

RESEARCH ARTICLE

Characterization of the C-terminal tail of the Arc protein

Melissa Boldridge¹, Jody Shimabukuro¹, Keith Nakamatsu², Christian Won², Chad Jansen³, Helen Turner³, Lei Wang^{1*}

1 Department of Natural Sciences, Hawaii Pacific University, Honolulu, Hawaii, United States of America, **2** Department of Natural Sciences, Windward Community College, Kaneohe, Hawaii, United States of America, **3** Laboratory of Immunology and Signal Transduction, School of Natural Sciences and Mathematics, Chaminade University, Honolulu, Hawaii, United States of America

* lwang@hpu.edu

**OPEN ACCESS**

Citation: Boldridge M, Shimabukuro J, Nakamatsu K, Won C, Jansen C, Turner H, et al. (2020) Characterization of the C-terminal tail of the Arc protein. PLoS ONE 15(9): e0239870. <https://doi.org/10.1371/journal.pone.0239870>

Editor: Patrick van der Wel, Rijksuniversiteit Groningen, NETHERLANDS

Received: June 18, 2020

Accepted: September 15, 2020

Published: September 29, 2020

Copyright: © 2020 Boldridge et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The NMR data are held at the Biological Magnetic Resonance Data Bank (<http://www.bmrb.wisc.edu/>). The accession numbers are: 50436, 50440, 50441.

Funding: L.W. National Institute of General Medical Sciences: <https://www.nigms.nih.gov/1R15GM126499-01>, P20GM103466. Hawaii Community Foundation: <https://www.hawaiicommunityfoundation.org/> 16ADVC-78888. NO - The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The activity-regulated cytoskeleton-associate protein Arc (or Arg3.1) is specifically linked to memory formation and a number of cognitive disorders, including Alzheimer's disease and schizophrenia. Since the discovery of Arc in 1995, extensive research has been conducted on the protein to identify its function and mechanisms of action, with solving the structure of Arc as a major goal. However, the Arc protein tends to self-oligomerize *in vitro*, and is difficult to crystallize. These properties have hindered efforts to obtain the structure of the full-length, whole protein Arc. As an alternative approach, we and others, have sought to solve the structures of various subdomain proteins of Arc, including the N-lobe, C-lobe, and capsid domain (N-lobe + C-lobe). In this study, we characterized the C-terminal tail of Arc using integrated bioinformatic and structural biology techniques. We compared the sequences of Arc proteins in different mammal species and found that the amino-acid composition in the C-terminal tail region has a significantly higher degree of variation rate than the rest of the protein. Structural prediction programs suggested that the C-terminal tail is structurally disordered. Chemical shift analysis based on solution NMR spectra confirmed that the C-terminal tail has a random coil (disordered) structure, and the tail starts from the residue D357. Furthermore, the NMR spectra showed that the C-terminal tail has minimum (if any) interaction with its neighboring capsid domain in Arc. This study fills gaps in our specific understanding of the structural nature and functional contributions of the Arc C-terminus.

Introduction

Long-term memory is believed to be encoded in long-term synaptic plasticity, which requires rapid synthesis of new mRNAs and proteins. While examining newly synthesized biomolecules during memory formation, two research groups independently discovered the 3.1 kb mRNA that encodes the activity-regulated cytoskeleton-associate protein Arc (or Arg3.1) [1, 2]. The Arc gene belongs to the immediate-early gene family, which mediates prompt transcriptional responses to neuronal activities [1]. Arc mRNA and protein are mostly expressed in the hippocampal and parietal cortex neurons that are responsible for learning and

Competing interests: The authors have declared that no competing interests exist.

memory formation [3]. In these neurons, Arc mRNA selectively accumulates at regions that contain recently activated synapses, and migrates away from inactive synapses [4]. Similar to its mRNA, Arc protein tends to accumulate at activated synapses, and is especially enriched in the postsynaptic density [5]. The distinctive pattern of Arc's selective accumulation at activated synapses suggests its connection with memory formation. Indeed, during behavioral learning, Arc was quickly and robustly induced in rat brains, and was used to mark neuronal networks involved in learning [6, 7]. Arc-knockout mice showed substantial deficits in long-lasting memories for implicit and explicit learning tasks, but their short-term task acquisition memories were not affected [8]. Moreover, antisense inhibition of Arc expression in rat hippocampi resulted in impaired long-term synaptic potentiation and long-term memory consolidation [9]. Extensive research has been conducted on the Arc protein to identify its functional mechanism at the molecular level. Studies have shown that Arc can interact with endocytic machinery, especially with dynamin and endophilin, to regulate neuronal trafficking of AMPA receptors [10]; Arc can interact with the γ -secretase subunit, presenilin, to regulate β -amyloid generation [11]; Arc binds to CaMKII and localizes at silenced synapses [12]; Triad3A can bind and ubiquitinate Arc and cause its rapid degradation [13]; the nuclear transportation of Arc is regulated by its nuclear localization, retention, and exportation domains [14]; Arc bears a capsid domain and can form retroviral capsid-like structures to transfer mRNA [15–17].

Solving the structure of Arc could help provide insight into its functions at the residue-specific level. The Arc protein in most mammal species has around 396 amino-acid residues; it tends to self-oligomerize in solution *in vitro* [18], and is difficult to crystallize. These properties have hindered the effort to obtain the structure of the full-length Arc as a whole protein using the traditional x-ray crystallography and solution NMR methods. As an alternative method to obtain Arc's structure, scientists have produced various subdomain proteins of Arc and solved their structures, respectively. For example, Zhang et al. produced the N-lobe subdomain of rat Arc (Arc²⁰⁷⁻²⁷⁷) and solved the crystal structure of the N-lobe complexed with the peptides derived from TARP γ 2 or CaMKII α . They found that the N-lobe folds into a four-helix bundle, and the peptide binds to a groove of the helix bundle that is formed by its N-terminal β -strand, the loop connecting α -2 and α -3 helices, and the N-terminus of the α -3 helix. They also produced the C-lobe subdomain of Arc (Arc²⁷⁸⁻³⁷⁰) and solved its crystal structure, which folds into a five-helix bundle. They further proposed that the two-lobe arrangement of Arc resembles the capsid domain of the retrovirus-like gag protein [15]. Later, Hallin et al. produced a full-length rat Arc protein by extracting the protein from inclusion bodies, and identified the relative positions of the subdomains of Arc using the small-angle X-ray scattering (SAXS) method. They found that Arc's coiled-coil subdomain (Arc²⁶⁻¹³⁰) lies above its bi-lobe subdomain (Arc²¹⁰⁻³⁶¹), and its N- and C-terminal tails lie at opposite ends. They also found that ligand binding to Arc's N-lobe did not cause major conformational changes to the rest of the protein [19]. Furthermore, Nielsen et al. produced the capsid domain of rat Arc (Arc²⁰⁶⁻³⁶⁴) that includes the N-lobe and the C-lobe, and solved its structure using the solution NMR method [20]. Recently, Cottee et al. and Erlendsson et al. studied the structure of Arc in fruit fly (*Drosophila*) using the x-ray crystallography and cryo-electron microscopy methods [21, 22]. Different from rat and other mammal species, the fruit fly has two copies of Arc protein: Arc1 and Arc2; Arc1 has 254 amino-acid residues and Arc2 has 193 amino-acid residues. Both Arc1 and Arc2 have a capsid domain, whose structure closely resembles the structure of the capsid domain of Arc in rat [21, 22]. In addition, Arc1 has a C-terminal tail, which is absent in Arc2. The tail contains a Cys-His zinc finger motif, and it can form an anti-parallel single-knuckle zinc finger with an adjacent Arc1 C-terminal tail [22]. However, the Arc protein in mammals does not have the zinc finger motif in its C-terminal tail.

In this study, we characterized the C-terminal tail of Arc in mammal species using the rat Arc protein as an example. We compared the sequences of Arc proteins in different mammals and found that the amino-acid composition in the C-terminal tail region has a significantly higher degree of variation rate than the rest of the protein. Structural prediction programs suggested that the C-terminal tail is structurally disordered. Chemical shift analysis based on solution NMR spectra confirmed that the C-terminal tail has a random coil (disordered) structure, and the tail starts from the residue D357. Furthermore, the NMR spectra showed that the C-terminal tail has minimum (if any) interaction with its neighboring capsid domain in Arc. Since the C-terminus of Arc in mammals has been relatively understudied, this study provides new information about a region of the protein that needs to be considered in the analysis of Arc structure-function relationships.

Materials and methods

Alignment of Arc protein sequences

The Arc protein sequences from different mammal species (human, chimpanzee, Rhesus monkey, rat, mouse, water buffalo, cattle, goat, and Arabian camel) were downloaded from the PubMed protein database (www.ncbi.nlm.nih.gov/pubmed) and compared with each other using the LALIGN algorithm [23]. In the alignment figure, the non-conserved amino-acid residues in the Arc protein sequence were highlighted in yellow color (the minority groups of residues were highlighted). The sequence segments and amino-acid residues that were found important for Arc functions were labeled above the protein sequence. The position of the secondary structures (mainly α -helices) in the Arc capsid domain (Arc²⁰⁶⁻³⁶⁴) were also labeled in the alignment figure.

Prediction of disorder probability of Arc structure

Three algorithms, DISpro [24], IUPred [25], and PONDR-FIT [26], were used to predict the structurally ordered and disordered regions in Arc. The protein sequence of rat Arc (NCBI ID: NP_062234.1) was submitted to and analyzed by each of the three algorithms. As results, the disorder probability of each amino-acid residue in Arc was plotted against the amino-acid residue number of Arc. The sequence segments that have lower disorder probability values (< 0.5) are suggested to be structurally ordered (having structural domains), and the sequence segments that have higher disorder probability values (> 0.5) are suggested to be structurally disordered (having random coil structures).

Production of Arc subdomain proteins

The DNA fragments encoding three rat Arc subdomain proteins (Arc²⁸⁰⁻³⁹⁶, Arc²⁰⁸⁻³⁹⁶, and Arc²⁰⁸⁻³⁶³) were respectively cloned into a pET32 vector using PCR methods. The pET32 vector adds a histidine-tag, a thioredoxin fusion protein, and a TEV cleavage sequence to the N-terminus of the produced Arc subdomain proteins. The DNA plasmids containing Arc subdomain sequences were transfected into BL21(DE3) cells, and the cells were grown in ¹³C-¹⁵N-labeled Spectra-9 media (Cambridge Isotope Laboratories, Inc.) at 37°C in a shaking incubator. When the optical density (OD₆₀₀) of the cell culture reached 0.8, the temperature of the incubator was decreased to 20°C, and 1 mM of IPTG (isopropyl- β -d-thiogalactoside) solution was added to the cell culture to induce protein expression. The cells were harvested 17 hours after the addition of IPTG using centrifugation, and they were lysed on ice using sonication. The lysed cells were centrifuged at 10,000 rpm, and the supernatant after centrifugation was loaded to an AKTA Prime FPLC instrument (GE Healthcare, Inc.). The Arc subdomain proteins were purified from the supernatant using the nickel affinity, size-exclusion, and ion-

exchange chromatography columns. Finally, the histidine-tag, thioredoxin fusion protein, and TEV (tobacco etch virus) protease cleavage sequence were removed from the Arc subdomain proteins using TEV protease.

NMR

The NMR samples contained 500 μM of ^{13}C - ^{15}N -labeled Arc subdomain proteins in a buffer with pH 7.0 containing 10 mM of Na_2HPO_4 , 50 mM of NaCl, 2mM of TCEP, 0.05% of NaN_3 , and 10% of D_2O . NMR experiments were performed on a Bruker 600 or 800 MHz spectrometer equipped with TCI triple-resonance cryogenic probes using standard Bruker pulse programs. ^1H , ^{13}C , and ^{15}N backbone resonances were assigned using standard triple resonance experiments, such as TROSY, HNCA, HNCACB, CBCA(CO)NH, and HN(CO)CA. All of the 3D experiments were collected with 16 transients at 293 K. All of the spectra were processed using the Topspin software (Bruker, Inc.) and analyzed using the Computer-Aided Resonance Assignment (CARA) software [27].

Chemical shift calculation

The chemical shift of the α -carbon of each amino-acid residue in $\text{Arc}^{280-396}$ was extracted from its NMR-HNCA spectra. The chemical shift was compared to the standard chemical shift of the α -carbon of the same amino-acid residue in a random coil structure. The difference in chemical shift (ΔC_α , ppm) of each amino-acid residue was calculated and plotted against the amino-acid residue number of $\text{Arc}^{280-396}$. The continuous and significantly positive ΔC_α values (> 0.05) suggest that the corresponding sequence segments have an α -helix secondary structure. The continuous and significantly negative ΔC_α values (< -0.05) suggest that the corresponding sequence segments have a β -strand secondary structure. The discrete and insignificant ΔC_α values (between -0.05 to 0.05) suggest that the corresponding sequence segments have a random coil structure.

Results

The amino-acid residue composition of Arc C-terminal tail is highly divergent

In this study, we aligned and compared the sequences of Arc proteins in different mammal species. The alignment result revealed the conserved and non-conserved amino-acid residues (Fig 1). It is noticeable that the amino-acid composition in the C-terminal tail region has a significantly higher degree of variation rate than the rest of the protein. Specifically, the variation rate is 39.1% in the C-terminal sequence segment $\text{Arc}^{351-396}$ (the residue variations highlighted in green color); whereas, the average variation rate is 12.3% in the rest of the protein (the residue variations highlighted in yellow color). It is commonly believed that the more functionally important regions in a protein are more structurally ordered (having structural domains), and their sequence compositions are more conserved during biological evolution; on the other hand, the less functionally important regions have more disordered structures (having random coil structures) and their sequence compositions are less conserved during evolution. The high degree of variation rate of amino-acid composition in $\text{Arc}^{351-396}$ could suggest that this region is less functionally important and is more structurally disordered. As a comparison to the C-terminal tail region, the variation rate of amino-acid composition of Arc's N-lobe ($\text{Arc}^{207-277}$) is only 7.0%. The N-lobe is involved in Arc's interaction with several other proteins, and it has a highly ordered structure. Furthermore, the alignment result revealed some features of Arc protein, such as the phosphorylation sites S84 and S170, that are common among mammal

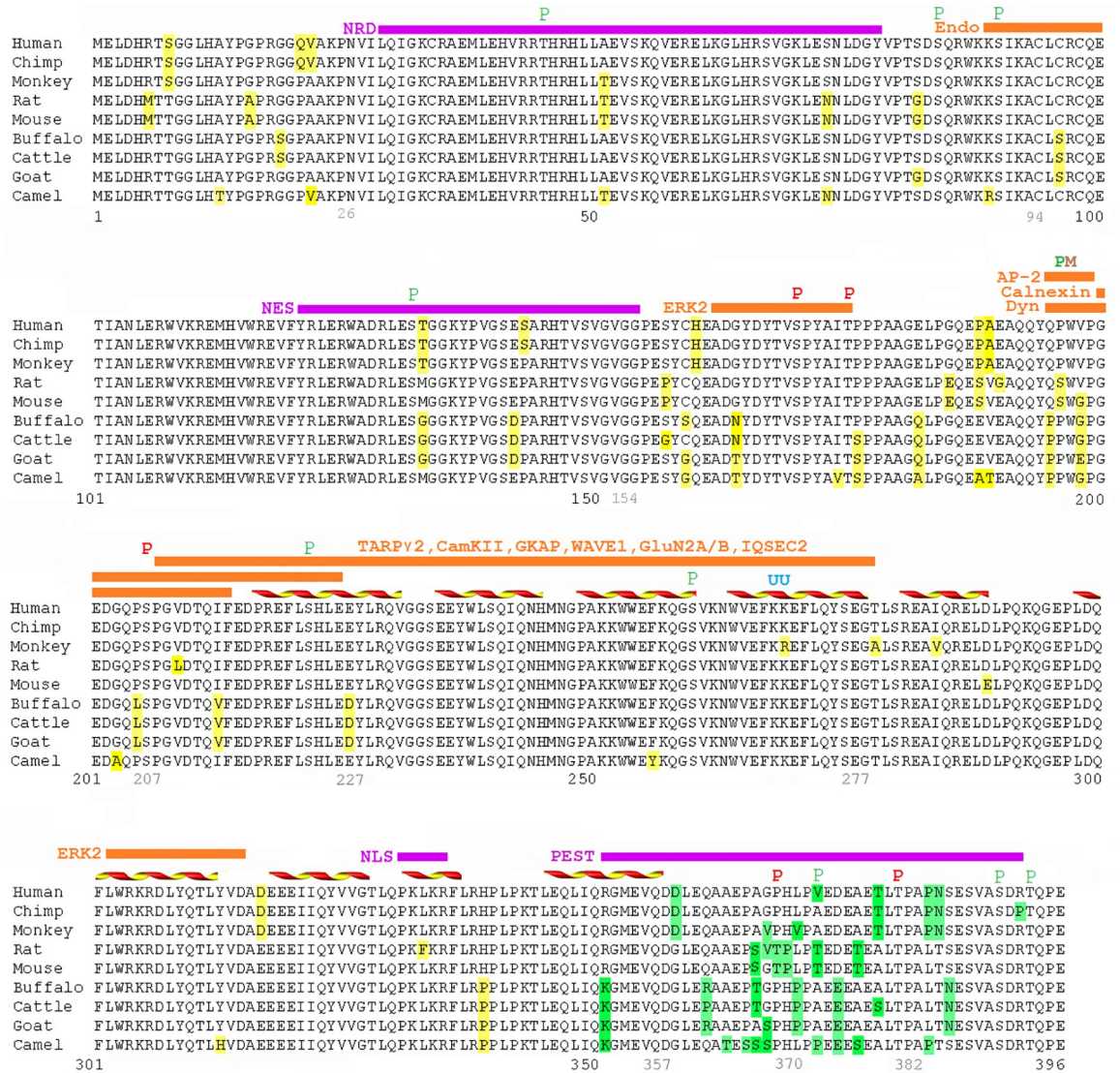


Fig 1. Sequence alignment of Arc proteins in different mammalian species. The sequence of Arc proteins in different mammalian species were aligned and compared with each other. The non-conserved amino-acid residues in the protein sequences were highlighted in green color (in the C-terminal sequence segment Arc³⁵¹⁻³⁹⁶) or yellow color (in the rest of the protein). The alignment result shows that the amino-acid residue composition in the C-terminal tail region of the Arc protein (~ Arc³⁵¹⁻³⁹⁶) is highly divergent compared to the rest of the protein. The purple bars above the protein sequence indicate the sequence regions that are important for Arc’s nuclear transportation (NRD, NES, NLS) [14] and degradation (PEST) [28]; the orange bars indicate the sequence regions that are important for Arc’s interaction with other proteins (these proteins’ names were labeled before/above the orange bars) [10, 11, 13, 15, 20, 29–33]; the colored letters indicate Arc’s phosphorylation sites (red “P”) [31, 34], putative phosphorylation sites (green “P”) [1, 2], ubiquitination sites (blue “U”) [13], and mutation site (brown “M”) [30]; the yellow-red ribbons indicate the α -helix secondary structures in Arc’s capsid domain [15, 20]. (See the Discussion section for more detail).

<https://doi.org/10.1371/journal.pone.0239870.g001>

species. This information adds to our understanding of how the sequences of Arc proteins from different animal classes compare with one other [18].

The C-terminal tail of Arc is predicted to be structurally disordered

Three prediction algorithms (DISpro, IUPred, and PONDR-FIT) were used to analyze the protein sequence of Arc (Fig 2). These algorithms predict each sequence segment’s probability of

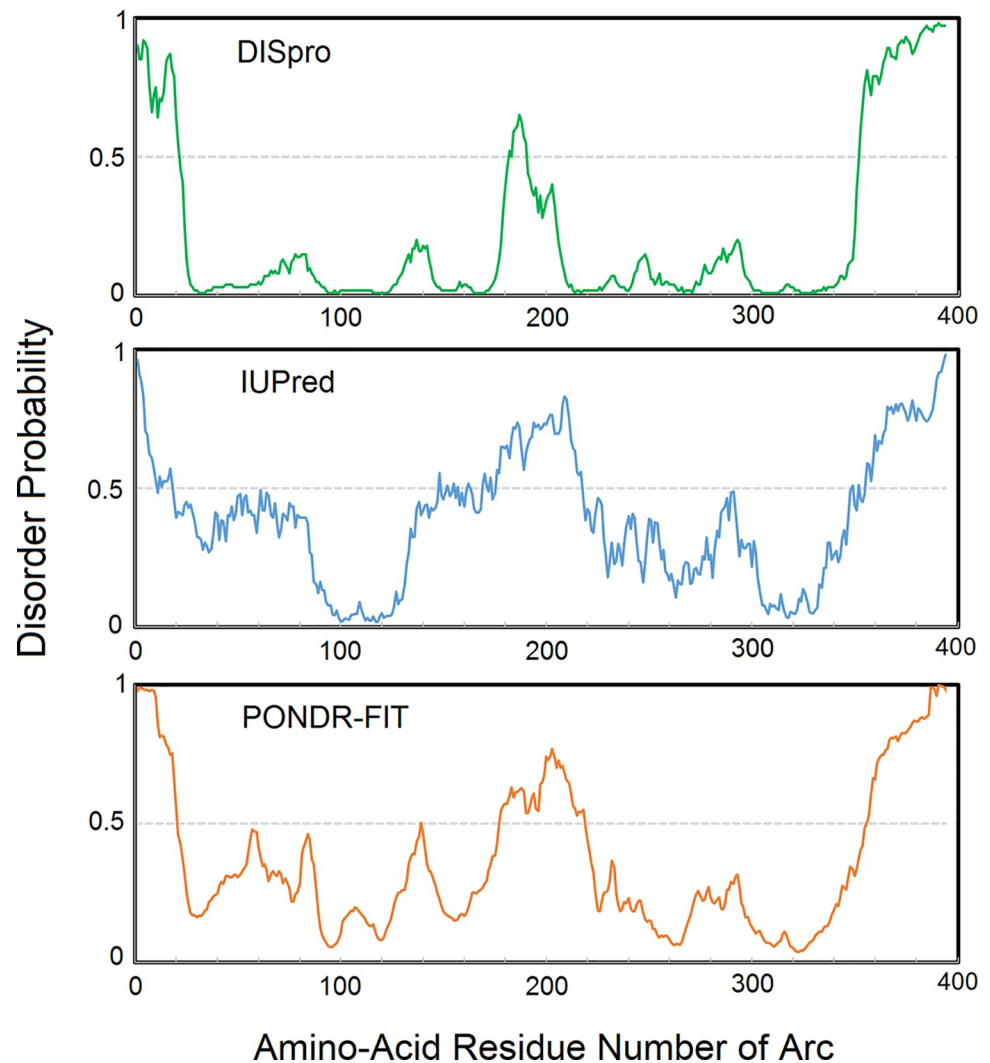


Fig 2. Prediction of structural disorder probability of Arc protein. The disorder probability of each amino-acid residue in Arc was plotted against the amino-acid residue number of Arc. The C-terminal tail of Arc was consistently predicted to have larger disorder probability values (> 0.5) by three prediction algorithms. The larger disorder probability suggests that the corresponding sequence segment is structurally disordered (having random coil structure).

<https://doi.org/10.1371/journal.pone.0239870.g002>

being structurally ordered or disordered. A disorder probability value higher than 0.5 suggests that the corresponding sequence segment of a protein tends to be structurally disordered; whereas, a disorder probability value lower than 0.5 suggests that the corresponding sequence segment tends to be structurally ordered. All three prediction algorithms consistently predicted that the C-terminal tail region of the Arc protein (approximately Arc³⁵⁶⁻³⁹⁶) has high disorder probability values (> 0.5), suggesting this region is likely to be structurally disordered. As a comparison to the C-terminal tail region, the prediction algorithms consistently predicted that the sequence segment Arc²¹⁰⁻³⁵⁰ has low disorder probability values (< 0.5), suggesting this region is structurally ordered. Indeed, the sequence segment Arc²¹⁰⁻³⁵⁰ largely overlaps with the structurally ordered capsid domain (Arc²⁰⁶⁻³⁶⁴). In addition, the prediction algorithms consistently predicted that the N-terminal tail and the middle region of the Arc protein have high disorder probability values (> 0.5), suggesting they could also be structurally

disordered. The prediction algorithms also consistently predicted that the region centered around the sequence segment Arc⁹⁰⁻¹²⁰ has low disorder probability values (< 0.5), suggesting this region could be structurally ordered. However, the exact sequence ranges of the N-terminal tail, the middle region, and the potentially structured region between them were not consistently defined. The prediction result adds to knowledge and complement with the previous study [18].

The C-terminal tail of Arc has a random coil structure and the tail starts at residue D357

In order to experimentally determine the structure of Arc C-terminal tail, we produced the ¹³C-¹⁵N-labeled Arc²⁸⁰⁻³⁹⁶ protein, which includes the C-lobe subdomain and the previously uncharacterized C-terminal tail. We collected a series of 3D NMR spectra on the protein, and the α -carbon chemical shift of each amino-acid residue in Arc²⁸⁰⁻³⁹⁶ was determined from the NMR spectra (except the first two amino-acid residues and the proline residues). The chemical shift was compared to the standard chemical shift of the α -carbon of the same amino-acid residue in a random coil structure. The difference in chemical shift (ΔC_{α} , ppm) of each amino-acid residue was calculated and plotted against the protein sequence of Arc²⁸⁰⁻³⁹⁶ (Fig 3). The continuous and significantly positive ΔC_{α} values (> 0.05) suggest that the corresponding sequence segments have an α -helix secondary structure, such as the five α -helices in Arc²⁸⁰⁻³⁹⁶ (grey bars, α 1- α 5) that agrees with the five-helix bundle structure of Arc C-lobe subdomain [15]. The discrete and insignificant ΔC_{α} values (between -0.05 to 0.05) suggest that the corresponding sequence segments have a random coil (disordered) structure, such as the Arc C-terminal tail Arc³⁵⁷⁻³⁹⁶ (green bar). Furthermore, the plot does not contain continuous and significantly negative ΔC_{α} values (< -0.05), which suggests that the Arc²⁸⁰⁻³⁹⁶ protein does not have β -strand secondary structure.

The C-terminal tail has minimal interaction with its neighboring capsid domain

We tested if there is any interaction between the C-terminal tail and the capsid domain adjacent to it. We successfully produced the capsid domain protein (Arc²⁰⁸⁻³⁶³), but our attempts to produce the C-terminal tail (Arc³⁵⁷⁻³⁹⁶) failed as this disordered protein was degraded

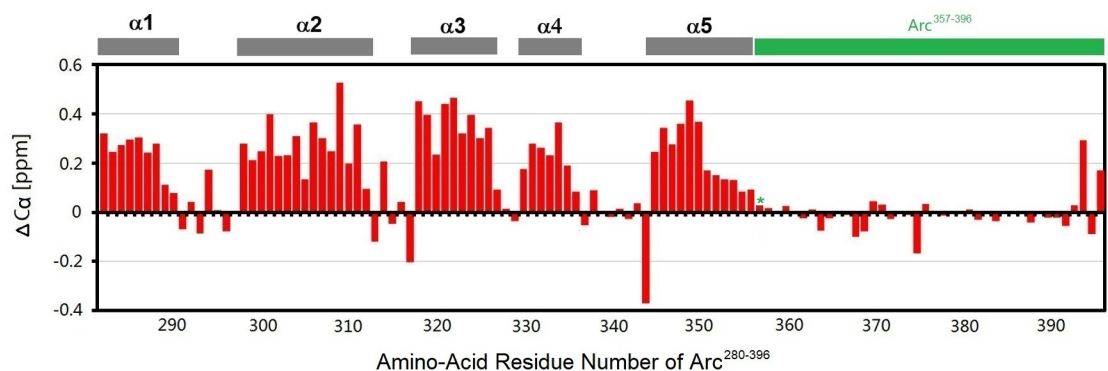


Fig 3. NMR chemical shift analysis of Arc²⁸⁰⁻³⁹⁶. The chemical shift differences of α -carbon (ΔC_{α}) were plotted against the amino-acid residue number of Arc²⁸⁰⁻³⁹⁶. The sequence segments with continuous and significantly positive ΔC_{α} values (> 0.05) correspond to the five α -helices (grey bars, α 1- α 5) in the C-lobe region. The C-terminal tail (green bar) has insignificant (-0.05 ~ 0.05) or discrete ΔC_{α} values, suggesting this region has a random coil structure. The chemical shift data of Arc²⁸⁰⁻³⁹⁶ was deposited in the Biological Magnetic Resonance Data Bank (BMRB) (ID: 50436).

<https://doi.org/10.1371/journal.pone.0239870.g003>

during bacterial expression. As an alternative approach, we produced the Arc²⁰⁸⁻³⁹⁶ protein, which includes the capsid domain and the C-terminal tail. We collected the NMR TROSY spectra of Arc²⁰⁸⁻³⁶³ and Arc²⁰⁸⁻³⁹⁶, and compared them with each other (Fig 4). The TROSY spectrum of Arc²⁰⁸⁻³⁶³ (blue) almost precisely overlaps with the TROSY spectrum of Arc²⁰⁸⁻³⁹⁶ (grey) (except the latter has additional cross-peaks from its C-terminal tail). In addition, we compared the NMR TROSY spectra of Arc²⁸⁰⁻³⁹⁶ and Arc²⁰⁸⁻³⁹⁶ (Fig 5A), and analyzed their chemical shifts of the backbone amide resonance (Fig 5B). Arc²⁸⁰⁻³⁹⁶ includes the C-lobe and C-terminal tail; Arc²⁰⁸⁻³⁹⁶ has an additional N-lobe than Arc²⁸⁰⁻³⁹⁶. The analysis result showed that the presence/removal of residues 208–279 (N-lobe) did not cause major chemical shift perturbation (CSP) in residues 280–396 (C-lobe and C-terminal), except for residues 282–285 that connect N-lobe and C-lobe (Fig 5B). It suggests that the N-lobe has minimum (if any) interaction with the C-terminal tail (and the C-lobe). We also compared the chemical shifts of the backbone amide resonances between Arc²⁸⁰⁻³⁹⁶ and Arc²⁸⁰⁻³⁶³ (Arc²⁸⁰⁻³⁹⁶ includes the C-lobe and C-terminal tail; Arc²⁸⁰⁻³⁶³ includes only the C-lobe) (Fig 5C). The comparison result showed that the presence/removal of residues 364–396 (C-terminal tail) did not cause major chemical shift perturbation (CSP) in residues 280–363 (C-lobe), except for residues 362–363 that connect C-lobe and C-terminal tail (Fig 5D). It suggests that the C-lobe also has minimum (if any) interaction with the C-terminal tail. Taken together, these data suggest that the C-terminal tail has minimum (if any) interaction with the N-lobe and C-lobe (together they form the capsid domain), and it does not affect the folding of the capsid domain. This observation agrees with a previous finding that the C-terminal tail lies outside the core of the Arc protein, which is consisted of the Arc N-terminus sitting above the capsid domain [19]. However, further experiments are required to examine the possible interaction between the C-terminal tail and the capsid domain.

Discussion

Arc was discovered in 1995, and it was found to play an important role in synaptic plasticity and memory formation. Solving the structure of Arc will help understand its functional

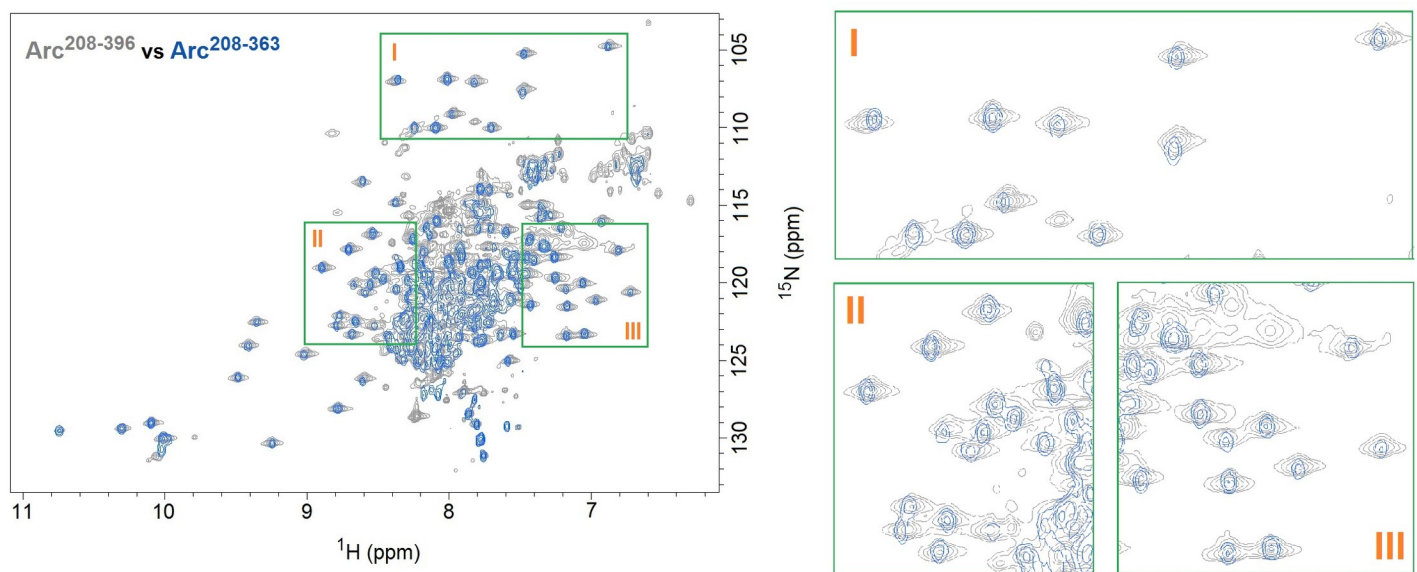


Fig 4. Comparison of NMR TROSY spectra of Arc²⁰⁸⁻³⁹⁶ and Arc²⁰⁸⁻³⁶³. The Arc²⁰⁸⁻³⁹⁶ protein (grey) has the additional C-terminal tail compared to the Arc²⁰⁸⁻³⁶³ protein (the capsid domain, blue). The presence of the C-terminal tail does not generate noticeable chemical shifts on the capsid domain peaks, suggesting the C-terminal tail has minimum (if any) interaction with the capsid domain. (The green boxes I, II, and III on the right are enlarged views of the example regions in the TROSY spectra on the left).

<https://doi.org/10.1371/journal.pone.0239870.g004>

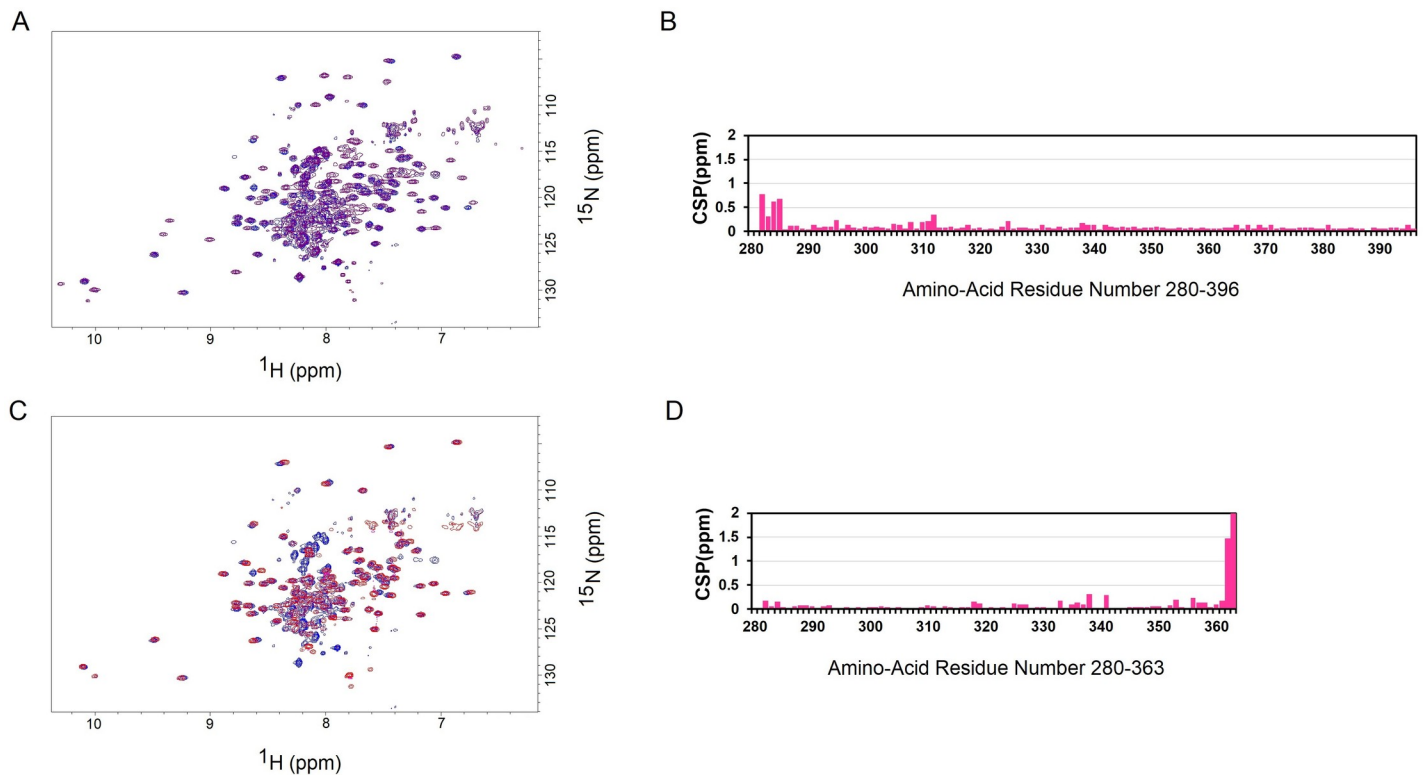


Fig 5. Changes of the chemical shifts of the backbone amide resonance between Arc²⁸⁰⁻³⁹⁶ and Arc²⁰⁸⁻³⁹⁶/Arc²⁸⁰⁻³⁶³. (A) The overlay of NMR TROSY spectra of Arc²⁸⁰⁻³⁹⁶ (blue) and Arc²⁰⁸⁻³⁹⁶ (purple). (B) The chemical shift perturbations (CSP) of residues 280–396 when the amide chemical shifts of Arc²⁰⁸⁻³⁹⁶ and Arc²⁸⁰⁻³⁹⁶ are compared with each other. (C) The overlay of NMR TROSY spectra of Arc²⁸⁰⁻³⁹⁶ (blue) and Arc²⁸⁰⁻³⁶³ (red). (D) The chemical shift perturbations (CSP) of residues 280–363 when the amide chemical shifts of Arc²⁸⁰⁻³⁶³ and Arc²⁸⁰⁻³⁹⁶ are compared with each other. The CSPs of backbone amides are calculated using the equation $[(\Delta H)^2 + (\Delta N)^2/5]^{1/2}$, where ΔH and ΔN are the changes in the 1H and ^{15}N chemical shifts respectively. The chemical shift data were deposited in the BMRB database (BMRB ID: 50436 (Arc²⁸⁰⁻³⁹⁶), 50440 (Arc²⁰⁸⁻³⁹⁶), and 50441 (Arc²⁸⁰⁻³⁶³)).

<https://doi.org/10.1371/journal.pone.0239870.g005>

mechanism. In this study, we characterized the C-terminal tail of Arc in mammals using the rat Arc protein as an example. We found that the C-terminal tail of Arc in mammal species has a significantly higher variation rate in amino-acid composition than the rest of the protein (39.1% vs. 12.3%), suggesting this region may lack important functional domains and/or ordered structures. The C-terminal tail was predicted to have high disorder probability, suggesting this region is structurally disordered. The NMR chemical shift analysis of Arc²⁸⁰⁻³⁹⁶ showed that the C-terminal tail indeed has a random coil (disordered) structure, and the tail starts from the residue D357. We further tested if the tail interacts with the neighboring capsid domain and found that it has minimum (if any) interaction with the latter.

In the recently characterized structure of *Drosophila* Arc1, 240 copies of Arc1 oligomerize into 12 pentameric and 30 hexameric capsomeres, and these capsomeres further form an icosahedral capsid sphere. Arc1's 48-residue C-terminal tail locates inside the capsid and below the capsomeres. The residues 224–252 of two C-terminal tails under each hexameric capsomere form anti-parallel zinc fingers; whereas, the other residues in the two C-terminal tails and the rest copies of C-terminal tail are structurally disordered [22]. The C-terminal tail of Arc1 also contains 12 basic residues; together with the zinc fingers, they may facilitate Arc1's mRNA recognition and binding, similar to the function of the nucleocapsid domain of the retrovirus HIV [22]. For the C-terminal tail of Arc in mammals, it may also locate inside the capsid as the *Drosophila* Arc1, and it could interfere with the capsid formation of the full-length Arc as the capsid domain alone was found unable to form the capsid layer [17]. But it lacks the

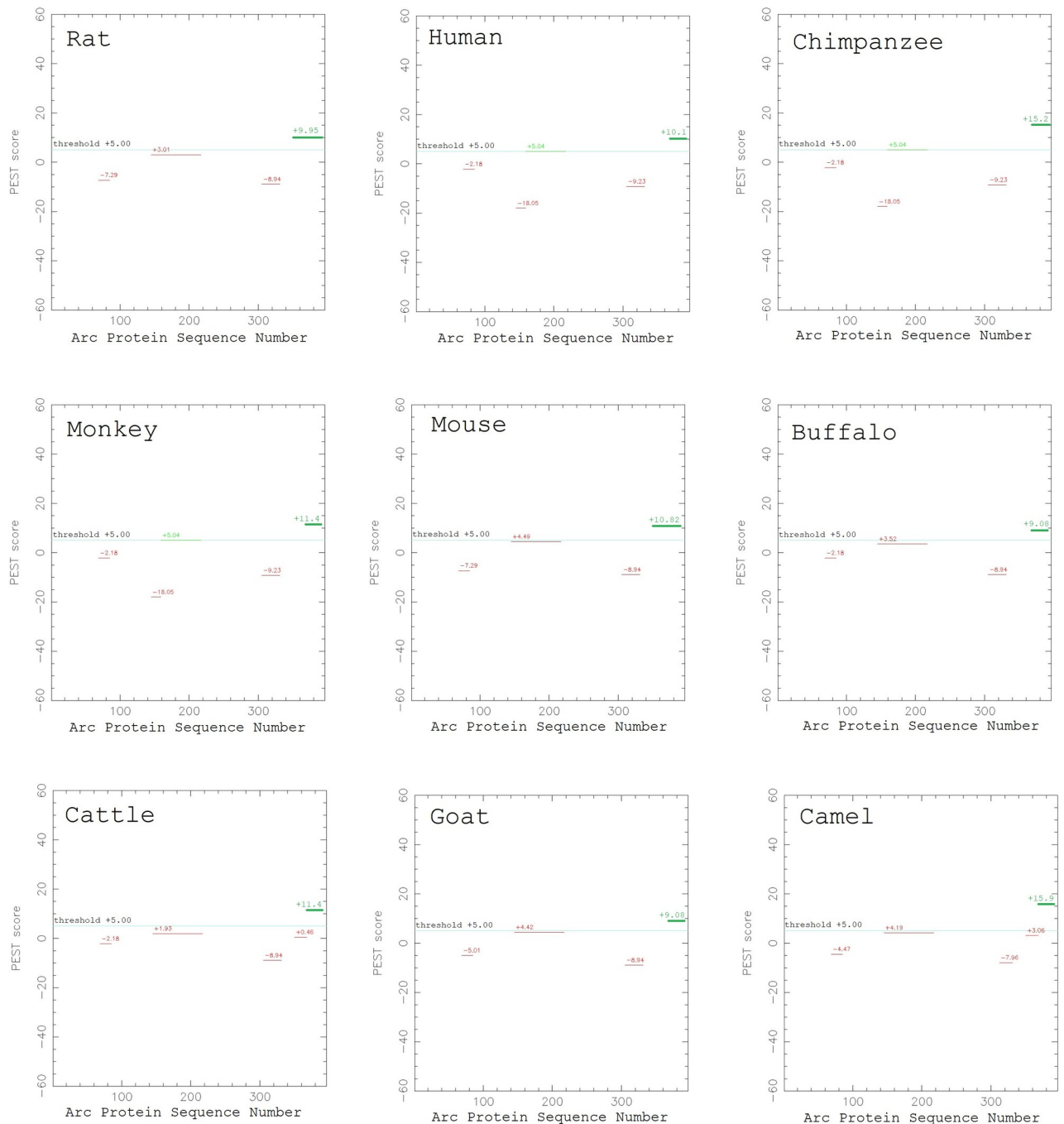


Fig 6. The PEST signal region in Arc C-terminal tail. The C-terminal tail of rat Arc contains a PEST signal region (the green bar that reads “+9.95” in the upper left plot). This region is largely preserved in the other examined mammal species (the green bars in the other 8 plots). The PEST score plots were generated using the ePESTfind tool at <https://emboss.bioinformatics.nl/cgi-bin/emboss/pepfind>.

<https://doi.org/10.1371/journal.pone.0239870.g006>

zinc fingers and basic residue patches for mRNA regulation, and no close homology to its sequence was found in other retrotransposons or retroviruses either.

To identify the possible functions of the C-terminal tail of Arc in mammals, we reviewed the previous studies that examined the functions of different sequence segments of Arc. These studies have shown that, within the C-terminal tail (Arc³⁵⁷⁻³⁹⁶), residues 351–392 of Arc were

predicted as a PEST signal region (a sequence segment that is rich in P, E, S, T residues) using the ePESTfind program [35], which may contribute to the protein's proteasome-dependent degradation [28]. Although the original PEST signal analysis was conducted on the Arc protein in rat, our sequence alignment result in this study showed that the PEST signal has been largely preserved in other mammal species despite the high variation rate of amino-acid composition at the C-terminal tail (Fig 6). In addition, phosphorylation sites have been identified in the C-terminal tail at residues T368, and T380 [31, 34]. It was also predicted that this region contains putative phosphorylation sites by the protein kinase C at residues S390, and by the casein kinase II at residues T372 and T393 [1, 2]. Among these phosphorylation sites, our sequence alignment result showed that the sites T380, S390, and T393 are preserved in the mammal species; whereas the sites T368 and T372 are only present in rat and mouse.

Furthermore, previous studies have shown that, for the Arc sequence segments that partially overlap with or include the C-terminal tail (Arc³⁵⁷⁻³⁹⁶), residues 228–380 of Arc share 20% sequence similarity with the 21st and 22nd repeats of the α -spectrin protein [1, 2]; since these structural motifs of α -spectrin tend to form oligomers [36], the corresponding 228–380 region of Arc may contribute to its self-oligomerization [18], and the Arc oligomers may further contribute to memory formation [37]. In addition, residues 155–396 of Arc bind to the dynamin 2 protein [10] and the clathrin-adaptor protein 2 (AP-2) [30] to regulate AMPA receptor endocytosis; residues 94–382 of Arc bind with the Triad3A protein, and Triad3A can ubiquitinate Arc at its residues K268 and K269 to help regulate synaptic strength [13]. Arc was also found to interact with Tip60 [38], PICK1 (especially its BAR domain, residues 152–278) [39], GSK3 α/β [34], and CaM kinase II [3, 12]; however, the specific sequence segments of Arc that regulate these interactions have not been identified. Therefore, the C-terminal tail of Arc may also contribute to Arc's self-oligomerization process as part of the spectrin-homologous domain, or in Arc's interaction with other proteins, such as dynamin 2, AP-2, Triad3A, Tip60, PICK1, GSK3 α/β , and/or CaM kinase II. Taken together, we speculate that, although the C-terminus is disordered, it may play important functional roles that vary between species.

Acknowledgments

We would like to thank Dr. Roland Riek, Dr. Christy Grace, Dr. David Horgen, Dr. Robert Nichols, Dr. Leticia Colmenares, Dr. Yongli Chen, and Dr. Gideon Berger for their valuable support.

Author Contributions

Conceptualization: Lei Wang.

Data curation: Melissa Boldridge, Jody Shimabukuro, Keith Nakamatsu, Christian Won, Lei Wang.

Formal analysis: Melissa Boldridge, Lei Wang.

Funding acquisition: Helen Turner, Lei Wang.

Investigation: Lei Wang.

Methodology: Lei Wang.

Project administration: Lei Wang.

Resources: Lei Wang.

Software: Lei Wang.

Supervision: Lei Wang.

Validation: Lei Wang.

Visualization: Lei Wang.

Writing – original draft: Chad Jansen, Helen Turner, Lei Wang.

Writing – review & editing: Chad Jansen, Helen Turner, Lei Wang.

References

1. Lyford GL, Yamagata K, Kaufmann WE, Barnes CA, Sanders LK, Copeland NG, et al. Arc, a growth factor and activity-regulated gene, encodes a novel cytoskeleton-associated protein that is enriched in neuronal dendrites. *Neuron*. 1995; 14(2):433–45. [https://doi.org/10.1016/0896-6273\(95\)90299-6](https://doi.org/10.1016/0896-6273(95)90299-6) PMID: 7857651
2. Link W, Konietzko U, Kauselmann G, Krug M, Schwanke B, Frey U, et al. Somatodendritic expression of an immediate early gene is regulated by synaptic activity. *Proc Natl Acad Sci U S A*. 1995; 92(12):5734–8. <https://doi.org/10.1073/pnas.92.12.5734> PMID: 7777577
3. Donai H, Sugiura H, Ara D, Yoshimura Y, Yamagata K, Yamauchi T. Interaction of Arc with CaM kinase II and stimulation of neurite extension by Arc in neuroblastoma cells expressing CaM kinase II. *Neurosci Res*. 2003; 47(4):399–408. <https://doi.org/10.1016/j.neures.2003.08.004> PMID: 14630344
4. Moga DE, Calhoun ME, Chowdhury A, Worley P, Morrison JH, Shapiro ML. Activity-regulated cytoskeletal-associated protein is localized to recently activated excitatory synapses. *Neuroscience*. 2004; 125(1):7–11. <https://doi.org/10.1016/j.neuroscience.2004.02.004> PMID: 15051140
5. Steward O, Worley PF. A cellular mechanism for targeting newly synthesized mRNAs to synaptic sites on dendrites. *Proc Natl Acad Sci U S A*. 2001; 98(13):7062–8. <https://doi.org/10.1073/pnas.131146398> PMID: 11416188
6. Guzowski JF, McNaughton BL, Barnes CA, Worley PF. Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. *Nat Neurosci*. 1999; 2(12):1120–4. <https://doi.org/10.1038/16046> PMID: 10570490
7. Guzowski JF, Setlow B, Wagner EK, McGaugh JL. Experience-dependent gene expression in the rat hippocampus after spatial learning: a comparison of the immediate-early genes Arc, c-fos, and zif268. *J Neurosci*. 2001; 21(14):5089–98. <https://doi.org/10.1523/JNEUROSCI.21-14-05089.2001> PMID: 11438584
8. Plath N, Ohana O, Dammermann B, Errington ML, Schmitz D, Gross C, et al. Arc/Arg3.1 is essential for the consolidation of synaptic plasticity and memories. *Neuron*. 2006; 52(3):437–44. <https://doi.org/10.1016/j.neuron.2006.08.024> PMID: 17088210
9. Guzowski JF, Lyford GL, Stevenson GD, Houston FP, McGaugh JL, Worley PF, et al. Inhibition of activity-dependent arc protein expression in the rat hippocampus impairs the maintenance of long-term potentiation and the consolidation of long-term memory. *J Neurosci*. 2000; 20(11):3993–4001. <https://doi.org/10.1523/JNEUROSCI.20-11-03993.2000> PMID: 10818134
10. Chowdhury S, Shepherd JD, Okuno H, Lyford G, Petralia RS, Plath N, et al. Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron*. 2006; 52(3):445–59. <https://doi.org/10.1016/j.neuron.2006.08.033> PMID: 17088211
11. Wu J, Petralia RS, Kurushima H, Patel H, Jung MY, Volk L, et al. Arc/Arg3.1 regulates an endosomal pathway essential for activity-dependent beta-amyloid generation. *Cell*. 2011; 147(3):615–28. <https://doi.org/10.1016/j.cell.2011.09.036> PMID: 22036569
12. Okuno H, Akashi K, Ishii Y, Yagishita-Kyo N, Suzuki K, Nonaka M, et al. Inverse synaptic tagging of inactive synapses via dynamic interaction of Arc/Arg3.1 with CaMKIIbeta. *Cell*. 2012; 149(4):886–98. <https://doi.org/10.1016/j.cell.2012.02.062> PMID: 22579289
13. Mabb AM, Je HS, Wall MJ, Robinson CG, Larsen RS, Qiang Y, et al. Triad3A regulates synaptic strength by ubiquitination of Arc. *Neuron*. 2014; 82(6):1299–316. <https://doi.org/10.1016/j.neuron.2014.05.016> PMID: 24945773
14. Korb E, Wilkinson CL, Delgado RN, Lovero KL, Finkbeiner S. Arc in the nucleus regulates PML-dependent GluA1 transcription and homeostatic plasticity. *Nat Neurosci*. 2013; 16(7):874–83. <https://doi.org/10.1038/nn.3429> PMID: 23749147
15. Zhang W, Wu J, Ward MD, Yang S, Chuang YA, Xiao M, et al. Structural basis of arc binding to synaptic proteins: implications for cognitive disease. *Neuron*. 2015; 86(2):490–500. <https://doi.org/10.1016/j.neuron.2015.03.030> PMID: 25864631

16. Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V, Thomson T. Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. *Cell*. 2018; 172(1–2):262–74 e11. <https://doi.org/10.1016/j.cell.2017.12.022> PMID: 29328915
17. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, et al. The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell*. 2018; 172(1–2):275–88 e18. <https://doi.org/10.1016/j.cell.2017.12.024> PMID: 29328916
18. Myrum C, Baumann A, Bustad HJ, Flydal MI, Mariaule V, Alvira S, et al. Arc is a flexible modular protein capable of reversible self-oligomerization. *Biochem J*. 2015; 468(1):145–58. <https://doi.org/10.1042/BJ20141446> PMID: 25748042
19. Hallin EI, Eriksen MS, Baryshnikov S, Nikolaienko O, Grodem S, Hosokawa T, et al. Structure of monomeric full-length ARC sheds light on molecular flexibility, protein interactions, and functional modalities. *J Neurochem*. 2018.
20. Nielsen LD, Pedersen CP, Erlendsson S, Teilum K. The Capsid Domain of Arc Changes Its Oligomerization Propensity through Direct Interaction with the NMDA Receptor. *Structure*. 2019; 27(7):1071–81 e5. <https://doi.org/10.1016/j.str.2019.04.001> PMID: 31080121
21. Cottee MA, Letham SC, Young GR, Stoye JP, Taylor IA. Structure of *Drosophila melanogaster* ARC1 reveals a repurposed molecule with characteristics of retroviral Gag. *Sci Adv*. 2020; 6(1):eaay6354. <https://doi.org/10.1126/sciadv.aay6354> PMID: 31911950
22. Erlendsson S, Morado DR, Cullen HB, Feschotte C, Shepherd JD, Briggs JAG. Structures of virus-like capsids formed by the *Drosophila* neuronal Arc proteins. *Nat Neurosci*. 2020; 23(2):172–5. <https://doi.org/10.1038/s41593-019-0569-y> PMID: 31907439
23. Huang X, Miller W. A time-efficient, linear-space local similarity algorithm. *Advances in Applied Mathematics*. 1991; 12(3):337–57.
24. Cheng J, Sweredoski MJ, Baldi P. Accurate Prediction of Protein Disordered Regions by Mining Protein Structure Data. *Data Mining and Knowledge Discovery*. 2005; 11(3):213–22.
25. Dosztanyi Z, Csizmek V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. 2005; 21(16):3433–4. <https://doi.org/10.1093/bioinformatics/bti541> PMID: 15955779
26. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*. 2010; 1804(4):996–1010. <https://doi.org/10.1016/j.bbapap.2010.01.011> PMID: 20100603
27. Keller R. The Computer Aided Resonance Assignment Tutorial. 2004.
28. Rao VR, Pintchovski SA, Chin J, Peebles CL, Mitra S, Finkbeiner S. AMPA receptors regulate transcription of the plasticity-related immediate-early gene Arc. *Nat Neurosci*. 2006; 9(7):887–95. <https://doi.org/10.1038/nn1708> PMID: 16732277
29. Byers CE, Barylko B, Ross JA, Southworth DR, James NG, Taylor CA, et al. Enhancement of dynamin polymerization and GTPase activity by Arc/Arg3.1. *Biochim Biophys Acta*. 2015; 1850(6):1310–8. <https://doi.org/10.1016/j.bbagen.2015.03.002> PMID: 25783003
30. DaSilva LL, Wall MJ, PdA L, Wauters SC, Januario YC, Muller J, et al. Activity-Regulated Cytoskeleton-Associated Protein Controls AMPAR Endocytosis through a Direct Interaction with Clathrin-Adaptor Protein 2. *eNeuro*. 2016; 3(3).
31. Nikolaienko O, Eriksen MS, Patil S, Bito H, Bramham CR. Stimulus-evoked ERK-dependent phosphorylation of activity-regulated cytoskeleton-associated protein (Arc) regulates its neuronal subcellular localization. *Neuroscience*. 2017; 360:68–80. <https://doi.org/10.1016/j.neuroscience.2017.07.026> PMID: 28736134
32. Myrum C, Soule J, Bittins M, Cavagnini K, Goff K, Ziemek SK, et al. Arc Interacts with the Integral Endoplasmic Reticulum Protein, Calnexin. *Front Cell Neurosci*. 2017; 11:294. <https://doi.org/10.3389/fncel.2017.00294> PMID: 28979192
33. Bloomer WA, VanDongen HM, VanDongen AM. Activity-regulated cytoskeleton-associated protein Arc/Arg3.1 binds to spectrin and associates with nuclear promyelocytic leukemia (PML) bodies. *Brain Res*. 2007; 1153:20–33. <https://doi.org/10.1016/j.brainres.2007.03.079> PMID: 17466953
34. Gozdz A, Nikolaienko O, Urbanska M, Cymerman IA, Sitkiewicz E, Blazejczyk M, et al. GSK3alpha and GSK3beta Phosphorylate Arc and Regulate its Degradation. *Front Mol Neurosci*. 2017; 10:192. <https://doi.org/10.3389/fnmol.2017.00192> PMID: 28670266
35. Rogers S, Wells R, Rechsteiner M. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*. 1986; 234(4774):364–8. <https://doi.org/10.1126/science.2876518> PMID: 2876518

36. Nicolas A, Delalande O, Hubert JF, Le Rumeur E. The spectrin family of proteins: a unique coiled-coil fold for various molecular surface properties. *J Struct Biol.* 2014; 186(3):392–401. <https://doi.org/10.1016/j.jsb.2014.03.011> PMID: 24657228
37. Day C, Shepherd JD. Arc: building a bridge from viruses to memory. *Biochem J.* 2015; 469(1):e1–3. <https://doi.org/10.1042/BJ20150487> PMID: 26173260
38. Wee CL, Teo S, Oey NE, Wright GD, VanDongen HM, VanDongen AM. Nuclear Arc Interacts with the Histone Acetyltransferase Tip60 to Modify H4K12 Acetylation(1,2,3). *eNeuro.* 2014; 1(1).
39. Goo B, Sanstrum BJ, Holden DZY, Yu Y, James NG. Arc/Arg3.1 has an activity-regulated interaction with PICK1 that results in altered spatial dynamics. *Sci Rep.* 2018; 8(1):14675. <https://doi.org/10.1038/s41598-018-32821-4> PMID: 30279480