

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra fyzikální chemie



MAPOVÁNÍ POLÁRNÍCH INTERAKCÍ V KOMPLEXECH PROTEINŮ S DNA

Bakalářská práce

Autor: Tomáš Nesvadba

Studijní program: B14107 Chemie

Studijní obor: Aplikovaná chemie

Vedoucí práce: doc. RNDr. Petr Jurečka, Ph.D.

Olomouc 2022

Prohlašuji, že jsem tuto práci vypracoval samostatně pod vedením doc. RNDr. Petra Jurečky, Ph.D. Veškeré literární a odborné zdroje jsem uvedl na konci práce.

V Olomouci dne.....

.....

podpis

Tímto bych chtěl poděkovat doc. RNDr. Petru Jurečkovi, Ph.D. za veškerou pomoc a cenné rady, které mi během psaní mé bakalářské práce poskytl. Dále bych chtěl poděkovat za obrovskou ochotu a trpělivost, kterou se mnou po celou dobu spolupráce měl.

Bibliografická identifikace

Jméno autora: Tomáš Nesvadba

Název práce: Mapování polárních interakcí v komplexech proteinů s DNA

Typ práce: Bakalářská

Pracoviště: Katedra fyzikální chemie, Přírodovědecká fakulta UP

Vedoucí práce: doc. RNDr. Petr Jurečka, Ph.D.

Rok obhajoby: 2022

Abstrakt:

Cílem práce bylo zmapovat stabilitu polárních interakcí u protein-DNA komplexů v molekulově dynamických simulacích. Celkem jsem studoval tři protein-DNA komplexy, u kterých jsem se zaměřil převážně na interakce, které se nazývají solné můstky, jež jsou zde formovány kladně nabitými aminokyselinami lysinem a argininem a záporně nabitými páteřními fosfáty v DNA. Simulace jsem realizoval v silovém poli OL15. Získané výsledky mohou v budoucnu pomoci při zpřesňování silových polí.

Klíčová slova: protein, DNA, interakce, solný můstek, molekulová dynamika

Počet stran: 50

Jazyk: čeština

Bibliographical identification

Author: Tomáš Nesvadba

Title: Mapping polar interactions in protein-DNA complexes

Type of thesis: Bachelor

Department: Department of physical Chemistry, Faculty of science UP

Supervisor: doc. RNDr. Petr Jurečka, Ph.D.

The year of presentation: 2022

Abstract:

The aim of this work was to map the stability of polar interactions in protein-DNA complexes in molecular dynamics simulations. In total, I studied three protein-DNA complexes, where I focused mainly on interactions called salt bridges, which are formed by the positively charged amino acids lysine and arginine and negatively charged backbone phosphates in DNA. The simulations were carried out in the OL15 force field. The results obtained may help to refine the force fields in the future.

Key words: protein, DNA, interaction, salt bridge, molecular dynamics

Number of pages: 50

Language: czech

OBSAH

OBSAH	6
1. ÚVOD	8
2. TEORETICKÁ ČÁST	9
2.1 Nukleové kyseliny	9
2.1.1 Stavba nukleových kyselin.....	10
2.1.2 Struktura DNA	10
2.1.3 Formy DNA	12
2.2 Proteiny	13
2.2.1 Struktura proteinů.....	13
2.2.2 Primární struktura.....	14
2.2.3 Sekundární struktura	14
2.2.4 Terciární struktura	15
2.2.5 Kvarterní struktura	16
2.3 Protein-DNA komplexy	16
2.4 Interakce v komplexech protein-DNA	16
2.4.1 1SKN.....	17
2.4.2 3EYI	18
2.4.3 1MNN	18
2.5 Molekulová mechanika	19
2.6 Molekulová dynamika.....	21
3. CÍLE PRÁCE	22
4. EXPERIMENTÁLNÍ ČÁST	23
4.1 Metody	23
4.2 Výsledky a diskuse.....	25
4.2.1 Stabilita studovaných komplexů v MD simulacích	25
4.2.2 Protein-DNA interakce.....	27

4.2.3 Interakce v 1SKN	30
4.2.4 Interakce v 3EYI	33
4.2.5 Interakce v 1MNN.....	35
4.2.6 Souhrnné poznatky	37
5. ZÁVĚR	43
6. SUMMARY	45
7. LITERATURA.....	48

1. ÚVOD

To, že proteiny jsou bílkoviny a DNA je nositelkou genetické informace, ví snad každý. Ještě aby ne, tyto biologické makromolekulární látky patří mezi jedny z nejnepostradatelnějších látek pro téměř všechny organismy. Proteiny plní v organismech funkci stavební, katalytickou, transportní, zásobní či pohybovou. DNA uchovává a přenáší genetické informace, čímž ovlivňuje vlastnosti celého organismu.

Spojení proteinu a DNA v komplex hraje v organismu důležitou roli. Tyto komplexy mají vliv na různé procesy, které se v buňce odehrávají. Nejčastěji ovlivňují transkripci, replikaci či opravu DNA. Dochází v nich také k řadě důležitých interakcí, které ovlivňují jejich celkovou strukturu a funkci. Interakce proteinů s DNA se dělí na sekvenčně specifické a nespecifické. Každá z těchto interakcí ovlivňuje protein-DNA komplexy jiným způsobem.

Díky neustálému zdokonalování a zrychlování molekulově dynamických simulací se teoretické studium těchto jedinečných struktur stává stále jednodušším a dostupnějším. Nemalý podíl na tom má také zlepšování silových polí pro zkoumání nukleových kyselin, hlavně DNA, které pomáhá získávat daleko přesnější a jednoznačnější výsledky.

Ve své práci se budu zajímat především nespecifickými interakcemi mezi kladně a záporně nabitými rezidui, tzv. solnými můstky, a jejich popisem současnými silovými poli. Pro analýzu jsem využil silové pole OL15 a zaměřím se v něm hlavně na stabilitu vybraných interakcí v molekulově dynamických simulacích.

2. TEORETICKÁ ČÁST

2.1 Nukleové kyseliny

Nukleové kyseliny jsou nepostradatelné biologické makromolekulární látky, které tvoří polynukleotidový řetězec. Při fyziologickém pH se chovají jako kyseliny, proto se nazývají nukleové kyseliny. Patří mezi jednu ze čtyř velkých skupin biomolekul, do kterých se řadí společně s bílkovinami, lipidy a sacharidy. Jejich základní funkcí je přenos a uchování genetické informace. Přítomny jsou ve všech buňkách, kde se nejčastěji vyskytují v jádře. Mohou se však vyskytovat i v malých množstvích mimo buněčná jádra, konkrétně v mitochondriích, cytoplazmě či chloroplastech. Dále je můžeme nalézt v bakteriích či virech.

Rozlišujeme dva typy nukleových kyselin, které se od sebe liší cukernou složkou a jednou ze čtyř bází. Rozdíl mezi sacharidy obsaženými v nukleových kyselinách je na druhém uhlíku, kdy na pozici druhého uhlíku u 2-deoxy-D-ribosy se nachází navázaný pouze vodík místo hydroxylové skupiny. Prvním typem je ribonukleová kyselina známá pod zkratkou RNA, ta ve své struktuře obsahuje D-ribosu a dusíkatou bázi uracil namísto thyminu, jenž je obsažen naopak u deoxyribonukleové kyseliny, známější pod svou zkratkou DNA. Pentóza obsažená ve struktuře DNA je 2-deoxy-D-ribosa.

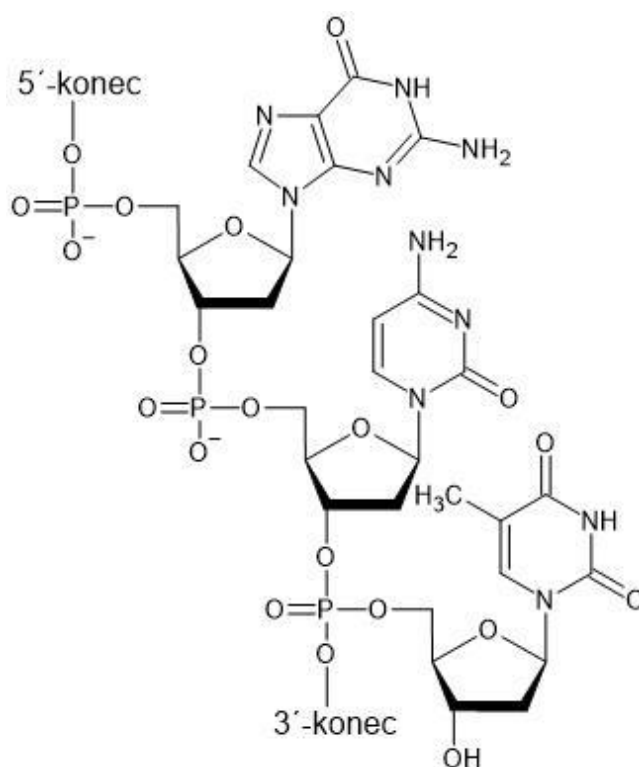
Za první významný krok lze v oblasti výzkumu nukleových kyselin považovat jejich objevení švýcarským lékařem a přírodovědcem Friedrichem Miescherem v roce 1869. Miescher se pokoušel izolovat buňky pro své experimenty z lymfatických uzlin, což bylo příliš náročné. Nakonec vzorky pro své experimenty získal z hnisu na čerstvých chirurgických obvazech, které sbíral na chirurgické klinice. DNA se mu povedlo izolovat z bílých krvinek, ale nebyl schopný ji dále zkoumat, jelikož neměl dostatečné množství. Tuto novou látku nazval „nuklein“, jelikož se vyskytovala v buněčných jádrech. [1] Dalším důležitým milníkem v oblasti zkoumání nukleových kyselin byly rentgenové difrakční snímky vláken DNA, jež se podařilo získat Rosalindě Franklinové a Maurici Wilkinsovi roku 1952. Tohoto v následujícím roce využili James D. Watson a Francis H. C. Crick, kteří stanovili strukturu DNA a sestavili její model, což publikovali v časopise Nature pod názvem „A Structure for Deoxyribose Nucleic Acid“. [2] V roce 1962 jim následně za to byla udělena Nobelova cena.

2.1.1 Stavba nukleových kyselin

Makromolekula DNA vytváří stabilní dvoušroubovici, která se skládá ze dvou komplementárních polynukleotidových vláken, které jsou k sobě antiparalelní, tzn. že polarita obou vláken je navzájem opačná. Konec prvního řetězce se označuje jako 3' konec a je na něm navázaná hydroxylová skupina. Druhý konec se značí 5' a je zakončen fosfátovou skupinou. RNA se většinou skládá pouze z jednoho různě stočeného vlákna. Můžeme se setkat i s A-RNA, což je označení pro dvoušroubovici RNA. [3]

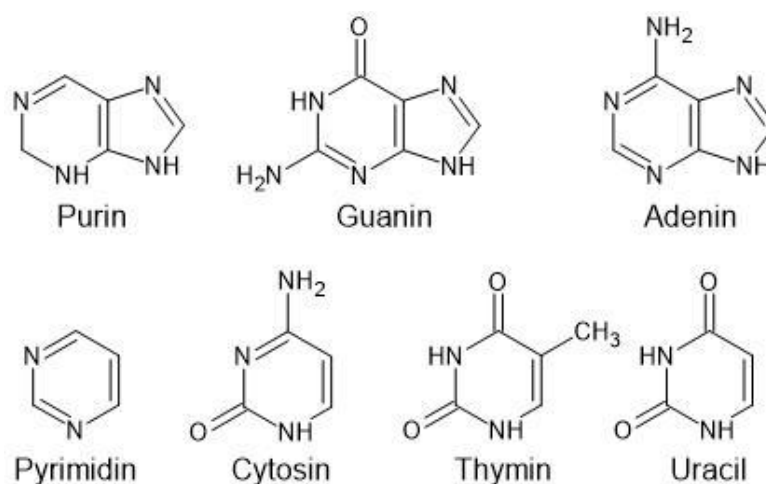
2.1.2 Struktura DNA

Základní stavební jednotkou DNA jsou nukleotidy, které se do polynukleotidového řetězce (Obrázek 1) formují pomocí fosfodiesterových vazeb, kdy se fosfátová skupina na 5. uhlíku pentosy prvního nukleotidu spojuje s hydroxylovou skupinou na 3. uhlíku pentosy druhého nukleotidu. Nukleotidy jsou monomerní jednotky skládající se ze tří základních složek. První částí nukleotidu je dusíkatá heterocyklická báze. Další částí je pěti uhlíkatý sacharid (pentosa). Dusíkatá báze a pentosa jsou spojeny pomocí N-glykosidové vazby do tzv. nukleosidu. Posledním dílkem nukleotidu je fosfát.



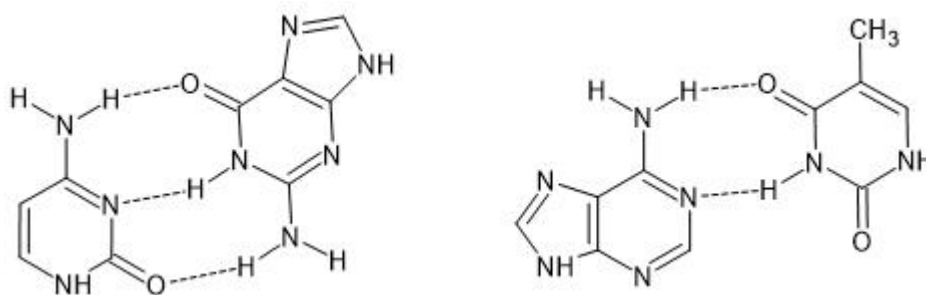
Obrázek 1: Nukleotidy GCT spojeny v polynukleotidový řetězec (vytvořeno v programu ChemSketch)

Základem nukleových bází jsou aromatické heterocyklické sloučeniny purin a pyrimidin, proto se nukleové báze dělí do dvou základních skupin podle toho, od které sloučeniny jsou odvozené. Mezi purinové báze se řadí adenin a guanin. Do pyrimidinových bází patří cytosin s thyminem, ale také uracil, který nahrazuje thymin ve struktuře RNA. Na obrázku 2 jsou pak znázorněny jednotlivé báze. Dusíkaté báze jsou spojeny se sacharidy N-glykosidovou vazbou. U purinových bází dochází ke vzniku vazby mezi dusíkem na pozici číslo 9 a uhlíkem C1' u sacharidu. Pyrimidinové báze se k sacharidovému uhlíku C1' váží přes dusík na první pozici.



Obrázek 2: Nukleové báze odvozené od purinu a pyrimidinu (vytvořeno v programu ChemSketch)

Nukleové báze ve své molekule obsahují donory i akceptory vodíkových vazeb. To hraje velkou roli pro stabilizaci dvoušroubovice DNA, jelikož dochází ke vzájemnému spojení dvou bází mezi sebou pomocí nekovalentní vodíkové vazby. Párování bází není nijak náhodné, ale podléhá Watson-Crickovským pravidlům komplementarity. U DNA se vyskytují dva páry komplementárních bází, při kterém vždy dochází k párování purinové báze s pyrimidinovou bází. První pár tvoří guanin s cytosinem, mezi kterými jsou přítomny tři vodíkové vazby. Ve druhém páru se vyskytují vodíkové vazby dvě a tento pár tvoří adenin s thyminem (viz Obrázek 3). Kromě Watson-Crickovských pravidel komplementarity se ve struktuře DNA setkáme ještě s dalším pravidlem, kterým je pravidlo Chargaffovo. To pojednává o obsahu nukleových bází v DNA a říká, že se počet guaninů rovná s počtem cytosinů, a to samé platí i pro adenin s thyminem.



Obrázek 3: Komplementární páry cytosinu s guaninem a adeninu s thyminem (vytvořeno v programu ChemSketch)

V nukleových kyselinách se sacharidy obecně vyskytují v cyklické furanózové formě. V polynukleotidovém řetězci DNA se jedná konkrétně o 2-deoxy-D-ribózu. Uhlíky tohoto sacharidu jsou číslovány po směru hodinových ručiček. Uhlíky, které jsou zapojeny do formování polynukleotidového řetězce jsou C1', který na sebe váže nukleovou bázi. Dalšími uhlíky jsou C3' a C5', ty zprostředkovávají vazbu s fosfáty.

Poslední složkou nukleotidu je zbytek kyseliny fosforečné, tzv. fosfát. Fosfát udává celkový záporný náboj makromolekule DNA, protože se na fosfátu vyskytují záporně nabitě kyslíky. Záporný náboj kyslíků je způsoben odštěpením vodíku z jedné OH skupiny ve vodném prostředí a delokalizací výsledného náboje na ostatní kyslíky. Záporný náboj fosfátu hraje důležitou roli v interakcích proteinu s DNA. S kladně nabitými aminokyselinovými zbytky argininu a lysinu se totiž mohou elektrostaticky vázat a vytvářet tzv. solný můstek.

2.1.3 Formy DNA

Nejstabilnější a nejběžnější formou DNA, která se ve vodném prostředí a v živých organismech může vyskytovat, je takzvaná B-DNA, označována též kanonická forma. B-DNA je pravotočivá dvoušroubovice. Jedna helikální otočka B-DNA je tvořena přibližně 10,5 párů bází, které směřují dovnitř šroubovice a k ose šroubovice jsou téměř kolmé. Vzdálenost mezi jednotlivými bázemi je pak 3,4 Å. [4]

Tato forma DNA byla dlouhou dobu považována za jedinou formu, avšak Watson s Crickem se domnívali, že DNA je ovlivňována prostředím, ve kterém se vyskytuje, díky čemuž se může měnit i její struktura či tvar. [5] Podle vlastností prostředí je DNA schopna se všelijak otáčet, roztahovat a jinak formovat, což potvrzuje domněnku Watsona s Crickem.

Další pravotočivou formou deoxyribonukleových kyselin je A-DNA, jež se lehce podobá B-DNA. Na první pohled si lze všimnout, že A-DNA je širší a plošší. To je

zapříčiněno rozdílnou konformací deoxyribosy v polynukleotidovém řetězci. Dalším rozdílem je, že na jednu otáčku připadá 11 nukleových bází. Pro svůj výskyt potřebuje A-DNA prostředí se sníženým obsahem vody, což lze zapříčinit zvýšením koncentrace solí nebo přidáním organického rozpouštědla do roztoku. [6]

Z-DNA se liší oproti výše zmíněným formám v mnoha ohledech. V polynukleotidovém řetězci obsahuje dinukleotidy, které se pravidelně opakují a na jednu otáčku připadá šest dinukleotidů. Dinukleotidy jsou tvořeny ze střídajících se guaninových a cytosinových nukleotidů. Zásadní rozdíl je pak ten, že se šroubovice Z-DNA otáčí na levou stranu a její páteř je nepravidelná. To způsobuje odlišná konformace nukleových bází. [4,5]

2.2 Proteiny

Proteiny neboli bílkoviny se řadí stejně jako nukleové kyseliny do skupiny velkých biomolekul, kam patří ještě sacharidy a lipidy. Jedná se o organické makromolekulární látky, které jsou základní součástí živých organismů. V organismu hrají důležitou roli, jelikož patří mezi jeden ze zdrojů energie a dusíku, působí jako enzymy a také udržují strukturu buněk a tkání. Proteiny se skládají z aminokyselinových zbytků. Tyto aminokyselinové zbytky jsou spojeny pomocí peptidových vazeb, které vznikají mezi karboxylovou skupinou první aminokyseliny s aminoskupinou následující aminokyselin. Spojením několika po sobě jdoucích aminokyselin vzniká polypeptidový řetězec. Podle počtu aminokyselin spojených do řetězce rozlišujeme oligopeptidy, kde se spojuje 2 - 10 aminokyselin, dále polypeptidy, obsahující 10 - 100 aminokyselin, a nakonec bílkoviny, jež jsou tvořeny 100 a více aminokyselinami.

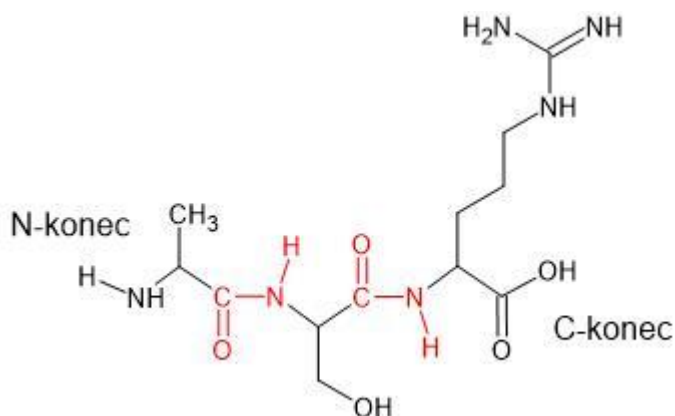
2.2.1 Struktura proteinů

O první úspěšné rozluštění struktury proteinů se v 50. letech 20. století zasloužil biochemik britského původu Frederick Sanger. Jemu se podařilo stanovit sekvenci aminokyselin v inzulinu. V roce 1958 za tento objev obdržel Nobelovu cenu v oblasti chemie. [7]

Ve struktuře proteinů se vyskytují atomy uhlíku, kyslíku, dusíku a vodíku. Malé procentuální zastoupení tvoří atomy síry a selenu. Struktura proteinu hraje důležitou roli v interakcích s jinými molekulami. Podle molekuly, se kterou proteiny interagují, se mohou dělit na glykoproteiny (interakce se sacharidy), lipoproteiny (interakce s lipidy), nukleoproteiny (interakce s nukleovými kyselinami) či metaloproteiny (obsahují kationty kovů). Interakci zprostředkovávají postranní řetězce jednotlivých aminokyselin, které určují i fyzikální a chemické vlastnosti proteinu.

2.2.2 Primární struktura

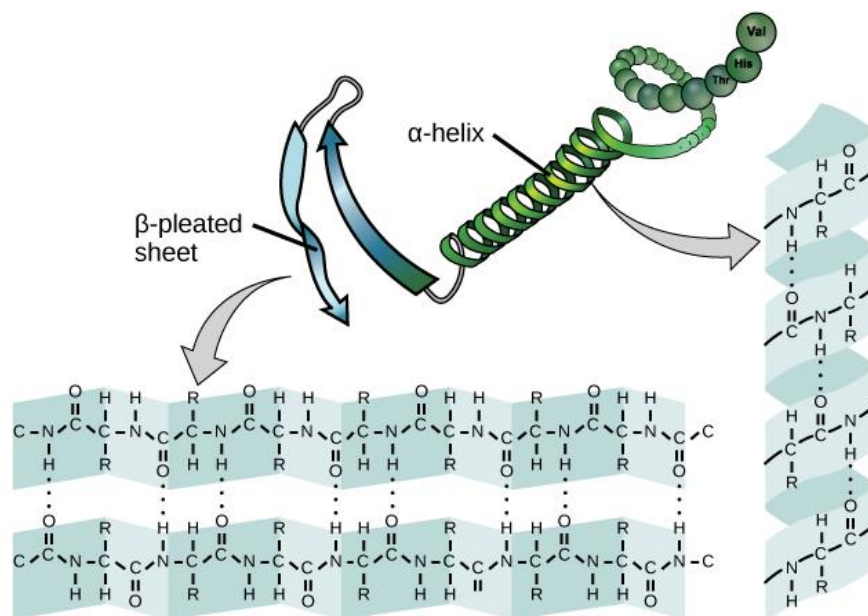
Primární strukturou (Obrázek 4) proteinu se označuje pořadí jednotlivých aminokyselin spojených v lineárním řetězci proteinu pomocí peptidových vazeb. Pořadí aminokyselin se čte od N-konce po C-konec. N-koncem rozumíme volnou aminoskupinu na jedné straně řetězce, na druhé straně řetězce se pak nachází volná karboxylová skupina čili C-konec. Z primární struktury se dále odvíjí vyšší stupeň organizace prostorového uspořádání molekuly. Primární struktura každého proteinu je jedinečná, což je způsobeno sekvencí nukleotidů v DNA při syntéze bílkovin. [8]



Obrázek 4: Sekvence aminokyselin ASR od N-konce po C-konec, peptidové vazby znázorněny červeně (vytvořeno v programu ChemSketch)

2.2.3 Sekundární struktura

Sekundární struktura proteinu se odvíjí od interakcí mezi částmi proteinového řetězce, které se nachází poblíž sebe. Vlivem těchto interakcí se protein skládá a formuje do trojrozměrného tvaru. Různé části proteinu se skládají do jedné ze dvou základních struktur. Těmito strukturami jsou alfa-helix a beta-skládaný list, které jsou zobrazeny na obrázku 5.



Obrázek 5: sekundární struktura proteinů; převzato a upraveno z [9]

Alfa-helix se nejčastěji vyskytuje ve formě pravotočivé šroubovice. Na jednu otáčku šroubovice připadá 3,6 aminokyselin. Formování do šroubovice je způsobeno vodíkovými vazbami, které se tvoří mezi C=O a N-H skupinami polypeptidové páteře, jež jsou od sebe vzdáleny o čtyři aminokyseliny. [8] Tyto vazby se také podílejí na stabilizaci šroubovice. Aminokyselinové postranní řetězce, které se nezapojují do stabilizace a formování šroubovice, vyčnívají směrem ven ze šroubovice, kde probíhají interakce s jinými molekulami.

Beta-skládaný list je tvořen dvěma řetězci, ty jsou navzájem rovnoběžné a antiparalelní. Jejich stabilizaci opět způsobuje přítomnost vodíkových vazeb mezi C=O a N-H skupinami polypeptidové páteře. Beta-skládaný list se dělí podle vzájemné orientace amino- a karboxylových skupin. Jestliže jsou amino- a karboxylové skupiny orientovány do stejného směru jedná se o paralelní β -list. Pokud jsou orientace obou skupin opačné, jedná se o antiparalelní β -list. [8]

2.2.4 Terciární struktura

Terciární struktura je konečná trojrozměrná forma, do které se protein tvaruje. Na jeho stabilizaci se podílí vodíkové vazby, které vznikají mezi vzdálenými atomy vedlejších řetězců aminokyselin, čímž se liší od sekundární struktury. Další interakce formující terciární strukturu jsou elektrostatické síly, disulfidové můstky či nepolární interakce. [10] Podle finálního tvaru rozlišujeme dva druhy proteinů: fibrilární proteiny (skleroproteiny) a globulární proteiny (sféroproteiny). Fibrilární proteiny jsou ve formě

podélných vláken. Globulární proteiny se skládají do kulovitého tvaru s nerovnoměrným povrchem.

2.2.5 Kvarterní struktura

Kvarterní struktura vzniká spojením dvou a více podjednotek proteinů pomocí nekovalentních vazeb, nikoliv peptidových vazeb. Podjednotkou se označuje jeden polypeptidový řetězec, který je do struktury zapojen. Místo, kde dochází ke vzniku nekovalentní vazby, se říká vazebné místo. Jestliže se v kvarterní struktuře vyskytují stejné podjednotky, nazýváme je homodimery. Rozdílné podjednotky jsou pak heterodimery. [11]

2.3 Protein-DNA komplexy

Protein-DNA komplexy obsahují ve struktuře alespoň jeden proteinový řetězec a jednu dvoušroubovici DNA, které spolu vzájemně interagují. Jejich interakce hraje důležitou roli v genetice organismu, jelikož mají vliv na různé procesy, které se v buňce odehrávají, jako je například transkripce, replikace či oprava DNA. [12]

V organismech se nachází celá řada proteinových struktur, které jsou schopné se vázat na DNA. Ve většině případů se jedná o enzymy. Podle funkce DNA vazebných proteinů je můžeme rozdělit do několika skupin. Největší skupinu tvoří transkripční faktory, které, jak už je z názvu zřejmé, ovlivňují transkripci. Další skupinou, ovlivňující genovou expresi jsou DNA replikační faktory, jejichž přítomnost hraje důležitou roli při replikaci DNA. DNA reparační faktory (proteiny) se dělí do několika skupin, jimiž jsou například proteiny opravující báze či celé nukleotidy v dvoušroubovici. Mezi DNA reparační proteiny se řadí také faktory opravující chybné vložení bází, jež může vzniknout při replikaci či rekombinaci DNA. [13] Další DNA vazebné proteiny patří do skupin jako jsou například telomerické proteiny, strukturní proteiny chromatinu nebo proteiny štěpící nukleové kyseliny. [14]

2.4 Interakce v komplexech protein-DNA

Interakce proteinů a DNA často doprovází změny ve struktuře obou makromolekul. Během interakcí dochází k tomu, že protein i DNA přizpůsobují svůj tvar tak, aby vzniklý kontakt mezi nimi byl optimální. Tyto změny ve struktuře a tvaru jsou stabilizovány základními silami, jimiž jsou elektrostatické síly, hydrofobní síly či van der Waalsovy síly. [15] Nejčastějším místem interakce je okolí velkého a malého žlábků nukleové kyseliny. Větší podíl interakcí je pak ve velkém žlábků, jelikož je lépe

přístupný pro interakci, a navíc je zde přítomno i větší množství donorů a akceptorů vodíkových vazeb ve formě aminoskupin či ketoskupin. [16]

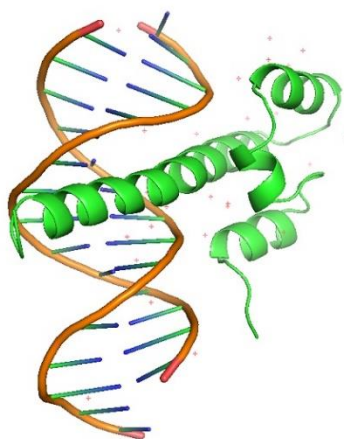
Při specifických interakcích dochází ke vzniku vodíkových vazeb mezi proteinem a nukleovou kyselinou. Specifické interakce v protein-DNA komplexech lze rozdělit na dvě skupiny podle volby vazebného místa. Base readout je případ, kdy protein rozpoznává specifické sekvence bází v DNA. Dále lze base readout rozdělit podle toho, jestli se interakce vyskytují ve velkém nebo malém žlábků. Druhým mechanismem je shape readout, v tomto případě protein rozpoznává tvar DNA. Analogicky se pak shape readout dělí na globální rozpoznávání tvaru a rozpoznání místního tvaru. [17]

Nespecifické interakce jsou na rozdíl od specifických interakcí formovány pomocí elektrostatických sil. Ke vzniku kontaktu dochází mezi kladně nabitými postranními řetězci argininu a lysinu v proteinovém řetězci a záporně nabitými fosfátovými skupinami, které tvoří páteř nukleové kyseliny. Tato forma interakce se také často nazývá solný můstek. [18]

V mé práci se budu zabývat hlavně nespecifickými interakcemi, zprostředkovanými často právě solnými můstky. Vybral jsem si následující tři komplexy, které podrobím molekulově dynamickým simulacím a následně analýzám stability polárních nekovalentních interakcí.

2.4.1 1SKN

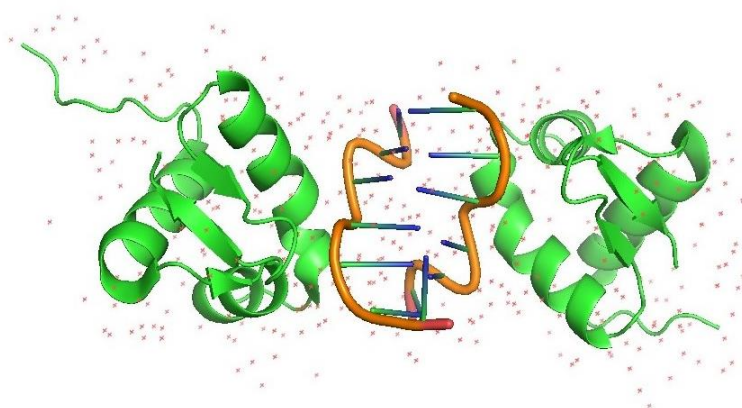
Prvním ze tří protein-DNA komplexů, kterým se budu v této práci zabývat, je zástupce transkripčních faktorů, který nese označení s PDB kódem 1SKN (Obrázek 6). Jedná se o základní mateřský vývojový transkripční faktor, který specifikuje mesodermální a endodermální fázi u embrya hád'átka obecného. DNA je ve formě B-DNA a je v interakci s jedním proteinovým řetězcem. Interakce probíhá převážně ve velkém žlábků, kde interaguje se šroubovicí proteinu podobně jako skupina proteinů bZip. [19]



Obrázek 6: Struktura ISKN (vytvořeno v programu PyMOL)

2.4.2 3EYI

Dalším studovaným komplexem, který jsem si zvolil je komplex s PDB označením 3EYI (Obrázek 7). Jedná se o DNA-dependentní aktivátor regulačních faktorů IFN. Fungují jako okamžitý buněčný obranný mechanismus v savčích buňkách, kde jsou v roli aktivátoru vrozené imunitní odpovědi v případě invaze cizích organismů dovnitř buňky. [20] V krystalové struktuře jsou na DNA, která je ve formě Z-DNA, vázány dva identické proteinové řetězce.

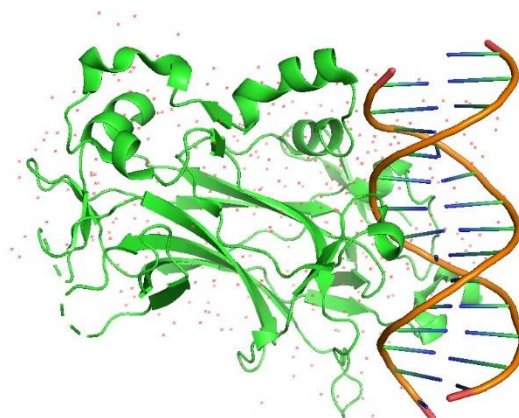


Obrázek 7: Struktura 3EYI (vytvořeno v programu PyMOL)

2.4.3 1MNN

Poslední protein-DNA komplex, jež jsem si vybral, nese PDB označení 1MNN (Obrázek 8). Tento komplex můžeme zařadit do skupiny transkripčních faktorů. 1MNN se podílí na průběhu sporulace u kvasinky pивní. Aktivace komplexu je zahájena

po dokončení meiotické rekombinace, kdy se protein naváže na specifickou sekvenci DNA. [21] Stejně jako u 1SKN je v interakci jeden proteinový řetězec s jednou nukleovou kyselinou ve formě B-DNA.



Obrázek 8: Struktura IMNN (vytvořeno v programu PyMOL)

2.5 Molekulová mechanika

Na rozdíl od kvantové mechaniky, molekulová mechanika ignoruje pohyby elektronů a počítá energii systému jako funkci jaderných poloh. Molekulární mechanika je založena na jednoduchých interakcích uvnitř systému jako je napínání vazeb, otevírání a uzavírání úhlů anebo rotace kolem jednoduchých vazeb. Energie těchto jevů je v molekulové mechanice popsána pomocí empirického potenciálu (silového pole). [22]

Jednotlivé interakce hrají v systému zásadní roli a dělí se na dva typy interakcí, které jsou zahrnuty ve výpočtu sil a energie celého systému. Prvním typem jsou vazebné (kovalentní) interakce, kam se řadí délka vazby, vazebné úhly anebo torzní úhly. Druhým typem interakcí jsou nevazebné (nekovalentní) interakce, mezi které jsou zahrnuty van der Waalovy síly a elektrostatické, nebo též Coulombické síly. Spojením příspěvků kovalentních a nekovalentních interakcí získáme rovnici 1, díky níž se vypočítá celková energie systému. [22]

$$E_{pot} = E_{bonds} + E_{angles} + E_{torsions} + E_{vdW} + E_C \quad (1)$$

Jednotlivé členy rovnice 1 lze pak rozdělit na pět rovnic, z nichž každá popisuje danou energii interakce. První člen rovnice 1 určuje vazebnou energii, ta se počítá pomocí rovnice 2,

$$E_{bonds} = \sum K_r(r - r_0)^2 \quad (2)$$

kde K_r je silová konstanta, r je vzdálenost vazby a r_0 je hodnota rovnovážné vzdálenosti.

Druhý člen udává energii vazebných úhlů, která lze spočítat pomocí rovnice 3,

$$E_{angles} = \sum K_\theta (\theta - \theta_0)^2 \quad (3)$$

kde K_θ je silová konstanta vazebného úhlu, θ je hodnota vazebného úhlu a θ_0 je rovnovážná hodnota vazebného úhlu.

Poslední člen, patřící mezi kovalentní interakce, vyjadřuje energii torzních úhlů a počítá se pomocí rovnice 4,

$$E_{torsions} = \sum_t \sum_n \frac{E_n}{2} [1 + \cos(n\theta - \theta_{n0})] \quad (4)$$

kde E_n popisuje výšku potenciální bariéry v n -té periodě, kde n je periodou jedné harmonické složky. θ je vazebná torze a θ_{n0} je fázový posun v n -té periodě.

V silových polích jsou nekovalentní interakce obvykle modelovány pomocí Coulombova potenciálu pro elektrostatické interakce a Lennard-Jonesovým potenciálem pro van der Waalovy síly. [22]

Rovnice 5 popisuje elektrostatický příspěvek energie. Elektrostatika patří mezi jedny z nejdůležitějších interakcí odehrávajících se v biomolekulách v roztoku.

$$E_C = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (5)$$

q_i a q_j jsou parciální náboje, ϵ_0 je permitivita vakua, r_{ij} je vzdálenost mezi atomy.

Finální člen rovnice a nekovalentních interakcí je vyjádřen v rovnici 6 a popisuje energetický příspěvek van der Waalových sil. Jedná se o přitažlivé a odpudivé síly, kdy při přiblížení dvou částic dochází k poklesu energie a následně k obrovskému vzrůstu energie.

$$E_{vdW} = \sum_{i < j} \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - \left(\frac{r_0}{r} \right)^6 \right] \quad (6)$$

E představuje hloubku minima křivky, r_0 je vzdálenost mezi dvěma částicemi, kde je minimální potenciál a r je vzdálenost dvou atomů.

V této práci se budu zabývat hlavně popisem solných můstků a jiných polárních interakcí mezi proteiny a DNA. Pro tyto interakce je důležitá hlavně rovnice 5, vyjadřující energii elektrostatické interakce mezi opačně nabitými rezidui, jako jsou např. fosfáty a argininy nebo lysiny. Důležitá je ale i rovnice 6, která určuje, na jakou nejbližší vzdálenost se mohou přiblížit opačně nabitě molekuly a spoluurčuje tak energii interakce v rovnovážné vzdálenosti. Je důležité poznamenat, že na stabilitu solných můstků a jiných polárních interakcí má vliv také přítomnost rozpouštědla (vody). Voda

totiž hydratuje ionty v nevázaném (volném) stavu a soutěží tedy s vazebnou energií mezi ionty. Pokud je součet hydratačních energií zúčastněných iontů větší než jejich interakční energie, nebude jejich solný můstek v roztoku stabilní. Interakce mezi ionty a solvatační energie jsou v nukleových kyselinách a proteinech v delikátní rovnováze a spoluurčují stabilitu všech nemolekulových komplexů.

2.6 Molekulová dynamika

Molekulová dynamika (MD) je počítačová simulace, díky níž jsme schopni předpovídat a popisovat pohyb atomů v určitém systému během časového úseku. Energie systému je určena silovým polem. Molekulovou dynamiku lze aplikovat pro studování velkých systémů, jako jsou biologické makromolekulární látky, kam se řadí jak proteiny, tak nukleové kyseliny. V molekulové dynamice se atomy považují za malé kuličky a jejich dynamika se počítá pomocí Newtonových pohybových zákonů. [23] Pohyb částice popisujeme pomocí rovnice 7.

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \quad (7)$$

F představuje sílu, která působí na částici ve směru pohybu, m_i je hmotnost částice, x_i je souřadnice, ve které se částice vyskytuje, t je čas. Sílu F můžeme vypočítat jako derivaci energie systému (popsaného silovým polem) podle souřadnic. S pomocí integrace rovnice 7 můžeme předpovídat budoucí polohu částice za určený časový interval (integrační krok). Soubor geometrií pro jednotlivé integrační kroky se nazývá trajektorie.

3. CÍLE PRÁCE

Cílem mé bakalářské práce je zmapování polárních interakcí v komplexech protein-DNA, kde se věnuji hlavně interakcím mezi záporně nabitými fosfátovými skupinami šroubovice DNA s kladně nabitými argininy a lysiny v proteinovém řetězci, které formují tzv. solné můstky. Na základě MD simulací se tedy snažím popsat, zda jsou vybrané interakce v těchto komplexech stabilní, částečně stabilní, nestabilní či dochází k alternaci mezi vazebnými partnery.

4. EXPERIMENTÁLNÍ ČÁST

4.1 Metody

Pro provádění simulace jsem si vybral dva zástupce B-DNA a jednoho zástupce Z-DNA z proteinové databanky www.rcsb.org. Všechny tři protein-DNA komplexy, které jsem vybral, lze v této databázi nalézt pod čtyřmístným kódem. Kritériem volby komplexů bylo vysoké rozlišení struktury do 2,5 Å.

První komplex nese označení 1SKN a má rozlišení 2,50 Å. Vycházel jsem z krystalové struktury komplexu, která byla zjištěna rentgenovou difrakční metodou při pH 6. Ve struktuře se kromě DNA a proteinu vyskytovala také molekula LDA neboli lauryl diemethylamin oxid. LDA je detergent, který je nezbytný pro krystalizaci daného komplexu. Pro mou prováděnou simulaci nebyl nijak významný, proto jsem jej ze struktury komplexu pomocí odstranil. Následně jsem náboj páteře zneutralizoval přidáním draselných iontů K^+ a poté jsem přidal ještě další K^+ a Cl^- ionty tak, aby prostředí v okolí komplexu mělo výslednou koncentraci 0,15M KCl, což odpovídá přibližně prostředí buňky. Všechny přidané ionty jsou parametrizovány dle Jounga a Cheatham. [24]

Druhým komplexem, kterým jsem se zabýval, je 3EYI, který má rozlišení 1,45 Å. Opět jsem vycházel z krystalové struktury, která byla rovněž zjištěna rentgenovou difrakční metodou ovšem při vyšším pH, jež se rovnalo hodnotě 8,5. Pro neutralizaci páteře jsem tentokrát místo draselných iontů použil ionty sodné (Na^+), protože v přítomnosti těchto iontů byl připraven monokrystal. Pro zajištění koncentrace prostředí 0,15 NaCl jsem následně přidal další Na^+ ionty a také Cl^- ionty.

Poslední komplex je 1MNN s rozlišením 1,40 Å. Stejně jako u předešlých komplexů jsem vycházel z krystalové struktury tohoto komplexu stanovené rentgenovou difrakční metodou, tentokrát při neutrálním pH 7. U této struktury bylo potřeba určit protonizační stavy histidinů obsažených v proteinu. K tomu jsem nejdříve využil internetový server PropKa [25], který předpovídá protonizační stav histidinu podle pH a struktury v jeho okolí. Následně jsem pro ověření správnosti strukturu zkontroloval vizuálně pomocí programu PyMOL [26], kdy jsem kontroloval přítomnost vodíkových vazeb u histidinů, které by mohly protonizaci ovlivnit. Kontrola hraje v tomto případě důležitou roli, jelikož PropKa není schopna pracovat s exitující sítí vodíkových vazeb uvnitř molekuly. Nakonec jsem po kontrole nahradil osm HIS forem histidinu za šest forem HIE a dvě

formy HID. U těchto forem se vodík vyskytuje navázaný na dusík v poloze ϵ nebo δ . [25]

Simulace jsem prováděl za izobaricko-izotermických podmínek (NPT), kdy jsem teplotu systému nastavil na 298 K (25 °C). Nastavený tlak odpovídal 1 Atm (101,325 kPa). Časový krok simulace odpovídal 4 fs [27], což bylo umožněno pomocí procedury HMR (hydrogen mass repartitioning) [28], kdy dochází k přesunutí části hmotnosti těžkých atomů na vodíkové atomy. Cutoff nevazebných interakcí odpovídal 10 Å a jednotlivé souřadnice atomů byly ukládány každých 10 ps. K solvataci komplexů jsem využil oktaedrální box s vodním modelem SPC/E [29]. Před začátkem produkční simulace byl systém i velikost boxu minimalizován a následně ekvilibrován s postupně se snižujícími velikostmi restrainů. Relaxační protokol byl převzat z literatury [30].

Pro realizaci simulací jsem použil programový balíček AMBER (Assisted Model Building with Energy Refinement) [31], konkrétně to byl kód PMEMD. Jedná se o programový balíček, jež umožňuje molekulovou dynamiku biomolekulárních systémů, kterými jsou např. nukleové kyseliny či proteiny. [32] Simulace jsem prováděl v jednom silovém poli, které se doporučuje pro simulaci DNA struktur, toto pole nese označení OL15 [33]. Pro analýzu trajektorií jsem využil program, spadající do již zmíněného balíčku AMBER, který se nazývá CPPTRAJ. Pro další analýzu a vizualizaci jsem využil program VMD (Visual Molecular Dynamics) [34].

Pomocí histogramů vzdáleností jsem si určil průměrnou vzdálenost minima histogramu, který odděluje oblast přímého kontaktu (4,3 Å) od oblasti vodou mediováného kontaktu (5,4 Å) u lysinu. Vzdálenost, při které interakce zaniká, odpovídala vzdálenosti za druhým píkem histogramu (>5,4 Å). U argininu bylo tyto hodnoty těžší určit, jelikož se překrývaly oblasti přímých a vodou mediováných kontaktů. Proto jsem určil jen přibližnou hodnotu, která určuje, jestli se arginin nachází v asociované či disociované formě, kdy v asociované formě je zahrnut přímý i vodou mediovaný kontakt (<6 Å).

Další důležitou částí mé práce bylo zvolit, mezi kterými atomy budu kontakty a vzdálenosti měřit. V programu LEAP, který patří do programového balíčku AMBER, jsem si našel pro jednotlivé molekuly, u kterých budu interakce měřit, z jakých atomů se skládají a jak jsou v programu označeny. Pro páteř dvoušroubovice jsem zvolil atom fosforu, který je v LEAPu označen jako P. Pro molekulu lysinu jsem zvolil atom dusíku, který je na konci aminokyselinové řetězce, tento atom dusíku je označen jako NZ.

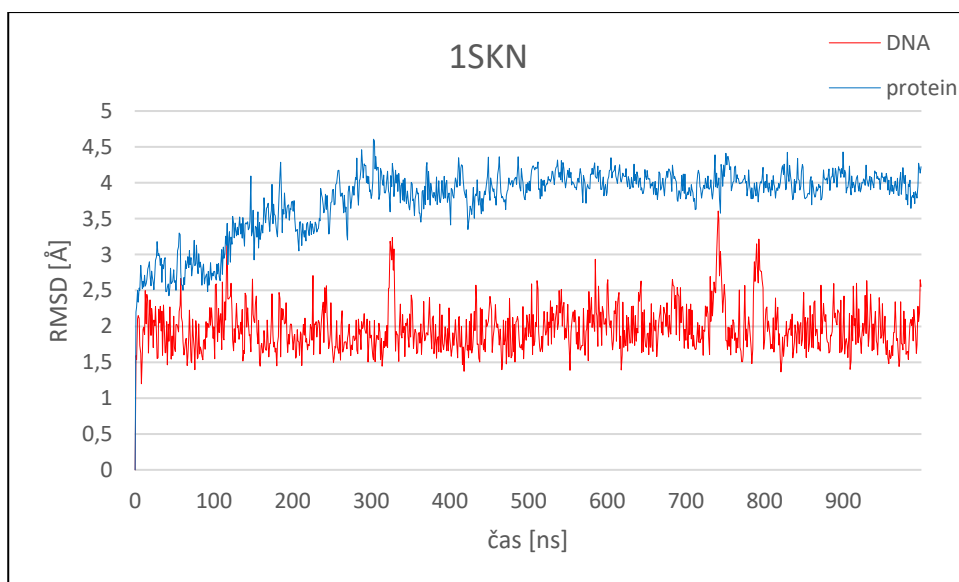
Pro arginin jsem vybral atom uhlíku, na kterém jsou navázány aminoskupiny, a jeho označení je CZ.

4.2 Výsledky a diskuse

4.2.1 Stabilita studovaných komplexů v MD simulacích

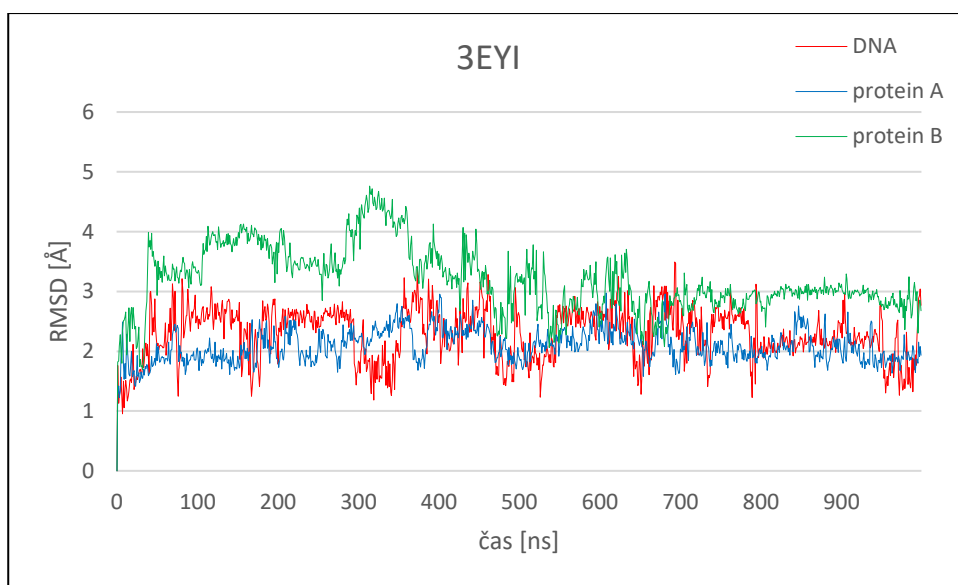
Pomocí RMSD (z angl. root mean square deviation) jsem porovnal struktury komplexu v průběhu simulace s výchozí krystalovou strukturou. Čím jsou hodnoty RMSD nižší, tím je komplex či molekula stabilnější, tzn., že při nulovém RMSD se struktura oproti krystalové struktuře neliší. RMSD jsem měřil pro jednotlivé složky komplexu a výsledky jsem zaznamenal do jednoho grafu pro každý komplex, aby šlo porovnat, jak moc jsou jednotlivé složky stabilní.

Na obrázku 9 můžeme vidět graf s hodnotami RMSD nukleové kyseliny a proteinu v simulacích u komplexu 1SKN. DNA je při simulacích výrazně stabilnější oproti proteinu a po většinu času se drží při hodnotách okolo 2 Å, až na několik kratších odchylek. Protein se naopak drží při hodnotách okolo 4 Å, kdy těchto hodnot začíná dosahovat přibližně v čase 250 ns. Tyto hodnoty jsou způsobeny volným C-koncem proteinu, který je ve výchozí krystalové struktuře v kontaktu se sousední molekulou. V simulaci původní krystalový kontakt neexistuje, takže konec má větší tendenci se v prostoru pohybovat, čímž ovlivňují hodnoty RMSD. 1SKN je navíc poměrně velký komplex, což také může způsobovat vyšší hodnoty RMSD.



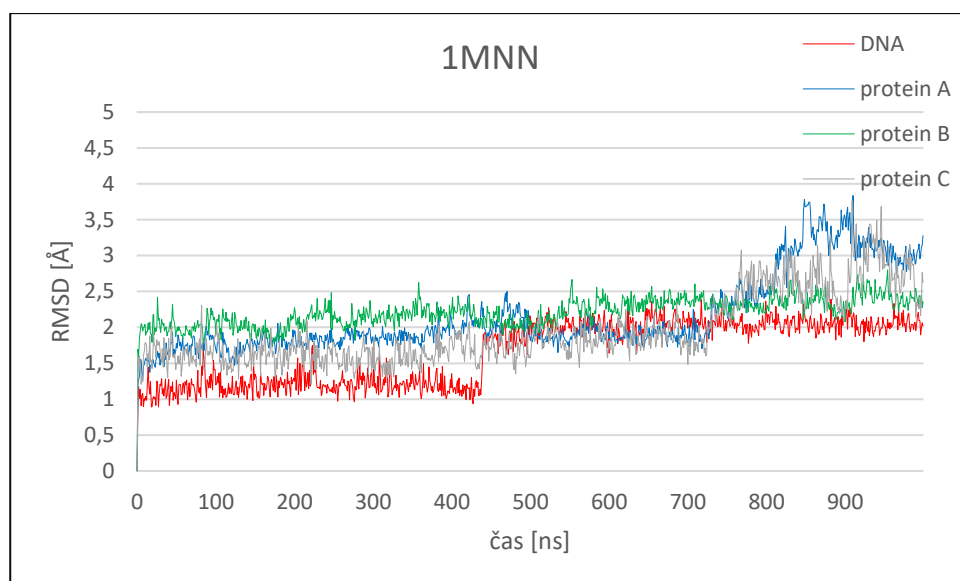
Obrázek 9: RMSD nukleové kyseliny a proteinu komplexu 1SKN

Další obrázek (Obrázek 10) popisuje hodnoty RMSD u komplexu 3EYI. Na obrázku jsou hodnoty RMSD pro DNA a dva proteinové řetězce, které jsou identické, ale v případě RMSD se poněkud liší. Proteinový řetězec B je totiž v průběhu simulací do protein-DNA interakcí zapojen méně, a tudíž se volně pohybuje během simulace v roztoku, což se odráží i na jeho RMSD. V průběhu prvních 400 ns se jeho RMSD pohybuje u 4 Å, poté se ale ustálí a klesne k hodnotě zhruba 3 Å, kde je už poměrně nízká hodnota. DNA a proteinový řetězec A se pohybují v rozmezí hodnot 1,5 - 3 Å, takže jsou v průběhu simulace stabilní.



Obrázek 10: RMSD nukleové kyseliny a proteinových řetězců komplexu 3EYI

Hodnoty RMSD pro poslední komplex 1MNN jsou popsány na obrázku 11, kde jsou znázorněny hodnoty pro nukleovou kyselinu a jeden proteinový řetězec, který na dvou místech obsahuje neurčená rezidua, takže je ve výchozí struktuře rozdělen na tři části. Tento komplex je ze všech mnou zvolených po celou dobu simulací nejstabilnější. Menší skok lze pozorovat přibližně v čase 750 ns, kde dochází k menšímu vzrůstu hodnot RMSD u proteinového řetězce A a C, tyto hodnoty ovšem nepřekračují 4 Å, takže nejde o nijak výraznou změnu. Po celou dobu simulace se hodnoty RMSD drží v rozmezí 1 - 2,5 Å, takže je komplex a jeho složky velmi stabilní. U řetězce DNA dochází přibližně v 340 ns k výraznějšímu zvýšení RMSD, které je způsobeno oddělením koncové nespárované báze od dvoušroubovice a vytvořením nového kontaktu s proteinem. Protože k tomu dochází na konci DNA, nemá to významnější vliv na zkoumané interakce.



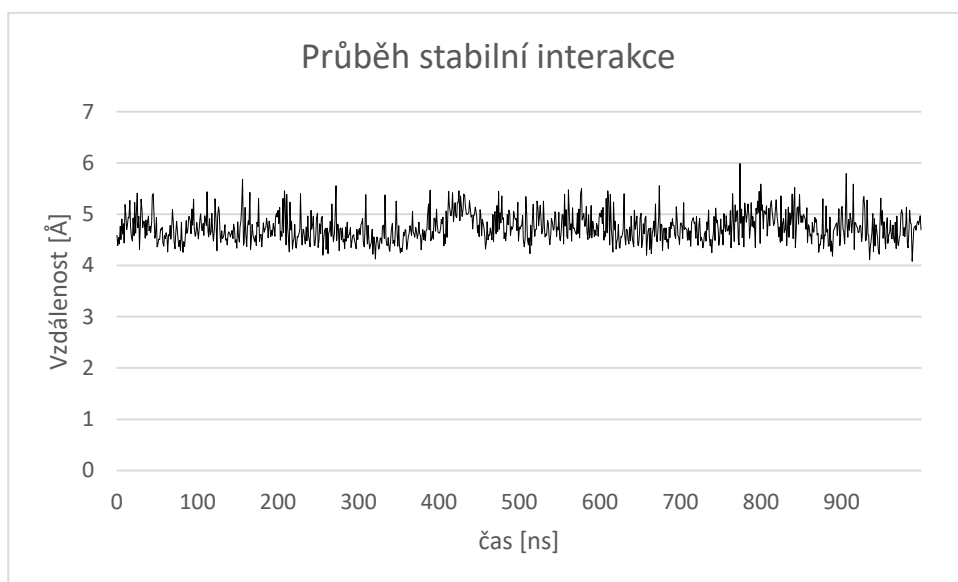
Obrázek 11: RMSD nukleové kyseliny a proteinových řetězců 1MNN

4.2.2 Protein-DNA interakce

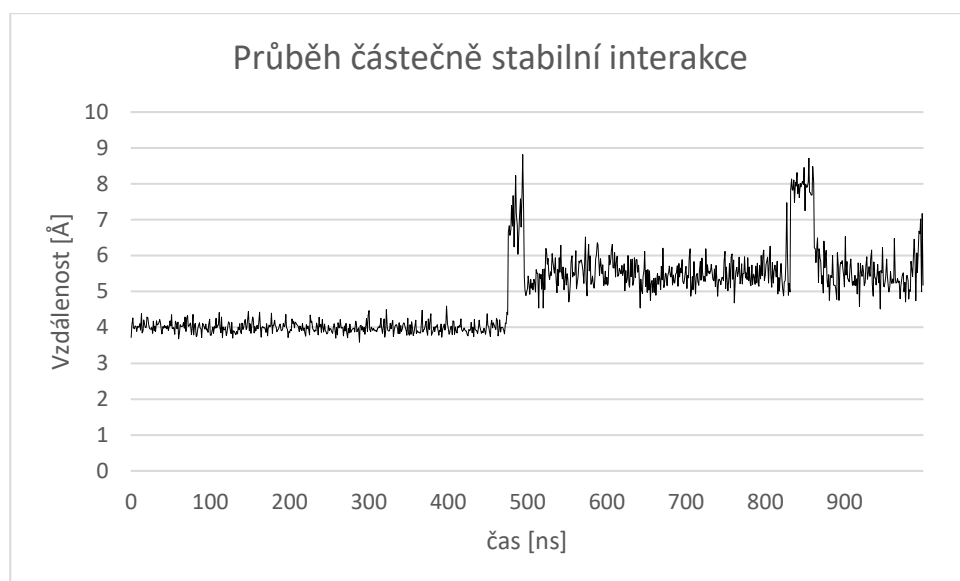
Cílem této práce bylo zmapovat protein-DNA interakce se zaměřením na interakce polární a zejména interakce nabitých reziduí proteinu s páteřními fosfátovými skupinami. U nalezených interakcí jsem zjišťoval, jestli jsou v MD průběhu simulace stabilní, částečně stabilní, nestabilní nebo dochází k alternaci vybraného rezidua mezi více vazebnými partnery. Pro nalezení interakcí jsem vycházel ze článků pro jednotlivé komplexy a také z programu VMD, kde jsem tyto interakce vizuálně vyhledával. Největší pozornost jsem věnoval interakcím, které byly zprostředkovány prostřednictvím elektrostatických sil, hlavně těmi mezi fosfátovou skupinou na páteři nukleové kyseliny, která má záporný náboj a kladně nabitými postranními řetězci proteinu, které jsou tvořeny argininem a lysinem. Vazba mezi těmito skupinami se nazývá solný můstek. Kritérium pro kontakt lysinu s jinou skupinou jsem zvolil tak, že vzdálenost mezi centrálními atomy byla do 4,3 Å, dále vzdálenost v rozmezí 4,3 - 5,4 Å odpovídala vodou mediovanému kontaktu a po překročení hranice 5,4 Å jsem určil, že se již nejedná o kontakt. U argininu bylo toto kritérium složitější určit kvůli většímu množství možností, jak může k interakci s guanidinovou skupinou docházet, což ztěžuje určení jednoznačných hraničních vzdáleností. Proto jsem určil kritérium pro jeho kontakt s jinou skupinou do vzdálenosti 6 Å, v níž je zahrnut i vodou mediovaný kontakt. Po překročení této vzdálenosti se již o kontakt nejedná.

Jak jsem již zmínil výše, v mnou prováděných simulacích jsem se setkal se čtyřmi možnými scénáři vývoje interakcí v čase. Prvním příkladem je stabilní interakce,

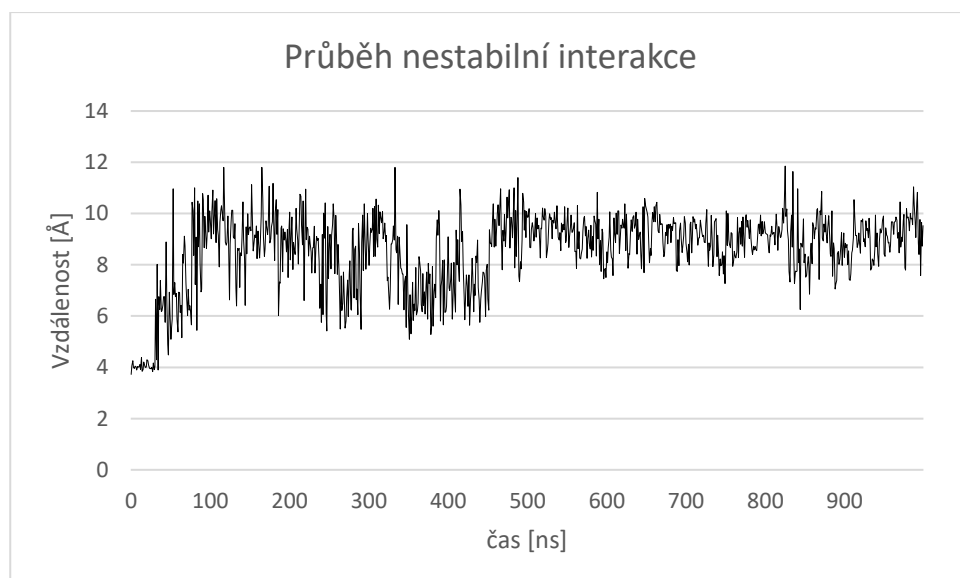
pro kterou je průběh mezimolekulové vzdálenosti ukázán na obrázku 12. Tuto interakci lze považovat za stabilní, jelikož nedochází k výrazné změně vzdáleností mezi centrálními atomy, u kterých dochází ke vzniku kontaktu. Druhým případem je interakce, která je částečně stabilní (Obrázek 13). Tato interakce je po určitý časový úsek stabilní, poté ale dochází ke zvětšení vzdálenosti mezi atomy a interakce se stává nestabilní (mezi 400 a 500 ns). Na dalším obrázku (Obrázek 14) je znázorněn třetí scénář interakce, jímž je nestabilní interakce, kdy dochází v průběhu začátku simulace k zániku původního kontaktu, který se již do původní formy nevrátí. Čtvrtý průběh interakce popisuje obrázek 15, kdy dochází k alternaci mezi původní a nově vzniklou interakcí. U tohoto typu interakce jsem také vypočítal, jaké procento času je interakce v původním kontaktu a novém kontaktu.



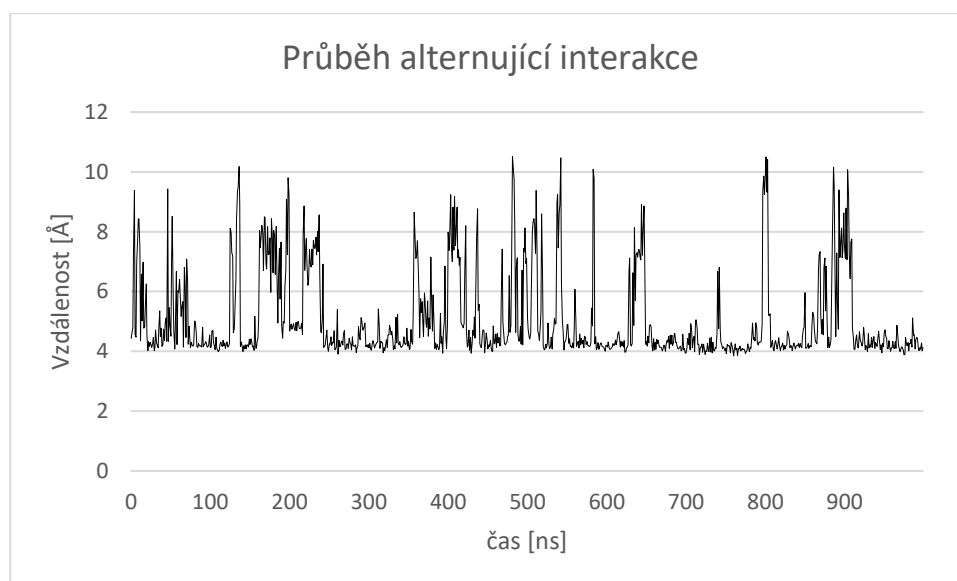
Obrázek 12: Průběh stabilní interakce



Obrázek 13: Průběh částečně stabilní interakce



Obrázek 14: Průběh nestabilní interakce



Obrázek 15: Průběh alternující interakce

4.2.3 Interakce v 1SKN

V protein-DNA komplexu s názvem 1SKN jsem sledoval celkem 21 interakcí. Největší podíl mají solné můstky, které tvoří 14 interakcí. Z těchto 14 interakcí jsou čtyři stabilní a nemění vazebného partnera. Ty jsou tvořeny pouze aminokyselinovým zbytkem argininu. Další čtyři solné můstky mění v průběhu simulací fosfátovou skupinu za jinou, se kterou tvoří nový solný můstek a zároveň jeden z těchto solných můstků přechází do disociovaného stavu. Vodou mediované solné můstky se zde nachází čtyři, kdy dva přecházejí z vodou mediovaného kontaktu na přímý kontakt a zbylé dva přechází do disociovaného stavu. Ze zbývajících dvou solných můstků jeden vytvoří vodíkovou vazbu s guaninovou bází a druhý alternuje mezi původní strukturou a disociovaného. Dále se v komplexu vyskytují interakce zprostředkované záporně nabitými aminokyselinami jako je kyselina asparagová, která z původní interakce s argininem vytváří v průběhu simulace kontakt s argininem novým a zároveň i s lysinem. Další iontovou interakci tvoří kyselina glutamová, která na začátku interaguje s argininem přes můstkovou vodu, ovšem v průběhu simulace dojde k přímému kontaktu. Dále se můžeme setkat se situací, kdy se kladně nabitá aminokyselina váže pomocí vodíkové vazby na dusíkatou bázi. V případě tohoto komplexu se váže arginin vodíkovou vazbou na guaninovou bázi a po celý průběh simulace je tato interakce stabilní. Další interakce s dusíkatými bázemi mohou tvořit i neutrální aminokyseliny, které se pomocí vodíkových vazeb také zapojují do protein-DNA interakcí. Tento případ se zde vyskytuje dvakrát, kdy se stejná

molekula asparaginu váže na dvě různé báze (thymín a cytosin) a po celou dobu je tento kontakt stabilní. Další vodíkové vazby mohou neutrální aminokyseliny tvořit s fosfátovou skupinou nebo s kladně nabitým reziduem. Konkrétně u 1SKN dochází k interakci serinu s lysinem, který v průběhu simulace vytvoří kromě přímého kontaktu i kontakt vodou mediováný. S fosfátovou skupinou interaguje threonin přes můstkovou vodu a poté přechází do disociovaného stavu.

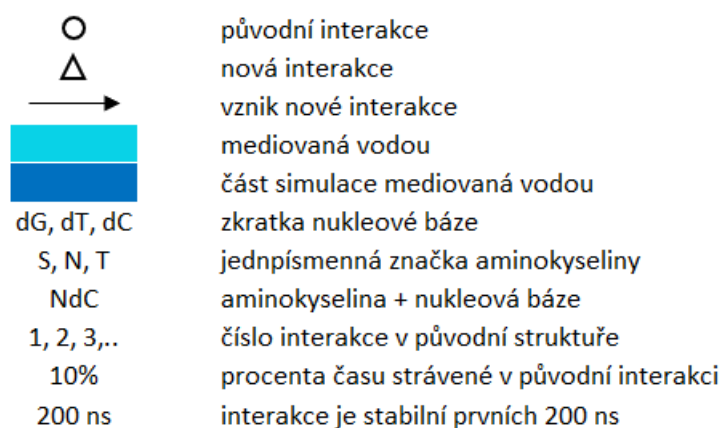
Popsané interakce a jejich změny jsou sumarizovány v tabulce 1. Na obrázku 16 je zobrazena legenda mapy pro lepší orientaci ve 2D mapách. Obrázek 17 představuje souhrn interakcí v komplexu 1SKN ve formě 2D mapy, do které jsem zaznamenal všechny interakce, včetně interakcí zmíněných v původní literatuře.

Tabulka 1: Souhrn interakcí v komplexu I SKN

Pořadí	Původní interakce		Nová interakce		Průběh
	interagující rezidua ^{a)}	typ int. ^{b)}	interagující rezidua	typ int.	
1	R457 B_dT13	sm	R457 B_dT12	sm	pohyb mezi dvěma fosfáty
2	R499 E496	vv	R499 E496	vv	alternace
3	K460 A_dC10	sm	K460 B_dT13	sm	pohyb mezi dvěma fosfáty
4	R516 A_dG8	sm	R516 A_G8	solvatace	alternace
5	K512 A_dT9	sm	K512 A_T9	sm	alternace
6	D461 R508	sm	D461 K512	sm	částečně stabilní
	R508 D461	sm	R508 A_T9	sm	částečně stabilní
7	R508 A_dC10	sm			stabilní
8	R507 A_dC10	sm	R507 B_dG5	vv	nestabilní
9	K460 B_dT13	sm	K460 B_dT13	solvatace	alternace
10	R503 B_dG4	sm			stabilní
11	R506 B_dG4	sm			nestabilní
12	R506 B_dG3	sm			stabilní
13	K510 B_dG5	sm	K510 B_dG5	sm	alternace
14	K510 S478	vv	K510 S478	vv	alternace
15	T517 B_dA6	vv	T517 voda	solvatace	alternace
16	R521 B_dT7	sm	R96 B_dA6	sm	pohyb mezi dvěma fosfáty
17	R522 B_dG8	sm			stabilní
18	R525 B_dT7	sm	R525 voda	solvatace	alternace
19	R519 A_dG8	vv			stabilní
20	N511 A_dC10	vv			stabilní
21	N511 B_dT7	vv			stabilní

^{a)} Rezidua aminokyselin jsou popsána jejich jednopísmennou značkou a pořadovým číslem v řetězci. Rezidua nukleové kyseliny jsou popsána jednopísmennou zkratkou řetězce a poté dvoupísmennou zkratkou a číslem, kde se na prvním místě nachází malé d, což je zkratka pro deoxy, po něm následuje zkratka báze a pořadí rezidua v řetězci.

^{b)} typ interakce; sm – solný můstek, vv – vodíková vazba



Obrázek 16: Legenda 2D map

	fosfát (PO ₄ ⁻)	lysín (K ⁺)	arginin (R ⁺)	kys. asparagová (D)	kys. glutamová (E)	protein	báze	voda
PO ₄ ⁻								
K ⁺								
R ⁺								
D ⁻								
E ⁻								
prot.								
báze								
voda								

Obrázek 17: 2D mapa komplexu 1SKN

4.2.4 Interakce v 3EYI

V komplexu s PDB označením 3EYI jsem sledoval celkem 13 interakcí. Solný můstek je v těchto 13 interakcích zastoupen jedenáctkrát. Žádná z těchto interakcí není stabilní. Fosfátovou skupinu při těchto interakcích vymění arginin i lysin vždy v jednom případě. Vodou mediované solné můstky převažují nad přímými kontakty aminokyselin

s fosfátem. Dohromady se zde nachází pět vodou mediováných kontaktů, které v průběhu simulace mění fosfátovou skupinu. Dva tyto kontakty vytváří lysin, jenž má na začátku vodou mediovaný kontakt s fosfátem, ale v průběh simulace vytvoří přímý kontakt s jinou fosfátovou skupinou. Jeden z těchto lysinů dokonce přechází do disociovaného stavu. Zbylé tři vodou mediované kontakty tvoří arginin, který ve dvou případech přechází z vodou mediovaného na přímý kontakt a jeden přechází z přímého kontaktu na vodou mediovaný. Dále se zde vyskytují tři interakce, kde kladně nabitá rezidua nemění fosfátovou skupinu, ale dochází ke změně zprostředkovatele kontaktu. Lysin je u těchto interakcí zastoupen dvakrát, kdy jednou tvoří na začátku přímý kontakt a přechází na vodou mediovaný, ve druhém případě je to naopak. Arginin je v tomto případě vodou mediovaný na začátku a přechází na přímý kontakt. Poslední solný můstek v tomto komplexu připadá argininu, který v průběhu simulace tvoří vodíkovou vazbu se serinem. Dále se zde vyskytují dvě interakce, které tvoří neutrálně nabitá aminokyselina s nukleovými bázemi, mezi nimiž se formuje vodíková vazba, jež je v průběhu simulace stabilní. Konkrétně se jedná o asparagin, který interaguje s cytosinovou bází a guaninovou bází.

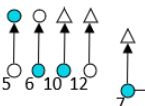
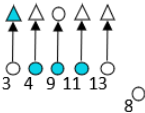
Všechny zmíněné interakce jsou shrnuty v tabulce 2 a na obrázku 18 podobně jako u komplexu 1SKN.

Tabulka 2: Souhrn interakcí v komplexu 3EYI

Pořadí	Původní interakce		Nová interakce		Průběh
	interagující rezidua ^{a)}	typ int. ^{b)}	interagující rezidua	typ int.	
1	A_N141 C_dC3	vv			stabilní
2	A_N141 C_dG4	vv			stabilní
3	A_R142 C_dC5	sm	A_R142 A_D139	sm	pohyb mezi dvěma fosfáty
4	A_R142 C_dG6	sm	A_R142 C_dG4	sm	pohyb mezi dvěma fosfáty
5	A_K138 C_dG4	sm	A_K138 C_dG4	sm	alternace
6	A_K138 D_dG6	sm	A_K138 D_dG6	sm	alternace
7	A_K160 C_dC3	sm	A_K160 C_dG2	sm	pohyb mezi dvěma fosfáty
	A_K160 C_dC3	sm	A_K160 voda	solvatace	alternace
8	A_R124 C_dG2	sm	A_R124 A_S159	vv	střídání kontaktů
9	B_R124 D_dG2	sm	B_R124 D_dG2	sm	alternace
10	B_K160 D_dC1	sm	B_K160 D_dC3	sm	pohyb mezi dvěma fosfáty
11	B_R142 D_dC5	sm	B_R142 D_dG4	sm	pohyb mezi dvěma fosfáty
12	B_K138 C_dG6	sm	B_K138 D_dG4	sm	pohyb mezi dvěma fosfáty
13	A_R146 A_D143	sm	A_R146 C_dC5	sm	pohyb mezi dvěma fosfáty

^{a)} Jelikož se v krystalové struktuře komplexu 3EYI vyskytovaly dva proteiny, jsou aminokyseliny odlišeny pomocí jednopísmenné zkratky podle toho, ke kterému proteinu patří. Rezidua nukleové kyseliny jsou popsána jednopísmennou zkratkou řetězce a poté dvoupísmennou zkratkou a číslem, kde se na prvním místě nachází malé d, což je zkratka pro deoxy, po něm následuje zkratka báze a pořadí rezidua v řetězci.

b) typ interakce; sm – solný můstek, vv – vodíková vazba

	fosfát (PO_4^-)	lysin (K^+)	arginin (R^+)	kys. asparagová (D^-)	kys. glutamová (E^-)	protein	báze	voda
PO_4^-								
K^+								
R^+								
D^-								
E^-								
prot.								
báze								
voda								

Obrázek 18: 2D mapa komplexu 3EYI

4.2.5 Interakce v 1MNN

V mnou posledním zkoumaném komplexu pod PDB označením 1MNN jsem pozoroval celkem 15 interakcí. Většina těchto interakcí je opět zastoupena solným můstkem, který se zde vyskytuje v 11 případech. Tři z těchto 11 interakcí jsou stabilní a nedochází u nich ke změně vazebného partnera. Dva zástupce těchto interakcí má arginin a jednoho lysin. Ke změně fosfátu dochází v případě komplexu 1MNN pouze jednou, kdy arginin z vodou mediovaneho kontaktu s jedním fosfátem vytváří přímý kontakt s novým fosfátem a zároveň i vodou mediovaný kontakt s dalším fosfátem. Arginin dále tvoří dva vodou zprostředkované solné můstky, kdy jeden přechází do přímého kontaktu a druhý zaniká a arginin místo něj tvoří vodíkovou vazbu se serinem. Kromě solných můstků je arginin také vázaný vodíkovou vazbou k dusíkatým bázím, dvakrát je vázán ke guaninové bázi a jednou k thyminové. Arginin vázaný v této formě je po celou dobu simulací stabilní. U lysinu dochází ve třech případech ke změně přímého kontaktu na vodou mediovaný a ve dvou případech z vodou zprostředkovaného

na přímý. Dále u jednoho solného můstku tvořeného lysinem dochází k disociaci lysinu. V jednom případě dochází dokonce ke vzniku solného můstku, kdy lysin přechází z vodou mediováného kontaktu s guaninovou bází k interakci s fosfátem, kdy dochází k přímému tak vodou mediovanému kontaktu.

Všechny interakce jsem opět shrnul v tabulce 3 a jsou také popsány na obrázku 19.

Tabulka 3: Souhrn interakcí v komplexu IMNN

Pořadí	Původní interakce		Nová interakce		Průběh
	interagující rezidua ^{a)}	typ int. ^{b)}	interagující rezidua	typ int.	
1	R326 B_dG4	sm	R326 S259	vv	částečně stabilní
2	R326 B_dC3	sm	R326 B_dC3	sm	alternace
3	K110 B_dC3	sm			stabilní
4	R111 C_dG10	vv			stabilní
5	R177 C_dG8	vv			stabilní
6	R58 C_dT6	vv			stabilní
7	K212 B_dC14	sm	K212 B_dC14	sm	alternace
8	R208 C_dT5	sm	R208 C_dT6	sm	pohyb mezi dvěma fosfáty
	R208 C_dT5	sm	R208 B_dC14	sm	pohyb mezi dvěma fosfáty
9	K176 C_T5	sm	K176 voda	solvatace	alternace
10	K176 C_T6	sm	K176 C_T6	sm	alternace
11	R254 C_dT7	sm			stabilní
12	R97 C_dG8	sm			stabilní
13	K50 C_dT9	sm	K50 C_dT9	sm	alternace
14	K54 C_dG10	sm	K54 C_dG10	sm	alternace
15	K264 B_dG2	vv	K264 B_dG2	sm	nestabilní

^{a)} Rezidua aminokyselin jsou popsána jejich jednopísmennou značkou a pořadovým číslem v řetězci. Rezidua nukleové kyseliny jsou popsána jednopísmennou zkratkou řetězce a poté dvoupísmennou zkratkou a číslem, kde se na prvním místě nachází malé d, což je zkratka pro deoxy, po něm následuje zkratka báze a pořadí rezidua v řetězci.

^{b)} typ interakce; sm – solný můstek, vv – vodíková vazba

	fosfát (PO_4^-)	lysin (K^+)	arginin (R^+)	kys. asparagová (D^-)	kys. glutamová (E^-)	protein	báze	voda
PO_4^-								
K^+								
R^+								
D^-								
E^-								
prot.								
báze								
voda								

Obrázek 19: 2D mapa komplexu 1MNN

4.2.6 Souhrnné poznatky

V mnou zkoumaných komplexech jsem se celkem zaměřil na 49 interakcí. Největší část těchto interakcí tvořily solné můstky, kterých bylo 36, tj. 73 %. Z těchto 36 solných můstků měl největší zastoupení arginin, jež v původním krystalu tvořil 21 těchto kontaktů, což je 58 %. Lysin tvořil zbývajících 15, tj. 42 %. Zachování těchto interakcí nastalo pouze v sedmi případech, kdy došlo k plnému zachování u argininu šestkrát (30 %) a u lysinu pouze jednou (5 %). Tam, kde k zachování nedošlo, docházelo nejvíce k pohybu kladně nabitých reziduí mezi blízkými fosfátovými skupinami. Takových interakcí jsem v případě argininu pozoroval sedm, tj. 33 %, a u lysinu čtyři, tj. 19 %, kde navíc ve dvou případech docházelo k přechodu lysinu do roztoku.

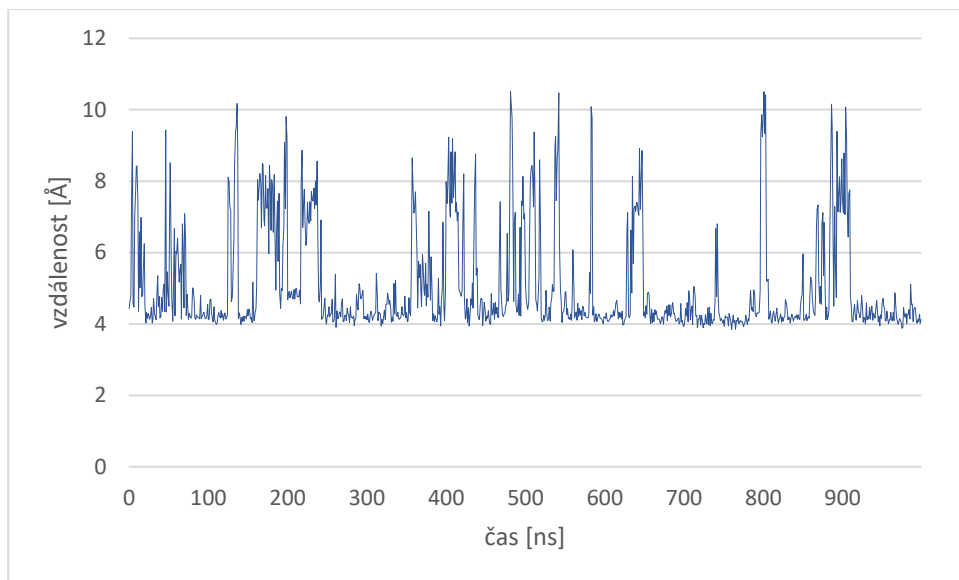
V komplexech jsem také pozoroval velké množství solných můstků, ve kterých nebyly ionty v přímém kontaktu, ale interakce byla zprostředkována molekulou vody nacházející se mezi kationtem a aniontem. Tuto interakci budu nadále označovat jako vodní můstek nebo vodou mediováný kontakt. Celkem jsem sledoval 15 kontaktů, které byly vodou mediované v původním krystalu, což tvoří 42 % všech solných můstků. Lysin a arginin mají téměř rovnocenné zastoupení při těchto interakcích. U argininu se

setkávám se sedmi interakcemi a u lysinu s osmi. To znamená, že ve všech argininem tvořených solných můstcích je přesně třetina vodou mediovaná, tj. 33 %. Z 15 solných můstků tvořených lysinem je 53 % vodou mediováných. Dále při interakcích docházelo ke změně zprostředkování vodou, kdy spolu původní rezidua tvořila solný můstek, ale buď mezi ně vstoupila voda, nebo byl vodou mediováný můstek nahrazen přímým kontaktem. Solných můstků, mezi které vstoupila voda, jsem pozoroval celkem pět, čtyři v případě lysinu a jediný u argininu. Solných můstků, které byly spojeny vodním můstkem a přešly na přímý kontakt, bylo šest. Lysin byl opět zastoupen čtyřmi interakcemi, arginin pouze dvěma. Důležitost vodou mediováných interakcí je v souladu s dřívějšími pozorováními v literatuře. [35]

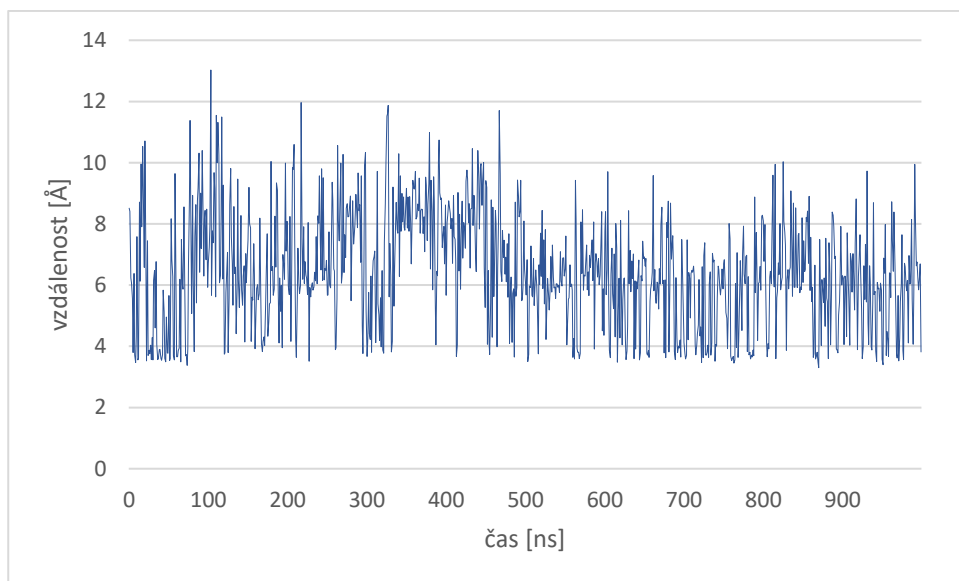
U argininu jsem dále pozoroval, že po zániku solného můstku vytvořil vodíkovou vazbu s nukleovou bází v jednom případě, ve druhém případě pak vytvořil vodíkovou vazbu s neutrální aminokyselinou. Lysiny ze solných můstků podobné interakce nevytvářely.

V případě několika solných můstků docházelo k tomu, že kladně nabitá aminokyselina ztrácela interakci s fosfátovou skupinou a přecházela do roztoku, ze kterého ale byla schopna se k původnímu fosfátu vrátit a na čas původní interakci obnovit. Tyto situace jsou zajímavé tím, že dochází k postupnému ustavování rovnováhy mezi dvěma stavy (vázaný kation vs. hydratovaný kation). Pokud by bylo dosaženo rovnováhy, můžeme z poměru časů strávených v simulaci v jednom a druhém stavu určit rovnovážnou konstantu, která tento děj charakterizuje. Proto se jimi budu zabývat podrobněji. V následujících grafech popisují jejich chování. Na obrázku 20 je znázorněn graf, který popisuje chování argininu R516, který na začátku simulace tvořil solný můstek s guaninem A_G8 a v průběhu simulace přecházel do roztoku a docházelo k alternaci mezi těmito dvěma stavy. Obrázek 21 popisuje alternaci kontaktu lysinu K460 s thyminem B_T13, jenž byl část simulačního času mediován vodou a přechodem do roztoku. Graf vyobrazený na obrázku 22 znázorňuje chování argininu R525, jenž v původním krystalu tvořil vodou mediováný solný můstek s thyminem B_T7 a v průběhu simulace alternoval mezi původním kontaktem a přechodem do roztoku. Všechny výše zmíněné situace se vztahují ke komplexu 1SKN. U komplexu 3EYI graficky popisují jedinou alternaci, která je zobrazena na obrázku 23. Ta probíhá v případě lysinu A_K160, který v původním krystalu tvoří vodou zprostředkovaný solný můstek s cytosinem C_C3 a v průběhu simulace přechází do disociovaného stavu a dochází k ustavování rovnováhy mezi vazebnými partnery. V komplexu 1MNN

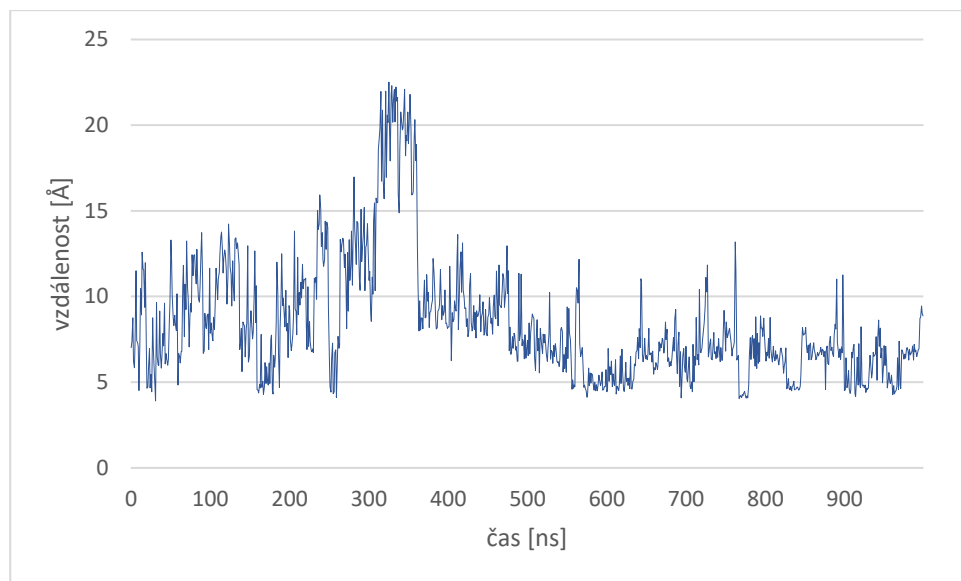
rovněž popisují jedinou interakci (Obrázek 24). V tomto případě lysin K176 formuje vodou mediovaný solný můstek s thyminem C_T5 a v průběhu simulace alternuje mezi původním kontaktem a disociovanou formou.



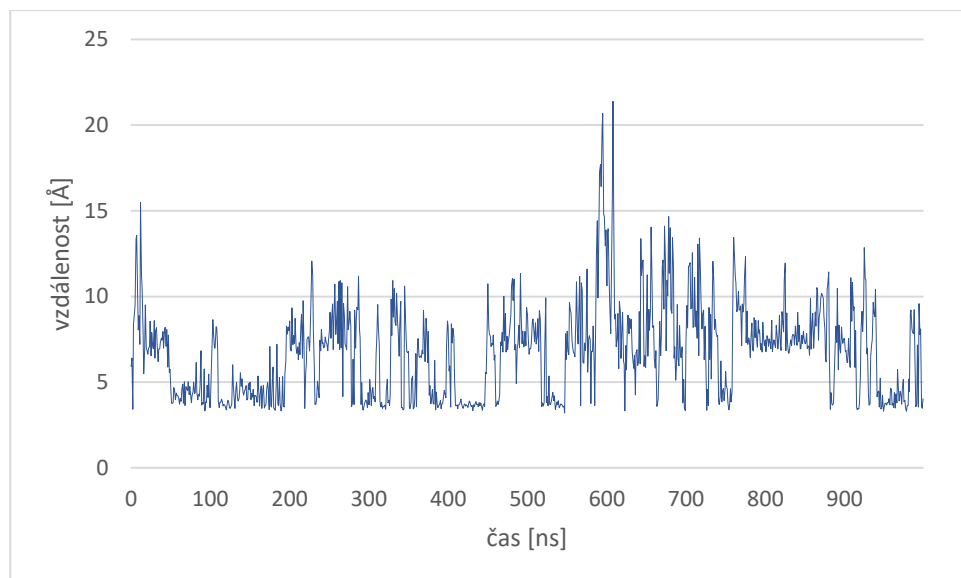
Obrázek 20: Alternace R516 - A_G8/voda



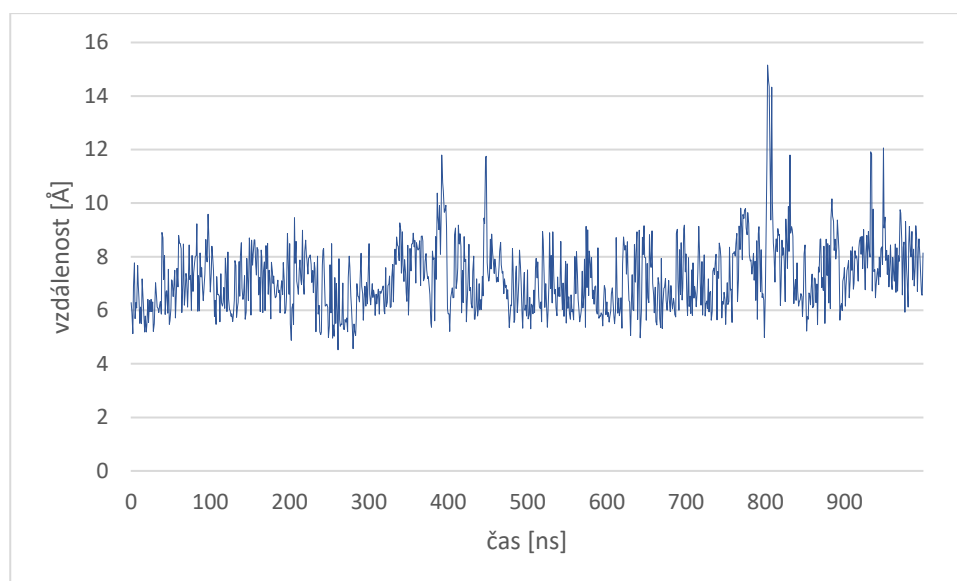
Obrázek 21: Alternace K460 - B_T13/voda



Obrázek 22: Alternace R525 - B_T7/voda



Obrázek 23: Alternace A_K160 - C_C3/voda



Obrázek 24: Alternace K176 - C_T5/voda

Kladně nabitá rezidua netvořila solné můstky pouze s fosfátovými skupinami. V mnou zkoumaných komplexech jsem také pozoroval solné můstky, jež se objevovaly mezi záporně nabitou aminokyselinou a kladně nabitou kyselinou. Jednalo se o dva případy, kdy arginin, jakožto kladně nabitá aminokyselina, tvořil solný můstek s kyselinou glutamovou a také s kyselinou asparagovou. Při poslední zmíněné interakci navíc došlo k tomu, že se v průběhu simulace kontakt argininu s kyselinou přerušil a místo něj se vytvořil solný můstek s fosfátovou skupinou dvoušroubovice DNA a zároveň s kyselinou asparagovou vytvořil solný můstek lysin.

Arginin a lysin se dále v několika případech vázaly vodíkovou vazbou k dusíkatým bázím či neutrálním aminokyselinám. Arginin byl ve čtyřech případech spojen vodíkovou vazbou s bázemi guaninu (3) a thyminu (1), ve kterých byl stabilní po celou dobu. Lysin byl takto vázán v jediném případě, a navíc skrze můstkovou vodu. Tato interakce v průběhu simulace zanikla a došlo ke vzniku solného můstku mezi lysinem a fosfátem nukleové kyseliny, jež byl střídavě v přímém i vodou mediovaném kontaktu. Dále se lysin vázal pomocí vodíkové vazby k neutrální aminokyselině, při simulaci však došlo k tomu, že mezi rezidua vstoupila voda a kontakt byl tudíž tvořen přes vodní můstek.

Následující interakce jsem sledoval, jelikož byly popsány v původní literatuře u jednotlivých komplexů. Jedná se o interakce, kdy neutrální aminokyseliny tvořily vodíkovou vazbu s dusíkatými bázemi nebo s fosfátem. Vodíkovou vazbou byl k bázím vázán asparagin, což se ukázalo jako stabilní interakce po celou dobu interakce.

K fosfátu se vodíkovou vazbou vázal threonin, jež v průběhu simulace oddisocioval do roztoku.

Všech 49 interakcí, které jsem sledoval a zmínil v této podkapitole, jsem shrnul do tabulky 4.

Tabulka 4: Souhrn všech interakcí

solný můstek	Arginin	Lysin	Asparagin	Threonin
celkem	21	15	-	-
stabilní	6	1	-	-
změna fosfátu ^{a)}	7	4	-	-
původní vm ^{b)}	7	8	-	-
nový vm	1	4	-	-
báze, AK ^{c)} , roztok	3	0	-	-
ostatní interakce				
solný můstek s AK	2	0	0	0
vv ^{d)} s báží	4	1	4	0
vv s AK	0	1	0	0
vv s fosfátem	0	0	0	1

^{a)} Zahrnuty i interakce, které byly vodou mediované v původním kontaktu (arginin - 3, lysin - 2)

^{b)} vm - vodou mediovaný kontakt

^{c)} AK - aminokyselina

^{d)} vv - vodíková vazba

5. ZÁVĚR

Proteiny a DNA patří mezi nejdůležitější látky pro všechny živé organismy. Společně vytvářejí komplexy, které ovlivňují mnoho pochodů v buňkách a velkou zásluhu na tom mají jejich vzájemné interakce. Pro svou bakalářskou práci jsem si vybral tři protein-DNA struktury s PDB označením 1SKN, 3EYI a 1MNN.

Vybrané komplexy jsem studoval pomocí molekulově dynamických simulací v silovém poli OL15. V každém komplexu jsem sledoval stabilitu původních kontaktů v průběhu MD simulací. Zaměřil jsem se převážně na solné můstky, u kterých jsem sledoval, jak se mění vzdálenost mezi centrálními atomy vazebných partnerů, podle čehož jsem dále vyhodnotil povahu interakce.

Nejprve jsem si u každého komplexu zjistil, jak jsou jejich jednotlivé složky v průběhu MD simulací stabilní. V případě všech komplexů bylo RMSD po většinu času simulace přijatelné velikosti. Odchyly byly většinou způsobené volnými konci struktur, které fluktovaly v roztoku. Jediné větší odchýlení jsem pozoroval v případě proteinu u komplexu 1SKN, které bylo způsobeno C-koncem, který v simulaci postrádá svůj původní krystalový kontakt, takže má větší tendenci se v prostoru pohybovat, čímž ovlivňuje hodnoty RMSD.

Ve své práci jsem pozoroval celkem 49 interakcí, jejichž stabilitu a změny jsem zobrazil v podobě 2D interakčních map. Z těchto map lze rychle vyčíst, zda je interakce stabilní, a pokud ne, jaký je její osud v průběhu MD simulace, tedy např. jaká nová interakce se místo ní vytváří.

Největší část ze 49 sledovaných interakcí představovaly solné můstky, které byly tvořeny více argininem než lysinem. Jen část z nich byla v průběhu interakce stabilní - arginin se jevil jako stabilnější, jelikož udržel daleko více původních kontaktů, než lysin (29 % vs. 5 %).

V situacích, kdy nedocházelo k zachování původních interakcí, jsem se dále nejčastěji setkával s pohybem kladně nabitých reziduí mezi blízkými fosfáty. Arginin byl v těchto interakcích opět zastoupen více, ale procentuálně se výskyt těchto interakcí v komplexech přibližně rovnal a objevoval se zhruba v 30 % procentech případů. To, že kladně nabitá rezidua proteinu netvořila specifické interakce s konkrétním fosfátem, ale naopak alternovala mezi blízkými fosfáty, je v souladu s předpokladem, že jsou tyto interakce nespecifické.

Rezidua, která nebyla v přímém kontaktu, se často objevovala v interakcích, které byly zprostředkovány pomocí vodního můstku. Vodou zprostředkované kontakty se častěji vyskytovaly u lysinu, který měl v původním krystalu i v průběhu simulací větší zastoupení než arginin. Lysin tedy ochotněji a častěji tvořil vodou mediované kontakty. Arginin u vodou mediovaných kontaktů často přecházel k jiné fosfátové skupině. Sledoval jsem však i případy, kdy se arginin raději pomocí vodíkové vazby navázal na bázi, neutrální aminokyselinu či přešel do roztoku. Dále je důležité zmínit, že vodou mediované kontakty byly v několika případech přerušeny a alternovaly mezi původním kontaktem a interakcí s roztokem (solvatací).

Další interakce, které jsem pozoroval, byly zprostředkovány vodíkovou vazbou. Vodíkovou vazbu jsem nejčastěji sledoval mezi dusíkatými bázemi DNA a aminokyselinovými zbytky proteinu. V menším zastoupení se také vodíková vazba formovala mezi aminokyselinami v proteinu. Těchto interakcí se v proteinu vyskytuje samozřejmě obrovské množství, jelikož formují strukturu proteinu. Ve své práci se však věnuji jen těm interakcím, které ovlivňují interakce proteinu s DNA.

Dva z mých studovaných komplexů byly transkripční faktory, které jsou typické výskytem sekvenčně specifických interakcí, kdy sekvenčně specifická aminokyselina proteinu reaguje s konkrétní dusíkatou bází. Tyto interakce byly po celou dobu simulací stabilní, což je zajímavé srovnat s výše popsány nespecifickými interakcemi, které často mění způsob vazby (přímý vs. vodou mediovaný), střídají nejbližší interagující fosfáty nebo zanikají na úkor plné solvatace interagujících partnerů. Výsledky mých simulací tedy naznačují, že sekvenčně specifické interakce jsou energeticky výhodné a stabilní, kdežto nespecifické interakce jsou neselektivní (alternace mezi dvěma sousedními fosfáty), vykazují energeticky blízké alternativní stavy (přímé vs. vodou mediované kontakty) a některé z nich jsou na hraně stability, protože alternují s disociovaným, plně hydratovaným stavem.

Výsledky popsané v této práci ukazují, že nespecifické mezimolekulové kontakty pozorované v krystalových geometriích mohou být v roztoku nestabilní a v rovnováze s jinými způsoby vazby. V případě navýšení počtu simulovaných komplexů, by mohlo dojít na základě statistické analýzy stability a rovnováh interakcí k odhadnutí rovnovážné interakční konstanty a porovnání s konstantami získanými z experimentu. Takové porovnání by mohlo ukázat, které interakce jsou v empirickém potenciálu popsány správně a které potřebují další upřesnění. Získané výsledky by mohly vést ke zpřesnění teoretického modelování interakcí mezi proteiny a nukleovými kyselinami.

6. SUMMARY

Proteins and DNA are among the most important substances for all living organisms. Together, they form complexes that influence many processes in cells, and their interactions with each other are of great importance. For my bachelor thesis I have chosen three protein-DNA structures with PDB designations 1SKN, 3EYI and 1MNN.

I studied the selected complexes using molecular dynamics simulations in the OL15 force field. In each complex, I monitored the stability of the original contacts during the MD simulations. I focused mainly on salt bridges, where I observed how the distance between the central atoms of the bonding partners changed, by which I further evaluated the nature of the interaction.

First, for each complex, I noted how stable their individual components were over the course of the MD simulations. For all complexes, the RMSD was of an acceptable magnitude for most of the simulation time. Deviations were mostly due to the free ends of the structures fluctuating in solution. The only major deviation I observed in the case of the protein for complex 1SKN was due to the C-terminus, which lacks its original crystal contact in the simulation, so it has a greater tendency to move in space, thus affecting the RMSD values.

In my work, I observed a total of 49 interactions, whose stability and changes were plotted as 2D interaction maps. From these maps one can quickly tell whether an interaction is stable, and if not, what its fate is during the MD simulation, e.g. what new interaction is created instead.

The largest fraction of the 49 observed interactions were salt bridges, which were formed more by arginine than by lysine. Only a fraction of these were stable during the interaction - arginine appeared to be more stable as it retained far more of the original contacts than lysine (29% vs. 5%).

Further, in situations where there was no maintenance of the original interactions, I most often encountered movement of positively charged residues between nearby phosphates. Arginine was again more abundant in these interactions, but the percentages of these interactions were approximately equal in the complexes, occurring in about 30% of the cases. The fact that the positively charged residues of the protein did not form specific interactions with a particular phosphate but instead alternated between nearby phosphates is consistent with the assumption that these interactions are nonspecific.

Residues that were not in direct contact often appeared in interactions that were mediated by a water bridge. Water-mediated contacts were more frequent for lysine, which was more abundant in the original crystal and during the simulations, than arginine. Thus, lysine more readily and frequently formed water-mediated contacts. Arginine at water-mediated contacts often switched to a different phosphate group. However, I have also observed cases where arginine preferred to bind to a base, a neutral amino acid via hydrogen bonding, or moved into solution. It is also important to note that water-mediated contacts were broken in several cases, alternating between the original contact and the interaction with the solution (solvation).

Other interactions that I observed were mediated by hydrogen bonding. The hydrogen bonding I observed most often was between the nitrogenous bases of DNA and amino acid residues of the protein. To a lesser extent, hydrogen bonding also formed between amino acids in the protein. These interactions are of course abundant in the protein, as they shape the structure of the protein. In my work, however, I focus only on those interactions that affect the protein-DNA interactions.

Two of the complexes I studied were transcription factors, which are characterized by the occurrence of sequence-specific interactions, where a sequence-specific amino acid of the protein reacts with a particular nitrogenous base. These interactions were stable throughout the simulations, which is interesting to compare with the nonspecific interactions described above, which often change the binding mode (direct vs. water-mediated), alternate the closest interacting phosphates, or disappear at the expense of full solvation of the interacting partners. Thus, the results of my simulations suggest that sequence-specific interactions are energetically favorable and stable, whereas non-specific interactions are non-selective (alternation between two adjacent phosphates), exhibit energetically close alternative states (direct vs. water-mediated contacts), and some are on the edge of stability because they alternate with a dissociated, fully hydrated state.

The results described in this work show that the non-specific intermolecular contacts observed in crystal geometries can be unstable in solution and in equilibrium with other modes of binding. If the number of simulated complexes is increased, a statistical analysis of the stability and equilibria of the interactions could be used to estimate the equilibrium interaction constant and compare it with those obtained from experiment. Such a comparison could show which interactions are correctly described in the empirical potential and which need further refinement. The results

obtained could lead to a refinement of the theoretical modelling of protein-nucleic acid interactions.

7. LITERATURA

- [1] DAHM, R. Friedrich Miescher and the discovery of DNA. *Developmental Biology*. 2005, **278**(2), 274-288.
- [2] WATSON, J. D. a F. H. C. CRICK. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*. 1953, **171**(4356), 737-738.
- [3] SAFAEE, N., A. M. NORONHA, D. RODIONOV, G. KOZLOV, C. J. WILDS, G. M. SHELDRIK a K. GEHRING. Structure of the Parallel Duplex of Poly(A) RNA: Evaluation of a 50 Year-Old Prediction. *Angewandte Chemie International Edition*. 2013, **52**(39), 10370-10373.
- [4] HARDINSON, R. 2.5: B-Form, A-Form, and Z-Form of DNA. *Biology LibreTexts*. [cit. 2022-04-10]. Dostupné z: [https://bio.libretexts.org/Bookshelves/Genetics/Book%3A_Working_with_Molecular_Genetics_\(Hardison\)/Unit_I%3A_Genes_Nucleic_Acids_Genomes_and_Chromosomes/2%3A_Structures_of_Nucleic_Acids/2.5%3A_B-Form_A-Form_and_Z-Form_of_DNA](https://bio.libretexts.org/Bookshelves/Genetics/Book%3A_Working_with_Molecular_Genetics_(Hardison)/Unit_I%3A_Genes_Nucleic_Acids_Genomes_and_Chromosomes/2%3A_Structures_of_Nucleic_Acids/2.5%3A_B-Form_A-Form_and_Z-Form_of_DNA)
- [5] GHOSH, A. a M. BANSAL. A glossary of DNA structures from A to Z. *Acta Crystallographica Section D Biological Crystallography*. 2003, **59**(4), 620-626.
- [6] POTAMAN, V. N. a R. R. SINDEN. DNA: Alternative Conformations and Biology: Vladimir N. Potaman and Richard R. Sinden. *National Library of Medicine*. [cit. 2022-04-10]. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK6545/>
- [7] ANANBEH, H. a H. KUPCOVÁ SKALNÍKOVÁ. Ze života bílkovin aneb Na tvaru záleží. 282-285. [cit. 2022-04-10]. Dostupné z: <https://ziva.avcr.cz/files/ziva/pdf/ze-zivota-bilkovin-aneb-na-tvaru-zalezi.pdf>
- [8] SUN, P. D., C. E. FOSTER a J. C. BOYINGTON. Overview of Protein Structural and Functional Folds. *Current Protocols in Protein Science*. 2004, **35**(1).
- [9] Orders of protein structure. *Khan Academy*. [cit. 2022-04-10]. Dostupné z: <https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/a/orders-of-protein-structure>
- [10] BLANCO, A. a G. BLANCO. Proteins. *Medical Biochemistry*. Elsevier, 2017, 2017, 21-71. ISBN 9780128035504.
- [11] Proteiny. *Univerzita Palackého v Olomouci: Přírodovědecká fakulta*. [cit. 2022-04-10]. Dostupné z: https://www.prf.upol.cz/fileadmin/userdata/PrF/katedry/kbb/Dokumenty/Materialy_k_vyuce/BB1P/BB1-3PR_2018.pdf

- [12] LUSCOMBE, N. M., S. E. AUSTIN, H. M. BERMAN a J. M. THORNTON. An overview of the structures of protein-DNA complexes. *Genome Biology*. 1(1).
- [13] ZHENG, Z. a Y. WANG. DNA binding proteins: outline of functional classification. *BioMolecular Concepts*. 2011, 2(4), 293-303.
- [14] STOJASPAL, M. *Kvantitativní popis interakcí DNA vazebných proteinů*. [cit. 2022-04-10]. Dostupné z: https://is.muni.cz/th/lcbp5/Bc._prace_-_skryte_vysledky_.pdf. Bakalářská práce. Masarykova univerzita, Přírodovědecká fakulta.
- [15] PANDEY, P., S. HASNAIN a S. AHMAD. *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, 2019. ISBN 9780128114322.
- [16] HARTEIS, S. a S. SCHNEIDER. Making the Bend: DNA Tertiary Structure and Protein-DNA Interactions. *International Journal of Molecular Sciences*. 2014, 15(7), 12335-12363.
- [17] ROHS, R., X. JIN, S. M. WEST, R. JOSHI, B. HONIG a R. S. MANN. Origins of Specificity in Protein-DNA Recognition. *Annual Review of Biochemistry*. 2010, 79(1), 233-269.
- [18] MARCOVITZ, A. a Y. LEVY. Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proceedings of the National Academy of Sciences*. 2011, 108(44), 17957-17962.
- [19] RUPERT, P. B., G. W. DAUGHDRILL, B. BOWERMAN a B. W. MATTHEWS. A new DNA-binding motif in the Skn-1 binding domain-DNA complex. *Nature Structural Biology*. 1998, 5(6), 484-491.
- [20] HA, S. C., D. KIM, H.-Y. HWANG, A. RICH, Y.-G. KIM a K. K. KIM. The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proceedings of the National Academy of Sciences*. 2008, 105(52), 20671-20676.
- [21] LAMOUREUX, J. S. Structure of the sporulation-specific transcription factor Ndt80 bound to DNA. *The EMBO Journal*. 21(21), 5721-5732.
- [22] LEACH, A. R. *Molecular modelling: principles and applications*. 2nd ed. Harlow: Prentice Hall, 2001. ISBN 05-823-8210-6.
- [23] ROY, K., S. KAR a R. N. DAS. Computational Chemistry. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Elsevier, 2015, 2015, 151-189. ISBN 9780128015056.
- [24] JOUNG, I. S. a T. E. CHEATHAM. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B*. 2008, 112(30), 9020-9041.

- [25] URANGA, J., P. MIKULSKIS, S. GENHEDEN a U. RYDE. Can the protonation state of histidine residues be determined from molecular dynamics simulations?. *Computational and Theoretical Chemistry*. 2012, **1000**, 75-84.
- [26] GRELL, L., C. PARKIN, L. SLATEST a P. A. CRAIG. EZ-Viz, a tool for simplifying molecular viewing in PyMOL. *Biochemistry and Molecular Biology Education*. 2006, **34**(6), 402-407.
- [27] JUREČKA, P., M. ZGARBOVÁ, F. ČERNÝ a J. SALOMON. *Continuous B-to A- Transition in Protein-DNA Binding - How Well Is It Described by Current AMBER Force Fields?*. 2022, 1-8.
- [28] HOPKINS, C. W., S. LE GRAND, R. C. WALKER a A. E. ROITBERG. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*. 2015, **11**(4), 1864-1874.
- [29] 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.
- [30] BANÁŠ, P., D. HOLLAS, M. ZGARBOVÁ, P. JUREČKA, M. OROZCO, T. E. CHEATHAM, J. ŠPONER a M. OTYEPKA. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *Journal of Chemical Theory and Computation*. 2010, **6**(12), 3836-3849.
- [31] CASE, D. A., T. E. CHEATHAM, T. DARDEN, et al. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. 2005, **26**(16), 1668-1688.
- [32] ŠPONER, J. a F. LANKAŠ. *Computational studies of RNA and DNA.: Challenges and Advances in Computational Chemistry and Physics*. ISBN 978-1-4020-4851-3.
- [33] ZGARBOVÁ, M., J. ŠPONER, M. OTYEPKA, T. E. CHEATHAM, R. GALINDO-MURILLO a P. JUREČKA. 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*. 2015, **11**(12), 5723-5736.
- [34] HUMPHREY, W., A. DALKE a K. SCHULTEN. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996, **14**(1), 33-38.
- [35] SCHNEIDER, B., J. ČERNÝ, D. SVOZIL, P. ČECH, J.-C. GELLY a A. G. DE BREVERN. Bioinformatic analysis of the protein/DNA interface. *Nucleic Acids Research*. 2014, **42**(5), 3381-3394.