MASARYKOVA UNIVERZITA

PŘÍRODOVĚDECKÁ FAKULTA

Analýza DNA vazebných proteinů se zaměřením na jejich interakce s lokálními strukturami DNA

Disertační práce

NATÁLIA VALKOVÁ

Vedoucí práce: prof. Václav Brázda, Ph.D.

Ústav experimentálnej biológie
Program Molekulární a buněčná biologie a genetika

Brno 2023

Bibliografický záznam

Autor: Natália Valková
Přírodovědecká fakulta
Masarykova univerzita
Ústav experimentálnej biológie

Název práce: Analýza DNA vazebných proteinů se zaměřením na jejich interakce s lokálními strukturami DNA

Studijní program: Molekulární a buněčná biologie a genetika

Vedoucí práce: prof. Václav Brázda, Ph.D.

Rok: 2023

Počet stran: 152

Klíčová slova: DNA, protein, virus, IFI16, p53, G-kvadruplex, G4Hunter, inverzní repetice

Bibliografický záznam

Autor: Natália Valková
Prírodovedecká fakulta
Masarykova univerzita
Ústav experimentálnej biológie

Názov práce: Analýza DNA väzbových proteínov so zameraním na ich interakcie s lokálnymi štruktúrami DNA

Študijný program: Molekulární a buněčná biologie a genetika

Vedúci práce: prof. Václav Brázda, Ph.D.

Rok: 2023

Počet strán: 152

Kľúčové slová: DNA, proteín, vírus, IFI16, p53, G-kvadruplex, G4Hunter, inverzná repetícia

Bibliographic record

Author: Natália Valková
Faculty of Science
Masaryk University
Department of Experimental Biology

Title of Thesis: Analysis of DNA binding proteins with a focus on their interaction with local DNA structures

Degree Programme: Molecular and Cell Biology and Genetics

Supervisor: prof. Václav Brázda, Ph.D.

Year: 2023

Number of Pages: 152

Keywords: DNA, protein, virus, IFI16, p53, G-quadruplex, G4Hunter, inverted repeat

Abstrakt

Interakce proteinů s nukleovými kyselinami je klíčovou vlastností při zásadních molekulárních dějích, jakými jsou replikace, genová exprese, oprava DNA či imunitní reakce. Dizertační práce je zaměřena na interakce dvou DNA-vazebných proteinů, senzoru cizorodé DNA interferonem γ indukovatelného proteinu 16 (IFI16) a transkripčního faktoru proteinu p53, s G-kvadruplexy (G4) a inverzními repeticemi. G4 a inverzní repetice byly detekovány ve všech doménách života a mají v genomu regulační úlohu. Představují nadějné terapeutické cíle v nádorové biologií, při různých virových infekcích a neurodegenerativních onemocněních.

Interakce proteinu p53 byly testovány prostřednictvím izogenního kvasinkového systému se sekvenci s potenciálem tvorby G4 začleněnou před promotorem reportérového genu. G4 způsoboval represi transaktivace proteinem p53 a moduloval interakci proteinu p53 s jeho izoformami, proteiny jeho rodiny a jejich mutantními variantami. Topologie DNA v okolí responzívní sekvenci p53 tak představuje další úroveň v regulaci signálních drah, do kterých je tento transkripční faktor zapojený.

Interakce IFI16 s DNA byly zkoumány pomocí mikroskopie atomárních sil. Typ pozorované interakci závisel na superhelicitě DNA. Lineární DNA indukovala tvorbu proteinových oligomerů, zatímco superhelikální DNA tvorbu globulárních komplexů proteinů s DNA o nižší molekulové hmotnosti. Vzhledem k tomu, že tvorba proteinového oligomeru je klíčová pro aktivaci nespecifické imunitní reakce proteinu IFI16, specifické interakce se superhelikální DNA mohou představovat jiný funkční mechanizmus zajišťující represi genové exprese proteinem IFI16.

Cílem druhé části dizertační práce bylo pomocí bioinformatických metod identifikovat skupiny vírů a cílové oblasti lidského genomu zajímavé z hlediska studia proteinů interagujících s DNA. Analýza 11000 virových genomů odhalila, že průměrná frekvence sekvence s potenciálem tvorby G4 je vyšší u virů, které mají tendenci způsobovat perzistentní infekce. Studie dále odhalila korelaci mezi frekvencí výskytu sekvencí s potenciálem tvorby G4 ve virových a hostitelských genomech, což poukazuje na evoluční konzervovanost G4. Pro výskyt G4 a inverzních repetic byly zkoumány rovněž dva chromozomy lidského genomu, jejichž kompletní sekvence byla zveřejněna v roce 2020.

Abstrakt

Interakcia proteínov s nukleovými kyselinami je kľúčovou vlastnosťou pri zásadných molekulárnych dejoch, akými sú replikácia, génová expresia, oprava DNA či imunitná reakcia. Dizertačná práca je zameraná na interakcie dvoch DNA-väzbových proteínov, senzora cudzorodej DNA interferónom γ indukovateľného proteínu 16 (IFI16) a transkripčného faktoru proteínu p53, s G-kvadruplexami (G4) a inverznými repetíciami. G4 a inverzné repetície boli detegované vo všetkých doménach života a majú v genóme regulačnú úlohu. Predstavujú nádejné terapeutické ciele v nádorovej biológií, pri rôznych vírusových infekciách a neurodegeneratívnych ochoreniach.

Interakcie proteínu p53 boli skúmané prostredníctvom izogénneho kvasinkového systému so sekvenciou s potenciálom tvorby G4 začlenenou pred promótorom reportérového génu. G4 spôsoboval represiu transaktivácie proteínom p53 a moduloval interakcie proteínu p53 s jeho izoformami, proteínmi jeho rodiny a ich mutantnými variantami. Topológia DNA v okolí responzívnej sekvencie p53 tak predstavuje ďalšiu úroveň v regulácii signálnych dráh, do ktorých je tento transkripčný faktor zapojený.

Interakcie IFI16 s DNA boli skúmané pomocou mikroskopie atomárnych síl. Typ pozorovanej interakcie závisel od superhelicity DNA. Lineárna DNA indukovala tvorbu proteínových oligomérov, zatiaľ čo superhelikálna DNA tvorbu globulárnych komplexov proteínu s DNA o nižšej molekulovej hmotnosti. Keďže tvorba proteínového oligoméru je kľúčovou pre aktiváciu nešpecifickej imunitnej reakcie proteínom IFI16, špecifické interakcie so superhelikálnou DNA môžu predstavovať iný funkčný mechanizmus zabezpečujúci represiu génovej expresie proteínom IFI16.

Cieľom druhej časti dizertačnej práce bolo pomocou bioinformatických metód identifikovať skupiny vírusov a cieľové oblasti ľudského genómu zaujímavé z hľadiska štúdia proteínov interagujúcich s DNA. Analýza 11000 vírusových genómov odhalila, že priemerná frekvencia sekvencií s potenciálom tvorby G4 je vyššia u vírusov, ktoré majú tendenciu spôsobovať perzistentné infekcie. Štúdia ďalej odhalila koreláciu medzi frekvenciou výskytu sekvencií s potenciálom tvorby G4 vo vírusových a hostiteľských genómoch, čo poukazuje na evolučnú konzervovanosť G4. Pre výskyt G4 a inverzných repetíc boli skúmané aj dva chromozómy ľudského genómu, ktorých kompletná sekvencia bola zverejnená v roku 2020.

Abstract

The interaction of proteins with nucleic acids is a key property of fundamental molecular processes, such as replication, gene expression, DNA repair, or immune response. The dissertation thesis is focused on the interactions of two DNA-binding proteins, the foreign DNA sensor interferon γ inducible protein 16 (IFI16) and the transcription factor p53, with G-quadruplexes (G4) and inverted repeats. G4 and inverted repeats were detected in all domains of life and have a regulatory role in the genome. They represent promising therapeutic targets in tumor biology, various viral infections, and neurodegenerative diseases.

Interactions of the p53 protein were investigated using an isogenic yeast reporter system with a putative G4-forming sequence inserted upstream of the reporter gene promoter. G4 inhibited the p53-mediated transactivation and modulated p53 interactions with its isoforms, its family proteins, and its mutant variants. The DNA topology around the responsive element of p53 thus represents another level in the regulation of signaling pathways in which this transcription factor is involved.

Interactions of IFI16 with DNA were investigated using atomic force microscopy. The type of interaction observed depended on the superhelicity of the DNA. Linear DNA induced the formation of protein oligomers, while superhelical DNA induced the formation of globular complexes of protein with DNA of lower molecular weight. Since the formation of a protein oligomer is crucial for the activation of a non-specific immune response by the IFI16 protein, specific interactions with superhelical DNA may represent a different functional mechanism, which could result in the repression of gene expression mediated by IFI16 protein.

The second part of the dissertation aimed to use bioinformatics methods to identify groups of viruses and target areas of the human genome which could be interesting in the studies of DNA-interacting proteins. An analysis of 11,000 viral genomes revealed higher frequency of sequences with the potential to form G4 in viruses that tend to cause persistent infections. The study further revealed a correlation between the frequency of occurrence of sequences with G4-forming potential in viral and host genomes, indicating tuhe evolutionary conservation of G4. Two chromosomes of the human genome were also examined for the occurrence of G4 and inverted repeats, the complete sequence of which was published in 2020.

Čestné prohlášení

Prohlašuji, že jsem svoji disertační práci vypracovala samostatně pod vedením vedoucího práce s využitím informačních zdrojů, které jsou v práci citovány.

V Brně 11. června 2023

.....
Natália Valková

Poděkování

Na tomto mieste by som sa chcela v prvom rade poďakovať môjmu školiteľovi, prof. Václavovi Brázdovi, Ph.D. za jeho cenné rady, nápady, podporu a ústretovosť.

Ďalej nemôžem nespomenúť Mgr. Jana Coufala, Ph.D. za trpezlivé odpovede na moje nekonečné otázky a veľkú pomoc pri realizácii experimentov.

Veľká vďaka patrí aj mojím súputníkom, Ing. Otílii Porubiakovej, Ph.D., MSc. Alessiovi Cantarovi, Ph.D., Ing. Ondrejovi Heskovi a MSc. Patrikovi Goswamimu, Ph.D. za zdieľanie radostí a strastí PhD života.

Ďalej by som sa chcela poďakovať celému oddeleniu Biofyzikálnej chémie a molekulárnej onkológie za podporu, príjemnú a priateľskú atmosféru počas môjho pôsobenia na tomto pracovisku.

Moja vďaka patrí aj všetkým pregraduálnym študentom, ktorých som mala možnosť počas môjho štúdia stretnúť, za ich záujem, tvrdú prácu a ich otázky, ktoré mi vždy dokázali koľko sa toho sama musím ešte naučiť.

Veľká vďaka patrí mojím rodičom za nekonečnú podporu a bezpodmienečnú vieru v to, že všetko zvládnem.

A v neposlednom rade môjmu manželovi Jozefovi Valkovi, s ktorým je všetko a vždy jednoducho lepšie.

Originální publikace a vymezení podílu autora dizertační práce

Práca vychádza z publikácií autora dizertačnej práce (Natália Valková, za slobodna Bohálová, NV).

Publikácia 1. (Príloha A. tejto dizertačnej práce)

Porubiaková O, **Bohálová N**, Inga A, et al. The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms. Int J Mol Sci. 2019;21(1). doi:10.3390/ijms21010127

NV vytvorila kvasinkové kmene použité v práci, podieľala sa na príprave plazmidov exprimujúcich izoformy p53, podieľala sa na transaktivačných experimentoch a príprave manuscriptu. Celkový podiel na príprave článku predstavuje 35 %. S koautorom OP NV zdieľa prvoautorstvo.

Publikácia 2. (Príloha B. tejto dizertačnej práce)

Monti P, Brazda V, **Bohálová N**, et al. Evaluating the Influence of a G-Quadruplex Prone Sequence on the Transactivation Potential by Wild-Type and/or Mutant P53 Family Proteins through a Yeast-Based Functional Assay. Ge-nes. 2021;12(2):277. doi:10.3390/genes12020277

NV vytvorila kvasinkové kmene použité v práci a podieľala sa na príprave manuscriptu. Celkový podiel na príprave článku je 10 %.

Publikácia 3. (Príloha C. tejto dizertačnej práce)

Valková N, et al. Dual mode of IFI16 binding to supercoiled and linear DNA: a closer insight

NV sa podieľala na experimentoch a príprave manuscriptu. Celkový podiel na príprave článku je 25 %.

Publikácia 4. (Príloha D. tejto dizertačnej práce)

Bohálová N, Mergny JL, Brázda V. Novel G-quadruplex prone sequences emerge in the complete assembly of the human X chromosome. *Biochimie*. 2021;191:87-90.doi:10.1016/j.biochi.2021.09.004

NV analyzovala nový chromozóm X, pripravila finálne obrázky a tabuľky a podieľala sa na písaní manuscriptu. Celkový podiel na príprave článku predstavuje 40 %.

Publikácia 5. (Príloha E. tejto dizertačnej práce)

Brazda V, **Bohalova N**, Bowater RP. New telomere to telomere assembly of human chromosome 8 reveals a previous underestimation of G-quadruplex forming sequences and inverted repeats. *Gene*. Published online November 1, 2021:146058. doi:10.1016/j.gene.2021.146058

NV analyzovala výskyt sekundárnych lokálnych štruktúr chromozómu 8, pripravila finálne obrázky a tabuľky a podieľala sa na písaní manuscriptu. Celkový podiel na príprave práce predstavuje 30 %.

Publikácia 6. (Príloha F. tejto dizertačnej práce)

Bohálová N, Cantara A, Bartas M, et al. Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. *Biochimie*. 2021;186:13-27. doi:10.1016/j.biochi.2021.03.017

NV v spolupráci s koautorom AC, s ktorým zdieľa prvoautorstvo práce, analyzovala 11 000 vírusových sekvencií, postulovala hlavnú hypotézu, pripravila finálne obrázky a tabuľky a podieľala sa na písaní manuscriptu. Celkový podiel na príprave článku je 35 %.

Publikácia 7. (Príloha G. tejto dizertačnej práce)

Bohálová N, Cantara A, Bartas M, et al. Tracing dsDNA Virus–Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences. Int J Mol Sci. 2021;22(7):3433. doi:10.3390/ijms22073433

NV uskutočnila hlavné štatistické analýzy, pripravila finálne obrázky a tabuľky a podieľala sa na písaní manuscriptu. Celkový podiel na príprave článku je 60 %.

Obsah

Zoznam obrázkov 19

Zoznam tabuliek 20

Zoznam pojmov a skratiek 21

Zoznam príloh 23

1 Úvod a teória 25

1.1 Sekundárne lokálne štruktúry nukleových kyselín 26

1.1.1 G-kvadruplex 26

1.1.2 Inverzné repetície 28

1.1.3 Ostatné sekundárne lokálne štruktúry nukleových kyselín 29

1.2 Sekundárne lokálne štruktúry a ich možné klinické aplikácie v humánnej medicíne. 30

1.2.1 Sekundárne lokálne štruktúry v ľudskom genóme 30

1.2.2 Sekundárne lokálne štruktúry vo vírusových genómoch 33

1.3 Interakcie proteínov so sekundárnymi lokálnymi štruktúrami nukleových kyselín 35

1.3.1 Väzba proteínov na nukleové kyseliny 35

1.3.2 Protein p53 36

1.3.3 Protein IFI16 38

1.4 Bioinformatické nástroje na analýzu sekundárnych lokálnych štruktúr 41

1.4.1 Bioinformatické nástroje založené na regulárnych výrazoch 41

1.4.2 Skórovacie metódy založené na regulárnych výrazoch 42

1.4.3 Metódy založené na prehľadávacom posúvnom okne a skórovaní 42

1.4.4 Orthogonálny systém 42

1.4.5 G4seq 43

1.4.6 Neurónové siete 43

1.4.7 Databázy 43

2 Ciele dizertačnej práce 44

3 Publikované výsledky 45

3.1 Publikácia 1. Vplyv G-kvadruplexu v blízkosti cieľových sekvencií p53 na transaktivačný potenciál izoforiem p53α 45

3.2 Publikácia 2. Hodnotenie vplyvu sekvencie s potenciálom tvorby G-kvadruplexu na transaktivačný potenciál proteínov z rodiny p53 prostredníctvom kvasinkového transaktivačného systému 47

3.3 Publikácia 3. Bližší pohľad na duálny režim väzby IFI16 na superhelikálnu a lineárnu DNA 48

3.4 Publikácia 4. V kompletnej sekvencii ľudského chromozómu X sa vyskytli predtým neidentifikované sekvencie s potenciálom tvorby G-kvadruplexu50

3.5 Publikácia 5. Kompletná sekvencia ľudského chromozómu 8 odhalila predchádzajúce podcenenie sekvencií s potenciálom tvorby G-kvadruplexu a inverzných repetícií 52

3.6 Publikácia 6. Analýza sekvencií potenciálne formujúcich G-kvadruplexy vo virusových genómoch odhalila ich koreláciu s typom infekcie 53

3.7 Publikácia 7. Pozorovanie koevolúcie dsDNA vírusov a ich hostiteľov skrz koreláciu sekvencii potenciálne formujúcich G-kvadruplexy 56

4 Záver 58

Použité zdroje 60

Přílohy 79

Příloha A The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms 80

Příloha B Evaluating the Influence of a G-Quadruplex Prone Sequence on the Transactivation Potential by Wild-Type and/or Mutant P53 Family Proteins through a Yeast-Based Functional Assay 93

Příloha C Dual mode of IFI16 binding to supercoiled and linear DNA: a closer insight 106

Příloha D Novel G-quadruplex prone sequences emerge in the complete assembly of the human X chromosome 113

Příloha E New telomere to telomere assembly of human chromosome 8 reveals a previous underestimation of G-quadruplex forming sequences and inverted repeats 118

Příloha F Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection 124

Příloha G Tracing dsDNA Virus-Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences 140

Zoznam obrázkov

Obr. 1: Štruktúra G4. 27
Obr. 2: Krížová štruktúra 28
Obr. 3: Reportérové kmene použité v transaktivačných experimentoch. 45
Obr. 4: Porovnanie transaktivačnej aktivity p53 v 4 kvasinkových kmeňoch a v koexpresii s N-koncovými izoformami. 46
Obr. 5: Mikroskopia atomárnych síl komplexov proteínu IFI16 a DNA. 48
Obr. 6: Porovnanie výskytu PQS v chromozóme X genómu GRCh38 a nového assembly T2T. 51
Obr. 7: Porovnanie výskytu PQS a inverzních repetíc v chromozóme 8 genómu GRCh38 a nového assembly T2T. 52
Obr. 8: Prehľad PQS frekvencií vo vírusových čeľadiach. 54
Obr. 9: Porovnanie frekvencií PQS vírusov spôsobujúcich akútne a latentné infekcie. 55
Obr. 10: Korelácia PQS frekvencií medzi vírusom a jeho hostiteľom. 57

Zoznam tabuliek

Tab. 1: Prehľad vírusov, pri ktorých prebieha výskum v spojení s G4 a skúmané ligandy 33

Zoznam pojmov a skratiek

AFM	–	mikroskopia atomárnych síl (z angl. atomic force microscopy)
bp	–	párov báz
CRISPR	–	skupina krátkych palindromatických opakovaní, ktoré sú pravidelne prerušované intervenujúcou sekvenciou (z angl. clustered regularly interspaced short palindromic repeats)
dsDNA	–	dvojvláknová DNA
EBNA1	–	antigén 1 vírusu Epsteina a Barrovej
EBV	–	vírus Epsteina a Barrovej
G4	–	G-kvadruplex
GOF	–	mutácia spôsobujúca novú funkciu génového produktu (z angl. gain of function mutation)
HAV	–	vírus hepatitídy A
HBV	–	vírus hepatitídy B
HCV	–	vírus hepatitídy C
HIV-1	–	ľudský vírus imunitnej nedostatočnosti 1 (z angl. human immunodeficiency virus)
HSV-1	–	herpes simples vírus 1
ICTV	–	medzinárodná komisia taxonómie vírusov (z angl. International Committee on Taxonomy of Viruses)
IFI16	–	Interferónom γ indukovateľný proteín 16
IR	–	inverzná repetícia
KSHV	–	herpesvírus Kaposiho sarkómu
NCBI	–	národné centrum pre biotechnologické informácie (z angl. national center for biotechnology information)
nt	–	nukleotid
p53	–	proteín 53
PRR	–	prirodzené rozpoznávacie receptory (z angl. innate pattern recognition receptors)

PQS	– sekvencie potenciálne tvoriace G4
RE	– responzívny element
ssDNA	– jednovláknová DNA
T2T	– konzorcium od teloméry po teloméru
wt	– kanonický proteín štandardného „divokého“ typu (z angl. wild-type)

Zoznam príloh

Prílohy v textu

Příloha A The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms 80

Příloha B Evaluating the Influence of a G-Quadruplex Prone Sequence on the Transactivation Potential by Wild-Type and/or Mutant P53 Family Proteins through a Yeast-Based Functional Assay 93

Příloha C Dual mode of IFI16 binding to supercoiled and linear DNA: a closer insight 106

Příloha D Novel G-quadruplex prone sequences emerge in the complete assembly of the human X chromosome 113

Příloha E New telomere to telomere assembly of human chromosome 8 reveals a previous underestimation of G-quadruplex forming sequences and inverted repeats 118

Příloha F Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection 124

Příloha G Tracing dsDNA Virus-Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences 140

1 Úvod a teória

Tri molekuly centrálnej molekulárnej dogmy – DNA, RNA a proteín tvoria základný kameň života. Pre funkčný celok sú ale nevyhnutné práve väzby a interakcie medzi týmito molekulami. Práve tie umožňujú reguláciu komplikovaných molekulárnych procesov, akými sú replikácia, génová expresia, bunková signalizácia, obrana pred patogénmi či oprava DNA. Interakcie týchto molekúl sú závislé nie len na primárnom chemickom zložení danej molekuly, ale aj na vyšších štruktúrach konkrétnych interakčných partnerov, či už ide o proteín alebo nukleovú kyselinu.

Cieľom tejto dizertačnej práce je výskum úlohy sekundárnych lokálnych štruktúr nukleových kyselín pri interakciách s proteínmi. Zatiaľ čo štruktúra proteínov je predmetom záujmu mnohých výskumných skupín a jej objasnenie viedlo k mnohým zásadným objavom vrátane vývoja viacerých liečiv, dlho nebolo známe či sú sekundárne lokálne štruktúry nukleových kyselín iba artefakty vyskytujúce sa výlučne *in vitro*, alebo sú to fyziologicky relevantné zmeny týchto molekúl. S vývojom technológií sa ich existencia potvrdila aj *in vivo* v mnohých organizmoch. Ich biologický význam sa stále skúma, ale dnes už je známe, že tvoria ďalší stupeň bunkovej regulácie. G-kvadruplexy a krížové štruktúry majú významný výskumný potenciál v oblasti biomedicíny, najmä v súvislosti s vývojom nových liekov proti nádorovým ochoreniam a vírusovým infekciám, aj keď oblasť ich využitia sa stále rozširuje.

Pre túto dizertačnú prácu boli vybrané dva proteíny záujmu – interferónom γ indukovateľný proteín 16 (IFI16) a proteín p53. Zatiaľ čo proteín IFI16 je zapojený hlavne v odpovedi prirodzeného imunitného systému na cudzorodú DNA, proteín p53 je tumor supresorový proteín, nepostrádateľný v regulácií bunkového cyklu. Oba proteíny svoju úlohu vykonávajú prostredníctvom interakcií s DNA. Oba proteíny sú zároveň senzormi DNA, IFI16 je senzorom DNA cudzorodej a p53 DNA poškodenej. Ďalšou ich spoločnou vlastnosťou je, že po rozpoznaní svojho cieľa aktivujú signálne dráhy vedúce k riešeniu problému. Buď k oprave a záchrane bunky, alebo riešeniu fatálnemu, a teda programovanej bunkovej smrti. Zatiaľ čo interakcia proteínu IFI16 s DNA je sekvenčne nešpecifická a plne závislá na elektrostatických interakciách negatívne nabitej cukorfosfátovej kostry nukleovej kyseliny s pozitívne nabitými doménami proteínov, väzba proteínu p53 je prevažne uskutočňovaná na špecifickú sekvenciu DNA. Čo ale ešte tieto dva proteíny spája, je preferenčná väzba na sekundárne lokálne štruktúry DNA, menovite G-kvadruplexy a krížové štruktúry tvorené v miestach genómu, v ktorých sa nachádzajú inverzné repetície.

Pochopenie významu štruktúry nukleových kyselín a proteínov, ktoré s nimi interagujú, môže viesť k účinnejším terapiám nielen v oblasti nádorovej biológie, ale aj v liečbe neurodegeneratívnych a infekčných ochorení.

1.1 Sekundárne lokálne štruktúry nukleových kyselín

V štandardných podmienkach sa ľudská DNA vyskytuje v B-forme (da Rosa et al., 2021; Travers and Muskhelishvili, 2015). B-forma predstavuje pravotočivú dvojzávitnicu s priemerným počtom 10,5 nukleotidového páru na jednu otáčku závitnice.

Lokálne môže vznikať Z-DNA, ľavotočivá dvojzávitnica formovaná negatívnou superhelicitou s väčším počtom páru báz na otáčku, priemerne 12 (Mitsui et al., 1970; Wang et al., 1979). Formovanie tejto štruktúry je štandardne energeticky nevýhodné, ale Z-DNA bola pozorovaná v sekvenciách bohatých na guanín a cytozín, v prítomnosti niektorých katiónov, napr. Na^+, K^+, Rb^+, Cs^+, Li^+, Mg^{2+} a Mn^{2+}. Bolo dokázané, že v sekvenciách tvoriacich Z-DNA dochádza k chromozomálnym zlomom a preto tieto miesta predstavujú miesta potenciálnych delécií a translokácií (Bansal et al., 2022).

Rozoznávame aj A-formu, pravotočivú dvojzávitnicu, ktorá je častejšie prítomná v dvojzávitnicovej RNA, kratších fragmentoch DNA a DNA/RNA hybridoch (Ivanov et al., 1974). Lokálne môžu vznikať aj iné štruktúry nukleových kyselín, akými sú G-kvadruplexy (G4), krížové štruktúry, vlásenky, triplexy a i-motívy (Bansal et al., 2022).

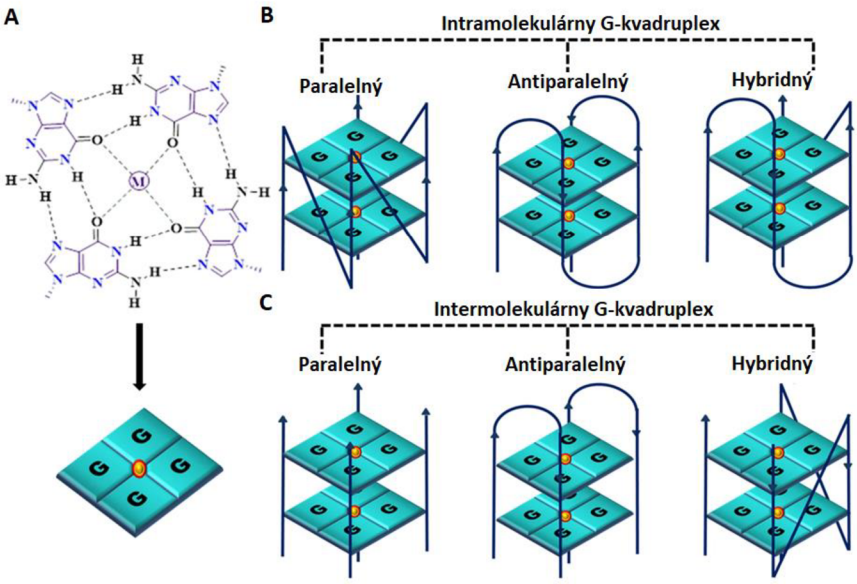
Všetky vyššie spomínané štruktúry DNA boli pozorované *in vitro* aj *in vivo*. Podieľajú sa na regulácii základných biologických procesov, akými sú replikácia, transkripcia a oprava DNA, ale môžu predstavovať aj komplikáciu pri replikácii DNA a tým potenciálne miesta genetickej nestability (Bansal et al., 2022). Napriek tomu, že tieto štruktúry predstavujú pre bunku riziko, sú evolučne uchovávané, čo poukazuje na ich biologicky relevantnú úlohu. V nasledujúcich podkapitolách budú predstavené jednotlivé sekundárne lokálne štruktúry nukleových kyselín s užším zameraním na G4 a krížové štruktúry.

1.1.1 G-kvadruplex

G4 je štruktúra tvorená v oblastiach DNA alebo RNA bohatých na guanín. Ide o tzv. nekanonickú štruktúru, ktorá sa tvorí spojením štyroch guanínových zvyškov. Tie sa ďalej spolu pomocou ôsmych vodíkových väzieb usporiadajú do planárnych štvorcových tetrád (Williamson et al., 1989). Nedochádza ku klasickému párovaniu báz, ktoré bolo popísané Jamesom Watesonom a Francisom Crickom (Watson and Crick, 1953), ale k párovaniu báz podľa Karsta Hogsteena (Hoogsteen, 1963). Guaníny v tomto alternatívnom párovaní využívajú N1H a N2H atómy tzv. Watson-Crickovskej akceptorovej strany a zároveň N7 a C6O atómy Hoogsteenovskej donorovej strany (Obr. 1 A). Hoogsteenove párovanie báz je typické nielen pre G4, ale aj pre triplexy a bolo pozorované aj v katalytických centrách ribozómov alebo v určitých typoch RNA, napríklad tRNA (Abrescia et al., 2004).

Podľa počtu planárnych tetrád guanínov sa G4 delia na 2-6-tetrádové G4. Termodynamická stabilita stúpa s počtom G-tetrád v G4 (Bartas et al., 2018). Efekt

na stabilitu má aj zloženie medzi G-blokmi, obecne G4 s kratšími slučkami vykazujú vyššiu stabilitu (Guédin et al., 2010). G4 je stabilizovaný monovalentným katiónom, ktorý je lokalizovaný v centrálnom kanále G4. Vytvorenie štruktúry bolo potvrdené v prítomnosti katiónov draslíka a sodíka (Ambrus et al., 2006; Gray and Chaires, 2008). V závislosti na orientácii reťazcov rozoznávame paralelné a antiparalelné G4, prípadne ich hybridnú formu. Pokiaľ G4 vzniká z jednej molekuly DNA, hovoríme o intramolekulárnom G4, pokiaľ je tvorený viacerými reťazcami, ide o intermolekulárny G4 (Majee et al., 2020). Štruktúra G4 sa nachádza na obrázku 1.



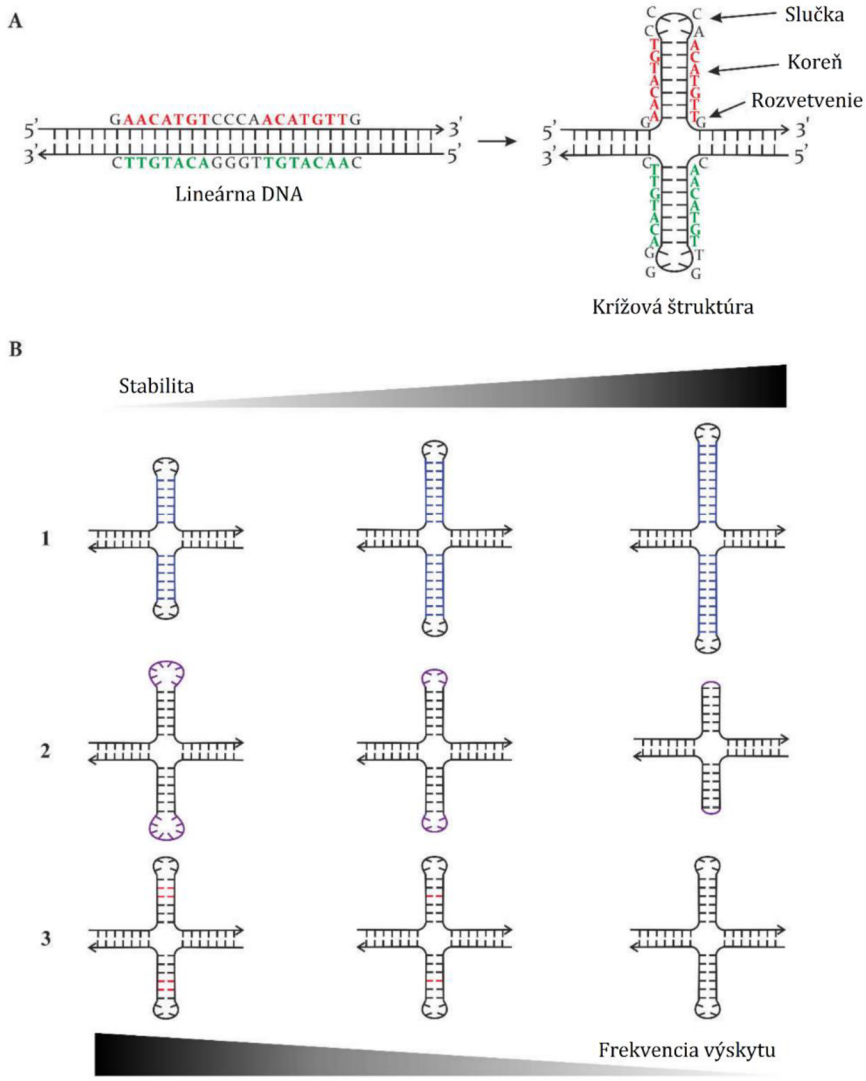
Obr. 1: Štruktúra G4.

(A) Planárna tetráda štyroch guanínových zvyškov navzájom interagujúcimi podľa Hoogsteenového párovania báz. Štruktúra G4 je stabilizovaná monovalentným iónom. (B) Intramolekulárny G4 v paralelnej, antiparalelnej a hybridnej forme. (C) Intermolekulárny G4 v paralelnej, antiparalelnej a hybridnej forme [15]upravené .

G4 nie sú stabilnejšie než duplex, ale ich tvorba je iniciovaná a stabilizovaná kompenzáciou nadzávitnicového vinutia, prítomnosťou proteínov interagujúcich s G4 alebo G4 ligandami (Awadasseid et al., 2021). G4 môžu byť využiteľné v biomedicíne dvoma prístupmi. Prvým sú vyššie spomínané G4 ligandy. Ide o malé aromatické molekuly, ktoré interagujú špecificky s G4 a nie duplexom. Častými ligandami sú deriváty akridínu (BRACO-19, RHPS4), deriváty nikotínamidu (PDS), deriváty bischinolínu (PhenDC3), porfyrínu (TMPyP4) a fluorochinolínu (CX-3543, CX5461) (Awadasseid et al., 2021). Druhým prístupom sú oligonukleotidy formujúce G4 nazývané aptaméry, ktoré sa môžu využiť na mimikciu signálnych dráh a špecifickú interakciu s proteínmi viažucimi sa na DNA alebo RNA G4 (Varizhuk et al., 2016). Viac o doterajšom využití G4 v klinickom výskume bude uvedené v kapitole 1.2.

1.1.2 Inverzné repetície

Inverzné repetície sú palindromatické sekvencie, úseky báz, ktoré sú čítané rovnako zľava doprava ako sprava doľava. V jednovláknovej DNA môže v palindromatických oblastiach dochádzať k tvorbe vláseniek. V prípade dvojvláknovej nukleovej kyseliny môže v palindromatickej oblasti vzniknúť krížová štruktúra, ktorá je znázornená na obrázku 2. Krížová štruktúra obsahuje miesto rozvetvenia, koreň a slučku. Stabilita a frekvencia výskytu týchto štruktúr je ovplyvnená primárnou sekvenciou, dĺžkou inverznej repetície, prítomnosťou a dĺžkou slučky a počtom nespárovaných báz v koreni krížovej štruktúry (Bowater et al., 2022).



Obr. 2: Krížová štruktúra.

(A) Krížová štruktúra vzniká v miestach inverzných repetícií a obsahuje miesto rozvetvenia, koreň a slučku. (B) Stabilita a frekvencia výskytu týchto štruktúr je ovplyvnená primárnou sekvenciou, (1) dĺžkou inverznej repetície, (2) prítomnosťou a dĺžkou slučky a (3)počtom nespárovaných báz (Bowater et al., 2022).

V genóme majú inverzné repetície rôzne funkcie. Napríklad, niektoré z nich môžu slúžiť ako miesta na väzbu proteínov, ktoré regulujú transkripciu génov. Iné sa môžu podieľať na oprave DNA, ktoré sú dôležité pre zachovanie integrity genómu (Bowater et al., 2022).

Inverzné repetície boli detegované vo všetkých formách života. Približne 36 % všetkých známych baktérií a 75 % archaea nesie vo svojom genóme prokaryotický imunitný systém známy ako CRISPR/Cas (Arroyo-Olarte et al., 2021; Pourcel et al., 2020). CRISPR je skratka pre genómový región, kde sa vkladajú 20nt dlhé sekvencie infikujúceho vírusu medzi pravidelne rozmiestnené krátke palindromické opakovania, zatiaľ čo Cas je názov skupiny proteínov, ktoré kooperujú s týmito oblasťami (Barrangou and Doudna, 2016; Charpentier and Doudna, 2013). Nukleázy Cas1 a Cas2 sú jediné Cas proteíny, ktoré sa vyskytujú vo všetkých systémoch CRISPR/Cas a obe nukleázy vyžadujú na integráciu nových intervenujúcich vírusových sekvencií negatívnu superhelicitu DNA. V podmienkach i*n vitro* je komplex Cas1-Cas2 schopný integrovať novú intervenujúcu sekvenciu aj mimo lokusu CRISPR, integrácia však nie je náhodná. Pri absencii lokusu CRISPR k integrácii dochádzalo prednostne v blízkosti inverzných repetícii schopných vytvárať krížové štruktúry (Moch et al., 2017; Nuñez et al., 2015). CRISPR/Cas systém získava široké uplatnenie v rôznych sférach výskumu a klinických aplikácií (Barrangou and Doudna, 2016).

1.1.3 Ostatné sekundárne lokálne štruktúry nukleových kyselín

Triplex je štruktúra DNA, v ktorej prostredníctvom vodíkových väzieb interagujú tri nukleotidové reťazce. V triplexe sú dva z týchto reťazcov navzájom spojené štandardným Watson-Crickovým párovaním, zatiaľ čo tretí reťazec sa k nim viaže pomocou Hoogsteenového spôsobu párovania báz (Mergny et al., 1991).

Existujú dva typy triplexu: triplex intramolekulárny (tretí reťazec je z rovnakej molekuly) a triplex intermolekulárny (tretí reťazec je z inej molekuly). V oboch prípadoch sa tretí reťazec viaže do velkého žliabku duplexu. Intramolekulárny triplex sa nazýva aj H-DNA a vzniká v homopurinových a homopyrimidinových sekvenciách (Frank-Kamenetskii and Mirkin, 1995). H-DNA hrá úlohu v expresii génov (Jain et al., 2008) a takisto môže predstavovať miesta genetickej nestability.

Ďalšou sekundárnou lokálnou štruktúrou nukleových kyselín je i-motív. Ten sa tvorí v oblastiach bohatých na cytozín. Vzniká rovnako ako G4 v oblastiach bohatých na GC bázy, ale pomocou párovania 4 cytozínov na základe Hoogsteenového párovania báz (Gehring et al., 1993). Interakcia C:C vzniká hemiprotonizáciou jedného z cytozínov. Stabilita štruktúry závisí od pH prostredia, počtu cytozínových zvyškov, ktoré spolu interagujú, teploty a koncentrácie solí. Zatiaľ čo mnohé i-motívy sú najstabilnejšie pri mierne kyslom pH, bola potvrdená ich existencia aj pri neutrálnom pH. Tento stav bol možný za určitých podmienok, vrátane nízkej teploty (4 °C), negatívnej superhelicity a prítomnosti katiónov striebra (Day et al.,

2014). Nahradenie cytozínu jeho halogénovanými analógmi (5-fluór, 5-bróm a 5-jód), takisto zvýšili stabilitu i-motívu DNA (Lannes et al., 2015). Takisto 5-metylácia cytozínu zvýšila pH, pri ktorom je i-motív stabilný a teplotu topenia i-motívov (Xu et al., 2015).

Zistilo sa, že i-motívy sa tvoria aj za biologických podmienok. Tieto štruktúry boli lokalizované v telomérach a promótoroch. Ich výskyt bol závislý na fáze bunkového cyklu, najvyšší výskyt bol zaznamenaný vo fáze G1 (Zeraati et al., 2018).

1.2 Sekundárne lokálne štruktúry a ich možné klinické aplikácie v humánnej medicíne.

1.2.1 Sekundárne lokálne štruktúry v ľudskom genóme

G4 majú významný výskumný potenciál v oblasti biomedicíny, najmä v súvislosti s vývojom nových liekov v liečbe nádorových ochorení, neurodegeneratívnych ochorení a vírusových infekcií (Awadasseid et al., 2021; Ruggiero and Richter, 2023). G4 sa totiž môžu podieľať na regulácii expresie génov, čím ovplyvňujú množstvo proteínov, ktoré bunka produkuje (Kim, 2019). Zatiaľ čo spočiatku sa myslelo, že G4 pôsobia na expresiu génu vždy represívne, dnes je jasné, že efekt môže byť aj aktivačný, v závislosti na orientácii reťazca, kde sa G4 nachádza (Kim, 2019). Formovanie v bunke je dynamické a mení sa v závislosti na bunkovom cykle. Nedávna štúdia popisuje tvorbu RNA G4 vplyvom bunkového stresu (Kharel et al., 2023).

V ľudskom genóme sa G4 vyskytujú hlavne v promótorových oblastiach a v telomérach (Huppert and Balasubramanian, 2007, 2005). Telomérne G4 sa podieľajú na udržiavaní genomickej stability. Teloméry chránia chromozómy pred poškodením a stratou génového materiálu počas replikácie DNA. Sekvencia telomerickej jednotky repetície sa líši v rôznych organizmoch. Nie všetky telomerické sekvencie tvoria G4, ale tvorba G4 v repetícii $(TTAGGG)_n$, ktorá sa vyskytuje v ľudskom genóme, bola potvrdená aj *in vivo* (Bazzicalupi et al., 2013; Jansson et al., 2019).

G4 sa nachádzajú častejšie v promótorových oblastiach génov, ako vo zvyšku genómu, čo bolo od počiatku ich výskumu nepriamym dôkazom, že G4 majú vplyv na génovú reguláciu. G4 boli predikované a ich existencia potvrdená v promótoroch onkogénov *c-myc* (Siddiqui-Jain et al., 2002), *k-RAS* (Cogoi and Xodo, 2006), *VEGF* (Sun et al., 2005), *b-RAF* (Recagni et al., 2019), *BCL-2* (Dexheimer et al., 2006).

G4 sú spojené aj s neurodegeneratívnymi ochoreniami so známou molekulárnou patogenézou založenou na expanzii oligonukleotidových repetícií (Teng et al., 2023). Ide o multiméry sekvencií CGG, GGGGCC a GGCCTG (Nordin et al., 2015), ktoré boli testované na vznik G4. Ďalšími potenciálnymi sekvenciami, ktoré prediktívne tvoria G4, sú AAGGG, TGGAA a CCTG. Expanzia týchto repetícií je typickým rysom rôznych

chorôb ako napríklad syndróm fragilného chromozómu X, amyotrofická laterálna skleróza, Parkinsonova choroba a spinocerebrálna ataxia SCA36 .

Niektoré lieky, ktoré boli vyvinuté na cielenie G4, preukázali protinádorovú aktivitu *in vitro* aj v *in vivo* nádorových modeloch a prešli do štádia klinického testovania. Napríklad liek Quarfloxin (CX-3543) zo skupiny fluorochinolónov bol vyvinutý na cielenie G4 štruktúr v promótorových oblastiach génu *c-myc* a bol testovaný v klinickej štúdii. Prešiel do II. fázy klinického testovania liečby neuroendokrinných nádorov (kód klinickej štúdie je NCT00780663), v ktorej bola štúdia ukončená z dôvodu prílišnej interakcie Quarfloxinu s albumínom. Efekt účinku bol popísaný ako vytesnenie nukleolínu z G4 sekvencie na netemplátovom vlákne DNA kódujúcej ribozomálnu RNA, čo spôsobí inhibíciu syntézy ribozomálnej RNA a apoptotickú smrť nádorových buniek (Xu and Hurley, 2022).

Ďalší liek, pidnarulex (CX-5461), je aktuálne testovaný v I. fáze klinického testovania v 2 klinických štúdiách. V prvej štúdii by mal byť využitý ako potenciálne terapeutikum pre tumory s mutáciou v géne *BRCA2* a/alebo *PALB2*, v ktorých nedochádza k homologickej rekombinácii (z angl. Homologous Recombination Deficiency HRD) (kód klinickej štúdie je NCT04890613) (Senhwa Biosciences, Inc., 2023). V druhej klinickej štúdii je pidnarulex využívaný na liečbu metastatických nádorov prostaty, pri ktorých nie je účinná kastrácia (kód klinickej štúdie je NCT05425862) (Lawrence et al., 2021; Peter MacCallum Cancer Centre, Australia, 2022).

Ďalšou možnosťou ako využiť štruktúru sekvencie na liečbu sú aptaméry (Platella et al., 2017; Roxo et al., 2019). Tieto krátke oligonukleotidy, ktoré môžu tvoriť sekundárne štruktúry, sa dajú využiť ako ligandy pre proteíny, patogénny a iné molekuly. V minulosti boli vhodné aptaméry identifikované metódou SELEX (Ellington and Szostak, 1990; Tuerk and Gold, 1990). Išlo o inkubáciu cieľového proteínu s knižnicou oligonukleotidov a následným screeningom viazaných sekvencií. Aptaméry formujúce G4 boli využité ako antikoagulanty väzbou na trombín, v liečbe nádorových ochorení interakciou s nukleolínom (AS1411), ako nanočastice a aptasenzory. Stabilita týchto molekúl môže byť regulovaná rôznymi modifikáciami oligonukleotidov (Roxo et al., 2019).

Výskum G4 sa v súčasnosti aktívne rozvíja a stále viac výskumníkov sa zameriava na hľadanie nových spôsobov, ako využiť tieto štruktúry na terapeutické účely, hlavne pre vývoj nových terapií pre nádorové a iné ochorenia súvisiace s telomérovou a genómovou stabilitou.

V ľudskom genóme sú inverzné repetície prevažne lokalizované v nekódujúcich oblastiach DNA, ktoré neobsahujú gény a sú charakterizované rôznymi dĺžkami opakujúcich sa úsekov báz. Inverzné repetície majú v ľudskom genóme fyziologický aj patologický význam. Podieľajú sa na oprave DNA, regulácie expresie, ale sú tiež spojené s dedičnými chorobami viazanými na poruchy helikáz, ako je Wernerov syndróm (Chan et al., 2012; Datta and Brosh, 2018), ochoreniami vznikajúcimi pri expanzii

trinukeotidov, ako napríklad syndróm fragilného X (Wells, 2009) alebo Sotosov syndróm (Kurotaki et al., 2005), ale aj nádorovými ochoreniami, hlavne v spojení s poruchami replikácie, transkripcie, chromozómovými prestavbami a amplifikáciami. Množstvo a stabilita štruktúr, ktoré je možné v rámci DNA uchovať, má dôležitý vplyv na integritu a plasticitu genómu.

Bolo pozorované, že krátke palindromatické sekvencie sú asociované s chromozómovými prestavbami a amplifikáciou pri viacerých typoch nádorov (Lu et al., 2015). Komplexná analýza sekvencie z 1000 ľudských genómov odhalila, že pravdepodobnosť výskytu patologických jednonukleotidových polymorfizmov je 14-krát vyššia v miestach inverzných opakovaní (Ganapathiraju et al., 2020). Jednonukleotidové polymorfizmy v inverzných opakovaniach sú spojené s mnohými ľudskými neurovými poruchami a rôznymi typmi nádorových ochorení (Guiblet et al., 2021).

Inverzné repetície tiež zohrávajú kľúčovú úlohu pri transpozícii a reorganizácii transponovateľných prvkov, ako sa ukázalo na niekoľkých modeloch chorôb, napríklad pri Williamsovom-Beurenovom syndróme, kde sú inzercie a delécie spojené s genómovými oblasťami, ktoré majú veľký počet inverzných repetícií (Savvateeva-Popova et al., 2017).

Inverzné opakovania sa často nachádzajú na miestach v genóme, ktoré sú náchylné na zlomy chromozómov, napríklad región FRA16D, kde inverzná repetícia tvorená AT bázami vytvára krížovú štruktúru, ktorá môže nadobúdať rôznu dĺžku a tým zastavovať replikáciu (Kaushal et al., 2019). Relatívna poloha a veľkosť inverzných repetíc je tiež dôležitá vo vzťahu k ich účinkom na stabilitu genómu. Tieto parametre ovplyvňujú frekvenciu translokácií, pričom veľkosť ramena repetície až do 100 bp koreluje s bodmi zlomu translokácií v bunkách po neoplastickej transformácií (Lu et al., 2015).

Spojenie niekoľkých ľudských chorôb s mutáciami DNA helikáz tiež naznačuje možnú úlohu krížových štruktúr pri týchto chorobách. Ak nastane porucha helikáz, krížové štruktúry tvoria prekážku pri replikácií a génovej expresii. Ochorenia, pri ktorých sú postihnuté práve helikázy patrí Wernerov syndróm, ktorý je spojený s mutáciami v helikáze *WRN* (Chan et al., 2012; Datta and Brosh, 2018).

Krížové štruktúry majú potenciál na využitie v medicíne. Napríklad krížová nanoštruktúra DNA napomáha cielenému dodávaniu chemoterapeutika doxorubicínu do nádorových buniek pri liečbe nádorov hrubého čreva (Abnous et al., 2018). Asociácia s reguláciou transkripcie otvára ďalšie terapeutické možnosti. Dôležitým nástrojom umožňujúcim takéto prístupy sú monoklonálne protilátky so špecificitou ku krížovej štruktúre, hoci doteraz boli využívané len na výskumné účely (Correll-Tash et al., 2021). V súčasnosti neexistujú žiadne malé molekuly, ktoré by špecificky rozpoznávali krížové štruktúry ako je tomu v prípade G4.

1.2.2 Sekundárne lokálne štruktúry vo vírusových genómoch

G4 a krížové štruktúry sa nachádzajú aj vo vírusových genómoch. Výskum G4 môže poskytnúť nové terapeutické ciele pre liečbu nielen nádorových ochorení, ale aj vírusových infekcií (Ruggiero and Richter, 2023). Výskyt, funkcia a vplyv G4 ligandov na životný cyklus bol skúmaný vo všetkých skupinách vírusov Baltimórovej klasifikácie. Vo väčšine prípadov G4 ligandy spôsobujú inhibíciu replikácie daného vírusu. Ich prehľad a známa funkcia G4 alebo G4 ligandu sa nachádza v tabuľke 1.

Tab. 1: Prehľad vírusov, pri ktorých prebieha výskum v spojení s G4 a skúmané ligandy

Vírus	BC	Známe funkčné prepojenie na G4 a G4 ligandy	Referencia
EBV	I.	represia EBNA1 translácie - PhenDC3, latencia	(Lista et al., 2017; Murat et al., 2014)
HSV-1	I.	Inhibícia replikácie - TMPyP4	(Frasson et al., 2021)
KSHV	I.	represia LANA translácie - TMPyP4, latencia	(Madireddy et al., 2016; Roy et al., 2016)
CMV	I.	Inhibícia replikácie - TMPyP4	(Ravichandran et al., 2018)
HPV	I.	Inhibícia replikácie - TMPyP4, PhenDC3	(Tlučková et al., 2013; Zahin et al., 2018)
SV40	I.	Inhibícia perylén diimid derivátov	(Fleming et al., 2019)
HCV	IV.	Inhibícia replikácie - PhenDC3, latencia	(Bian et al., 2019)
ZIKA	IV.	Inhibícia replikácie - TMPyP4	(Fleming et al., 2019, 2016)
SARS-CoV-2	IV.	Inhibícia replikácie - TMPyP4, PhenDC3	(Bartas et al., 2021, 2020; Goswami et al., 2021; Hognon et al., 2020; Ji et al., 2021; Liu et al., 2022; Qin et al., 2022)
EBOLA	V.	Inhibícia replikácie - TMPyP4	(Krafčíková et al., 2017; Wang et al., 2016)
IAV	V.	Inhibícia replikácie - TMPRSS2	(Brázda et al., 2021)
Nipah	V.	Inhibícia replikácie - TMPyP4, Braco-19, HGQ gény	(Majee et al., 2020)
HIV-1	VI.	Latencia	(Amrane et al., 2014; Bian et al., 2019; Fleming et al., 2019; Perrone et al., 2013a, 2013b; Piekna-Przybylska et al., 2014; Tassinari et al., 2020a; Tosoni et al., 2015)
HBV	VII.	Latencia	(Biswas et al., 2017)

Zvláštna pozornosť je venovaná výskumu zapojenia G4 do regulácie latentného štádia infekcie. Zistilo sa, že G4 regulujú aj životný cyklus vírusov. Bolo potvrdené, že sa môžu podieľať na prechode infekcie do latentného štádia vírusov ako je napríklad ľudský vírus imunitnej nedostatočnosti 1 (HIV-1) alebo rôzne herpesvírusy (Ajoge et al., 2022; Lista et al., 2017; Madireddy et al., 2016). Podľa svetovej zdravotníckej organizácie má až 70 % celosvetovej populácie latentnú infekciu herpes simplex vírusom 1 (HSV-1) (https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus, 2023). Keďže ide o oblasti, pri ktorých nie je možné trvalé vyliečenie, ale iba tlmenie príznakov, G4 sa považujú za nádejný terapeutický cieľ (Ruggiero and Richter, 2023).

G4 sa nachádzajú v kritických oblastiach genómu HIV-1, ktoré sú dôležité pre jeho replikáciu. Ide o väzbové miesto pre sp1 transkripčný faktor, kde sa môžu vyskytovať až 4 rôzne G-kvadruplexy. V závislosti na tom, ktorý G4 je sformovaný, sa mení dostupnosť pre sp1 faktor a tým aj génová expresia (Perrone et al., 2013a, 2017; Tassinari et al., 2020a; Tosoni et al., 2015). Špecifické ligandy by tak mohli ovplyvniť topológiu tohto regiónu a tým inhibovať replikáciu vírusu (Tassinari et al., 2020a).

Vírus Epsteina a Bárrovej (EBV) spôsobuje infekčnú mononukleózu, ale u imunodeficientních pacientov môže spôsobiť aj malígne lymfómy, napríklad Hodgkinov lymfóm. Niektoré štúdie tiež naznačujú, že G4 v genóme hostiteľa môžu ovplyvniť replikáciu a perzistenciu EBV prostredníctvom regulácie translácie EBNA1 mRNA (z angl. Epstein-Barr virus nuclear antigen 1). Gén *EBNA1* patrí do kategórie génov so skorou ranou expresiou. Proteín EBNA1 je kritický pre jeho replikáciu a prechod do perzistentnej infekcie v hostiteľskej bunke. Expresia EBNA1 musí byť prísne regulovaná, pretože proteín je vysoko imunogénny. EBV vírus obsahuje v mRNA EBNA1 18 nukleotidovú sekvenciu GGGGCAGGAGCAGGAGGA, ktorá tvorí G4. Pokiaľ bola tvorba G4 znemožnená komplementárnym oligonukleotidom, mRNA EBNA1 sa translatovala a dochádzalo k antigénnej prezentácii a zamedzeniu prechodu do perzistentného štádia infekcie. Naopak stabilizácia tohto G4 špecifickým ligandom pyridostatinom mala opačný účinok (Murat et al., 2014). Stabilizácia G4 nukleolinom takisto viedla k nižšej translácii mRNA EBNA1 (Lista et al., 2017). Obdobný mechanizmus bol pozorovaný aj v regulácií LANA mRNA vírusu spojeného s kaposiho sarkómom (KSHV) (Dabral et al., 2020; Madireddy et al., 2016; Uppal et al., 2014)

Podobný mechanizmus regulácie expresie génov skorého raného štádia bol nedávno popísaný aj pri ďalšom type vírusu – vírusu hepatitídy typu C (HCV). Tento vírus je hlavným pôvodcom chronickej hepatitídy a hepatocelulárneho karcinómu. Sekvencia tvoriaca G4 sa nachádza v mRNA kódujúcej jadrový proteín a selektívna väzba nukleolínu na tento element potlačila replikáciu vírusu (Bian et al., 2019).

Krížové štruktúry sú u vírovej menej preštudované. U Adeno asociovaného vírusu je známe, že sa prostredníctvom 4 inverzných repetíc tvoria vlásenkové štruktúry v tvare T v terminálnom regióne o dĺžke 125 bp. Táto terminálna inverzná repetícia AAV je väzbovým miestom pre transkripčné faktory a uľahčuje rekombináciu vírusového genómu s celulárnym genómom (Berns, 2020).

1.3 Interakcie proteínov so sekundárnymi lokálnymi štruktúrami nukleových kyselín

1.3.1 Väzba proteínov na nukleové kyseliny

Interakcia s nukleovými kyselinami je základnou vlastnosťou mnohých proteínov vrátane transkripčných faktorov, DNA reparačných proteínov, nukleáz, topoizomeráz či histónových proteínov. Väzba proteínov na nukleovú kyselinu umožňuje reguláciu génovej exprese, replikáciu DNA, opravu poškodenej DNA a mnoho ďalších biologických funkcií. Špecifita väzby proteín-DNA sa spočiatku pripisovala hlavne primárnej štruktúre DNA, sekvencii nukleotidov. Ukazuje sa, že aj trojrozmerná sekundárna štruktúra nukleovej kyseliny zásadne ovplyvňuje interakciu proteín-DNA(Bowater et al., 2022; Nakanishi and Seimiya, 2020; Ruggiero and Richter, 2020).

Lokálne sekundárne štruktúry DNA sú rozpoznávané systémami na opravu DNA. Tieto štruktúry môžu byť zdrojom rôznych preskupení genómu, ktoré sú genetickým základom mnohých chorôb (Correll-Tash et al., 2021; Lu et al., 2015). Lokálne štruktúry DNA existujú vo všetkých formách života (Bowater et al., 2022; Ding et al., 2018; Ruggiero and Richter, 2020) a napríklad nadmerné zastúpenie inverzných repetícií, bolo zaznamenané v blízkosti promótorových oblastí, a v miestach iniciácie replikácie. Telomérové oblasti mnohých organizmov pozostávajú zo sekvencií tvoriacich G4, kde sa špecifické proteíny s G4 afinitou môžu viazať a regulovať procesivitu telomerázy. V ľudskom genóme bolo preukázané, že proteíny podieľajúce sa na ochrane a replikácii ľudských telomér CTC1-STN1-TEN1, POT1 a RPA majú pozoruhodnú afinitu ku G4 DNA (Chaires et al., 2020; Olson et al., 2023).

Proteíny sa viažu na DNA pomocou elektrostatických interakcií, vodíkových väzieb a hydrofóbnych interakcií. DNA-väzbové proteíny majú rozličné funkcie, ktoré vyplývajú aj z mechanizmu interakcie. Najčastejšie sú to proteíny, ktoré majú špecifické väzbové domény (napr. helix-otáčka-helix alebo zinkový prst), ktoré sa priamo viažu na konkrétnu sekvenciu DNA. Tieto proteíny spadajú pod transkripčné faktory a sú dôležité pre reguláciu génovej expresie. Ďalšie proteíny sa môžu viazať na DNA nešpecificky, teda na akejkoľvek sekvencii DNA. Tieto proteíny sa zúčastňujú napríklad na oprave poškodenej DNA alebo reorganizácii chromozómov a ide o skupiny nukleáz, helikáz, topoizomeráz a ďalších.

Väzba proteínov na DNA môže byť tiež regulovaná rôznymi chemickými modifikáciami DNA, ako sú napríklad metylácia, acetylácia a fosforylácia. Tieto modifikácie môžu ovplyvniť konformáciu DNA a umožniť alebo brániť väzbe proteínov na DNA.

Celkovo je väzba proteínov na DNA kritická pre mnohé biologické procesy a je dôležitá pre pochopenie fungovania bunky a jej reakcie na rôzne podnety a stresové situácie.

1.3.2 Proteín p53

TP53 je známy tumor supresorový gén. Je to najčastejšie mutovaný gén nádorov ľudí (Leroy et al., 2014; Rivlin et al., 2011). Tento gén sa nachádza na krátkom ramienku chromozómu 17 a je tvorený 11 exónmi a 10 intrónmi. Proteín p53 sa skladá z transaktivačnej domény, DNA-viažucej domény a tetramerizačnej domény. Proteín p53 pôsobí ako senzor poškodenia DNA. Na základe rozsahu poškodenia je schopný aktivovať sériu signálnych dráh, ktoré vedú buď k zastaveniu bunkového cyklu a oprave poškodenej DNA alebo, ak je rozsah poškodenia príliš vysoký a poškodenie nie je možné opraviť, k programovanej bunkovej smrti formou apoptózy. Proteín p53 tým pomáha chrániť bunky pred vznikom nádorov a udržiava genetickú stabilitu (Hadj Amor et al., 2008). Mutácie v géne p53 môžu viesť k jeho nefunkčnosti a nedostatočnej kontrole bunkového cyklu a hromadeniu mutácií, čo zvyšuje riziko neoplastickej transformácie bunky (Freed-Pastor and Prives, 2012).

V štandardných podmienkach je hladina proteínu p53 udržiavaná v bunke pomocou ubikvitinligázy MDM2. Expresia génu *MDM2* je čiastočne regulovaná aj transkripčným faktorom p53. Vysoké hladiny proteínu p53 aktivujú expresiu MDM2, čím vytvoria spätnoväzbovú slučku pre udržiavanie hladiny proteínu p53 v bunke. V stresových podmienkach sa proteín p53 stabilizuje fosforyláciou až na 12 aminokyselinových zvyškoch prostredníctvom kináz ako napríklad ATM, ATR, DNA-PK, Chk1, Chk2, ktoré reagujú na poškodenie DNA. Proteín p53 stráca fosforyláciou schopnosť interakcie s ubikvitinligázou MDM2, na základe čoho sa aktivuje a začne sa hromadiť v bunkovom jadre (Honda et al., 1997; Nag et al., 2013; Wu et al., 1993).

Aktívny proteín p53 vo forme tetraméru interaguje so špecifickou sekvenciou DNA, ktorá je tvorené dvomi dekamérmi s konsenznou sekvenciou RRRCWWGYYY oddelenou medzerníkom v dĺžke 0-13 bp. Okrem sekvenčne špecifickej väzby na responzívny element (RE) sa p53 viaže na sekundárne lokálne štruktúry, ako sú krížové štruktúry, teloméry, triplexy a G4 (Brázda and Fojta, 2019; Coufal et al., 2013; Quante et al., 2012). p53 sa viaže s vyššou afinitou na negatívne aj pozitívne superhelikálne DNA, ktoré stabilizujú lokálne non-B DNA štruktúry. Sekvencie tvoriace non-B DNA štruktúry sa nachádzajú v mnohých génových promótoroch, napríklad *IL2R*, *DNA-POL1* a *c-myc* (Brázda and Fojta, 2019; Quante et al., 2012; Stansel et al., 2002).

Mechanizmami alternatívneho zostrihu, alternatívneho promótora a interného miesta (IRES) vzniká 12 izoforiem tohto proteínu (Anbarasan and Bourdon, 2019). Delia sa na N-koncové izoformy, ktoré označujeme Δ40, Δ133 a Δ160 a C-koncové izoformy, ktoré pomenuvávame α, β a γ. Funkcia jednotlivých izoforiem stále nie je objasnená, ale predpokladá sa, že narozdiel od kanonickej izoformy p53 môžu mať aj proliferatívne účinky a tak sa podieľať na neoplasmatickej transformácii (Anbarasan and Bourdon, 2019; Kazantseva et al., 2018).

Gén *TP53* môže byť mutovaný rôznymi spôsobmi; najčastejšími sú bodové nonsense mutácie, ale vyskytujú sa aj génové delécie, inverzie alebo mutácie s posunom rámca. Väčšina tumorogénnych mutácií *TP53* je popísaná v jeho centrálnej doméne v aminokyselinových zvyškoch priamo zapojených do väzby p53 na cieľovú sekvenciu DNA (Freed-Pastor and Prives, 2012). Zdedená mutácia *TP53* spôsobuje Li-Fraumeniho syndróm spojený so zvýšeným výskytom nádorov ľudí v mladom veku (Lang et al., 2004).

Vo svojej forme štandardného typu (wt) je p53 hlavným supresorom nádorov, ktorého funkcia je rozhodujúca pre ochranu pred neoplasmatickou transformáciou bunky. Hetero- aj homozygotná delécia *TP53* potenciuje rýchlosť tvorby nádorov myší. Avšak transgénne myši bez delécie *TP53*, obsahujúce bodové mutácie *TP53* pozorované pri ľudských nádorových ochoreniach vykazujú významne vyššiu rýchlosť vývoja nádoru ako myši s homozygotnou deléciou *TP53* (Lang et al., 2004; Olive et al., 2004). Dokonca bolo zistené, že v závislosti na konkrétnom type mutácie, vzniká rôzne spektrum nádorov. Wtp53 teda bráni tvorbe nádoru a jeho strata predisponuje k neoplasmatickej transformácií, zatiaľ čo bodovými mutáciami proteín nielenže stráca túto aktivitu, ale získava aj neoonkogénnu vlastnosť, čím ďalej podporuje tvorbu nádoru. Mutovaná forma p53 hrá aktívnu úlohu vo vývoji konkrétnych nádorov, vrátane zvýšenej rezistencie na liečbu. Tieto neomorfné mutantné aktivity p53 sa označujú ako GOF mutácie (z angl. „gain-of-function“) (Oren and Rotter, 2010).

Mutovaná forma GOF p53 môže byť zapojená do mnohých bunkových procesov, ako je zvýšená genetická nestabilita, antiapoptotická aktivita a zvýšenie invazívnosti (Vaughan et al., 2017). Na klinickej úrovni sú mutácie p53 spojené so zlou prognózou, skorším nástupom ochorenia, aneuploidiou a nežiaducimi odpoveďami na chemoterapeutiká (Robles and Harris, 2010; Short et al., 2020). Molekulárne mechanizmy, ktoré vedú k GOF proteínu, nie sú úplne jasné. Onkogénna vlastnosť mutantných foriem p53 môže byť čiastočne spôsobená aj dominantnými negatívnymi účinkami závislými na hetero-oligomerizácii mutantných a wt foriem proteínov, ktoré vytvoria plne dysfunkčný komplex (Oren and Rotter, 2010). Okrem toho je mutantný p53 často sprevádzaný stratou heterozygotnosti lokusu p53, čo spôsobuje stratu alely wtp53.

Spektrum mutácií proteínu p53 je veľmi široké a väčšinou nespôsobia jeho kompletný zánik, ale len viac či menej závažné zmeny v jeho štruktúre a vlastnostiach. Najbežnejšie mutácie sa vyskytujú v centrálnej časti proteínu a sú to

buď takzvané kontaktné mutanty zabraňujúce väzbe na RE (Eldar et al., 2013), alebo štrukturálne mutanty – meniace štruktúru proteínu, čím sa nepriamo zhoršuje väzba na cieľovú sekvenciu (Lukman et al., 2013). Proteín p53 potrebuje nielen neporušenú centrálnu doménu, ale pre účinnú interakciu s DNA je dôležitá aj tetramerizačná a C-koncová doména.

Mutantné proteíny p53 modulujú transkripciu v globálnom meradle prostredníctvom svojej väzby na intrónové a intergénne sekvencie predisponované k tvorbe non-B DNA štruktúr. Zistilo sa, že mutantné formy p53 interagujú vo väčšej miere s oblasťami 1 kb od miesta začiatku transkripcie v oboch smeroch (Kurotaki et al., 2005). Spektroskopia cirkulárneho dichroizmu ukazuje, že p53 môže tiež stabilizovať skladanie G4 *in vitro*. p53 sa tiež viaže na telomerické G4 a väzba je silnejšia s narastajúcim počtom telomerických opakovaní (Adámik et al., 2016). Medzi novonadobudnutú vlastnosť GOF mutantov p53 tak môže patriť aj zavedenie nerovnováhy medzi väzbou na štandardný RE proteínu p53 a sekundárne lokálne štruktúry.

1.3.3 Proteín IFI16

Primárnym krokom v obrane hostiteľa pred patogénnom je jeho rozpoznanie. Jedine po rozpoznaní patogénneho agens je možná aktivácia mechanizmov prirodzeného imunitného systému a následné zabránenie šíreniu infekcie. Rozpoznávanie patogénne špecifickej molekuly v ľudských bunkách zabezpečujú proteíny, ktoré sa spoločne nazývajú prirodzené rozpoznávacie receptory (PRR, z angl. innate pattern recognition receptors). PRR rozpoznávajú molekuly, ktoré sú vysoko konzervované a nie sú typické pre konkrétneho patogéna.

Medzi PRR radíme aj interferónom γ indukovateľný proteín 16 (IFI16), ktorý plní úlohu senzora cudzorodej nukleovej kyseliny (Unterholzner et al., 2010). Po rozpoznaní proteín IFI16 prostredníctvom signálnej kaskády aktivuje tvorbu prozápalových cytokínov, ktoré napomáhajú v boji organizmu proti infekcii. Okrem toho bolo dokázané, že proteín IFI16 reguluje replikáciu, expresiu vírusových génov priamou väzbou na DNA a podieľa sa aj na regulácii latentného štádia daného vírusu (Roy et al., 2016).

Interferónom γ indukovateľný proteín (IFI16) je senzor cudzorodých nukleových kyselín (Caposio et al., 2007; Unterholzner et al., 2010). Po rozpoznaní cudzorodej nukleovej kyseliny, proteín stimuluje odpoveď prirodzeného imunitného systému (Caposio et al., 2007), ale podieľa sa aj na regulácii transkripcie (Johnstone et al., 1998; Song et al., 2010) či už bunkových alebo vírusových génov, a bunkovej diferenciácie a proliferácii (Piccaluga et al., 2015; Zhao et al., 2015). IFI16 patrí do vysoko homológnej rodiny proteínov HIN-200, charakterizovanej 200 aminokyselinovým motívom obsahujúcim DNA väzbovú doménu na C-konci a PYRIN doménu na N-konci, ktoré sa podieľajú najmä pri interakciách proteín-proteín. Ľudská rodina HIN-200 sa

skladá zo štyroch charakterizovaných členov; proteín v melanóme 2 (AIM2), interferónom indukovateľnom proteíne X (IFIX), antigéne nukleárnej diferenciácie myeloidných buniek (MNDA) a IFI16 (Dawson and Trapani, 1996; Zhao et al., 2015). IFI16 sa od ostatných členov líši prítomnosťou dvoch HIN domén a intracelulárnou lokalizáciou. Bol detegovaný nielen v jadre, ale aj v cytoplazme. Vnútrobunková lokalizácia IFI16 je ovplyvnená posttranslačnými modifikáciami (Li et al., 2012), typom bunky (Kimkong et al., 2009), UV žiarením (Costa et al., 2011) a prítomnosťou infekčného agens (Dell'Oste et al., 2014).

IFI16 stimuluje tvorbu zápalov v určitých typoch buniek počas infekcie vírusom HSV-1 (Howard et al., 2022), EBV (Pisano et al., 2017), ľudským cytomegalovírusom (HCMV) (Griffante et al., 2022), KSHV (Roy et al., 2016), HIV-1 (Jakobsen et al., 2013), grampozitívnymi baktériami Listeria monocytogenes (Hansen et al., 2014) a Francisella novicida (Storek et al., 2015).

Proteín IFI16 rozoznáva aj RNA vírusy. Bolo potvrdené, že proteín vie rozpoznať aj chikungunya vírus a zika vírus (Wichit et al., 2019). Pri infekcii vírusom chrípky proteín IFI16 pozitívne reguluje expresiu proteínu RIG1, ktorý predstavuje senzor cudzorodej RNA (Jiang et al., 2021).

Schopnosť proteínu IFI16 rozlišovať medzi vnútornou a cudzou DNA v bunke hrá kľúčovú úlohu vo vrodenej imunite a je udržiavaná minimálnou dĺžkou DNA na spustenie imunitnej odpovede, ktorá je 70 bp (Unterholzner et al., 2010). Táto dĺžka je dlhšia ako voľná DNA medzi nukleozómami. Aktivácia zápalu, ku ktorej dochádza pri detekcii cudzej alebo poškodenej vlastnej DNA, spúšťa prozápalové procesy, ktoré môžu vyvolať pyroptózu v určitých typoch buniek (Mishra et al., 2022). Po rozpoznaní cudzej DNA proteín IFI16 aktivuje signálne dráhy, ktoré spúšťajú viaceré nešpecifické mechanizmy imunitnej odpovede. Po rozpoznaní cudzej DNA sa signalizácia šíri stimulátorom interferónového génu (STING), kinázou 1 viažucou serín/treonín proteín TANK (TBK-1) a regulačným faktorom interferónu 3 (IRF3). Ten dimerizuje a prechádza do jadra, kde zahŕňa regulátory transkripcie a indukuje expresiu cytokínov v infikovanom jadre a subcelulárnom prostredí (Almine et al., 2017). IFI16 sa podieľa na reštrikcii vírusovej infekcie okrem navodenia nešpecifickej imunitnej reakcie aj represiou génov. Ako represor pôsobí aj na gény vírusu, aj gény hostiteľskej bunky (Johnson et al., 2014; Johnstone et al., 1998).

Zvýšené hladiny IFI16 a autoprotilátky IFI16 boli zistené v sére pacientov trpiacich rôznymi autoimunitnými ochoreniami, ako je zápalové ochorenie čriev (Vanhove et al., 2015), systémový lupus erythematosus (Antiochos et al., 2022), psoriáza (Cao et al., 2016), reumatoidná artritída (Alunno et al., 2016), systémová sklerodermia (McMahan et al., 2016) alebo Sjögrenov syndróm (Baer et al., 2016). IFI16 by sa využiť ako marker na predpovedanie progresie ochorenia alebo špecifickej liečby, aj keď jeho presná charakterozácia v molekulárnej patogenéze vyššie spomínaných autoimunitných ochorení nie je známa.

IFI16 prvýkrát identifikovali ako proteín viažuci DNA Dawson a Trapani v roku 1995 (Dawson and Trapani, 1995). Neskôr bol IFI16 HIN-A opísaný ako proteín podobný RPA, ktorý zdieľa rovnaký oligonukleotid/oligosacharid viažuci (z angl. OB fold) (Yan et al., 2008). HIN-A doména sa viaže na jednovláknovú (ssDNA) a dvojvláknovú (dsDNA) *in vitro* s podobnou afinitou. HIN-B vykazuje vyššiu afinitu k DNA bohatej na GC (Ni et al., 2016). Väzba IFI16 na DNA nie je sekvenčne špecifická, vzniká elektrostatickými interakciami medzi negatívne nabitou cukorfosfátovou kostrou DNA a kladným nábojom aminokyselinovými zvyškami HIN domén. Prostredníctvom PYD domény proteín oligomerizuje a vzniká kooperatívne zostavenie filamentov IFI16 na dvojvláknovej DNA (Morrone et al., 2014). IFI16 preferenčne interaguje so superhelikálnou DNA a stabilizuje krížové štruktúry a G4 (Brázda et al., 2012). Okrem toho bol IFI16 lokalizovaný v oblasti NHEIII promótora génu *c-myc* obsahujúceho sekvencie s potenciálom tvorby G4 (Mathad et al., 2011). IFI16 by však mohol obmedziť vírusovú infekciu nezávisle od priamej interakcie proteínu s DNA. Bosso et al opísali dôležitosť interakcie domény PYRIN s transkripčným faktorom sp1 pri reštrikcii HIV-1, ktorá bola nezávisla na DNA-viažucich HIN doménach (Bosso et al., 2020). Responzívny element pre transkripčný faktor sp1 obsahuje sekvenciu s potenciálom tvorby G4 a preto môže tento proteín prepájať G4, proteín IFI16 a reštrikciu vírusovej infekcie (Bosso et al., 2020; Hároníková et al., 2016; Tassinari et al., 2020b).

1.4 Bioinformatické nástroje na analýzu sekundárnych lokálnych štruktúr

Aj keď prvé nástroje na predikciu sekundárnych štruktúr vznikli už v roku 2004, do dnešného dňa pribúdajú nové spôsoby ako predikovať tvorbu G4 v genómoch, prípadne ju bioinformaticky detegovať na genomickej úrovni z dát sekvenovania novej generácie. Existuje mnoho metód na predikciu G4, avšak každá z nich má svoje obmedzenia a nedokáže poskytnúť kompletnú informáciu o týchto štruktúrach. V súčasnosti vznikajú aj prehľadové komparatívne štúdie, ktoré porovnávajú výkonnosť týchto metód (Puig Lombardi and Londoño-Vallejo, 2020). Existencia viacerých prístupov komplikuje vzájomné porovnávanie *in silico* predikcií sekundárnych štruktúr DNA.

1.4.1 Bioinformatické nástroje založené na regulárnych výrazoch

Najstaršou metódou je detekcia vzorca sekvencie s potenciálom tvorby G4 (PQS) (Huppert and Balasubramanian, 2005). Napríklad regulárny výraz G2-4+N1−7G2-4+N1−7G2-4+N1−7G2-4 deteguje sekvencie, kde sú 4 skupiny guanínov v počte 2-4 oddelené navzájom medzerníkovou sekvenciou tvoriacou slučku o dĺžke 1-7 nukleotidov. Medzi bioinformatické nástroje na báze regulárnych výrazov pre detekciu PQS patrí Quadparser (Huppert and Balasubramanian, 2005), Quadruplexes (Todd et al., 2005), AllQuads (Kudlicki, 2016), ImGQfinder (Varizhuk et al., 2017).

Medzi nástroje založené na regulárnych výrazoch môžeme zaradiť aj Palindrome Analyzer a DetectIR (Brázda et al., 2016; Ye et al., 2014). Palindrome Analyzer je nástroj, ktorý sa používa na detekciu inverzných repetícií (Brázda et al., 2016). Nástroj identifikuje palindromické sekvencie a okrem informácie o ich polohe a veľkosti poskytuje užívateľom možnosť nastaviť aj počet tolerovaných nespárovaných báz a veľkosť slučky. Okrem toho nástroj počíta aj teoretickú voľnú Gibbsovu energiu potrebnú na tvorbu duplexu alebo krížovej štruktúry metódou najbližšieho suseda (SantaLucia, 1998).

1.4.2 Skórovacie metódy založené na regulárnych výrazoch

Medzi metódy založené na regulárnych výrazoch so skórovacím systémom patrí QGRS mapper (Kikin et al., 2006), pqsfinder(Hon et al., 2017) a G4Catchall (Doluca, 2019). PQSfinder umožňuje okrem klasického nastavenia dĺžky G-tetrády, slučky aj nastavenie počtu mismatchov a individuálnu dĺžku každej G-tetrády a slučky zvlášť. Výsledné skóre zohľadňuje penalizácie za počet mismatchov a dĺžku slučky.

QGRS Mapper umožňuje používateľom zvoliť rôzne nastavenia pre analýzu, vrátane dĺžky G-skupiny, minimálneho počtu guanínových zvyškov v G-skupine a minimálneho počtu nukleotidových zvyškov medzi G-skupinami.

QPARSE algoritmus kombinuje metódu založenú na regulárnych výrazoch so skórovacích systémom a m-fold predikciu (Berselli et al., 2020). Navyše umožňuje detekciu multimérnych PQS s dlhou slučkou.

1.4.3 Metódy založené na prehľadávacom posúvnom okne a skórovaní

Ďalšou skupinov metód sú metódy využívajú posúvne prehľadávacie okno. Počítajú skóre v určenej dĺžke sekvencie. Medzi tieto nástroje patrí G4Hunter (Brázda et al., 2019), G4P , cG/cC (Beaudoin et al., 2014), G4RNA Screener softvér (Garant et al., 2018).

G4Hunter bol vyvinutý na základe algoritmu, ktorý kombinuje rôzne prediktívne prvky, ako sú počet guanínových zvyškov, ich rozmiestnenie a sekvenčná entropia(Brázda et al., 2019). Prítomnosť adenínu a tymínu nemá vplyv na výsledné skóre. Prítomnosť guanínu poskytuje pozitívny prídavok do výpočtu, ktorého hodnota závisí od počtu guanínou za sebou do maximálnej výšky +4. Prítomnosť cytozínu má opačný efekt a teda negatívne skóre, ktorého absolútna hodnota závisí od počtu cytozínov v bezprostrednej blízkosti. S nárastom absolútnej hodnoty G4Hunter skóre narastá pravdepodobnosť tvorby G4. Prednastavené skóre 1,2 sa ukázalo ako vhodný kompromis medzi falošne pozitívnymi a falošne negatívnymi výsledkami (Bedrat et al., 2016).

1.4.4 Orthogonálny systém

Novou metódou na predikciu PQS a inverzných repetíc zároveň je použitie orthogonálneho systému nukleových kyselín (Víglaský, 2022). Adenín a tymín sú vyobrazené v rovine xy, bázy cytozín a guanín v rovine xz. Komplementárne bázy majú opačnú hodnotu. Výpočet skóre je určený priemetom konkrétnej postupnosti nukleotidov a výpočtom jeho plochy buď v rovine xy alebo xz. Vydelením plochy počtom nukleotidov dostaneme samostatné skóre pre rovinu xy a rovinu xz. Antiparalelné vlákno predstavuje zrkadlový obraz pre oba komponenty (Víglaský, 2022). Aj keď tento softvér nebol ešte robustne testovaný, má potenciál v predikcii štruktúr.

1.4.5 G4seq

G4-seq je technika využívajúca bioinformatickú analýzu sekvenácie novej generácie na predikovanie PQS. Táto metóda kombinuje experimentálne prístupy s využitím schopnosti G4 štruktúr viazať sa na G4 ligandy v prítomnosti katiónov draslíka, metódy sekvenácii novej generácie, ktorá umožňuje identifikovať a mapovať G4 v celých genómoch s následnou bioinformatickou analýzou (Chambers et al., 2015; Nawy, 2015).

G4-seq umožňuje identifikovať G4 štruktúry v rôznych bunkových kultúrach a tkanivách a kvantitatívnu analýzu G4 tvorby za rôznych podmienok, napríklad po expozícii buniek na rôzne liečivá, v rôznom štádiu bunkového cyklu a podobne. Metódou G4seq bolo zanalyzovaných 12 druhov, ktorých výsledky sú verejne dostupné (Marsico et al., 2019). G4-seq môže byť užitočným nástrojom pre štúdium biologických procesov spojených s tvorbou a rozkladom G4, ako aj pre identifikáciu nových cielených oblastí pre liečbu ochorení súvisiacich s G4 štruktúrami.

1.4.6 Neurónové siete

Medzi prístupy založené na strojovom učení patrí DeepG4 (Rocher et al., 2021), Quadron (Sahakyan et al., 2017), G4RNA screener (Garant et al., 2018), G4Boost (Cagirici et al., 2022) a G4detector (Barshai et al., 2022). Podmienkou vyššie spomínaných prístupov založených na neurónových sieťach je využitie trénovacieho datasetu. Problémom vyššie spomenutých prístupov je, že ako trénovací dataset často využívajú sekvencie, ktorých tvorba G4 bola potvrdená *in vitro.* Tie boli prevažne identifikované staršími metódami založenými na regulárnych výrazoch, preto pravdepodobne dochádza k podceneniu výskytu nepravidelných a multimérnych G4. Tento prístup má ale vysoký potenciál do budúcnosti.

1.4.7 Databázy

Dôležitou súčasťou bioinformatickej analýzy sú databázy. Tie slúžia napríklad ako trénovací súbor dát pre algoritmy založené na neurónových sieťach, ktoré boli spomínané v predchádzajúcej kapitole. Databázy môžeme rozdeliť na tie, ktoré evidujú iba predikované regióny nukleovej kyseliny alebo iba sekvencie, ktoré boli testované biofyzikálnymi metódami (Miskiewicz et al., 2021). Samostatnou kategóriou sú databázy, ktoré združujú aj proteíny, ktoré s nimi interagujú.

GRSDB2 a GRS_UTRdb sú databázy PQS detegovaných QGRS softvérom v mRNA eukaryotických organizmov (Kikin et al., 2008). Ďalšou databázou transkriptomických predikovaných G4 je G4Atlas (Yu et al., 2023). Databázou zameranou výlučne na RNA G4 je QUADRatlas. Okrem sekvencií je v nej uložený aj zoznam proteínov interagujúcich s RNA G4 (Bourdon et al., 2023). G4IPDB je databáza proteínov interagujúcich s DNA aj RNA G4 (Mishra et al., 2016).

2 Ciele dizertačnej práce

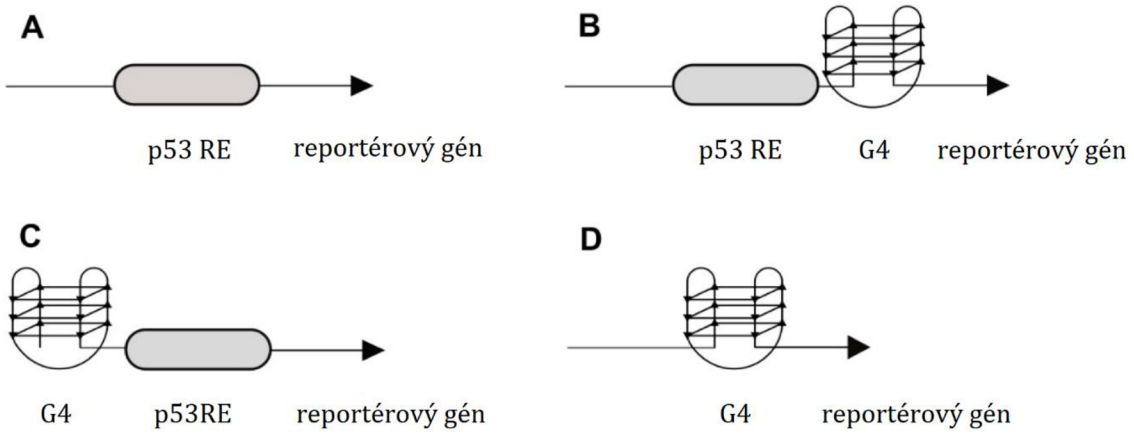
Ciele dizertačnej práce možno rozdeliť do dvoch oblastí:

1. Analýza DNA-väzbových proteínov p53 a IFI16 molekulárne-biologickými a biofyzikálnymi metódami so zameraním na sekundárne lokálne štruktúry DNA. Proteín IFI16 a p53 interagujú s lokálnymi sekundárnymi štruktúrami DNA (Brázda et al., 2012; Hároníková et al., 2016). Keďže G4 a inverzné repetície sa často nachádzajú práve v regulačných elementoch DNA, vzniká otázka či práve väzba proteínov na tieto lokálne štruktúry môže ovplyvňovať funkciu skúmaných proteínov. Pre analýzu proteínu p53 bude využitá najmä kvasinková reportérová esej s cieľom definovať vplyv a) G4 v blízkosti p53 RE, b) izoforiem a mutantných variant proteínu p53 na transaktivačnú schopnosť proteínu wtp53. Väzbové vlastnosti proteínu IFI16 budú analyzované využitím gélovej retardačnej eseji a mikroskopie atomárnych síl za účelom posúdenia rozdielu vo väzbe na lineárnu a superhelikálnu DNA.
2. Bioinformatická analýza sekundárnych lokálnych štruktúr v ľudskom genóme a v genómoch vírusov. Cieľom tejto časti je identifikovať skupiny vírusov a cieľové oblasti ľudského genómu, ktoré by mohli byť zaujímavé z hľadiska štúdia proteínov interagujúcich s DNA.

3 Publikované výsledky

3.1 Publikácia 1. Vplyv G-kvadruplexu v blízkosti cieľových sekvencií p53 na transaktivačný potenciál izoforiem p53α

Metódou klonovania v kvasinkách na princípe homologickej rekominácie Delitto Perfetto (Monti et al., 2019) boli v rámci publikácie v prílohe A pripravené 4 izogénne kvasinkové kmene, ktoré sa líšili iba v oblasti pred promótorom reportérového génu (v smere transkripcie), ktorým bol gén kódujúci luciferázu (Obr. 3). Išlo o kombináciu p53 responzívneho elementu z promótora proapoptotického génu *PUMA* a sekvenciu, ktorá tvorí G4 *in vitro*. Transformáciou príslušného kvasinkového kmeňa plazmidom obsahujúcim gén pre transkripčný faktor interagujúci s miestom pred promótorom reportérového génu môžeme aktivovať expresiu luciferázového reportérového génu.



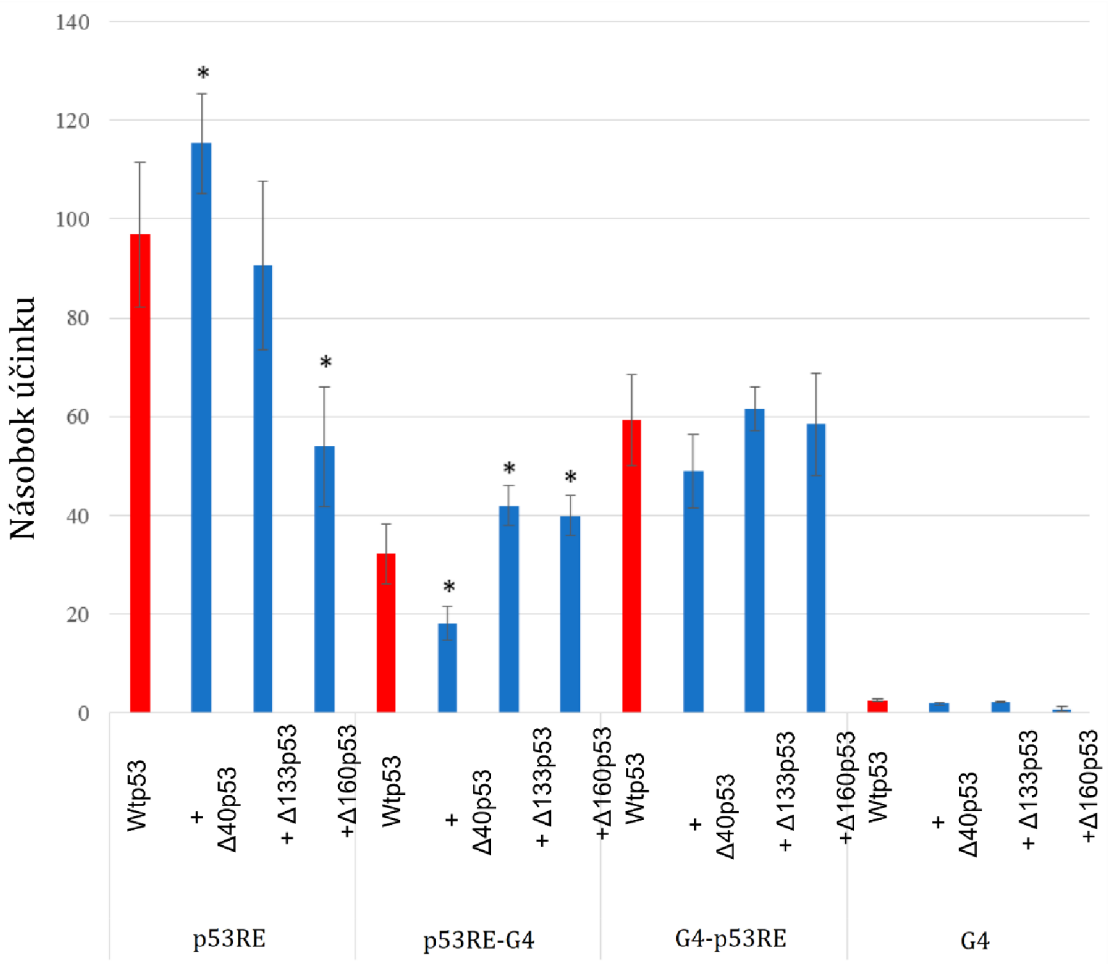
Obr. 3: Reportérové kmene použité v transaktivačných experimentoch.

(A) Začlenený bol samostatný p53 RE. (B) Sekvencia s potenciálom tvorby G4 začlenená pred p53 RE v smere transkripcie. (C) Sekvencia s potenciálom tvorby G4 začlenená za p53 RE v smere transkripcie. (D) Do kvasinkového izogénneho kmeňa bola vložená samostatná sekvencia s potenciálom tvorby G4.

Prítomnosť G4 v okolí p53 responzívneho elementu mala represívny vplyv na transaktiváciu reportérového génu prostredníctvom p53 (Obr 4). Vyšší vplyv mal G4 začlenený za RE v smere transkripcie. Samotná sekvencia G4 nebola dostatočná na aktiváciu transkripcie proteínom p53 (Porubiaková et al., 2019).

Výsledky ukázali, že Δ160p53α exprimovaný spolu s wtp53 znížil transaktiváciu oproti p53 samotnému v kmeni, ktorý obsahoval iba p53 RE. Pozorovanie je v súlade s predtým publikovanými štúdiami, kde Δ160p53 inhiboval apoptózu, na rozdiel od Δ133p53 (Candeias et al., 2016). Tento jav ale nebol pozorovaný na druhom kvasinkovom kmeni, kde bol za RE začlenený G4. Izoformy Δ133p53α a Δ160p53α

v tomto kmeni neznížili transaktiváciu p53; v skutočnosti došlo k jej zvýšeniu (obr. 4). Tento výsledok naznačuje, že G4 v blízkosti RE môže zásadne ovplyniť správanie heterotetramérov p53. Našou hypotézou je, že zatiaľ čo bez G4 v prítomnosti p53 RE je heterotetramerizácia Δ133p53α alebo Δ160p53α s wtp53 v porovnaní so samotným p53 nevýhodná a inhibuje transaktiváciu na cieľových RE. Naopak, v spojení so štruktúrou G4 dochádza k stérickej výhode menšieho heterotetramérneho komplexu a tým môže byť aktivácia reportérového génu účinnejšia.



Obr. 4: Porovnanie transaktivačnej aktivity p53 v 4 kvasinkových kmeňoch a v koexpresii s N-koncovými izoformami.

Vplyv izoforiem Δp53α s indukovateľnou expresiou na transaktivačnú aktivitu konštitutívne exprimovaného proteínu p53. Intenzita signálu je vyhodnotená ako násobok účinku v porovnaní s prázdnym vektorom neobsahujúcim proteín p53 ani jeho izoformy. (A) v médiu s 0,2% galaktózou; (B) v médiu s 2% galaktózou. Boli použité štyri izogénne kvasinkové kmene, s cieľovým miestom p53 (PUMA), s cieľovým miestom p53 po sekvencii tvoriacej G4 (G4-PUMA), s cieľovým miestom p53 pred sekvenciou tvoriacou G-kvadruplex (PUMA-G4) a G4 tvoriaca sekvencia upstream od luciferázového génu. Hviezdičky označujú štatisticky významnú zmenu transaktivácie vplyvom danej izoformy od samostatného proteínu p53 ($p < 0,05$).

3.2 Publikácia 2. Hodnotenie vplyvu sekvencie s potenciálom tvorby G-kvadruplexu na transaktivačný potenciál proteínov z rodiny p53 prostredníctvom kvasinkového transaktivačného systému

V publikácií, ktorá je súčasťou prílohy B, boli využité kvasinkové izogénne kmene s p53 RE z promótora génu PUMA a sekvenciu, ktorá tvorí G4 *in vivo* začleneným pred promótorom reportérového génu znázornené na obr. 3 (Monti et al., 2021). Tentokrát sme sledovali vplyv proteínu proteínu p63, p73 a mutantných foriem proteínu p53 na transaktivačnú schopnosť wtp53.

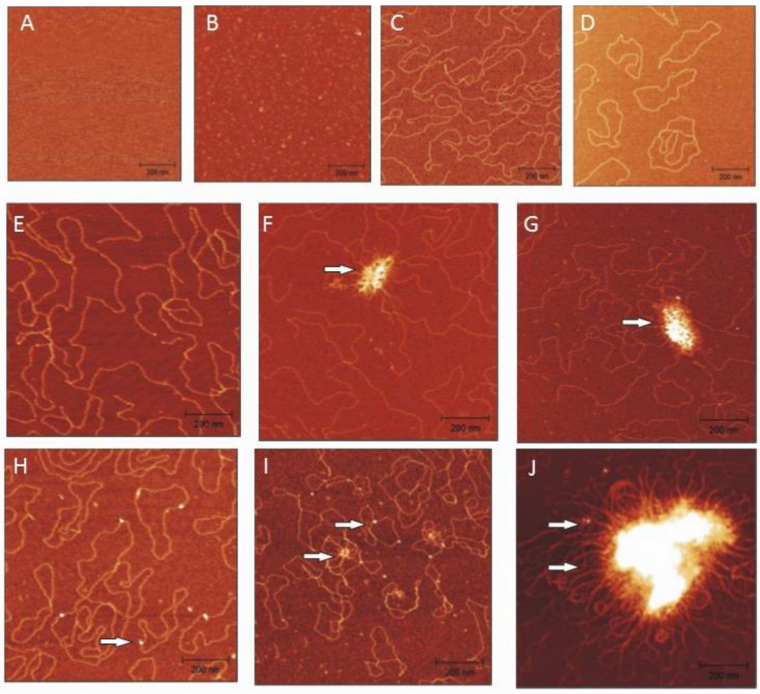
Pozorovali sme, že prítomnosť PQS proti smeru a výraznejšie po smere traskripcie vzhľadom na p53 RE spôsobuje inhibíciu transaktivácie sprostredkovanej proteínmi rodiny p53, ich N-koncovými delečnými izoformami aj mutantnými variantami. Je zaujímavé, že napriek tomu, že N-koncovým delečným izoformám chýba transaktivačná doména plnej dĺžky, tieto proteíny sú transkripčne aktívne a ukázalo sa, že regulujú expresiu vlastného súboru génov (Kartasheva et al., 2003; Kazantseva et al., 2018; King et al., 2003). Rozdiel v relatívnej transaktivácii medzi wtp53 a delečnými N-koncových izoformami proteínov p53, p63 a p73 bol významne znížený prítomnosťou sekvencie s potenciálom tvorby G4 začlenenej v smere transkripcie od p53 RE. Tento fakt naznačuje, že inhibícia transaktivácie pomocou G4 nepôsobí na všetky transkripčné faktory rovnako, ale je závislá na konkrétnom transkripčnom faktore. Silnejšia inhibícia bola zaznamenaná u wtp53 v porovnaní s delečnými N-koncovými izoformami. Je možné, že rovnako ako v kapitole 3.1 ide aj o stérickú výhodu menšieho heterotetramérneho komplexu v prítomnosti G4 v blízkosti p53 RE.

Ďalej sme rozšírili štúdiu o proteínové mutanty rodiny p53. Boli vybrané mutantné varianty proteínov p53 a p63, ktoré sú funkčne heterogénne z hľadiska schopnosti transaktivácie a často sa vyskytujú v ľudských nádoroch. Prítomnosť sekvencie s potenciálom tvorby G4 (najmä v smere transkripcie od p53 RE) znižovala rozdiel medzi aktivitou transaktivácie prostredníctvom wtp53 a mutantného proteínu. Simulácia heterozygotného stavu kotransformáciou plazmidov exprimujúcich súčasne wt aj mutantnú variantu proteínu p53 alebo p63 neodhalila očakávaný dominantne negatívne efekt mutantnej varianty. Čiastočná aktivita wtp53 proteínu bola zachovaná a G4 v blízkosti p53 RE nemal na relatívnu expresiu wtp53 s mutantnými variantami významný efekt.

Topológia DNA v okolí responzívnej sekvencie p53 tak môže predstavovať ďalšiu úroveň v regulácii signálnych dráh, do ktorých je tento transkripčný faktor zapojený.

3.3 Publikácia 3. Bližší pohľad na duálny režim väzby IFI16 na superhelikálnu a lineárnu DNA

Bolo zistené, že protein IFI16 preferenčne interaguje so superhelikálnou DNA a stabilizuje G4 a krížové štruktúry (Brázda et al., 2012; Hároníková et al., 2016). Hároníková a kolektív odhalili nielen preferenciu, ale aj náznak rozdielneho spôsobu väzby na dsDNA v porovnaní s väzbou na G4 alebo krížovú štruktúru (Hároníková et al., 2016). Protein IFI16 s oligonukleotidom vo forme dsDNA interaguje už pri nižších vzájomných molárnych pomeroch vo forme proteínového oligoméru. Pri rovnakých podmienkach komplex IFI16 s G4 alebo krížovou štruktúrou vykazoval výrazne nižšiu molárnu hmotnosť. Väzba bola skúmaná iba prostredníctvom elektroforetickej retardačnej analýzy. Preto bolo cieľom práce, ktorá je súčasťou prílohy C, bližšie preštudovať pozorovaný rozdiel vo väzbe na lineárnu a superhelikálnu DNA pokročilejšími metódami, konkrétne mikroskopiaou atomárnych síl (AFM) (Obr. 5) a charakterizovať domény zodpovedné na tento rozdiel (Valková et al., 2023).



Obr. 5: Mikroskopia atomárnych síl komplexov proteínu IFI16 a DNA.

(A) Kontrolná vzorka - čistá sľuda (B) Protein IFI16 nanesený na sľudu v depozičnom pufri. (C) Lineárna DNA. Linearizovaný pBluescript s koncentráciou 1 ng/ul DNA. (D) Superhelikálna DNA. Superhelikálny pBluescript - koncentrácia 1 ng/ul. (E) Lineárna DNA a IFI16 v molárnom pomere DNA:protein 1:10 vo väzbovom pufri (F) Lineárna DNA a IFI16 v molárnom pomere 1:20. (G) Lineárna DNA a IFI16 v molárnom pomere 1:50. (H) Superhelikálna DNA a IFI16 v molárnom pomere 1:10. (I) Superhelikálna DNA a IFI16 v molárnom pomere 1:20. (J) Superhelikálna DNA a IFI16 v molárnom pomere 1:50. Podmienky merania boli pre všetky vzorky rovnaké. Všetky obrázky sú skeny 1 µm x 1 µm a biele šípky poukazujú na pozorované komplexy DNA s proteínom.

Metódou AFM bol potvrdený fenomén pozorovaný pri elektroforetických retardačných analýzach (Hároníková et al., 2016). Potvrdil sa rozdielny mechanizmus väzby pri nízkych molárnych pomeroch DNA:proteín do výšky 1:20. Výsledky ukazujú, že duálny režim väzby IFI16 závisí od topológie DNA. IFI16 sa pri nižších molárnych pomeroch do 1:10 viaže buď kooperatívnym spôsobom, ktorý vyústi do tvorby proteínového oligoméru na lineárnu DNA (panel E a F obrázku 5) alebo vo forme globulárnych komplexov v prítomnosti negatívnej superhelicity DNA (panel H a G obrázku 5).

Pozorovaný rozdiel vo väzbe je pravdepodobne spôsobený úlohou jednotlivých domén proteínu IFI16. HIN domény interagujú s DNA, zatiaľ čo C-koncová PYRIN doména oligomerizuje. Jediná molekula proteínu naviazaná na dsDNA teda môže prostredníctvom PYD domény interagovať s ďalšími molekulami IFI16. Zároveň bolo zistené, že proteín skenuje DNA nezávisle na PYD doméne, čo urýchľuje stret jednotlivých IFI16 molekúl a tvorbu proteínového oligoméru. Superhelicita DNA tak zdá sa vo výsledku podnecuje vznik väčších globulárnych komplexov a urýchľuje tento proces. Z našich výsledkov sme dospeli k záveru, že proteín pravdepodobne najskôr tvorí špecifické komplexy s určitými oblasťami DNA s negatívnou superhelicitou a následne v prípade zvýšenia koncentrácie IFI16 v bunke, napríklad v dôsledku infekcie, tvorí oligomér a stimuluje odpoveď nešpecifického imunitného systému. Ďaľšou hypotézou je, že duálny spôsob väzby proteínu IFI16 môže byť zásadný aj pre jeho funkciu. Proteín po interakcii s lineárnou DNA navodí nešpecifickú imunitnú reakciu prostredníctvom tvorby proteínového oligoméru, ale pri špecifickej interakcii so sekundárnymi lokálnymi štruktúrami zabezpečí represiu génovej expresie.

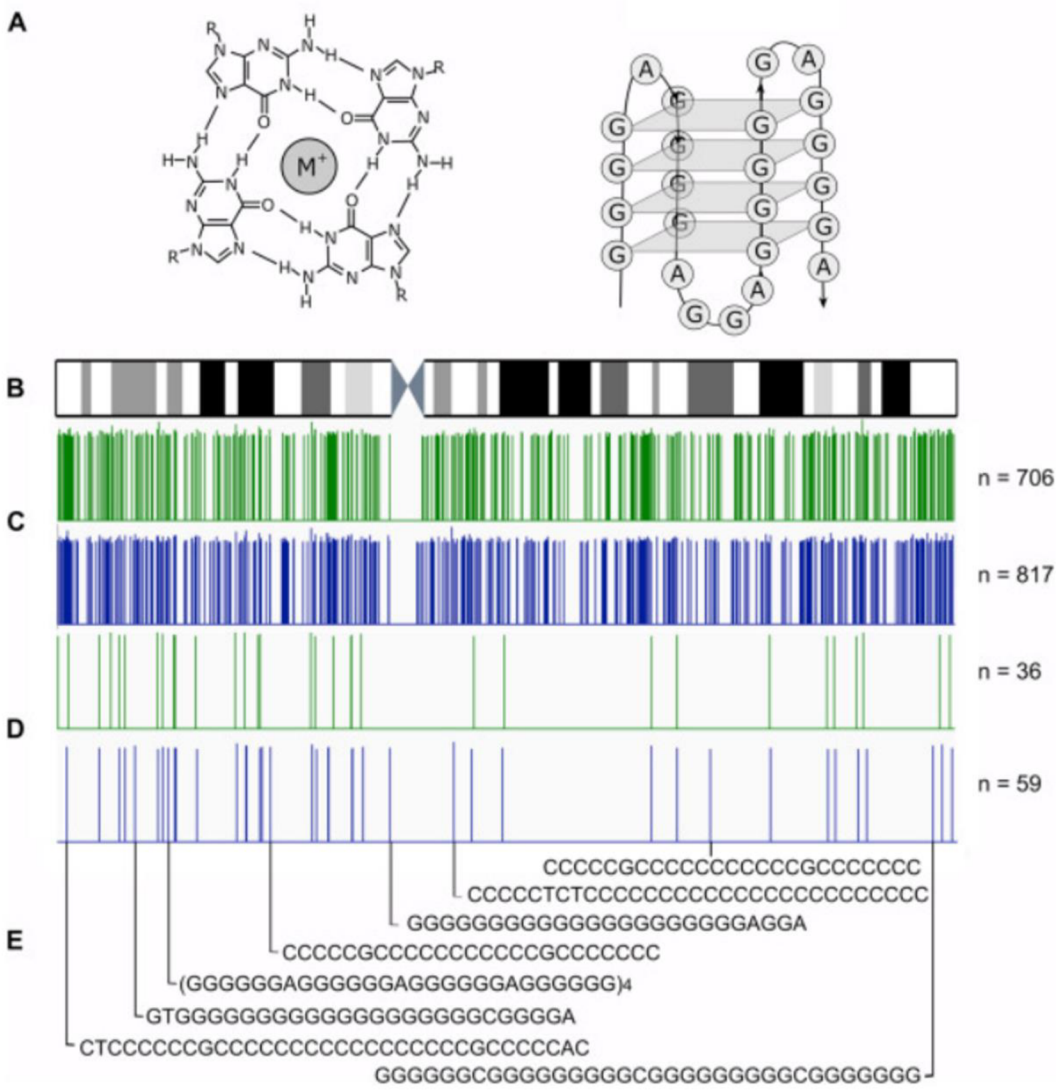
Zo štúdie ale nie je jasné či samotná superhelicita alebo konkrétna sekundárna štruktúra DNA spôsobuje rozdiel vo väzbe. Podľa prediktívnych algoritmov obsahuje použitý plazmid 3 PQS a 4 inverzné repetície, ktoré predstavujú potenciálne interaktory IFI16.

3.4 Publikácia 4. V kompletnej sekvencii ľudského chromozómu X sa vyskytli predtým neidentifikované sekvencie s potenciálom tvorby G-kvadruplexu

Aj keď prvá verzia sekvencie ľudského genómu bola zverejnená už v roku 2003 a odvtedy sa neustále vylepšovala až do široko rozšírenej verzie GRCh38, stále ostávalo ešte 8 % genómu prevažne v heterochromatínových úsekoch, ktorých sekvencia nebola rozlúštená (Miga et al., 2020). Išlo hlavne o nekódujúce oblasti repetícií, chromozomálne oblasti a chromozóm Y, ktorých sekvenovanie a následná bioinformatická analýza bola a do dnešného dňa je náročná.

Sekvenovaniu predtým neidentifikovaných repetitívnych oblastí výrazne napomohla kombinácia dvoch typov sekvenovania novej generácie – Pacific Biosciences (PacBio) a Oxford Nanopore Technologies, ktorých výstup predstavujú dlhé čítania do 1000 bp. Z celý projektom stojí konzorcium od teloméry po teloméru (T2T). Kombináciou týchto dvoch prístupov sa podarilo rozlúštiť zvyšné regióny ľudského genómu, opraviť predchádzajúce chyby, predikovať nové gény, vrátane tých proteín kódujúcich a doplniť chýbajúce úseky, ktoré zahŕňajú centroméry a krátke ramienka akrocentrických chromozómov. Jedinou dodnes chýbajúcou časťou genómu je chromozóm Y (Altemose et al., 2022; Nurk et al., 2022).

Prvým chromozómom so zverejnenou kompletnou sekvenciou od teloméry krátkeho ramienka po teloméru dlhého ramienka bol chromozóm X. Cieľom článku, ktorý je v prílohe D, bolo porovnanie výskytu PQS verzie GRCh38 s novou T2T sekvenciou. Nová sekvencia chromozómu X je kratšia o 1,14 % s miernym poklesom v zastúpení GC báz (Bohálová et al., 2021c). Zastúpenie PQS je pri prednastavenom G4Hunter skóre 1,2 je takmer identické, ale rozdiel medzi GRCh38 a T2T sekvenciami sa priamo úmerne zvyšuje s narastajúcim G4Hunter skóre v prospech novozostaveného T2T genómu. To poukazuje na podcenenie výskytu PQS v predchádzajúcich verziách genómu v dôsledku nedokonalých sekvenačných metód. Na obrázku 6 sa nachádza porovnanie rozloženia PQS sekvencií v genóme GRCh38 a T2T pri nastavení G4Hunter skóre na 3 a 3,5.



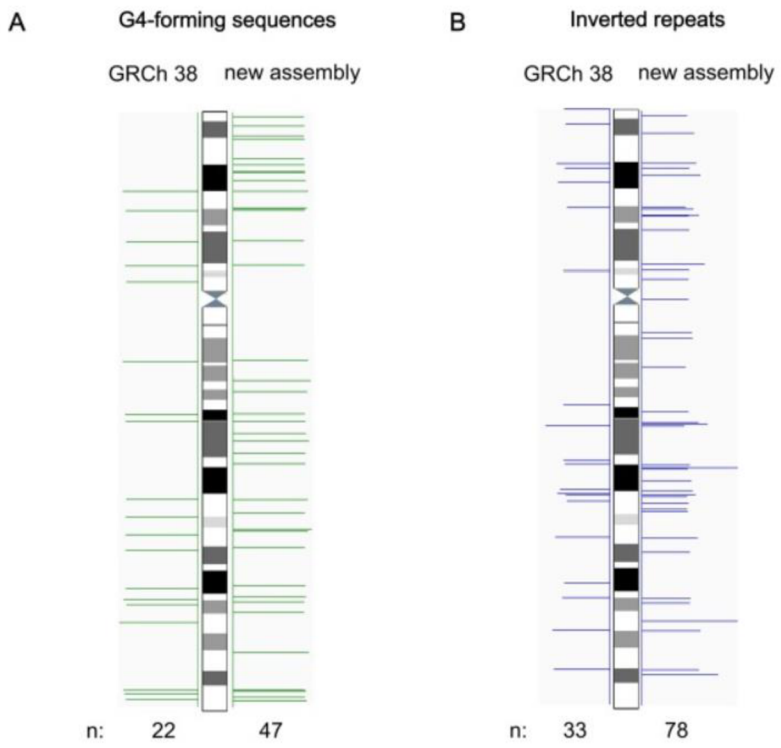
Obr. 6: Porovnanie výskytu PQS v chromozóme X genómu GRCh38 a nového assembly T2T.

(A) Schematické znázornenie G-tetrády a G4. Symbol M^+ označuje monovalentný ión. (B) Idiogram chromozómu X. (C) Lokalizácia PQS s nastavením parametrov G4Hunter softvéru s minimálním G4Hunter skóre 3,0 pre genóm GCRh38 (zelená) a T2T (modrá). (D) Lokalizácia PQS s nastavením parametrov minimálního G4Hunter skóre 3,5 pre genóm GRCh38 (zelená) a T2T (modrá). (E) Novoidentifikované PQS oblasti.

3.5 Publikácia 5. Kompletná sekvencia ľudského chromozómu 8 odhalila predchádzajúce podcenenie sekvencií s potenciálom tvorby G-kvadruplexu a inverzných repetícií.

Ďalším ľudským chromozómom, ktorého sekvencia bola plne prečítaná konzorciom T2T a zverejnená, bola sekvencia chromozómu 8 (Logsdon et al., 2021). Tentokrát sme sa rozhodli okrem analýzy PQS predikovať aj tvorbu inverzných repetíc a porovnať ich výskyt so staršou verziou ľudského genómu GRCh38 (Brazda et al., 2021). Publikovaný článok je súčasťou prílohy E.

Predchádzajúce závery z chromozómu X sa potvrdili nielen pre výskyt PQS, ale aj inverzné repetície na chromozóme 8. Článok je uvedený v prílohe B. V oboch prípadoch dochádzalo k väčším rozdielom medzi GRCh38 a novou genomickou sekvenciou pri nastavení prísnejších parametrov analýzy. Takýmto postupom sme schopní eliminovať falošne pozitívne výsledky a identifikovať stabilnejšie štruktúry. Lokalizácia PQS a IR na chromozóme 8 je znázornená na obrázku 7.



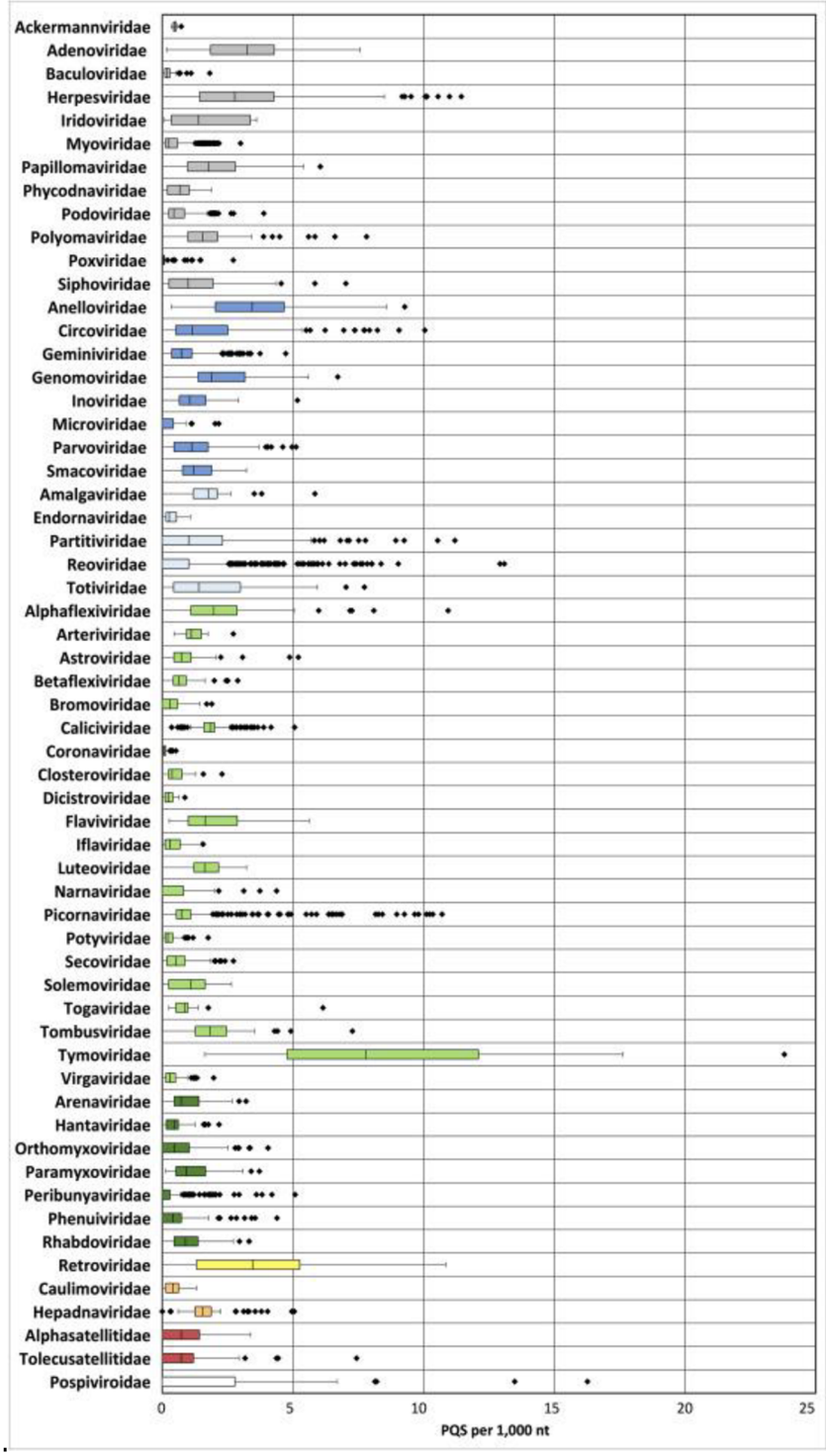
Obr. 7: Porovnanie výskytu PQS a inverzních repetíc v chromozóme 8 genómu GRCh38 a nového assembly T2T.

(A) Porovnanie výskytu PQS v chromozóme 8 genómu GRCh38 a nového assembly T2T s nastavením minimálního G4Hunter skóre na 3,5. (B) Porovnanie výskytu inverzních repetíc v chromozóme 8 genómu GRCh38 a nového assembly T2T s nastavením minimálnej dĺžky repetície na .

3.6 Publikácia 6. Analýza sekvencií potenciálne formujúcich G-kvadruplexy vo virusových genómoch odhalila ich koreláciu s typom infekcie

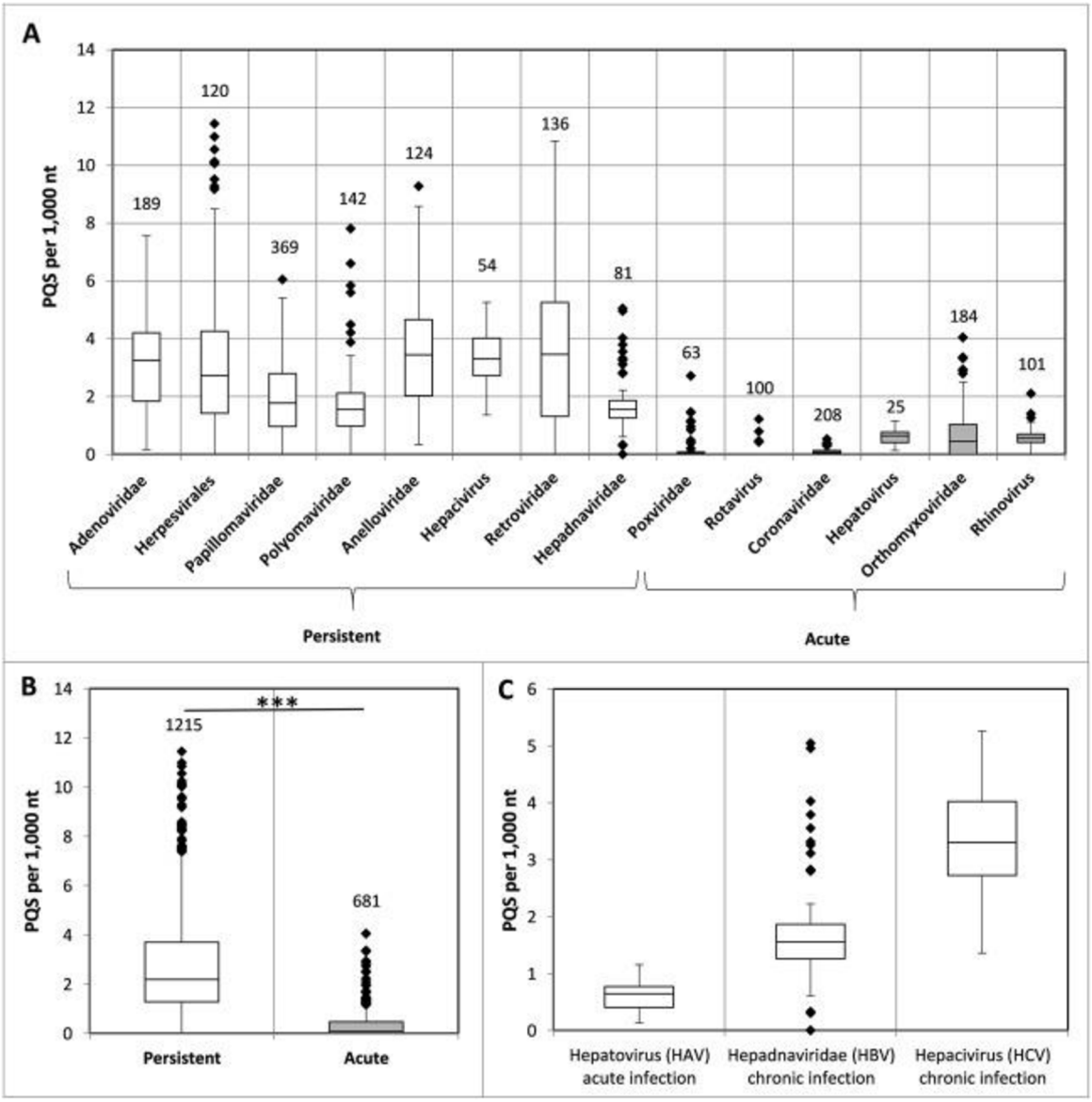
Cieľom publikácie, ktorá sa nachádza v prílohe F, bolo analyzovať výskyt PQS vo všetkých vírusových genómoch, ktorých sekvencia je dostupná v databáze NCBI (Bohálová et al., 2021b). Pomocou G4Hunter softvéru bolo zanalyzovaných 11 000 vírusových genómov, rozdelených podľa Baltimórovej klasifikácie a súčasnej klasifikácie medzinárodnej komisie taxonómie vírusov (ICTV) do fylogenetických skupín, ktoré sú znázornené na obrázku 8.

PQS sa častejšie vyskytovali v repetíciách, 5' a 3' neprekladaných oblastiach. Priemerné hodnoty frekvencií PQS sa líšili medzi skupinami Baltimórovej klasifikácie. Pozoruhodné je, že v rámci jedného typu nukleovej kyseliny sa vyskytovali fylogenetické skupiny s veľkými rozdielmi v priemerných PQS frekvenciách. Našou prvou hypotézou bolo, že fenomén môže byť spôsobený intracelulárnym výskytom vírusu v hostiteľskej bunke alebo bunkovou afinitou vírusu. Oba varianty boli testované, ale nepreukázal sa štatisticky významný rozdiel medzi týmito skupinami. Vzhľadom na predchádzajúce znalosti o zapojení G4 v regulácii latentného štádia, ktoré boli zhrnuté v tejto dizertačnej práci v časti 1.2.2, bolo našou ďalšou hypotézou rozdelenie podľa typu infekcie, ktorú skupina vírusov dominantne spôsobuje. Ako je diskutované v článku, delenie vírusov na tie, ktoré spôsobujú nejaké formu perzistentnej infekcie alebo výlučne akútne infekcie nie je a nemôže byť jednoznačné, keďže typ infekcie môže závisieť aj na bunkovej afinite vírusu, imunitnom systéme hostiteľa a v niektorých prípadoch aj na sérotype daného druhu. Z tohto dôvodu boli vybraté iba extrémne skupiny, pri ktorých je preukázateľne častejší jeden typ infekcie. Ich prehľad je zhrnutý na obrázku 9. Následne bolo potvrdené, že frekvencia PQS skutočne koreluje s typom infekcie; vyššie frekvencie PQS boli nájdené v skupinách vírusov, o ktorých je známe, že spôsobujú perzistentné, latentné alebo chronické infekcie pri hostiteľoch z *Metazoa*. Do tejto skupiny vírusov patria rády *Anelloviridae, Retroviridae, Papillomaviridae* či *Hepadnaviridae*. Na druhej strane frekvencia PQS bola veľmi nízka pri skupinách vírusov spôsobujúcich akútne infekcie, ako je *Rotavirus, Hepatovirus, Coronaviridae*. Porovnanie frekvencie PQS v týchto skupinách infikujúcich hostiteľov *Metazoa* je na obrázku 9. Najlepšie sa demonštruje tento fenomén na vírusoch spôsobujúcich hepatitídu, pri ktorých je známe, že vírus hepatitídy A (HAV) spôsobuje akútne infekcie, zatiaľ čo vírusy hepatitídy B (HBV) a C (HCV) spôsobujú u ľudí hepatitídu chronického typu.



Obr. 8: Prehľad PQS frekvencií vo vírusových čeľadiach.

Čeľade sú farebne rozlíšené na satelity, viroidy a vírusy, ktoré sú ďalej rozdelené do tried podľa Baltimórovej klasifikácie: I. šedá, II. tmavomodrá, III. svetlomodrá, IV. svetlozelená, V. tmavozelená. VI. žltá, VII. oranžová, satelity – červená, viroidy-biela. Os x predstavuje priemerný počet PQS na 1000 nt.



Obr. 9: Porovnanie frekvencií PQS vírusov spôsobujúcich akútne a latentné infekcie.

Os y predstavuje priemerný počet PQS na 1000 nt. (A)Vybrané skupiny vírusov boli kategorizované podľa dominantného typu infekcie. (B) Vírusy spôsobujúce väčšinou perzistentnú infekciu (biele) obsahujú štatisticky významne vyššie frekvencie PQS v porovnaní s akútne infikujúcimi vírusmi (sivé). (C) Frekvencie PQS troch vírusových skupín spôsobujúcich akútnu alebo chronickú hepatitídu.

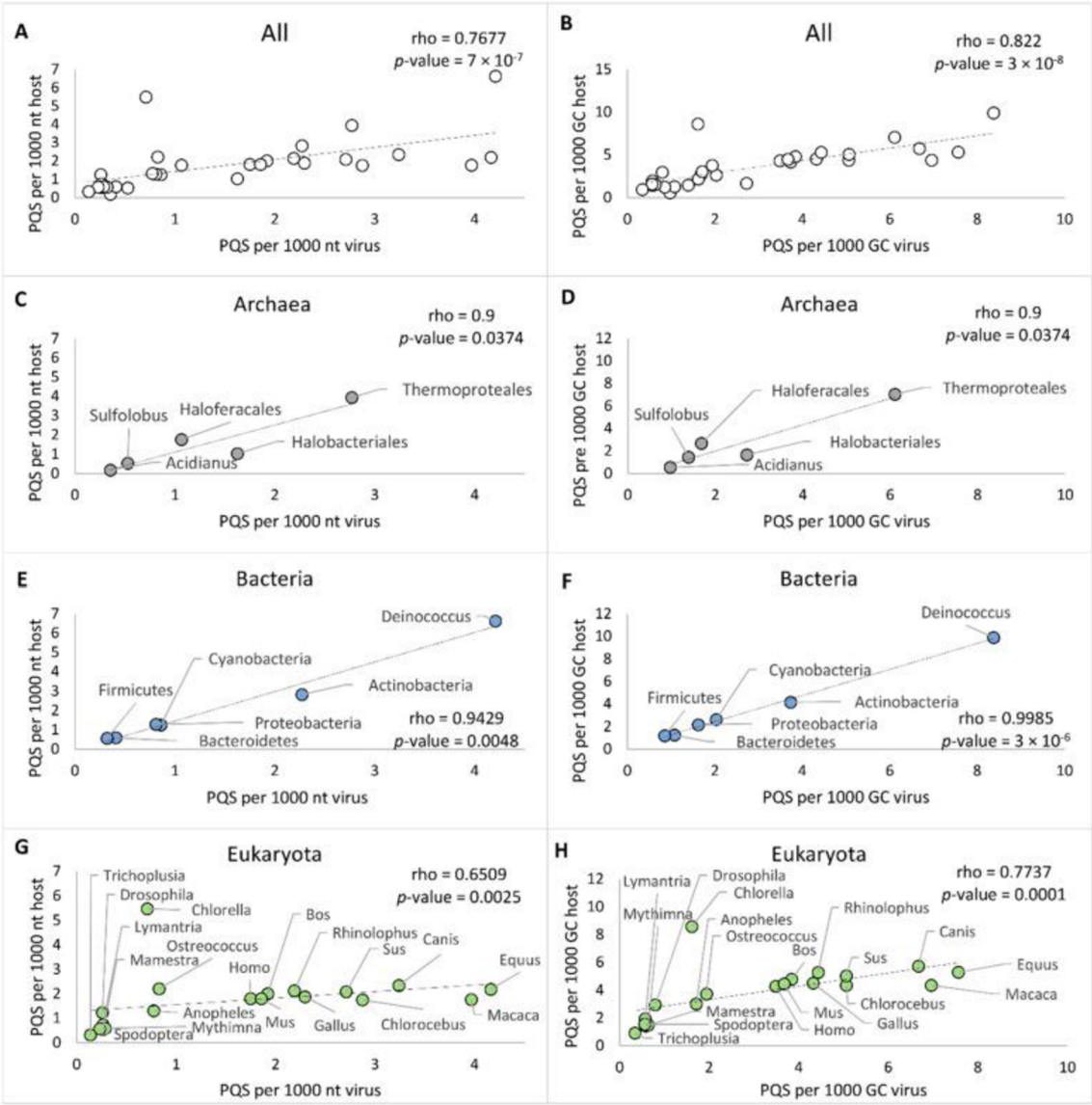
Našou hypotézou je, že eliminácia G4s z genómu vírusu môže predstavovať výhodu pre jednoduchšiu replikáciu v hostiteľskej bunke. Na druhej strane, G4 vo vírusovom genóme môže mimikovať genóm hostiteľa, a tak napomôcť jeho tolerancii a pri rozvoji perzistentných, latentných a chronických infekcií.

3.7 Publikácia 7. Pozorovanie koevolúcie dsDNA vírusov a ich hostiteľov skrz koreláciu sekvencii potenciálne formujúcich G-kvadruplexy

Keďže interakcie patogéna s hostiteľom predstavujú silný mechanizmus selekčného tlaku, hypotézou ktorá viedla k príprave článku v prílohe G bolo, že výskyt PQS vírusov kopíruje vzorec výskytu PQS hostiteľa (Bohálová et al., 2021a). Predchádzajúce publikácie popisovali rovnaké sekvencie v slučkách PQS v genómoch zástupcov čeľade *Herpesviridae* a príslušných hostiteľov z podkmeňa stavovce (Puig Lombardi et al., 2019).

Výber sme obmedzili na rovnakú nukleovú kyselinu hostiteľa a patogéna, a teda boli analyzované iba dsDNA vírusy z I. triedy Baltimórovej klasifikácie. Prvým dôkazom potvrdzujúcim našu hypotézu bola korelujúca PQS frekvencia v genómoch bakteriofágov a baktérií. Prehľadová štúdia PQS v genómoch baktérií, ktorá bola na porovnanie použitá, bola pripravená a publikovaná už v roku 2019 rovnakou metodológiou (Bartas et al., 2019). Neskôr sa rovnaký fenomén potvrdil aj v doménach archaea a vybraných zástupcov eukaryota. Prehľadová štúdia výskytu PQS v genómoch všetkých sekvenovaných zástupcov domény archaea bola takisto v čase prípravy tohto článku už publikovaná (Brázda et al., 2020). Zástupcovia domény eukaryota boli vybratý na základe výsledkov z Virus-Host databázy (Mihara et al., 2016). Pre každú vírusovú čeľaď bolo vybraných 10 najfrekventovanejších hostiteľov a tie boli analyzované, pokiaľ ich sekvencie boli v databáze NCBI dostupné.

Okrem počtu PQS na 1000 bp sme sa rozhodli porovnať aj počet PQS vztiahnutý na počet GC báz vírusu a príslušného hostiteľa (Obr. 10). Aj po normalizácií na obsah GC dochádzalo k významnej korelácii všetkých pozorovaných skupín. Silný vzťah medzi PQS vo vírusovom a jeho hostiteľskom genóme naznačuje ich úzku koevolúciu a evolučne recipročné napodobňovanie organizácie genómu.



Obr. 10: Korelácia PQS frekvencií medzi vírusom a jeho hostiteľom.

Korelácia medzi vírusom a hostiteľom boli vyhodnocovaná buď na základe frekvencie PQS na 1000 bp alebo na základe frekvencie PQS na 1000 GC báz. (A) Všetky páry hostiteľ-vírus, počet PQS na 1000 bp; (B) Všetky páry hostiteľ-vírus, počet PQS na 1000 GC; (C) páry Archaea-vírus, počet PQS na 1000 bp; (D) páry Archaea-vírus, počet PQS na 1000 GC; (E) páry baktérie – vírus, počet PQS na 1000 bp; (F) páry baktérie-vírus, počet PQS na 1000 GC; (G) páry eukaryota-vírus, počet PQS na 1000 bp; (H) Páry Eukaryota-vírus, počet PQS na 1000 GC.

4 Záver

Teoretická časť dizertačnej práce *Analýza DNA väzbových proteínov so zameraním na ich interakcie s lokálnymi sekundárnymi štruktúrami DNA* bola venovaná popisu sekundárnych lokálnych štruktúr s užším zameraním na G4 a krížové štruktúry v ľudskom genóme a vírusoch. Obe štruktúry majú vplyv na zásadné molekulárne procesy. Predstavujú nádejné terapeutické ciele v nádorovej biológií, pri rôznych vírusových infekciách a neurodegeneratívnych ochoreniach. Boli diskutované klinické aplikácie sekundárnych štruktúr a s nimi súvisiacich molekúl v humánnej medicíne.

Interakcia proteínov s nukleovými kyselinami je kľúčovou vlastnosťou pri zásadných molekulárnych dejoch, akými sú replikácia, génová expresia, oprava DNA či imunitná reakcia. Podrobne boli popísané dva proteíny, senzory nukleových kyselín, preferenčne viažuce sa na sekundárne lokálne štruktúry nukleových kyslín, senzor cudzorodej DNA IFI16 a transkripčný faktor proteín p53.

Praktická časť dizertačnej práce bola rozdelená na experimentálne a bioinformatické analýzy.

Dve publikácie boli zamerané na experimenty proteínu p53 v kvasinkovom izogénnom reportérovom systéme. G4 začlenený pred promótorom reportérového génu podľa očakávaní spôsoboval represiu transaktivácie reportérového génu proteínom p53. Bližšia pozornosť bola venovaná vplyvu izoforiem p53, členom rodiny p53 a mutantným formách proteínu p53 na transaktivačný potenciál p53. G4 začlenený pred RE v smere transkripcie zásadne ovplyvnil správanie heterotetramérov Δ133p53α a Δ160p53α s wtp53. Topológia DNA v okolí responzívnej sekvencie p53 tak predstavuje ďalšiu úroveň v regulácii signálnych dráh, do ktorých je tento transkripčný faktor zapojený.

Interakcie IFI16 s DNA boli skúmané pomocou mikroskopie atomárnych síl. Typ pozorovanej interakcie závisel od superhelicity DNA. Lineárna DNA indukovala tvorbu proteínových oligomérov, zatiaľ čo superhelikálna DNA tvorbu globulárnych komplexov proteínu s DNA o nižšej molekulovej hmotnosti. Keďže tvorba proteínového oligoméru je kľúčovou pre aktiváciu nešpecifickej imunitnej reakcie proteínom IFI16, interakcie so superhelikálnou DNA môžu predstavovať mechanizmus, pomocou ktorého proteín špecificky reguluje génovú expresiu. Našou hypotézou je, že proteín IFI16 môže vykonávať reštrikciu vírusovej infekcie v duálnom režime, buď navodí nešpecifickú imunitnú reakciu prostredníctvom tvorby proteínového oligoméru alebo špecifickou väzbou na sekundárne lokálne štruktúry zabezpečí represiu génovej expresie.

Cieľom druhej časti dizertačnej práce bolo pomocou bioinformatických metód identifikovať skupiny vírusov a cieľové oblasti ľudského genómu zaujímavé z hľadiska štúdia proteínov interagujúcich s DNA. Bioinformatická časť

obsahovala 4 články, v ktorých boli publikované prehľadové štúdie putatívnych sekundárnych lokálnych štruktúr nukleových kyselín. Dve z nich sa zameriavali na novo sekvenovaný ľudský chromozóm X a chromozóm 8 (Bohálová et al., 2021c; Brazda et al., 2021), ktorých sekvencia bola zverejnená T2T konzorciom (Logsdon et al., 2021; Miga et al., 2020). Bolo zistené, že v chýbajúcich úsekoch DNA predchádzajúcej genómovej sekvencii boli vysoko zastúpené stabilné PQS a inverzné repetície.

Zvyšné bioinformatické práce popisovali výskyt G4 a inverzných repetíc vo vírusových genómoch. Bola publikovaná prehľadová štúdia výskytu PQS vo všetkých sekvenovaných genómoch vírusov. Bolo zistené, že priemerná frekvencia PQS je vyššia vo vírusových genómoch, ktoré majú tendenciu spôsobovať perzistentné infekcie v jedincoch bez imunitnej nedostatočnosti. Ďalšia publikácia popisovala koevolúciu vírusu a hostiteľa z pohľadu PQS. Bola zistená korelácia medzi PQS frekvenciou vo vírusovom a hostiteľskom genóme, čo poukazuje na evolučnú konzervovanosť G4.

Použité zdroje

Abnous, K., Danesh, N.M., Ramezani, M., Charbgoo, F., Bahreyni, A., Taghdisi, S.M., 2018. Targeted delivery of doxorubicin to cancer cells by a cruciform DNA nanostructure composed of AS1411 and FOXM1 aptamers. Expert Opin. Drug Deliv. 15, 1045–1052. https://doi.org/10.1080/17425247.2018.1530656

Abrescia, N.G.A., González, C., Gouyette, C., Subirana, J.A., 2004. X-ray and NMR Studies of the DNA Oligomer d(ATATAT): Hoogsteen Base Pairing in Duplex DNA. Biochemistry 43, 4092–4100. https://doi.org/10.1021/bi0355140

Adámik, M., Kejnovská, I., Bažantová, P., Petr, M., Renčiuk, D., Vorlíčková, M., Brázdová, M., 2016. p53 binds human telomeric G-quadruplex in vitro. Biochimie 128–129, 83–91. https://doi.org/10.1016/j.biochi.2016.07.004

Ajoge, H.O., Kohio, H.P., Paparisto, E., Coleman, M.D., Wong, K., Tom, S.K., Bain, K.L., Berry, C.C., Arts, E.J., Barr, S.D., 2022. G-Quadruplex DNA and Other Non-Canonical B-Form DNA Motifs Influence Productive and Latent HIV-1 Integration and Reactivation Potential. Viruses 14, 2494. https://doi.org/10.3390/v14112494

Almine, J.F., O'Hare, C.A.J., Dunphy, G., Haga, I.R., Naik, R.J., Atrih, A., Connolly, D.J., Taylor, J., Kelsall, I.R., Bowie, A.G., Beard, P.M., Unterholzner, L., 2017. IFI16 and cGAS cooperate in the activation of STING during DNA sensing in human keratinocytes. Nat. Commun. 8. https://doi.org/10.1038/ncomms14392

Altemose, N., Logsdon, G.A., Bzikadze, A.V., Sidhwani, P., Langley, S.A., Caldas, G.V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., Sauria, M.E.G., Borchers, M., Gershman, A., Mikheenko, A., Shepelev, V.A., Dvorkina, T., Kunyavskaya, O., Vollger, M.R., Rhie, A., McCartney, A.M., Asri, M., Lorig-Roach, R., Shafin, K., Lucas, J.K., Aganezov, S., Olson, D., de Lima, L.G., Potapova, T., Hartley, G.A., Haukness, M., Kerpedjiev, P., Gusev, F., Tigyi, K., Brooks, S., Young, A., Nurk, S., Koren, S., Salama, S.R., Paten, B., Rogaev, E.I., Streets, A., Karpen, G.H., Dernburg, A.F., Sullivan, B.A., Straight, A.F., Wheeler, T.J., Gerton, J.L., Eichler, E.E., Phillippy, A.M., Timp, W., Dennis, M.Y., O'Neill, R.J., Zook, J.M., Schatz, M.C., Pevzner, P.A., Diekhans, M., Langley, C.H., Alexandrov, I.A., Miga, K.H., 2022. Complete genomic and epigenetic maps of human centromeres. Science 376, eabl4178. https://doi.org/10.1126/science.abl4178

Alunno, A., Caneparo, V., Bistoni, O., Caterbi, S., Terenzi, R., Gariglio, M., Bartoloni, E., Manzo, A., Landolfo, S., Gerli, R., 2016. Circulating Interferon-Inducible Protein IFI16 Correlates With Clinical and Serological Features in Rheumatoid Arthritis. Arthritis Care Res. 68, 440–445. https://doi.org/10.1002/acr.22695

Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A., Yang, D., 2006. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res. 34, 2723–2735. https://doi.org/10.1093/nar/gkl348

Amrane, S., Kerkour, A., Bedrat, A., Vialet, B., Andreola, M.-L., Mergny, J.-L., 2014. Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a

potential target for anti-HIV drug development. J. Am. Chem. Soc. 136, 5249–5252. https://doi.org/10.1021/ja501500c
Anbarasan, T., Bourdon, J.-C., 2019. The Emerging Landscape of p53 Isoforms in Physiology, Cancer and Degenerative Diseases. Int. J. Mol. Sci. 20, 6257. https://doi.org/10.3390/ijms20246257
Antiochos, B., Trejo-Zambrano, D., Fenaroli, P., Rosenberg, A., Baer, A., Garg, A., Sohn, J., Li, J., Petri, M., Goldman, D.W., Mecoli, C., Casciola-Rosen, L., Rosen, A., 2022. The DNA sensors AIM2 and IFI16 are SLE autoantigens that bind neutrophil extracellular traps. eLife 11, e72103. https://doi.org/10.7554/eLife.72103
Arroyo-Olarte, R.D., Bravo Rodríguez, R., Morales-Ríos, E., 2021. Genome Editing in Bacteria: CRISPR-Cas and Beyond. Microorganisms 9, 844. https://doi.org/10.3390/microorganisms9040844
Awadasseid, A., Ma, X., Wu, Y., Zhang, W., 2021. G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy. Biomed. Pharmacother. 139, 111550. https://doi.org/10.1016/j.biopha.2021.111550
Baer, A.N., Petri, M., Sohn, J., Rosen, A., Casciola-Rosen, L., 2016. Association of Antibodies to Interferon-Inducible Protein-16 With Markers of More Severe Disease in Primary Sjögren's Syndrome. Arthritis Care Res. 68, 254–260. https://doi.org/10.1002/acr.22632
Bansal, A., Kaushik, S., Kukreti, S., 2022. Non-canonical DNA structures: Diversity and disease association. Front. Genet. 13, 959258. https://doi.org/10.3389/fgene.2022.959258
Barrangou, R., Doudna, J.A., 2016. Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. 34, 933–941. https://doi.org/10.1038/nbt.3659
Barshai, M., Aubert, A., Orenstein, Y., 2022. G4detector: Convolutional Neural Network to Predict DNA G-Quadruplexes. IEEE/ACM Trans. Comput. Biol. Bioinform. 19, 1946–1955. https://doi.org/10.1109/TCBB.2021.3073595
Bartas, M., Brazda, V., Bohálová, N., Cantara, A., Volná, A., Stachurová, T., Malachová, K., Jagelská, E.B., Porubiaková, O., Červeň, J., Pečinka, P., 2020. In-depth Bioinformatic Analyses of Nidovirales Including Human SARS-CoV-2, SARS-CoV, MERS-CoV Viruses Suggest Important Roles of Noncanonical Nucleic Acid Structures in Their Lifecycles. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.01583
Bartas, M., Brázda, V., Karlický, V., Červeň, J., Pečinka, P., 2018. Bioinformatics analyses and in vitro evidence for five and six stacked G-quadruplex forming sequences. Biochimie 150, 70–75. https://doi.org/10.1016/j.biochi.2018.05.002
Bartas, M., Čutová, M., Brázda, V., Kaura, P., Šťastný, J., Kolomazník, J., Coufal, J., Goswami, P., Červeň, J., Pečinka, P., 2019. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. Molecules 24, 1711. https://doi.org/10.3390/molecules24091711
Bartas, M., Goswami, P., Lexa, M., Červeň, J., Volná, A., Fojta, M., Brázda, V., Pečinka, P., 2021. Letter to the Editor: Significant mutation enrichment in inverted repeat

sites of new SARS-CoV-2 strains. Brief. Bioinform. bbab129. https://doi.org/10.1093/bib/bbab129

Bazzicalupi, C., Ferraroni, M., Bilia, A.R., Scheggi, F., Gratteri, P., 2013. The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. Nucleic Acids Res. 41, 632–638. https://doi.org/10.1093/nar/gks1001

Beaudoin, J.-D., Jodoin, R., Perreault, J.-P., 2014. New scoring system to identify RNA G-quadruplex folding. Nucleic Acids Res. 42, 1209–1223. https://doi.org/10.1093/nar/gkt904

Bedrat, A., Lacroix, L., Mergny, J.-L., 2016. Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res 44, 1746–1759. https://doi.org/10.1093/nar/gkw006

Berns, K.I., 2020. The Unusual Properties of the AAV Inverted Terminal Repeat. Hum. Gene Ther. 31, 518–523. https://doi.org/10.1089/hum.2020.017

Berselli, M., Lavezzo, E., Toppo, S., 2020. QPARSE: searching for long-looped or multimeric G-quadruplexes potentially distinctive and druggable. Bioinforma. Oxf. Engl. 36, 393–399. https://doi.org/10.1093/bioinformatics/btz569

Bian, W.-X., Xie, Y., Wang, X.-N., Xu, G.-H., Fu, B.-S., Li, S., Long, G., Zhou, X., Zhang, X.-L., 2019. Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. Nucleic Acids Res. 47, 56–68. https://doi.org/10.1093/nar/gky1177

Biswas, B., Kandpal, M., Vivekanandan, P., 2017. A G-quadruplex motif in an envelope gene promoter regulates transcription and virion secretion in HBV genotype B. Nucleic Acids Res. 45, 11268–11280. https://doi.org/10.1093/nar/gkx823

Bohálová, N., Cantara, A., Bartas, M., Kaura, P., Šťastný, J., Pečinka, P., Fojta, M., Brázda, V., 2021a. Tracing dsDNA Virus–Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences. Int. J. Mol. Sci. 22, 3433. https://doi.org/10.3390/ijms22073433

Bohálová, N., Cantara, A., Bartas, M., Kaura, P., Šťastný, J., Pečinka, P., Fojta, M., Mergny, J.-L., Brázda, V., 2021b. Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. Biochimie 186, 13–27. https://doi.org/10.1016/j.biochi.2021.03.017

Bohálová, N., Mergny, J.-L., Brázda, V., 2021c. Novel G-quadruplex prone sequences emerge in the complete assembly of the human X chromosome. Biochimie 191, 87–90. https://doi.org/10.1016/j.biochi.2021.09.004

Bosso, M., Prelli Bozzo, C., Hotter, D., Volcic, M., Stürzel, C.M., Rammelt, A., Ni, Y., Urban, S., Becker, M., Schelhaas, M., Wittmann, S., Christensen, M.H., Schmidt, F.I., Gramberg, T., Sparrer, K.M.J., Sauter, D., Kirchhoff, F., 2020. Nuclear PYHIN proteins target the host transcription factor Sp1 thereby restricting HIV-1 in human macrophages and CD4+ T cells. PLOS Pathog. 16, e1008752. https://doi.org/10.1371/journal.ppat.1008752

Bourdon, S., Herviou, P., Dumas, L., Destefanis, E., Zen, A., Cammas, A., Millevoi, S., Dassi, E., 2023. QUADRatlas: the RNA G-quadruplex and RG4-binding proteins

database. Nucleic Acids Res. 51, D240–D247. https://doi.org/10.1093/nar/gkac782
Bowater, R.P., Bohálová, N., Brázda, V., 2022. Interaction of Proteins with Inverted Repeats and Cruciform Structures in Nucleic Acids. Int. J. Mol. Sci. 23, 6171. https://doi.org/10.3390/ijms23116171
Brazda, V., Bohalova, N., Bowater, R.P., 2021. New telomere to telomere assembly of human chromosome 8 reveals a previous underestimation of G-quadruplex forming sequences and inverted repeats. Gene 146058. https://doi.org/10.1016/j.gene.2021.146058
Brázda, V., Coufal, J., Liao, J., Arrowsmith, C., 2012. Preferential binding of IFI16 protein to cruciform structure and superhelical DNA. Biochem. Biophys. Res. Commun. 422, 716–20. https://doi.org/10.1016/j.bbrc.2012.05.065
Brázda, V., Fojta, M., 2019. The Rich World of p53 DNA Binding Targets: The Role of DNA Structure. Int. J. Mol. Sci. 20, E5605. https://doi.org/10.3390/ijms20225605
Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Šťastný, J., Mergny, J.-L., 2019. G4Hunter web application: a web server for G-quadruplex prediction. Bioinformatics 35, 3493–3495. https://doi.org/10.1093/bioinformatics/btz087
Brázda, V., Kolomazník, J., Lýsek, J., Hároníková, L., Coufal, J., Šťastný, J., 2016. Palindrome analyser - A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. Biochem. Biophys. Res. Commun. 478, 1739–1745. https://doi.org/10.1016/j.bbrc.2016.09.015
Brázda, V., Luo, Y., Bartas, M., Kaura, P., Porubiaková, O., Šťastný, J., Pečinka, P., Verga, D., Da Cunha, V., Takahashi, T.S., Forterre, P., Myllykallio, H., Fojta, M., Mergny, J.-L., 2020. G-Quadruplexes in the Archaea Domain. Biomolecules 10, 1349. https://doi.org/10.3390/biom10091349
Brázda, V., Porubiaková, O., Cantara, A., Bohálová, N., Coufal, J., Bartas, M., Fojta, M., Mergny, J.-L., 2021. G-quadruplexes in H1N1 influenza genomes. BMC Genomics 22, 77. https://doi.org/10.1186/s12864-021-07377-9
Cagirici, H.B., Budak, H., Sen, T.Z., 2022. G4Boost: a machine learning-based tool for quadruplex identification and stability prediction. BMC Bioinformatics 23, 240. https://doi.org/10.1186/s12859-022-04782-z
Candeias, M.M., Hagiwara, M., Matsuda, M., 2016. Cancer-specific mutations in p53 induce the translation of Δ160p53 promoting tumorigenesis. EMBO Rep. 17, 1542–1551. https://doi.org/10.15252/embr.201541956
Cao, T., Shao, S., Li, B., Jin, L., Lei, J., Qiao, H., Wang, G., 2016. Up-regulation of Interferon-inducible protein 16 contributes to psoriasis by modulating chemokine production in keratinocytes. Sci. Rep. 6, 25381. https://doi.org/10.1038/srep25381
Caposio, P., Gugliesi, F., Zannetti, C., Sponza, S., Mondini, M., Medico, E., Hiscott, J., Young, H.A., Gribaudo, G., Gariglio, M., Landolfo, S., 2007. A Novel Role of the Interferon-inducible Protein IFI16 as Inducer of Proinflammatory Molecules in

Endothelial Cells. J. Biol. Chem. 282, 33515–33529. https://doi.org/10.1074/jbc.M701846200
Chaires, J.B., Gray, R.D., Dean, W.L., Monsen, R., DeLeeuw, L.W., Stribinskis, V., Trent, J.O., 2020. Human POT1 unfolds G-quadruplexes by conformational selection. Nucleic Acids Res. 48, 4976–4991. https://doi.org/10.1093/nar/gkaa202
Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., Balasubramanian, S., 2015. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. 33, 877–881. https://doi.org/10.1038/nbt.3295
Chan, N.L.S., Hou, C., Zhang, T., Yuan, F., Machwe, A., Huang, J., Orren, D.K., Gu, L., Li, G.-M., 2012. The Werner syndrome protein promotes CAG/CTG repeat stability by resolving large (CAG)(n)/(CTG)(n) hairpins. J. Biol. Chem. 287, 30151–30156. https://doi.org/10.1074/jbc.M112.389791
Charpentier, E., Doudna, J.A., 2013. Biotechnology: Rewriting a genome. Nature 495, 50–51. https://doi.org/10.1038/495050a
Cogoi, S., Xodo, L.E., 2006. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. 34, 2536–2549. https://doi.org/10.1093/nar/gkl286
Correll-Tash, S., Lilley, B., Salmons Iv, H., Mlynarski, E., Franconi, C.P., McNamara, M., Woodbury, C., Easley, C.A., Emanuel, B.S., 2021. Double strand breaks (DSBs) as indicators of genomic instability in PATRR-mediated translocations. Hum. Mol. Genet. 29, 3872–3881. https://doi.org/10.1093/hmg/ddaa251
Costa, S., Borgogna, C., Mondini, M., De Andrea, M., Meroni, P. l., Berti, E., Gariglio, M., Landolfo, S., 2011. Redistribution of the nuclear protein IFI16 into the cytoplasm of ultraviolet B-exposed keratinocytes as a mechanism of autoantigen processing. Br. J. Dermatol. 164, 282–290. https://doi.org/10.1111/j.1365-2133.2010.10097.x
Coufal, J., Jagelská, E.B., Liao, J.C.C., Brázda, V., 2013. Preferential binding of p53 tumor suppressor to p21 promoter sites that contain inverted repeats capable of forming cruciform structure. Biochem. Biophys. Res. Commun. 441, 83–88. https://doi.org/10.1016/j.bbrc.2013.10.015
da Rosa, G., Grille, L., Calzada, V., Ahmad, K., Arcon, J.P., Battistini, F., Bayarri, G., Bishop, T., Carloni, P., Cheatham III, T., Collepardo-Guevara, R., Czub, J., Espinosa, J.R., Galindo-Murillo, R., Harris, S.A., Hospital, A., Laughton, C., Maddocks, J.H., Noy, A., Orozco, M., Pasi, M., Pérez, A., Petkevičiūtė-Gerlach, D., Sharma, R., Sun, R., Dans, P.D., 2021. Sequence-dependent structural properties of B-DNA: what have we learned in 40 years? Biophys. Rev. 13, 995–1005. https://doi.org/10.1007/s12551-021-00893-8
Dabral, P., Babu, J., Zareie, A., Verma, S.C., 2020. LANA and hnRNP A1 Regulate the Translation of LANA mRNA through G-Quadruplexes. J. Virol. 94. https://doi.org/10.1128/JVI.01508-19
Datta, A., Brosh, R.M., 2018. New Insights Into DNA Helicases as Druggable Targets for Cancer Therapy. Front. Mol. Biosci. 5, 59. https://doi.org/10.3389/fmolb.2018.00059

Dawson, M.J., Trapani, J.A., 1996. HIN-200: a novel family of IFN-inducible nuclear proteins expressed in leukocytes. J. Leukoc. Biol. 60, 310–316. https://doi.org/10.1002/jlb.60.3.310

Dawson, M.J., Trapani, J.A., 1995. The Interferon Inducible Autoantigen, IFI 16: Localization to the Nucleolus and Identification of a DNA-Binding Domain. Biochem. Biophys. Res. Commun. 214, 152–162. https://doi.org/10.1006/bbrc.1995.2269

Day, H.A., Pavlou, P., Waller, Z.A.E., 2014. i-Motif DNA: Structure, stability and targeting with ligands. Bioorg. Med. Chem., Unlocking Nature through Chemistry 22, 4407–4418. https://doi.org/10.1016/j.bmc.2014.05.047

Dell'Oste, V., Gatti, D., Gugliesi, F., Andrea, M.D., Bawadekar, M., Cigno, I.L., Biolatti, M., Vallino, M., Marschall, M., Gariglio, M., Landolfo, S., 2014. Innate Nuclear Sensor IFI16 Translocates into the Cytoplasm during the Early Stage of In Vitro Human Cytomegalovirus Infection and Is Entrapped in the Egressing Virions during the Late Stage. J. Virol. 88, 13.

Dexheimer, T.S., Sun, D., Hurley, L.H., 2006. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2 P1 Promoter. J. Am. Chem. Soc. 128, 5404–5415. https://doi.org/10.1021/ja0563861

Ding, Y., Fleming, A.M., Burrows, C.J., 2018. Case studies on potential G-quadruplex-forming sequences from the bacterial orders Deinococcales and Thermales derived from a survey of published genomes. Sci. Rep. 8, 15679. https://doi.org/10.1038/s41598-018-33944-4

Doluca, O., 2019. G4Catchall: A G-quadruplex prediction approach considering atypical features. J. Theor. Biol. 463, 92–98. https://doi.org/10.1016/j.jtbi.2018.12.007

Eldar, A., Rozenberg, H., Diskin-Posner, Y., Rohs, R., Shakked, Z., 2013. Structural studies of p53 inactivation by DNA-contact mutations and its rescue by suppressor mutations via alternative protein-DNA interactions. Nucleic Acids Res. 41, 8748–8759. https://doi.org/10.1093/nar/gkt630

Ellington, A.D., Szostak, J.W., 1990. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818–822. https://doi.org/10.1038/346818a0

Fleming, A.M., Ding, Y., Alenko, A., Burrows, C.J., 2016. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. ACS Infect. Dis. 2, 674–681. https://doi.org/10.1021/acsinfecdis.6b00109

Fleming, A.M., Nguyen, N.L.B., Burrows, C.J., 2019. Colocalization of m6A and G-Quadruplex-Forming Sequences in Viral RNA (HIV, Zika, Hepatitis B, and SV40) Suggests Topological Control of Adenosine N6-Methylation. ACS Cent. Sci. 5, 218–228. https://doi.org/10.1021/acscentsci.8b00963

Frank-Kamenetskii, M.D., Mirkin, S.M., 1995. Triplex DNA structures. Annu. Rev. Biochem. 64, 65–95. https://doi.org/10.1146/annurev.bi.64.070195.000433

Frasson, I., Soldà, P., Nadai, M., Lago, S., Richter, S.N., 2021. Parallel G-quadruplexes recruit the HSV-1 transcription factor ICP4 to promote viral transcription in

herpes virus-infected human cells. Commun. Biol. 4, 1–13. https://doi.org/10.1038/s42003-021-02035-y
Freed-Pastor, W.A., Prives, C., 2012. Mutant p53: one name, many proteins. Genes Dev. 26, 1268–1286. https://doi.org/10.1101/gad.190678.112
Ganapathiraju, M.K., Subramanian, S., Chaparala, S., Karunakaran, K.B., 2020. A reference catalog of DNA palindromes in the human genome and their variations in 1000 Genomes. Hum. Genome Var. 7, 40. https://doi.org/10.1038/s41439-020-00127-5
Garant, J.-M., Perreault, J.-P., Scott, M., 2018. G4RNA screener web server: User focused interface for RNA G-quadruplex prediction. Biochimie 151. https://doi.org/10.1016/j.biochi.2018.06.002
Gehring, K., Leroy, J.-L., Guéron, M., 1993. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. Nature 363, 561–565. https://doi.org/10.1038/363561a0
Goswami, P., Bartas, M., Lexa, M., Bohálová, N., Volná, A., Červeň, J., Červeňová, V., Pečinka, P., Špunda, V., Fojta, M., Brázda, V., 2021. SARS-CoV-2 hot-spot mutations are significantly enriched within inverted repeats and CpG island loci. Brief. Bioinform. 22, 1338–1345. https://doi.org/10.1093/bib/bbaa385
Gray, R.D., Chaires, J.B., 2008. Kinetics and mechanism of K+- and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures. Nucleic Acids Res. 36, 4191–4203. https://doi.org/10.1093/nar/gkn379
Griffante, G., Hewelt-Belka, W., Albano, C., Gugliesi, F., Pasquero, S., Castillo Pacheco, S.F., Bajetto, G., Porporato, P.E., Mina, E., Vallino, M., Krapp, C., Jakobsen, M.R., Purdy, J., von Einem, J., Landolfo, S., Dell'Oste, V., Biolatti, M., 2022. IFI16 Impacts Metabolic Reprogramming during Human Cytomegalovirus Infection. mBio 13, e0043522. https://doi.org/10.1128/mbio.00435-22
Guédin, A., Gros, J., Alberti, P., Mergny, J.-L., 2010. How long is too long? Effects of loop size on G-quadruplex stability. Nucleic Acids Res. 38, 7858–7868. https://doi.org/10.1093/nar/gkq639
Guiblet, W.M., Cremona, M.A., Harris, R.S., Chen, D., Eckert, K.A., Chiaromonte, F., Huang, Y.-F., Makova, K.D., 2021. Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. Nucleic Acids Res. 49, 1497–1516. https://doi.org/10.1093/nar/gkaa1269
Hadj Amor, I.Y., Smaoui, K., ChaabÃ¨ne, I., Mabrouk, I., Djemal, L., Elleuch, H., Allouche, M., Mokdad-Gargouri, R., Gargouri, A., 2008. Human p53 induces cell death and downregulates thioredoxin expression in *Saccharomyces cerevisiae*. FEMS Yeast Res. 8, 1254–1262. https://doi.org/10.1111/j.1567-1364.2008.00445.x
Hansen, K., Prabakaran, T., Laustsen, A., Jørgensen, S.E., Rahbæk, S.H., Jensen, S.B., Nielsen, R., Leber, J.H., Decker, T., Horan, K.A., Jakobsen, M.R., Paludan, S.R., 2014. Listeria monocytogenes induces IFNβ expression through an IFI16-, cGAS- and STING-dependent pathway. EMBO J. 33, 1654–1666. https://doi.org/10.15252/embj.201488029

Hároníková, L., Coufal, J., Kejnovská, I., Jagelská, E.B., Fojta, M., Dvořáková, P., Muller, P., Vojtesek, B., Brázda, V., 2016. IFI16 Preferentially Binds to DNA with Quadruplex Structure and Enhances DNA Quadruplex Formation. PLOS ONE 11, e0157156. https://doi.org/10.1371/journal.pone.0157156

Hognon, C., Miclot, T., Iriepa, C.G., Francés-Monerris, A., Grandemange, S., Terenzi, A., Marazzi, M., Barone, G., Monari, A., 2020. Role of RNA Guanine Quadruplexes in Favoring the Dimerization of SARS Unique Domain in Coronaviruses (preprint). Biophysics. https://doi.org/10.1101/2020.04.07.029447

Hon, J., Martínek, T., Zendulka, J., Lexa, M., 2017. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. Bioinformatics 33, 3373–3379. https://doi.org/10.1093/bioinformatics/btx413

Honda, R., Tanaka, H., Yasuda, H., 1997. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. FEBS Lett. 420, 25–27. https://doi.org/10.1016/s0014-5793(97)01480-4

Hoogsteen, K., 1963. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. Acta Crystallogr. 16, 907–916. https://doi.org/10.1107/S0365110X63002437

Howard, T.R., Lum, K.K., Kennedy, M.A., Cristea, I.M., 2022. The Nuclear DNA Sensor IFI16 Indiscriminately Binds to and Diminishes Accessibility of the HSV-1 Genome to Suppress Infection. mSystems 7, e00198-22. https://doi.org/10.1128/msystems.00198-22

https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus, 2023. https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus [WWW Document]. URL https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus (accessed 6.11.23).

Huppert, J.L., Balasubramanian, S., 2007. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 35, 406–413. https://doi.org/10.1093/nar/gkl1057

Huppert, J.L., Balasubramanian, S., 2005. Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 33, 2908–2916. https://doi.org/10.1093/nar/gki609

Ivanov, V.I., Minchenkova, L.E., Minyat, E.E., Frank-Kamenetskii, M.D., Schyolkina, A.K., 1974. The B to A transition of DNA in solution. J. Mol. Biol. 87, 817–833. https://doi.org/10.1016/0022-2836(74)90086-2

Jain, A., Wang, G., Vasquez, K.M., 2008. DNA Triple Helices: biological consequences and therapeutic potential. Biochimie 90, 1117–1130. https://doi.org/10.1016/j.biochi.2008.02.011

Jakobsen, M.R., Bak, R.O., Andersen, A., Berg, R.K., Jensen, S.B., Jin, T., Laustsen, A., Hansen, K., Østergaard, L., Fitzgerald, K.A., Xiao, T.S., Mikkelsen, J.G., Mogensen, T.H., Paludan, S.R., 2013. IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. Proc. Natl. Acad. Sci. U. S. A. 110, E4571–E4580. https://doi.org/10.1073/pnas.1311669110

Jansson, L.I., Hentschel, J., Parks, J.W., Chang, T.R., Lu, C., Baral, R., Bagshaw, C.R., Stone, M.D., 2019. Telomere DNA G-quadruplex folding within actively extending human telomerase. Proc. Natl. Acad. Sci. 116, 9350–9359. https://doi.org/10.1073/pnas.1814777116

Ji, D., Juhas, M., Tsang, C.M., Kwok, C.K., Li, Y., Zhang, Y., 2021. Discovery of G-quadruplex-forming sequences in SARS-CoV-2. Brief. Bioinform. 22, 1150–1160. https://doi.org/10.1093/bib/bbaa114

Jiang, Z., Wei, F., Zhang, Y., Wang, T., Gao, W., Yu, S., Sun, H., Pu, J., Sun, Y., Wang, M., Tong, Q., Gao, C., Chang, K.-C., Liu, J., 2021. IFI16 directly senses viral RNA and enhances RIG-I transcription and activation to restrict influenza virus infection. Nat. Microbiol. 6, 932–945. https://doi.org/10.1038/s41564-021-00907-x

Johnson, K.E., Bottero, V., Flaherty, S., Dutta, S., Singh, V.V., Chandran, B., 2014. IFI16 Restricts HSV-1 Replication by Accumulating on the HSV-1 Genome, Repressing HSV-1 Gene Expression, and Directly or Indirectly Modulating Histone Modifications. PLOS Pathog. 10, e1004503. https://doi.org/10.1371/journal.ppat.1004503

Johnstone, R.W., Kerry, J.A., Trapani, J.A., 1998. The human interferon-inducible protein, IFI 16, is a repressor of transcription. J. Biol. Chem. 273, 17172–17177. https://doi.org/10.1074/jbc.273.27.17172

Kartasheva, N.N., Lenz-Bauer, C., Hartmann, O., Schäfer, H., Eilers, M., Dobbelstein, M., 2003. DeltaNp73 can modulate the expression of various genes in a p53-independent fashion. Oncogene 22, 8246–8254. https://doi.org/10.1038/sj.onc.1207138

Kaushal, S., Wollmuth, C.E., Das, K., Hile, S.E., Regan, S.B., Barnes, R.P., Haouzi, A., Lee, S.M., House, N.C.M., Guyumdzhyan, M., Eckert, K.A., Freudenreich, C.H., 2019. Sequence and Nuclease Requirements for Breakage and Healing of a Structure-Forming (AT)n Sequence within Fragile Site FRA16D. Cell Rep. 27, 1151-1164.e5. https://doi.org/10.1016/j.celrep.2019.03.103

Kazantseva, M., Mehta, S., Eiholzer, R.A., Hung, N., Wiles, A., Slatter, T.L., Braithwaite, A.W., 2018. A mouse model of the Δ133p53 isoform: roles in cancer progression and inflammation. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 29, 831–842. https://doi.org/10.1007/s00335-018-9758-3

Kharel, P., Fay, M., Manasova, E.V., Anderson, P.J., Kurkin, A.V., Guo, J.U., Ivanov, P., 2023. Stress promotes RNA G-quadruplex folding in human cells. Nat. Commun. 14, 205. https://doi.org/10.1038/s41467-023-35811-x

Kikin, O., D'Antonio, L., Bagga, P.S., 2006. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 34, W676–W682. https://doi.org/10.1093/nar/gkl253

Kikin, O., Zappala, Z., D'Antonio, L., Bagga, P.S., 2008. GRSDB2 and GRS_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. Nucleic Acids Res. 36, D141-148. https://doi.org/10.1093/nar/gkm982

Kim, N., 2019. The Interplay between G-quadruplex and Transcription. Curr. Med. Chem. 26, 2898–2917. https://doi.org/10.2174/0929867325666171229132619

Kimkong, I., Avihingsanon, Y., Hirankarn, N., 2009. Expression profile of HIN200 in leukocytes and renal biopsy of SLE patients by real-time RT-PCR. Lupus 18, 1066–1072. https://doi.org/10.1177/0961203309106699

King, K.E., Ponnamperuma, R.M., Yamashita, T., Tokino, T., Lee, L.A., Young, M.F., Weinberg, W.C., 2003. deltaNp63alpha functions as both a positive and a negative transcriptional regulator and blocks in vitro differentiation of murine keratinocytes. Oncogene 22, 3635–3644. https://doi.org/10.1038/sj.onc.1206536

Krafčíková, P., Demkovičová, E., Víglaský, V., 2017. Ebola virus derived G-quadruplexes: Thiazole orange interaction. Biochim. Biophys. Acta BBA - Gen. Subj., G-quadruplex 1861, 1321–1328. https://doi.org/10.1016/j.bbagen.2016.12.009

Kudlicki, A.S., 2016. G-Quadruplexes Involving Both Strands of Genomic DNA Are Highly Abundant and Colocalize with Functional Sites in the Human Genome. PloS One 11, e0146174. https://doi.org/10.1371/journal.pone.0146174

Kurotaki, N., Stankiewicz, P., Wakui, K., Niikawa, N., Lupski, J.R., 2005. Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. Hum. Mol. Genet. 14, 535–542. https://doi.org/10.1093/hmg/ddi050

Lang, G.A., Iwakuma, T., Suh, Y.-A., Liu, G., Rao, V.A., Parant, J.M., Valentin-Vega, Y.A., Terzian, T., Caldwell, L.C., Strong, L.C., El-Naggar, A.K., Lozano, G., 2004. Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. Cell 119, 861–872. https://doi.org/10.1016/j.cell.2004.11.006

Lannes, L., Halder, S., Krishnan, Y., Schwalbe, H., 2015. Tuning the pH Response of i-Motif DNA Oligonucleotides. ChemBioChem 16, 1647–1656. https://doi.org/10.1002/cbic.201500182

Lawrence, M.G., Porter, L.H., Choo, N., Pook, D., Grummet, J.P., Pezaro, C.J., Sandhu, S., Ramm, S., Luu, J., Bakshi, A., Goode, D.L., Sanij, E., Pearson, R.B., Hannan, R.D., Simpson, K.J., Taylor, R.A., Risbridger, G.P., Furic, L., 2021. CX-5461 Sensitizes DNA Damage Repair-proficient Castrate-resistant Prostate Cancer to PARP Inhibition. Mol. Cancer Ther. 20, 2140–2150. https://doi.org/10.1158/1535-7163.MCT-20-0932

Leroy, B., Anderson, M., Soussi, T., 2014. TP53 mutations in human cancer: database reassessment and prospects for the next decade. Hum. Mutat. 35, 672–688. https://doi.org/10.1002/humu.22552

Li, T., Diner, B.A., Chen, J., Cristea, I.M., 2012. Acetylation modulates cellular distribution and DNA sensing ability of interferon-inducible protein IFI16. Proc. Natl. Acad. Sci. U. S. A. 109, 10558–10563. https://doi.org/10.1073/pnas.1203447109

Lista, M.J., Martins, R.P., Billant, O., Contesse, M.-A., Findakly, S., Pochard, P., Daskalogianni, C., Beauvineau, C., Guetta, C., Jamin, C., Teulade-Fichou, M.-P.,

Fåhraeus, R., Voisset, C., Blondel, M., 2017. Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. Nat. Commun. 8, 16043. https://doi.org/10.1038/ncomms16043

Liu, G., Du, W., Sang, X., Tong, Q., Wang, Y., Chen, G., Yuan, Y., Jiang, L., Cheng, W., Liu, D., Tian, Y., Fu, X., 2022. RNA G-quadruplex in TMPRSS2 reduces SARS-CoV-2 infection. Nat. Commun. 13, 1444. https://doi.org/10.1038/s41467-022-29135-5

Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., de Lima, L.G., Dvorkina, T., Porubsky, D., Harvey, W.T., Mikheenko, A., Bzikadze, A.V., Kremitzki, M., Graves-Lindsay, T.A., Jain, C., Hoekzema, K., Murali, S.C., Munson, K.M., Baker, C., Sorensen, M., Lewis, A.M., Surti, U., Gerton, J.L., Larionov, V., Ventura, M., Miga, K.H., Phillippy, A.M., Eichler, E.E., 2021. The structure, function and evolution of a complete human chromosome 8. Nature 593, 101–107. https://doi.org/10.1038/s41586-021-03420-7

Lu, S., Wang, G., Bacolla, A., Zhao, J., Spitser, S., Vasquez, K.M., 2015. Short Inverted Repeats Are Hot Spots For Genetic Instability: Relevance to Cancer Genomes. Cell Rep. 10, 1674–1680. https://doi.org/10.1016/j.celrep.2015.02.039

Lukman, S., Lane, D.P., Verma, C.S., 2013. Mapping the Structural and Dynamical Features of Multiple p53 DNA Binding Domains: Insights into Loop 1 Intrinsic Dynamics. PLOS ONE 8, e80221. https://doi.org/10.1371/journal.pone.0080221

Madireddy, A., Purushothaman, P., Loosbroock, C.P., Robertson, E.S., Schildkraut, C.L., Verma, S.C., 2016. G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. Nucleic Acids Res. 44, 3675–3694. https://doi.org/10.1093/nar/gkw038

Majee, P., Kumar Mishra, S., Pandya, N., Shankar, U., Pasadi, S., Muniyappa, K., Nayak, D., Kumar, A., 2020. Identification and characterization of two conserved G-quadruplex forming motifs in the Nipah virus genome and their interaction with G-quadruplex specific ligands. Sci. Rep. 10, 1477. https://doi.org/10.1038/s41598-020-58406-8

Marsico, G., Chambers, V.S., Sahakyan, A.B., McCauley, P., Boutell, J.M., Antonio, M.D., Balasubramanian, S., 2019. Whole genome experimental maps of DNA G-quadruplexes in multiple species. Nucleic Acids Res. 47, 3862–3874. https://doi.org/10.1093/nar/gkz179

Mathad, R.I., Hatzakis, E., Dai, J., Yang, D., 2011. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. Nucleic Acids Res. 39, 9023–9033. https://doi.org/10.1093/nar/gkr612

McMahan, Z.H., Shah, A.A., Vaidya, D., Wigley, F.M., Rosen, A., Casciola-Rosen, L., 2016. Anti–Interferon-Inducible Protein 16 Antibodies Associate With Digital Gangrene in Patients With Scleroderma. Arthritis Rheumatol. 68, 1262–1271. https://doi.org/10.1002/art.39558

Mergny, J.L., Sun, J.S., Rougee, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J., Helene, C., 1991. Sequence specificity in triple helix formation: experimental and theoretical studies of the effect of mismatches on triplex stability. Biochemistry 30, 9791–9798. https://doi.org/10.1021/bi00104a031

Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., Schneider, V.A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., Markovic, C., Maduro, V., Dutra, A., Bouffard, G.G., Chang, A.M., Hansen, N.F., Wilfert, A.B., Thibaud-Nissen, F., Schmitt, A.D., Belton, J.-M., Selvaraj, S., Dennis, M.Y., Soto, D.C., Sahasrabudhe, R., Kaya, G., Quick, J., Loman, N.J., Holmes, N., Loose, M., Surti, U., Risques, R. ana, Graves Lindsay, T.A., Fulton, R., Hall, I., Paten, B., Howe, K., Timp, W., Young, A., Mullikin, J.C., Pevzner, P.A., Gerton, J.L., Sullivan, B.A., Eichler, E.E., Phillippy, A.M., 2020. Telomere-to-telomere assembly of a complete human X chromosome. Nature 585, 79–84. https://doi.org/10.1038/s41586-020-2547-7

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., Ogata, H., 2016. Linking Virus Genomes with Host Taxonomy. Viruses 8, 66. https://doi.org/10.3390/v8030066

Mishra, S., Raj, A.S., Kumar, A., Rajeevan, A., Kumari, P., Kumar, H., 2022. Innate immune sensing of influenza A viral RNA through IFI16 promotes pyroptotic cell death. iScience 25, 103714. https://doi.org/10.1016/j.isci.2021.103714

Mishra, S.K., Tawani, A., Mishra, A., Kumar, A., 2016. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. Sci. Rep. 6, 38144. https://doi.org/10.1038/srep38144

Miskiewicz, J., Sarzynska, J., Szachniuk, M., 2021. How bioinformatics resources work with G4 RNAs. Brief. Bioinform. 22, bbaa201. https://doi.org/10.1093/bib/bbaa201

Mitsui, Y., Langridge, R., Shortle, B.E., Cantor, C.R., Grant, R.C., Kodama, M., Wells, R.D., 1970. Physical and enzymatic studies on poly d(I-C)-poly d(I-C), an unusual double-helical DNA. Nature 228, 1166–1169. https://doi.org/10.1038/2281166a0

Moch, C., Fromant, M., Blanquet, S., Plateau, P., 2017. DNA binding specificities of Escherichia coli Cas1-Cas2 integrase drive its recruitment at the CRISPR locus. Nucleic Acids Res. 45, 2714–2723. https://doi.org/10.1093/nar/gkw1309

Monti, P., Bosco, B., Gomes, S., Saraiva, L., Fronza, G., Inga, A., 2019. Yeast As a Chassis for Developing Functional Assays to Study Human P53. J. Vis. Exp. JoVE. https://doi.org/10.3791/59071

Monti, P., Brazda, V., Bohálová, N., Porubiaková, O., Menichini, P., Speciale, A., Bocciardi, R., Inga, A., Fronza, G., 2021. Evaluating the Influence of a G-Quadruplex Prone Sequence on the Transactivation Potential by Wild-Type and/or Mutant P53 Family Proteins through a Yeast-Based Functional Assay. Genes 12, 277. https://doi.org/10.3390/genes12020277

Morrone, S.R., Wang, T., Constantoulakis, L.M., Hooy, R.M., Delannoy, M.J., Sohn, J., 2014. Cooperative assembly of IFI16 filaments on dsDNA provides insights

into host defense strategy. Proc. Natl. Acad. Sci. 111, E62–E71. https://doi.org/10.1073/pnas.1313577111

Murat, P., Zhong, J., Lekieffre, L., Cowieson, N.P., Clancy, J.L., Preiss, T., Balasubramanian, S., Khanna, R., Tellam, J., 2014. G-quadruplexes regulate Epstein-Barr virus–encoded nuclear antigen 1 mRNA translation. Nat. Chem. Biol. 10, 358–364. https://doi.org/10.1038/nchembio.1479

Nag, S., Qin, J., Srivenugopal, K.S., Wang, M., Zhang, R., 2013. The MDM2-p53 pathway revisited. J. Biomed. Res. 27, 254. https://doi.org/10.7555/JBR.27.20130030

Nakanishi, C., Seimiya, H., 2020. G-quadruplex in cancer biology and drug discovery. Biochem. Biophys. Res. Commun., G Quadruplexes 531, 45–50. https://doi.org/10.1016/j.bbrc.2020.03.178

Nawy, T., 2015. Where the G-quadruplexes are. Nat. Methods 12, 807–807. https://doi.org/10.1038/nmeth.3572

Ni, X., Ru, H., Ma, F., Zhao, L., Shaw, N., Feng, Y., Ding, W., Gong, W., Wang, Q., Ouyang, S., Cheng, G., Liu, Z.-J., 2016. New insights into the structural basis of DNA recognition by HINa and HINb domains of IFI16. J. Mol. Cell Biol. 8, 51–61. https://doi.org/10.1093/jmcb/mjv053

Nordin, A., Akimoto, C., Wuolikainen, A., Alstermark, H., Jonsson, P., Birve, A., Marklund, S.L., Graffmo, K.S., Forsberg, K., Brännström, T., Andersen, P.M., 2015. Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. Hum. Mol. Genet. 24, 3133–3142. https://doi.org/10.1093/hmg/ddv064

Nuñez, J.K., Lee, A.S.Y., Engelman, A., Doudna, J.A., 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature 519, 193–198. https://doi.org/10.1038/nature14237

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G.V., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., de Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V.V., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sović, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'Neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H., Phillippy, A.M., 2022. The complete sequence of a human genome. Science 376, 44–53. https://doi.org/10.1126/science.abj6987

Olive, K.P., Tuveson, D.A., Ruhe, Z.C., Yin, B., Willis, N.A., Bronson, R.T., Crowley, D., Jacks, T., 2004. Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. Cell 119, 847–860. https://doi.org/10.1016/j.cell.2004.11.004

Olson, C.L., Barbour, A.T., Wieser, T.A., Wuttke, D.S., 2023. RPA engages telomeric G-quadruplexes more effectively than CST. Nucleic Acids Res. 51, 5073–5086. https://doi.org/10.1093/nar/gkad315

Oren, M., Rotter, V., 2010. Mutant p53 gain-of-function in cancer. Cold Spring Harb. Perspect. Biol. 2, a001107. https://doi.org/10.1101/cshperspect.a001107

Perrone, R., Lavezzo, E., Palù, G., Richter, S.N., 2017. Conserved presence of G-quadruplex forming sequences in the Long Terminal Repeat Promoter of Lentiviruses. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-02291-1

Perrone, R., Nadai, M., Frasson, I., Poe, J.A., Butovskaya, E., Smithgall, T.E., Palumbo, M., Palù, G., Richter, S.N., 2013a. A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter. J. Med. Chem. 56, 6521–6530. https://doi.org/10.1021/jm400914r

Perrone, R., Nadai, M., Poe, J.A., Frasson, I., Palumbo, M., Palù, G., Smithgall, T.E., Richter, S.N., 2013b. Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 nef Coding Region: Implications for Antiviral Activity. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0073121

Peter MacCallum Cancer Centre, Australia, 2022. Phase 1 Trial of Pidnarulex and Talazoparib in Patients With Metastatic Castration Resistant Prostate Cancer (Clinical trial registration No. NCT05425862). clinicaltrials.gov.

Piccaluga, P.P., Agostinelli, C., Fuligni, F., Righi, S., Tripodo, C., Re, M.C., Clò, A., Miserocchi, A., Morini, S., Gariglio, M., Ferri, G.G., Rinaldi-Ceroni, A., Piccin, O., De Andrea, M., Pileri, S.A., Landolfo, S., Gibellini, D., 2015. IFI16 Expression Is Related to Selected Transcription Factors during B-Cell Differentiation. J. Immunol. Res. 2015. https://doi.org/10.1155/2015/747645

Piekna-Przybylska, D., Sullivan, M.A., Sharma, G., Bambara, R.A., 2014. U3 Region in the HIV-1 Genome Adopts a G-Quadruplex Structure in Its RNA and DNA Sequence. Biochemistry 53, 2581–2593. https://doi.org/10.1021/bi4016692

Pisano, G., Roy, A., Ahmed Ansari, M., Kumar, B., Chikoti, L., Chandran, B., 2017. Interferon-γ-inducible protein 16 (IFI16) is required for the maintenance of Epstein-Barr virus latency. Virol. J. 14. https://doi.org/10.1186/s12985-017-0891-5

Platella, C., Riccardi, C., Montesarchio, D., Roviello, G.N., Musumeci, D., 2017. G-quadruplex-based aptamers against protein targets in therapy and diagnostics. Biochim. Biophys. Acta Gen. Subj. 1861, 1429–1447. https://doi.org/10.1016/j.bbagen.2016.11.027

Porubiaková, O., Bohálová, N., Inga, A., Vadovičová, N., Coufal, J., Fojta, M., Brázda, V., 2019. The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms. Int. J. Mol. Sci. 21. https://doi.org/10.3390/ijms21010127

Pourcel, C., Touchon, M., Villeriot, N., Vernadet, J.-P., Couvin, D., Toffano-Nioche, C., Vergnaud, G., 2020. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. Nucleic Acids Res. 48, D535–D544. https://doi.org/10.1093/nar/gkz915

Puig Lombardi, E., Londoño-Vallejo, A., 2020. A guide to computational methods for G-quadruplex prediction. Nucleic Acids Res. 48, 1–15. https://doi.org/10.1093/nar/gkz1097

Puig Lombardi, E., Londoño-Vallejo, A., Nicolas, A., 2019. Relationship Between G-Quadruplex Sequence Composition in Viruses and Their Hosts. Molecules 24, 1942. https://doi.org/10.3390/molecules24101942

Qin, G., Zhao, C., Liu, Y., Zhang, C., Yang, G., Yang, J., Wang, Z., Wang, C., Tu, C., Guo, Z., Ren, J., Qu, X., 2022. RNA G-quadruplex formed in SARS-CoV-2 used for COVID-19 treatment in animal models. Cell Discov. 8, 1–13. https://doi.org/10.1038/s41421-022-00450-x

Quante, T., Otto, B., Brázdová, M., Kejnovská, I., Deppert, W., Tolstonog, G.V., 2012. Mutant p53 is a transcriptional co-factor that binds to G-rich regulatory regions of active genes and generates transcriptional plasticity. Cell Cycle 11, 3290–3303. https://doi.org/10.4161/cc.21646

Ravichandran, S., Kim, Y.-E., Bansal, V., Ghosh, A., Hur, J., Subramani, V.K., Pradhan, S., Lee, M.K., Kim, K.K., Ahn, J.-H., 2018. Genome-wide analysis of regulatory G-quadruplexes affecting gene expression in human cytomegalovirus. PLoS Pathog. 14, e1007334. https://doi.org/10.1371/journal.ppat.1007334

Recagni, M., Tassinari, M., Doria, F., Cimino-Reale, G., Zaffaroni, N., Freccero, M., Folini, M., Richter, S.N., 2019. The Oncogenic Signaling Pathways in BRAF-Mutant Melanoma Cells Are Modulated by Naphthalene Diimide-Like G-Quadruplex Ligands. Cells 8, 1274. https://doi.org/10.3390/cells8101274

Rivlin, N., Brosh, R., Oren, M., Rotter, V., 2011. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. Genes Cancer 2, 466–474. https://doi.org/10.1177/1947601911408889

Robles, A.I., Harris, C.C., 2010. Clinical Outcomes and Correlates of TP53 Mutations and Cancer. Cold Spring Harb. Perspect. Biol. 2, a001016. https://doi.org/10.1101/cshperspect.a001016

Rocher, V., Genais, M., Nassereddine, E., Mourad, R., 2021. DeepG4: A deep learning approach to predict cell-type specific active G-quadruplex regions. PLOS Comput. Biol. 17, e1009308. https://doi.org/10.1371/journal.pcbi.1009308

Roxo, C., Kotkowiak, W., Pasternak, A., 2019. G-Quadruplex-Forming Aptamers—Characteristics, Applications, and Perspectives. Molecules 24, 3781. https://doi.org/10.3390/molecules24203781

Roy, A., Dutta, D., Iqbal, J., Pisano, G., Gjyshi, O., Ansari, M.A., Kumar, B., Chandran, B., 2016. Nuclear Innate Immune DNA Sensor IFI16 Is Degraded during Lytic Reactivation of Kaposi's Sarcoma-Associated Herpesvirus (KSHV): Role of IFI16 in Maintenance of KSHV Latency. J. Virol. 90, 8822–8841. https://doi.org/10.1128/JVI.01003-16

Ruggiero, E., Richter, S.N., 2023. Targeting G-quadruplexes to achieve antiviral activity. Bioorg. Med. Chem. Lett. 79, 129085. https://doi.org/10.1016/j.bmcl.2022.129085

Ruggiero, E., Richter, S.N., 2020. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy, in: Annual Reports in Medicinal Chemistry. Elsevier, pp. 101–131. https://doi.org/10.1016/bs.armc.2020.04.001

Sahakyan, A.B., Chambers, V.S., Marsico, G., Santner, T., Di Antonio, M., Balasubramanian, S., 2017. Machine learning model for sequence-driven DNA G-quadruplex formation. Sci. Rep. 7, 14535. https://doi.org/10.1038/s41598-017-14017-4

SantaLucia, J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. 95, 1460–1465. https://doi.org/10.1073/pnas.95.4.1460

Savvateeva-Popova, E.V., Zhuravlev, A.V., Brázda, V., Zakharov, G.A., Kaminskaya, A.N., Medvedeva, A.V., Nikitina, E.A., Tokmatcheva, E.V., Dolgaya, J.F., Kulikova, D.A., Zatsepina, O.G., Funikov, S.Y., Ryazansky, S.S., Evgen'ev, M.B., 2017. Drosophila Model for the Analysis of Genesis of LIM-kinase 1-Dependent Williams-Beuren Syndrome Cognitive Phenotypes: INDELs, Transposable Elements of the Tc1/Mariner Superfamily and MicroRNAs. Front. Genet. 8, 123. https://doi.org/10.3389/fgene.2017.00123

Senhwa Biosciences, Inc., 2023. Phase Ib Expansion Study of CX-5461 in Patients With Solid Tumours and BRCA2 and/or PALB2 Mutation (Clinical trial registration No. NCT04890613). clinicaltrials.gov.

Short, N.J., Montalban-Bravo, G., Hwang, H., Ning, J., Franquiz, M.J., Kanagal-Shamanna, R., Patel, K.P., DiNardo, C.D., Ravandi, F., Garcia-Manero, G., Takahashi, K., Konopleva, M., Daver, N., Issa, G.C., Andreeff, M., Kantarjian, H., Kadia, T.M., 2020. Prognostic and therapeutic impacts of mutant TP53 variant allelic frequency in newly diagnosed acute myeloid leukemia. Blood Adv. 4, 5681–5689. https://doi.org/10.1182/bloodadvances.2020003120

Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., Hurley, L.H., 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc. Natl. Acad. Sci. 99, 11593–11598. https://doi.org/10.1073/pnas.182256799

Song, L.L., Ponomareva, L., Shen, H., Duan, X., Alimirah, F., Choubey, D., 2010. Interferon-Inducible IFI16, a Negative Regulator of Cell Growth, Down-Regulates Expression of Human Telomerase Reverse Transcriptase (hTERT) Gene. PLoS ONE 5, e8569. https://doi.org/10.1371/journal.pone.0008569

Stansel, R.M., Subramanian, D., Griffith, J.D., 2002. p53 binds telomeric single strand overhangs and t-loop junctions in vitro. J. Biol. Chem. 277, 11625–11628. https://doi.org/10.1074/jbc.C100764200

Storek, K.M., Gertsvolf, N.A., Ohlson, M.B., Monack, D.M., 2015. cGAS and Ifi204 cooperate to produce type I IFNs in response to Francisella infection. J. Immunol. Baltim. Md 1950 194, 3236–3245. https://doi.org/10.4049/jimmunol.1402764

Sun, D., Guo, K., Rusche, J.J., Hurley, L.H., 2005. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. Nucleic Acids Res. 33, 6070–6080. https://doi.org/10.1093/nar/gki917

Tassinari, M., Zuffo, M., Nadai, M., Pirota, V., Sevilla Montalvo, A.C., Doria, F., Freccero, M., Richter, S.N., 2020a. Selective targeting of mutually exclusive DNA G-quadruplexes: HIV-1 LTR as paradigmatic model. Nucleic Acids Res. 48, 4627–4642. https://doi.org/10.1093/nar/gkaa186

Tassinari, M., Zuffo, M., Nadai, M., Pirota, V., Sevilla Montalvo, A.C., Doria, F., Freccero, M., Richter, S.N., 2020b. Selective targeting of mutually exclusive DNA G-quadruplexes: HIV-1 LTR as paradigmatic model. Nucleic Acids Res 48, 4627–4642. https://doi.org/10.1093/nar/gkaa186

Teng, Y., Zhu, M., Qiu, Z., 2023. G-Quadruplexes in Repeat Expansion Disorders. Int. J. Mol. Sci. 24, 2375. https://doi.org/10.3390/ijms24032375

Tlučková, K., Marušič, M., Tóthová, P., Bauer, L., Šket, P., Plavec, J., Viglasky, V., 2013. Human papillomavirus G-quadruplexes. Biochemistry 52, 7207–7216. https://doi.org/10.1021/bi400897g

Todd, A.K., Johnston, M., Neidle, S., 2005. Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. 33, 2901–2907. https://doi.org/10.1093/nar/gki553

Tosoni, E., Frasson, I., Scalabrin, M., Perrone, R., Butovskaya, E., Nadai, M., Palù, G., Fabris, D., Richter, S.N., 2015. Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. Nucleic Acids Res. 43, 8884. https://doi.org/10.1093/nar/gkv897

Travers, A., Muskhelishvili, G., 2015. DNA structure and function. FEBS J. 282, 2279–2295. https://doi.org/10.1111/febs.13307

Tuerk, C., Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505–510.

Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T., Latz, E., Xiao, T.S., Fitzgerald, K.A., Paludan, S.R., Bowie, A.G., 2010. IFI16 is an innate immune sensor for intracellular DNA. Nat. Immunol. 11, 997–1004. https://doi.org/10.1038/ni.1932

Uppal, T., Banerjee, S., Sun, Z., Verma, S.C., Robertson, E.S., 2014. KSHV LANA—The Master Regulator of KSHV Latency. Viruses 6, 4961–4998. https://doi.org/10.3390/v6124961

Valková, N., Kratochvilová, L., Martinková, L., Brázda, V., 2023. Dual mode of IFI16 binding to supercoiled and linear DNA: A closer insight. Biochem. Biophys. Res. Commun. 667, 89–94. https://doi.org/10.1016/j.bbrc.2023.05.049

Vanhove, W., Peeters, P.M., Staelens, D., Schraenen, A., Van der Goten, J., Cleynen, I., De Schepper, S., Van Lommel, L., Reynaert, N.L., Schuit, F., Van Assche, G., Ferrante, M., De Hertogh, G., Wouters, E.F.M., Rutgeerts, P., Vermeire, S., Nys, K., Arijs, I., 2015. Strong Upregulation of AIM2 and IFI16 Inflammasomes in

the Mucosa of Patients with Active Inflammatory Bowel Disease: Inflamm. Bowel Dis. 21, 2673–2682. https://doi.org/10.1097/MIB.0000000000000535
Varizhuk, A., Ilyinsky, N., Smirnov, I., Pozmogova, G., 2016. G4 Aptamers: Trends in Structural Design. Mini Rev. Med. Chem. 16, 1321–1329. https://doi.org/10.2174/1389557516666160321114715
Varizhuk, A., Ischenko, D., Tsvetkov, V., Novikov, R., Kulemin, N., Kaluzhny, D., Vlasenok, M., Naumov, V., Smirnov, I., Pozmogova, G., 2017. The expanding repertoire of G4 DNA structures. Biochimie 135, 54–62. https://doi.org/10.1016/j.biochi.2017.01.003
Vaughan, C.A., Singh, S., Grossman, S.R., Windle, B., Deb, S.P., Deb, S., 2017. Gain-of-function p53 activates multiple signaling pathways to induce oncogenicity in lung cancer cells. Mol. Oncol. 11, 696–711. https://doi.org/10.1002/1878-0261.12068
Víglaský, V., 2022. Hidden Information Revealed Using the Orthogonal System of Nucleic Acids. Int. J. Mol. Sci. 23, 1804. https://doi.org/10.3390/ijms23031804
Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G., Rich, A., 1979. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature 282, 680–686. https://doi.org/10.1038/282680a0
Wang, S.-R., Zhang, Q.-Y., Wang, J.-Q., Ge, X.-Y., Song, Y.-Y., Wang, Y.-F., Li, X.-D., Fu, B.-S., Xu, G.-H., Shu, B., Gong, P., Zhang, B., Tian, T., Zhou, X., 2016. Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. Cell Chem. Biol. 23, 1113–1122. https://doi.org/10.1016/j.chembiol.2016.07.019
Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737–738. https://doi.org/10.1038/171737a0
Wells, R.D., 2009. Mutation Spectra in Fragile X Syndrome Induced by Deletions of CGG·CCG Repeats. J. Biol. Chem. 284, 7407–7411. https://doi.org/10.1074/jbc.R800024200
Wichit, S., Hamel, R., Yainoy, S., Gumpangseth, N., Panich, S., Phuadraksa, T., Saetear, P., Monteil, A., Morales Vargas, R., Missé, D., 2019. Interferon-inducible protein (IFI) 16 regulates Chikungunya and Zika virus infection in human skin fibroblasts. EXCLI J. 18, 467–476. https://doi.org/10.17179/excli2019-1271
Williamson, J.R., Raghuraman, M.K., Cech, T.R., 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell 59, 871–880. https://doi.org/10.1016/0092-8674(89)90610-7
Wu, X., Bayle, J.H., Olson, D., Levine, A.J., 1993. The p53-mdm-2 autoregulatory feedback loop. Genes Dev. 7, 1126–1132. https://doi.org/10.1101/gad.7.7a.1126
Xu, B., Devi, G., Shao, F., 2015. Regulation of telomeric i-motif stability by 5-methylcytosine and 5-hydroxymethylcytosine modification. Org. Biomol. Chem. 13, 5646–5651. https://doi.org/10.1039/C4OB02646B

Xu, H., Hurley, L.H., 2022. A first-in-class clinical G-quadruplex-targeting drug. The bench-to-bedside translation of the fluoroquinolone QQ58 to CX-5461 (Pidnarulex). Bioorg. Med. Chem. Lett. 77, 129016. https://doi.org/10.1016/j.bmcl.2022.129016

Yan, H., Dalal, K., Hon, B.K., Youkharibache, P., Lau, D., Pio, F., 2008. RPA nucleic acid-binding properties of IFI16-HIN200. Biochim. Biophys. Acta BBA - Proteins Proteomics 1784, 1087–1097. https://doi.org/10.1016/j.bbapap.2008.04.004

Ye, C., Ji, G., Li, L., Liang, C., 2014. detectIR: A Novel Program for Detecting Perfect and Imperfect Inverted Repeats Using Complex Numbers and Vector Calculation. PLoS ONE 9, e113349. https://doi.org/10.1371/journal.pone.0113349

Yu, H., Qi, Y., Yang, B., Yang, X., Ding, Y., 2023. G4Atlas: a comprehensive transcriptome-wide G-quadruplex database. Nucleic Acids Res. 51, D126–D134. https://doi.org/10.1093/nar/gkac896

Zahin, M., Dean, W.L., Ghim, S.-J., Joh, J., Gray, R.D., Khanal, S., Bossart, G.D., Mignucci-Giannoni, A.A., Rouchka, E.C., Jenson, A.B., Trent, J.O., Chaires, J.B., Chariker, J.H., 2018. Identification of G-quadruplex forming sequences in three manatee papillomaviruses. PloS One 13, e0195625. https://doi.org/10.1371/journal.pone.0195625

Zeraati, M., Langley, D.B., Schofield, P., Moye, A.L., Rouet, R., Hughes, W.E., Bryan, T.M., Dinger, M.E., Christ, D., 2018. I-motif DNA structures are formed in the nuclei of human cells. Nat. Chem. 10, 631–637. https://doi.org/10.1038/s41557-018-0046-3

Zhao, H., Gonzalezgugel, E., Cheng, L., Richbourgh, B., Nie, L., Liu, C., 2015. The roles of interferon-inducible p200 family members IFI16 and p204 in innate immune responses, cell differentiation and proliferation. Genes Dis. 2, 46–56. https://doi.org/10.1016/j.gendis.2014.10.003

Přílohy

Příloha A

The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms

International Journal of
Molecular Sciences

MDPI

Article

The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms

Otília Porubiaková [1,2,†], Natália Bohálová [2,3,†], Alberto Inga [4], Natália Vadovičová [2,5], Jan Coufal [2], Miroslav Fojta [2] and Václav Brázda [1,2,*]

1 Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 61200 Brno, Czech Republic; o.porubiakova@gmail.com
2 Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 61265 Brno, Czech Republic; nataliabohalova@gmail.com (N.B.); vadovicovan@mail.muni.cz (N.V.); jac@ibp.cz (J.C.); fojta@ibp.cz (M.F.)
3 Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic
4 Laboratory of Transcriptional Networks, Department CIBIO, University of Trento, via Sommarive 9, 38123 Trento, Italy; alberto.inga@unitn.it
5 Department of Biology, Faculty of Medicine, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic
* Correspondence: vaclav@ibp.cz; Tel.: +420541517231; Fax: +420541211293
† These authors contributed equally to this work.

Received: 30 October 2019; Accepted: 23 December 2019; Published: 24 December 2019

Abstract: p53 is one of the most studied tumor suppressor proteins that plays an important role in basic biological processes including cell cycle, DNA damage response, apoptosis, and senescence. The human *TP53* gene contains alternative promoters that produce N-terminally truncated proteins and can produce several isoforms due to alternative splicing. p53 function is realized by binding to a specific DNA response element (RE), resulting in the transactivation of target genes. Here, we evaluated the influence of quadruplex DNA structure on the transactivation potential of full-length and N-terminal truncated p53α isoforms in a panel of *S. cerevisiae* luciferase reporter strains. Our results show that a G-quadruplex prone sequence is not sufficient for transcription activation by p53α isoforms, but the presence of this feature in proximity to a p53 RE leads to a significant reduction of transcriptional activity and changes the dynamics between co-expressed p53α isoforms.

Keywords: p53 protein; protein-DNA interaction; transactivation potential

1. Introduction

The tumor suppressor protein, p53, is called the "guardian of the genome" due to its crucial role in maintaining genetic stability and inhibiting cancer formation [1,2]. To exert this role, once activated after cell injury, p53 induces a number of cellular processes, resulting in cell repair and survival or in programmed cell death [3–5]. The canonical p53 protein, also named p53α, FLp53α, or TAp53alpha (hereafter referred to as FLp53α), was the first identified p53 form [6]. Human FLp53α is 393-amino acids long and has seven functional domains. The N-terminal domain contains two transactivation (TA) domains, which are required to induce a distinct subset of p53-target genes. Other domains are a proline-rich domain (PRD), a DNA-binding domain (DBD), a hinge domain (HD), and a C-terminal domain composed of an oligomerization domain (OD) and a negative regulation domain (α) [7]. The negative regulation domain is rich in lysine and undergoes many posttranslational modifications that regulate FLp53α activity and stability [8]. The DBD contains several conserved cysteines

and histidines that coordinate Zn^{2+} or Mg^{2+} ions, which are essential for FLp53α conformation and DNA-binding activity [9]. Different N-terminal isoforms of p53α have been identified due to alternative translation initiation, splicing sites, or alternative promoter usage: Δ40p53α, Δ133p53α, and Δ160p53α lack the 39, 132, and 159 N-terminal amino acids, respectively, compared with FLp53α [10,11]. As a consequence, Δ40p53α lacks one of the two TA domains while the other two isoforms lack both TA and the PR domains, plus part of the conserved cysteine box in the DBD [12]. Based on experiments over the past ten years, it has been shown that p53 isoforms are physiologically active proteins. Misregulation of p53 isoform expression can lead to cancer, premature aging, neurodegenerative diseases, or even embryo malformations [13,14].

p53 is part of an extensive transcriptional network that coordinates the response to intracellular and extracellular stresses or damage [5]. The main function of p53 is provided by its interaction with DNA [15–19]. p53 regulates target gene expression mainly by activation of p53-responsive promoters. The DNA response element (RE) for p53 binding comprises two copies of a 5′- RRRC(A/T)(T/A)GYYY-3′ sequence [15,20,21] accommodating the binding of two p53 dimers to form a p53 tetramer that is considered the functional unit for transcriptional modulation [16]. The domain responsible for sequence-specific DNA binding is the core DBD, even though the OD is critical for tetramer formation and modifications to the C-terminal domain influence binding affinity and specificity [22]. p53-DNA interactions with p53 REs are sensitive to DNA topology and this is a key parameter contributing to p53-DNA affinity and specificity [18,23]. It was demonstrated that p53 also binds to various local DNA structures stabilized by DNA topological stress such as cruciforms [24,25], quadruplex [26], triplex [27], bulged [28], and hemicatenate [29] DNAs.

The unicellular yeast *Saccharomyces cerevisiae* has been previously employed to study the transcriptional activity of many human transcription factors including p53 and its isoforms [30–32]. Here, we have engineered yeast reporter strains to study the impact of positioning a G-quadruplex (G4) prone sequence alone or in proximity (upstream or downstream) of a p53 RE on the transactivation induced by FLp53α and the N-terminally truncated isoforms (Δ40p53α, Δ133p53α, and Δ160p53α), expressed both individually and in combination.

In particular, we investigated whether G4 prone sequences are capable of inducing p53-dependent transactivation per se, and/or whether they modify transcription when present in close proximity to an established p53 binding site. We also investigated whether G4 prone sequences impact on the crosstalk between co-expressed p53 isoforms and mapped the presence of G4 forming sequences nearby p53 PUMA RE in genomic context. Our results further emphasize the potential role of structural DNA features as modifiers of p53 protein functions at target promoter sites.

2. Results

2.1. Construction of Isogenic Yeast Strains

To elucidate the influence of a G4 on p53α transcriptional activity, we exploited yeast isogenic reporters. We used the following G-rich DNA sequence GGGGCGGGGGACGGGGGAGGGG, which is very highly prone to form a G4, based on the propensity score given by the G4Hunter tool [33,34] (G4Hunter score 3.182), which is even higher than the sequence from the c-Myc promoter region (G4Hunter score 2.941) where the presence of the G4 structure has been evaluated both in vitro and in vivo [35,36]. We confirmed the propensity of this sequence to form G4 by CD spectroscopy (Figure 1). The measurements showed that the G-rich sequence forms a hybrid type of G4 with dominant parallel G4 represented by the peak at 264 nm and an antiparallel G4 structure resulting in the secondary peak at 295 nm. The slow drop off of the curve after the typical 264 nm peak is in keeping with the evidence that topologically different G4 intermediates may coexist [37,38]. Sequences with an additional PUMA p53RE region showed higher preference for the antiparallel G4 structure with a more prominent peak around 295 nm.

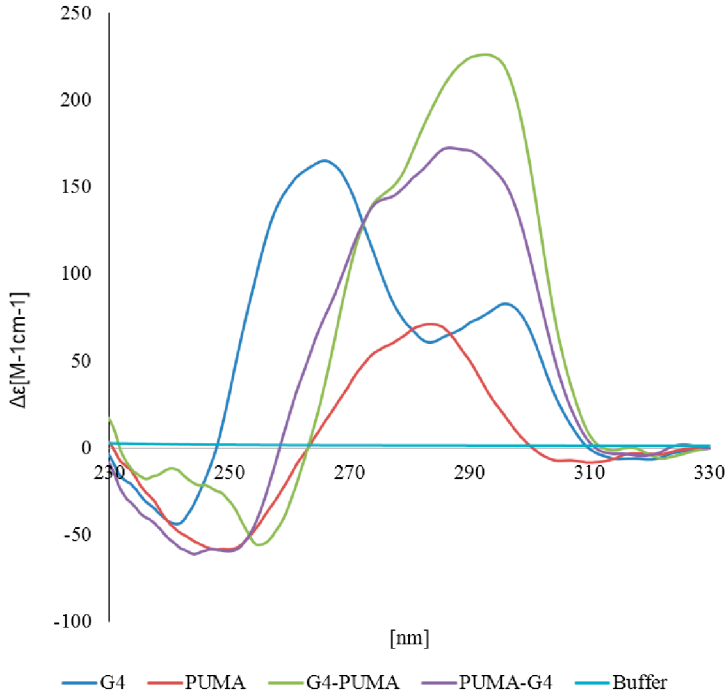


Figure 1. Circular Dichroism }CD] spectroscopy of used DNA sequences. CD spectra of the buffer (light blue), and oligonucleotides from the Table 1 (G4, blue, PUMA-red, G4-PUMA-green, PUMA-G4 violet).

Next, we integrated the p53 RE derived from the human PUMA/BBC3 promoter and the G4 sequence alone or combined upstream of a minimal promoter driving the luciferase reporter gene at the *ade2* locus in yeast. Two versions of the combined element were constructed, differing in the position of the G4 sequence either upstream or downstream of the p53 RE (Figure 2).

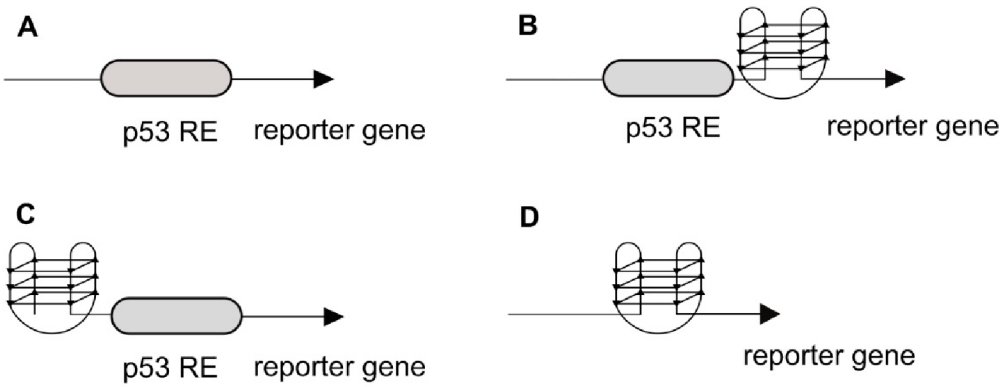


Figure 2. Scheme of the tested sequences in the luciferase reporter promoter region.

2.2. Transactivation Activity of p53α

The reporter yeast strains were used to measure the transactivation potential of four p53α isoforms. First, exploiting the galactose inducible system to control p53 expression, we analyzed the level of transcription of the reporter in the presence of the PUMA p53RE without galactose and with 0.2% or 2% galactose. The results showed that both FLp53α and Δ40p53α transactivate the reporter, although to different extents (Figure 3). Increasing the amount of galactose led to a proportional increase in transactivation for both isoforms. The Δ133α and Δ160α isoforms did not induce transactivation of the PUMA p53 RE.

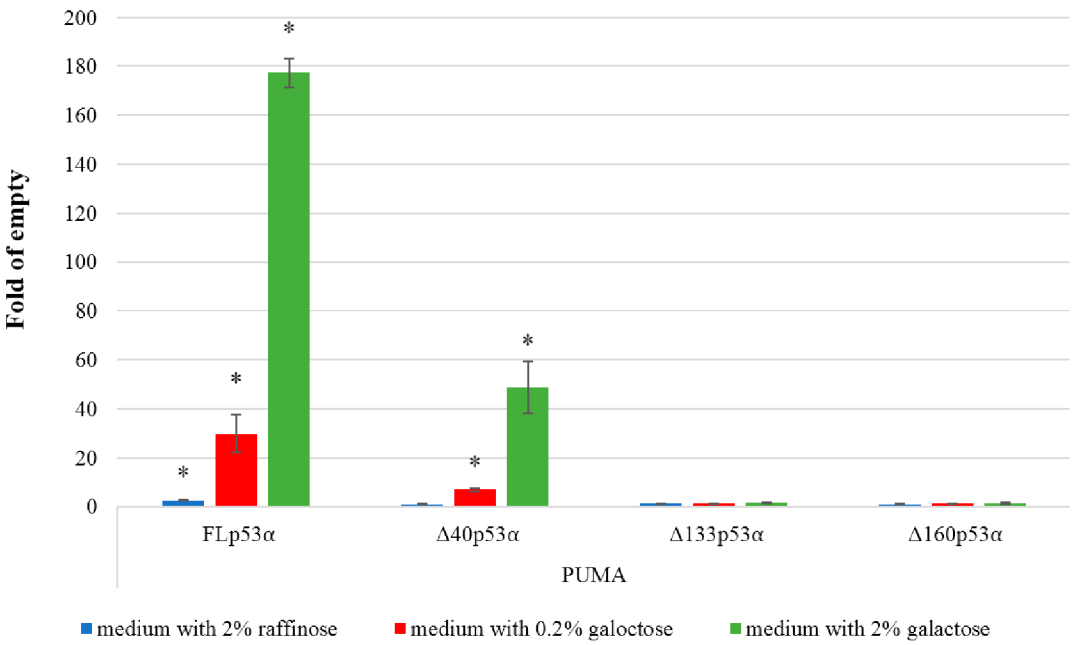


Figure 3. p53-dependent transactivation potential in yeast. All p53α isoforms are expressed under an inducible GAL1 promoter. Histograms show the average fold induction over empty vector in three biological replicates (mean ± S.D.). The results with three levels of p53 induction (no induction, moderate, high) obtained after 24 h in inducing media are presented. Asterisks indicate a significant induction of p53 dependent transactivation ($p < 0.05$).

Similarly, the transactivation potential of constitutively expressed p53 (GPD promoter) was significantly higher for the FLp53α isoform compared to the Δ40p53α isoform, while Δ133 and Δ160 isoforms were not able to transactivate the reporter (Figure 4).

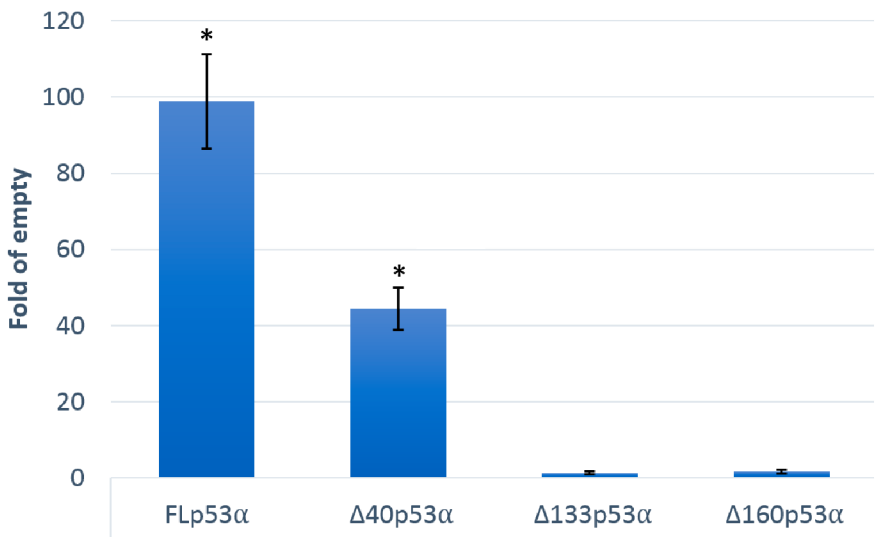


Figure 4. p53-dependent transactivation potential in yeast. All p53α isoforms are expressed under a constitutive glyceraldehyde-3-phosphate dehydrogenase (*GPD*) promoter. The results for the indicated p53α isoforms obtained after 24 h in media without induction are presented. Asterisks indicate a significant induction of p53 dependent transactivation ($p < 0.05$).

To elucidate the role of G4 structure on the transcriptional activity p53α isoforms, we tested three additional yeast isogenic strains containing the G4 alone or combinations of the p53 RE with the G4 sequence upstream or downstream. All strains were co-transformed so that the activity of FLp53α expressed alone or combined with the other p53α isoforms could be assessed in the various reporter strains. FLp53α was expressed under the constitutive *GPD* promoter while Δp53α isoforms were under the *GAL1* promoter and were expressed both at moderate (Figure 5A) and high levels (Figure 5B). Performing western blot of p53 isoforms is challenging due to the lack of commercially available isoform-specific antibodies, but western blot with the DO-1 antibody that detects an N-terminal epitope (residues 11–25) in FLp53α has shown that expression of full-length p53 by the constitutive *GPD* promoter in yeast was not dramatically affected by the co-selection of expression plasmids for p53α isoforms (Figure S1). FLp53α induced transactivation in the strain with just the p53 RE upstream of the luciferase reporter, but had no transactivation activity on G4 alone. The transactivational activity of FLp53α was affected by the G4 sequence placed either upstream or downstream of the p53RE. Interestingly, the presence of the G4 in close proximity to the p53 RE decreased p53-dependent transactivation (Figure 5, red bars), but the position of the G4 sequence influenced this effect. The inhibitory effect was greater with the G4 inserted after the p53 RE (i.e., closer to the TSS) than when the G4 was positioned upstream of the p53 RE. None of the Δp53α isoforms impacted the low transcription activity of the reporter containing the G4 sequence only. In the p53 PUMA RE reporter strain, Δ160p53α decreased transactivation by FLp53α, particularly when expressed at high levels (Figure 5B). Such a decrease was not observed with Δ40p53α (consistent with the residual transactivation potential of this isoform), but it slightly potentiated FLp53α transactivation activity. However, placing the G4 sequence downstream of the p53 RE led to changes in the apparent functional interaction between co-expressed p53α isoforms, and Δ40p53α gained an inhibitory effect over FLp53α, while Δ133p53α and Δ160p53α lost that property. Indeed, when expressed alone, Δ40p53α was impacted by the presence of the G4 sequence in a manner similar to FLp53α (Figure 6).

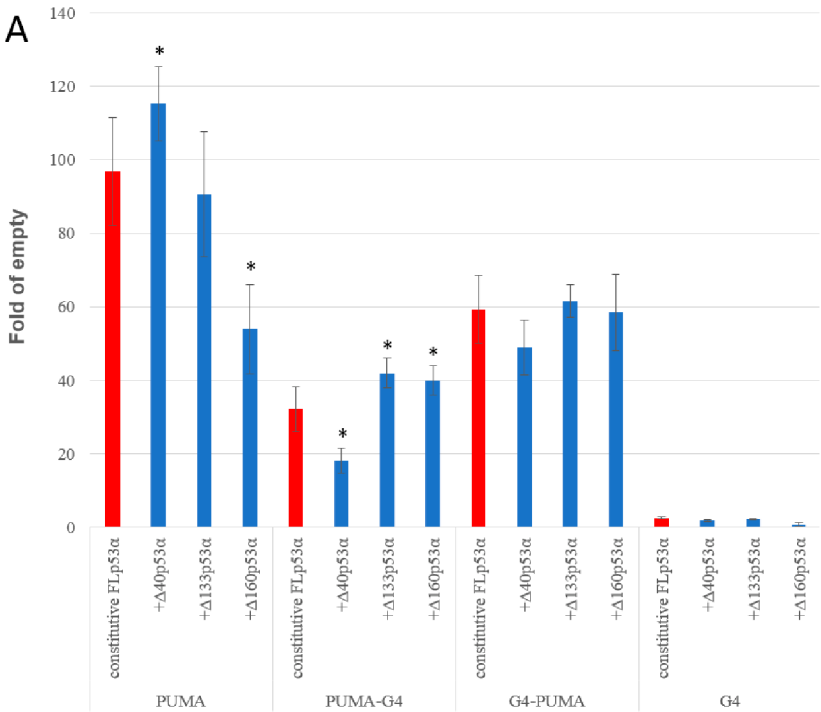


Figure 5. *Cont.*

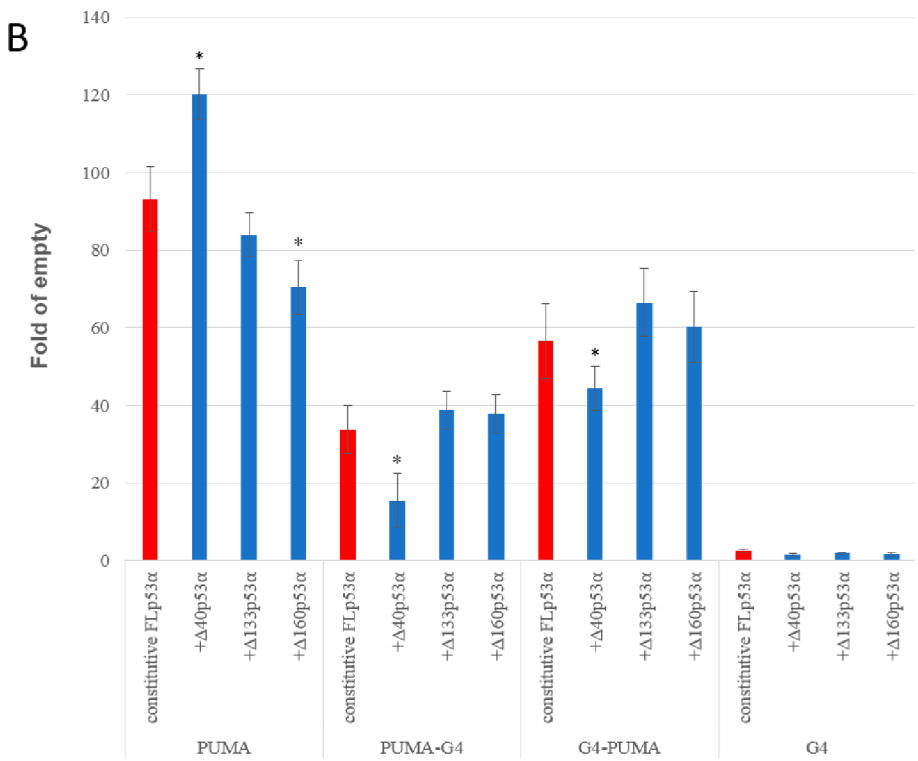


Figure 5. Influence of Δp53α isoforms with inducible expression on transactivation activity of constitutively expressed FLp53α. (**A**) in media with 0.2 % galactose; (**B**) in media with 2% galactose. Four isogenic yeast strains were used, with the p53 target site (PUMA), with the p53 target site after G4 forming sequence (G4-PUMA), with the p53 target site before G-quadruplex forming sequence (PUMA-G4), and the G4 forming sequence upstream of the luciferase gene. Asterisks indicate a significant induction of p53-dependent transactivation ($p < 0.05$).

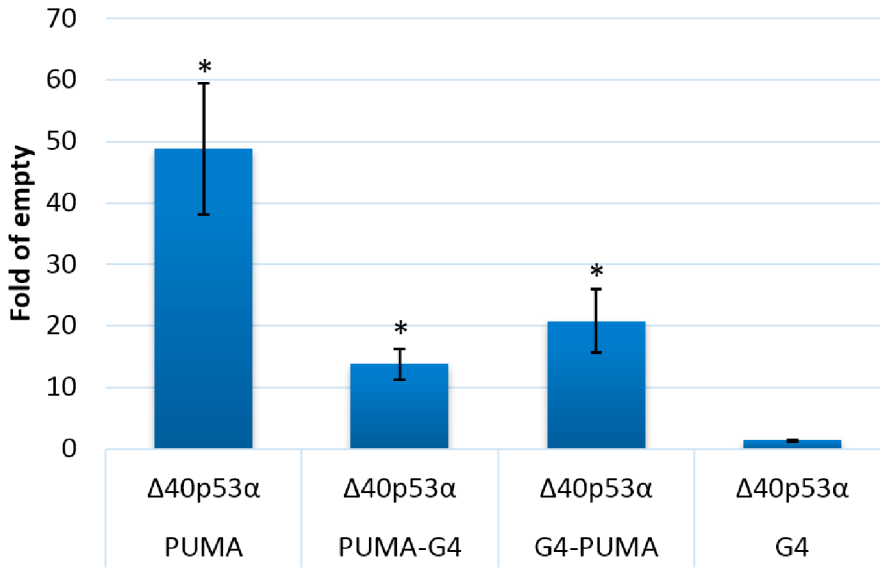


Figure 6. Impact of the G4 prone sequence on Δ40p53α transactivation activity from the PUMA RE. Δ40p53α was expressed under an inducible *GAL1* promoter. Histograms show average fold induction over empty vector in three biological replicates (mean ± S.D.). The results obtained after 24 h in 2% galactose inducing media are presented. Asterisks indicate a significant induction of p53-dependent transactivation ($p < 0.05$).

3. Discussion

p53 is a transcription factor that recognizes a 20-bp long DNA motif. However, chromatin immunoprecipitation has shown that many p53 targets do not contain a classical full-length p53 RE, but can be formed by half-site [21], or do not contain classical target sequences [39]. Non-canonical DNA motifs are transcriptionally active for wild type and mutant p53 proteins [40] and local DNA structures are important determinants for protein-DNA binding [41]. Recently, the interaction of p53 with G4s has been demonstrated [26]. Even if it was demonstrated that G4 structures are often located in gene regulatory sequences in the human genome [42] and there are many studies of p53 target genes [16,39], a combined study of both features is missing. Therefore, we performed additional analyses of 100 bp sequence surrounding the p53-target sequence in the PUMA gene promoter. Interestingly, there are several potential G4-prone sequences in close proximity to the PUMA p53-target sequence (Figure 7). The G4-prone sequence is located tightly before p53 RE (−33 to −1 before p53 RE, max. G4Hunter score in this area 1.84) and several G4-prone sequences are located after the p53 RE including a G4Hunter score of 1.32 immediately after the p53 RE—location 0–25—and another further downstream (starting either 21, 45, and 58 nucleotides after the p53 RE; highest G4Hunter score of 3.2 for the sequence: GGGGGCGGGG CGGGGCGGGG CGGGG, peak at 71 nucleotides after p53 RE).

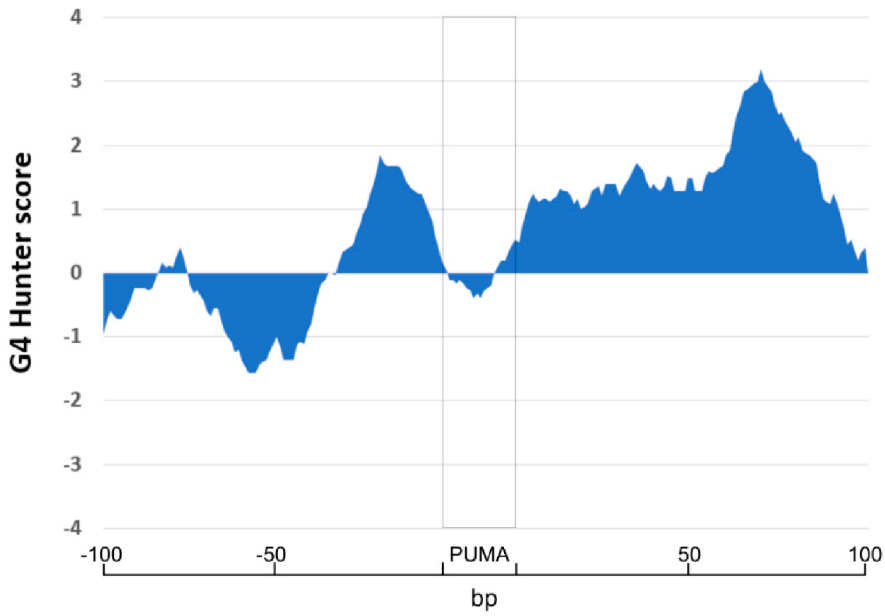


Figure 7. Localization of G4-prone sequences around p53 PUMA RE sequence (100 bp before and after p53 PUMA RE). The analysis of G4-prone sequences has shown that p53 PUMA RE (rectangle) in the human genome is surrounded by G4-prone sequences with peaks at 1.84 before p53 RE and long G4-prone sequence with the peak at 1.32 just after p53 RE and with a maximum peak with G4Hunter score 3.2).

Even though the localization of both p53 RE and G4 sequences have been shown in the genome, the roles of G-quadruplexes in regulating transcription by p53 isoforms have not been evaluated. Therefore, we prepared a model system and analyzed the impact of a sequence endowed with high propensity to adopt a G4 structure positioned either upstream or downstream of a moderately active p53 RE using yeast reporter strains. FLp53α protein and its Δ-isoforms failed to transactivate a minimal promoter when only a G4-prone sequence was inserted at the site. It has been shown recently that G4s have an inhibitory effect on translation in vivo in the yeast system [43]. Our results showed that Δ160p53α expressed together with FLp53α decreased transactivation at the p53 RE. These new data are in agreement with previously published apoptosis assays, where Δ160p53 inhibits apoptosis, in contrast to Δ133p53 [44]. On the other hand, the Δ133p53α and Δ160p53α isoforms failed

to decrease transactivation of the p53 RE presented together with a G4-prone sequence in front of the RE; in fact, there was a slight increase in transactivation (Figure 5A). This result suggests that hetero-tetramerization of Δ133p53α or Δ160p53α with FLp53α (contrary to Δ40p53α) does not inhibit transactivation at p53 targets associated with a G4 structure, while in the case of Δ40p53α, competition between isoform specific homo-tetramers or the formation of hetero-tetramers can lead to the inhibition of the transactivation potential of FLp53α at these sites (Figure 8).

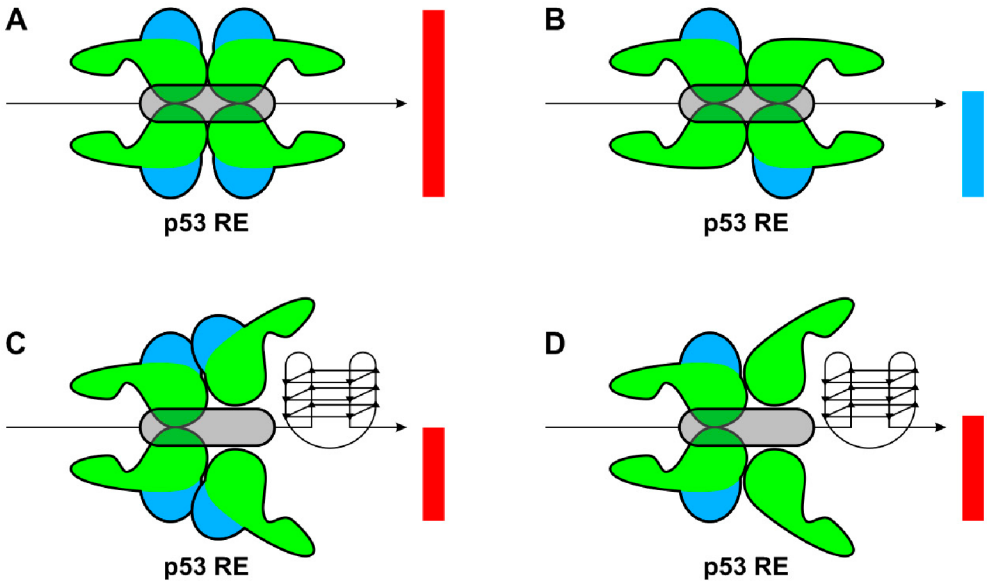


Figure 8. Schematic model of p53 isoforms binding to a RE associated with G4 sequence. (**A**) FLp53α or Δ40p53α bind effectively to the RE and there is a high or moderate level of transactivation. (**B**) Δ133p53α and Δ160p53α inhibit FLp53α transactivation, (**C**) the presence of a G4 close to the RE decreases accessibility of the TA domains and FLp53α transactivation, (**D**) which is not more inhibited by Δ133p53α and Δ160p53α, although steric protein orientation is impaired due to the G4 structure. TA is the blue domain, rest of the protein is in green, column represent transactivation induced by p53 complex (red column FLp53α, blue FLp53α with Δ133p53α and Δ160p53α isoforms).

Therefore, it appears that the composition of the p53 isoforms could be a selective determinant in p53 transactivation specificity, resulting not only from the p53 RE sequence, but also from structural DNA features, particularly a G4 upstream or downstream of the p53 RE. The G4-prone sequences localized in close proximity to the PUMA p53 RE suggests that G4 formation could be an additional feature that determines the effectiveness of p53 transcriptional regulation. The co-expression of different p53 isoforms may increase plasticity through a compromise between effective FLp53 homotetramers at RE sites embedded in structurally favorable contexts and less effective, but sterically more beneficial heterotetramers, at RE sites flanked by structured motifs such as G4.

4. Methods

4.1. Preparation of Plasmids to Express p53α Isoforms

Vectors containing the coding sequences of p53α isoforms were prepared by the Gateway cloning system (detailed in [45]). As the destination vector, pAG414GALccdB-HA containing the inducible GAL promoter and pAG415GPDccdB-HA with the constitutive GPD promoter were used. Destination vectors containing the cDNAs of p53α isoforms were isolated from *E. coli* STBL3 strain using a commercial plasmid extraction kit (Omega-Biotek, Norcross, USA).

4.2. Preparation of Yeast Isogenic Strains by Delitto Perfetto Homologous Recombination

S. cerevisiae haploid strain yLFM-ICORE (MATα leu2–3nic strains, 112 trp1–1 his3–11,15 can1–100; ura3–1; ade2:RE:pCyc1::LUC1) was used for deriving a panel of isogenic reporter strains, which differ in the presence of a p53 RE and a G4 prone sequence (Table 1). The double counterselectable- REporter ICORE cassette was replaced by a targeting oligonucleotide, consisting of 30 nt flanking homology and the RE + G4 as an intervening sequence, following the protocol described in [46]. Replacement was facilitated by induction of a single site-specific DNA double strand break at the ICORE site by the homing endonuclease I-SceI, selected by exploiting resistance to 5-fluoro-orotic acid caused by loss of the ICORE cassette and confirmed by colony PCR and Sanger sequencing. The obtained yeast reporter strains differing in the p53 target site were purified and transformed with a plasmid for the expression of specific p53α isoforms.

Table 1. Sequences cloned into luciferase promoter regions into yeast isogenic reporter strain (PUMA sequence – highlighted by grey, G-repeats – bold).

Region	Sequence 5′–3′
PUMA	CTGCAAGTCCTGACTTGTCC
PUMA–G4	CTGCAAGTCCTGACTTGTCC**GGGGCGGGGGACGGGGGAGGGG**
G4–PUMA	**GGGGCGGGGGACGGGGGAGGGG** CTGCAAGTCCTGACTTGTCC
G4	**GGGGCGGGGGACGGGGGAGGGG**

4.3. Circular Dichroism (CD) Spectroscopy

CD measurements were carried out in a Jasco 815 (Jasco International Co. Ltd., Tokyo, Japan) dichrograph in 1 cm path-length quartz Hellma microcells placed in a thermostatically regulated cell holder at 23 °C. A set of four scans was averaged for each sample with a data pitch of 0.5 nm and 100 nm/min scan speed. CD signal was expressed as the difference in the molar absorption, Δε of the left- and right-handed circularly polarized light, molarity being related to DNA strands; buffer: 50 mM KCl, 5 mM Tris/HCl pH 8.

4.4. Transformation of Yeast Strains

Yeast were transformed by a method based on mixing cells and DNA in the presence of lithium acetate, TE, PEG, DMSO and performing heat shock, starting from saturated overnight cultures [47]. Double transformants were selected by auxotrophic selection on plates lacking both tryptophan and leucine.

4.5. Luciferase Assay

Purified transformant colonies were inoculated on 96-well plates in 120 μL selective media containing 2% raffinose as a carbon source and different concentrations of galactose to induce p53α isoform expression from the GAL promoter of the pAG414GAL vector. Luciferase was measured as described [40]. To ascertain p53 protein expression, samples used for the transcription analysis were also used to prepare protein extracts for immunodetection by western blotting.

4.6. Western Blot

Yeast cell lysis was performed as described [48]. Protein extracts were quantified using the Bradford assay. Proteins (80 μg) were electrophoresed using 12.5% acrylamide sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane. Specific antibodies directed against p53 were donated by Dr. Vojtěšek and the membranes were incubated as described [49–51]. The signal was detected using the ECL Select reagent (Pierce Fast Western Blot Kit, Thermo Fisher, WA, USA) and results were visualized as chemiluminiscence on LAS 3000. Results are shown in Figure S1.

4.7. Statistical Analysis

Transactivation data were plotted as fold luciferase induction relative to a control reporter activity, measured with cells that do not contain a p53 expression plasmid and cultured under the same conditions. Mean and standards deviation of at least three biological replicates are presented. Statistical significance was evaluated using Student's t-test.

4.8. G4Hunter Analyses

The DNA sequence of the p53RE that regulates PUMA on chromosome 19 including 100 bp before and 20-bp after the p53RE was downloaded in FASTA format from the National Center for Biotechnology Information (NCBI) [52]. The sequence was analyzed by G4Hunter web [34] for the presence and localization of G-quadruplex prone sequences with parameters of "25" for window and G4Hunter score above 1.2.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/21/1/127/s1.

Author Contributions: Conceptualization, V.B. and A.I.; Data curation, O.P. and N.B.; Formal analysis, O.P. and N.B.; Funding acquisition, V.B. and M.F.; Investigation, O.P. and N.B.; Methodology, N.V., J.C., and V.B.; Project administration, V.B.; Resources, V.B., A.I., and M.F.; Supervision, V.B. and A.I.; Validation, J.C.; Visualization, O.P. and N.B.; Writing–original draft, O.P., N.B., and V.B.; Writing—review & editing, V.B. and A.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by The Czech Science Foundation (18-15548S), by project FCH-S-18-5334 and by the SYMBIT project reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477 financed from the ERDF.

Acknowledgments: We thank Borivoj Vojtesek for providing the p53 antibodies and Philip Coates for proofreading and editing the manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interests. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Lane, D.P. Cancer. p53, guardian of the genome. *Nature* **1992**, *358*, 15–16. [CrossRef] [PubMed]
2. Oren, M. Decision making by p53: Life, death and cancer. *Cell Death Differ.* **2003**, *10*, 431–442. [CrossRef] [PubMed]
3. Suzuki, K.; Dashzeveg, N.; Lu, Z.G.; Taira, N.; Miki, Y.; Yoshida, K. Programmed cell death 6, a novel p53-responsive gene, targets to the nucleus in the apoptotic response to DNA damage. *Cancer Sci.* **2012**, *103*, 1788–1794. [CrossRef] [PubMed]
4. Wang, X.; Simpson, E.R.; Brown, K.A. p53: Protection against Tumor Growth beyond Effects on Cell Cycle and Apoptosis. *Cancer Res.* **2015**, *75*, 5001–5007. [CrossRef]
5. Levine, A.J. p53, the Cellular Gatekeeper for Growth and Division. *Cell* **1997**, *88*, 323–331. [CrossRef]
6. Khoury, M.P.; Bourdon, J.-C. p53 Isoforms: An Intracellular Microprocessor? *Genes Cancer* **2011**, *2*, 453–465. [CrossRef]
7. Joruiz, S.M.; Bourdon, J.-C. p53 Isoforms: Key Regulators of the Cell Fate Decision. *Cold Spring Harb. Perspect. Med.* **2016**, *6*, a026039. [CrossRef]
8. Meek, D.W.; Anderson, C.W. Posttranslational Modification of p53: Cooperative Integrators of Function. *Cold Spring Harb. Perspect. Biol.* **2009**, *1*, a000950. [CrossRef]
9. Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* **1994**, *265*, 346–355. [CrossRef] [PubMed]
10. Ghosh, A.; Stewart, D.; Matlashewski, G. Regulation of Human p53 Activity and Cell Localization by Alternative Splicing. *Mol. Cell Biol.* **2004**, *24*, 7987–7997. [CrossRef] [PubMed]
11. Marcel, V.; Tran, P.L.T.; Sagne, C.; Martel-Planche, G.; Vaslin, L.; Teulade-Fichou, M.-P.; Hall, J.; Mergny, J.-L.; Hainaut, P.; Van Dyck, E. G-quadruplex structures in TP53 intron 3: Role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **2011**, *32*, 271–278. [CrossRef] [PubMed]
12. Pavletich, N.P.; Chambers, K.A.; Pabo, C.O. The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. *Genes Dev.* **1993**, *7*, 2556–2564. [CrossRef] [PubMed]

13. Nutthasirikul, N.; Limpaiboon, T.; Leelayuwat, C.; Patrakitkomjorn, S.; Jearanaikoon, P. Ratio disruption of the Δ133p53 and TAp53 isoform equilibrium correlates with poor clinical outcome in intrahepatic cholangiocarcinoma. *Int. J. Oncol.* **2013**, *42*, 1181–1188. [CrossRef] [PubMed]
14. Chambers, S.K.; Martinez, J.D. The significance of p53 isoform expression in serous ovarian cancer. *Future Oncol.* **2012**, *8*, 683–686. [CrossRef]
15. El-Deiry, W.S.; Kern, S.E.; Pietenpol, J.A.; Kinzler, K.W.; Vogelstein, B. Definition of a consensus binding site for p53. *Nat. Genet.* **1992**, *1*, 45. [CrossRef]
16. Weinberg, R.L.; Veprintsev, D.B.; Bycroft, M.; Fersht, A.R. Comparative Binding of p53 to its Promoter and DNA Recognition Elements. *J. Mol. Biol.* **2005**, *348*, 589–596. [CrossRef]
17. Vyas, P.; Beno, I.; Xi, Z.; Stein, Y.; Golovenko, D.; Kessler, N.; Rotter, V.; Shakked, Z.; Haran, T.E. Diverse p53/DNA binding modes expand the repertoire of p53 response elements. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10624–10629. [CrossRef]
18. Brázda, V.; Coufal, J. Recognition of local DNA structures by p53 protein. *Int. J. Mol. Sci.* **2017**, *18*, 375. [CrossRef]
19. Brázda, V.; Fojta, M. The Rich World of p53 DNA Binding Targets: The Role of DNA Structure. *Int. J. Mol. Sci.* **2019**, *20*, 5605. [CrossRef]
20. Qian, H.; Wang, T.; Naumovski, L.; Lopez, C.D.; Brachmann, R.K. Groups of p53 target genes involved in specific p53 downstream effects cluster into different classes of DNA binding sites. *Oncogene* **2002**, *21*, 7901–7911. [CrossRef]
21. Menendez, D.; Inga, A.; Resnick, M.A. The expanding universe of p53 targets. *Nat. Rev. Cancer* **2009**, *9*, 724–737. [CrossRef] [PubMed]
22. Göhler, T.; Reimann, M.; Cherny, D.; Walter, K.; Warnecke, G.; Kim, E.; Deppert, W. Specific Interaction of p53 with Target Binding Sites Is Determined by DNA Conformation and Is Regulated by the C-terminal Domain. *J. Biol. Chem.* **2002**, *277*, 41192–41203. [CrossRef] [PubMed]
23. Jagelska, E.B.; Brazda, V.; Pecinka, P.; Palecek, E.; Fojta, M. DNA topology influences p53 sequence-specific DNA binding through structural transitions within the target sites. *Biochem. J.* **2008**, *412*, 57–63. [CrossRef] [PubMed]
24. Coufal, J.; Jagelská, E.B.; Liao, J.C.C.; Brazda, V. Preferential binding of p53 tumor suppressor to p21 promoter sites that contain inverted repeats capable of forming cruciform structure. *Biochem. Biophys. Res. Commun.* **2013**, *441*, 83–88. [CrossRef] [PubMed]
25. Brazda, V.; Čechová, J.; Battistin, M.; Coufal, J.; Jagelská, E.B.; Raimondi, I.; Inga, A. The structure formed by inverted repeats in p53 response elements determines the transactivation activity of p53 protein. *Biochem. Biophys. Res. Commun.* **2017**, *483*, 516–521. [CrossRef] [PubMed]
26. Petr, M.; Helma, R.; Polaskova, A.; Krejci, A.; Dvorakova, Z.; Kejnovska, I.; Navratilova, L.; Adamik, M.; Vorlickova, M.; Brazdova, M. Wild-type p53 binds to MYC promoter G-quadruplex. *Biosci. Rep.* **2016**, *36*, e00397. [CrossRef]
27. Brazdova, M.; Tichy, V.; Helma, R.; Bazantova, P.; Polaskova, A.; Krejci, A.; Petr, M.; Navratilova, L.; Ticha, O.; Nejedly, K.; et al. p53 Specifically Binds Triplex DNA In Vitro and in Cells. *PLoS ONE* **2016**, *11*, e0167439. [CrossRef]
28. Degtyareva, N.; Subramanian, D.; Griffith, J.D. Analysis of the binding of p53 to DNAs containing mismatched and bulged bases. *J. Biol. Chem.* **2001**, *276*, 8778–8784. [CrossRef]
29. Stros, M.; Muselikova-Polanska, E.; Pospisilova, S.; Strauss, F. High-affinity binding of tumor-suppressor protein p53 and HMGB1 to hemicatenated DNA loops. *Biochemistry* **2004**, *43*, 7215–7225. [CrossRef]
30. Spradling, A.; Ganetsky, B.; Hieter, P.; Johnston, M.; Olson, M.; Orr-Weaver, T.; Rossant, J.; Sanchez, A.; Waterston, R. New roles for model genetic organisms in understanding and treating human disease: Report from the 2006 Genetics Society of America meeting. *Genetics* **2006**, *172*, 2025–2032.
31. Lion, M.; Raimondi, I.; Donati, S.; Jousson, O.; Ciribilli, Y.; Inga, A. Evolution of p53 Transactivation Specificity through the Lens of a Yeast-Based Functional Assay. *PLoS ONE* **2015**, *10*. [CrossRef] [PubMed]
32. Guaragnella, N.; Palermo, V.; Galli, A.; Moro, L.; Mazzoni, C.; Giannattasio, S. The expanding role of yeast in cancer research and diagnosis: Insights into the function of the oncosuppressors p53 and BRCA1/2. *FEMS Yeast Res.* **2014**, *14*, 2–16. [CrossRef] [PubMed]
33. Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [CrossRef] [PubMed]

34. Brázda, V.; Kolomazník, J.; Lýsek, J.; Bartas, M.; Fojta, M.; Šťastný, J.; Mergny, J.-L. G4Hunter web application: A web server for G-quadruplex prediction. *Bioinformatics* **2019**, *35*, 3493–3495. [CrossRef]
35. Siddiqui-Jain, A.; Grand, C.L.; Bearss, D.J.; Hurley, L.H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593–11598. [CrossRef]
36. Yang, D.; Hurley, L.H. Structure of the biologically relevant G-quadruplex in the c-MYC promoter. *Nucleosides Nucleotides Nucleic Acids* **2006**, *25*, 951–968. [CrossRef]
37. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R.A.; Yang, D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **2006**, *34*, 2723–2735. [CrossRef]
38. Del Villar-Guerra, R.; Trent, J.O.; Chaires, J.B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 7171–7175. [CrossRef]
39. Nguyen, T.-A.T.; Grimm, S.A.; Bushel, P.R.; Li, J.; Li, Y.; Bennett, B.D.; Lavender, C.A.; Ward, J.M.; Fargo, D.C.; Anderson, C.W.; et al. Revealing a human p53 universe. *Nucleic Acids Res.* **2018**, *46*, 8153–8167. [CrossRef]
40. Jordan, J.J.; Menendez, D.; Inga, A.; Nourredine, M.; Bell, D.; Resnick, M.A. Noncanonical DNA Motifs as Transactivation Targets by Wild Type and Mutant p53. *PLoS Genet.* **2008**, *4*, e1000104. [CrossRef]
41. Brázda, V.; Laister, R.C.; Jagelská, E.B.; Arrowsmith, C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* **2011**, *12*, 33. [CrossRef] [PubMed]
42. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007**, *35*, 406–413. [CrossRef] [PubMed]
43. Tokan, V.; Puterova, J.; Lexa, M.; Kejnovsky, E. Quadruplex DNA in long terminal repeats in maize LTR retrotransposons inhibits the expression of a reporter gene in yeast. *BMC Genom.* **2018**, *19*, 184. [CrossRef] [PubMed]
44. Candeias, M.M.; Hagiwara, M.; Matsuda, M. Cancer-specific mutations in p53 induce the translation of Δ160p53 promoting tumorigenesis. *EMBO Rep.* **2016**, *17*, 1542–1551. [CrossRef] [PubMed]
45. User guide: Gateway Technology with Clonase II-A universal technology to clone DNA sequences for functional analysis and expression in multiple systems. Available online: http://tools.thermofisher.com/content/sfs/manuals/gateway_clonaseii_man.pdf (accessed on 10 January 2019).
46. Storici, F.; Resnick, M.A. The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzym.* **2006**, *409*, 329–345.
47. Sharma, V.; Monti, P.; Fronza, G.; Inga, A. Human transcription factors in yeast: The fruitful examples of P53 and NF-κB. *FEMS Yeast Res.* **2016**, *16*. [CrossRef] [PubMed]
48. Andreotti, V.; Ciribilli, Y.; Monti, P.; Bisio, A.; Lion, M.; Jordan, J.; Fronza, G.; Menichini, P.; Resnick, M.A.; Inga, A. p53 Transactivation and the Impact of Mutations, Cofactors and Small Molecules Using a Simplified Yeast-Based Screening System. *PLoS ONE* **2011**, *6*, e20643. [CrossRef]
49. Vojtesek, B.; Dolezalova, H.; Lauerova, L.; Svitakova, M.; Havlis, P.; Kovarik, J.; Midgley, C.A.; Lane, D.P. Conformational changes in p53 analysed using new antibodies to the core DNA binding domain of the protein. *Oncogene* **1995**, *10*, 389–393.
50. Brazda, V.; Muller, P.; Brozkova, K.; Vojtesek, B. Restoring wild-type conformation and DNA-binding activity of mutant p53 is insufficient for restoration of transcriptional activity. *Biochem. Biophys. Res. Commun.* **2006**, *351*, 499–506. [CrossRef]
51. Pospísilová, S.; Brázda, V.; Amrichová, J.; Kamermeierová, R.; Palecek, E.; Vojtesek, B. Precise characterisation of monoclonal antibodies to the C-terminal region of p53 protein using the PEPSCAN ELISA technique and a new non-radioactive gel shift assay. *J. Immunol. Methods* **2000**, *237*, 51–64. [CrossRef]
52. Sayers, E.W.; Agarwala, R.; Bolton, E.E.; Brister, J.R.; Canese, K.; Clark, K.; Connor, R.; Fiorini, N.; Funk, K.; Hefferon, T.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2019**, *47*, D23–D28. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Příloha B

Evaluating the Influence of a G-Quadruplex Prone Sequence on the Transactivation Potential by Wild-Type and/or Mutant P53 Family Proteins through a Yeast-Based Functional Assay

genes

MDPI

Article

Evaluating the Influence of a G-Quadruplex Prone Sequence on the Transactivation Potential by Wild-Type and/or Mutant P53 Family Proteins through a Yeast-Based Functional Assay

Paola Monti [1], Vaclav Brazda [2,3], Natália Bohálová [2,4], Otília Porubiaková [2,3], Paola Menichini [1], Andrea Speciale [1], Renata Bocciardi [5,6], Alberto Inga [7,*] and Gilberto Fronza [1,*]

1 Mutagenesis and Cancer Prevention Unit, IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy; paola.monti@hsanmartino.it (P.M.); paola.menichini@hsanmartino.it (P.M.); andrea.speciale@hsanmartino.it (A.S.)
2 Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 61265 Brno, Czech Republic; vaclav@ibp.cz (V.B.); natalia.bohalova@ibp.cz (N.B.); o.porubiakova@gmail.com (O.P.)
3 Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 61200 Brno, Czech Republic
4 Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic
5 Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DiNOGMI), University of Genoa, Largo P. Daneo 3, 16132 Genoa, Italy; bocciardi@unige.it
6 Medical Genetics Unit, IRCCS Istituto Giannina Gaslini, via G. Gaslini 5, 16147 Genoa, Italy
7 Laboratory of Transcriptional Networks, Department of Cellular, Computational and Integrative Biology, CIBIO, University of Trento, via Sommarive 9, 38123 Trento, Italy
* Correspondence: alberto.inga@unitn.it (A.I.); gilberto.fronza@hsanmartino.it (G.F.); Tel.: +39-0461-283714 (A.I.); +39-010-5558225 (G.F.)

Abstract: P53, P63, and P73 proteins belong to the P53 family of transcription factors, sharing a common gene organization that, from the P1 and P2 promoters, produces two groups of mRNAs encoding proteins with different N-terminal regions; moreover, alternative splicing events at C-terminus further contribute to the generation of multiple isoforms. P53 family proteins can influence a plethora of cellular pathways mainly through the direct binding to specific DNA sequences known as response elements (REs), and the transactivation of the corresponding target genes. However, the transcriptional activation by P53 family members can be regulated at multiple levels, including the DNA topology at responsive promoters. Here, by using a yeast-based functional assay, we evaluated the influence that a G-quadruplex (G4) prone sequence adjacent to the p53 RE derived from the apoptotic *PUMA* target gene can exert on the transactivation potential of full-length and N-terminal truncated P53 family α isoforms (wild-type and mutant). Our results show that the presence of a G4 prone sequence upstream or downstream of the P53 RE leads to significant changes in the relative activity of P53 family proteins, emphasizing the potential role of structural DNA features as modifiers of P53 family functions at target promoter sites.

Keywords: P53 family; yeast; G-quadruplex (G4) prone sequence; wild-type and mutant P53/P63 proteins; transactivation potential

check for updates

Citation: Monti, P.; Brazda, V.; Bohálová, N.; Porubiaková, O.; Menichini, P.; Speciale, A.; Bocciardi, R.; Inga, A.; Fronza, G. Evaluating the Influence of a G-Quadruplex Prone Sequence on the Transactivation Potential by Wild-Type and/or Mutant P53 Family Proteins through a Yeast-Based Functional Assay. *Genes* **2021**, *12*, 277. https://doi.org/10.3390/genes12020277

Academic Editor: Alvaro Galli

Received: 21 January 2021
Accepted: 10 February 2021
Published: 15 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The P53 family of transcription factors (TFs) is composed of P53, P63, and P73 proteins [1–3] that share an N-terminal transactivation domain, a central sequence-specific DNA-binding domain, and a C-terminal tetramerization domain. At the C-terminus, a sterile α motif domain (SAM), probably involved in protein–protein interactions, is present only in P63 and P73 proteins. These TFs, by inducing a plethora of target genes, can influence different cellular pathways including proliferation, apoptosis, DNA repair, angiogenesis, senescence, metabolism, and differentiation [4,5].

To further complicate the scenario, multiple P53, P63, and P73 isoforms are generated from mechanisms of alternative usage of promoters, splicing sites, and/or translation initiation sites [6]. All P53 family members can be transcribed starting from different promoters resulting in variants with different N-terminal sequences (i.e., full-length P53, ΔN40P53, TAP63, and TAP73 from the P1 promoter, and ΔN133P53, ΔN160P53, ΔNP63, and ΔNP73 from the P2 promoter). Moreover, C-terminal variations occur as a result of alternative splicing, giving rise to at least three different isoforms (i.e., α, β, γ for P53 and P63; α, β, γ, δ, ε, ζ, η for P73).

Despite these similarities, the overlap in cellular functions between P53, P63, and P73 proteins is limited. *TP53* is the most frequently mutated gene in human sporadic cancers and *TP53* germ-line mutations are associated with the development of the cancer-prone Li-Fraumeni and Li-Fraumeni-like syndromes [7]. Conversely, the *TP63* gene, being critical for the correct development of ectodermal-derived tissues, is associated with the occurrence of a subset of ectodermal dysplasia syndromes (i.e., P63-associated disorders) due to *TP63* germ-line mutations [3,8,9]. Lastly, P73 contributes to neural and immune system functions but no genetic disorder has been linked to the gene [10,11], possibly because any such *TP73* mutation might produce severe defects that are embryonically lethal.

To regulate genes expression, the P53 family members, acting mainly as TFs, share a DNA response element (RE) which consists of two degenerate decameric sequences separated by a variable spacer [RRRCWWGYYY-(n)-RRRCWWGYYY (*R* = purine; *W* = A/T; *Y* = pyrimidine; *n* = 0–13)] [12–14]. Even though the DNA-binding specificities for P53, P63, and P73 appear comparable, differences for certain DNA sequences have been reported [15–17]. Moreover, gene expression and genome-wide occupancy studies reveal a partial overlap between gene networks of P53 family members [18]; in fact, many genes appeared to be exclusively targeted by one of the P53 family members [19–22].

Recently, it was shown that DNA topology can contribute to regulating P53 DNA affinity and specificity [23,24]. It was demonstrated that P53 protein binds to various DNA structures stabilized by DNA topological stress such as G-quadruplex (G4)-forming sequences, cruciforms, and other local DNA structures [24–26], as recently reviewed [27]. Furthermore, some P53 mutants have a high predisposition for binding to Transcription Start Site-associated G/C-rich regions, particularly G4s, which are prone to form non-B DNA structures [28].

The yeast *Saccharomyces cerevisiae* is a recognized model system for understanding different aspects of human biology. Thanks to the evolutionary conservation of basic components of the transcription machinery, many TFs, including P53 family proteins, when ectopically expressed in yeast cells can modulate the expression rates of reporter genes, acting through promoters engineered to contain cognate target REs [29–31]. Moreover, a particularly versatile approach based on the *Delitto Perfetto* technique is available in yeast to target the selected genomic loci through homologous recombination [32], leading to the development of a matrix of results where the functional interaction of TFs binding sites with other nearby cis-elements can also be evaluated [33].

Based on the assumption that DNA topology can be an important factor in establishing P53-dependent transactivation at target genomic sites, we previously evaluated the influence of a G4 DNA prone sequence in the proximity of a P53 RE on the transactivation potential of full-length and N-terminal truncated P53 α isoforms [34]. The results showed that a G4 prone sequence alone is not sufficient for transcriptional activation by the different P53 α isoforms (i.e., full length, ΔN40, ΔN133, and ΔN160); however, its presence in proximity to a P53 RE leads to significantly different fold changes in transcriptional activity and dynamics between the co-expressed P53 isoforms.

Here, by using the same experimental approach in yeast, we extended this study to ΔN and TA α variants of P63 and P73 proteins. Moreover, since human germ-line *TP53* and *TP63* mutations are responsible for the development of specific genetic disorders and considering that *TP53* somatic mutations are frequent in human cancers, we also evaluated the effect of the presence of the same G4 prone sequence on the functional features of

mutant P53 and P63 proteins. The results obtained highlight how the presence of a G4 prone sequence adjacent to a P53 RE impact the transcriptional activity of wild-type and mutant proteins from all P53 family members.

2. Materials and Methods

2.1. Yeast Strains and Media

A panel of isogenic reporter strains, which differ only in the presence and position of a P53 RE [from *PUMA* (p53 up-regulated modulator of apoptosis) target gene: 5′-CTGCAAGTCCTGACTTGTCC-3′] and a G4 prone sequence [from KSHV (Kaposi sarcoma-associated herpes virus): 5′-GGGGCGGGGGACGGGGGAGGGG-3′] both located upstream of the luciferase reporter gene (LUC1) were used [34–36]. The *PUMA* RE was selected in virtue of its moderate strength and for the presence of physiological G4 sequences around the RE in the human promoter. This panel includes yLFM-PUMA (PUMA RE), yLFM-KSHV (G4), yLFM-KSHV-PUMA (G4 upstream of PUMA RE), and yLFM-PUMA-KSHV (G4 downstream of PUMA RE) strains. Yeast cells were grown in YPDA medium (1% yeast extract, 2% peptone, 2% dextrose, 200 mg/L adenine) or selective medium (with or without 2% agar), containing dextrose or raffinose as a carbon source plus adenine (200 mg/L), but in the absence of tryptophan and/or leucine (Sigma-Aldrich, Saint Louis, MO, USA; Biokar Diagnostics, Allonne, France). Galactose (Sigma-Aldrich, Saint Louis, MO, USA) was added to the medium for the modulation of P53 family proteins expression under the inducible GAL1, 10 promoter [37]. Yeast manipulations were performed, as previously described [33,38].

2.2. Yeast Vectors

Wild-type P53 [full length (i.e., P53α corresponding to the well-known 393 amino-acids long protein) and ΔN40α], P63 (ΔNα and TAα) and P73 (ΔNα and TAα) proteins were expressed by yeast pTSG-based vectors (inducible GAL1,10 promoter, TRP1) [39–41] as well as mutant full-length P53 (i.e., R175H and R282W) and ΔNP63α (i.e., G134V and R204W) proteins [41,42]. For co-expression experiments vectors expressing wild-type full-length P53 and ΔNP63α under the constitutive ADH1 promoter [pLS76 (LEU2) and pLS-ΔNP63α (LEU2), respectively] were used along with the yeast plasmids described above [41,43]. Empty vectors pRS314 (TRP1) and pRS315 (LEU2) were available.

2.3. Yeast Functional Assay

Quantitative functional assays were performed according to the miniaturized protocol we developed [33,38]. Briefly, yeast transformants were grown at 30 °C in a selective medium containing 1% Galactose in order to modulate P53 family proteins expression; after 8 h of growth OD (600 nm) was measured. Twenty μL of cell suspension was transferred into a white plate and mixed with an equal volume of PLB buffer 2X (Promega Italia, Milan, Italy) to obtain the lysis of yeast cells. After 15 min of shaking at room temperature, firefly luciferase substrate (20 μL, Bright Glo, Promega) was added. Luciferase activity was measured using a multilabel plate reader (Mithras LB940, Berthold Technologies, Calmbacher, Germany). The transactivation ability of wild-type and mutant P53 family proteins was measured as relative light units (RLUs) and normalized first to the cultures' absorbance (600 nm). Then the fold changes were calculated using as reference the normalized RLUs measured from cultures of each yeast reporter strains transformed with the appropriate empty vector(s) (pRS314 or pRS314 + pRS315 depending on the experiments). The relative activity from the same fold change data was calculated by comparing the activity of a protein/allele of interest with the activity of a chosen reference. Data derive from two technical replicates with at least two biological replicates except for yLFM-KSHV strain measurements (i.e., three biological replicates).

2.4. Statistical Analysis

Statistical analysis was performed by using two-way ANOVA and Tukey's multiple comparisons test or one-way ANOVA and Dunnett's multiple comparisons test (Graphpad Prism 9) (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; ns: not significant).

3. Results

3.1. The Presence of a G4 Prone Sequence Adjacent to a P53 RE Alters the Relative Activity of Wild-Type P53 Family Proteins

To elucidate the role of a G4 prone sequence on the transcriptional activity of P53 family members, we focused our analysis on α isoforms including ΔN40P53, full-length P53, ΔNP63, TAP63, ΔNP73, and TAP73 (ΔN40P53, ΔNP63, and ΔNP73 were indicated as ΔN variants; full-length P53, TAP63, and TAP73 were indicated as TA variants) by using a yeast-based functional assay. Consistent with previous observations regarding P53, wild-type P63 and P73 proteins (ΔN and TA variants) were unable to activate transcription (measured as fold change over empty vector) from a G4 prone sequence (yLFM-KSHV) (Figure 1A, right panel) compared to the control yLFM-PUMA (Figure 1A, left panel). Similarly, the presence of the same G4 prone sequence upstream (yLFM-KSHV-PUMA) or downstream (yLFM-PUMA-KSHV) of the PUMA P53 RE led to a significant decrease in transactivation (Figure 1B), especially in the strain where the G4 prone sequence is downstream of the RE; in this strain, the full-length P53, TAP63α, and TAP73α showed an eleven-, seven- and four-fold decrease in activity, respectively (Figure 1B, right panel).

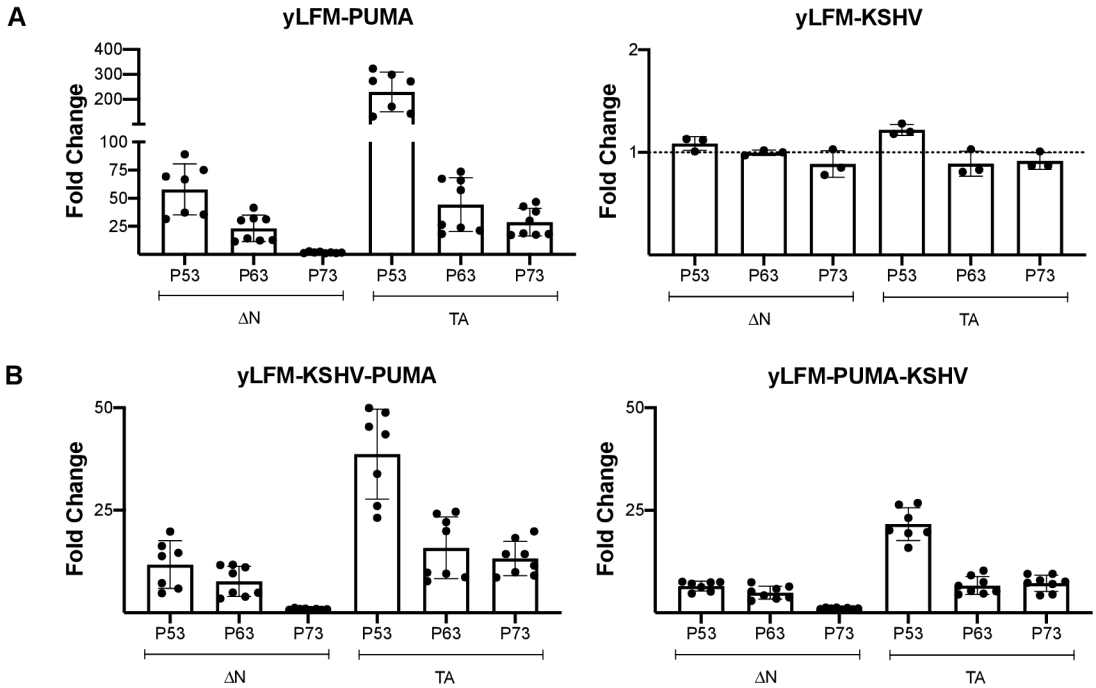


Figure 1. Fold change transactivation by wild-type P53 family proteins. Yeast cells expressing different P53 family isoforms by an inducible GAL1,10 promoter were grown for 8 h in Galactose 1% to evaluate their transactivation ability. (**A**) Evaluation of wild-type P53, P63, and P73 (ΔN variants: ΔN40P53α, ΔNP63α, and ΔNP73α; TA variants: full-length P53, TAP63α, and TAP73α) transcriptional activity as fold change over the empty vector in presence of the P53 response element (RE) from the *PUMA* target gene (yLFM-PUMA) or a G4 prone sequence (yLFM-KSHV). (**B**) Evaluation of transcriptional activity as in panel A in the presence of the G4 prone sequence upstream (yLFM-KSHV-PUMA) or downstream (yLFM-PUMA-KSHV) of the P53 RE from the *PUMA* target gene.

The results also showed that, independently from the presence of the G4 prone sequence, the TA variant from P53, P63, or P73 proteins is always significantly more active than the corresponding ΔN variant (Figure 1A,B). However, the positioning of the

G4 prone sequence downstream of the P53 RE (yLFM-PUMA-KSHV) leads to a slight but significant decrease in the relative activity of TA over ΔN variants for both P53 (full-length P53/ΔN40P53α) and P63 (TAP63α/ΔNP63α) (Figure 2A). Further, by comparing across P53 family proteins, the presence of the same G4 prone sequence either upstream or downstream of the RE (yLFM-KSHV-PUMA or yLFM-PUMA-KSHV, compared to yLFM-PUMA) significantly changes all relative activities, reducing the differences in transcriptional activity between different proteins of the P53 family, as highlighted by the significant increase in ΔNP63α/full-length P53, TAP63α/full-length P53, and TAP73α/full-length P53 relative activity (Figure 2B).

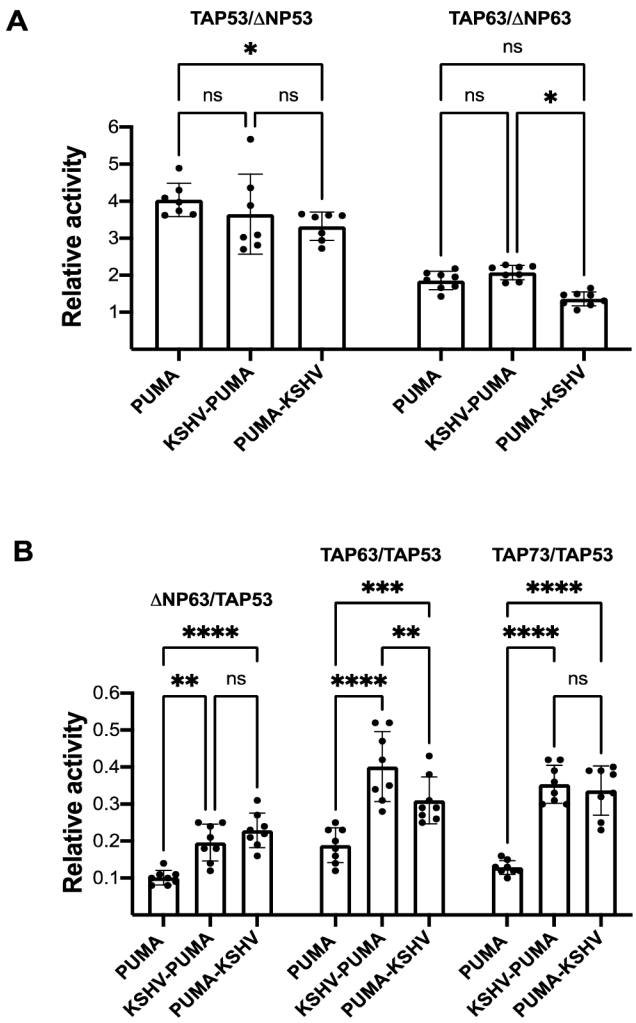


Figure 2. Relative activity of wild-type P53 family α isoforms. (**A**) Comparison of P53 (full-length P53) or P63 (TAP63α) TA variant activity with respect to the corresponding ΔN variant (ΔN40P53α and ΔNP63α, respectively) in yLFM-PUMA, yLFM-KSHV-PUMA, and yLFM-PUMA-KSHV strains. (**B**) Comparison of ΔNP63α, TAP63α, or TAP73α variant activity with respect to TAP53 (full-length P53) variant in yLFM-PUMA, yLFM-KSHV-PUMA, and yLFM-PUMA-KSHV strains. ns: not significant, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

3.2. The Presence of a G4 Prone Sequence Adjacent to a P53 RE Determines a Variation in the Relative Functionality of Mutant P53 and P63 Proteins

Since somatic and germ-line mutations are common to the *TP53* gene in human cancers, while germ-line mutations at the *TP63* locus, mainly affecting at protein level the ΔNP63α isoform, are involved in the development of P63-associated genetic disorders, we decided to extend the functional studies to mutant P53 family proteins. To this aim, the P53 R175H and R282W proteins were selected (as full-length variants), being hot spot

mutations both in sporadic cancers and Li-Fraumeni syndrome; the two mutants are also representative of a loss of function (i.e., R175H) and a partial function (i.e., R282W) mutant, respectively [42,44]. With regard to *TP63* mutations, we selected the two ΔNP63α G134V and ΔNP63α R204W variants that we identified in patients affected by P63-associated disorders and that are representative of mutant proteins with greater or lesser functionality, respectively [41]. We previously confirmed that in our yeast model the observed differences in functional features of the selected P53 or P63 mutant proteins are not due to different levels of protein expression [39,42].

The P53 R175H and ΔNP63α R204W proteins are characterized by consistent loss of activity in all yeast strains independently from the presence of a G4 prone sequence (Figure 3A,B). The presence of the G4 prone sequence also causes an absolute reduction in the transactivation ability of the partial function mutant TAP53 R282W and ΔNP63α G134V proteins. However, the presence of the G4 prone sequence downstream of the P53 RE significantly lowers the difference in activity with respect to the wild-type protein, especially for ΔNP63α G134V mutant (i.e., 71% of wild-type functionality in yLFM-PUMA-KSHV with respect to 28% in yLFM-PUMA) (Supplementary Figure S1).

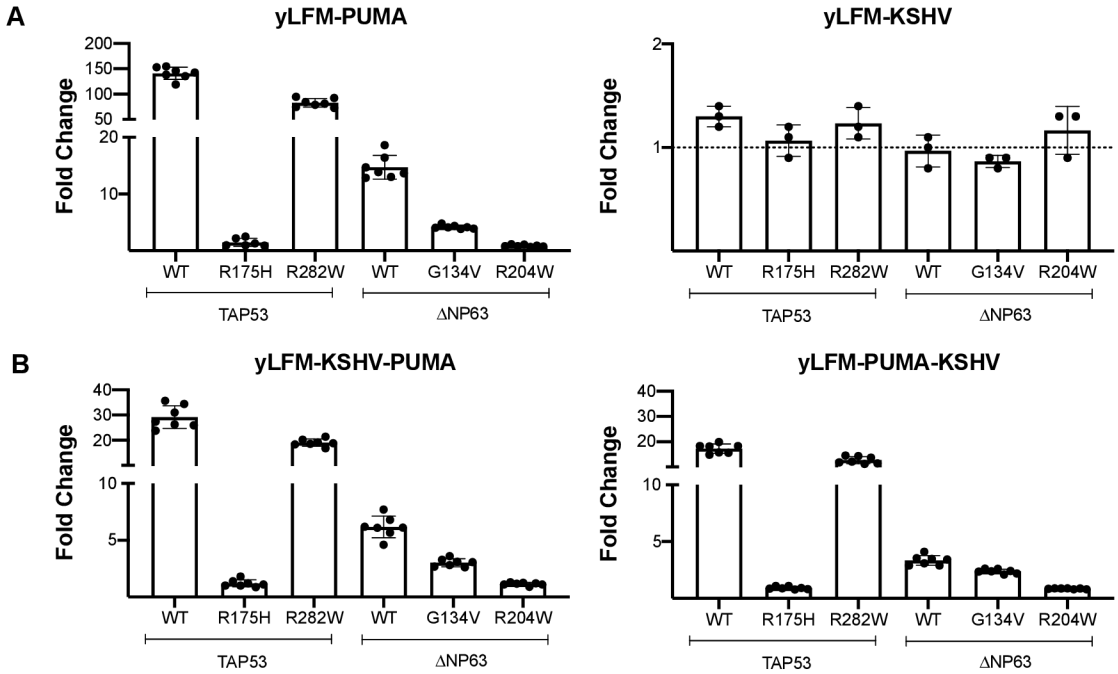


Figure 3. Fold change transactivation by mutant P53 family proteins. Yeast cells expressing wild-type and mutant TAP53 or ΔNP63 proteins by an inducible GAL1,10 promoter were grown for 8 h in Galactose 1% to evaluate the transactivation ability. (**A**) Evaluation of wild-type (TAP53: full-length P53; ΔNP63: ΔNP63α) or mutant (TAP53: R175H and R282W; ΔNP63: G134V and R204W) P53 and P63 proteins transcriptional activity as fold change over the empty vector in presence of the P53 RE from the *PUMA* target gene (yLFM-PUMA) or a G4 prone sequence (yLFM-KSHV). (**B**) Evaluation of transcriptional activity as in panel (**A**) in presence of the G4 prone sequence upstream (yLFM-KSHV-PUMA strain) or downstream (yLFM-PUMA-KSHV) of the P53 RE from the *PUMA* target gene. WT, wild-type.

As previously described, germ-line heterozygous mutations in the *TP53* or *TP63* gene are the molecular basis of the specific P53- and P63-related genetic disorders; then, it is conceivable that in the cells of affected patients, being P53 and P63 tetrameric TFs, wild-type, mixed, and mutant tetramers are formed. Based on this consideration, we simulated the heterozygous status in our yeast model by co-expressing wild-type and mutant P53 or P63 variants, and we evaluated the influence of the G4 prone sequence in this setting.

Although the presence of a G4 prone sequence causes the expected reduction in transactivation ability (Supplementary Figure S2; Supplementary Figure S3), it does not alter the

potential for mutant P53 or P63 proteins to inhibit the corresponding wild-type-mediated transactivation, a feature known as dominant-negative or interfering ability [45]; in fact, the P53 R175H (full-length P53) and P63 R204W (ΔNP63α) mutants that we previously described as dominant negative [41–43,46] retain a comparable potential to inhibit wild-type activity also in the presence of the G4 prone sequence adjacent (upstream or downstream) to the P53 RE (Figure 4A,B).

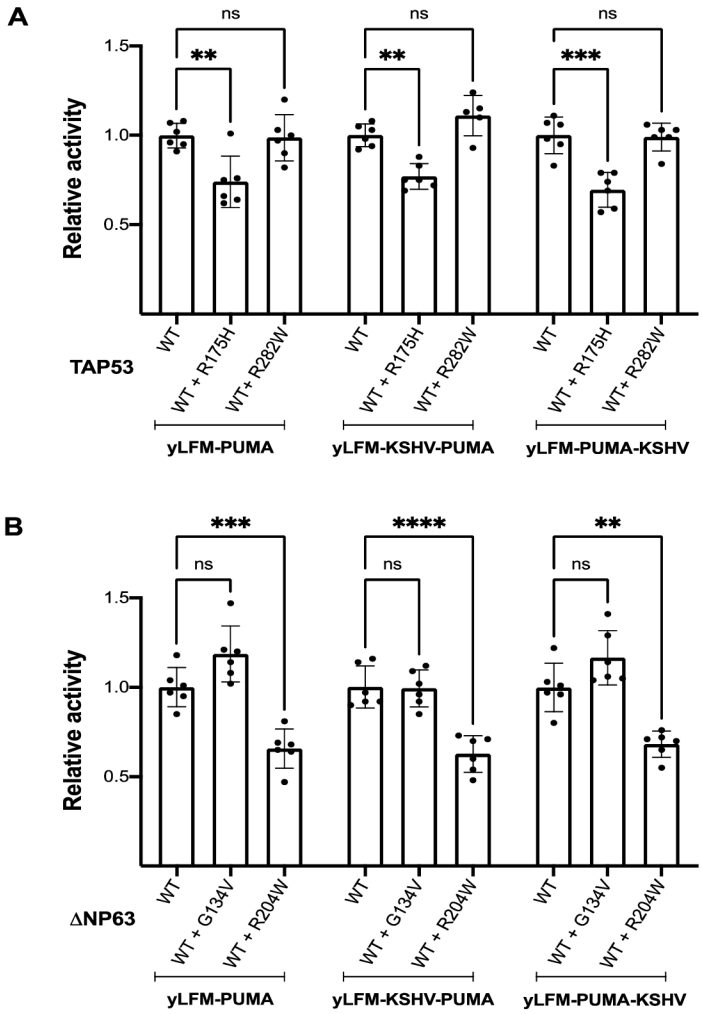


Figure 4. Relative activity of co-expressed wild-type and mutant TAP53 or ΔNP63 proteins compared to the corresponding single protein expression. (**A**) Comparison of wild-type and mutant TAP53 co-expression activity (i.e., constitutive pADH1-wild-type TAP53 + inducible pGAL1,10 TAP53 R175H or R282W) with that of wild-type TAP53 expressed alone (i.e., constitutive pADH1-wild-type TAP53) in presence of the P53 RE from the *PUMA* target gene (yLFM-PUMA) or the same p53 RE along with a G4 prone sequence upstream (yLFM-KSHV-PUMA) or downstream (yLFM-PUMA-KSHV). (**B**) Comparison of wild-type and mutant ΔNP63 co-expression activity (i.e., constitutive pADH1-wild-type ΔNP63 + inducible pGAL1,10 ΔNP63 G134V or R204W) with that of wild-type ΔNP63 expressed alone (i.e., constitutive pADH1 wild-type ΔNP63) in presence of the P53 RE from the *PUMA* target gene (yLFM-PUMA) or the same P53 RE along with a G4 prone sequence upstream (yLFM-KSHV-PUMA) or downstream (yLFM-PUMA-KSHV). WT, wild-type. ns: not significant, ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

4. Discussion

P53 family comprises a group of potent tetrameric TFs (i.e., P53, P63, and P73) sharing a conserved DNA binding domain responsible for the binding to sequence-specific DNA REs located near promoters of the target genes. Nevertheless, the transcriptional networks

regulated by these proteins are largely distinct [47]; this feature is associated with the plasticity provided by the promoters (P1 and P2) and the splicing regulation of P53 family genes (N- and C-terminal isoforms), by the modulation of cofactors and companion TFs, and by the impact of quaternary structure of P53 family proteins at promoter target sites [4,6,48].

Many studies have highlighted that RE recognition by P53 is regulated by the direct contacts with target DNA sequence; however, the so-called indirect readout provided by nucleotides within the RE that are not directly contacted by the P53 protein is also important [49,50]. This indirect readout, also referred to as "shape readout" since it appears to be related to structural properties at the P53 DNA target sites, has been recently extended to sequences surrounding the P53 RE [51].

Previously, we adapted a well-established yeast-based functional assay to evaluate the influence of a G4 DNA prone sequence in proximity of a P53 RE on the transactivation potential of full-length and N-terminal truncated P53 α isoforms [34]. Structurally, G4s are four-stranded nucleic acids structures held together by non-canonical Hoogsteen G-G base pairs [52,53]. G4 motifs are evolutionarily conserved from bacteria to human, confirming the importance of their formation in vivo [54]. In yeast and human, G4s are enriched in telomeric and ribosomal DNA, at transcriptional regulatory units, and at mitotic and meiotic double-strand break sites [53,55,56]. G4s can also form in RNA, where they can affect its stability or translation [57–60].

Indeed, contemporary research has demonstrated the importance of G4s in various cellular processes, including interactions with TFs [60–63]. Moreover, the targeting of the unique structure of G4s is proposed as a good tool for various diseases therapies including cancer [64–68].

Interestingly, G4s are often located in promoter sequences as shown for many human genes [69–71], including P53 targets; in fact, the presence of G4 prone sequences has been found around P53 RE in the promoter of the P53-induced apoptotic PUMA protein [34]. Therefore, the presence and the formation of a G4 structure could be an important transcriptional regulatory element, as previously observed in various organisms including humans [72–74].

In the present paper we extended for the first time the evaluation of the role of G4s in the transcription regulation to all P53 family proteins (wild-type and mutant α isoforms) by using the yeast *Saccharomyces cerevisiae* as model system; the impact of a sequence element prone to adopt a G4 secondary structure located upstream or downstream of the moderately active P53 RE from apoptotic *PUMA* target was evaluated. Moreover, the presence of G4 prone sequence is one of the possible reasons for binding of P53 to sites without a P53 RE as detected by ChIP studies [75]. However, we observed no significant activation of transcription by all analyzed proteins in presence of the G4 prone sequence alone, in agreement with the previous analysis regarding P53 protein [34]. On the contrary, we observed that the presence of the G4 prone sequence upstream, and more significantly downstream, of the RE causes inhibition of both ΔN and TA variants P53 family protein-mediated transactivation. Interestingly, despite lacking a full-length transactivation domain, ΔN isoforms are not transcriptionally inactive and have been shown to regulate the expression of their own set of genes [76–85]. The difference in relative transactivation between TA and ΔN isoforms was, however, significantly reduced by the presence of the G4 prone sequence downstream of RE, which suggests that inhibition of transcription activity by G4 in this sequence context could be variant-dependent (i.e., stronger on TA than on ΔN).

Then, we decided to extend the study to P53 family protein mutants. To this aim, P53 and P63 mutations associated with specific genetic syndromes and functionally heterogeneous in terms of transactivation ability were selected; moreover, the chosen P53 mutations are also hot spot codons in human cancers. The presence of the G4 prone sequence (especially downstream of the P53 RE) tends to lower the difference between wild-type and mutant protein transactivation activity. Conversely, the simulation of heterozygous con-

dition of P53- and P63-associated disorders by co-expressing wild-type and mutant P53 or P63 variants failed to reveal an impact of the G4 prone sequence on the dominant negative properties of the P53 R175H and P63 R204W mutations. However, it is to be taken into consideration that, P53 family proteins acting as tetramers, there is a variation in the modality by which tetramers are constituted and in the possible stoichiometry of hetero-tetramers comprising the combination of wild-type and missense mutations; thus, the adaptability of our yeast assay will allow further insights in this direction.

5. Conclusions

Overall, our data suggest that G4 prone sequences proximal to a P53 RE lead to an overall reduction of P53 family proteins-dependent transactivation. Still, at the same time, their presence can modulate the interplay between isoforms within the same protein and between different members of P53 family. Moreover, some functional features of mutant P53 family proteins can also be affected by the placement of a G4 prone sequence adjacent to a P53 RE. Hence, we propose that the sequence context surrounding a P53 RE can contribute to tuning P53 family proteins functions and should be considered as an important variable to fully characterize the P53 family cistrome. Furthermore, given that the net transcriptional effect of the P53 family proteins can be dependent on the ratio TA/ΔN variants of P53, P63, and P73 isoforms and on wild-type and mutant P53/P63/P73 interactions and binding to the promoters of target genes, our finding also paves the way for future studies involving the use of available small molecules that can modulate the conformations of G4s structure-forming DNA sequences [86].

Lastly, the yeast *Saccharomyces cerevisiae* has reconfirmed itself as a robust model system that can be turned into a sort of in vivo test-tube for the functional analysis of human TFs. Although the placement of a G4 prone sequence adjacent to a P53 RE within a minimal promoter can impact the site's global transactivation potential, our approach in yeast mainly focuses on the relative effect towards specific P53 family proteins or on their functional interactions. The otherwise isogenic nature of the reporter yeast strains and the regulated systems for ectopic protein expression we used give us confidence that the observed relative differences are not dependent on other variables.

Supplementary Materials: The following are available online at https://www.mdpi.com/2073-4425/12/2/277/s1, Figure S1: Relative activity of mutant P53 and P63 proteins, Figure S2: Fold change transactivation by co-expressed wild-type and mutant P53 proteins, Figure S3: Fold change transactivation by co-expressed wild-type and mutant P63 proteins.

Author Contributions: Conceptualization, P.M. (Paola Monti), V.B., A.I., G.F.; Writing—original draft preparation, P.M. (Paola Monti), V.B., A.I., G.F.; Methodology, P.M. (Paola Monti), V.B., N.B., O.P., P.M. (Paola Menichini), A.S., R.B., A.I., G.F.; Validation, P.M. (Paola Monti), V.B., N.B., O.P., P.M. (Paola Menichini), A.S., R.B., A.I., G.F.; Data curation, P.M. (Paola Monti), V.B., N.B., O.P., P.M. (Paola Menichini), A.S., R.B., A.I., G.F.; Writing—review and editing, P.M. (Paola Monti), V.B., N.B., O.P., P.M. (Paola Menichini), A.S., R.B., A.I., G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by AIRC I.G. n. 5506 (to G.F.), Italian Ministry of Health 5×1000 funds 2013, 2015, and 2016 (to G.F.), Current Research 2016 (to G.F.), Compagnia S. Paolo Turin Italy project 2017.0526 (to G.F.); by AIRC I.G. n.18985 (to A.I.); by the Czech Science Foundation n. 18-15548S (to V.B.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lane, D.P.; Crawford, L.V. T Antigen is bound to a host protein in SV40-transformed cells. *Nature* **1979**, *278*, 261–263. [CrossRef]
2. Kaghad, M.; Bonnet, H.; Yang, A.; Creancier, L.; Biscan, J.-C.; Valent, A.; Minty, A.; Chalon, P.; Lelias, J.-M.; Dumont, X.; et al. Monoallelically expressed gene related to P53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **1997**, *90*, 809–819. [CrossRef]
3. Yang, A.; Kaghad, M.; Wang, Y.; Gillett, E.; Fleming, M.D.; Dötsch, V.; Andrews, N.C.; Caput, D.; McKeon, F. P63, a P53 homolog at 3q27–29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol. Cell* **1998**, *2*, 305–316. [CrossRef]
4. Collavin, L.; Lunardi, A.; Del Sal, G. P53-family proteins and their regulators: Hubs and spokes in tumor suppression. *Cell Death Differ.* **2010**, *17*, 901–911. [CrossRef]
5. Wei, J.; Zaika, E.; Zaika, A. P53 family: Role of protein isoforms in human cancer. *J. Nucleic Acids* **2012**, *2012*, 687359. [CrossRef] [PubMed]
6. Bourdon, J.-C. P53 family isoforms. *CPB* **2007**, *8*, 332–336. [CrossRef]
7. Malkin, D. Li-Fraumeni syndrome. *Genes Cancer* **2011**, *2*, 475–484. [CrossRef]
8. Mills, A.A.; Zheng, B.; Wang, X.-J.; Vogel, H.; Roop, D.R.; Bradley, A. P63 is a P53 homologue required for limb and epidermal morphogenesis. *Nature* **1999**, *398*, 708–713. [CrossRef]
9. Rinne, T.; Brunner, H.G.; van Bokhoven, H. P63-associated disorders. *Cell Cycle* **2007**, *6*, 262–268. [CrossRef]
10. Rufini, A.; Agostini, M.; Grespi, F.; Tomasini, R.; Sayan, B.S.; Niklison-Chirou, M.V.; Conforti, F.; Velletri, T.; Mastino, A.; Mak, T.W.; et al. P73 in cancer. *Genes Cancer* **2011**, *2*, 491–502. [CrossRef] [PubMed]
11. Nemajerova, A.; Moll, U.M. Tissue-specific roles of P73 in development and homeostasis. *J. Cell Sci.* **2019**, *132*, jcs233338. [CrossRef] [PubMed]
12. Menendez, D.; Inga, A.; Resnick, M.A. The expanding universe of P53 targets. *Nat. Rev. Cancer* **2009**, *9*, 724–737. [CrossRef] [PubMed]
13. Riley, T.; Sontag, E.; Chen, P.; Levine, A. Transcriptional control of human P53-regulated genes. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 402–412. [CrossRef] [PubMed]
14. El-Deiry, W.S.; Kern, S.E.; Pietenpol, J.A.; Kinzler, K.W.; Vogelstein, B. Definition of a consensus binding site for P53. *Nat. Genet.* **1992**, *1*, 45–49. [CrossRef] [PubMed]
15. Brandt, T.; Petrovich, M.; Joerger, A.C.; Veprintsev, D.B. Conservation of DNA-binding specificity and oligomerisation properties within the P53 family. *BMC Genom.* **2009**, *10*, 628. [CrossRef]
16. Perez, C.A.; Ott, J.; Mays, D.J.; Pietenpol, J.A. P63 Consensus DNA-binding site: Identification, analysis and application into a P63MH algorithm. *Oncogene* **2007**, *26*, 7363–7370. [CrossRef] [PubMed]
17. Osada, M.; Park, H.L.; Nagakawa, Y.; Yamashita, K.; Fomenkov, A.; Kim, M.S.; Wu, G.; Nomoto, S.; Trink, B.; Sidransky, D. Differential recognition of response elements determines target gene specificity Forp53 and P63. *MCB* **2005**, *25*, 6077–6089. [CrossRef]
18. Moll, U.M.; Slade, N. P63 and P73: Roles in development and tumor formation. *Mol. Cancer Res.* **2004**, *2*, 371–386.
19. Zheng, X.; Chen, X. Aquaporin 3, a glycerol and water transporter, is regulated by P73 of the P53 family. *FEBS Lett.* **2001**, *489*, 4–7. [CrossRef]
20. Nakagawa, T.; Takahashi, M.; Ozaki, T.; Watanabe, K.; Todo, S.; Mizuguchi, H.; Hayakawa, T.; Nakagawara, A. Autoinhibitory regulation of P73 by ΔNp73 to modulate cell survival and death through a P73-specific target element within the ΔNp73 promoter. *MCB* **2002**, *22*, 2575–2585. [CrossRef]
21. Sasaki, Y.; Ishida, S.; Morimoto, I.; Yamashita, T.; Kojima, T.; Kihara, C.; Tanaka, T.; Imai, K.; Nakamura, Y.; Tokino, T. The P53 family member genes are involved in the notch signal pathway. *J. Biol. Chem.* **2002**, *277*, 719–724. [CrossRef]
22. Harms, K.; Nozell, S.; Chen, X. The common and distinct target genes of the P53 family transcription factors. *Cell. Mol. Life Sci.* **2004**, *61*, 822–842. [CrossRef]
23. Brázda, V.; Coufal, J. Recognition of local DNA structures by P53 protein. *Int. J. Mol. Sci.* **2017**, *18*, 375. [CrossRef]
24. Jagelská, E.B.; Brázda, V.; Pečinka, P.; Paleček, E.; Fojta, M. DNA Topology influences P53 sequence-specific DNA binding through structural transitions within the target sites. *Biochem. J.* **2008**, *412*, 57–63. [CrossRef]
25. Coufal, J.; Jagelská, E.B.; Liao, J.C.C.; Brázda, V. Preferential binding of P53 tumor suppressor to P21 promoter sites that contain inverted repeats capable of forming cruciform structure. *Biochem. Biophys. Res. Commun.* **2013**, *441*, 83–88. [CrossRef] [PubMed]
26. Petr, M.; Helma, R.; Polášková, A.; Krejčí, A.; Dvořáková, Z.; Kejnovská, I.; Navrátilová, L.; Adámik, M.; Vorlíčková, M.; Brázdová, M. Wild-type P53 binds to MYC promoter G-quadruplex. *Biosci. Rep.* **2016**, *36*, e00397. [CrossRef]
27. Brázda, V.; Fojta, M. The rich world of P53 DNA binding targets: The role of DNA structure. *Int. J. Mol. Sci.* **2019**, *20*, 5605. [CrossRef] [PubMed]
28. Quante, T.; Otto, B.; Brázdová, M.; Kejnovská, I.; Deppert, W.; Tolstonog, G.V. Mutant P53 is a transcriptional co-factor that binds to g-rich regulatory regions of active genes and generates transcriptional plasticity. *Cell Cycle* **2012**, *11*, 3290–3303. [CrossRef] [PubMed]
29. Kennedy, B.K. Mammalian transcription factors in yeast: Strangers in a familiar land. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 41–49. [CrossRef] [PubMed]

30. Schärer, E.; lggo, R. Mammalian P53 can function as a transcription factor in yeast. *Nucleic Acids Res.* **1992**, *20*, 1539–1545. [CrossRef]
31. Sharma, V.; Monti, P.; Fronza, G.; Inga, A. Human transcription factors in yeast: The fruitful examples of P53 and NF-KB. *FEMS Yeast Res.* **2016**, *16*, fow083. [CrossRef] [PubMed]
32. Storici, F.; Lewis, L.K.; Resnick, M.A. In vivo site-directed mutagenesis using oligonucleotides. *Nat. Biotechnol.* **2001**, *19*, 773–776. [CrossRef]
33. Monti, P.; Bosco, B.; Gomes, S.; Saraiva, L.; Fronza, G.; Inga, A. Yeast as a chassis for developing functional assays to study human P53. *J. Vis. Exp.* **2019**, *150*. [CrossRef]
34. Porubiaková, O.; Bohálová, N.; Inga, A.; Vadovičová, N.; Coufal, J.; Fojta, M.; Brázda, V. The influence of quadruplex structure in proximity to P53 target sequences on the transactivation potential of P53 alpha isoforms. *Int. J. Mol. Sci.* **2019**, *21*, 127. [CrossRef]
35. Storici, F.; Resnick, M.A. Delitto perfetto targeted mutagenesis in yeast with oligonucleotides. *Genet. Eng. N. Y.* **2003**, *25*, 189–207.
36. Resnick, M.A.; Inga, A. Functional mutants of the sequence-specific transcription factor P53 and implications for master genes of diversity. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9934–9939. [CrossRef] [PubMed]
37. Inga, A.; Storici, F.; Darden, T.A.; Resnick, M.A. Differential transactivation by the P53 transcription factor is highly dependent on P53 level and promoter target sequence. *MCB* **2002**, *22*, 8612–8625. [CrossRef]
38. Andreotti, V.; Ciribilli, Y.; Monti, P.; Bisio, A.; Lion, M.; Jordan, J.; Fronza, G.; Menichini, P.; Resnick, M.A.; Inga, A. P53 Transactivation and the impact of mutations, cofactors and small molecules using a simplified yeast-based screening system. *PLoS ONE* **2011**, *6*, e20643. [CrossRef]
39. Monti, P.; Russo, D.; Bocciardi, R.; Foggetti, G.; Menichini, P.; Divizia, M.T.; Lerone, M.; Graziano, C.; Wischmeijer, A.; Viadiu, H.; et al. EEC- and ADULT-associated TP63 mutations exhibit functional heterogeneity toward P63 responsive sequences. *Hum. Mutat.* **2013**, *34*, 894–904. [CrossRef]
40. Ciribilli, Y.; Monti, P.; Bisio, A.; Nguyen, H.T.; Ethayathulla, A.S.; Ramos, A.; Foggetti, G.; Menichini, P.; Menendez, D.; Resnick, M.A.; et al. Transactivation specificity is conserved among P53 family proteins and depends on a response element sequence code. *Nucleic Acids Res.* **2013**, *41*, 8637–8653. [CrossRef]
41. Monti, P.; Ciribilli, Y.; Bisio, A.; Foggetti, G.; Raimondi, I.; Campomenosi, P.; Menichini, P.; Fronza, G.; Inga, A. ΔN-P63α and TA-P63α exhibit intrinsic differences in transactivation specificities that depend on distinct features of DNA target sites. *Oncotarget* **2014**, *5*, 2116–2130. [CrossRef] [PubMed]
42. Monti, P.; Lionetti, M.; De Luca, G.; Menichini, P.; Recchia, A.G.; Matis, S.; Colombo, M.; Fabris, S.; Speciale, A.; Barbieri, M.; et al. Time to first treatment and P53 dysfunction in chronic lymphocytic leukaemia: Results of the O-CLL1 study in early stage patients. *Sci. Rep.* **2020**, *10*, 18427. [CrossRef] [PubMed]
43. Monti, P.; Perfumo, C.; Bisio, A.; Ciribilli, Y.; Menichini, P.; Russo, D.; Umbach, D.M.; Resnick, M.A.; Inga, A.; Fronza, G. Dominant-negative features of mutant TP53 in germline carriers have limited impact on cancer outcomes. *Mol. Cancer Res.* **2011**, *9*, 271–279. [CrossRef]
44. Zhang, Y.; Coillie, S.V.; Fang, J.-Y.; Xu, J. Gain of function of mutant P53: R282W on the peak? *Oncogenesis* **2016**, *5*, e196. [CrossRef]
45. Ko, L.J.; Prives, C. P53: Puzzle and paradigm. *Genes Dev.* **1996**, *10*, 1054–1072. [CrossRef] [PubMed]
46. Monti, P.; Campomenosi, P.; Ciribilli, Y.; Iannone, R.; Inga, A.; Abbondandolo, A.; Resnick, M.A.; Fronza, G. Tumour P53 mutations exhibit promoter selective dominance over wild type P53. *Oncogene* **2002**, *21*, 1641–1648. [CrossRef]
47. Dötsch, V.; Bernassola, F.; Coutandin, D.; Candi, E.; Melino, G. P63 and P73, the ancestors of P53. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a004887. [CrossRef]
48. Ferraiuolo, M.; Di Agostino, S.; Blandino, G.; Strano, S. Oncogenic intra-P53 family member interactions in human cancers. *Front. Oncol.* **2016**, *6*, 77. [CrossRef] [PubMed]
49. Kitayner, M.; Rozenberg, H.; Rohs, R.; Suad, O.; Rabinovich, D.; Honig, B.; Shakked, Z. Diversity in DNA recognition by P53 revealed by crystal structures with hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **2010**, *17*, 423–429. [CrossRef]
50. Beno, I.; Rosenthal, K.; Levitine, M.; Shaulov, L.; Haran, T.E. Sequence-dependent cooperative binding of P53 to DNA targets and its relationship to the structural properties of the DNA targets. *Nucleic Acids Res.* **2011**, *39*, 1919–1932. [CrossRef]
51. Senitzki, A.; Safieh, J.; Sharma, V.; Golovenko, D.; Danin-Poleg, Y.; Inga, A.; Haran, T.E. The complex architecture of P53 binding sites. *Nucleic Acids Res.* **2021**. [CrossRef]
52. Lipps, H.J.; Rhodes, D. G-quadruplex structures: In vivo evidence and function. *Trends Cell Biol.* **2009**, *19*, 414–422. [CrossRef]
53. Marsico, G.; Chambers, V.S.; Sahakyan, A.B.; McCauley, P.; Boutell, J.M.; Antonio, M.D.; Balasubramanian, S. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **2019**, *47*, 3862–3874. [CrossRef] [PubMed]
54. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The structure and function of DNA G-quadruplexes. *Trends Chem.* **2020**, *2*, 123–136. [CrossRef] [PubMed]
55. Huppert, J.L. Structure, location and interactions of G-quadruplexes. *FEBS J.* **2010**, *277*, 3452–3458. [CrossRef] [PubMed]
56. Capra, J.A.; Paeschke, K.; Singh, M.; Zakian, V.A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput. Biol.* **2010**, *6*, e1000861. [CrossRef]
57. Huppert, J.L. Hunting G-quadruplexes. *Biochimie* **2008**, *90*, 1140–1148. [CrossRef]
58. Morris, M.J.; Basu, S. An unusually stable G-quadruplex within the 5′-UTR of the MT3 matrix metalloproteinase MRNA represses translation in eukaryotic cells. *Biochemistry* **2009**, *48*, 5313–5319. [CrossRef]

59. Agarwala, P.; Pandey, S.; Mapa, K.; Maiti, S. The G-quadruplex augments translation in the 5′ untranslated region of transforming growth factor B2. *Biochemistry* **2013**, *52*, 1528–1538. [CrossRef]
60. Dumas, L.; Herviou, P.; Dassi, E.; Cammas, A.; Millevoi, S. G-quadruplexes in RNA biology: Recent advances and future directions. *Trends Biochem. Sci.* **2020**. [CrossRef]
61. Rhodes, D.; Lipps, H.J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015**, *43*, 8627–8637. [CrossRef]
62. Hároníková, L.; Coufal, J.; Kejnovská, I.; Jagelská, E.B.; Fojta, M.; Dvořáková, P.; Muller, P.; Vojtesek, B.; Brázda, V. IFI16 preferentially binds to DNA with quadruplex structure and enhances DNA quadruplex formation. *PLoS ONE* **2016**, *11*, e0157156. [CrossRef]
63. Dhamodharan, V.; Pradeepkumar, P.I. Specific recognition of promoter G-quadruplex DNAs by small molecule ligands and light-up probes. *ACS Chem. Biol.* **2019**, *14*, 2102–2114. [CrossRef] [PubMed]
64. Kharel, P.; Balaratnam, S.; Beals, N.; Basu, S. The role of RNA G-quadruplexes in human diseases and therapeutic strategies. *WIREs RNA* **2020**, *11*, e1568. [CrossRef]
65. Chaudhuri, R.; Bhattacharya, S.; Dash, J.; Bhattacharya, S. Recent update on targeting *c-MYC* G-quadruplexes by small molecules for anticancer therapeutics. *J. Med. Chem.* **2020**, *64*, 42–70. [CrossRef] [PubMed]
66. Sanchez-Martin, V.; Lopez-Pujante, C.; Soriano-Rodriguez, M.; Garcia-Salcedo, J.A. An updated focus on quadruplex structures as potential therapeutic targets in cancer. *Int. J. Mol. Sci.* **2020**, *21*, 8900. [CrossRef]
67. Asamitsu, S.; Yabuki, Y.; Ikenoshita, S.; Wada, T.; Shioda, N. Pharmacological prospects of G-quadruplexes for neurological diseases using porphyrins. *Biochem. Biophys. Res. Commun.* **2020**, *531*, 51–55. [CrossRef] [PubMed]
68. Kawauchi, K.; Urano, R.; Kinoshita, N.; Kuwamoto, S.; Torii, T.; Hashimoto, Y.; Taniguchi, S.; Tsuruta, M.; Miyoshi, D. Photosensitizers based on G-quadruplex ligand for cancer photodynamic therapy. *Genes* **2020**, *11*, 1340. [CrossRef]
69. Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [CrossRef]
70. Lago, S.; Nadai, M.; Ruggiero, E.; Tassinari, M.; Marušič, M.; Tosoni, B.; Frasson, I.; Cernilogar, F.M.; Pirota, V.; Doria, F.; et al. The MDM2 inducible promoter folds into four-tetrad antiparallel G-quadruplexes targetable to fight malignant liposarcoma. *Nucleic Acids Res.* **2021**, *49*, 847–863. [CrossRef]
71. Da Ros, S.; Nicoletto, G.; Rigo, R.; Ceschi, S.; Zorzan, E.; Dacasto, M.; Giantin, M.; Sissi, C. G-quadruplex modulation of SP1 functional binding sites at the KIT proximal promoter. *Int. J. Mol. Sci.* **2020**, *22*, 329. [CrossRef]
72. Brázda, V.; Hároníková, L.; Liao, J.; Fojta, M. DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.* **2014**, *15*, 17493–17517. [CrossRef] [PubMed]
73. Gazanion, E.; Lacroix, L.; Alberti, P.; Gurung, P.; Wein, S.; Cheng, M.; Mergny, J.-L.; Gomes, A.R.; Lopez-Rubio, J.-J. Genome wide distribution of G-quadruplexes and their impact on gene expression in malaria parasites. *PLoS Genet.* **2020**, *16*, e1008917. [CrossRef]
74. Chashchina, G.V.; Beniaminov, A.D.; Kaluzhny, D.N. Stable G-quadruplex structures of oncogene promoters induce potassium-dependent stops of thermostable DNA polymerase. *Biochem. Mosc.* **2019**, *84*, 562–569. [CrossRef] [PubMed]
75. Nguyen, T.-A.T.; Grimm, S.A.; Bushel, P.R.; Li, J.; Li, Y.; Bennett, B.D.; Lavender, C.A.; Ward, J.M.; Fargo, D.C.; Anderson, C.W.; et al. Revealing a human P53 universe. *Nucleic Acids Res.* **2018**, *46*, 8153–8167. [CrossRef] [PubMed]
76. Kartasheva, N.N.; Lenz-Bauer, C.; Hartmann, O.; Schäfer, H.; Eilers, M.; Dobbelstein, M. ΔNp73 can modulate the expression of various genes in a P53-independent fashion. *Oncogene* **2003**, *22*, 8246–8254. [CrossRef] [PubMed]
77. King, K.E.; Ponnamperuma, R.M.; Yamashita, T.; Tokino, T.; Lee, L.A.; Young, M.F.; Weinberg, W.C. ΔNp63α functions as both a positive and a negative transcriptional regulator and blocks in vitro differentiation of murine keratinocytes. *Oncogene* **2003**, *22*, 3635–3644. [CrossRef] [PubMed]
78. Kazantseva, M.; Mehta, S.; Eiholzer, R.A.; Hung, N.; Wiles, A.; Slatter, T.L.; Braithwaite, A.W. A Mouse model of the Δ133p53 isoform: Roles in cancer progression and inflammation. *Mamm. Genome* **2018**, *29*, 831–842. [CrossRef]
79. Ghioni, P.; Bolognese, F.; Duijf, P.H.G.; van Bokhoven, H.; Mantovani, R.; Guerrini, L. Complex transcriptional effects of P63 isoforms: Identification of novel activation and repression domains. *MCB* **2002**, *22*, 8659–8668. [CrossRef]
80. Barton, C.E.; Johnson, K.N.; Mays, D.M.; Boehnke, K.; Shyr, Y.; Boukamp, P.; Pietenpol, J.A. Novel P63 target genes involved in paracrine signaling and keratinocyte differentiation. *Cell Death Dis.* **2010**, *1*, e74. [CrossRef]
81. Dohn, M.; Zhang, S.; Chen, X. P63α and ΔNp63α can induce cell cycle arrest and apoptosis and differentially regulate P53 target genes. *Oncogene* **2001**, *20*, 3193–3205. [CrossRef] [PubMed]
82. Romano, R.-A.; Ortt, K.; Birkaya, B.; Smalley, K.; Sinha, S. An active role of the ΔN isoform of P63 in regulating basal keratin genes K5 and K14 and directing epidermal cell fate. *PLoS ONE* **2009**, *4*, e5623. [CrossRef]
83. Wu, G.; Osada, M.; Guo, Z.; Fomenkov, A.; Begum, S.; Zhao, M.; Upadhyay, S.; Xing, M.; Wu, F.; Moon, C.; et al. DeltaNp63alpha up-regulates the Hsp70 gene in human cancer. *Cancer Res.* **2005**, *65*, 758–766.
84. Higashikawa, K.; Yoneda, S.; Tobiume, K.; Saitoh, M.; Taki, M.; Mitani, Y.; Shigeishi, H.; Ono, S.; Kamata, N. ΔNp63α-dependent expression of Id-3 distinctively suppresses the invasiveness of human squamous cell carcinoma. *Int. J. Cancer* **2009**, *124*, 2837–2844. [CrossRef] [PubMed]
85. Kommagani, R.; Leonard, M.K.; Lewis, S.; Romano, R.-A.; Sinha, S.; Kadakia, M.P. Regulation of VDR by Np63 is associated with inhibition of cell invasion. *J. Cell Sci.* **2009**, *122*, 2828–2835. [CrossRef] [PubMed]
86. del Mundo, I.M.A.; Vasquez, K.M.; Wang, G. Modulation of DNA structure formation using small molecules. *Biochim. Biophys. Acta Mol. Cell Res.* **2019**, *1866*, 118539. [CrossRef] [PubMed]

Příloha C

Dual mode of IFI16 binding to supercoiled and linear DNA: a closer insight

Biochemical and Biophysical Research Communications 667 (2023) 89–94

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Dual mode of IFI16 binding to supercoiled and linear DNA: A closer insight

Natália Valková [a, b], Libuše Kratochvilová [c], Lucia Martinková [d], Václav Brázda [a, *]

[a] Institute of Biophysics, Academy of Sciences of the Czech Republic, 612 65, Brno, Czech Republic
[b] Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 625 00, Brno, Czech Republic
[c] Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 612 00, Brno, Czech Republic
[d] RECAMO, Masaryk Memorial Cancer Institute, Zluty kopec 7, 656 53, Brno, Czech Republic

ARTICLE INFO

Article history:
Received 9 May 2023
Accepted 14 May 2023
Available online 15 May 2023

Keywords:
IFI16
Superhelicity
DNA
AFM
G-quadruplex
Inverted repeat

ABSTRACT

IFI16 (Interferon inducible protein 16) is a DNA sensor responsible for innate immune response stimulation and a direct viral restriction by modulating gene expression and replication. Many IFI16-DNA binding properties were described - length-dependent and sequence-independent binding, oligomerization of IFI16 upon recognition, sliding on the DNA, and preference for supercoiled DNA. However, the question of the role of IFI16-DNA binding in distinct IFI16 functions remains unclear. Here we demonstrate two modes of IFI16 binding to DNA using atomic force microscopy and electrophoretic mobility shift assays. In our study, we show that IFI16 can bind to DNA in the form of globular complexes or oligomers depending on DNA topology and molar ratios. The stability of the complexes is different in higher salt concentrations. In addition, we observed no preferential binding with the HIN-A or HIN-B domains to supercoiled DNA, revealing the importance of the whole protein for this specificity. These results provide more profound insight into IFI16-DNA interactions and may be important in answering the question of self- and non-self-DNA binding by the IFI16 protein and potentially could shed light on the role of DNA binding in distinct IFI16 functions.

© 2023 Elsevier Inc. All rights reserved.

1. Introduction

Interferon gamma-inducible protein 16 (IFI16) is a hematopoietic antigen that contains two 200 amino acid motifs referred to as HIN domains [1]. This, along with human AIM2 (absent in melanoma 2), IFIX (interferon-inducible protein X) and MNDA (myeloid cell nuclear differentiation antigen), places the protein into a family of proteins called HIN-200 [2]. Oligosaccharide binding folds in the two HIN domains (HIN-A and HIN-B) are responsible for protein binding to dsDNA by electrostatic interactions of negative DNA charge and positive residues of HIN domains (3), while the N-terminal IFI16 protein domain - PYRIN domain - is mainly responsible for protein-protein interactions [3]. The HIN-A domain binds to ssDNA and dsDNA *in vitro* with similar affinity. The HIN-B show a higher affinity for GC-rich DNA [4]. The spacer region of the IFI16 protein is rich in serine, threonine and proline. Due to the alternative splicing of this region, different isoforms of the IFI16 protein arise (IFI16 A, B, C) arise [5]. Its subcellular localization was first considered as solely nuclear, by it is known that could be alternate by cell type [6], posttranslational modification [7], the presence of infectious agents [8] and exposure to UV radiation [9].

IFI16 cooperatively assembles on dsDNA [10] but preference for GC rich ssDNA over dsDNA has also been described [11]. This result is consistent with our previous study where we observed a higher affinity of IFI16 for superhelical DNA and local secondary DNA structures, particularly G-quadruplexes (G4) and inverted repeats compared to dsDNA [12,13]. Furthermore, IFI16 was located in the NHEIII region of the *c-myc* promotor containing the putative G4 structure sequence confirmed *in vitro* [14,15]. G4s are secondary non-canonic local structures of DNA or RNA, which are rich in G bases. G4s are easily formed under physiological conditions. G4s are commonly present in important regulatory regions, such as telomers, promoters of oncogenes, origins of replication, and play a significant role in regulation processes such as replication and gene expression [16]. According to recent findings, stabilization of G4 negative supercoiling could have a potential impact on viral regulation of replication, transcription, or latent stage of infection [17].

The interferon-inducible protein 16 (IFI16) has multiple

* Corresponding author.
E-mail address: vaclav@ibp.cz (V. Brázda).

https://doi.org/10.1016/j.bbrc.2023.05.049
0006-291X/© 2023 Elsevier Inc. All rights reserved.

biological functions. In a cell, the IFI16 protein recognizes foreign or damaged DNA, suppresses the expression of cellular and viral genes [18,19], significantly participates in cell proliferation [20], differentiation [21] and inflammation activation [22]. The ability of the IFI16 protein to distinguish between intrinsic and foreign DNA in a cell plays a key role in innate immunity and is maintained by minimal DNA length to trigger immune response around 70 bp [23], which is longer than the free DNA between nucleosomes. Inflammasome activation, which occurs when foreign or damaged self-DNA is detected, triggers pro-inflammatory processes that can induce pyroptosis in certain cell types [24]. Upon recognition of foreign DNA IFI16 protein activates signaling pathways that trigger multiple non-specific immune response mechanisms. After recognition of foreign DNA, signaling is propagated by interferon gene stimulator (STING), serine/threonine protein TANK-binding kinase 1 (TBK-1) and interferon regulatory factor 3 (IRF3). The latter dimerizes and passes into the nucleus, where it involves transcriptional regulators and induces cytokine expression in the infected nucleus and subcellular environments [25]. However, IFI16 could restrict viral infection independently by stimulating the innate immune response. Bosso et al. described importance in the interaction of the PYRIN domain with the sp1 transcription factor in HIV restriction [26,27].

Elevated IFI16 levels and IFI16 autoantibodies have been detected in serum from patients suffering from various autoimmune diseases such as inflammatory bowel disease [28], systematic lupus erythematosus [29], psoriasis [30], rheumatoid arthritis [31], systemic scleroderma [32] or Sjögren syndrome [33]. The exact involvement of IFI16 in the molecular pathogenesis of these autoimmune diseases is unknown, but it could be used as a marker to predict disease progression or specific treatment. IFI16 has also been shown to be involved in the regulation of the latent stage of viral infection [34,35]. IFI16 recognizes infection of Herpes simplex virus 1 [19], Epstein-Barr virus [34], Human cytomegalovirus [36,37], Kaposi Sarcoma-associated herpesvirus [35], Human immunodeficiency virus [38], gram-positive bacteria Listeria monocytogenes and Francisella novicida [39]. The IFI16 protein also seems to be the protein sensor for RNA viruses [24].

Here, we demonstrate different mechanisms of binding of IFI16 to linear and supercoiled DNA. This dual mode of IFI16 binding could be explained by different cellular functions of IFI16. Together, these results suggest a new understanding of the binding properties of IFI16 with potential applications in virology and antiviral therapy.

2. Materials and methods

2.1. Plasmid DNA

Supercoiled plasmid DNA of pBluescript II SK (-), was isolated from bacterial strain DH5α as described in the QIAGEN protocol (QIAGEN GmbH, Germany). *EcoRI* restriction enzyme (New England Biolabs, UK) was used for linearization of plasmid.

2.2. Protein purification

The coding region for IFI16, HIN-A and HIN-B protein fragments were PCR amplified from human IFI16 cDNA, subcloned into the pET15b expression vector (Novagen), expressed in E.Coli BL-21-CodonPLus cells (Stratagene) with N-terminal His6-tag. Protein purification was performed by Co2+ affinity column (Talon).

2.3. Gel electrophoretic mobility shift assays on agarose gel

DNA (100 ng) and proteins were mixed at different molar ratios in 10 μl of DNA binding buffer (5 mM Tris-HCl, pH 7.0, 1 mM EDTA, 50 mM KCl and 0.01% Triton X-100). The samples were incubated for 15 min at 4 °C and loaded on 1% (IFI16) or 1.5% (HIN-A and HIN-B) agarose gel containing 0.33x Trisborate-EDTA buffer. Electrophoresis was performed for 3h at 100 V at low temperature (4 °C). The gels were stained with ethidium bromide and visualized on UV-transilluminator.

2.4. Atomic force microscopy

The protein-DNA complexes were prepared in 10 μl of DNA binding buffer (5 mM Tris-HCl, pH 7.0, 1 mM EDTA, 50 mM KCl and 0.01% Triton X-100) and were incubated for 15 min at 4 °C. AFM imaging was performed on Grade V4 mica discs (SPI supplies, USA). The samples were deposited on mica in buffer containing 5 mM Na-Hepes pH 7.5, 10 mM MgCl2 in the concentration of 1 ng/μl DNA and incubated for 5 min, followed by rinsing with deionized water and air-dried. The images were obtained using AFM/STM Multimode 8 electrochemical system, (Veeco, USA), operating in ScanAsyst mode in room temperature in air. The cantilever SCANASYST-AIR (Bruker) had a nominal spring constant of 0.4 N/m and the nominal scanning rate was set as 1Hz. Obtained images were then analyzed using Gwyddion software package [34].

2.5. Bioinformatic analysis

Sequence of pBluescript was analyzed by the web application Palindrome analyzer [40] for presence of inverted repeats. The size of IRs was set to 10–30 nt, size of spacers to 0–10 nt, and maximally one mismatch was allowed. Putative G-quadruplex forming structures were identified with G4Hunter [41] with default setting; 25 nt long sliding window and 1.2 for a G4Hunter threshold.

3. Results

The IFI16 preference for binding to superhelical DNA was described earlier [12]. However, we wanted to further investigate the observations from Hároníková et al., 2016 [13]. It was shown, that IFI16 binds cooperatively only to dsDNA oligonucleotides on polyacrylamide gels, not to local DNA secondary structures. When the formation of cruciform or G-quadruplex structures was induced, the protein was bound in a form of globular complexes with lower molecular weight. Whereas with dsDNA oligonucleotides, DNA:protein complexes 'hanging' in gel wells and not penetrating into the gel, in the case of oligonucleotides induced to form secondary structures, the protein:DNA complex migrated into the gel as expected. Since the phenomenon was observed in both G-quadruplex and cruciform structures without any preference, we took the superhelical plasmid pBluescript which has putative secondary structures (Fig. 1A). G4Hunter and PalindromeAnalyzer detected 3 putative G4 with a G4Hunter score greater than 1.2 and 4 inverted repeats with a repeat length of 10 and longer [40,41]. We used an elecromobility shift assay on agarose gels to confirm the observation with the full length IFI16 and HIN domains separately (Fig. 1B). In the case of the supercoiled structure of pBluescript, the full length IFI16 protein was found to bind already at a DNA:protein molar ratio of 1:2.5 and its linearized form did not show a loss of free DNA up to a DNA:protein ratio of 1:20 (Fig. 2A). Except molar ratios, one other important difference between interaction with superhelical and linear DNA was noted, linear DNA:protein complexes migrated as higher molecular weight complexes, proving the oligomerization of protein. Arrows on Fig. 2A point to the IFI16-DNA complexes. With superhelical DNA we can see the specific binding. The formed cooperative binding of the IFI16 protein to the DNA structure suggests its tendency to oligomerize only on linear

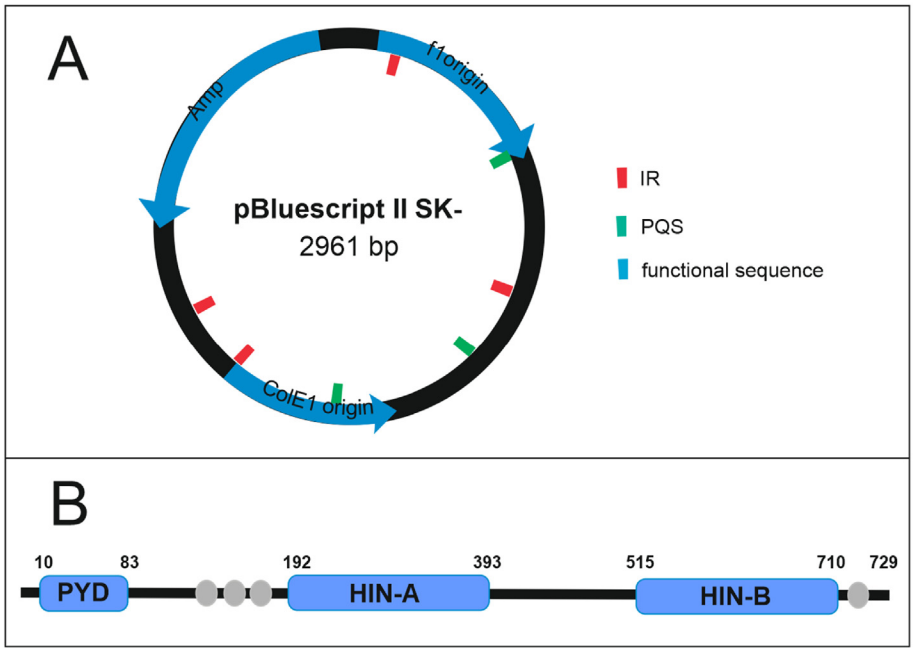


Fig. 1. (A) Location of sequences capable of forming non-B DNA structures in pBluescript as analyzed by the Palindrome analyzer and G4Hunter, Inverted repeats 10bp and longer in red (IR), Potential G-quadruplex-forming sequences with G4Hunter score greater than 1.2 in green (PQS) (B) Scheme of IFI16 protein and its domains. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

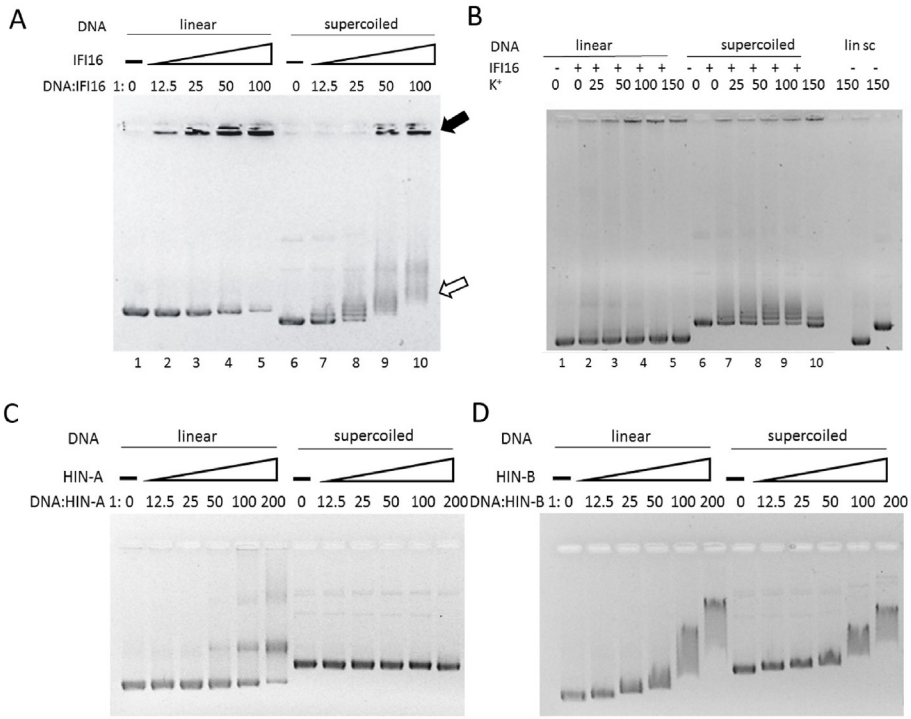


Fig. 2. Electromobility shift assay. (A) IFI16 binding to linear and supercoiled DNA. 100 ng of linear (lane 1-5) and supercoiled (lane 6-10) DNA was incubated with increasing protein concentration (molar ratio DNA to protein 0 (lane 0/6), 12.5 (lane 2/7), 25 (lane 3/8), 50 (lane 4/9), 100 (lane 5/10). Black arrows show the DNA:protein complexes. (B) IFI16 binding to linear and supercoiled DNA in different K+ concentrations. 100 ng of linear (lane 1-6) and supercoiled (7-12) DNA was incubated with identical protein concentration (molar ratio DNA: protein 1:25 lane 2-6 and 8-12) in different KCl concentration (0 mM in lane 1/2/7/8, 25 mM in lane 3/9, 50 mM in lane 4/10, 100 mM in lane 5/11 and 150 mM in lane 6/12) in binding buffer (C) HIN-A binding to linear and supercoiled DNA. 100 ng of linear (lane 1-6) and supercoiled (lane 7-12) DNA was incubated with increasing protein concentration (molar ratio DNA to protein was 0 (lane 0/7), 12.5 (lane 2/8), 25 (lane 3,9), 50 (lane 4/10), 100 (lane 5/11), 200 (lane 6/12) for 15 min on ice in binding buffer. Black arrows show the complexes. (D) HIN-B binding to linear and supercoiled DNA. The HIN-B in the same experiment design as in (C).

DNA molecules.

Furthermore, the effect of increasing concentration of potassium ions K+ on the binding of IFI16 protein to both forms of DNA – linear and supercoiled pBluescript was studied (Fig. 2B). As can be

seen in both sets of samples, the binding of the IFI16 protein to DNA is affected not only by the structure of the DNA molecule, but also by the concentration of the monovalent K+ ion in the samples. In addition, the effects of increasing concentrations of K+ ions differ for linear and supercoiled DNA. Free DNA intensities in protein-containing samples were determined against protein-free free DNA intensities without KCl. In the linear form of pBluescript, there was no significant change in IFI16 binding to DNA after 25 mM KCl addition compared to the 0 mM KCl sample. However, higher additions of K+ ions had an inhibitory effect on protein binding, and with increasing concentrations of K+ ions, there was an almost linear decrease in the concentration of the DNA:protein complex. This may be caused by non-covalent interactions between K+ ions and the sugar-phosphate backbone of the linear dsDNA. The addition of a higher concentration of monovalent K+ cation can compensate for the negatively charged charge of the sugar-phosphate backbone or directly weakly bind it to a small groove of DNA, which can block IFI16 protein binding sites. Thus, with increasing concentration of competitor (K+ ions), there would be a direct increase in free DNA and a decrease in DNA:IFI16 binding. On the contrary, in the case of samples containing supercoiled DNA pBluescript, the addition of K+ ions first leads to a reduction in the intensity of free DNA (Fig. 2B lanes 9 and 10) and thus to the promotion of protein binding to scDNA (Fig. 2C). The intensity of free DNA reaches a minimum value in the concentration range of 50–100 mM K+ and further decreases. Potassium ions at a certain concentration probably stabilize the secondary structures of the pBluescript scDNA and thus facilitate the binding of the IFI16 protein, while K+ ions of a certain value rather support the oligomerization of the protein itself, as can be seen in protein complexes formed in gel starters.

The binding of both “DNA binding conserved HIN domains” (HIN-A and HIN-B) to the linear and supercoiled DNA structure was also studied at higher molar ratios (Fig. 2C).Samples of both domains were prepared in the same manner. Linear and supercoiled DNA was incubated with increasing concentrations of IFI16 protein for 15 min on ice in 50 mM KCl binding buffer. When the HIN-A domain was added to the linear pBluescript samples, protein-DNA complexes formed and free DNA decreased proportionally to the concentration of the HIN-A domain. The presence of the DNA:HIN-A complex was recorded at a molar ratio of 1:25, but free DNA detected even at a molar ratio of 1:200 indicates only weak DNA binding. At a molar ratio of 1:100, a second band of higher molecular weight DNA:HIN-A complex appears, which can be formed largely by grouping the HIN-A domains on the DNA molecule. After adding the HIN-A domain to the supercoiled pBluescript samples, no DNA:HIN-A complex was formed even at a molar ratio of 1:200, indicating the absence of any binding to the supercoiled DNA structure. Furthermore, the binding of the HIN-B domain to linear and supercoiled DNA structures was monitored by the same EMSA. As seen in Fig. 2D, the addition of the HIN-B domain to the linear pBluescript samples causes binding to DNA at a DNA:HIN-B molar ratio 1:25, which manifested as a free DNA band shift on the gel. The complete disappearance of free DNA occurred after the addition of HIN-B in a DNA:HIN-B ratio of 1:50.

To compare binding of the IFI16 protein to different DNA structures, we visualized the complexes by atomic force microscopy. The cooperative binding of protein IFI16 to linear DNA supervene in the molar ratio DNA: protein in the form of protein oligomer, and with increasing concentration of protein IFI16, the complex remained similar (Fig. 3G). In the case of supercoiled DNA, the first signs of protein binding to DNA are evident already at molar ratio 1:10 (Fig. 3E) and with increasing concentration of protein significant changes are observed. In Fig. 3I, at molar ratio superhelical DNA:protein 1:20, many local conglomerations of protein IFI16 on DNA are apparent. Under the same conditions, the protein binds to linear DNA in one cluster (Fig. 3F) and other complexes are not evident. The concentration of supercoiled DNA: protein 1:50 shows a large local structure of DNA and a cooperatively binding protein and other smaller complexes as emerging centers suitable for the formation of other oligomers. This indicates that IFI16 the mechanism of its cooperative binding to DNA depends on the structure of DNA. The protein binds to structurally more complex structures such as the supercoiled DNA of pBluescript first in the form of specific one-to-one complexes, which could further facilitate the formation of IFI16 oligomers if the IFI16 concentration increases.

4. Discussion

The results we presented show dual mode of IFI16 binding is depending on DNA topology. IFI16 binds either in a cooperative manner on linear DNA or in a form of globular complexes in presence of negative superhelicity up to certain protein:DNA ratio. Cooperative binding is caused by the nature of the domains in the IFI16 protein. HIN domains interact with DNA, while the protein C-terminal PYRIN oligomerizes. Thus, a single protein molecule bound to dsDNA can "recruit" other IFI16 molecules and, with an increasing number of bound protein molecules, increase the likelihood of IFI16 protein molecules binding to each other and promoting other molecules to DNA through protein-protein interactions between their PYRIN domains. Analysis of the kinetics of protein oligomerization to dsDNA concluded that oligomerization was affected not only by cooperative binding itself, but also by the ability of the IFI16 protein to scan along the dsDNA strand, increasing the probability of finding an already bound IFI16 molecule on dsDNA [42]. This 1D diffusion of the protein along the dsDNA was neither continuous nor dependent on the presence of the PYRIN domain. By this mechanism, the IFI16 protein is likely to increase the potential to form inflamasomes associated with the immune response of the body. From our results we concluded that protein likely forms specific complexes with certain DNA regions with negative superhelicity first, and then could speed up the oligomer formation in the case IFI16 concentration increase in the cell, for example due to infection.

Further light could be shed on the fact whether the superhelicity itself or a particular DNA secondary structure is causing the difference in the binding pattern. According to prediction algorithms, the plasmid used contains 3 G4s and 4 inverted repeats, which are all potentially sources of IFI16 target interactors. The arguments for this phenomenon to be structure-specific are: i. it was described with short oligonucleotides forming both G-quadruplex and cruciform structures [13], ii. we observed dependence on the K ions, the crucial stabilizator of G-quadruplexes, iii. Only the HIN-B domain, which was previously described to bind GC rich regions stronger, showed binding to superhelical DNA. Therefore, we propose that specific recognition of G4 and crucifom could induce a different binding pattern, not a cooperative assembly and oligomerization through the PYD domain of IFI16, but a specific recognition and stabilization of G4 and cruciform. However, the distinction between linear and superhelical DNA, nor HIN-A or HIN-B alone, is sufficient. A recent study describes synergistic binding of p204 HINa and HINb to dsDNA, while the DNA binding affinity of HIN-AB is higher than that of HIN-A or HIN-B alone [43].

One thing to discuss is whether the DNA binding is really a crucial step in restriction of viral infection as proposed by Hotter et al. [27]. They observed that restriction of HIV infection is also possible with the protein construct without HIN domains, just with the PYD and nuclear localization signal region, and independently of immune sensing. HIN domains could interact with DNA whereas

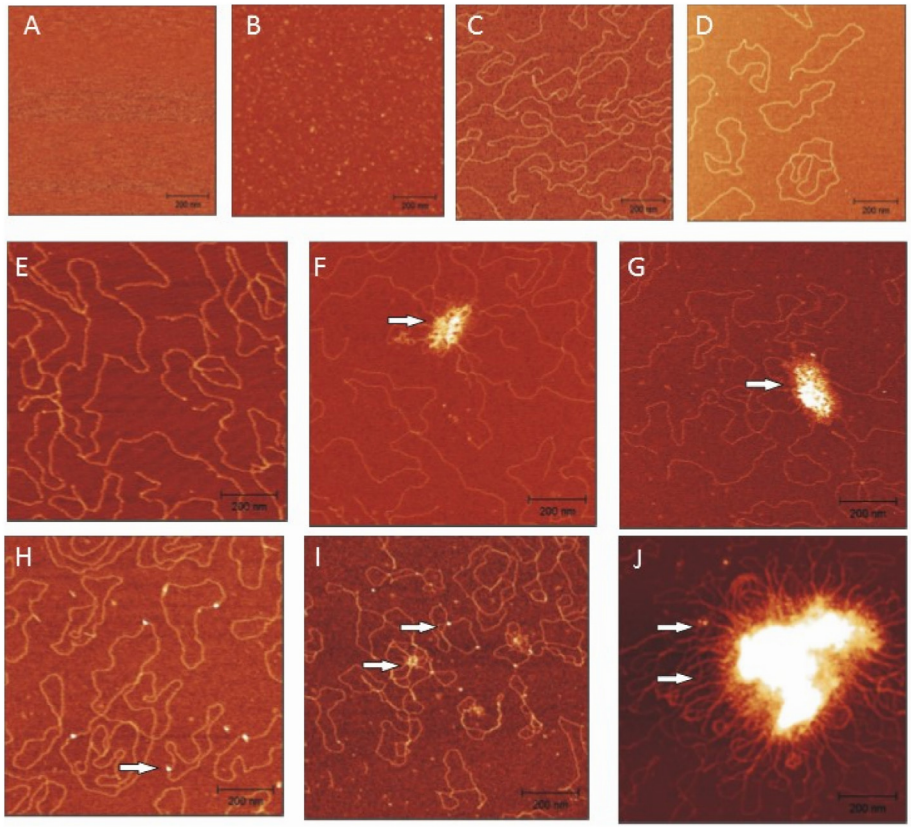


Fig. 3. Atomic force microscopy analysis of IFI16-DNA complexes. (A) Control sample. Pure mica after the whole preparation as with other samples. (B) IFI16 protein. The IFI16 protein was deposited onto mica in the deposition buffer with concentration as for the DNA-protein complexes. (C) Linear DNA. The linearized pBluescript was deposited onto mica - concentration 1ng/ul of DNA. (D) Supercoiled DNA. The supercoiled pBluescript -concentration 1ng/ul of DNA. (E) Linear DNA – IFI16 complex 1:10. The sample of linear DNA and IFI16 mixed in molar ratio DNA:protein 1:10 in binding buffer (incubation 15 min, 4 °C), diluted to 1ng/ul in deposition buffer and deposited onto mica. (F) Linear DNA – IFI16 complex 1:20. (G) Linear DNA – IFI16 complex 1:50. (H) Supercoiled DNA – IFI16 complex 1:10. (I) Supercoiled DNA – IFI16 complex 1:20. (J) Supercoiled DNA – IFI16 complex 1:50. The conditions for measurement were identical for all samples. All images are 1 μm × 1 μm scans and white arrows show the complexes.

PYD domain is adapted to catching up Sp1 transcription factor. Interestingly both Sp1 binding site and *c-myc* NHEIII region, where IFI16 was localized are covered with several putative G-quadruplex forming sites. Sp1 binding site in the U3 region of HIV virus was extensively studied by Richter et al. [44]. Supposedly 4 distinct G-quadruplex structures exist in the region and modulate viral life cycle. We propose that dual mode of IFI16 binding could play a role in distinct IFI16 roles, the protein is bind to DNA in lower molecular ratios during PYD domain is catching up transcription factors or another IFI16 molecules to stimulate innate immune response.

Author contributions

Conceived and designed the experiments: VB, LH. Performed the experiments: LH, LK. Analyzed the data: LH, NV, LK. Contributed reagents/materials/analysis tools: VB. Wrote the paper: NV, LK, LH VB.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by the Czech Science Foundation [No. 22-21903S] to V.B. L.M. was supported by the Ministry of Health, Czech Republic [Conceptual Development of Research Organization MMCI, 00209805].

References

[1] JosephA. Trapani, M. Dawson, VickiA. Apostolidis, KylieA. Browne, Genomic organization of IFI16, an interferon-inducible gene whose expression is associated with human myeloid cell differentiation: correlation of predicted protein domains with exon organization, Immunogenetics 40 (1994), https://doi.org/10.1007/BF00177824.

[2] M.J. Dawson, J.A. Trapani, HIN-200: a novel family of IFN-inducible nuclear proteins expressed in leukocytes, J. Leukoc. Biol. 60 (1996) 310–316, https://doi.org/10.1002/jlb.60.3.310.

[3] J. Bertin, P.S. DiStefano, The PYRIN domain: a novel motif found in apoptosis and inflammation proteins, Cell Death Differ. 7 (2000) 1273–1274, https://doi.org/10.1038/sj.cdd.4400774.

[4] X. Ni, H. Ru, F. Ma, L. Zhao, N. Shaw, Y. Feng, W. Ding, W. Gong, Q. Wang, S. Ouyang, G. Cheng, Z.-J. Liu, New insights into the structural basis of DNA recognition by HINa and HINb domains of IFI16, J. Mol. Cell Biol. 8 (2016) 51–61, https://doi.org/10.1093/jmcb/mjv053.

[5] D. Li, L. Xie, Z. Qiao, J. Zhu, H. Yao, Y. Qin, Y. Yan, Z. Chen, F. Ma, IFI16 isoforms with cytoplasmic and nuclear locations play differential roles in recognizing invaded DNA viruses, J. Immunol. 207 (2021) 2699–2709, https://doi.org/10.4049/jimmunol.2100398.

[6] I. Kimkong, Y. Avihingsanon, N. Hirankarn, Expression profile of HIN200 in leukocytes and renal biopsy of SLE patients by real-time RT-PCR, Lupus 18 (2009) 1066–1072, https://doi.org/10.1177/0961203309106699.

[7] T. Li, B.A. Diner, J. Chen, I.M. Cristea, Acetylation modulates cellular distribution and DNA sensing ability of interferon-inducible protein IFI16, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 10558–10563, https://doi.org/10.1073/pnas.1203447109.

[8] V. Dell'Oste, D. Gatti, F. Gugliesi, M.D. Andrea, M. Bawadekar, I.L. Cigno, M. Biolatti, M. Vallino, M. Marschall, M. Gariglio, S. Landolfo, Innate nuclear

sensor IFI16 translocates into the cytoplasm during the early stage of in vitro human cytomegalovirus infection and is entrapped in the egressing virions during the late stage, J. Virol. 88 (2014) 13.
[9] S. Costa, C. Borgogna, M. Mondini, M. De Andrea, P.L. Meroni, E. Berti, M. Gariglio, S. Landolfo, Redistribution of the nuclear protein IFI16 into the cytoplasm of ultraviolet B-exposed keratinocytes as a mechanism of autoantigen processing: UVB-induced IFI16 redistribution, Br. J. Dermatol. 164 (2011) 282–290, https://doi.org/10.1111/j.1365-2133.2010.10097.x.
[10] S.R. Morrone, T. Wang, L.M. Constantoulakis, R.M. Hooy, M.J. Delannoy, J. Sohn, Cooperative assembly of IFI16 filaments on dsDNA provides insights into host defense strategy, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E62–E71, https://doi.org/10.1073/pnas.1313577111.
[11] H. Yan, K. Dalal, B.K. Hon, P. Youkharibache, D. Lau, F. Pio, RPA nucleic acid-binding properties of IFI16-HIN200, Biochim. Biophys. Acta, Proteins Proteomics 1784 (2008) 1087–1097, https://doi.org/10.1016/j.bbapap.2008.04.004.
[12] V. Brázda, J. Coufal, J.C.C. Liao, C.H. Arrowsmith, Preferential binding of IFI16 protein to cruciform structure and superhelical DNA, Biochem. Biophys. Res. Commun. 422 (2012) 716–720, https://doi.org/10.1016/j.bbrc.2012.05.065.
[13] L. Hároníková, J. Coufal, I. Kejnovská, E.B. Jagelská, M. Fojta, P. Dvořáková, P. Muller, B. Vojtesek, V. Brázda, IFI16 preferentially binds to DNA with quadruplex structure and enhances DNA quadruplex formation, PLoS One 11 (2016), e0157156, https://doi.org/10.1371/journal.pone.0157156.
[14] L. Egistelli, S. Chichiarelli, E. Gaucci, M. Eufemi, M.E. Schininà, A. Giorgi, I. Lascu, C. Turano, A. Giartosio, L. Cervoni, IFI16 and NM23 bind to a common DNA fragment both in the P53 and the cMYC gene promoters, J. Cell. Biochem. 106 (2009) 666–672, https://doi.org/10.1002/jcb.22053.
[15] R.I. Mathad, E. Hatzakis, J. Dai, D. Yang, c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability, Nucleic Acids Res. 39 (2011) 9023–9033, https://doi.org/10.1093/nar/gkr612.
[16] C. Nakanishi, H. Seimiya, G-quadruplex in cancer biology and drug discovery, Biochem. Biophys. Res. Commun. 531 (2020) 45–50, https://doi.org/10.1016/j.bbrc.2020.03.178.
[17] E. Ruggiero, S.N. Richter, Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy, in: Annual Reports in Medicinal Chemistry, Elsevier, 2020, pp. 101–131, https://doi.org/10.1016/bs.armc.2020.04.001.
[18] R.W. Johnstone, J.A. Kerry, J.A. Trapani, The human interferon-inducible protein, IFI 16, is a repressor of transcription, J. Biol. Chem. 273 (1998) 17172–17177, https://doi.org/10.1074/jbc.273.27.17172.
[19] K.E. Johnson, V. Bottero, S. Flaherty, S. Dutta, V.V. Singh, B. Chandran, IFI16 restricts HSV-1 replication by accumulating on the HSV-1 genome, repressing HSV-1 gene expression, and directly or indirectly modulating histone modifications, PLoS Pathog. 10 (2014), e1004503, https://doi.org/10.1371/journal.ppat.1004503.
[20] L.-T. Ong, W.C. Lee, S. Ma, G. Oguz, Z. Niu, Y. Bao, M. Yusuf, P.L. Lee, J.Y. Goh, P. Wang, K.S.M. Yong, Q. Chen, W. Wang, A. Ramasamy, D.S.B. Hoon, H.J. Ditzel, E.Y. Tan, S.C. Lee, Q. Yu, IFI16-dependent STING signaling is a crucial regulator of anti-HER2 immune response in HER2+ breast cancer, Proc. Natl. Acad. Sci. U. S. A. 119 (2022), e2201376119, https://doi.org/10.1073/pnas.2201376119.
[21] P.P. Piccaluga, C. Agostinelli, F. Fuligni, S. Righi, C. Tripodo, M.C. Re, A. Clò, A. Miserocchi, S. Morini, M. Gariglio, G.G. Ferri, A. Rinaldi-Ceroni, O. Piccin, M. De Andrea, S.A. Pileri, S. Landolfo, D. Gibellini, IFI16 expression is related to selected transcription factors during B-cell differentiation, J. Immunol. Res. 2015 (2015), https://doi.org/10.1155/2015/747645.
[22] M.A. Ansari, V.V. Singh, S. Dutta, M.V. Veettil, D. Dutta, L. Chikoti, J. Lu, D. Everly, B. Chandran, Constitutive interferon-inducible protein 16-inflammasome activation during epstein-barr virus latency I, II, and III in B and epithelial cells, J. Virol. 87 (2013) 8606–8623, https://doi.org/10.1128/JVI.00805-13.
[23] L. Unterholzner, S.E. Keating, M. Baran, K.A. Horan, S.B. Jensen, S. Sharma, C.M. Sirois, T. Jin, E. Latz, T.S. Xiao, K.A. Fitzgerald, S.R. Paludan, A.G. Bowie, IFI16 is an innate immune sensor for intracellular DNA, Nat. Immunol. 11 (2010) 997–1004, https://doi.org/10.1038/ni.1932.
[24] S. Mishra, A.S. Raj, A. Kumar, A. Rajeevan, P. Kumari, H. Kumar, Innate immune sensing of influenza A viral RNA through IFI16 promotes pyroptotic cell death, iScience 25 (2022), 103714, https://doi.org/10.1016/j.isci.2021.103714.
[25] J.F. Almine, C.A.J. O'Hare, G. Dunphy, I.R. Haga, R.J. Naik, A. Atrih, D.J. Connolly, J. Taylor, I.R. Kelsall, A.G. Bowie, P.M. Beard, L. Unterholzner, IFI16 and cGAS cooperate in the activation of STING during DNA sensing in human keratinocytes, Nat. Commun. 8 (2017), 14392, https://doi.org/10.1038/ncomms14392.
[26] M. Bosso, C. Prelli Bozzo, D. Hotter, M. Volcic, C.M. Stürzel, A. Rammelt, Y. Ni, S. Urban, M. Becker, M. Schelhaas, S. Wittmann, M.H. Christensen, F.I. Schmidt, T. Gramberg, K.M.J. Sparrer, D. Sauter, F. Kirchhoff, Nuclear PYHIN proteins target the host transcription factor Sp1 thereby restricting HIV-1 in human macrophages and CD4+ T cells, PLoS Pathog. 16 (2020), e1008752, https://doi.org/10.1371/journal.ppat.1008752.
[27] D. Hotter, M. Bosso, K.L. Jønsson, C. Krapp, C.M. Stürzel, A. Das, E. Littwitz-Salomon, B. Berkhout, A. Russ, S. Wittmann, T. Gramberg, Y. Zheng, L.J. Martins, V. Planelles, M.R. Jakobsen, B.H. Hahn, U. Dittmer, D. Sauter, F. Kirchhoff, IFI16 targets the transcription factor Sp1 to suppress HIV-1 transcription and latency reactivation, Cell Host Microbe 25 (2019) 858–872, https://doi.org/10.1016/j.chom.2019.05.002, e13.
[28] W. Vanhove, P.M. Peeters, D. Staelens, A. Schraenen, J. Van der Goten, I. Cleynen, S. De Schepper, L. Van Lommel, N.L. Reynaert, F. Schuit, G. Van Assche, M. Ferrante, G. De Hertogh, E.F.M. Wouters, P. Rutgeerts, S. Vermeire, K. Nys, I. Arijs, Strong upregulation of AIM2 and IFI16 inflammasomes in the mucosa of patients with active inflammatory bowel disease, Inflamm. Bowel Dis. 21 (2015) 2673–2682, https://doi.org/10.1097/MIB.0000000000000535.
[29] B. Antiochos, D. Trejo-Zambrano, P. Fenaroli, A. Rosenberg, A. Baer, A. Garg, J. Sohn, J. Li, M. Petri, D.W. Goldman, C. Mecoli, L. Casciola-Rosen, A. Rosen, The DNA sensors AIM2 and IFI16 are SLE autoantigens that bind neutrophil extracellular traps, Elife 11 (2022), e72103, https://doi.org/10.7554/eLife.72103.
[30] T. Cao, S. Shao, B. Li, L. Jin, J. Lei, H. Qiao, G. Wang, Up-regulation of Interferon-inducible protein 16 contributes to psoriasis by modulating chemokine production in keratinocytes, Sci. Rep. 6 (2016), 25381, https://doi.org/10.1038/srep25381.
[31] A. Alunno, V. Caneparo, O. Bistoni, S. Caterbi, R. Terenzi, M. Gariglio, E. Bartoloni, A. Manzo, S. Landolfo, R. Gerli, Circulating interferon-inducible protein IFI16 correlates with clinical and serological features in rheumatoid arthritis, Arthritis Care Res. 68 (2016) 440–445, https://doi.org/10.1002/acr.22695.
[32] Z.H. McMahan, T.R. Cottrell, F.M. Wigley, B. Antiochos, E.T. Zambidis, T. Soon Park, M.K. Halushka, L. Gutierrez-Alamillo, R. Cimbro, A. Rosen, L. Casciola-Rosen, Enrichment of scleroderma vascular disease–associated autoantigens in endothelial lineage cells, Arthritis Rheumatol. 68 (2016) 2540–2549, https://doi.org/10.1002/art.39743.
[33] A.N. Baer, M. Petri, J. Sohn, A. Rosen, L. Casciola-Rosen, Association of antibodies to interferon-inducible protein-16 with markers of more severe disease in primary sjögren's syndrome, Arthritis Care Res. 68 (2016) 254–260, https://doi.org/10.1002/acr.22632.
[34] G. Pisano, A. Roy, M. Ahmed Ansari, B. Kumar, L. Chikoti, B. Chandran, Interferon-γ-inducible protein 16 (IFI16) is required for the maintenance of Epstein-Barr virus latency, Virol. J. 14 (2017) 221, https://doi.org/10.1186/s12985-017-0891-5.
[35] A. Roy, D. Dutta, J. Iqbal, G. Pisano, O. Gjyshi, M.A. Ansari, B. Kumar, B. Chandran, Nuclear innate immune DNA sensor IFI16 is degraded during lytic reactivation of kaposi's sarcoma-associated herpesvirus (KSHV): role of IFI16 in maintenance of KSHV latency, J. Virol. 90 (2016) 8822–8841, https://doi.org/10.1128/JVI.01003-16.
[36] G. Griffante, W. Hewelt-Belka, C. Albano, F. Gugliesi, S. Pasquero, S.F. Castillo Pacheco, G. Bajetto, P.E. Porporato, E. Mina, M. Vallino, C. Krapp, M.R. Jakobsen, J. Purdy, J. von Einem, S. Landolfo, V. Dell'Oste, M. Biolatti, IFI16 impacts metabolic reprogramming during human cytomegalovirus infection, mBio 13 (2022), e0043522, https://doi.org/10.1128/mbio.00435-22.
[37] T. Li, J. Chen, I.M. Cristea, Human cytomegalovirus tegument protein pUL83 inhibits IFI16-mediated DNA sensing for immune evasion, Cell Host Microbe 14 (2013) 591–599, https://doi.org/10.1016/j.chom.2013.10.007.
[38] M.R. Jakobsen, R.O. Bak, A. Andersen, R.K. Berg, S.B. Jensen, J. Tengchuan, T. Jin, A. Laustsen, K. Hansen, L. Ostergaard, K.A. Fitzgerald, T.S. Xiao, J.G. Mikkelsen, T.H. Mogensen, S.R. Paludan, IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E4571–E4580, https://doi.org/10.1073/pnas.1311669110.
[39] K. Hansen, T. Prabakaran, A. Laustsen, S.E. Jørgensen, S.H. Rahbæk, S.B. Jensen, R. Nielsen, J.H. Leber, T. Decker, K.A. Horan, M.R. Jakobsen, S.R. Paludan, Listeria monocytogenes induces IFNβ expression through an IFI16-, cGAS- and STING-dependent pathway, EMBO J. 33 (2014) 1654–1666, https://doi.org/10.15252/embj.201488029.
[40] V. Brázda, J. Kolomazník, J. Lýsek, L. Hároníková, J. Coufal, J. Šťastný, Palindrome analyser - a new web-based server for predicting and evaluating inverted repeats in nucleotide sequences, Biochem. Biophys. Res. Commun. 478 (2016) 1739–1745, https://doi.org/10.1016/j.bbrc.2016.09.015.
[41] V. Brázda, J. Kolomazník, J. Lýsek, M. Bartas, M. Fojta, J. Šťastný, J.-L. Mergny, G4Hunter web application: a web server for G-quadruplex prediction, Bioinformatics 35 (2019) 3493–3495, https://doi.org/10.1093/bioinformatics/btz087.
[42] S.A. Stratmann, S.R. Morrone, A.M. van Oijen, J. Sohn, The innate immune sensor IFI16 recognizes foreign DNA in the nucleus by scanning along the duplex, Elife 4 (2015), e11721, https://doi.org/10.7554/eLife.11721.
[43] X. Fan, J. Jiang, D. Zhao, F. Chen, H. Ma, P. Smith, L. Unterholzner, T.S. Xiao, T. Jin, Structural mechanism of DNA recognition by the p204 HIN domain, Nucleic Acids Res. 49 (2021) 2959–2972, https://doi.org/10.1093/nar/gkab076.
[44] R. Perrone, M. Nadai, I. Frasson, J.A. Poe, E. Butovskaya, T.E. Smithgall, M. Palumbo, G. Palù, S.N. Richter, A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter, J. Med. Chem. 56 (2013) 6521–6530, https://doi.org/10.1021/jm400914r.

Příloha D

Novel G-quadruplex prone sequences emerge in the complete assembly of the human X chromosome

Biochimie 191 (2021) 87–90

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Novel G-quadruplex prone sequences emerge in the complete assembly of the human X chromosome

Natália Bohálová [a, b], Jean-Louis Mergny [a, c], Václav Brázda [a, *]

[a] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65, Brno, Czech Republic
[b] Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 62500, Brno, Czech Republic
[c] Laboratoire d'Optique & Biosciences, Ecole Polytechnique, CNRS, Inserm, Institut Polytechnique de Paris, Palaiseau, France

ARTICLE INFO

Article history:
Received 3 August 2021
Received in revised form
1 September 2021
Accepted 5 September 2021
Available online 9 September 2021

Keywords:
G-quadruplex
Chromosome X
G4Hunter
Bioinformatics
Non-B DNA

ABSTRACT

G-quadruplexes are non-B secondary structures with regulatory functions and therapeutic potential. Improvements in sequencing methods recently allowed the completion of the first human chromosome which is now available as a gapless, end-to-end assembly, with the previously remaining spaces filled and newly identified regions added. We compared the presence of G-quadruplex forming sequences in the current human reference genome (GRCh38) and in the new end-to-end assembly of the X chromosome constructed by high-coverage ultra-long-read nanopore sequencing. This comparison revealed that, even though the corrected length of the chromosome X assembly is surprisingly 1.14% shorter than expected, the number of G-quadruplex forming sequences found in this gapless chromosome is significantly higher, with 493 new motifs having G4Hunter scores above 1.4 and 23 new sequences with G4Hunter scores above 3.5. This observation reflects an improved precision of the new sequencing approaches and points to an underestimation of G-quadruplex propensity in the previous, widely used version of the human genome assembly, especially for motifs with a high G4Hunter score, expected to be very stable. These G-quadruplex forming sequences probably remained undiscovered in earlier genome datasets due to previously unsolved G-rich and repetitive genomic regions. These observations allow a precise targeting of these important regulatory regions.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The contemporary reference assembly of the human genome (GRCh38) does not contain any complete chromosome, *i.e.,* fully sequenced from telomere to telomere. The current reference assembly of chromosome X in the Genome database contains 39 missing regions, which sequences were not determined and include more than 1 million (1,147,861) unidentified bases (0.74%). Recently, the first full end-to-end sequence of the human X chromosome was released with reconstructed centromeric satellite DNA, all previously remaining gaps filled, and including newly identified sequences from human pseudo autosomal regions and cancer-testis ampliconic gene families [1]. Advanced assembly approaches and the combination of several sequencing methods, including ultra-long nanopore sequence reading, enabled resolving the most repetitive regions as well as the mapping of methylation patterns across complex tandem repeats and satellite arrays [1]. Finally, completion of the entire human genome sequence is now within reach, and this will significantly improve our current knowledge of the human genome and, as a consequence, of previously hidden structural and regulatory regions.

While double-stranded DNA is predominantly arranged into a right-handed B-DNA conformation, a variety of local DNA structures has been found playing critical roles in the regulation of key cellular processes. The formation of G-quadruplexes (G4s) in cells is now firmly demonstrated [2–4], and these structures may contribute to normal and pathogenic processes [5]. G4s are formed by G-rich nucleic acid sequences and result from the stacking of two or more G-quartets, in which four guanines establish a cyclic array of Hoogsteen hydrogen bonds. G4-prone motifs are often located in promoters, regulatory regions and telomeres [5,6]. G4s play roles in transcription, translation, genomic stability, and telomere maintenance [7,8]. Moreover, some G4s can be more stable than B-DNA and constitute possible targets for antiviral and antiproliferative therapeutics [9,10].

* Corresponding author.
E-mail address: vaclav@ibp.cz (V. Brázda).

https://doi.org/10.1016/j.biochi.2021.09.004
0300-9084/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2. Materials and methods

To compare the X chromosome assemblies, we downloaded the full chromosome sequences from the National Center for Biotechnology Information (NCBI) database: the current reference sequence GRCh38 (NCBI ID NC_000023.11) and the recently published new end-to-end X chromosome assembly (NCBI ID CP068255.2). G4Hunter analyses were conducted with default parameters as previously described [11]. The G4Hunter algorithm finds all sequences with a score above selected G4Hunter threshold; some G or C rich regions can be found in several overlapping windows when long runs are considered [12]. For this reason, we performed postprocessing analyses of data to evaluate not only the total number of PQS in the dataset, but also the total coverage of PQS sequences without overlaps (*i.e.*, the fraction of nucleotides in the genomes which are present in a window with a G4Hunter score above chosen threshold).

3. Results and discussion

Accurate assessment of G4 propensity at the whole genome level is crucially important. These unusual DNA structures are often found in repetitive or GC-rich sequences, which frequently correspond to regions where gaps or errors are present [12,13]. We reasoned that these gaps in the contemporary GRCh38 assembly may have a disproportionally strong effect on the estimation of G4 content. Therefore, we analyzed the new gapless assembly of the X chromosome to evaluate the presence of G-quadruplex forming sequences and compared it to GRCh38.

Comparing the basic parameters of the current reference and the new assembly revealed that, while the GC content of these sequences was almost identical (39.53% vs. 39.51%), the fully assembled chromosome X sequence was 1.8 Mbp shorter (1.14%) than the reference GRCh38 X assembly (Table 1, upper part). The total number of GC bases in the new assembly dropped by only 0.45%, suggesting that those nucleotides previously assigned incorrectly were predominantly AT bases.

Table 1
Comparison of G4-prone sequence prevalence in X chromosome (assembly GRCh38, new assembly, raw data in Supplementary material 1). Upper part: Length and GC content comparison; Δ (difference in numbers: X new − X GRCh38), Δ % ((X new − X GRCh38)/X GRCh38). Middle part: G4-forming sequences numbers comparison. Lower part: G4-forming sequences non-overlapping length in bp comparison.

	X assembly GRCh38	X − new assembly	Δ	Δ%
Length	156,040,895	154,259,566	−1,781,329	−1.14
GC	61,221,521	60,948,906	−272,615	−0.45
GC (%)	39.53	39.51	−0.02	−0.05
G4Hunter score	X assembly GRCh 38	X − new assembly	Δ	Δ%
1.2	224,745	224,875	+130	+0.06
1.4	118,819	119,312	+493	+0.41
1.6	65,675	66,331	+656	+1.00
1.8	37,200	38,776	+1576	+4.24
2.0	21,349	22,323	+974	+4.56
2.5	5439	5613	+174	+3.20
3.0	706	817	+111	+15.72
3.5	36	59	+23	+63.89
G4Hunter score	X assembly GRCh 38	X − new assembly	Δ	Δ%
1.2	6,148,804	6,184,086	+35,282	+0.57
1.4	3,299,278	3,334,399	+35,121	+1.06
1.6	1,831,728	1,862,441	+30,713	+1.68
1.8	1,042,172	1,070,476	+28,304	+2.72
2.0	597,314	617,800	+20,486	+3.43
2.5	151,482	156,610	+5128	+3.39
3.0	17,682	20,208	+2526	+14.29
3.5	912	1334	+422	+46.27

Even more surprising was the analysis of these sequences using G4Hunter, a predictive algorithm for G4 propensity with unprecedented accuracy [11,14]. Despite its shorter length and slightly lower GC content, the gapless X chromosome bears a higher number of G4-forming sequences (Table 1, middle part). More G4-forming sequences were found for all G4Hunter score parameters, and this difference was proportionally greater when considering only the most stable motifs (*i.e.*, those with high G4Hunter scores; Table 1, middle part). This difference was already noteworthy for sequences with G4Hunter score >1.8, in which case 1576 new G4-prone sequences were detected (+4.24% compared to reference assembly 38), and even more dramatic for G4Hunter scores above 3.0 (+15.7%) or 3.5 (+64%; Table 1). In other words, the higher the G4Hunter score, the larger the difference between the two versions of the X-chromosome. This is an important observation, as the most stable quadruplexes, which may pose problems during replication, transcription (or sequencing) tend to have high G4Hunter scores. A similar trend was observed also when considering G4-forming sequences coverage without overlapping PQS (Table 1, lower part). All found PQS are shown in Supplementary material 2; among them, we found 23 novel PQS sequences with a G4Hunter score above 3.5. These novel motifs emerged at different locations of the X chromosome (Fig. 1D). Quadruplex formation for very G-rich sequences such as those identified in our analyses (Fig. 1E) was experimentally confirmed *in vitro* [15,16]. Unfortunately, the new gapless assembly is not fully annotated yet, so the comparison of newly identified G4s locations with the older assembly is currently difficult. Primarily annotation is accessible now using a completely different annotation system. Nerveless, preliminary analyses and comparison of G4-forming sequences presence in these annotations (Supplementary material 2) shows a strong abundance of G4-forming sequences before and after "*gene*" annotation in both old and new assemblies, pointing to the importance of the regulatory functions of G4 for gene expression. The highest density of G4-forming sequences in the old assembly is in "*protein bind*" region; unfortunately, this region is not annotated yet in the new assembly. As higher G4Hunter scores correspond to higher probability of motifs formation and higher stability, our analyses show that the widely used GRCh38 human assembly underestimates the density of these important regulatory sequences. While in GRCh38 the source of DNA was derived from thirteen anonymous volunteers [17], the new gapless assembly has been done on a single human cell line (CHM13) uniformly homozygous for one set of alleles [1]. While differences in genomic sequences are expected among all individuals and are within 0.3% of sequences [18], we found a significant underestimation of stable PQS sequences (from ~4% for PQS with G4Hunter scores above 1.8 up to 15% and 64% abundance for G4Hunter scores above 3 and 3.5 respectively). An increasing number of studies use G4s as targets for gene regulation, and accurate genome- or transcriptome-wide analyses of the effects of G4 ligands require full characterization of G4 propensity in the target genome. The current release of gapless assemblies enables a proper appreciation of the cellular regulatory processes regulated by these structures. Our analyses of the recently published end-to-end assembly of the human X chromosome demonstrate that these non-B DNA structures occur more frequently than estimated from previous studies on the current reference GRCh38 genome. Besides improved G4 prediction, our results point to the necessity of complete sequence knowledge for G4 experimental validation methods such as G4seq to avoid discrepancies between predictions and experimental characterization. As this difference is especially striking for the most stable candidates, this observation may also be

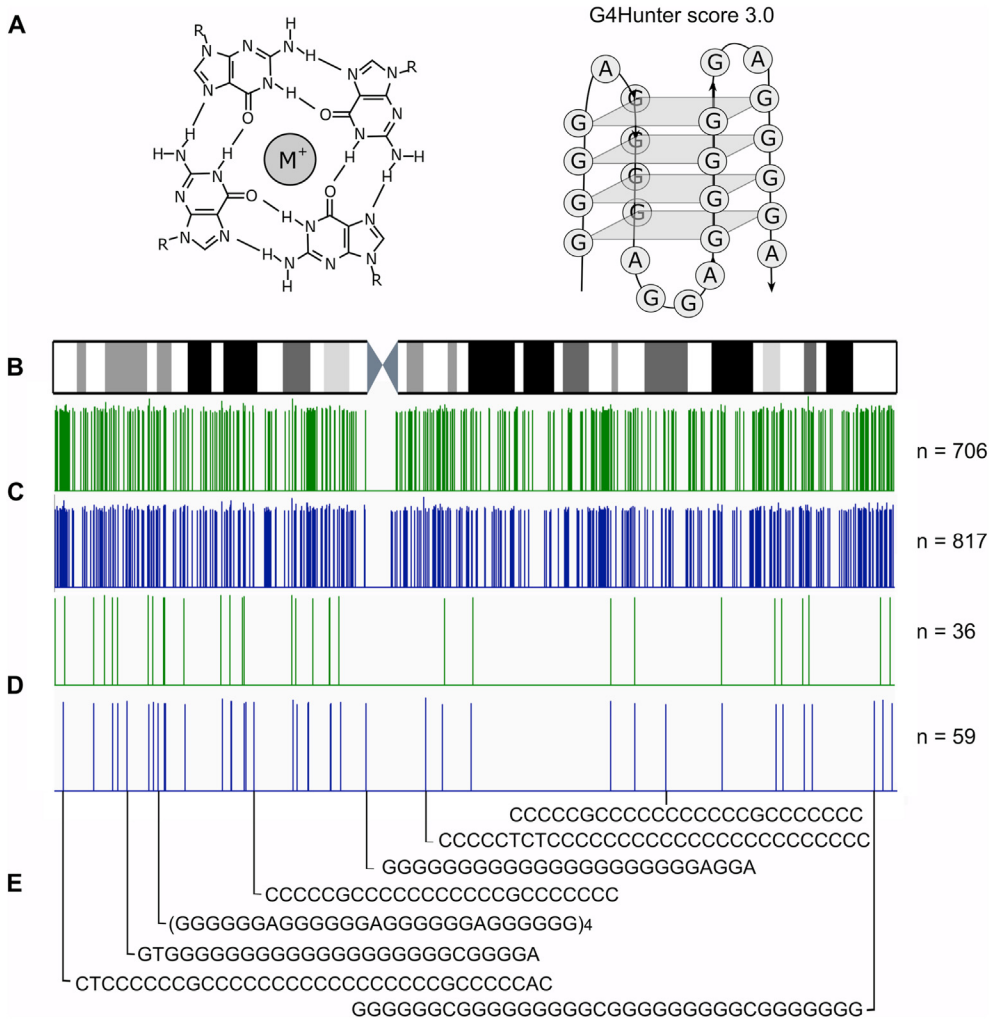


Fig. 1. Analyses of G4-forming sequences on human chromosome X assembly GRCh38 and new telomere-to-telomere assembly. (**A**) G-quartet stabilized by monovalent ion (*left*) and G-quadruplex structure of a sequence with G4Hunter score of 3.0 (*right*). (**B**) G-banding of chromosome X. (**C**) Potential G4-forming sequences with G4Hunter score above a threshold of 3.0 in the GRCh38 assembly (green) and the new telomere-to-telomere assembly (blue). (**D**) Localization of potential G4-forming sequences with G4Hunter score above a threshold of 3.5 on chromosome X GRCh38 assembly (green) and new telomere-to-telomere assembly (blue), sequences and their comparison accessible at Supplementary material 1. The higher the peak, the higher the G4Hunter score; higher *n* indicates a higher number of G4-forming sequences on the human X chromosome. (**E**) Examples of newly identified sequences.

relevant for i-DNA, another unusual motif formed by C-rich DNA sequences, for which only long and highly skewed motifs have been found to be stable in cells [19,20].

4. Conclusion

The search for G-quadruplex sequences in the human genome has been performed several times. However, until very recently, it was not possible to perform an exhaustive search due to missing information in the non-complete assemblies. Our results allow to re-evaluate the number of PQS in humans, allowing a better understanding of the targeting of these structures. A similar re-evaluation will be important in other species once their gapless genome is available, as G4-prone motifs are found in all domains of life [21–23].

Author contributions

Data was collected and analyzed by VB and NB. Figures and tables were prepared by VB and NB, with input from JLM. All authors wrote the manuscript.

Declaration of competing interest

The authors declare no conflict of interests.

Acknowledgements

We thank Gale A. Kirking for proofreading of this article. This work was funded by the Agence Nationale de la Recherche (*G4Access* grant ANR-20-CE12-0023) and the SYMBIT project (Reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477) financed by the ERDF.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biochi.2021.09.004.

References

[1] K.H. Miga, S. Koren, A. Rhie, M.R. Vollger, A. Gershman, A. Bzikadze, S. Brooks, E. Howe, D. Porubsky, A.M. Phillippy, Telomere-to-telomere assembly of a complete human X chromosome, Nature 585 (2020) 79–84, https://doi.org/

10.1038/s41586-020-2547-7.
[2] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H.J. Lipps, A. Plückthun, In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei, Proc. Natl. Acad. Sci. Unit. States Am. 98 (2001) 8572–8577, https://doi.org/10.1073/pnas.141229498.
[3] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells, Nat. Chem. 5 (2013) 182–186, https://doi.org/10.1038/nchem.1548.
[4] G.F. Salgado, C. Cazenave, A. Kerkour, J.-L. Mergny, G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy, Chem. Sci. 6 (2015) 3314–3320, https://doi.org/10.1039/C4SC03853C.
[5] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, The regulation and functions of DNA and RNA G-quadruplexes, Nat. Rev. Mol. Cell Biol. 21 (2020) 459–474, https://doi.org/10.1038/s41580-020-0236-x.
[6] V. Brázda, M. Bartas, R.P. Bowater, Evolution of diverse strategies for promoter regulation, Trends Genet. 37 (2021) 730–744, https://doi.org/10.1016/j.tig.2021.04.003.
[7] S. Ravichandran, J.-H. Ahn, K.K. Kim, Unraveling the regulatory G-quadruplex puzzle: lessons from genome and transcriptome-wide studies, Front. Genet. 10 (2019) 1002, https://doi.org/10.3389/fgene.2019.01002.
[8] L.I. Jansson, J. Hentschel, J.W. Parks, T.R. Chang, C. Lu, R. Baral, C.R. Bagshaw, M.D. Stone, Telomere DNA G-quadruplex folding within actively extending human telomerase, Proc. Natl. Acad. Sci. Unit. States Am. 116 (2019) 9350–9359, https://doi.org/10.1073/pnas.1814777116.
[9] E. Ruggiero, S.N. Richter, Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy, Annu. Rep. Med. Chem. 54 (2020) 101–131, https://doi.org/10.1016/bs.armc.2020.04.001.
[10] J. Carvalho, J.-L. Mergny, G.F. Salgado, J.A. Queiroz, C. Cruz, G-quadruplex, friend or foe: the role of the G-quartet in anticancer strategies, Trends Mol. Med. 26 (2020) 848–861, https://doi.org/10.1016/j.molmed.2020.05.002.
[11] V. Brázda, J. Kolomazník, J. Lýsek, M. Bartas, M. Fojta, J. Šťastný, J.-L. Mergny, G4Hunter web application: a web server for G-quadruplex prediction, Bioinformatics 35 (2019) 3493–3495, https://doi.org/10.1093/bioinformatics/btz087.
[12] T.J. Treangen, S.L. Salzberg, Repetitive DNA and next-generation sequencing: computational challenges and solutions, Nat. Rev. Genet. 13 (2012) 36–46, https://doi.org/10.1038/nrg3117.
[13] J. Spiegel, S. Adhikari, S. Balasubramanian, The structure and function of DNA G-quadruplexes, TRECHEM 2 (2020) 123–136, https://doi.org/10.1016/j.trechm.2019.07.002.
[14] A. Bedrat, L. Lacroix, J.-L. Mergny, Re-evaluation of G-quadruplex propensity with G4Hunter, Nucleic Acids Res. 44 (2016) 1746–1759, https://doi.org/10.1093/nar/gkw006.
[15] M. Bartas, V. Brázda, V. Karlický, J. Červeň, P. Pečinka, Bioinformatics analyses and in vitro evidence for five and six stacked G-quadruplex forming sequences, Biochimie 150 (2018) 70–75, https://doi.org/10.1016/j.biochi.2018.05.002.
[16] A. Sengar, B. Heddi, A.T. Phan, Formation of G-quadruplexes in poly-G sequences: structure of a propeller-type parallel-stranded G-quadruplex formed by a G_{15} stretch, Biochemistry 53 (2014) 7718–7723, https://doi.org/10.1021/bi500990v.
[17] A.P. Read, Stewart Scherer, A short guide to the human genome, Hum. Genet. 125 (2009), https://doi.org/10.1007/s00439-008-0609-2, 231–231.
[18] A. Auton, G.R. Abecasis, D.M. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, A.G. Clark, P. Donnelly, et al., A global reference for human genetic variation, Nature 526 (2015) 68–74, https://doi.org/10.1038/nature15393.
[19] S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J.-L. Mergny, S. Foldynova-Trantirkova, L. Trantirek, Evaluation of the stability of DNA i-motifs in the nuclei of living mammalian cells, Angew Chem. Int. Ed. Engl. 57 (2018) 2165–2169, https://doi.org/10.1002/anie.201712284.
[20] M. Cheng, D. Qiu, L. Tamon, E. Ištvánková, P. Víšková, S. Amrane, A. Guédin, J. Chen, L. Lacroix, H. Ju, L. Trantírek, A.B. Sahakyan, J. Zhou, J.-L. Mergny, Thermal and pH stabilities of i-DNA: confronting in vitro experiments with models and in-cell NMR data, Angew Chem. Int. Ed. Engl. 60 (2021) 10286–10294, https://doi.org/10.1002/anie.202016801.
[21] A. Abiri, M. Lavigne, M. Rezaei, S. Nikzad, P. Zare, J.-L. Mergny, H.-R. Rahimi, Unlocking G-quadruplexes as antiviral targets, Pharmacol. Rev. 73 (2021) 897–923.
[22] M. Bartas, M. Čutová, V. Brázda, P. Kaura, J. Šťastný, J. Kolomazník, J. Coufal, P. Goswami, J. Červeň, P. Pečinka, The presence and localization of G-quadruplex forming sequences in the domain of bacteria, Molecules 24 (2019) 1711, https://doi.org/10.3390/molecules24091711.
[23] V. Brázda, Y. Luo, M. Bartas, P. Kaura, O. Porubiaková, J. Šťastný, P. Pečinka, D. Verga, V. Da Cunha, et al., G-quadruplexes in the archaea domain, Biomolecules 10 (2020) 1349, https://doi.org/10.3390/biom10091349.

Příloha E

New telomere to telomere assembly of human chromosome 8 reveals a previous underestimation of G-quadruplex forming sequences and inverted repeats

Gene 810 (2022) 146058

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

New telomere to telomere assembly of human chromosome 8 reveals a previous underestimation of G-quadruplex forming sequences and inverted repeats

Václav Brázda [a,*], Natália Bohálová [a,b], Richard P. Bowater [c,*]

[a] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, Brno 612 65, Czech Republic
[b] Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, Brno 62500, Czech Republic
[c] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

ARTICLE INFO

Keywords:
G-quadruplex
Inverted repeat
Genome sequence of human chromosome 8
Non-B DNA structures

ABSTRACT

Taking advantage of evolving and improving sequencing methods, human chromosome 8 is now available as a gapless, end-to-end assembly. Thanks to advances in long-read sequencing technologies, its centromere, telomeres, duplicated gene families and repeat-rich regions are now fully sequenced. We were interested to assess if the new assembly altered our understanding of the potential impact of non-B DNA structures within this completed chromosome sequence. It has been shown that non-B secondary structures, such as G-quadruplexes, hairpins and cruciforms, have important regulatory functions and potential as targeted therapeutics. Therefore, we analysed the presence of putative G-quadruplex forming sequences and inverted repeats in the current human reference genome (GRCh38) and in the new end-to-end assembly of chromosome 8. The comparison revealed that the new assembly contains significantly more inverted repeats and G-quadruplex forming sequences compared to the current reference sequence. This observation can be explained by improved accuracy of the new sequencing methods, particularly in regions that contain extensive repeats of bases, as is preferred by many non-B DNA structures. These results show a significant underestimation of the prevalence of non-B DNA secondary structure in previous assembly versions of the human genome and point to their importance being not fully appreciated. We anticipate that similar observations will occur as the improved sequencing technologies fill in gaps across the genomes of humans and other organisms.

1. Introduction

In the current and most complete reference assembly of the human genome (GRCh38), there are no chromosomes that are fully sequenced from telomere-to-telomere, meaning there are many gaps and missing regions that are not determined at the base sequence level. Recently, using ultra-long Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) high-fidelity (HiFi) data, the telomere-to-telomere sequence of human chromosome 8 was published: it is 146,259,670 bases in length, including 3,334,256 bases that are missing from GRCh38, with an estimated base accuracy in excess of 99.99% (Logsdon et al., 2021). The use of such advanced sequencing technologies mean that completion of the entire human genome sequence and the associated updated knowledge about complex genome-directed regulatory pathways is close.

Whereas DNA molecules mostly adopt double-stranded, right-handed B-helical conformations, a range of alternative (non-B) structures can also occur. It has been shown that such non-B DNA structures can play crucial roles in many basic cellular processes and can exist in vivo (Kouzine et al., 2019; Nakamura et al., 2017; Hänsel-Hertsch et al., 2020; Poggi and Richard, 2021). Alternative structures, such as G-quadruplexes (G4) and cruciforms, can be favoured under specific environmental (and cellular) conditions due to their improved thermodynamic stability compared to B-form DNA (Bowater et al., 2014; Matsumoto and Sugimoto, 2021). G4 structures regularly arise in G-rich regions of DNA sequence, mainly because the typical building block of a G4 is a G-quartet that emerges via Hoogsteen base pairing of four guanines. Two or more G-quartets stack on top of each other and are

Abbreviations: G4, G-quadruplexes; HiFi, high-fidelity; ONT, Oxford Nanopore Technologies; PacBio, Pacific Biosciences.

* Corresponding authors.
E-mail addresses: vaclav@ibp.cz (V. Brázda), R.Bowater@uea.ac.uk (R.P. Bowater).

https://doi.org/10.1016/j.gene.2021.146058
Received 31 August 2021; Received in revised form 14 October 2021; Accepted 29 October 2021
Available online 1 November 2021
0378-1119/© 2021 Elsevier B.V. All rights reserved.

physiologically stabilized by monovalent cations (Harkness and Mittermaier, 2017). Sites of inverted repeats are able to form hairpin stem-loop secondary structures in a single-stranded section of DNA or a cruciform structure if they occur on opposing strands of a double-stranded DNA (Brazda et al., 2020).

We decided to analyze the new telomere to telomere assembly of human chromosome 8 to evaluate the presence of sequences that have potential to form well characterised local non-B DNA structures, focusing on G4-forming and inverted repeat sequences. Potential G4 structures were identified by G4Hunter software, which calculates the G4Hunter score of sequences depending on the asymmetrical distribution of G and C bases on each strand (Brázda et al., 2019; Bedrat et al., 2016). The presence of inverted repeats was assessed by Palindrome analyser (Brázda et al., 2016). We reveal that the number of potential G4 and inverted repeats sequences is significantly increased in the new telomere to telomere assembly of human chromosome 8 compared to the same analyses of GRCh38. As we go on to highlight, these results suggest that the number of these types of regulatory sequences is strongly underestimated in the previous broadly used reference assemblies of the human genome.

2. Results

To compare the human chromosome 8 assemblies, we downloaded full chromosome sequences (NCBI ID CM000670.2 of GRCh38 human genome assembly and CP061028.1 of the telomere to telomere chromosome 8 assembly) and analyzed them. The fully assembled chromosome 8 sequence (156,259,670 bp) is 1,121,034 bp longer than the older chromosome 8 assembly within GRCh38 (145,138,636 bp), with a 1.21% increase in GC content. While the changes in total number of bases are below 1%, analyses by G4Hunter software revealed a remarkable increase in the abundance of G4-forming sequences (Table 1). In general, the higher the G4Hunter score, the more likely it is that a stable G4 will be formed. Excitingly, although the new genome assembly contained higher G4Hunter scores at all levels, the increase is most dramatic at the higher scores, which are those where there is greatest confidence that they can adopt a stable G4. For example, for G4Hunter scores up to 2.0 there is an increase of up to 10% in the G4-prone sequences in the new assembly, but for higher G4Hunter scores of 2.5 and 3 the increase is 13% and 25%, respectively. For a G4Hunter score of 3.5 there are more than twice as many G4-prone sequences in the new gapless assembly of chromosome 8 compared to GRCh38 (Table 1).

An even more obvious trend was observed with inverted repeats as analyzed by Palindrome finder (Table 2). The number of shorter inverted repeats (6–9 bases as the length of each half of the repeat) is similar in both assemblies (0.73% increase in new assembly) and corresponds well to the change of the corrected chromosome 8 length (0,77% increase). However, longer inverted repeats (with a size of >9 bases in each half of the repeat) are much more prevalent in the newly assembled chromosome 8 sequence than expected (1.90% increase for inverted repeats 10–14). Importantly, the length of the inverted repeats are connected to their proposed biological functions, with the highest abundance of inverted repeats immediately before transcription start sites being for repeats of 8 bp or longer (Brázda et al., 2020). The difference between the two assemblies increases with the length of the inverted repeat and there are almost twice as many inverted repeats with the repeat size of 30 in the previous newly assembled sequence compared to GRCh38 (Table 2). Such long inverted repeats are more prone to be a threat for genetic instability and recombination (Brázda et al., 2011; Wang and Leung, 2006).

Table 1
Comparison of the prevalence of G4-forming sequences in human chromosome 8. G4Hunter web applications with default parameters (window 25, G4Hunter score indicated in the first column) were used to identify G4-forming sequences in the current reference genome assembly ("GRCh 38") and the new telomere to telomere assembly of chromosome 8 ("chr8 new"). First column: G4Hunter score; second column: number of G4-forming sequences in chromosome 8 GRCh38 assembly; third column: number of G4-forming sequences in new chromosome 8 assembly; forth column: change in number of G4-forming sequences (new assembly-GRCh38), fifth column: difference in % (the number of G4-forming sequences in CRCh 38 was taken as 100%).

G4Hunter score	GRCh 38	new assembly	Δ	Δ%
1.2	232,871	237,055	+4184	**+1.80**
1.4	121,355	124,004	+2649	**+2.18**
1.6	64,699	66,638	+1939	**+3.00**
1.8	34,514	36,008	+1494	**+4.33**
2.0	18,368	19,526	+1158	**+6.30**
2.5	3746	4259	+513	**+13.69**
3.0	520	650	+130	**+25**
3.5	22	47	25	**+113.63**

Table 2
Comparison of the inverted repeats presence in human chromosome 8. Palindrome finder with default parameters (Length of one repeat unit is indicated in the first column) were used to identify inverted repeats in the current reference genome assembly ("GRCh 38") and the new telomere to telomere assembly of chromosome 8 ("chr8 new"). First column: Length of repeat unit; second column: number of inverted repeats in chromosome 8 GRCh38 assembly; third column: number of inverted repeats in new chromosome 8 assembly; fourth column: change in number of inverted repeats (new assembly-GRCh38); fifth column: difference in % (with the number of inverted repeats in GRCh 38 taken as 100%).

Inverted repeat size range	GRCh 38	new assembly	Δ	Δ%
6–9	5,023,663	5,060,277	+36,614	**+0.73**
10–14	114,281	116,455	+2174	**+1.90**
15–19	5914	6501	+587	**+9.93**
20–24	1670	2081	+411	**+24.61**
25–29	526	668	+142	**+27.00**
30–44	364	528	+164	**+45.05**
Over 45	33	78	+45	**+136.36**

Then we visualize the localization of G4-prone sequences and inverted repeats according to Giemsa banding (Speicher and Carter, 2005) of chromosome 8 (Fig. 1). This highlights the non-random enrichment of these non-B DNA structures in the fully assembled human chromosome 8 and points to the previous underestimation of non-B DNA structures in the human genome. It is important to appreciate that the sequences that are highly prone to adopt G4 structures have always been present within the human genome but they were not identified in genome databases due to incomplete sequence information.

We further investigated the localization of the highest-scoring G4-prone sequences (with a G4Hunter score above 3.5) and the longest inverted repeats (over 45 bp in the repeat unit) in their genomic context (Supplementary Material 01). The nearest *gene* and *repeat* that were identified in the complete sequence of chromosome 8 (Logsdon et al., 2021) were assigned to each G4-prone and inverted repeat sequence identified in this study. Interestingly, the majority of the G4-prone sequences (55%) and inverted repeats (59%) are present upstream of *gene* regions compared to downstream of the transcription start site. Moreover, all highest-scoring G4-prone sequences and the longest inverted repeats that are present within *genes* are located in introns.

3. Discussion

Although the current reference human genome (GRCh38) is the latest example as part of the process that continually improves it compared to the initial sequencing (Lander et al., 2001), several problems still persist (Ballouz et al., 2019). The new completed human chromosome 8 assembly sequence (published in April 2021) is 1.1 Mbp longer than GRCh38, but this comprehensive dataset is also more accurate about its number of open reading frames and regulatory regions (Logsdon et al., 2021). Whilst it is not surprising that completion of a

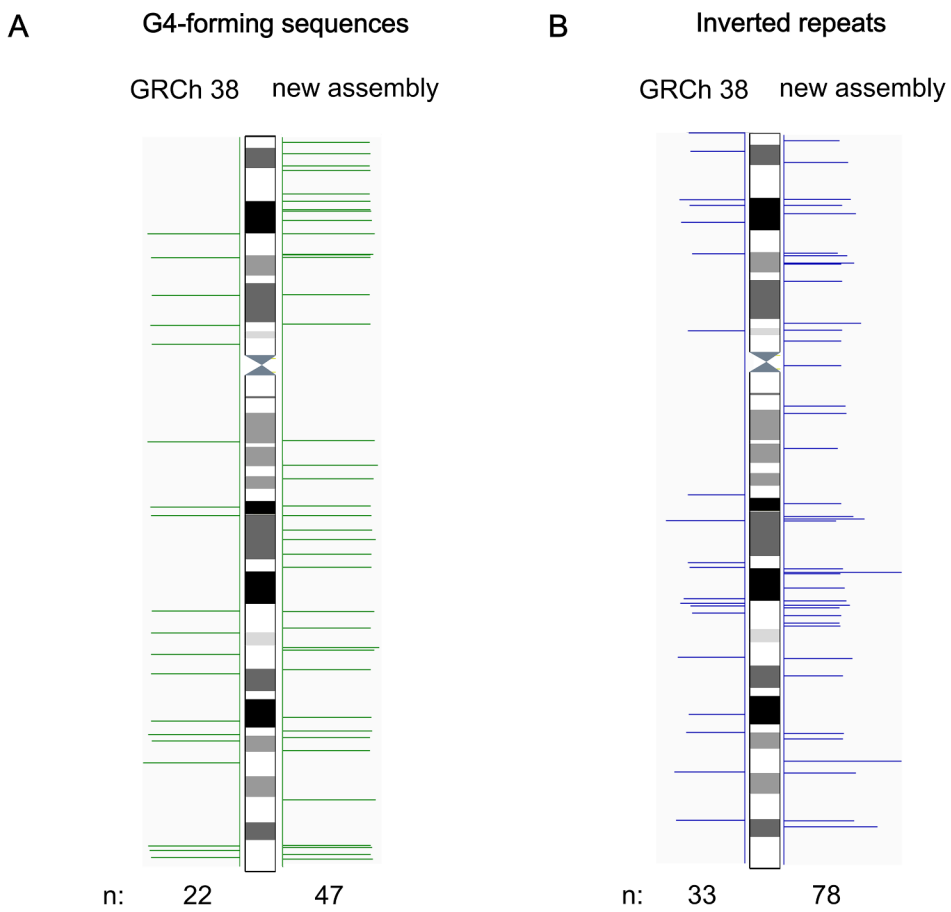


Fig. 1. Comparison of the localization of G4-forming sequences with a G4Hunter score threshold of 3.5 (A) and inverted repeats with an individual size for each half of the repeat to be 45 bases and higher (B) in human chromosome 8. The length of green dashes denotes the actual G4Hunter score, the blue dashes denotes the overall length of inverted repeat. For both parts, "n" reports the number of respective sequences found in each corresponding assembly. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

genome sequence will alter its overall size (either longer or shorter), the effects on its gene number and their expression are less predictable. For example, the completed genomes can have more or less genes (open reading frames), and even shorter genomic sequences may be able to express more genes due to reorganisation of the bases. Altered arrangement of the sequence can also lead to altered propensity to form local non-B DNA structures, which can have effects on the numbers of regulatory regions, such as promoters and terminators (Brazda et al., 2020). Contemporary sequencing methods with long-reads and combinations of several sequencing methods allow previously unimaginable progress in accuracy of genome assemblies. This has dramatic impacts on some types of sequences, particularly those that are repetitive in nature. Previously, many repetitive sequences were missed entirely because the sequencing methods did not allow their detection, which led to "gaps" in the genome assembly. Furthermore, even when repetitive sequences were detected by sequencing methods they were often deleted from assemblies due to the impossibility of determining their exact location in the DNA molecule.

Recent advances in DNA sequencing technologies allow determination of the complete information without any gaps and ambiguities (Phillippy, 2017). GRCh38 is the most widely used reference assembly of the human genome, but it has 370,500 unknown (N) bases in the sequence of chromosome 8, which point to many uncertainties with its usage and it should be replaced by the fully assembled sequence in the near future. As we demonstrate, the new human chromosome 8 telomere-to-telomere assembly fixes obvious problems with the reference genome and it also brings many unexpected findings like the strong underestimation of non-B DNA structures in the human genome, as demonstrated here for chromosome 8. Local non-B DNA structures are involved in the regulation of basic molecular processes (Brázda et al., 2021; Guiblet et al., 2021). Bioinformatics studies have demonstrated that G4-forming sequences are enriched in promoters (Brázda et al., 2021) and other regulatory regions of the human genome, such as untranslated regions (UTRs) and telomeres (Kharel et al., 2020; Wu et al., 2021; Nakanishi and Seimiya, 2020). Inverted repeats have been detected in human promoters and represent targets for regulatory proteins (Brázda et al., 2020; Brázda and Coufal, 2017; Fleming et al., 2020) and the impact of G4 structural variations on gene activity was also shown (Gong et al., 2021). As highlighted in Fig. 2, it has been demonstrated that these non-B DNA structures play a role in replication, transcription, translation, genomic stability, and telomere maintenance (Brázda et al., 2011; Ravichandran et al., 2019; Spiegel et al., 2020; De Magis et al., 2019). Data included in Supplementary Material 01 demonstrates that the most stable of the observed G4-prone sequences (with a G4Hunter score above 3.5) and inverted repeats (over 45 bp in the repeat unit) are located before genes or in introns. An important potential conclusion from this is that if such sequences form stable non-B DNA structures then they could impair transcription machinery or have a role in alternative splicing (Gomez et al., 2004; Marcel et al., 2011). Moreover, the majority of these observed G4-prone and inverted repeat sequences are located within "*repeat*" regions, such as "*simple repeats*", which are known to be regulatory elements of gene expression (Brazda et al., 2020). Together, these observations warrant further investigation to the potential effects of non-B DNA structures on gene expression. An increasing number of studies have reported on the cellular significance of these non-B DNA structures, especially for G4 due to their possible utilization in antiviral and antiproliferative therapeutic targeting and because of their great stability under physiological conditions (Puig Lombardi et al., 2019; Ruggiero and Richter, 2020; Asamitsu et al., 2019). Diverse roles of G4 could be the result of a synergic effect with other structurally compatible non-B DNA structures, such as an R loop, which is formed by one DNA strand annealed to an

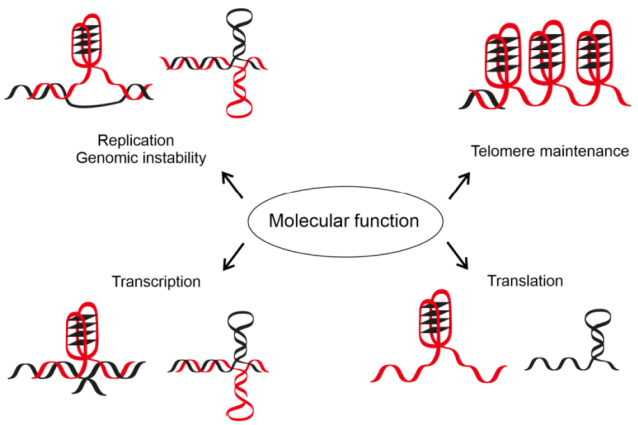


Fig. 2. Non-B DNA structures in the regulation of basic molecular processes. Both G4 and cruciforms play a role in replication initiation, but could also impair replication machinery and thus be a source of genomic instability (Prioleau, 2017; Wang et al., 2019; Inagaki et al., 2009; Takahashi and Sugimoto, 2019). Stable G4s form in human telomere sequences and are involved in the protection of telomere ends (Ravichandran et al., 2019; Jansson et al., 2019). G4 or cruciform formation in promoter regions could modulate the initiation of transcription (Brázda et al., 2021). G4 or hairpin formation in mRNA could affect mRNA stability and prematurely terminate translation (Brázda et al., 2011; Kamura et al., 2020).

RNA, thus forming a DNA:RNA hybrid (De Magis et al., 2019).

In summary, the exciting findings from the recently published telomere to telomere assembly of chromosome 8 uncovered an increased abundance of G4-forming sequences and inverted repeats compared to the current reference genome sequence (GRCh38). We found significantly more putative G4-forming sequences and inverted repeats in the newly assembled sequence, which strongly highlights the previous underestimation of the prevalence of potential non-B secondary structures in the human genome. Similar findings appear to be valid also for a new telomere to telomere assembly of chromosome X (Miga et al., 2020), as demonstrated for G4-prone sequences (Bohálová et al., 2021). These potential regulatory sequences within newly-assembled genomes should be evaluated using methods that allow targeted detection of the non-B DNA structures, particularly within cells (Tassinari et al., 2020; Di Antonio et al., 2020). This new complete assembly of human chromosome 8 highlights the exciting findings that await as we approach the accurate reading and interpretation of the full human genome with its broad variety of regulatory sequences, including local non-B DNA structures, such as G4-forming sequences and inverted repeats.

4. Materials and methods

To compare the human chromosome 8 assemblies, we downloaded full chromosome sequences (NCBI ID CM000670.2 of GRCh38 human genome assembly and CP061028.1 of the telomere to telomere chromosome 8 assembly) and analyzed them by Palindrome finder and G4Hunter web applications with default parameters (Brázda et al., 2019; Brázda et al., 2016). *Gene* and *repeat* annotations were provided by the authors of the original paper that described the telomere to telomere assembly of chromosome 8 (Logsdon et al., 2021).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by The Czech Science Foundation (No. 18-15548S) to VB.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2021.146058.

References

Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., de Lima, L.G., Dvorkina, T., Porubsky, D., Harvey, W.T., Mikheenko, A., Bzikadze, A.V., Kremitzki, M., Graves-Lindsay, T.A., Jain, C., Hoekzema, K., Murali, S.C., Munson, K.M., Baker, C., Sorensen, M., Lewis, A.M., Surti, U., Gerton, J.L., Larionov, V., Ventura, M., Miga, K.H., Phillippy, A.M., Eichler, E.E., 2021. The structure, function and evolution of a complete human chromosome 8. Nature 593 (7857), 101–107. https://doi.org/10.1038/s41586-021-03420-7.

Kouzine, F., Wojtowicz, D., Yamane, A., Casellas, R., Przytycka, T.M., Levens, D.L., 2019. In Vivo Chemical Probing for G-Quadruplex Formation. Methods Mol Biol. 2035, 369–382. https://doi.org/10.1007/978-1-4939-9666-7_23.

Nakamura, T., Okabe, S., Yoshida, H., Iida, K., Ma, Y., Sasaki, S., Yamori, T., Shin-ya, K., Nakano, I., Nagasawa, K., Seimiya, H., 2017. Targeting glioma stem cells in vivo by a G-quadruplex-stabilizing synthetic macrocyclic hexaoxazole. Sci Rep. 7, 3605. https://doi.org/10.1038/s41598-017-03785-8.

Hänsel-Hertsch, R., Simeone, A., Shea, A., Hui, W.W.I., Zyner, K.G., Marsico, G., Rueda, O.M., Bruna, A., Martin, A., Zhang, X., Adhikari, S., Tannahill, D., Caldas, C., Balasubramanian, S., 2020. Landscape of G-quadruplex DNA structural regions in breast cancer. Nat Genet. 52 (9), 878–883. https://doi.org/10.1038/s41588-020-0672-8.

Poggi, L., Richard, G.-F., 2021. Alternative DNA Structures In Vivo: Molecular Evidence and Remaining Questions. Microbiol Mol Biol Rev. 85, e00110–20. https://doi.org/10.1128/MMBR.00110-20.

R.P. Bowater, Z.A. Waller, DNA Structure, in: ELS, John Wiley & Sons, Ltd: Chichester, 2014. 10.1002/9780470015902.a0006002.pub2.

Matsumoto, S., Sugimoto, N., 2021. New Insights into the Functions of Nucleic Acids Controlled by Cellular Microenvironments. Top Curr Chem (Cham). 379, 17. https://doi.org/10.1007/s41061-021-00329-7.

Harkness, R.W., Mittermaier, A.K., 2017. G-quadruplex dynamics. Biochim Biophys Acta Proteins Proteom. 1865 (11), 1544–1554. https://doi.org/10.1016/j.bbapap.2017.06.012.

Brazda, V., Fojta, M., Bowater, R.P., 2020. Structures and stability of simple DNA repeats from bacteria. Biochem J. 477, 325–339. https://doi.org/10.1042/BCJ20190703.

Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Šťastný, J., Mergny, J.-L., 2019. G4Hunter web application: a web server for G-quadruplex prediction. Bioinformatics 35, 3493–3495. https://doi.org/10.1093/bioinformatics/btz087.

Bedrat, A., Lacroix, L., Mergny, J.-L., 2016. Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. 44 (4), 1746–1759. https://doi.org/10.1093/nar/gkw006.

Brázda, V., Kolomazník, J., Lýsek, J., Hároníková, L., Coufal, J., Šťastný,, J., 2016. Palindrome analyser – A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. Biochem. Biophys. Res. Commun. 478, 1739–1745. https://doi.org/10.1016/j.bbrc.2016.09.015.

Brázda, Václav, Bartas, M., Lýsek, J., Coufal, J., Fojta, M., 2020. Global analysis of inverted repeat sequences in human gene promoters reveals their non-random distribution and association with specific biological pathways. Genomics 112 (4), 2772–2777. https://doi.org/10.1016/j.ygeno.2020.03.014.

Brázda, V., Laister, R.C., Jagelská, E.B., Arrowsmith, C., 2011. Cruciform structures are a common DNA feature important for regulating biological processes. BMC Mol Biol. 12, 33. https://doi.org/10.1186/1471-2199-12-33.

Wang, Y., Leung, F.C.C., 2006. Long inverted repeats in eukaryotic genomes: Recombinogenic motifs determine genomic plasticity. FEBS Lett. 580, 1277–1284. https://doi.org/10.1016/j.febslet.2006.01.045.

Speicher, M.R., Carter, N.P., 2005. The new cytogenetics: blurring the boundaries with molecular biology. Nat Rev Genet. 6, 782–792. https://doi.org/10.1038/nrg1692.

E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.-F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R.A. Gibbs, D.M. Muzny, S.E.

Scherer, J.B. Bouck, E.J. Sodergren, K.C. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R.S. Kucherlapati, D.L. Nelson, G.M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R.W. Davis, N.A. Federspiel, A.P. Abola, M.J. Proctor, B.A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W.R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.-C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G.R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F.A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K.H. Wolfe, S.-P. Yang, R.-F. Yeh, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, R.M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M.V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, A. Patrinos, M.J. Morgan, International Human Genome Sequencing Consortium, C. for G.R. Whitehead Institute for Biomedical Research, The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, I. of M.B. Department of Genome Analysis, GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, T.I. for S.B. Multimegabase Sequencing Center, Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, L.A.H. G.C. Cold Spring Harbor Laboratory, GBF—German Research Centre for Biotechnology:, also includes individuals listed under other headings): *Genome Analysis Group (listed in alphabetical order, U.N.I. of H. Scientific management: National Human Genome Research Institute, Stanford Human Genome Center:, University of Washington Genome Center:, K.U.S. of M. Department of Molecular Biology, University of Texas Southwestern Medical Center at Dallas:, U.D. of E. Office of Science, The Wellcome Trust:, Initial sequencing and analysis of the human genome, Nature. 409 (2001) 860–921. 10.1038/35057062.

Ballouz, S., Dobin, A., Gillis, J.A., 2019. Is it time to change the reference genome? Genome Biol. 20, 159. https://doi.org/10.1186/s13059-019-1774-4.

A.M. Phillippy, New advances in sequence assembly, Genome Res. 27 (2017) xi–xiii. 10.1101/gr.223057.117.

Brázda, Václav, Bartas, M., Bowater, R.P., 2021. Evolution of Diverse Strategies for Promoter Regulation. Trends Genet. 37 (8), 730–744. https://doi.org/10.1016/j.tig.2021.04.003.

Guiblet, W.M., Cremona, M.A., Harris, R.S., Chen, D., Eckert, K.A., Chiaromonte, F., Huang, Y.-F., Makova, K.D., 2021. Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. Nucleic Acids Res. 49, 1497–1516. https://doi.org/10.1093/nar/gkaa1269.

Kharel, P., Becker, G., Tsvetkov, V., Ivanov, P., 2020. Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back. Nucleic Acids Res. 48, 12534–12555. https://doi.org/10.1093/nar/gkaa1126.

Wu, F., Niu, K., Cui, Y., Li, C., Lyu, M., Ren, Y., Chen, Y., Deng, H., Huang, L., Zheng, S., Liu, L., Wang, J., Song, Q., Xiang, H., Feng, Q., 2021. Genome-wide analysis of DNA G-quadruplex motifs across 37 species provides insights into G4 evolution. Communications Biology. 4, 1–11. https://doi.org/10.1038/s42003-020-01643-4.

Nakanishi, C., Seimiya, H., 2020. G-quadruplex in cancer biology and drug discovery. Biochem. Biophys. Res. Commun. 531 (1), 45–50. https://doi.org/10.1016/j.bbrc.2020.03.178.

Brázda, V., Coufal, J., 2017. Recognition of Local DNA Structures by p53 Protein. Int J Mol Sci. 18, 375. https://doi.org/10.3390/ijms18020375.

Fleming, A.M., Zhu, J., Jara-Espejo, M., Burrows, C.J., 2020. Cruciform DNA Sequences in Gene Promoters Can Impact Transcription upon Oxidative Modification of 2′-Deoxyguanosine. Biochemistry 59 (28), 2616–2626. https://doi.org/10.1021/acs.biochem.0c0038710.1021/acs.biochem.0c00387.s001.

Gong, J.-Y., Wen, C.-J., Tang, M.-L., Duan, R.-F., Chen, J.-N., Zhang, J.-Y., Zheng, K.-W., He, Y., Hao, Y.-H., Yu, Q., Ren, S.-P., Tan, Z., 2021. G-quadruplex structural variations in human genome associated with single-nucleotide variations and their impact on gene activity. Proc Natl Acad Sci U S A. 118 https://doi.org/10.1073/pnas.2013230118.

Ravichandran, S., Ahn, J.-H., Kim, K.K., 2019. Unraveling the Regulatory G-Quadruplex Puzzle: Lessons From Genome and Transcriptome-Wide Studies. Front Genet. 10, 1002. https://doi.org/10.3389/fgene.2019.01002.

Spiegel, J., Adhikari, S., Balasubramanian, S., 2020. The Structure and Function of DNA G-Quadruplexes. TRECHEM. 2, 123–136. https://doi.org/10.1016/j.trechm.2019.07.002.

De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O., Capranico, G., 2019. DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. Proc Natl Acad Sci U S A. 116 (3), 816–825. https://doi.org/10.1073/pnas.1810409116.

Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J.-L., Riou, J.-F., 2004. Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. Nucleic Acids Res. 32, 371–379. https://doi.org/10.1093/nar/gkh181.

Marcel, V., Tran, P.L.T., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.-P., Hall, J., Mergny, J.-L., Hainaut, P., Van Dyck, E., 2011. G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. Carcinogenesis 32, 271–278. https://doi.org/10.1093/carcin/bgq253.

E.P. Lombardi, Allyson Holmes, D. Verga, M.-P. Teulade-Fichou, A. Nicolas, A. Londoño-Vallejo, Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species, Nucleic Acids Res. 47 (2019) 6098–6113. 10.1093/nar/gkz463.

Ruggiero, E., Richter, S.N., 2020. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. Annu Rep Med Chem. 54, 101–131. https://doi.org/10.1016/bs.armc.2020.04.001.

Asamitsu, S., Obata, S., Yu, Z., Bando, T., Sugiyama, H., 2019. Recent Progress of Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy. Molecules 24. https://doi.org/10.3390/molecules24030429.

Prioleau, M.-N., 2017. G-Quadruplexes and DNA Replication Origins. Adv Exp Med Biol. 1042, 273–286. https://doi.org/10.1007/978-981-10-6955-0_13.

Wang, Y., Yang, J., Wild, A.T., Wu, W.H., Shah, R., Danussi, C., Riggins, G.J., Kannan, K., Sulman, E.P., Chan, T.A., Huse, J.T., 2019. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nat Commun. 10, 943. https://doi.org/10.1038/s41467-019-08905-8.

Inagaki, H., Ohye, T., Kogo, H., Kato, T., Bolor, H., Taniguchi, M., Shaikh, T.H., Emanuel, B.S., Kurahashi, H., 2009. Chromosomal instability mediated by non-B DNA: Cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. Genome Res. 19, 191–198. https://doi.org/10.1101/gr.079244.108.

Takahashi, S., Sugimoto, N., 2019. Quantitative Analysis of Stall of Replicating DNA Polymerase by G-Quadruplex Formation. Methods Mol Biol. 2035, 257–274. https://doi.org/10.1007/978-1-4939-9666-7_15.

Jansson, L.I., Hentschel, J., Parks, J.W., Chang, T.R., Lu, C., Baral, R., Bagshaw, C.R., Stone, M.D., 2019. Telomere DNA G-quadruplex folding within actively extending human telomerase. PNAS 116 (19), 9350–9359. https://doi.org/10.1073/pnas.1814777116.

Kamura, T., Katsuda, Y., Kitamura, Y., Ihara, T., 2020. G-quadruplexes in mRNA: A key structure for biological function. Biochem. Biophys. Res. Commun. 526 (1), 261–266. https://doi.org/10.1016/j.bbrc.2020.02.168.

Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., Schneider, V.A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., Markovic, C., Maduro, V., Dutra, A., Bouffard, G.G., Chang, A.M., Hansen, N.F., Wilfert, A.B., Thibaud-Nissen, F., Schmitt, A.D., Belton, J.-M., Selvaraj, S., Dennis, M. Y., Soto, D.C., Sahasrabudhe, R., Kaya, G., Quick, J., Loman, N.J., Holmes, N., Loose, M., Surti, U., Risques, R.ana., Graves Lindsay, T.A., Fulton, R., Hall, I., Paten, B., Howe, K., Timp, W., Young, A., Mullikin, J.C., Pevzner, P.A., Gerton, J.L., Sullivan, B.A., Eichler, E.E., Phillippy, A.M., 2020. Telomere-to-telomere assembly of a complete human X chromosome. Nature 585 (7823), 79–84. https://doi.org/10.1038/s41586-020-2547-7.

Bohálová, N., Mergny, J.-L., Brázda, V., 2021. Novel G-quadruplex prone sequences emerge in the complete assembly of the human X chromosome. Biochimie 191, 87–90. https://doi.org/10.1016/j.biochi.2021.09.004.

Tassinari, M., Zuffo, M., Nadai, M., Pirota, V., Sevilla Montalvo, A.C., Doria, F., Freccero, M., Richter, S.N., 2020. Selective targeting of mutually exclusive DNA G-quadruplexes: HIV-1 LTR as paradigmatic model. Nucleic Acids Res. 48, 4627–4642. https://doi.org/10.1093/nar/gkaa186.

Di Antonio, M., Ponjavic, A., Radzevičius, A., Ranasinghe, R.T., Catalano, M., Zhang, X., Shen, J., Needham, L.-M., Lee, S.F., Klenerman, D., Balasubramanian, S., 2020. Single-molecule visualization of DNA G-quadruplex formation in live cells. Nat. Chem. 12 (9), 832–837. https://doi.org/10.1038/s41557-020-0506-4.

Příloha F

Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection

Biochimie 186 (2021) 13–27

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection

Natália Bohálová [a, b, 1], Alessio Cantara [a, b, 1], Martin Bartas [c], Patrik Kaura [d], Jiří Šťastný [d, e], Petr Pečinka [c], Miroslav Fojta [a], Jean-Louis Mergny [a, f], Václav Brázda [a, *]

[a] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, Brno, 612 65, Czech Republic
[b] Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 62500, Brno, Czech Republic
[c] Department of Biology and Ecology/Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Ostrava, 710 00, Czech Republic
[d] Brno University of Technology, Faculty of Mechanical Engineering, Technická 2896/2, 616 69, Brno, Czech Republic
[e] Department of Informatics, Mendel University in Brno, Zemědělská 1, Brno, 613 00, Czech Republic
[f] Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, 91128, Palaiseau, France

ARTICLE INFO

Article history:
Received 18 January 2021
Received in revised form
30 March 2021
Accepted 31 March 2021
Available online 9 April 2021

Keywords:
G-quadruplex
Viral genome
Bioinformatics
Persistent infection
Acute infection
G4Hunter

ABSTRACT

G-quadruplexes contribute to the regulation of key molecular processes. Their utilization for antiviral therapy is an emerging field of contemporary research. Here we present comprehensive analyses of the presence and localization of putative G-quadruplex forming sequences (PQS) in all viral genomes currently available in the NCBI database (including subviral agents). The G4Hunter algorithm was applied to a pool of 11,000 accessible viral genomes representing 350 Mbp in total. PQS frequencies differ across evolutionary groups of viruses, and are enriched in repeats, replication origins, 5′UTRs and 3′UTRs. Importantly, PQS presence and localization is connected to viral lifecycles and corresponds to the type of viral infection rather than to nucleic acid type; while viruses routinely causing persistent infections in Metazoa hosts are enriched for PQS, viruses causing acute infections are significantly depleted for PQS. The unique localization of PQS identifies the importance of G-quadruplex-based regulation of viral replication and life cycle, providing a tool for potential therapeutic targeting.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

G-quadruplexes (G4s) are stacked secondary nucleic acid structures formed by G-rich DNA or RNA sequences. The building block of a G4 is a G-quartet that arises via Hoogsteen base pairing of four guanines. Two or more G-quartets are stacked on top of each other and physiologically stabilized mainly by monovalent cations, such as potassium [1]. G4s have been found in various genome locations such as telomeres, origins of replication and promoter regions [2,3]. With respect to their presence in these particular areas of genomes, G4s are most likely involved in different cellular events such as modulation of telomeric functions or regulation of gene expression [2,4].

In recent years, increasing evidence showed that G4s are also involved in the control of key viral processes, such as replication or transcription and influence viral latency [5]. Stabilization of G4 by specific G4 ligands, such as bisquinolinium derivatives (PhenDC3), porphyrin derivatives, perylenes and naphthalene diimides (PIPER), pyridostatin (PDS) or acridine derivatives (*e.g.*, BRACO-19) can modulate viral gene expression and, moreover, viral infection [6]. Therefore, G4 ligands can be considered as promising anti-proliferative agents [7] and also as potential antiviral therapeutic compounds [6].

Under physiological conditions, G4 formation is regulated by G4-binding proteins [8]. This suggests their crucial roles in a number of signaling pathways, including those relevant for viral infection (*e.g.*, the immune response). For example nucleolin, the most abundant nucleolar phosphoprotein known to promote the formation of G4 and inhibit the activity of the *c-Myc* gene [9], has also been reported to stabilize the formation of G4 in LTR repeats of the human immunodeficiency virus 1 (HIV-1) virus and thus silence viral transcription [10]. Nucleolin also mediates immune evasion of Epstein-Barr virus (EBV) by stabilizing G4 in EBV-encoded nuclear antigen-1 (EBNA1) mRNA [11] and suppresses replication of Hepatitis C Virus (HCV) by binding to RNA G4 [12].

* Corresponding author.
E-mail address: vaclav@ibp.cz (V. Brázda).
[1] These authors contributed equally.

https://doi.org/10.1016/j.biochi.2021.03.017
0300-9084/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Furthermore, viral G4-binding proteins are crucial for the viral life cycle; for example nucleocapsid proteins of the HIV-1 virus can function as molecular chaperones for G4 [13] and it the G4-binding macrodomain within the SARS unique domain (SUD) of SARS-CoV Nsp3 protein is essential for viral replication and transcription [14]. Targeting of viral G4 binding proteins is thus a prospective antiproliferative and antiviral therapeutic strategy. For this purpose, antiviral DNA aptamers have been designed [15].

To assess the propensity of a sequence to form G4 and distribution of these putative G4-forming sequences (PQS) in genomes, several algorithms have already been described. Chronologically, the first generation algorithm searches for the [GnNmGnNoGnNpGn] motif in a sequence [16]. The second generation algorithm takes into consideration occurrence of repeating units of Gn ($n \geq 2$) [17]. Nevertheless, both algorithms produce binary (yes/no; match/no match) results, rather than the qualitative analyses that are necessary for correlation of G4 strength metrics. G4Hunter was developed to overcome this limitation, and it calculates G4 propensity depending on the asymmetrical distribution of G and C (GC-skewness) [18].

Our aim was to analyze for the first time a pool of all accessible 11,000 viral genomes searching for the presence of PQS with the G4Hunter web application [19]. Viruses are small infectious agents that replicate inside a living cell of an organism. According to the Baltimore classification they can be divided into classes according to their genome: double-stranded DNA (dsDNA) viruses, single-stranded DNA (ssDNA) viruses, double-stranded RNA (dsRNA) viruses, positive sense RNA (ssRNA+) viruses, negative stranded RNA (ssRNA-) viruses, RNA reverse transcribing viruses and DNA reverse transcribing viruses. In addition, there are also subviral agents such as viroids and satellites. Viroids are small circular ssRNA molecules, which do not encode any protein. All known viroids infect higher plants and represent dangerous agricultural pathogens [20]. Satellites are also mainly found in plants; their replication depends on co-infection with an independent helper virus [21]. Evidence of the presence of PQS has already been reported in several viruses from different families; for example dsDNA viruses from *Herpesviridae* or *Papillomaviridae*, ssRNA+ viruses from *Flaviviridae* or *Filoviridae*, RNA viruses with reverse transcriptase (RT) *Retroviridae* and also dsDNA with RT *Hepadnaviridae* [6,22]. The presence of PQS and G4 functions in viruses remains poorly understood and until now has been mainly limited to studies of eukaryotic viruses by first generation algorithms [23,24]. In our study, we comprehensively analyzed the presence and locations of PQS in all accessible viral genomes with diverse hosts using G4Hunter. For classification purposes, we categorized viruses according to their types of genome and current family assignment in the National Center for Biotechnology Information (NCBI) database. These data bring basic information about evolutionary changes of PQS frequency among the phyla and for the first time link G4 with the type of viral infection and thus provide possible new strategies for their targeting in therapeutics.

2. Materials and methods

2.1. Selection of viral sequences

All 13,809 accessible viral genomic sequences of 11,000 unique species were downloaded from the genome database of the NCBI on the 10th of January 2020. The phylogenetic classification was performed according to NCBI and Virus-Host database [25]. The accession codes and phylogenetic classification of all viral complete genomic sequences are shown in Supplementary Material 1.

2.2. Process of analysis

We used the computational core of our DNA analyzer software written in Java with G4Hunter algorithm implementation [19]. Parameters for G4Hunter were set to 25 nucleotides for window size and a threshold score of 1.2 or above. A separate list of PQS in each of the sequences and an overall report were obtained. The overall results for each species group contained a list of species with size of genomic DNA, GC content, number of PQS found, frequency normalized for 1,000 nt and length of the sequence covered with PQS (Supplementary Material 2). PQS were also grouped depending on their G4Hunter score in the five following intervals: 1.2–1.4, 1.4–1.6, 1.6–1.8, 1.8–2.0 and 2.0 or more (the higher the score, the more likely/more stable is the G4).

2.3. Statistical analysis

Statistical evaluations of differences in PQS between phylogenetic groups, viruses grouped by the type of infection and features were made by Kruskal–Wallis test. Post-hoc multiple pairwise comparison by Dunn's test with Bonferroni correction of the significance level was applied with *p*-value cut-off 0.05. Data are available in Supplementary Material 3. A cluster dendrogram of PQS characteristics was constructed in the program R, version 3.6.3, using the pvclust package to further reveal and graphically depict similarities between particular viral families. The following values were used as input data: Mean f (mean of predicted PQS per 1,000 nt), Min f (the lowest frequency of predicted PQS per 1,000 nt), Max f (the highest frequency of predicted PQS per 1,000 nt) and Cov % (% of the genome covered by PQS with a threshold of 1.2 or more) (Supplementary Material 4). The following parameters were used for both analyses: cluster method 'ward.D2', distance 'euclidean', number of bootstrap resampling was 10,000. Statistically significant clusters (based on approximately unbiased values above 95, equivalent to p-values less than 0.05) are highlighted by rectangles marked with broken red lines. R code is provided in Supplementary Material 4. Receiver operating characteristic (ROC) curves were constructed to determine the area under the curve (AUC) for genome groups and viral families causing either persistent or acute type of infection through web-based ROC [26]. Correlation was assessed by non-parametric Spearman's correlation coefficient. Normality of datasets was tested by d'Agostino-Pearson test.

2.4. Analysis of PQS around annotated NCBI features

Accessible features tables of all viral genomes were downloaded from the NCBI database. Features were grouped by their name stated in the feature table file. We analyzed PQS occurrence inside and around (before and after) predefined featured neighborhood (±100 nt). From this analysis, we obtained a file with feature names and number of PQS found inside and around features. Further processing was performed in Microsoft Excel and the data are available as Supplementary Material 5.

3. Results

3.1. Variation in PQS frequency in viruses and subviral agents

According to NCBI taxonomy classification, the analyzed genomes were divided into three main categories, viruses (10,613 genomes), satellites (321 genomes) and viroids (66 genomes) (Supplementary Material 6). Viruses were further divided according to their nucleic acid type into 7 groups (3,787 genomes of dsDNA viruses; 1,736 ssDNA; 438 dsRNA; 2,538 ssRNA+ viruses;

631 ssRNA-viruses; 132 ssRNA+ viruses with RT; 131 dsDNA viruses with RT) [27,28]. These groups were divided into 56 families as shown in the phylogenetic tree (Supplementary Material 6). Satellites were categorized into ssDNA (272 genomes) and ssRNA satellites (27 genomes). For further statistical analyses, only categories with 20 or more sequenced genomes were evaluated. Detailed phylogenetic classification for every analyzed virus and corresponding host is presented in Supplementary Material 1.

We analyzed the presence of PQS by G4Hunter in a pool of 13,809 sequences belonging to 11,000 unique viruses. All accessible genotypes and segments of each virus were considered. The detailed overall data are available in Supplementary Material 2. The smallest accessible complete genome was *Rice yellow mottle virus satellite* with the length of 220 nt, the longest one *Pandoravirus salinus*, a giant virus belonging to dsDNA viruses with a genome size of 2,473,870 bp [29]. The statistical review of results is shown in Table 1. The average GC content in all analyzed sequences was 45.29%; however, the GC content varied remarkably among the groups and families (Table 1) with a maximum of 57.98% GCs in *Pospiviroidae* (viroids) and a minimum of 33.72% GCs in *Poxviridae* (dsDNA viruses). For the whole pool of viruses, the mean PQS frequency with the default parameters of the analyses (window of 25 nt and score threshold of 1.2) was 1.13 per 1,000 nt and PQS cover on average 2.8% of the genomes. In comparison, the analysis of the human genome with the same parameters shows a PQS frequency of 1.71 with an average PQS coverage of 4.8%. The lowest frequency among the group category was found for ssRNA-viruses (0.67), followed by ssDNA satellites (0.96) and dsDNA viruses with RT (1.06). The highest frequency among the groups was observed for ssRNA+ viruses with RT (3.55) and ssRNA satellites (2.19). At the family level, the lowest mean PQS frequency was found in *Coronaviridae* (0.10) belonging to ssRNA+ viruses and *Poxviridae* (0.22) belonging to the dsDNA viruses and the highest for plant-infecting ssRNA+ viruses from the *Tymoviridae* (8.69) family, *Anelloviridae* (3.57) belonging to the ssDNA viruses and the *Retroviridae* family of ssRNA viruses with RT, where HIV-1 belongs (3.55). Detailed statistical analyses of inter group and inter family comparisons are provided in Supplementary Material 3.

Detailed statistical characteristics for PQS frequencies per 1,000 nt (including mean, variance, outliers) are depicted in boxplots for all inspected viral families (Fig. 1). The highest heterogeneity was found within the ssRNA+ viruses, where frequencies differed remarkably from 0.1 for the family of *Coronaviridae*; a family of important human pathogens to 8.69 inside the *Tymoviridae* family (*i.e.*, a 87-fold difference in PQS density). Within the ssDNA group, the highest PQS frequency was found for *Anelloviridae* (3.57), the family chronically infecting humans without known link to any disease [28] and the lowest for *Microviridae* (0.24), a family of bacteria-infecting viruses. A higher PQS frequency was observed for ssRNA+ viruses compared to ssRNA-ones. Two subgroups of dsDNA viruses with RT differed significantly: for *Caulimoviridae,* the plant infecting family of dsDNA viruses with RT, we observed a mean frequency of 0.44 PQS per 1,000 nt, while it was 1.75 for *Hepadnaviridae,* dsDNA viruses with RT to which hepatitis B virus (HBV) belongs.

To further demonstrate and graphically depict similarities among individual viral families we created a cluster dendrogram (Fig. 2). Input data and R code are provided in Supplementary Material 4. Interestingly, the families did not cluster according to their type of nucleic acid, as could be expected, but rather according to other phenotypic characteristics such as the dominant type of infection in Metazoa host. The significance of this clustering was confirmed by ROC analyses; while AUC values found for families causing either persistent or acute type of infection were set to 0.93 and 0.87 respectively, for genome groups they differed from 0.50 to 0.63 (Supplementary Material 4). Two evolutionary distinct families that typically cause persistent types of infection in *Metazoa* hosts, *Herpesviridae* (dsDNA) and *Retroviridae* (ssRNA with RT) [30–32], clustered together on the right side of the dendrogram with statistically significant similarity, whereas *Poxviridae* (dsDNA) clustered on the left side with another acutely infecting family *Coronaviridae* (ssRNA+) [33,34].

The virus with the highest PQS frequency (23.8 per 1,000 nt which translates into one PQS every 42 nt) was the *Grapevine fleck virus* with a GC content of 66.23%. This virus belongs to the *Maculavirus* genus in the order *Tymovirales*, in the family *Tymoviridae*. The RNA *Grapevine fleck virus* genome is linear, 7,564-nt long, and replicated by a positive stranded RNA virus replication model. Strikingly, PQS cover 66.17% of its genome when using a default threshold of 1.2 and 50.91% using a more stringent threshold of 1.4 (Fig. 3). In other words, the majority of the genome of this virus may potentially adopt a G4 fold. All PQS found in *Grapevine fleck virus* are located on the G-rich negative strand. ssRNA viruses are often non-symmetrically distributed with guanines and cytosines; in this case, guanines represent almost 50% of all bases on the negative intermediate. Moreover, PQS found in the genome of *Grapevine fleck virus* are remarkably long, with one PQS identified with a scanning window of 1,000 nt! Interestingly, this extreme density does not simply result from a high GC content, as viruses with an even higher proportion of guanines and cytosines, such as *Papiine alphaherpesvirus* (76.09%) have a lower PQS coverage (32.84%). This example again illustrates that the GC content is not strongly correlated with G4 propensity, as shown in other organisms such as Archaea [35]: in the *Grapevine fleck virus* this extremely unusual density in PQS is also the result a strong GC-skewness. Spearman's correlation coefficient for PQS frequency and GC content for viral sequences showed only a moderate positive correlation with a value of 0.65 (p-value < 0.001). After the G4Hunter analyses, the results were clustered according to the following intervals of G4Hunter score: 1.2–1.4, 1.4–1.6, 1.6–1.8, 1.8–2.0 and above 2.0 (Supplementary Material 7). PQS with higher G4Hunter score exhibit a higher probability of G4 formation [18,19]. The graphic representation of the GC contents and PQS frequencies is shown in Fig. 4. Presented PQS frequencies are normalized by virus with the highest PQS frequency in the corresponding G4Hunter score interval. All viruses with a high abundance of PQS relative to their GC content (over 70% of the maximal observed PQS frequency) are reported in Fig. 4. Different outliers could be seen depending on the analyzed intervals in Fig. 4: *Grapevine fleck virus*, *Cacao yellow mosaic virus* and *Grapevine asteroid mosaic associated virus, Grapevine red globe virus* and *Okra mosaic virus* which belong to *Tymoviridae* family of ssRNA+ viruses. The other outliers with more than 70% PQS belong to other classes of viruses: *Citrus exocortis viroid* and *Grapevine yellow speckle viroid 1* and *2* (viroids), *Human gyrovirus, Avian gyrovirus*, *Gyrovirus Tu789 and Micro torque teno virus* (ssDNA), *Equine encephalosis* (dsRNA) and *Ambe virus* (ssRNA-).

3.2. Localization of PQS in genomes of viruses

To evaluate the localization of PQS in viral genomes, we downloaded the genome annotations of all publicly accessible viral genomes and analyzed the presence of all PQS in each annotated *feature* and in its close proximity (100 nt before and after feature annotation). The pool of viral sequences was divided into 7 groups according to the type of nucleic acid as shown in Supplementary Material 6. Detailed results are summarized in Supplementary Material 5. The best annotated features in viral genomes are coding sequence (*CDS)* and *gene* (more than 371,000 annotated genes in viral genomes). The mean PQS frequency in *gene* was 0.9 per 1,000 nt, PQS frequency found in region 100 nt before *gene* was 1.02

Table 1
Genomic length, PQS frequencies and total counts. Seq (total number of sequences), Median (median length of sequences), GC % (average GC content), PQS (total number of predicted PQS), Mean f (mean of predicted PQS per 1,000 nt), Min f (the lowest frequency of predicted PQS per 1,000 nt), Max f (the highest frequency of predicted PQS per 1,000 nt) and Cov % (% of genome covered by PQS). Colors correspond to phylogenetic tree shown in Supplementary Material 6.

Domain	Seq	Median	GC%	PQS	Mean	Min	Max	Cov%
All	13,809	7,149	45.29	342,830	1.13	0	23.80	2.84
Viruses	13,405	7,328	45.33	342,130	1.13	0	23.80	2.83
Satellites	334	1,346	41.79	356	1.08	0	12.72	2.81
Viroids	66	339.5	56.68	49	2.09	0	16.26	2.83
Group	**Seq**	**Median**	**GC%**	**PQS**	**Mean**	**Min**	**Max**	**Cov%**
dsDNA	4,081	43,734	46.80	285,263	1.15	0	11.51	2.80
ssDNA	2,006	2,711	44.91	7,413	1.27	0	10.04	3.28
dsRNA	1,569	2,157	44.97	4,790	1.09	0	13.08	2.66
ssRNA+	2,818	7,511	45.97	25,570	1.21	0	23.80	3.10
ssRNA-	1,265	4,568	40.52	6,385	0.67	0	8.92	1.79
ssRNA+ RT	136	8,399.5	47.93	3,699	3.55	0	10.84	8.75
dsDNA RT	172	6,991	44.45	787	1.06	0	5.05	2.72
ssDNA satellites	285	1,350	40.21	296	0.96	0	7.43	2.50
ssRNA satellites	27	457	53.31	34	2.19	0	12.72	5.86
Family	**Seq**	**Median**	**GC%**	**PQS**	**Mean**	**Min**	**Max**	**Cov%**
Ackermannviridae	21	157,498	46.46	1589	0.48	0.35	0.72	1.23
Adenoviridae	189	35,172	52.55	21,945	3.27	0.16	7.57	7.91
Baculoviridae	91	131,160	41.74	2,903	0.25	0	1.81	0.57
Herpesviridae	120	153,119	53.14	66,125	3.43	0	11.44	8.91
Iridoviridae	40	114,073	46.49	8,323	1.82	0.08	3.62	4.31
Myoviridae	677	147,303	42.98	31,959	0.43	0	2.99	1.06
Papillomaviridae	369	7,673	43.34	5,618	1.97	0	6.04	5.00
Phycodnaviridae	36	192,531	41.35	6,031	0.69	0	1.88	1.69
Podoviridae	431	42,492	47.71	12,318	0.60	0	3.88	1.45
Polyomaviridae	142	5,088	42.04	1,277	1.68	0	7.81	4.63
Poxviridae	63	192,353	33.72	2,210	0.22	0	2.71	0.51
Siphoviridae	1,419	47,342	51.96	93,824	1.23	0	7.02	2.88
Anelloviridae	124	2,909	48.03	1,364	3.57	0.34	9.27	9.09
Circoviridae	245	1,901	46.17	891	1.87	0	10.04	4.76
Geminiviridae	894	2,737	43.72	1,999	0.83	0	4.72	2.18
Genomoviridae	117	2,199	51.54	601	2.33	0	6.71	6.15
Inoviridae	56	6,988	47.45	467	1.18	0	5.17	2.83
Microviridae	102	5,386	43.26	129	0.24	0	2.16	0.57
Parvoviridae	181	5,053	44.93	1,205	1.32	0	5.11	3.37
Smacoviridae	57	2,575	47.49	199	1.36	0	3.22	3.44
Amalgaviridae	21	3,412	48.75	137	1.91	0	5.84	5.07
Endornaviridae	38	14,157.5	41.54	199	0.34	0	1.08	0.85
Partitiviridae	162	1,832	46.48	569	1.79	0	11.19	4.38
Reoviridae	1,060	1,876.5	42.95	1,737	0.84	0	13.08	2.06
Totiviridae	99	5,193	51.51	1,034	1.86	0	7.73	4.61
Alphaflexiviridae	61	6,656	50.08	991	2.45	0	10.93	5.92
Arteriviridae	24	15,235.5	51.70	433	1.20	0.45	2.72	3.10
Astroviridae	70	6,613.5	47.19	421	0.93	0	5.21	2.26
Betaflexiviridae	120	8,390	43.90	658	0.70	0	2.88	1.67

Bromoviridae	134	2,862	43.95	128	0.36	0	1.89	0.94
Caliciviridae	399	7,511	50.41	5,662	1.89	0.35	5.06	5.09
Coronaviridae	208	29,695	39.78	575	0.10	0	0.52	0.23
Closteroviridae	72	13,227	40.99	480	0.51	0	2.28	1.33
Dicistroviridae	27	9,275	39.05	67	0.27	0	0.86	0.71
Flaviviridae	221	10,560	51.58	4,446	2.04	0.26	5.63	5.03
Iflaviridae	45	9,932	38.87	204	0.46	0	1.56	1.27
Luteoviridae	55	5,808	49.07	527	1.68	0	3.23	4.66
Narnaviridae	63	2,718	41.22	105	0.56	0	4.37	1.34
Picornaviridae	435	7,413	45.04	4,282	1.25	0	10.70	3.34
Potyviridae	185	9,689	42.51	452	0.28	0	1.75	0.70
Secoviridae	138	5,841.5	44.10	498	0.63	0	2.72	1.72
Solemoviridae	20	4,172.5	50.09	93	1.09	0	2.65	2.67
Togaviridae	44	11,620	51.23	452	0.91	0.25	6.15	2.25
Tombusviridae	84	4,016	49.54	601	1.93	0	7.27	4.81
Tymoviridae	46	6,334	55.50	2363	8.69	1.63	23.80	21.86
Virgaviridae	85	6,311	42.39	165	0.38	0	1.96	0.88
Arenaviridae	96	3,534	42.09	450	0.96	0	3.20	2.32
Hantaviridae	105	6,637	38.54	174	0.49	0	2.17	1.39
Orthomyxoviridae	184	1,721	43.49	214	0.68	0	4.05	1.65
Paramyxoviridae	92	15,621	42.91	1,754	1.20	0.13	3.71	2.97
Peribunyaviridae	197	4,527	36.75	182	0.37	0	5.08	1.09
Phenuiviridae	144	3,402.5	41.19	302	0.60	0	4.39	2.03
Rhabdoviridae	184	11,902.5	41.66	2,123	0.98	0	3.32	2.53
Retroviridae	136	8,399.5	47.93	3,699	3.55	0	10.84	8.75
Caulimoviridae	91	7,586	41.05	309	0.44	0	1.31	1.09
Hepadnaviridae	81	3,215	48.27	478	1.75	0	5.05	4.55
Alphasatellitidae	102	1,361	41.59	105	0.80	0	3.38	1.96
Tolecusatellitidae	148	1,351.5	39.18	129	0.72	0	7.43	1.97
Pospiviroidae	52	329.5	57.98	36	1.99	0	16.26	4.45

and 100 nt after *gene* 1.04 PQS per 1.000 nt. The best annotated in viral genomes is the "gene" feature. To compare localization in various features we took PQS frequency in genes as "1" and compare enrichment or depletion in other locations relative to this normalized value. PQS frequency in every feature type was calculated first for each virus separately and then the average was made. PQS frequencies in comparison to PQS frequencies in *gene* in annotated genome sites and in their close proximity are shown in Fig. 5. Statistically significant differences between PQS frequency in gene and other features in the corresponding category are highlighted by asterisk. In the first group, dsDNA viruses, the most notable enrichment was noted inside and around *repeat regions* (5.1- and 1.6-times more frequent compared to gene). A higher PQS frequency in comparison to *gene* average PQS abundance was observed also inside and after *misc_RNAs* (2.9- and 3.5-times, respectively), *tmRNA* (2-times) and *exons* (2.2-times more frequent than gene). In ssDNA viruses, similarly to dsDNA group, the highest PQS frequency was found inside *repeat regions* (9.2-times higher compared to *gene* region). In contrary to dsDNA group, but in agreement with dsRNA viruses, an abundance of PQS was observed also in *stem_loops* (6.5- and 2.1-times more) and *5′UTRs* (2.5- and 1.6-times). Higher frequency in comparison to PQS frequency in *genes* was found also inside the *introns* (2.5-times higher) and *rep_origins* (1.8-times). For ssRNA+ group, the most significant enhancement was found around *exon* (1.7- and 3.4-times, respectively), before *regulatory elements* (1.9) and inside *miscRNAs* (1.6). For ssRNA-group, no PQS were found inside *5′UTRs,* but the enrichment was noted also inside *stem_loops* (35.3-times), *misc_RNA* (2.1-times), *regulatory elements* (2.2-times) and before and in *exons* (4.4- and 2.3-times). For ssRNA viruses with RT, the highest PQS frequency compared to *gene* was observed before *exons* (3.3-times), however, the abundance of PQS was observed also inside *misc_binding* sites (2.3-times) *3′UTRs* and *5′UTRs* (1.4-times). In dsDNA viruses with RT the most notable enrichment was before *mRNAs* (4-times) after *regulatory elements (1.8-times respectively) and* around *misc_RNAs (2.1- and 3.3-times)*. Satellites are enriched for PQS in *stem_loops* (14.9-times) and *repeats* (22.7-times).

3.3. Comparison of PQS frequency in viruses with dominant acute and persistent type of the infection

The cluster and ROC analyses (Fig. 2) revealed that the type of nucleic acid is not the main factor influencing the PQS prevalence, as might be expected, but viruses rather cluster according to other phenotypic characteristics, such as dominant type of infection [31,32,36,37]. In general, a viral infection is considered acute when onset and resolution of infection is rapid, or persistent when primary infection is not cleared by the immune response and virus

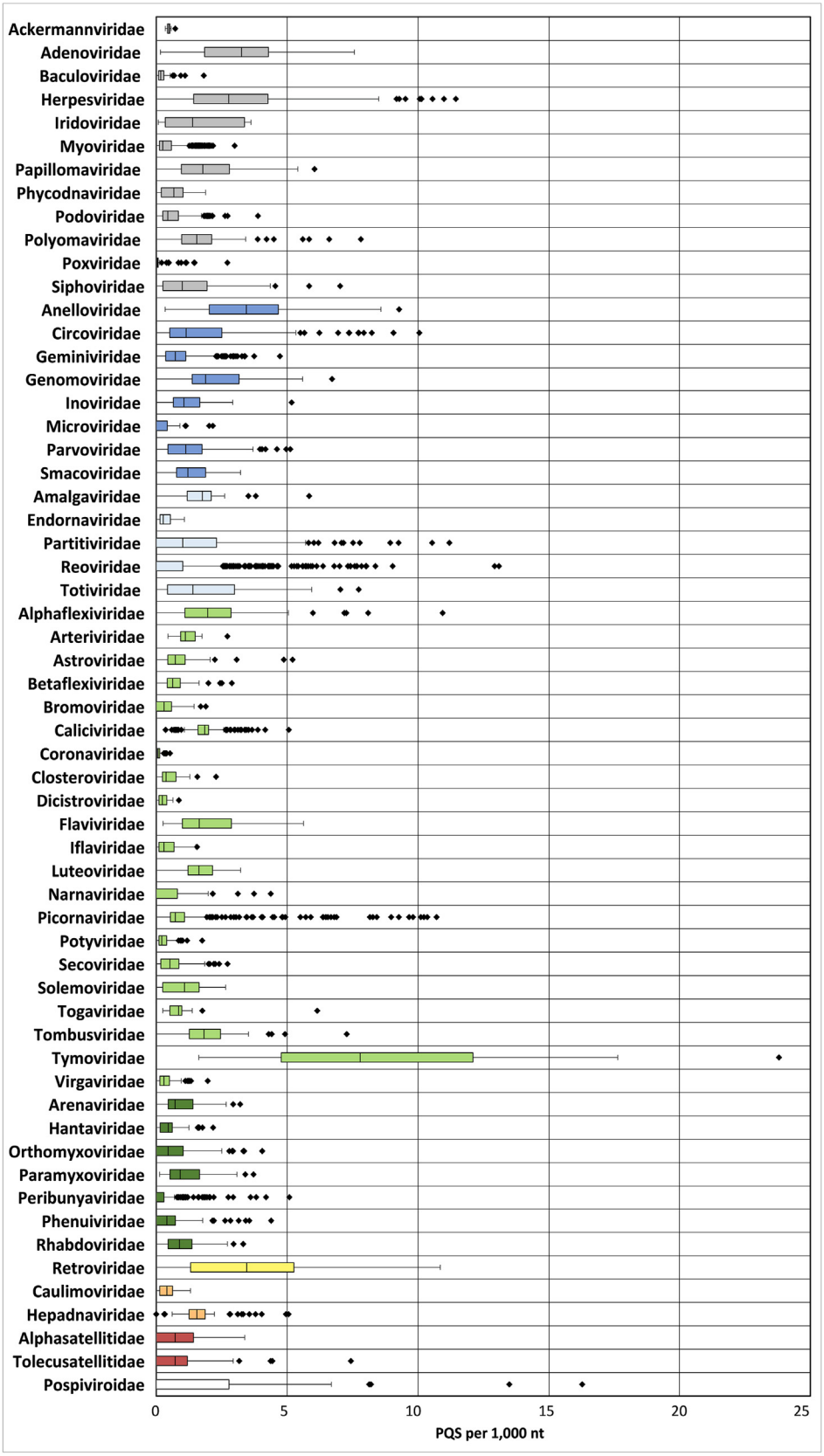


Fig. 1. Frequencies of PQS in families of the analyzed viral genomes. Data within boxes span the interquartile range and whiskers show the lowest and highest values within 1.5 interquartile range. Black diamonds denote outliers. Colors correspond to the phylogenetic tree shown in Supplementary Material 6.

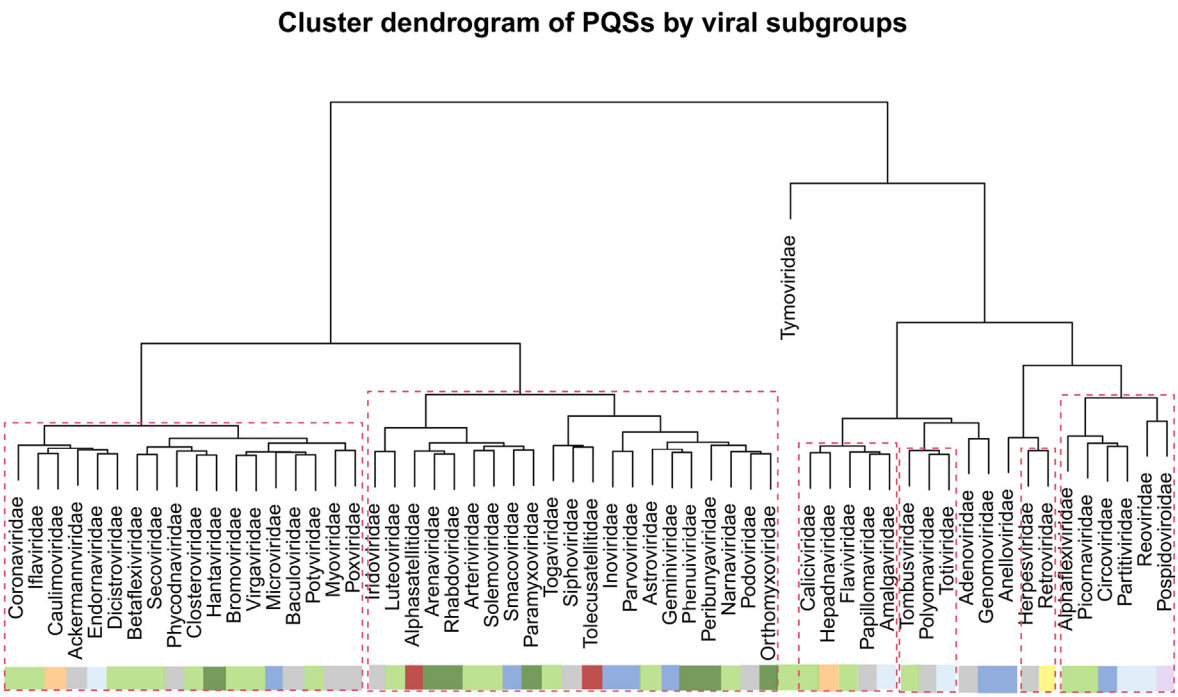


Fig. 2. Cluster dendrogram based on PQS characteristics in all viral families. Input data are listed in Supplementary Material 4. Statistically significant clusters (based on AU values above 95, equivalent to p-values lower than 0.05) are highlighted by rectangles drawn with dashed lines. Colors corresponds to phylogenetic tree in Supplementary Material 6.

remains in host cells either as a provirus or episome. This distinction is not always straightforward, as some viruses causing acute infection may also establish persistency in immunocompromised hosts [38]. Based on the literature [30–34,36–49] we chose only extreme groups of viruses which typically cause either persistent or acute type of infection in *Metazoa* hosts and separated them in two main groups: 1. those causing persistent infections (*Adenoviridae, Herpesviridae, Papillomaviridae, Polyomaviridae, Anelloviridae, Hepacivirus* genus of *Flaviviridae* family, *Retroviridae*, and *Hepadnaviridae*); 2. viruses causing mainly acute infections (*Poxviridae, Rotavirus* genus of *Reoviridae*, *Coronaviridae*, *Hepatovirus* genus and Rhinoviruses of *Picornaviridae* family and *Orthomyxoviridae*). In case the family of viruses was too diverse, we selected lower taxonomic units. For example in *Flaviviridae* family; while viruses from genus *Hepacivirus* are a common cause of chronic infections, the genus *Flavivirus* contain viruses with a much more diverse behavior [50,51]. Similarly *Reoviridae* family includes the acutely infecting genus *Rotavirus,* but also other genera of plant pathogens [52]. For further statistical analyses, only groups with 20 or more sequenced genomes were evaluated. PQS frequencies are shown in Fig. 6, panel A. While the persistently infecting viruses have an average PQS frequency of 2.66 PQS per 1,000 nt with five groups with a PQS frequency above 3 PQS per 1000 nt (*Adenoviridae* (3.27), *Herpesviridae* (3.55), *Anelloviridae* (3.57), *Hepacivirus* (3.31), *Retroviridae* (3.55), all average PQS frequencies of acutely infecting groups of viruses are below 0.7 per 1,000 nt (average of 0.33, with values ranging from 0.04 for *Rotaviruses* and 0.68 for *Orthomyxoviridae*; *Poxviridae* (0.22), *Coronaviridae* (0.097), *Hepatoviruses* (0.6), and Rhinoviruses (0.57) falling in between). The values found for the viruses in the second group are statistically significantly different from the density of group 1 viruses, independently of the type of nucleic acid (Fig. 6, panel B and Supplementary Material 3). PQS frequencies therefore correlate with the dominant type of infection rather than with the type of nucleic acid (dsDNA virus from *Adenoviridae, Herpesviridae* vs dsDNA viruses of *Poxviridae* as shown in Fig. 2), intracellular localization (nucleus localization of persistently infecting *Herpesviruses* and acutely infecting *Orthomyxoviridae* vs cytoplasmatic localization of acutely infecting *Poxviridae* and chronically infecting *Hepacivirus*) or tissue tropism. Such example is viruses causing hepatitis; whereas hepatitis A virus (HAV) from *Hepatovirus* of *Picornaviridae* family with ssRNA+ genomes cause only acute hepatitis with average PQS frequency of 0.6, HBV from *Hepadnavirus* of *Hepadnaviridae* (dsDNA RT) and hepatitis C virus (HCV) from *Hepacivirus* of *Flaviriridae* (ssRNA+) are causative agents of chronic hepatitis with average PQS frequency of 1.75 and 3.31, respectively (Fig. 6, panel C). Moreover, both HAV and HCV are ssRNA+ viruses with statistically significant difference in PQS frequency. Hepatitis D virus (HDV) and hepatitis E virus (HEV) were not included in the division as HDV is a satellite of HBV as it requires the HBV envelope proteins to form virus particles, and the infection type of HEV is not straightforward [53–55].

4. Discussion

The presence of the G4s in viral genomes has attracted eminent interest due to their potential in antiviral therapies [6]. The most comprehensive viral genome analyses to date of PQS have been done using a pattern-based algorithm [23,24]. However, pattern-based algorithms do not consider atypical G4 forming structures, including bulges for example [56]. In this study, we have broadened the range of analyses significantly by looking at *all* sequenced viral genomes, not only human-hosted viruses (among 11,000 viruses studied, 1,184 are human hosted). The presence and locations of PQS in more than thirteen thousand viral sequences were analyzed using G4Hunter, which allows the identification of non-canonical PQS and the categorization of results based on G4Hunter score values. By expanding the host spectrum to all organisms, we gained crucial additional information about possible zoonotic viral pathogens, conservation of PQS among different species and

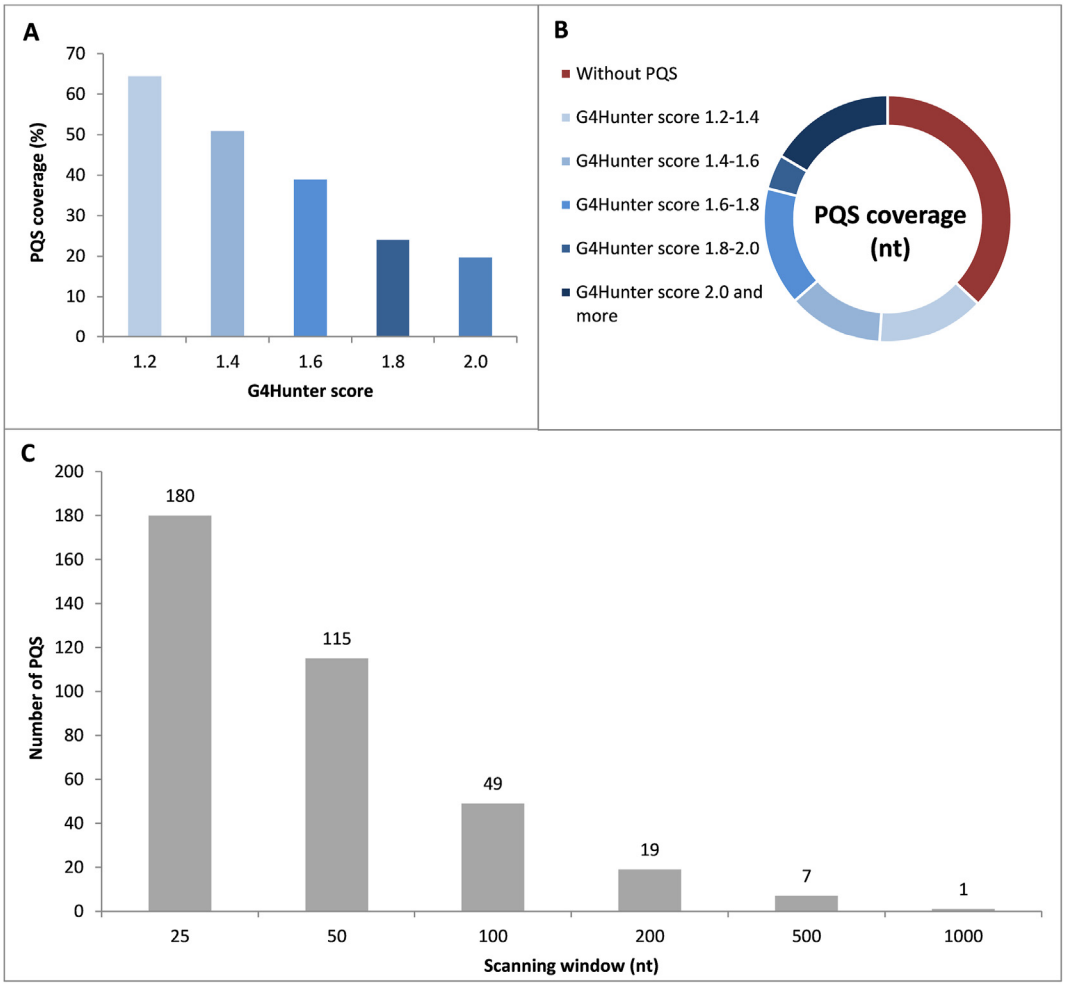


Fig. 3. Graphical visualization of PQS densities in the *Grapevine fleck virus*. (A) PQS coverage of the *Grapevine fleck virus* genome with the different G4Hunter score threshold. (B) Representation of the PQS length in the *Grapevine fleck virus* according to different G4hunter score intervals' (C) Number of PQS found according to different scanning window length using a threshold of 1.2.

interspecies variability in PQS frequency. Moreover, we broadened the analyses also to subviral agents. On average, viroids showed higher PQS frequency (2.09) than viruses (1.13) and satellites (1.08). Since viroids do not code for any protein, it has been proposed that their RNA secondary structures may play important roles in their life cycle [20].

Grapevine fleck virus, *Cacao yellow mosaic virus, Grapevine red globe virus* and *Okra mosaic virus* are four outliers found in the GC content analyses (Fig. 4). All of them belong to the *Tymoviridae* family of ssRNA+ viruses, the family with the highest PQS frequency found (8.69). *Grapevine fleck virus* has the highest frequency of 23.8 PQS per 1,000 nt among all analyzed genomes. Such frequency is extraordinary, especially for ssRNA+ viruses, where the average PQS frequency is only 1.21 PQS per 1,000 nt. The *Grapevine fleck virus* is strictly confined to the phloem of infected hosts. Infected cells contain multivesiculate bodies, cytopathic structures derived from mitochondria, and a great quantity of accumulated viral particles, sometimes also in crystalline arrays [57]. The *Anelloviridae* family has the second highest PQS frequency (3.57) among all families (Table 1, Fig. 1). Viruses of the *Anelloviridae* family cause chronic human infections without any known direct clinical symptoms [30]. Four viruses of this family, *Human gyrovirus*, *Avian gyrovirus, Gyrovirus Tu789* and *Micro Torque teno virus,* exhibited extremely high PQS frequencies (Fig. 4). The third highest PQS frequency among genome subgroups was found for the *Retroviridae* family (3.55), which includes RNA viruses with reverse transcriptase, as shown in Table 1. G4s in genomes of the *Retroviridae* family, especially in HIV-1 virus from the genus *Lentivirus*, have been intensively studied with respect to promising antiviral therapy via G4s targeting. Three potential G4-forming sequences, LTR-II, LTR-III and LTR-IV, are localized in highly conserved sp1 and NFκB binding sites of the U3 region of the long terminal repeat (LTR). Viral transcription can be regulated by G4 stabilizing proteins and ligands. Whereas selective stabilization of LTR-III leads to inhibition of viral transcription, stabilization of LTR-IV has the opposite effect [58,59]. Our analyses confirm that PQS of *Retroviridae* family are enriched for PQS in 3′*UTRs and* 5′*UTRs* compared to *genes*, as shown in Fig. 5. PQS was found also in the *nef* gene of HIV virus, an important virulence factor, responsible for T-cell activation and maintenance of a persistent state of infection [60].

Among dsDNA viruses, the highest observed PQS frequency was found inside the *Herpesviridae* family (3.43). The importance of the roles of G4s in the regulation of basic molecular mechanisms was reported for several species of this family [11,61–64]. G4 stabilizing compounds could inhibit replication, gene expression and importantly alter viral life cycle [5,65]. Our analyses indicate that PQS are mostly found in *repeat regions, misc_RNA* and *tmRNA*s. Another dsDNA family with significantly higher PQS frequency is

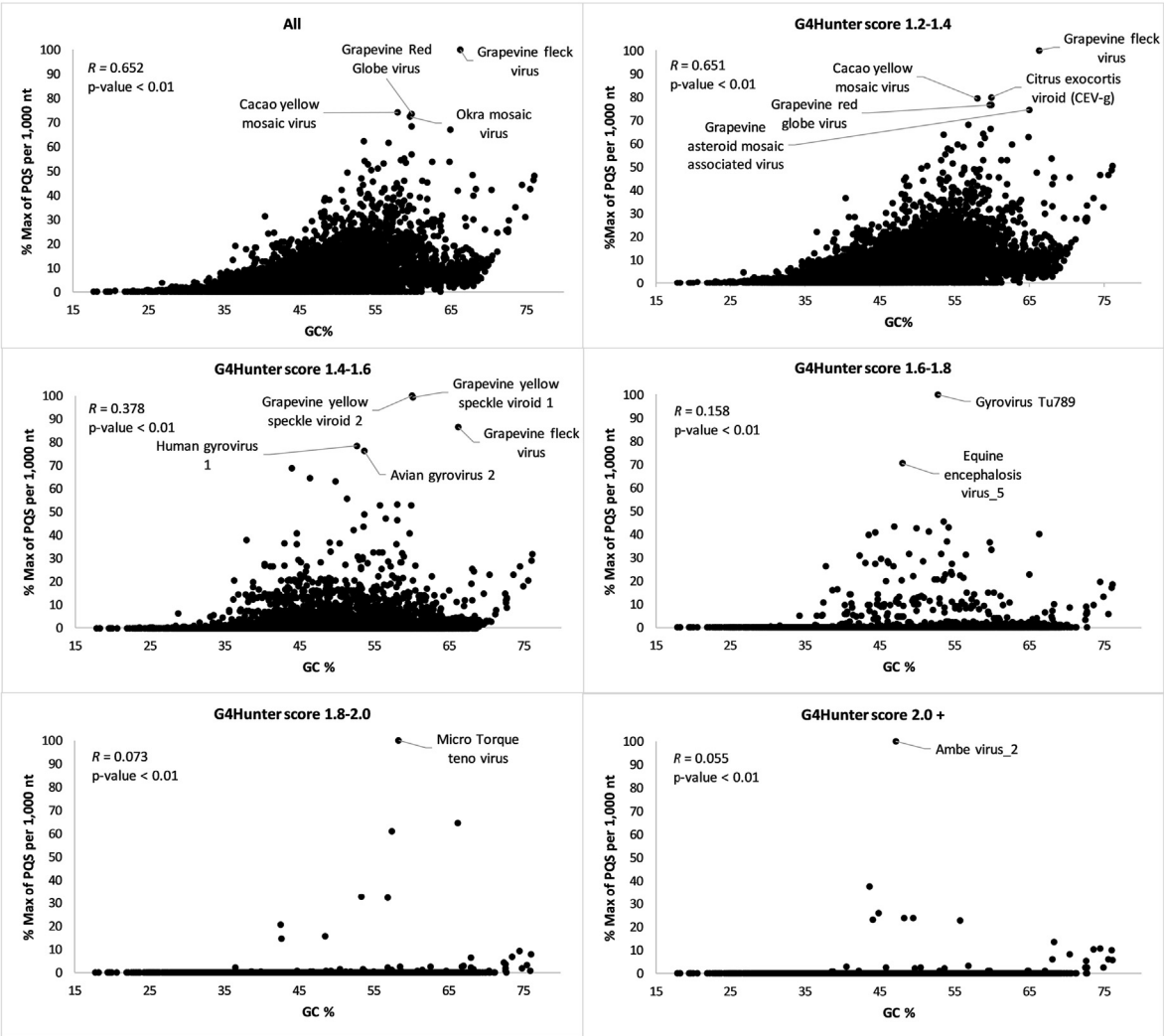


Fig. 4. Relationship between observed frequency of PQS per 1,000 nt and GC content in all analyzed viral sequences in various G4 Hunter score intervals. In each G4Hunter score interval miniplot, frequencies were normalized according to the highest observed frequency of PQS. The names of viruses exhibiting a relative maximal frequency per 1,000 nt greater than 70% are reported. Information about statistical evaluation by Spearman's correlation analyses are shown in the graphs.

Adenoviridae. Genome-wide analyses revealed 15 G4s to be conserved across different species of human adenoviruses [66]. Moreover, the G4 ligand BRACO-19 inhibited multiplication of the virus in human cells [66]. For both families, *Herpesviridae* and *Adenoviridae*, the presence of G4s was found in early-expressed genes, which encode regulatory proteins crucial for the next phases of viral life cycles [66,67]. Inside the *Papillomaviridae* family (where we noted the third highest PQS frequency (1.97) among dsDNA viruses), G4 forming sequences were found by pattern based algorithm in 8 out of the 120 identified HPV genotypes, including oncogenic high risk *Human Papillomavirus type 52* (HPV-52), HPV-58 [68]. Our analyses identified only 5 human genotypes (out of 188) completely lacking PQS. It has recently been shown that PQS frequency in dsDNA viruses correlate with PQS frequencies in corresponding hosts, pointing to evolutionarily reciprocal mimicking of virus-host genome organization [69].

A higher frequency of PQS was found for ssRNA+ viruses in comparison to ssRNA-viruses. For both groups, both positive and negative strands were analyzed for the presence of PQS as both strands are essential in the life cycle of the virus. For example, the stabilization of G4 formation on the negative intermediate of HCV (belonging to the *Flaviviridae* family of ssRNA+ viruses) can modulate RNA synthesis [70]. Localization of PQS was evaluated in all genome groups with emphasis on annotated features. Several groups of viruses exhibited a PQS abundance in *5′UTRs*, *3′UTRs*, *repeats*, or *replication origin* compared to gene locations (Fig. 5). Enrichment in these particular areas of the genomes suggest important roles of G4s in regulation of different basic molecular processes [6].

Based on our analyses we demonstrate that PQS frequency in viral genomes is correlated to the type of viral infection in *Metazoa* hosts. We observed statistically significant higher PQS frequencies (as shown in Fig. 6, Supplementary Material 3) in subgroups of viruses that have previously been reported [30–32,36,40–43,49] to cause persistent, chronic or latent infections in *Metazoa* hosts such as *Adenoviridae*, *Herpesviridae*, *Papillomaviridae*, *Polyomaviridae*, *Anelloviridae*, *Hepacivirus*, *Retroviridae* or *Hepadnaviridae*. On the contrary, low PQS frequencies were observed in subgroups causing mostly acute infections [33,34,36,37,45,48], such as *Poxviridae*, *Rotavirus* genus of *Reoviridae* family, *Hepatovirus* genus and Rhinoviruses of *Picornaviridae* family, the *Coronaviridae* family of ssRNA+ viruses, as well as the *Orthomyxoviridae* family of ssRNA-viruses. Interestingly, PQS frequency correlates with the type of infection rather than with the type of nucleic acid, phylogenetic

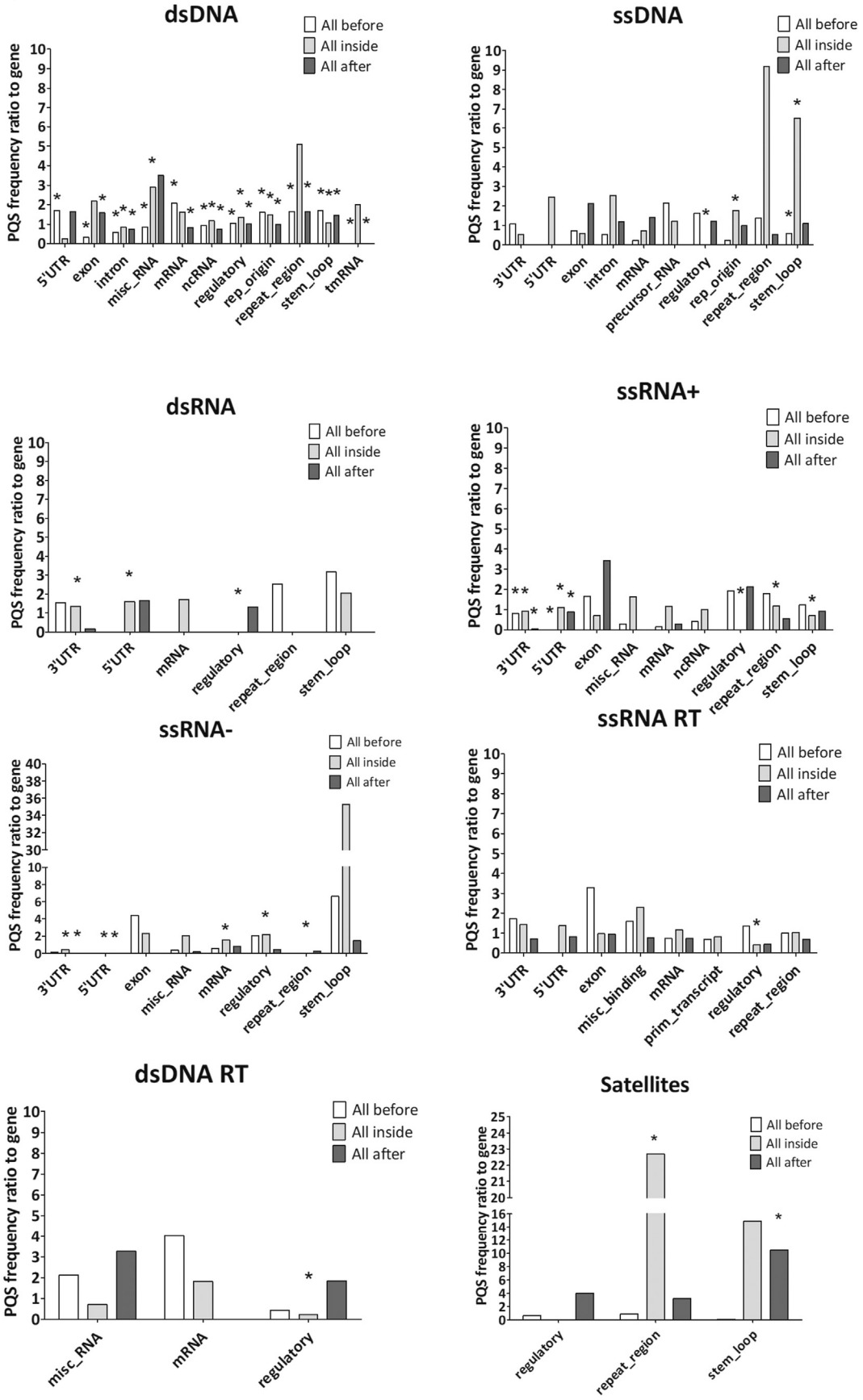


Fig. 5. Differences in PQS frequency by annotated locus. The chart shows the ratio of PQS frequencies per 1,000 nt between *gene* annotation and other annotated locations from the NCBI database. We analyzed the frequencies of all PQS within (inside), before (100 nt) and after (100 nt) annotated locations. Detailed results are summarized in Supplementary Material 5. Asterisks denote statistically significant differences (p-value < 0.05) between PQS frequency in *gene* and other feature in corresponding category evaluated by non-parametric Kruskal-Wallis test and the results are presented in Supplementary Material 3.

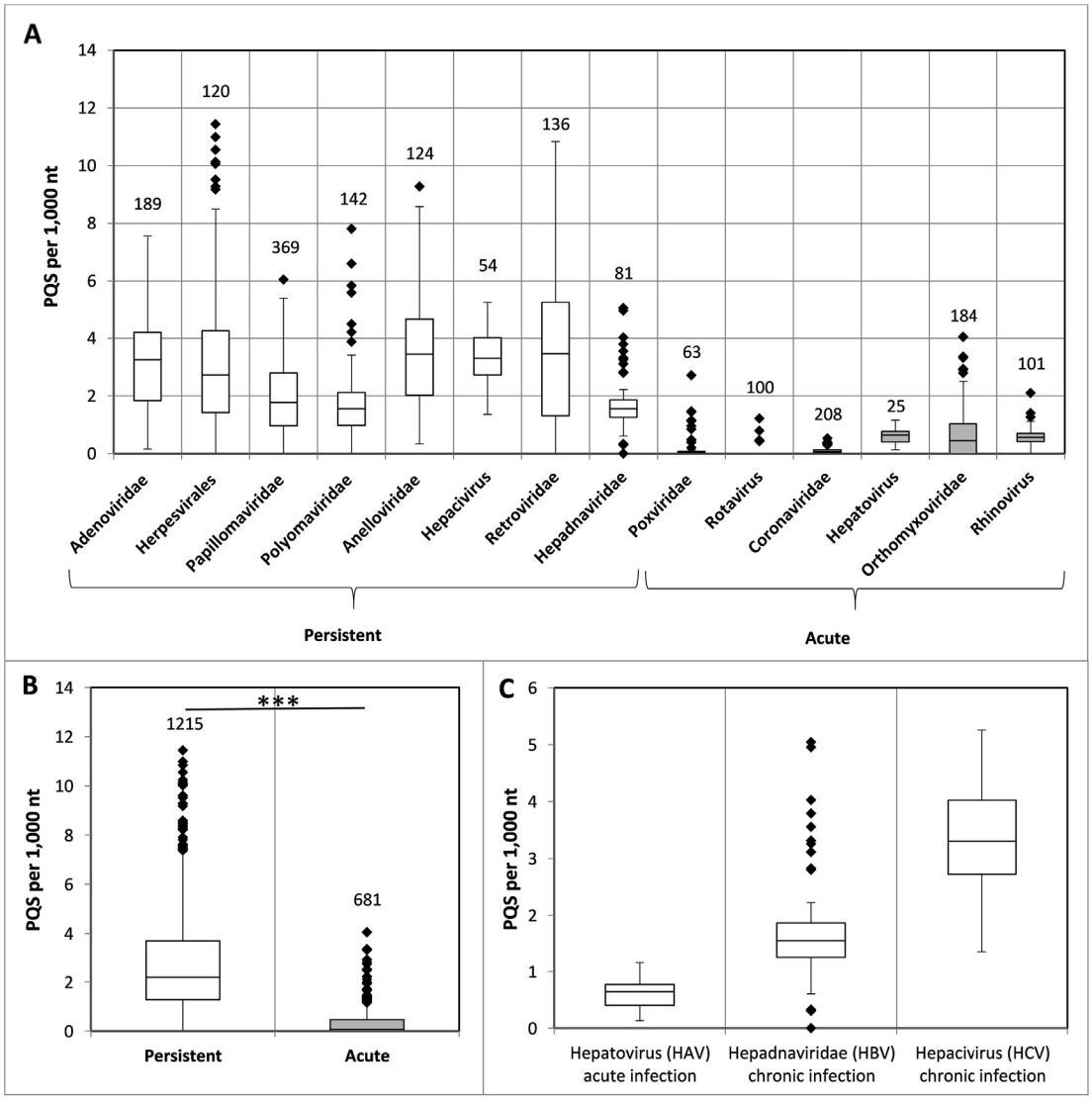


Fig. 6. Comparison of PQS frequency in viruses with dominant persistent and acute types of the infection. Data within boxes span the interquartile range and whiskers show the lowest and highest values within 1.5 interquartile range. Black diamonds denote outliers and the digits indicate the number of viruses in each category. Selected groups of viruses were categorized according dominant type of the infection (panel A). Viruses causing mostly persistent infection (white) are statistically significantly (p-value 3E-219) enriched in PQS compared to acutely infecting viruses (grey) (panel B). PQS frequencies of three viral groups causing either acute or chronic hepatitis (panel C). Statistical evaluation of the data set has been made through Kruskal-Wallis test and the results are presented in Supplementary Material 3.

categorization, or tissue tropism. For example, viruses belonging to the *Poxviridae* family with a dsDNA genome that cause acute infections have a significantly lower PQS frequency than other dsDNA subgroups that are known to cause persistent infections (such as *Adenoviridae* and *Papillomaviridae* family or order *Herpesviridae*) [31,34]. ssRNA+ viruses from genus *Hepacivirus* that cause chronic hepatitis and hepatocellular carcinoma in humans [66] have significantly higher PQS than ssRNA+ viruses from genus *Hepatovirus*, acutely infecting viruses with same tissue tropism [36]. Moreover, cluster analyses (Fig. 2) revealed that viruses causing persistent infections clustered together regardless of the nucleic acid type and current phylogenetic classification. On top of that, two evolutionary distinct subgroups that typically cause persistent types of infection, *Herpesviridae* and *Retroviridae*, clustered together with statistically significant similarity (Fig. 2). However, identifying the primary type of infection is not always straightforward, as found for the *Filoviridae* family to which Ebola and Marburg virus belong. Members of this family with an average PQS frequency of 0.95 PQS per 1,000 nt cause severe acute infections with extreme mortality rates over 50%. However, both Ebola and Marburg virus RNA have been found in semen, cerebrospinal and intraocular fluids of survivors long after blood samples tested negative [71,72]. Stabilization of G4 inhibits L gene expression, which encodes for RNA-directed RNA polymerase of Zaire ebolavirus [73,74]. Another example of not clear category of viruses could be Measles virus of *Paramyxoviridae* family with the 2.33 PQS per 1,000 nt causing mostly acute infection, but it can persist in the nervous system and was linked to chronic disease such as Paget's disease and otosclerosis [75]. The presence and high prevalence of G4 thus cannot be considered as the only indicator of the type of infection, but particular G4s could play a role in a general viral persistency in Metazoan hosts.

Evolutionary elimination of G4s may represent a selective advantage for actively replicating viruses, for at least two reasons: i) this would avoid any problem in replication, transcription or reverse transcription, as stable DNA and RNA G4s may interfere

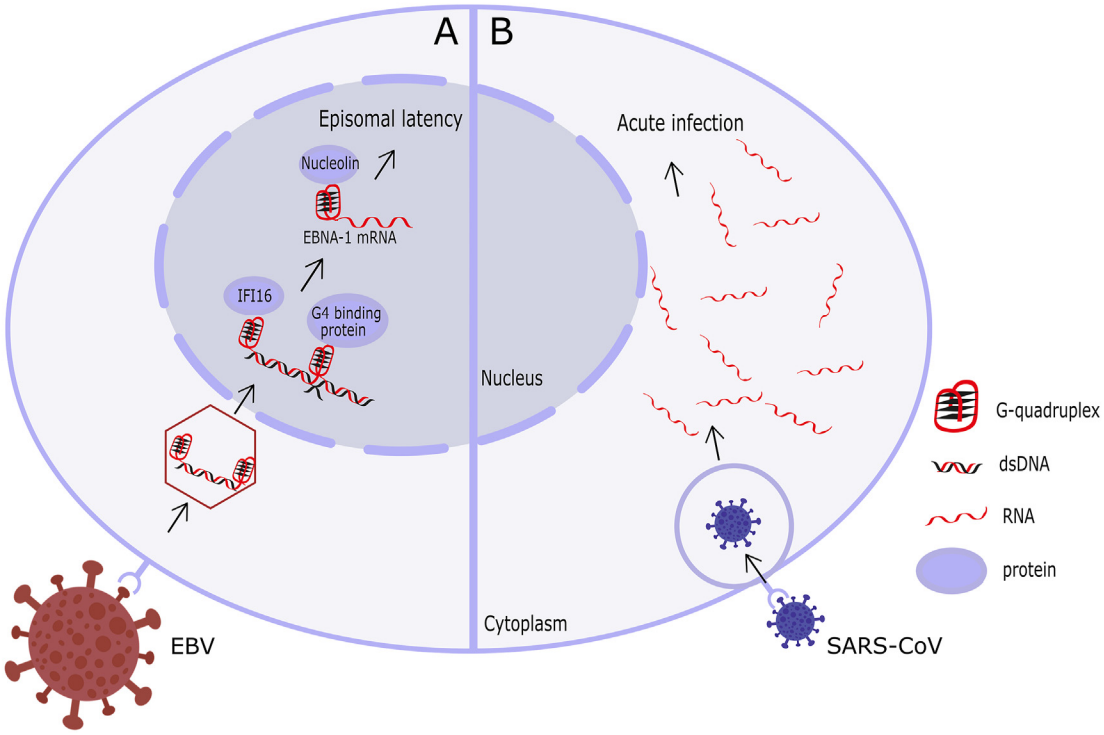


Fig. 7. Schematic model of G4 influence on the viral life cycle. Viruses that cause persistent infections, such as EBV, have high potential PQS frequency (panel A) that may be recognized by DNA sensors (e.g. IFI16) [76,77] and other G4-binding proteins (e.g. nucleolin) [11]. Binding to a G4-binding protein such as nucleolin can influence latency and immune evasion, as found for EBNA-1. In contrast, viruses that cause acute infections, such as SARS-CoV, tend to have a low PQS frequency (panel B).

with these processes; ii) their genome would not be recognized by cellular G4 binding proteins and alien DNA sensors of the host cells, such as interferon-γ-inducible protein 16 (IFI16) [11,76], a protein which preferentially binds to G4 and is critical for maintenance of EBV latency [11,76,77]. G4s depletion thus could be one of the reasons why these viruses are able to quickly develop an acute phase of infection. On the other hand, viruses with high PQS frequency would be forced to "hide" their genomes into the host genome, which leads to development of the persistent phase of the infection (Fig. 7). Acutely infecting viruses are the causes of seasonal epidemics. This hypothesis confirmed also contemporary analyses of H1N1 flu genome [78] and SARS-CoV-2 genome in which PQS are rare [79,80]. So far, the involvement of G4s in the regulation of viral life cycles has been demonstrated on HIV-1 virus from *Retroviridae* family, Kaposi's sarcoma-associated herpes virus (KSHV) and EBV virus belonging to *Herpesviridae* family. In the case of HIV-1 virus, treatment of latently infected cells with G4 ligands BRACO-19 and TMPyP4 led to inhibition of HIV-1 replication and increased apoptosis rate [81]. Moreover, formation of the three conserved G4s has been confirmed in the *nef* gene of HIV-1, an important factor in maintenance of the persistent state of infection. Treatment with a G4 ligand reduced *nef* expression and repressed nef-dependent HIV-1 infectivity [60]. G4 is also located inside the regulator of latency and immune evasion of another virus – herpesvirus EBV. EBV, formally known as a *Human gammaherpesvirus 4,* is a causative agent of different lymphoproliferative disease, such as Burkitt lymphoma or Hodgkin lymphoma. EBNA-1 is the only EBV protein required for replication and segregation of episomal genomes in latency, resulting in the maintenance of the EBV genomes. EBNA-1 is also highly antigenic; destabilization of the G4 structures through codon-modification significantly enhanced *in vivo* antigen presentation and activation of virus-specific T cells [82]. Formation of RNA G4 in EBNA1 mRNA of EBV inhibits its translation and, as a result, also the presentation of antigen. G4 formation in mRNA of EBNA1 therefore represents a mechanism of how the virus balances between latency, immune evasion and stable copy number of genomes [11,65]. The analogous gene in the KSHV genome, *LANA-1*, is the most abundantly expressed protein during latency which is required for the persistence of KSHV in the host cell. It has been demonstrated that stabilization of G4 in LANA-1 mRNA inhibits its translation and immune recognition [83]. KSHV virus, also known as *Human gammaherpesvirus 8* causes Kaposi sarcoma and other malignancies, mostly in immunodeficient patients. G4 ligands could eliminate even extrachromosomal episomes of KSHV [5]. Overall, while the presence of PQS has been associated with latency for some viruses, our study provides for the first time a statistical link between G4 propensity and persistence of the virus.

5. Conclusions

In this article, we analyzed the presence of PQS in all accessible viral genome sequences. The number of PQS differs remarkably among individual groups and families, suggesting evolutionary adaptations connected with G4. The most striking result was the observation that the PQS frequency varied according to the type of infection, rather than to nucleic acid type of the genome: while viruses typically causing persistent infections in *Metazoa* hosts require a similar G4 background as the host and are enriched for the presence of PQS, the viruses causing acute phase only are often reduced or absent in PQS. We have shown that PQS are non-randomly distributed in viral genomes and associated with specific regions in their genomes, which have mainly regulatory functions. These data allow use of the unique and non-random localization of PQS in viral genomes in their characterization and points to G4 importance for viral lifecycles.

Funding

This work was supported by The Czech Science Foundation (18-15548S), the ANR Flash-Covid 2021 call and by the SYMBIT project Reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477 financed from the ERDF.

Author contributions

Conceptualization, V.B.; Data curation, N.B and A.C.; Formal analysis, N.B. and M.B.; Funding acquisition, V.B. and M.F.; Investigation, N.B and A.C.; Methodology, V.B.; Project administration, V.B.; Resources, P.K. and J.Š.; Software, P.K. and J.Š.; Supervision, V.B., M.F., and P.P.; Validation, V.B.; Visualization, N.B., A.C. J.L.M. and M.B.; Writing—original draft, N.B., A.C, M.B. and V.B.; and Writing—review and editing, P.P., J.L.M. and M.F.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

We thank Prof. Richard Bowater, University of East Anglia, for proofreading and providing feedback on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biochi.2021.03.017.

References

[1] R.W. Harkness, A.K. Mittermaier, G-quadruplex dynamics, Biochim. Biophys. Acta Protein Proteonomics 1865 (2017) 1544–1554, https://doi.org/10.1016/j.bbapap.2017.06.012.
[2] D. Rhodes, H.J. Lipps, G-quadruplexes and their regulatory roles in biology, Nucleic Acids Res. 43 (2015) 8627–8637, https://doi.org/10.1093/nar/gkv862.
[3] Y. Ding, A.M. Fleming, C.J. Burrows, Case studies on potential G-quadruplex-forming sequences from the bacterial orders Deinococcales and Thermales derived from a survey of published genomes, Sci. Rep. 8 (2018) 1–11, https://doi.org/10.1038/s41598-018-33944-4.
[4] M. Falabella, J.E. Kolesar, C. Wallace, D. de Jesus, L. Sun, Y.V. Taguchi, C. Wang, T. Wang, I.M. Xiang, J.K. Alder, R. Maheshan, W. Horne, J. Turek-Herman, P.J. Pagano, C.M. St Croix, N. Sondheimer, L.A. Yatsunyk, F.B. Johnson, B.A. Kaufman, G-quadruplex dynamics contribute to regulation of mitochondrial gene expression, Sci. Rep. 9 (2019) 5605, https://doi.org/10.1038/s41598-019-41464-y.
[5] A. Madireddy, P. Purushothaman, C.P. Loosbroock, E.S. Robertson, C.L. Schildkraut, S.C. Verma, G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV, Nucleic Acids Res. 44 (2016) 3675–3694, https://doi.org/10.1093/nar/gkw038.
[6] E. Ruggiero, S.N. Richter, Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy, Annu. Rep. Med. Chem. 54 (2020) 101–131, https://doi.org/10.1016/bs.armc.2020.04.001.
[7] S. Asamitsu, S. Obata, Z. Yu, T. Bando, H. Sugiyama, Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy, Molecules 24 (2019) 429, https://doi.org/10.3390/molecules24030429.
[8] S.K. Mishra, A. Tawani, A. Mishra, A. Kumar, G4IPDB: a database for G-quadruplex structure forming nucleic acid interacting proteins, Sci. Rep. 6 (2016) 38144, https://doi.org/10.1038/srep38144.
[9] V. González, L.H. Hurley, The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity, Biochemistry 49 (2010) 9706–9714, https://doi.org/10.1021/bi100509s.
[10] E. Tosoni, I. Frasson, M. Scalabrin, R. Perrone, E. Butovskaya, M. Nadai, G. Palù, D. Fabris, S.N. Richter, Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription, Nucleic Acids Res. 43 (2015) 8884, https://doi.org/10.1093/nar/gkv897.
[11] M.J. Lista, R.P. Martins, O. Billant, M.-A. Contesse, S. Findakly, P. Pochard, C. Daskalogianni, C. Beauvineau, C. Guetta, C. Jamin, M.-P. Teulade-Fichou, R. Fåhraeus, C. Voisset, M. Blondel, Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA, Nat. Commun. 8 (2017) 16043, https://doi.org/10.1038/ncomms16043.
[12] W.-X. Bian, Y. Xie, X.-N. Wang, G.-H. Xu, B.-S. Fu, S. Li, G. Long, X. Zhou, X.-L. Zhang, Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication, Nucleic Acids Res. 47 (2019) 56–68, https://doi.org/10.1093/nar/gky1177.
[13] A.D. Cian, J.-L. Mergny, Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation, Nucleic Acids Res. 35 (2007) 2483–2493, https://doi.org/10.1093/nar/gkm098.
[14] Y. Kusov, J. Tan, E. Alvarez, L. Enjuanes, R. Hilgenfeld, A G-quadruplex-binding macrodomain within the "SARS-unique domain" is essential for the activity of the SARS-coronavirus replication–transcription complex, Virology 484 (2015) 313–322, https://doi.org/10.1016/j.virol.2015.06.016.
[15] C. Platella, C. Riccardi, D. Montesarchio, G.N. Roviello, D. Musumeci, G-quadruplex-based aptamers against protein targets in therapy and diagnostics, Biochim. Biophys. Acta Gen. Subj. 1861 (2017) 1429–1447, https://doi.org/10.1016/j.bbagen.2016.11.027.
[16] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res. 33 (2005) 2908–2916, https://doi.org/10.1093/nar/gki609.
[17] J. Eddy, N. Maizels, Gene function correlates with potential for G4 DNA formation in the human genome, Nucleic Acids Res. 34 (2006) 3887–3896, https://doi.org/10.1093/nar/gkl529.
[18] A. Bedrat, L. Lacroix, J.-L. Mergny, Re-evaluation of G-quadruplex propensity with G4Hunter, Nucleic Acids Res. 44 (2016) 1746–1759, https://doi.org/10.1093/nar/gkw006.
[19] V. Brázda, J. Kolomazník, J. Lýsek, M. Bartas, M. Fojta, J. Šťastný, J.-L. Mergny, G4Hunter web application: a web server for G-quadruplex prediction, Bioinformatics 35 (2019) 3493–3495, https://doi.org/10.1093/bioinformatics/btz087.
[20] B. Ding, Viroids: self-replicating, mobile, and fast-evolving noncoding regulatory RNAs, WIREs RNA 1 (2010) 362–375, https://doi.org/10.1002/wrna.22.
[21] P. Gnanasekaran, S. Chakraborty, Biology of viral satellites and their role in pathogenesis, Curr. Opin. Virol. 33 (2018) 96–105, https://doi.org/10.1016/j.coviro.2018.08.002.
[22] A.M. Fleming, Y. Ding, A. Alenko, C.J. Burrows, Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the Flaviviridae family, ACS Infect. Dis. 2 (2016) 674–681, https://doi.org/10.1021/acsinfecdis.6b00109.
[23] E. Puig Lombardi, A. Londoño-Vallejo, A. Nicolas, Relationship between G-quadruplex sequence composition in viruses and their hosts, Molecules 24 (2019) 1942, https://doi.org/10.3390/molecules24101942.
[24] E. Lavezzo, M. Berselli, I. Frasson, R. Perrone, G. Palù, A.R. Brazzale, S.N. Richter, S. Toppo, G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide, PLoS Comput. Biol. 14 (2018), e1006675, https://doi.org/10.1371/journal.pcbi.1006675.
[25] T. Mihara, Y. Nishimura, Y. Shimizu, H. Nishiyama, G. Yoshikawa, H. Uehara, P. Hingamp, S. Goto, H. Ogata, Linking virus genomes with host taxonomy, Viruses 8 (2016) 66, https://doi.org/10.3390/v8030066.
[26] ROC analysis: web-based calculator for ROC curves. http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html#exportAnchor. (Accessed 29 March 2021) accessed.
[27] P.J. Walker, S.G. Siddell, E.J. Lefkowitz, A.R. Mushegian, D.M. Dempsey, B.E. Dutilh, B. Harrach, R.L. Harrison, R.C. Hendrickson, S. Junglen, N.J. Knowles, A.M. Kropinski, M. Krupovic, J.H. Kuhn, M. Nibert, L. Rubino, S. Sabanadzovic, P. Simmonds, A. Varsani, F.M. Zerbini, A.J. Davison, Changes to virus taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2019), Arch. Virol. 164 (2019) 2417–2429, https://doi.org/10.1007/s00705-019-04306-w.
[28] D. Baltimore, Expression of animal virus genomes, Bacteriol. Rev. 35 (1971) 235–241.
[29] N. Philippe, M. Legendre, G. Doutre, Y. Couté, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J.-M. Claverie, C. Abergel, Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes, Science 341 (2013) 281–286, https://doi.org/10.1126/science.1239181.
[30] S. Spandole, D. Cimponeriu, L.M. Berca, G. Mihăescu, Human anelloviruses: an update of molecular, epidemiological and clinical aspects, Arch. Virol. 160 (2015) 893–908, https://doi.org/10.1007/s00705-015-2363-9.
[31] P.M. Lieberman, Epigenetics and genetics of viral latency, Cell Host Microbe 19 (2016) 619–628, https://doi.org/10.1016/j.chom.2016.04.008.
[32] M. Kane, T. Golovkina, Common threads in persistent viral infections, J. Virol. 84 (2010) 4116–4123, https://doi.org/10.1128/JVI.01905-09.
[33] Coronaviridae, Fenners Vet. Virol. (2011) 393–413, https://doi.org/10.1016/B978-0-12-375158-4.00024-9.
[34] R.M. Buller, G.J. Palumbo, Poxvirus pathogenesis, Microbiol. Rev. 55 (1991) 80–122.
[35] V. Brázda, Y. Luo, M. Bartas, P. Kaura, O. Porubiaková, J. Šťastný, P. Pečinka, D. Verga, V. Da Cunha, T.S. Takahashi, P. Forterre, H. Myllykallio, M. Fojta, J.-L. Mergny, G-quadruplexes in the archaea domain, Biomolecules 10 (2020) 1349, https://doi.org/10.3390/biom10091349.
[36] E.-C. Shin, P.S. Sung, S.-H. Park, Immune responses and immunopathology in acute and chronic viral hepatitis, Nat. Rev. Immunol. 16 (2016) 509–523, https://doi.org/10.1038/nri.2016.69.
[37] C.L. Charlton, E. Babady, C.C. Ginocchio, T.F. Hatchette, R.C. Jerris, Y. Li, M. Loeffelholz, Y.S. McCarter, M.B. Miller, S. Novak-Weekley, A.N. Schuetz, Y.-W. Tang, R. Widen, S.J. Drews, Practical guidance for clinical microbiology laboratories: viruses causing acute respiratory tract infections, Clin. Microbiol. Rev. 32 (2018), https://doi.org/10.1128/CMR.00042-18.

[38] E. van der Vries, K.J. Stittelaar, G. van Amerongen, E.J.B. Veldhuis Kroeze, L. de Waal, P.L.A. Fraaij, R.J. Meesters, T.M. Luider, B. van der Nagel, B. Koch, A.G. Vulto, M. Schutten, A.D.M.E. Osterhaus, Prolonged influenza virus shedding and emergence of antiviral resistance in immunocompromised patients and ferrets, PLoS Pathog. 9 (2013), e1003343, https://doi.org/10.1371/journal.ppat.1003343.
[39] R. Gonzalez-Dosal, K.A. Horan, S.H. Rahbek, H. Ichijo, Z.J. Chen, J.J. Mieyal, R. Hartmann, S.R. Paludan, HSV infection induces production of ROS, which potentiate signaling from pattern recognition receptors: role for S-glutathionylation of TRAF3 and 6, PLoS Pathog. 7 (2011), e1002250, https://doi.org/10.1371/journal.ppat.1002250.
[40] G.A. Maglennon, J. Doorbar, The biology of papillomavirus latency, Open Virol. J. 6 (2012) 190–197, https://doi.org/10.2174/1874357901206010190.
[41] T. Lion, Adenovirus persistence, reactivation, and clinical management, FEBS Lett. 593 (2019) 3571–3582, https://doi.org/10.1002/1873-3468.13576.
[42] J.S. Niczyporuk, Adenoviruses and their diversity in poultry, Appl. Genet. Genomics Poult. Sci. 7 (2018) 103–118, https://doi.org/10.5772/intechopen.77131.
[43] Family - Polyomaviridae, in: A.M.Q. King, M.J. Adams, E.B. Carstens, E.J. Lefkowitz (Eds.), Virus Taxon., Elsevier, San Diego, 2012, pp. 279–290, https://doi.org/10.1016/B978-0-12-384684-6.00027-6.
[44] T.L. Goldberg, V.L. Clyde, A. Gendron-Fitzpatrick, S.D. Sibley, R. Wallace, Severe neurologic disease and chick mortality in crested screamers (Chauna torquata) infected with a novel Gyrovirus, Virology 520 (2018) 111–115, https://doi.org/10.1016/j.virol.2018.05.014.
[45] K. Dhama, R.S. Chauhan, M. Mahendran, S.V.S. Malik, Rotavirus diarrhea in bovines and other domestic animals, Vet. Res. Commun. 33 (2009) 1–23, https://doi.org/10.1007/s11259-008-9070-x.
[46] S. Roy, L.H. Vandenberghe, S. Kryazhimskiy, R. Grant, R. Calcedo, X. Yuan, M. Keough, A. Sandhu, Q. Wang, C.A. Medina-Jaszek, J.B. Plotkin, J.M. Wilson, Isolation and characterization of adenoviruses persistently shed from the gastrointestinal tract of non-human primates, PLoS Pathog. 5 (2009), e1000503, https://doi.org/10.1371/journal.ppat.1000503.
[47] F. Krammer, G.J.D. Smith, R.A.M. Fouchier, M. Peiris, K. Kedzierska, P.C. Doherty, P. Palese, M.L. Shaw, J. Treanor, R.G. Webster, A. García-Sastre, Influenza, Nat. Rev. Dis. Primer. 4 (2018) 3, https://doi.org/10.1038/s41572-018-0002-y.
[48] F.B. Lestari, S. Vongpunsawad, N. Wanlapakorn, Y. Poovorawan, Rotavirus infection in children in Southeast Asia 2008–2018: disease burden, genotype distribution, seasonality, and vaccination, J. Biomed. Sci. 27 (2020) 66, https://doi.org/10.1186/s12929-020-00649-8.
[49] M. Weidner-Glunde, E. Kruminis-Kaszkiel, M. Savanagouder, Herpesviral latency—common themes, Pathogens 9 (2020) 125, https://doi.org/10.3390/pathogens9020125.
[50] K. Murray, C. Walker, E. Herrington, J.A. Lewis, J. McCormick, D.W.C. Beasley, R.B. Tesh, S. Fisher-Hoch, Persistent infection with west nile virus years after initial infection, J. Infect. Dis. 201 (2010) 2–4, https://doi.org/10.1086/648731.
[51] L. Mlera, W. Melik, M.E. Bloom, The role of viral persistence in flavivirus biology, Pathog. Dis. 71 (2014) 137–163, https://doi.org/10.1111/2049-632X.12178.
[52] Family - Reoviridae, in: A.M.Q. King, M.J. Adams, E.B. Carstens, E.J. Lefkowitz (Eds.), Virus Taxon., Elsevier, San Diego, 2012, pp. 541–637, https://doi.org/10.1016/B978-0-12-384684-6.00051-3.
[53] A. Mrzljak, I. Tabain, H. Premac, M. Bogdanic, L. Barbic, V. Savic, V. Stevanovic, A. Jelic, D. Mikulic, T. Vilibic-Cavlek, The role of emerging and neglected viruses in the etiology of hepatitis, Curr. Infect. Dis. Rep. 21 (2019) 51, https://doi.org/10.1007/s11908-019-0709-2.
[54] F.A. Lempp, S. Urban, Hepatitis delta virus: replication strategy and upcoming therapeutic options for a neglected human pathogen, Viruses 9 (2017) 172, https://doi.org/10.3390/v9070172.
[55] T. Horvatits, J. Schulze zur Wiesch, M. Lütgehetmann, A.W. Lohse, S. Pischke, The clinical perspective on hepatitis E, Viruses 11 (2019) 617, https://doi.org/10.3390/v11070617.
[56] H.L. Lightfoot, T. Hagen, N.J. Tatum, J. Hall, The diverse structural landscape of quadruplexes, FEBS Lett. 593 (2019) 2083–2102, https://doi.org/10.1002/1873-3468.13547.
[57] G.P. Martelli, S. Sabanadzovic, N. Abou Ghanem-Sabanadzovic, P. Saldarelli, Maculavirus, a new genus of plant viruses, Arch. Virol. 147 (2002) 1847–1853, https://doi.org/10.1007/s007050200046.
[58] R. Perrone, M. Nadai, I. Frasson, J.A. Poe, E. Butovskaya, T.E. Smithgall, M. Palumbo, G. Palù, S.N. Richter, A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter, J. Med. Chem. 56 (2013) 6521–6530, https://doi.org/10.1021/jm400914r.
[59] A.M. Fleming, N.L.B. Nguyen, C.J. Burrows, Colocalization of m6A and G-quadruplex-forming sequences in viral RNA (HIV, Zika, hepatitis B, and SV40) suggests topological control of adenosine N6-methylation, ACS Cent. Sci. 5 (2019) 218–228, https://doi.org/10.1021/acscentsci.8b00963.
[60] R. Perrone, M. Nadai, J.A. Poe, I. Frasson, M. Palumbo, G. Palù, T.E. Smithgall, S.N. Richter, Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity, PloS One 8 (2013), e73121, https://doi.org/10.1371/journal.pone.0073121.
[61] B. Biswas, P. Kumari, P. Vivekanandan, Pac1 signals of human Herpesviruses contain a highly conserved G-quadruplex motif, ACS Infect. Dis. 4 (2018) 744–751, https://doi.org/10.1021/acsinfecdis.7b00279.
[62] S. Artusi, M. Nadai, R. Perrone, M.A. Biasolo, G. Palù, L. Flamand, A. Calistri, S.N. Richter, The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: implications for the antiviral activity of a G-quadruplex ligand, Antivir. Res. 118 (2015) 123–131, https://doi.org/10.1016/j.antiviral.2015.03.016.
[63] H. Deng, B. Gong, Z. Yang, Z. Li, H. Zhou, Y. Zhang, X. Niu, S. Liu, D. Wei, Intensive distribution of G2-quaduplexes in the pseudorabies virus genome and their sensitivity to cations and G-quadruplex ligands, Molecules 24 (2019) 774, https://doi.org/10.3390/molecules24040774.
[64] S. Kumar, D. Choudhary, A. Patra, N.S. Bhavesh, P. Vivekanandan, Analysis of G-quadruplexes upstream of herpesvirus miRNAs: evidence of G-quadruplex mediated regulation of KSHV miR-K12-1-9,11 cluster and HCMV miR-US33, BMC Mol. Cell Biol. 21 (2020) 67, https://doi.org/10.1186/s12860-020-00306-w.
[65] P. Murat, J. Zhong, L. Lekieffre, N.P. Cowieson, J.L. Clancy, T. Preiss, S. Balasubramanian, R. Khanna, J. Tellam, G-quadruplexes regulate Epstein-Barr virus–encoded nuclear antigen 1 mRNA translation, Nat. Chem. Biol. 10 (2014) 358–364, https://doi.org/10.1038/nchembio.1479.
[66] P. Majee, U. Shankar, S. Pasadi, K. Muniyappa, D. Nayak, A. Kumar, Genome-wide analysis reveals a regulatory role for G-quadruplexes during Adenovirus multiplication, Virus Res. (2020) 197960, https://doi.org/10.1016/j.virusres.2020.197960.
[67] B. Biswas, M. Kandpal, U.K. Jauhari, P. Vivekanandan, Genome-wide analysis of G-quadruplexes in herpesvirus genomes, BMC Genom. 17 (2016) 949, https://doi.org/10.1186/s12864-016-3282-1.
[68] K. Tlučková, M. Marušič, P. Tóthová, L. Bauer, P. Šket, J. Plavec, V. Viglasky, Human papillomavirus G-quadruplexes, Biochemistry 52 (2013) 7207–7216, https://doi.org/10.1021/bi400897g.
[69] N. Bohálová, A. Cantara, M. Bartas, P. Kaura, J. Šťastný, P. Pečinka, M. Fojta, V. Brázda, Tracing dsDNA virus–host coevolution through correlation of their G-quadruplex-forming sequences, Int. J. Mol. Sci. 22 (2021) 3433, https://doi.org/10.3390/ijms22073433.
[70] C. Jaubert, A. Bedrat, L. Bartolucci, C. Di Primo, M. Ventura, J.-L. Mergny, S. Amrane, M.-L. Andreola, RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand, Sci. Rep. 8 (2018) 8120, https://doi.org/10.1038/s41598-018-26582-3.
[71] S.T. Jacob, I. Crozier, W.A. Fischer, A. Hewlett, C.S. Kraft, M.-A. de L. Vega, M.J. Soka, V. Wahl, A. Griffiths, L. Bollinger, J.H. Kuhn, Ebola virus disease, Nat. Rev. Dis. Primer. 6 (2020) 13, https://doi.org/10.1038/s41572-020-0147-3.
[72] K.M. Coffin, J. Liu, T.K. Warren, C.D. Blancett, K.A. Kuehl, D.K. Nichols, J.J. Bearss, C.W. Schellhase, C.J. Retterer, J.M. Weidner, S.R. Radoshitzky, J.M. Brannan, A.P. Cardile, J.M. Dye, G. Palacios, M.G. Sun, J.H. Kuhn, S. Bavari, X. Zeng, Persistent Marburg virus infection in the testes of nonhuman primate survivors, Cell Host Microbe 24 (2018) 405–416, https://doi.org/10.1016/j.chom.2018.08.003, e3.
[73] S.-R. Wang, Q.-Y. Zhang, J.-Q. Wang, X.-Y. Ge, Y.-Y. Song, Y.-F. Wang, X.-D. Li, B.-S. Fu, G.-H. Xu, B. Shu, P. Gong, B. Zhang, T. Tian, X. Zhou, Chemical targeting of a G-quadruplex RNA in the Ebola virus L gene, Cell Chem. Biol. 23 (2016) 1113–1122, https://doi.org/10.1016/j.chembiol.2016.07.019.
[74] P. Krafčíková, E. Demkovičová, V. Víglaský, Ebola virus derived G-quadruplexes: thiazole orange interaction, Biochim. Biophys. Acta BBA - Gen. Subj. 1861 (2017) 1321–1328, https://doi.org/10.1016/j.bbagen.2016.12.009.
[75] D.E. Griffin, W.-H.W. Lin, A.N. Nelson, Understanding the causes and consequences of measles virus persistence, F1000Research 7 (2018) 237, https://doi.org/10.12688/f1000research.12094.1.
[76] L. Hároníková, J. Coufal, I. Kejnovská, E.B. Jagelská, M. Fojta, P. Dvořáková, P. Muller, B. Vojtesek, V. Brázda, IFI16 preferentially binds to DNA with quadruplex structure and enhances DNA quadruplex formation, PloS One 11 (2016), e0157156, https://doi.org/10.1371/journal.pone.0157156.
[77] G. Pisano, A. Roy, M. Ahmed Ansari, B. Kumar, L. Chikoti, B. Chandran, Interferon-γ-inducible protein 16 (IFI16) is required for the maintenance of Epstein-Barr virus latency, Virol. J. 14 (2017) 221, https://doi.org/10.1186/s12985-017-0891-5.
[78] V. Brázda, O. Porubiaková, A. Cantara, N. Bohálová, J. Coufal, M. Bartas, M. Fojta, J.-L. Mergny, G-quadruplexes in H1N1 influenza genomes, BMC Genom. 22 (2021) 77, https://doi.org/10.1186/s12864-021-07377-9.
[79] M. Bartas, V. Brázda, N. Bohálová, A. Cantara, A. Volná, T. Stachurová, K. Malachová, E.B. Jagelská, O. Porubiaková, J. Červeň, P. Pečinka, In-depth bioinformatic analyses of nidovirales including human SARS-CoV-2, SARS-CoV, MERS-CoV viruses suggest important roles of non-canonical nucleic acid structures in their lifecycles, Front. Microbiol. 11 (2020) 1583, https://doi.org/10.3389/fmicb.2020.01583.
[80] D. Ji, M. Juhas, C.M. Tsang, C.K. Kwok, Y. Li, Y. Zhang, Discovery of G-quadruplex-forming sequences in SARS-CoV-2, brief, Bioinformation 22 (2021) 1150–1160, https://doi.org/10.1093/bib/bbaa114.
[81] D. Piekna-Przybylska, G. Sharma, S.B. Maggirwar, R.A. Bambara, Deficiency in DNA damage response, a new characteristic of cells infected with latent HIV-1, Cell Cycle 16 (2017) 968–978, https://doi.org/10.1080/

15384101.2017.1312225.

[82] J.T. Tellam, J. Zhong, L. Lekieffre, P. Bhat, M. Martinez, N.P. Croft, W. Kaplan, R.L. Tellam, R. Khanna, mRNA Structural constraints on EBNA1 synthesis impact on in vivo antigen presentation and early priming of CD8+ T cells, PLoS Pathog. 10 (2014), e1004423, https://doi.org/10.1371/journal.ppat.1004423.

[83] P. Dabral, J. Babu, A. Zareie, S.C. Verma, LANA and hnRNP A1 regulate the translation of LANA mRNA through G-quadruplexes, J. Virol. 94 (2019) e01508–e01519, https://doi.org/10.1128/JVI.01508-19.

Příloha G

Tracing dsDNA Virus-Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences

International Journal of
Molecular Sciences

MDPI

Article

Tracing dsDNA Virus–Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences

Natália Bohálová [1,2], Alessio Cantara [1,2], Martin Bartas [3], Patrik Kaura [4], Jiří Šťastný [4,5], Petr Pečinka [3], Miroslav Fojta [1] and Václav Brázda [1,*]

1 Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic; natalia.bohalova@ibp.cz (N.B.); alexcantara41@gmail.com (A.C.); fojta@ibp.cz (M.F.)
2 Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic
3 Department of Biology and Ecology, Institute of Environmental Technologies, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic; martin.bartas@osu.cz (M.B.); petr.pecinka@osu.cz (P.P.)
4 Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896/2, 616 69 Brno, Czech Republic; 160702@vutbr.cz (P.K.); stastny@fme.vutbr.cz (J.Š.)
5 Department of Informatics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
* Correspondence: vaclav@ibp.cz; Tel.: +420-541517231; Fax: +420-541211293

Abstract: The importance of gene expression regulation in viruses based upon G-quadruplex may point to its potential utilization in therapeutic targeting. Here, we present analyses as to the occurrence of putative G-quadruplex-forming sequences (PQS) in all reference viral dsDNA genomes and evaluate their dependence on PQS occurrence in host organisms using the G4Hunter tool. PQS frequencies differ across host taxa without regard to GC content. The overlay of PQS with annotated regions reveals the localization of PQS in specific regions. While abundance in some, such as repeat regions, is shared by all groups, others are unique. There is abundance within introns of Eukaryota-infecting viruses, but depletion of PQS in introns of bacteria-infecting viruses. We reveal a significant positive correlation between PQS frequencies in dsDNA viruses and corresponding hosts from archaea, bacteria, and eukaryotes. A strong relationship between PQS in a virus and its host indicates their close coevolution and evolutionarily reciprocal mimicking of genome organization.

Keywords: G-quadruplex; virus; bioinformatics; coevolution; host; dsDNA; G4Hunter

check for updates

Citation: Bohálová, N.; Cantara, A.; Bartas, M.; Kaura, P.; Šťastný, J.; Pečinka, P.; Fojta, M.; Brázda, V. Tracing dsDNA Virus–Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences. *Int. J. Mol. Sci.* **2021**, *22*, 3433. https://doi.org/10.3390/ijms22073433

Academic Editor: James K. Bashkin

Received: 2 March 2021
Accepted: 23 March 2021
Published: 26 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Viruses are intracellular parasites closely coevolving with their host organisms and thus shaping genotypic diversity [1,2]. The interplay between a virus and its host constitutes a powerful mechanism of reciprocal selection pressure. Coevolution of the two can be traced by nucleic acid sequence, protein tertiary structure, and also at the whole-function level. For example, hosts' antiviral defense mechanisms often originate from viruses [2–4]. The study of reciprocal coevolutionary adaptations between a virus and the host immune system could provide new insights and potential strategies in developing antiviral treatments [5].

G-quadruplexes (G4s) are noncanonical, local secondary structures of nucleic acids that have been identified as having regulatory roles within cells in gene expression, replication, and telomere maintenance [6–8]. A G4 consists of stacked planar G-quartets, which are built by Hoogsteen hydrogen bond-based pairing of four guanines. It has been demonstrated that G4s are very often targets for various cellular proteins [9–11], and a specific domain for G4 recognition has been shown [12,13]. Moreover, several proteins are also capable of stabilizing the G4 structure [14,15]. Recently, it has been demonstrated that the G4 binding domain is also conserved between SARS-CoV and SARS-CoV-2 genomes [16], and it was proven to be crucial for the SARS-CoV life cycle [17]. G4s can be found in all domains of

life [18–21], and they have been described as constituting an important structural genomic feature with various functions in several viral classes [21,22]. The G4 formation was shown to limit the replication and transcription of the Ebola virus, hepatitis B virus (HBV), hepatitis C virus (HCV), human immunodeficiency virus (HIV), and several viruses from the *Herpesviridae* family [23–26]. In the life cycles of the Epstein–Barr virus (EBV) and Kaposi sarcoma herpesvirus (KSHV), moreover, RNA G4 has been described as a cis-acting regulatory element that downregulates the translation of highly antigenic proteins and thus influences the immune evasion of the virus and eases transit and persistence in the latent period of infection [27,28]. Importantly, the functions of G4 may be modified via their stabilization by proteins or small-molecule ligands [29,30]. Therefore, stabilizing G4 ligands are considered promising antiviral and antibacterial drugs [29]. Coevolution of viral and host loop sequences of the G-quadruplex-forming sequences in human *Herpesviridae* viruses was recently proposed [31].

G4Hunter is one of several software programs available for predicting putative G-quadruplex forming sequences (PQS) [32–34]. The G4Hunter algorithm searches for Gs/Cs and sums up the scores for the groups of bases. The final score is thus a combination of G-richness and G-skewness and the presence of G-blocks. The default threshold is set to a G4Hunter score of 1.2, which has proven to be a reasonable compromise between false-negative and false-positive results. The higher the score, the higher the probability for a G4 structure to form [33]. G4Hunter provides the benefit of targeting even atypical G4s that could not be found by pattern-based algorithms [35,36].

Here, we present an extensive analysis of 2903 viruses across a diverse range of host organisms. Our goals were to identify PQS occurrence and localization in the genomes of viruses infecting a given host group, study the evolutionary differences related to PQS, and describe the potential dependence of PQS frequency between a virus and the corresponding host.

2. Results and Discussion

2.1. Variation in Frequency for G4-Forming Sequences in dsDNA Viruses Grouped by Host

Using the National Center for Biotechnology Information (NCBI) taxonomy classification, the analyzed viruses were divided into three domains according to their host organisms: Archaea, Bacteria, and Eukaryota. The domains were further divided into 23 groups (12 with five or more sequenced genomes) as shown in the phylogenetic tree in Figure 1. Phylogenetic classification of the viruses and corresponding host is presented in Supplementary Materials S1. All hosts were assigned by the Virus-Host database without further modification [37], which could have limited the potential host range, especially in arboviruses. Whereas 95% of all known bacteriophages and archaea-infecting viruses have dsDNA genomes [38], eukaryotes could be infected by all classes of the Baltimore virus classification, which means, in addition to dsDNA viruses, also ssDNA viruses, dsRNA viruses, ssRNA viruses, ssRNA reverse-transcribing viruses, and dsDNA reverse-transcribing viruses. We therefore restricted the analyses of PQS occurrence to only dsDNA viruses, although they have not been found to infect higher plants belonging to the Streptophyta but only lower species of plants belonging to Chlorophyta [39,40].

For further statistical analyses, only those groups with five or more sequenced genomes were included. We analyzed the PQS occurrence in all 2903 reference dsDNA viral genomes divided according to their host organisms (Supplementary Materials S2). A summary of all PQS found in ranges of the G4Hunter score intervals (1.2–1.4, 1.4–1.6, 1.6–1.8, 1.8–2.0, and 2.0 and higher) and precomputed PQS frequencies per 1000 nt is shown in Table 1.

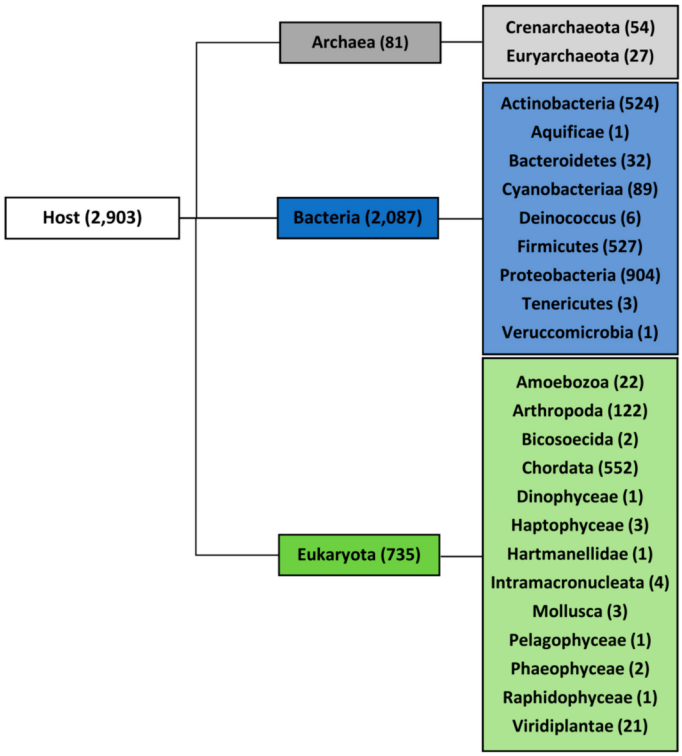


Figure 1. Full set of viral genomes divided according to host. The number of accessible unique genomes for each domain and group is shown in brackets.

Table 1. Total number of putative G-quadruplex-forming sequences (PQS) and their resulting frequencies per 1000 nt in all 2903 viral genomes and host categories, grouped by G4Hunter score. Frequencies were calculated as the total number of PQS in each category divided by the total length of all analyzed sequences, multiplied by 1000 and normalized by the number of viruses infecting one genus.

G4Hunter Score	PQS Frequency per 1000 nt			
	All	Archaea	Bacteria	Eukaryota
1.2–1.4	1.27	1.74	0.88	1.46
1.4–1.6	0.039	0.025	0.026	0.047
1.6–1.8	0.0042	0	0.00088	0.0062
1.8–2.0	0.00025	0	0.000041	0.00038
2.0 and more	0.00021	0	0.000050	0.00031

The mean frequency for all viral genomes in G4Hunter score interval 1.2–1.4 was 1.27 PQS per 1000 nt (see above). The lowest frequency in the same interval was observed for bacteriophages (0.88 PQS per 1000 nt), whereas the highest frequency was detected for archaea-infecting viruses (1.74). Surprisingly, there was not one PQS with a G4Hunter score higher than 1.4 found in the archaea host domain. In the Bacteria and Eukaryota host domains, by contrast, there were some PQS found even with G4Hunter scores higher than 2.0.

The numbers of analyzed viral sequences, grouped by their host phylogenetic categories; median genome length; mean, minimum, and maximum observed frequency of PQS per 1000 nt; and total PQS counts are summarized in Table 2. Just four viral groups (viruses infecting Euryarchaeota, Actinobacteria, *Deinococcus*, and Proteobacteria) showed >50% GC content. On the other hand, three groups (viruses infecting Bacteroidetes, Firmicutes, and Arthropoda) showed <40% GC content. Detailed statistical characteristics for PQS frequencies per 1000 nt (including mean, variance, and outliers) are depicted in boxplots for all inspected host groups in Figure 2. The mean

frequency for all viral genomes was 1.32 PQS per 1000 nt. Detailed statistical analyses of host inter-domain and intergroup comparisons are presented in Supplementary Materials S3. We observed the highest mean frequency in the archaea host domain (1.76), followed by viruses infecting Eukaryota (1.52), whereas the lowest was noted in bacteriophages (0.89). At the group level, the most extreme values were found within the viruses infecting the bacteria domain. The lowest mean frequency was found in viruses infecting the Firmicutes (0.32) followed by Bacteroidetes (0.41). The highest PQS frequency was observed in the *Deinococcus* host group (4.21), followed by Actinobacteria (2.27). In viruses infecting the archaea domain, notable enrichment relative to the average was found for both the Euryarchaeota (1.69) and Crenarchaeota (1.85) groups. Within the Eukaryota host domain, the highest PQS frequency was observed for Chordata (2.18), the lowest for Arthropoda (0.30). The mean PQS frequency found in viruses infecting humans was lower (1.75) than the average PQS in the Chordata host group as normalized by the number of viruses infecting one host genus (2.18). We created a cluster dendrogram, as shown in Figure 3, to further reveal and graphically depict similarities among host groups. The input data and R code are listed in Supplementary Materials S4. Viruses infecting humans are notably clustered together with other viruses infecting Chordata on the left side of the dendrogram. Other viruses infecting eukaryotes are clustered in the second branch on the right.

Table 2. Distribution of PQS frequencies in viruses according to host organisms. Genomic length, PQS frequencies, and total counts. Seq (total number of sequences), Median (median length of sequences), GC% (average GC content), PQS (total number of predicted PQS), Mean f (mean frequency of predicted PQS per 1000 nt normalized by the number of viruses infecting one genus), Min f (the lowest frequency of predicted PQS per 1000 nt), Max f (the highest frequency of predicted PQS per 1000 nt), and Cov (% of genome covered by PQS).

All	Seq	Median	GC%	PQS	Mean f	Min f	Max f	Cov
All	3134	44,746.5	44.94	220,569	1.32	0	11.51	3.34
Domain	**Seq**	**Median**	**GC%**	**PQS**	**Mean f**	**Min f**	**Max f**	**Cov**
Archaea	81	33,356	48.92	3137	1.76	0	4.80	4.32
Bacteria	2087	49,639	48.10	112,664	0.89	0	11.51	2.11
Eukaryota	966	7951.5	43.09	104,768	1.52	0	11.44	3.93
Group	**Seq**	**Median**	**GC%**	**PQS**	**Mean f**	**Min f**	**Max f**	**Cov**
Crenarchaeota	54	32,047.5	40.91	1012	1.85	0	4.80	4.76
Euryarchaeota	27	49,107	54.92	2125	1.69	0.28	3.75	3.99
Actinobacteria	524	53,403.5	60.90	61,313	2.27	0.33	7.02	5.12
Bacteroidetes	32	47,060	38.12	477	0.41	0.03	1.14	1.01
Cyanobacteria	89	174,079	43.33	3875	0.82	0.06	3.88	2.10
Deinococcus	6	61,150	50.26	726	4.21	0.33	11.51	10.45
Firmicutes	527	41,843	38.14	7886	0.32	0	1.39	0.78
Proteobacteria	904	49,035	50.07	38,334	0.80	0	4.55	1.90
Amoebozoa	22	495,022	42.47	21,931	0.66	0	1.89	1.60
Arthropoda	345	7276	38.77	4957	0.30	0	1.92	0.73
Chordata	561	7852	45.48	72,420	2.18	0	11.44	5.65
Viridiplantae	21	193,301	46.91	3542	1.06	0	2.01	2.54
Subgroup	**Seq**	**Median**	**GC%**	**PQS**	**Mean f**	**Min f**	**Max f**	**Cov**
Humans	120	7344	42.55	15,996	1.75	0	11.44	4.48

The colors correspond to phylogenetic tree depiction in Figure 1 (Grey—Archaea, Blue—Bacteria, Green—Eukaryota as host organisms).

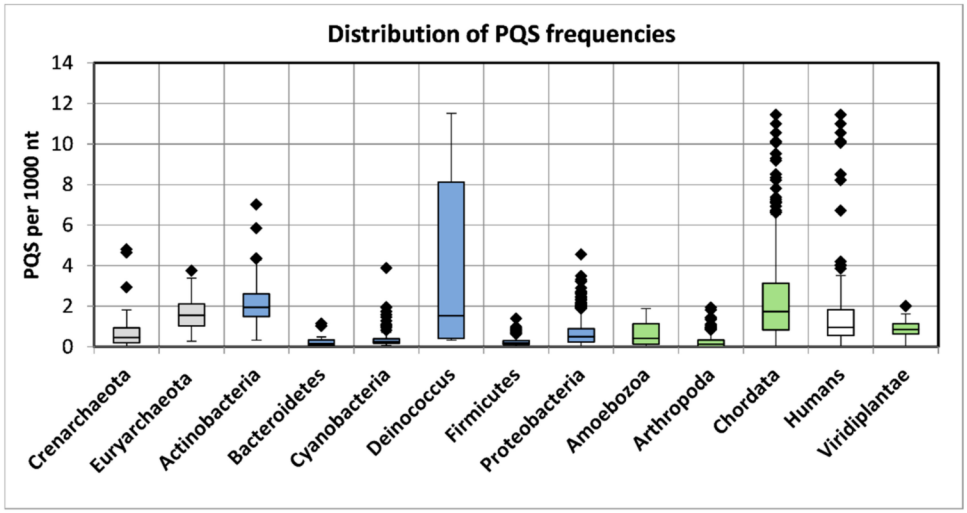


Figure 2. Frequencies of PQS in host groups of the analyzed viral genomes. Data within boxes span the interquartile range and whiskers show the lowest and highest values within the 1.5 interquartile range. Black diamonds denote outliers. The colors correspond to phylogenetic tree depiction.

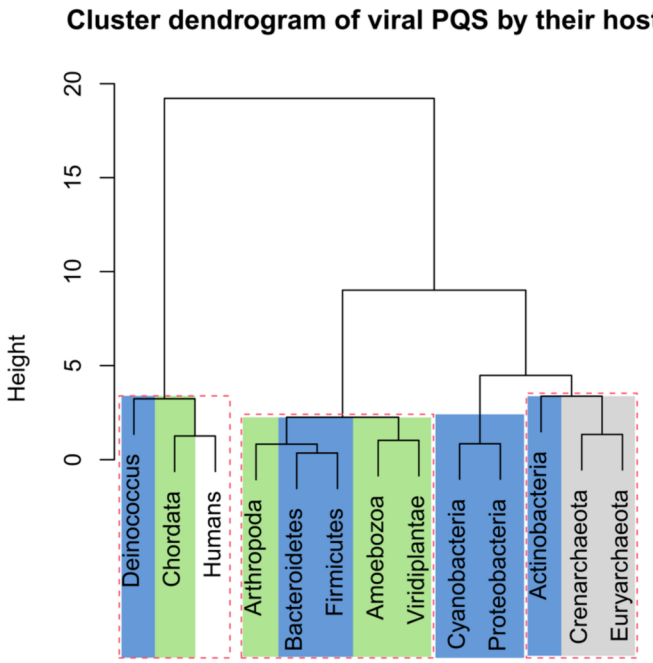


Figure 3. Cluster dendrogram based on PQS characteristics in all viral species by their host. Input data are listed in Supplementary Materials S4. Statistically significant clusters (based upon approximately unbiased *p*-values above 95, equivalent to *p*-values lower than 0.05) are highlighted by rectangles drawn with broken red lines. The colors correspond to phylogenetic tree depiction.

2.2. Features Characteristic for Hosts Are Enriched for PQS in Corresponding dsDNA Viral Genomes

To evaluate the localization of PQS within viral genomes, we overlapped PQS with annotation regions extracted from the NCBI database (Supplementary Materials S5). We took the PQS frequency per 1000 nt in genes as a reference and plotted the ratio of the PQS frequency in features to that in genes (Figure 4). PQS frequencies differ depending on the annotated motif and across different hosts. In the archaea domain, the most notable enrichment was found inside and 100 bp after *stem_loops* (4.2× and 10.2× enrichment) and *mobile_elements* (3.5× and 3.4× enhancement). Predictably, abundance was also found in the archaea-infecting viruses' *repeat_regions* (2.9×). The *repeat_regions* were also enriched for

PQS inside bacteria-infecting (3.3×) and Eukaryota-infecting viruses (4.4×). The highest relative frequency inside bacteria was found in *tmRNA* (3.13) and *ncRNA* (1.9×), followed by a region 100 bp long before *misc_RNA* and *misc_recomb* (1.8 and 1.5× abundance). In addition to *repeat_regions*, we noted PQS enrichment in *misc_RNA* (6.7×), *variation* (5.2×), *protein_bind* (3.8×), and *introns* (1.9×) of Eukaryota-infecting viruses. Notably, the PQS frequency was increased in comparison to *genes* inside *introns* only in Eukaryota-infecting viruses (1.9× enrichment), whereas *introns* in bacteria-infecting viruses were depleted for PQS presence (0.14× lower PQS frequency in comparison to genes). This indicates that the prevalence of PQS in specific viral features is important for the host's cellular machinery. A G4 located in an intron could affect the expression profile; it was shown to regulate the splicing of alternative isoforms of a p53 protein in the human genome [41].

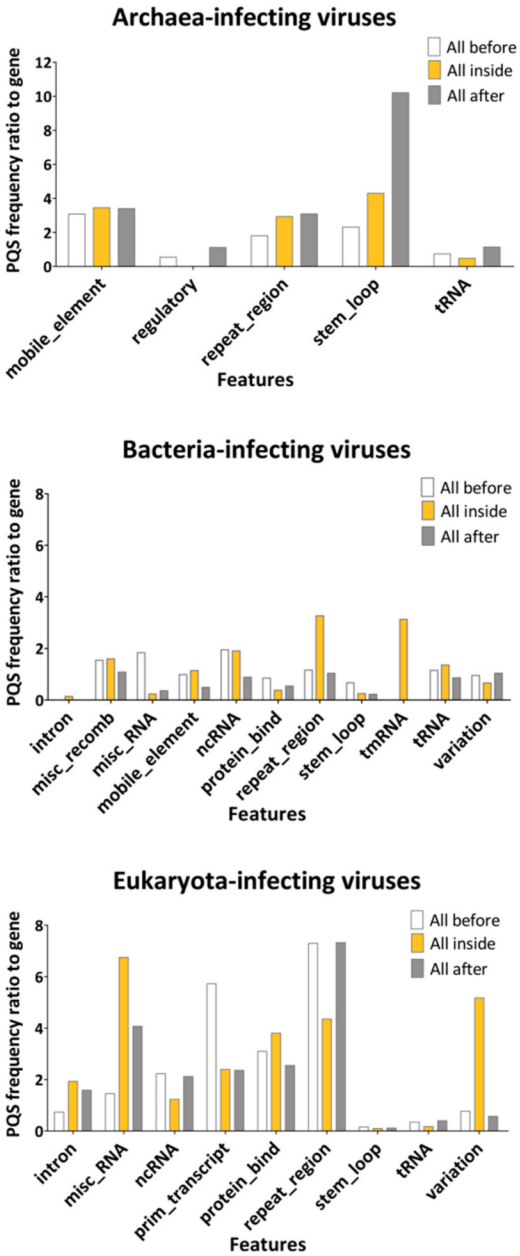


Figure 4. The ratio of PQS frequencies per 1000 nt between gene annotation and other annotated locations from the NCBI database. PQS frequencies within (inside), before (100 nt), and after (100 nt) annotated locations were analyzed. Detailed results are summarized in Supplementary Materials S5.

2.3. PQS Frequencies of dsDNA Viruses Correlate with Their Hosts' Genomes

Next, to evaluate the relationship between PQS frequencies of the virus and the host, we analyzed selected genomes of host organisms for PQS presence. For hosts from archaea and bacteria, we utilized the previously published results of our group on PQS occurrence in all accessible archaeal and bacterial genomes [19,42]. In all analyses, we used the same workflow and same parameters for G4Hunter and data processing. Reference genomes of the Eukaryota hosts were retrieved from the NCBI database, and their list together with correlation analyses is available in Supplementary Materials S1. We selected all available reference genomes of hosts belonging to Viridiplantae, and for the remaining Eukaryota groups (Arthropoda and Chordata), we selected the 10 most frequent hosts in each corresponding category. There is no reference genome, however, for the *Acanthamoeba* genus, the only host of Amoebozoa-infecting viruses. We always compared a single eukaryotic host genome to all corresponding dsDNA viruses. The overall results of the correlation analyses are presented in Figure 5 and in the Supplementary Material S6. Spearman's correlation coefficient for the average of PQS frequencies in all investigated virus-host pairs was determined to be 0.7677, with a statistically significant p-value of 7×10^{-7} (Figure 5A). To exclude the GC content as a bias factor, we plotted also the average PQS/GC per 1000 nt. The correlation coefficient for the average of PQS/GC per 1000 nt then increased to 0.822, with p-value of 3×10^{-8} (Figure 5B).

The strongest correlation was found between virus–bacteria pairs. Spearman's correlation coefficient for PQS frequency of bacteria-infecting viruses and their hosts showed a strong, statistically significant (p-value < 0.01) positive correlation (Figure 5E,F). The correlation coefficient was 0.9429 for PQS frequency and 0.9985 for GC/PQS, with p-value < 0.01. Our previously published PQS frequencies of all known bacterial genomes [19] correspond to the frequencies determined here for PQS in bacteriophage genomes. In all virus–bacteria pairs, the mean PQS frequency was higher in the bacterial group than in the viral host group. A statistically significant positive correlation (p-value < 0.05) was observed, also with PQS frequencies grouped by G4Hunter score intervals and PQS frequencies identified by the Tetraplex Finder module of QuadBase2 software with default low stringency parameters (Supplementary Materials S7). Dispersion of PQS frequencies among bacteriophages was more diverse than inside other viruses, and the same observation (higher diversity in PQS frequencies) has been reported for bacteria compared to other hosts [18]. The corresponding frequencies of virus and bacteria hosts confirmed by correlation analyses pointed to their having close coevolutionary processes. The second highest correlation coefficient was found for the Archaea subgroup, with a value of 0.9 for PQS average frequency and PQS/GC (p-value < 0.05) (Figure 5C,D). To distinguish several different phyla, we further divided Crenarchaeota into two subgroups (*Sulfolobus*, Thermoproteales) and Euryarchaeota into three subgroups (*Acidianus*, Halobacteriales, Haloferacales). Because of the high diversity and the low number of sequenced genomes in several categories, the minimum number of viral or host genomes for statistical analyses was set to four.

Inside the eukaryote domain, we noted lower but still statistically significant (p-value < 0.01) correlation coefficients of 0.6509 for PQS frequencies and 0.7737 for PQS/GC. This finding could be attributed to two main causes (Figure 5G,H). First, the statistical sample for Eukaryota host genomes was significantly smaller, in comparison to that for bacteria and archaea host domains; an average of six host genomes were analyzed for each group of Eukaryota, shown in Table 2. Second, with the increasing complexity of the organisms, genomic duplications, and extensive repetitions, the correlation could be less obvious and significant.

Recently, coevolution of G4 sequence composition between dsDNA viruses from the *Herpesviridae* family and host has been proposed, as herpesviruses are often enriched for C-rich looped G4s, which are binding sites for host transcription factors, and TTA-looped G4s identical to the telomeric motif of vertebrates [31]. Mimicking the genome organization of the host could influence the PQS prevalence in dsDNA viral genomes and vice versa.

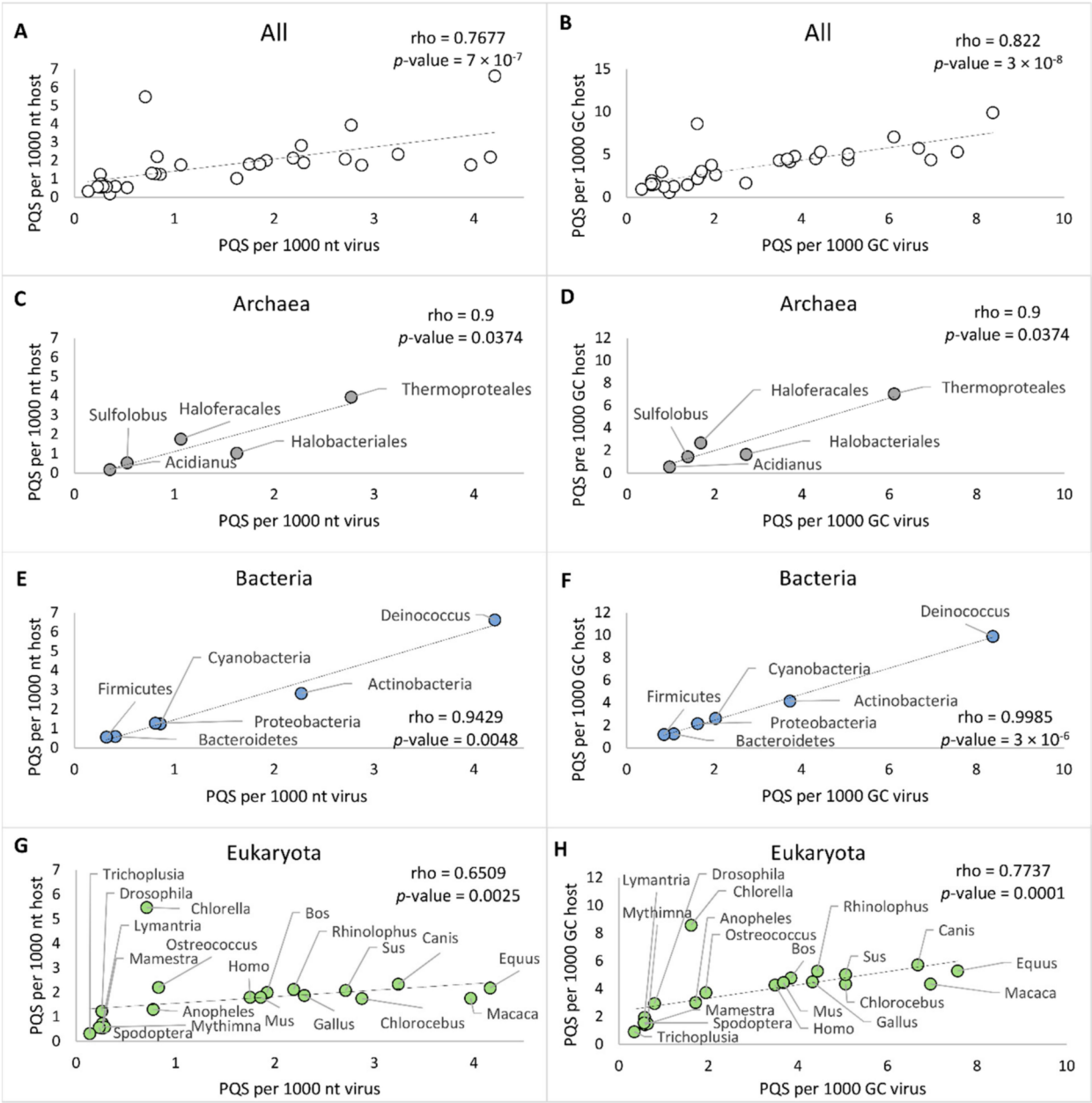


Figure 5. Relationships between virus and various hosts as measured by observed PQS frequency per 1000 nt and PQS frequency per 1000 GC. (**A**) All host-virus pairs, PQS frequencies; (**B**) All host–virus pairs, PQS per 1000 GC; (**C**) Archaea–virus pairs, PQS frequencies; (**D**) Archaea–virus pairs, PQS per 1000 GC; (**E**) Bacteria–virus pairs, PQS frequencies; (**F**) Bacteria–virus pairs, PQS per 1000 GC; (**G**) Eukaryota–virus pairs, PQS frequencies; (**H**) Eukaryota–virus pairs, PQS per 1000 GC of the archaea–virus pairs.

3. Materials and Methods

3.1. Viral and Host Sequences

A total of 3134 sequences of 2903 unique viral genomes were downloaded from the genome database of the National Center for Biotechnology Information (NCBI). Where more than one sequence was available, only the reference genome was used in the analyses. Hosts were assigned according to the NCBI and the Virus-Host database [37]. Subviral agents were assigned to the host of the coinfected virus as stated in the database [37]. A full list of NCBI IDs and host assignments are presented in Supplementary Materials S1. For hosts from archaea and bacteria, we utilized the previously published results of our group

on PQS occurrence in all accessible archaeal and bacterial genomes [19,42]. For eukaryote groups, we selected the 10 most frequented hosts for each viral group, and these are also listed in Supplementary Materials S1.

3.2. PQS Analyses

Analyses were run using the computational core of DNA analyzer software written in Java [43] with G4Hunter algorithm implementation [32] and default parameters (25 nt for window size, G4 score threshold 1.2). The overall results as to the number of PQS found together with the size of genomic DNA, GC content, PQS frequency normalized for 1000 nt, and lengths of sequences covered with PQS are summarized in Supplementary Materials S2. The average PQS frequency of host groups, shown in Table 1, was normalized by the number of viruses infecting each genus to avoid sampling bias due to the overabundance of viruses infecting specific species (such as a human). PQS were also classified into the five intervals of the G4Hunter score: 1.2–1.4, 1.4–1.6, 1.6–1.8, 1.8–2.0, and 2.0 and more. To confirm the results acquired by G4Hunter, bacteriophages and corresponding hosts were selected and analyzed by the Tetraplex Finder module of QuadBase2 software with default low stringency parameters (nonoverlapping PQS with minimum two-tracked PQS and loop length of 1–12 nt), and the results are listed in Supplementary Materials S7 [44].

3.3. Statistical Analysis

The normality of data was tested by the Shapiro–Wilk test. The nonparametric Kruskal–Wallis test was utilized for statistical evaluations of the differences in PQS among host phylogenetic groups. Post hoc multiple pairwise comparisons, using Dunn's test with Bonferroni correction of the significance level, were applied with the p-value cutoff set at 0.05; data are available in Supplementary Materials S3 for sequences grouped by their host organism and their comparison to the PQS frequency of host groups. For correlation analyses, the two-tailed Spearman's correlation coefficient was considered. To further reveal and graphically depict similarities between viral hosts, we constructed a cluster dendrogram of PQS characteristics in program R, version 3.6.3, using the pvclust package. The following values were used as input data: Mean f (mean of predicted PQS per 1000 nt), Min f (the lowest frequency of predicted PQS per 1000 nt), Max f (the highest frequency of predicted PQS per 1000 nt), and Cov % (% of genome covered by PQS) (Supplementary Materials S4). The following parameters were used for both analyses, the cluster method "ward.D2", distance "euclidean", and the number of bootstrap resampling 10,000. Statistically significant clusters (based on approximately unbiased p-values values above 95, equivalent to p-values less than 0.05) are highlighted in Figure 3 by rectangles marked with broken red lines. R code is provided in Supplementary Materials S4.

3.4. Overlay of PQS with Annotated Features from NCBI

Annotated feature tables of all viral genomes were downloaded from the NCBI database. Features were grouped by their names as stated in the feature table file. PQS occurrence was analyzed inside and around (before and after) a predefined featured neighborhood (±100 nt). From this analysis, we obtained a file with feature names and numbers of PQS found inside and around features. Further processing was performed in Microsoft Excel and the data are available as Supplementary Material S5.

4. Conclusions

PQS frequencies in viral genomes differ across host taxa and correspond to the PQS frequencies of the host organism. The overlay of PQS with annotated regions revealed nonrandom localization of G4 sequences and their abundance in various regions, such as repeat regions, stem-loops, mobile elements, protein-binding regions, RNA, etc. While abundance and depletion in some locations are shared by viruses of different hosts, others are unique. For example, there is an abundance of PQS in introns of Eukaryota-infecting viruses, but depletion of PQS in introns of bacteria-infecting viruses. Our study revealed

the correlation between PQS frequencies of dsDNA viruses and corresponding hosts from archaea, bacteria, and even eukaryotes, which indicate their close coevolution and evolutionarily reciprocal mimicking of genome organization.

Supplementary Materials: Supplementary Materials can be found at https://www.mdpi.com/article/10.3390/ijms22073433/s1.

Author Contributions: Conceptualization, V.B.; methodology, V.B.; software, P.K. and J.Š.; validation, V.B.; formal analysis, N.B., A.C. and M.B.; investigation, N.B. and A.C.; resources, P.K. and J.Š.; data curation, N.B. and A.C.; writing—original draft preparation, N.B., A.C., M.B. and V.B.; writing—review and editing, P.P. and M.F.; visualization, N.B., A.C. and M.B.; supervision, V.B., M.F. and P.P.; project administration, V.B.; and funding acquisition, V.B. and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by The Czech Science Foundation (18-15548S) and by the SYMBIT project Reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477 financed from the ERDF.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Gale A. Kirking for proofreading and providing feedback on the manuscript.

Conflicts of Interest: The authors declare no competing interests.

Abbreviations

EBV	Epstein–Barr virus
G4s	G-quadruplexes
HBV	hepatitis B virus
HCV	hepatitis C virus
HIV	human immunodeficiency virus
KSHV	Kaposi sarcoma herpes virus
PQS	putative G-quadruplex-forming sequences
NCBI	National Center for Biotechnology Information

References

1. McLaughlin, R.N.; Malik, H.S. Genetic conflicts: The usual suspects and beyond. *J. Exp. Biol.* **2017**, *220*, 6–17. [CrossRef]
2. Kaján, G.L.; Doszpoly, A.; Tarján, Z.L.; Vidovszky, M.Z.; Papp, T. Virus–Host Coevolution with a Focus on Animal and Human DNA Viruses. *J. Mol. Evol.* **2020**, *88*, 41–56. [CrossRef]
3. Charpentier, E.; Doudna, J.A. Rewriting a genome. *Nat. Cell Biol.* **2013**, *495*, 50–51. [CrossRef]
4. Moelling, K.; Broecker, F.; Russo, G.; Sunagawa, S. RNase H as Gene Modifier, Driver of Evolution and Antiviral Defense. *Front. Microbiol.* **2017**, *8*, 1745. [CrossRef] [PubMed]
5. Woolhouse, M.E.J.; Webster, J.P.; Domingo, E.; Charlesworth, B.; Levin, B.R. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat. Genet.* **2002**, *32*, 569–577. [CrossRef] [PubMed]
6. Lemarteleur, T.; Gomez, D.; Paterski, R.; Mandine, E.; Mailliet, P.; Riou, J.-F. Stabilization of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase. *Biochem. Biophys. Res. Commun.* **2004**, *323*, 802–808. [CrossRef] [PubMed]
7. Patel, D.J.; Phan, A.T.; Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **2007**, *35*, 7429–7455. [CrossRef]
8. Rhodes, D.; Lipps, H.J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015**, *43*, 8627–8637. [CrossRef] [PubMed]
9. Mishra, S.K.; Tawani, A.; Mishra, A.; Kumar, A. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* **2016**, *6*, 38144. [CrossRef] [PubMed]
10. Brázda, V.; Hároníková, L.; Liao, J.C.C.; Fojta, M. DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.* **2014**, *15*, 17493–17517. [CrossRef] [PubMed]
11. Alavi, S.; Ghadiri, H.; Dabirmanesh, B.; Moriyama, K.; Khajeh, K.; Masai, H. G-quadruplex binding protein Rif1, a key regulator of replication timing. *J. Biochem.* **2021**, *169*, 1–14. [CrossRef] [PubMed]
12. Brázda, V.; Červeň, J.; Bartas, M.; Mikysková, N.; Coufal, J.; Pečinka, P. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **2018**, *23*, 2341. [CrossRef]

13. Masuzawa, T.; Oyoshi, T. Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. *ACS Omega* **2020**, *5*, 5202–5208. [CrossRef]
14. Brázda, V.; Coufal, J.; Liao, J.C.; Arrowsmith, C.H. Preferential binding of IFI16 protein to cruciform structure and superhelical DNA. *Biochem. Biophys. Res. Commun.* **2012**, *422*, 716–720. [CrossRef] [PubMed]
15. Tosoni, E.; Frasson, I.; Scalabrin, M.; Perrone, R.; Butovskaya, E.; Nadai, M.; Palù, G.; Fabris, D.; Richter, S.N. Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* **2015**, *43*, 8884–8897. [CrossRef] [PubMed]
16. Bartas, M.; Brázda, V.; Bohálová, N.; Cantara, A.; Volná, A.; Stachurová, T.; Malachová, K.; Jagelská, E.B.; Porubiaková, O.; Červeň, J.; et al. In-Depth Bioinformatic Analyses of Nidovirales Including Human SARS-CoV-2, SARS-CoV, MERS-CoV Viruses Suggest Important Roles of Non-canonical Nucleic Acid Structures in Their Lifecycles. *Front. Microbiol.* **2020**, *11*, 1583. [CrossRef]
17. Kusov, Y.; Tan, J.; Alvarez, E.; Enjuanes, L.; Hilgenfeld, R. A G-quadruplex-binding macrodomain within the "SARS-unique domain" is essential for the activity of the SARS-coronavirus replication–transcription complex. *Virology* **2015**, *484*, 313–322. [CrossRef]
18. Ding, Y.; Fleming, A.M.; Burrows, C.J. Case studies on potential G-quadruplex-forming sequences from the bacterial orders Deinococcales and Thermales derived from a survey of published genomes. *Sci. Rep.* **2018**, *8*, 15679. [CrossRef]
19. Bartas, M.; Čutová, M.; Brázda, V.; Kaura, P.; Šťastný, J.; Kolomazník, J.; Coufal, J.; Goswami, P.; Červeň, J.; Pečinka, P. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* **2019**, *24*, 1711. [CrossRef]
20. Brazda, V.; Fojta, M.; Bowater, R.P. Structures and stability of simple DNA repeats from bacteria. *Biochem. J.* **2020**, *477*, 325–339. [CrossRef]
21. Ruggiero, E.; Richter, S.N. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.* **2020**, *54*, 101–131. [CrossRef] [PubMed]
22. Saranathan, N.; Vivekanandan, P. G-Quadruplexes: More Than Just a Kink in Microbial Genomes. *Trends Microbiol.* **2019**, *27*, 148–163. [CrossRef]
23. Wang, S.-R.; Zhang, Q.-Y.; Wang, J.-Q.; Ge, X.-Y.; Song, Y.-Y.; Wang, Y.-F.; Li, X.-D.; Fu, B.-S.; Xu, G.-H.; Shu, B.; et al. Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. *Cell Chem. Biol.* **2016**, *23*, 1113–1122. [CrossRef] [PubMed]
24. Jaubert, C.; Bedrat, A.; Bartolucci, L.; Di Primo, C.; Ventura, M.; Mergny, J.-L.; Amrane, S.; Andreola, M.-L. RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. *Sci. Rep.* **2018**, *8*, 1–9. [CrossRef]
25. Frasson, I.; Nadai, M.; Richter, S.N. Conserved G-Quadruplexes Regulate the Immediate Early Promoters of Human Alphaherpesviruses. *Molecules* **2019**, *24*, 2375. [CrossRef] [PubMed]
26. Liu, Y.; Le, C.; Tyrrell, D.L.; Le, X.C.; Li, X.-F. Aptamer Binding Assay for the E Antigen of Hepatitis B Using Modified Aptamers with G-Quadruplex Structures. *Anal. Chem.* **2020**, *92*, 6495–6501. [CrossRef] [PubMed]
27. Murat, P.; Zhong, J.; Lekieffre, L.; Cowieson, N.P.; Clancy, J.L.; Preiss, T.; Balasubramanian, S.; Khanna, R.; Tellam, J. G-quadruplexes regulate Epstein-Barr virus–encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* **2014**, *10*, 358–364. [CrossRef]
28. Dabral, P.; Babu, J.; Zareie, A.; Verma, S.C. LANA and hnRNP A1 Regulate the Translation of LANA mRNA through G-Quadruplexes. *J. Virol.* **2020**, *94*, 94. [CrossRef] [PubMed]
29. Ruggiero, E.; Richter, S.N. G-quadruplexes and G-quadruplex ligands: Targets and tools in antiviral therapy. *Nucleic Acids Res.* **2018**, *46*, 3270–3283. [CrossRef]
30. De Cian, A.; Gros, J.; Guédin, A.; Haddi, M.; Lyonnais, S.; Guittat, L.; Riou, J.-F.; Trentesaux, C.; Saccà, B.; Lacroix, L.; et al. DNA and RNA Quadruplex ligands. *Nucleic Acids Symp. Ser.* **2008**, *52*, 7–8. [CrossRef]
31. Lombardi, E.P.P.; Londoño-Vallejo, A.; Nicolas, A. Relationship Between G-Quadruplex Sequence Composition in Viruses and Their Hosts. *Molecules* **2019**, *24*, 1942. [CrossRef]
32. Brázda, V.; Kolomazník, J.; Lýsek, J.; Bartas, M.; Fojta, M.; Šťastný, J.; Mergny, J.-L. G4Hunter web application: A web server for G-quadruplex prediction. *Bioinformatics* **2019**, *35*, 3493–3495. [CrossRef]
33. Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [CrossRef] [PubMed]
34. Lombardi, E.P.; Londoño-Vallejo, A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* **2020**, *48*, 1–15. [CrossRef]
35. Lightfoot, H.L.; Hagen, T.; Tatum, N.J.; Hall, J. The diverse structural landscape of quadruplexes. *FEBS Lett.* **2019**, *593*, 2083–2102. [CrossRef]
36. Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **2010**, *38*, 7858–7868. [CrossRef] [PubMed]
37. Mihara, T.; Nishimura, Y.; Shimizu, Y.; Nishiyama, H.; Yoshikawa, G.; Uehara, H.; Hingamp, P.; Goto, S.; Ogata, H. Linking Virus Genomes with Host Taxonomy. *Viruses* **2016**, *8*, 66. [CrossRef] [PubMed]
38. Ofir, G.; Sorek, R. Contemporary Phage Biology: From Classic Models to New Insights. *Cell* **2018**, *172*, 1260–1270. [CrossRef] [PubMed]
39. Van Etten, J.L.; Dunigan, D.D. Chloroviruses: Not your everyday plant virus. *Trends Plant. Sci.* **2012**, *17*, 1–8. [CrossRef] [PubMed]
40. Hull, R. *Comparative Plant. Virology*, 2nd ed.; Elsevier: Amsterdam, The Netherlands; Academic Press: Boston, MA, USA, 2009; ISBN 9780123741547.

41. Marcel, V.; Tran, P.L.; Sagne, C.; Martel-Planche, G.; Vaslin, L.; Teulade-Fichou, M.-P.; Hall, J.; Mergny, J.-L.; Hainaut, P.; Van Dyck, E. G-quadruplex structures in TP53 intron 3: Role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **2010**, *32*, 271–278. [CrossRef]
42. Brázda, V.; Luo, Y.; Bartas, M.; Kaura, P.; Porubiaková, O.; Šťastný, J.; Pečinka, P.; Verga, D.; Da Cunha, V.; Takahashi, T.S.; et al. G-Quadruplexes in the Archaea Domain. *Biomolecules* **2020**, *10*, 1349. [CrossRef] [PubMed]
43. Brázda, V.; Kolomazník, J.; Lýsek, J.; Hároníková, L.; Coufal, J.; Šťʼ Astný, J. Palindrome analyser—A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. *Biochem. Biophys. Res. Commun.* **2016**, *478*, 1739–1745. [CrossRef] [PubMed]
44. Dhapola, P.; Chowdhury, S. QuadBase2: Web server for multiplexed guanine quadruplex mining and visualization. *Nucleic Acids Res.* **2016**, *44*, 277–283. [CrossRef] [PubMed]