

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta
Katedra fyzikální chemie



Studium nekanonických konformačních stavů páteře v DNA a v komplexech proteinů s DNA

DIPLOMOVÁ PRÁCE

Autor: **Elizabeth Kolářová**
Studijní program: B1407 – Chemie
Studijní obor: Fyzikální chemie
Forma studia: Prezenční
Vedoucí práce: **doc. RNDr. Petr Jurečka, Ph.D.**
Rok: : 2024

Prohlašuji, že jsem bakalářskou práci vypracoval/a samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním bakalářské práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byl/a jsem seznámen/a s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne *podpis bakaláře*

Bibliografická identifikace

Jméno a příjmení autor	Elizabeth Kolářová
Název práce	Studium nekanonických konformačních stavů páteře v DNA a v komplexech proteinů s DNA
Typ práce	Diplomová práce
Pracoviště	Katedra fyzikální chemie
Vedoucí práce	doc. RNDr. Petr Jurečka, Ph.D..
Rok obhajoby práce	2024
Klíčová slova	DNA, protein, interakce, konformer, stabilita, rekonformace, potenciál, torze
Počet stran	76
Jazyk	Český (anglický)

Bibliographical identification

Autor's first name and surname	Elizabeth Kolářová
Title	Study of non-canonical backbone conformational substates in DNA and protein-DNA complexes
Type of thesis	Master thesis
Department	Department of physical chemistry
Supervisor	doc. RNDr. Petr Jurečka, Ph.D.
The year of presentation	2024
Keywords	DNA, protein, interaction, conformer, stability, reconfiguration, potential, torsion
Number of pages	76
Language	Czech (English)

Abstrakt

Tato diplomová práce se bude zabývat molekulově dynamickými simulacemi DNA a protein-DNA komplexů, které poskytují přínosné informace o stabilitě, energetice či dynamice s cílem hlouběji porozumět interakcím mezi těmito molekulami. Pro simulace je využívána klasická molekulová dynamika umožňující detailní analýzy. Práce je zaměřena na čtyři různé struktury samostatné DNA a dva protein-DNA komplexy, které obsahují nekanonické konformace s tím, že hlavní zaměření bude na nekanonické α/γ stavy cukr-fosfátové páteře. Bude se jednat o konformace BB02, BB03, BB12, BB13 ale také o nedefinované konformace NANT, u nichž bude snaha o jejich definování pomocí devíti torzních úhlů. Molekulově dynamické simulace byly prováděny v silových polích z rodiny AMBER, konkrétně pole bsc1, OL15 a nově vyvinuté silové pole OL21. Výsledky této práce poukazují na možný stabilní výskyt některých nekanonických α/γ stavů u neobvyklých struktur DNA a u komplexů protein-DNA ve funkčních segmentech. Nově definované konformační stavy mohou přispět k zpřesnění popisu nekanonických konformerů v krystalografických databázích.

Abstract

This thesis will look at molecular dynamic's simulations of DNA and protein-DNA complexes that provide useful information on stability, energetics or dynamics in order to gain a deeper understanding of the interactions between these molecules. Classical molecular dynamics is used for the simulations to allow detailed analyses. The work focuses on four different structures of single DNA and two protein-DNA complexes that contain non-canonical conformations, with the main focus on the non-canonical α/γ states of the sugar-phosphate backbone. This will include the BB02, BB03, BB12, BB13 conformations but also the unassigned NANT conformations for which an attempt will be made to define them using nine torsion angles. Molecular dynamics simulations were performed in force fields from the AMBER family, namely the bsc1, OL15 and the newly developed OL21 force field. The results of this work point to the possible stable occurrence of some non-canonical α/γ states in unusual DNA structures and in protein-DNA complexes in functional segments. The newly defined conformational states may contribute to the refinement of the description of non-canonical conformers in crystallographic databases.

Obsah

1	ÚVOD	9
2	STRUKTURA PROTEINU A DNA.....	11
2.1	Struktura DNA	11
2.1.1	DNA konformery	12
2.2	Struktura proteinu	13
3	KONFORMACE DNA	15
3.1	GAA hairpin, 1PQT	15
3.2	Levotočivý G-kvadruplex, 4U5M.....	17
3.3	Interkalovaný duplex d(GCGAAAGC), 1UE2	18
3.4	GCA hairpin, 1ZHU.....	19
4	KOMPLEXY PROTEIN-DNA.....	21
4.1	DNA glykosyláza, 2F5N	22
4.2	DNA Endonukleáza, 2GIG	23
5	MOLEKULOVÁ DYNAMIKA (MD)	25
5.1	Klasická molekulová dynamika	25
5.2	Silové pole.....	26
6	CÍLE PRÁCE	29
7	METODY	31
7.1	Parametrizace silového pole	32
7.2	Příprava výchozí struktury	34
8	VÝSLEDKY A DISKUZE	36
8.1	RMSD	36
8.1.1	RMSD GAA hairpinu, 1PQT	36
8.1.2	RMSD GCA hairpinu, 1ZHU.....	37
8.1.3	RMSD G-kvadruplexu, 4U5M	38
8.1.4	RMSD interkalovaného duplexu, 1UE2	39
8.1.5	RMSD DNA glykosylázy, 2F5N	40
8.1.6	RMSD DNA endonukleázy, 2GIG	41
8.2	NOVÉ KONFORMAČNÍ STAVY	44
8.2.1	Konformační stavy GAA hairpinu, 1PQT	45
8.2.2	Konformační stavy GCA hairpinu, 1ZHU.....	48
8.2.3	Konformační stavy G-kvadruplexu, 4U5M	50
8.2.4	Konformační stavy interkalovaného duplexu, 1UE2	53
8.2.5	Konformační stavy DNA glykosylázy, 2F5N	56

8.2.6	Konformační stavy DNA endonukleázy, 2GIG	59
9	SHRNUTÍ.....	62
10	ZÁVĚR.....	67
11	SUMMARY	69
12	REFERENCE	71

1 ÚVOD

V rámci této práce se budu soustředit na teoretické simulace DNA a protein-DNA komplexů, jež umožňují porozumět interakcím mezi těmito molekulami. Tyto teoretické simulace poskytují hodnotné informace o stabilitě, energetice či dynamice různých stavů, a přispívají tak k lepšímu pochopení biologických mechanismů.

Práce bude využívat klasickou molekulovou dynamikou, která představuje v dnešní době jeden z klíčových nástrojů, jak detailně analyzovat simulace pohybu atomů a molekul v čase na atomární úrovni. Využívá se v široké škále různých vědních disciplín jako jsou fyzika, chemie, biologie, farmakologie či bioinformatika. Poskytuje hlubší porozumění jednotlivých interakcí mezi samotnými atomy, celkové dynamiky a funkce různých zkoumaných systémů.

Hlavním úkolem bude za pomoci empirických potenciálů určit stabilitu vybraných nekanonických konformerů cukrfosfátové páteře DNA a zároveň zpřesnit jejich popis. Konformerem se myslí označení různých prostorových uspořádání molekuly. Pro výběr a porovnání konformerů byla použita databáze DNATCO^[1], která obsahuje 96 klasifikovaných DNA/RNA konformerů, tzv. NtC (nucleotid conformers), které byly definovány na základě 12 geometrických parametrů. Každý konformer je identifikován čtyřpísmenným symbolem a reprezentován 9 torzními úhly. Pokud daný konformer není možné přiřadit ani k jedné z výše uvedených 96 konformací, pak spadají do „97. konformace“ označené jako NANT (unassigned conformations).

Mezi vybrané nekanonické konformace byly zahrnuty konformace obsahující nekanonické hodnoty páteřních úhlů α/γ . K takovýmto konformacím patří například BB12, BB13 nebo BB03 z databáze DNATCO, obsahující celkem 96 možných konformací vycházejících z experimentálně získaných dat. Dále se bude práce zabývat konformacemi označovaných jako NANT, které nejsou v databázi zařazeny do žádné ze zmíněných 96 konformací. Zde bude hlavní snahou popsat a klasifikovat nový konformační stav, do něhož by mohly tyto NANT konformace spadaly.

Simulace byly provedeny s využitím silových polí z rodiny AMBER, a to konkrétně silová pole bsc1, OL15 a OL21. Pole OL15 a OL21 byla vyvinuta v laboratoři na Katedře fyzikální chemie Univerzity Palackého v Olomouci (OL je zkratka pro Olomouc). Parametrizace jsou zaměřeny na dihedrální úhly, a využívají parametrizačního postupu zahrnujícího konformačně závislé solvatační efekty. Pole OL15^[2] je odvozeno od pole ff99 a bsc0^[3] a obsahuje modifikace ϵ/ζ , χ a β torzních potenciálů pro zpřesnění simulací DNA (i RNA). K poli

OL21^[4] byla navíc přidána modifikace parametrizace torzních úhlů α/γ (pouze pro simulace DNA), která vede ke zlepšení popisu nekanonických molekul DNA.

2 STRUKTURA PROTEINU A DNA

Proteiny a DNA jsou jedny ze základních konceptů v molekulární biologii a genetice. Hrají klíčovou roli v biologických procesech. Jejich struktura je úzce spjata s jejich funkcemi a schopnostmi vykonávat různé úkony v buňce. Porozumění jejich struktury je klíčové k vysvětlení mnoha biologických procesů na molekulární úrovni.

2.1 Struktura DNA

Zkratka DNA pochází od anglického slovního spojení deoxyribonucleic acid, tedy deoxyribonukleová kyselina. Základní stavební jednotkou DNA je nukleotid, který se skládá z pětiuhlíkatého cukru (2-deoxyribóza - aldopentóza) obsahující funkční hydroxylovou skupinu, dále z fosfátových zbytků, a v poslední řadě ze čtyř bází (adenin, guanin, cytosin a thymin). Tyto dusíkaté báze jsou spárovány v komplementární dvojici. Řetěžením nukleotidů propojených díky fosfodiesterové vazbě vzniká polynukleotidové vlákno. Dvě takováto polynukleotidová vlákna tvoří vůči sobě antiparalelní a komplementární strukturu zvanou DNA šroubovice. Struktura byla objevena a popsána v roce 1953 J. D. Watsonem a F. Crickem^[5], kteří za tento převratný objev získali o necelých deset let později Nobelovu cenu.

Na prvním uhlíku C5 sacharidu je glykosidickou vazbou připojena dusíkatá báze, která se komplementárně, tedy vždy purinová báze pojící se s pyrimidinovou bází, váže na další dusíkatou bázi za pomoci vodíkových můstků. Pyrimidinové báze jsou cytosin a thymin (uracil), do purinových bází patří guanin a adenin. Propojení cytosinu s guaninem zprostředkovávají tři vodíkové vazby a adenin s thyminem jsou propojeny přes dva vodíkové můstky. Tento párovací mechanismus zajišťuje při replikaci vytvoření dvou identických molekul DNA z jednoho vlákna původní matrice DNA.

Na posledním, tedy pátém uhlíku C5 sacharidu je esterovou vazbou navázán fosfát. Celá tato struktura, umožňuje tvorbu celkového dlouhého řetězce DNA nesoucího genetickou informaci živého organismu.^{[6][7]}

Struktura DNA je často klasifikována do tří tříd dle sekundární struktury či funkční charakteristiky. První třídou je pravotočivá kanonická B-DNA, v níž se vyskytují dva žlábký (velký a malý) a na jednu její otočku připadá 10,5 bází. V této formě DNA jsou báze uspořádány rovnoměrně podél osy spirály.^[8]

Druhou možností je opět pravotočivá forma zvaná A-DNA, která je velmi podobná klasické kanonické B-DNA, ale liší se tvarem, kdy jsou báze posunuty ven směrem od osy spirály. Na otočku zde připadá 11 bází a struktura A-DNA není tolik zakřivená jako B-DNA.

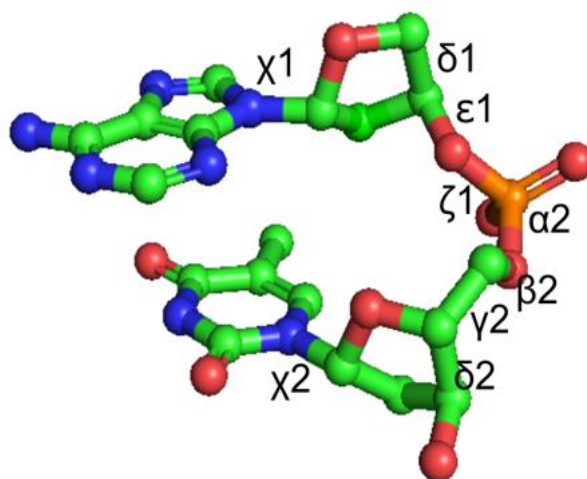
Třetí formou je levotočivá Z-DNA s 12 páry bazí na otočku, která má charakteristickou „cik-cak“ strukturu, kde se střídají purinové a k nim komplementární pyrimidinové báze. Jedním z největších rozdílů mezi těmito strukturami jsou jejich konformační stavy páteře. Forma B-DNA obsahuje kanonické konformační stavy páteře, avšak dojde-li ke změně konformace, dochází tak ke vzniku nekanonické konformace. U forem A- a Z-DNA nalezneme pouze nekanonické konformace.^[8]

2.1.1 DNA konformery

Jak již bylo výše v textu zmíněno, vybrané konformery DNA byly čerpány z databáze DNATCO, která se snaží klasifikovat geometrii páteře DNA prostřednictvím devíti torzních úhlů mezi sousedními nukleotidy (Obrázek 1). Tyto dihedrální úhly popisují prostorové uspořádání páteře DNA a jejich hodnoty ovlivňují schopnost tvorby DNA strukturních motivů.

Klasická kanonická B-DNA je v databázi zastoupena jako NtC konformační třída BB00, též je označovaná jako BI. Každá konformace se liší v daných dihedrálních úhlech, jež ji charakterizují. V práci se budou objevovat pouze nekanonické konformace obsahující α/γ rekonformace. V tabulce (Tabulka 1) jsou uvedeny vybrané konformace vyskytující se v této práci a porovnány s klasickou kanonickou B-DNA (NtC – BB00). Konformace BB02 je popsána jako BI s úhly α/γ rovny $g+/g-$ (*gauche+/gauche-*), BB03 je popsána obdobně jako BI, ale s α/γ rekonformací nabývající hodnot t/t (*trans/trans*), konformace BB12 a BB13 mají dihedrální úhel $\gamma = trans$ a velmi nízkou hodnotu úhlu β .^{[9], [10][11][12]}

Jedním z nejdiskutovanějších témat ohledně změn konformací je částečný přechod z B-DNA formy na formu A-DNA. Přechod je spojen s rekonformací deoxyribózy a glykosidického úhlu χ (chí). Změna konformace B-DNA ovlivňuje prostorovou orientaci a uspořádání dusíkatých bází. Tento přechod může být způsoben mnohými faktory, jako jsou změny v specifických sekvencích bází DNA, koncentrace iontů nebo interakce s proteinem.^{[13][14]}



Obrázek 1: Páteř DNA popsaná devíti torzními úhly.
Vytvořeno v PyMOL a upraveno v Inscape.

Tabulka 1 Vybrané NtC konformace [°], obsahující α/γ rekonformaci, v porovnání s kanonickou B-DNA. Nekanonické hodnoty jsou zvýrazněny barevně.

NtC	$\delta 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 1$	$\chi 2$
BB00	138	183	258	304	180	44	138	253	258
BB02	141	194	246	31	195	297	150	252	253
BB03	145	175	274	163	166	175	146	241	233
BB12	140	196	280	257	76	171	140	269	205
BB13	143	187	293	219	98	161	146	253	219

Existují i další (a častější) varianty konformací DNA, např. konformace BI (NtC BB00) a BII (NtC BB07). Konformace se liší v uspořádání ϵ a ζ úhlů. V konformaci BI je $\epsilon/\zeta = t/g$ - a v konformaci BII je $\epsilon/\zeta = g/t$. Varianta BI konformace je běžnější a stabilnější, varianta DNA s konformací BII se kromě menší populace v B-DNA dvoušroubovici dále objevuje v jednovláknové DNA, na konečných reziduí řetězců DNA, či v komplexech protein-DNA. Přechod mezi těmito konformacemi, dle některých nových studií, poukazuje na důležitý mechanismus při interakci protein-DNA. ^[15]

2.2 Struktura proteinu

Proteiny jsou biomakromolekulární látky jejichž základní stavební jednotkou jsou aminokyseliny (AK) jež jsou substitučními deriváty karboxylových kyselin. Existuje až 23 aminokyselin, které se mohou spojovat do různých kombinací a podílet se tak na tvorbě proteinů. Proteiny vznikají na základě kondenzace karboxylové skupiny AK spolu s aminovou

skupinou jiné AK za vzniku peptidické kovalentní vazby CO-NH-. Dle počtu AK vznikají buď oligopeptidy (s až 10 AK), polypeptidy (až 100 AK) nebo bílkoviny (s více než 100 AK).

Struktura proteinů se skládá z několika úrovní určujících výsledný tvar a funkci. Primární strukturou proteinů se myslí pořadí AK v jejich řetězci, které je kódováno nukleotidy v DNA, což proteinům udává jejich originální vlastnosti a funkce.

Sekundární struktura zahrnuje lokální geometrické uspořádání, kdy proteiny nabývají dvou základních struktur, jimiž jsou α -helixy (pravotočivá šroubovice) nebo β -sheet (skládaný list) - antiparalelní a paralelní. Tyto struktury jsou navzájem propojené za pomoci vodíkových můstků. Mimo to se řetězec může skládat z nerepetitivních struktur: otoček (turns), β -výdutí (bulge) či smyček (loops).

Terciární struktura je pak následné celkové, globální uspořádání výše zmíněných sekundárních struktur do určitého prostorového tvaru například fibrilárního (vlákenného) nebo globulárního (klubko) tvaru.

Poslední úrovní pro popis proteinů je kvartérní struktura, což je vzájemné prostorové uspořádání podjednotek, to se týká například hemoglobinu, který se skládá ze čtyř podjednotek. ^{[16], [17], [18]}

3 KONFORMACE DNA

Struktura DNA může existovat v mnoha konformacích, které ovlivňují její schopnost vázat se na proteiny či provádět nejrůznější biologické procesy v buňce. Konformace, v kterých se DNA může nacházet jsou ovlivňovány nejrůznějšími změnami podmínek. Různé strukturní variace polynukleotidového vlákna hrají roli například ve specifických interakcích. Interakce ovlivňují změny ve struktuře páteře DNA. ^[19]

Nejznámější variantou DNA je dvoušroubovice, nacházející se ve třech hlavních formách (A-, B- a Z-DNA). Avšak v buňkách se struktura DNA vyskytuje i v jiných nekanonických formách. Příkladem takovýchto struktur může být bulge (vydutina), hairpin (vlásenka), kvadruplex, triplex, Hollidayova struktura, R loop nebo i-motiv. Formy typu bulge, hairpin a R loop jsou převážně jednovláknové struktury, kvadruplex je speciální struktura, která se skládá z guaninových bází, k němuž je obdobná struktura i-motivu, která je bohatá na cytosinové báze. Triplex a Hollidayova struktura vznikají propojením více vláken DNA dohromady. Všechny vyjmenované struktury jsou zástupci nekanonických konformací, kterých může DNA nabývat. ^[20]

Molekula DNA má schopnost vytvářet různé nekanonické struktury díky svým unikátním chemickým vlastnostem, právě tyto nekanonické typy struktur se staly předmětem studování v této práci.

3.1 GAA hairpin, 1PQT

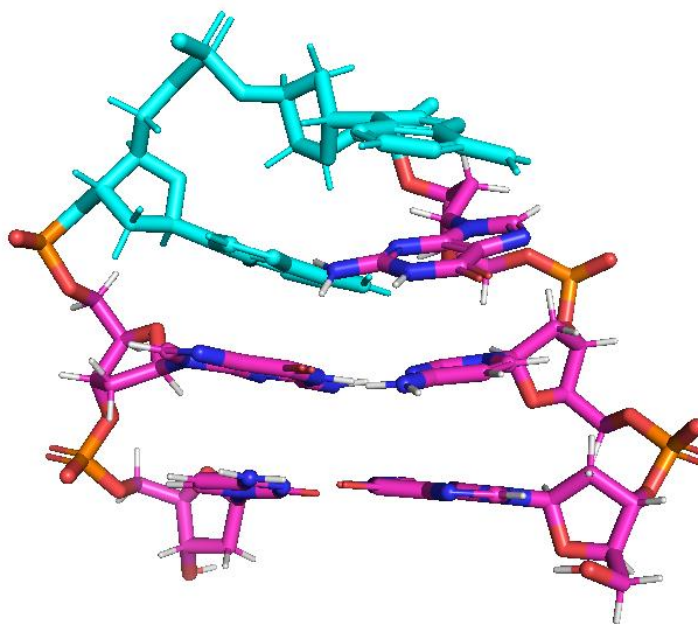
První vybraná struktura s PDB ID: 1PQT (Obrázek 2) byl hairpin (vlásenka), tedy jedno vláknenná struktura nukleové kyseliny s vnitro řetězcovým párováním bází. V tomto případě se jednalo o DNA s fragmentem d(GCGAAGC) obsahujícím sedm nukleotidů.

Experimentálně byla tato struktura získána pomocí NMR (Nuclear magnetic resonance) metody. Vlášeny mají významnou roli v expanzi tripletových repetice u nemocí jako jsou Huntingtonova choroba, Friedreichova ataxie či syndrom X.

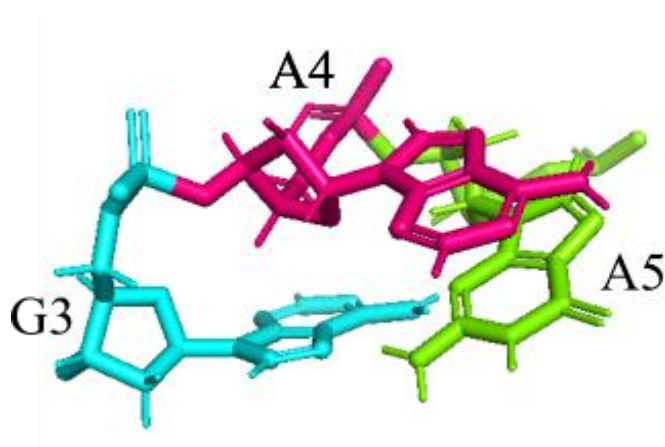
Ve fragmentu se nacházejí dva páry vázaného guaninu s cytosinem (G1-C7, C2-G6) a trinukleotidový loop (smyčka) (Obrázek 3) tvořený guaninem (G3) a dvěma adeniny (A4, A5). Samotný heptamer vykazuje několik zvláštních vlastností. Centrální triplet (GAA) a deoxyribóza nacházející se na páteři, mají vysoký vliv na stabilizaci celkové struktury vlášeny DNA ve vodných roztocích, jedná se tak o jeden z termodynamicky nejstabilnějších fragmentů, dokonce stabilnější než korespondující RNA hairpin r(GCGAAGC). Také je vysoce odolný vůči nukleázám a této vlastnosti se využívá při in vitro syntéze proteinů. ^{[21], [22], [23]}

Tato konkrétní sekvence se vyskytuje v mnoha důležitých biologických regionech. Tento konkrétní segment se našel například v replikačních počátcích viru herpes simplex, který způsobuje orofaryngeální infekce (v dutině ústní) nebo také promotorové části *E. coli*.^[21]

Struktura obsahovala v GAA triloopu nedefinovaný NANT konformer, který obsahoval nekanonické torzní úhly $\alpha/\gamma = g+/t$. Bude zde tedy snaha nasimulovat tento step, aby bylo dokázáno, zda se α/γ udrží v nativních hodnotách a zároveň tento konformer klasifikovat.



Obrázek 2: Struktura 1PQT GAA hairpinu, tyrkysová barva vyznačuje step s nekanonickou konformací. Vytvořeno v PyMOL.



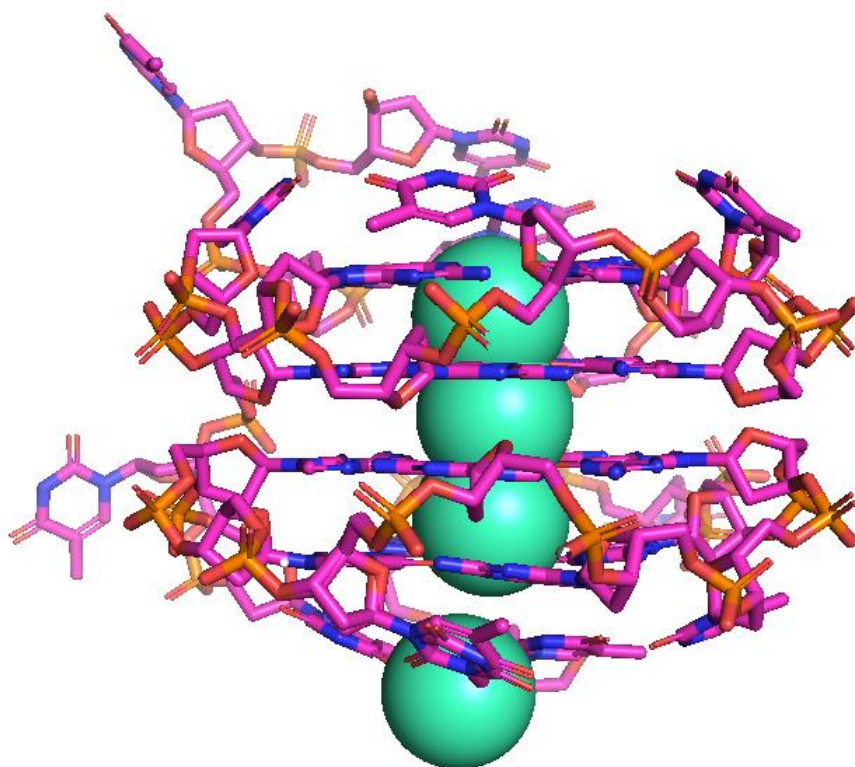
Obrázek 3: Struktura triloopu (GAA). Guanin tyrkysová, centrální adenin růžový a postranní adenin zelený. Vytvořeno v PyMOL.

3.2 Levotočivý G-kvadruplex, 4U5M

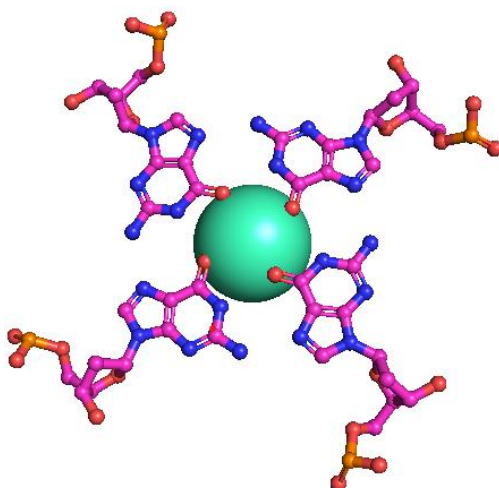
Další vybranou strukturou je struktura s PDB ID: 4U5M, tedy levotočivý G-kvadruplex DNA (Obrázek 4). Tato struktura byla získána metodou X-ray difrakce (rentgenová difrakce) s rozlišením 1,5 Å. G-kvadruplex je známá čtyřvláknová struktura, tento obsahuje 28 nukleotidů a čtyři draselné ionty. Doposud byly známy převážně pravotočivé formy kvadruplexů, avšak tato struktura je zajímavá právě tím, že je levotočivá. Celé struktuře přiřadil server DNATCO NANT konformace a předběžná analýza ukázala populace $\alpha/\beta = trans/high-trans$ stavů, proto byla tato struktura vybrána k podrobnějšímu zkoumání.

Guaniny jsou v tomto komplexu vázány Hoogsteenovým párováním v koplanárním uspořádání, jež jsou propojeny centrálním linkerem (T13-T14), umožňuje to tak vysokou polymorfnost struktury a přispívá to k jejím jedinečným vlastnostem. Sekvence G-kvadruplexu se nacházejí v mnoha lidských genomech a podílejí se na buněčných regulacích, replikaci, translaci či transkripci.

Ve struktuře se objevuje uspořádání head-to-head a tail-to-tail umožňující snazší přístup tyminových zbytků ze smyček. Díky těmto uspořádáním se v komplexu vyskytují čtyři draselné ionty zasunuté mezi G-tetradové vrstvy obdobně jako u pravotočivých ekvivalentů kvadruplexů (Obrázek 5). Potenciální využití takovýchto struktur je například v nanotechnologii. [24], [25]



Obrázek 4: Struktura levotočivého G-kvadruplexu 4U5M. Vytvořeno v PyMOL.



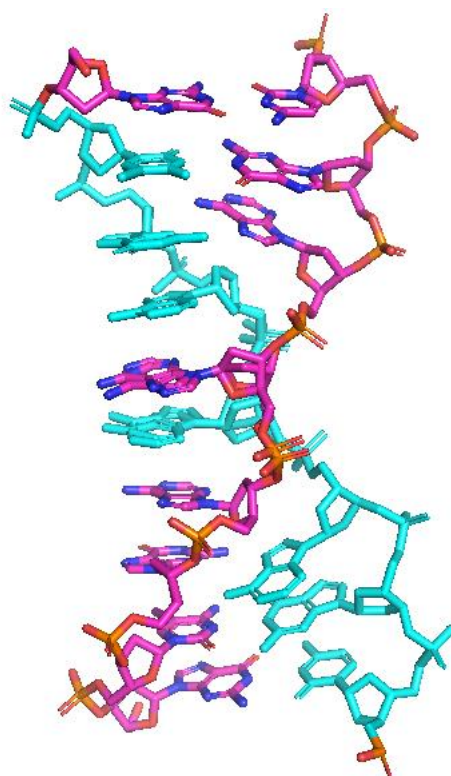
Obrázek 5: G-tetráda. Vytvořeno v PyMOL.

3.3 Interkalovaný duplex d(GCGAAAGC), 1UE2

Jako předposlední struktura samostatné DNA byla vybrána struktura interkalovaného duplexu s PDB ID: 1UE2 (Obrázek 6), složená ze dvou oktamerů, experimentálně změřena metodou X-ray difrakce. V tomto případě se jedná o zvláštní druh duplexu, tedy o dvojici bázově interkalovaných oktamerů, které jsou na obou koncích propojeny klasickým Watson-Crickovým párováním bází a uprostřed jsou báze interkalovány (zipper-like duplex).^[26]

Interkalací bází je obvykle myšleno vkládání malých organických molekul, jako například aromatických jader, mezi nukleové kyseliny, což dodává struktuře stabilitu. Uprostřed struktury 1UE2 se báze adeninu (A5) interkalují navzájem a interagují stackovou interakcí.^[27]

V páteři se v blízkosti interkalovaných bází vyskytují dvojí okupance, tedy dvě možné zastoupené konformace. Právě tato část páteře bude studována, protože obsahuje NANT konformace s nekanonickými hodnotami torzních úhlů $\alpha/\gamma = \text{high-trans/trans}$ a konformer BB12.

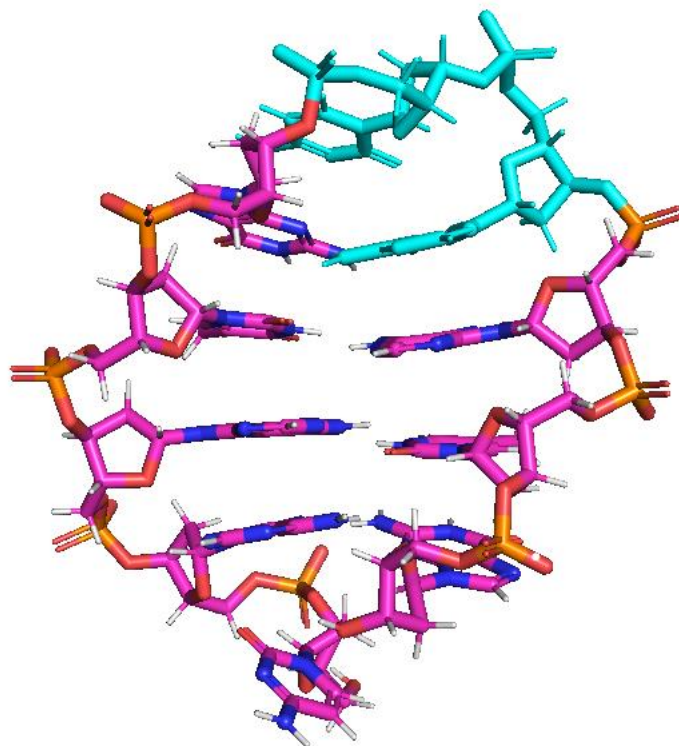


Obrázek 6: Struktura interkalovaného duplexu 1UE2. Tyrkysové jsou vyznačeny stepy obsahující nekanonické konformace. Vytvořeno v PyMOL.

3.4 GCA hairpin, 1ZHU

Poslední vybranou strukturou DNA je další hairpin s PDB ID: 1ZHU (Obrázek 7). Jedná se o jednocytidinovou vlásenkovou smyčku s tripletem GCA, který byl experimentálně získán metodou NMR. Vlásenková struktura GCA je složena z krátkého helixu o čtyřech párech bází s GA párem bází na obou jeho koncích a dvěma páry v centru propojenými klasickým Watson-Crickovým párováním. Na 3'-konci duplex je navíc nespárovaný cytidine. Triplet GCA vyskytující se v této struktuře je kodon kódující aminokyselinu s názvem alanin. Jako jeden triplet z mála může mutovat na své třetí pozici a stále bude kódovat alanin. Triplet GCA není schopen tvořit motiv $(GCA)_2$ na rozdíl od (GGA) nebo (GAA) , tedy vlastnost párovat se sám se sebou v antiparalelní orientaci a vytvářet tak například motiv $(GGA)_2$. Avšak, narozdíl od výše zmíněných, má schopnost tvorby neobvyklé vlásenkové struktury s pouze jedním reziduem ve smyčce. [28], [29]

Zde se nacházel jeden step obsahující $\alpha/\gamma = g-t$, nacházel se v loopu GCA, jednalo se zde tedy o obdobný případ jako u struktury 1PQT.



Obrázek 7: Struktura 1ZHU hairpin d(CAATGCAATG), tyrkysové jsou vyznačeny stepy s nekanonickými konformacemi. Vytvořeno v PyMOL.

4 KOMPLEXY PROTEIN-DNA

Interakce mezi proteiny a DNA představuje základní mechanismus regulace genetické exprese a mnoha dalších biologických procesů v buňkách. Tyto interakce jsou nezbytné k řízení transkripce, replikace či oprav DNA a dalších funkčních aspektů týkajících se genetické informace. Interakcemi strukturních motivů proteinů s DNA se myslí např. α -helixy, které upřednostňují interakci ve velkém žlábků DNA a β -sheet, jenž se nespecificky váže do malého žlábků.

Specifické interakce, mezi které můžeme zařadit např. vodíkové vazby, interakce aminokyselinových reziduí s bázemi v prohlubních DNA (malý a velký žlábek). Nejvýhodnější interakce probíhá mezi více vazebnými aminokyselinami. ^[30] Specifické interakce se dělí na dvě třídy, a to na sekvenčně specifickou vazbu a nepřímé čtení. Sekvenčně specifická vazba je založena na přesném rozpoznání báze DNA. Zde se protein váže na specifické sekvence nukleotidů DNA, patří sem například helix-turn-helix (interakce ve velkém žlábků). Naproti tomu se u nepřímého čtení jedná o interakci mezi páteří DNA a proteinem, jež už je na DNA vázaný. Protein také může nabývat sekvenční specifity. Jako příklad nepřímého čtení můžeme uvést interakce transkripčních faktorů, či enzymy opravující DNA. Nepřímé čtení je nerozlučně spjato s konformačním chováním páteře nukleových kyselin. ^{[31], [32]}

U nespecifický interakcí oproti specifickým nezáleží na tom, o jakou sekvenci se jedná, jsou tedy méně selektivní. Často se zde proteiny nacházejí v kombinaci s několika různými vazebnými místy DNA. Kladně nabitě aminokyseliny proteinu se váží na záporně nabitě fosfátové skupiny páteře DNA, jedná se tak o interakce elektrostatické. Ke slovu ale mohou přijít i interakce van der Waalsovy či hydrofobní. Nespecifické interakce přispívají k silnější vazebné afinitě proteinu na DNA. K takovým to interakcím patří například interakce histonů v chromatinu. Specifické i nespecifické interakce přispívají například k regulaci buněčných procesů. ^{[33], [34]}

Konformační chování cukr-fosfátové páteře ale také přispívá k interakcím proteinů s nukleovými kyselinami. Ve vzniklých komplexech je totiž páteř často více či méně deformována a může obsahovat i nekanonické konformační stavy. Protože nekanonické stavy mají obecně jinou energii než kanonické, musí být stabilita protein-DNA komplexu přítomností neobvyklých konformací páteře ovlivněna. Nejčastější konformací je částečný přechod z B-DNA směrem k A-DNA formě, ale může přecházet i na jiné méně časté

nekanonické formy. Protože cílem této práce je studovat nekanonické α/γ stavy, budu se zabývat také jejich možným příspěvkem ke stabilitě a struktuře komplexů proteinů s DNA. ^[35]

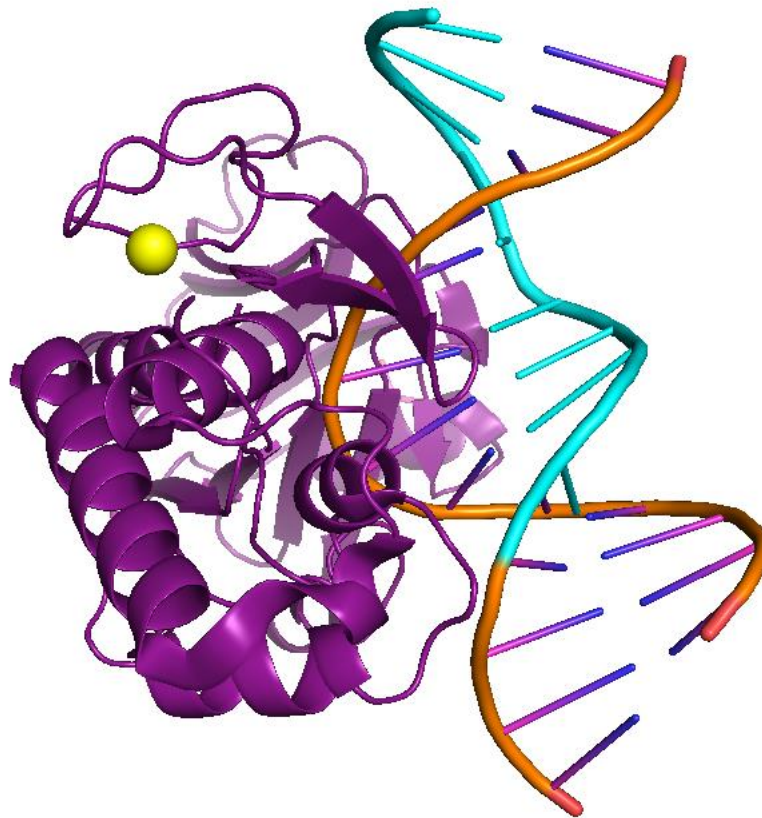
4.1 DNA glykosyláza, 2F5N

Struktura s PDB ID: 2F5N (Obrázek 8) je prvním zástupcem komplexů protein-DNA. Struktura byla získána experimentálně za pomoci X-ray difrakce a její rozlišení činí 2 Å. V komplexu se nachází jeden zinečnatý iont. Protein, jenž se v tomto komplexu nachází je enzym glykosyláza – formamidopyrimidine-DNA glykosyláza, jejímž úkolem je se aktivně podílet na opravě DNA, tím že prohledává páry bází DNA do té doby, dokud nenajde poškozenou nukleobázi a tu následně odstraní.

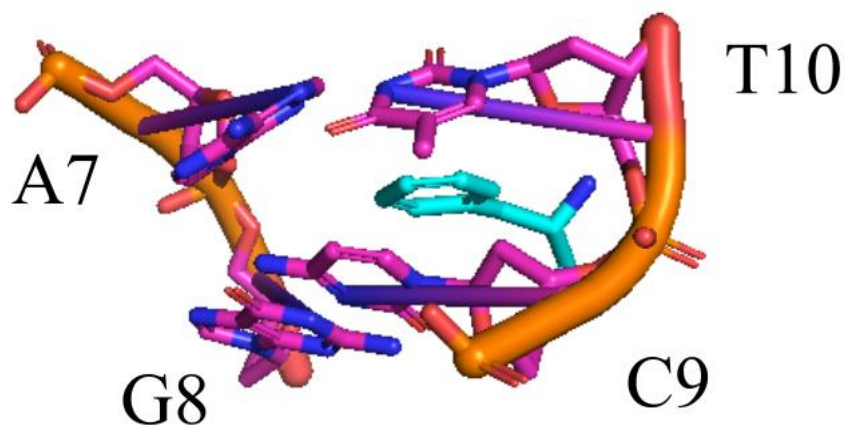
Výše zmíněnou nukleobází je adenin párovaný s mutagenním 7,8-dihydro-8-oxoguaninem (8-oxoG), který je velmi podobný klasické dusíkaté bázi guaninu. Mutagenní 8-oxoG je jedním z nejčastějších oxidativních poškození DNA. Vytváří transverzní mutaci G-C na T-A (pokud není opravena). V tomto případě se jedná o adenin a guanin z řetězce A a thymin s cytosinem z řetězce B, mezi něž je vmezeřený fenylalanin (Obrázek 9). ^[36]

Dalšími možnými způsoby poškození DNA jsou alkylace, depurinace či deaminace v prostředí z chemických reakcí do nichž má tendenci se DNA zapojovat. Glykosylázy tyto poškození opravují takovým způsobem, že danou poškozenou bázi odvrátí do apurinového či apyrimidinového aktivního místa (AP-místo) enzymu a zahájí specifické modifikace poškozené báze. ^[37]

V tomto komplexu se nacházelo hned několik stepů obsahující nekanonické konformace BB02, BB02, BB12, BB13 a nedefinované NANT. Zde bude hlavní snaha zjistit, zda se v komplexu tyto konformace s α/γ rekonformacemi udrží v průběhu MD simulací. Z toho jeden vybraný step obsahující nekanonickou konformaci je interkalovaný fenylalaninem, což by mohlo udržení se tohoto konformeru podpořit.



Obrázek 8: Komplex DNA glykosylázy 2F5N, tyrkysově vyznačené jsou nekanonické stepy, vytvořeno v PyMOL.



Obrázek 9: Interkalovaný fenylalanin mezi báze, vytvořeno v PyMOL.

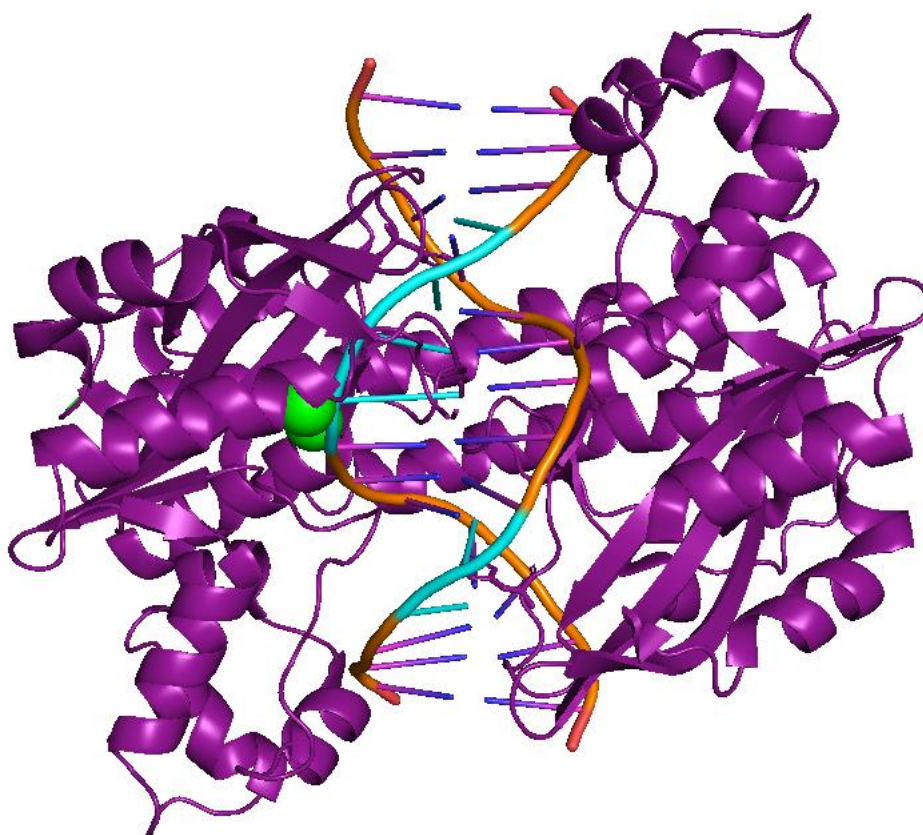
4.2 DNA Endonukleáza, 2GIG

Poslední vybranou strukturou obsahující nekanonické α/γ rekonformace je protein-DNA komplex s PDB ID: 2GIG (Obrázek 10). Jedná se o restriční endonukleázu typu II. HincII v komplexu s DNA, která byla experimentálně získána pomocí rentgenové difrakce s rozlišením 1,8 Å. Endonukleázy mají vysokou specifičnost, fungují na bázi vyhledávání

konkrétní sekvence v DNA, u nichž se následně naváže na molekulu DNA a provede řez v cukr-fosfátové páteři. Hlavní funkcí HincII je rozpoznat nepřímým čtením a vzápětí rozštěpit sekvenci GTYRAC, kde Y může být C nebo T a R buď to A nebo G báze. ^[38]

Struktura DNA je interkalovaná dvěma fenylalaniny na pozicích 138 v řetězcích proteinu A a B. Tyto interkalace zasahují právě do stepů s nekanonickými konformacemi v DNA, dalo by se tedy předpokládat, že mohou způsobit vynucení nekanonické konformace namísto konformace kanonické.

Komplex obsahuje dva sodné ionty, které se také nacházejí v místě dalšího nekanonického stepu, opět by se tedy dalo předpokládat, že by si mohly vynucovat konformační změny páteře DNA vedoucí k nekanonickým konformacím.



Obrázek 10: Komplex DNA endonukleázy 2GIG, kde jsou nekanonické stepy označeny tyrkysově. Vytvořeno pomocí PyMOL.

5 MOLEKULOVÁ DYNAMIKA (MD)

První molekulová simulace byla provedena již na začátku druhé poloviny 20. století, avšak její základy sahají do dřívější doby. V 19. století J. C. Maxwell a L. Boltzmann představili základní koncepty statistické mechaniky, roku 1905 A. Einstein poskytl pochopení Brownova pohybu, tedy náhodného pohybu částic a zároveň objasnil fotoelektrický jev vedoucí k základům kvantové mechaniky. V první polovině 20. století se vyvinula kvantová mechanika, ta vedla k lepšímu porozumění chování částic na mikroskopické úrovni.

V dnešní době má molekulová dynamika široké spektrum využití v různých oblastech vědních oborů, nejčastěji je využívána například v molekulární biologii, bioinformatice, farmaceutickém průmyslu či chemii. Lze díky ní studovat nově vyvinutá léčiva jejich interakce, studovat vlastnosti a chování materiálů na atomární úrovni, dynamiku biomolekul, pochopení chemických reakcí a mnoha dalších.

Molekulová dynamika se zabývá studiem pohybu atomů a molekul, a vývojem systému v určitém čase na základě obecného fyzikálního modelu meziatomových interakcí. Snahou MD je nasimulovat, zanalyzovat a předpovídat tak chování atomů nebo molekul za různých podmínek (teplota, tlak) a vyskytujících se v různých prostředích, které by je mohly nějakým způsobem, ať už pozitivně či negativně ovlivňovat, za pomoci výpočetních metod. MD simulace jsou využívány mnohdy v kombinaci s řadou experimentálních technik jakož jsou rentgenová krystalografie, kryoelektronová mikroskopie (kryoEM), nukleární magnetická rezonance (NMR) atd. ^{[39], [40]}

Existují různé typy modelů MD jako jsou například model tuhé koule s nespojitým potenciálem, klasická molekulová dynamika se spojitým potenciálem, metoda dráhového integrálu či kvantové simulace. ^[41]

5.1 Klasická molekulová dynamika

Pro systém na atomární či molekulární úrovni, kde si můžeme atomy představit jako jednotlivé body s určitou polohou a hybností. Interakce mezi částicemi jsou popsány empirickými mezimolekulovými potenciály, tedy silovými poli.

Klasický pohyb atomů v silovém poli popisují klasické pohybové rovnice dle druhého Newtonova zákona. Síla \vec{f}_i působící na částici je rovna změně její hybnosti v čase (Rovnice 1), kde $i = 1, \dots, N$. Tato rovnice může být také zapsána jako analog ke kvantové Schrodingerově rovnici - Liouvilleova rovnice (Rovnice 2), kde $[q, p]$ je vektor souřadnic a hybnosti a L je Liouvilleův operátor, kde $\Gamma = [q, p]$.

Rovnice 1: Druhý Newtonův zákon

$$\frac{d^2 \vec{r}}{dt^2} = \frac{\vec{f}_i}{m_i}$$

Rovnice 2: Liouvilleova formulace klasické mechaniky

$$\frac{i\partial\Gamma}{\partial t} = L\Gamma$$

Rovnice 1 se pak aplikuje s následnou integrací na každou částici v systému s tím, že m_i je hmotnost dané částice, r je její poloha jako funkce v čase t . K vyřešení těchto rovnic je potřeba znát počáteční polohy a rychlosti, následně bude rovnice řešena metodou konečných diferencí k získání výsledné trajektorie. Síly mezi částicemi jsou určeny například z empirických potenciálů, jakým může být např. Lennard-Jonesův potenciál (Rovnice 3).^[41]

Rovnice 3: Lennard-Jonesův potenciál

$$V(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right]$$

Kde ε je hloubka potenciálové jámy (jak moc jsou částice přitahovány), σ je vzdálenost při níž je mezimolekulový potenciál mezi částicemi nulový a r je vzdálenost mezi částicemi.

5.2 Silové pole

Pod pojmem silové pole (force field) v MD se myslí popis interakcí nastávajících v daném systému. Silové pole využívá potenciálových funkcí s jejichž parametry získává schopnost k výpočtu potenciální energie jakožto součet energií všech interakcí v molekule a interakcí s ostatními molekulami v systému. Silová pole využívaná v této práci patří do rodiny AMBER (Assisted Model Building with Energy Refinement).^[42]

Parametry pro modelovaný systém se získávají z obecné rovnice potenciální energie (E_P). Rovnice se skládá z pěti členů (Rovnice 4), dále se dělí na vazebné a nevazebné členy. Do vazebných členů patří kovalentní vazby, úhly a dihedrály a do nevazebných elektrostatika a van der Waalovy interakce (Obrázek 11). Prvním vazebným členem je suma kovalentních vazeb, kde K_r je silová konstanta vazby, r je aktuální vzdálenost mezi atomy a r_0 rovnovážná vzdálenost, obdobně je tomu u sumy vazebných úhlů. Posledním vazebným členem je suma torzí, tedy rotace kolem chemických vazeb, kde n je periodičita funkce, V_n je bariérou rotace, φ je rotační (torzní) úhel a γ značí úhel v kterém potenciál dosahuje svého minima. První nevazebným členem jsou van der Waalovy slabé interakce. Využívá se proto Lennard-Jonesova potenciálu, kdy jeho rovnici najdeme o podkapitolu výše (5.1), odpovídá za disperzní přitažlivost mezi atomy ($-1/r^6$) a zároveň odpuzování jader při přiblížení atomů ($+1/r^{12}$). A_{ij} a B_{ij} jsou parametry hloubky polohy minima křivky a R značí vzdálenost mezi atomy. Pro výpočet posledního elektrostatického členu, se využívá Coulombova zákona. Elektrostatický potenciál je pro atomové náboje se stejným znaménkem odpuzivý, a naopak pro stejná znaménka přitažlivý. V sumě je q_i a q_j jež jsou parciálními náboji atomů a ε je

dielektrická konstanta. Výpočtem této rovnice získáváme potenciální energii, její derivací pak síly. [41], [43]

Rovnice 4: Rovnice výpočtu celkové potenciální energie.

$$E_P = E_v + E_n$$

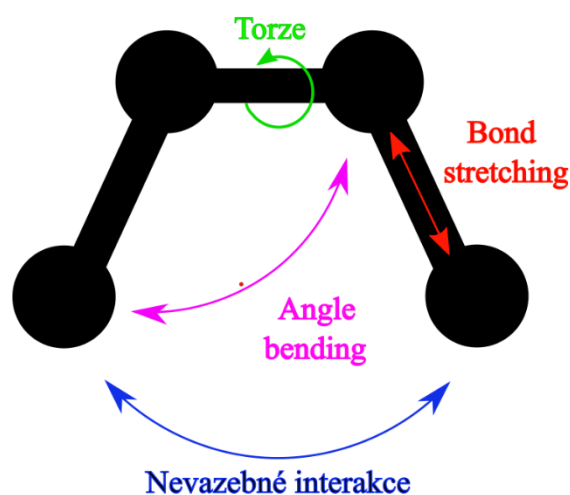
$$E_P = \sum_{\text{kovalentní vazby}} K_r (r - r_0)^2$$

$$+ \sum_{\text{úhly}} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{torze}} \frac{V_n}{2} (1 + \cos[n\varphi - \gamma])$$

$$+ \sum_{i>j}^{\text{atomy}} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \quad \text{van der Waals}$$

$$+ \sum_{i>j}^{\text{atomy}} \frac{q_i q_j}{\epsilon R_{ij}} \quad \text{elektrostatika}$$



Obrázek 11: Vazebné a nevazebné členy (vytvořeno v Inkscape).

Rozlišujeme tři základní typy silových polí a těmi jsou all atom, united atom a coarse grained pole. Silové pole all atom (AA) je první možný teoretický model MD popisující interakce na atomární úrovni s vysokou přesností a zahrnuje explicitní popis, tj. každý atom, včetně vodíku, je zastoupen svými vlastními parametry (náboj, poloha, hmotnost). Využívá se pro simulace MD v oblastech výzkumu biomolekul či v oblasti farmakologie. [44]

Oproti tomu silové pole United atom (UA) bylo vyvinuto hlavně k složitějším simulacím například skládání proteinů. V tom to silovém poli jsou vodíky vázané na alifatických uhlících v aminokyselinách sjednoceny dohromady, dochází tak díky tomu to sjednocení k minimalizaci narušení rozložení různých konformačních stavů proteinu. Skupina atomů C a H je tak chápána jako jednotné centrum interakce. [45]

Nejčastějším problémem MD simulací je jejich časová náročnost. Coarse grained (CG) je speciálně vytvořen pro dlouhé simulace. Silové pole představuje skupiny atomů jako

„efektivní kuličky, CG tak snižuje výpočetní náklady na výpočty, tím že jsou snižovány počty stupňů volnosti daného modelu. Jedním z nejvíce využívaných CG modelů je model Martini, využívá schéma mapování čtyři ku jedné (čtyři těžké atomy a přidružené vodíky v jedné „GC kuličce“).^[46]

6 CÍLE PRÁCE

Hlavními cíli této práce bude analyzovat, popsat a klasifikovat nekanonické konformační stavy cukrfosfátové páteře v několika neobvyklých DNA molekulách a v komplexech proteinu s DNA a dále prozkoumat, jak dobře jsou tyto stavy modelovány současnými silovými poli z rodiny AMBER. Pro tuto práci byly vybrány specifické nekanonické konformace torzních úhlů α a γ . Nekanonické konformace jako je tato jsou jednou z nejčastějších komplikací molekulově dynamických simulací. Klasická kanonická konformace α/γ , která se vyskytuje v konformaci BB00, a odpovídá tak kanonické B-formě DNA, má hodnotu α dle databáze DNATCO 304° a odpovídá tedy g^- (*gauche*-). Úhel γ dle databáze odpovídá 44° , tedy g^+ (*gauche*+). Pokud bychom tedy chtěli zkráceně popsat konformaci α/γ pro kanonickou formu, zapsali bychom tento stav jako $\alpha/\gamma = g^-/g^+$.

Avšak, jak bylo výše zmíněno, budou zde studovány především nekanonické konformace. Do výzkumu byly vybrány kombinace α/γ , které obsahovaly α či γ v *trans* oblasti, jež se vyskytuje okolo 180° . Tyto konformace byly zkoumány na jednotlivých modelových DNA molekulách, ale i v komplexech protein-DNA, kde předpokládáme, že daný stav přispívá ke stabilitě komplexu a v některých případech může být proteinem i vynucen.

Konformery, které by takové nekanonické stavy měly obsahovat jsou například BB12, BB13, BB03, BB02 a mnoho dalších. Experimentálně jsou popsány pro nespočet struktur, avšak snahou této práce je zjistit, zda jsou stabilní a správně popsány i v našem teoretickém modelování.

Kromě dříve popsaných konformerů byly vybrány i stavy páteře, které neodpovídají žádným definovaným konformacím (označované v DNATCO jako “NANT“ stavy), obsahující právě výše zmíněné nekanonické α/γ stavy. U nich bylo hlavní snahou, nalézt a klasifikovat nové konformační stavy (označované NS), které by jim nejlépe odpovídaly, aby byly jasně definovány a mohly tak být případně dále použity pro zpřesnění klasifikace nekanonických struktur v databázi DNATCO.

Pro simulace byla použita silová pole OL15, OL21 a bsc1 z rodiny AMBER. Silové pole OL21 má rozdíl od starších OL15 a bsc1 modifikaci α/γ potenciálu pro DNA, a mělo by tedy některé nekanonické regiony, zejména $\alpha = t$ a $\gamma = t$, lépe stabilizovat. To by mělo vést ke zvýšené stabilitě přirozených nekanonických stavů v MD simulacích a snížené snaze navštěvovat regiony, které by v realitě neměly být populovány. Pro porovnání s tímto vylepšeným silovým polem byla použita pole OL15 a bsc1 bez zmíněné α/γ modifikace. Další snahou je tedy dokázat, že silové pole OL21 s modifikovaným α/γ potenciálem

zpřesňuje popis nekanonických stavů v průběhu molekulově dynamických simulací oproti polím OL15 a bsc1.

7 METODY

V následující tabulce (Tabulka 2) jsou uvedeni jednotliví zástupci konformerů, kteří byli vybráni pro simulace s α/γ úhly, či jinými nekanonickými hodnotami. Termínem „konformer“ se myslí popis prostorového uspořádání molekul různých forem, které může molekula zaujmout při rotaci kolem vazeb. Jednotlivé konformace se liší v geometrii a energii, při jejich studování získáváme informace o vlastnostech daných struktur či jejich interakcích.

Výběr jednotlivých konformerů byl proveden na základě několika kritérií, a to dostatečného rozlišení jejich nativní struktury (do 2 Å), které udává míru detailnosti, s jakou jsou atomy a jejich vzájemné vzdálenosti identifikovány ve struktuře. Velikost struktury hraje v MD simulacích také svou roli, pokud by byla zvolena struktura příliš velká, nebyla by optimální pro rychlost simulace. Z toho důvodu byly vybrány struktury s takovou velikostí, aby simulace probíhaly v čase několika dní. A nakonec byl výběr ovlivněn dostatečným obsahem nekanonických α/γ rekonformací či jiných nekanonických hodnot vhodných pro analýzu v dané molekule.

Následně byly struktury kontrolovány dle hodnot RMSD (root-mean-square deviation), tedy střední kvadratické odchylky, poskytující informace o podobnosti s geometrií daného NtC konformeru. V tabulce je také uveden confal (konformační validační skóre), který vyhodnocuje správnost zařazení do určitých konformačních tříd NtC, nebo do jaké míry se shoduje s referenčními strukturami.

Tabulka 2: Zástupci vybraných NtC konformerů obsahující nekanonické konformace.

PDB ID	NtC	Krok konformeru (řetězec_číslo reziduí)	RMSD	Confal
1PQT	NANT	A_dA4_dA5	1,04	0
4U5M	NANT	A_dG5_dG6	1,52	0
		A_dT7_dG8	1,04	
		A_dG8_dG9	0,85	
		A_dT10_dG11	1,09	
		A_dG11_dG12	0,85	
		A_dT16_dG17	1,01	
		A_dG17_dG18	0,86	
		A_dT19_dG20	0,99	
		A_dG20_dG21	0,85	
		A_dT22_dG23	0,96	
		A_dG23_dG24	0,85	
A_dT25_dG26	0,96			
1UE2	NANT	A_C382_dG3	0,52	0
		A_dA4_dA5	0,40	
	BB12	A_dA5_dA6	0,71	90
	A_dG7_dC8	0,36		
1ZHU	NANT	A_dC6_dA7	1,45	0
2F5N	BB02	B_dA5_dG6	0,25	87
	BB13	C_dC5_dC6	0,40	59
	NANT	C_dA7_dG8	0,95	0
	BB12	C_dG9_dT10	0,41	65
	NANT	C_dT10_dC11	0,56	0
	BB12	C_dT11_dT12	0,54	18
	BB03	C_dT12_dA13	0,47	43
	BB02	C_dA13_dC14	0,39	22
2GIG	BB12	C_dC14_dC15	0,42	70
	NANT	E_dC10_dC11	0,46	0
	NANT	F_dG8_dA9	0,56	
NANT	F_dC10_dC11	0,53		

7.1 Parametrizace silového pole

Jak již bylo vícekrát výše zmíněno, v práci byla použita tři silová pole známá jako OL15, OL21 a bsc1 z rodiny AMBER, modifikovaná a odvozená od původního pole Amber ff99. Použitím více silových polí vede k rozdílným predikcím chování molekul v průběhu simulací v daném čase, což je zapříčiněno různými parametry jednotlivých polí. Porovnání výsledků z

těchto tří polí poskytne užitečné informace k zjištění jejich spolehlivosti a vhodnosti pro konkrétní aplikace na systém.

Parametrizací silového pole se rozumí definování parametrů (úhly, torze, vazebné potenciály) a interakčních potenciálů. Parametry určují chování molekul a možných interakcí v rámci simulace molekulární dynamiky. Silové pole se vybírá na základě vhodnosti pro daný simulovaný systém. V tom to případě byla vybrána optimální silová pole pro simulace nukleových kyselin a proteinů. Nejvhodnějším polem v dnešní době pro nukleové kyseliny by mělo být silové pole OL21. Hlavní snahou je tak vybrat takové silové pole, které by pak nejlépe reprodukovalo experimentální data.^[14]

Pro modelování energie systému získáváme parametry z obecné rovnice potenciální energie (Rovnice 4). Použité dihedrální parametry v polích, které jsou pro výsledné simulace této práce důležité, jsou následující. Pro silové pole OL15, parametry pro α/γ byly použity z pole bsc0 ($g+/t$, Barcelona α/γ modifikace) avšak hlavní modifikací tohoto pole je zpřesnění torzního úhlu β cukr-fosfátové páteře (β_{OL1}), vylepšení $\epsilon/\zeta = g-/t$ ($\epsilon\zeta_{OL1}$) a glykosidické torze χ (χ_{OL4}).

Obdobně je tomu u pole bsc1, ale již bez modifikace β dihedrálního úhlu, zde je parm99. Pro ϵ/ζ je zde bsc1 ϵ/ζ^2 a pro glykosidické torze bsc1 χ .

V silovém poli OL21 byl modifikován parametr α/γ (OL21 α/γ), zatímco z pole OL15 byla převzata modifikace β , ϵ , ζ , α χ dihedrálních úhlů (β_{OL1}). Pro modifikace parametrů α/γ se využilo metody, která zohledňuje konformačně závislé solvatační efekty, aniž by docházelo k dvojímu započítání solvatační energie.^{[3], [4]}

Pro popis proteinů bylo použito silové pole ff14SB.^[47] Pro předpověď fyzikálních vlastností vody v molekulově dynamické simulace byl využit vodní model SPC/E^[48] (extended simple point model) jakož to reparametrizace od původního SPC (simple point charge) modelu vody. Model je založen na popisu vzájemné interakce mezi bodovými náboji a vdW kuličkami. Voda je zde modelována jako tří bodový (tři atomový) rovnoramenný trojúhelník, kde pro každý bod jsou přiřazeny parciální náboje a pro kyslík i Lennard-Jonesovy parametry; délka vazeb a úhly jsou pevně dané. SPC/E zahrnuje průměrné účinky polarizace (average effects of polarization). Dalším možným vodním modelem by mohl být využit například TIP3P nebo čtyřbodový OPC, který je ale výpočetně velmi náročný. Pro model SPC/E byly publikovány přehledové články, kde se navzájem porovnávají jednotlivé vodní modely pro různé aplikace, kde jsem se inspirovala právě pro výběr tohoto modelu, pro struktury nukleových kyselin.

Jednotlivé MD simulace na GPU byly spuštěny v programu CUDA PMEMD (Particle Mesh Ewald Molecular Dynamics)^[49] z balíčku AMBER, který je k tomu primárně určen. Program CUDA výrazně snižuje čas potřebný k provedení simulací, protože výpočet probíhá na grafické kartě. Časový krok simulace byl nastaven na 4 fs.^[3] K optimalizaci integračního kroku simulace bylo využito metody HMR schéma přerozdělení hmotnosti atomů na atomu vodíku (Hydrogen mass repartitioning). Dochází k přerozdělení hmotnosti z těžších atomů (uhlík, dusík) na atomy vodíku, jejichž hmotnost tak zvětší, čímž se sníží jejich rychlost a zvětší se příspěvek k celkové kinetické energii systému. Technika HMR zlepšuje efektivitu MD simulací a umožňuje dosažení výrazného zrychlení dynamiky.^[50]

Simulace byly prováděny za konstantního tlaku, teploty a počtu částic (NPT). Tlak byl regulován Monte Carlo barostatem a byl nastaven na 1 atmosféru (101 kPa), MC barostat generuje náhodné změny objemu v systému a vyhodnocuje jejich efekt na tlak. Teplota byla konstantně udržována Langevinovým termostatem na 298 K (25 °C). Na vazby mezi vodíky s jinými atomy byla použita metoda SHAKE s tolerancí 0,00001 Å. Použití metody SHAKE umožňuje simulovat s delším časovým krokem, s větší integrační stabilitou a přesností. Souřadnice nukleových kyselin a proteinů byly ukládány každých 10 ps. Výsledné trajektorie byly analyzovány pomocí programu cpptraj z balíčku AMBER. U struktur samotné DNA bez proteinu byla celková délka simulace 5 ns, abych získala podrobnější a komplexnější informace. Pro komplexy protein-DNA byla délka simulace nastavena na 1 ns.^[3]

7.2 Příprava výchozí struktury

Výchozí struktura byla ve formě pdb a musela být nejprve upravena, aby mohla být načtena do programu LEaP. Nejprve bylo nutné u všech struktur zneutralizovat nábojové interakce v systému, aby nedocházelo mezi periodickými oktaedrickými boxy k repulzím a minimalizovaly se artefakty. Následně bylo dodáno potřebné množství iontů k dosažení výsledné koncentrace 0,15 M KCl s parametry iontů dle Jouna a Cheatham.^[51] Tímto způsobem byly připraveny i ostatní struktury.

K solvaci struktur byl využit oktaedrický box s vodním modelem SPC/E, jehož vzdálenost mezi vnějšími atomy a okrajem boxu byla minimálně 10 Å. Pokud by byla struktura větší, musí se její velikost vzít v potaz a zvětšit tak oktaedrický box, aby nedocházelo k nechtěným kontaktům s identickými strukturami ve vedlejších periodických boxech.

V uvedených strukturách se vyskytovaly i jiné ionty než jen draselné a chloridové. Jednalo se o ionty zinku vyskytující se ve struktuře DNA glykosylázy. Pro parametrizaci takového iontu byl využit AMBER tutoriál, popisující použití pole ZAFF.^[52] Z webové stránky

(<https://ambermd.org/tutorials/advanced/tutorial20/ZAFF.php>) byly staženy potřebné parametry jako ZEFF.prep a ZAFF.frcmod, obsahující parametry pro modelování – vazby, úhly, informace o náboji atd. (parametry byly staženy v lednu 2024). Dle tabulky uvedené na stránkách musely být vytvořeny nové vazby na čtyřnásobně koordinovaný iont zinku (Zn). V tomto případě u 2F5N se jednalo o Zn-CCCC (center ID: 1). Dále byly spolu s parametrem pro zinek dodány i příslušné parametry pro rezidua, jež jsou na něj navázaná, tedy v tom to případě čtyři cysteiny vázající se ke kovu, které se nadále označovaly jako CYM. Následně musely být ve výchozích PDB přepsána jména reziduí na jména modifikovaných reziduí ze ZAFF, a to následovně: rezidua iontu zinku na ZN1 a rezidua čtyř cysteinů na CY1.

U PDB ID: 1UE2, která v původním pdb byla jen jednořetězcovou strukturou, bylo potřeba vygenerovat komplementární řetězcovou strukturu, aby bylo možné ji simulovat. Toho jsem docílila za pomoci programu PyMol, kde bylo možno ze symetrie vygenerovat nejbližší sousedy řetězce.

V neposlední řadě bylo nutné určit protonizační stavy histidinu, jež se nachází ve třech formách. U histidinových zbytků získaných rentgenovou difrakcí nejsou uvedeny pozice atomů vodíku, tudíž nemůžeme s přesností určit o jaký protonizační stav histidinu se jedná. Pro určení, o jaký typ histidinu se jedná, jsem využila server PropKa^[53], který navrhuje protonizační stav dle zadaného pH a struktury v okolí histidinu. Pro ujištění se, že je navržený protonizační stav dle serveru PropKa správný a konzistentní se sítí vodíkových vazeb v krystalu, bylo potřeba vizuálně zkontrolovat protonizaci za pomoci programu PyMol, spolu se změřením délek vodíkových vazeb. Díky těmto dvěma ověřeními protonizačního stavu jsme schopni určit o jaký typ protonizace se jedná. V původním PDB se histidin zapisuje třípísmenným označením HIS, které neříká nic o svém protonizačním stavu. Po určení protonizace programem PropKa bylo možno volit jeden z následujících typů. Prvním typem histidinu je HID (deprotonovaný), kde je vodík umístěn na δ dusíku imidazolu, tedy blíže k postrannímu řetězci. Druhým typem je HIE (zvláštní forma), kde je vodík umístěn vzdáleněji od postranního řetězce, a posledním možným typem je HIP protonován na dusíku imidazolu, tj. nese kladný náboj.^[54] Dle dvojitého ověření typu histidinu, byly pak v původním PDB histidiny označené jako HIS, přepsány na jednotlivé typy HIP, HID či HIE.

8 VÝSLEDKY A DISKUZE

Hlavním cílem této práce bylo sledovat chování a stabilitu vybraných nekanonických konformací cukrfosfátové páteře v MD simulacích neobvyklých DNA motivů a komplexů proteinů s DNA. V první řadě se jednalo o simulace konformací obsahujících nekanonické hodnoty torzních úhlů α/γ , prováděné s novým silovým polem OL21, jehož modifikovaný potenciál α/γ by měl lépe popisovat právě tyto konformace. Následně budou výsledky porovnány s výsledky získanými s pomocí starších silových polí OL15 a bsc1.

Dalším úkolem bylo s pomocí MD simulací popsat a definovat nové konformační stavy páteře DNA, které nebyly popsány v databázi DNATCO (kde jsou označeny jako NANT). Tyto nové konformační stavy jsem popsala hodnotami torzních úhlů, které dobře charakterizovaly nativní geometrii v MD simulacích. Chování těchto stavů bylo opět porovnáno v simulacích s různými silovými poli.

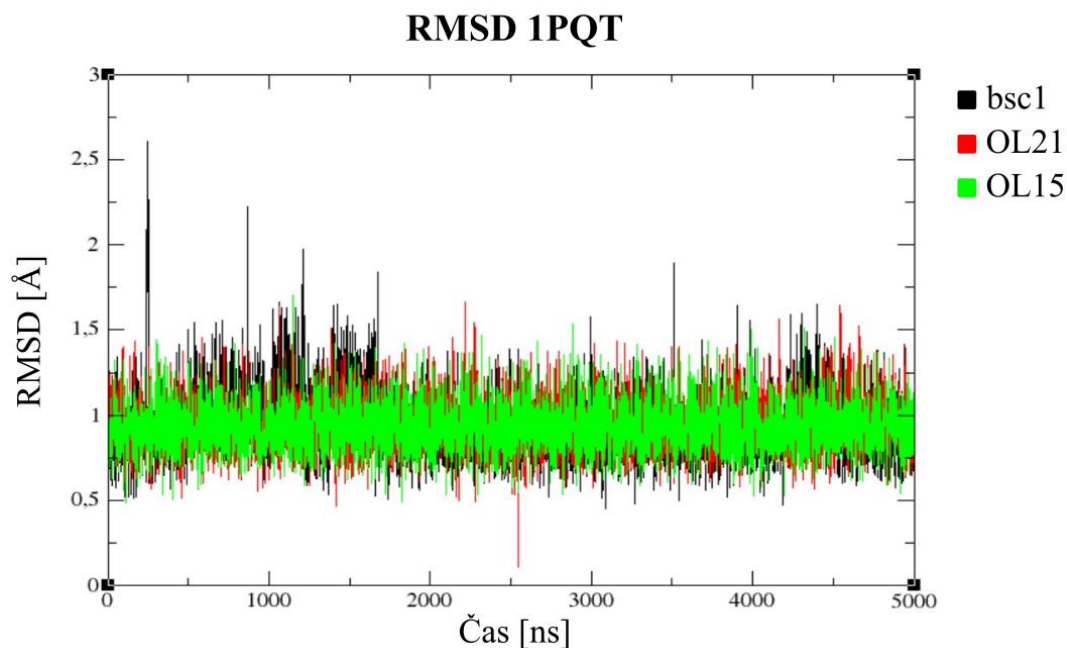
8.1 RMSD

Kontrola molekulově dynamických simulací byla prováděna vizuálně v programu VMD a pomocí grafů RMSD (root-mean-square deviation), tedy dle střední kvadratické odchylky souřadnic. Hodnoty RMSD zprostředkovávají informaci o tom, jak moc se během simulace jednotlivé snímky odchylojí od výchozí geometrie struktury, získané z experimentální analýzy. Z grafů RMSD můžeme usuzovat, zda byla MD simulace stabilní, nebo v ní došlo k významné rekonformaci molekuly, kterou doprovázejí vysoké hodnoty RMSD. Vysoké hodnoty RMSD mohou být indikací různých druhů problémů v simulacích a naznačují potřebu posouzení použitelnosti simulované trajektorie. Pokud se hodnoty nijak extrémně neodchylojí od původní experimentální geometrie, pak můžeme usoudit, že simulace probíhala správně a můžeme s ní nadále pracovat a analyzovat ji. Mírné odchylky RMSD se většinou připisují lehkým deformacím geometrie, které nijak neovlivňují použitelnost simulace. RMSD může nabývat vyšších hodnot pro molekuly větších rozměrů, případně kvůli fluktuacím volných konců nukleové kyseliny nebo proteinu.

8.1.1 RMSD GAA hairpinu, 1PQT

První zkoumanou strukturou je DNA hairpin 1PQT. RMSD (Obrázek 12) naměřené pro strukturu 1PQT v polích bsc1, OL21 a OL15 se drží okolo hodnot 1 Å, struktura je tedy v průběhu simulace velmi stabilní a obdobná geometrii NMR nativní struktury. V časovém průběhu záznamu RMSD se nachází několik odchylek okolo 2-2,5 Å, po kterých se ale RMSD vrací zpět k nižším hodnotám. Tyto odchylky jsou způsobeny volnými konci nukleové

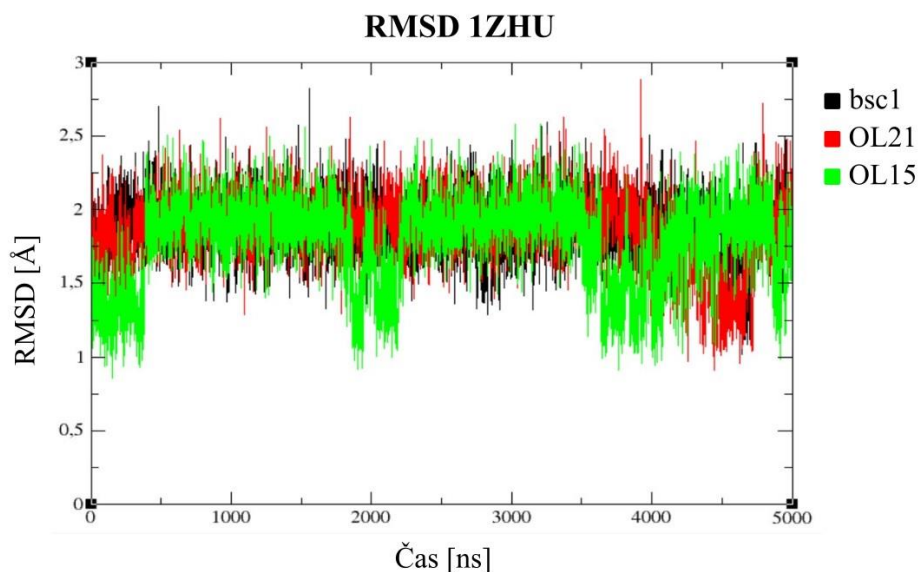
kyseliny. Celkově byly geometrie ve všech těchto třech silových polích stabilní po celou dobu simulace.



Obrázek 12: RMSD struktury GAA hairpinu 1PQT naměřené v polích bsc1, OL21 a OL15.

8.1.2 RMSD GCA hairpinu, 1ZHU

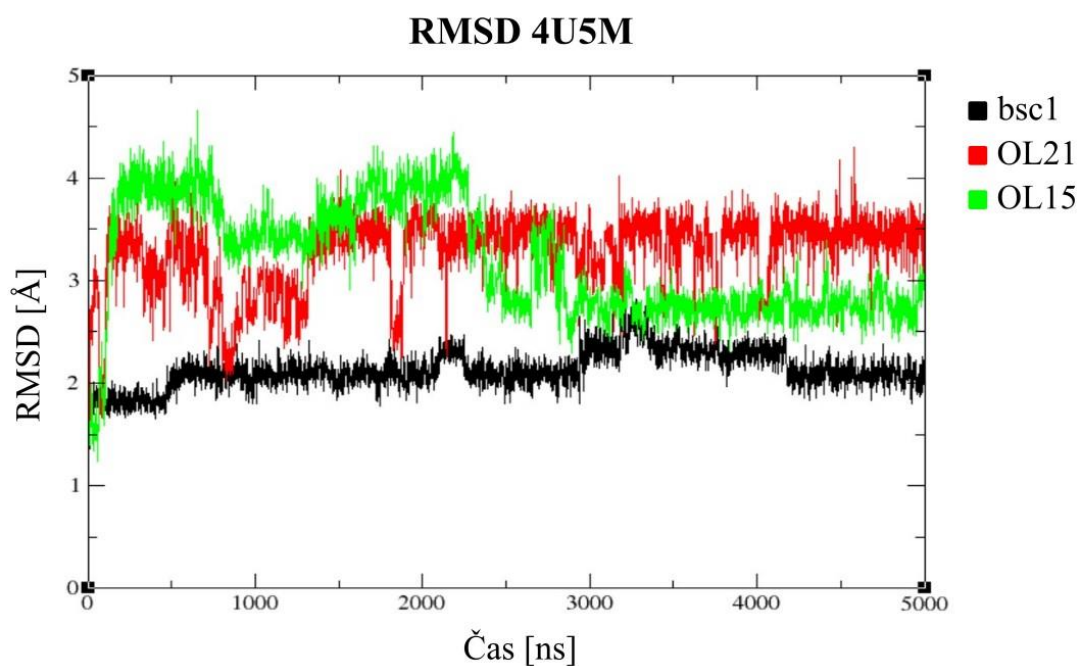
RMSD (Obrázek 13) pro strukturu GCA hairpinu 1ZHU nepřekročil ani v jednom ze tří silových polí hodnotu 3 Å. Od NMR původní struktury se v průběhu MD simulace nijak výrazně neodlišuje. Lehké odchylky, které můžeme vidět u silového pole OL15 jsou způsobeny odchlípnutím nespárované báze dC1, která volně fluktuuje v roztoku a báze v loopu GCA dC6 se v průběhu simulace také na menší chvíli odklopí, ale vrací se zpět. Obdobně je tomu u silového pole OL21, kde odchýlení od NMR nativní struktury není tak časté. Celkově jsou grafy RMSD ve všech polích velmi stabilní a nepřekračují hodnoty 3 Å. Můžeme tedy říct, že simulace proběhly v pořádku a můžeme tak nadále analyzovat data získaná z nich.



Obrázek 13: Graf RMSD pro strukturu GCA hairpinu 1ZHU ve všech silových polích.

8.1.3 RMSD G-kvadruplexu, 4U5M

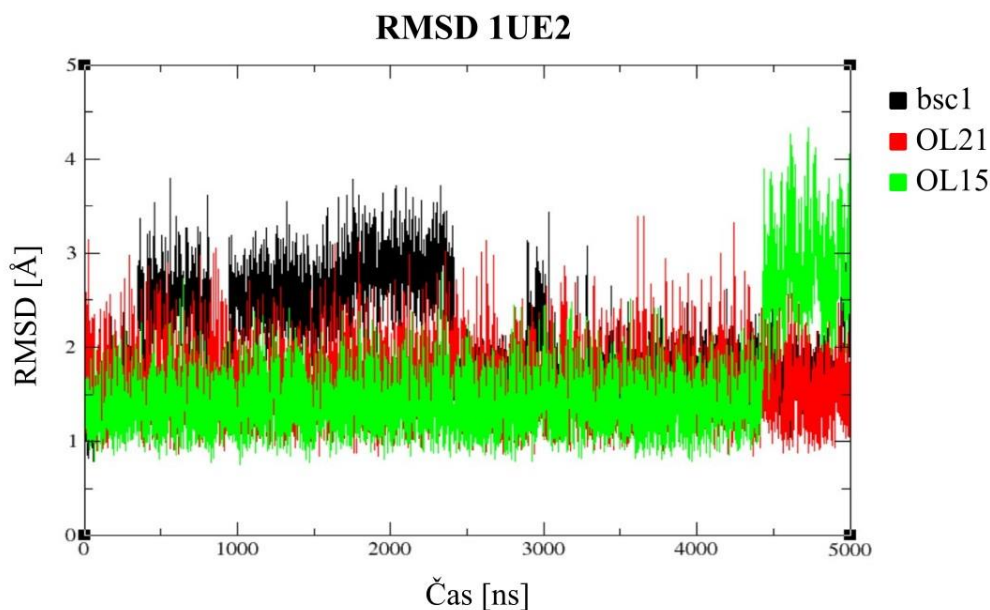
Hodnoty RMSD (Obrázek 14) struktury G-kvadruplexu 4U5M se pohybovaly nejčastěji v oblasti okolo 3 Å. Nejblíže se k původní krystalové geometrii přiblížila simulace silového pole bsc1 RMSD okolo 2 Å. Stemy kvadruplexů bývají většinou relativně stabilní, také díky stabilizaci ionty K^+ , nacházejícími se v kanálu G-kvadruplexu. To platilo i zde, avšak vyšší RMSD odchylky od původní geometrie byly zapříčiněny dlouhými volnými konci nukleové kyseliny. Pro větší kontrolu jsem si ověřila fluktuaci výše zmíněných volných konců v průběhu simulace v programu VMD, kde bylo zřetelně vidět. Dalším faktorem ovlivňujícím vyšší hodnoty RMSD, byly dvojí okupance torzního úhlu χ nacházejících se na čtyřech místech G-kvadruplexu, jež nabývaly střídavých hodnot *syn* a *anti*. Tyto torzní úhly měly v průběhu simulací tendenci odchýlit se od původních nativních hodnot, což může naznačovat nepřesnost parametrizace glykosidického úhlu nebo nesprávné určení konformace v X-ray struktuře. Průběh jednotlivých RMSD v každém silovém poli je až na občasné odchylky relativně stabilní, a nepřesahuje 5 Å. RMSD pro silové pole OL15 je zpočátku simulace lehce vyšší, avšak poté se hodnoty také ustálí v okolí 3 Å. Celkově se jedná o stabilní simulace.



Obrázek 14: RMSD G-kvadruplexu 4U5M v silových polích bsc1, OL21 a OL15.

8.1.4 RMSD interkalovaného duplexu, 1UE2

Hodnoty RMSD (Obrázek 15) v simulacích struktury interkalovaného duplexu 1UE2 se pohybovaly okolo 2 Å. V poli OL21 je struktura po celou dobu simulace relativně stabilní a nenabývá žádných odchylek. Oproti tomu v poli OL15 dojde v čase kolem 4500 ns ke zvýšení odchylky k 4 Å z důvodu, že se báze dA5 a dA13, které jsou ve stackingu, odklopují a opouštějí své místo. Dochází ke ztrátě stackingu s okolními bázemi. Podobnou odchylku vykazuje i pole bsc1 ve větší části první poloviny simulace ze stejného důvodu, odklopování bází dA4 a dA13. V poli OL21 k tomuto odklopování bází nedochází.

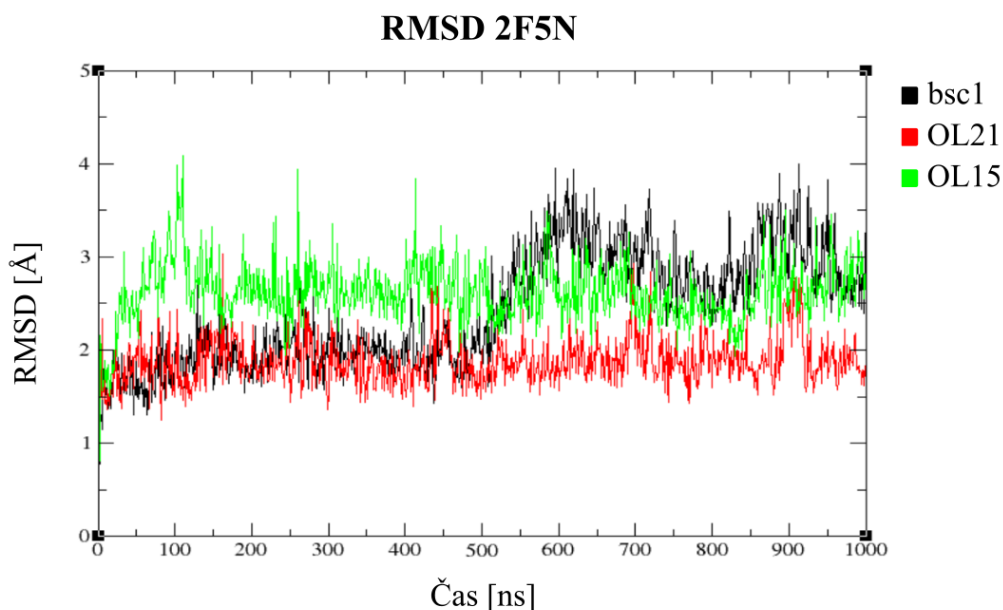


Obrázek 15: RMSD struktury interkalovaného duplexu 1UE2 v silových polích bsc1, OL21 a OL15.

8.1.5 RMSD DNA glykosylázy, 2F5N

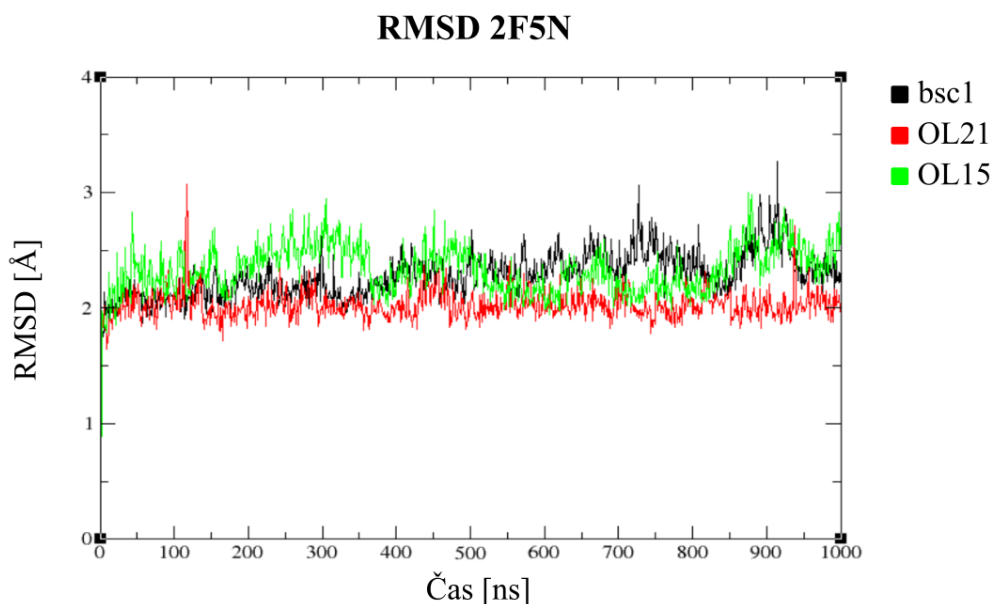
Stabilita dle RMSD se pro komplex DNA glykosylázy 2F5N prováděla zvlášť pro nukleovou kyselinu a zvlášť pro protein.

Na prvním RMSD (Obrázek 16) se nachází pouze RMSD pro nukleovou kyselinu. RMSD zde opět nepřesahuje hodnot 4 Å, a drželo se ve všech třech polích spíše okolo 3 Å. Nejvíce stabilní a nejbližší experimentální krystalové struktúře bylo pole OL21, které nemá v průběhu simulace žádné vyšší odchylky. V poli bsc1 a OL15 došlo k menším odchylkám, které připisují velikosti nukleové kyseliny a jejím volným koncům, které fluktuovaly v průběhu simulace, kde báze A_dA1 není párována, a tudíž v průběhu simulace docházelo k častým fluktuacím.



Obrázek 16: Graf RMSD nukleové kyseliny struktury DNA glykosylázy 2F5N.

Protein dle RMSD (Obrázek 17) vypadá velmi stabilně a mírné odchylky budou dány velikostí proteinu. Po shlédnutí simulace v programu VMD připisují mírné odchýlení fluktuacím volných konců. Můžeme tedy říci, že dle kontroly stability RMSD proběhly simulace vydařeně.

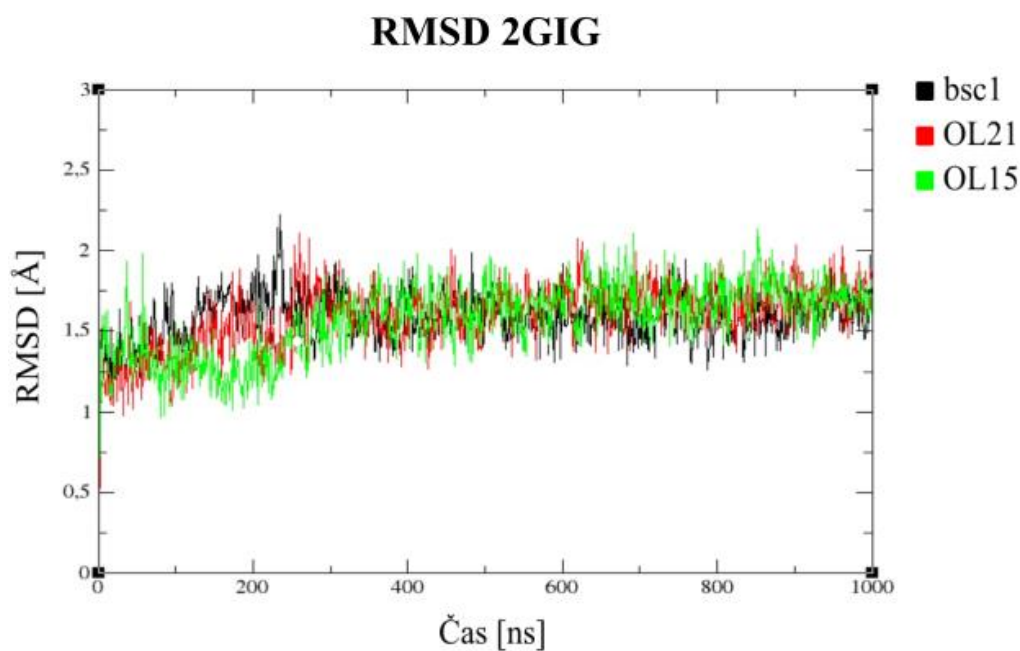


Obrázek 17: Graf RMSD proteinu struktury DNA glykosylázy 2F5N.

8.1.6 RMSD DNA endonukleázy, 2GIG

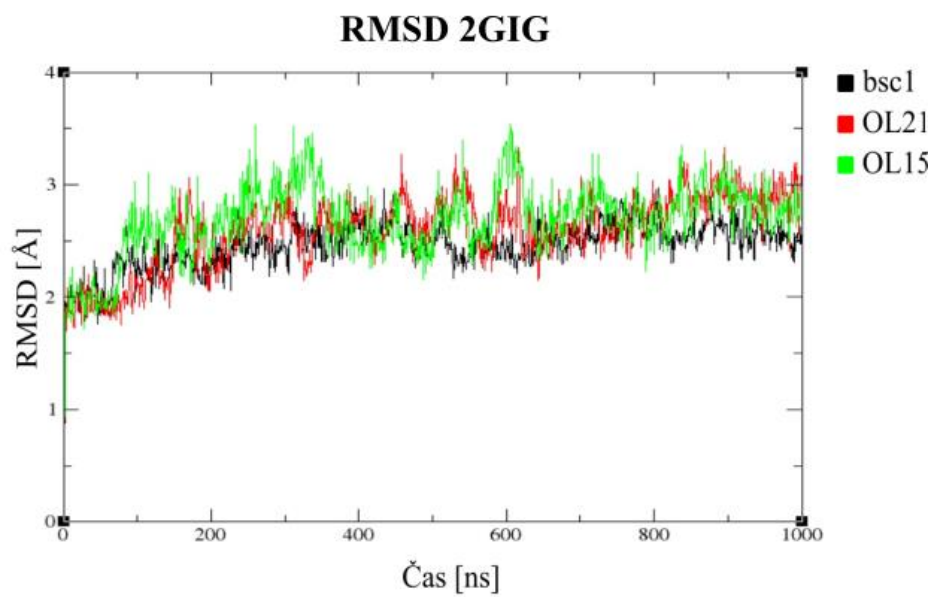
Kontrola stability u tohoto komplexu proběhla také zvlášť pro nukleovou kyselinu DNA endonukleázy a pro protein tohoto komplexu.) RMSD nukleové kyseliny, která vypadá velmi stabilně (Obrázek 18). Mírné odchylky od původní struktury jsou způsobeny velikostí

samotné nukleové kyseliny a fluktuací volných konců v roztoku. Hodnoty RMSD se drží pod 2 Å.



Obrázek 18: RMSD nukleové kyseliny DNA endonukleázy 2GIG.

RMSD proteinu (Obrázek 19) je opět stabilní, mírné odchylky způsobují volné konce proteinu, které fluktuují v průběhu MD simulace. Celkově hodnoty RMSD nepřesahují hodnot 3,5 Å až na pár výjimek, z nichž se opět vrací k nižším hodnotám okolo 2,5 Å, kde setrvávají většinu simulace. Obě provedená RMSD poukazují na stabilitu MD simulace.



Obrázek 19: RMSD proteinu struktury DNA endonukleázy 2GIG.

8.2 NOVÉ KONFORMAČNÍ STAVY

Jelikož se ve studovaných molekulách DNA vyskytovalo mnoho nepřirazených konformačních stavů NANT, bylo jedním z hlavních úkolů tyto konformace popsat a definovat pro ně nový konformační stav NtC, který by mohl být vhodným kandidátem na rozšíření databáze konformačních stavů DNATCO.

Východiskem byla stávající databáze DNATCO, ze které bylo čerpáno přiřazení v ní definovaných stavů NtC dle devíti torzních úhlů. S pomocí těchto devíti torzí byly také definovány nové konformace, které byly v této práci označovány zkratkou NS (new state).

Nový konformační stav byl vytvářen na základě dat získaných z molekulově dynamických simulací. Histogramy jednotlivých torzních úhlů naměřených v průběhu simulace byly vytvářeny zvlášť pro jednotlivé torze každého studovaného DNA stepu, a v některých případech jen pro torze α a γ , pokud byla potřeba analýza jen těchto úhlů. Nejpoptulovanější regiony histogramů byly charakterizovány průměrnou hodnotou torzního úhlu v tomto regionu. Pokud bylo navštěvováno více regionů, byly nové konformační stavy vytvořeny zároveň s nejpoptulovanějším i pro ně.

Navštěvované regiony různých úhlů byly následně přiřazeny k sobě podle toho, zda se v časovém průběhu úhlů z MD simulace vyskytovaly zároveň. Například pokud se α v pozici *gauche+* vyskytovala zároveň s úhlem γ , který se nacházel v pozici *trans*, pak byly ve výsledné konformaci přiřazeny k sobě a popsány jako $\alpha/\gamma = g+/t$.

Způsob nalezení nejpoptulovanějších regionů, v nichž se dané torze nacházejí, byl aplikován zvlášť v každém silovém poli, kde byl region definován pomocí histogramů. Pokud byly histogramy obdobné či totožné v jednotlivých silových polích, pak byl proveden průměr mezi všemi poli a bylo dosaženo zisku regionů definující dané torze. Tímto způsobem byl definován nový konformační stav (NS) o devíti dihedrálních úhlech popisujících lokální konformaci páteře DNA.

V některých případech byly rekonformace úhlů ϵ/ζ nezávislé na rekonformacích úhlů α , β a γ . Proto pro dosažení výsledné definice nového stavu byl vybrán pouze nejnavštěvovanější region úhlu ϵ a ζ , i když se ne vždy vyskytoval zároveň s definovaným stavem α , β a γ .

Následně byly nové konformační stavy společně s NtC konformery známými z databáze DNATCO přiřazovány awk skriptem. Tento skript analyzoval hodnoty devíti torzních úhlů pro každý snímek MD simulace a přiřadil jim NtC konformaci nebo nový stav NS na základě nejmenší vzdálenosti v 9D prostoru torzí. Při přiřazování byl použit cutoff na vzdálenost analyzovaného snímku od referenční struktury, aby byly snímky přiřazeny pouze, pokud jsou

dostatečně blízké k referenční struktuře. Cutoff měl hodnotu 80° , takže snímky, jejichž vzdálenost od referenční struktury byla větší než 80° (RMSD) nebyly přiřazeny.

Kontrola molekulově dynamických simulací byla prováděna vizuálně v programu VMD (Visual Molecular Dynamics) a pomocí grafů RMSD (root-mean-square deviation), tedy dle střední kvadratické odchylky souřadnic. Hodnoty RMSD zprostředkovávají informaci o tom, jak moc se během simulace jednotlivé snímky odchyľují od výchozí geometrie struktury, získané z experimentální analýzy. Z grafů RMSD můžeme usuzovat, zda byla MD simulace stabilní, nebo v ní došlo k významné rekonformaci molekuly, kterou doprovázejí vysoké hodnoty RMSD. Vysoké hodnoty RMSD mohou být indikací různých druhů problémů v simulacích a naznačují potřebu posouzení použitelnosti simulované trajektorie. Pokud se hodnoty nijak extrémně neodchyľují od původní experimentální geometrie, pak můžeme usoudit, že simulace probíhala správně a můžeme s ní nadále pracovat a analyzovat ji. Mírné odchylky RMSD se většinou připisují lehkým deformacím geometrie, které nijak neovlivňují použitelnost simulace. RMSD může nabývat vyšších hodnot pro molekuly větších rozměrů, případně kvůli fluktuacím volných konců nukleové kyseliny nebo proteinu.

8.2.1 Konformační stavy GAA hairpinu, 1PQT

V GAA smyčce této struktury byl zkoumán step A_dA4_dA5, který je v databázi DNATCO definován jako nepopsaný NANT konformační stav. Proto byl zde úkol následující: zjistit, o jaký konformační stav se zde jedná, a pokusit se ho popsat charakteristickými hodnotami torzních úhlů. V tabulce (Tabulka 3) níže jsou uvedeny původní nativní hodnoty z NMR struktury, ke kterým jsem se snažila přiblížit. Nejvíce jim odpovídal nový konformační stav NS1 a již existující NtC konformace OP19, kde NS1 byl vytvořen jako průnik jednotlivých torzních úhlů mezi všemi silovými poli, tj. utvořen z nejpobulovanějších regionů torzních úhlů, na základě jejich současného výskytu v daných regionech v časovém průběhu přibližně celé MD simulace, pokud tomu tak bylo.

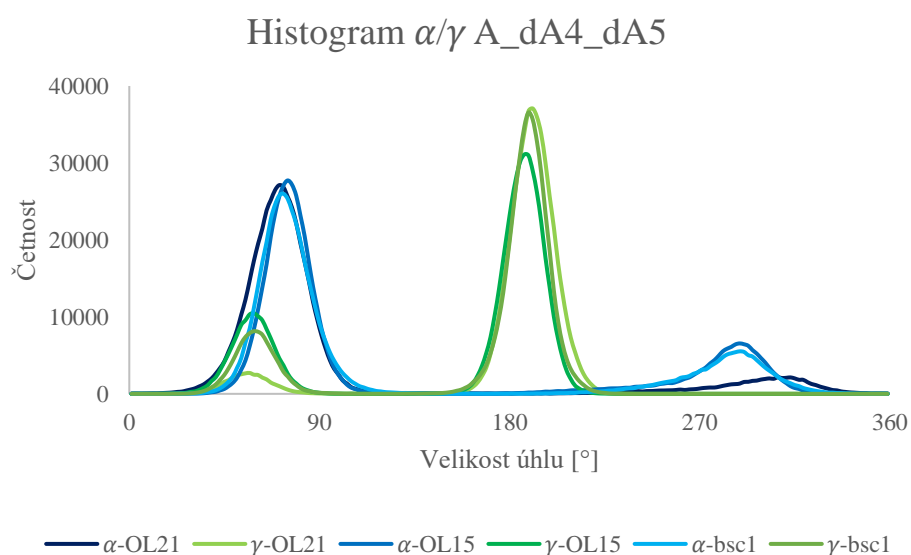
Tyto nejpobulovanější nekanonické regiony α/γ rekonformací jsou vyobrazeny v grafu (Graf 1) níže, kde jsou úhly α vyznačeny zelenou barvou a úhly γ modrou. Torze α i γ navštěvovaly také méně obsazené kanonické regiony, avšak ty se v průběhu simulace vyskytovaly jen minimálně (pod 10 % celkového výskytu). V konformeru OP19 byly úhly α/γ nekanonické $\alpha/\gamma = g+/t$ (*gauche+/trans*). Tyto dva stavy se vyskytovaly v průběhu celé simulace ve všech třech polích. Jednotlivé nové konformační stavy se v průběhu simulace velmi často střídaly. Jediným rozdílem mezi novými konformačními stavy byly hodnoty torzí ϵ/ζ , v grafu (Graf 2) jsou úhly ϵ vyznačeny modrou barvou a ζ oranžovou. V simulaci

navštěvovaly kanonickou i nekanonickou formu nezávisle na konformacích ostatních okolních úhlů. Hodnoty torzí ϵ/ζ zastoupené nejčetnějším regionem odpovídaly stavu NS1.

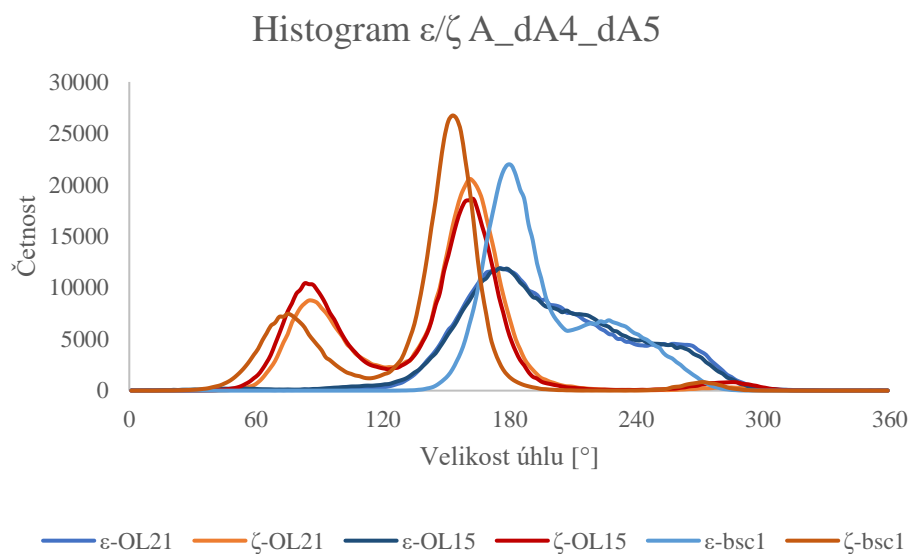
Tabulka 3: Hodnoty devíti torzních úhlů [°] pro nativní strukturu a nové stavy (NS) získané z MD simulace, společně s NiC konformací OP19 struktury GAA hairpinu.

	$\delta 1$	$\chi 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 2$
nativní	146	245	210	103	76	210	185	149	280
NS1	138	253	175	155	72	227	190	146	279
OP19	145	250	225	64	74	186	188	126	255

Graf 1: Histogram pro α/γ stepu dA4_dA5 ve všech silových polích GAA hairpinu.



Graf 2: Histogram ϵ/ζ stepu A_d14_dA5 ve všech silových polích GAA hairpinu.



Silové pole OL21, ve kterém je nově modifikován α/γ potenciál a torze ϵ/ζ jsou přebrány ze silového pole OL15, nejlépe popsalo nekanonické stavy OP19 a NS1 jejichž geometrie se

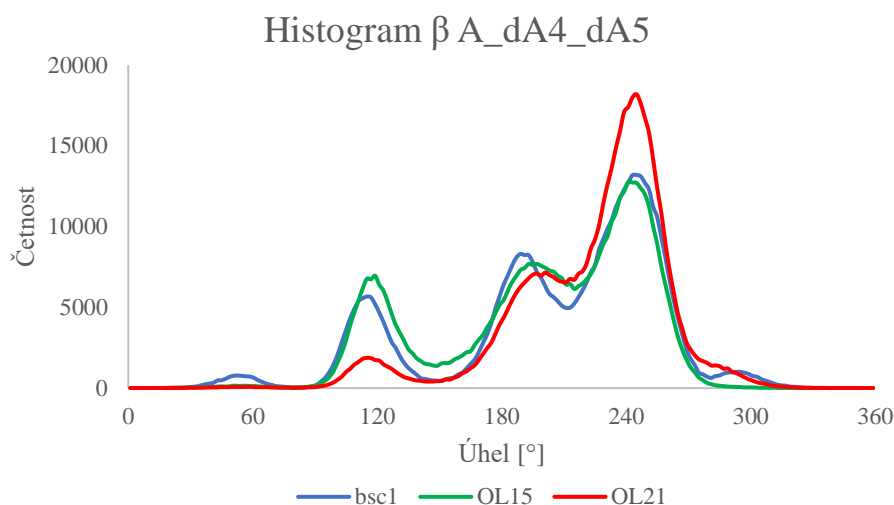
nejvíce blíží nativní NMR konformaci. Nový konformační stav NS1 se zde vyskytoval v 60 %, dále se tento stav udržel v 46 % v poli OL15 a nejhůře na tom bylo silové pole bsc1, které tento stav udrželo pouze v 48 %. Konformační stav OP19 se nejvíce vyskytoval v silovém poli OL15 až v 26 %. V následující tabulce (Tabulka 4) je celkový procentuální přehled nových konformačních stavů jednotlivě v každém silovém poli. Jelikož si tyto konformační stavy jsou v tomto případě velmi podobné (odlišné pouze v ε/ζ), z pohledu úhlů α/γ , které mě zajímají nejvíce, by se dalo říci, že se jedná o tentýž stav. A tudíž celkové procento obsažené těmito stavy v daném silovém poli je v tuto chvíli nejdůležitějším faktorem pro posouzení, jak přesně silové pole popisuje nekanonické stavy α/γ . Tabulka 4 ukazuje, že silové pole OL21 dokázalo nekanonický stav relativně dobře popsat a stabilizovat až na necelých 90 % doby simulace, což naznačuje, že modifikovaný α/γ potenciál v poli OL21 popisuje konformaci GAA smyčky lépe než ostatní testovaná silová pole.

Tabulka 4: Procentuální podíl nových konformačních stavů GAA hairpinu v jednotlivých silových polích.

	OL21 [%]	OL15 [%]	Bsc1 [%]
NS1	62,0	46,1	48,7
OP19	25,5	26,7	27,3
Součet	87,5	72,8	76,0

Pro step A_dA4_dA5 bylo hlavním cílem zanalyzovat konformer s nekanonickými hodnotami α/γ , avšak v této konformaci se vyskytuje i nekanonická β nabývající hodnot vyšších, než je kanonická *trans* okolo 180° (*high-trans*). Proto jsem zkusila porovnat torzní úhel β pomocí histogramů v jednotlivých silových polích a pozorovat tak její chování v průběhu molekulově dynamických simulací. Na tomto grafu (Graf 3) je vidět evoluční posun, kdy silové pole bsc1 je velmi obdobné silovému poli OL15, a oproti tomu novější pole OL21 již lépe stabilizuje nekanonický konformační stav torzního úhlu β . V poli OL21 úplně vymizel stav okolo 60°, navštěvovaný v bsc1, a raději úhel β navštěvoval nekanonický stav, který se nacházel v původní NMR struktuře, tedy okolo 230°. V silových polích OL15 a bsc1 β navštěvovala také stav okolo 110°, kterýžto v poli OL21 také skoro vymizel. Z tohoto histogramu můžeme tedy usoudit, že novější silové pole OL21 dokáže lépe stabilizovat nekanonický stav i pro torzní úhel β .

Graf 3: Histogram β stepu A_dA4_dA5 GAA hairpinu ve všech silových polích.



8.2.2 Konformační stavy GCA hairpinu, 1ZHU

Tato struktura byla vybrána, protože se zde nachází step A_dC6_dA7, který obsahuje nekanonický $\alpha/\gamma = g-/t$ a tento step je v databázi DNATCO zařazen do třídy NANT. Struktura 1ZHU obsahovala ještě jeden nedefinovaný NANT step a to A_dC1_dA2, avšak jak již bylo zmíněno u ověřování stability dle RMSD, tato rezidua se nacházejí na konci nukleové kyseliny a dochází u nich k vysoké fluktuaci a dochází tak k deformacím páteře. Proto tento step nebudu analyzovat.

Hlavní snahou bude analyzovat α/γ rekonformaci v tomto konformačním kroku. Předpokládám obdobné chování jako u předchozí struktury GAA hairpinu 1PQT. Na základě nalezení nejpopulovanějších regionů devíti torzí se nalezený konformační stav shodoval s NS1 v 1PQT. Byl zde tedy zaveden nový konformační stav NS1 definovaný v předchozí kapitole 8.2.1 struktury GAA hairpinu. Tento nový stav NS1 obsahuje hodnoty úhlů $\alpha/\gamma = g+/t$, které se velmi podobají hodnotám nativním dvojí okupance (B). Hodnoty všech torzí jsou uvedeny v porovnání s původními NMR hodnotami v tabulce (Tabulka 5) níže. Torzní úhel α si vždy hned po začátku MD simulace našel výhodnější stav, který je v tomto případě nekanonický *gauche+* a v tomto stavu setrval po zbytek simulace, nebo se jen minimálně vracel do kanonického *gauche-*.

V MD simulaci se objevovala i konformace OP19, která také obsahuje nekanonické hodnoty pro $\alpha/\gamma = g+/t$, avšak neodpovídá původním hodnotám ϵ/ζ . Tyto dva stavy si jsou

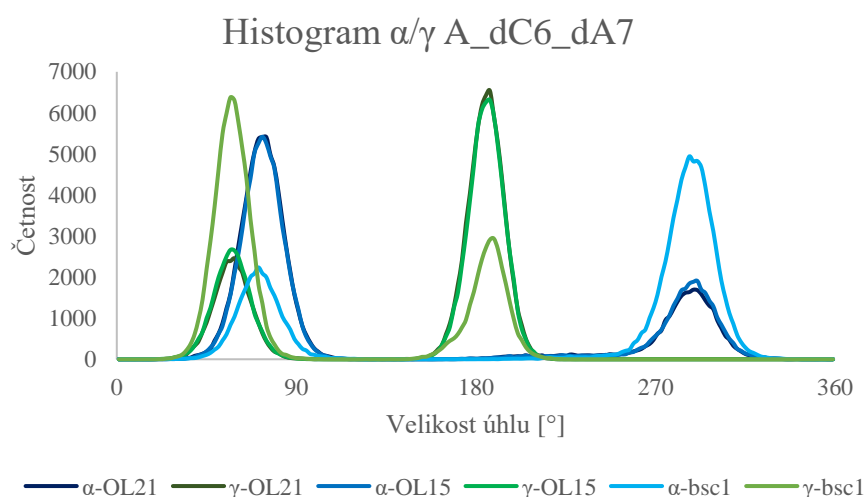
velmi podobné až na zmíněné torze ε/ζ , proto jsem je ve výsledcích uvedla oba a sečetla je tak dohromady jakožto podobnost k nativním hodnotám.

Tabulka 5: Hodnoty nativní struktury GCA hairpinu a nových konformačních stavů [°].

	$\delta 1$	$\chi 1$	$\varepsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 2$
A_dC6_dA7 (A)	142	237	175	250	265	57	166	123	263
A_dC6_dA7 (B)	151	239	183	130	80	259	184	122	262
NS1	138	253	175	155	72	227	190	146	279
OP19	145	250	225	64	74	186	188	126	255

Pro porovnání chování úhlů α/γ během MD simulace bylo využito histogramů jednotlivých torzních úhlů, proveden pro všechna tři silová pole. Z grafu (Graf 4) je viditelné, že torze navštěvovaly i jiné než-li jen nekanonické regiony. Torzní úhel γ pro pole OL21 a OL15 nejčastěji setrval v nekanonické *trans* poloze, oproti tomu v poli bsc1, kde úhel γ setrval v kanonické *gauche+*. Torzní úhel α navštěvoval nejčastěji nekanonický region *gauche+* v polích OL21 a OL15, v poli bsc1 tento stav spíše setrval v kanonické *gauche-*. Z histogramu α/γ vyplývá, že nejlépe nekanonickou konformaci udržela pole OL21 a OL15, to můžeme připsat jejich evoluční vyspělosti.

Graf 4: Histogram α/γ stepu A_dC6_dA7 GCA hairpinu ve všech silových polích.



Výsledné procentuální zastoupení nového stavu NS1 je uvedené v tabulce (Tabulka 6) níže. Celkově si jsou výsledky podobné pro silová pole OL21 a OL15. Silové pole bsc1, které v MD simulaci spíše upřednostňovalo kanonické konformace, se k původní nekanonické γ *trans* nepřiblížilo, podobnost s nativními hodnotami pole udrželo pouze u α *gauche-*. Pole OL21 a OL15 hodnoty úhlů α/γ v novém konformačním stavu NS1 udržely v nekanonických

hodnotách na více jak 50 % celé MD simulace. Můžeme tedy říci, že si torzní úhel α raději vyhledá nekanonické hodnoty než-li kanonické které po zbytek simulace navštěvuje současně s nekanonickou γ v regionu *trans*.

Stavy OP19 se vyskytovaly u polí OL21 a OL15 procentuálně nad 17 %, u pole bsc1 se tento stav vyskytoval minimálně, z toho důvodu, že bsc1 nebylo schopné správně nasimulovat a udržet nekanonické hodnoty torzních úhlů α/γ . Celkově součtem stavů NS1 a OP19 získáváme v polích OL21 a OL15 výsledné hodnoty okolo 70 %.

Kdybychom započítaly k těmto hodnotám ještě kanonický druhý nejnavštěvovanější region úhlů, dosáhlo by se hodnot u polí OL21 a OL15 až 85 %, pro pole bsc1 by jen samotné kanonické konformace utvořily více jak 60 % MD simulace.

Tabulka 6: Výsledné procentuální zastoupení NS1 a OP19 v GCA hairpinu.

	OL21 [%]	OL15 [%]	Bsc1 [%]
NS1	52,7	52,1	23,9
OP19	18,2	17,0	0
Součet	70,9	69,1	28,1

Výsledná procenta poukazují na shodu mezi strukturami hairpinů 1PQT a 1ZHU, kdy v nedefinovaných stepech se nacházely stejné konformace, které mezi sebou v průběhu MD simulace přecházely.

8.2.3 Konformační stavy G-kvadruplexu, 4U5M

Struktura levotočivého G-kvadruplexu je zajímavá sama o sobě, avšak pro tuto práci byla především zajímavá v tom, že všechny její stavy jsou v databázi DNATCO označeny jako nedefinované NANT konformace. Zde byly studovány stavy nacházející se v tabulce 7. Všechny tyto stavy obsahovaly v nativní struktuře stejné konformace, charakteristické hodnotami torze α v poloze *trans* a β v poloze *high-trans*, kde torze γ se objevovala buď jako kanonická *gauche+* anebo jako nekanonická *gauche-*. Takováto konformace nebyla před tím v databázi DNATCO popsána, a proto je zde uvedena jako nedefinovaná. Pokud by uvedená mnohonásobně se opakující konformace byla v MD simulacích stabilní, bylo by možné o ní přemýšlet jako o nové konformaci, jež by mohla rozšířit databázi DNATCO. Zde se tedy zaměřím hlavně na nekanonické úhly α/β a jejich stabilitu v různých silových polích.

Silová pole OL15 a OL21 by měla tento nekanonický stav stabilizovat lépe, díky modifikaci torzního úhlu β , která lépe stabilizuje hodnoty úhlu β v oblasti *high-trans*. Pole OL21 s modifikovaným α/γ potenciálem by mělo stabilizovat úhel $\alpha = trans$ lépe než pole

OL15 a bsc1. Na základě MD simulací ve všech třech silových polích byly vytvořeny nové konformační stavy vycházející z hodnot maxim distribucí torzních úhlů v jejich histogramech v jednotlivých silových polích. V tabulce níže (Tabulka 7) jsou uvedeny nové konformační stavy (NS) spolu s nativními hodnotami torzních úhlů získanými experimentální metodou rentgenové difrakce. Z tabulky můžeme vyčíst, že nový konformační stav NS1 je populovaný především v polích OL15 a OL21 a je svými hodnotami velmi podobný hodnotám nativním, proto ho také můžeme označovat za nativní.

Tabulka 7: Hodnoty torzních úhlů nových konformačních stavů pro 4U5M spolu s nativními hodnotami z krystalové struktury. Hodnoty úhlů jsou uváděny v °.

	$\delta 1$	$\chi 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 2$
nativní	143	234	303	92	171	229	55	143	234
NS1	143	222	307	84	176	224	55	143	222
NS2	141	228	307	133	304	160	307	133	228
NS3	146	235	279	145	295	155	320	146	235

Nekanonické stavy obsahující torze α/β mají nativní hodnoty $\alpha/\beta = trans/high-trans$. V MD simulaci se se hodnoty pohybovaly především v $\alpha/\beta = t/high$ (*tras/high-trans*) (Tabulka 8), tyto hodnoty udržela pole OL21 a OL15. V silovém poli bsc1 se hodnoty držely $\alpha/\beta = g-/t$, tedy v klasických kanonických hodnotách.

Tabulka 8: Hodnoty torzních úhlů α/β ve vybraných nekanonických stavech struktury 4U5M. Jsou-li pro jednu torzi uvedeny dva stavy (např. $t/(g-)$), pak první je dominantní a druhý méně populovaný.

Step	Úhel	Nativní struktura (α/β)	OL21	OL15	bsc1
A_dG5_dG6	$\alpha 6, \beta 6$	t; high	t; high	$t/(g-)$; high	$g-; t$
A_dT7_dG8	$\alpha 8, \beta 8$		t; high	t; high	$g-; t$
A_dG8_dG9	$\alpha 9, \beta 9$		t; high	t; high	$g-; t$
A_dT10_dG11	$\alpha 11, \beta 11$		$t/(g-)$; high	$t/(g-)$; high	$g+; t$
A_dG11_dG12	$\alpha 12, \beta 12$		$t/(g-)$; high	t; high	$g-; t$
A_dT16_dG17	$\alpha 17, \beta 17$		$g-/t; t/(high)$	$g-/(t)$; high	$g-; t$
A_dG17_dG18	$\alpha 18, \beta 18$		t; high	t; high	$g-; t$
A_dT19_dG20	$\alpha 20, \beta 20$		t; high	t; high	$g-; t$
A_dG20_dG21	$\alpha 21, \beta 21$		t; high	$t/(g-)$; high	$g-; t$
A_dT22_dG23	$\alpha 23, \beta 23$		$g-/(t); t/(high)$	$g-/t; high/t$	t; t
A_dG23_dG24	$\alpha 24, \beta 24$		$g-/t; t/high$	$g-/(t); t/(high)$	$g-; t$
A_dT25_dG26	$\alpha 26, \beta 26$		$t/g-; high/t$	$g-/(t); t/high$	$g-; t$

Hlavní snahou bylo, aby silová pole udržela nejlépe po celou dobu simulace nativní hodnoty, tedy hodnoty nového konformačního stavu NS1. Konformační stavy NS2 a NS3 jsou

charakterizovány kanonickými hodnotami torzí α a β , avšak v experimentální struktuře se vyskytují nekanonické hodnoty γ (*gauche*-). V tabulce (Tabulka 9) jsou uvedeny výsledky s cutoffem nových konformačních stavů v jednotlivých silových polích. Výsledná procenta představují průměr hodnot všech vybraných stepů, viz Tabulka 10.

Jedinou výjimkou, u které došlo v MD simulaci se silovým polem OL21 k výraznému odchýlení od stavu NS1, byl step A_dT10_dG11, kde došlo ke změně úhlu χ v dT10 z konformace *anti* na *syn*. Tymin dT10 interaguje stackovou interakcí s koncovým kvartetem kvadruplexu, ale jeho překryv a interakce s kvartetem není optimální, a proto dochází relativně snadno k uvolnění dT10 do roztoku a k případné rekonformaci úhlu χ . Protože tato změna nesouvisela s konformací páteře, zahrnula jsem obě konformace (s $\chi_1 = 235^\circ$ i s $\chi_1 = 58^\circ$) do stavu NS1. To bylo prakticky provedeno tak, že byla zavedena nová speciální konformace s $\chi_1 = \textit{syn}$ ($\chi_1 = 58^\circ$) a její populace byla započítána do populace stavu NS1. Celková procentuální populace stavu NS1 pro step A_dT10_dG11 v poli OL21 se započítáním hodnot χ jak *anti*, tak *syn* byla 87,3 %. V tabulce 10 je tato hodnota označena hvězdičkou s poznámkou.

U ostatních silových polí tomu tak nebylo, jejich simulovaná chí hned na začátku simulace přeskočila do polohy *anti*. Tento přeskok do jiné polohy byl zapříčiněn tím, že struktura má v tomto konformačním kroku dvojí okupanci, což znamená, že může nabývat, jak hodnot *syn*, tak hodnot *anti* (pro chí) v poměru 50:50. Další dvojí okupance se nacházely u χ_4 , χ_{13} a χ_{14} , avšak ty zde nijak nadále neovlivňovaly vybrané konformační stepy.

Silové pole OL21 dokázalo udržet stav NS1 v necelých 70 % celé simulace, pole OL15 v 68 % simulace, celkový rozdíl mezi poli je přibližně o 2 %. V porovnání mezi jednotlivými stepy, silové pole OL21 dokázalo lépe definovat nové konformační stavy a také dokázalo nasimulovat možnost dvojí okupance oproti poli OL15 a bsc1, což by mělo být díky novější modifikaci α/γ potenciálu pole OL21 pro vyšší stabilizaci. V polích OL15 a OL21 se v průběhu zbylé simulace navštěvovaly konformace NS2 a NS3, a jen malé procento jiných konformačních stavů. Naopak v silovém poli bsc1 se převládal konformační stav NS3, který se odchyluje od nativních hodnot, přičemž pole jen minimálně konvergovalo k těmto hodnotám po dobu celé MD simulace. Tato skutečnost by mohla být způsobena neschopností silového pole bsc1 adekvátně stabilizovat α/γ potenciál.

Tabulka 9: Procentuální podíl nových konformačních stavů v jednotlivých silových polích.

(α / β)	OL21 [%]	OL15 [%]	bsc1 [%]
NS1(t/high)	69,8	68,0	2,2
NS2 (g-/t)	10,3	5,4	11,8
NS3 (g-/t)	11,1	14,9	49,5
Jiné stavy	8,8	11,8	36,5

Tabulka 10: Procentuální podíl nových konformačních stavů 4U5M v jednotlivých silových polích.

	OL21 [%]			OL15 [%]			bsc1 [%]		
	NS1	NS2	NS3	NS1	NS2	NS3	NS1	NS2	NS3
A_dG5_dG6	85,4	7,8	5,3	84,7	6,7	4,6	0,0	12,8	87,1
A_dT7_dG8	91,1	1,1	3,8	95,3	0,0	0,3	5,4	0,0	4,2
A_dG8_dG9	84,5	10,2	4,3	76,3	12,8	9,6	0,0	9,9	89,4
A_dT10_dG11	87,3*	4,5	19,6	49,4	1,6	17,6	0,0	0,0	0,0
A_dG11_dG12	79,8	13,0	4,8	86,5	2,8	2,5	2,4	26,3	71,0
A_dT16_dG17	34,3	10,9	35,3	13,6	14,4	71,2	1,5	5,2	85,5
A_dG17_dG18	84,2	9,0	3,4	84,5	7,9	3,9	5,2	26,9	67,7
A_dT19_dG20	83,6	2,3	7,8	94,9	0,1	1,2	8,1	0,5	19,4
A_dG20_dG21	81,4	12,8	3,8	87,2	7,4	3,4	0,3	15,0	84,3
A_dT22_dG23	43,1	9,2	33,8	32,8	8,9	51,6	0,2	0,1	9,4
A_dG23_dG24	45,9	42,3	8,1	80,9	0,9	2,5	2,6	44,2	52,0
A_dT25_dG26	37,1	0,1	3,0	29,4	0,9	10,4	0,4	0	24,2

* Populace stavu NS1 zahrnovala χ_1 s hodnotou *anti* i *syn*, viz text.

8.2.4 Konformační stavy interkalovaného duplexu, 1UE2

U struktury interkalovaného duplexu 1UE2 se vyskytovalo několik nedefinovaných NANT stepů, u nichž jsem se rozhodla je nadefinovat novým konformačním stavem. Struktura obsahovala dvě vlákna, která byla identická, z toho důvodu jsem analyzovala pouze jedno a předpokládala jsem obdobné chování obou řetězců. Jednalo se o tři stavy A_C382_dG3, A_dA4_dA5 a A_dA5_dA6, každý z nich měl unikátní hodnoty torzních úhlů, tudíž se muselo nadefinovat více nových konformačních stavů. Nové konformační stavy se získaly z průměru nejpůvodnějších hodnot jednotlivých torzních úhlů, mezi všemi silovými poli. U stavů A_dA4_dA5 a A_dA5_dA6 byla v krystalové struktuře dvojí okupance, tudíž jsem zde pracovala i s možností více okupovaných regionů úhly. Nové konformační stavy jsou uvedeny spolu s nativními hodnotami (i hodnoty okupance) v tabulce (Tabulka 11) níže. Nové konformační stavy se snaží co nejlépe reprezentovat naměřené experimentální geometrie.

Dále bude studován koncový step A_dG7_dC8 definován databází DNATCO jako konformer BB12, kde budou pozorovány hodnoty torze $\alpha/\gamma = g-/t$ a budu zjišťovat jejich stabilitu v průběhu MD simulací.

Tabulka 11: Hodnoty torzních úhlů nových konformačních stavů 1UE2 a hodnoty nativní ve vybraných nekanonických stepech [°].

Step		$\delta 1$	$\chi 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 2$
A_C382_dG3	Nativní	123	254	168	268	310	161	55	149	264
	BB00	138	253	183	258	304	180	44	138	258
A_dA4_dA5	Nativní (A)	88	213	221	299	293	216	57	147	286
	Nativní (B)	90	213	239	296	218	151	164	126	237
	IC01	83	203	220	290	297	223	54	145	283
	NS1	90	217	218	290	179	185	187	115	233
A_dA5_dA6	Nativní (A)	147	286	218	166	293	199	50	145	266
	Nativní (B)	126	237	236	202	290	199	50	145	266
	IC05	141	273	255	174	289	178	50	147	275
	NS2	115	233	238	183	62	194	299	144	271

Stepu A_C382_dG3 byl skriptem přiřazen konformační stav BB00, který je svými hodnotami kanonický, přestože step je databází DNATCO definován jako NANT. To je pravděpodobně způsobeno méně přísným cutoffem pro přiřazení použitým v našem skriptu ve srovnání s obdobným cutoffem v DNATCO. Hodnoty $\alpha/\beta/\gamma$ odpovídají kanonickým *gauche-/trans/gauche+*.

Pro step A_dA4_dA5 byl zaveden nový konformační stav NS1, který by se měl přibližovat (B) nativním hodnotám definujícím step. Torze jsou nekanonické $\alpha/\gamma = t/t$. Zároveň byl stepu skriptem přiřazen již existující NtC konformer IC01.

Obdobně tomu bylo u stepu A_dA5_dA6, kde byl vytvořen nový konformační stav NS2 na základě nalezení druhých nejpobulovanějších regionů, a skriptem byl přiřazen NtC konformer IC05, který odpovídal nativnímu konformeru (A). Konaformační stav NS2 má hodnoty $\alpha/\gamma = g+/g-$ nekanonické.

V tabulce (Tabulka 12) níže je již evidováno výsledné procentuální zastoupení jednotlivých nových konformačních stavů v silových polích. Step A_C382_dG3 jemuž byla přiřazena kanonická konformace BB00 vyšel u všech silových polí nad 70 %.

Silové pole OL21 u stepu A_dA4_dA5 se nejvíce přiblížilo zastoupení nativních stavů A i B (dvojí okupance) a dokázalo tak simulovat i nekanonické hodnoty α/γ . Pole OL15 definovalo tento step spíše na základě konformace IC01. Nejproblematictější situace nastala u silového pole bsc1, které nebylo schopné step popsat vůbec. IC01 je velmi podobná nativní

konformaci (A) v tom smyslu, že všechny torze jsou ve stejných minimech. Hodnoty těchto torzí se ale od IC01 někdy výrazněji liší, což vede k tomu, že struktury v MD jsou ve velké části snímků od referenční IC01 vzdálenější než cutoff. Proto jako IC01 je kategorizována jen část snímků simulace, přičemž další část je sice strukturně blízká, ale již nekategorizovaná .

Poslední nedefinovaný step A_dA5_dA6 byl nejlépe popsán polem OL15 až na 84 % celé simulace. V průběhu simulací se nekanonická konformace, kterou popisuje konformační stav NS2, výrazně neprojevila. Všechna pole z velké většiny populovala existující konformační stav IC05, který obsahoval kanonické hodnoty α/γ , jen v poli bsc1 se v 15 % simulace objevila nekanonická varianta s $\alpha/\gamma = g+/g-$.

Celkově nejlépe dokázalo popsat stepy silové pole OL15. Oproti tomu silové pole bsc1 nedokázalo u dvou stepů ze tří studovaných správně popsat ani polovinu simulace. Připisuji to špatné stabilitě simulace, kterou jsem objevila na základě RMSD a vizuální kontroly ve VMD, což se projevilo na nesprávných populacích stepu A_dA4_dA5, kde v poli bsc1 ve více jak polovině simulace docházelo k odklápění bazí do roztoku, v důsledku čehož nemohlo pole správně popsat jednotlivé torze nativního konformačního stavu.

Tabulka 12: Procentuální zastoupení nových konformačních stavů v jednotlivých silových polích u struktury interkalovaného duplexu..

Step		OL15 [%] Součet		OL21 [%] Součet		Bsc1 [%] Součet	
A_C382_dG3	BB00	70,1		75,0		70,7	
A_dA4_dA5	IC01	43,8	57,8	31,2	55,2	3,1	3,1
	NS1	14,0		24,0		0	
A_dA5_dA6	IC05	75,6	84,0	57,1	63,5	8,2	23,5
	NS2	8,4		11,1		15,3	

Step A_dG7_dC8 byl definován jako konformer BB12, tedy nekanonický konformer obsahující $\alpha/\gamma = g-/t$ (Tabulka 13). Zde bylo hlavní snahou zjistit, zda se daný nekanonický stav udrží po dobu MD simulace a potvrdí tak experimentálně naměřená nekanonická data.

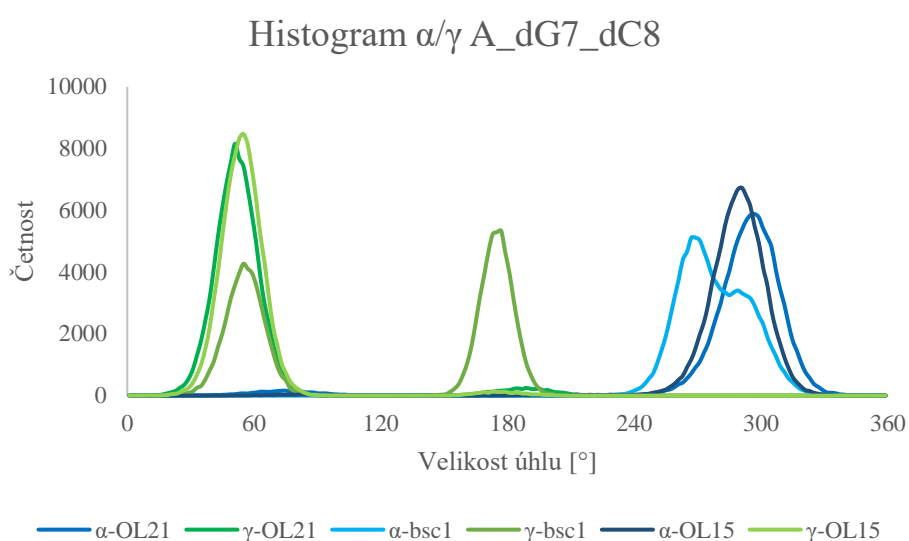
Tabulka 13: Nativní hodnoty torzních úhlů A_dG7_dC8 stepu interkalovaného duplexu [°].

Step	$\delta 1$	$\chi 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 2$
A_dG7_dC8	133	268	195	269	260	69	168	150	206

Pro ověření, zda se stav udrží bylo využito grafů četnosti, které ukazují jednotlivé zastoupení regionů, které torzní úhly α a γ navštěvují v průběhu simulace. V grafu (Graf 5) jsou úhly α vyznačeny modrými barvami a zelenými jsou označeny úhly γ .

Bohužel dokázalo v průběhu simulace udržet nekanonické hodnoty α/γ pouze pole bsc1, zatímco u ostatních silových polí si v průběhu simulace torzní úhly vynutily kanonické hodnoty, ve kterých následně setrvaly po většinu simulace. Výskyt nekanonických hodnot může být dán tím, že je step koncovým reziduem nukleové kyseliny, tudíž může být v krystalu deformován krystalovými kontakty. Tento step ale obsahuje nízkou hodnotu torze β ($\beta/\gamma = g+/t$), přičemž obdobné konformace byly již dříve v práci Chad R. Simmonse et al. [55] označeny za artefakt pole bsc1. To může být důvodem, proč se v poli bsc1 dokázala tato nekanonická konformace udržet.

Graf 5: Histogram α/γ stepu A_dG7_dC8 interkalovaného duplexu v jednotlivých silových polích.



8.2.5 Konformační stavy DNA glykosylázy, 2F5N

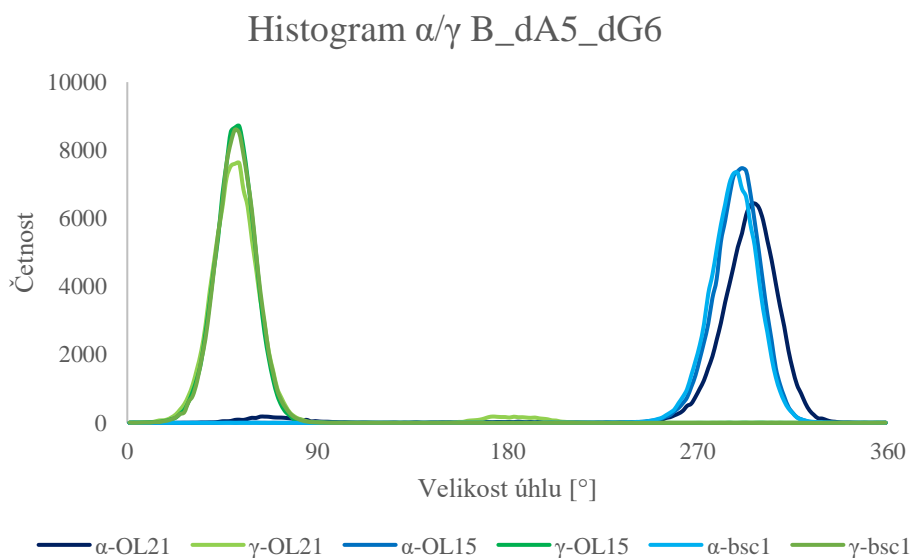
V komplexu DNA glykosylázy bylo vybráno hned několik nekanonických konformací obsahujících α/γ rekonformace. Nachází se zde NtC konformace BB02, BB13, BB12, BB03 a nedefinované NANT. Všechny konformace jsou známé právě tím, že obsahují nekanonické úhly α/γ . Hlavním úkolem zde tedy bylo zjistit, zda se v struktuře tohoto komplexu tyto nekanonické konformace udrží, jak bylo předpovězeno experimentálními daty. Vybrané stepy s jejich nativními hodnotami jsou vypsány v tabulce (Tabulka 14) níže.

Tabulka 14: Nativní hodnoty devíti torzních úhlů vybraných konformačních kroků struktury DNA glykosylázy [°].

NtC	Step	$\delta 1$	$\chi 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 2$
BB02	B_dA5_dG6	142	265	189	242	30	189	303	155	256
BB13	C_dC5_dC6	136	261	196	307	210	69	168	141	232
NANT	C_dA7_dG8	140	296	234	220	130	176	180	148	275
BB12	C_dG9_dT10	133	272	182	289	235	77	171	150	210
NANT	C_dT10_dC11	140	210	211	252	29	151	320	154	256
BB12	C_dT11_dT12	154	256	187	306	229	56	179	142	213
BB03	C_dT12_dA13	142	213	199	278	145	186	176	142	246
BB02	C_dA13_dC14	142	246	193	242	28	150	325	145	270
BB12	C_dC14_dC15	145	270	211	289	247	70	163	156	220

Prvním konformačním krokem byl B_dA5_dG6, který dle databáze DNATCO odpovídá konformeru BB02. Pro analýzu úhlů α/γ bylo využito histogramů (Graf 6), které ukazují, že pole OL21, OL15 ani bsc1 neprokázaly udržení nekanonických hodnot α/γ . Torzní úhly α i γ setrvaly hned od počátku simulace v kanonických *gauche-/gauche+*. Jen minimálně je na grafu vidět, že úhel γ v poli OL21 navštívil i region *trans* a α region *gauche+*, jiná pole tyto stavy nezaznamenala.

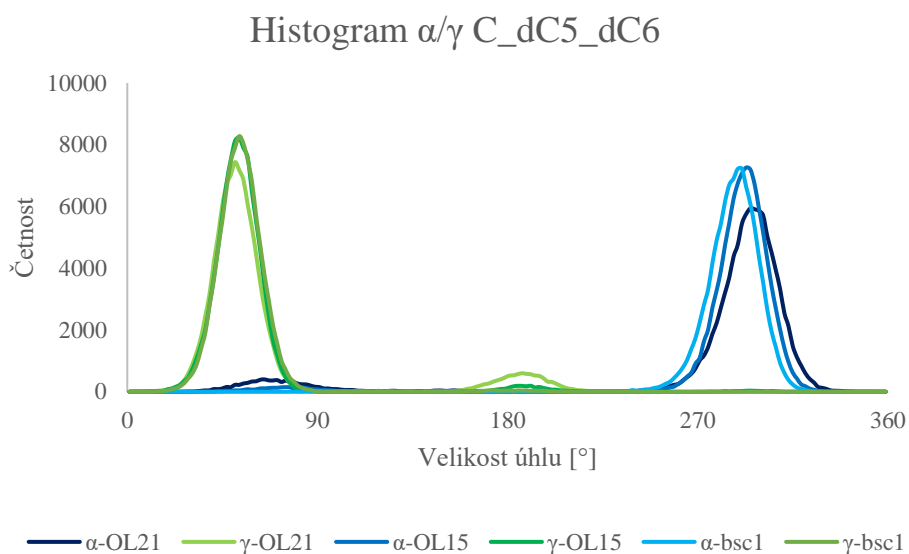
Graf 6: Histogram α/γ stepu B_dA5_dG6 (BB02) DNA glykosylázy.



Dalším konformačním krokem byl C_dC5_dC6 v druhém řetězci komplexu. Zde hodnoty α/γ odpovídaly $\alpha/\gamma = \text{high-trans/trans}$. Hned v prvních snímcích MD simulace se úhly odklonily od nativních hodnot a našly si jiné stavy v kterých setrvaly (Graf 7). V průběhu simulace se zde udržel nekanonický stav o chvíli déle než u předchozího konformačního kroku, ale opět se prosadily kanonické regiony pro úhly α/γ . Nekanonické hodnoty α/γ se

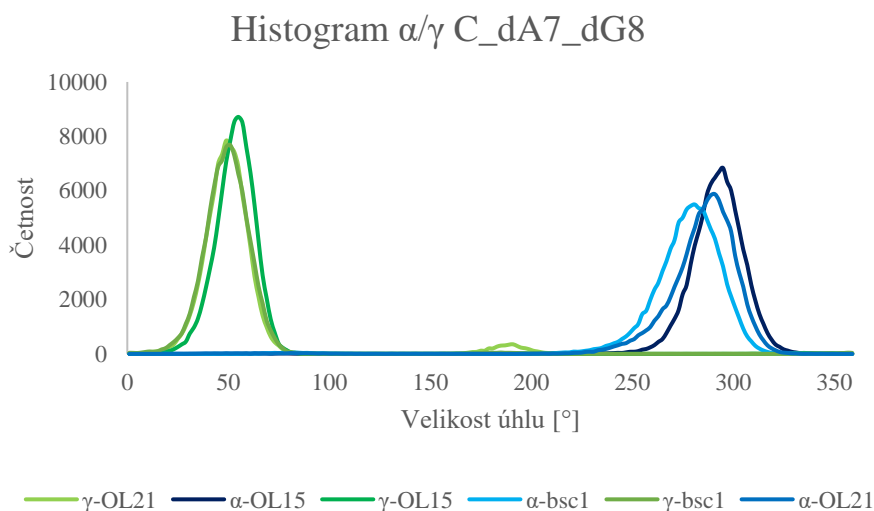
objevily na malou chvíli v polích OL21 a OL15. Silové pole bsc1, nedokázalo udržet nekanonické nativní hodnoty.

Graf 7: Histogram α/γ stepu C_dc5_dc6 (BB13) DNA glykosylázy.



Krok C_da7_dG8 měl hodnoty $\alpha/\gamma = t/t$. Tento konformační krok se se svými nekanonickými hodnotami také neudržel. Opět, jak je patrné z grafu (Graf 8), se pouze v silovém poli OL21 na chvíli udržel nekanonický stav γ *trans*. Z těchto tří histogramů je jasně vidět, že v žádném z uvedených vybraných stepů se nekanonické konformace neudržely na delší dobu MD simulace.

Graf 8: Histogram α/γ stepu C_da7_dG8 DNA glykosylázy.



Celkově se po pečlivém zanalyzování všech histogramů provedených ve všech silových polích neobjevilo jediné udržení nekanonického stavu a z histogramů je viditelné, že si ve struktuře DNA glykosylázy všechny stepy našly regiony s kanonickými hodnotami, ve kterých setrvaly po zbytek simulací. Pouze silové pole OL21 dokázalo alespoň po

minimální čas zaznamenat a udržet tyto konformace, avšak také si posléze našly kanonické hodnoty.

Není vyloučeno, že se zde nekanonické hodnoty nemohou nacházet, protože teoretické výpočty nejsou vždy zcela spolehlivé. Pokud ale tyto stavy nedokázalo udržet ani jedno pole ze tří, přiklání se k tomu, že silová pole modelovala strukturu správně. Neshodu mezi experimentem a simulací tedy připisují možné nepřesné interpretaci experimentálních dat.

8.2.6 Konformační stavy DNA endonukleázy, 2GIG

V tomto komplexu se nacházely tři konformační kroky obsahující nekanonické hodnoty úhlů α/γ . Hlavní snahou zde tedy bylo zjistit, zda se tyto nekanonické konformace α/γ udrží v MD simulacích. Nativní hodnoty jednotlivých konformačních kroků jsou uvedeny v tabulce (Tabulka 15) níže.

Tabulka 15: Nativní hodnoty devíti torzních úhlů vybraných konformačních kroků DNA endonukleázy [°].

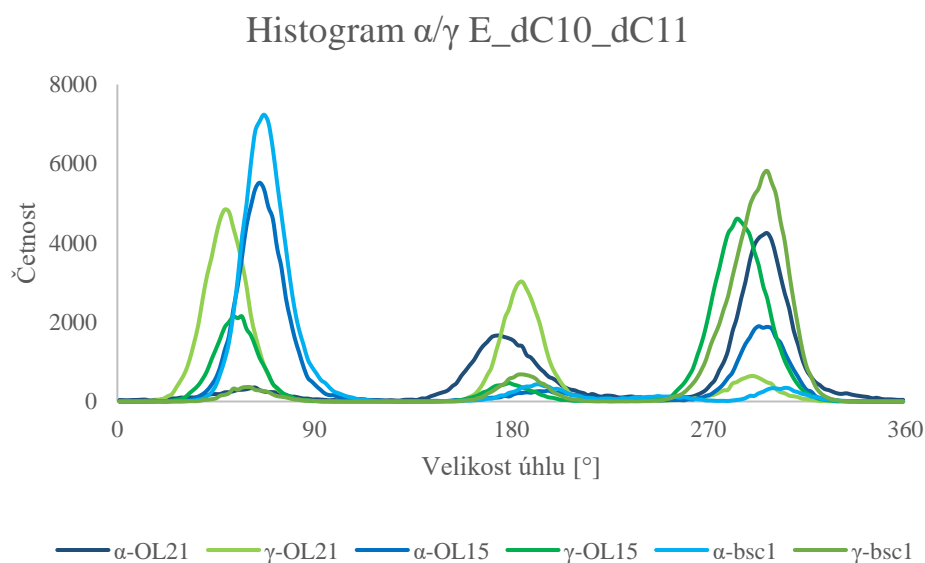
NtC	Krok konformeru	$\delta 1$	$\chi 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 2$
NANT	E_dC10_dC11	147	237	201	303	174	132	188	135	248
NANT	F_dG8_dA9	101	233	2230	295	259	83	167	138	228
NANT	F_dC10_dC11	131	249	210	231	60	179	295	148	269

V prvním stepu E_dC10_dC11 byly hodnoty $\alpha/\gamma = t/t$. Tento step je interkalován fenylalaninem. Pro zjištění, zda se hodnoty udržely v průběhu MD simulace bylo využito histogramů (Graf 9). Úhel α nabýval v průběhu simulace hodnot *gauche+*, *gauche-* i *trans*. V silovém poli OL21 se úhel α vyskytoval převážně v kanonickém regionu *gauche-*, ale i v nekanonickém regionu *trans*. Úhel γ byl v kanonickém regionu *gauche+* a v menším zastoupením v nekanonickém *trans*. V silových polích OL15 a bsc1 se úhly α/γ nacházely ve všech regionech, nejčastěji α/γ nabývaly nekanonických hodnot $\alpha/\gamma = gauche+/gauche-$, ale nenavštěvovaly nativní nekanonické $\alpha/\gamma = trans/trans$ jako v poli OL21. Dalo by se říci, že nekanonické hodnoty mohou být vyvolány interkalací fenylalaninem, kde by tato interkalace mohla vyžadovat nekanonické hodnoty *trans/trans* či *gauche+/gauche-*.

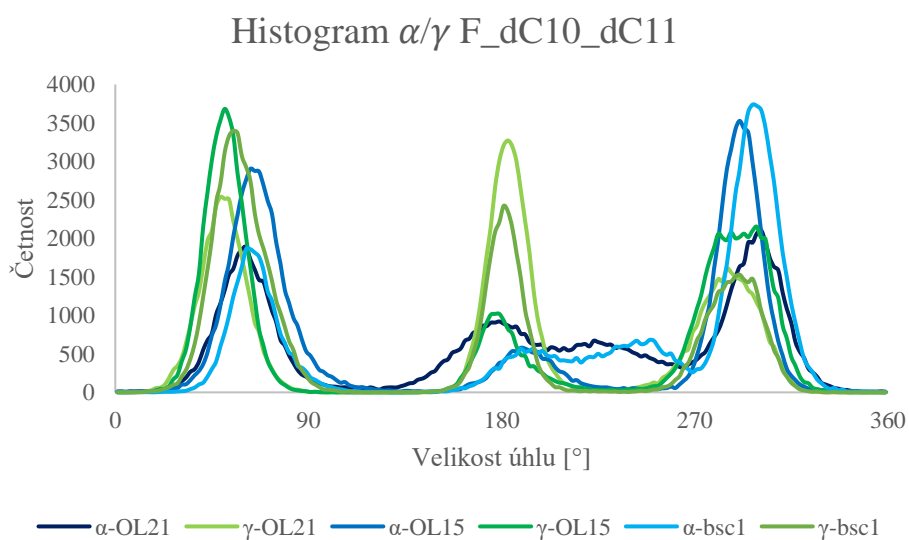
V dalším konformačním kroku F_dC10_dC11 druhého řetězce, který byl taktéž interkalován fenylalaninem, byly hodnoty $\alpha/\gamma = g+/g-$. Histogram (Graf 10) ukazuje, že v silových polích OL21 a bsc1 se torzní úhel γ nacházel nejčastěji v regionu *trans* a *gauche+*, v poli OL15 se úhel γ objevoval ve všech regionech s nejčastějším zastoupením v regionu *gauche+*. Torzní úhel α se pak ve všech třech polích nacházel proměnlivě mezi všemi regiony, z toho nejvíce byl populován v kanonickém regionu *gauche-*. Není tedy vyloučeno, že

interkalace fenylalaninem vyžaduje úhly α/γ v nekanonických hodnotách, kde se také v průběhu simulací objevovaly.

Graf 9: Histogram α/γ v interkalovaném stepu E_dC10_dC11 DNA endonukleázy.



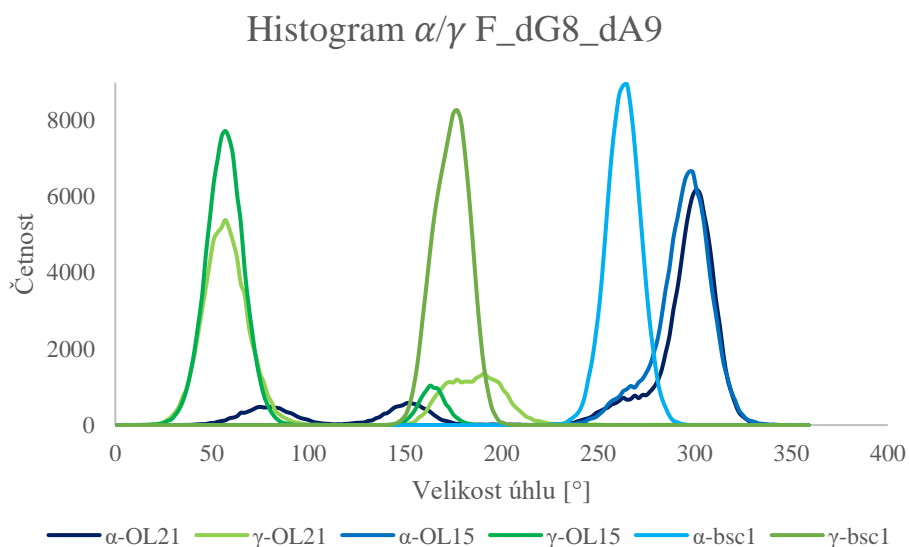
Graf 10: Histogram α/γ v interkalovaném stepu F_dC10_dC11 fenylalaninem DNA endonukleázy.



Následující konformační krok F_dG8_dA9 obsahující $\alpha/\gamma = g-/t$ se nacházel v přítomnosti dvou sodných iontů (aktivní místo). Z grafu (Graf 11) vychází, že nejlépe byla α/γ rekonformace popsána silovým polem bsc1, kde úhly setrvaly v nativních nekanonických hodnotách po celou dobu MD simulace. Torzní úhly α/γ v silovém poli OL21 setrvaly nejčastěji v kanonických hodnotách *gauche-/gauche+*, avšak úhel γ také navštěvoval region nekanonických hodnot *trans*. Silové pole OL15 se zdržovalo

více v hodnotách kanonických regionů a nekanonické hodnoty navštěvovalo pouze minimálně.

Graf 11: Histogram α/γ stepu F_dG8_dA9 v aktivním místě DNA endonukleázy.

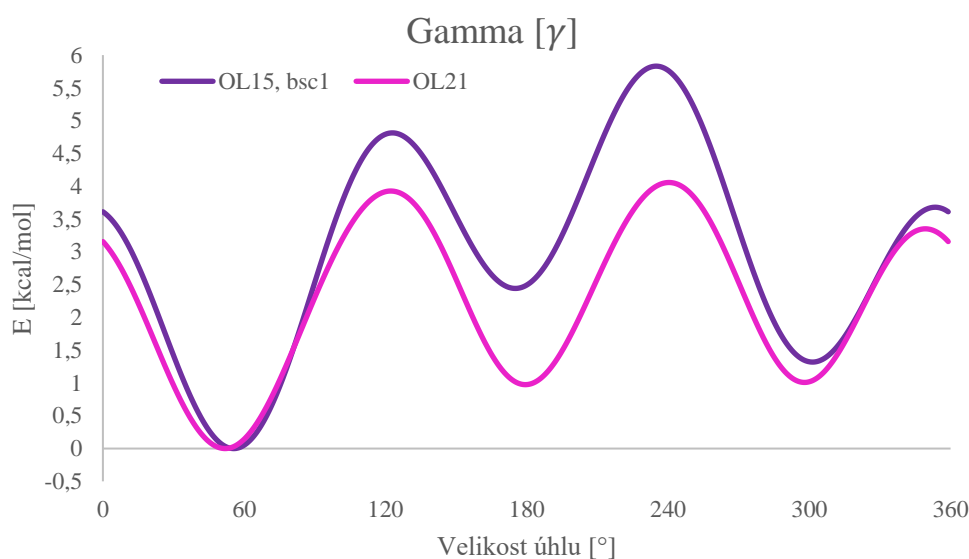


V mé předešlé bakalářské práci^[56] jsem se zabývala jinou modelovou strukturou DNA endonukleázy, kde se taktéž v MD simulaci v aktivním místě objevily nekanonické hodnoty α/γ úhlů v konformaci BB12. Přestože ve zde studované endonukleáze byla konformace neurčená (NANT), hodnoty všech úhlů se blížily konformaci BB12. Po důkladném prohledání databáze DNATCO jsem našla další DNA endonukleázy, které v aktivním místě obsahují nativní nekanonické hodnoty, a to jak BB12, tak i další konformace BB03 a NANT konformace s nekanonickými hodnotami torzí α/γ . Můžeme tedy předpokládat, že tyto nekanonické hodnoty úhlů α/γ se zde opravdu mohou vyskytovat.

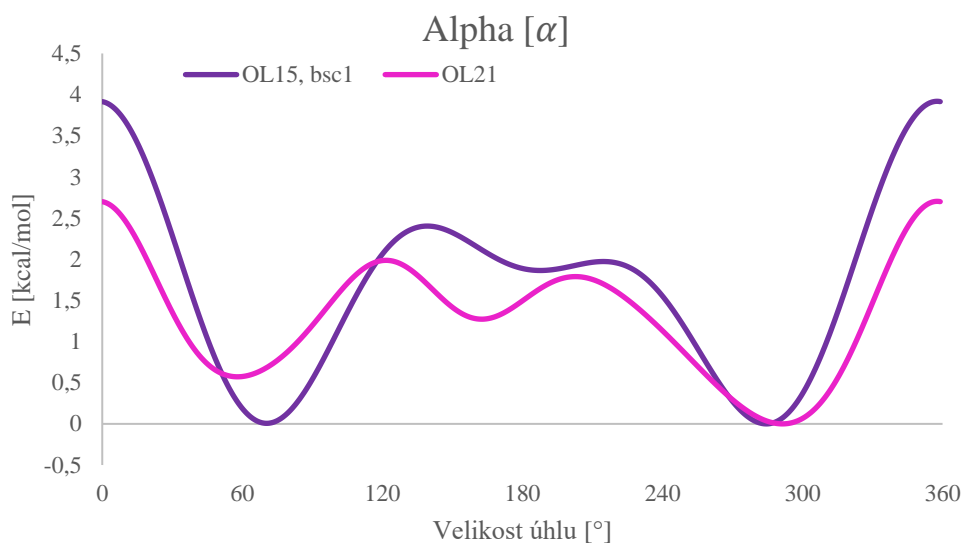
8.3 SHRNU TÍ

Silová pole použita v této práci mají různé popisy potenciálů α a γ . Na grafech (Graf 12 a Graf 13) jsou potenciály α/γ pro silová pole OL21 kde je potenciál α/γ nově modifikovaný, zatímco silová pole OL15 a bsc1 používají stejný starší potenciál původem z bsc0. Potenciály uvedené v grafech značí, jaké regiony jsou energeticky nejvýhodnější pro úhly α/γ . Pro úhel α i γ by měly být nejvýhodnější kanonické regiony, není však vyloučeno, aby oba úhly navštěvovaly i region *trans*. Pro úhel α je energetická bariéra relativně nízká, takže by neměl být problém ji překonat. Obdobně je tomu pro torzní úhel γ , kde je bariéra jen o něco vyšší. Pro oba úhly je patrné, že pravděpodobnost udržení stavu okolo 180° je v modifikovaném poli OL21 vyšší. Z výsledných dat v této práci je zřejmé, že nově modifikovaný potenciál α/γ přispívá k stabilitě torzí α/γ , přestože v některých případech si byla silová pole, i když mají jiné potenciály, velmi podobná a simulovala tedy podobné výsledky.

Graf 12: Potenciál γ silových polí OL21, OL15 a bsc1.



Graf 13: Potenciál α silových polí OL21, OL15 a bsc1.



Konformery vybrané pro tuto práci, obsahovaly v krystalových strukturách neobvyklé nekanonické konformace α/γ torzních úhlů. To, že databáze evidují dané konformery v publikovaných krystalových strukturách, neznamená že zde tyto konformery musí být skutečně stabilní. Molekulově dynamické simulace potvrdily výskyt některých nekanonických konformerů, ale zároveň zpochybnily jejich výskyt v jiných kontextech, například u struktury DNA glykosylázy (2F5N).

V tabulce (Tabulka 16) níže jsou uvedeny nativní hodnoty úhlů α/γ z krystalových a NMR struktur v porovnání s nejčastěji navštěvovanými regiony v průběhu jednotlivých MD simulací ve všech silových polích. Uvedeny jsou i nativní hodnoty dvojích okupancí v krystalových strukturách. Hodnoty získané z průběhu simulací jsou tabelovány dle nejvyšší populace (tj. nejčastěji navštěvovaný region jako první a druhý nejčastěji navštěvovaný region jako druhý, oddělen středníkem). Červeně vyznačené hodnoty jsou totožné s hodnotami nativními. Nejvíce se nativním hodnotám dokázalo přiblížit silové pole OL21 oproti tomu starší pole bsc1 nedokázalo udržet nativní hodnoty skoro v žádném konformačním kroku.

Tabulka 16: Nativní hodnoty α/γ v porovnání s nejčastěji navštěvovanými regiony v průběhu MD simulací. Jsou-li v simulaci populovány dva regiony, je populovanější uveden jako první a méně populovaný jako druhý, oddělený středníkem. Červeně jsou označeny hodnoty shodné s nativními.

PDB ID	NtC	Krok konformeru	Nativní α/γ (A;B)	OL21 α/γ	OL15 α/γ	bsc1 α/γ
1PQT	NANT	A_dA4_dA5	g+/t	g+/t	g+/t	g+/t
4U5M	NANT	A_dG5_dG6	t/g+	t/g+	t/g+	g-/g-
		A_dT7_dG8	t/g+	t/g+	t/g+	g-/g+
		A_dG8_dG9	t/g+	t/g+	t/g+	g-/g-
		A_dT10_dG11	t/g+	t;g-/g+	t;g-/g+	g+/g+
		A_dG11_dG12	t/g+	t/g+	t/g+	g-/g-
		A_dT16_dG17	t/g+	g-/t/g+	g-/g-	g-/g-
		A_dG17_dG18	t/g+	t/g+	t/g+	g-/g-
		A_dT19_dG20	t/g+	t/g+	t/g+	g-/g+
		A_dG20_dG21	t/g+	t/g+	t/g+	g-/g-
		A_dT22_dG23	t/g+	g-/t/g+	g-/t/g-/g+	t/g+
		A_dG23_dG24	t/g+	g-/t/g-/g+	t/g+	g-/g-
A_dT25_dG26	t/g+	t;high/g+	t;g-/g+	g-/g+		
1UE2	NANT	A_C382_dG3	g-/g+	g-/g+	g-/g+	g-/g+
		A_dA4_dA5	g-/g+ ; high/t	g-/t/g+;t	g-/g+	g-/g+
	BB12	A_dA5_dA6	g-/g+ ; g-/t	g-/g+	g-/g+	g-/g-
1ZHU	NANT	A_dG7_dC8	high/t	g-/g+	g-/g+	g-/t;g+
		A_dC6_dA7	high/t	g+/t	g+/t	g-/g+
2F5N	BB02	B_dA5_dG6	g+/g-	g-/g+	g-/g+	g-/g+
	BB13	C_dC5_dC6	high/t	g-/g+	g-/g+	g-/g+
	NANT	C_dA7_dG8	t/t	g-/g+	g-/g+	g-/g+
	BB12	C_dG9_dT10	high/t	g-/g+	g-/g+	g-/g+
	NANT	C_dT10_dC11	g+/g-	g-/g+	g-/g+	g-/g+
	BB12	C_dT11_dT12	high/t	g-/g+	g-/g+	g-/g+
	BB03	C_dT12_dA13	t/t	g-/g+	g-/g+	g-/g+
	BB02	C_dA13_dC14	g+/g-	g-/g+	g-/g+	g-/g+
2GIG	BB12	C_dC14_dC15	high/t	g-/g+	g-/g+	g-/g+
	NANT	E_dC10_dC11	t/t	g-/t/g+;t	g+/g-	g+/g-
	NANT	F_dG8_dA9	high/t	g-/g+;t	g-/g+	g-/g+
NANT	F_dC10_dC11 ^{ad1)}	g+/g-	g-/g+;t	g-/g+;t	g-/g+;t	

Ad1) v silových polích u tohoto konformačního kroku úhly α/γ navštěvovaly všechny regiony

V strukturách hairpinů GAA a GCA (1PQT, 1ZHU) se objevovaly nekanonické rekonformace $\alpha/\gamma = gauche+/trans$. Tyto struktury byly definovány stavy OP19 a nově vytvořeným stavem NS1, kde se stavy lišily pouze v hodnotách ϵ/ζ . Obě struktury byly silovými poli adekvátně popsány, zejména pole OL21 dokázalo nekanonický stav velmi dobře popsat a stabilizovat, díky nové modifikaci α/γ potenciálu oproti polím OL15 a bsc1.

Struktura levotočivého guaninového kvadruplexu obsahovala nekanonickou konformaci $\alpha/\beta = trans/high-trans$. Tuto konformaci definovaly nové konformační stavy NS1, NS2 a NS3, kde NS1 nejlépe reprezentovalo geometrii blízkou krystalové geometrii. Silová pole OL21 a OL15 dokázala udržet stav NS1 v necelých 70 % celé simulace oproti poli bsc1, které nedokázalo správně popsat nativní nekanonické hodnoty. Důvodem optimálnějších výsledků silových polí OL21 a OL15 je modifikace torzního úhlu β .

Molekula interkalovaného duplexu obsahovala několik stepů s nekanonickou α/γ rekonformací. V poli OL21 byla molekula, dle RMSD a vizuální kontroly ve VMD, stabilnější oproti zbylým polím. Nekanonické konformace nejlépe popsal nové silové pole OL21.

Zajímavé je, že v MD simulacích DNA endonukleázy (2GIG) se nekanonické hodnoty α/γ vyskytovaly ve vysoké míře. Můžeme tedy předpokládat, že proteiny typu endonukleázy mohou podporovat výskyt nekanonických hodnot torzních úhlů α/γ . Toto tvrzení bylo potvrzeno i v mé předešlé bakalářské práci u protein-DNA komplexu s PDB ID: 1T9I, což je jiný typ endonukleázy, u které byla pozorována stabilita konformeru BB12, u kterého byly hodnoty úhlů $\alpha/\gamma = high-trans/trans$. Rozhodla jsem se tedy pro prostudování databáze DNATCO, nalezení podobných komplexů obsahujících endonukleázu a zkontrolování hodnot úhlů α/γ , zda neobsahují ve svých nativních geometriích nekanonické hodnoty α/γ . Například u struktur s PDB ID: 1A73, 1R7M, 1BHM, 1CW0 a mnoho dalších, jsem v databázi zaznamenala několik konformačních kroků obsahující buďto NANT konformery, v nichž jsou hodnoty α/γ nekanonické, nebo obsahovaly konformery jako jsou BB12 či BB13. Dalo by se tedy předpokládat, že tyto nekanonické hodnoty α/γ by se měly v aktivním místě opravdu nacházet a pro enzym endonukleázu by mohly být strukturně a funkčně důležité.

Komplex DNA glykosylázy (2F5N) nedokázala silová pole udržet v nativních nekanonických hodnotách ani u jednoho stepu. V každém stepu došlo v MD simulaci k rekonformaci do klasických kanonických hodnot úhlů $\alpha/\gamma = g-/g+$. Protože se toto chování projevilo u všech použitých silových polí, je zde vysoká pravděpodobnost, že by u tohoto komplexu mohly tyto kanonické hodnoty dobře reprezentovat realitu. Je tedy možné, že

experimentální hodnoty pro tento komplex mohly být určeny nepřesně. Obdobně tomu bylo i v mé předešlé bakalářské práci, kde jsem studovala dvě konformační rodiny BB15 a BB12, kde u tří modelových komplexů protein-DNA byl NtC konformer BB15 v simulacích nestabilní, což jsme vysvětlili možnou chybou v experimentálním určení geometrie. ^[56]

9 ZÁVĚR

Hlavním cílem této práce bylo prozkoumat stabilitu a kvalitu popisu nekanonických motivů cukrfofátové páteře ve třech současných empirických potenciálech z rodiny AMBER: bsc1, OL15 a OL21. Zaměřila jsem se při tom na konformery obsahující nekanonické hodnoty torzních úhlů α nebo γ , protože nejnovější z testovaných silových polí, OL21, obsahuje modifikaci právě těchto torzních úhlů, která by měla stabilizovat především $\alpha = t$ a $\gamma = t$ konformace. Bylo vybráno několik konformerů z databáze DNATCO, konkrétně NtC konformery BB02, BB03, BB12 a BB13, obsahující nekanonické kombinace torzních úhlů α nebo γ . Kromě nich jsem ale do studie zahrnula i několik dalších, v databázi DNATCO neurčených (NANT) konformerů, které také obsahovaly nekanonické kombinace torzních úhlů α/γ . Pro testovací simulace byly vybrány čtyři neobvyklé molekuly DNA, a to levotočivý guaninový kvadruplex, GAA hairpin, GCA hairpin a interkalovaný duplex a také dva komplexy proteinu s DNA, a to DNA glykosyláza a DNA endonukleáza s navázanými DNA duplexy.

Ve většině z provedených simulací se ukázalo, že nové silové pole OL21 s modifikovaným potenciálem α/γ popisuje nekanonické stavy páteře lépe, než starší silová pole bsc1 a OL15, která sdílejí stejný původní α/γ potenciál. V molekule levotočivého guaninového kvadruplexu (4U5M) popsalo silové pole OL21 všech 12 neurčených (NANT) stavů páteře jako stabilní s geometrií blízkou krystalové geometrii. Protože silové pole OL15 poskytlo podobně dobré výsledky a pole bsc1 naopak nedokázalo nativní stav stabilizovat, domnívám se, že důvodem zlepšení je modifikace torzního úhlu β v OL15 a OL21. V molekulách hairpinů GAA a GCA poskytlo silové pole OL21 průměrně větší populace nativních nekanonických stavů než bsc1 i OL15. Konečně, v molekule interkalovaného duplexu jsem pozorovala větší procentuální populaci nativních nekanonických α/γ stavů v poli OL21 než ve zbylých dvou variantách a simulovaná molekula byla v OL21 také viditelně stabilnější.

Schopnost silových polí populovat nativní nekanonické α/γ stavy páteře se potvrdila i v komplexu DNA endonukleázy 2GIG, kde byla pozorována zvýšená populace nekanonických α/γ stavů jednak v oblasti aktivního místa komplexu a také v oblasti interkalace fenylalaninem. Zatímco silové pole OL21 předpovídalo vyšší populace nekanonických stavů v interkalované struktuře, v aktivním místě předpovídalo větší populace pole bsc1. Podobné nekanonické stavy jsem pozorovala již ve své bakalářské práci v aktivním místě jiného typu endonukleázy a z databázové studie vyplývá, že jsou nativně přítomny i v dalších

endonukleázách. Dá se tedy předpokládat, že tyto nekanonické α/γ stavy (v aktivním místě i v oblasti interkalace) jsou důležité pro strukturu a funkci endonukleáz.

Naopak u DNA glykosylázy (2GIG) byly všechny nekanonické α/γ stavy nestabilní ve všech testovaných silových polích. Podobné chování jsem pozorovala již ve své bakalářské práci u tří dalších komplexů proteinů s DNA. Proto se domnívám, že v těchto případech popisovala silová pole páteř DNA správně a nekanonické stavy α/γ byly v krystalových strukturách určeny nepřesně.

Dalším cílem bylo pokusit se popsat, pokud možno také charakterizovat v databázi DNATCO neurčené (NANT) konformační stavy ve výše uvedených molekulách DNA. V několika případech se to podařilo a v MD simulacích jsem pozorovala stabilní dosud nepopsané konformery, které byly svou geometrií blízké výchozím nativním krystalovým nebo NMR strukturám. Tyto konformery jsem pojmenovala zkratkami NSx, jako Nová Struktura s pořadovým číslem x. Celkem byly identifikovány 6 nové konformačních stavy, z nichž tři se vyskytovaly v levotočivém guaninovém kvadruplexu, jeden v molekulách GAA a GCA hairpinů a dva v interkalovaném duplexu. Tyto nové struktury by mohly být vzaty jako základ pro rozšíření databáze DNATCO o další přirozeně se vyskytující, dobře charakterizované a stabilní NtC konformery.

10 SUMMARY

The main goal of this work was to investigate the stability and quality of the description of non-canonical sugar-phosphate backbone motifs in three recent empirical potentials from the AMBER family: bsc1, OL15 and OL21. In doing so, I focused on conformers containing non-canonical values of the torsion angles α or γ , since the newest of the force fields tested, OL21, contains a modification of these very torsion angles that should stabilize in particular the $\alpha = t$ and $\gamma = t$ conformations. Several conformers were selected from the DNATCO database, namely the NtC conformers BB02, BB03, BB12 and BB13, containing non-canonical combinations of torsional angles α or γ . However, in addition to these, I also included in the study several other unassigned (NANT) conformers in the DNATCO database that also contained non-canonical combinations of α/γ torsion angles. Four unusual DNA molecules, namely the left-handed guanine quadruplex, the GAA hairpin, the GCA hairpin and the intercalated duplex, as well as two protein-DNA complexes, namely DNA glycosylase and DNA endonuclease with bound DNA duplexes, were selected for the test simulations.

In most of the simulations performed, the new OL21 force field with a modified α/γ potential was found to describe non-canonical backbone states better than the older bsc1 and OL15 force fields, which share the same original α/γ potential. In a left-handed guanine quadruplex molecule (4U5M), the OL21 force field described all 12 unassigned (NANT) backbone states as stable with a geometry close to crystal geometry. Since the OL15 force field provided similarly good results and the bsc1 field, on the other hand, failed to stabilize the native state, I suggest that the reason for the improvement is due to the modification of the torsion angle β in OL15 and OL21. In the GAA and GCA hairpin molecules, the OL21 force field provided, on average, a larger population of native noncanonical states than both bsc1 and OL15. Finally, in the intercalated duplex molecule, I observed a larger percentage population of native noncanonical α/γ states in the OL21 field than in the other two variants, and the simulated molecule was also visibly more stable in OL21.

The ability of the force fields to populate native non-canonical α/γ backbone states was confirmed in the 2GIG DNA endonuclease complex, where an increased population of non-canonical α/γ states was observed both in the active site region of the complex and in the phenylalanine intercalation region. While the OL21 force field predicted a higher population of non-canonical states in the intercalated structure, the bsc1 field predicted a higher population in the active site. Similar non-canonical states were observed in my undergraduate thesis in the active site of another type of endonuclease, and the database study suggests that

they are natively present in other endonucleases. Thus, it can be assumed that these non-canonical α/γ states (both in the active site and in the intercalation region) are important for the structure and function of endonucleases.

In contrast, for DNA glycosylase (2GIG), all non-canonical α/γ states were unstable in all force fields tested. Similar behavior was observed in my undergraduate thesis for three other protein-DNA complexes. Therefore, I believe that in these cases the force fields described the DNA backbone correctly and the non-canonical α/γ states were inaccurately determined in the crystal structures.

Another goal was to try to describe, if possible also to characterize in the DNATCO database, the unassigned (NANT) conformational states in the above DNA molecules. In several cases this was successful and in MD simulations I observed stable previously undescribed conformers that were close in geometry to the initial native crystal or NMR structures. I named these conformers with the abbreviations NS x , as New Structure with serial number x . In total, six new conformational states were identified, three of which occurred in the left-handed guanine quadruplex, one in the GAA and GCA hairpin molecules, and two in the intercalated duplex. These new structures could be taken as a basis for expanding the DNATCO database to include other naturally occurring, well-characterized and stable NtC conformers.

11 REFERENCE

- [1] Černý, J., Božíková, P., Svoboda, J., & Schneider, B. (2020). **A unified dinucleotide alphabet describing both RNA and DNA structures.** *Nucleic Acids Research*, 48(11), 6367–6381. <https://doi.org/10.1093/nar/gkaa383>
- [2] Zgarbová, M., Šponer, J., Otyepka, M., Cheatham, T. E., Galindo-Murillo, R., & Jurečka, P. (2015). **Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA.** *Journal of Chemical Theory and Computation*, 11(12), 5723–5736. <https://doi.org/10.1021/acs.jctc.5b00716>
- [3] Banáš, P., Hollas, D., Zgarbová, M., Jurečka, P., Orozco, M., Cheatham, T. E., Šponer, J., & Otyepka, M. (2010). Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *Journal of Chemical Theory and Computation*, 6(12), 3836–3849. <https://doi.org/10.1021/ct100481h>
- [4] Zgarbová, M., Šponer, J., & Jurečka, P. (2021). **Z-DNA as a Touchstone for Additive Empirical Force Fields and a Refinement of the Alpha/Gamma DNA Torsions for AMBER.** *Journal of Chemical Theory and Computation*, 17(10), 6292–6301. <https://doi.org/10.1021/acs.jctc.1c00697>
- [5] Watson J. D., Crick F. H. C. (1953) **The structure of DNA.** *Cold Spring Harbor Symposia on Quantitative Biology*. 1953, roč. 18, s. 123-131. ISSN 0091-7451. <https://doi.org/10.1101/SQB.1953.018.01.020>
- [6] Chaffey, N. (2003). Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. **Molecular biology of the cell. 4th edn.** *Annals of Botany*, 91(3), 401–401. <https://doi.org/10.1093/aob/mcg023>
- [7] Packer, M. J., & Hunter, C. A. (1998). **Sequence-dependent DNA structure: the role of the sugar-phosphate backbone** 1. Edited by T. Richmond. *Journal of Molecular Biology*, 280(3), 407–420. <https://doi.org/10.1006/jmbi.1998.1865>
- [8] Zimmerman, S. B. (1982). **The Three-Dimensional Structure of DNA.** Annual Review of Biochemistry, 51(1), 395–427. <https://doi.org/10.1146/annurev.bi.51.070182.002143>
- [9] Schneider, B., Božíková, P., Nečasová, I., Čech, P., Svozil, D., & Černý, J. (2018). **A DNA structural alphabet provides new insight into DNA flexibility.** *Acta Crystallographica Section D Structural Biology*, 74(1), 52–64. <https://doi.org/10.1107/S2059798318000050>
- [10] Černý, J., Božíková, P., Svoboda, J., & Schneider, B. (2020). **A unified dinucleotide alphabet describing both RNA and DNA structures.** *Nucleic Acids Research*, 48(11), 6367–6381. <https://doi.org/10.1093/nar/gkaa383>

- [11] Travers, A. A. (1989). **DNA conformation and protein binding**. *Annual Review of Biochemistry*, 58(1), 427–452. <https://doi.org/10.1146/annurev.bi.58.070189.002235>
- [12] Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981). **Structure of a B-DNA dodecamer: conformation and dynamics**. *Proceedings of the National Academy of Sciences*, 78(4), 2179–2183. <https://doi.org/10.1073/pnas.78.4.2179>
- [13] Vargason, J. M., Henderson, K., & Ho, P. S. (2001). **A crystallographic map of the transition from B-DNA to A-DNA**. *Proceedings of the National Academy of Sciences*, 98(13), 7265–7270. <https://doi.org/10.1073/pnas.121176898>
- [14] Jurečka, P., Zgarbová, M., Černý, F., & Salomon, J. (2022). **Continuous B- to A-Transition in Protein-DNA Binding - How Well Is It Described by Current AMBER Force Fields?** <https://doi.org/10.1101/2022.01.13.476176>
- [15] Schneider, B., Černý, J., Svozil, D., Čech, P., Gelly, J.-C., & de Brevern, A. G. (2014). **Bioinformatic analysis of the protein/DNA interface**. *Nucleic Acids Research*, 42(5), 3381–3394. <https://doi.org/10.1093/nar/gkt1273>
- [16] Efimov, A. V. (1993). **Standard structures in proteins**. *Progress in Biophysics and Molecular Biology*, 60(3), 201–239. [https://doi.org/10.1016/0079-6107\(93\)90015-C](https://doi.org/10.1016/0079-6107(93)90015-C)
- [17] Branden, C. I., & Tooze, J. (2012). **Introduction to Protein Structure**. *Garland Science*. <https://doi.org/10.1201/9781136969898>
- [18] Voet, D., Voetová, J. G. **Biochemie. 1. vyd.** Praha: Victoria Publishing, 1995. 1325 s. ISBN 80-856-0544-9
- [19] Eun, H.-M. (1996). **Enzymes and Nucleic Acids**. In *Enzymology Primer for Recombinant DNA Technology* (pp. 1–108). Elsevier. <https://doi.org/10.1016/B978-012243740-3/50004-1>
- [20] Belmont, P., Constant, J. F., & Demeunynck, M. (2001). **Nucleic acid conformation diversity: from structure to function and regulation**. *Chemical Society Reviews*, 30(1), 70–81. <https://doi.org/10.1039/a904630e>
- [21] Padrta, P., Štefl, R., Králík, L., Žídek, L., & Sklenář, V. (2002). **Nuclease resistance of an extraordinarily thermostable mini-hairpin DNA fragment, d(GCGAAGC) and its application to in vitro protein synthesis**. *Journal of Biomolecular NMR*, 24(1), 1–14. <https://doi.org/10.1023/a:1020632900961>
- [22] Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K., & Miura, K. (1994). **Most compact hairpin-turn structure exerted by a short DNA fragment**,

- d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat.** *Nucleic Acids Research*, 22(4), 576–582. <https://doi.org/10.1093/nar/22.4.576>
- [23] Padrta, P., Štefl, R., Králík, L., Žídek, L., & Sklenář, V. (2002). **Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings.** *Journal of Biomolecular NMR*, 24(1), 1–14. <https://doi.org/10.1023/A:1020632900961>
- [24] Chung, W. J., Heddi, B., Schmitt, E., Lim, K. W., Mechulam, Y., & Phan, A. T. (2015). **Structure of a left-handed DNA G-quadruplex.** *Proceedings of the National Academy of Sciences of the United States of America*, 112(9), 2729–2733. <https://doi.org/10.1073/PNAS.1418718112/-/DCSUPPLEMENT>
- [25] Lipps, H. J., & Rhodes, D. (2009). **G-quadruplex structures: in vivo evidence and function.** *Trends in Cell Biology*, 19(8), 414–422. <https://doi.org/10.1016/j.tcb.2009.05.002>
- [26] Sunami, T., Kondo, J., Hirao, I., Watanabe, K., Miura, K., & Takénaka, A. (2004). **Structure of d(GCGAAAGC) (hexagonal form): a base-intercalated duplex as a stable structure.** *Acta Crystallographica Section D Biological Crystallography*, 60(1), 90–96. <https://doi.org/10.1107/S0907444903024703>
- [27] Kondo, J., Umeda, S. -i., Fujita, K., Sunami, T., & Takenaka, A. (2003). **The base-intercalated duplexes of d(GCGAXAGC) with mutation at X (X=G, T or C).** *Nucleic Acids Symposium Series*, 3(1), 175–176. <https://doi.org/10.1093/nass/3.1.175>
- [28] Zhu, L., Chou, S.-H., Xu, J., & Reid, B. R. (1995). **Structure of a single-cytidine hairpin loop formed by the DNA triplet GCA.** *Nature Structural & Molecular Biology*, 2(11), 1012–1017. <https://doi.org/10.1038/nsb1195-1012>
- [29] Yoshizawa, S., Ueda, T., Ishido, Y., Miura, K., Watanabe, K., & Hirao, I. (1994). **Nuclease resistance of an extraordinarily thermostable mini-hairpin DNA fragment, d(GCGAAGC) and its application to in vitro protein synthesis.** *Nucleic Acids Research*, 22(12), 2217–2221. <https://doi.org/10.1093/nar/22.12.2217>
- [30] FOJTA, Miroslav. (2013) *Interakce DNA s proteiny*. Masarykova Univerzita, z: https://is.muni.cz/el/sci/podzim2013/Bi7015/um/zkrac-prot_noncovalent.pdf.
- [31] Stormo, G. D., & Zhao, Y. (2010). **Determining the specificity of protein–DNA interactions.** *Nature Reviews Genetics*, 11(11), 751–760. <https://doi.org/10.1038/nrg2845>
- [32] Frykholm, K., Nyberg, L. K., & Westerlund, F. (2017). **Exploring DNA–protein interactions on the single DNA molecule level using nanofluidic tools.** *Integrative Biology*, 9(8), 650–661. <https://doi.org/10.1039/C7IB00085E>

- [33] Revzin, Arnold. (1990). **The Biology of nonspecific DNA-protein interactions** (Arnold. Revzin, Ed.) [Book]. CRC Press.
- [34] Steffen, N. R., Murphy, S. D., Toller, L., Hatfield, G. W., & Lathrop, R. H. (2002). **DNA sequence and structure: direct and indirect recognition in protein-DNA binding.** *Bioinformatics*, 18(suppl_1), S22–S30.
https://doi.org/10.1093/bioinformatics/18.suppl_1.S22
- [35] Mládek, A., Šponer, J. E., Jurečka, P., Banáš, P., Otyepka, M., Svozil, D., & Šponer, J. (2010). **Conformational Energies of DNA Sugar–Phosphate Backbone: Reference QM Calculations and a Comparison with Density Functional Theory and Molecular Mechanics.** *Journal of Chemical Theory and Computation*, 6(12), 3817–3835.
<https://doi.org/10.1021/ct1004593>
- [36] Banerjee, A., Santos, W. L., & Verdine, G. L. (2006). **Structure of a DNA Glycosylase Searching for Lesions.** *Science*, 311(5764), 1153–1157.
<https://doi.org/10.1126/science.1120288>
- [37] Jacobs, A. L., & Schär, P. (2012). **DNA glycosylases: in DNA repair and beyond.** *Chromosoma*, 121(1), 1–20. <https://doi.org/10.1007/s00412-011-0347-4>
- [38] Joshi, H. K., Etkorn, C., Chatwell, L., Bitinaite, J., & Horton, N. C. (2006). **Alteration of Sequence Specificity of the Type II Restriction Endonuclease HincII through an Indirect Readout Mechanism.** *Journal of Biological Chemistry*, 281(33), 23852–23869.
<https://doi.org/10.1074/jbc.M512339200>
- [39] KOLAFA, Jiří. (2023) *Molekulové modelování a simulace*. Ústav Fyzikální chemie, VŠCHT Praha.
- [40] Hollingsworth, S. A., & Dror, R. O. (2018). **Molecular Dynamics Simulation for All.** *Neuron*, 99(6), 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
- [41] JUNGWIRTH, Pavel. *Klasická a kvantová molekulová dynamika*. Ústav organické chemie a biochemie Akademie věd České republiky.
- [42] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzes, G. Giamba, su, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O’Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J.

- Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, Q. Zhu, and P.A. Kollman (2023), **Amber 2023**, University of California, San Francisco.
- [43] Mayo, S. L., Olafson, B. D., & Goddard, W. A. (1990). **DREIDING: a generic force field for molecular simulations**. *The Journal of Physical Chemistry*, 94(26), 8897–8909. <https://doi.org/10.1021/j100389a010>
- [44] Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). **An all atom force field for simulations of proteins and nucleic acids**. *Journal of Computational Chemistry*, 7(2), 230–252. <https://doi.org/10.1002/jcc.540070216>
- [45] Yang, L., Tan, C., Hsieh, M.-J., Wang, J., Duan, Y., Cieplak, P., Caldwell, J., Kollman, P. A., & Luo, R. (2006). **New-Generation Amber United-Atom Force Field**. *The Journal of Physical Chemistry B*, 110(26), 13166–13176. <https://doi.org/10.1021/jp060163v>
- [46] Souza, P. C. T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünewald, F., Patmanidis, I., Abdizadeh, H., Bruininks, B. M. H., Wassenaar, T. A., Kroon, P. C., Melcr, J., Nieto, V., Corradi, V., Khan, H. M., Domański, J., Javanainen, M., Martinez-Seara, H., Reuter, N., ... Marrink, S. J. (2021). **Martini 3: a general purpose force field for coarse-grained molecular dynamics**. *Nature Methods*, 18(4), 382–388. <https://doi.org/10.1038/s41592-021-01098-3>
- [47] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- [48] Berendsen, H. J. C., Grigera, J. R., & Straatsma, T. P. (1987). **The missing term in effective pair potentials**. *The Journal of Physical Chemistry*, 91(24), 6269–6271. <https://doi.org/10.1021/j100308a038>
- [49] Li, P., Roberts, B. P., Chakravorty, D. K., & Merz, K. M. (2013). **Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent**. *Journal of Chemical Theory and Computation*, 9(6), 2733–2748. <https://doi.org/10.1021/ct400146w>
- [50] Hopkins, C. W., le Grand, S., Walker, R. C., & Roitberg, A. E. (2015). **Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning**. *Journal of Chemical Theory and Computation*, 11(4), 1864–1874. <https://doi.org/10.1021/ct5010406>
- [51] Joung, I. S., & Cheatham, T. E. (2008). **Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations**.

- The Journal of Physical Chemistry B*, 112(30), 9020–9041.
<https://doi.org/10.1021/jp8001614>
- [52] Peters, M. B., Yang, Y., Wang, B., Füsti-Molnár, L., Weaver, M. N., & Merz, K. M. (2010). **Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF)**. *Journal of Chemical Theory and Computation*, 6(9), 2935–2947. <https://doi.org/10.1021/ct1002626>
- [53] Li, H., Robertson, A. D., & Jensen, J. H. (2005). **Very fast empirical prediction and rationalization of protein pK_a values**. *Proteins: Structure, Function, and Bioinformatics*, 61(4), 704–721. <https://doi.org/10.1002/prot.20660>
- [54] Kim, M. O., Nichols, S. E., Wang, Y., & McCammon, J. A. (2013). **Effects of histidine protonation and rotameric states on virtual screening of M. tuberculosis RmlC**. *Journal of Computer-Aided Molecular Design*, 27(3), 235–246. <https://doi.org/10.1007/s10822-013-9643-9>
- [55] Simmons, C. R., MacCulloch, T., Krepl, M., Matthies, M., Buchberger, A., Crawford, I., Šponer, J., Šulc, P., Stephanopoulos, N., & Yan, H. (2022). **The influence of Holliday junction sequence and dynamics on DNA crystal self-assembly**. *Nature Communications*, 13(1), 3112. <https://doi.org/10.1038/s41467-022-30779-6>
- [56] KOLÁŘOVÁ, Elizabeth. (2022). *Minoritní stavy páteře B-DNA a jejich vliv na interakci s proteiny*. Bakalářská práce. Olomouc: Univerzita Palackého.