

# UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta  
Katedra fyzikální chemie



## Minoritní stavy páteře B-DNA a jejich vliv na interakci s proteiny

### BAKALÁŘSKÁ PRÁCE

Autor:	<b>Elizabeth Kolářová</b>
Studijní program:	B1407 – Chemie
Studijní obor:	Aplikovaná chemie
Forma studia:	Prezenční
Vedoucí práce:	<b>doc. RNDr. Petr Jurečka, Ph.D.</b>
Rok:	2022

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 autora	Elizabeth Kolářová
Název práce	Minoritní stavy páteře B-DNA a jejich vliv na interakci s proteiny
Typ práce	Bakalářská
Pracoviště	Katedra fyzikální chemie
Vedoucí práce	doc. RNDr. Petr Jurečka, Ph.D.
Rok obhajoby práce	2022
Klíčová slova	DNA, protein, molekulová dynamika, simulace, interakce, konformer, stabilita
Počet stran	57
Jazyk	Český

## **Bibliographical identification**

Autor's first name and surname	Elizabeth Kolářová
Title	Minority states of the B-DNA backbone and their effect on protein interaction
Type of thesis	Bachelor
Department	Department of physical chemistry
Supervisor	doc. RNDr, Petr Jurečka, Ph.D.
The year of presentation	2022
Keywords	DNA, protein, molecular dynamic, simulation, interaction, conformer, stability
Number of pages	57
Language	Czech

## **Abstrakt**

Tato bakalářská práce se zabývá studiem dvou nekanonických konformerů páteře DNA, které jsou relativně často zastoupeny v komplexech DNA s proteiny. Celkově jsem se zaměřila na čtyři zástupce protein-DNA komplexů, kteří ve své struktuře obsahují alespoň jeden takovýto nekanonický konformer. Ověřuji zde jejich stabilitu a chování v molekulově dynamických (MD) simulacích a zaměřuji se na to, jestli je popis jejich geometrie a stability ve shodě s popisem v krystalografických datech. Dále s pomocí MD simulací zjišťuji, jak ovlivňují interakce s proteinem konformaci páteře DNA. Simulace byly prováděny ve třech silových polích z rodiny AMBER, OL15, bsc1 a nově vyvinutém poli OL21. Výsledky ve všech silových polích shodně ukazují, že dva studované stavy se výrazně liší svou stabilitou. Zatímco jeden z nich je v přítomnosti proteinu zcela stabilní, druhý není stabilní v žádné ze studovaných struktur. Moje výsledky mohou pomoci zpřesnit popis nekanonických konformerů v krystalografických databázích.

## **Abstract**

This bachelor thesis focuses on study of two non-canonical DNA backbone conformers, that are relatively frequently represented in DNA-protein complexes. Overall, I have focused on four representatives of protein-DNA complexes, that contain at least one such non-canonical conformer in their structure. Here I verify their stability and behavior in molecular dynamics (MD) simulations and focus on whether the description of their geometry and stability is consistent with that in crystallographic data. Furthermore, I use MD simulations to determine how interactions with the protein affect the conformation of the DNA backbone. The simulations were performed in three force fields from the AMBER family, OL15, bsc1 and the newly developed OL21. The results in all force fields consistently show that the two studied states differ significantly in their stability. While one is completely stable in the presence of the protein, the other is not stable in any of the studied structures. My results may help to refine the description of non-canonical conformers in crystallographic databases.

## Obsah

1. ÚVOD.....	9
2. STRUKTURA .....	10
2.1. Struktura proteinu.....	10
2.1.1. Primární struktura.....	10
2.1.2. Sekundární struktura .....	10
2.1.3. Dihedrální úhel.....	10
2.1.4. Terciární struktura.....	12
2.1.5. Kvartérní struktura .....	12
2.2. Struktura DNA .....	12
2.2.1. Struktura A-, B- a Z-DNA .....	13
3. INTERAKCE PROTEINŮ S DNA .....	15
3.1. Specifická a nespecifická interakce proteinu s DNA .....	15
4. DNA KONFORMERY .....	17
4.1. Konformery BB00, BB15 a BB12 .....	18
5. KOMPLEXY PROTEIN-DNA .....	19
5.1. 3JTG, Ets doména vázající TGF- $\beta$ receptor.....	19
5.2. 1ZS4, $\lambda$ cII-aktivátor transkripce.....	21
5.3. 1G2F, TATA box komplex .....	22
5.4. 1T9I, LAGLIDADG homing endonukleáza.....	23
6. CÍLE PRÁCE.....	25
7. METODY .....	26
7.1. Silová pole a parametrizace.....	26
7.2. Příprava výchozích struktur.....	28
7.2.1. Protonizační stavy histidinu.....	30
8. VÝSLEDKY A DISKUZE.....	31
8.1. Stabilita MD simulací.....	31
8.1.1. Ets doména vázající TGF- $\beta$ receptor, 3JTG .....	31
8.1.2. $\lambda$ cII-aktivátor transkripce, 1ZS4 .....	34
8.1.3. TATA box komplex, 1G2F.....	36
8.1.4. LAGLIDADG homing endonukleáza, 1T9I .....	37
8.2. Stabilita konformerů.....	39
8.2.1. Stabilita konformeru BB15 .....	40
8.2.2. Stabilita konformeru BB12 .....	44
8.3. MM potenciál a referenční křivka QM.....	46

9.	ZÁVĚR .....	50
10.	SUMMARY .....	52
11.	REFERENCE.....	54



# 1. ÚVOD

V této práci se zabývám klasickou molekulovou dynamikou (MD), která dokáže nasimulovat a analyzovat pohyb stovek až tisíců molekul, v časech několika nanosekund či dokonce mikrosekund. Tyto molekuly mohou měnit konformaci, interagovat mezi sebou nebo vytvářet a přerušovat nekovalentní vazby s okolními molekulami. MD nám tak může pomoci na atomární úrovni objasnit původ a průběh dějů, které by bylo obtížné studovat pomocí experimentu.

V případě této práce se bude jednat o teoretické simulace komplexů protein-DNA. Interakce proteinů s DNA jsou jedny z nejzákladnějších a nejdůležitějších biologických interakcí. Tyto interakce zprostředkovávají funkci enzymů, jako jsou nukleázy, udržují histony v chromatinu, kontrolují expresi genů, funkci polymerázy či opravy samotné DNA a mnoho dalších důležitých procesů.

Gavezzotti<sup>54(1)</sup> popsal jako první molekulovou mechaniku s použitím silových polí. První simulace Monte Carlo (MD) dle Metropolis a spol.<sup>(2)</sup> byla provedena již v roce 1950 a její cíl spočíval v určení střední hodnoty. Následující silová pole v sedmdesátých letech dvacátého století byla založena na harmonických potenciálech. Tato řada silových polí se označovala jako 1. generace, za níž následovaly další dvě generace silových polí, ze kterých se postupem času vyvinula současná silová pole, s jejichž pomocí můžeme objasňovat dynamiku zkoumaných molekul.

Mým úkolem bude s pomocí empirických potenciálů určit stabilitu vybraných konformací v několika simulovaných komplexech proteinů s DNA. Existuje mnoho různých silových polí, jako např. AMBER, CHARMM, OPLS, GROMOS a mnoho dalších. V současnosti jsou nejvyužívanějšími silovými poli verze z rodin AMBER a CHARMM. Ve své práci budu využívat různé verze silových polí z rodiny AMBER<sup>(1)</sup> (z anglického Assisted Model Building with Energy Refinement) a to silová pole OL15, bsc1 a nově vyvinuté pole OL21. Je ale třeba poznamenat, že vývoj a zpřesňování silových polí nadále pokračuje.

Konkrétně se budu zajímat o neobvyklé (nekanonické) stavy páteře DNA a působení proteinu na ně. Budu studovat, jak přítomnost proteinu ovlivňuje jejich stabilitu, jestli dochází k jejich deformaci, nebo naopak zda je protein stabilizuje natolik, že bude jejich geometrie zachována tak, jak je popsána v krystalografických strukturách. Stabilitu těchto nekanonických struktur budu sledovat v závislosti na čase v průběhu MD simulací právě ve zmiňovaných silových polích z rodiny AMBER.

## 2. STRUKTURA

### 2.1. Struktura proteinu

Základní strukturu proteinů určují aminokyseliny (označované též dvoupísmennou zkratkou AK) propojené peptidovou vazbou. Svým specifickým uspořádáním pořadí v řetězci a jejich originálními vlastnostmi, dají proteinům jejich potřebnou specifitu. <sup>(4)(5)</sup>

#### 2.1.1. Primární struktura

Primární struktura je nejjednodušší struktura bílkovin, která je kódována pořadím nukleotidů, tedy v DNA. Toto pořadí udává aminokyselinovou sekvenci v proteinovém řetězci, rovněž tak i charakter postranním řetězcům AK. Každý protein je jedinečný, a to právě díky primární struktuře. Aminokyseliny jsou v řetězci propojeny kovalentní peptidovou vazbou –CO-NH– (tedy kontaktem vznikajícím mezi aminovou skupinou –NH<sub>2</sub> náležící první aminokyselině poskytující –H, a karboxylovou skupinou –COOH patřící následující aminokyselině z níž získáme –OH skupinu, následně dojde k odštěpení vody a vzniku peptidu). Jejich pořadí se zapisuje vždy od N-konce k C-konci. Zbylé strukturální uspořádání (sekundární, terciární a kvartérní) se oproti primární struktuře vyznačuje i nekovalentními interakcemi mezi řetězci. <sup>(4)(5)</sup>

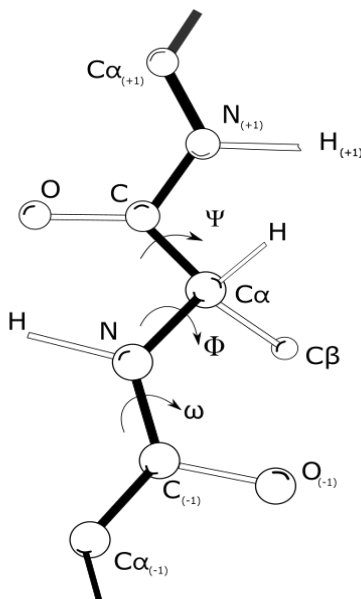
#### 2.1.2. Sekundární struktura

Sekundární struktura je postavena spíše na geometrickém, prostorovém uspořádání řetězce bílkovin. Existují zde dvě struktury polypeptidových řetězců, a to skládaný beta list ( $\beta$  – list), či sheet a alfa helix ( $\alpha$  – helix) neboli pravotočivá šroubovice. Jejich tvar stanovují torzní neboli dihedrální úhly. Zbytek řetězce se skládá z otoček, výdutí a spirál. Rozdíl mezi první strukturou a touto je, že vazby jsou zde tvořeny nekovalentně, přesněji řečeno vodíkovými můstky. <sup>(4)(5)(6)</sup>

#### 2.1.3. Dihedrálční úhel

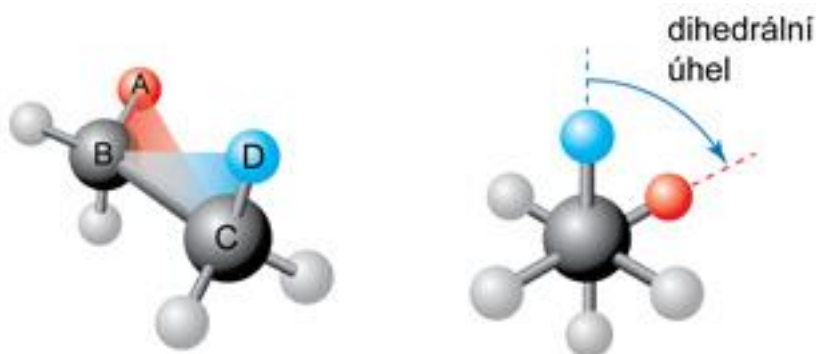
Sekundární struktura je určena dihedrálními úhly (Obrázek 1). Charakteristika trojrozměrného uspořádání jednoduché vazby nacházející se mezi dvěma atomy B a C definuje úhel svíraný mezi rovinami. První rovina je u atomů ABC a protilehlá druhá rovina

u atomů BCD (Obrázek 2). Celkové úhlové rozpětí může dosahovat od  $-180^\circ$  do  $+180^\circ$  úhlových stupňů.<sup>(7)</sup>



Obrázek 1: Zobrazení torzních úhlů  $\Phi$  a  $\Psi$  na peptidovém řetězci. Upravený obrázek v inkscape převzat z:<sup>(8)</sup>

Torzní úhel  $\Phi$  (phi) je úhel vazby N- $C_\alpha$  (centrální uhlík) k sousední vazbě a úhel  $\Psi$  (psi) je mezi vazbou C- $C_\alpha$  (tedy karbonylový uhlík a centrální uhlík.) a opět přilehlou vazbou.  $\omega$  (omega) udává planaritu vazby, pokud je úhel  $\omega$  roven  $180^\circ$ , pak se jedná o tzv. konfiguraci trans naproti tomu, pokud bude roven 0, tak se jedná o konfiguraci opačnou, tedy cis.



Obrázek 2: Dihedrání úhel, obrázek převzat z:<sup>(7)</sup>

Planarita úhlu  $\omega$  je způsobena zvyšováním polaritě (rozdíl elektronegativit) peptidové vazby, tj. od nepolární až k iontové vazbě, které je zapříčiněno rezonančními efekty, způsobujícími dvojný charakter této vazby. Dochází tak k rovinnému uspořádání centrálního a karbonylového atomu uhlíku, kyslíku a amidového dusíku. Avšak konfigurace cis není tak častá. Konformace  $\Psi$  vs.  $\Phi$  určuje typ sekundární struktury a na této závislosti torzních úhlů

je založen Ramachandranův diagram zobrazující všechny možné kombinace mezi těmito úhly v peptidovém řetězci. Ramachandranův diagram je využíván například k ověření strukturních modelů. <sup>(7)</sup>

#### **2.1.4. Terciární struktura**

Pro to abychom popsali správnou strukturu molekuly bílkoviny v jednom uceleném polypeptidovém řetězci, tedy v její 3D podobě, pak je zapotřebí vyjít ze sekundární struktury. Jedná se o struktury  $\alpha$ -helixu a  $\beta$ -listu o nichž je zmínka výše (viz 2.1.2.).

U terciární struktury klasifikujeme tři třídy. První z nich je fibrilární uspořádání, název je odvozen od slova fibrila, tedy jemné vlákno, což nám napovídá, že se jedná o vláknitou strukturu. Druhá třída se nazývá globulární (tvarem připomínající klubko) a třetí třídou jsou domény, jež jsou vždy uspořádány tak, aby byly prostorově nezávislé na zbytku molekuly. Domény mají svou vlastní klasifikaci, řadí se sem tři základní typy ( $\alpha$ -domény,  $\beta$ -domény a  $\alpha/\beta$  domény). Domény se dají brát jako jeden ze základních stavebních kamenů proteinů.

Pro tuto proteinovou strukturu jsou ve vazbách mezi sekundárními elementy zakomponovány i van der Waalsovy přitažlivé síly, disulfidické a vodíkové můstky, které mají funkci stabilizace celkové struktury. <sup>(5)(6)</sup>

#### **2.1.5. Kvartérní struktura**

Kvartérní struktura popisuje, jak se několik řetězců proteinu skládá do jednoho celku, ve formě tzv. podjednotek, které již nejsou propojeny peptidovými vazbami, ale nekovalentními vazbami. V prostoru spolu tyto podjednotky nadále interagují. Kvartérní struktura tedy určuje a popisuje umístění podjednotek vůči ostatním podjednotkám. Většinou jde o dimery či tetramery. K proteinům s kvartérní strukturou se řadí například enzymy a klasickým příkladem pro tento typ struktury je metaloprotein hemoglobin. <sup>(4)(5)</sup>

### **2.2. Struktura DNA**

Základní stavební jednotkou DNA (deoxyribonukleová kyselina) je nukleotid, jehož řetězením, tj. spojením více nukleotidů pomocí fosfodiesterové vazby (dvě esterové vazby a fosfát), vzniká polynukleotidové vlákno či řetězec. Tato vlákna jsou dále propojena vodíkovými můstky, na základě párování bazí. Obě polynukleotidová vlákna jsou vůči sobě

antiparalelní a komplementární. Tuto strukturu popsali již v roce 1953 pánové James D. Watson a Francis Crick, kteří za to o devět let později obdrželi Nobelovu cenu.<sup>(9)</sup>

Chemická struktura jednoho nukleotidu se skládá ze tří složek, a to z pěti-uhlíkatého cukru (aldopentóza), fosfátu a dusíkaté báze. Sacharid, jenž se zde objevuje, je 2-deoxyribóza. Na prvním uhlíku tohoto sacharidu je díky glykosidické vazbě napojena jedna ze čtyř dusíkatých bází, specifických pro strukturu DNA, jimiž jsou: adenin (A), cytosin (C), guanin (G) a thymin (T). Oproti tomu v porovnání se strukturou RNA, se místo thyminu (T), vyskytuje dusíkatá báze zvaná uracil (U), a místo deoxyribózy je zde pouze ribóza. Báze se řídí zákonem komplementarity, kdy se mezi sebou váže vždy jedna pyrimidinová a jedna purinová báze, tuto interakci zprostředkovávají vodíkové můstky. Adenin a thymin jsou propojeny dvěma vodíkovými můstky a cytosin a guanin třemi vodíkovými můstky. V neposlední řadě je na pátém uhlíku esterovou vazbou navázán fosfát.<sup>(9)(10)</sup>

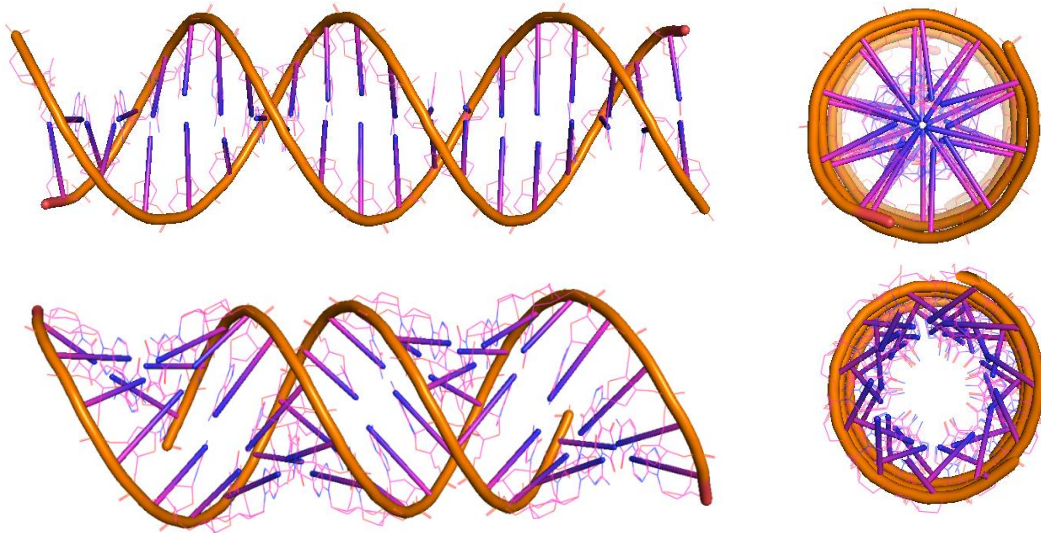
Na povrchu dvoušroubovice nacházíme dva žlábký, které jsou jedním z možných míst pro interakci s nukleovou kyselinou. První nese název „velký“ a jak již název napovídá, tak bude větší a tím i lépe přístupný pro interakce, tudíž se sem přesně vejde  $\alpha$ -šroubovice. Druhý se nazývá „malý“ žlábek, ten naopak lépe interaguje s  $\beta$ -listy.<sup>(9)(11)</sup>

### 2.2.1. Struktura A-, B- a Z-DNA

Forma B-DNA (Obrázek 3) je nejběžnější pravotočivá dvojitá šroubovice. Tuto formu a její typické páteřní stavy budeme nazývat kanonickými. Obsahuje dva žlábký: velký žlábek, který má šířku asi 12 Å (Angström) a malý žlábek, který je asi poloviční, avšak skoro stejně hluboký. Skládá se ze dvou antiparalelních vláken, tvořících spirálovité závity, díky nimž dostává DNA svůj specifický tvar. Jeden závit obsahuje v průměru okolo 10,5 párů bází, které jsou od sebe vzdáleny asi 3,4 Å a jsou téměř kolmé k ose šroubovice. V těchto parametrech popisujících uspořádání bází, se B-DNA liší od A- a Z-DNA forem.<sup>(12)</sup>

Formy A- a Z-DNA jsou též dvoušroubovice, nicméně A-DNA obsahuje 11 párů bází na jeden závit, které jsou nakloněné o 20° k ose, a žlábký u této formy DNA zde nejsou tak hluboké. Z-DNA je nápadně odlišná od zbylých dvou forem. Je to levotočivá nekanonická šroubovice s tzv. „cik-cak“ vzorem, kdy se střídají purinové a pyrimidinové repetice.<sup>(13)(14)</sup>

Hlavní rozdíl mezi B-DNA a formami A- a Z-DNA spočívá v konformačních stavech páteře. Klasická B-DNA obsahuje kanonické stavy, ale pokud dojde k jakékoli změně konformace páteře DNA, ke které může docházet i díky interakci mezi proteinem a DNA, pak se jedná o tzv. nekanonické stavy páteře. A- a Z-DNA mají pouze nekanonické stavy páteře. <sup>(13)(14)</sup>



*Obrázek 3: Strukturální rozdíl mezi A- a B-DNA. Nahoře se nachází B-DNA a dole A-DNA. Obrázek byl vytvořen v programu PyMOL.*

### 3. INTERAKCE PROTEINŮ S DNA

Pod pojmem interakce si představujeme vzájemné působení mezi dvěma a více činiteli, v tomto případě mezi DNA a proteinem.

#### 3.1. Specifická a nespecifická interakce proteinu s DNA

Specifické interakce proteinu s DNA klasifikujeme do dvou tříd. První z nich se nazývá přímé čtení kontaktů neboli sekvenčně specifická vazba. Kontaktem se zde myslí místo styku pro interakci. U přímého čtení se tento kontakt týká proteinu s okrajovým lemem páru bází. Když se protein spojí s bází, dojde k tzv. „base readout“ což znamená, že se protein snaží rozeznat o jakou bázi DNA se jedná. Zde tedy záleží, o jakou sekvenci bází jde, narozdíl od nespecifických interakcí, kde na sekvenci nezáleží. Často se stává, že tato vazba ovlivní strukturu DNA, a může tak dojít například ke změně kinetických vlastností. <sup>(15)</sup>

Druhou třídou specifických interakcí je nepřímé čtení, které se vztahuje k interakci mezi proteinem a páteří DNA. Zde přichází ke slovu mnoho aspektů ovlivňujících interakci jako jsou: deformace DNA, forma DNA, hloubka žlábků, ohnutí struktury páteře a mnoho dalších. Díky těmto aspektům se vytváří jedinečná specifická vazba k sekvenci DNA, tudíž se mohou tvořit kontakty jen s určitými energeticky výhodnějšími strukturami. <sup>(15)(16)</sup>

Nespecifické interakce proteinu s DNA jsou založeny na interakci záporně nabitě cukr – fosfátové páteře s proteinem. Většinou se jedná o kontakty vytvořené páteří DNA na kladně nabitě aminokyseliny.

Proteiny, které se vážou se stejnou vysokou afinitou ke všem sekvencím DNA, se nazývají nespecifické vazebné proteiny. Oproti tomu proteiny, vážající se na určitou sekvenci s nejvyšší afinitou, jsou specifické vazebné proteiny. Dále nás u interakce zajímá, jak je daná afinita rozložena mezi sekvence. Jako příklad lze uvést restriční enzymy mající vysokou specifitu vůči konkrétním sekvencím. Slouží k ochraně buňky před invazivními sekvencemi DNA jako jsou viry, kdy se snaží modifikovat tyto sekvence za pomoci metyláz. Dalšími specifickými vazebnými proteiny jsou transkripční faktory (TFs) regulující různé geny. TFs mají vysokou specifitu s nejvyšší afinitou pouze k jedné vybrané sekvenci. To je velkou výhodou, pokud daný gen chceme exprimovat. Tím že se TFs vážou na určitá místa silněji, díky vyšší afinitě, pak lze takovéto geny snadno regulovat na různých úrovních. <sup>(16)</sup> Další možnou variantou pro vazbu jsou žlábkové (již popsáno výše) anebo možnost spojení za pomoci vazebných domén. Pro  $\alpha$ -šroubovici to jsou například leucinový zip, helix-otočka-

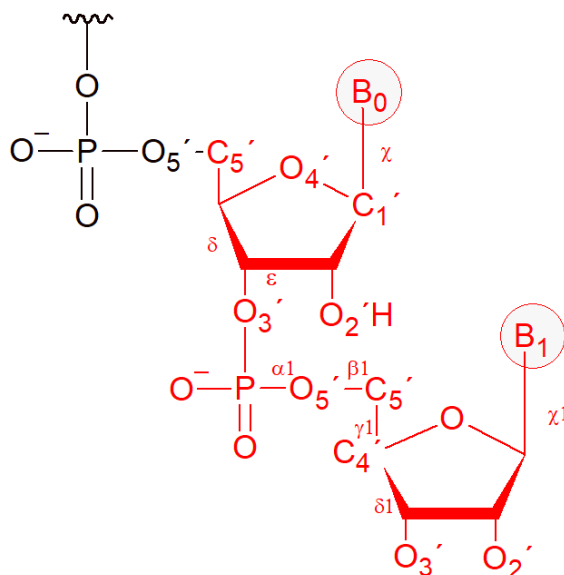
helix (HTH-helix-turn-helix) a zinkový zip. Pro  $\beta$ -listy to jsou histony a HMG-box (high mobility group box). <sup>(16)(17)</sup>



## 4. DNA KONFORMERY

Projekt DNATCO se snaží klasifikovat geometrii páteře DNA (Assignment of DNA and RNA conformations). Všeobecně známé rozdělení DNA na formy A, B a Z může být na základě klasifikace konformací nukleotidových kroků v DNA s pomocí algoritmu přiřazujícího určitým charakteristickým hodnotám torzních úhlů jednotlivé třídy, zvané NtC (nucleotide conformers), rozšířeno o další konformery, označované čtyřpísmennými symboly. Například v NtC BB00 první dvě písmena definují, o jakou formu DNA se v konformaci jedná (A, B nebo Z), a třetí a čtvrtá pozice kódu jsou jedinečná čísla zajišťující originalitu NtC. Nachází se zde ještě speciální třída NANT, která je určena pro konformace, nespádající do žádné jiné třídy NtC. Takovýchto tříd konformerů NtC je  $57 + 1$  (NANT), a jsou v původní literatuře uvedeny v přehledné tabulce, spolu se svými charakteristikami. Konformace NtC se mohou dále sdružovat do písmen strukturní abecedy DNA, NtA, skládající se z 12 členů vycházejících z podobnosti struktur NtC. <sup>(18)</sup>

Krok konformeru NtC je jednoznačně popsán všemi šesti torzními úhly páteře DNA, spolu s dalším torzním úhlem  $\delta$  (delta) následujícího nukleotidu a dále dvěma hodnotami torzních úhlů popisujícími rotaci glykosidové vazby obou bazí, tj. oblast od prvního torzního úhlu  $\delta$  až po druhý torzní úhel  $\delta_1$ . Celý tento krok je zakreslen červeně (viz Obrázek 4). <sup>(18)</sup>



Obrázek 4: Krok konformeru ntC. Obrázek byl převzat a zpracován v programu ChemSketch z <https://doi.org/10.1093/nar/gkw381>. <sup>(18)</sup>

#### 4.1. Konformery BB00, BB15 a BB12

Nejpřirozenějším a také nejčastějším konformerem je NtC BB00, též nazývaný BI. Je to kanonická B forma DNA, kdy její hodnoty páteřních úhlů jsou  $\alpha \sim 304^\circ$ ,  $\beta \sim 180^\circ$  a  $\gamma \sim 44^\circ$ . Oproti tomu BB15 je velmi zajímavý a neobvyklý stav, kdy  $\beta$  je na stejných hodnotách jako u kanonické DNA, ale  $\alpha$  a  $\gamma$  se výrazně liší, neboť nabývají hodnot,  $\alpha \sim 340^\circ$  a  $\gamma$  se blíží  $0^\circ$ . Protože tyto hodnoty jsou v blízkosti maxim energie potenciálů  $\alpha$  a  $\gamma$ , mohly by mít v MD simulacích tendenci být nestabilní a snahu vracet se zpět do kanonických hodnot, jak budu diskutovat níže. A právě tento nevšední konformer mě v této práci bude zajímat.<sup>(19)</sup> Druhým neobvyklým konformerem je BB12, opět jeden z frekventovanějších nekanonických konformerů, kdy jeho úhly jsou rovny hodnotám  $\alpha \sim 257^\circ$ ,  $\beta \sim 76^\circ$  a  $\gamma \sim 171^\circ$ , které jsou také velmi odlišné od kanonické BB00, ale už nemají tak kriticky vysoké hodnoty, a tudíž by mohly být v MD simulacích stabilní.<sup>(19)</sup>

## 5. KOMPLEXY PROTEIN-DNA

Interakce proteinů s DNA je nejzákladnější a jednou z nejdůležitějších biologických interakcí. Tato interakce je důležitá pro funkci enzymů, jako jsou nukleázy a metylázy, udržuje histony v chromatinu, zhušťuje a uspořádává chromozomy, kontroluje expresi genů, funkci polymeráz a hraje roli při opravě poškození samotné DNA, kdy se proteiny dají do pohybu, aby se DNA co nejrychleji opravila a v mnoha dalších důležitých procesech. <sup>(16)(23)</sup>

Pro každou interakci je specifický jiný protein. Například pro regulaci transkripce u prokaryot je zapotřebí interakce s regulačními proteiny, které se váží v oblasti promotoru DNA. Pak může dojít ke dvěma účinkům proteinu, a to k pozitivnímu, kdy je vyvolána transkripce, tak k negativnímu účinku, kdy dochází k inaktivaci transkripce, díky vazbě represoru na operátor. <sup>(16)(21)</sup>

Jak již bylo zmíněno výše, interakce v těchto komplexech proteinů s DNA je buď sekvenčně specifická, nebo strukturně specifická, a proteiny mohou s DNA interagovat za pomoci malého a velkého žlábků. <sup>(22)</sup>

V této práci jsem se zaměřila na následující tři komplexy proteinů s DNA.

### 5.1. 3JTG, Ets doména vázající TGF- $\beta$ receptor

Krystalová struktura myši Elf3 C-terminální DNA-vázající domény v komplexu s DNA promotorem TGF- $\beta$  receptoru typu II DNA s rozlišením krystalové struktury 2.20 Å, je zobrazena na obrázku 7. Elf3 (E74-like factor 3) je jedním z třicetičlenné rodiny transkripčních faktorů zvané Ets (E-twenty-six specific). Celá rodina Ets, stejně tak i Elf3, když je správně aktivován, tak je pak je schopen se specificky vázat na promotorovou sekvenci DNA. Tato specifická vazba Ets domény na jádro GGAA/T motivu DNA se často nazývá zkratkou EBS (Ets-binding site). Každý faktor z rodiny Ets vyžaduje po aktivaci parametry na rozpoznání správného místa EBS.

Elf3 je zastoupen v mnoha buňkách epitelu či karcinomů, a také se podílí na nádorovém bujení a metastázách včetně nádorů prsu. Právě Elf3 by mělo silně stimulovat, na dvou vazebných místech, promotor genu TGF- $\beta$  receptoru typu II (T $\beta$ R-II), který se chová jako nádorový supresorový gen. Tato vazebná místa sekvence mT $\beta$ R-II DNA (viz Obrázek 5) označujeme jako: místo B, kde se Elf3 dotýká jádra GGAA/T motivu DNA ve velkém žlábků, což je pro rodinu Ets normální a Elf3 má k tomuto místu i vyšší afinitu, a dále vazebné místo A, kde se Elf3 dotýká sekvence z menšího žlábků DNA, což už není pro

rodinu Ets tak obvyklé. Interakce Elf3 s místem B vede k lokálním konformačním změnám, což vede k posílení vazebné afinity k místu A. Tudíž dochází k postupné interakci, kde dojde jako první k interakci s místem B a následně díky konformaci, k interakci s místem A. Elf3 díky těmto vazebným místům může například v buňkách karcinomu prsu zvyšovat expresi TβR-II a snižovat tak nádorovou aktivitu těchto buněk. Přerušení vazby mezi místy B a A s Elf3 vede k drastickému snížení aktivity promotoru genu TβR-II.



Obrázek 5: Sekvence mTβR-II DNA obsahující obě vazebná místa A a B, žlutě zvýrazněné G je nahrazeno C pro zvýšení krystalizace. Upravený obrázek v inkscape převzat z: <sup>(22)</sup>

Elf3 se skládá z pěti specifických domén (Obrázek 7): ostrá a háková doména (PNT a AT/NLS), transaktivační doména (TAD), doména obsahující serin a kyselinu asparagovou



Obrázek 7: Specifické domény Elf3, upraveno v programu inkscape a převzato z: <sup>(22)</sup>

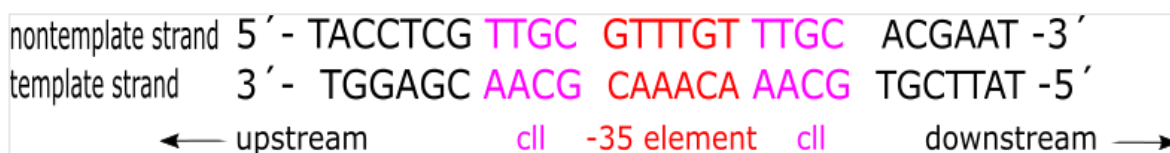
(SAR) a Ets doména (ETS). C – terminální doména Ets, kterou se ve své práci zabývám, je velmi důležitá pro správnou funkci Elf3, a také je mimo jiné nezbytná pro vazbu Elf3 na TβR-II gen jak in vivo, tak in vitro. <sup>(22)</sup> Konformery BB15 jsou tyrkysově vyznačeny na obrázku 7, kde je vyobrazena celá struktura komplexu Ets domény vázající TGF-β receptor.



Obrázek 6: Struktura 3JTG vytvořena v programu PyMOL. Růžová barva znázorňuje protein a tyrkysová konformery BB15.

## 5.2. 1ZS4, $\lambda$ cII-aktivátor transkripce

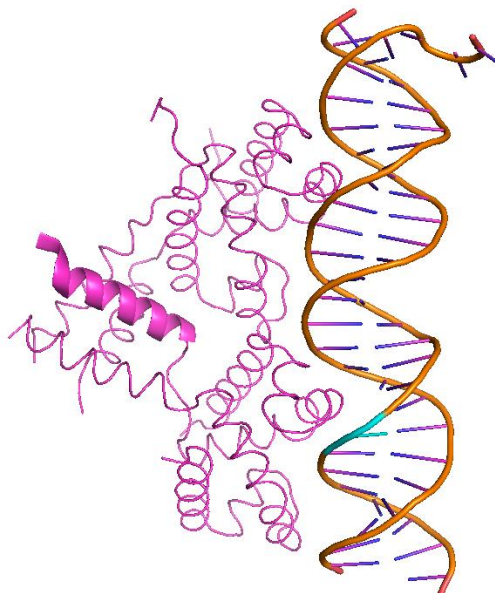
Tetramerní protein cII bakteriofága  $\lambda$  (Obrázek 9) je aktivátorem transkripce z promotorů  $P_{RE}$ ,  $P_I$  a  $P_{AQ}$  bakteriofága. Řídí lytický a lyzogenní vývoj bakteriofága. Aktivuje se vazbou na dvě repetice, které se nacházejí na lemu promotoru elementu -35 (viz Obrázek 8). CII náleží do rodiny proteinů, která dokáže díky motivu HTH (helix-turn-helix) jednoho monomeru, vytvořit z každého dimeru sekvenčně specifickou vazbu na repetice DNA, a prostřednictvím druhého monomeru přispívá k vazbě na DNA interakcí fosfát/páteř.



Obrázek 8: Pozice DNA vůči počátku transkripce (na +1 pro  $P_{RE}$ ), červeně je označen element -35 a purpurově  $\lambda$  cII specifické přímé repetice. Obrázek byl převzat a upraven v inkscape. <sup>(23)</sup>

$\lambda$  cII má dvě funkce, za prvé funguje jako homodimer, rozpoznávající operátory DNA s invertovanou repeticí na stejné straně jako RNA polymeráza (RNAP). To cII usnadňuje vznik simultánní interakce protein/protein s C-terminální doménou podjednotky  $\alpha/\sigma$  RNAP/CTD. RNAP a  $\alpha$ CTD vytvářejí vícestranné interakce s doménami a  $\alpha$ CTD formuje interakci s DNA v malém žlábků, těsně před elementem -35. Ztráta nebo jakákoli mutace  $\alpha$ CTD způsobuje poškození v aktivaci promotoru  $\lambda P_{RE}$  v závislosti na cII.

Takto  $\alpha$ CTD slouží jako proteinový můstek, kde na jedné straně dochází k interakci s cII a na druhé s RNAP a přes interakci s  $\sigma_4$  na další. Většinou aktivátory třídy II přímo utváří kontakt protein/protein s  $\sigma_4$  doménou vázanou na -35 element. Pokud se cII naváže, pak dokáže ohýbat operátor DNA takovým způsobem, který posílí interakci  $\sigma_4$  s elementem -35 to vede k snížení aktivačního účinku cII. Druhou vykonávanou funkcí  $\lambda$  cII je úloha homotetrameru 97 aminokyselinových monomerů. Diagnostikuje tak přímé repetice, protínající hexamer promotoru, tetranukleotidových sekvencí na protější straně duplexu DNA oproti RNAP. <sup>(23)</sup> Celkové rozlišení z krystalografické analýzy struktury je 1,70 Å.



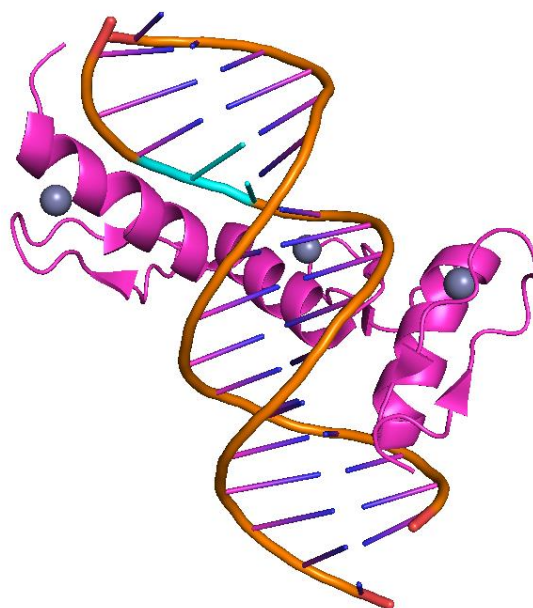
Obrázek 9: Struktura 1ZS4 (vytvořeno pomocí PyMOL). Růžová barva znázorňuje protein a tyrkysová konformery BB15.

### 5.3. 1G2F, TATA box komplex

Struktura dvou Cys<sub>2</sub>His<sub>2</sub> Zinc finger/TATA box komplexů je ukázána na obrázku 10, kde jsou rezidua konformeru BB15 vyznačena tyrkysově a nacházejí se v blízkosti alfa helixu proteinu. Celý komplex se skládá z jednoho proteinového řetězce a dvou řetězců nukleových kyselin a jeho rozlišení je 2,00 Å. <sup>(25)</sup>

Cys<sub>2</sub>His<sub>2</sub> je složen ze dvou cysteinových a dvou histidinových zbytků, které napomáhají k vázání zinku a stabilizaci struktury. Je to jeden z rozšířenějších typů zinkového prstu (Zif268 z rodiny EGR), jenž funguje jako vazebný proteinový motiv rozpoznávající sekvence DNA, v tomto případě k rozpoznání sekvence TATA boxu promotoru adenoviru vazebného místa p53 (Hognessova boxu). Cys<sub>2</sub>His<sub>2</sub> je většinou transregulační regulátor genové exprese, což znamená, že může mít velký vliv na diferenciaci a potlačení maligních transformací buněk. <sup>(24)(25)(26)</sup>

Zinkový prst se skládá z dvou či tří β-listů a jednoho α-helixu. V N-koncové části α-helixu se nacházejí čtyři zbytky aminokyselin, které zapříčiňují vazbu s DNA v jejím velkém žlábků, díky interakci s donory a akceptory vodíkové vazby. Zinkový prst neslouží pouze a jen k rozpoznávání sekvencí DNA, ale také k zprostředkování interakce protein/protein nebo RNA/protein. <sup>(24)(25)</sup>

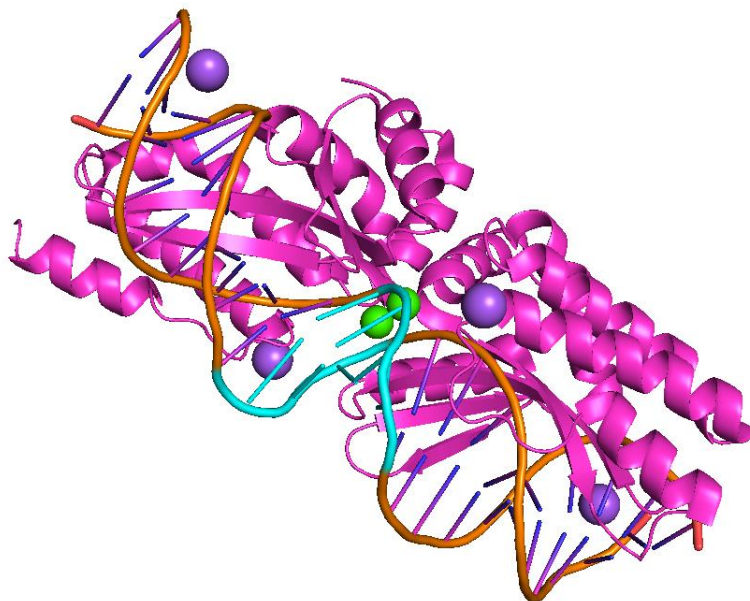


Obrázek 10: Struktura 1G2F (vytvořena v programu PyMOL). Růžová představuje protein, šedá zinek a tyrkysová konformer BB15.

#### 5.4. 1T9I, LAGLIDADG homing endonukleáza

Struktura 1T9I je komplexem I-CreI LAGLIDADG homing (naváděcí) endonukleázy s DNA a jejíž rozlišení je 1.60 Å (Obrázek 11). Proces „Homing“ v názvu znamená v tomto případě, že dokáže duplikovat určité introny či inteiny do některých typů alel, jež tyto integrované oblasti ve své struktuře samy neobsahují. To vše je iniciováno endonukleázou I – CreI kódovanou sekvencí ORF, která rozpozná určité místo pro duplikaci a vložení. Enzym I – CreI v podobě volných homodimerů se nachází jako samostatné geny uvnitř intervenční sekvence intronů a inteinů. Hlavním úkolem endonukleázy je oprava dvouvláknové DNA, kdy přes intron jako její hostitelskou buňku, spustí buněčné mechanismy opravy DNA. V případě tohoto komplexu se jako „naváděcí“ enzym bere LAGLIDADG endonukleáza (I – CreI), která je považována za největší rodinu „naváděcích“ enzymů. Endonukleázy se obecně vyskytují pouze jen na velmi málo místech, nebo dokonce jen na jednom. V tomto komplexu se nachází, jak již bylo zmíněno homodimerní I – CreI a monomerní I – SceI, které se váží na DNA, a mezi nimiž jsou na aktivních místech, s kontaktem na kyselinu asparagovou, situovány ionty dvojmocného kovu s funkcí štěpení vláken. Jak je vidět z obrázku 11, mnou sledovaný stav BB12 se nachází v blízkosti

aktivního centra a lze očekávat, že jeho přítomnost bude důležitá pro funkci endonukleázy.<sup>(27)</sup>



*Obrázek 11: Struktura 1T9I (vytvořena v programu PyMOL).  
Růžová představuje protein, zelená vápník, fialová sodík a  
tyrkysová konformer BB12.*



## 6. CÍLE PRÁCE

Konformace nukleových kyselin mohou být klasifikovány do tzv. NtC tříd. Zaměřila jsem se na dvě z těchto NtC tříd, označené zkratkami BB15 a BB12, které patří mezi nejčastěji zastoupené nekanonické konformery. Cílem této práce bylo zjistit, jak stabilní jsou v molekulově dynamických simulacích a jestli je jejich geometrie správně popsána dnes používanými silovými poli. Tyto konformery se od kanonické B-DNA odlišují hlavně hodnotami páteřních úhlů  $\alpha$  a  $\gamma$ , proto jsem se na ně zaměřila.

Pro konformer BB15 jsou dle klasifikace NtC (uvedené na webové stránce DNATCO) hodnoty úhlů:  $\alpha \sim 345^\circ$  a  $\gamma \sim 350^\circ$ , což jsou relativně vysoké hodnoty, u nichž jsem se domnívala, že by nemusely být v průběhu molekulové dynamiky stabilní. Oproti tomu konformer BB12, kde jsou hodnoty úhlů:  $\alpha \sim 257^\circ$  a  $\gamma \sim 171^\circ$ , jsem si vybrala z důvodu porovnání, protože u těchto hodnot předpokládám, že by mohly být stabilnější a měly by tedy v simulaci setrvat. V práci byly testovány dvě široce používané varianty silových polí, bsc1 a OL15 z roku 2015 a jedno nově vyvinuté pole, OL21.

## 7. METODY

Vybrala jsem si tři zástupce dinukleotidu NtC konformery označené BB15 a jednoho zástupce NtC konformeru BB12. Výběr zástupců byl založen na dostatečném rozlišení krystalové struktury a nízkém RMSD (root-mean-square derivation), tedy střední kvadratické odchylky od reprezentativní struktury daného NtC (podobnost s geometrií NtC). Z důvodu rychlosti simulace byly voleny komplexy spíše menší velikosti. Dále byl výběr založen na konformačním validačním skóre (confal) hodnotící jistoty zařazení do daného NtC.<sup>(28)</sup> Všechny zvolené zástupce jsem uvedla do jedné přehledné tabulky (Tab.1).

Tabulka 1: Jednotliví zástupci NtC konformerů BB15 a BB12.

PDB ID	NtC	Rozlišení [Å]	Krok konformeru (řetězec_číslo reziduí)	RMSD <sup>a)</sup>	Confal
3JTG	BB15	2,20	B_dT9_dC10	0,19	85
			B_dC10_dC11	0,23	77
			C_dA104_dC105	0,21	64
			C_dA111_dC112	0,17	72
			C_dC114_dC115	0,25	91
1ZS4	BB15	1,70	T_dG8_dC9	0,00	95
1G2F	BB15	2,00	A_dG5_dC6	0,23	79
1T9I	BB12	1,60	C_dG511_dT512	0,39	63
			C_dG513_dA514	0,25	72
			D_dT561_dC562	0,22	92
			D_dA563_dC564	0,33	57

a) Střední kvadratická odchylka od reprezentativní struktury daného NtC.

### 7.1. Silová pole a parametrizace

V této práci se zabývám současnými silovými poli z rodiny AMBER. Konkrétně jsem se rozhodla otestovat tři silová pole vhodná pro popis DNA: OL15, OL21 a bsc1. Tato tři různá silová pole byla zvolena proto, že odlišná pole mohou někdy poskytovat různé předpovědi a jejich porovnání zvýší vypovídací hodnotu teoretických simulací protein-DNA komplexů, Nároky na přesnost silových polí jsou u simulací nukleových kyselin vysoké, jinak by nemohly správně předpovídat všechny konformační změny a chování DNA, či jiných

organických látek vyskytujících se ve vybraných komplexech. V případech, kdy se ukáže, že silová pole nejsou dostatečně přesná, je potřeba provést další úpravy a publikovat opravu pole. Každé ze tří vybraných polí obsahuje jiné modifikace vzhledem k původnímu silovému poli Amber ff99, které byly zavedeny pro zlepšení popisu různých DNA forem. Například u pole bsc1 byly modifikovány parametry pro konformaci deoxyribózy, které byly oproti tomu u pole OL15 byly ponechány beze změny. Pro moji práci bude důležité hlavně to, že u pole OL21 byly modifikovány parametry  $\alpha$  a  $\gamma$ , které jsou u verzí OL15 a bsc1 stejné jako u výchozího pole ff99.<sup>(29)</sup>

Aby silová pole z rodiny AMBER dokázala správně modelovat energii systému, musíme mít k dispozici parametry pro modelovaný systém. Tyto parametry se vyskytují v obecné rovnici potenciální energie. Rovnice se skládá z vazebných (kovalentní vazby) a nevazebných (elektrostatika a van der Waalovy interakce) členů (Rovnice 1)

*Rovnice 1: Základní obecná rovnice celkové potenciální energie.*

$$\begin{aligned}
 E_{\text{celková}} &= E_{\text{vazeb.}} + E_{\text{nevazeb.}} \\
 E_{\text{celková}}(R) &= \sum_{\text{kovalentní vazby}} K_r (r - r_{eq})^2 \\
 &\quad + \sum_{\text{úhly}} K_\theta (\theta - \theta_{eq})^2 \\
 &\quad + \sum_{\text{torze}} \frac{V_n}{2} (1 + \cos[n\varphi - \gamma]) \\
 &\quad + \sum_{i < j}^{\text{atomy}} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \quad \text{van der Waals} \\
 &\quad + \sum_{i < j}^{\text{atomy}} \frac{q_i q_j}{\epsilon R_{ij}} \quad \text{elektrostatika}
 \end{aligned}$$

Jak již bylo zmíněno, rovnice se skládá ze dvou částí. Do první části vazebných členů patří suma kovalentních vazeb, kde  $K_r$  je empirická konstanta možné roztažnosti vazby,  $r$  je vazebná vzdálenost a  $r_{eq}$  je přirozená (původní) délka vazby. Dále vzorec sumy úhlů obsahuje empirickou konstantu ohybové síly úhlu  $K_\theta$ ,  $\theta$  je vazebný úhel a  $\theta_{eq}$  značí rovnovážnou hodnotu úhlu.  $K_\theta$  a  $K_r$  konstanty reprodukovují vibrační frekvence. Posledním vazebným členem je suma torzních úhlů, kde  $V_n$  značí bariéru rotace přirozených vazeb,  $n$  je periodičita rotace,  $\varphi$  značí torzní úhel a  $\gamma$  úhel v němž potenciál dosahuje svého minima.

Zbylé dvě sumy se již týkají nevazebných členů rovnice, a to rovnice vyjadřující van der Waalovu energii, kde  $A_{ij}$  a  $B_{ij}$  jsou parametry obsahující informaci o hloubce a poloze

minima na křivce potenciální energie dané dvojice interagujících atomů a  $R_{ij}$  je vzdálenost mezi atomy (Lennard-Jonesův potenciál). Druhá suma nevazebného členu udává součet elektrostatických energií mezi všemi atomy nesoucími statické náboje, jež mezi sebou interagují. Malé  $q_i$  a  $q_j$  jsou parciální náboje atomů a  $\epsilon$  je dielektrická konstanta.

MD simulace vycházejí z druhého Newtonova zákona, kdy na základě znalosti síly, která působí na atomy, můžeme pomocí integrace Newtonových rovnic určit polohy atomů pohybujících se v čase (polohy, kde se atomy budou nacházet po uplynutí zvoleného časového intervalu). Sboru časových snímků (koordinát atomů) se říká trajektorie. Síly působící na atomy získáme derivací rovnice 1.

## 7.2. Příprava výchozích struktur

Výchozí struktury byly nejdříve neutralizovány a poté byly přidány další ionty tak, aby výsledná koncentrace byla 0,15 M KCl s parametry iontů dle Jounga a Cheathama <sup>(30)</sup>.

Neutralizace boxů je potřeba k tomu, aby nedocházelo mezi periodicky se opakujícími oktaedrickými boxy k repulzi a dodání dalších iontů bylo provedeno proto, aby se podmínky simulace přiblížily přirozenému prostředí v buňce. Ionty byly přidány v programu LEaP z programového balíku AMBER. Například pro neutralizaci páteře komplexu 3JTG jsem musela do systému dodat 23  $K^+$  iontů a následně 22  $Cl^-$  iontů 22  $K^+$  iontů pro dosažení zvolené koncentrace 0,15 M KCl. Stejným postupem jsem neutralizovala i zbylé komplexy.

Všechny komplexy byly solvatovány za pomoci oktaedrického boxu s vodním modelem SPC/E <sup>(31)</sup>, jehož velikost byla určena tak, aby vzdálenost mezi vnějšími atomy molekul a okrajem boxu byla alespoň 10 Å. Pouze u struktury 1ZS4 se kvůli její velikosti a objemnosti proteinu musel tento parametr upravit na 12 Å, aby nedocházelo k nežádoucím kontaktům mezi komplexy v sousedních boxech.

U komplexu 1G2F jsem z původního PDB zjistila, že základní buňka obsahuje dvě identické struktury komplexu proteinu s DNA. Rozhodla jsem se pro simulaci využít jenom jednu z nich, a to tu první, pro snazší orientaci v číslování reziduí. Dále se v krystalické struktuře tohoto komplexu vyskytovaly tři stejné strukturní motivy zinkových prstů umožňující vazbu proteinu na DNA, pro které bylo nutné dodat parametry, protože parametry zinkových prstů nejsou automaticky přiřazovány programem LEaP. Toho jsem docílila využitím silového pole Zinc AMBER neboli ZAFF (zinc AMBER force field), které bylo vyvinuto panem Martinem B. Petersem a kol. <sup>(32)</sup> a patří mezi relativně dobře ověřené

parametrizace umožňující spolehlivý popis několika chemicky odlišných variant koordinovaného zinku. Pro toto silové pole bylo nutno dodat soubory ZAFF.prep a ZAFF.frcmod, které jsem získala z oficiální stránky programu AMBER (<https://ambermd.org/tutorials/advanced/tutorial20/ZAFF.htm>, soubory jsem stahovala v lednu 2022).

ZAFF.prep obsahuje informace o atomových typech atomů, konektivitě molekuly, či parciálních nábojích atomů. Díky tomuto souboru máme připravenou část potřebných informací, ještě je ale zapotřebí parametrů pro vazby atomů, jejich vzájemné vzdálenosti, úhly, Van der Waalsových parametrů a některé další potřebné informace pro správný chod simulace. Tyto parametry obsahuje druhý soubor ZAFF.frcmod.

Centrum zinku Zn-CCHH komplexu 1G2F bylo uspořádáno tak, že v okolí každého iontu zinku (Zn) se nacházely dvě molekuly cysteinu (C) označující se CYM (zkratka pro cystein vázající atom kovu) a další dvě molekuly histidinu (H) označující se HID. Aby došlo ke správnému přiřazení parametrů programem LEaP, bylo nutno ve výchozí PDB souboru přepsat jména reziduí iontů zinku na ZN4 a dále musely být přepsány rezidua cysteinu na CY4 a v neposlední řadě rezidua histidinu na HD2. Postup je popsán na veřejně dostupné webové stránce programu AMBER (<https://ambermd.org/tutorials/advanced/tutorial20/ZAFF.htm>), kde je uvedena tabulka, jak správně parametrizovat čtyřnásobně koordinované zinečnaté ionty. Tímto zápisem jsem parametrizovala ionty zinku pro následnou simulaci.

U komplexu 1T9I se nacházely další ionty, pro které je potřeba vybrat parametry, a to ionty  $\text{Ca}^{2+}$  a  $\text{Na}^+$ . Pro ionty  $\text{Na}^+$  jsem, stejně jako u neutralizujícího KCl, vybrala parametry Jounga a Cheathama. Dvojmocné  $\text{Ca}^{2+}$  ionty parametrizují dle Kenneth M. Merz Jr.<sup>(33)</sup>

Pro spuštění jednotlivých simulací byl využit program PMEMD z balíčku AMBER. Parametry simulace jsou následující. Tlak byl nastaven na 1 atmosféru (101 kPa) a byl regulován za pomoci Monte Carlo barostatu. Teplota byla nastavena na 298 K (25 °C) a byla kontrolována Langevinovým termostatem. Časový krok simulace byl 4 fs.<sup>(29)</sup> Aby bylo možno použít takto dlouhý integrační krok, muselo být použito schématu přerozdělení hmotností atomů na atomu vodíku (Hydrogen mass repartitioning<sup>(34)</sup>) Pro zpracování a analýzu výsledných trajektorií, pro něž byly souřadnice ukládány každých 10 ps, byl využit program cpptraj opět spadající do programového balíčku od AMBER.

### 7.2.1. Protonizační stavy histidinu

U výchozích struktur nejprve musely být u histidinových zbytků určeny jejich protonizační stavy. Ve strukturách získaných rentgenovou diffrakcí (XRD z anglického X-ray diffraction) nejsou uvedeny pozice vodíků, takže např. v případě histidinu není ze struktury jasné, jaký je jeho protonizační stav a musí se tedy zjistit. U původního PDB se histidin zapisuje třípísmenným označením HIS, ale program AMBER používá tři možné typy zápisu histidinu, které osvětlují specifickou strukturu izomeru histidinu, a tedy i jeho protonizační stav. Prvním typem je zkratka HID, kde se vodík na histidinu nachází blíže postrannímu řetězci, tedy na  $\delta$  dusíku. Druhá možnost je HIE, kde se nachází vzdáleněji od postranního řetězce, tedy na  $\epsilon$  dusíku. A třetí poslední možnou variantou je HIP, kdy vodíky histidinu jsou na obou dusících a zároveň je i kladně nabitý. <sup>(35)</sup>

Ke správnému určení, o jaký typ histidinu se jedná byl použit server PropKa <sup>(36)</sup>, který navrhuje protonizační stav na základě zadaného pH a struktury v okolí histidinu. V některých případech může ale PropKa navrhnout protonizaci, která je v rozporu se sítí vodíkových vazeb přítomných v krystalové geometrii komplexu. Proto byla struktura následně vizuálně zkontrolována v programu PyMOL se změřením na délky vodíkových vazeb histidinu s okolními molekulami a případně byl protonizační stav upraven, tak aby respektoval existující síť vodíkových vazeb. Například u komplexu s PDB ID 3JTG bylo původní krystalizační pH roztoku 6,5 a tak byly tři histidiny zapsány jako dvě protonované formy HIP a jeden HIS nahrazen jako HIE.

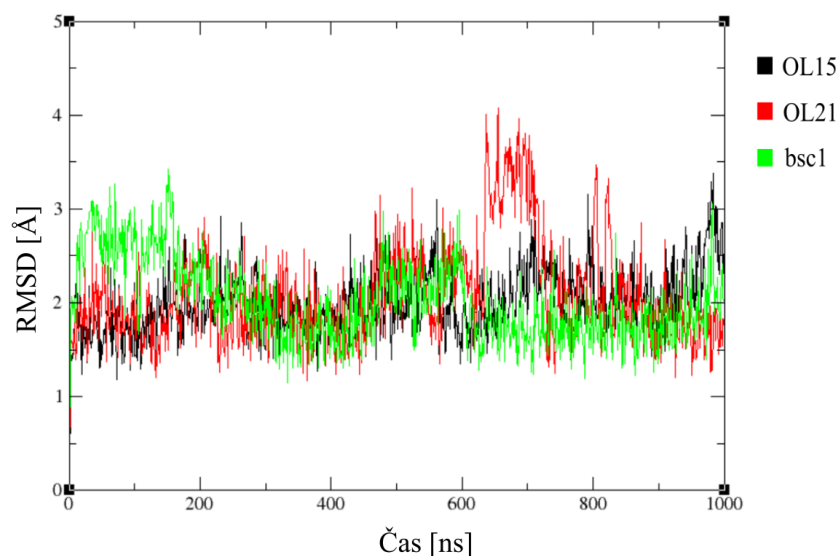
## 8. VÝSLEDKY A DISKUZE

### 8.1. Stabilita MD simulací

Pro první orientační kontrolu stability komplexů v MD simulacích jsem použila RMSD (root-mean-square derivation, střední kvadratická odchylka). RMSD udává pro každý snímek simulace odchylku okamžité geometrie od výchozí geometrie z krystalografické analýzy. Pokud RMSD nabývá vysokých hodnot, ve kterých setrvává, pak je odchylka od původní krystalické struktury velmi vysoká a značí buď špatně provedenou simulaci, nebo to, že komplex je v simulaci velice nestabilní a jeho geometrie se značně deformuje. Naopak pokud se hodnota RMSD drží na nízkých hodnotách (např. pod 2 až 3 Å, v závislosti na velikosti molekuly), znamená to, že simulovaná struktura se příliš neodlišuje od výchozí a nedochází k výrazné nechtěné deformaci. Menší krátkodobé změny (zvýšení) RMSD nejsou většinou důležité, zvláště pokud se struktura samovolně vrací k nižším hodnotám RMSD. Taková simulace je stabilní a vhodná k další analýze. Hodnotu RMSD jsem měřila zvláště pro protein a zvláště pro nukleovou kyselinu (DNA).

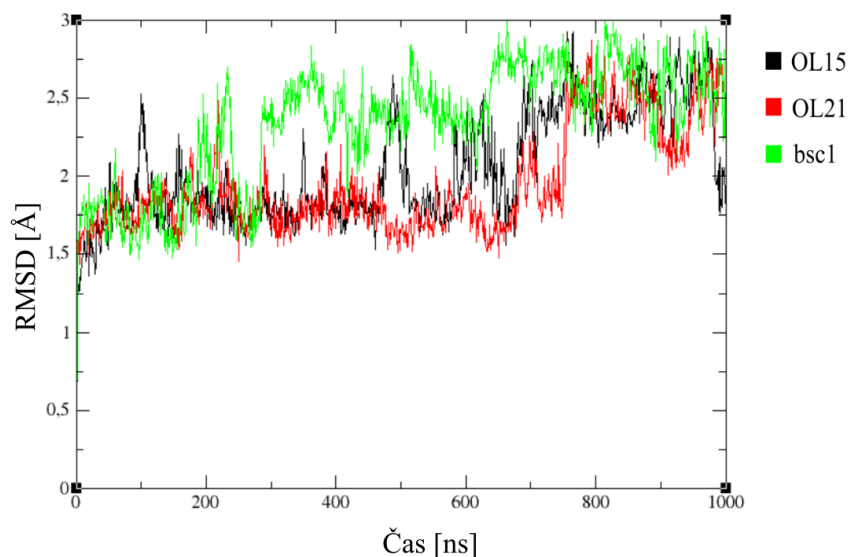
#### 8.1.1. Ets doména vázající TGF- $\beta$ receptor, 3JTG

První graf RMSD (Obrázek 12) je pro komplex vázající TGF- $\beta$  receptor, kde můžeme vidět hodnoty RMSD nukleové kyseliny v simulacích se všemi testovanými silovými poli. Pro všechna pole je RMSD spíše nižší a odpovídá stabilní simulaci. Pouze u silového pole OL21 se RMSD v čase 600 ns lehce zvýší a trochu se tedy liší od zbylých dvou silových polí. Připisuji to fluktuaci volných konců DNA, jež jsem viděla při kontrole simulované dynamiky v programu VMD (z anglického Visual Molecular Dynamics). Tyto volné konce mají větší tendenci se pohybovat a tím zvyšovat hodnoty RMSD. Protože je zvýšení RMSD dočasné a brzy se vrátí k původní nižší hodnotě, nenaznačuje to problémy v MD simulaci.



Obrázek 12: RMSD nukleové kyseliny v komplexu *Elf3* s *TGF-β* receptorem (3JTG) pro všechna silová pole.

RMSD pro protein (Obrázek 13) je zpočátku nižší než pro DNA, ale postupně se v průběhu simulace zvýší, nejdříve v poli *bsc1* a pak také v polích *OL15* a *OL21*. Opět to způsobují volné konce proteinu, které fluktuují v prostoru boxu, avšak tyto konce nejsou v blízkosti DNA a nijak neovlivňují interakci proteinu s DNA a nedochází tak k deformacím páteře DNA. Hodnoty RMSD se drží do 3 Å a struktura se tedy významně neodchyluje od původní krystalové struktury. Celkově jsou MD simulace komplexu vázajícího *TGF-β* receptor podle RMSD stabilní a výrazně se v jednotlivých silových polích neliší.

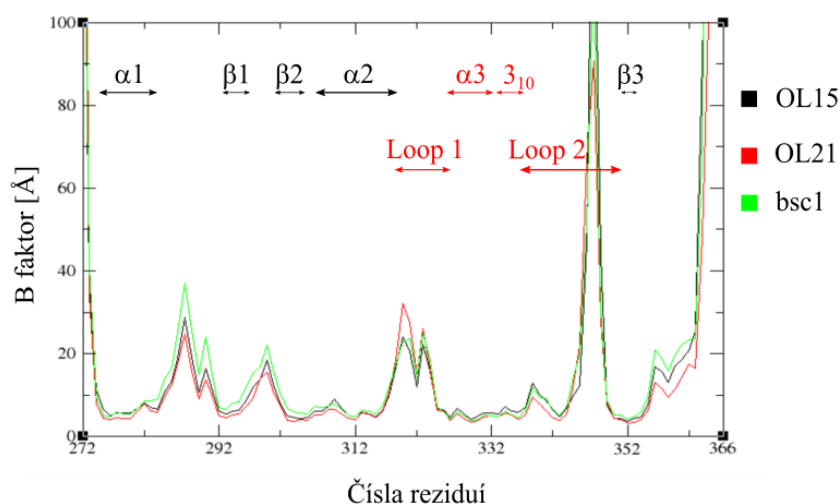


Obrázek 13: RMSD proteinu v komplexu *Elf3* s *TGF-β* receptorem (3JTG) pro všechna silová pole.



Dalším potvrzením stability je graf B faktorů, které zjišťují dynamické chování atomů proteinu, tedy jejich lokální pohyb a flexibilitu. Vždy jsem sledovala především chování proteinu poblíž páteře DNA, kde by mohly velké fluktuace znamenat problematický popis interakce proteinu s DNA. Na prvním grafu B faktoru (Obrázek 14) je Ets doména vázající TGF- $\beta$  receptor. Jak lze v grafu vidět, v oblastech  $\alpha$  helixů a  $\beta$  listů se nenacházejí vysoké hodnoty B faktorů, které by značily vysokou fluktuaci fragmentů. V počáteční oblasti prvních reziduí, kde začíná první volný konec proteinu, je vysoká hodnota B faktoru, a to samé v koncové oblasti reziduí dochází k značnému nárůstu hodnot. Toto je běžná situace, protože volné konce proteinu a smyčky mnohou v roztoku relativně volně rekonformovat. Lze to pozorovat i u Loopů 1 a 2, kde je opět zvýšená hodnota B faktoru. Tyto hodnoty B faktoru jsou shodné ve všech třech polích, ve kterých byl daný komplex simulován.

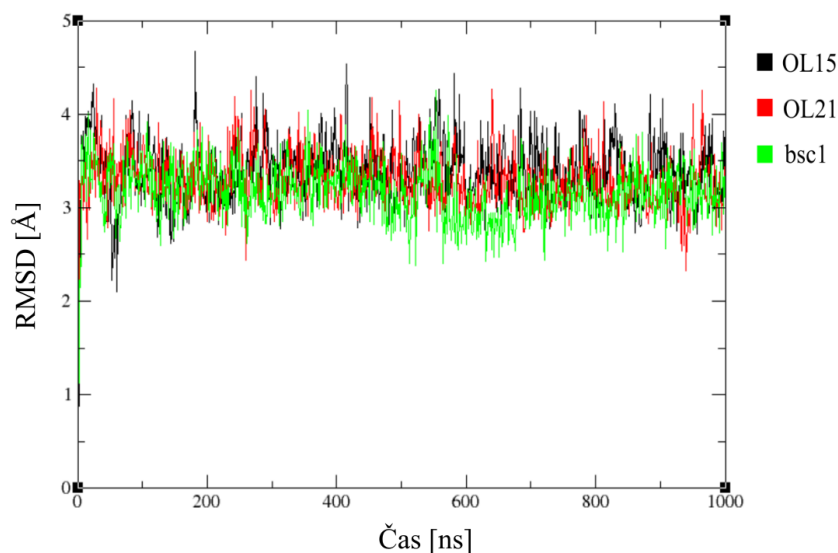
V červeně vyznačených úsecích je protein v kontaktu s páteří DNA. Hodnoty B faktoru zde nejsou nijak vysoké a naznačují spíše stabilní rozhraní proteinu s DNA. Loop 2 má sice vysoké hodnoty, ale koncová rezidua, která Loop 2 obsahuje, již nijak nezasahují do blízkosti páteře DNA, a tudíž ji nijak neovlivňují. Celkově je dle grafu B faktoru tento komplex spíše stabilní



Obrázek 14: B faktor komplexu *Elf3* s TGF- $\beta$  receptorem (3JTG) pro všechna silová pole.

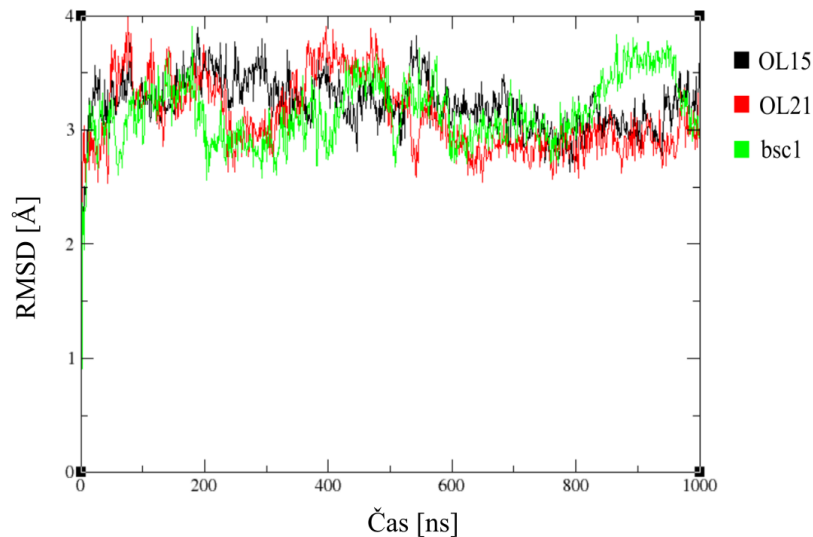
### 8.1.2. $\lambda$ cII-aktivátor transkripce, 1ZS4

Následují grafy RMSD pro  $\lambda$ cII-aktivátor transkripce. Prvním z nich je RMSD nukleové kyseliny (Obrázek 15), na němž je vidět, že je ve všech ohledech stabilní, drží se na hodnotách od 3 do 4 Å. Hodnoty jsou vyšší jednak proto, že se jedná o dlouhý fragment DNA (27 bází v jednom řetězci) a také kvůli volnému pohybu nespárovaných reziduí na koncích duplexu.



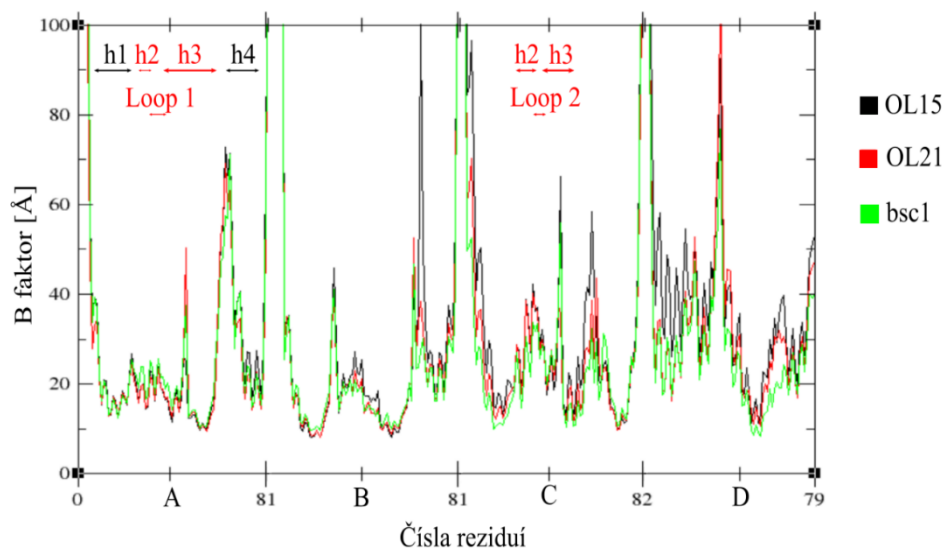
Obrázek 15: RMSD nukleové kyseliny v komplexu  $\lambda$ cII-aktivátor transkripce (1ZS4) pro všechna silová pole.

Další graf (Obrázek 16) ukazuje hodnoty RMSD pro protein tohoto komplexu. Zde jsou hodnoty RMSD od původního krystalu opět vyšší, což je dáno velikostí samotného proteinu skládajícího se ze čtyř podjednotek, který má tedy i více volných konců, které se během simulace náhodně pohybují. Opět tyto výkyvy nemají vliv na celkovou stabilitu komplexu a z grafů je patrné, že hodnoty pro všechna silová pole jsou podobné.



Obrázek 16: RMSD proteinu v komplexu  $\lambda$ cII-aktivátor transkripce (1ZS4) pro všechna silová pole.

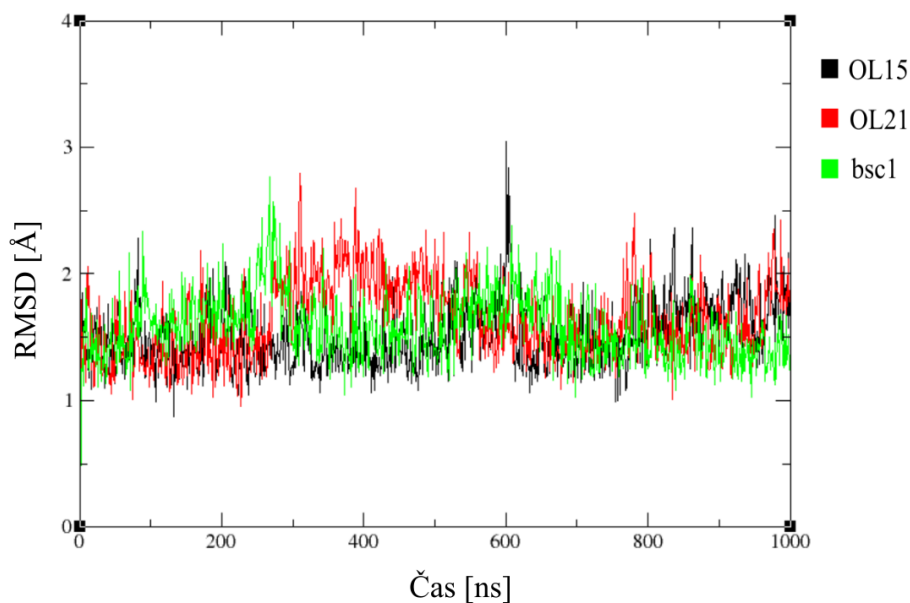
Graf B faktorů pro  $\lambda$ cII-aktivátor transkripce je ukázán na obrázku 17. Tento komplex je největší zkoumanou strukturou a jeho protein se skládá ze čtyř samostatných podjednotek (označených velkými písmeny A, B, C, D) obsahující mnoho volných konců, které by teoreticky mohly ovlivňovat páteř DNA. Avšak jak se dá z grafu vyčíst, v oblastech, kde se protein přibližuje páteři (oblasti jsou opět červeně označeny) nejsou hodnoty B faktoru nijak vysoké. V počátečních a koncových úsecích každé z podjednotek proteinu jsou opět vysoké hodnoty B faktoru, z důvodu jejich volných konců, které volně fluktuují v prostoru. Celkově můžeme tedy říci, že komplex je z tohoto pohledu stabilní. To platí shodně ve všech silových polích.



Obrázek 17: B faktor komplexu  $\lambda$ cII-aktivátor transkripce (1ZS4) pro všechna silová pole.

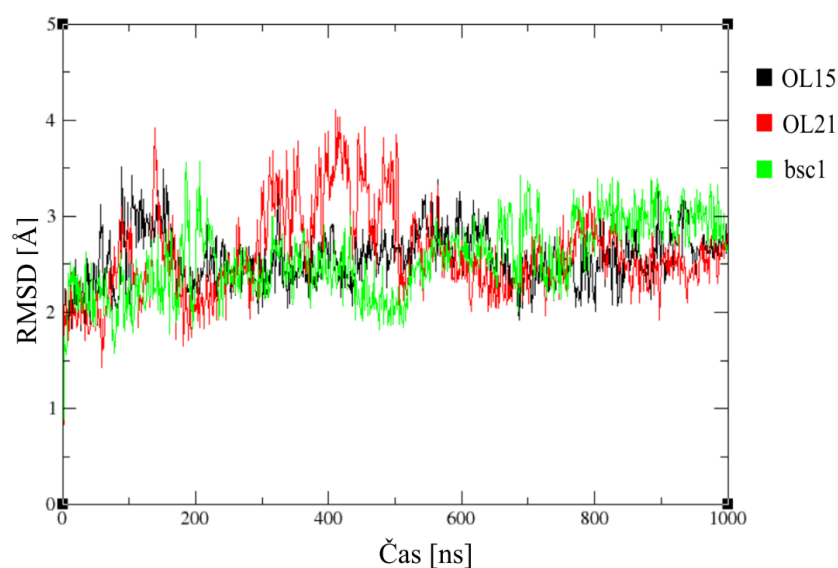
### 8.1.3. TATA box komplex, 1G2F

Následují poslední grafy RMSD pro komplexy obsahující konformer BB15, a to grafy pro TATA box komplex. Prvním z nich je opět graf RMSD pro nukleovou kyselinu (Obrázek 18), které je stabilní a nepřesahuje hodnotu 3 Å kromě jednoho výkyvu v poli OL15, kdy se volné konce DNA na chvíli rozpojily, ale následně se zase vrátily do původního stavu.



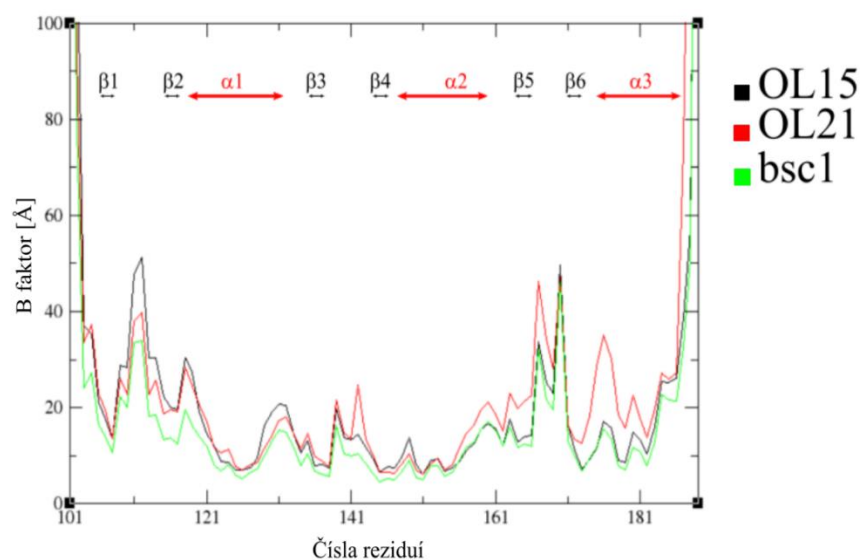
Obrázek 18: RMSD nukleové kyseliny v komplexu TATA boxu (1G2F) pro všechna silová pole.

Hodnoty RMSD pro protein (Obrázek 19) nepřesahují cca. 3-4 Å a jsou relativně stabilní okolo 2,5 Å, kam se také po odchýlení k hodnotě 4 Å vracují.



Obrázek 19: RMSD proteinu v komplexu TATA boxu (1G2F) pro všechna silová pole.

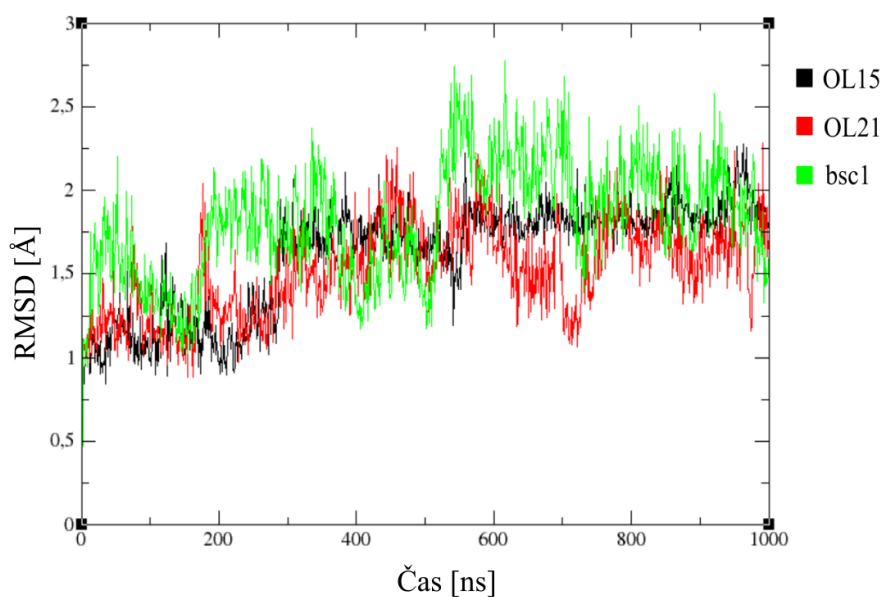
Z pohledu na graf B faktorů pro komplex TATA boxu (Obrázek 20) je patrné, že je opět stabilní. Vyšší hodnoty B faktorů v tomto grafu patří smyčce mezi rezidui  $\beta 1$  a  $\beta 2$  a smyčce mezi  $\beta 5$  a  $\beta 6$  listu a opět počátečnímu a koncovému volnému konci proteinu. V červeně vyznačených oblastech fragmentů  $\alpha$  helixů, které jsou v blízkosti páteře DNA, je míra fluktuace proteinu velmi nízká. U tohoto grafu B faktoru jako u jediného z prezentovaných se hodnoty liší pro různá silová pole, rozdíly jsou však spíše malé.



Obrázek 20: B faktor komplexu TATA boxu (1G2F) pro všechna silová pole.

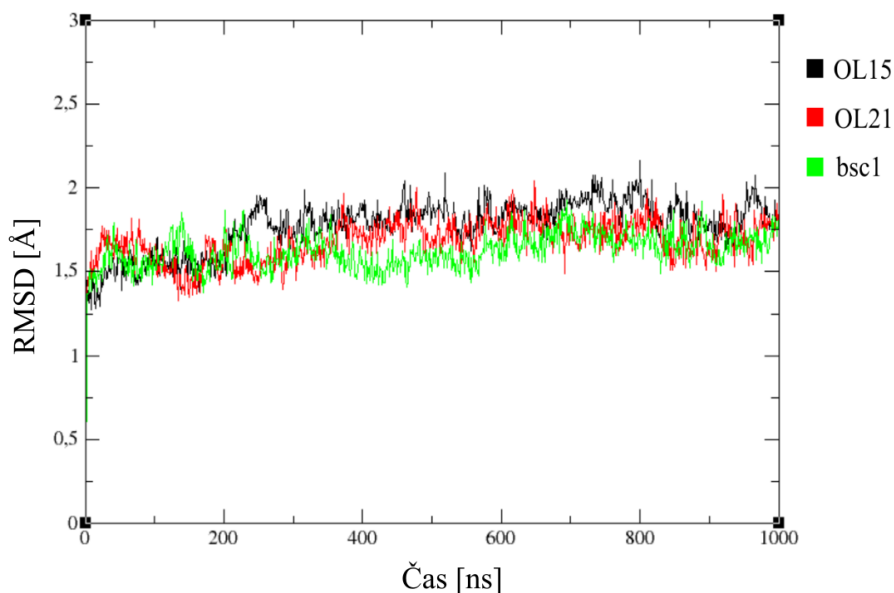
#### 8.1.4. LAGLIDADG homing endonukleáza, 1T9I

Hodnoty RMSD pro komplex endonukleázy LAGLIDADG (jediný zvolený zástupce třídy NtC BB12) jsou stabilnější a nižší oproti komplexům, které obsahovaly NtC konformer BB15. Prvním grafem je RMSD nukleové kyseliny (Obrázek 21), které i když se mírně liší pro jednotlivá silová pole, od začátku do asi poloviny simulace postupně vzrůstalo a poté se ustálilo a nepřesáhlo hodnotu 3 Å. Nejčastěji se RMSD pohybovalo na hodnotě okolo 2 Å, jedná se tedy o jedno z nejstabilnějších a nejnižších RMSD mezi mými komplexy.



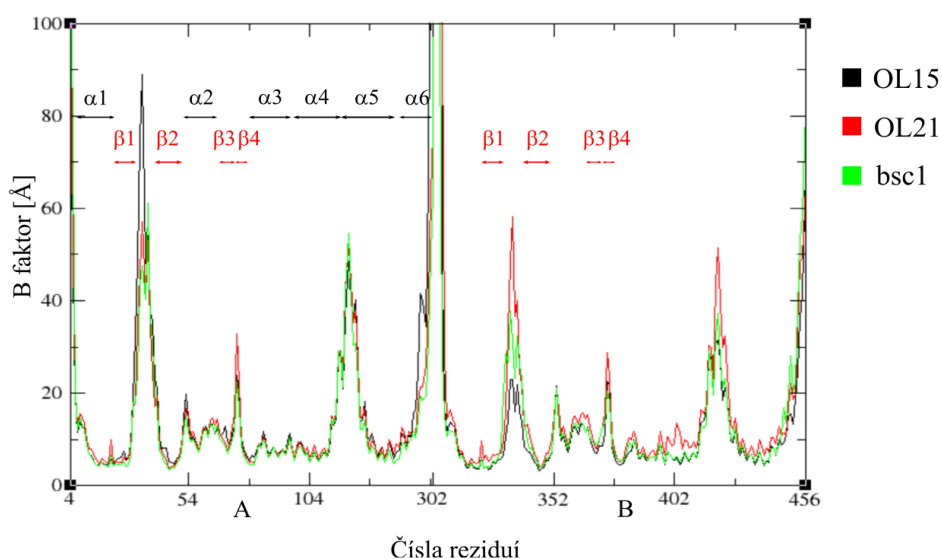
Obrázek 21: RMSD nukleové kyseliny v komplexu LAGLIDADG homing endonukleázy (1T9I) pro všechna silová pole.

Graf RMSD pro protein struktury 1T9I (Obrázek 22) je jedním s nejnižšími hodnotami pohybujícími se maximálně do 2 Å, což vypovídá o jeho největší stabilitě v průběhu daných simulací, a tedy se protein od původního krystalu odchyluje nejméně. To nám shodně potvrdilo RMSD pro simulace se všemi třemi silovými poli.



Obrázek 22: RMSD proteiny v komplexu LAGLIDADG homing endonukleázy (1T9I) pro všechna silová pole.

U tohoto komplexu se jeho protein skládá ze dvou identických podjednotek, u nichž jsou nejvyšší hodnoty B faktoru (Obrázek 23) v počátečních a koncových oblastech proteinu, kde jeho volné konce výrazně fluktuují v prostoru, a v oblasti, kde se nachází dlouhá smyčka mezi fragmenty  $\beta 1$  a  $\beta 2$  listu a smyčka mezi helixy  $\alpha 4$  a  $\alpha 5$ . Tyto fragmenty však nejsou v blízkosti páteře DNA, což znamená, že ji nijak neovlivňují, a tudíž je na základě grafu B faktoru i tento komplex stabilní a můžeme ho tedy dále analyzovat.



Obrázek 23: Graf B faktoru komplexu TATA boxu (1T9I) pro všechna silová pole.

## 8.2. Stabilita konformerů

Cílem mé práce bylo zjistit, jak stabilní jsou vybrané konformační rodiny páteře nukleových kyselin (NtC), BB15 a BB12, v MD simulacích komplexů proteinů s DNA. Začnu konformerem BB15, pro který jsou význačné úhly  $\alpha$  a  $\gamma$ , které mají oba neobvyklé hodnoty – úhel  $\gamma$  má hodnotu okolo  $360^\circ$  a úhel  $\alpha$  hodnotu okolo hodnot  $340^\circ$ . Pokud se hodnoty těchto úhlů v simulaci udrží (budou fluktuovat v okolí výchozí hodnoty z krystalové struktury), budu moci konformer BB15 považovat za stabilní.

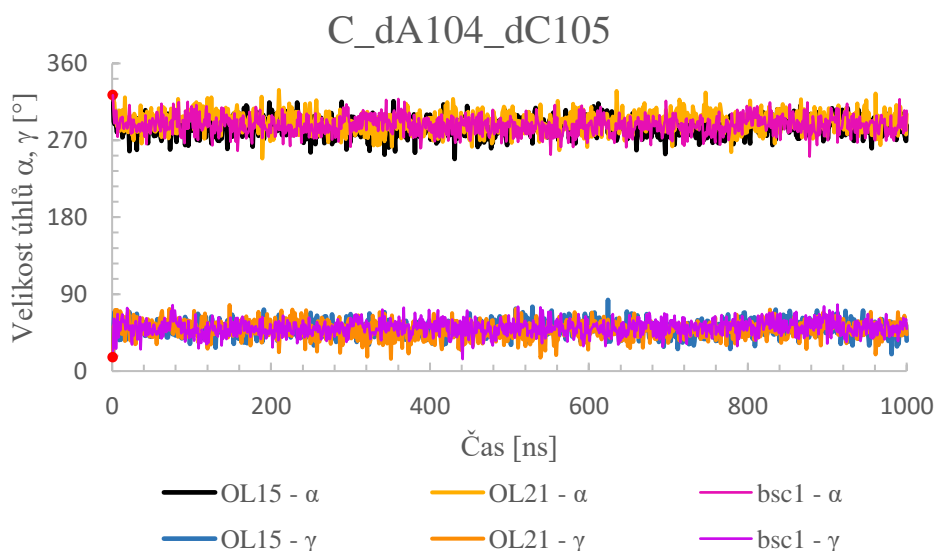
Následující grafy (graf č. 24-30) ukazují časové průběhy obou úhlů,  $\alpha$  i  $\gamma$ , ve třech různých komplexech proteinu s DNA v průběhu simulací dlouhých 1  $\mu\text{s}$  a prováděných ve všech třech testovaných silových polích. Z grafů je vidět, že ani jeden úhel se v žádném ze simulovaných komplexů neudržel a našel si energeticky výhodnější stav. Ve všech případech měl nový stav hodnoty úhlů  $\alpha \sim 300^\circ$  a  $\gamma \sim 60^\circ$ , což odpovídá geometrii nejběžnější (kanonické) DNA, která je označována jako NtC konformer BB00. K přechodu ze stavu BB15 do kanonického BB00 došlo ve všech případech velice rychle, již v prvních snímcích

MD simulace. Výchozí hodnoty úhlů  $\alpha$  a  $\gamma$  se pak už v simulaci neobjevily, ani na velmi krátkou chvíli. To by mohlo naznačovat, že stav BB15 je energeticky velice nevýhodný.

### 8.2.1. Stabilita konformeru BB15

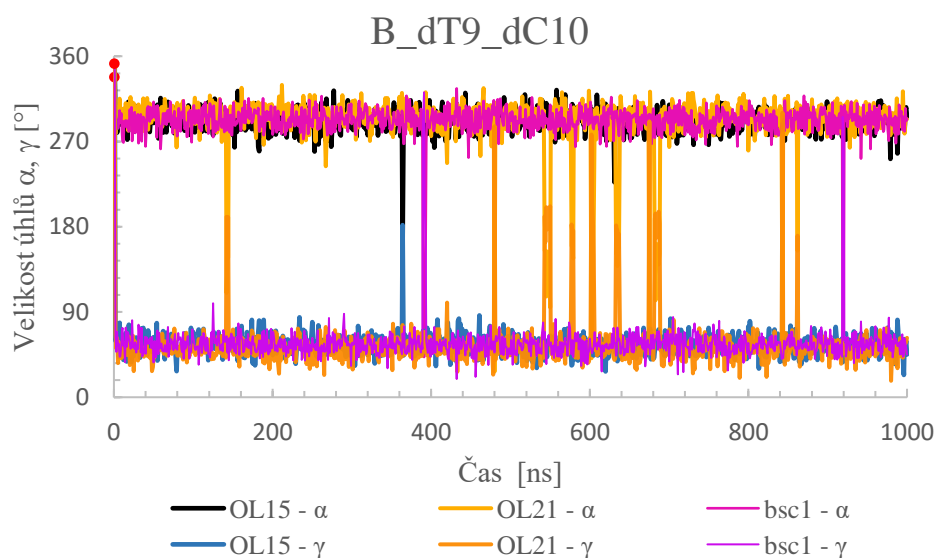
Ve většině simulací, ale byly pozorovány krátké časové intervaly, ve kterých úhly  $\alpha$  a  $\gamma$  navštěvovaly jiné regiony než kanonické BB00. Výjimkou je jen simulace dinukleotidu dA104\_dC105 (Obrázek 24) v řetězci C komplexu 3JTG, kde je kanonický stav BB00 zcela stabilní. Na jiném místě tohoto komplexu, v dinukleotidech dT9\_dC10 a dC10\_dC11 (Obrázek 25 a Obrázek 26) v řetězci B byly nakrátko navštěvovány oblasti s  $\alpha \sim 180^\circ$  a  $\gamma \sim 180^\circ$ , a to v malé míře v polích OL15 a bsc1 a o něco častěji v poli OL21. Stav s  $\alpha/\gamma \sim t/t$  se relativně často nachází v krystalových strukturách RNA (NtC AA01) a v menší míře i v DNA (BB03). Výskyt v našich simulacích tedy není pravděpodobně nežádoucí a jeho nízká frekvence odpovídá tomu, že jde o méně častý stav páteře. Tento stav se vyskytoval v dinukleotidu dA111\_dC112 v řetězci B (Obrázek 27). Také se zde našel stav s  $\alpha/\gamma \sim g+/t$  u dinukleotidu dC114\_dC115 v řetězci C (Obrázek 28). Také tyto stavy se dle databáze DNATCO v krystalových strukturách vyskytují, i když řídko.

Následují grafy (Obrázky 24-28) struktury Ets domény vázající TGF- $\beta$  receptor (3JTG) obsahující konformer BB15.

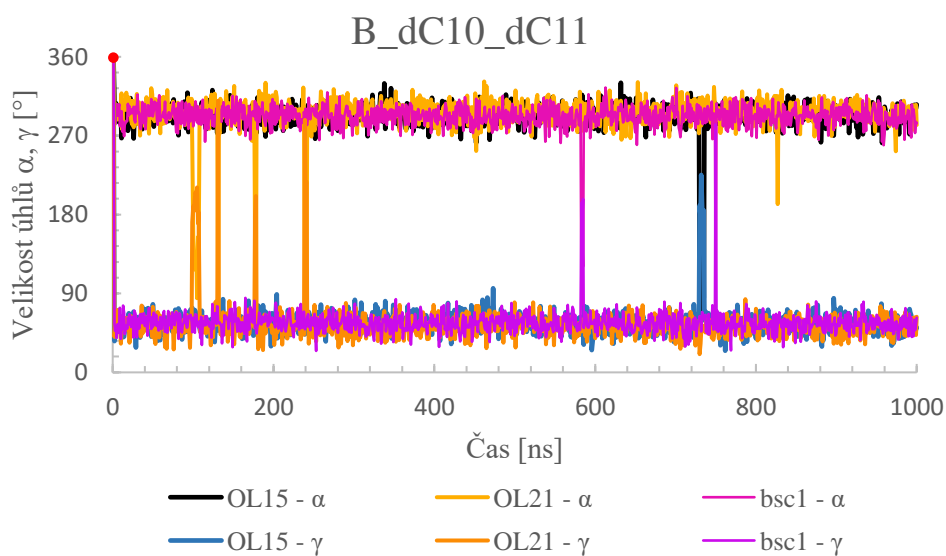


Obrázek 24: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB15 (rezidua: dA104\_dC105) pro strukturu Ets domény vázající TGF- $\beta$  receptor (3JTG). Tento dinukleotid si jako jediný udržel, bez jediného výkyvu, ve všech polích nové hodnoty okolo  $\alpha \sim 300^\circ$  a  $\gamma \sim 45^\circ$ , které odpovídají konformeru BB00.

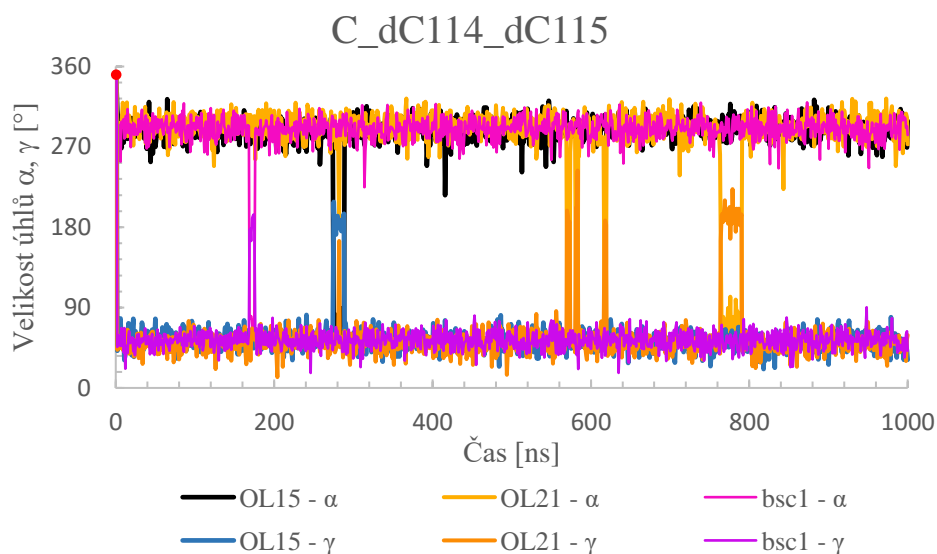




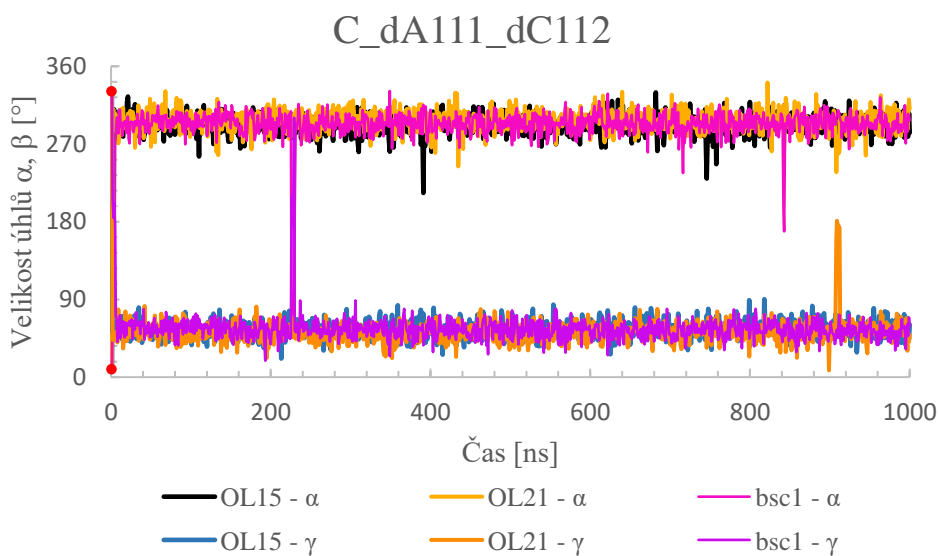
Obrázek 25: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB15 (rezidua: dT9\_dC10) pro strukturu Ets domény vazající TGF- $\beta$  receptor (3JTG). Východí krystalové hodnoty se v simulaci neudržely a byly nahrazeny hodnotami  $\alpha \sim 300^\circ$  a  $\gamma \sim 80^\circ$  (BB00).



Obrázek 26: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB15 (rezidua: dT10\_dT11) pro strukturu Ets domény vazající TGF- $\beta$  receptor (3JTG). Východí krystalové hodnoty se v simulaci neudržely a byly nahrazeny hodnotami  $\alpha \sim 300^\circ$  a  $\gamma \sim 70^\circ$  (BB00).

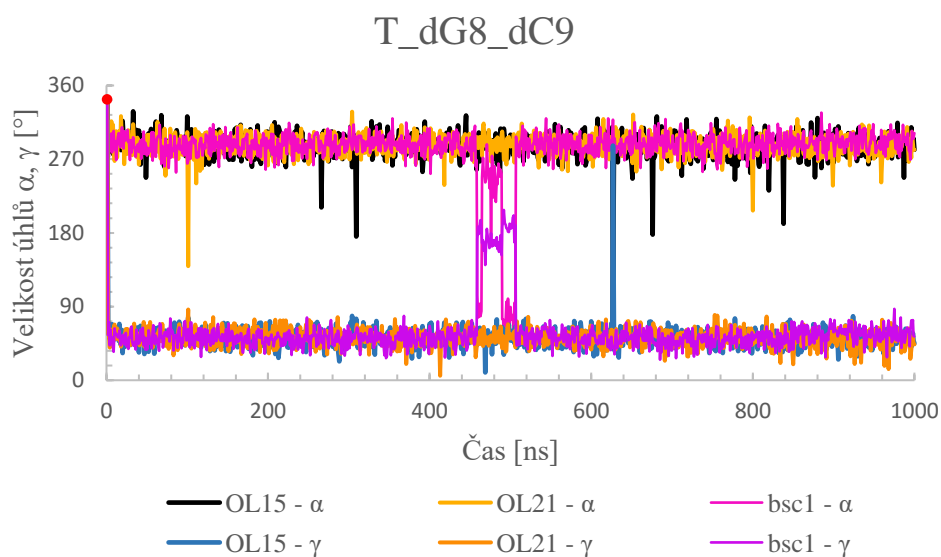


Obrázek 27: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB15 (rezidua: dT111\_dT112) pro strukturu Ets domény vázající TGF- $\beta$  receptor (3JTG). Zde se hodnoty opět podobají předešlým grafům, tedy  $\alpha \sim 300$  a  $\gamma \sim 80$ , bez návštev jiných oblastí.



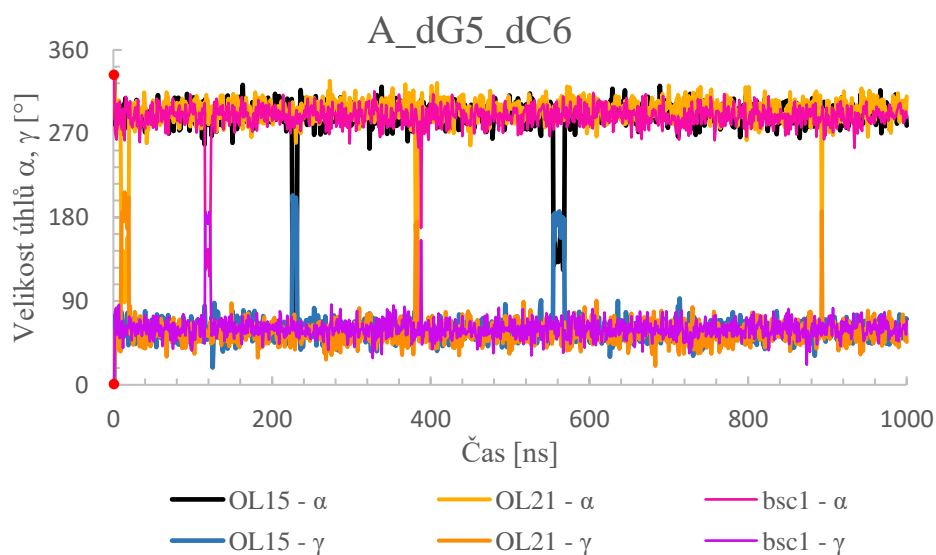
Obrázek 28: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB15 (rezidua: dC114\_dC115) pro strukturu Ets domény vázající TGF- $\beta$  receptor (3JTG). Poslední graf patřící struktuře 3JTG potvrzuje stejně jako grafy předešlé, že se hodnoty odpovídající konformeru BB15 neudržely.

Dalším zkoumaným zástupcem obsahujícím stav BB15 je struktura  $\lambda$ cII-aktivátoru transkripce (1ZS4). Ani v této struktuře se stav BB15 v MD simulaci neudržel a v prvních krocích simulace přešel na hodnoty blízké konformeru kanonické B-DNA (BB00). Kromě toho se v průběhu simulace v malé míře objevoval i stav s  $\alpha/\gamma \sim g/t$  u dinukleotidu dG8\_dC9 v řetězci T (Obrázek 29), ale pouze v poli bsc1.



Obrázek 29: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB15 (rezidua: dG8\_dC9) pro strukturu  $\lambda$ cII-aktivátor transkripce (1ZS4).

Posledním zástupcem obsahujícím konformer BB15 byla struktura TATA box komplexu (1G2F), která stejně jako předchozí struktury potvrzuje nestabilitu tohoto konformeru. Hodnoty obou úhlů dinukleotidu dG5\_dT6 řetězce A (Obrázek 30) zde častěji navštěvovaly oblasti  $\alpha \sim 180$  a  $\gamma \sim 180$ , kterým lze přiřadit již zmiňovaný konformer BB03.

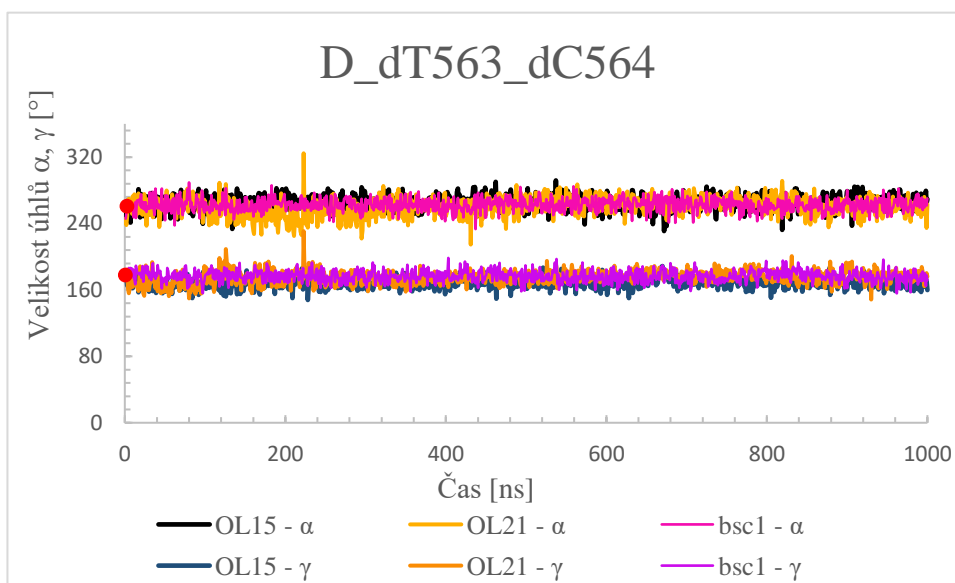


Obrázek 30: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB15 (rezidua: dG5\_dC6) pro strukturu TATA box komplexu (1G2F).

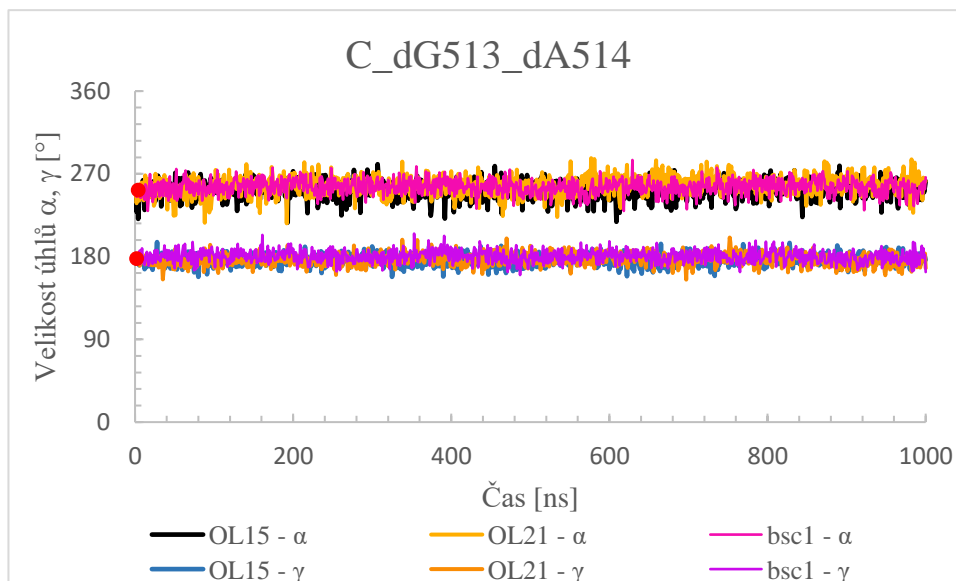
### 8.2.2. Stabilita konformeru BB12

Pro zkoumání stability nekanonických konformerů jsem si vybrala ještě jednoho zástupce, konformer třídy NtC BB12. Ten byl vybrán i proto, že již při pohledu na hodnoty úhlů výchozí geometrie  $\alpha \sim 260^\circ$  a  $\gamma \sim 170^\circ$ , se dalo předpokládat, že bude stabilnější než předchozí konformer BB15. Obě hodnoty se totiž nacházejí v oblasti, kde mají úhly  $\alpha$  a  $\gamma$  na svém potenciálu lokální minima. Komplex obsahuje čtyři výskyty stavu BB12.

Následují grafy (Obrázky 31-34) zobrazující hodnoty úhlů  $\alpha$  a  $\gamma$  struktury LAGLIDADG homing endonukleázy v závislosti na probíhajících simulacích, z nichž jasně vyplývá udržitelnost předpovězeného stavu konformeru BB12. Z grafů je vidět, že se úhly  $\alpha$  a  $\gamma$  prakticky po celou dobu molekulově dynamické simulace držely v blízkosti výchozích hodnot z krystalové struktury. Dinukleotidy dT563\_dC564 (Obrázek 31) a dG513\_dA514 (Obrázek 32) řetězce C jako jediné výlučně setrvaly v hodnotách konformeru BB12 a nenavštěvovaly žádné jiné stavy a potvrzují tak, že tento stav může být v našich MD simulacích zcela stabilní.

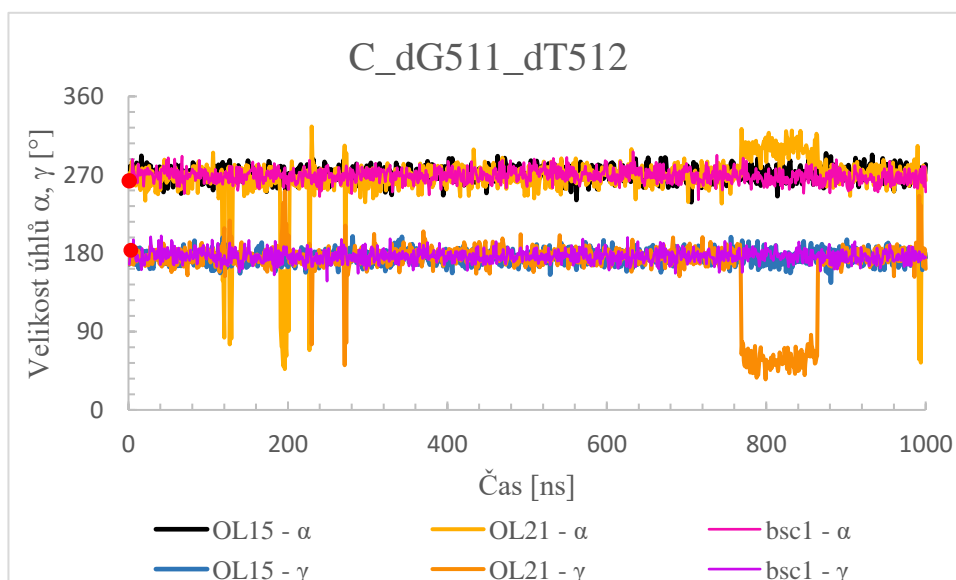


Obrázek 31: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB12 (rezidua: dT563\_dC564) pro strukturu LAGLIDADG homing endonukleázy (1T9I). Stav dinukleotidu se až na jediný výkyv udržel v hodnotách původního krystalu.



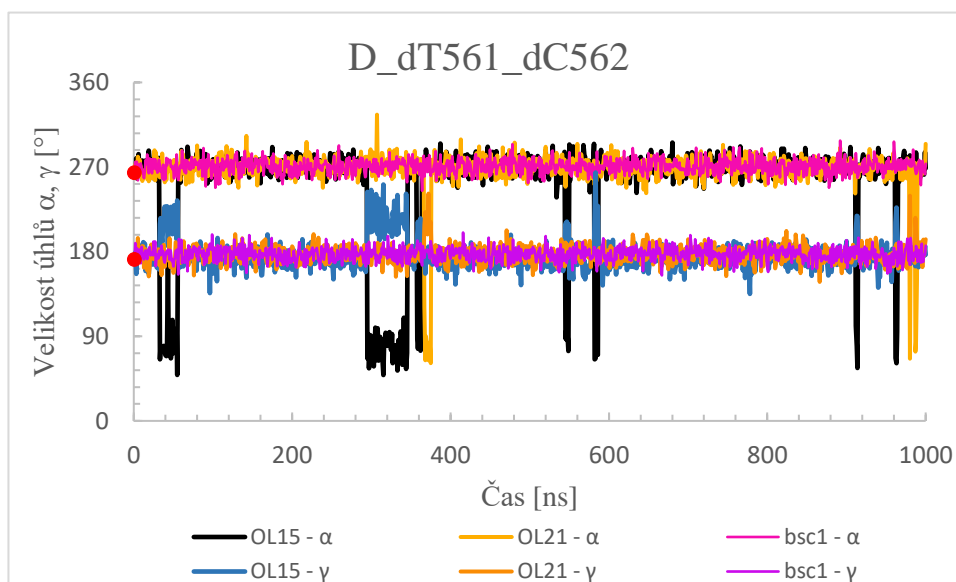
Obrázek 32: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB12 (rezidua: dG513\_dA514) pro strukturu LAGLIDADG homing endonukleázy (1T9I). Stav dinukleotidu se udržel v hodnotách původního krystalu.

U dinukleotidu dG511\_dT512 řetězce C (Obrázek 33), pouze v poli OL21, byla často navštěvována oblast s hodnotami  $\alpha \sim 300^\circ$  a  $\gamma \sim 55^\circ$ , která odpovídá konformeru NtC AB05, který je jedním z nejčastějších konformerů RNA. Avšak úhly se vždy vrátily do původních výchozích hodnot byly tak stabilní po většinu simulace.



Obrázek 33: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB12 (rezidua: dG511\_dT512) pro strukturu LAGLIDADG homing endonukleázy (1T9I). Dinukleotid navštěvuje oblasti s hodnotami okolo  $\alpha \sim 300^\circ$  a  $\gamma \sim 55^\circ$ , avšak se vždy vrátí do původních hodnot, tudíž je stav hodnot výchozí struktury udržen.

Dinukleotid dT561\_dC562 řetězce D (Obrázek 34) navštěvoval hodnoty  $\alpha \sim 70^\circ$  a  $\gamma \sim 220^\circ$ , které přibližně odpovídají konformeru NtC BB05 (konformer přemostující formy BI a BII).



Obrázek 34: Graf úhlů  $\alpha$  a  $\gamma$  konformeru BB12 (rezidua: dT561\_dC562) pro strukturu LAGLIDADG homing endonukléazy (1T9I). Z grafu je patrné, že dinukleotid chvilkově navštěvuje oblasti konformeru NtC BB05, ale po většinu času simulace se vrací do hodnot výchozí struktury.

### 8.3. MM potenciál a referenční křivka QM

Molekulově mechanický (MM) potenciál, popisující závislost energie na hodnotě úhlů  $\alpha$  a  $\gamma$ , může být srovnán s referenční potenciálovou křivkou získanou kvantově mechanickým (QM) výpočtem. MM i referenční potenciál jsou pro oba zkoumané úhly ukázány v Obrázcích 35 a 36. Jak MM, tak i QM křivka obsahuje vliv solvatační energie, jak je popsáno v referenci od P. Jurečka, M. Zgarbová a spol.<sup>(29)</sup>, ze které byly křivky převzaty.

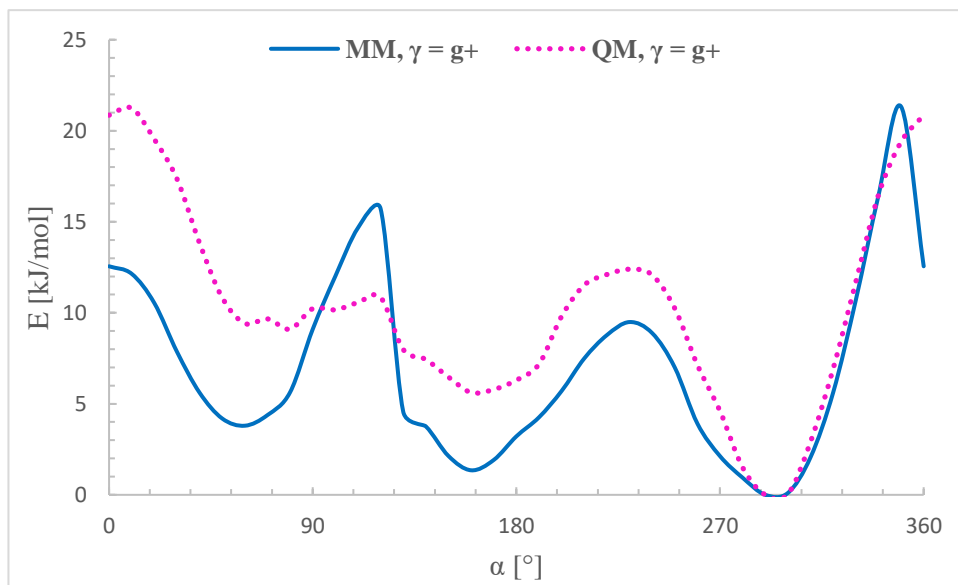
Zobrazené potenciály udávají, jaké hodnoty úhlů jsou pro dihedrály  $\alpha$  a  $\gamma$  energeticky nejvýhodnější. Na obou křivkách vidíme lokální minima v okolí  $180^\circ$  (trans, t) a  $+60^\circ$  (gauche+, g+) a  $-60^\circ$  (gauche-, g-). I když se MM a QM křivky v několika ohledech liší, obě vykazují tato tři minima s přibližně podobným pořadím energií.

Z předešlých grafů bylo zjištěno, že úhly s hodnotami  $\alpha \sim 340^\circ$  a  $\gamma \sim 360^\circ$ , které se nacházejí v NtC konformeru BB15, nebyly udržitelné ani v jedné z provedených simulací. Zde je díky křivkám MM a QM možné ukázat, proč tomu tak bylo. Úhel  $\alpha$  s hodnotou okolo

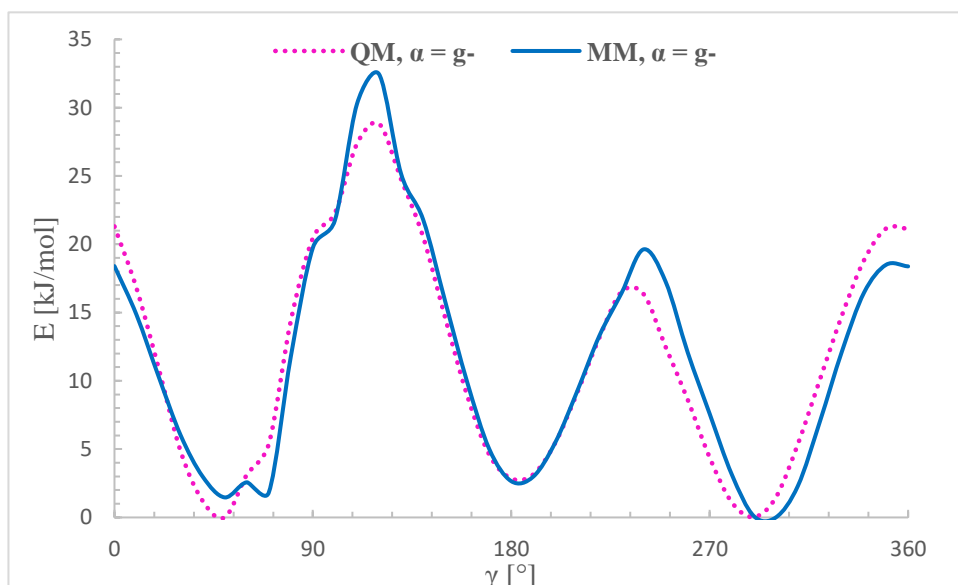
340° odpovídá velmi vysokým hodnotám energie a pro udržení tohoto stavu v simulaci by tak musela být vynaložena značná energie zvenčí – více než 20 kJ/mol. I když protein může svým působením stabilizovat energeticky vyšší nekanonické stavy, je nepravděpodobné, že by byl schopen stabilizovat stav, který je energeticky takto nevýhodný. Proto se stav BB15 v v simulacích neudržel a úhel  $\alpha$  si našel výhodnější hodnoty okolo 300°. Úhel 300° odpovídá globálnímu minimu s nejnižší možnou energií (Obrázek 35). To samé platí i pro úhel  $\gamma$ , který si také našel novou hodnotu v minimu 60°, tedy g- (Obrázek 36). Dle Boltzmannovy rovnice jsme zjistili, že pro energie odpovídající krystalovým geometriím dle uvedených grafů by pravděpodobnost výskytu stavu BB15 ve volných simulacích (bez ovlivnění proteinem) byla při pokojové teplotě nižší než 0,1 %.

Pro konformer BB12 je tomu naopak. Z předchozích MD simulací bylo zjištěno, že se hodnoty charakteristické pro tento konformer udržely bez větších potíží. Pro úhel  $\gamma \sim 170^\circ$ , který je blízký energetickému minimu, to nebyl takový problém. Úhel  $\alpha \sim 260^\circ$  se opět blížil energetickému minimu. Pro stabilizaci těchto úhlů nemuselo být vynaloženo tolik energie, kolik by bylo potřeba pro stabilizaci konformeru BB15.

Je také důležité poznamenat, že simulace byly provedeny s různými verzemi silových polí, které se lišily právě v tom, jak popisují potenciál  $\alpha$  a  $\gamma$ . Přesto byly výsledky simulací ve všech silových polích velice podobné. To naznačuje, že nestabilita stavu BB15 není působena nepřesností silového pole. Nepřesnosti silových polí jsou známé tím, že působí významné a nežádoucí odchylky simulací od reálného chování molekul a pole jsou proto často viněny z nespolehlivosti MD simulací. V tomto případě se však s největší pravděpodobností o takovou nepřesnost nejedná, protože výsledky různých parametrizací jsou v jednoznačné shodě.



Obrázek 35: Torzní potenciály MM a QM pro torzi  $\alpha$  ( $s$   $\gamma$  nastavenou na  $g+$ ). Energetická minima v hodnotách okolo 60°, 170° a 280°.



Obrázek 36: Torzní potenciály MM a QM pro torzi  $\gamma$  ( $s$   $\alpha$  nastavenou na  $g-$ ). Energetická minima v hodnotách okolo 60°, 180° a 280°.

Skutečnost, že se daný konformer nachází v mnoha výskytech v publikovaných krystalových strukturách, ještě nemusí znamenat, že se skutečně jedná o skutečně stabilní strukturu. Jedním problémem při interpretaci elektronových hustot je jejich malé rozlišení. V této práci jsem vybrala struktury s relativně vysokým rozlišením (3JTG 2,2 Å, 1ZS4 1,7 Å, 1G2F 2,0 Å a 1T9I 1,6 Å), přesto není toto rozlišení dostatečné k jednoznačnému určení poloh všech atomů s dostatečně malou chybou. Dalším problémem je použití (nepřesných) empirických potenciálů při fitování elektronové hustoty získané z rozptylových dat. Krystalografické struktury proto mohou někdy



obsahovat nerealistické konformace. Z předešlé diskuze vyplývá, že konformer BB15 by mohl být příkladem nepřesně určené krystalografické struktury.

## 9. ZÁVĚR

Vybraní zástupci nejčastějších nekanonických konformerů, nazvaní v NtC kategorizaci BB15 a BB12, byli podrobeni molekulově dynamickým simulacím ve třech silových polích z rodiny AMBER, OL15, bsc1 a nově vyvinutém OL21. Pole OL15 a bsc1 si jsou velmi podobná, oproti tomu silové pole OL21 se liší parametrizací dihedrálních úhlů cukr-fosfátové páteře  $\alpha$  a  $\gamma$ , což může mít vliv na chování konformerů BB15 a BB12 v komplexech proteinů s DNA.

Konformer NtC třídy BB15 se vyznačuje páteřními úhly  $\alpha \sim 340^\circ$  a  $\gamma \sim 0^\circ$ , a konformer BB12 úhly  $\alpha \sim 257^\circ$  a  $\gamma \sim 171^\circ$ . Zatímco hodnoty úhlů v BB15 jsou velmi neobvyklé a očekávala jsem, že by v MD simulaci mohly být nestabilní, hodnota úhlu  $\gamma$  v konformeru BB12 je sice nekanonická, ale odpovídá lokálnímu minimu na potenciálu  $\gamma$ , a mohla by tedy být stabilnější. Stabilitu obou konformerů jsem se snažila prozkoumat s pomocí MD simulací celkem sedm výskytů BB15 a čtyř výskytů BB12 v celkově čtyřech různých protein-DNA komplexech.

Nejdříve bylo nutno ověřit, zda jsou všechny simulace stabilní. To mi potvrdila analýza závislosti RMSD a B faktorů proteinu na čase, která ukázala relativně nízké RMSD odchylky i nízké B faktory pro proteinová rezidua v blízkosti nukleové kyseliny.

Pro všechny tři komplexy, které obsahovaly celkem sedm dinukleotidů s konformací BB15 potvrdil jednotný trend, kdy hned v prvním kroku simulace se hodnota obou úhlů,  $\alpha$  i  $\gamma$ , odchýlila od výchozí hodnoty z krystalografické analýzy a vždy přešla na energeticky výhodnější stav. Tím byl u většiny dinukleotidů stav odpovídající konformeru BB00, typickému pro kanonickou dvoušroubovici B-DNA. V této nové konformaci setrval dinukleotid po většinu simulace. V několika případech úhly  $\alpha$  a  $\gamma$  navštěvovaly i jiné regiony, jako jsou AA01 nebo BB03. Tento trend byl patrný pro všechna tři silová pole, nezávisle na jejich odlišných parametrech. Z toho lze usoudit na jednoznačnou nestabilitu konformeru BB15 v MD simulacích.

Pro jediný vybraný protein-DNA komplex, který obsahoval čtyři výskyty konformeru BB12, se ve všech výskytech potvrdila stabilita tohoto stavu ve všech provedených MD simulacích. Všechna tři silová pole tedy nezávisle potvrdila, že se hodnoty torzních úhlů se během simulací mohou udržet v blízkosti výchozích hodnot z krystalografické analýzy, a stav BB12 je tedy v mých simulacích zcela stabilní.

Tato pozorování jsou ve shodě s tvary kvantově mechanických a molekulově mechanických potenciálů pro úhly  $\alpha$  a  $\gamma$  dostupnými v literatuře. Křivky těchto potenciálů nám udávají závislost energie na hodnotě zvoleného úhlu a energeticky nejvýhodnější stavy pro  $\alpha$  a  $\gamma$ . Z nichž je patrné, že hodnoty v konformeru BB15 ( $\alpha \sim 340^\circ$  a  $\gamma \sim 0^\circ$ ) odpovídají vysokým hodnotám energie a na udržení takového stavu v dynamice by musel interagující protein vynaložit neobvykle vysokou interakční energii. Naproti tomu na stabilizaci hodnot úhlů v konformeru BB12 ( $\alpha \sim 257^\circ$  a  $\gamma \sim 171^\circ$ ), které jsou blízké energetickému minimu, není potřebná tak vysoká energie. Tvary MM i QM potenciálů tedy vysvětlují pozorovanou nestabilitu konformer BB15 a stabilitu konformeru BB12.

Výsledky mé práce mohou posloužit ke zlepšení strukturní interpretace krystalografických dat a zvýšení spolehlivosti krystalografických databází. Naznačují, že zatímco některé popsané konformační stavy páteře, jako je BB12, jsou odůvodnitelné a stabilní, jiné jako je BB15, jsou pravděpodobně nerealistické a mohou být výsledkem nesprávné interpretace mapy elektronových hustot.

## 10. SUMMARY

Selected representatives of the most common non-canonical conformers, named BB15 and BB12 in the NtC classification, were subjected to molecular dynamics simulations in three force fields from the AMBER family, OL15, bsc1 and the newly developed OL21. The OL15 and bsc1 force fields are very similar to each other, in contrast, the OL21 force field differs in the parameterization of the dihedral angles of the sugar-phosphate backbone  $\alpha$  and  $\gamma$ , which may affect the behavior of the BB15 and BB12 conformers in protein-DNA complexes.

The NtC class BB15 conformer is characterized by backbone angles of  $\alpha \sim 340^\circ$  and  $\gamma \sim 0^\circ$ , and the BB12 conformer by angles of  $\alpha \sim 257^\circ$  and  $\gamma \sim 171^\circ$ . While the values of the angles in BB15 are very unusual and I expected that they might be unstable in the MD simulation, the value of the angle  $\gamma$  in the BB12 conformer, although non-canonical, corresponds to a local minimum at the potential  $\gamma$ , and thus might be more stable. I tried to investigate the stability of both conformers using MD simulations of a total of seven occurrences of BB15 and four occurrences of BB12 in four different protein-DNA complexes.

First, it was necessary to verify that all simulations were stable. This was confirmed by my analysis of the time dependence of the RMSD and B factors of the protein, which showed relatively low RMSD deviations as well as low B factors for protein residues near the nucleic acid.

For all three complexes, which contained a total of seven dinucleotides with the BB15 conformation, a uniform trend was confirmed, where in the very first step of the simulation the value of both angles  $\alpha$  and  $\gamma$ , deviated from the initial value from the crystallographic analysis and always switched to the more energetically favorable state. This was, for most dinucleotides, the state corresponding to the BB00 conformer, typical of the canonical B-DNA double helix. The dinucleotide remained in this new conformation for most of the simulation. In a few cases, angles  $\alpha$  and  $\gamma$  visited other regions such as AA01 or BB03. This trend was evident for all three force fields, independent of their different parameters. This suggests a clear instability of the BB15 conformer in the MD simulations.

For a single selected protein-DNA complex containing four occurrences of the BB12 conformer, the stability of this state was confirmed in all MD simulations performed. Thus, all three force fields independently confirmed that the torsion angle values can be maintained close to the initial values from the crystallographic analysis during the simulations, and the BB12 state is therefore completely stable in my simulations.

These observations are in agreement with the shapes of the quantum mechanical and molecular mechanical potentials for the  $\alpha$  and  $\gamma$  angles available in the literature. The curves of these potentials give us the dependence of the energy on the value of the chosen angle and the most energetically favorable states for  $\alpha$  and  $\gamma$ . These show that the values in the BB15 conformer ( $\alpha \sim 340^\circ$  a  $\gamma \sim 0^\circ$ ) correspond to high energy values, and the interacting protein would have to expend unusually high interaction energy to maintain such a state in the dynamics. In contrast, such high energy is not required to stabilize the values of the angles in conformer BB12 ( $\alpha \sim 257^\circ$  a  $\gamma \sim 171^\circ$ ), which are close to the energy minimum. Thus, the shapes of the MM and QM potentials explain the observed instability of the BB15 conformer and the stability of the BB12 conformer.

The results of my work can be used to improve the structural interpretation of crystallographic data and increase the reliability of crystallographic databases. They suggest that while some of the described backbone conformers, such as BB12, are reasonable and stable, others, such as BB15, are likely unrealistic and may be the result of misinterpretation of the electron density map.

## 11. REFERENCE

- 1) GAVEZZOTTI, Angelo. *Molecular aggregation: Structure analysis and molecular simulation of crystals and liquids*. United Kingdom: Oxford university press, 2007. ISBN 978-0-19-967365-0.
- 2) METROPOLIS, Nicholas, Arianna W. ROSENBLUTH, Marshall N. ROSENBLUTH, Augusta H. TELLER, Edward TELLER a Jiří ČERNÝ. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*. 1953, **21**(6), 1087-1092. Dostupné z: doi:10.1063/1.1699114
- 3) CORNELL, Wendy D., Piotr CIEPLAK, Christopher I. BAYLY, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*. 1995, **117**(19), 5179-5197. Dostupné z: doi:10.1021/ja00124a002
- 4) Marečková, J. Skládání proteinů. Bakalářská práce, Univerzita Pardubice, Fakulta chemicko-technologická, Pardubice, 2019.
- 5) SÝSOVÁ, Kateřina. *Metody studia struktury proteinů*. Pardubice, 2017. Dostupné z: [https://dk.upce.cz/bitstream/handle/10195/71684/SysovaK\\_Struktury\\_proteinu\\_KV\\_2017.pdf?sequence=1&isAllowed=y](https://dk.upce.cz/bitstream/handle/10195/71684/SysovaK_Struktury_proteinu_KV_2017.pdf?sequence=1&isAllowed=y). Bakalářská práce. Univerzita Pardubice. Vedoucí práce Mgr. Katarína Vorčáková.
- 6) HOŠŤÁKOVÁ, Nina. *METODY PREDIKCE SEKUNDÁRNÍ STRUKTURY PROTEINŮ*. Brno, 2012. Dostupné z: <https://core.ac.uk/download/pdf/30281499.pdf>. Bakalářská práce. Vysoké učení technické v Brně. Vedoucí práce Ing. Denisa Madeřánková.
- 7) KODÍČEK, Milan. *Biochemické pojmy: výkladový slovník: Dihedrání úhel* [online]. 2007. Praha: VŠCHT, 2007. ISBN 978-80-7080-669-2. Dostupné z: [https://vydavatelstvi-old.vscht.cz/knihy/uid\\_es-002\\_v1/hesla/uhel\\_dihedralni.html](https://vydavatelstvi-old.vscht.cz/knihy/uid_es-002_v1/hesla/uhel_dihedralni.html)
- 8) RĘDZIKOWSKI, Adam, ed. *Ramachandranův diagram: Torzní úhly  $\phi$  and  $\psi$  ( $\omega$ ) hlavního řetězce peptidu*. 2011, 2011. Dostupné z: [https://cs.wikipedia.org/wiki/Ramachandran%C5%AFv\\_diagram#/media/Soubor:Protein\\_backbone\\_PhiPsiOmega\\_drawing.svg](https://cs.wikipedia.org/wiki/Ramachandran%C5%AFv_diagram#/media/Soubor:Protein_backbone_PhiPsiOmega_drawing.svg)
- 9) BERDÁR, Daniel. *Struktura DNA a nástroje její analýzy: nukleotid* [online]. Praha, 2021. Bakalářská práce. Karlova univerzita. Vedoucí práce Prof. Ing. Bohdan Schneider, CSc., DSc.
- 10) BENDA, Vladimír, Ivan BABŮREK a Pavel KOTRBA. *Základy biologie*. Praha: Vydavatelství VŠCHT, 2005. ISBN isbn80-7080-587-0.
- 11) BOSÁKOVÁ, Veronika. *Interakce proteinů s DNA se zaměřením na i-motivy: Ústav experimentální biologie*. Brno, 2020. Dostupné z: [https://is.muni.cz/th/usrgl/Diplomova\\_prace\\_Veronika\\_Bosakova.pdf](https://is.muni.cz/th/usrgl/Diplomova_prace_Veronika_Bosakova.pdf). Diplomová práce. Masarykova univerzita v Brně. Vedoucí práce Doc. Mgr. Václav Brázda, Ph.D.

- 12) POTAMAN, Vladimir N., SINDEN, Richard R., *DNA: Alternative Conformations and Biology*. In: Madame Curie Bioscience Database. Austin (TX): Landes Bioscience; 2000-2013. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK6545/>
- 13) WANG, Guliang a Karen M. VASQUEZ. Z-DNA, an active element in the genome. *Frontiers in Bioscience*. 2007, **12**(8-12), 6367-6381. Dostupné z: doi:10.2741/2399
- 14) HARDISON, Ross. *B-Form, A-Form, and Z-Form of DNA* [online]. The Pennsylvania State University, 2021. Dostupné z: <https://bio.libretexts.org/@go/page/307>
- 15) FRYKHOLM, Karolin, Lena K. NYBERG a Fredrik WESTERLUND. Exploring DNA–protein interactions on the single DNA molecule level using nanofluidic tools. *Integrative Biology*. 2017, **9**(8), 650-661. Dostupné z: doi:10.1039/C7IB00085E
- 16) GARRY, D a STORMO. *Introduction to protein-DNA interactions: structure, thermodynamics, and bioinformatics*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, 2013. ISBN 978-193-6113-507.
- 17) VON HIPPEL, Peter H. a James D. MCGHEE. *DNA-PROTEIN INTERACTIONS*. Institute of Molecular Biology and Departments of Chemistry and Biology University of Oregon, Eugene. Dostupné z: doi:10.1146/annurev.bi.41.070172.001311
- 18) Jiří Černý, Paulína Božíková, Bohdan Schneider, DNATCO: assignment of DNA conformers at [dnatco.org](http://dnatco.org), *Nucleic Acids Research*, Volume 44, Issue W1, 8 July 2016, Pages W284–W287, <https://doi.org/10.1093/nar/gkw381>
- 19) ČERNÝ, Jiří, Paulína BOŽÍKOVÁ, Jakub SVOBODA a Bohdan SCHNEIDER. A unified dinucleotide alphabet describing both RNA and DNA structures. *Nucleic Acids Research*. 2020, **48**(11), 6367-6381. Dostupné z: doi:10.1093/nar/gkaa383
- 20) *Regulace genové exprese u prokaryot: pozitivní a negativní regulace*. Brno: Mendelova univerzita v Brně, 2016. Dostupné z: [http://user.mendelu.cz/urban/vsg1/molekul/mol\\_regul2.html#](http://user.mendelu.cz/urban/vsg1/molekul/mol_regul2.html#)
- 21) BOSÁKOVÁ, Veronika. *Interakce proteinů s DNA se zaměřením na i-motivy: Ústav experimentální biologie*. Brno, 2020. Dostupné z: [https://is.muni.cz/th/usrgl/Diplomova\\_prace\\_Veronika\\_Bosakova.pdf](https://is.muni.cz/th/usrgl/Diplomova_prace_Veronika_Bosakova.pdf). Diplomová práce. Masarykova univerzita v Brně. Vedoucí práce Doc. Mgr. Václav Brázda, Ph.D.
- 22) AGARKAR, Vinod B., Nigar D. BABAYEVA, Phillip J. WILDER, Angie RIZZINO a Tahir H. TAHIROV. Crystal Structure of Mouse Elf3 C-terminal DNA-binding Domain in Complex with Type II TGF- $\beta$  Receptor Promoter DNA. *Journal of Molecular Biology*. 2010, **397**(1), 278-289. Dostupné z: doi:10.1016/j.jmb.2010.01.017
- 23) JAIN, Deepti, Youngchang KIM, Karen L. MAXWELL, Steven BEASLEY, Rongguang ZHANG, Gary N. GUSSIN, Aled M. EDWARDS a Seth A. DARST. Crystal Structure of Bacteriophage  $\lambda$ cII and Its DNA Complex. *Molecular Cell*. 2005, **19**(2), 259-269. Dostupné z: doi:10.1016/j.molcel.2005.06.006

- 24) RAZIN, S. V., V. V. BORUNOVA, O. G. MAKSIMENKO a O. L. KANTIDZE. Cys2His2 zinc finger protein family: Classification, functions, and major members. *Biochemistry (Moscow)*. 2012, **77**(3), 217-226. Dostupné z: doi:10.1134/S0006297912030017
- 25) WOLFE, Scot A., Robert A. GRANT, Monicia ELROD-ERICKSON a Carl O. PABO. Beyond the “Recognition Code”: structures of two Cys2His2 zinc finger/TATA box complexes. *Structure*. 2001, **9**(8), 717-723. Dostupné z: doi:10.1016/S0969-2126(01)00632-3
- 26) GODBEY, W.T. *Genes. An Introduction to Biotechnology*. Elsevier, 2014, 2014, s. 65-105. ISBN 9781907568282. Dostupné z: doi:10.1016/B978-1-907568-28-2.00004-6
- 27) CHEVALIER, Brett, Django SUSSMAN, Christian OTIS, et al. Metal-Dependent DNA Cleavage Mechanism of the I-Cre I LAGLIDADG Homing Endonuclease. *Biochemistry*. 2004, **43**(44), 14015-14026. Dostupné z: doi:10.1021/bi048970c
- 28) *Acta Crystallographica Section D Structural Biology*. 74. 2018. Dostupné z: <http://scripts.iucr.org/cgi-bin/paper?S2059798318000050>
- 29) JUREČKA, Petr, Marie ZGARBOVÁ, Filip ČERNÝ a Jan SALOMON. *Continuous B- to A-Transition in Protein-DNA Binding - How Well Is It Described by Current AMBER Force Fields?*. Department of Physical Chemistry, Faculty of Science, Palacky University, 17. listopadu 12, 77146 Olomouc, Czech Republic, 2022, 1-8. Dostupné z: doi:10.1101/2022.01.13.476176
- 30) JOUNG, In Suk a Thomas E. CHEATHAM, III. *Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations*. Department of Bioengineering, College of Engineering, and Departments of Medicinal Chemistry and of Pharmaceutics and Pharmaceutical Chemistry, January 8, 2008. 112, 9020–9041. Dostupné z: <https://pubs.acs.org/doi/pdf/10.1021/jp8001614>
- 31) BERENDSEN, H. J. C., J. R. GRIGERA a T. P. STRAATSMA. The missing term in effective pair potentials. *The Journal of Physical Chemistry*. 1987, **91**(24), 6269-6271.. Dostupné z: doi:10.1021/j100308a038
- 32) PETERS, Martin B., Yue YANG, Bing WANG, László FÜSTI-MOLNÁR, Michael N. WEAVER a Kenneth M. MERZ. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *Journal of Chemical Theory and Computation*. 2010, **6**(9), 2935-2947. Dostupné z: doi:10.1021/ct1002626
- 33) LI, Pengfei, Benjamin P. ROBERTS, Dhruva K. CHAKRAVORTY a Kenneth M. MERZ. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for 2 Metal Cations in Explicit Solvent. *Journal of Chemical Theory and Computation*. 2013, **9**(6), 2733-2748. ISSN 1549-9618. Dostupné z: doi:10.1021/ct400146w
- 34) HOPKINS, Chad W., Scott LE GRAND, Ross C. WALKER a Adrian E. ROITBERG. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*. 2015, **11**(4), 1864-1874. Dostupné z: doi:10.1021/ct5010406



- 35) HOM. Amber Histidine HIS-HID-HIE-HIP Problem. *Hom's Blog*. 2015, 2015-12-26. Dostupné z: <http://platinhom.github.io/2015/12/26/HIS-HID-HIE-HIP/>
- 36) LI, Hui, Andrew D. ROBERTSON a Jan H. JENSEN. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics*. 2005, **61**(4), 704-721. ISSN 08873585. Dostupné z: doi:10.1002/prot.20660