



**HAL**  
open science

## Identification of Proteins Specifically Assembled on a Stem-Loop Composed of a CAG Triplet Repeat

Robert Fuchs, Asako Isogawa, Joao Paulo, Shingo Fujii

► **To cite this version:**

Robert Fuchs, Asako Isogawa, Joao Paulo, Shingo Fujii. Identification of Proteins Specifically Assembled on a Stem-Loop Composed of a CAG Triplet Repeat. *DNA*, 2023, 3 (2), pp.109-118. 10.3390/dna3020009 . hal-04234597

**HAL Id: hal-04234597**

**<https://hal.science/hal-04234597v1>**

Submitted on 10 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Identification of Proteins Specifically Assembled on a Stem-Loop Composed of a CAG**  
2 **Triplet Repeat**

3

4 Robert P. Fuchs<sup>1,4\*</sup>, Asako Isogawa<sup>2</sup>, Joao A. Paulo<sup>3</sup>, and Shingo Fujii<sup>2,\*</sup>

5

6 <sup>1</sup> Marseille Medical Genetics, UMR1251, Marseille, France

7 <sup>2</sup> Cancer Research Center of Marseille, CNRS, UMR7258, Marseille, France

8 <sup>3</sup> Department of Cell Biology, Harvard Medical School, Boston, USA

9 <sup>4</sup> Department of BCMP, Harvard Medical School, Boston, USA

10 \* Correspondence: [roberpipo@gmail.com](mailto:roberpipo@gmail.com) (RF); [shingo.fujii@inserm.fr](mailto:shingo.fujii@inserm.fr) (SF)

11

12 **Abstract:** Human genomic DNA contains a number of diverse repetitive sequence motifs, often  
13 identified as fragile sites leading to genetic instability. Among them, expansion events occurring at  
14 triplet repeats have been extensively studied due to their association with neurological disorders  
15 including Huntington's disease (HD). In the case of HD, expanded CAG triplet repeats in *HTT*  
16 gene are thought to cause the onset. Expansion of CAG triplet repeats is believed to be triggered  
17 by the emergence of stem-loops composed of CAG triplet repeats while the underlying molecular  
18 mechanisms are largely unknown. Therefore, identification of proteins recruited on such stem-  
19 loops would be useful to understand the molecular mechanisms leading to genetic instability of  
20 CAG triplet repeats. We previously developed a plasmid DNA pull-down methodology that  
21 captures proteins specifically assembled on any sequence-of-interest using nuclear extracts.  
22 Analysis by Mass Spectrometry revealed that among the proteins specifically bound to a stem-  
23 loop composed of CAG triplet repeats, many turned out to belong to DNA repair pathways. We  
24 expect our data set to represent a useful entry point for the design of assays allowing the molecular  
25 mechanisms of genetic instability at CAG triplet repeats to be explored.

26

1 **Keywords:** plasmid DNA pull-down; CAG triplet repeat; MS analysis

2

3

#### 4 **1. Introduction**

5         Maintaining integrity of genomic DNA is an essential demand for healthy life in any species.  
6         However, certain specific DNA sequence contexts (e.g., triplet repeats, G-rich sequences inducing  
7         G-quadruplexes) are intrinsically fragile and potentially harmful during life-span [1, 2]. Among them,  
8         genetic instability of triplet repeats is extensively studied because of their causality to hereditary  
9         or sporadic neurological diseases (e.g., Huntington's disease (HD), fragile X syndrome (FXS),  
10         myotonic dystrophy type 1 (DM1), amyotrophic lateral sclerosis (ALS)) [3]. For a long time, the  
11         pathogenesis of HD is known to be closely correlated to expansion of CAG triplet repeat in *HTT*  
12         the gene encoding huntingtin protein. Onset of HD is thought to be mediated by the expanded  
13         polyglutamine stretches (CAG encodes glutamine) present in the mutant forms of HD protein  
14         leading to protein aggregation and the so-called notion of polyglutamine disorders. Such a  
15         fundamental premise for the onset of HD has, however, recently been challenged since the  
16         expanded length of CAG repeat *per se* was found to be positively correlated to the onset of HD  
17         rather than the length of the polyglutamine track in the HD protein. Indeed, it was shown that  
18         interruption of expanded CAG repeats by CAA triplets, that also encode glutamine residue, delays  
19         the timing of disease onset [4]. This finding indicates that onset of HD directly relies on the length  
20         of the DNA (or RNA) uninterrupted CAG repeats *per se* rather than on the length of polyglutamine  
21         tract in the resulting protein. In fact, such a possibility is likely to be shared by other DNA repeat  
22         expansion diseases for which the culprit *loci* are located outside coding regions (untranslated  
23         region (UTR), introns) [3]. In order to tackle diseases involving structured DNA motifs such as HD,  
24         it is important to discover the proteins that bind to DNA intermediates that arise during the  
25         expansion process *per se*. For instance, it was proposed that the DNA repair enzyme OGG1 (7,8-  
26         dihydro-8-oxoguanine-DNA glycosylase), when acting at an oxidized DNA damage located within

1 a triplet repeat, can trigger repeat expansion via formation of a CAG stem-loop intermediate  
2 structure in mouse cell lines [5].

3 We have developed a novel plasmid DNA pull-down methodology to capture specific  
4 nucleoprotein complexes from nuclear extracts in a DNA sequence-specific manner (termed IDAP:  
5 Isolation of DNA Associated Proteins), which turned out to be highly efficient and versatile [6, 7].  
6 *Xenopus* egg extracts are well known to support various DNA transactions (e.g., Replication,  
7 Repair, DNA damage response) [8–10]. For example, by using such extracts, IDAP was applied  
8 to capture and identify proteins recruited on DNA lesions induced by N-methyl-N-nitrosourea  
9 (MNU), a mimic of chemotherapeutic agent, temozolomide (TMZ) [11, 12]. Based on the MS data  
10 and subsequent biochemical validation, we proposed that the cytotoxic effect of TMZ in non-  
11 dividing cells may be triggered by DNA double-strand break (DSB) generated via accidental  
12 encounter of base excision repair (BER) and mismatch repair (MMR) processes at closely-spaced  
13 lesions [11, 12].

14 As mentioned above, it is believed that the critical feature of triplet repeats relies on their  
15 capacity to form stem-loops, when a triplet repeat sequence is exposed as single-stranded DNA  
16 during transactions such as replication, transcription, repair [13]. As the plasmid DNA pull-down  
17 approach is an efficient protein discovery tool, we applied this approach to identify proteins that  
18 specifically bind to a CAG stem-loop in *Xenopus* egg extracts. For this purpose, we constructed a  
19 plasmid DNA possessing a stem-loop composed of a CAG triplet repeat on one strand [7]. In the  
20 present paper, by using this newly constructed plasmid, we implemented the plasmid DNA pull-  
21 down approach to capture and identify specific proteins assembled on the stem-loop. We report  
22 characteristic features of the identified proteins which appear to, not only help us understand the  
23 molecular events occurring at the stem-loop, but also provide potentially novel clues for developing  
24 effective treatments for patients.

25

## 26 **2. Materials and Methods**

## 1 2.1. Chemicals and Materials

2 Chemicals were from VWR (France) or Thermo Fisher Scientific (France), and all enzymes  
3 were from New England Biolabs (France). Oligo DNA and also TFO were from Eurogentec  
4 (Belgium). A TFO probe, TFO-1, is used [6]. This TFO-1 is composed of a psoralen residue, a 22-  
5 mer locked nucleic acid (LNA)/DNA mixed oligonucleotide, a spacer arm composed of tandemly  
6 oriented hexa-ethylene glycol, and a terminal desthiobiotin residue (see Figure 1 for its schematic  
7 drawing). Sequence context of the 22-mer oligonucleotide is 5'-tTtTcTtTtCtCCtCtTcCct (LNA and  
8 DNA residues are shown in small and capital letters, respectively). Capital C represents 5-methyl  
9 dC residue.

10

## 11 2.2. Plasmid DNA Substrates

12 A CAG triplet repeat-containing plasmid (termed pAS203) is derived from a vector,  
13 pAS200.2, that contains a tag sequence, TFT-1, able to form triple helix with TFO-1 probe [7]. The  
14 sequence of TFT-1 is 5'-AAAAGAAAAGAGGAGAAGAGGA. Plasmid pAS203 contains a stretch  
15 of seven tandemly repeated CAG triplets surrounded by two nickases recognition sites (Figure 2).  
16 Three forms of pAS203 have been generated, a relaxed closed circular form (rcc), a gapped form  
17 (gap), and a closed stem-loop form (cSL)) are termed pAS203(rcc), pAS203(gap), and  
18 pAS203(cSL), respectively (Figure 3). Detailed protocols to prepare all plasmid DNA substrates  
19 are described in [7].

20

## 21 2.3. TFO-Conjugated Plasmid on Magnetic Bead

22 Constructed plasmid substrates (250 ng of pAS203(rcc), pAS203(gap), or pAS203(cSL))  
23 were conjugated with TFO-1 following triple helix formation as previously reported [6, 7]. Each of  
24 the TFO-conjugated plasmid was mixed with 25 µl of Dynabeads M-280 Streptavidin (Invitrogen)  
25 as in [6, 7].

26

1 *2.4. Xenopus Egg Extracts*

2 We used nucleoplasmic extract (NPE) prepared from *Xenopus* eggs [14] as *Xenopus* egg  
3 extracts. Preparation of NPE is described in [15].

4  
5 *2.5. Plasmid DNA Pull-Down and Protein Analysis*

6 Each of the TFO-conjugated plasmid (250 ng) immobilized on M-280 beads was incubated  
7 with 16  $\mu$ l of NPE for 17 min at room temperature (RT) in the following buffer (10 mM HEPES-  
8 KOH (7.7), 50 mM KCl, 2.5 mM MgCl<sub>2</sub>, 0.25 M sucrose, 0.1% NP40). Proteins bound on plasmid  
9 were crosslinked by addition of 320  $\mu$ l of a buffer containing 0.8% formaldehyde for 10 min at RT.  
10 Subsequent steps of beads wash and protein elution were the same procedure as previously  
11 reported [12]. The eluted proteins were analyzed by silver staining following polyacrylamide gel  
12 electrophoresis or by mass spectrometry (MS). For silver staining, an aliquot of each sample  
13 corresponding to 30 ng of plasmid was treated by heat to reverse the crosslinking condition. The  
14 samples were loaded and run on a 4-20% gradient gel (Bio-Rad) at 200 volts for 23 min, and  
15 stained with a silver staining kit (Silver Stain Plus kit, Bio-Rad). MS and its data analysis were the  
16 same procedure as previously reported [12].

17

18 **3. Results**

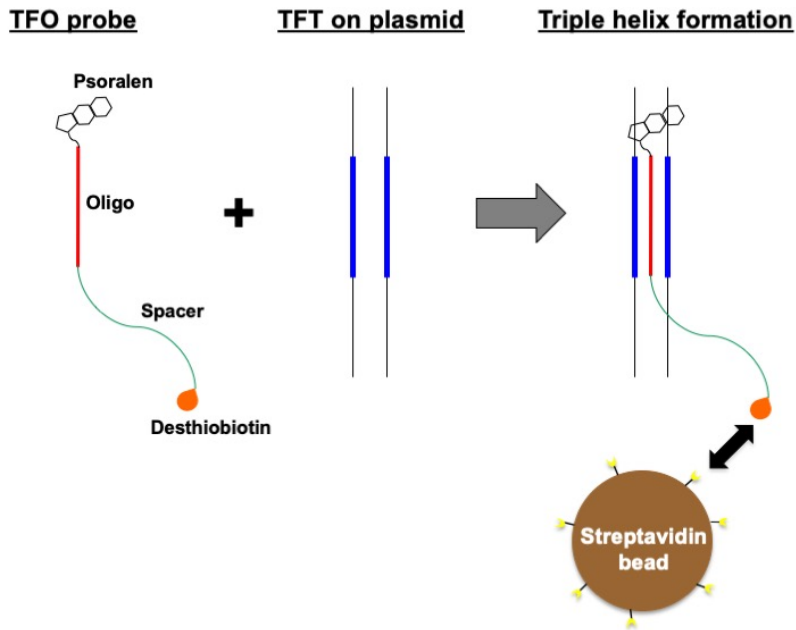
19 Whereas precise molecular mechanisms of triplet repeat expansion are largely unknown,  
20 it has been presumed that stem-loop formation relying on intrinsic properties of triplet repeats  
21 would eventually trigger cascades leading to triplet repeat expansion (or contraction) in dividing  
22 or non-dividing cells [2, 3]. In order to understand the molecular events occurring at stem-loops,  
23 identification of proteins recruited on stem-loops will be instrumental to help understand the  
24 mechanisms leading to genetic instability at triplet repeats.

25

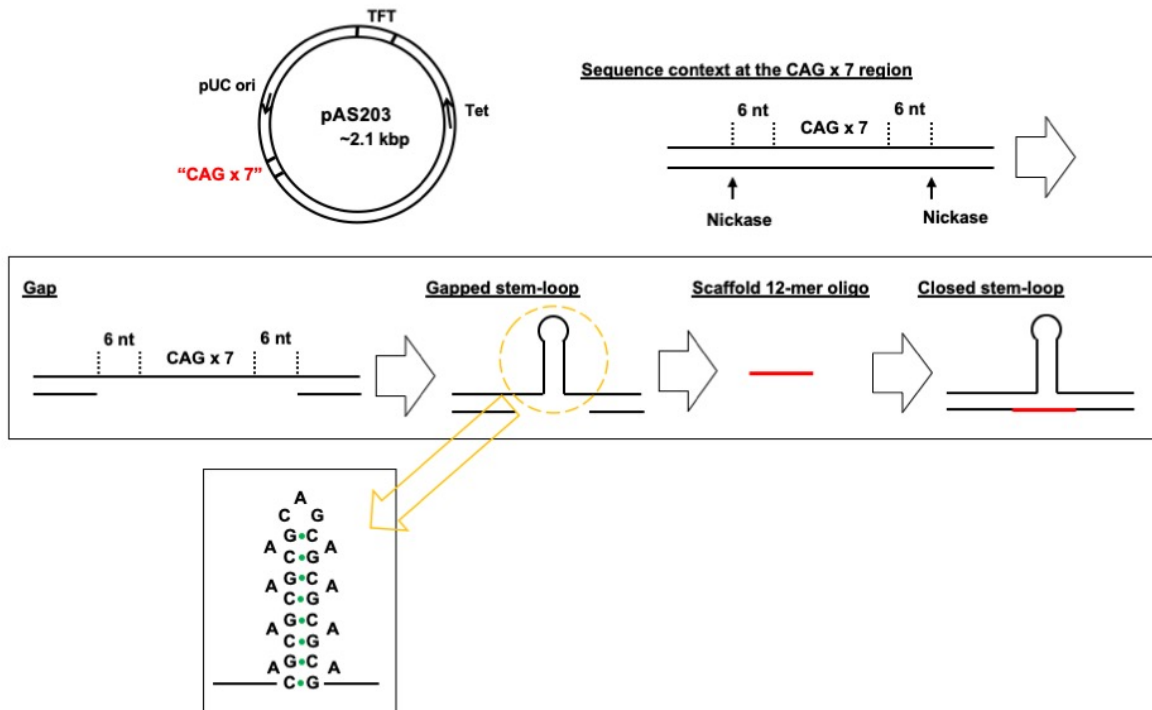
26 *3.1. Assay Design to Capture and Identify Proteins Specifically Assembled on Structural DNA*

1           Our experimental goal is to identify proteins specifically recruited on a CAG triplet repeat  
2 stem-loop. As described in the introduction, we have developed a plasmid DNA pull-down  
3 methodology [6, 7], to discover proteins that specifically associate with any DNA element-of-  
4 interest, such as CAG stem-loops. The core part of the methodology relies on the way the DNA  
5 element-of-interest is immobilized on magnetic beads. This is achieved by the use of a specific  
6 oligonucleotide (TFO probe) that forms triple helix with a cognate double-stranded DNA (dsDNA)  
7 sequence (termed TFT: Triple helix Forming Tag) located in the same plasmid as the DNA element-  
8 of-interest (Figure 1). As the TFO probe carries a desthiobiotin moiety, the binary complex (i.e.,  
9 dsDNA with the TFO probe) can be conjugated to streptavidin-coated magnetic bead. As DNA  
10 substrates for the pull-down experiments, we constructed a plasmid pAS203 containing a stretch  
11 of seven tandemly repeated CAG triplet sequence, and flanked by two nickases' recognition sites  
12 (Figure 2). By utilizing the nickases (Figure 2), three derivatives of pAS203 were constructed  
13 (Figure 3): a relaxed closed circular form, pAS203(rcc), a plasmid that contains the CAG repeat  
14 in its classical double-stranded form; a gapped form, pAS203(gap), a plasmid carrying a 33-nt  
15 long gap that encompasses the (CAG)<sub>7</sub> sequence; a closed stem-loop form, pAS203(cSL) bearing  
16 a (CAG)<sub>7</sub> stem-loop [7]. These constructs were immobilized on beads and employed in the context  
17 of the plasmid DNA pull-down approach.

18



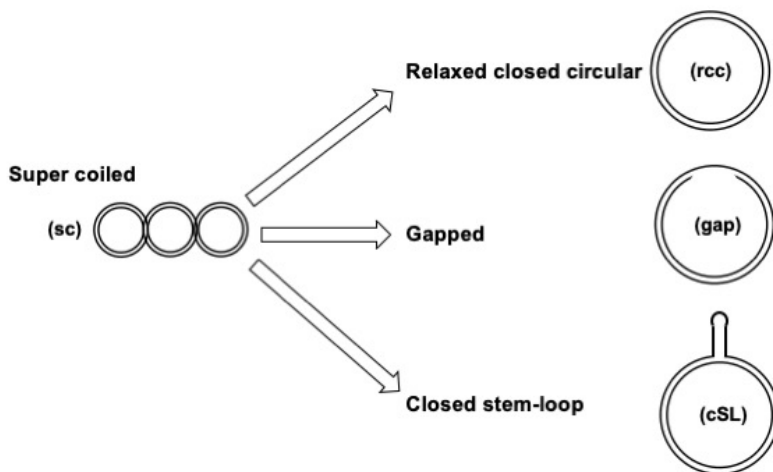
1  
2  
3 **Figure 1.** Strategy to capture a plasmid DNA via triple helix formation. TFO: Triple helix Forming  
4 Oligonucleotide. TFT: Triple helix Forming Tag.





1  
2  
3  
4  
5  
6  
7  
8  
9

**Figure 2.** Construction of a plasmids harboring a CAG-containing gap and stem-loop. Two single-stranded nicks introduced in pAS203 within the same strand by two sequence-specific ‘nickases’ led to the excision of a short DNA fragment that can be released by heating (pAS203(gap)). Formation of a stem-loop composed of the CAG triplet repeat is achieved, following heat treatment, by ligation of a complementary oligonucleotide (12-mer) (pAS203(cSL)). pUC ori: high copy number pUC origin in *E. coli*. Tet: tetracycline resistance gene. CAG x 7: seven tandemly repeated CAG triplets.



10  
11

**Figure 3.** Constitution of DNA substrates. All three plasmid (rcc, gap, cSL) are derived from the same super coiled (sc) plasmid. The rcc is constructed by self-ligation following introduction of a nick by a nickase.

15

### 3.2. Protein Capture in the Context of the Plasmid DNA Pull-Down Approach

17

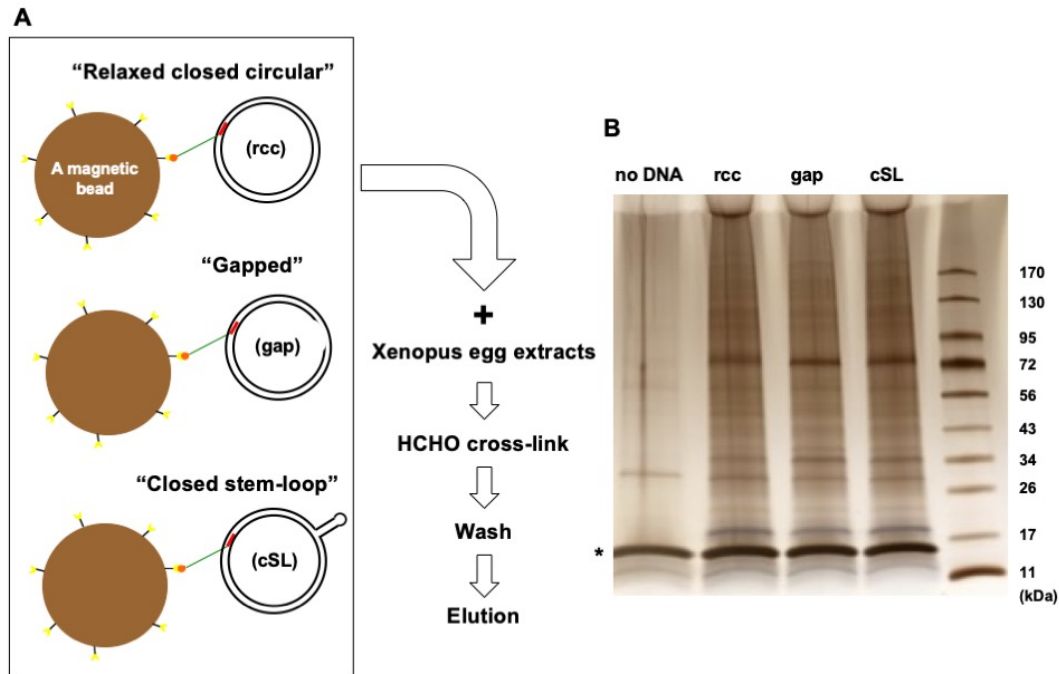
Xenopus egg extracts turned out to be particularly instrumental in the context of our plasmid pull-down methodology. Indeed, within the context of methylated DNA damage, we captured and identified, from frog extracts, specific proteins that specifically interact with alkylating

19

1 lesions. These experiments led us to propose a novel mode of action of the drug Temozolomide  
2 used in the treatment of glioblastomas [12]. Consequently, we decided to utilize *Xenopus* egg  
3 extracts, to discover the proteins that specifically interact with CAG-containing DNA intermediates  
4 related to HD. Each of the three constructed plasmids (i.e., pAS203(rcc), pAS203(gap),  
5 pAS203(cSL)) were individually incubated with TFO-probe to form a triple helix with the cognate  
6 TFT sequence in the plasmid and then mixed with magnetic beads (Figure 4A). Subsequently, the  
7 TFO-conjugated plasmid beads were incubated with *Xenopus* egg extracts, followed by  
8 formaldehyde crosslink treatment. Subsequently, the beads were washed and eluted from the  
9 plasmid after cross-link reversion by heat. Eluted proteins are analyzed by electrophoresis on a  
10 silver-stained PAGE gel. Irrespective of the DNA substrates, the pattern of eluted proteins in lanes  
11 rcc, gap and cSL (Figure 4B) revealed numerous proteins with a similar pattern of bands, a similar  
12 total amount and a broadly distributed size range.

13           Importantly, in contrast to all other lanes, the “no DNA” control lane (Figure 4B) appears to  
14 be essentially free of protein signal reflecting low non-specific background protein carry-over. The  
15 high specificity of the plasmid DNA pull-down procedure assesses the high efficiency of the bead  
16 washing procedure and guarantees optimal capture of specific DNA-binding proteins and their  
17 subsequent identification by MS analysis.

18



1  
2  
3 **Figure 4.** Protein capture by means of the plasmid DNA pull-down approach. (A) Schematic  
4 representation of the plasmid DNA pull-down approach. HCHO: formaldehyde. (B) Eluted proteins  
5 from 30 ng of each plasmid in the context of the plasmid DNA pull-down approach is loaded on a  
6 polyacrylamide gel following reverse crosslink reaction. After electrophoresis, the gel is visualized  
7 by silver staining. The “no DNA” control lane is prepared in the same way but without addition of  
8 plasmid DNA; \*: streptavidin.

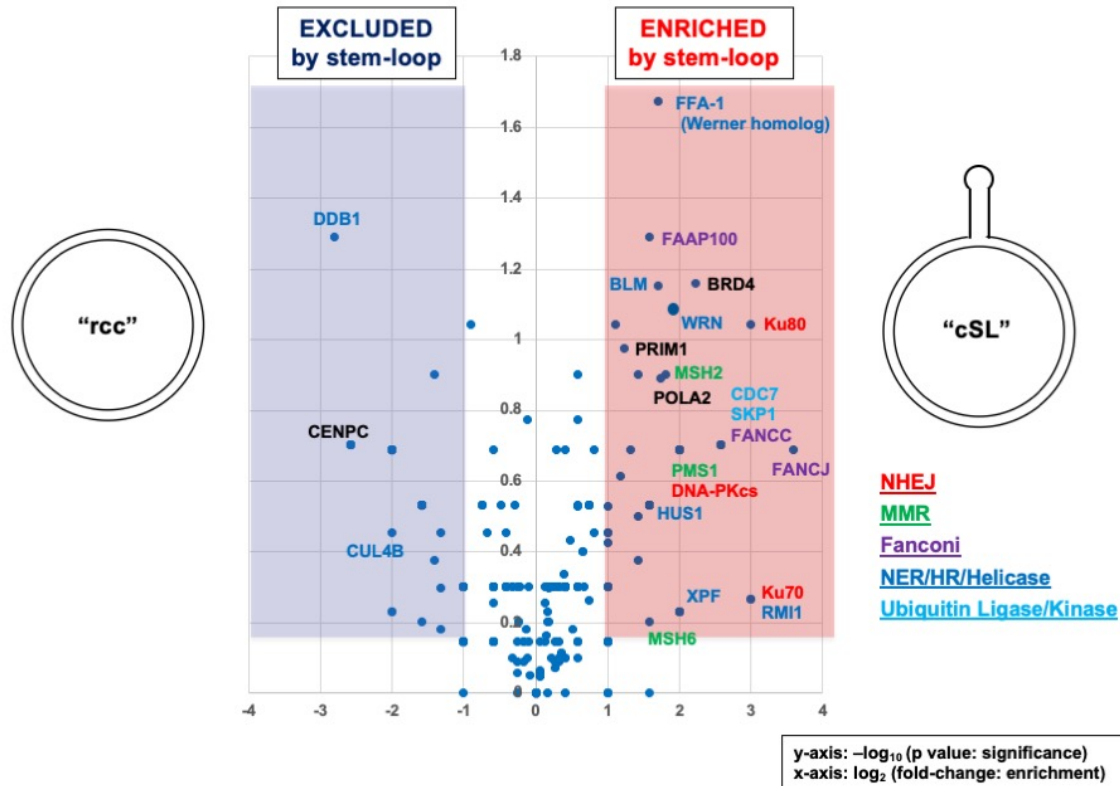
9  
10 **3.3. Identification of Proteins Specifically Recruited on the Stem-Loop**

11 Proteins bound to all three constructs (pAS203(rcc), pAS203(gap), pAS203(cSL)) were  
12 eluted from plasmid, 250ng DNA equivalent each and identified by MS analysis. As expected from  
13 the patterns observed in the silver-stained gel (Figure 4B), numerous proteins were identified by  
14 MS analysis (Table S1). In order to discover proteins that specifically bind to the CAG stem-loop  
15 structure, we compared the proteins eluted from the plasmid containing the stem-loop  
16 (pAS203(cSL)) to the proteins eluted from the plasmid containing the CAG sequence in its classical

1 double-stranded form (pAS203(rcc)). Comparison was achieved by plotting the MS data in the  
2 form of a 'Volcano' plot. Briefly, for each protein the average spectral count value found in the  
3 stem-loop containing plasmid was divided by the average spectral count value in the  
4 corresponding double-stranded plasmid pAS203(rcc). The resulting ratio is plotted as its  $\log_2$  value  
5 along x-axis. The statistical significance of the data is estimated by the p-value in the Student's t-  
6 test and plotted as  $-\log_{10}$  value along y-axis. Proteins that are specifically enriched on the stem-  
7 loop structure appear in the right top-corner of the Volcano plot (Figure 5). Interestingly, among  
8 the enriched proteins many appear to belong to well-characterized DNA repair pathways such as  
9 non-homologous end joining (NHEJ), mismatch repair (MMR), nucleotide excision repair (NER).  
10 In contrast, few proteins appear to be excluded by the stem-loop (left side corner). Similarly, using  
11 the 'Volcano' plot representation, we analyzed the proteins differentially bound to the gap-plasmid  
12 in comparison fully double-stranded DNA (Figure S1 for pAS203(rcc) vs pAS203(gap)) or to the  
13 stem-loop construct (Figure S2 for pAS203(gap) vs pAS203(cSL)).

14

15



1  
2  
3 **Figure 5.** Identification of proteins assembled on the stem-loop. The volcano plot highlights  
4 proteins significantly enriched on pAS203(rcc) and pAS203(cSL) on the left- and right-side top  
5 corners, respectively. All proteins identified in the MS analysis are listed in Table S1. x-axis:  $\log_2$   
6 (fold-change: enrichment); y-axis:  $-\log_{10}$  (p value: significance). NHEJ: non-homologous end  
7 joining. MMR: mismatch repair. NER: nucleotide excision repair. HR: homologous recombination.

8  
9 **4. Discussion**

10 Onset of Huntington’s disease (HD) is closely linked to expansion of CAG triplet repeat in  
11 the *HTT* gene [3]. Importance of the length of CAG triplet repeat to HD onset was demonstrated  
12 [4]. Ways to shorten expanded CAG triplet repeats would be instrumental in the context of  
13 effective therapies towards HD. Along these lines, it was found that a chemical compound,  
14 naphthyridine-azaquinolone (NA), specifically binds to stem-loops composed of CAG triplet

1 repeats but not to canonical dsDNA composed of CAG·CTG repeats. The structure-specific  
2 binding DNA compound NA was demonstrated to lead to contraction of expanded CAG triplet  
3 repeats in HD mice models irrespective of the growing phases (non-dividing or dividing) [16]. Of  
4 interest, such contraction events *in vivo* depend on the presence of functional MutS $\beta$  (a  
5 component of MMR) that recognizes small loops in dsDNA [17]. It was also reported that  
6 introduction of TALEN (a transcription activator-like effector) leads to contraction of expanded  
7 triplet repeats in model yeast strains [18]. For this contraction event, DNA double-strand break  
8 repair (DSBR) related proteins were shown to be indispensable. Therefore, one potential  
9 therapeutic strategy towards alleviating HD is to harness intrinsic DNA repair pathways by  
10 artificially stimulating their activities. Recently, in studies of post-translational levels, cytotoxic  
11 effects of polyglutamine was shown to be dependent upon the presence of a prion, Rnq1, in yeast  
12 model systems [19]. This study indicated that an aggregated, pathogenically toxic form of  
13 polyglutamine may be formed via an active process assisted by a chaperon such as a prion, rather  
14 than via spontaneous formation. This finding may be able to explain the cell-type specificity of HD  
15 occurring in the brain. Taken together, auxiliary factors such as prion may be potential targets for  
16 HD therapy.

17         The main goal of the present paper is to reveal clues that may prove to be instrumental in  
18 the design of novel and effective therapeutic strategies to HD. By utilizing the plasmid DNA pull-  
19 down approach, MS analysis revealed many interesting proteins specifically recruited by the  
20 presence of the stem-loop (Figure 5, Table S1). For example, the following DNA repair-related  
21 proteins were identified: NHEJ-related Ku70/80 and DNA-PKcs [20]; MMR-related MSH2, MSH6  
22 and PMS1 [17]; Fanconi-related FAAP100, FNACC and FANCI [21]; NER/HR/Helicase-related  
23 WRN, BLM, XPF, HUS1 and RMI1 [22–24]. In human cell extracts, factors such as DNA-PKcs and  
24 Ku heterodimer, along with other NHEJ proteins, were previously identified by proteomics as  
25 binding to hairpins in single-stranded DNA [25, 26].

26

1           Some “cSL” specific factors are commonly found in the “gap” containing plasmid (e.g.,  
2 FANCC, MSH6, BLM, WRN); these proteins may be recruited on the stem-loop structure that can  
3 transiently forms in the gapped plasmid (Figure S1, Table S1). When comparing “gap” with “cSL”  
4 (Figure S2, Table S1), some factors are enriched in “cSL” (e.g., Ku80, BRD4). It may indicate that  
5 recruitment of these factors specifically need a stable stem-loop. In any case, the biological  
6 significance of proteins identified by MS as being recruited on given DNA substrate requires further  
7 validation by additional biochemical or genetic assays.

8           A limitation of our approach relates to the fact that we cannot ascertain if all specifically  
9 identified proteins are simultaneously recruited on a single plasmid DNA molecule. Nevertheless,  
10 from the present data set, we can derive molecular events that may occur at the stem-loop as a  
11 working model. For instance, in the case of MMR-related proteins, several models were proposed  
12 in which MMR activities contribute to expansion (or contraction) of triplet repeats [3, 27]. In these  
13 models, MutS $\beta$  (MSH2/MSH3) and MutL $\alpha$  (MLH1/PMS2) access stem-loops in a manner similar  
14 to the way they recognize small DNA loops in regular double-stranded DNA. On the other hand,  
15 we detected MSH2, MSH6 and PMS1 as MMR-related proteins but not MSH3 and PMS2. Namely,  
16 it may indicate that MutS $\beta$  and MutL $\alpha$  do not participate in the early phase of stem-loop recognition  
17 or processing. PMS1 is a component of MutL $\beta$  (MLH1/PMS1) the role of which in MMR is largely  
18 unknown [17]. As a working model, we can speculate that either MutL $\beta$ , PMS1 *per se* or a PMS1-  
19 containing complex associating with or without MutS $\alpha$  (MSH2/MSH6) may be adapted to  
20 recognize structural DNA such as the ones found in stem-loops. Intriguingly, MSH6 and PMS1 are  
21 enriched in “gap” (Figure S1, Table S1). It implies that these protein factors are recruited on  
22 unstable stem-loops that will be formed in an early stage during events of genetic instability at  
23 CAG triplet repeats. Identification of unexpected (or unknown) proteins, thus, stimulates ideas for  
24 designing assays in order to explore new mechanisms.

25

## 26 **5. Patents**

1 R.F. and S.F. hold a patent, 100007204, covering conceptualization and methodology  
2 described in this manuscript.

3  
4 **Supplementary Materials:** The following supporting information can be downloaded at:  
5 ([www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1)), Table S1: Resource data of MS analysis, containing three tabs; Figure  
6 S1: Volcano plot, rcc vs gap; Figure S2: Volcano plot, gap vs cSL.

7  
8 **Author Contributions:** Conceptualization, R.F. and S.F.; methodology, R.F., A.I., J.P. and S.F.;  
9 formal analysis, R.F. and S.F.; data curation, R.F., A.I., J.P. and S.F.; writing—original draft  
10 preparation, S.F.; writing—review and editing, R.F. and S.F. All authors have read and agreed to  
11 the published version of the manuscript.

12  
13 **Funding:** This research was funded by National Institutes of Health, grant number R01  
14 GM132129 to J.P..

15  
16 **Institutional Review Board Statement:** Not applicable.

17  
18 **Informed Consent Statement:** Not applicable.

19  
20 **Data Availability Statement:** All data were provided in the manuscript and Supplementary  
21 Materials.

22  
23 **Acknowledgments:** We thank Johannes Walter (Harvard Medical School, USA) for providing  
24 space, support and materials.

25



1 **Conflicts of Interest:** R.F. and S.F. are co-founders and consultants for bioHalosis, which  
2 provides a custom service of specific nucleoprotein capture used herein. The company had no  
3 role in the design of the study; in the collection, analyses, or interpretation of data; in the writing  
4 of the manuscript, or in the decision to publish the results. All other authors declare no conflict of  
5 interest.

6

7

## 8 **References**

- 9 [1] Haeusler, A. R.; Donnelly, C. J.; Periz, G.; Simko, E. A. J.; Shaw, P. G.; Kim, M.-S.;  
10 Maragakis, N. J.; Troncoso, J. C.; Pandey, A.; Sattler, R.; et al. C9orf72 Nucleotide Repeat  
11 Structures Initiate Molecular Cascades of Disease. *Nature*, **2014**, *507* (7491), 195–200.  
12 <https://doi.org/10.1038/nature13124>.
- 13 [2] Yum, K.; Wang, E. T.; Kalsotra, A. Myotonic Dystrophy: Disease Repeat Range,  
14 Penetrance, Age of Onset, and Relationship between Repeat Size and Phenotypes.  
15 *Current Opinion in Genetics & Development*, **2017**, *44*, 30–37.  
16 <https://doi.org/10.1016/j.gde.2017.01.007>.
- 17 [3] Iyer, R. R.; Pluciennik, A.; Napierala, M.; Wells, R. D. DNA Triplet Repeat Expansion and  
18 Mismatch Repair. *Annu. Rev. Biochem.*, **2015**, *84* (1), 199–226.  
19 <https://doi.org/10.1146/annurev-biochem-060614-034010>.
- 20 [4] Lee, J.-M.; Correia, K.; Loupe, J.; Kim, K.-H.; Barker, D.; Hong, E. P.; Chao, M. J.; Long, J.  
21 D.; Lucente, D.; Vonsattel, J. P. G.; et al. CAG Repeat Not Polyglutamine Length  
22 Determines Timing of Huntington’s Disease Onset. *Cell*, **2019**, *178* (4), 887-900.e14.  
23 <https://doi.org/10.1016/j.cell.2019.06.036>.
- 24 [5] Kovtun, I. V.; Liu, Y.; Bjoras, M.; Klungland, A.; Wilson, S. H.; McMurray, C. T. OGG1  
25 Initiates Age-Dependent CAG Trinucleotide Expansion in Somatic Cells. *Nature*, **2007**, *447*  
26 (7143), 447–452. <https://doi.org/10.1038/nature05778>.
- 27 [6] Isogawa, A.; Fuchs, R. P.; Fujii, S. Versatile and Efficient Chromatin Pull-down  
28 Methodology Based on DNA Triple Helix Formation. *Sci Rep*, **2018**, *8* (1), 5925.  
29 <https://doi.org/10.1038/s41598-018-24417-9>.
- 30 [7] Isogawa, A.; Fuchs, R. P.; Fujii, S. Chromatin Pull-Down Methodology Based on DNA Triple  
31 Helix Formation. In *DNA Electrophoresis*; Hanada, K., Ed.; Methods in Molecular Biology;  
32 Springer US: New York, NY, 2020; Vol. 2119, pp 183–199. [https://doi.org/10.1007/978-1-0716-0323-9\\_16](https://doi.org/10.1007/978-1-0716-0323-9_16).
- 34 [8] Cupello, S.; Richardson, C.; Yan, S. Cell-Free Xenopus Egg Extracts for Studying DNA  
35 Damage Response Pathways. *Int. J. Dev. Biol.*, **2016**, *60* (7-8–9), 229–236.  
36 <https://doi.org/10.1387/ijdb.160113sy>.
- 37 [9] Gillespie, P. J.; Gambus, A.; Blow, J. J. Preparation and Use of Xenopus Egg Extracts to  
38 Study DNA Replication and Chromatin Associated Proteins. *Methods*, **2012**, *57* (2), 203–  
39 213. <https://doi.org/10.1016/j.ymeth.2012.03.029>.
- 40 [10] Wu, R. A.; Semlow, D. R.; Kamimae-Lanning, A. N.; Kochenova, O. V.; Chistol, G.;  
41 Hodskinson, M. R.; Amunugama, R.; Sparks, J. L.; Wang, M.; Deng, L.; et al. TRAIIP Is a  
42 Master Regulator of DNA Interstrand Crosslink Repair. *Nature*, **2019**, *567* (7747), 267–272.  
43 <https://doi.org/10.1038/s41586-019-1002-0>.

- 1 [11] Fujii, S.; Sobol, R. W.; Fuchs, R. P. Double-Strand Breaks: When DNA Repair Events  
2 Accidentally Meet. *DNA Repair*, **2022**, *112*, 103303.  
3 <https://doi.org/10.1016/j.dnarep.2022.103303>.
- 4 [12] Fuchs, R. P.; Isogawa, A.; Paulo, J. A.; Onizuka, K.; Takahashi, T.; Amunugama, R.; Duxin,  
5 J. P.; Fujii, S. Crosstalk between Repair Pathways Elicits Double-Strand Breaks in Alkylated  
6 DNA and Implications for the Action of Temozolomide. *eLife*, **2021**, *10*, e69544.  
7 <https://doi.org/10.7554/eLife.69544>.
- 8 [13] Richard, G.-F.; Kerrest, A.; Dujon, B. Comparative Genomics and Molecular Dynamics of  
9 DNA Repeats in Eukaryotes. *Microbiol Mol Biol Rev*, **2008**, *72* (4), 686–727.  
10 <https://doi.org/10.1128/MMBR.00011-08>.
- 11 [14] Walter, J.; Sun, L.; Newport, J. Regulated Chromosomal DNA Replication in the Absence  
12 of a Nucleus. *Molecular Cell*, **1998**, *1* (4), 519–529. [https://doi.org/10.1016/S1097-  
13 2765\(00\)80052-0](https://doi.org/10.1016/S1097-2765(00)80052-0).
- 14 [15] Lebofsky, R.; Takahashi, T.; Walter, J. C. DNA Replication in Nucleus-Free *Xenopus* Egg  
15 Extracts. In *DNA Replication*; Vengrova, S., Dalgaard, J. Z., Eds.; Methods in Molecular  
16 Biology; Humana Press: Totowa, NJ, 2009; Vol. 521, pp 229–252.  
17 [https://doi.org/10.1007/978-1-60327-815-7\\_13](https://doi.org/10.1007/978-1-60327-815-7_13).
- 18 [16] Nakamori, M.; Panigrahi, G. B.; Lanni, S.; Gall-Duncan, T.; Hayakawa, H.; Tanaka, H.; Luo,  
19 J.; Otabe, T.; Li, J.; Sakata, A.; et al. A Slipped-CAG DNA-Binding Small Molecule Induces  
20 Trinucleotide-Repeat Contractions in Vivo. *Nat Genet*, **2020**, *52* (2), 146–159.  
21 <https://doi.org/10.1038/s41588-019-0575-8>.
- 22 [17] Jiricny, J. The Multifaceted Mismatch-Repair System. *Nat Rev Mol Cell Biol*, **2006**, *7* (5),  
23 335–346. <https://doi.org/10.1038/nrm1907>.
- 24 [18] Mosbach, V.; Poggi, L.; Viterbo, D.; Charpentier, M.; Richard, G.-F. TALEN-Induced  
25 Double-Strand Break Repair of CTG Trinucleotide Repeats. *Cell Reports*, **2018**, *22* (8),  
26 2146–2159. <https://doi.org/10.1016/j.celrep.2018.01.083>.
- 27 [19] Gropp, M. H. M.; Klaips, C. L.; Hartl, F. U. Formation of Toxic Oligomers of PolyQ-Expanded  
28 Huntingtin by Prion-Mediated Cross-Seeding. *Molecular Cell*, **2022**, *82* (22), 4290-  
29 4306.e11. <https://doi.org/10.1016/j.molcel.2022.09.031>.
- 30 [20] Scully, R.; Panday, A.; Elango, R.; Willis, N. A. DNA Double-Strand Break Repair-Pathway  
31 Choice in Somatic Mammalian Cells. *Nat Rev Mol Cell Biol*, **2019**, *20* (11), 698–714.  
32 <https://doi.org/10.1038/s41580-019-0152-0>.
- 33 [21] Semlow, D. R.; Walter, J. C. Mechanisms of Vertebrate DNA Interstrand Cross-Link Repair.  
34 *Annu. Rev. Biochem.*, **2021**, *90* (1), 107–135. [https://doi.org/10.1146/annurev-biochem-  
35 080320-112510](https://doi.org/10.1146/annurev-biochem-080320-112510).
- 36 [22] Datta, A.; Brosh, R. M. New Insights Into DNA Helicases as Druggable Targets for Cancer  
37 Therapy. *Front. Mol. Biosci.*, **2018**, *5*, 59. <https://doi.org/10.3389/fmolb.2018.00059>.
- 38 [23] Compe, E.; Egly, J.-M. Nucleotide Excision Repair and Transcriptional Regulation: TFIIH  
39 and Beyond. *Annu. Rev. Biochem.*, **2016**, *85* (1), 265–290. [https://doi.org/10.1146/annurev-  
40 biochem-060815-014857](https://doi.org/10.1146/annurev-biochem-060815-014857).
- 41 [24] Mehta, A.; Haber, J. E. Sources of DNA Double-Strand Breaks and Models of  
42 Recombinational DNA Repair. *Cold Spring Harbor Perspectives in Biology*, **2014**, *6* (9),  
43 a016428–a016428. <https://doi.org/10.1101/cshperspect.a016428>.
- 44 [25] Uliel, L.; Weisman-Shomer, P.; Oren-Jazan, H.; Newcomb, T.; Loeb, L. A.; Fry, M. Human  
45 Ku Antigen Tightly Binds and Stabilizes a Tetrahelical Form of the Fragile X Syndrome  
46 d(CGG) Expanded Sequence. *Journal of Biological Chemistry*, **2000**, *275* (42), 33134–  
47 33141. <https://doi.org/10.1074/jbc.M005542200>.
- 48 [26] Janel-Bintz, R.; Kuhn, L.; Frit, P.; Chicher, J.; Wagner, J.; Haracska, L.; Hammann, P.;  
49 Cordonnier, A. M. Proteomic Analysis of DNA Synthesis on a Structured DNA Template in  
50 Human Cellular Extracts: Interplay Between NHEJ and Replication-Associated Proteins.  
51 *Proteomics*, **2020**, *20* (3–4), 1900184. <https://doi.org/10.1002/pmic.201900184>.

1 [27] Usdin, K.; House, N. C. M.; Freudenreich, C. H. Repeat Instability during DNA Repair:  
2 Insights from Model Systems. *Critical Reviews in Biochemistry and Molecular Biology*, **2015**,  
3 50 (2), 142–167. <https://doi.org/10.3109/10409238.2014.999192>.  
4