VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA CHEMICKÁ

FACULTY OF CHEMISTRY

ÚSTAV CHEMIE POTRAVIN A BIOTECHNOLOGIÍ

INSTITUTE OF FOOD SCIENCE AND BIOTECHNOLOGY

EUKARYOTICKÉ BUNĚČNÉ SYSTÉMY A JEJICH BIOTECHNOLOGICKÉ VYUŽITÍ

EUCARYOTIC CELLS SYSTEMS AND THEIR BIOTECHNOLOGYCAL APPLICATIONS

DIZERTAČNÍ PRÁCE
DOCTORAL THESIS

AUTOR PRÁCE
AUTHOR

Ing. Otília Porubiaková

ŠKOLITEL
SUPERVISOR

doc. Mgr. Václav Brázda, Ph.D.

BRNO 2022

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA CHEMICKÁ

Zadání dizertační práce

Ústav:	Ústav chemie potravin a biotechnologií	Akademický rok: 2021/22
Studentka:	**Ing. Otília Porubiaková**	
Studijní program:	Chemie a technologie potravin	
Studijní obor:	Potravinářská chemie	
Vedoucí práce:	**doc. Mgr. Václav Brázda, Ph.D.**	

Název dizertační práce:

Eukaryotické buněčné systémy a jejich biotechnologické využití

Zadání dizertační práce:

• Studium literatury a literární přehled k dané problematice se zaměřením na buněčné systémy a jejich biotechnologické využití
• Biochemická charakterizace chemoterapeutik a přírodních látek při interakci s biologickými molekulami se zaměřením na DNA a jejich strukturní motivy, analýza strukturních motivů DNA v genomech
• Kultivace bakteriálních a kvasinkových kultur a testování viability těchto kultur, funkční testovaní, charakterizace a biotechnologické využití
• Kultivace eukaryotických nenádorových a nádorových kultur a testování viability těchto kultur po působení chemoterapeutik a přírodních látek

Termín odevzdání dizertační práce: 5.5.2022:

Ing. Otília Porubiaková studentka	doc. Mgr. Václav Brázda, Ph.D. vedoucí práce	prof. RNDr. Ivana Márová, CSc. vedoucí ústavu

V Brně dne 3.9.2021

prof. Ing. Martin Weiter, Ph.D.
děkan

ABSTRAKT

Predložená dizertačná práca je rozdelená na niekoľko častí. Prvá časť sa zaoberá interakciami proteínu p53 a jeho izoformami s rôznymi potenciálnymi DNA substrátmi za rôznych experimentálnych podmienok. Jedná sa prevažne o DNA a jej nekanonické štruktúrne motívy, ako sú G-kvadruplexy alebo krížové štruktúry, ktorých interakcie boli skúmané v izogénnych kvasinkových systémoch alebo technikami *in vitro*. Druhá časť práce sa zaoberá bioinformatickou analýzou spomínaných sekundárnych DNA štruktúr v rôznych skupinách organizmov, pričom výsledkom je súbor publikácií, ktoré vypovedajú o ich nenáhodnej distribúcii v genóme a ich vzťahu k regulácii. Posledná časť práce obsahuje doposiaľ nepublikované výsledky, vrátane výsledkov testovania vplyvu prírodných a syntetických látok na starnutie v modelových ľudských bunkách.

ABSTRACT

The submitted dissertation is divided into several parts. The first part deals with the interactions of the p53 protein and its isoforms with different potential DNA substrates under different experimental conditions. These are predominantly DNA and its non-canonical structural motifs, such as G-quadruplexes or cruciform structures, whose interactions have been studied in yeast isogenic systems or by *in vitro* methods. The seconds part deals with the bioinformatic analysis of the mentioned secondary DNA structures in different organismal groups, and the result is a set of publications that show their non-random distribution in the genome and their relationship with regulation. The last part of the work contains unpublished results, including the results of testing the effect of natural and synthetic substances on aging in model human cells.

Kľúčové slová

Proteín p53, väzba p53-DNA, nekanonické sekundárne štruktúry nukleových kyselín, bunkové starnutie.

Key words

Protein p53, p53-DNA binding, non-canonical secondary structures of nucleic acids, cell senescence.

Poďakovanie

Na prvom mieste by som chcela poďakovať svojmu školiteľovi, doc. Mgr. Václavovi Brázdovi, Ph.D. za možnosť pracovať pod jeho vedením na Biofyzikálnom ústave Akadémie vied Českej republiky, za predanie cenných rád a skúseností, trpezlivosť a podporu počas celého doktorandského štúdia. Veľké poďakovanie patrí pani prof. RNDr. Ivane Márovej, CSc., za vedenie, výnimočný prístup, podporu a pomoc nie len počas doktorandského programu, ale skrz celým mojím vysokoškolským štúdiom.

Ďalej by som rada poďakovala všetkým bývalým a súčasným kolegom z Oddelenia biofyzikálnej chémie a molekulárnej onkológie za ich čas, podporu, ochotu pomôcť, cenné rady a vytvorenie príjemného pracovného prostredia za celú spoločne strávenú dobu. Za spoluprácu by som rada poďakovala najmä doc. RNDr. Miroslavovi Fojtovi, Mgr. Natálii Bohálovej, Mgr. Alessiovi Cantarovi, a celému nášmu výskumného tímu, a tiež kolegom z oddelenia Mgr. Daniela Renčiuka, Ph.D. Veľká vďaka patrí aj prof. Johannesovi Grillarimu a celého jeho výskumnému tímu z výskumného Inštitútu Experimentálnej a klinickej traumatológie vo Viedni za jeho ochotu, profesionálny ale aj priateľský prístup počas doby pracovnej stáže.

Nakoniec by som rada poďakovala svojim najbližším priateľom a rodine za podporu, zázemie a trpezlivosť.

PORUBIAKOVÁ, Otília. Eukaryotické buněčné systémy a jejich biotechnologické využití. Brno, 2022. Dostupné také z: https://www.vutbr.cz/studenti/zav-prace/detail/143950. Dizertační práce. Vysoké učení technické v Brně, Fakulta chemická, Ústav chemie potravin a biotechnologií. Vedoucí práce Václav Brázda.

Prohlášení

Prohlašuji, že jsem disertační práci vypracovala samostatně, a že všechny použité literární zdroje jsem citovala správné a úplně. Disertační práce jez hlediska obsahu majetkem Fakulty chemické VUT v Brně a může být využita ke komerčním účelům jen se souhlasem vedoucího disertační práce a děkana FCH VUT.

……..

podpis studenta

OBSAH

ÚVOD

„Biotechnológie sú technológie využívajúce biologické systémy, živé organizmy alebo z nich odvodené biologické systémy k produkcii alebo modifikácii výrobkov či procesov pre špecifické použitie“ [1]. V molekulárnej biológii alebo medicínskom výskume k tomuto účelu slúžia modelové organizmy, využívané s cieľom porozumieť špecifickým biologickým procesom, a modelové bunkové systémy, ak experimenty na zvieratách alebo ľuďoch sa považujú za neuskutočniteľné alebo neetické [2].

Nádorové ochorenia stále patria medzi najčastejšie príčiny úmrtí a sú aj v súčasnosti veľkým medicínskym problémom. Častým javom pri tumorogenéze je inaktivácia nádorových supresorov alebo ich signálnych dráh. Medzi najčastejšie potlačované nádorové supresory patrí proteín p53, ktorý zohráva úlohu transkripčného faktoru. Reguláciou expresie svojich cieľových génov ako odpoveď na stres môže viesť k trvalému zastaveniu bunkového cyklu alebo spustiť mechanizmy programovanej bunkovej smrti [3, 4]. Regulácia väčšiny zmienených génov je podmienená špecifickou väzbou centrálnej DNA-väzbovej domény proteínu p53 na responzívny element (RE) DNA [5]. Okrem sekvenčne špecifickej väzby môže dochádzať aj k väzbe štruktúrne selektívnym spôsobom prostredníctvom C-terminálnej domény, ktorá napomáha regulácii tohto druhu väzby a zároveň rozpoznáva rôzne druhy nekanonickej formy DNA [6].

Dizertačná práca sa zameriava na niekoľko oblastí, ktorých hlavným spojením je interakcia proteínu p53 so štruktúrnymi DNA motívmi nukleových kyselín, ktoré boli v rámci práce študované *in vitro* a *in vivo*, v bakteriálnych, kvasinkových a ľudských bunkových systémoch.

V rámci teoretickej časti dizertačnej práce bol vypracovaný teoretický základ popisujúci základné vedomosti a konkrétne zvýrazňuje podstatu práce. Zameriava sa na eukaryotické organizmy a ich využitie v biotechnológiách s dôrazom na ich využitie vo forme modelových systémov, proteín p53, jeho izoformy a ich funkcie, väzbové vlastnosti k DNA a jej lokálnym štruktúrnym motívom, ďalej rozoberá jednotlivé štruktúrne motívy a bioinformatické prístupy na ich vyhľadávanie a lokalizáciu

v rôznych genómoch. Experimentálna časť obsahuje predovšetkým súbor publikačných výstupov.

Dizertačná práca vznikla spoluprácou VUT s Biofyzikálnym ústavom akadémie vied Českej Republiky. Vďaka tejto spolupráci bolo možné nabrať medziodborové skúsenosti, a výskum sa týmto mohol uberať slobodne a bez väčších materiálových obmedzení.

1 TEORETICKÁ ČASŤ

1.1 Eukaryotické organizmy v biotechnológiách

Eukaryotické organizmy sú dominantne využívanými systémami v rôznych oblastiach biotechnológií, molekulárnej biológie, medicínskeho výskumu či farmácie [7]. Hlavnou výhodou ich použitia je robustný proteosyntetický aparát umožňujúci komplexné posttranslačné modifikácie a správne zbalenie proteínov [8, 9]. Najčastejšie sa používajú ako modelové organizmy k štúdiu expresie rekombinantných proteínov, genotoxicity, signálnych dráh a podobne [10].

V súčasnosti sa v tomto odbore využíva široká škála modelových systémov, vrátane kvasiniek, *Arabidopsis thaliana, Drosophila melanogaster, Caenorhabditis elegans,* myší a rôznych druhov kultivovateľných ľudských buniek [11], ktoré sú stručne popísané v nasledujúcich kapitolách. Platí, že výber experimentálnych organizmov na základe konkrétnych fyziologických vlastností alebo praktickej vhodnosti pre danú techniku je kľúčom k zodpovedaniu konkrétnych biologických otázok.

1.1.1 Kvasinky

Kvasinky sú jednobunkové heterotrofné eukaryotické organizmy patriace medzi vyššie huby, ktoré zdieľajú vysoko konzervované molekulárne a bunkové mechanizmy s ľudskými bunkami [12]. Dávno predtým, ako našli svoje uplatnenie vo výskume, boli používané pri výrobe potravín, vrátane fermentácie pív, vín, syrov a ďalších potravinárskych produktov [13]. Aj keď sú v týchto oblastiach využívané aj v súčasnosti, nové biomedicínske aplikácie kvasiniek vytvorili nové možnosti výskumu.

1.1.1.1 Využitie kvasiniek ako modelových organizmov

Kvasinka *Saccharomyces cerevisiae*, tiež známa ako pekárenské droždie, je najznámejším a najštudovanejším kvasinkovým organizmom. Práve tieto kvasinkové modely pomohli k pochopeniu konzervovaných bunkových mechanizmov, ako je bunkové delenie, replikácia DNA, metabolizmus, skladanie proteínov a intracelulárny transport [14]. S ich životným cyklom, ktorý zahŕňa haploidné aj diploidné formy, je

štúdium mutácií extrémne zjednodušené. Navyše, klasické genetické manipulácie sú uľahčené párením haploidných kmeňov a sporulujúcich diploidných kmeňov pri vysokej účinnosti transformácie týchto buniek. Preto táto účinná rekombinantná dráha umožňuje relatívne ľahké vloženie, odstránenie alebo mutáciu akejkoľvek sekvencie [15]. Na základe vyššie spomínaných zistení, bol *S. cerevisiae* prvým úplne sekvenovaným eukaryotickým organizmom [16], čo otvorilo mnohé ďalšie oblasti výskumu. Tým došlo k objaveniu množstva funkčných genomických nástrojov, vrátane komplexnej zbierky kvasinkových mutantov [17, 18], zbierky celogenómovej nadmernej expresie [19] a kvasinkové kmene značené zeleným fluorescenčným proteínom (GFP) [20, 21]. Vyspelosť genetického a molekulárneho súboru nástrojov postavila kvasinky na vyššiu úroveň pre vývoj mnohých vysokovýkonných technológií, vrátane transkriptómových [22–24], proteómových [25] a metabolómových skríningov [26, 27]. S príchodom možností modelovania boli vyvinuté rôzne metódy študujúce interakcie proteín-proteín [28, 29], proteín-DNA [30, 31] a genetické interakcie [32, 33].

Kvasinky zdieľajú s ľuďmi významnú časť svojich funkčných dráh, vrátane bunkového cyklu [34], metabolizmu [35], programovanej bunkovej smrti [36, 37], skladania a degradácie proteínov [38], vezikulárny transport [39] a mnohých signálnych dráh, ako je mitogénom aktivovaná proteínkináza (MAPK) [40, 41], cieľ rapamycínu (TOR) [42] a inzulín/IGF-I [43]. Tieto konzervované biochemické a metabolické dráhy okrem iného riadia bunkový rast, delenie, transport látok, odpoveď na stres a sekréciu, pričom sú spojené s rôznymi ľudskými chorobami. Práve preto sú kvasinky významným modelovým organizmom pre štúdium ľudských patológií [44–46], vrátane pochopenia vzniku rakoviny [47] a neurodegradatívnych ochorení [48].

1.1.1.2 Využitie kvasiniek v potravinárskom priemysle

Kvasinky majú široké spektrum využitia vrátane potravinového priemyslu (výroba vína, piva, destilátov and pečiva) a produkcie biomasy. Ich vlastnosť metabolizovať sacharidy, etanol, oxid uhličitý a iné metabolity [49] prispieva k chemickému zloženiu,

senzorickej kvalite a typickým vlastnostiam fermentovaných potravín a nápojov [50, 51].

Väčšina kvasinkových kmeňov dokáže fungovať pod aeróbnymi aj anaeróbnymi podmienkami prostredia, vrátane prepínania typu metabolizmu [52]. Aj keď priebeh hlavných metabolických dráh ostáva zachovaný, prejavuje nezvyčajnú flexibilitu vďaka niektorým regulačným mechanizmom [53]. V potravinárskom priemysle je kvasinka *S. cerevisiae* používaná najčastejšie, pričom rody *Candida, Endomycopsis* a *Kluyveromyces* sú využívanými kmeňmi pre produkciu biomasy [54].

S. cerevisiae/ S. pasterianus sa používajú na zakvasenie mladiny pri výrobe piva, pričom dochádza k premene skvasiteľných sacharidov v mladine na alkohol a oxid uhličitý procesom anaeróbneho kvasenia [55, 56]. Podľa druhu použitých kvasiniek dochádza k vrchnému („Ale“) alebo spodnému (ležiaky) kvaseniu. K spodnému kvaseniu sa používajú kvasinky *S. pasterianus*, ktorých teplotné rozmedzie sa pohybuje od 7 do 15°C. *S. cerevisiae* sa používajú pre tzv. vrchné kvasenie, pri ktorom dochádza k vynášaniu kvasiniek do kvasnej deky v poslednej fáze kvasiaceho procesu, pri teplotách 18-22 °C [57]. Samotná fermentácia je jedným z časovo najnáročnejších krokov pri výrobe piva, pričom jej rýchlosť závisí od uhlíkových a dusíkových zdrojov a rýchlosti ich využitia. Atraktívnou alternatívnou možnosťou je príprava nových kvasných kmeňov pomocou genetických manipulácii, ktorá ponúka nové možnosti tvorby kvasníc s rôznymi technologickými a senzorickými vlastnosťami prípadne tvorbu úplne nových druhov pív [58].

Medzi vínne kvasinky sa radia kvasinky rodu *Saccharomyces* (*S. cerevisiae*, *S. oviformis*, *S. beticus,* a iné), ktoré pri fermentácii produkujú žiadané charakteristické senzorické látky. Kvasinky sa do hroznového muštu dostávajú buď z povrchu hrozien (spontánne kvasenie) alebo, v druhom prípade, sa do muštu pridávajú čisté kultúry ušľachtilých kvasiniek (pravá fermentácia) [55, 56]. Okrem hlavných produktov fermentácie, ktorými sú oxid uhličitý a etanol, vznikajú prchavé kyseliny, rôzne organické kyseliny vyšších alkoholov, esterov, aldehydov, ketónov a sírnych a amónnych zlúčenín, ktoré ovplyvňujú senzorické vlastnosti výsledného produktu [56]. Rovnako ako u pivovarských kvasiniek je žiadúca príprava špecializovaných

kmeňov kvasiniek so širokým spektrom optimalizovaných vlastností. Medzi najdôležitejšie ciele génového inžinierstva patrí tvorba kmeňov so zlepšenou výrobnou technológiou a kvalitou, ako je zvýšenie fermentačnej výkonnosti, vyššia tolerancia kvasiniek k etanolu, lepšie využitie sacharidov a zlepšené organoleptické vlastnosti [49, 59, 60].

Vo väčšine vyrábaného pečiva, sú kvasným činidlom kvasnice (droždie) – *S. cerevisiae.* V procese výroby sa používa tzv. kvások, ktorý predstavuje kašovitá zmes múky, vody, sacharidov a kvasiniek. Často sa okrem *S. cereviasiae*, do kvásku pridávajú aj iné mikroorganizmy, ako napríklad *Kluyveromyces marxianus,* pri nutnosti spracovania laktózy a srvátkovej bielkoviny, alebo *Kazachstania exigua* či *Candida humilis*, ktoré dokážu hydrolyzovať maltózu [61].

V liehovarníctve na výsledný liehový produkt významne vplýva výber vhodných kvasinkových kmeňov. V súčasnosti sú čisté kultúry komerčne voľne dostupné a často obsahujú rôzne genetické modifikácie umožňujúce fermentovať širokú škálu sacharidov (napr. maltotetralózu). Môžu byť modifikované za účelom získania nových funkcií alebo výhod, ako je zvýšená tolerancia k prostrediu, osmotickému tlaku, teplote alebo pH [62, 63]. S výnimkou voľného prirodzeného kvasenia sa pre výrobu liehovín používajú prevažne kvasinky rodu *S. cerevisiae* (napr. pre výrobu rumu sa používajú v kombinácii s kvasinkami *Schizosaccharomyces pombe*) [31].

Kvasinky nachádzajú svoje uplatnenie aj na poli probiotík. Dlhodobo bola označovaná ako jediná kvasinka s probiotickými účinkami *S. cerevisiae* var*. boulardii* [64]. Táto problematika je stále málo študovaná, aj keď sú kvasinky veľmi dôležité pre udržanie rovnováhy gastrointestinálneho traktu. Prirodzená gastrointestinálna mikroflóra obsahuje menej ako 0,1 % kvasiniek, pričom *C. albicans* v najväčšom zastúpení [65]. Na základe účinnosti mnohé klinické štúdie odporúčajú *S. boulardii* ako prostriedok k prevencii a liečbe niekoľkých črevných ochorení [66]. Medzi ďalšie kvasinky s probiotickými účinkami patrí napríklad *Kluyveromyces marxianus* s vlastnosťami modulácie imunitnej odpovede [67–69].

1.1.2 *Arabidopsis thaliana*

Arabidopsis thaliana patrí do čeľade Brassicaceae a tiež medzi základné genetické nástroje k štúdiu rastu a vývojových procesov v rastlinách po mnoho desaťročí [70]. Jej použitie umožnilo rozsiahle pokroky v chápaní biologických procesov v rastlinách, ako je vývoj, reakcia na abiotický stres, hormonálna biológia a signalizácia [71].

Vývin *Arabidopsis* zo semena na zrelú rastlinu prebieha v krátkom časovom období a na rozdiel od iných rastlín je možné tieto rastliny pestovať v interiéri pod slabým fluorescenčným osvetlením bez potreby kokultivácie iných druhov, čo zvyšuje možnosť kontroly aseptických rastových podmienok [72]. Použitie *Arabidopsis* je tým pádom rýchly a relatívne málo nákladný proces [73] zvýhodnený relatívne malou veľkosťou plne sekvenovaného genómu (približne 132 Mbp) [74]. Okrem genetiky je tento rastlinný model využívaný aj na zodpovedanie otázok v biochémii, molekulárnej biológii a fyziológii [75].

1.1.3 *Drosophila melanogarster*

Drosophila melanogaster, známa ako ovocná muška, je ďalším z najštudovanejších modelových organizmov v biomedicínskom výskume. Po prvý krát bola použitá k preukázaniu chromozomálnej teórie dedičnosti [76]. Nasledovalo definovanie mnohých princípov genetiky, vrátane účinkov röntgenových lúčov alebo mutácií [77]. Na základe týchto zistení vzišla nová generácia špecializovaných chromozómov, ktoré bránia rekombinácii prostredníctvom série inverzií DNA [78].

Moderná éra výskumu sa začala po dôkladnej analýze génov zapojených do vývoja embrya [79]. Prelomovým objavom bolo zistenie, že jednotlivé gény regulujú rôzne aspekty vývoja, a mnohé z nich sú homológne s génmi zapojenými do ľudského vývoja. V minulosti sa chemická mutagenéza používala k generovaniu nových mutácií, ktoré boli podrobené skríningu na zaujímavé fenotypy, nasledované genetickým mapovaním a klonovaním [80]. V súčasnosti sa na zacielenie všetkých génov používa transpozónový systém MiMIC („*Minos* mediated integration casette“ – *Minos* sprostredkovaná integračná kazeta), ktorý poskytuje nulové mutácie a platformu na

označovanie proteínov, sledovanie génovej expresie a mnoho ďalších funkcií prostredníctvom prístupu výmeny exónov [81]. Toto, v spojení s CRISPR/Cas9 a stratégiou nadmernej expresie dovoľuje inaktiváciu, značenie a nadmernú expresiu akéhokoľvek génu v priebehu niekoľkých týždňov [82].

1.1.4 *Caenorhabditis elegans*

Dôležitými vlastnosťami, ktoré stavajú *C. elegans* medzi atraktívne a efektívne modelové organizmy, je ľahká manipulácia, minimálne požiadavky na výživu a rast, samooplodnením produkované veľké potomstvo v priebehu niekoľkých dní a znalosť sekvencie celého genómu [83].

Definovaná vývojová línia a dynamická zárodočná línia obsahujúca priestorovo rozlíšené mitotické a meiotické bunkové delenia umožňuje študovať stabilitu genómu a mechanizmy opravy DNA [84]. Tento mnohobunkový eukaryotický organizmus nachádza svoje uplatnenie v neurobiológii, vývojovej biológii, genetike, toxikológii a vysokovýkonných skríningových prístupoch [85].

1.1.5 Myšacie modelové systémy

Za posledných 30 rokov genetické, biochemické a fyziologické štúdie na myšiach zmenili porovnávaciu biológiu u cicavcov a poskytli základný náhľad pre pochopenie ľudskej fyziológie a mechanizmov chorôb [86] na základe ich fylogenetickej a fyziologickej príbuznosti. Genomické štúdie spolu s vývojom nových metód k príprave transgénnych, „knockout“ a „knockin“ myší, a vývoj výkonných nástrojov pre ich výskum viedli k nárastu používania práve týchto organizmov a prispeli k pochopeniu biológie človeka [87]. Myši ale reagujú na experimentálne zásahy odlišne ako ľudia. Preto sa v skutočnosti mnoho liečiv nedostane na trh, čo je v istej miere zapríčinené obmedzením zvieracích modelov pri testovaní [88].

1.1.6 Bunkové línie

Rozvoj bunkových kultúr výrazne ovplyvnil oblasť biologických vied a prispel k pokroku v medicíne. Ich použitie je základným postupom pre modelovanie chorôb,

výskum kmeňových buniek a rakoviny, a zavádzanie terapií [89]. Okrem toho hrajú významnú úlohu pri štúdiu fyziologických, patologických a diferenciačných procesoch a umožňujú skúmanie postupných zmien v štruktúre, biológii a genetickom zložení bunky v kontrolovanom prostredí [90]. K tomuto účelu slúžia nesmrteľné, primárne a kmeňové bunkové línie.

1.1.6.1 Imortalizované bunkové línie

Nesmrteľné, alebo imortalizované, bunkové línie sú bunky upravené k neobmedzenej proliferácii s cieľom dlhodobej kultivácie. Sú odvodené z rôznych zdrojov s chromozomálnymi abnormalitami, ktoré im umožňujú kontinuálne delenie, napríklad z nádorov [91]. Ponúkajú niekoľko výhod, vrátane nákladovej efektivity, ľahkého použitia a obchádzajú etické obavy spojené s používaním zvieracích a ľudských tkanív.

V súčasnosti je možné generovať akékoľvek imortalizované bunkové línie pomocou manipulácie s genómom. K imortalizácii môže dochádzať spontánne (HeLa) alebo inhibíciou normálnych kontrolných bodov bunkového cyklu, ako je napríklad proteín p53 alebo pRb (retinoblastómový proteín), alebo zavedenie expresie vírusových onkogénov s cieľom zabrániť iniciáciu apoptotických dráh a tým umožniť nepretržité delenie [92]. Ďalšou možnosťou je nadmerná expresia telomerázy a hTERT (telomerázová reverzná transkriptáza) prostredníctvom retrovírusového vektora, ktorá je schopná predĺžiť teloméry a tým oddialiť replikačnú senescenciu [93].

Bunkové línie priniesli revolúciu na poli vývoja vakcín, testovania metabolizmu liekov a cytotoxicity, produkcii protilátok, štúdiu génov, vytvárania umelých tkanív a syntéze rôznych biologických zlúčenín [94, 95].

1.1.6.2 Primárne bunkové línie

Aj keď sú bunkové línie široko používané na skúmanie rôznych bunkových procesov, môžu podliehať spontánnej mutácii alebo môžu podliehať chromozomálnym, morfologickým alebo tumorogénnym zmenám [96]. Na rozdiel od imortalizovaných bunkových kultúr, primárna bunková kultúra je *ex vivo* kultúra čerstvo získaná z mnohobunkového organizmu [97]. Kvôli náročnosti kultivácie sa primárne

bunkové línie rozšírili až po roku 2000, ale poskytujú lepšie znázornenie bunkovej heterogenity tkanív, transkriptomického a proteomického profilu a realistickejšie funkčné reakcie vrátane reakcií na liečivá [98–100].

1.1.6.3 Kmeňové bunkové línie

Hlavným problémom výskumu kmeňových buniek bola ich identifikácia a charakterizácia *in vivo* a úspešná kultivácia *in vitro*. Zárodočné bunky, ktoré dávajú vzniknúť novým embryám a tým aj kmeňovým bunkám, je taktiež možné zachovať *in vitro* [101]. Ľudské embryonálne línie kmeňových buniek boli po prvý krát úspešne kultivované v roku 1998 a môžu byť udržiavané *in vitro* po neurčitú dobu [102, 103]. Významnou vlastnosťou týchto bunkových línií je, že môžu mať potenciál vytvárať nové tkanivá pre akúkoľvek časť tela, čo podnietilo rozsiahly výskum. Práca s nimi je ale nesmierne náročná, a to z dôvodu rýchlej diferenciácie, čo z nich robí extrémnu výzvu na rast a spoľahlivú prípravu [104].

1.2 Proteín p53

Proteín p53 je jedným z najštudovanejších tumor supresorových proteínov vďaka jeho kľúčovým úlohám pri udržovaní genetickej stability a inhibícii tvorby nádorových ochorení [105]. Pri jeho aktivácii, po poškodení bunky, proteín p53 indukuje expresiu génov, podieľajúcich sa na oprave poškodenia DNA, bunkovom raste, zastavení bunkového cyklu [106] alebo programovanej bunkovej smrti [107–109].

1.2.1 Štruktúra proteínu p53

Štruktúra proteínu p53 je rozdelená na domény, ktoré zahrňujú transaktivačnú doménu, doménu bohatú na prolín, centrálnu doménu, oblasť obsahujúcu signál jadrovej lokalizácie a oligomerizačnú doménu [110].

Centrálna doména je zodpovedná za väzbu na sekvenčne špecifické prvky DNA umiestnené v blízkosti promótorov cieľových génov proteínu p53 [111]. Dve dekamerické sekvencie, 5‘-RRRC(A/T)(T/A)GYYY-3‘ (pričom R predstavuje purínovú a Y pyrimidínovú bázu), tvoria RE [112–114]. Rovnako ako iné transkripčné faktory,

proteín p53 obsahuje druhé DNA väzbové miesto v C-terminálnej doméne (oligomerizačnej doméne), ktoré tvorí stabilné komplexy s DNA sekvenčne nešpecifickým spôsobom, vrátane nesprávne priradenej DNA, dvojvláknových zlomov a jednovláknovej DNA [115, 116]. Z tohto je zrejmé, že C-terminálna oblasť rovnako ovplyvňuje účinnosť p53 ako transkripčného faktoru a jeho posttranslačné modifikácie a interakcie s inými proteínmi modulujú stabilitu jeho komplexov s DNA [117]. Tieto poznatky vedú k hypotéze, že regulačná súhra medzi centrálnou a C-terminálnou doménou proteínu p53 môže ovplyvňovať globálnu štruktúru molekuly proteínu po jeho aktivácii [118].

Transaktivačná doména (TAD, N-terminálna doména) je nevyhnutná pre interakcie s transkripčnými kofaktormi a korepresormi. Pozostáva z dvoch homológnych subdomén, TAD I a TAD II, pričom obe subdomény obsahujú sekvenčne konzervované motívy, Φ – X – X – Φ – Φ (Φ je hydrofóbna a X je akákoľvek aminokyselina), ktorých výskyt je obvyklý pre mnoho proteínov regulujúcich transkripciu [119]. Transaktivačnú doménu nasleduje doména bohatá na prolín (PRD, „proline rich domain“) [120], ktorá slúži k regulácii transaktivácie apoptotických génov.

Úloha proteínu p53 ako transkripčného faktoru je za normálnych podmienok potlačovaná proteínom MDM2, ktorý maskuje N-terminálnu doménu p53 a tiež prispieva k jeho degradácii prostredníctvom ubikvitinácie [121–123].

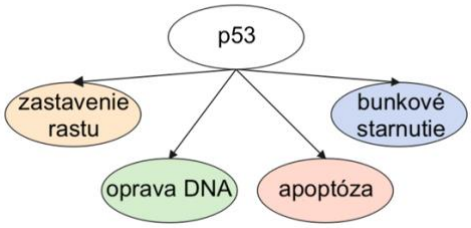
1.2.2 Aktivácia proteínu p53

Aktivácia proteínu p53 ako odpoveď na bunkový stres, zahŕňa tri základné kroky: stabilizáciu proteínu, sekvenčne špecifickú väzbu na DNA a transaktiváciu cieľových génov [124]. K stabilizácii proteínu zväčša dochádza prostredníctvom dejov, ktoré narúšajú jeho interakciu s MDM2, ako je napríklad fosforylácia N-terminálnej domény [125]. Po stabilizácii nasleduje sekvenčne špecifická väzba proteínu p53 na DNA prostredníctvom DNA väzbovej domény a následná aktivácia alebo represia cieľových génov. Proteín p53 môže interagovať s rôznymi transkripčnými aktivátormi, ako je napríklad histón acetyltransferáza CBP/p300 [126], alebo transkripčnými represormi a korepresormi, ako sú histónové deacetylázy a sin3 [127], k modulácii transkripcie.

Interakcia p53 s CBP/p300 uľahčuje acetyláciu histónov a súčasne proteínu p53, čo vedie k otvorenejšej konformácii chromatínu v blízkosti cieľových sekvencií proteínu a k jeho vyššej aktivite [128].

1.2.3 Funkcie proteínu p53

Proteín p53 je kľúčovým regulátorom ktorý zabezpečuje bunkovú odpoveď na stresové signály indukciou zastavenia bunkového cyklu, bunkového starnutia alebo apoptózy. Konečný funkčný výsledok určuje povaha a intenzita stresového signálu, typ bunky a kontext [129]. Schematické rozdelenie funkcií proteínu p53 je znázornené na Obrázok 1.



Obrázok 1. Funkcie proteínu p53. Upravené z [130].

1.2.3.1 Bunkové starnutie (senescencia)

Bunkové starnutie je permanentná forma zastavenia bunkového cyklu, ktorá bola prvý krát popísaná v roku 1965 [131]. Bunkové starnutie môže indukovať množstvo stresových faktorov, vrátane dysfunkčných telomér, netelomerického poškodenia DNA, nadmernej mitogénnej signalizácie a odchýlok v organizácii chromatínu [132].
Na rozdiel od normálnych buniek sú „senescentné“ bunky relatívne stabilné, bez proliferačnej kapacity, ale zachovávajú si metabolickú aktivitu. Sú charakteristické veľkou sploštenou morfológiou, zvýšenou aktivitou β-galaktozidázy a SASP („senescence-associated secretory phenotype“), charakteristický zvýšenou expresiou cytokínov a chemokínov. P53 hrá v tomto procese významnú úlohu, pretože niektoré z týchto látok môžu byť indukované priamo týmto proteínom [132, 133].

Všeobecne bunkový cyklus pozostáva z S fázy (DNA syntéza), M fázy (mitóza) a dvoch G fáz (G0 a G1), ktoré sú regulované proteínmi bunkového cyklu (cyklíny),

cyklín dependentnými kinázami (CDK) a cyklín dependentnými inhibítormi kináz (CDKI). CDKI môžu byť radené do dvoch kategórií, CIP rodina ($p21^{CIP1}$) a INK4 rodina ($p16^{INK4a}$). Ako je možné vidieť na Obrázok 2, prostredníctvom udalostí vyvolaných genomickým stresom, stimulovaný p53 transaktivuje $p21^{CIP1}$, ktorý naopak inhibuje CDK a tým zachováva pRb v neaktívnom, nefosforylovanom stave. Nefosforylovaný pRb vytvára komplex s proteínom E2F, ktorý je zodpovedný za prechod G1/S, čím dochádza k zastaveniu proliferácie a oprave poškodenia DNA [134–136]. Rovnako $p16^{INK4a}$ hrá dôležitú úlohu ako centrálny modulátor zastavenia bunkového cyklu. $p16^{INK4a}$ môže inhibovať CDK a viesť k rovnakému výsledku, ktorým je nefosforylovaný, a tým pádom v tomto procese neaktívny, pRb. Navyše, $p14^{ARF}$ je schopný detegovať rôzne signály starnutia a väzbou na MDM2 aktivovať p53 [137]. Vzájomná komunikácia a redundancia medzi týmito dráhami zdôrazňuje dôležitosť bunkového starnutia ako biologického tumor supresorového faktoru.

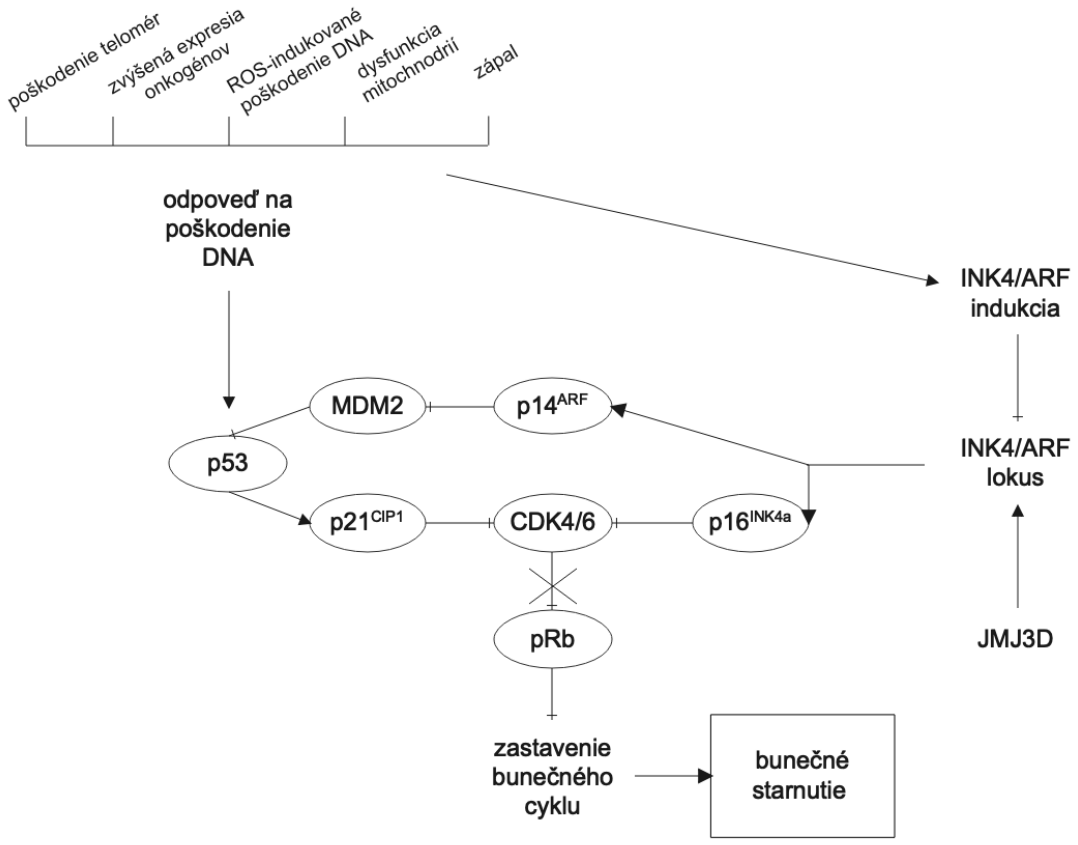
Za normálnych podmienok sa pri bunkových cykloch teloméry, repetitívne sekvencie DNA na konci chromozómov, skracujú. Ak dĺžka telomér dosiahne určitú hranicu, bunková proliferácia sa zastaví, a vyvolá replikatívne starnutie [138]. Okrem tohto, bunkové starnutie môže nastať v dôsledku poškodenia DNA, poškodením telomér, následkom zvýšenej expresie onkogénov alebo iných faktorov, čo sa označuje ako predčasné starnutie [139–141].

Keďže je bunkové starnutie trvalou formou zastavenia bunkového cyklu, sú kľúčové faktory kontrolných bodov, ako napríklad p53, $p21^{CIP1}$, $p16^{INK4a}$ a pRB, tiež kľúčovými regulátormi starnutia. V ľudských bunkách je replikatívne starnutie závislé od dráhy p53, zatiaľ čo predčasné starnutie môže byť sprostredkované oboma typmi dráh [138, 142]. Princíp výberu dráhy vedúcej k bunkového starnutiu doposiaľ nie je známy, ale predpokladá sa významný vplyv druhu a intenzity stresového signálu.

1.2.3.2 Kontrolné body bunkového cyklu

Kontrolné body bunkového cyklu sú dôležitými riadiacimi bodmi, ktoré zaisťujú presnosť bunkového delenia pomocou systému overovania dokončenia predchádzajúcej fázy cyklu. V reakcii na rôzne bunkové stresové podnety môže dôjsť

k zastaveniu rastu bunky s cieľom zabrániť šíreniu mutácií v DNA. Napríklad, pri poškodení DNA môže dochádzať k aktivácii ATM/ATR, tým k aktivácii p53 a následnému zastaveniu cyklu v G1 fáze. K tomuto dochádza prevažne transaktiváciou $p21^{CIP1}$ (cyklín dependentný inhibítor kinázy) [143]. Prechod buniek z G2 fázy do mitózy je riadený MPF (faktor podporujúci dozrievanie) s obsahom komplexu cyklínu B1 a cdc2, pričom bolo dokázané, že k zastaveniu G2/M prechodu dochádza narušením funkcie práve tohoto komplexu. Konkrétne, po poškodení DNA p53 potláča funkciu cdc25c fosfatázy, ktorej funkciou je podpora mitózy [144]. Ďalej sa ukázalo, že p53 môže tiež po poškodení transkripčne aktivovať proteín 14-3-3σ, ktorý zabraňuje správnej lokalizácii komplexu cyklínu B1/cdc2, a tým spôsobuje podobný efekt, ktorým je zastavenie bunkového rastu [145].



Obrázok 2. p53/$p21^{CIP1}$/pRb a $p16^{INK4a}$/pRb signálne dráhy. INK4/ARF kóduje ARF („alternate reading frame“/ alternatívny čítací rámec) aj $p16^{INK4a}$ proteín. ARF môže stimulovať p53 prostredníctvom inhibície a degradácie MDM2. Aktivovaný p53 transaktivuje $p21^{CIP1}$, ktorý inhibuje CDK (cyklín dependentnú kinázu) a následne CDK závislú fosforyláciu pRb. Nefosforylovaný pRb potláča funkciu

G1/S fázy ovplyvňujúceho faktoru a výsledkom je zastavenie proliferácie a oprava DNA. Podobne, p16INK4a môže inhibovať CDK, tým redukovať fosforyláciu pRb a viesť k zastaveniu bunkového cyklu. Tým dochádza k zastaveniu bunkového cyklu v G1 fáze [136]. JMJ3D patrí medzi demetylázy a môže ovplyvňovať bunkové starnutie skrz INK4/ARF lokus. Prevzaté a upravené z [146].

1.2.3.3 Autofágia

Autofágia je katabolický proces zahrňujúci degradáciu vlastných bunkových komponent primárne prostredníctvom lyzozomálneho aparátu. Môže mať pro- alebo anti-onkogénne funkcie, ktoré môžu odrážať jej pôsobenie ako mechanizmu prežitia alebo smrti [147]. Hlavným aktérom autofágie je konzervovaný proteín Beclin 1, ktorý tiež plní funkciu nádorového supresoru [148]. Úlohou proteínu p53 je transkripčne aktivovať poškodený gén modulátora autofágie regulovaného poškodením (DRAM), ktorý produkuje lyzozomálny proteín, a indukovať autofágiu spôsobom závislým od DRAM [149]. Rovnako je DRAM zásadný pre apoptózu sprostredkovanú p53.

1.2.3.4 Apoptóza

Apoptóza je aktívny proces, ktorý hrá kľúčovú úlohu v procese bunkového starnutia a predstavuje homeostatický mechanizmus na udržanie pravidelných bunkových populácií [150]. Jednou z hlavných biologických úloh proteínu p53 je jeho schopnosť indukovať tento proces v geneticky nestabilnej bunke [151]. Apoptóza je charakterizovaná sekvenciou krokov, ktoré vedú k programovanej smrti buniek. Všeobecne prebieha v niekoľkých fázach, najskôr kondenzáciou jadrového chromatínu, potom jadrovým štiepením a nakoniec rozdelením bunkového obsahu na odlišné vezikuly uzatvorené v membráne, nazývané apoptotické telieska [152, 153]. Následné odstránenie apoptotického materiálu fagocytmi je veľmi rýchle, preto je prítomnosť apoptotických teliesok *in vivo* obmedzená [154]. V cicavčích bunkách dochádza k apoptóze, alebo tiež programovanej bunkovej smrti, dvomi dráhami [155]; vnútornou a vonkajšiu.

Vnútorná dráha sa nazýva dráhou regulovanou BCL-2 (tiež mitochondriálnou alebo stresovou dráhou) a je aktivovaná napríklad depriváciou cytokínov, stresom endoplazmatického retikula alebo poškodením DNA. Bunková smrť je iniciovaná

transkripčnou alebo posttranskripčnou zvýšenou expresiou tzv. proapoptotických proteínov BH3 – členov BCL-2 proteínovej rodiny (BIM, PUMA, BID, BMF, BAD, BIK, NOXA,HRK). Tieto proteíny viažu a inhibujú BCL-2 „proteíny prežitia“ (BCL-2,BCL-XL, MCL-1, BCL-W a A1/BFL1), čím uvoľňujú efektory bunkovej smrti BAX a BAK. Aktivácia BAX a BAK spôsobuje mitochondriálnu permeabilizáciu vonkajšej membrány s následnou aktiváciu kaskády aspartátovo špecifických cysteínových proteáz (kaspáz, pričom táto dráha je iniciovaná pomocou kaspázy-9 a aktivátorom je APAF-1) [156–158], ktoré spôsobujú rozpad bunky [159].

Vonkajšia dráha je tiež nazývaná „dráhou receptora smrti“ a na jej aktiváciu je potrebná spolupráca viacerých členov rodiny receptorov tumor nekrotizujúcich faktorov (TNFR – „Tumour necrosis factor receptor“) nesúcich doménu intracelulárnej smrti. Na rozdiel od vnútornej dráhy aktivuje apoptózu náborom a aktiváciou pro-kaspázy-8 prostredníctvom adaptérov FADD a v niektorých prípadoch aj TRADD na ligovaných receptoroch smrti na plazmatickej membráne [160, 161]. Napríklad pre tymocyty (druh leukocytov) platí, že aktivácia kaspázy-8 následne vedie aktivácii efektorových kaspáz, kaspáz-3 a -7, čo dostačuje k úspešnej indukcii apoptózy. Naopak v hepatocytoch vyžaduje účinné usmrtenie bunky amplifikáciu kaspázovej kaskády krížovou aktiváciou apoptotickej dráhy regulovanej BCL-2, ku ktorej dochádza proteolytickou aktiváciou sprostredkovanou pomocou kaspázy-8, inak inertným, BID proteínom [162–166].

P53 zjavne podporuje apoptózu prostredníctvom transkripčne závislých (skrz proteíny BCL-2 rodiny) a nezávislých mechanizmov s cieľom dosiahnutia efektívneho procesu bunkovej smrti [167]. Rozhodnutie medzi nastolením apoptózy a prežitím buniek závisí od členov BCL-2 rodiny, ktoré sú regulované proteínom p53 [168].

1.2.4 Proteíny rodiny p53

Rodina p53 v ľudských organizmoch zahŕňa homológne gény *TP53*, *TP63* a *TP73*. Každý z členov produkuje niekoľko proteínových izoforiem ako výsledok použitia alternatívneho promótora a/alebo alternatívneho mRNA zostrihu. Expresia množstva tkanivovo špecifických proteínov, ktoré sa chovajú ako transkripčné faktory, vyúsťuje

v správnu diferenciáciu, metabolizmus, ochranu pred tumorogenézou a v mnohé ďalšie funkcie [169–171].

Gény lokalizované na chromozómoch 17p13.1, 3q27-29 a 1p36.2-3 kódujú proteíny s podobnou výslednou štruktúrou a významnou aminokyselinovou homológiou v transaktivačnej, DNA väzbovej a oligomerizačnej doméne. Evolučne najkonzervovanejšou doménou je DNA väzbová doména, čo naznačuje, že regulácia transkripcie hrá významnú úlohu v mnohých funkciách prisúdených rodine p53 [172].

Transkripcia všetkých troch proteínov p53 rodiny je regulovaná podobnými mechanizmami. Ku kontrole dochádza prostredníctvom P1 a P2 promótorov, pričom P2 je alternatívnym vnútorným promótorom. Na tomto základe môžu byť generované proteíny rozdelené do dvoch hlavných skupín, TA a ΔN [173, 174], pričom TA varianty obsahujú transaktivačnú doménu a variantom ΔN táto doména čiastočne alebo úplne chýba v závislosti od P2 promótoru [175].

Dodatočná diverzita je tvorená pomocou alternatívnych zostrihov na 3’ konci transkriptu. Kombináciou alternatívneho zostrihu na 5‘ a 3‘ konci, alternatívnou iniciáciou translácie a použitím alternatívneho promótora je možné ešte navýšiť diverzitu proteínov [176–178].

Proteíny p53, TAp63 a TAp73 zdieľajú funkčné podobnosti. Všetky spomínané varianty sa podieľajú na zastavení bunkového cyklu, apoptóze alebo bunkovom starnutí. Tieto podobnosti môžu byť čiastočne zapríčinené transaktiváciou rovnakých transkripčných cieľových génov. Vo všeobecnosti sa tieto proteíny viažu špecificky ku konzervovaným cieľovým sekvenciám p53 na DNA pomocou ich DNA väzbovej domény.

ΔN varianty fungujú ako dominantné negatívne inhibítory TA variant, kedy tieto dve varianty „súťažia“ o cieľové miesto na promótori [175, 179, 180]. Sú regulované negatívnym spätno-väzbovým mechanizmom, pri ktorom TA izoformy dokážu indukovať transkripciu ΔN aktiváciou P2 promótorov. Naopak, indukcia ΔN izoforiem inhibuje TA [176, 181, 182].

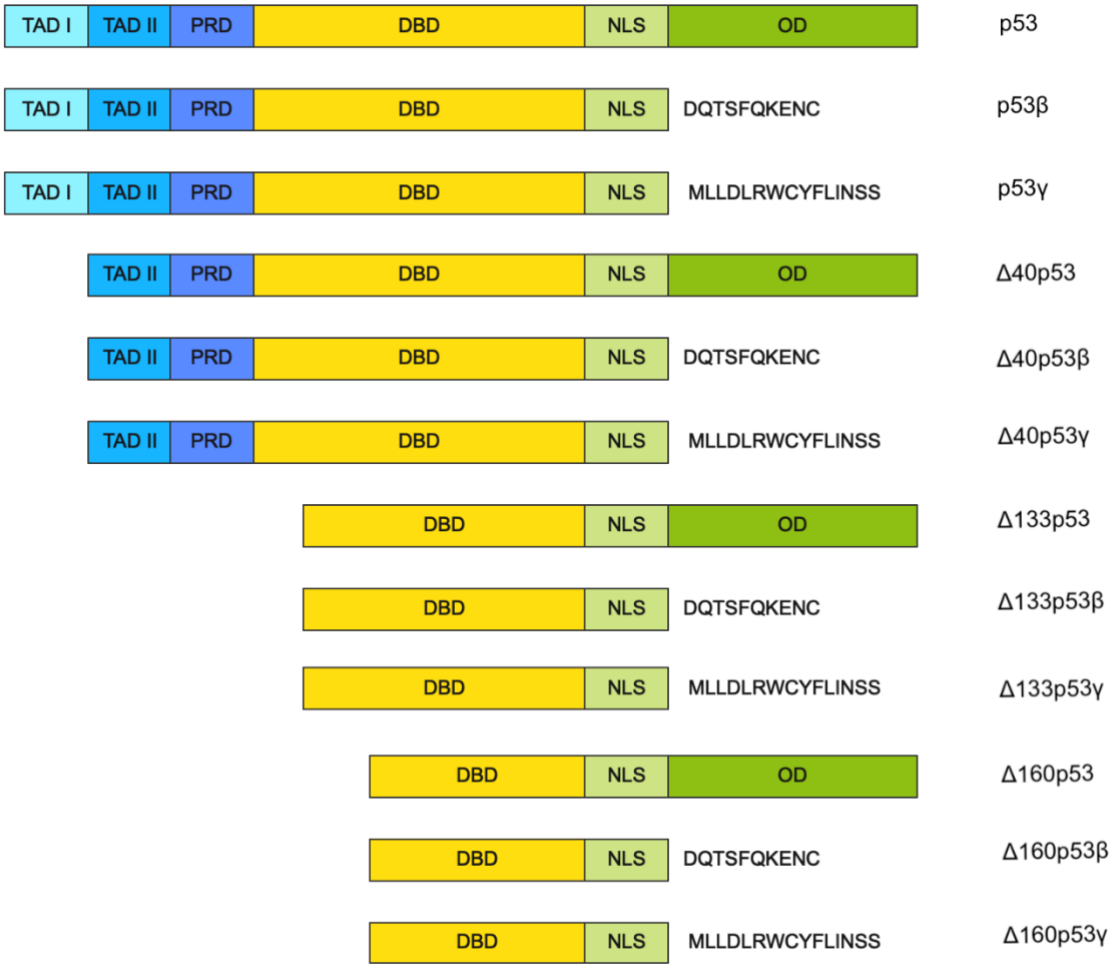
1.2.5 Izoformy proteínu p53

Ľudský gén *TP53* obsahuje 13 exónov (z ktorých dva sa označujú ako alternatívne spojené exóny) lokalizované na chromozóme 17p13.1 [183]. Bolo objavených a popísaných niekoľko variant zostrihu génu *TP53* a bol stanovený ich biologických a klinický význam [184, 185]. Alternatívnym zostrihom intrónu 9 a alternatívnym promótorom vzniká, v ľudskom organizme, dvanásť p53 proteínových izoforiem, ktoré sú kódované deviatimi transkriptmi mRNA génu *TP53*, exprimovanými v tkanivách rozdielne (Obrázok 3) [176, 186–189].

Izoformy proteínu p53 sú generované kombináciou použitia alternatívnych promótorov (P1 a P2), alternatívneho zostrihu a alternatívnej iniciácie translácie [188]. Alternatívny zostrih intrónu 9 môže viesť k vzniku mRNA varianty s exónom 9β alebo 9γ, pričom vznikajú izoformy β alebo γ. mRNA traskribovaná z promótoru P1 so zostrihnutým intrónom 2 môže byť preložená z prvého AUG kodónu (lokalizovaného v exóne 2) za vzniku p53, p53β, p53γ, a z AUG40 kodónu v dôsledku vnútorného vstupného miesta pre ribozómy v 5'UTR z exón 1 – exón 2 vedúceho k expresii izoforiem Δ40p53, Δ40p53β, Δ40p53γ [190, 191]. Transkripcia mRNA *TP53* môže byť iniciovaná z vnútorného promótora P2 umiestneného na intróne 4. Translácia týchto transkriptov z iniciačných kodónov 133 a 160 môže generovať izoformy Δ133p53 a Δ160p53, a vzhľadom na alternatívne zostrihy v intróne 9 môže dochádzať k vzniku Δ133p53, Δ133p53β, Δ133p53γ a Δ160p53, Δ160p53β, Δ160p53γ [188, 192].

N-terminálnym izoformám chýba časť alebo celá transaktivačná doména. Na tomto základe sa predpokladalo, že ich hlavnou funkciou bude pôsobiť ako dominantne negatívne regulátory aktivity proteínu p53. Niekoľko štúdií ale naznačuje, že Δ40p53 inhibuje bazálnu aktivitu proteínu p53 počas progresie bunkového cyklu [187, 189], ale nie je jasné, či táto izoforma môže fungovať bez závislosti na p53. Keďže obsahuje nenarušenú DNA väzbovú doménu, môže dochádzať k súťaženiu s p53 o väzbu na cieľové sekvencie tohto proteínu, a tým k modulácii ich dostupnosti pre iné transkripčné faktory. Navyše v Δ40p53 chýba prvá časť (TAD I), ale zachováva si druhú časť transaktivačnej domény (TAD II), pomocou ktorej je schopná regulovať

génovú expresiu [193, 194]. Na rozdiel od Δ40p53 sa izoformy Δ133p53, Δ160p53 neviažu na cieľové sekvencie p53, keďže im chýba časť DNA väzbovej domény, a fungujú ako negatívny inhibítor proteínu p53 [195, 196] alebo sa môžu správať ako dominantný mutantný p53 proteín [182].



Obrázok 3. Schematické znázornenie domén ľudských p53 izoforiem vrátane dvoch transaktivačných domén (TAD I – svetlomodrá, TAD II – tmavomodrá), domény bohatej na prolín (PRD – fialová), DNA väzbovej domény (DBD – žltá), C-terminálnej domény obsahujúcej signál jadrovej lokalizácie (NLS – svetlozelená) a oligomerizačnú doménu (OD – tmavozelená). Upravené z [192, 197].

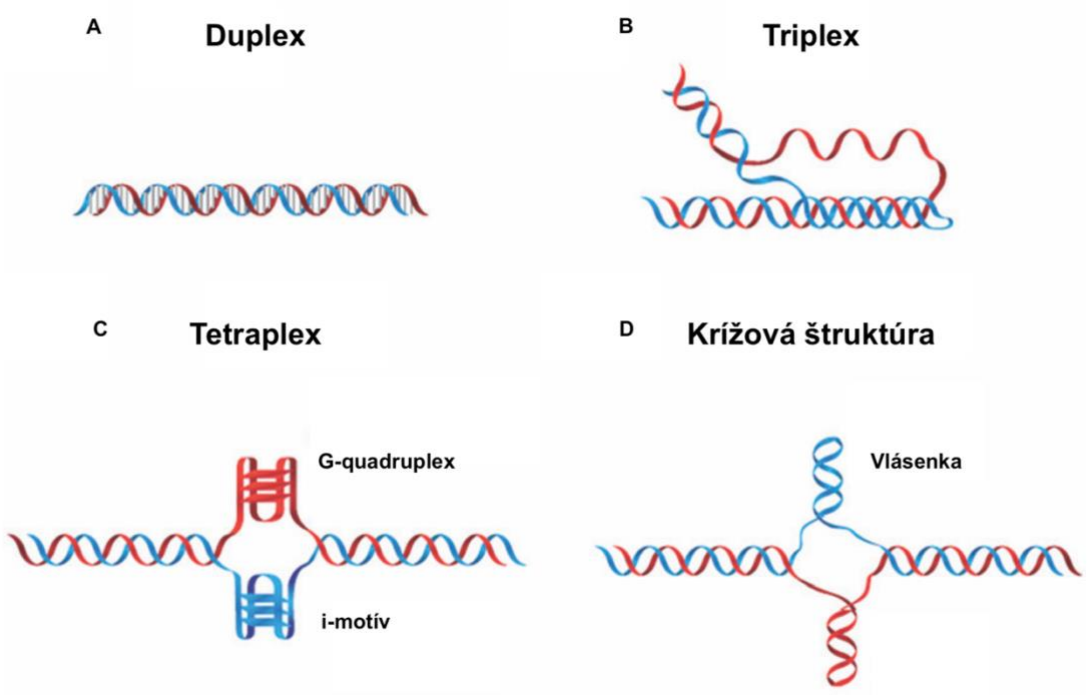
Biologické funkcie ďalších izoforiem sú popísané slabo. Niekoľko štúdií naznačuje spoluúčasť p53β izoformy na bunkovom starnutí v ľudských fibroblastoch [198].

Celkovo, súčasné experimentálne údaje o biologických úlohách p53 izoforiem sú neúplné. Na základe toho, že sa tieto izoformy líšia v troch funkčných doménach (TAD,

DBD a OD), je možné, že ich potenciál modulovať reakcie závislé od p53 bude rozmanitý a závislý od typu bunky.

1.3 Nekanonické štruktúry nukleových kyselín

Nukleové kyseliny sú polymorfné molekuly, ktoré môžu v bunke zaujať rôzne konformácie. Medzi najčastejšie vyskytujúce sa formy DNA patrí pravotočivá dvojzávitnica, označovaná ako B-DNA, popísaná Watsonom a Crickom v roku 1953 [199]. Nie je prekvapujúce, že okrem kanonickej štruktúry (Obrázok 4A), ktorá je v bunkách prevládajúcou konfiguráciou, môžu nukleové kyseliny tvoriť nekanonické sekundárne štruktúry ako triplexy [200], G-quadruplexy [201, 202], i-motívy [203] alebo krížové štruktúry [204] (Obrázok 4). Tvorba nekanonických štruktúr závisí od nukleotidovej sekvencie a okolitých podmienok, ako je relatívna vlhkosť prostredia, pH alebo väzba proteínov [205, 206].



Obrázok 4. Schémy rôznych foriem DNA štruktúr. Kanonická DNA (A) duplexová štruktúra a nekanonické štruktúry; triplex (B), tetraplex (C) a krížová štruktúra (D). Prevzaté z [208].

Mnoho cieľových génov pre proteín p53 obsahuje kanonický RE, no okrem toho sa p53 viaže tiež na DNA v miestach, kde sa táto sekvencia nevyskytuje, a to v miestach výskytu práve nekanonických DNA štruktúr [207].

1.3.1 Triplex

Triple-helix, alebo triplex, zaujíma štruktúru charakterizovanú tretím vláknom DNA, bohatým na pyrimidín alebo purín, umiestneným vo veľkom žliabku homopurínového/ homopyrimidínového úseku duplexu DNA [200, 207, 209, 210]. Stabilná interakcia tretieho vlákna je dosiahnutá buď špecifickou Hoogsteenovou alebo reverznou Hoogsteenovou vodíkovou väzbou na homopurínové vlákno duplexnej DNA. Boli charakterizované intermolekulárne triplexy, kde tretie vlákno pochádza zo samostatnej molekuly DNA, alebo intramolekulárne triplexy, v ktorých tretie vlákno pochádza z rovnakej DNA molekuly [200, 207, 209].

Predpokladá sa, že intramolekulárne triplexy sa *in vivo* vyskytujú len za vhodných podmienok, ako je dostatočne veľký negatívny superhelikálny stres a podieľajú sa na niekoľkých bunkových procesoch, vrátane transkripcie, replikácie a rekombinácie [200]. Na druhej strane intermolekulárne triplexy sú uznávanými nástrojmi genetickej manipulácie vrátane génovej regulácie a mutagenézy [211, 212].

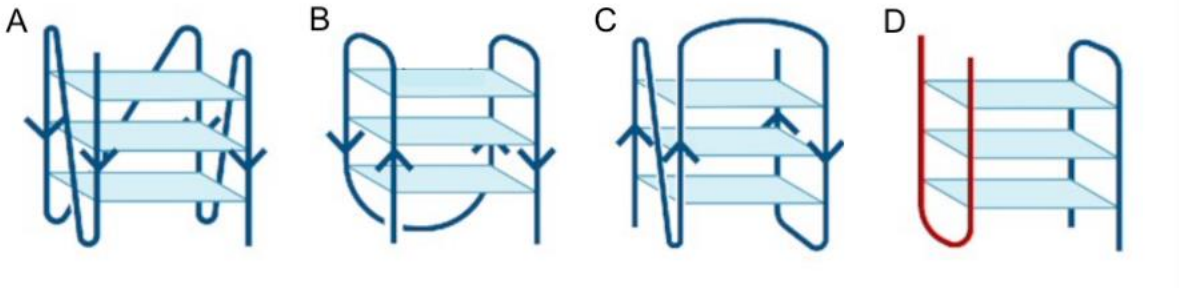
Napriek korelácii medzi genomickou nestabilitou a tvorbou triplexnej DNA je funkcia proteínov, ktoré tieto štruktúry rozpoznávajú, stále málo objasnená. Ukázalo sa, že niekoľko proteínov sa na triplexovú DNA viaže [207, 213].

1.3.2 Tetraplex

1.3.2.1 G-kvadruplex

G-kvadruplexy (G4) sú sekundárne štruktúry nukleových kyselín, ktoré môžu byť formované ako v DNA tak v RNA [214], v oblastiach bohatých na guanín za fyziologických podmienok [215–218]. K tvorbe dochádza, ak dôjde k uloženiu dvoch alebo viacerých guanínových tetrád nad seba koordinovaných pomocou monokovalentných katiónov, ako je draslík alebo sodík. Každá tetráda pozostáva zo štyroch guanínových zvyškov, ktoré sa viažu na cukor-fosfátovú kostru a navzájom sú

pospájané pomocou Hoogsteenových vodíkových väzieb. G4 sú polymorfné štruktúry, ktorých topológia môže byť ovplyvnená stechiometriou a polaritou reťazca, tiež dĺžkou slučiek medzi jednotlivými tetrádami a ich umiestnením v sekvencii [219]. G4 sa môžu skladať intermolekulárne, z jedného vlákna bohatého na guaníny, alebo intramolekulárne, prostredníctvom dimerizácie alebo tetramerizácie oddelených vlákien [216, 220–222]. Orientácia reťazcov môže byť definovaná paralelnou, antiparalelnou alebo kombinovanou topológiou (hybrid), ktorá priamo koreluje konformačným stavom glykozidickej väzby medzi guanínom a cukrom [214](**Obrázok 5**).



Obrázok 5. Schematická reprezentácia G4 štruktúry v paralelnej (A), anti-paralelnej (B) alebo hybridnej (C) forme, v závislosti na relatívnej orientácii reťazca DNA v štruktúre. Môže dochádzať k formácii intermolekulárnej štruktúry (D), a to v prípade, ak sa na tvorbe štruktúry podieľa viac ako jeden reťazec DNA [223].

Použitím bioinformatických nástrojov bolo zistené, že ľudský genóm obsahuje viac ako 700 000 oblastí potenciálne tvoriacich G4 štruktúry [224, 225]. Rovnako aj v iných organizmoch ako sú vírusy [116–118, 119], archea [230], baktérie [231] alebo kvasinky [232] boli *in silico* analýzou nájdené potenciálne G4 motívy. Okrem analýzy celkového počtu sekvencií potenciálne tvoriacich G4 bola analyzovaná aj ich lokalizácia v genóme. Na základe týchto štúdií je možné konštatovať, že G4 motívy nie sú v genóme distribuované náhodne, ale vyskytujú sa v určitých špecifických regiónoch, ako sú promótory, teloméry alebo väzbové miesta pre transkripčné faktory [233, 234]. Evolučná konzervovanosť a špecifická lokalizácia v genómoch, ako aj rôzne biochemické a molekulárne experimenty dokazujú, že G4 sa tvoria v živých bunkách, kde ovplyvňujú rôzne biologické dráhy [201, 217, 235].

Fyziologická dôležitosť G4 štruktúr podporuje existencia proteínov, ktoré sa na tieto štruktúry viažu [236]. Existuje niekoľko tried takýchto proteínov, ktoré viažu, stabilizujú, alebo rozkladajú tieto štruktúry [237]. Delécia alebo mutácia v génoch kódujúcich tieto proteíny môže viesť k zmenám vo formácii G4, ktoré môžu viesť až k zmenám v transkripcii a viesť k zníženiu genómovej stability [238–242]. Spomínané zmeny môžu byť cielené chemickou cestou, pomocou tzv. G4 ligandov, ktoré sa líšia v špecifite, väzbovom povrchu alebo schopnosti bunkovej permeabilizácie [243].
Bez ohľadu na to, že cielenie pomocou niektorých G4 ligandov vykazuje sľubné výsledky v kontexte nových terapeutík [219, 220], ich klinické použitie zatiaľ nie je schválené. To je spôsobené prevažne problémami so selektivitou, ale v posledných rokoch bolo vynaložené veľké úsilie na vývoj takých G4 ligandov, ktoré by mali vysokú protinádorovú aktivitu so znížením vedľajších účinkov [243, 244].

1.3.2.2 I-motív

I-motívy sú DNA sekundárne štruktúry tvorené v sekvenciách bohatých na cytozín [245] s možným širokým využitím v nanotechnológiách a s potenciálom ovplyvniť biologické procesy [203, 246]. I-motív pozostáva z interkalovaných párov protónovaných cytozínov (cytozín$^+$-cytozín) medzi štyrmi vláknami nukleových kyselín [206, 247]. Vzhľadom k nutnosti protonizácie cytozínov bolo predpokladom tvorby i-motívov kyslé pH prostredie, ale stabilná tvorba týchto štruktúr bola dokázaná aj pri zásaditom [248], neutrálnom pH [198, 245, 249] a v podmienkach pripomínajúcich fyziologické prostredie [250]. Medzi doposiaľ odhalené funkcie týchto štruktúr patrí schopnosť inhibovať DNA polymerázu [251] a tvorba i-motívov bola zaznamenaná v promótorových oblastiach niekoľkých génov spojených s tvorbou rakoviny, medzi ktoré patria napríklad transkripčné faktory alebo onkogény [252–254].

V zásade komplementárne vlákno akejkoľvek sekvencie tvoriacej G4 je náchylné k tvorbe i-motívu [203]. Bolo dokázané, že tvorba jednej štruktúry destabilizuje druhú na komplementárne vlákno [255]. Na tomto základe je možné predpokladať odlišné funkcie v transkripčnej regulácii.

1.3.3 Krížové štruktúry

Krížové štruktúry sa tvoria v dvojvláknovej DNA z palindromatických sekvencií, ktoré obsahujú úplne alebo čiastočne identické obrátené repetície. Ak sú tieto sekvencie úplne identické (neobsahujú medzerník) nazývajú sa perfektnými. V prípadne neperfektných repetícií, medzerníky tvoria slučku [256].

Palindromatické sekvencie boli nájdené vo všetkých genómoch, kde zohrávajú dôležité úlohy ako väzbové miesta, sú súčasťou promótorov, replikačných počiatkov alebo iných regulačných sekvencií [204]. Tieto sekvencie, ktoré majú potenciál vytvárať sekundárne štruktúry v DNA, predstavujú riziko pre stabilitu genómu a môžu vyústiť do dvojvláknových zlomov DNA, ktoré vedú ku genetickej rekombinácii, čo môže potenciálne viesť k translokáciám, deléciám alebo amplifikácii génov s rôznymi dôsledkami na organizmus [257–260].

2 CIELE DIZERTAČNEJ PRÁCE

V rámci dizertačnej práce boli spracované nasledujúce čiastočné úkony:

a) Štúdium literatúry a literárny prehľad danej problematiky so zameraním na bunkové systémy a ich biotechnologické využitie;
b) Biochemická charakterizácia chemoterapeutík a prírodných látok pri interakcii s biologickými molekulami so zameraním na DNA a jej štruktúrne motívy, analýza štruktúrnych motívov DNA v genómoch;
c) Kultivácia bakteriálnych a kvasinkových kultúr a testovanie viability týchto kultúr, funkčné testovanie, charakterizácia a biotechnologické využitie;
d) Kultivácia eukaryotických nenádorových a nádorových kultúr a testovanie viability týchto kultúr po pôsobení chemoterapeutík a prírodných látok.

K dosiahnutiu vopred vytýčených cieľov bolo potrebné zvládnuť rôzne laboratórne postupy a metódy:

a) Pokročilé molekulárne biologické metodiky vrátane izolácií nukleových kyselín, PCR techník, kultivácií a transformácií/ transfekcií bunkových kultúr, western blot analýzy, elektroforetických metód a iných;
b) Fluorescenčná a konfokálna mikroskopia pre stanovenie viability buniek. a indukcia apoptózy po pôsobení látok na rôzne bunkové kultúry/ pre lokalizáciu aktivovaných proteínov a štúdium zmien v molekulách nukleových kyselín, luciferázová analýza a ďalšie;
c) Využitie fluorescenčnej anizotropie a ďalších techník pre štúdium interakcií testovaných látok s proteínmi a nukleovými kyselinami, testovanie viability a genotoxicity na bunkových líniách v závislosti na statuse p53 izoforiem.

3 EXPERIMENTÁLNA ČASŤ

Experimentálna časť práce pozostáva z komentovaného súboru publikácií, ktoré sú rozdelené na tri základné časti. Prvá časť sa zaoberá štúdiom nekanonických lokálnych štruktúr nukleových kyselín prevažne v kvasinkových izogénnych systémoch alebo *in vitro* s ohľadom na tumor supresorový proteín p53, druhá časť obsahuje súbor publikácií zameraných na štúdium nekanonických štruktúr nukleových kyselín v rôznych skupinách organizmov pomocou bioinformatických nástrojov. Posledná časť obsahuje doposiaľ nepublikované výsledky.

Práca je koncipovaná tak, že publikované práce sú komentované prehľadne s odkazom na príslušnú prílohu. Prílohy obsahujú recenzované postupy a výstupy, ktoré už boli posúdené v rámci recenzného riadenia.

3.1 Zoznam vedeckých publikácií

3.1.1 Chronologicky zoradené vedecké publikácie

[P1] **PORUBIAKOVÁ, Otília**, Natália BOHÁLOVÁ, Alberto INGA, Natália VADOVIČOVÁ, Jan COUFAL, Miroslav FOJTA a Václav BRÁZDA. The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms. *International Journal of Molecular Sciences* [online]. 2019, **21**(1), 127. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms21010127

[P2] ČUTOVÁ, Michaela, Jacinta MANTA**, Otília PORUBIAKOVÁ**, Patrik KAURA, Jiří ŠŤASTNÝ, Eva B. JAGELSKÁ, Pratik GOSWAMI, Martin BARTAS a Václav BRÁZDA. Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae. *Genomics* [online]. 2020, **112**(2), 1897–1901. ISSN 08887543. Dostupné z: doi:10.1016/j.ygeno.2019.11.002

[P3] BARTAS, Martin, Václav BRÁZDA, Natália BOHÁLOVÁ, Alessio CANTARA, Adriana VOLNÁ, Tereza STACHUROVÁ, Kateřina MALACHOVÁ, Eva B. JAGELSKÁ, **Otília PORUBIAKOVÁ**, Jiří ČERVEŇ a Petr 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. *Frontiers in Microbiology* [online]. 2020, **11**, 1583. ISSN 1664-302X. Dostupné z: doi:10.3389/fmicb.2020.01583

[P4] BRÁZDA, Václav, Yu LUO, Martin BARTAS, Patrik KAURA, **Otilia PORUBIAKOVÁ**, Jiří ŠŤASTNÝ, Petr PEČINKA, Daniela VERGA, Violette DA CUNHA, Tomio S. TAKAHASHI, Patrick FORTERRE, Hannu MYLLYKALLIO, Miroslav FOJTA a Jean-Louis MERGNY. G-Quadruplexes in the Archaea Domain.

[P5] BRÁZDA, Václav, **Otília PORUBIAKOVÁ**, Alessio CANTARA, Natália BOHÁLOVÁ, Jan COUFAL, Martin BARTAS, Miroslav FOJTA a Jean-Louis MERGNY. G-quadruplexes in H1N1 influenza genomes. *BMC Genomics* [online]. 2021, **22**(1), 77. ISSN 1471-2164. Dostupné z: doi:10.1186/s12864-021-07377-9

[P6] MONTI, Paola, Vaclav BRAZDA, Natália BOHÁLOVÁ, **Otília PORUBIAKOVÁ**, Paola MENICHINI, Andrea SPECIALE, Renata BOCCIARDI, Alberto INGA a Gilberto FRONZA. 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* [online]. 2021, **12**(2). ISSN 2073-4425. Dostupné z: doi:10.3390/genes12020277

3.1.2 Chronologicky zoradené konferenčné príspevky

[C1] **PORUBIAKOVÁ Otília,** Natália BOHÁLOVÁ a Václav BRÁZDA, V. *P53alpha protein isoforms binds preferentially to cruciforms in transactivation Yeast isogenic system.* Prezentované na 46th Annual Conference on Yeast, Smolenice, Slovensko, 7. – 10.5.2019.

[C2] BRÁZDA Václav, **Otília PORUBIAKOVÁ,** Natália BOHÁLOVÁ, Alberto INGA a Miroslav FOJTA. *Attenuation of p53-alpha isoform transactivation by inverted repeat sequences in p53 target sites.* Konferenčný príspevok, 8th Mutant Workshop & p53 Isoforms, Lyon, Francúzsko, 15. – 18.5.2019.

[C3] **PORUBIAKOVÁ Otília,** Natália BOHÁLOVÁ A Václav BRÁZDA. *Attenuation of p53-alpha isoform binding to DNA by inverted repeat abolishment in p53 target.* Prezentované na FEBS 2019, Krakow, Poľsko, 6. – 11.7.2019.

[C4] **PORUBIAKOVÁ Otília**, Natália BOHÁLOVÁ a Vácalv BRÁZDA. *G-quadruplex structure in proximity of p53 target sequence causes significant inhibition on transactivation potential of p53 isoforms.* Konferenčný príspevok prezentovaný na konferencii ŠVK- Chémia a technológia pre život, Bratislava, Slovenská republika, 25.11.2020.

[C5] **PORUBIAKOVÁ Otília,** Natália BOHÁLOVÁ a Václav BRÁZDA. *The influence of non-canonical structure on the p53 isoforms binding to DNA.* Konferenčný príspevok prezentovaný na konferencii Chemie je život, Brno, Česká republika, 27.11.2020.

[C6] **PORUBIAKOVÁ Otília**, Alessio CANTARA, Veronika Přepechalová a BRÁZDA Václav. *Comparison between synthehic and natural G4-ligands*

and the effects of their binding to DNA. Prezentované na FEBS2021, Ľubľana, Slovinsko, 8.7.2021.

3.1.3 Zoznam ďalších výsledkov pripravovaných k publikácií

[NP1] GOSWAMI Pratik, **Otília PORUBIAKOVÁ**, Jakub VILÍMEK, Jan Coufal, Miroslav FOJTA a Václav BRÁZDA. *P53-C-terminal isoforms vary in their binding preferences to p53 response element and G-quadruplex structures in DNA.*

[NP2] **PORUBIAKOVÁ Otília**, Michal ŠEDÝ, Veronika PŘEPĚCHALOVÁ, Martin BARTAS, Stefan BIDULA, Miroslav FOJTA a Václav BRÁZDA. *Variability of inverted repeats in all accessible genomes of bacteria.*

[NP3] **PORUBIAKOVÁ Otília**, Ingo LÄMMERMANN, Alessio CANTARA, Johannes GRILLARI a Václav BRÁZDA. Blocking negative effects of senescence using G4-ligands

[NP4] **PORUBIAKOVÁ Otília**, Alessio CANTARA, Aleš DAŇHEL, Jean-Louis MERGNY a Vácalv BRÁZDA. *Unlocking of new G4-ligands.*

3.2 Komentovaný súbor publikačných príspevkov

3.2.1 Prvá časť

3.2.1.1 P1: Vplyv kvadruplexovej štruktúry v blízkosti p53 cieľovej sekvencie na transaktivačný potenciál p53alfa izoforiem

Proteín p53 je jedným z najštudovanejších tumor supresorových proteínov, ktorý hrá kľúčovú úlohu v základných biologických procesoch vrátane bunkového cyklu, bunkovej odpovede na poškodenie DNA, apoptózy a bunkového starnutia. Gén *TP53* kódujúci tento proteín obsahuje alternatívne promótory produkujúce skrátené proteíny v N-terminálnej doméne, ktoré môžu vytvárať niekoľko izoforiem v dôsledku

alternatívneho zostrihu. Proteín p53, ako transkripčný faktor, realizuje svoju funkciu väzbou na špecifické DNA RE s výslednou transaktiváciou cieľových génov. V tejto publikácii bol vyhodnotený vplyv G-kvadruplexovej štruktúry na transaktivačný potenciál tzv. „wild type“ p53 (wtp53)proteínu a jeho izoforiem modifikovaných na N-konci v paneli luciferázových reportérových kmeňov *S. cerevisiae.* Bolo potvrdené, že G-kvadruplex hrá dôležitú úlohu pri interakcii proteín-DNA.

V prvom rade bola vykonaná bioinformatická analýza 100 bp dlhých sekvencií obklopujúcich cieľovú sekvenciu p53 v promótori PUMA. V tesnej blízkosti tohto promótoru bolo nájdených niekoľko potenciálnych sekvencií tvoriacich G4.

Na základe zistení, že v okolí PUMA promótoru sa za bežných podmienok G4 nachádzajú, bol pripravený vhodný modelový systém, v ktorom bol analyzovaný vplyv sekvencie s vysokou pravdepodobnosťou tvorby G4 štruktúry umiestnený pred alebo za stredne aktívnym p53 RE (PUMA promótor) pomocou kvasinkového reportérového systému. Transaktivačný potenciál proteínu p53 a jeho izoforiem bol analyzovaný samostatne a tiež v kombinácii kratších izoforiem s wtp53.

Na základe výsledkov je možné usudzovať, že G4 lokalizované v tesnej blízkosti p53 RE môžu určovať účinnosť transkripčnej regulácie proteínu p53. Kombináciou izoforiem je možné docieliť zvýšenie plasticity kompromisom medzi wtp53 v RE miestach a menej účinnými, ale stéricky výhodnejšími heterotretramérmi v cieľových miestach obklopených štruktúrnymi motívmi ako G4.

Publikácia v Prílohe P1 (The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms). Prvoautorka sa podieľala na experimentálnej časti, formálnej analýze a vizualizácií výsledkov, napísala originálnu verziu manuskriptu (podiel 30 %).

3.2.1.2 P6: Vyhodnotenie vplyvu sekvencie náchylnej k tvorbe G-kvadruplexu na transaktivačný potenciál prírodným a/ alebo mutantnými proteínmi rodiny p53 prostredníctvom funkčného testu v kvasinkách

Proteíny rodiny p53, ako sú p53, p63 a p73, patria medzi transkripčné faktory, ktoré zdieľajú spoločnú génovú organizáciu, konzervatívnu DNA väzbovú doménu,

zodpovedajúcu za sekvenčne špecifickú väzbu na DNA. Tieto proteíny môžu ovplyvňovať mnoho bunkových dráh hlavne cez špecifickú väzbu na DNA sekvencie známe ako RE, a transaktiváciu prislúchajúcich cieľových génov. V tomto článku bola použitá funkčná analýza v kvasinkových izogénnych systémoch, kde bol hodnotený vplyv G4 sekvencie susediacej s p53 RE odvodenej od apoptotického cieľového génu PUMA na transaktivačný potenciál izoforiem proteínov p53 rodiny modifikovaných na N-konci, wtp53 a ich mutantov.

Výsledky ukazujú, že prítomnosť G4 štruktúr pred alebo za p53 RE vedie k významným zmenám v relatívnej aktivite proteínov rodiny p53, čo zdôrazňuje úlohu štruktúrnych znakov DNA ako modifikátorov funkcií rodiny p53 na cieľových promótorových miestach.

Príloha P2 (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). Autorka dodala poznatky o metodológií a validácií, podieľala sa na spracovaní dát a editácii (podiel 10 %).

3.2.2 Druhá časť

3.2.2.1 P2: Rozdielna distribúcia invertovaných opakovaní a G-kvadruplex formujúcich sekvencií v Saccharomyces cerevisiae

Predložená práca je zameraná na štúdium nekanonických štruktúr DNA, spomedzi ktorých inverzné repetície (IR), ktoré môžu tvoriť krížové štruktúry, a sekvencie bohaté na guanín, ktoré môžu tvoriť G-kvadruplexy, sú široko rozšírené v prokaryotických a eukaryotických organizmoch a sú cieľmi mnohých regulačných proteínov. Tieto DNA štruktúry boli analyzované v genóme najvýznamnejšieho biotechnologického mikroorganizmu, kvasinky *S. cerevisiae*.

Významnou úlohou inverzných repetícií je ich účasť na regulácii transkripcie, a to interakciou s rôznymi proteínmi vrátane ľudského tumor supresorového proteínu p53. Sú to práve bioinformatické nástroje, ktoré umožňujú analýzu kompletných genómov a prinášajú komplexnejší pohľad na štruktúru a reguláciu DNA. Analýzou úplne sekvenovaného genómu *S. cerevisiae* (z NCBI, členený na 11 chromozómov

a mtDNA) pomocou *Palindrome analyseru* [261] bolo identifikovaných 8 951 IR s frekvenciou 0,74 IR/kbp a bolo zistené, že krátke IR sú v mtDNA hojnejšie zastúpené ako v chromozomálnej DNA, a že väčšina IR sa nachádza v centromérických oblastiach chromozómov.

Ďalším krokom bola analýza prítomnosti G4 štruktúr pomocou online nástroja *G4Hunter* [262] na základe objavenia niekoľkých proteínov, ktoré interagujú s G4 v *S. cerevisiae*. Predpokladanou úlohou proteínov, ktoré viažu G4, je regulácia dĺžky telomér a transkripčné umlčanie alebo aktivácia niektorých proteínov, čo robí z G4 účinný regulačný nástroj. V práci bolo zistené, že G4 štruktúry sa hojne vyskytujú v telomérických oblastiach a v tesnej blízkosti niekoľkých dôležitých regulačných oblastí.

Pomocou štatistickej analýzy sme demonštrovali nenáhodné umiestnenie IR a G4 pričom k obohateniu dochádza v špecifických chromozómoch a určitých regiónoch jadrového a mitochondriálneho genómu *S. cerevisiae*. Výsledky poukazujú na odlišné funkcie týchto štruktúr a rozdiely vo frekvenciách lokálnych štruktúr naznačujú odlišnosť regulácie expresie DNA v kruhových a lineárnych DNA.

Publikácia sa nachádza v Prílohe P4 (Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae). Autorka sa podieľala na bioinformatických analýzach a štatistickom vyhodnocovaní (podiel 10 %).

3.2.2.2 P3: Podrobná bioinformatická analýza Nidovirales vrátane ľudského SARS-CoV-2, SARS-CoV, MERS-CoV vírusov naznačuje dôležité úlohy nekanonických štruktúr nukleových kyselín v ich životnom cykle

Nekanonické štruktúry nukleových kyselín hrajú dôležitú úlohu v regulácii bunkových procesov. Vzhľadom na pretrvávajúcu koronavírusovú krízu, sme sa v tomto článku zamerali na analýzu týchto štruktúr vo všetkých dostupných sekvenovaných genómoch radu *Nidovirales* z NCBI („National Center for Biotechnology Information“) databázy.

Analýza potenciálnych G-kvadruplexových štruktúr (PQS) vo vírusoch demonštruje ich konzervovanosť a tým možnosť ich použitia ako cieľov pre antivírusové terapie.

Tento článok je zameraný na analýzu všetkých dostupných genómov *Nidovirales*, vrátane SARS-CoV-2 genómu. Analýzou pomocou nástroja *G4Hunter* bola nájdená v SARS-CoV-2 iba jedna PQS, ktorá ale nebola nájdená pomocou iných bioinformatických algoritmov.

Boli skúmané tri patogénne ľudské koronavírusy (SARS-CoV, SARS-CoV-2 a MERS-CoV), aby bolo možné porovnať kľúčové rozdiely v ich hlavných doménach.

Analýzou IR bolo nájdené vyššie množstvo IR v 5'UTR regióne *Nidovirales* genómov. Táto oblasť je vo všeobecnosti dôležitým miestom pre reguláciu replikácie vírusu a génovú expresiu a tým pádom aj ďalším možným cieľom pre vírusovú reguláciu.

Boli nájdené významné rozdiely v distribúcii PQS a IR v rámci rôznych vírusov z danej skupiny, čo naznačuje, že ich organizácia a regulácia genómu je rozdielna, a že u niektorých vírusov z tejto skupiny pravdepodobne nehrá prítomnosť G4 zásadnú úlohu v biologickej regulácii. Tiež bolo dokázané, že G4 boli evolučne eliminované v niektorých genómoch skupiny *Nidovirales*, pričom predpokladom je možná evolučná výhoda, keďže množstvo hostiteľských bunkových proteínov interaguje s týmito sekundárnymi štruktúrami. Na druhej strane, prítomnosť IR predstavuje neoddeliteľnú súčasť genómov a to predovšetkým kvôli umožneniu správneho skladania proteínov a štruktúrnej špecifickej regulácie ich funkcií.

Dostupné v prílohe P5 (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). Autorka sa podieľala na príprave originálnej verzie manuskriptu, pracovala na vizualizácií a poskytla zdroje k analýze (podiel 10 %).

3.2.2.3 P4: G-kvadruplexy v doméne Archea

Dôležitosť nekanonických štruktúr v regulácii základných bunkových procesov je stále rozvíjajúcou sa oblasťou výskumu. Spomedzi lokálnych nekanonických DNA štruktúr sú G4 najatraktívnejšími štruktúrami vzhľadom na ich prítomnosť a funkčnú dôležitosť v DNA a RNA, ktoré boli demonštrované v mnohých vírusových,

bakteriálnych a eukaryotických genómoch, vrátane ľudského. Táto štúdia bola zameraná na systematickú analýzu G4 formujúcich sekvencií (PQS) vo všetkých genómoch *Archea* dostupných z NCBI databázy, pričom výskyt a lokalizácia týchto štruktúr bola uskutočnená pomocou *G4Hunter* algoritmu [262].

Celkovou analýzou bolo zistené, že frekvencia výskytu PQS sa extrémne líši v skupinách a podskupinách archeálnych analyzovaných genómov, pričom tento jav môže byť vysvetlený rozdielmi v obsahu GC báz. Na jednej strane existujú organizmy, v ktorých PQS zastúpenie predstavuje menej ako 1 % ich genómu, a na druhej strane existujú také organizmy, v ktorých bolo nájdené významné obohatenie PQS a to prevažne pre skupiny, ktoré žijú v extrémnych podmienkach. Preto bola dodatkovo uskutočnená analýza pomocou BioSample NCBI databázy [263], ktorá umožnila porovnať životné prostredie vybraných arecheí s PQS frekvenciami. Väčšina organizmov s obsahom extrémne vysokých PQS frekvencií bola nájdená v sedimentoch horúcich prameňov alebo v sedimentoch hlbokomorských hydrotermálnych prieduchov, pričom prítomnosť vysokých PQS frekvencií môže byť spojená s ich extrémofilným životom.

Dostupné v Prílohe P6 (G-Quadruplexes in the Archaea Domain). Autorka sa podieľala na príprave, prevedení a vyhodnocovaní analýz (podiel 10 %).

3.2.2.4 P5: G-kvadruplexy v H1N1 chrípkových genómoch

Vírusy chrípky patria medzi nebezpečné patogény, preto je potrebné hľadať nové terapeutické prístupy k protivírusovej liečbe. V tejto publikácii sme analyzovali výskyt a distribúciu PQS v genóme chrípkového vírusu A subtypu H1N1 ako potenciálnych terapeutických cieľov.

Podstatou štúdie bolo analyzovať všetky dostupné reťazce G4-EA-H1N1 („genotype 4 reassortant Eurasian avian-like H1N1 virus“) pomocou bioinformatického nástroja *G4Hunter* [262]. Genóm tohto vírusu je 13 133 nt dlhý a rozdelený do ôsmich segmentov: PB1, PB2, M, HA, NP, NS, PA a NA. Zatiaľ čo najvyššia priemerná frekvencia PQS bola zistená v NP segmente, ktorý kóduje proteín hrajúci ústrednú úlohu v replikácii vírusu, v NS segmente, ktorý kóduje neštrukturálny proteín NS a je

najhojnejšie vyskytujúcim sa proteínom v infikovaných bunkách a zároveň je považovaný za najsľubnejší terapeutický cieľ, nebola nájdená prítomnosť PQS.

Pre vyhodnotenie prítomnosti PQS v jednotlivých fragmentoch boli vybrané dve skupiny H1N1 (s najvyššou a najnižšou PQS frekvenciou), ktoré boli päťkrát randomizované a štatisticky vyhodnotené. Bol zistený významný rozdiel medzi frekvenciami pôvodných a randomizovaných sekvencií pre NA segment, čo naznačuje jeho dôležitú evolučne konzervovanú funkciu.

Tvorba G4 štruktúr bola experimentálne potvrdená dvoma biofyzikálnymi metódami, a to pomocou CD spektroskopie a ThT kompetitívneho testu. Podľa predpokladu, sekvencie s nižším G4Hunter skóre netvorili G4 štruktúru, a na druhej strane sekvencie schopné vytvárať G-kvadruplex vykazovali teoretické G4Hunter skóre vyššie ako 1,2.

Taktiež analýzou oboch vláken, genómového (s negatívnym zmyslom) a mRNA (s pozitívnym zmyslom), vzhľadom na ich účasť v životnom cykle vírusu, bolo zistené, že PQS nie sú rovnomerne rozložené, ale prevažne sa vyskytujú na RNA pozitívnom vlákne, čím sa môžu podieľať na regulácii translácie. Genóm H1N1 nie je stabilný a výrazne sa líši medzi jednotlivými kmeňmi. Bolo nájdených niekoľko vysoko konzervovaných PQS v segmentoch M a HA, ktoré môžu byť považované za sľubné kandidátske terapeutické ciele.

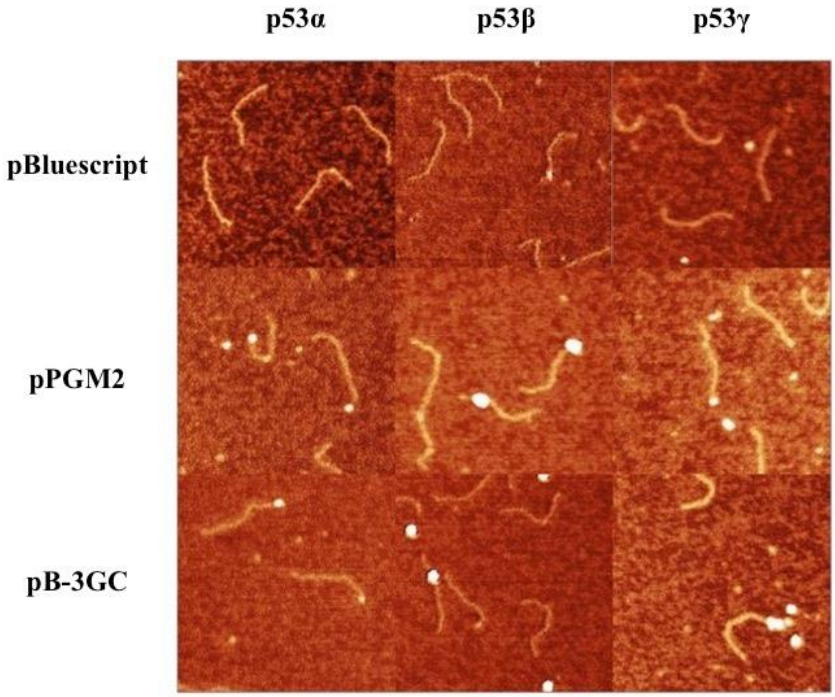
Príloha P7 (G-quadruplexes in H1N1 influenza genomes). Prvoautorka pripravila a interpretovala výsledky analýz a napísala originálnu verziu manuskriptu (podiel 25%).

3.2.3 Doposiaľ nepublikované experimentálne výsledky

3.2.3.1 NP1: C-terminálne izoformy proteínu p53 sa líšia v preferenciách väzby k p53 responzívnemu elementu a G-kvadruplexovým štruktúram v DNA

Proteín p53 je jedným z najvýznamnejších regulačných proteínov ovplyvňujúcich prevenciu rakoviny u ľudí. Zatiaľ čo klasický, wtp53, chráni proti rakovine, jeho mutácie sú často spájané s jej progresiou. Po strate bunkovej homeostázy sú úlohy proteínu p53 primárne sprostredkované jeho väzbou na DNA, kde v aktivovanej forme pôsobí ako transkripčný faktor. Okrem klasickej formy proteínu dochádza vplyvom

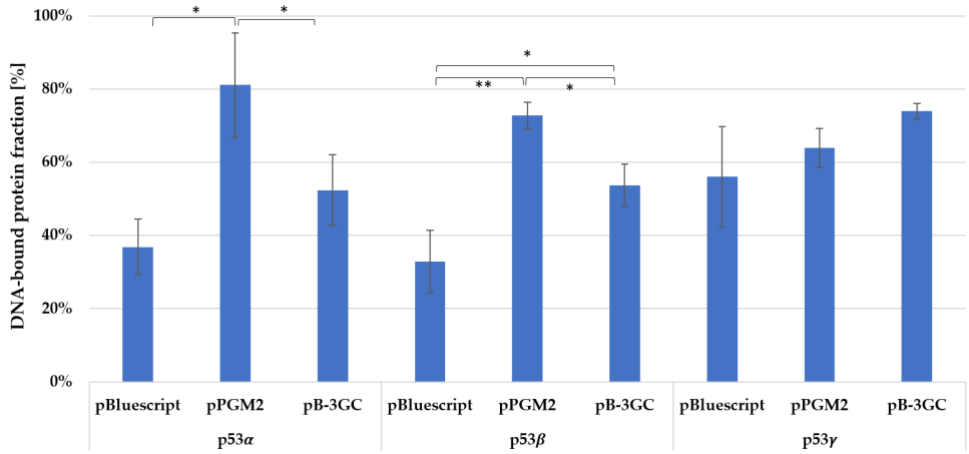
alternatívneho zostrihu exónu 9 k produkcii C-terminálnych izoforiem, a to p53α (wtp53), p53β a p53γ. Pomocou mikroskopie atomárnych síl (AFM) bola skúmaná väzba izoforiem líšiacich sa v C-terminálnej doméne na fragmenty DNA obsahujúce rôzne cieľové sekvencie.



Obrázok 6. Obrázky z AFM pre izoformy proteínu p53 (p53α, p53β and p53γ) viažúce sa na DNA fragmenty pBluescript, pPGM2 a PB-3GC. Meranie bolo prevedené pomocou Veeco Multimode VIII SPM (Digital Instruments, Bruker) s konzolou Scanasyst-AIR s rezonančnou frekvenciou f0 50-90 kHz a konštantou pružiny k-0,4N/m. Zobrazovanie bolo získané v režime Scanasyst-AIR s nasledujúcimi parametrami: 1024 x 1024 pixelov, veľkosť skenu – 2 μm s veľkosťou mierky 500 nm pre všetky obrázky. Obrázky boli spracované pomocou Gwyddion a ImageJ.

Sledované fragmenty DNA boli pripravované reštrikčným štiepením, ktoré umožňovalo upresniť lokalizáciu sledovanej sekvencie (prázdny, RE, G4) na jeden koniec daného fragmentu. Proteín p53α (Obrázok 6) sa viazal náhodne na lineárny (prázdny) fragment DNA (pBluescript). Na druhej strane dochádzalo k výraznejšej väzbe na koniec fragmentu pPGM2 (obsahujúci RE) a pB-3GC (s obsahom G4).

Väzba ďalších dvoch izoforiem p53 na ktorýkoľvek z plazmidových fragmentov (Obrázok 6) bola pozorovaná ako náhodnejšia v porovnaní s p53α.



Obrázok 7. Percentuálne vyjadrené väzby izoforiem proteínu p53α viažuce sa na DNA fragmenty. Štatistická významnosť je vyjadrená ako hviezdička ($p < 0,05$ *; $p < 0,01$ **).

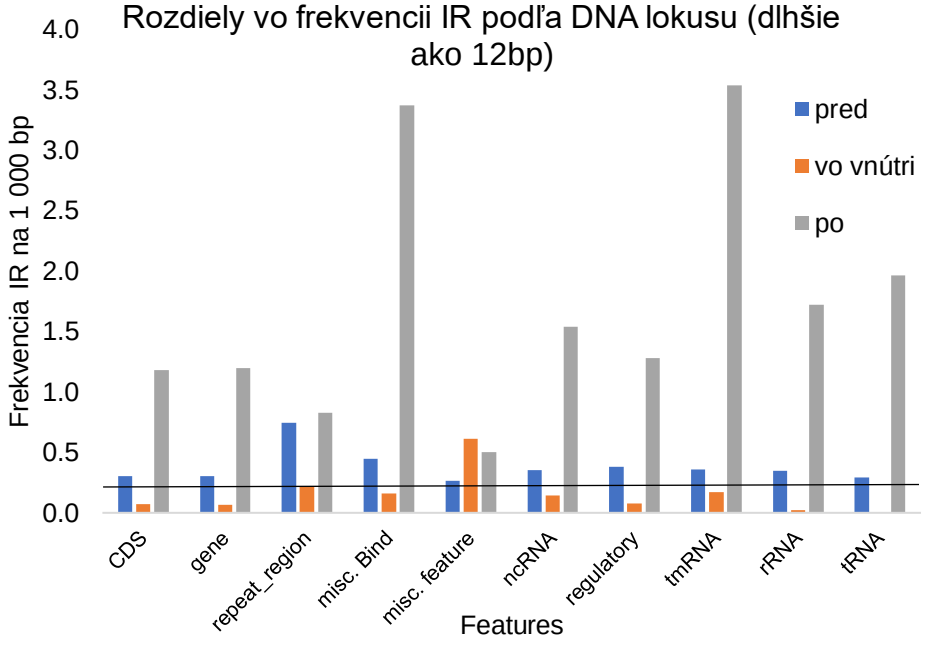
Zatiaľ čo v prípade fragmentu pBluescript sa len 36,8 % molekúl p53α naviazalo na koniec fragmentov, 81,1 % p53α bolo naviazaných na RE na konci fragmentov pPGM2 (reprezentatívne zobrazené na Obrázok 7). Je zaujímavé, že väzbová frekvencia p53α na G4 sekvenciu bola vyššia v porovnaní s väzbou p53α na pBluescript, ale oveľa nižšia v porovnaní s väzbou rovnakých proteínov na RE (52,4 % p53α sa viazalo na koniec fragmentu pB-3GC). Relatívne afinity väzby p53 na RE a sekvenciu náchylnú na G4 boli rôzne pre izoformy p53p a p53y. Zatiaľ čo izoforma p53β sa viazala podobne ako p53α na koniec príslušných fragmentov, rozdiely vo väzbe p53γ na jednotlivé ciele boli oveľa menšie. Prekvapivo bola G4 sekvencia viazaná izoformou p53y s vyššou frekvenciou ako RE (73,9 %, resp. 63,9 %).

Táto porovnávacia väzbová analýza odhalila rôzne väzbové mechanizmy izoforiem proteínu p53, čo naznačuje, že ich regulačné aktivity sú komplexné alebo komplementárne.

Autorka sa podieľala na plánovaní experimentov a štatistickej analýze.

3.2.3.2 NP2: Variabilita inverzných opakovaní vo všetkých dostupných bakteriálnych genómoch

Dôležité funkcie lokálnych DNA štruktúr v bunkových procesoch boli v posledných rokoch intenzívne študované a demonštrované. Význam krížových štruktúr formovaných IR a G4 boli nájdené v rôznych organizmoch, vrátane ľudského genómu. Použitím *Palindrome analyseru* bola analyzovaná prítomnosť IR vo všetkých dostupných bakteriálnych genómových sekvenciách s cieľom zistenia ich frekvencie, dĺžky a lokalizácie. IR boli identifikované vo všetkých druhoch, ale ich frekvencia sa medzi jednotlivými druhmi evolučných skupín výrazne líšila. V 1 565 bakteriálnych genómoch bolo nájdených 242 373 717 IR. Najvyššia priemerná frekvencia bola nájdená v skupine Tenericutes (61,89 IR/kbp) a najnižšia v skupine Alphaproteobacteria (27,08 IR/kbp). Nájdené IR boli najhojnejšie zastúpené v niekoľkých oblastiach, vrátane génov a okolia regulačných oblastí, tRNA, tmRNA a rRNA oblastí (Obrázok 8). Je pozoruhodné, že IR boli nájdené prevažne v regiónoch



Obrázok 8. Rozdiely vo frekvencii IR podľa DNA lokusu. IR dlhšie ako 12 bp v rámci anotovaných miest a 100 bp pred alebo po nich. Čiara označuje priemernú frekvenciu pre IR dlhé 12 bp a dlhšie.

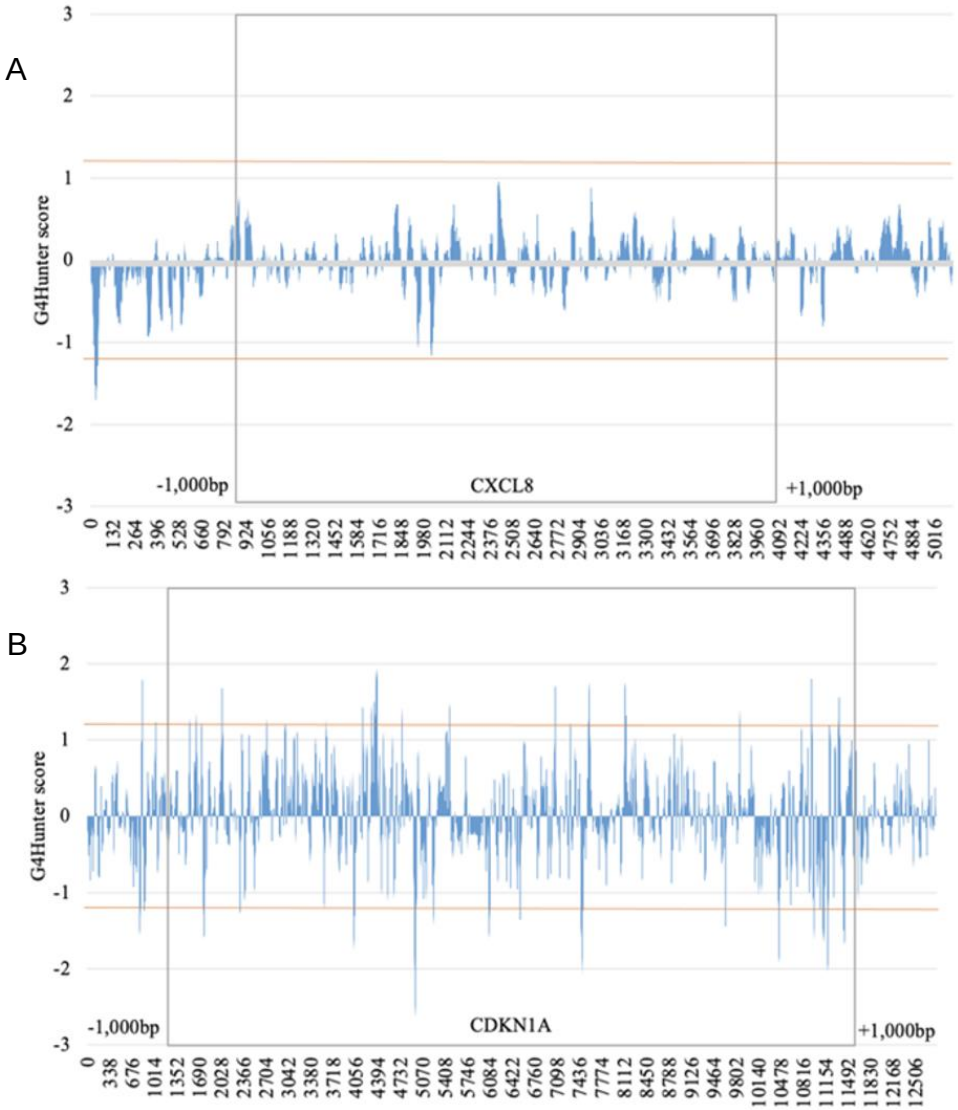
kódujúcich proteíny a génoch pre RNA, čo poukazuje na dôležitosť v základných bunkových procesoch ako je udržovanie genómu, replikácia DNA a transkripcia.

Okrem toho bolo zistené, že organizmy s vysokou frekvenciou IR sú pravdepodobnejšie endosymbiotické, produkujúce antibiotiká alebmo patogénne. Naopak, organizmy s nízkou frekvenciou boli pravdepodobnejšie termofilné. Táto prvá komplexná analýza všetkých dostupných bakteriálnych genómov na prítomnosť a lokalizáciu IR poukazuje na ich výskyt vo všetkých použitých genómoch, ich nenáhodnú distribúciu a obohatenie v regulačných oblastiach genómov.

Prvoautorka previedla prvotnú analýzu setu dát, zozbierala a interpretovala výsledky analýz a napísala prvú verziu manuskriptu.

3.2.3.3 NP3: Blokovanie negatívnych efektov bunkového starnutia pomocou G4-ligandov

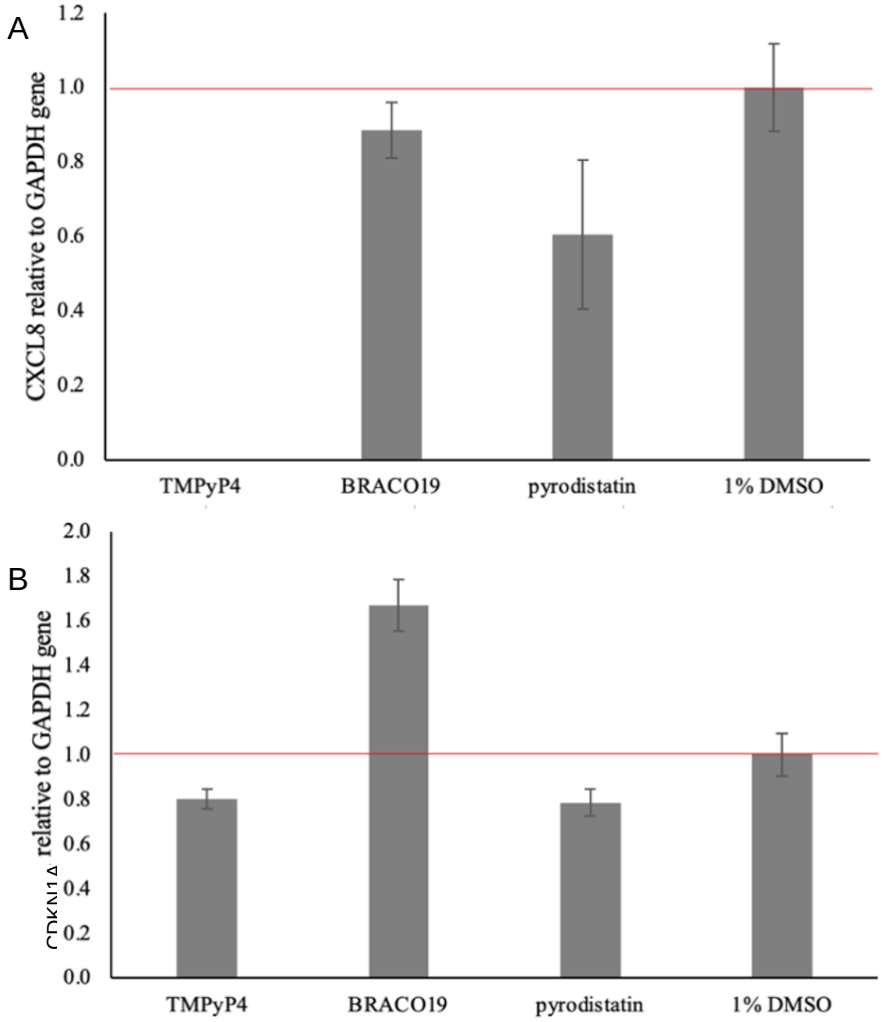
Dôkazy o senescentných bunkách ako hybnej sile rôznych patológií súvisiacich s vekom rastú a ich selektívna eliminácia môže viesť k predĺženiu života a zlepšeniu zdravia. Senescentné bunky negatívne ovplyvňujú okolité prostredie a tkanivá vylučovaním charakteristického sekrečného fenotypu (SASP). Nové prístupy k cieleniu a zvráteniu senescencie by mohli zahŕňať stabilizáciu sekvencií bohatých na guanín, ktoré tvoria G4. Molekuly stabilizujúce tieto štruktúry majú antiproliferačný potenciál k mechanizmom závislým alebo nezávislým na telomérach, ale ich účinok na starnutie ostáva do značnej miery nepreskúmaný. V koncepte starnutia sme identifikovali niektoré potenciálne sekvencie tvoriace G4 v a v okolí génov kódujúcich senescentne asociované markery pomocou G4Hunter-u (Obrázok 9). Sekvencie odpovedajúce týmto génom boli stiahnuté z NCBI a analyzované s parametrami 25 ako veľkosť analyzovaného okna a 1,2 alebo vyššie G4Hunter skóre. Bolo nájdených niekoľko potenciálne významných G4 motívov obklopujúcich gén CDKN1A, čo naznačovalo možnosť regulácie expresie prostredníctvom stabilizácie G4-ligandmi. V ostatných sledovaných znakoch nebola nájdená žiadna potenciálna G4 sekvencia na mRNA reťazci, takže bolo predpokladané, že dané znaky sa s prídavkom G4-ligandov nebudú meniť.



Obrázok 9. Lokalizácia PQS v a v okolí markerov asociovaných so senescenciou (1 000 bp pred a po CXCL8 a CDKN1A sekvencii). Analýza potenciálnych G4 sekvencií ukázali, že sekvencia CXCL8 (A) (v obdĺžniku) nie je obklopená PQ. Na druhú stranu gén CDKN1A (B) obsahuje PQS vo svojej sekvencii a v promótorovej oblasti vykazuje zaujímavé PQS a píkom pri 1,86 a 1,24.

V predloženom článku ukazujeme, že pôsobenie G4 ligandov spôsobuje inhibíciu expresie niektorých významných senescentných zápalových markerov (prevažne chemokíny (Obrázok 10A.)), čo vedie k oddialeniu senescentného fenotypu. Došlo k poklesu expresie o 20 % oproti negatívnej kontrole v géne CDKN1A (kódujúcom $p21^{CIP}$) po ošetrení TMPyP4 a pyridostatínom, čo ukazuje na možné potlačenie expresie prostredníctvom stabilizácie G4 štruktúry v promótorovej oblasti. Rovnako bol zistený pokles expresie CXCL8 závislý na použitom G4-ligande. Najvýraznejší efekt

bol sledovaný na hladine CXCL8 po pôsobení TMPyP4, kedy došlo k redukcii expresie o 99,83 %. Vzhľadom k skutočnosti, že v okolí génu CXCL8 neboli nájdené žiadne potenciálne G4 štruktúry, jediným vysvetlením môže byť kaskádový efekt pôsobenia. Bol dokázaný inhibičný efekt TMPyP4 na expresiu c-MYC proteínu, ktorý by mohol ovplyvňovať hladinu spomínaných chemokínov/cytokínov cez kontrolu hTERT.



Obrázok 10. Vyhodnotenie hladín CXCL8 (A) a CDKN1A (B) v senescentných bunkách ošetrených vybranými G4-ligandmi. RNA bola izolovaná zo senescentných buniek HDF164. Každá fyziologická podmienka bola analyzovaná v biologických triplikátoch. Hladiny expresie boli kvantifikované pomocou qPCR a normalizované na GAPDH.

Záverom bolo demonštrované, že účinok G4-ligandov spôsobuje zníženie génovej expresie niektorých významných markerov spojených so starnutím, čo vedie k zdržaniu senescentného fenotypu. Bol identifikovaný G4-ligand, TMPyP4, ktorý bol schopný blokovať negatívne účinky starnutia buniek a udržiavať ich funkčnosť. Okrem

toho výsledky poukazujú na významný a nedostatočne pochopený dôkaz o vzťahu medzi starnutím a G4 štruktúrami, ktorý bude ďalej študovaný.

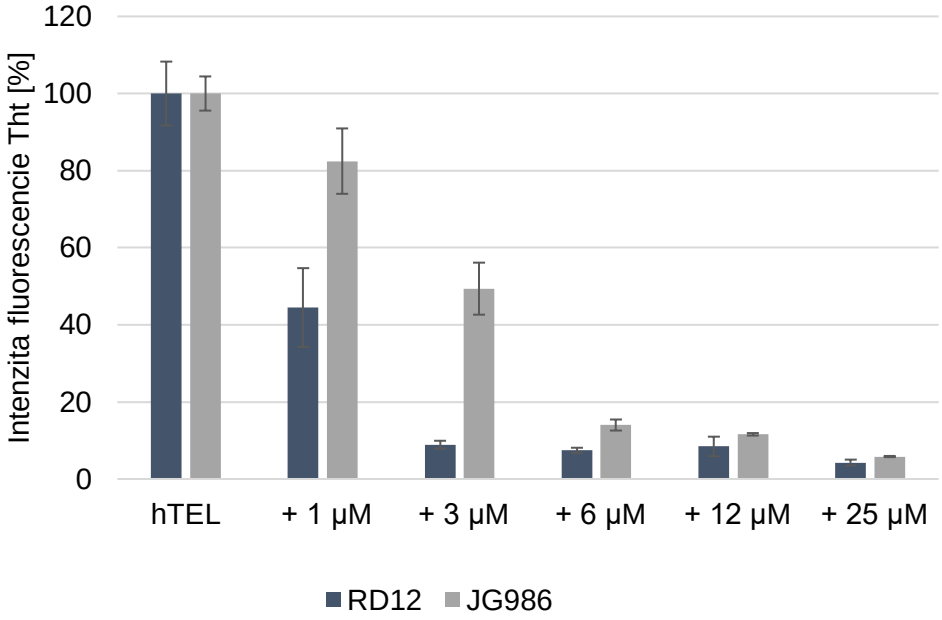
Prvoautorka v rámci pracovnej stáže previedla experimentálnu časť, vyhodnotila výsledky a spracovala prvú verziu manuskriptu.

3.2.3.4 NP4: Charakterizácia nových G4-ligandov

Z pomedzi lokálnych nekanonických DNA štruktúr sú práve sekvencie, ktoré tvoria G4, široko rozšírené v prokaryotických a eukaryotických organizmoch, a sú cieľmi regulačných proteínov. Tieto štruktúry sa vyskytujú v sekvenciách bohatých na guanín a sú formované v promótorových regiónoch určitých génov alebo ich blízkosti. Existuje niekoľko dôkazov z viac ako troch dekád výskumu potvrdzujúcich funkcie G4 v dôležitých bunkových procesoch ako je transkripcia a replikácia.

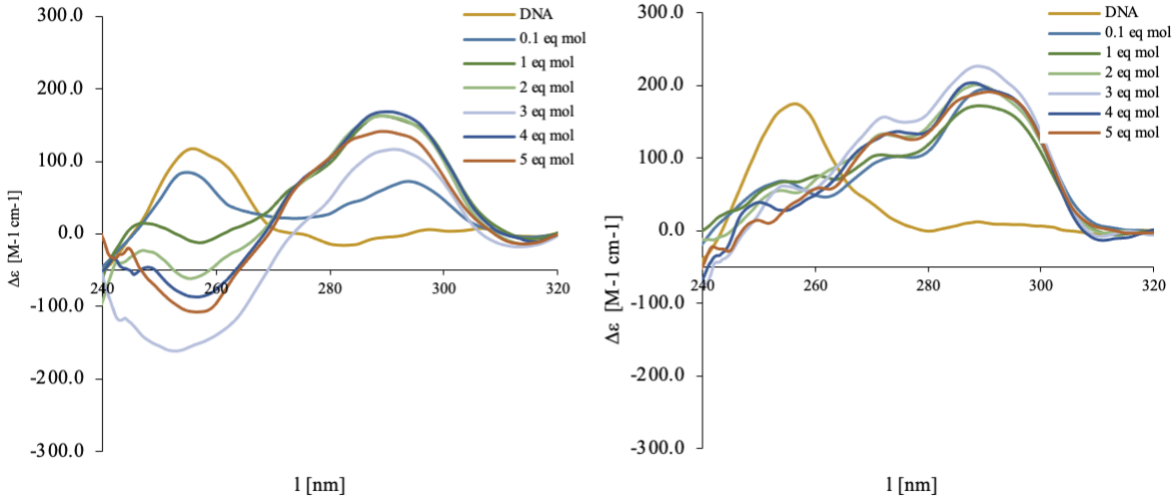
G4 sa stali významnými cieľmi v protinádorových terapií vďaka G4-ligandom, ktoré sa špecificky a s vysokou afinitou viažu a stabilizujú tieto sekundárne DNA štruktúry a tým ovplyvňujú zmeny v genóme, ako je napríklad expresia génov regulujúcich rast buniek. Využitie týchto malých molekúl k stabilizácii sekundárnych štruktúr v promótoroch onkogénov a modulácia ich transkripcie je sľubným prístupom k liečbe niekoľkých druhov rakoviny.

K vyhodnoteniu termodynamickej stability a afinity väzby telomérickej G-kvadruplexovej DNA na syntetizované molekuly ED12 a JG986 sme použili CD spektroskopiu a ThT test. Stabilizačné účinky vybraných látok na DNA boli merané v prítomnosti katiónu draslíka s prídavkom zvyšujúcej sa koncentrácie G4-ligandov pomocou ThT testu (Obrázok 11). Po pridaní ThT k oligonukleotidu bol zaznamenaný pík pri 490 nm odpovedajúci väzbe ThT na G4-motív. Po pridaní G4-ligandov bol pozorovaný veľký pokles fluorescencie a to vytlačením ThT z väzby. Už po pridaní 1 μM oboch G4-ligandov sa fluorescencia komplexu ThT a hTEL znížila, pre prídavok RD12 o viac než polovicu. Prídavok už 3 μM RD12 spôsobilo kompletné vyviazanie ThT z komplexu. K úplnému vyviazaniu ThT z komplexu bolo pottebné pridať niekoľko násobne vyššie množstvo JG986, čo naznačuje silnejšiu väzbovú afinitu pre RD12.



Obrázok 11. Fluorescencia ThT (excitácia pri 440 nm a emisia medzi 470 – 700 nm) pred-inkubovanej hTEL sekvencie v molárnom pomere 0,5:1. Oligonukleotid, hTEL, bol titrovaný RD12 a JG986 v rôznych koncentráciách (0 – 25 µM) a emitovaná fluorescencia bola meraná po 5 minútovej inkubácií pri laboratórnej teplote. Intenzita fluorescencie bola počítaná v percentách, kde 100 % predstavuje komplex oligonukleotidu a ThT.

K štúdiu konformácie G-kvadruplexu po väzbovom účinku G4-ligandov bola prevedená CD spektroskopia (**Error! Reference source not found.**). Rovnako ako v predchádzajúcom experimente, hTEL sekvencia bola podrobená titrácií pomocou vybraných G4-ligandov. hTEL vytvára antiparalelnú konformáciu G4 s negatívnym píkom pri 255 nm a pozitívnym vrcholom pri 290 nm. K meraniu bola sekvencia pripravená v pufre obsahujúcom Tris-HCI (10 mM, pH 7,4) bez pridania KCI. Pridanie JG986 (Obrázok 12A.) preukázateľne spôsobuje tvorbu antiparalelnej G4-štruktúry už po pridaní dvojnásobne vyššej molárnej koncentrácie DNA a ešte vyššie molárne pomery spôsobujú vyššiu stabilizáciu. Pozitívny vrchol sa ukázal pri 290 a negatívny vrchol pri 255 nm. Na druhej strane pridanie 0,1. krát RD12 (Obrázok 12B.) k DNA spôsobilo vytvorenie hybridnej konformácie s pozitívnym vrcholom pri 290 nm a pri 275 nm. Tieto výsledky naznačujú, že vybrané G4-ligandy sú schopné stabilizovať telomerickú DNA bez prítomnosti stabilizačného katiónu.



Obrázok 12. Indukované CD spektrá získané po titrácií G4-ligandmi, JG986 (A) a RD12 (B), k hTEL sekvencii. Ligandy boli pridávané z alikvótov látok rozpustených v DMSO (dĺžka dráhy 1 cm, t = 20 °C). Molárny CD absorpčný koeficient (Δε [$M^{-1}.cm^{-1}$]) bol počítaný ako totálna koncentrácia DNA vo vzorkách.

Rôznymi technickými prístupmi bolo teda možné dokázať, že zvolené G4-ligandy, RD12 a JG986, stabilizujú G4 štruktúry a zároveň sa chovajú rozdielne v rôznych typoch ľudských buniek. Výsledky tejto práce zdôrazňujú pochopenie väzbovej selektivity použitých G4-ligandov na kvadruplexovú DNA a pomáhajú pri budúcom vývoji liečiv pre kvadruplexovo-špecifické ligandy.

Prvoautorka sa podieľala na experimentálnom prevedení analýz, spracovala výsledky a pripravila prvú verziu manuskriptu.

4 ZÁVER

Hlavné výsledky dizertačnej práce je možné zhrnúť nasledovne:

Bola vypracovaná literárna rešerš zameraná na eukaryotické bunkové systémy používané ako modelové organizmy v biotechnológií a molekulárnej biológii, tiež DNA a jej nekanonické sekundárne štruktúry. Na základe rešerše boli vybrané vhodné modelové systémy k skúmaniu spomínaných sekundárnych štruktúr a ich úlohách v základných biologických procesoch.

Praktická časť práce bola spracovaná ako komentovaný súbor publikácií a experimentálnych výstupov. Táto časť bola rozdelená do troch hlavných častí, zaoberajúcich sa štúdiom sekundárnych štruktúr nukleových kyselín *in vivo* a *in vitro*.

Prvá časť experimentálnej práce bola zameraná na štúdium nekanonických štruktúr prevažne v kvasinkových izogénnych systémoch alebo *in vitro* s ohľadom na proteín p53 a obsahuje dva publikačné výstupy. Dôraz bol kladený na vplyv prítomnosti G-kvadruplexových sekvencií v blízkosti responzívneho elementu pre tumor supresorových proteín p53 v kvasinkovom izogénnom systéme. Výsledkom bolo zistenie, že prítomnosť G-kvadruplexovej štruktúry v blízkosti cieľového génu môže určovať účinnosť transkripčnej regulácie proteínu p53, čo zdôrazňuje úlohu štruktúrnych znakov DNA ako modifikátorov funkcií rodiny p53 na cieľových promótorových miestach.

Druhá časť obsahuje súbor publikácií zameraných na štúdium nekanonických DNA štruktúr v rôznych skupinách organizmov pomocou bioinformatických nástrojov vyvinutých Biofyzikálnym ústavom Akadémie vied ČR v spolupráci s informatickým oddelením Mendelovej univerzity, prístupných online [261, 262, 264]. Z publikačných výsledkov je možné zhrnúť, že bol nájdený významný rozdiel medzi obsahom PQS

a IR, a tiež, že ich distribúcia v rôznych skupinách genómov nie je náhodná ale mieste špecifická.

Posledná časť obsahuje doposiaľ nepublikované výsledky, zamerané tiež na nekanonické štruktúry. Dôraz sa kladie na úlohu G4-ligandov, ako stabilizátorov G4, v ľudských nádorových a nenádorových bunkových líniách.

K publikačným výstupom boli počas štúdia získané rozsiahle skúsenosti s rôznymi laboratórnymi technikami a bola vybudovaná medzinárodná spolupráca s Univerzitou prírodných zdrojov a spoločenských vied (BOKU) a Inštitútom experimentálnej a klinickej traumatológie Ludwiga Boltzmanna vo Viedni, v Rakúsku.

5 LITERATÚRA

[1] HÚNGARO, H.M., W.E.L. PEÑA, N.B.M. SILVA, R.V. CARVALHO, V.O. ALVARENGA a A.S. SANT'ANA. Food Microbiology. In: *Encyclopedia of Agriculture and Food Systems* [online]. B.m.: Elsevier, 2014 [vid. 2021-08-30], s. 213–231. ISBN 978-0-08-093139-5. Dostupné z: doi:10.1016/B978-0-444-52512-3.00059-0

[2] PANDEY, G. Model organisms used in molecular biology or medical research. *International Research Journal of Pharmacy*. 2011, **2**, 62–65.

[3] RIVLIN, N., R. BROSH, M. OREN a V. ROTTER. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes & Cancer* [online]. 2011, **2**(4), 466–474. ISSN 1947-6027. Dostupné z: doi:10.1177/1947601911408889

[4] TESFAYE, E., E. MARTINEZ-TERROBA, J. BENDOR, L. WINKLER, Ch. OLIVERO, Kevin CHEN, David M. FELDSER, Jesse R. ZAMUDIO a Nadya DIMITROVA. The p53 transcriptional response across tumor types reveals core and senescence-specific signatures modulated by long noncoding RNAs. *Proceedings of the National Academy of Sciences* [online]. 2021, **118**(31), e2025539118. ISSN 0027-8424, 1091-6490. Dostupné z: doi:10.1073/pnas.2025539118

[5] MILLAU, J.-F., N. BASTIEN a R. DROUIN. P53 transcriptional activities: A general overview and some thoughts. *Mutation Research/Reviews in Mutation Research* [online]. 2009, **681**(2–3), 118–133. ISSN 13835742. Dostupné z: doi:10.1016/j.mrrev.2008.06.002

[6] GÖHLER, T., M. REIMANN, D. CHERNY, K. WALTER, G. WARNECKE, E. KIM a W. DEPPERT. Specific Interaction of p53 with Target Binding Sites Is Determined by DNA Conformation and Is Regulated by the C-terminal Domain. *Journal of Biological Chemistry* [online]. 2002, **277**(43), 41192–41203. ISSN 0021-9258, 1083-351X. Dostupné z: doi:10.1074/jbc.M202344200

[7] VELLAI, T. a G. VIDA. The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proceedings of the Royal Society of London. Series B: Biological Sciences* [online]. 1999, **266**(1428), 1571–1577. ISSN 1471-2954. Dostupné z: doi:10.1098/rspb.1999.0817

[8] CALVO, P. a F. BALUŠKA. Conditions for minimal intelligence across eukaryota: a cognitive science perspective. *Frontiers in Psychology* [online]. 2015, **6** [vid. 2019-04-23]. ISSN 1664-1078. Dostupné z: doi:10.3389/fpsyg.2015.01329

[9] R. BURGESS. *Guide to Protein Purification.* B.m.: Elsevier Science, 2009. ISBN 978-0-08-092317-8.

[10] LACROIX, B. a V. CITOVSKY. Transfer of DNA from Bacteria to Eukaryotes. *mBio* [online]. 2016, **7**(4), e00863-16, /mbio/7/4/e00863-16.atom. ISSN 2150-7511. Dostupné z: doi:10.1128/mBio.00863-16

[11] RUSSELL, J. J., J. A. THERIOT, P. SOOD, W. F. MARSHALL, L. F. LANDWEBER a ET ALL. Non-model model organisms. *BMC Biology* [online]. 2017, **15**(1), 55, s12915-017-0391–5. ISSN 1741-7007. Dostupné z: doi:10.1186/s12915-017-0391-5

[12] TENREIRO, S. a T. F. OUTEIRO. Simple is good: yeast models of neurodegeneration: Yeast as a model for neurodegeneration. *FEMS Yeast Research* [online]. 2010, **10**(8), 970–979. ISSN 15671356. Dostupné z: doi:10.1111/j.1567-1364.2010.00649.x

[13] JOHNSON, E. A. a C. ECHAVARRI-ERASUN. Yeast Biotechnology. In: *The Yeasts* [online]. B.m.: Elsevier, 2011 [vid. 2021-11-06], s. 21–44. ISBN 978-0-444-52149-1. Dostupné z: doi:10.1016/B978-0-444-52149-1.00003-3

[14] FIELDS, S. a M. JOHNSTON. Cell biology. Whither model organism research? *Science (New York, N.Y.)* [online]. 2005, **307**(5717), 1885–1886. ISSN 1095-9203. Dostupné z: doi:10.1126/science.1108872

[15] SUGIYAMA, M., K. YAMAGISHI, Y.-H. KIM, Y. KANEKO, M. NISHIZAWA a S. HARASHIMA. Advances in molecular methods to alter chromosomes and genome in the yeast Saccharomyces cerevisiae. *Applied Microbiology and Biotechnology* [online]. 2009, **84**(6), 1045–1052. ISSN 0175-7598, 1432-0614. Dostupné z: doi:10.1007/s00253-009-2144-z

[16] GOFFEAU, A., B. G. BARRELL, H. BUSSEY, R. W. DAVIS, B. DUJON, H. FELDMANN a ET ALL. Life with 6000 genes. *Science (New York, N.Y.)*. 1996, **274**(5287), 546, 563–567. ISSN 0036-8075.

[17] GIAEVER, G., A. M. CHU, L. NI, C. CONNELLY, L. RILES, S. VÉRONNEAU, S. DOW, A. LUCAU-DANILA a ET ALL. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* [online]. 2002, **418**(6896), 387–391. ISSN 0028-0836, 1476-4687. Dostupné

z: doi:10.1038/nature00935

[18] WINZELER, E. A., D. D. SHOEMAKER, A. ASTROMOFF, H. LIANG, K. ANDERSON, B. ANDRE, R. BANGHAM, R. BENITO a ET ALL. Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis. *Science* [online]. 1999, **285**(5429), 901–906. ISSN 0036-8075, 1095-9203. Dostupné z: doi:10.1126/science.285.5429.901

[19] JONES, G. M., J. STALKER, S. HUMPHRAY, A. WEST, T. COX, J. ROGERS, I. DUNHAM a G. PRELICH. A systematic library for comprehensive overexpression screens in Saccharomyces cerevisiae. *Nature Methods* [online]. 2008, **5**(3), 239–241. ISSN 1548-7091, 1548-7105. Dostupné z: doi:10.1038/nmeth.1181

[20] GHAEMMAGHAMI, S., W.-K. HUH, K. BOWER, R. W. HOWSON, A. BELLE, N. DEPHOURE, E. K. O'SHEA a J. S. WEISSMAN. Global analysis of protein expression in yeast. *Nature* [online]. 2003, **425**(6959), 737–741. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/nature02046

[21] HUH, W.-K., J. V. FALVO, L. C. GERKE, A. S. CARROLL, R. W. HOWSON, J. S. WEISSMAN a E. K. O'SHEA. Global analysis of protein localization in budding yeast. *Nature* [online]. 2003, **425**(6959), 686–691. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/nature02026

[22] LASHKARI, D. A., J. L. DERISI, J. H. MCCUSKER, A. F. NAMATH, C. GENTILE, S. Y. HWANG, P. O. BROWN a R. W. DAVIS. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proceedings of the National Academy of Sciences of the United States of America* [online]. 1997, **94**(24), 13057–13062. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.94.24.13057

[23] DERISI, J. L., V. R. IYER a P. O. BROWN. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* [online]. 1997, **278**(5338), 680–686. ISSN 0036-8075, 1095-9203. Dostupné z: doi:10.1126/science.278.5338.680

[24] CHO, R. J., M. J. CAMPBELL, E. A. WINZELER, L. STEINMETZ, A. CONWAY, L. WODICKA, T. G. WOLFSBERG, A. E. GABRIELIAN, D. LANDSMAN, D. J. LOCKHART a R. W. DAVIS. A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular Cell* [online]. 1998, **2**(1), 65–73. ISSN 1097-2765. Dostupné z: doi:10.1016/s1097-

2765(00)80114-8

[25] ZHU, H., M. BILGIN, R. BANGHAM, D. HALL, A. CASAMAYOR, P. BERTONE, N. LAN, R. JANSEN a ET ALL. Global Analysis of Protein Activities Using Proteome Chips. *Science* [online]. 2001, **293**(5537), 2101–2105. ISSN 0036-8075, 1095-9203. Dostupné z: doi:10.1126/science.1062191

[26] VILLAS-BÔAS, S.G., J.F. MOXLEY, M. ÅKESSON, G. STEPHANOPOULOS a J. NIELSEN. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochemical Journal* [online]. 2005, **388**(2), 669–677. ISSN 0264-6021, 1470-8728. Dostupné z: doi:10.1042/BJ20041162

[27] JEWETT, M.C., G. HOFMANN a J. NIELSEN. Fungal metabolite analysis in genomics and phenomics. *Current Opinion in Biotechnology* [online]. 2006, **17**(2), 191–197. ISSN 09581669. Dostupné z: doi:10.1016/j.copbio.2006.02.001

[28] ITO, T., T. CHIBA, R. OZAWA, M. YOSHIDA, M. HATTORI a Y. SAKAKI. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences* [online]. 2001, **98**(8), 4569–4574. ISSN 0027-8424, 1091-6490. Dostupné z: doi:10.1073/pnas.061034498

[29] UETZ, P., L. GIOT, G. CAGNEY, T. A. MANSFIELD, R. S. JUDSON, J. R. KNIGHT, D. LOCKSHON, V. NARAYAN a ET ALL. A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature* [online]. 2000, **403**(6770), 623–627. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/35001009

[30] LIEB, J.D., X. LIU, D. BOTSTEIN a P.O. BROWN. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nature Genetics* [online]. 2001, **28**(4), 327–334. ISSN 1061-4036, 1546-1718. Dostupné z: doi:10.1038/ng569

[31] IYER, V.R., C.E. HORAK, C.S. SCAFE, D. BOTSTEIN, M. SNYDER a P.O. BROWN. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* [online]. 2001, **409**(6819), 533–538. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/35054095

[32] TONG, A.H.Y., M. EVANGELISTA, A.B. PARSONS, H. XU, G.D. BADER, N. PAGÉ, M. ROBINSON, S. RAGHIBIZADEH a ET ALL. Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. *Science* [online]. 2001, **294**(5550), 2364–2368. ISSN 0036-

8075, 1095-9203. Dostupné z: doi:10.1126/science.1065810

[33] TONG, A.H.Y, G. LESAGE, G.D. BADER, H. DING, H. XU, X. XIN, J. YOUNG, G.F. BERRIZ a ET ALL. Global Mapping of the Yeast Genetic Interaction Network. *Science* [online]. 2004, **303**(5659), 808–813. ISSN 0036-8075, 1095-9203. Dostupné z: doi:10.1126/science.1091317

[34] HARTWELL, L.H. Nobel Lecture. Yeast and cancer. *Bioscience Reports* [online]. 2002, **22**(3–4), 373–394. ISSN 0144-8463. Dostupné z: doi:10.1023/a:1020918107706

[35] PETRANOVIC, D., K. TYO, G.N. VEMURI a J. NIELSEN. Prospects of yeast systems biology for human health: integrating lipid, protein and energy metabolism: Yeast systems biology for human health. *FEMS Yeast Research* [online]. 2010, **10**(8), 1046–1059. ISSN 15671356. Dostupné z: doi:10.1111/j.1567-1364.2010.00689.x

[36] MUNOZ, A.J., K. WANICHTHANARAK, E. MEZA a D. PETRANOVIC. Systems biology of yeast cell death. *FEMS Yeast Research* [online]. 2012, **12**(2), 249–265. ISSN 15671356. Dostupné z: doi:10.1111/j.1567-1364.2011.00781.x

[37] CARMONA-GUTIERREZ, D, C RUCKENSTUHL, M A BAUER, T EISENBERG, S BÜTTNER a F MADEO. Cell death in yeast: growing applications of a dying buddy. *Cell Death & Differentiation* [online]. 2010, **17**(5), 733–734. ISSN 1350-9047, 1476-5403. Dostupné z: doi:10.1038/cdd.2010.10

[38] BRODSKY, J.L. a W.R. SKACH. Protein folding and quality control in the endoplasmic reticulum: Recent lessons from yeast and mammalian cell systems. *Current Opinion in Cell Biology* [online]. 2011, **23**(4), 464–475. ISSN 09550674. Dostupné z: doi:10.1016/j.ceb.2011.05.004

[39] BONIFACINO, J.S. a B.S. GLICK. The mechanisms of vesicle budding and fusion. *Cell* [online]. 2004, **116**(2), 153–166. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(03)01079-1

[40] CHEN, R.E. a J. THORNER. Function and regulation in MAPK signaling pathways: Lessons learned from the yeast Saccharomyces cerevisiae. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* [online]. 2007, **1773**(8), 1311–1340. ISSN 01674889. Dostupné z: doi:10.1016/j.bbamcr.2007.05.003

[41] WIDMANN, C., S. GIBSON, M. B. JARPE a G. L. JOHNSON. Mitogen-activated

protein kinase: conservation of a three-kinase module from yeast to human. *Physiological Reviews* [online]. 1999, **79**(1), 143–180. ISSN 0031-9333. Dostupné z: doi:10.1152/physrev.1999.79.1.143

[42] DE VIRGILIO, C. a R. LOEWITH. The TOR signalling network from yeast to man. *The International Journal of Biochemistry & Cell Biology* [online]. 2006, **38**(9), 1476–1481. ISSN 13572725. Dostupné z: doi:10.1016/j.biocel.2006.02.013

[43] BARBIERI, M., M. BONAFÈ, C. FRANCESCHI a G. PAOLISSO. Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. *American Journal of Physiology-Endocrinology and Metabolism* [online]. 2003, **285**(5), E1064–E1071. ISSN 0193-1849, 1522-1555. Dostupné z: doi:10.1152/ajpendo.00296.2003

[44] SMITH, M.G. a M. SNYDER. Yeast as a Model for Human Disease. *Current Protocols in Human Genetics* [online]. 2006, **48**(1) [vid. 2021-11-06]. ISSN 1934-8266, 1934-8258. Dostupné z: doi:10.1002/0471142905.hg1506s48

[45] PEROCCHI, F., E. MANCERA a L.M. STEINMETZ. Systematic screens for human disease genes, from yeast to human and back. *Mol. BioSyst.* [online]. 2008, **4**(1), 18–29. ISSN 1742-206X, 1742-2051. Dostupné z: doi:10.1039/B709494A

[46] PETRANOVIC, D. a J. NIELSEN. Can yeast systems biology contribute to the understanding of human disease? *Trends in Biotechnology* [online]. 2008, **26**(11), 584–590. ISSN 01677799. Dostupné z: doi:10.1016/j.tibtech.2008.07.008

[47] GUARAGNELLA, Nicoletta, Vanessa PALERMO, Alvaro GALLI, Loredana MORO, Cristina MAZZONI a Sergio GIANNATTASIO. The expanding role of yeast in cancer research and diagnosis: insights into the function of the oncosuppressors p53 and BRCA1/2. *FEMS Yeast Research* [online]. 2014, **14**(1), 2–16. ISSN 15671356. Dostupné z: doi:10.1111/1567-1364.12094

[48] KHURANA, V. a S. LINDQUIST. Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? *Nature Reviews Neuroscience* [online]. 2010, **11**(6), 436–449. ISSN 1471-003X, 1471-0048. Dostupné z: doi:10.1038/nrn2809

[49] PRETORIUS, I. S. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast (Chichester, England)* [online]. 2000, **16**(8), 675–729. ISSN 0749-503X. Dostupné z: doi:10.1002/1097-0061(20000615)16:8<675::AID-

YEA585>3.0.CO;2-B

[50] PÉREZ-COELLO, M.S., A.I. BRIONES PÉREZ, J.F. UBEDA IRANZO a P.J. MARTIN ALVAREZ. Characteristics of wines fermented with different Saccharomyces cerevisiae strains isolated from the La Mancha region. *Food Microbiology* [online]. 1999, **16**(6), 563–573. ISSN 07400020. Dostupné z: doi:10.1006/fmic.1999.0272

[51] ROMANO, P., C. FIORE, M. PARAGGIO, M. CARUSO a A. CAPECE. Function of yeast species and strains in wine flavour. *International Journal of Food Microbiology* [online]. 2003, **86**(1–2), 169–180. ISSN 0168-1605. Dostupné z: doi:10.1016/s0168-1605(03)00290-3

[52] FELDMANN, H., ed. Yeast Metabolism. In: H. FELDMANN, ed. *Yeast* [online]. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2012 [vid. 2021-10-27], s. 25–58. ISBN 978-3-527-65918-0. Dostupné z: doi:10.1002/9783527659180.ch3

[53] BARNETT, J.A. a K.-D. ENTIAN. A history of research on yeasts 9: regulation of sugar metabolism. *Yeast* [online]. 2005, **22**(11), 835–894. ISSN 0749-503X, 1097-0061. Dostupné z: doi:10.1002/yea.1249

[54] ŻYMAŃCZYK-DUDA, E., M. BRZEZIŃSKA-RODAK, M. KLIMEK-OCHAB, M. DUDA a A. ZERKA. Yeast as a Versatile Tool in Biotechnology. In: A. MORATA a I. LOIRA, ed. *Yeast - Industrial Applications* [online]. B.m.: InTech, 2017 [vid. 2021-11-01]. ISBN 978-953-51-3599-9. Dostupné z: doi:10.5772/intechopen.70130

[55] DRDÁK, M., J. STUDNICKÝ, E. MÓROVÁ a J. KAROVIČOVÁ. *Základy potravinárskych technológií: spracovanie rastlinných a živočíšnych surovín ; cereálne a fermentačné technológie ; uchovávanie, hygiena a ekológia potravín*. Bratislava: Malé centrum, 1996. ISBN 978-80-967064-1-9.

[56] DOYLE, M.P., L.P. BEUCHAT a T.J. MONTVILLE, ed. *Food microbiology: fundamentals and frontiers*. 2nd ed. Washington, D.C: ASM Press, 2001. ISBN 978-1-55581-208-9.

[57] REED, G. a T.W. NAGODAWITHANA. *Yeast Technology*. [online]. Dordrecht: Springer Netherlands, 2012 [vid. 2021-11-01]. ISBN 978-94-011-9771-7. Dostupné z: http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=3567189

[58] NIELSEN, J. Production of biopharmaceutical proteins by yeast: Advances through metabolic engineering. *Bioengineered* [online]. 2013, **4**(4), 207–211. ISSN 2165-5979, 2165-

5987. Dostupné z: doi:10.4161/bioe.22856

[59] DEQUIN, S. The potential of genetic engineering for improving brewing, wine-making and baking yeasts. *Applied Microbiology and Biotechnology*. 2001, **56**(5–6), 577–588. ISSN 0175-7598.

[60] PRETORIUS, I.S. a F.F. BAUER. Meeting the consumer challenge through genetically customized wine-yeast strains. *Trends in Biotechnology*. 2002, **20**(10), 426–432. ISSN 0167-7799.

[61] KULP, K. a K.J. LORENZ, ed. *Handbook of dough fermentations*. New York: Marcel Dekker, 2003. Food science and technology, 127. ISBN 978-0-8247-4264-5.

[62] BAMFORTH, C.W. *Food, fermentation, and micro-organisms*. Oxford ; Ames, Iowa: Blackwell Science, 2005. ISBN 978-0-632-05987-4.

[63] HUI, Y. H. a E. Özgül EVRANUZ, ed. *Handbook of plant-based fermented food and beverage technology*. 2nd ed. Boca Raton, FL: CRC Press, 2012. ISBN 978-1-4398-4904-0.

[64] MORÉ, M.I. a A. SWIDSINSKI. Saccharomyces boulardii CNCM I-745 supports regeneration of the intestinal microbiota after diarrheic dysbiosis - a review. *Clinical and Experimental Gastroenterology* [online]. 2015, **8**, 237–255. ISSN 1178-7023. Dostupné z: doi:10.2147/CEG.S85574

[65] CZERUCKA, D., T. PICHE a P. RAMPAL. Review article: yeast as probiotics -- Saccharomyces boulardii. *Alimentary Pharmacology & Therapeutics* [online]. 2007, **26**(6), 767–778. ISSN 0269-2813. Dostupné z: doi:10.1111/j.1365-2036.2007.03442.x

[66] KELESIDIS, T. a C. POTHOULAKIS. Efficacy and safety of the probiotic Saccharomyces boulardii for the prevention and therapy of gastrointestinal disorders. *Therapeutic Advances in Gastroenterology* [online]. 2012, **5**(2), 111–125. ISSN 1756-2848. Dostupné z: doi:10.1177/1756283X11428502

[67] MACCAFERRI, S., A. KLINDER, P. BRIGIDI, P. CAVINA a A. COSTABILE. Potential Probiotic Kluyveromyces marxianus B0399 Modulates the Immune Response in Caco-2 Cells and Peripheral Blood Mononuclear Cells and Impacts the Human Gut Microbiota in an *In Vitro* Colonic Model System. *Applied and Environmental Microbiology* [online]. 2012, **78**(4), 956–964. ISSN 0099-2240, 1098-5336. Dostupné z: doi:10.1128/AEM.06385-11

[68] MORADI, R., R. NOSRATI, H. ZARE, T. TAHMASEBI, H. SADERI a P. OWLIA.

Screening and characterization of in-vitro probiotic criteria of Saccharomyces and Kluyveromyces strains. *Iranian Journal of Microbiology*. 2018, **10**(2), 123–131. ISSN 2008-3289.

[69] QVIRIST, L.A., C. DE FILIPPO, F. STRATI, I. STEFANINI, M. SORDO, T. ANDLID, G.E. FELIS, P. MATTARELLI a D. CAVALIERI. Isolation, Identification and Characterization of Yeasts from Fermented Goat Milk of the Yaghnob Valley in Tajikistan. *Frontiers in Microbiology* [online]. 2016, **7**, 1690. ISSN 1664-302X. Dostupné z: doi:10.3389/fmicb.2016.01690

[70] CHANG, C., J.L. BOWMAN a E.M. MEYEROWITZ. Field Guide to Plant Model Systems. *Cell* [online]. 2016, **167**(2), 325–339. ISSN 00928674. Dostupné z: doi:10.1016/j.cell.2016.08.031

[71] PROVART, N.J., J. ALONSO, S.M. ASSMANN, D. BERGMANN, S.M. BRADY, J. BRKLJACIC, J. BROWSE, C. CHAPPLE a ET ALL. 50 years of Arabidopsis research: highlights and future directions. *New Phytologist* [online]. 2016, **209**(3), 921–944. ISSN 0028-646X, 1469-8137. Dostupné z: doi:10.1111/nph.13687

[72] WOODWARD, A.W. a B. BARTEL. Biology in Bloom: A Primer on the *Arabidopsis thaliana* Model System. *Genetics* [online]. 2018, **208**(4), 1337–1349. ISSN 1943-2631. Dostupné z: doi:10.1534/genetics.118.300755

[73] KOORNNEEF, M. a D. MEINKE. The development of Arabidopsis as a model plant. *The Plant Journal* [online]. 2010, **61**(6), 909–921. ISSN 09607412, 1365313X. Dostupné z: doi:10.1111/j.1365-313X.2009.04086.x

[74] ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* [online]. 2000, **408**(6814), 796–815. ISSN 0028-0836. Dostupné z: doi:10.1038/35048692

[75] KELLOGG, E.A. *Brachypodium distachyon* as a Genetic Model System. *Annual Review of Genetics* [online]. 2015, **49**(1), 1–20. ISSN 0066-4197, 1545-2948. Dostupné z: doi:10.1146/annurev-genet-112414-055135

[76] KOHLER, R.E. *Lords of the fly: Drosophila genetics and the experimental life*. Chicago: University of Chicago Press, 1994. ISBN 978-0-226-45062-9.

[77] MULLER, H. J. The Production of Mutations by X-Rays. *Proceedings of the National*

Academy of Sciences of the United States of America [online]. 1928, **14**(9), 714–726. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.14.9.714

[78] LINDSLEY, D.L., G.G. ZIMM a D.L. LINDSLEY. *The genome of Drosophila melanogaster*. San Diego: Academic Press, 1992. ISBN 978-0-12-450990-0.

[79] NÜSSLEIN-VOLHARD, C. a E. WIESCHAUS. Mutations affecting segment number and polarity in Drosophila. *Nature* [online]. 1980, **287**(5785), 795–801. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/287795a0

[80] RIGGLEMAN, B., E. WIESCHAUS a P. SCHEDL. Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. *Genes & Development* [online]. 1989, **3**(1), 96–113. ISSN 0890-9369. Dostupné z: doi:10.1101/gad.3.1.96

[81] NAGARKAR-JAISWAL, S., S.Z. DELUCA, P.-T. LEE, W.-W LIN, H. PAN, Z. ZUO, J. LV, A.C. SPRADLING a H.J BELLEN. A genetic toolkit for tagging intronic MiMIC containing genes. *eLife* [online]. 2015, **4**. ISSN 2050-084X. Dostupné z: doi:10.7554/eLife.08469

[82] EWEN-CAMPEN, B., D. YANG-ZHOU, V.R. FERNANDES, D.P. GONZÁLEZ, Lu-Ping LIU, R. TAO, X. REN, J. SUN a ET ALL. Optimized strategy for in vivo Cas9-activation in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* [online]. 2017, **114**(35), 9409–9414. ISSN 1091-6490. Dostupné z: doi:10.1073/pnas.1707635114

[83] MENEELY, P.M., C.L. DAHLBERG a J.K. ROSE. Working with Worms: *Caenorhabditis elegans* as a Model Organism. *Current Protocols Essential Laboratory Techniques* [online]. 2019, **19**(1) [vid. 2021-11-01]. ISSN 1948-3430, 1948-3430. Dostupné z: doi:10.1002/cpet.35

[84] ZHANG, S., F. LI, T. ZHOU, G. WANG a Z. LI. Caenorhabditis elegans as a Useful Model for Studying Aging Mutations. *Frontiers in Endocrinology* [online]. 2020, **11**, 554994. ISSN 1664-2392. Dostupné z: doi:10.3389/fendo.2020.554994

[85] LEUNG, M.C.K., P.L. WILLIAMS, A. BENEDETTO, C. AU, K.J. HELMCKE, M. ASCHNER a J.N. MEYER. Caenorhabditis elegans: An Emerging Model in Biomedical and Environmental Toxicology. *Toxicological Sciences* [online]. 2008, **106**(1), 5–28. ISSN 1096-

6080, 1096-0929. Dostupné z: doi:10.1093/toxsci/kfn121

[86] BROWN, S. D. M. Advances in mouse genetics for the study of human disease. *Human Molecular Genetics* [online]. 2021, **30**(20), R274–R284. ISSN 1460-2083. Dostupné z: doi:10.1093/hmg/ddab153

[87] FOX, J.G., ed. *The mouse in biomedical research*. 2nd ed. Amsterdam ; Boston: Elsevier, AP, 2007. American College of Laboratory Animal Medicine series. ISBN 978-0-12-369454-6.

[88] ELSEA, S.H. a R.E. LUCAS. The mousetrap: what we can learn when the mouse model does not mimic the human disease. *ILAR journal* [online]. 2002, **43**(2), 66–79. ISSN 1084-2020. Dostupné z: doi:10.1093/ilar.43.2.66

[89] JEDRZEJCZAK-SILICKA, M. History of Cell Culture. In: S.J.T. GOWDER, ed. *New Insights into Cell Culture Technology* [online]. B.m.: InTech, 2017 [vid. 2021-11-07]. ISBN 978-953-51-3133-5. Dostupné z: doi:10.5772/66905

[90] ULRICH, A.B. a P.M. POUR. Cell Lines. In: *Encyclopedia of Genetics* [online]. B.m.: Elsevier, 2001 [vid. 2021-11-07], s. 310–311. ISBN 978-0-12-227080-2. Dostupné z: doi:10.1006/rwgn.2001.0173

[91] CARTER, M. a J.C. SHIEH. Cell Culture Techniques. In: *Guide to Research Techniques in Neuroscience* [online]. B.m.: Elsevier, 2010 [vid. 2021-11-07], s. 281–296. ISBN 978-0-12-374849-2. Dostupné z: doi:10.1016/B978-0-12-374849-2.00013-6

[92] MAQSOOD, M.I., M.M. MATIN, A.R. BAHRAMI a M.M. GHASROLDASHT. Immortality of cell lines: challenges and advantages of establishment. *Cell Biology International* [online]. 2013, **37**(10), 1038–1045. ISSN 1095-8355. Dostupné z: doi:10.1002/cbin.10137

[93] OUELLETTE, M. M., L. D. MCDANIEL, W. E. WRIGHT, J. W. SHAY a R. A. SCHULTZ. The establishment of telomerase-immortalized cell lines representing human chromosome instability syndromes. *Human Molecular Genetics* [online]. 2000, **9**(3), 403–411. ISSN 0964-6906. Dostupné z: doi:10.1093/hmg/9.3.403

[94] GÓMEZ-LECHÓN, M. J., M. T. DONATO, J. V. CASTELL a R. JOVER. Human hepatocytes as a tool for studying toxicity and drug metabolism. *Current Drug Metabolism* [online]. 2003, **4**(4), 292–312. ISSN 1389-2002. Dostupné z: doi:10.2174/1389200033489424

[95] SCHURR, M.J., K.N. FOSTER, J.M. CENTANNI, A.K. COMER, A. WICKS, A.L. GIBSON, C.L. THOMAS-VIRNIG, Sandy J. SCHLOSSER a ET ALL. Phase I/II clinical evaluation of StrataGraft: a consistent, pathogen-free human skin substitute. *The Journal of Trauma* [online]. 2009, **66**(3), 866–873; discussion 873-874. ISSN 1529-8809. Dostupné z: doi:10.1097/TA.0b013e31819849d6

[96] PASTOR, D.M., L.S. PORITZ, T.L. OLSON, C.L. KLINE, L.R. HARRIS, W.A. KOLTUN, V.M. CHINCHILLI a R.B. IRBY. Primary cell lines: false representation or model system? a comparison of four human colorectal tumors and their coordinately established cell lines. *International Journal of Clinical and Experimental Medicine*. 2010, **3**(1), 69–83. ISSN 1940-5901.

[97] GERAGHTY, R. J., A. CAPES-DAVIS, J. M. DAVIS, J. DOWNWARD, R. I. FRESHNEY, I. KNEZEVIC, R. LOVELL-BADGE, J. R. W. MASTERS, J. MEREDITH, G. N. STACEY, P. THRAVES, M. VIAS, a CANCER RESEARCH UK. Guidelines for the use of cell lines in biomedical research. *British Journal of Cancer* [online]. 2014, **111**(6), 1021–1046. ISSN 1532-1827. Dostupné z: doi:10.1038/bjc.2014.166

[98] GILLET, J.-P., S. VARMA a M.M. GOTTESMAN. The clinical relevance of cancer cell lines. *Journal of the National Cancer Institute* [online]. 2013, **105**(7), 452–458. ISSN 1460-2105. Dostupné z: doi:10.1093/jnci/djt007

[99] CREE, I.A., S. GLAYSHER a A.L. HARVEY. Efficacy of anti-cancer agents in cell lines versus human primary tumour tissue. *Current Opinion in Pharmacology* [online]. 2010, **10**(4), 375–379. ISSN 14714892. Dostupné z: doi:10.1016/j.coph.2010.05.001

[100] TIRIAC, H., P. BELLEAU, D.D. ENGLE, D. PLENKER, A. DESCHÊNES, T.D.D. SOMERVILLE, F.E.M. FROELING, R.A. BURKHART a ET ALL. Organoid Profiling Identifies Common Responders to Chemotherapy in Pancreatic Cancer. *Cancer Discovery* [online]. 2018, **8**(9), 1112–1129. ISSN 2159-8290. Dostupné z: doi:10.1158/2159-8290.CD-18-0349

[101] LABOSKY, P. A., D. P. BARLOW a B. L. HOGAN. Embryonic germ cell lines and their derivation from mouse primordial germ cells. *Ciba Foundation Symposium* [online]. 1994, **182**, 157–168; discussion 168-178. ISSN 0300-5208. Dostupné z: doi:10.1002/9780470514573.ch9

[102] PERA, M. F., B. REUBINOFF a A. TROUNSON. Human embryonic stem cells. *Journal of Cell Science*. 2000, **113 (Pt 1)**, 5–10. ISSN 0021-9533.

[103] HEINS, N., M.C.O. ENGLUND, C. SJÖBLOM, U. DAHL, A. TONNING, C. BERGH, A. LINDAHL, C. HANSON a H. SEMB. Derivation, characterization, and differentiation of human embryonic stem cells. *Stem Cells (Dayton, Ohio)* [online]. 2004, **22**(3), 367–376. ISSN 1066-5099. Dostupné z: doi:10.1634/stemcells.22-3-367

[104] JOHN WILEY & SONS, LTD, ed. *eLS*. 1. vyd. B.m.: Wiley, 2001. ISBN 978-0-470-01617-6.

[105] LANE, D. P. p53, guardian of the genome. *Nature* [online]. 1992, **358**(6381), 15–16. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/358015a0

[106] PALEČEK, E., V. BRÁZDA, E. JAGELSKÁ, P. PEČINKA, L. KARLOVSKÁ a M. BRÁZDOVÁ. Enhancement of p53 sequence-specific binding by DNA supercoiling. *Oncogene* [online]. 2004, **23**(12), 2119–2127. ISSN 0950-9232, 1476-5594. Dostupné z: doi:10.1038/sj.onc.1207324

[107] LEVINE, A.J. p53, the Cellular Gatekeeper for Growth and Division. *Cell* [online]. 1997, **88**(3), 323–331. ISSN 00928674. Dostupné z: doi:10.1016/S0092-8674(00)81871-1

[108] SUZUKI, K., N. DASHZEVEG, T.-G. LU, N. TAIRA, Y. MIKI a K. YOSHIDA. Programmed cell death 6, a novel p53-responsive gene, targets to the nucleus in the apoptotic response to DNA damage. *Cancer Science* [online]. 2012, **103**(10), 1788–1794. ISSN 1349-7006. Dostupné z: doi:10.1111/j.1349-7006.2012.02362.x

[109] PORUBIAKOVÁ, O., N. BOHÁLOVÁ, A. INGA, N. VADOVIČOVÁ, J. COUFAL, M. FOJTA a V. BRÁZDA. The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms. *International Journal of Molecular Sciences* [online]. 2019, **21**(1), E127. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms21010127

[110] OKOROKOV, A.L., M.B. SHERMAN, C. PLISSON, V. GRINKEVICH, K. SIGMUNDSSON, G. SELIVANOVA, J. MILNER a E.V. ORLOVA. The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity. *The EMBO journal* [online]. 2006, **25**(21), 5191–5200. ISSN 0261-4189. Dostupné z: doi:10.1038/sj.emboj.7601382

[111] CHO, Y., S. GORINA, P. D. JEFFREY a N. P. PAVLETICH. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science (New York, N.Y.)* [online]. 1994, **265**(5170), 346–355. ISSN 0036-8075. Dostupné z: doi:10.1126/science.8023157

[112] KERN, S. E., K. W. KINZLER, A. BRUSKIN, D. JAROSZ, P. FRIEDMAN, C. PRIVES a B. VOGELSTEIN. Identification of p53 as a sequence-specific DNA-binding protein. *Science (New York, N.Y.)* [online]. 1991, **252**(5013), 1708–1711. ISSN 0036-8075. Dostupné z: doi:10.1126/science.2047879

[113] EL-DEIRY, W.S., S.E. KERN, J.A. PIETENPOL, K.W. KINZLER a B. VOGELSTEIN. Definition of a consensus binding site for p53. *Nature Genetics* [online]. 1992, **1**(1), 45–49. ISSN 1061-4036, 1546-1718. Dostupné z: doi:10.1038/ng0492-45

[114] TOKINO, T., S. THIAGALINGAM, W. S. EL-DEIRY, T. WALDMAN, K. W. KINZLER a B. VOGELSTEIN. p53 tagged sites from human genomic DNA. *Human Molecular Genetics* [online]. 1994, **3**(9), 1537–1542. ISSN 0964-6906. Dostupné z: doi:10.1093/hmg/3.9.1537

[115] BAKALKIN, G., G. SELIVANOVA, T. YAKOVLEVA, E. KISELEVA, E. KASHUBA, K. P. MAGNUSSON, L. SZEKELY, G. KLEIN, L. TERENIUS a K. G. WIMAN. p53 binds single-stranded DNA ends through the C-terminal domain and internal DNA segments via the middle domain. *Nucleic Acids Research* [online]. 1995, **23**(3), 362–369. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/23.3.362

[116] LEE, S., B. ELENBAAS, A. LEVINE a J. GRIFFITH. p53 and its 14 kDa C-terminal domain recognize primary DNA damage in the form of insertion/deletion mismatches. *Cell* [online]. 1995, **81**(7), 1013–1020. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(05)80006-6

[117] VOGELSTEIN, B., D. LANE a A. J. LEVINE. Surfing the p53 network. *Nature* [online]. 2000, **408**(6810), 307–310. ISSN 0028-0836. Dostupné z: doi:10.1038/35042675

[118] HUPP, T. R. a D. P. LANE. Allosteric activation of latent p53 tetramers. *Current biology: CB* [online]. 1994, **4**(10), 865–875. ISSN 0960-9822. Dostupné z: doi:10.1016/s0960-9822(00)00195-0

[119] JENKINS, L. M. M., S. R. DURELL, S. J. MAZUR a E. APPELLA. p53 N-terminal

phosphorylation: a defining layer of complex regulation. *Carcinogenesis* [online]. 2012, **33**(8), 1441–1449. ISSN 0143-3334, 1460-2180. Dostupné z: doi:10.1093/carcin/bgs145

[120] RAJ, N. a L.D. ATTARDI. The Transactivation Domains of the p53 Protein. *Cold Spring Harbor Perspectives in Medicine* [online]. 2017, **7**(1), a026047. ISSN 2157-1422. Dostupné z: doi:10.1101/cshperspect.a026047

[121] OLINER, J. D., J. A. PIETENPOL, S. THIAGALINGAM, J. GYURIS, K. W. KINZLER a B. VOGELSTEIN. Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature* [online]. 1993, **362**(6423), 857–860. ISSN 0028-0836. Dostupné z: doi:10.1038/362857a0

[122] KUSSIE, P. H., S. GORINA, V. MARECHAL, B. ELENBAAS, J. MOREAU, A. J. LEVINE a N. P. PAVLETICH. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science (New York, N.Y.)* [online]. 1996, **274**(5289), 948–953. ISSN 0036-8075. Dostupné z: doi:10.1126/science.274.5289.948

[123] KUBBUTAT, M. H., S. N. JONES a K. H. VOUSDEN. Regulation of p53 stability by Mdm2. *Nature* [online]. 1997, **387**(6630), 299–303. ISSN 0028-0836. Dostupné z: doi:10.1038/387299a0

[124] YEE, K.S. a K.H. VOUSDEN. Complicating the complexity of p53. *Carcinogenesis* [online]. 2005, **26**(8), 1317–1322. ISSN 0143-3334. Dostupné z: doi:10.1093/carcin/bgi122

[125] SHERR, C. J. Divorcing ARF and p53: an unsettled case. *Nature Reviews. Cancer* [online]. 2006, **6**(9), 663–673. ISSN 1474-175X. Dostupné z: doi:10.1038/nrc1954

[126] IYER, N.G., H. OZDAG a C. CALDAS. p300/CBP and cancer. *Oncogene* [online]. 2004, **23**(24), 4225–4231. ISSN 0950-9232. Dostupné z: doi:10.1038/sj.onc.1207118

[127] ZILFOU, J. T., W. H. HOFFMAN, M. SANK, D. L. GEORGE a M. MURPHY. The corepressor mSin3a interacts with the proline-rich domain of p53 and protects p53 from proteasome-mediated degradation. *Molecular and Cellular Biology* [online]. 2001, **21**(12), 3974–3985. ISSN 0270-7306. Dostupné z: doi:10.1128/MCB.21.12.3974-3985.2001

[128] CHAN, H. M. a N. B. LA THANGUE. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *Journal of Cell Science*. 2001, **114**(Pt 13), 2363–2373. ISSN 0021-9533.

[129] HAUPT, Y., A.I. ROBLES, C. PRIVES a V. ROTTER. Deconstruction of p53 functions and regulation. *Oncogene* [online]. 2002, **21**(54), 8223–8231. ISSN 0950-9232,

1476-5594. Dostupné z: doi:10.1038/sj.onc.1206137

[130] MOULDER, D., D. HATOUM, E. TAY, Y. LIN a E. MCGOWAN. The Roles of p53 in Mitochondrial Dynamics and Cancer Metabolism: The Pendulum between Survival and Death in Breast Cancer? *Cancers* [online]. 2018, **10**(6), 189. ISSN 2072-6694. Dostupné z: doi:10.3390/cancers10060189

[131] HAYFLICK, L. THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. *Experimental Cell Research* [online]. 1965, **37**, 614–636. ISSN 0014-4827. Dostupné z: doi:10.1016/0014-4827(65)90211-9

[132] XUE, W., L. ZENDER, C. MIETHING, R.A. DICKINS, E. HERNANDO, V. KRIZHANOVSKY, C. CORDON-CARDO a S.W. LOWE. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* [online]. 2007, **445**(7128), 656–660. ISSN 1476-4687. Dostupné z: doi:10.1038/nature05529

[133] GORGOULIS, V.G., P. ZACHARATOS, A. KOTSINAS, D. KLETSAS, G. MARIATOS, V. ZOUMPOURLIS, K.M. RYAN, C. KITTAS a A.G. PAPAVASSILIOU. p53 activates ICAM-1 (CD54) expression in an NF-kappaB-independent manner. *The EMBO journal* [online]. 2003, **22**(7), 1567–1578. ISSN 0261-4189. Dostupné z: doi:10.1093/emboj/cdg157

[134] DULIĆ, V., W. K. KAUFMANN, S. J. WILSON, T. D. TLSTY, E. LEES, J. W. HARPER, S. J. ELLEDGE a S. I. REED. p53-dependent inhibition of cyclin-dependent kinase activities in human fibroblasts during radiation-induced G1 arrest. *Cell* [online]. 1994, **76**(6), 1013–1023. ISSN 0092-8674. Dostupné z: doi:10.1016/0092-8674(94)90379-4

[135] KUERBITZ, S. J., B. S. PLUNKETT, W. V. WALSH a M. B. KASTAN. Wild-type p53 is a cell cycle checkpoint determinant following irradiation. *Proceedings of the National Academy of Sciences of the United States of America* [online]. 1992, **89**(16), 7491–7495. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.89.16.7491

[136] LI, X., H. XU, C. XU, M. LIN, X. SONG, F. YI, Y. FENG, K. COUGHLAN a ET ALL. The Yin-Yang of DNA Damage Response: Roles in Tumorigenesis and Cellular Senescence. *International Journal of Molecular Sciences* [online]. 2013, **14**(2), 2431–2448. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms14022431

[137] BROWN, J. P. Bypass of Senescence After Disruption of p21CIP1/WAF1 Gene in

Normal Diploid Human Fibroblasts. *Science* [online]. 1997, **277**(5327), 831–834. ISSN 00368075, 10959203. Dostupné z: doi:10.1126/science.277.5327.831

[138] HERBIG, U., W.A. JOBLING, B.P.C. CHEN, D.J. CHEN a J.M. SEDIVY. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Molecular Cell* [online]. 2004, **14**(4), 501–513. ISSN 1097-2765. Dostupné z: doi:10.1016/s1097-2765(04)00256-4

[139] SERRANO, M., A. W. LIN, M. E. MCCURRACH, D. BEACH a S. W. LOWE. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* [online]. 1997, **88**(5), 593–602. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(00)81902-9

[140] BARTEK, J., J. LUKAS a J. BARTKOVA. DNA damage response as an anti-cancer barrier: damage threshold and the concept of „conditional haploinsufficiency". *Cell Cycle (Georgetown, Tex.)* [online]. 2007, **6**(19), 2344–2347. ISSN 1551-4005. Dostupné z: doi:10.4161/cc.6.19.4754

[141] CAMPISI, J. Cellular senescence as a tumor-suppressor mechanism. *Trends in Cell Biology* [online]. 2001, **11**(11), S27-31. ISSN 0962-8924. Dostupné z: doi:10.1016/s0962-8924(01)02151-1

[142] CAMPISI, J. Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors. *Cell* [online]. 2005, **120**(4), 513–522. ISSN 0092-8674. Dostupné z: doi:10.1016/j.cell.2005.02.003

[143] GIONO, L.E. a J.J. MANFREDI. The p53 tumor suppressor participates in multiple cell cycle checkpoints. *Journal of Cellular Physiology* [online]. 2006, **209**(1), 13–20. ISSN 0021-9541. Dostupné z: doi:10.1002/jcp.20689

[144] ST CLAIR, S. a J.J. MANFREDI. The dual specificity phosphatase Cdc25C is a direct target for transcriptional repression by the tumor suppressor p53. *Cell Cycle (Georgetown, Tex.)* [online]. 2006, **5**(7), 709–713. ISSN 1551-4005. Dostupné z: doi:10.4161/cc.5.7.2628

[145] HERMEKING, H., C. LENGAUER, K. POLYAK, T. C. HE, L. ZHANG, S. THIAGALINGAM, K. W. KINZLER a B. VOGELSTEIN. 14-3-3sigma is a p53-regulated inhibitor of G2/M progression. *Molecular Cell* [online]. 1997, **1**(1), 3–11. ISSN 1097-2765. Dostupné z: doi:10.1016/s1097-2765(00)80002-7

[146] WILLIAMS, K., J. CHRISTENSEN, J. RAPPSILBER, A.L. NIELSEN, J.V. JOHANSEN a K. HELIN. The Histone Lysine Demethylase JMJD3/KDM6B Is Recruited to p53 Bound Promoters and Enhancer Elements in a p53 Dependent Manner. *PLoS ONE* [online]. 2014, **9**(5), e96545. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0096545

[147] EISENBERG-LERNER, A. a A. KIMCHI. The paradox of autophagy and its implication in cancer etiology and therapy. *Apoptosis: An International Journal on Programmed Cell Death* [online]. 2009, **14**(4), 376–391. ISSN 1573-675X. Dostupné z: doi:10.1007/s10495-008-0307-5

[148] LIANG, X. H., S. JACKSON, M. SEAMAN, K. BROWN, B. KEMPKES, H. HIBSHOOSH a B. LEVINE. Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature* [online]. 1999, **402**(6762), 672–676. ISSN 0028-0836. Dostupné z: doi:10.1038/45257

[149] CRIGHTON, D., S. WILKINSON, J. O'PREY, N. SYED, P. SMITH, P.R. HARRISON, M. GASCO, O. GARRONE a ET ALL. DRAM, a p53-induced modulator of autophagy, is critical for apoptosis. *Cell* [online]. 2006, **126**(1), 121–134. ISSN 0092-8674. Dostupné z: doi:10.1016/j.cell.2006.05.034

[150] ATKIN-SMITH, G.K. a I. K. H. POON. Disassembly of the Dying: Mechanisms and Functions. *Trends in Cell Biology* [online]. 2017, **27**(2), 151–162. ISSN 1879-3088. Dostupné z: doi:10.1016/j.tcb.2016.08.011

[151] YONISH-ROUACH, E., D. RESNFTZKY, J. LOTEM, L. SACHS, A. KIMCHI a M. OREN. Wild-type p53 induces apoptosis of myeloid leukaemic cells that is inhibited by interleukin-6. *Nature* [online]. 1991, **352**(6333), 345–347. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/352345a0

[152] ELMORE, S. Apoptosis: a review of programmed cell death. *Toxicologic Pathology* [online]. 2007, **35**(4), 495–516. ISSN 0192-6233. Dostupné z: doi:10.1080/01926230701320337

[153] POVEA-CABELLO, S., M. OROPESA-ÁVILA, P. DE LA CRUZ-OJEDA, M. VILLANUEVA-PAZ, M. DE LA MATA, J.M. SUÁREZ-RIVERO, M. ÁLVAREZ-CÓRDOBA, I. VILLALÓN-GARCÍA a ET ALL. Dynamic Reorganization of the Cytoskeleton during Apoptosis: The Two Coffins Hypothesis. *International Journal of Molecular Sciences*

[online]. 2017, **18**(11), E2393. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms18112393

[154] AKERS, J.C., D. GONDA, R. KIM, B.S. CARTER a C.C. CHEN. Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies. *Journal of Neuro-Oncology* [online]. 2013, **113**(1), 1–11. ISSN 1573-7373. Dostupné z: doi:10.1007/s11060-013-1084-8

[155] STRASSER, A., L. O'CONNOR a V. M. DIXIT. Apoptosis signaling. *Annual Review of Biochemistry* [online]. 2000, **69**, 217–245. ISSN 0066-4154. Dostupné z: doi:10.1146/annurev.biochem.69.1.217

[156] HAKEM, R., A. HAKEM, G. S. DUNCAN, J. T. HENDERSON, M. WOO, M. S. SOENGAS, A. ELIA, J. L. DE LA POMPA, D. KAGI, W. KHOO, J. POTTER, R. YOSHIDA, S. A. KAUFMAN, S. W. LOWE, J. M. PENNINGER a T. W. MAK. Differential requirement for caspase 9 in apoptotic pathways in vivo. *Cell* [online]. 1998, **94**(3), 339–352. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(00)81477-4

[157] KUIDA, K., T. F. HAYDAR, C. Y. KUAN, Y. GU, C. TAYA, H. KARASUYAMA, M. S. SU, P. RAKIC a R. A. FLAVELL. Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* [online]. 1998, **94**(3), 325–337. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(00)81476-2

[158] LI, P., D. NIJHAWAN, I. BUDIHARDJO, S. M. SRINIVASULA, M. AHMAD, E. S. ALNEMRI a X. WANG. Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* [online]. 1997, **91**(4), 479–489. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(00)80434-1

[159] GREEN, D.R. Apoptotic pathways: ten minutes to dead. *Cell* [online]. 2005, **121**(5), 671–674. ISSN 0092-8674. Dostupné z: doi:10.1016/j.cell.2005.05.019

[160] STRASSER, A., P.J JOST a S. NAGATA. The many roles of FAS receptor signaling in the immune system. *Immunity* [online]. 2009, **30**(2), 180–192. ISSN 1097-4180. Dostupné z: doi:10.1016/j.immuni.2009.01.001

[161] VARFOLOMEEV, E. E., M. SCHUCHMANN, V. LURIA, N. CHIANNILKULCHAI, J. S. BECKMANN, I. L. METT, D. REBRIKOV, V. M. BRODIANSKI a ET ALL. Targeted disruption of the mouse Caspase 8 gene ablates cell death induction by the TNF receptors, Fas/Apo1, and DR3 and is lethal prenatally. *Immunity* [online].

1998, **9**(2), 267–276. ISSN 1074-7613. Dostupné z: doi:10.1016/s1074-7613(00)80609-3

[162] SCHUTGENS, R. B. a F. A. BEENER. [Letter: Zonular cataract in an infant]. *Nederlands Tijdschrift Voor Geneeskunde*. 1975, **119**(23), 934–935. ISSN 0028-2162.

[163] LI, H., H. ZHU, C. J. XU a J. YUAN. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* [online]. 1998, **94**(4), 491–501. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(00)81590-1

[164] LUO, X., I. BUDIHARDJO, H. ZOU, C. SLAUGHTER a X. WANG. Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell* [online]. 1998, **94**(4), 481–490. ISSN 0092-8674. Dostupné z: doi:10.1016/s0092-8674(00)81589-5

[165] YIN, X. M., K. WANG, A. GROSS, Y. ZHAO, S. ZINKEL, B. KLOCKE, K. A. ROTH a S. J. KORSMEYER. Bid-deficient mice are resistant to Fas-induced hepatocellular apoptosis. *Nature* [online]. 1999, **400**(6747), 886–891. ISSN 0028-0836. Dostupné z: doi:10.1038/23730

[166] AUBREY, B.J., G.L. KELLY, A. JANIC, M.J. HEROLD a A. STRASSER. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death & Differentiation* [online]. 2018, **25**(1), 104–113. ISSN 1350-9047, 1476-5403. Dostupné z: doi:10.1038/cdd.2017.169

[167] FRIDMAN, Jordan S a S. W. LOWE. Control of apoptosis by p53. *Oncogene* [online]. 2003, **22**(56), 9030–9040. ISSN 0950-9232, 1476-5594. Dostupné z: doi:10.1038/sj.onc.1207116

[168] FEROZ, W. a A.M.A. SHEIKH. Exploring the multiple roles of guardian of the genome: P53. *Egyptian Journal of Medical Human Genetics* [online]. 2020, **21**(1), 49. ISSN 2090-2441. Dostupné z: doi:10.1186/s43042-020-00089-x

[169] NAPOLI, M., X. LI, H.D. ACKERMAN, A.A. DESHPANDE, I BARANNIKOV, M.A. PISEGNA, I. BEDROSIAN, J. MITSCH a ET ALL. Pan-cancer analysis reveals TAp63-regulated oncogenic lncRNAs that promote cancer progression through AKT activation. *Nature Communications* [online]. 2020, **11**(1), 5156. ISSN 2041-1723. Dostupné z: doi:10.1038/s41467-020-18973-w

[170] BASU, S. a M.E. MURPHY. p53 family members regulate cancer stem cells. *Cell*

Cycle (Georgetown, Tex.) [online]. 2016, **15**(11), 1403–1404. ISSN 1551-4005. Dostupné z: doi:10.1080/15384101.2016.1171649

[171] NICOLAI, S., A. ROSSI, N. DI DANIELE, G. MELINO, M. ANNICCHIARICO-PETRUZZELLI a G. RASCHELLÀ. DNA repair and aging: the impact of the p53 family. *Aging* [online]. 2015, **7**(12), 1050–1065. ISSN 1945-4589. Dostupné z: doi:10.18632/aging.100858

[172] MELINO, G. Functional regulation of p73 and p63: development and cancer. *Trends in Biochemical Sciences* [online]. 2003, **28**(12), 663–670. ISSN 09680004. Dostupné z: doi:10.1016/j.tibs.2003.10.004

[173] YANG, A. a F. MCKEON. p63 and p73: p53 mimics, menaces and more. *Nature Reviews Molecular Cell Biology* [online]. 2000, **1**(3), 199–207. ISSN 1471-0072, 1471-0080. Dostupné z: doi:10.1038/35043127

[174] VILGELM, A.E., M.K. WASHINGTON, J. WEI, H. CHEN, V.S. PRASSOLOV a A.I. ZAIKA. Interactions of the p53 Protein Family in Cellular Stress Response in Gastrointestinal Tumors. *Molecular Cancer Therapeutics* [online]. 2010, **9**(3), 693–705. ISSN 1535-7163, 1538-8514. Dostupné z: doi:10.1158/1535-7163.MCT-09-0912

[175] STIEWE, T., C.C. THESELING a B.M. PÜTZER. Transactivation-deficient ΔTA-p73 Inhibits p53 by Direct Competition for DNA Binding. *Journal of Biological Chemistry* [online]. 2002, **277**(16), 14177–14185. ISSN 00219258. Dostupné z: doi:10.1074/jbc.M200480200

[176] MARCEL, V., S. PERRIER, M. AOUBALA, S. AGEORGES, M.J. GROVES, A. DIOT, K. FERNANDES, S. TAURO a J.-C. BOURDON. Δ160p53 is a novel N-terminal p53 isoform encoded by Δ133p53 transcript. *FEBS letters* [online]. 2010, **584**(21), 4463–4468. ISSN 1873-3468. Dostupné z: doi:10.1016/j.febslet.2010.10.005

[177] UEDA, Y., M. HIJIKATA, S. TAKAGI, T. CHIBA a K. SHIMOTOHNO. New p73 variants with altered C-terminal structures have varied transcriptional activities. *Oncogene* [online]. 1999, **18**(35), 4993–4998. ISSN 0950-9232, 1476-5594. Dostupné z: doi:10.1038/sj.onc.1202817

[178] ZAIKA, A., M. IRWIN, C. SANSOME a U. M. MOLL. Oncogenes Induce and Activate Endogenous p73 Protein. *Journal of Biological Chemistry* [online]. 2001, **276**(14),

11310–11316. ISSN 00219258. Dostupné z: doi:10.1074/jbc.M005737200

[179] ZAIKA, A., N. SLADE, S.H. ERSTER, C. SANSOME, T.W. JOSEPH, M. PEARL, E. CHALAS a U.M. MOLL. ΔNp73, A Dominant-Negative Inhibitor of Wild-type p53 and TAp73, Is Up-regulated in Human Tumors. *Journal of Experimental Medicine* [online]. 2002, **196**(6), 765–780. ISSN 1540-9538, 0022-1007. Dostupné z: doi:10.1084/jem.20020179

[180] NAKAGAWA, T., M. TAKAHASHI, T. OZAKI, K.-i. WATANABE, S. TODO, H. MIZUGUCHI, T. HAYAKAWA a A. NAKAGAWARA. Autoinhibitory Regulation of p73 by ΔNp73 To Modulate Cell Survival and Death through a p73-Specific Target Element within the Δ *Np73* Promoter. *Molecular and Cellular Biology* [online]. 2002, **22**(8), 2575–2585. ISSN 0270-7306, 1098-5549. Dostupné z: doi:10.1128/MCB.22.8.2575-2585.2002

[181] MOORE, H C, L B JORDAN, S E BRAY, L BAKER, P R QUINLAN, C A PURDIE, A M THOMPSON, J-C BOURDON a F V FULLER-PACE. The RNA helicase p68 modulates expression and function of the Δ133 isoform(s) of p53, and is inversely associated with Δ133p53 expression in breast cancer. *Oncogene* [online]. 2010, **29**(49), 6475–6484. ISSN 0950-9232, 1476-5594. Dostupné z: doi:10.1038/onc.2010.381

[182] MARCEL, V, V VIJAYAKUMAR, L FERNÁNDEZ-CUESTA, H HAFSI, C SAGNE, A HAUTEFEUILLE, M OLIVIER a P HAINAUT. p53 regulates the transcription of its Δ133p53 isoform through specific response elements contained within the TP53 P2 internal promoter. *Oncogene* [online]. 2010, **29**(18), 2691–2700. ISSN 0950-9232, 1476-5594. Dostupné z: doi:10.1038/onc.2010.26

[183] LANE, D.P., C.F. CHEOK, C. BROWN, A. MADHUMALAR, F.J. GHADESSY a C. VERMA. Mdm2 and p53 are highly conserved from placozoans to man. *Cell Cycle* [online]. 2010, **9**(3), 540–547. ISSN 1538-4101, 1551-4005. Dostupné z: doi:10.4161/cc.9.3.10516

[184] MATLASHEWSKI, G., P. LAMB, D. PIM, J. PEACOCK, L. CRAWFORD a S. BENCHIMOL. Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. *The EMBO Journal* [online]. 1984, **3**(13), 3257–3262. ISSN 02614189. Dostupné z: doi:10.1002/j.1460-2075.1984.tb02287.x

[185] WOLF, D, N HARRIS, N GOLDFINGER a V ROTTER. Isolation of a full-length mouse cDNA clone coding for an immunologically distinct p53 molecule. *Molecular and Cellular Biology* [online]. 1985, **5**(1), 127–132. ISSN 0270-7306, 1098-5549. Dostupné

z: doi:10.1128/mcb.5.1.127-132.1985

[186] BOURDON, J.-C. p53 isoforms can regulate p53 transcriptional activity. *Genes & Development* [online]. 2005, **19**(18), 2122–2137. ISSN 0890-9369. Dostupné z: doi:10.1101/gad.1339905

[187] COURTOIS, S., G. VERHAEGH, S. NORTH, M.-G. LUCIANI, P. LASSUS, U. HIBNER, M. OREN a P. HAINAUT. DeltaN-p53, a natural isoform of p53 lacking the first transactivation domain, counteracts growth suppression by wild-type p53. *Oncogene* [online]. 2002, **21**(44), 6722–6728. ISSN 0950-9232. Dostupné z: doi:10.1038/sj.onc.1205874

[188] JORUIZ, S.M. a J.-C. BOURDON. p53 Isoforms: Key Regulators of the Cell Fate Decision. *Cold Spring Harbor Perspectives in Medicine* [online]. 2016, **6**(8), a026039. ISSN 2157-1422. Dostupné z: doi:10.1101/cshperspect.a026039

[189] YIN, Y., C.W. STEPHEN, M.G. LUCIANI a R. FÅHRAEUS. p53 stability and activity is regulated by Mdm2-mediated induction of alternative p53 translation products. *Nature Cell Biology* [online]. 2002, **4**(6), 462–467. ISSN 1465-7392, 1476-4679. Dostupné z: doi:10.1038/ncb801

[190] GROVER, R., P.S. RAY a S. DAS. Polypyrimidine tract binding protein regulates IRES-mediated translation of p53 isoforms. *Cell Cycle (Georgetown, Tex.)* [online]. 2008, **7**(14), 2189–2198. ISSN 1551-4005. Dostupné z: doi:10.4161/cc.7.14.6271

[191] SHARATHCHANDRA, A., A. KATOCH a S. DAS. IRES mediated translational regulation of p53 isoforms: Regulation of p53 translation. *Wiley Interdisciplinary Reviews: RNA* [online]. 2014, **5**(1), 131–139. ISSN 17577004. Dostupné z: doi:10.1002/wrna.1202

[192] ANBARASAN, T. a J.C. BOURDON. The Emerging Landscape of p53 Isoforms in Physiology, Cancer and Degenerative Diseases. *International Journal of Molecular Sciences* [online]. 2019, **20**(24), 6257. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms20246257

[193] BRADY, C.A., D. JIANG, S.S. MELLO, T.M. JOHNSON, L.A. JARVIS, M.M. KOZAK, D.M. BROZ, S. BASAK a ET ALL. Distinct p53 Transcriptional Programs Dictate Acute DNA-Damage Responses and Tumor Suppression. *Cell* [online]. 2011, **145**(4), 571–583. ISSN 00928674. Dostupné z: doi:10.1016/j.cell.2011.03.035

[194] OHKI, R., T. KAWASE, T. OHTA, H. ICHIKAWA a Y. TAYA. Dissecting functional roles of p53 N-terminal transactivation domains by microarray expression analysis.

Cancer Science [online]. 2007, **98**(2), 189–200. ISSN 1347-9032, 1349-7006. Dostupné z: doi:10.1111/j.1349-7006.2006.00375.x

[195] BOURDON, J.-C. p53 isoforms can regulate p53 transcriptional activity. *Genes & Development* [online]. 2005, **19**(18), 2122–2137. ISSN 0890-9369. Dostupné z: doi:10.1101/gad.1339905

[196] MURRAY-ZMIJEWSKI, F., D. P. LANE a J.-C. BOURDON. p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. *Cell Death and Differentiation* [online]. 2006, **13**(6), 962–972. ISSN 1350-9047. Dostupné z: doi:10.1038/sj.cdd.4401914

[197] KHOURY, M. P. a J.-C. BOURDON. p53 Isoforms: An Intracellular Microprocessor? *Genes & Cancer* [online]. 2011, **2**(4), 453–465. ISSN 1947-6019, 1947-6027. Dostupné z: doi:10.1177/1947601911408893

[198] FUJITA, K., A.M. MONDAL, I. HORIKAWA, G.H. NGUYEN, K. KUMAMOTO, J.J. SOHN, E.D. BOWMAN, E.A. MATHE a ET ALL. p53 isoforms Δ133p53 and p53β are endogenous regulators of replicative cellular senescence. *Nature Cell Biology* [online]. 2009, **11**(9), 1135–1142. ISSN 1465-7392, 1476-4679. Dostupné z: doi:10.1038/ncb1928

[199] WATSON, J. D. a F. H. C. CRICK. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* [online]. 1953, **171**(4356), 737–738. ISSN 0028-0836, 1476-4687. Dostupné z: doi:10.1038/171737a0

[200] FRANK-KAMENETSKII, M. D. a S. M. MIRKIN. Triplex DNA structures. *Annual Review of Biochemistry* [online]. 1995, **64**, 65–95. ISSN 0066-4154. Dostupné z: doi:10.1146/annurev.bi.64.070195.000433

[201] BOCHMAN, M.L., K. PAESCHKE a V.A. ZAKIAN. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* [online]. 2012, **13**(11), 770–780. ISSN 1471-0056, 1471-0064. Dostupné z: doi:10.1038/nrg3296

[202] MENDOZA, O., A. BOURDONCLE, J.-B. BOULÉ, R.M. BROSH a J.-L. MERGNY. G-quadruplexes and helicases. *Nucleic Acids Research* [online]. 2016, **44**(5), 1989–2006. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkw079

[203] ABOU ASSI, H., M. GARAVÍS, C: GONZÁLEZ a M.J. DAMHA. i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Research* [online]. 2018,

46(16), 8038–8056. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gky735

[204] BRÁZDA, V., R.C. LAISTER, E.B. JAGELSKÁ a C. ARROWSMITH. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Molecular Biology* [online]. 2011, **12**(1), 33. ISSN 1471-2199. Dostupné z: doi:10.1186/1471-2199-12-33

[205] PALECEK, E. Premelting changes in DNA conformation. *Progress in Nucleic Acid Research and Molecular Biology*. 1976, **18**, 151–213. ISSN 0079-6603.

[206] GEHRING, K., J. L. LEROY a M. GUÉRON. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature* [online]. 1993, **363**(6429), 561–565. ISSN 0028-0836. Dostupné z: doi:10.1038/363561a0

[207] BRÁZDA, V. a M. FOJTA. The Rich World of p53 DNA Binding Targets: The Role of DNA Structure. *International Journal of Molecular Sciences* [online]. 2019, **20**(22). ISSN 1422-0067. Dostupné z: doi:10.3390/ijms20225605

[208] TATEISHI-KARIMATA, H. a N. SUGIMOTO. Chemical biology of non-canonical structures of nucleic acids for therapeutic applications. *Chemical Communications* [online]. 2020, **56**(16), 2379–2390. ISSN 1359-7345, 1364-548X. Dostupné z: doi:10.1039/C9CC09771F

[209] MIRKIN, S. M. a M. D. FRANK-KAMENETSKII. H-DNA and related structures. *Annual Review of Biophysics and Biomolecular Structure* [online]. 1994, **23**, 541–576. ISSN 1056-8700. Dostupné z: doi:10.1146/annurev.bb.23.060194.002545

[210] FELSENFELD, G. a A. RICH. Studies on the formation of two- and three-stranded polyribonucleotides. *Biochimica Et Biophysica Acta* [online]. 1957, **26**(3), 457–468. ISSN 0006-3002. Dostupné z: doi:10.1016/0006-3002(57)90091-4

[211] BUSKE, F.A., D.C. BAUER, J.S. MATTICK a T.L. BAILEY. Triplex-Inspector: an analysis tool for triplex-mediated targeting of genomic loci. *Bioinformatics* [online]. 2013, **29**(15), 1895–1897. ISSN 1460-2059, 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btt315

[212] VASQUEZ, K.M. a P.M. GLAZER. Triplex-forming oligonucleotides: principles and applications. *Quarterly Reviews of Biophysics* [online]. 2002, **35**(1), 89–107. ISSN 0033-5835. Dostupné z: doi:10.1017/s0033583502003773

[213] BUSKE, F.A., J.S. MATTICK a T.L. BAILEY. Potential in vivo roles of nucleic acid triple-helices. *RNA biology* [online]. 2011, **8**(3), 427–439. ISSN 1555-8584. Dostupné z: doi:10.4161/rna.8.3.14999

[214] BURGE, S., G.N. PARKINSON, P. HAZEL, A.K. TODD a S. NEIDLE. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research* [online]. 2006, **34**(19), 5402–5415. ISSN 0305-1048, 1362-4962. Dostupné z: doi:10.1093/nar/gkl655

[215] MUKHERJEE, A.K., S. SHARMA a S. CHOWDHURY. Non-duplex G-Quadruplex Structures Emerge as Mediators of Epigenetic Modifications. *Trends in genetics: TIG* [online]. 2019, **35**(2), 129–144. ISSN 0168-9525. Dostupné z: doi:10.1016/j.tig.2018.11.001

[216] KWOK, C.K. a C.J. MERRICK. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends in Biotechnology* [online]. 2017, **35**(10), 997–1013. ISSN 1879-3096. Dostupné z: doi:10.1016/j.tibtech.2017.06.012

[217] SPIEGEL, J., S. ADHIKARI a S. BALASUBRAMANIAN. The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry* [online]. 2020, **2**(2), 123–136. ISSN 2589-5974. Dostupné z: doi:10.1016/j.trechm.2019.07.002

[218] RAGUSEO, F., S. CHOWDHURY, A. MINARD a M. DI ANTONIO. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chemical Communications* [online]. 2020, **56**(9), 1317–1324. ISSN 1359-7345, 1364-548X. Dostupné z: doi:10.1039/C9CC09107F

[219] RUGGIERO, E. a S.N. RICHTER. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Research* [online]. 2018, **46**(7), 3270–3283. ISSN 0305-1048, 1362-4962. Dostupné z: doi:10.1093/nar/gky187

[220] CARVALHO, J., J.-L. MERGNY, G.F. SALGADO, J.A. QUEIROZ a C. CRUZ. G-quadruplex, Friend or Foe: The Role of the G-quartet in Anticancer Strategies. *Trends in Molecular Medicine* [online]. 2020, **26**(9), 848–861. ISSN 1471-499X. Dostupné z: doi:10.1016/j.molmed.2020.05.002

[221] KUDLICKI, A.S. G-Quadruplexes Involving Both Strands of Genomic DNA Are Highly Abundant and Colocalize with Functional Sites in the Human Genome. *PloS One* [online]. 2016, **11**(1), e0146174. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0146174

[222] PARROTTA, L., F. ORTUSO, F. MORACA, R. ROCCA, G. COSTA, S. ALCARO a A. ARTESE. Targeting unimolecular G-quadruplex nucleic acids: a new paradigm for the drug discovery? *Expert Opinion on Drug Discovery* [online]. 2014, **9**(10), 1167–1187. ISSN 1746-045X. Dostupné z: doi:10.1517/17460441.2014.941353

[223] ROBINSON, J., F. RAGUSEO, S.P. NUCCIO, D. LIANO a M. DI ANTONIO. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Research* [online]. 2021, **49**(15), 8419–8431. ISSN 0305-1048, 1362-4962. Dostupné z: doi:10.1093/nar/gkab609

[224] CHAMBERS, V.S., G. MARSICO, J.M. BOUTELL, M. DI ANTONIO, G.P. SMITH a S. BALASUBRAMANIAN. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature Biotechnology* [online]. 2015, **33**(8), 877–881. ISSN 1546-1696. Dostupné z: doi:10.1038/nbt.3295

[225] MARSICO, G., V.S. CHAMBERS, A.B. SAHAKYAN, P. MCCAULEY, J.M. BOUTELL, M.D. ANTONIO a S. BALASUBRAMANIAN. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Research* [online]. 2019, **47**(8), 3862–3874. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkz179

[226] BARTAS, M., V. BRÁZDA, N. BOHÁLOVÁ, A. CANTARA, A. VOLNÁ, T. STACHUROVÁ, K. MALACHOVÁ, E.B. JAGELSKÁ a ET ALL. *In-depth Bioinformatic Analyses of Human SARS-CoV-2, SARS-CoV, MERS-CoV, and Other* Nidovirales *Suggest Important Roles of Noncanonical Nucleic Acid Structures in Their Lifecycles* [online]. preprint. B.m.: Bioinformatics. 2020 [vid. 2020-07-14]. Dostupné z: doi:10.1101/2020.04.09.031252

[227] BOHÁLOVÁ, N., A. CANTARA, M. BARTAS, P. KAURA, J. ŠŤASTNÝ, P. PEČINKA, M. FOJTA, J.-L. MERGNY a V. BRÁZDA. Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. *Biochimie* [online]. 2021, **186**, 13–27. ISSN 1638-6183. Dostupné z: doi:10.1016/j.biochi.2021.03.017

[228] BOHÁLOVÁ, N., A: CANTARA, M. BARTAS, P. KAURA, J. ŠŤASTNÝ, P. PEČINKA, M. FOJTA a V. BRÁZDA. Tracing dsDNA Virus-Host Coevolution through Correlation of Their G-Quadruplex-Forming Sequences. *International Journal of Molecular Sciences* [online]. 2021, **22**(7), 3433. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms22073433

[229] BRÁZDA, V., O. PORUBIAKOVÁ, A. CANTARA, N. BOHÁLOVÁ, J. COUFAL,

M. BARTAS, M. FOJTA a J.-L. MERGNY. G-quadruplexes in H1N1 influenza genomes. *BMC Genomics* [online]. 2021, **22**(1), 77. ISSN 1471-2164. Dostupné z: doi:10.1186/s12864-021-07377-9

[230] BRÁZDA, V., Y. LUO, M. BARTAS, P. KAURA, O. PORUBIAKOVÁ, J. ŠŤASTNÝ, P. PEČINKA, D. VERGA a ET ALL. G-Quadruplexes in the Archaea Domain. *Biomolecules* [online]. 2020, **10**(9), 1349. ISSN 2218-273X. Dostupné z: doi:10.3390/biom10091349

[231] BARTAS, M., M. ČUTOVÁ, V. BRÁZDA, P. KAURA, J. ŠŤASTNÝ, J. KOLOMAZNÍK, J. COUFAL, P. GOSWAMI a ET ALL. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* [online]. 2019, **24**(9), 1711. ISSN 1420-3049. Dostupné z: doi:10.3390/molecules24091711

[232] ČUTOVÁ, Michaela, Jacinta MANTA, Otília PORUBIAKOVÁ, Patrik KAURA, Jiří ŠŤASTNÝ, Eva B. JAGELSKÁ, Pratik GOSWAMI, Martin BARTAS a Václav BRÁZDA. Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae. *Genomics* [online]. 2020, **112**(2), 1897–1901. ISSN 08887543. Dostupné z: doi:10.1016/j.ygeno.2019.11.002

[233] TODD, A.K., M. JOHNSTON a S. NEIDLE. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Research* [online]. 2005, **33**(9), 2901–2907. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gki553

[234] HUPPERT, J.L. a S. BALASUBRAMANIAN. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research* [online]. 2007, **35**(2), 406–413. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkl1057

[235] VARSHNEY, D., J. SPIEGEL, K. ZYNER, D. TANNAHILL a S. BALASUBRAMANIAN. The regulation and functions of DNA and RNA G-quadruplexes. *Nature Reviews Molecular Cell Biology* [online]. 2020, **21**(8), 459–474. ISSN 1471-0072, 1471-0080. Dostupné z: doi:10.1038/s41580-020-0236-x

[236] BRÁZDA, V., L. HÁRONÍKOVÁ, J. LIAO a M. FOJTA. DNA and RNA Quadruplex-Binding Proteins. *International Journal of Molecular Sciences* [online]. 2014, **15**(10), 17493–17517. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms151017493

[237] SAUER, M. a K. PAESCHKE. G-quadruplex unwinding helicases and their function

in vivo. *Biochemical Society Transactions* [online]. 2017, **45**(5), 1173–1182. ISSN 1470-8752. Dostupné z: doi:10.1042/BST20170097

[238] SCHIAVONE, D., G. GUILBAUD, P. MURAT, C. PAPADOPOULOU, P. SARKIES, M.-N. PRIOLEAU, S. BALASUBRAMANIAN a J.E. SALE. Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *The EMBO journal* [online]. 2014, **33**(21), 2507–2520. ISSN 1460-2075. Dostupné z: doi:10.15252/embj.201488398

[239] PAESCHKE, K., J.A. CAPRA a V.A. ZAKIAN. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell* [online]. 2011, **145**(5), 678–691. ISSN 1097-4172. Dostupné z: doi:10.1016/j.cell.2011.04.015

[240] DE MAGIS, A., S.G. MANZO, M. RUSSO, J. MARINELLO, R. MORIGI, O. SORDET a G. CAPRANICO. DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* [online]. 2019, **116**(3), 816–825. ISSN 1091-6490. Dostupné z: doi:10.1073/pnas.1810409116

[241] LOPES, J., A. PIAZZA, R. BERMEJO, B. KRIEGSMAN, A. COLOSIO, M.-P. TEULADE-FICHOU, M. FOIANI a A. NICOLAS. G-quadruplex-induced instability during leading-strand replication. *The EMBO journal* [online]. 2011, **30**(19), 4033–4046. ISSN 1460-2075. Dostupné z: doi:10.1038/emboj.2011.316

[242] PUIG LOMBARDI, E., A. HOLMES, D. VERGA, M.-P. TEULADE-FICHOU, A. NICOLAS a A. LONDOÑO-VALLEJO. Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species. *Nucleic Acids Research* [online]. 2019, **47**(12), 6098–6113. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkz463

[243] HAIDER, S.M., S. NEIDLE a G.N. PARKINSON. A structural analysis of G-quadruplex/ligand interactions. *Biochimie* [online]. 2011, **93**(8), 1239–1251. ISSN 1638-6183. Dostupné z: doi:10.1016/j.biochi.2011.05.012

[244] FELSENSTEIN, K.M., L.B. SAUNDERS, J.K. SIMMONS, E. LEON, D.R. CALABRESE, S. ZHANG, A. MICHALOWSKI, P. GAREISS a ET ALL. Small Molecule Microarrays Enable the Identification of a Selective, Quadruplex-Binding Inhibitor of MYC Expression. *ACS Chemical Biology* [online]. 2016, **11**(1), 139–148. ISSN 1554-8929, 1554-8937. Dostupné z: doi:10.1021/acschembio.5b00577

[245] DAY, H.A., P. PAVLOU a Z.A.E. WALLER. i-Motif DNA: Structure, stability and targeting with ligands. *Bioorganic & Medicinal Chemistry* [online]. 2014, **22**(16), 4407–4418. ISSN 09680896. Dostupné z: doi:10.1016/j.bmc.2014.05.047

[246] ALBERTI, P., A. BOURDONCLE, B. SACC?, L. LACROIX a J.-L. MERGNY. DNA nanomachines and nanostructures involving quadruplexes. *Organic & Biomolecular Chemistry* [online]. 2006, **4**(18), 3383. ISSN 1477-0520, 1477-0539. Dostupné z: doi:10.1039/b605739j

[247] ŠKOLÁKOVÁ, P., D. RENČIUK, J. PALACKÝ, D. KRAFČÍK, Z. DVOŘÁKOVÁ, I. KEJNOVSKÁ, K. BEDNÁŘOVÁ a M. VORLÍČKOVÁ. Systematic investigation of sequence requirements for DNA i-motif formation. *Nucleic Acids Research* [online]. 2019, **47**(5), 2177–2189. ISSN 0305-1048, 1362-4962. Dostupné z: doi:10.1093/nar/gkz046

[248] ZHOU, J., C. WEI, G. JIA, X. WANG, Z. FENG a C. LI. Formation of i-motif structure at neutral and slightly alkaline pH. *Mol. BioSyst.* [online]. 2010, **6**(3), 580–586. ISSN 1742-206X, 1742-2051. Dostupné z: doi:10.1039/B919600E

[249] WRIGHT, E.P., J.L. HUPPERT a Z.A.E. WALLER. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Research* [online]. 2017, **45**(6), 2951–2959. ISSN 0305-1048, 1362-4962. Dostupné z: doi:10.1093/nar/gkx090

[250] RAJENDRAN, A., S. NAKANO a N. SUGIMOTO. Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chemical Communications* [online]. 2010, **46**(8), 1299. ISSN 1359-7345, 1364-548X. Dostupné z: doi:10.1039/b922050j

[251] TAKAHASHI, S., J.A. BRAZIER a N. SUGIMOTO. Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. *Proceedings of the National Academy of Sciences* [online]. 2017, **114**(36), 9605–9610. ISSN 0027-8424, 1091-6490. Dostupné z: doi:10.1073/pnas.1704258114

[252] BRAZIER, J.A., A. SHAH a G.D. BROWN. I-Motif formation in gene promoters: unusually stable formation in sequences complementary to known G-quadruplexes. *Chemical Communications* [online]. 2012, **48**(87), 10739. ISSN 1359-7345, 1364-548X. Dostupné z: doi:10.1039/c2cc30863k

[253] BROOKS, T.A., S. KENDRICK a L. HURLEY. Making sense of G-quadruplex and

i-motif functions in oncogene promoters: G-quadruplex and i-motif in oncogene promoters. *FEBS Journal* [online]. 2010, **277**(17), 3459–3469. ISSN 1742464X. Dostupné z: doi:10.1111/j.1742-4658.2010.07759.x

[254] LI, H., J. HAI, J. ZHOU a G. YUAN. The formation and characteristics of the i-motif structure within the promoter of the c-myb proto-oncogene. *Journal of Photochemistry and Photobiology B: Biology* [online]. 2016, **162**, 625–632. ISSN 10111344. Dostupné z: doi:10.1016/j.jphotobiol.2016.07.035

[255] KING, J.J., K.L. IRVING, C.W. EVANS, R.V. CHIKHALE, R. BECKER, C.J. MORRIS, C.D. PEÑA MARTINEZ, P. SCHOFIELD a ET ALL. DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *Journal of the American Chemical Society* [online]. 2020, **142**(49), 20600–20604. ISSN 1520-5126. Dostupné z: doi:10.1021/jacs.0c11708

[256] SVETEC MIKLENIĆ, M. a I.K. SVETEC. Palindromes in DNA—A Risk for Genome Stability and Implications in Cancer. *International Journal of Molecular Sciences* [online]. 2021, **22**(6), 2840. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms22062840

[257] SINDEN, R.R., M.J. PYTLOS-SINDEN a V.N. POTAMAN. Slipped strand DNA structures. *Frontiers in Bioscience: A Journal and Virtual Library* [online]. 2007, **12**, 4788–4799. ISSN 1093-9946. Dostupné z: doi:10.2741/2427

[258] SINDEN, R. R., G. X. ZHENG, R. G. BRANKAMP a K. N. ALLEN. On the deletion of inverted repeated DNA in Escherichia coli: effects of length, thermal stability, and cruciform formation in vivo. *Genetics* [online]. 1991, **129**(4), 991–1005. ISSN 0016-6731. Dostupné z: doi:10.1093/genetics/129.4.991

[259] LISNIĆ, B., I.-K. SVETEC, A. ŠTAFA a Z. ZGAGA. Size-dependent palindrome-induced intrachromosomal recombination in yeast. *DNA Repair* [online]. 2009, **8**(3), 383–389. ISSN 15687864. Dostupné z: doi:10.1016/j.dnarep.2008.11.017

[260] CHALKER, A F, E A OKELY, A DAVISON a D R LEACH. The effects of central asymmetry on the propagation of palindromic DNA in bacteriophage lambda are consistent with cruciform extrusion in vivo. *Genetics* [online]. 1993, **133**(2), 143–148. ISSN 1943-2631. Dostupné z: doi:10.1093/genetics/133.2.143

[261] BRÁZDA, V., J. KOLOMAZNÍK, J. LÝSEK, L. HÁRONÍKOVÁ, J. COUFAL a J.

ŠT'ASTNÝ. Palindrome analyser – A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. *Biochemical and Biophysical Research Communications* [online]. 2016, **478**(4), 1739–1745. ISSN 0006291X. Dostupné z: doi:10.1016/j.bbrc.2016.09.015

[262] BRÁZDA, V., J. KOLOMAZNÍK, J. LÝSEK, M. BARTAS, M. FOJTA, J. ŠŤASTNÝ a J.-L. MERGNY. G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* [online]. 2019, **35**(18), 3493–3495. ISSN 1367-4803, 1460-2059. Dostupné z: doi:10.1093/bioinformatics/btz087

[263] WU, Z., J. LIU, H. YANG, H. LIU a H. XIANG. Multiple replication origins with diverse control mechanisms in Haloarcula hispanica. *Nucleic Acids Research* [online]. 2014, **42**(4), 2282–2294. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gkt1214

[264] BEDRAT, A., L. LACROIX a J.-L. MERGNY. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Research* [online]. 2016, **44**(4), 1746–1759. ISSN 0305-1048, 1362-4962. Dostupné z: doi:10.1093/nar/gkw006

6 Zoznam skratiek a symbolov

AFM	spektroskopia atomárnych síl
C. elegnas	Caenorhabditis elegans
CD	cirkulárny dichroizmus
CDK	cyklín dependentné kinázy
CIP	cyklín dependetný inhibítor kinázy
DBD	DNA väzbová doména
DNA	deoxyribonukleová kyselina
DRAM	gén modulátora autofágie regulovaného poškodením
G4	G-kvadruplex
GAL1	galaktózový promótor
GFP	zelený fluorescenčný proteín
GPD	glyceraldehyd-3-fosfát dehydrogenázový promótor
hTERT	telomerázová reverzná traskriptáza
IR	inverzné repetície
IR/kbp	inverzné repetície na kilobázu
NCBI	Národné centrum pre biotechnologické informácie
MAPK	mitogénom aktvivovaná proteínkináza
MiMIC	Minosom sprostredkovaná integračná kazeta
MPF	faktor podporujúci dozrievanie
mtDNA	mitochondriálna DNA
NLS	signál jadrovej lokalizácie
nt	nukleotid
OD	oligomerizačná doména
PQS	potenciálne G-kvadruplexové sekvencie
PRD	doména bohatá na prolín
RE	responzívny element
RNA	ribonukleová kyselina
S.cerevisiae	Saccharomyces cerevisiae

SASP	sekrečný fenotyp asociovaný so starnutím
TA	transaktivačný variant proteínu
TAD	transaktivačná doména
TOR	cieľ rapamycínu
tmRNA	transférová-mediátorová RNA
tRNA	transférová RNA
wtp53	„wild type“ p53
ΔN	skrátený variant proteínu

7 ŽIVOTOPIS

Meno a priezvisko: Ing. Otília Porubiaková

Dátum narodenia: 11. 02. 1994

Národnosť: Slovenská

Trvalá adresa: Budimír 145, 044 43 Košice, Slovenská republika

Prechodná adresa: Úvoz 525/21, 60200 Brno-střed-Staré Brno, Česká republika

Telefón: +420 776 830 578

e-mail: o.porubiakova@gmail.com/xcporubiakova@vutbr.cz

Vzdelanie:

2018 – súčasnosť: Doktorské štúdium

Vysoké učení technické v Brně, Fakulta chemická, ČR

Odbor: Potravinářská chemie

Názov práce: Eukaryotické buněčné systémy a jejich biotechnologické využití

2016–2018: Magisterské štúdium

Vysoké učení technické v Brně, Fakulta chemická, ČR

Odbor: Potravinářská chemie a biotechnologie

Diplomová práca: Sledování změn obsahu proteinů lepku v průběhu technologie výroby piva

2013–2016: Bakalárske štúdium

Vysoké učení technické v Brně, Fakulta chemická, ČR

Odbor: Biotechnologie

Bakalárska práca: Sledování obsahu gliadínu v pivu

Pracovné skúsenosti:

2018 – súčasnosť: Prax, Akademie věd České republiky, Brno (doktorand)

Jún 2014: Prax, Výskumný ústav pivovarský a sladařský, Brno

Absolvované stáže:

Jún 2021 – Január 2022:

Ludwig Boltzmann Institute for Experimental and Clinical Traumatology (LBI Trauma)/ University of Natural Resources and Life Sciences (BOKU), Viedeň (ERASMUS+)

Jazykové znalosti:

Slovenský	rodný jazyk
Anglický jazyk	B2
Nemecký jazyk	A2
Český jazyk	B2
Francúzsky jazyk	A2

Pedagogická činnosť v priebehu doktorského štúdia:

- konzultant bakalárskych prác:
    - Patrik Bušanski: Srovnání biotechnologických postupů přípravy čistých proteinů (2018/2019)
    - Lucie Šislerová: Aflatoxiny v potravinách a jejich vliv na DNA a buněčné linie (2019/2020)
    - Matúš Vojsovič: Izolace a charakterizace alfa izoforem proteinu p53 (2019/2020)
    - Zuzana Gardošová: Vplyv podmienok kultivácie na produkciu rekombinantných proteínov (2020/2021)
    - Zdeněk Veselý: Vliv makroelementů z potravy na DNA a epigenetický profil (2021/2022)
    - Juan Osorio: Polyfenoly ve výživě a jejich vliv na DNA (2021/2022)
- Konzultant diplomových prác:
    - Bc. Patrik Bušanski: Vplyv prírodných polyfenolických látok na expresiu proteínu p53 (2020/2021)
    - Bc. Matúš Vojsovič: Mutantní protein p53 a jeho vazebné a transaktivační vlastnosti (2021/2022)
    - Bc. Lucia Šislerová: Vliv skladování na mikrobiální složení francouzského sýru Saint-nectaire (2021/2022)
    - Bc. Kristýna Perná: Vliv přírodních a syntetických ligandů na transaktivaci p53 (2021/2022)

- spolupráca na praktickej výuke:
    - PISA – Úloha 4: Elektroforéza na čipu (2017-2021)
    - Praktikum z molekulární biotechnologie – Úloha 3: Izolace DNA a její analýza pomocí RT-PCR (2017-2021)
    - Laboratorní projekt (2017-2022)

8 PREHĽAD PUBLIKAČNEJ ČINNOSTI

8.1 Chronologicky zoradené vedecké publikácie

PORUBIAKOVÁ, Otília, Natália BOHÁLOVÁ, Alberto INGA, Natália VADOVIČOVÁ, Jan COUFAL, Miroslav FOJTA a Václav BRÁZDA. The Influence of Quadruplex Structure in Proximity to P53 Target Sequences on the Transactivation Potential of P53 Alpha Isoforms. *International Journal of Molecular Sciences* [online]. 2019, **21**(1), 127. ISSN 1422-0067. Dostupné z: doi:10.3390/ijms21010127

ČUTOVÁ, Michaela, Jacinta MANTA**, Otília PORUBIAKOVÁ**, Patrik KAURA, Jiří ŠŤASTNÝ, Eva B. JAGELSKÁ, Pratik GOSWAMI, Martin BARTAS a Václav BRÁZDA. Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae. *Genomics* [online]. 2020, **112**(2), 1897–1901. ISSN 08887543. Dostupné z: doi:10.1016/j.ygeno.2019.11.002

BARTAS, Martin, Václav BRÁZDA, Natália BOHÁLOVÁ, Alessio CANTARA, Adriana VOLNÁ, Tereza STACHUROVÁ, Kateřina MALACHOVÁ, Eva B. JAGELSKÁ, **Otília PORUBIAKOVÁ**, Jiří ČERVEŇ a Petr 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. *Frontiers in Microbiology* [online]. 2020, **11**, 1583. ISSN 1664-302X. Dostupné z: doi:10.3389/fmicb.2020.01583

BRÁZDA, Václav, Yu LUO, Martin BARTAS, Patrik KAURA, **Otilia PORUBIAKOVÁ**, Jiří ŠŤASTNÝ, Petr PEČINKA, Daniela VERGA, Violette DA CUNHA, Tomio S. TAKAHASHI, Patrick FORTERRE, Hannu MYLLYKALLIO, Miroslav FOJTA a Jean-Louis MERGNY. G-Quadruplexes in the Archaea Domain. *Biomolecules* [online]. 2020, **10**(9), 1349. ISSN 2218-273X. Dostupné z: doi:10.3390/biom10091349

BRÁZDA, Václav, **Otília PORUBIAKOVÁ**, Alessio CANTARA, Natália BOHÁLOVÁ, Jan COUFAL, Martin BARTAS, Miroslav FOJTA a Jean-Louis MERGNY. G-quadruplexes in H1N1 influenza genomes. *BMC Genomics* [online]. 2021, **22**(1), 77. ISSN 1471-2164. Dostupné z: doi:10.1186/s12864-021-07377-9

MONTI, Paola, Vaclav BRAZDA, Natália BOHÁLOVÁ, **Otília PORUBIAKOVÁ**, Paola MENICHINI, Andrea SPECIALE, Renata BOCCIARDI, Alberto INGA a Gilberto FRONZA. 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* [online]. 2021, **12**(2). ISSN 2073-4425. Dostupné z: doi:10.3390/genes12020277

8.2 Chronologicky zoradené konferenčné príspevky

PORUBIAKOVÁ Otília, Natália BOHÁLOVÁ a Václav BRÁZDA, V. *P53alpha protein isoforms bind preferentially to cruciforms in transactivation Yeast isogenic system.* Prezentované na 46th Annual Conference on Yeast, Smolenice, Slovensko, 7. – 10.5.2019.

BRÁZDA Václav, **Otília PORUBIAKOVÁ,** Natália BOHÁLOVÁ, Alberto INGA a Miroslav FOJTA. *Attenuation of p53-alpha isoform transactivation by inverted repeat sequences in p53 target sites.* Konferenčný príspevok, 8th Mutant Workshop & p53 Isoforms, Lyon, Francúzsko, 15. – 18.5.2019.

PORUBIAKOVÁ Otília, Natália BOHÁLOVÁ A Václav BRÁZDA. *Attenuation of p53-alpha isoform binding to DNA by inverted repeat abolishment in p53 target.* Prezentované na FEBS 2019, Krakow, Poľsko, 6. – 11.7.2019.

PORUBIAKOVÁ Otília, Natália BOHÁLOVÁ a Václav BRÁZDA. *G-quadruplex structure in proximity of p53 target sequence causes significant inhibition on transactivation potential of p53 isoforms.* Konferenčný príspevok prezentovaný na ŠVK - Chémia a technológia pre život, Bratislava, Slovensko, 25.11.2020.

PORUBIAKOVÁ Otília, Natália BOHÁLOVÁ a Václav BRÁZDA. *The influence of non-canonical structure on the p53 isoforms binding to DNA. .* Konferenčný príspevok prezentovaný na Chemie je život, Brno, Česká Republika, 27.11.2020.

PORUBIAKOVÁ Otília, Alessio CANTARA, Veronika Přepechalová a BRÁZDA Václav. *Comparison between synthetic and natural G4-ligands and the effects of their binding to DNA.* Prezentované na FEBS2021, Ľubľana, Slovinsko, 8.7.2021.

8.3 Zoznam ďalších výsledkov pripravovaných k publikácií

GOSWAMI Pratik, **Otília PORUBIAKOVÁ**, Jakub VILÍMEK, Jan Coufal, Miroslav FOJTA a Václav BRÁZDA. *P53-C-terminal isoforms vary in their binding preferences to p53 response element and G-quadruplex structures in DNA.*

PORUBIAKOVÁ Otília, Michal ŠEDÝ, Veronika PŘEPĚCHALOVÁ, Martin BARTAS, Stefan BIDULA, Miroslav FOJTA a Václav BRÁZDA. *Variability of inverted repeats in all accessible genomes of bacteria.*

PORUBIAKOVÁ Otília, Ingo LÄMMERMANN, Alessio CANTARA, Johannes GRILLARI a Václav BRÁZDA. Blocking negative effects of senescence using G4-ligands.

PORUBIAKOVÁ Otília, Alessio CANTARA, Aleš DAŇHEL, Jean-Louis MERGNY a Václav BRÁZDA. *Unlocking of new G4-ligands.*

9 PRÍLOHY

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.

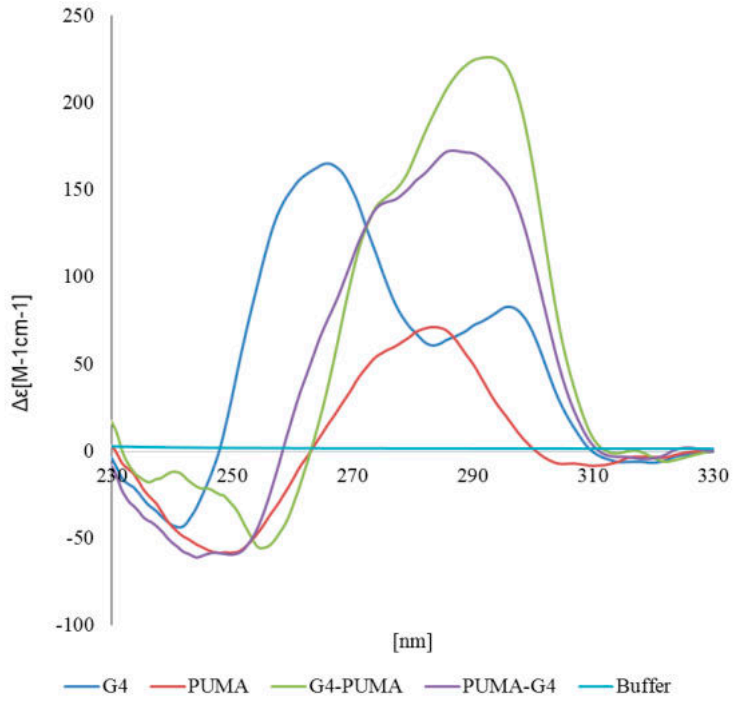


Figure 1. Circular Dichroism }CD] spectroscopy of used DNA sequences. CD spectra of the buffer (light blue), and oligonucleotides from the Table 1 (G4, blue, PUMA-red, G4-PUMA-green, PUMA-G4 violet).

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).

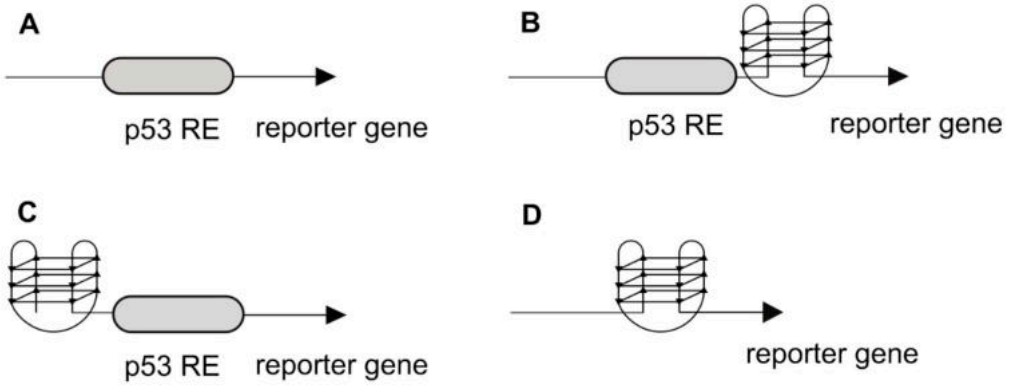


Figure 2. Scheme of the tested sequences in the luciferase reporter promoter region.

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.

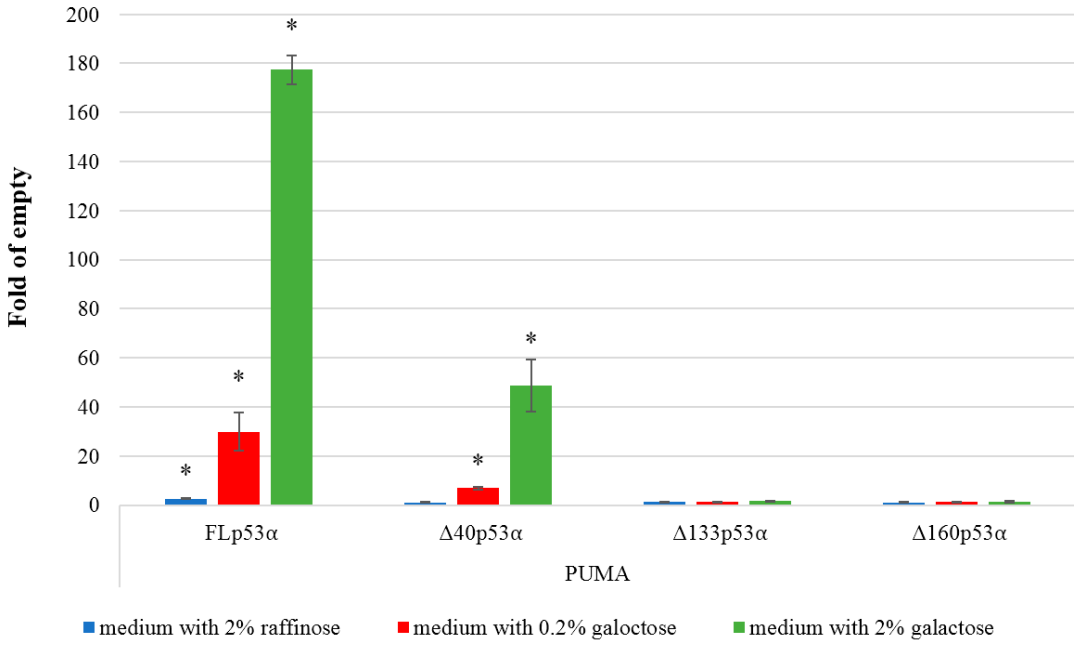


Figure 3. p53-dependent transactivation potential in yeast. All p53α isoforms are expressed under an inducible GAL1 promoter. Histograms show the average fold induction over empty vector in three biological replicates (mean ± S.D.). The results with three levels of p53 induction (no induction, moderate, 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).

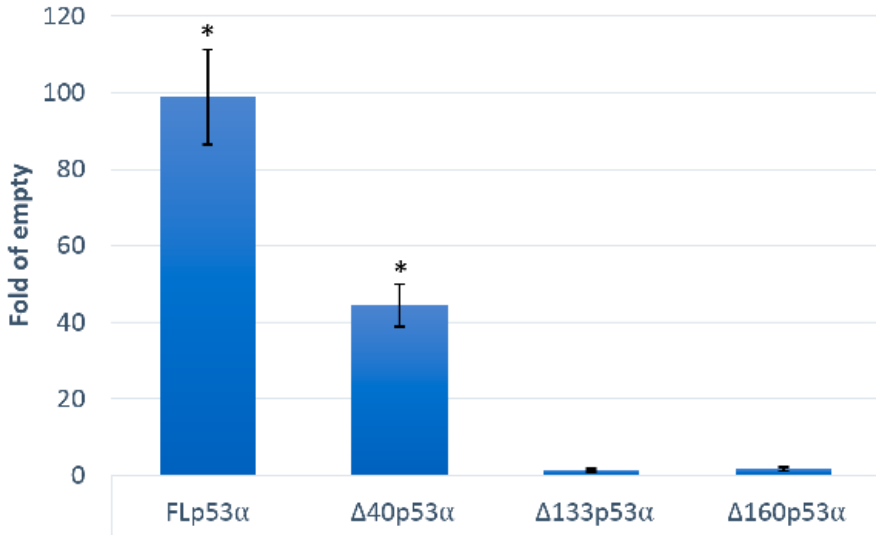


Figure 4. p53-dependent transactivation potential in yeast. All p53α isoforms are expressed under a constitutive glyceraldehyde-3-phosphate dehydrogenase (*GPD*) promoter. The results for the indicated p53α isoforms obtained after 24 h in media without induction are presented. Asterisks indicate a significant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

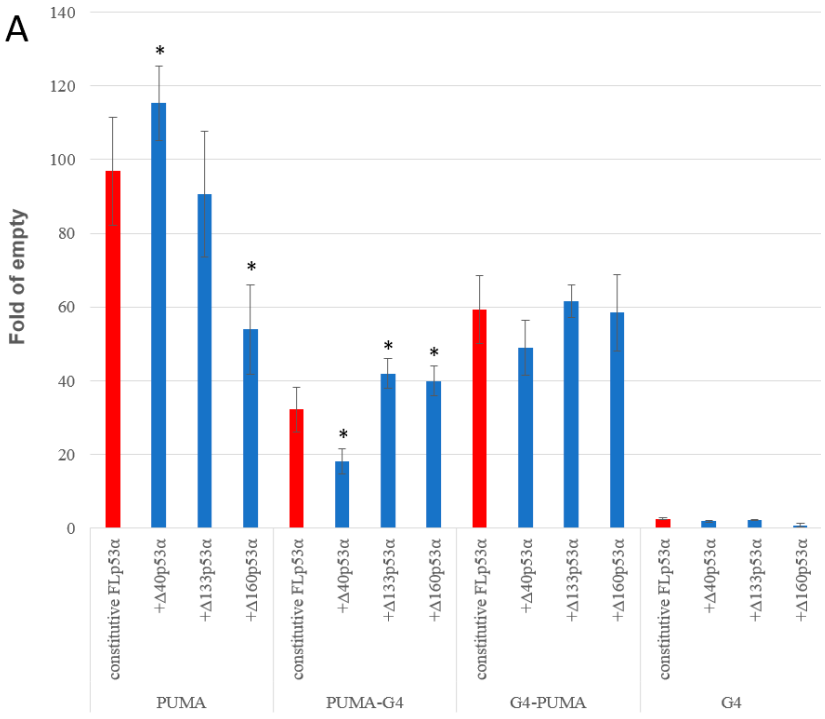


Figure 5. *Cont.*

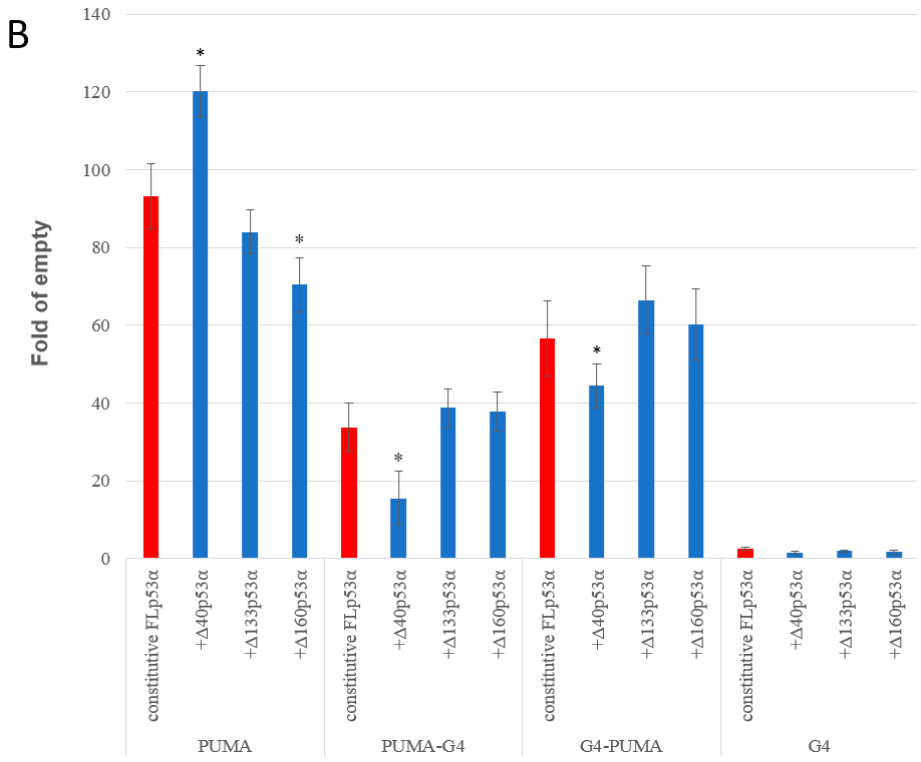


Figure 5. Influence of Δp53α isoforms with inducible expression on transactivation activity of constitutively expressed FLp53α. (**A**) in media with 0.2 % galactose; (**B**) in media with 2% galactose. Four isogenic yeast strains were used, with the p53 target site (PUMA), with the p53 target site after 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 ($p < 0.05$).

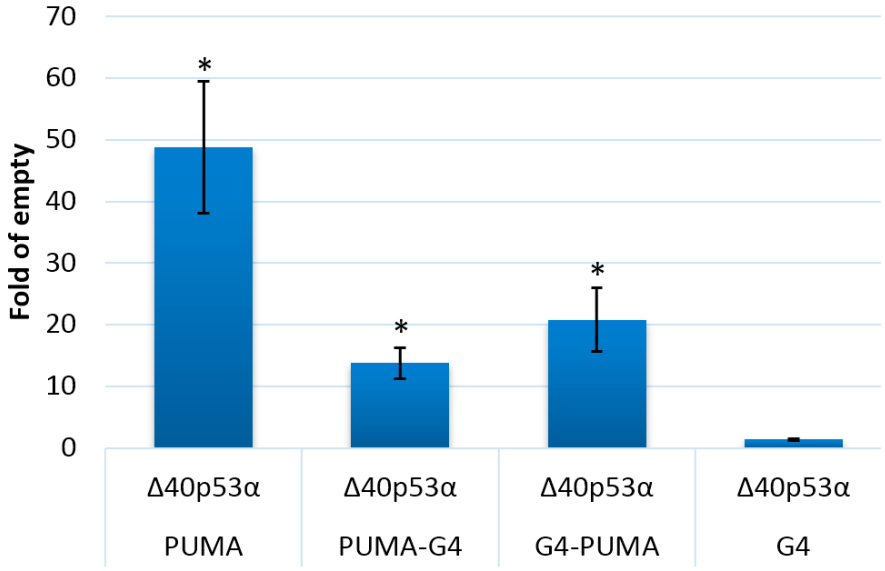


Figure 6. Impact of the G4 prone sequence on Δ40p53α transactivation activity from the PUMA RE. Δ40p53α was expressed under an inducible *GAL1* promoter. Histograms show average fold induction over empty vector in three biological replicates (mean ± S.D.). The results obtained after 24 h in 2% galactose 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).

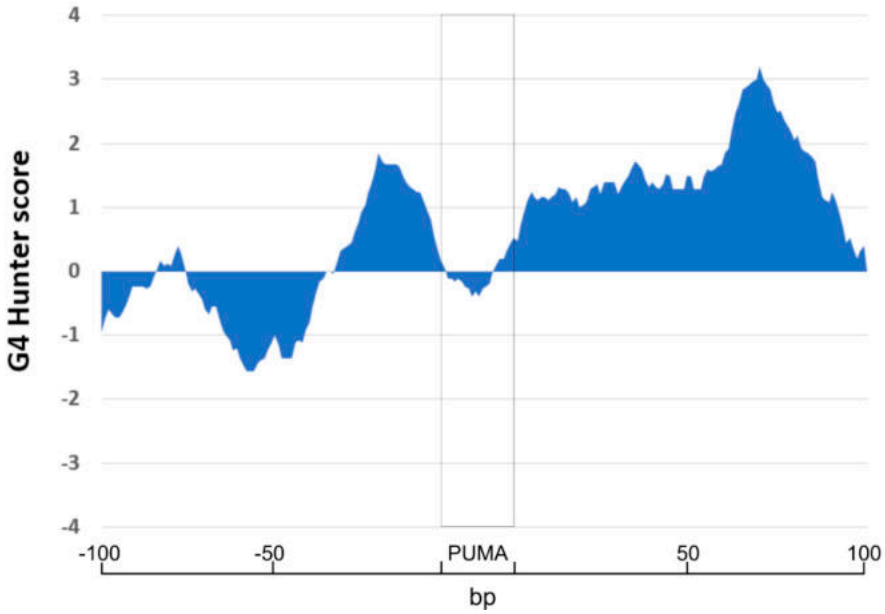


Figure 7. Localization of G4-prone sequences around p53 PUMA RE sequence (100 bp before and after p53 PUMA RE). The analysis of G4-prone sequences has shown that p53 PUMA RE (rectangle) in the human genome is surrounded by G4-prone sequences with peaks at 1.84 before p53 RE and long G4-prone sequence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

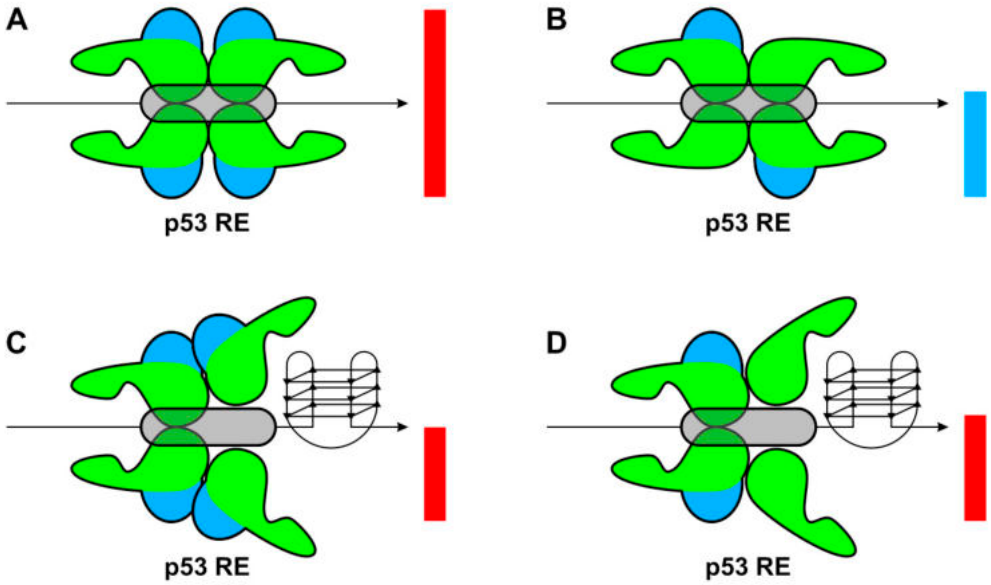


Figure 8. Schematic model of p53 isoforms binding to a RE associated with G4 sequence. (**A**) FLp53α or Δ40p53α bind effectively to the RE and there is a high or moderate level of transactivation. (**B**) Δ133p53α and Δ160p53α inhibit FLp53α transactivation, (**C**) the presence of a G4 close to the 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, $\Delta\varepsilon$ 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]

Genomics 112 (2020) 1897–1901

Contents lists available at ScienceDirect

Genomics

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

Divergent distributions of inverted repeats and G-quadruplex forming sequences in *Saccharomyces cerevisiae*

Michaela Čutová[a], Jacinta Manta[a], Otília Porubiaková[a], Patrik Kaura[b], Jiří Šťastný[b,c], Eva B. Jagelská[d], Pratik Goswami[d], Martin Bartas[e], Václav Brázda[a,d,*]

[a] Brno University of Technology, Faculty of Chemistry, Purkyňova 118, 612 00 Brno, Czech Republic
[b] Brno University of Technology, Faculty of Mechanical Engineering, Technická 2896/2, 616 69 Brno, Czech Republic
[c] Mendel University in Brno, Zemědělská 1665/1, 61300 Brno, Czech Republic
[d] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic
[e] Department of Biology and Ecology/Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Ostrava 710 00, Czech Republic

ARTICLE INFO

Keywords:
Inverted repeat
G-quadruplex
Saccharomyces cerevisiae

ABSTRACT

The importance of DNA structure in the regulation of basic cellular processes is an emerging field of research. Among local non-B DNA structures, inverted repeat (IR) sequences that form cruciforms and G-rich sequences that form G-quadruplexes (G4) are found in all prokaryotic and eukaryotic organisms and are targets for regulatory proteins. We analyzed IRs and G4 sequences in the genome of the most important biotechnology microorganism, *S. cerevisiae*. IR and G4-prone sequences are enriched in specific genomic locations and differ markedly between mitochondrial and nuclear DNA. While G4s are overrepresented in telomeres and regions surrounding tRNAs, IRs are most enriched in centromeres, rDNA, replication origins and surrounding tRNAs. Mitochondrial DNA is enriched in both IR and G4-prone sequences relative to the nuclear genome. This extensive analysis of local DNA structures adds to the emerging picture of their importance in genome maintenance, DNA replication and transcription of subsets of genes.

1. Introduction

DNA primarily exists as a double helix [1]. However, it has been demonstrated that DNA can form local non-B DNA structures including cruciforms [2], triplexes [3] and G-quadruplexes (G4s) [4], which regulate important biological processes [5] including transcriptional activation, telomere maintenance and immune response [6]. Formation of these structures is induced by negative supercoiling of DNA, which induces local nucleotide sequence-dependent conformational changes [7], and/or by protein binding [8]. The genomes and transcriptomes of various organisms, including model organisms *E. coli* [9] *and S. cerevisiae* [10], and of humans cells [11], contain numerous inverted repeats (IRs) and G-rich sequences. Local DNA structures are formed in the presence of sequence motifs, among the most studied structures are cruciforms (Fig. 1A) that can form from an IR, and G4s that form in G-rich sequences where guanines are presented in specific G-tracts (Fig. 1B). The formation of cruciforms is mainly dependent on base sequence and requires perfect or imperfect IRs of 6 or more nucleotides [12,13]. IRs involved in regulatory functions have been evolutionarily conserved, on the other hand, long IRs could be a source of genetic instability [14,15]. Importantly, IRs are often found in close vicinity to polyadenylation (poly(A)) sites [16], replication sites and transcription start sites [2]. G4s are stabilized by potassium or sodium ions and can assume various conformations involving one, two or four DNA molecules [17,18]. G4 sequences are found in telomeres, promoters and other biologically important regions of mammalian genomes [19,20] and are formed by G:G Hoogsteen base pairing [21]. Clusters of G4-forming sequences induce gene expression and are found in regulatory regions of oncogenes [22]. Knowledge of IR and potential quadruplex-forming sequences (PQS) is therefore important for understanding their roles in DNA maintenance and regulation.

Previous studies have shown significant enrichment of IRs in the *S. cerevisiae* genome by IRFinder [23], as well as enrichment of PQS and their functional relevance with transcription [10]. The importance of these local DNA structures is supported also by observations that several *S. cerevisiae* proteins such as Rap1p, Hop1p, Mre11p and Sep1p interact with G4s [24], in agreement with recent studies showing the importance of G4-binding proteins in humans [25–27]. *S. cerevisiae,* as a model eukaryotic organism, offers several genetic systems that facilitate exploration of the in vivo functions of local DNA structures. For

* Corresponding author at: Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic.
E-mail address: vaclav@ibp.cz (V. Brázda).

https://doi.org/10.1016/j.ygeno.2019.11.002
Received 6 August 2019; Received in revised form 13 September 2019; Accepted 1 November 2019
Available online 06 November 2019

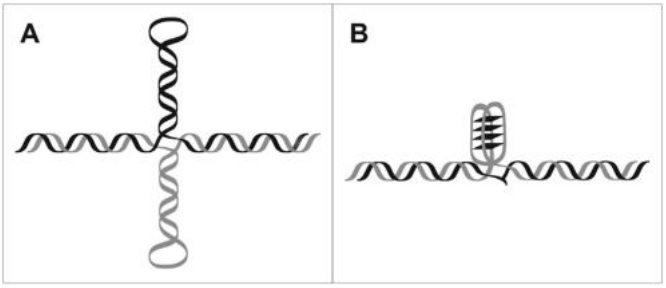


Fig. 1. Schemes of local DNA structures (A) cruciform (B) G-quadruplex.

example, the 3′ ends of most eukaryotic telomeres terminate in a G-rich single stranded overhang with high G4-forming potential [28] and the telomere repeat DNA in *S. cerevisiae* is formed by the consensus sequence 5′-[$(TG)_{0-6}$TGGGTGTG(G)]$_n$-3′ [29], which has been shown to form a G4 structure [30]. In addition to telomeres, numerous sequences have the potential to form intramolecular G4s. For examples, promoter regions [31], ribosomal DNA [32], minisatellites [33] and the immunoglobulin (Ig) heavy chain genes [34]. Knowledge of the presence and localization of local DNA structures in *S. cerevisiae* is therefore important for their effective usage as a tool for functional studies.

Here, we analyzed for the first time differences in IR and PQS localization in the *S. cerevisiae* nuclear and mitochondrial genomes using the new tools Palindrome analyser and G4Hunter web server. Our results show the presence, frequency and localization of these sequence motifs and their dissimilar localizations, suggesting their distinct functional roles in the genome.

2. Results

The fully sequenced genome of *S. cerevisiae* in the NCBI database consists of 16 chromosomes and one mtDNA. *S. cerevisiae* chromosomes vary from 230 kbp for chromosome I to 1.53 Mbp for chromosome IV and mtDNA is 85.78 kbp in length (NCBI IDs are provided in Supplementary material 1).

2.1. Analysis of IRs

We analyzed the *S. cerevisiae* genome for the presence of IRs of length 10–30, spacer size 0–10 bp and 0 or 1 mismatch. In total, we identified 8951 IRs in 12157105 bp (overall IR frequency = 0.74 IR/kbp). The overall number of IRs and the longest IR found in each chromosome and mtDNA of *S. cerevisiae* are shown in Table 1. While the frequency of IRs in chromosomal DNA varied from 0.42 to 0.83 IRs per kbp (mean = 0.56 IR/kbp), the frequency of IRs in mtDNA is 45-times higher at 25.02 IR/kbp. Shapiro-Wilk test of IR frequencies in chromosomal DNAs shows that these data are not normally distributed: W = 0.86912, *p*-value = .02637, and Wilcoxon signed rank test indicates that IR frequencies in chromosomal DNA are significantly different ($p = .0004793$). A graphical representation of IR frequencies is shown in Fig. 2.

We also compared IR frequencies according to length (Table 2). IRs with the shortest length (10 bp) are the most abundant and their numbers and frequencies decrease with increasing IR length. While all chromosomes and mtDNA have at least one IR of length 16 bp, longer IRs were rare and individual IRs with length 19 bp or more represent < 10% of all IRs. Although the IR frequency is higher in mtDNA, longer IRs are present in chromosome XII (24 bp) and chromosome II (29 bp long).

The majority of IRs are without any spacer, followed by IRs with 1 bp between repeats (> 3-fold less frequent for 1 bp spacer). Interestingly, mtDNA has a relatively high frequency of IRs with 1 bp spacer (only 2-fold less frequent than IRs without spacer). The third most often spacer in chromosomal DNA is 3 bp, whereas the third most often spacer is 5 bp in mtDNA (all data for spacers are in Supplementary material 2). The total number of IRs without any mismatch is 876, the remaining 8602 IRs contain one mismatch in the sequence. Numbers and frequencies for all analyzed lengths, spacers, mismatches and their statistical comparisons are detailed in Supplementary material 2.

Table 1
IR characteristics in chromosomes and mtDNA.

DNA sequence	Size (bp)	Number of IRs	Frequency (IR/kbp)	Longest IR (bp)
Chr I	230218	96	0.42	16
Chr II	813184	476	0.59	23
Chr III	316620	262	0.83	29
Chr IV	1531933	866	0.57	23
Chr V	576874	337	0.58	19
Chr VI	270161	172	0.64	16
Chr VII	1090940	662	0.61	21
Chr VIII	562643	321	0.57	19
Chr IX	439888	253	0.58	18
Chr X	745751	392	0.53	20
Chr XI	666816	414	0.62	20
Chr XII	1078177	602	0.56	24
Chr XIII	924431	448	0.48	18
Chr XIV	784333	418	0.53	22
Chr XV	1091291	605	0.55	20
Chr XVI	948066	481	0.51	20
Total nuclear	*12071326*	*6805*	*0.56*	*29*
mtDNA	85779	2146	25.02	23

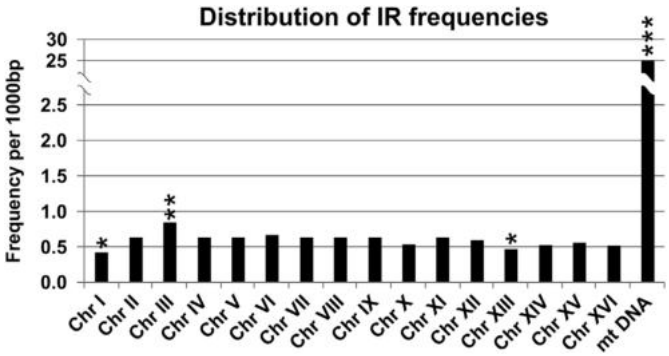


Fig. 2. IR frequencies in chromosomes and mtDNA, Outliers: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 2
Numbers and frequencies of IRs according to size.

IR size	Number in data set	IR/kbp	IR size	Number in data set	IR/kbp
10	4610	0.3792	21	14	0.0012
11	2017	0.1659	22	10	0.0008
13	947	0.0779	23	4	0.0003
14	540	0.0444	24	1	0.0001
15	321	0.0264	25	0	0
16	187	0.0154	26	1	0.0001
17	73	0.0060	27	0	0
18	59	0.0049	28	0	0
19	35	0.0029	29	1	0.0001
20	27	0.0022	30	0	0

2.2. PQS analysis

We analyzed *S. cerevisiae* DNA sequences for PQS by G4Hunter. G4Hunter results are presented as “individual G4Hunter score” (for exact size of the analyses – in our case 25 bp long window) and as “grouped G4Hunter score” (for sequences with G4Hunter above the defined score of 1.2 – which includes all 25 bp long individual sequences grouped to one long sequence). PQS is usually associated with

Table 3
Numbers of PQS, highest G4Hunter score, GC content and %PQS in chromosome/mtDNA. Analysis conditions: Window 25, Threshold 1.2, grouped G4Hunter score.

DNA sequence	Number of PQS	Highest G4Hunter score	%GC content	%PQS
Chr I	87	1.38	39	1.16%
Chr II	248	1.37	38	0.90%
Chr III	138	1.37	39	1.34%
Chr IV	429	1.54	38	0.85%
Chr V	197	1.44	39	1.03%
Chr VI	86	1.38	39	1.03%
Chr VII	334	1.59	38	0.89%
Chr VIII	157	1.37	38	0.88%
Chr IX	162	1.42	39	1.11%
Chr X	192	1.61	38	0.77%
Chr XI	214	1.41	38	0.96%
Chr XII	347	1.47	39	0.97%
Chr XIII	258	1.45	38	0.94%
Chr XIV	256	1.32	39	1.01%
Chr XV	314	1.56	38	0.86%
Chr XVI	306	1.41	38	0.96%
Total nuclear	*3725*	*1.44*	*39*	*0.94%*
mtDNA	43	1.37	17	1.65%

high GC content and *S. cerevisiae* DNA has a low GC content, varying from 37.9% to 39.3% in chromosomes and only 17.1% in mtDNA. Using standard values for G4Hunter - window size 25 and G4Hunter score above 1.2 - we identified a total of 3768 grouped PQS in *S. cerevisiae* DNA. Surprisingly, given that mtDNA has the lowest GC content, mtDNA showed the highest frequency of PQS. The overall number of PQS (score 1.2 or above), the highest G4Hunter scores, GC content and length of all PQS (all base pairs with potential to form G4) divided by total number of bp in the DNA (%PQS) for each chromosome and mtDNA of *S. cerevisiae* are shown in Table 3. Detailed results (Threshold 1.2) are provided in Supplementary Material 3.

A graphical representation of PQS frequencies is shown in Fig. 3. Using Wilcoxon signed rank test, the frequencies of PQS in chromosomal DNA are significantly different (p = .00003052). Considering the very low GC content in mtDNA [35], the highest frequency of PSQ in mtDNA points to specific organization of GC-rich regions in specific G-prone sites. Higher G4Hunter scores indicate a higher probability of a G4 forming inside the PQS [40,41]. We therefore also analyzed the sequences using various values (G4Hunter score from 1.2 to 2) and fixed window size 25, a summary of these data is shown in Table 4.

2.3. IR frequencies according to sequence annotations

The NCBI genome database contains chromosome and mtDNA annotations. The best described are gene (9026186), mRNA (8587452), repeat region (111937), replication origin (107618), tRNA (19630), ncRNA (33448), centromere (3702), mobile_element (291801),

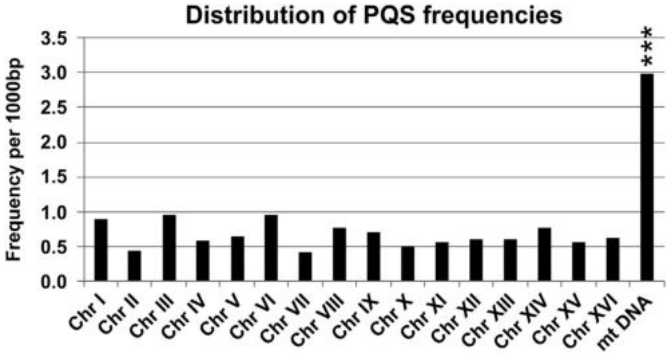


Fig. 3. PQS frequencies in chromosomes and mtDNA, Outliers - see Fig. 2.

Table 4
PQS numbers and frequencies per kbp in *S.cerevisiae* based on G4 hunter score (grouped PQS).

G4Hunter threshold	Number of PQS	PQS frequency (/kbp)
1.2–1.4	2490	0.21
1.4–1.6	865	0.073
1.6–1.8	260	0.022
1.8–2.0	91	0.0077
2.0–more	62	0.0052

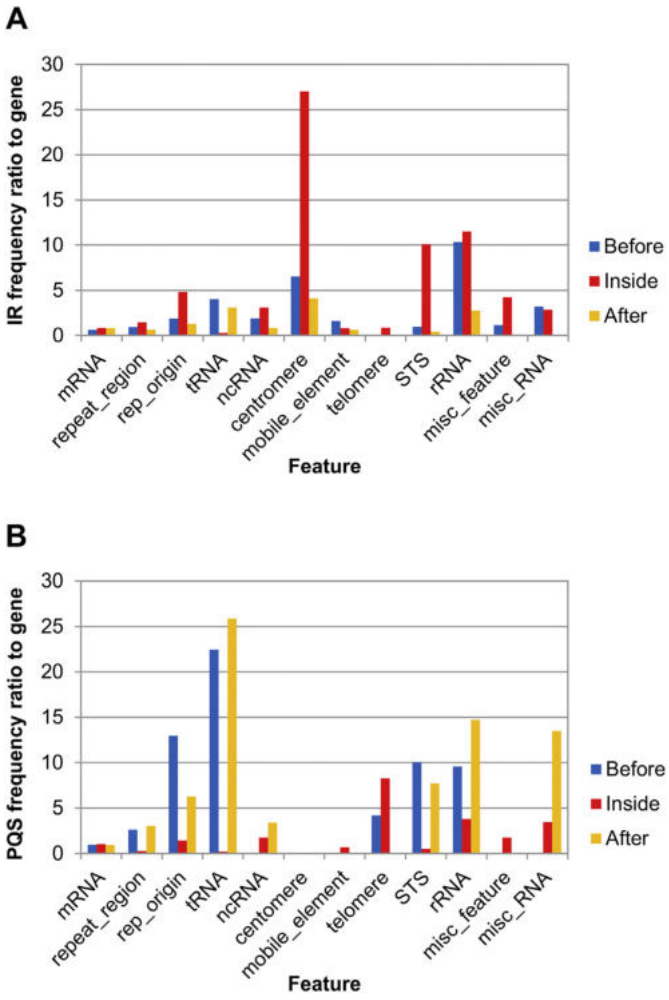


Fig. 4. (A) Differences in IR frequency by DNA locus. The chart shows IR frequencies relative to their frequency in "gene" locations. We analyzed frequencies of IRs with lengths 10 bp and longer within annotated locations (inside) and 100 bp before or after annotated locations. (B) Differences in PQS frequency by DNA locus. The chart shows PQS frequencies relative to the frequency in "gene locations". We analyzed frequencies of all PQS within annotated locations (inside) and 100 bp before or after annotated locations.

telomere (140679), sequence-tagged site (STS) (7297), rRNA (15327), misc_feature (12429) and misc_RNA (16714). To compare IR frequencies at different locations we used the most common annotation "gene" as a standard for comparison with others. The frequency of IRs in centromeric regions is 27-fold higher than IRs located in gene regions. IRs are also abundant in and before rRNA and inside STS locations (> 10-times more abundant than genes), in replication origins, and before rRNA and centromere regions. IRs are relatively less frequent inside tRNA regions and more abundant before and after tRNA regions. The ratios of IR frequencies in annotated features are shown in Fig. 4A. All data of frequencies and ratios for IRs in various regions are

provided in Supplementary material 4.

2.4. PQS frequencies according to sequence annotations

The ratio of PQS frequencies in different annotated features are shown in Fig. 4B. The highest occurrence of PQS within features was observed in telomere sequences (> 8-times more often than in gene regions), followed by rRNA and misc_feature. PQS are abundant also within rRNA (3.8-times higher compared to gene regions) and replication origins (1.4-fold higher). We also found high levels of PQS before and after features, most notably tRNA regions; even though PQS inside the tRNA region is 5-times lower than genes, PQS are > 20-times more abundant 100 bp before and after tRNA regions. Similarly, there is an abundance before and after rRNA regions (9.6-fold higher before and 14.7-fold higher after this feature), and before and after STS, replication origins and repeat regions (Fig. 4B, Supplementary material 4).

3. Discussion

The importance of local DNA structures for DNA maintenance and regulation has emerged in the last decade. It has been demonstrated that IRs able to form cruciforms play important roles in transcription regulation by interaction with many proteins [2] including, for example, the p53 human tumor suppressor protein [36]. Contemporary bioinformatics tools allow analyses of complete genomes, which brings a more complete view of DNA structure and regulation [37,38]. Analyses of mitochondrial and plastid DNAs of all sequenced organisms has shown the presence of IRs in replication origins and other regulatory regions [39,40]. A previous study using a different algorithm (Inverted Repeat Finder) and with different parameters, demonstrated that *S. cerevisiae* chromosomal IRs are clustered in intergenic regions [23] and a correlation between IRs and the polyadenylation signal was shown recently [16]. Here, using Palindrome analyser applied to nuclear and mtDNA, we show that short IRs are more abundant in *S. cerevisiae* mtDNA than chromosomal DNA and that the majority of IRs are located in centromeric regions of chromosomal DNA. The relative abundance in replication origins suggest that IRs play an important role not only in viral and bacterial replication [41, 42], but also in controlling eukaryotic replication.

In addition to IRs, we also investigated PQS. Several proteins have been described to interact with G4s in *S. cerevisiae* [24], including helicases that play important roles in genome maintenance during replication, for example Sgs1p [43,44]. Other G4-binding proteins are proposed to have roles in telomere capping, such as Cdc13p [45], or telomere length regulation and transcriptional silencing/activation, such as Rap1p [46]. Recently, 29 novel G4 binding proteins from *S. cerevisiae* were identified by affinity purification and quantitative LC-MS/MS [47] and a novel G4 binding protein in *S. cerevisiae*, Slx9 [48], was demonstrated to bind in vitro to G4 and ChIP-seq analyses demonstrated that its binding increased in the absence of G4 helicase [48]. Moreover, several molecular and genetic studies show that G4s can act as a regulatory tool [49]. The most recent papers on the functional roles of cruciform-forming IRs and G4 binding proteins in *S. cerevisiae* provide evidence that these structures play regulatory roles [16,47,48]. While direct influences of G4 structures on gene transcription have been shown in bacteria [50] and vertebrates [51], the roles of G4s in *S. cerevisiae* have yet to be fully evaluated. Here, we demonstrated that PQS are highly abundant in telomere regions and, moreover, in close proximity to several important regulatory regions of the *S. cerevisiae* genome. The presence of PQS in these regions suggests their function in *S. cerevisiae* genome regulation. Interestingly, we show that both IRs and PQS are not randomly located, but are enriched in specific chromosomes and specific locations of the *S. cerevisiae* nuclear and mitochondrial genomes. These results point to distinct functional usages of IRs and G4s. The differences in frequencies of local DNA structures in mitochondrial and chromosomal DNA suggest that regulation of DNA expression is different in circular and linear DNAs. With this knowledge, the future perspective will be to manipulate these sequences, either genetically or pharmacologically, and investigate the functional effects in *S. cerevisiae* on e.g. DNA replication, genome stability, tRNA production/activity and protein synthesis, either as a model system or indeed for biotechnology purposes.

4. Methods

4.1. Genome sequences and bioinformatic analyses

The complete DNA sequence of the *S. cerevisiae S288C* nuclear and mitochondrial genomes were downloaded (May 20, 2019) in FASTA format from the National Center for Biotechnology Information (NCBI) [52]. The genomes were analyzed by Palindrome analyser [38] and G4Hunter web [53] for the presence and localization of IRs and PQS, respectively. For IR analyses, the parameters were; size of IR 10 to 30 bp; spacer size 0 to 10 bp; one mismatch was allowed. We produced a separate list of IRs found in each chromosome and mtDNA sequences. PQS analyses used parameters of “25” for window and G4Hunter score above 1.2.

4.2. Analysis of IRs and PQS around annotated NCBI features

We also downloaded tables containing feature annotations from NCBI (e.g. transcription start sites, polyA sites, centromere, telomere, etc). These data were used to investigate the occurrence of IRs of various lengths and of PQS before and after (± 100 bp) and inside features. Further processing was performed in Microsoft Excel.

4.3. Statistical evaluation

Statistical evaluations of normality were made by Shapiro-Wilk test. Because data do not have normal distributions, we additionally analyzed our data by Wilcoxon signed rank test (with *p*-value cut-off 0.05). For statistical evaluation of outliers, chisq.out.test (from R package “outliers”) was used.

Acknowledgments

We thank Dr. Philip Coates for proofreading and editing the manuscript. This work was supported by The Czech Science Foundation (18-15548S).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2019.11.002.

References

[1] J.D. Watson, F.H.C. Crick, A structure for deoxyribose nucleic acid, Nature 171 (1953) 737–738.
[2] V. Brázda, R.C. Laister, E.B. Jagelská, C. Arrowsmith, Cruciform structures are a common DNA feature important for regulating biological processes, BMC Mol. Biol. 12 (2011) 33.
[3] L.D. Nelson, C. Bender, H. Mannsperger, D. Buergy, P. Kambakamba, G. Mudduluru, U. Korf, D. Hughes, M.W. Van Dyke, H. Allgayer, Triplex DNA-binding proteins are associated with clinical outcomes revealed by proteomic measurements in patients with colorectal cancer, Mol. Cancer 11 (2012) 38.
[4] Z.-Y. Sun, X.-N. Wang, S.-Q. Cheng, X.-X. Su, T.-M. Ou, Developing novel G-Quadruplex ligands: from interaction with nucleic acids to interfering with nucleic acid–protein interaction, Molecules 24 (2019).
[5] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, 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 (2002) 11593–11598.
[6] J. Zhao, A. Bacolla, G. Wang, K.M. Vasquez, Non-B DNA structure-induced genetic instability and evolution, Cell. Mol. Life Sci. 67 (2010) 43–62.
[7] E. Palecek, Local supercoil-stabilized DNA structures, Crit. Rev. Biochem. Mol. Biol. 26 (1991) 151–226.

[8] V. Brázda, J. Coufal, Recognition of Local DNA Structures by p53 Protein, Int. J. Mol. Sci. 18 (2017) pii: E375.

[9] P. Rawal, V.B.R. Kummarasetti, J. Ravindran, N. Kumar, K. Halder, R. Sharma, M. Mukerji, S.K. Das, S. Chowdhury, Genome-wide prediction of G4 DNA as regulatory motifs: Role in *Escherichia coli* global regulation, Genome Res. 16 (2006) 644–655.

[10] S.G. Hershman, Q. Chen, J.Y. Lee, M.L. Kozak, P. Yue, L.-S. Wang, F.B. Johnson, Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae, Nucleic Acids Res. 36 (2008) 144–156.

[11] J.L. Huppert, S. Balasubramanian, Prevalence of quadruplexes in the human genome, Nucleic Acids Res. 33 (2005) 2908–2916.

[12] A.L. Mikheikin, A.Y. Lushnikov, Y.L. Lyubchenko, The effect of DNA supercoiling on geometry of holliday junctions, Biochemistry 45 (2006) 12998–13006.

[13] O.I. Limanskaia, A.P. Limanskiĭ, Distribution of potentially hairpin-loop structures in the genome of bovine retroviruses, Vopr. Virusol. 54 (2009) 27–32.

[14] E.A. Wojcik, A. Brzostek, A. Bacolla, P. Mackiewicz, K.M. Vasquez, M. Korycka-Machala, A. Jaworski, J. Dziadek, Direct and inverted repeats elicit genetic instability by both exploiting and eluding DNA double-strand break repair systems in mycobacteria, PLoS One 7 (2012) e51064.

[15] S. Lu, G. Wang, A. Bacolla, J. Zhao, S. Spitser, K.M. Vasquez, Short inverted repeats are hot spots for genetic instability: relevance to cancer genomes, Cell Rep. 10 (2015) 1674–1680.

[16] O. Miura, T. Ogake, H. Yoneyama, Y. Kikuchi, T. Ohyama, A strong structural correlation between short inverted repeat sequences and the polyadenylation signal in yeast and nucleosome exclusion by these inverted repeats, Curr. Genet. 65 (2019) 575–590.

[17] M. Vorlíčková, I. Kejnovská, J. Sagi, D. Renčiuk, K. Bednářová, J. Motlová, J. Kypr, Circular dichroism and guanine quadruplexes, Methods 57 (2012) 64–75.

[18] V. Tokan, J. Puterova, M. Lexa, E. Kejnovsky, Quadruplex DNA in long terminal repeats in maize LTR retrotransposons inhibits the expression of a reporter gene in yeast, BMC Genomics 19 (2018) 184.

[19] 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.

[20] O. Kikin, L. D'Antonio, P.S. Bagga, QGRS mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences, Nucleic Acids Res. 34 (2006) W676–W682.

[21] R.W. Harkness, A.K. Mittermaier, G-quadruplex dynamics, Biochim. Biophys. Acta Protein Proteomics 1865 (2017) 1544–1554.

[22] W. Yoshida, H. Saikyo, K. Nakabayashi, H. Yoshioka, D.H. Bay, K. Iida, T. Kawai, K. Hata, K. Ikebukuro, K. Nagasawa, I. Karube, Identification of G-quadruplex clusters by high-throughput sequencing of whole-genome amplified products with a G-quadruplex ligand, Sci. Rep. 8 (2018) 1–8.

[23] E.M. Strawbridge, G. Benson, Y. Gelfand, C.J. Benham, The distribution of inverted repeat sequences in the Saccharomyces cerevisiae genome, Curr. Genet. 56 (2010) 321–340.

[24] J.E. Johnson, J.S. Smith, M.L. Kozak, F.B. Johnson, In vivo veritas: using yeast to probe the biological functions of G-quadruplexes, Biochimie 90 (2008) 1250–1263.

[25] V. Brázda, L. Hároníková, J.C.C. Liao, M. Fojta, DNA and RNA Quadruplex-binding proteins, Int. J. Mol. Sci. 15 (2014) 17493–17517.

[26] 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.

[27] V. Brázda, J. Červeň, M. Bartas, N. Mikysková, J. Coufal, P. Pečinka, The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors, Molecules 23 (2018) (pii: E2341).

[28] 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. 98 (2001) 8572–8577.

[29] K. Förstemann, J. Lingner, Molecular basis for telomere repeat divergence in budding yeast, Mol. Cell. Biol. 21 (2001) 7277–7286.

[30] P.K. Patel, R.V. Hosur, NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by Saccharomyces cerevisiae telomere repeats, Nucleic Acids Res. 27 (1999) 2457–2464.

[31] M.L. Duquette, P. Handa, J.A. Vincent, A.F. Taylor, N. Maizels, Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA, Genes Dev. 18 (2004) 1618–1629.

[32] L.A. Hanakahi, H. Sun, N. Maizels, High affinity interactions of nucleolin with G-G-paired rDNA, J. Biol. Chem. 274 (1999) 15908–15912.

[33] G.N. Parkinson, M.P.H. Lee, S. Neidle, Crystal structure of parallel quadruplexes from human telomeric DNA, Nature 417 (2002) 876–880.

[34] M.N. Weitzmann, K.J. Woodford, K. Usdin, DNA secondary structures and the evolution of hypervariable tandem arrays, J. Biol. Chem. 272 (1997) 9517–9523.

[35] J.A. Capra, K. Paeschke, M. Singh, V.A. Zakian, G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae, PLoS Comput. Biol. 6 (2010) e1000861.

[36] V. Brázda, J. Čechová, M. Battistin, J. Coufal, E.B. Jagelská, I. Raimondi, A. Inga, The structure formed by inverted repeats in p53 response elements determines the transactivation activity of p53 protein, Biochem. Biophys. Res. Commun. 483 (2017) 516–521.

[37] A. Bedrat, L. Lacroix, J.-L. Mergny, Re-evaluation of G-quadruplex propensity with G4Hunter, Nucleic Acids Res. 44 (2016) 1746–1759.

[38] 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.

[39] J. Čechová, J. Lýsek, M. Bartas, V. Brázda, Complex analyses of inverted repeats in mitochondrial genomes revealed their importance and variability, Bioinformatics 34 (2018) 1081–1085.

[40] V. Brázda, J. Lýsek, M. Bartas, M. Fojta, Complex analyses of short inverted repeats in all sequenced chloroplast DNAs, Biomed. Res. Int. 2018 (2018).

[41] S. Chodavarapu, J.M. Kaguni, Replication initiation in bacteria, Enzymes 39 (2016) 1–30.

[42] W. Fuchs, C. Ehrlich, B.G. Klupp, T.C. Mettenleiter, Characterization of the replication origin (Ori(S)) and adjoining parts of the inverted repeat sequences of the pseudorabies virus genome, J. Gen. Virol. 81 (2000) 1539–1543.

[43] H. Sun, R.J. Bennett, N. Maizels, The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs, Nucleic Acids Res. 27 (1999) 1978–1984.

[44] M.D. Huber, M.L. Duquette, J.C. Shiels, N. Maizels, A conserved G4 DNA binding domain in RecQ family helicases, J. Mol. Biol. 358 (2006) 1071–1080.

[45] Y.C. Lin, J.W. Shih, C.L. Hsu, J.J. Lin, Binding and partial denaturing of G-quartet DNA by Cdc13p of Saccharomyces cerevisiae, J. Biol. Chem. 276 (2001) 47671–47674.

[46] R. Giraldo, D. Rhodes, The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA, EMBO J. 13 (1994) 2411–2420.

[47] J. Gao, A.K. Byrd, B.L. Zybailov, J.C. Marecki, M.J. Guderyon, A.D. Edwards, S. Chib, K.L. West, Z.J. Waldrip, S.G. Mackintosh, Z. Gao, A.A. Putnam, E. Jankowsky, K.D. Raney, DEAD-box RNA helicases Dbp2, Ded1 and Mss116 bind to G-quadruplex nucleic acids and destabilize G-quadruplex RNA, Chem. Commun. 55 (2019) 4467–4470.

[48] S. Götz, S. Pandey, S. Bartsch, S. Juranek, K. Paeschke, A novel G-Quadruplex binding protein in yeast—Slx9, Molecules 24 (2019) 1774.

[49] D. Rhodes, H.J. Lipps, G-quadruplexes and their regulatory roles in biology, Nucleic Acids Res. 43 (2015) 8627–8637.

[50] Z.A.E. Waller, B.J. Pinchbeck, B.S. Buguth, T.G. Meadows, D.J. Richardson, A.J. Gates, Control of bacterial nitrate assimilation by stabilization of G-quadruplex DNA, Chem. Commun. 52 (2016) 13511–13514.

[51] A.P. David, E. Margarit, P. Domizi, C. Banchio, P. Armas, N.B. Calcaterra, G-quadruplexes as novel cis-elements controlling transcription during embryonic development, Nucleic Acids Res. 44 (2016) 4163–4173.

[52] E.W. Sayers, R. Agarwala, E.E. Bolton, J.R. Brister, K. Canese, K. Clark, R. Connor, N. Fiorini, K. Funk, T. Hefferon, J.B. Holmes, S. Kim, A. Kimchi, P.A. Kitts, S. Lathrop, Z. Lu, T.L. Madden, A. Marchler-Bauer, L. Phan, V.A. Schneider, C.L. Schoch, K.D. Pruitt, J. Ostell, Database resources of the National Center for Biotechnology Information, Nucleic Acids Res. 47 (2019) D23–D28.

[53] 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 (2019), https://doi.org/10.1093/bioinformatics/btz087.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 July 2020
doi: 10.3389/fmicb.2020.01583

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

Martin Bartas[1†], Václav Brázda[2,3†], Natália Bohálová[2,4], Alessio Cantara[2,4], Adriana Volná[5], Tereza Stachurová[1], Kateřina Malachová[1], Eva B. Jagelská[2], Otília Porubiaková[2,3], Jiří Červeň[1] and Petr Pečinka[1]*

[1] Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czechia, [2] Department of Biophysical Chemistry and Molecular Oncology, Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czechia, [3] Faculty of Chemistry, Brno University of Technology, Brno, Czechia, [4] Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia, [5] Department of Physics, Faculty of Science, University of Ostrava, Ostrava, Czechia

Non-canonical nucleic acid structures play important roles in the regulation of molecular processes. Considering the importance of the ongoing coronavirus crisis, we decided to evaluate genomes of all coronaviruses sequenced to date (stated more broadly, the order *Nidovirales*) to determine if they contain non-canonical nucleic acid structures. We discovered much evidence of putative G-quadruplex sites and even much more of inverted repeats (IRs) loci, which in fact are ubiquitous along the whole genomic sequence and indicate a possible mechanism for genomic RNA packaging. The most notable enrichment of IRs was found inside 5′UTR for IRs of size 12+ nucleotides, and the most notable enrichment of putative quadruplex sites (PQSs) was located before 3′UTR, inside 5′UTR, and before mRNA. This indicates crucial regulatory roles for both IRs and PQSs. Moreover, we found multiple G-quadruplex binding motifs in human proteins having potential for binding of SARS-CoV-2 RNA. Non-canonical nucleic acids structures in *Nidovirales* and in novel SARS-CoV-2 are therefore promising druggable structures that can be targeted and utilized in the future.

Keywords: coronavirus, genome, RNA, G-quadruplex, inverted repeats

OPEN ACCESS

Edited by:
Ilaria Frasson,
University of Padova, Italy

Reviewed by:
Naoki Sugimoto,
Frontier Institute for Biomolecular Engineering Research, Japan
Emanuela Ruggiero,
University of Padova, Italy

****Correspondence:***
Petr Pečinka
petr.pecinka@osu.cz

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Antimicrobials, Resistance and Chemotherapy, a section of the journal Frontiers in Microbiology

Received: *13 April 2020*
Accepted: *17 June 2020*
Published: *03 July 2020*

Citation:
Bartas M, Brázda V, Bohálová N, Cantara A, Volná A, Stachurová T, Malachová K, Jagelská EB, Porubiaková O, Červeň J and Pečinka P (2020) 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. 11:1583. doi: 10.3389/fmicb.2020.01583

INTRODUCTION

The order *Nidovirales* is a monophyletic group of animal RNA viruses. This group can be divided into the six distinct families of *Arteriviridae, Coronaviridae*, *Mesnidovirineae*, *Mononiviridae*, *Ronidovirineae,* and *Tobaniviridae*. All known *Nidovirales* have single-stranded, polycistronic RNA genomes of positive polarity (Modrow et al., 2013). Due to the Severe acute respiratory syndrome-related coronavirus (SARS-CoV) epidemic (November 2002–July 2003, Southern China), Middle East respiratory syndrome-related coronavirus (MERS-CoV) outbreaks (January 2014–May 2014,

Saudi Arabia and May 2015–June 2015, South Korea), and the most recent Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) worldwide pandemic (starting in November 2019 in Wuhan, China), this viral group is now extensively studied (Hung, 2003; Cowling et al., 2015; Oboho et al., 2015; Li et al., 2020).

The small size of RNA virus genomes is in principle linked to their limited ability for RNA synthesis, which is directly connected to a replication complex containing RNA-dependent RNA polymerase (RdRp) without reparation mechanisms (Gorbalenya et al., 2006). These RNA viruses can generate one mutation per genome per replication round (Drake and Holland, 1999). This combination of features means that RNA viruses are able to adapt to new environmental conditions, but they are limited in expanding their genomes because they must keep their mutation load low so that their survival is possible (Eigen and Schuster, 1977; Lauring et al., 2013; Carrasco-Hernandez et al., 2017). *Nidovirales* bind to their host cell receptors on the cell surface, after which fusion of the viral and cellular membranes is mediated by one of the surface glycoproteins. This mechanism, which progresses in the cytoplasm or endosomal membrane, releases the nucleocapsid into the host cell's cytoplasm. After genome "transportation," translation of two replicase open reading frames (ORFs) is initiated by host ribosomes. This results in large polyprotein precursors that undergo autoproteolysis to eventually produce a membrane-bound replicase/transcriptase complex. This complex initiates synthesis of the genome RNA and controls the synthesis of structural and some other proteins. New virus particles are assembled by association of the new genomes with the cytoplasmic nucleocapsid protein and subsequent envelopment of the nucleocapsid structure. Subsequently, the viral envelope proteins are inserted into intracellular membranes and targeted to the site of virus assembly (most often membranes between the endoplasmic reticulum and Golgi complex) and then they meet up with the nucleocapsid and trigger the budding of virus particles into the lumen of the membrane compartment. The newly formed virions then leave the cell by following the exocytic pathway toward the plasma membrane (Lai and Cavanagh, 1997; Gorbalenya, 2001; Snijder et al., 2003; Ziebuhr, 2004; Gorbalenya et al., 2006).

Today, SARS-CoV-2 is being studied intensively by scientists all over the world due to the ongoing 2020 coronavirus pandemic (Cohen, 2020). The origin of this virus is unknown, but recently a two-hit scenario was proposed wherein SARS-CoV-2 ancestors in bats first acquired genetic characteristics of SARS-CoV by incorporating a SARS-like receptor-binding domain through recombination prior to 2009; subsequently, the lineage that led to SARS-CoV-2 accumulated further, unique changes specifically within this domain (Patino-Galindo et al., 2020). As true of SARS-CoV, cell entry by SARS-CoV-2 depends upon angiotensin-converting enzyme 2 (ACE2) and transmembrane protease, serine 2 (TMPRSS2) for viral spike protein priming (Hoffmann et al., 2020). The genome of SARS-CoV-2 was sequenced and annotated in early January 2020 (Wu et al., 2020). A recent study revealed that the transcriptome of SARS-CoV-2 is highly complex due to numerous canonical and non-canonical recombination events. Moreover, it was found that SARS-CoV-2 produces transcripts encoding unknown ORFs and at least 41 potential RNA modification sites with an AAGAA motif were discovered in its RNA (Kim et al., 2020).

G-quadruplex binding proteins (QBPs) play crucial roles in many signaling pathways, including such biologically highly relevant activities as cell division, dysregulations of which lead to cancer development (Wu et al., 2008; Brázda et al., 2014). QBPs have been found to be involved in various viral infection pathways. An interesting example is HIV-1 nucleocapsid protein NCp7. It has been described how this protein helps to resolve an otherwise very stable G-quadruplex structure in viral RNA that stalls reverse transcription (Butovskaya et al., 2019). Various G-quadruplex-forming aptamers are used as drugs against many different viral proteins, suggesting a prominent role for QBP-mediated regulation. Among many other viruses, in particular Hepatitis C virus, HIV-1, and SARS-CoV are targets of these G-quadruplex-forming aptamers (Platella et al., 2017). Quadruplex binding domain has been found in non-structural protein 3 (Nsp3) (Lei et al., 2018). This so-called SARS-UNIQUE-DOMAIN (SUD), and especially its M subdomain, was observed to be essential for SARS replication in host cells. Deletion or even substitution mutations in key RNA-interacting amino acids were shown to result in viral inability to replicate within host cells (Kusov et al., 2015). Moreover, this subdomain was found also in MERS and several other coronaviruses (Kusov et al., 2015).

G-quadruplexes are secondary nucleic acid structures formed in guanine-rich strands (Burge et al., 2006; Vorlíčková et al., 2012; Kolesnikova and Curtis, 2019). These have been detected in various genomes, but most extensively they have been described in human genomes (Chambers et al., 2015; Bedrat et al., 2016; Hänsel-Hertsch et al., 2016). They are present also in viruses (Lavezzo et al., 2018; Frasson et al., 2019). G-quadruplexes probably play an important role in regulating replication in most viral nucleic acids (Lavezzo et al., 2018), and these structures have been suggested as targets for antiviral therapy (Métifiot et al., 2014; Ruggiero and Richter, 2018, 2020). Along with cruciforms and hairpins, which can be formed in nucleic acids by inverted repeats (IRs), G-quadruplexes are significant genome elements playing specific biological roles. They are involved, for example, in the regulation of DNA replication and transcription (Bagga et al., 1990; Yu, 2009; Brázda et al., 2011). It has been demonstrated that IRs are important for various processes in viruses, including their genome organization (Li and Li, 2010; Ishimaru et al., 2013; Xie et al., 2017; Bridges et al., 2019). Another interesting RNA motif that has been used as a drug target and was found in SARS-CoV targeted by 1,4-diazepame is the "slippery sequence" followed by a pseudoknot (Plant et al., 2005). This structure, common among all coronaviruses, works based on ribosomal −1 frameshifting that switches on viral fusion proteins (Plant et al., 2005).

In all prokaryotic and eukaryotic cells, as well as viruses, there have been found sequence motifs such as IR sequences forming cruciforms and hairpins or G-rich sequences that form G-quadruplexes (Brázda et al., 2011; Lavezzo et al., 2018; Bartas et al., 2019). In the present research, we conducted a systematic and comprehensive bioinformatic study searching

for the occurrence of IRs and putative quadruplex sites (PQSs) within the genomes of all known *Nidovirales*. The aim was to find one or more potential druggable RNA targets to address the present COVID-19 threat (**Figure 1**).

MATERIALS AND METHODS

Genome Source

Linear genomes of 109 viruses from the order *Nidovirales* were downloaded from the genome database of the National Center for Biotechnology Information (NCBI) (Federhen, 2011). Full names, phylogenetic groups, exact NCBI accession numbers, and further information are summarized in **Supplementary Material 1**.

Nidovirales Phylogenetic Tree Construction

Exact taxonomic identifiers of all analyzed *Nidovirales* species (obtained from Taxonomy Browser via NCBI Taxonomy Database Drake and Holland, 1999 were downloaded to phyloT: a tree generator (http://phylot.biobyte.de) and a phylogenetic tree was constructed using the function "Visualize in iTOL" in the Interactive Tree of Life environment (Letunic and Bork, 2019).

Analyses of PQSs Occurrence in *Nidovirales* Sequences

Nidovirales sequences were analyzed using our G4Hunter Web Tool (Brázda et al., 2019), and selected sequences were verified also by QGRS mapper (Kikin et al., 2006) and G4screener (Garant et al., 2018). G4Hunter web is a more recent tool compared to Quadparser and QGRS mapper, and this algorithm allows quantitative analyses whereby G4 propensity is calculated depending on G richness and G skewness. Moreover, a new implementation of G4Hunter allows for performing batch analyses. In general, the earlier algorithms did not consider atypical quadruplex-forming structures, as have been described in various outstanding articles (Guedin et al., 2010; Kocman and Plavec, 2017; Lightfoot et al., 2019). The G4Hunter Web software is capable to read the NCBI identifier of the sequences uploaded in a.csv file. The parameters for G4Hunter were set to 25 as window size and G4Hunter score above 1.2. The results for each analyzed sequence contained information about the size of the genome and number of putative PQSs. All the results were merged into a single Microsoft Excel file where statistical analysis was then made. We also downloaded features tables of each virus from the NCBI database. These tables contain information about known features in the genome of each species. We searched the occurrence of G-quadruplex-forming sequences before, inside, and after the specific features of each genome using a predefined feature neighborhood ±100 nucleotides (nt). Data were then plotted in Excel, where the statistical analysis was made. Complete analyses of PQS occurrence in *Nidovirales*, including a list of those PQSs found, are provided in **Supplementary Material 2**.

Analyses of IRs Occurrence in *Nidovirales* Sequences

All *Nidovirales* genomes were analyzed by the core of our Palindrome analyzer webserver (Brázda et al., 2016). The

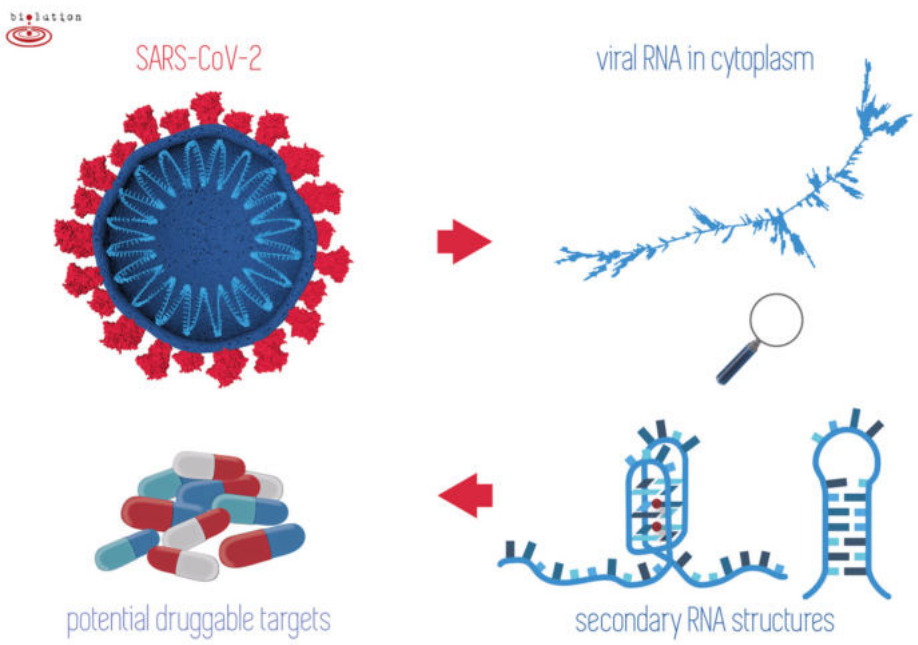


FIGURE 1 | When outside host cells, the RNA of SARS-CoV-2 is highly spiralized by means of nucleocapsid phosphoproteins (top left). Once RNA is released into the cytoplasm and during replication/transcription processes, the formation of G-quadruplex and/or hairpins may take place (right). Stabilization of these structures by existing drugs may play a crucial role in inhibiting the viral lifecycle (bottom left).

software was modified to read NCBI identifiers of sequences from text files. The size of IRs was set to 6–30 nt, size of spacers to 0–10 nt, and maximally one mismatch was allowed. A separate list of IRs in each of the 109 sequences and an overall report were exported. The overall report contained the lengths of the analyzed sequences, total number of IRs found, numbers of IRs grouped by size of IR (6–30 nt), and sum of IRs longer than 8, 10, and 12 nt. The software also counted frequency of IRs in each sequence. Frequencies of IRs were normalized per 1,000 nt. Features tables of 109 *Nidovirales* genomes were downloaded from the NCBI database and grouped by their names as stated in the feature table file. Analyses of IRs occurrence inside and around (before and after) these features was performed. The search for IRs took place in predefined feature neighborhoods ±100 nt around and inside feature boundaries. We calculated the numbers of all IRs and of those longer than 8, 10, and 12 nt in regions before, inside, and after the features. The categorization of an IR according to its overlap with a feature or feature neighborhood is demonstrated by the example shown in **Supplementary Material 3**. Complete analyses of IRs occurrence in *Nidovirales* are provided in **Supplementary Material 4**.

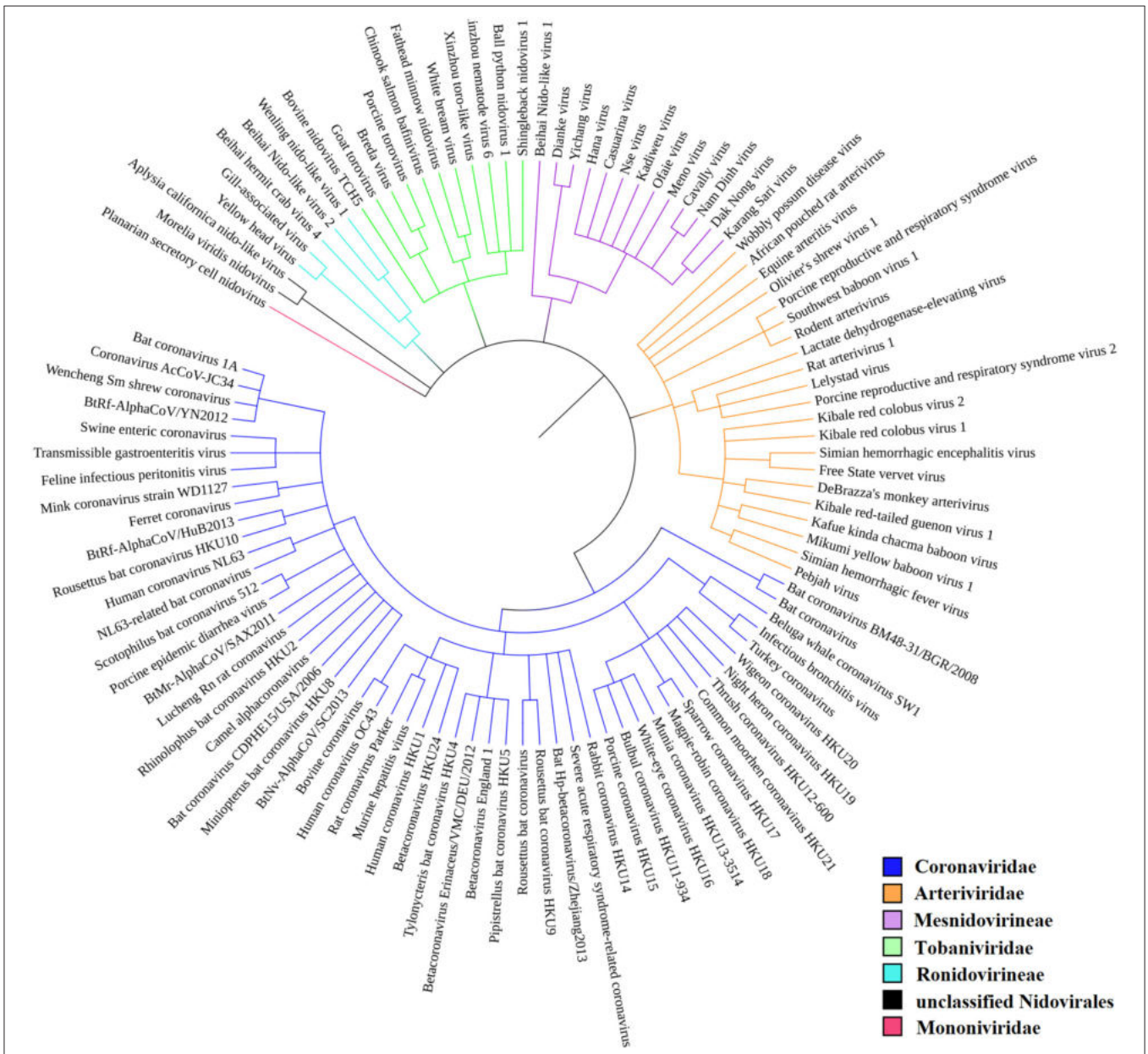


FIGURE 2 | Phylogenetic relationships in *Nidovirales*. SARS-CoV-2 is not yet recognized by iToL, but, based upon current evidence, it probably will be placed close to the Bat Hp-betacoronavirus/Zheijang2013 and Severe acute respiratory syndrome-related coronavirus (SARS-nCoV) branches (Lu et al., 2020).

RNA Fold Predictions

In order to be able to display higher structures of the coronavirus genome, we used Galaxy's free-online webserver (Afgan et al., 2018) and its RNA fold tool (Lorenz et al., 2011). This tool allows quick calculation of minimum free energy of secondary structures. We left the default parameters (Temperature 37°C, Unpaired bases to participate in all dangling ends, Naview layout). We used SARS-CoV-2 genomic RNA sequence (NC_045512.2) in FASTA format as the input format. The output data were then displayed using the RNA plot tool (Lorenz et al., 2011). We again left the default parameters (Naview layout, Output format Postscript.ps). We then downloaded the displayed secondary structures in high-resolution format. The raw data are provided in **Supplementary Material 5**.

Multiple Alignment of SUD Domains (M Regions) in Nsp3 of Pathogenic Species

Multiple protein alignment was done using MUSCLE (Edgar, 2004) under default parameters [UGENE (Okonechnikov et al., 2012) workflow was used]. The following accessions were used: NP_828862.2 (Nsp3 SARS-CoV), YP_009047231.1 (Nsp3 MERS-CoV), and YP_009725299.1 (Nsp3 SARS-CoV-2). Conserved regions were added according to a graph published previously (Kusov et al., 2015).

Prediction of Human RNA-Binding Proteins Sites in SARS-CoV-2 RNA

The human SARS-CoV-2 RNA sequence was downloaded from NCBI (accession NC_045512.2) in FASTA format and inserted into the RBPmap (Mapping Binding Sites of RNA-binding proteins) web-based tool (Paz et al., 2014). The database of 114 human experimentally validated motifs was used. Both the "High stringency" and "Apply conservation filter" options were used. The output was further filtered in Excel to keep only those hits below $p = 1.10^{-6}$. The complete results are provided in **Supplementary Material 6**.

Statistical Analyses

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 *Nidovirales* species (**Figure 4**). The following values were used as input data: frequency of PQS per 1,000 nt with threshold G4Hunter score of 1.2; frequency of PQS per 1,000 nt with threshold G4Hunter score of 1.4; and % PQS in genome (coverage) (**Supplementary Material 7**). A cluster dendrogram of IRs (**Figure 6**) was constructed from these values: frequency of IRs per 1,000 nt; frequency of IRs per 1,000 nt of length 8+; frequency of IRs per 1,000 nt of length 10+; and frequency of IRs per 1,000 nt of length 12+. 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 AU values above 95, equivalent to $p < 0.05$) are highlighted by rectangles marked with broken red lines. R code is provided in **Supplementary Material 7**.

To determine whether SARS-CoV-2 significantly differs in frequencies of PQSs and IRs compared to randomly shuffled sequences ($N = 10$) of the same length and nucleotide composition, we performed a two-sided Wilcoxon signed-rank test (with p-value cutoff at 0.05). This test was run for plus and minus strands separately.

RESULTS

Phylogenetic Relationships in *Nidovirales*

According to the current state of knowledge, the order *Nidovirales* can be divided into six distinct families (Liò and Goldman, 2004; Hanada et al., 2005; Chen et al., 2020). There are just two unclassified species among all those species recorded within NCBI Genome (**Figure 2**). *Coronaviridae* is the largest family and consists of 56 species. Among these are 12 species able

TABLE 1 | Genomic sequence sizes, PQS frequencies, and total counts.

All	Seq	Median	Short	Long	PQS	Mean f	Min f	Max f	GC%
Nidovirales	109	26,975	12,704	41,178	1,021	0.48	0	9.07	42.15
Family									
Arteriviridae	22	15,236	12,704	15,728	392	1.18	0.45	2.72	51.60
Coronaviridae	56	28,345	20,398	31,686	162	0.10	0	0.52	39.39
Mesnidovirineae	12	20,117	19,867	20,949	250	1.03	0	9.07	38.75
Ronidovirineae	5	26,253	24,648	29,385	51	0.39	0.04	1.07	44.79
Tobaniviridae	11	27,318	20,261	33,452	129	0.43	0	1.48	40.94
By hosts									
Invertebrates	17	20,307	19,917	41,178	298	0.82	0	9.07	39.76
Vertebrates (including Humans)	83	27,608	12,704	33,452	591	0.38	0	1.75	42.74
Humans	13	29,751	27,317	31,028	21	0.06	0	0.32	37.88
SARS-CoV-2	1	29,903	–	–	1	0.03	–	–	37.97

Seq (total number of sequences), Median (median length of sequences), Short (shortest sequence), Long (longest sequence), PQS (total number of predicted PQSs), Mean f (mean frequency of predicted PQS per 1,000 nt), Min f (lowest frequency of predicted PQS per 1,000 nt), Max f (highest frequency of predicted PQS per 1,000 nt), GC% (mean GC content). Note that 132 PQSs were in Nidovirales with Undefined host.

to infect humans, including SARS-CoV, MERS-CoV, and SARS-CoV-2. Second largest is the *Arteriviridae* family with 22 species, third largest is the *Mesnidovirineae* family with 13 species, fourth is the *Tobaniviridae* family with 11 species, including 1 species able to infect humans. Fifth largest is the *Ronidovirineae* family with 5 species, and sixth is the *Mononiviridae* family with only 1 species.

Variation in PQS Frequency in *Nidovirales*

We analyzed the occurrence of PQSs using G4Hunter in all 109 known genomes of *Nidovirales*. The length of genomes in the dataset varies from 12,704 nt (Equine arteritis virus) to 41,178 nt (Planarian secretory cell nidovirus). The mean GC content is 42.15%, the minimum is 27.50% for Planarian secretory cell nidovirus (*Mononiviridae* family), and the maximum is 57.35% for Beihai Nido-like virus 1 (*Mesnidovirineae* family). Using standard values for the G4Hunter algorithm (i.e., window size 25 and G4Hunter score above 1.2), we found 1,021 PQSs among all 109 *Nidovirales* genomes. The most abundant PQSs are those with G4Hunter scores of 1.2–1.4 (98.24% of all PQS), much less abundant are PQSs with G4Hunter scores 1.4–1.6 (1.76% of all PQS), and there were no PQSs above the 1.6 G4Hunter score threshold. In general, a higher G4Hunter score means a higher probability of G-quadruplexes forming inside the PQS (Bedrat et al., 2016). Genomic sequence sizes, total PQS counts, and PQS frequencies characteristics are summarized in **Table 1**.

The highest PQS frequencies were found in Beihai Nido-like virus 1, which is an intracellular parasite of sea snails in genus *Turritella*. Beihai Nido-like virus 1 has in total 184 PQSs in its genomic sequence 20,278 nt long and PQS frequency of 9.07 PQS per 1,000 nt. On the other hand, no PQSs were found in the genomic sequences of 16 other *Nidovirales* species. The mean PQS value for all *Nidovirales* was 0.48 PQS per 1,000 nt. By viral families, the highest mean PQS frequency per 1,000 nt was in *Arteriviridae* (1.18), followed by in *Mesnidovirineae* (1.03), much lower in *Tobaniviridae* (0.43) and in *Ronidovirineae* (0.39), and the lowest was in *Coronaviridae* (0.10). When grouped and analyzed by host organisms, new information became apparent. The highest mean PQS frequency was in *Nidovirales* infecting Invertebrates (0.80), then in *Nidovirales* infecting Vertebrates including Humans (0.38), and the last in *Nidovirales* infecting Humans (0.06). In pathogenic human coronaviruses, the total PQSs counts were as follow: 4 PQS in SARS-CoV, 0 PQS in MERS-CoV, and 1 PQS in SARS-CoV-2. To test whether the PQS frequency per 1,000 nt in SARS-CoV-2 is significantly different than in a randomly shuffled sequence, we generated 10 scrambled sequences with length and nucleotide content the same as in SARS-CoV-2. In that test, the mean PQS frequency per 1,000 nt was 0.191 and standard deviation was 0.101. We performed Wilcoxon signed-rank test with continuity correction and the resulting p-value was 0.008 Thus, the PQS frequency per 1,000 nt in SARS-CoV-2 (0.033) is significantly lower than expected (only about one-sixth).

All PQSs found in ranges of G4Hunter score intervals and precomputed PQS frequencies per 1,000 nt are summarized in **Table 2**. The relationships between observed PQS frequency per 1,000 nucleotides and GC content in all analyzed *Nidovirales* sequences are depicted in **Figure 3**. The Beihai Nido-like virus has both the highest GC content and highest PQS frequency. Within *Nidovirales* with Humans as host, however, the highest PQS frequency was found in Breda virus, which does not have the highest GC content in the dataset.

TABLE 2 | Total number of PQSs and their resulting frequencies per 1,000 nt in all 109 genomes of *Nidovirales* and in particular categories according to their hosts, grouped by G4Hunter score.

Interval of G4Hunter Score	Number of PQSs in dataset	PQS frequency per 1,000 nt
All		
1.2	1,003	0.47
1.4	18	0.01
Arteriviridae		
1.2	386	1.17
1.4	6	0.02
Coronaviridae		
1.2	157	0.10
1.4	5	0.003
Mesnidovirineae		
1.2	245	1.01
1.4	5	0.02
Ronidovirineae		
1.2	50	0.38
1.4	1	0.01
Tobaniviridae		
1.2	129	0.43
1.4	0	0
Humans only		
1.2	20	0.05
1.4	1	0.003
Vertebrate (Including humans)		
1.2	580	0.37
1.4	11	0.01
Invertebrates		
1.2	292	0.80
1.4	6	0.02

Frequency was computed using total number of PQSs in each category divided by total length of all analyzed sequences and multiplied by 1,000. Note that 132 PQSs were in Nidovirales with Undefined host.

Most of the PQSs have G4Hunter scores between 1.2 and 1.4. Only a few sequences have G4Hunter scores above 1.4. In comparison with other *Nidovirales*, the members of the *Coronaviridae* and especially pathogenic human *Coronaviridae* have the lowest PQS frequency in the dataset.

A cluster dendrogram based on the PQS characteristics (**Figure 4**) further revealed interesting information. Beihai Nido-like virus 1 was confirmed as an outlier from the rest of the *Nidovirales*. All viral species belonging to the largest *Coronaviridae* family are located together in the middle of the cluster dendrogram. The relatively abundant family *Arteriviridae* clustered together mainly on the right side of the cluster dendrogram. Based

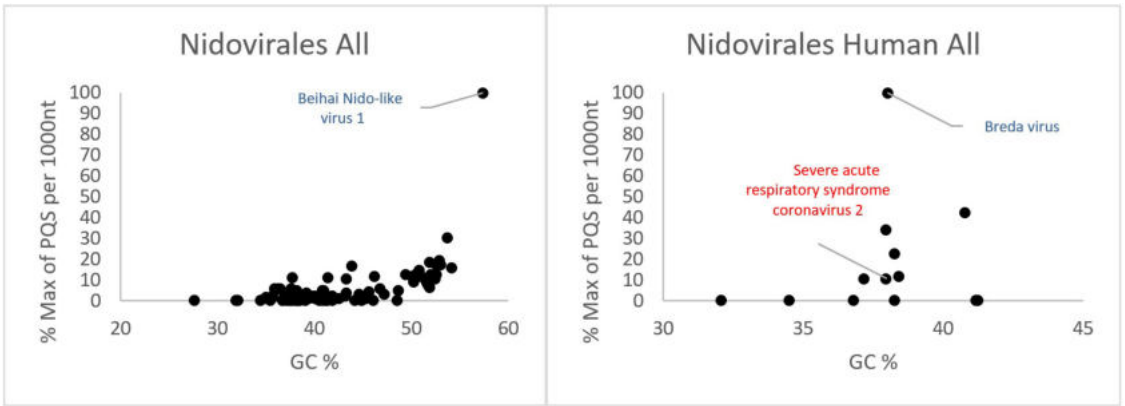


FIGURE 3 | Relationship between observed PQS frequency per 1,000 nucleotides and GC content in all analyzed *Nidovirales* sequences. In each G4Hunter score interval miniplot, frequencies were normalized according to the highest observed PQS frequency. Organisms with maximum frequency per 1,000 nt >50% are described and highlighted in color.

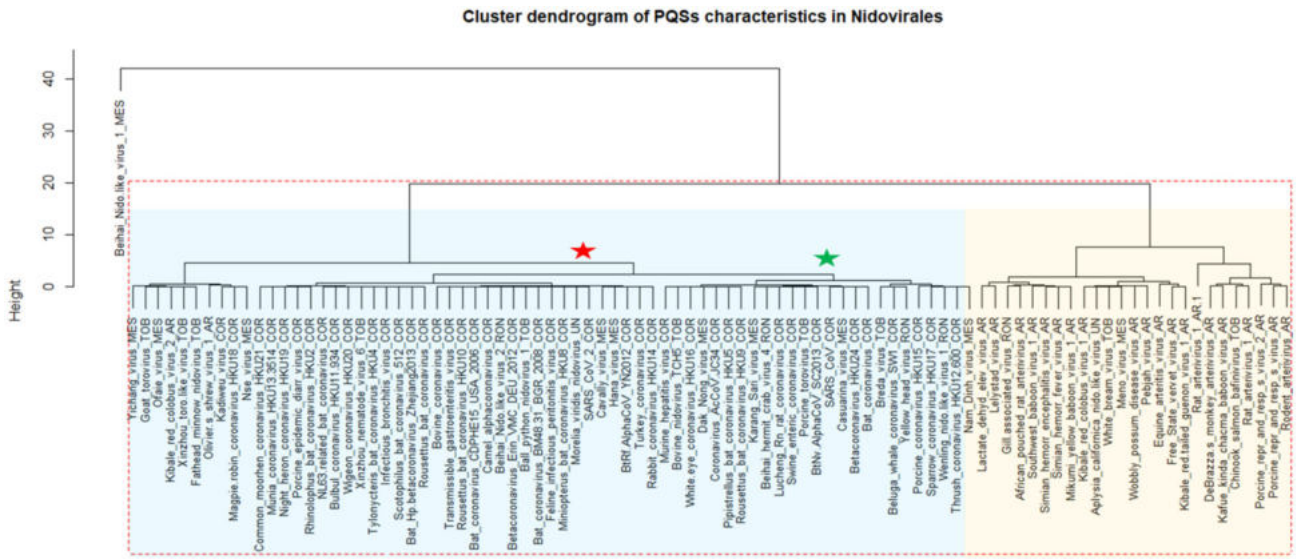


FIGURE 4 | Cluster dendrogram based on PQS characteristics in *Nidovirales*. Input data are listed in **Supplementary Material 7**. Statistically significant clusters (based on AU values above 95, equivalent to *p*-values lower than 0.05) are highlighted by rectangles drawn with broken red lines. Two main branches are highlighted by light blue and light orange backgrounds. In creating the PQS diagram, only those viruses with at least one predicted PQS were used. *Nidovirales* families are marked by suffixes AR (*Arteriviridae*), COR (*Coronaviridae*), MES (*Mesnidovirineae*), MON (*Mononiviridae*), RON (*Ronidovirineae*), TOB (*Tobaniviridae*), or UN (unclassified Nidovirales). The green star indicates SARS-CoV and the red star indicates SARS-CoV-2.

on dendrogram distances, we highlighted two main branches—one consisting mainly of *Coronaviridae* (light blue background) and the second of *Arteriviridae* (light orange background).

Figure 5 shows the differences in PQS frequency by annotated loci. We downloaded features for every virus genome and analyzed the presence of PQS in each annotated sequence and in its proximity (before and after). The most notable enrichment of PQSs was located inside 5′UTR and before and inside 3′UTR. The lowest PQS frequencies were found after 3′UTR, before 5′UTR, and after miscellaneous features.

Variation in Frequency of Inverted Repeats in *Nidovirales*

To find short IRs in *Nidovirales* genomic sequences, we utilized the core of the *Palindrome analyzer* (Brázda et al., 2016). We used values for *Palindrome analyzer* returning inverted repeats capable to form stable hairpins (size of IRs was set to 6–30 nt, size of spacer to 0–10 nt, maximally one mismatch). Genomic sequence sizes, total IR counts, and IR frequencies characteristics are summarized in **Table 3**. In total, 93,369 IRs were found and the mean IRs frequency was 33.90 per 1,000 nt. The maximal IR frequency was found in Planarian secretory cell nidovirus (56.17),

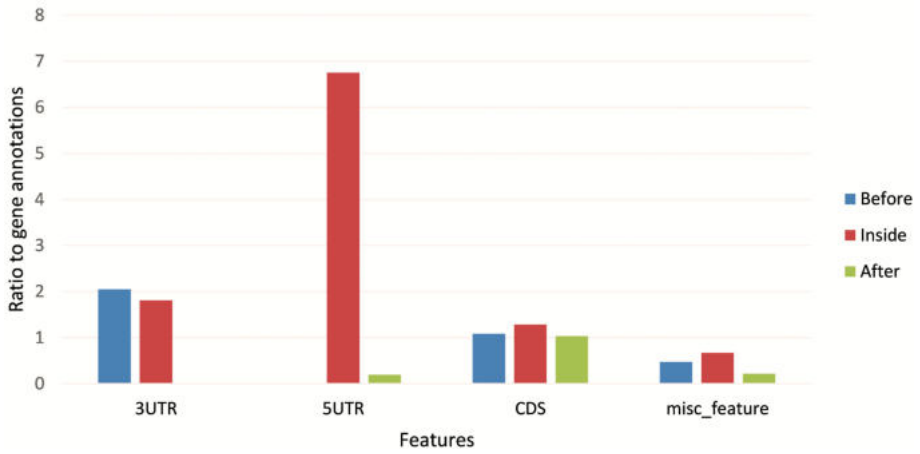


FIGURE 5 | Differences in PQS frequency by annotated locus. The figure shows the PQS frequencies between annotations from the NCBI database. We analyzed frequencies of all PQSs inside, before, and after the annotations.

TABLE 3 | Genomic sequence sizes, IRs frequencies and total counts.

Order	Seq	Median	Short	Long	IRs	Mean f	Min f	Max f	GC%
Nidovirales	109	26,975	12,704	41,178	93,369	33.90	19.23	56.17	42.15
Family									
Arteriviridae	22	15,236	12,704	15,728	9,695	29.44	26.47	33.57	51.60
Coronaviridae	56	28,345	20,398	31,686	58,943	37.08	31.13	48.16	39.39
Mesnidovirineae	12	20,117	19,867	20,949	8,012	33.04	19.23	39.06	38.75
Ronidovirineae	5	26,253	24,648	29,385	3,474	26.37	22.20	29.62	44.79
Tobaniviridae	11	27,318	20,261	33,452	9,082	30.24	23.96	37.95	40.94
Host									
Invertebrates	17	20,307	19,917	4,178	13,233	31.77	19.23	56.17	39.76
Vertebrates (including Humans)	83	27,608	12,704	33,452	45,844	34.25	24.18	48.16	42.74
Humans	13	29,751	27,317	31,028	14,408	38.13	33.89	43.44	37.88
SARS-CoV-2	1	29,903	–	–	1,203	40.23	–	–	37.97

Seq (total number of sequences), Median (median length of sequences), Short (shortest sequence), Long (longest sequence), GC% (mean GC content), IRs (total number of predicted IRs), Mean f (mean frequency of predicted IRs per 1,000 nt), Min f (lowest frequency of predicted IRs per 1,000 nt), and Max f (highest frequency of predicted PQS per 1,000 nt) are shown for all 109 Nidovirales sequences, families, host groups, and for SARS-CoV-2 alone.

the species with the largest genome among *Nidovirales*. It has been suggested that Planarian secretory cell nidovirus diverged early from multi-ORF *Nidovirales* and acquired additional genes, including those typical of large DNA viruses or hosts (RNAse T2, Ankyrin, and Fibronectin type II), which might modulate virus–host interactions (Saberi et al., 2018). The lowest IR frequency was noticed in Beihai Nido-like virus 1 (19.23). Noteworthy is that this is the species of *Nidovirales* with the highest GC content and PQS frequency.

A cluster dendrogram based on the IR characteristics (**Figure 6**) further revealed significant differences in IR presence within *Nidovirales* genomes. Based on dendrogram distances, we highlighted two main branches—one consisting mainly of *Coronaviridae* (light blue background) and the second of *Arteriviridae* (light orange background). It is noteworthy that SARS-CoV and MERS-CoV are clustered adjacent to one another (green asterisk), but the novel coronavirus SARS-CoV-2 is located relatively far away from them (red asterisk).

Differences in IR frequency by annotated locus are depicted in **Figure 7**. By families, the highest mean IRs frequency was found in *Coronaviridae* (37.08) and lowest in *Ronidovirineae* (26.37). Novel human coronavirus SARS-CoV-2 has relatively high IR frequency per 1,000 nt in comparison with other *Nidovirales*. To test if the frequency of IRs per 1,000 nt in SARS-CoV-2 is significantly greater than in randomly shuffled sequences, we generated 10 scrambled sequences with length and nucleotide content the same as in SARS-CoV-2 (for both positive and negative strands). For both positive and negative strands, the results were very similar (mean frequencies of IRs per 1,000 nt were 34.43 $\pm$ 0.54 and 34.47

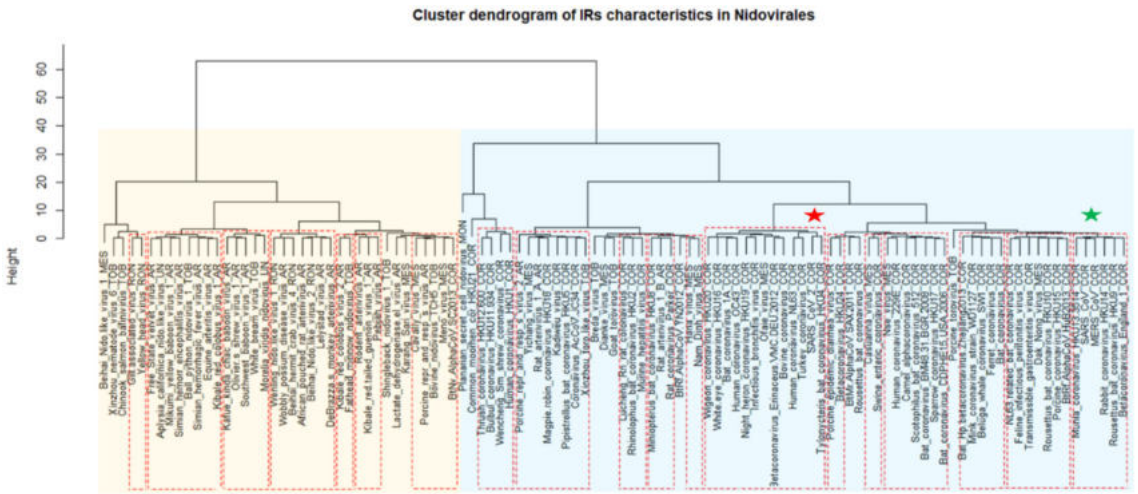


FIGURE 6 | Cluster dendrogram based on IR characteristics in *Nidovirales*. Input data are listed in **Supplementary Material 7**. Statistically significant clusters (based on AU values above 95, equivalent to *p*-values lower than 0.05) are highlighted by rectangles drawn with broken red lines. Two main branches are highlighted by light blue and light orange backgrounds. *Nidovirales* families are marked by suffixes AR (*Arteriviridae*), COR (*Coronaviridae*), MES (*Mesnidovirineae*), MON (*Mononiviridae*), RON (*Ronidovirineae*), TOB (*Tobaniviridae*), or UN (unclassified Nidovirales). The green star indicates SARS-CoV and MERS and red star indicates SARS-CoV-2.

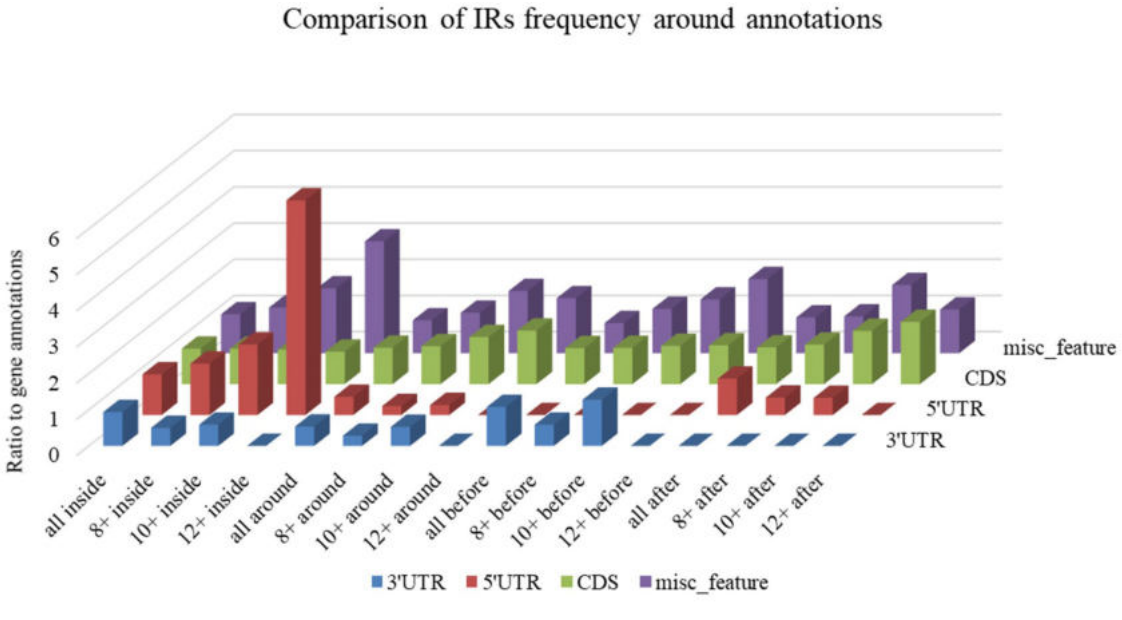


FIGURE 7 | Differences in IR frequency by annotated locus. The chart compares IR frequencies per 1,000 nt between "gene" annotation and other annotated locations from the NCBI database that were found in genomes more than 10 times. We analyzed frequencies of all IRs (all) and of IRs with lengths 8 nt and longer (8+), 10 nt and longer (10+,) and 12 nt and longer (12+) within annotated locations (inside) as well as before and after annotated locations.

± 0.61, respectively). Wilcoxon signed-rank test with continuity correction showed the IR frequency per 1,000 nt in SARS-CoV-2 (40.23) to be significantly higher than expected (*p*-value 0.003). When we inspected differences of IR frequencies according to hosts, we found the highest mean IR frequency per 1,000 nt to be in *Nidovirales* infecting Humans (38.13), then in Vertebrates (34.25), and the lowest mean frequency is in Invertebrates (31.77).

A summary of all IRs found in ranges of different IR sizes and precomputed IR frequencies per 1,000 nt is shown in **Table 4**. Although generally the frequency of IR presence decreases with the IR length, there are notable differences between groups and also between viruses with different hosts. The most as well as longest IRs occur in the *Coronaviridae* group and in viruses having humans as a host. IRs 12 bp long and longer are very rare in the *Ronidovirineae* group. The relationship between IRs

length and stability of resulting secondary structure is not simple. While some authors believe that longer IRs are more stable, others suggest that there is an energy optimum defined by arm and spacer length (Sinden et al., 1991; Brázda et al., 2016; Georgakopoulos-Soares et al., 2018).

Differences in IR frequency according to annotated loci are shown in **Figure 5**. The most notable enrichment of IRs was found inside 5′UTR for IRs of size 12+ nt, and this is the most frequently occurring location of 12+ IRs in *Nidovirales* genomes. Noteworthy is that there are no 12+ IRs around 5′UTR loci, but there is an abundance of IRs 10+ nt long in these locations. The 5′UTR are abundant for 12+ IR, but there are no 12+ IRs within 3′UTR. This points to functional relevance of these IRs in viral genomes.

Prediction of Human SARS-CoV-2 RNA Folding

It is very likely that coronavirus RNA in the viral capsid is compactly folded into a spiral-like structure due to nucleocapsid phosphoprotein N, as was proposed earlier for SARS-CoV (Chang et al., 2014). But what structure does the RNA takes after the capsid content is released into the cytoplasm of the host cell? We made a global prediction of RNA folding for SARS-CoV-2 and in a negative control (random RNA sequences with the same GC content and length). It was clearly visible that there arose a much more compact RNA structure in SARS-CoV-2 (**Figure 8**). The RNA folding algorithm shows that inverted repeats forming hairpins have the main importance for its folding. Due to the relative lack of PQSs with high G4Hunter scores in the SARS-CoV-2 positive strand, we assume that G-quadruplexes are not important for its folding. For the 10 longest IRs obtained from Palindrome analyzer, we prepared local predictions of their 2D structures using the Mfold webserver (Zuker, 2003) (**Supplementary Material 8**).

Comparison of SUD Domain M Region of Nsp3 in Three Pathogenic Human Coronaviruses

Consisting of nearly 2,000 amino acids, Nsp3 is the largest multi-domain protein in *Nidovirales*. Nevertheless, Nsp3's role is still largely unknown. It is believed that the protein plays various roles in coronavirus infection. Nsp3 interacts with other viral Nsps as well as RNA to form the replication/transcription complex. It also acts on posttranslational modifications of host proteins to antagonize the host innate immune response. On the other hand, Nsp3 is itself modified in host cells by N-glycosylation and can interact with host proteins to support virus survival (Lei et al., 2018). Both SARS-CoVs (2003 and 2019) have PQSs in their genomes and have a retained M region of the SUD domain in protein Nsp3 that is reported to be critical for interacting with G-quadruplexes, and particularly all G-quadruplex interacting residues, as proposed by Kusov et al. (2015), are 100% conserved (**Figure 9**). MERS has no PQS in its RNA sequence and, strikingly, it has a deletion in this region of the SUD domain suggesting parallel evolution simplifications.

TABLE 4 | Total number of IRs and their resulting frequencies per 1,000 nt grouped by size of IR.

Size of IRs	Number of IRs in dataset	IRs frequency per 1,000 nt
All		
All	93,369	33.90
8+	12,669	4.54
10+	1,872	0.65
12+	244	0.08
Arteriviridae		
All	9,695	29.44
8+	1,127	3.41
10+	103	0.31
12+	8	0.02
Coronaviridae		
All	58,943	37.08
8+	8,527	5.36
10+	1,393	0.88
12+	190	0.12
Mesnidovirineae		
All	8,012	33.04
8+	1,081	4.46
10+	137	0.56
12+	13	0.05
Ronidovirineae		
All	3,474	26.37
8+	320	2.44
10+	25	0.18
12+	2	0.01
Tobaniviridae		
All	9,082	30.24
8+	1,060	3.54
10+	128	0.43
12+	15	0.05
Humans only		
All	14,408	38.13
8+	2,048	5.40
10+	344	0.91
12+	42	0.11
Vertebrate (Including humans)		
All	72,838	34.25
8+	10,097	4.66
10+	1,549	0.69
12+	207	0.09
Invertebrates		
All	13,233	31.77
8+	1,686	4.06
10+	218	0.51
12+	24	0.05
SARS-CoV-2		
All	1,203	40.23
8+	203	6.79
10+	37	1.24
12+	5	0.17

Shown are data for all 109 representative Nidovirales sequences, for families, host groups, and for SARS-CoV-2 alone. Frequency was computed using total number of IRs in each category divided by total length of all analyzed sequences and multiplied by 1,000.

FIGURE 8 | RNA fold (Lorenz et al., 2011) prediction for SARS-CoV-2 RNA molecule **(Left)** and random sequence negative control **(Right)**. This figure shows a high level of SARS-CoV-2 genome folding via complementarity of particular RNA regions and forming of hairpins and/or cruciforms. RNA fold prediction was carried out using default parameters via Galaxy webserver (Afgan et al., 2018), which enables queries of lengths longer than 10,000 nucleotides.

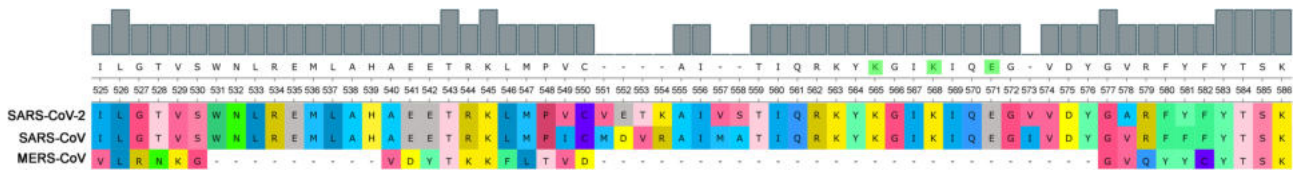


FIGURE 9 | Comparison of SUD domain M region in three pathogenic human coronaviruses. Nsp3 proteins of three pathogenic human viruses (SARS-CoV-2, SARS-CoV, and MERS-CoV) were aligned and part of the critical G-quadruplex binding M region of the SUD domain (525–577 aa according to the SARS-CoV reference sequence) was visualized. As predicted according to Kusov et al. (2015), the G-quadruplex-interacting residues K565, K568, and E571 are highlighted in green (upper consensus panel).

Prediction of Human RNA-Binding Proteins Sites in SARS-CoV-2 RNA

Using the RBPmap tool, we predicted human RNA-binding proteins sites in SARS-CoV-2 RNA. While holding to very stringent thresholds, three highly promising candidate human RNA-binding proteins were predicted. SRSF7 was predicted to bind viral RNA in nucleotide position 26,194–26,199 (NC_045512.2), which is in fact exactly the binding motif for this protein (ACGACG). Protein HNRNPA1 was predicted to bind viral RNA in nucleotide position 23,090–23 097 (exact binding motif GUAGUAGU). The last protein found was TRA2A in nucleotide position 3,056–3,065 and position 3,074–3,083, in both cases with one mismatch (the motifs found were GAAGAAGAAG and the experimentally validated motif for TRA2A is GAAGAGGAAG). All three of these proteins share multiple RGG-rich novel interesting quadruplex interaction (NIQI) motifs, which are common to most human G-quadruplex-binding proteins (Brázda et al., 2018; Huang et al., 2018; Masuzawa and Oyoshi, 2020). All find individual motif occurrence (FIMO) hits below $p = 1.10^{-4}$ are enclosed in **Supplementary Material 9**. The proteins SRSF7, HNRNPA1, and TRA2A are involved in mRNA splicing via the spliceosome pathway (Szklarczyk et al., 2015).

DISCUSSION

Local DNA structures have been shown to play important roles in basic biological processes, including replication and transcription (Brázda et al., 2011; Travers and Muskhelishvili, 2015; Surovtsev and Jacobs-Wagner, 2018). PQS analyses of human viruses have clearly demonstrated that these sequences are conserved also in the viral genomes (Lavezzo et al., 2018) and could be targets for antiviral therapies (Wang et al., 2015; Krafčíková et al., 2017; Ruggiero and Richter, 2018). For example, the conserved PQS sequence present in the L gene of Zaire ebolavirus and related to its replication is inhibited by interaction with G-quadruplex ligand TMPyP4, and this finding has led to suggestions that G-quadruplex RNA stabilization could constitute a new strategy against Ebola virus disease (Wang et al., 2016). A similar strategy has been proposed against Hepatitis C virus belonging to the *Flaviviridae* family of positive ssRNA viruses. Stabilization of G4s by Phen-DC3 ligand has been shown to lead to inhibited

replication of this virus (Jaubert et al., 2018). Viruses with single-stranded genomes are often non-symmetrically distributed with guanines and cytosines. For example, Zika viral genome (*Flaviviridae* family) has a G-rich positive-sense genome and a C-rich negative-sense strand. Fleming et al. found more than 60 PQSs in the Zika virus genome on the positive strand but no PQSs on the negative strand. This observation identifies a large asymmetry with respect to PQS content between the two strands. The strand asymmetry for PQS sites likely results from the high guanine content relative to cytosine content in the positive-sense strands (Fleming et al., 2016). There are many PQS searching algorithms (reviewed by Puig Lombardi and Londoño-Vallejo, 2020). We used the G4Hunter algorithm. Because it had originally been written to analyze DNA (Bedrat et al., 2016), the G4Hunter algorithm analyzed both strands simultaneously (Brázda et al., 2019). Analyses of both strands are therefore presented in our manuscript. The results with positive G4Hunter scores show PQSs in the plus strand and the results with negative G4Hunter scores show PQSs in the minus strand. In our study, we analyzed all accessible genomes of *Nidovirales,* thus including also the contemporary pandemic SARS-CoV-2 genome. Our analyses found only one PQS in the genomic sequence of SARS-CoV-2, that having a G4Hunter score of −1.24 and the following sequences: 5′-CCCCAAAAUCAGC GAAAUGCACCCC-3′ for a positive-strand intermediate and 5′-GGGGUGCAUUUCGCUGAUUUUGGGG-3′ for a negative-strand intermediate (where G-quadruplex could theoretically arise). This sequence was not identified by other prediction algorithms that use searching based upon regex sequence or that are pattern-based (PQSfinder or QGRS mapper), and it has been given only a low score by the new machine-learning G4screener algorithm (Garant et al., 2018). Nevertheless, none of these algorithms consider possible substitutions of guanine by adenine in quadruplex tetrads, as has been described by Kocman and Plavec (2017), or other untypical quadruplexes as reviewed by Lightfoot et al. (2019) and stable quadruplexes with long loops that have been described (Guedin et al., 2010). Moreover, there is another GGG track in the proximity of our predicted sequence. On the other hand, analysis of the SARS-CoV-2 genome by QGRS algorithms showed 25 hits on the positive and 12 hits on the negative RNA strand (**Supplementary Material 10**). These hits are almost exclusively with two G-repeats only and have relatively low scores (maximum 19 for the QGRS algorithm and 1.111 for the G4Hunter algorithm). The best PQS suggested by the G4Hunter algorithm is located in the position 28,289–28,313 within the nucleocapsid phosphoprotein coding sequence. It is noteworthy that all viral RNAs are produced through negative-strand intermediates, which are only about 1% as abundant as their positive-sense counterparts (Fehr and Perlman, 2015). It would be of great value to know whether TMPyP4 or other G-quadruplex stabilizing compounds can inhibit replication processes of SARS-CoV-2. We have aligned three pathogenic human coronaviruses (SARS-CoV, SARS-CoV-2, and MERS-CoV) to see if there are differences in the SUD domain, which was earlier proposed to be G-quadruplex binding (Kusov et al., 2015). Comparison of three key amino acid residues involved in G-quadruplex binding revealed that the SUD domain of MERS-CoV lacks these residues. This correlates with the fact that no G-quadruplex was predicted that is within its genome. It has been demonstrated, however, that the RGG domain can play roles in various nucleic acid and protein interactions (Thandapani et al., 2013), so this correlation can be G4-independent for SARS-CoV-2. RNA hairpins, which are formed by IRs, are basic structural elements of RNA and play crucial roles in gene expression and intermolecular recognition. Conserved palindromic RNA structures have been found in many viral genomes, including HIV-1, and play a crucial role in their replication (Liu et al., 2018). We have found an abundance of IRs inside 5′UTR in *Nidovirales* genomes. In general, 5′UTR is an important locus for regulation of viral replication and gene expression. It has been demonstrated that stem integrity of phylogenetically conserved stem-loop structure located in 5′UTR of the PRRSV virus from the *Arteriviridae* family is crucial for viral replication and subgenomic mRNA synthesis. Similar secondary structures have been proposed for several viruses from the *Arteriviridae* and *Coronaviridae* families (Lu et al., 2011). Our analyses of annotated features in **Figure 5** therefore support this report. The discrete locations of specific IRs in viral genomes could therefore be additional targets for their regulation.

Significant differences of PQS and IR frequencies among various ssRNA viruses in *Nidovirales* groups show that their genome organization and regulation are not identical but that for some *Nidovirales* the presence of the G-quadruplexes most probably does not play an essential role in their biological regulation. Moreover, our analyses suggest that G-quadruplexes have been evolutionally eliminated in some genomes of *Nidovirales*. This is quite surprising, considering that G-quadruplexes have been found in all evolutionary groups, including such CG low-level organisms as *Saccharomyces cerevisiae* (Bartas et al., 2019; Brázda et al., 2019; Singh and Lakhanpaul, 2019; Gage and Merrick, 2020). In fact, it could be an evolutionary advantage not to present G-quadruplex in viral genome because a number of cellular proteins interact with G-quadruplexes (Brázda et al., 2014; Mishra et al., 2016) and therefore the presence of G-quadruplex in viral genome could serve as a structure recognized by the innate immune system (Unterholzner et al., 2010; Hároníková et al., 2016; Voter et al., 2020). Their RNAs are therefore not recognized as alien nucleic acids and are processed by the cellular machinery. On the other hand, presence of IRs in *Nidovirales* genomes constitutes an inseparable part of their genomes and allows their correct folding and structure-specific regulation of their functions (Lorenz et al., 2011; Dutkiewicz et al., 2016).

Targeting viral proteins is usually effective only against specific viral strains and fails even for closely related viral species. It appears that targeting host proteins should be able to provide a response toward a wider spectrum of viruses inasmuch as different viruses exploit common cellular pathways. Many cellular RNA-binding proteins (RBPs) containing well-established RNA-binding domains (RBDs) are known to be critical for infection by different viruses. Recently, 472 RBPs were reviewed for their linkage to viruses (Garcia-Moreno et al., 2018). It has been demonstrated that G-quadruplex formation in HIV-1 viral genome stalls RNA polymerase, thereby limiting

viral replication in host cells. HIV-1 nucleocapsid protein NCp7 helps to resolve G-quadruplex formation and therefore enables virus to spread. Stabilization of this quadruplex has been targeted by several experimental compounds. This treatment slowed or inhibited viral growth (Butovskaya et al., 2019).

Virus reproduction is dependent on the cellular transcription machinery, and therefore the interaction of the cellular proteins with viral RNA could be another target for antiviral therapy (Roberts et al., 2009). Our analysis predicted several QBPs to be capable to bind SARS-CoV-2 RNA. One of these, HNRNPA1 protein, is responsible for nuclear–cytoplasm shuttling (Garcia-Moreno et al., 2018). Like some other RNA-binding proteins, HNRNPA1 forms so-called membraneless organelles. These organelles are assemblies of proteins along with RNA or DNA that condense in specific cellular loci. The organelles can undergo transition from liquid-like droplets to amyloid fibrils, and mutations in so-called low complexity domains of these proteins lead to formation of amyloid aggregations that are found in many neurodegenerative diseases (Gui et al., 2019). HNRNPA1 is involved in many different cellular processes and has been targeted in various diseases. One such example is the varicella zoster virus that causes chicken pox. Moreover, this response to varicella zoster virus has been connected with autoimmune disease that complicates multiple sclerosis (Kattimani and Veerappa, 2018). HNRNPA1 is known to bind and resolve G-quadruplex formed in TRA2B promoter and to promote its transcription. Dysregulation of this binding leads to progression of colon cancer. This interaction has been targeted by the well-known G-quadruplex stabilizer pyridostatin, which led to decreased transcription from the TRA2B promoter (Nishikawa et al., 2019). Furthermore, HNRNPA1 is co-expressed with bromodomain and extraterminal domain protein BRD4 in human tumor samples. It has been shown experimentally that the well-described, naturally occurring polyphenolic flavonoid quercetin inhibited this protein and thereby led to better susceptibility to treatment in cancer patients (Pham et al., 2019). Both SRSF7 and TRA2A proteins, which are predicted to interact with SARS-CoV-2 RNA, play roles in alternative splicing (Ghosh et al., 2016). SRSF7 is a serine and arginine-rich splicing factor and is part of the spliceosome. Its expression has been connected to several types of cancer and it has been shown that its knockdown induced p21 expression and thus reduced cancer development (Saijo et al., 2016). Little is known about TRA2A protein. It was first identified in insects together with its paralog TRA2B, which has been studied to a greater extent (Tan et al., 2018). It has been found that TRA2A can promote paclitaxel therapy and promote cancer progression in triple-negative breast cancers (Liu et al., 2017). A role of TRA2A protein in regulation of HIV1 virus replication has been described. Both TRA2A and TRA2B bind to a specific HIV1 sequence and regulate its replication within the cell through alternative splicing of viral RNA (Erkelenz et al., 2013).

Scientists around the world are united in their efforts to find an effective therapy against coronavirus disease (COVID-19). Among the most promising candidates are remdesivir and chloroquine. Remdesivir is an adenosine analog that incorporates into nascent viral RNA chains and results in premature termination. Preliminary data have shown that remdesivir effectively inhibited virus infection in a human liver cancer cell line (Wang et al., 2020). Chloroquine is a potential broad-spectrum antiviral drug and it already has been widely used as a low-cost and safe anti-malarial and autoimmune disease drug for more than 70 years. Application of chloroquine causes elevation of endosomal pH and also interferes with terminal glycosylation of the cellular receptor ACE2. This probably has a negative influence on virus-receptor binding and abrogates the infection (Vincent et al., 2005). Drug repurposing seems like a very good strategy for quickly and at low cost finding a new therapy for new human diseases (Oprea and Mestres, 2012). To date, there are few substances with G-quadruplex stabilizing features. One example is topotecan (also known by its brand name Hycamtin®), which frequently is used for treating ovary cancer. If everything else fails, it seems to be an option. On the other hand, topotecan cause unpleasant side effects, such as nausea, vomiting, and diarrhea (Topotecan - Chemotherapy Drugs – Chemocare, n.d.). Many studies have described strong G-quadruplex stabilization effects (Li et al., 2018; Satpathi et al., 2018), which might be one possible mode of action. Moreover, it has been proven that higher structures of nucleic acids, and G-quadruplex especially, might be stabilized by use of various natural substances. Berbamine is one such substance and is a component of traditional Chinese medicine. It frequently is used for treating chronic myeloid leukemia or melanoma and has strong binding affinity to G-quadruplex structures (Tan et al., 2014). Viral nucleic acids and their loci with G-quadruplex-forming potential are in all cases very specific and are promising molecular targets for treating serious diseases. Evidence of G-quadruplex formation as a potential target for therapy was proven for Hepatitis A, flu virus, and HIV-1 (Métifiot et al., 2014). Because all the aforementioned cases concern RNA viruses, the viral G-quadruplexes are localized in the cytoplasm of the host cell. Our comparative analyses of IRs and PQSs in *Nidovirales* show that, in contrast to IRs, which are presented abundantly in all *Nidovirales* genomes, the sequences able to form G-quadruplex structures are very unequally distributed and are very rare especially in *Nidovirales* species capable to infect humans. This suggests intentional suppression during evolution in order to simplify viral RNA replication. Finding the proper stabilizers of viral higher RNA structures might be crucial for inhibiting or stopping viral RNA replication in order to gain time for the immune system to deal successfully with an infection.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MB and VB contributed to the conceptualization, formal analysis, and methodology. NB, AC, AV, and TS helped with the data curation. PP was involved in with funding acquisition

and project administration. MB, VB, AV, and JČ carried out the investigation. MB, NB, OP, VB, and EJ helped with the resources. VB, KM, and PP supervised the study. VB and EJ validated the study. MB, NB, and AC worked on the visualization. MB, VB, NB, OP, and JČ wrote the original draft. VB, JČ, KM, and PP reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic in the National Feasibility Program I (LO1208 TEWEP); by the EU structural funding Operational Programme Research and Development for innovation, project No. CZ.1.05/2.1.00/19.0388; and by the projects SGS/01/PrF/2020 and SGS/07/PrF/2020 financed by University of Ostrava.

ACKNOWLEDGMENTS

We would like to express our gratitude to Alena Volná, M.Sc. for her time spent in preparing the draft illustrations accompanying this work. **Figure 1** is reproduced with permission from Biolution GmbH.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01583/full#supplementary-material

REFERENCES

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucl. Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379

Bagga, R., Ramesh, N., and Brahmachari, S. K. (1990). Supercoil-induced unusual DNA structures as transcriptional block. *Nucl. Acids Res.* 18, 3363–3369. doi: 10.1093/nar/18.11.3363

Bartas, M., Cutová, M., Brázda, V., Kaura, P., Šťastný, J., Kolomazník, J., et al. (2019). The presence and localization of G-quadruplex forming sequences in the domain of bacteria. *Molecules* 24:1711. doi: 10.3390/molecules24091711

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

Brázda, V., Cerven, J., Bartas, M., Mikysková, N., Coufal, J., Pečinka, P., et al. (2018). The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors. *Molecules* 23:2341. doi: 10.3390/molecules23092341

Brázda, V., Hároníková, L., Liao, J. C., and Fojta, M. (2014). DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.* 15, 17493–17517. doi: 10.3390/ijms151017493

Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Šťastný, J., et al. (2019). G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* 35, 3493–3495. doi: 10.1093/bioinformatics/btz087

Brázda, V., Kolomazník, J., Lýsek, J., Hároníková, L., Coufal, J., and 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. doi: 10.1016/j.bbrc.2016.09.015

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

Bridges, R., Correia, S., Wegner, F., Venturini, C., Palser, A., White, R. E., et al. (2019). Essential role of inverted repeat in Epstein–Barr virus IR-1 in B cell transformation; geographical variation of the viral genome. *Philos. Trans. R. Soc. B* 374:20180299. doi: 10.1098/rstb.2018.0299

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucl. Acids Res.* 34, 5402–5415. doi: 10.1093/nar/gkl655

Butovskaya, E., Solda, P., Scalabrin, M., Nadai, M., and Richter, S. N. (2019). HIV-1 nucleocapsid protein unfolds stable RNA G-quadruplexes in the viral genome and is inhibited by G-quadruplex ligands. *ACS Infect. Dis.* 5, 2127–2135. doi: 10.1021/acsinfecdis.9b00272

Carrasco-Hernandez, R., Jácome, R., López Vidal, Y., and Ponce de León, S. (2017). Are RNA viruses candidate agents for the next global pandemic? A review. *ILAR J.* 58, 343–358. doi: 10.1093/ilar/ilx026

Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33:877. doi: 10.1038/nbt.3295

Chang, C., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., and Huang, T. (2014). The SARS coronavirus nucleocapsid protein – Forms and functions. *Antiviral Res.* 103, 39–50. doi: 10.1016/j.antiviral.2013.12.009

Chen, N., Li, X., Li, S., Xiao, Y., Ye, M., Yan, X., et al. (2020). How related is SARS-CoV-2 to other coronaviruses? *Vet. Rec.* 186, 496–496. doi: 10.1136/vr.m1452

Cohen, J. (2020). New coronavirus threat galvanizes scientists. *Science* 367, 492–493. doi: 10.1126/science.367.6477.492

Cowling, B. J., Park, M., Fang, V. J., Wu, P., Leung, G. M., and Wu, J. T. (2015). Preliminary epidemiologic assessment of MERS-CoV outbreak in South Korea, May–June 2015. *Euro Surveill.* 20:21163. doi: 10.2807/1560-7917.ES2015.20.25.21163

Drake, J. W., and Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S.A.* 96, 13910–13913. doi: 10.1073/pnas.96.24.13910

Dutkiewicz, M., Stachowiak, A., Swiatkowska, A., and Ciesiołka, J. (2016). Structure and function of RNA elements present in enteroviral genomes. *Acta Biochim. Polonica* 63, 623–630. doi: 10.18388/abp.2016_1337

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Eigen, M., and Schuster, P. (1977). A principle of natural self-organization: Part A: emergence of the hypercycle. *Naturwissenschaften* 64, 541–565. doi: 10.1007/BF00450633

Erkelenz, S., Poschmann, G., Theiss, S., Stefanski, A., Hillebrand, F., Otte, M., et al. (2013). Tra2-mediated recognition of HIV-1 5 ' splice site D3 as a key factor in the processing of vpr mRNA. *J. Virol.* 87, 2721–2734. doi: 10.1128/JVI.02756-12

Federhen, S. (2011). The NCBI taxonomy database. *Nucl. Acids Res.* 40, D136–D143. doi: 10.1093/nar/gkr1178

Fehr, A. R., and Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* 1282, 1–23. doi: 10.1007/978-1-4939-2438-7_1

Fleming, A. M., Ding, Y., Alenko, A., and Burrows, C. J. (2016). Zika virus genomic RNA possesses conserved G-quadruplexes characteristic of the flaviviridae family. *ACS Infect. Dis.* 2, 674–681. doi: 10.1021/acsinfecdis.6b00109

Frasson, I., Nadai, M., and Richter, S. N. (2019). Conserved G-quadruplexes regulate the immediate early promoters of human Alphaherpesviruses. *Molecules* 24:2375. doi: 10.3390/molecules24132375

Gage, H. L., and Merrick, C. J. (2020). Conserved associations between G-quadruplex-forming DNA motifs and virulence gene families in malaria parasites. *BMC Genomics* 21:236. doi: 10.1186/s12864-020-6625-x

Garant, J.-M., Perreault, J.-P., and Scott, M. S. (2018). G4RNA screener web server: user focused interface for RNA G-quadruplex prediction. *Biochimie* 151, 115–118. doi: 10.1016/j.biochi.2018.06.002

Garcia-Moreno, M., Jaervelin, A. I., and Castello, A. (2018). Unconventional RNA-binding proteins step into the virus-host battlefront. *Wiley Interdiscipl. Rev. RNA* 9:e1498. doi: 10.1002/wrna.1498

Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M., and Nik-Zainal, S. (2018). Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome Res.* 28, 1264–1271. doi: 10.1101/gr.231688.117

Ghosh, P., Grellscheid, S. N., and Sowdhamini, R. (2016). A tale of two paralogs: human Transformer2 proteins with differential RNA-binding affinities. *J. Biomol. Struct. Dyn.* 34, 1979–1986. doi: 10.1080/07391102.2015.1100551

Gorbalenya, A. E. (2001). Big nidovirus genome. When count and order of domains matter. *Adv. Exp. Med. Biol.* 494, 1–17. doi: 10.1007/978-1-4615-1325-4_1

Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., and Snijder, E. J. (2006). Nidovirales: evolving the largest RNA virus genome. *Virus Res.* 117, 17–37. doi: 10.1016/j.virusres.2006.01.017

Guedin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. *Nucl. Acids Res.* 38, 7858–7868. doi: 10.1093/nar/gkq639

Gui, X., Luo, F., Li, Y., Zhou, H., Qin, Z., Liu, Z., et al. (2019). Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* 10:2006. doi: 10.1038/s41467-019-09902-7

Hanada, K., Suzuki, Y., Nakane, T., Hirose, O., and Gojobori, T. (2005). The origin and evolution of porcine reproductive and respiratory syndrome viruses. *Mol. Biol. Evol.* 22, 1024–1031. doi: 10.1093/molbev/msi089

Hänsel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., et al. (2016). G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* 48:1267. doi: 10.1038/ng.3662

Hároníková, L., Coufal, J., Kejnovská, I., Jagelská, E. B., Fojta, M., Dvoráková, P., et al. (2016). IFI16 preferentially binds to DNA with quadruplex structure and enhances DNA quadruplex formation. *PLoS ONE* 11:e0157156. doi: 10.1371/journal.pone.0157156

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 181, 271–280.e8 doi: 10.1016/j.cell.2020.02.052

Huang, Z.-L., Dai, J., Luo, W.-H., Wang, X.-G., Tan, J.-H., Chen, S.-B., et al. (2018). Identification of G-quadruplex-binding protein from the exploration of RGG Motif/G-quadruplex interactions. *J. Am. Chem. Soc.* 140, 17945–17955. doi: 10.1021/jacs.8b09329

Hung, L. S. (2003). The SARS epidemic in Hong Kong: what lessons have we learned? *J. R. Soc. Med.* 96, 374–378. doi: 10.1258/jrsm.96.8.374

Ishimaru, D., Plant, E. P., Sims, A. C., Yount, B. L. Jr., Roth, B. M., Eldho, N. V., et al. (2013). RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucl. Acids Res.* 41, 2594–2608. doi: 10.1093/nar/gks1361

Jaubert, C., Bedrat, A., Bartolucci, L., Di Primo, C., Ventura, M., Mergny, J.-L., et al. (2018). RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-018-26582-3

Kattimani, Y., and Veerappa, A. M. (2018). Complex interaction between mutant HNRNPA1 and gE of varicella zoster virus in pathogenesis of multiple sclerosis. *Autoimmunity* 51, 147–151. doi: 10.1080/08916934.2018.1482883

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

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *bioRxiv.* 181, 914–921.e10. doi: 10.1016/j.cell.2020.04.011

Kocman, V., and Plavec, J. (2017). Tetrahelical structural family adopted by AGCGA-rich regulatory DNA regions. *Nat. Commun.* 8:15355. doi: 10.1038/ncomms15355

Kolesnikova, S., and Curtis, E. A. (2019). Structure and function of multimeric G-quadruplexes. *Molecules* 24:3074. doi: 10.3390/molecules24173074

Krafčíková, P., Demkovičová, E., and Víglaský, V. (2017). Ebola virus derived G-quadruplexes: thiazole orange interaction. *Biochim. Biophys. Acta General Subj.* 1861, 1321–1328. doi: 10.1016/j.bbagen.2016.12.009

Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., and Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the "SARS-unique domain" is essential for the activity of the SARS-coronavirus replication–transcription complex. *Virology* 484, 313–322. doi: 10.1016/j.virol.2015.06.016

Lai, M. M. C., and Cavanagh, D. (1997). "The molecular biology of coronaviruses," in *Advances in Virus Research* (Elsevier), 1–100. Available online at: https://linkinghub.elsevier.com/retrieve/pii/S0065352708602869

Lauring, A. S., Frydman, J., and Andino, R. (2013). The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microbiol.* 11, 327–336. doi: 10.1038/nrmicro3003

Lavezzo, E., Berselli, M., Frasson, I., Perrone, R., Palù, G., Brazzale, A. R., et al. (2018). G-quadruplex forming sequences in the genome of all known human viruses: a comprehensive guide. *PLOS Comput. Biol.* 14:e1006675. doi: 10.1371/journal.pcbi.1006675

Lei, J., Kusov, Y., and Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res.* 149, 58–74. doi: 10.1016/j.antiviral.2017.11.001

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucl. Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, F., Zhou, J., Xu, M., and Yuan, G. (2018). Exploration of G-quadruplex function in c-Myb gene and its transcriptional regulation by topotecan. *Int. J. Biol. Macromol.* 107, 1474–1479. doi: 10.1016/j.ijbiomac.2017.10.010

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N. Engl. J. Med*. 382, 1199–1207. doi: 10.1056/NEJMoa2001316

Li, R.-F., and Li, H. (2010). Study on the influences of palindromes in protein coding sequences on the folding rates of peptide chains. *Protein Peptide Lett.* 17, 881–888. doi: 10.2174/092986610791306652

Lightfoot, H. L., Hagen, T., Tatum, N. J., and Hall, J. (2019). The diverse structural landscape of quadruplexes. *FEBS Lett.* 593, 2083–2102. doi: 10.1002/1873-3468.13547

Liò, P., and Goldman, N. (2004). Phylogenomics and bioinformatics of SARS-CoV. *Trends Microbiol.* 12, 106–111. doi: 10.1016/j.tim.2004.01.005

Liu, T., Sun, H., Zhu, D., Dong, X., Liu, F., Liang, X., et al. (2017). TRA2A promoted paclitaxel resistance and tumor progression in triple-negative breast cancers via regulating alternative splicing. *Mol. Cancer Therap.* 16, 1377–1388. doi: 10.1158/1535-7163.MCT-17-0026

Liu, Y., Chen, J., Nikolaitchik, O. A., Desimmie, B. A., Busan, S., Pathak, V. K., et al. (2018). The roles of five conserved lentiviral RNA structures in HIV-1 replication. *Virology* 514, 1–8. doi: 10.1016/j.virol.2017.10.020

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA package 2.0. *Algorith. Mol. Biol.* 6:26. doi: 10.1186/1748-7188-6-26

Lu, J., Gao, F., Wei, Z., Liu, P., Liu, C., Zheng, H., et al. (2011). A 5'-proximal stem-loop structure of 5' untranslated region of porcine reproductive and respiratory syndrome virus genome is key for virus replication. *Virol. J.* 8:172. doi: 10.1186/1743-422X-8-172

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8

Masuzawa, T., and Oyoshi, T. (2020). Roles of the RGG domain and RNA recognition motif of nucleolin in G-quadruplex stabilization. *ACS Omega*. 5, 5202–5208. doi: 10.1021/acsomega.9b04221

Métifiot, M., Amrane, S., Litvak, S., and Andreola, M.-L. (2014). G-quadruplexes in viruses: function and potential therapeutic applications. *Nucl. Acids Res.* 42, 12352–12366. doi: 10.1093/nar/gku999

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

Modrow, S., Falke, D., Truyen, U., and Schätzl, H. (2013). "Viruses with single-stranded, positive-sense RNA genomes," in *Molecular Virology* (Berlin: Springer Berlin Heidelberg), 185–349. Available online at: http://link.springer.com/10.1007/978-3-642-20718-1_14

Nishikawa, T., Kuwano, Y., Takahara, Y., Nishida, K., and Rokutan, K. (2019). HnRNPA1 interacts with G-quadruplex in the TRA2B promoter and stimulates its transcription in human colon cancer cells. *Sci. Rep.* 9:10276. doi: 10.1038/s41598-019-46659-x

Oboho, I. K., Tomczyk, S. M., Al-Asmari, A. M., Banjar, A. A., Al-Mugti, H., Aloraini, M. S., et al. (2015). 2014 MERS-CoV outbreak in Jeddah—a link to health care facilities. *N. Engl. J. Med.* 372, 846–854. doi: 10.1056/NEJMoa1408636

Okonechnikov, K., Golosova, O., Fursov, M., and Team, U. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091

Oprea, T. I., and Mestres, J. (2012). Drug repurposing: far beyond new targets for old drugs. *AAPS J.* 14, 759–763. doi: 10.1208/s12248-012-9390-1

Patino-Galindo, J. A., Filip, I., AlQuraishi, M., and Rabadan, R. (2020). Recombination and lineage-specific mutations led to the emergence of SARS-CoV-2. *bioRxiv [Preprint]*. doi: 10.1101/2020.02.10.942748

Paz, I., Kosti, I., Ares, M. Jr., Cline, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucl. Acids Res.* 42, W361–W367. doi: 10.1093/nar/gku406

Pham, T. N. D., Stempel, S., Shields, M. A., Spaulding, C., Kumar, K., Bentrem, D. J., et al. (2019). Quercetin enhances the anti-tumor effects of BET inhibitors by suppressing hnRNPA1. *Int. J. Mol. Sci.* 20:4293. doi: 10.3390/ijms20174293

Plant, E., Perez-Alvarado, G., Jacobs, J., Mukhopadhyay, B., Hennig, M., and Dinman, J. (2005). A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol.* 3, 1012–1023. doi: 10.1371/journal.pbio.0030172

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

Puig Lombardi, E., and Londoño-Vallejo, A. (2020). A guide to computational methods for G-quadruplex prediction. *Nucl. Acids Res.* 48, 1–15. doi: 10.1093/nar/gkaa033

Roberts, L. O., Jopling, C. L., Jackson, R. J., and Willis, A. E. (2009). "Chapter 9 viral strategies to subvert the mammalian translation machinery," in *Progress in Molecular Biology and Translational Science, Translational Control in Health and Disease*. (Academic Press), 313–367. Available online at: http://www.sciencedirect.com/science/article/pii/S1877117309900096

Ruggiero, E., and Richter, S. N. (2018). G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucl. Acids Res.* 46, 3270–3283. doi: 10.1093/nar/gky187

Ruggiero, E., and Richter, S. N. (2020). Viral G-quadruplexes: new frontiers in virus pathogenesis and antiviral therapy. *Annu. Rep. Med. Chem.* doi: 10.1016/bs.armc.2020.04.001. [Epub ahead of print].

Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A., and Gorbalenya, A. E. (2018). A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog.* 14:e1007314. doi: 10.1371/journal.ppat.1007314

Saijo, S., Kuwano, Y., Masuda, K., Nishikawa, T., Rokutan, K., and Nishida, K. (2016). Serine/arginine-rich splicing factor 7 regulates p21-dependent growth arrest in colon cancer cells. *J. Med. Invest.* 63, 219–226. doi: 10.2152/jmi.63.219

Satpathi, S., Singh, R. K., Mukherjee, A., and Hazra, P. (2018). Controlling anticancer drug mediated G-quadruplex formation and stabilization by a molecular container. *Phys. Chem. Chem. Phys.* 20, 7808–7818. doi: 10.1039/C8CP00325D

Sinden, R., Zheng, G., Brankamp, R., and Allen, K. (1991). On the deletion of inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform formation *in viv*o. *Genetics* 129, 991–1005.

Singh, A., and Lakhanpaul, S. (2019). Genome-wide analysis of putative G-quadruplex sequences (PGQSs) in onion yellows phytoplasma (strain OY-M): an emerging plant pathogenic bacteria. *Indian J. Microbiol.* 59, 468–475. doi: 10.1007/s12088-019-00831-z

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L. M., et al. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331, 991–1004. doi: 10.1016/S0022-2836(03)00865-9

Surovtsev, I. V., and Jacobs-Wagner, C. (2018). Subcellular organization: a critical feature of bacterial cell replication. *Cell* 172, 1271–1293. doi: 10.1016/j.cell.2018.01.014

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucl. Acids Res.* 43, D447–52. doi: 10.1093/nar/gku1003

Tan, W., Zhou, J., and Yuan, G. (2014). Electrospray ionization mass spectrometry probing of binding affinity of berbamine, a flexible cyclic alkaloid from traditional Chinese medicine, with G-quadruplex DNA. *Rapid Commun. Mass Spectr.* 28, 143–147. doi: 10.1002/rcm.6763

Tan, Y., Hu, X., Deng, Y., Yuan, P., Xie, Y., and Wang, J. (2018). TRA2A promotes proliferation, migration, invasion and epithelial mesenchymal transition of glioma cells. *Brain Res. Bull.* 143, 138–144. doi: 10.1016/j.brainresbull.2018.10.006

Thandapani, P., O'Connor, T. R., Bailey, T. L., and Richard, S. (2013). Defining the RGG/RG motif. *Mol. Cell* 50, 613–623. doi: 10.1016/j.molcel.2013.05.021

Topotecan - Chemotherapy Drugs – Chemocare. Available online at: http://chemocare.com/chemotherapy/drug-info/Topotecan.aspx

Travers, A., and Muskhelishvili, G. (2015). DNA structure and function. *FEBS J.* 2279–2295. doi: 10.1111/febs.13307

Unterholzner, L., Keating, S. E., Baran, M., Horan, K. A., Jensen, S. B., Sharma, S., et al. (2010). IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.* 11, 997–1004. doi: 10.1038/ni.1932

Vincent, M. J., Bergeron, E., Benjannet, S., Erickson, B. R., Rollin, P. E., Ksiazek, T. G., et al. (2005). Chloroquine is a potent inhibitor of SARS coronavirus infection and spread. *Virol. J.* 2:69. doi: 10.1186/1743-422X-2-69

Vorlíčková, M., Kejnovská, I., Bednárová, K., Renčiuk, D., and Kypr, J. (2012). Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality* 24, 691–698. doi: 10.1002/chir.22064

Voter, A. F., Callaghan, M. M., Tippana, R., Myong, S., Dillard, J. P., and Keck, J. L. (2020). Antigenic variation in neisseria gonorrhoeae occurs independently of RecQ-mediated unwinding of the pilE G quadruplex. *J. Bacteriol.* 202:e00607-19. doi: 10.1128/JB.00607-19

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., et al. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. *Cell Res.* 30, 269–271. doi: 10.1038/s41422-020-0282-0

Wang, S.-K., Wu, Y., and Ou, T.-M. (2015). RNA G-quadruplex: the new potential targets for therapy. *Curr. Topics Med. Chem.* 15, 1947–1956. doi: 10.2174/1568026615666150515145733

Wang, S.-R., Zhang, Q.-Y., Wang, J.-Q., Ge, X.-Y., Song, Y.-Y., Wang, Y.-F., et al. (2016). Chemical targeting of a G-quadruplex RNA in the Ebola virus L gene. *Cell Chem. Biol.* 23, 1113–1122. doi: 10.1016/j.chembiol.2016.07.019

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3

Wu, Y., Shin-ya, K., and Brosh, R. M. Jr. (2008). FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.* 28, 4116–4128. doi: 10.1128/MCB.02210-07

Xie, J., Mao, Q., Tai, P. W., He, R., Ai, J., Su, Q., et al. (2017). Short DNA hairpins compromise recombinant adeno-associated virus genome homogeneity. *Mol. Ther.* 25, 1363–1374. doi: 10.1016/j.ymthe.2017.03.028

Yu, L. O. (2009). Bioinformatic analysis of inverted repeats of coronaviruses genome. *Biopolym. Cell* 25, 307–314. doi: 10.7124/bc.0007EA

Ziebuhr, J. (2004). Molecular biology of severe acute respiratory syndrome coronavirus. *Curr. Opin. Microbiol.* 7, 412–419. doi: 10.1016/j.mib.2004.06.007

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

biomolecules

MDPI

Article

G-Quadruplexes in the Archaea Domain

Václav Brázda [1,*], Yu Luo [2], Martin Bartas [3], Patrik Kaura [4], Otilia Porubiaková [1,5], Jiří Šťastný [4,6], Petr Pečinka [3], Daniela Verga [2], Violette Da Cunha [7], Tomio S. Takahashi [7], Patrick Forterre [7], Hannu Myllykallio [8], Miroslav Fojta [1] and Jean-Louis Mergny [1,8,*]

1 Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic; o.porubiakova@gmail.com (O.P.); fojta@ibp.cz (M.F.)
2 Institut Curie, CNRS UMR9187, INSERM U1196, Universite Paris Saclay, 91400 Orsay, France; yu.luo@curie.fr (Y.L.); Daniela.Verga@curie.fr (D.V.)
3 Department of Biology and Ecology/Institute of Environmental Technologies, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic; dutartas@gmail.com (M.B.); petr.pecinka@osu.cz (P.P.)
4 Faculty of Mechanical Engineering, Brno University of Technology, Technicka 2896/2, 616 69 Brno, Czech Republic; 160702@vutbr.cz (P.K.); stastny@fme.vutbr.cz (J.Š.)
5 Faculty of Chemistry, Brno University of Technology, Purkyňova 464/118, 612 00 Brno, Czech Republic
6 Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
7 Institut de Biologie Intégrative de la Cellule (I2BC), CNRS, Université Paris-Saclay, CEDEX, 91198 Gif-sur-Yvette, France; violette.da.cunha.vdc@gmail.com (V.D.C.); tomio.takahashi@i2bc.paris-saclay.fr (T.S.T.); patrick.forterre@pasteur.fr (P.F.)
8 Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, Institut Polytechnique de Paris, 91128 Palaiseau, France; hannu.myllykallio@polytechnique.edu
* Correspondence: vaclav@ibp.cz (V.B.); jean-louis.mergny@inserm.fr (J.-L.M.); Tel.: +42-05-4151-7231 (V.B.); Fax: +42-05-4121-1293 (V.B.)

Received: 11 August 2020; Accepted: 18 September 2020; Published: 21 September 2020

Abstract: The importance of unusual DNA structures in the regulation of basic cellular processes is an emerging field of research. Amongst local non-B DNA structures, G-quadruplexes (G4s) have gained in popularity during the last decade, and their presence and functional relevance at the DNA and RNA level has been demonstrated in a number of viral, bacterial, and eukaryotic genomes, including humans. Here, we performed the first systematic search of G4-forming sequences in all archaeal genomes available in the NCBI database. In this article, we investigate the presence and locations of G-quadruplex forming sequences using the G4Hunter algorithm. G-quadruplex-prone sequences were identified in all archaeal species, with highly significant differences in frequency, from 0.037 to 15.31 potential quadruplex sequences per kb. While G4 forming sequences were extremely abundant in *Hadesarchaea archeon* (strikingly, more than 50% of the *Hadesarchaea archaeon* isolate WYZ-LMO6 genome is a potential part of a G4-motif), they were very rare in the *Parvarchaeota* phylum. The presence of G-quadruplex forming sequences does not follow a random distribution with an over-representation in non-coding RNA, suggesting possible roles for ncRNA regulation. These data illustrate the unique and non-random localization of G-quadruplexes in Archaea.

Keywords: G4-forming motif; genome analysis; Archaea; unusual nucleic acid structures; sequence prediction

1. Introduction

The Archaea domain was classified separately from Bacteria by Carl Woese and George Fox in 1977 [1]. Later on, it was found that all major molecular machinery, such as DNA replication, transcription, and translation, of archaea are much more similar to those of eukaryotes than to those of

bacteria [2,3]. This is also true for some important membrane proteins, such as ATP synthases and proteins of the Sec transport system [4,5], or for some proteins involved in cell division and vesicle trafficking [6]. Thus, the archaeal domain occupies a key position in the Tree of Life, and there is currently a hot debate about their exact relationships with eukaryotes [7,8]. A schematic phylogenic tree for the Archaea domain is proposed in Figure 1; this phylogeny is rapidly evolving with many new phyla recently identified via the accumulation of metagenome associated genomes (MAGs) and various new proposals for phylum definition and nomenclature [9,10]. The first detected archaea were isolated in harsh environments but later found in almost every environment, including the human microbiota, where they play important roles in the gut, mouth, and on the skin [11,12]. It has been hypothesized that archaea found in oceans are one of the most abundant groups of organisms on the planet with important roles both in the carbon and the nitrogen cycle [13]. The Archaea domain has several unique features, such as *ether*-linked lipids, while eukaryotes and most of the bacteria have ester-linked lipids [14]. Moreover, the stereochemistry of archaeal lipids has the opposite configuration as compare to the ones of eukaryotic and bacterial origin. Interestingly, methanogenesis, the production of greenhouse methane gas as a metabolic by-product, occurs only in the archaeal domain [15,16].

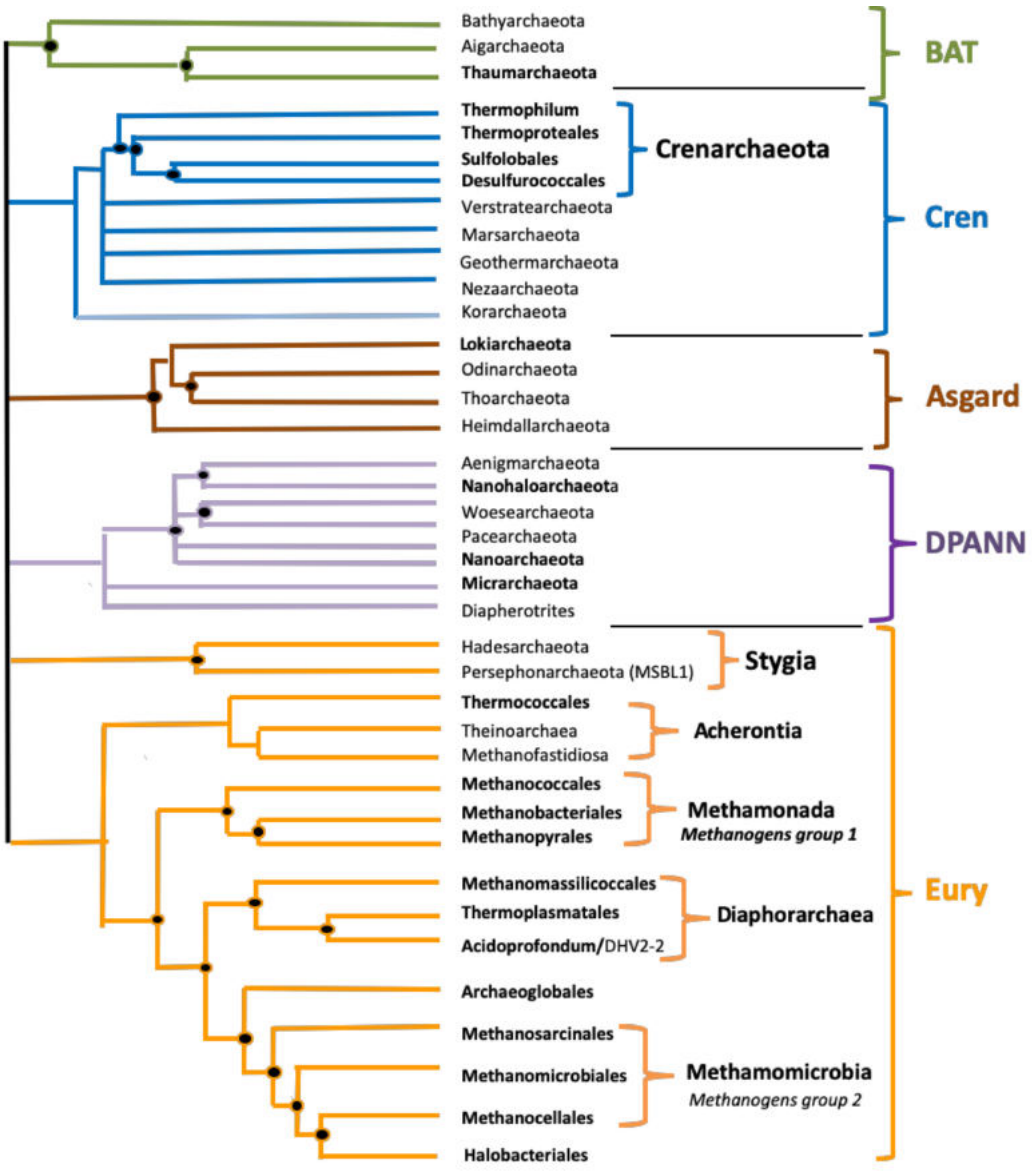


Figure 1. A schematic phylogenic tree for Archaea. This unrooted evolutionary tree of Archaea is based

on the schematic tree of Forterre (2015) [17] updated according to recent phylogenetic analyses [9,18]. BAT stands for Bathyarchaeota, Aigarchaeota, and Thaumarchaeota. DPANN is an acronym based on the first five groups discovered: *Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanoarchaeota*, and *Nanohaloarchaeota*. The term BAT superphylum has been proposed by Gaia et al. in 2018 [19], and the terms Eury and Cren superphyla are suggested here. The terms Cren superphylum is suggested here because the phyla *Crenarchaeota*, *Verstratearchaeota Marsarchaeota*, *Nezaarchaeota*, and Geothermarchaeota form a consensus monophyletic clade in all archaeal phylogeny. We included *Korarchaeota* in this superphylum because they often branch as sister groups of the above phyla in archaeal phylogenies, although the fast evolutionary rate made their positioning sometimes difficult. We suggested in parallel the term Eury superphylum because Euryarchaeota includes very diverse groups of cultivated and uncultivated Archaea which are difficult to the group in a single phylum, especially considering that phyla, such as *Verstratearchaeota Marsarchaeota*, or *Nezaarchaeota* only contain few uncultivated species only defined by a few metagenome associated genomes (MAGs). Names in bold letters correspond to subgroups that include cultivated species; names in thin letters correspond to subgroups that include only MAGs.

G-quadruplex structures (G4) formed by guanine rich sequences are among the most intensively studied local DNA/RNA structures [20]. G4s are formed by G:G Hoogsteen base pairing in a guanine quartet, and their formation requires the presence of stabilizing cations, such as potassium [21] (Figure 2). In both bacteria and eukaryotes, G4 formation regulates various processes, including gene expression [22], protein translation [23], and proteolysis [24]. G4 have been identified in a number of pathogens, including viruses, eukaryotes (e.g., *Plasmodium falciparum*) [25,26] or prokaryotes (e.g., *Neissseria gonorrhoeae* [27], and *Mycobacterium tuberculosis*) [28,29]. Moreover, many G4-binding proteins are conserved in all organisms highlighting the importance of the G4 structure regulations [30], and novel G4 binding proteins have been identified, sharing the NIQI amino acid motif (RGRGRRGGGSGGSGGRGRG) [31]. Specific helicases have been identified both in eukaryotes and bacteria to unfold these structures, which can be extremely stable and would be problematic for the transcription or replication of G-rich motifs (e.g., the Pif1 or RecQ family helicases) [32]. Recently, G4Hunter was successfully used for the prediction of G-quadruplex-forming sequences in all complete bacterial genomes [33]. These results showed that G-quadruplex-forming sequences are present in all species with the highest frequencies in some extremophiles. In contrast to RNA, there is no correlation between genomic DNA GC% in Archaea (and in Bacteria) and the optimal growth temperature. This is likely because DNA in vivo is topologically closed, and topologically closed DNA is stable at least up to 107 °C [34]. We therefore cannot anticipate a higher density of G4-prone motifs in thermophiles, due to a GC-bias. A comparison with Extremophiles in bacteria is interesting [35]. Ding et al. hypothesized that stress-resistant bacteria found in the Deinococcales may utilize putative quadruplex sequences (PQS) for gene regulatory purposes. An enrichment in prokaryote PQS has been found in thermophilic organisms [33] but also in organisms with resistance to other stress factors, such as radiation [36,37]; thus, a direct correlation between temperature and G4 presence is not supported by these findings. In addition, while bacteria in the Deinococcus-Thermus group are the most abundant for PQS, it is striking that the mostly thermophilic and hyperthermophilic bacteria in the Thermotogae phylum have one of the lowest PQS frequencies. Correlation among thermophiles and G4s, therefore, depends on the phylum (Gram-negative vs. Gram-positive bacteria).

Due to the roles of G4s in the regulation of basic cellular processes, it is important to identify their location in genomes. Several algorithms are available to predict G-quadruplex-forming sequences [38–41]. Among them, the G4Hunter application was developed to provide quantitative analyses giving a propensity score as an output [41], and the G4Hunter web tool allows effective and fast analyses of PQS in large datasets [42].

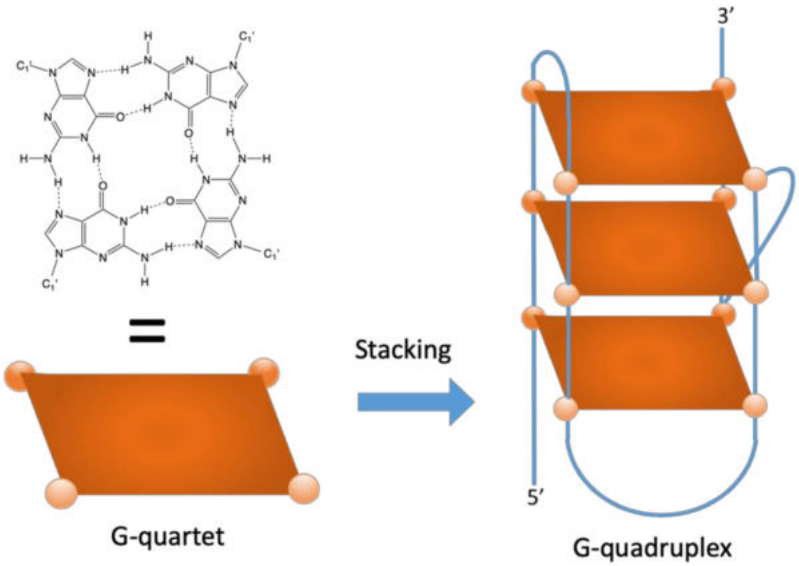


Figure 2. A G-quartet involves four coplanar guanines establishing a cyclic array of H-bonds (left). Stacking of two or more (three in this example) quartets leads to the formation of a G-quadruplex structure (right), stabilized by cations, such as potassium (not shown).

The prokaryotic genetic material is generally stored in circular chromosomes and plasmids [43]. The presence of quadruplex-prone motifs in over a hundred of bacterial genomes was determined over a decade ago [44]. In bacterial genomes, PQS are located non-randomly with a higher relative abundance in non-coding RNA (ncRNA), mRNA, and regions around tRNA and regulatory sequences. PQS also play roles in nitrate assimilation in *Paracoccus denitrificans* [45]. PQS in the hsdS, recD, and pmrA genes of *Streptococcus pneumoniae* contributes to host–pathogen interactions [46]. Such observations show the significant role of G4 in bacteria. The importance of another local DNA structure, the cruciform formed by inverted repeats, has been shown as an important regulatory feature of eukaryotic cell organelles, such as chloroplasts and mitochondria with circular DNA genomes [47,48]. Overall, the role of G4s in bacteria [27,49] and eukaryotes [50] is increasingly recognized.

In contrast, little is currently known regarding the abundancy and location of PQS in the archaeal domain. Ding et al. performed an initial search on bacterial and archaeal genomes using a modified Quadparser algorithm with relaxed parameters allowing long loops (up to 12 nucleotides) [35]. They found that thermophilic microorganisms (both archaea and bacteria) appear to favor PQS in their genomes. Dhapola et al. created the Quadbase2 web server, in which G4 motifs found in a variety of organisms, including archaea, may be searched but did not analyze G4 propensity in archaea [51]. Because G4s play many important biological roles in bacterial and eukaryotic cells, we assume that G4s are also likely to have important functions in archaea. Therefore, we comprehensively analyzed the presence and locations of PQS in all sequenced archaeal genomes by G4Hunter [41,42]. These data provide the first study analyzing the presence of G4-prone sequences in this important domain of life.

2. Materials and Methods

2.1. Selection of the DNA Sequences

The set of all archaeal genomic DNA sequences was downloaded from the Genome database of the National Center for Biotechnology Information [52]. We have used for our analyses all accessible archaeal genomes, including contig and scaffold sequences (3387 genomes), and we have selected one representative genome for each species (Supplementary Table S1). For PQS analyses of features, we restricted our analysis to the subset of 140 completely assembled genomes. In total, we have

analyzed the presence of G4 forming sequences in 3387 genomes from the archaeal Domain representing a total of 6423 Mbps.

2.2. Process of Analysis

We used the computational core of our DNA analyzer software written in Java programming language [53]. For our analyses, we used a new G4Hunter algorithm implementation [42]. Default parameters for G4Hunter were set to "25" for window size and 1.2 or above for the G4H score (G4HS). PQS score was grouped to the five intervals: 1.2–1.4, 1.4–1.6, 1.6–1.8,1.8–2.0, and 2.0 and more. Overall results for each species group contained a list of species with size of its genomic DNA sequence and number of putative G4 sequences found (Supplementary Table S2A); for clarity, the results for Groups and Subgroups are in separate files (Supplementary Table S2B,C). These data were processed by python jupyter using pandas with statistical tools [54]. Graphs were generated from the pandas tables using the "seaborn" graphical library. Note that the distinction between overlapping or discrete (non-overlapping) G4 motifs may create issues in the way potential motifs are counted. For this reason, we also provide a % PQS factor, which corresponds to the probability that any given nucleotide in the group or subgroup belongs to a G4-prone region (G4H > 1.2).

The default window value for G4Hunter has been discussed and tested in previous publications [41]. The value is chosen here (25 nt) corresponds more or less to the size of a typical intramolecular quadruplex. We considered shorter windows (20 nt) in previous studies. However, we noticed that for low thresholds (<1.2), a single GGGGGG run would give a hit; while intermolecular G4 formation is indeed possible with this motif, we hypothesized that intramolecular structures would be more relevant.

A slightly longer window (e.g., 30 nucleotides) further contributes to eliminating such motifs, but at the cost of significantly decreasing the number of hits (by a factor of 2 to 3; see Table 1): This larger window would, therefore, increase the number of false negatives, i.e., miss "real" intramolecular G4. On the other hand, a much larger window (50–100 nt) would be interesting to identify "G4 clusters" in which multiple tandem quadruplexes may be formed. We present the number of sequences found in three different complete archaeal genomes using four different window sizes and a threshold of 1.2:

Table 1. A number of putative quadruplex sequences (PQS) were found using four different window sizes in three complete archaeal genomes.

Archaea (GC %)	Number of G4 Sequences Found for a Window of:			
	25 nt	**30 nt**	**50 nt**	**100 nt**
Methanococcus maripaludis C7 (33.3%)	558	171	3	0
Cenarchaeum symbiosum A (57.3%)	6019	3197	324	5
Halobacterium salinarum NRC (65.9%)	4738	2313	262	4

As shown in Table 1, long G-rich prone regions, potentially supporting the formation of multiple quadruplexes, are present, but far less frequent (by a factor of 19 to 186 for a window of 50 vs. 25) than the classically defined G4Hunter motifs. In these three genomes, a large majority (95–99%) of the G4-prone regions would only support the formation of a single individual quadruplex.

2.3. Analysis of Putative G4 Sequences Around Annotated NCBI Features

We downloaded feature tables from the NCBI database along with genomic DNA sequences. Feature tables contain annotations of known features found in DNA sequences. We performed an analysis of G4-prone sequences occurrence inside recorded features. Features were grouped by their name stated in the feature table file (gene, rRNA, tRNA, ncRNA, and repeat region). From this analysis, we obtained a file with feature names and numbers of putative G4 forming sequences found inside and around features for each group of species analyzed. Search for putative G4 forming sequences

took place inside feature boundaries; note that frequencies of inverted repeats in mitochondrial DNA (mtDNA) [48], as well in the G4 prone sequences in bacteria [33], are distributed with different frequencies in close proximity to specific features. Further processing was performed in Microsoft Excel and the data are available as Supplementary Table S3.

2.4. Statistical Analysis

A cluster dendrogram of PQS characteristics was constructed in program R, version 3.6.3, library *pvclust* [55], to further reveal and graphically depict similarities between particular archaeal subgroups. Mean, Min, Max, and % PQS values were used as input data (Supplementary Table S4). The following parameters were used for analysis: Cluster method 'ward.D2', distance 'Euclidean', number of bootstrap resampling was set to 10,000. Statistically significant clusters (based on AU values (blue) 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 Table S4). Statistical evaluations of differences in G4 forming sequences presence in various phylogenetic groups were made by a Kruskal–Wallis test with a Bonferroni adjustment in STATISTICA, with p-value cut-off 0.05; data are available in Supplementary Table S5.

2.5. Quadruplex Formation In Vitro

Representative examples of the candidate sequences identified by G4Hunter were experimentally tested for G4 formation using different techniques: Isothermal difference spectra (IDS) and Circular dichroism (CD as described previously [41]).

2.5.1. Samples

Oligonucleotides were purchased from Eurogentec, Belgium, as dried samples purified by RP cartridge purification. Stock solutions were prepared at 250 µM strand concentration in ddH_2O.

2.5.2. Experimental Conditions

Most experiments were performed in a 10 mM Lithium Cacodylate pH 7.1 buffer supplemented with 100 mM KCl (since *Hadesarchaea* has not been cultivated, it is impossible to know their intracellular potassium concentration. However, this is in the range of intracellular potassium concentration for other archaea, such as *Thermococcales*).

2.5.3. Isothermal Spectra

2.5 µM oligonucleotide solutions were prepared in 10 mM Lithium Cacodylate buffer at pH 7.1. The solutions were kept at 95 °C for 5 min and slowly cooled to room temperature and kept at 4 °C overnight. Absorbance spectra were recorded on a Cary 300 (Agilent Technologies, France) spectrophotometer at 37 °C (scan range: 500–200 nm; scan rate: 600 nm/min; automatic baseline correction). After recording these first series of spectra (unfolded as no potassium was present) 1 M KCl (100 µL) was added to the samples, and UV-absorbance spectra were recorded after 15 min equilibration, and corrected for dilution. Each IDS corresponds to the arithmetic difference between the initial (unfolded) and final (folded, corrected for dilution) spectra.

2.5.4. Circular Dichroism

2.5 µM oligonucleotide solutions were prepared in 10 mM lithium cacodylate buffer at pH 7.1 supplemented with 100 mM KCl. The solutions were kept at 95 °C for 5 min and slowly cooled to room temperature and kept at 4 °C overnight. CD spectra were recorded on a JASCO J-1500 (France) spectropolarimeter at room temperature or at 80 °C, using a scan range of 400–210 nm, a scan rate of 200 nm/min, and averaging four accumulations (Supplementary Figure S1).

2.6. G-Quadruplex Binding Proteins Prediction

For G-quadruplex binding proteins prediction, based on previously published G-quadruplex binding motif (RGRGRGRGGGSGGSGGRGRG) [31], the BLASTp algorithm was used [56]. The target organisms were limited to the Archaea domain (NCBI taxid ID: 2157). E-value cut-off was set to 0.05. For similarity search of RecQ helicase from *Escherichia coli* (UNIPROT ID: P15043), BLASTp algorithm [56] was used with an E-value cut-off of 0.0001 and the same restriction to the Archaea domain, as above. BLASTp analyses are enclosed in Supplementary Table S6. FIMO search [57,58] for G-quadruplex binding motif (RGRGRGRGGGSGGSGGRGRG) [31] in *Methanosarcina mazei* complete proteome was carried out on a set of 15722 known protein sequences downloaded from NCBI, with q-value (*p*-value corrected for multiple testing by Benjamini and Hochberg method) cut-off of 0.05 (Supplementary Table S7). The most similar protein of RecQ helicase from *Escherichia coli* (UNIPROT ID: P15043) in *Hadesarchaea archaeon* isolate WYZ-LMO6 was searched using tBLASTn [59], and the resulting best hit was translated using Expasy Translate Tool [60,61] and functional domain were visualized using NCBI CDD [62] (Supplementary Table S8).

3. Results

3.1. Prediction of G4 Forming Sequences in Archaea

We analyzed the occurrence of putative G4 sequences (PQS) with G4Hunter in 3387 archaeal genomes. The length of sequenced archaeal genomes in our dataset varied from 100 kbps to 13.4 Mbps (list provided in Supplementary Table S1). The average GC content was 46.51%, with a minimum of 24.30% for *Nanobsidianus stetteri* isolate SCGC AB-777 (*Nanoarchaeota*) and a maximum of 70.95% for *Halobacteriales archaeon* SW_7_71_33 (phylum *Euryarchaeota*). Using standard parameters for the G4Hunter search algorithm (window size of 25 and G4HS $\geq$ 1.2) we found 4,470,813 PQS in these 3387 archaeal genomes using a default threshold of 1.2. The higher the G4HS score is, the higher the stability of the structure. Over 90% and 98% of sequences with a score above 1.2 or 1.5, respectively, were experimentally demonstrated to form a stable quadruplex in vitro [41]. Figure 3A provides an example of G-rich motifs found in archaea with G4HS between 1.32 and 3.0. As expected from previous analyses on eukaryotes and bacteria, most (97%) PQS have a relatively low (1.2 to 1.4) G4Hunter score. More stable motifs are rarer, with a sharp decrease in the number of retrieved sequences with scores above 1.4, as shown in Table 2. Only 132 PQS with a G4Hunter score of 2 or more were found. A summary of all PQS found in ranges of G4Hunter score intervals and precomputed PQS frequencies per 1000 bp is provided in Table 2.

Table 2. Number of PQS found and their frequencies per 1000 bp in all 3387 archaeal genomes, grouped by G4Hunter score (1.2-1.4 means any sequence with a score between 1.2 and 1.399; 1.4 between 1.4 and 1.599, etc.).

G4HS	Number of PQS in Dataset	Fraction of All PQS	PQS Frequency Per kbp
1.2–1.4	4,344,917	0.9718	1.19
1.4–1.6	119,233	0.0267	1.8×10^{-2}
1.6–1.8	6357	0.00142	9.9×10^{-4}
1.8–2.0	174	0.0000389	2.5×10^{-5}
>2.0	132	0.0000295	2.2×10^{-5}
Total	4,470,813	1	

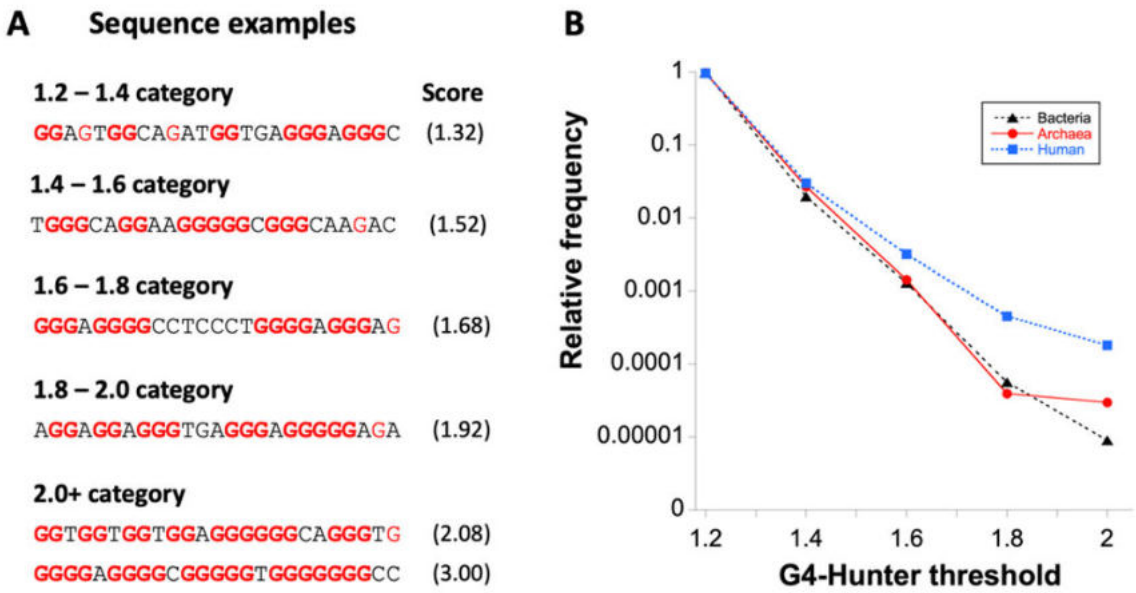


Figure 3. Examples of sequences with different G-quadruplexes (G4) Hunter scores (G4HS) and distribution of PQS according to threshold category. (**A**) Examples of archaea 25-nt long sequences (corresponding to the window size chosen for the analysis) for which G4Hunter scores are provided within parentheses. Isolated guanines are shown in red, all other guanines in bold red characters. Longer archaea motifs with high G4H scores are provided in Table 3. (**B**) Distribution of G4-prone motifs according to the G4Hunter score. 1.2 means any sequence with a score between 1.2 and 1.399; 1.4 between 1.4 and 1.599, etc. These numbers are normalized by the total number of PQS found in bacteria, archaea, and compared with *Homo sapiens*. The first category represents 97.9% and 97.2% of all PQS sequences in bacteria and archaea, respectively. Note the log scale on the Y-axis.

Table 3. Genomic sequences sizes, GC%, total count of PQS, and mean frequencies of quadruplex motifs. Seq (total number of sequences), Median (median length of sequences), Short. (shortest sequence), Long. (longest sequence), GC % (average GC content), PQS (total number of predicted PQS), Mean f (mean frequency of predicted PQS per 1000 bp), Min f (lowest frequency of predicted PQS per 1000 bp), Max f (highest frequency of predicted PQS per 1000 bp). %PQS corresponds to the probability that any given nucleotide in the group or subgroup belongs to a G4-prone region (G4H > 1.2). Colors correspond to phylogenetic tree depiction.

Kingdom	**Seq.**	**Median**	**Short**	**Long**	**GC %**	**PQS**	**Mean f**	**Min f**	**Max f**	**% PQS**
Archeae	3387	1,686,930	100,212	13,399,915	46.51	7,927,775	1.21	0.04	15.31	3.58
Superphylum	**Seq.**	**Median**	**Short**	**Long**	**GC %**	**PQS**	**Mean f**	**Min f**	**Max f**	**% PQS**
BAT	320	1,180,629	164,795	3,506,105	43.07	421,678	1.16	0.05	8.42	3.49
Cren	379	1,808,184	210,860	6,451,204	43.05	1,009,660	1.56	0.09	9.44	4.75
Asgard	71	2,322,715	291,515	5,684,038	38.75	74,647	0.47	0.12	1.50	1.39
DPANN	309	832,169	100,212	6,604,953	39.22	219,058	0.70	0.08	4.20	2.18
Eury	2308	1,826,841	137,797	13,399,915	48.77	6,202,732	1.25	0.04	15.31	3.68
Phylum	**Seq.**	**Median**	**Short**	**Long**	**GC %**	**PQS**	**Mean f**	**Min f**	**Max f**	**% PQS**
Bathyarchaeota	128	1,208,976.5	200,493	3,506,105	46.29	245,162	1.54	0.23	8.42	3.00
Thaumarchaeota	192	1,173,909.5	164,795	3,441,569	40.93	176,516	0.91	0.05	5.32	2.73
Thermoproteales	147	1,581,744	242,587	3,969,448	45.86	513,053	2.07	0.11	7.38	6.31
Sulfolobales	118	2,223,757.5	210,860	3,034,024	38.20	200,842	0.79	0.34	4.58	2.38
Desulfurococcales	29	1,580,347	807,477	2,148,448	46.99	99,211	2.29	0.40	6.37	6.95
Verstraetearchaeota	18	1,171,913.5	419,172	1,937,662	46.76	40,586	1.83	0.10	3.43	5.50

Table 3. *Cont.*

Marsarchaeota	15	1,915,630	351,358	3,731,392	46.72	52,853	1.64	0.47	2.94	5.01
Geothermarchaeota	6	1,183,145.5	803,797	1,671,866	42.72	16,582	2.15	0.96	7.03	6.65
Nezhaarchaeota	2	1,332,140.5	1,315,707	1,348,574	43.53	2016	0.76	0.75	0.77	2.27
Korarchaeota	18	1,542,873	834,209	2,942,065	48.39	68,434	2.63	1.05	9.44	7.95
Unclassified Crenarchaeota	27	1,203,892	301,027	6,451,204	37.01	19,361	0.44	0.09	1.49	1.29
Lokiarchaeota	29	1,892,624	320,847	5,143,417	32.77	25,479	0.41	0.21	1.50	1.24
Odinarchaeota	1	1,460,710	1,460,710	1,460,710	38.05	1038	0.71	0.71	0.71	2.16
Thorarchaeota	29	2,770,204	291,515	4,389,059	46.55	40,006	0.60	0.24	1.18	1.76
Heimdallarchaeota	12	2,167,091	432,340	5,684,038	34.42	8124	0.27	0.12	0.50	0.82
Aenigmarchaeota	35	751,672	248,182	1,410,470	39.33	17,990	0.71	0.11	3.78	2.12
Nanohaloarchaeota	17	815,638	565,289	1,480,846	44.53	8672	0.48	0.09	1.82	1.50
Woesearchaeota	72	966,794.5	518,295	2,944,567	40.77	57,833	0.66	0.08	3.92	1.96
Pacearchaeota	60	719,507	279,432	6,604,953	33.74	37,675	0.56	0.08	2.99	1.73
Nanoarchaeota	25	577,110	204,081	1,162,239	32.83	9940	0.59	0.13	4.20	1.70
Micrarchaeota	39	887,931	658,716	1,333,875	50.41	42,298	1.17	0.15	2.86	3.47
Diapherotrites	19	568,419	302,064	1,130,899	37.42	6077	0.49	0.11	2.33	1.46
Unclassified DPANN	40	858,043.5	100,212	3,188,023	35.57	33,846	0.67	0.15	2.39	2.04
Hadesarchaeota	12	857,575	451,393	1,241,441	53.77	56,369	4.61	1.26	15.31	14.55
Persephonarchaeota	33	637,942	137,797	1,412,535	44.06	34,905	1.49	0.59	2.36	4.49
Thermococcales	60	1,867,904.5	207,909	2,388,527	46.77	191,492	1.72	0.47	7.53	5.15
Theinoarchaeota	2	4,165,806	3,559,548	4,772,064	41.57	5480	0.66	0.65	0.67	1.94
Methanofastidiosa	96	992,372	156,656	13,399,915	40.71	141,192	0.83	0.08	3.64	2.54
Methanococcales	24	1,717,483	1,207,361	1,936,387	32.01	15,065	0.39	0.20	0.86	1.19
Methanobacteriales	224	2,001,036	1,157,521	3,466,370	33.62	175,191	0.39	0.04	2.32	1.14
Methanopyrales	3	1,430,309	1,421,621	1,694,969	58.94	10,798	2.34	1.97	3.00	6.84
Methanomassiliicoccales	91	1,404,109	640,223	2,641,216	56.22	257,340	1.85	0.22	4.41	5.38
Thermoplasmatales	135	1,621,237	593,453	2,816,557	42.71	246,832	1.13	0.11	7.03	3.42
Acidoprofondum/DHV2-2	11	1,731,076	519,420	2,981,805	40.55	16,609	1.21	0.29	4.12	3.59
Archaeoglobales	53	1,901,943	478,535	3,408,041	42.98	117,470	1.22	0.57	3.29	3.66
Methanosarcinales	279	2,913,215	208,261	5,751,492	44.99	845,394	1.19	0.15	7.52	3.54
Methanomicrobiales	146	2,228,967.5	622,799	3,978,804	54.97	783,172	2.38	0.23	7.20	7.07
Methanocellales	5	2,957,635	1,465,272	3,243,770	50.96	16,825	1.21	0.41	1.88	3.51
Halobacteriales	440	3,585,981	397,623	5,605,381	63.95	2,271,600	1.56	0.08	4.25	4.50
Unclassified Diaforarchaea	97	1,460,542	233,168	2,294,894	47.38	136,115	1.03	0.18	2.55	3.02
Unclassified other	597	1,400,198	258,312	7,416,915	46.88	862,962	1.02	0.07	5.16	3.00

The comparison of G4 prone sequences found in archaea with bacteria genomes revealed that in both domains, frequencies sharply decreased with G4HS as compared to the human genome, in which highly stable G4s are relatively more frequent (see Figure 3B). This result indicates an overall stronger relative selection pressure against stable G4 motifs in both archaea and bacteria as compared to humans, and likely most eukaryotes, as the relative number of G4Hunter high scoring motifs is even higher in yeast [63]. Guo and Bartel suggested that eukaryotes have robust machinery that globally unfolds RNA G-quadruplexes, whereas some bacteria have instead undergone evolutionary depletion of G-quadruplex-forming sequences [64]. Our analysis suggests that archaea behave like bacteria, except for the slight difference found for the most stable motifs (G4HS >2), which were less selected against in archaea than in bacteria.

3.2. Variation in Frequency for G4 Forming Sequences in Archaea

The total number of analyzed sequences in particular phylogenetic categories, together with a median length of the genome, shortest genome, longest genome, mean, minimal, and maximal observed frequency PQS per kbp, and total PQS counts are shown in Table 3. For this analysis, Archaea have been divided into five superphyla that form monophyletic assemblages (clades) in the most recent phylogenetic analysis and 41 subgroups that correspond to different taxonomic ranks (suffix *aeota* for phylum, candidate phylum, suffix *ales* for orders). Seven subgroups have an average GC content above 50%, the highest GC content being observed in *Halobacteriales* (63.95%), which is also the archaeal group containing the highest number of available genome sequences–440), all other groups have average GC contents below 50%.

The mean frequency of PQS per kbp for all archaeal genomes was 1.207. The lowest mean frequency was for the *Heimdallarchaeota* (0.273), followed by *Methanococcales* and *Methanobacteriales* (0.39). The highest density of PQS was found in the *Hadesarchaea* subgroup (4.607), followed by *Korarchaeota* (2.626). The highest absolute frequency of PQS was found in *Hadesarchaea archaeon isolate* WYZ-LMO6 with 15.3 PQS per 1000bp (i.e., one quadruplex every 65 bp), and the lowest frequency was found in *Methanobrevibacter* sp. 87.7: Interestingly, only 71 PQS were found in its 1.92 Mb long genome (Supplementary Table S2A). Detailed statistical characteristics for PQS frequencies per kbp (including mean, variance, outliers) are depicted in boxplots for all inspected subgroups (Figure 4). The *Hadesarchaea* subgroup has a higher PQS frequency in comparison to other subgroups. The comparison of the five main superphyla BAT, Cren, Asgard, Eury, and DPANN (*Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanoarchaeota*, and *Nanohaloarchaeota*) (Figure 1) revealed the highest mean PQS frequency in Cren superphylum (1.15) and the lowest in Asgard superphylum (0.48). However, the *Hadesarchaea* subgroup, which exhibits the highest frequency among subgroups, is found in the Eury superphylum. The detailed data for superphyla are in Supplementary Table S2B, for subgroups in Supplementary Table S2C.

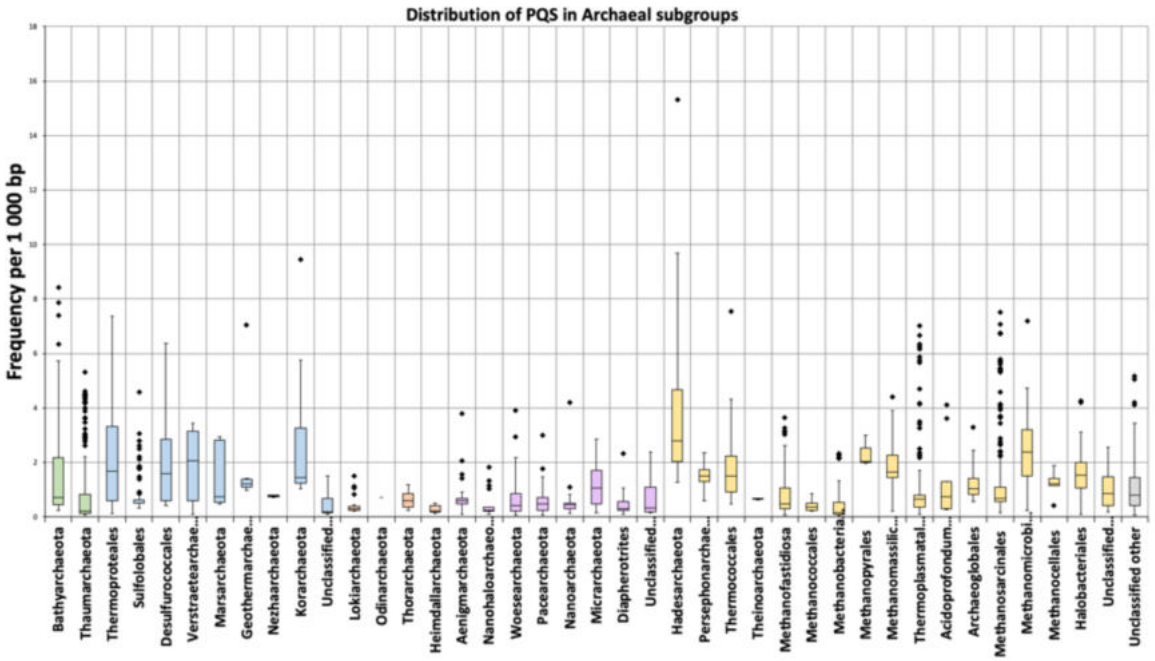


Figure 4. Frequencies of PQS in subgroups of analyzed archaeal genomes. Data within boxes span the interquartile range, and whiskers show the lowest and highest values within 1.5 interquartile range. Black points denote outliers. Horizontal black lines inside boxplots are median values.

A cluster dendrogram shows the similarities among subgroups based on the PQS data (Figure 5). This dendrogram shows that the *Hadesarchaeota* subgroup is the most distant one (the shortest branch length) compare to other subgroups. The cluster dendrogram based on PQS characteristics is similar to the phylogenetic relationships (see Figure 1). For example, all of the Asgard subgroups (*Odinarchaeota*, *Heimdallarchaeota*, *Thorarchaeota*, and *Lokiarchaeota*) lie close together, in one bigger cluster (Figure 5, left part). Other examples are the *Woesearchaeota*, *Aenigmarchaeota*, and *Nanoarchaeota* subgroups, which are members of the DPANN superphylum, and lie adjacent to each other in PQS based cluster tree. On the other hand, all of the subgroups with the prefix "-thermo", indicative of high-temperature environments, are clustered together (*Thermoplasmatales*, *Thermococcales*, *Thermoproteales*, and *Geothermarchaeota*). These subgroups are relatively PQS rich, but lack phylogenetical proximity, suggesting that PQS richness does not rely on evolutionary proximity.

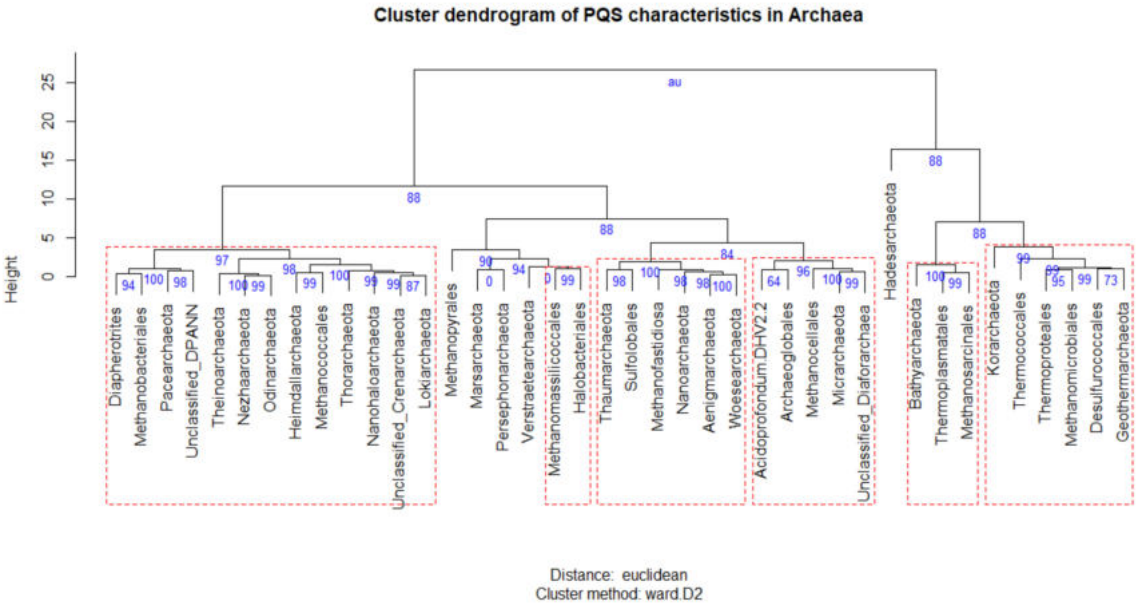


Figure 5. Cluster dendrogram of PQS characteristics of archaeal subgroups. Cluster dendrogram of PQS characteristics (Supplementary Table S4) was made in R v. 3.6.3 (code provided in Supplementary Table S4) using pvclust package with these parameters: Cluster method 'ward.D2', distance 'euclidean', number of bootstrap resamplings was 10,000. AU values are in blue and indicate the statistical significance of particular branching (values above 95 are equivalent to *p*-values lesser than 0.05). Statistically significant clusters are highlighted by red dashed rectangles.

We then analyzed the relationship between overall % GC content and PQS frequency (Figure 6). PQS frequencies tend to correlate with GC content as G4-prone motifs need to be relatively G-rich; however, there are interesting exceptions to this rule, and this correlation is poorer than anticipated. Ding et al. already noticed that *Methanomicrobia* and *Thermococci* have greater densities of PQS than the theoretical values based on the GC % of their genomes [35]. Organisms with higher than expected PQS frequencies based on their GC content (over 50% of the maximal observed PQS frequency, Figure 6) are highlighted in color; the whole figure is separated into smaller segments according to inspected G4Hunter score intervals. The most extreme outlier is *Hadesarchaea* archaeon, for which 51% of its genome has a G4Hunter score above 1.2, despite a GC content of 54%, i.e., only modestly above the 46.5% average for all sequences tested here, and far below the most GC rich archaea genomes. Cherry-picked examples of G-rich motifs with high G4 Hunter scores (G4HS) in *Hadesarchaea archaeon* are provided in Table 4. We have also carried out additional statistical evaluation of PQS differences between all groups and subgroups; detailed results are found in Supplementary Table S5. Nearly all comparisons were significant, i.e., there are significant differences between PQS frequencies of particular groups and subgroups.

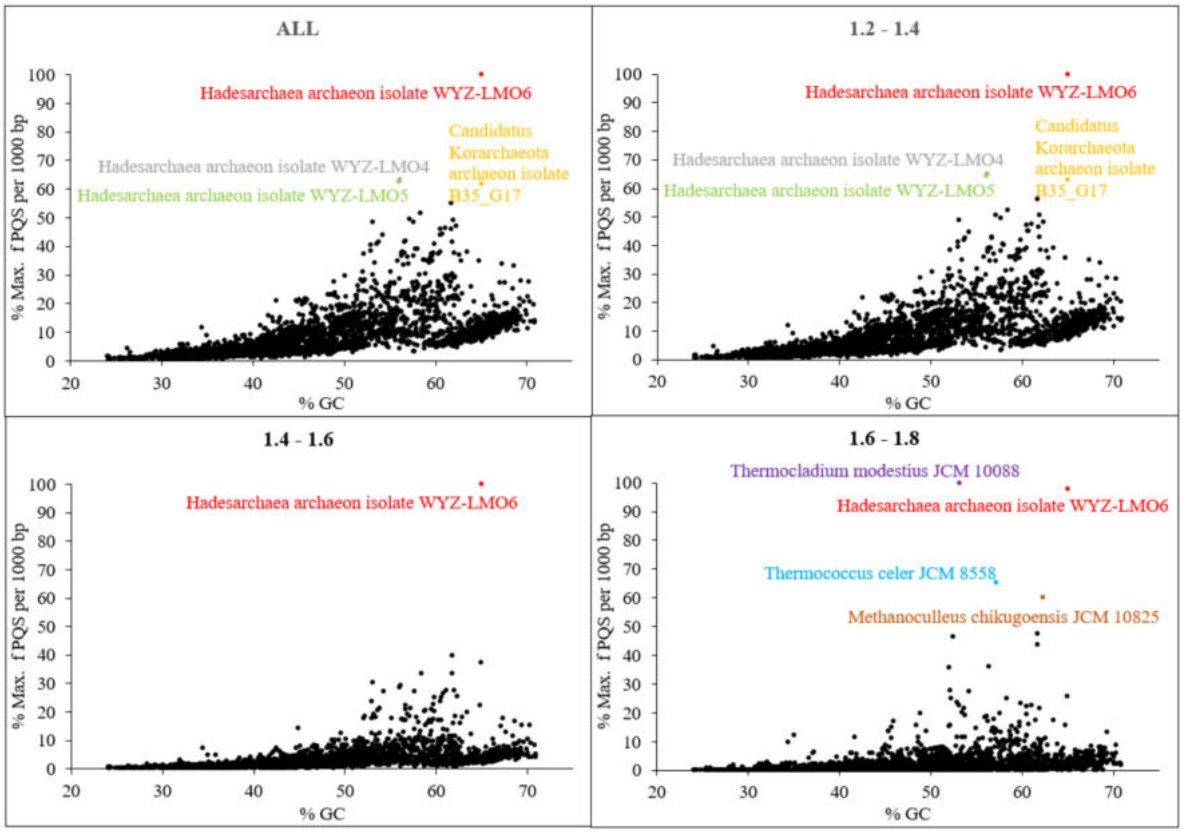


Figure 6. Relationship between the observed frequency of PQS per 1000 bp and GC content. Different G4Hunter score intervals are considered. In each G4Hunter score interval miniplot, frequencies were normalized according to the highest observed frequency of PQS. Organisms with max. frequency per 1000 bp greater than 50% are described and highlighted in color.

Table 4. Long G4-prone motifs with high G4HS found in *Hadesarchea archeon*.

Name	Sequences (5′ to 3′)	G4 Hunter Score	IDS	CD
038_K	AGGCTGGGGGTGAGGGCGGTGGTGGGGAAGGGAGGGGTGGGGGAGAAAACGAAGGGGGT	2.07	G4	Parallel
086_K	TGGGGAGGAGGGGAGGGGAGGTGGGCTGGGGGGGGCT	2.57	G4	Parallel
174_K	AGGGTGAGGGAGGAGGTGCTGGGGGGAAGGGAGGTGGGGGAGGGGGAGGTGGAGGGGCTGGTGAGGGA	2.07	G4	Parallel
175_K	AGGGGAGGAGGGTGGCCGTGGTGGGGGCGGGGGGAGGGGCGGGGGTGGGGGGGCCTGGGGGGA	2.54	G4	Parallel
176_K	AGGAGGAGGGTGAGGGACCAGGGGAGGAGGGAGGGGAGGGGGGGAAGGAGGAGGGAGAGGAGGAGGGA	1.93	G4	Parallel
178_K	TGGTGGGGGCGGGGGGAGGGGCGGGGGTGGGGGGGCCTGGGGGGA	2.89	G4	Parallel
195_K	AGGGGAGGAGGGTGGCCGTGGTGGGGGCGGGGGGAGGGGCGGGGGTGGCCTCCACGGA	1.91	G4	Parallel
196_K	AGGGGAGGAGGGAGGGGAGGGGGGGAAGGAGGAGGGAGAGGAGGAGGGA	2.22	G4	Parallel
245_K	GGGGTCGTCGGGGGGGAGAGCTGGGGAGGAGGGGAGGGGAGGTGGGCTGGGGGGGGCTGGGGAGGGAGGAGGTGAGGGG	2.33	G4	Parallel
640_K	AGGGAGGTGGGGGAGGGGGAGGTGGAGGGGCT	2.38	G4	Parallel
642_K	TGGTGGGGGCGGGGGGAGGGGCGGGGGT	2.93	G4	Hybrid*
643_K	AGGCTGGGGGTGAGGGCGGTGGTGGGGAAGGGAGGGGTGGGGGAGAAAACGAAGGGGGT	2.07	G4	Parallel
644_K	AGGGCGGTGGTGGGGAAGGGAGGGGTGGGGGA	2.41	G4	Parallel
645_K	GGCGGGGGGGGAGTCCTTCATCCTGGGGTAGGGG	1.74	G4	Parallel

* Sequence 642_K adopts a hybrid structure at room temperature, which is converted to a parallel conformation at high temperatures.

Figure 7 shows the relationship between GC percentage and mean PQS frequencies (or mean percentage of PQS length of the genome) in particular archaeal subgroups. Overall, we found some correlation (although far from perfect, as shown by $R^2 = 0.7$) between mean PQS frequencies (expressed as the mean fraction of nucleotides of the genomes involved a PQS motif) and increasing GC % content.

The highest mean percentage of PQS length of the genomes was found in subgroup *Hadesarchaea*, in which more than 10% of their genomes are involved in a potential PQS.

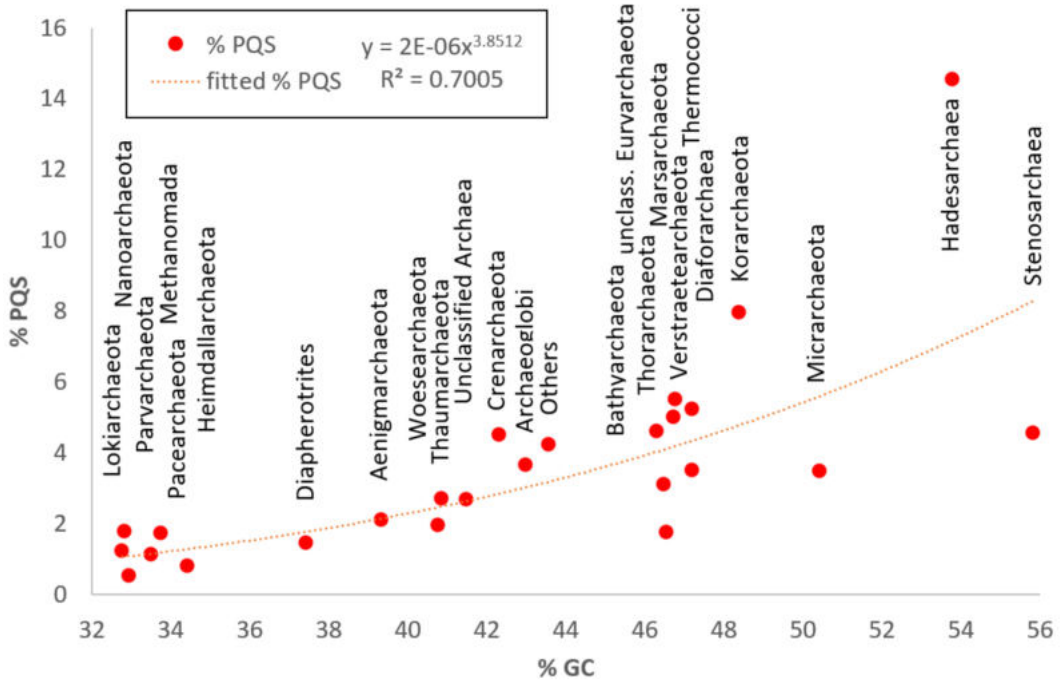


Figure 7. Relationship between GC percentage and % of PQS in genomes of particular archaeal subgroups. The Fitted equation with the R^2 coefficient is depicted on the top side of the plot.

3.3. Localization of PQS in Genomes

To evaluate the position of PQS in archaeal genomes, we downloaded the described "features" of all archaeal genomes and analyzed the presence of all PQS in annotated sequences (Figure 8). Overall, we find a higher density of G4-prone motifs in non-protein coding RNAs (tRNA, rRNA, and other ncRNA) than in protein-coding genes. G4 density in ncRNA is clearly above average genomic G4 density, while mRNA G4 density is close to the genomic average. This may derive in part from the observation that rRNA and tRNA genes are especially GC-rich in hyperthermophilic archaea, in order to stabilize folding under harsh conditions [65]. On the other hand, we can probably expect a stronger selection pressure against the formation of intramolecular quadruplexes within the relatively small tRNA core, as this would disrupt its three-dimensional shape and alter its biological function. In line with this hypothesis, the PQS frequencies are actually lower in tRNA than in ncRNA and rRNA [66]. Interestingly, the 5′ end of some human tRNA genes is often G-rich and has been reported to allow G4 formation: Ivanov and colleagues have shown that mature cytoplasmic tRNAs are cleaved during stress response to produce tRNA fragments that function to repress translation in vivo and that these bioactive tRNA fragments assemble into intermolecular RNA G4s [67]. The 5′ fragment of $tRNA^{Ala}$ involves a predominant hairpin structure that starts with the 5′-GGGGGU motif, allowing the formation of tetramolecular quadruplex structures with five tetrad layers. Interestingly, tRNA-derived fragments have also been described in archaea. For example, a 26-residue-long fragment (5′ GGGUUGGUGGUCUAGUCUGGUUAUGA) originating from the 5′ part of valine tRNA is the most abundant tRNA fragment in *Haloferax volcanii* [68]. This fragment, while exhibiting a relatively G-rich 5′ end (starting with GGGUUGG), may, in principle, allow intermolecular quadruplex formation as well.

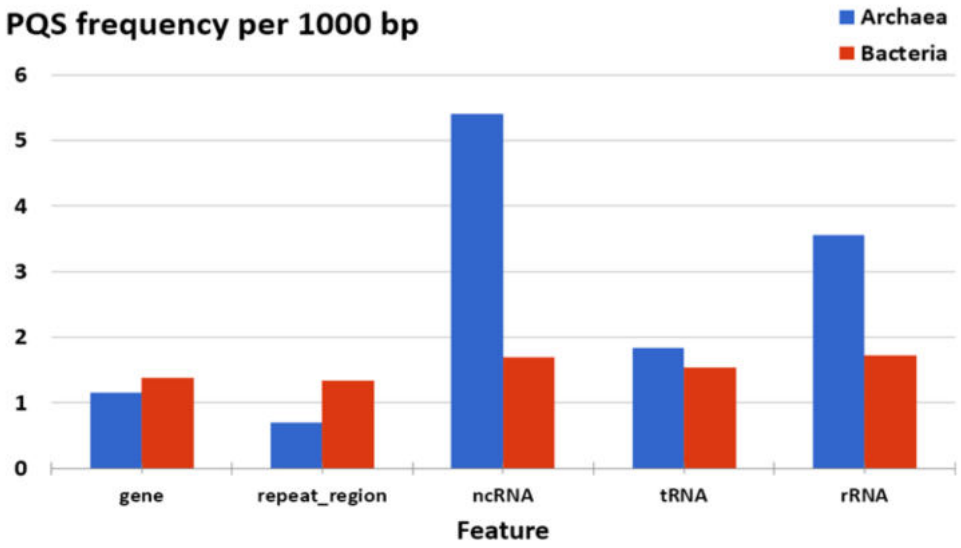


Figure 8. Differences in PQS frequency by DNA locus. The chart shows PQS frequencies normalized per 1000 bp annotated locations from the NCBI database and shows a comparison between Archaea and Bacteria. Archaea G4-prone motifs are strongly over-represented in ncRNA and rRNA compared to the average G4 density in Archaea (mean f = 1.207), but also compared to bacteria. PQS count is provided in Supplementary Table S3 Excel file.

Unfortunately, other features in archaeal genomes are so poorly annotated that we cannot use these data for evaluation. Comparison of PQS frequencies in annotated sequences with analyses of Bacteria shows the same trend for ncRNA, rRNA, protein-coding gene, and tRNA features. In contrast, the frequency in bacteria for ncRNA is 1.7 per kbp, and the frequency in archaea for ncRNA is 5.3 per kbp. On the other hand, the PQS frequency in repeat regions is lower in archaea than in the bacteria genome. We have to take into account that the data could be influenced by poor annotation in archaea genomes, and also by a low number of annotated sequences in Archaea; only 141 representative archaeal genomes are annotated, compared to 1627 representative bacteria annotated genomes. The strong abundance of the PQS in ncRNA compare to other locations pointing to its functional relevance. ncRNAs are present in the cells as single-stranded molecules in contrast to DNA, and therefore, they can easily adopt the G4 structures as a part of their 3D arrangement similarly to mRNAs [69,70]. It has been shown that ncRNAs play important roles in many cellular processes, including the regulation of gene transcription, post-transcriptional, and epigenetic regulations [71,72].

Other specific regions, such as replication origins or promoter regions, were not included in this graph. The oriC 10.0 database (http://tubic.org/doric/public/index.php) contains 226 archaeal origins of replication obtained by both in vitro studies and in silico predictions ([73]), prediction and experimental data are available for the *Thermococcales* [74,75], the *Haloarchaea*, and the *Sulfolobales* [76]. Archaeal replicators, as in bacteria, are composed of three main elements: A cluster of binding sites for the initiator Cdc6, the DNA unwinding element (DUE), and binding sites for regulatory proteins [75]. Interestingly, it was found in several *Haloarchaea* species that a specific (TGGGGGGG) motif occurs in one of the two origins of replication (oriC1) [77]. This long G-rich motif was shown to be necessary for efficient replication initiation in *Haloarcula hispanica* [78,79] and predicted to be prone to inter-molecular quadruplex formation.

3.4. Experimental Demonstration of Quadruplex Formation In Vitro

Next, we selected a few DNA G4-prone motifs found in *Hadesarchaea* and experimentally tested if they formed a G4 structure under classical conditions. As inferred from isothermal difference spectra (IDS) (Figure 9a) and circular dichroism (CD) spectra (Figure 9b), all motifs clearly formed G-quadruplexes at room temperature. However, as these motifs are found in an archeon expected to live

at a high temperature, we also recorded the spectra at 80 °C. As shown in Figure 9c, these quadruplexes were thermally stable and still formed at high temperatures. Of note, most spectra are indicative of a parallel fold. This bias is the result of a high threshold for G4Hunter (all motifs have scores > 1.7). As a consequence, these motifs are very G-rich, with runs of G separated by short spacers, often 1–2 nt. As short loops tend to be propeller-type, this sequence bias will favor a parallel conformation.

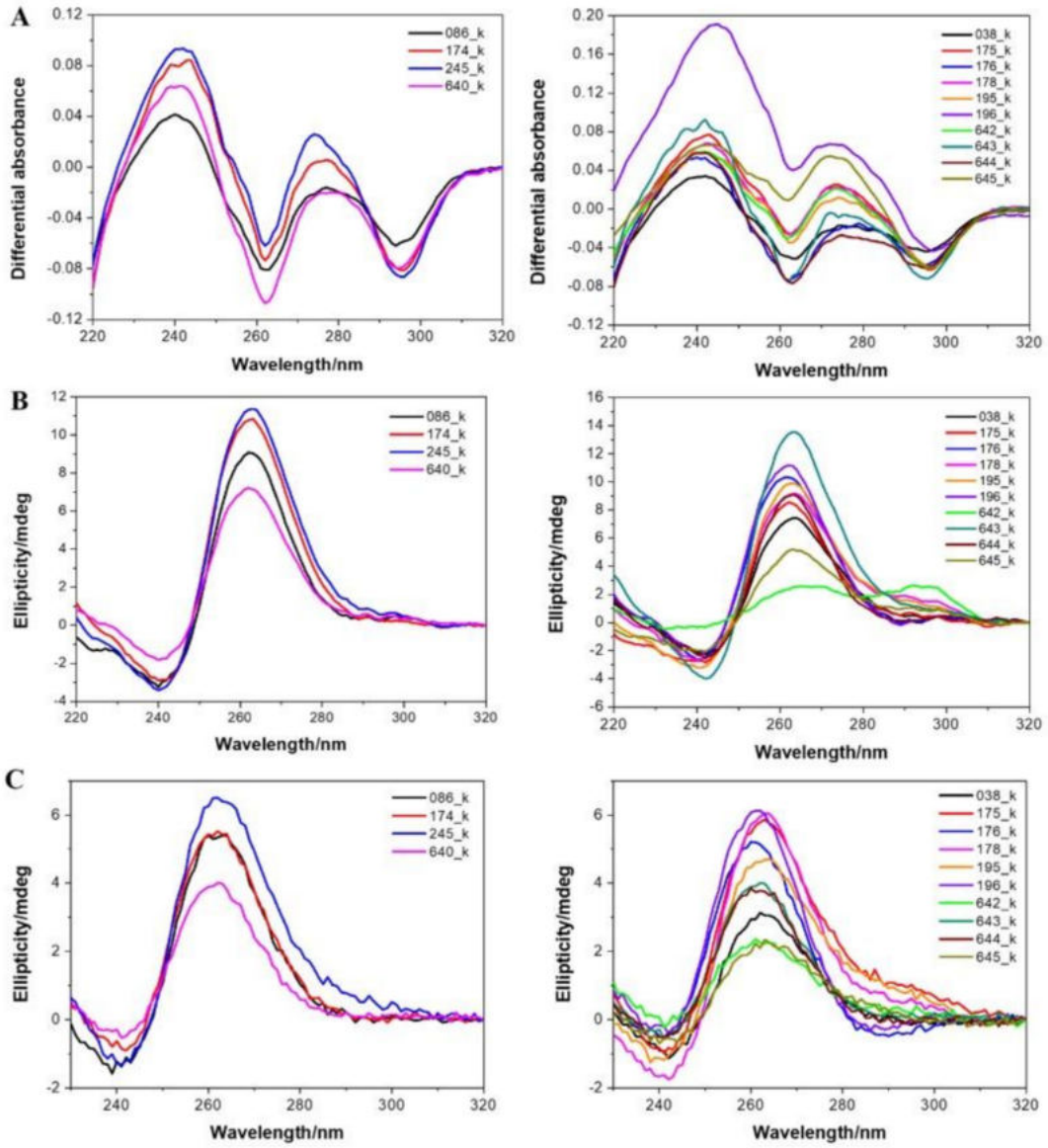


Figure 9. Experimental evidence for quadruplex formation with archaea sequences. Isothermal differential absorbance (IDS; panel **A**) and circular dichroism (CD; panels **B** and **C**) spectra of *Hadesarchaea archeon* DNA sequences were recorded at 20 °C (panels **A** and **B**) or at a high temperature (80 °C) for CD (panel **C**).

3.5. G4-Binding Proteins from Archaea

Given that G4-prone motifs are found in Archaea, and actually extremely abundant in some subgroups, it was interesting to check if potential helicases are present to solve these structures. A number of DNA and RNA G4-helicases have been identified in eukaryotes, e.g., Pif1, DOG, Rhau/DHX36, WRN, BLM; for a review [80]. Little or no experimental data is currently available on archaea enzymes able to unfold G-quadruplexes. As RecQ has been reported to unfold G4 structures

in bacteria, we searched for RecQ homologs in Archaea. A BLASTp search using RecQ (UNIPROT ID: P15043) from *E. coli* as a query revealed 1206 homologous protein sequences in a archaeal domain with an E-value cut-off = 0.0001. A listing of all candidates identified is presented in supplementary information (Supplementary Table S6). Five proteins have an identity with G-quadruplex RecQ resolvase higher than 50%, and 312 proteins have more than 50% aa positives hits in the sequence, suggesting that they share the G4 unfolding functionality in archaea genomes. Besides protein actively unfolding G4 structures, other peptides may actually bind to single-strand G-rich sequences and passively contribute to G4 unfolding by conformational selection. This is the case for a single-strand binding protein isolated from *Methanococcus jannaschii*, which was used to design an assay to detect G4 formation [79]. Apart from proteins that actively or passively unfold quadruplexes, others may bind to and sometimes promote G4 formation. The amino acid composition of 77 G-quadruplex binding proteins from *Homo sapiens* revealed unique features of quadruplex binding proteins, with prominent enrichment for glycine (G) and arginine I [31]. Human-binding proteins share a 20 amino acid long motif/domain (RGRGR GRGGG SGGSG GRGRG), which is similar to the previously described RG-rich domain of the FMR1 G-quadruplex binding protein. The search for this 20 amino acid-long motif in archaea proteome found 23 hits/potential G-quadruplex binding proteins with an E-value threshold of 0.05; the identity was found, e.g., for RNA DEAD box helicase or for two 30S ribosomal proteins S4 (Supplementary Table S6, list 2). We searched protein sequences in the proteome of the mesophylic archaeon *Methanosarcina mazei* (for which the largest amount of proteins is known) for the presence of this motif. For highly significant p values ($p < 10^{-6}$), we found four proteins with a potential quadruplex-binding motif (Supplementary Table S7), while significantly more (193) hits were found for p-values $< 1 \times 10^{-5}$. Three of them are without any known function (DUF134 domain-containing protein, PGF-pre-PGF domain-containing protein, and DUF5320 domain-containing protein). Even if the full proteome of *Hadesarchaea archaeon* is not known, it is interesting to note that this RG-domain is present in a number of putative proteins. In addition, while a true RecQ homolog was not found, one *Hadesarchaea archaon* 600aa-polypeptide has a good similarity with RecQ in its N-terminal half (Supplementary Table S8). The presence of the NIQI motif in the "DNA-directed RNA polymerase subunit" is also interesting and possibly logical, given the necessity of unraveling G-quadruplexes during transcription. The presence in archaeal genomes of potential G4-binding and G4-unfolding proteins supports the formation of quadruplex structures in archaeal cells.

4. Discussion

We provide here the first comprehensive study of PQS occurrences, frequencies, and distributions in archaeal genomes. The overall analysis made on global frequency hides extreme differences between species and subgroups, which can be explained by differences in GC content and possibly codon usage.

At one end of the G4 spectrum, some subgroups of archaea, such as *Parvarchaeota* or *Heimdallarchaeota*, have very low PQS frequencies, and PQS cover 1% or less of their genomes. In sharp contrast, we found an unprecedented enrichment of PQS for some subgroups, often living under extreme conditions. For example, over 50% of the genome of *Hadesarchaea archaeon* may potentially adopt a quadruplex fold. This *Hadesarchaea* is living under extreme conditions, as it was found in South African gold mines 3 km underground, without light and oxygen (*Hades* is the Greek god of the underworld). Following this analysis, we used the BioSample NCBI database [78] to compare the living environment of the archaea organisms with the highest PQS frequencies. Data for all genomes with PQS frequency above 6 per kbp are shown in Table 5. A majority of organisms with extremely high PQS frequencies are found in hot springs sediments or in deep-sea hydrothermal vent sediments, and this high PQS frequency may be associated with their extremophilic life, although more work will be necessary to compare G4 density in acidophilic, thermophilic, halophilic and psychrophilic organisms. For example, in bacteria, in the Gram-positive subgroup *Deinococcus-Thermus*, a high PQS frequency was associated with their extremophilic origin [35,81], while the gram-negative extremophilic bacteria subgroup *Thermotogae* are among organisms with a low PQS frequency [33]. We suggest

that the high stability of G4 structures compare to dsDNA structure could play important roles in archaea and Gram-positive extremophiles organisms. We then experimentally confirmed G4 formation with a few archaea sequences to confirm that our in silico predictions are verified: All predicted experimentally tested formed stable G-quadruplexes in vitro. This absence of false positives is hardly surprising given that we chose high scoring motifs. From our published [41] and unpublished data on now over 500 sequences, false positives for sequences with scores above 1.5 are extremely rare (<1.5%), and we have yet to find a false positive with a score > 1.75. Some of the sequences considered were long and may even allow the formation of two juxtaposed G4 structures. In a few cases, we can even propose a topology, as for example, TGGTGGGGGCGGGGGGAGGGGCGGGGGT (642K), in which the predicted guanine tracks (underlined) may either be: TGGTG**GGGG**CG**GGGG**GA**GGGG**CG**GGGG**T or TGGTG**GGGG**C**GGGG**GGA**GGGG**C**GGGG**GT, and different folds may result from these possibilities (the latter would be likely parallel, as experimentally observed at 80 °C, while the former may adopt a non-parallel fold, as observed at room temperature). Note, however, that G4 hunter does not make any hypothesis on the G tracts involved in G4 formation, in contrast with Quadparser, for example, where one actively seeks the four runs of G involved in G-quartet formation. G4 formation is (still) full of surprises, and correctly predicting which runs (or individual guanines) participate in G-quartet formation is far from trivial and requires extensive experimental validation.

The extreme enrichment found in some archaea challenges our existing views on "noncanonical" DNA structures to which G-quadruplexes belong, as it is plausible that a substantial part of the *Hadesarchaea* genome may be packed into G-quadruplex structures. The complementary C-rich strand may also fold into a different quadruplex structure called the i-motif [82] that is favored by acidic pH. Further studies will be dedicated to i-DNA formation in Archaea.

Table 5. Detailed characteristics of archaeal species with PQS frequency per 1000 bp greater than 6.00. Living environments data were obtained from the BioSample NCBI database [83].

Organism Name	GC Content	PQS f	% PQS	Living Environment (Isolated from)
Hadesarchaea archaeon isolate WYZ-LMO6	65.01	15.310	51.15	Hot springs sediment, Yellowstone NP, USA
Hadesarchaea archaeon isolate WYZ-LMO4	56.17	9.685	31.10	Hot springs sediment, Jinze hot spring, China
Hadesarchaea archaeon isolate WYZ-LMO5	56.04	9.581	30.69	Hot springs sediment, Jinze hot spring, China
Korarchaeota archaeon isolate B35_G17	65.01	9.445	28.80	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Bathyarchaeota archaeon B23	61.78	8.418	26.12	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Bathyarchaeota archaeon isolate M10_bin139	58.42	7.858	24.55	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Thermococcus celer JCM 8558	57.21	7.534	24.52	Solfataric marine water hole on a beach of Vulcano, Italy
Methanosaeta harundinacea isolate UBA152	62.01	7.518	23.12	Waste water, Suncor tailings pond 6, Canada
Bathyarchaeota archaeon isolate B23_G15	57.67	7.397	22.90	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Thermocladium modestius JCM 10088	53.14	7.381	25.59	Mud from a spring pool, Noji-onsen, Fukushima, Japan
Methanoculleus chikugoensis JCM 10825	62.36	7.198	22.90	Paddy field soil, Chikugo, Fukuoka, Japan

Table 5. *Cont.*

Organism Name	GC Content	PQS f	% PQS	Living Environment (Isolated from)
Methanosaeta harundinacea isolate UBA281	61.14	7.089	21.80	Wastewater, North Alberta, Canada
Geothermarchaeota archaeon ex4572_27	60.54	7.032	22.01	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Thermoplasmata archaeon isolate CSSed11_322R1	61.82	7.028	22.57	Hypersaline soda lake sediment, Kulunda Steppe, Russia
Methanosarcinales archaeon Methan_02	60.8	6.738	20.67	Anaerobic digester metagenome, Australia
Methanosaeta harundinacea 6Ac	60.6	6.721	20.66	isolated from an upflow anaerobic sludge blanket reactor treating beer-manufacture wastewater in Beijing, China. (ref PMID:16403877)
Thermoplasmatales archaeon ex4484_36	54.25	6.673	21.15	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Aeropyrum camini SY1 = JCM 12091	56.73	6.370	19.72	Deep-sea hydrothermal vent chimney, the Suiyo Seamount in the Izu-Bonin Arc, Japan
Bathyarchaeota archaeon isolate B46_G17	61.92	6.332	19.03	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Thermoplasmata archaeon isolate B14_G15	53.83	6.327	20.11	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Thermoplasmata archaeon isolate B23_G1	53.66	6.240	19.72	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico
Pyrobaculum neutrophilum V24Sta	59.91	6.233	19.52	isolated from a hot spring in Iceland
Thermoplasmata archaeon isolate B23_G9	52.98	6.164	19.65	Deep-sea hydrothermal vent sediments, Guaymas Basin, Gulf of California, Mexico

Hadesarchaea archaeon isolates WYZ-LMO4, WYZ-LMO5, WYZ-LMO6 are archaeal species isolated from hydrothermal spring sediments. Besides high temperatures, often above 50 °C, these ecological niches usually have high salinity. Interestingly, most G-quadruplexes withstand high temperatures (their melting point is often above 70 °C) and are further stabilized by positively charged ions such a K^+ and Na^+ [84,85]. Such conditions may have naturally favored G-quadruplexes over duplexes. It also highlights one of the consequences of a high GC %: G4-prone motifs become more frequent (Figure 5). In addition, all hyperthermophilic organism genomes encode a reverse gyrase, which positively supercoil DNA, possibly to protect the genome [86]. In future studies, it would be very interesting to carry out a genome-wide wet-lab experiment, for example, direct DNA sequencing of G-quadruplex loci as described in [87,88] or direct visualization of G-quadruplexes in living cells using specific antibodies, such as BG4 [89].

5. Conclusions

Overall, our results indicate that archaea are, like eukaryotes and bacteria, prone to G-quadruplex formation: G-quadruplexes are here, there, and everywhere! Important differences in G4 densities were found among species, and experimental validation was obtained in vitro for a few candidate sequences. Follow-up studies may check if specific archaeal PQS loci—for example, in important genes, show some phylogenetic conservation. If confirmed, this could serve as a new (additional) phylogenetic marker and give us some extended clues about the evolution and function of G-quadruplex forming sequences in Archaea. This study will stimulate further studies on G4 presence in Archaea, and help to establish whether some regulatory mechanisms may only apply to a given domain or be truly universal.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-273X/10/9/1349/s1/, Figure S1: Experimental evidence for G4 formation with Hadesarchaea sequences at high temperature; Table S1: The accession codes and phylogenetic classification of all archaeal genomic DNA sequences, Table S2: Overall results of PQS frequencies found in each analyzed genomic sequence (all (A), superphylum (B) or phylum (C)) together with GC content, sequence length and other parameters, Table S3: Feature counts, Table S4: PQS characteristics used for the dendrogram shown in Figure 6, Table S5: Statistical evaluation, Table S6: BLASTp search for RecQ and NIQI in Archaea, Table S7: FIMO search for putative quadruplex binding motif, Table S8: The most similar protein of RecQ (E. coli) in Hadesarchaea archaeon.

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

Funding: This work was funded 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.

Conflicts of Interest: The authors declare no conflict of interest. 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. Woese, C.R.; Fox, G.E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Sci. Acad. USA* **1977**, *74*, 5088–5090. [CrossRef]
2. Olsen, G.J.; Woese, C.R. Archaeal genomics: An overview. *Cell* **1997**, *89*, 991–994. [CrossRef]
3. Forterre, P. Archaea: What can we learn from their sequences? *Curr. Opin. Genet. Dev.* **1997**, *7*, 764–770. [CrossRef]
4. Grüber, G.; Manimekalai, M.S.S.; Mayer, F.; Müller, V. ATP synthases from archaea: The beauty of a molecular motor. *Biochim. Biophys. Acta* **2014**, *1837*, 940–952. [CrossRef]
5. Bolhuis, A. The archaeal Sec-dependent protein translocation pathway. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2004**, *359*, 919–927. [CrossRef]
6. Samson, R.Y.; Dobro, M.J.; Jensen, G.J.; Bell, S.D. The Structure, Function and Roles of the Archaeal ESCRT Apparatus. *Subcell. Biochem.* **2017**, *84*, 357–377. [CrossRef]
7. Spang, A.; Eme, L.; Saw, J.H.; Caceres, E.F.; Zaremba-Niedzwiedzka, K.; Lombard, J.; Guy, L.; Ettema, T.J.G. Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet.* **2018**, *14*, e1007080. [CrossRef] [PubMed]
8. Da Cunha, V.; Gaia, M.; Nasir, A.; Forterre, P. Asgard archaea do not close the debate about the universal tree of life topology. *PLoS Genet.* **2018**, *14*, e1007215. [CrossRef]
9. Adam, P.S.; Borrel, G.; Brochier-Armanet, C.; Gribaldo, S. The growing tree of Archaea: New perspectives on their diversity, evolution and ecology. *ISME J.* **2017**, *11*, 2407. [CrossRef]
10. Spang, A.; Caceres, E.F.; Ettema, T.J.G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **2017**, *357*. [CrossRef]
11. Pennisi, E. Survey of archaea in the body reveals other microbial guests. *Science* **2017**, *358*, 983. [CrossRef] [PubMed]
12. Chaudhary, P.P.; Conway, P.L.; Schlundt, J. Methanogens in humans: Potentially beneficial or harmful for health. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 3095–3104. [CrossRef]
13. Vuillemin, A.; Wankel, S.D.; Coskun, Ö.K.; Magritsch, T.; Vargas, S.; Estes, E.R.; Spivack, A.J.; Smith, D.C.; Pockalny, R.; Murray, R.W. Archaea dominate oxic subseafloor communities over multimillion-year time scales. *Sci. Adv.* **2019**, *5*, eaaw4108. [CrossRef]
14. Jain, S.; Caforio, A.; Driessen, A.J.M. Biosynthesis of archaeal membrane ether lipids. *Front. Microbiol.* **2014**, *5*, 641. [CrossRef] [PubMed]
15. Nobu, M.K.; Narihiro, T.; Kuroda, K.; Mei, R.; Liu, W.-T. Chasing the elusive Euryarchaeota class WSA2: Genomes reveal a uniquely fastidious methyl-reducing methanogen. *ISME J.* **2016**, *10*, 2478–2487. [CrossRef] [PubMed]

16. Aouad, M.; Borrel, G.; Brochier-Armanet, C.; Gribaldo, S. Evolutionary placement of Methanonatronarchaeia. *Nat. Microbiol.* **2019**, *4*, 558–559. [CrossRef]
17. Forterre, P. The universal tree of life: An update. *Front. Microbiol.* **2015**, *6*. [CrossRef]
18. Dombrowski, N.; Lee, J.-H.; Williams, T.A.; Offre, P.; Spang, A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol. Lett.* **2019**, *366*, fnz008. [CrossRef]
19. Gaia, M.; Forterre, P. The Tree of Life. In *Molecular Mechanisms of Microbial Evolution (Grand Challenges in Biology and Biotechnology)*; Rampelotto, P.H., Ed.; Springer: New York, NY, USA, 2018.
20. Sun, Z.-Y.; Wang, X.-N.; Cheng, S.-Q.; Su, X.-X.; Ou, T.-M. Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to Interfering with Nucleic Acid–Protein Interaction. *Molecules* **2019**, *24*, 396. [CrossRef]
21. Harkness, R.W.; Mittermaier, A.K. G-quadruplex dynamics. *BBA Proteins Proteomics* **2017**, *1865*, 1544–1554. [CrossRef]
22. 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] [PubMed]
23. Lee, S.C.; Zhang, J.; Strom, J.; Yang, D.; Dinh, T.N.; Kappeler, K.; Chen, Q.M. G-Quadruplex in the NRF2 mRNA 5′ Untranslated Region Regulates De Novo NRF2 Protein Translation under Oxidative Stress. *Mol. Cell. Biol.* **2016**, *37*. [CrossRef] [PubMed]
24. Crenshaw, E.; Leung, B.P.; Kwok, C.K.; Sharoni, M.; Olson, K.; Sebastian, N.P.; Ansaloni, S.; Schweitzer-Stenner, R.; Akins, M.R.; Bevilacqua, P.C.; et al. Amyloid Precursor Protein Translation is Regulated by a 3′UTR Guanine Quadruplex. *PLoS ONE* **2015**, *10*. [CrossRef] [PubMed]
25. Gage, H.L.; Merrick, C.J. Conserved associations between G-quadruplex-forming DNA motifs and virulence gene families in malaria parasites. *BMC Genomics* **2020**, *21*, 236. [CrossRef]
26. Gazanion, E.; Lacroix, L.; Alberti, P.; Gurung, P.; Wein, S.; Cheng, M.; Mergny, J.; Gomes, A.; Lopez-Rubio, J. Genome wide distribution of G-quadruplexes and their impact on gene expression in malaria parasites. *PLoS Genetics* **2020**. [CrossRef]
27. Cahoon, L.A.; Seifert, H.S. An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science* **2009**, *325*, 764–767. [CrossRef]
28. Thakur, R.S.; Desingu, A.; Basavaraju, S.; Subramanya, S.; Rao, D.N.; Nagaraju, G. Mycobacterium tuberculosis DinG is a structure-specific helicase that unwinds G4 DNA implications for targeting g4 dna as a novel therapeutic approach. *J. Biol.* **2014**, *289*, 25112–25136.
29. Mishra, S.K.; Shankar, U.; Jain, N.; Sikri, K.; Tyagi, J.S.; Sharma, T.K.; Mergny, J.-L.; Kumar, A. Characterization of G-Quadruplex Motifs in espB, espK, and cyp51 Genes of Mycobacterium tuberculosis as Potential Drug Targets. *Mol. Ther. Nucleic. Acids* **2019**, *16*, 698–706. [CrossRef]
30. Brazda, V.; Haronikova, L.; Liao, J.C.; Fojta, M. DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.* **2014**, *15*, 17493–17517. [CrossRef]
31. 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]
32. Ribeyre, C.; Lopes, J.; Boulé, J.-B.; Piazza, A.; Guédin, A.; Zakian, V.A.; Mergny, J.-L.; Nicolas, A. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet.* **2009**, *5*, e1000475. [CrossRef]
33. 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] [PubMed]
34. Marguet, E.; Forterre, P. DNA stability at temperatures typical for hyperthermophiles. *Nucleic Acids Res.* **1994**, *22*, 1681–1686. [CrossRef] [PubMed]
35. 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**. [CrossRef]
36. Kota, S.; Dhamodharan, V.; Pradeepkumar, P.I.; Misra, H.S. G-quadruplex forming structural motifs in the genome of Deinococcus radiodurans and their regulatory roles in promoter functions. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 9761–9769. [CrossRef] [PubMed]

37. Mishra, S.; Chaudhary, R.; Singh, S.; Kota, S.; Misra, H.S. Guanine Quadruplex DNA Regulates Gamma Radiation Response of Genome Functions in the Radioresistant Bacterium Deinococcus radiodurans. *J. Bacteriol.* **2019**, *201*. [CrossRef] [PubMed]
38. Todd, A.K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005**, *33*, 2901–2907. [CrossRef]
39. Huppert, J.L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005**, *33*, 2908–2916. [CrossRef]
40. Eddy, J.; Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **2006**, *34*, 3887–3896. [CrossRef] [PubMed]
41. Bedrat, A.; Lacroix, L.; Mergny, J.L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016**. [CrossRef]
42. 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] [PubMed]
43. Finan, T.M. The divided bacterial genome: Structure, function, and evolution. *Microbiol. Mol. Biol. Rev.* **2017**, *81*, e00019-17.
44. Yadav, V.K.; Abraham, J.K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. QuadBase: Genome-wide database of G4 DNA-occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.* **2008**, *36*, D381–D385. [CrossRef] [PubMed]
45. Waller, Z.A.; Pinchbeck, B.J.; Buguth, B.S.; Meadows, T.G.; Richardson, D.J.; Gates, A.J. Control of bacterial nitrate assimilation by stabilization of G-quadruplex DNA. *Chem. Commun.* **2016**, *52*, 13511–13514. [CrossRef] [PubMed]
46. Rawal, P.; Kummarasetti, V.B.R.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S.K.; Chowdhury, S. Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res.* **2006**, *16*, 644–655. [CrossRef]
47. Brázda, V.; Lýsek, J.; Bartas, M.; Fojta, M. Complex Analyses of Short Inverted Repeats in All Sequenced Chloroplast DNAs. *BioMed Res. Int.* **2018**, *2018*, 1097018. [CrossRef] [PubMed]
48. Čechová, J.; Lýsek, J.; Bartas, M.; Brázda, V. Complex analyses of inverted repeats in mitochondrial genomes revealed their importance and variability. *Bioinformatics* **2018**, *34*, 1081–1085. [CrossRef] [PubMed]
49. Cahoon, L.A.; Seifert, H.S. Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in Neisseria gonorrhoeae. *PLoS Pathog.* **2013**, *9*, e1003074. [CrossRef] [PubMed]
50. Neidle, S. The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* **2009**, *19*, 239–250. [CrossRef]
51. Dhapola, P.; Chowdhury, S. QuadBase2: Web server for multiplexed guanine quadruplex mining and visualization. *Nucleic Acids Res.* **2016**, *44*, W277–W283. [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]
53. Brázda, V.; Kolomazník, J.; Lỳsek, J.; Hároníková, L.; Coufal, J.; Šťastnỳ, 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]
54. Computational Tools—Pandas 0.25.1 Documentation. Available online: https://pandas.pydata.org/pandas-docs/stable/user_guide/computation.html (accessed on 16 October 2019).
55. Suzuki, R.; Shimodaira, H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **2006**, *22*, 1540–1542. [CrossRef]
56. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
57. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017–1018. [CrossRef] [PubMed]
58. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]
59. Gertz, E.M.; Yu, Y.-K.; Agarwala, R.; Schäffer, A.A.; Altschul, S.F. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol.* **2006**, *4*, 41. [CrossRef]

60. Wernersson, R. Virtual Ribosome—A comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Res.* **2006**, *34*, W385–W388. [CrossRef]
61. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; De Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [CrossRef]
62. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [CrossRef]
63. Čutová, M.; Manta, J.; Porubiaková, O.; Kaura, P.; Šťastný, J.; Jagelská, E.B.; Goswami, P.; Bartas, M.; Brázda, V. Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae. *Genomics* **2020**, *112*, 1897–1901. [CrossRef] [PubMed]
64. Guo, J.U.; Bartel, D.P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **2016**, *353*. [CrossRef] [PubMed]
65. Galtier, N.; Tourasse, N.; Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science* **1999**, *283*, 220–221. [CrossRef]
66. Klein, R.J.; Misulovin, Z.; Eddy, S.R. Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc. Natl. Sci. Acad. USA* **2002**, *99*, 7542–7547. [CrossRef] [PubMed]
67. Lyons, S.M.; Gudanis, D.; Coyne, S.M.; Gdaniec, Z.; Ivanov, P. Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nat. Commun.* **2017**, *8*, 1127. [CrossRef]
68. Gebetsberger, J.; Zywicki, M.; Künzi, A.; Polacek, N. tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloferax volcanii*. *Archaea* **2012**, *2012*, 260909. [CrossRef]
69. Magnus, M.; Kappel, K.; Das, R.; Bujnicki, J.M. RNA 3D structure prediction guided by independent folding of homologous sequences. *BMC Bioinf.* **2019**, *20*, 512. [CrossRef]
70. Kamura, T.; Katsuda, Y.; Kitamura, Y.; Ihara, T. G-quadruplexes in mRNA: A key structure for biological function. *Biochem. Biophys. Res. Commun.* **2020**. [CrossRef]
71. Qu, Z.; Adelson, D.L. Evolutionary conservation and functional roles of ncRNA. *Front. Genet.* **2012**, *3*. [CrossRef]
72. Buddeweg, A.; Daume, M.; Randau, L.; Schmitz, R.A. Noncoding RNAs in Archaea: Genome-Wide Identification and Functional Classification. *Meth. Enzymol.* **2018**, *612*, 413–442. [CrossRef]
73. Luo, H.; Gao, F. DoriC 10.0: An updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.* **2019**, *47*, D74–D77. [CrossRef] [PubMed]
74. Cossu, M.; Da Cunha, V.; Toffano-Nioche, C.; Forterre, P.; Oberto, J. Comparative genomics reveals conserved positioning of essential genomic clusters in highly rearranged Thermococcales chromosomes. *Biochimie* **2015**, *118*, 313–321. [CrossRef] [PubMed]
75. Matsunaga, F.; Forterre, P.; Ishino, Y.; Myllykallio, H. In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11152–11157. [CrossRef] [PubMed]
76. Dueber, E.C.; Costa, A.; Corn, J.E.; Bell, S.D.; Berger, J.M. Molecular determinants of origin discrimination by Orc1 initiators in archaea. *Nucleic Acids Res.* **2011**, *39*, 3621–3631. [CrossRef]
77. Norais, C.; Hawkins, M.; Hartman, A.L.; Eisen, J.A.; Myllykallio, H.; Allers, T. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet.* **2007**, *3*, e77. [CrossRef]
78. Wu, Z.; Liu, J.; Yang, H.; Liu, H.; Xiang, H. Multiple replication origins with diverse control mechanisms in *Haloarcula hispanica*. *Nucleic Acids Res.* **2013**, *42*, 2282–2294. [CrossRef]
79. Zhuang, X.; Tang, J.; Hao, Y.; Tan, Z. Fast detection of quadruplex structure in DNA by the intrinsic fluorescence of a single-stranded DNA binding protein. *J. Mol. Recognit.* **2007**, *20*, 386–391. [CrossRef]
80. Mendoza, O.; Bourdoncle, A.; Boulé, J.-B.; Brosh, R.M.; Mergny, J.-L. G-quadruplexes and helicases. *Nucleic Acids Res.* **2016**, *44*, 1989–2006. [CrossRef]
81. Beaume, N.; Pathak, R.; Yadav, V.K.; Kota, S.; Misra, H.S.; Gautam, H.K.; Chowdhury, S. Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: Radioresistance of D. radiodurans involves G4 DNA-mediated regulation. *Nucleic Acids Res.* **2013**, *41*, 76–89. [CrossRef]
82. Gehring, K.; Leroy, J.-L.; Guéron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **1993**, *363*, 561–565. [CrossRef]

83. Barrett, T.; Clark, K.; Gevorgyan, R.; Gorelenkov, V.; Gribov, E.; Karsch-Mizrachi, I.; Kimelman, M.; Pruitt, K.D.; Resenchuk, S.; Tatusova, T.; et al. BioProject and BioSample databases at NCBI: Facilitating capture and organization of metadata. *Nucleic Acids Res.* **2012**, *40*, D57–D63. [CrossRef] [PubMed]
84. Bartas, M.; Brázda, V.; Karlický, V.; Červeň, J.; Pečinka, P. Bioinformatics analyses and in vitro evidence for five and six stacked G-quadruplex forming sequences. *Biochimie* **2018**, *150*, 70–75. [CrossRef]
85. Risitano, A.; Fox, K.R. Stability of Intramolecular DNA Quadruplexes: Comparison with DNA Duplexes. *Biochemistry* **2003**, *42*, 6507–6513. [CrossRef]
86. Couturier, M.; Gadelle, D.; Forterre, P.; Nadal, M.; Garnier, F. The reverse gyrase TopR1 is responsible for the homeostatic control of DNA supercoiling in the hyperthermophilic archaeon Sulfolobus solfataricus. *Mol. Microbiol.* **2020**, *113*, 356–368. [CrossRef] [PubMed]
87. Chambers, V.S.; Marsico, G.; Boutell, J.M.; Di Antonio, M.; Smith, G.P.; Balasubramanian, S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **2015**, *33*, 877. [CrossRef]
88. Hänsel-Hertsch, R.; Spiegel, J.; Marsico, G.; Tannahill, D.; Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **2018**, *13*, 551. [CrossRef] [PubMed]
89. Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S. DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell. Biol.* **2017**, *18*, 279. [CrossRef]

Brázda et al. BMC Genomics (2021) 22:77
https://doi.org/10.1186/s12864-021-07377-9

BMC Genomics

RESEARCH ARTICLE Open Access

G-quadruplexes in H1N1 influenza genomes

Václav Brázda[1,2*†], Otília Porubiaková[1,2†], Alessio Cantara[1,3], Natália Bohálová[1,3], Jan Coufal[1], Martin Bartas[4], Miroslav Fojta[1] and Jean-Louis Mergny[1*]

Abstract

Background: Influenza viruses are dangerous pathogens. Seventy-Seven genomes of recently emerged genotype 4 reassortant Eurasian avian-like H1N1 virus (*G4*-EA-H1N1) are currently available. We investigated the presence and variation of potential G-quadruplex forming sequences (PQS), which can serve as targets for antiviral treatment.

Results: PQS were identified in all 77 genomes. The total number of PQS in *G4*-EA-H1N1 genomes was 571. Interestingly, the number of PQS per genome in individual close relative viruses varied from 4 to 12. PQS were not randomly distributed in the 8 segments of the *G4*-EA-H1N1 genome, the highest frequency of PQS being found in the NP segment (1.39 per 1000 nt), which is considered a potential target for antiviral therapy. In contrast, no PQS was found in the NS segment. Analyses of variability pointed the importance of some PQS; even if genome variation of influenza virus is extreme, the PQS with the highest G4Hunter score is the most conserved in all tested genomes. G-quadruplex formation in vitro was experimentally confirmed using spectroscopic methods.

Conclusions: The results presented here hint several G-quadruplex-forming sequences in *G4*-EA-H1N1 genomes, that could provide good therapeutic targets.

Keywords: Influenza virus, G-quadruplex, G4Hunter

Background

Influenza viruses are deadly pathogens for humans, and more generally mammals, as well as avian species. They belong to the *Orthomyxoviridae* family and are classified into three types termed Influenza A, B and C. Among these, influenza A viruses (IAVs) pose the greatest threat to human and animal health. IAV genome is divided to 8 segments of negative-sense RNA that encodes 11 proteins [1]. Subtype classification of *G4*-EA-H1N1 is based on the antigenicity of the two major cell surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). HA protein facilitates binding of the virus to host cell receptors and subsequent endosomal fusion [2], and NA protein is responsible for binding to cellular receptors and fusion of the viral membranes, causing replication and transcription of viral RNAs in the infected host [3, 4]. The viral RNA genome (gRNA) is transcribed into mRNA and replicated through an intermediate RNA to produce a large quantity of progeny gRNA. These NAs are synthesized by the viral RNA-dependent RNA polymerase complex – polymerase basic protein 2 (PB2), polymerase basic protein 1 (PB1) and polymerase acidic protein (PA), the nucleoprotein (NP), the matrix protein (M) and the non-structural protein (NS) [5, 6].

Roots of virus H1N1 can be traced to 1918, when an avian virus overcame the species barrier to infect humans [7]. That was the beginning of a pandemic that resulted in an estimated 50 to 100 million deaths. Thereafter, influenza viruses rapidly diverged antigenically and three years later this virus was replaced by a new strain. Reassortment of influenza viruses is a major mechanism

* Correspondence: vaclav@ibp.cz; mergny@ibp.cz
†Václav Brázda and Otília Porubiaková contributed equally to this work.
[1]Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic
Full list of author information is available at the end of the article

to generate progeny viruses with novel antigenic and biological characteristics [8, 9]. The emerged genotype 4 reassortant Eurasian avian-like H1N1 virus (*G4*-EA-H1N1) has become predominant in swine populations since 2016 [10] and is a new cause of concern.

Guanine quadruplexes (G4) are local nucleic acid structures formed by G-rich DNA and RNA in which four guanines fold in a planar arrangement through Hoogsteen hydrogen bonds [11, 12]. Putative quadruplex sequences (PQSs) contribute to the regulation of key biological processes [13] and have been found in the genomes of viruses (reviewed in: [14]). For example, it has been demonstrated that G-quadruplexes regulate HIV transcription and can be targeted by small compounds called G4 ligands. A comprehensive database of PQS in human all human viruses found with the Quadparser algorithm has been published [15] but these new H1N1 strains were not available at that time.

Here we analyzed 77 newly sequenced variations of H1N1 influenza virus emerged during the last years with a different algorithm, G4Hunter. There are accessible several tools to analyze PQS in genomic sequences (reviewed in [16]). We used the G4Hunter algorithm where G4 propensity is calculated depending on G richness and G/C skewness and PQS are evaluated quantitatively [17] and validated experimentally [17, 18]. We used a new G4Hunter algorithm implementation, which is suitable for batch and full genomes analyses [19, 20] and accessible as the web-tool G4Hunter web [21]. Analyses of the human genome revealed the presence of many G4-prone sequences and G4 presence has been demonstrated in a variety of species, including eukaryotes, bacteria, archaea or viruses both in silico [19, 20, 22] and confirmed experimentally [17, 23, 24]. G4 have been shown to participate in cellular and viral replication, recombination and control of gene expression [25–27]. In addition, DNA aptamers that adopt a quadruplex fold have been described as inhibitors and diagnostic tools to detect viruses [28].

In this article, we analyzed 77 *G4*-EA-H1N1 virus genomes for G-quadruplex occurrence, localization and variance to provide a rational background for PQS targeting in antiviral influenza therapy approaches.

Results

We analyzed 616 sequences in total belonging to 77 strains of *G4*-EA-H1N1. The genome of *G4*-EA-H1N1 is 13,133 nt long and consists of 8 different segments: PB1, PB2, M, HA, NP, NS, PA and NA. PQS frequencies were analyzed according to individual *G4*-EA-H1N1 strains, and for statistical comparison we have grouped genomes according to regions of origin (10 groups based on [10]) and also according to their genomic segments (8 segments). The average GC content for the entire list of viruses is 43.37%, with minimal differences between strains, from 43.20% in the Heilongjiang strain to 43.44% in the Shandong strain. Using standard default values for the G4Hunter algorithm (window size of 25 nucleotides and G4Hunter score above 1.2), 571 PQSs were found among all genomes and all fragments. Mean PQS frequency for the whole set of sequences was 0.56 PQS per 1000 nt and PQSs cover an average of 1.58% of *G4*-EA-H1N1 genomes. The mean number of PQS per *G4*-EA-H1N1 genome was 7.42. The highest number of PQS was found in Swine Beijing 0301 2018 strain with a total of 12 PQSs, giving a PQS frequency of 0.91 PQS per 1000 nt. The lowest frequency (0.30 PQS per 1000 nt) was found in Swine Shandong S113 2014 and Swine Shandong JM78 2017 strains, where only 4 PQS with a G4Hunter score above 1.2 were found. Genomic sequence sizes, GC count, and PQS characteristics are summarized in **Table 1**, all results for individual species and groups are in SM_02A.

Our analyses showed that PQS frequencies of *G4*-EA-H1N1 were significantly different for the Shandong group (compared to Hebei ($p = 0.016$), Jiangsu ($p = 0.047$), Liaoning ($p = 0.0041$) groups), and for the Liaoning group (compared to Henan ($p = 0.025$) and Heilongijang groups ($p = 0,031$)) (available in SM_03). Graphical representation of PQS frequencies is shown in Fig. 1.

We also performed PQS analyses of individual segments of influenza genomes (**Table 2.**); all results for segments are shown in SM_02B. Even if the global GC content in all species is very conserved, the GC content within each segment is more variable - from 41.16% in the HA segment to 47.34% in the M segment. Despite the highest CG content in HA segment, the highest mean PQS frequency was found in the NP segment (with a GC content of 46.23%), with the highest number of PQS (160). It was followed by segments NA (149 PQS) and PB2 (79 PQS). On the other hand, no PQS was found in the NS segment (which codes the non-structural protein) with a GC content of 41.52%. These data are pointing to possible functional importance of G-quadruplex in IAV genomes. All the species have 1, 2 or 4 PQS in segment NP, except for Swine Shandong LY142 2017, which does not contain any PQS with a G4Hunter score above 1.2. IAV belong to the negative-sense single-stranded RNA viruses group. Interestingly, the PQS were not distributed equally among minus gRNA which is copied for protein production (mRNA). Most of the PQSs are located in its mRNA (498 compare to 73 in gRNA). Moreover, in PB1, PB2, NP and NA segments PQS are exclusively found in mRNA (**Table 2**).

The distribution of G4Hunter score parameters for all PQSs found in *G4*-EA-H1N1 segments is summarized in **Table 3**. As previously found in eukaryotes, bacteria and

Table 1 Strains of *G4*-EA-H1N1: genomic sequences sizes, PQS frequency and total counts of PQS. Seq (number of strains), Length (length of the sequence, nt), GC % (average GC content), PQS (total number of predicted PQS), Mean PQS (mean number of predicted PQS), Min PQS (lowest number of predicted PQS), Max PQS (highest number of predicted PQS), PQS frequency (PQS frequency per 1000 nt), Cov% (% of genome covered by PQS).

	Seq	Length	GC%	PQS	Mean PQS	Min PQS	Max PQS	PQS frequency	Cov%
All	77	13 133	43.37	571	7.42	4	12	0.56	1.58
Strain	**Seq**	**Length**	**GC%**	**PQS**	**Mean PQS**	**Min PQS**	**Max PQS**	**PQS frequency**	**Cov%**
Anhui	3	13 133	43.25	24	8.00	6	9	0.61	1.64
Beijing	7	13 133	43.39	55	7.86	6	12	0.60	1.70
Hebei	11	13 133	43.35	87	7.91	6	10	0.60	1.68
Heilongjiang	3	13 133	43.20	19	6.33	6	7	0.48	1.30
Henan	13	13 133	43.38	92	7.08	5	9	0.54	1.51
Jiangsu	4	13 133	43.29	33	8.25	6	9	0.63	1.71
Jilin	6	13 133	43.39	45	7.5	6	9	0.57	1.61
Liaoning	6	13 133	43.41	53	8.83	7	10	0.67	1.88
Shandong	17	13 133	43.44	108	6.35	4	9	0.48	1.34
Tianjin	7	13 133	43.30	55	7.86	5	9	0.60	1.70

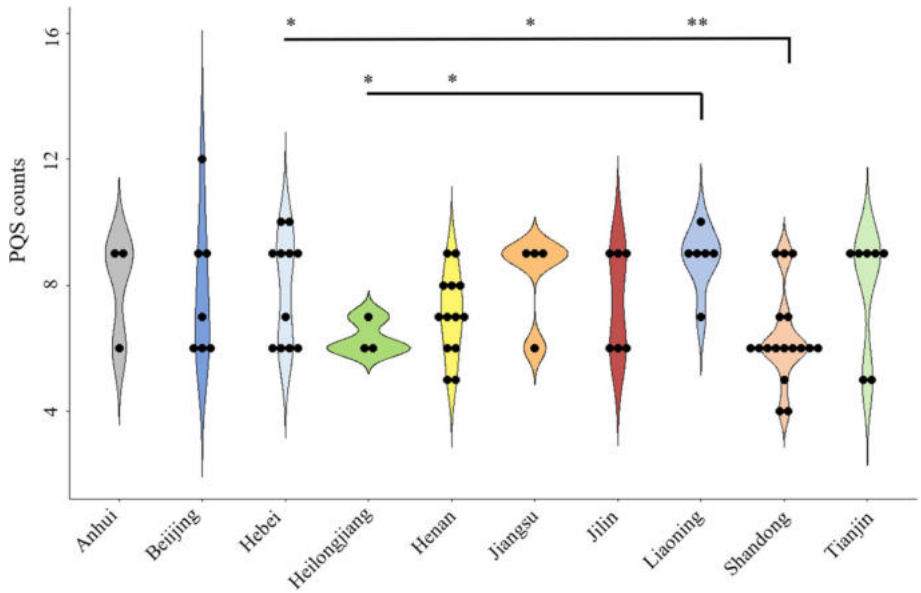


Fig. 1 Violin plot of PQS number in *G4*-EA-H1N1 groups (SM_03). The significant differences between groups are depicted by asterisks (*p*-value < 0.05 is *; *p*-value < 0.01 is **)

Table 2 Segments of *G4*-EA-H1N1: genomic sequences sizes, PQS frequencies and total counts of PQS. Seq (total number of sequences), Length (median length of sequences), GC % (average GC content), PQS (total number of predicted PQS), Mean PQS (mean number of predicted PQS per sequence), Min PQS (lowest number of predicted PQS per sequence), Max PQS (highest number of predicted PQS per sequence), mRNA-gRNA (viral messenger RNA-viral genome RNA), PQS frequency (PQS frequency per 1000 nt), Cov% (% of genome covered by PQS).

	Seq	Length	GC%	PQS	Mean PQS	Min PQS	Max PQS	PQS frequency	Cov%
All	616	13133	43.37	571	0.93	0	4	0.56	1.58
Segment	**Seq**	**Length**	**GC%**	**PQS**	**Mean PQS**	**Min PQS**	**Max PQS (mRNA-gRNA)**	**PQS frequency**	**Cov%**
PB1	77	2274	42.39	41	0.53	0	1 (1-0)	0.23	0.73
PB2	77	2280	43.51	79	1.03	0	3 (3-0)	0.45	1.23
M	77	982	47.34	69	0.9	0	2 (0-2)	0.91	2.29
HA	77	1701	41.16	27	0.35	0	2 (1-1)	0.21	0.55
NP	77	1497	46.23	160	2.08	0	4 (4-0)	1.39	3.69
NS	77	838	43.36	0	0	0	0	0	0
PA	77	2151	43.42	46	0.60	0	3 (2-1)	0.28	0.71
NA	77	1410	41.52	149	1.94	1	3 (3-0)	1.37	4.25

viruses [19, 20, 22], most of the PQS have relatively low G4Hunter scores (in the 1.2–1.4 range). Only 10 / 571 motifs have a G4Hunter score above 1.4 (all in the HA segment), and no PQS was found with a G4Hunter above 1.6.

Detailed statistical characteristics for PQS frequencies per 1000 nt, including mean, variance, and outliers, are depicted in boxplots for segments are shown in Fig. 2. Statistical evaluation of PQS in IAV segments showed the statistical differences for all comparisons except for three cases (PB1 *vs*. HA, PB1 *vs*. PA, and PB2 *vs*. M) for which differences were not significantly different.

We evaluated the localization of G4 prone sequences in the genome of Swine Beijing 0301 2018, where we found the highest number of PQS (Fig. 3.). From a total of 12 PQS found, 3 PQSs were in the PB2 and NA segments, 2 PQSs were located in the NP and PA segments and 1 PQS was found in the M and HA segments. The majority of PQS were found in mRNA. Ten out of all PQS were located in mRNA (with positive G4Hunter

Table 3 PQSs in *G4*-EA-H1N1 segments grouped by G4Hunter score (absolute values). Frequency was computed using total number of PQSs in each category divided by total length of all analyzed sequences and multiplied by 1000, the total number of PQS are in brackets.

PQS frequency per 1,000 nt (number of PQS)								
	PB1	**PB2**	**M**	**HA**	**NP**	**NS**	**PA**	**NA**
1.2-1.4	0.23 (41)	0.45 (79)	0.91 (69)	0.13 (17)	1.39 (160)	0	0.28 (46)	1.37 (149)
1.4-1.6	0	0	0	0.08 (10)	0	0	0	0
>1.6	0	0	0	0	0	0	0	0

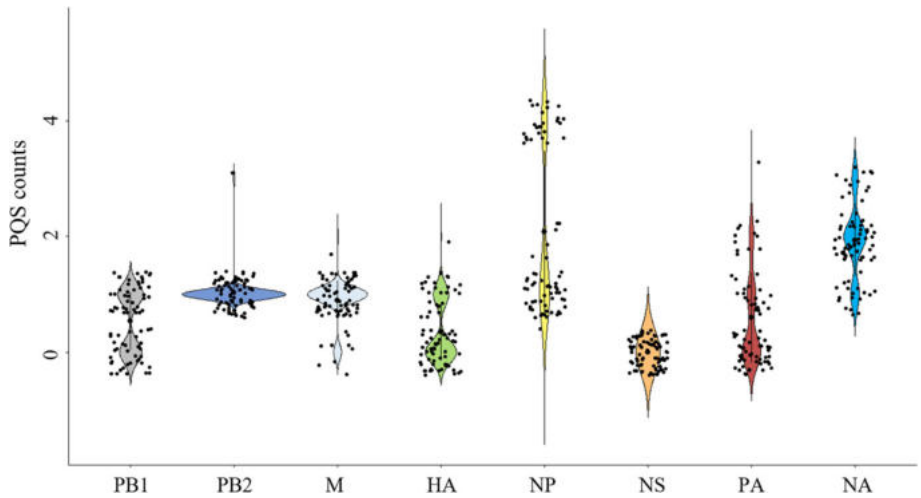


Fig. 2 Violin plots of PQS number in *G4*-EA-H1N1 segments (SM_03). All 28 inter comparisons were significant with *p*-values < 0.05, except for PB1 vs. HA, PB1 vs. PA, and PB2 vs. M

score), whereas only 2 PQS were located in negative genomic RNA (with negative G4Hunter score). Interestingly, in segment M, one PQS was located at the 3′ end of intron in negative-sense genomic RNA, near the splicing site of mRNA, which encodes M2 protein. M segment codes 2 matrix proteins – M1, which is coded by whole segment and spliced protein M2 [29]. All 10 conserved PQSs located in positive-sense RNA completely span coding regions; this is hardly surprising, as the vast majority of RNA segments are protein coding, except for short 3′ and 5′ UTRs.

A comparison of genomes revealed that some, but not all, PQS motifs were highly conserved. We align all predicted PQS and generate their LOGO representation (SM_04). Selected LOGO sequences with the highest positive and negative G4Hunter scores and with the most variable nucleotides are shown in Fig. 4**.** For example, in the M and HA segments, we found PQS in which only 1 nucleotide (out of 25 and 27, respectively) is variable within the PQS motif among all 77 strains. In contrast, other PQS sequences were poorly conserved / extremely variable (for example, the PQS sequence "**C**" in the NP segment has 12 / 26 variable nucleotides in its PQS; this can lead to significant variations in G4Hunter score and quadruplex propensity).

Overall, *G4*-EA-H1N1 genomic sequences are very variable. The analyses of 77 *G4*-EA-H1N1 genomes show a global variation of 23.4%. Therefore, the high sequence conservation of some PQS (two of them have a variation < 4.0% in Fig. 4) suggests they play crucial roles in influenza virus. The PQS sequence with the highest G4Hunter score is also the most conserved among all found PQS. Similarly, another sequence with two GGGG runs (Fig. 4d), which could form bimolecular G4, has 100% conservation within the G-tracts.

We then determined if the quadruplex-prone sequences identified in silico actually form G4 in vitro. This experimental confirmation is important for these motifs, as their G4-Hunter scores are relatively low, and some candidate sequences may prefer formation of other structures and/or fail to form stable G4 (100% confidence in predicted motifs can only be achieved for relatively high scores, typically above 1.6). To confirm the ability of the most conserved PQS to form G4 in vitro, we used a combination of two biophysical methods, circular dichroism (CD) spectroscopy and the Thioflavin T (ThT) fluorescent assay [30, 31], results are shown in SM_06. We tested nine synthetic oligonucleotides derived from the LOGO sequence listed in Fig. 4. For sequences A, C and E we analyzed two variants, one with the highest and one with the lowest possible G4Hunter score. Quadruplex formation was confirmed for 5 out of 8 analyzed sequences (**Table 4**). G-quadruplex formation in vitro was confirmed by CD spectroscopy as the shift of the peak from 270 to 264 nm and a stronger signal in the presence of K^+ ions (potassium ions stabilize the G4 structure). An example of positive result is presented in Fig. 5**, part A** for a conserved sequence derived from HA fragment and in Fig. 5**, part C** for the sequence from NP fragment with the highest possible G4Hunter score. An example of negative result acquired by CD spectroscopy is shown in Fig. 5**, part B** for a negative control sequence with the G4Hunter score of 0.37 and in Fig. 5**, part D** for the sequence derived from the NP fragment with the lowest possible G4Hunter score.

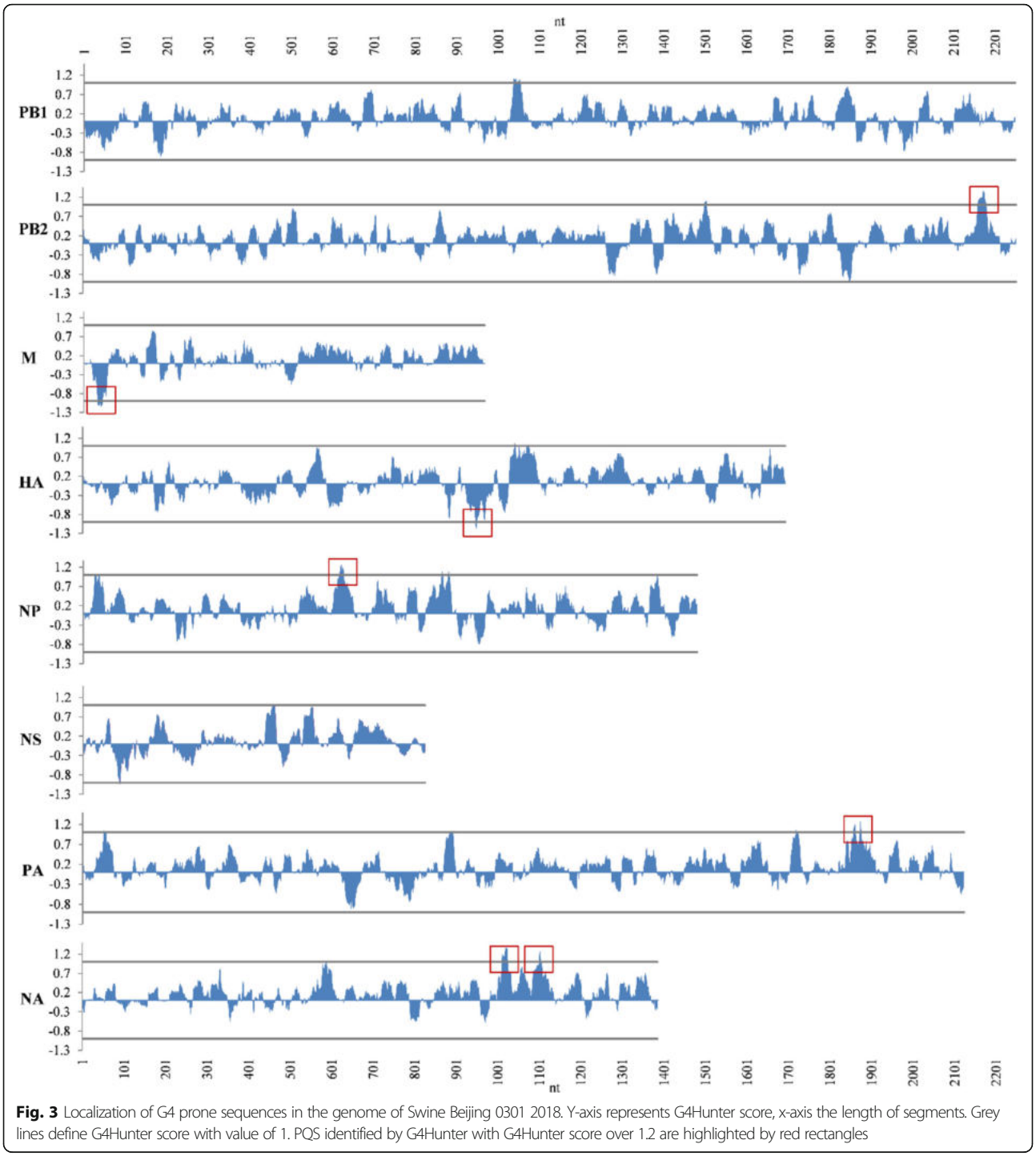


Fig. 3 Localization of G4 prone sequences in the genome of Swine Beijing 0301 2018. Y-axis represents G4Hunter score, x-axis the length of segments. Grey lines define G4Hunter score with value of 1. PQS identified by G4Hunter with G4Hunter score over 1.2 are highlighted by red rectangles

Discussion and conclusions

The influenza viruses pose a global public health concern. Influenza claims 250,000–500,000 lives annually, even though vaccines and antiviral drugs are available. There is therefore an urgent need to develop antiviral drugs with novel mechanisms of action. Noncanonical nucleic acid structures play an important role in basic biological processes [32] and it has been shown that G4s may be used as targets for therapy [33, 34]. Therefore, noncanonical structures in the H1N1 viral genome could serve as possibly targets for antiviral therapy. In this study, we provide a detailed analysis of PQSs occurrences, frequencies and distributions in the contemporary emerged *G4*-EA-H1N1 strains.

We found a total number of 571 PQS in all 77 *G4*-EA-H1N1 genomes. Interestingly, the number of PQS in

		LOGO sequence	Total length	Conserved	Variable	Variable %
A	HA		27	26	1	3.7
B			26	16	10	38.5
C	NP		28	16	12	42.9
D	NA		31	26	5	19.2
E	M		25	24	1	4.0

Fig. 4 Examples of PQS motifs and their variation presented as LOGO sequences **a**. PQS with the highest G4Hunter score (1.4), **b**. PQS with the lowest G4Hunter score (– 1.2; a negative score indicates that the G-rich motif is located in negative gRNA), **c**. PQS with the most variable sequence (G4Hunter score 1.2) from NP segment, **d**. PQS with conserved GGGG-tracks (1.1) and **e**. PQS conserved sequence (– 1.2,). Perfectly conserved nucleotides are represented by full size letters. All sequence logos are shown in SM_04

close *G4*-EA-H1N1 relatives varied from 4 to 12. Analyses of variability pointed to the importance of some PQS: even if genome variation of influenza virus is extreme, the PQS with the highest G4Hunter score is nearly perfectly conserved in all tested genomes. Comparison of segments shows significant differences among individual *G4*-EA-H1N1 segments. While the highest mean PQS frequency was found in the NP segment (1.39), which codes for a protein playing a central role in viral replication [35], the most abundant viral protein in infected cells [36, 37] and the most promising drug target [37] – no PQS was found in the NS segment (which codes for the non-structural NS protein).

To evaluate the presence of the PQS in individual fragments we randomize five-times the RNA sequences of the Liaoming group (the group with highest PQS frequency) and as well in the Shandong group (the group with the lowest PQS frequency). A significant difference

Table 4 Summary of the in vitro G4 formation analyses by CD spectroscopy and ThT fluorescent assay in vitro. Sequences are shown in the 5′ to 3′ direction; all oligonucleotides are RNA. For G4 formation by CD, "Yes" indicates that a CD signature typical of a parallel G4 structure in the presence of K^+. The result of CD spectroscopy was considered positive in the case of a blue-shift of the positive ellipticity peak (from 270 to 264 nm) and a stronger signal in the presence of K^+ ions. Ratio between ThT fluorescence in the presence of oligonucleotide and background fluorescence of ThT alone is presented in the last column. The light-up effect (("Fold of ThT") refers to fold increase in Thioflavin T fluorescence emission when the candidate sequence is added: the higher this increase, the more likely is the structure to form a G4 motif.

Name	Sequence	G4Hunter score	G4 formation by CD	Fold of ThT
A0	GGGGAAGGAAGUGCUUGUAAUUUGGGG	1.407	Yes	7.23
A1	GGGAAAGGAAGUGCUUGUAAUUUGGGG	1.148	Yes	8.79
B0	UGGGGCAUUUUCCGAGGGUGACGGGA	1.154	Yes	7.90
C0	GAAGGGGGUAGGGACAAUGGUGAUGGAG	1.393	Yes	12.06
C1	AAAAGGAAUAGGAACAAUGGUAAUGGAG	0.571	No	6.23
D0	UGGGGCAAAUGGGGU	2.067	Yes	4.95
E0	GGCUUUGAGGGGGCCUGACGGG	1.318	No	4.28
E1	GGCUUUGAGGGGGCCUGACGGU	1.091	No	4.40
NK	UCUGACCGACCUAGAU	0.37	No	1.69

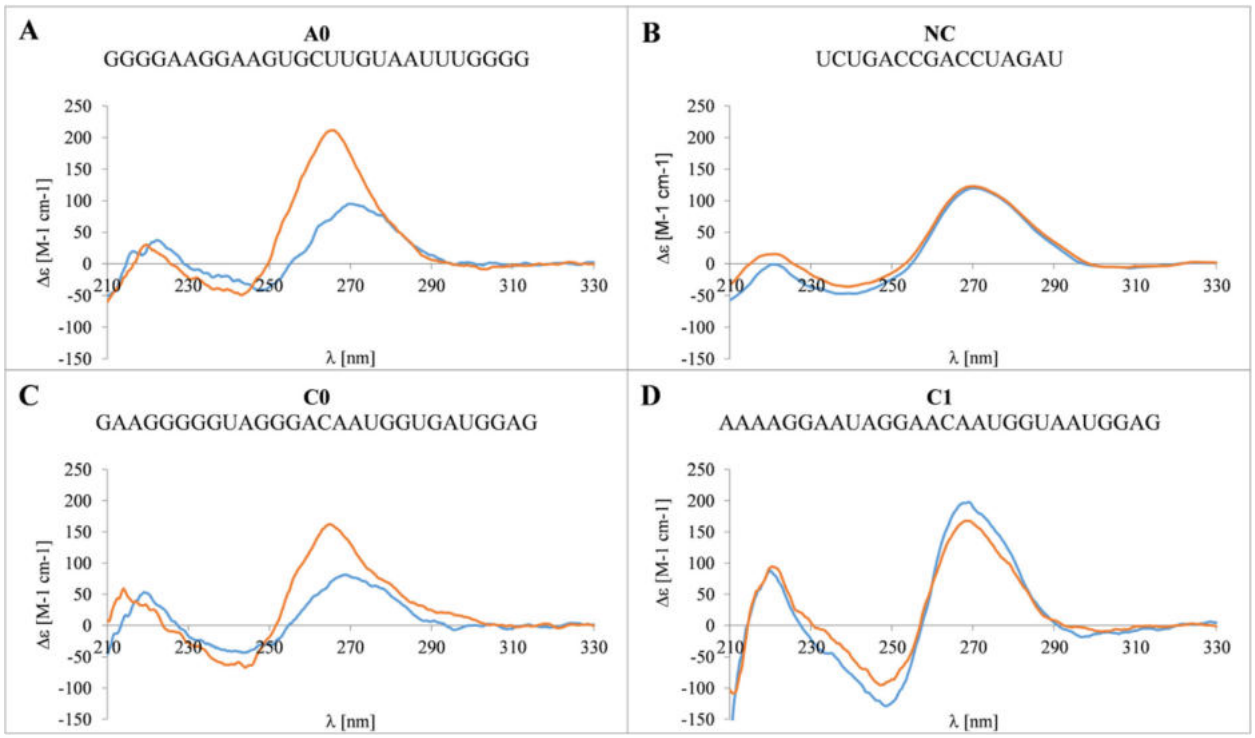


Fig. 5 Circular dichroism (CD) spectra of selected PQS in 1 mM sodium phosphate buffer (pH 7) (blue lines) or in 1 mM sodium phosphate (pH 7), 10 mM potassium phosphate (pH 7), 90 mM KCl buffer (orange lines); **a**. Oligonucleotide AO **b**. Negative control (NC); **c** Oligonucleotide CO; **d** Oligonucleotide C1

in PQS frequency was found between reference and randomized sequences for the NA segment of both groups (SM_05). On the other hand, the PQS frequency was not significantly different for other frogments, except for a depletion in the PA fragment in the Shandong group. These results are in agreement with recently proposed hypotheses that viruses causing acute infection are depleted in PQS (Bohálová N, Cantara A, Bartas M, Kaura P, Šťastný J, Pečinka P, Fojta M., Mergny J-L., Brázda V.: Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection (submitted). A similar finding was published for SARS-CoV-2 [22]. On the other hand, the abundance of PQS in in the NA segment suggests its important evolutionarily conserved function.

Of note, none of the PQS identified here match a classical quadruplex consensus, in which four runs of three or more guanines are separated by 7 nucleotides or less, as predicted by Quadparser [38] with default parameters. As RNA G4 structures tend to be more stable than DNA, some of the motifs found here are still likely to form quadruplexes under physiological conditions, and this was experimentally confirmed using a combination of two biophysical methods. Given that all G4Hunter scores were relatively low, G4 formation was not a given, and needed the experimental confirmation. Our results show that the most conserved PQS in HA fragment, one with a conserved G run in NA fragment, as well as some others are capable to form G-quadruplex structure in vitro, as shown by CD spectroscopy and by Fluorescence light-up measurements, while two sequence with low G4Hunter scores (< 0.6) did not form stable quadruplexes at room temperature. Interestingly, and as observed previously, the "grey zone" for which a sequence may well form a quadruplex or not seems to be centered around 1.1–1.2, and we have several sequences with relatively similar scores (between 1.09 and 1.31) which give different outcomes. G4Hunter is therefore not perfect – as all current prediction tools – and we are currently working on modifying parameters to improve accuracy. This may prove more difficult for RNA than for DNA, as we currently have access to far more experimental data on DNA than on RNA oligonucleotides.

In contrast to Quadparser, G4Hunter does not pick individual G-tracts to propose a core quadruplex with three loops. As demonstrated by a number of studies, the universe of G4-forming sequences is very diverse, and may involve bulges or snapback motifs, allowing individual, isolated guanines to participate in G-quadruplex formation. There are, of course, specific cases in which G participating to G-quartets can reasonably be assumed. For example, a bimolecular four-layer G4 motif can be predicted within the D motif for the GGGGCAAAUGGGG region. In addition, for all motifs, one can imagine an intramolecular structure, provided

that *i)* two-layered RNA G-quadruplexes are stable, and that *ii)* zero-nucleotide loops are allowed (such propeller loops have been found in a limited number of cases). In addition, one cannot exclude that isolated G also contribute to the core quartets. For this reason, in the absence of high resolution structures, it is rather premature to propose which G within these motifs are involved in G4 formation. These observations further illustrate that it is not possible to cover all G4-forming motifs with a single general consensus sequence.

These structures may offer opportunities for regulation and targeting by G4 ligands. Interestingly, several conserved PQS contain two GGGG runs, which may allow stable bimolecular G4 formation, as suggested for genome organization in other viruses including SARS-CoV-2 [39, 40].

Both strands, negative-sense genomic RNA and positive-sense mRNA, were analyzed for the presence and distribution of PQS as both RNAs are involved in lifecycle of the virus. Our result show that PQS are not evenly distributed but are mostly located on the RNA positive strand, thereby they may be involved in translation and splicing regulation. The genome of IAV is not stable and varies remarkably among strains [10, 41]. Comparison of PQS in various strains demonstrated that several PQS in the M segment and HA segment are highly conserved and therefore may be considered as suitable candidate targets with therapeutic potential. The HA segment (hemagglutinin) codes for a primary viral protein, which is recognized by the immune system and also is the primary target for vaccine design [42]. HA contains two subunits: HA1, which is responsible for receptor binding and HA2, which function is to support HA1 and mediates membrane fusion during viral entry [43]. Moreover, the conserved PQS in the HA sequence has the highest G4Hunter score among all found PQS. Another highly conserved PQS (just 1 variable nucleotide as shown in Table 3) was found at the 3′ end of an intron in negative-sense genomic RNA, near the splicing site of M2 protein. M1 is the only viral structural component which plays a major role in virus particle assembly [44]. M2 is a transmembrane ion channel protein which plays an important role in early stages of viral entry. Moreover, the M segment of 2009 H1N1 pandemic influenza virus was derived from the Eurasian avian-like swine lineage and was shown to affect neuraminidase activity and therefore might also have a potent effect on transmissibility [45].

Compared to strong depletion of PQS in the contemporarily outbreaking SARS-CoV-2 virus [22, 46], the genomes of IAV contain a highly conserved PQS, which could serve as a selective target. Our comprehensive analyses confirmed that several candidates adopt a quadruplex fold, arguing for the therapeutic potential of these PQS as targets for specific ligands.

Methods

Source of DNA sequences

The complete set of 616 sequences of 77 *G4*-EA-H1N1 genomes (each genome is divided into 8 segments) was downloaded on July 3, 2020 from the Genome database of the National center for Biotechnology Information (NCBI) [47]. NCBI accessions are shown in SM_01.

Process of analysis

All sequences belonging to 77 strains of *G4*-EA-H1N1 were analyzed with the G4Hunter Web tool [21], which is capable to read the NCBI identifier of the sequences uploaded in a .csv files. Default parameters for G4Hunter were set to "25" for window size and 1.2 or above for G4Hunter threshold score. G4Hunter score was grouped to the five intervals: 1.2–1.4, 1.4–1.6, 1.6–1.8, 1.8–2.0 and 2.0 and more, as previously performed for the Bacteria domain [19]. PQS frequencies were analyzed according to individual *G4*-EA-H1N1 strains, grouped according to regions of isolation and according to their eight genomic segments. All results including information about the size of genomic DNA sequence, number of PQS and statistical data are shown (SM_02A – grouped by region and SM_02B – grouped by segments).

Statistical evaluation

Statistical analysis of normality was made with the Shapiro-Wilk test. Since it was found that the data do not have a normal distribution, we used Kruskal-Wallis signed rank test to evaluate significant differences among strains and segments. 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 SM_03.

The statistical significance analysis of the found PQS with respect to the same genome composition in a scrambled order was performed. The strains with the highest and lowest PQS frequency were selected. Reference sequences of individual segments of selected strains were five times randomized by the program Sequence Manipulation Suite – Shuffle DNA [48]. PQS frequencies were evaluated equally in reference and randomized sequences and plotted (SM_05).

Construction of LOGO sequences

NCBI sequences by list from SM_01 in FASTA format were downloaded and the dataset was uploaded to SnapGene (Align Multiple DNA Sequences) program. For every PQS we used the corresponding sequences from all *G4*-EA-H1N1 genomes and alignments by Clustal Omega tool [49] were generated. All found PQS were

searched in aligned sequences and WebLogo 3 [50] was used for generating LOGO sequences, all predicted PQS and LOGO sequences for PQS are available in SM_04.

CD spectroscopy

Synthetic oligonucleotides were purchased from Sigma-Aldrich and diluted in water. Oligonucleotides were heated at 85 °C for 3 min in 1 mM sodium phosphate buffer pH 7 or 1 mM sodium phosphate pH 7, 10 mM potassium phosphate pH 7 and 90 mM KCl buffer and slowly cooled down to room temperature. CD measurements were carried out with a Jasco 815 (Jasco International Co., Ltd.,Tokyo, Japan) dichrograph in 1 cm path-length microcells at 23 °C. A set of four scans with a data pitch of 0.5 nm and 200 nm/min scan speed was averaged for each sample. CD signal is expressed as the difference in molar absorption, Δε, of the left- and right-handed circularly polarized light [31].

ThT fluorescent assay

Synthetic oligonucleotides were purchased from Sigma Aldrich and diluted in water. Oligonucleotides were diluted to 2 μM final concentration in 100 mM Tris-HCl pH 7.5 and 100 mM KCl buffer, heated at 85 °C for 3 min and slowly cooled down to room temperature. ThT was diluted in water to 1 μM final concentration and 100 mM KCl was added. Experiments were performed in a 384-well microplate from CORNING (Flat Bottom Black Polyester). Each condition was tested in triplicate. Measurements were performed at room temperature. Oligonucleotides and ThT were mixed at 0.5:1 M ratio to a final volume of 20 μl. Fluorescent emission was collected at 490 nm after excitation at 425 nm with a microplate reader (Spark, Tecan) [30].

Abbreviations

G4-EA-H1N1: Genotype 4 reassortant Eurasian avian-like H1N1 virus; PQS: Potential quadruplex-forming sequence; G4: Quadruplex; g-RNA: Viral RNA genome; IAV: Influenza A virus; HA: Hemagglutinin; NA: Neuraminidase; PB1: Polymerase basic protein 1; PB2: Polymerase basic protein 2; PA: Polymerase acidic protein; NP: Neuroprotein; M: The matrix protein; NS: The non-structural protein; CD: Circular dichroism; ThT: Thioflavin T

Acknowledgements

Not Applicable.

Authors' contributions

AC and NB performed analysis of dataset. OP collected and interpreted results of analysis and wrote the initial draft of the manuscript. NB and JC performed experimental methods. VB and MB revised the draft and wrote notes of the final content. JLM and MF designed the study and contributed to writing and critically edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the SYMBIT project (reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477) financed by the ERDF.

Availability of data and materials

Data availability: all data are enclosed in the Supplementary materials (SM).

SM_01: NCBI accession numbers for all tested sequences.
SM_02: G4Hunter analyses results (A) strains (B) fragments.
SM_03: Statistical evaluation.
SM_04: PQS LOGO representation.
SM_05: Statistical analysis of randomized sequences (A) Shandong (B) Liaoning.
SM_06: Circular dichroism (CD) spectroscopy and the Thioflavin T (ThT) fluorescent assay results.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic. [2]Faculty of Chemistry, Brno University of Technology, Purkyňova 118, 612 00 Brno, Czech Republic. [3]Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic. [4]Department of Biology and Ecology/Institute of Environmental Technologies, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic.

Received: 6 September 2020 Accepted: 8 January 2021
Published online: 23 January 2021

References

1. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza a epidemic. Nature. 2009;459(7250):1122–5.
2. Hoffmann M, Pöhlmann S. Cell entry of influenza a viruses: sweet talk between HA and CaV1.2. Cell Host Microbe. 2018;23(6):697–9.
3. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature. 2009;459(7249):931–9.
4. Zhang R, Xu C, Duan Z. Novel antigenic shift in HA sequences of H1N1 viruses detected by big data analysis. Infect Genet Evol. 2017;51:138–42.
5. Mostafa A, Kanrai P, Ziebuhr J, Pleschka S. The PB1 segment of an influenza a virus H1N1 2009pdm isolate enhances the replication efficiency of specific influenza vaccine strains in cell culture and embryonated eggs. J Gen Virol. 2016;97(3):620–31.
6. Lee J-Y, Ouh I-O, Cho S-D, Cho I-S, Park CK, Song J-Y. Complete genome sequence of H1N1 swine influenza virus from pigs in the Republic of Korea in 2016. Baltrus DA, editor. Microbiol Resour Announc 2018;7(23):e01229–e01218, e01229-18.
7. Sullivan SJ, Jacobson RM, Dowdle WR, Poland GA. 2009 H1N1 Influenza. Mayo Clin Proc. 2010;85(1):64–76.
8. Zimmer SM, Burke DS. Historical perspective — emergence of influenza a (H1N1) viruses. N Engl J Med. 2009;361(3):279–85.
9. Shope RE. The incidence of neutralizing antibodies for swine influenza virus in the sera of human beings of different ages. J Exp Med. 1936;63(5):669–84.
10. Sun H, Xiao Y, Liu J, Wang D, Li F, Wang C, et al. Prevalent Eurasian avian-like H1N1 swine influenza virus with 2009 pandemic viral genes facilitating human infection. Proc Natl Acad Sci U S A. 2020;29:201921186.
11. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006;34(19):5402–15.
12. Malgowska M, Czajczynska K, Gudanis D, Tworak A, Gdaniec Z. Overview of the RNA G-quadruplex structures. Acta Biochim Pol. 2016;63(4):609–21.
13. Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. The regulation and functions of DNA and RNA G-quadruplexes. Nat Rev Mol Cell Biol. 2020 Aug;21(8):459–74.
14. Ruggiero E, Richter SN. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. In: Annual Reports in Medicinal Chemistry. Elsevier; 2020 [cited 2020 Jul 29]. p. S0065774320300142.
15. Lavezzo E, Berselli M, Frasson I, Perrone R, Palù G, Brazzale AR, et al. G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. PLoS Comput Biol. 2018 13; 14(12).

16. Lombardi EP, Londoño-Vallejo A. A guide to computational methods for G-quadruplex prediction. Nucleic Acids Res. 2020;48(3):1603.
17. Bedrat A, Lacroix L, Mergny J-L. Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. 2016;44(4):1746–59.
18. Gazanion E, Lacroix L, Alberti P, Gurung P, Wein S, Cheng M, et al. Genome wide distribution of G-quadruplexes and their impact on gene expression in malaria parasites. Di Antonio M, editor. PLoS Genet. 2020;16(7):e1008917.
19. Bartas M, Čutová M, Brázda V, Kaura P, Šťastný J, Kolomazník J, et al. The presence and localization of G-Quadruplex forming sequences in the domain of Bacteria. Molecules. 2019;24(9):1711.
20. Čutová M, Manta J, Porubiaková O, Kaura P, Šťastný J, Jagelská EB, et al. Divergent distributions of inverted repeats and G-quadruplex forming sequences in Saccharomyces cerevisiae. Genomics. 2019: S0888754319305269.
21. Brázda V, Kolomazník J, Lýsek J, Bartas M, Fojta M, Šťastný J, et al. G4Hunter web application: a web server for G-quadruplex prediction. Hancock J, editor. Bioinformatics. 2019;35(18):3493–5.
22. Bartas M, Brázda V, Bohálová N, Cantara A, Volná A, Stachurová T, 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.
23. Di Antonio M, Ponjavic A, Radzevičius A, Ranasinghe RT, Catalano M, Zhang X, et al. Single-molecule visualization of DNA G-quadruplex formation in live cells. Nat Chem [Internet]. 2020 20.
24. Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, et al. Involvement of G-quadruplex regions in mammalian replication origin activity. Nat Commun. 2019;10(1):3274.
25. Artusi S, Perrone R, Lago S, Raffa P, Di Iorio E, Palù G, et al. Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. Nucleic Acids Res. 2016;44(21):10343–53.
26. Callegaro S, Perrone R, Scalabrin M, Doria F, Palù G, Richter SN. A core extended naphtalene diimide G-quadruplex ligand potently inhibits herpes simplex virus 1 replication. Sci Rep. 2017;7(1):2341.
27. Métifiot M, Amrane S, Litvak S, Andreola M-L. G-quadruplexes in viruses: function and potential therapeutic applications. Nucleic Acids Res. 2014;42(20):12352–66.
28. González VM, Martín ME, Fernández G, García-Sacristán A. Use of Aptamers as Diagnostics Tools and Antiviral Agents for Human Viruses. Pharmaceuticals. 2016;9(4):78.
29. Majerciak V, Zheng Z-M. Detection of Viral RNA Splicing in Diagnostic Virology. In: Tang Y-W, Stratton CW, editors. Advanced Techniques in Diagnostic Microbiology. Cham: Springer International Publishing; 2018. p. 345–402.
30. Renaud de la Faverie A, Guédin A, Bedrat A, Yatsunyk LA, Mergny J-L. Thioflavin T as a fluorescence light-up probe for G4 formation. Nucleic Acids Res. 2014;42(8):e65.
31. Vondrušková J, Kypr J, Kejnovská I, Fialová M, Vorlíčková M. Guanine quadruplex formation by RNA/DNA hybrid analogs of Oxytricha telomere G4T4G4 fragment. Biopolymers. 2008;89(10):797–806.
32. Brázda V, Laister RC, Jagelská EB, Arrowsmith C. Cruciform structures are a common DNA feature important for regulating biological processes. BMC Mol Biol. 2011;12(1):33.
33. Ruggiero E, Richter SN. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. Nucleic Acids Res. 2018;46(7):3270–83.
34. Shen L-W, Qian M-Q, Yu K, Narva S, Yu F, Wu Y-L, et al. Inhibition of influenza a virus propagation by benzoselenoxanthenes stabilizing TMPRSS2 gene G-quadruplex and hence down-regulating TMPRSS2 expression. Sci Rep. 2020;10(1):7635.
35. Eisfeld AJ, Neumann G, Kawaoka Y. At the Centre: influenza a virus ribonucleoproteins. Nat Rev Microbiol. 2015;13(1):28–41.
36. Compans RW, Content J, Duesberg PH. Structure of the ribonucleoprotein of influenza virus. J Virol. 1972;10(4):795–800.
37. Hu Y, Sneyd H, Dekant R, Wang J. Influenza A Virus Nucleoprotein: A Highly Conserved Multi-Functional Viral Protein as a Hot Antiviral Drug Target. Curr Top Med Chem. 2017;17(20):2271–85.
38. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 2006; 34(Web Server):W676–W682.
39. Dolinnaya NG, Ogloblina AM, Yakubovskaya MG. Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: overview 50 years after their discovery. Biochem Moscow. 2016;81(13):1602–49.
40. Hognon C, Miclot T, García-Iriepa C, Francés-Monerris A, Grandemange S, Terenzi A, et al. Role of RNA guanine Quadruplexes in Favoring the dimerization of SARS unique domain in coronaviruses. J Phys Chem Lett 2020;11(14):5661–5667.
41. Speranskaia AS, Mel'nikova NV, Belenkin MS, Dmitriev AA, Oparina NI, Kudriavtseva AV. Genetic diversity and evolution of the influenza C virus. Genetika. 2012 Jul;48(7):797–805.
42. Xu R, Ekiert DC, Krause JC, Hai R, Crowe JE, Wilson IA. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. Science. 2010;328(5976):357–60.
43. Kuenstling TE, Sambol AR, Hinrichs SH, Larson MA. Oligomerization of bacterially expressed H1N1 recombinant hemagglutinin contributes to protection against viral challenge. Sci Rep. 2018;8(1):11856.
44. Gómez-Puertas P, Albo C, Pérez-Pastrana E, Vivo A, Portela A. Influenza virus matrix protein is the major driving force in virus budding. J Virol. 2000; 74(24):11538–47.
45. Campbell PJ, Danzy S, Kyriakis CS, Deymier MJ, Lowen AC, Steel J. The M segment of the 2009 pandemic influenza virus confers increased neuraminidase activity, filamentous morphology, and efficient contact transmissibility to a/Puerto Rico/8/1934-based Reassortant viruses. J Virol. 2014;88(7):3802–14.
46. Ji D, Juhas M, Tsang CM, Kwok CK, Li Y, Zhang Y. Discovery of G-quadruplex-forming sequences in SARS-CoV-2. Brief Bioinform. 2020: bbaa114.
47. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the National Center for biotechnology information. Nucleic Acids Res. 2019;47(D1):D23–8.
48. Stothard P. The sequence manipulation suite: JavaScript programs for Analyzing and formatting protein and DNA sequences. BioTechniques. 2000; 28(6):1102–4.
49. Sievers F, Higgins DG. Clustal omega. Curr Protoc Bioinformatics. 2014;48:3. 13.1–16.
50. Crooks GE. WebLogo: A Sequence Logo Generator. Genome Res. 2004;14(6): 1188–90.

Publisher's Note

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

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).

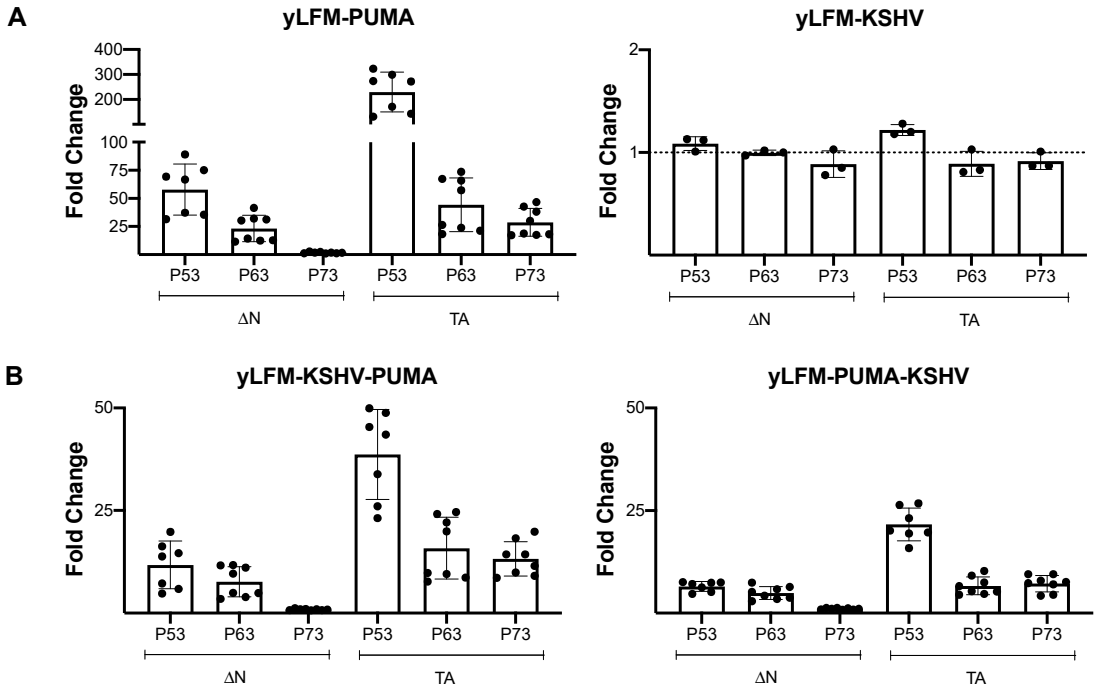


Figure 1. Fold change transactivation by wild-type P53 family proteins. Yeast cells expressing different P53 family isoforms by an inducible GAL1,10 promoter were grown for 8 h in Galactose 1% to evaluate their transactivation ability. (**A**) Evaluation of wild-type P53, P63, and P73 (ΔN variants: ΔN40P53α, 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).

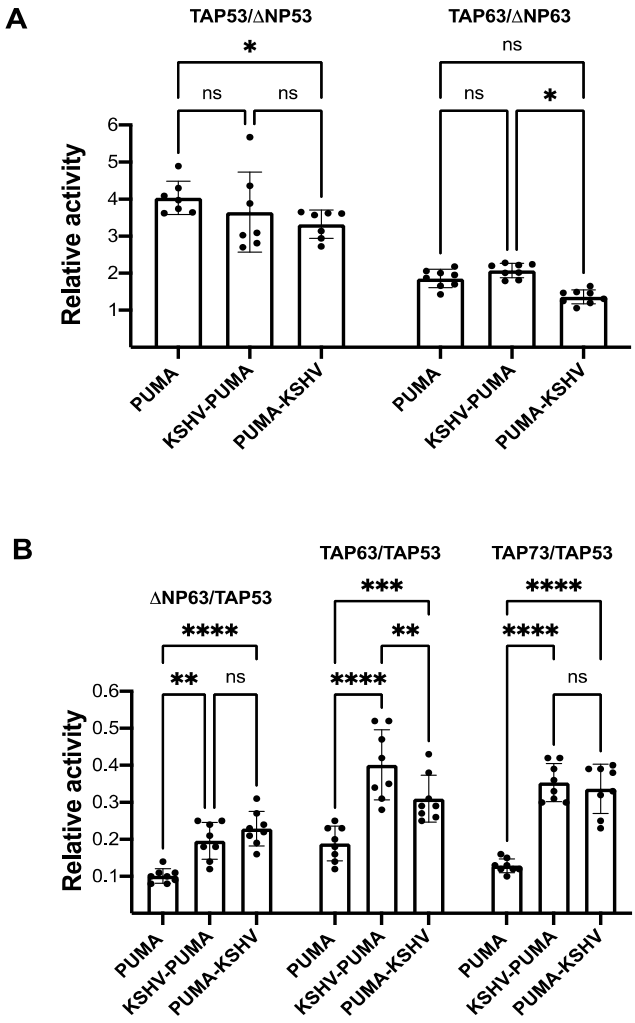


Figure 2. Relative activity of wild-type P53 family α isoforms. (**A**) Comparison of P53 (full-length P53) or P63 (TAP63α) TA variant activity with respect to the corresponding ΔN variant (ΔN40P53α and ΔNP63α, respectively) in yLFM-PUMA, yLFM-KSHV-PUMA, and yLFM-PUMA-KSHV strains. (**B**) Comparison 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).

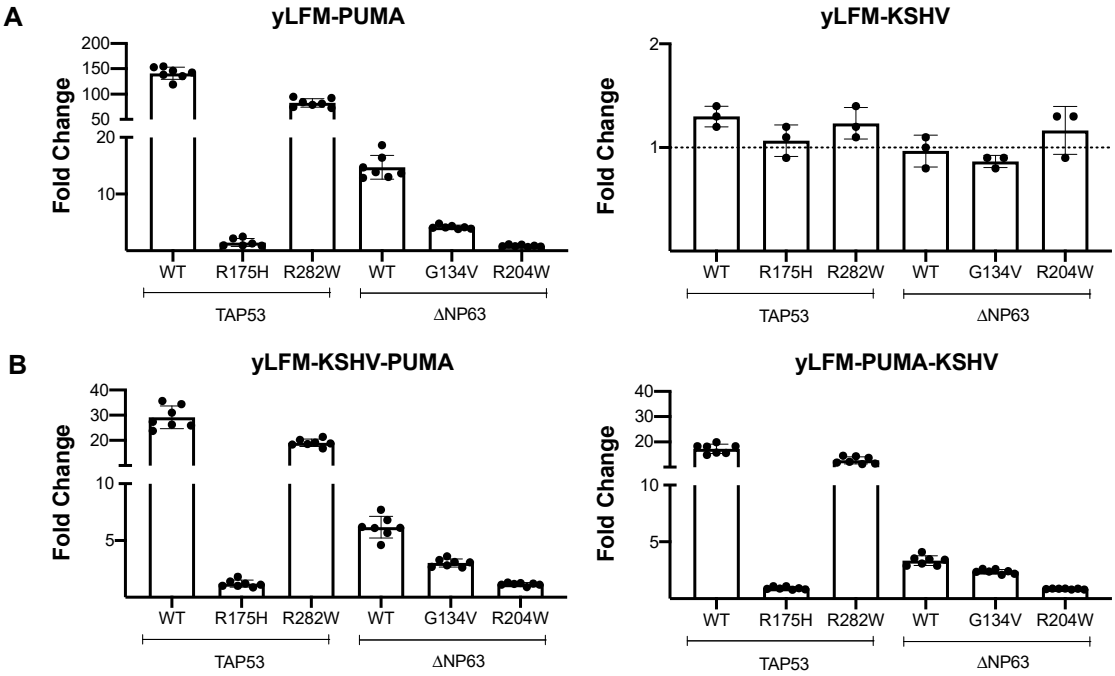


Figure 3. Fold change transactivation by mutant P53 family proteins. Yeast cells expressing wild-type and mutant TAP53 or ΔNP63 proteins by an inducible GAL1,10 promoter were grown for 8 h in Galactose 1% to evaluate the transactivation ability. (**A**) Evaluation of wild-type (TAP53: full-length P53; 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).

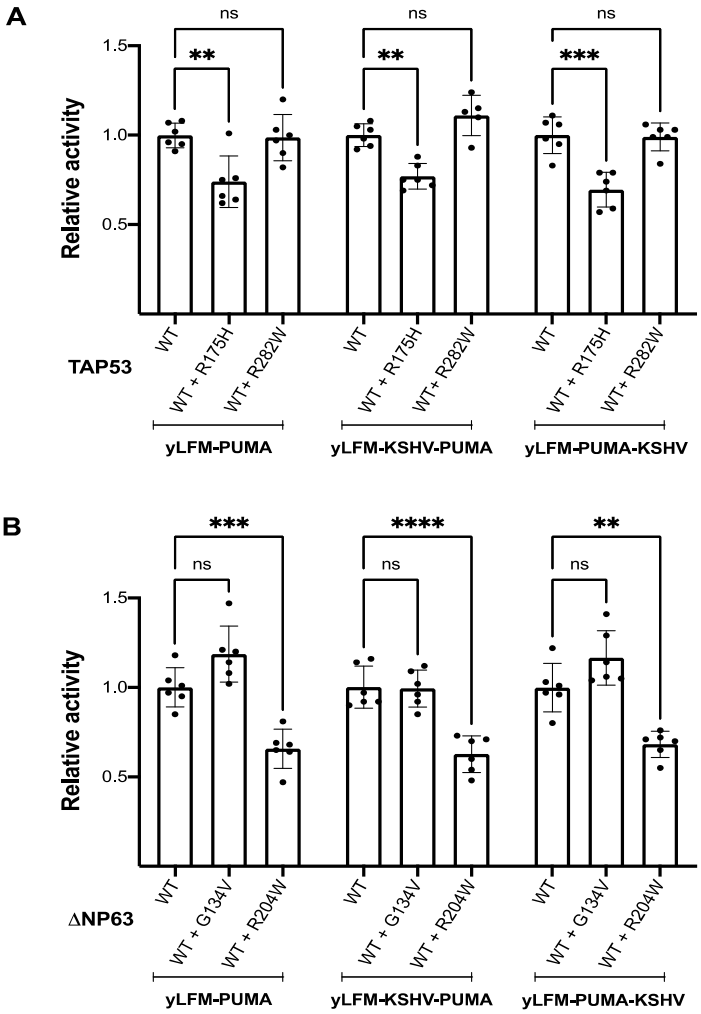


Figure 4. Relative activity of co-expressed wild-type and mutant TAP53 or ΔNP63 proteins compared to the corresponding single protein expression. (**A**) Comparison of wild-type and mutant TAP53 co-expression activity (i.e., constitutive pADH1-wild-type TAP53 + inducible pGAL1,10 TAP53 R175H or R282W) 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]