

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
ZDRAVOTNĚ SOCIÁLNÍ FAKULTA

Bakalářská práce

Software pro molekulární dynamiku

Autor: Kristýna Šilhavá

Vedoucí práce: RNDr. Milan Předota, Ph.D.

2011

Abstract

The software for molecular dynamics provides the ability of simulating systems that consist of millions of particles, just for a few nanoseconds. This method allows for the carrying out of virtual experiments without real laboratory equipment and only with computer technology. Molecular dynamics play an important role in understanding the structure and function of biological, organic and inorganic systems.

The objective of this thesis was to introduce the basics of molecular dynamics, solve model tasks in Amber software and Gromacs and to compare their capabilities.

I have worked on the Hermes Computer Cluster that belongs to the Faculty of Science at the University of South Bohemia and this was part of the MetaCentrum Project. I have worked with both programs using the command line of the Linux Operating System.

This thesis describes all of the file formats that are needed for working with these programs, and how they are to be used. I have built up several systems in Gromacs and Amber: 1) water 2) water and ions 3) solvated lysozyme. Based upon these sample systems, I have compared the basic utilities for preparation, the execution and the analysis of these simulations. This thesis should serve as a quick introduction and familiarization with Amber software and Gromacs.

Prohlašuji, že svoji bakalářskou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

.....

datum

.....

podpis studenta

Chtěla bych tímto poděkovat vedoucímu práce RNDr. Milanu Předotovi, Ph.D. za pomoc a veškerý čas věnovaný konzultacím. Také bych chtěla poděkovat Ing. Ondřeji Kroutilovi za rady a připomínky k dané tématice. V neposlední řadě patří poděkování MetaCentru.

OBSAH:

ÚVOD.....	8
1. SOUČASNÝ STAV.....	9
1.1. Molekulární mechanika.....	9
1.1.1. <i>Silové pole a potenciálová funkce</i>	10
1.1.2. <i>Potenciály</i>	11
1.1.2.1. <i>Potenciály vazebné</i>	11
1.1.2.2. <i>Potenciály nevazebné</i>	12
1.1.3. <i>Parametrizace, typy silových polí</i>	13
1.2. Integrace pohybových rovnic.....	14
1.2.1. <i>Verletův algoritmus</i>	14
1.3. Technické detaily.....	15
1.3.1. <i>Periodické okrajové podmínky</i>	15
1.3.2. <i>Ořznutí interakčního potenciálu (cutoff)</i>	16
1.3.3. <i>Explicitní a implicitní solvatace</i>	16
1.3.4. <i>Termostatování a barostatování</i>	17
1.3.4.1. <i>Kanonický soubor (NVT)</i>	17
1.3.4.2. <i>Izotermicko-izobarický soubor (NPT)</i>	17
1.4. Software.....	18
1.4.1. <i>Amber a Gromacs</i>	18
2. CÍLE PRÁCE A HYPOTÉZY.....	20
3. METODIKA.....	21
4. VÝSLEDKY.....	22
4.1. Základní schéma MD a práce se soubory.....	22
4.1.1. <i>Počáteční konfigurace</i>	22
4.1.2. <i>Generování topologie, výběr silového pole a parametrů</i>	23
4.1.3. <i>Solvatace</i>	24
4.1.4. <i>Přidání iontů</i>	25

4.1.5.	<i>Minimalizace energie</i>	25
4.1.6.	<i>Ekvilibrace</i>	26
4.1.7.	<i>Produkční fáze</i>	27
4.1.8.	<i>Analýza</i>	28
4.2.	Formáty vstupních a výstupních souborů	31
4.2.1.	<i>Vstupní souřadnice</i>	31
4.2.1.1.	<i>Formát (.pdb)</i>	31
4.2.1.2.	<i>Formát (.gro)</i>	31
4.2.1.3.	<i>Formát (.inpcrd)</i>	32
4.2.1.4.	<i>Formát (.mol2)</i>	32
4.2.2.	<i>Topologický soubor</i>	33
4.2.2.1.	<i>Gromacsovské topologie</i>	33
4.2.2.2.	<i>Amberovské topologie</i>	38
4.2.2.3.	<i>Převaděč topologie a souřadnic Amber → Gromacs</i>	41
4.2.3.	<i>Parametrické soubory</i>	41
4.2.3.1.	<i>Parametry pro Gromacs (.mdp)</i>	41
4.2.3.2.	<i>Parametry pro Amber (.in)</i>	45
4.3.	Práce s utilitami	48
4.3.1.	<i>Gromacs</i>	48
4.3.1.1.	<i>Vybrané utility pro přípravu systému</i>	49
4.3.1.2.	<i>Výpočetní utilita</i>	52
4.3.1.3.	<i>Vybrané utility pro analýzu</i>	52
4.3.2.	<i>Amber</i>	54
4.3.2.1.	<i>Vybrané utility pro přípravu systému</i>	54
4.3.2.2.	<i>Výpočetní utility</i>	56
4.3.2.3.	<i>Vybrané utility pro analýzu</i>	57
4.4.	Modelové úlohy	58
4.4.1.	<i>Simulace boxu s vodou</i>	58
4.4.2.	<i>Simulace boxu s vodou a ionty</i>	60
4.4.3.	<i>Simulace lysozymu</i>	61

4.4.4. Propojení sekvencí.....	63
5. DISKUZE.....	65
6. ZÁVĚR.....	66
7. SEZNAM POUŽITÉ LITERATURY.....	68
8. KLÍČOVÁ SLOVA.....	69
9. PŘÍLOHY.....	70

Úvod

Molekulární dynamika (MD) spadá pod oblast molekulového modelování. Spojuje v sobě poznatky chemie, fyziky, matematiky a výpočetní techniky. Velmi zjednodušeně na ní lze pohlížet jako na virtuální mikroskop s vysokým časovým a prostorovým rozlišením. Pomocí výpočetní techniky se provádí simulace řádově až 10^6 částic. Simulace systémů, které jsem prováděla, dosahovaly řádově rozměrů *nm* a času *ps*.

Hlavním účelem MD je sledovat pohyb a vlastnosti částic v čase. Umožňuje vědcům experimentovat, aniž by k tomu potřebovali drahé vybavení laboratoře. Platí se pouze za počítačové vybavení a software. Metoda nalézá uplatnění v širokém spektru oblastí - biochemie a biofyziky, při výzkumu materiálů a léčiv, ale i ekologie. Velkou výhodou je bezpečné experimentování s radioaktivními a toxickými látkami.

Metoda Monte Carlo je stochastická metoda počítačové simulace, využívající generátor náhodných čísel. Z počáteční konfigurace se náhodným krokem dostaneme do nové konfigurace, která je přijata nebo odmítnuta v souladu s Boltzmannovským vzorkováním $\sim \exp(-E_p/kT)$.

Molekulární dynamika je deterministická metoda počítačové simulace. Dělí se na tři odvětví - modelování molekulově mechanické (MM), modelování kvantově mechanické (QM) a modelování hybridní (MM/QM). MM přístup dovoluje simulovat větší počet částic, je to rychlá metoda zahrnující více aproximací. Vyžaduje rozsáhlé parametrizace. Nelze ji ovšem použít standardně k simulaci chemických reakcí, kde vznikají nebo zanikají kovalentní vazby. Ve své práci se budu zabývat právě tímto klasickým přístupem. QM je přesnější metoda, ale náročná na výpočet. Hodí se pro menší systém částic, řádově 10^2 částic. Hybridní modelování je spojením předchozích přístupů. Na malou zkoumanou část systému se aplikuje QM a na zbytek MM.

1. Současný stav

1.1. Molekulární mechanika

Klasická molekulární dynamika se opírá o principy molekulární mechaniky. Modelem atomu je koule s van der Waalsovským poloměrem, nábojem ve svém středu, který reprezentuje celkový nebo parciální náboj atomu. Atomy jsou rozděleny do atomových typů podle své chemické značky, vaznosti a geometrie vázaných atomů. Vazba se modeluje v nejjednodušším a nejběžnějším případě jako harmonická pružina s tuhostí k . Kompletním popisem interakcí v systému je silové pole, které tvoří potenciálové funkce (podle nichž se počítá potenciální energie) a kompletní sada parametrů vystupujících v potenciálových funkcích. Tyto parametry jsou zpravidla určeny jen pro čisté látky, křížové členy určující interakci mezi různými typy molekul jsou zpravidla dopočítány podle kombinačních pravidel.

Lorentz-Berthelotovo kombinační pravidlo určuje potenciálové parametry σ a ε pro heteroatomické páry (ij , $i \neq j$) tak, že kombinuje analogické potenciálové parametry pro pár identických molekul (ii a jj).

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj})$$

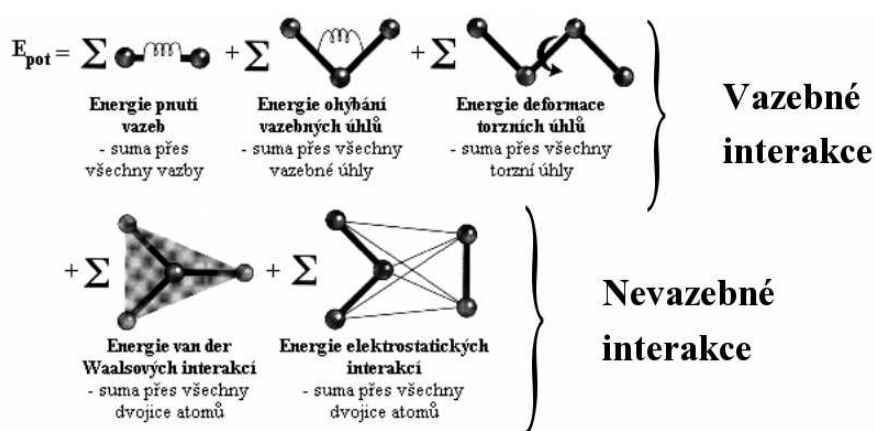
$$\varepsilon = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$$

Dalším pravidlem je Berthelotovo, které má tvar pouze geometrických průměrů. Potenciální energie molekuly E_p závisí na geometrickém uspořádání (konformaci) molekuly a její interakcích s okolím. Pokud je hodnota potenciální energie nízká, molekula je stabilní. Potenciálová funkce počítá E_p jako součet energií všech interakcí v molekule a interakcí s ostatními molekulami. Interakce dělíme na vazebné a nevazebné. Vazebné interakce říkají, jak se liší geometrie od „ideálního stavu“. V rámci silového pole jsou definované ideální délky vazeb, vazebných a torzních úhlů. Čím víc se určitý parametr liší od ideálního, tím větší má daná interakce energii. Nevazebné interakce jsou elektrostatické a neelektrostatické.

Metodou molekulární mechaniky je minimalizace energie. Metodou největšího spádu nebo konjugovaných gradientů se struktura následně optimalizuje. Kinetická energie se zanedbává – výpočet probíhá pro teplotu absolutní nuly.[3].

1.1.1. Silové pole a potenciálová funkce

Interakce molekul jsou modelovány pomocí silového pole. Silové pole se dělí na interakce vazebné a nevazebné. Interakce jsou matematicky vyjádřené pomocí soustavy rovnic (nazývané také potenciální funkce) a parametrů, které jsou v rovnicích vystupují.



Obr. 1.1. Příklad potenciálové funkce [2].

Z rovnic se generují potenciální energie a jejich derivace, síly. Pro vytvoření dobrého popisu silového pole pracují týmy vědců i několik let. Volba silového pole je významným faktorem ovlivňujícím přesnost výsledků i efektivitu simulace. První otázkou je, které silové pole obsahuje parametry pro molekuly, které jsou v našem systému. Tři základní typy silového pole:

- 1) all atom parametry jsou stanoveny pro každý jednotlivý atom v systému, včetně vodíku
- 2) united atom skupina atomů je považována za jedno interakční centrum, např. atomy H a C v methyl a methylenových skupinách

3) coarse grained se používají pro velmi dlouhé simulace bílkovin a jiných polymerů, abstraktně reprezentují molekuly seskupováním atomů do superatomů za účelem vyšší výpočetní výkonnosti a možnosti simulovat systémy na větší prostorové i časové škále.

1.1.2. *Potenciály*

Pro výpočet energie specifické interakce využívají silová pole různé popisy. Uvedeny jsou nejčastější vyjádření.

1.1.2.1. *Potenciály vazebné*

Pnutí vazeb – harmonický potenciál

Představuje interakce mezi párem atomů, které jsou spojeny jednou kovalentní vazbou (může být jednoduchá i vícenásobná). Je to přiblížení energie vazby jako funkce výchylky od ideální délky vazby, b_0 . Silová konstanta K_b určuje sílu vazby. b_0 a K_b jsou specifické pro každý pár vázaných atomů, závisí na typu atomů a vazby.

$$E_b = \sum_{1,2} K_b (b - b_0)^2$$

Energie ohýbání úhlů – harmonický potenciál

Souvisí se změnou vazebných úhlů od ideálních hodnot, což je reprezentováno harmonickým potenciálem. Hodnoty θ_0 a K_θ závisí na typu atomů, které tvoří úhel. Energie ohýbání úhlů popisuje odchylku od ideální geometrie – v dokonale optimalizované struktuře by měla být blízká nule.

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

Energie torzí

Reprezentuje energii spojenou s rotací úseku atomů (A-B-C-D) kolem vazby B-C. Lze též interpretovat jako potenciál závislý na dihedralním úhlu Φ mezi rovinami ABC a BCD. Potenciál se předpokládá periodický a je často vyjadřován rozvojem v kosinových funkcích.

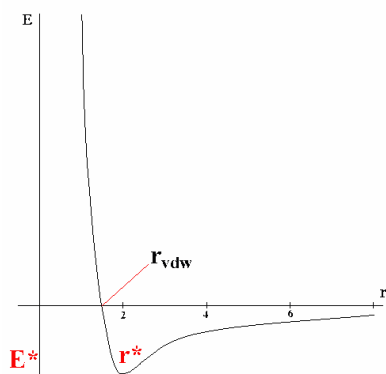
Parametry pro tyto termy (K_Φ , ...) jsou získané ze studií malých modelů a porovnáním s geometrií a vibračním spektrem v plynné fázi (IR a Ramanova spektroskopie), doplněné kvantovými výpočty *ab initio*.

$$E_{\phi} = \sum_{1,4 \text{ páry}} K_{\phi} (1 - \cos(n\phi))$$

1.1.2.2. Potenciály nevazebné

Energie van der Waalsových interakcí

Van der Waalsovy interakce mezi dvěma atomy vznikají z bilance mezi odpuzivými a přitažlivými silami. Odpuzivá síla vzniká na krátkých vzdálenostech, kde je silná repulze mezi elektrony. Přitažlivá síla (také označované jako disperzní síla) vzniká z fluktuační v nábojovém rozložení v elektronových mracích. Fluktuační v elektronové distribuci vede ke vzniku okamžitého dipólu, který indukuje dipól v druhém atomu či molekule, což dává vznik přitažlivým silám. Tyto síly klesají s šestou mocninou vzdálenosti. Polohy atomů v optimální vzdálenosti stabilizují systém. Hodnota minimální energie E a optimální vzdálenosti atomů r (což je zhruba součet van der Waalsových poloměrů atomu) závisí na parametrech silového pole těchto atomů.



Obr. 1.2. Přitažlivé a repulzní síly [4].

Van der Waalsovy interakce jsou nejčastěji modelovány pomocí Lennard–Jones 12-6 potenciálu. Vyjadřuje interakční energii pomocí atomově-závislých konstant A a C . Hodnoty A a C mohou být stanoveny řadou hodnot, jako jsou nevazebné vzdálenosti v krystalech nebo měřené vzdálenosti v plynné fázi. Van der Waalsovy interakce jsou jedním z nejdůležitějších pro stabilitu biologických makromolekul.

$$E_{LJ} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6}$$

Alternativně se Lennard-Jonesův potenciál vyjadřuje pomocí geometrického parametru σ vyjadřujícího vzdálenost, ve které je potenciál nulový a energetického parametru ϵ , který určuje hloubku potenciálového minima.

$$V_{LJ}(\mathbf{r}_{ij}) = 4\pi\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right]$$

Energie elektrostatických interakcí – Coulombův potenciál

Elektrostatické interakce mezi páry atomů jsou popsány Coulombovým potenciálem.

$$E_e = \sum \frac{q_i q_j}{4\pi\epsilon_0}$$

[4]

Kromě uvedených potenciálů se lze setkat i se složitějšími potenciálovými funkcemi, např.:

<i>Druh interakce</i>	<i>Možnosti potenciálů</i>
Pnutí vazeb	Harmonický, kubický, Morseho, FENE
Ohýbání úhlů	Harmonický, kosinový, Urey-Bradleyho
Torze	Ryckaert-Bellemansův, periodický, Fourierův
Van der Waalsovy interakce	Lennard-Jonesova 12-6, Buckingham (exp-6)
Elektrostatické interakce	Coulombův

Tab. 1.1. Možnosti vyjádření vazebných a ne vazebných potenciálů.

1.1.3. Parametrizace, typy silových polí

Silové pole kromě potenciálů definuje i sadu parametrů pro každý typ atomu. Například zpravidla silové pole obsahuje jiné parametry pro atom kyslíku v karbonylové skupině a v hydroxylové skupině. Typická sada atomárních parametrů obsahuje atomovou hmotnost, van der Waalsův poloměr, náboj, rovnovážnou délku vazeb, rovnovážné vazebné úhly, parametry torzních úhlů a hodnoty odpovídající efektivní tuhosti pružiny pro každý potenciál.

Všechna silová pole jsou založena na řadě numerických zjednodušení a pochází z různých typů experimentálních dat. Proto se jim říká empirická. Různá silová pole byla navržena pro různé účely. Klasická silová pole jsou AMBER (užívané pro proteiny a DNA), CHARMM (pro malé molekuly a makromolekuly), GROMACS (optimalizované pro program Gromacs), GROMOS (biomolekulární systémy), OPLS (Optimized Potentials for Liquid Simulations) – a jeho varianty: all atom OPLS-AA, united atom OPLS – UA a další silová pole.

1.2. Integrace pohybových rovnic

Základem metody MD je deterministické určení nové konfigurace v čase $t+h$ na základě znalostí polohy, rychlosti, sil, případně vyšších derivací polohy, v čase t .

1.2.1. Verletův algoritmus

Záporný gradient potenciálu se rovná síle působící na molekulu či atom. Pohybové rovnice jsou vyřešeny integrací – za pomoci jednoho z integračních algoritmů.

Síla, která působí na částici je záporný gradient potenciálu U :

$$\mathbf{F}_i = -grad_i U \quad (1)$$

Z druhého Newtonova zákona plyne diferenciální rovnice druhého řádu:

$$\ddot{\mathbf{r}}_i = \frac{\mathbf{f}_i}{m_i}, i = 1, 2, \dots, N \quad (2)$$

„K této soustavě musíme ještě přidat počáteční podmínky, tj. (obvykle) znalost všech poloh \mathbf{r}_i , a rychlostí \mathbf{r}_i' , pro jistý počáteční čas $t=t_0$. Řešením soustavy je tzv. trajektorie, tedy funkce $\mathbf{r}_i(t)$, $i = 1, 2, \dots, N$ v nějakém časovém intervalu (t_0, t_{max}) . Funkce spojitého argumentu budeme na počítači zaznamenávat v diskrétních bodech ve tvaru $t=t_0+ih$, kde h je jistý malý časový úsek (integrační krok) a i je nezáporné celé číslo. Krok h musí být natolik krátký, aby se během něj změnily polohy částic jakož i síly jen málo. Tento postup se nazývá *metoda konečných diferencí* [1,s.28]”. Pro atomistické simulace je h řádově 1 fs.

Nejběžnějším řešením metody konečných diferencí je Verletův algoritmus. Je založen na Taylorově rozvoji souřadnic do druhého řádu se středem v t .

$$\ddot{\mathbf{r}}_i(t) = \frac{\mathbf{r}_i(t-h) - 2\mathbf{r}_i(t) + \mathbf{r}_i(t+h)}{h^2} + O(h^2) \quad (3)$$

K potřebnému výpočtu pozice v bodě $\mathbf{r}_i(t+h)$ se rovnice (3) dosadí do (2).

$$\frac{\mathbf{r}_i(t-h) - 2\mathbf{r}_i(t) + \mathbf{r}_i(t+h)}{h^2} + O(h^2) = \frac{\mathbf{f}_i(t)}{m_i} \quad (4)$$

Zanedbáním chyby $O(h^2)$ a vyjádřením $\mathbf{r}_i(t+h)$ na levé straně dostaneme vztah

$$\mathbf{r}_i(t+h) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-h) + h^2 \frac{\mathbf{f}_i(t)}{m_i} \quad (5)$$

Podle rovnice ze známých hodnot – polohy \mathbf{r}_i v čase t , m_i vypočítáme nové hodnoty poloh v čase $t+h$. Pro první integrační krok simulace chybí předchozí poloha $\mathbf{r}_i(t-h)$. Pro tu lze použít rozvoj

$$\mathbf{r}_i(t-h) = \mathbf{r}_i(t) - h\dot{\mathbf{r}}_i(t) + \frac{h^2\ddot{\mathbf{r}}_i(t)}{2} + O(h^3) \quad (6)$$

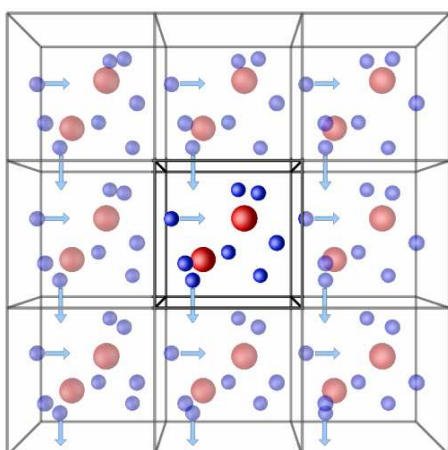
Pro zjednodušení lze člen s h^2 zanedbat.[1]

Dalšími početními metodami jsou např. Eulerova metoda, Leap-frog (alternativní zápis ekvivalentní Verletově metodě), rychlostní Verlet.

1.3. Technické detaily

1.3.1. Periodické okrajové podmínky

Částice uzavřeme do pracovní oblasti (box), která bude obsahovat N částic. Box obklopíme ze všech stran jeho obrazy, virtuálními kopiemi. V trojrozměrném prostoru je nejbližších obrazů 26. Kopie budou svázané s pracovní oblastí a tudíž každý pohyb částice z pracovní oblasti se promítne do všech kopií. Pokud částice opustí box, z druhé strany do boxu ekvivalentně částice vstoupí. Jsou ale případy, kdy je lepší PBC nevyužívat – v některých případech vyžadují použití velkého počtu molekul vody. Schůdnější varianta je obklopit protein, nebo jen jeho část vrstvou vody. Pro PBC se nejčastěji používá pracovní box ve tvaru krychle nebo oktahedronu [5].



Obr. 1.3. Periodické okrajové podmínky [7].

1.3.2. Oříznutí interakčního potenciálu (cutoff)

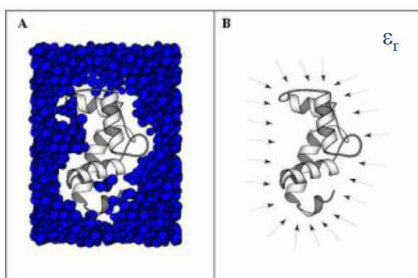
Slouží k urychlení výpočtů, jelikož právě nevazebné parametry molekuly jsou početně náročnou záležitostí – bez oříznutí potenciálu rostou s kvadrátem počtu částic. Je-li vzdálenost dvou atomů (nebo referenčních atomů v případě skupinového cutoffu) větší než zvolená vzdálenost r_{cut} , není tato interakce (elektrostatická nebo van der Waalsova) počítána v jednotlivých krocích, ale pouze započtena jako korekce celkové energie ve formě dlouhodobé korekce. Vzdálenosti r_{cut} musí splňovat minimální obrazovou konvenci – pokud chceme uvažovat jen nejbližší obrazy, nejdelší cutoff nesmí být delší než polovina nejkratší strany boxu.

1.3.3. Explicitní a implicitní solvatace

Explicitní solvatace využívá atomární model vody, kde se používají aproximace molekulární mechaniky. Bylo navrženo mnoho modelů, mohou být klasifikovány podle počtu bodů, sloužících k definování modelu, zda je struktura pevná nebo flexibilní, a zda obsahuje polarizační účinky (SPC, SPC/E, TIP3P, ...). Model vody je definován jeho geometrií a parametry jako je atomový náboj a parametry Lennardova – Jonesova potenciálu.

Implicitní solvatace je metoda reprezentování rozpouštědla jako kontinuálního média, které reprezentuje statistický průměr všech možných stavů rozpouštědla, namísto jednotlivých explicitních molekul. Zanedbávají se tedy atomové vlastnosti rozpouštědla.

Tato metoda se často používá k odhadu volné energie v strukturálních interakcích rozpouštědla-rozpuštěné látky a chemických procesech, jako je skládání nebo konformační přechody bílkovin, DNA, RNA, polysacharidů, sdružení biologických makromolekul s ligandy, nebo transport léků před biologické membrány. Pro simulace biomolekul je vhodnější implicitní rozpouštědlo v rámci QM. Pro MM je vhodné použít explicitní solvataci.



Obr. 1.4. Model prostředí A-explicitní B-implicitní solvatace [6].

1.3.4. Termostatování a barostatování

1.3.4.1. Kanonický soubor (NVT)

Různé metody termostatu “přidávají a odebírají energii” z hranic či celého objemu systému více či méně realistickým způsobem. Snaží se přiblížit kanonickému souboru, který má konstantní látkové množství, objem a teplotu. Metodami jsou přeskálování rychlostí, Nosé-Hooverův termostat, Berendsenův termostat nebo Langevinův termostat. Požadavkem je, aby teplota systému dosáhla stabilní hladiny okolo zvolené hodnoty [9].

1.3.4.2. Izotermicko-izobarický soubor (NPT)

Cílem je stabilizovat průměrnou hodnotu tlaku (většinou na 1 bar). V izotermicko-izobarickém souboru je konstantní látkové množství, tlak a teplota. Kromě termostatu je potřeba zapojit barostat. Pak systém nejvíce odpovídá laboratorním podmínkám, pokojové teplotě a tlaku. Metodami barostatování jsou například Berendsenův nebo Parrinello-Rahmanův barostat [9].

1.4. Software

Pro vykonávání molekulární dynamiky existuje řada softwarových balíčků. Ve své práci budu porovnávat dva z nich - Amber a Gromacs.

1.4.1. *Amber a Gromacs*

Amber se skládá ze dvou různě distribuovaných částí. První je neplacená AmberTools pro přípravu systému a analýzu výsledků. Druhá je Amber, jehož cena je 400 \$ pro akademické účely. Celkově však obsahuje asi 50 podprogramů.

Hlavními jsou NAB (Nucleic Acid Builder), který umožňuje sestavovat nukleové kyseliny, LEaP ve shellové a grafické verzi, a antechamber pro vytvoření a editaci systému před spuštěním výpočtu (v rámci AmberTools). Hlavními výpočetními programy pro MD jsou sander a pmemd, což je verze sanderu, která je optimalizovaná pro rychlé paralelní výpočty.

K analýze slouží ptraj, pbsa, a mm-pbsa. Aktuální verze programu je Amber 11 a AmberTools 1.5. Stránky programu jsou (<http://ambermd.org>).

Gromacs není členěn na přípravnou a simulační část, je to ucelený program pro vykonávání MD s řadou utilit. Pracuje se pouze na příkazové řádce. Gromacs je bezplatný software určený především pro biochemické molekuly jako jsou bílkoviny, lipidy a nukleové kyseliny, které mají hodně složité molekulární interakce. Gromacs je extrémně rychlý při výpočtu ne vazebných interakcí (které obvykle dominují simulaci) a tak se používá i pro výzkum nebiologických systémů, např. polymerů.

Gromacs obsahuje hlavní výpočetní program mdrun. Pro něj vždy slouží jako vstup jeden jediný soubor, který se předzpracovává programem grompp. Aktuální verze programu je Gromacs 4.5.3. Stránky Gromacsu jsou (<http://www.gromacs.org>).

		Gromacs	Amber
veličina	symbol	jednotky	
délka	r	nm	Å
hmotnost	m	amu	
čas	t	ps	
náboj	q	e	
teplota	T	K	
energie	E,V	kJ / mol	kcal/mol
síla	F	kJ / (mol x nm)	kcal/(mol x Å)
tlak	p	bar	
rychlost	v	nm / ps	Å / ps
objem	V	nm³	Å³

Tab. 1.2. Základní fyzikální jednotky programů Gromacs a Amber.

2. Cíle práce a hypotézy

1. Seznámení se základy molekulární dynamiky
2. Seznámení s prací v několika programech pro molekulární dynamiku (Gromacs, Amber)
3. Vyřešení několika modelových úloh v těchto programech (simulace biomolekul a jejich interakcí s povrchy)
4. Zhodnocení jednotlivých programů k řešení konkrétních úkolů

V této práci není obsaženo potvrzování nebo vyvracení hypotéz.

3. Metodika

S oběma programy jsem pracovala pomocí příkazové řádky v OS Linux Ubuntu 9.10. Registrací na stránkách MetaCentra (http://metavo.metacentrum.cz/cs/about/ngi_a_vo.html) jako student Jihočeské univerzity jsem získala přístup ke klastru Hermes, který JU vlastní. Klastry umožňují používání výpočetně náročných aplikací, právě jako jsou simulační programy.

V aktuální složce, kde byly všechny potřebné soubory pro následnou práci s Gromacsem nebo Amberem, jsem mohla spustit program. Seznam používaných příkazů v rámci linux:

příkaz	funkce	použití
mkdir	vytvoření adresáře	mkdir voda
ls	výpis aktuálního adresáře	ls
cd	přejít do adresáře	cd voda
cd ..	přejít do adresáře výš	cd ..
rm	odstranit soubor	rm voda
pwd	zobrazení aktuálního adresáře	pwd
tar cfz	zabalit archiv	tar cfz voda*
tar xfz	rozbalit archiv	tar xfz voda
scp	zabezpečené kopírování	scp voda.pdb student@prf.jcu.cz:
cp	kopírování	cp voda.pdb voda
whoami	aktuální uživatel	whoami
mc	midnight commander	mc
rmdir	smazání adresáře	rmdir voda
mv	přesun souboru do adresáře	mv voda.pdb voda
history	výpis práce	history
exit, quit	ukončení	exit, quit

Tab 3.1. Seznam nejpoužívanějších příkazů pro práci na příkazové řádce – manipulace se soubory a složkami.

4. Výsledky

4.1. Základní schéma MD a práce se soubory

Při provádění molekulární dynamiky se pracuje s mnoha soubory. Koncovky souborů budou uváděny v závorkách pro lepší přehlednost.

Dva typy souborů jsem získala externě – prvním je souřadnicový soubor, který definuje kartézské souřadnice pro atomy ve formátu (.pdb) z Protein Data Bank. Druhým je parametrický soubor. V Gromacsu má koncovku (.mdp), v Amberu (.in). Soubor určuje typ výpočtu, jeho délku, časový krok a další parametry simulace. Všechny ostatní soubory budou výsledkem činností programů Amber a Gromacs.

Příkladem k popsání základních fází simulace MD a k tomu, jaké soubory jsou v tomto procesu potřebné, bude simulace proteinu ve vodě za laboratorních podmínek. (podrobněji viz část 4.4.3). Popsány jsou především utility, které se hodí k řešení tohoto příkladu.

Gromacs má speciální utilitu **grompp**, která slouží pro předzpracování. Vstupními soubory jsou parametrický soubor (.mdp), souřadnicový soubor (.gro) a topologie (.top). Utilita provede potřebné kontroly a výsledkem je jediný binární soubor (.tpr), který se předává další utilitě nebo hlavnímu simulačnímu programu **mdrun**.

4.1.1. Počáteční konfigurace

Prvním krokem je výběr počáteční konfigurace. Mohou to být data experimentálně získaná rentgenovou difrakcí nebo NMR. Také se může použít struktura vytvořená modelováním. Mnoho proteinových struktur lze stáhnout ve formátu (.pdb) z Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>). Dalšími zdroji struktur jsou Nucleic Acid Database (<http://ndbserver.rutgers.edu>) nebo Cambridge Structural Database (<http://www.ccdc.cam.ac.uk/products/csd>).

Gromacs

Častým startovním formátem je (.pdb), který se konvertuje do formátu (.gro), který je Gromacsu vlastní. Konverze se provádí pomocí utilit jako **editconf** nebo **pdb2gmx**. Vizualizovat strukturu lze pomocí jiného software, např. VMD (<http://www.ks.uiuc.edu/Research/vmd>).

Amber

Pro sestavení malých systémů lze použít **xleap**. Pro velké struktury, jako jsou proteiny, je vhodnější opatřit si experimentální strukturu z externí databáze, opět v (.pdb) formátu, která se do LEaPu načte pomocí příkazu **loadpdb**. Pokud není dostupná experimentální struktura, další možností je strukturu modelovat pomocí programů, užívajících homologické modelování a techniky predikce. Pro nukleové kyseliny je to NAB (Nucleic Acid Builder), který je součástí AmberTools. Nástroje pro predikci proteinů jsou ve vývoji. Strukturu můžeme vizualizovat pomocí LEaP editoru nebo pomocí jiného softwaru.

4.1.2. Generování topologie, výběr silového pole a parametrů

Gromacs

Pdb2gmx je utilita ke generování topologie pro systémy složené z residuů, jako je právě protein. Residua jsou dílčí části, ze kterých se struktura skládá – v tomto případě aminokyseliny. Topologie pro residua jsou obsaženy v databázi. Utilita konvertuje souřadnice z (.pdb) formátu do (.gro). Volbou silového pole určíme, jaké parametry se mají načíst z knihovny Gromacsu. Program také přidává strukturu chybějící vodíky. Nekompletní vnitřní struktura nebo AK residua, kterým chybí atomy, mohou vést k chybě. Vstupním souborem je (.pdb). Výstupním souborem je topologie (.top), souřadnice (.gro) a soubor pozičního omezení posre.itp, který bude využit v pozdější části – fázi ekvilibrace. Pro molekuly, které nejsou zahrnuté v databázi residuů, je určen nástroj **g_x2top**.

Amber

V Amberu se topologie generuje pomocí utility **saveamberparm** v části LEaP. Volba silového pole není defaultní, ale stejně jako u Gromacsu ji lze defaultně nastavit.

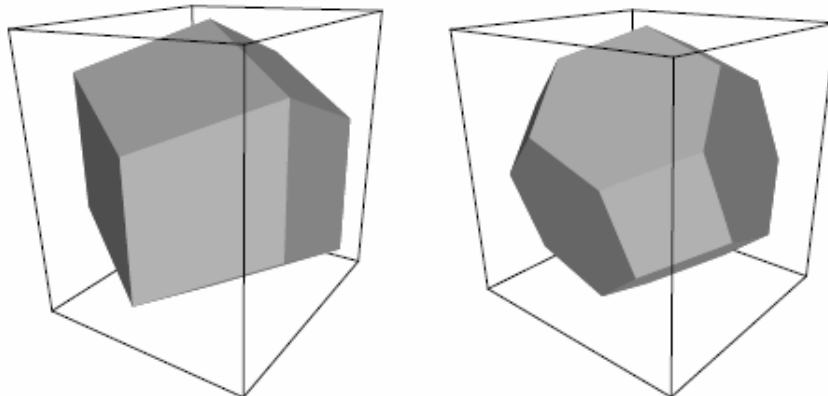
Volbou silového pole se načtou databáze (knihovny) residuí a parametry. Program očekává, že všechny zbytky v (.pdb) souboru jsou připojeny v pořadí, ve kterém jsou uvedeny. Symbol TER ukončuje řetězec. Do LEaP se soubor (.pdb) načte pomocí loadpdb. Načtený soubor definujeme pro LEaP jako jednotku, se kterou dále pracujeme. Při použití **saveamaberparm** se vygeneruje topologie (.prmtop) a počáteční souřadnice (.inpcrd). Pro nestandardní residua jako např. koenzym musíme použít **antechamber**.

4.1.3. Solvatace

Většina biochemických reakcí se odehrává ve viskózním prostředí. Proto se provádí solvatace, která přidá k látce explicitní molekuly rozpouštědla. Molekuly vody se získávají z velkého boxu vody, která byla již ekvilibrována. Celý box vody je překryt na protein a ty molekuly, které ho překrývají, jsou odstraněny.

Gromacs

Utilita **editconf** definuje rozměr boxu a **genbox** ho naplní rozpouštědlem. Tvar boxu může být krychle, nebo jiný, jako např. kosočtverný dvanáctistěn, který má jen 71% objemu krychle – ušetří počet molekul, které jsou třeba k solvataci a tím snižuje výpočetní náročnost úlohy. Vstupním a výstupním souborem pro editconf je (.gro), ke kterému se přidá informace o velikosti boxu. Pro genbox jsou vstupním souborem také souřadnice, výstupem jsou nové souřadnice s přidáním molekulami rozpouštědla a aktualizovaná topologie. Konfigurace rozpuštěné látky se generovala pomocí **pdb2gmx**, konfigurace vybraných rozpouštědel (různé modely vody, methanol, močovina, v některých silových polích i ethanol a 1-propanol) je součástí Gromacsu.



Obr. 4.1. Tvary solvatačních boxů – kosočtverečný dvanáctistěn a komolý osmistěn [8, s.13].

Amber

Solvatace se provádí pomocí **solvatebox** nebo **solvateoct**. Vytvoří se tím box s rozpouštědlem. Souřadnice a topologie aktualizovaného systému se uloží pomocí **saveamberparm**, neukládají se automaticky po vykonání příkazu jako u Gromacsu.

4.1.4. Přidání iontů

Gromacs

O náboji systému informuje topologie nebo řádka z výstupu **pdb2gmx**. Utilita **genion** přidává do systému ionty, či systém neutralizuje. Pro **genion** je potřeba jako vstup binární soubor (.tpr), sestavený pomocí **grompp**. Výstupem je nová konfigurace (.gro), textové soubory (.log) a aktualizovaná topologie (.top).

Amber

O celkovém náboji jednotky informuje příkaz **charge**. Přidání iontů nebo neutralizace se provede příkazem **addions**. Utilita přidává ionty dle potřeby na místo nejnižšího nebo nejvyššího elektrostatického potenciálu.

4.1.5. Minimalizace energie

Než se začne s dynamikou, je nutné zajistit, aby systém neměl žádné prostorové střety nebo nevhodnou geometrii. Struktura je optimalizována pomocí minimalizace

energie. Prohledává se křivka PES (Potential Energy Surface) a hledají se lokální minima. Proces odstraňuje všechny silné van der Waalsovy repulze, které by mohly vést k deformaci struktury.

Gromacs

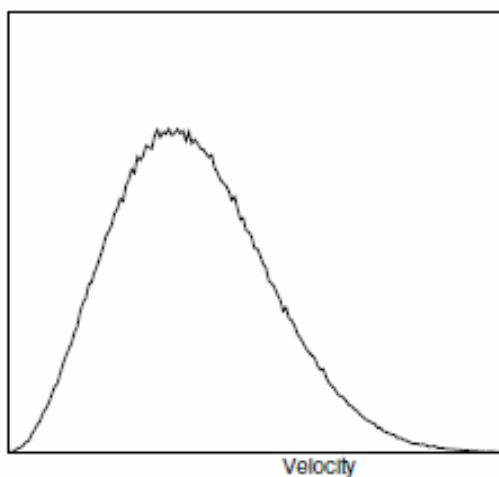
Prvním krokem je sestavení binárního souboru pro výpočet (.tpr). Pro minimalizaci energie se používá program **mdrun**. Vstupním souborem je soubor (.tpr). Výstupními soubory jsou minimalizovaná struktura (.gro), binární soubor energií (.edr), textový soubor (.log), případně binární soubor s trajektorií (.trr).

Amber

Minimalizace energie se vykonává utilitou **sander**. Z počátečních souřadnic (.inpcrd), topologie (.prmtop) a parametrického souboru (.in) dostaneme konečné souřadnice (.rst), a informace o průběhu minimalizace (.out).

4.1.6. Ekvilibrace

Tento proces probíhá jako krátký běh MD. Pokud nejsou počáteční rychlosti známy, přiřadí se podle Maxwell-Boltzmanovy distribuce.



Obr. 4.2. Tvar Maxwell-Boltzmanovy distribuce[8, s.17].

System se nejdříve přivede k požadované teplotě a pak k tlaku. Během této fáze je sledováno několik vlastností – zejména, tlak, teplota, hustota a energie. Cílem ekvilibrace je provést simulaci tak, aby byly tyto veličiny stabilní v čase.

Gromacs

Vstupem je soubor (.tpr). Výstupem **mdrun** jsou kontrolní soubor, který se píše ve stanovených intervalech a obsahuje informace o stavových veličinách v systému (.cpt), soubor s energiemi (.edr), výsledné souřadnice (.gro) a textový soubor (.log).

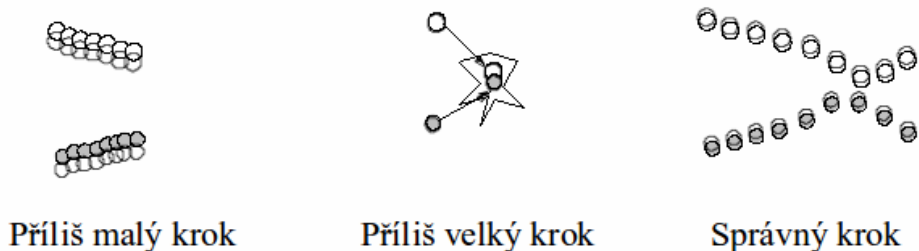
Amber

Topologie (.prmtop), souřadnice (.rst), souřadnice pro poziční omezení (.rst) a parametry (.in) jsou předány **sanderu**. Výstupem jsou nové souřadnice (.rst), trajektorie (.mdcrd) a textový soubor (.out).

4.1.7. Produkční fáze

Po stabilizaci teploty a tlaku systému krátkým během MD se přechází k produkční fázi MD.

Volba časového kroku nesmí být malá, pak MD pokrývá jen omezenou část konformačního prostoru a během dlouhého času výpočtu pokryje jen krátký čas vývoje systému. Příliš velký krok zase způsobí nestabilitu trajektorie a vzniklou energií překrývajících se atomů. Správný časový krok je 1/10 nejkratšího vibračního pohybu. Příklad - vazba C-H vibruje s periodou 10 fs, správný integrační krok $h = 1$ fs.[2]



Obr. 4.3. Důsledky volby časového kroku [2]

Gromacs

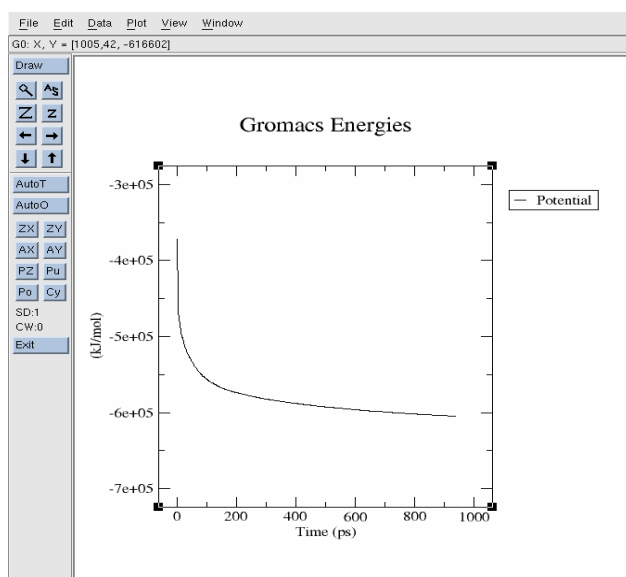
Vstupem je binární soubor (.tpr). Výpočet MD provádí program **mdrun**. Výpočet se může provádět paralelně na více procesorech nebo jádrech pro zvýšení rychlosti. Výstupem budou, tak jako v přechozím případě, čtyři soubory.

Amber

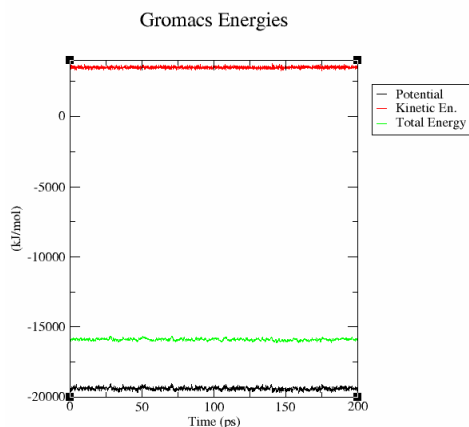
Hlavním výpočetní utilitou je opět **sander**. Z topologie (.prmtop), souřadnic z předchozího běhu (.rst) a parametrů (.in) dostaneme nové souřadnice (.rst), trajektorii (.mdcrd) a textový soubor (.out).

4.1.8. Analýza

Výsledkem minimalizace energie je postupné snižování E_p v čase. Analýza ekvilibrace se provede kontrolou energie, teploty, tlaku, objemu a hustoty systému, které by se měly v čase stabilizovat. K analýze produkční fáze se používá široká řada utilit.



Obr. 4.4. Graf časového vývoje E_p v programu Xmgrace.



Obr. 4.5. Hodnoty energií při produkční fázi MD.

Gromacs

Nástroj **g_energy** vytvoří ze souboru (.edr) soubor (.xvg) se zvolenými informacemi. Software Xmgrace (<http://plasma-gate.weizmann.ac.il/Grace>) pomocí vygenerovaného (.xvg) souboru vykreslí graf. Časový vývoj molekulární struktury lze zobrazit pomocí utility **ngmx**.či jiného softwaru, jako např. VMD. Gromacs obsahuje širokou paletu dalších analyzačních utilit:

UTILITA	FUNKCE
g_energy	zobrazuje průměrné hodnoty veličin, jejich časovou závislost (drift) a odchylky
g_rms	počítá odchylky (rmsd)
g_bond	analýza vazeb
g_rdf	počítá radiální distribuční funkci
g_angle	analýza úhlů
g_hbond	analýza vodíkových vazeb
g_tcaf	počítá viskozitu tekutin
g_potential	počítá elektrostatický potenciál
g_density	počítá lokální hustotu systému
g_angle	počítá úhel a vzdálenost mezi skupinami
g_dipole	spočítá celkový dipól
g_spol	analyzuje dipóly kolem solutu
g_sorient	analyzuje orientaci solventu kolem solutu

Tab 4.1. Příklady analyzačních nástrojů

Amber

Výstupní soubor (.out) obsahuje přehledné informace o hodnotách energie. Analýza může proběhnout za pomoci skriptu process_mdout.perl, který lze nalézt na

internetových

stránkách

(http://www.rosswalker.co.uk/tutorials/amber_workshop/Tutorial_two/files/process_md_out.perl) nebo pomocí utility **ptraj**. Strukturu lze také zobrazit pomocí VMD.

4.2. Formáty vstupních a výstupních souborů

4.2.1. Vstupní souřadnice

Souřadnicové soubory informují o počtu, poloze a jménech atomů a residuí. Dále mohou obsahovat rychlosti nebo velikost boxu.

4.2.1.1. formát (.pdb)

Je vstupním souborem pro Gromacs i Amber. Zdrojem mohou být internetové databáze. Pro jednoduché systémy můžeme využít software pro molekulární modelování.

```
TITLE Phenol
REMARK THIS IS A SIMULATION BOX
CRYST1 24.992 24.992 24.992 90.00 90.00 90.00 P 1 1
MODEL 1
ATOM 1 C1 MOL 1 4.160 25.190 24.530 1.00 0.00
ATOM 2 H1 MOL 1 3.360 24.470 24.410 1.00 0.00
ATOM 3 C2 MOL 1 4.600 25.880 23.390 1.00 0.00
ATOM 4 H2 MOL 1 4.360 25.480 22.420 1.00 0.00
ATOM 5 C3 MOL 1 5.100 27.180 23.530 1.00 0.00
ATOM 6 H3 MOL 1 5.440 27.710 22.650 1.00 0.00
ATOM 7 C4 MOL 1 5.070 27.780 24.800 1.00 0.00
ATOM 8 H4 MOL 1 5.350 28.820 24.870 1.00 0.00
ATOM 9 C5 MOL 1 4.720 27.050 25.940 1.00 0.00
ATOM 10 H5 MOL 1 3.910 27.540 26.450 1.00 0.00
ATOM 11 C6 MOL 1 4.410 25.700 25.810 1.00 0.00
ATOM 12 O1 MOL 1 5.180 24.870 26.570 1.00 0.00
ATOM 13 H6 MOL 1 5.670 25.330 27.240 1.00 0.00
TER
ENDMDL
```

Obr 4.6. Příklad (.pdb) souboru –molekula fenolu.

4.2.1.2. formát (.gro)

Souřadnicový soubor vlastní pro Gromacs. Generovaný pomocí **editconf** nebo vzniklý jako výstup simulace.

```
Phenol
13
1MOL C1 1 0.416 2.519 2.453
1MOL H1 2 0.336 2.447 2.441
1MOL C2 3 0.460 2.588 2.339
1MOL H2 4 0.436 2.548 2.242
1MOL C3 5 0.510 2.718 2.353
```

1MOL	H3	6	0.544	2.771	2.265
1MOL	C4	7	0.507	2.778	2.480
1MOL	H4	8	0.535	2.882	2.487
1MOL	C5	9	0.472	2.705	2.594
1MOL	H5	10	0.391	2.754	2.645
1MOL	C6	11	0.441	2.570	2.581
1MOL	O1	12	0.518	2.487	2.657
1MOL	H6	13	0.567	2.533	2.724
2.49920	2.49920	2.49920			

Obr. 4.7. Příklad (.gro) souboru – molekula fenolu.

4.2.1.3. formát (.inpcrd)

Amberovské souřadnice generované pomocí LEaP.

```

MOL
13
4.1600000 25.1900000 24.5300000 3.3600000 24.4700000 24.4100000
4.6000000 25.8800000 23.3900000 4.3600000 25.4800000 22.4200000
5.1000000 27.1800000 23.5300000 5.4400000 27.7100000 22.6500000
5.0700000 27.7800000 24.8000000 5.3500000 28.8200000 24.8700000
4.7200000 27.0500000 25.9400000 3.9100000 27.5400000 26.4500000
4.4100000 25.7000000 25.8100000 5.1800000 24.8700000 26.5700000
5.6700000 25.3300000 27.2400000

```

Obr 4.8. Příklad (.inpcrd) souboru –molekula fenolu.

4.2.1.4. formát (.mol2)

Amberovské souřadnice generované antechamberem.

```

@<TRIPOS>MOLECULE
MOL
13 13 1 0 0
SMALL
bcc

@<TRIPOS>ATOM
1 C1 4.1600 25.1900 24.5300 ca 1 MOL -0.156600
2 H1 3.3600 24.4700 24.4100 ha 1 MOL 0.149900
3 C2 4.6000 25.8800 23.3900 ca 1 MOL -0.097400
4 H2 4.3600 25.4800 22.4200 ha 1 MOL 0.133200
5 C3 5.1000 27.1800 23.5300 ca 1 MOL -0.165000
6 H3 5.4400 27.7100 22.6500 ha 1 MOL 0.134200
7 C4 5.0700 27.7800 24.8000 ca 1 MOL -0.091900
8 H4 5.3500 28.8200 24.8700 ha 1 MOL 0.131600
9 C5 4.7200 27.0500 25.9400 ca 1 MOL -0.213200
10 H5 3.9100 27.5400 26.4500 ha 1 MOL 0.132700
11 C6 4.4100 25.7000 25.8100 ca 1 MOL 0.123100
12 O1 5.1800 24.8700 26.5700 oh 1 MOL -0.498800

```



```

13 H6      5.6700 25.3300 27.2400 ho    1 MOL    0.418300
@<TRIPOS>BOND
1 1 2 1
2 1 3 ar
3 1 11 ar
4 3 4 1
5 3 5 ar
6 5 6 1
7 5 7 ar
8 7 8 1
9 7 9 ar
10 9 10 1
11 9 11 ar
12 11 12 1
13 12 13 1
@<TRIPOS>SUBSTRUCTURE
1 MOL      1 TEMP      0 **** 0 ROOT

```

Obr. 4.9. Příklad (.mol2) souboru –molekula fenolu.

4.2.2. Topologický soubor

Popisuje propojení mezi atomy, přiřazení typů jednotlivým atomům a jejich interakční parametry. Ke generování topologie je potřeba souřadnicový soubor s kartézskými souřadnicemi všech atomů. Jako další krok následuje definice silového pole. Výsledný soubor obsahuje informace o atomech, vazbách, úhlech, torzních úhlech, nábojích a dalším.

4.2.2.1. Gromacsovské topologie

V Gromacsu jsou tři typy topologických souborů:

1) systémová topologie (.top)

Soubor popisuje topologii celého systému. Systémovou topologii musíme vygenerovat. Na začátku je zahrnuta informace o zvoleném silovém poli, viz obr. 4.11. Každá sekce obsahuje informace ve sloupcích. Systémová topologie musí obsahovat právě jednu sekci [system], která definuje název systému a sekci [molecules], ve které se udává počet molekul jednotlivých typů molekul. Detailní topologie typů molekul je buď obsažena v samostatných (.itp) souborech a vložena instrukcí include, nebo přímo obsažena v sekcích [moleculetype] definujících v dalších podsekcích topologii molekuly; [atoms] definuje atomy – nr (číslo atomu), type (přiřazený typ na základě

výběru silového pole), náboj a nábojové skupiny. Nábojové skupiny by měly být voleny tak, aby jejich celkový náboj byl blízký nule. Sekce [bonds] určuje vazby - funct 1 popisuje vazbu harmonickým potenciálem. Sekce [pairs] popisuje atomy, které jsou od sebe odděleny třemi vazbami, tzv. 1-4 páry. Další sekce jsou úhly [angles], torze [dihedrals]. Příklad topologického souborů viz obr. 4.10, 4.11.

2) vložená topologie (.itp)

Soubor je zahrnut v rámci systémové topologie, definuje jen část systému topologie jako samostatný soubor, např různé druhy rozpouštědla, ionty nebo jednotlivé molekuly. Najdeme ho v gromacsovské databázi pro určené silové pole nebo použijeme uživatelsky vytvořené. Také definují vazebné a nevazebné interakce pro atomy.

3) reziduová topologie (.rtp)

Residuum je stavebním blokem většího celku. Program **pdb2gmx** používá informace ve své databázi residuů k sestavení systémové topologie.

Topologie se generuje pomocí souřadnicového souboru a definováním silového pole. Program přiřazuje pro molekuly molekulový typ, náboje a další parametry podle své vnitřní databáze. Příkladem pro generování topologie pro utility bude 1) molekula fenolu (fenol.pdb), 2) protein lyzozym (1AKI.pdb).

Utilita **pdb2gmx** převádí souřadnicový soubor (.pdb) do formátu (.gro) a generuje topologii (.top). Je vhodný pro molekuly složených s residuů, které jsou předdefinované v databázi. Pro malé molekuly, které nejsou explicitně zahrnuty, vhodný není. Při generování topologie pro molekulu lyzozymu je úspěšný. Topologie se generuje příkazem:

```
pdb2gmx -f 1AKI.pdb -o 1AKI.gro
```

Po zobrazení nabídky silových polí vybereme např. OPLS-AA. Topologie se vygenerovala úspěšně jako soubor topol.top. Při stejném postupu pro molekulu fenolu skončí proces chybou, protože tato molekula není v databázi jako residuum.

Nástroj **x2top** (**g_x2top** pro aktuální verzi 4.3.5) je další možností, jak generovat topologii. Příkaz pro vytvoření topologie pro molekulu fenolu:

```
x2top -f fenol.pdb -o fenol.top
```

```

;      File 'phenol_x2top.top' was generated
;
;      By user: fey (62843)
;      On host: hermes.prf.jcu.cz
;      At date: Tue Mar 29 14:24:12 2011
;
;      This is your include topology file
;      Generated by x2top
;
; Include forcefield parameters
#include "ffoplsaa.itp"

[ moleculetype ]
; Name      nrexcl
ICE          3

[ atoms ]
; nr  type resnr residue atom  cgnr  charge  mass typeB  chargeB  massB
1  opls_145  1  MOL  C1  1  -0.12  12.011 ; qtot -0.12
2  opls_140  1  MOL  H1  1  0.06   1.008 ; qtot -0.06
3  opls_145  1  MOL  C2  1  -0.12  12.011 ; qtot -0.18
4  opls_140  1  MOL  H2  1  0.06   1.008 ; qtot -0.12
5  opls_145  1  MOL  C3  1  -0.12  12.011 ; qtot -0.24
6  opls_140  1  MOL  H3  1  0.06   1.008 ; qtot -0.18
7  opls_145  1  MOL  C4  1  -0.12  12.011 ; qtot -0.3
8  opls_140  1  MOL  H4  1  0.06   1.008 ; qtot -0.24
9  opls_145  1  MOL  C5  1  -0.12  12.011 ; qtot -0.36
10 opls_140  1  MOL  H5  1  0.06   1.008 ; qtot -0.3
11 opls_145  1  MOL  C6  1  -0.12  12.011 ; qtot -0.42
12 opls_154  1  MOL  O1  1  -0.683 15.9994 ; qtot -1.103
13 opls_155  1  MOL  H6  1  0.418  1.008 ; qtot -0.685

[ bonds ]
; ai  aj funct  c0  c1  c2  c3
1  2  1  1.080000e-01  4.000000e+05  1.080000e-01  4.000000e+05
1  3  1  1.400000e-01  4.000000e+05  1.400000e-01  4.000000e+05
1  11 1  1.400000e-01  4.000000e+05  1.400000e-01  4.000000e+05
3  4  1  1.080000e-01  4.000000e+05  1.080000e-01  4.000000e+05
3  5  1  1.400000e-01  4.000000e+05  1.400000e-01  4.000000e+05
5  6  1  1.080000e-01  4.000000e+05  1.080000e-01  4.000000e+05
5  7  1  1.400000e-01  4.000000e+05  1.400000e-01  4.000000e+05
7  8  1  1.080000e-01  4.000000e+05  1.080000e-01  4.000000e+05
7  9  1  1.400000e-01  4.000000e+05  1.400000e-01  4.000000e+05
9  10 1  1.080000e-01  4.000000e+05  1.080000e-01  4.000000e+05
9  11 1  1.390000e-01  4.000000e+05  1.390000e-01  4.000000e+05
11 12 1  1.360000e-01  4.000000e+05  1.360000e-01  4.000000e+05
12 13 1  9.490000e-02  4.000000e+05  9.490000e-02  4.000000e+05

[ pairs ]
; ai  aj funct  c0  c1  c2  c3
1  6  1
1  7  1
1  10 1
1  13 1
2  4  1
2  5  1
2  9  1
2  12 1
3  8  1
3  9  1
3  12 1
4  6  1
4  7  1
4  11 1
5  10 1
5  11 1
6  8  1
6  9  1
7  12 1

```

```

8 10 1
8 11 1
9 13 1
10 12 1

[ angles ]
; ai aj ak funct c0 c1 c2 c3
2 1 3 1 1.180000e+02 4.000000e+02 1.180000e+02 4.000000e+02
2 1 11 1 1.180000e+02 4.000000e+02 1.180000e+02 4.000000e+02
3 1 11 1 1.200000e+02 4.000000e+02 1.200000e+02 4.000000e+02
1 3 4 1 1.190000e+02 4.000000e+02 1.190000e+02 4.000000e+02
1 3 5 1 1.190000e+02 4.000000e+02 1.190000e+02 4.000000e+02
4 3 5 1 1.210000e+02 4.000000e+02 1.210000e+02 4.000000e+02
3 5 6 1 1.190000e+02 4.000000e+02 1.190000e+02 4.000000e+02
3 5 7 1 1.190000e+02 4.000000e+02 1.190000e+02 4.000000e+02
6 5 7 1 1.220000e+02 4.000000e+02 1.220000e+02 4.000000e+02
5 7 8 1 1.180000e+02 4.000000e+02 1.180000e+02 4.000000e+02
5 7 9 1 1.210000e+02 4.000000e+02 1.210000e+02 4.000000e+02
8 7 9 1 1.210000e+02 4.000000e+02 1.210000e+02 4.000000e+02
7 9 10 1 1.100000e+02 4.000000e+02 1.100000e+02 4.000000e+02
7 9 11 1 1.190000e+02 4.000000e+02 1.190000e+02 4.000000e+02
10 9 11 1 1.090000e+02 4.000000e+02 1.090000e+02 4.000000e+02
1 11 9 1 1.190000e+02 4.000000e+02 1.190000e+02 4.000000e+02
1 11 12 1 1.130000e+02 4.000000e+02 1.130000e+02 4.000000e+02
9 11 12 1 1.140000e+02 4.000000e+02 1.140000e+02 4.000000e+02
11 12 13 1 1.130000e+02 4.000000e+02 1.130000e+02 4.000000e+02

[ dihedrals ]
; ai aj ak al funct c0 c1 c2 c3 c4 c5
11 1 3 5 3 1.490000e+02 5.000000e+00 3.000000e+00 1.490000e+02 5.000000e+00 3.000000e+00
3 1 11 9 3 2.370000e+02 5.000000e+00 3.000000e+00 2.370000e+02 5.000000e+00 3.000000e+00
1 3 5 7 3 1.680000e+02 5.000000e+00 3.000000e+00 1.680000e+02 5.000000e+00 3.000000e+00
3 5 7 9 3 2.050000e+02 5.000000e+00 3.000000e+00 2.050000e+02 5.000000e+00 3.000000e+00
5 7 9 11 3 1.820000e+02 5.000000e+00 3.000000e+00 1.820000e+02 5.000000e+00 3.000000e+00
7 9 11 1 3 1.360000e+02 5.000000e+00 3.000000e+00 1.360000e+02 5.000000e+00 3.000000e+00
1 11 12 13 3 2.820000e+02 5.000000e+00 3.000000e+00 2.820000e+02 5.000000e+00 3.000000e+00

[ system ]
; Name
ICE

[ molecules ]
; Compound #mols
ICE 1

```

Obr 4.10. Příklad (.top) souboru generovaného příkazem *x2top* – molekula fenolu

Program vytvoří topologii avšak varuje, že náboje generované tímto procesem nejsou přesné a je dobré si je ověřit podle jiného zdroje. Pro velké systémy není **x2top** vhodný. Pokus o vytvoření topologie pro lysozym se nezdařil.

Další možnost, jak generovat topologii nabízí skript **mktop** (není součástí Gromacsu). Skript lze stáhnout (<http://labmm.iq.ufrj.br/mktop/>). Příkazy ke generování topologie:

```
chmod u+rxw mktop.pl
```

```
./mktop.pl -i fenol.pdb -c charges.txt -t topology.top
```

```
; Include forcefield parameters
#include "/usr/local/bin/gromacs-4.0.4/share/top/ffoplsaa.itp"

[ moleculetype ]
; Name          nrexcl
MOL             3

[ atoms ]
; nr   type  resnr residue  atom  cgnr  charge  mass  typeB  chargeB  massB
  1  opls_145  1  MOL    C1    1 -0.22805  12.011
  2  opls_146  1  MOL    H1    1  0.17609   1.008
  3  opls_145  1  MOL    C2    2 -0.19686  12.011
  4  opls_146  1  MOL    H2    2  0.17725   1.008
  5  opls_145  1  MOL    C3    3 -0.14955  12.011
  6  opls_146  1  MOL    H3    3  0.14961   1.008
  7  opls_145  1  MOL    C4    4 -0.19686  12.011
  8  opls_146  1  MOL    H4    4  0.17725   1.008
  9  opls_145  1  MOL    C5    5 -0.22805  12.011
 10  opls_146  1  MOL    H5    5  0.17609   1.008
 11  opls_220  1  MOL    C6    6  0.30608  12.011
 12  opls_154  1  MOL    O1    6 -0.54018  15.9994
 13  opls_155  1  MOL    H6    6  0.37716   1.008

[ bonds ]
 1 11 1  0.140 265265.6
 13 1  0.140 392459.2
 1 2 1  0.108 307105.6
 3 5 1  0.140 392459.2
 3 4 1  0.108 307105.6
 5 7 1  0.140 392459.2
 5 6 1  0.108 307105.6
 7 9 1  0.140 392459.2
 7 8 1  0.108 307105.6
 9 11 1  0.139 265265.6
 9 10 1  0.108 307105.6
 11 12 1  0.136 267776.0
 12 13 1  0.095 462750.4

[ angles ]
 1 11 9 1  118.594 334.720
 1 11 12 1  112.866 418.400
 13 5 1  119.167 527.184
 13 4 1  118.650 292.880
 2 1 11 1  118.379 513.800
 2 1 3 1  117.937 292.880
 3 1 11 1  120.497 585.760
 3 5 7 1  118.645 527.184
 3 5 6 1  119.062 292.880
 4 3 5 1  120.990 292.880
```

```

5 7 9 1 121.292 527.184
5 7 8 1 117.685 292.880
6 5 7 1 122.187 292.880
7 9 11 1 119.092 585.760
7 9 10 1 109.715 292.880
8 7 9 1 121.007 292.880
9 11 12 1 114.377 418.400
10 9 11 1 108.581 513.800
11 12 13 1 112.955 460.240

```

```

[ dihedrals ]
3 1 11 9 3
2 1 11 9 3
3 1 11 12 3
2 1 11 12 3
11 1 3 5 3
2 1 3 5 3
11 1 3 4 3
2 1 3 4 3
1 3 5 7 3
4 3 5 7 3
1 3 5 6 3
4 3 5 6 3
3 5 7 9 3
6 5 7 9 3
3 5 7 8 3
6 5 7 8 3
5 7 9 11 3
8 7 9 11 3
5 7 9 10 3
8 7 9 10 3
7 9 11 1 3
10 9 11 1 3
7 9 11 12 3
10 9 11 12 3
1 11 12 13 3
9 11 12 13 3

[ dihedrals ]
11 1 3 2 1 improper_Z_CA_X_Y
1 3 5 4 1 improper_Z_CA_X_Y
3 5 7 6 1 improper_Z_CA_X_Y
5 7 9 8 1 improper_Z_CA_X_Y
7 9 11 10 1 improper_Z_CA_X_Y

[ pairs ]
9 3 1
9 2 1
12 3 1
12 2 1
5 11 1
5 2 1
4 11 1
4 2 1
7 1 1
7 4 1
6 1 1
6 4 1
9 3 1
9 6 1
8 3 1
8 6 1
11 5 1
11 8 1
10 5 1
10 8 1
1 7 1
1 10 1
12 7 1
12 10 1
13 1 1
13 9 1

[ system ]
; Name
MKTOP

[ molecules ]
; Compound #mols
MOL 1

```

Obr 4.11. Příklad (.top) souboru generovaného skriptem mktop – molekula fenolu.

4.2.2.2. Amberovská topologie

Amberovská systémová topologie také popisuje parametry pro systém, ale soubor je nepřehledný, není tak dobře čitelný jako gromacsovský (.top). Pro lepší přehled dat se používá ke zpracování (.prmtop) utilita **ptraj**.

```

%VERSION VERSION_STAMP = V0001.000 DATE = 04/12/11 00:09:53
%FLAG TITLE
%FORMAT(20a4)

%FLAG POINTERS
%FORMAT(10i8)
  2202  16 1201 1021 2166 1381 4229 3433  0  0
 11053 207 1021 1381 3433  43  89  42 30  1
   0  0  0  0  0  0  0  0 24  0
  0
%FLAG ATOM_NAME
%FORMAT(20a4)
N H1 H2 H3 CA HA CB HB2 HB3 CG HG2 HG3 CD HD2 HD3 CE HE2 HE3 NZ HZ1
HZ2 HZ3 C O N H CA HA CB HB CG1 HG11HG12HG13CG2 HG21HG22HG23C O
N H CA HA CB HB2 HB3 CG CD1 HD1 CE1 HE1 CZ HZ CE2 HE2 CD2 HD2 C O
N H CA HA2 HA3 C O N H CA HA CB HB2 HB3 CG HG2 HG3 CD HD2 HD3
NE HE CZ NH1 HH11HH12NH2 HH21HH22C O N H CA HA CB HB2 HB3 SG HG
C O N H CA HA CB HB2 HB3 CG HG2 HG3 CD OE1 OE2 C O N H CA
HA CB HB2 HB3 CG HG CD1 HD11HD12HD13CD2 HD21HD22HD23C O N H CA HA
CB HB1 HB2 HB3 C O N H CA HA CB HB1 HB2 HB3 C O N H CA HA
CB HB1 HB2 HB3 C O N H CA HA CB HB2 HB3 CG HG2 HG3 SD CE HE1 HE2
HE3 C O N H CA HA CB HB2 HB3 CG HG2 HG3 CD HD2 HD3 CE HE2 HE3 NZ

```

Obr. 4.12. Příklad části (.prmtop) souboru – molekula lysozymu.

```

1962: CD2 -0.41630 12.0 (129:LEU) CT BLA
1963: HD21 0.10380 1.0 (129:LEU) HC BLA
1964: HD22 0.10380 1.0 (129:LEU) HC BLA
1965: HD23 0.10380 1.0 (129:LEU) HC BLA
1966: C 0.83260 12.0 (129:LEU) C BLA
1967: O -0.81990 16.0 (129:LEU) O2 BLA
1968: OXT -0.81990 16.0 (129:LEU) O2 BLA
1969: O -0.83400 16.0 (130:WAT) OW BLA
1970: H1 0.41700 1.0 (130:WAT) HW BLA
1971: H2 0.41700 1.0 (130:WAT) HW BLA
1972: O -0.83400 16.0 (131:WAT) OW BLA
1973: H1 0.41700 1.0 (131:WAT) HW BLA
1974: H2 0.41700 1.0 (131:WAT) HW BLA
1975: O -0.83400 16.0 (132:WAT) OW BLA
1976: H1 0.41700 1.0 (132:WAT) HW BLA
1977: H2 0.41700 1.0 (132:WAT) HW BLA

```

Obr. 4.13. Zobrazení části (.prmtop) souboru pomocí ptraj – molekula lysozymu.

Utilita **saveamberparm** generuje souřadnice (.inpcrd) a topologii (.prmtop) ze souřadnicového souboru (.pdb). Je vhodná pro systémy složené z residuů. Před použitím **saveamberparm** nejdříve pomocí textového editoru editujeme (.pdb) soubor - odstraníme řádky s REMARK, jinak se generování topologie nezdaří.

1. Spuštění LEaP, načtení parametrů pro silové pole

```
tleap -s -f leaprc.ff99SB
```

2. Definování jednotky

```
sys = loadpdb 1AKI.pdb
```

3. Uložení topologie a souřadnic

```
saveamberparm sys 1AKI.prmtop 1AKI.inpcrd
```

Vytvoření topologie pro lysozym bylo úspěšné. Při generování topologie pro fenol už se při načtení molekuly objeví upozornění, že atomy nejsou v rámci residua. Topologie se nevytvoří, proto musíme použít jiný postup. **Antechamber** generuje topologie pro molekuly, které nejsou v databázi residuí. Čerpala jsem z tutorialu (<http://ambermd.org/tutorials/basic/tutorial4b/>). Topologie pro molekulu fenolu se vytvoří příkazy:

1. Generování mol2 souboru:

```
antechamber -i fenol.pdb -fi pdb -o fenol.mol2 -fo mol2 -c bcc -s 2
```

2. Kontrola parametrů, vytvoření parametrického souboru frcmod pro LEaP

```
parmchk -i fenol.mol2 -f mol2 -o fenol.frcmod
```

3. Spuštění LEaPu

```
tleap -f leaprc.ff99SB
```

4. Načtení parametrů pro silové pole GAFF

```
source leaprc.gaff
```

5. Definování jednotky do LEaP např. sys, kterou je molekula fenolu ve formátu mol2

```
SYS = loadmol2 fenol.mol2
```

6. Kontrola parametrů

```
check SYS
```

7. Načtení parametrů

```
loadamberparams fenol.frcmod
```

8. Vytvoření knihovny

```
saveoff SYS sys.lib
```

9. Vytvoření topologie a souřadnic

```
saveamberparm SYS fenol.prmtop fenol.inpcrd
```


4.2.2.3. Převaděč topologie a souřadnic Amber → Gromacs

Pearl skript amb2gmx (<http://ffamber.cnsm.csulb.edu/amb2gmx.pl>) umožňuje vytvořit z (.prmtop) a (.inpcrd) souborů gromacsovské (.top) a (.gro)

1. Vytvoření spustitelného programu

```
chmod a+x amb2gmx.pl
```

2. Spuštění amb2gmx

```
./amb2gmx.pl --prmtop voda_ionty.prmtop --crd voda_ionty.inpcrd --outname voda_ionty
```

4.2.3. Parametrické soubory

Soubory slouží k nastavení parametrů pro každý typ výpočtu. V Gromacsu mají koncovku (.mdp), v Amberu (.in). Příklady souborů jsou čerpány z tutoriálů (viz část 4.4.3). Jsou chronologicky seřazeny tak, jak se používají v procesu MD.

4.2.3.1. Parametry pro Gromacs (.mdp)

Minimalizace energie – minim.mdp

```
; minim.mdp - used as input into grompp to generate em.tpr
; Parameters describing what to do, when to stop and what to save
integrator      = steep          ; Algorithm (steep = steepest descent minimization)
emtol           = 1000.0        ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep         = 0.01          ; Energy step size
nsteps         = 50000         ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist        = 1             ; Frequency to update the neighbor list and long range forces
ns_type        = grid          ; Method to determine neighbor list (simple, grid)
rlist          = 1.0           ; Cut-off for making neighbor list (short range forces)
coulombtype    = PME           ; Treatment of long range electrostatic interactions
rcoulomb       = 1.0           ; Short-range electrostatic cut-off
rvdw           = 1.0           ; Short-range Van der Waals cut-off
pbc            = xyz           ; Periodic Boundary Conditions (yes/no)
```

Obr. 4.14. Příklad (.mdp) souboru pro minimalizaci energie.

Soubor určuje typ algoritmu pro výpočet, zde metodu největšího spádu (integrator=steep). Dále při jaké hodnotě maximální síly se má minimalizace ukončit (emtol) [kJ/(mol x nm)] a maximální velikost kroku (emstep) [nm] a maximální počet

kroků minimalizace (nsteps). Nstlist určuje frekvenci aktualizování seznamu sousedních částic. Rlist pro něj určuje vzdálenost cutoffu. Coulombtype říká, jak modelovat elektrostatiku. Rcoulomb určuje vzdálenost pro Coulombický cutoff. Rvdw určuje vzdálenost cutoffu pro Lennard-Jonesův nebo Buckingham potenciály. Uvedené periodické okrajové podmínky (pbc) jsou pro 3-D rozměr.

NVT Ekvilibrace – nvt.mdp

```

title           = OPLS Lysozyme NVT equilibration
define          = -DPOSRES           ; position restrain the protein
; Run parameters
integrator      = md                  ; leap-frog integrator
nsteps         = 50000                ; 2 * 50000 = 100 ps
dt              = 0.002                ; 2 fs
; Output control
nstxout        = 100                  ; save coordinates every 0.2 ps
nstvout        = 100                  ; save velocities every 0.2 ps
nstenergy      = 100                  ; save energies every 0.2 ps
nstlog         = 100                  ; update log file every 0.2 ps
; Bond parameters
continuation    = no                  ; first dynamics run
constraint_algorithm = lincs          ; holonomic constraints
constraints     = all-bonds           ; all bonds (even heavy atom-H bonds) constrained
lincs_iter     = 1                    ; accuracy of LINCS
lincs_order    = 4                    ; also related to accuracy
; Neighborsearching
ns_type        = grid                 ; search neighboring grid cels
nstlist        = 5                    ; 10 fs
rlist          = 1.0                  ; short-range neighborlist cutoff (in nm)
rcoulomb       = 1.0                  ; short-range electrostatic cutoff (in nm)
rvdw           = 1.0                  ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype    = PME                  ; Particle Mesh Ewald for long-range electrostatics
pme_order      = 4                    ; cubic interpolation
fourierspacing = 0.16                 ; grid spacing for FFT
; Temperature coupling is on
tcoupl         = V-rescale             ; modified Berendsen thermostat
tc-grps        = Protein Non-Protein   ; two coupling groups - more accurate
tau_t          = 0.1 0.1              ; time constant, in ps
ref_t          = 300 300              ; reference temperature, one for each group, in K
; Pressure coupling is off
pcoupl         = no                   ; no pressure coupling in NVT
; Periodic boundary conditions
pbc            = xyz                  ; 3-D PBC
; Dispersion correction
DispCorr       = EnerPres             ; account for cut-off vdW scheme
; Velocity generation
gen_vel        = yes                  ; assign velocities from Maxwell distribution
gen_temp       = 300                  ; temperature for Maxwell distribution
gen_seed       = -1                   ; generate a random seed
  
```

Obr. 4.15. Příklad souboru (.mdp) pro nvt fázi ekvilibrace.

Pomocí NVT souboru má systém dosáhnout stabilní hodnoty teploty. Soubor definuje použitý algoritmus (integrator), počet kroků (nsteps) a časový krok (dt) [ps]. Nastavuje i výstupní parametry - po kolika krocích se ukládají informace o souřadnicích (nstxtout), rychlostech (nstvout), energii (nstenergy).

V sekci termostatování je určena teplota, ke které chceme náš systém přivést (ref_t) [K], termostatované skupiny (tc_grps), typ termostatu (tcoupl) a jeho relaxační čas (tau_t) [ps]. Rychlosti jsou přiřazeny z Maxwellovy distribuce (gen_vel) při teplotě 300 K (gen_temp), barostat je vypnutý (pcoupl=no).

NPT Ekvilibrace – npt.mdp

```

title          = OPLS Lysozyme NPT equilibration
define         = -DPOSRES          ; position restrain the protein
; Run parameters
integrator     = md                 ; leap-frog integrator
nsteps        = 50000              ; 2 * 50000 = 100 ps
dt            = 0.002              ; 2 fs
; Output control
nstxout       = 100                ; save coordinates every 0.2 ps
nstvout       = 100                ; save velocities every 0.2 ps
nstenergy     = 100                ; save energies every 0.2 ps
nstlog        = 100                ; update log file every 0.2 ps
; Bond parameters
continuation   = yes               ; Restarting after NVT
constraint_algorithm = lincs       ; holonomic constraints
constraints    = all-bonds         ; all bonds (even heavy atom-H bonds) constrained
lincs_iter    = 1                  ; accuracy of LINCS
lincs_order   = 4                  ; also related to accuracy
; Neighborsearching
ns_type       = grid               ; search neighboring grid cels
nstlist       = 5                  ; 10 fs
rlist         = 1.0                ; short-range neighborlist cutoff (in nm)
rcoulomb      = 1.0                ; short-range electrostatic cutoff (in nm)
rvdw         = 1.0                 ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype   = PME                ; Particle Mesh Ewald for long-range electrostatics
pme_order     = 4                  ; cubic interpolation
fourierspacing = 0.16             ; grid spacing for FFT
; Temperature coupling is on
tcoupl        = V-rescale          ; modified Berendsen thermostat
tc-grps       = Protein Non-Protein ; two coupling groups - more accurate
tau_t         = 0.1 0.1           ; time constant, in ps
ref_t         = 300 300           ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl        = Parrinello-Rahman  ; Pressure coupling on in NPT
pcoupltype    = isotropic          ; uniform scaling of box vectors
tau_p         = 2.0                ; time constant, in ps

```

```

ref_p          = 1.0          ; reference pressure, in bar
compressibility = 4.5e-5      ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc           = xyz          ; 3-D PBC
; Dispersion correction
DispCorr      = EnerPres     ; account for cut-off vdW scheme
; Velocity generation
gen_vel       = no          ; Velocity generation is off

```

Obr. 4.16. Příklad souboru (.mdp) pro npt fázi ekvilibrace.

Pomocí NPT souboru se stabilizuje tlak a tím i hustota systému. Oproti NVT souboru má několik změn: definované parametry pro barostatování - průměrnou hodnotu tlaku, které chceme dosáhnout (ref_p) [bar], typ barostatu (pcoupl), izotropický typ barostatování (pcoupltype) - box bude expandovat či se stahovat rovnoměrně ve všech směrech k zachování správného tlaku, s časovou konstantou (tau_p)[ps], stlačitelností solventu použitého v simulaci (compressibility)[bar⁻¹], zde defaultní nastavení pro vodu při 300 K na 1 bar. Rychlosti se negenerují (gen_vel=no). Pokračuje se v simulaci z NVT fáze ekvilibrace (continuation=yes).

Produkční fáze – md.mdp

```

title          = OPLS Lysozyme MD
; Run parameters
integrator     = md          ; leap-frog integrator
nsteps        = 500000; 2 * 500000 = 1000 ps, 1 ns
dt            = 0.002       ; 2 fs
; Output control
nstxout       = 1000        ; save coordinates every 2 ps
nstvout       = 1000        ; save velocities every 2 ps
nstxtcout     = 1000        ; xtc compressed trajectory output every 2 ps
nstenergy     = 1000        ; save energies every 2 ps
nstlog        = 1000        ; update log file every 2 ps
; Bond parameters
continuation   = yes        ; Restarting after NPT
constraint_algorithm = lincs ; holonomic constraints
constraints    = all-bonds   ; all bonds (even heavy atom-H bonds) constrained
lincs_iter    = 1           ; accuracy of LINCS
lincs_order   = 4           ; also related to accuracy
; Neighborsearching
ns_type       = grid        ; search neighboring grid cels
nstlist      = 5           ; 10 fs
rlist        = 1.0         ; short-range neighborlist cutoff (in nm)
rcoulomb     = 1.0         ; short-range electrostatic cutoff (in nm)
rvdw        = 1.0         ; short-range van der Waals cutoff (in nm)

```

```

; Electrostatics
coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling is on
tcoupl = V-rescale ; modified Berendsen thermostat
tc-grps = Protein Non-Protein ; two coupling groups - more accurate
tau_t = 0.1 0.1 ; time constant, in ps
ref_t = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling is on
pcoupl = Parrinello-Rahman ; Pressure coupling on in NPT
pcoupltype = isotropic ; uniform scaling of box vectors
tau_p = 2.0 ; time constant, in ps
ref_p = 1.0 ; reference pressure, in bar
compressibility = 4.5e-5 ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc = xyz ; 3-D PBC
; Dispersion correction
DispCorr = EnerPres ; account for cut-off vdW scheme
; Velocity generation
gen_vel = no ; Velocity generation is off

```

Obr. 4.17. Příklad (.mdp) souboru pro produkční fázi MD

Hodnoty pro termostatování a barostatování jsou obdobné jako v NVT a NPT souborech, pouze doba simulace i intervaly ukládání konfigurací a mezivýsledků jsou prodlouženy.

4.2.3.2. *parametry pro Amber (.in)*

Minimalizace energie -min1.in, min2.in

```

energy minimization stage 1
&cntrl
imin=1, maxcyc=5000, ncyc=2500,
cut=10.0, ntb=1,
ntc=1, ntf=1,
ntpr=10,
ntr=1,
restraintmask=':1-500',
restraint_wt=2.0
/

energy minimization stage 2
&cntrl
imin=1, maxcyc=10000, ncyc=5000,
cut=10.0, ntb=1,
ntc=1, ntf=1,
ntpr=10,
/

```

Obr. 4.18. Příklad (.in) souborů pro minimalizaci energie

První soubor aplikuje poziční omezení na molekulu rozpuštěné látky (specifikováno čísla atomů, zde 1-500), minimalizuje se solvent. Druhý soubor je předpisem pro minimalizaci energie celého systému. Vykonává se minimalizace ($imin=1$) s maximálním počtem kroků ($maxcyc$). Nejdříve se provádí metoda největšího spádu, která se po zvoleném počtu cyklů ($ncyc$) přepne na konjugované gradienty. Velikost interakčního ořezu určí (cut) [Å]. Používají se periodické okrajové podmínky ($ntb=1$), algoritmus SHAKE není aktivní ($ntc=1$). Informace o hodnotě energie budou zaznamenávány do souboru (.out) ve zvoleném intervalu ($ntpr$). Poziční omezení se provádí pomocí harmonického potenciálu ($ntr=1$). Čísla residuí, která budou pozičně omezena, budou definována v *restraintmask*. Silová konstanta pro omezení je dána (*restraint_wt*).

Ohřev systému – heat.in

```
heating
&cntrl
imin=0,irest=0,ntx=1,
nstlim=25000,dt=0.002,
ntc=2,ntf=2,
cut=10.0, ntb=1,
ntpr=500, ntwx=500,
ntt=3, gamma_ln=2.0,
temp0=300.0, temp1=0.0,
ntr=1, restraintmask=':1-500',
restraint_wt=1.0,
nmropt=1
/
&wt TYPE='TEMP0', istep1=0, istep2=25000,
value1=0.1, value2=300.0, /
&wt TYPE='END' /
```

Obr. 4.19. Příklad (.in) souboru pro ohřev.

Cílem je postupný ohřev systému z 0 na 300 K. Užívá se poziční omezení na molekulu rozpuštěné látky. Vykonává se MD ($imin=0$), bez počátečních rychlostí ($ntx=1$). Soubor určuje simulaci počet kroků ($nstlim$) a časový krok (dt). Algoritmus SHAKE je aktivní pro vazby zahrnující vodík ($ntc=2$). Informace o souřadnicích se

zapisují do (.mdcrd) souboru po zvoleném počtu kroků (ntwx). Používá se Langevinův termostat (ntt=3), což vyžaduje definici parametru (gamma_ln). Je určena počáteční teplota (temp1) a konečná teplota, které má systém dosáhnout (temp0). Podmínky pozičního omezení určuje (nmropt=1).

Ekvilibrace – equil.in

```
equilibration
&cntrl
imin=0, irest=1, ntx=5,
nstlim=250000, dt=0.002,
ntc=2, ntf=2,
cut=10.0, ntb=2, ntp=1, taup=2.0,
ntpr=500, ntwx=500, ntwr=5000,
ntt=3, gamma_ln=2.0,
temp0=300.0,
/
```

Obr. 4.20. Příklad (.in) souboru pro ekvilibraci.

Předpis určuje konstantní tlakové podmínky a izotropické škálování pozic (ntp=1). Počáteční rychlosti jsou zapsány ve vstupním souřadnicovém souboru (irest=1). Konstantní tlak je určen (ntb =2). Teplota je udržována na 300 K pomocí Langevinova termostatu s časovou konstantou (taup). Algoritmus SHAKE je použit na omezení vodíků (ntc=2). Po určitém počtu kroků (ntwr) bude psán (.rst) soubor.

Produkční fáze

```
production dynamics
&cntrl
imin=0, irest=1, ntx=5,
nstlim=5000000, dt=0.002,
ntc=2, ntf=2,
cut=10.0, ntb=2, ntp=1, taup=2.0,
ntpr=1000, ntwx=1000, ntwr=50000,
ntt=3, gamma_ln=2.0,
temp0=300.0,
/
```

Obr. 4.21. Příklad (.in) souboru pro produkční fázi MD

Předpis pro produkční fázi obsahuje stejné parametry jako soubor equil.in, pouze má zvolen delší časový krok.

4.3. Práce s utilitami

Následující výčet utilit Gromacsu a Amberu není kompletním výčtem všech možností, které jsou velice široké. Zahrnuje však základní postupy a volby.

4.3.1. Gromacs

Pracuje se na příkazové řádce. Zadáním názvu utility, spolu s příkazovými volbami se vykoná požadovaný proces. Každá utilita v gromacsu s volbou `-h` (např. `editconf -h`) vypíše nápovědu programu, základní informace se zobrazí i zadáním názvu programu bez argumentů. Menu dostupných voleb ukazuje, jak je specifikovat, které jsou povinné, zda jsou parametry vstupní nebo výstupní, jejich defaultní hodnoty a povolené formáty souborů.

`:-) g_bond (-:`

Option	Filename	Type	Description
<code>-f</code>	<code>traj.xtc</code>	Input	Trajectory: xtc trr trj gro g96 pdb cpt
<code>-n</code>	<code>index.ndx</code>	Input	Index file
<code>-s</code>	<code>topol.tpr</code>	Input, Opt.	Structure+mass(db): tpr tpb tpa gro g96 pdb
<code>-o</code>	<code>bonds.xvg</code>	Output	xvgr/xmgr file
<code>-l</code>	<code>bonds.log</code>	Output, Opt.	Log file
<code>-d</code>	<code>distance.xvg</code>	Output, Opt.	xvgr/xmgr file

Option	Type	Value	Description
<code>-[no]h</code>	bool	no	Print help info and quit
<code>-[no]version</code>	bool	no	Print version info and quit
<code>-nice</code>	int	19	Set the nicelevel
<code>-b</code>	time	0	First frame (ps) to read from trajectory
<code>-e</code>	time	0	Last frame (ps) to read from trajectory
<code>-dt</code>	time	0	Only use frame when $t \text{ MOD } dt = \text{first time (ps)}$
<code>-[no]w</code>	bool	no	View output xvg, xpm, eps and pdb files
<code>-xvg</code>	enum	xmgrace	xvg plot formatting: xmgrace, xmgr or none
<code>-blen</code>	real	-1	Bond length. By default length of first bond
<code>-tol</code>	real	0.1	Half width of distribution as fraction of blen
<code>-[no]aver</code>	bool	yes	Average bond length distributions
<code>-[no]averdist</code>	bool	yes	Average distances (turns on -d)

Obr. 4.22. Možnosti utility g_bond.

4.3.1.1. Vybrané utility pro přípravu systému

pdb2gmx

Utilita čte souřadnicový soubor a generuje pro něj topologii a souřadnicový soubor (.gro). Doplnuje též strukturu chybějící vodíky. Po vyvolání se zobrazí nabídka silového pole a po jeho výběru se načtou příslušné soubory z databáze. Generuje se topologie pouze pro strukturu, která je složená z popsaných residuí. Pdb2gmx čte soubor s atomovými typy s koncovkou (.atp) a databázi residuí residuetypes.dat. Pokud se residuum nenajde, topologie se nevytvoří. Jsou metody, jak přidat vlastní residua do databáze, ale nejsou prováděny v rámci pdb2gmx. Volitelně program umí nastavit podmínky pro vazbu vodíku dvěma geometrickými kritérii – specifikováním maximální donor-akceptorové vzdálenosti a úhlu.

Nabídka silových polí:

```
Opening library file /afs/ics.muni.cz/software/gromacs-4.0.7/amd64_linux26/share
/gromacs/top/FF.dat

Select the Force Field:
0: GROMOS96 43a1 force field
1: GROMOS96 43a2 force field (improved alkane dihedrals)
2: GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
3: GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
4: GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
5: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
6: [DEPRECATED] Gromacs force field (see manual)
7: [DEPRECATED] Gromacs force field with hydrogens for NMR
8: Encad all-atom force field, using scaled-down vacuum charges
9: Encad all-atom force field, using full solvent charges
```

Obr.4.23. Nabídka silových polí Gromacs verze 4.0.

```

Select the Force Field:
From '/opt/gromacs-4.5.3/bin/share/gromacs/top':
 1: AMBER03 force field (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
 2: AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)
 3: AMBER96 force field (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)
 4: AMBER99 force field (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000)
 5: AMBER99SB force field (Hornak et al., Proteins 65, 712-725, 2006)
 6: AMBER99SB-ILDN force field (Lindorff-Larsen et al., Proteins 78, 1950-58, 20
10)
 7: AMBERGS force field (Garcia & Sanbonmatsu, PNAS 99, 2782-2787, 2002)
 8: CHARMM27 all-atom force field (with CMAP) - version 2.0
 9: GROMOS96 43a1 force field
10: GROMOS96 43a2 force field (improved alkane dihedrals)
11: GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
12: GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
13: GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
14: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)
15: [DEPRECATED] Encad all-atom force field, using full solvent charges
16: [DEPRECATED] Encad all-atom force field, using scaled-down vacuum charges
17: [DEPRECATED] Gromacs force field (see manual)
18: [DEPRECATED] Gromacs force field with hydrogens for NMR

```

Obr. 4.24. Nabídka silových polí verze 4.5.3 – došlo k rozšíření o CHARMM27 a pole programu Amber.

V dalším kroku program nabízí výběr modelu vody SPC, SPC/E, TIP3P, TIP4P, TIP5P.

g_x2top

Vytvoří jednoduchou topologii ze souřadnicového souboru. Program předpokládá, že jsou v systému všechny vodíky. Defaultně je zvoleno silové pole OPLS-AA, ale volitelně můžeme silové pole vybrat interaktivně, jako u pdb2gmx. Program umí také vytvořit (.rtp), které můžeme poté přidat do databáze residuí. Podporovaná silová pole jsou G53a5 GROMOS96 53a5 a OPLS-AA.

editconf

Program umí konvertovat strukturní formáty (.gro), (.g96) nebo (.pdb). Upravuje tvar, velikost a úhly boxu. Tvar může být buď kubický nebo trojklonný, kosočtverečný dvanáctistěn a komolý osmistěn (což jsou dva speciální případy trojklonného boxu). Volitelně umístí solut do středu boxu. Velikost boxu pro krychli, dvanáctistěn a osmistěn se udává jako jedna hodnota. Pro tyto boxy můžeme volitelně nastavit velikost podle solutu – největší vzdálenost mezi atomy plus dvakrát specifikovaná

vzdálenost, typicky se použije hodnota oříznutí potenciálu (cutoff) pro zamezení interakce solutu se svým periodickým obrazem. Pro trojklonný určíme velikosti os x, y a z. Pro tento typ boxu lze manipulovat i s velikostmi úhlů. Volitelně také lze rotovat souřadnicemi solventu, nebo požadovat, aby se osy systému zarovnaly podél os souřadnic, což umožňuje snížit objem boxu.

genbox

Genbox vytvoří box solventu s definovanou velikostí. Dále solvuje molekulu, přičemž kontroluje vzdálenosti mezi molekulami solvent-solvent a solvent-solut. Pokud jsou menší než definovaná suma van der Waalových poloměrů obou molekul, jsou molekuly solventu z boxu odstraněny. Tyto poloměry jsou čteny z databáze. Pro atomy, které nejsou v databázi je použita defaultní hodnota 0,105 nm. Hodnotu můžeme změnit. Program také může vložit zvolený počet molekul na náhodné pozice. Platí přitom stejná podmínka s poloměrem, jako v předchozím případě. Defaultním solventem je Simple Point Charge voda (SPC), se souřadnicemi v souboru spc216.gro. Program též umí molekulu solutu obalit vrstvou solventu zadané tloušťky. Program aktualizuje systémovou topologii, přidá informaci o počtu molekul solventu.

genion

Nahrazuje molekuly solventu jednoatomovými ionty. Specifikují se jména a počet iontů, náboj. Program umí neutralizovat systém nebo dosáhnout zvolené koncentrace soli (mol/l), nastavit minimální vzdálenost mezi ionty. Pokud se zvolí aktualizace topologie, zapíše se informace o počtech přidávaných iontů a sníží počet molekul solventu.

grompp

Utilita čte topologii a kontroluje její platnost. Také čte z parametrického souboru (počet kroků, časový krok, atd.). **Grompp** používá jména atomů z topologie. Pokud se tato jména neshodují se jmény v souřadnicovém souboru, je na tento nesoulad upozorněno.

4.3.1.2. *Výpočetní utilita*

mdrun

Je hlavním výpočetním programem Gromacsu. Provádí simulaci molekulární dynamiky, ale i stochastickou dynamiku, minimalizaci energie. Výstupem jsou nejméně čtyři soubory – trajektorie, souřadnice konečné struktury, soubor s časovým vývojem energie a dalších veličin a textový soubor shrnující průběh simulace společně s analýzou využití výpočetního času. **Mdrun** může provádět i paralelní výpočet na více vláknech. Lze nastavit dynamické vyvažování zátěže (automaticky je nastaveno, pokud je nerovnováha zatížení 5%) a řadu dalších voleb umožňujících zlepšit výkon.

4.3.1.3. *Vybrané utility pro analýzu*

g_energy

Utilita extrahuje data ze souboru (.edr), uživatel si vybere jaké informace chce zpracovat do formátu (.xvg), který si může zobrazit pomocí jiného software, např. Xmgrace. Soubory (.xvg) obsahují úvodní metapříkazy pro vytvoření legendy a popisu os grafu, ale jinak se jedná o textové soubory s prvním sloupcem obsahujícím čas a hodnotami zvolených veličin v dalších sloupcích.

trjconv

Slouží ke konvertování souborů trajektorie. Může vybrat podskupinu (ukládat jen atomy určitých skupin), časový interval (např. vynechat úvodní část odpovídající ekvilibraci), změnit časový krok snímků. Podporuje sedm formátů (.xtc), (.trr), (.trj), (.gro), (.g96), (.pdb) a (.g87).

make_ndx

Jako volitelný vstup pro téměř každou utilitu může sloužit indexový soubor. Programy defaultně vytvářejí určité indexové skupiny – např. pro celý systém, protein, residua, pro solvent. Použitím **make_ndx** lze vytvořit vlastní indexovou skupinu. Po spuštění

program zobrazí defaultní indexové skupiny systému. Pomocí editoru lze vytvořit novou skupinu podle čísla nebo typu atomu, residua nebo jména řetězce.

```
0 System          : 34440 atoms
1 Protein         : 1960 atoms
2 Protein-H      : 1001 atoms
3 C-alpha        : 129 atoms
4 Backbone       : 387 atoms
5 MainChain      : 517 atoms
6 MainChain+Cb   : 634 atoms
7 MainChain+H    : 646 atoms
8 SideChain      : 1314 atoms
9 SideChain-H    : 484 atoms
10 Prot-Masses   : 1960 atoms
11 non-Protein   : 32480 atoms
12 Water         : 32472 atoms
13 SOL          : 32472 atoms
14 non-Water     : 1968 atoms
15 Ion           : 8 atoms
16 CL-          : 8 atoms
17 Water_and_ions : 32480 atoms

nr : group      ! 'name' nr name  'splitch' nr  Enter: list groups
'a': atom      & 'del' nr      'splitres' nr 'l': list residues
't': atom type | 'keep' nr     'splitat' nr  'h': help
'r': residue   'res' nr     'chain' char
"name": group  'case': case sensitive  'q': save and quit
'ri': residue index
```

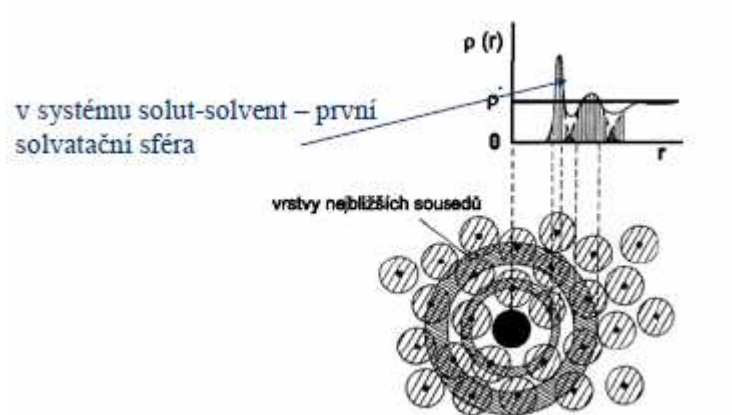
Obr. 4.25. Defaultní indexové skupiny pro systém lyzozym+voda, následované přehledem příkazů programu make_ndx.

g_bond

Nástroj pro analýzu vazeb. Při použití harmonického potenciálu by měla mít distribuce délek vazeb tvar Gaussovy distribuce.

g_rdf

Tato utilita počítá radiální distribuční funkci (RDF), která popisuje pravděpodobnost, že se v určité vzdálenosti bude nacházet atom nebo těžiště molekuly. Výsledek uloží ve formátu (.xvg) pro Xmgrace. Normální metoda je výpočet RDF mezi atomy dvou specifikovaných skupin (mohou být identické), jiná metoda je výpočet RDF kolem těžiště atomů či residuí. Pro takový proces je potřeba soubor (.tpr). Výpočet probíhá pouze pro atomy ze zvolené indexové skupiny. [8]



Obr.4.26. Schéma výpočtu radiální distribuční funkce a její typický tvar [6].

4.3.2. Amber

Příprava systému se provádí v programu LEaP. Můžeme pracovat na příkazové řádce – spustíme příkazem `tleap` nebo `xleap` pro práci v grafickém okně, kde můžeme systém zobrazit a editovat. Pro seznam dostupných příkazů stačí napsat `help`. Pro informace o vybraném příkazu zadáme např. `help` a příkaz, např. `help additions`. Pokud napíšeme příkaz chybně zobrazí se řádka se správnou syntaxí. Povinné argumenty jsou v ostrých závorkách, nepovinné v hranatých.

Vnitřní databáze silových polí, residuů a parametrů:

```
hermes.prf.jcu.cz$ ls $AMBERHOME/dat/leap/cmd
leaprc.ff02polEP.r1  leaprc.ff03.r1    leaprc.ff99SB    leaprc.gaff      oldff
leaprc.ff02pol.r1  leaprc.ff03ua    leaprc.ffAM1    leaprc.GLYCAM_04EP  README
leaprc.ff03        leaprc.ff99bsc0  leaprc.ffPM3    leaprc.GLYCAM_06
hermes.prf.jcu.cz$ █

ls $AMBERHOME/dat/leap/
```

Obr. 4.27. Nabídka silových polí v Amberu verze 10.

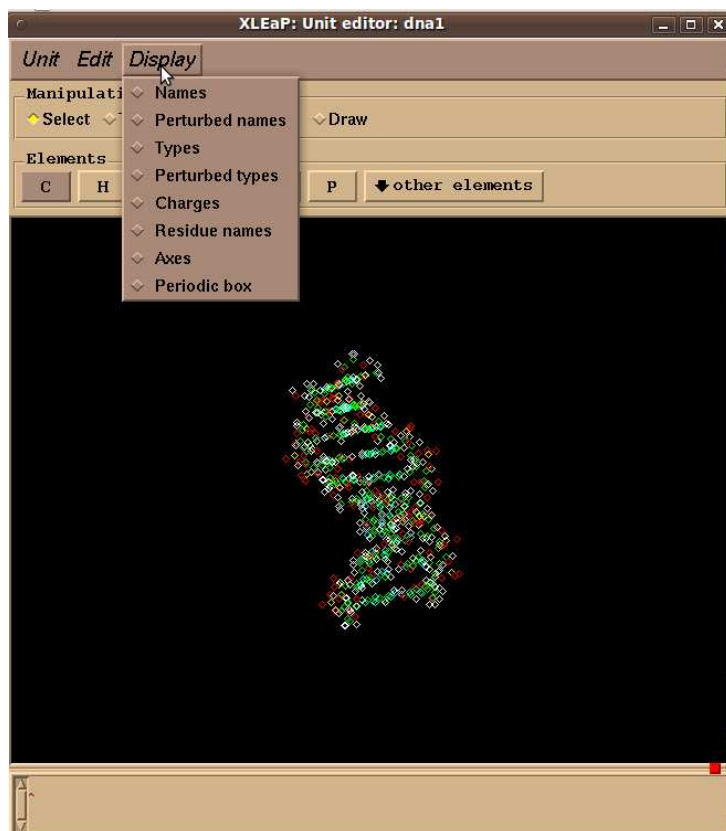
Výše diskutované programy **sander** a **antechamber** se spouštějí svými jmény. Pro **nab** je postup specifitější (viz část 4.4.4).

4.3.2.1. Vybrané utility pro přípravu systému

edit

Program **edit** funguje pouze pod `xleapem`. Zobrazí se grafické okno pro editaci systému či residua, což musíme definovat jako jednotku. Název jednotky se vždy zadává při

dalších procesech. Editor má nástroje pro výběr, přidávání a mazání molekul, zobrazení boxu, otáčení, přiblížení a oddálení, atd.



Obr. 4.28. Příklad práce v editoru – zobrazení části DNA

solvatebox, solvateoct, solvateshell

První dva příkazy vytvoří periodický box okolo solvatované jednotky. Tvar pro solvatebox je krychle, pro solvateoct je osmistěn. Velikost boxu je stanovena tak, aby vzdálenosti okrajů boxu od kteréhokoliv atomu solutu byla větší než zadaný parametr. Volitelně můžeme nastavit nejmenší vzdálenost mezi atomem solventu a atomem solutu. Defaultní hodnota je 1.0 [Å]. Vzdálenost by neměla být menší než suma van der Waalsových poloměrů atomů. Solvateshell obalí solut molekulami solventu do určené vzdálenosti. Výsledkem bude nepravidelný tvar sledující obrysy solutu.

addions, addions2, charge

Addions přidává ionty do systému zadáním jména, náboje a počtu. Umí také systém neutralizovat, pokud se zadá místo počtu 0. Pokud je přítomný solvent, je ignorován. Příkaz **addions2** solvent neignoruje. Příkaz funguje pro celá čísla (pokud má systém náboj -17,99 tak se přidá jen 17 kationtů, 18 se musí zadat ručně). Příkaz **charge** vypíše celkový náboj systému.

check

Je to utilita ke kontrole jednotky. Kontroluje vnitřní nejednotnosti, což by mohlo vést k problémům při výpočtech. Kontroluje dlouhé a krátké vazby, chybějící atomové typy, blízké kontakty (menší než 1,5 Å) mezi ne vazebnými atomy. Příkaz by se měl používat před uložením pomocí **saveamberparm**.

saveamberparm, savemol2, savepdb

První utilita ukládá amberovskou topologii a souřadnicový soubor pro definovanou jednotku. LEaP hledá ve své databázi parametry definující interakce mezi atomy. Druhá utilita uloží souřadnice ve formátu (.mol2), třetí uloží souřadnice ve formátu (.pdb).

loadamberparam, loadpdb, loadMol2

Příkazy k načtení příslušných souborů. V prvním případě parametrických, v druhém souřadnic ve formátu (.pdb), ve třetím souřadnic ve formátu (.mol2). [11]

4.3.2.2. Výpočetní utility

sander

Je hlavním výpočetním nástrojem Amberu. Vykonává minimalizaci energie, ekvilibraci a produkční fázi MD. LEaP produkuje jen dva druhy boxu – kubický a osmistěn, ale sander sám může pracovat i s jinými typy boxu. Particle-mesh Ewald (PME), nebo volitelně Ewaldova sumace se používá k přesnějšímu započtení dlouhodobých interakcí. Volitelně definujeme poziční omezení pro vazby, valenční úhly a torze. Umí provádět QM/MM výpočty. Provádí výpočet i pro implicitní solvent.

pmemd

Program obsahuje samostatnou a vylepšenou (výpočetně optimalizovanou) část programu sander pro provádění molekulární dynamiky s použitím particle-mesh Ewaldovy sumace.

4.3.2.3. Vybrané utility pro analýzu

ambmask

Tato utilita má funkci filtru, který z topologie a souřadnic vybere uživatelem zadané atomy nebo residua. Residua jsou vybírána podle jejich jmen nebo čísel. Atomy mohou být vybrány podle čísla, jména nebo atomového typu. Seznam vybraných atomů lze zobrazit ve třech možných variantách – amber, short, pdb. Pro výběr se užívá tzv. masky:

residua se určují :	:1-10	:	:1,3-5,9	:	:LYS, ALA, GLY
atomy se vybírají @	@12,17		@54-66		@CA, O, H
všechna residua :*	:*				
všechny atomy @*	@				
atomový typ @%	@%CT		@%N*		

Tento základní výběr může být kombinovaný ještě s operátory & (a), | (nebo), ! (negace) a operátory pro vzdálenost <:, >:, <@, >@ a závorkami. Operátor = vybere residua nebo atomy které začínají písmeny před ním, například :AS= vybere residua ASP a ASN. [10]

ptraj

Ptraj je nejen nástrojem pro analýzu trajektorie, ale i nástrojem k přehlednému čtení z topologického souboru. Lze si nechat vypsát přehledně ve sloupcích informace o atomech, vazbách, úhlech, torzích atd. Nápověda zobrazí dostupné utility. Příklad:

```
rdparm fenol.prmtop  
printAtoms
```

```

RDPARM MENU: printAtoms
Mask [*] represents 13 atoms
Number: Atom      Charge Mass ( Residue ) Type Tree
1:  C1    -0.15660 12.0 ( 1:MOL ) ca   BLA
2:  H1     0.14990  1.0 ( 1:MOL ) ha   BLA
3:  C2    -0.09740 12.0 ( 1:MOL ) ca   BLA
4:  H2     0.13320  1.0 ( 1:MOL ) ha   BLA
5:  C3    -0.16500 12.0 ( 1:MOL ) ca   BLA
6:  H3     0.13420  1.0 ( 1:MOL ) ha   BLA
7:  C4    -0.09190 12.0 ( 1:MOL ) ca   BLA
8:  H4     0.13160  1.0 ( 1:MOL ) ha   BLA
9:  C5    -0.21320 12.0 ( 1:MOL ) ca   BLA
10: H5     0.13270  1.0 ( 1:MOL ) ha   BLA
11: C6     0.12310 12.0 ( 1:MOL ) ca   BLA
12: O1    -0.49880 16.0 ( 1:MOL ) oh   BLA
13: H6     0.41830  1.0 ( 1:MOL ) ho   BLA

```

Obr. 4.29. Seznam atomů v souboru *fenol.prmtop*, zobrazení pomocí *ptraj*.

Ptraj umí konvertovat trajektorie, počítat délky vazeb, velikosti úhlů, torzí, hodnoty odchylek RMSD a další charakteristiky. Jako vstup slouží topologický soubor (*.prmtop*), skript (*.in*), který v sobě obsahuje příkaz pro načtení trajektorie a následuje řada uživatelem definovaných příkazů – dle toho, jaké vlastnosti chceme analyzovat. V příkazech lze určit, s jakým časovým krokem se budou data analyzovat, jména výstupních souborů a také jaké atomy či residua do naší analýzy chceme zahrnout. Pro výběr atomů a residuí se používá stejné značení jako v **ambmask**. [11]

Příklad skriptu pro analýzu:

```

trajin protein_md2.mdcrd
radial rdf 0.1 20 :WAT@O
distance vzdalenost :1 :2 out vzdalenost.dat time 50
angle @1 @2 @3 out angle.dat time 10

```

4.4. Modelové úlohy

V této kapitole dokumentuji možnosti nástrojů programů Amber a Gromacs v typických jednoduchých úlohách. Úlohy zahrnují přípravy a spuštění simulace systému.

4.4.1. Simulace boxu s vodou

Posutp v Gromacsu:

1. Vytvoření kubického boxu s vodou o velikosti hrany 2.4 nm, zaplněného SPC vodou

```
genbox -cs spc216.gro -box 2.4 -o w.gro
```

2. Generování topologie s interaktivním výběrem silového pole OPLS-AA

```
pdb2gmx -f w.gro -p w.top -o w.gro
```

3. Předzpracování s parametry k minimalizaci+minimalizace

```
a) grompp -c w.gro -p w.top -o minim.tpr -f minim.mdp
```

```
b) mdrun -s minim.tpr -v -deffnm minim
```

4. Předzpracování s parametry NVT souboru+krátký běh MD

```
a) grompp -c minim.gro -p w.top -o nvt.tpr -f nvt.mdp
```

```
b) mdrun -s nvt.tpr -v -deffnm nvt
```

5. Předzpracování s parametry NPT souboru+krátký běh MD

```
a) grompp -c nvt.gro -p w.top -o npt.tpr -f npt.mdp
```

```
b) mdrun -s npt.tpr -v -deffnm npt
```

5. Předzpracování s parametry pro produkční fázi +produkční běh MD

```
a) grompp -c npt.gro -p w.top -o md.tpr -f md.mdp
```

```
b) mdrun -s md.tpr -v -deffnm md
```

Postup v Amberu:

1. Načtení parametrů a silového pole ff99SB

```
xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99SB
```

2. Načtení residua TP3 (molekula vody TIP3P) do jednotky sys

```
sys = copy TP3
```

3. Solvatace kubického boxu s parametrem 15 Å vodou TIP3P

```
solvatebox sys TIP3PBOX 15
```

4. Vytvoření amberovské topologie a souřadnic

```
saveamberparm sys voda.prmtop voda.inpcrd
```

5. Minimalizace energie systému

```
sander -O -i min2.in -p voda.prmtop -c voda.inpcrd -o voda_min.out -r voda_min.rst
```

6. Ohřev z 0 na 300 K

```
sander -O -i heat.in -p voda.prmtop -c voda_min.rst -o voda_heat.out -r voda_heat.rst -x voda_heat.mdcrd
```

7. Ekvilibrace systému

```
sander -O -i equil.in -p voda.prmtop -c voda_heat.rst -o voda_equil.out -r voda_equil.rst
```

```
-x voda_equil.mdcrd
```

8. Produkční fáze MD

```
sander -O -i prod.in -p voda.prmtop -c voda_equil.rst -o voda_prod.out -r voda_prod.rst  
-x voda_prod.mdcrd
```

4.4.2. Simulace boxu s vodou a ionty

Postup v Gromacsu:

1. Vytvoření kubického boxu s vodou o hraně 2.4 nm, zaplněného SPC vodou

```
genbox -cs spc216.gro -box 2.4 -o i.gro
```

2. Generování topologie s interaktivním výběrem silového pole OPLS-AA

```
pdb2gmx -f i.gro -p i.top -o i.gro
```

3. Přidání 3 iontů Na⁺ a 3 iontů Cl⁻ namísto molekul vody (výběr skupiny 2)

```
a) grompp -f minim.mdp -c i.gro -p i.top -o i.tpr
```

```
b) genion -s i.tpr -o ionty.gro -p i.top -pname NA -np 3 -nname CL -nn 3
```

4. Provedení minimalizace energie

```
a) grompp -c ionty.gro -p i.top -o minim.tpr -f minim.mdp
```

```
b) mdrun -s minim.tpr -v -deffnm minim
```

5. NVT ekvilibrace

```
a) grompp -c minim.gro -p i.top -o nvt.tpr -f nvt.mdp
```

```
b) mdrun -s nvt.tpr -v -deffnm nvt
```

6. NPT ekvilibrace

```
a) grompp -c nvt.gro -p i.top -o npt.tpr -f npt.mdp
```

```
b) mdrun -s npt.tpr -v -deffnm npt
```

7. Provedení produkční fáze MD

```
a) grompp -c npt.gro -p i.top -o md.tpr -f md.mdp
```

```
b) mdrun -s md.tpr -v -deffnm md
```

Postup v Amberu:

1. Načtení parametrů a silového pole ff99SB

```
xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99SB
```

2. Načtení residua TP3 jako jednotky sys

```
sys = copy TP3
```

3. Solvatace kubického boxu s parametrem 15 Å

```
solvatebox sys TIP3PBOX 15
```

4. Přidání iontů

```
addions2 sys Na+ 3 Cl- 3
```

5. Vytvoření amberovské topologie a souřadnic

```
saveamberparm sys ions.prmtop ions.inpcrd
```

6. Minimalizace energie

```
sander -O -i min2.in -p ions.prmtop -c ions.inpcrd -o ions_min.out -r ions_min.rst
```

7. Ohřev systému

```
sander -O -i heat.in -p ions.prmtop -c ions_min.rst -o ions_heat.out -r ions_heat.rst -x ions_heat.mdcrd
```

8. Ekvilibrace

```
sander -O -i equil.in -p ions.prmtop -c ions_heat.rst -o ions_equil.out -r ions_equil.rst -x ions_equil.mdcrd
```

9. Produkční fáze

```
sander -O -i prod.in -p ions.prmtop -c ions_equil.rst -o ions_prod.out -r ions_prod.rst -x ions_prod.mdcrd
```

4.4.3. Simulace lysozymu

Pro simulaci enzymu lysozymu získáme počáteční strukturu 1AKI.pdb na stránkách Protein Data Bank. (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1AKI>)

Postup v Gromacsu

Tutoriál pro tuto simulaci v programu Gromacs je na stránkách

(<http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/lysozyme/>)

1. Načtení parametrů, interaktivní výběr silového pole OPLS-AA a generování topologie pro systém lysozym+voda SPCE

```
pdb2gmx -f 1AKI.pdb -o 1AKI_processed.gro -water spce
```

2. Vytvoření kubického boxu s lysozymem uprostřed

```
editconf -f 1AKI_processed.gro -o 1AKI_newbox.gro -c -d 1.0 -bt cubic
```

3. Solvatace lysozymu SPC vodou, aktualizování topologie

```
genbox -cp 1AKI_newbox.gro -cs spc216.gro -o 1AKI_solv.gro -p topol.top
```

4. Přidání 8 iontů Cl⁻ pro neutralizaci (výběr skupiny 13), aktualizování topologie

```
a) grompp -f minim.mdp -c 1AKI_solv.gro -p topol.top -o ions.tpr
```

b) genion -s ions.tpr -o 1AKI_solv_ions.gro -p topol.top -nname CL -nn 8

5. Provedení minimalizace energie

a) grompp -f minim.mdp -c 1AKI_solv_ions.gro -p topol.top -o em.tpr

b) mdrun -v -deffnm em

6. Ekvilibrace systému, nejprve s parametry NVT, poté s parametry NPT

a) grompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr

b) mdrun -deffnm nvt

c) grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr

d) mdrun -deffnm npt

7. Provedení produkčního běhu

a) grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr

b) mdrun -deffnm md_0_1

Pro paralelní výpočet na 4 vláknech

b) mpirun -n 4 mdrun -deffnm md_0_1

Postup v Amberu

Čerpala jsem z tutoriálu na stránkách

(http://enzyme.fbb.msu.ru/Tutorials/Tutorial_1/Sec_1.htm)

Pro tento úkol je potřeba pozměnit v souboru min1.in čísla residuí v restraintmask – úloha je nastavena pro bílkovinu složenou z 500 residuí, lysozym má residuí 129.

1. Načtení silového pole ff99SB

xleap -s -f \$AMBERHOME/dat/leap/cmd/leaprc.ff99SB

2. Načtení lysozymu v (.pdb) souboru do jednotky „sys“

sys = loapdb 1AKI.pdb

3. Solvatace jednotky sys v kubickém boxu s parametrem 15 Å a rozpouštědlem TIP3P

solvatebox sys TIP3PBOX 15

4. Přidání iontů pro dosažení neutrálního systému

addions sys Cl- 0

5. Vytvoření topologie a souřadnic

saveamberparm sys protein.prmtop protein.inpcrd

Pokud chceme systém zobrazit, vytvoříme (.pdb) v tleap

```
ambpdb -p protein.prmtop < protein.inpcrd > protein_solvated.pdb
```

6. Minimalizace energie s fixovaným proteinem na svém místě

```
sander -O -i min1.in -p protein.prmtop -c protein.inpcrd -o protein_min1.out -r protein_min1.rst  
-ref protein.inpcrd
```

7. Minimalizace celého systému

```
sander -O -i min2.in -p protein.prmtop -c protein_min1.rst -o protein_min2.out -r protein_min2.rst
```

8. Ohřev z 0 na 300 K, poziční omezení proteinu

```
sander -O -i heat.in -p protein.prmtop -c protein_min2.rst -o protein_heat.out -r protein_heat.rst  
-x protein_heat.mdcrd -ref protein_min2.rst
```

9. Ekvilibrace za konst. tlaku

```
sander -O -i equil.in -p protein.prmtop -c protein_heat.rst -o protein_equil.out -r protein_equil.rst  
-x protein_equil.mdcrd
```

10. Spuštění produkční fáze MD

```
sander -O -i prod.in -p protein.prmtop -c protein_equil.rst -o protein_prod.out -r protein_prod.rst  
-x protein_prod.mdcrd
```

4.4.4. *Propojení sekvencí*

Postup v Gromacsu

Nenalezla jsem utilitu, která by umožňovala propojování aminokyselin či vytvořila sekvenci bazí.

Postup v Amberu

Aminokyseliny jako residua je možné propojit v LEaP příkazem `sequence`. Seznam residuí pro dané silové pole obdržíme příkazem `list` (po načtení silového pole).

```
xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99SB  
pept = sequence { ALA GLY PRO }  
saveamberparm pept pept.prmtop pept.inpcrd
```

Sekvenci bazí vytvoříme pomocí programu NAB. K tomu potřebujeme nejdříve soubor, který bude v jazyku NAB popisovat molekulu. Lze jej vytvořit v textovém editoru, uložit soubor jako (.nab) nebo stáhnout a editovat, např. z (<http://ambermd.org/tutorials/basic/tutorial1/section2.htm>).

```
molecule m;  
  
m = bdna( "cgcg" );  
putpdb( "cg4.pdb", m );  
exit( 0 );
```

Spuštěním programu NAB s uvedením jména souboru popisujícího molekulu

```
nab cg4.nab
```

vznikne spustitelný soubor a.out, který po jeho provedení

```
./a.out
```

vygeneruje soubor (.pdb). Pomocí xleapu můžeme molekulu zobrazit a editovat.

```
xleap -s -f $AMBERHOME/dat/leap/cmd/leaprc.ff99SB  
sys = loadpdb cg4.pdb
```

NAB vytvoří kompletní dvojšroubovici včetně komplementárních bazí. Pokud chceme ponechat jen jedno vlákno, nejsnadnější způsob je v editoru xleapu druhé vlákno smazat.

5. Diskuze

Při řešení modelových úloh jsem porovnávala práci v programech a měla jsem možnost srovnání jednotlivých utilit. Oba programy obsahují příkazy pro přípravu systému (generování topologie a vstupních souřadnic), utilitu přidávající ionty do systému nebo k jeho neutralizaci, solvataci různými způsoby s určením druhu rozpouštědla (v Gromacsu je defaultní voda SPC, v Amberu TIP3P), výpočetní utility, analyzační programy k analýze vazeb, úhlů, trajektorie, odchylek a další. Programy také nabízejí možnost aplikování pozičního omezení např. na molekulu rozpuštěné látky. Mnoho utilit má tedy podobnou funkci:

Gromacs	Amber
pdb2gmx	saveamberparm (LEaP)
x2top	antechamber
genion	addions2 (LEaP)
genbox -box	solvatebox (LEaP)
genbox -shell	solvateshell (LEaP)
make_ndx	ambmask
mdrun	sander
trjconv	trajout (ptraj)
g_bond	bond (ptraj)
g_angle	angle (ptraj)
g_rms	rms (ptraj)
g_rdf	radial (ptraj)

Patrnými rozdíly mezi programy jsou jejich struktura a vyjádření některých fyzikálních veličin (viz tab. 1.2). Gromacs má ucelenou formu, kdy utility pro přípravu systému, simulaci i analýzu voláme stejným způsobem z příkazové řádky.

Amber je rozdělen na několik částí - na přípravné, simulační a analyzační. V rámci každé této jednotlivé části Amber obsahuje několik utilit. Utility se také spouštějí z příkazové řádky, ale právě v rámci jedné části, např. LEaP, antechamber.

6. Závěr

Cílem této práce bylo porovnat programy Amber a Gromacs, které slouží k simulaci molekulární dynamiky. V obou programech jsem pro porovnání a diskusi jejich možností simulovala systémy: 1) voda 2) voda s ionty 3) solvatovaný lysozym s ionty.

U Gromacsu je jeho největší předností rychlost a přehlednost výpočetní části mdrun. Volitelně lze v terminále sledovat, kolik času zbývá do dokončení simulace a v jakém časovém kroku se simulace nachází. Gromacs nabízí oproti Amberu širší výběr silových polí, utilitu `make_ndx`, která pomáhá přehledně pracovat jen se skupinou atomů – buď s jednou z defaultních skupin nebo s uživatelem specifikovanou skupinou. Utilita `editconf` nabízí větší výběr solvatačních boxů. Program `genion`, sloužící k přidávání iontů do systému, nahrazuje ionty za molekuly vody oproti amberovskému `addions`, který pouze do systému ionty přidá.

Velikou výhodou Gromacsu jsou jeho možnosti pracovat se systémy obsahujícími periodická rovinná rozhraní – např. vodné roztoky molekul interagující s anorganickými nebo organickými povrchy (lipidové dvojvrstvy, povrchy membrán). Gromacs nabízí možnost výpočtu elektrostatických interakcí modifikovanou Ewaldovou sumací pro 2D periodické systémy. Tomu odpovídají i možnosti programů pro analýzu (např. `g_density`, `g_potencial`), které poskytují informace o průběhu veličin jako funkce vzdálenosti od povrchu.

Za nevýhodu Gromacsu považuji formu utilit – každá utilita má mnoho voleb, ve kterých se hůře orientuje. V Gromacsu se před výpočtem musí provést o jeden krok navíc, kdy je nutné sestavit vstupní binární soubor (`.tpr`) pro mdrun.

Pro uživatele, který se seznamuje s molekulární dynamikou, považuji Amber za vhodnější. Program má jasně definované části. Funkce utilit v přípravné části je většinou jen jedna a nemají tolik volitelných parametrů jako v Gromacsu. V přípravné části nabízí Amber grafické prostředí pro editaci systému. Pomocí `xleap` a `NAB` lze sestavit vstupní struktury, což Gromacs neumožňuje.

Za nevýhodu Amberu považuji nepřehlednost utility `ptraj`, která má dvě funkce, ale nápověda se vztahuje pouze k `rdparm`.

Za nevýhodu obou programů považuji, že si uživatel v určitém okamžiku jen s těmito programy nevystačí – u Gromacsu je potřeba jiný software, abychom mohli strukturu zobrazit. Oba programy obsahují analyzační utility, které zpracovávají data do výstupních formátů, které se mohou vykreslit do grafu, ale žádná utilita pro výrobu grafů v nich obsažená není. Uživatel musí dohledávat externí skripty či použít jiný program.

Programy nabízejí opravdu velmi široké možnosti. Popis všech utilit a přehled všech formátů, které jsou potřebné k práci s těmito programy by jistě byl přínosný, ale přesáhl by rámec této práce. Ke kompletnímu výčtu možností těchto programů také patří to, že Amber i Gromacs vykonávají kromě klasických i QM/MM výpočty.

7. Seznam použité literatury

1. NEZBEDA, Ivo; KOLAFA, Jiří; KOTRLA, Miroslav. *Úvod do počítačových simulací, Metody Monte Carlo a molekulární dynamiky*. Praha: Univerzita Karlova Karolinum, 1998, ISBN - neuvedeno

Internetové zdroje:

2. Počítačová chemie, Masarykova univerzita Brno
http://ncbr.chemi.muni.cz/~n19n/vyuka/pocitacova_chemie/new/ [2010-01]
3. HUML, K.; VONDRÁŠEK, J.; *Tři přístupy k počítačovému modelování struktury a interakcí velkých molekul*
<http://www.xray.cz/ms/bul2000-1/huml.pdf> [2010-01]
4. Stránky programu CHARMM
http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html [2010-03]
5. HRACH, Rudolf; *Počítačová fyzika I*
physics.ujep.cz/~mmaly/vyuka/poc_fyz_1/pocitacova_fyzika_1.pdf [2010-01]
6. Katedra fyzikální chemie, Univerzita Palackého v Olomouci
http://fch.upol.cz/skripta/chs/qm_V.pdf [2010-02]
7. <http://isaacs.sourceforge.net/phys/pbc.html> [2011-04]
8. Uživatelský manuál Gromacs
<http://www.gromacs.org/Documentation/Manual> [2010-03]
9. http://en.wikipedia.org/wiki/Molecular_dynamics [2010-05]
10. Uživatelský manuál Amber
<http://ambermd.org/doc11/Amber11.pdf> [2011-04]
11. Uživatelský manuál AmberTools
<http://ambermd.org/doc11/AmberTools.pdf> [2011-04]

8. Klíčová slova

Simulace molekulární dynamiky

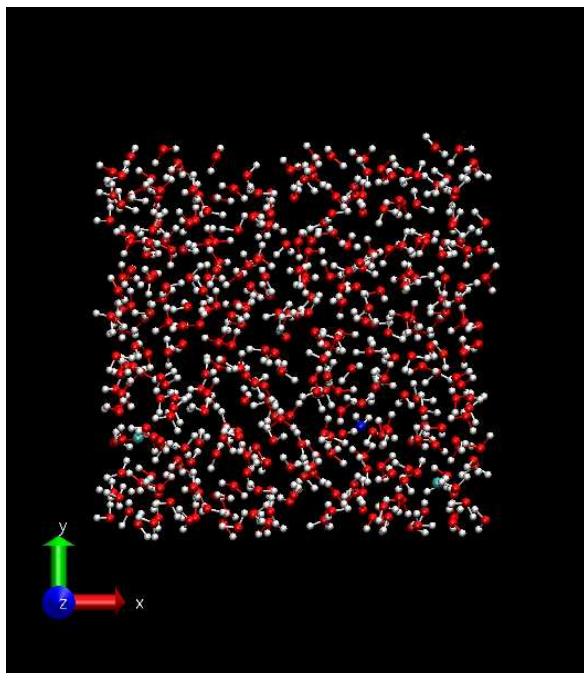
Amber

Gromacs

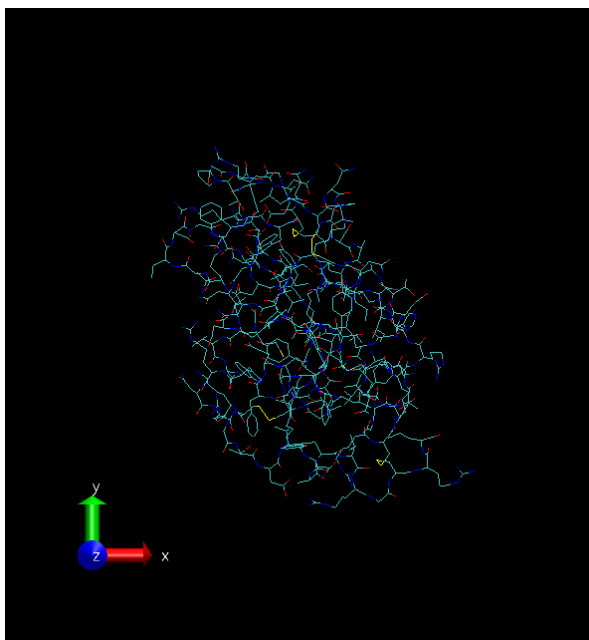
Silové pole

Potenciálová funkce

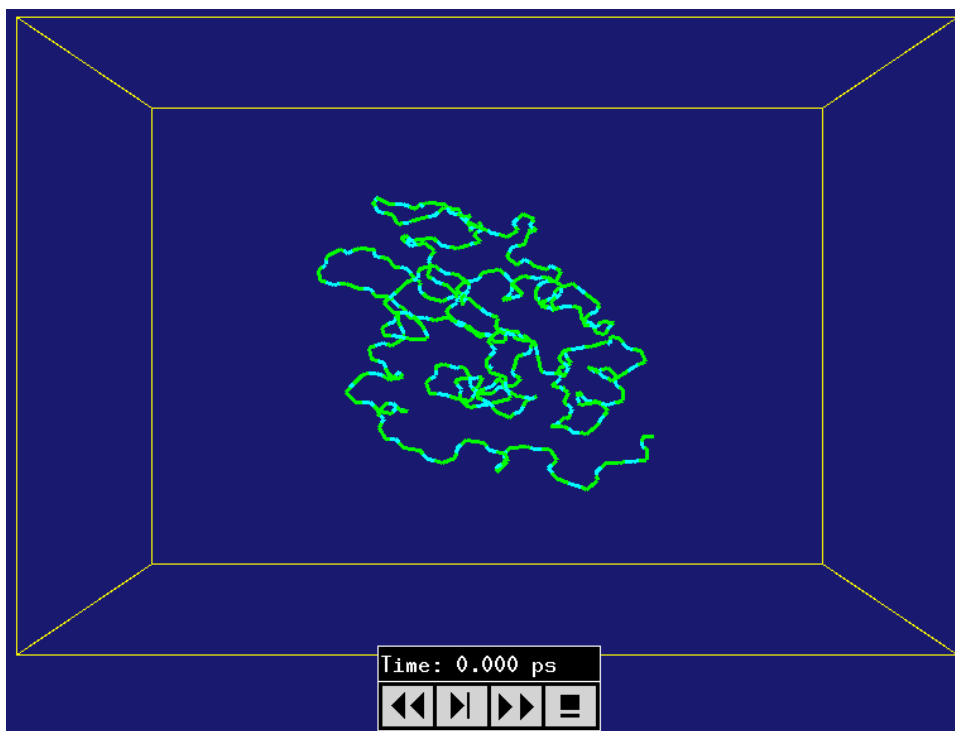
9. Přílohy



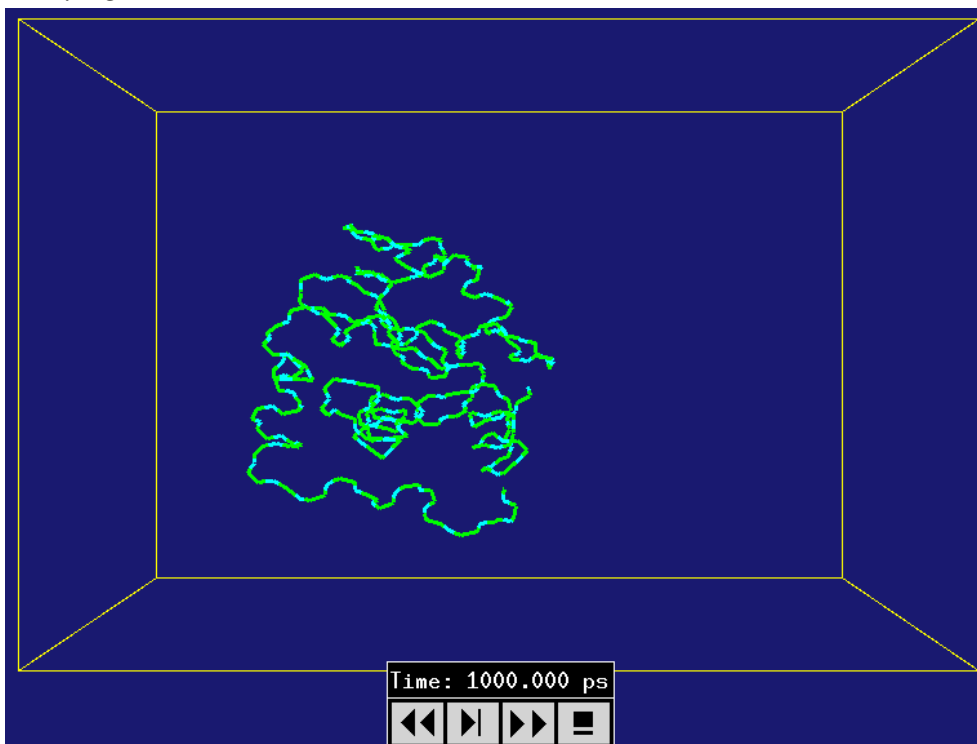
Obr. 9.1 Kubický box naplněný vodou s ionty, zobrazené pomocí VMD. Ionty jsou značeny modře a zeleně.



Obr. 9.2 Molekula lysozymu – zobrazeno pomocí VMD.



Obr 9.3. Molekula lysozymu na počátku simulace, zobrazeno pomocí gromacsovské utility ngmx.



Obr. 9.4. Molekula lysozymu na konci 1000 ps simulace, zobrazeno pomocí gromacsovské utility ngmx.