



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA CHEMICKÁ

FACULTY OF CHEMISTRY

ÚSTAV FYZIKÁLNÍ A SPOTŘEBNÍ CHEMIE

INSTITUTE OF PHYSICAL AND APPLIED CHEMISTRY

LOKÁLNÍ STRUKTURY V GENOMECH RETROVIRŮ

LOCAL STRUCTURES IN RETROVIRUS GENOMES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Adéla Sádlová

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Mgr. Václav Brázda, Ph.D.

BRNO 2024

Zadání bakalářské práce

Číslo práce: FCH-BAK1973/2023

Akademický rok: 2023/24

Ústav: Ústav fyzikální a spotřební chemie

Studentka: **Adéla Sádlová**

Studijní program: Chemie pro medicínské aplikace

Studijní obor: bez specializace

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

Název bakalářské práce:

Lokální struktury v genomech retrovirů

Zadání bakalářské práce:

Rešerše o retrovirech a DNA strukturách

Analýza pro přítomnost G–kvadruplexů v genomech retrovirů

Srovnání lokalizace lokálních struktur

Srovnání retrovirů mezi sebou a s jinými organismy

Termín odevzdání bakalářské práce: 20.5.2024:

Bakalářská práce se odevzdává v děkanem stanoveném počtu exemplářů na sekretariát ústavu.

Toto zadání je součástí bakalářské práce.

Adéla Sádlová
studentka

prof. Mgr. Václav Brázda, Ph.D.
vedoucí práce

prof. Ing. Miloslav Pekař, CSc.
vedoucí ústavu

V Brně dne 1.2.2024

prof. Ing. Michal Veselý, CSc.
děkan

ABSTRAKT

Tato bakalářská práce se zaměřuje na studium lokálních struktur v genomech retrovirů. Retroviry jsou skupinou virů s RNA genomem, které mají schopnost integrovat svůj genom do hostitelské DNA. Lokální struktury v genomu retrovirů hrají důležitou roli v procesech replikace, regulace genové exprese a interakcí s hostitelskými proteiny. Cílem této práce bylo analyzovat lokální struktury u vybraných retrovirů pomocí bioinformatických metod, přičemž klíčovým nástrojem byl G4 hunter, což je nástroj pro identifikaci G-kvadruplexů v genomu. G-kvadruplexy jsou struktury DNA nebo RNA, které se skládají z guaninových základů uspořádaných do čtveřic a jsou spojeny vrstvením guaninových tetrad. G-kvadruplexy jsou často lokalizovány v telomerách a promočních oblastech genů, a mohou hrát důležitou roli v regulaci genové exprese, replikaci DNA a dalších buněčných procesech. Analýza se zaměřila především na přítomnost potenciálních sekvencí tvořící G-kvadruplexy (PQS). Výsledky této práce poskytují důležité informace o distribuci a frekvenci PQS v genomech vybraných retrovirů a přispívají k lepšímu porozumění jejich role v biologii virů. Výsledky ukazují, že G-kvadruplexy jsou významné lokální struktury a často se nacházejí v regulačních oblastech genomu, což naznačuje jejich klíčovou roli v procesu virové infekce.

KLÍČOVÁ SLOVA

G-kvadruplex, retroviry, bioinformatická analýza, G4hunter

ABSTRACT

This bachelor thesis focuses on the study of local structures in the genomes of retroviruses. Retroviruses are a group of viruses with an RNA genome that have the ability to integrate their genome into the host DNA. Local structures in the retroviral genome play an important role in the processes of replication, gene expression regulation, and interactions with host proteins. The aim of this work was to analyze the local structures in selected retroviruses using bioinformatics methods, with G4 Hunter being the key tool for identifying G-quadruplexes in the genome. G-quadruplexes are DNA or RNA structures composed of guanine bases arranged into quartets and are connected by stacking guanine tetrads. G-quadruplexes are often located in telomeres, gene promoter regions, and can play an important role in regulating gene expression, DNA replication, and other cellular processes. The analysis focused primarily on the presence of potential G-quadruplex-forming sequences (PQS). The results of this work provide important information on the distribution and frequency of PQS in the genomes of selected retroviruses and contribute to a better understanding of their role in viral biology. The results show that G-quadruplexes are significant local structures and are often found in regulatory areas of the genome, indicating their key role in the viral infection process.

KEYWORDS

G-quadruplex, retroviruses, bioinformatics analysis, G4hunter

SÁDLOVÁ, Adéla. *Lokální struktury v genomech retrovirů* [online]. Brno, 2024 [cit. 2024-04-27]. Dostupné z: <https://www.vut.cz/studenti/zav-prace/detail/156573>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta chemická, Ústav fyzikální a spotřební chemie. Vedoucí práce Václav Brázda.

PROHLÁŠENÍ

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a že všechny použité literární zdroje jsem správně a úplně citovala. Bakalářská práce je z hlediska obsahu majetkem Fakulty chemické VUT v Brně a může být využita ke komerčním účelům jen se souhlasem vedoucího bakalářské práce a děkana FCH VUT.

.....
podpis studenta

PODĚKOVÁNÍ

Tímto bych chtěla moc poděkovat za vedení mé bakalářské práce prof. Mgr. Václavu Brázdovi, PhD. Moc děkuji Ing. Michaele Dobrovolné za její trpělivost při zpracování dat a pomoc při vypracovávání.

OBSAH

1.ÚVOD	7
1.TEORETICKÁ ČÁST.....	8
1.1. Charakteristika.....	8
1.1.1. Struktura	8
1.1.2. Obal retrovirů	8
1.1.3. Virový genom.....	9
1.1.4. Proteiny u retrovirů	9
1.1.4.1. Gag-protein.....	9
1.1.4.2. NC-protein.....	9
1.1.4.3. Pol-protein.....	9
1.1.4.4. Proteáza	10
1.1.4.5. Env-protein.....	10
1.1.4.6. Nef-protein	10
1.2. Životní cyklus retrovirů	10
1.2.1. Reverzní transkripce.....	11
1.2.2. Integrace	12
1.3. Rozdělení retrovirů.....	13
1.3.1. Orthoretrovirinae.....	13
1.3.1.1. Virus leukózy ptačího sarkomu.....	13
1.3.1.2. Virus lidské T-buněčné leukémie.....	13
1.3.1.3. Virus myši leukémie.....	13
1.3.1.4. HIV.....	14
1.3.1.5. FIV	14
1.3.2. Spumaretrovirinae	15
1.3.2.1. Feline foamy.....	15
1.4. Využití retrovirů	15
2.LOKÁLNÍ STRUKTURY DNA	15
2.1. Z-DNA	15
2.2. Triplex	16
2.3. G-kvadruplexy	17
2.4. Křížová struktura.....	19
2.5. I-motiv	19
3.EXPERIMENTÁLNÍ ČÁST	21
3.1. Analyzované organismy.....	21
3.2. Metody.....	21
3.2.1. Software G4hunter	21
3.2.1.1. G4hunter skóre	22
3.3. Analýza genomu HIV pomocí webové aplikace G4 hunter	23
3.4. Výpočty.....	24

4.VÝSLEDKY	25
4.1. <i>Analýza PQS</i>	25
4.1.1. Analýza PQS v genomu Human immunodeficiency virus 1.....	26
4.1.2. Srovnání výskytu PQS v genomech retrovirů s genomem Homo sapiens	30
5.ZÁVĚR.....	32
6.ZDROJE	34
7.POUŽITÉ ZTKRATKY	43

1. ÚVOD

Retroviry představují fascinující skupinu virů s RNA genomem, které mají schopnost integrovat svůj genetický materiál do hostitelské DNA. Tyto viry jsou známé pro svou schopnost způsobovat vážná onemocnění u lidí a zvířat, včetně AIDS a několika typů nádorových onemocnění [1].

První část této bakalářské práce obsahuje literární rešerši o retrovirech. Zahrnuje analýzu jejich struktury, důležitých proteinů této skupiny, životního cyklu a různých typů onemocnění, které způsobují. Práce je zaměřena na lokální struktury v genomech retrovirů. Lokální struktury jsou specifické strukturální prvky v sekvencích DNA nebo RNA, které mohou mít významný vliv na biologické funkce těchto molekul. G-kvadruplexy jsou lokální struktury, na které se tato práce soustředí. Tyto struktury se skládají z guaninů uspořádaných do čtveřic, vrstvením 4 guaninů vznikají guaninové tetrády. G-kvadruplexy mohou vznikat v jednořetězcové DNA nebo RNA sekvenci, kde jsou guaniny uspořádány do specifických motivů. Tyto struktury mohou vznikat za určitých podmínek, jako je například vyšší koncentrace výskytu guaninů a přítomnost kationtů, jako je draslík nebo sodík [2, 3, 4].

V experimentální části práce byly identifikovány a analyzovány G-kvadruplexy v genomech vybraných retrovirů pomocí bioinformatických metod. Využit byl nástroj G4hunter, který umožňuje identifikaci potenciálních sekvencí tvořící G-kvadruplex (PQS) v genomech. Analýza těchto struktur poskytla informace o jejich šíření, frekvenci a možných funkcích v retrovirových genomech. Studium G-kvadruplexů v genomech retrovirů má potenciál poskytnout nové poznatky o biologii těchto virů a může vést k objevům nových terapeutických cílů pro léčbu retrovirových infekcí [5].

Cílem práce bylo analyzovat skupinu retrovirů pomocí bioinformatické analýzy a zjistit výskyt PQS v jejich genomech. Dále bylo provedeno srovnání těchto výsledků s genomem člověka, aby byly zjištěny případné odlišnosti v rozložení G4 struktur. Zvláštní pozornost byla věnována také analýze jednoho konkrétního retrovirového genomu – HIV-1, který je zodpovědný za vznik syndromu AIDS [6]. Tento výzkum zahrnoval anotaci genomu pomocí bioinformatických analýz, které poskytly informace o potenciálních funkčních oblastech virového genomu.

Tato bakalářská práce se snaží přispět k lepšímu porozumění funkcí retrovirů pomocí kombinace literární rešerše a experimentální bioinformatické analýzy.

1. TEORETICKÁ ČÁST

1.1. Charakteristika

Skupina retrovirů byla poprvé objevena okolo roku 1900. Původní snaha studie této skupiny virů byla založena na využití zvířecích modelů k pochopení lidských nádorových onemocnění. Dnes jsou hlavními cíli retrovirologie léčba a prevence AIDS a využití virů jako nosičů genů. Zkoumání retrovirů také zahrnuje zaměření na vir HIV nebo k vytváření virových vektorů [7, 8]

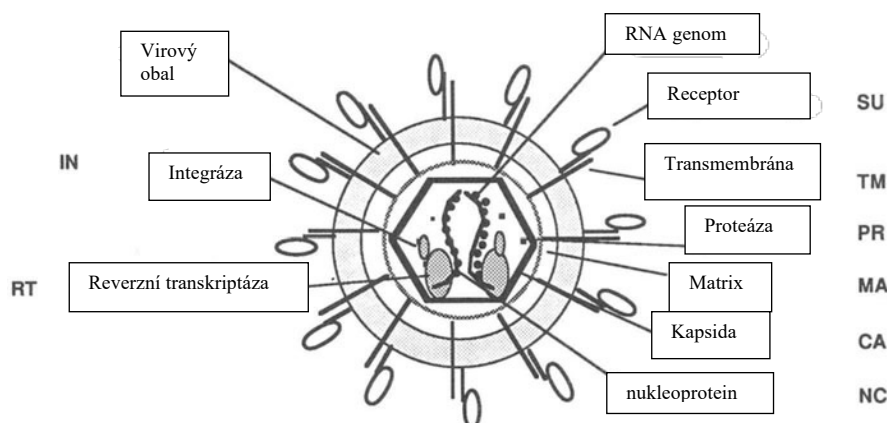
Tyto viry zahrnují velkou skupinu původců (*retroviridae*), kterou spojuje společná struktura virionu a způsob replikace. Retroviry byly izolovány z většiny druhů obratlovců [1]. Je to skupina virů, která je schopna měnit genetický materiál hostitelské buňky. Radíme je mezi RNA-viry, mají tedy namísto DNA genetickou informaci v podobě RNA. Charakteristickým rysem těchto virů je přepis své RNA na DNA pomocí reverzní transkriptázy procesem zvaným reverzní transkripce a následnou integrací virové DNA do genomové RNA. Tímto je jim umožněno uložit svůj genetický materiál do hostitelského genomu [10, 8].

1.1.1. Struktura

Viriony retrovirů jsou obalené částice o průměru asi 100–200 nm s vnitřním kulovitým nebo kuželovitým jádrem a dimerním genomem polyadenylované RNA o délce 7–0 kilobází (kb) [1, 9]. Nezralé částice vykazují zřetelnou morfologii ve tvaru koule. U nezralých částic retroviru je hustá vrstva pod membránou nezralého retroviru tvořena polyproteiny Gag, které vytváří parakrystalické uspořádání [11].

Jádro viru je tvořeno nukleovou kyselinou (2 molekuly jednovláknové RNA), která je chráněná kapsidou, tvořenou virovými proteiny a dále je obalena fosfolipidovou dvovrstvou a glykoproteinovým virovým obalem, který je také dále využit při vstupu virionu do buňky.

Na *Obrázku 1* vidíme popsany virion retroviru. Obal virionu chrání matrix a pod ním kapsidu, která obsahuje RNA genom, reverzní transkriptázu, která umožňuje přepis RNA do DNA a je zároveň klíčovým enzymem pro replikaci retrovirů [10]. Na obrázku je také zaznačen enzym proteáza a integráza. Integráza umožňuje integraci virové DNA do genomu hostitelské buňky, což vede k replikaci viru [8].



Obrázek 1: Virion retroviru. Obrázek je schematický [8]

1.1.2. Obal retrovirů

Obal retrovirů je tvořen lipidovou dvovrstvou, která vzniká pučením plazmatické membrány. Proteiny obalu specificky rozpoznávají proteiny buněčného povrchu a pomocí endocytózy se

dostávají do buněk. Vstup iniciuje přeměna kladně obalené částice na nukleoproteinový komplex, který vykonává své funkce, jako je reverzní transkripce a integrace. Tyto úkony vykonává s minimální pomocí od hostitelské buňky [10, 15, 12, 16].

1.1.3. Virový genom

Genom retroviru se skládá z dimeru identických jednovláknových molekul RNA o délce 7-10 kb. Genomová RNA je modifikována způsobem, který odráží její syntézu a zpracování buněčnými mechanismy: na 5' konci je uzavřena, na 3' konci je polyadenylována. Tento dimerní genom retrovirů je schopen sloužit jako posel pro kapsidové proteiny a enzymy. Sekvence, které regulují strukturální transformace virového genomu během životního cyklu, jsou shlukovány na konci RNA (signály pro zahájení a průběh syntézy DNA, pro integraci, pro přepis viru do RNA, pro zpracování RNA a pro balení RNA do částic). Během syntézy virové DNA jsou sekvence, které se nacházejí na koncích virové RNA, duplikovány, což vede k vytvoření dlouhých terminálních repetitiv. Tyto repetice mají délku několika set bazových párů a jsou umístěny na obou koncích virové DNA [8].

Kromě genomu obsahují viriony retrovirů řadu malých molekul RNA (a dokonce i některé molekuly DNA), z nichž většina nemá zjevnou úlohu při replikaci a předpokládá se, že jde o náhodné buněčné složky obsažené ve virionu. Jedinou malou molekulou je molekula tRNA, která je specificky spojena s genomem v rámci párování bází, konkrétně zajišťuje přenos aminokyselin do ribozomů a hraje velmi důležitou roli v procesu translace [16, 17].

1.1.4. Proteiny u retrovirů

1.1.4.1. *Gag-protein*

Tyto polyproteinové prekurzory jsou syntetizovány z nesestříhané genomové RNA na ribozomech bez membrány v cytoplazmě. Poté jsou transportovány na plazmatickou membránu a indukují uvolnění vznikajících částic v hostitelské buňce, kde jsou nastříhány pomocí proteázy [18]. Gag protein zastává ve virionu retrovirů dvě velmi důležité funkce. Je součástí hlavního prekurzoru, který se podílí na sestavování, pučení a řídí tyto kroky. Dále jako jednotlivý protein, který podporuje strukturu virionu a hraje důležitou roli (v syntéze DNA, integraci) na počátku infekce. Tyto dvě funkce dávají za výsledek to, že lze dosáhnout spojení protein-membrána, které způsobí pučení. Pučení napomáhá proteinu matrixu, který lemuje vnitřní stranu obalu virionu [7, 8].

Všechny retroviry mají gen *Gag* na 5' konci svého genomu. Protein Gag je primární translační produkt tohoto genu. V okamžiku, kdy je protein Gag zmutován, ztrácejí retroviry svou schopnost se přemístit z plazmatické membrány hostitelské buňky [19, 18].

1.1.4.2. *NC-protein*

NC-protein neboli HIV-1 je malý základní protein, který vzniká štěpením polyproteinu Gag virovou proteázou během sestavování viru v infikované buňce. Retrovirální NC proteiny jsou proteiny vázající nukleové kyseliny, které mají silnou kondenzační a chaperonovou aktivitu. NC protein řídí kritické strukturální změny genomové RNA, zejména dimerizaci RNA v průběhu sestavování viru a annealing virové nukleové kyseliny, který je nutný pro replikaci genomové RNA virovou reverzní transkriptázou [20, 21]. NC podporuje tvorbu duplexů RNA-RNA, což je pravděpodobně nezbytné pro uskutečnění asociace genom-primer a dimerizaci genomu. Je také zodpovědný za rozlišování mezi sestříženými a nesestříženými virovými RNA [8, 9, 20, 21].

1.1.4.3. *Pol-protein*

Úlohou proteinů Pol je syntéza virové DNA a její integrace do hostitelské DNA brzy po infekci. Ačkoli jsou tyto proteiny schopny vykonávat potřebné enzymatické funkce *in vitro*,

skutečné reakce *in vivo* probíhají v kontextu složitější struktury odvozené od virové kapsidy, která může být nezbytná pro zajištění specifičnosti [1, 8]

1.1.4.4. Proteáza

Retrovirové proteázy jsou klíčovými enzymy při virovém šíření a jsou zpočátku syntetizovány s jinými virovými proteiny jako polyproteinové prekurzory, které jsou následně štěpeny virovou proteázovou aktivitou na specifických místech za vzniku zralých, funkčních jednotek [25, 26]. Oblast genomu kódující proteázu leží vždy mezi proteiny Gag a Pol. Je to asparagová proteáza, jejíž funkční formou je dimer. Je zodpovědná za všechna proteolytická štěpení vytvářející zralé Gag a Pol proteiny během zrání virionu [8, 27].

1.1.4.5. Env-protein

Gen *Env* je překládán na polysomech vázaných na membránu, čímž vzniká prekurzor, který je následně modifikován glykosylací a štěpením na dva proteiny. Na rozdíl od proteinů Gag a Pol jsou proteiny *Env* zpracovávány buněčnými systémy. Jedinou úlohou *Env* proteinů je zprostředkování asociace virionu s hostitelskou buňkou a vstup do ní [7, 8].

Gen *Env* obvykle překrývá 3' konec čtecího rámce *Pol* (s výjimkou lentivirů a spumavirů) a akceptor sestřihu pro *Env* obvykle leží v kódující oblasti *Pol*. Míra využití tohoto akceptoru sestřihu určuje poměr genomu a *Env* mRNA. Využití tohoto proteinu je určeno sekvencemi v blízkosti míst sestřihu a uvnitř genů *Gag* a *Pol* [30].

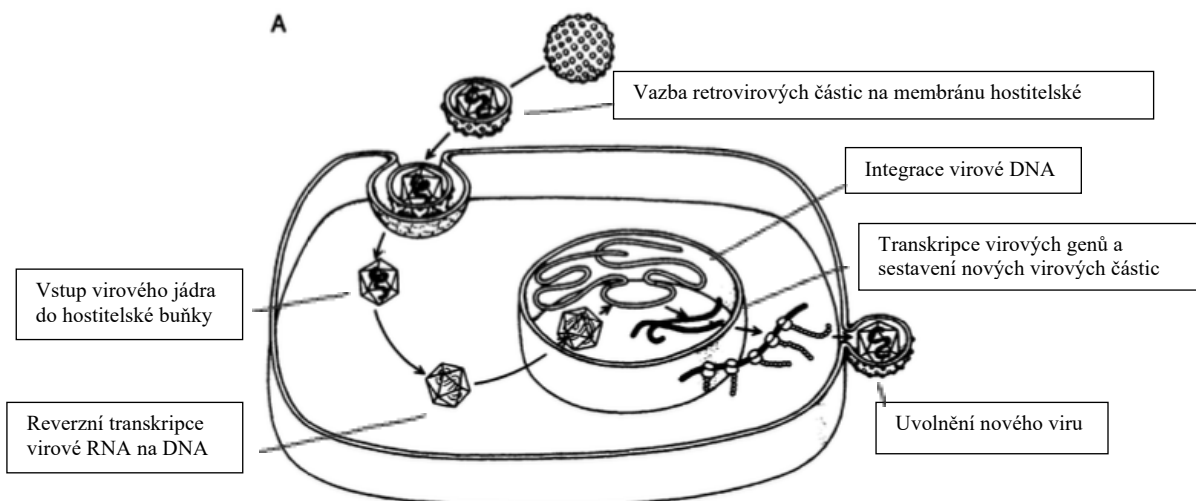
1.1.4.6. Nef-protein

Protein *Nef* je vylučován všemi lentivirovými viry primátů – HIV-1, HIV-2 a SIV. Jeho exprese v raných fázích životního cyklu viru zajišťuje dva základní atributy infekce HIV – je to aktivace T-buněk a vytvoření trvalého stavu infekce. *Nef* má pozitivní vliv na virovou infekci a replikaci tím, že podporuje přežívání infikovaných buněk [31]. *Nef* je výrazně exprimován na počátku infekce a předpokládá se, že optimalizuje buněčné prostředí pro virovou replikaci. Kontroluje zároveň úroveň exprese různých povrchových buněčných molekul, které hrají důležitou roli v imunitě hostitele a životním cyklu viru [32].

1.2. Životní cyklus retrovirů

Životní cyklus retrovirů začíná reverzní transkripcí a integrací do genomu hostitele, následuje vznik mRNA, která vzniká v procesu transkripce, během kterého RNA polymeráza přepisuje genetickou informaci z DNA do RNA, mRNA následně putuje do cytoplazmy, kde probíhá druhý krok, který se nazývá translace. Reverzní transkripce probíhá ve virové partikuli v cytoplazmě. Vznikne výsledná cDNA, která je poté pomocí virové integrázy začleněna do genomu hostitele [8, 33, 34]. V této fázi viry využijí hostitelské mechanismy pro expresi genů.

Na *Obrázku 2.* můžeme vidět životní cyklus těchto virů. Část viru vstupující do oblasti v levé horní části buňky se rozbaluje a vytváří proteinový komplex, ve kterém se nachází virová RNA, následuje proces reverzní transkripce, kdy je přepisována virová RNA do DNA pomocí reverzní transkriptázy. Po migraci do buňky se komplex integruje do chromozomu hostitelské buňky. Následně je zde zobrazena syntéza virové RNA a proteinů, což vede k seskupení částic, které opouštějí buňku přes plazmatickou membránu. Cyklus končí uvolněním nových virů z hostitelské buňky, což umožňuje infikovat další organismy [8,1].

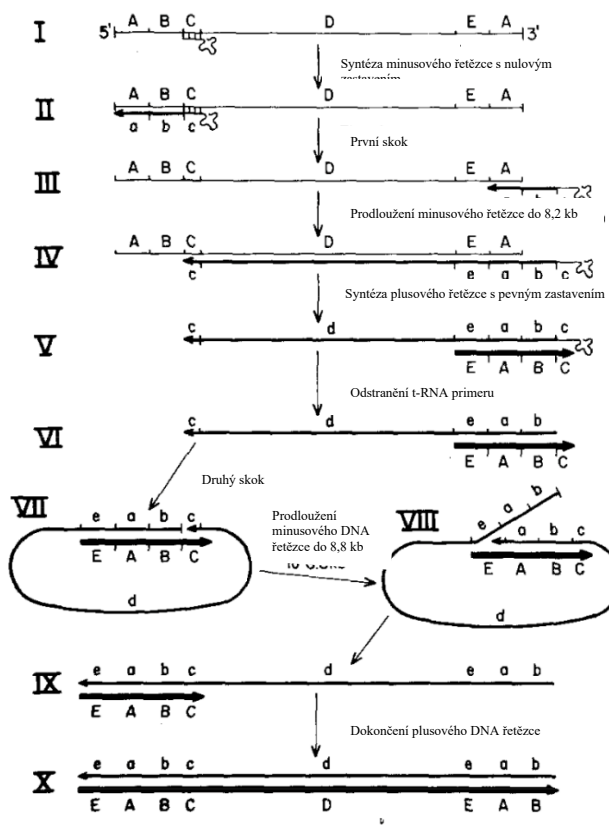


Obrázek 2: Životní cyklus retrovirů [8]

1.2.1. Reverzní transkripce

Proces transkripce začíná syntézou jednořetězcové DNA (proto také reverzní transkripce, protože přepis probíhá z RNA do DNA). Molekula tRNA zahájí reverzní transkripci. Je komplementární k místu vazby primeru (PBS). Poté přichází reverzní transkriptáza, která syntetizuje 3' konec virové DNA, RN-áza rozloží virovou RNA, to umožní párování reverzních sekvencí nově vzniklé DNA s 3' koncem RNA viru. Nakonec se reverzní oblast DNA řetězce přesune z 5' konce na 3' konec RNA viru. Poté reverzní transkriptáza dokončuje tvorbu kopie s využitím 5' konce. V dalším kroku RN-áza rozloží veškerou RNA v duplexu RNA-DNA s výjimkou PPT (Primer-Template-Pairing), což v reverzní transkripci označuje párování předlohy (template) s počátečním primerem během procesu syntézy komplementárního DNA (cDNA) z RNA pomocí reverzní transkriptázy [8, 11, 28].

Tento úsek je využíván k zahájení syntézy dalšího řetězce DNA, který odpovídá 3' polovině retrovirového genomu. Po odstranění tRNA a genomové RNA z RNA-DNA dochází ke druhému „skoku“, během kterého se PBS na 5' konci druhého řetězce DNA hybridizuje s komplementárním PBS na 5' konci prvního řetězce DNA. 3' hydroxylové konce obou řetězců pak slouží k zahájení syntézy DNA, aby mohla být dokončena dvouřetězcová molekula. Konverze retrovirové RNA na DNA vede k tvorbě dlouhých koncových repetitivních sekvencí na obou koncích molekuly [8, 35, 37, 67]. Mechanismus zobrazuje *Obrázek 3*.



Obrázek 3: Mechanismus reverzní transkripce [67]

1.2.2. Integrace

Integrace kopie DNA virového genomu do DNA hostitelské buňky je základním krokem v životním cyklu většiny, ne-li všech retrovirů. Retroviróvé molekuly DNA nejsou obvykle schopny se samostatně replikovat jako epizomy, jejich stabilní udržení v dělicích se buňkách závisí na integraci. Po integraci se genom viru (RNA) neboli provirus přenáší jako integrální prvek hostitelského genomu. Integrace slouží k přepisu virové DNA do nových kopií virového genomu a mRNA, které kódují virové proteiny. Integrace proviru tedy definuje kritický přechod v životním cyklu viru od pouhého přežívání k množení [42, 43, 24].

Začlenění retroviróvé DNA do hostitelského genomu je katalyzováno enzymem integrázou, který má endonukleázovou aktivitu. Štěpí se konce dlouhých koncových repetit a integráza vytvoří 5' konce. Ty jsou pak použity k ataku fosfodiesterových vazeb v cílové DNA sekvenci hostitele, katalyzovaných pomocí integrázy. Toto vede k vytvoření vazeb mezi retroviróvým řetězcem a řetězcem hostitele. Provirus je v místě začlenění duplikován a stává se tak úplnou součástí hostitelské buňky, ve které se replikuje jako každý jiný segment její vlastní DNA [24, 44, 37, 35, 8].

1.3. Rozdělení retrovirů

Podle aktuálního rozdělení z roku 2022 podle ICTV (International Committee on Taxonomy of Viruses) dělíme retroviry na 2 podčeledi. *Orthoretrovirinae* je podčeď, která má ve své skupině 6 druhů retrovirů. Druhá podčeď je *Spumaretrovirinae*, která je rozdělena do 5 druhů retrovirů.

1.3.1. Orthoretrovirinae

Do této podčeledi jsou zařazeny viry, které způsobují velmi závažná onemocnění jak u člověka, tak u zvířete. Mezi nejznámější *Orthoretrovirinae* patří virus způsobující onemocnění AIDS – virus HIV. V *Tabulce 1* vidíme rozdělení podčeledi *Orthoretrovirinae* a druhy této skupiny. U každého druhu jsou vypsány nejznámější zástupci retrovirů.

Tabulka 1: Rozdělení *Orthoretrovirinae* podle ICTV

ORTHORETROVIRINAE					
ALPHARETRIVIRUS	BETARETROVIRUS	DELTAETROVIRUS	EPSILONRETROVIRUS	GAMMARETROVIRUS	LENTIVIRUS
ASLV (Virus leukózy ptačího sarkomu) RSV	MMTV	Viry typu HTLV (Virus lidské T-buněčné leukémie)	WEHV	MLV (Virus myší leukémie)	HIV SIV FIV

1.3.1.1. Virus leukózy ptačího sarkomu

Virus ptačí leukózy (ALV) je nejběžnějším přirozeně se vyskytujícím ptačím retrovirem, který může u kuřat způsobit řadu neoplastických onemocnění. Kromě toho, že způsobuje neoplazii (tvoření nové tkáně ve spojení s bujením nádorů), je známo, že ALV je spojen se sníženou produktivitou a produkčními problémy celkově v postižených hejnech. Typ nádoru vyvolaného ALV je ovlivněn kmenem viru, dávkou expozice, genotypem, pohlavím hostitele a věkem při expozici [46, 47].

Viriony virů ptačí leukózy a sarkomu se od virionů jiných obalených živočišných RNA virů liší složitější strukturou bez jasně pozorovatelné vnitřní symetrie a tím, že mají více proteinů [24].

1.3.1.2. Virus lidské T-buněčné leukémie

Kolem roku 1980 byla v Japonsku popsána „nová“ leukémie označovaná jako T-buněčná leukémie, jinak nazýván lymfom dospělých T-buněk (ATL). 10 let poté byl izolován retrovirus, virus lidské T-buněčné leukémie typu 1 (HTLV-1), který toto onemocnění způsobuje. Jedná se o infekčnost HTLV, kde onkoprotein HTLV-1 Tax iniciuje transformaci vytvořením buněčného prostředí podporujícího aneuploidii a poškození DNA [47]. Je to retrovirus, kterým je infikováno 20 milionů lidí na celém světě a zároveň byl prvním retrovirem, u něhož byla prokázána příčinná souvislost s lidským nádorovým onemocněním, lymfomu dospělých T-buněk [46]

Jeho virový onkoprotein Tax je potřebný k zahájení, ale nikoli k udržení buněčné transformace. Tax transformuje buňky různými mechanismy, včetně vytváření chromozomální nestability, amplifikace centrozomů, poškození DNA a aktivace cyklin-dependentních kináz [52].

1.3.1.3. Virus myší leukémie

Viry myší leukémie (MLV) patří mezi nejjednodušší retroviry. Jsou to viry způsobující leukémii a nádorová onemocnění u infikovaných myší. MLV jsou studovány již mnoho let,

počínaje 50. lety 20. století, kdy se zjistilo, že leukemie může být přenášena na novorozené myši. Genom MLV byl také použit jako výchozí materiál při vývoji vektorů pro genovou terapii. V neposlední řadě byly MLV často považovány za „modelové“ retroviry [53, 54]. Na buněčné úrovni se zdá být infekce MLV téměř zcela neškodná, obecně nejsou patrné žádné účinky produktivní infekce MLV na růst, fyziologii nebo morfologii buněk [53, 54, 55].

1.3.1.4. HIV

V této době žije přibližně 37 milionů osob s touto infekcí. Neléčená replikace HIV způsobuje postupný úbytek CD4⁺T buněk, což vede ke zvýšenému riziku infekčních, onkologických a imunitních komplikací. Infekce HIV také přispívá ke vzniku kardiovaskulárních onemocnění, onemocnění kostí, poruchám funkce ledvin nebo jater [6].

Onemocnění AIDS je způsobeno virem lidské imunodeficiencie (HIV). U léku Retrovir (AZT) bylo prokázáno, že se díky němu prodlouží život pacientů s nákazou HIV [58, 57, 59, 60].

1.3.1.5. FIV

Virus kočičí imunodeficiencie (FIV), dříve kočičí T-lymfotropní lentivirus, byl poprvé izolován ze skupiny koček v Kalifornii v roce 1986. Replikuje se přednostně v kočičích T-lymfoblastických buňkách, kde způsobuje charakteristický cytopatický účinek. U koťat infikovaných FIV se infekce projevuje horečkou. To je spojeno s generalizovanou lymfadenopatií. Většina koček se z této počáteční fáze onemocnění zotaví a stane se celoživotními nositeli viru. K úplnému uzdravení nedochází v žádném rozsahu v přírodě ani v laboratorních podmínkách [58, 62, 61].

1.3.2. Spumaretrovirinae

Do této počeledi řadíme viry, které způsobují závažná onemocnění hlavně u zvířat, jako je Feline foamy virus napadající kočky nebo koně. Řadíme sem i viry, které způsobují onemocnění u lidí – HFV. V *Tabulce 2* je sepsáno rozdělení čeledi *Spumaretrovirinae* a skupiny tohoto druhu. V každé skupině je zmíněn jeden hlavní zástupce.

Tabulka 2.: Rozdělení spumaretrovirinae podle ICTV

SPUMARETROVIRINAE				
BOVISPUMAVIRUS	EQUISPUMAVIRUS	FELISPUMAVIRUS	PROSIMIISPUMAVIRUS	SIMIISPUMAVIRUS
Bovine foamy (virus napadající skot)	Equine foamy (virus napadající koně)	Feline foamy (Virus napadající kočky)	Brown greater galago prosimian foamy (virus napadající komby)	HFV (foamy virus napadající lidi)

1.3.2.1. Feline foamy

Tyto viry jsou možné izolovat z většiny tkání infikovaného zvířete. Jsou to viry napadající hlavně kočky. DNA těchto virů jsou přítomny v lymfocytech zvířete. Přenos viru probíhá nejspíše kousnutím nebo vzduchem. Tento virus negativně ovlivňuje imunitní systém kočky, napadá určitý orgán nebo část těla [64].

1.4. Využití retrovirů

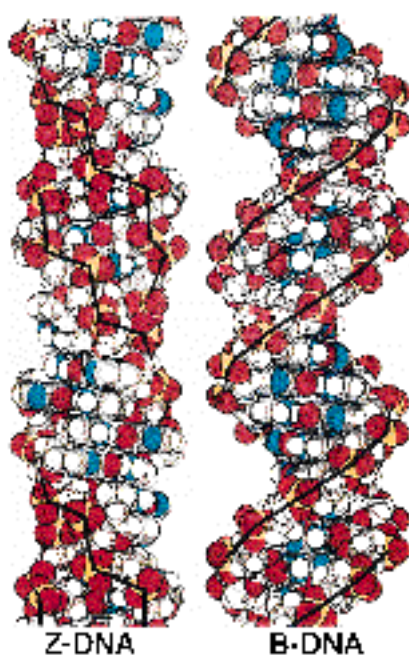
Použití retrovirů jako vektorů se osvědčilo v mnoha genových terapiích. Použití retrovirových vektorů ke změně genové exprese se ukázalo být cennou alternativou k transgenním přístupům. Mnohé z genetických manipulací, které se tak osvědčily u myši, lze napodobit pomocí retrovirových vektorů. Použití retrovirů ke změně genové exprese má několik výhod vyplývajících ze snadného vytváření různých expresních vzorců pomocí stejného viru. Retroviry mohou být použity k přenesení specificky zaměřeného genetického materiálu do buněk pro studium funkce genů nebo regulaci genové exprese [65, 66].

V genové terapii najdou tedy retroviry také své místo. Genová terapie je léčebná metoda s potenciálem léčit dědičné a získané choroby. K přenosu genů do savčích buněk bylo použito několik virových a fyzikálně-chemických nosičů, ale ve většině klinických studií genové terapie se používají rekombinantní retroviry [67, 68].

2. LOKÁLNÍ STRUKTURY DNA

2.1. Z-DNA

DNA kóduje biologickou informaci nejméně dvěma různými způsoby. Zaprvé, prostřednictvím lineárního pořadí nukleotidů určuje složení proteinů. Zadruhé, prostřednictvím svého tvaru může DNA poskytovat informace k regulaci sestavení buněčných struktur DNA. Z-DNA je schopna nabývat mnoha tvarů. Jedna ze změn tvaru je ta, ke které dochází při přechodu od pravotočivé šroubovice B-DNA k protáhlejší levotočivé konformaci Z-DNA [69, 71, 72]. Na *Obrázku 4* můžeme vidět rozdíl mezi Z-DNA a B-DNA.



Obrázek 4:: Srovnání struktury B-DNA a Z-DNA [71]

Z-DNA je konformace s vyšší energií než B-DNA. Z-DNA mohou tvořit sekvence jak GC, tak GT. Přeměna na Z-DNA závisí na sekvenci, nejlepší sekvence pro tvorbu tohoto typu DNA je sekvence d(GC)_n, tato zkratka znamená, že sekvence obsahuje dvojici bází G a C, „n“ udává počet opakování této dvojice. Další sekvence, která tvoří Z-DNA je d(TG)_n, opakování d(GGGC)_n nebo d(TA)_n [71, 69, 73, 74, 75, 76, 72].

Energie potřebná k tvorbě Z-DNA *in vivo* vzniká transkripcí. Torzní napětí generované průchodem RNA polymerázy se pak stává silným zdrojem energie pro stabilizaci Z-DNA. K překlacení z formy B do formy Z dochází, když procesní enzymy, jako jsou polymerázy a helikázy, vytvářejí nedostatečně navinutou DNA [71, 72].

Důležitým inflexním bodem byla identifikace vazebné domény Z-DNA pojmenované Z α 1,2 DNA z dsRNA editačního proteinu ADAR. Interakce mezi Z α a Z-DNA jsou konformačně specifické. Bylo zjištěno, že doména Z α váže levotočivou konformaci Z-RNA dsRNA (dvojvláknová RNA) bez sekvenci specifity, což poskytuje mechanismus pro zacílení regulačního proteinu ADAR na substráty upravující dsRNA. Uvádí se tedy, že Z-DNA podporuje transkripci DNA [69, 70, 73, 74, 75, 76, 77].

2.2. Triplex

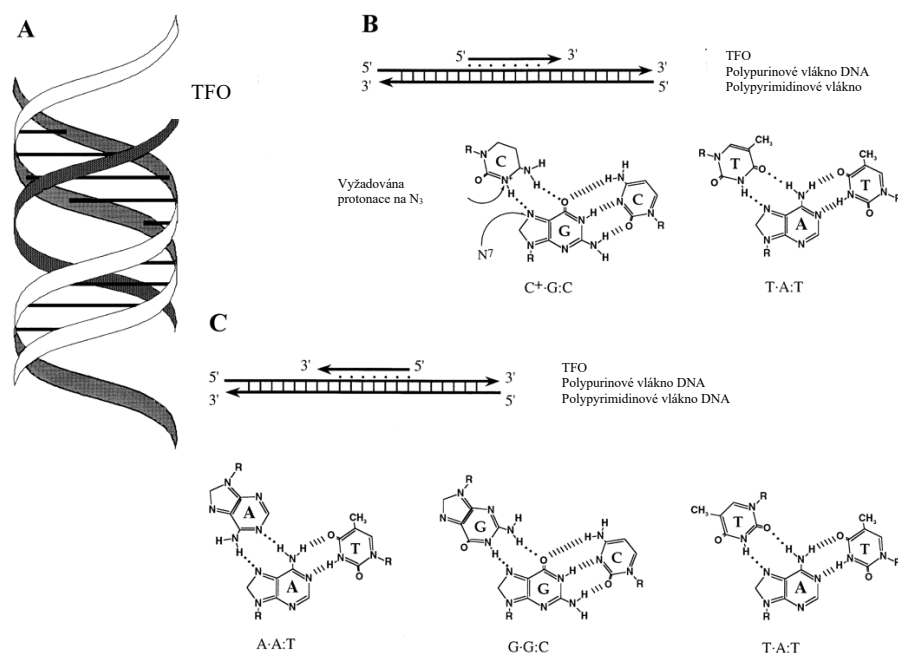
Tvorba třířetězcové, resp. trojitě helikální struktury nukleových kyselin byla poprvé pozorována v roce 1957, kdy prokázali stabilní a specifickou vazbu jednořetězcového polyuridinového oligonukleotidu na polyuridin/polyadenosinový duplex [78].

Triplex DNA vzniká, když pyrimidinové nebo purinové báze obsadí hlavní drážku dvojité šroubovice DNA a tvoří Hoogsteenovy páry s puriny Watson-Crickových bází. Mezimolekulární triplexy vznikají mezi triplexotvornými oligonukleotidy (TFO) a cílovými sekvencemi na duplexní DNA. Intramolekulární triplexy jsou hlavními prvky H-DNA (H-DNA zahrnuje triplex jako hlavní strukturální prvek), které se tvoří v homopurin-homopyrimidinových oblastech superzávitové DNA [77].

Během studia triplexů se ukázalo, že TFO by mohly být univerzálními léčivy (regulace genové exprese, léčba genetických chorob nebo antivirová léčba), které vykazují sekvenci specifické rozpoznávání duplexní DNA [72].

Na *Obrázku 5* vidíme 3 části. V sekci A model DNA, tmavá páska znázorňující relativní

polohu TFO (černé vlákno) v hlavní drážce DNA (šedé a bílé vlákno); plně černé čáry jsou Watsonovy-Crickovy páry bází. V části B je pyrimidinový vazebný motiv. Nahoře vazba TFO v paralelní orientaci na polypurinové vlákno DNA; šipky označují jednotlivá vlákna trojitě šroubovice a směřují ve směru 5' až 3'. Dole dva triplety tohoto motivu. Protonizace N3 cytosinu v tripletu je nutná pro správnou vodíkovou vazbu s guaninem. V sekci C je purinový vazebný motiv. Nahoře vazba TFO na polypurinové vlákno DNA v antiparalelní orientaci. Dole tři kanonické triplety bází v tomto motivu-Hoogsteenovy vodíkové vazby a Watsonovy-Crickovy vodíkové vazby [82, 81]



Obrázek 5: Vznik triplexu, základní rysy interakce trojitě šroubovice založené na oligonukleotidech [78]

2.3. G-kvadruplexy

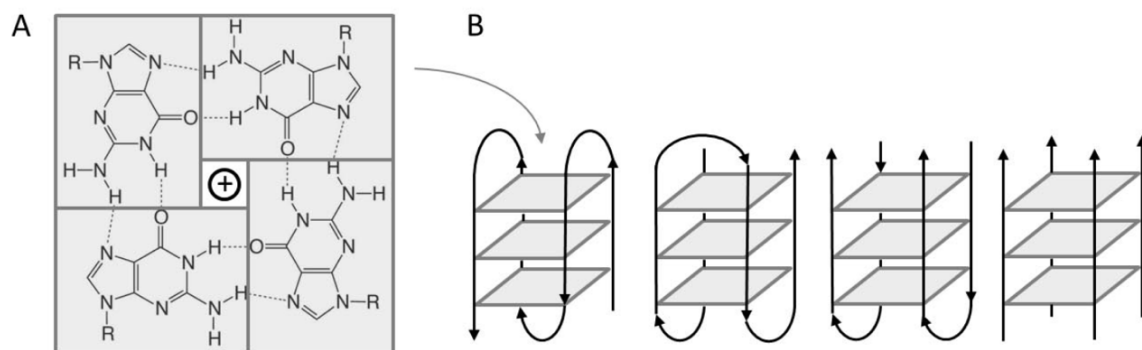
G-kvadruplexy jsou specifické struktury, které mohou vznikat v sekvencích bohatých na guanin v DNA nebo RNA. Mohou se vytvářet intramolekulárně v jednotlivém řetězci s vysokým obsahem guaninu nebo intermolekulárně, což vede k různým trojrozměrným konformacím [2]. Tyto struktury jsou umístěny v klíčových oblastech genomu, jako jsou telomery, promotory onkogenů, vazebná místa transkripčních faktorů a rekombinační místa [81].

Intramolekulární skládání vyžaduje přítomnost čtyř nebo více G-traktů v jednom řetězci, zatímco intermolekulární skládání může vzniknout ze dvou nebo čtyř řetězců, což vede k paralelním nebo antiparalelním strukturám v závislosti na orientaci řetězců v G-kvadruplexu [85, 86]. Tyto struktury často nacházejí společně s regulačními oblastmi genů a jsou asociovány s chromosomálními zlomy, translokacemi a genetickými poruchami. [87, 88, 89].

Sekvence kódující G-kvadruplexy se nachází v regulačních oblastech, hrají roli v transkripci a také regulaci RNA. Sekvence G-kvadruplexů jsou zastoupeny v promotorech onkogenů (KIT, MYC, KRAS). Jsou tedy významné pro rozvoj nádorových onemocnění, kde jsou terapeutickými cíli pro léčbu [90, 91].

Na Obrázku 6 v části A je zobrazena G-tetráda, což je plochá struktura, kde každá guaninová báze slouží jako donor i akceptor dvou vodíkových vazeb [92]. Tyto struktury jsou tvořeny čtyřmi guaninovými bázemi, které jsou uspořádané do rovinného kvartetu, kde každý guanin je vodíkově vázán na sousední guanin. Kvartety se pak mohou skládat nad sebou, čímž vzniká stabilní G-kvadruplexová struktura. G-kvadruplexy jsou charakteristické tím, že

obsahují vrstvy dvou nebo více G-tetrád, ve kterých jsou čtyři guaniny propojeny Hoogsteenovými H-vazbami a jsou centrálně koordinovány monovalentními kationty. Obecně platí, že stabilita G-kvadruplexu roste s počtem obsažených G-tetrád [93, 94].

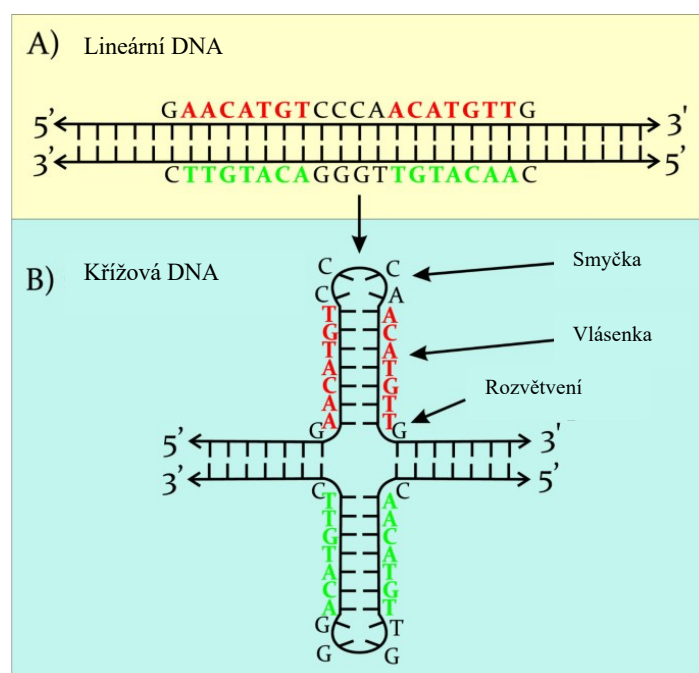


Obrázek 6: Obrázek 5: Guaninová tetráda a příklady G-kvadruplexu tvořeného jedním nebo více vláknů nukleových kyselin. Obrázek A – guaninová tetráda vytvořená na základě Hoogsteenova párování bází, Obrázek B – vrstvení guaninových tetrád a příklady G-kvadruplexů z jednoho, dvou a čtyř vláken nukleové kyseliny [89]

2.4. Křížová struktura

Tyto struktury vznikají z inverzně umístěných opakování, která mohou mít délku od 6 do několika stovek nukleotidů. Důležitou roli zde hraje složení repetice, což přispívá k vyšší stabilitě páru cytosin-guanin, stejně jako k párování mezi opakováními, která tvoří smyčku vlásenky [77]. Pro dosažení termodynamické stability křížové struktury je nezbytné, aby inverzní opakování mělo délku minimálně 6 bází [66, 97].

Na *Obrázku 7* je znázorněn přechod z inverzní repetice jako lineární DNA do křížové struktury DNA.

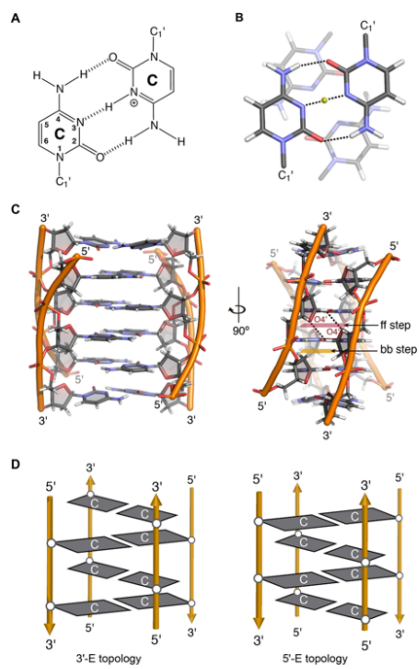


Obrázek 7: dlouhá inverzní repetice. Obrázek A zobrazuje lineární DNA a Obrázek B inverzní repetici jako křížovou strukturu [77].

2.5. I-motiv

I-motivy jsou čtyř vláknové sekundární struktury DNA, které se mohou tvořit v sekvencích bohatých na cytosin. Jsou stabilizovány kyselými podmínkami při pH menší než 7 a skládají se ze dvou paralelně vedených DNA duplexů, které jsou drženy pohromadě v antiparalelní orientaci interkalovanými páry bází cytosin-cytosin⁺ [99].

I-motivové struktury mohou být tvořeny v regulačních oblastech lidského genomu a podporují, že G4 struktury mohou hrát komplementární roli v regulaci genové exprese [100, 101]. Na *Obrázku 8* vidíme v sekci (A) C:C⁺ základní pár cytosinu, je zde zobrazena chemická struktura včetně umístění atomů vodíku, uhlíku, dusíku a kyslíku. V sekci (B) vidíme vrstvení mezi interagujícími páry bází C:C⁺, ukazuje interakce mezi dvěma cytosinovými bázemi. Tyto báze jsou vázány vodíkovými interakcemi. V sekci (C) jsou zobrazeny trojrozměrné struktury I-motivu v DNA. Na levé straně je pohled z boku, zatímco na pravé straně je tentýž motiv otočen o 90°, což ukazuje schodovitou strukturu. Na obrázku (D) je schéma znázorňující 3'E (vlevo) a 5'E (vpravo) interkalační topologie intermolekulární struktury I-motivu, znázorňuje uspořádání cytosinových bází a vazeb mezi nimi [101, 77].



Obrázek 8.: Zobrazení I-motivu [101]

3. EXPERIMENTÁLNÍ ČÁST

3.1. Analyzované organismy

Pro analýzu byly vybrány retroviry. Pomocí webové stránky NCBI (Národní centrum pro biotechnologické informace), která je veřejně dostupnou databází, byly získány genomy těchto virů. Celkový počet získaných genomů retrovirů činil přibližně 8000, nicméně kvůli nekompletnosti některých genomů bylo pro analýzu využito pouze 5778 z nich.

3.2. Metody

3.2.1. Software G4hunter

G4hunter je bioinformatický nástroj navržený k identifikaci a predikci potenciálních G-kvadruplexů v DNA sekvencích. Pro analýzu PQS sekvencí, které mohou tvořit kvadruplexy („putative quadruplex sequence“) byl použit webový nástroj G4hunter, který je součástí webového serveru DNA Analyser. K identifikaci potenciálních G-kvadruplexů G4hunter využívá algoritmy pro vyhledání sekvencí v DNA, které mají potenciál tvořit G-kvadruplexové struktury [103, 104, 105]. Výsledky jsou prezentovány pomocí G4hunter skóre.

Byl vytvořen vstupní soubor s koncovkou .csv. Tento soubor obsahoval název viru, NCBI ID viru a tag, který obsahoval podčeleď, ve které je retrovirus zařazen. NCBI ID je unikátní kód každého viru, pomocí něhož se z databáze NCBI importují sekvence nukleových kyselin, které právě G4hunter analyzuje [103].

Vstupní soubor byl nahrán na disk Google a ve skriptu (*Obrázek 10*) v Google colabory byla provedena samotná analýza. Byly nastaveny parametry pro analýzu následujícím způsobem, velikost okna pro zobrazení byla stanovena na 25 nt (počtu nukleotidů) a G4hunter skóre (prahová hodnota) na 1,2. Výstupní soubor ze skriptu obsahuje informace o počtu guaninů a cytosinů v DNA, informace zahrnuje i cytosin, protože je vždy analyzováno jedno vlákno molekuly DNA a cytosin je komplementární s guaninem. Soubor dále obsahuje informace o délkách daných sekvencí, počet PQS a také informace o frekvenci nálezu PQS v daném úseku G4hunter skóre.

Výsledky analýzy jsou ukládány do webové aplikace G4hunter, která nám poskytuje grafické zobrazení výsledků, počet PQS, frekvenci PQS, informace o délce sekvence a poměr GC. Tyto výsledky lze vidět na *Obrázku 9*.

Position	Length	Sequence	Score chart	G4Hunter score (abs)
359	28	AGCGGGGGA GAATTA GATCGATGGGAAA		1,107
4 325	51	ATTTTAAAA GAAAAGGGGGGATTGGGGGTACAGTGCAGGGGAAAGAATAG		1,314
5 250	29	TTATGGGGATACTTGGGCAGGAGTGGAAAG		1,138
5 792	26	AGCACTTGTGGAGATGGGGTGGAGA		1,154
5 795	38	ACTTGTGGAGATGGGGTGGAGATGGGCACCATGCTC		1,026
5 986	35	GCCTGTGTACCCACAGACCCCAACCCACAAGAAGT		-0,971
7 544	30	TCCTGGGGATTTGGGGTTGCTCTGGAAAAC		1
7 914	30	ATCGTTTCAGACCCACCTCCCAACCCGAG		-1,1
8 112	34	TGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCC		1,029
8 612	36	TTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTAC		0,944
9 005	39	GGGAGGCGTGGCTGGGCGGGACTGGGAGTGGCGAGCC		1,154

Obrázek 9: Snímek obrazovky aplikace G4 hunter, vstupní okno pro HIV 1 genom, zobrazení výskytu PQS, výsledky analýzy

```

complete analyses 05 MD.ipynb ☆
Soubor Upravit Zobrazit Vložit Běh Nástroje Nápověda Naposledy upraveno 27. března

+ Kód + Text

[ ] # we have to make pairs of same files
import glob
results = glob.glob('/content/drive/MyDrive/BP_Praktická/HIV/*.csv')
annotations = glob.glob('/content/drive/MyDrive/BP_Praktická/Anotace/*.csv')

# we have to sort them by name
results.sort()
annotations.sort()

[ ] results

['/content/drive/MyDrive/BP_Praktická/HIV/virus5098_d1973805-36cd-4559-95ac-f404c8ca99b0_result.csv']

[ ] annotations

['/content/drive/MyDrive/BP_Praktická/Anotace/Human_immunodeficiency_virus_1.csv']

[ ] # now lets create pairs, you can see that if we zip two lists of filepaths together
# we will make pairs of g4hunter result and annotation files :)
# you have to check it because of known server bug currently in progress of fixing! :)

for result, annotation in zip(results, annotations):
    print(result)
    print(annotation)
    print('-----')

/content/drive/MyDrive/BP_Praktická/HIV/virus5098_d1973805-36cd-4559-95ac-f404c8ca99b0_result.csv
/content/drive/MyDrive/BP_Praktická/Anotace/Human_immunodeficiency_virus_1.csv
-----

[ ] # now run overlay
# you have to again provide folder where you want to store overlay results

```

Obrázek 10: Snímek obrazovky skriptu G4 hunteru Google colabatory

3.2.1.1. G4hunter skóre

G4hunter je nástroj pro predikci G-kvadruplexů v nukleových kyselinách. Algoritmus, který G4hunter používá, počítá s oblastmi DNA nebo RNA, ve kterých se guanin vyskytuje se zvýšenou, nebo sníženou četností [103].

Při výpočtu G4hunter skóre mají G kladnou hodnotu a C hodnotu zápornou, přičemž A a T

mají hodnoty rovny 0. Pokud máme v sekvenci více G za sebou, hodnota skóre se tímto zvyšuje. Na druhé straně, když je sekvence s více C za sebou, hodnota se tedy snižuje. Součet skóre je pak dělen délkou okna. Nejvyšší dosažitelné skóre je 4/-4 [106] . Při zahájení analýzy je třeba na webové aplikaci nastavit nejmenší velikost G4hunter skóre, pro kterou budeme G4 nacházet, tato velikost byla nastavena na 1,2.

Příklady výpočtu G4 hunter skóre [105]:

- CCC

$$\Rightarrow -3-3-3 = -9 \Rightarrow -9 / 3 = -3$$

- GGGGGG

$$\Rightarrow 4+4+4+4+4+4 = 24 \Rightarrow 24 / 6 = 4$$

- ATCCCAAGGGAA

$$\Rightarrow 0+0-3-3-3+0+0+3+3+3+0+0 = 0 \Rightarrow 0 / 12 = 0$$

- ATGGATGGATGATGAT

$$\Rightarrow 0+0+2+2+0+0+2+2+0+0+1+0+0+1+0+0 = 10 \Rightarrow 10 / 16 = 0.625$$

3.3. Analýza genomu HIV pomocí webové aplikace G4 hunter

Pro analýzu konkrétního jednoho genomu byl vytvořen vstupní soubor ve formátu .csv., který obsahoval název viru, NCBI ID a podčeleď, do jaké je retrovirus zařazen. Následně byl soubor zanalyzován pomocí webové aplikace G4hunter (*Obrázek 11*). Jako referenční genom byl nastaven virus s NCBI ID AF033916, tedy *Human Immunodeficiency virus 1*.

NAME,NCBI_ID,TAG
Human_immunodeficiency_virus_1,AF033819,Lentivirus

Obrázek 11.: Snímek obrazovky vstupního souboru pro anotace

Výstupní soubory z tohoto kroku obsahovaly podrobné informace o genomu HIV. V tomto genomu byly získány informace o genu, CDS (kódující oblast genu). Ve výstupním souboru také nalezneme informace o 3' UTR a 5'UTR, což jsou nepřekládané oblasti na konci (3'UTR) a na začátku (5'UTR) mRNA. Tyto nepřekládané oblasti jsou důležité pro regulaci genové exprese a mohou hrát klíčovou roli při kontrole, kdy, kde a v jaké míře je gen exprimován [107].

Jsou zde také informace o vazebném místě primeru a ostatních nespecifikovaných oblastech HIV. U všech těchto vlastností G4hunter vygeneroval celkovou délku těchto oblastí a počet těchto oblastí v genomu. Další .csv soubor obsahoval informace o počtu PQS v jednotlivých intervalech G4hunter skóre.

3.4. Výpočty

Výpočet procentuálního obsahu GC [%]

$$GC_{\%} = \frac{\text{počet GC [nt]}}{\text{délka sekvence [nt]}} \cdot 100$$

Výpočet frekvence PQS [kb]

$$PQS_{fr} = \frac{\text{počet PQS}}{\text{délka sekvence [nt]}} \cdot 1000 [nt]$$

Pearsonův korelační koeficient

Nejčastěji se termín korelace používá v souvislosti s lineárním vztahem mezi 2 spojitými proměnnými a vyjadřuje se jako Pearsonova součinná korelace. Korelační koeficient se pohybuje v rozmezí od **-1 do +1**, kde 0 znamená, že neexistuje žádná lineární nebo monotónní asociace a vztah sílí a nakonec se blíží přímce (Pearsonova korelace) nebo stále rostoucí nebo klesající křivce (Spearmanova korelace), jak se koeficient blíží absolutní hodnotě 1 [108]. Pearsonův korelační koeficient se značí „r“.

Vzorec pro výpočet Pearsonova korelačního koeficientu [108]:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

Kde X_i a Y_i jsou jednotlivé hodnoty dvou proměnných, \bar{X} a \bar{Y} jsou průměrné hodnoty X a Y

Míra korelace dat v závislosti na Pearsonovým korelačním koeficientu: [108]

0,00 - 0,19- velmi slabá

0,20 - 0,39- slabá

0,40 - 0,59- střední

0,60 - 0,79- silná

0,80- 1,00- velmi silná

4. VÝSLEDKY A DISKUZE

Celkem bylo analyzováno 5778 genomů retrovirů.

4.1. Analýza PQS

V softwaru G4 hunter byly analyzovány genomy retrovirů. Výsledky této analýzy jsou shrnuty v *Tabulce 3*.

Tabulka 3: Výsledky obecné analýzy genomů retrovirů. Sekvence (počet zanalyzovaných sekvencí), Nejdelší (nejdelší analyzovaná sekvence), Nejkratší (nejkratší analyzovaná sekvence), PQS (počet nalezených PQS), GC (průměrný obsah guaninu a c cytosinu sekvencích, udáno v procentech), f_{min} . (minimální frekvence PQS na 1000 nt), f_{max} . (maximální frekvence PQS na 1000 nt), $f_{průměr}$ (průměrná frekvence PQS na 1000 nt)

Obecná analýza genomů retrovirů	
Sekvence	5778
Celková délka všech analyzovaných sekvencí [nt]	51917997
Nejdelší [nt]	13885
Nejkratší [nt]	6529
Počet PQS	81912
GC [%]	42,17
f_{min}	0
f_{max}	10,27
$f_{průměr}$	1,59

V celkové délce 51917997 nt (nukleotidů) analyzovaných genomů retrovirů bylo nalezeno 81912 PQS. Výskyt PQS souvisí s vyšším obsahem guaninu/cytosinu (G/C). PQS byly identifikovány v G4 hunteru při hodnotách G4hunter skóre 1,2 a vyšší. Velikost čtecího okna byla nastavena 25 nt. Obsah guaninu a cytosinu tvoří v průměru 42,11 %. Nejkratší genom o délce 6529 nt má virus s názvem *Human immunodeficiency virus 1*, zatímco nejdelší z analyzovaných genomů má *Simian foamy virus* s délkou 13885 nt. Průměrná frekvence výskytu PQS na 1000 nt je 1,59 u zjišťovaných genomů retrovirů. Maximální frekvence výskytu PQS byla nalezena u *Human T-lymphotropic viru 2*, u tohoto genomu bylo nalezeno 92 PQS v jeho celkové délce 8958 nt, frekvence byla vypočtena na 10,27 na 1000 nt.

Tabulka 4: Celkový počet PQS v jednotlivých intervalech G4hunter skóre, jejich průměrná frekvence výskytu na 1000 nt a celková délka PQS

G4 hunter skóre	Počet PQS	Frekvence PQS [kb]	Celková délka [nt]	Zastoupení PQS [%]
1,2-1,4	42073	0,810	1146863	51,36
1,4-1,6	23782	0,458	822826	29,03
1,6-1,8	7720	0,149	299418	9,42
1,8-2	3120	0,060	137158	3,81
>2	5217	0,100	259771	6,37

Z Tabulky 4 lze vyčíst, že se zvyšujícím se G4hunter skóre se snižuje frekvence výskytu PQS. Největší zastoupení PQS má G4hunter skóre v rozmezí od 1,2-1,4, toto rozmezí zahrnuje konkrétně intervaly od 1,2-1,399, stejně to je i u ostatních skupin skóre. Nejvyšší frekvence zastoupení PQS je opět ve skupině s G4hunter skóre od 1,2-1,4. Frekvence výskytu PQS jsou v retrovirech je obecně vyšší. Vyšší frekvence PQS v určitých oblastech genomu může naznačovat, že tyto oblasti mají zvláštní biologickou funkci, která souvisí s guaninovými kvadruplexy. Z Tabulky 4 lze také vysledovat, že nejdelsí součet délek je opět ve skupině, která tvoří PQS o skóre 1,2-1,4. Tento fakt souvisí s tím, že PQS jsou zde objeveny v největším množství a také s nejvyšší frekvencí.

Z dat byl také vypočten Pearsonův korelační koeficient, který byl vypočten z dat frekvencí a dat obsahů GC [%]. Tento koeficient byl stanoven na 0,466, což značí, že korelace mezi daty je střední. Právě střední Pearsonův korelační koeficient je nejspíše způsoben tím, že bylo analyzováno velké množství dat, která mohou obsahovat například odlehle hodnoty, které mohou způsobit střední korelaci.

4.1.1. Analýza PQS v genomu Human immunodeficiency virus 1

Výskyt PQS byl analyzován ve funkčních oblastech genomu *Human immunodeficiency viru 1*. Funkční oblasti, které také můžeme nazvat jako anotovaná místa, byly nacházeny v okolí genomu HIV-1. Výskyt PQS byl analyzován konkrétně před, uvnitř a za anotovanými místy v genomu HIV-1. Nalezeny byly tyto funkční oblasti genomu: gen, CDS (*kódující sekvence*), misc_feature (*„miscellaneous feature“*), regulatory (*regulační oblasti*), 5'UTR (*nepřekládaná oblast na konci 5'*), 3'UTR (*nepřekládaná oblast na konci 3'*) a primer bind (*vazba primeru*).

Tabulka 5: Seznam nalezených anotovaných znaků v genomu HIV 1, jejich počet a celková délka [nt]

Funkční oblast	Počet PQS	Celková délka znaků (nt)
gene	9	13801
CDS	11	9135
misc_feature	2	192
regulatory	1	6
5'UTR	1	85
3'UTR	1	455
primer bind	1	18

V Tabulce 5 vidíme, že nejvíce PQS v genomu HIV-1 bylo nalezeno ve funkční oblasti CDS, což je kódující sekvence. Druhá největší četnost PQS byla nalezena v okolí genu, který měl zároveň i nejdelší celkovou délku v genomu ze všech anotovaných funkčních oblastí.

Tabulka 6: Počet PQS před, uvnitř a za oblastí anotovaného genomu HIV 1

Funkční oblast	Počet PQS		
	před	uvnitř	za
gene	1	13	1
CDS	1	11	1
misc_feature	0	0	2
regulatory	0	0	2
5'UTR	0	2	0
3'UTR	0	2	0
primer bind	2	0	0

V tabulce 6 můžeme vidět, že nejvíce funkčních oblastí se nachází uvnitř anotovaného genomu HIV 1. V genomu HIV-1 zastává konkrétně gen 13 PQS. Hojně PQS je také uvnitř anotované oblasti CDS.

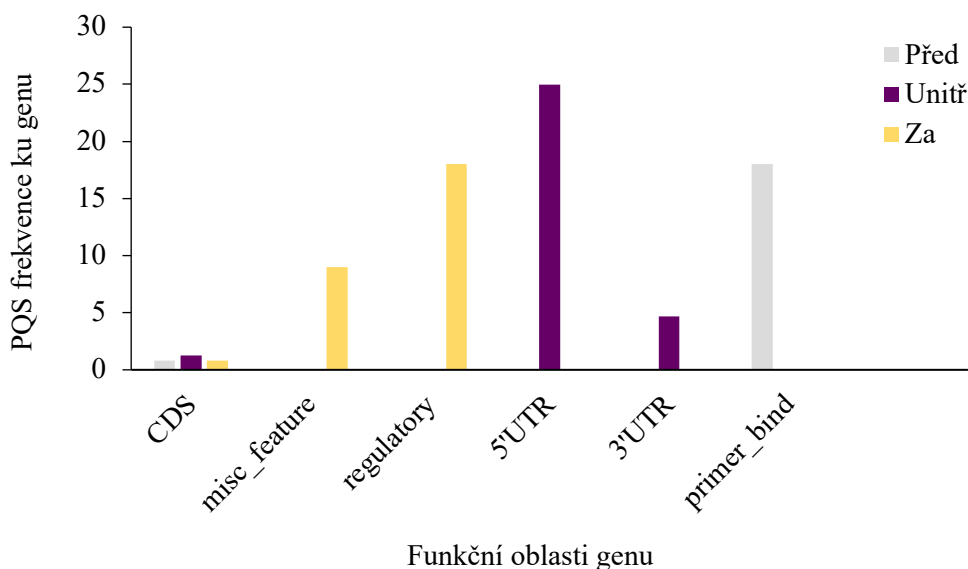
Tabulka 7: Frekvence PQS/kb před, uvnitř a za oblastí anotovaného genomu HIV-1

Frekvence PQS/kb			
Funkční oblast	před	uvnitř	za
gene	1,11	0,94	1,11
CDS	0,91	1,20	0,91
misc_feature	0,00	0,00	10,00
regulatory	0,00	0,00	20,00
5'UTR	0,00	23,53	0,00
3'UTR	0,00	4,40	0,00
primer bind	20,00	0,00	0,00

V *Tabulce 7* můžeme vidět jednotlivé frekvence výskytu PQS v jednotlivých oblastech genomu pro nalezené funkční oblasti. Vzhledem k faktu, že gen se nachází nejhojněji ve všech genomech a je také jeden z nejčastěji analyzovaných anotačních znaků, byl vytvořen graf frekvence výskytu PQS v jednotlivých funkčních oblastech právě vzhledem ke genu. Gen byl tedy zvolen jako standard pro porovnání všech dalších frekvencí pro ostatní funkční oblasti. Největší frekvence výskytu ku genu byla pozorována uvnitř anotovaného místa pro 5'UTR. U oblasti primer bind byla nalezena vysoká frekvence výskytu PQS před anotovaným místem v HIV-1 a u regulatory za anotovaným místem. U funkční oblasti CDS byly pozorovány podobné frekvence výskytu PQS ve funkční oblasti před, uvnitř i za anotovanou oblastí.

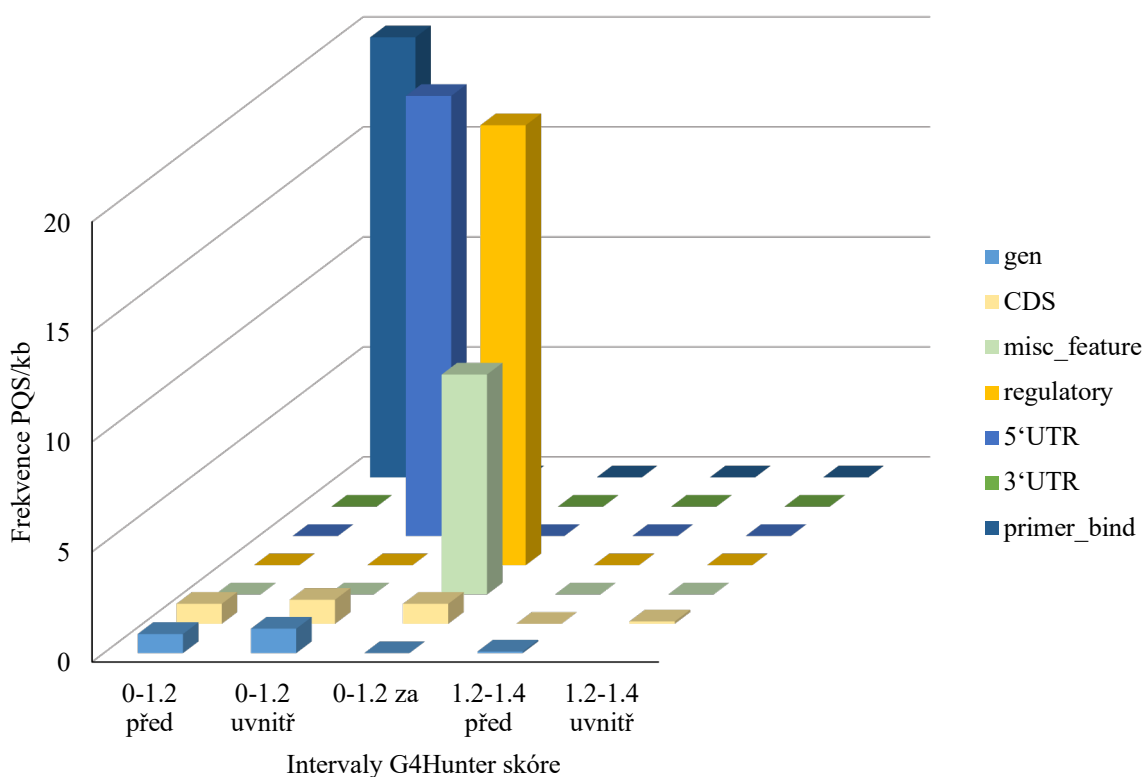
Výpočty pro jednotlivé frekvence byly provedeny jako podíl frekvence PQS dané funkční oblasti (před, uvnitř i za) a frekvence PQS genu dané oblasti. Výpočty frekvencí PQS jsou uvedeny v *Tabulce 7*.

Celkové grafické znázornění výskytu PQS v oblastech jednotlivých frekvencí ku genu viru HIV 1 je možné vidět na *Obrázku 12*. Podle analýzy ve skriptu jsme také zjistili, kolik je v genomu jak dlouhých oblastí a kolik PQS v nich máme.



Obrázek 12: Srovnání frekvencí PQS jednotlivých funkčních oblastí genu HIV 1 vzhledem k frekvenci PQS genu

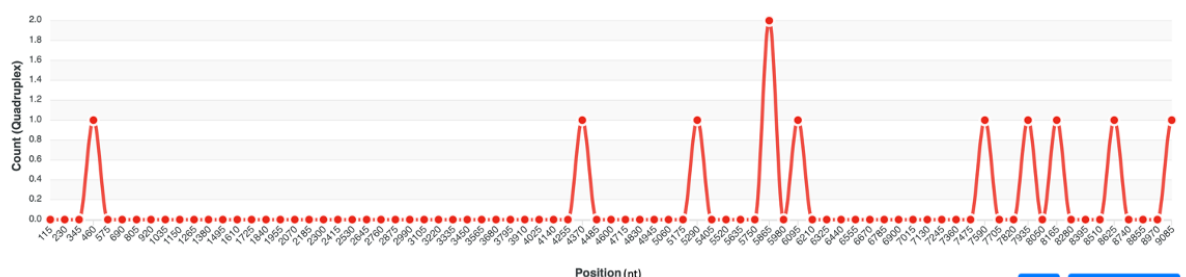
Na Obrázku 13 je zobrazena závislost frekvence výskytu PQS k jednotlivým intervalům G4hunter skóre. Zobrazeny byly všechny námi anotované funkční oblasti před, uvnitř a za těmito místy.



Obrázek 13: Celkové grafické znázornění frekvence PQS/kb v genomu viru HIV 1

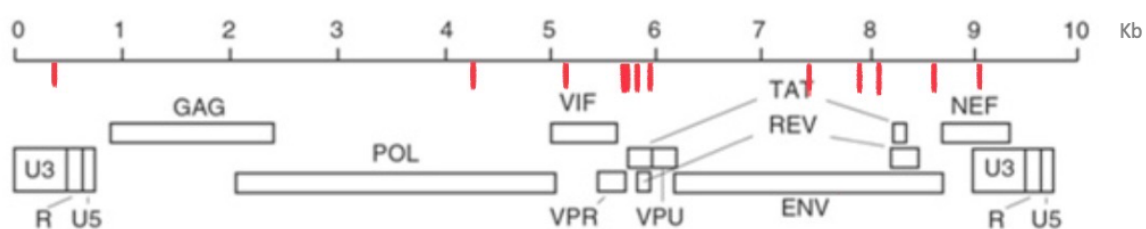
Na Obrázku 14 je vidět graf závislosti počtu G-kvadruplexů na pozici, kde se daný PQS nachází. V HIV-1 genomu bylo nalezeno 11 G4, přičemž na pozici 5865 nt byly nalezeny

2 G4. Na jednom místě může být nalezeno více G-kvadruplexů kvůli tomu, že výskyt těchto sekvencí přímo koreluje s výskytem G. Je-li v určité oblasti vyšší četnost výskytu guaninových sekvencí, může být vytvořeno více G-kvadruplexů najednou. Kvadruplexy se také mohou překrývat nebo být umístěny těsně vedle sebe [84].



Obrázek 14: Závislost počtu G4 na pozici

Na Obrázku 15 je graficky znázorněno umístění sekvencí daného proteinu HIV-1. Červeně označené čárky značí přítomnost G-kvadruplexů. Vidíme tedy, jak se v HIV-1 genomu vyskytují PQS a geny kódující proteiny v závislosti na sobě. Vidíme, že největší zastoupení na kb (kilobázi) genomu HIV-1 je obsažen gen kódující protein Pol. V úseku, kde se tento gen nachází je 1 G-kvadruplex. Frekvence výskytu PQS v genomu HIV-1 je přibližně 2.78 PQS na 1000 nt. Tato frekvence byla vypočtena jako podíl počtu PQS v genomu HIV 1 (27 PQS) ku celkové délce genomu (9700 nt) vynásobené 1000.



Obrázek 15: Zobrazení poloh proteinů v HIV 1 genomu a poloh PQS (červeně zobrazeno)

4.1.2. Srovnání výskytu PQS v genomech retrovirů s genomem *Homo sapiens*

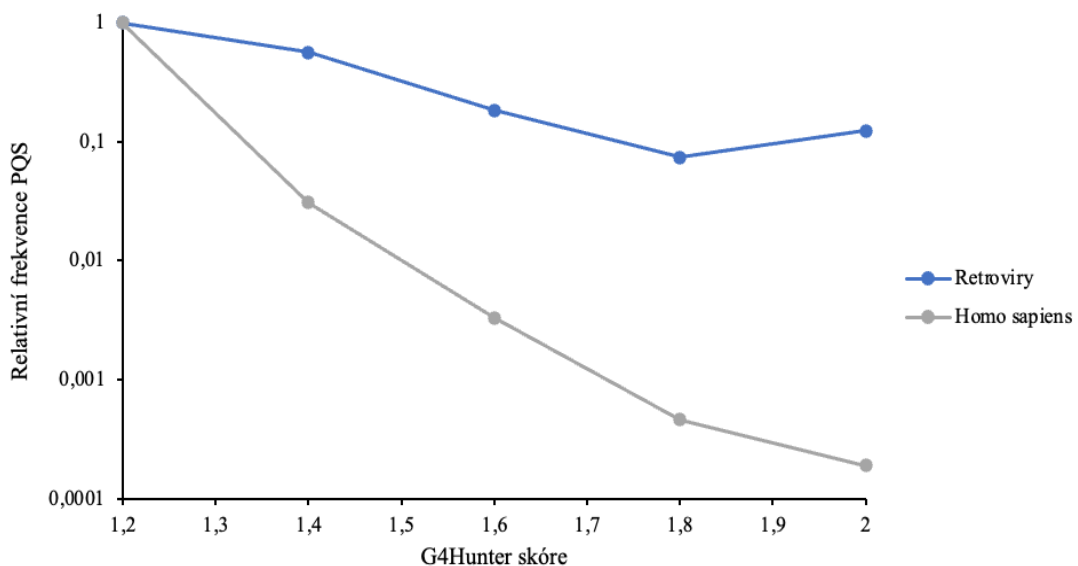
Genomy všech analyzovaných retrovirů byly srovnávány s genomy *Homo sapiens*. Data pro *Homo sapiens* byla získána ze zdroje [65]. V Tabulce 8 jsou zobrazeny počty PQS v jednotlivých intervalech jak u *Homo sapiens*, tak u retrovirů. Úsek 1,2 značí interval G4Hunter skóre 1,2-1,399, 1,4 značí interval G4Hunter skóre 1,4-1,599, stejně je to i u dalších skupin G4hunter skóre.

Tabulka 8: Počty PQS v jednotlivých intervalech G5 hunter skóre pro genomy *Homo sapiens* a genomy retrovirů. Frekvence značí relativní normalizovanou frekvenci pro konkrétní úsek G4 hunter skóre.

Kategorie	<i>Homo sapiens</i>	Retroviry
PQS 1,2-1,4	5112318	42073
Frekvence	1	1
PQS 1,4-1,6	158102	23782
Frekvence	0,03093	0,5653
PQS 1,6-1,8	16953	7720
Frekvence	0,00332	0,1835
PQS 1,8-2,0	2372	3120
Frekvence	0,00046	0,0742
PQS >2,0	951	5217
Frekvence	0,00019	0,1240

Podle frekvencí v intervalu 1,2-1,4 byly normalizovány frekvence pro další skupiny. Relativní frekvence pro jednotlivé rozmezí G4hunter skóre byly tedy vypočteny jako poměr frekvence v konkrétním úseku ku frekvenci pro 1,2-1,4. Právě ve skupině G4hunter skóre od 1,2-1,4 se u genomu *Homo sapiens* nachází 96,6 % PQS, zatímco u retrovirů je procentuální nález PQS mnohem menší (51,36 %). Z tohoto tvrzení vyplývá, že retroviry obsahují mnohem stabilnější PQS než *Homo sapiens*.

Na Obrázku 16 vidíme tendenci klesající křivky u obou analyzovaných skupin. Křivka s daty *Homo sapiens* je klesající mnohem strměji než křivka s daty retrovirů. Z grafu lze vyčíst, že se zvyšujícím se G4hunter skóre klesá relativní frekvence výskytu PQS.



Obrázek 16: Graf závislosti relativní frekvence PQS na G4 hunter skóre

5. ZÁVĚR

Bakalářská práce je zaměřena na prozkoumání lokálních struktur v genomech retrovirů. Tato práce se zabývá G-kvadruplexy, což jsou struktury, které mohou mít zásadní vliv na regulaci genové exprese a na proces replikace. Využitím bioinformatického nástroje G4 Hunter byly analyzovány specifické sekvence DNA a RNA, které představují potenciální G-kvadruplexové motivy v genomu vybraných retrovirů.

Výsledky ukázaly, že tyto struktury se objevují na klíkových místech genomů, což naznačuje, že by mohly hrát důležitou roli v životním cyklu retrovirů. To otevírá nové možnosti pro další výzkum, který by mohl přispět k objevu nových terapeutických cest pro léčbu retrovirových infekcí.

Nejprve byla provedena analýza genomů retrovirů, která byla zaměřena na studium PQS, což jsou sekvence s potenciálem tvořit G-kvadruplexy v jednotlivých genomech. Nejvyšší frekvence výskytu PQS byla nalezena ve skupině s G4hunter skóre od 1,2-1,399. V tomto úseku G4hunter skóre bylo nalezeno 41831 PQS. Fakt, že retroviry jsou obecně velmi bohaté na tyto úseky naznačuje, že hrají velmi důležitou biologickou roli. V práci bylo zjištěno, že výskyt PQS v retrovirových genomech se shoduje se zjištěním z článku [4], že retroviry mají vyšší frekvenci výskytu PQS než viry způsobující akutní infekce. Platí, že čím více PQS obsahují genomy virů, tím horší infekci viry způsobují. Viry s vyšší frekvencí výskytu PQS jsou s velkou pravděpodobností spojeny se závažnými infekcemi, právě tento fakt může být důležitý pro vývoj antivirových terapií, které jsou zaměřeny na G-kvadruplexy [4].

V dalším kroku je experimentální část zaměřena na anotaci jednoho genomu retrovirů. Byl vybrán genom HIV-1, což je jeden z nejvíce prostudovaných retrovirů jako potenciální antivirový terapeutický cíl [2]. G-kvadruplexy se velmi často vyskytují v regulovaných oblastech genomů, jako jsou promotory nebo telomery. Z databáze NCBI bylo zjištěno, že genom HIV-1 obsahuje několik funkčních oblastí, konkrétně gen, CDS, misc_feature, regulatory, 5'UTR, 3'UTR a primer bind. Toto odpovídá tvrzení z článku, že PQS jsou bohatě zastoupeny v oblastech jako 5'UTR a 3'UTR, které jsou klíčové pro regulaci genové exprese a replikaci virů [4, 90]. V genomu HIV-1 bylo nalezeno nejvíce anotovaných oblastí genu, konkrétně 13 PQS uvnitř anotované oblasti, 1 PQS před a 1 PQS za anotovanou oblastí genu. Gen *Nef* je gen kódující protein, který je zastoupen v genomu HIV-1. Podle článku [90] můžeme tvrdit, že stabilizace G-kvadruplexů v *Nef* oblasti může inhibovat expresi *Nef* proteinu a následně potlačit infekčnost HIV-1. Toto je důkaz, jak G-kvadruplexy mohou regulovat genovou expresi a infekčnost HIV-1. Proteiny, které jsou v genomu HIV-1 obsaženy hrají velmi důležitou funkci. *Nef* hraje zásadní roli v úpravě hostitelské buňky a umožňuje viru efektivněji se replikovat a šířit, zatímco *Pol* a *Gag* proteiny se mnohem více podílejí na samotném procesu replikace a sestavení viru [90].

Ve srovnání genomu HIV-1 s genomem HDV (*Hepatitis delta virus*) z článku [109] ukazuje, že vysoký výskyt PQS v genomech HIV-1 (27 PQS) a HDV (6675 PQS) naznačuje jejich klíčovou roli v regulaci virové replikace a genové exprese. HDV má výrazně vyšší frekvenci PQS ve svém genomu (8.25 PQS na 1000 nt) ve srovnání s HIV-1 (2.78 PQS na 1000 nt), což ukazuje rozdíly v jejich genomových strukturách a životních cyklech. Při srovnání genomu HIV 1 s genomy SARS-CoV-2 (Koronavirus 2 způsobující těžký akutní respirační syndrom), SARS-CoV (Koronavirus způsobující těžký akutní respirační syndrom) a MERS-CoV (Koronavirus způsobující respirační syndrom na Blízkém východě) [110]. Při srovnání dat z článku s genomem HIV 1 bylo zjištěno, že genom HIV 1 má mnohem vyšší frekvenci výskytu PQS než analyzované genomy z článku. Toto naznačuje, že nízký počet PQS v SARS-CoV-2 (0,033 PQS na 1000 nt) a absence PQS v MERS-CoV ukazují na odlišné strategie těchto virů při regulaci genomu a možnou eliminaci těchto struktur, může to souviset s jeho mechanismy replikace. Vysoký výskyt PQS v genomu HIV-1 naznačuje, že

G-kvadruplexy mohou hrát klíčovou roli v regulaci replikaci a genové exprese, což z nich činí potenciální cíle pro antivirovou terapii.

V závěrečném kroku experimentální části byla srovnána naměřená data pomocí bioinformatické analýzy a data *Homo sapiens*. Srovnání frekvence PQS ukázala, že poměrově retroviry obsahují ve svých genomech mnohem více stabilních PQS než *Homo sapiens*. To znovu potvrzuje fakt, že tyto struktury jsou vhodné jako terapeutické cíle.

6. ZDROJE

- [1] LEVY, Jay A. The Retroviridae, Structure and Classification of Retroviruses. In: *The Retroviridae*. Springer Science & Business Media, 2013, Chapter 2. ISBN 1489917217, 9781489917218.
- [2] RUGGIERO, Emanuela a RICHTER, Sara N. Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. Online. *Quadruplex Nucleic Acids As Targets For Medicinal Chemistry. Annual Reports in Medicinal Chemistry*. 2020, s. 101-131. ISBN 9780128210178. Dostupné z: <https://doi.org/10.1016/bs.armc.2020.04.001>. [cit. 2024-05-01].
- [4] BOHÁLOVÁ, Natália; CANTARA, Alessio; BARTAS, Martin; KAURA, Patrik; ŠŤASTNÝ, Jiří et al. Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. Online. *Biochimie*. 2021, roč. 186, s. 13-27. ISSN 03009084. Dostupné z: <https://doi.org/10.1016/j.biochi.2021.03.017>. [cit. 2024-05-01].
- [5] BRÁZDA, Václav; ČERVENĚ, Jiří; BARTAS, Martin; MIKYSKOVÁ, Nikol; COUFAL, Jan et al. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. Online. *Molecules*. 2018, roč. 23, č. 9. ISSN 1420-3049. Dostupné z: <https://doi.org/10.3390/molecules23092341>. [cit. 2024-05-16].
- [6] DEEKS, Steven G.; OVERBAUGH, Julie; PHILLIPS, Andrew a BUCHBINDER, Susan. HIV infection. Online. *Nature Reviews Disease Primers*. 2015, roč. 1, č. 1. ISSN 2056-676X. Dostupné z: <https://doi.org/10.1038/nrdp.2015.35>. [cit. 2024-05-14].
- [7] SMALLA, Kornelia; JECHALKE, Sven a TOP, Eva M. Plasmid Detection, Characterization, and Ecology. Online. *Microbiology spectrum*. 2015, roč. 3, č. 1, s. PLAS-0038-2014. ISSN 2165-0497. Dostupné z: <https://doi.org/10.1128/microbiolspec.plas-0038-2014>. [cit. 2024-04-27].
- [8] COFFIN, John M. Structure and Classification of Retroviruses. Online. *The Retroviridae*. 1992, s. 19-49. ISBN 978-1-4613-6488-7. Dostupné z: https://doi.org/10.1007/978-1-4615-3372-6_2. [cit. 2024-05-14].
- [9] LORSON, Christian a KUMAR, Sudhir. Genomes . Second Edition. By T A Brown. Published by Wiley-Liss, New York, by arrangement with BIOS Scientific, Oxford . \$97.50. xxvii + 572 p; ill.; index. ISBN: 0-471-25046-5. 2002. Online. *The Quarterly review of biology*. 2003, roč. 78, č. 2, s. 225-225. ISSN 0033-5770. Dostupné z: <https://doi.org/10.1086/377958>. [cit. 2024-04-27].
- [10] CALARGE, Chadi a MILLER, Del. Predictors of Risperidone and 9-Hydroxyrisperidone Serum Concentration in Children and Adolescents. Online. *Journal of Child and Adolescent Psychopharmacology; New Rochelle*. April 2001, roč. 21, č. 2, s. 163-9. Dostupné z: <https://doi.org/10.1089/cap.2010.0038>. [cit. 2021-04-26].
- [11] ZHANG, Wei; CAO, Sheng; L. MARTIN, Jessica; D. MUELLER, Joachim a M. MANSKY, Louis. Morphology and ultrastructure of retrovirus particles. Online. *AIMS Biophysics*. 2015, roč. 2, č. 3, s. 343-369. ISSN 2377-9098. Dostupné z: <https://doi.org/10.3934/biophy.2015.3.343>. [cit. 2024-05-14].
- [12] The Cell: A Molecular Approach . Seventh Edition. By Geoffrey M. Cooper and Robert E. Hausman. Sunderland (Massachusetts): Sinauer Associates. \$155.95. xxv + 832 p.; ill.; index. ISBN: 978-1-60535-290-9. [A companion website is available.] 2016. Online. *The Quarterly Review of Biology*. 2017, roč. 92, č. 2, s. 202-202. ISSN 0033-5770. Dostupné z: <https://doi.org/10.1086/692220>. [cit. 2024-04-27].

- [13] GULLEDGE, Thomas. What is integration? Online. 2006, roč. 106, č. 1, s. 5-20. ISSN 0263-5577. Dostupné z: <https://doi.org/10.1108/02635570610640979>. [cit. 2024-04-27].
- [14] KUZMINOV, Andrei. The Precarious Prokaryotic Chromosome. Online. *Journal of Bacteriology*. 2014, roč. 196, č. 10, s. 1793-1806. ISSN 0021-9193. Dostupné z: <https://doi.org/10.1128/JB.00022-14>. [cit. 2024-05-01].
- [15] TOUCHON, Marie a ROCHA, Eduardo P.C. Coevolution of the Organization and Structure of Prokaryotic Genomes. Online. *Cold Spring Harbor Perspectives in Biology*. 2016, roč. 8, č. 1. ISSN 1943-0264. Dostupné z: <https://doi.org/10.1101/cshperspect.a018168>. [cit. 2024-04-27].
- [16] JOHNS, Paul; XING, Dong; CHANG, Chi-Han; LI, Heng a XIE, X. Sunney. *Cellular mechanisms of neurological disease*. Online. Clinical Neuroscience. Elsevier, 2014. ISBN 9780443103216. Dostupné z: <https://doi.org/10.1016/B978-0-443-10321-6.00008-4>. [cit. 2024-05-16].
- [17] NURK, Sergey; KOREN, Sergey; RHIE, Arang; RAUTIAINEN, Mikko; BZIKADZE, Andrey V et al. The complete sequence of a human genome. Online. *Science (American Association for the Advancement of Science)*. 2022, roč. 376, č. 6588, s. 44-53. ISSN 0036-8075. Dostupné z: <https://doi.org/10.1126/science.abj6987>. [cit. 2024-04-27].
- [18] WILLS, John W. a CRAVEN, Rebecca C. Form, function, and use of retroviral Gag proteins. In: *AIDS*. 5(6). 1991, s. 639-654.
- [19] FISCHLE, Wolfgang; WANG, Yanming a ALLIS, C David. Histone and chromatin cross-talk. Online. *Current opinion in cell biology*. 2003, roč. 15, č. 2, s. 172-183. ISSN 0955-0674. Dostupné z: [https://doi.org/10.1016/S0955-0674\(03\)00013-9](https://doi.org/10.1016/S0955-0674(03)00013-9). [cit. 2024-04-27].
- [20] MURIAUX, Delphine a DARLIX, Jean-Luc. Properties and functions of the nucleocapsid protein in virus assembly. Online. *RNA Biology*. 2014, roč. 7, č. 6, s. 744-753. ISSN 1547-6286. Dostupné z: <https://doi.org/10.4161/rna.7.6.14065>. [cit. 2024-05-14].
- [21] DARLIX, Jean-Luc; LAPADAT-TAPOLSKY, Mary; DE ROCQUIGNY, Hugues a ROQUES, Bernard P. First Glimpses at Structure-function Relationships of the Nucleocapsid Protein of Retroviruses. Online. *Journal of Molecular Biology*. 1995, roč. 254, č. 4, s. 523-537. ISSN 00222836. Dostupné z: <https://doi.org/10.1006/jmbi.1995.0635>. [cit. 2024-05-14].
- [22] KOUZARIDES, Tony. Chromatin Modifications and Their Function. Online. *Cell*. 2007, roč. 128, č. 4, s. 693-705. ISSN 0092-8674. Dostupné z: <https://doi.org/10.1016/j.cell.2007.02.005>. [cit. 2024-04-27].
- [23] MÜLLER, Manuel M a MUIR, Tom W. Histones: At the Crossroads of Peptide and Protein Chemistry. Online. *Chemical reviews*. 2015, roč. 115, č. 6, s. 2296-2349. ISSN 0009-2665. Dostupné z: <https://doi.org/10.1021/cr5003529>. [cit. 2024-04-27].
- [24] BOWERMAN, B; BROWN, P O; BISHOP, J M a VARMUS, H E. A nucleoprotein complex mediates the integration of retroviral DNA. Online. 1989, roč. 3, č. 4, s. 469-478. ISSN 0890-9369. Dostupné z: <https://doi.org/10.1101/gad.3.4.469>. [cit. 2024-05-14].
- [25] DUNN, Ben M; GOODENOW, Maureen M; GUSTCHINA, Alla a WLODAWER, Alexander. Online. *Genome Biology*. Roč. 3, č. 4. ISSN 14656906. Dostupné z: <https://doi.org/10.1186/gb-2002-3-4-reviews3006>. [cit. 2024-05-14].
- [26] BURGER, Gertraud; GRAY, Michael W a LANG, B. Franz. Mitochondrial genomes:

- anything goes. Online. *Trends in genetics*. 2003, roč. 19, č. 12, s. 709-716. ISSN 0168-9525. Dostupné z: <https://doi.org/10.1016/j.tig.2003.10.012>. [cit. 2024-04-27].
- [27] OROSZLAN, S. a LUFTIG, R. B. Retroviral Proteinases. Online. *Retroviruses. Current Topics in Microbiology and Immunology*. 1990, s. 153-185. ISBN 978-3-642-75220-9. Dostupné z: https://doi.org/10.1007/978-3-642-75218-6_6. [cit. 2024-05-14].
- [28] TAANMAN, Jan-willem. The mitochondrial genome: structure, transcription, translation and replication. Online. *BBA - Bioenergetics*. 1999, roč. 1410, č. 2, s. 103-123. ISSN 0005-2728. Dostupné z: [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3). [cit. 2024-04-27].
- [30] BERNT, Matthias; BRABAND, Anke; SCHIERWATER, Bernd a STADLER, Peter F. Genetic aspects of mitochondrial genome evolution. Online. *Molecular phylogenetics and evolution*. 2013, roč. 69, č. 2, s. 328-338. ISSN 1055-7903. Dostupné z: <https://doi.org/10.1016/j.ympev.2012.10.020>. [cit. 2024-04-27].
- [31] DAS, Suman Ranjan a JAMEEL, Shahid. Biology of the HIV Nef protein. Online. *Indian journal of medical research (New Delhi, India: 1994)*. 2005, roč. 121, č. 4, s. 315. ISSN 0971-5916. [cit. 2024-05-01].
- [32] PEREIRA, Estela A. a DASILVA, Luis L. P. HIV-1 Nef: Taking Control of Protein Trafficking. Online. *Traffic (Copenhagen, Denmark)*. 2016, roč. 17, č. 9, s. 976-996. ISSN 1398-9219. Dostupné z: <https://doi.org/10.1111/tra.12412>. [cit. 2024-05-01].
- [33] BURK, Robert D; HARARI, Ariana a CHEN, Zigui. Human papillomavirus genome variants. Online. *Virology (New York, N.Y.)*. 2013, roč. 445, č. 1, s. 232-243. ISSN 0042-6822. Dostupné z: <https://doi.org/10.1016/j.virol.2013.07.018>. [cit. 2024-04-27].
- [34] TAN, Longzhi; XING, Dong; CHANG, Chi-han; LI, Heng a XIE, X Sunney. Three-dimensional genome structures of single diploid human cells. Online. *Science (American Association for the Advancement of Science)*. 2018, roč. 361, č. 6405, s. 924-928. ISSN 0036-8075. Dostupné z: <https://doi.org/10.1126/science.aat5641>. [cit. 2024-04-27].
- [35] HU, Wei-Shau a TEMIN, Howard M. Retroviral Recombination and Reverse Transcription. Online. *Science*. 1990, roč. 250, č. 4985, s. 1227-1233. ISSN 0036-8075. Dostupné z: <https://doi.org/10.1126/science.1700865>. [cit. 2024-05-14].
- [37] WILHELM, M. a WILHELM, F.-X. Reverse transcription of retroviruses and LTR retrotransposons. Online. *Cellular and Molecular Life Sciences*. 2001, roč. 58, č. 9, s. 1246-1262. ISSN 1420-682X. Dostupné z: <https://doi.org/10.1007/PL00000937>. [cit. 2024-05-14].
- [38] ZHENG, Zhi-Ming; BRABAND, Anke a SCHIERWATER, Bernd. Papillomavirus genome structure, expression, and post-transcriptional regulation: structure, transcription, translation and replication. Online. In: *Frontiers in Bioscience*. S. 328-338. ISBN 9780443103216. Dostupné z: [doi:10.2741/1971](https://doi.org/10.2741/1971). [cit. 2024-04-27].
- [39] MARAMOROSCH, Karl; SHATKIN, Aaron J; MURPHY, Frederick A a MURPHY, Frederick A. Replication and Partitioning of Papillomavirus Genomes. Online. *Advances in Virus Research*. 2008, roč. 72, s. 155-205. ISBN 0123743222. ISSN 0065-3527. Dostupné z: [https://doi.org/10.1016/S0065-3527\(08\)00404-1](https://doi.org/10.1016/S0065-3527(08)00404-1). [cit. 2024-04-27].
- [40] SWOFFORD, Sharon E. Infections in Transplant and Oncology Patients, An Issue of Infectious Disease Clinics/Infectious Disease Clinics of North America. Online. *Nephrology nursing journal: journal of the American Nephrology Nurses' Association*. 2010, roč. 37, č. 6, s. 669. ISSN 1526-744X. [cit. 2024-04-27].
- [42] CHABEDA, Aleyo; YANEZ, Romana J.R.; LAMPRECHT, Renate; MEYERS, Ann

- E.; RYBICKI, Edward P. et al. Therapeutic vaccines for high-risk HPV-associated diseases. Online. *Papillomavirus research*. 2018, roč. 5, s. 46-58. ISSN 2405-8521. Dostupné z: <https://doi.org/10.1016/j.pvr.2017.12.006>. [cit. 2024-04-27].
- [43] HIRANO, Tatsuya. Chromosome Dynamics during Mitosis. Online. *Cold Spring Harbor perspectives in biology*. 2015, roč. 7, č. 6, s. a015792. ISSN 1943-0264. Dostupné z: <https://doi.org/10.1101/cshperspect.a015792>. [cit. 2024-04-27].
- [44] BROWN, P. O. Strategies of Replication. Online. In: *Retroviruses: Strategies of Replication*. Swanstrom: Vogt, PK, Eds, 1990, s. 19-48. [cit. 2024-05-14].
- [45] LAN, Ke a LUO, Min-hua. Herpesviruses: epidemiology, pathogenesis, and interventions. Online. *Virologica Sinica*. 2017, roč. 32, č. 5, s. 347-348. ISSN 1674-0769. Dostupné z: <https://doi.org/10.1007/s12250-017-4108-2>. [cit. 2024-04-27].
- [46] FADLY, Aly M. Isolation and identification of avian leukosis viruses: A review. Online. *Avian Pathology*. 2010, roč. 29, č. 6, s. 529-535. ISSN 0307-9457. Dostupné z: <https://doi.org/10.1080/03079450020016760>. [cit. 2024-05-14].
- [47] PAYNE, L. N a FADLY, A. M. Leukosis/sarcoma group. In: *Diseases of Poultry*. 1997, s. 414-466.
- [48] HUPPERT, Julian L. a BALASUBRAMANIAN, Shankar. Prevalence of quadruplexes in the human genome. Online. *Nucleic acids research*. 2005, roč. 33, č. 9, s. 2908-2916. ISSN 0305-1048. Dostupné z: <https://doi.org/10.1093/nar/gki609>. [cit. 2024-04-27].
- [51] BOCHMAN, Matthew L; PAESCHKE, Kotrin a ZAKIAN, Virginia A. DNA secondary structures: stability and function of G-quadruplex structures. Online. *Nature reviews. Genetics*. 2012, roč. 13, č. 11, s. 770-780. ISSN 1471-0056. Dostupné z: <https://doi.org/10.1038/nrg3296>. [cit. 2024-04-27].
- [52] MATSUOKA, Masao a JEANG, Kuan-Teh. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. Online. *Nature Reviews Cancer*. 2007, roč. 7, č. 4, s. 270-280. ISSN 1474-175X. Dostupné z: <https://doi.org/10.1038/nrc2111>. [cit. 2024-05-14].
- [53] KOLESNIKOVA, Sofia a CURTIS, Edward A. Structure and Function of Multimeric G-Quadruplexes. Online. *Molecules (Basel, Switzerland)*. 2019, roč. 24, č. 17, s. 3074. ISSN 1420-3049. Dostupné z: <https://doi.org/10.3390/molecules24173074>. [cit. 2024-04-27].
- [54] REIN, Alan. Murine Leukemia Viruses: Objects and Organisms. Online. *Advances in Virology*. 2011, roč. 2011, s. 1-14. ISSN 1687-8639. Dostupné z: <https://doi.org/10.1186/1743-422X-1-14>. [cit. 2024-05-13].
- [55] SLIVA, Katja; ERLWEIN, Otto; BITTNER, Alexandra a SCHNIERLE, Barbara S. Online. *Virology Journal*. Roč. 1, č. 1. ISSN 1743422X. Dostupné z: <https://doi.org/10.1186/1743-422X-1-14>. [cit. 2024-05-13].
- [57] SIPE, Theresa Ann; BARHAM, Terrika L.; JOHNSON, Wayne D.; JOSEPH, Heather A.; TUNGOL-ASHMON, Maria Luisa et al. Structural Interventions in HIV Prevention: A Taxonomy and Descriptive Systematic Review. Online. *AIDS and Behavior*. 2017, roč. 21, č. 12, s. 3366-3430. ISSN 1090-7165. Dostupné z: <https://doi.org/10.1007/s10461-017-1965-5>. [cit. 2024-05-13].
- [58] GAINES, H.; VON SYDOW, M.; PEHRSON, P. O. a LUNDBEGH, P. Clinical picture of primary HIV infection presenting as a glandular-fever-like illness. Online. *BMJ*. 1988, roč. 297, č. 6660, s. 1363-1368. ISSN 0959-8138. Dostupné z: <https://doi.org/10.1136/bmj.297.6660.1363>. [cit. 2024-05-14].
- [59] CANG, Xiaohui; ŠPONER, Jiří a CHEATHAM, 3rd. Explaining the varied glycosidic

- conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. Online. *Nucleic acids research*. 2011, roč. 39, č. 10, s. 4499-4512. ISSN 0305-1048. Dostupné z: <https://doi.org/10.1093/nar/gkr031>. [cit. 2024-04-27].
- [60] KUKHANOVA, M. K. Anti-HIV nucleoside drugs: A retrospective view into the future. Online. *Molecular Biology*. 2012, roč. 46, č. 6, s. 768-779. ISSN 0026-8933. Dostupné z: <https://doi.org/10.1134/S002689331206012X>. [cit. 2024-05-13].
- [61] HOOKS, John J.; GIBBS, Clarence J.; CUTCHINS, E. C.; ROGERS, N. G.; LAMPERT, P. et al. Characterization and distribution of two new foamy viruses isolated from chimpanzees. Online. *Archiv für die gesamte Virusforschung*. 1972, roč. 38, č. 1, s. 38-55. ISSN 0304-8608. Dostupné z: <https://doi.org/10.1007/BF01241354>. [cit. 2024-05-13].
- [62] PEDERSEN, N.C.; YAMAMOTO, J.K.; ISHIDA, T. a HANSEN, H. Feline immunodeficiency virus infection. Online. *Veterinary Immunology and Immunopathology*. 1989, roč. 21, č. 1, s. 111-129. ISSN 01652427. Dostupné z: [https://doi.org/10.1016/0165-2427\(89\)90134-7](https://doi.org/10.1016/0165-2427(89)90134-7). [cit. 2024-05-13].
- [64] YU, Shuyuan F.; BALDWIN, David N.; GWYNN, Samuel R.; YENDAPALLI, Suneetha a LINIAL, Maxine L. Human Foamy Virus Replication: A Pathway Distinct from That of Retroviruses and Hepadnaviruses. Online. *Science*. 1996, roč. 271, č. 5255, s. 1579-1582. ISSN 0036-8075. Dostupné z: <https://doi.org/10.1126/science.271.5255.1579>. [cit. 2024-05-13].
- [65] BRÁZDA, Václav, Yu LUO, Martin BARTAS, et al. G-Quadruplexes in the Archaea Domain. *Biomolecules* [online]. 2020, 10(9) [cit. 2024-04-20]. ISSN 2218-273X. Dostupné z: [doi:10.3390/biom10091349](https://doi.org/10.3390/biom10091349).
- [66] RUGGIERO, Emanuela; ZANIN, Irene; TERRERI, Marianna a RICHTER, Sara N. G-Quadruplex Targeting in the Fight against Viruses: An Update. Online. *International journal of molecular sciences*. 2021, roč. 22, č. 20, s. 10984. ISSN 1422-0067. Dostupné z: <https://doi.org/10.3390/ijms222010984>. [cit. 2024-04-27].
- [67] GILBOA, Eli; MITRA, Sudha W.; GOFF, Stephen a BALTIMORE, David. A detailed model of reverse transcription and tests of crucial aspects. Online. *Cell*. 1979, roč. 18, č. 1, s. 93-100. ISSN 00928674. Dostupné z: [https://doi.org/10.1016/0092-8674\(79\)90357-X](https://doi.org/10.1016/0092-8674(79)90357-X). [cit. 2024-05-14].
- [68] HUANG, Wei-chun; TSENG, Ting-yuan; CHEN, Ying-ting; CHANG, Cheng-chung; WANG, Zi-fu et al. Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. Online. *Nucleic acids research*. 2015, roč. 43, č. 21, s. 10102-10113. ISSN 0305-1048. Dostupné z: <https://doi.org/10.1093/nar/gkv1061>. [cit. 2024-04-27].
- [69] PHAN, Anh Tuân. Human telomeric G-quadruplex: structures of DNA and RNA sequences. Online. *The FEBS journal*. 2010, roč. 277, č. 5, s. 1107-1117. ISSN 1742-464X. Dostupné z: <https://doi.org/10.1111/j.1742-4658.2009.07464.x>. [cit. 2024-04-27].
- [70] BIDZINSKA, Joanna; CIMINO-REALE, Graziella; ZAFFARONI, Nadia a FOLINI, Marco. G-quadruplex structures in the human genome as novel therapeutic targets. Online. *Molecules*. 2013, roč. 18, č. 10, s. 12368-12395. ISSN 1420-3049. Dostupné z: <https://doi.org/10.3390/molecules181012368>. [cit. 2024-04-27].
- [71] HERBERT, Alan a RICH, Alexander. The Biology of Left-handed Z-DNA. Online. *Journal of Biological Chemistry*. 1996, roč. 271, č. 20, s. 11595-11598. ISSN 00219258. Dostupné z: <https://doi.org/10.1074/jbc.271.20.11595>. [cit. 2024-05-14].

- [72] HERBERT, Alan. Z-DNA and Z-RNA in human disease. Online. *Communications Biology*. 2019, roč. 2, č. 1. ISSN 2399-3642. Dostupné z: <https://doi.org/10.1038/s42003-018-0237-x>. [cit. 2024-05-14].
- [73] ONEL, Buket; LIN, Clement a YANG, Danzhou. DNA G-quadruplex and its potential as anticancer drug target. Online. *Science China. Chemistry*. 2014, roč. 57, č. 12, s. 1605-1614. ISSN 1674-7291. Dostupné z: <https://doi.org/10.1007/s11426-014-5235-3>. [cit. 2024-04-27].
- [74] BISSLER, John J, Jiří ŠPONER, Thomas E. CHEATHAM, et al. DNA inverted repeats and human disease: structures of DNA and RNA sequences. *Frontiers in Bioscience* [online]. Elsevier, 1998, 2008, 3(4), d408-418 [cit. 2023-04-01]. *Advances in Virus Research*. ISBN 9780123743220. ISSN 10939946. Dostupné z: doi:10.2741/A284.
- [75] VOINEAGU, Irina; NARAYANAN, Vidhya; LOBACHEV, Kirill S a MIRKIN, Sergei M. Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. Online. *Proceedings of the National Academy of Sciences - PNAS*. 2008, roč. 105, č. 29, s. 9936-9941. ISSN 0027-8424. Dostupné z: <https://doi.org/10.1073/pnas.0804510105>. [cit. 2024-04-27].
- [76] BRÁZDA, Václav; KOLOMAZNÍK, Jan; LÝSEK, Jiří; HÁRONÍKOVÁ, Lucia; COUFAL, Jan et al. Palindrome analyser – A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. Online. *Biochemical and biophysical research communications*. 2016, roč. 478, č. 4, s. 1739-1745. ISSN 0006-291X. Dostupné z: <https://doi.org/10.1016/j.bbrc.2016.09.015>. [cit. 2024-04-27].
- [77] BRÁZDA, Václav; LAISTER, Rob C; JAGELSKÁ, Eva B a ARROWSMITH, Cheryl. Cruciform structures are a common DNA feature important for regulating biological processes. Online. *BMC molecular biology*. 2011, roč. 12, č. 1, s. 33-33. ISSN 1471-2199. Dostupné z: <https://doi.org/10.1186/1471-2199-12-33>. [cit. 2024-04-27].
- [78] FELSENFELD, G.; DAVIES, David R. a RICH, Alexander. FORMATION OF A THREE-STRANDED POLYNUCLEOTIDE MOLECULE. Online. *Journal of the American Chemical Society*. 1957, roč. 79, č. 8, s. 2023-2024. ISSN 0002-7863. Dostupné z: <https://doi.org/10.1021/ja01565a074>. [cit. 2024-05-14].
- [79] IRVING, Kelly L.; KING, Jessica J.; WALLER, Zoë A.E.; EVANS, Cameron W. a SMITH, Nicole M. Stability and context of intercalated motifs (i-motifs) for biological applications. Online. *Biochimie*. 2022, roč. 198, s. 33-47. ISSN 0300-9084. Dostupné z: <https://doi.org/10.1016/j.biochi.2022.03.001>. [cit. 2024-04-27].
- [80] DZATKO, Simon; KRAFCIKOVA, Michaela; HÄNSEL-HERTSCH, Robert; FESSL, Tomas; FIALA, Radovan et al. Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. Online. *Angewandte Chemie (International ed.)*. International ed. in English. 2018, roč. 57, č. 8, s. 2165-2169. ISSN 1433-7851. Dostupné z: <https://doi.org/10.1002/anie.201712284>. [cit. 2024-04-27].
- [81] OU, Tian-miao; LU, Yu-jing; TAN, Jia-heng; HUANG, Zhi-shu; WONG, Kwok-Yin et al. G-Quadruplexes: Targets in Anticancer Drug Design. Online. *ChemMedChem*. 2008, roč. 3, č. 5, s. 690-713. ISSN 1860-7179. Dostupné z: <https://doi.org/10.1002/cmdc.200700300>. [cit. 2024-05-14].
- [82] IRVING, Kelly L.; KING, Jessica J.; WALLER, Zoë A.E.; EVANS, Cameron W. a SMITH, Nicole M. Stability and context of intercalated motifs (i-motifs) for biological applications. Online. *Biochimie*. 2022, roč. 198, s. 33-47. ISSN 03009084. Dostupné z: <https://doi.org/10.1016/j.biochi.2022.03.001>. [cit. 2024-05-01].
- [83] BOHÁLOVÁ, Natália; CANTARA, Alessio; BARTAS, Martin; KAURA, Patrik; ŠŤASTNÝ, Jiří et al. Analyses of viral genomes for G-quadruplex forming sequences

- reveal their correlation with the type of infection. Online. *Biochimie*. 2021, roč. 186, s. 13-27. ISSN 0300-9084. Dostupné z: <https://doi.org/10.1016/j.biochi.2021.03.017>. [cit. 2024-04-27].
- [84] RUGGIERO, Emanuela a Sara N. RICHTER. Viral G-quadruplexes. Quadruplex Nucleic Acids As Targets For Medicinal Chemistry [online]. Elsevier, 2020, 101-131 [cit. 2022-06-23]. Annual Reports in Medicinal Chemistry. ISBN 9780128210178. Dostupné z: doi:10.1016/bs.armc.2020.04.001.
- [85] SONG, Jingwen; PERREAULT, Jean-pierre; TOPISIROVIC, Ivan a RICHARD, Stéphane. RNA G-quadruplexes and their potential regulatory roles in translation. Online. *Translation (Austin, Tex.)*. 2016, roč. 4, č. 2, s. e1244031-e1244031. ISSN 2169-074X. Dostupné z: <https://doi.org/10.1080/21690731.2016.1244031>. [cit. 2024-04-27].
- [86] HISCOX, Julian A. RNA viruses: hijacking the dynamic nucleolus. Online. *Nature reviews. Microbiology*. 2007, roč. 5, č. 2, s. 119-127. ISSN 1740-1526. Dostupné z: <https://doi.org/10.1038/nrmicro1597>. [cit. 2024-04-27].
- [87] BARR, J.N. a FEARNES, R. Genetic Instability of RNA Viruses. Online. *Genome Stability*. 2016, s. 21-35. ISBN 9780128033098. Dostupné z: <https://doi.org/10.1016/B978-0-12-803309-8.00002-1>. [cit. 2024-04-27].
- [88] JOHNSON, Margaret A.; CHATTERJEE, Amarnath; NEUMAN, Benjamin W. a WÜTHRICH, Kurt. SARS Coronavirus Unique Domain: Three-Domain Molecular Architecture in Solution and RNA Binding. Online. *Journal of molecular biology*. 2010, roč. 400, č. 4, s. 724-742. ISSN 0022-2836. Dostupné z: <https://doi.org/10.1016/j.jmb.2010.05.027>. [cit. 2024-04-27].
- [89] RHODES, Daniela a LIPPS, Hans J. G-quadruplexes and their regulatory roles in biology. Online. *Nucleic acids research*. 2015, roč. 43, č. 18, s. 8627-8637. ISSN 0305-1048. Dostupné z: <https://doi.org/10.1093/nar/gkv862>. [cit. 2024-04-27].
- [90] PERRONE, Rosalba; NADAI, Matteo; POE, Jerrod A.; FRASSON, Ilaria; PALUMBO, Manlio et al. Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 nef Coding Region: Implications for Antiviral Activity. Online. *PLoS ONE*. 2013, roč. 8, č. 8. ISSN 1932-6203. Dostupné z: <https://doi.org/10.1371/journal.pone.0073121>. [cit. 2024-05-01].
- [91] PAYNE, Susan. Introduction to RNA Viruses. Online. *Viruses*. 2017, s. 97-105. ISBN 9780128031094. Dostupné z: <https://doi.org/10.1016/B978-0-12-803109-4.00010-6>. [cit. 2024-04-27].
- [92] ZYNER, Katherine G; MULHEARN, Darcie S; ADHIKARI, Santosh; MARTÍNEZ CUESTA, Sergio; DI ANTONIO, Marco et al. Genetic interactions of G-quadruplexes in humans. Online. *ELife*. 2019, roč. 8. ISSN 2050-084X. Dostupné z: <https://doi.org/10.7554/elife.46793>. [cit. 2024-04-27].
- [93] HOWARD. Flaviviruses. Online. *Perspectives in Medical Virology*. 2005, roč. 11, s. 13-51. ISBN 0444506608. ISSN 0168-7069. Dostupné z: [https://doi.org/10.1016/S0168-7069\(05\)80003-5](https://doi.org/10.1016/S0168-7069(05)80003-5). [cit. 2024-04-27].
- [94] Simonsson, Tomas. "G-Quadruplex DNA Structures Variations on a Theme" *Biological Chemistry*, vol. 382, no. 4, 2001, pp. 621-628. <https://doi.org/10.1515/BC.2001.073>.
- [95] LAURETI, Mathilde; NARAYANAN, Divya; RODRIGUEZ-ANDRES, Julio; FAZAKERLEY, John K a KEDZIERSKI, Lukasz. Flavivirus Receptors: Diversity, Identity, and Cell Entry. Online. *Frontiers in immunology*. 2018, roč. 9, s. 2180-2180. ISSN 1664-3224. Dostupné z: <https://doi.org/10.3389/fimmu.2018.02180>. [cit. 2024-

- 04-27].
- [96] FLEMING, Aaron M.; DING, Yun; ALENKO, Anton a BURROWS, Cynthia J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. Online. *ACS Infectious Diseases*. 2016, roč. 2, č. 10, s. 674-681. ISSN 2373-8227. Dostupné z: <https://doi.org/10.1021/acsinfecdis.6b00109>. [cit. 2024-04-27].
- [97] BOWATER, Richard P; BOHÁLOVÁ, Natália a BRÁZDA, Václav. Interaction of Proteins with Inverted Repeats and Cruciform Structures in Nucleic Acids. Online. *International journal of molecular sciences*. 2022, roč. 23, č. 11, s. 6171. ISSN 1422-0067. Dostupné z: <https://doi.org/10.3390/ijms23116171>. [cit. 2024-04-27].
- [98] BIAN, Wen-xiu; XIE, Yan; WANG, Xiao-ning; XU, Guo-hua; FU, Bo-shi et al. Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. Online. *Nucleic acids research*. 2019, roč. 47, č. 1, s. 56-68. ISSN 0305-1048. Dostupné z: <https://doi.org/10.1093/nar/gky1177>. [cit. 2024-04-27].
- [99] DAY, Henry A.; PAVLOU, Pavlos a WALLER, Zoë A.E. I-Motif DNA: Structure, stability and targeting with ligands. Online. 2014, roč. 22, č. 16, s. 4407-4418. ISSN 09680896. Dostupné z: <https://doi.org/10.1016/j.bmc.2014.05.047>. [cit. 2024-05-14].
- [100] JAUBERT, Chloé; BEDRAT, Amina; BARTOLUCCI, Laura; DI PRIMO, Carmelo; VENTURA, Michel et al. RNA synthesis is modulated by G-quadruplex formation in Hepatitis C virus negative RNA strand. Online. *Scientific reports*. 2018, roč. 8, č. 1, s. 8120-9. ISSN 2045-2322. Dostupné z: <https://doi.org/10.1038/s41598-018-26582-3>. [cit. 2024-04-27].
- [101] ABOU ASSI, Hala; GARAVÍS, Miguel; GONZÁLEZ, Carlos a DAMHA, Masad J. I-Motif DNA: structural features and significance to cell biology. Online. *Nucleic Acids Research*. 2018, roč. 46, č. 16, s. 8038-8056. ISSN 0305-1048. Dostupné z: <https://doi.org/10.1093/nar/gky735>. [cit. 2024-05-14].
- [102] Gelderblom HR. Structure and Classification of Viruses. In: Baron S, editor. *Medical Microbiology*. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 41. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK8174/>. Online. [cit. 2024-04-27].
- [103] AMINA, BEDRAT. *G4-Hunter: a new algorithm for G-quadruplexes prediction's*. PhD Thesis. Université de Bordeaux, 2015.
- [104] BARTAS, Martin; ČUTOVÁ, Michaela; BRÁZDA, Václav; KAURA, Patrik; ŠŤASTNÝ, Jiří et al. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. Online. *Molecules*. 2019, roč. 24, č. 9. ISSN 1420-3049. Dostupné z: <https://doi.org/10.3390/molecules24091711>. [cit. 2024-04-27].
- [105] BRÁZDA, Václav; KOLOMAZNÍK, Jan; LÝSEK, Jiří; BARTAS, Martin; FOJTA, Miroslav et al. G4Hunter web application: a web server for G-quadruplex prediction. Online. *Bioinformatics*. 2019, roč. 35, č. 18, s. 3493-3495. ISSN 1367-4803. Dostupné z: <https://doi.org/10.1093/bioinformatics/btz087>. [cit. 2024-04-27].
- [106] *DNA analyser*. Online. Dostupné z: <https://bioinformatics.ibp.cz/#/help/quadruplex>. [cit. 2024-04-27].
- [107] LEE, Inhan; AJAY, Subramanian S; YOON, Jong In; KIM, Hyun Sil; HONG, Su Hyung et al. New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. Online. *Genome Research*. 2009, roč. 19, č. 7, s. 1175-1183. ISSN 1088-9051. Dostupné z: <https://doi.org/10.1101/gr.089367.108>. [cit. 2024-04-27].

- [108] SCHOBER, Patrick; BOER, Christa a SCHWARTE, Lothar A. Correlation Coefficients: Appropriate Use and Interpretation. Online. 2018, roč. 126, č. 5, s. 1763-1768. ISSN 0003-2999. Dostupné z: <https://doi.org/10.1213/ANE.0000000000002864>. [cit. 2024-04-27].
- [109] BRÁZDA, Václav; VALKOVÁ, Natália; DOBROVOLNÁ, Michaela a MERGNY, Jean-Louis. Abundance of G-Quadruplex Forming Sequences in the Hepatitis Delta Virus Genomes. Online. ACS Omega. ISSN 2470-1343. Dostupné z: <https://doi.org/10.1021/acsomega.3c09288>. [cit. 2024-05-17].
- [110] BARTAS, Martin; BRÁZDA, Václav; BOHÁLOVÁ, Natália; CANTARA, Alessio; VOLNÁ, Adriana 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. Online. Frontiers in Microbiology. 2020, roč. 11. ISSN 1664-302X. Dostupné z: <https://doi.org/10.3389/fmicb.2020.01583>. [cit. 2024-05-17].

7. POUŽITÉ ZTKRATKY

A – adenin

ADAR – adenosin deamináza působící na RNA

AIDS – akviziční syndrom získané imunodeficience

bp – párů bází

C – cytosin

cDNA – komplementární DNA

CDS – kódující sekvence

DNA – deoxyribonukleová kyselina

dsRNA – dvojitá deoxyribonukleová kyselina

G – guanin

G4 – G-kvadruplex

H – vodík

HIV – virus lidské imunodeficience

misc_feature – miscellaneous feature

MMTV – virus myšího mléčného nádoru

mRNA – mediátorová ribonukleová kyselina

NC-protein – nucleokapsid protein

NCBI – Národní centrum pro biotechnologické informace

nt – počtu nukleotidů

O – kyslík

PQS – sekvence s potenciálem tvořit G-kvadruplexy (potentially Quadruplex-forming Sequences)

KIT – tyrosin-proteinkináza (tyrosine-protein kinase)

KRAS – homolog virového onkogenu Kirstenova krysího sarkomu (Kirsten rat sarcoma viral oncogene homolog)

MYC – onkogen myelocytomatózy (myelocytomatosis oncogene)

kb – kilobáží

RNA – ribonukleová kyselina

rRNA – ribozomální ribonukleová kyselina

T – thymin

misc_difference – miscellaneous difference feature

misc_feature – miscellaneous feature

regulatory – regulační oblasti

3'UTR – nepřekládaná oblast na 3' konci

5'UTR – nepřekládaná oblast na 5' konci

repeat_region – repetitivní sekvence

rep_origin – počátek replikace

3'UTR – nepřekládaná oblast na 3' konci

5'UTR - nepřekládaná oblast na 5' konci

LTR – dlouhé terminální opakování (long terminal repeat)