PLOS ONE

Crystal and solution structures reveal oligomerization of individual capsid homology domains of Drosophila Arc

Erik I. Hallin[1][◉], Sigurbjörn Markússon[◉][1][◉], Lev Böttger[2], Andrew E. Torda[2], Clive R. Bramham[1,3], Petri Kursula[◉][1,4] *

1 Department of Biomedicine, University of Bergen, Bergen, Norway, 2 Centre for Bioinformatics (ZBH), University of Hamburg, Hamburg, Germany, 3 KG Jebsen Centre for Neuropsychiatric Disorders, University of Bergen, Bergen, Norway, 4 Faculty of Biochemistry and Molecular Medicine & Biocenter Oulu, University of Oulu, Oulu, Finland

◉ These authors contributed equally to this work.
* petri.kursula@uib.no

Abstract

Synaptic plasticity is vital for brain function and memory formation. One of the key proteins in long-term synaptic plasticity and memory is the activity-regulated cytoskeleton-associated protein (Arc). Mammalian Arc forms virus-like capsid structures in a process requiring the N-terminal domain and contains two C-terminal lobes that are structural homologues to retroviral capsids. *Drosophila* has two isoforms of Arc, dArc1 and dArc2, with low sequence similarity to mammalian Arc, but lacking a large N-terminal domain. Both dArc isoforms are related to the Ty3/gypsy retrotransposon capsid, consisting of N- and C-terminal lobes. Structures of dArc1, as well as capsids formed by both dArc isoforms, have been recently determined. We carried out structural characterization of the four individual dArc lobe domains. As opposed to the corresponding mammalian Arc lobe domains, which are monomeric, the dArc lobes were all oligomeric in solution, indicating a strong propensity for homophilic interactions. A truncated N-lobe from dArc2 formed a domain-swapped dimer in the crystal structure, resulting in a novel dimer interaction that could be relevant for capsid assembly or other dArc functions. This domain-swapped structure resembles the dimeric protein C of flavivirus capsids, as well as the structure of histones dimers, domain-swapped transcription factors, and membrane-interacting BAK domains. The strong oligomerization properties of the isolated dArc lobe domains explain the ability of dArc to form capsids in the absence of any large N-terminal domain, in contrast to the mammalian protein.

Introduction

Memory formation in the brain is dependent on synaptic plasticity, and the activity-regulated cytoskeleton-associated protein (Arc) plays an important role in this process [1, 2]. Arc promotes the endocytosis of AMPA receptors located on the post-synaptic membrane [3–5], regulates actin cytoskeletal dynamics and dendritic spine structure [6–8], and enters the nucleus to

regulate gene expression [4, 9, 10]. The targeting of AMPA receptors may involve direct interactions of stargazin (TARPγ2) with both AMPA receptors and Arc [11–13]. Due to its many interaction partners, Arc regulates several neuronal signalling processes as well as the structure of the postsynaptic density scaffold [2, 14].

Arc forms capsid-like structures that may transfer information from one neuron to another [15, 16]. Mammalian Arc (mArc) has a C-terminal domain (Arc-CT) with close structural homology to the C-terminal domain (CA-CTD) of the retroviral capsid (CA) protein [12], and mArc-CT consists of two structurally similar lobe domains, N-lobe (NL) and C-lobe (CL) [12, 17]. Viral CA has in addition an N-terminal domain (CA-NTD), and both CA-NTD and CA-CTD are involved in viral capsid assembly. mArc has a large N-terminal domain (Arc-NT) of unknown structure, which is absent in dArc. The Arc-NT is predicted to have homology to the retroviral matrix domain and is required for the formation of large mArc oligomers. Without its N-terminal domain, mArc is monomeric in solution [18]. In mArc, it is likely that the presence of both mArc-NT and mArc-CT are required for high-order oligomerization and capsid formation [19, 20].

Drosophila has two Arc isoforms (dArc1 and dArc2), which share high sequence similarity. *Drosophila* Arc (dArc) isoforms have a CT domain, containing tandem N-and C-lobes, but lack an Arc-NT found in mArc. However, dArc forms capsids [16], whose structure has been determined by electron cryomicroscopy (cryo-EM) [21]. In addition, the crystal structure of dimeric dArc1, containing both N-and C-lobes, was recently determined [22]. Whether dArc functions similarly to mArc in neurons, even if the functionally important mArc-NT is missing and the sequence similarity to mArc is low, is currently unknown. Mammalian Arc also forms capsid-like structures [16], but the high-resolution structure remains to be solved.

We set out to determine structures of the individual dArc lobe domains. The CL of both dArc1 and dArc2 is structurally homologous to the mArc-CL, confirming the connection of dArc to both mArc and retroviral capsids. The structure of dArc2-NL showed a domain-swapped dimer, resulting in a structure similar to the flavivirus capsid protein and resembling histones as well as membrane-interacting BAK domains. All individual dArc lobes were oligomeric in solution, in contrast to the monomeric mArc-CT. Such oligomeric units, reflective of the different evolutionary pathways leading to mammalian and insect Arc, could be building blocks during the assembly of virus-like capsids by full-length dArc, or they could relate to other functions of dArc.

Materials and methods

Recombinant protein production

Proteins were expressed in *Escherichia coli* BL21(DE3) with a TEV protease-cleavable His tag-maltose binding protein (MBP) fusion at the N terminus. Cells were grown at +37˚C until an A_{600} of 0.6 was reached. 1 mM isopropyl β-D-1-thiogalactopyranoside was added to start the induction, lasting 4 h at +30˚C. The cells were lysed in HBS (40 mM Hepes, pH 7.5, 100 mM NaCl) containing 0.1 mg/ml lysozyme, by one freeze-thaw cycle followed by sonication. The lysate was centrifuged at 16 000 g for 30 min at +4˚C and loaded onto a Ni-NTA resin. After washing with HBS containing 20 mM imidazole, the protein was eluted with HBS containing 300 mM imidazole. His-tagged TEV protease [23] was added to the eluate, and the sample was dialyzed against 20 mM Hepes (pH 7.5), 100 mM NaCl, and 1 mM dithiothreitol for 20 h at +4˚C. The sample was passed through a Ni-NTA resin again to remove the TEV protease and the cleaved His-MBP tag.

For the purification of dArc1-CL and dArc2-CL, an additional step was required to remove remains of the cleaved MBP tag. This was done by passing the sample through an amylose resin, equilibrated with HBS containing 1 mM EDTA.

The Ni-NTA or amylose flow-through was loaded on a Superdex S200 16/600 column, equilibrated with TBS (20 mM Tris-HCl (pH 7.4), 150 mM NaCl). All proteins gave one major peak in the chromatogram. Selected fractions were concentrated using spin concentrators to a final concentration of 10 mg/ml. Protein purity was analyzed using sodium dodecyl sulphate–polyacrylamide gel electrophoresis, giving one strong Coomassie-stained band of the expected size. Protein identity was confirmed using mass spectrometry of trypsin-digested in-gel samples, as described [24].

The details of the protein constructs are given in S1 Table. The expression and purification of human Arc NL and CL have been described before [18].

Size exclusion chromatography—Multi-angle light scattering

The absolute mass of the proteins was determined by SEC-MALS, using a miniDawn Treos instrument (Wyatt). A Superdex S200 Increase 10/300 equilibrated with TBS was used for sample separation. The system was calibrated using bovine serum albumin, and protein concentration was measured using an online refractometer. Data were analysed with ASTRA (Wyatt).

Circular dichroism spectroscopy

The ellipticity of the proteins was recorded using a Jasco J-810 spectropolarimeter and a 1-mm quartz cuvette. The protein concentration was 0.2 mg/ml in 20 mM phosphate (pH 7). The experiments were done at +20˚C.

Crystal structure determination

Crystals were obtained by sitting-drop vapor diffusion at +20˚C. The crystals of dArc2-NL were grown by mixing 150 nl of dArc2-NL at 8 mg/ml with 150 nl of reservoir solution (200 mM ammonium chloride, 100 mM sodium acetate (pH 5), 20% PEG 6000). The crystals of dArc2-CL were made by mixing 200 nl of protein at 15 mg/ml with 100 nl of reservoir solution (1.25 M ammonium sulphate, 100 mM Tris (pH 8.5), 200 mM lithium sulphate). The crystals of dArc1-CL were made by mixing 150 nl of the protein at 12 mg/ml with 150 nl of a reservoir, consisting of 100 mM MIB buffer (malonic acid, imidazole, boric acid) (pH 5) and 25% PEG 1500. The crystals of dArc1-CL used for phasing were grown by mixing 2 μl of the purified protein at 12 mg/ml with 2 μl of a reservoir solution, consisting of 20% PEG 3350, by hanging-drop vapour diffusion at +20˚C. These crystals were soaked in a solution of 20% PEG 3350 with 500 mM NaI for 20 s.

Crystals were mounted in loops and snap-cooled in liquid nitrogen. X-ray diffraction data for dArc2-NL were collected on the I03 beamline at Diamond Light Source (Oxfordshire, UK), while the data for dArc1-CL and dArc2-CL were collected on the P11 beamline [25] at PETRAIII/DESY (Hamburg, Germany). All data were processed using XDS [26].

Phasing of dArc2-NL was done with molecular replacement in AMPLE [27] and *ab initio* models generated by QUARK [28], using the CCP4 online server [29, 30]. The phasing of dArc1-CL was done *via* iodine single-wavelength anomalous dispersion (SAD) and the Auto-Rickshaw pipeline [31], with the combined use of SHELX [32], PHASER [33], PARROT [34], and BUCCANEER [35]. The resulting near-complete model was taken as a template for molecular replacement in PHASER [33], using atomic-resolution data from a native crystal with a different space group. The phasing of dArc2-CL was done using the dimeric structure of dArc1-CL as a search model in PHASER [33]. All structures were refined with phenix.refine

Table 1. Data processing and structure refinement.

Sample	dArc1-CL (derivative)	dArc1-CL (native)	dArc2-NL	dArc2-CL
Space group	$P2_12_12$	$C222_1$	$P222_1$	$P4_22_12$
Unit cell dimensions (Å)	38.2, 82.5, 29.8	37.6, 51.1, 82.4	24.6, 33.4, 84.9	190.0, 190.0, 40.4
Wavelength (Å)	2.0662	1.0332	0.976	1.0332
Resolution range (Å)	50–2.30 (2.44–2.30)	50–1.05 (1.08–1.05)	50–1.90 (1.95–1.90)	50–2.80 (2.87–2.80)
$\langle I/\sigma(I)\rangle$	24.3 (16.0)	26.1 (1.6)	9.9 (1.0)	8.7 (0.9)
R_{sym} (%)	8.4 (10.3)	2.9 (29.5)	10.3 (143.2)	34.4 (303.6)
R_{meas} (%)	8.8 (10.8)	3.1 (39.0)	11.1 (156.8)	35.7 (315.2)
Completeness (%)	99.4 (96.6)	92.6 (45.5)	100 (100)	99.8 (98.3)
Redundancy	12.2 (10.9)	5.9 (1.6)	6.8 (6.1)	14.3 (13.9)
$CC_{1/2}$ (%)	99.8 (99.7)	100 (91.3)	99.8 (39.5)	99.2 (43.3)
Wilson B factor (Å²)	30.9	14.9	40.8	49.8
R_{cryst} (%)	-	13.0	23.8	23.1
R_{free} (%)	-	15.1	27.1	28.0
rmsd bond lengths (Å)	-	0.019	0.013	0.013
rmsd bond angles (°)	-	1.5	1.2	1.3
Ramachandran favoured/outliers (%); Molprobity score (percentile)	-	100/0; 1.77 (38th)	93.4/0; 2.22 (49th)	97.9/0; 2.02 (99th)
PDB entry	-	6SID	6SIB	6SIE

The values in parentheses refer to the highest-resolution shell.

[36], and model building was done in Coot [37]. The quality of the structures was assessed using MolProbity [38]. Data processing and refinement statistics are given in Table 1.

Structure analysis

The PISA server [39] was used to calculate probable oligomeric states from crystal symmetry, and PDBsum [40] and PISA were used for structural analysis and in-depth analysis of dimer interfaces. Structural homologues were searched using DALI [41] and SALAMI [42], in addition to known homologues from literature and manual searches. Electrostatic potential maps were calculated with PDB2PQR and APBS [43] and visualized in UCSF Chimera [44] or PyMOL. Sequence identity between the dArc CA lobes and structural homologues was calculated using the EMBOSS Needle server [45], and the dArc1-NL homology model was generated using SWISS-MODEL [46], with the dArc2-NL crystal structure as template.

Small-angle X-ray scattering

Synchrotron SAXS data for dArc1-NL and dArc2-NL were collected on the B21 beamline at Diamond Light Source (Oxfordshire, UK) using a SEC-SAXS setup, where SAXS frames are collected as the sample elutes from a SEC column. A Superdex S200 Increase 3.2 column equilibrated with TBS was used. The injected sample was at 5 mg/ml, and the measurements were done at +10˚C. SEC-SAXS data for dArc1-CL and dArc2-CL were similarly collected on the P12 beamline [47] at EMBL/DESY (Hamburg, Germany).

SAXS data were processed using ATSAS [48], and the frames showed no signs of aggregation or radiation damage. *Ab initio* dummy atom and chain-like SAXS models were built with DAMMIN [49] and GASBOR [50], respectively. CRYSOL [51] was used to generate SAXS scattering profiles from 3D protein structures and compare these to experimental SAXS data.

Isothermal titration calorimetry

A MicroCal iTC200 instrument (Malvern, UK) was used to determine the binding affinity of a stargazin peptide (RIPSYRYR with N-terminal acetylation and C-terminal amidation) to the NL of *Drosophila* and human Arc. Arc in the cell had a concentration of 0.5 mM (dArc N-lobes) or 0.25 mM (hArc N-lobe); the peptide concentration in the syringe was 10-fold higher. The peptide was injected in 26 3-μl injections, with an initial injection of 0.5 μl and a second 0.5-μl injection after 14 injections due to syringe refill. Both the protein and peptide were in TBS buffer. The experiments were done at +25˚C, and the data were analyzed with MicroCal Origin 7, using a one-site binding model.

Sequence comparisons

For conservation calculations, homologues from the non-redundant sequence database were collected using BLAST, accepting hits with an *e*-value less than 10^{-10} [52]. For larger sets and phylogenetic speculation, homologues were collected using iterative PSI-BLAST in up to three stages, each with no more than four iterations, accepting homologues with *e*-values less than 10^{-20}, 10^{-10}, and 10^{-8} [53]. Full length proteins or the candidate ranges were re-aligned using MAFFT in its most accurate mode, with up to 200 iterations [54]. Before alignment, redundancy amongst the sequences was removed by calculating an alignment in fast mode, saving the matrix of distances between sequences, and sorting the list of pair distances. Starting from the smallest distance, one member of each pair was removed until the target number was reached. This removes redundancy and ensures the most even spread of sequences within a set of homologues. For conservation calculations, search starting points were NP_610955.1 (dArc1-NL and full-length dArc1), as well as PDB codes 6sib for dArc2-NL, 6sid for dArc1-CL, and 6sie for dArc2-CL.

Sequence conservation/variability was calculated from the alignments using entropy,

$$S = \Sigma_{i=1}^{20} p_i \log_{20} p_i$$

where p_i is the frequency of amino acid type *i* at a given alignment position.

Results and discussion

Although mArc and dArc both contain similar lobe domains (Fig 1), in mArc, low solubility is linked to the presence of the NT [18]. A construct containing both the NL and CL is soluble and fully monomeric for hArc [18]. dArcs lack an Arc-NT, but they are predicted to have an N-terminal helix that might play a role in oligomerization; the N-terminal region forms spike-like structures on the outside of the dArc capsids [21].

The isolated NL and CL domains of both dArc1 and dArc2 could be produced in soluble form. However, our dArc constructs containing both lobes were poorly soluble; this low solubility resembles that of full-length mArc. While we could obtain pure dArc1 containing both lobes, the protein formed a mixture of large oligomeric states upon storage; dArc2 with both lobes was insoluble in our hands. These problems are likely linked to the missing dArc N-terminal tail in our constructs, which recently was shown to be an integral part of the dArc-NL fold [21, 22], as opposed to mArc. These aspects are discussed more in the sections below, and our experimental work here focuses on the individual dArc lobes in isolation.

The different behaviour of the isolated lobe domains suggests that mArc and dArc differ in the mechanisms, through which they form larger structures, such as capsids. Our aim was to understand the structural basis of these differences, using the individual dArc1 and dArc2 lobe domains for structural studies.
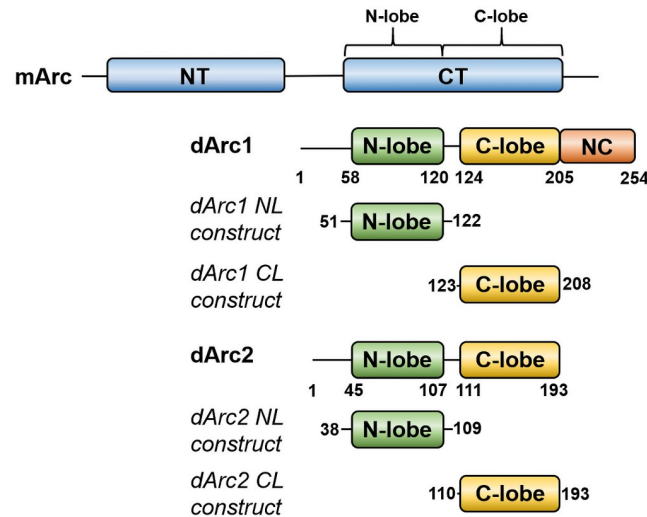
Fig 1. Comparison of mArc and dArc domain structure. The constructs used for structural studies on the individual dArc lobe domains are indicated.

Crystal structure of dArc2-NL presents a domain-swapped dimer similar to nucleotide- and membrane-interacting proteins

It has been suggested that viral CA-CTD domains employ different dimerization modes during capsid assembly [55]. In our 1.90-Å crystal structure, dArc2-NL is a domain-swapped dimer, in which the second and third helix of the canonical lobe domain join to form an extended helix (Fig 2A–2C). One layer of the dimer is formed by α2 of each subunit, which pack at a cross angle of 140.7˚. The other layer is formed by the remaining helices, and α1 lies in a groove formed between α2 and α3. The subunit interface of the dimer spans 3440 Å^2 of buried surface area. The interface consists exclusively of π-π and van der Waals interactions, and the helices encapsulate a hydrophobic core between the monomers, formed by Phe51, Val55, Pro74, Phe77, Ile80, Trp84, Trp95, Leu99, Leu102, and Phe106 (Fig 2B). The solvent-exposed surface of the dimer is charged and polarized (Fig 2D); the electrostatic surface potential of the side formed by the two coiled α2 helices is highly positive, while the opposite surface formed by α1 and α3 is mainly negative. In the extended crystal lattice (S1 Fig), these opposing charges take part in crystal contacts. Crystal packing does not, however, explain the observed homotetrameric oligomeric state of dArc2-NL in solution (see below).

The dArc2-NL domain-swapped dimer differs from the dimers of retroviral capsid CA-CTD seen with crystallography and NMR [56] and might be functionally relevant for dArc oligomerization and capsid formation. However, the cryo-EM structures of the dArc1 and dArc2 capsids [21] do not show such domain swapping, and the dArc-NL forms penta- and hexameric rings instead (Fig 3A). In both capsids, a kink is present between α2 and α3 (at Leu76, Phe77, and Lys78 in dArc2), resulting in a canonical lobe domain fold. As a result, the surface charges of the lobe are re-oriented such that the negatively charged surface of α1 can interact with α2 (Fig 3A). Therefore, the turn seems vital to the formation of the capsid hexa- and pentamers. An interesting discrepancy between the domain-swapped dimer, the capsid structure, and the recently published crystal structure of the dArc1 CT domain [22], is the presence of the N-terminal tail preceding the NL (residues 29–44 in dArc2, residues 41–57 in dArc1). In the capsid structures of both dArc isoforms, the tail packs into the exposed hydrophobic core of the lobe (Fig 3B), and Phe32 and Phe39 are observed in two hydrophobic
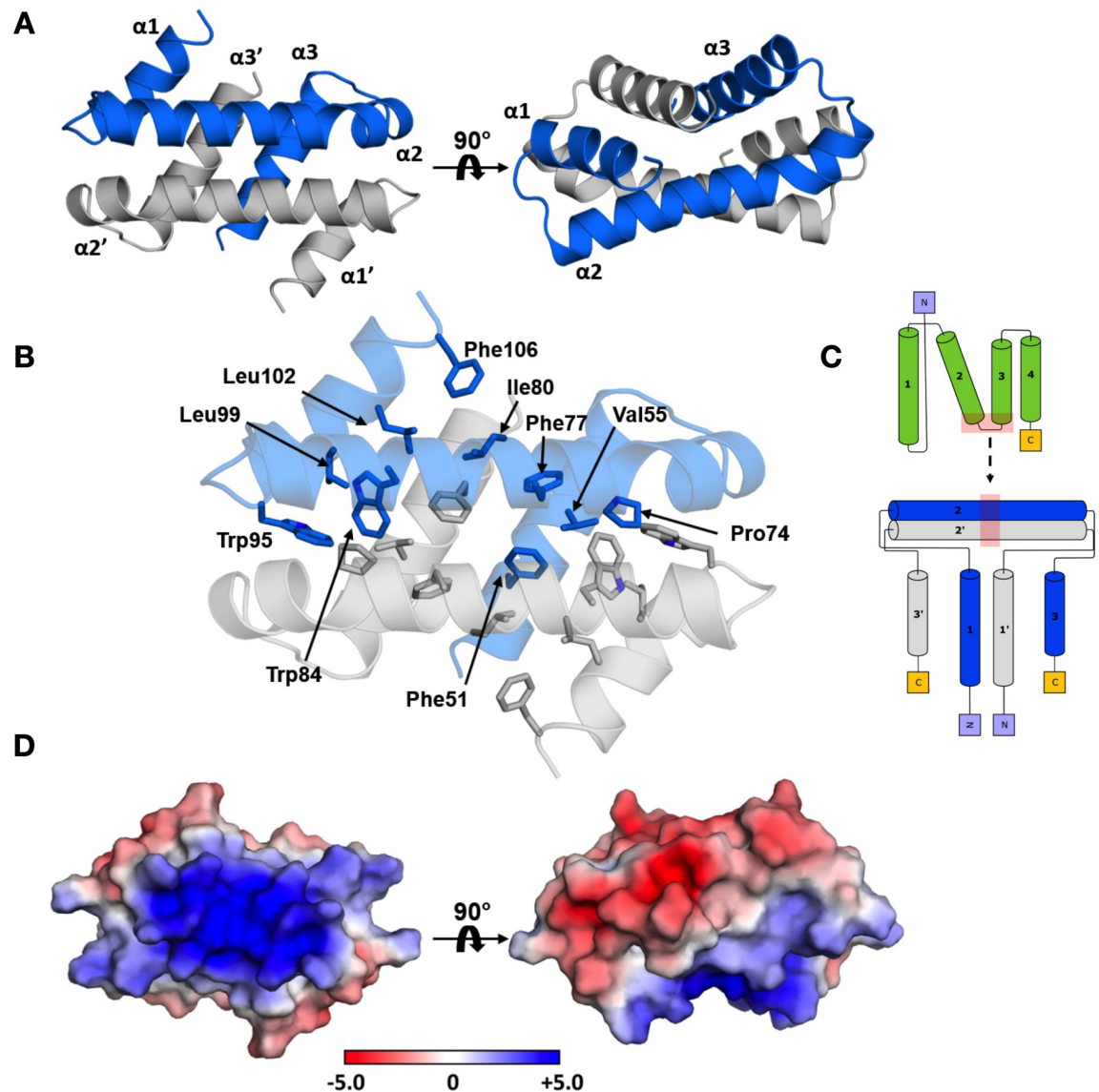
Fig 2. The crystal structure of dArc2-NL. (A) The domain-swapped dimer observed in the crystal, with the three α-helices labeled. (B) The folded dimer encapsulates an extensive hydrophobic core, with no polar interactions connecting the two monomers. Residues are only labelled in subunit A, but also seen in subunit B. (C) Topology diagram of domain swapping. The location of the conformational change is indicated by red shading. (D) The electrostatic surface of the dimer. The α2 and α2' have a positive surface potential, in contrast to the surface formed by α1, α3, α1', and α3'. The protein is in the same orientations as in (A).

https://doi.org/10.1371/journal.pone.0251459.g002

pockets and Ser40 interacts directly with Lys78 and Ser79, of the α2 kink, *via* hydrogen bonding. Furthermore, the binding site for the N-terminal tail in dArc-NL corresponds to the peptide binding pocket of mammalian Arc-NL [12, 17]. Only a part of the tail (residues 37–45) was included in the dArc2-NL crystallized here, reflecting the fact that our experiments were carried out before structural information for the capsids and the importance of the N-terminal tail was available. In our structure, this fragment could not be resolved, due to flexibility. Therefore, it seems that the full N-terminal fragment, including Phe32, is needed for the packing of the tail into the exposed hydrophobic core of the canonical N-lobe and the formation of the penta- and hexameric forms of dArc-NL found in capsids. The binding surface for the
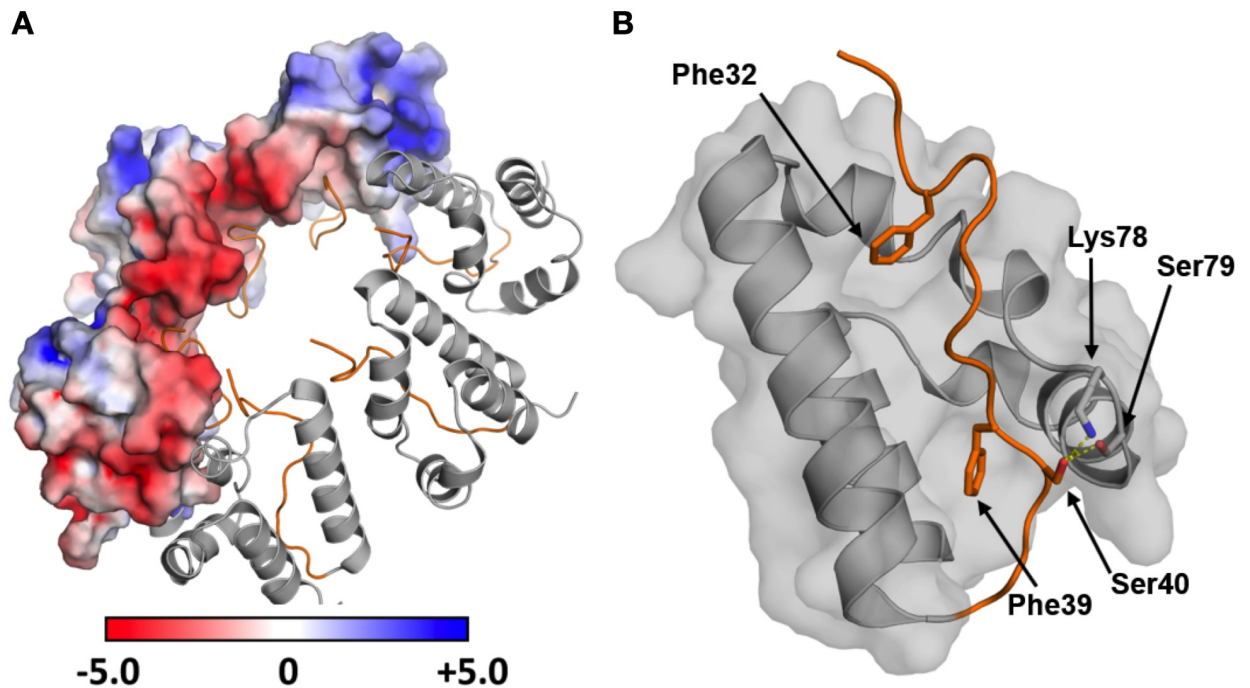
Fig 3. The N-terminal region preceding dArc-NL packs into the hydrophobic core of the domain and leads to the formation of the capsid hexamer. (A) The hexameric form of dArc2-NL observed in the capsid (PDB: 6TAQ; [21]), showing the electrostatic surface potential for half of the monomers. The canonical fold enables contact formation between the oppositely charged surfaces of each monomer. The N-terminal tail is showed in orange. (B) Residues contributing to the packing of the N-terminal tail (orange) into the capsid hexamer. Phe32 and Phe39 pack into two exposed pockets in the hydrophobic core. Further interactions are observed for Ser40, which hydrogen bonds directly with Lys78 and Ser79 in the α2 kink (yellow dashed lines).

N-terminal segment is buried at the interface of the dArc2-NL domain-swapped dimer (Fig 2A and 2B). Taken together, the structure of dArc2-NL reveals an intrinsic property of the Arc lobe domain to form alternate dimers *via* domain swapping, possibly regulated through interactions of the folded dArc2-NL core domain with the N-terminal tail. The biological relevance of domain swapping in dArc-NL remains to be determined.

In exploring the domain-swapped dArc2-NL dimer, we found that the overall fold of the dimer resembles retroviral proteins known to exhibit domain swapping. dArc2-NL resembles the flaviviral capsid C protein [57], which also forms domain-swapped dimers [58]. Despite low sequence similarity with the Dengue virus 2 C protein (16.2%) and the core protein of the Kunjin subtype West-Nile virus (9.4%), the overall fold is surprisingly similar (Fig 4A and 4B) [57, 59]. Both the Dengue and West-Nile virus are enveloped RNA viruses, and these proteins are essential for the formation of the viral capsid. Despite different helix topology, the two proteins share fold similarity with dArc2-NL and have highly positive electrostatic surfaces, suggested to have a role in the binding of encapsulated genomic RNA [57, 59]. Interestingly, in the crystal, the West-Nile virus core protein forms tetramers [59]. In the tetrameric form, long helices from each monomer (homologous to α2 in dArc2-NL) form a four-helix bundle subunit interface (Fig 4A). This could be similar to the dArc2-NL tetrameric form in solution (see below). Moreover, the arrangement of dArc2-NL in the crystal bears some similarity to the domain-swapped dimer of the HIV CA lobe domain, induced by the deletion of a single residue [60].

Domain swapping of modular proteins emerges as a common theme in capsid-forming proteins. The dArc2-NL structure, with similarity to both retroviral and flaviviral capsid
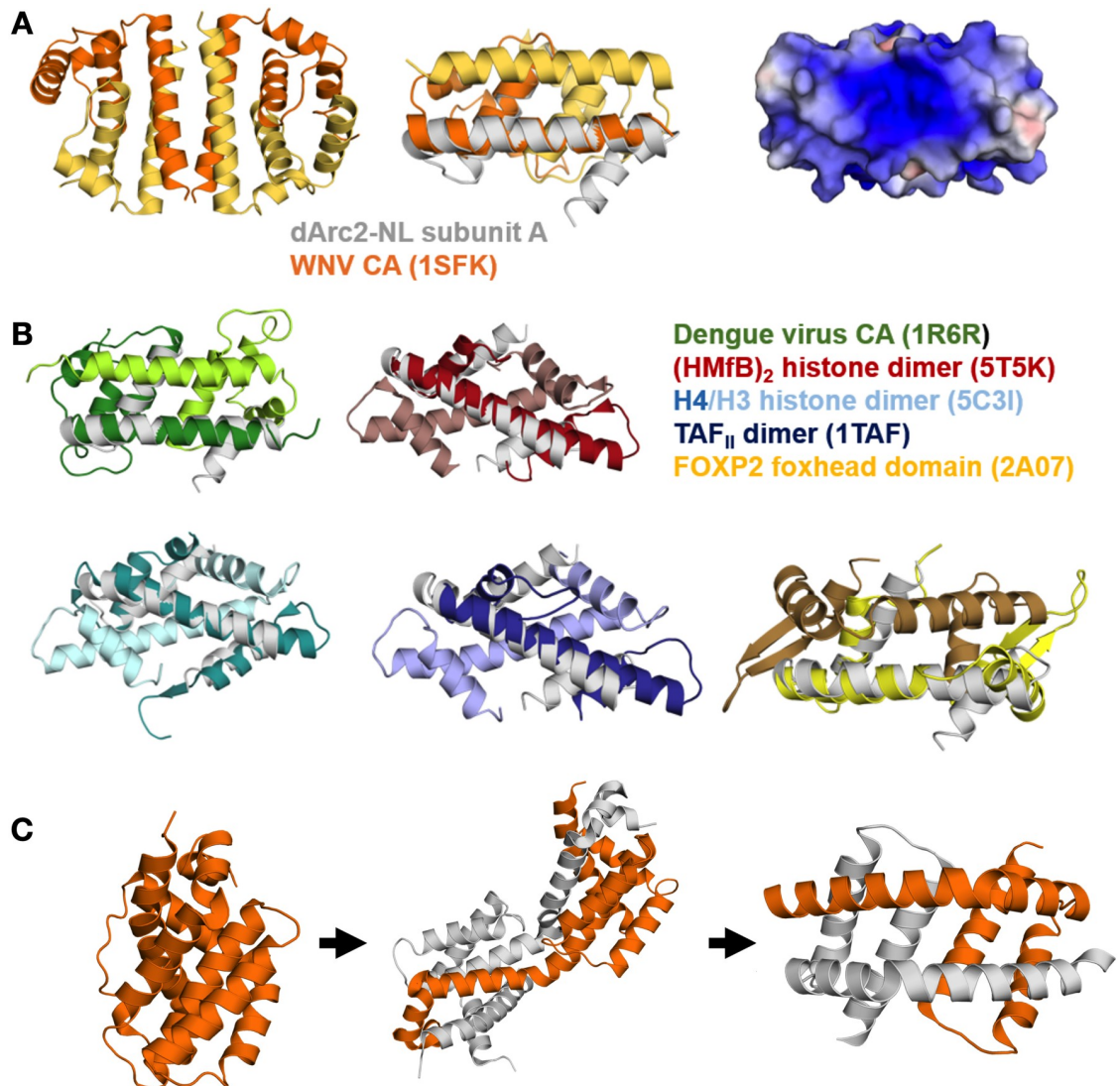
Fig 4. The dArc2-NL domain-swapped dimer resembles flaviviral coat proteins and DNA-binding proteins. (A) The tetrameric coat protein of the Kunjin subtype West-Nile virus (WNc), where the longest helix of each monomer (analogous to α2 of dArc2-NL) contributes to a four-helix bundle interface (PDB: 1SFK [59]) (left). Middle: a single dimer of the tetramer (yellow/orange) overlaid with subunit A from dArc2-NL (grey). Right: the electrostatic surface potential of a WNc dimer, which resembles that of dArc2-NL. (B) Structural comparison between the dArc2-NL and similar domain-swapping proteins. Shown are the retroviral Dengue virus CA (green; PDB: 1R6R [57]) and the DNA binding dimers of (HMfb)$_2$ histone (red; PDB: 5T5K [61]), a dimer of histones H3 and H4 (cyan; PDB: 5C3I [62]), TAF$_{II}$ transcription factor (blue; PDB: 1TAF [63]) and the foxhead domain of the FoxP2 transcription factor (yellow; PDB: 2A07 [64]). Each chain in a dimer is coloured with a different shade, and a dArc2-NL monomer is superimposed and shown in grey. (C) Domain swapping and conformational selection in the apoptosis-induced BAK protein. Shown on the left is the inactive monomeric form of BAK (PDB: 2IMT [65]), which has an orthogonal bundle fold similar to Arc-NL. Binding of a BH3 domain causes partial unfolding and opening of the hinge region (middle, PDB: 4U2U [66]), which leads to the formation of a membrane-binding domain-swapped dimer (right, PDB: 4U2V [66]). Panel C is based on [67]. The two chains in the BAK dimer are coloured grey and orange.

domain structures, shows that it is possible, *via* an extended helix, to transform the canonical capsid domain to a domain-swapped dimer. Whether such structures are related to the evolutionary history of capsid proteins, remains to be studied. It is interesting to note that the nucleocapsid protein from SARS coronavirus [68] also dimerizes *via* domain swapping, while the sequence and structure are not similar to Arc.

In addition to viral capsid proteins, the structure of a single chain in the domain-swapped structure of dArc2-NL resembles the histone core protein monomer, as well as that of TATA box-binding protein-associated factors and the foxhead domain FoxP transcription factors (Fig 4B) [14, 63, 64, 69, 70]. The foxhead domain exists as both monomers and DNA binding domain-swapped dimers [64], which share similar fold topology with the dArc2-NL dimer. Moreover, upon replacement of a crucial alanine residue in the hinge region with proline (A39P), the foxhead domain lost all domain swapping ability [71]. The histone protein forms dimers, which combine to form tetramers and finally an octamer, to which DNA binds to form the nucleosome [72]. The histone and dArc2-NL dimer arrangements are different (Fig 4B), but the monomer structures are strikingly similar. This observation could be related to either the propensity of certain protein sequences to form domain-swapped structures or a functional similarity. The above observations on dArc2-NL are interesting in light of the histone mimicry by Dengue virus protein C [73], which interferes with host histones to inhibit nucleosome formation and gene transcription [9]. Whether such a mechanism could be important for Arc function, as mArc accumulates in the nucleus, associates with specific histone-modifying complexes, and is implicated in regulation of chromatin state and transcription [9, 10, 14, 74], is a subject for future studies.

Similar domain swapping has been observed for the membrane binding core domains of BAK and BAX. BAK and BAX are members of the Bcl2 protein family and are important mediators of apoptosis. In its inactive form, BAK is monomeric and fully soluble, and the core domain has an orthogonal bundle fold. Upon activation, mediated by binding of certain BH3-only proteins into a hydrophobic groove in the core domain, the protein is partially unfolded, which leads to separation of the core and latch domains. The core domain then dimerizes to form amphipathic domain-swapped dimers [66, 75]. These dimers can further oligomerize and partition to the outer mitochondrial membrane where they bind and cause permeabilization, leading to the release of apoptotic factors, such as cytochrome c, into the cytosol [67]. Both the inactive and active forms of BAK show similar fold topology to the monomeric and dimeric state of the dArc2-NL, respectively. Additionally, the structure of the BAK intermediate, with the hinge region not fully open, could suggest a similar mechanism for conformational selection in dArc (Fig 4C). This could give relevant insight into the mechanism of domain swapping of dArc2-NL, in which interactions in the hydrophobic peptide binding groove seem of importance. Specific interactions in the groove might lead to dimerization, upon which the protein surface potential is rearranged to accommodate for nucleic acid or membrane binding.

Evolutionary aspects of dArc2-NL domain swapping

Fig 5A shows the sequence entropy (opposite of conservation) of Arc NL and CL, based on hundreds of aligned sequence homologues. In dArc2-NL, the most conserved residues are Ala81, Trp84, and Trp85, which sit on the hydrophobic side of the long α2 helix, corresponding to a conserved hydrophobic core in the domain family. Most interesting is Ser79 of dArc2, given its possible role in domain swapping. In all Arc-CL structures, the corresponding residue is a glycine in a β-turn, with a positive φ angle. In the dArc2-NL structure, there is no β-turn, and the corresponding residue, Ser79, is in the middle of a long regular α-helix. Overall, the analysis shows lower levels of conservation (higher entropy) of Arc-NL, when mArc is included in the search; this is an indication of higher conservation of Arc-CL than Arc-NL between insects and vertebrates. The result suggests that the function of Arc-NL may not have been fully conserved between mArc and dArc.
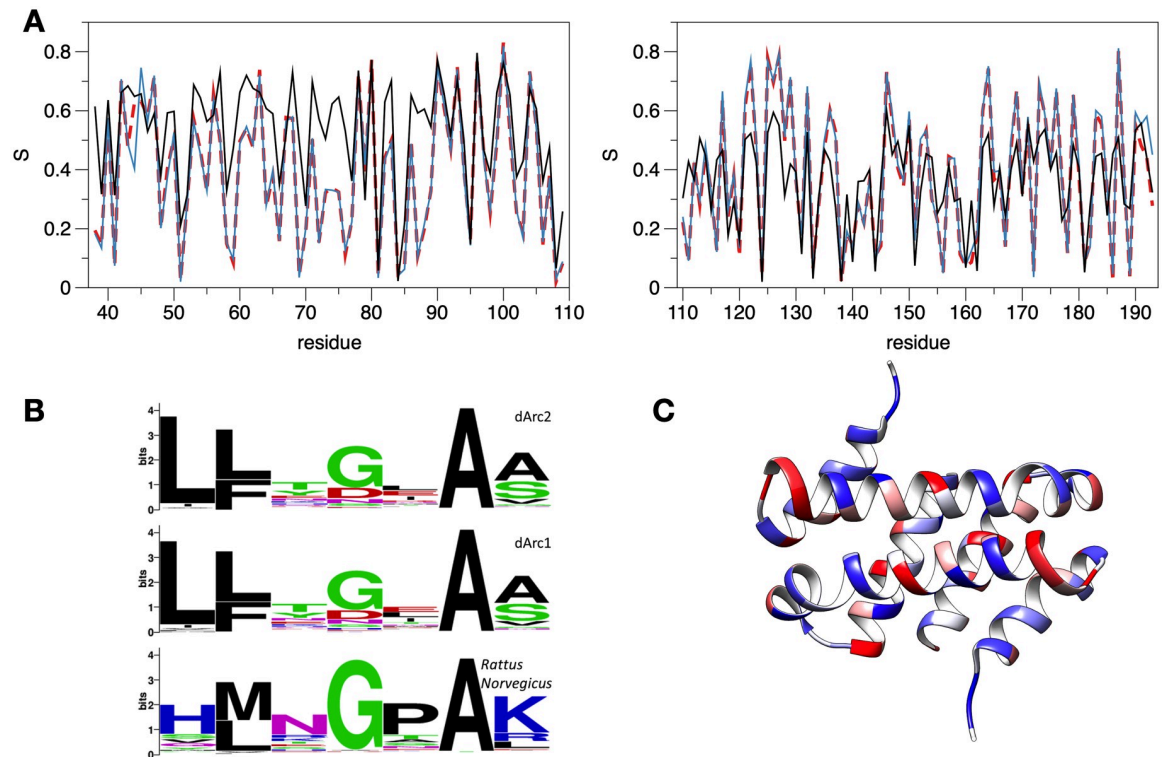
Fig 5. Sequence conservation analysis of the central lobe region. (A) Sequence variability in Arc N- and C-lobes (left and right, respectively). *S* is sequence entropy / variability within the seach results at each position (see Methods for details). Numbering follows dArc2. Red dashed lines: a search with dArc2 gave 220 homologues. Blue: a search with dArc1 and dArc2 resulted in a combined group of 250 homologues. Black: 699 sequences resulting from a search with dArc1, dArc2, and rat Arc. (B) Sequence logo for the region centered at Ser79 of dArc2-NL compared to corresponding residues from dArc1 and *Rattus norvegicus* homologues. While Gly is the most conserved residue at this position, starting searches with dArc1 and dArc2 indicates variability also at this position, unlike a search with rat Arc. (C) Mapping of conservation onto the dArc2-NL dimer. Blue corresponds to conserved and red to non-conserved sites.

https://doi.org/10.1371/journal.pone.0251459.g005

In evolutionary terms, Ser79 of dArc2-NL is an outlier, as clearly shown by a sequence logo of the region (Fig 5B). The sequence window around Ser79 in dArc2 is lfkSiav, but whether one looks at homologues of dArc1, dArc2, or mArc, the site corresponding to Ser79 is most often a Gly, Asp, or Asn. These are common residues in a β turn [76], but Ser and Thr are also possible [77]; this changes the interpretation. In the dArc2-NL crystal structure, one has a domain-swapped dimer and an α helix, where related structures have a β turn. Looking at the sequence homologues, the proteins have kept residues, which can adopt positive φ angles and are likely to adopt turns. Looking at the conservation mapped onto the dArc2-NL structure, no clear cues are observed; rather, conserved residues are evenly dispersed along the folded structure (Fig 5C). In the structures of the dArc1 and dArc2 capsid [21] as well as the crystal structure of dArc1 [22], the NL has the canonical fold without domain swapping. These features imply that the domain-swapped dArc2-NL structure might be due to crystallization of one domain alone, possibly linked to the deletion of the N-terminal tail segment from the construct, but the result does confirm the general capability of CA domains to dimerize through different modes, including domain swapping [55, 56, 60, 78].

Structures of dArc C-lobes

Both dArc1 and dArc2 CL crystallized as homodimers (Fig 6A and 6B, S2 and S3 Figs), and the structures were refined at resolutions of 1.05 and 2.80 Å, respectively. Each monomer
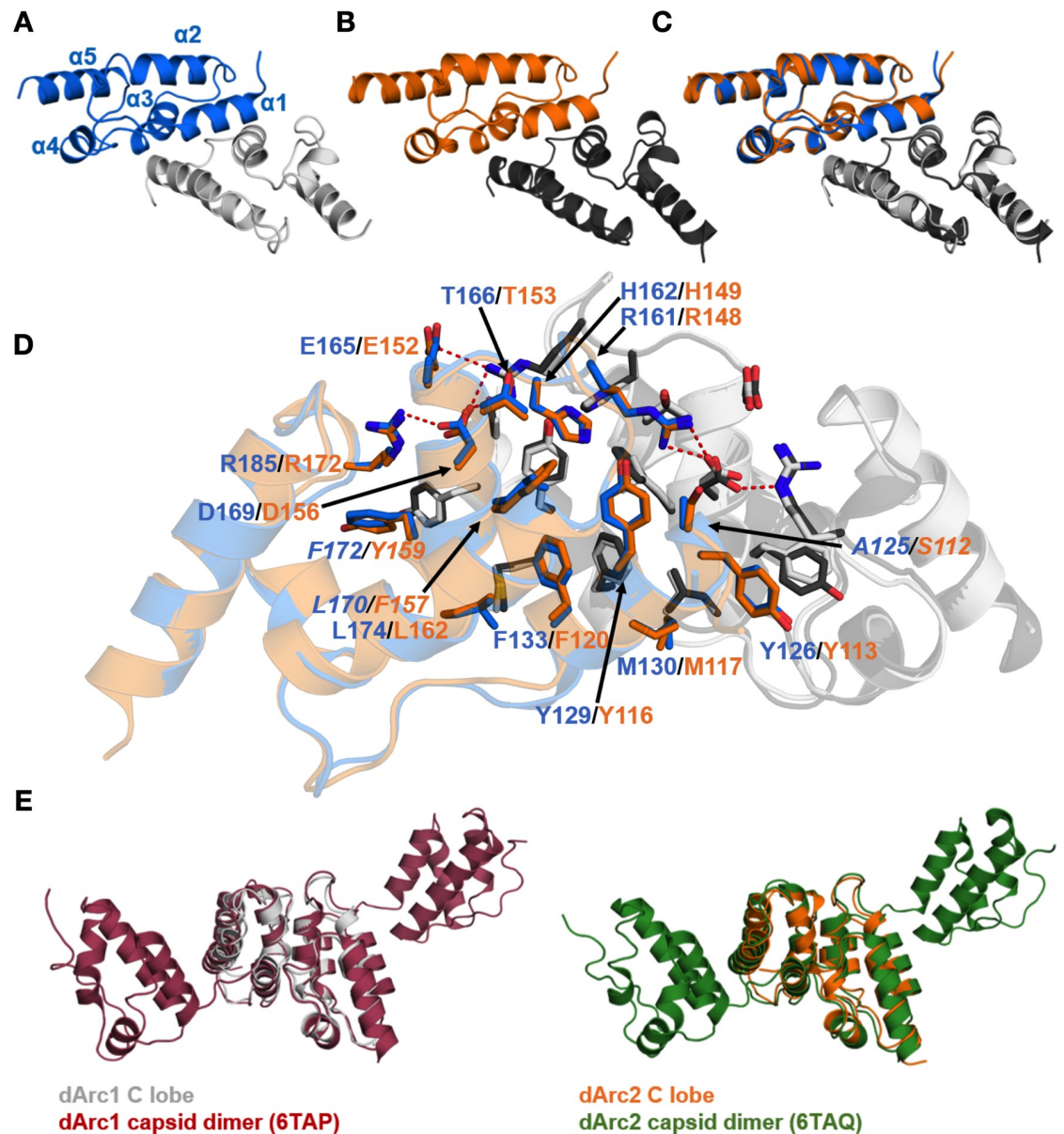
Fig 6. Crystal structures of the dArc1 and dArc2 C-lobes. (A) dArc1-CL. (B) dArc2-CL. (C) The two structures, which deviate with an all-atom RMSD of 0.48 Å, superimposed. (D) Residues contributing to the dimer interface in dArc-CL. dArc1-CL residues are marked in blue, and residues of dArc2-CL are marked in orange. Variable residues are indicated in italics. All residues contributing to the dimer interface are conserved, with the exception of A125 (dArc1) which corresponds to S112 (dArc2). Polar interactions are shown with red dashed lines. (D) A comparison of the dArc-CL crystal structures with the same domains in dArc capsids. Both the dArc1 and dArc2 C-lobes closely resemble their counterparts in the capsids, with an all-atom RMSD of 1.75 Å² and 1.13 Å², respectively.

consists of five helices in an orthogonal bundle fold, and the structures are highly similar to each other (Fig 6C). The dimer interface is in both cases formed by α1 and α3 from each monomer, and the total buried surface area at the interface is ~1400 Å² (Fig 6D). Both interfaces contain four hydrogen bonds and four salt bridges. The interface is conserved, displaying only three conservative replacements between the isoforms (A125/L170/F172 in dArc1 to S112/F157/Y159 in dArc2). The dArc-CL dimers resemble the corresponding domains in the

dArc capsids (Fig 6E). The CA domain of dArc1 is dimeric in solution [22]. Conservation of the dimer interface suggests a vital function of this mode of oligomerization in dArc function, as both a capsid and a dimer in solution.

The dArc CL domains are dimeric also in solution (see below). This behaviour of the dArc C-lobes is different to the monomeric mArc C-lobe [18], while both share the same core structure [12]. The dimer interface in dArc-CL, which corresponds to that in retroviral CA-CTD [56], contains mainly hydrophobic interactions; half of these hydrophobic residues are polar in the rat Arc-CL, and the first helix of the dArc-CL, a major part of the dimer interface, is tilted away in mArc, possibly explaining the monomeric state of mArc-CL in solution [18].

The crystal structures of the dArc C-lobes resemble those of mArc and retroviral capsid proteins (Fig 7A). The dimerization of the retroviral CA-CTD is similar to that of dArc-CL; α1 and α3 of each five-helix bundle contribute to the subunit interface (Fig 7B). However, in both HIV and bovine leukemia virus (BLV) CTDs, the N-terminal segment differs from Arc, consisting of a seven- and six-helix orthogonal bundle, respectively. HIV forms elongated conical capsids, and HIV-1 CA assembles spontaneously into helical tubes *in vitro* [79]. Thus, despite high similarity of individual domains within the fold family, the assembly mechanisms into larger structures may be different and depend on additional domain modules in the corresponding protein.

Both the intact mArc-CT and the mArc-CL alone are monomeric in solution [12, 18, 84]. The crystal structure of the rat Arc CL suggests a monomeric state [12], and a dimer similar to dArc-CL cannot be found in the crystal symmetry. However, a likely dimeric state of the protein was found in the crystal lattice by PISA (Fig 7B). Despite the high structural similarity to both the dArc C-lobes, oligomerization differs in mArc. In this putative dimer, the interface is formed by α1 and α2 of each monomer. The total buried surface area at the interface is similar to both dArc1-CL and dArc2-CL, being composed of 75 van der Waals and π-π contacts, 2 hydrogen bonds, and 4 salt bridges.

Conservation within the Arc-CL (Fig 5A) raises some questions. The most conserved residue (Gln124 in dArc2) is structurally important, forming hydrogen bonds and contacts with many neighbours, including the conserved residues Phe133 and Met162. Arg138 and Asp151 are surface-exposed, but highly conserved in both insects and mammals. Therefore, they could be central in a network of salt bridge interactions on the CL surface. It is likely that such conserved residues are required for the correct folding of the Arc lobe structure.

Comparison to the crystal structure of dimeric dArc1

At the time our experiments were planned and carried out, no data had been published on dArc structure. Since then, both the crystal structure of a dimeric, bilobar dArc1 construct [22] as well as cryoEM structures of dArc1 and dArc2 capsids [21] have been revealed. While these recent data provide central information on structure-function relationships in dArc, our data complements this work by highlighting unique properties of the individual dArc lobes compared to mArc; all the dArc lobes are oligomeric, while mArc lobes and the mArc-CT are monomeric [17, 18]. This likely reflects the fact that the dArcs lack a domain corresponding to the mArc-NT; this domain is required for higher-order oligomerization and capsid formation by mArc [18, 19]. Hence, as we await the full structure of mArc capsids, it is already evident that the mechanisms of capsid formation must be different at the molecular level between dArc and mArc.

Above, we compared the current structures mainly to the dArc capsids and homologous proteins. Importantly, the crystal structure of dimeric dArc [22], with both lobes and a longer
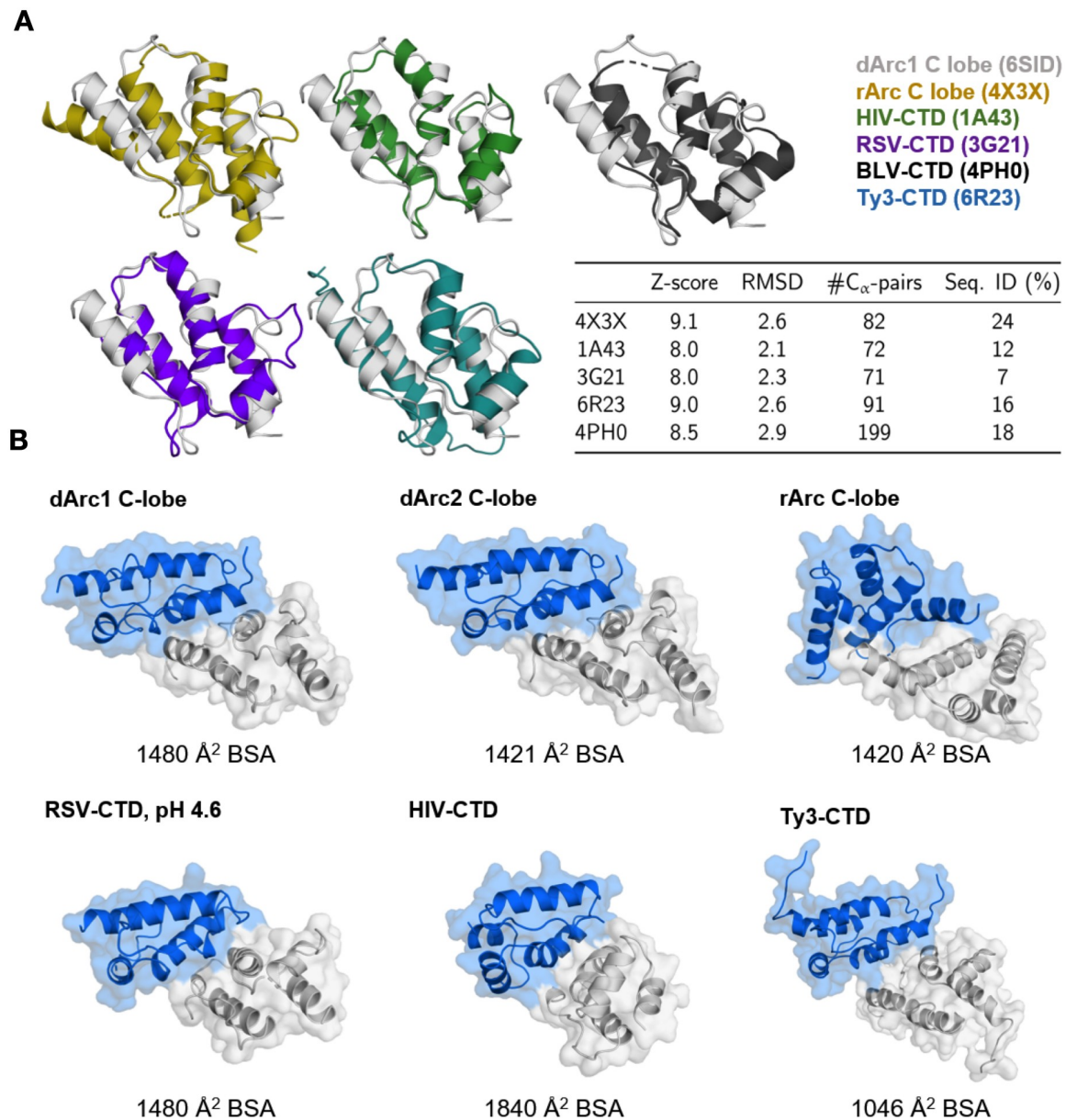
Fig 7. The structure of dArc-CL resembles that of mArc and retroviral capsid proteins. (A) Structures similar to dArc1-CL. dArc1-CL (grey) is shown superimposed with crystal structures of the rat Arc C-lobe (yellow; PDB: 4X3X) [12], HIV CA-CTD (green; PDB:1A43) [80], bovine leukemia virus (BLV) C-terminal domain (black; PDB:4PH0) [81], the rous sarcoma virus (RSV) C-terminal domain crystallized at pH 4.6 (purple; PDB: 3G21) [82], and the C-terminal domain of the Ty3 retrotransposon capsid (cyan; PDB: 6R23) [83]. Also shown are the scoring criteria obtained from the Dali server. (B) Comparison of CT dimerization. Shown are homodimers of the structural homologues in (A) and dArc2-CL, as calculated by PISA [39] from the crystalline states, apart from the BSV-CTD, which was not dimeric. Buried surface area (BSA) of each interface is shown below each structure. Note that even though a homodimer is predicted for the rat Arc-CL, this domain is monomeric in solution, and the predicted dimer is arranged differently from dArc.

N-terminal tail, provides additional information (Fig 8). As for the capsids, both dArc1-CL and dArc2-CL superimpose well on the CL dimer seen in the dArc1 crystal structure.

Interesting differences arise, however, when we compare different Arc-NL structures to each other. The dArc1-NL in the structure of Cottee *et al.* [22] is monomeric and in the
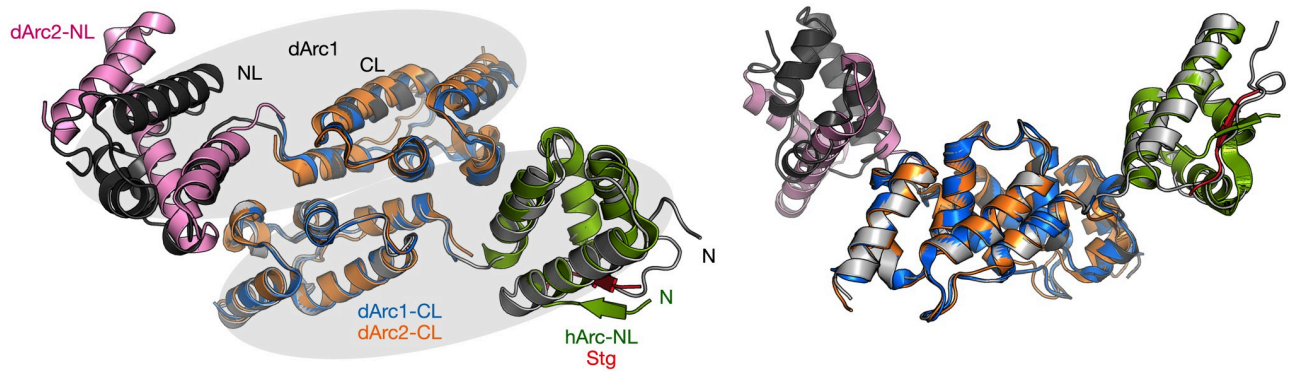
Fig 8. Comparison to the dArc1 crystal structure. Left: top view of the dArc1 dimer (gray); the two monomers are highlighted by ellipsoids in light gray. The superposed structures on the CL dimer and the two NL domains are indicated, and include dArc2-NL monomer (pink; this work), dArc1-CL dimer (blue; this work), dArc2-CL dimer (orange; this work), and hArc-NL (green) complexed with the Stg ligand peptide (red) [17]. Right, the same structures viewed from the side of the dArc1 dimer.

https://doi.org/10.1371/journal.pone.0251459.g008

canonical fold; the N-terminal tail is inserted into a hydrophobic groove (Fig 8), similarly to the capsid structures (Fig 3B). Superimposing the structure of hArc-NL in complex with a ligand peptide [17] clearly shows the same pocket is used for ligand peptide recognition in mammalian Arc. Hence, a similar binding property does exist in mArc and dArc, but according to current data, dArc binds its own tail with this pocket, while mArc uses this site to bind ligand proteins. In the current study, this part of the tail was missing from the constructs, which may be linked to the formation of the domain-swapped dimers with a different folding of the monomers (Fig 8).

As we were unable to produce dArc constructs with both lobes preset for structural studies, a brief comparison of the protein production methods to recent studies is warranted. Cottee *et al.* [22] successfully produced dimeric, soluble dArc1 for crystallization using a C-terminal His tag, while we had a cleavable His-MBP tag at the N terminus. For capsid formation, full-length dArc1 and dArc2 were produced as soluble GST fusions [21], and capsids were spontaneously formed upon tag cleavage. Additionally, our constructs lacked several residues of the N-terminal tail shown to bind to the dArc hydrophobic pocket in both the crystal state and capsids [21, 22]. It is possible that the larger oligomeric species we observed for both dArc1 and dArc2 carrying both lobes were in fact capsid-like structures; however, we decided to focus the current work on the individual lobe domains.

All dArc lobe domains are oligomeric in solution

The structure and oligomeric state of the dArc lobe domains were analyzed in solution by SEC-MALS, SAXS, and CD (Fig 9, Table 2). Both dArc1-CL and dArc2-CL are compact and slightly elongated, fitting the crystallographic dimers (Fig 9A and 9B). dArc1-NL is similar, being the size of a dimer. dArc2-NL is twice the size of dArc1-NL in solution, indicating a tetramer. The details of the latter arrangement are currently unknown, since no symmetric tetrameric assemblies can be deduced from the crystal structure, but the assembly could be similar to the West Nile virus C protein [59]. The tetrameric C protein resembles the dArc2-NL (Fig 4A), in that it is a dimer of domain-swapped dimers similar to the dArc2-NL homodimer. SAXS data for dArc2-NL in solution fit the structure of a similar tetrameric assembly (Fig 9C and 9D). Hence, it is possible that the observed dArc2-NL tetramers in solution assemble in the same way as those for the C protein, but higher-resolution data would be required to confirm this hypothesis.
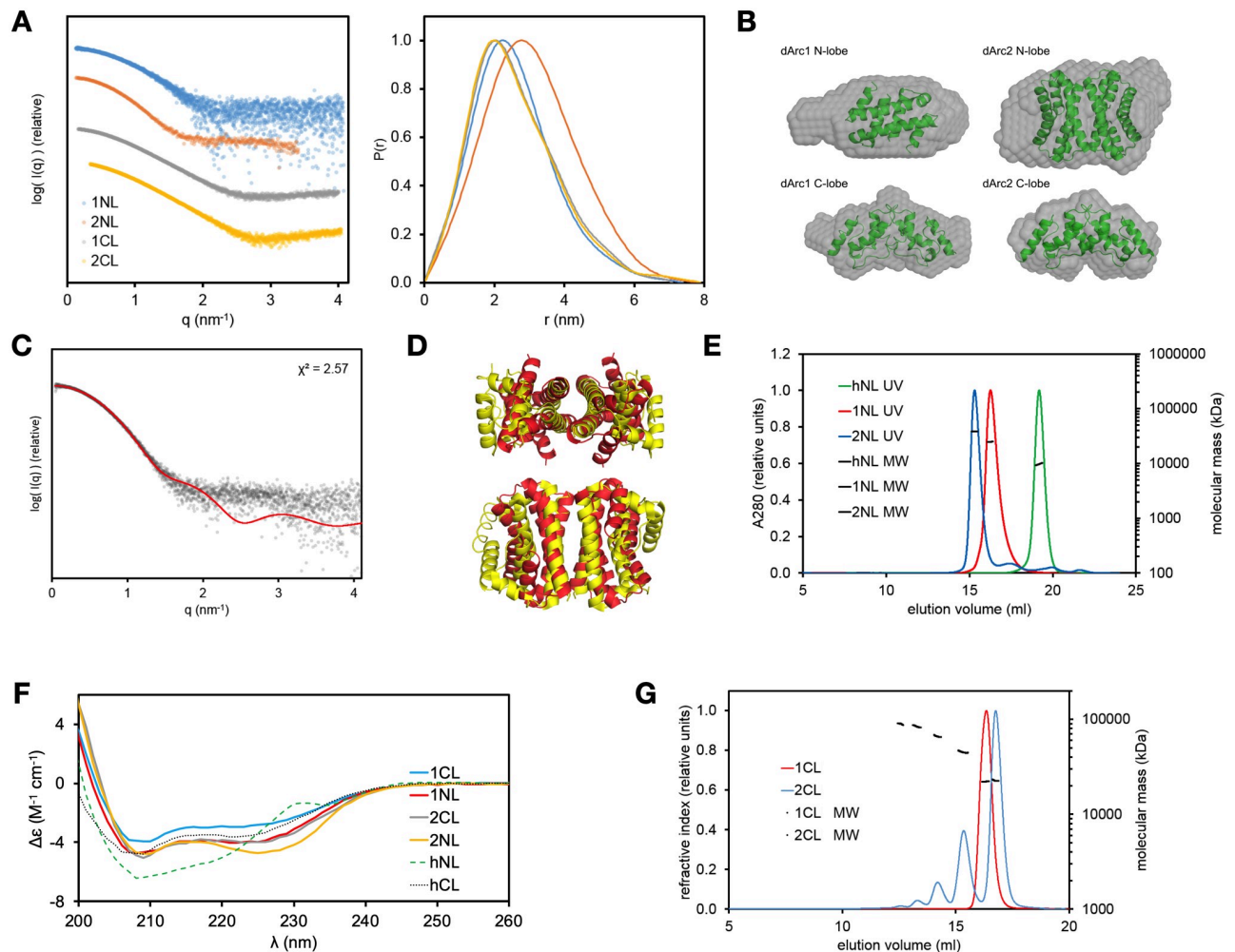
Fig 9. Solution structures of dArc lobe domains. (A) SAXS data for dArc lobes in solution (left) and distance distribution plots (right). (B) *Ab initio* models (grey spheres) of dArc lobes. The models are superimposed with the following structures (shown as cartoons): dArc2-NL dimer (overlaid on dArc1-NL), dArc1-CL dimer (on dArc1-CL), a dimer of dArc2-NL dimers (on dArc2-NL), and dArc2-CL dimer (on dArc2-CL). (C) Fit of the SAXS data for dArc2-NL (dots) with the possible tetramer of dArc2-NL seen in panel (D). (D) Tetrameric assembly of the West Nile virus protein C (yellow) and an aligned structure of two dArc2-NL dimers (red) showing a possible tetrameric structure. (E) SEC-MALS for human and *Drosophila* N-lobes. (F) CD data for dArc and hArc lobes. (G) SEC-MALS for human and *Drosophila* C-lobes.

Arc-NL domains have various oligomeric states. mArc-NL is monomeric in solution [18], whereas the dArc-NL forms dimers (dArc1) and tetramers (dArc2) (Fig 9E). While the crystal structure of dArc2-NL shows a dimer, a tetrameric assembly is not present in the crystal. This is remarkable, as the sequences of dArc1-NL and dArc2-NL are very similar (Fig 10A). The

Table 2. Dimensions and oligomeric state for different dArc constructs.

Protein	R_g (nm)	D_{max} (nm)	MM from SAXS envelope volume (kDa)	MW from Bayesian estimation based on SAXS data (kDa)	MM from MALS (kDa)	Monomeric mass (kDa)	Oligomeric state
dArc1-NL	2.0	7.4	19.6	18.7	24	8.2	Dimer
dArc2-NL	2.4	7.8	34.0	31.7	37	8.4	Tetramer
dArc1-CL	2.0	7.1	17.6	17.4	22	11	Dimer
dArc2-CL	2.0	7.9	16.2	19.9	23	10	Dimer

A

```
dArc1NL  FGGTRDHDVVEEFIGNIETYKDVEGISDENALKGISLLFYGMASTWWQGVRKEATTWKEAIALIREHFSPT
dArc2NL  FSGQRDHDAVDEFINAVETYKEVEGISDKDALKGLPLLFKSIAVVWWKGVRRDAKTWSDALQLLRDHFSPT
         *.* ****.*:***. .:****:******:.:****. *** .:* .**:***::*.**.:*: *:*:*****
```

B

C
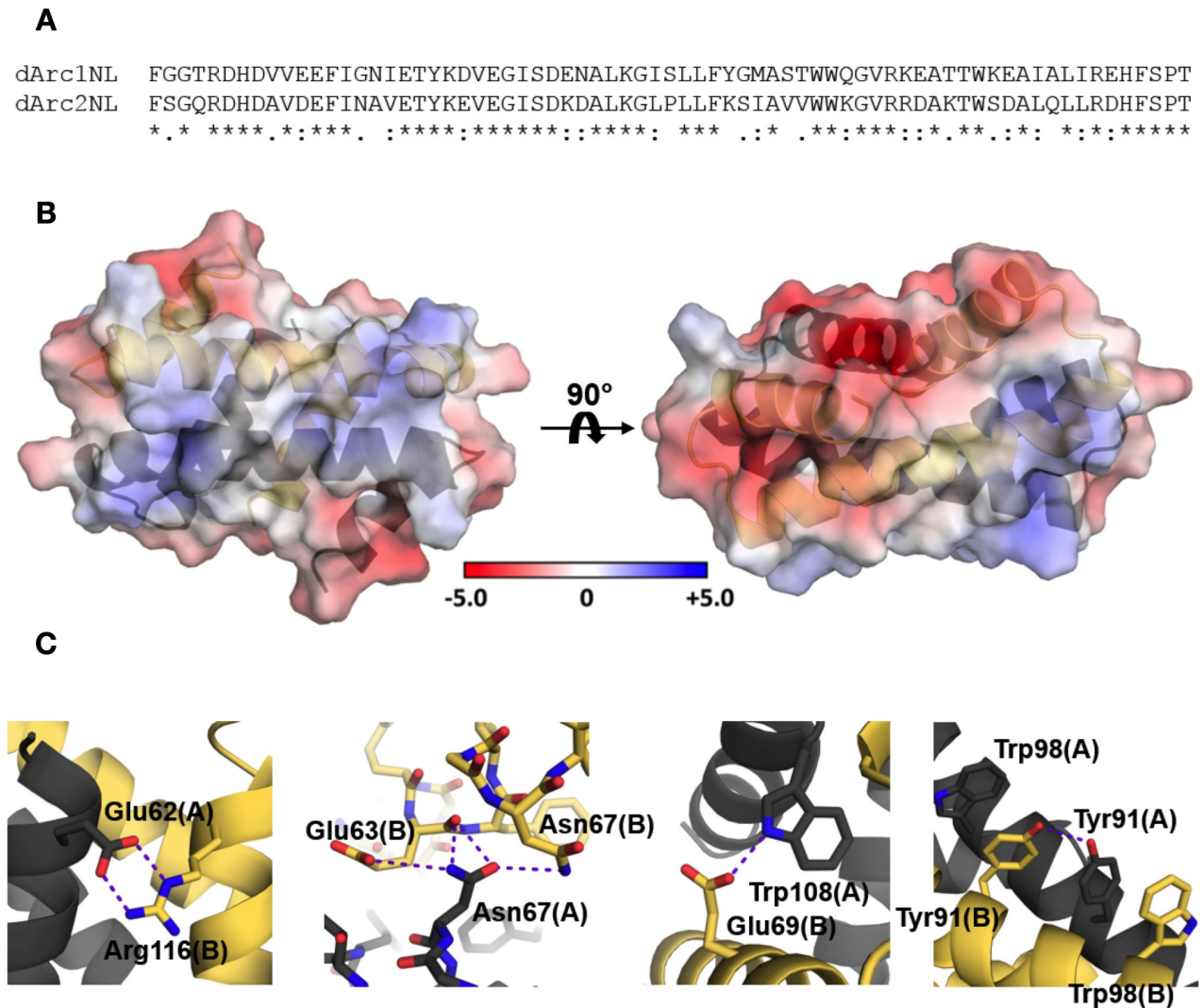
Fig 10. dArc1-NL homology model. (A) Sequence alignment between dArc1-NL and dArc2-NL. (B) The homology model of dArc1-NL displays contrasting electrostatic surface potential, where the highly positive character of dArc2-NL along α2 and α2' (Fig 2) is replaced with a more modest surface potential. (C) Additional monomer-monomer interactions observed in the dArc1-NL model, not observed in the dArc2-NL crystal structure. Polar interactions are shown with purple dashed lines.

https://doi.org/10.1371/journal.pone.0251459.g010

dArc2-NL dimer surface is electrostatically polarized (Fig 2). By threading the sequence of dArc1 onto the dArc2-NL crystal structure, the variable residues are mainly located on the surface of the long helices (α2 and α2', Fig 2), suggesting that this region is responsible for dArc2-NL tetramerization. Presumably, a tetramer of dArc2-NL is achieved either by formation of a four-helix interface bundle, similar to the West-Nile virus coat protein (Figs 4A and 9D) or *via* interaction of the contrasting surface potentials on each side of the dimer. The lack of this electrostatic polarization in dArc1-NL could be linked to oligomerization (Fig 10B). Furthermore, additional polar interactions are observed at the interface of a dArc1-NL domain-swapped homology model (Fig 10C), compared to the dArc2-NL interface, which only consists of nonpolar contacts. These different interactions could also be linked to dArc-NL isoform-specific oligomerization.

CD spectroscopy showed that all four dArc lobes are α-helical (Fig 9F), with some variations in spectral shape and amplitude. dArc2-NL has a higher 222-to-208-nm ratio compared to dArc1-NL. This could be related to differences in dimerization (domain swapping) or tetramer formation. Tetramerization may involve interactions between the long helices of dArc2-NL, and coiled-coil interactions increase the 222-to-208-nm ratio [85–87]. CD spectra of the dArc C-lobes are similar but differ in intensity, suggesting that dArc1-CL is less folded in solution, despite the very similar crystal structures. As the spectra have similar shapes and peak positions, the difference in amplitude could also reflect inaccurate concentration, for example caused by aggregation of one of the samples during the measurement. However, in line with the CD data, SEC showed (Fig 9G) a higher hydrodynamic radius for the dArc1-CL dimer. Kratky plots also indicate that dArc1-CL is more flexible than dArc2-CL. The CD spectrum of monomeric hArc-CL is similar to dArc-CL but shows less helical structure (Fig 9F). The monomeric hArc-NL has unique CD features, possibly arising from interactions between aromatic side chains. Taken together, the above data show that each of the dArc lobes has unique properties compared to each other, and to homologues.

Understanding higher-order oligomerization

We determined the crystal structure of three of the four dArc lobe domains: both C-lobes and dArc2-NL; a crystal structure for dArc1-NL could not be obtained. However, the high sequence similarity between dArc1-NL and dArc2-NL (Fig 10A) suggests that the structure of dArc1-NL is similar to dArc2-NL. Furthermore, secondary structure analysis using CD shows similar spectra for both proteins (Fig 9F), and the structure of the dArc2-NL dimer fits well with the SAXS data for dArc1-NL (Fig 9C and 9D). However, other dimeric arrangements for dArc1-NL are possible. In this respect, it is interesting to note the loss of a conserved Gly residue in the canonical α2-α3 loop in dArc2-NL, which could be related to the extension of the dArc2-NL helix. Replacement of a Gly residue in such a loop is a common means to induce domain swapping [88, 89]. The recent crystal structure of dArc1-CT containing both lobes showed dArc1-NL in the orthogonal bundle fold, being similar to mArc-NL [22]. No significant interactions were observed between the NL and CL in dArc1-CA, and interlobal interactions are an unlikely cause of the different fold. Neither CD nor SAXS can determine if dArc1-NL has the same domain-swapped structure as dArc2-NL or a non-domain-swapped dimer as seen for dArc-CL.

All four lobe domains of dArc are homo-oligomeric in solution. In full-length dArc, the NL is connected with the CL, and to test for interactions between the dArc N- and C-lobes, we mixed the individual lobes and looked for complexes using SEC. No new complexes of higher molecular weight were observed (S4 Fig), indicating that the isolated NL and CL do not interact with high affinity. However, when both lobes are within the same polypeptide chain, larger assemblies do form (S4 Fig)–reflected by the insolubility of the corresponding constructs in our hands and the ability of full-length dArc to form capsids spontaneously [15, 21].

dArc sequence properties

Arc may not be a universal protein, but its history is ancient. Related proteins appear in eukaryotes, from insects to fungi and plants [90]. At the same time, Arc-like proteins are coded for by the Ty3/gypsy transposons, and its relatives appear in viral capsids. This means the domain is widespread because of duplications and movements within and between genomes, rather than its age. This invites some speculation about the history of Arc, or at least the history of the N- and C-lobes.

The NL and CL are sequence-related, suggesting a duplication. They are related to viral capsid (Gag) proteins, but the Gag protein in flavi- and other viruses has only one unit, or lobe. One might expect to see either the NL or CL by itself in some cellular organism. A long, iterated search starting from either *Drosophila* or *Rattus* full sequences only gives proteins with both NL and CL, even amongst distantly related proteins from plants. This is not surprising, also given earlier studies linking Arc evolution to the Ty3/gypsy family of retrotransposons, which have a capsid protein containing two lobe domains [83]. Database scores are such that a long weak similarity will score higher than a short hit, and one will see proteins with both lobes. The correct procedure is to do a comprehensive database search starting from the CL, retrieve and align full length sequences, and see if any are missing the NL. This should then be repeated starting from the NL. Unfortunately, this does not give a clear result. Starting from the rat Arc-CL, one can collect a set of 706 sequences with an *e*-value $\leq 1.4 \times 10^{-5}$. The set runs from mammals to insects and even the first homologues from plants (*Oryza sativa* and *Nicotiana tabacum*). We find 26 sequences with an incomplete NL. More than a third (9) of these are annotated as partial sequences. Of the remaining sequences, none are confirmed to exist, and there is no clear domain boundary in the alignment. Similar results are obtained starting from an NL sequence. This does, however, not prove conclusively that the NL and CL are always found together in eukaryotes, but according to current data, a double-lobed Ty3/gypsy capsid protein is the ancestor of Arc in the animal kingdom.

In eukaryotes from mammals to insects and plants, the overwhelming majority of Arc-like proteins have both NL and CL. This has implications for the Arc evolutionary history. It appears that a functional protein has both lobes, and there are no clear examples of a system with two copies of just an NL or CL. The obvious interpretation is that after a duplication event, the two halves adopted different roles and sites in the two halves have experienced differential evolutionary pressure, as suggested by the conservation plots (Fig 5A) and the observation [21] that both dArc lobes are necessary for capsid assembly.

It was suggested that Arc entered the realm of animals *via* two separate integration events of distinct Ty3/gypsy retrotransposons into vertebrates and insects [16]. The earlier work relied on a small set of DNA sequences. When we align larger sets of proteins from mammals, insects, birds, and plants, one always finds the insect sequences forming a close group including both dArc1 and dArc2. Even a cursory look at the alignment (Fig 9A) shows the similarity of dArc1 and dArc2. This leads to a parsimonious interpretation. A protein containing both NL and CL evolved in retrotransposons, leading to the tandem arrangement always seen in Arc. Arc has had much time to evolve between insects and mammals; a duplication has led to the multiple copies seen in insects, and mArc has gained an NT from a currently unknown source. The properties of the individual lobes in dArc and mArc have diverged, despite having the same fold, and mArc-NT has taken over some of the functions intrinsic to the lobes of dArc, such as oligomerization.

Functional considerations

Arc is critical to the nervous system, but the protein fold is related to capsid proteins and, at least at the sequence level, even to proteins found in plants. A functional property of the mArc-NL is a peptide binding site, shown to interact with several proteins [12, 17]. Protein ligand binding could be a way of regulating Arc oligomerization [84], its function in the PSD organization, and/or capsid assembly. Using ITC, we tested if the dArc N-lobes bind the stargazin peptide like the mArc-NL (Fig 11) [12]. The peptide binds to hArc-NL, but not to the two dArc N-lobes, and the peptide binding site is not conserved. In the domain-swapped structure of dArc2-NL, the putative peptide binding site would be buried within the fold.
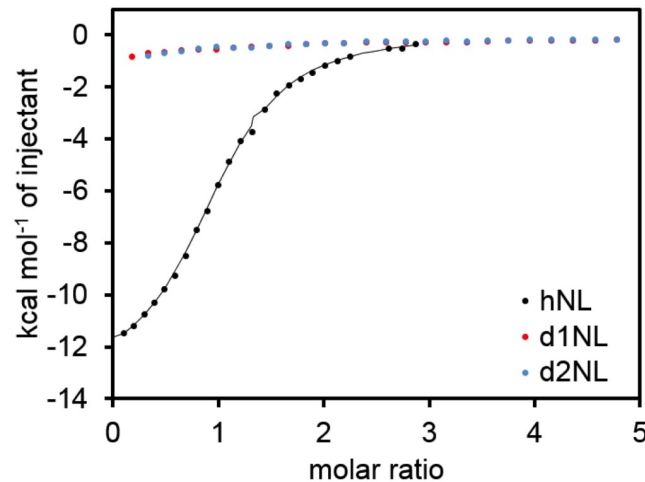
Fig 11. Binding of human and *Drosophila* Arc N-lobes to a Stargazin peptide.

Whether dArc N-lobes bind other peptides/proteins apart from the dArc N-terminal tail, and how this process might affect capsid formation, remains to be studied. Given that the dArc N-lobe forms pentamers and hexamers in the capsid [21], and that the site corresponding to the mArc peptide binding site is in the middle of these assemblies, it seems likely that external protein ligand binding to the same site would be mutually exclusive with capsid formation. Such aspects remain to be studied in detail for both dArc and mArc.

Conclusions

We have shown that both lobes of dArc1 and dArc2 are oligomeric in solution. The isolated lobes of the C-terminal domain of mammalian Arc do not exhibit the same propensity for oligomerization [18]. Absent in both dArc isoforms is the Arc-NT, which is involved in capsid formation and likely mediates interactions of mammalian Arc with negatively charged membranes [18, 19]. Therefore, oligomerization of the dArc lobes likely represents functional compensation for the lack of the N-terminal domain. On the other hand, it is possible that the N-terminal predicted helix in dArc1 and dArc2, which forms spike structures in the capsid [21], shares functions with the mArc N-terminal domain, such as lipid membrane binding as shown for hArc-NT [18].

Furthermore, we presented a novel dimeric state of dArc2-NL. This domain-swapped dimer could have a role in the non-capsid functions of dArc. It shares structural similarity with nucleotide- and membrane-interacting proteins, suggesting a related function. How homotetramerization of this lobe, observed in solution, might affect its function, remains to be studied. Overall, the strikingly different behaviour of the purified lobe domains from dArc and mArc points towards different mechanisms in their molecular function and oligomeric assembly. Our data shed light on the individual lobes as small building blocks of dArc capsids, and they complement seminal recent work on dArc structure within capsids [21] and in isolation, as a dimer [22].

Supporting information

S1 Table. The expression constructs used in the current study.
(PDF)

S1 Fig. Crystal packing in dArc2-NL.
(PDF)

S2 Fig. Crystal packing in dArc1-CL.
(PDF)

S3 Fig. Crystal packing in dArc2-CL.
(PDF)

S4 Fig. Mixing of dArc-NL and dArc-CL does not result in larger oligomers.
(PDF)

S1 File.
(PDF)

S2 File.
(PDF)

S3 File.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Erik I. Hallin, Andrew E. Torda, Clive R. Bramham, Petri Kursula.

Data curation: Sigurbjörn Markússon, Lev Böttger, Andrew E. Torda, Petri Kursula.

Formal analysis: Erik I. Hallin, Sigurbjörn Markússon, Lev Böttger, Andrew E. Torda, Petri Kursula.

Funding acquisition: Clive R. Bramham, Petri Kursula.

Investigation: Erik I. Hallin, Sigurbjörn Markússon, Lev Böttger, Andrew E. Torda, Petri Kursula.

Methodology: Erik I. Hallin, Lev Böttger, Andrew E. Torda.

Project administration: Clive R. Bramham, Petri Kursula.

Resources: Andrew E. Torda.

Software: Andrew E. Torda.

Supervision: Andrew E. Torda, Clive R. Bramham, Petri Kursula.

Validation: Erik I. Hallin, Sigurbjörn Markússon, Lev Böttger, Andrew E. Torda, Petri Kursula.

Visualization: Erik I. Hallin, Sigurbjörn Markússon, Andrew E. Torda.

Writing – original draft: Erik I. Hallin, Andrew E. Torda, Petri Kursula.

Writing – review & editing: Sigurbjörn Markússon, Lev Böttger, Clive R. Bramham, Petri Kursula.

References

1. Bramham CR, Alme MN, Bittins M, Kuipers SD, Nair RR, Pai B et al. (2010) The Arc of synaptic memory. Exp Brain Res 200: 125–140. https://doi.org/10.1007/s00221-009-1959-2 PMID: 19690847

2. Shepherd JD, Bear MF (2011) New views of Arc, a master regulator of synaptic plasticity. Nat Neurosci 14: 279–284. https://doi.org/10.1038/nn.2708 PMID: 21278731

3. Chowdhury S, Shepherd JD, Okuno H, Lyford G, Petralia RS, Plath N et al. (2006) Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. Neuron 52: 445–459. https://doi.org/10.1016/j.neuron.2006.08.033 PMID: 17088211

4. Rial Verde EM, Lee-Osbourne J, Worley PF, Malinow R, Cline HT (2006) Increased expression of the immediate-early gene arc/arg3.1 reduces AMPA receptor-mediated synaptic transmission. Neuron 52: 461–474. https://doi.org/10.1016/j.neuron.2006.09.031 PMID: 17088212

5. Shepherd JD, Rumbaugh G, Wu J, Chowdhury S, Plath N, Kuhl D et al. (2006) Arc/Arg3.1 mediates homeostatic synaptic scaling of AMPA receptors. Neuron 52: 475–484. https://doi.org/10.1016/j.neuron.2006.08.034 PMID: 17088213

6. Messaoudi E, Kanhema T, Soulé J, Tiron A, Dagyte G, da Silva B et al. (2007) Sustained Arc/Arg3.1 synthesis controls long-term potentiation consolidation through regulation of local actin polymerization in the dentate gyrus in vivo. J Neurosci 27: 10445–10455. https://doi.org/10.1523/JNEUROSCI.2883-07.2007 PMID: 17898216

7. Peebles CL, Yoo J, Thwin MT, Palop JJ, Noebels JL, Finkbeiner S (2010) Arc regulates spine morphology and maintains network stability in vivo. Proc Natl Acad Sci U S A 107: 18173–18178. https://doi.org/10.1073/pnas.1006546107 PMID: 20921410

8. Zhang H, Bramham CR (2020) Arc/Arg3.1 function in long-term synaptic plasticity: Emerging mechanisms and unresolved issues. Eur J Neurosci https://doi.org/10.1111/ejn.14958 PMID: 32888346

9. Korb E, Wilkinson CL, Delgado RN, Lovero KL, Finkbeiner S (2013) Arc in the nucleus regulates PML-dependent GluA1 transcription and homeostatic plasticity. Nat Neurosci 16: 874–883. https://doi.org/10.1038/nn.3429 PMID: 23749147

10. Wee CL, Teo S, Oey NE, Wright GD, VanDongen HM, VanDongen AM (2014) Nuclear Arc Interacts with the Histone Acetyltransferase Tip60 to Modify H4K12 Acetylation(1,2,3). eNeuro 1: https://doi.org/10.1523/ENEURO.0019-14.2014 PMID: 26464963

11. Jackson AC, Nicoll RA (2011) Stargazing from a new vantage—TARP modulation of AMPA receptor pharmacology. J Physiol 589: 5909–5910. https://doi.org/10.1113/jphysiol.2011.223495 PMID: 22174138

12. Zhang W, Wu J, Ward MD, Yang S, Chuang YA, Xiao M et al. (2015) Structural basis of arc binding to synaptic proteins: implications for cognitive disease. Neuron 86: 490–500. https://doi.org/10.1016/j.neuron.2015.03.030 PMID: 25864631

13. Zhao Y, Chen S, Yoshioka C, Baconguis I, Gouaux E (2016) Architecture of fully occupied GluA2 AMPA receptor-TARP complex elucidated by cryo-EM. Nature 536: 108–111. https://doi.org/10.1038/nature18961 PMID: 27368053

14. Nikolaienko O, Patil S, Eriksen MS, Bramham CR (2018) Arc protein: a flexible hub for synaptic plasticity and cognition. Semin Cell Dev Biol 77: 33–42. https://doi.org/10.1016/j.semcdb.2017.09.006 PMID: 28890419

15. Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V, Thomson T (2018) Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. Cell 172: 262–274.e11. https://doi.org/10.1016/j.cell.2017.12.022 PMID: 29328915

16. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J et al. (2018) The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. Cell 173: 275. https://doi.org/10.1016/j.cell.2018.03.024 PMID: 29570995

17. Hallin EI, Bramham CR, Kursula P (2021) Structural properties and peptide ligand binding of the capsid homology domains of human Arc. Biochem Biophys Rep https://doi.org/10.1016/j.bbrep.2021.100975 PMID: 33732907

18. Hallin EI, Eriksen MS, Baryshnikov S, Nikolaienko O, Grødem S, Hosokawa T et al. (2018) Structure of monomeric full-length ARC sheds light on molecular flexibility, protein interactions, and functional modalities. J Neurochem 147: 323–343. https://doi.org/10.1111/jnc.14556 PMID: 30028513

19. Eriksen MS, Nikolaienko O, Hallin EI, Grødem S, Bustad HJ, Flydal MI et al. (2020) Arc self-association and formation of virus-like capsids are mediated by an N-terminal helical coil motif. FEBS J https://doi.org/10.1111/febs.15618 PMID: 33175445

20. Zhang W, Chuang YA, Na Y, Ye Z, Yang L, Lin R et al. (2019) Arc Oligomerization Is Regulated by CaMKII Phosphorylation of the GAG Domain: An Essential Mechanism for Plasticity and Memory Formation. Mol Cell 75: 13–25.e5. https://doi.org/10.1016/j.molcel.2019.05.004 PMID: 31151856

21. Erlendsson S, Morado DR, Cullen HB, Feschotte C, Shepherd JD, Briggs JAG (2020) Structures of virus-like capsids formed by the Drosophila neuronal Arc proteins. Nat Neurosci 23: 172–175. https://doi.org/10.1038/s41593-019-0569-y PMID: 31907439

22. Cottee MA, Letham SC, Young GR, Stoye JP, Taylor IA (2020) Structure of Drosophila melanogaster ARC1 reveals a repurposed molecule with characteristics of retroviral Gag. Sci Adv 6: eaay6354. https://doi.org/10.1126/sciadv.aay6354 PMID: 31911950

23. van den Berg S, Löfdahl PA, Härd T, Berglund H (2006) Improved solubility of TEV protease by directed evolution. J Biotechnol 121: 291–298. https://doi.org/10.1016/j.jbiotec.2005.08.006 PMID: 16150509

24. Raasakka A, Myllykoski M, Laulumaa S, Lehtimäki M, Härtlein M, Moulin M et al. (2015) Determinants of ligand binding and catalytic activity in the myelin enzyme 2',3'-cyclic nucleotide 3'-phosphodiesterase. Sci Rep 5: 16520. https://doi.org/10.1038/srep16520 PMID: 26563764

25. Burkhardt A, Pakendorf T, Reime B, Meyer J, Fischer P, Stübe N et al. (2016) Status of the crystallography beamlines at PETRA III. The European Physical Journal Plus 131: 56.

26. Kabsch W (2010) XDS. Acta Crystallogr D Biol Crystallogr 66: 125–132. https://doi.org/10.1107/S0907444909047337 PMID: 20124692

27. Bibby J, Keegan RM, Mayans O, Winn MD, Rigden DJ (2012) AMPLE: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models. Acta Crystallogr D Biol Crystallogr 68: 1622–1631. https://doi.org/10.1107/S0907444912039194 PMID: 23151627

28. Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80: 1715–1735. https://doi.org/10.1002/prot.24065 PMID: 22411565

29. Krissinel E, Uski V, Lebedev A, Winn M, Ballard C (2018) Distributed computing for macromolecular crystallography. Acta Crystallogr D Struct Biol 74: 143–151. https://doi.org/10.1107/S2059798317014565 PMID: 29533240

30. Potterton L, Agirre J, Ballard C, Cowtan K, Dodson E, Evans PR et al. (2018) CCP4i2: the new graphical user interface to the CCP4 program suite. Acta Crystallogr D Struct Biol 74: 68–84. https://doi.org/10.1107/S2059798317016035 PMID: 29533233

31. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA (2005) Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. Acta Crystallogr D Biol Crystallogr 61: 449–457. https://doi.org/10.1107/S0907444905001307 PMID: 15805600

32. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64: 112–122. https://doi.org/10.1107/S0108767307043930 PMID: 18156677

33. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674. https://doi.org/10.1107/S0021889807021206 PMID: 19461840

34. Cowtan K (2010) Recent developments in classical density modification. Acta Crystallogr D Biol Crystallogr 66: 470–478. https://doi.org/10.1107/S090744490903947X PMID: 20383000

35. Cowtan K (2012) Completion of autobuilt protein models using a database of protein fragments. Acta Crystallogr D Biol Crystallogr 68: 328–335. https://doi.org/10.1107/S0907444911039655 PMID: 22505253

36. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr 68: 352–367. https://doi.org/10.1107/S0907444912001308 PMID: 22505256

37. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66: 486–501. https://doi.org/10.1107/S0907444910007493 PMID: 20383002

38. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN et al. (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci 27: 293–315. https://doi.org/10.1002/pro.3330 PMID: 29067766

39. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797. https://doi.org/10.1016/j.jmb.2007.05.022 PMID: 17681537

40. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM (2018) PDBsum: Structural summaries of PDB entries. Protein Sci 27: 129–134. https://doi.org/10.1002/pro.3289 PMID: 28875543

41. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38: W545–9. https://doi.org/10.1093/nar/gkq366 PMID: 20457744

42. Margraf T, Schenk G, Torda AE (2009) The SALAMI protein structure search server. Nucleic Acids Res 37: W480–4. https://doi.org/10.1093/nar/gkp431 PMID: 19465380

43. Unni S, Huang Y, Hanson RM, Tobias M, Krishnan S, Li WW et al. (2011) Web servers and services for electrostatics calculations with APBS and PDB2PQR. J Comput Chem 32: 1488–1491. https://doi.org/10.1002/jcc.21720 PMID: 21425296

44. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25: 1605–1612. https://doi.org/10.1002/jcc.20084 PMID: 15264254

45. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N et al. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47: W636–W641. https://doi.org/10.1093/nar/gkz268 PMID: 30976793

46. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46: W296–W303. https://doi.org/10.1093/nar/gky427 PMID: 29788355

47. Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM et al. (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). J Appl Crystallogr 48: 431–443. https://doi.org/10.1107/S160057671500254X PMID: 25844078

48. Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT et al. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Crystallogr 50: 1212–1225. https://doi.org/10.1107/S1600576717007786 PMID: 28808438

49. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 76: 2879–2886. https://doi.org/10.1016/S0006-3495(99)77443-6 PMID: 10354416

50. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80: 2946–2953. https://doi.org/10.1016/S0006-3495(01)76260-1 PMID: 11371467

51. Svergun DIBC, Barberato C, Koch MHJ (1995) CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. Journal of applied crystallography 28: 768–773.

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

53. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694

54. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30: 772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

55. Wong HC, Shin R, Krishna NR (2008) Solution structure of a double mutant of the carboxy-terminal dimerization domain of the HIV-1 capsid protein. Biochemistry 47: 2289–2297. https://doi.org/10.1021/bi7022128 PMID: 18220423

56. Byeon IJ, Meng X, Jung J, Zhao G, Yang R, Ahn J et al. (2009) Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. Cell 139: 780–790. https://doi.org/10.1016/j.cell.2009.10.010 PMID: 19914170

57. Ma L, Jones CT, Groesch TD, Kuhn RJ, Post CB (2004) Solution structure of dengue virus capsid protein reveals another fold. Proc Natl Acad Sci U S A 101: 3414–3419. https://doi.org/10.1073/pnas.0305892101 PMID: 14993605

58. Oliveira ERA, Mohana-Borges R, de Alencastro RB, Horta BAC (2017) The flavivirus capsid protein: Structure, function and perspectives towards drug design. Virus Res 227: 115–123. https://doi.org/10.1016/j.virusres.2016.10.005 PMID: 27751882

59. Dokland T, Walsh M, Mackenzie JM, Khromykh AA, Ee KH, Wang S (2004) West Nile virus core protein; tetramer structure and ribbon formation. Structure 12: 1157–1163. https://doi.org/10.1016/j.str.2004.04.024 PMID: 15242592

60. Ivanov D, Tsodikov OV, Kasanov J, Ellenberger T, Wagner G, Collins T (2007) Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. Proc Natl Acad Sci U S A 104: 4353–4358. https://doi.org/10.1073/pnas.0609477104 PMID: 17360528

61. Mattiroli F, Bhattacharyya S, Dyer PN, White AE, Sandman K, Burkhart BW et al. (2017) Structure of histone-based chromatin in Archaea. Science 357: 609–612. https://doi.org/10.1126/science.aaj1849 PMID: 28798133

62. Wang H, Wang M, Yang N, Xu RM (2015) Structure of the quaternary complex of histone H3-H4 hetero-dimer with chaperone ASF1 and the replicative helicase subunit MCM2. Protein Cell 6: 693–697. https://doi.org/10.1007/s13238-015-0190-0 PMID: 26186914

63. Xie X, Kokubo T, Cohen SL, Mirza UA, Hoffmann A, Chait BT et al. (1996) Structural similarity between TAFs and the heterotetrameric core of the histone octamer. Nature 380: 316–322. https://doi.org/10.1038/380316a0 PMID: 8598927

64. Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S et al. (2006) Structure of the forkhead domain of FOXP2 bound to DNA. Structure 14: 159–166. https://doi.org/10.1016/j.str.2005.10.005 PMID: 16407075

65. Moldoveanu T, Liu Q, Tocilj A, Watson M, Shore G, Gehring K (2006) The X-ray structure of a BAK homodimer reveals an inhibitory zinc binding site. Mol Cell 24: 677–688. https://doi.org/10.1016/j.molcel.2006.10.014 PMID: 17157251

66. Brouwer JM, Westphal D, Dewson G, Robin AY, Uren RT, Bartolo R et al. (2014) Bak core and latch domains separate during activation, and freed core domains form symmetric homodimers. Mol Cell 55: 938–946. https://doi.org/10.1016/j.molcel.2014.07.016 PMID: 25175025

67. Cowan AD, Smith NA, Sandow JJ, Kapp EA, Rustam YH, Murphy JM et al. (2020) BAK core dimers bind lipids and can be bridged by them. Nat Struct Mol Biol https://doi.org/10.1038/s41594-020-0494-5 PMID: 32929280

68. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH (2014) The SARS coronavirus nucleocapsid protein—forms and functions. Antiviral Res 103: 39–50. https://doi.org/10.1016/j.antiviral.2013.12.009 PMID: 24418573

69. Sandman K, Reeve JN (2006) Archaeal histones and the origin of the histone fold. Curr Opin Microbiol 9: 520–525. https://doi.org/10.1016/j.mib.2006.08.003 PMID: 16920388

70. Venkatesh S, Workman JL (2015) Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol 16: 178–189. https://doi.org/10.1038/nrm3941 PMID: 25650798

71. Chu YP, Chang CH, Shiu JH, Chang YT, Chen CY, Chuang WJ (2011) Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: insight into its domain swapping and DNA binding. Protein Sci 20: 908–924. https://doi.org/10.1002/pro.626 PMID: 21416545

72. Volle C, Dalal Y (2014) Histone variants: the tricksters of the chromatin world. Curr Opin Genet Dev 25: 8–14,138. https://doi.org/10.1016/j.gde.2013.11.006 PMID: 24463272

73. Colpitts TM, Barthel S, Wang P, Fikrig E (2011) Dengue virus capsid protein binds core histones and inhibits nucleosome formation in human liver cells. PLoS One 6: e24365. https://doi.org/10.1371/journal.pone.0024365 PMID: 21909430

74. Salery M, Dos Santos M, Saint-Jour E, Moumné L, Pagès C, Kappès V et al. (2017) Activity-Regulated Cytoskeleton-Associated Protein Accumulates in the Nucleus in Response to Cocaine and Acts as a Brake on Chromatin Remodeling and Long-Term Behavioral Alterations. Biol Psychiatry 81: 573–584. https://doi.org/10.1016/j.biopsych.2016.05.025 PMID: 27567310

75. Czabotar PE, Westphal D, Dewson G, Ma S, Hockings C, Fairlie WD et al. (2013) Bax crystal structures reveal how BH3 domains activate Bax and nucleate its oligomerization to induce apoptosis. Cell 152: 519–531. https://doi.org/10.1016/j.cell.2012.12.031 PMID: 23374347

76. Sibanda BL, Thornton JM (1985) Beta-hairpin families in globular proteins. Nature 316: 170–174. https://doi.org/10.1038/316170a0 PMID: 4010788

77. Duddy WJ, Nissink JW, Allen FH, Milner-White EJ (2004) Mimicry by asx- and ST-turns of the four main types of beta-turn in proteins. Protein Sci 13: 3051–3055. https://doi.org/10.1110/ps.04920904 PMID: 15459339

78. Ivanov D, Stone JR, Maki JL, Collins T, Wagner G (2005) Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. Mol Cell 17: 137–143. https://doi.org/10.1016/j.molcel.2004.12.015 PMID: 15629724

79. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J et al. (2013) Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature 497: 643–646. https://doi.org/10.1038/nature12162 PMID: 23719463

80. Worthylake DK, Wang H, Yoo S, Sundquist WI, Hill CP (1999) Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. Acta Crystallogr D Biol Crystallogr 55: 85–92. https://doi.org/10.1107/S0907444998007689 PMID: 10089398

81. Obal G, Trajtenberg F, Carrión F, Tomé L, Larrieux N, Zhang X et al. (2015) STRUCTURAL VIROL-OGY. Conformational plasticity of a native retroviral capsid revealed by x-ray crystallography. Science 349: 95–98. https://doi.org/10.1126/science.aaa5182 PMID: 26044299

82. Bailey GD, Hyun JK, Mitra AK, Kingston RL (2009) Proton-linked dimerization of a retroviral capsid protein initiates capsid assembly. Structure 17: 737–748. https://doi.org/10.1016/j.str.2009.03.010 PMID: 19446529

83. Dodonova SO, Prinz S, Bilanchone V, Sandmeyer S, Briggs JAG (2019) Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. Proc Natl Acad Sci U S A 116: 10048–10057. https://doi.org/10.1073/pnas.1900931116 PMID: 31036670

84. Nielsen LD, Pedersen CP, Erlendsson S, Teilum K (2019) The Capsid Domain of Arc Changes Its Oligomerization Propensity through Direct Interaction with the NMDA Receptor. Structure 27: 1071–1081. e5. https://doi.org/10.1016/j.str.2019.04.001 PMID: 31080121

85. Alfadhli A, Steel E, Finlay L, Bächinger HP, Barklis E (2002) Hantavirus nucleocapsid protein coiled-coil domains. J Biol Chem 277: 27103–27108. https://doi.org/10.1074/jbc.M203395200 PMID: 12019266

86. Kammerer RA, Schulthess T, Landwehr R, Lustig A, Engel J, Aebi U et al. (1998) An autonomous folding unit mediates the assembly of two-stranded coiled coils. Proc Natl Acad Sci U S A 95: 13419–13424. https://doi.org/10.1073/pnas.95.23.13419 PMID: 9811815

87. Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH et al. (2007) Molecular basis of coiled-coil formation. Proc Natl Acad Sci U S A 104: 7062–7067. https://doi.org/10.1073/pnas.0700321104 PMID: 17438295

88. Bhargav SP, Vahokoski J, Kallio JP, Torda AE, Kursula P, Kursula I (2015) Two independently folding units of Plasmodium profilin suggest evolution via gene fusion. Cell Mol Life Sci 72: 4193–4203. https://doi.org/10.1007/s00018-015-1932-0 PMID: 26012696

89. Han H, Kursula P (2014) Periaxin and AHNAK nucleoprotein 2 form intertwined homodimers through domain swapping. J Biol Chem 289: 14121–14131. https://doi.org/10.1074/jbc.M114.554816 PMID: 24675079

90. Smyth DR, Kalitsis P, Joseph JL, Sentry JW (1989) Plant retrotransposon from Lilium henryi is related to Ty3 of yeast and the gypsy group of Drosophila. Proc Natl Acad Sci U S A 86: 5015–5019. https://doi.org/10.1073/pnas.86.13.5015 PMID: 2544887