PLOS ONE

RESEARCH ARTICLE

Evolving Notch polyQ tracts reveal possible solenoid interference elements

Albert J. Erives*

Department of Biology University of Iowa Iowa City, IA, United States of America

* albert-erives@uiowa.edu

Abstract

Polyglutamine (polyQ) tracts in regulatory proteins are extremely polymorphic. As functional elements under selection for length, triplet repeats are prone to DNA replication slippage and indel mutations. Many polyQ tracts are also embedded within intrinsically disordered domains, which are less constrained, fast evolving, and difficult to characterize. To identify structural principles underlying polyQ tracts in disordered regulatory domains, here I analyze deep evolution of metazoan Notch polyQ tracts, which can generate alleles causing developmental and neurogenic defects. I show that Notch features polyQ tract turnover that is restricted to a discrete number of conserved "polyQ insertion slots". Notch polyQ insertion slots are: (*i*) identifiable by an amphipathic "slot leader" motif; (*ii*) conserved as an intact C-terminal array in a 1-to-1 relationship with the N-terminal solenoid-forming ankyrin repeats (ARs); and (*iii*) enriched in carboxamide residues (Q/N), whose sidechains feature dual hydrogen bond donor and acceptor atoms. Correspondingly, the terminal loop and β-strand of each AR feature conserved carboxamide residues, which would be susceptible to folding interference by hydrogen bonding with residues outside the ARs. I thus suggest that Notch polyQ insertion slots constitute an array of AR interference elements (ARIEs). Notch ARIEs would dynamically compete with the delicate serial folding induced by adjacent ARs. Huntingtin, which harbors solenoid-forming HEAT repeats, also possesses a similar number of polyQ insertion slots. These results suggest that intrinsically disordered interference arrays featuring carboxamide and polyQ enrichment may constitute coupled proteodynamic modulators of solenoids.

Citation: Erives AJ (2017) Evolving Notch polyQ tracts reveal possible solenoid interference elements. PLoS ONE 12(3): e0174253. https://doi.org/10.1371/journal.pone.0174253

Editor: Miguel A Andrade-Navarro, Johannes Gutenberg Universitat Mainz, GERMANY

Received: October 21, 2016

Accepted: March 6, 2017

Published: March 20, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The author(s) received no specific funding for this work.

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

Introduction

Polyglutamine (polyQ) tracts are functional features of many conserved transcriptional regulators and allow for complex conformational dynamics and interactions with other polyQ factors [1–9]. A prominent polyQ tract was first identified in the neurogenic gene *Notch* [10]. The Notch protein is central to a signaling pathway guiding patterning and cell fate decisions during metazoan development [11]. This polyQ tract is embedded in the Notch intracellular domain (NICD), which when cleaved leads to nuclear import and activation via DNA-bound CSL proteins. This polyQ tract is highly polymorphic in *Drosophila melanogaster* and can generate new alleles that cause developmental and neurogenic defects [7]. All of the identified

polymorphic alleles are specific to *Drosophila melanogaster* because this same tract is uniquely configured in most other *Drosophila* species as determined by the placement of an intervening histidine residue and the underlying CAX triplet nucleotide repeat pattern [7].

The single polyQ tract of *Drosophila* Notch proteins is embedded in a much larger intrinsically disordered protein (IDP) region, which is common to many transcriptional activators and co-activators, including NICD, TBP, and many of the Mediator co-activator subunits [12, 13]. It is also common to large scaffolding proteins that serve as signal integration platforms that are regulated by multiple protein-protein interactions. On such example is Huntingtin (Htt), which is conserved in humans, flies [14], and non-metazoan eukaryotes such as social slime molds [15]. Polyglutamine (polyQ) expansions in Htt [1, 8, 9, 16, 17], TBP [18, 19], and other conserved loci in humans underlie various progressive neurodegenerative diseases [1].

IDP regions of regulator and scaffolding proteins provide conformational flexibility and allow the formation of transient regulator complexes under specific conditions [12, 13]. IDPs form random coils or molten globules with very little protein secondary structure and this makes them exceedingly difficult to study biophysically. As such, these regions are typically removed in proteins subjected to structural studies. These regions are also difficult to align with orthologous sequences encoded in other genomes, and are a major impediment to accurate gene annotation in whole genome sequence assemblies. Furthermore, short read assemblies are intractable at loci encoding lengthy polyQ tracts.

Although some eukaryotes, such as the amoebozoan slime mold *Dictyostelium*, exhibit genome-wide trends in polyQ content [20–22], in general encoded polyQ tracts of conserved regulators expand and contract in a locus-specific manner. The N-terminal polyQ tract in human Htt is highly polymorphic and is expanded relative to early branching primates (*e.g.*, tarsiers), and other mammals (*e.g.*, mouse) **(Fig 1A, top)**. However, the *Drosophila* Htt ortholog features only a single glutamine residue at this position **(Fig 1A, top)**. Nonetheless, a different region of *Drosophila* Htt features a prominent polyQ tract, which is absent in vertebrates, including humans, despite the region of its appearance being conserved **(Fig 1A, bottom)**. Both of these polyQ "insertion slots" are preceded by a short leader motif that is predicted to form an amphipathic helix. In general, polyQ tracts evolve in an evolutionary turnover process, the study of which can lead to biophysical insights not immediately accessible via biochemical characterization.

To understand the biochemical constraints governing polyQ evolution, I considered the long polyQ tract in the Notch intracellular domain in *Drosophila*. We previously found that this polyQ tract is highly unstable and variable within the *Drosophila melanogaster* population [7]. This tract has been continuously evolving since the *Drosophila* radiation with different species exhibiting distinct polyQ tract configurations within the large IDP region of NICD [7]. Here I analyzed Notch proteins from diverse dipterans (flies) and discovered a previously unknown level of Notch polyQ evolution that is functionally revealing about the Notch intrinsically ordered region and its relation to the adjacent solenoid-forming ankyrin repeats (AR). I show that a polyQ slot insertion model provides a powerful interpretation of the functional interactions between polyQ tracts constrained to a limited number of slots in solenoid-type proteins such as the Notch intracellular domain and Huntingtin. The polyQ insertion model can be informed by use of the fast-evolving dipteran orthologs of human disease-related proteins featuring expanded polyQ tracts.

Results

I define "polyQ insertion slots" as specific positions in a protein where polyQ tracts are favored to occur during evolution. Strong support for a polyQ insertion model of protein evolution

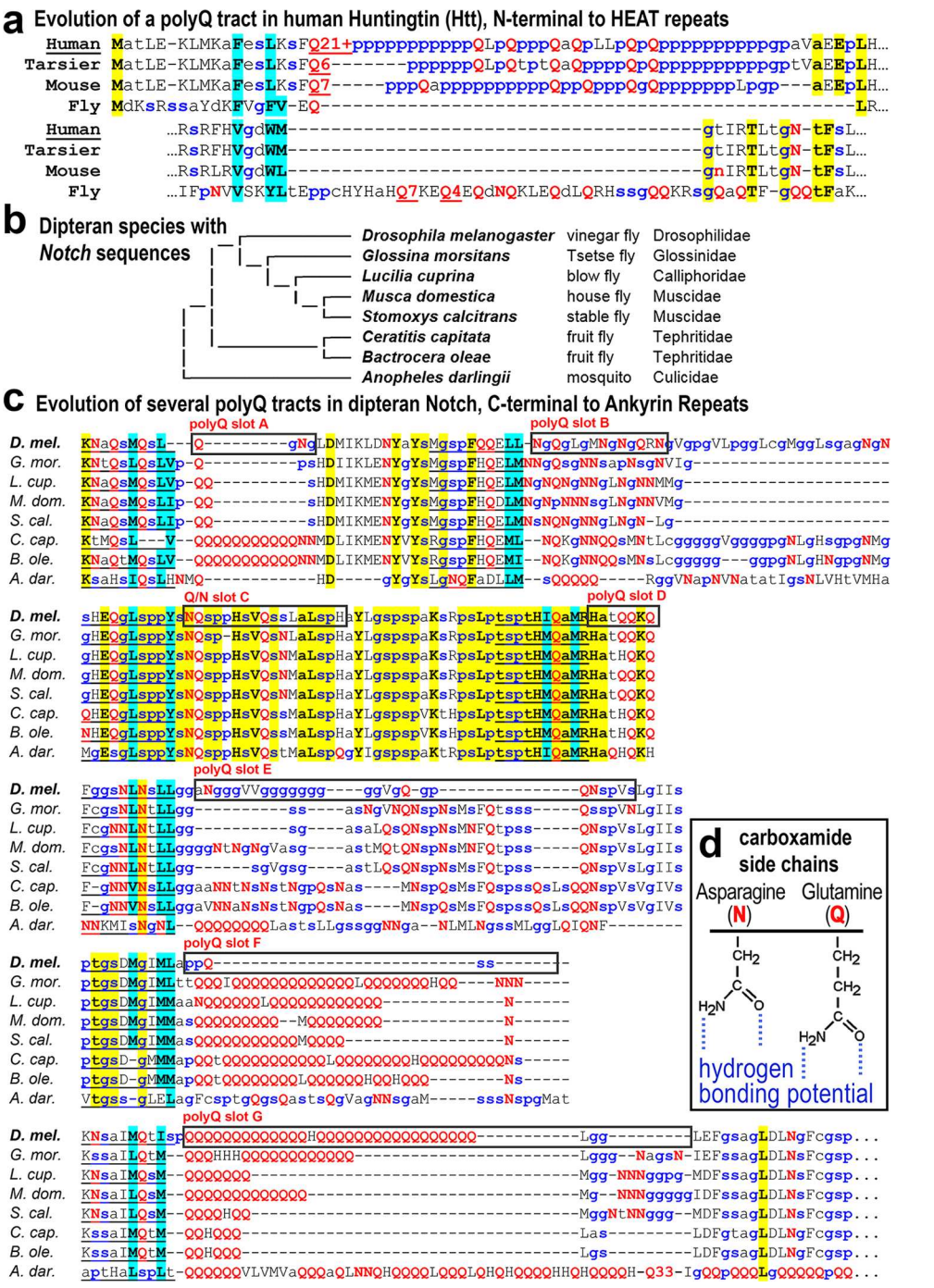

Fig 1. PolyQ tracts evolve in well-defined slots in Htt and Notch proteins. (a) Shown is the N-terminal region of Huntingtin (Htt) from humans, tarsier, mouse, and *Drosophila*. This region evolved a polyglutamine (polyQ) tract at a well-defined location in mammals, which later expanded in the evolution of humans, and is still evolving. While fly Htt is missing the N-terminal polyQ tract, it features a separate tract elsewhere in the protein as shown. Small amino acid residues that are secondary structure breakers are highlighted in blue. Glutamine residues are highlighted in red. Hydrophobic amino acid residues preceding the polyQ tract insertion site are highlighted in cyan. Conserved residues are highlighted in yellow. See Fig 3 for a diagram showing the location of polyQ tracts relative to other NICD domains. **(b)** Shown is an evolutionary tree of various dipteran genera for which Notch sequences were analyzed in this study. The cladogram is a simplified version of the comprehensive fly tree based on multiple nuclear and mitochondrial genes and morphological markers [42]. The six different families to which these species belong are listed. **(c)** Shown is the polyQ tract

regions of fly Notch proteins. Several sites (lettered slots) have independently evolved a polyQ tract in different fly genera. Most slots are preceded by a leader motif sequence, even when a polyQ tract is not present. **(d)** Inset shows the high hydrogen bonding potential of carboxamide amino acid side chains.

https://doi.org/10.1371/journal.pone.0174253.g001

could include (***i***) evidence of a non-random or constrained distribution of polyQ tracts across a set of diverged orthologs, (***ii***) evidence of a polyQ turnover process whereby polyQ tracts have been lost at one position and gained at another, (***iii***) identification of structural or peptide sequence motifs serving as polyQ slot leaders or trailers, and (***iv***) evidence of a functional relationship of polyQ slots to other functional domains within the protein or within a protein interactor.

To identify polyQ insertion slots in the Notch intracellular domain, I analyzed predicted Notch proteins from eight different genera in six different families of Diptera (**Fig 1B**). I find that polyQ tracts in dipteran Notch proteins have evolved in seven different well-defined polyQ insertion slots (**Fig 1C**). Each slots is preceded by a conserved polyQ slot leader motif similar to the Htt amphipathic helix, further confirming the utility of the polyQ insertion slot model. Even in the absence of polyQ tracts, slot leaders precede a region enriched in the carboxamide residues, asparagine (N) and glutamine (Q), whose sidechains feature hydrogen bond donor and acceptor groups (**Fig 1D**). These same slot regions are also enriched in the small amino acid residues that are secondary structure breakers ("g", "p", and "s").

Using the polyQ slot leader organizational model of the IDP region of dipteran NICD and a few reliably homologous slot leader motifs, I identified seven slots in the human Notch1 protein, as shown by an alignment to *Drosophila* NICD (**Fig 2A**). As in dipteran Notch proteins, these slots in human Notch1 are also enriched in both carboxamide residues and small secondary structure breakers that lead to IDPs. Thus, carboxamide-enriched IDPs with slot leaders likely represent a deeper organizational principle underlying many polyQ-rich regulators. I demonstrate the utility of this point as follows.

The seven N-terminal ankyrin repeats (ARs) in NICD form a solenoid-like domain, which fold by delicate intra-repeat interactions. Given the approximate 1-to-1 conservation of ARs to the IDP repeats with polyQ slots, it is possible that the IDP array is a kinetically active ensemble on one side of the linear AR solenoid. To investigate whether the polyQ IDP slots could be proteodynamic modulators of ARs, I analyzed the AR structural elements for carboxamide residues that could be hydrogen bonding with the carboxamide and polyQ rich IDPs. I find that the terminal halves of ARs are >10-fold enriched (13-to-1) in conserved carboxamide residues relative to the first half (**see boxed residues in Fig 2A, and labeled residues in Fig 2B**). Additional non-conserved carboxamide residues are also enriched in the terminal halves of ARs of both species, particularly in the terminal loop and strand elements (**red Q's and N's in Fig 2A**).

The terminal β-strand of each AR forms a hairpin with the starting β-strand of the next AR, and is important for induced propagation of solenoid formation [23]. I thus propose that the conserved array of seven polyQ insertion slots, which feature polyQ and/or carboxamide enriched IDPs, constitute AR interference elements (ARIEs). A pair of adjacent asparagines are also conserved in AR-5, which could be available for hydrogen bonding interactions in unfolded Notch AR solenoids (**Fig 2C**). Furthermore, the approximate 1-to-1 conservation of ARs to ARIEs in Notch suggest that ARIEs are proteodynamic modulators can be associated with specific solenoid repeats or pairs of repeats. The proteodynamic ARIEs array would attenuate AR solenoid formation via kinetic coupling and constitute a critical aspect of NICD regulation at multiple steps in the Notch signaling pathway.

Analogous to Notch and other AR-based solenoids, the proteins Htt, EF3, the regulatory A subunit of PP2A, and TOR1 (target of rapamycin) are solenoids based on HEAT repeats with

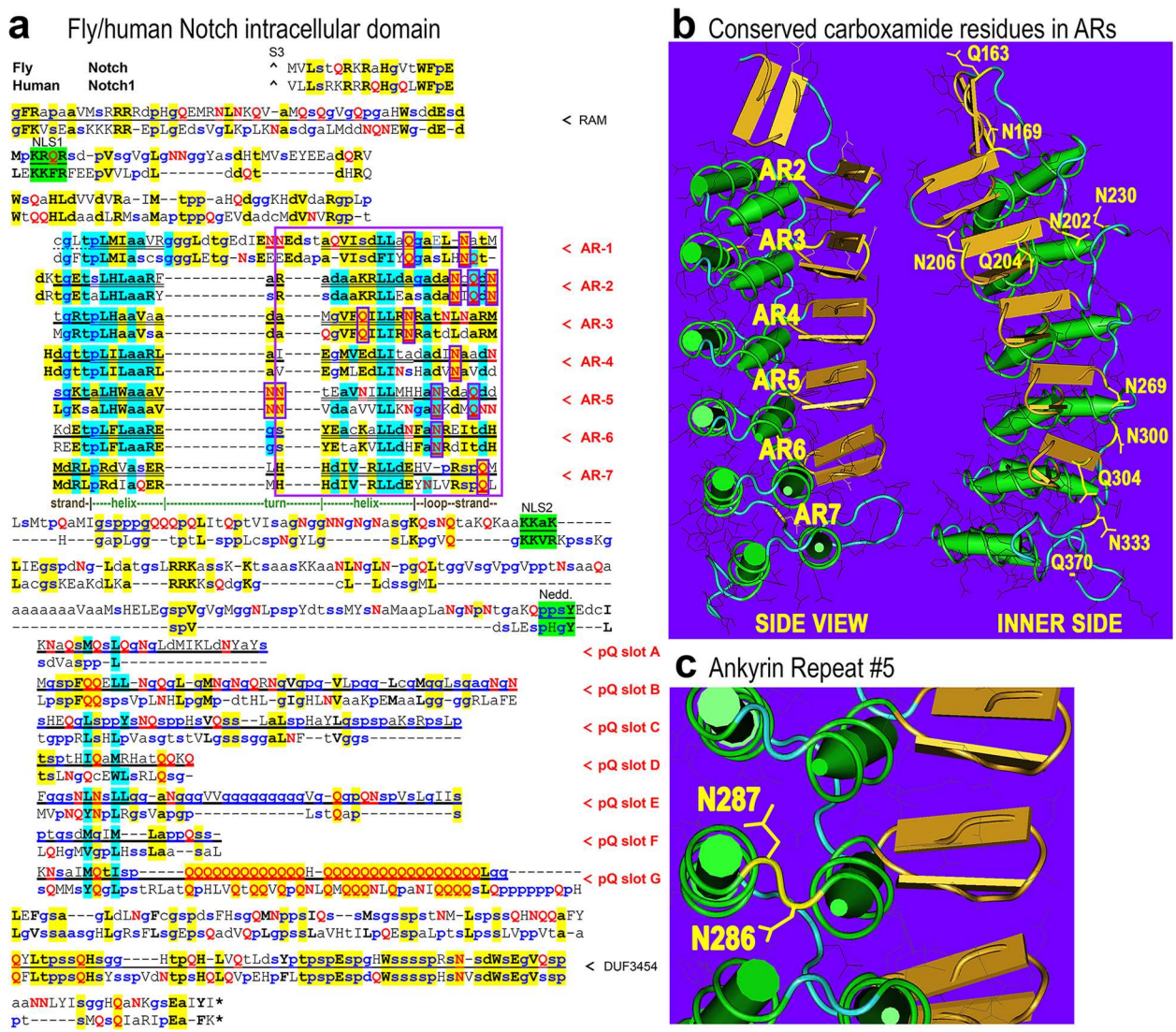

Fig 2. Notch polyQ slots are conserved IDP interference elements of ankyrin repeats. (a) Shown is an alignment of *Drosophila* Notch and human Notch1 intracellular domains (NICD) beginning at the S3 cleavage site (caret). The CSL/RBPJ-associated molecule (RAM) region is located at the N-terminus of NICD. Two nuclear localization signals (green with wavy underlining) flank the seven ankyrin repeats (ARs). The 33 residues of each AR features a helix-turn-helix peptide motif (helical regions in double underlining). This is followed by a conserved neddylation site and the six or seven polyQ insertion slots each starting with a polyQ slot leader sequence motif. A conserved domain of unknown function (DUF3454) follows the polyQ insertion slots. The polyQ slots fill up the region that is intrinsically disordered, but some slots are more conserved in sequence than others, such as slot-B, slot-D, and slot-G. See Fig 3 for a diagram showing the location of polyQ slots relative to other NICD domains. **(b)** The crystal structure of the human AR2–AR7 region is annotated to show the carboxamide residues that are conserved between fly Notch and human Notch1. These conserved Q/N residues and others not conserved are featured in the terminal loop and strand elements of each ankyrin repeat. **(c)** Shown is a close-up of ankyrin repeat 5 and the conserved double asparagines in the turn element.

https://doi.org/10.1371/journal.pone.0174253.g002

Htt having three HEAT repeats [24]. Thus, I further suggest that solenoid interference arrays featuring polyQ turnover dynamics are functionally-coupled components of solenoids. Solenoid folding is a delicate regulatory-prone process and is distinct from the folding seen in stable globular structures driven by hydrophobic packing and long-distance interactions. Thus, a folding funnel landscape view of NICD and Htt might be described by bi-stable minima involving a lower moat ARIEs-ARs interference state and a higher dimple solenoid state [25].

Alternatively, or possibly occurring only in special contexts, the AR solenoid of NICD may "organize" the folding of the intrinsically disordered region contains the ARIEs array.

The conserved carboxamide residues of NICD occupy the inner side of the AR solenoid **(Fig 2B, numbered residues on inner side)**. This side of the AR solenoid corresponds to the interaction surface in transcriptional complexes with Mastermind and CSL [23, 26]. Possibly, the polyQ richness of Mastermind (Mam/Maml) [4, 5, 27], a dedicated Notch-coactivator, functions as an AR solenoid chaperone by competing with the same solenoid surface entangled with the ARIEs array. Similarly, the Notch interacting protein Deltex features polyQ tracts in *Drosophila*, which has a longer polyQ tract than human Notch proteins, but not in human Deltex homologs (not shown). As such these functional intermolecular interactions would stem directly from AR-ARIE intramolecular coupling. Intermolecular ARIE-ARIE interactions in transcriptional NICD multimers [28, 29] may also relieve intramolecular inhibition of AR solenoid formation by the ARIEs array. Thus, polyQ expansions and contractions are likely to be under complex selection for their regulatory interactions with the adjacent solenoid repeats and their regulatory modulators.

Analysis of the evolutionary turnover of polyQ tracts in dipteran NICD moieties provides a novel perspective of the IDP region of NICD, and possibly completes the functional domain inventory of a key developmental signaling molecule **(Fig 3A)**. ARIEs are evident by evolutionary polyQ slot turnover and by the subsequently identified common leader motif resembling the amphipathic leader in Huntingtin. To see if additional peculiarities pertain to the NICD slot leaders, I derived an NICD-specific polyQ slot leader motif by taking the nine residue peptide leader sequences from *Drosophila* and *Stomoxys* slot leaders except for those from the hypothesized polyQ slot-C, for which polyQ insertions have not yet been seen **(Fig 3B)**. Not only is this a short amphipathic helix with hydrophobic residues on one side but there is typically at least one glutamine residue adjacent to the hydrophobic side **(Fig 3B)**. Thus, the leader motif itself may function to both display and interact with polyQ epitopes or carboxamide-rich IDPs in slots, or with the carboxamide rich terminal elements of associated ankyrin repeats.

Discussion

Altogether, the dipteran Notch proteins reveal the evolutionary appearance of lengthy polyQ tracts in five of the seven ARIEs and constitute an evolutionary Rosetta Stone for understanding NICD and its disordered carboxamide interference elements **(Fig 3C–3F)**. The single polyQ tract of *Drosophila melanogaster* is polymorphic and homologous to the single polyQ tract that is configured differently in other *Drosophila* species. However, order-wide patterns across multiple dipteran genera are needed to reveal the entire expanse within which polyQ turnover happens. Similar deep evolutionary profiling of polyQ turnover in other conserved regulators may reveal additional solenoid-modulating interference arrays. Thus fast evolving dipteran genera may provide an ideal evolutionary system for understanding mutant human proteins associated with proteotoxicity diseases.

These results also raise the likelihood of blind spots in our understanding of solenoids formed by ankyrin repeats in Notch and other AR solenoids, HEAT repeats in Htt and other HEAT solenoids, and others. Much, much more work will be needed to understand how these delicate structures are coupled in turn to intrinsically disordered modules. Biophysical studies that necessarily start with truncated solenoid domains are unlikely to approximate the more dynamic lives these domains have in their full-length proteins and within the cell. Furthermore, studies focused on the biochemistry of Q/N-rich and polyQ-rich polypeptides are providing evidence of complex conformational dynamics involving multiple structural states including coiled-coils, α-helices, and β-sheets, and distinct properties promoting or

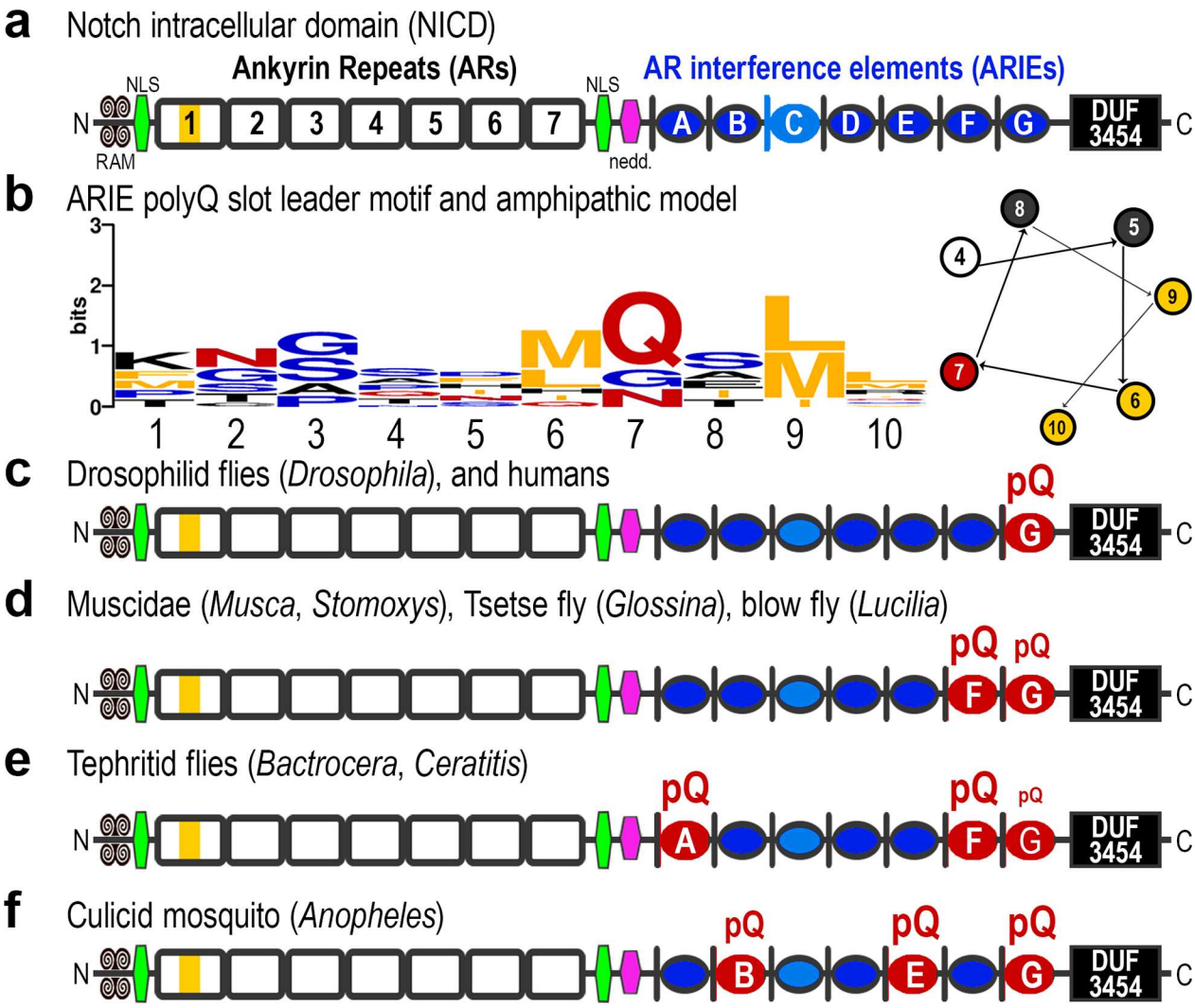

Fig 3. Notch AR interference elements (ARIEs) are present as a distinct array within NICD. (a) The polyQ slot leader motif defines 7 possible insertion sites, for which six have independently evolved polyQ tracts in distinct lineages of flies. The number of slots is also suggestive of an interference with the seven ankyrin repeats or the six pairs of adjacent repeats. **(b)** The leader sequences from *Drosophila* and *Stomoxys* polyQ insertion slots (except for slot C) were used to produce the motif logo shown. Also shown is an amphipathic helix starting with residue number four and showing how the hydrophobic amino acids at positions 6, 9, and 10 of the slot leader motif are segregated to one side of the helix (yellow circles). Interestingly, position 7 frequently features a single glutamine residue, which could interact with an adjacent polyQ tract. **(c)** *Drosophila* Notch and vertebrate Notch1 feature a polyQ tract in slot-G, although it is much more expanded in *Drosophila* than humans. **(d, e)** Unlike *Drosophila* Notch, the Notch proteins from several other fly genera, including *Musca*, *Stomoxys*, *Glossina*, *Lucilia*, *Bactrocera*, and *Ceratitis*, feature a more prominent polyQ tract in slot-F. These flies also have a much smaller polyQ tract in slot-G. The tephritid genera also feature a tract in slot-A, demonstrating that N-terminal slots can also accept polyQ tracts. **(f)** The Notch protein from *Anopheles darlingii* features polyQ tracts in slot-B and slot-E in addition to a very long tract in slot-G.

https://doi.org/10.1371/journal.pone.0174253.g003

disallowing multimerization and/or aggregation [30–32]. Theoretical modeling has also provided some support for the ability of these conformational states to be mechanically coupled to the cell cytoskeleton [33]. In this regard, a recent study has tested the complex dynamics bound to emerge in a full-length Huntingtin protein and has found evidence for the proposed intra-repeat modulatory effect on solenoid formation by its N-terminal polyQ tract [34]. Additional studies in which these expanded polyQ tracts are moved either to polyQ insertion slots

identified by evolutionary comparisons or to negative control slots could highlight the extent to which the concept of polyQ insertion slots is productive. If so, the identification of such slots in fast-evolving dipteran orthologs of key human proteins should be prioritized.

The identification of satisfying multiple sequence alignments (MSA) for fast-evolving proteins enriched in intrinsically-disordered peptide regions is an important but difficult problem. The absence of protein-folding domains in these regions exacerbates the underlying problem. Furthermore, such regions are highly tolerant of insertions and deletions (indels), particularly for homopolymeric runs. To this end, newer advanced MSA methods, such as Bayesian Markov chain Monte Carlo samplers that can infer variable gap penalties (HMM transition probabilities for indels) from sequence data, could be extremely useful [35]. Such methods could also be applied to fast-evolving *cis*-regulatory sequences that tend to become enriched in microsatellite repeats [36, 37]. In the case of the NICD ARIEs region, polyQ slot leaders, if well conserved, can serve as ungapped alignment blocks separated by gap-tolerant linkers. Similarly, in the case of *cis*-regulatory DNAs, transcription factor binding motifs can serve as ungapped alignment blocks separated by gap-tolerant and MSR-enriched linkers. Thus, examples such as the NICD ARIEs can serve as a valuable data set for the development of newer MSA methods that can learn complex patterns of variable gap tolerance.

Methods

Protein structures were analyzed and annotated using NCBI's Cn3D 4.3.1 software (http://www.ncbi.nih.gov/Structure) and high resolution structures of the ankyrin repeats of human Notch1 (PDB ID: 2F8Y, MMDB: 38239) [26]. The motif logo of Fig 3B was constructed using WebLogo (http://weblogo.berkeley.edu/logo.cgi). The source sequences for dipteran Notch proteins are from the following accessions and additional analyses: AAC36153.1, AAC36151.1 (*Lucilia cuprina*), XP_011292982.1 *(Musca domestica*), XP_013109509.1 (*Stomoxys calcitrans*), XP_004535280.1 (*Ceratitis capitata*), and XP_014087559.1 (*Bactrocera oleae*). The NICD sequence of *Glossina morsitans* (Tsetse fly) is a conceptual translation of CCAG010014983. The NICD sequence of *Anopheles darlingii* is a combination of ETN64594 and a splice corrected sequence from ADMH02000947. Alignments of Notch intrinsically disordered protein regions were explored by extensive variation of parameter space using both CLUSTALW [38] and MUSCLE [39, 40] on MEGA7 [41], but these were incongruent with each other and frequently subject to polyQ alignment artifact. Computed alignments improved when the sequences were anchored and or trimmed to include only the N-terminal ARIEs slot leaders to the highly conserved C-terminal DUF3454 sequence. Alignments in their final form shown in the figures were constructed manually and represent a maximization of the number of optimal local alignments. S1 and S2 Files contain text editable versions of the alignments of Figs 1 and 2.

Supporting information

S1 File. Sequences used in alignment of polyQ tracts in fly NICD proteins.
(DOCX)

S2 File. Full-length alignment of NICD for fly, bee, and human Notch proteins.
(DOCX)

Acknowledgments

I thank members of the Erives Laboratory for feedback at lab meetings and for asking good questions. I thank members of the Kreinitz Laboratory for the same. I thank Dr. Jan Fassler for providing critical feedback on an earlier version of this manuscript.

Author Contributions

Conceptualization: AJE.

Data curation: AJE.

Formal analysis: AJE.

Investigation: AJE.

Methodology: AJE.

Visualization: AJE.

Writing – original draft: AJE.

Writing – review & editing: AJE.

References

1. Trottier Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G, et al. Polyglutamine Expansion as a Pathological Epitope in Huntingtons-Disease and 4 Dominant Cerebellar Ataxias. Nature. 1995; 378 (6555):403–6. https://doi.org/10.1038/378403a0 PMID: 7477379
2. Orr HT. Polyglutamine neurodegeneration: expanded glutamines enhance native functions. Curr Opin Genet Dev. 2012; 22(3):251–5. https://doi.org/10.1016/j.gde.2012.01.001 PMID: 22284692
3. Tautz D. Hypervariability of Simple Sequences as a General Source for Polymorphic DNA Markers. Nucleic Acids Res. 1989; 17(16):6463–71. PMID: 2780284
4. Newfeld SJ, Schmid AT, Yedvobnick B. Homopolymer length variation in the Drosophila gene mastermind. J Mol Evol. 1993; 37(5):483–95. PMID: 8283480
5. Newfeld SJ, Tachida H, Yedvobnick B. Drive-selection equilibrium: homopolymer evolution in the Drosophila gene mastermind. J Mol Evol. 1994; 38(6):637–41. PMID: 8083889
6. Schlotterer C, Vogl C, Tautz D. Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural Drosophila melanogaster populations. Genetics. 1997; 146(1):309–20. PMID: 9136020
7. Rice C, Beekman D, Liu L, Erives A. The Nature, Extent, and Consequences of Genetic Variation in the opa Repeats of Notch in Drosophila. G3 (Bethesda). 2015; 5(11):2405–19. PubMed Central PMCID: PMCPMC4632060.
8. Andresen JM, Gayan J, Cherny SS, Brocklebank D, Alkorta-Aranburu G, Addis EA, et al. Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds. J Med Genet. 2007; 44(1):44–50. PubMed Central PMCID: PMCPMC2597910. https://doi.org/10.1136/jmg.2006.045153 PMID: 17018562
9. Andresen JM, Gayan J, Djousse L, Roberts S, Brocklebank D, Cherny SS, et al. The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. Ann Hum Genet. 2007; 71(Pt 3):295–301. https://doi.org/10.1111/j.1469-1809.2006.00335.x PMID: 17181545
10. Wharton KA, Yedvobnick B, Finnerty VG, Artavanis-Tsakonas S. opa: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in D. melanogaster. Cell. 1985; 40(1):55–62. PMID: 2981631
11. Guruharsha KG, Kankel MW, Artavanis-Tsakonas S. The Notch signalling system: recent insights into the complexity of a conserved pathway. Nat Rev Genet. 2012; 13(9):654–66. https://doi.org/10.1038/nrg3272 PMID: 22868267
12. Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ. Malleable machines take shape in eukaryotic transcriptional regulation. Nat Chem Biol. 2008; 4(12):728–37. PubMed Central PMCID: PMCPMC2921704. https://doi.org/10.1038/nchembio.127 PMID: 19008886
13. Toth-Petroczy A, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, et al. Malleable machines in transcription regulation: the mediator complex. PLoS Comput Biol. 2008; 4(12):e1000243. PubMed Central PMCID: PMCPMC2588115. https://doi.org/10.1371/journal.pcbi.1000243 PMID: 19096501
14. Li Z, Karlovich CA, Fish MP, Scott MP, Myers RM. A putative Drosophila homolog of the Huntington's disease gene. Human Molecular Genetics. 1999; 8(9):1807–15. PMID: 10441347

15. Myre MA, Lumsden AL, Thompson MN, Wasco W, MacDonald ME, Gusella JF. Deficiency of Huntingtin Has Pleiotropic Effects in the Social Amoeba Dictyostelium discoideum. Plos Genet. 2011; 7(4).
16. Shelbourne PF, Keller-McGandy C, Bi WL, Yoon SR, Dubeau L, Veitch NJ, et al. Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. Hum Mol Genet. 2007; 16(10):1133–42. https://doi.org/10.1093/hmg/ddm054 PMID: 17409200
17. Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, et al. Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. Nat Genet. 1993; 5(2):174–9. https://doi.org/10.1038/ng1093-174 PMID: 8252043
18. Zuhlke C, Hellenbroich Y, Dalski A, Kononowa N, Hagenah J, Vieregge P, et al. Different types of repeat expansion in the TATA-binding protein gene are associated with a new form of inherited ataxia. Eur J Hum Genet. 2001; 9(3):160–4. https://doi.org/10.1038/sj.ejhg.5200617 PMID: 11313753
19. Rubinsztein DC, Leggo J, Crow TJ, Delisi LE, Walsh C, Jain S, et al. Analysis of polyglutamine-coding repeats in the TATA-Binding protein in different human populations and in patients with schizophrenia and bipolar affective disorder. Am J Med Genet. 1996; 67(5):495–8. https://doi.org/10.1002/(SICI)1096-8628(19960920)67:5<495::AID-AJMG12>3.0.CO;2-I PMID: 8886170
20. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, et al. The genome of the social amoeba Dictyostelium discoideum. Nature. 2005; 435(7038):43–57. https://doi.org/10.1038/nature03481 PMID: 15875012
21. Scala C, Tian XJ, Mehdiabadi NJ, Smith MH, Saxer G, Stephens K, et al. Amino Acid Repeats Cause Extraordinary Coding Sequence Variation in the Social Amoeba Dictyostelium discoideum. Plos One. 2012; 7(9).
22. Santarriaga S, Petersen A, Ndukwe K, Brandt A, Gerges N, Scaglione JB, et al. The Social Amoeba Dictyostelium discoideum Is Highly Resistant to Polyglutamine Aggregation. J Biol Chem. 2015; 290 (42):25571–8. https://doi.org/10.1074/jbc.M115.676247 PMID: 26330554
23. Ehebauer MT, Chirgadze DY, Hayward P, Martinez Arias A, Blundell TL. High-resolution crystal structure of the human Notch 1 ankyrin domain. Biochem J. 2005; 392(Pt 1):13–20. PubMed Central PMCID: PMCPMC1317659. https://doi.org/10.1042/BJ20050515 PMID: 16011479
24. Andrade MA, Bork P. HEAT repeats in the Huntington's disease protein. Nat Genet. 1995; 11(2):115–6. https://doi.org/10.1038/ng1095-115 PMID: 7550332
25. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol. 1997; 4(1):10–9. PMID: 8989315
26. Nam Y, Sliz P, Song LY, Aster JC, Blacklow SC. Structural basis for cooperativity in recruitment of MAML coactivators to Notch transcription complexes. Cell. 2006; 124(5):973–83. https://doi.org/10.1016/j.cell.2005.12.037 PMID: 16530044
27. Newfeld SJ, Smoller DA, Yedvobnick B. Interspecific comparison of the unusually repetitive Drosophila locus mastermind. J Mol Evol. 1991; 32(5):415–20. PMID: 1904096
28. Vasquez-Del Carpio R, Kaplan FM, Weaver KL, VanWye JD, Alves-Guerra MC, Robbins DJ, et al. Assembly of a Notch Transcriptional Activation Complex Requires Multimerization. Mol Cell Biol. 2011; 31(7):1396–408. https://doi.org/10.1128/MCB.00360-10 PMID: 21245387
29. Kelly DF, Lake RJ, Walz T, Artavanis-Tsakonas S. Conformational variability of the intracellular domain of Drosophila Notch and its interaction with Suppressor of Hairless. Proc Natl Acad Sci U S A. 2007; 104(23):9591–6. PubMed Central PMCID: PMCPMC1887580. https://doi.org/10.1073/pnas.0702887104 PMID: 17535912
30. Fiumara F, Fioriti L, Kandel ER, Hendrickson WA. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. Cell. 2010; 143(7):1121–35. PubMed Central PMCID: PMCPMC3472970. https://doi.org/10.1016/j.cell.2010.11.042 PMID: 21183075
31. Petrakis S, Schaefer MH, Wanker EE, Andrade-Navarro MA. Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. Bioessays. 2013; 35(6):503–7. PubMed Central PMCID: PMCPMC3674527.
32. Pelassa I, Cora D, Cesano F, Monje FJ, Montarolo PG, Fiumara F. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. Hum Mol Genet. 2014; 23(13):3402–20. ubMed Central PMCID: PMCPMC4049302. https://doi.org/10.1093/hmg/ddu049 PMID: 24497578
33. Chen M, Zheng W, Wolynes PG. Energy landscapes of a mechanical prion and their implications for the molecular mechanism of long-term memory. Proc Natl Acad Sci U S A. 2016; 113(18):5006–11. https://doi.org/10.1073/pnas.1602702113 PMID: 27091989
34. Vijayvargia R, Epand R, Leitner A, Jung TY, Shin B, Jung R, et al. Huntingtin's spherical solenoid structure enables polyglutamine tract-dependent modulation of its structure and function. Elife. 2016; 5. PubMed Central PMCID: PMCPMC4846397.

35. Neuwald AF, Altschul SF. Bayesian Top-Down Protein Sequence Alignment with Inferred Position-Specific Gap Penalties. PLoS Comput Biol. 2016; 12(5):e1004936. PubMed Central PMCID: PMCPMC4871425. https://doi.org/10.1371/journal.pcbi.1004936 PMID: 27192614

36. Brittain A, Stroebele E, Erives A. Microsatellite repeat instability fuels evolution of embryonic enhancers in Hawaiian Drosophila. Plos One. 2014; 9(6):e101177. PubMed Central PMCID: PMCPMC4076327. https://doi.org/10.1371/journal.pone.0101177 PMID: 24978198

37. Crocker J, Potter N, Erives A. Dynamic evolution of precise regulatory encodings creates the clustered site signature of enhancers. Nat Commun. 2010; 1:99. PubMed Central PMCID: PMCPMC2963808. https://doi.org/10.1038/ncomms1102 PMID: 20981027

38. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22(22):4673–80. PubMed Central PMCID: PMCPMC308517. PMID: 7984417

39. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004; 5:113. PubMed Central PMCID: PMCPMC517706. https://doi.org/10.1186/1471-2105-5-113 PMID: 15318951

40. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32(5):1792–7. PubMed Central PMCID: PMCPMC390337. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

41. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016.

42. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, et al. Episodic radiations in the fly tree of life. Proc Natl Acad Sci U S A. 2011; 108(14):5690–5. PubMed Central PMCID: PMCPMC3078341. https://doi.org/10.1073/pnas.1012675108 PMID: 21402926