RESEARCH PAPER
https://doi.org/10.1071/CH23084

AUSTRALIAN JOURNAL OF CHEMISTRY

Polyadenine complexed to polyglutamine suggests the peptide backbone has a *cis* conformation

Anthony B. Bransgrove[A,*] and Louise Amelia Anderson[B,C]

For full list of author affiliations and declarations see end of paper

***Correspondence to:**
Anthony B. Bransgrove
National Diagnostic Products, The SAND Institute, 7–9 Merriwa Street, Gordon, NSW 2072, Australia
Email: anthony@ndp.com.au

Handling Editor:
John Wade

ABSTRACT

The polyadenine (poly(A)) tail of mRNA is a homopolymer and as such is a potential H-bonding partner for other cellular homopolymers. The secondary structure of proteins and peptides employs the polar groups of the homopolymer backbone to bind with themselves or other polymers to respectively form an alpha helix or a beta sheet. These same backbone polar groups appear to be suitably positioned to bind with poly(A) but only when the backbone has rotated to the all-*cis* conformation. However the all-*cis* conformation of peptides or proteins is rarely encountered and so they would seem to be unlikely binding partners. Here we show that a homopolymer of glutamine (poly(Q)) may be an exception to this rarity because its default conformation seems to be all-*cis*. We found that exposure of poly(A) to increasing amounts of poly(Q) causes a progressive loss of soluble poly(A) migrating on gel electrophoresis, indicating binding between the two polymers. Stereochemical modelling of the likely complex has suggested the particular polar groups that are responsible for binding of the two polymers, along with an explanation for the all-*cis* conformation of poly(Q).

Keywords: carboxamide chain cooperativity, *cis* conformation of peptide backbone, mRNA polyadenine tail, nascent peptides as target for polyadenine, neurodegenerative disease, peptide bound adenine, polyglutamine aggregates, polyglutamine binding to promoter opens double helix.

Introduction

Functions

Polyadenylation is the non-template addition of a polyadenine (poly(A)) tail to the 3′ end of an RNA transcript, giving rise to the mature mRNA. Poly(A) is known to play a role in the nuclear export of RNA, to act as a binding site for poly(A) binding proteins (PABP) and to be involved in translation initiation, and its diminishing length in the cytoplasm is linked to mRNA decay.[1]

Lineage

Polyadenylation is a near universal modification present in all forms of life – eukaryotes, bacteria and archaea.[2] Its presence in prokaryotes suggests its origins are ancient and probably extend back to the mid Precambrian era (2.5–3.5 Ma).[3]

Evolutionary selection pressures over such a prolonged period would suggest that the homopolymer of adenine may itself be of primary importance, whereas the well-known poly(A) binding proteins and other functions have adapted to bind with the stable and enduring poly(A) tail.

Purpose

What purpose then might be served by an adenine homopolymer? In this communication, evidence is presented for credible poly(A) binding partners, their source and their composition. Experimental data are then given indicating the formation of a complex

Received: 4 May 2023
Accepted: 26 June 2023
Published: 18 July 2023

Cite this:
Bransgrove AB and Anderson LA (2023) *Australian Journal of Chemistry* **76**(9), 590–599. doi:10.1071/CH23084

between the binding partner and poly(A). Next, a functional molecular structure is proposed based wholly on stereochemical considerations. Finally, a discussion encompasses downstream implications of poly(A) binding as a complex.

Partnering possibilities

In the case of proteins, the homopolymer backbone carries a regular repeat of polar groups able to form H bonds with similar groups on the same or separate chains, establishing the secondary structure of an alpha helix or a beta sheet.[4] Nucleic acids, on the other hand, only show a regular repeat of polar groups when formed as a homopolymer such as poly (A). The regular repeat in poly(A) suggests that, like proteins, it may have the capacity to interact with other homopolymers[5] – obviously other nucleic acids, including itself, but other plausible partners for this role such as homorepeat-containing peptides or proteins themselves.

Peptide homopolymers

Microsatellite DNA, also known as simple sequence repeats (SSRs), are tracts of 1–6 base pairs, repeated typically 5–50 times.[6] Of these simple sequences, trinucleotide repeats are codon length, and when present in coding regions, specify homopolymer tracts of amino acids carried within the host protein. These amino acid repeats are found in a large number of eukaryotic and prokaryotic proteins including approximately 20% of human proteins,[7] suggesting they serve a normal and necessary functional role.

Mutability

Microsatellites are highly mutable,[8] particularly in many cancers for which instability giving rise to expansions can be both predictive and prognostic for the disease.[9] In non-malignant disorders, expansions of these tracts can become pathological when a threshold size is reached,[10,11] and they are associated with inherited diseases, particularly neurodegenerative states where intracellular aggregates of the homopolymer, along with its harbouring protein, appear. It has recently been reported that microsatellite expansions are responsible for more than 50 human genetic diseases.[12]

Expansion disorders

Prominent among the expansion disorders are the CAG repeats, coding for glutamine, GCC repeats, coding for alanine and CGG repeats, coding for glycine. There are 10 known polyglutamine expansion disorders, exemplified by Huntington's disease. All have a pathological length threshold of ~37 repeats,[10,13] all are within transcription factors (TFs) and all are neurodegenerative diseases.[14] In the case of polyalanine there are nine known expansion disorders with a pathological length threshold of ~20 repeats.[15] Of these, eight involve developmentally important transcription factors and are responsible for causing congenital malformations.[16] Polyglycine expansion disorders exceed five and have a pathological length threshold of ~50 repeats.[12] They are noteworthy because they frequently arise by non-canonical replication from non-coding gene regions.[17] Foremost within this group are Fragile X mental retardation (FXMR) and Fragile X tremor with ataxia (FXTAS).[18]

Structure

Homopolymer interactions, if they were to occur, between poly(A) and peptides or proteins might involve the amino acid side-chain repeats, or equally, the peptide backbone of a homopolymer, which alone is the common element linking the amino acids of the repeat (or the non-repeats of a natural protein) and which presents a recurrent pattern of polar groups potentially able to pair with those of the nucleic acid.

The peptide bond

The peptide bond has a partial double bond character that restricts free rotation around the bond to two torsion angles of ~0 and 180°, causing the peptide group of six surrounding atoms all lie in the one plane. Peptide bonds are mostly found in the *trans* configuration with a torsion angle ω close to 180°; however, a small number in the non-proline *cis* conformation also occur,[19] with ω angles close to 0°.

Experimental

Polymer length is nominated in mer number when relevant, next to each monomer abbreviation.

Peptides

The following peptides were used from the indicated sources: poly(Q15) (Ac-Gln15), poly(QKK) (Ac-Gln15LysLys) and poly(Q8) (Ac-Gln8Lys), purchased from GenScript Inc; poly(Ala) (polyalanine15) purchased from Sigma/Merck.

Polynucleotides

The following polynucleotides were used from the indicated sources: poly(A20) (polyadenine 20) and poly(dA)–poly(dT) (polydeoxyadenine20–polydeoxythymine20) duplex purchased from Integrated DNA Technologies (IDT) and Sigma–Aldrich.

Electrophoresis

Gels were cast in empty cassettes (Bio-Rad) in acrylamide concentrations of 16% w/v, and run using the Bio-Rad Mini PROTEAN Tetra Cell system.

Electrophoretic mobility shift assay (EMSA)

Nucleic acid–protein interactions are a widespread and crucial part of the many processes required for gene expression. The EMSA is perhaps one of the most informative and frequently used methods to identify and characterise these interactions. Reagents and protocols for this study were taken from Hellman and Fried,[20] with silver staining from Tang *et al.*[21]

Molecular modelling

Skeletal modelling for stereochemical work employed the Orbit Molecular Building System:

> Sufficiently accurate and rigid enough to make satisfactory models. [L. Bragg in a 1970 foreword to a Cochranes of Oxford manual on building molecular models with the orbit molecular building system]

Results

Polyacrylamide gel electrophoresis

EMSA has been used in the present investigation in an attempt to detect binding between poly(A) and two (glycine and alanine) of the three most commonly encountered repeat amino acids that are frequently associated with genetic expansion disorders (i.e. homopolymers of glycine, alanine and glutamine). Each peptide of glycine and alanine was prepared as a 0.5-mM stock in dH_2O. Aliquots were added to microtubes in amounts of 0, 2, 3, 4, 5, 6, 7, 8, 9 and 10 µL, then each was supplemented with dH_2O to a final volume of 10 µL. To each microtube was then added a 10-µL aliquot of stock 15.0-µM poly(A) in 2× binding buffer (20 mM of Tris-acetate pH 7.8 at 20°C, 2 mM of EDTA, 10% glycerol). Each mixture was incubated for 30 min at room temperature. Samples of 10 µL were then removed and resolved on a 16% acrylamide gel. Binding experiments were run at least in triplicate.

Fig. 1 shows a gel segment for the poly(Ala) binding reaction. Following electrophoresis and staining, all poly(A) bands, including the control, are equivalent in size and intensity with no response to increasing poly(Ala) concentrations.

The second homopolymer investigated was polyglutamine (poly(Q)). Significantly, a clear response was observed upon incubating with poly(A). After electrophoresis and silver staining, a stepwise gradual loss of the poly(A) band size and intensity occurs as seen in Fig. 2, and this accompanies the incremental increase of the poly(Q) content.

The inverse relationship between the poly(Q) content and the poly(A) band size and intensity does suggest the binding of poly(A) by poly(Q), but a second slower band attributable to a migrating complex is not apparent. It is not uncommon

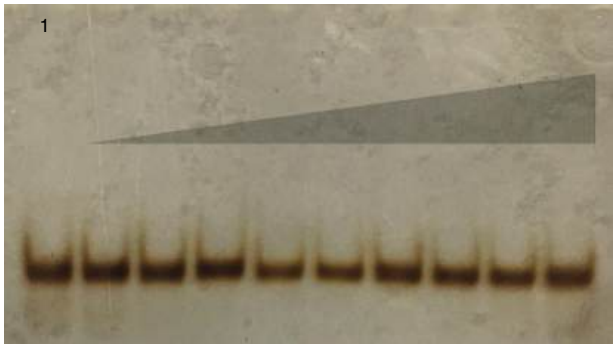

Fig. 1. Gel image of binding reaction between poly(Ala) and poly (A), showing the same poly(A) band intensity for all reaction mixtures containing varied poly(Ala) amounts. The peptide concentration is represented by the grey wedge for each of the gels.

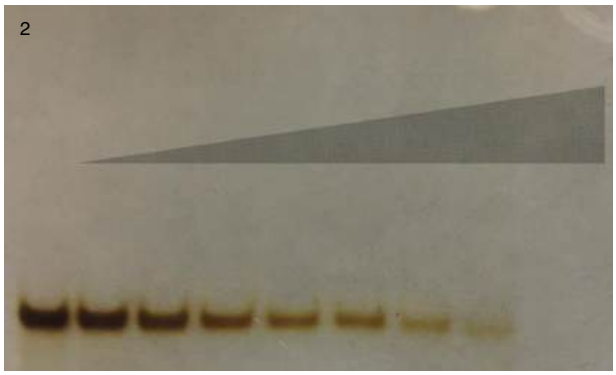

Fig. 2. Gel image of binding reaction between poly(Q) and poly(A), showing progressive loss of band intensity proportional to the binding mix poly(Q) content. No second slower group of bands is observed to indicate a complex is formed.

for a complex to be excluded from the gel, usually because its size is too large for the gel pores,[22] and in the present study this was suspected to be the case because of the aggregation prone nature of poly(Q). An attempt was made to reduce aggregation by using a more soluble form of poly(Q) that had been terminated with two lysine residues poly(Q_{LL}). This failed to reveal a slower second band, and it did not alter the poly(A) band loss pattern (Fig. 3).

The robustness of the binding reaction was further tested by using poly(A) when in the form of a double helix. This was thought to possibly isolate it from the action of poly(Q). Poly(A) in the binding reaction was replaced with a poly (dA)–poly(dT) duplex, the deoxy form of poly(A) also being a feasible target in the binding assay.

Fig. 4 reveals that, unexpectedly, the poly(dA)–poly(dT) duplex, used to contain the double helical form of poly(A), appears to be susceptible to the same actions of increasing amounts of poly(Q), showing a similar pattern of band loss to poly(A). The gel also shows a very pale band that increases with intensity along with the increase in poly(Q) content, but then there is a sudden fall of intensity in the last

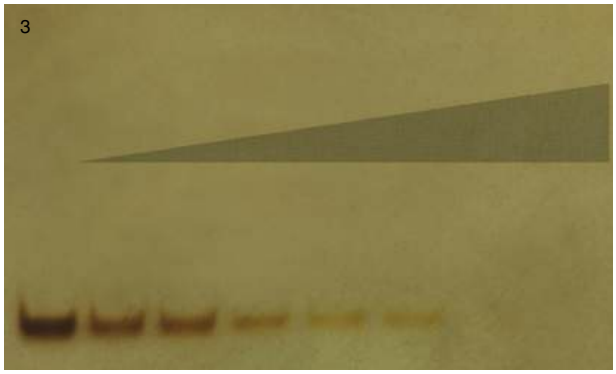

Fig. 3. Gel image of binding reaction between poly(QKK) and poly(A). Despite solubility enhancement the loss of band intensity is clearly apparent and similar to that seen using poly(Q) in Fig. 2.

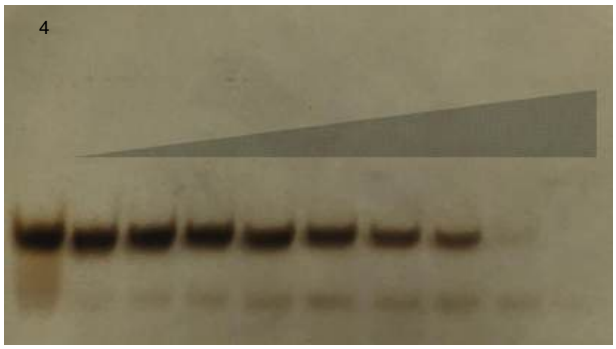

Fig. 4. Gel image of the binding reaction between the poly(dA)–poly(dT) duplex and poly(Q). There is a progressive loss of poly(dA)–poly(dT) duplex band intensity with increasing poly(Q) content.

two lanes, not thought to be a random effect because it is repeatable. The pale band is poly(T), possibly displaced from the duplex by poly(Q). Thymine is reportedly not stained by silver[21]; however, we have found that some weak staining occurs, and with excess sample it can be observed. The smudge in the first lane is probably excess sample.

Stereochemical modelling

Binding of poly(A) to poly(Q) is likely to have a structural basis that can be identified at the molecular level. Preliminary modelling with alanine di- and tripeptides failed to reveal any suitable peptide backbone polar group alignments with adenine able to explain the binding between poly(A) and poly(Q). Nevertheless, it was considered important to examine all aspects of the physical geometry of the peptides, and with this in mind, the orthodox peptide bond in the *trans* conformation was rotated to the *cis* form. Significantly, this conformation of the same peptide did show polar group alignments appropriate for H-bond formation between the dipeptide backbone and adenine.

In the *trans* conformation, successive side chains are positioned on opposite sides of the peptide backbone and are directed away from each other, similarly for the polar groups, which alternate between sides. In the *cis* conformation, successive side chains are positioned on one side of the backbone with the same directionality, while polar groups are positioned together in pairs, carbonyl and amine, on the opposite side.

This assembly of side chains to the one side of the backbone with a *cis* conformation presents a unique opportunity to explore side-chain–side-chain interactions not previously entertained with *trans* conformers. In particular, aromatic or polar side chains might be expected to interact intimately with one another[23] in a manner not sterically possible with the *trans* peptide.

Energy barriers for ω torsion

The negative gel results obtained for poly(A) exposure to alanine peptides is not surprising because conversion of the *trans* peptide into the energetically less favourable *cis* conformer is governed by the amide partial double bond with a rotational barrier of ~20 kcal mol^{-1}.[24] Equally the energy barrier from *cis* to *trans* is of the same order of magnitude at ~18 kcal mol^{-1}.[24] Adenine base pairing with uracil or thymine has been variously estimated at 12.8–18 kcal mol^{-1}.[25] A similar energy might be expected for adenine pairing with the *cis* peptide, but this is insufficient to overcome the ω rotational barrier and flip from *trans* to *cis*. Consequently, in the absence of suitable isomerases or other energy inputs, poly(A) would be restricted to binding (if at all) only to pre-existing *cis*-peptide chains. Possibly, this is the conformation adopted by poly(Q).

Adenine: *cis* construct

Because both the peptide group and adenine base are planar, it is possible to show a two-dimensional schematic of their interaction in some detail (Fig. 5*a*, *b*). The atomic coordinates cited are taken from Pauling[4] with minor refinements to adenine from Voet and Rich.[26]

On the left side of Fig. 5*a* is shown the peptide backbone in the *trans* conformation ($\omega = 180^\circ$). It appears to lack any suitable polar group pair in an appropriate position or orientation to engage with adenine, confirming preliminary findings with molecular models. On the right side (Fig. 5*b*) the last peptide bond has been rotated through 180° to the *cis* conformation ($\omega = 0^\circ$) This brings about several changes to the peptide. As mentioned above, the side chains are on the same side of the backbone, and the polar groups are positioned together on the opposite side now with similar directional orientation but diverging a little from each other.

These changes are relevant because they appear to allow possible H-bond interactions with adenine (Fig. 5*b*): the

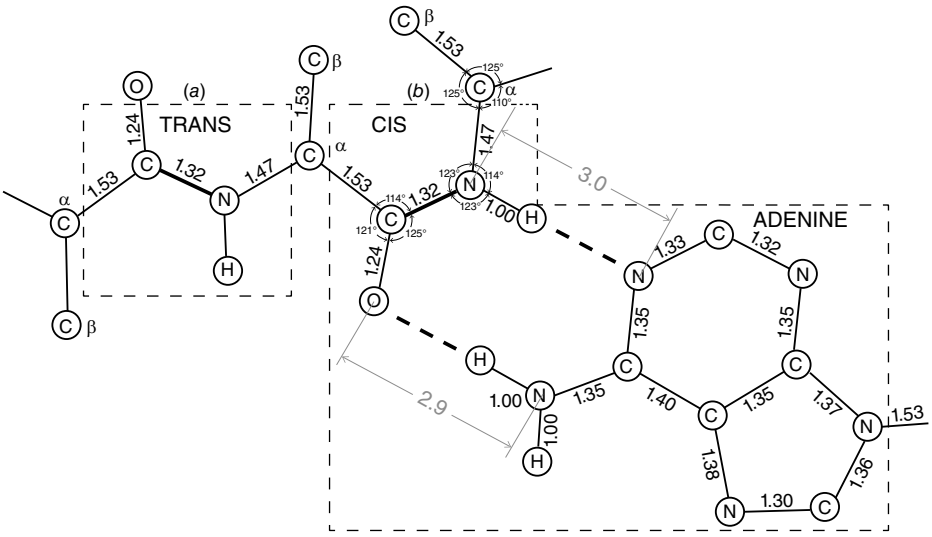

Fig. 5. Schematic showing peptide backbone polar group alignments with those of adenine for *trans* (*a*) and *cis* (*b*) peptides.

carbonyl oxygen with adenine N6 amine and the peptide amine with the adenine ring nitrogen N1. The bond to the oxygen targets one lone pair, situated to the side, seen also with a nucleic acid base pair. H-bond distances are set at ~2.9 Å, or a little less than the sum of the van der Waals radii, both H-bond angles appear to be close to 180° with donor, hydrogen and acceptor all in a straight line maximising the bond strength. Importantly, like the *trans* isomer, the *cis* peptide is planar, the base is planar and the H bonds unite the two giving a coplanar structure (Fig. 5*b*). The *cis* peptide appears to be complementary to adenine in the same fashion as is the base uracil. A comparison of the *cis* peptide and uracil suggests that the relevant bond lengths and angles are broadly similar and so they might well be interchangeable. The schematic suggests an adenine monomer can pair with and potentially bind the *cis* dipeptide but it doesn't indicate whether the adenine polymer can bind to an all-*cis* peptide polymer. This is an important consideration because successive polar groups along the two chains must be in allignment to generate a stable interaction between the two polymers.

Interaction of poly(Q) and poly(A)

A construct of an all-*cis* poly(Q) was assembled independently of poly(A). It was then used as a scaffold against which a poly(A) chain was positioned and polar group alignment could be assessed for potential bond formation as described above for the monomeric adenine.

Cis polyglutamine

Modelling of a hexapeptide of glutamine in the all-*cis* conformation revealed a helical structure, as previously predicted,[27] in which adjacent side chain carboxamide groups are close enough to permit H-bond formation. These multiple inter-amide side chain links are able to extend the full length of the peptide forming a second chain parallel to the *cis* backbone. The second chain appears to play a significant role in fixing and stabilising the geometry of the poly(Q) helix, particularly because it is able to exploit the cooperatively augmented H-bond strength (as outlined in the *Discussion*). Stereochemically, the construct indicates that the backbone of this *cis* poly(Q) structure forms a regular right hand helix ~6.4-Å diameter with a pitch of ~28 Å and 9–10 residues per turn. The omega torsion angle is 0°, phi and psi torsion angles are close to −150 and +150° respectively, placing them in the beta zone of a Ramachandran and Mitra plot.[19] The backbone polar groups NH and CO are paired and directed radially from the exterior of the helix, and the whole appears to be free from van der Waals contacts. The carboxamide H-bonded chain is ~10.5 Å in diameter and its added H-bond strength might play a role in constraining the backbone as the *cis* isomer, as well as setting other parameters of the helix. The rotamer angles positioning the donor and acceptor groups at 2.6–2.8 Å apart (implying bonding) are: chi 1 (Cα–Cβ) = +10°, chi 2 (Cβ–Cγ) = 180° and chi 3 (Cγ–Cδ) = +65°.

The complex

Single stranded poly(A) is part helical and part random coil,[28] with freedom of rotation around bonds of the sugar phosphate backbone and the glycosidic bond. This essential flexibility allows the helical configuration of poly (A) to pair its base polar groups with those of the *cis* peptide backbone recurrently and extending for the length of the two bound chains. As such, their respective orientation is: poly(Q) is NH → CO to poly(A) 5′ → 3′. The plane of the peptide bond in the poly(Q) helix is ~30° to the axis, which means for coplanarity with adenine, the base must roll to the same plane by rotation around the glycosidic bond, which is appropriately oriented with ribose almost perpendicular to

the base. The poly(A) helix is right handed and coaxial with poly(Q), which it surrounds. Its diameter is 23.5 Å with 9.5 residues per turn. The pitch is 28.0 Å with a residue rise of almost 3.0 Å. This resembles a single strand of the A form of DNA,[29] with a larger diameter and a shorter pitch than the B form, also a base inclined to the axis and a central void, which in the present proposal is occupied by the poly(Q) helix. It differs from the A form with a lower number of bases in a full turn (9.5 *v*. 11).

Fig. 6 shows a schematic projection of the complex normal to the common axis. The sugar phosphate backbone is represented by the heavy circle. Internal to this is the base, then the *cis* peptide, both represented in part as blank plates or plaques[30] to reduce the information and aid clarity of understanding the structure.[31] Molecular details are indicated in one quadrant. The position of the side chain carboxamides in an H-bonded train is shown, except for the quadrant with molecular detail, and the coplanar base and peptide group are represented on an oblique plane.

The stability of this complex appears to rely on the number of inter-carboxamide H bonds that can be formed, and these are set by the poly(Q) chain length. The ability to form the complex was tested with poly(Q8) in the binding mix instead of poly(Q15). The result, as seen in Fig. 7, indicates that eight mers is not able to bind poly(A).

Discussion

This investigation began as a search for a direct physicochemical function for the poly(A) tails of mRNA. The outcome is a suggestion that poly(A) can bind in a complementary manner to a backbone segment of any protein or peptide in an all-*cis* conformation, with torsion angle $\omega = 0$.

This is a somewhat radical departure from conventional thinking, particularly because *cis* peptide-containing proteins are extremely rare, ~0.05%,[30] and this refers to single peptides only, not repetitive all-*cis* tracts. This

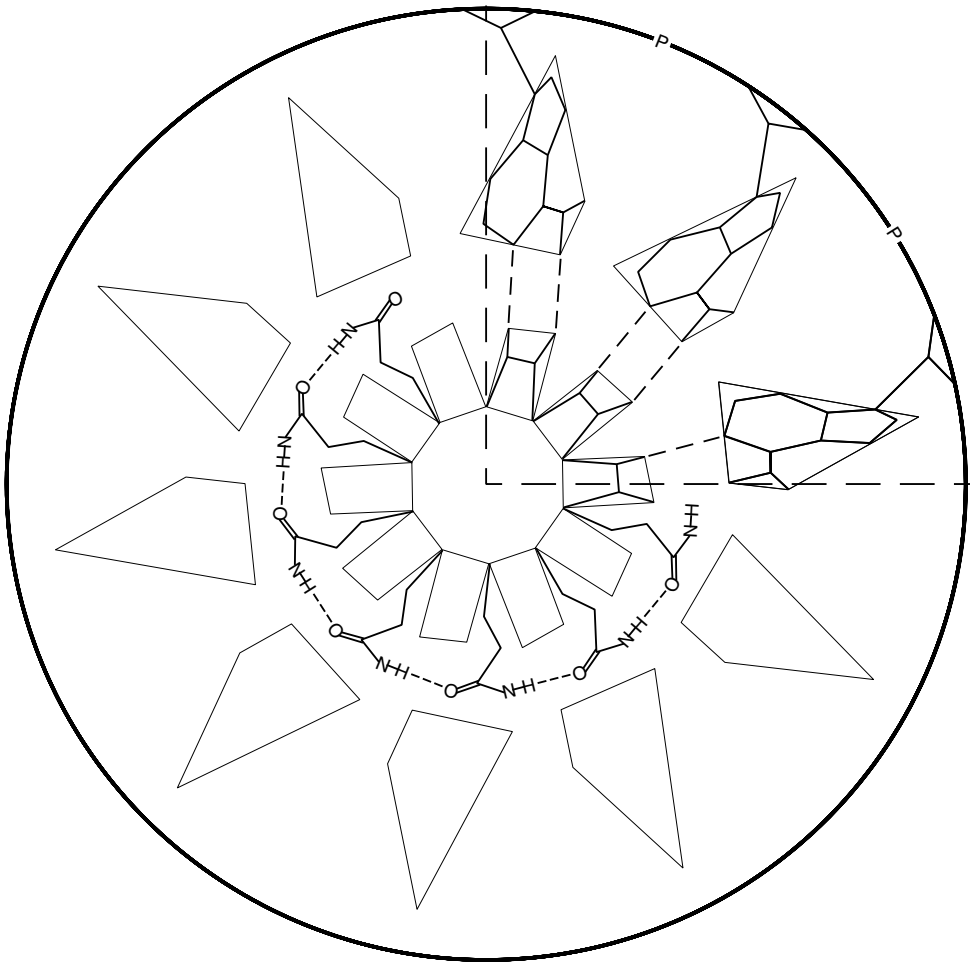

Fig. 6. Schematic showing relative coaxial positions and geometry of helical poly(Q10) encircled by helical poly(A10). The glutamine side chain carboxamide groups lie between the base and the peptide backbone. Potential H bonds between the two helices are indicated by broken lines. Viewing is down the axis.

scarcity, together with misconceptions about adjacent side chain clashes[24] (only possible if phi and psi torsion angles were not free to rotate), and the energetic cost of ~2 kcal mol^{-1} to stabilise the *cis* over the *trans* conformation, are the three main reasons[27] why the *cis* peptide conformation has been mostly considered irrelevant.

Rarity of *cis* backbone

The scarcity, however, may be more apparent than real; it has been suggested recently in a review of the Protein Data Bank using an unambiguous assessment method that almost 5000 *cis* peptide bonds have been incorrectly assigned as *trans*.[32] If poly(Q) tracts in proteins are in the default *cis* conformation but unrecognisable as such because they are within intrinsically disordered domains, or because of the aggregation prone nature of poly(Q), then a relatively large cohort of *cis* peptides may exist and be part of the above misassigned group.

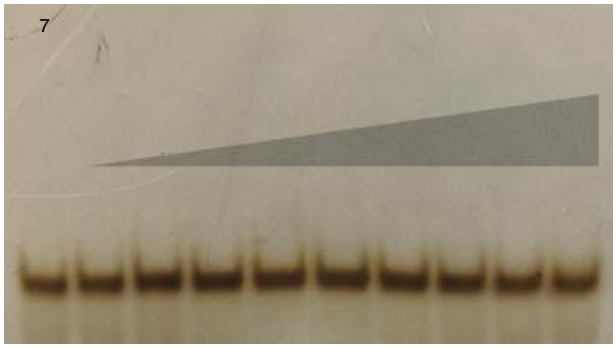

Fig. 7. Gel images of binding reaction between poly(Q8) and poly (A). There is no loss of band intensity for the low and high amounts of poly(Q8).

An important outcome of this study, in the context of a main-chain *cis* conformation of poly(Q), has been the probability of H bonding between carboxamide groups of adjacent side chains which may link them together over the length of the *cis* peptide. Side-chain–side-chain preferred bonding sets them apart from main-chain polar groups and would reduce the possibility of interaction between the two. In support of this notion are reports emphasising selective intrastrand side-chain–side-chain bonding.[33–35]

Cooperativity energy levels

H-bonded chains of formamide closely resemble those found in proteins (Fig. 8*a*). Importantly, formamide derived data show that linear extension of an H-bonded amide chain has a cooperative effect,[35] giving an energetic enhancement for each additional bond.

Depending on the chain length, the H-bond strength can increase by greater than 200%,[36] and in proteins this sets the stage for a selection process giving rise to inter-side chain amide H bonding in preference to bonding with multiple other polar groups in the vicinity. The energies of individual bonds in a chain depend on the chain length and position of the bond in the chain,[37] and, although they are encountered in natural proteins, the *cis* stereoisomer of poly(Q) does appear to enhance their formation. Experimental data for the decamer chain of formamide[36] show the central H-bond strength as 12.87 kcal mol^{-1}, which is ~2.5× the single H bond of 4.49 kcal mol^{-1} in the dimer.[37] Bond strength enhancement of this magnitude would be sufficient to selectively build polycarboxamide side chains. But its loss or considerable reduction might disrupt carboxamide chain formation and explain the lack of binding activity found for poly(Q8) seen in Fig. 7. But

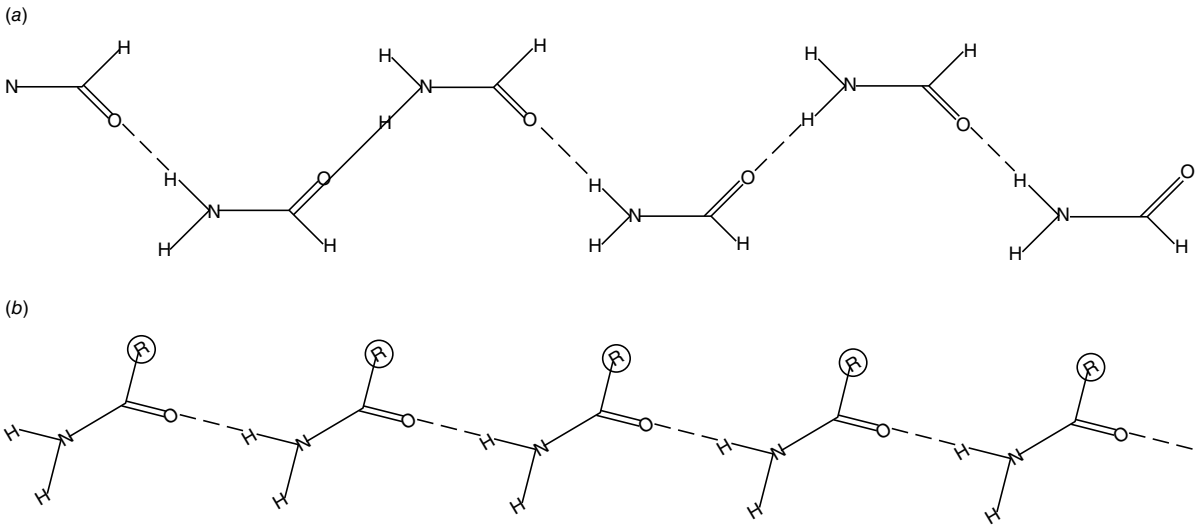

Fig. 8. Schematic representation of (*a*) formamide and (*b*) *cis* carboxamide of poly(Q) inter-side chain linkage. Part (*a*) is based on part extract of schematic, figure 1 in Kobko *et al.*[36]

even a 250% bond strength enhancement will be less than the 20 kcal mol^{-1} energy barrier needed to be overcome to flip the *trans* amide to *cis*. Nevertheless, the fixed and proximate position of the polar groups of poly(Q) (not available to freely diffusing formamide), and chains longer than 10 residues, could provide the added impetus to breach that energy barrier, suggesting that *cis* poly(Q) might well self assemble.

Protease sensitivity

Poly(Q) expansion disorders generate cellular aggregates composed of the host protein and other proteins along with the embedded poly(Q) tract. These components resist breakdown by both the ubiquitin proteasome system and the autophagy and lysosomal system[38]; however, the normal protein components are processed leaving behind poly(Q) tracts that resist degradation.[39] Proline peptides have a high proportion of bonds as the *cis* isomer, and this has served as a test system for the protease sensitivity of *cis* peptides. It has been reported that exopeptidases, prolidase and aminopeptidase P, are unable to hydrolyse the *cis* isomer but instead are specific for the *trans* isomer of the X-L-proline peptide bond. The *cis* form must isomerise to the *trans* before it can be hydrolysed.[40] Prospective studies need to determine if the *trans* peptide bond is a general requirement for proteolytic activity, and if so, it might explain why the poly(Q) peptide as a *cis* isomer is resistant to, if not unaffected by, proteolysis.[39]

Nascent peptide

Returning to the function of the poly(A) tail of mRNA *in vivo*, it is of course in close proximity to the ribosome with its protein synthesising machinery. All surrounding proteins on or off the ribosome would be established as secondary or tertiary structures based on a *trans* backbone. With respect to a *cis* backbone able to bind with the poly(A) tail, the only candidate likely to fill the role would be the unfolded nascent peptide as it exits the ribosome tunnel. Peptides are synthesised at the peptidyl transferase centre (PTC) of the ribosome by successive addition of aminoacyl tRNA residues to the c-terminus of the nascent chain. If each new aminoacyl tRNA enters the A site of the PTC with the same orientation, then newly synthesised peptide bonds at the P site should all be stereochemically identical, suggesting *de novo* origin of all *cis* bonds with no requirement for energy expenditure creating the alternating bonds of the *trans* isomer. Because of the 14 kcal mol^{-1} or more energy barrier for the *cis* into *trans* conversion, it is likely that the *cis* peptide isomer once formed in the PTC would be sufficiently stable to negotiate the full length of the exit tunnel. The all-*cis* peptide has extended backbone torsion angles placing it in the B zone of the Ramachandran and Mitra plot, suggesting that it can form either a B strand or, as with poly(Q), an extended helix (structurally not greatly dissimilar). There are many reports describing peptide secondary structure within the ribosome exit tunnel, mostly of compaction, thought to be probable alpha helices[41] but also of B strands.[42] It seems that helices can form when close to the PTC but change to B structures on movement towards the exit.[42] Moreover, it has been suggested that all nascent peptides, regardless of sequence, emerge from the tunnel with the same uniform B-strand like conformation,[43] which equates to extended backbone torsion angles resembling a *cis* conformation. The poly(A) tail is in the immediate vicinity of the lead ribosome (of a polysome) with its emerging (possibly *cis*) nascent peptide a possible binding partner for the poly(A) tail. Binding to the nascent peptide backbone would very likely affect the translation process – speeding, slowing or halting it. In addition, it might assist in the process of peptide folding leading eventually to a *trans* conformation and disengagement from the poly(A) tail, in turn voiding its effect on the translation process.

Transcription factors (TFs)

Apart from the mRNA tail, other poly(A) and poly(dA) sites exist within the cell, particularly in the genetic regulatory regions of promoter and enhancer of most organisms at the 5′ end of the gene.[44] These are short, usually a run of 4–6 bases[45] and flanked by alternating thymine residues as seen in the native TATA box sequence, which is one of several distinct elements of the core promoter responsible for assembling the transcription complex with RNA polymerase II. An oligomer TATATAAAA is known to adopt the DNA A form structure,[29] which interestingly is similar to the conformation of the poly(A) of the proposed complex with poly(Q). Promoters are targeted by TFs which, on binding, activate the transcription process. TFs appear to consist of a DNA binding domain (DBD) and one or more intrinsically disordered regions (IDRs).[46] Many TFs contain poly(Q) tracts[47] that appear to be part of the IDRs and have been shown to stimulate transcription.[48] The 10 known poly(Q) neurodegenerative disorders all have TFs carrying expansions of their normal poly(Q) complement. However, normal length poly(Q) tracts in a TF may be suitably positioned to bind with the multiple adenine residues present in promoters.

The finding in this study that poly(Q) is able to bind with poly(dA) when formed in a double helix, by displacing poly(dT), suggests this as a possible mechanism TFs might use to open a double helical section of the promoter. But why would poly(A) bind to poly(Q) in preference to binding with poly(T)? Duplex oligonucleotides are prone to separate because of the mutual repulsion of their backbone charged phosphate groups, often able to be masked with salt solutions. In the case of poly(Q), which is electrically neutral, a complex with poly(A) would be devoid of the repulsion

element and so would bind more strongly than would the two chains of a double helix.

Conclusion

Gel electrophoresis indicates poly(A) binds to poly(Q). Stereochemical modelling suggests poly(A) is H bonding to the all-*cis* conformation of the poly(Q) peptide backbone, possibly by formation of a coaxial helix encircling the smaller helical conformation of *cis* poly(Q). The complex appears to be particularly integral to the mechanisms of translation and transcription. Non-poly(Q) peptides, if in the all-*cis* conformation, would be expected to also bind with poly(A); nascent peptides being a possible target group, but there may even be others – concealed among the many reported misassigned *trans* peptides.

Poly(Q) peptides bind to one another with an unusual tenacity, as exemplified by the irreversible nature of aggregations seen with pathological expansions of poly(Q), this is possibly assisted by the concentration of polar groups to one side of a *cis* backbone. Future studies into the actions of poly(A) might be of benefit for those strategies being developed to oppose intra-cellular aggregate formation or even to assist with its breakdown.

References

[1] Passmore LA, Coller J. Roles of mRNA poly (A) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol* 2022; 23: 93–106. doi:10.1038/s41580-021-00417-y

[2] Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res* 2002; 30: 1427–1464. doi:10.1093/nar/30.7.1427

[3] Schopf JW. Fossil evidence of archaean life. *Phil Trans R Soc B Biol Sci* 2006; 361: 869–885. doi:10.1098/rstb.2006.1834

[4] Pauling L. *The Nature of the Chemical Bond*, 3rd edn. Cornell University Press; 1960.

[5] Chavali S, Singh AK, Santhanam B, Babu MM. Amino acid homorepeats in proteins. *Nat Rev Chem* 2020; 4: 420–434. doi:10.1038/s41570-020-0204-1

[6] Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* 1984; 12: 4127–4138. doi:10.1093/nar/12.10.4127

[7] Mularoni L, Ledda A, Toll-Riera M, Albà MM. Natural selection drives the accumulation of amino acid tandem repeats in human proteins. *Genome Res* 2010; 20: 745–754. doi:10.1101/gr.101261.109

[8] Darling AL, Uversky VN. Intrinsic disorder in proteins with pathogenic repeat expansions. *Molecules* 2017; 22: 2027. doi:10.3390/molecules22122027

[9] Amato M, Franco R, Facchini G, Addeo R, Ciardiello F, Berretta M, Vita G, Sgambato A, Pignata S, Caraglia M, Accardo M, Zito Marino F. Microsatellite instability: from the implementation of the detection to a prognostic and predictive role in cancers. *Int J Mol Sci* 2022; 23: 8726. doi:10.3390/ijms23158726

[10] Adegbuyiro A, Sedighi F, Pilkington IV AW, Groover S, Legleiter J. Proteins containing expanded polyglutamine tracts and neurodegenerative disease. *Biochemistry* 2017; 56: 1199–1217. doi:10.1021/acs.biochem.6b00936

[11] Perutz MF, Johnson T, Suzuki M, Finch JT. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Nat Acad Sci USA* 1994; 91: 5355–5358. doi:10.1073/pnas.91.12.5355

[12] Boivin M, Charlet-Berguerand N. Trinucleotide CGG repeat diseases: an expanding field of polyglycine proteins. *Front Genet* 2022; 13: 843014. doi:10.3389/fgene.2022.843014

[13] Wen J, Scoles DR, Facelli JC. Molecular dynamics analysis of the aggregation propensity of polyglutamine segments. *PLoS One* 2017; 12: e0178333. doi:10.1371/journal.pone.0178333

[14] Wetzel R. Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence. *J Mol Biol* 2012; 421: 466–490. doi:10.1016/j.jmb.2012.01.030

[15] Oma Y, Kino Y, Toriumi K, Sasagawa N, Ishiura S. Interactions between homopolymeric amino acids (HPAAs). *Protein Sci* 2007; 16: 2195–2204. doi:10.1110/ps.072955307

[16] Li L, Ng NKL, Koon AC, Chan HYE. Expanded polyalanine tracts function as nuclear export signals and promote protein mislocalization via eEF1A1factor. *J Biol Chem* 2017; 292: 5784–5800. doi:10.1074/jbc.M116.763599

[17] Boivin M, Deng J, Pfister V, Grandgirard E, Oulad-Abdelghani M, Morlet B, Ruffenach F, Negroni L, Koebel P, Jacob H, Riet F, Dijkstra AA, McFadden K, Clayton WA, Hong D, Miyahara H, Iwasaki Y, Sone J, Wang Z, Charlet-Berguerand N. Translation of GGC repeat expansions into a toxic polyglycine protein in NIID defines a novel class of human genetic disorders: the polyG diseases. *Neuron* 2021; 109: 1825–1835.e5. doi:10.1016/j.neuron.2021.03.038

[18] Liufu T, Zheng Y, Yu J, Yuan Y, Wang Z, Deng J, Hong D. The polyG diseases: a new disease entity. *Acta Neuropathol Commun* 2022; 10: 79. doi:10.1186/s40478-022-01383-y

[19] Ramachandran GN, Mitra AK. An explanation for the rare occurrence of cis peptide units in proteins and polypeptides. *J Mol Biol* 1976; 107: 85–92. doi:10.1016/S0022-2836(76)80019-8

[20] Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat Protoc* 2007; 2: 1849–1861. doi:10.1038/nprot.2007.249

[21] Tang W, Zhou H, Li W. Silver and cyanine staining of oligonucleotides in polyacrylamide gel. *PLoS One* 2015; 10: e0144422. doi:10.1371/journal.pone.0144422

[22] Seo M, Lei L, Egli M. Label-free electrophoretic mobility shift assay (EMSA) for measuring dissociation constants of protein–RNA complexes. *Curr Protoc Nucl Acid Chem* 2019; 76: e70. doi:10.1002/cpnc.70

[23] Jabs A, Weiss MS, Hilgenfeld R. Non-proline *cis* peptide bonds in proteins. *J Mol Biol* 1999; 286: 291–304. doi:10.1006/jmbi.1998.2459

[24] Mathieu S, Poteau R, Trinquier G. Estimating the “steric clash” at *cis* peptide bonds. *J Phys Chem B* 2008; 112: 7894–7902. doi:10.1021/jp711082d

[25] Jurecka P, Hobza P. True Stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine···cytosine, adenine···thymine, and their 9- and 1-methyl derivatives: complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. *J Am Chem Soc* 2003; 125: 15608–15613. doi:10.1021/ja036611j

[26] Voet D, Rich A. The crystal structures of purines, pyrimidines and their intermolecular complexes. *Prog Nucleic Acid Res Mol Biol* 1970; 10: 183–265. doi:10.1016/S0079-6603(08)60565-6

[27] Poteau R, Trinquier G. All-*cis* helical polypeptides. *J Org Chem* 2007; 72: 8251–8258. doi:10.1021/jo0711790

[28] Saenger W, Riecke J, Suck D. A structural model for the polyadenylic acid single helix. *J Mol Biol* 1975; 93: 529–534. doi:10.1016/0022-2836(75)90244-2

[29] Dickerson RE, Ng H-L. DNA structure from A to B. *Proc Natl Acad Sci USA* 2001; 98: 6986–6988. doi:10.1073/pnas.141238898

[30] Poteau R, Trinquier G. All-*cis* cyclic peptides. *J Am Chem Soc* 2005; 127: 13875–13889. doi:10.1021/ja052342g

[31] Perutz MF, Finch JT, Berriman J, Lesk A. Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci USA* 2002; 99: 5591–5595. doi:10.1073/pnas.042681399

[32] Touw WG, Joosten RP, Vriend G. Detection of *trans–cis* flips and peptide–plane flips in protein structures. *Acta Crystallogr D Biol Crystallogr* 2015; 71: 1604–1614. doi:10.1107/S1399004715008263

[33] Ellisdon AM, Thomas B, Bottomley SP. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J Biol Chem* 2006; 281: 16888–16896. doi:10.1074/jbc.M601470200

[34] Marchut AJ, Hall CK. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. *Biophys J* 2006; 90: 4574–4584. doi:10.1529/biophysj.105.079269
[35] Plumley JA, Dannenberg JJ. The importance of hydrogen bonding between the glutamine side chains to the formation of amyloid VQIVYK parallel β-sheets: an ONIOM DFT/AM1 study. *J Am Chem Soc* 2010; 132: 1758–1759. doi:10.1021/ja909690a
[36] Kobko N, Paraskevas L, del Rio E, Dannenberg JJ. Cooperativity in amide hydrogen bonding chains: implications for protein-folding models. *J Am Chem Soc* 2001; 123: 4348–4349. doi:10.1021/ja0042711
[37] Kobko N, Dannenberg JJ. Cooperativity in amide hydrogen bonding chains. Relation between energy, position, and H-bond chain length in peptide and protein folding models. *J Phys Chem A* 2003; 107: 10389–10395. doi:10.1021/jp0365209
[38] Venkatraman P, Wetzel R, Tanaka M, Nukina N, Goldberg AL. Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. *Mol Cell* 2004; 14: 95–104. doi:10.1016/S1097-2765(04)00151-0
[39] Nijholt DAT, De Kimpe L, Elfrink HL, Hoozemns JJM, Scheper W. Removing protein aggregates: the role of proteolysis in neurodegeneration. *Curr Med Chem* 2011; 18: 2459–2476. doi:10.2174/092986711795843236
[40] Lin L-N, Brandts JF. Role of *cis–trans* isomerism of the peptide bond in protease specificity: kinetic studies on small proline-containing peptides and on polyproline. *Biochemistry* 1979; 18: 5037–5042. doi:10.1021/bi00590a002
[41] Zhang Y, Wölfle T, Rospert S. Interaction of nascent chains with the ribosomal tunnel proteins Rpl4, Rpl17, and Rpl39 of *Saccharomyces cerevisiae*. *J Biol Chem* 2013; 288: 33697–33707. doi:10.1074/jbc.M113.508283
[42] Komar AA. From alpha to beta – a co-translational way to fold? *Cell Cycle* 2022; 21: 1663–1666. doi:10.1080/15384101.2022.2062186
[43] Syroegin EA, Aleksandrova EV, Polikanov YS. Insights into the ribosome function from the structures of non-arrested ribosome–nascent chain complexes. *Nat Chem* 2023; 15: 143–153. doi:10.1038/s41557-022-01073-1
[44] Orsi RH, Bowen BM, Wiedmann M. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. *BMC Genomics* 2010; 11: 102. doi:10.1186/1471-2164-11-102
[45] Haran TE, Mohanty U. The unique structure of A-tracts and intrinsic DNA bending. *Q Rev Biophys* 2009; 42: 41–81. doi:10.1017/S0033583509004752
[46] Brodsky S, Jana T, Barkai N. Order through disorder: the role of intrinsically disordered regions in transcription factor binding specificity. *Curr Opin Struct Biol* 2021; 71: 110–115. doi:10.1016/j.sbi.2021.06.011
[47] Gemayel R, Chavali S, Pougach K, Legendre M, Zhu B, Boeynaems S, van der Zande E, Gevaert K, Rousseau F, Schymkowitz J, Babu MM, Verstrepen KJ. Variable glutamine-rich repeats modulate transcription factor activity. *Mol Cell* 2015; 59: 615–627. doi:10.1016/j.molcel.2015.07.003
[48] Atanesyan L, Günther V, Dichtl B, Georgiev O, Schaffner W. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biol Chem* 2012; 393: 63–70. doi:10.1515/BC-2011-252

Data availability. Data used for generating results in this paper are available from A. B. Bransgrove.

Conflicts of interest. The authors declare that they have no conflicts of interest.

Declaration of funding. This work was supported by funding from National Diagnostic Products (Aus.).

Acknowledgements. The authors thank Prof. Richard Christopherson and Dr Leo Phillips for their comments and discussion and Dr Nicholas Proschugo for Mass Spectrometry control poly(Q) status.

Author contributions. The authors contributed equally to all aspects of this submission.

Author affiliations
[A]National Diagnostic Products, The SAND Institute, 7–9 Merriwa Street, Gordon, NSW 2072, Australia.
[B]School of Molecular Bioscience, University of Sydney, Sydney, NSW 2006, Australia.
[C]Present address: Neutral Bay Medical Centre, Sydney, NSW 2089, Australia. Email: lbra1246@gmail.com