

G-KVADRUPLEXY A JEJICH ROLE V REGULACI TRANSKRIPCE

ERIKA KUŽMOVÁ

Ústav organické chemie a biochemie AV ČR, v. v. i.,
Flemingovo náměstí 542/2, 166 10 Praha 6
erika.kuzmova@uochb.cas.cz

Došlo 14.11.18, přijato 25.2.19.

Klíčová slova: G-kvadruplex, transkripce, DNA, RNA, hybridní DNA:RNA G-kvadruplex, sekundární struktura

Obsah

1. Úvod
2. Co jsou G-kvadruplexy
3. Kde se G-kvadruplexy vyskytují
4. G-kvadruplexy a regulace transkripce
 - 4.1. G-kvadruplexy jako fyzická překážka pro RNA-polymerasu
 - 4.2. G-kvadruplexy jako vazebná místa pro transkripční faktory
 - 4.3. DNA G-kvadruplexy jako senzory oxidativního stresu
 - 4.4. DNA G-kvadruplexy a nukleosomy
 - 4.5. Rozplétání G-kvadruplexů v genomu
 - 4.6. Hybridní G-kvadruplexy
5. Závěr

1. Úvod

DNA se v živých buňkách nachází převážně ve formě pravotočivé dvoušroubovice, avšak některé sekvence DNA mohou za určitých podmínek vytvářet nekanonické struktury, tzv. „non-B-DNA“. Mezi tyto struktury patří například levotočivá Z-DNA, triplexy, H-DNA, křížové struktury, R-smyčky, vlásenky, G-kvadruplexy a i-motivy¹. V současné době jsou tyto struktury velice dobře definovány v podmínkách *in vitro*². Díky stále se zlepšujícím metodickým přístupům vznikla možnost studovat tyto nekanonické struktury DNA i v endogenním prostředí a potvrdit tak jejich existenci a funkci *in vivo*^{3–7}. Takto byla zjištěna spojitost „non-B-DNA“ struktur s určitými typy rakoviny a jinými lidskými chorobami^{8–10}.

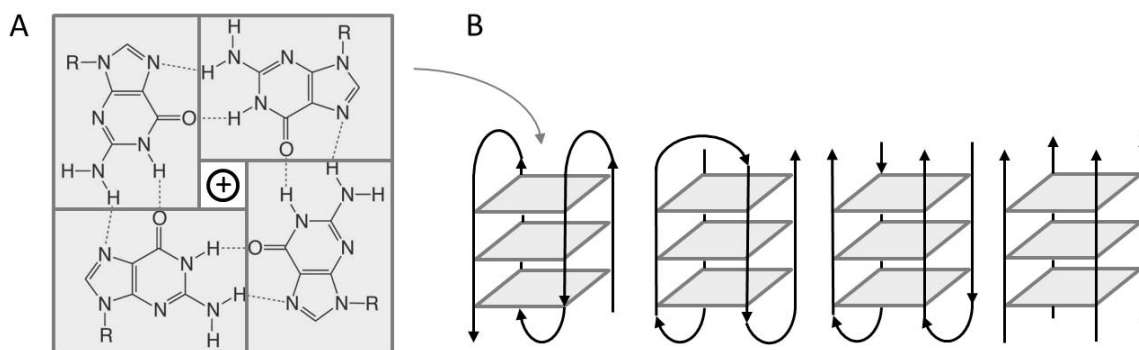
DNA v buňce je velmi dynamická molekula a životnost různých lokálních „non-B-DNA“ struktur je mnohdy pouze přechodná, což znemožňuje jejich studium a uspo-

kojivou interpretaci výsledků. Výskyt „non-B-DNA“ struktur je vázaný na nespárované nukleotidy a specifickou sekvenci, přičemž tvorba jedné struktury často znemožňuje tvorbu jiné¹¹. Mnohé lokální sekundární struktury DNA jsou stabilizovány pomocí nadšroubovicového vinutí (tzv. negativní superhelicita), které vzniká také během transkripce, při které RNA-polymerasa rozplétá dvoušroubovici DNA a způsobuje její pnutí. Vzniklé pnutí je částečně uvolněno právě vznikem sekundárních struktur¹². Bioinformatické analýzy ukázaly, že distribuce těchto sekundárních struktur v genomu není náhodná. Tyto struktury často kolokalizují s regulačními oblastmi genů, s chromosomálními zlomy a translokacemi a jsou asociovány s genetickými poruchami^{1,8,13–15}. Sekundární struktury DNA proto tvoří atraktivní terapeutický cíl, ačkoliv otázka potenciální selektivity doposud není dostatečně objasněna. Na druhé straně tvorba nekanonických sekundárních struktur DNA zejména v oblasti promotorů onkogenů a genů podílejících se na opravách DNA dynamicky sleduje aktivitu genu, sehrává důležitou roli v komplexní transkripční regulaci těchto genů a podílí se tak na udržení stability genomu^{1,12,14}. Jednou z nejvíce studovaných sekundárních struktur nukleových kyselin jsou G-kvadruplexy (G4).

2. Co jsou G-kvadruplexy

G4 jsou sekundární struktury nukleových kyselin, které se tvoří v oblastech bohatých na guaniny (G) (cit.¹⁶). Guaniny na základě Hoogsteenova párování bází tvoří guaninové tetrády, které se na sebe vrství a jsou stabilizovány monovalentními kationty v následujícím pořadí $K^+ > Na^+ > NH_4^+ > Li^+$ (cit.¹⁷). G4 se tvoří buď z jednoho vlákna nukleové kyseliny, nebo i z více (2–4) samostatných vláken (obr. 1). K tvorbě G4 může docházet na DNA, RNA, ale také na jejich kombinaci. Pak hovoříme o tzv. hybridním DNA:RNA G4 (HQ) (cit.¹⁸). Vzhledem k absenci kompetujícího komplementárního vlákna je sekvence RNA bohatá na guaniny mnohem náchylnější k tvorbě G4 struktur než DNA (cit.¹⁹).

Strukturní polymorfismus G4 kontrastuje s typickou dvoušroubovicí DNA. G4 existují v mnoha různých uspořádáních v důsledku variabilní délky smyček, orientace jednotlivých vláken a samotné sekvence tvořící G4 (cit.^{20,21}). Struktura a stabilita G4 závisí na mnoha faktorech, jako je například teplota, pH, koncentrace solí, aktivita vody, hustota roztoku, ve kterém se G4 nachází, délka smyček či postranní sekvence²². Změnou některé z podmínek se může výrazně změnit uspořádání G4 (cit.²²). G4 jsou *in vitro* velmi stabilní struktury, přičemž obecně lze tvrdit, že G4 s kratšími smyčkami (≤ 4 nukleotidy) a s větším počtem guaninových tetrad jsou stabilnější²³.



Obr. 1. Guaninová tetráda a příklady G-kvadruplexu tvořeného jedním nebo více vlákní nukleových kyselin. A – guaninová tetráda vytvořená na základě Hoogsteenova párování bází stabilizovaná monovalentním kationtem, B – vrstvení guaninových tetrád a příklady G-kvadruplexů z jednoho, dvou a čtyř vláken nukleové kyseliny. Přepřacováno podle¹⁵

RNA G4 jsou stabilnější než DNA G4 (cit.¹⁹). HQ vykazují také velmi vysokou stabilitu a odolnost vůči RNAsam¹⁸.

V biologickém kontextu se na stabilizaci G4 významně podílí i nadšroubovicové vinutí a některé proteiny včetně helikas, které budou diskutovány dále v textu. O stabilitě G4 v různých biologických kontextech se však ví velmi málo. Experimentálně bylo dokázáno, že paralelní G4 představují pro DNA-polymerasu větší překážku než dynamické G4 struktury, které balancují mezi paralelní a antiparalelní konfigurací¹⁴. G4 s kratšími smyčkami korelují s vyšší úrovní genomové nestability²⁴. Experimenty s vysoce stabilním HQ a nestabilním vnitromolekulárním DNA G4 té samé sekvence naznačují, že kompetice a konverze mezi jednotlivými typy G4 může mít regulační funkci, jako je např. terminace transkripce²⁵. Pro detailnější obeznámení se s problematikou G4 doporučují některý z přehledových článků^{15,17,26,27}.

3. Kde se G-kvadruplexy vyskytují

Jedním z nejdůležitějších nástrojů studia výskytu G4 jsou bioinformatika a *in silico* analýzy. Otázka existence G4 *in vivo* a jejich role v biologických procesech zůstávala dlouhou dobu pouze předmětem diskusí. Systematické prohledávání genomů různých organismů ukázalo, že sekvence, které mohou potenciálně tvořit G4 (pG4s) se nacházejí jak u prokaryot, tak u všech eukaryot^{28–35}, a to i v mitochondriální DNA (cit.^{36,37}), ale také u nebuněčných virů³⁸. pG4s však výrazně převládají v eukaryotických genomech, a to zejména u savců, u kterých jsou pG4s koncentrovány v regulačních oblastech promotorů³⁵. Tyto oblasti jsou vysoce konzervovány mezi druhy, což indikuje pozitivní selekční tlak pro zachování G4 sekvencí³⁹. Naopak pokud vezmeme v úvahu i hybridní G4 s polovičním počtem G tripletů potřebných pro tvorbu G4, sekvence, které mohou potenciálně tvořit HQ (pHQs) se začaly hromadit v genomech obojživelníků a byly dále pozitivně selektovány v genech teplotekrevných živočichů

především na netemplátovém (kódujícím) vlákně v oblastech za transkripčním počátkem⁴⁰. Problematice HQ a pHQs se podrobněji věnuje kapitola 4.6.

Nejpoužívanějším modelem pro zjištění, zda daná sekvence potenciálně tvoří intramolekulární G4, se stal zjednodušený sekvenční motiv $G_m X_n G_m X_o G_m X_p G_m$, kde m je počet guaninů a nabývá hodnot 3 až 5 a X_n , X_o a X_p označují smyčku, kterou může tvořit kterákoliv báze včetně guaninu. Délka smyčky nabývá hodnot 1–7, respektive až 12 (cit.⁴¹). Na základě tohoto modelu dvě nezávislé bioinformatické studie lidského genomu popsaly přes 350 000 G4 motivů o délce smyček 1–7 a až přes 700 000 o délce smyček 1–12 (cit.^{41,42}). Obě studie poukazují na signifikantní a nenáhodné rozložení pG4s v genomu. Nicméně skupina okolo Shankara Balasubramaniana metodou G4-ChIP Seq založenou na detekci sekundárních struktur DNA pomocí G4 specifické protilátky a následovanou vysoce výkonným sekvenováním („antibody-based G4 chromatin immunoprecipitation and high-throughput sequencing“) potvrdila pouze 10 000 G4 v lidském genomu, a to zejména v regulačních a aktivně prepisovaných oblastech chromatinu s absencí nukleosomů⁵. Výše popsaný kanonický vzorec pro pG4s pokryl v uvedené studii jen 21 % detegovaných G4-ChIP Seq vrcholů. Rozšíření modelu o sekvence s delšími smyčkami a o sekvence zahrnující nepárový nukleotid v tetrádě guaninů tvořící G4 pravděpodobně lépe vystihuje biologicky relevantní G4 (cit.⁴³).

Nejvyšší výskyt pG4s je na telomerech a v oblastech spojených s regulací replikace, transkripce a translace^{41,42}. pG4s jsou koncentrovány především v promotorových oblastech onkogenů a genů podílejících se na opravách DNA⁴⁴. Naopak v oblastech tumor supresorových a provozních (tzv. „house keeping“) genů jsou zastoupeny jen velmi málo^{35,46}. Toto zjištění logicky propojilo G4 s karcinogenezí a se snahou o přípravu ligandů stabilizujících G4 struktury. pG4s jsou časté i na hranici mezi introny a exony, a to především v oblasti prvního exonu, v oblastech genomové nestability, zejména v blízkosti chromosomálních translokací, mutací, variabilních počtů

kopii segmentů DNA (tzv. „somatic copy number variations“), v oblastech nadšroubovicového stresu a v oblastech R-smyček^{35,41,45,47}. Zvýšená koncentrace pG4s byla zaznamenána také na 5' a 3' nepřekládaných koncích mRNA (5'-UTR a 3'-UTR), kde by mohly G4 hrát úlohu při sestřihu pre-mRNA nebo při regulaci translace⁴⁸. Lokalizace pG4s zejména v důležitých genomových oblastech podporuje předpoklady o regulační a stabilizační roli G4 v genomu.

Tvorba G4 je mimo jiné stimulována negativním nadšroubovicovým vinutím, zvýšením hustoty makromolekul v okolí G4 (tzv. „molecular crowding“) a specifickými proteiny vázajícími se na G4 (cit.^{46,49}). Výskyt G4 se zvyšuje v průběhu S-fáze^{3,9}. Bylo identifikováno množství helikas, které jsou zapojeny do rozplétání G4 (cit.⁵⁰). Vzhledem k bohatému zastoupení a regulační roli G4 v genomu mají poruchy jejich tvorby i rozplétání patologický potenciál a G4 představují atraktivní terapeutický cíl. Avšak žádný z dosavadních G4 stabilizujících ligandů se zatím nedostal do posledních fází klinického testování.

4. G-kvadruplexy a regulace transkripce

První důkaz o vlivu G4 na transkripci genů pochází z prací na genu *c-Myc* (cit.⁵¹) a ze studií zkoumajících ovlivnění transkripce jiných genů pomocí známých ligandů stabilizujících G4, zejména onkogenů obsahujících pG4s ve svém promotoru^{10,51}. Předpokládán byl inhibiční vliv stabilizace G4 na expresi daného genu^{10,51,52}. Regulace transkripce *c-Myc* genu přes G4 je však mnohem komplexnější než se původně předpokládalo a je pravděpodobně výsledkem koordinované regulace více sekundárních struktur na promotoru a enhanceru *c-Myc* genu, které se dynamicky mění a odráží tak stav a míru transkripce⁵³.

Výskyt pG4s v genomu předurčuje G4 ke globální regulaci genové exprese. Dnes je role G4 jako řídicích transkripčních prvků obecně uznávána a objevují se stále nové mechanismy, jimiž G4 regulují transkripci^{14,18,53}. Ukazuje se, že G4 mohou ovlivňovat transkripci pozitivně i negativně, a to v závislosti na kontextu a vlákně, na kterém se nacházejí⁴⁵. Také mohou být součástí vrozené regulace transkripce, která je podmíněná sekvencí¹⁸, či naopak epigenetickou značkou sledující oxidativní stres a transkripci v závislosti na typu genu¹⁴. Množí se experimentální důkazy o tom, že G4 jsou úzce spjaty také s transkripčními R-smyčkami a s terminací transkripce¹². V následujícím textu budou zmíněny alespoň některé z uvedených regulací.

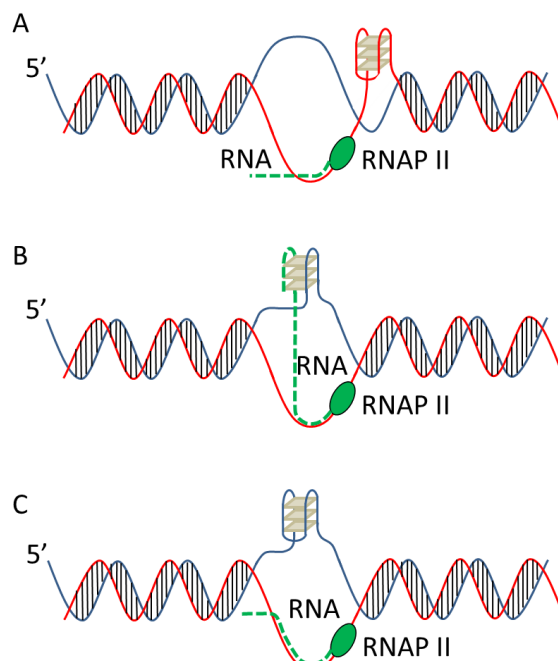
4.1. G-kvadruplexy jako fyzická překážka pro RNA-polymerasu

Původně se předpokládalo, že při transkripci slouží G4 jako fyzická překážka pro RNA-polymerasu^{10,53}. Vznik a stabilizace G4 brání postupu RNA-polymerasy a způsobí její pozastavení nebo rozpadnutí celého transkripčního komplexu^{10,51}. Existuje několik mechanismů vzniku tran-

skripčního bloku. Inhibiční G4 se tvoří buď na templátovém nebo na netemplátovém vlákně⁵². Tvorba G4 na netemplátovém vlákně zabráňuje opětovnému vzniku duplexu DNA, udržuje jednovláknovou DNA a umožňuje tak nascentnímu vlákně RNA hybridizovat s templátovou DNA. Nascentní vlákno RNA může vytvořit i HQ s netemplátovým vláknem^{18,53}. Není také vyloučeno, že nascentní RNA transkript tvoří RNA G4, a dále že kompetice mezi jednotlivými G4 (DNA, RNA a HQ) tvoří důležitou součást regulace transkripce, sestřihu pre-mRNA či translace. Příklad vzniku uvedených typů G4 je vyobrazen na obr. 2.

Většina výše popsaných G4 vzniká v transkripční „bublině“ a způsobuje transkripční blok zabráňující dalšímu kolu transkripce^{40,53}. Avšak bližší analýza lidských genů obsahujících G4 motivy v blízkosti transkripčních počátků ukázala, že více než polovina genů obsahuje alespoň jeden G4 v rozmezí 500 nukleotidů za transkripčním počátkem na netemplátovém vlákně a toto zjištění naopak koreluje se signifikantně zvýšenou expresí genů³². Jedno z možných mechanistických vysvětlení předpokládá, že tvorba G4 na netemplátovém vlákně udržuje templátové vlákno v jednořetězcové podobě a naopak tím usnadňuje následnou transkripci³².

V poslední době se ukazuje, že u aktivně transkribovaných genů pG4s úzce souvisejí s transkripčně podmíně-



Obr. 2. Možné mechanismy transkripčního bloku pomocí G-kvadruplexu. A – vytvoření G-kvadruplexu na templátovém vlákně, B – vytvoření hybridního DNA:RNA G-kvadruplexu na netemplátovém vlákně spolu s nascentní RNA, C – vytvoření G-kvadruplexu na netemplátovém vlákně. 5' konec označuje netemplátové vlákno. RNA-polymerasa II (RNAP II). Přepracováno podle⁵²

ným vznikem R-smyček, s GC vláknovou asymetrií (tzv. „GC skew“), CpG ostrůvky a nemethylovaným stavem těchto oblastí^{54,55}.

Pozitivní „GC skew“ (nadbytek G nad C na netemplátovém vlákně) za transkripčním počátkem a nízký obsah pG4s v promotorových oblastech je charakteristický pro konstitutivně přepisované provozní (tzv. „house keeping“) geny³⁵. Naopak výrazný negativní „GC skew“ (nadbytek G nad C v templátovém vlákně) v oblasti promotorů je charakteristický například pro geny související s transkripční regulací, jako jsou mnohé protoonkogeny (*c-Myc*, *c-Kit*, *VEGF*, *KRAS*), které mají zdokumentovaný regulační G4 ve svém promotoru^{10,54}. Pozitivní „GC skew“ se vyskytuje i na 3' a 5' koncích genů⁵⁴. Ukázalo se, že takto ohraničené geny korelují se zvýšenou hustotou genů na chromosomu, kde hrozí přechod transkripčního komplexu do dalšího genu a jen potvrzuje souvislost těchto oblastí s potřebou přesné a účinné regulace iniciace a terminace transkripce⁵⁴. Za připomenutí stojí, že pG4s se často nacházejí právě na 3'-UTR a 5'-UTR koncích mRNA a tedy na netemplátovém (kódujícím) vlákně DNA, kde pravděpodobně výše popsanými mechanismy přispívají k přesné regulaci iniciace a terminace transkripce⁴⁸.

4.2. G-kvadruplexy jako vazebná místa pro transkripční faktory

Další představa fungování G4 v transkripci je podpořena výskytem proteinů, které se přednostně a s vysokou afinitou vážou na G4 strukturu, stabilizují ji, nebo svou vazbou napomáhají jejímu vytvoření⁵⁶. Pro ilustraci budou zmíněny jen některé proteiny, i když jsou identifikovány stále nové proteiny, které se vážou na G4 a dá se předpokládat, že tato skupina proteinů se bude dále rozšiřovat.

Dobře známým proteinem vázajícím se s vysokou afinitou na DNA G4 je nukleolin¹⁰. Nukleolin sehrává důležitou roli při maturaci ribosomální RNA a u B-lymfocytů, kde slouží jako protein vázající RNA. Jeho nadměrná produkce a přemístění je častým znakem onkologických onemocnění⁵⁷. Interakce nukleolinu s DNA G4 je však nejednoznačná. V případě vazby na promotorový G4 lidského genu *VEGF* působí jako transkripční aktivátor. Vazba na nedávno identifikovaný LTR promotor lidského HIV-1 viru slouží jako molekulární „chaperon“ umožňující vytvoření DNA G4 (cit.⁵⁸). V neposlední řadě vazba nukleolinu na sekvenci G4 v promotoru *c-Myc* genu vede k represí transkripce⁵³, což vedlo ke snaze o vytvoření specifických ligandů cílících na G4 mnohých protoonkogenů.

Naopak příkladem typického transkripčního aktivátoru, vázajícího se s vysokou afinitou ke G4, je lidský SP1. Jde o transkripční faktor se strukturálním motivem zinkového prstu. Původně se předpokládalo, že se váže na dvouvláknovou DNA s minimálním motivem 5'-GGGCGG-3', avšak celogenomový ChIP-Seq experiment ukázal, že 36 % SP1 vazebných míst vůbec neobsahuje SP1 konsenzální motiv, nýbrž jeden nebo více G4 motivů⁵⁹.

Příkladem dalšího transkripčního aktivátoru je MAZ

protein rovněž s motivem zinkového prstu. MAZ se váže na promotor *KRAS* genu, který obsahuje podobně jako *c-Myc* ve svém nukleasovém hypersenzitivním místě G4. Vazba mezi MAZ a G4 vede k aktivaci transkripce, zatímco mutace znemožňující tvorbu G4 vede k inhibici exprese genu *KRAS*⁶⁰.

Způsob regulace uvedených genů závisí nejen na sekvenci DNA, ale také na její konformaci a dalších faktorech, které sekundární konformaci DNA ovlivňují. To naznačuje existenci strukturně specifických transkripčních trans-aktivátorů^{12,53}.

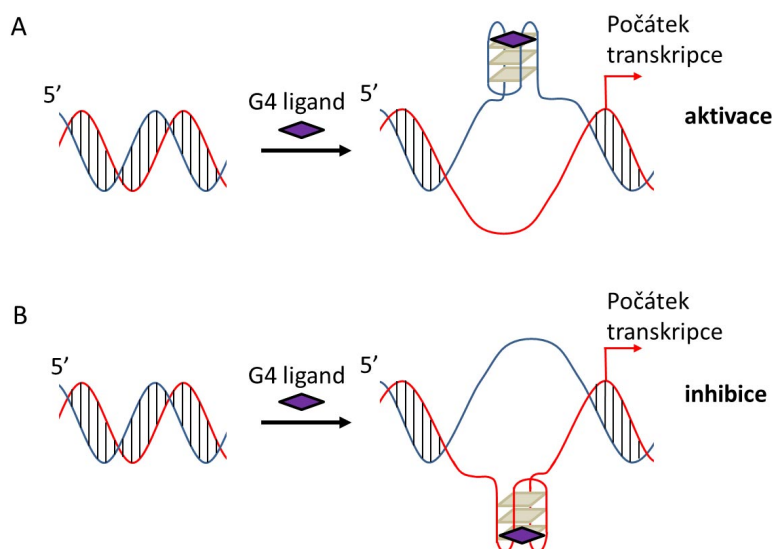
4.3. DNA G-kvadruplexy jako senzory oxidativního stresu

V posledních dvou letech ukázala skupina okolo Cyn-tie Burrowsové v několika desítkách článků, že G4 sekvence hrají důležitou roli v reakci na oxidativní stres^{14,45,61,62}. Sekvence bohaté na G jsou obecně náchylné k oxidaci. Velmi snadno zde dochází ke vzniku 8-oxo-7,8-dihydroguaninu (OG), který se nedokáže párovat pomocí Hoogsteenových vodíkových vazeb. Ukazuje se, že OG reprezentuje epigenetickou modifikaci sloužící jako regulační prvek v genomu⁶². Nedávná bioinformatická analýza ukázala zvýšenou koncentraci pG4s v promotorové oblasti lidských genů podílejících se na opravách DNA, přičemž jejich zastoupení na templátovém a netemplátovém vlákně bylo zhruba stejné⁴⁴. Mnohé geny obsahují sekvence minimálně 4 a více tripletů G. Při modifikaci konkrétních pozic G dochází k destabilizaci a přestavbě původního G4 a využití některého z dalších tripletů G jako tzv. „náhradní pneumatiky“ pro tvorbu nového G4. Přestavba G4 vede k „vystavení“ vzniklé modifikace ve smyčce nově vytvořeného G4, což rekrutuje příslušné proteiny, které ji rozeznávají, a dochází ke změně úrovně transkripce^{14,62}.

Transkripční vzorec také závisí na typu genu a jeho roli v oxidativním stresu⁶¹. Experimenty založené na luciferasovém expresním systému s různou pozicí G4 sekvence v regulační oblasti luciferasového genu, na templátovém či netemplátovém vlákně, a se známými G4 ligandy pyridostatinem (PDS), Phen-DC3 a BRACO-19 potvrzují aktivaci nebo inhibici transkripce vzhledem k pozici G4 na vlákně. U studovaných genů, jejichž produkty se podílejí na opravách DNA (BLM, MGMT, NEIL1 a NEIL3) stabilizace G4 na templátovém vlákně vedla k inhibici exprese luciferasy, zatímco stabilizace na netemplátovém vlákně vedla k její stimulaci⁴⁵ (obr. 3).

4.4. DNA G-kvadruplexy a nukleosomy

Oblasti bez nukleosomů (tzv. „nucleosome-free regions“ – NFRs) jsou charakteristické pro promotory a zesilovače transkripce (enhancery a superenhancery) aktivně transkribovaných genů¹². NFRs vykazují hypersenzitivitu k DNase I, dále je pro ně charakteristický vysoký obsah GC bází, CpG ostrůvků⁶³, sekvencních motivů pro transkripční faktory, ale i pG4s (cit.⁴⁶). Nedávné mapování výskytu G4 v lidských keratinocytech ukázalo, že až 98 %



Obr. 3. Vliv pozice G-kvaruplexu na transkripci. A – umístění na netemplátovém vlákně (značené od 5' konce) vede k aktivaci transkripce, B – umístění na templátovém vlákně vede k inhibici transkripce. Přepracováno podle⁴⁴

signálu pro G4 se nachází v NFRs (cit.⁵). Tato místa vysoce koreluje s aktivní transkripcí a zastavením RNAP II, což je v souladu s výše popsanou teorií, že G4 hrají důležitou roli při regulaci transkripce. Ukazuje se, že NFRs jsou primárně určeny sekvencí DNA a jen do jisté míry transkripční aktivitou⁶³. Oblasti bohaté na GC a pravděpodobně i tvorba nekanonických sekundárních struktur DNA přirozeně brání vzniku nukleosomů a dále se podílejí na udržení rozvolněné struktury chromatinu⁶³. U některých druhů kvasinek promotorové oblasti bohaté na poly-G zabraňují v daných místech tvorbě nukleosomů⁶⁴. Nedávné mapování „non-B DNA“ sekundárních struktur pomocí vysokokapacitního sekvenování poukazuje na vysoký výskyt sekundárních struktur DNA obecně v myších a lidských buňkách. Ty jsou asociované s časoprostorovou funkcí genů, buněčným stavem a s potenciálem přestavět pozici nukleosomů¹. Přesný mechanismus, jakým je pozice nukleosomů regulovaná, není zatím objasněn.

4.5. Rozplétání G-kvadruplexů v genomu

ChIP-Seq experimenty s G4 specifickými protilátkami ukazují, že množství G4 motivů predikovaných *in silico*, které se ve skutečnosti vyskytuje v G4 konformaci *in vivo*, je minimální (cit.^{5,55}). To naznačuje, že k tvorbě G4 jsou zapotřebí i jiné faktory, nejen samotná sekvence nukleové kyseliny. Bez ohledu na biologickou funkci G4 je zřejmé, že vzhledem k jejich důležitosti v regulaci buněčných procesů musí být jejich tvorba a rozplétání účinné a kontrolované a že deregulace může vést k poruchám stability genomu a nemocem⁵⁰. Dá se proto předpokládat, že všechny proteiny, které ovlivňují rozvolnění dvoušroubovice DNA na jednovláknovou DNA a následně tvorbu G4 struktury, budou mít zásadní vliv na transkripci.

Jednou ze skupin proteinů úzce souvisejících s transkripcí jsou helikasy BLM a WRN, které se účastní rozplétání G4. Mutace v těchto helikásách vedou k vážným onemocněním, jako jsou Bloomův syndrom a Wernerův syndrom (tzv. adultní progerie)⁶⁵. Bylo experimentálně dokázáno, že v případě inaktivace těchto dvou helikas se snížila exprese genů majících více G4 v promotorových oblastech a prvním intronu a naopak zvýšila exprese genů majících méně G4 motivů⁶⁶. Tento trend se neprojevil u buněk pacientů s Rothmundovým-Thomsonovým syndromem, který je způsoben defektem v příbuzné helikase, která ale neobsahuje RQC doménu potřebnou pro interakci s G4 (cit.⁶⁶). Homologní helikasa k BLM RecQ u kvasinek je helikasa Sgs1. Delece Sgs1 vede také k selektivnímu snížení exprese genů s G4 motivem v promotorové oblasti³⁴.

Nejlépe dokumentovaný vliv G4 na transkripci pochází ze studií *c-Myc* protoonkogenu. Zatím však byly pro *c-Myc* G4 popsány pouze tři proteiny, které hrají roli v jeho rozplétání. Těmito proteiny jsou CNBP, PARP-1 a helikasa Pif1 (cit.⁵⁰). Bohužel detailní analýzy těchto proteinů chybí. Pif1 je evolučně konzervovaná helikasa, která preferenčně rozvolňuje RNA:DNA substráty a zároveň aktivně rozplétá i G4 (cit.^{67,68}). Předpokládá se i role Pif1 při rozplétání HQ v mitochondriích¹⁸.

Dalšími důležitými G4 helikásami v procesu transkripce jsou helikasy XPD a XPB, přičemž až 40 % jejich vazebných míst se překrývá s pG4s, a to zejména v promotorech aktivně přepisovaných genů⁶⁹. Zajímavou helikásou je RNA helikasa DHX36, která rozplétá DNA G4 nacházející se v promotoru transkripčního faktoru YY1, ale již nerozplétá stejný RNA G4 motiv na 5'-UTR konci přepsané mRNA⁷⁰. Mechanismus této selektivity není zatím zcela objasněn. Za zmínku dále stojí helikasa DHX9, která se

zdá být dobrým kandidátem na rozplétání již zmíněných HQ a RNA G4, které mimo jiné mohou přispívat také k regulaci terminace transkripce (viz níže). Zajímavostí je, že tato helikasa rozplétá RNA G4 téměř dvakrát rychleji než DNA G4 (cit.⁷¹).

4.6. Hybridní G-kvadruplexy

Většina dosavadních studií se soustředila na DNA G4 tvořené převážně čtyřmi a více tripletami G na jedné molekule nukleové kyseliny. Ve fyziologicky relevantních procesech byl však jejich výskyt studován jen zřídka. Mnohem častěji může v buňce docházet ke vzniku HQ, jenž je tvořený polovinou konvenční G4 sekvence, a to minimálně dvěma tripletami G za sebou^{40,72,73}. pHQs odpovídá motivu $G_{\geq 3}(N_{1-7}G_{\geq 3})_{\geq 1}$, ve kterém G je guanin a N je jakýkoliv nukleotid včetně G. Tato sekvence může na netemplátovém vlákne tvořit HQ s nascentní RNA. Bylo zjištěno, že pHQs se vyskytují až v 97 % lidských genů, přičemž v průměru se v každém genu vyskytuje až 73 pHQs (cit.⁷²). Experimenty Zheng Tana prokázaly, že HQ jsou schopny přímo inhibovat transkripci jak v podmínkách *in vitro*, tak *in vivo*²⁵. Jde zjevně o regulační vliv charakteristický pro transkripci, protože zahrnuje transkribovanou DNA, RNA a RNA-polymerasu bez účasti dalších faktorů. Regulace je daná sekvencí samotného genu a uskutečňuje se pouze, pokud se gen transkribuje a vzniká nascentní RNA. Jedná se tedy o jednoduchou, přímou a ekonomickou autologní kontrolu genové exprese⁷². Celogenomová analýza ukázala, že regulace transkripce přes HQ není omezená na specifické geny či dráhy, nýbrž jde o obecný řídicí *cis*-element na základní úrovni transkripce, který limituje expresní potenciál daného genu⁴⁰. Výše zmíněnou roli v regulaci transkripce potvrzuje i fakt, že pHQs se přednostně vyskytují na netemplátovém vlákne a jsou mnohem více zastoupené v transkribovaných oblastech, zejména do 1000 nt od transkripčního počátku⁴⁰.

Předčasná terminace transkripce pomocí tvorby HQ byla dokumentována u mitochondrií¹⁸. Vlákno nascentní RNA nejdříve hybridizuje s templátovým vlákem a vytvoří R-smyčku. Při dalším kole transkripce je vlákno RNA uvolněno z R-smyčky a vzniká ssRNA, která se ochotně váže s netemplátovou DNA obsahující G triplety za tvorby stabilního HQ (cit.⁷³). Obdobný mechanismus by mohl fungovat i v genomové DNA, kde takto vzniklý HQ pozastaví RNAP II, nebo způsobí terminaci transkripce (obr. 2). K tvorbě HQ dochází také na telomerách, kde zatím jejich přesná úloha není známa⁷⁴.

5. Závěr

Existence G-kvadruplexů v podmínkách *in vivo* byla dlouhou dobu předmětem diskusí. Dnes jsou G-kvadruplexy uznávanou součástí regulačního systému genomu a je známo, že jejich tvorba a rozplétání podléhá striktní kontrole. Deregulace tvorby či rozplétání G4 obec-

ně, ale i G4-závislé transkripce, je patologická, a proto G4 představují atraktivní terapeutický cíl. Bioinformatické analýzy predikují velké množství sekvencí s potenciálem tvořit G4, což je však v kontrastu se skutečně detegovanými G4. Informace o tom, kdy, proč, s jakou frekvencí a které G4 se tvoří v biologickém kontextu, jsou zatím jen útržkovité a mnohé otázky zůstávají nezodpovězené.

V poslední době se ukazuje, že tvorba G4 je komplexnější, než se původně myslelo a bylo nutné proto definovat méně striktní pravidla pro potenciální sekvence tvořící G4. Rozšíření potenciální „G-kvadruplexovosti“ by mohlo přinést nový pohled na staré výsledky. Množí se také důkazy o tom, že G4 úzce souvisí i s ostatními sekundárními „non-B DNA“ strukturami, které hrají důležitou roli v regulaci transkripce. Nedávná studie lokusu lidského *c-Myc* ukazuje, že sekundární struktury DNA dynamicky sledují aktivitu genu. Zvyšující se transkripční aktivita genu vede k postupné formaci odlišných sekundárních struktur, na které se vážou další transkripční faktory a v závislosti na kontextu ovlivňují zpětně transkripci¹². Nadšroubovicové vinutí se ukazuje jako jedna z nejdůležitějších hybných sil vedoucích ke vzniku „non-B DNA“ struktur.

Základnější úroveň regulace transkripčního potenciálu genu představují kratší sekvence G tripletů. Ty jsou schopné tvořit hybridní DNA:RNA G-kvadruplex s nascentní RNA a vést k terminaci transkripce. Tato *cis*-regulace je závislá pouze na samotné transkripci. V posledních dvou letech se nahromadily i důkazy o G4 jako senzorech oxidativního stresu, jenž vede k ovlivnění transkripce příslušného genu. Není vyloučena ani dynamická transkripční regulace pomocí různých forem a typů G4. Jediná sekvence DNA s potenciálem formovat G4 skýtá prostor pro množství různých topologií a konformací a zajisté se tak můžeme těšit na nové spojitosti a poznatky v oblasti regulace transkripce pomocí G4.

Vypracováno s finanční podporou Ministerstva školství, mládeže a tělovýchovy České republiky v rámci Národního programu udržitelnosti (NPU), Projekt InterBio-Med LO1302.

Seznam zkratk

G	guanin
G4	G-kvadruplex
pG4s	potenciální G4 tvořící sekvence
HQ	hybridní DNA:RNA G-kvadruplex
pHQs	potenciální HQ tvořící sekvence
5'-UTR	5' nepřekládaný konec (five prime untranslated region)
3'-UTR	3' nepřekládaný konec (three prime untranslated region)
NFRs	oblasti bez nukleosomů (nucleosome-free regions)
RNAP II	RNA-polymerasa II

LITERATURA

- Kouzine F., Wojtowicz D., Baranello L., Yamane A., Nelson S., Resch W., Kieffer-Kwon K., Benham C. J., Casellas R., Przytycka T. M., Levens D.: *Cell Syst.* 4, 344 (2017).
- Kypr J., Kejnovská I., Renčičuk D., Vorlíčková M.: *Nucleic Acids Res.* 37, 1713 (2009).
- Biffi G., Tannahill D., McCafferty J., Balasubramanian S.: *Nat. Chem.* 5, 182 (2013).
- Chambers V. S., Marsico G., Boutell J. M., Di Antonio M., Smith G. P., Balasubramanian S.: *Nat. Biotechnol.* 33, 877 (2015).
- Hänsel-Hertsch R., Beraldi D., Lensing S. V., Marsico G., Zyner K., Parry A., Di Antonio M., Pike J., Kimura H., Narita M., Tannahill D., Balasubramanian S.: *Nat. Genet.* 48, 1267 (2016).
- Xu Y., Komiya M.: *Chem. Bio. Chem.* 14, 927 (2013).
- Zeraati M., Langley D. B., Schofield P., Moye A. L., Rouet R., Hughes W. E., Bryan T. M., Dinger M. E., Christ D.: *Nat. Chem.* 10, 631 (2018).
- Simone R., Fratta P., Neidle S., Parkinson G. N., Isaacs A. I.: *FEBS Lett.* 589, 1653 (2015).
- Murat P., Balasubramanian S.: *Curr. Opin. Genet. Dev.* 25, 22 (2014).
- Balasubramanian S., Hurley L. H., Neidle S.: *Nat. Rev. Drug Discov.* 10, 261 (2011).
- Dhokal S., Yu Z., Konik R., Cui Y., Koirala D., Mao H.: *Biophys. J.* 102, 2575 (2012).
- Zaytseva O., Quinn L. M.: *BioEssays* 40, 1 (2018).
- Marnef A., Cohen S., Legube G.: *J. Mol. Biol.* 429, 1277 (2017).
- Omega C. A., Fleming A. M., Burrows C. J.: *Biochemistry* 57, 2958 (2018).
- Rhodes D., Lipps H. J.: *Nucleic Acids Res.* 43, 8627 (2015).
- Gellert M., Lipsett M. N., Davies D. R.: *Proc. Natl. Acad. Sci. U.S.A.* 48, 2013 (1962).
- Davis J. T.: *Angew. Chem., Int. Ed.* 43, 668 (2004).
- Wanrooij P. H., Uhler J. P., Shi Y., Westerlund F., Falkenberg M., Gustafsson C. M.: *Nucleic Acids Res.* 40, 10334 (2012).
- Zaccaria F., Guerra C. F.: *Chem. – Eur. J.* 24, 16315 (2018).
- Burge S., Parkinson G. N., Hazel P., Todd A. K., Neidle S.: *Nucleic Acids Res.* 34, 5402 (2006).
- Webba da Silva M.: *Chem. – Eur. J.* 13, 9738 (2007).
- Lane A. N., Chaires J. B., Gray R. D., Trent J. O.: *Nucleic Acids Res.* 36, 5482 (2008).
- Bugaut A., Balasubramanian S.: *Biochemistry* 47, 689 (2008).
- Piazza A., Adrian M., Samazan F., Heddi B., Hamon F., Serero A., Lopes J., Teulade-Fichou M. P., Phan A. T., Nicolas A.: *EMBO J.* 34, 1718 (2015).
- Zheng K. W., Wu R. Y., He Y. D., Xiao S., Zhang J. Y., Liu J. Q., Hao Y. H., Tan Z.: *Nucleic Acids Res.* 42, 10832 (2014).
- Dolinnaya N. G., Ogloblina A. M., Yakubovskaya M. G.: *Biochemistry (Moscow)* 81, 1602 (2016).
- Bochman M. L., Paeschke K., Zakian V. A.: *Nat. Rev. Genet.* 13, 770 (2012).
- Rawal P., Bhadra V., Kummasetti R., Kumar N., Halder K., Sharma R., Mukerji M., Das S. K., Chowdhury S.: *Genome Res.* 16, 644 (2006).
- Yadav V. K., Abraham J. K., Mani P., Kulshrestha R., Chowdhury S.: *Nucleic Acids Res.* 36, 381 (2008).
- Yadav V., Kim N., Tuteja N., Yadav P.: *Front. Plant Sci.* 8, 1163 (2017). doi: 10.3389/fpls.2017.01163.
- Capra J. A., Paeschke K., Singh M., Zakian V. A.: *PLoS Comput. Biol.* 6, 1 (2010), doi:10.1371/journal.pcbi.1000861.
- Du Z., Zhao Y., Li N.: *Nucleic Acids Res.* 37, 6784 (2009).
- Griffin B. D., Bass H. W.: *Plant Sci.* 269, 143 (2018).
- Hershman S. G., Chen Q., Lee J. Y., Kozak M. L., Yue P., Wang L. S., Johnson F. B.: *Nucleic Acids Res.* 36, 144 (2008).
- Huppert J. L., Balasubramanian S.: *Nucleic Acids Res.* 35, 406 (2007).
- Huang W. C. a 11 spoluautorů: *Nucleic Acids Res.* 43, 10102 (2015).
- Falabella M., Fernandez R., Johnson B., Kaufman B. A.: *Curr. Med. Chem.* 25, PMID: 29493440 (2018). doi: 10.2174/0929867325666180228165527.
- Ruggiero E., Richter S. N.: *Nucleic Acids Res.* 46, 3270 (2018).
- Frees S., Menendez C., Crum M., Bagga P. S.: *Hum. Genomics* 8, 1 (2014).
- Xiao S., Zhang J. Y., Zheng K. W., Hao Y. H., Tan Z.: *Nucleic Acids Res.* 41, 10379 (2013).
- Huppert J. L., Balasubramanian S.: *Nucleic Acids Res.* 33, 2908 (2005).
- Hubáľková P.: *Chem. Listy* 109, 918 (2015).
- Todd A. K., Johnston M., Neidle S.: *Nucleic Acids Res.* 33, 2901 (2005).
- Guédin A., Gros J., Alberti P., Mergny J. L.: *Nucleic Acids Res.* 38, 7858 (2010).
- Fleming A. M., Zhu J., Ding Y., Visser J. A., Zhu J., Burrows C. J.: *Biochemistry* 57, 991 (2018).
- Lipps H. J., Rhodes D.: *Trends Cell Biol.* 19, 414 (2009).
- De S., Michor F.: *Nat. Struct. Mol. Biol.* 18, 950 (2011).
- Agarwala P., Pandey S., Maiti S.: *Optoelectron. Adv. Mater. Rapid Commun.* 13, 5570 (2015).
- Zhou J., Tateishi-Karimata H., Mergny J. L., Cheng M., Feng Z., Miyoshi D., Sugimoto N., Li C.: *Biochimie* 121, 204 (2016).
- Sauer M., Paeschke K.: *Biochem. Soc. Trans.* 45, 1173 (2017).
- Siddiqui-Jain A., Grand C. L., Bearss D. J., Hurley L. H.: *Proc. Natl. Acad. Sci. U.S.A.* 99, 11593 (2002).
- Brown R. V., Danford F. L., Gokhale V., Hurley L. H., Brooks T. A.: *J. Biol. Chem.* 286, 41018 (2011).
- Kim N.: *Curr. Med. Chem.* 25, 1 (2017).

54. Ginno P. A., Lim Y. W., Lott P. L., Korf I., Chédin F.: *Genome Res.* 23, 1590 (2013).
55. Hänsel-Hertsch R., Di Antonio M., Balasubramanian S.: *Nat. Rev. Mol. Cell Biol.* 18, 279 (2017).
56. Brázda V., Hároníková L., Liao J., Fojta M.: *Int. J. Mol. Sci.* 15, 17493 (2014).
57. Berger C. M., Gaume X., Bouvet P.: *Biochimie* 113, 78 (2015).
58. Tosoni E., Frasson I., Scalabrin M., Perrone R., Butovskaya E., Nadai M., Palù G., Fabris D., Richter S. N.: *Nucleic Acids Res.* 43, 8884 (2015).
59. Raiber E., Kranaster R., Lam E., Nikan M., Balasubramanian S.: *Nucleic Acids Res.* 40, 1499 (2012).
60. Kerkour A., Marquevielle J., Ivashchenko S., Yatsunyk L. A., Mergnyet J. L.: *J. Biol. Chem.* 292, 8082 (2017).
61. Fleming A. M., Zhu J., Ding Y., Burrowse C. J.: *ACS Chem. Biol.* 12, 2417 (2017).
62. Fleming A. M., Ding Y., Burrows C. J.: *Proc. Natl. Acad. Sci. U.S.A.* 114, 2604 (2017).
63. Fenouil R. a 11 spoluautorů: *Genome Res.* 22, 2399 (2012).
64. Tsankov A., Yanagisawa Y., Rhind N., Regev A., Rando O. J.: *Genome Res.* 21, 1851 (2011).
65. Johnson J. E., Cao K., Ryvkin P., Wang L. S., Johnson F. B.: *Nucleic Acids Res.* 38, 1114 (2010).
66. Smestad J. A., Maher L. J.: *BMC Med. Genet.* 16, 1 (2015), DOI 10.1186/s12881-015-0236-4.
67. Boule J. B., Zakian V. A.: *Nucleic Acids Res.* 35, 5809 (2007).
68. Sanders C. M.: *Biochem. J.* 430, 119 (2010).
69. Gray L. T., Vallur A. C., Eddy J., Maizels N.: *Nat. Chem. Biol.* 10, 313 (2014).
70. Huang W. a 12 spoluautorů: *Nucleic Acids Res.* 40, 1033 (2012).
71. Lee T., Pelletier J.: *Oncotarget* 7, 42716 (2016).
72. Zheng K. W., Xiao S., Liu J. Q., Zhang J. Y., Hao Y. H., Tan Z.: *Nucleic Acids Res.* 41, 5533 (2013).
73. Zhang J. Y., Zheng K. W., Xiao S., Hao Y. H., Tan Z.: *J. Am. Chem. Soc.* 136, 1381 (2014).
74. Xu Y. Q., Liu Z. C.: *Stem Cell Rev.* 4, 101 (2008).

E. Kužmová (*Institute of Organic Chemistry and Biochemistry of the CAS, Prague*): **The Role of G-quadruplexes in Transcriptional Regulation**

G-quadruplexes belong to noncanonical secondary structures of the nucleic acids commonly referred to as non-B DNA. In recent years there is growing evidence about the existence of G-quadruplex structures *in vivo* and their important regulatory roles in biological processes. This mini-review brings up the latest results on the transcriptional regulation by G-quadruplexes. On one hand G-quadruplexes might inhibit general expression potential of the genes as a part of a basal transcription activity, which is not specific to any genes or pathways. On the other hand, G-quadruplexes might dynamically monitor the activity of a gene together with other secondary structures localized predominantly in regulatory regions such as promoters, transcription start sites and transcription termination sites. Last but not least, it has been shown that G-quadruplexes serve as sensors of oxidative stress and subsequently modulate the transcription. The inhibitory or stimulatory effect of the G-quadruplexes on a gene transcription depends on its position on the template or coding strand.

Keywords: G-quadruplex, transcription, DNA, RNA, hybrid DNA:RNA G-quadruplex, secondary structure

Acknowledgements

This work was supported by grants from the Ministry of Education, Youth and Sports of the Czech Republic (Project InterBioMed LO1302).