crov539, Crov563, Crov564, and Crov588 as His₁₄-bdSUMO or H14-zz-bdSumo-Crov fusion proteins. The His-tagged proteins were purified using Ni(II) chelate beads, using extensive washing in buffer A (50 mM Tris pH8.0, 300 mM NaCl, 10 mM Imidazole, 1 mM DTT), high salt buffer B (50 mM Tris pH8.0, 1 M NaCl, 10 mM Imidazole, 1 mM DTT), ATP wash buffer C (10 mM Tris pH8.0, 100 mM KCl, 10 mM Imidazole, 5 mM MgCl₂, 1.5 mM ATP, 1 mM DTT), low salt buffer D (10% Glycerol, 10 mM Imidazole, 1 mM DTT). Elution was by on-column-tag-cleavage with a SUMO tag-cleaving protease in elution buffer E (50 mM Tris pH8.0, 300 mM NaCl, 5 mM Imidazole, 10% glycerol, 1 mM DTT) as described before^{52,53}. We were able to crystallize the Crov539, Crov588 and Crov564 proteins. However, crystals of Crov564 observed in HTS trials were not reproducible.

Differential scanning fluorometry (DSF/Thermofluor). Proteins were diluted to 1 mg/ml in 20 μ l 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 1 \times SYPRO Orange (Life Technologies). Experiments were performed in a Hard-Shell® 96-well plate (Bio-Rad) sealed with transparent MicroSeal “B” Seal (Bio-Rad), using the CFX96 Real-Time System (C1000 Thermal Cycler, Bio-Rad). The samples were incubated for 5 min at 20 °C before the temperature was gradually increased to 95 °C with 1-K increments and 45 s for each incubation step. At the end of each step, the SYPRO Orange fluorescence was measured using the HEX channel. Non-linear fitting of truncated fluorescence data and T_m value determination was carried out using GraphPad Prism.

Crystallization and structure determination of Crov588. Several initial crystallization conditions for Crov588 were identified by high-throughput screening with 100 nL + 100 nL drop volumes using a Crystal Gryphon robot (Art Robbins Instruments) and commercial grid and sparse matrix screens (Hampton Research, Qiagen, Jena Biosciences, Molecular Dimensions). The best crystals were reproduced and grown via hanging drop vapor diffusion by equilibrating a 1.5 μ l drop of protein with 1.5 μ l of reservoir solution. Final crystallization conditions were 0.1 M MES pH 6.0, 20% PEG3000, 0.1 M MgCl₂. Crystals were cryoprotected gradually by soaking in reservoir solution containing increasing amounts of ethylene glycol up to 25–30% for several minutes, before being mounted directly from the drop in a Mitogen loop and flash frozen in liquid nitrogen. For the Anderson – Evans polyoxotungstate [TeW₆O₂₄]⁶⁻ (TEW) derivative, a final concentration of 1 mM was added to the cryoprotectant solution from a 100 mM stock (Jena Biosciences).

Diffraction data were collected at 1 Å wavelength for native data and 1.2108 Å for the Anderson tungstate derivative using a PILATUS 6M (Dectris) detector at the PXII beamline (Swiss Light Source, Villigen, Switzerland) and all data processed with XDS⁵⁴ (for statistics see Table 1). Native Crov588 crystals showed good diffraction to 2.4 Å and belong to the space group P1211 with two molecules in the asymmetric unit and 63% solvent. In contrast, the TEW soaked crystals showed weaker diffraction to 3.0 Å resolution; soaking also appeared to alter the symmetry of the crystals to be space group P2₁2₁. The structure was initially solved in this space group with 1 molecule in the asymmetric unit and 63% solvent using the data from two TEW soaked crystals. Initial phases were obtained by W-SAD using the Phaser-SAD routine running Hyss⁵⁵ which identified several individual heavy atom sites. Tungstate clusters were positioned manually and recycled again using the Phaser-SAD routine, this procedure was repeated several times before 4 clusters could be accurately positioned on the concave surface of the Crov588 repeat.

After density modification with Resolve^{56,57}, the map was easily interpretable and the central 60% portion of the repeat sequence could be automatically traced using Buccaneer⁵⁸. This partial model was then used to solve the higher resolution native dataset by molecular replacement using Molrep⁵⁸. The remainder of the structure was built by multiple cycles of manual rebuilding in COOT⁵⁹ with refinement using both REFMAC5⁶⁰ and Phenix⁵⁶. Final refinement of the structure, converged with R_{cryst} = 22% and R_{free} = 25%. The final Crov588 model was then used to refine against the W-SAD data and at this stage it became clear that the true space group of the tungsten data was also P1211, with a β-angle of 90.13 compared to the β-angle of 94 for the native data. The structure was then resolved by molecular replacement and re-refined. Additional residual tungsten sites were also visible at this stage, final refinement was checked for both structures using the PDB-REDO server⁶¹ (for statistics see Table 1).

Crystallization and structure determination of Crov539. Crystallization conditions were found after setting up a single commercial screen, namely the Hampton Research PEGRx screen. Clusters of thin plate crystals were grown via hanging drop vapor diffusion by equilibrating a 1.5 µl drop of protein with 1.5 µl of reservoir solution (PEGRx2 42) containing 0.1 M Bis-TRIS pH 6.5, 10% PEG10,000, 0.2 M Potassium Sodium Tartrate tetrahydrate. Crystals were cryoprotected gradually by soaking in reservoir solution containing increasing amounts of ethylene glycol up to 25% for several minutes, before being mounted directly from the drop in a Mitogen loop and flash frozen in liquid nitrogen.

Diffraction data were collected at 1 Å wavelength using a EIGER 2 16 M detector at the PXII beamline (Swiss Light Source, Villigen, Switzerland) and all data processed with XDS⁵⁴ (for statistics see Table 1). Native Crov539 crystals showed good diffraction to 2.73 Å and belong to the space group P2₂2₁ with 6 molecules in the asymmetric unit and 63% solvent.

The Crov539 structure was solved by molecular replacement with Phaser⁶² using a homology model generated with Modeller⁶³ from the previously solved Crov588 structure. The structure required multiple cycles of manual rebuilding in COOT⁵⁹ with refinement using both REFMAC5⁶⁰ and Phenix⁵⁶. Final refinement of the structure, converged with R_{cryst} = 21% and R_{free} = 26%.

Structural and bioinformatic analysis. Structural alignments were carried out in cool⁵⁹, centroid distances and angles were calculated using pseudo atoms and all figures prepared using Pymol (Schrödinger).

Data availability

GenBank accession numbers for the Crov FGxxFN LRR proteins are ADO67561.1 (Crov527), ADO67573.1 (Crov539), QCH59643.1 (Crov588). The crystal structures have been deposited in the protein data bank with the following accession codes; Crov588 (6NYR), Anderson–Evans polyoxotungstate [TeW₆O₂₄]⁶⁻ (TEW) derivative (6NYS) and Crov539 (7ZAL).

Received: 11 April 2022; Accepted: 19 September 2022

Published online: 30 September 2022

References

- Fischer, M. G., Allen, M. J., Wilson, W. H. & Suttle, C. A. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. USA* **107**, 19508–19513 (2010).
- Fischer, M. G., Kelly, I., Foster, L. J. & Suttle, C. A. The virion of *Cafeteria roenbergensis* virus (CroV) contains a complex suite of proteins for transcription and DNA repair. *Virology* **466–467**, 82–94 (2014).
- Garza, D. R. & Suttle, C. A. Large double-stranded DNA viruses which cause the lysis of a marine heterotrophic nanoflagellate (*Bodo* sp.) occur in natural marine viral communities. *Aquat. Microb. Ecol.* **9**, 203–210 (1995).
- Colson, P., Gimenez, G., Boyer, M., Fournous, G. & Raoult, D. The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. *PLoS ONE* **6**, e18935 (2011).
- Fischer, M. G. & Hackl, T. Host genome integration and giant virus-induced reactivation of the viroplasm. *Nature* **540**, 288–291 (2016).
- Shukla, A., Chatterjee, A. & Kondabagil, K. The number of genes encoding repeat domain-containing proteins positively correlates with genome size in amoebal giant viruses. *Virus Evol.* **4**, vex039 (2018).
- Suhre, K. Gene and genome duplication in *Acanthamoeba polyphaga* Mimivirus. *J. Virol.* **79**, 14095–14101 (2005).
- Van Etten, J. L. Another really, really big virus. *Viruses* **3**, 32–46 (2011).
- Xiao, C. *et al.* Cryo-EM reconstruction of the *Cafeteria roenbergensis* virus capsid suggests novel assembly pathway for giant viruses. *Sci. Rep.* **7**, 5484 (2017).
- Hackl, T. *et al.* Four high-quality draft genome assemblies of the marine heterotrophic nanoflagellate *Cafeteria roenbergensis*. *Sci. Data* **7**, 29 (2020).
- Hurt, E. C. A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of *Saccharomyces cerevisiae*. *EMBO J.* **7**, 4323–4334 (1988).

12. Wente, S. R., Rout, M. P. & Blobel, G. A new family of yeast nuclear pore complex proteins. *J. Cell Biol.* **119**, 705–723 (1992).
13. Schmidt, H. B. & Görlich, D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *Elife* **4**, e04251 (2015).
14. Schmidt, H. B. & Görlich, D. Transport selectivity of nuclear pores, phase separation, and membraneless organelles. *Trends Biochem. Sci.* **41**, 46–61 (2016).
15. Frey, S., Richter, R. P. & Görlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–817 (2006).
16. Frey, S. & Görlich, D. A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* **130**, 512–523 (2007).
17. Frey, S. & Görlich, D. FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties. *EMBO J.* **28**, 2554–2567 (2009).
18. Kobe, B. & Deisenhofer, J. The leucine-rich repeat: A versatile binding motif. *Trends Biochem. Sci.* **19**, 415–421 (1994).
19. Buchanan, S. G. & Gay, N. J. Structural and functional diversity in the leucine-rich repeat family of proteins. *Prog. Biophys. Mol. Biol.* **65**, 1–44 (1996).
20. Enkhbayar, P., Kamiya, M., Osaki, M., Matsumoto, T. & Matsushima, N. Structural principles of leucine-rich repeat (LRR) proteins. *Proteins* **54**, 394–403 (2004).
21. Kajava, A. V. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **277**, 519–527 (1998).
22. Matsushima, N., Miyashita, H., Mikami, T. & Kuroki, Y. A nested leucine rich repeat (LRR) domain: The precursor of LRRs is a ten or eleven residue motif. *BMC Microbiol.* **10**, 235 (2010).
23. Xia, Y. *et al.* Role of an FNIP repeat domain-containing protein encoded by *Megavirus baoshan* during viral infection. *J. Virol.* **96**, e0081322 (2022).
24. O'Day, D. H., Suhre, K., Myre, M. A., Chatterjee-Chakraborty, M. & Chavez, S. E. Isolation, characterization, and bioinformatic analysis of calmodulin-binding protein cmbB reveals a novel tandem IP22 repeat common to many Dictyostelium and Mimivirus proteins. *Biochem. Biophys. Res. Commun.* **346**, 879–888 (2006).
25. Catalano, A., Luo, W., Wang, Y. & O'Day, D. H. Synthesis and biological activity of peptides equivalent to the IP22 repeat motif found in proteins from Dictyostelium and Mimivirus. *Peptides* **31**, 1799–1805 (2010).
26. Zhang, Y. *et al.* High-pressure NMR and SAXS reveals how capping modulates folding cooperativity of the pp32 leucine-rich repeat protein. *J. Mol. Biol.* **430**, 1336–1349 (2018).
27. Dao, T. P., Majumdar, A. & Barrick, D. Capping motifs stabilize the leucine-rich repeat protein PP32 and rigidify adjacent repeats. *Protein Sci.* **23**, 801–811 (2014).
28. Huyton, T. & Wolberger, C. The crystal structure of the tumor suppressor protein pp32 (Anp32a): Structural insights into Anp32 family of proteins. *Protein Sci.* **16**, 1308–1315 (2007).
29. Bijelic, A. & Rompel, A. T. Good reasons for the use of the tellurium-centered Anderson–Evans polyoxotungstate in protein crystallography. *Acc. Chem. Res.* **50**, 1441–1448 (2017).
30. Kobe, B. & Kajava, A. V. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **11**, 725–732 (2001).
31. Marino, M., Braun, L., Cossart, P. & Ghosh, P. Structure of the InlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. *Mol. Cell* **4**, 1063–1072 (1999).
32. Choe, J., Kelker, M. S. & Wilson, I. A. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* **309**, 581–585 (2005).
33. He, X. L. *et al.* Structure of the Nogo receptor ectodomain: A recognition module implicated in myelin inhibition. *Neuron* **38**, 177–185 (2003).
34. Yamagata, A. *et al.* Structural insights into modulation and selectivity of transsynaptic neuroligin-LRRTM interaction. *Nat. Commun.* **9**, 3964 (2018).
35. Chelli, R., Gervasio, F. L., Procacci, P. & Schettino, V. Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *J. Am. Chem. Soc.* **124**, 6133–6143 (2002).
36. Holm, L. Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
37. Paracuellos, P. *et al.* Structural and functional analysis of two small leucine-rich repeat proteoglycans, fibromodulin and chondroadherin. *Matrix Biol.* **63**, 106–116 (2017).
38. Bublitz, M. *et al.* Crystal structure and standardized geometric analysis of InlJ, a listerial virulence factor and leucine-rich repeat protein with a novel cysteine ladder. *J. Mol. Biol.* **378**, 87–96 (2008).
39. Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. Unusual molecular architecture of the Yersinia pestis cytotoxin YopM: A leucine-rich repeat protein with the shortest repeating unit. *J. Mol. Biol.* **312**, 807–821 (2001).
40. Kobe, B. & Deisenhofer, J. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**, 751–756 (1993).
41. Kim, H. M. *et al.* Structural diversity of the hagfish variable lymphocyte receptors. *J. Biol. Chem.* **282**, 6726–6732 (2007).
42. Fridmann-Sirkis, Y. *et al.* Efficiency in complexity: Composition and dynamic nature of mimivirus replication factories. *J. Virol.* **90**, 10039–10047 (2016).
43. Rämisch, S., Weininger, U., Martinsson, J., Akke, M. & André, I. Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc. Natl. Acad. Sci. USA* **111**, 17875–17880 (2014).
44. Bell, J. K., Askins, J., Hall, P. R., Davies, D. R. & Segal, D. M. The dsRNA binding site of human Toll-like receptor 3. *Proc. Natl. Acad. Sci. USA* **103**, 8792–8797 (2006).
45. Kim, J. I. *et al.* Crystal structure of CD14 and its implications for lipopolysaccharide signaling. *J. Biol. Chem.* **280**, 11347–11351 (2005).
46. Aramburu, I. V. & Lemke, E. A. Floppy but not sloppy: Interaction mechanism of FG-nucleoporins and nuclear transport receptors. *Semin. Cell Dev. Biol.* **68**, 34–41 (2017).
47. Chourasia, M., Sastry, G. M. & Sastry, G. N. Aromatic–Aromatic Interactions Database, A(2)ID: An analysis of aromatic π -networks in proteins. *Int. J. Biol. Macromol.* **48**, 540–552 (2011).
48. Hunter, C. A., Singh, J. & Thornton, J. M. Pi-pi interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **218**, 837–846 (1991).
49. Singh, J. & Thornton, J. M. The Interaction between phenylalanine rings in proteins. *FEBS* **191**, 1–6 (1985).
50. Singh, J. & Thornton, J. M. SIRIUS. An automated method for the analysis of the preferred packing arrangements between protein groups. *J. Mol. Biol.* **211**, 595–615 (1990).
51. Hughes, M. P. *et al.* Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* **359**, 698–701 (2018).
52. Frey, S. & Görlich, D. A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J. Chromatogr. A* **1337**, 95–105 (2014).
53. Frey, S. & Görlich, D. Purification of protein complexes of defined subunit stoichiometry using a set of orthogonal, tag-cleaving proteases. *J. Chromatogr. A* **1337**, 106–115 (2014).
54. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
55. Read, R. J. & McCoy, A. J. Using SAD data in Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 338–344 (2011).

56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
57. Terwilliger, T. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52 (2004).
58. Collaborative Computational Project Number. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
60. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195 (2004).
61. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCr* **1**, 213–220 (2014).
62. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
63. Eswar, N. *et al.* Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.* **31**, 3375–3380 (2003).

Acknowledgements

We wish to thank Ulrich Steuerwald and Jürgen Wawrzinek at the MPINAT crystallization facility for their support. We also thank the Paul Scherer Institute for synchrotron radiation beamtime and the Max-Planck-Gesellschaft and the Deutsche Forschungsgemeinschaft (SFB860, B11 and SFB1190, P08/P14) for funding this work.

Author contributions

T.H. and W.T. cloned, expressed and purified proteins and performed crystallization. T.H. collected X-ray data, solved and refined structures. The manuscript was prepared by T.H., M.J. and D.G. Project management was carried out by D.G. All authors contributed to the final editing and approval of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-20758-8>.

Correspondence and requests for materials should be addressed to T.H. or D.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022