



HAL
open science

Why does the pUG tail curl?

David Monchaud

► **To cite this version:**

David Monchaud. Why does the pUG tail curl?. *Molecular Cell*, 2023, 83 (3), pp.330-331.
10.1016/j.molcel.2022.12.010 . hal-03971075

HAL Id: hal-03971075

<https://hal.science/hal-03971075>

Submitted on 3 Feb 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.

Why does the pUG tail curl?

David Monchaud*

Institut de Chimie Moléculaire, ICMUB CNRS UMR 6302, UBFC Dijon, France.

*Correspondence to: david.monchaud@cnr.fr

Roschdi et al.¹ report on a new higher-order RNA structure folding from an alternating uridine (U)/guanosine (G) repeated sequence—the pUG tail—into a peculiar G-quadruplex structure—the pUG fold—found to orchestrate the gene-silencing activity of pUGylated RNAs.

G-quadruplexes (G4s) are everywhere. These four-stranded nucleic acid structures fold from guanine (or guanosine, G)-rich sequences, as a result of the self-assembly of Gs to form G-quartets, and the self-stacking of G-quartets to form the thermodynamically stable G4 core (Figure 1A,B). G4s can be found in both genomes (G4-DNA)² and transcriptomes (G4-RNA)³ of a variety of species including parasites (*e.g.*, *Leishmania*), bacteria (*e.g.*, *Rhodobacter*), nematodes (*e.g.*, *C. elegans*) plants and humans.⁴ They are located in either coding regions (*e.g.*, gene bodies for G4-DNA, coding sequences (CDS) of mRNA) or non-coding regions (*e.g.*, intronic DNA, UTRs of mRNAs, ncRNAs), thus exerting repressive or activating cellular roles depending on their location, mostly studied in the general context of gene expression regulation⁵ and G4-mediated genetic instability.⁶ G4s are now considered as key players in genetics because of their genomic and transcriptomic prevalence: hundreds of thousands of G4-prone DNA and RNA sequences have been detected *in silico*⁷ and hundreds of them shown to fold into G4s *in vitro* and *in vivo*.^{2,3} By their nature, these sequences display distinctive G4 signatures, *i.e.*, stretches of consecutive Gs (G-runs, usually between 2 to 5 Gs) separated by diverse intervening sequences (G4 loops, between 0 to >20 nucleotides), which translates into topological diversity. All known G4-forming sequences share the common feature of exhibiting G-runs, which makes them readily detectable (by predictive algorithms) and identifiable (by sequencing). All of them... at least, until only very recently.¹

In 2020, Kennedy and coworkers⁸ described in *C. elegans* an enzyme, RDE-3 (also known as MUT-2), which adds a series of uridine (U)/guanosine (G) dinucleotides to the 3' end of RNAs, thus decorating them with a poly(UG) tail, also referred to as a pUG tail. A tail with >8 UG pairs

was found to be functional, which led the authors to surmise that “*the precise length and sequence requirements for pUG tail function suggest that long pUG tail may (...) form a structure that helps to recruit, a possibly prime, RdRPs [RNA-dependent RNA polymerases]*”.⁸ Indeed, pUG tails were thought to act as docking sites for RdRPs, which then use them as a template for short interference RNA (siRNA) production, subsequently involved in gene silencing pathways. Normally, G-rich RNA sequences such as the pUG tails would have promptly been spotted as likely G4-folding sequences; however, none of the reported G4-forming sequences display interrupted G-runs as pUG tails do, making them poor G4 candidates. All predictive algorithms would have set them aside: as an example, the 12-unit pUG tail (UG)₁₂ has a G4Hunter score of 0.5, far from the 1-1.2 threshold below which G4 folding is considered unreliable.⁹ In spite of this, Kennedy, Butcher and coworkers¹ used a series of biophysical methods to establish the G4 signature of pUG tails both in solution (CD, TDS, 1D- and 2D-NMR) and in the solid state, resolving the X-ray crystallographic structure of (UG)_{11.5} and (UG)₁₂ in complex with the G4 ligands N-methyl-mesoporphyrin IX (NMM) (Figure 1C). Beyond their alternating Gs, the pUG folds differ from classical G4 structure by the presence of a U-quartet beneath the G4 core and, above all, the inverted helicity (left-handed) of the overall structure, thus adding another type of topology to an already rich portfolio of G4 structures, even for G4-RNA.¹⁰

When considering the biological function of this pUG structure, the authors show that RDE-3 adds pUG tails to various RNAs in *C. elegans*; this pUGylation offers RdRPs a docking site (the pUG fold) from which they synthesize gene-silencing siRNAs (Figure 1D). While the direct interaction between RdRPs and pUG fold is still to be confirmed, strong lines of evidence support the pUG fold involvement, notably the demonstration that 1) 12 or more UG repeats are required to exert siRNA-mediated gene-silencing activity, and 2) the binding of NMM alters this ability, which strongly suggests that the G4 ligand, by stacking atop the accessible G-quartet of pUG fold, hinders proper interaction between the enzyme (RdRP) and its substrate (pUG fold). The authors also suggest that pUGylated RNAs are vectors for transgenerational epigenetic inheritance, allowing *C. elegans* to remember past gene-silencing landscapes and pass it on to the next generation. This feed-forward and transgenerational pUG-mediated gene-silencing mechanism is appealing but cannot yet be suggested in humans as our cells do not possess RdRPs, which makes the functional relevance

of pUG RNAs to mammalian biology questionable. However, pUG fold motifs are particularly enriched in both *C. elegans* and mammalian introns, with >20,000 pUG-folds in the human genome, >95% of which are located in introns. These sequences are particularly found near splice sites, hinting towards key roles in splicing processes, but pUG folds are also present in ncRNA (>300 motifs) and mRNA (>500 motifs), which suggests a broader, yet-to-be-determined cellular role. Deciphering these new roles will require additional effort but this study provides a new G4 folding pattern, which leads G4 scientists to rethink their predictive algorithms as the already high transcriptomic and genomic prevalence of G4s might turn out to be even higher than anticipated. As a result, G4 biology, which already stands out for its fine-lace intricacy, still holds a lot of surprises.

Declaration of Interests

None declared.

References

1. Roschdi, S., Yan, J., Nomura, Y., Escobar, C.A., Petersen, R.J., Bingman, C.A., Tonelli, M., Vivek, R., Montemayor, E.J., Wickens, M., *et al.* (2022). An atypical RNA quadruplex marks RNAs as vectors for gene silencing. *Nat. Struct. Mol. Biol.* *29*, 1113-1121.
2. Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. *Trends Chem.* *2*, 123-136.
3. Lyu, K., Chow, E.Y.-C., Mou, X., Chan, T.-F., and Kwok, Chun K. (2021). RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Res.* *49*, 5426-5450.
4. Marsico, G., Chambers, V.S., Sahakyan, A.B., McCauley, P., Boutell, J.M., Antonio, M.D., and Balasubramanian, S. (2019). Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* *47*, 3862-3874.
5. Robinson, J., Raguseo, F., Nuccio, S.P., Liano, D., and Di Antonio, M. (2021). DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* *49*, 8419-8431.
6. Zell, J., Rota Sperti, F., Britton, S., and Monchaud, D. (2021). DNA folds threaten genetic stability and can be leveraged for chemotherapy. *RSC Chem. Biol.* *2*, 47-76.
7. Puig Lombardi, E., and Londoño-Vallejo, A. (2019). A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* *48*, 1-15.

8. Shukla, A., Yan, J., Pagano, D.J., Dodson, A.E., Fei, Y., Gorham, J., Seidman, J.G., Wickens, M., and Kennedy, S. (2020). poly(UG)-tailed RNAs in genome protection and epigenetic inheritance. *Nature* 582, 283-288.
9. Bedrat, A., Lacroix, L., and Mergny, J.-L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746-1759.
10. Banco, M.T., and Ferré-D'Amaré, A.R. (2021). The emerging structural complexity of G-quadruplex RNAs. *RNA* 27, 390-402.

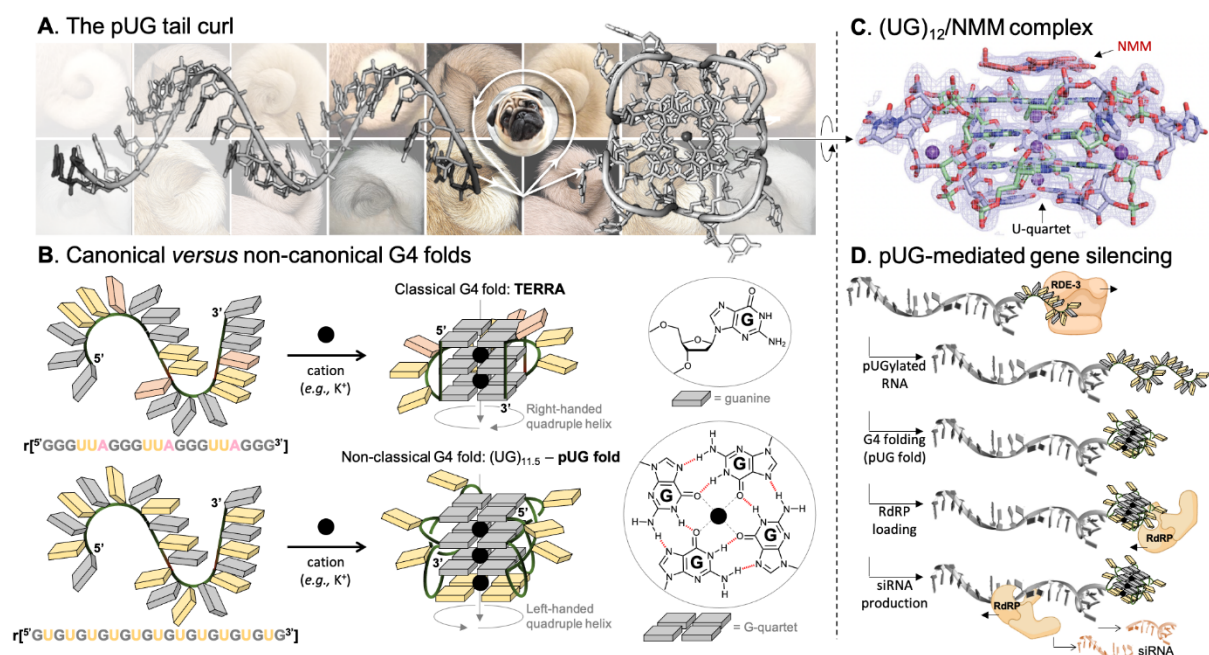


Figure 1. **A,B.** Stylized (**A**) and schematic representation (**B**) of the way the pUG tail folds upon itself to form a G-quadruplex structure, along with the comparison of classical (TERRA) versus non-classical (pUG fold) G4 folding. **C.** X-ray crystal structure of pUG fold $((UG)_{12})$ /N-methylmesoporphyrin (NMM) complex (1.97 Å-resolution). **D.** Gene-silencing mechanism involving the pUGylation of RNAs by RDE-3 which allows the pUG tails (>12 repeats) to fold into a G4 structure (pUG fold) that acts as a docking site for RNA-dependent RNA polymerases (RdRPs), which then used the pUGylated RNAs as a template for short interference RNA (siRNA) synthesis.