Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/1/eaay6354/DC1

Supplementary Materials for

Structure of *Drosophila melanogaster* ARC1 reveals a repurposed molecule with characteristics of retroviral Gag

Matthew A. Cottee, Suzanne C. Letham, George R. Young, Jonathan P. Stoye, Ian A. Taylor*

*Corresponding author. Email: ian.taylor@crick.ac.uk

Published 1 January 2020, *Sci. Adv.* **6**, eaay6354 (2020)
DOI: 10.1126/sciadv.aay6354

This PDF file includes:

Table S1. dARC1 CA statistics of data collection, phasing, and refinement.

	dARC1 CA oP (Se-Met, peak)	dARC1 CA oP (Se-Met, lrem)	dARC1 CA oP (Se-Met, lrem STARANISO)	dARC1 CA hP (Se-Met, anom)
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P6_122$
Cell dimensions				
a, b, c (Å)	62.43, 70.715, 142.07	62.35, 70.39, 142.84	62.33, 70.39, 142.82	62.18, 62.18, 401.02
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 120
Wavelength (Å)	0.9794	0.9840	0.9840	0.9794
Resolution (Å)	47.36-2.06 (2.12-2.06)	142.84-1.70 (1.74-1.70)	71.41-1.55 (1.70-1.55)	57.29-2.14 (2.20-2.14)
Anisotropic diff. limits (a*, b*, c*)	-	-	1.55, 2.40, 1.62	
Unique reflections	39409 (2872)†	70026 (5125)	48232 (2412)	26972 (1886)
R_{meas} (%)	21.5 (254.6)‡	13.5 (105.8)	9.3 (150.1)	12.2 (111.0)‡
R_{pim} (%)	8.2 (106.1)‡	4.3 (33.9)	2.9 (47.8)	2.5 (31.6)‡
$CC_{1/2}$	0.998 (0.595)	0.998 (0.568)	0.999 (0.557)	1.000 (0.456)
I/σ(I)	10.3 (1.1)	9.1 (1.5)	16.4 (1.5)	22.0 (2.7)
Completeness (spherical, %)	99.5 (95.4)	100.0 (100.0)	52.9 (11.5)	99.9 (99.3)
Completeness (ellipsoidal, %)	-	-	84.1 (66.0)	-
Multiplicity	12.9 (10.5)	9.6 (9.5)	10.0 (9.7)	40.8 (22.4)
Anom Multiplicity	6.8 (5.5)	-	-	22.4 (11.8)
Phasing				
Significant anomalous signal to: (Aimless)	2.82 Å	-		2.59 Å
No. sites (found/expected)	7/4	-		10/4
Mean FOM (Phenix)	0.28	-		0.324
Local rms density corr (Phenix)	0.89			0.83
Refinement				
Resolution (Å)			71.41-1.70 (1.74-1.70)	53.86-2.30 (2.36-2.30)
refl working/free			41243/2241 (1680/69)	36685/1844§ (2798/121)
R_{work}/R_{free}/Test set size (%)			18.1/21.3/4.9 (25.2/28.2/3.9)	24.2/27.8/4.8 (31.6/31.8/4.1)
No residues/atoms				
Protein			328/2730	321/2658
Ligands			5	
Water			391	20
B-factors (Å²)				
Wilson			19.74	44.15
Protein			23.40	66.04
Ligands			33.17	
Water			31.52	46.10
Overall			24.43	65.89
Geometry				
Bond lengths RMSD (Å)			0.010	0.004
Bond angles RMSD (deg)			1.71	0.697
Ramachandran Outliers			0 %	0 %
Ramachandran Favoured			99.38 %	97.79 %
Molprobity score (N number/percentile)			0.92 (9248/100th)	1.44 (8909/99th)

†Values in parenthesis refer to the highest resolution shell, ‡Values quoted for within I^+/I^-, §Friedel pairs separated.

Table S2. Hydrodynamic parameters of dARC1 CA.

Protein	dARC1 CA
Protein Parameter	
v ($mL.g^{-1}$)	0.731
ρ ($g.mL^{-1}$)	1.005
[a]M_r	19,633
ε_{280} (M^{-1} cm^{-1})	23,600
Sed velocity	
C_{range} (μM)	25-100
[b]$S_{20,w}$ ($x10^{13}$) sec	2.92±0.03
[c]$S_{20,w}$ (calc) ($x10^{13}$) sec	2.93
[d]$D_{20,w}$ ($x10^{7}$) cm^2 sec^{-1}	6.95±0.05
[e]$D_{20,w}$ (calc) ($x10^{7}$) cm^2 sec^{-1}	6.77
[f]M_w (S/D) kD	38.2±1
[g]f/f_0 ($S_{20,w}$)	1.41
[h]f/f_0 ($D_{20,w}$)	1.37
[i]M_w C(S) kD	38.7±3
[j]f/f_0 C(S)	1.41
[k]RMSD C(S)	0.005 – 0.008
Sed eqm	
C_{range} (μM)	25-70
[l]M_w kD	38.9±1
[m]RMSD	0.002 – 0.004
[n]χ^2	0.31

[a]Molar mass calculated from the protein sequence after removal of the C-terminal His-tag
[b]Sedimentation coefficient at standard conditions (water @ 20° C) determined by combining all data from discrete component and C(S) analysis.
[c]Sedimentation coefficient calculated using HYDROPRO-10 from the co-ordinates of the dARC1 CA crystal structure using an atomic element radius (AER) of 2.84 Å.
[d]Translational diffusion coefficient for dARC1 CA determined from combined data from discrete component analysis.
[e]Translational diffusion coefficient calculated from the crystal structure using HYDROPRO-10, AER = 2.84 Å.
[f]The weight averaged molecular weight from discrete component analysis. Error is the range of three measurements
[g]The frictional ratio calculated from $S_{20,w}$ from discrete component analysis.
[h]The frictional ratio calculated from $D_{20,w}$ from discrete component analysis.
[i]The weight averaged molecular weight from C(S) analysis. Error is the range of three measurements
[j]The weight-averaged frictional ratio from the best fit C(S) distribution function.
[k]The range of the rms deviations observed when data were fitted using a continuous sedimentation coefficient distribution model.
[l]The weight averaged molecular weight from Global SE analysis. The error is the range of molecular weights observed for each multi-speed sample when fitted individually.
[m]The range of the rms deviations observed for each multi-speed sample when fitted individually.
[n]The reduced chi-squared for the global fitting of all multispeed data.

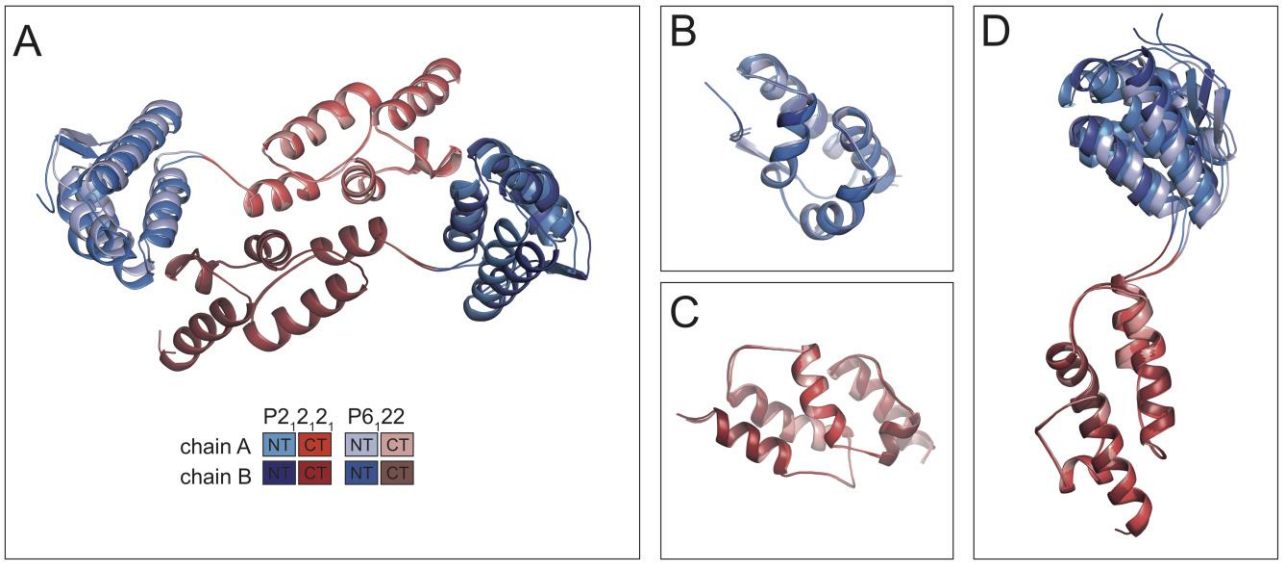

Fig. S1. Crystal structures of dARC1 CA. (**A**) Structural alignment of orthorhombic ($P2_12_12_1$) and hexagonal ($P6_122$) crystal forms of dARC1 CA that contain an almost identical dimer. The ASUs are shown in cartoon representation, aligned using the dimeric CTDs (RMSD of 0.247 Å across 133 Cα pairs). NTDs are coloured in blue shades, CTDs in red shades, according to the legend. (**B**) Alignment of all four NTDs only, average RMSD 0.34 ± 0.09Å over 66±2 Cαs. (**C**) Alignment of all four CTDs only, average rmsd 0.32 ± 0.12 Å over 74±4 Cαs. (**D**) Alignment of all four chains through the CTD, highlighting that structural differences between chains are due to slight flexibility of the NTD-CTD linker region.

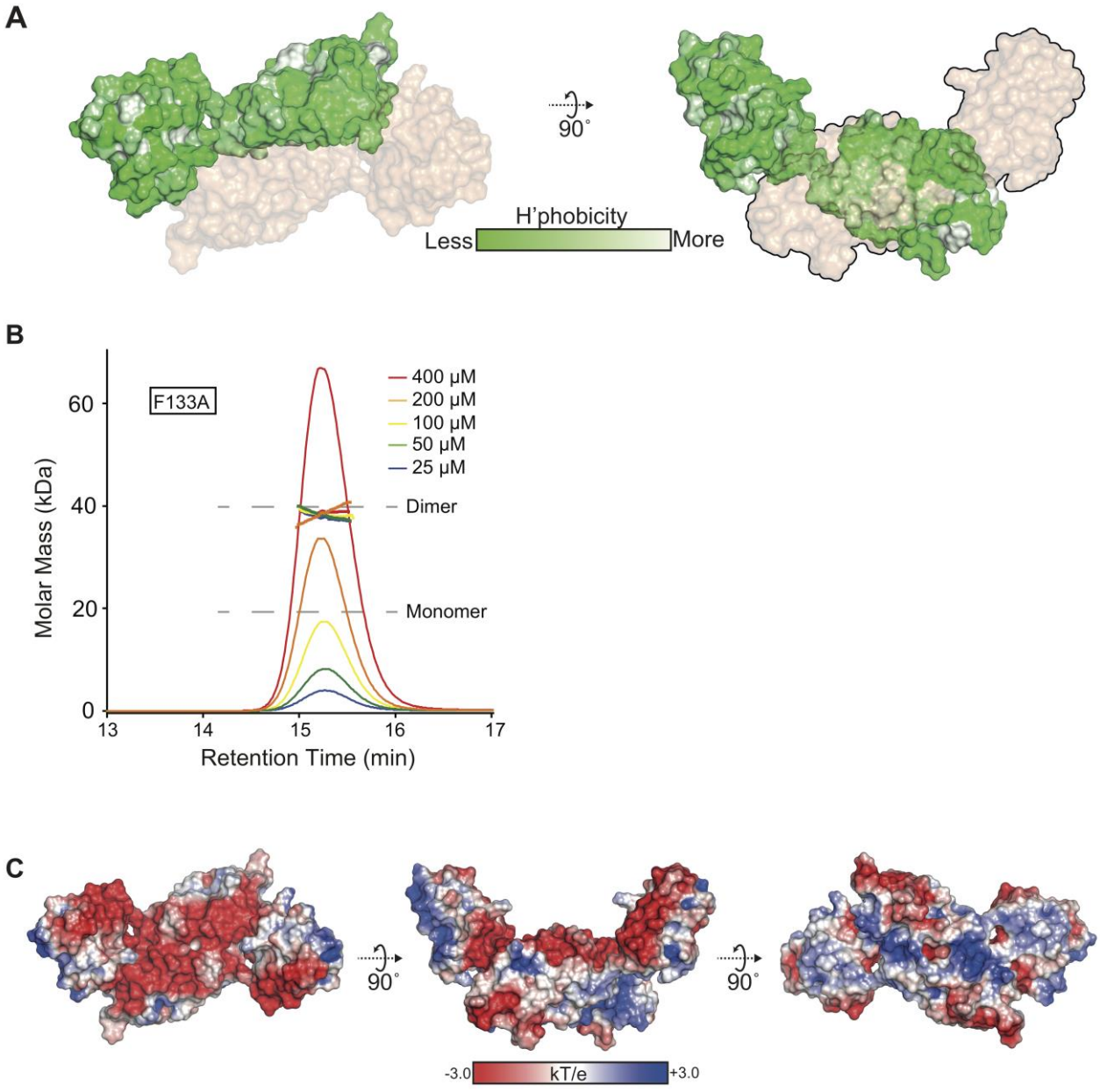

Fig. S2. dARC1 CA dimer. (**A**) Surface representation of dARC1 CA-CTD dimer displaying the distribution of surface hydrophobicity/hydrophilicity calculated using the pymol script (https://pymolwiki.org/index.php/Color_h). Greater Hydrophilicity is represented by darker green shading, the lighter and non-coloured regions represent the most hydrophobic areas of the molecule. The orientation in the left- and right-hand panels is the same as in **Fig. 1A** with monomer A displaying surface hydrophobicity, and monomer B shown as a wheat surface. **(B)** SEC-MALLS analysis of dARC1 CA (F133A). The sample loading concentrations were 400 µM (8 mg/mL) (red), 200 µM (4 mg/mL)

(orange), 100 µM (2 mg/mL) (yellow), 50 µM (1 mg/mL) (green) and 25 µM (0.5 mg/mL) (blue). The differential refractive index (dRI) is plotted against column retention time and the molar mass, determined at 1-second intervals throughout the elution of each peak, is plotted as points. The dARC1 CA monomer and dimer molecular mass is indicated with the grey dashed lines. (**C**) Electrostatic surface potential of the dARC1 CA dimer, calculated with APBS. The model was modified to replace SeMet with native Met residues. The view in the left-hand and centre panels is the same as in **A**. The right-hand panel is a view from the C-terminal “underside” of the dimer.

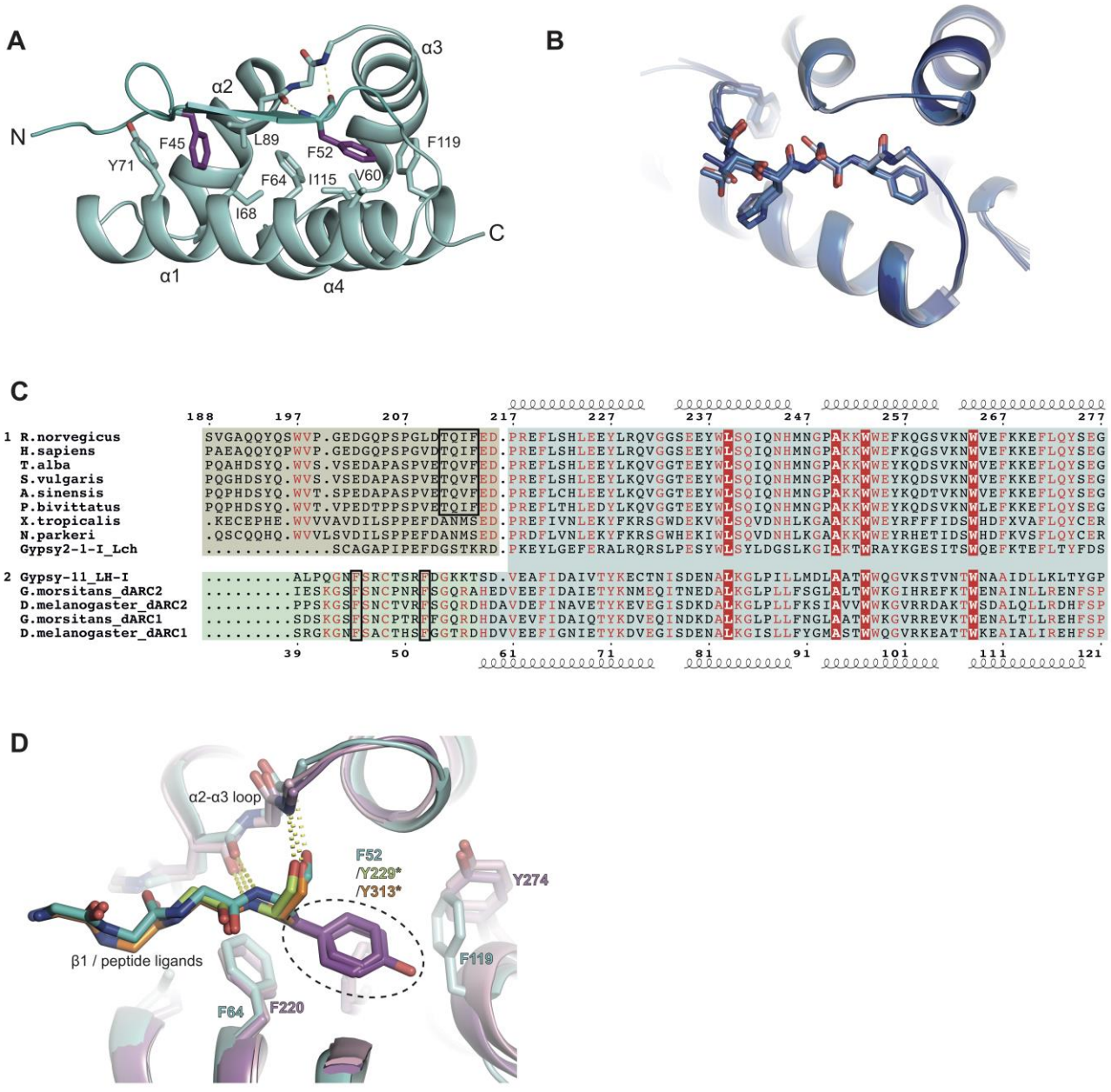

Fig. S3. Comparison of dARC1and mam-ARC CA-NTDs. (**A**) The backbone of the dARC1 CA-NTD is shown in cartoon representation in cyan. Residues that contribute to the hydrophobic interface between the core domain, and the native N-terminal strand are shown in stick representation, coloured by atom type, with the conserved aromatic residues F45 and F52 that are buried in the interface coloured purple. The mainchain hydrogen bonding interactions between the F52 backbone amide and carbonyl with the carbonyl of L89 and the amide of Y91 are shown as dashed lines. (**B**) Overlay of the four dARC1 CA NTDs from the two crystal structures, shown in different shades of blue as in

fig S1B, highlighting the identical conformation of the native N-terminal strand (sticks). (**C**) Multiple sequence alignment of NTDs of ARC and dARC CA sequences. Group 1 contains tetrapod ARC (tARC) sequences and the closely related *L. ch* Gypsy2 retrotransposon. Above, rARC secondary structure, numbers according to the rARC (*R. norvegicus*) sequence). Group 2 contains dARC CA sequences and the closely related *L. h* Gypsy11 retrotransposon. Below, dARC1 secondary structure, numbers according to the dARC1 (*D. melanogaster*) sequence). Red box, white text, invariant residues shared between groups. Red text, residues conserved within a group. Blue highlight, alpha-helical core which is conserved between the two groups. Yellow highlight, native N-terminal strand which is conserved within group 1. TQIF motif is boxed in black. Green highlight, native N-terminal strand which is conserved within group 2. Conserved, strand-burying aromatic residues are boxed in black. (**D**) Close-up view of the dARC1 CA-NTD (cyan), aligned with the rARC CA NTD-CaMK2B (light pink-orange) and rARC CA NtD-TARPγ2 (dark pink-lime) structures, showing identical mechanisms of strand burial in the core domain. The native strand (dARC1) or rARC-bound peptides (TARPγ2 and CaMK2B) as well as the NTD α2-α3 loop are shown in stick representation, the equivalent strand-anchoring aromatic residue (circled) in each case is coloured purple.

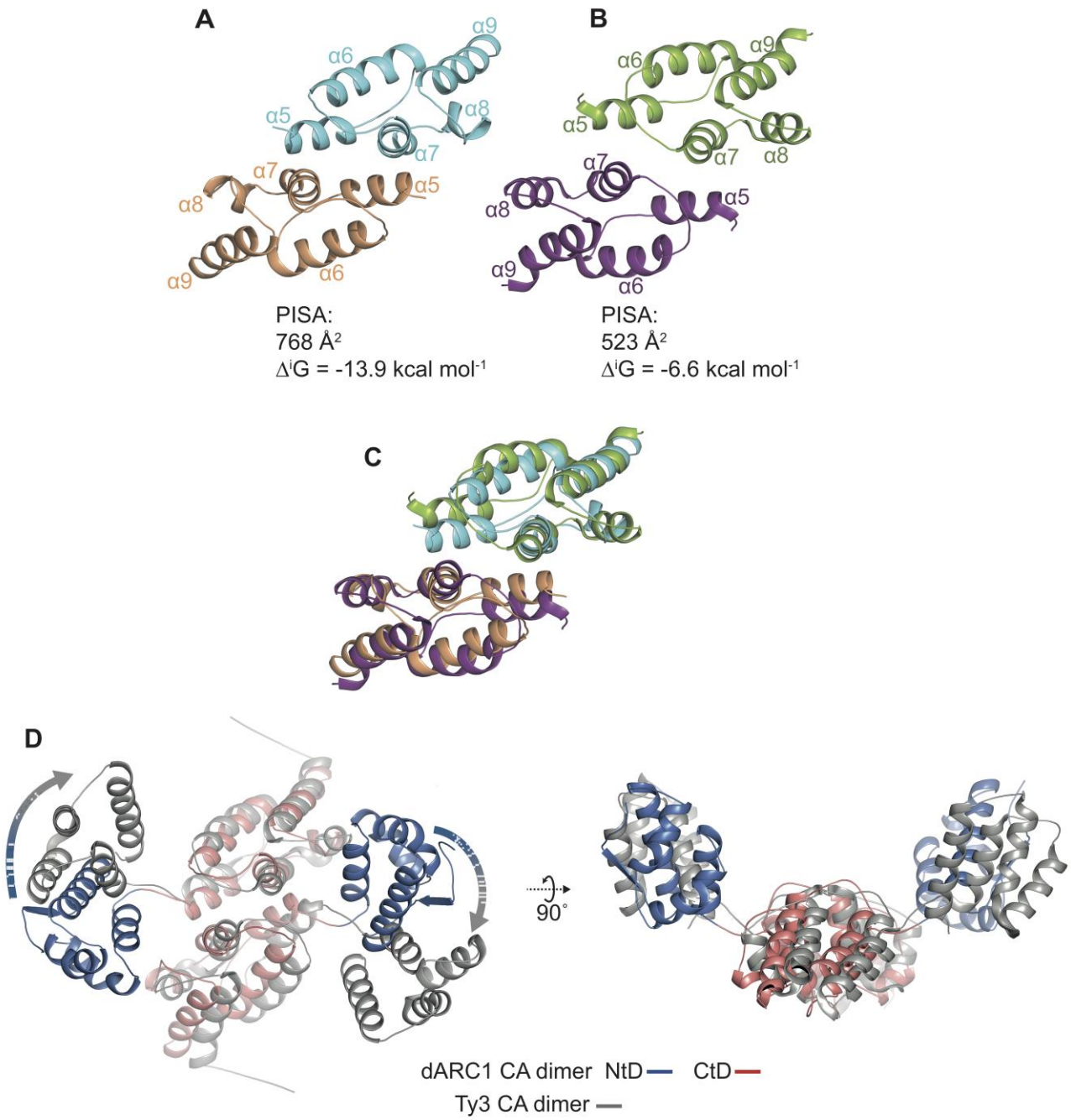

Fig. S4. Comparison of dARC1 and Ty3 CA. (A and B) Cartoon representations of CA-CTD dimers. (**A**) dARC1, cyan and wheat, (**B**) Ty3 (PDB: 6r23), magenta and pale green. CTD helices are labelled α5 to α9 according to the dARC1 crystal structure. The buried surface area (Å^2) and free energy of interaction (Δ^iG) of each interface, calculated in PDBePISA is displayed below each structure. (**C**) Structural alignment of dARC1 CA-CTD and Ty3 CA-CTD. The structures are aligned with respect to the dARC1 dimer (RMSD = 3.3 Å, 143 Cα), protein backbones are coloured as in **A and B**. (**D**) View of the dARC1 and

Ty3 dimers showing the positioning of NTDs with respect to the CTD dimer. The structures are aligned on the CTD dimer as in **C**. Ty3 is coloured in grey, dARC1 CA-NTD and CA-CTD are coloured blue and red respectively. The dARC1 CA-NTDs have to undergo a slight rotational operation to match the conformation seen in the Ty3 dimer (represented by the shaded arrows). Despite this difference, both dimers have the same “glacial trough” arrangement and the dARC1 crystallographic dimer has an NTD-CTD orientation very close to that observed Ty3 icosahedral particles.