Molecular Neurobiology (2019) 56:3664–3675
https://doi.org/10.1007/s12035-018-1269-4

Polyglutamine Repeats in Viruses

Catherine H. Schein [1]

Received: 15 March 2018 / Accepted: 19 July 2018 / Published online: 4 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

This review explores the presence and functions of polyglutamine (polyQ) in viral proteins. In mammals, mutations in polyQ segments (and CAG repeats at the nucleotide level) have been linked to neural disorders and ataxias. PolyQ regions in normal human proteins have documented functional roles, in transcription factors and, more recently, in regulating autophagy. Despite the high frequency of polyQ repeats in eukaryotic genomes, little attention has been given to the presence or possible role of polyQ sequences in virus genomes. A survey described here revealed that polyQ repeats occur rarely in RNA viruses, suggesting that they have detrimental effects on virus replication at the nucleotide or protein level. However, there have been sporadic reports of polyQ segments in potyviruses and in reptilian nidoviruses (among the largest RNA viruses known). Conserved polyQ segments are found in the regulatory control proteins of many DNA viruses. Variable length polyQ tracts are found in proteins that contribute to transmissibility (cowpox A-type inclusion protein (ATI)) and control of latency (herpes viruses). New longer-read sequencing methods, using original biological samples, should reveal more details on the presence and functional role of polyQ in viruses, as well as the nucleotide regions that encode them. Given the known toxic effects of polyQ repeats, the role of these segments in neurovirulent and tumorigenic viruses should be further explored.

Keywords Neurotropic viruses · Glutamine repeat diseases · A-type inclusion protein · Deoxyuridine 5′-triphosphate nucleotide hydrolase (DUT) · Herpes virus latency · Cowpox virus · RNA viruses · Virus transmissibility · Protein inclusions containing virus · Beclin-1 control of autophagy · Kaposi's sarcoma

Introduction

Mutations in human proteins that result in longer polyQ repeat sequences have been linked to dementias and ataxias [1]. Their toxicity has been attributed, at the protein level, to aggregation of long polyQ protein tracts, interference with autophagy [2] and to their ability to bind RNA in several model organisms, including marmosets [3], *Drosophila*, and *E. coli* [4–7]. Proteins containing mutated longer repeats may also lose their function. For example, expanded polyQ repeats in ataxin-3 may interfere with miRNA function in Machado-Josef disease [8] and expansion of the polyQ tract in the androgen receptor reduces its DNA binding capacity [9]. The mechanisms of polyQ toxicity are dependent on the protein encoded, and even alternative reading frames of the DNA [10, 11]. For example, aggregated polyQ containing protein, huntingtin, is found in the brain of Huntington's disease (HD) victims. However, a rare disease similar in clinical appearance to HD, Huntington's disease-like 2 (HDL2), has been linked to repeat regions in RNA and alternative transcripts causing loss of expression of other proteins, such as junctophilin-3 [12]. Targeting such repeats at the protein or RNA level may provide novel therapies for these diseases [13–15].

While the mechanisms for the function and toxicity of extended polyQ segments (or the nucleic regions that encode them) in eukaryotic proteins continue to be actively studied [16], there has been little exploration of their occurrence and possible roles, even in neurovirulent viruses. This is particularly curious, in light of the documented role of polyQ tracts in transcription factors (TFs) and their abundance in eukaryotic genomes [17], even constituting a "polyQ interactome" [16]. The first goal of this work was to determine whether viral proteins contain polyQ repeats at all. One might anticipate that longer polyQ sequences, based on their tendency to aggregate

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12035-018-1269-4) contains supplementary material, which is available to authorized users.

✉ Catherine H. Schein
chschein@utmb.edu

[1] Department of Biochemistry and Molecular Biology, Institute for Human Infection and Immunity, University of Texas Medical Branch, Galveston, TX, USA

and to interfere with transcription, would be selected against in rapidly replicating viruses under extreme environmental pressure. Sequence selection in DNA viruses during chronic infections, on the other hand, would favor compatibility with host cell transcription and translation mechanisms and immune evasion [18, 19], rather than rapid growth [20–23]. As this study has found, polyQ segments are indeed rare in the catalogued sequences of smaller RNA viruses, but even very long repeats have been found in several large RNA and DNA viruses.

The second goal is to suggest what functions, if any, such repeat sequences, at the protein or nucleotide level, could play in viral replication, chronic infection, or neuro-pathogenesis. Clues for the potential role of the repeats could be gleaned from their roles in eukaryotic proteins, where they are present in many transcription factors. In addition, Q-rich repeats in the N-terminus of the Argonaute-2 protein of *Drosophila* and other insects [24, 25] are essential for antiviral activity [26] and one in a cellular protein, TLE2, contributes to this protein's ability to control lytic reactivation of Kaposi's sarcoma-associated herpesvirus [27]. As discussed below, the polyQ segments found in several viral proteins could indeed affect the ability of viruses to control the activities or transcription of their own or cellular proteins, while their possible role in neurovirulence remains to be established.

While it may be surprising that polyQ sequences in neurovirulent viruses have not been a major topic of study, it should be emphasized that the extent of very long repeat segments would be difficult to detect by short-read sequencing of the large viruses in which they have been found. In addition, CAG triplet repeats are known to be unstable [28] and may be specifically excised during the transition from latency to active growth, or after adaptation to cell culture. Newer methods, designed to specifically determine repeat sequences in direct isolates from infected tissues, should reveal more details about the presence and roles of repeated sequences.

Section 1: Long PolyQ Segments in Larger RNA and DNA Viruses

Searching for polyQ tracts in viruses Searching the published sequences of many different virus families revealed that while they are not present in smaller RNA viruses, surprisingly long tracts of polyQ have been found in larger RNA and DNA viruses. The search also suggested that repeats may be much more common in viruses than is indicated by the currently archived sequences.

At the start of this work, the ViPR database [29], which allows rapid access to the published sequences of over 75,000 viral genomes or genome segments, was used to determine which RNA and DNA viruses contain polyQ repeats. A new resource, the Influenza research database [30], was used to screen influenza virus sequences. Once Q-rich sequences were identified, BLAST searches starting from the viral proteins that contained them were used to determine the extent of their conservation in the same virus family and to find other virus proteins containing similar tracts. BLAST was also used to find viral proteins containing repeats similar to those of the Argonaut-2 proteins.

The Vast Majority of Published RNA Virus Genomes Contain No Extended PolyQ Repeats If long polyQ repeats are intrinsically toxic for the function of proteins, or stimulate aggregation [31], one would expect that rapid evolving RNA viruses would selectively eliminate them. Table 1 summarizes searches of over 20,000 genome sequences of many families of pathogenic viruses, including single-strand RNA viruses (Flavivirus, Reoviruses, Picornaviruses, Bunyaviridae, etc.), 43,000 segments of the dsRNA Reoviridae from the ViPR database and over 100,000 strains of Influenza from the Influenza research database [30]. This revealed that only a few RNA viruses contain even a QQQQ sequence. Longer polyQ sequences, which would be anticipated to cause aggregation of the viral proteins (or, as discussed later, interfere with autophagy), were not found. As long repeats of many other amino acids (especially D, T, L, E, P) and mixed basic or acidic residues occur very frequently, this would suggest that there is some selection against longer polyQ tracts, either at the RNA or protein level. Literature searches have revealed sporadic reports of polyQ segments in some small RNA viruses, including potyviruses [32, 33], and even a coxsackie A24 isolate [34], whereby the lack of consistency among closely related viruses suggests these have no functional role.

More meaningfully, BLAST searches beginning with a Q-rich sequence from DNA viruses (see below) identified a polyQ sequence in the first open reading frame of a nidovirus isolated from a python, representing a novel genus of Torovirus [35]. A similar polyQ sequence is also found in the ORF1 of the *Morelia viridis* (*Boa constrictor*) nidovirus, but not in that of a nidovirus isolated from lizards [36]. Nidoviruses (which include the Coronaviridae) have the longest known RNA virus genome, with continuous positive sense strands of 26–32 kBases [37]. In contrast, other +-strand RNA viruses range from 7.5 to 12 kb, and negative-strand RNA viruses have genome lengths ranging from 7 to 19 kb. Bunyavirdae can be up to 22.7 kb in total length, but their longest (L) segments do not exceed 12 kb.

It is possible that these long polyQ insertions may play a role, at the RNA level, during genome replication or adapting to changing environments [38]. As for other RNA virus families [39, 40], several studies have indicated the importance of dsRNA folding domains near the 5′ end of coronavirus genomes [41]. Formation of dsRNA intermediates [42], important for the interferon response [43–46], as well as viral enzymes that interfere with the OAS/RNaseL system that would

Table 1 Maximum length of polyQ repeats (Qn) found in published genomes of mammalian RNA and DNA virus groups. The second column shows the number of genomes searched for each group of viruses, and the last column lists some of the proteins that contain the longer polyQ repeats. See Table 2 for examples of herpes proteins with polyQ repeats and Fig. 1 for longer repeats

Group	Genomes	Q_n	Found in
+-strand RNA			
Coronavirus	1727	4	GKGQQQQGQ is conserved in the nucleocapsid of Bat corona virus and SARS
Flavivirus	10,242	4	Hepacivirus NS4B (22 total)
Caliciviridae	1215	4	Norwalk p22, (15 total)
Hepeviridae	316	3	Hepatitis E (13 total)
Picornaviridae	3704	4	Sapelovirus; human parechovirus (17 total)
Togaviridae	1342	3	Alphaviruses including VEEV, Ross River, Sindbis, Semliki Forest and Aura (542 total)
-strand RNA			
Arenaviridae	961	4	Sabia virus nucleocapsid (3 total)
Bunyaviridae	6273	4	Brazoran nucleocapsid, Enseada polymerase, Southbay virus L Protein (6 total)
Filoviridae	497	3	Ebolaviruses (several times; 1726 total)
Paramyxoviridae	2590	4	Mumps, Tuhoko, Newcastle, Avian and Bat paramyxovirus nucleocapsid (106 total)
Influenza	107,759	3	Many virus proteins
Rhabdoviridae	1136	4	Rice yellow stunt virus nucleocapsid (1 result)
dsRNA			
Reoviridae	43,913 segments	4	In 16 sequences: Rotavirus NSP3, orthoreovirus cell attachment factor sigma 1, Cypovirus VP4, Eyach VP8
DNA viruses:			
Herpesviridae	796	33	Tupaiid T2; RF1 of Rhadinovirus type 1 (Fig. 1)
Poxviridae	391	22	Cowpox virus A-type inclusion body protein (ATI) (Table 3)

target these [47, 48], are characteristic of infection by several different nidoviruses. These include the coronaviruses, MERS and SARS. The repeat RNA sequence encoding the polyQ region in the 5′ region of the python virus may fold as an independent domain containing extended segments of dsRNA (Fig. S1, [49, 50]), whereby the low energy of folding generates a dynamic, unstable structure [51]. The CAG repeat region might be excised completely during rapid growth, as CAG repeats are known to be unstable [28, 52, 53]. Alternatively, it may be transcribed past during generation of subgenomic RNAs, which in nidoviruses proceeds by selective transcription of parts of the open reading frames [54].

PolyQ Repeats in DNA Viruses Searches within two diverse and well-studied DNA virus families, poxviridae and herpes, indicated that several of these large viruses, known to cause chronic neurotropic infections, contain long polyQ segments (Tables 1, 2 and Fig. 1). These DNA virus genomes are 145–200 kbp, 5–10 times longer than those of the RNA viruses. Many herpes virus proteins contain variable length polyQ repeats in conserved regions (Table 2 shows some examples), and even longer polyQ repeats have also been found (Fig. 1). In addition to direct polyQ repeats, there are long, Q-rich repeats in other viral proteins. For example, there is a long, variable length, Q-rich repeat in the MC006L protein of the pox virus, *Molluscum contagiosum* [55]. This virus causes the formation of wart-like blisters on the skin of infected individuals, and characteristic cellular inclusions.

As discussed below, the longest repeats were found in DNA virus proteins that function in enhancing transmissibility (cowpox ATI) or contribute to viral latency (herpes viruses).

Section 2: Exploring the Function of Glutamine Repeats in Viral Proteins

The RNA virus results, coupled with the fact that polyQ expansions in human proteins can lead to disease, suggest that polyQ segments are probably selected against in rapidly growing viruses. This leads to the question: what possible functions could they serve for the virus itself or interaction with host cells? This is an important question to answer as the repeats occur in proteins from viruses triggering hard-to-treat neuropathies and epilepsy [56] and isolated from latently infected tissues and tumors.

PolyQ Repeats Serve Important Functions in Mammalian Proteins Although studied for their role in disease, polyQ segments in mammalian proteins have important regulatory

functions. After a polyQ repeat was found to be an activation domain [57] of the TF, SP1, the role of such segments in TFs was extensively studied [58] long before they were documented to have pathogenic characteristics possibly related to folding and formation of aggregates in cells (e.g., [59]). More recently, the length of polyQ inserts was directly related to their ability to enhance TF activity [9, 38]. Variable length Q-rich repeats may also modulate TF activity in eukaryotic cells by modulating their solubility [60] or by recruiting other factors to the DNA binding complex. A polyQ repeat in murine SRY (sex determining region on the Y chromosome) both stabilizes the protein and serves as a transactivation domain [61]. However, the region is found only in rodent SRY and can be replaced by an irrelevant protein (mCherry).

Consistent with a possible role for polyQ tracts in viral proteins in controlling transcription, the first report of a polyQ tract in a DNA virus was in a baculovirus regulatory protein, where the authors noted the similarity of the amino acid repeats to those in SP1 [62]. As Table 2 illustrates, polyQ tracts are present in several regulatory proteins of herpes viruses.

Further evidence for a functional roles in controlling virus replication is that Q-rich tetratricopeptide repeats are upregulated during bovine leukemia virus infection [63], as well as in human breast cancer cells [64]. The Q-rich N-terminal region of a cellular protein, transducing inhibitor of SPLIT (TLE2), contributes to this protein's ability to control lytic reactivation of Kaposi's sarcoma-associated herpesvirus [27].

PolyQ Regions in Viral Proteins May Mediate Neurovirulence Through Interference with Autophagy Recent reports suggest that polyQ segments may also serve to downregulate autophagy, which serves as a barrier to the growth of neurovirulent herpes viruses (whereby RNA viruses may use the membranous structures characteristic of autophagy for their own replication). These examples suggest possible roles for the longer repeats in proteins of viruses that typically cause latent infections, including herpes simplex, Epstein Barr, β- and ɣ-herpes viruses (Fig. 1). Mutations in beclin-1, a protein which triggers the process, were previously linked to development of neurodegenerative diseases [65]. Neurovirulent herpes simplex virus produces a protein that specifically binds to and interferes with beclin-1 function [66], called neurovirulence factor ICP34.5 (or gamma1 34.5, ɣ34.5).

Figure 2 (based on [2, 67]) shows how an expansion of the polyQ repeat in mutant ataxin-3, as well as excess polyQ from other cellular (or viral) proteins, could interfere with the interaction of ataxin-3 and beclin-1 to inhibit autophagy. The

Table 2 Examples of herpes proteins containing polyQ repeat segments

Virus	Protein	Residues	Sequence
Human herpesvirus 5	Multifunctional expression regulator	703–713	QQQQQQQQQQQ
Human herpesvirus 5	Protein UL133	247–257	QQQQQQQHQTG
Human herpesvirus 5	Tegument protein pp150	399–409	RQQNLQQRQQQ
Elephant endotheliotropic herpesvirus 4	Protein ORF-S	316–326	QQQQQQQQQQQ
Elephant endotheliotropic herpesvirus 4	Protein U59	74–84	QQQQQQQQQRQ
Tupaiid herpesvirus 1 2	T2 (see also Fig. 1)	496–506	QQQQQQQQQQQ
Murid herpesvirus 1 C4A	m18	60–70	QQQQQQQQQQE
Murid herpesvirus 1 C4A	M25	335–345	QRQQQQQQQQQ
Murid herpesvirus 1 C4A	M34	176–186	REQQHQQQQQG
Murid herpesvirus 1 K181	Apoptosis inhibitor	112–122	QQQQEKQQQQQ
Equid herpesvirus 2 86/67	Capsid maturation protease	606–616	QPQQQQQPQQQ
Equid herpesvirus 2 86/67	Capsid scaffold protein	299–309	QPQQQQQPQQQ
Equid herpesvirus 5 2–141/67	DNA packaging protein UL32	248–258	KQQQGQGQRQQ
Equid herpesvirus 5 2–141/67	DNA packaging tegument protein UL25	415–425	KQQQSQQQQQS
Equid herpesvirus 5 2–141/67	Uracil-DNA glycosylase (UDG)	12–22	QQQQQQPQDDQ
Equid herpesvirus 5 2–141/67	Envelope glycoprotein B	789–799	QQQQQQQQQQQ
Equid herpesvirus 5	Glycoprotein B	790–800	QQQQQQQQQQQ
Suid alphaherpesvirus 1	VP1/2	2258–2268	QQQQQQQQQRQ
Suid herpesvirus 1	Protein V57	106–116	QQQQQQQQQQR
Suid alphaherpesvirus 1	ICP27	62–72	QRQQQQQRQQQ
Suid herpesvirus 1	Early regulation protein UL54	64–74	QRQQQQQQRQQ
Suid herpesvirus 1	UL3.5	106–116	QQQQQQQQQQR

Fig. 1 Extensive polyQ repeats and Q-rich (underlined) regions are present in several different herpes and pox virus proteins

Tupaiid herpesvirus 1 protein T2:
RRRRRQRRSSSSRSSRRRPLLRPPSPDLPQAPPRPRR[**Q33**]PPPPQKQ
QPRPPPLPSRPSEEPSEEPSEEPSEDSPPILSSSPIQPVPVPIPPPPPP
PPPAFHD

Retroperitoneal fibromatosis-associated herpesvirus protein RF1:
LSMIACCVYKWLMMRQQQQQQQQPQQQPQQQQQHHQQQHHQQQQHHQQQ
QHHQQQQHHQQQQRQQQQQQHHHQQQQQQQP[**Q17**]LPDYMPLLDRQPS
LYNIASS

Equid herpesvirus 5:2-141/67 Name:envelope glycoprotein B:
LIVGGIIVLYLFITRSRTVYQAPIRMLYPEVDRAPQQNVQPIPEDQVRS
ILLAMHQFQQQQQQQQQQQQEEHTQRRSIFDTIRESTSNILRRRRGGGG
YTRLRQR

Human herpesvirus 5/BE/33/2011 protein UL69:
PPSPPAPLAGVRSHRGELNLMTPSPSHGGSPPQVPHKQPIIPVQSANGN
HSTTATQQQQQQQQQQQQQQPPPPPPVPQEDDSVVMRCQTPDYEDMLCY
SDDMDD

Molluscum contagiosum virus Protein MC006L:
PTQQMQLQQHLRQLQQQLQFPPPYPSQLQLQLPHAAQAPAQAPPRRAQR
PQGPQATKAQTGKARAPTTKPQSAKVRAQAAKEQAEARAETRATRVRAA
KAREMQSQAQTRTQV<u>QAQAQQAQAQAQQAQAQQAQAQQAQAQQAQAQQA</u>
<u>QAQQARAQQAQAQQAQQAQAQQAQQAQQAQQAQQAQQAQQAQAQQAQQA</u>
<u>QAQQAQQAQAQQARAQQA</u>QVQARSQTQALVQARSQAQAQRQAHVIQPPQ
LSLDTPGPSPLEDAFEISYAP

polyQ region of wt-ataxin-3, a deubiquitinase, is expanded in spinocerebellar ataxia type 3. The normal length polyQ region mediates binding of ataxin-3 to beclin-1, preventing its degradation and allowing it to stimulate autophagy (Fig. 2, top line). Soluble, mutated polyQ segments can inhibit this binding, thus preventing beclin-1 degradation and upregulation of autophagy, preventing efficient clearance of aging cellular, as well as viral, proteins. Another indication that polyQ sequences from the virus, or some other repeat in ɣ34.5, may also be involved in this regulation is that there is a discontinuity within the (otherwise well conserved) ɣ34.5 sequence in many herpes isolates (see supplementary material). Such discontinuities usually indicate repeat insertions [68].

This finding ties in with many years of research on the effect of inhibiting autophagy on replication and neurovirulence of various viruses [69]. While neurovirulent viruses such as herpes are indeed held in check by autophagy, some RNA viruses subvert the process for their own replication (e.g., picornaviruses [70], dengue [71]). Although poliovirus requires autophagy for non-lytic spread, its replication is not affected by beclin-1 inhibition [72], suggesting it uses other ways to trigger the process.

Role of PolyQ Regions in Maintaining Latency As is probably the case with the polyQ repeat in murine SRY, polyQ repeat regions in viral proteins are generally variable in length and may be unstructured or "disordered" [73]. However, some of the examples where the repeats are found suggest they have important functions that would not be obvious during in vitro replication. Once they have infected a cell, viruses enter different growth phases, ranging from almost no replication to rapid growth leading to cell lysis. A herpes virus-infected ganglion may contain less than 1000 copies of the virus/cell in the latent state and still successfully reactivate after stress (from heat, UV light exposure or infection with, for example, a rhinovirus) [74]. Although RNA viruses are generally considered to be "hit and run", with rapid clearance from the serum, recent experience with Zika [75–77] and Ebola [78, 79] viruses has shown that some may also persist within body compartments where they are protected from the immune response.

This leads us to a complicated equation: a virus seeking to survive intracellularly must sacrifice rapid growth for its ability to evade immune detection. Variable polyQ repeats may allow a virus to adjust to changing levels of required cellular factors [80], and determine whether the virus is able to actively replicate, or assume a lysogenic state. Herpes viruses in particular are known to incorporate genes from the cells they infect into their genomes that may aid in maintaining lysogeny. PolyQ insertions at the amino acid or RNA level may directly contribute to viral latency by lowering the transcription or activity of the affected proteins. Alternatively, their presence, or the RNA tracts encoding them, could contribute to neurovirulence by mechanisms demonstrated for human proteins (e.g., huntingtin).

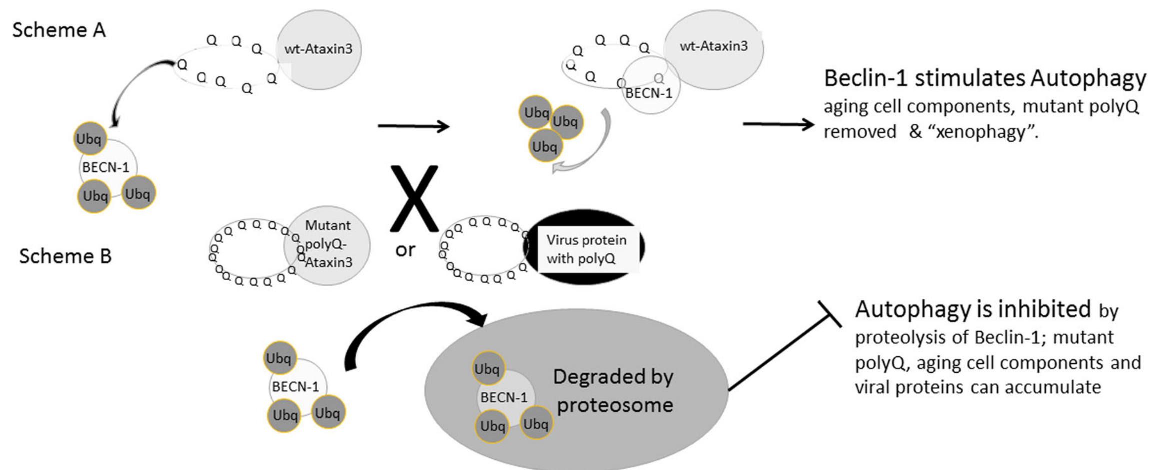

Fig. 2 Soluble polyQ segments (of cell or viral origin) may prevent beclin-1-induced autophagy, which depends on the DNA binding ability of the polyQ segment of wt-ataxin-3 (based on [2, 67]). Scheme A shows that under normal cell conditions, ataxin-3 binding (mediated by its polyQ region) to beclin-1 (BECN) protects it from proteosomal degradation. This allows beclin-1 to stimulate autophagy, which eliminates both aging cellular proteins and those of viral invaders. Scheme B suggests that viral proteins' polyQ, similar to the extended polyQ loop of mutant ataxin-3, can interfere with this control by preventing ataxin-3 from binding. Beclin-1 is now degraded and cannot stimulate autophagy, resulting in even more accumulation of polyQ tracts, defective cellular, and viral proteins that will interfere with normal metabolism

Many herpes virus proteins contain conserved, variable length polyQ segments (Table 2 gives a sampling), including regulatory proteins, an apoptosis inhibitor, and uracil-DNA glycosylase (UDG), all factors that may affect viral replication positively or negatively. It may also be pertinent that a long, Q-rich repeat is present immediately after the catalytic domain of the deoxyuridine 5′-triphosphate nucleotide hydrolase (DUT) gene of the red deer parapox virus (RDPV). Similar sequences have not been reported in the DUT enzymes of other viruses, nor has the role of the polyQ sequence been determined in RDPV. However, UDG and DUT enzymes, which remove or prevent insertion of U residues in viral DNA, are found in all herpes viruses [81]. Their enzymatic activity is essential for neurovirulence, neuroinvasion, and escape from latency of herpes viruses [82]. Mutation of the virus encoded DUT inhibits transcription of equine infectious anemia virus (EILV, a lentivirus and retrovirus). On the other hand, EILV can replicate in non-dividing cells [83] if it allows incorporation of U into its DNA [84].

Accordingly, insertion or amplification of the polyQ segment in UDG or DUT could slow replication to help maintain a latent state. As single point mutations (D71E in the active site, or those preventing phosphorylation of S187 [85]) are sufficient to reduce neurovirulence, DUT may also be a target for antiviral drug design [86]. However, such inhibitors must be very efficient, as residual low levels of the enzyme might have the negative effect of prolonging viral latency (analogous to antibiotic treatment selecting for slow-growing bacterial persister cells [87]).

The long polyQ repeats in other herpes virus proteins (Fig. 1) may also help to suppress virus growth during latency. These include the direct repeats of polyQ that occur in the low complexity C-terminal regions of the Tupaiid T2 protein (β-Herpes group F, isolated from a lymphoma in a tree shrew [88]) and the RF1 protein of Radinovirus type 1 (γ-Herpesvirus), isolated from a Kaposi's sarcoma-like lesion in a macaque [89]. It is possible that these polyQ repeats were directly incorporated from the host cell genes, as their sequences are quite similar to some host proteins (Fig. 3). Further evidence that these polyQ repeats were incorporated in an adventitious fashion from the host cell is that repeats are not found in the published sequences of the (otherwise similar) N1 proteins of Radinoviruses type 2 [91]. Longer repetitive regions could slow growth by decreasing transcription of an essential enzyme, making its RNA more vulnerable to cellular nucleases, and at the protein level, reducing its solubility [92] or enhancing its degradability. Under growth conditions allowing the virus to resume lytic growth, where the enzyme activity is required to ensure efficient replication, the region

```
Tupaiid herpesvirus 1 protein T2:
RSSSSRSSRRRPLLRPPSPDLPQAPPRPRR[Q33]PPPPQKQQPRPPPL
Human huntingtin fragment:
      MATLEKLMKAFESLKSF[Q45]PPPPPPPPPPPPQLPQP
Pig huntingtin:
      MKAFESLKSF[Q24]PPPPPPQPPQPPPQTQPPPQPPP
Human ataxin 2 fragment:
                  [Q45]PPPAAANVRKPG
Human TATA box binding protein:
      PQPIQNTNSLSILEEQQR[Q47]AVAAAAVQQSPS
Human Ataxin 3 variant:
NLTSEELRKRREAYFEKQQQK[Q66]
```

Fig. 3 The polyQ region in the Tupaiid T2 protein (herpes virus group F, isolated from a lymphoma in a tree shrew) is flanked by poly-prolines (P), similar to polyQ expansions in huntingtn, and ataxins associated with neurological disease. Proline residues may also affect protein solubility [90]. Two other mammalian proteins that also contain long polyQ repeats are shown for comparison. The T2 repeat is encoded primarily by CAG codons, as is the case with huntingtn, and the nidovirus repeat (Fig. S1)

encoding the polyQ segment could be rapidly removed at the gene level.

PolyQ Repeats in Proteins that Mediate Virus Transmissibility

As with the RNA viruses, the published sequences of only a few Poxviridae proteins contain even a tetrad QQQQ repeat. However, there is a conserved polyQ insertion of variable length in the A-type inclusion proteins (ATI) of cowpox (CWPX) viruses (Table 3). As with the Q-rich repeat in the MC006L protein of *Molluscum contagiosum* (Fig. 1), this repeat is a variable area in an otherwise well-conserved protein (Supplementary material). The ATI with the longest polyQ segment is in strain FM2292, isolated from a lesion in a vole, which causes skin lesions and mild symptoms in its host. Although the length of the polyQ segment in the CWPX strains in Table 3 is not directly related to pathogenicity, ATI plays a role in a more difficult to measure parameter: transmissibility. The ATI protein, together with the p4C protein [93, 94], allows CWPX to form protein inclusions that, when excreted from the animal, protect the virus from the elements. Inclusions that sequester the virus (V^+ phenotype) contribute to the high transmissibility of CWPX in the wild. As Jennings noted centuries ago, nearly all milkmaids had been infected with CWPX, as were probably most cows. It is significant that in a comparison of three CWPX strains, only the FM2292 virus, which contained the longest polyQ insertion in its ATI, made V^+ inclusions containing virus particles [95]. Two strains with shorter polyQ segments, the index strain Brighton Red and a similar strain from rat, formed inclusions that contained no internalized virus particles (V^0). The ATIs of these three strains differ primarily in their polyQ repeat region length (Table 3 and supplementary).

Growth in cell culture alone does not indicate that ATI is an essential gene [96], although it is one of the most abundant CWPX proteins, amounting to as much as half of all protein synthesis in the "late-late stages" of replication [97]. Deleting the ATI gene leads to a faster growing virus [98]. However, as discussed above, ATI enhances transmissibility from animal to animal, as well as virus survival outside the host. The presence of a longer polyQ sequence could reduce its transcription, synthesis, or solubility during restrictive growth in an organism, where ATI's activity is not required.

Smallpox and vaccinia virus (VV) strains lack polyQ segments in their ATIs and form only virus-free inclusions (V^0 phenotype). CWPX and VV strains also differ in the ATI interacting protein, p4C, in that only CWPX strains contain long repeats (up to 28) of aspartate (D) residues. These results suggest strongly that this amino acid repeat, together with the polyQ segment in the ATI, aid in sequestering virus particles into the V^+ inclusions, which further the extracorporeal survival of the virus.

Recent direct, deep sequencing of fresh CWPX isolates from diseased animals indicated diversity in both genome length and coding areas from the Brighton Red reference

Table 3 Variable length polyQ repeat region in the highly conserved A-type inclusion proteins of cowpox strains. The last three lines show data from Hoffman et al. 2015, where the ability of three strains to form virus containing inclusion bodies (V^+ phenotype), which aid in transmissibility, was compared

Cowpox strain	PolyQ region and surrounding area of the ATI
HumGri07/1Russia, 1990	ATGGDKEEQEQQHQQQQPVKVVQTQPDDDG
HumBer07/1	ATGGDKEEQEQQHQQQQQQQQPVKVVQTQPDDDG
EleGri07/1	ATGGDKEEQEQQHQQQQPVKVVQTQPDDDG
CatBer07/1	ATGGDKEEQEQQHQQQQQQQQPVKVVQTQPDDDG
Cowpox virus MonKre08/4	ATGGDKEEQEQQQHQQQQQQQQPVKVVQTQPDDDG
JagKre08/2	ATGGDKEEQEQQQHQQQQQQQQPVKVVQTQPDDDG
JagKre08/1	ATGGDKEEQEQQQHQQQQQQQQPVKVVQTQPDDDG
HumMag07/1	ATGGDKEEQEQQHQEQQHQQQQQQQQQPVKVVQTQPDDDG
HumLan08/1	ATGGDKEEQEQQQHQQQQQQQQPVKVVQTQPDDDG
BeaBer04/1	ATGGDKEEQEQQHQQQQQQQQQQPVKVVQTQPDDDG
BH71/10	ATGGDKEEQEQQQQQQQQQQQQQPVKVVQSQPDDG
Germany_2002_MKY (marmoset, fatal)	ATGGDKEEQEQQQQQQPVKVVQTQPDDDGI
Germany_1998_2	ATGGDKEEQEQQHQQQPVKVVQTQPDDDDG
Germany_1990_2 (human, fatal)	ATGGDKEEQEQQQQQQQQQQQQPVKVVQSQPDDD
Germany_1980_EP4 (Elephant, 1980)	ATGGDKEEQQQQQQQQQQQQQQQPVKVVQTQPDDDG
CPR06	ATGGDKEEQEQQPVKVVQSKPDDGITPYN
CPXV Amadeus 2015	ATGGDKEEQEQQHQQQQQQQQPVKVVQTQPDDDG
RatHei09/1 V^0	ATGGDK**EEQEQQQHQQQQQQQQ**PVKVVQTQPDDDG
Brighton Red V^0	ATGGDK**EEQEQQ**PVKVVQSKPDDGITPYN
FM2292: V^+	ATGGDK**EQQQQQQQQQQQQQQQQQQQQQQ**PVKVVQSQPDDG

strain, including an additional 6000 bp ORF [95]. As Table 3 illustrates, freshly isolated strains have the longest polyQ region in the ATI, which makes it difficult to determine a "wild type" length of polyQ. It is, for example, possible that the polyQ repeats in CWPX ATI can be selected against during growth in tissue culture. In keeping with this, there is no polyQ repeat in the (extensively passaged [98]) Brighton Red strain, first isolated in 1937 in England from human lesions. This strain would be expected to transmit poorly in the wild, thanks to its V^0 phenotype. More recently isolated German strains (1998 and 2002) have shorter polyQ regions than isolates from 1980 or 1990, but it is unknown how often these have been transferred in cell culture [99].

PolyQ Repeats as a Key to Antiviral Therapy As noted in the introduction to this article, a primary reason for documenting the presence of polyQ segments in viruses is the role polyQ sequences in human proteins have been shown to play in human neurological syndromes [100–102]. Considering the importance of glutamine metabolism for central nervous system function, it would be instructive to specifically test the role of the Q-rich regions on virus latency or replication in neuronal cells. Glutamine itself is extremely important in brain chemistry, and inhibitors similar to this amino acid have antiviral activity. A Q analogue, 6-diazo-5-oxo-l-norleucine (DON), can delay encephalitis caused by alphaviruses, such as Sindbis, by reducing the amount of glutamate synthesized from glutamine [103]. Adding polyQ tracts to the antiviral agent zanamivir greatly enhanced its anti-influenza activity [104].

As Fig. 3 shows, the viral proteins that contain long polyQ segments are very similar to those implicated in Huntington's disease and human ataxias, and may thus be targeted by protein- [16] or gene-based [15, 101, 105–107] therapies similar to those now being tested. Going forward, diagnostics should, as much as possible, distinguish polyQ sequences due to a latent virus from those indicating a mutation in a human gene. The flanking regions, which contain proline repeats (PolyP), may also affect the solubility of the proteins [90]. To date, there have been few investigations of a direct role for these polyQ repeats in initiating neural damage. Aiding in establishing a latent infection could, in itself, contribute to neurovirulence, due to the presence of viral products [108].

Conclusions

PolyQ repeats in viruses could play important roles in controlling transcription, latency, transmissibility, and neurovirulence, whereby the latter three aspects of virus pathogenicity are independent of the ability of the virus to grow to high titer in cell culture. Long polyQ tracts in the protein products of neurotropic and cancer-related DNA viruses could chronically disturb their host cells, by mechanisms similar to those identified for huntingtin and other ataxia-related proteins that contain similar repeats.

Just as B cells and other somatic cells may change their genome structure upon differentiation, it is probable that rapidly growing viruses (and those adapted to tissue culture) have different sequences than those in a latent state. Serial cultivation can favor rapid growth and the loss of pathogenic characteristics, an attenuation process used since the first vaccines against Yellow Fever [109] and poliovirus [110]. The instability of repeated CAG regions that encode polyQ repeat sequences might be a mechanism for adapting virus replication to changes in environmental factors [38]. This means that they may be selectively excised during generation of subgenomic RNAs or resumption of active growth after latent periods. Thus, rational reference sequences of viruses should be based on those obtained from direct isolates of diseased tissue or consensus sequences covering many isolates [111–113].

As the Brighton Red example illustrates, historical reference strains, many of which have been transferred multiple times in labs across the globe, may have long ago eliminated their unstable polyQ repeat regions. As more direct sequences from infected tissues become available, it is possible that polyQ repeats will be found in many other viral proteins. Several methods have been validated for identifying such long repeat sequences [114], which may be difficult to identify with more traditional methods. For example, sequences up to 20 kb can be generated from a single read using "PacBio" or MinIon technology and related methods. This should allow further determination of the accurate length of repeat regions, and better characterization of their importance for neurovirulent virus infections.

Acknowledgements The author thanks all those who sent papers and apologizes in advance for any oversight of relevant literature. The resources of the Department of Biochemistry and Molecular Biology at the University of Texas Medical Branch were used in preparing this manuscript. Dr. Tom Horvath helped prepare Fig. 2.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Research Involving Human Participants and/or Animals There are no human or animal participants.

Informed Consent None required.

References

1. Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature 447(7143):447–452. https://doi.org/10.1038/nature05778

2. Ashkenazi A, Bento CF, Ricketts T, Vicinanza M, Siddiqi F, Pavel M, Squitieri F, Hardenberg MC et al (2017) Polyglutamine tracts regulate beclin 1-dependent autophagy. Nature 545(7652):108–111. https://doi.org/10.1038/nature22078
3. Tomioka I, Ishibashi H, Minakawa EN, Motohashi HH, Takayama O, Saito Y, Popiel HA, Puentes S et al (2017) Transgenic monkey model of the polyglutamine diseases recapitulating progressive neurological symptoms. eNeuro 4(2):ENEURO.0250–ENEU16.2017. https://doi.org/10.1523/ENEURO.0250-16.2017
4. Colton CA, Xu Q, Burke JR, Bae SY, Wakefield JK, Nair A, Strittmatter WJ, Vitek MP (2004) Disrupted spermine homeostasis: a novel mechanism in polyglutamine-mediated aggregation and cell death. J Neurosci 24(32):7118–7127. https://doi.org/10.1523/JNEUROSCI.1233-04.2004
5. Moulder KL, Onodera O, Burke JR, Strittmatter WJ, Johnson EM Jr (1999) Generation of neuronal intranuclear inclusions by polyglutamine-GFP: analysis of inclusion clearance and toxicity as a function of polyglutamine length. J Neurosci 19(2):705–715
6. Onodera O, Burke JR, Miller SE, Hester S, Tsuji S, Roses AD, Strittmatter WJ (1997) Oligomerization of expanded-polyglutamine domain fluorescent fusion proteins in cultured mammalian cells. Biochem Biophys Res Commun 238(2):599–605. https://doi.org/10.1006/bbrc.1997.7337
7. Onodera O, Roses AD, Tsuji S, Vance JM, Strittmatter WJ, Burke JR (1996) Toxicity of expanded polyglutamine-domain proteins in Escherichia coli. FEBS Lett 399(1–2):135–139
8. Carmona V, Cunha-Santos J, Onofre I, Simoes AT, Vijayakumar U, Davidson BL, Pereira de Almeida L (2017) Unravelling endogenous microRNA system dysfunction as a new pathophysiological mechanism in Machado-Joseph disease. Mol Ther 25(4):1038–1055. https://doi.org/10.1016/j.ymthe.2017.01.021
9. Belikov S, Bott LC, Fischbeck KH, Wrange O (2015) The polyglutamine-expanded androgen receptor has increased DNA binding and reduced transcriptional activity. Biochem Biophys Rep 3:134–139. https://doi.org/10.1016/j.bbrep.2015.07.014
10. Ratovitski T, Chaerkady R, Kammers K, Stewart JC, Zavala A, Pletnikova O, Troncoso JC, Rudnicki DD et al (2016) Quantitative proteomic analysis reveals similarities between Huntington's disease (HD) and Huntington's disease-like 2 (HDL2) human brains. J Proteome Res 15(9):3266–3283. https://doi.org/10.1021/acs.jproteome.6b00448
11. Lanz RB, Wieland S, Hug M, Rusconi S (1995) A transcriptional repressor obtained by alternative translation of a trinucleotide repeat. Nucleic Acids Res 23(1):138–145
12. Seixas AI, Holmes SE, Takeshima H, Pavlovich A, Sachs N, Pruitt JL, Silveira I, Ross CA et al (2012) Loss of junctophilin-3 contributes to Huntington disease-like 2 pathogenesis. Ann Neurol 71(2):245–257. https://doi.org/10.1002/ana.22598
13. Monteys AM, Ebanks SA, Keiser MS, Davidson BL (2017) CRISPR/Cas9 editing of the mutant huntingtin allele in vitro and in vivo. Mol Ther 25(1):12–23. https://doi.org/10.1016/j.ymthe.2016.11.010
14. Keiser MS, Monteys AM, Corbau R, Gonzalez-Alegre P, Davidson BL (2016) RNAi prevents and reverses phenotypes induced by mutant human ataxin-1. Ann Neurol 80(5):754–765. https://doi.org/10.1002/ana.24789
15. Batra R, Nelles DA, Pirie E, Blue SM, Marina RJ, Wang H, Chaim IA, Thomas JD et al (2017) Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. Cell 170(5):899–912 **e810**. https://doi.org/10.1016/j.cell.2017.07.010
16. Ripaud L, Chumakova V, Antonin M, Hastie AR, Pinkert S, Korner R, Ruff KM, Pappu RV et al (2014) Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome. Proc Natl Acad Sci U S A 111(51):18219–18224. https://doi.org/10.1073/pnas.1421313111
17. Bettencourt BR, Hogan CC, Nimali M (2007) Polyglutamine expansion in Drosophila: thermal stress and Hsp70 as selective agents. J Biosci 32(3):537–547
18. Bugert JJ, Darai G (2000) Poxvirus homologues of cellular genes. Virus Genes 21(1–2):111–133
19. Cao H, Dai P, Wang W, Li H, Yuan J, Wang F, Fang CM, Pitha PM et al (2012) Innate immune response of human plasmacytoid dendritic cells to poxvirus infection is subverted by vaccinia E3 via its Z-DNA/RNA binding domain. PLoS One 7(5):e36823. https://doi.org/10.1371/journal.pone.0036823
20. Bahr U, Darai G (2004) Re-evaluation and in silico annotation of the Tupaia herpesvirus proteins. Virus Genes 28(1):99–120. https://doi.org/10.1023/B:VIRU.0000012267.97659.e0
21. Brennan G, Kitzman JO, Shendure J, Geballe AP (2015) Experimental evolution identifies vaccinia virus mutations in A24R and A35R that antagonize the protein kinase R pathway and accompany collapse of an extragenic gene amplification. J Virol 89(19):9986–9997. https://doi.org/10.1128/JVI.01233-15
22. Orvedahl A, Levine B (2009) Autophagy in mammalian antiviral immunity. Curr Top Microbiol Immunol 335:267–285. https://doi.org/10.1007/978-3-642-00302-8_13
23. Park S, Buck MD, Desai C, Zhang X, Loginicheva E, Martinez J, Freeman ML, Saitoh T et al (2016) Autophagy genes enhance murine gammaherpesvirus 68 reactivation from latency by preventing virus-induced systemic inflammation. Cell Host Microbe 19(1):91–101. https://doi.org/10.1016/j.chom.2015.12.010
24. Hain D, Bettencourt BR, Okamura K, Csorba T, Meyer W, Jin Z, Biggerstaff J, Siomi H et al (2010) Natural variation of the amino-terminal glutamine-rich domain in Drosophila argonaute2 is not associated with developmental defects. PLoS One 5(12):e15264. https://doi.org/10.1371/journal.pone.0015264
25. Meyer WJ, Schreiber S, Guo Y, Volkmann T, Welte MA, Muller HA (2006) Overlapping functions of argonaute proteins in patterning and morphogenesis of Drosophila embryos. PLoS Genet 2(8):e134. https://doi.org/10.1371/journal.pgen.0020134
26. Palmer WH, Obbard DJ (2016) Variation and evolution in the glutamine-rich repeat region of Drosophila Argonaute-2. G3 (Bethesda) 6(8):2563–2572. https://doi.org/10.1534/g3.116.031880
27. He Z, Liu Y, Liang D, Wang Z, Robertson ES, Lan K (2010) Cellular corepressor TLE2 inhibits replication-and-transcription-activator-mediated transactivation and lytic reactivation of Kaposi's sarcoma-associated herpesvirus. J Virol 84(4):2047–2062. https://doi.org/10.1128/JVI.01984-09
28. Perutz MF (1999) Glutamine repeats and neurodegenerative diseases: molecular aspects. Trends Biochem Sci 24(2):58–63
29. Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S et al (2012) ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res 40(Database issue):D593–D598. https://doi.org/10.1093/nar/gkr859
30. Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S et al (2017) Influenza research database: an integrated bioinformatics resource for influenza virus research. Nucleic Acids Res 45(D1):D466–D474. https://doi.org/10.1093/nar/gkw857
31. Schein CH (2010) Protein aggregation and precipitation, measurement and control. In: Flickinger MC (ed) Encyclopedia of industrial biotechnology. John Wiley & Sons, Inc., Hoboken. https://doi.org/10.1002/9780470054581.eib052
32. Kidanemariam DB, Abraham AD, Sukal AC, Holton TA, Dale JL, James AP, Harding RM (2016) Complete genome sequence of a novel zantedeschia mild mosaic virus isolate: the first report from Australia and from Alocasia sp. Arch Virol 161(4):1079–1082. https://doi.org/10.1007/s00705-015-2745-z

33. Huang CH, Chang YC (2005) Identification and molecular characterization of Zantedeschia mild mosaic virus, a new calla lily-infecting potyvirus. Arch Virol 150(6):1221–1230. https://doi.org/10.1007/s00705-004-0488-3
34. Li H, Xu CP, Yan JY, Lu YY, Jin QQ, Feng Y, Mo SH (2013) Study on the complete sequence of CA24 variant isolated during the acute hemorrhagic conjunctivitis outbreaks in Zhejiang province during 2002 to 2010. Zhonghua Liu Xing Bing Xue Za Zhi 34(5):496–502
35. Uccellini L, Ossiboff RJ, de Matos RE, Morrisey JK, Petrosov A, Navarrete-Macias I, Jain K, Hicks AL et al (2014) Identification of a novel nidovirus in an outbreak of fatal respiratory disease in ball pythons (Python regius). Virol J 11:144. https://doi.org/10.1186/1743-422X-11-144
36. O'Dea MA, Jackson B, Jackson C, Xavier P, Warren K (2016) Discovery and partial genomic characterisation of a novel nidovirus associated with respiratory disease in wild shingleback lizards (Tiliqua rugosa). PLoS One 11(11):e0165209. https://doi.org/10.1371/journal.pone.0165209
37. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ (2006) Nidovirales: evolving the largest RNA virus genome. Virus Res 117(1):17–37. https://doi.org/10.1016/j.virusres.2006.01.017
38. Atanesyan L, Gunther V, Dichtl B, Georgiev O, Schaffner W (2012) Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. Biol Chem 393(1–2):63–70. https://doi.org/10.1515/BC-2011-252
39. Ng WC, Soto-Acosta R, Bradrick SS, Garcia-Blanco MA, Ooi EE (2017) The 5′ and 3′ untranslated regions of the flaviviral genome. Viruses 9(6). https://doi.org/10.3390/v9060137
40. Garcia-Blanco MA, Vasudevan SG, Bradrick SS, Nicchitta C (2016) Flavivirus RNA transactions from viral entry to genome replication. Antivir Res 134:244–249. https://doi.org/10.1016/j.antiviral.2016.09.010
41. Madhugiri R, Karl N, Petersen D, Lamkiewicz K, Fricke M, Wend U, Scheuer R, Marz M et al (2017) Structural and functional conservation of cis-acting RNA elements in coronavirus 5′-terminal genome regions. Virology 517:44–55. https://doi.org/10.1016/j.virol.2017.11.025
42. Rappe JCF, de Wilde A, Di H, Muller C, Stalder H, V'Kovski P, Snijder E, Brinton MA et al (2018) Antiviral activity of K22 against members of the order Nidovirales. Virus Res 246:28–34. https://doi.org/10.1016/j.virusres.2018.01.002
43. Schein CH (1997) From housekeeper to microsurgeon: the diagnostic and therapeutic potential of ribonucleases. Nat Biotechnol 15(6):529–536. https://doi.org/10.1038/nbt0697-529
44. Schein CH (2001) Producing soluble recombinant RNases and assays to measure their interaction with interferon-gamma in vitro. Methods Mol Biol 160:113–137. https://doi.org/10.1385/1-59259-233-3:113
45. Schein CH, Haugg M (1995) Deletions at the C-terminus of interferon gamma reduce RNA binding and activation of double-stranded-RNA cleavage by bovine seminal ribonuclease. Biochem J 307(Pt 1):123–127
46. Schein CH, Haugg M, Benner SA (1990) Interferon-gamma activates the cleavage of double-stranded RNA by bovine seminal ribonuclease. FEBS Lett 270(1–2):229–232
47. Zhao L, Jha BK, Wu A, Elliott R, Ziebuhr J, Gorbalenya AE, Silverman RH, Weiss SR (2012) Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. Cell Host Microbe 11(6):607–616. https://doi.org/10.1016/j.chom.2012.04.011
48. Kindler E, Gil-Cruz C, Spanier J, Li Y, Wilhelm J, Rabouw HH, Zust R, Hwang M et al (2017) Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. PLoS Pathog 13(2):e1006195. https://doi.org/10.1371/journal.ppat.1006195
49. Lorenz R, Luntzer D, Hofacker IL, Stadler PF, Wolfinger MT (2016) SHAPE directed RNA folding. Bioinformatics 32(1):145–147. https://doi.org/10.1093/bioinformatics/btv523
50. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA Package 2.0. Algorithms Mol Biol 6:26. https://doi.org/10.1186/1748-7188-6-26
51. Barik S (2017) Amino acid repeats avert mRNA folding through conservative substitutions and synonymous codons, regardless of codon bias. Heliyon 3(12):e00492. https://doi.org/10.1016/j.heliyon.2017.e00492
52. Gao R, Matsuura T, Coolbaugh M, Zuhlke C, Nakamura K, Rasmussen A, Siciliano MJ, Ashizawa T et al (2008) Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. Eur J Hum Genet 16(2):215–222. https://doi.org/10.1038/sj.ejhg.5201954
53. Su XA, Freudenreich CH (2017) Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 114(40):E8392–E8401. https://doi.org/10.1073/pnas.1711283114
54. Schutze H, Ulferts R, Schelle B, Bayer S, Granzow H, Hoffmann B, Mettenleiter TC, Ziebuhr J (2006) Characterization of white bream virus reveals a novel genetic cluster of nidoviruses. J Virol 80(23):11598–11609. https://doi.org/10.1128/JVI.01758-06
55. Senkevich TG, Koonin EV, Bugert JJ, Darai G, Moss B (1997) The genome of molluscum contagiosum virus: analysis and comparison with other poxviruses. Virology 233(1):19–42. https://doi.org/10.1006/viro.1997.8607
56. Bartolini L, Libbey JE, Ravizza T, Fujinami RS, Jacobson S, Gaillard WD (2018) Viral triggers and inflammatory mechanisms in pediatric epilepsy. Mol Neurobiol. https://doi.org/10.1007/s12035-018-1215-5
57. Courey AJ, Tjian R (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55(5):887–898
58. Gerber HP, Seipel K, Georgiev O, Hofferer M, Hug M, Rusconi S, Schaffner W (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 263(5148):808–811
59. Perutz MF, Johnson T, Suzuki M, Finch JT (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proc Natl Acad Sci U S A 91(12):5355–5358
60. Gemayel R, Chavali S, Pougach K, Legendre M, Zhu B, Boeynaems S, van der Zande E, Gevaert K et al (2015) Variable glutamine-rich repeats modulate transcription factor activity. Mol Cell 59(4):615–627. https://doi.org/10.1016/j.molcel.2015.07.003
61. Zhao L, Ng ET, Davidson TL, Longmuss E, Urschitz J, Elston M, Moisyadi S, Bowles J et al (2014) Structure-function analysis of mouse Sry reveals dual essential roles of the C-terminal polyglutamine tract in sex determination. Proc Natl Acad Sci U S A 111(32):11768–11773. https://doi.org/10.1073/pnas.1400666111
62. Carson DD, Summers MD, Guarino LA (1991) Molecular analysis of a baculovirus regulatory gene. Virology 182(1):279–286
63. Reichert M (2017) Proteome analysis of sheep B lymphocytes in the course of bovine leukemia virus-induced leukemia. Exp Biol Med (Maywood) 242(13):1363–1375. https://doi.org/10.1177/1535370217705864
64. Zhu T, Ji Z, Xu C, Peng Z, Gu L, Zhang R, Liu Y (2014) Expression and prognostic role of SGTA in human breast carcinoma correlates with tumor cell proliferation. J Mol Histol 45(6):665–677. https://doi.org/10.1007/s10735-014-9586-z

65. Ranaghan MJ, Durney MA, Mesleh MF, McCarren PR, Garvie CW, Daniels DS, Carey KL, Skepner AP et al (2017) The autophagy-related beclin-1 protein requires the coiled-coil and BARA domains to form a homodimer with submicromolar affinity. Biochemistry 56(51):6639–6651. https://doi.org/10.1021/acs.biochem.7b00936
66. Orvedahl A, Alexander D, Talloczy Z, Sun Q, Wei Y, Zhang W, Burns D, Leib DA et al (2007) HSV-1 ICP34.5 confers neurovirulence by targeting the beclin 1 autophagy protein. Cell Host Microbe 1(1):23–35. https://doi.org/10.1016/j.chom.2006.12.001
67. Ashkenazi A, Bento CF, Ricketts T, Vicinanza M, Siddiqi F, Pavel M, Squitieri F, Hardenberg MC et al (2017) Polyglutamine tracts regulate autophagy. Autophagy 13(9):1613–1614. https://doi.org/10.1080/15548627.2017.1336278
68. Utturkar SM, Klingeman DM, Hurt RA Jr, Brown SD (2017) A case study into microbial genome assembly gap sequences and finishing strategies. Front Microbiol 8:1272. https://doi.org/10.3389/fmicb.2017.01272
69. Kirkegaard K (2009) Subversion of the cellular autophagy pathway by viruses. Curr Top Microbiol Immunol 335:323–333. https://doi.org/10.1007/978-3-642-00302-8_16
70. Delorme-Axford E, Abernathy E, Lennemann NJ, Bernard A, Ariosa A, Coyne CB, Kirkegaard K, Klionsky DJ (2018) The exoribonuclease Xrn1 is a post-transcriptional negative regulator of autophagy. Autophagy 14(5):898–912. https://doi.org/10.1080/15548627.2018.1441648
71. Mateo R, Nagamine CM, Spagnolo J, Mendez E, Rahe M, Gale M Jr, Yuan J, Kirkegaard K (2013) Inhibition of cellular autophagy deranges dengue virion maturation. J Virol 87(3):1312–1321. https://doi.org/10.1128/JVI.02177-12
72. Bird SW, Kirkegaard K (2015) Escape of non-enveloped virus from intact cells. Virology 479-480:444–449. https://doi.org/10.1016/j.virol.2015.03.044
73. Rees M, Gorba C, de Chiara C, Bui TT, Garcia-Maya M, Drake AF, Okazawa H, Pastore A et al (2012) Solution model of the intrinsically disordered polyglutamine tract-binding protein-1. Biophys J 102(7):1608–1616. https://doi.org/10.1016/j.bpj.2012.02.047
74. Sawtell NM, Poon DK, Tansky CS, Thompson RL (1998) The latent herpes simplex virus type 1 genome copy number in individual neurons is virus strain specific and correlates with reactivation. J Virol 72(7):5343–5350
75. de Laval F, Matheus S, Labrousse T, Enfissi A, Rousset D, Briolant S (2017) Kinetics of Zika viral load in semen. N Engl J Med 377(7):697–699. https://doi.org/10.1056/NEJMc1612600
76. Oliveira Souto I, Alejo-Cancho I, Gascon Brustenga J, Peiro Mestres A, Munoz Gutierrez J, Martinez Yoldi MJ (2018) Persistence of Zika virus in semen 93 days after the onset of symptoms. Enferm Infecc Microbiol Clin 36(1):21–23. https://doi.org/10.1016/j.eimc.2016.10.009
77. Atkinson B, Thorburn F, Petridou C, Bailey D, Hewson R, Simpson AJ, Brooks TJ, Aarons EJ (2017) Presence and persistence of Zika virus RNA in semen, United Kingdom, 2016. Emerg Infect Dis 23(4):611–615. https://doi.org/10.3201/eid2304.161692
78. Sissoko D, Keita M, Diallo B, Aliabadi N, Fitter DL, Dahl BA, Akoi Bore J, Raymond Koundouno F et al (2017) Ebola virus persistence in breast milk after no reported illness: a likely source of virus transmission from mother to child. Clin Infect Dis 64(4):513–516. https://doi.org/10.1093/cid/ciw793
79. Deen GF, Broutet N, Xu W, Knust B, Sesay FR, McDonald SLR, Ervin E, Marrinan JE et al (2017) Ebola RNA persistence in semen of Ebola virus disease survivors-final report. N Engl J Med 377(15):1428–1437. https://doi.org/10.1056/NEJMoa1511410
80. Garcia-Blanco MA, Cullen BR (1991) Molecular basis of latency in pathogenic human viruses. Science 254(5033):815–820
81. Kudelova M, Rajcani J (2009) Gammaherpesviruses and oncogenesis. In: Gluckman TR (ed) Herpesviridae: viral structure, life cycle and infections. Nova Science Publishers, Hauppauge, pp. 187–226
82. Pyles RB, Sawtell NM, Thompson RL (1992) Herpes simplex virus type 1 dUTPase mutants are attenuated for neurovirulence, neuroinvasiveness, and reactivation from latency. J Virol 66(11):6706–6713
83. Payne SL, Elder JH (2001) The role of retroviral dUTPases in replication and virulence. Curr Protein Pept Sci 2(4):381–388
84. Steagall WK, Robek MD, Perry ST, Fuller FJ, Payne SL (1995) Incorporation of uracil into viral DNA correlates with reduced replication of EIAV in macrophages. Virology 210(2):302–313. https://doi.org/10.1006/viro.1995.1347
85. Kato A, Arii J, Koyanagi Y, Kawaguchi Y (2015) Phosphorylation of herpes simplex virus 1 dUTPase regulates viral virulence and genome integrity by compensating for low cellular dUTPase activity in the central nervous system. J Virol 89(1):241–248. https://doi.org/10.1128/JVI.02497-14
86. Topalis D, Gillemot S, Snoeck R, Andrei G (2016) Distribution and effects of amino acid changes in drug-resistant alpha and beta herpesviruses DNA polymerase. Nucleic Acids Res 44(20):9530–9554. https://doi.org/10.1093/nar/gkw875
87. Van den Bergh B, Fauvart M, Michiels J (2017) Formation, physiology, ecology, evolution and clinical importance of bacterial persisters. FEMS Microbiol Rev 41(3):219–251. https://doi.org/10.1093/femsre/fux001
88. Bahr U, Darai G (2001) Analysis and characterization of the complete genome of tupaia (tree shrew) herpesvirus. J Virol 75(10):4854–4870. https://doi.org/10.1128/JVI.75.10.4854-4870.2001
89. Bruce AG, Thouless ME, Haines AS, Pallen MJ, Grundhoff A, Rose TM (2015) Complete genome sequence of pig-tailed macaque rhadinovirus 2 and its evolutionary relationship with rhesus macaque rhadinovirus and human herpesvirus 8/Kaposi's sarcoma-associated herpesvirus. J Virol 89(7):3888–3909. https://doi.org/10.1128/JVI.03597-14
90. Schein CH (1990) Solubility as a function of protein structure and solvent components. Bio/Technology 8(4):308–317
91. Bruce AG, Ryan JT, Thomas MJ, Peng X, Grundhoff A, Tsai CC, Rose TM (2013) Next-generation sequence analysis of the genome of RFHVMn, the macaque homolog of Kaposi's sarcoma (KS)-associated herpesvirus, from a KS-like tumor of a pig-tailed macaque. J Virol 87(24):13676–13693. https://doi.org/10.1128/JVI.02331-13
92. Schein CH (1994) Controlling oligomerization of pharmaceutical proteins. Pharm Acta Helv 69(3):119–126
93. Okeke MI, Okoli AS, Nilssen O, Moens U, Tryland M, Bohn T, Traavik T (2014) Molecular characterization and phylogenetics of Fennoscandian cowpox virus isolates based on the p4c and atip genes. Virol J 11:119. https://doi.org/10.1186/1743-422X-11-119
94. Okeke MI, Adekoya OA, Moens U, Tryland M, Traavik T, Nilssen O (2009) Comparative sequence analysis of A-type inclusion (ATI) and P4c proteins of orthopoxviruses that produce typical and atypical ATI phenotypes. Virus Genes 39(2):200–209. https://doi.org/10.1007/s11262-009-0376-8
95. Hoffmann D, Franke A, Jenckel M, Tamosiunaite A, Schluckebier J, Granzow H, Hoffmann B, Fischer S et al (2015) Out of the reservoir: phenotypic and genotypic characterization of a novel cowpox virus isolated from a common vole. J Virol 89(21):10959–10969. https://doi.org/10.1128/JVI.01195-15
96. Xu Z, Zikos D, Osterrieder N, Tischer BK (2014) Generation of a complete single-gene knockout bacterial artificial chromosome library of cowpox virus and identification of its essential genes. J Virol 88(1):490–502. https://doi.org/10.1128/JVI.02385-13

97. Funahashi S, Sato T, Shida H (1988) Cloning and characterization of the gene encoding the major protein of the A-type inclusion body of cowpox virus. J Gen Virol 69(Pt 1):35–47. https://doi.org/10.1099/0022-1317-69-1-35
98. Kastenmayer RJ, Maruri-Avidal L, Americo JL, Earl PL, Weisberg AS, Moss B (2014) Elimination of A-type inclusion formation enhances cowpox virus replication in mice: implications for orthopoxvirus evolution. Virology 452-453:59–66. https://doi.org/10.1016/j.virol.2013.12.030
99. Carroll DS, Emerson GL, Li Y, Sammons S, Olson V, Frace M, Nakazawa Y, Czerny CP et al (2011) Chasing Jenner's vaccine: revisiting cowpox virus classification. PLoS One 6(8):e23086. https://doi.org/10.1371/journal.pone.0023086
100. Perutz MF, Pope BJ, Owen D, Wanker EE, Scherzinger E (2002) Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid beta-peptide of amyloid plaques. Proc Natl Acad Sci U S A 99(8):5596–5600. https://doi.org/10.1073/pnas.042681599
101. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, Stevanin G, Koutsis G et al (2016) DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol 79(6):983–990. https://doi.org/10.1002/ana.24656
102. Cortes CJ, La Spada AR (2018) X-linked spinal and bulbar muscular atrophy: from clinical genetic features and molecular pathology to mechanisms underlying disease toxicity. Adv Exp Med Biol 1049:103–133. https://doi.org/10.1007/978-3-319-71779-1_5
103. Baxter VK, Glowinski R, Braxton AM, Potter MC, Slusher BS, Griffin DE (2017) Glutamine antagonist-mediated immune suppression decreases pathology but delays virus clearance in mice during nonfatal alphavirus encephalomyelitis. Virology 508:134–149. https://doi.org/10.1016/j.virol.2017.05.013
104. Weight AK, Belser JA, Tumpey TM, Chen J, Klibanov AM (2014) Zanamivir conjugated to poly-L-glutamine is much more active against influenza viruses in mice and ferrets than the drug itself. Pharm Res 31(2):466–474. https://doi.org/10.1007/s11095-013-1175-4
105. Yang S, Chang R, Yang H, Zhao T, Hong Y, Kong HE, Sun X, Qin Z et al (2017) CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. J Clin Invest 127(7):2719–2724. https://doi.org/10.1172/JCI92087
106. Moore LR, Rajpal G, Dillingham IT, Qutob M, Blumenstein KG, Gattis D, Hung G, Kordasiewicz HB et al (2017) Evaluation of antisense oligonucleotides targeting ATXN3 in SCA3 mouse models. Mol Ther Nucleic Acids 7:200–210. https://doi.org/10.1016/j.omtn.2017.04.005
107. Keiser MS, Kordasiewicz HB, McBride JL (2016) Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. Hum Mol Genet 25(R1):R53–R64. https://doi.org/10.1093/hmg/ddv442
108. Wang Y, Santerre M, Tempera I, Martin K, Mukerjee R, Sawaya BE (2017) HIV-1 Vpr disrupts mitochondria axonal transport and accelerates neuronal aging. Neuropharmacology 117:364–375. https://doi.org/10.1016/j.neuropharm.2017.02.008
109. Beck A, Tesh RB, Wood TG, Widen SG, Ryman KD, Barrett AD (2014) Comparison of the live attenuated yellow fever vaccine 17D-204 strain to its virulent parental strain Asibi by deep sequencing. J Infect Dis 209(3):334–344. https://doi.org/10.1093/infdis/jit546
110. Sanders BP, Liu Y, Brandjes A, van Hoek V, de Los Rios Oakes I, Lewis J, Wimmer E, Custers JH et al (2015) Brunenders: a partially attenuated historic poliovirus type I vaccine strain. J Gen Virol 96(9):2614–2622. https://doi.org/10.1099/vir.0.000197
111. Schein CH, Bowen DM, Lewis JA, Choi K, Paul A, van Noort GJV, Lu WZ, Filippov DV (2012) Physicochemical property consensus sequences for functional analysis, design of multivalent antigens and targeted antivirals. BMC Bioinformatics 13:S9. https://doi.org/10.1186/1471-2105-13-s13-s9
112. Danecek P, Lu W, Schein CH (2010) PCP consensus sequences of flaviviruses: correlating variance with vector competence and disease phenotype. J Mol Biol 396(3):550–563. https://doi.org/10.1016/j.jmb.2009.11.070
113. Schein CH, Ye M, Paul AV, Oberste MS, Chapman N, van der Heden van Noort GJ, Filippov DV, Choi KH (2015) Sequence specificity for uridylylation of the viral peptide linked to the genome (VPg) of enteroviruses. Virology 484:80–85. https://doi.org/10.1016/j.virol.2015.05.016
114. Doi K, Monjo T, Hoang PH, Yoshimura J, Yurino H, Mitsui J, Ishiura H, Takahashi Y et al (2014) Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. Bioinformatics 30(6):815–822. https://doi.org/10.1093/bioinformatics/btt647