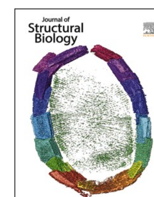
Journal of Structural Biology 215 (2023) 108001

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi



The repetitive structure of DNA clamps: An overlooked protein tandem repeat☆

Paula Nazarena Arrías [a], Alexander Miguel Monzon [b], Damiano Clementel [a], Soroush Mozaffari [a], Damiano Piovesan [a], Andrey V. Kajava [c], Silvio C.E. Tosatto [a,*]

[a] *Department of Biomedical Sciences, University of Padova, via U. Bassi 58/b, 35121 Padova, Italy*
[b] *Department of Information Engineering, University of Padova, via Giovanni Gradenigo 6/B, 35131 Padova, Italy*
[c] *Centre de Recherche en Biologie cellulaire de Montpellier (CRBM), UMR 5237 CNRS, Université Montpellier, 1919 Route de Mende, Cedex 5, 34293 Montpellier, France*

ARTICLE INFO

Keywords:
Tandem repeat protein
Box repeats
Structured tandem repeat

ABSTRACT

Structured tandem repeats proteins (STRPs) are a specific kind of tandem repeat proteins characterized by a modular and repetitive three-dimensional structure arrangement. The majority of STRPs adopt solenoid structures, but with the increasing availability of experimental structures and high-quality predicted structural models, more STRP folds can be characterized. Here, we describe "Box repeats", an overlooked STRP fold present in the DNA sliding clamp processivity factors, which has eluded classification although structural data has been available since the late 1990s. Each Box repeat is a βαβββ module of about 60 residues, which forms a class V "beads-on-a-string" type STRP. The number of repeats present in processivity factors is organism dependent. Monomers of PCNA proteins in both Archaea and Eukarya have 4 repeats, while the monomers of bacterial beta-sliding clamps have 6 repeats. This new repeat fold has been added to the RepeatsDB database, which now provides structural annotation for 66 Box repeat proteins belonging to different organisms, including viruses.

1. Introduction

Tandem repeat proteins (TRPs) are a ubiquitous type of non-globular proteins characterized by repetitive sequence elements arranged in tandem (Kajava and Tosatto, 2018). The length of these repetitive stretches, or "repetitive units", can range from just a few residues to more than a hundred. This distinct organization can generate a modular 3D protein structure made up of repetitions of the same structural unit denominated STRPs (Structural Tandem Repeats Proteins) (Monzon et al., 2023). The repetitive units are defined as the smallest structural building block that make up the repetitive region (Di Domenico et al., 2014). Repetitive regions however, are usually not perfect and can include insertions, i.e. segments that do not belong to the repetitive units, which can be found inside the units or between them.

TPRs are reported to be highly prevalent in eukaryotes, but they are also present in bacteria and archaea, as well as in some viruses (Marcotte et al., 1999; Delucchi et al., 2020; Moore et al., 2008).

Kajava (Kajava, 2012) proposed a classification for TRPs based on their architecture and the length of their units, grouping tandem repeat protein structures into five distinct classes. Class I groups proteins that form crystalline aggregates, in which the repetitive regions have very short repeat units (1 or 2 amino acids). Class II gathers fibrous structures that require interchain interactions for stabilization, and in which the repeats have a length of 3 to 7 residues. Class III is mainly composed of elongated structures (mostly different types of solenoids), with repeats in the range from 5 to 45 amino acids. The units in this class require one another to maintain the structure. Class IV of closed structures, has a similar repeat length to Class III (each repeat averaging 30–60 amino acids in length), but in contrast to elongated repeats, have a fixed number of units due to the circular nature of the structure, although they still need one another to maintain the stability. Lastly, Class V is mainly dominated by structures of different "beads-on-a-string" repeats, in which the units have average lengths of over 50 residues. These longer values in repeat length, makes it possible for each unit to fold into small

☆ This Special Issue is edited by Andrey V. Kajava and Silvio C.E Tosatto, and is published as part of the activities of the MSCA RISE consortia "REFRACT: Repeat protein Function, Refinement,Annotation and Classification of Topologies" funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 823886

* Corresponding author.
E-mail address: silvio.tosatto@unipd.it (S.C.E. Tosatto).

https://doi.org/10.1016/j.jsb.2023.108001
Received 28 April 2023; Received in revised form 12 July 2023; Accepted 16 July 2023
Available online 17 July 2023
1047-8477/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

"globular" domains loosely connected to each other, like the repeats present in, for example, different types of zinc fingers and annexin repeats. However, the increased structural data deposited into the Protein Data Bank has led to the identification of other types or structures that belong to Class V. Included within these structures are spectrin repeats (IPR002017) and sushi repeats (IPR000436), among others (Kajava, 2012).

Identification, classification and annotation of new types of STRPs is a continuous effort. The RepeatsDB database (Paladin et al., 2021) collects experimental STRPs information, and provides both classification (based on Kajava's proposal) and annotation of tandem repeat protein structures.

DNA replication is a crucial process in all domains of life, including some viruses, which is carried out by a multi-protein complex denominated DNA replisome. A quintessential protein of the replisome is the DNA polymerase (DNApol). DNA polymerases synthesize complementary DNA in a 5′ to 3′ direction, and present outstanding fidelity which assures precise replication of the genomes (Lehman et al., 1958). In contraposition to their high fidelity, DNA polymerases, in general, have very low processivity. Processivity refers to the ability of the DNApol to synthesize DNA continuously on a template without dissociating from it, and it is usually measured in terms of the number of nucleotides that are incorporated into the growing DNA molecule per binding event (Zhuang and Ai, 2010). For the complete replication of large DNA genomes, DNApol processivity is not enough. However, this has been made possible due to the existence of other proteins; the processivity factors (Mace and Alberts, 1984). A specific type of processivity factor, known as DNA sliding clamps, are proteins that enhance the processivity of the DNApol by holding it together with the DNA.

Although structural information of DNA sliding clamps has been around since the early 1990s (Kong et al., 1992), they have not been previously extensively classified as STRPs, despite the fact that, at least for some of them, their repeat-containing nature has been suggested before (Parra et al., 2013). Here we describe the repetitive nature of these processivity factors following Kajava's scheme. We offer a comprehensive analysis encompassing multiple facets, such as the length of the repeat regions and units, their structural and sequence similarities, and their distribution across diverse domains of life. By examining these characteristics, our aim is to advance the understanding of the repetitive elements found within clamps.

2. Methods

The RepeatsDB database (Paladin et al., 2021) was used to initially retrieve a dataset of protein structures with "Box" repeats, which had no curation but were present in the database as result of automatic detection of the repeats by the RepeatsDB-lite predictor (Hirsh et al., 2018). In addition, DNA clamp protein sequences with experimental structures were retrieved from UniProt (The UniProt Consortium, 2023) by using the CATH superfamily identifier 3.70.10.10 which contains the fold denominated "box". These initial sets of structures were grouped by protein sequence with the UniProt accession number, and by protein family/domain with Pfam (Mistry et al., 2021). To identify repeated units and insertions, we followed a rigorous biocuration protocol, similar to the one used for the RepeatsDB database. The RepeatsDB-lite (Hirsh et al., 2018) predictions were visually examined by experienced biocurators to ensure the accuracy of boundaries for regions, units, and insertions. In case of any discrepancies, the boundaries were modified. The classification terms were also meticulously curated by the biocurators. In total, we annotated 66 PDB structures, each representing a unique protein sequence and Pfam family/domain (Supplementary Table 1).

Structural similarity between repeat units was calculated by performing pairwise structural alignments with the software TM-align (Zhang and Skolnick, 2005). The sequences of the repeated units were extracted based on the boundaries defined in the previous step. Then a MSA was performed by using the default settings of ClustalW (Thompson et al., 1994) and visualized with the software JalView (Waterhouse et al., 2009). The alignment consensus was calculated using the EMBOSS Cons tool (Madeira et al., 2022). The clustering analysis was performed by using the DBSCAN algorithm implemented in the Scikit-learn python library (Pedregosa et al., 2011; Ester et al., 1996). The FoldSeek server (van Kempen et al., 2023) was used to search the PDB for similar folds to Box repeats. PyMOL Molecular Graphics System was used for protein structure representation (Schrödinger, LLC, 2015). The residue contacts of representative structures were analyzed using RING-PyMOL (Del Conte et al., 2023), a PyMOL plugin, along with the RING software (Clementel et al., 2022). RING-PyMOL facilitated the calculation and visualization of residue interaction networks, providing insights into contacts between units and insertions.

3. Results and discussion

3.1. DNA clamps are composed of structural tandem repeats

3.1.1. Structural architecture and unit definition

Most DNA clamps form ring-shaped oligomers with six domains and have been previously described to have 6-fold pseudosymmetry (Oakley, 2016). The domains are superimposable with each other and they are composed of two α-helices and eight β-strands. However, looking into the arrangement of the secondary structure elements in each domain it is possible to define two repetitions of βαβββ modules. These repetitions, or units, are arranged in tandem (Fig. 1A, Supplementary Figure S4A). Although the structural similarity of the units belonging to the same protein is high, sequentially this is not always the case (Fig. 1B and 1C).

The overall arrangement of the units gives rise to a structure denominated as "box" in CATH/Gene3D database, and all the structures of DNA clamps analyzed (Table 1) belong to the superfamily 3.70.10.10 (Sillitoe et al., 2021).

3.1.2. Insertions

Tandem repeats are usually not perfect, and their modular structure

Table 1
Experimental structures analyzed in this work. *PDB codes and chain IDs grouped by domain of life. All of the entries were added to the RepeatsDB database.*

Domain of Life							
Bacteria			Eukarya		Archaea		Viruses
6degA	3t0pA	6ptvA	3g65A	4ztdA	3hi8A	2hiiA	1b77A
4k74A	4tr6A	2avtA	3a1jC	6qh1A	5a6dA	2hikC	1czdA
6amqC	6dj8A	2awaA	3a1jB	1plqA	6t8hE	1rwzA	2z0lA
6ap4A	5wceA	6d47A	7wp3E	7bupA	6aigA	1iz4A	3hslX
4tr7A	5x06A	4trtA	3k4xA	7ep8A	1ud9A	3aixA	1t6lA
3p16A	4n96A	1vpkA	5tupA	7o1eA	3lx1A	3lx2A	1dmlA
5ah2A	4rkiA		3p91A	2zvwA	2hiiB	3aixB	
6dm6A	4tr8A		4cs5A	7sh2H			
6d46A	6manA		2zvvA	7sh2G			
7rzmA	7evpA		4hk1A	7sh2F			

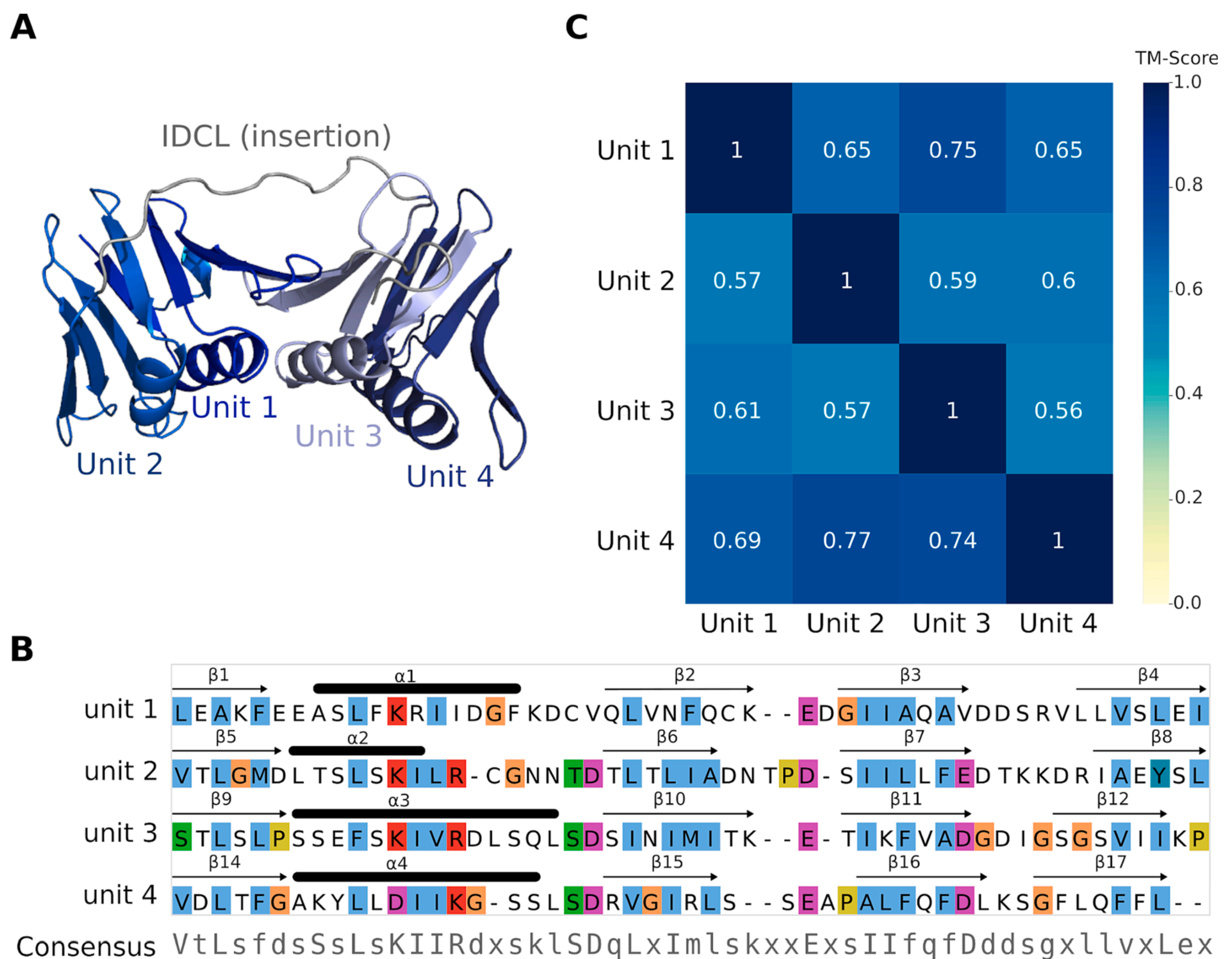


Fig. 1. Repetitive units definition and organization in DNA clamps. (A) Unit organization and insertion in S. cerevisiae PCNA (PDB code: 1plqA). (B) Unit multiple sequence alignment. Arrows represent β-strands and black cylinders represent α-helices. (C) Unit structural similarity matrix (TM-score) in S. cerevisiae PCNA (PDB code: 1plqA).

makes them prone to deletions, duplications and insertions. Insertions are elements that do not belong to the repeat units, and can appear inside units or between them. They usually do not contribute to the stability of the repeat, but they might have a functional role, such as binding (Paladin et al., 2020). Such is the case of DNA clamps, where in every two units there is a non-repetitive segment, composed of a loop (IDCL, inter-domain connecting loop), which connects the last unit of one domain to the first one of the other (Fig. 1A). Residue contact analysis shows that the IDCL insertion has few intra-chain contacts with the repetitive units, supporting the notion that it is not needed for the stability of this kind of repeats. The interaction of DNA clamps with multiple of their partners, such as p21, Fen1 and DNA ligase, among others, are mediated through residues of the IDCL (Pascal et al., 2006; Sakurai et al., 2005; Warbrick et al., 1995).

3.2. The structural organization of DNA sliding clamps varies across the domains of life

3.2.1. Bacterial DNA sliding clamps

The first structural information of these processivity factors came out in 1992 with the determination of *Escherichia coli* beta sliding clamp (coded by gene dnaN) structure (Kong et al., 1992). Up to this day, the same protein (UniProt ID P0A988) has been crystallized multiple times, and the structures of other 23 bacterial sliding clamps have been solved (Table 1 and Supplementary Table 1).

Bacterial sliding clamp monomers have about 378 amino acids. They present three distinct Pfam domains; PF00712, PF02768, PF02767 (Fig. 2A). The structures of these domains are highly similar, showing an average TM-score of 0.75 (Supplementary Fig. 1).

Two monomers assemble in a head-to-tail fashion to form a donut-shaped homodimer that accommodates double stranded DNA (dsDNA) in a central cavity of around 35 Å in diameter (Kong et al., 1992) (Fig. 2B). The dimer interacts with DNA polymerase III and is "loaded" onto DNA by the action of the clamp loader complex.

3.2.2. Eukaryotic PCNA and 9-1-1 clamp

Homo sapiens DNA sliding clamp (UniProt ID P12004), commonly known as proliferating cell nuclear antigen (PCNA), was originally discovered as an antigen reacting with antibodies derived from systemic lupus erythematosus patients' sera (Miyachi et al., 1978). Homologs in other eukaryotic organisms were described afterwards. Although there is virtually no sequence similarity between eukaryotic and bacterial DNA sliding clamps (Acharya et al., 2021), their structures are highly similar. Each PCNA monomer has an average length of 260 residues, and presents two distinct Pfam domains; PF00705 and PF02747. As in the case of their bacterial counterparts, both domains share structural similarity, with an average TM-score of 0.79 (Supplementary Fig. 1), but in contraposition to the DNA sliding clamps of bacteria, PCNA monomers assemble into a trimer (Gulbis et al., 1996). The PCNA trimer increases DNApol processivity by directly interacting with it and holding it to the DNA during chromosome replication.

In addition to PCNA, eukaryotic cells also have another structurally similar DNA sliding clamp (Venclovas and Thelen, 2000) that is involved in replication checkpoint control (S-phase progression, G2/M arrest)

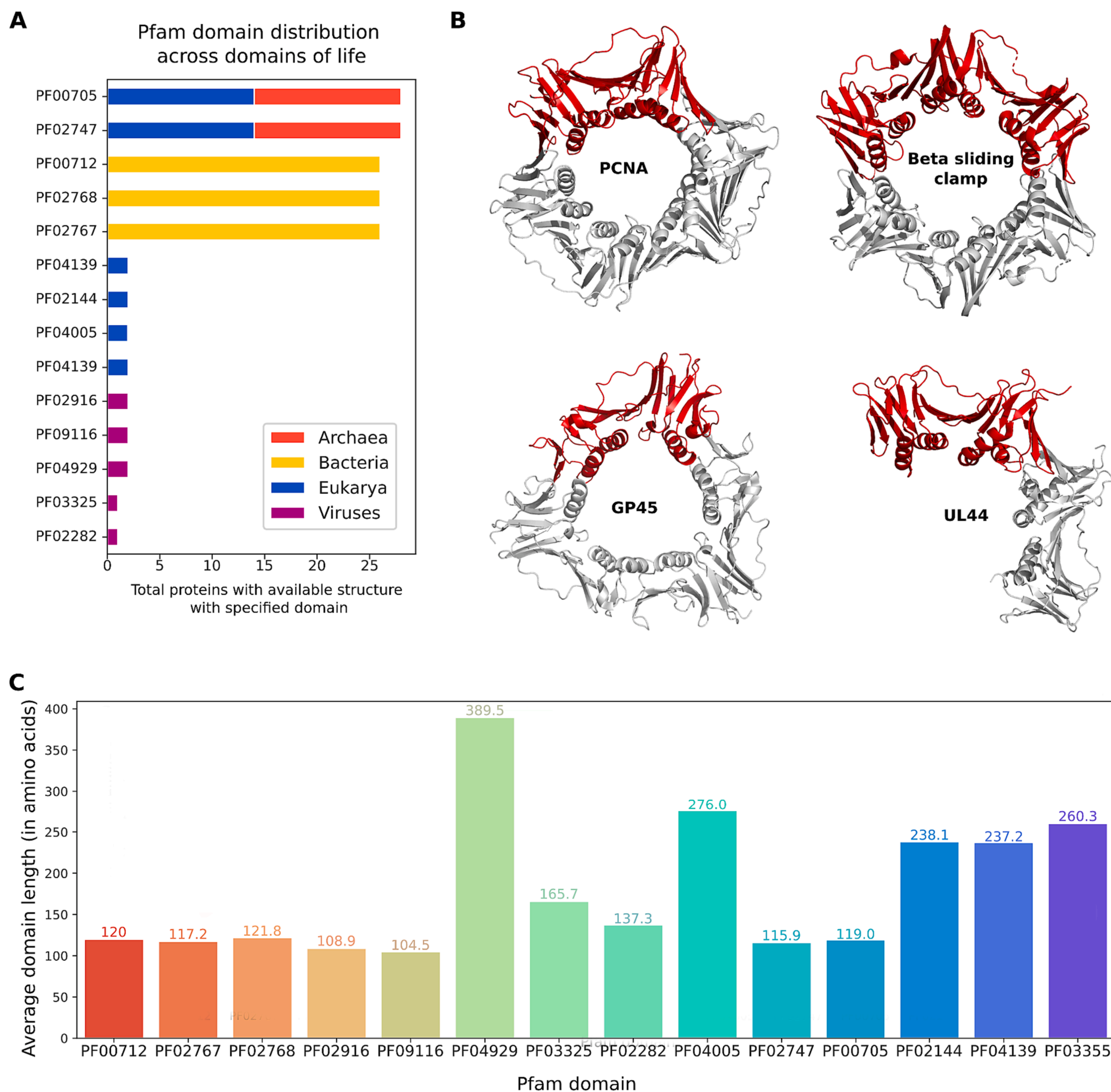


Fig. 2. Structural organization of DNA sliding clamps across the domains of life (A) Pfam domains present in DNA clamps and their distribution across domains of life (including viruses). (B) Oligomerization states of DNA clamps. In red one monomer. (C) Average length of the Pfam domains present in DNA clamps. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and DNA repair that has been denominated the 9-1-1 complex (Parrilla-Castellar et al., 2004). A fundamental difference between PCNA and the 9-1-1 complex is that the latter is a heterotrimer made up from 3 distinct proteins that in humans are called RAD9 (UniProt ID Q99638), RAD1 (UniProt ID O60671) and HUS1 (UniProt ID O60921).

RAD9 of *Homo sapiens* is a protein of 391 amino acids, almost the average length of the bacterial sliding clamps, but it is structurally similar to PCNA, with the difference that Pfam assigns only one domain to the entire protein (PF04139).

RAD1 and HUS1 have lengths of 280 and 282 amino acids respectively, more similar to the average length of the rest of the eukaryotic PCNAs. As with RAD9, Pfam only assigns one domain for each protein, being PF02144 for RAD1 and PF04005 for HUS1. In the yeast *Saccharomyces cerevisiae,* the proteins DDC1 (UniProt ID Q08949), MEC3 (UniProt ID Q02574) and RAD17 (UniProt ID P48581) make up the 9–1-1 clamp. The structures of 20 different proteins have been deposited into the PDB so far (Table 1 and Supplementary Table 1).

3.2.3. Archaeal PCNA

The DNA sliding clamps of Archaea resemble eukaryotic PCNA. Each monomer has an average length of 248 residues and presents the two distinct domains of eukaryotic PCNA; PF00705 and PF02747. In contrast to eukaryotic PCNA, some archaeans present two (*Thermococcus kodakarensis*) and even three (*Sulfolobus solfataricus* and *Sulfurisphaera tokodaii)* PCNA homologs, which assemble as heterotrimers, or even as a heterotetramer which has been proposed can accommodate in its cavity a Holliday junction (Kawai et al., 2011). There are currently 14 different archaeal PCNAs structures in the PDB (Table 1 and Supplementary Table 1).

3.2.4. Viral processivity factors

Some double stranded DNA viruses and bacteriophages also have proteins that act as processivity factors. Currently, the structure of only 6 of these viral proteins (Table 1 and Supplementary Table 1) has been elucidated.

The protein GP45 of two species of bacteriophages; Mosigvirus RB69 (previously known as Escherichia phage RB69) and Tequatrovirus T4 (also known as T4 bacteriophage) have a length of 228 amino acids. Similarly to eukaryotic and archaeal DNA sliding clamps, GP45 forms a homotrimeric ring, and each monomer has two distinct domains; PF02916, PF09116, which are structurally similar although sequentially there is no apparent similarity.

Human alphaherpesvirus 1 (HHV-1) processivity factor, a protein of 488 amino acids denominated UL42 (UniProt ID P10226), differs from the previously mentioned DNA sliding clamps. Structurally, the N-terminal domain of UL42 adopts a similar conformation as archaeal and eukaryotic clamps but the C-terminal domain is predicted to be disordered by MobiDB (Piovesan et al., 2018). Pfam assigns a sole domain for the N-terminal region, PF02282, which is a member of the CL0060 clan (same as the other sliding clamps). In contrast to the other sliding clamps, UL42 apparently binds DNA as a monomer and not as a ring-shaped oligomer (Randell and Coen, 2004).

Similarly to HHV-1 UL42, the processivity factor of another herpesvirus, UL44 (UniProt ID P16790) of human betaherpesvirus 5 (HHV-5), a protein of 433 residues, presents a N-terminal domain (PF03325) that is structurally similar to PCNA. However, the crystal structure suggests that it exerts its function as a C-shaped dimer, in which both monomers interact in a head-to-head fashion, instead of a ring formed by the heat-to-tail interaction of the monomers (Appleton et al., 2004). This situation also applies for HHV-8 protein PF-8 (UniProt ID Q77ZG5) (Baltz et al., 2009) and BMRF1 (UniProt ID P03191) of HHV-4 (Murayama et al., 2009), which have a length of 396 and 404 residues, respectively. The N-terminal domain of both proteins is classified as PF04929.

The length of the monomers defines the number of units within the repetitive region; in the case of eukaryotic, archaeal and viral DNA

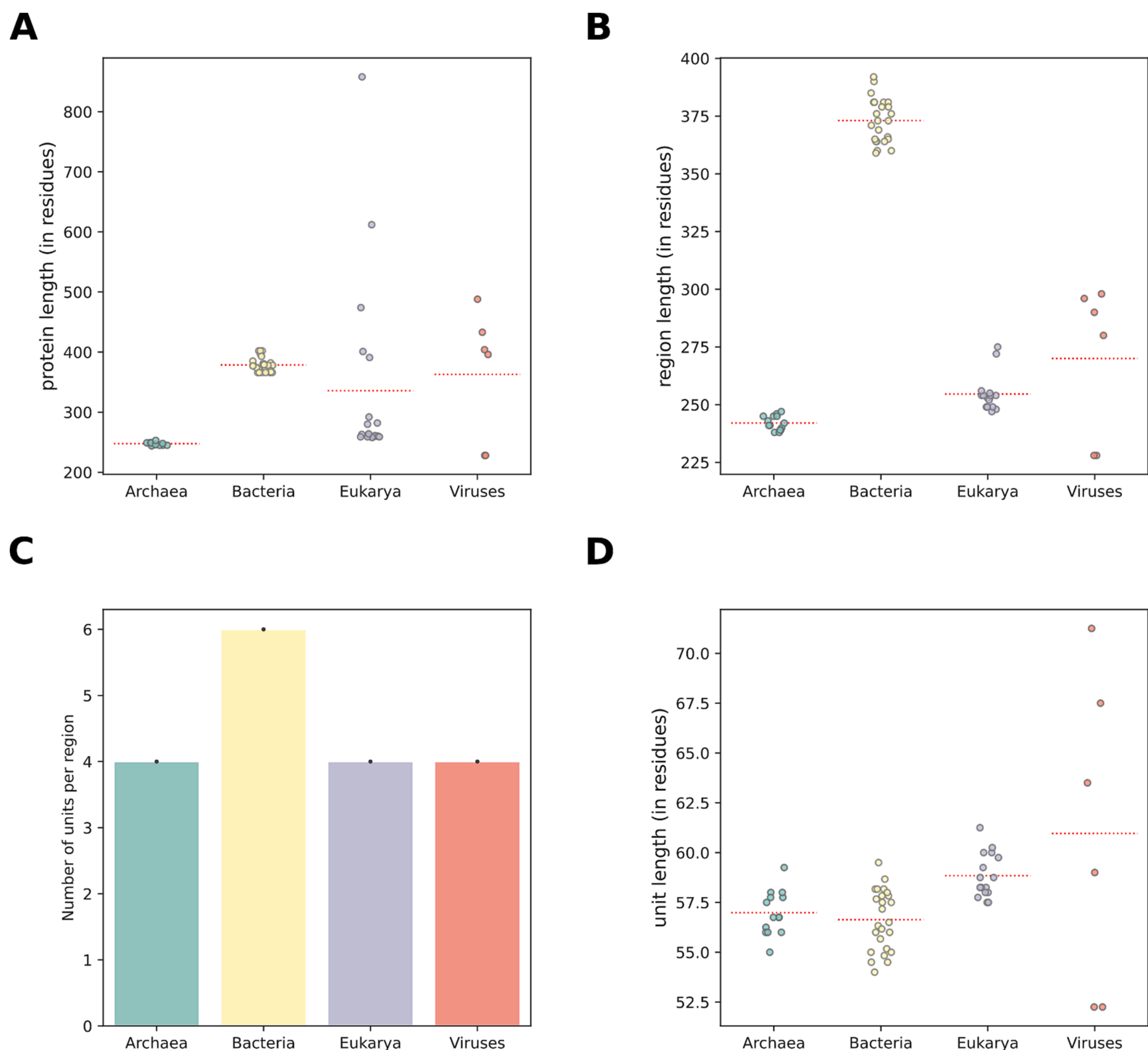


Fig. 3. The repeats present in DNA clamps vary across domains of life. (A) Protein average length in amino acids (B) Repetitive region average length in amino acids (C) Total number of units per region in the different domains of life. (D) Unit length distribution (in residues). In all cases, the red dotted line represents mean values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

clamps, the number of units in each monomer is four, while the bacterial clamps have six (Fig. 3A and 3C). The average length of each unit is quite homogeneous in the case of Archaea, Bacteria and Eukarya but for viruses this number can vary quite a lot (Fig. 3D). However, the number of experimental structures for viral processivity factors available on the PDB is small (Table 1 and Supplementary Table 1).

3.3. Structural Tandem Repeat classification of DNA clamps: "Box" repeats

3.3.1. General classification

Box repeats are about 60 residue long. Given the average unit length of the repeats as well as the tight connection between units, box repeats can be classified as belonging to either Class IV or Class V structures. Kajava's classification is based not only on the length of the repeat, but also on the context under which a repetitive unit of such a length can fold into a stable structure. The individual repeats of the Class IV structures cannot form a stable structure and become stable in the context of closed (ring-like) structures. At the same time, the repeats from Class V are large enough to fold independently into stable domains. Thus, evidence about the ability of a single Box repeat to fold or not to fold into the stable structure becomes essential for the classification. We searched the PDB database to find single structural domains similar to the βαβββ modules of Box repeats. Some proteins, such as the Ragulator complex protein LAMTOR4 (Q0VGL1, PDB code: 5VOK_A) and the polyprotein of Hepacivirus C (H9XGD6, PDB code: 4UOI_A) have single domains with the structures (Rasheed et al., 2019; El Omari et al., 2014) similar to one Box repeat, suggesting that Box repeats have a potential to fold independently. Therefore, Box repeat proteins can be assigned to Class V rather than to Class IV.

3.3.2. RepeatsDB classification

RepeatsDB (Paladin et al., 2021) is a database that provides annotation and classification of STRPs. The classification is based on Kajava's scheme, with the first level of classification being "Class". However, it incorporates 4 extra levels of classification; "Topology" (second level of classification), which is distinguished by the general path of the polypeptide chain and the type of secondary structure in the repeat units, "Fold" (third level of classification), which comprises variants of a certain topology that present structural differences in the number of secondary structure elements, additional structural elements and overall structural arrangement, "Clan" (fourth level of classification) which is a subfold, and lastly "Family" (fifth level of classification), which groups proteins with a common ancestor based on sequence similarity.

Within this classification, Box repeats comprise a unique fold within the Alpha/beta beads topology of Class V. Structural clustering analysis of the regions employing DBSCAN shows that the box repeats from bacteriophages form a separate cluster in the range of 73%-76% of structural similarity and hence we propose they form a separate clan (Supplementary Figure S2). The level of family has not yet been implemented into the current version of the database, and thus remains unassigned for box repeats.

Eukaryotic PCNA exon arrangement does not exhibit the periodicity of their repeat units.

It has been suggested that in eukaryotes, repeat segments could correspond to exons, thus allowing their easy duplication and shuffling (Schaper and Anisimova, 2015). The instances in which the structural symmetry of the tandem repeat regions is also seen in their exon arrangement supports the notion that STRPs can evolve through duplication of their coding exons. Such is the case of some repeat types, such as ankyrins and leucine-rich repeats (Paladin et al., 2020).

We examined this periodicity in the exon arrangement for the eukaryotic DNA clamps with available structure, and employed the pipeline developed by Paladin and collaborators (Paladin et al., 2020) to produce their "repeat/exon plots", which allow the visualization of the alignment between the structural units and exons.

We found that for eukaryotic clamps there is not a uniform pattern. For example, both *Saccharomyces cerevisiae* PCNA (P15873) and the three proteins that compose its 9-1-1 complex are encoded by a single exon (Supplementary Fig. S3A), while *Homo sapiens* PCNA (P12004) is encoded by 6, but shows a complex pattern (Supplementary Fig. S3B). *Arabidopsis thaliana* PCNA1 (Q9M7Q7) and PCNA2 (Q9ZW35) are, however, encoded by 4 exons, but these do not correspond to the structural units, just partial parts of them.

4. Conclusion

Here we have provided a classification for the structural tandem repeats present in the DNA clamp processivity factors within Kajava's scheme. We have manually annotated 66 proteins that contain box repeats from different organisms, including DNA viruses, for some of which there are more than 50 structures available. This new data and structural description will help to increase and improve the coverage of the "Box" repeats in the RepeatsDB database, as well as improve the identification and study of these proteins.

CRediT authorship contribution statement

Paula Nazarena Arrías: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Methodology, Data curation. **Alexander Miguel Monzon:** Supervision, Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Damiano Clementel:** Formal analysis, Software. **Soroush Mozaffari:** Formal analysis. **Damiano Piovesan:** Formal analysis. **Andrey V. Kajava:** Formal analysis, Writing – original draft. **Silvio C.E. Tosatto:** Supervision, Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 823886 (H2020 MSCA-RISE "REFRACT") and based upon work from COST Action ML4NGP, CA21160, supported by COST (European Cooperation in Science and Technology).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2023.108001.

References

Acharya, S., Dahal, A., Bhattarai, H.K., 2021. Evolution and origin of sliding clamp in bacteria, archaea and eukarya. PLoS One 16 (8), e0241093.

Appleton, B.A., Loregian, A., Filman, D.J., Coen, D.M., Hogle, J.M., 2004. The cytomegalovirus DNA polymerase subunit UL44 forms a C clamp-shaped dimer. Mol. Cell 15 (2), 233–244. https://doi.org/10.1016/j.molcel.2004.06.018.

Baltz, J.L., Filman, D.J., Ciustea, M., Silverman, J.E.Y., Lautenschlager, C.L., Coen, D.M., Ricciardi, R.P., Hogle, J.M., 2009. The crystal structure of PF-8, the DNA polymerase accessory subunit from Kaposi's sarcoma-associated herpesvirus. J. Virol. 83 (23), 12215–12228. https://doi.org/10.1128/JVI.01158-09.

Clementel, D., Del Conte, A., Monzon, A.M., Camagni, G.F., Minervini, G., Piovesan, D., Tosatto, S.C.E., 2022. RING 3.0: Fast generation of probabilistic residue interaction

networks from structural ensembles. Nucleic Acids Res. 50 (W1), W651–W656. https://doi.org/10.1093/nar/gkac365.

Del Conte, A., Monzon, A.M., Clementel, D., Camagni, G.F., Minervini, G., Tosatto, S.C.E., Piovesan, D., 2023. RING-PyMOL: Residue interaction networks of structural ensembles and molecular dynamics. Bioinformatics 39 (5), btad260. https://doi.org/10.1093/bioinformatics/btad260.

Delucchi, M., Schaper, E., Sachenkova, O., Elofsson, A., Anisimova, M., 2020. A New Census of Protein Tandem Repeats and Their Relationship with Intrinsic Disorder. Genes 11 (4), 407. https://doi.org/10.3390/genes11040407.

Di Domenico, T., Potenza, E., Walsh, I., Parra, R.G., Giollo, M., Minervini, G., Piovesan, D., Ihsan, A., Ferrari, C., Kajava, A.V., Tosatto, S.C.E., 2014. RepeatsDB: A database of tandem repeat protein structures. Nucleic Acids Res. 42 (Database issue), D352–D357. https://doi.org/10.1093/nar/gkt1175.

El Omari, K., Iourin, O., Kadlec, J., Fearn, R., Hall, D.R., Harlos, K., Grimes, J.M., Stuart, D.I., 2014. Pushing the limits of sulfur SAD phasing: De novo structure solution of the N-terminal domain of the ectodomain of HCV E1. Acta Crystallograph. Sect. D Biol. Crystallogr. 70 (Pt 8), 2197–2203. https://doi.org/10.1107/S139900471401339X.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining.

Gulbis, J.M., Kelman, Z., Hurwitz, J., O'Donnell, M., Kuriyan, J., 1996. Structure of the C-Terminal Region of p21WAF1/CIP1 Complexed with Human PCNA. Cell 87 (2), 297–306. https://doi.org/10.1016/S0092-8674(00)81347-1.

Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., 2018. RepeatsDB-lite: A web server for unit annotation of tandem repeat proteins. Nucleic Acids Res. 46 (W1), W402–W407. https://doi.org/10.1093/nar/gky360.

Kajava, A.V., 2012. Tandem repeats in proteins: From sequence to structure. J. Struct. Biol. 179 (3), 279–288. https://doi.org/10.1016/j.jsb.2011.08.009.

Kajava, A.V., Tosatto, S.C.E., 2018. Editorial for special issue "Proteins with tandem repeats: Sequences, structures and functions"☆. J. Struct. Biol. 201 (2), 86–87. https://doi.org/10.1016/j.jsb.2017.12.011.

Kawai, A., Hashimoto, H., Higuchi, S., Tsunoda, M., Sato, M., Nakamura, K.T., Miyamoto, S., 2011. A novel heterotetrameric structure of the crenarchaeal PCNA2-PCNA3 complex. J. Struct. Biol. 174 (3), 443–450. https://doi.org/10.1016/j.jsb.2011.02.006.

Kong, X.P., Onrust, R., O'Donnell, M., Kuriyan, J., 1992. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: A sliding DNA clamp. Cell 69 (3), 425–437. https://doi.org/10.1016/0092-8674(92)90445-i.

Lehman, I.R., Bessman, M.J., Simms, E.S., Kornberg, A., 1958. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. J. Biol. Chem. 233 (1), 163–170.

Mace, D.C., Alberts, B.M., 1984. T4 DNA polymerase. Rates and processivity on single-stranded DNA templates. J. Mol. Biol. 177 (2), 295–311. https://doi.org/10.1016/0022-2836(84)90458-3.

Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., Lopez, R., 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 50 (W1), W276–W279. https://doi.org/10.1093/nar/gkac240.

Marcotte, E.M., Pellegrini, M., Yeates, T.O., Eisenberg, D., 1999. A census of protein repeats. J. Mol. Biol. 293 (1), 151–160. https://doi.org/10.1006/jmbi.1999.3136.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: The protein families database in 2021. Nucleic Acids Res. 49 (D1), D412–D419. https://doi.org/10.1093/nar/gkaa913.

Miyachi, K., Fritzler, M.J., Tan, E.M., 1978. Autoantibody to a nuclear antigen in proliferating cells. Journal of Immunology (Baltimore, Md.: 1950), 121(6), 2228–2234.

Monzon, A.M., Arrías, P.N., Elofsson, A., Mier, P., Andrade-Navarro, M.A., Bevilacqua, M., Clementel, D., Bateman, A., Hirsh, L., Fornasari, M.S., Parisi, G., Piovesan, D., Kajava, A.V., & Tosatto, S.C.E. (2023). A STRiP-ed definition of Structured Tandem Repeats in Proteins. Manuscript Submitted for Publication.

Moore, A.D., Björklund, A.K., Ekman, D., Bornberg-Bauer, E., Elofsson, A., 2008. Arrangements in the modular evolution of proteins. Trends Biochem. Sci. 33 (9), 444–451. https://doi.org/10.1016/j.tibs.2008.05.008.

Murayama, K., Nakayama, S., Kato-Murayama, M., Akasaka, R., Ohbayashi, N., Kamewari-Hayami, Y., Terada, T., Shirouzu, M., Tsurumi, T., Yokoyama, S., 2009. Crystal structure of epstein-barr virus DNA polymerase processivity factor BMRF1. J. Biol. Chem. 284 (51), 35896–35905. https://doi.org/10.1074/jbc.M109.051581.

Oakley, A.J., 2016. Dynamics of Open DNA Sliding Clamps. PLoS One 11 (5), e0154899.

Paladin, L., Necci, M., Piovesan, D., Mier, P., Andrade-Navarro, M.A., Tosatto, S.C.E., 2020. A novel approach to investigate the evolution of structured tandem repeat protein families by exon duplication. J. Struct. Biol. 212 (2), 107608 https://doi.org/10.1016/j.jsb.2020.107608.

Paladin, L., Bevilacqua, M., Errigo, S., Piovesan, D., Mičetić, I., Necci, M., Monzon, A.M., Fabre, M.L., Lopez, J.L., Nilsson, J.F., Rios, J., Menna, P.L., Cabrera, M., Buitron, M.G., Kulik, M.G., Fernandez-Alberti, S., Fornasari, M.S., Parisi, G., Lagares, A., Tosatto, S.C.E., 2021. RepeatsDB in 2021: Improved data and extended classification for protein tandem repeat structures. Nucleic Acids Res. 49 (D1), D452–D457. https://doi.org/10.1093/nar/gkaa1097.

Parra, R.G., Espada, R., Sánchez, I.E., Sippl, M.J., Ferreiro, D.U., 2013. Detecting Repetitions and Periodicities in Proteins by Tiling the Structural Space. J. Phys. Chem. B 117 (42), 12887–12897. https://doi.org/10.1021/jp402105j.

Parrilla-Castellar, E.R., Arlander, S.J.H., Karnitz, L., 2004. Dial 9–1-1 for DNA damage: The Rad9-Hus1-Rad1 (9–1-1) clamp complex. DNA Repair 3 (8–9), 1009–1014. https://doi.org/10.1016/j.dnarep.2004.03.032.

Pascal, J.M., Tsodikov, O.V., Hura, G.L., Song, W., Cotner, E.A., Classen, S., Tomkinson, A.E., Tainer, J.A., Ellenberger, T., 2006. A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. Mol. Cell 24 (2), 279–291. https://doi.org/10.1016/j.molcel.2006.08.015.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12 (85), 2825–2830.

Piovesan, D., Tabaro, F., Paladin, L., Necci, M., Micetic, I., Camilloni, C., Davey, N., Dosztányi, Z., Mészáros, B., Monzon, A.M., Parisi, G., Schad, E., Sormanni, P., Tompa, P., Vendruscolo, M., Vranken, W.F., Tosatto, S.C.E., 2018. MobiDB 3.0: More annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Res. 46 (D1), D471–D476. https://doi.org/10.1093/nar/gkx1071.

Randell, J.C.W., Coen, D.M., 2004. The herpes simplex virus processivity factor, UL42, binds DNA as a monomer. J. Mol. Biol. 335 (2), 409–413. https://doi.org/10.1016/j.jmb.2003.10.064.

Rasheed, N., Lima, T.B., Mercaldi, G.F., Nascimento, A.F.Z., Silva, A.L.S., Righetto, G.L., Bar-Peled, L., Shen, K., Sabatini, D.M., Gozzo, F.C., Aparicio, R., Smetana, J.H.C., 2019. C7orf59/LAMTOR4 phosphorylation and structural flexibility modulate Ragulator assembly. FEBS Open Bio 9 (9), 1589–1602. https://doi.org/10.1002/2211-5463.12700.

Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., Hakoshima, T., 2005. Structural basis for recruitment of human flap endonuclease 1 to PCNA. EMBO J. 24 (4), 683–693. https://doi.org/10.1038/sj.emboj.7600519.

Schaper, E., Anisimova, M., 2015. The evolution and function of protein tandem repeats in plants. New Phytol. 206 (1), 397–410. https://doi.org/10.1111/nph.13184.

Schrödinger, LLC, 2015. The PyMOL Molecular Graphics System, Version 1.8.

Sillitoe, I., Bordin, N., Dawson, N., Waman, V.P., Ashford, P., Scholes, H.M., Pang, C.S.M., Woodridge, L., Rauer, C., Sen, N., Abbasian, M., Le Cornu, S., Lam, S.D., Berka, K., Varekova, I.H., Svobodova, R., Lees, J., Orengo, C.A., 2021. CATH: Increased structural coverage of functional space. Nucleic Acids Res. 49 (D1), D266–D273. https://doi.org/10.1093/nar/gkaa1079.

The UniProt Consortium, 2023. UniProt: The Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 51 (D1), D523–D531. https://doi.org/10.1093/nar/gkac1052.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22 (22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673.

van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., Steinegger, M., 2023. Fast and accurate protein structure search with Foldseek. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-01773-0.

Venclovas, C., Thelen, M.P., 2000. Structure-based predictions of Rad1, Rad9, Hus1 and Rad17 participation in sliding clamp and clamp-loading complexes. Nucleic Acids Res. 28 (13), 2481–2493. https://doi.org/10.1093/nar/28.13.2481.

Warbrick, E., Lane, D.P., Glover, D.M., Cox, L.S., 1995. A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21WAF1 and proliferating cell nuclear antigen. Curr. Biol. 5 (3), 275–282. https://doi.org/10.1016/s0960-9822(95)00058-3.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. Bioinformatics 25 (9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Zhang, Y., Skolnick, J., 2005. TM-align: A protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 33 (7), 2302–2309. https://doi.org/10.1093/nar/gki524.

Zhuang, Z., Ai, Y., 2010. Processivity factor of DNA polymerase and its expanding role in normal and translesion DNA synthesis. BBA 1804 (5), 1081–1093. https://doi.org/10.1016/j.bbapap.2009.06.018.