

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra fyzikální chemie



**Návrh struktur vybraných makromolekul  
metodami výpočetní chemie**

Magisterská práce

Autor:	Jan Pauswang
Vedoucí magisterské práce:	Mgr. Petra Kůhrová, Ph.D.
Studijní program:	Chemie
Studijní obor:	Materiálová chemie
Forma studia:	Prezenční

Olomouc 2018

Prohlašuji, že jsem předloženou magisterskou práci vypracoval samostatně pod vedením Mgr. Petry Kührové, Ph.D. a že jsem použil zdrojů, které cituji a uvádím v seznamu použitých pramenů.

V Olomouci dne.....

.....  
vlastnoruční podpis

## Poděkování

Chtěl bych poděkovat vedoucí mé magisterské práce Mgr. Petře Kührové, Ph.D. za odborné vedení, cenné rady, pomoc a trpělivost při řešení problémů souvisejících s magisterskou prací. Dále bych chtěl poděkovat doc. Mgr. Pavlovi Banášovi, Ph.D. a Mgr. Michalu Janečkovi za podmínky, připomínky a rady v souvislosti s magisterskou prací.

Děkuji mojí rodině za poskytnuté zázemí pro vytváření magisterské práce. A také oceňuji, že mě upozornili na některé jazykové gramatické jevy související s textem magisterské práce.

## **Bibliografická identifikace:**

Jméno a příjmení autora	Jan Pauswang
Název práce	Návrh struktur vybraných makromolekul metodami výpočetní chemie
Typ práce	Magisterská práce
Pracoviště	Katedra fyzikální chemie
Školitelka magisterské práce	Mgr. Petra Kührová, Ph.D.
Rok obhajoby práce	2018
Abstrakt	<p>Práce se zabývá vytvořením modelů polymerázy cytomegaloviru. Při vytváření modelu byl pro návrh struktury použit program I-TASSER. Simulace modelů probíhaly pomocí výpočtů založených na principech molekulové mechaniky v programu Amber. Templáty, na jejichž základě byly pomocí programu I-TASSER vytvořeny modely polymerázy, odpovídaly kódům 2gv9, 3iay a 4flw v PDB databázi. Byly vytvořeny dva modely polymerázy cytomegaloviru obsahující DNA. Jeden model obsahoval DNA v exonukleázové formě. Druhý model obsahoval DNA v polymerázové formě.</p>
Klíčová slova	molekulová mechanika, molekulová dynamika, Amber, polymeráza, cytomegalovirus, model polymerázy
Počet stran	47
Jazyk	Český

**Bibliographical identification:**

Author's first name and surname	Jan Pauswang
Title	Design of macromolecular structures using computational methods
Type of thesis	Master thesis
Department	Department of Physical Chemistry
Supervisor	Mgr. Petra Kührová, Ph.D.
The year of presentation	2018
Abstract	In this work the generating of model of polymerase of cytomegalovirus was studied. I-TASSER program was used to prediction of polymerase structure. Simulation of models were performed with program Amber using molecular mechanism principles. Structure of models is based on concrete structures of herpesvirus polymerase (2gv9), yeast (3iay) and archaeobacteria (4flw). Models were prepared with DNA in two modes. First model was prepared with DNA in the polymerizing mode. Second model was prepared with DNA in editing mode.
Keywords	molecular mechanics, molecular dynamics, Amber, polymerase, cytomegalovirus, polymerase's model
Number of pages	47
Language	Czech

Vysoká škola: Univerzita Palackého v Olomouci      Fakulta: Přírodovědecká  
Katedra: Katedra fyzikální chemie      Školní rok: 2016/2017

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Jméno: Bc. **JAN PAUSWANG**

Studijní obor: Materiálová chemie

Název práce

česky: Návrh struktur vybraných makromolekul metodami výpočetní chemie  
anglicky: Design of macromolecular structures using computational methods

Cíl práce: Seznámení se s oblastí „soft-matter“ chemie a návržení struktury DNA polymerázy cytomegaloma viru nástroji výpočetní chemie.

### Zásady pro vypracování:

1. Vypracování literární rešerše
2. Prostudování uvedené odborné literatury
3. Bioinformatická studie
4. Návrh a výstavba systémů
5. Výpočet simulací
6. Interpretace a diskuze výsledků
7. Vypracování diplomové práce
8. Příprava a prezentace multimediální prezentace diplomové práce

Rozsah grafických prací: podle potřeby

Rozsah původní zprávy: 40-70 stran


- Seznam odborné literatury:
- původní časopisecká sdělení dle vlastní literární rešerše
  - Molecular Modeling and Simulation: An Interdisciplinary Guide, T. Schlick, 2010, Springer-Verlag New York
  - Innovations in Biomolecular Modeling and Simulations, T. Schlick, 2012, RCS Publishing

Vedoucí diplomové práce: Mgr. Petra Kührová, Ph.D.

Datum zadání diplomové práce: 25. 10. 2016

Termín odevzdání diplomové práce: duben 2018

  
.....  
podpis studenta

  
.....  
podpis školitele

  
.....  
podpis vedoucího katedry

## Obsah:

1.	Úvod	9
2.	Teoretická část	10
2.1	Výpočetní metody chemie a fyziky	10
2.1.1	Dynamické rovnice	10
2.1.2	Klasická mechanika	11
2.1.3	Molekulová mechanika	12
2.1.4	Energie určená silovým polem	12
2.1.5	Další důležité pojmy z výpočetní chemie	14
2.2	Struktura biomakromolekul	16
2.2.1	Proteiny	16
2.3	Cytomegalovirus	17
2.4	Polymeráza DNA	18
2.4.1	Polymeráza DNA - herpesviry	21
2.5	Použité programy, databáze a nástroje	23
3.	Výpočetní část	26
3.1	Nastavení výpočtů v programu Amber	26
3.2	Výchozí struktury použité pro výstavbu polymerázy cytomegaloviru	26
3.3	Příprava templátu pro výstavbu polymerázy cytomegaloviru v programu I-TASSER	29
3.4	Předpovězené struktury z programu I-TASSER	31
3.5	Optimalizace struktur polymerázy cytomegaloviru	33
3.6	Výsledné modely polymerázy cytomegaloviru, možné pokračování	35
4.	Závěr	38
5.	Summary	39
6.	Příloha	40
7.	Seznam použité literatury	45



## 1. Úvod

V dobách dávno minulých byli lidé decimováni kromě hladu a ozbrojených konfliktů i mnohými nemocemi. V dnešní době už umírá lidí na podvýživu výrazně méně. Zranění a nemoci způsobené nepřeborným množstvím důvodů a událostí je možné léčit. Medicína je opravdu velmi daleko. Lékaři dokáží transplantovat orgány, působit proti účinku potenciálních epidemií (chřipka, obezita). V některých případech jsou dříve obávaná onemocnění zcela nebo aspoň zčásti vymýcena. To vše je možné díky široké škále léčebných postupů a medikamentů. Naneštěstí ne všechny medikamenty lze opakovaně využívat. U některých léčiv byly postupně zjištěny neblahé vedlejší účinky. Na jiné léky si cílové organismy zvykly a vyvinula se u nich rezistence. To znamená, že nemoci způsobující některé bakterie a viry prošly evolučním vývojem, přizpůsobily se prostředí a staly se odolnějšími.

Cytomegalovirus je virus, jenž je přítomen u většiny světové populace. Jeho přítomnost v těle však zdraví lidé nepocítí. Zdraví nebezpečná infekce tímto virem je problémem hlavně pro imunodeficientní pacienty. Jejich oslabená imunita nedokáže dostatečně reagovat a nebezpečná infekce tímto virem může naplno propuknout. Infekce cytomegalovirem může vést např. k zánětům plic, trávicí soustavy, houbovým a bakteriálním infekcím. Naštěstí je možné některé virové infekce léčit pomocí antivirotik. K léčbě se používají zejména léky ganciclovir, foscarnet, cidofovir, vanganciclovir a acyclovir.[1, 2, 3, 4, 5] Některým pacientům, jenž jsou imunodeficientní (např. pacienti požívající v rámci potransplantační léčby imunosupresiva nebo pacienti trpící nemocí AIDS) a jsou postiženi infekcí tímto virem, se však léčba uvedenými léky může zkomplikovat. Vůči cytomegaloviru jsou totiž některé léky neúčinné. Cytomegalovirus je na ně rezistentní.[6, 7]

Cílem, na který antivirotika působí, je polymeráza DNA viru.[4,8] Antivirotika ji vyřadí z provozu a tím se přestane virus replikovat. Rezistence cytomegaloviru je způsobena takovými mutacemi v polymeráze viru, že antivirotika již vyřazení polymerázy DNA z provozu nezpůsobí. Pro studium interakce molekuly léčiva s polymerázou cytomegaloviru, je nezbytné znát strukturu polymerázy cytomegaloviru. Struktura polymerázy však není známa. Respektive známe jen strukturu posledních 22 aminokyselin z celkových 1242.[9] To však pro studium interakce s léčivem nestačí. Z tohoto důvodu byla snaha v této práci vystavět strukturální model polymerázy cytomegaloviru.

Při výstavbě daného modelu bylo postupováno tak, že se v databázi UniProt našla polymeráza s co nejvíce podobnou sekvencí aminokyselin. V našem případě to byla polymeráza herpes simplex viru-1. Následně byly na této struktuře pomocí serveru I-TASSER vystavěny strukturální modely samotné polymerázy cytomegaloviru. Poté, aby byly modely otestovány a ohodnoceny, byly použity jako vstupy do molekulové dynamiky. Výpočetní simulace molekulové dynamiky probíhaly na základě teorií klasické mechaniky v programu Amber. Dle výsledků molekulové dynamiky byly modely následně upravovány a hodnoceny.

## 2. Teoretická část

### 2.1 Výpočetní metody chemie a fyziky

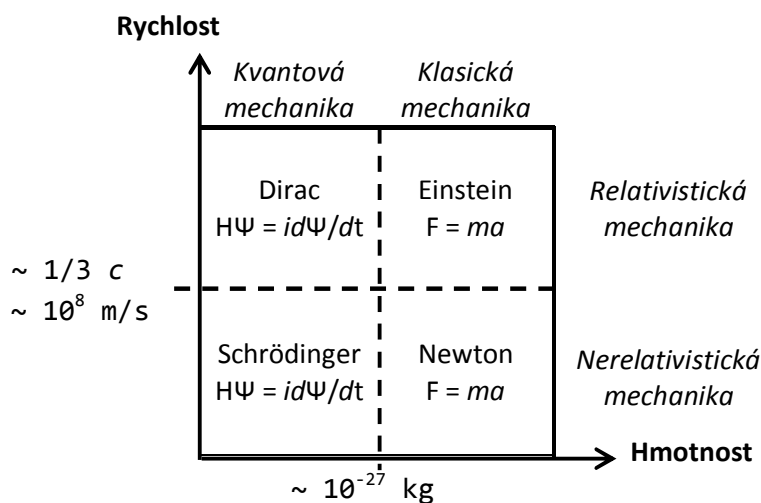
Výpočetní chemie je název pro vědní disciplínu, jež se vyskytuje na pomezí působnosti fyziky a chemie. Výpočetní chemie nám může pomoci zjistit, jaké jsou energie a vlastnosti různých konformací molekul. Dále jakým způsobem částice mezi sebou interagují a jak se struktura molekul a její vlastnosti mění v čase. Vývoj systému v čase popisuje dynamická rovnice.

Existují teorie, jejichž úkolem je popsat částice nebo látky. Tento popis je uskutečněn pomocí matematických rovnic. Při tzv. mnohočásticovém problému (systému) se rovnice popisující tento systém neřeší analyticky, ale numericky. Proto je výhodné pro výpočty použít výpočetní kapacity počítačů, jež jsou schopny provést milióny matematických operací za zlomek sekundy.

#### 2.1.1 Dynamické rovnice

Jak již bylo řečeno, vývoj systému v čase popisuje dynamická rovnice. Pro spuštění výpočtu simulace v čase je nezbytné vědět, jakou molekulou se chceme zabývat, jaká je počáteční rychlost, umístění této částice a jaká je matematická formulace sil, jež působí v a mezi částicemi. Dynamická rovnice je diferenciální rovnice, která obsahuje derivaci dle času a prostoru s přesným vyjádřením závislosti na hmotnosti a rychlosti částice.

Dynamická rovnice má různý tvar podle toho, jakou má uvažovaný systém hmotnost a rychlost (viz obr. 1). V práci byly využívány dynamické rovnice spadající do klasické nerelativistické mechaniky.



Obr. 1: Schéma znázorňuje různé dynamické rovnice podle toho, jakou má uvažovaný systém hmotnost a rychlost.  $F$  je síla,  $m$  hmotnost,  $a$  zrychlení,  $H$  Hamiltonián,  $\Psi$  vlnová funkce a  $t$  čas. Převzato.[10]

### 2.1.2 Klasická mechanika

Klasická mechanika je oblast fyziky, která se zabývá popisem pohybu hmotných těles s rychlostmi menšími, než je rychlost světla  $c$ . V klasické mechanice platí Newtonův druhý pohybový zákon

$$F = \frac{dp}{dt}, \quad 1)$$

kde  $F$  je síla,  $p$  je hybnost a  $t$  znamená čas. Zákon je možné přeformulovat za předpokladu, že hmotnost  $m$  je konstantní na tvar

$$F = \frac{dp}{dt} = m \frac{dv}{dt} = ma, \quad 2)$$

kde  $a$  znamená zrychlení. Jelikož víme, že derivace potenciálu  $V$  podle polohy  $r$  je síla

$$F(r) = -\frac{\partial V}{\partial r} \quad 3)$$

a druhá derivace polohy podle času je zrychlení

$$a = -\frac{d^2r}{dt^2}, \quad 4)$$

můžeme Newtonův druhý pohybový zákon přeformulovat do diferenciální podoby

$$-\frac{\partial V}{\partial r} = m \frac{\partial^2 r}{\partial t^2}. \quad 5)$$

Máme tedy dynamickou rovnici v diferenciální podobě, jež je funkcí času a jejímž řešením lze získat informace o poloze jakékoliv částice v čase.

Při uvažování částic s velice malou hmotností se místo Newtonovy druhé pohybové rovnice musí použít Schrödingerova časově závislá rovnice

$$H\Psi = i\frac{\partial\Psi}{\partial t}, \quad 6)$$

kde  $H$  je Hamiltonián. Druhou mocninou vlnové funkce, získané řešením této rovnice, se zjistí pravděpodobnost výskytu částice v určitém prostoru a čase. Hamiltonián skrývá mnoho členů, kdy ze všech je nejvíce obtížně popsatelné působení mezi elektrony navzájem. Výpočet elektronické energie je tedy nejvíce problematickou součástí Schrödingerovy rovnice.

Řešení diferenciální dynamické rovnice založené na Newtonově druhém pohybovém zákoně, při úvaze dvou částic s jednoduchým potenciálem mezi sebou, je analytické. Výsledkem je určitá matematická funkce  $r(t)$ . Ta samá rovnice při úvaze mnohočasticového systému se musí řešit numericky s využitím sledu malých časových kroků  $\Delta t$ . Tento časový krok je význačný parametr, který se používá v chemických výpočtech. Vždy by měl být minimálně o řád menší než je časová škála sledované vlastnosti systému nebo kritické (nejhlavnější) vlastnosti systému.[10]

### 2.1.3 Molekulová mechanika

Metody molekulové mechaniky neboli metody silového pole jsou metody výpočetní chemie, které obcházejí složitý výpočet elektronické energie pro dané rozložení jader takovým způsobem, že elektronická energie je pouze parametrickou funkcí daného uspořádání jader. Parametry se určují vhodným napasováním na experimentální data nebo na data z výpočtů o hodně vysoké úrovni kvality.

Dynamika atomů, molekul se v molekulové mechanice popisuje pomocí klasické mechaniky. Pro časově nezávislé případy je problém výpočtu zjednodušen na výpočet energie dané geometrie molekuly. Tímto způsobem můžeme zjistit, která z konformací molekuly je stabilnější.

Molekuly jsou v molekulové mechanice vystavěny ze stavebních bloků, ne z elektronů. Základní stavební blok (též nazývaný atomový typ) může být atom, ale i funkční skupina, aminokyselina nebo nukleotid. Každý atomový typ má své označení (např. číslo, písmenný symbol) a dané vlastnosti (hmotnost, velikost, náboj, síla vazby s konkrétním vazebným partnerem – vazebné informace). Vlastnosti atomových typů se liší podle toho, jaké silové pole použijeme. Platí, že jiné silové pole má jiné vlastnosti atomových typů.

Metody silového pole je výhodné použít pro výpočet na velkých systémech (např. biomakromolekulách). Jsou primárně určeny k předpovědi geometrií a relativních energií molekul. Nelze je použít k popisu vzniku a rozpadu vazeb nebo pro popis přesunu elektronů.

Podle toho co nás zajímá, volíme méně či více dokonalé silové pole. Důkladně sestavené silové pole je nezbytné pro zkoumání relativních energií různých konformací malých molekul. Zatímco hrubě sestavená silová pole se používají při studiu dynamiky proteinu sestaveného ze stovek aminokyselin.

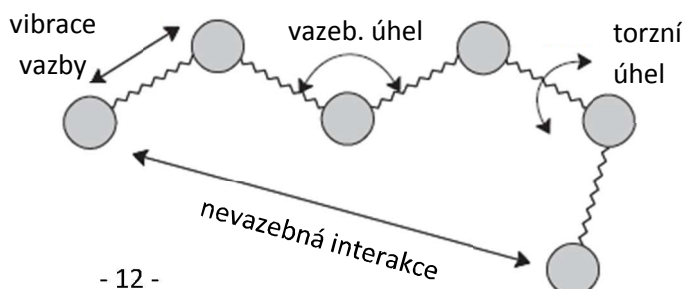
Například existuje silové pole MM2 (Molecular Mechanics versions 2) určené pro uhlovodíky. Rozlišuje 71 různými atomových typů. Tyto typy rozlišují atom uhlíku například v hybridizaci  $sp^3$  a  $sp^2$ . Atom uhlíku s hybridizací  $sp^2$  je ve více různých atomových typech lišících se povahou systému, ve kterém je tento atom přítomen. Např. je rozdíl mezi  $sp^2$  atomem uhlíku v cyklobutenu, aromátech, cyklopropenu a alkenech.

V molekulové mechanice jsou molekuly popisovány modelem „koulí a pružinek“. Koule představují atomy a pružiny představují vazby (viz obr. 2). Atomy se liší např. hmotností, vazby délkou a tuhostí pružiny.

### 2.1.4 Energie určená silovým polem

Energie daná určitým silovým polem je souhrnem matematických výrazů popisujících energii, která je potřebná pro deformaci molekuly určitým způsobem (viz obr. 2, rovnice 7).

*Obr. 2: Schéma znázorňující různé druhy silového působení v molekule. Znárodněná je vibrace vazby, vazebný úhel, torzní úhel a nevazebná interakce. Převzato.*



$$E_{sp} = E_{vib} + E_{úhel} + E_{tor} + E_{vdw} + E_{el} + E_3 \quad 7)$$

Energie molekuly popsaná silovým polem  $E_{sp}$  je dána součtem energie vibrace vazby  $E_{vib}$ , energie deformace vazebného úhlu  $E_{úhel}$ , energie deformace torzního úhlu  $E_{tor}$ , energie působení van der Waalsových sil  $E_{vdw}$ , energie elektrostatického působení  $E_{el}$  a energie  $E_3$ , která určitým způsobem popisuje spřažení  $E_{vib}$ ,  $E_{úhel}$  a  $E_{tor}$ . [10, 12, 13]

Stabilním molekulám, jež se na ploše potenciální energie nacházejí v minimu, odpovídá určité rozmístění atomů popsatelné prostorovými souřadnicemi jader. Minimalizací  $E_{sp}$  lze toto stabilní rozložení jader obdržet. Při výpočtech hovoříme o tzv. optimalizaci, tedy minimalizaci  $E_{sp}$  a změně geometrie jader.

$E_{vib}$  je energie vibrace vazby mezi dvěma atomovými typy A, B. Tuto energii lze základně popsat pomocí harmonického oscilátoru

$$E_{vib}(R^{AB} - R_0^{AB}) = k^{AB}(R^{AB} - R_0^{AB})^2 = k^{AB}(\Delta R^{AB})^2, \quad 8)$$

kde  $R_0$  je rovnovážná vazebná délka,  $R$  je aktuální délka vazby,  $k$  je silová konstanta harmonického oscilátoru a  $\Delta R$  je výchylka z rovnováhy. Správnější popis energie vibrace poskytuje neharmonický Morseho potenciál  $E_{Morse}$  (rovnice 9). Ten navíc lépe popisuje působení mezi atomy při velké vazebné vzdálenosti.

$$E_{Morse}(\Delta R) = D(1 - e^{-\alpha\Delta R})^2, \quad \alpha = \sqrt{\frac{k}{2D}} \quad 9)$$

V rovnici 9 znamená  $D$  disociační energii,  $\alpha$  je zabudovaná silová konstanta.

$E_{úhel}$  je energie odpovídající úhlu, který je vymezen třemi atomy A, B a C. Tuto energii lze popsat opět pomocí rovnice harmonického oscilátoru

$$E_{úhel}(\theta^{ABC} - \theta_0^{ABC}) = k^{ABC}(\theta^{ABC} - \theta_0^{ABC})^2, \quad 10)$$

kde  $\theta$  znamená aktuální úhel a  $\theta_0$  je úhel v rovnovážném stavu.

$E_{tor}$  je energie, která odpovídá změně torzního úhlu. Máme řadu atomů A, B, C, D, které jsou v této řadě spojené chemickou vazbou. Torzní úhel poté odpovídá rotaci kolem vazby BC a úhlu mezi vazbami AB a CD. Silové působení při změně torzního úhlu výrazně souvisí s nevazebnými interakcemi a stérickými efekty, proto pro jeho popis funkci harmonického oscilátoru nejde použít. Místo toho se je možné použít základní výraz

$$E_{tor}(\omega) = \sum_{n=1} V_n \cos(n\omega), \quad 11)$$

kde  $\omega$  znamená torzní úhel,  $V_n$  je výraz pro bariéru, která rotaci překáží a  $n$  udává velikost jedné periody při rotaci. Funkce popisující torzi je na rozdíl od předešlých periodická.

$E_{vdw}$  je energie, která odpovídá vzájemnému působení elektronových „mraků“. Podle vzdálenosti elektronových „mraků“, může být  $E_{vdw}$  kladné, nulové nebo záporné. Záporná  $E_{vdw}$  odpovídá vzniku přitažlivé interakce. To nastane, když se k sobě přiblíží na ideální vzdálenost dva elektronové „mraky“ a navzájem si sobě indukují časově proměnný slabý dipól. Vznikne tak slabá dipól-dipól interakce. Van der Waalsovskou interakci neboli také Londonovské nebo disperzní síly je možné základně popsat pomocí Lennardova-Jonesova potenciálu

$$E_{LJ}(R) = \varepsilon \left[ \left( \frac{R_0}{R} \right)^{12} - 2 \left( \frac{R_0}{R} \right)^6 \right], \quad (12)$$

kde  $R_0$  je vzdálenost při minimální energii,  $\varepsilon$  odpovídá velikosti energetického minima a  $R$  je aktuální vzdálenost. Londonovské interakce působí např. mezi nepolárními alkany.

$E_{el}$  je energie, která odpovídá vzájemnému silovému působení mezi náboji. Takové působení mezi dvěma náboji A a B lze popsat pomocí Coulombického potenciálu

$$E_{el}(R^{AB}) = \frac{Q^A Q^B}{\varepsilon R^{AB}}, \quad (13)$$

kde  $R^{AB}$  je vzdálenost dvou nábojů,  $Q$  je velikost náboje a  $\varepsilon$  je dielektrická konstanta.

$E_3$  je energie, která souvisí s tím, že  $R$ ,  $\theta$  a  $\omega$  se při svých změnách navzájem ovlivňují. Tedy tyto parametry výrazně spolu souvisí. Jejich vzájemné spřažení můžeme popsat rovnicí kombinující rovnice pro výpočet  $E_{vib}$ ,  $E_{úhel}$  a  $E_{tor}$ . To jestli a jaké parametry se do členu  $E_3$  kombinují, závisí na druhu silového pole. Není výjimkou, že se kombinují i jen dva parametry.

Příspěvky nevazebných energií  $E_{nevaz}$  jsou příspěvky, jež spotřebovávají výpočetní čas nejvíce. Na rozdíl od vazebných příspěvků totiž jejich počet neroste s velikostí molekuly lineárně, ale s druhou mocninou. Pro zjednodušení je proto výhodné uvažovat nevazebná působení jen na malou vzdálenost pomocí seznamu všech dvojic atomů do této vzdálenosti. To sníží počet nevazebných příspěvků do  $E_{sp}$ . Pro van der Waalsovské působení bývá tato vzdálenost 10 Å, pro elektrostatische 30 Å a pro interakce sítě bodových nábojů je ještě mnohem větší.

### 2.1.5 Další důležité pojmy z výpočetní chemie

**Optimalizace** mnohoprostorové funkce je jeden z problémů ve výpočetní chemii. Optimalizace znamená, že funkci hledáme její stacionární body. To jsou body funkce, jejichž první derivace se rovná nule. Tyto body představují minimum a maximum funkce. Většinou nás zajímá však minimum funkce. Např. hledáme energetické minimum dané molekuly (tedy její stabilní stav), kdy energie je funkcí všech souřadnic (mnohoprostorová funkce) všech atomů molekuly. Různé optimalizační techniky mohou poskytovat různé druhy minim funkce (lokální, globální). Podle toho v jakém tvaru jsou parametry zadávající funkci, může být tato funkce vyšetřována pomocí sestavy lineárních rovnic, které je možné vyřešit maticovými technikami. Častěji však jsou tyto parametry složitější a funkce se vyšetřuje iterativními metodami.

Hledání globálního energetického minima pro velké makromolekuly může záviset na počáteční geometrii makromolekuly. Pro malé systémy je možné použít při optimalizačních výpočtech takový systém, že systematicky vzorkujeme celou plochu potenciální energie a hledáme nejhlubší energetické minimum. Tento způsob však je pro velké makromolekuly nepoužitelný. Pro velké makromolekuly se používají metody, jež jsou schopné narušovat geometrii makromolekuly a tím se přemísťovat z jednoho lokálního minima do druhého. Zmíněny budou metody Monte-Carlo, molekulová dynamika. A dále metody simulace žíhání, metadynamika a metoda molekulové dynamiky založená na výměně replik, jež se označují jako takzvané pokročilé techniky molekulové dynamiky.[14]

Metoda Monte-Carlo (MC) je založená na tom, že náhodně posune pozici atomu nebo atomů v molekule, vypočítá energii nové konformace, porovná ji s předchozí nebo jinak zhodnotí a podle zhodnocení ne/použije novou konformaci jako základ pro další náhodné posunutí atomů.

Další metoda nalezení minima je molekulová dynamika (MD). Ta, jak bylo řečeno výše, používá Newtonovy pohybové zákony k popisu pohybu atomů. energii molekuly lze zde rozdělit na kinetickou a potenciální (silové působení mezi atomy). Během dynamiky je molekula schopná přeskocit přes energetickou bariéru do jiného minima jen, když energie (výška) bariéry je menší než celková energie zmenšená o potenciální.

Jak u metody Monte-Carlo, tak u molekulové dynamiky, je možné simulaci začít při vysoké teplotě a postupně ji snižovat. Jelikož vyšší teplota dává molekule schopnost přeskakovat větší bariéry a tím pádem více se pohybovat po ploše potenciální energie. Je pravděpodobnější, že ke konci simulace zůstane molekula blíže ke globálnímu energetickému minimu, než kdyby se vysoká teplota nepoužila. Metoda se nazývá simulace žíhání nebo chlazení.

Metadynamika je metoda, jež obsahuje paměť s již prozkoumanou plochou potenciální energie. Metoda funguje tak, že dochází ke vzorkování prostoru kolem energetického minima. To vede k nalezení bodu (vrcholu bariéry), ze kterého je možné jít do ještě nižšího energetického minima. Postup se stále opakuje. A jelikož je v celém procesu zabudována práce s pamětí, může dojít vždy k rozhodnutí, jestli je dané energetické minimum ještě hlubší než předchozí. Tím dochází k přeskakování čím dál tím vyšších energetických bariér a pohybu v širokém prostoru plochy potenciální energie.

Metoda molekulové dynamiky založená na výměně replik je metoda, jež zahrnuje více na sobě nezávislých souběžných simulací s metodou Monte Carlo. Kde každá simulace zvlášť je prováděna při jiné teplotě. Pozice atomů (tedy stav systému) jsou pak závislé na teplotě a energetických rozdílnostech mezi jednotlivými simulacemi. Podle daného předpisu si tak simulace mohou půjčovat, vzájemně vyměňovat své stavy (informace o pozicích atomů) a tím snadněji docílit přesunu stavu mezi dvěma energetickými minimy i přes vysokou bariéru.

**Solvatační modely vody**, které popisují obalení části rozpuštěné látky částicemi rozpouštědla, jsou nezbytné při popisu makromolekul. Ty se totiž ve svém přirozeném prostředí vyskytují v prostředí polárního rozpouštědla - vody a jejich nativní struktura je tedy na přítomnosti vody závislá. Molekuly vody s biomakromolekulami interagují pomocí slabých nevazebných interakcí a tvoří v blízkosti biomakromolekul určitou slupku.

## 2.2 Struktura biomakromolekul

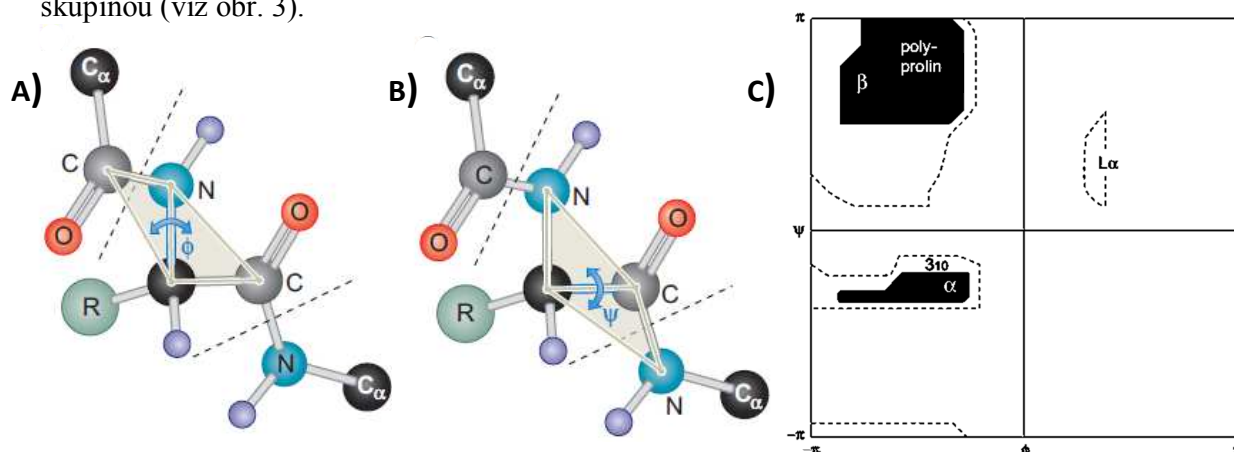
Biomakromolekuly jsou stvořeny ze základních jednotek takovým způsobem, že kovalentní vazba spojuje mezi sebou základní jednotky a jednotlivé atomy v základních jednotkách. Aminokyseliny tvoří proteiny. Nukleové kyseliny tvoří polynukleotidy. Kovalentní vazba je silná a ve fyziologických podmínkách buňky se samovolně nerozpadá. Toho jsou schopny slabé nevazebné interakce. Ty jsou sice řádově slabší než kovalentní vazba, ale jsou nezbytné pro vznik správné struktury a biologickou funkci biomakromolekul. Biomakromolekuly pomocí nich interagují mezi sebou, v rámci své molekuly nebo s okolím (rozpouštědly, ionty). Slabé nevazebné interakce oproti kovalentní vazbě mohou působit i na delší vzdálenosti. Mají více variací, jež se týkají vazebných možností (vazebné úhly, vaznosti). Můžou mít méně omezení pro orientaci vázaných atomů oproti např. peptidické vazbě, kde atomy leží v jedné nehybné rovině. To všechno znamená, že jich je v biomakromolekulách velké množství a i přesto, že jejich energie je relativně malá, jsou velmi významné.

Hlavní typy slabých nevazebných interakcí, které jsou přítomny v biomakromolekulách, jsou van der Waalsovské interakce, vodíkové vazby, interakce mezi ionty (solný můstek) a hydrofobní interakce (jeden z hlavních efektů, jež mají vliv a jsou důvodem sbalení proteinu).

Může docházet k efektům přenosů náboje a vzájemnému ovlivnění pohybu (dynamiky) mezi solventem a solutem. Molekuly vody mohou být polarizovány a orientovány podle náboje přítomného na biomakromolekulách.

### 2.2.1 Proteiny

Proteiny jsou polymery, jejichž základní stavební jednotka je aminokyselina. Existuje dvacet základních aminokyselin. Každá je tvořena z atomu uhlíku  $C_{\alpha}$ , na kterém jsou čtyři substituenty. Vodík, aminoskupina, karboxylová skupina a postranní řetězec, jímž se aminokyseliny od sebe liší (viz obr. 3). Aminokyseliny se spojují pomocí peptidové vazby do řetězců. Pro popis celkové konformace řetězce stačí znát úhel  $\Phi$ ,  $\Psi$  pro každou aminokyselinu.  $\Phi$  popisuje rotaci kolem vazby mezi NH skupinou a atomem  $C_{\alpha}$ .  $\Psi$  popisuje rotaci kolem vazby mezi atomem  $C_{\alpha}$  a CO skupinou (viz obr. 3).



Obr. 3: Na obrázcích A), B) jsou znázorněny dva významné úhly ( $\Phi$ ,  $\Psi$ ) v aminokyselinách, jež se používají pro popis konformací řetězce aminokyselin. Čárkovaně je znázorněna peptidická vazba. Obrázek C) je tzv. Ramachandranův diagram, jež znázorňuje typy sekundárních struktur, jež vzniknou vhodnou kombinací hodnot úhlů  $\Phi$  a  $\Psi$ . Symbol  $\alpha$  znamená  $\alpha$ -helix,  $\beta$  znamená  $\beta$ -skládaný list. Poly-prolin,  $3_{10}$  a  $L_{\alpha}$  jsou další typy helixů. Převzato.



Strukturu proteinů můžeme rozdělit do čtyř úrovní. Zaprvé, primární struktura udává pořadí aminokyselin v proteinu. Zadruhé, sekundární struktura udává, jaké prostorové uspořádání zaujímá krátká část řetězce aminokyselin (např.  $\alpha$ -helix,  $\beta$ -skládaný list). Zatřetí, terciální struktura udává, jaké je prostorové uspořádání celého složeného aminokyselinového řetězce nebo jedné domény. Doména je část polypeptidického řetězce, která je sbalená, a co je důležité, její stabilita nezávisí na zbytku řetězce. Začtvrté, kvartérní struktura znamená, jak je k sobě prostorově orientováno (naskládáno) více již sbalených polypeptidických řetězců. Struktura proteinu nemusí být však jen jedna. Např. proteinové enzymy se mohou částečně přeuspořádávat do jiné struktury při své práci.

Do jaké prostorové struktury se řetězec aminokyselin sbalí, je dáno sekvencí aminokyselin v tomto řetězci.[18] Víme, že nativní struktura proteinu má nejmenší energii. Z toho vyplývá, že bychom mohli být schopni **předpovědět strukturu proteinu** jen na základě sekvence aminokyselin.[15, 16] Stačilo by, aby počítač pomocí vhodných nástrojů a programů propočítal energii pro všechny konformace proteinu a zvolil strukturu proteinu s nejnižší energií. Takhle by to mohlo fungovat. Bohužel i jen malý aminokyselinový řetězec by tvořil tak velké množství konformací, že by to žádný počítač nikdy nespočítal. Naštěstí je možné pro předpověď struktury proteinu použít tzv. homologní modelování, tedy modelování struktury proteinu založené na podobnosti struktur proteinů s podobnou sekvencí.[15, 19, 20]

Existují dvě základní metody určení struktury proteinu. Je to metoda rentgenové krystalografie a nukleární magnetické resonance.[17, 19] Bohužel tyto metody nelze použít pro zjištění struktury všech proteinů. Protože vzorek proteinu musí splnit určité požadavky, aby se pomocí těchto metod mohl zkoumat. Pro metody rentgenové difrakce musí být vzorek krystalem a pro metody nukleární magnetické resonance musí být schopný se rozpustit v rozpouštědle.[20]

### 2.3 Cytomegalovirus

Skupina herpesvirů je skupina virů, jež svou genetickou informaci nosí ve formě dvouvláknové DNA. Tyto viry jsou přítomné všude (zvířata, lidé) a mohou způsobit infekce v rozsahu od kožních vředů po encefalitidu. Viry v této skupině mají podobnou morfologii a relativně rozsáhlý genom. Podstatně se liší v genetické sekvenci. Většina herpesvirů ovlivňuje metabolismus buňky a tím může způsobit dokonce smrt buňky. Výjimkou je cytomegalovirus, jenž napodobuje syntézu proteinů a nukleových kyselin. Herpesviry můžeme rozdělit do tří podskupin  $\alpha$ ,  $\beta$  a  $\gamma$  (viz tab. 1).

Tab. 1: Tabulka ukazuje klasifikaci herpesvirů. Osm herpesvirů je rozděleno do tří podskupin. Cytomegalovirus je zde označen jako HHV-5 a patří do podskupiny herpesvirů  $\beta$ . Převzato.[3]

<b>Zástupci podskupiny herpesvirů <math>\alpha</math></b>			
<b>Běžný název</b>	herpes simplex virus-1	herpes simplex virus-2	virus varicella-zoster
<b>Označení</b>	HHV-1	HHV-2	HHV-3
<b>Zástupci podskupiny herpesvirů <math>\beta</math></b>			
<b>Běžný název</b>	cytomegalovirus	lidský herpesvirus-6	lidský herpesvirus-7
<b>Označení</b>	HHV-5	HHV-6	HHV-7
<b>Zástupci podskupiny herpesvirů <math>\gamma</math></b>			
<b>Běžný název</b>	Epsteinův-Barrův virus		herpesvirus Kaposiho sarkomu
<b>Označení</b>	HHV-4		HHV-8

Cytomegalovirus (CMV), nazývaný též také lidský cytomegalovirus (HCMV, angl. zkratka human cytomegalovirus) má největší genom z herpesvirů. Jeho název vychází z toho, že tento virus způsobuje zvětšení buňky, kterou infikoval. Také je důležité zdůraznit, že existují nespočetné různé kmeny tohoto viru, které se liší v genetické informaci a fenotypu.[3]

Jak již bylo řečeno, cytomegalovirus je všudypřítomný. Infikuje většinu světové populace.[21] Virus je však přítomen často v neaktivní podobě a zvýšené riziko nemoci způsobené tímto virem mají lidé s oslabenou imunitou. Např. lidé, kteří podstupují transplantaci, jež musí brát imunosupresiva, aby jejich těla cizí orgán přijala nebo lidé trpící nemocí AIDS.[3, 1] V případě aktivní infekce může cytomegalovirus způsobit neurologické poruchy, záněty trávicí soustavy, sítnice a plic. Cytomegalovirus může infikovat i monocyty (typ bílých krvinek). To ve svém důsledku může způsobit větší náchylnost k velkým bakteriálním a houbovým infekcím.

Infekce dané cytomegalovirem se léčí pomocí léků gancicloviru[1, 2] a foscarnetu, které blokují polymerázu DNA viru. Dále je možné pro léčbu použít léky cidofovir[3], valganciclovir[4] a acyclovir[5]. V různých fázích vývoje jsou léky maribavir, brincidofovir, letermovir.[6] Při dlouhodobém současném používání gancicloviru a silných imunosupresiv se na ganciclovirus vyvíjí rezistence.[7] Rezistence vůči cidofoviru nebyla ještě registrována, ale vůči foscarnetu je vyvinuta také.[6]

Genom cytomegaloviru kóduje více než 150 proteinů. Dva z nich jsou extrémně důležité pro činnost antivirových přípravků. Jsou to proteiny cytomegalová fosfotransferáza (kódovaná genem UL97) a cytomegalová DNA polymeráza (kódovaná genem UL54).[4, 22] Fosfotransferáza fosforyluje antivirová léčiva do jejich aktivní podoby. Polymeráza DNA je homologní s polymerázami ostatních herpesvirů a je to element, na který antivirotika běžně cílí.[4, 8]

## 2.4 Polymeráza DNA

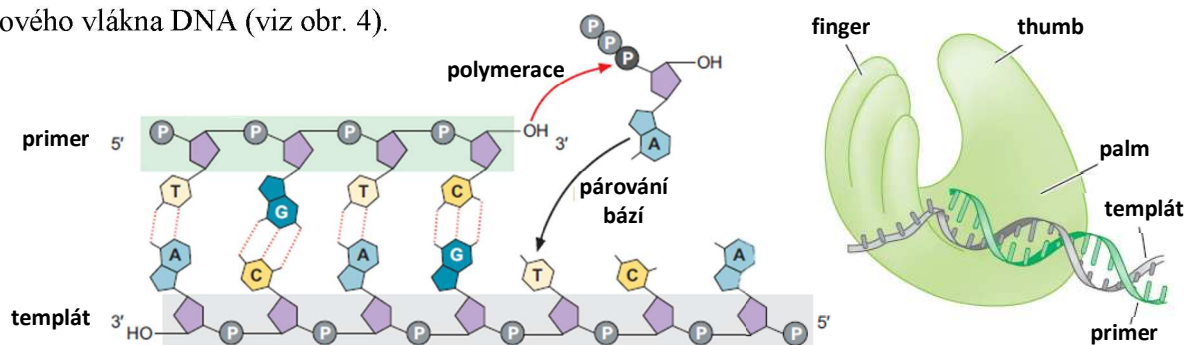
DNA polymerázy jsou enzymy, které katalyzují syntézu polydeoxyribonukleotidů z deoxyribonukleosid trifosfátů (dNTP). Kromě toho mají i další kontrolní a opravné funkce syntézy DNA, funkce syntézy primeru RNA a např. funkci vyplňování spojů v nově syntetizované DNA.[23] Podle toho jaké funkce polymeráza opravdu má a z důvodu existence různých druhů organismů můžeme polymerázy dělit na různé typy (viz tab. 2). Všechny typy polymeráz lze dále zařadit do polymerázových rodin (A, B, C, D, X, Y, RT) dle sekvenčních a strukturních podobností.[24, 25, 26] Polymerázy virů mají značnou sekvenční podobnost s polymerázami typu pol  $\alpha$ , pol  $\beta$ , pol  $\epsilon$ , pol II a T4,  $\Phi$ 29 (polymerázy určitých bakteriofágů).[27]

*Tab. 2: Tabulka ukazuje základní typy polymeráz. Existuje šest typů prokaryotických polymeráz a třináct typů eukaryotických polymeráz. Převzato.[15, 28]*

<b>Základní typy polymeráz</b>	
<b>Prokaryotické typy</b>	Pol I, Pol II, Pol III – jaderná, Pol III – holoenzym, Pol IV, Pol V
<b>Eukaryotické typy</b>	Pol $\alpha$ , Pol $\beta$ , Pol $\gamma$ , Pol $\delta$ , Pol $\epsilon$ , Pol $\theta$ , Pol $\zeta$ , Pol $\lambda$ , Pol $\mu$ , Pol $\kappa$ , Pol $\eta$ , Pol $\iota$ , Rev1

Kromě polymeráz se na replikaci (zdvojení) DNA podílejí i další enzymy. Nejdůležitější jsou helikázy, jež rozmotávají dvoušroubovici DNA a vytváří replikační oko, topoisomerázy, které uvolňují torzní napětí v rozmotané DNA, primázy DNA, jež syntetizují RNA primery, proteiny SSB (single strand binding), jež brání předčasnému uzavření replikačního oka a ligázy DNA, jež spojují fragmenty DNA pomocí fosfodiesterové vazby.

Syntéza DNA probíhá vždy podle vzoru již existující DNA. Pro syntézu je nezbytné, aby byly k dispozici dNTP a komplex primer-templát. Templát je část dlouhé jednovláknové DNA, podle níž (komplementárně) vzniká nová jednovláknová DNA. Primer je na rozdíl od templátu krátký a jeho báze se párují s bázemi dlouhého vlákna DNA těsně před oblastí templátu. Volná skupina 3'-OH na primeru přiléhá těsně k vláknu templátu a připojují se na ni nukleotidy při vzniku nového vlákna DNA (viz obr. 4).



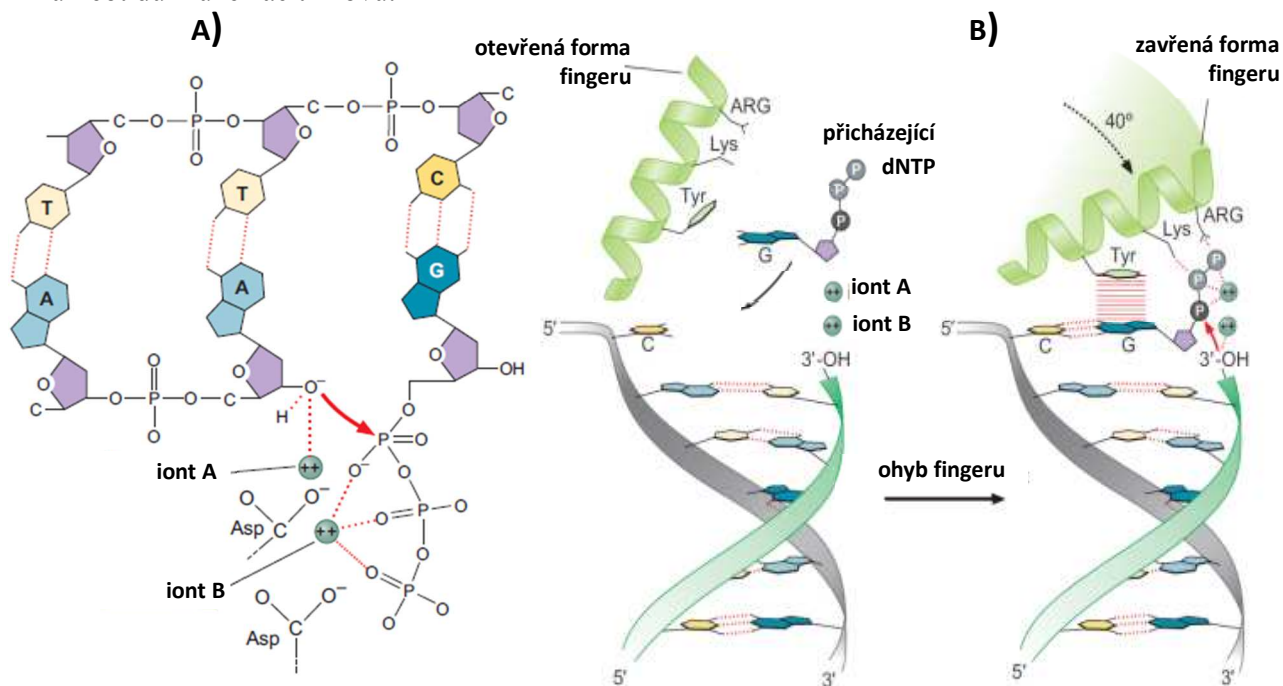
Obr. 4: Na obrázku vlevo vidíte komplex primer-templát a schematicky vyznačené připojení dNTP na primer. Na obrázku vpravo vidíte schematicky zaznačenou polymerázu DNA (zeleně) se svými doménami a komplex primer-templát, jež je touto polymerázou zpracováván. Převzato.

Při syntéze nového vlákna DNA 3'-OH skupina primeru působí na  $\alpha$ -fosfát přistupujícího dNTP. Následně proběhne reakce, která zapříčiní, že primer je o jeden nukleotid delší (přes báze spárovány s vláknem templátu) a vznikne pyrofosfát (zbylé dva fosfáty z původně přistupujícího dNTP). Dále je nutné dodat, že nově vzniklé vlákno DNA je vzhledem k templátu orientováno antiparalelně (obráceně).

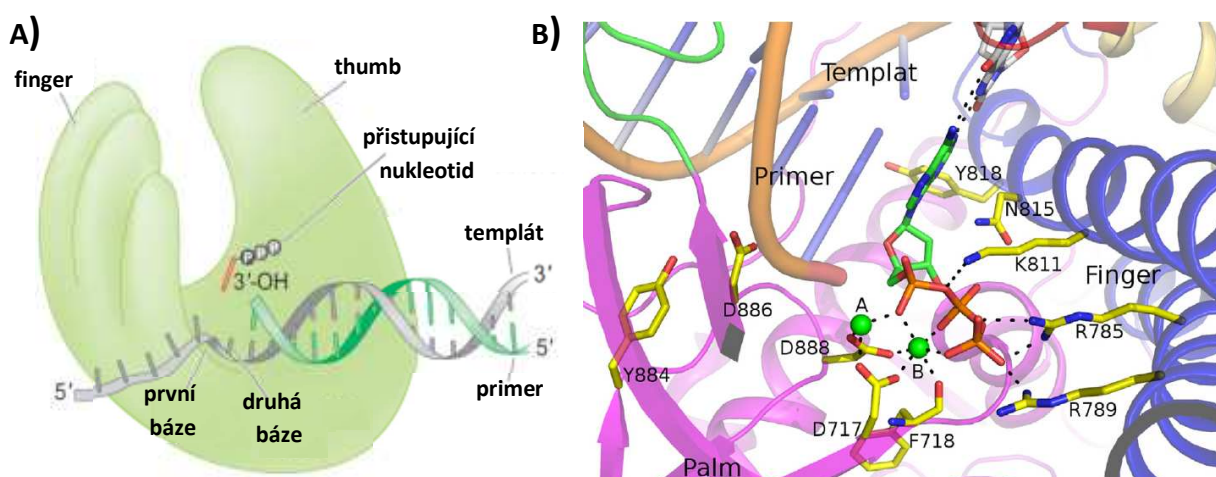
Jelikož DNA polymeráza obaluje komplex primer-templát jako pravá lidská ruka, jsou její tři domény pojmenovány podle toho (viz obr. 4). Jmenují se thumb (angl. palec), finger (angl. prst) a palm (angl. dlaň).

Doména palm je složena převážně z  $\beta$ -skládaných listů a její část je hlavní složkou katalytického místa enzymu. Toto katalytické místo obsahuje dva dvouvalné kovové ionty (např. dva  $Mg^{2+}$  nebo dva  $Zn^{2+}$ ), jež ovlivňují prostředí kolem dvojice dNTP a 3'-OH (viz obr. 5-A). Jeden kationt oslabuje vazbu H-O na skupině 3'-OH takovým způsobem, že vzniká 3'-O<sup>-</sup> skupina, která má záporný náboj, a může lépe napadat atom fosforu (kladný náboj) na  $\alpha$ -fosfátu. Druhý kationt stabilizuje zbylé dva fosfáty ( $\beta$ ,  $\gamma$ ) na přichozím dNTP a následně i pyrofosfát. Dále část domény palm vytváří, v případě správně párovaných bází, množství vodíkových vazeb s malým žlábkem vznikající DNA. V případě nesprávného párování jsou tyto vodíkové vazby určitým způsobem ovlivněny a dochází ke zpomalení syntézy.

Doména finger se na katalytické aktivitě polymerázy podílí takovým způsobem, že některé její aminokyseliny se vážou na přicházející dNTP. Jakmile se vytvoří správné párování bází (dNTP-templát), je dNTP na místě uzavřeno pomocí ohybu domény finger (viz obr. 5-B). Dále doména finger způsobuje ohyb mezi první a druhou bází na vlákně templátu (viz obr. 6-A). To vede k tomu, že v katalytickém místě enzymu je jen jedna báze a přicházející dNTP se orientuje jen k ní a žádné jiné. Po úspěšném navázání dNTP na primer se doména finger vrací do původní polohy. Celý komplex primer-templát se posune o jeden nukleotid a celý proces navázání nukleotidu může začít znovu.



Obr. 5: Na obrázku A) vidíte, jakým způsobem interagují kationty kovu s dNTP a 3'-OH v doměně palm. Červená šipka znázorňuje reakci mezi dNTP a 3'-OH. Na obrázku B) vidíte doménu finger jak se váže na přicházející dNTP a uzavírá ho na místě. Převzato.



Obr. 6: Na obrázku A) můžete vidět ohyb na vlákně templátu mezi první a druhou bází způsobený doménou finger. To dává více místa k navázání dNTP na správnou (první) bázi templátu. Na obrázku B) můžete vidět katalytické místo polymerázy herpes simplex viru. Pro polymerázovou aktivitu jsou nejdůležitější rezidua D888, D717, F718, K811, R785 a R789, jež s pomocí kationtů kovů A, B interagují s přicházejícím nukleotidem. Převzato.

Doména thumb se sice přímo nepodílí na syntéze DNA, interaguje však s již vytvořenou DNA. To má dva důsledky. Zaprvé, umístění primeru a aktivního místa se díky rigidnější vazbě mezi DNA a polymerázou lépe udržuje ve vzájemné optimální pozici. Zadruhé, má to vliv na zlepšení vlastností polymerázy v tom smyslu, že váže lépe mnoho dNTP na primer vždy, když se vytvoří komplex primer-templát.

DNA polymerázy můžeme rozdělit do dvou skupin (procesivní, neprocesivní). Neprocesivní polymerázy se vždy po navázání jednoho nukleotidu na primer odpojí od DNA. Rychlost syntézy je tedy pomalá. Naopak procesivní polymerázy se po každém navázání nukleotidu na primer neodpojují od DNA, ale naopak navazují na nově vznikající vlákno DNA tolik nukleotidů, kolik můžou. Rychlost syntézy je tedy velká. To je umožněno hladkým posouváním DNA v polymeráze. Toto hladké posouvání je způsobeno tím, že komplex primer-templát je k polymeráze vázán nespécificky, elektrostatisky doménou thumb a pomocí vodíkových vazeb doménou palm. Při posouvání polymerázy po DNA je pak polymeráza stále uchycena minimálně právě přes elektrostatische interakce s doménou thumb. To umožňuje, aby mohla být polymeráza těsně spojena s DNA a mohla se plně věnovat replikaci bez nutnosti rozvázání od DNA. Ještě větší rychlost posouvání polymerázy kolem DNA může být dosažena spřáhnutím polymerázy a některých pomocných enzymů dohromady.

Přestože polymerázy pracují velice přesně, co se týče párování bází, občas se chyby vyskytnou. Tyto chyby mohou odstranit enzymy, které se jmenují nukleázy. Jsou dva typy nukleáz.[15, 23] Exonukleázy (první typ) odbourávají nově vzniklé vlákno DNA na jejím začátku, tedy na 3' konci. Endonukleázy (druhý typ) odbourávají nově vzniklé vlákno DNA i uvnitř něho. Funkce exonukleázy je spolu s funkcí polymerázy DNA přítomna na tomtéž řetězci polypeptidu. To exonukleáze umožňuje odstranit chybný nukleotid hned po jeho navázání na primer. Funguje to tak, že výskytem chybné báze na 3' konci primeru se oslabí interakce domény palm s DNA. To má za následek změnu geometrie mezi skupinou 3'-OH a přístupujícím nukleotidem. Dojde k destabilizaci aktivního místa polymerázy komplexu primer-templát a následně k rozpárování několika koncových bází nového řetězce DNA s řetězcem templátu. Volnější (nespárovaný) konec primeru se přesune do aktivního místa exonukleázy. Exonukleáza splní svou funkci a zbaví primer špatného nukleotidu. Poté se řetězec primeru opět svým 3'-OH koncem vrátí na aktivní místo polymerázy a dojde k opětovnému spárování několika koncových nukleotidů primeru s vláknem templátu. V případě procesivní polymerázy se toto vše děje bez rozrušení všech spojení (interakcí) mezi DNA a polymerázou. Toto opravování chyb zvyšuje přesnost syntézy DNA z 1 chyby za  $10^5$  přidaných nukleotidů na 1 chybu za  $10^7$  přidaných nukleotidů.

#### **2.4.1 Polymeráza DNA – herpesvirů[30]**

Polymerázy herpesvirů patří do polymerázové rodiny B (pol  $\alpha$ , pol  $\delta$ , pol  $\zeta$ , pol  $\epsilon$  a některé polymerázy DNA bakteriofágů, ostatních DNA virů, rostlin a hub). Polymerázová rodina B se od ostatních polymerázových rodin liší v tom, že na rozdíl od nich tvoří mnohapodjednotkové komplexy z DNA polymeráz a dalších doprovodných enzymů, které replikaci provázejí. Polymerázy DNA v této rodině jsou také charakteristické v tom, že mají i opravnou funkci a jsou spojeny s procesivním faktorem. Všechny polymerázy DNA herpesvirů obsahují homologní konzervované oblasti I-VII (dle klesající podobnosti) a  $\delta$ -region C, jež je k polymerázám DNA

herpesvirů homologní taky.[31] Mechanismus polymerace v polymerázové rodině B (stejně jako ve všech ostatních polymerázových rodinách) probíhá pomocí interakce dvou iontů kovů a dvou reziduí kys. asparagové (viz výše).[32]

Na replikaci DNA cytomegaloviru se podílí šest proteinů. Jsou to DNA polymeráza a její protein procesivního faktoru, SSB protein, dále trojčlenný komplex tvořený dvěma primázami (HP1, HP2) a helikázou (HP3). Všechny tyto proteiny jsou v uvedeném pořadí v DNA kódovány geny UL54, UL44, UL57, UL105, UL70 a UL102.[30] Další velké množství enzymů se podílí na zahajovacích a ukončovacích mechanizmech replikace. Polymeráza herpes simplex viru je kódována pomocí UL30.

Funkce šesti proteinů nutných pro replikaci je následující. Komplex polymerázy a procesivního faktoru má za úkol syntézu nového vlákna DNA na primeru. SSB protein má za úkol udržet dvě vlákna DNA (během procesu syntézy nového vlákna DNA) od sebe a rozmotaná. Komplex polymeráz a helikázy má za úkol syntetizovat primer.

Polymerázy cytomegaloviru (CMV) a herpes simplex viru (HSV) nemají jen jednu funkci. Prodlužují řetězec primeru, mají exo- a endo- (enzym ribonukleáza H) nukleázovou aktivitu. V případě, že C-konec jejich katalytických podjednotek interaguje s procesivním faktorem (UL42 u HSV, UL44 u CMV), může docházet ke zvýšení procesivity při vzniku DNA. To je dáno tím, že UL42 brání předčasnému odstoupení polymerázy DNA od replikovaného řetězce. UL44 je složen z 290 aminokyselin a je to homodimer. UL42 je na rozdíl od UL44 složen z 488 aminokyselin a je heterodimerem.

Jelikož zatím nebylo možné polymerázu DNA cytomegaloviru vykrytalizovat, neznáme její krytalografická data a tedy ani její strukturu. Polymeráza (apoenzym) herpes simplex viru-1 však vykrytalizovaná byla, známe tedy její prostorovou strukturu.[33] Pro polymerázy DNA herpesvirů je dokonce tato struktura považovaná za vhodný vzor.[30] Zkoumáním této struktury bylo zjištěno, že obsahuje pět domén (obdobné v celé polymerázové rodině B). Jsou to domény pre-NH<sub>2</sub> a NH<sub>2</sub> doména, 3'-5'exonukleázová doména<sup>1</sup> a domény palm, finger a thumb. Domény palm, finger a thumb jsou katalytickým místem polymerázy (viz výše, viz obr. 6-vpravo).

### **Domény polymerázy herpes simplex viru**

Doména pre-NH<sub>2</sub> tvoří začátek polymerázy herpes simplex viru. Není přesně známo, jaký je její význam, ale předpokládá se, že pomocí motivu FYNPYL interaguje s doménou exo. Také se předpokládá, že interaguje s obslužnými proteiny replikace (helikázo-primázový komplex) a podílí se na jejich zformování do replisomu (komplex obslužných proteinů při replikaci).

Doména NH<sub>2</sub> obsahuje tři motivy. Zaprvé, dva β-skládané listy (β) a helix (α). Zadruhé, sekundární struktury βαββαβ. Zatřetí, dva helixy. Předpokládá se, že tato doména by mohla hrát hlavní úlohu v endonukleázové funkci polymerázy.

---

<sup>1</sup> Doména se jmenuje 3'-5'exonukleázová doména nebo i jen exonukleázová doména. V dalším textu bude označována jako doména exo.

Doména exo obsahuje konzervované oblasti IV a  $\delta$ -region C. Dále čtyři rezidua, která mají negativní náboj a interagují s dvěma kovovými ionty (D368, E370, D471, D581). V doméně exo je obsažena i funkce ribonukleázy H, jež odděluje primázu RNA od nově syntetizované DNA, aby mohly vzniknout Okazakiho fragmenty.[34]

Doména palm obsahuje konzervované oblasti VII, I a II. Oblast I (DTDS) je umístěna za doménou finger, oblast II (DXXLYPS) před doménou finger a společně tvoří tři antiparalelní  $\beta$ -skládané listy. Rezidua F718, D717 a D888 (oblasti I, II) stabilizují  $\alpha$ -fosfát příchozího nukleotidu. Dále doména palm obsahuje motiv KKRY stabilizující interakci bází primeru a templátu. Předpokládá se, že motiv KKRY by spolu s Y818, Y884 a D886 mohl interagovat s předposledním párem bází na primeru DNA a tím cítit možné chyby v párování nukleotidů v nově vzniklém řetězci DNA.

Doména finger obsahuje dva navzájem antiparalelní  $\alpha$ -helixy (N, P) a konzervované oblasti III (KXXXNSXYG) a VI. Doména finger dále obsahuje rezidua R785, R789 a K811 jež interagují s fosfátovými skupinami příchozího nukleotidu. Konzervovaná rezidua N815 a Y818 interagují s příchozím nukleotidem, podílí se na tvorbě komplementárních vodíkových vazeb a tím na správném výběru nukleotidu. Doména thumb obsahuje velké množství  $\alpha$ -helixů a konzervovanou oblast V.

### **Model polymerázy cytomegaloviru**

Model polymerázy cytomegaloviru byl již v určité formě vystavěn. Pomocí programu MODELLER a 3D-Jury Meta server, byly vytvořeny modely polymerázy cytomegaloviru v otevřené a uzavřené formě obsahující 602 a 604 reziduí.[35] Pro porovnání, modely polymerázy cytomegaloviru sestavené v této práci jsou tvořeny z cca 980 aminokyselin (dle typu modelu) a při jejich sestavování byly použity nástroje serveru I-TASSER.

### **2.5 Použité programy, databáze a nástroje**

V této magisterské práci byly použity konkrétní programy, databáze, nástroje a výpočetní zařízení. Informace o nich jsou uvedeny níže. Dané nástroje, programy a databáze ulehčovaly zpracování dat poskytnutých výpočetním softwarem a dovolovaly systematické hledání informací nutných pro vytvoření vstupů do výpočetních softwarů.

PyMol je počítačový program, jehož je možné použít k vizualizaci a úpravě molekul. Obsahuje základní nástroje pro analýzu strukturních vlastností molekul.[36]

Amber je software, který se ve výpočetní chemii používá k simulaci biomakromolekul. Jeho výpočty jsou založeny na molekulové mechanice s využitím různých typů silových polí. Obsahuje množství doprovodných obslužných programů.[37]

WinSCP a Putty jsou programy, které dokáží propojit osobní počítač se vzdáleným serverem. Výpočetní servery (spravované katedrou fyzikální chemie), jež jsem ke své práci využíval, se jmenovaly elwe a aule. Tyto názvy odkazují na jména Elwë (jeden z vůdců elfů) a Āule (bůh, kovář, mistr řemesel).[38]

VMD (dle Visual Molecular Dynamics) je program, který je možné použít k vizualizaci makromolekul. Hlavně je však přímo navržen pro vizualizaci trajektorií, jež pocházejí ze simulací molekulové dynamiky.[39]

PDB databáze (dle Protein Data Bank) je běžně používaný název databáze volně přístupné na internetu, jež obsahuje strukturní informace o biomakromolekulách (každá má svůj vlastní identifikační pdb kód). Tyto informace a struktury lze hledat a jsou seřazeny podle různých kategorií.[40]

UniProt (dle Universal Protein resource) je databáze sekvencí proteinů, která umožňuje hledat proteiny s podobnou sekvencí (nástroj BLAST) a porovnávat sekvence proteinů. Obsahuje navíc další související informace o proteinech.[41]

## I-TASSER

I-TASSER (dle Iterative Threading Assembly Refinement algorithm) je server, který je možné použít pro předpověď strukturního modelu a funkce proteinu dle sekvence aminokyselin.[42, 43, 44]

Proces návrhu struktur pomocí I-TASSER je složen ze čtyř kroků. Prvním krokem je hledání vhodných templátů, na jejichž základě bude model vystavěn. K tomuto hledání se používá program LOMETS jež prohledává pdb databáze. Druhým krokem je sestavení modelu z celistvých fragmentů templátů pomocí simulací metodou Monte Carlo s výměnou replik. Části modelu bez vhodného templátu (hlavně nestrukturované smyčky) jsou vytvořeny na základě *ab initio* modelování. Třetím krokem je simulace těchto fragmentů znovu, ale tentokrát se začátkem v místech prostorových střetů a omezení s cílem odstranění těchto potíží. Následně jsou vybrány struktury s nejnižší energií a ve čtvrtém kroku proběhne optimalizace sítě vodíkových vazeb v těchto strukturách.

V případě, že o modelovaném proteinu jsou již známy určité strukturní informace, je možné některé do vstupu zadat. Ve vstupu je možné určit vzdálenost mezi dvěma konkrétními atomy. Dále je možné zadat templát, na kterém chceme, aby byl model vystavěn. A to v podobě kódu již existujícího proteinu v PDB databázi nebo vlastní struktury ve formátu pdb. U preferovaných templátů, je možné zadat, jakým způsobem mají být sekvence templátu a navrhovaného proteinu srovnány. Naopak je možné i vyřadit proteiny (přes pdb kód), u kterých nechceme, aby byly jako templáty použity.

Výstup ze serveru I-TASSER obsahuje (mimo jiné). Zaprvé, předpovězenou sekundární strukturu s parametrem, jež popisuje jistotu předpovědi. Zadruhé, graf předpovězené hodnoty B-faktoru pro každou aminokyselinu. Vyšší hodnota B-faktoru znamená větší tepelnou mobilitu aminokyseliny. Zatřetí, deset templátů, které jsou ke vstupní sekvenci aminokyselin nejvíce podobné. Tato podobnost je hodnocena pomocí tzv. Z-score, kde vyšší hodnota znamená vhodnější templát. Hodnoty nad 1 znamenají dobrou shodu. Templátů je deset, protože se využívá deseti různých programů. Začtvrté, pět předpovězených modelů. Celková správnost modelu je zhodnocena pomocí parametru C-score, kde vyšší hodnota znamená více kvalitní model. Hodnota C-score nad -1,5 indikuje globálně správnou topologii modelu. Modely jsou řazeny dle klesající hodnoty



C-score. Nemusí to tak být ovšem vždy, protože na řazení pěti modelů má vliv i jejich velikost. To však nemění nic na tom, že první model by měl být průměrně nejspolehlivějším. Ke každému modelu je navíc přiřazen odhad lokální správnosti modelu. Zápáté, deset strukturně nejvíce podobných proteinů k prvnímu modelu. Podobnost je zde vyjádřena v parametru TM-score. Hodnota TM-score nad 0,5 napovídá, že obecně jsou porovnávané proteiny složeny stejně.

### 3. Výpočetní část

#### 3.1 Nastavení výpočtů v programu Amber

Pro přípravu vstupních souborů do programu Amber (za účelem následných výpočetních simulací) byly použity následující silová pole. Silové pole OL15, jež popisuje RNA a DNA.[45] Silové pole ff14SB, jež popisuje proteiny.[46] A silové pole SPC/E (dle extended simple point charge), jež popisuje model molekuly vody. Byly použity také speciální parametry pro DCP. Struktury připravené pro výpočetní simulace obsahovaly solvatační obal ve tvaru oktaedru s přítomnou sítí  $K^+$  a  $Cl^-$  iontů. Celková koncentrace těchto iontů byla taková, že odpovídala koncentraci fyziologického roztoku. Délka molekulové dynamiky byla vždy nastavena na 1  $\mu$ s tvořenou ze 100 cyklů. Jeden cyklus simuloval 10 ns s časovým krokem 0,004 ps. Takto velký časový krok byl umožněn pomocí přesunu hmotnosti těžkého atomu na atom vodíku, jež byl s tímto těžkým atomem svázán vazbou.[47]

Struktura vznikajícího modelu polymerázy cytomegaloviru byla často upravována. Proto bylo nutné před molekulovou dynamiku zařadit proces minimalizace, který záměrně pozměněnou strukturu zrelaxoval. K tomu bylo použito následujících deset kroků. Prvním krokem byla relaxace všech atomů vodíku (kromě vody). Druhým byla relaxace vody a iontů. Třetím byla krátká molekulová dynamika všeho kromě proteinu a nukleotidů. Čtvrtý až sedmý krok zahrnoval relaxaci nepáteřních atomů proteinů a nukleových kyselin. Zároveň však docházelo ke snižování potenciálu vloženého na páteřní atomy. Z tohoto důvodu (ke konci této minimalizační fáze) docházelo taky k relaxaci páteřních atomů proteinů a nukleových kyselin. V osmém kroku proběhla minimalizace všech atomů. Během posledních dvou kroků probíhala termalizace. Tzn. krátká molekulová dynamika (nejprve za konstantního objemu, poté tlaku) během níž vzrůstala teplota. Pomocí statistického zpracování rychlostí atomů při daných teplotách, byla každému atomu přiřazena rychlost. Tato rychlost byla použita jako počáteční rychlost atomů na začátku molekulové dynamiky.

#### 3.2 Výchozí struktury použité pro výstavbu polymerázy cytomegaloviru

Při sestavování modelu polymerázy cytomegaloviru bylo vycházeno ze tří různých polymeráz (viz tab. 3). Byly to polymerázy herpes simplex viru-1, kvasinky a archebakterie, u nichž známe jejich strukturu.[33, 48, 49] Polymeráza herpes simplex viru-1 byla použita z několika důvodů. Spolu s cytomegalovirem patří do skupiny herpesvirů. Nástrojem BLAST byla identifikována jako sekvenčně nejvíce podobná k polymeráze cytomegaloviru. Je považovaná jako dobrý vzor pro polymerázy herpesvirů.[30] Polymerázy kvasinky a archebakterie byly použity z několika důvodů. Podobně jako polymeráza herpes simplex viru-1 byly nástrojem BLAST identifikovány jako sekvenčně podobné k polymeráze cytomegaloviru. Dalším důvodem jejich použití bylo, že polymeráza archebakterie a kvasinky obsahuje DNA. V polymeráze kvasinky je DNA přítomna ve formě polymerázové. V polymeráze archebakterie je DNA přítomna ve formě exonukleázové. Struktury vstupních polymeráz jsou zobrazeny dále (viz obr. 7).

Všechny tři výchozí struktury jsou polymerázy. Z tohoto důvodu všechny tři obsahují ekvivalentní domény. Zčásti je doména pre-NH<sub>2</sub> zobrazena ve struktuře polymerázy herpes simplex viru-1. V ostatních není zobrazena vůbec. Platí, že ve všech třech polymerázách začíná

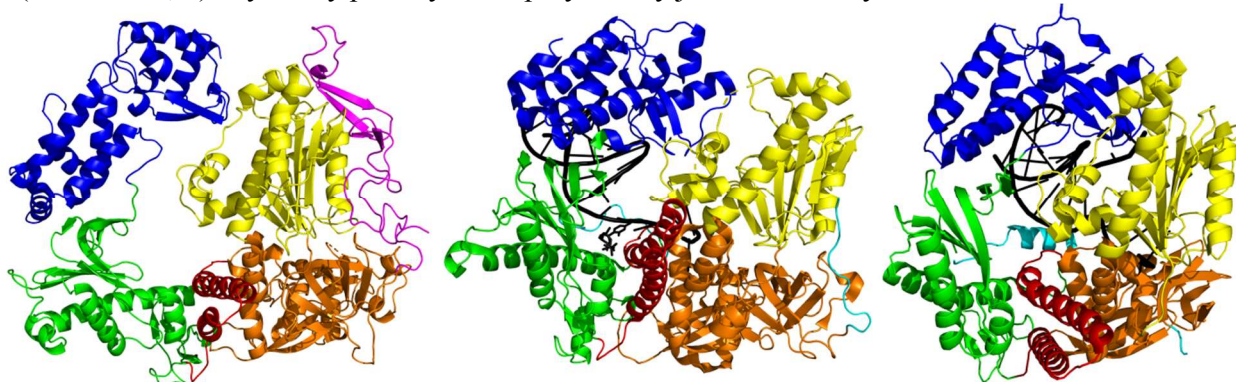
aminokyselinový řetězec doménou pre-NH<sub>2</sub>, popřípadě NH<sub>2</sub> doménou. Poté postupuje přes část NH<sub>2</sub> domény a celou exo doménu, aby se vrátil do zbytku domény NH<sub>2</sub>. Následně pokračuje řetězec přes dlouhý nestrukturovaný spoj do části domény palm. A poté přes domény finger a zbytek domény palm přechází do poslední domény thumb.

Tab. 3: Tabulka popisuje informace o výchozích strukturách polymeráz použitých pro vznik modelu polymerázy cytomegaloviru.[40] Všechny čísla, kromě číselných údajů o rozlišení a kódu v databázi UniProt, označují pořadí nebo počet reziduí (aminokyselin). Sekvence domén byla v případě polymerázy herpes simplex viru-1 a archebakterie vyjádřena ze zdrojových publikací.[33, 49] Sekvence domén polymerázy kvasinky byla odvozena pomocí srovnání sekvencí aminokyselin mezi polymerázou herpes simplex viru-1 a kvasinky.[41]

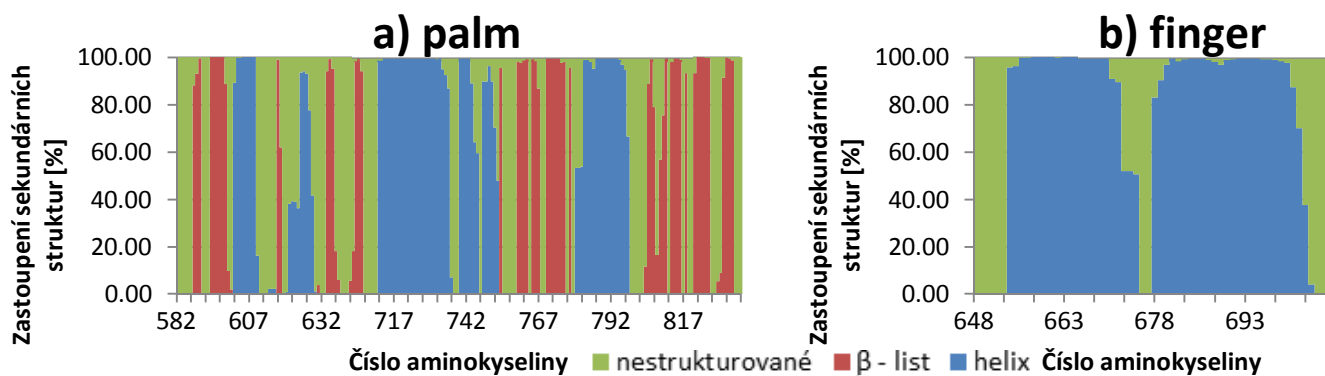
<b>Výchozí polymerázy - informace</b>			
informace	kód pdb		
	2gv9	3iay	4flw
organismus	herpes simplex virus-1	saccharomyces cerevisiae	pyrococcus abyssi
<b>typ organismu</b>	herpesvirus	kvasinka	archebakterie
<b>polymerázová rodina</b>	B	B	B
metoda zisku struktury	rentgenová difrakce	rentgenová difrakce	rentgenová difrakce
rozlišení [Å]	2,68	2	2,15
DNA ve struktuře	ne	ano	ano
forma polymerázy	-	polymerázová	exonukleázová
velikost proteinu (počet AK)	1235	1097	771
kód UniProt	P04292	P15436	P0CL77
další důležité části	-	DNA, DCP, 3xCa <sup>2+</sup>	DNA
části se známou strukturou (č. AK)	59-640, 701-1097, 1136-1197	95-490, 497-985	1-214, 216-387, 390-757
<b>Sekvence domén<sup>2</sup> se známou strukturou</b>			
pre-NH <sub>2</sub> (č. AK)	59-140	-	-
NH <sub>2</sub> (č. AK)	141-362, 594-639	107-315, 533-578	1-130, 327-369
exo (č. AK)	363-593	316-532	131-326
palm (č. AK)	701-766, 826-956	592-655, 716-836	390-445, 503-609
finger (č. AK)	767-825	656-715	446-502
thumb (č. AK)	957-1197	837-985	610-757

<sup>2</sup> Číselné označení sekvencí domén v této tabulce vychází z pdb souborů a jejich literárních zdrojů. Číslování sekvence domén a velikost domén se může v následujícím textu částečně lišit. Důvodem je upravování polymeráz v editačních programech a systém, v rámci něhož byly struktury polymeráz upravovány pro vstup do výpočetní simulace.

U všech tří polymeráz bylo vyzkoušeno, jestli je možná jejich kombinace s použitými silovými poli a nastavením. Proto byly dané struktury polymeráz upraveny takovým způsobem, že byly očištěny o vše kromě polypeptidů, DNA a DCP s jeho protiiionty. V případě existence více různě dlouhých segmentů stejných částí polymerázy (v krystalové struktuře, v pdb souboru) byly tyto části zkompletovány do co nejucelenější struktury. Na takto upravených strukturách byly vystavěny vstupy do simulací a spuštěna molekulová dynamika (nastavení viz oddíl 3.1). Výsledky z molekulové dynamiky jsou pro herpes simplex virus-1 znázorněny dále (viz obr. 8-a, b). Výsledky pro zbylé dvě polymerázy jsou znázorněny v Příloze A.

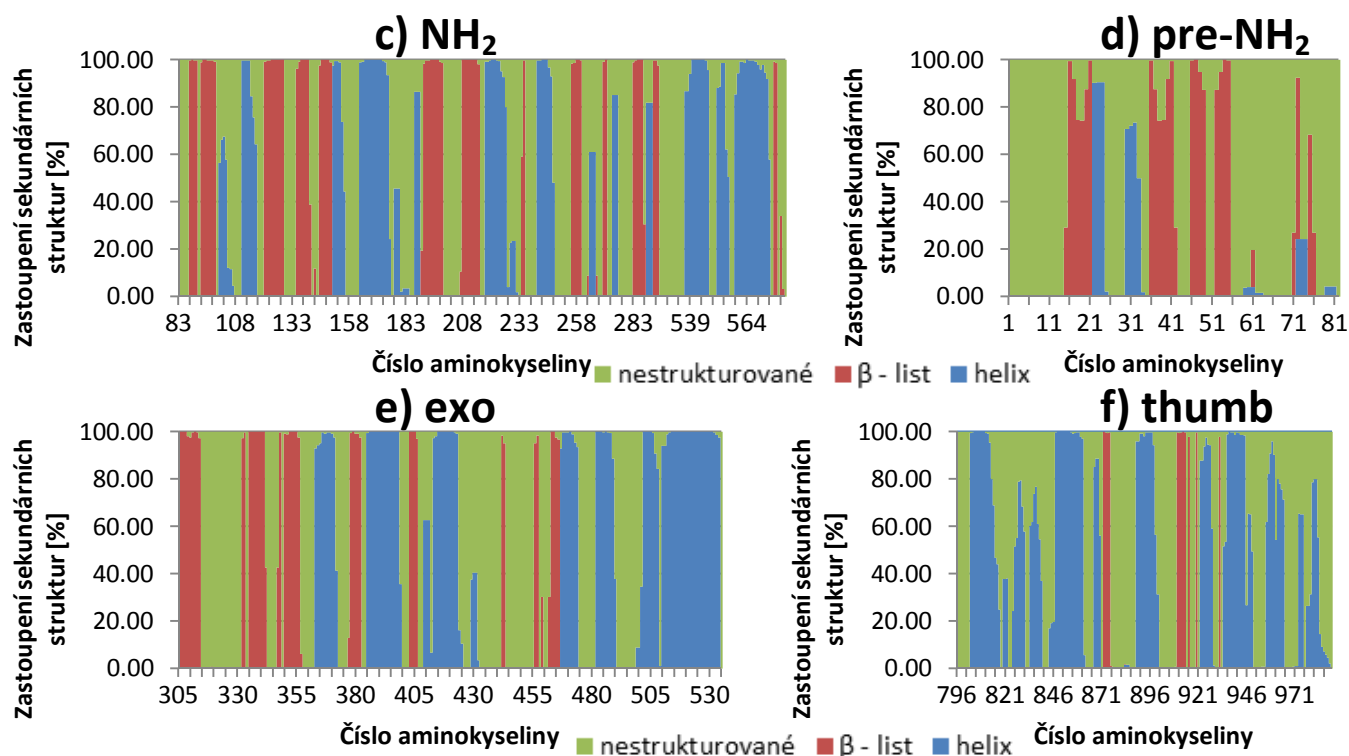


Obr. 7: Obrázky výchozích struktur polymeráz. Vlevo polymeráza herpes simplex viru-1, uprostřed polymeráza kvasinky a vpravo polymeráza archebakterie. Barvy fialová, oranžová, žlutá, zelená, červená a modrá znázorňují domény (v pořadí) pre-NH<sub>2</sub>, NH<sub>2</sub>, exo, palm, finger, thumb. Azurová barva znázorňuje část proteinu, jež není zařazena do žádné domény. Černá barva znázorňuje DNA, na prostředním obrázku i molekulu DCP.



Obr. 8-a: Grafy (a-b) sekundárních struktur pro polymerázu herpes simplex viru-1. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě  $\beta$ -listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální<sup>3</sup> zastoupení sekundárních struktur.

<sup>3</sup> Popisek svislé osy grafů sekundárních struktur je Zastoupení sekundárních struktur v procentech. Číselná hodnota procenta pro danou aminokyselinu určuje, jak velká část celkového času molekulové dynamiky byla konkrétní aminokyselina přítomna ve formě daného typu sekundární struktury. Např. sekundární struktura aminokyseliny může být z 85 % ve formě helixu a z 15 % v nestrukturované formě. Znamená to, že v 85 % celkového času molekulové dynamiky byla tato aminokyselina ve formě helixu a z 15 % celkového času ve formě nestrukturované. To znamená, že uvedené grafy sekundárních struktur nevyjadřují sekundární strukturu pro každou aminokyselinu v konkrétním čase molekulové dynamiky. Naopak, tyto grafy vyjadřují sumu daných typů sekundárních struktur (v procentech) přes celý čas molekulové dynamiky pro každou aminokyselinu.



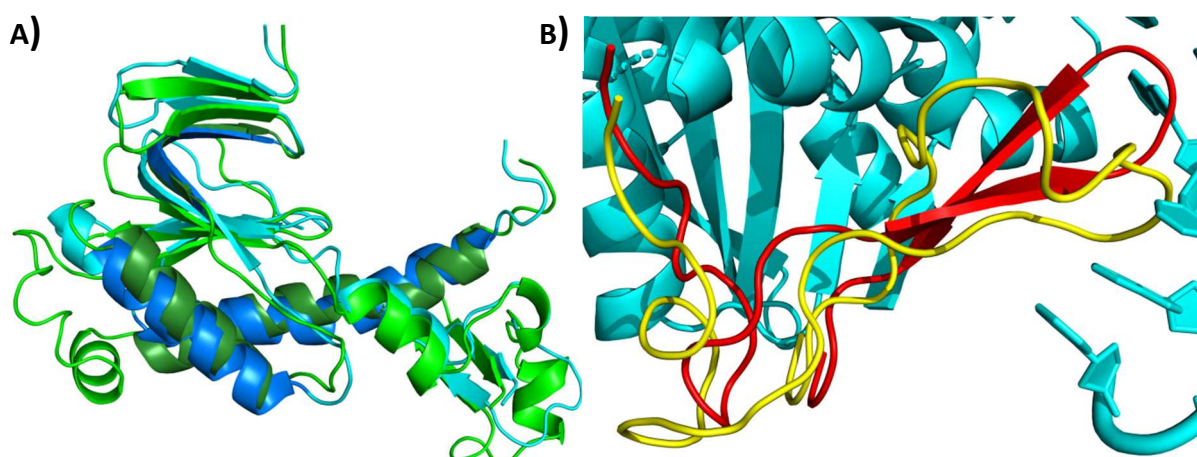
Obr. 8-b: Grafy (c-f) sekundárních struktur pro polymerázu herpes simplex viru-1. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě  $\beta$ -listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální zastoupení sekundárních struktur.

Z výsledků molekulových dynamik pro vstupní struktury vyplynulo, že je možné použít zvolené nastavení a silová pole pro výpočty v Amberu dohromady s našimi strukturami polymeráz. A tudíž je možné (z hlediska výpočetního prostředí) pokračovat ve výstavbě modelu polymerázy cytomegaloviru na základě těchto struktur. Zanalyzováním grafů sekundárních struktur všech tří polymeráz bylo zjištěno také, že v průběhu simulace jsou nejméně stále sekundární struktury v doméně thumb v polymeráze herpes simplex viru-1.

### 3.3 Příprava templátu pro výstavbu polymerázy cytomegaloviru v programu I-TASSER

Pomocí funkce pair\_fit v programu PyMol byla struktura polymerázy herpes simplex viru-1 nafitována (přes atomy páteře polypeptidu) na strukturu polymerázy kvasinky a archebakterie s přítomnou DNA. Struktura polymerázy herpes simplex viru-1 byla přetvářena do struktury polymeráz kvasinky a archebakterie postupně po doménách. Nebyla nafitována jako celek naráz, protože bylo zapotřebí změnit strukturu polymerázy herpes simplex viru-1 tak, aby zaujímala podobu domén polymeráz kvasinky a archebakterie kolem DNA. Zároveň bylo zapotřebí udržet strukturu domén, jako základních stabilních samostatných částí proteinu. Proto fitování zpravidla neprobíhalo po úsecích menších než doména. Příklad překlopení struktury polymerázy herpes simplex viru-1 do struktury polymerázy s DNA je znázorněn dále (viz obr. 9-A).

Příprava templátů pro výstavbu polymerázy cytomegaloviru nespočívala však pouze v nafitování polypeptidického řetězce. Do templátů byla umístěna také DNA. Templátu vystavěném na základě struktury polymerázy archebakterie byla ponechána  $\beta$ -smyčka v doméně *exo* z polymerázy archebakterie, protože struktura polymerázy herpes simplex viru-1 neměla tuto  $\beta$ -smyčku strukturovanou (viz obr. 9-B). Potřeba mít  $\beta$ -smyčku strukturovanou vychází z toho, že DNA v tomto templátu je v exonukleázové formě. A také, že  $\beta$ -smyčka hraje důležitou roli při interakci DNA a polymerázy během navazování primer DNA do aktivního místa exonukleázy.[50] V případě templátu vzniklého na základě struktury polymerázy kvasinky, měl výsledný templát obsaženou též molekulu DCP stabilizovanou třemi ionty  $Mg^{2+}$  (změna typu iontů oproti výchozí polymeráze). Tento templát neobsahuje strukturovanou  $\beta$ -smyčku v doméně *exo*. To ale nevadí. Je to totiž polymerázová forma polymerázy DNA a tudíž nepotřebuje strukturovanou  $\beta$ -smyčku v doméně *exo*.



Obr. 9: Obrázek A je schéma popisující nafitování domény *palm* polymerázy herpes simplex viru-1 na doménu *palm* polymerázy archebakterie. Zelené barvy označují doménu *palm* polymerázy herpes simplex viru-1. Modré barvy označují doménu *palm* polymerázy archebakterie. Tmavě zelená barva a tmavě modrá barva označují úseky polymeráz, přes které bylo provedeno fitování. Obrázek B znázorňuje DNA, část domény *exo* (azurová) a v ní část s  $\beta$ -smyčkou (červeně, žlutě). Žlutě je znázorněna nestrukturovaná část, pocházející ze struktury polymerázy herpes simplex viru-1. Červeně je znázorněna strukturovaná část pocházející ze struktury polymerázy archebakterie, jež byla použita pro přípravu templátu vystavěného na základě struktury polymerázy archebakterie.

Jelikož byla struktura polymerázy herpes simplex viru-1 v některých částech výrazněji upravována, bylo nutné předpřipravené templáty zbavit překryvů mezi jejich částmi. To bylo řešeno ruční úpravou polypeptidického řetězce, ale i funkcí *sculpting* v programu PyMol. Výsledkem bylo, že templáty už žádné překryvy mezi částmi polymerázy neobsahovaly.

Souhrnem k této kapitole bych chtěl pro přehlednost zdůraznit, že byly připraveny dva templáty pro program I-TASSER. Byly to polymerázy herpes simplex viru-1 ve dvou různých formách. Z důvodu neúplnosti struktury polymerázy herpes simplex viru-1 obsahovaly oba templáty v jejich polypeptidickém řetězci mezery. V obou případech chyběl řetězec aminokyselin mezi doménami  $NH_2$  a *palm*, střední část domény *thumb* a také zcela na začátku a na konci polypeptidického řetězce celé polymerázy. Všechny domény však byly aspoň z části přítomny.

### 3.4 Předpovězené struktury z programu I-TASSER

Dva templáty, jejichž příprava byla popsána v předchozí kapitole, byly použity v programu I-TASSER jako vzory pro návrh struktury polymerázy cytomegaloviru v polymerázové (1039 aminokyselin v templátu) a exonukleázové formě (1031 aminokyselin v templátu). V obou případech bylo v programu I-TASSER nastaveno vyloučení použití struktury 2gv9 (pdb kód) z návrhu polymerázy cytomegaloviru. Tato možnost byla zvolena, protože templáty jsou již její formou (formou struktury 2gv9). Jako sekvence, jejíž strukturu na základě templátů chceme předpovědět, byla zvolena sekvence 1242 aminokyselin odpovídající polymeráze cytomegaloviru (kód UniProt: P08546). Výsledky z návrhu struktury polymeráz jsou k dispozici dále (viz tab. 4-a, b, c).

*Tab. 4-a: Tabulka ukazuje deset nejvíce podobných templátů k sekvenci aminokyselin polymerázy cytomegaloviru použitých při návrhu modelů programem I-TASSER. Je rozdělena do dvou hlavních částí. Vlevo je část týkající se exonukleázové, vpravo polymerázové formy navrhovaného modelu. Kromě templátů, jež byly popsány v předchozí kapitole, jsou v tabulce uvedeny pdb kódy všech templátů. V tabulce jsou též uvedeny hodnoty Z-score týkající se kvality templátů.*

Deset nejvíce podobných templátů použitých při návrhu modelu							
exonukleázová forma				polymerázová forma			
pdb kód	Z-score	pdb kód	Z-score	pdb kód	Z-score	pdb kód	Z-score
templát	10,00	3iayA	3,15	templát	10,00	5mdnA	3,83
3iayA	4,59	3iayA	4,18	3iayA	3,32	3iayA	3,15
3iayA	3,32	3iayA	10,96	5mdnA	3,54	5mdnA	2,52
3iayA	6,92	3iayA	6,25	3iayA	3,05	3iayA	4,59
3iayA	8,10	3iayA	3,96	3iayA	6,92	3iayA	8,10

*Tab. 4-b: Tabulka ukazuje informace o navržených modelech polymeráz programem I-TASSER. Je rozdělena do dvou hlavních oblastí. Vlevo je oblast týkající se exonukleázové a vpravo polymerázové formy navrhovaného modelu. V tabulce je zobrazena informace o hodnotách C-score pro každý z pěti navržených modelů.*

Návrh struktur polymeráz dle I-TASSER							
exonukleázová forma				polymerázová forma			
Pět navržených modelů							
číslo modelu	C-score	číslo modelu	C-score	číslo modelu	C-score	číslo modelu	C-score
1.	-1,54	4.	-2,22	1.	-1,45	4.	-2,67
2.	-2,12	5.	-1,70	2.	-2,11	5.	-2,68
3.	-2,31	-	-	3.	-2,54	-	-

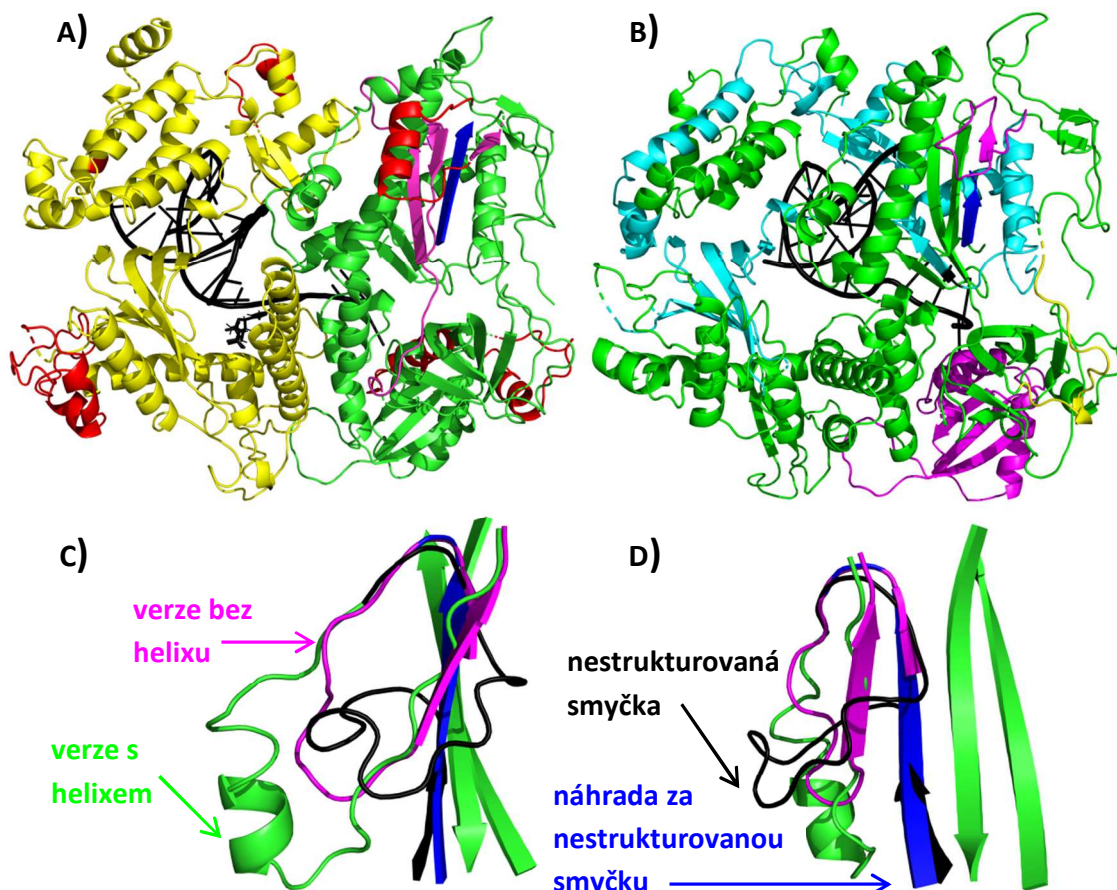
Tab. 4-c: Tabulka ukazuje informace o navržených modelech polymeráz programem I-TASSER. Je rozdělena do dvou hlavních oblastí. Vlevo je oblast týkající se exonukleázové a vpravo polymerázové formy navrhovaného modelu. V tabulce jsou informace (pdb kód, TM-score) o deseti strukturně nejvíce podobných proteinech k prvnímu modelu. TM-score je parametr udávající míru podobnosti proteinů.

Návrh struktur polymeráz dle I-TASSER							
exonukleázová forma				polymerázová forma			
Deset nejvíce podobných proteinů k prvnímu modelu							
pdb kód	TM-core	pdb kód	TM-core	pdb kód	TM-core	pdb kód	TM-core
3iayA	0,69	1ig9A	0,56	2gv9B	0,64	1qhtA	0,53
2gv9B	0,66	2vwkA	0,54	2vwjA	0,55	4q5vA	0,51
4m8oA	0,61	1d5aA	0,53	3iayA	0,55	1d5aA	0,50
4qclA	0,61	5mdnA	0,52	4m8oA	0,55	4fvmA	0,50
3k5aA	0,57	4fvmA	0,51	5mdnA	0,54	3k5aA	0,48

Z dat v tabulce 4-a vyplynulo, že nástroje I-TASSER nenašly množství různých proteinů sekvenčně podobných k sekvenci požadovaného modelu. Pro model exonukleázové formy to byl pouze jeden protein. Pro model polymerázové formy to byly pouze dva různé proteiny. Z toho je možné usoudit, že buď jako nejlepší možné vzory existují právě tyto dvě struktury (3iay, 5mdn). Nebo jiné struktury prostě neexistují. Z dat v tabulce 4-b vyplývá, že do pořadí modelů nepromluvily pouze hodnoty C-score, ale i velikost modelů. Obvykle by totiž měly být modely seřazené podle klesající kvality zároveň s klesajícím C-score. Což se nestalo. Údaje z deseti nejvíce podobných proteinů k modelu číslo jedna (viz tab. 4-c) potvrzují, že výslednému modelu jsou velmi podobné proteiny 2gv9, 3iay (vysoký TM-score).

Z modelů vystavených programem I-TASSER byly sestaveny celkem čtyři modely polymerázy cytomegaloviru. Po dvou v každé z forem (exonukleázová, polymerázová). Při sestavování zmíněných čtyř modelů bylo vycházeno z částí modelů (vystavených programem I-TASSER), jež si byly vzájemně strukturně podobné, a které byly více strukturované. Nebyly použity části modelů (vystavených programem I-TASSER), jež nebyly strukturované nebo by zavazely molekule DNA. Program I-TASSER navrhl i části polymerázy, jež byly vymodelované zcela náhodně. To platí pro části modelů polymerázy cytomegaloviru, pro něž v templátu vstupujícím do programu I-TASSER chyběly strukturní předlohy (tzn. v sekvenci polypeptidu polymerázy cytomegaloviru bylo více reziduí než v předloze templátu). Některé modely a jejich důležité rysy jsou zobrazeny dále (viz obr. 10).





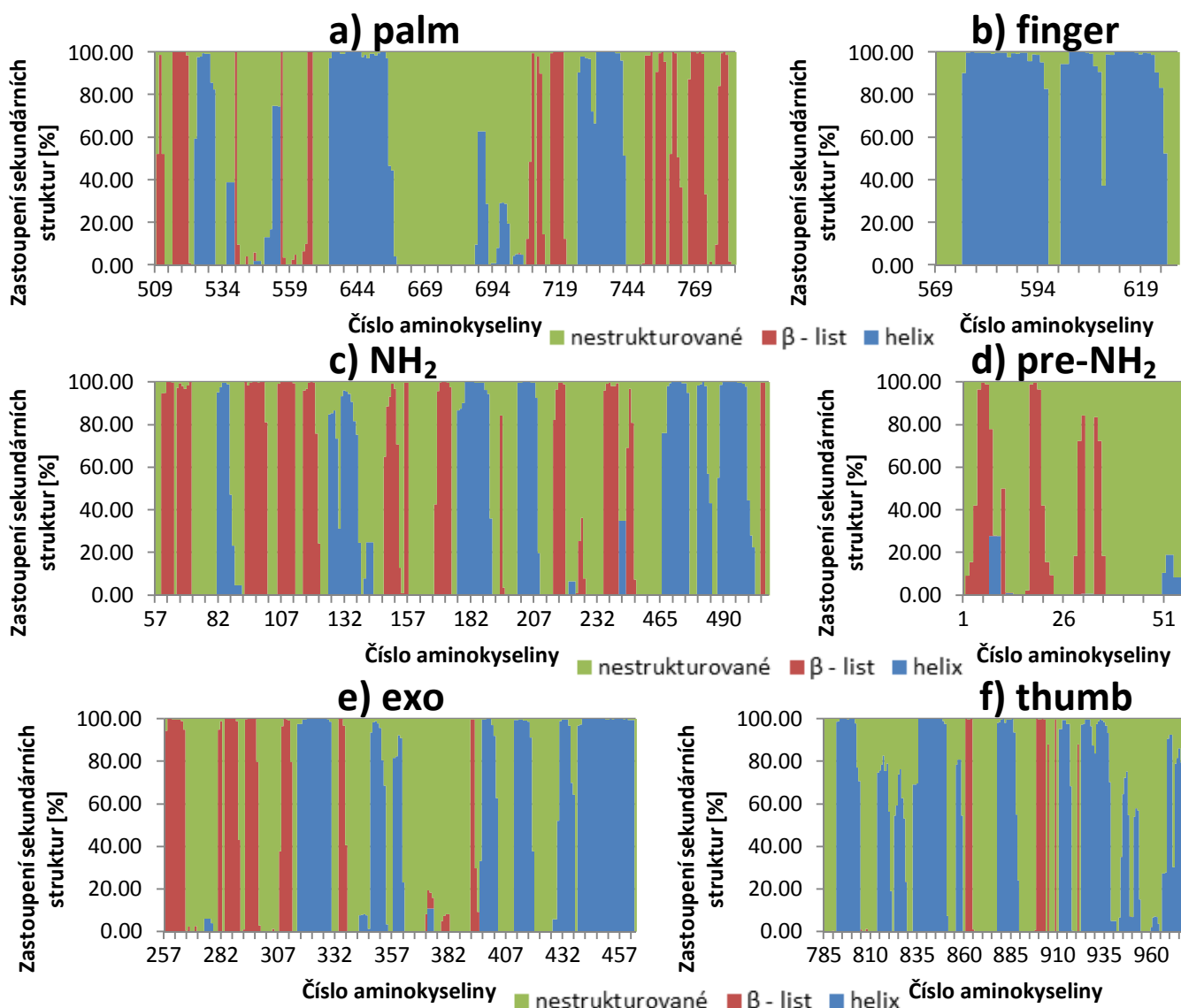
Obr. 10: Na obrázcích A, B jsou znázorněny hlavní dva modely postavené na základě výsledků z programu I-TASSER. Na obrázcích C, D je znázorněna část domény *exo* ze dvou různých pohledů. Na obrázku C lze vidět důvod, proč byly pro každý model (exonukleázová, polymerázová forma) polymerázy cytomegaloviru sestaveny dvě verze. Verze s helixem (vlevo, zeleně) a verze bez helixu (vpravo od helixu, růžově). Obrázek D představuje důvod, proč byla do všech čtyř modelů dodána strukturovaná část, jejíž předobraz pochází z polymerázy herpes simplex viru-1 (modře). Ve všech modelech vystavěných programem I-TASSER byla část odpovídající modrému  $\beta$ -listu totiž vystavěna s vloženou nestrukturovanou smyčkou (černě). Různé barvy označují části z různých modelů vystavěných programem I-TASSER.

Čtyři modely polymeráz cytomegaloviru byly sestaveny z různých modelů (viz obr. 10-A, B) předpovězených programem I-TASSER. Navíc obsahují část sestavenou přímo na základě části  $\beta$ -skládaného listu z *exo* domény polymerázy herpes simplex viru-1 (viz obr. 10-modrá barva). Přesto nebylo možné sestavit celý model polymerázy cytomegaloviru. Vymodelovat se dostatečně nepodařil začátek domény pre-NH<sub>2</sub>, dvě části polymerázy v doméně *exo*, spoj mezi doménami NH<sub>2</sub> a palm, střední část a konec domény thumb.

### 3.5 Optimalizace struktur polymerázy cytomegaloviru

Jelikož čtyři modely sestavené v předchozí kapitole jsou jen obrazem vzešlým na základě teoretického přístupu programu I-TASSER. Bylo nutné všechny modely optimalizovat. Během optimalizace bylo zjištěno, že některé konce polypeptidických řetězců nejsou nijak vázané na zbytek polymerázy. A že během molekulové dynamiky jsou tyto konce volně pohyblivé. Také bylo zjištěno, že některé části řetězců aminokyselin si během molekulové dynamiky neudrží svoji sekundární strukturu, jež se rozpadá. Z těchto důvodů byly některé konce řetězců zkráceny. A pro

vývoj v molekulové dynamice atomy některých reziduí pozičně uzamčeny. Jeden z výsledků (jeden ze čtyř modelů) takovéto molekulové dynamiky je ve formě grafů sekundárních struktur zobrazen dále (viz obr. 11). Výsledky molekulové dynamiky zbylých tří modelů jsou k dispozici v příloze B. To, které části polymerázy byly pozičně uzamčeny, je znázorněno obrázkem dále (viz obr. 12).



Obr. 11: Grafy (a-f) domén sekundárních struktur pro model polymerázy cytomegaloviru ve formě exonukleázové bez helixu. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě  $\beta$ -listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální zastoupení sekundárních struktur.



Obr. 12: Obrázek znázorňuje, které části polymerázy cytomegaloviru byly pozičně uzamčeny (černá) během molekulové dynamiky s cílem udržení sekundární struktury. Na obrázku je znázorněna struktura polymerázy cytomegaloviru v exonukleázové formě bez helixu. Pozičně uzamčené části polymerázy byly obdobné ve všech zmíněných čtyřech formách (viz výše). Barvy fialová, oranžová, žlutá, zelená, červená a modrá znázorňují domény (v pořadí) pre-NH<sub>2</sub>, NH<sub>2</sub>, exo, palm, finger a thumb. Azurová barva znázorňuje DNA.

Z obrázků jedenáct, dvanáct a přílohy B vyplývá následující. Sekundární struktura domén zůstala průběhem molekulové dynamiky většinou nezměněna. Problémy s rozpadem sekundární struktury použitím pozičního zámku nebyly vyřešeny. Výsledky však naznačují, že by mohly být vyloučeny polymerázy s helixem v části domény exo (viz obr. 10-c, d). Dle grafů sekundárních struktur totiž převažuje nestrukturovaná forma v této části polymerázy. Sekvence teoretického helixu navíc obsahuje glycin, který je možné považovat za aminokyselinu, která se oproti jiným aminokyselinám v helixech příliš často nevyskytuje.[51] Pro jistější rozhodnutí ohledně přítomnosti nebo nepřítomnosti sekundárních struktur (jež byla snaha udržet) v modelech, by však bylo vhodné vyzkoušet i simulaci ne s pozičním zámkem, ale se zamknutými vzdálenostmi mezi atomy, jež by odpovídaly charakteristickým vodíkovým vazbám držící sekundární strukturu pohromadě.

### 3.6 Výsledné modely polymerázy cytomegaloviru, možné pokračování

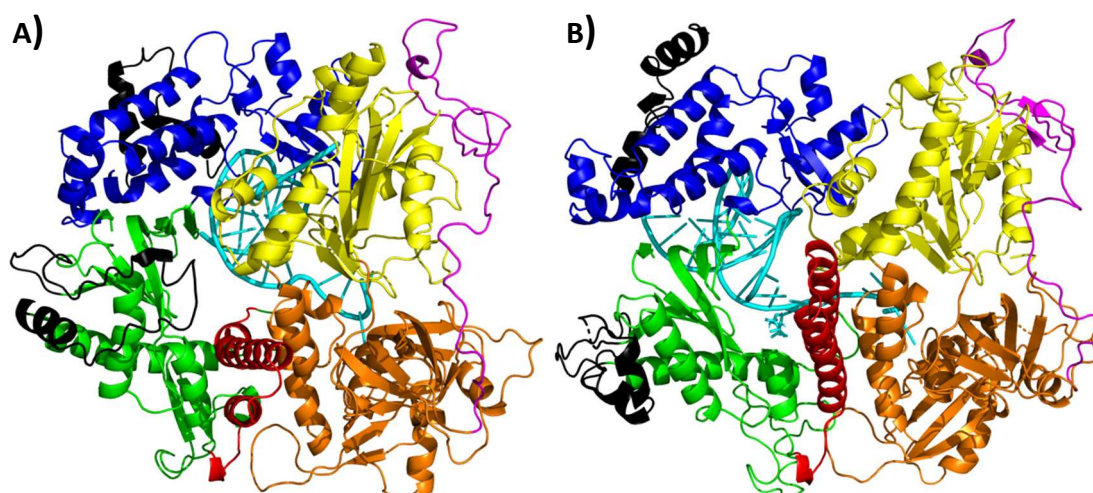
Výsledkem této práce jsou dva modely polymerázy cytomegaloviru. Oba dva modely nejsou vzhledem k sekvenci polymerázy cytomegaloviru úplné. Některé části polypeptidického řetězce chybí (viz tab. 5-a, b). Model polymerázy s DNA ve formě exonukleázové je znázorněn na obrázku (viz obr. 13-a). Model polymerázy s DNA ve formě polymerázové je znázorněn na obrázku (viz obr. 13-b).

Tab. 5-a: Tabulka ukazuje, z jakých částí se model polymerázy v exonukleázové formě skládá. Informace v tabulce jsou vztaženy ke každé doméně zvlášť. Informace se týkají modelu polymerázy

Části modelu polymerázy cytomegaloviru s DNA v exonukleázové formě			
celková délka polypeptidického řetězce polymerázy cytomegaloviru: 1242 reziduí			
domény	sekvence reziduí	domény	sekvence reziduí
per-NH <sub>2</sub>	40:95	palm	706:765, 825:980
NH <sub>2</sub>	96:296, 554:598	finger	766:824
exo	297:329, 360:438, 465:553	thumb	981:1104, 1171:1242

Tab. 5-b: Tabulka ukazuje, z jakých částí se model polymerázy skládá. Informace v tabulce jsou vztaženy ke každé doméně zvlášť. Informace se týkají modelu polymerázy v polymerázové formě.

Části modelu polymerázy cytomegaloviru s DNA v polymerázové formě			
celková délka polypeptidického řetězce polymerázy cytomegaloviru: 1242 reziduí			
domény	sekvence reziduí	domény	sekvence reziduí
per-NH <sub>2</sub>	40:95	palm	706:765, 825:980
NH <sub>2</sub>	96:296, 554:598	finger	766:824
exo	297:329, 360:438, 460:553	thumb	981:1104, 1171:1242



Obr. 13: Oba obrázky znázorňují model polymerázy cytomegaloviru. Obrázek A) polymerázu v exonukleázové formě. Obrázek B) v polymerázové formě. Barvy fialová, oranžová, žlutá, zelená, červená a modrá znázorňují domény (v pořadí) pre-NH<sub>2</sub>, NH<sub>2</sub>, exo, palm, finger a thumb. Azurová barva znázorňuje DNA, molekulu DCP. Černá barva znázorňuje části v doméně palm (45 reziduí) a thumb (36 reziduí), které jsou porovnáním mezi dvěma formami polymeráz nejvíce rozdílné.

Z tabulky 5 vyplývá, jak moc jsou modely polymerázy cytomegaloviru úplné. Model polymerázy cytomegaloviru s DNA v exonukleázové formě obsahuje 974 reziduí. To odpovídá výstavbě na úrovni 78,4 %. Model polymerázy cytomegaloviru s DNA v polymerázové formě obsahuje 979 reziduí. To odpovídá výstavbě modelu z 78,8 %. Může být tedy řečeno, že modely polymerázy byly vymodelovány z cca 78,6 %. To však nic neříká o kvalitě jednotlivých částí modelů. Obrázek 13 obsahuje černě vyznačené části polymeráz. Tyto části se mezi modely výrazně liší. To je

důsledkem toho, že tyto části byly předpovězeny bez vhodného templátu. V případě, že bychom tyto části modelů neuvažovali, musíme poupravit procenta, jež vyjadřují úplnost modelů. Úpravou zjistíme, že připravené modely by poté zaujímaly jen cca 71,2 % sekvence polymerázy cytomegaloviru.

### **Možné pokračování**

V této práci byly sestaveny modely polymerázy cytomegaloviru. Neznamená to však, že není možné dále pokračovat ve zlepšování kvality modelu. Je možné uvažovat o dvou základních cestách, kterými by bylo možné pokračovat. První cestou je snaha, která by vedla k nalezení jiných nebo dalších templátů. Databáze UniProt byla dosud prohledávána s cílem najít struktury, jež jsou podobné k sekvenci celé polymerázy cytomegaloviru. Tudiž by bylo možné onu databázi prohledávat i podle menších částí sekvence polymerázy cytomegaloviru (například podle částí velikosti domén). Byly vytvořeny dva modely. Proto by bylo možné využít část jednoho modelu ke zlepšení modelu druhého a naopak. Druhou cestou by bylo sestavení polymerázového hybridu. Části polymerázy, které nejsou v modelech přítomny, by teoreticky bylo možné nahradit. Použit by se k tomu mohly části z jiných polymeráz, jejichž strukturu známe. Tato známá struktura by se poté použila a zabudovala na prázdná místa modelů. Ověření, jenž by řeklo, jestli je tato náhrada možná, by bylo provedeno následujícím způsobem. Jelikož je možné syntetizovat polymerázu s libovolnou polypeptidickou sekvencí a taky odzkoušení, jak účinně polymeráza pracuje. Bylo by možné tímto způsobem ohodnotit, jaké hybridní polymerázy je možné sestavit.

#### 4. Závěr

Úkolem v této magisterské práci bylo sestavit model polymerázy cytomegaloviru. Důvodem bylo, že strukturu polymerázy neznáme, ale bylo by výhodné ji použít pro modelování mutací, jež způsobují rezistenci vůči některým antivirotikům. Model polymerázy byl sestaven ve dvou formách. Ve formě s DNA v exonukleázové poloze a s DNA v polymerázové poloze. Literatura o struktuře polymerázy cytomegaloviru uvádí, že je známa struktura pouze 22 aminokyselin na samotném konci polypeptidického řetězce polymerázy. Dále bylo zjištěno, že model polymerázy cytomegaloviru již modelován byl. Úplnost tohoto modelu však dosahovala cca jen 48,5 %.

Pro vymodelování struktury polymerázy cytomegaloviru (protein) bylo vycházeno z následujícího principu, který se týká struktury proteinu. Informace o prostorové struktuře proteinu je dána v samotné sekvenci polypeptidu. Díky tomuto principu bylo možné, pomocí známých struktur polymeráz se sekvencí polypeptidického řetězce podobného k sekvenci polymerázy cytomegaloviru, navrhnout požadovaný model. Jako polymerázy podobné k sekvenci polymerázy cytomegaloviru byly nalezeny a zvoleny následující struktury. Polymeráza herpes simplex viru-1. Polymeráza určitého typu kvasinky a archebakterie. Pomocí uvedených tří polymeráz a za využití programu PyMol a I-TASSER byly navrženy a v programu Amber simulovány dva modely polymerázy cytomegaloviru. Oba dva s přítomnou DNA ve dvou různých formách. První model s DNA v polymerázové formě. Druhý model s DNA v exonukleázové formě. Simulace modelů probíhaly na základě výpočtů založených na teorii molekulové mechaniky.

Vystavěné dva modely polymerázy nejsou úplné. Oba dva byly sestaveny z cca 78,6 % a obsahují části o různé kvalitě. Kroky, které měly na kvalitu modelů polymerázy kritický vliv, jsou následující. Zaprvé, volba vstupních polymeráz se známou strukturou, jež jsou sekvenčně nejpodobnější k sekvenci polymerázy cytomegaloviru. Zadruhé, způsob jakým byla struktura nejpodobnější polymerázy (pdb kód: 2gv9) nabitována na polymerázy obsahující DNA. Zatřetí, proces, jež vedl k odstranění nechtěných překryvů různých částí domén. Začtvrté, sestavení modelů z výstupů poskytnutých programem I-TASSER.

Jak již bylo řečeno, dva modely polymerázy cytomegaloviru byly sestaveny. Avšak možnost pokračovat v doplňování a zpřesňování modelů existuje. V úvahu připadá následující. Hledání dalších vhodných templátů, na jejichž základě by bylo možné části polymerázy vystavět. Dále je možná náhrada chybějících částí polymerázy strukturou z jiné polymerázy. Kdy jedna z kontrol takového hybridního modelu by proběhla na základě zkoušky aktivity tohoto uměle syntetizovaného modelu. Poslední možností je příprava vzorku polymerázy cytomegaloviru ve stavu, jenž by dovoľoval použití fyzikálně-chemických metod pro určení struktury polymerázy. Tento přístup by však byl spíše kontrolou k přístupu teoretickému, který byl použit v této práci.

## 5. Summary

My goal in this work was to build model of cytomegalovirus polymerase. Models with DNA were prepared in two modes. First model was prepared with DNA in the polymerizing mode. Second model was prepared with DNA in editing mode. The literature indicates that structure of 22 ends amino acids is known. It was also found that model of polymerase of cytomegalovirus was build already. But completeness of the model was 48,5 % only.

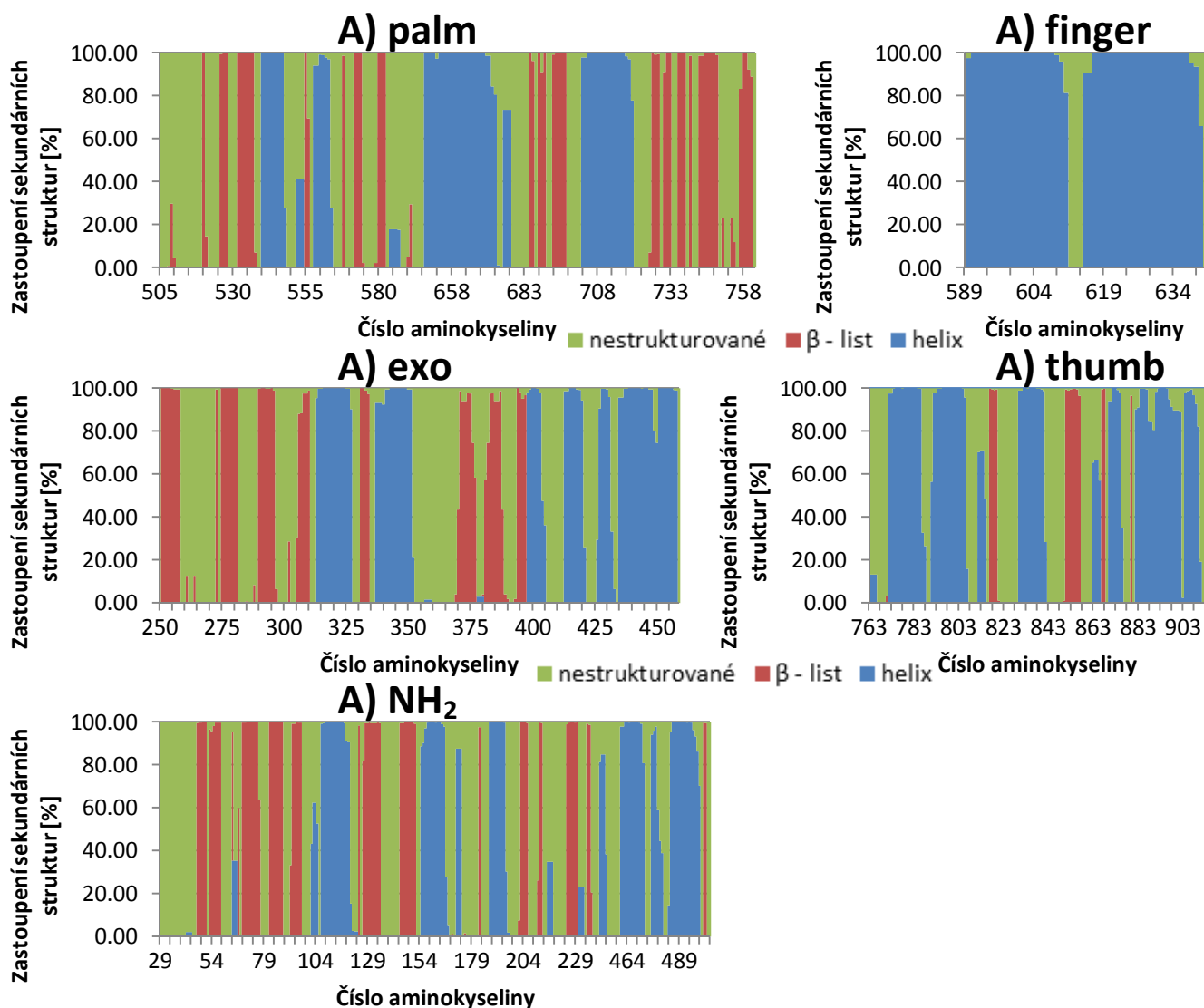
Building of polymerase models was based on the principle that protein structure is determined by the sequence of amino acids. It is reason why structures with similar amino acids sequence to sequence of polymerase of cytomegalovirus were used. Generating of polymerase models was based on three polymerases with known structure. These were polymerase of herpes simplex virus type-1 and concrete polymerases of yeast and archaebacterial. Structure models of cytomegalovirus polymerase were prepared using PyMol and I-TASSER programs. Amber program was used to computation simulations of models of polymerase. All calculations were performed using molecular mechanism theory.

Two prepared models of cytomegalovirus polymerase are not complete. Models were prepared from 78,6 % in comparison with sequence of cytomegalovirus polymerase. Critical moments of preparing models, which had an impact on quality of models are following. Firstly, selection of entry polymerases with known structure, which were sequentially similar to the sequence of cytomegalovirus polymerase. Secondly, the way in which the most similar polymerase structure (pdb code: 2gv9) was fitted to structure of polymerase containing DNA. Thirdly, the removal of unwanted overlaps between different parts of domains. Fourthly, generating models from I-TASSER outputs.

Two models of cytomegalovirus polymerase were built. That does not mean we can not continue in generating. We can look for other possible templates on the level of domain to generate the missing parts of the polymerase. Missing parts of models can be replaced by structure of different polymerase. The resulting hybrid then could be evaluated by its polymerase activity. Alternatively, we can prepare a sample of cytomegalovirus polymerase in a state that allows the usage of physicochemical methods to determine structure of the polymerase. However, this approach would rather was used to control the theoretical approach used in this work.

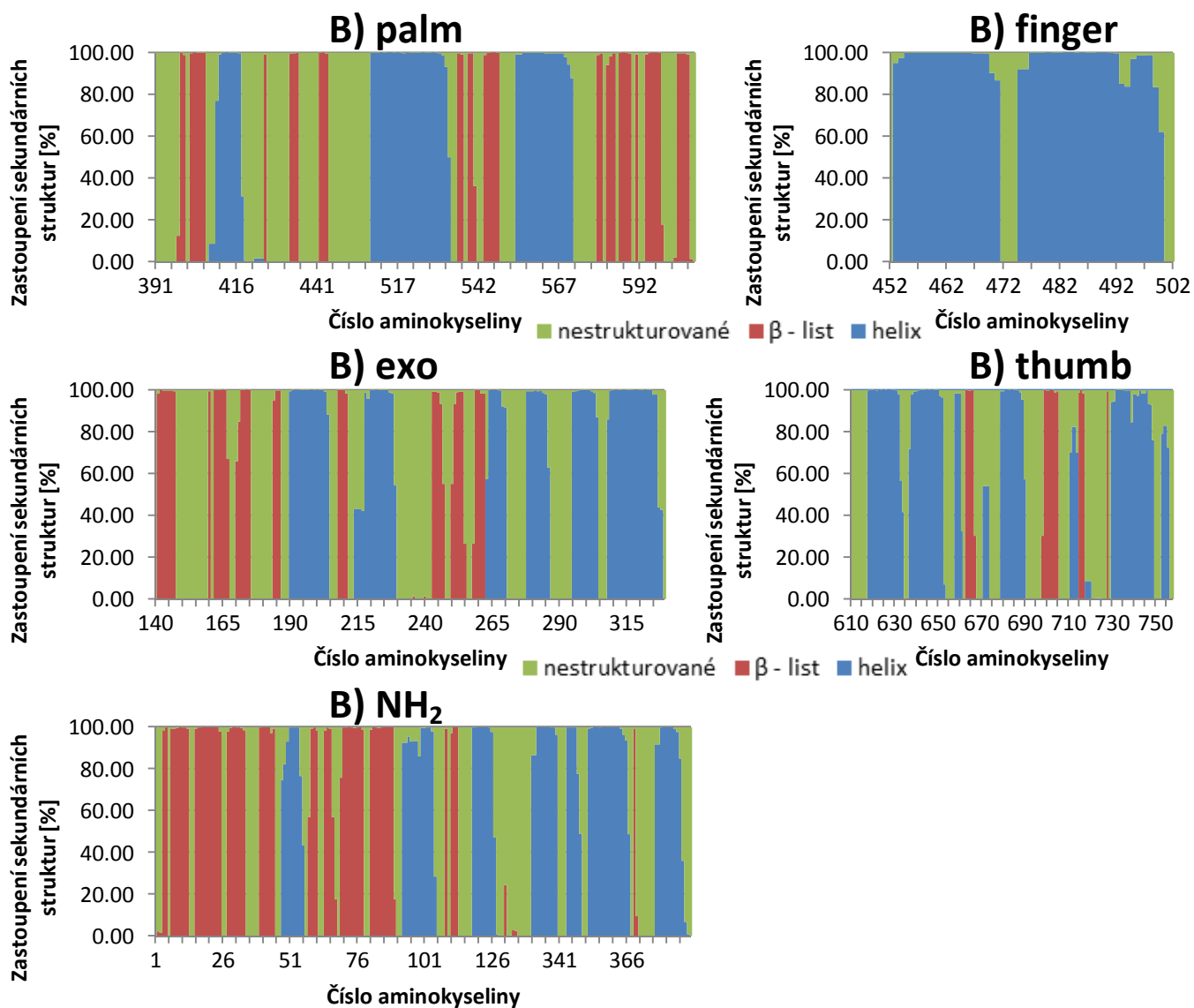
## 6. Příloha

**Příloha A:** Výsledky molekulové dynamiky pro struktury vstupních polymeráz kvasinky (A) a archebakterie (B).



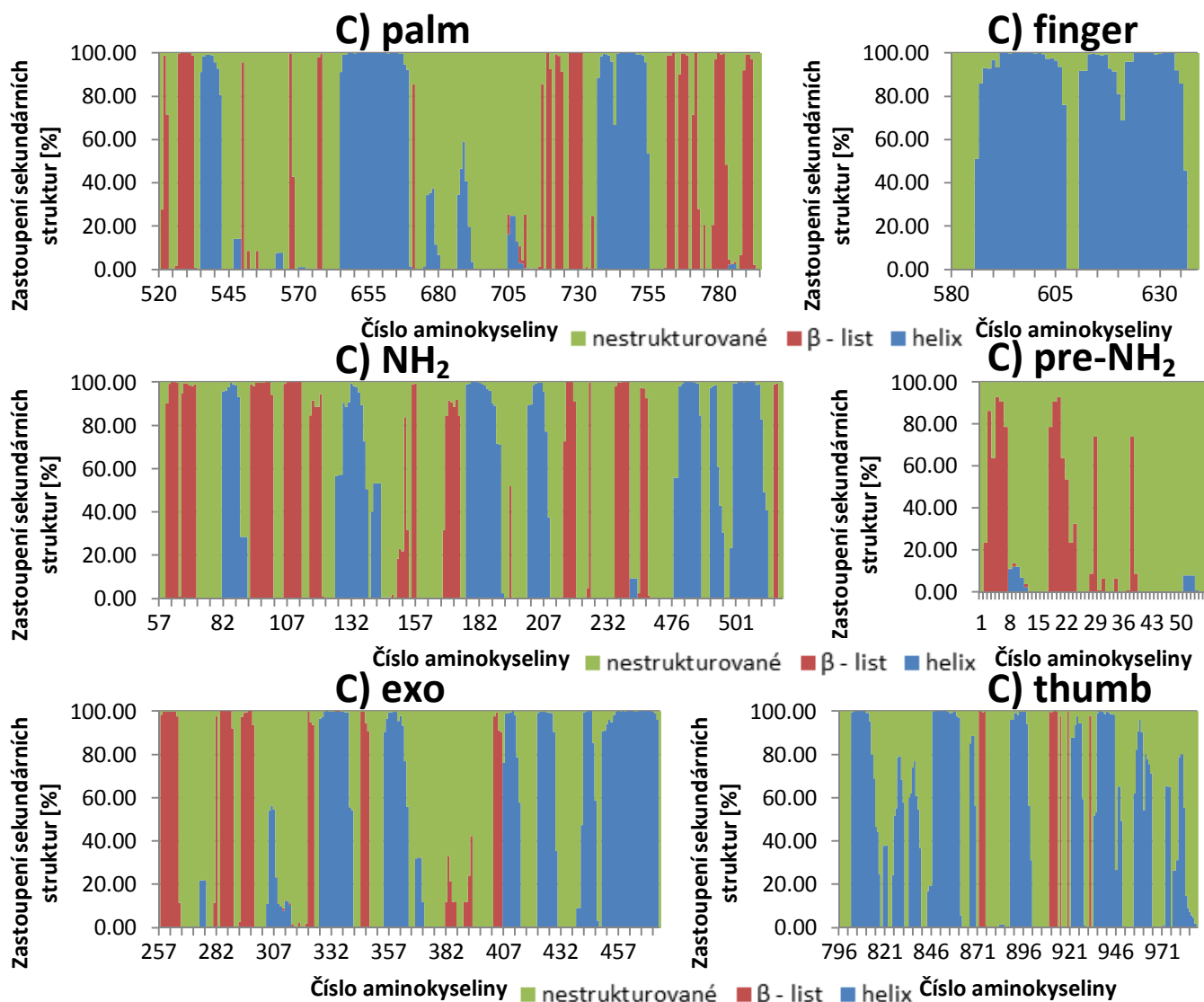
Obr. 14: Grafy sekundárních struktur domén polymerázy kvasinky. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě  $\beta$ -listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální zastoupení sekundárních struktur.



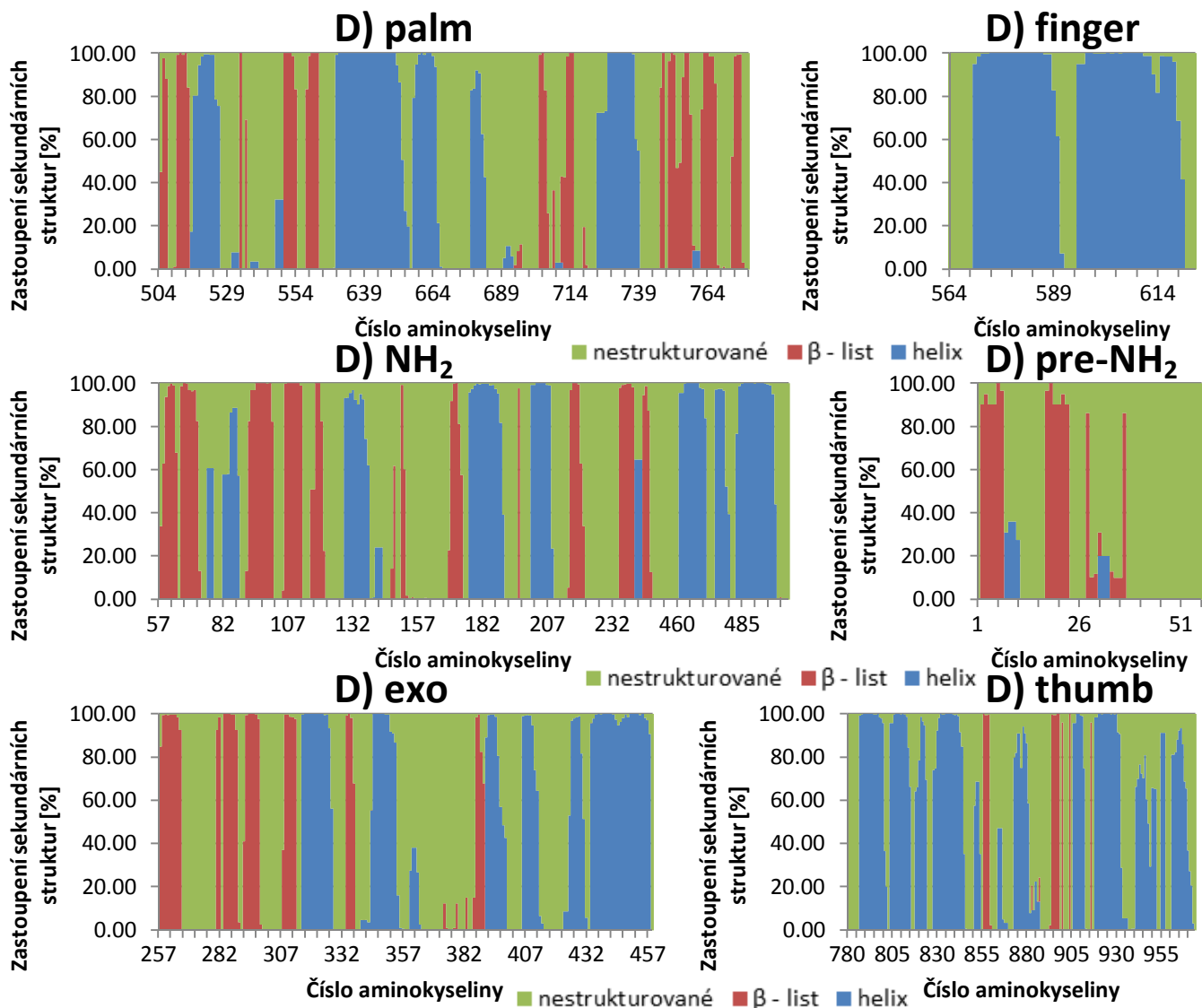


Obr. 15: Grafy sekundárních struktur domén polymerázy archebakterie. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě β-listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální zastoupení sekundárních struktur.

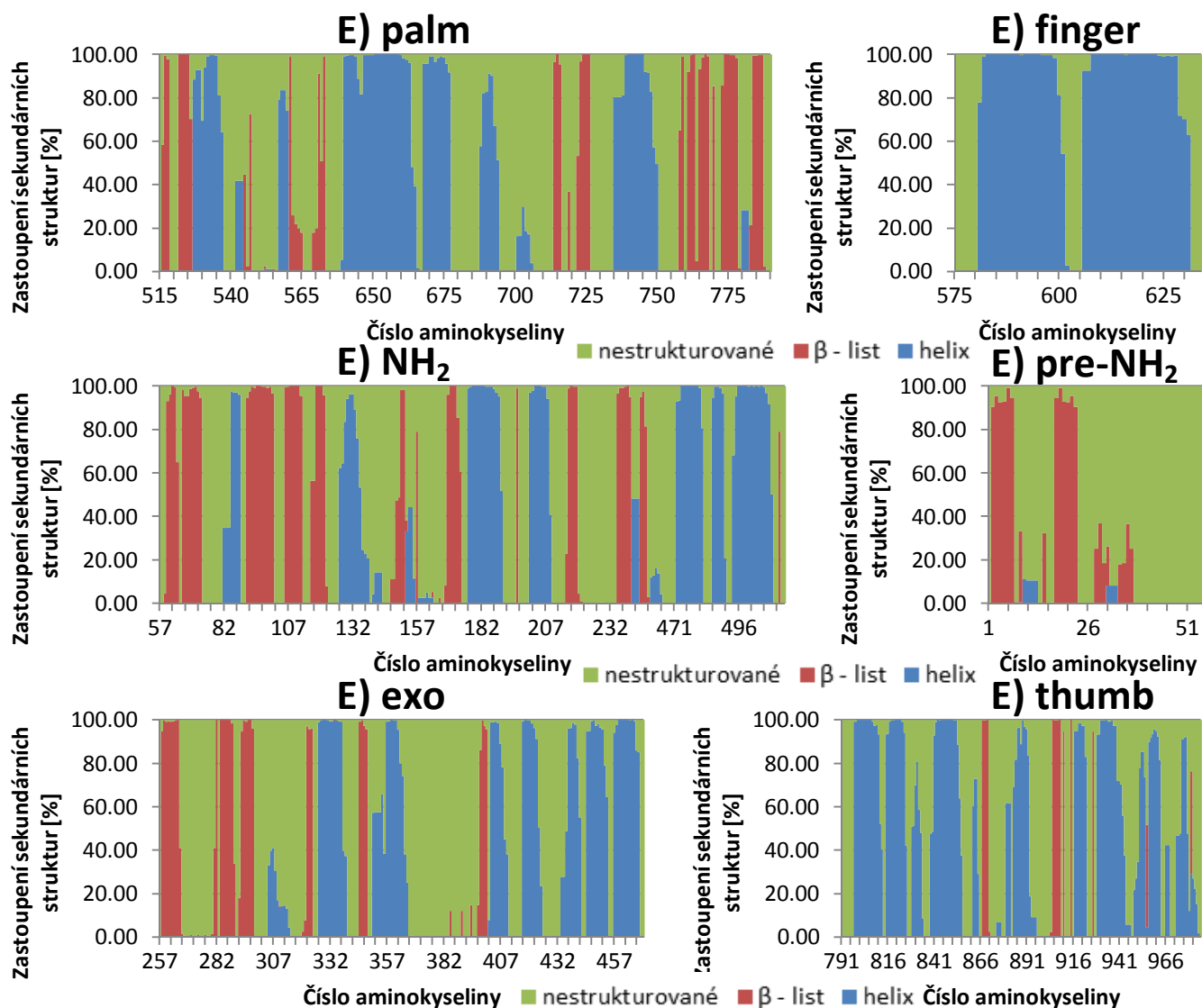
**Příloha B:** Výsledky molekulové dynamiky pro struktury polymerázy cytomegaloviru ve formě exonukleázové s helixem (C), ve formě polymerázové bez helixu (D), ve formě polymerázové s helixem (E).



Obr. 16: Grafy domén sekundárních struktur pro model polymerázy cytomegaloviru ve formě exonukleázové s helixem. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě  $\beta$ -listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální zastoupení sekundárních struktur.



Obr. 17: Grafy domén sekundárních struktur pro model polymerázy cytomegaloviru ve formě polymerázové bez helixu. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě  $\beta$ -listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální zastoupení sekundárních struktur.



Obr. 18: Grafy domén sekundárních struktur pro model polymerázy cytomegaloviru ve formě polymerázové s helixem. Sekundární struktury jsou barevně rozlišeny. Zelená znamená část řetězce bez sekundární struktury. Červená znamená řetězec ve formě  $\beta$ -listu. Modrá znamená řetězec ve formě helixu. Na vodorovné ose grafů je znázorněno číslo aminokyseliny. Svislá osa grafů popisuje procentuální zastoupení sekundárních struktur.

## 7. Seznam použité literatury

- 1 M. MACH, T. STAMMINGER, G. JAHN, Human Cytomegalovirus: Recent Aspects from Molecular Biology, *J. gen. Virol.*, 1989, 70, 3117-3146
- 2 M. K. GANDHI, R. KHANNA, Human cytomegalovirus: clinical aspects, immune regulation, and emerging treatments, *Lancet Infect. Dis.*, 2004, 4, 725-738
- 3 K. J. RYAN, C. G. RAY, J. C. SHERRIS, *Sherris medical microbiology: an introduction to infectious diseases. 4th edition*, New York, McGraw-Hill, 2004
- 4 A. AHMED, Antiviral Treatment of Cytomegalovirus Infection, *Infect. Disord. Drug Targets*, 2011, 11, 475-503
- 5 K. K. BIRON, Antiviral drugs for cytomegalovirus diseases, *Antiviral Res.*, 2006, 71, 154-163
- 6 M. MCINTOSH, B. HAUSCHILD, V. MILLER, Human cytomegalovirus and transplantation: drug development and regulatory issues, *J. Virus Erad.*, 2016, 2, 143-148
- 7 A. P. LIMAYE, L. COREY, D. M. KOELLE, C. L. DAVIS, M. BOECKH, Emergence of ganciclovir-resistant cytomegalovirus disease among recipients of solid-organ transplants, *The Lancet*, 2000, 356, 645-649
- 8 C. GILBERT, G. BOIVIN, Human Cytomegalovirus Resistance to Antiviral Drugs, *Antimicrob. Agents and Chemother.*, 2005, 49, 873-883
- 9 B. A. APPLETON, J. BROOKS, A. LOREGIAN, D. J. FILMAN, D. M. COEN, J. M. HOGLE, Crystal structure of the cytomegalovirus DNA polymerase subunit UL44 in complex with the C terminus from the catalytic subunit. Differences in structure and function relative to unliganded UL44, *J. Biol. Chem.*, 2006, 281, 5224-5232
- 10 F. JENSEN, *Introduction to computational chemistry. 2nd edition*, Chichester, Hoboken, New Jersey, Wiley, 2007
- 11 N. L. ALLINGER, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing  $V_1$  and  $V_2$  torsional terms, *J. Am. Chem. Soc.*, 1977, 99 (25), 8127-8134
- 12 C. J. CRAMER, *Essentials of computational chemistry: theories and models*, Chichester, Hoboken, New Jersey, Wiley, 2004
- 13 D. C. YOUNG, *Computational chemistry: a practical guide for applying techniques to real-world problems*, New York, Wiley, 2001
- 14 R. C. BERNARDI, M. C. R. MELO, K. SCHULTEN, Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems, *Biochim. Biophys. Acta*, 2015, 1850, 872-877
- 15 J. D. WATSON, A. GANN, T. A. BAKER, M. LEVINE, S. P. BELL, R. LOSICK, S. C. HARRISON, *Molecular biology of the gene. Seventh edition*, Boston, Pearson, 2014
- 16 G. PETSKO, D. RINGE, *Protein structure and function*, London, New Science Press, 2004
- 17 M. KODÍČEK, V. KARPENKO, *Biofyzikální chemie*, Praha, Česká republika, Academia, 2013
- 18 C. B. ANFINSEN, Principles that Govern the Folding of Protein Chains, *Science*, 1973, 181, 223-230
- 19 A. M. LESK, *Introduction to bioinformatics*, Oxford, Oxford University Press, 2002
- 20 A. FERSHT, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, New York, W.H. Freeman, 1998

- 21 P. GRIFFITHS, I. BARANIAK, M. REEVES, The pathogenesis of human cytomegalovirus, *J. Pathol.*, 2015, 235, 288-297
- 22 S. CHOU, Antiviral drug resistance in human cytomegalovirus, *Transpl. Infect. Dis.*, 1999, 1, 105-114
- 23 R. K. MURRAY, D. A. BENDER, K. M. BOTHAM, P.J. KENNELLY, V. W. RODWELL, P. A. WEIL, *Harperova ilustrovaná biochemie*, Praha, Česká republika, Galén, 2012
- 24 S. WU, W. A. BEARD, L. G. PEDERSEN, S. H. WILSON, Structural Comparison of DNA Polymerase Architecture Suggest a Nucleotide Gateway to the Polymerase Active Site, *Chem. Rev.*, 2014, 114(5), 2759-2774
- 25 M. GARCIA-DIAZ, K. BEBENEK, Multiple functions of DNA polymerases, *Critical Reviews in Plant Sciences*, 2007, 26(2): 105-122
- 26 J. FILEE, P. FORTERRE, T. SEN-LIN, J. LAURENT, Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins, *J. Mol. Evol.*, 2002, 54, 763-773
- 27 M. L. DEPAMPHILIS, *DNA replication in eukaryotic cells*, NY, Cold Spring Harbor Laboratory Press, 1996
- 28 H. M. EUN, *Enzymology primer for recombinant DNA technology*, San Diego, Academic Press, 1996
- 29 D. SOFROVÁ a kolektiv autorů, *Biochemie - Základní kurz*, Praha, Karolinum, 2009
- 30 K. ZARROUK, J. PIRET, G. BOIVIN, Herpesvirus DNA polymerases: Structures, functions and inhibitors, *Virus Research*, 2017, 234, 177-192
- 31 J. WANG, A. K. M. A. SATTAR, C. C. WANG, J. D. KARAM, W. H. KONIGSBERG, T. A. STEITZ, Crystal Structure of a pol  $\alpha$  Family Replication DNA Polymerase from Bacteriophage RB69, *Cell*, 1997, 89, 1087-1099
- 32 T. A. STEITZ, Structural biology: A mechanism for all polymerases, *Nature*, 1998, 391, 231-232
- 33 S. LIU, J. D. KNAPFELS, J. S. CHANG, G. A. WASZAK, E. T. BALDWIN, M. R. DEIBEL JR., D. R. THOMSEN, F. L. HOMA, P. A. WELLS, M. C. TORY, R. A. POORMAN, H. GAO, X. QIU, A. P. SEDDON, Crystal Structure of the Herpes Simplex Virus 1 DNA Polymerase, *J. Biol. Chem.*, 2006, 281(26), 18193-18200
- 34 I. R. LEHMAN, P. E. BOEHMER, Replication of Herpes Simplex Virus DNA, *J. Biol. Chem.*, 1999, 274, 28059-28062
- 35 R. SHI, A. AZZI, C. GILBERT, G. BOIVIN, S. LIN, Three-dimensional modeling of cytomegalovirus DNA polymerase and preliminary analysis of drug resistance, *Proteins*, 2006, 64, 301-307
- 36 The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC
- 37 D.A. CASE, D.S. CERUTTI, T.E. CHEATHAM, III, T.A. DARDEN, R.E. DUKE, T.J. GIESE, H. GOHLKE, A.W. GOETZ, D. GREENE, N. HOMEYER, S. IZADI, A. KOVALENKO, T.S. LEE, S. LEGRAND, P. LI, C. LIN, J. LIU, T. LUCHKO, R. LUO, D. MERMELSTEIN, K.M. MERZ, G. MONARD, H. NGUYEN, I. OMELIAN, A. ONUFRIEV, F. PAN, R. QI, D.R. ROE, A. ROITBERG, C. SAGUI, C.L. SIMMERLING, W.M. BOTELLO-SMITH, J. SWAILS, R.C. WALKER, J. WANG, R.M. WOLF, X. WU, L. XIAO, D.M. YORK, P.A. KOLLMAN, *AMBER 2017*, University of California, San Francisco, 2017

- 38 J. R. R. TOLKIEN, *Silmarillion: Mýty a legendy Středozeemě*, Praha, Česká republika, ARGO, 2008
- 39 W. HUMPHREY, A. DALKE, K. SCHULTEN, VMD: Visual Molecular Dynamics, *J. Molec. Graphics*, 1996, 14, 33-38
- 40 P. W. ROSE, A. PRLIĆ, A. ALTUNKAYA, C. BI, A. R. BRADLEY, C. H. CHRISTIE, L. D. COSTANZO, J. M. DUARTE, S. DUTTA, Z. FENG, R. KRAMER GREEN, D. S. GOODSSELL, B. HUDSON, T. KALRO, R. LOWE, E. PEISACH, C. RANDLE, A. S. ROSE, C. SHAO, Y.-P. TAO, Y. VALASATAVA, M. VOIGT, J. D. WESTBROOK, J. WOO, H. YANG, J. Y. YOUNG, C. ZARDECKI, H. M. BERMAN, S. K. BURLEY, The RCSB protein data bank: integrative view of protein, gene and 3D structural information, *Nucleic Acids Res.*, 2017, 45, D271-D281
- 41 THE UNIPROT CONSORTIUM, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.*, 2017, 45, D158-D169
- 42 J. YANG, R. YAN, A. ROY, D. XU, J. POISSON, Y. ZHANG, The I-TASSER Suite: Protein structure and function prediction, *Nature Methods*, 2015, 12, 7-8
- 43 A. ROY, A. KUCUKURAL, Y. ZHANG, I-TASSER: a unified platform for automated protein structure and function prediction, *Nature Protocols*, 2010, 5, 725-738
- 44 Y. ZHANG, I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics*, 2008, 9
- 45 M. ZGARBOVÁ, J. ŠPONER, M. OTYEPKA, T. E. CHEATHAM, R. GALINDO-MURILLO, P. JUREČKA, Refinement of the Sugar-Phosphate Backbone Torsion Beta for the AMBER Force Fields Improves the Description of Z-DNA and B-DNA, *J. Chem. Theory Comput.*, 2015, 11, 5723-5736
- 46 J. A. MAIER, C. MARTINEZ, K. KASAVAJHALA, L. WICKSTROM, K. E. HAUSER, C. SIMMERLING, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.*, 2015, 11, 3696-3713
- 47 C. W. HOPKINS S. LE GRAND, R. C. WALKER, A. E. ROITBERG, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, *J. Chem. Theory Comput.*, 2015, 11, 1864-1874
- 48 M. K. SWAN, R. E. JOHNSON, L. PRAKASH, S. PRAKASH, A. K. AGGARWAL, Structural basis of high fidelity DNA synthesis by yeast DNA polymerase delta, *Nat. Struct. Mol. Biol.*, 2009, 16, 979-986
- 49 J. GOUGE, C. RALEC, G. HENNEKE, M. DELARUE, Molecular Recognition of Canonical and Deaminated Bases by P. abyssi Family B DNA Polymerase, *J. Mol. Biol.*, 2012, 423, 315-336
- 50 M. HOGG, P. ALLER, W. KONIGSBERG, S. S. WALLACE, S. DOUBLIE, Structural and Biochemical Investigation of the Role in Proofreading of a  $\beta$  Hairpin Loop Found in the Exonuclease Domain of a Replicative DNA Polymerase of the B Family, *J. Biol. Chem.*, 2007, 282, 1432-1444
- 51 Y. OKAMOTO, Helix-Forming Tendencies of Nonpolar Amino Acids Predicted by Monte Carlo Simulated Annealing, *Proteins*, 1994, 19, 14-23