



# 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

# SEKVENCE TVOŘÍCÍ G-KVADRUPLEXY V LIDSKÉM GENU A JEHO HOMOLOZÍCH PRO AMYLOID BETA PREKURSOR

SEQUENCES FORMING G-QUADRUPLEXES IN THE AMYLOID BETA PRECURSOR HUMAN GENE AND ITS HOMOLOGUES

## BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

## AUTOR PRÁCE

AUTHOR

Anna Stránská

## VEDOUCÍ PRÁCE

SUPERVISOR

doc. Mgr. Václav Brázda, Ph.D.

BRNO 2022

## Zadání bakalářské práce

Číslo práce: FCH-BAK1680/2021 Akademický rok: 2021/22  
Ústav: Ústav fyzikální a spotřební chemie  
Studentka: **Anna Stránská**  
Studijní program: Chemie a chemické technologie  
Studijní obor: Chemie pro medicínské aplikace  
Vedoucí práce: **doc. Mgr. Václav Brázda, Ph.D.**

### Název bakalářské práce:

Sekvence tvořící G–kvadruplexy v lidském genu a jeho homolozích pro amyloid beta prekursor

### Zadání bakalářské práce:

Získání sekvencí APP genu  
Analýza pro přítomnost G–kvadruplexových sekvencí (PQS)  
Srovnání lokalizace PQS  
Srovnání s přítomností v jiných organismech  
Otestování tvorby G–kvadruplexu ve vybrané sekvenci

### Termín odevzdání bakalářské práce: 27.5.2022:

Bakalářská práce se odevzdává v děkanem stanoveném počtu exemplářů na sekretariát ústavu. Toto zadání je součástí bakalářské práce.

-----  
Anna Stránská  
studentka

-----  
doc. Mgr. Václav Brázda, Ph.D.  
vedoucí práce

-----  
prof. Ing. Miloslav Pekař, CSc.  
vedoucí ústavu

V Brně dne 1.2.2022

-----  
prof. Ing. Michal Veselý, CSc.  
děkan

## ABSTRAKT

*APP* gen kóduje transmembránový protein amyloid beta prekursor, který je exprimován v mnoha typech buněk, včetně neuronů. Jeho funkce nebyly zatím zcela popsány, je ale zřejmé, že u něj probíhá štěpení před exportem do extracelulárního prostoru. Dochází k jeho degradaci po různých rozkladných drahách a homodimerizaci, při které mohou vzniknout částice s protektivní funkcí neuronů stejně tak jako fragmenty, které jsou toxické a způsobují smrt nervových buněk. Vznik škodlivých amyloidních beta plaků a jejich hromadění mezi neurony v mozku je úzce spjato se vznikem a postupem Alzheimerovy choroby, neurodegenerativního onemocnění mozku projevujícího se smrtí a ztrátou neuronů, která vede k nástupu demence, tedy ztráty kognitivních funkcí. V současné době existuje mnoho výzkumů, které se zabývají souvislostmi mezi neurodegenerativními onemocněními a výskytem G-kvadruplexů v genech, které jsou zapojeny do projevů nemocí. G-kvadruplexy jsou nekanonické sekundární struktury nukleových kyselin DNA a RNA, které vznikají v oblastech bohatých na guanin. Jsou významné především z hlediska jejich spojitosti s biologickými procesy jako je regulace genové exprese v genech a převážně v onkogenech, protože se vyskytují ve významných oblastech genu jako je promotor. Je možné je stabilizovat malými molekulami, a právě této schopnosti je využíváno ve výzkumu terapeutické léčby různých onemocnění. Ke zjištění důležitosti vyplývající z lokalizace a konzervovanosti G-kvadruplexů vyskytujících se v lidském *APP* genu, byla provedena bioinformatická analýza jak genu člověka, tak 346 dalších genových homologů. Pro tento účel byl použit program G4Hunter, který poskytl informace o nalezených sekvencích s potenciálem tvořit G-kvadruplex, jako je jejich umístění či G4Hunter skóre. Dále byla provedena *in vitro* analýza pomocí činidla thioflavin T, která otestovala schopnost nalezených sekvencí tvořit G-kvadruplexy za fyziologických podmínek. Výsledky potvrdily přítomnost a evoluční důležitost G-kvadruplexů nalezených v *APP* genu *Homo sapiens* a jejich schopnost skládat se v přítomnosti solí jako je sodík a draslík do kvadruplexních struktur.

## KLÍČOVÁ SLOVA

*APP* gen, G-kvadruplex, analýza genu, G4Hunter, genová homologie, lokalizace, konzervovanost

## ABSTRACT

The *APP* gene encodes the transmembrane protein amyloid beta precursor, which is expressed in many cell types, including neurons. Its functions have not yet been fully described, but it is clear that it is being cleaved before being exported to the extracellular space. It is degraded by various degradation pathways and also undergoes homodimerization, which can produce particles with protective neuronal function as well as fragments that are toxic and cause nerve cell death. The formation of harmful amyloid beta plaques and their accumulation among neurons in the brain is closely linked to the onset and progression of Alzheimer's disease, a neurodegenerative brain disease manifested by death and loss of neurons, which leads to dementia, i.e. loss of cognitive functions. There is currently a lot of research that deals with the links between neurodegenerative diseases and the occurrence of G-quadruplexes in genes that are involved in disease manifestations. G-quadruplexes are non-canonical DNA and RNA nucleic acid secondary structures that arise in guanine-rich regions. They are important mainly in terms of their connection with biological processes such as the regulation of gene expression in genes and mainly in oncogenes because they occur in important regions of the gene such as the promoter. It is possible to stabilize them with small molecules, and it is this ability that is used in research into the therapeutic treatment of various diseases. Bioinformatics analysis of both the human gene and 346 other gene homologs was performed to determine the importance of G-quadruplexes localization and conservation in the human *APP* gene. For this purpose, the G4Hunter program was used, which provided information about the found sequences with the potential to form a G-quadruplex, such as their location or G4Hunter score. *In vitro* analysis was performed using thioflavin T reagent, which tested the ability of the found sequences to form G-quadruplexes under physiological conditions. The results confirmed the presence and evolutionary importance of G-quadruplexes found in the *APP* gene of *Homo sapiens* and their ability to assemble into quadruplex structures in the presence of salts such as sodium and potassium.

## KEYWORDS

*APP* gene, G-quadruplex, gene analysis, G4Hunter, gene homology, localization, conservation

STRÁNSKÁ, Anna. *Sekvence tvořící G-kvadruplexy v lidském genu a jeho homolozích pro amyloid beta prekursor*. Brno, 2022. Dostupné také z: <https://www.vutbr.cz/studenti/zav-prace/detail/139032>.  
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 použita ke komerčním účelům jen se souhlasem vedoucího bakalářské práce a děkana FCH VUT.

.....

podpis studentky

## PODĚKOVÁNÍ

Mé poděkování patří panu doc. Mgr. Václavu Brázdovi, Ph.D. za vedení mé bakalářské práce a cenné rady. Obrovské díky věnuji také své konzultantce Mgr. Natálii Bohálové za vždy ochotný a vstřícný přístup při sběru dat a vyhodnocování výsledků, za její čas a pomoc při kompletaci práce.

# OBSAH

1	Úvod.....	7
2	Teoretická část .....	8
2.1	Obecně o nukleových kyselinách .....	8
2.1.1	Sekundární struktury DNA.....	8
2.2	G-kvadruplexy .....	9
2.2.1	Struktura.....	9
2.2.2	Topologie .....	10
2.2.3	Stabilita .....	11
2.2.4	Výskyt a terapeutický potenciál .....	11
2.2.5	Interakce G4 s ligandy.....	12
2.3	Analýza G4 <i>in silico</i> .....	13
2.3.1	G4Hunter.....	13
2.4	Analýza G4 konkrétního genu .....	15
2.4.1	Gen pro amyloid beta prekursor.....	15
2.4.2	Alzheimerova choroba – současné přístupy léčby a budoucnost .....	16
3	Experimentální část.....	19
3.1	Materiály pro <i>in silico</i> analýzu .....	19
3.2	Získávání dat a jejich následná analýza v G4Hunter .....	19
3.3	Důkaz tvorby G4 <i>in vitro</i> .....	20
3.3.1	ThT analýza.....	20
3.3.2	Materiály pro <i>in vitro</i> analýzu .....	21
3.3.3	Oligonukleotidy.....	21
3.3.4	Chemikálie .....	22
3.3.5	Přístroje .....	22
3.3.6	Příprava roztoků .....	22
3.3.7	Příprava oligonukleotidů a měření .....	22
4	Výsledky a diskuze .....	24
4.1	Analýza G4Hunter <i>in silico</i> .....	24
4.1.1	Fylogenetický strom.....	24
4.1.2	Konzervovanost PQS v oblastech genu.....	28
4.1.3	Konzervovanost sekvencí z <i>APP</i> genu <i>Homo sapiens</i> .....	29
4.1.4	Výskyt a frekvence PQS u konkrétních homologů .....	31
4.2	Analýza G4 pomocí ThT <i>in vitro</i> .....	34
5	Závěr .....	40
6	Seznam zkratk .....	42
7	Seznam použité literatury .....	45

# 1 ÚVOD

Gen pro amyloid beta prekursor (*APP*) kóduje buněčný povrchový receptor a transmembránový prekursorový protein, který je enzymy sekretázy štěpen za vzniku několika peptidů. Tyto peptidy mají různou funkci a jednou z nich je tvorba amyloidních plaků, které je možné nalézt v mozku pacientů s Alzheimerovou chorobou (AD) [1]. AD je fyzické progresivní onemocnění, které postihuje mozkové buňky, jež mezi sebou ztrácejí schopnost komunikace, až nakonec odumírají [2]. V současné době neexistuje lék na AD, ale pacientům lze poskytnout mnoho podpůrných prostředků zlepšujících kvalitu života. V rámci drah neurologických poruch se čím dál více hovoří o jejich možné léčbě spojené s využitím DNA struktur G-kvadruplexů (G4). Konkrétně se jedná o nemoci jako syndrom fragilního X, Parkinsonova choroba, frontotemporální demence a amyotrofická laterální skleróza (ALS) [3].

G4 jsou sekundární struktury nukleových kyselin tvořící se v oblastech s velkou koncentrací purinové báze guanin [4]. Mohou se skládat do polymorfních struktur a ve své centrální dutině většinou vážou jednomocné i dvojmocné kationty, které strukturu stabilizují [5]. G4 přijímají širokou škálu různých topologií dle kombinací směru vlákna, délky a složení smyčky a vyskytují se ve všech doménách života. Tvorba G4 byla zkoumána v eukaryotických telomerách, ale i v netelomerické genomové DNA. Mimo jiné byla věnována pozornost G4 vyskytujících se v promotorech onkogenů [6] a také v dalších oblastech, které regulují některé z významných biologických dějů [7] jako je transkripce, zpracování mRNA, translace, replikace DNA a funkce telomer a mitochondrií [8]. Z jejich vlastností vyplynulo jejich možné využití jako terapeutických cílů v léčbě různých onemocnění. Vzrostl také zájem o studium proteinů a malých molekul, které se mohou s vysokou afinitou a selektivitou vázat na G4 a tím je stabilizovat [9].

Kvůli důležitosti G4 byly v uplynulých letech vyvinuty různé biofyzikální metody, které *in vitro* identifikují G4 a charakterizují jejich skládání do různých topologií. Tyto poznatky otevřely cestu predikcím založeným na výpočetní technice [10]. Jedná se o algoritmy, které na základě zvolených kritérií hledají sekvence se sklonem tvořit G4 v podmínkách živých soustav [11]. Mezi takové programy patří i G4Hunter, nástroj pro predikci G4, který bere v úvahu G-bohatost sekvencí nukleových kyselin. Výsledky uvádí jako skóre sekvence vyjadřující její náchylnost k tvorbě G4 [12].

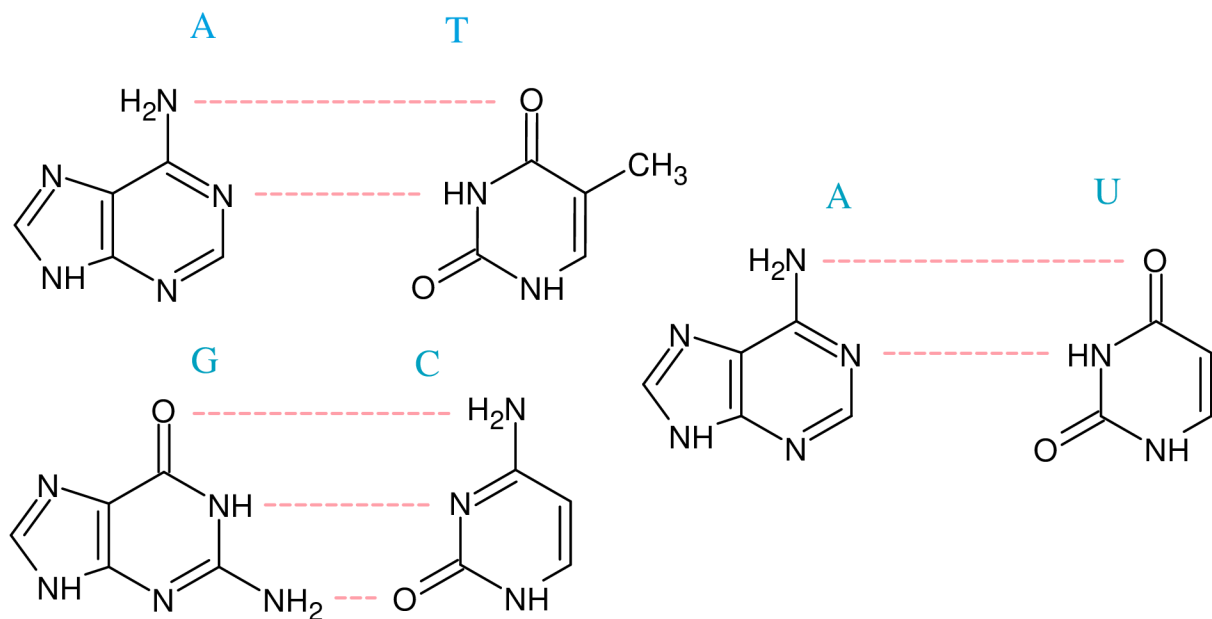
V této bakalářské práci byla provedena *in silico* analýza *APP* genu člověka z pohledu výskytu a lokalizace G4. Pro pochopení významnosti nalezených sekvencí byly analýze podrobeny i *APP* geny 346 dalších genových homologů *Homo sapiens*. Byla porovnána frekvence výskytu PQS u homologů a v genu člověka. Dále byl sestaven fylogenetický strom, vytvořený na základě frekvence výskytu PQS, který byl porovnán s fylogenetickým stromem založeným na taxonomických jednotkách. Byla provedena *in silico* analýza konzervovanosti 10 PQS sekvencí *APP* genu *Homo sapiens* s nejvyšším G4Hunter skóre. Nakonec byly vybrány zajímavé sekvence z *APP* genu člověka s potenciálem tvořit G4. Ty byly následně podrobeny *in vitro* analýze pomocí thioflavinu T (ThT), při níž byla u těchto sekvencí potvrzena tvorba G4. Výsledky by mohly odhalit významnost nalezených G4 v *APP* genu člověka a zjistit tím možné pozice, které by se daly využít v budoucí genové léčbě AD.

## 2 TEORETICKÁ ČÁST

### 2.1 Obecně o nukleových kyselinách

Nukleové kyseliny, deoxyribonukleová kyseliny DNA a ribonukleová kyselina RNA, jsou nositelkami genetické informace. Struktura dvoušroubovité DNA umožňuje tuto informaci kopírovat, předávat dalším generacím a tím tvořit základ pro veškerý život [13].

Když Watson a Crick roku 1953 popsali 3D strukturu DNA [14], zdůraznili její dva hlavní rysy: komplementaritu základních sekvencí na dvou vláknech a přirozenou povahu tvořit polymer tvaru dvoušroubovice. Byla zjištěna komplementarita bází adeninu k thyminu a guaninu k cytosinu [15]. V případě RNA se adenin komplementárně váže k uracilu [16].



Obrázek 1: Kanonické párování bází pro DNA i RNA, v případě DNA (nalevo) se A (adenin) páruje s T (thyminem), G (guanin) s C (cytosinem), v případě RNA (napravo) se A (adenin) páruje s U (uracilem). Přepracováno podle [17].

#### 2.1.1 Sekundární struktury DNA

DNA se kromě kanonické dvoušroubovice může složit i do dalších struktur. Dlouho se předpokládalo, že tyto struktury mohou existovat pouze *in vitro*, avšak výsledky *in vivo* analýz ukazují, že se vyskytují i v živých organismech [18].

Dvojřetězcová pravotočivá dsDNA neboli B-DNA je tvořena dvěma polynukleotidovými řetězci, které jsou svinuty do šroubovice. Mezi bázemi obou vláken, které jsou orientovány dovnitř šroubovice, se tvoří vodíkové můstky podle principu párování komplementárních bází s tím, že řetězce jsou uspořádány antiparalelně [14].

Mezi další sekundární struktury kromě klasické dvoušroubovice patří levotočivá šroubovice Z-DNA [19], která uvolňuje torzní napětí vyvolané transkripcí [20]. Dále lze zmínit například



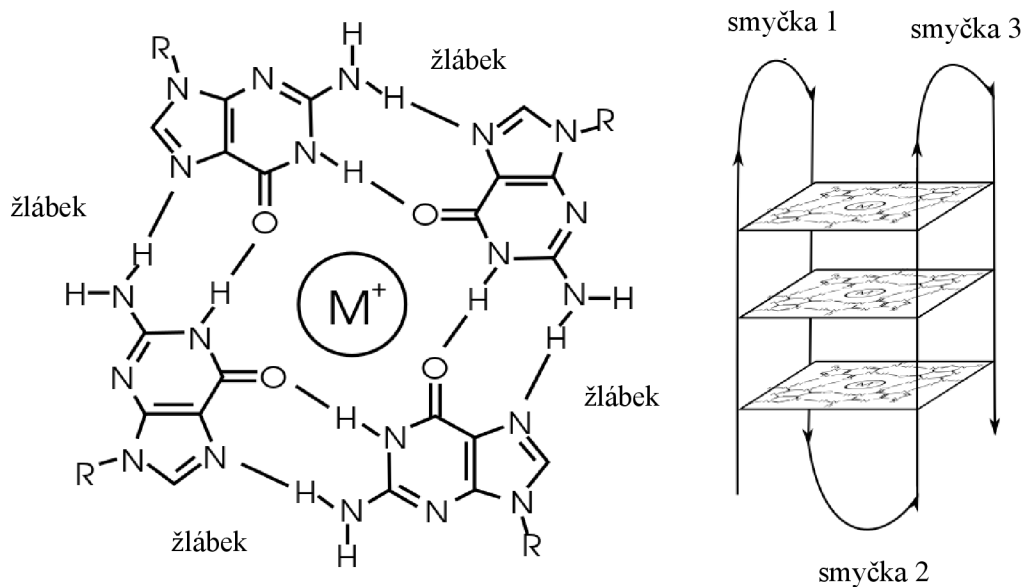
čtyřramenné křížové struktury, které se nacházejí v blízkosti míst počátku replikace a promotorů v různých organismech [21]. Další sekundární strukturou je triplexní DNA, která vzniká v případě vytvoření Hoogsteenovy vodíkové vazby mezi jednořetězcovou DNA a duplexní B-DNA bohaté na purin [22]. Tyto struktury jsou běžné u eukaryot, u prokaryot jsou mnohem vzácnější [23]. U savců se triplexotvorné motivy vyskytují v intronech různých základních genů [24].

## 2.2 G-kvadruplexy

### 2.2.1 Struktura

G-kvadruplexy jsou vysoce polymorfni intra- nebo intermolekulární struktury nukleových kyselin, které se mohou tvořit ve specifických sekvencích DNA nebo RNA bohatých na purinovou bázi guanin (G) [4]. Intramolekulární G4 se tvoří v rámci jednoho vlákna nukleové kyseliny [25], jejich tvorbu řídí délka smyček a stabilní uspořádání zaujímají zejména v přítomnosti draselného kationu [26]. Intermolekulární G4 se formují v rámci několika vláken v prostředí o vyšší iontové síle, kde jsou G patra oddělena jen jedním či dvěma nukleotidy [27].

Strukturu G4 tvoří čtyři guaninové báze vázáné na sebe vodíkovými můstky, které tvoří Hoogsteenovu vazbu [28]. Takto vzniklá stabilní planární uspořádání se nazývají G-kvartety, jejichž stohy jsou poté stabilizovány jednomocným kationtem jako je  $\text{Na}^+$  nebo  $\text{K}^+$  [28]. Zmíněné kationty obsazují centrální dutiny mezi vrstvami a neutralizují elektrostatické odpuzování dovnitř směřujících guaninových kyslíků [29]. Ve struktuře se mohou objevit i dvojmocné kationty, např.  $\text{Ca}^{2+}$  nebo  $\text{Mg}^{2+}$  [30].



Obrázek 2: Struktura G4, nalevo: G-kvartet složený z guaninových bází,  $M^+$  představuje jednomocný kation, dále jsou vyznačené žlábků zmiňované v textu, napravo: intramolekulární G4 komplex se třemi smyčkami, upraveno podle [31].

## 2.2.2 Topologie

G4 jsou většinou tvořeny jedním, dvěma nebo čtyřmi samostatnými řetězci DNA či RNA a mohou vykazovat širokou škálu topologií, což je způsobeno různými kombinacemi směru řetězců a variacemi ve velikosti smyčky a sekvence [32].

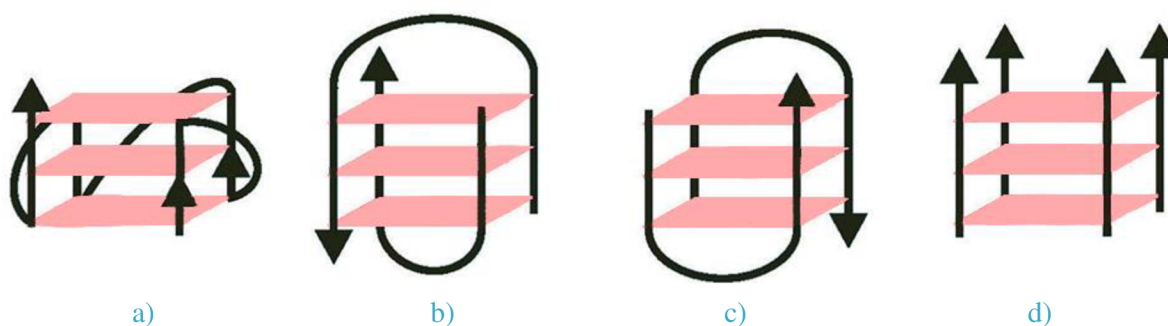
Obecně lze unimolekulární (tj. intramolekulární) sekvenci G4 definovat jako:



kde  $m$  je počet G v každém G-traktu a  $X_{n,o,p}$  jsou smyčky, které může tvořit jakákoliv báze včetně guaninu. G-trakty nemusí být vždy stejně dlouhé, což platí u většiny netelomerické genové sekvence organismů. Bimolekulární a tetramolekulární G4 mohou být složeny z asociace různě dlouhých sekvencí [32]. Nedávnými studiemi bylo zjištěno, že G-trakty s pěti až šesti opakováními mají schopnost tvořit stabilní intramolekulární interakce za fyziologických podmínek, což naznačuje jejich biologickou roli [33].

Struktura G4 může být klasifikována podle polaritě vláken a umístění smyček, které spojují guaninová vlákna. Antiparalelní jsou takové, které mají alespoň jedno z vláken v antiparalelní pozici k ostatním. Tento typ se nachází ve většině uni- a bimolekulárních G4. Dále lze rozlišit paralelní a hybridní G4. Všechny paralelní G4 mají guaninové glykosidické úhly v *anti* konformaci. Antiparalelní G4 mají *syn* i *anti* konformaci. Dále mají všechny G4 čtyři žlábký, což jsou dutiny ohraničené fosfodiesterovými páteřemi [32].

Unimolekulární a bimolekulární G4 jsou spojeny smyčkami, které jsou převážně tvořeny thyminovými bázemi. Lze rozlišit smyčky diagonální, laterální a externí. Diagonální a laterální smyčky spojují dva antiparalelní protilehlé řetězce s tím, že diagonální smyčka přechází přes oblast centrálního kanálu. Laterální smyčka tvoří strukturu známou jako vlásenka. Externí smyčka spojuje dva paralelní řetězce přes boční část molekuly v oblasti žlábků [34].



Obrázek 3: Schématické znázornění některých topologií G4: a) unimolekulární paralelní se třemi externími smyčkami, b) bimolekulární antiparalelní se dvěma diagonálními smyčkami, c) bimolekulární antiparalelní se dvěma laterálními smyčkami, d) lineární paralelní, upraveno podle [34].

### 2.2.3 Stabilita

Stabilita G4 je jednou z jejich klíčových vlastností. G4 obsahují na jedné straně neobyčejně stabilní a rigidní guaninové jádro a centrální kanál s ionty, na straně druhé velmi flexibilní smyčky, které mohou mít různá geometrická uspořádání a ovlivňují tak interakce G4 s okolím. Vliv na jejich stabilitu má množství faktorů, mezi které patří např. délka a typ smyček, počet tetrad či vzájemná orientace řetězců. Lze říct, že jsou stabilní za fyziologických podmínek a teplota tání převyšuje 37 °C. Na stabilitu má vliv i přítomnost centrálního iontu, jehož cílenou volbou lze zřejmě ovlivnit topologii G4 [34].

Vliv na stabilitu G4 mohou mít specificky interagující proteiny, které se dělí dle jejich funkce do několika odlišných skupin. Proteiny první skupiny po navázání zvyšují stabilitu G4. Mezi významné zástupce této třídy patří mutantní protein p53, který se přednostně a samovolně váže na oblast počátku transkripce bohatou na G či C. Často je asociován s RNA polymerázu II [35], což je enzym katalyzující syntézu hnRNA a malých RNA v eukaryotické buňce [36]. Proteiny druhé skupiny přes vazbu na DNA indukují tvorbu G4 struktury. Příkladem této skupiny proteinů je lidský nukleolin, což je multifunkční fosfoprotein. Ten hraje roli například při dekonenzaci chromatinu, regulaci transkripce, biogenezi ribozomů, buněčné proliferaci a apoptóze [36]. Poslední zmíněné jsou proteiny třetí skupiny, které destabilizují kvadruplexní struktury. Patří sem enzymy helicázy odvíjející DNA. Bylo zjištěno, že některé z nich přednostně rozvíjejí G4 struktury [36].

### 2.2.4 Výskyt a terapeutický potenciál

V lidském genomu se G4 nacházejí například i v telomerické DNA. Telomery jsou lokalizovány na konci chromozomů. Skládají se z opakujících se sekvencí DNA a proteinů a chrání chromozomy před degradací [37]. U lidí se jedná o repetice TTAGGG, které se nacházejí na konci telomer a s věkem se postupně zkracují [38]. Naproti tomu telomery zárodečných a rakovinových buněk zkracování nepodléhají z důvodu jejich udržování enzymem telomerázou [39]. Aktivita telomerázy byla nalezena u 85 až 90 % všech lidských nádorů, kdežto v normálních buňkách nalezena nebyla. Předpokládá se proto, že telomeráza může hrát roli v nekontrolovaném růstu nádorových buněk [40]. Bylo objeveno velké množství sloučenin, které stabilizují G4 a inhibují právě zmíněný enzym [41].

G4 mohou být také často lokalizovány v promotorových oblastech onkogenů. Ty mohou způsobit vznik zhoubného bujení [42], kde G4 fungují jako regulátory transkripce. Konkrétně bylo dokázáno, že se vyskytují v lidských genech jako je *c-Myc*, *VEGF*, *KRAS*, *c-Kit* nebo *c-Myb* [43]. Lokalizace PQS je mnohdy na hranici mezi introny a exony, což jsou nekódující a kódující oblasti DNA. Vyskytují se také na 5' a 3' nepřekládaných koncích mRNA, což jsou oblasti, které se sice transkribují do RNA, ale nepřekládají se do pořadí aminokyselin při translaci. V těchto částech by G4 mohly hrát úlohu při sestřihu pre-mRNA nebo regulaci translace [44].

Zmíněné poznatky naznačují, že obecně by struktury G4 mohly zaujímat hlavní roli v řízení řady buněčných procesů, včetně udržování telomer, replikace, transkripce a translace. Vysoce polymorfní povaha G4 konformací umožňuje navázání malých molekul, které pravděpodobně mají schopnost selektivně rozpoznávat a stabilizovat tuto strukturu [45]. Ve srovnání onkogenů a nádorových supresorů, tedy normálních genů, které zpomalují buněčné dělení [46], je tvorba G4 u onkogenů významně vyšší. Proto jsou G4 považovány za perspektivní cíl pro onkogenní léčbu [47].

G4 struktury se nacházejí i v RNA a předpokládá se, že ovlivňují různé aspekty biologie zahrnující transkripci, sestřih, translaci a poločas rozpadu [48]. Kromě terapeutického potenciálu existuje značný zájem o použití G4 pro biomateriály, biosenzory a biokatalyzátory [43].

G4 ovlivňují transkripci pozitivně i negativně. Povaha závisí na kontextu a na vlákně, na kterém se nacházejí [49]. Jeden ze způsobů regulace může být využití G4 jako fyzické překážky pro RNA polymerázu, kde vznik a stabilizace sekvence brání postupu enzymu. Tím způsobí pozastavení jeho činnosti nebo rozpad celého transkripčního komplexu [50]. Pokud se G4 se zmíněnou funkcí vytvoří na netemplátovém, tedy kódujícím vlákně, zamezuje tím opakovanému zdvojení řetězce udržováním ssDNA (jednovláknová DNA), čímž je umožněna hybridizace vznikajícího RNA vlákna s templátovou DNA. Vznikající RNA transkript může vytvořit RNA G4 nebo hybridní DNA-RNA G4 [51]. Všechny jmenované pozice PQS a jejich vzájemné interakce mohou být důležitou součástí regulace transkripce, sestřihu pre-mRNA či translace [44].

Bližší analýza lidských genů, které obsahují G4 lokalizované v oblasti počátků transkripce ukázala, že ve většině genů se alespoň jeden G4 vyskytuje v rozmezí 500 nukleotidů za počátkem transkripce na kódujícím vlákně, což poukazuje na významně zvýšenou expresi genů. Předpokladem je, že G4 usnadňuje transkripci mechanismem udržování ssDNA na templátovém vlákně [52].

### 2.2.5 Interakce G4 s ligandy

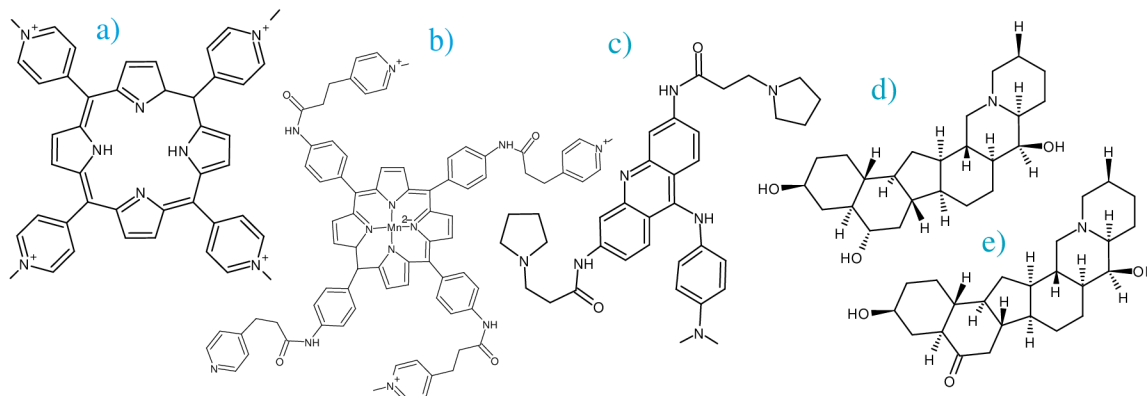
Jak bylo výše zmíněno, G4 působí jako důležité regulační prvky v DNA i RNA. Z toho důvodu jsou studovány za účelem možného využití jejich interakce s malými molekulami v terapeutických aplikacích. Doposud bylo pomocí výpočetních a chemicko-biologických přístupů objasněno několik způsobů interakce komplexů G4-ligand.

Stabilizací pomocí jednomocných kationtů byly inspirovány návrhy malých molekul, které by se zaměřovaly na jedinečnou strukturu G4. Aby byly účinné, musí mít vysokou afinitu ke G4 a vysokou selektivitu k dsDNA. Mezi místa, kde nejvíce dochází k interakcím, patří terminální a střední G-kvartety, se kterými reagují aromatické planární molekuly, dále žlábký, smyčkový, páteřový a centrální kanál. Se žlábkými, smyčkovými a páteřovými reagují nejvíce aminoskupiny ligandů [53].

První důkaz popsala studie z roku 1997, která prokázala, že stabilizace struktury G4 *in vitro* pomocí 2,6-diamidoantrachinonu vedla k inhibici aktivity telomerázy [54]. Od té doby bylo navrženo velké množství molekul specificky interagujících s G4 DNA. K identifikaci selektivnějších sloučenin cílících právě na G4 jsou v dnešní době využívány vysoce výkonné *in silico* metody modelování molekul [55].

Konkrétním příkladem aromatického planárního ligandu je TMPyP4, který inhibuje telomerázovou aktivitu a reguluje expresi onkogenů, jako je například *c-Myc*. Ačkoliv se však vyznačuje vysokou afinitou ke G4, v otázce preference G4 oproti dsDNA není vhodným selektivním činidlem [56]. K překonání tohoto omezení byl navržen porfyrin Mn<sup>III</sup>, který vykazuje jak vysokou afinitu, tak selektivitu k G4 [57]. K dalším studovaným ligandům, které reagují s koncovými G-kvartety, patří například akridinový derivát BRACO-19 [58].

Mezi žlábký G4 a dsDNA jsou významné rozdíly ve velikosti a tvaru, proto jsou vhodné k interakci s ligandy, které vůči G4-žlábkům vykazují selektivitu. Jako příklad je možné uvést peimin a peiminin, které mají neplanární aromatickou strukturu a vykazují vysokou afinitu a selektivitu ke G4-žlábkům [59].



Obrázek 4: G4-ligandy: a) TMPyP4, b) porfyrin  $Mn^{III}$ , c) BRACO-19, d) peimin, e) peiminin, upraveno podle [60].

## 2.3 Analýza G4 *in silico*

V posledních letech byly vyvinuty různé algoritmy k předpovědím potenciální tvorby G4 přímo z primární sekvence DNA či RNA. Ačkoliv se současné studie zaměřují primárně na unimolekulární záhyby G4, nedávno byly popsány intermolekulární struktury, které se pravděpodobně podílejí na klíčových buněčných funkcích [61].

Existuje celá řada popsaných experimentálních technik, které lze použít k ověření schopnosti tvořit G4 specifické sekvence. Patří mezi ně metody poskytující informace o struktuře, např. NMR [62], rentgenová krystalografie [63] nebo měření spekter cirkulárního dichroismu [64] a také metody, díky nimž lze ověřit tepelnou stabilitu G4 [65]. Žádná z těchto metod ovšem není vhodná pro sekvenování a identifikaci nových G4 na úrovni genomu, proto je nezbytné použít určité formy prediktivního algoritmu, který vyhledává dané struktury v měřítku celého genomu [61].

Dosud se ke charakterizaci struktur G4 používalo přístupu kombinujícího předpovědi *in silico* se skládáním *in vitro*. První algoritmy pro detekci *in silico* byly založeny na predikcích podložených biochemickými a biofyzikálními experimenty. Zpočátku byly používány algoritmy zakládající se na shodě regulárních výrazů, které byly dále vylepšeny o funkce používající k predikci výpočet G4Hunter skóre (bodování). Ty lze dále kombinovat s algoritmy posuvného okna či přístupy strojového učení [61].

### 2.3.1 G4Hunter

Jedná se o výkonný a široce používaný nástroj pro predikci náchylnosti DNA (RNA) tvořit G4 [66]. Při analýze je možné importovat sekvence pomocí NCBI ID, přímo nahrát FASTA nebo .txt soubor či vložit sekvenci ze schránky [12]. Dále je nutné nastavit velikost okna a prahovou hodnotu G4Hunter skóre. Velikost okna se může pohybovat v rozmezí 10 až 100 nukleotidů a doporučuje se nastavit

na hodnotu 25. Prahová hodnota se pohybuje v intervalu od 0,8 do 4,0 a je doporučeno použít hodnotu 1,4 [12].

Algoritmus programu počítá s oblastmi, které jsou bohaté na G a podává informaci o náchylnosti sekvence tvořit G4. Při počítání G4Hunter skóre vyhledává program sekvence tvořící řadu guaninů nebo cytosinů a sečte hodnoty, které dává každá skupina těchto bází. G bázím je přiřazena kladná hodnota a C bázím záporná. Ostatní báze (T, A, U) mají hodnotu nula. V případě více stejných G nebo C bází, které jdou bezprostředně za sebou, se hodnota zvyšuje či snižuje až do hodnoty 4 či -4. Takto vypočtená suma se vydělí hodnotou velikostí okna [66].

Příklady výpočtu G4Hunter skóre jsou následující:

$$CCC \rightarrow (-3-3-3)/3 = 3$$

$$GGGGGG \rightarrow (4+4+4+4+4+4)/6 = 4$$

$$ATCCCAAGGGAA \rightarrow (0+0-3-3-3+0+0+3+3+3+0+0)/10 = 0,625$$

Některé výsledky mohou vést k analýze sekvencí delších, než je zadaná velikost okna a nižším výsledkům G4Hunter skóre, než je zadaná prahová hodnota. Děje se to z důvodu automatického seskupování shodných částí. Pokud program najde dvě nebo více sekvencí s G4Hunter skóre, které je vyšší než prahová hodnota, seskupí tyto sekvence dohromady a analyzuje je jako jednu [66]. Pro příklad byla zadána velikost okna 25 a prahová hodnota 1:

Tabulka 1: Příklad výpočtu a seskupení sekvencí, převzato z [66], upraveno.

pozice	sekvence	G4Hunter skóre
349	GATCCACTTGGCTACATCCGCCCCC	-1
350	ATCCACTTGGCTACATCCGCCCCCT	-1,04
351	TCCACTTGGCTACATCCGCCCCCTT	-1,04
352	CCACTTGGCTACATCCGCCCCCTTA	-1,04
výsledná sekvence, délka 28 bp	GATCCACTTGGCTACATCCGCCCCCTTA	-0,893

Výsledky z analýzy se zobrazí jako teplotní mapa a seznam výpočtů G4Hunter skóre analyzovaných sekvencí. Teplotní mapa zobrazuje počet nalezených PQS v definovaných částech. Mapu lze zobrazit ve dvou režimech – jako absolutní počet nalezených PQS nebo procentuální pokrytí analyzované sekvence oblastmi PQS. Seznam výsledků obsahuje pozici segmentu, kde byl G4 nalezen, jeho délku a G4Hunter skóre [66]. Primárně jsou zobrazovány agregované výsledky, ale lze zobrazit i oddělené výsledky o délce velikosti okna. Aplikace poskytuje tři formáty výsledků. Prvním z nich je bedgraph soubor, což je standardizovaný formát souborů používaný v genomických zobrazovacích programech,

jako je IGV (integrační genomový prohlížeč) či Genome Browser. Dalšími dvěma typy formátů jsou agregované a neagregované sekvence v souborech CSV. Tento formát obsahuje v řádcích informace o analyzovaných sekvencích, a to pozici, délku nejdelší souvislé sekvence, G4Hunter skóre nad prahovou hodnotou a sekvenci vypsanou v pořadí bází. SUB\_SCORE zobrazuje G4Hunter skóre pro jednotlivé pozice uvnitř agregované sekvence [12].

## 2.4 Analýza G4 konkrétního genu

### 2.4.1 Gen pro amyloid beta prekursor

*APP* gen se nachází na chromozomu 21 a má rozpětí 290 586 bp genomové DNA. Bývá exprimován v široké škále lidských tkání, nejvíce v lidském mozku. Protein amyloid beta prekursor, který tento gen kóduje, je transmembránový protein typu I. Je štěpen enzymem  $\alpha$ -sekretázou za vzniku sekretovaného N-koncového rozpustného APP $\alpha$  (sAPP $\alpha$ ) a C-koncového fragmentu o délce 83 aminokyselin [67]. Stejně tak ale může být štěpen  $\beta$ -sekretázou za vzniku rozpustných fragmentů, které mohou být toxické a zapříčinit smrt neuronů [68].

APP protein má tři převládající varianty: APP695, APP751 a APP770. Majoritní izoformou je APP695, která je exprimována v neuronech, zatímco varianta APP751 bývá detekována zejména v astrocytech [69]. Protein je proteolyticky štěpen enzymy sekretázy a vzniklé fragmenty jsou sekretovány do extracelulárního prostoru. Mezi nimi jsou i neurotoxické A $\beta$  peptidy, sAPP $\beta$  a CTF (C-terminální fragmenty) [70].

Studiemi bylo zjištěno, že mutace v APP a dalších (PSEN1 a PSEN2) způsobují časný nástup dědičné Alzheimerovy choroby (AD), tj. ve věku do šedesáti let [70]. Nejběžnější mutace *APP* mění jeden ze stavebních bloků bílkoviny. Tato mutace nahrazuje aminokyselinu valin za izoleucin na pozici 717. Mutace v genu mohou vést ke zvýšenému množství amyloidních  $\beta$  peptidů nebo o něco delší a mazlavější formy peptidu, které se mohou hromadit v mozku a vytvářet shluky zvané amyloidní plaky. Hromadění toxického amyloidního  $\beta$  peptidu a vzniklých plaků může vést ke smrti neuronů a progresivním příznakům a symptomům AD [71].

Analýzou *in silico* bylo zjištěno, že v lidském *APP* genu se nachází sekvence se šesti opakováními G-traktů, které jsou odděleny jedním nukleotidem (jiným než G). Obecné schéma této sekvence je G<sub>6</sub>H<sub>1</sub>G<sub>6</sub>H<sub>1</sub>G<sub>6</sub>H<sub>1</sub>G<sub>6</sub>. Zmíněný G4 se nachází na pozici 26 016 931 bp [33]. Na základě bioinformatických analýz lokalizace zmíněných sekvencí v genomu se předpokládá, že pravděpodobně regulují genovou expresi a tím biologické procesy. Proto byly provedeny biofyzikální experimenty se sekvencemi G5 a G6 *in vitro* a bylo potvrzeno, že tvoří stabilní intramolekulární interakce za fyziologických podmínek, což naznačuje jejich biologickou roli a jejich možné využití jako terapeutických cílů [33].

*APP* gen se, stejně jako jiné geny, skládá z několika funkčních složek, z nichž každá je zapojena do jiné části genové exprese. Při genové expresi dochází k přepisu neboli transkripci DNA genu do mRNA, která je následně při translaci přeložena do pořadí aminokyselin, ze kterých se tvoří proteiny. Z obecné roviny existují dvě hlavní funkční jednotky: regulační promotorová a přepisovaná oblast. Promotorová oblast řídí, kdy a v jaké tkáni je gen exprimován pomocí speciálních elementů, nukleotidových sekvencí. Struktury označující se jako *cis*- a *trans*- na sebe vážou proteinové faktory, které jsou zodpovědné

za iniciaci tvorby mRNA a transkripci genu. U eukaryot jsou přepisující se oblasti genů většinou nespojitě a střídají se exony a introny. Funkce intronů zatím nebyla zcela objasněna, ale jelikož jsou v rámci evoluce zachovány, předpokládá se jejich specifický účel [72].

## 2.4.2 Alzheimerova choroba – současné přístupy léčby a budoucnost

V současné době neexistuje lék proti demenci způsobené AD hlavně kvůli tomu, že projevy klinických příznaků se objeví až desítky let po rozsáhlých a zřejmě nevratných změnách, které v organismu probíhají na úrovních molekulových, buněčných či systémových. Moderní léčba se zaměřuje na potlačování symptomů, zmírnění zátěže nemoci a udržení kvality života. Využívá se dlouhodobé farmakoterapie, z počátku inhibitory cholinesterázy (ChEI) a ke konci kombinací dvou léčiv ChEI a memantinem. Léčba medikamenty se kombinuje s psychoedukací [73], kdy jsou člověku systematicky a strukturovaně předávány informace a znalosti o nemoci, které pacientovi pomohou vyrovnat se s ní a zlepšit tak účinnost celé léčby [74].

Lék na AD ve stádiích demence je navzdory neustálému úsilí v blízké budoucnosti nereálný. Společným cílem nefarmakologické a farmakologické léčby proto zůstává snaha o minimalizaci kognitivního a funkčního poklesu vedoucího k invaliditě pacienta. Léky schválené FDA (Food and Drug Administration) ChEI – donepezil, galantamin, rivastigmin a antagonisty N-methyl-D-AP5 (NMDA) memantin mohou snížit progresi klinických příznaků a invalidity [75].

ChEI a memantin mají komplementární mechanismus působení a potenciálně přídavný efekt, zároveň jsou relativně dobře snášenlivé a bezpečné [76]. Medikamenty, ať už podávané samostatně či kombinovaně, prokázaly schopnost z krátkodobého i dlouhodobého hlediska zmírnit pokles poznávacích funkcí, zpomalit vznik neuropsychiatrických projevů a oddálit umístění v pečovatelském domě, bohužel stále bez prodloužení doby života [77].

ChEI usnadňují centrální cholinergní aktivitu snížením fyziologického odbourávání acetylcholinu enzymem acetylcholinesterázou v synaptické štěrbině. Zastavením funkčnosti enzymu pomocí inhibitoru se zvyšuje cholinergní neurotransmise. Memantin blokuje otevřené kanály NMDA receptorů a ovlivňuje tím glutamatergní přenos [73].

Ačkoliv od roku 2002 nebyly pro léčbu AD schváleny žádné nové léky, předpokládá se pozitivní účinek nově objevených sloučenin s různým mechanismem působení. Ty se dají rozdělit do tří skupin: terapie modifikující nemoc (DMT), symptomatické kognitivní zesilovače a symptomatická činidla zaměřená na symptomy demence [78]. Předpokládá se, že látky klasifikované jako DMT ovlivňují proteiny signálních drah zapojených v patobiologii AD, která souvisí s proteiny A $\beta$  a tau a mohou zpomalit progresi onemocnění. A $\beta$ <sub>42</sub> peptid, který při akumulaci může tvořit neurotoxické extracelulární plaky, je tvořen z APP štěpením enzymů  $\beta$ - a  $\gamma$ -sekretáz.

Inhibitory  $\beta$ - a  $\gamma$ -sekretázy se jeví jako slibné, objevily se ale obavy ohledně nežádoucích účinků. Mezi inhibitory  $\gamma$ -sekretázy patří semagecestat, jehož vývoj byl zastaven kvůli zvýšení výskytu rakoviny kůže a infekcí [79]. Mezi další patří avagecestat či tarenflurbil, u nichž byly také zjištěny vedlejší negativní účinky [73]. Vývoj některých inhibitorů  $\beta$ -sekretázy stále probíhá, jedná se o látky elenbecestat a CNP520, avšak u většiny látek této skupiny, jmenovitě verubecestat, atabecestat a LY3202606, se



vyskytly problémy, které vedly k přerušení klinických testů. Jednalo se hlavně o nedostatečné účinky v případě střední až těžké AD demence, vyrážky, pády pacientů a neuropsychiatrické symptomy [80].

Redukce A $\beta$ -plaku pomocí inhibitorů agregace je jednou z dalších perspektivních možností léčby. Konkrétně látka ELND005 v preklinických studiích zpomaluje progresi patologie AD neutralizací oligomerů A $\beta$ <sub>42</sub> a inhibicí jejich toxických účinků. Ovšem studie fáze II neprokázala účinnost u mírné a středně těžké AD a je možné, že ve vyšších dávkách by mohla být toxická [81].

Poznatky ze studií založených na aktivní imunoterapii vykazovaly ze začátku převážně negativní výsledky v důsledku aktivace T buněk. Proto byl změněn přístup směrem k pasivní imunoterapii. Ta se jeví jako velice slibná a v dnešní době stále probíhají klinické studie ve fázi III. Patří sem látky aducanumab, gantenerumab a BAN2401 [73]. Mezi další přístupy patří zabránění fosforylace zmiňovaného proteinu tau, který také souvisí s rozvojem AD. Tyto přístupy zahrnují inhibitory glykogensyntéazy kinázy-3, inhibitory agregace či stabilizátory mikrotubulů [73].

Jak již bylo dříve zmíněno, velké množství funkcí G4 v živých soustavách nastiňuje jejich využití při léčbě různých onemocnění. Proto se v posledních letech zvýšil zájem o výzkum těchto struktur, hlavně ve spojitosti s onkologickou léčbou. Zatímco se ale při léčbě rakoviny studie zaměřují především na regulaci udržování telomer, genovou expresi a duplikaci genomu [82][83], jsou zmiňovány i důkazy naznačující funkční roli G4 při neurologických a virových onemocnění [84][85]. Nejvíce vyskytujícím se jevem v případě neurologických poruch je expanze opakující se sekvence v genech přítomných v nervovém systému. V případě neurologických poruch byly navrženy dva mechanismy, kterými G4 k dané problematice přispívají. První zahrnuje vnitřní vliv G4 na své okolí, tvorbu a hromadění toxických transkribovaných RNA ložisek a dipeptidových repetitivních proteinů, které je způsobeno rozpínáním sekvencí bohatých na guanin. Druhý mechanismus představuje regulační procesy související s G4 a to mutace genu ovlivňující expresi patologických genů, vazebnou afinitu G4 k proteinům a následnou regulací vad [86]. Mezi neurologickými onemocněními studovanými pro jejich možnost léčby s využitím G4 patří mimo jiné ALS a frontotemporální demence. Zde v genu *C9orf72* G4 v místě intronu snižuje efektivitu replikace [87], brání transkripci [88], ovlivňuje sestřih mRNA [89] a indukuje tvorbu RNA ložisek [90]. V případě onemocnění syndromu fragilního X je funkcí G4 v genu *FMR1* na pozici 5'UTR především represe translace [91] a snížení translace proteinu RAN [92]. Dále na pozici exonu působí přítomný G4 na sestřih mRNA [93] a v genu *PSD-95* na pozici 3'UTR ovlivňuje represí translace [94]. Přítomnost G4 v genu *SNCA* v případě Parkinsonovy choroby na pozici 5'UTR má vliv na represí transkripce a snížení translace proteinu RAN [95]. Nakonec je samozřejmě již zmiňovaná AD, kterou ovlivňuje přítomnost více G4 v genech. V genu *APP* byla prokázána přítomnost G4 na pozici 3'UTR ovlivňující represí translace [96]. Podobný vliv má i G4 v genu *ADAM10* na pozici 5'UTR [97]. Dále v genech *BACE1* a *APOE* v místě exonu G4 ovlivňují sestřih mRNA [98] a genovou expresi [99].

Z důvodu výše zmíněné lokalizace G4 v genech ovlivňujících neurodegenerativní onemocnění, kam patří i AD, a vlastností z toho odvozených, byla v experimentální části zjišťována konzervovanost a z toho vyplývající důležitost G4 jak u člověka, tak u jeho homologů. Z hlediska bioinformatiky se homologie vyskytuje mezi sekvencemi nukleových kyselin nebo proteinů dvou organismů a odkazuje se na společný evoluční původ dvou sekvencí. Důvodem výskytu homologie může být speciace, která

probíhá díky vzniku mutací v genetickém kódu, čímž vznikají nové biologické druhy. Takové homology se nazývají ortology. Při horizontálním přenosu genů, tedy když jedinec přijme genetický materiál jiného jedince, vznikají homology označovány jako xenology. Třetím způsobem vzniku homologů je duplikace, při které se zdvojuje jakákoliv oblast DNA obsahující gen, čímž vznikají paralogy. Z hlediska významu nalezených G4 u člověka, například ve studiích [33] a [96], byl v následující praktické části sledován výskyt a umístění PQS struktur napříč homology člověka a *Homo sapiens* samotného. Z výsledků by poté bylo možné odvodit důležitost výskytu G4 v regulačních oblastech *APP* genu člověka a možný budoucí terapeutický potenciál právě v souvislosti s AD.

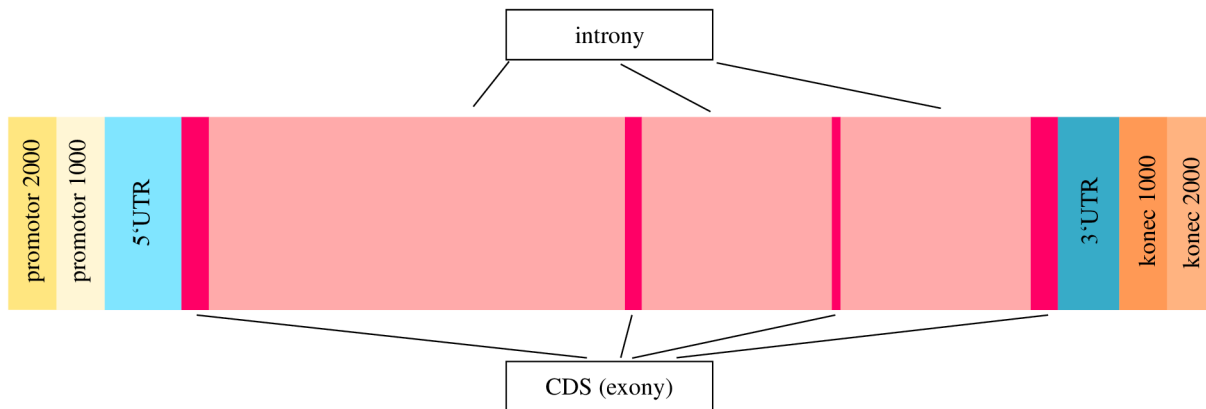
### 3 EXPERIMENTÁLNÍ ČÁST

#### 3.1 Materiály pro *in silico* analýzu

Pro tuto práci byla vybrána sekvence genu pro amyloid beta prekurzor u *Homo sapiens* a jeho genových homologů. Celkově bylo z internetové databáze National Center for Biotechnology Information (NCBI) získáno 347 sekvencí pro různé organismy. Organismy byly vybrány za účelem zjištění lokalizace PQS v rámci genu *APP* a jejich konzervovanosti u homologů.

#### 3.2 Získávání dat a jejich následná analýza v G4Hunter

S pomocí NCBI byl vytvořen dokument obsahující názvy jednotlivých homologů a jejich ID pro sekvence genu *APP*. Oblast byla navíc rozšířena o  $\pm 2000$  bází před a za genem, aby bylo možné sledovat výskyt G4 i v promotorech a koncových oblastech. Z databáze byly získávány jak textové soubory anotovaných oblastí, tak i FASTA soubory se sekvencí DNA. Tato data byla analyzována programem G4Hunter. Byly nastaveny podmínky velikosti okna na 25 a prahové hodnoty G4Hunter skóre na 1,2. Výsledkem analýzy byly soubory ve formátu CSV. Pro porovnání lokalizace nalezených PQS mezi homology bylo definováno následujících 8 oblastí: promotor\_2000 a promotor\_1000, což jsou oblasti  $-2000$  bp (páru bází) před genem rozdělené na dvě části, konec\_1000 a konec\_2000, což jsou oblasti  $+2000$  bp za genem také rozděleny na dvě části, 3' a 5'UTR, introny a CDS. Oblast CDS je kódující částí genu obsahující exony.



Obrázek 5: Schéma obecného genu a anotovaných oblastí: promotor 2000 bp před genem – tmavě žlutá barva, promotor 1000 bp před genem – světle žlutá barva, 5'UTR – světle modrá barva, CDS (exon) – tmavě růžová barva, intron – světle růžová barva (a střídání těchto dvou oblastí), 3'UTR - tmavě modrá barva, konec 1000 bp za genem – tmavě oranžová barva, konec 2000 bp za genem – světle oranžová barva.

Mezi dalšími informacemi získanými analýzou G4Hunter ve formátu CSV byly pro konkrétní organismy délka *APP* genu, obsah GC, frekvence výskytu a počet nalezených PQS. Dále, jak již bylo zmíněno, počty nalezených PQS v daných oblastech genu, délky oblastí, a to jak pro jednotlivé organismy, tak pro celou skupinu homologů.

Dále bylo opět z databáze NCBI ke každému organismu na základě jeho ID získáno jeho taxonomické zařazení pro pozdější vytvoření fylogenetického stromu. Fylogenetický strom graficky znázorňuje

evoluční vztahy organismů a podobnost objektů, které vycházejí z jejich historie [100]. Homology byly následně rozděleny do skupin právě dle taxonomických jednotek, které měly být v rozsahu 5 až 35 organismů. Pro velkou diverzitu v počtech zástupců v jednotlivých skupinách se toho ale nepodařilo dosáhnout, a proto jsou v některých krajních případech ve skupině pouze dva organismy. Příkladem toho je řád *Monotremata*, neboli ptakořitní, kde jedinými zástupci skupiny jsou ptakopysk a ježura australská. Dalším případem je třída *Chondrichthyes*, neboli paryby, která obsahuje pouze zástupce chimérovka tasvánská a rejnok hvězdnatý. U celků, které byly velmi početně obsáhlé a obsahovaly větší množství podcelků se skupinami alespoň po 10, byly tyto podcelky odděleny a zbylé organismy, které by byly ve svém podcelku osamoceny, byly ponechány ve zmíněném celku. Takovým příkladem jsou organismy z podtřídy *Neognathae*, neboli letci, kde byly vyčleněny objemné skupiny - čeleď *Falconiformes* (sokoli), nadřád *Galloanserae* (vrubozubí a hrabaví) a řády *Passeriformes* (pěvci) a *Pelecaniformes* (brodiví a pelikáni), ale zbylé podskupiny podtřídy *Neognathae* byly ponechány společně. Dalšími dvěma ponechanými nadskupinami byl klad *Eutheria* (placentálové) a nadřád *Teleostei* (kostnatí). Z takto definovaných taxonomických skupin organismů byla z každé určena průměrná hodnota frekvence výskytu PQS, nejvyšší a nejnižší hodnota a hodnota překryvu. Hodnota překryvu vyjadřuje, kolik plochy genu je zabráno PQS a byla vypočítána jako podíl hodnoty délky všech PQS bez překryvů a délky celé sekvence, vynásobené číslem 100 a zprůměrované pro celou skupinu organismů:

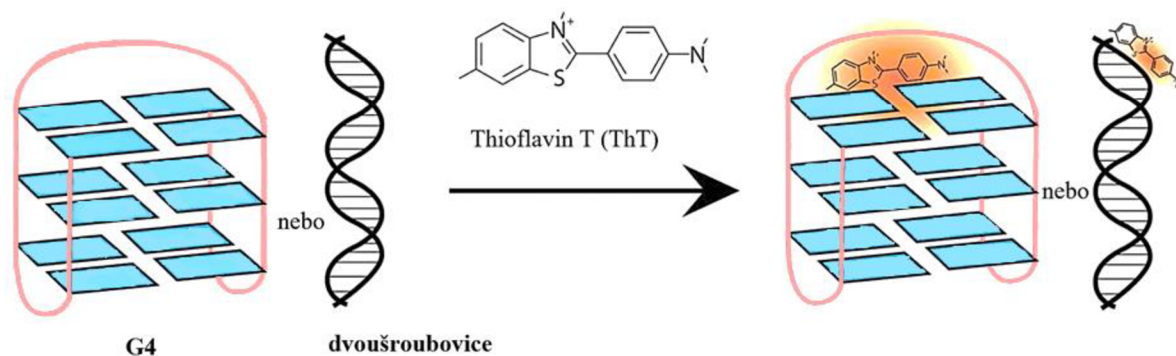
$$\text{Cov (\%)} = \frac{\text{délka všech PQS bez překryvů}}{\text{délka celé sekvence}} \cdot 100$$

Frekvence v rámci jednotlivých homologů byla vypočtena jako podíl počtu nalezených PQS a celkovou délkou *APP* genu pro daný organismus. Průměrná hodnota frekvence v daných skupinách byla vypočtena pomocí excelovské funkce PRŮMĚR( $\bar{x}$ ), stejně jako maximum a minimum podle funkce MIN( $\bar{x}$ ) a MAX( $\bar{x}$ ). Vytvořená tabulka byla přetvořena ve fylogenetický strom založený na hodnotách popsaných výše.

### 3.3 Důkaz tvorby G4 *in vitro*

#### 3.3.1 ThT analýza

Thioflavin T je molekula benzothiazol, která rozpoznává lidskou telomerickou sekvenci a lze ji použít pro rozlišení této struktury od ssDNA či dsDNA [101]. Jedná se o levný a rychlý test k určení schopnosti oligonukleotidu přijímat konformaci G4 [102]. Princip je v měření fluorescence molekuly ThT po navázání na oligonukleotid, kdy se intenzita zvýší, pokud je sekvence schopná tvořit G4. Předpokládá se, že za vysoký nárůst fluorescence může imobilizace molekuly ThT. Ta má dva kruhy, benzylaminový a benzathiolový, které volně rotují kolem osy jejich sdílené vazby uhlík-uhlík. Tato rotace zhasí excitované fotony, což způsobuje nízkou emisi volné molekuly. Pokud je ale rotace imobilizována interakcí, je zachován excitovaný stav a kvantový výtěžek fluorescence vzroste [103].



Obrázek 6: Princip ThT analýzy, kdy se ThT specificky váže na G4 a jeho fluorescence je zesílena. Pro porovnání, pokud je oligonukleotid jednořetězcový či dvojřetězcový, je pozorována slabší fluorescence, upraveno podle [102].

Pro další potvrzení tvorby G4 ve vybraných vzorcích, tedy kromě jejich vazby na ThT a následné zesílení signálu fluorescence oproti negativním kontrolám, byl každý vzorek připraven ve 12 různých prostředích soli KCl či NaCl a koncentracích od 0 do 100 mM. Ze studií plyne, že se G4 formují lépe v prostředí kationtu draselného než sodného, a to z důvodu velikosti kationtu a jeho hydratační energie. Tato energie potřebná pro koordinaci  $K^+$  s G4 je menší než energie pro  $Na^+$  [104]. I v této analýze se tedy předpokládá, že bude u studovaných oligonukleotidů vztah mezi intenzitou fluorescence a vybraným kationtem či jeho vzrůstající koncentrací jiný, než u negativních kontrol.

### 3.3.2 Materiály pro *in vitro* analýzu

Veškeré experimenty spojené s *in vitro* důkazem tvorby G4 ve vybraných sekvencích byly provedeny na Biofyzikálním ústavu Akademie věd ČR, kde byly poskytnuty i veškeré chemikálie a materiál.

### 3.3.3 Oligonukleotidy

Z výsledků analýzy G4Hunter byly vybrány sekvence z různých oblastí *APP* genu *Homo sapiens*, které vykazovaly vysoké G4Hunter skóre a měly tedy potenciál tvořit G4, a ty byly podrobeny analýze pomocí ThT. Celkem byly zvoleny 4 sekvence z oblastí 3'UTR, 5'UTR, promotoru a sekvence s nejvyšším G4Hunter skóre obsahující 6 opakování G-traktů, o které byla zmínka ve studii [33], s umístěním v intronu. Dále byly pro srovnání vytvořeny 3 sekvence jako negativní kontroly. Sekvence 25\_random byla generována náhodně. Pouze byla stanovena délka 25 bp, aby byla zachována podobná délka u všech oligonukleotidů. Sekvence 29\_random byla stejně jako předešlá generována náhodně. Byl u ní ale definován podíl GC (%) a délka 29 nukleotidů, oba parametry byly totožné se sekvencí 3\_UTR. Oligonukleotidy 29\_random a 3\_UTR byly podobné, ale bylo vycházeno z předpokladu, že budou mít jinou schopnost tvořit G4 s ohledem na různé rozmístění bází v sekvenci. Poslední negativní kontrolou byla NK6G s T-smyčkou. Ta má sice vysoké G4Hunter skóre, ale právě kvůli zmíněné smyčce tvořené thyminy by neměla mít potenciál G4 tvořit. Sekvence pro promotor a 6GC byly nalezeny v oblasti bohaté na C, tedy se záporným G4Hunter skóre. Ale při předpokladu, že je druhé vlákno DNA komplementární, byly sekvence převedeny pomocí softwaru Reverse Complement [105], tedy místo báze C je nyní G a místo A je T (a naopak). Syntetizované oligonukleotidy byly objednány od firmy Merck v čistotě HPLC a přehledně jsou vypsány v následující tabulce:

Tabulka 2: Použité oligonukleotidy pro ThT analýzu rozdělené na testované a kontrolní sekvence, zvláště seřazeny podle klesajícího G4Hunter skóre.

sekvence	název	umístění	G4Hunter skóre
TTGGGGGGCGGGGGCGGGGGCGGGGGTT	6GC	intron	3.000
GGGGCGGGTGGGGAGGGGTGCTCTGCTGG	3_UTR	3' nepřekládaná oblast	2.034
GGGGGTGGTCA GTGTGGATTTTGGGGG	promotor	promotor 1000 bp před	1.607
GGACGGCTCGGGCGCTGGAGGTGGG	5_UTR	5' nepřekládaná oblast	1.200
TTGGGGGGCTTTTTTCGGGGGGCGGGGGTT	NK6G	negativní kontrola	2.226
GTCTCATTATGATCGTACGCTATTC	25_random	negativní kontrola	-0.080
CGGATACGCCTGTTCTGTCGTGCCCGTA	29_random	negativní kontrola	-0.207

### 3.3.4 Chemikálie

Použité chemikálie jsou shrnuty v tabulce 3. Další, běžně se vyskytující chemikálie v biotechnologických a molekulárně-biologických laboratořích, nejsou v seznamu uvedeny a byly pořízeny v nejvyšší možné čistotě.

Tabulka 3: Souhrn použitých chemikálií v ThT analýze a jejich výrobci.

Chemikálie	výrobce
3,6-dimethyl-2-(4-dimethylaminofenyl)benzothiazoliový kation (ThT)	Sigma-Aldrich
Tris(hydroxymethyl)aminomethan (Tris)	Duchefa Biochemie
kyselina chlorovodíková (HCl)	Sigma-Aldrich
chlorid draselný (KCl)	Penta
chlorid sodný (NaCl)	Sigma-Aldrich

### 3.3.5 Přístroje

Pro měření intenzity fluorescence byl použit přístroj na detekci fluorescenčních emisních spekter značky TECAN Spark. Pro přípravu a ředění potřebných roztoků byly použity běžně se vyskytující přístroje v biotechnologických a molekulárně-biologických laboratořích.

### 3.3.6 Příprava roztoků

Byl připraven 1 mM zásobní roztok pufru Tris/HCl s pH = 7. Dále byly připraveny zásobní roztoky solí KCl a NaCl s koncentrací 3 a 2,5 M. Pro přípravu 1 mM zásobního roztoku ThT bylo naváženo 31,89 mg práškového ThT doplněného destilovanou vodou na objem 100 ml. Z takto vzniklého 1 mM roztoku bylo odpipetováno 10 µl do 10 ml vody pro přípravu 1 µM roztoku. Tento roztok bylo kvůli jeho fluorescenčním vlastnostem a možné degradaci na světlo nutno uchovávat ve tmě. Ze zmíněných roztoků byly následně pipetovány vypočítané objemy do zkumavek.

### 3.3.7 Příprava oligonukleotidů a měření

Obdržené lyofilizované oligonukleotidy od firmy Merck byly nejprve rozpuštěny v destilované vodě na 100 µM koncentraci a důkladně promíchány pomocí vortexu. Poté bylo připraveno 12 vzorků

pro každý oligonukleotid, které se lišily v koncentraci použité soli KCl a NaCl v rozmezí 0 až 200  $\mu\text{M}$ . Výsledná koncentrace oligonukleotidů ve vzorcích byla 2  $\mu\text{M}$ . Výsledná koncentrace Tris/HCl byla 100  $\mu\text{M}$ .

Takto získané vzorky byly promíchány na vortexu a stočeny na centrifuze. Následně byly na termobloku zahřáty na 90 °C po dobu 5 minut, kdy ve směsích proběhla denaturace, tedy rozpletení možných sekundárních struktur DNA. Vzorky byly ponechány při pokojové teplotě renaturaci přes noc, kdy se díky přítomnosti kationtů  $\text{K}^+$  a  $\text{Na}^+$  v sekvencích náchylných k tvorbě G4 právě tyto motivy měly vytvořit.

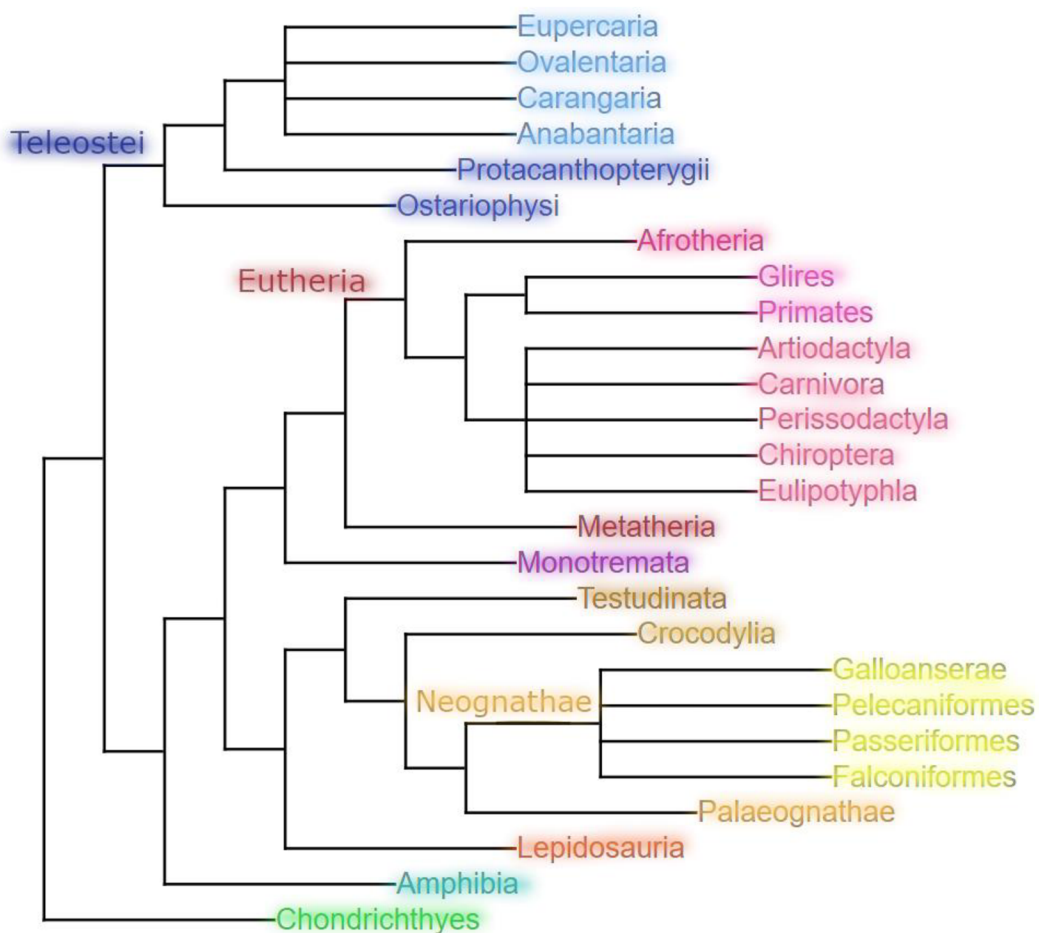
Po renaturaci bylo stejné množství z každé zkumavky nanášeno na černou destičku značky CORNING s 384 jamkami, do kterých byl přidán roztok ThT o koncentraci 1  $\mu\text{M}$ . Celkově byly analyzované vzorky DNA ku ThT v molárním poměru 2:1. Směsi se ponechaly hodinu inkubovat. Pomocí přístroje TECAN byl snímán emisní signál při 490 nm a po excitaci při 425 nm. Kromě měření fluorescence jedné vlnové délky byl proveden i sken emisních spekter shromážděných mezi 470 a 600 nm s krokem 2 nm.

## 4 VÝSLEDKY A DISKUZE

### 4.1 Analýza G4Hunter *in silico*

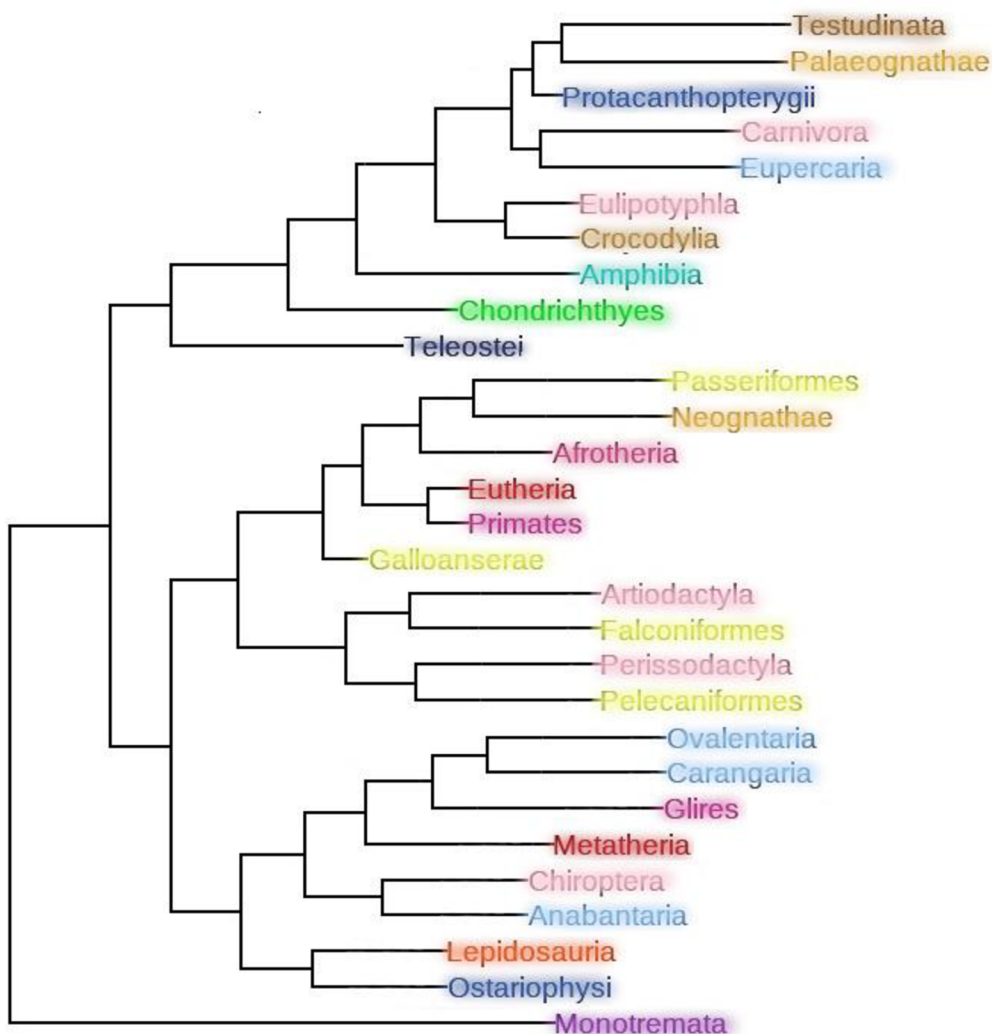
#### 4.1.1 Fylogenetický strom

Na základě zmíněného seznamu homologů a jejich taxonomického zařazení byla vytvořena tabulka s hodnotami průměru frekvence PQS na 1000 bp, nejnižší a nejvyšší hodnotou frekvence a hodnotou procentuálního pokrytí genomu PQS. Z tabulky byl vytvořen fylogenetický strom na základě podobnosti výskytu PQS Wardovou metodou shlukování (obrázek 8), která byla použita na základě článku [106]. Pro porovnání byl vytvořen i druhý strom dle oficiálního taxonomického zařazení (obrázek 7). V prvním zmíněném stromě jsou navíc podbarveny příbuzné taxonomické skupiny, stejné podbarvení skupin bylo poté zachováno i pro druhý fylogenetický strom.



Obrázek 7: Fylogenetický strom vybraných genetických homologů, založený na taxonomických jednotkách, byl vytvořen pomocí webové stránky NCBI a ETE toolkit [107]. Pro snadnější orientaci a porovnávání v dalších analýzách jsou barevně označeny taxonomické skupiny: ryby (modrá barva), obojživelníci (tyrkysová), plazi (hnědá), ptáci (žlutá), savci (červená/růžová), ptakořitní (fialová) a paryby (zelená).





Obrázek 8: Fylogenetický strom vybraných genetických homologů založený na podobnosti výskytu PQS. Pro snadnější orientaci a porovnání jsou barevně označeny taxonomické skupiny stejně jako na obrázku 7: ryby (modrá barva), dále obojživelníci (tyrkysová), plazi (hnědá), ptáci (žlutá), savci (červená/růžová), ptakořitní (fialová) a paryby (zelená).

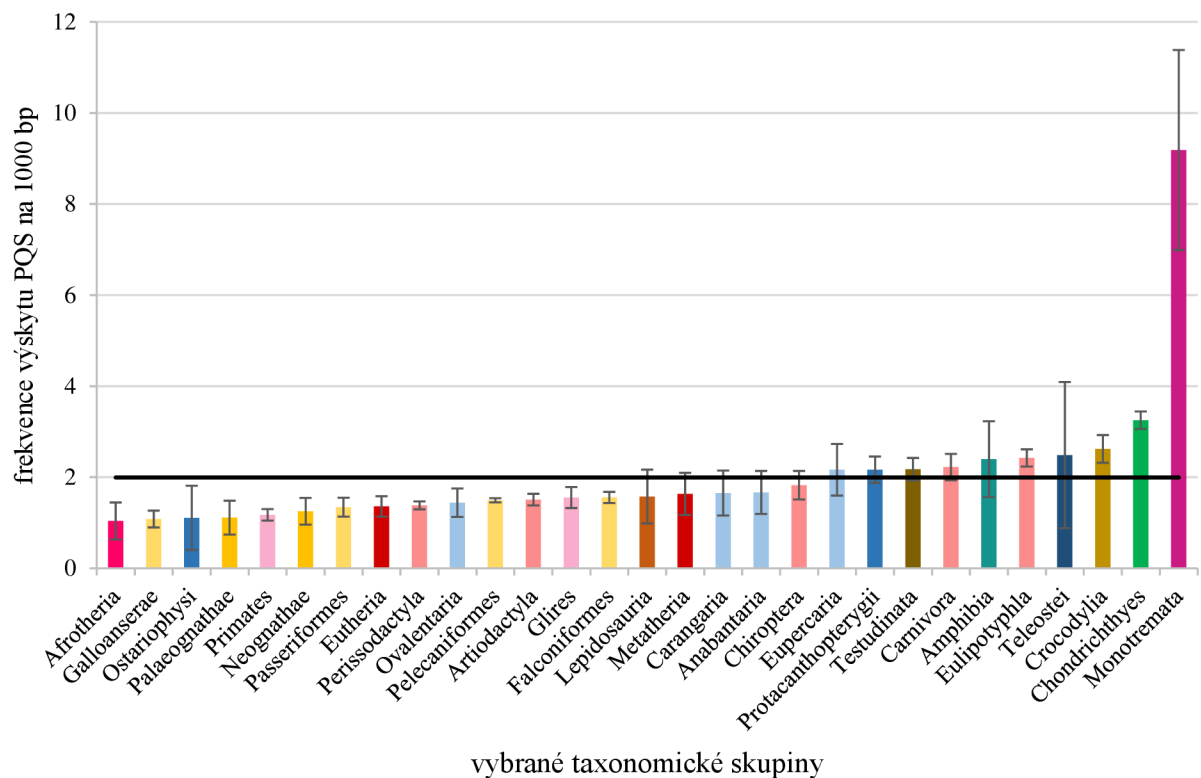
Porovnáním obou stromů na obrázcích 7 a 8 lze pozorovat, že se významně liší. Při evoluci DNA sekvencí organismů na úrovni celého genomu se v průběhu historie jednotlivé skupiny změnami ve své genetické informaci rozřazovaly do dnešních taxonů. Toto rozřazení ale nemusí platit pro všechny genomové lokusy. Shlukování taxonů na základě některých lokusů nemusí souhlasit se standardně akceptovaným fylogenetickým stromem. Mezi důvody způsobující odlišné signály rozdílnosti mezi lokusy patří rychlá divergence, imunitní odpověď, adaptace, gen pod evoluční selekcí nebo neúplné oddělení linií. Odlišnosti jednotlivých druhů v průběhu evoluce probíhají akumulací mutací, které se v populacích organismů šíří [100].

Pro diskuzi jsem si vybrala skupinu *Primates*, do které patří i *Homo sapiens* (člověk). Na obrázku 7 je patrné, že primáti patří do jedné větve společně se skupinou *Glires*, což je skupina savců spojující hlodavce a zajícovce. Dalšími příbuznými skupinami jsou i *Artiodactyla* (sudokopytníci), *Carnivora* (šelmy), *Perissodactyla* (lichokopytníci), *Chiroptera* (letouni) a *Eulipotyphla* (hmyzožravci), které patří

do třídy savců. V rámci fylogenetického stromu založeného na podobnosti výskytu PQS (obrázek 8) je skupina primáti ve stejné větvi jako anotovaná skupina *Eutheria*, která je z hlediska taxonomie klad nadřazený řádu primáti a v této práci byla zavedena z důvodu popsaného v kapitole 3.2. Nejbližší příbuznou skupinu dle taxonomie – *Glires*, zařadil algoritmus do jiné větve společně s organismy ze skupin *Ovalentaria* a *Carangaria*, což jsou klady třídy paprskovitých ryb. Na sousední větvi se společným uzlem se nachází nadřád *Afrotheria*, starobylá skupina placentálních savců zahrnující chobotnatce, damany, sirény, hrabáče a dva řády hmyzožravců – bércouny a afrosoricidy. V rámci vzdálenějších podobností jsou tu skupiny se společným uzlem *Passereiformes* (pěvci), *Neognathae* (letci) a *Galloanserae* (vrubozobí a hrabaví). Všechny tři skupiny přitom patří do třídy ptáci. Zbylé zmíněné skupiny třídy savců jsou rozptýleny v rámci celého stromu, není mezi nimi tedy větší podobnost. Třída ptáci a třída savci, kam patří řád primáti, náleží jako všechny ostatní studované homology do společného kladu *Euteleostomi*, neboli kostnatí obratlovci, jedinou výjimku tvoří třída *Chondrichthyes* (paryby). Z pohledu podobnosti výskytu PQS primátů a řádů ptáků koreluje fylogenetický strom na obrázku 7 s fylogenetickým stromem na obrázku 8 v tom smyslu, že ptáci a savci společně s obojživelníky a plazy jsou ve společné větvi, ale ryby se v případě druhého hlavního uzlu od ostatních oddělují.

Dále byl vytvořen sloupcový graf, který shrnuje informace o průměrné frekvenci PQS pro definované fylogenetické skupiny (obrázek 9). Velmi zajímavou skupinu tvoří řád *Monotremata*, neboli ptakořitní, kteří průměrně dosahují 4,6násobku průměrné hodnoty, která činí 1,99 PQS na 1000 bp. Ze statistického hlediska je ovšem tato skupina problematická, protože obsahuje pouze dva zástupce. Mezi další nadprůměrnou skupinu patří třída *Chondrichthyes*, která je z evolučního hlediska od ostatních tříd velmi vzdálená. Paryby mají oproti průměru 1,6krát větší frekvenci výskytu PQS. Nadprůměrných hodnot dosahují také skupiny *Teleostei*, série *Eupercaria* patřící pod kostnaté a nadtříd *Protacanthopterygii*, *Amphibia* (třída obojživelníci), *Crocodylia* (řád krokodýli), *Palaeognathae* (podtřída běžci), *Testudinata* (nadřád řádu želvy), *Carnivora* a *Eulipotyphla*. Naopak nejnižší hodnoty byly zaznamenány u nadřádu *Afrotheria*, *Galloanserae* a *Ostariophysi*, které dosahují okolo poloviny hodnoty průměru.

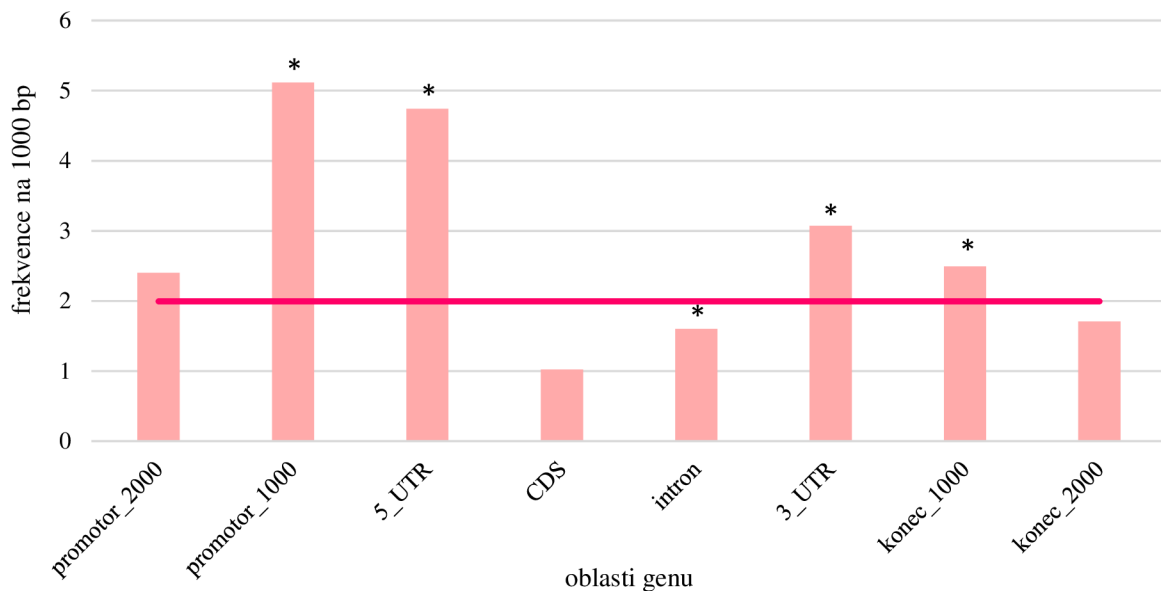
Zajímavé jsou skupiny ryb *Anabantaria*, *Carangaria* a *Eupercaria*, které se v grafu na obrázku 9 shlukly k sobě, což značí, že frekvence výskytu PQS sekvencí je v této evoluční větvi téměř identická. Naproti tomu homology savců *Perissodactyla*, *Artiodactyla*, *Chiroptera*, *Carnivora* a *Eulipotyphla* jsou od sebe odděleny a nachází se na celém spektru grafu frekvence nalezených PQS a jejich evoluční významnost není pro savce pozorovatelná. Pokud jde o ptačí homology, je většina skupin zařazena do první poloviny, co se frekvence PQS týče.



Obrázek 9: Graf zobrazující průměrnou frekvenci výskytu PQS pro jednotlivé vybrané taxonomické skupiny, které jsou seřazeny vzestupně. Černá vodorovná linie znázorňuje průměrnou hodnotou. Skupiny jsou zbarveny podle taxonomického rozdělení: ryby (modrá barva), obojživelníci (tyrkysová), plazi (hnědá), ptáci (žlutá), savci (červená/růžová), ptakořitní (fialová) a paryby (zelená). Dále jsou černou barvou zobrazeny směrodatné odchylky pro jednotlivé skupiny.

#### 4.1.2 Konzervovanost PQS v oblastech genu

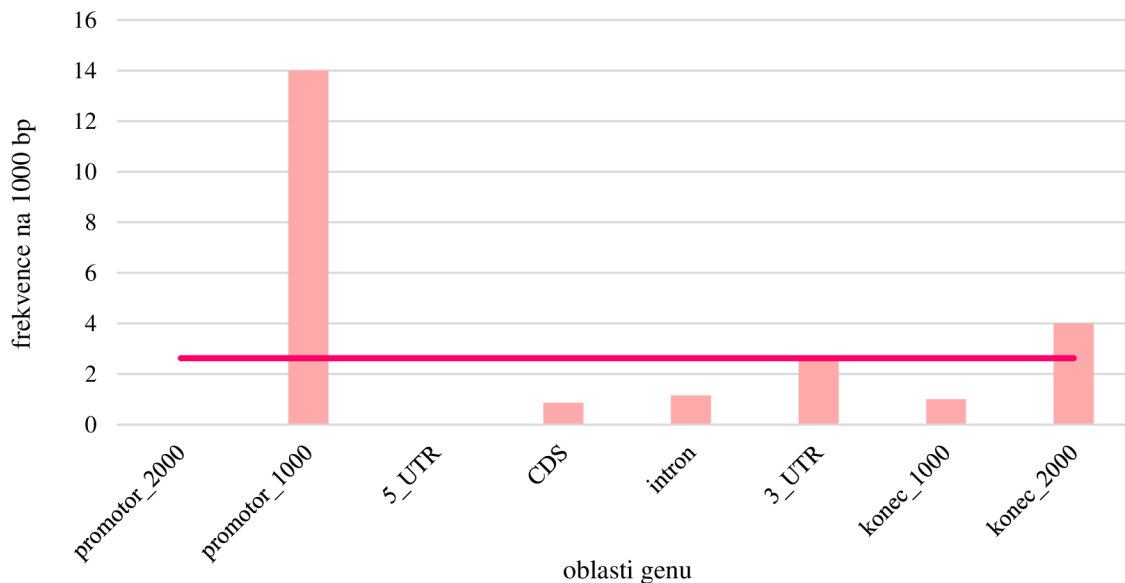
Po získání výsledků z G4Hunter analýzy bylo zjišťováno, ve kterých oblastech genu homologů *Homo sapiens* se PQS nejhojněji vyskytují. Frekvence výskytu PQS ve zmíněných oblastech byla vypočítána pomocí programu Excel. Graf byl sestaven z výsledků frekvence výskytu PQS vztahených na 1000 bp jak pro všechny homology dohromady, tak konkrétně pro *Homo sapiens*.



Obrázek 10: Frekvence PQS v homologních genech k lidskému APP genu. Frekvence vztahená na 1000 bp dané oblasti. Sloupcová světle růžová část grafu znázorňuje frekvenci nalezených PQS ve všech APP genech pro jednotlivé oblasti genu, vodorovná tmavě růžová linie je průměrná hodnota frekvence výskytu PQS pro skupiny homologů. Hvězdičkami jsou označeny statisticky významně odlišné oblasti.

V grafu na obrázku 10 je zaznamenána frekvence výskytu PQS v anotovaných oblastech pro celou skupinu homologů. Průměrná hodnota frekvence vztahená na oblasti je 1,99 a je patrné, že největší frekvence výskytu PQS je v promotorové oblasti 1000 bp před genem a čítá hodnotu 5,12 a v oblasti 5'UTR s hodnotou 4,74. Obě hodnoty jsou tedy zhruba 2,5násobně vyšší než průměr. Mezi další oblasti, jejichž frekvence je významně nadprůměrná, patří i 3'UTR a konec 1000 bp za genem. Tento výsledek odpovídá teoretickému předpokladu, že se G4 nacházejí v promotorových a koncových oblastech, kde mohou ovlivňovat genovou expresi skrz řízení transkripce [43] a v 5' a 3' nepřekládaných oblastech, kde ovlivňují stabilitu mRNA, transkripci a translaci, čímž regulují genovou expresi jak z prostorového, tak časového hlediska [107][109]. Velmi nízká frekvence PQS je naproti tomu v CDS a intronech.

Byla zhotovena statistika pro anotované oblasti. Pro otestování normality dat byl použit Shapiro-Wilkův test, který nepotvrdil normální rozložení. Byl proto využit neparametrický Kruskal-Wallisův test s Bonferroniho korekcí p-hodnoty. Jako post-hoc test pro párové porovnání skupin byl využit Dunnův test. Do úvahy byla brána statistická významnost každé anotované skupiny k celému genu.



Obrázek 11: Frekvence PQS v lidském *APP* genu, frekvence vztážená na 1000 bp pro anotované oblasti. Sloupcová světle růžová část grafu znázorňuje frekvenci nalezených PQS v *APP* genu *Homo sapiens*, vodorovná tmavě růžová linie je průměrná hodnota frekvence výskytu PQS pro *Homo sapiens*.

Při porovnání konzervovanosti PQS v lidském *APP* a mezi jeho homology se G4 v lidském genu nejvíce vyskytují v oblasti promotoru 1000 bp před genem. Hodnota 14,00 je 5,3násobně vyšší, než je průměrná hodnota 2,63. V porovnání s grafem na obrázku 10 je frekvence v této oblasti vyšší. Korelace výskytu G4 v promotoru u *Homo sapiens* a jeho homologů naznačuje evoluční důležitost těchto sekvencí v dané anotované oblasti pro tento konkrétní gen. Druhou nadprůměrnou hodnotu dosahuje oblast konec 2000 bp za genem s hodnotou 4,00, tedy asi 1,5krát vyšší než průměr. Oblast terminátorů, tedy koncových oblastí, se nachází na opačné straně sekvence nukleotidů. Ke konci procesu transkripce tato oblast ovlivňuje oddělení RNA polymerázy od templátového řetězce DNA [110]. I ta tedy může hrát roli v regulaci genové exprese. Naopak v oblastech promotor 2000 bp před genem a 5'UTR se u *Homo sapiens* nevyskytují PQS, které by měly frekvenci nad prahovou hodnotu 1,2.

#### 4.1.3 Konzervovanost sekvencí z *APP* genu *Homo sapiens*

Z výsledků G4Hunter analýzy pro *Homo sapiens* byla vytvořena tabulka s deseti sekvencemi s nejvyšší hodnotou G4Hunter skóre v absolutní hodnotě, tedy pro oba řetězce, jak paralelní, tak antiparalelní. Sekvence byly zapsány ve zkráceném zápisu, kde dolní index značí počet opakujících se nukleotidů. Tabulka dále obsahuje informace o pozici v rámci lidského *APP* genu, lokalizace sekvence v anotovaných oblastech, délku sekvence, označení řetězce, na kterém se sekvence nachází, a hodnotu jejího G4Hunter skóre v absolutní hodnotě.

Tabulka 4: Sekvence s nejvyšším G4Hunter skóre v absolutní hodnotě nalezené v lidském APP genu, tmavě růžově jsou zvýrazněné guaninové nukleotidy, světle růžové jsou cytosinové.

sekvence	pozice	lokalizace	délka (bp)	řetězec	G4Hunter skóre – absolutní hodnota
ACT <sub>5</sub> ACTGAGCAC <sub>3</sub> TGTGAC <sub>5</sub> AC <sub>1</sub> TGC <sub>6</sub> AGAGA <sub>2</sub> CA <sub>2</sub> C <sub>3</sub> GA <sub>1</sub> CTG	91 334	intron 3	64	–	1.64
TC <sub>2</sub> AG <sub>2</sub> T <sub>2</sub> CTG <sub>2</sub> AG <sub>2</sub> TG <sub>5</sub> A <sub>2</sub> TG <sub>4</sub> AGTG <sub>2</sub> AG <sub>2</sub> TG <sub>5</sub> A <sub>2</sub> TG <sub>4</sub> AGTG <sub>2</sub> AG <sub>2</sub> TG <sub>5</sub> A <sub>2</sub> TG <sub>4</sub> AGT <sub>2</sub> GT <sub>3</sub> A <sub>2</sub> TG <sub>2</sub> TGT	106 241	intron 3	86	+	1.63
TATATCTC <sub>3</sub> A <sub>2</sub> TGC <sub>1</sub> TATC <sub>1</sub> TC <sub>6</sub> T <sub>6</sub> AC <sub>4</sub> AC <sub>3</sub> ACAGTC <sub>4</sub> AGAGTGTGA	248 498	intron 13	61	–	1.62
C <sub>4</sub> GCA <sub>4</sub> TC <sub>2</sub> ACA <sub>1</sub> CTGAC <sub>2</sub> AC <sub>5</sub>	1 243	promotor 1000	28	–	1.61
TCTGC <sub>3</sub> AC <sub>3</sub> T <sub>2</sub> G <sub>2</sub> CG <sub>2</sub> CAC <sub>3</sub> GC <sub>6</sub> GC <sub>6</sub> ACA <sub>3</sub> GTGC TG <sub>3</sub> A	155 794	intron 6	60	–	1.52
T <sub>2</sub> GAG <sub>5</sub> TG <sub>2</sub> AGTG <sub>2</sub> C <sub>2</sub> AGT <sub>3</sub> G <sub>6</sub> CAGA	279 523	intron 16	34	+	1.50
CACAC <sub>5</sub> AG <sub>3</sub> C <sub>2</sub> TGT <sub>2</sub> G <sub>7</sub> TG <sub>6</sub> T <sub>2</sub> G <sub>5</sub> AG <sub>3</sub> A <sub>3</sub> GCATCAG <sub>2</sub> A GA <sub>3</sub>	167 938	intron 6	57	+	1.49
G <sub>4</sub> CAG <sub>2</sub> AT <sub>2</sub> GA <sub>3</sub> GATATG <sub>2</sub> CG <sub>4</sub>	224 278	intron 13	27	+	1.48
TG <sub>4</sub> AGTAT <sub>2</sub> GAGCAG <sub>2</sub> TG <sub>2</sub> CAG <sub>4</sub> AG	66 989	intron 2	29	+	1.45
AG <sub>2</sub> AGTGC <sub>3</sub> AC <sub>6</sub> AGT <sub>2</sub> CTAC <sub>6</sub> AC <sub>2</sub> AC <sub>3</sub> TAGC <sub>2</sub> T <sub>3</sub> GT G <sub>2</sub> T <sub>2</sub>	136 464	intron 5	54	–	1.44

Byla zkoumána konzervovanost výše zmíněných deseti sekvencí u všech homologů za použití programu Unipro UGENE [111]. Protože jsou ale sekvence různě dlouhé a některé dosahují délky až 86 bp, pravděpodobnost nalezení totožné sekvence není příliš vysoká a shoda byla nalezena pouze u zástupců řádu *Primates*, kam patří i člověk. Nalezené shody jsou popsány v následující tabulce, kde je pořadí sekvencí shodné s tabulkou 4.

Tabulka 5: Tabulka shrnující nalezené shody deseti sekvencí v genových homologích *Homo sapiens* s nejvyšším G4Hunter skóre, v prvním sloupci je pozice sekvence v APP genu člověka, ve druhém, zda byla nalezena shoda u jiných homologů, třetí sloupec jmenovitě uvádí organismy, ve kterých byla shoda nalezena, a pozici, kde se nachází.

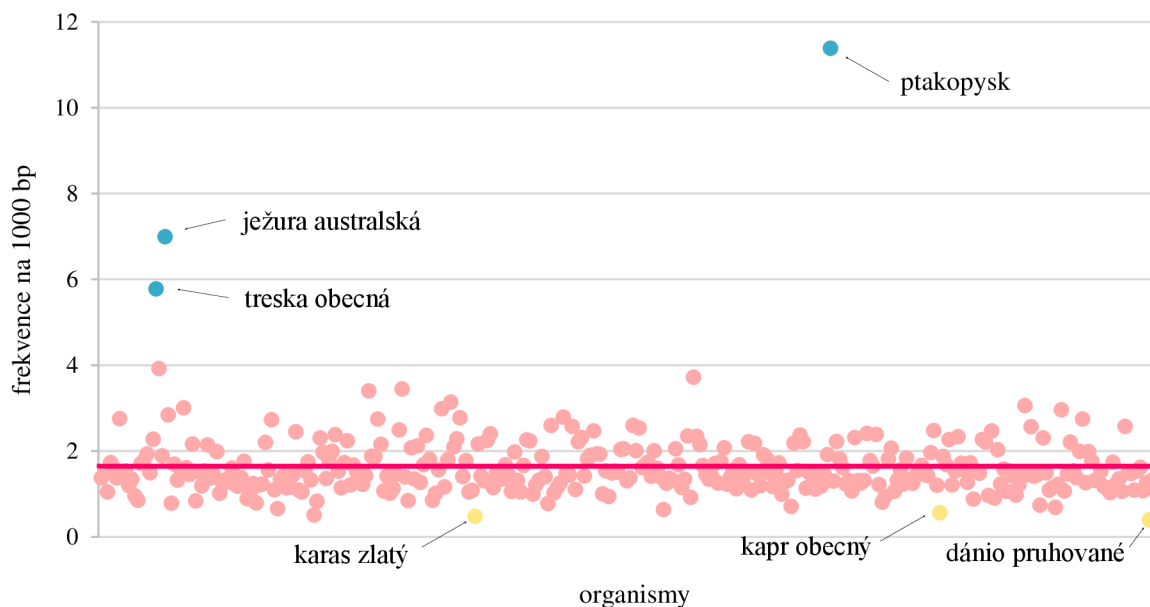
pozice	shoda	homolog a pozice shody (bp)					
91 334	ne						
106 241	ne						
248 498	ne						
1 243	ano	šimpanz učenílivý (1 296)	gibon bělolící (1 254)	šimpanz bonobo (1 238)	gibon stříbrný (1 254)	orangutan sumaterský (1 524)	gorila nížinná (1 587)
155 794	ne						
279 523	ano	šimpanz učenílivý (274 155)	šimpanz bonobo (273 631)	gorila nížinná (273 504)			
167 938	ne						
224 278	ne						
66 989	ano	šimpanz učenílivý (67 146)	šimpanz bonobo (66 937)				
136 464	ne						

Z tabulky 5 je patrné, že se některé ze sekvencí nejvíce náchylných ke tvorbě G4 nacházejí jak u člověka, tak i u jeho homologů ze skupiny *Primates*. Pozice nalezených shod jsou téměř totožné či podobné, což je z evolučního hlediska pochopitelné. Sekvence obsahující G-kvadruplex 6GC, která byla nalezena u *Homo sapiens* v intronu 6 na pozici 155 794 bp, nebyla ani u příbuzných ze skupiny *Primates*, ani u zbylých organismů nalezena. Lze tedy soudit, že je v rámci genu *APP* specifická pouze pro člověka.

Stejným způsobem jako v případě tabulky 5 byla zkoumána konzervovanost sekvencí, které byly podrobeny ThT analýze, tedy promotoru, 6GC, 3\_UTR a 5\_UTR. V případě sekvence 6GC nebyla nalezena shoda u žádného z homologů, jinak tomu však bylo u zbylých tří. V případě sekvence promotoru byla nalezena shoda u zástupců skupiny *Primates*, o čemž je již zmínka v předešlém odstavci. Mnohem větší počet shod byl nalezen se sekvencí 3\_UTR, kde bylo objeveno 11 homologů s identickou sekvencí v *APP* genu. Konkrétně se jednalo o organismy ze skupiny *Primati*, kteří dále náleží do oddělení *Catarrhini* (úzkonosí). V případě *Homo sapiens* byla 3\_UTR sekvence nalezena v 3' nepřekládané oblasti na pozici 291 820 bp. U všech nalezených homologů byla stejná posloupnost lokalizována v podobných oblastech v rozmezí 284 000 až 315 000 bp. Násobně větší shoda byla v případě 5\_UTR sekvence, kdy bylo objeveno 44 shodných sekvencí. 5\_UTR sekvence byla nalezena v 5' nepřekládané oblasti a u *Homo sapiens* je lokalizována na pozici 2 187 bp. Všechny homology s identickou posloupností náleží do třídy savců, konkrétně jsou zde zastoupeny skupiny *Glires*, *Chiroptera*, *Perissodactyla*, *Eulipotyphla*, *Eutheria* a samozřejmě *Primates*, ve které bylo nalezeno nejvíce takových organismů. Sekvence byla nalezena na pozicích obdobných jako v případě *Homo sapiens*, tedy okolo 2 000 bp.

#### 4.1.4 Výskyt a frekvence PQS u konkrétních homologů

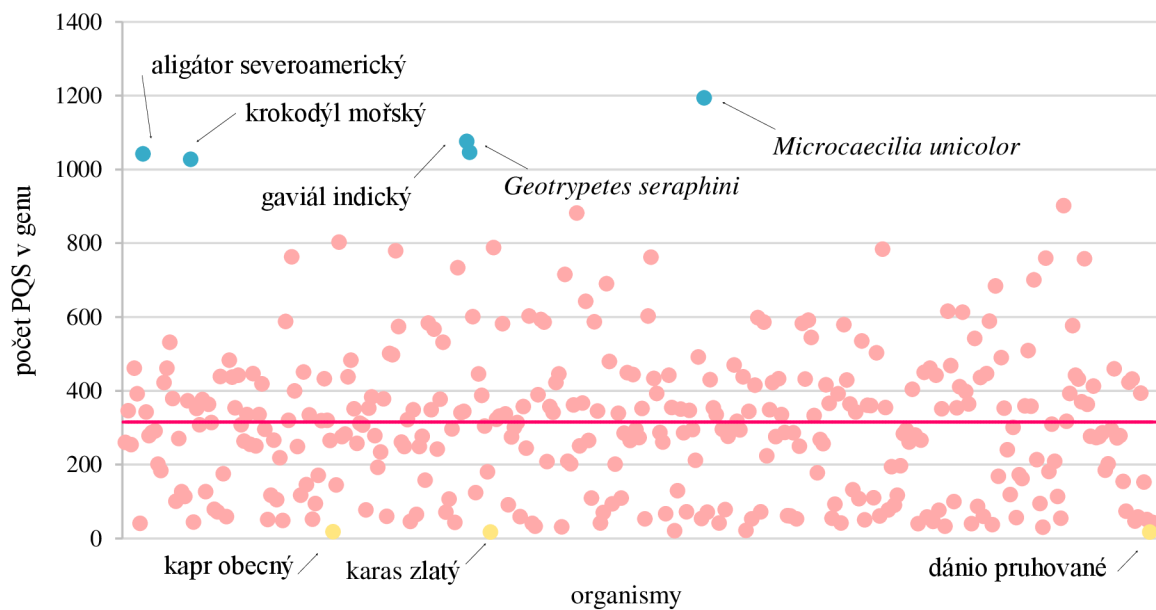
Byla vypočtena frekvence výskytu PQS pro jednotlivé homology a tyto hodnoty byly vyneseny do následujícího grafu na obrázku 12, aby byly zřetelné některé nalezené extrémní hodnoty. Jedná se o organismy ptakopysk s frekvencí 11,38, ježura australská s hodnotou 6,99 a treska obecná s hodnotou 5,78, které se významně liší od celkového průměru 1,65. Dva největší extrémy tohoto pozorování jsou organismy z řádu *Monotremata*, jež jsou jedineční tím, že ačkoliv kladou vajíčka, svá mláďata kojí a jedná se proto o savce. Pocházejí z australské pevniny, Tasmánie a Nové Guineje [112], jedná se tedy o endemity tamního území. Z hlediska nejnižší frekvence výskytu PQS je na prvním místě dánío pruhované s frekvencí 0,40, na druhém karas zlatý s frekvencí 0,47 a dále kapr obecný s frekvencí 0,51. Všechny tyto organismy patří do skupiny *Ostariophysi*, což je jeden z nadřádů ryb skupiny *Teleostei*. Ptakopysk, tedy homolog s nejvyšší frekvencí výskytu PQS, má asi 28,5x vyšší frekvenci než dánío pruhované, tedy homolog s nejnižší frekvencí. Délka *APP* genu pro ptakopyska je 34 444 bp a pro dánía pruhovaného 42 995 bp, je tedy zajímavé, že *APP* gen ptakopyska je 0,8x kratší než u dánía pruhovaného.



Obrázek 12: Celková frekvence PQS v APP genu u jednotlivých homologů znázorněna světle růžovými body, s vyznačenými extrémními hodnotami, pro maximum modrou barvou a to pro organismy ptakopysk, ježura australská a treska obecná, pro extrémní minimální hodnoty znázorněné žlutou barvou dánio pruhované, kapr obecný a karas zlatý. Tmavě růžovou vodorovnou linií je znázorněna průměrná hodnota frekvence nalezených PQS v APP genu u jednotlivých homologů.

Obdobný graf na obrázku 13 obsahuje také srovnání všech organismů, ale pro počty PQS nevztážené na délku genu homologů. Nejvyšší počet nalezených PQS je u obojživelníka *Microcaecilia unicolor*, kde dosahuje hodnoty 1194. Dalšími extrémními, tedy až třináásobku průměru 314,87, dosahuje obojživelník *Geotrypetes seraphini* s hodnotou 1076, plaz gaviál indický 1046, aligátor severoamerický 1042 a krokodýl mořský 1027. V případě počtu PQS lze pozorovat větší diverzitu mezi jednotlivými homology než u frekvence nalezených PQS. Je také zajímavé, že z taxonomického hlediska jsou kromě prvního zmíněného všechny zbylé extrémní v počtu PQS mezi zástupci z řádu krokodýli. Z pohledu nejnižšího počtu PQS se opakují stejné organismy jako v případě frekvence, tedy dánio pruhované s počtem 17, karas zlatý také s počtem 17 a kapr obecný s počtem 18. Při porovnání nejvyššího nalezeného počtu PQS u *Microcaecilia unicolor* a nejnižšího počtu u dánie pruhovaného se v prvním zmíněném organismu nachází 70x více PQS než u organismu druhého. Pro porovnání, délka sekvenční APP genu u *Microcaecilia unicolor* je 321 332 bp a u dánie pruhovaného 42 995 bp. Délka je tedy asi jen 7,5x větší než u dánie pruhovaného.

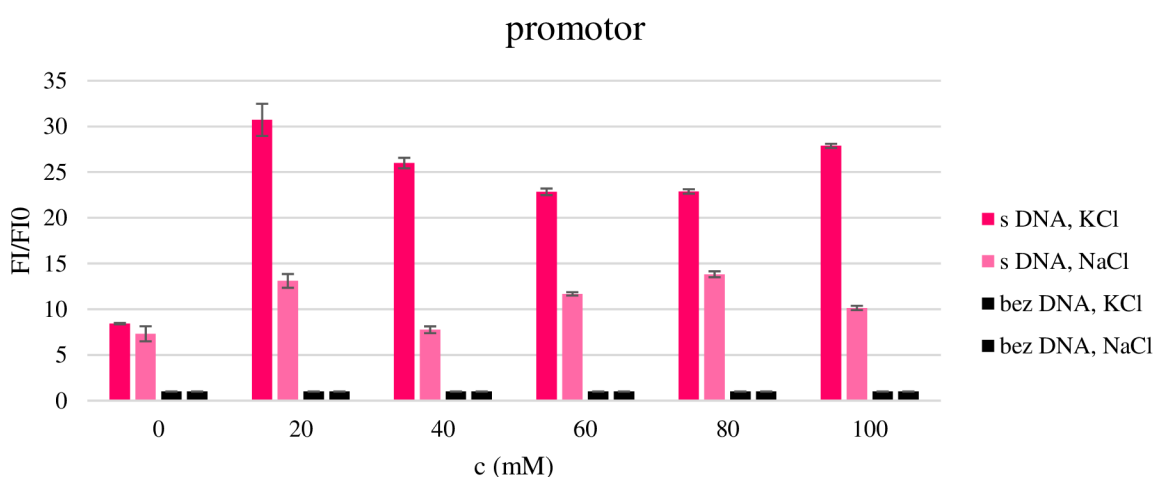




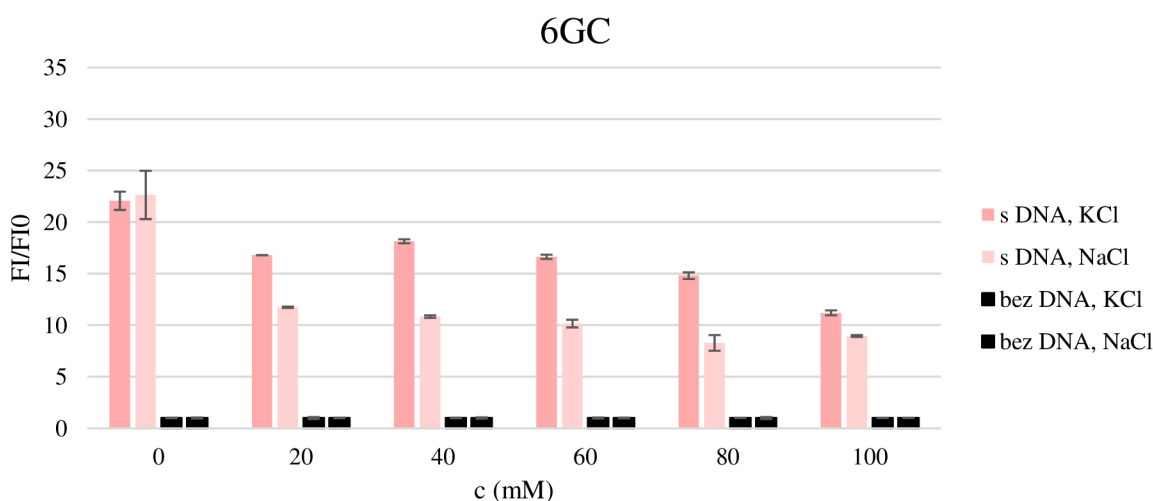
Obrázek 13: Celkový počet PQS v APP genu u jednotlivých homologů znázorněna světle růžovými body, s vyznačenými extrémními hodnotami, pro maximum modrou barvou pro organismy *Microcaecilia unicolor*, *Geotrypetes seraphini*, gaviál indický, aligátor severoamerický a krokodýl mořský, pro extrémní minimální hodnoty znázorněné žlutou barvou dánio pruhované, kapr obecný a karas zlatý. Tmavě růžovou vodorovnou linií je znázorněna průměrná hodnota nalezených PQS v APP genu u homologů.

## 4.2 Analýza G4 pomocí ThT *in vitro*

Sekvence popsané v kapitole 3.3.3 byly analyzované pomocí ThT. Naměřené hodnoty intenzity fluorescence pro každý vzorek v různých prostředích byly měřeny třikrát, z těchto hodnot byl vypočten průměr a hodnoty byly vztaženy k intenzitě fluorescence pufru. Testované sekvence byly porovnány s negativními kontrolami a mezi sebou. Byla vypočtena směrodatná odchylka pro každé měření. Všechny výpočty a grafy byly vytvořeny v programu Excel. Pomocí vzorce pro orientační vyloučení extrémních hodnot byly vyřazeny hodnoty, jejichž odchylka od průměru intenzity pro dané prostředí byla větší než trojnásobek směrodatné odchylky zbylých dvou záznamů měření.



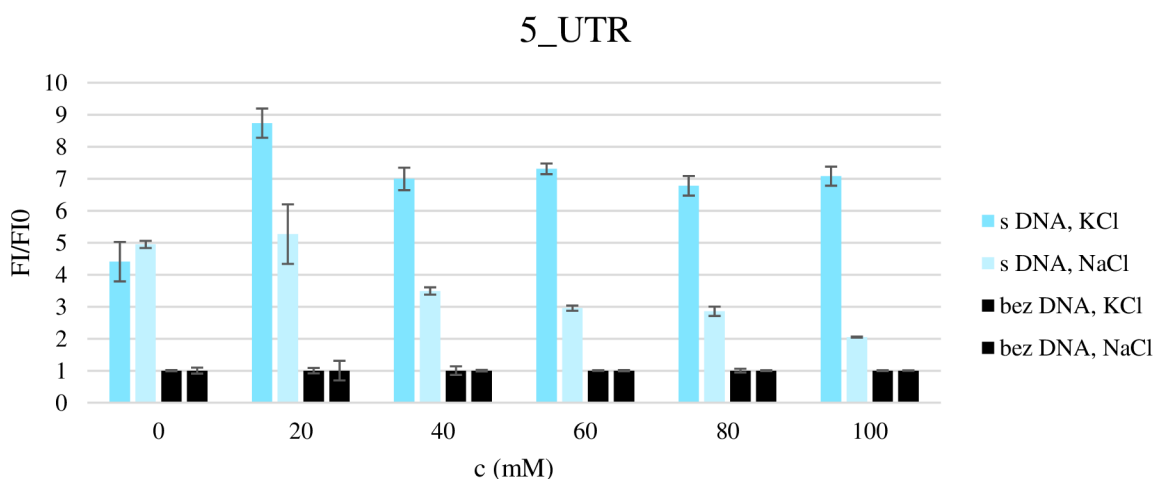
Obrázek 14: Měření intenzity fluorescence sekvence promotoru v prostředí KCl a NaCl o koncentracích 0,20,40,60,80 a 100 mM v porovnání s intenzitou fluorescence samotného ligandu (ThT), hodnoty normalizované na intenzitu pufru. Chybové úsečky vypočtené funkcí  $SMODCH.P(\bar{x})$  jsou značeny černými svislými ohraničenými čarami.



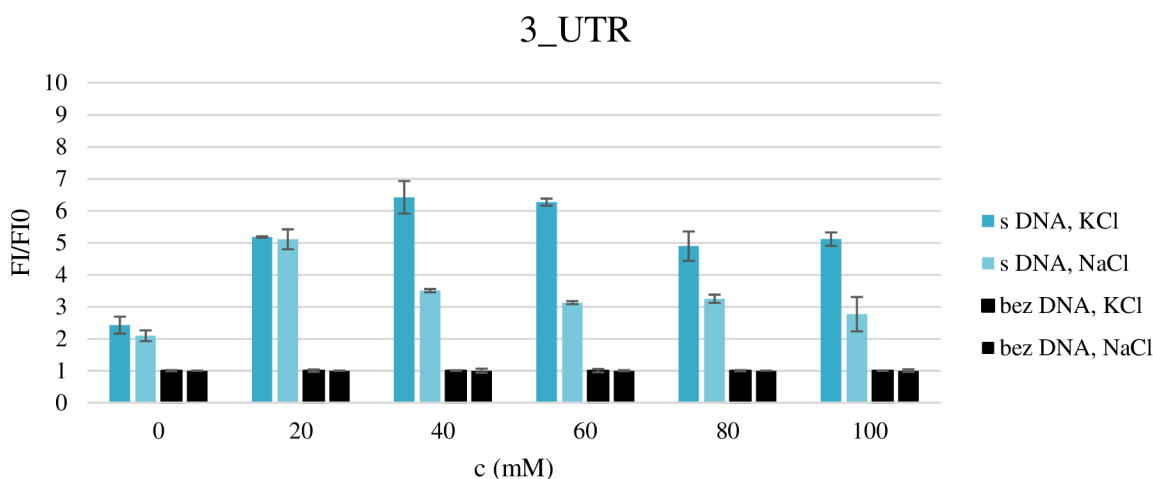
Obrázek 16: Měření intenzity fluorescence sekvence 6GC v prostředí KCl a NaCl o koncentracích 0,20,40,60,80 a 100 mM v porovnání s intenzitou fluorescence samotného ligandu (ThT), hodnoty normalizované na intenzitu pufru. Chybové úsečky vypočtené funkcí  $SMODCH.P(\bar{x})$  jsou značeny černými svislými ohraničenými čarami.

Z výsledků analýzy sekvencí pomocí programu G4Hunter vyplývá, že právě sekvence promotoru a 6GC by měly mít nejvyšší potenciál tvorby G4 struktury. Sekvence promotoru dosahující hodnoty G4Hunter skóre 1,607 vykazovala při měření intenzity fluorescence nejvyšší hodnoty. Maximální hodnota dosahuje 30,7násobku fluorescence samotného ligandu. Tohoto maxima dosáhla sekvence při koncentraci 40 mM KCl. Na obrázku 15 lze také pozorovat vzrůstající trend při měření vzorků o vyšší koncentraci soli a zároveň i vyšší hodnoty v prostředí chloridu draselného.

Pro sekvenci 6GC bylo vypočítáno G4Hunter skóre 3,00 a při *in vitro* analýze dosáhla ve svém maximum 0 mM soli hodnoty 22,6násobku intenzity fluorescence samotného ligandu. V prostředí draselné soli je intenzita vyšší, než v prostředí soli sodné. Zvyšující se iontová závislost není v případě 6GC pozorovatelná a dochází zde spíše ke snižujícímu trendu. Může to být z důvodu výskytu silného G4 a rozprostření náboje. Tvorba takového G4 nejspíše není závislá na iontové síle, a proto byl při měření získán nejvyšší signál fluorescence při nulové koncentraci soli.

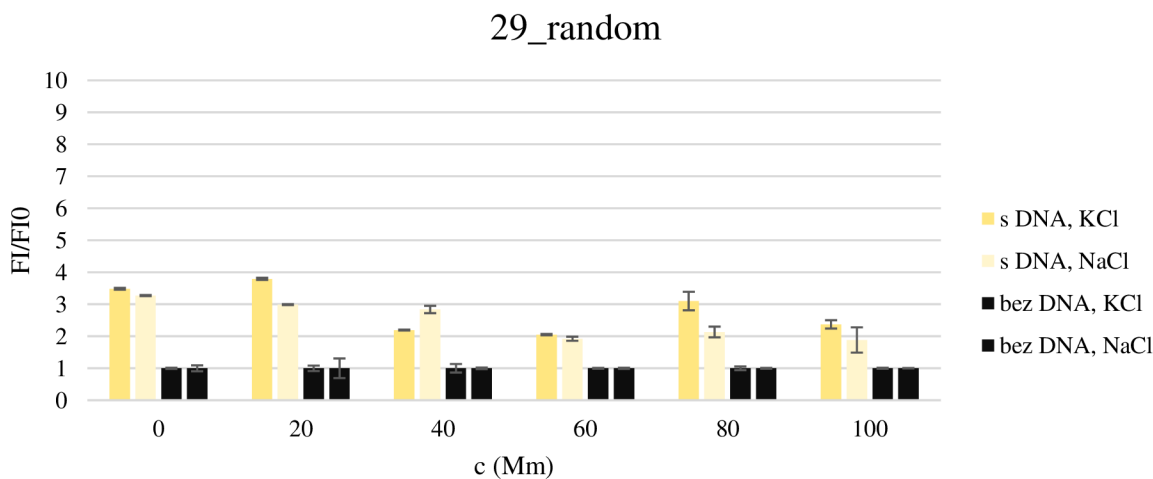


Obrázek 17: Měření intenzity fluorescence sekvence 5\_UTR v prostředí KCl a NaCl o koncentracích 0,20,40,60,80 a 100 mM v porovnání s intenzitou fluorescence samotného ligandu (ThT), hodnoty normalizované na intenzitu pufru. Chybové úsečky vypočtené funkcí  $SMODCH.P(\bar{x})$  jsou značeny černými svislými ohraničenými čarami.



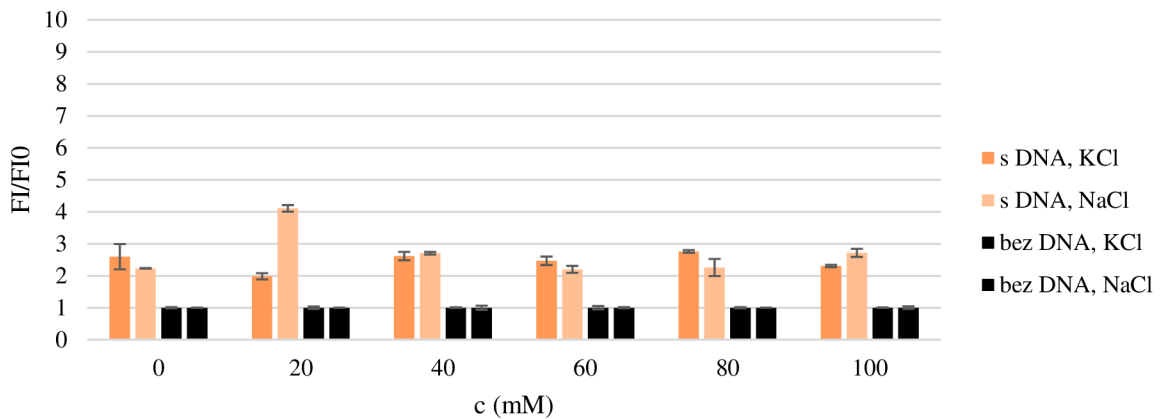
Obrázek 18: Měření intenzity fluorescence sekvence 3\_UTR v prostředí KCl a NaCl o koncentracích 0,20,40,60,80 a 100 mM v porovnání s intenzitou fluorescence samotného ligandu (ThT), hodnoty normalizované na intenzitu pufru. Chybové úsečky vypočtené funkcí  $SMODCH.P(\bar{x})$  jsou značeny černými svislými ohraničenými čarami.

Sekvencím 5\_UTR a 3\_UTR bylo vypočteno G4Hunter skóre 1.200 a 2.034. Sekvence 5\_UTR dosáhla při ThT analýze vyšších hodnot a maximální fluorescence byla zaznamenána při 40 mM koncentraci KCl, hodnota dosahuje 8,74násobku fluorescence ligandu. I v tomto případě lze pozorovat trend zvýšené intenzity v prostředí draselné soli oproti soli sodné. Maximum při sekvenci 3\_UTR bylo změřeno v 40 mM KCl a to 6,42násobek fluorescence ligandu. Intenzita byla naměřena vyšší v prostředí KCl.



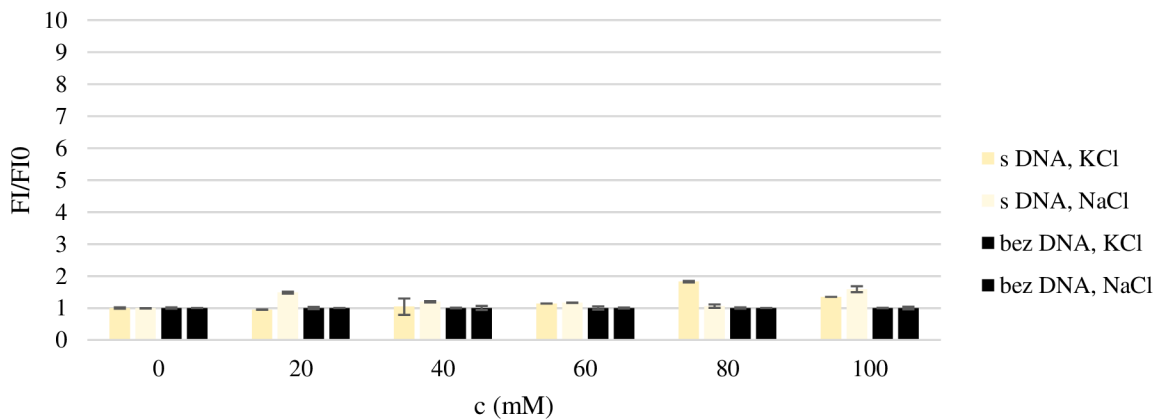
Obrázek 19: Měření intenzity fluorescence sekvence 29\_random v prostředí KCl a NaCl o koncentracích 0,20,40,60,80 a 100 mM v porovnání s intenzitou fluorescence samotného ligandu (ThT), hodnoty normalizované na intenzitu pufru. Chybové úsečky vypočtené funkcí  $SMODCH.P(\bar{x})$  jsou značeny černými svislými ohraničenými čarami.

## NK6G



Obrázek 20: Měření intenzity fluorescence sekvence NK6G v prostředí KCl a NaCl o koncentracích 0,20,40,60,80 a 100 mM v porovnání s intenzitou fluorescence samotného ligandu (ThT), hodnoty normalizované na intenzitu pufru. Chybové úsečky vypočtené funkcí  $SMODCH.P(\bar{x})$  jsou značeny černými svislými ohraničenými čarami.

## 25\_random

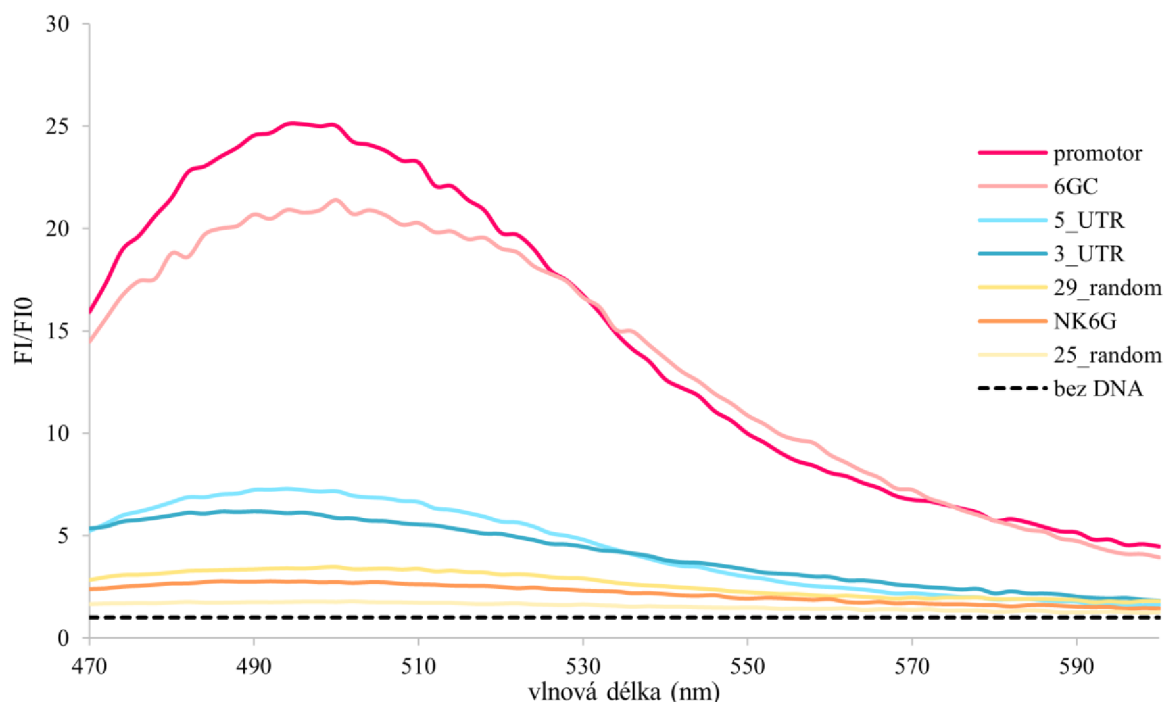


Obrázek 21: Měření intenzity fluorescence sekvence 25\_random v KCl a NaCl o koncentracích 0,20,40,60,80 a 100 mM v porovnání s intenzitou fluorescence samotného ligandu (ThT), hodnoty normalizované na intenzitu pufru. Chybové úsečky vypočtené funkcí  $SMODCH.P(\bar{x})$  jsou značeny černými svislými ohraničenými čarami.

Tři negativní kontroly, 29\_random, NK6G a 25\_random vykazují oproti sekvencím testovaným na tvorbu G4 jak malou závislost na iontové síle, tak téměř žádnou preferenci jedné z použitých solí. Oligonukleotid 29\_random dosahuje svého maxima 3,80násobek fluorescence samotného rozpouštědla při 40 mM KCl, NK6G při 20 mM NaCl dosahuje maxima 4,10násobek rozpouštědla a 25\_random 1,82násobku v 160 mM KCl.

Aby mohl být potenciál tvořit G4 v sekvencích lépe porovnatelný mezi sebou, byla pro každou sekvenci určena maximální hodnota poměru intenzity vzorku ku intenzitě samotného rozpouštědla v daném prostředí. Maximální hodnoty skenu emisních spekter pro vlnové délky od 470 do 600 nm s krokem 2 nm pro každou sekvenci byly vyneseny do jednotného grafu, který je zobrazen na obrázku 22.

Protože byly provedeny dvě ThT analýzy, bylo nutné i skeny normalizovat na intenzitu fluorescence rozpouštědla v daném prostředí.

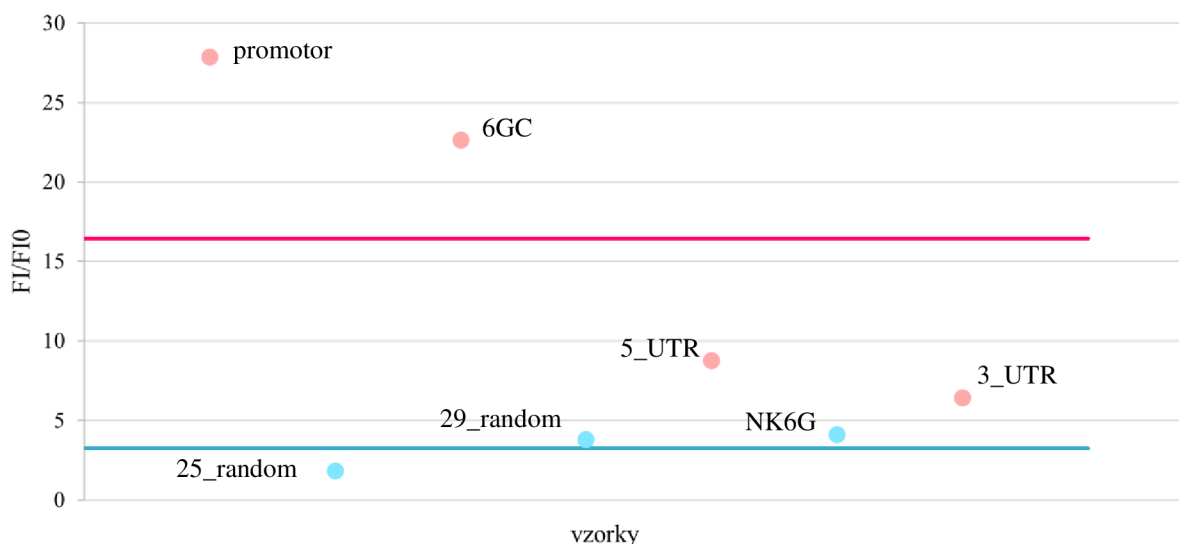


Obrázek 22: Graf emisních spekter všech oligonukleotidů při maximální fluorescenci, křivky viz legenda. Hodnoty spekter byly vztaženy na spektra pro roztok bez DNA.

Lze pozorovat, že výsledky určitou mírou korespondují s *in silico* predikcemi. Největší tendenci tvořit G4 má sekvence promotoru, která dosahuje okolo 25násobku intenzity fluorescence samotného ligandu. Následuje ji sken sekvence 6GC dosahující ve svém maximu okolo 21násobku intenzity fluorescence ligandu. Oligonukleotidy označené 3\_UTR a 5\_UTR jsou téměř identické, sekvence 5\_UTR dosahuje 7násobku a 3\_UTR 6násobku intenzity. Negativní kontroly měly nejnižší sílu fluorescence, sekvence 29\_random hodnotu 3,5násobku, NK6G 2,8násobku a 25\_random 1,8násobku.

Ačkoliv výsledky přesně nekorelují s hodnotami G4Hunter skóre, všechny domnělé, G4 tvořící sekvence prokázaly schopnost tvořit zmíněnou sekundární strukturu DNA. Vyplývá to z dat měření negativních kontrol, kdy jak sekvence vytvořené náhodně s G4Hunter skóre blízké nule, tak negativní kontrola NK6G s thyminovou smyčkou měly intenzitu fluorescence blízkou hodnotě intenzity ligandu. V průměru dosažených maxim mají sekvence domněle tvořící G4 struktury 15,6násobně vyšší signál fluorescence oproti rozpouštědlu, kdežto negativní kontroly pouze 2,8násobek. Dokazované sekvence měly při ThT analýze v průměru 5,6krát vyšší intenzitu fluorescence oproti negativním kontrolám.

Dalším faktem, který dokazuje tvorbu G4, je trend zvyšující se intenzity jak při zvětšování iontové síly, tedy koncentrace soli, tak i v případě použitého kationtu. U grafů jednotlivých sekvencí lze pozorovat, že má ve většině případů vyšší poměr fluorescence vzorek obsahující draselné ionty, což koreluje s předpovědí zmíněnou ve zdroji [104]. U téměř většiny domnělých G4 tvořících oligonukleotidů lze pozorovat závislost intenzity na zvyšující se koncentraci soli, tedy zvyšující se iontové síle. Naproti tomu negativní kontroly mají stejné zvýšení intenzity v případě většiny prostředí.



Obrázek 23: Bodový graf zvýšení fluorescence ThT v přítomnosti studovaných sekvencí. Vyneseny jsou maximální hodnoty pro každý z oligonukleotidů. Růžově jsou označeny hodnoty G4 tvořících oligonukleotidů, modře negativní kontroly. Obdobně jsou označeny průměry pro obě skupiny, v grafu se jedná o vodorovné linie – tmavě růžová je průměr G4 tvořících sekvencí a tmavě modrá je průměr negativních kontrol.

Rozdíl mezi vybranými sekvencemi a negativními kontrolami je pozorovatelný také na obrázku 23. Do bodového grafu byly vyneseny maximální hodnoty signálu při měření pro všechny sekvence a průměr jak G4 tvořících, tak negativních kontrolních oligonukleotidů. Průměrná hodnota zvýšení fluorescence u prvních zmíněných je 16,42, u druhých 3,24. Lze tedy konstatovat, že navýšení intenzity u mnou studovaných oligonukleotidů je 5x vyšší, než u negativních kontrol.

Je třeba si uvědomit, že ThT analýza nemusí plně korelovat s predikcemi *in silico*. Ve studii [102] se zaměřovali na důkaz specifičnosti ThT jakožto účinného nástroje pro selektivní rozpoznávání struktur schopných tvořit G4 oproti ssDNA, dsDNA a triplexním strukturám, ale jejich výsledky nebyly rozšířeny o výpočet G4Hunter skóre. Intenzita fluorescence ve zmíněném článku nekoreluje s výsledky G4Hunter skóre u PQS mezi sebou. Proto i v tomto testu nevykazuje sekvence 6GC nejvyšší fluorescenci, ačkoliv má nejvyšší G4Hunter skóre. Každopádně tvorba G4 byla potvrzená u všech PQS sekvencí predikovaných G4Hunterem. ThT analýza tedy slouží pouze jako nástroj pro potvrzení tvorby G4.

## 5 ZÁVĚR

V teoretické části této práce byly stručně shrnuty základní informace týkající se DNA a především nekanonické sekundární struktury G-kvadruplexů. Byla popsána jejich struktura, topologie, stabilita, výskyt a terapeutický potenciál, jejich interakce s ligandy a *in silico* analýza zabývající se predikcí tvorby zmíněných struktur. G4 jsou klíčové pro některé důležité biologické procesy, jako například regulaci replikace, transkripce, translace či údržbu telomer. Dále byl představen *APP* gen a jeho souvislost s onemocněním Alzheimerovou chorobou. *APP* gen kóduje transmembránový glykoprotein amyloidový prekurzor, který se vyskytuje v mnoha tkáních a orgánech, především v centrálním nervovém systému. V organismu probíhá jeho štěpení na menší fragmenty po dvou odlišných katabolických drahách. Při rozkladu  $\alpha$ -sekretázou mají vzniklé zlomy neuroprotektivní funkci, zatímco rozpustné fragmenty vznikající štěpením  $\beta$ -sekretázou mohou být toxické a zapříčinit degeneraci axonů a smrt nervových buněk. Protein APP a jeho zpracování v organismu se jeví jako kritické při nástupu a progresi Alzheimerovy choroby, při které dochází k hromadění toxických peptidických amyloidních beta plaků a jejich ukládání v mozku.

V experimentální části práce byl *in silico* studován lidský *APP* gen pomocí programu G4Hunter spolu s dalšími 346 genovými homology *Homo sapiens*. Pomocí programu byla provedena bioinformatická analýza s cílem vyhledat oblasti potenciálně tvořící G4, určit frekvenci jejich výskytu, lokalizace a konzervovanost napříč homology. Organismy byly na základě taxonomie rozděleny do taxonomických celků a byly vytvořeny dva fylogenetické stromy, jeden standardně akceptovaný a druhý shrnující informace o frekvenci výskytu PQS. Stromy se do značné míry lišily, přesto byly zachovány některé příbuznosti korelující s klasickým fylogenetickým stromem.

V rámci genu byly anotovány oblasti jako je promotor, 3'UTR a 5'UTR, introny, exony (CDS) a koncová oblast. Lokalizace G4, vyjádřena jako frekvence v rámci celé skupiny homologů, byla porovnána s lokalizací nalezených G4 u *Homo sapiens*. Bylo potvrzeno, že nejvyšší frekvence výskytu G4 v *APP* genu je jak u všech homologů, tak u člověka v promotorové oblasti 1000 bp před genem, což z evolučního hlediska poukazuje na jejich význam při regulaci genové exprese. Naproti tomu oblast 5'UTR, která měla v rámci všech homologů druhou nejvyšší frekvenci výskytu PQS, u *Homo sapiens* nebyla nalezena žádná PQS v této lokalizaci. Dále byly zkoumány frekvence a počty výskytu PQS také i jednotlivě mezi homology.

Z nalezených sekvencí v lidském *APP* genu bylo vybráno 10 sekvencí s nejvyšším G4Hunter skóre v absolutní hodnotě, tedy jak na paralelním, tak antiparalelním řetězci. Tyto sekvence byly hledány napříč homology. Shoda byla nalezena u tří sekvencí a všechny homology patřily do řádu *Primates*, kam náleží i člověk. Jednou ze sekvencí je i G4 nalezený v promotoru 1000 bp před genem, který značí evoluční důležitost například v regulaci transkripce. Nakonec byly pro potvrzení tvorby G4 *in vitro* vybrány 4 sekvence z lidského *APP* genu a určeny tři negativní kontroly. Vzorby byly podrobeny analýze pomocí ThT. Všechny testované sekvence s potenciálem tvořit G4 prokázaly svoji schopnost tvořit danou strukturu ve fyziologických podmínkách.



Analýzou *in silico* byla potvrzena přítomnost G4 v *APP* genu v blízkosti významných regulačních pozic. Dále byla potvrzena také evoluční důležitost nalezených PQS a jejich schopnost tvořit nekanonické sekundární kvadruplexní struktury. Tato zjištění mohou být přínosná pro budoucí výzkum zaměřující se na vývoj medicínských prostředků pro potlačení projevů AD či její úplnou léčbu, což je v dnešní době stále aktuální problém. Dnešní farmakoterapie je bohužel schopná pouze tlumit příznaky a zlepšovat kvalitu života pacientů s touto nemocí, nikoliv ji však úspěšně léčit. Zaměřením se na G4 nalezené v blízkosti oblastí genu podílejících se na regulaci genové exprese by mohlo napomoci omezení tvorby nebezpečných amyloidních beta plaků způsobujících progresi AD, například prostřednictvím interagujících ligandů.

## 6 SEZNAM ZKRATEK

.txt	standardní textový dokument, který obsahuje neformátovaný text
A	purinová báze adenin
AD	Alzheimerova choroba (z angl. Alzheimer's disease)
<i>Adam10</i>	ADAM metalopeptidázová doména 10 je gen kódující protein buněčného povrchu s unikátní strukturou, která má potenciálně jak adhezni, tak proteázové domény
ALS	amyotrofická laterální skleróza
<i>ApoE</i>	gen kódující apolipoprotein E, který se podílí na metabolismu tuků v těle savců, podtyp se podílí na Alzheimerově chorobě a kardiovaskulárním onemocnění
<i>APP</i>	gen kódující amyloidní prekurzorový protein, což je glykoprotein zanořený do membrány nervových buněk
APP695, APP751, APP770	majoritní izofomy proteinu APP
A $\beta$	amyloid beta, hlavní složka senilních plaků a neurofibrilárních klubek, které byly nalezeny v mozku pacientů s AD
<i>Bace</i>	gen kódující enzym beta-sekretáza 1 (proteináza kyseliny asparagové), který štěpí amyloidní prekurzorový protein beta-místa 1
BAN2401	humanizovaná IgG1 verze myši monoklonální protilátky mAb158, která se selektivně váže na velké, rozpustné A $\beta$ protofibrily
bp	pár bází, základní jednotka dvouřetězcových nukleových kyselin (z angl. base pair)
BRACO-19	selektivní inhibitor telomerázy s 3,6,9-trisubstituovanou akridinovou strukturou, systematický název N- , N' - (9 - {[4- (dimethylamino)fenyl]amino}akridin- 3,6 - diyl)bis(3- pyrrolidin - 1 - ylpropanamid)
C	pyrimidinová báze cytosin
<i>C9orf72</i>	gen kódující protein nacházející se v mnoha oblastech mozku, v cytoplazmě neuronů a také v presynaptických zakončeních
CDS	kódující oblast genu (z angl. coding region)
<i>c-Kit</i>	proto-onkogen kódující receptorové tyrosinkinázy
<i>c-Myb</i>	proto-onkogen kóduje protein, který zprostředkovává homotypické a heterotypické proteinové interakce
<i>c-Myc</i>	hlavní regulátor buněčného dělení a proliferace
CNP520	perorální inhibitor aspartylproteázy určený ke snížení produkce A $\beta$ za účelem prevence nebo léčby AD
CSV	soubor sestávající z řádků, ve kterých jsou jednotlivé položky odděleny čárkou či uvozovkami (z angl. Comma-separated values)

CTF	cytosinový terminální fragment
DMT	terapie modifikující nemoc (z angl. disease-modifying therapies)
dsDNA	dvouvláknová DNA (z angl. double-stranded DNA)
ELND005	orální stereoisomer inositolu, neutralizuje toxické A $\beta$ oligomery s nízkým obsahem N a brání jim v agregaci
FASTA	textový formát pro reprezentaci buď nukleotidových sekvencí nebo aminokyselinových sekvencí, ve kterých jsou nukleotidy nebo aminokyseliny reprezentovány pomocí jednopísmenných kódů
FDA	Úřad pro kontrolu potravin a léčiv (z angl. Food and Drug Administration)
<i>Fmr1</i>	gen kódující protein zvaný fragilní X mentální retardační protein nebo FMRP, který se nejčastěji vyskytuje v mozku; je nezbytný pro normální kognitivní vývoj a ženské reprodukční funkce
G	purinová báze guanin
G4	G-kvadruplex
GC	obsah GC je podíl guanino-cytosinového páru (GC) v rámci určité sekvence
hnRNA	heterogenní jaderná RNA, prekurzorová mRNA, která vzniká v jádře eukaryotické buňky a obsahuje strukturální geny
IGV	integrační genomový prohlížeč (z angl. Integrative Genomics Viewer)
<i>KRAS</i>	gen vytvářející protein, který se účastní buněčných signálních drah, řídí buněčný růst, zrání buněk a buněčnou smrt (z angl. Kirsten rat sarcoma virus)
LY3202606	inhibitor $\beta$ -sekretázy
mRNA	jednovláknová nukleová kyselina vznikající během transkripce DNA, která slouží jako předpis pro translaci
NCBI	Národní centrum pro biotechnologické informace (z angl. National Center for Biotechnology Information)
NCBI ID	identifikační číslo používané pro označení např. organismu či taxonomické skupiny v databázi NCBI
NMDA receptor	<i>N</i> -metyl-D-aspartát receptor, jeden z typů ionotropních glutamátových receptorů; aktivuje se, když se naváže glutamát a glycin a umožní pozitivně nabitým iontům proudit skrz buněčnou membránu
NMR	nukleární magnetická rezonance
PQS	v molekulární biologii sekvence DNA nebo RNA schopná vytvořit G-kvadruplex (z angl. Potential Quadruplex-forming Sequence)
pre-mRNA	primární transkript, první forma RNA vytvořena po transkripci při syntéze proteinů

<i>Psd95</i>	gen kódující protein (z angl. postsynaptic density protein 95), který je členem rodiny membránově asociovaných guanylátkináz
PSEN1 a PSEN2	formy proteinu presenilin, který je součástí komplexu $\gamma$ -sekretáza
RAN	malý G protein (GTPáza) z Ras genové nadrodiny, který se účastní membránového transportu přes jadernou membránu
sAPP $\beta$	rozpusný APP beta protein (z angl. soluble APP beta protein)
<i>Snca</i>	gen kódující alfa-synuklein, protein hojně exprimovaný v mozku, může sloužit k integraci presynaptické signalizace a membránového přenosu (z angl. synuclein alpha)
ssDNA	jednovláknová DNA (z angl. single-stranded DNA)
T	pyrimidinová báze thymin
tau	označení pro skupinu příbuzných proteinů, které se váží na mikrotubuly a hrají důležitou roli v nervové tkáni
ThT	benzothiazolová sůl thioflavin T používaná k vizualizaci a kvantifikaci v molekulární biologii
TMPyP4	kationtový porfyrin, systematický název 5,10,15,20-tetrakis-(N-methyl-4-pyridyl)porfin
<i>VEGF</i>	vaskulární endotelový růstový faktor (z angl. Vascular Endothelial Growth Factor A)

## 7 SEZNAM POUŽITÉ LITERATURY

- [1] APP amyloid beta precursor protein [ Homo sapiens (human) ]. *National Library of Medicine* [online]. [cit. 2022-04-13]. Dostupné z: <https://www.ncbi.nlm.nih.gov/gene/351#summary>
- [2] Alzheimer's disease. *Alzheimer's Society* [online]. [cit. 2022-04-13]. Dostupné z: <https://www.alzheimers.org.uk/about-dementia/types-dementia/alzheimers-disease>
- [3] WANG, Ernest, Ravi THOMBRE, Yajnas SHAH, Rachel LATANICH a Jiou WANG, 2021. G-Quadruplexes as pathogenic drivers in neurodegenerative disorders. *Nucleic Acids Research* [online]. **49**(9), 4816-4830 [cit. 2022-04-13]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkab164
- [4] BOCHMAN, Matthew L., Katrin PAESCHKE a Virginia A. ZAKIAN, 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* [online]. **13**(11), 770-780 [cit. 2021-12-06]. ISSN 1471-0056. Dostupné z: doi:10.1038/nrg3296
- [5] SEN, Dipankar a Walter GILBERT, 1990. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* [online]. **344**(6265), 410-414 [cit. 2022-04-19]. ISSN 0028-0836. Dostupné z: doi:10.1038/344410a0
- [6] BURGE, Sarah, Gary N. PARKINSON, Pascale HAZEL, Alan K. TODD a Stephen NEIDLE, 2006. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research* [online]. **34**(19), 5402-5415 [cit. 2022-04-19]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkl655
- [7] SIDDIQUI-JAIN, Adam, Cory L. GRAND, David J. BEARSS a Laurence H. HURLEY, 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences* [online]. **99**(18), 11593-11598 [cit. 2022-04-19]. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.182256799
- [8] JOHNSON, F. Brad, 2020. Fundamentals of G-quadruplex biology. *Quadruplex Nucleic Acids As Targets For Medicinal Chemistry* [online]. Elsevier, 2020, 3-44 [cit. 2022-04-21]. Annual Reports in Medicinal Chemistry. ISBN 9780128210178. Dostupné z: doi:10.1016/bs.armc.2020.06.004
- [9] LIPPS, Hans J. a Daniela RHODES, 2009. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology* [online]. **19**(8), 414-422 [cit. 2022-04-21]. ISSN 09628924. Dostupné z: doi:10.1016/j.tcb.2009.05.002
- [10] TODD, Alan K., 2007. Bioinformatics approaches to quadruplex sequence location. *Methods* [online]. **43**(4), 246-251 [cit. 2022-04-21]. ISSN 10462023. Dostupné z: doi:10.1016/j.ymeth.2007.08.004
- [11] HUPPERT, J.L., 2008. Hunting G-quadruplexes. *Biochimie* [online]. **90**(8), 1140-1148 [cit. 2022-04-21]. ISSN 03009084. Dostupné z: doi:10.1016/j.biochi.2008.01.014
- [12] BRÁZDA, Václav, Jan KOLOMAZNÍK, Jiří LÝSEK, Martin BARTAS, Miroslav FOJTA, Jiří ŠŤASTNÝ, Jean-Louis MERGNY a John HANCOCK, 2019. G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* [online]. **35**(18), 3493-3495 [cit. 2022-04-21]. ISSN 1367-4803. Dostupné z: doi:10.1093/bioinformatics/btz087

- [13] MINCHIN, Steve a Julia LODGE, 2019. Understanding biochemistry: structure and function of nucleic acids. *Essays in Biochemistry* [online]. **63**(4), 433-456 [cit. 2021-4-30]. ISSN 0071-1365. Dostupné z: doi:10.1042/EBC20180038
- [14] WATSON, J. D. a F. H. C. CRICK, 1953. Molecular Structure of Nucleic Acids: a Structure for Deoxyribose Nucleic Acid. *Nature* [online]. **171**(4356), 737-738 [cit. 2021-12-06]. ISSN 0028-0836. Dostupné z: doi:10.1038/171737a0
- [15] AVERY, Oswald T., Colin M. MACLEOD a Maclyn MCCARTY, 1944. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *Journal of Experimental Medicine* [online]. **79**(2), 137-158 [cit. 2021-12-06]. ISSN 1540-9538. Dostupné z: doi:10.1084/jem.79.2.137
- [16] RICH, Alexander a David R. DAVIES, 1956. a NEW TWO STRANDED HELICAL STRUCTURE: POLYADENYLIC ACID AND POLYURIDYLIC ACID. *Journal of the American Chemical Society* [online]. **78**(14), 3548-3549 [cit. 2022-03-14]. ISSN 0002-7863. Dostupné z: doi:10.1021/ja01595a086
- [17] DNA/RNA Base Pairs. *ChemTube3D* [online]. [cit. 2022-03-02]. Dostupné z: <https://www.chemtube3d.com/dnabases-2/>
- [18] BOCHMAN, Matthew L., Katrin PAESCHKE a Virginia A. ZAKIAN, 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* [online]. **13**(11), 770-780 [cit. 2021-4-30]. ISSN 1471-0056. Dostupné z: doi:10.1038/nrg3296
- [19] GESSNER, R.V., C.A. FREDERICK a G.J. QUIGLEY, 1989. The molecular structure of the left-handed Z-DNA double helix at 1.0-Å atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *J Biol Chem* . [online]. 264(14):7921-35 [cit. 2021-12-06]. Dostupné z: doi:10.2210/pdb1dcg/pdb
- [20] HA, Sung Chul, Ky LOWENHAUPT, Alexander RICH, Yang-Gyun KIM a Kyeong Kyu KIM, 2005. Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* [online]. **437**(7062), 1183-1186 [cit. 2021-12-06]. ISSN 0028-0836. Dostupné z: doi:10.1038/nature04088
- [21] VAN HOLDE, Kensal a Jordanka ZLATANOVA, 1994. Unusual DNA structures, chromatin and transcription. *BioEssays* [online]. **16**(1), 59-68 [cit. 2021-12-06]. ISSN 0265-9247. Dostupné z: doi:10.1002/bies.950160110
- [22] JAIN, Aklank, Guliang WANG a Karen M. VASQUEZ, 2008. DNA triple helices: Biological consequences and therapeutic potential. *Biochimie* [online]. **90**(8), 1117-1130 [cit. 2021-12-06]. ISSN 03009084. Dostupné z: doi:10.1016/j.biochi.2008.02.011
- [23] MANOR, Haim, B. Sridhara RAO a Robert G. MARTIN, 1988. Abundance and degree of dispersion of genomic d(GA)<sub>n</sub>·d(TC)<sub>n</sub> sequences. *Journal of Molecular Evolution* [online]. **27**(2), 96-101 [cit. 2021-12-06]. ISSN 0022-2844. Dostupné z: doi:10.1007/BF02138367
- [24] BACCOLLA, A., J. R. COLLINS, B. GOLD, et al., 2006. Long homopurine\*homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region. *Nucleic Acids Research* [online]. **34**(9), 2663-2675 [cit. 2021-12-06]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkl354

- [25] HAZEL, P., 2006. Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures. *Nucleic Acids Research* [online]. **34**(7), 2117-2127 [cit. 2022-04-04]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkl182
- [26] TLUČKOVÁ, Katarína, Maja MARUŠIČ, Petra TÓTHOVÁ, Lubos BAUER, Primož ŠKET, Janez PLAVEC a Viktor VIGLASKY, 2013. Human Papillomavirus G-Quadruplexes. *Biochemistry* [online]. **52**(41), 7207-7216 [cit. 2022-04-04]. ISSN 0006-2960. Dostupné z: doi:10.1021/bi400897g
- [27] VÍGLASKÝ, Viktor, Luboš BAUER a Katarína TLUČKOVÁ, 2010. Structural Features of Intra- and Intermolecular G-Quadruplexes Derived from Telomeric Repeats. *Biochemistry* [online]. **49**(10), 2110-2120 [cit. 2022-04-04]. ISSN 0006-2960. Dostupné z: doi:10.1021/bi902099u
- [28] GÓMEZ-MÁRQUEZ, Jaime, 2010. DNA G-quadruplex: structure, function and human disease. *FEBS Journal* [online]. **277**(17), 3451-3451 [cit. 2021-12-06]. ISSN 1742464X. Dostupné z: doi:10.1111/j.1742-4658.2010.07757.x
- [29] BIDZINSKA, Joanna, Graziella CIMINO-REALE, Nadia ZAFFARONI a Marco FOLINI, 2013. G-Quadruplex Structures in the Human Genome as Novel Therapeutic Targets. *Molecules* [online]. **18**(10), 12368-12395 [cit. 2021-12-06]. ISSN 1420-3049. Dostupné z: doi:10.3390/molecules181012368
- [30] BHATTACHARYYA, Debmalya, Gayan MIRIHANA ARACHCHILAGE a Soumitra BASU, 2016. Metal Cations in G-Quadruplex Folding and Stability. *Frontiers in Chemistry* [online]. **4** [cit. 2021-12-06]. ISSN 2296-2646. Dostupné z: doi:10.3389/fchem.2016.00038
- [31] G-quadruplex, 2001-. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation [cit. 2021-12-06]. Dostupné z: <https://en.wikipedia.org/wiki/G-quadruplex>
- [32] BURGE, Sarah, Gary N. PARKINSON, Pascale HAZEL, Alan K. TODD a Stephen NEIDLE, 2006. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research* [online]. **34**(19), 5402-5415 [cit. 2022-01-05]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkl655
- [33] BARTAS, Martin, Václav BRÁZDA, Václav KARLICKÝ, Jiří ČERVENĚ a Petr PEČINKA, 2018. Bioinformatics analyses and in vitro evidence for five and six stacked G-quadruplex forming sequences. *Biochimie* [online]. **150**, 70-75 [cit. 2022-01-05]. ISSN 03009084. Dostupné z: doi:10.1016/j.biochi.2018.05.002
- [34] ŠPAČKOVÁ, Naděžda, 2009. Tři jsou málo, pět je moc aneb seznamte se s G4. *Živa* [online]. **57**(3), 98-100 [cit. 2022-01-05]. ISSN 0044-4812. Dostupné z: <https://ziva.avcr.cz/files/ziva/pdf/tri-jsou-malo-pet-je-moc-aneb-seznamte-se-s-kvadru.pdf>
- [35] RNA-polymeráza II. *Velký lékařský slovník* [online]. [cit. 2022-02-02]. Dostupné z: <https://lekarske.slovníky.cz/pojem/rna-polymeraza-ii>
- [36] QIU, Jun, Mingxue WANG, Yan ZHANG, et al., 2015. Biological Function and Medicinal Research Significance of G-Quadruplex Interactive Proteins. *Current Topics in Medicinal Chemistry* [online]. **15**(19), 1971-1987 [cit. 2021-5-4]. ISSN 15680266. Dostupné z: doi:10.2174/1568026615666150515150803

- [37] SANDELL, L., 1993. Loss of a yeast telomere: Arrest, recovery, and chromosome loss. *Cell* [online]. **75**(4), 729-739 [cit. 2022-01-05]. ISSN 00928674. Dostupné z: doi:10.1016/0092-8674(93)90493-A
- [38] MINHAS, Gurpreet Singh, Daniel S. PILCH, John E. KERRIGAN, Edmond J. LAVOIE a Joseph E. RICE, 2006. Synthesis and G-quadruplex stabilizing properties of a series of oxazole-containing macrocycles. *Bioorganic & Medicinal Chemistry Letters* [online]. **16**(15), 3891-3895 [cit. 2022-01-05]. ISSN 0960894X. Dostupné z: doi:10.1016/j.bmcl.2006.05.038
- [39] LINGNER, Joachim, Julia Promisel COOPER a Thomas R. CECH, 1995. Telomerase and DNA End Replication: No Longer a Lagging Strand Problem? *Science* [online]. **269**(5230), 1533-1534 [cit. 2022-01-05]. ISSN 0036-8075. Dostupné z: doi:10.1126/science.7545310
- [40] WHITE, Laura K., Woodring E. WRIGHT a Jerry W. SHAY, 2001. Telomerase inhibitors. *Trends in Biotechnology* [online]. **19**(3), 114-120 [cit. 2022-01-05]. ISSN 01677799. Dostupné z: doi:10.1016/S0167-7799(00)01541-9
- [41] NEIDLE, S., READ M.A., 2001. G-quadruplex as therapeutic targets. *Biopolymers* [online]. **56**, 195 [cit. 2022-01-05]. PMID: 11745111. Dostupn z:doi:10.1002/1097-0282(2000)56:3<195::AID-BIP10009>3.0.CO;2-5
- [42] Onkogen. *Velký lékařský slovník* [online]. [cit. 2022-02-02]. Dostupné z: <https://lekarske.slovníky.cz/pojem/onkogen>
- [43] YANG, Danzhou, 2019. G-Quadruplex DNA and RNA. *G-Quadruplex Nucleic Acids* [online]. New York, NY: Springer New York, 2019-08-23, 1-24 [cit. 2021-5-5]. Methods in Molecular Biology. ISBN 978-1-4939-9665-0. Dostupné z: doi:10.1007/978-1-4939-9666-7\_1
- [44] KUŽMOVÁ, Erika, 2019. G-G4 a jejich role v regulaci transkripce. *Chemické listy* [online]. **113**(5) 320-327 [cit. 2022-06-01]. Dostupné z: <http://www.chemicke-listy.cz/ojs3/index.php/chemicke-listy/article/view/3358>
- [45] BOCHMAN, Matthew L., Katrin PAESCHKE a Virginia A. ZAKIAN, 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics* [online]. **13**(11), 770-780 [cit. 2021-12-06]. ISSN 1471-0056. Dostupné z: doi:10.1038/nrg3296
- [46] Oncogenes and Tumor Suppressor Genes. *American Cancer Society* [online]. [cit. 2022-02-02]. Dostupné z: <https://www.cancer.org/cancer/cancer-causes/genetics/genes-and-cancer/oncogenes-tumor-suppressor-genes.html>
- [47] EDDY, Johanna a Nancy MAIZELS, 2006. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Research* [online]. **34**(14), 3887-3896 [cit. 2021-12-06]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkl5297
- [48] VARSHNEY, Dhaval, Jochen SPIEGEL, Katherine ZYNER, David TANNAHILL a Shankar BALASUBRAMANIAN, 2020. The regulation and functions of DNA and RNA G-quadruplexes. *Nature Reviews Molecular Cell Biology* [online]. **21**(8), 459-474 [cit. 2022-01-05]. ISSN 1471-0072. Dostupné z: doi:10.1038/s41580-020-0236-x
- [49] FLEMING, Aaron M., Judy ZHU, Yun DING, Joshua A. VISSER, Julia ZHU a Cynthia J. BURROWS, 2018. Human DNA Repair Genes Possess Potential G-Quadruplex Sequences



- in Their Promoters and 5'-Untranslated Regions. *Biochemistry* [online]. **57**(6), 991-1002 [cit. 2022-01-06]. ISSN 0006-2960. Dostupné z: doi:10.1021/acs.biochem.7b01172
- [50] SIDDIQUI-JAIN, A., C. L. GRAND, D. J. BEARSS a L. H. HURLEY, 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences* [online]. **99**(18), 11593-11598 [cit. 2022-01-06]. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.182256799
- [51] KIM, Nayun, 2019. The Interplay between G-quadruplex and Transcription. *Current Medicinal Chemistry* [online]. **26**(16), 2898-2917 [cit. 2022-01-06]. ISSN 09298673. Dostupné z: doi:10.2174/0929867325666171229132619
- [52] ZHAO, Yongxi, Xiaolei ZUO, Qian LI, et al., 2021. Nucleic Acids Analysis. *Science China Chemistry* [online]. **64**(2), 1-33. [cit. 2022-01-06]. ISSN 1674-7291. Dostupné z: doi:10.1007/s11426-020-9864-7
- [53] MURAT, Pierre, Yashveer SINGH a Eric DEFRANCQ, 2011. Methods for investigating G-quadruplex DNA/ligand interactions. *Chemical Society Reviews* [online]. **40**(11) [cit. 2022-01-05]. ISSN 0306-0012. Dostupné z: doi:10.1039/c1cs15117g
- [54] SUN, Daekyu, Brian THOMPSON, Brian E. CATHERS, et al., 1997. Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *Journal of Medicinal Chemistry* [online]. **40**(14), 2113-2116 [cit. 2022-02-17]. ISSN 0022-2623. Dostupné z: doi:10.1021/jm970199z
- [55] ALCARO, Stefano, Caterina MUSETTI, Simona DISTINTO, et al., 2013. Identification and Characterization of New DNA G-Quadruplex Binders Selected by a Combination of Ligand and Structure-Based Virtual Screening Approaches. *Journal of Medicinal Chemistry* [online]. **56**(3), 843-855 [cit. 2022-02-17]. ISSN 0022-2623. Dostupné z: doi:10.1021/jm3013486
- [56] DE ARMOND, Richard, Stacey WOOD, Daekyu SUN, Laurence H. HURLEY a Scot W. EBBINGHAUS, 2005. Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1 $\alpha$  Promoter. *Biochemistry* [online]. **44**(49), 16341-16350 [cit. 2022-01-05]. ISSN 0006-2960. Dostupné z: doi:10.1021/bi051618u
- [57] DIXON, Isabelle M., Frédéric LOPEZ, Agueda M. TEJERA, Jean-Pierre ESTÈVE, Maria A. BLASCO, Geneviève PRATVIEL a Bernard MEUNIER, 2007. a G-Quadruplex Ligand with 10000-Fold Selectivity over Duplex DNA. *Journal of the American Chemical Society* [online]. **129**(6), 1502-1503 [cit. 2022-01-05]. ISSN 0002-7863. Dostupné z: doi:10.1021/ja065591t
- [58] CAMPBELL, Nancy H., Gary N. PARKINSON, Anthony P. RESZKA a Stephen NEIDLE, 2008. Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *Journal of the American Chemical Society* [online]. **130**(21), 6722-6724 [cit. 2022-01-05]. ISSN 0002-7863. Dostupné z: doi:10.1021/ja8016973
- [59] LI, Qian, Junfeng XIANG, Xudong LI, et al., 2009. Stabilizing parallel G-quadruplex DNA by a new class of ligands: Two non-planar alkaloids through interaction in lateral grooves. *Biochimie* [online]. **91**(7), 811-819 [cit. 2022-01-05]. ISSN 03009084. Dostupné z: doi:10.1016/j.biochi.2009.03.007

- [60] ZHANG, Sulin, Yanling WU a Wen ZHANG, 2014. G-Quadruplex Structures and Their Interaction Diversity with Ligands. *ChemMedChem* [online]. **9**(5), 899-911 [cit. 2022-01-25]. ISSN 18607179. Dostupné z: doi:10.1002/cmdc.201300566
- [61] PUIG LOMBARDI, Emilia a Arturo LONDOÑO-VALLEJO, 2020. a guide to computational methods for G-quadruplex prediction. *Nucleic Acids Research* [online]. **48**(1), 1-15 [cit. 2021-4-30]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkz1097
- [62] WEBBA DA SILVA, Mateus, 2007. NMR methods for studying quadruplex nucleic acids. *Methods* [online]. **43**(4), 264-277 [cit. 2022-04-05]. ISSN 10462023. Dostupné z: doi:10.1016/j.ymeth.2007.05.007
- [63] CAMPBELL, Nancy H. a Gary N. PARKINSON, 2007. Crystallographic studies of quadruplex nucleic acids. *Methods* [online]. **43**(4), 252-263 [cit. 2022-04-05]. ISSN 10462023. Dostupné z: doi:10.1016/j.ymeth.2007.08.005
- [64] DEL VILLAR-GUERRA, Rafael, John O. TRENT a Jonathan B. CHAIRES, 2018. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angewandte Chemie International Edition* [online]. **57**(24), 7171-7175 [cit. 2022-04-05]. ISSN 14337851. Dostupné z: doi:10.1002/anie.201709184
- [65] RACHWAL, Phillip A. a Keith R. FOX, 2007. Quadruplex melting. *Methods* [online]. **43**(4), 291-301 [cit. 2022-04-05]. ISSN 10462023. Dostupné z: doi:10.1016/j.ymeth.2007.05.004
- [66] *DNA analyser* [online]. [cit. 2022-01-06]. Dostupné z: <https://bioinformatics.ibp.cz/#/>
- [67] WU, Yili, Si ZHANG, Qin XU, Haiyan ZOU, Weihui ZHOU, Fang CAI, Tingyu LI a Weihong SONG, 2016. Regulation of global gene expression and cell proliferation by APP. *Scientific Reports* [online]. **6**(1) [cit. 2022-01-06]. ISSN 2045-2322. Dostupné z: doi:10.1038/srep22460
- [68] SAVONENKO, Alena V., Tatiana MELNIKOVA, Tong LI, Donald L. PRICE a Philip C. WONG, 2015. Alzheimer Disease. *Neurobiology of Brain Disorders* [online]. Elsevier, 2015, 321-338 [cit. 2022-05-05]. ISBN 9780123982704. Dostupné z: doi:10.1016/B978-0-12-398270-4.00021-5
- [69] YOSHIKAI, Shun-ichi, Hiroyuki SASAKI, Katsumi DOH-URA, Hirokazu FURUYA a Yoshiyuki SAKAKI, 1990. Genomic organization of the human amyloid beta-protein precursor gene. *Gene* [online]. **87**(2), 257-263 [cit. 2022-01-06]. ISSN 03781119. Dostupné z: doi:10.1016/0378-1119(90)90310-N
- [70] CRUCHAGA, Carlos, Sumitra CHAKRAVERTY, Kevin MAYO, et al., 2012. Rare Variants in APP, PSEN1 and PSEN2 Increase Risk for AD in Late-Onset Alzheimer's Disease Families. *PLoS ONE* [online]. **7**(2) [cit. 2022-01-06]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0031039
- [71] *MedlinePlus Trusted Health Information for You* [online], 2020. [cit. 2022-01-06]. Dostupné z: <https://medlineplus.gov/genetics/gene/app/#references>
- [72] POLYAK, Kornelia a Matthew MEYERSON. Overview: Gene Structure. *Holland-Frei Cancer Medicine* [online]. **6**. [cit. 2022-01-31]. ISBN 1550092499. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK12983/>

- [73] ATRI, Alireza, 2019. Current and Future Treatments in Alzheimer's Disease. *Seminars in Neurology* [online]. **39**(02), 227-240 [cit. 2022-02-17]. ISSN 0271-8235. Dostupné z: doi:10.1055/s-0039-1678581
- [74] EKHTIARI, Hamed, Tara REZAPOUR, Robin L. AUPPERLE a Martin P. PAULUS, 2017. Neuroscience-informed psychoeducation for addiction medicine: a neurocognitive perspective. *Brain Research in Addiction* [online]. Elsevier, 2017, 239-264 [cit. 2022-02-17]. Progress in Brain Research. ISBN 9780128135013. Dostupné z: doi:10.1016/bs.pbr.2017.08.013
- [75] CAPPELL, Jaclyn, Nathan HERRMANN, Stephen CORNISH a Krista L. LANCTÔT, 2010. The Pharmacoeconomics of Cognitive Enhancers in Moderate to Severe Alzheimer's Disease. *CNS Drugs* [online]. **24**(11), 909-927 [cit. 2022-02-17]. ISSN 1172-7047. Dostupné z: doi:10.2165/11539530-000000000-00000
- [76] ATRI, Alireza, José L MOLINUEVO, Ole LEMMING, Yvonne WIRTH, Irena PULTE a David WILKINSON, 2013. Memantine in patients with Alzheimer's disease receiving donepezil: new analyses of efficacy and safety for combination therapy. *Alz Res Therapy* [online]. **5**(1) [cit. 2022-02-17]. ISSN 1758-9193. Dostupné z: doi:10.1186/alzrt160
- [77] ROUNTREE, Susan D., Alireza ATRI, Oscar L. LOPEZ a Rachelle S. DOODY, 2013. Effectiveness of antidementia drugs in delaying Alzheimer's disease progression. *Alzheimer's & dementia : the journal of the Alzheimer's Association* [online]. **9**(3), 338-345 [cit. 2022-02-17]. ISSN 1552-5260. Dostupné z: doi:10.1016/j.jalz.2012.01.002
- [78] CUMMINGS, Jeffrey, Garam LEE, Aaron RITTER a Kate ZHONG, 2018. Alzheimer's disease drug development pipeline: 2018. *Alzheimers Dement* [online]. **4**(1), 195-214 [cit. 2022-02-17]. ISSN 2352-8737. Dostupné z: doi:10.1016/j.trci.2018.03.009
- [79] DOODY, Rachelle S., Rema RAMAN, Martin FARLOW, et al., 2013. a Phase 3 Trial of Semagacestat for Treatment of Alzheimer's Disease. *New England Journal of Medicine* [online]. **369**(4), 341-350 [cit. 2022-02-17]. ISSN 0028-4793. Dostupné z: doi:10.1056/NEJMoa1210951
- [80] EGAN, Michael F., James KOST, Pierre N. TARIOT, et al., 2018. Randomized Trial of Verubecestat for Mild-to-Moderate Alzheimer's Disease. *New England Journal of Medicine* [online]. **378**(18), 1691-1703 [cit. 2022-02-17]. ISSN 0028-4793. Dostupné z: doi:10.1056/NEJMoa1706441
- [81] SALLOWAY, S., R. SPERLING, R. KEREN, et al., 2011. a phase 2 randomized trial of ELND005, scyllo-inositol, in mild to moderate Alzheimer disease. *Neurology* [online]. **77**(13), 1253-1262 [cit. 2022-02-17]. ISSN 0028-3878. Dostupné z: doi:10.1212/WNL.0b013e3182309fa5
- [82] BALASUBRAMANIAN, Shankar, Laurence H. HURLEY a Stephen NEIDLE, 2011. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?. *Nature Reviews Drug Discovery* [online]. **10**(4), 261-275 [cit. 2022-03-11]. ISSN 1474-1776. Dostupné z: doi:10.1038/nrd3428
- [83] DÜCHLER, Markus, 2012. G-quadruplexes: targets and tools in anticancer drug design. *Journal of Drug Targeting* [online]. **20**(5), 389-400 [cit. 2022-03-11]. ISSN 1061-186X. Dostupné z: doi:10.3109/1061186X.2012.669384
- [84] SIMONE, Roberto, Pietro FRATTA, Stephen NEIDLE, Gary N. PARKINSON a Adrian M. ISAACS, 2015. G-quadruplexes: Emerging roles in neurodegenerative diseases

- and the non-coding transcriptome. *FEBS Letters* [online]. **589**(14), 1653-1668 [cit. 2022-03-11]. ISSN 00145793. Dostupné z: doi:10.1016/j.febslet.2015.05.003
- [85] WANG, Ernest, Ravi THOMBRE, Yajnas SHAH, Rachel LATANICH a Jiou WANG, 2021. G-Quadruplexes as pathogenic drivers in neurodegenerative disorders. *Nucleic Acids Research* [online]. **49**(9), 4816-4830 [cit. 2022-03-11]. ISSN 0305-1048. Dostupné z: doi:10.1093/nar/gkab164
- [86] XU, Jinglei, Haiyan HUANG a Xiang ZHOU, 2021. G-Quadruplexes in Neurobiology and Virology: Functional Roles and Potential Therapeutic Approaches. *JACS Au* [online]. **1**(12), 2146-2161 [cit. 2022-03-11]. ISSN 2691-3704. Dostupné z: doi:10.1021/jacsau.1c00451
- [87] THYS, Ryan Griffin a Yuh-Hwa WANG, 2015. DNA Replication Dynamics of the GGGGCC Repeat of the C9orf72 Gene. *Journal of Biological Chemistry* [online]. **290**(48), 28953-28962 [cit. 2022-03-11]. ISSN 00219258. Dostupné z: doi:10.1074/jbc.M115.660324
- [88] HAEUSLER, Aaron R., Christopher J. DONNELLY, Goran PERIZ, et al., 2014. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* [online]. **507**(7491), 195-200 [cit. 2022-03-11]. ISSN 0028-0836. Dostupné z: doi:10.1038/nature13124
- [89] VATOVEC, Sabina, Anja KOVANDA a Boris ROGELJ, 2014. Unconventional features of C9ORF72 expanded repeat in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neurobiology of Aging* [online]. **35**(10), 2421.e1-2421.e12 [cit. 2022-03-11]. ISSN 01974580. Dostupné z: doi:10.1016/j.neurobiolaging.2014.04.015
- [90] DEJESUS-HERNANDEZ, Mariely, Ian R. MACKENZIE, Bradley F. BOEVE, et al., 2011. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* [online]. **72**(2), 245-256 [cit. 2022-03-11]. ISSN 08966273. Dostupné z: doi:10.1016/j.neuron.2011.09.011
- [91] SANTORO, Michael R., Steven M. BRAY a Stephen T. WARREN, 2012. Molecular Mechanisms of Fragile X Syndrome: a Twenty-Year Perspective. *Annual Review of Pathology: Mechanisms of Disease* [online]. **7**(1), 219-245 [cit. 2022-03-11]. ISSN 1553-4006. Dostupné z: doi:10.1146/annurev-pathol-011811-132457
- [92] ASAMITSU, Sefan, Yasushi YABUKI, Susumu IKENOSHITA, et al., 2021. CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Science Advances* [online]. **7**(3) [cit. 2022-03-11]. ISSN 2375-2548. Dostupné z: doi:10.1126/sciadv.abd9440
- [93] SANTORO, Michael R., Steven M. BRAY a Stephen T. WARREN, 2012. Molecular Mechanisms of Fragile X Syndrome: a Twenty-Year Perspective. *Annual Review of Pathology: Mechanisms of Disease* [online]. **7**(1), 219-245 [cit. 2022-03-11]. ISSN 1553-4006. Dostupné z: doi:10.1146/annurev-pathol-011811-132457
- [94] DEMARCO, Brett, Snezana STEFANOVIC, Allison WILLIAMS, Kathryn R. MOSS, Bart R. ANDERSON, Gary J. BASSELL, Mihaela Rita MIHAILESCU a Yoon Ki KIM, 2019. FMRP - G-quadruplex mRNA - miR-125a interactions: Implications for miR-125a mediated translation regulation of PSD-95 mRNA. *PLOS ONE* [online]. **14**(5) [cit. 2022-03-11]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0217275

- [95] KOUKOURAKI, Pelagia a Epaminondas DOXAKIS, 2016. Constitutive translation of human  $\alpha$ -synuclein is mediated by the 5'-untranslated region. *Open Biology* [online]. **6**(4) [cit. 2022-03-11]. ISSN 2046-2441. Dostupné z: doi:10.1098/rsob.160022
- [96] CRENSHAW, Ezekiel, Brian P. LEUNG, Chun Kit KWOK, et al., 2015. Amyloid Precursor Protein Translation Is Regulated by a 3'UTR Guanine Quadruplex. *PLOS ONE* [online]. **10**(11) [cit. 2022-03-11]. ISSN 1932-6203. Dostupné z: doi:10.1371/journal.pone.0143160
- [97] LAMMICH, Sven, Frits KAMP, Judith WAGNER, Brigitte NUSCHER, Sonja ZILOW, Ann-Katrin LUDWIG, Michael WILLEM a Christian HAASS, 2011. Translational Repression of the Disintegrin and Metalloprotease ADAM10 by a Stable G-quadruplex Secondary Structure in Its 5'-Untranslated Region. *Journal of Biological Chemistry* [online]. **286**(52), 45063-45072 [cit. 2022-03-11]. ISSN 00219258. Dostupné z: doi:10.1074/jbc.M111.296921
- [98] SINGER, Oded, Robert a MARR, Edward ROCKENSTEIN, Leslie CREWS, Nicole G COUFAL, Fred H GAGE, Inder M VERMA a Eliezer MASLIAH, 2005. Targeting BACE1 with siRNAs ameliorates Alzheimer disease neuropathology in a transgenic model. *Nature Neuroscience* [online]. **8**(10), 1343-1349 [cit. 2022-03-11]. ISSN 1097-6256. Dostupné z: doi:10.1038/nn1531
- [99] CHAUDHARY, Swati, Mahima KAUSHIK, Saami AHMED, Ritushree KUKRETI a Shrikant KUKRETI, 2018. Structural Switch from Hairpin to Duplex/Antiparallel G-Quadruplex at Single-Nucleotide Polymorphism (SNP) Site of Human Apolipoprotein E ( APOE ) Gene Coding Region. *ACS Omega* [online]. **3**(3), 3173-3182 [cit. 2022-03-11]. ISSN 2470-1343. Dostupné z: doi:10.1021/acsomega.7b01654
- [100] *Fylogenetická analýza* [online]. [cit. 2022-03-31]. Dostupné z: [https://is.muni.cz/www/98951/41610771/43823411/43823458/Analýza\\_genomick/55450383/Fylogenetika150227.pdf](https://is.muni.cz/www/98951/41610771/43823411/43823458/Analýza_genomick/55450383/Fylogenetika150227.pdf)
- [101] MOHANTY, Jyotirmayee, Nilotpal BAROOAH, V. DHAMODHARAN, S. HARIKRISHNA, P. I. PRADEEPKUMAR a Achikanath C. BHASIKUTTAN, 2013. Thioflavin T as an Efficient Inducer and Selective Fluorescent Sensor for the Human Telomeric G-Quadruplex DNA. *Journal of the American Chemical Society* [online]. **135**(1), 367-376 [cit. 2022-01-31]. ISSN 0002-7863. Dostupné z: doi:10.1021/ja309588h
- [102] RENAUD DE LA FAVERIE, Amandine, Aurore GUÉDIN, Amina BEDRAT, Liliya A. YATSUNYK a Jean-Louis MERGNY, 2014. Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Research* [online]. **42**(8), e65-e65 [cit. 2022-01-31]. ISSN 1362-4962. Dostupné z: doi:10.1093/nar/gku111
- [103] KHURANA, Ritu, Chris COLEMAN, Cristian IONESCU-ZANETTI, Sue A. CARTER, Vinay KRISHNA, Rajesh K. GROVER, Raja ROY a Shashi SINGH, 2005. Mechanism of thioflavin T binding to amyloid fibrils. *Journal of Structural Biology* [online]. **151**(3), 229-238 [cit. 2022-01-31]. ISSN 10478477. Dostupné z: doi:10.1016/j.jsb.2005.06.006
- [104] BHATTACHARYYA, Debmalya, Gayan MIRIHANA ARACHCHILAGE a Soumitra BASU, 2016. Metal Cations in G-Quadruplex Folding and Stability. *Frontiers in Chemistry* [online]. **4** [cit. 2022-01-31]. ISSN 2296-2646. Dostupné z: doi:10.3389/fchem.2016.00038

- [105] Reverse Complement. *Bioinformatics* [online]. [cit. 2022-05-17]. Dostupné z: [https://www.bioinformatics.org/sms/rev\\_comp.html](https://www.bioinformatics.org/sms/rev_comp.html)
- [106] BARTAS, Martin, Václav BRÁZDA, Natália BOHÁLOVÁ, et al., 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. *Frontiers in Microbiology* [online]. **11** [cit. 2022-04-01]. ISSN 1664-302X. Dostupné z: doi:10.3389/fmicb.2020.01583
- [107] Phylogenetic tree (newick) viewer, 2016. *ETE Toolkit* [online]. *MOI Biol Evol* [cit. 2022-04-01]. Dostupné z: doi:10.1093/molbev/msw046
- [108] WILKIE, Gavin S., Kirsten S. DICKSON a Nicola K. GRAY, 2003. Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends in Biochemical Sciences* [online]. **28**(4), 182-188 [cit. 2022-03-14]. ISSN 09680004. Dostupné z: doi:10.1016/S0968-0004(03)00051-3
- [109] CHATTERJEE, Sangeeta a Jayanta K. PAL, 2009. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biology of the Cell* [online]. **101**(5), 251-262 [cit. 2022-03-14]. ISSN 02484900. Dostupné z: doi:10.1042/BC20080104
- [110] ZERBINI, Francisco Murilo, Fábio Nascimento da SILVA, Gloria Patricia Castillo URQUIZA a Marcos Fernando BASSO, 2014. Transgenic Plants. *Biotechnology and Plant Breeding* [online]. Elsevier, 2014, **2014**, 179-199 [cit. 2022-03-14]. ISBN 9780124186729. Dostupné z: doi:10.1016/B978-0-12-418672-9.00008-8
- [111] Unipro UGENE. *A unified bioinformatics toolkit* [software]. Konstantin Okonechnikov, Olga Golosova, Mikhail Fursov, the UGENE team. *Bioinformatics* 2012 **28**: 1166-1167 [cit. 2022-22-04]. Dostupné z: doi:10.1093/bioinformatics/bts091
- [112] Fowler's Zoo and Wild Animal Medicine, 2015. *Fowler's Zoo and Wild Animal Medicine*. 8. University of Missouri, Columbia, Missouri, 743 - 773. ISBN 978-1-4557-7397-8.