

Univerzita Karlova
Přírodovědecká fakulta

Studijní program: Chemie
Studijní obor: Chemie



Vojtěch Keprta

Modelování acidobazických vlastností kopolyptidů glutamátu a
lysinu s tyrosinem

Modeling of acid-base properties of copolypeptides of glutamate
and lysine with tyrosine

Vedoucí práce: doc. RNDr. Peter Košovan, Ph.D.

Praha, 2024

Chtěl bych poděkovat doc. RNDr. Peteru Košovanovi, Ph.D. za vedení práce a podporu a odborné rady při jejím zpracování.

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 20. 8. 2024

Vojtěch Keprta

Abstrakt: Polypeptidy jsou základní součástí biologických systémů. Pokud obsahují aminokyseliny s kyselými nebo zásaditými postranními řetězci, mohou v důsledku disociačních reakcí v závislosti na pH nést náboj. Polypeptidy tak mohou být považovány za slabé polyelektrolyty, a jejich pH-responzivní chování, které je důležité z hlediska popisu jejich biologického fungování i potencionálních aplikací, může být studováno v tomto kontextu. Molekulové simulace specificky mají v tomto ohledu dlouhou historii. V této práci je vytvořen zhrubený počítačový model polypeptidu a je zkoumáno, jak parametry modelu jako úroveň detailu v něm nebo parametry systému jako délka řetězce a koncentrace soli ovlivňují chování polypeptidu. Cílem je získat výsledky ze simulací, které by v budoucnu mohli být porovnány s experimentálními výsledky, aby šlo určit, zda je vytvořený model vhodný pro studium těchto systémů i nadále.

Klíčová slova: Polypeptid, slabý polyelektrolyt, počítačová simulace, disociace, modelování

Abstract: Polypeptides are an essential part of biological systems. If they contain amino acids with acidic or basic side chains, they can carry a charge due to dissociation reactions depending on the pH. Polypeptides can thus be considered weak polyelectrolytes, and their pH-responsive behavior, which is important in terms of describing their biological functioning as well as potential applications, can be studied in this context. Molecular simulations in particular have a long history in this regard. In this work, a coarse-grained computer model of the polypeptide is created and it is investigated how parameters of the model such as the level of detail in it or system parameters such as chain length and salt concentration affect the behavior of the polypeptide. The goal is to get results from simulations that could later be compared with experimental results in order to determine whether the created model is suitable for the study of these systems in the future.

Keywords: Polypeptide, weak polyelectrolyte, computer simulation, dissociation, modelling

Obsah

1	Motivace	2
2	Teoretický Úvod	3
2.1	Acidobazické rovnováhy	3
2.1.1	Kyseliny a zásady	3
2.1.2	Stupeň ionizace kyselin a zásad	5
2.2	Polypeptidy	6
2.2.1	Aminokyseliny a peptidová vazba	6
2.2.2	Gyrační poloměr polypeptidu	9
2.2.3	Polypeptid jako polyelektrolyt	9
3	Metodická část	11
3.1	Model polypeptidu	11
3.1.1	Reprezentace aminokyseliny v použitém modelu	11
3.1.2	Reprezentace ostatních složek systému v použitém modelu	12
3.1.3	Interakce v použitém modelu	12
3.2	Molekulové simulace	13
3.2.1	Molekulová dynamika	14
3.2.2	Monte Carlo	16
3.2.3	Simulační box	17
3.2.4	Jednotky v simulaci	18
3.2.5	Sběr a zpracování dat	18
3.3	Simulované systémy	19
3.4	Simulační protokol	20
4	Výsledky a diskuze	21
4.1	Polypeptid EEEY	24
4.2	Polypeptid KKKY	28
4.3	Vliv zvoleného modelu	32
4.4	Vliv délky řetězce	38
4.5	Vliv koncentrace přidané soli	43
5	Závěr a výhled do budoucna	49
	Literatura	50

1 Motivace

Peptidy jsou biomolekuly tvořené aminokyselinami propojenými peptidovou vazbou. Konkrétně se tato práce zaměřuje na polypeptidy, což jsou specificky peptidy tvořené deseti nebo více aminokyselinami. V biologických systémech peptidy hrají důležitou roli jako hormony, neurotransmitery nebo jako součást imunitního systému [1]. Proteiny, tvořené jedním nebo více dlouhými polypeptidovými řetězci, jsou pak základní složkou všech živých organismů, kde fungují jako biokatalyzátory, strukturní prvky a mají mnoho dalších funkcí.

Náboj polypeptidů a jejich prostorová konformace jsou zásadními faktory ovlivňujícími jejich interakce s ostatními molekulami [2], což je podstatné např. pro vývoj léčiv [3], přičemž tyto dvě veličiny jsou vzájemně provázané. Náboj polypeptidu je pak obecně dán stupněm ionizace aminokyselin, ze kterých se skládá. Specificky jde primárně o ionizaci jejich postranních řetězců a to u těch aminokyselin, jejichž postranní řetězce obsahují skupiny chovající se jako slabé kyseliny nebo zásady. Disociace slabých kyselin a zásad je řízena pH systému a pro ideální systém (bez mezimolekulových interakcí) ji lze jednoduše analyticky popsat známou Hendersonovou-Hasselbalchovou rovnicí (rov. 2.20 a 2.21).

V reálných systémech však existují mezi náboji vzniklými disociací kyselin a zásad dalekosahové elektrostatické interakce. Tyto interakce nelze zanedbat obzvláště pro náboje nacházející se na polypeptidovém řetězci v relativně malé vzdálenosti od sebe. Vlivem těchto interakcí dochází k odchýlení od ideálního chování a ke zvýšení nebo snížení stupně ionizace aminokyselin v rámci regulace náboje polypeptidu [4]. Teoretický popis těchto jevů je často omezen jen na systémy se specifickou geometrií nebo je jeho přesnost snížena zavedením aproximací pro umožnění výpočtů. Ke studiu takových systémů jsou pak vhodné molekulové simulace, které stojí na pomezí mezi teorií a experimentem. V této práci jsou používány zhrubené (coarse-grained) simulační modely. Tyto modely reprezentují reálný systém v nižším detailu než atomistické modely, což zvyšuje výpočetní účinnost simulací. Otázkou je pak, do jaké míry dokáže i tento zjednodušený model správně popsat reálné chování systému, respektive jak velké zjednodušení si můžeme dovolit.

Cílem této práce je vytvořit zhrubený simulační model polypeptidového řetězce v roztoku, který by mohl sloužit k předpovídání změny náboje polypeptidu a jeho konformace v závislosti na pH, včetně odchylek od ideality v důsledku elektrostatických interakcí. Konkrétně jsou v této práci modelovány kopolypeptid aminokyselin glutamové kyseliny a tyrosinu a kopolypeptid lysinu a tyrosinu. Důvodem pro zvolení těchto dvou polypeptidů je možnost získání experimentálních dat studiem dostupných reálných polypeptidových vzorků s tímto složením v rámci výzkumné skupiny. Tyrosin je fluorofor, což umožňuje sledování jeho stupně ionizace pomocí spektroskopie [5]. Porovnání simulačních výsledků z této práce s, v budoucnu získanými, experimentálními výsledky by pak mohlo vést k validaci modelu nebo být podmětem k jeho dalším úpravám.

2 Teoretický Úvod

2.1 Acidobazické rovnováhy

Obecné zákonitosti řídící disociaci slabých kyselin a zásad jsou základem pro popis ionizace aminokyselin v molekule polypeptidu.

2.1.1 Kyseliny a zásady

Kyseliny a zásady jsou příkladem elektrolytů. Elektrolyty jsou chemické látky, které ve vodném roztoku disociují na ionty. Lze je rozdělit na silné a slabé podle toho, jestli jsou v roztoku plně nebo jen částečně disociovány.

Kyseliny a zásady lze definovat podle Brøntedovy-Lowryho teorie následovně. Kyselina (HA) je chemická látka schopná odštěpit proton a rovnice pro její disociaci je pak



Zásada (B) je naopak látkou, která je proton schopná přijmout.



Molekuly vody jako takové také disociují podle rovnice autoprotolýzy vody



Disociační konstanta

Obecně je chemická rovnováha pro konkrétní systém definována podmínkou [6]

$$\sum_i \nu_i \mu_i = \sum_i \nu_i (\mu_i^{ref} + \mu_i^{id} + \mu_i^{ex}) = 0, \quad (2.4)$$

kde ν_i je stechiometrický koeficient složky i , μ_i je chemický potenciál složky i , který lze rozložit na chemický potenciál referenčního stavu (μ_i^{ref}), chemický potenciál ideálního plynu (μ_i^{id}) a chemický potenciál daný intermolekulárními interakcemi v systému (μ_i^{ex}). Ideální chemický potenciál pro složku v roztoku je pak definován jako [6]

$$\mu_i^{id} = k_B T \ln \left(\frac{c_i}{c^{ref}} \right), \quad (2.5)$$

kde c_i je koncentrace složky i a c^{ref} je referenční koncentrace (typicky 1 mol/dm³), k_B je Boltzmanova konstanta a T je teplota. Pro popis systémů odchylojících se od ideálního chování ($\mu_i^{ex} \neq 0$) je zavedena aktivita (a) ve tvaru

$$a_i = \gamma_i \left(\frac{c_i}{c^{ref}} \right), \quad (2.6)$$

kde γ_i je aktivitní koeficient, který má pro ideální stav hodnotu 1. Chemický potenciál i -té složky z rovnice (2.4) pak lze vyjádřit jako

$$\mu_i = \mu_i^{ref} + k_B T \ln(a_i) = \mu_i^{ref} + k_B T \ln \left(\frac{c_i}{c^{ref}} \right) + k_B T \ln(\gamma_i). \quad (2.7)$$

Pro popis chemické rovnováhy je používána rovnovážná konstanta K definována jako

$$\ln(K) = -\frac{1}{k_B T} \sum_i \nu_i \mu_i^{ref}. \quad (2.8)$$

S použitím rovnic 2.7 a 2.4 ji lze vyjádřit ve tvaru [4]

$$K = \prod_i a_i^{\nu_i} = (c^{ref})^{-\bar{\nu}} \prod_i (\gamma_i c_i)^{\nu_i}, \quad (2.9)$$

kde $\bar{\nu} = \sum_i \nu_i$.

Pro charakterizaci kyseliny nebo zásady je typicky používána aciditní disociační konstanta (K_A), což je rovnovážná konstanta reakce 2.1 resp. 2.2. U kyselin i zásad je K_A definována pro reakci formulovanou ve směru deprotonace

$$K_A(\text{HA}) = \frac{a_{\text{H}^+} a_{\text{A}^-}}{a_{\text{HA}}} = (c^{ref})^{-3} \frac{\gamma_{\text{H}^+} \gamma_{\text{A}^-} c_{\text{H}^+} c_{\text{A}^-}}{\gamma_{\text{HA}} c_{\text{HA}}} \quad (2.10)$$

$$K_A(\text{B}) = \frac{a_{\text{H}^+} a_{\text{B}}}{a_{\text{BH}^+}} = (c^{ref})^{-3} \frac{\gamma_{\text{H}^+} \gamma_{\text{B}} c_{\text{H}^+} c_{\text{B}}}{\gamma_{\text{BH}^+} c_{\text{BH}^+}}. \quad (2.11)$$

Vzhledem k rozpětí číselných hodnot, které K_A různých kyselin a zásad nabývá, je také používán její záporný dekadický logaritmus označovaný jako $\text{p}K_A$.

$$\text{p}K_A = -\log_{10}(K_A) \quad (2.12)$$

pH

Rovnovážný stav pro autoprotolýzu vody (reakce 2.3) je popsán iontovým součinem vody

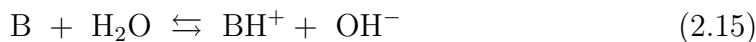
$$K_W = a_{\text{H}^+} a_{\text{OH}^-}, \quad (2.13)$$

který má při 25 °C hodnotu 10^{-14} ($\text{p}K_W = 14$). Míra kyselosti vodného roztoku je popsána veličinou pH definovanou jako

$$\text{pH} = -\log_{10}(a_{\text{H}^+}). \quad (2.14)$$

Konstanta bazicity

Jiný způsob jak charakterizovat zásadu je konstanta bazicity K_B . Je to rovnovážná konstanta pro reakci



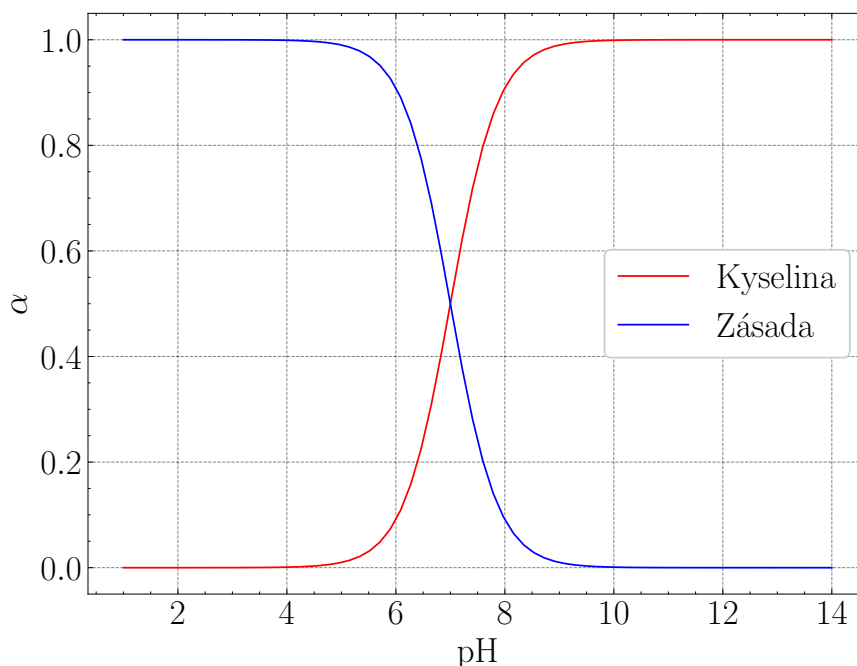
ve tvaru

$$K_B = \frac{a_{\text{BH}^+} a_{\text{OH}^-}}{a_{\text{B}}}. \quad (2.16)$$

Převod mezi K_B a K_A je dán vztahem

$$K_A = \frac{K_W}{K_B}. \quad (2.17)$$

Dále v práci je používána pouze konstanta K_A .



Obrázek 2.1 Ideální titrační křivky (závislosti stupně ionizace α na pH) podle vztahů 2.20 a 2.21 pro kyselinu a zásadu s $pK_A = 7$

2.1.2 Stupeň ionizace kyselin a zásad

Elektrostatické interakce hrají podstatnou roli v chování systémů obsahujících ionty. Proto je vhodné zavést veličinu, která přímo udává zlomek částic nesoucích náboj v systému obsahujícím slabé kyseliny a zásady. Stupeň ionizace je pro kyselinu definován takto:

$$\alpha_A = \frac{c_{A^-}}{c_{A^-} + c_{HA}} \quad (2.18)$$

a pro zásadu takto:

$$\alpha_B = \frac{c_{BH^+}}{c_{BH^+} + c_B}. \quad (2.19)$$

Ideální titrační křivka

Pro systém bez intermolekulárních interakcí, kde aktivitní koeficienty všech složek jsou jednotkové, lze kombinací vztahů pro α , pH a K_A (pK_A) odvodit Hendersonovu-Hasselbalchovu rovnici jako vztah mezi stupněm ionizace slabé kyseliny resp. zásady v závislosti na pH systému.

$$\text{pH} - pK_A = \log_{10} \left(\frac{\alpha_A}{1 - \alpha_A} \right) \quad (2.20)$$

$$\text{pH} - pK_A = \log_{10} \left(\frac{1 - \alpha_B}{\alpha_B} \right). \quad (2.21)$$

Protože kyselina je nabitá v disociovaném stavu, zatímco zásada je nabitá v asociovaném stavu, má Hendersonova-Hasselbalchova rovnice jiný tvar pro slabou kyselinu než pro slabou zásadu. Závislost stupně ionizace na pH danou výše uvedenými vztahy označujeme jako ideální titrační křivku (Obr. 2.1).

Vliv interakcí

Vztah pro ideální titrační křivku je pro interagující systém nutné upravit zahrnutím členu Δ [4]

$$\text{pH} - \text{p}K_A = \log_{10} \left(\frac{\alpha_A}{1 - \alpha_A} \right) + \Delta, \quad (2.22)$$

kde Δ má obecný tvar

$$\Delta = \frac{1}{k_B T \ln(10)} \sum_{i \neq \text{H}^+} \nu_i \mu_i^{ex}, \quad (2.23)$$

kde μ^{ex} pro H^+ iont je již zahrnutý v definici pH (2.14). Odchýlení od ideální titrační křivky pak lze ekvivalentně popsat jako efektivní změnu hodnoty v $\text{p}K_A$ kyseliny nebo pH systému [4]:

$$\text{pH} - \text{p}K_A - \Delta = \text{pH} - \text{p}K_{A,\text{eff}} = \text{pH}_{\text{eff}} - \text{p}K_A = \log_{10} \left(\frac{\alpha_A}{1 - \alpha_A} \right). \quad (2.24)$$

Úprava pro ideální titrační křivku slabé zásady je analogická.

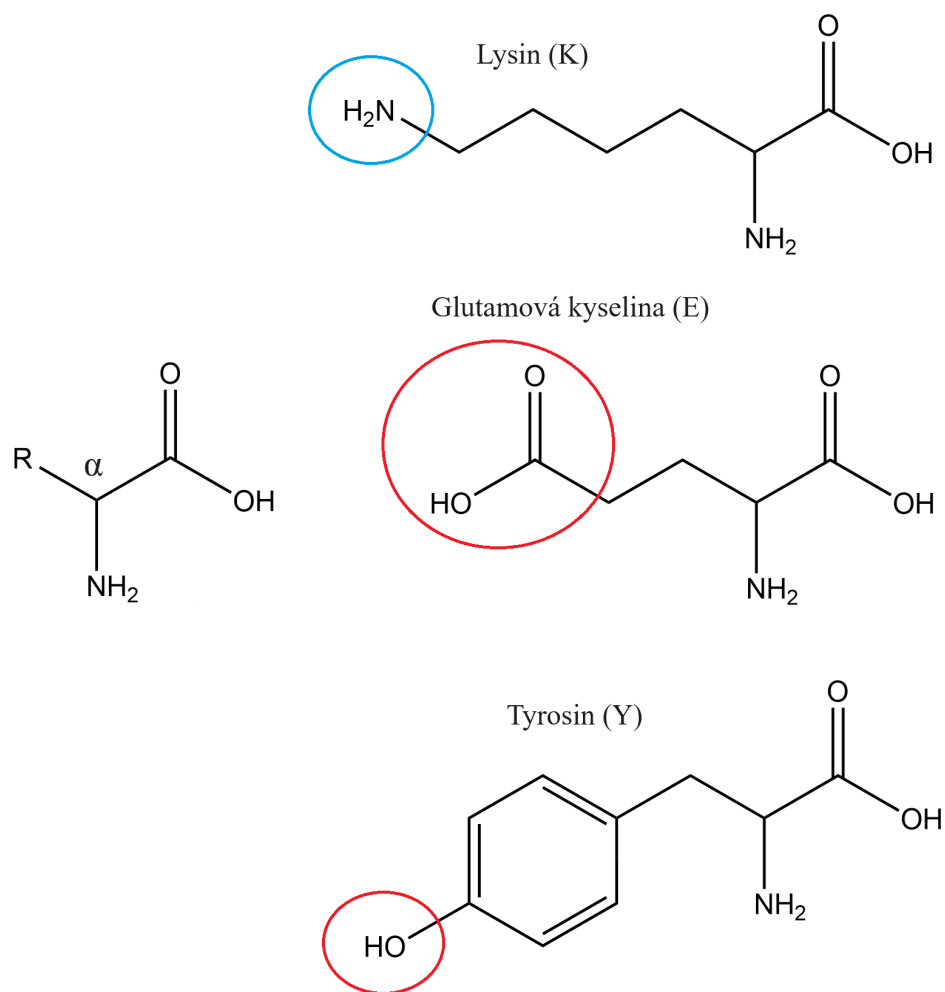
2.2 Polypeptidy

Struktura a vlastnosti polypeptidu jsou dány aminokyselinami, ze kterých je složen, takže je nejprve nutné uvést základní popis aminokyselin jako takových. Chování samotných polypeptidů je pak dále v práci diskutováno v kontextu známých poznatků z polymerních věd a hlavně z oblasti výzkumu polyelektrolytů, které jsou zde stručně nastíněny.

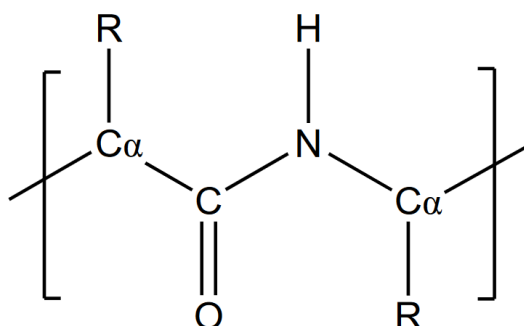
2.2.1 Aminokyseliny a peptidová vazba

Aminokyseliny jsou obecně organické sloučeniny obsahující zároveň aminovou skupinu ($-\text{NH}_2$) a karboxylovou skupinu ($-\text{COOH}$). V přírodě jich existuje mnoho se širokou škálou funkcí. Jako součást proteinů se ale typicky vyskytuje pouze 20 konkrétních aminokyselin, přičemž všechny tyto aminokyseliny jsou α -aminokyseliny, jejich aminová skupina je navázána na α -uhlík, tedy uhlík, na který je zároveň navázána i skupina karboxylová. Kromě glycinu obsahují tyto aminokyseliny na α -uhlíku centrum chiraloty, jehož konfigurace ve Fisherově projekci je pro přírodní aminokyseliny vždy L [7].

Zbytek molekuly vycházející z α -uhlíku je pak typicky označován jako postranní řetězec. Specifické vlastnosti aminokyselin jsou dány jejich postranními řetězci, a proto je běžné dělit aminokyseliny podle struktury a polaroty postranních řetězců. Pro tuto práci je relevantní rozdělení podle $\text{p}K_A$ postranního řetězce na aminokyseliny s neutrálním, kyselým nebo zásaditým postranním řetězcem. Co se týče námi studovaných aminokyselin (znázorněných na obr. 2.2), lysin má zásaditý postranní řetězec, glutamová kyselina kyselý a, ačkoliv je tyrosin typicky řazen mezi neutrální aminokyseliny, jeho hydroxylová skupina může při dostatečně vysokém pH disociovat jako kyselina [8].



Obrázek 2.2 Vlevo: Vzorec obecné α -aminokyseliny s postranním řetězcem označeným R, Vpravo: Aminokyseliny, z nichž jsou složeny studované polypeptidy, s označenými kyselými (červená) nebo zásaditými (modrá) skupinami na postranních řetězcích



Obrázek 2.3 Schéma peptidové vazby uvnitř molekuly peptidu s vyznačenými α -uhlíky (C_α) a postranními řetězci označenými jako R

Peptidová vazba

Aminokyseliny se spolu spojují kondenzační reakcí mezi karboxylovou skupinou na jedné molekule a aminovou skupinou na druhé molekule za vzniku amidu a molekuly vody. Tato amidová vazba se v kontextu aminokyselin označuje jako peptidová. Vzniklý peptid má na jednom konci molekuly volnou aminovou skupinu – N-konec a na druhém volnou karboxylovou skupinu – C-konec. Struktura peptidové vazby je zobrazena na obr. 2.3. Vazba C-N má částečný charakter dvojné vazby kvůli částečnému sdílení elektronů z volného páru na amidovém dusíku s karbonylovým uhlíkem, takže atomy O, C, N a H existují v jedné rovině s minimální rotací kolem této vazby [7]. K rotaci může volně docházet kolem vazeb C_α -C a N- C_α , což dává polypeptidu jeho konformační stupně volnosti. Pro postranní řetězce je pak v molekule typicky preferována trans konformace [2].

Izoelektrický bod a náboj peptidu

Izoelektrický bod (pI) je obecně pH, při kterém je celkový náboj molekuly nulový. Je to relevantní veličina například pro separaci peptidů a proteinů pomocí elektroforézy [8]. Samotné aminokyseliny všechny obsahují alespoň jednu kyselou a jednu zásaditou skupinu v podobě karboxylové a aminové skupiny, jejichž stupeň ionizace závisí na pH podle rovnic 2.20 a 2.21 (v aproximaci bez interakcí). Při $\text{pH} = \text{pI}$ v roztoku obě skupiny existují převážně v nabitě formě, takže se jejich náboje vzájemně kompenzují. Peptidy kromě těchto skupin na C- a N-koncích mohou obecně obsahovat mnoho ionizovatelných skupin z postranních řetězců různých aminokyselin, kde každá z nich podléhá vlastní ionizační rovnováze. Celkový náboj molekuly peptidu lze pak vyjádřit jako funkci pH následovně [9]:

$$z(\text{pH}) = \sum_i z_i n_i \alpha_i(\text{pH}), \quad (2.25)$$

kde z je celkový náboj peptidu, z_i je náboj skupiny typu i (pro postranní řetězce aminokyselin vždy $+1 e$ nebo $-1 e$) a α_i je průměrný stupeň ionizace skupiny typu i a n_i je počet skupin typu i . Pro obecný peptid je $z(\text{pH})$ nerostoucí funkcí, která prochází nulou při $\text{pH} = \text{pI}$.

2.2.2 Gyrační poloměr polypeptidu

Polymery včetně polypeptidů mají v porovnání s malými molekulami mnohem větší množství možných konformačních stavů. To je primárně díky počtu jednoduchých vazeb v makromolekule, kolem kterých mohou její segmenty rotovat. V polymerních vědách existuje více veličin, které lze použít pro popis prostorových konformací makromolekul. Nejjednodušší veličinou je vzdálenost konců (\vec{R}_{ee}) definována jako

$$\vec{R}_{ee} = \vec{r}_N - \vec{r}_0 \quad (2.26)$$

kde \vec{r}_N je polohový vektor poslední monomerní jednotky (pro řetězec délky N) a \vec{r}_0 je polohový vektor první monomerní jednotky. Protože molekula neustále přechází mezi možnými konformacemi, jsou typicky používány střední hodnoty veličin (označené pomocí $\langle \rangle$). Specificky se u vzdálenosti konců používá střední kvadratická vzdálenost konců $\langle R_{ee}^2 \rangle = \langle \vec{R}_{ee} \cdot \vec{R}_{ee} \rangle$ resp. odmocnina z ní $\langle R_{ee}^2 \rangle^{1/2}$ [2]. Protože je vzdálenost konců dána polohami pouze dvou monomerních jednotek, má relativně velký rozptyl, a zároveň je problém ji definovat pro rozvětvené polymery.

Pro charakterizaci konformace polypeptidu je v práci použita veličina gyrační poloměr (R_g), který obecně udává rozložení hmoty tělesa kolem jeho těžiště a lze ho interpretovat jako poloměr duté koule se stejným momentem setrvačnosti, jako má molekula. Pro homopolymer o délce N ho lze spočítat následovně [10]:

$$\langle R_g^2 \rangle^{1/2} = \sqrt{\frac{1}{N} \sum_i^N \langle (\vec{r}_i - \vec{r}_{com})^2 \rangle}, \quad (2.27)$$

kde \vec{r}_i je polohový vektor monomerní jednotky i a \vec{r}_{com} je polohový vektor těžiště.

2.2.3 Polypeptid jako polyelektrolyt

Polyelektrolyty jsou polymery, ve kterých velká část monomerních jednotek obsahuje skupiny, které mohou nést náboj v důsledku disociačních reakcí – jsou ionizovatelné. Polyelektrolyty obsahující pouze záporně nabitě skupiny se označují jako polyanionty, zatímco ty obsahující pouze kladně nabitě skupiny jako polykationty. Pokud obsahují oba typy skupin, jsou to polyamfolyty. Analogicky s nízkomolekulárními elektrolyty (viz. 2.1.1) lze polyelektrolyty rozdělit na silné a slabé s tím, že náboj slabého polyelektrolytu není konstantní, ale závisí na pH. Volné ionty vzniklé disociačními reakcemi jsou označovány jako proti-ionty, přičemž roztok zároveň může obsahovat kromě molekul polyelektrolytu a proti-iontů i další volné ionty vzniklé disociací přidané nízkomolekulární soli. Polyelektrolyty mohou být umělé jako polyakrylová kyselina nebo přírodní jako DNA. Polypeptidy, které obsahují aminokyseliny s kyselými nebo zásaditými postranními řetězci, lze považovat za slabé polyelektrolyty.

Vliv elektrostatických interakcí na chování polyelektrolytu

Specifické chování polyelektrolytů se v mnoha ohledech liší jak od nenabitých polymerů, tak od nízkomolekulárních elektrolytů [11]. Nabitě skupiny na polyelektrolytu interagují mezi sebou a s okolními ionty skrz coulombický potenciál a tyto elektrostatické interakce ovlivňují konformaci molekuly, kde polyelektrolyty obecně zaujímají v prostoru více nataženou konformaci než nenabitě polymery

se vzdáleností konců řetězce rostoucí s jejich stupněm ionizace [12]. Zároveň tyto interakce mění disociační rovnováhu ionizovatelných skupin na slabém polyelektrolytu. Relevantní veličinou pro roztoky polyelektrolytů je Bjerrumova délka l_B

$$l_B = \frac{e^2}{4\pi\epsilon k_B T}, \quad (2.28)$$

kde e je elementární náboj a ϵ je permitivita prostředí. Pokud se dva elementární náboje nacházejí ve vzdálenosti l_B jejich interakční energie odpovídá termální energii $k_B T$. Kvalitativně lze říct, že pro vzdálenosti menší než l_B dominují elektrostatické interakce, zatímco pro vzdálenosti větší než l_B převládají termální fluktuace. Ve vodě o teplotě 298 K má l_B hodnotu 0,71 nm [2].

Proti-ionty v roztoku jsou ovlivněny dvěma protichůdnými působeními: entropické termální fluktuace a přitažlivé elektrostatické interakce s řetězcem polyelektrolytu. Jednoduchá teorie popisující jejich chování udává, že přitažlivé interakce převládají, když efektivní vzdálenost nábojů na řetězci je menší než l_B [2]. Proti-ionty pak kondenzují na řetězci a dochází k částečnému potlačení dalšího vlivu elektrostatických interakcí (například na konformaci řetězce) v důsledku jejich odstínění.

Člen Δ v rovnici 2.22, který obecně popisuje odchylky od ideální disociace kyselin a zásad, lze pro neideální chování způsobené elektrostatickými interakcemi v polyelektrolytu vyjádřit jako [4]

$$\Delta = \frac{\langle e\psi(\vec{r}) \rangle}{k_B T \ln(10)}, \quad (2.29)$$

kde $\langle \psi(\vec{r}) \rangle$ je střední elektrostatický potenciál v bodě daném polohovým vektorem \vec{r} . Neideální disociaci kyselých a zásaditých skupin v polyelektrolytu lze tedy v souladu s rovnicí 2.24 brát jako danou lokálními hodnotami jejich $pK_{A,\text{eff}}$ na a v okolí polyelektrolytu, popř. „lokálním pH“ (správněji lokální koncentrací H^+ iontů). Při titraci slabého polyelektrolytu je velikost elektrostatického potenciálu (a tedy i hodnota $pK_{A,\text{eff}}$) závislá na stupni ionizace skupin v polyelektrolytu, který se v průběhu titrace mění v závislosti na pH. Zároveň, pokud není valná většina skupin na řetězci ionizovaná, elektrostatický potenciál nemusí být obecně napříč celým řetězcem stejný. Titrační křivka pak není posunuta o konstantní hodnotu, ale celkově deformována [4].

Tento popis je založen na tzv. mean-field aproximaci, uvažující střední hodnotu potenciálu bez jeho fluktuací. Teoretické modely založené na mean-field aproximaci (využívající Poissonovu-Boltzmannovu rovnici nebo sofistikovanější přístupy) dokáží popsat rozložení iontů kolem polyelektrolytu i jeho konformaci, ale stále mají nedostatky jako např. omezení primárně na systémy s výhodnou symetrií, což dává prostor molekulovým simulacím [4].

3 Metodická část

Reálný systém, který má být reprezentován a studován v simulaci, se skládá z polypeptidu o dané aminokyselinové sekvenci v roztoku pufru o konstantním pH a iontové síle. Počítačový model proto musí obsahovat reprezentaci aminokyselin spojených chemickou vazbou a nabitých částic z pufru a přidané soli. Zároveň musí správně vystihovat acidobazické chování polypeptidu s disociačními reakcemi kyselých a zásaditých postranních řetězců aminokyselin, při kterých dochází ke změně jejich náboje a které jsou závislé na okolním pH. Počítačový model musí být zvolen s tak detailním rozlišením, aby všechny výše zmíněné požadavky byly splněny a zároveň tak, aby výpočetní náročnost simulace byla dostatečně nízká pro získání požadovaného množství nekorelovaných dat ve vyhrazeném čase s dostupným hardwarem.

3.1 Model polypeptidu

Reálný chemický systém sestávající z elementárních částic je pro účely simulace potřeba více či méně zjednodušit na vhodný model. Zvolená míra zjednodušení se odvíjí od dostupných výpočetních zdrojů, konkrétní simulační metody a charakteru zkoumaných veličin. Pro simulační metody založené na klasické mechanice používané v této práci jsou typické modely atomistické nebo zhrubené (coarse-grained).

V atomistických simulacích jsou základním strukturním prvkem atomy. Interakční potenciály mezi nimi jsou pak založeny na hodnotách získaných z kvantově-chemických výpočtů nebo na empirických parametrech. Takto detailní model poskytuje přesné výsledky, ale výpočetní náročnost spojená s množstvím interagujících atomů vytváří omezení na velikost systému.

Pro snížení výpočetních nároků na úkor přesnosti lze použít zhrubený model. Ve zhrubeném modelu jsou některé skupiny atomů nebo celé molekuly aproximovány jako jedna částice. Částice ve zhrubeném modelu jsou typicky měkké koule, jejichž efektivní velikost je dána odpudivými ne vazebnými interakcemi.

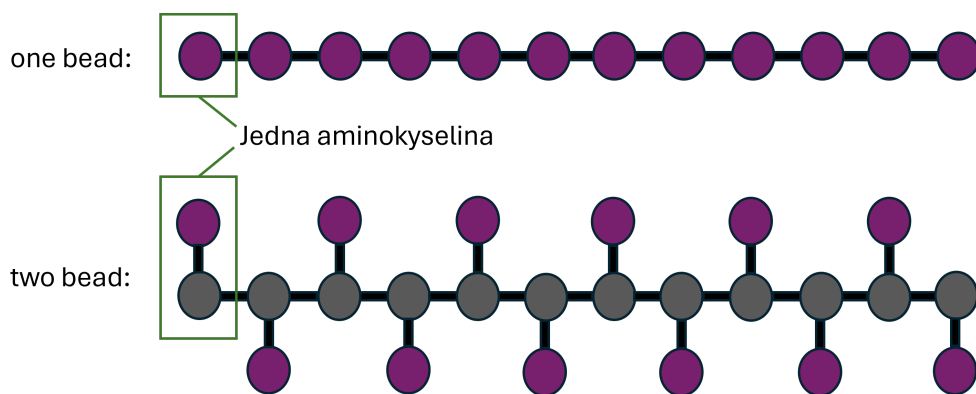
Vzhledem k výpočetní náročnosti simulování všech atomů rozpouštědla a polypeptidového řetězce byl pro reprezentaci systému v simulaci zvolen zhrubený model s implicitním rozpouštědlem.

3.1.1 Reprezentace aminokyseliny v použitém modelu

Při použití zhrubeného modelu pro reprezentaci aminokyselin v polypeptidovém řetězci je nutné zvolit, kolik částic připadne na jednu aminokyselinu. Při modelování studovaného systému byly zvoleny dva přístupy znázorněné na obr. 3.1, které jsou v této práci označovány jako one bead model a two bead model.

Nejjednodušší možností je použít pro celou aminokyselinu pouze jednu částici – one bead model. Polypeptid tvořený N aminokyselinami se pak skládá z N částic spojených chemickou vazbou. Případný náboj který aminokyselina může nést v důsledku disociačních reakcí je umístěn do středu částice, která ji reprezentuje.

Konformační vlastnosti polypeptidu (viz. 2.2.1) lépe vystihuje two bead model [13], kde je každá aminokyselina tvořena dvěma chemickou vazbou spojenými částicemi:



Obrázek 3.1 Ilustrace polypeptidového řetězce modelovaného pomocí one bead nebo two bead modelu

1. neutrální centrální částicí, která je pro všechny aminokyseliny stejná a reprezentuje tu část molekuly, která je spojená peptidovou vazbu s ostatními aminokyselinami
2. částicí reprezentující postranní řetězec aminokyseliny, která nese informaci o acidobazickém chování dané aminokyseliny a její případný náboj.

Hodnoty pK_A postranních řetězců aminokyselin byly převzaty z práce Hasse a Muldera [14]. Karboxylová skupina a aminoskupina aminokyselin na C-konci a N-konci polypeptidu nejsou vzhledem k délce studovaných polypeptidů v modelech nijak reprezentovány.

3.1.2 Reprezentace ostatních složek systému v použitém modelu

Kromě částic reprezentujících aminokyseliny v polypeptidovém řetězci jsou do simulace zahrnuty další dva typy částic – obecný monovalentní kation a anion, které reprezentují všechny ionty z přidané soli i pufru.

Molekuly rozpouštědla nejsou v simulaci reprezentovány pomocí částic, ale pouze implicitně skrz relativní permitivitu prostředí.

3.1.3 Interakce v použitém modelu

Nevazebné interakce

Všechny částice spolu interagují skrz Weeksův–Chandlerův–Andersenův (WCA) potenciál [15], což je varianta Lennard-Jonesova (LJ) potenciálu obecně definovaného jako

$$U_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (3.1)$$

kde r je vzdálenost částic, ϵ je interakční energie minima a σ je vzdálenost částic s nulovou interakční energií. V simulaci je nutné působení Lennard-Jonesova potenciálu omezit tak, že za jistou konečnou vzdáleností r_{cut} je potenciál nulový. Aby potenciál zůstal spojitý je vzniklý skok kompenzován posunutím potenciálu o konstantu c . Pro získání WCA potenciálu je hodnota r_{cut} zvolena jako $2^{1/6}\sigma$, což

odpovídá vzdálenosti v minimu LJ potenciálu, a posun c jako ϵ . Výsledný WCA potenciál má pak čistě odpudivý charakter a je definován takto [16]:

$$U_{\text{WCA}} = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] + \epsilon & \text{pro } 0 < r < 2^{1/6}\sigma \\ 0 & \text{pro } r > 2^{1/6}\sigma \end{cases} \quad (3.2)$$

Parametry WCA potenciálu byly zvoleny pro všechny částice v simulaci podle práce Lunkada et al. [9] jako $\epsilon = k_B T$, kde k_B je Boltzmanova konstanta, T je teplota simulovaného systému (25 °C) a $\sigma = 0.35$ nm

Vazebné interakce

Chemické vazby mezi částicemi jsou v simulaci reprezentovány harmonickým potenciálem

$$U_b = \frac{1}{2}k(r - r_0)^2 \quad (3.3)$$

kde r_0 je rovnovážná délka vazby a k je tuhost vazby.

V one bead modelu byly parametry všech vazeb mezi částicemi reprezentujícími aminokyseliny zvoleny jako $k = 400k_B T \text{ nm}^{-2}$ a $r = 0,4$ nm. V two bead modelu existuje více typů vazeb s různou délkou (centrální částice - postranní řetězec (rozdílné délky pro každou aminokyselinu), centrální částice - centrální částice), jejichž hodnoty parametru r byly opět převzaty z práce Lunkada et al. [9], ve které byly získány s pomocí atomistických simulací.

Elektrostatické interakce

Nabitě částice spolu navíc interagují skrz elektrostatický potenciál daný Coulombovým zákonem

$$U_e = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_1 q_2}{r}, \quad (3.4)$$

kde r je vzdálenost mezi nabitými částicemi, q_1 a q_2 jsou náboje částic, ϵ_0 je permitivita vakua a ϵ_r je relativní permitivita. Za ϵ_r je dosazena relativní permitivita vody při 25 °C ($\epsilon_r = 78,5$ [17]) pro simulování vlivu rozpouštědla.

Vzhledem k tomu, že coulombické interakce mají nekonečný dosah a jejich velikost klesá pouze s r^{-1} , by použití obdobného přístupu jako u ne vazebných interakcí s useknutím potenciálu pro $r > r_{\text{cut}}$ vedlo k velké systematické chybě v hodnotě potenciálu. V simulaci je pro jeho vyčíslení používán P3M (Particle-Particle-Particle-Mesh) algoritmus [16], který při výpočtu používá Ewaldovu sumaci [18].

3.2 Molekulové simulace

Počítačové simulace propojují experimentální a teoretické přístupy, protože umožňují provedení experimentu se systémem, jehož chování je zcela určeno zvoleným teoretickým modelem. Simulace tak může sloužit k prozkoumání vlastností daného modelu a, pokud je model dostatečně přesný, i k rozšíření poznání o reálném světě. Výše popsany model námi studovaného systému, použitý v simulacích,

je založen na klasické mechanice a teoretickým základem pro takové simulace je proto klasická statistická mechanika.

Stav systému N částic, který se chová podle zákonů klasické mechaniky, lze úplně popsat N trojrozměrnými polohovými vektory \vec{r} a N trojrozměrnými vektory hybnosti \vec{p} . Tato sada vektorů se označuje jako mikrostav $\psi = (\vec{r}_1, \dots, \vec{r}_N, \vec{p}_1, \dots, \vec{p}_N)$. Všechny takto definované mikrostavy, ve kterých se systém může nacházet, pak tvoří $6N$ -rozměrný fázový prostor. Konfigurační prostor je $3N$ -rozměrný prostor tvořený pouze možnými polohovými vektory [19].

Konkrétní makroskopický stav (makrostav) systému popsany termodynamickými veličinami jako teplota (T), objem (V) a počet částic (N) lze realizovat obecně velkým množstvím mikrostavů vyskytujících se s různou pravděpodobností. Termodynamický soubor je podprostorem fázového prostoru a obsahuje v sobě všechny mikrostavy, zastoupené s frekvencí podle pravděpodobnosti jejich výskytu, které odpovídají jednomu zvolenému makrostavu. Označení a vlastnosti souboru se odvíjí od veličin, které ho definují. Například mikrokanonický soubor je definován konstantním počtem částic, objemem a energií $[N, V, E]$, makrokanonický soubor je definován konstantním počtem částic, objemem a teplotou $[N, V, T]$.

Vzhledem k množství dostupných mikrostavů, mezi kterými systém neustále přechází, můžeme při měření veličiny $X(\psi)$, jejíž hodnota závisí na konkrétním mikrostavu, získat pouze průměr z jejích okamžitých hodnot. Pro systém v rovnováze fluktuují okamžité hodnoty $X(\psi)$ kolem její souborové střední hodnoty

$$\langle X \rangle = \int_{\Omega} X(\psi) \pi(\psi) d\psi \quad (3.5)$$

kde $\pi(\psi)$ je hustota pravděpodobnosti nalezení systému v mikrostavu ψ , jejíž hodnota závisí na tom, jaký termodynamický soubor odpovídá podmínkám měření a Ω je celá podoblast fázového prostoru příslušící tomuto souboru.

V molekulových simulacích vzorkujeme fázový nebo konfigurační prostor pro daný systém, omezený na vybraný soubor. Vzorkování znamená generování odpovídajících mikrostavů způsobem daným konkrétní simulační metodou. Pro každý mikrostav lze spočítat hodnotu sledované veličiny $X(\psi)$ a její střední hodnotu (3.5) pak můžeme odhadnout pomocí průměru

$$\langle X \rangle \approx \bar{X} = \frac{\sum_{i=1}^n X(\psi_i)}{n} \quad (3.6)$$

kde n je počet vygenerovaných vzorků.

3.2.1 Molekulová dynamika

Při simulaci založené na molekulové dynamice je vzorkována jedna trajektorie (posloupnost bodů ve fázovém prostoru), kterou systém projde za určitý čas. Časový vývoj systému z jednoho bodu trajektorie do následujícího je založen na řešení Newtonových pohybových rovnic

$$\frac{d\vec{r}(t)}{dt} = \vec{v}(t) \quad (3.7)$$

$$\frac{d\vec{v}(t)}{dt} = \frac{\vec{F}(\vec{r}(t))}{m}, \quad (3.8)$$

kde \vec{r} je poloha částice, t je čas, \vec{v} je rychlost částice, \vec{F} je celková síla působící na částici a m je hmotnost částice. Obecný postup při realizaci MD simulace je

1. Zvolení počátečních pozic a rychlostí, obecně náhodně.
2. Výpočet sil působících na částice jako

$$\vec{F}(\vec{r}(t)) = -\frac{\partial U(\vec{r}(t))}{\partial \vec{r}(t)}, \quad (3.9)$$

kde $U(\vec{r})$ je celkový potenciál, ve kterém se částice nachází, daný interakčními potenciály použitými v modelu

3. Určení nových pozic a rychlostí částic v čase $t + dt$ integrací rovnic 3.7 a 3.8.
4. Kroky 2 a 3 jsou opakovány po dobu trvání simulace.

Integrace pohybových rovnic

Newtonovy pohybové rovnice je v praxi nutné integrovat numericky metodou konečných diferencí. Populární metodou, která je zároveň časově reverzibilní, a tudíž zachovává celkovou energii, je Verletova metoda [19].

Konkrétní algoritmus použitý v našich simulacích je Velocity Verlet [16] sestávající se z následujících kroků

1. výpočet rychlosti v polovině integračního kroku

$$\vec{v}(t + dt/2) = \vec{v}(t) + \frac{\vec{F}(\vec{r}(t))}{m} dt/2 \quad (3.10)$$

2. určení nové pozice

$$\vec{r}(t + dt) = \vec{r}(t) + \vec{v}(t + dt/2) dt \quad (3.11)$$

3. výpočet sil působících na částici nacházející se na nové pozici podle rovnice 3.9
4. výpočet nové rychlosti

$$\vec{v}(t + dt) = \vec{v}(t + dt/2) + \frac{\vec{F}(\vec{r}(t + dt))}{m} dt/2. \quad (3.12)$$

Termostat

Výše uvedený postup umožňuje provádět simulace systému s konstantní energií. Pro simulování systému za konstantní teploty je nutné modifikovat pohybové rovnice přidáním termostatu. V našich simulacích byl použit termostat založený na Langevinově rovnici [20]. Síly spočítané podle rovnice 3.9 jsou upraveny následovně [16]

$$\vec{F}(\vec{r}, t) = \vec{F}(\vec{r}(t)) - \gamma \vec{v}(t) + \sqrt{2\gamma k_B T} \eta(t), \quad (3.13)$$

kde $\vec{F}(\vec{r}, t)$ je výsledná síla použitá při integraci pohybových rovnic, γ je koeficient reprezentující tření v kapalině - její viskozitu, \vec{v} je rychlost částice, k_B je Boltzmannova konstanta, T je zvolená konstantní teplota a η je náhodná síla reprezentující srážky částice s molekulami rozpouštědla o teplotě T .

3.2.2 Monte Carlo

V Monte Carlo simulacích je vzorkován konfigurační prostor pomocí generátoru náhodných čísel. V tom nejjednodušším provedení jsou všechny konfigurace generovány se stejnou pravděpodobností, bez ohledu na celkovou potenciální energii systému. Při simulaci N částic v krychli o objemu V by každá konfigurace byla vygenerována jako $3N$ náhodných souřadnic z rovnoměrného rozdělení na intervalu $(0, \sqrt[3]{V})$. Pro každou konfiguraci by pak bylo možné spočítat hodnotu sledované veličiny, která je funkcí souřadnic částic.

Typicky je ale cílem simulace odhadnout střední hodnotu veličiny v konkrétním souboru (3.5), kde některé konfigurace mohou být pravděpodobnější než jiné. Například v kanonickém $[N, V, T]$ souboru je pravděpodobnost výskytu konfigurace i s energií E_i úměrná Boltzmannovu faktoru

$$P(i) \approx \exp\left(-\frac{E_i}{k_B T}\right). \quad (3.14)$$

Správný odhad střední hodnoty v kanonickém souboru by bylo možné získat z váženého průměru, kde je hodnota veličiny v každé konfiguraci vážena příslušným Boltzmannovým faktorem. Pokud jsou však konfigurace stále generovány rovnoměrně, je velká část výpočetního času využita na konfigurace, které k souborové střední hodnotě přispívají jen minimálně (hodnota jejich Boltzmannova faktoru je zanedbatelná), a ke správnému výsledku by simulace konvergovala velmi pomalu.

Mnohem efektivnějším přístupem je preferenční vzorkování (importance sampling), kde pravděpodobnost vygenerování konfigurace přímo odpovídá jejímu Boltzmannovu faktoru. Základním algoritmem pro preferenční vzorkování v kanonickém souboru je Metropolisova metoda [21].

Metropolisova metoda

Při použití této metody je nutné začít se systémem v nějaké počáteční konfiguraci. Každá další konfigurace je pak odvozena od té předchozí. Postup při generování nové konfigurace je následující

1. z předchozí konfigurace s potenciální energií U_p je vygenerována zkušební konfigurace náhodným pohybem částice (nebo částic)
2. je určena potenciální energie zkušební konfigurace U_{zk}
3. zkušební konfigurace je přijata jako nová počáteční konfigurace s pravděpodobností

$$P = \min\left[1, \exp\left(-\frac{\Delta U}{k_B T}\right)\right], \quad (3.15)$$

kde $\Delta U = U_{zk} - U_p$.

Pravděpodobnost přijetí je tedy 1 pokud je $U_{zk} \leq U_p$ nebo $\exp(-(U_{zk} - U_p)/k_B T)$ pokud je $U_{zk} > U_p$. Při nepřijetí je jako nová konfigurace použita ta předchozí

Opakováním těchto kroků lze získat posloupnost konfigurací odpovídajících kanonickému souboru.

Chemické reakce pomocí MC

Pro simulaci systému s chemickými reakcemi nelze použít kanonický soubor, protože počet částic není v důsledku reakcí konstantní. Existuje několik souborů, které reprezentují různé varianty chemických systémů s reakcemi. Vzhledem k tomu, že je cílem reprezentovat disociační reakce postranních řetězců aminokyselin v roztoku pufru o konstantním pH, byl pro naše simulace zvolen Constant-pH soubor [22], který je založený na modifikaci Metropolisovy Monte Carlo metody.

Při realizaci reakčního kroku v Constant-pH souboru je nejprve zvolen směr reakce. Všechny reakce jsou definovány ve směru deprotonace a pravděpodobnost zvolení dopředné resp. zpětné reakce odpovídá celkovému počtu protonovaných resp. deprotonovaných skupin.

$$P_{\text{vpřed}} = \frac{N_{\text{prot}}}{N_{\text{prot}} + N_{\text{deprot}}} \quad (3.16)$$

$$P_{\text{zpět}} = \frac{N_{\text{deprot}}}{N_{\text{prot}} + N_{\text{deprot}}} = 1 - P_{\text{vpřed}}, \quad (3.17)$$

kde $P_{\text{vpřed}}$ je pravděpodobnost zvolení dopředné reakce, $P_{\text{zpět}}$ je pravděpodobnost zvolení zpětné reakce, N_{prot} je počet protonovaných skupin a N_{deprot} je počet deprotonovaných skupin.

Při zvolení dopředné reakce je náhodně vybraná částice reprezentující protonovanou aminokyselinu (nebo její postranní řetězec v two bead modelu) nahrazena korespondující deprotonovanou verzí a na náhodné místo je vložen obecný kation pro zachování elektroneutality. U zpětné reakce je naopak deprotonovaná verze nahrazena tou protonovanou a náhodný kation je odstraněn. Takto vygenerovaná zkušební konfigurace je pak přijata s pravděpodobností

$$P = \min \left[1, \exp \left(-\frac{\Delta U}{k_B T} + \xi(\text{pH} - \text{p}K_A)\ln(10) \right) \right], \quad (3.18)$$

kde ξ je rozsah reakce ($\xi = +1$ pro dopřednou reakci a $\xi = -1$ pro zpětnou reakci).

V Constant-pH souboru vstupuje pH systému do simulace pouze implicitně jako parametr v rovnici 3.18. Ionty H^+ a OH^- a k nim korespondující proti-ionty, které by byly přítomny v nezanedbatelném množství při nízkých (< 3) a vysokých (> 11) hodnotách pH nejsou v simulaci nijak reprezentovány. To má za následek, že při těchto hodnotách pH se iontová síla v simulaci může výrazně lišit od iontové síly v reálném roztoku [23]. Tento nedostatek je kompenzován zavedením dostatečně vysoké konstantní koncentrace přidané soli [4].

3.2.3 Simulační box

I přes všechna zjednodušení použitá při modelování systému je počet částic, které je možné simulovat, o mnoho řádů nižší než počet částic v reálném roztoku. Zlomek částic, které se nachází na povrchu systému je tak mnohem větší v simulaci, než jaký by byl v experimentu [21]. Efektivně by tak byla simulována pouze malá kapka, jejíž vlastnosti se mohou výrazně lišit od makroskopického roztoku. Aby bylo možné aproximovat reálný systém bez velkého navýšení výpočetních nároků, jsou zavedeny periodické okrajové podmínky.

Simulace jsou prováděny v kubickém simulačním boxu, který je obklopený nekonečným počtem svých kopií. Počítačem jsou sledovány pouze částice z primárního boxu a jejich periodické kopie vystupují jen v interakcích s částicemi v primárním boxu. Důsledkem periodických okrajových podmínek je, že pokud částice opustí box skrz nějakou stěnu, objeví se na jeho opačné straně.

3.2.4 Jednotky v simulaci

Simulace využívá interní systém redukováných jednotek. Jako redukováná délka byl zvolen parametr WCA potenciálu $\sigma = 0.35$ nm (efektivní velikost coarse-grained částice). Jako jednotka energie byla použita termální energie $k_B T$ při teplotě 25 °C, takže 1 redukováná energie = $4,116 \cdot 10^{-21}$ J. Jako jednotka náboje byla použita velikost elementárního náboje $1,602 \cdot 10^{-19}$ C. Všechny částice mají jednotkovou hmotnost.

3.2.5 Sběr a zpracování dat

Výstupem simulace je časová řada hodnot sledovaných veličin (vzorků) zapisovaných v jejím průběhu a popřípadě i přímo zaznamenané konfigurace systému. Tato data je nutné dále statisticky zpracovat.

Ekvilibrace

Počáteční stav systému může být zvolen v principu libovolně, takže není zaručené, že bude ležet blízko rovnováhy. Ekvilibrace je pak ta část simulace, kde se systém vyvíjí směrem k rovnováze a pozorované veličiny mají viditelný trend místo toho, aby fluktovaly kolem souborové střední hodnoty. Protože cílem simulace je získat informace o rovnovážném stavu systému, je nutné ekvilibraci před dalším zpracováním dat odstranit.

Statistická analýza korelovaných dat

Odhadem souborové střední hodnoty pro veličinu X z nasbíraných dat je podle rovnice 3.6 aritmetický průměr \bar{X} . Pro charakterizaci přesnosti výsledku je použita standardní chyba průměru definována jako

$$\sigma_{\bar{X}}^2 = \langle (\bar{X} - \langle X \rangle)^2 \rangle. \quad (3.19)$$

Standardní chybu průměru lze pro nekorelovaná data odhadnout pomocí vzorce

$$\sigma_{\bar{X}} = \frac{\sigma_{X_i}}{\sqrt{N}}, \quad (3.20)$$

kde N je počet vzorků a σ_{X_i} je výběrová směrodatná odchylka

$$\sigma_{X_i} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2}. \quad (3.21)$$

Pro odhad standardní chyby průměru ze simulace ovšem nelze přímo použít rovnici 3.20, protože data získaná ze simulace jsou korelovaná. Obě použité simulační metody (molekulová dynamika i Monte Carlo podle Metropolisovy

metody) generují následující konfiguraci systému vždy na základě té předchozí, takže nasbíraná data nejsou nezávislá.

Jednou z možností jak získat standardní chybu průměru pro korelovaná data je použít blokovou analýzu [24]. Data jsou rozdělena do b bloků s k vzorky ($N = kb$) a pro každý blok je spočítán blokový průměr \bar{X}_b . S Blokovými průměry je pak nakládáno jako s nezávislými vzorky a jejich výběrovou směrodatnou odchylku $\sigma_{\bar{X}_b}$ lze spočítat podle rovnice 3.21 s dosazením b za N a $\bar{X}_{b,i}$ za X_i . Standardní chyba průměru je pak

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma_{\bar{X}_b}^2}{b}} = \sqrt{\frac{\sigma_{\bar{X}_i}^2}{N_{\text{eff}}}}, \quad (3.22)$$

kde $N_{\text{eff}} = b(\sigma_{\bar{X}_i}^2/\sigma_{\bar{X}_b}^2)$ je efektivní počet nekorelovaných vzorků. Zároveň lze jako míru korelací odhadnout autokorelační čas, což je doba, po které vymizí vliv předchozí konfigurace na tu aktuální [19].

$$\tau = \Delta t \frac{N}{2N_{\text{eff}}}, \quad (3.23)$$

kde Δt je čas, který v simulaci uplyne mezi zapsáním dvou hodnot.

Celý tento postup je založený na předpokladu, že blokové průměry nejsou korelované. Analýzu považujeme za platnou, pokud na jeden blok vychází alespoň 5 efektivních vzorků. Pro analýzu našich simulací byl použit počet bloků $b = 16$, takže výše uvedenou podmínku lze vyjádřit jako

$$N_{\text{eff}} > 80. \quad (3.24)$$

3.3 Simulované systémy

V našich simulacích byly studovány systémy obsahující vždy jeden ze dvou konkrétních typů polypeptidových řetězců:

1. polypeptid označovaný EEEY, který je tvořen opakujícími se podjednotkami složenými ze 3 glutamátových reziduí (E) a jednoho tyrosinového rezidua (Y). Glutamátová rezidua mají kyselý charakter s $pK_A = 4,4$ pro postranní řetězec aminokyseliny. Tyrosinová rezidua mají také kyselý charakter s $pK_A = 9,6$ pro postranní řetězec aminokyseliny.
2. polypeptid označovaný KKKY, který je tvořen opakujícími se podjednotkami složenými ze 3 lysinových reziduí (K) a jednoho tyrosinového rezidua (Y). Na rozdíl od předchozí sekvence mají lysinová rezidua zásaditý charakter s postranním řetězcem s $pK_A = 10,4$, takže tento typ polypeptidu je polyamfolyt s kyselými i zásaditými skupinami.

Oba tyto typy polypeptidů byly modelovány s použitím one bead i two bead modelu pro aminokyseliny (viz 3.1.1). Zároveň byly simulovány dvě různé délky řetězce: 24 aminokyselin ($N = 24$) a 48 aminokyselin ($N = 48$), tedy polypeptidy složené ze 6 nebo 12 opakujících se podjednotek: $(\text{EEEY})_6$, $(\text{EEEY})_{12}$, $(\text{KKKY})_6$ a $(\text{KKKY})_{12}$.

Byly prováděny simulace izolovaného řetězce, takže do simulačního boxu byl vždy umístěn pouze jeden polypeptidový řetězec, a ve všech simulacích byl objem simulačního boxu nastaven tak, aby koncentrace polypeptidu byla 1 g/l.

Většina simulací byla provedena s koncentrací přidané soli $C_s = 32$ mM (odpovídající pětinásobku koncentrace ionizovatelných skupin v polypeptidu) zajištěnou přidáním odpovídajícího množství částic reprezentujících monovalentní kationty a anionty (viz 3.1.2). Tato koncentrace soli slouží jako reprezentace iontů nacházejících se v reálném roztoku a zároveň zajišťuje dostatečně vysokou iontovou sílu pro použití Constant-pH souboru na námi studovaném rozsahu hodnot pH, který je 2–13. Simulace polypeptidů s délkou $N = 24$ byly také provedeny s fyziologickou koncentrací soli ($C_s = 154$ mM).

Základními parametry všech simulací je tedy použitý model (one bead nebo two bead), délka řetězce ($N = 24$ nebo $N = 48$) a koncentrace přidané soli ($C_s = 32$ mM nebo $C_s = 154$ mM).

3.4 Simulační protokol

Simulace výše popsaných systémů je inicializována umístěním polypeptidového řetězce v natažené konformaci do středu simulačního boxu a náhodným rozmístěním iontů soli okolo. Pro odstranění potenciálních překryvů mezi částicemi jsou pozice částic upraveny pomocí 100 iterací steepest descent algoritmu [16] (minimalizujícího energii systému) ve tvaru

$$\vec{r}_{i+1} = \vec{r}_i + \min(\gamma \vec{F}_i, \vec{r}_{\max}), \quad (3.25)$$

kde \vec{r}_i je pozice částice v kroku i , \vec{r}_{i+1} je pozice částice v kroku $i+1$, γ je konstanta s hodnotou 0,1, \vec{F}_i je síla působící na částici daná rovnicí 3.9 (bez elektrostatických interakcí) a \vec{r}_{\max} je maximální posunutí povolené v jednom kroku nastavené jako 0,1 v redukované délce.

Hlavní simulační smyčka pak sestává ze dvou střídajících se fází:

1. simulace za konstantního složení pomocí molekulové dynamiky (3.2.1)
2. série reakčních kroků v Constant-pH souboru (3.2.2) při konstantní prostorové konformaci

. Po každé iteraci simulační smyčky jsou zaznamenány stupně ionizace přítomných aminokyselin, gyrační poloměr polypeptidu a celkový náboj polypeptidu. Celkový počet iterací (a zároveň nasbíraných vzorků sledovaných veličin, zpracovaných viz. 3.2.5) ve všech simulacích je 12 500.

Integrační krok dt v molekulové dynamice má hodnotu 0,01 v simulačních časových jednotkách. Délka jedné simulace je 50 000 časových jednotek pro polypeptidy s $N = 24$ a 500 000 časových jednotek pro polypeptidy s $N = 48$. Při výše zmíněném počtu iterací pak připadá na jednu iteraci simulační smyčky 400 integračních kroků molekulové dynamiky pro $N = 24$ a 4 000 pro $N = 48$. Počet reakčních kroků v jedné iteraci je vždy dvojnásobek počtu ionizovatelných skupin v polypeptidu (48 pro $N = 24$ a 96 pro $N = 48$).

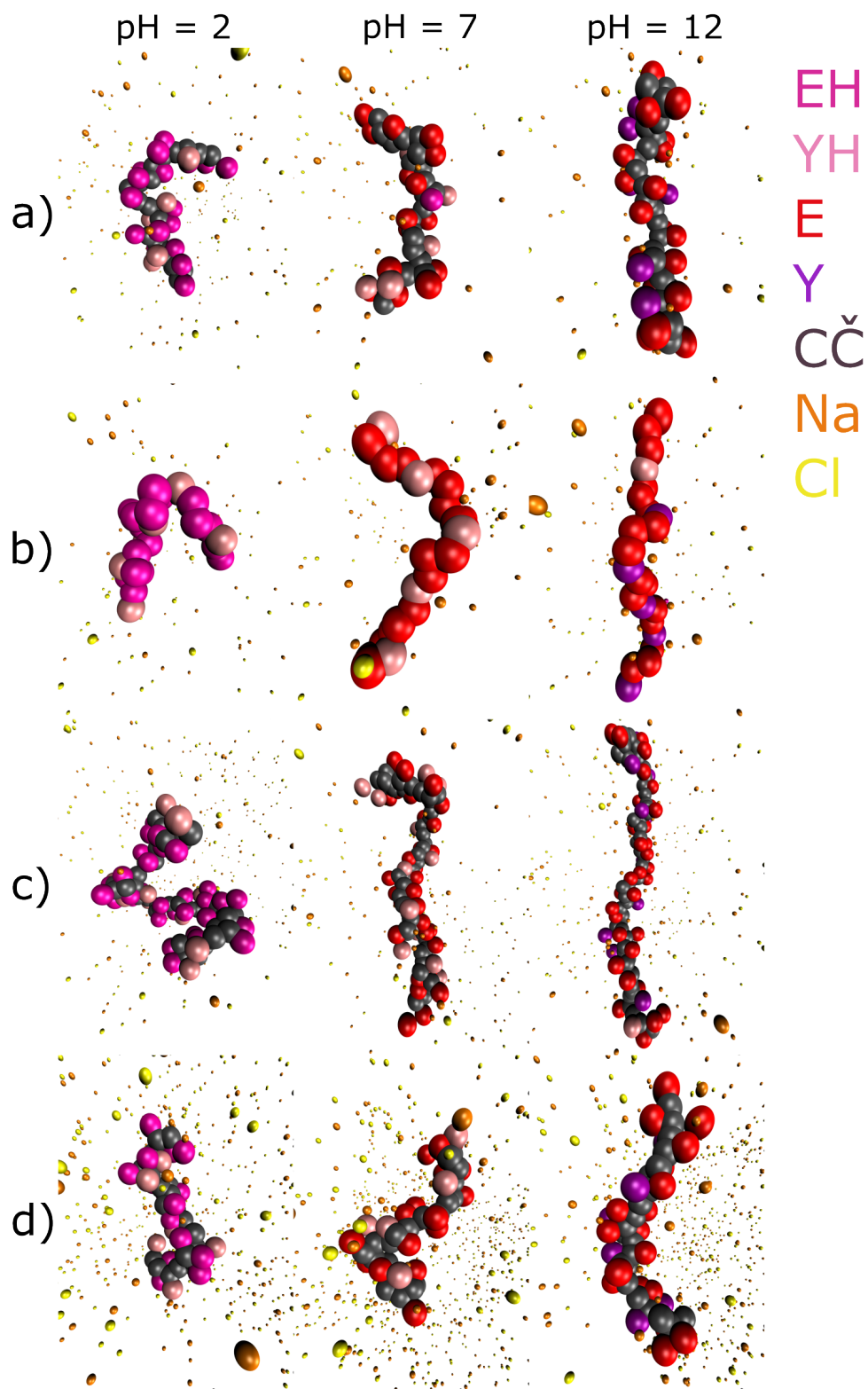
Všechny simulace byly realizovány v simulačním softwaru ESPResSo [16] s knihovnou pyMBE [25] využívanou při modelování systému.

4 Výsledky a diskuze

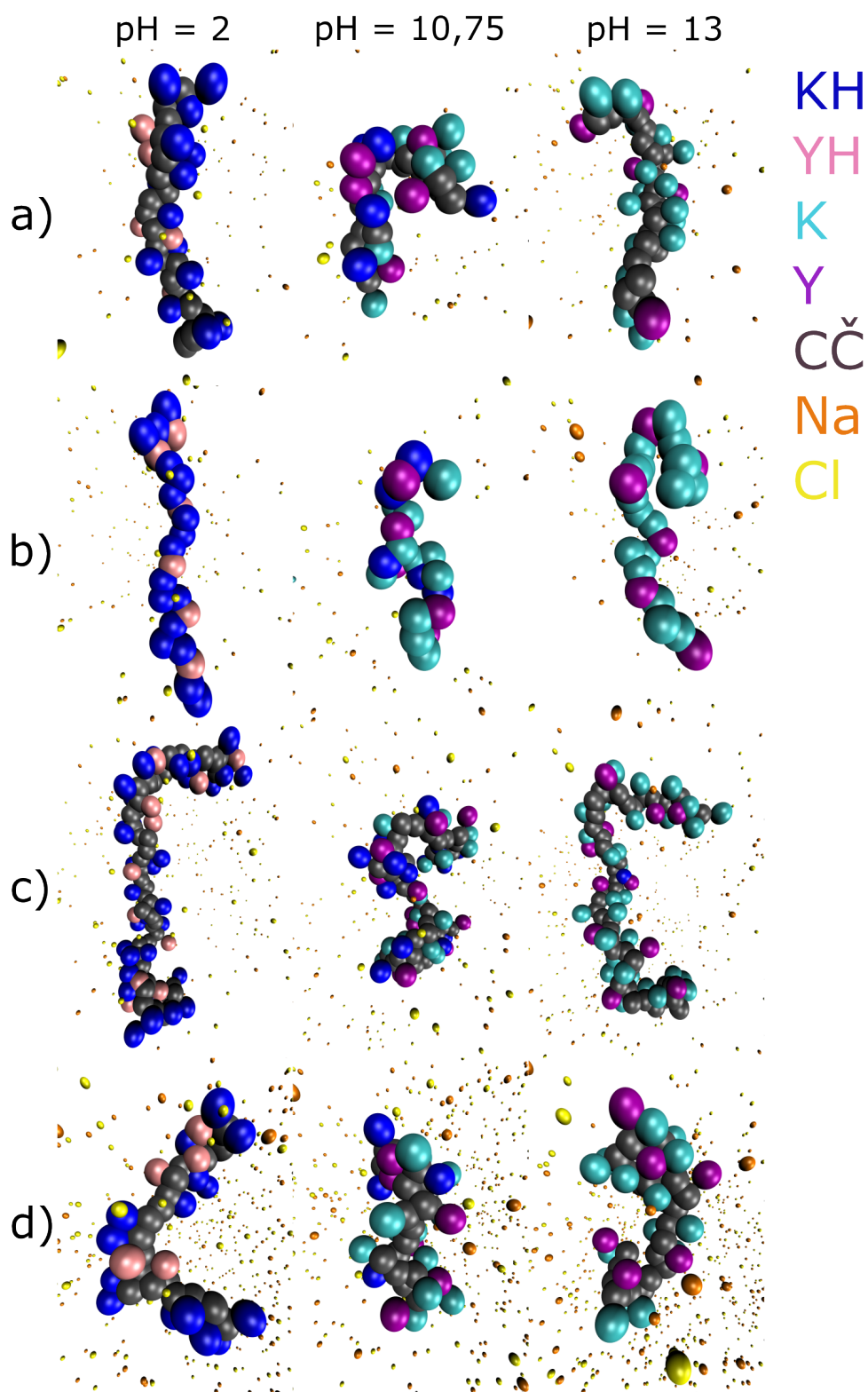
Základními výstupy ze simulací jsou titrační křivky polypeptidů (závislost náboje polypeptidu na pH), titrační křivky jednotlivých aminokyselin vázaných v polypeptidech (závislost průměrného stupně ionizace aminokyselin daného typu na pH) a závislost gyračního poloměru polypeptidu na pH. V rámci těchto výsledků jsou nejprve separátně diskutovány oba studované polypeptidy, jak se jejich titrační křivky odchyľují od těch ideálních, jak je ovlivněn jejich gyrační poloměr přítomností nábojů na řetězci a v čem se ve svém chování od sebe polypeptidy liší vzhledem k tomu, že jsou složeny z jiných aminokyselin a polypeptid EEEY je polykyselina, zatímco polypeptid KKKY je polyamfolyt.

Tyto jevy jsou demonstrovány na výsledcích ze simulací s parametry: two bead model polypeptidu, délka řetězce 24 aminokyselin ($N = 24$) a koncentrace soli $C_s = 32$ mM. Tyto výsledky jsou dále porovnány s výsledky ze simulací, ve kterých je vždy jeden z těchto parametrů změněn. Takto je možné pozorovat vliv na chování polypeptidů v simulaci, pokud jsou modelovány pomocí one bead modelu namísto two bead modelu, pokud je jejich délka zvětšena na 48 aminokyselin a pokud je koncentrace přidané soli v roztoku zvýšena na $C_s = 154$ mM.

Simulační snapshoty ze všech výše zmíněných variant simulovaných systémů jsou vyobrazeny na obr. 4.1 pro polypeptid EEEY a na obr. 4.2 pro polypeptid KKKY a je na nich zároveň ilustrováno jak se mění stupeň ionizace a konformace polypeptidů při různých pH.



Obrázek 4.1 Snapshots vytvořené pomocí VMD [26] ze simulací polypeptidu EEEY při různých pH s parametry: a) two bead model, $N = 24$, $C_s = 32$ mM b) one bead model, $N = 24$, $C_s = 32$ mM c) two bead model, $N = 48$, $C_s = 32$ mM d) two bead model, $N = 24$, $C_s = 154$ mM; jednopísmenná zkratka aminokyseliny + písmeno H značí protonovanou variantu aminokyseliny, samotná zkratka značí deprotonovanou aminokyselinu, CČ je neutrální centrální částice (viz. 3.1.1), Na a Cl jsou kladně (Na) a záporně (Cl) ionty soli a proti-ionty (částice polypeptidu jsou pro účely vizualizace zvětšeny)



Obrázek 4.2 Snapshots vytvořené pomocí VMD ze simulací polypeptidu KKKY při různých pH s parametry: a) two bead model, $N = 24$, $C_s = 32$ mM b) one bead model, $N = 24$, $C_s = 32$ mM c) two bead model, $N = 48$, $C_s = 32$ mM d) two bead model, $N = 24$, $C_s = 154$ mM; jednopísmenná zkratka aminokyseliny + písmeno H značí protonovanou variantu aminokyseliny, samotná zkratka značí deprotonovanou aminokyselinu, CČ je neutrální centrální částice (viz. 3.1.1), Na a Cl jsou kladně (Na) a záporně (Cl) nabitě ionty soli a proti-ionty (částice polypeptidu jsou pro účely vizualizace zvětšeny)

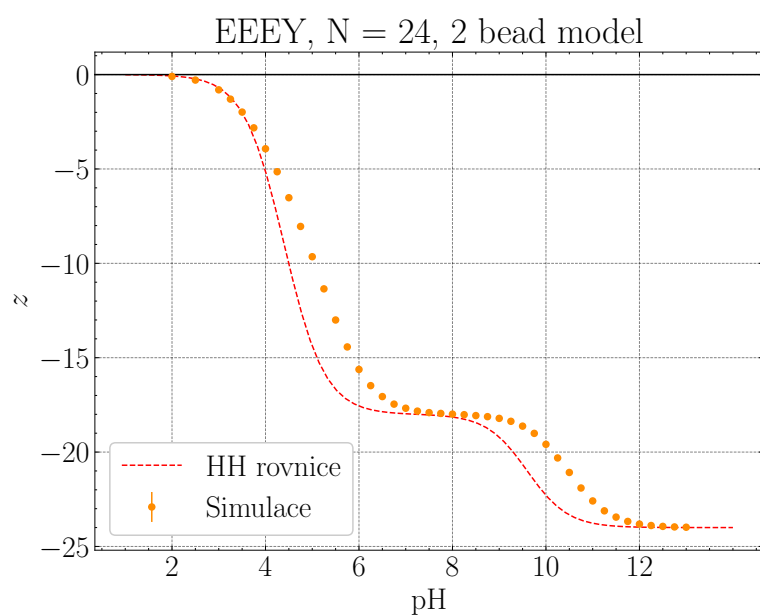
4.1 Polypeptid EEEY

Ideální křivka spočítaná podle Hendersonovy–Hasselbalchovy rovnice (2.20) pro polypeptid EEEY na grafu 4.3 má dvě vlny, což odpovídá tomu, že je polypeptid tvořen dvěma aminokyselinami, jejichž postranní řetězce mají vzdálené hodnoty pK_A ($pK_A(E) = 4,4$; $pK_A(Y) = 9,6$). Každá vlna pak odpovídá ionizaci jedné aminokyseliny. Izoelektrický bod polypeptidu leží daleko v kyselé oblasti pH, protože neobsahuje žádné zásadité skupiny.

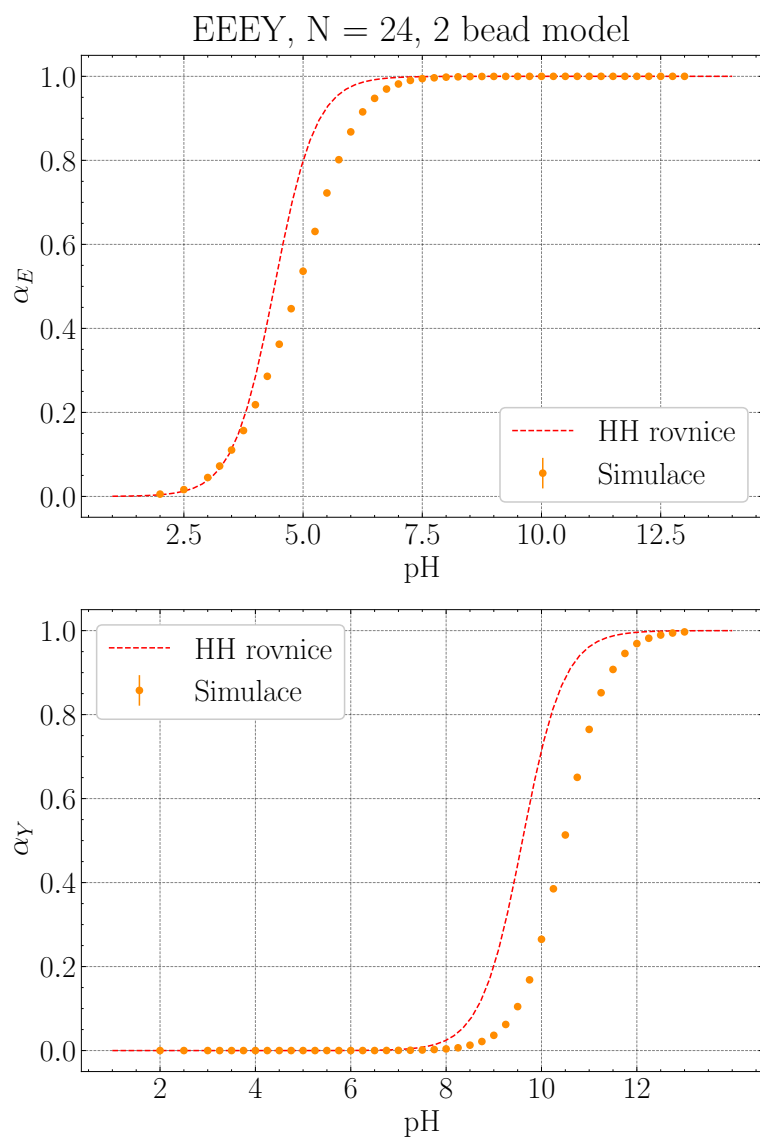
Na výsledcích ze simulací (také v grafu 4.3) je vidět odchýlení od ideální titrační křivky směrem vpravo, tedy pro konkrétní hodnotu pH nese simulovaný polypeptid vždy nižší absolutní náboj než v ideálním případě. To je v souladu s očekávaným snížením ionizace postranních řetězců aminokyselin v důsledku energeticky nevýhodné elektrostatické repulze. Ideální titrační křivka a simulační data jsou totožná pro silně kyselé pH, kde je pH příliš nízké pro ionizaci obou aminokyselin, pro silně zásadité pH, kde už jsou obě aminokyseliny plně ionizované, a pro oblast okolo $pH = 7,5$, kde je glutamát již plně ionizován, ale tyrosin se ještě ionizovat nezačal.

Z grafu 4.3 je patrné, že odchylka od ideality v oblasti ionizace glutamátu není konstantní, ale roste s pH – simulovaná titrační křivka je deformovaná. Zatímco v oblasti ionizace tyrosinu je odchylka od ideality v průběhu titrace konstantní a titrační křivka je pouze posunutá. Tento jev lze dobře pozorovat na grafu 4.4, kde jsou vidět samostatné titrační křivky obou aminokyselin uvnitř polypeptidu. Toto pozorování lze vysvětlit na základě geometrie molekuly polypeptidu. V polypeptidové sekvenci se vždy nachází tři glutamátová rezidua vedle sebe. I když na počátku titrace se mohou ionizovat glutamáty, které jsou daleko od sebe, a tudíž se ovlivňují pouze minimálně a odchylka od ideality je malá, s rostoucím pH (a tedy i počtem ionizovaných glutamátů) se další glutamáty, které se mají ionizovat, už nacházejí v prostředí, kde jejich sousedé nesou náboj. A každý další ionizovaný glutamát tento efekt zvyšuje, takže odchylka od ideality roste s pH. Tyrosinová rezidua jsou v polypeptidu vždy na obou stranách obklopena třemi glutamátovými rezidui. Ionizace jednoho tyrosinu nemá pak prakticky žádný vliv na ostatní tyrosiny. Protože k ionizaci tyrosinu dochází až v oblasti titrační křivky, kde je glutamát plně ionizován, tyrosiny se efektivně ionizují v neměnném prostředí záporně nabitých glutamátů, což vede ke konstantní odchylce od ideálního chování.

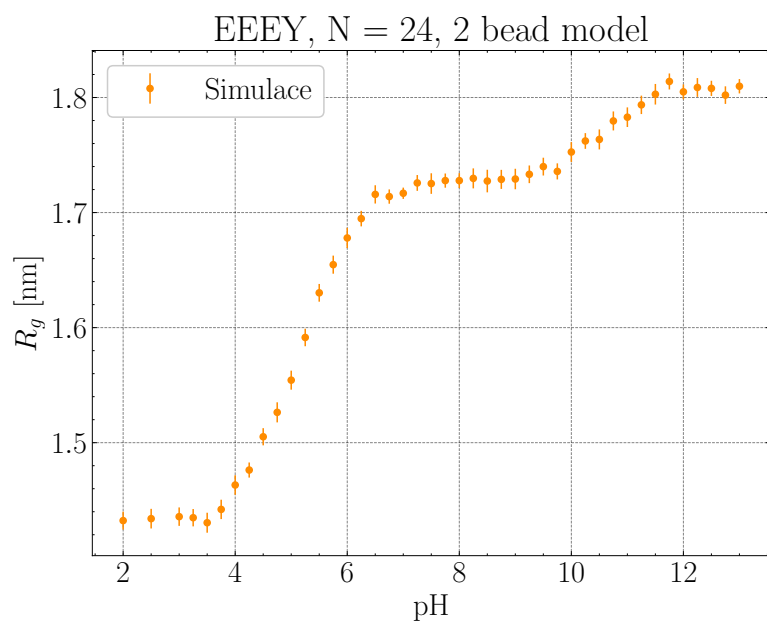
Na grafu 4.5 je vidět závislost gyračního poloměru polypeptidu EEEY na pH. Minimum gyračního poloměru se nachází v silně kyselé oblasti pH, což odpovídá polypeptidu, který je prakticky nenabitý. Gyrační poloměr pak roste s pH ve dvou vlnách, které odpovídají vlnám na titrační křivce polypeptidu (graf 4.3). Polypeptid tedy zaujímá více rozbalenou konformaci s tím jak roste jeho náboj, což lze jednoduše vysvětlit repulzivními elektrostatickými interakcemi mezi nabitými aminokyselinami. Nárůst v první vlně je větší než při druhé vlně, protože při ionizaci všech glutamátů v molekule, která této vlně koresponduje, se zvětší náboj polypeptidu více než při ionizaci tyrosinů, kterých je v molekule méně.



Obrázek 4.3 Závislost celkového náboje molekuly (z) na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEY), o délce 24 aminokyselin, modelovaný two bead modelem; porovnání výsledků ze simulace s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.4 Závislost stupně ionizace (α) aminokyselin glutamátu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci EEEEY, o délce 24 aminokyselin, modelovaný two bead modelem; porovnání výsledků ze simulace s ideální titrační křivkou podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.5 Závislost gyračního poloměru (R_g) v nm na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEY), o délce 24 aminokyselin, modelovaný two bead modelem

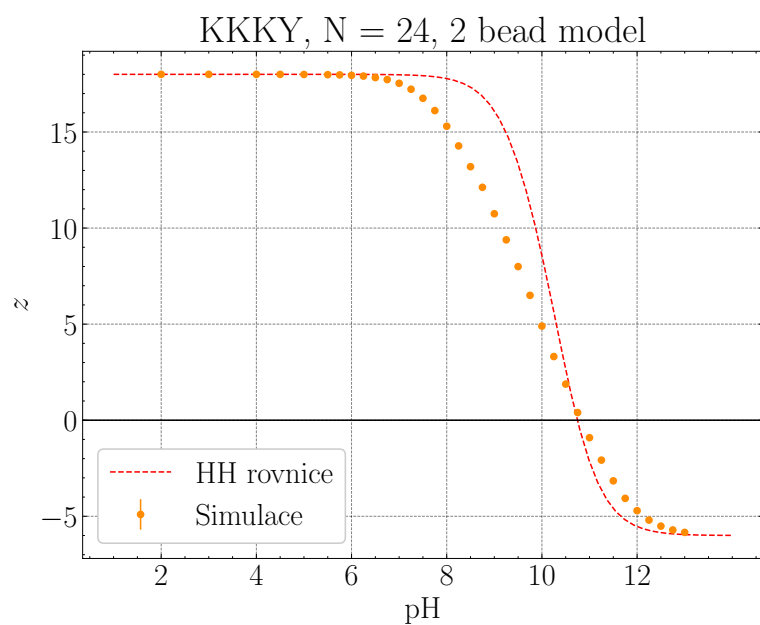
4.2 Polypeptid KKKY

Graf 4.6 ukazuje ideální a simulované titrační křivky polypeptidu KKKY. Titrační křivky jsou na rozdíl od polypeptidu EEEY tvořené pouze jednou vlnou. Postranní řetězce lysinu a tyrosinu mají pK_A s podobnými hodnotami ($pK_A(K) = 10,4$; $pK_A(Y) = 9,6$), které obě leží v zásadité oblasti pH. Polypeptid obsahuje kyselé i zásadité skupiny, takže vlna na titrační křivce prochází nulou, a jeho izoelektrický bod leží okolo $pH = 10,75$. Stejně jako u polypeptidu EEEY je absolutní náboj polypeptidu KKKY vždy stejný nebo menší než podle ideální titrační křivky, což už ale nelze vysvětlit prostým snížením ionizace skupin na polypeptidu v důsledku elektrostatické repulze. Protože některé tyto skupiny mohou nést kladný náboj a některé mohou nést záporný náboj, můžou se náboje na polypeptidu vzájemně kompenzovat.

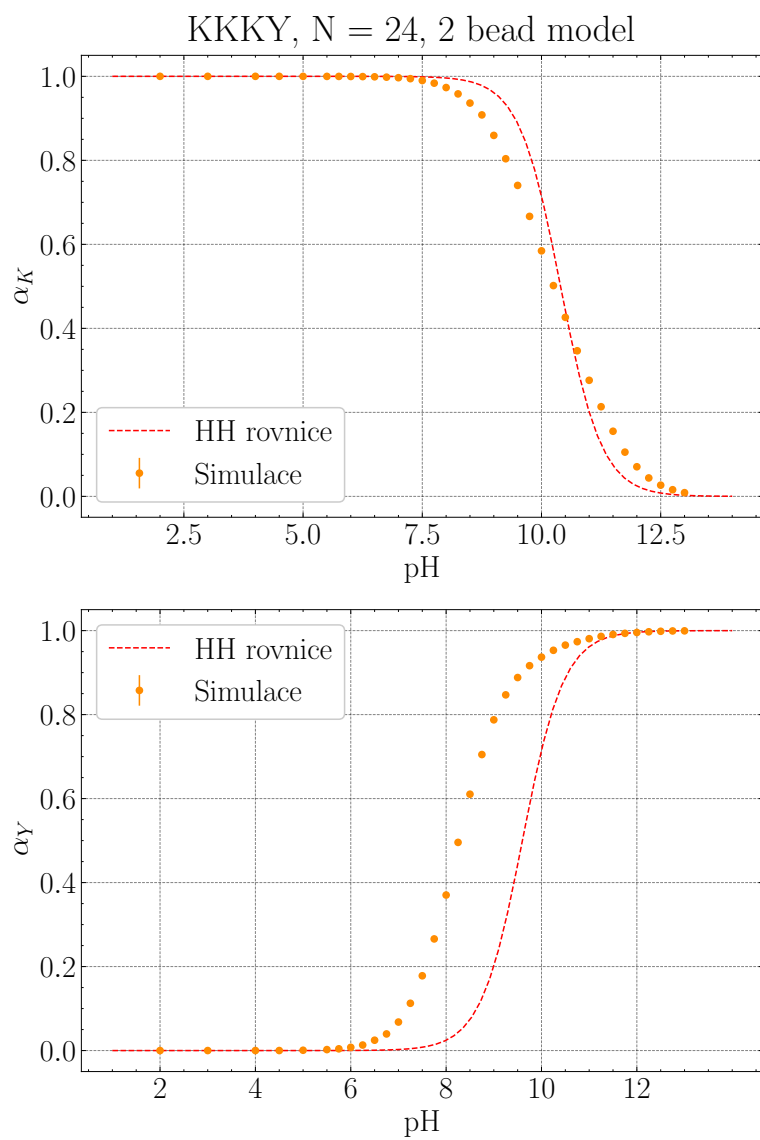
Z grafu 4.7, který ukazuje individuální titrační křivky aminokyselin z polypeptidu, je patrné, že v polypeptidu KKKY elektrostatické interakce způsobují v různých oblastech pH jak snížení tak i zvýšení ionizace aminokyselin. V oblasti $pH < pI$ je ionizace lysinu snížena. Sekvence polypeptidu má stejnou strukturu jako u polypeptidu EEEY, takže je vidět podobný efekt, kde je odchylka od ideálního chování jedné aminokyseliny závislá na pH a druhá nikoliv. V důsledku repulze mezi nabitými lysiny, které se nacházejí dost blízko sebe, aby se mohli navzájem ovlivnit, je ionizace lysinů snížena. Odchylkou od ideality se pak zvětšuje s rostoucím stupněm ionizace lysinu. V případě lysinu ovšem tato odchylka s pH klesá, zatímco u glutamátu roste. To je dáno tím, že lysin má na rozdíl od glutamátu zásaditý postranní řetězec a také tím, že s rostoucím pH stoupá stupeň ionizace tyrosinu, jehož záporný náboj kompenzuje kladný náboj lysinu.

Ionizace tyrosinu je systematicky vyšší než v ideálním případě a jeho simulovaná titrační křivka je prakticky nedeformovaná. Toto lze vysvětlit tím, že tyrosin se ionizuje v prostředí dominovaném kladně nabitými lysiny, které stabilizuje záporný náboj na ionizovaném tyrosinu. Díky stejným přitažlivým elektrostatickým interakcím je v oblasti $pH > pI$ naopak mírně zvýšena ionizace lysinu. Tyrosin je v této oblasti plně ionizován ale tyrosinů je v polypeptidu relativně méně, takže tento opačný efekt není tak silný. V obou případech zvýšená ionizace jedné aminokyseliny částečně kompenzuje náboje vzniklé ionizací té druhé a celkově snižuje náboj polypeptidu.

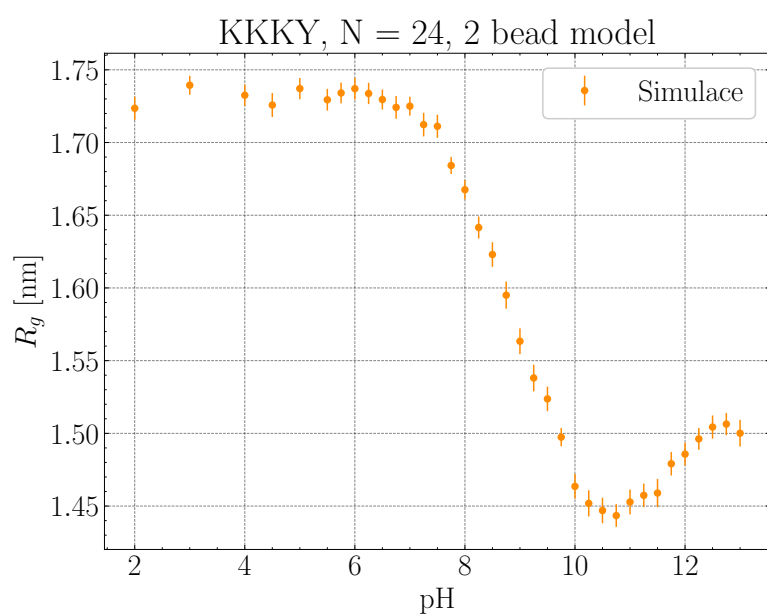
V grafu 4.8 je ukázána závislost gyračního poloměru polypeptidu KKKY na pH. Gyrační poloměr opět obecně roste s celkovým absolutním nábojem polypeptidu. V izoelektrickém bodě má gyrační poloměr polypeptidu minimum, protože molekula má efektivně nulový náboj.



Obrázek 4.6 Závislost celkového náboje molekuly (z) na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), o délce 24 aminokyselin, modelovaný two bead modelem; porovnání výsledků ze simulace s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.7 Závislost stupně ionizace (α) aminokyselin lysinu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci KKKY, o délce 24 aminokyselin, modelovaný two bead modelem; porovnání výsledků ze simulace s ideální titrační křivkou podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.8 Závislost gyračního poloměru (R_g) v nm na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), o délce 24 aminokyselin, modelovaný two bead modelem

4.3 Vliv zvoleného modelu

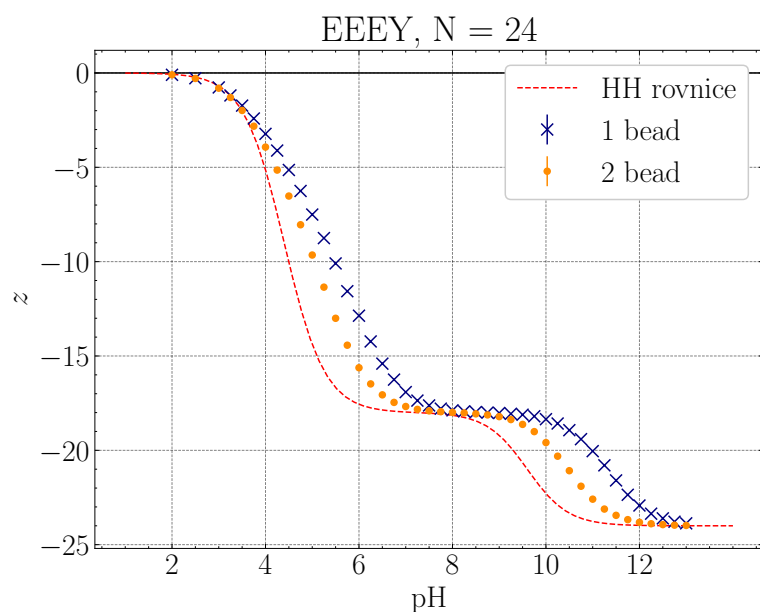
V grafech 4.9 a 4.10 jsou porovnány titrační křivky polypeptidu EEEY a individuálních aminokyselin v něm z předchozích sekcí, kde byl použit two bead model, se simulacemi, kde byl použit one bead model. Kvalitativně oba modely ukazují podobné chování, ale na obou grafech je vidět, že odchýlení od ideálního chování je pro one bead model větší než pro two bead model. Celkový absolutní náboj polypeptidu (graf 4.9) je pro one bead model nižší, protože aminokyseliny jsou méně ionizované (graf 4.10).

Stejná srovnání mezi modely pro závislosti náboje polypeptidu a stupně ionizace aminokyselin na pH jsou pro polypeptid KKKY ukázána v grafech 4.11 a 4.12. Kvantitativně má opět polypeptid modelovaný one bead modelem nižší celkový absolutní náboj a z grafu 4.12 je patrné, že jak snížení ionizace lysinu v důsledku repulze, tak zvýšení ionizace tyrosinu a lysinu v důsledku přitažlivých interakcí je výraznější u one bead modelu. Obecně tedy one bead model předpovídá více extrémní chování polypeptidu oproti two bead modelu.

Tento rozdíl mezi one bead a two bead modelem je způsoben jejich rozdílnou geometrií, která má vliv na schopnost polypeptidu rozložit náboj v prostoru. Two bead model dává polypeptidu více konformačních stupňů volnosti, takže postranní řetězce aminokyselin mohou volně přejít do trans konformace a výrazně tak zvětšit vzdálenost mezi sousedními náboji oproti fixní vazebné délce mezi částicemi ve one bead modelu (viz. obr. 3.1). V trans konformaci je vzdálenost sousedních postranních řetězců jen o málo menší než vzdálenost mezi postranními řetězci ob jednu aminokyselinu (s druhým nejbližším sousedem). Větší vzdálenosti mezi náboji snižují sílu odpuzivých i přitažlivých elektrostatických interakcí a disociace aminokyselin je tedy pro two bead model méně odchýlená od ideality.

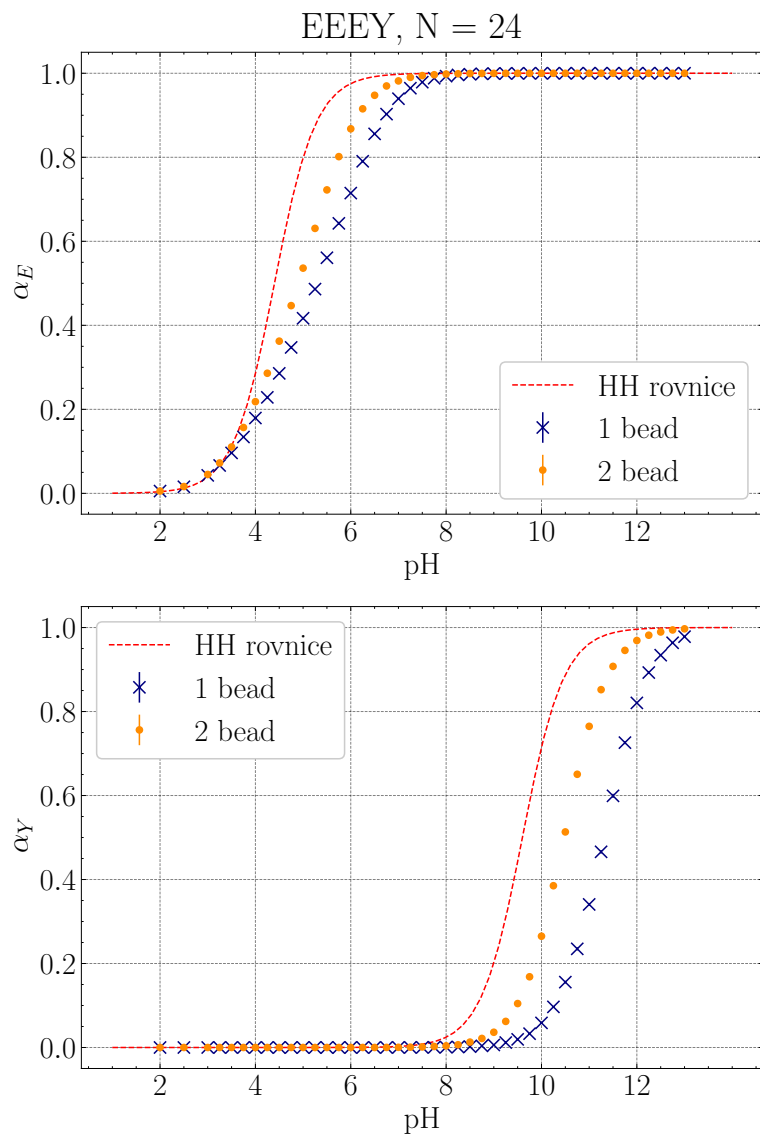
Použití těchto dvou modelů vychází z prací [9] a [13], kde jsou tímto způsobem modelovány oligopeptidy a parametry těchto zhubených modelů jsou založeny na pozorovaných vzdálenostech atomů v atomistických simulacích. Protože two bead model je svou geometrií blíže reálným polypeptidům, je zvolení vazebných délek pro tento model relativně přímočaré a v této práci bylo postupováno stejně jako v původních publikacích. Zvolení smysluplných parametrů, které nějakým způsobem odrážejí fyzikální realitu polypeptidu, je pro one bead model více problematické. Ve výše zmíněných publikacích byla vazebná délka mezi částicemi v modelu zvolena jako vzdálenost ionizovatelných skupin na sousedních postranních řetězcích v trans konformaci, což ale bez fixace vazebných úhlů vede ke zvětšení vzdáleností mezi postranními řetězci více vzdálených aminokyselin (druhý nejbližší soused) oproti polypeptidu v trans konformaci. Tento one bead model má pak naopak celkově větší vzdálenosti mezi náboji než two bead model. V této práci byla na druhou stranu jako vazebná délka ve one bead modelu zvolena prakticky stejná vazebná délka jako mezi centrálními částicemi v two bead modelu, takže titrační křivky pro one bead model lze interpretovat jako náležící polypeptidům, jejichž postranní řetězce jsou efektivně zafixované v cis konformaci.

Na grafech 4.13 a 4.14 jsou porovnány závislosti gyračního poloměru na pH mezi one bead two bead modely pro oba polypeptidy. Obecně mají polypeptidy modelované one bead modelem menší gyrační poloměr. Naším původním předpokladem bylo, že díky lepší schopnosti rozložit náboj bude mít two bead model celkově menší gyrační poloměr, ale z porovnání hodnot gyračního poloměru

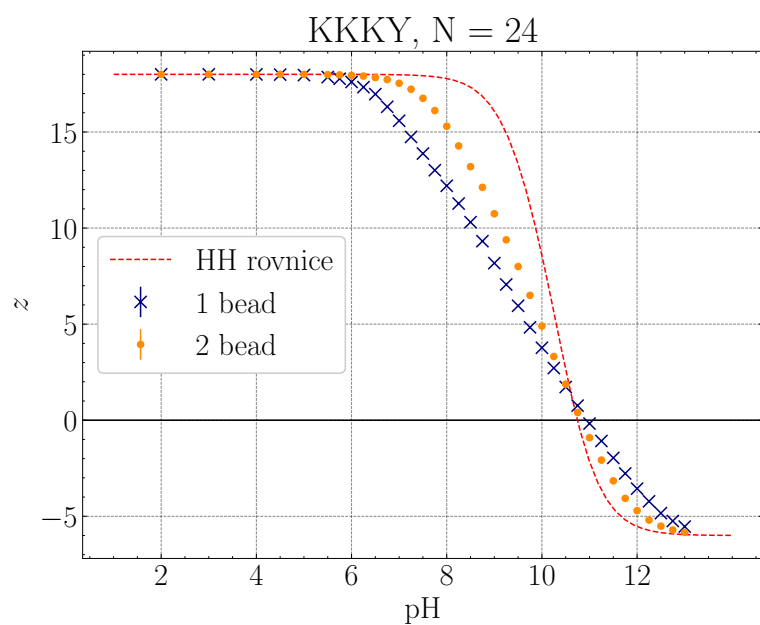


Obrázek 4.9 Závislost celkového náboje molekuly (z) na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEEY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných one bead a two bead modelem s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice

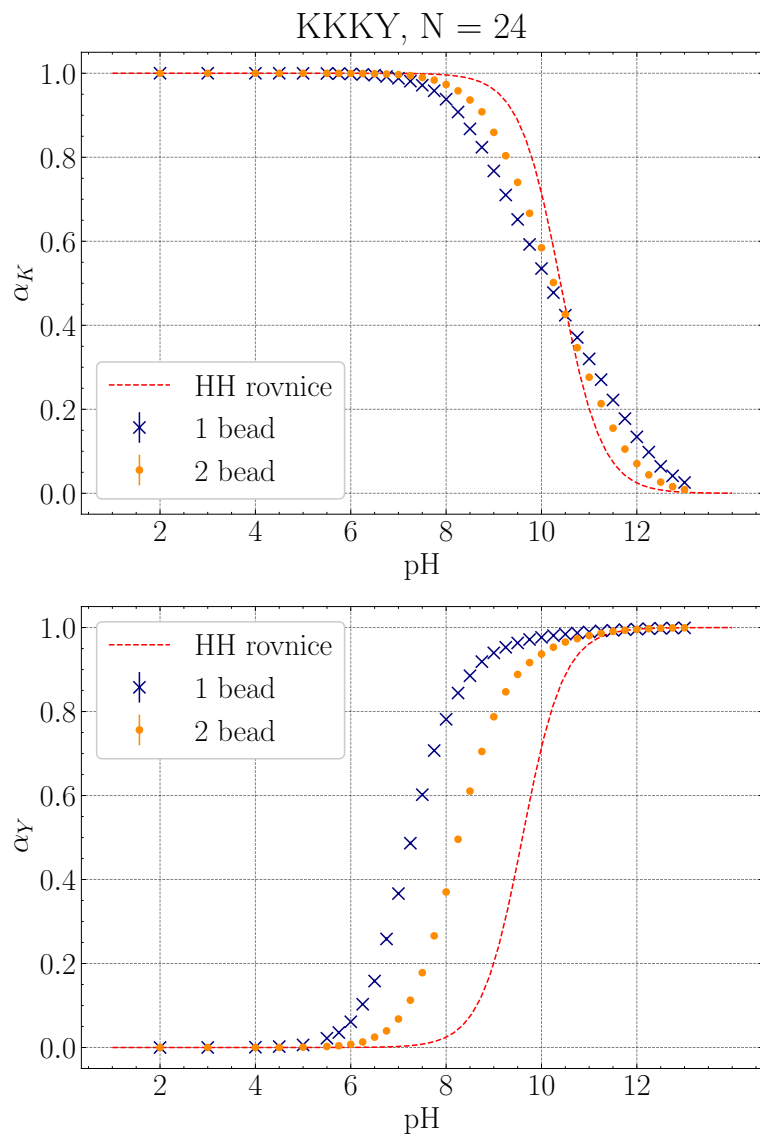
z oblastí pH, kde jsou polypeptidy nenabitě, je vidět, že přítomnost dalších částic reprezentujících postranní řetězce v two bead modelu sama o sobě způsobuje, že má takto modelovaný polypeptid větší gyrační poloměr než v případě one bead modelu. Na druhou stranu je z obou grafů patrné, že relativní zvětšení gyračního poloměru mezi nenabitým a nabitým stavem polypeptidu je významnější pro one bead model, což je v souladu s původní hypotézou, že lepší schopnost rozložit náboj umožní polypeptidu zaujmout sbalenější konformaci.



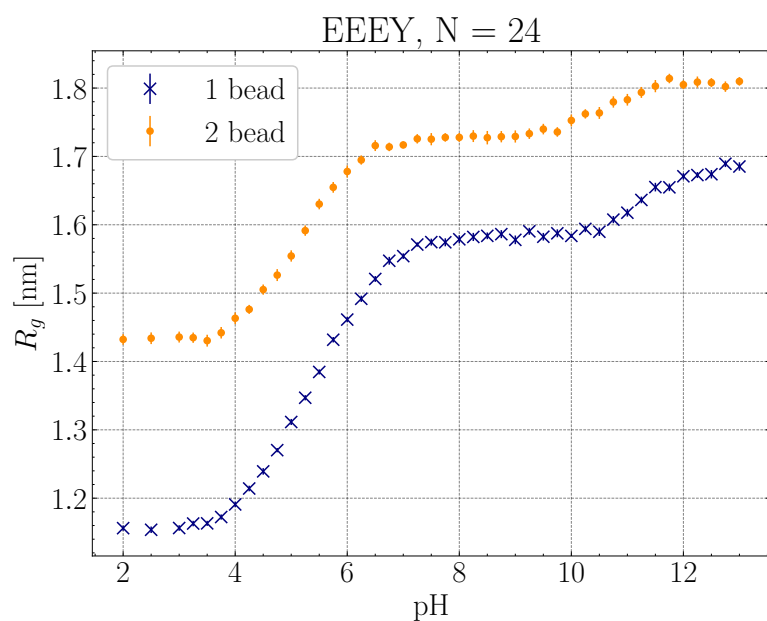
Obrázek 4.10 Závislost stupně ionizace (α) aminokyselin glutamátu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci EEEY, o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných one bead a two bead modelem s ideální titrační křivkou podle Hendersonovy–Hasselbalchovy rovnice



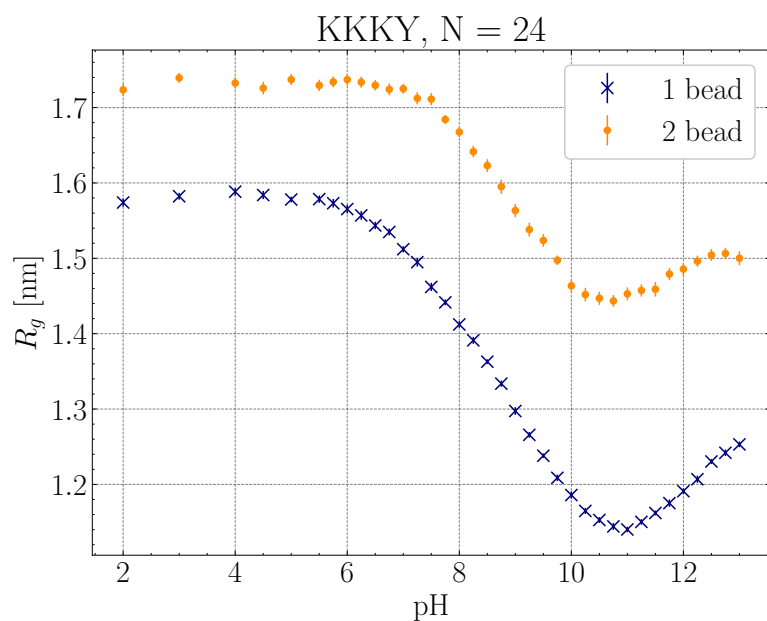
Obrázek 4.11 Závislost celkového náboje molekuly (z) na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných one bead a two bead modelem s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice



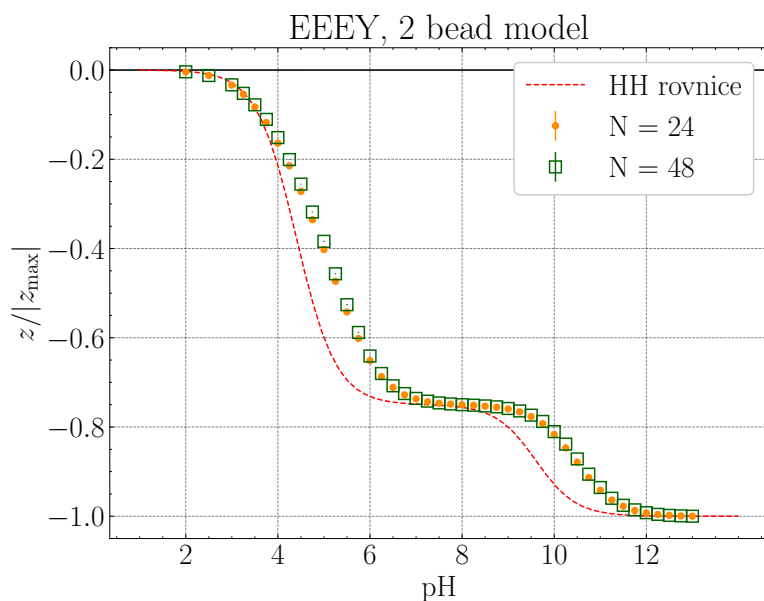
Obrázek 4.12 Závislost stupně ionizace (α) aminokyselin lysinu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci KKKY, o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných one bead a two bead modelem s ideální titrační křivkou podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.13 Závislost gyračního poloměru (R_g) v nm na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných one bead a two bead modelem



Obrázek 4.14 Závislost gyračního poloměru (R_g) v nm na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných one bead a two bead modelem

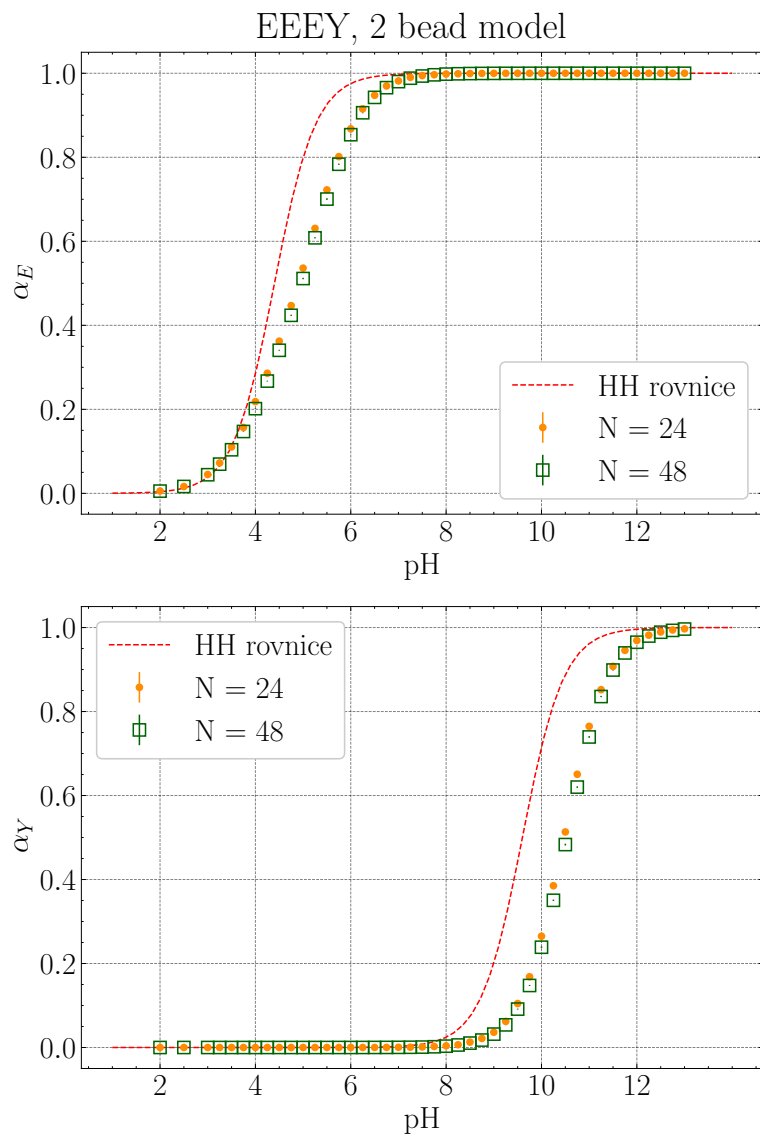


Obrázek 4.15 Závislost celkového náboje molekuly (z) děleného absolutní hodnotou maxima ($|z_{max}|$) na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEY), modelovaný two bead modelem; porovnání výsledků ze simulací polypeptidů o délce 24 a 48 aminokyselin s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice

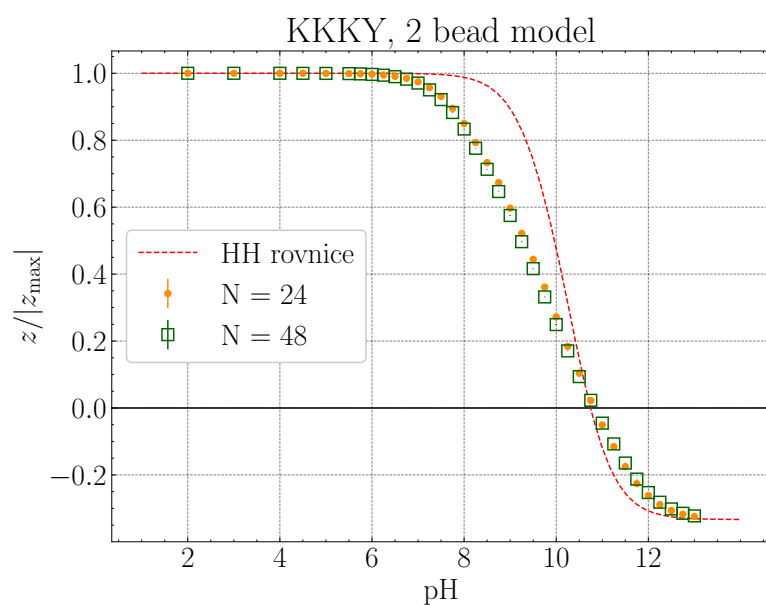
4.4 Vliv délky řetězce

Dalším zkoumaným parametrem byla délka polypeptidového řetězce. Na grafech 4.15 a 4.16 jsou titrační křivky polypeptidu a aminokyselin v něm porovnané pro polypeptidy EEEY dlouhé 24 a 48 aminokyselin. Pro polypeptidy KKKY jsou stejným způsobem porovnány titrační křivky v grafech 4.17 a 4.18. Titrační křivky polypeptidů jsou normalizované jejich maximálním absolutním nábojem, aby bylo možné provést porovnání. Ze všech čtyř grafů vyplývá, že rozdíly v ionizačním chování aminokyselin způsobené délkou polypeptidu, jsou velmi malé. Obecně vykazují delší polypeptidové řetězce ($N = 48$) ve všech směrech o něco větší odchylku od ideálního chování oproti kratším řetězcům. toto pozorování je konzistentní s předchozími studii vlivu délky slabých polyelektrolytů na jejich stupeň ionizace [27], kde zvyšující se délka vedla k větším odchylkám od ideality až po limit $N = 50$, po kterém už délka řetězce přestala mít vliv.

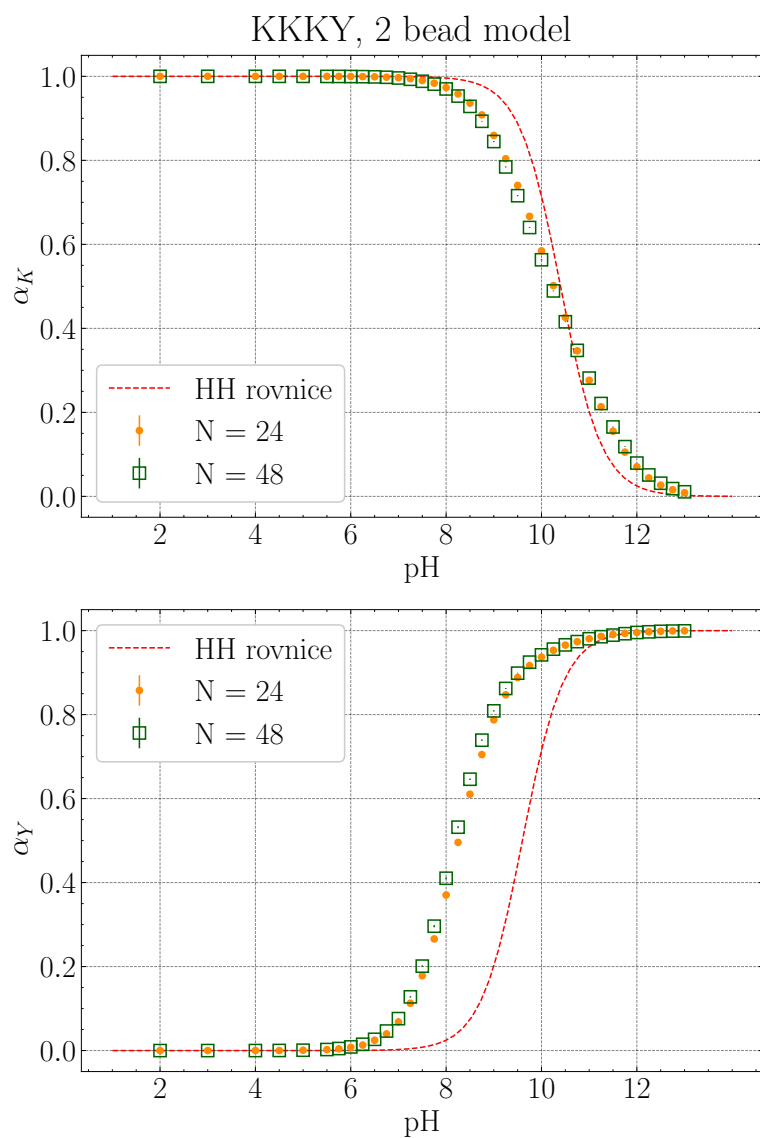
Z grafů 4.19 a 4.20 zobrazujících rozdíly v závislosti relativního gyračního poloměru ($R_g/R_{g,max}$) na pH pro polypeptidy EEEY a KKKY o různých délkách je patrné, že v oblastech pH, kde jsou polypeptidy silně nabitě, jsou pozorované hodnoty relativního gyračního poloměru pro různě dlouhé polypeptidy ve shodě. Když však náboj polypeptidu klesne, nabývají delší řetězce relativně více sbalenou konformaci (v porovnání s jejich maximálním natažením) oproti kratším řetězcům. Možným vysvětlením by mohlo být, že malé množství nábojů, které polypeptidy v těchto oblastech pH nesou, se může v delších řetězcích rozmístit výhodněji a způsobit tak pozorovanou nižší hodnotu relativního gyračního poloměru. Úplné objasnění jevu však v průběhu práce získáno nebylo.



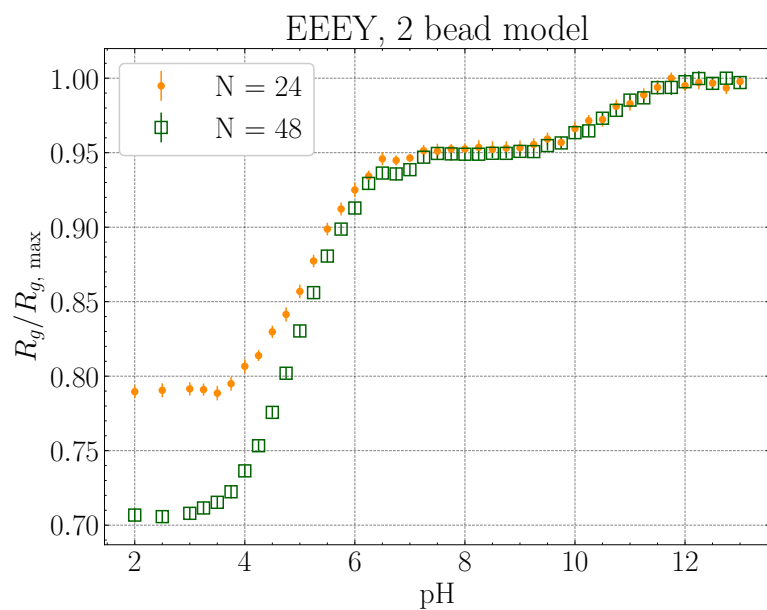
Obrázek 4.16 Závislost stupně ionizace (α) aminokyselin glutamátu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci EEEY, modelovaný two bead modelem; porovnání výsledků ze simulací polypeptidů o délce 24 a 48 aminokyselin s ideální titrační křivkou podle Hendersonovy–Hasselbalchovy rovnice



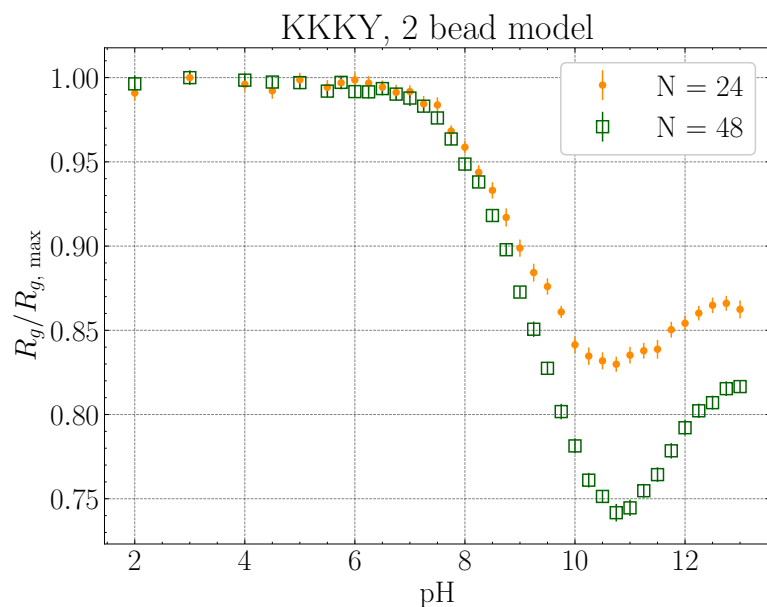
Obrázek 4.17 Závislost celkového náboje molekuly (z) děleného absolutní hodnotou maxima ($|z_{max}|$) na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), modelovaný two bead modelem; porovnání výsledků ze simulací polypeptidů o délce 24 a 48 aminokyselin s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.18 Závislost stupně ionizace (α) aminokyselin lysinu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci KKKY, modelovaný two bead modelem; porovnání výsledků ze simulací polypeptidů o délce 24 a 48 aminokyselin s ideální titrační křivkou podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.19 Závislost gyačního poloměru (R_g) děleného hodnotou maxima ($R_{g,max}$) na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEY), modelovaný two bead modelem; porovnání výsledků ze simulací polypeptidů o délce 24 a 48 aminokyselin

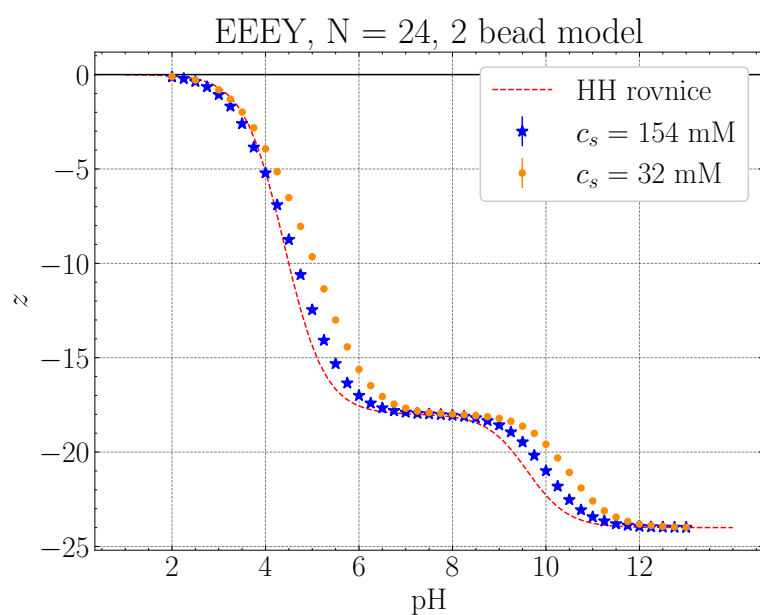


Obrázek 4.20 Závislost gyačního poloměru (R_g) děleného hodnotou maxima ($R_{g,max}$) na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), modelovaný two bead modelem; porovnání výsledků ze simulací polypeptidů o délce 24 a 48 aminokyselin

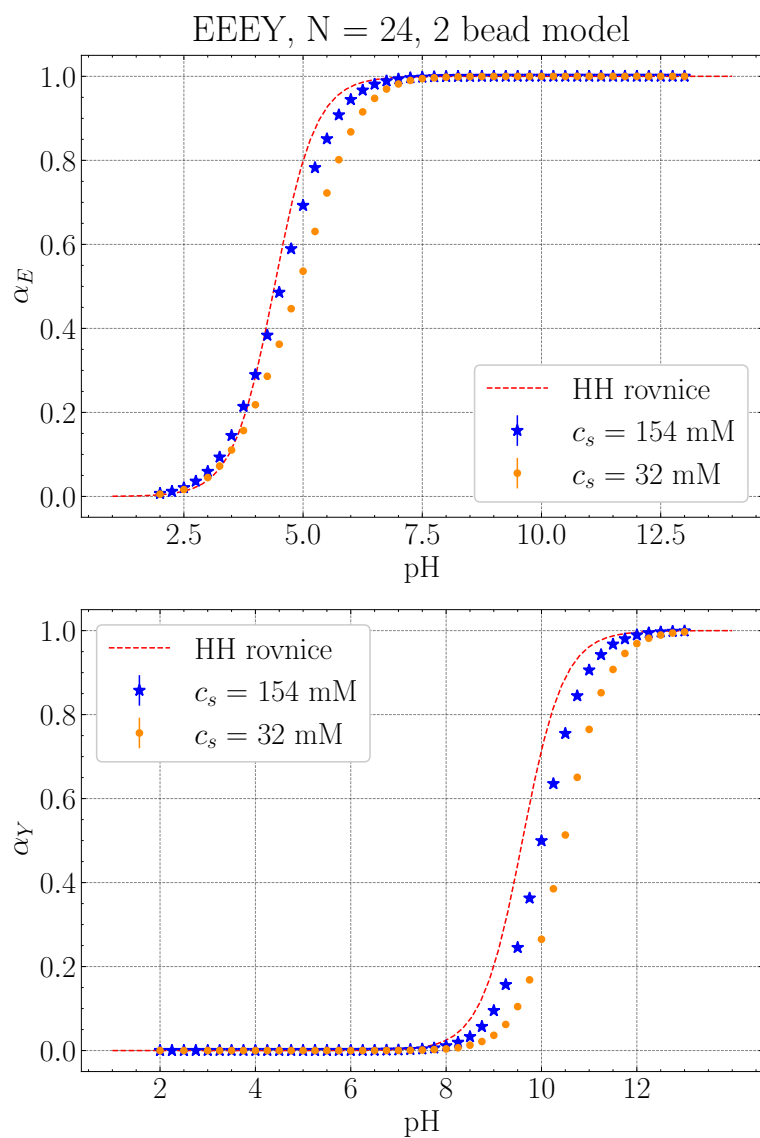
4.5 Vliv koncentrace přidané soli

Posledním studovaným parametrem byla koncentrace soli v roztoku. Na grafech 4.21 a 4.22 jsou titrační křivky polypeptidu EEEY a jeho individuálních aminokyselin v roztocích s koncentrací soli 32 mM a 154 mM (fyziologický roztok). Na grafech 4.23 a 4.24 jsou pak stejné titrační křivky pro polypeptid KKKY. Obecně lze říct, že vyšší koncentrace soli vede k posunu titračních křivek blíže k ideálním titračním křivkám, což odpovídá dřívějším studiím [4]. Oba polypeptidy mají vyšší celkový absolutní náboj v roztoku s vyšší koncentrací soli a ionizace aminokyselin je blíže Hendersonově–Hasselbalchově rovnici. Ionty ze soli, které mají opačný náboj než polypeptid, stíní náboje na řetězci, takže tyto náboje se pak ovlivňují méně silně. Efekt je poněkud slabší pro polypeptid KKKY (v oblasti $\text{pH} < \text{pI}$), což lze vysvětlit tím, že záporné ionty ze soli nemohou tak efektivně stínit kladně nabitě lysiny, protože na řetězci jsou zároveň přítomny záporně nabitě tyrosiny. U polypeptidu EEEY je v oblasti pH 3–4 (tedy v oblasti s nízkým stupněm ionizace polypeptidu) pozorovatelná naopak odchylka od ideální titrační křivky směrem k většímu stupni ionizace, která je způsobená stabilizací vyšší ionizace glutamátu přitažlivými interakcemi s kladně nabitými ionty soli.

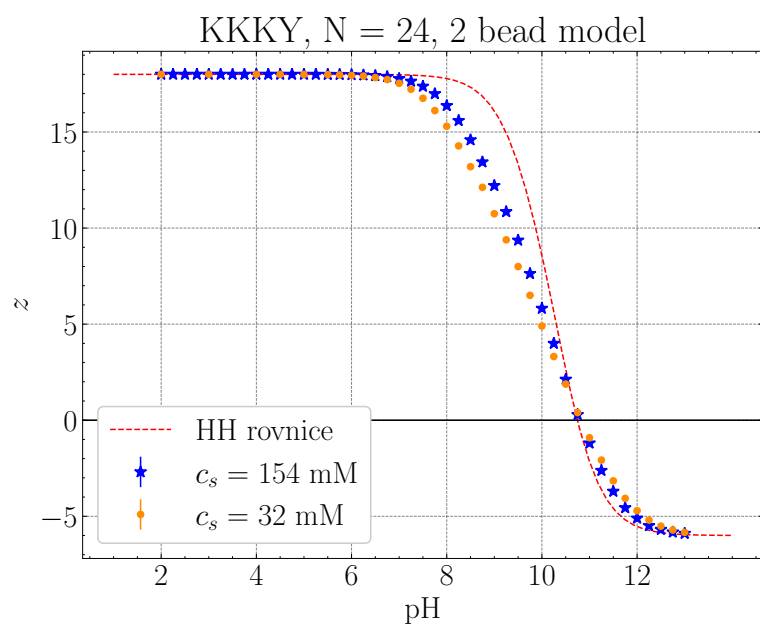
Na grafech 4.25 a 4.26 jsou zobrazeny závislosti gyračního poloměru polypeptidů EEEY a KKKY na pH opět v roztocích s různou koncentrací soli. V oblastech pH , kde nesou polypeptidy nízký náboj, panuje mezi pozorovanými hodnotami gyračního poloměru relativní shoda. Pro oblasti s více nabitými polypeptidy ale zvýšené stínění od iontů přidané soli způsobuje částečný kolaps řetězce. Polypeptidové řetězce v roztoku s $C_s = 154$ mM tedy mají menší gyrační poloměr i přesto, že jsou obecně více nabitě než polypeptidy v roztoku s $C_s = 32$ mM. Elektrostatické stínění iontů ze soli překoná vyšší náboj polypeptidu a umožní mu existovat ve více sbalení konformaci.



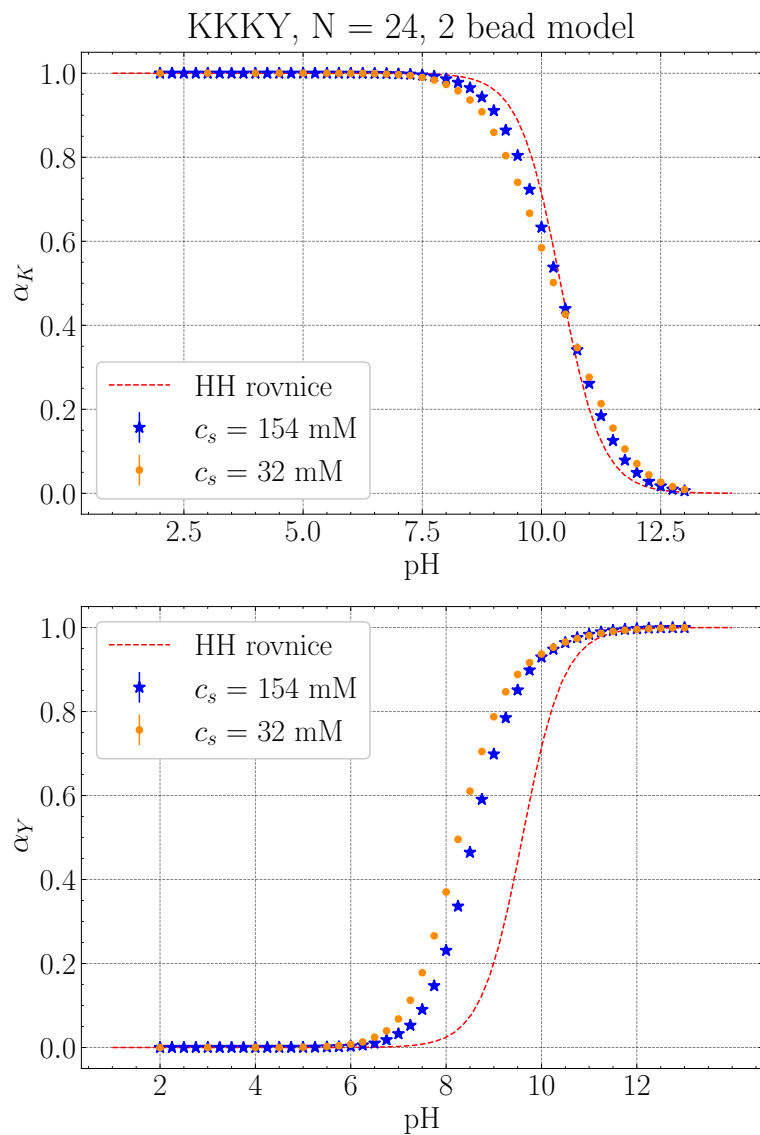
Obrázek 4.21 Závislost celkového náboje molekuly (z) na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů s $C_s = 32$ mM a $C_s = 154$ mM s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice



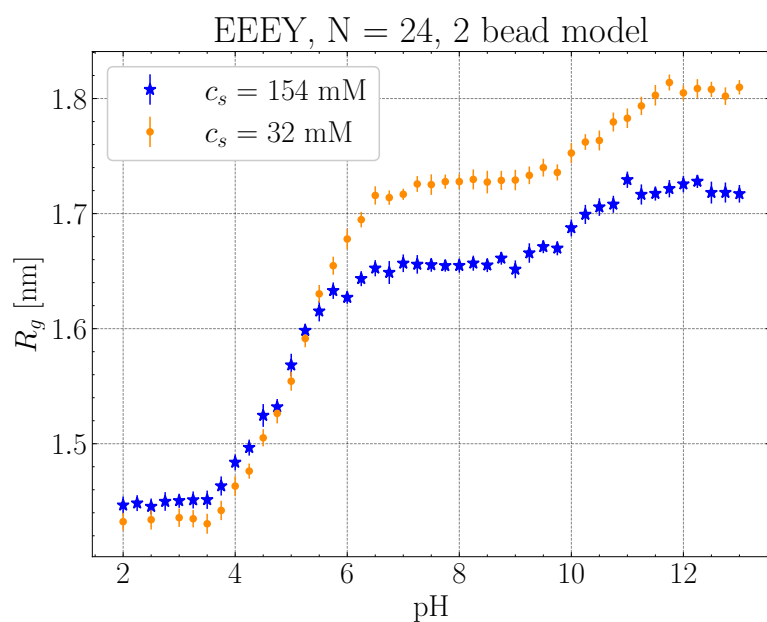
Obrázek 4.22 Závislost stupně ionizace (α) aminokyselin glutamátu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci EEEEY, o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů s $C_s = 32$ mM a $C_s = 154$ mM s ideální titrační křivkou podle Hendersonovy–Hasselbalchovy rovnice



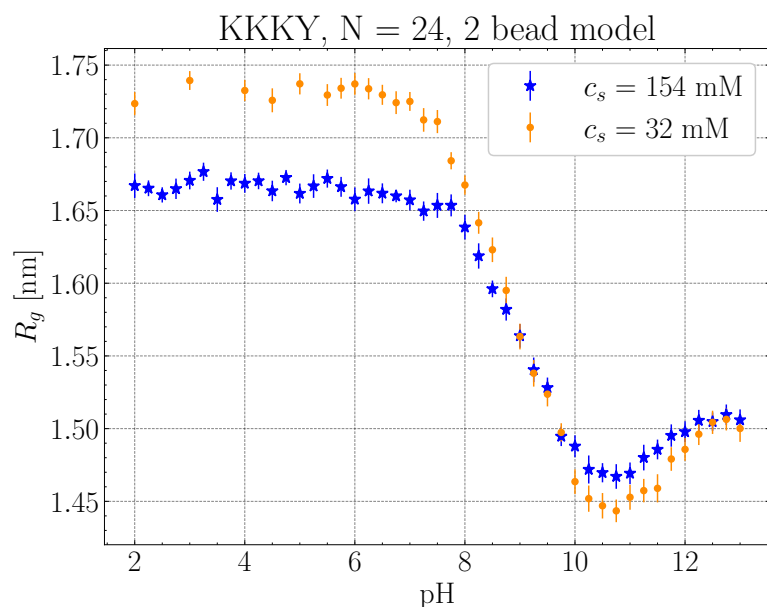
Obrázek 4.23 Závislost celkového náboje molekuly (z) na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných s $C_s = 32$ mM a $C_s = 154$ mM s nábojem molekuly podle ideální titrační křivky podle Hendersonovy–Hasselbalchovy rovnice



Obrázek 4.24 Závislost stupně ionizace (α) aminokyselin lysinu (nahore) a tyrosinu (dole) tvořících polypeptid složený z opakujících se podjednotek o sekvenci KKKY, o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných s $C_s = 32$ mM a $C_s = 154$ mM s ideální titrační křivkou podle Hendersony–Hasselbalchovy rovnice



Obrázek 4.25 Závislost gyračního poloměru (R_g) v nm na pH pro polypeptid složený z opakujících se podjednotek obsahujících glutamát a tyrosin (sekvence podjednotky: EEEY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů s $C_s = 32$ mM a $C_s = 154$ mM



Obrázek 4.26 Závislost gyračního poloměru (R_g) v nm na pH pro polypeptid složený z opakujících se podjednotek obsahujících lysin a tyrosin (sekvence podjednotky: KKKY), o délce 24 aminokyselin; porovnání výsledků ze simulací polypeptidů modelovaných s $C_s = 32$ mM a $C_s = 154$ mM

5 Závěr a výhled do budoucna

Simulace polypeptidů modelovaných v této práci podaly výsledky, které se alespoň kvalitativně shodují s obecnými poznatky o slabých polyelektrolytech. Elektrostatické interakce mezi náboji aminokyselin v polypeptidech vytváří odchylky od ideální disociace kyselin a zásad popsané Hendersonovou–Hasselbalchovou rovnicí. Zároveň závislost konformace polypeptidu na pH kopíruje jeho celkový náboj a při jeho zvětšení se polypeptid více rozbalí.

Rozdíly v chování dvou studovaných polypeptidových sekvencí (označovaných EEEY a KKKY) lze do velké míry vysvětlit tím, že polypeptid EEEY je čistá polykyselina, zatímco polypeptid KKKY je polyamfolyt, který může nést kladné i záporné náboje. V polypeptidu KKKY pak hrají roli jak odpudivé tak přitažlivé elektrostatické interakce. Sekvence polypeptidu také hraje roli v tom, jak budou vypadat titrační křivky jednotlivých typů aminokyselin. Zda bude titrační křivka aminokyseliny pouze posunutá nebo i deformovaná je dáno typem ionizovatelných skupin v jejím okolí.

Polypeptidy byly modelovány pomocí dvou zhrubených modelů – one bead modelu a two bead modelu. Ačkoliv one bead model sestává z méně částic, a tudíž klade menší výpočetní nároky, konformační stupně volnosti, které poskytuje two bead model, mají zásadní vliv na rozmístění náboje v molekule polypeptidu. To pak celkově vede ke změně stupně ionizace polypeptidu, takže predikce podané těmito modely se shodují pouze kvalitativně ale ne kvantitativně.

Simulace polypeptidů (primárně využívající two bead model) pak byly schopné reprodukovat výsledky z literatury popisující vliv délky řetězce slabého polyelektrolytu na jeho stupeň ionizace [27] – delší řetězce (stále kratší než 50 aminokyselin) vykazovaly malé zvětšení v jejich odchylce od ideálního chování oproti kratším – a vliv elektrostatického stínění od iontů soli na stupeň ionizace polyelektrolytu [4], kde zvýšená koncentrace přidané soli v simulačním boxu vedla k posunu titračních křivek blíže k ideálnímu chování.

Jedním z nedostatků použitého simulačního protokolu bylo umístování polypeptidového řetězce do simulačního boxu na základě generátoru náhodných čísel. V některých případech způsobilo nevhodné umístění s velkým množstvím překryvů mezi částicemi pád simulace. V dalších simulacích by bylo vhodné začínat vždy ze stejného předem připraveného stabilního stavu systému. Hlavním cílem do budoucna je pak porovnání výsledků ze simulací s experimentálními daty. To by mohlo ukázat, zda je použitý model polypeptidu vhodný pro další využití.

Literatura

1. KASTIN, Abba J. (ed.). *Handbook of biologically active peptides*. Second edition. Elsevier/Academic Press, 2013. ISBN 978-0-12-385096-6.
2. MAAREL, Johan R. C. van der. *Introduction To Biopolymer Physics*. World Scientific Publishing Company, 2007. ISBN 978-981-310-143-2.
3. ULBRICH, Karel; ŠUBR, Vladimír. Polymeric anticancer drugs with pH-controlled activation. *Advanced Drug Delivery Reviews*. 2004, roč. 56, č. 7, s. 1023–1050. ISSN 0169-409X. Dostupné z DOI: <https://doi.org/10.1016/j.addr.2003.10.040>.
4. LANDSGESELL, Jonas; NOVÁ, Lucie; RUD, Oleg; UHLÍK, Filip; SEAN, David; HEBBEKER, Pascal; HOLM, Christian; KOŠOVAN, Peter. Simulations of ionization equilibria in weak polyelectrolyte solutions and gels. *Soft Matter*. 2019, roč. 15, č. 6, s. 1155–1185. ISSN 1744-6848. Dostupné z DOI: [10.1039/C8SM02085J](https://doi.org/10.1039/C8SM02085J).
5. WHITE, Audrey. Effect of pH on fluorescence of tyrosine, tryptophan and related compounds. *Biochemical Journal*. 1959, roč. 71, č. 2, s. 217–220. ISSN 0264-6021. Dostupné z DOI: [10.1042/bj0710217](https://doi.org/10.1042/bj0710217).
6. ATKINS, Peter; PAULA, Julio de. *Physical Chemistry*. 8th edition. W. H. Freeman, 2006. ISBN 978-0-7167-8759-4.
7. NELSON, David L.; COX, Michael M. *Lehninger's Principles of biochemistry*. Fourth Edition. W. H. Freeman, 2004. ISBN 978-0-7167-4339-2.
8. MCMURRY, John E. *Organic Chemistry*. 8th edition. Brooks Cole, 2011. ISBN 978-0-8400-5444-9.
9. LUNKAD, Raju; MURMILIUK, Anastasiia; HEBBEKER, Pascal; BOUBLÍK, Milan; TOŠNER, Zdeněk; ŠTĚPÁNEK, Miroslav; KOŠOVAN, Peter. Quantitative prediction of charge regulation in oligopeptides. *Molecular Systems Design & Engineering*. 2021, roč. 6, č. 2, s. 122–131. ISSN 2058-9689. Dostupné z DOI: [10.1039/D0ME00147C](https://doi.org/10.1039/D0ME00147C).
10. RUBINSTEIN, M.; COLBY, Ralph H. *Polymer Physics*. Oxford University Press, 2003. ISBN 978-0-19-852059-7.
11. NAGASAWA, Mitsuru; RICE, Stuart A.; DINNER, Aaron R. *Advances in Chemical Physics, Volume 158: Physical Chemistry of Polyelectrolyte Solutions*. Wiley, 2015. ISBN 978-1-119-05708-6.
12. ULLNER, Magnus; JÖNSSON, Bo; WIDMARK, Per-Olof. Conformational properties and apparent dissociation constants of titrating polyelectrolytes: Monte Carlo simulation and scaling arguments. *The Journal of Chemical Physics*. 1994, roč. 100, č. 4, s. 3365–3366. ISSN 0021-9606. Dostupné z DOI: [10.1063/1.466378](https://doi.org/10.1063/1.466378).
13. LUNKAD, Raju; MURMILIUK, Anastasiia; TOŠNER, Zdeněk; ŠTĚPÁNEK, Miroslav; KOŠOVAN, Peter. Role of pKA in Charge Regulation and Conformation of Various Peptide Sequences. *Polymers*. 2021, roč. 13, č. 2, s. 214. ISSN 2073-4360. Dostupné z DOI: [10.3390/polym13020214](https://doi.org/10.3390/polym13020214).

14. HASS, Mathias A. S.; MULDER, Frans A. A. Contemporary NMR Studies of Protein Electrostatics. *Annual Review of Biophysics*. 2015, roč. 44, s. 53–75. ISSN 1936-1238. Dostupné z DOI: 10.1146/annurev-biophys-083012-130351.
15. WEEKS, John D.; CHANDLER, David; ANDERSEN, Hans C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *The Journal of Chemical Physics*. 1971, roč. 54, č. 12, s. 5237–5247. ISSN 0021-9606. Dostupné z DOI: 10.1063/1.1674820.
16. *ESPResSo documentation*. [B.r.]. Dostupné také z: <https://espressomd.github.io/doc4.2.1/index.html>.
17. ARCHER, Donald G.; WANG, Peiming. The Dielectric Constant of Water and Debye-Hückel Limiting Law Slopes. *Journal of Physical and Chemical Reference Data*. 1990, roč. 19, č. 2, s. 371–411. ISSN 0047-2689. Dostupné z DOI: 10.1063/1.555853.
18. WELLS, Brad A.; CHAFFEE, Alan L. Ewald Summation for Molecular Simulations. *Journal of Chemical Theory and Computation*. 2015, roč. 11, č. 8, s. 3684–3695. ISSN 1549-9618. Dostupné z DOI: 10.1021/acs.jctc.5b00093.
19. KOLAFA, Jiří. *Molekulové modelování a simulace*. 2021.
20. GRETT, Gary S.; KREMER, Kurt. Molecular dynamics simulation for polymers in the presence of a heat bath. *Physical Review A*. 1986, roč. 33, č. 5, s. 3628–3631. Dostupné z DOI: 10.1103/PhysRevA.33.3628. Publisher: American Physical Society.
21. FRENKEL, Daan; SMIT, Berend. *Understanding Molecular Simulation*. Academic Press, 2002. Dostupné z DOI: 10.1016/B978-0-12-267351-1.X5000-7.
22. REED, Christopher E.; REED, Wayne F. Monte Carlo study of titration of linear polyelectrolytes. *The Journal of Chemical Physics*. 1992, roč. 96, č. 2, s. 1609–1620. ISSN 0021-9606. Dostupné z DOI: 10.1063/1.462145.
23. LANDSGESELL, Jonas; HOLM, Christian; SMIA TEK, Jens. Simulation of weak polyelectrolytes: a comparison between the constant pH and the reaction ensemble method. *The European Physical Journal Special Topics*. 2017, roč. 226, č. 4, s. 725–736. ISSN 1951-6401. Dostupné z DOI: 10.1140/epjst/e2016-60324-3.
24. JANKE, Wolfhard. Statistical Analysis of Simulations: Data Correlations and Error Estimation. 2009. Dostupné také z: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.physik.uni-leipzig.de/~janke/Paper/nic10_423_2002.pdf&ved=2ahUKEwixn_nU5JqGAXV8rsIHRnOD6UQFnoECBcQAQ&usg=A0vVaw2ZFXtxvkq64PYzwnsf2fZ
25. BEYER, David; TORRES, Paola B.; PINEDA, Sebastian P.; NARAMBUENA, Claudio F.; GRAD, Jean-Noël; KOŠOVAN, Peter; BLANCO, Pablo M. *pyMBE: the Python-based Molecule Builder for ESPResSo*. 2024. Dostupné z arXiv: 2401.14954 [cond-mat.soft].
26. HUMPHREY, William; DALKE, Andrew; SCHULTEN, Klaus. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*. 1996, roč. 14, s. 33–38.

27. NOVÁ, Lucie; UHLÍK, Filip; KOŠOVAN, Peter. Local pH and effective pK_A of weak polyelectrolytes – insights from computer simulations. *Physical Chemistry Chemical Physics*. 2017, roč. 19, č. 22, s. 14376–14387. ISSN 1463-9084. Dostupné z DOI: 10.1039/C7CP00265C. Publisher: The Royal Society of Chemistry.