Faculty of Science
Charles University

DIPLOMA THESIS

Bc. Magdaléna Nejedlá

Molecular simulation of acid-base properties of short peptides and the Aβ-(1–42) peptide

Molekulové simulace acidobazických vlastností krátkých peptidů a peptidu Aβ-(1-42)

Department of Physical and Macromolecular Chemistry

Supervisor: doc. RNDr. Peter Košovan, Ph.D.

Study programme: Physical chemistry

Prague 2024

I would like to thank to my supervisor, Peter Košovan, who provided me with a lot of useful advice and was very helpful in drafting the text of the thesis itself. I would also like to thank Pablo Blanco who was very involved in the beginning of the amyloid peptide project and always helped in the actual set up of the simulation. I would also like to thank my friends, Katarína Markušová and Samuel van Overloop, who were instrumental in finalizing this work. Last but not least, a big thank you also goes to my dearest, Martin Šesták, who helped with the programming part of the work, was always supportive and always patient with me.

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ..... Podpis autora

Title: Molecular simulation of acid-base properties of short peptides and the Aβ-(1–42) peptide

Author: Bc. Magdaléna Nejedlá

Department: Department of Physical and Macromolecular Chemistry

Supervisor: doc. RNDr. Peter Košovan, Ph.D.

Abstract:

This thesis deals with molecular simulations of peptides in acid-base equilibrium. In the first part we focused on 4 pentapeptides. Using a coarse-grained simulation we determined the degree of ionization of their side chains, values of their effective pK_a and Hill coefficients. As a result, our CG model provides only a rough estimate of pK_a values. The results depend on the chosen model parameters, the effect of which on the final value is not yet known. In the second part of our work, we studied the adsorption of the Aβ-(1–42) peptide onto a charged surface which for our purposes represents a charged nanoparticle. Successful adsorption could be a way to prevent the Aβ-(1–42) peptide aggregation, which is suspected of causing Alzheimer's disease. Our preliminary results show that adsorption of the Aβ-(1–42) peptide onto the charged surface of a negatively charged nanoparticle occurs at low pH, $1 < \mathrm{pH} < 4$.

Keywords: acid, base, acid-base equilibrium, pH, peptide, amyloid, charged surface

Abstrakt:

Tato práce se zabývá molekulovými simulacemi peptidů v acidobazické rovnováze. V první části jsme se zaměřili na 4 pentapeptidy. Pomocí simulace hrubozrnným modelem jsme určili stupeň ionizace jejich postranních řetězců, hodnoty jejich efektivních pK_a a Hillovy koeficienty. Výsledkem je, že náš CG model poskytuje pouze hrubý odhad hodnot pK_a. Výsledky závisí na zvolených parametrech modelu, jejichž vliv na konečnou hodnotu zatím není znám. Ve druhé části naší práce jsme studovali adsorpci peptidu Aβ-(1-42) na nabitý povrch, který pro naše účely představuje nabitou nanočástici. Úspěšná adsorpce by mohla být způsobem, jak zabránit agregaci peptidu Aβ-(1-42), který je podezřelý z toho, že způsobuje Alzheimerovu chorobu. Naše předběžné výsledky ukazují, že k adsorpci peptidu Aβ-(1-42) na nabitý povrch záporně nabité nanočástice dochází při nízkém pH, $1 < \mathrm{pH} < 4$.

Klíčová slova: kyselina, báze, acidobazická rovnováha, pH, peptid, amyloid, nabitý povrch

Contents

List of abbreviations

AA - all-atom model

BAM - *N*-*tert*-butylacrylamide

CG - coarse-grained model

cpH - constant-pH method

DHc - Debye–Hückel correction

HH - Henderson-Hasselbalch equation

LJ - Lennard-Jones potential

MC - Monte Carlo methods

MD - molecular dynamics

NIPAM - *N*-isopropylacrylamide

NPCE - non-permanently charged ends

PCE - permanently charged ends

RxMC - reaction ensemble MC method

1. Motivation

Peptides and proteins are abundant in all living organisms and perform many functions, such as biocatalysis, regulatory and protective functions. These substances are found in an aqueous environment and fulfill their function under specific conditions, defined by for example pressure, temperature, pH and salt concentration. The appropriate setting of these conditions can be used for many purposes, such as removal of waste substances from blood during dialysis, protein-mediated drug delivery to target sites in the body, or improvement of food quality. However, to achieve these goals, it is necessary to understand the effect of the above-mentioned quantities on the functions of a certain peptide. We focus on studying the acid-base behavior of molecules in equilibrium, specifically the degree of ionization of individual side chains in peptides depending on pH.

1.1 Historical overview of acid-base behavior studies

The first widely accepted acid-base theory was described by Arrhenius as early as 1884. This theory described an acid as a substance releasing the proton H^+ and a base as a substance releasing the hydroxide ion OH^- in an aqueous solution.

In 1923 a more accurate theory dealing with the acid-base behavior of simple acids and bases, the Brønsted theory[1], was published. This theory defines an acid as a substance that is able to donate an H^+ proton and a base as a substance that can accept this H^+ proton.

Around the same time period, in 1908, the work of Henderson, Hasselbalch and Sörenson resulted in the definition of the equation known today as the Henderson-Hasselbalch equation[2]. This formula relates 3 quantities - pH, pK_a and the ratio of the dissociated form of an acid to its undissociated form. The advantage of this approach is in its simplicity and it is therefore widely used in analytical chemistry. However, this formula can only be used for dilute solutions. In order to calculate the percentage of the ionized form of the acid in a certain solution, it is necessary to know, in addition to the measured pH value, also the dissociation constant pK_a. Nevertheless, the pK_a value is dependent on temperature, ionic strength and the dielectric constant of the solvent.

Debye attempted to describe the effect of ionic strength on the acid-base behavior of simple molecules in collaboration with Huckel, resulting in the formation of the Debye-Hückel theory[3] in 1923. Their theory results in several expressions for calculating the activity coefficient, such as the McInnes approximation or the Davies equation. Using this obtained coefficient, we can calculate the acid-base behavior of molecules even in solutions of higher concentrations.

However, these aforementioned analytical methods are no longer applicable for more complex systems, such as peptides or polyelectrolytes. Computer molecular simulations

were invented to investigate these systems. The biggest development of these methods came in the second half of the 20th century[4][5][6][7]. Main groups of computational methods include Monte Carlo[4][5] methods and Molecular dynamics[6][7].

Any molecular simulation needs a model of the given molecules of the system. The main categories of these models include atomistic and coarse-grained models.

In atomistic models, each particle represents one atom of a molecule. Simulations of atomistic models usually provide comparable results to experiment[8], but suffer from high computational and time requirements.

That is why the development of coarse-grained models began, where one particle of the model represents a certain group of atoms in a given molecule, for example one amino acid in a peptide can be represented by one particle. In principle, these rough models do not contain all the properties of the real molecule and therefore can provide a different result from the experiment. However, the advantage of these coarse-grained models is that we can tune the complexity of the model by turning various interactions "on and off" and observe the effect of individual simplifications on the result. In this way, we can obtain tools for theoretical prediction of acid-base behavior and gain acceptable results in a sufficiently short time.

One of the freely available softwares is pepKalc, which works on the principle of Gaussian-chain model for the disordered state and a hybrid mean-field approximation treatment[9]. The main advantage of this method is its speed. We have not yet verified the accuracy of this approach by comparing its results with other computational or experimental methods.

In our research group, led by Peter Košovan, we study the acid-base equilibrium of peptides and polymers using the coarse-grained model[10][11][12][13]. Our colleague Pablo Blanco is working on the development of the software pyMBE: the Python-based Molecule Builder for ESPResSo[14]. This tool is widely used to study acid-base behavior throughout our group[15].

1.2 This work

Overall, this work consists of two parts - *Short peptides* and *The Aβ-(1–42) peptide at a charged surface*. The Short peptides project is an extension of the topic of my bachelor's thesis, and the Aβ-(1–42) peptide at a charged surface is a new project where we use our coarse-grained model to study an even more complex system.

In the Short peptides project, we investigated the degree of ionization of the side chains in four pentapeptides using the coarse-grained model. The question was whether CG models can predict the degree of ionization of individual side chains in a peptide with accuracy comparable to atomistic models. Our hypothesis was that our model is able to predict the degree of ionization of given side chains with comparable accuracy to the AA model and that it performs this calculation in significantly less time. In order to test our

hypothesis, we have compared our data with the AA model data and also with the NMR experiment data from Dobrev et al.[8].

In our second project, we used the coarse-grained model to study an amyloid peptide, specifically the Aβ-(1–42) peptide. It is a 42 amino acid long molecule that is the main component of amyloid plaques found in patients suffering from Alzheimer's disease[16]. The plaques are aggregates of these peptides and may be the cause of this disease. However, it is not yet clear whether the aggregation of amyloid peptides is a consequence or a cause of Alzheimer's disease. How to effectively and safely prevent the aggregation of these peptides is now one of the main subjects of worldwide research.

Many studies have focused on the structural changes of amyloid peptides during aggregation[17][18]. Kirkitadze et al.[17] investigated the structural changes of several amyloid proteins during fibrillogenesis at pH 7.5 using circular dichroism. Their study shows that fibrils are rich in beta sheet structure, which is formed under a transient increase in alpha helix structure. They also found that fibrillation was significantly retarded in Aβ-(1–40) peptide compared to Aβ-(1–42) peptide. They further monitored these structural changes as a function of pH and found that the rapid acceleration of helical intermediate formation occurred between pH 3.86 and 6.0. They attributed this acceleration to the pH regime in which the β-carboxyl group of Asp and the imidazole ring of His are charged.

Some studies show that the presence of charged nanoparticles significantly inhibits the aggregation of amyloid peptides[19][20]. Cabaleiro-Lago et al.[20] investigated the effect of copolymer NiPAM:BAM nanoparticles on Aβ-(1–40) peptide fibrillation at pH 7.4. They found that the presence of these nanoparticles predominantly affects the nucleation phase of fibrillation and that the elongation phase is no longer affected by these nanoparticles. Analysis of their data suggests that binding of monomeric Aβ and prefibrillated oligomers to the nanoparticle prevent fibrillation. Two years later, Cabaleiro-Lago et al.[19] published a study of amino-modified polystyrene nanoparticles and their effect on Aβ-(1–40) protein fibrillation. These amino modified particles were 120 nm in size and were cationic polymer nanoparticles. The main finding of this study was that amyloid peptide fibrillation is directly dependent on the ratio of protein to nanoparticle concentration. At a constant peptide concentration (16 μM), fibrillation is accelerated at low nanoparticle concentration (0.05 mg/ml), whereas fibrillation is retarded at high nanoparticle concentration (1.1 mg/ml).

In contrast, Linse et al.[21] suggest that the nucleation of protein fibrillation may be enhanced by the presence of a nanoparticle. They argue that several layers of peptide on the surface of the nanoparticle, providing a local increase in protein concentration, can promote oligomer formation.

The combination of these studies suggests that adsorption of amyloid peptides onto nanoparticles does occur, but this process is very sensitive to individual conditions such as pH or nanoparticle concentration. We offer the possibility to study these conditions of peptide adsorption on nanoparticles by molecular simulations, specifically using a coarse-grained model. These methods are significantly cheaper and more flexible in terms of changing various properties of the system, such as the charge density on the nanoparticle or the length of the amyloid peptide. In this work, we focused on the effect of pH on the Aβ-(1–42) peptide adsorption onto the negatively charged surface.

2. Theoretical introduction

2.1 Acid-base behavior of a simple molecule

The term simple molecule here means a molecule with one ionizable group, i.e. a monoprotic acid or base.

According to the Brönsted theory[1], an acid is characterized by dissociating in an aqueous solution to form a proton H^+ and a charged acid molecule, conjugate base A^-

$$HA + H_2O \rightleftharpoons H_3O^+ + A^- \tag{2.1}$$

In contrast, in an aqueous solution a base as a neutral substance is able to bind a proton and this proton-binding base, conjugate acid BH^+, is charged. However, we can also describe acid-base behavior of the base using dissociation of the charged (proton-binding) form BH^+ as

$$BH^+ + H_2O \rightleftharpoons H_3O^+ + B \tag{2.2}$$

The difference between these two dissociation equations is, among other things, that in Equation 2.1, two charged molecules are formed from neutral ones and in Equation 2.2 the charge is simply transferred from one molecule to the other. Both these equations (2.1 and 2.2) are characterized by their own equilibrium constant, for an acid as $K(\text{acid})$

$$K(\text{acid}) = \frac{a_{(A^-)}a_{(H_3O^+)}}{a_{(HA)}a_{(H_2O)}} \tag{2.3}$$

and for a base $K(\text{base})$

$$K(\text{base}) = \frac{a_{(B)}a_{(H_3O^+)}}{a_{(BH^+)}a_{(H_2O)}} \tag{2.4}$$

which tells us something about the tendency of the acid or base to dissociate. We can also include water activity in these constants K and then we call these new constants the acidity constant $K_a(\text{acid})$ for an acid

$$K_a(\text{acid}) = \frac{a_{(A^-)}a_{(H_3O^+)}}{a_{(HA)}} \tag{2.5}$$

and acidity constant $K_a(\text{base})$ for a base.

$$K_a(\text{base}) = \frac{a_{(B)}a_{(H_3O^+)}}{a_{(BH^+)}} \tag{2.6}$$

All these constants are defined using activities a_i of given substances

$$a_i = e^{\frac{\mu_i - \mu_i^\ominus}{k_B T}} \tag{2.7}$$

that are defined using the difference between chemical potential of substance μ_i and chemical potential of its standard state $\mu_i^\ominus$.

The activity of individual substances can also be expressed using their concentration c_i relative to a certain standard concentration $c^{\ominus}$ and a quantity called the activity coefficient γ_i

$$a_i = \frac{c_i}{c^{\ominus}}\gamma_i = c_{(i,rel)}\gamma_i \tag{2.8}$$

The standard concentration $c^{\ominus}$ is here to ensure the dimensionlessness of the activity a_i and its value is $c^{\ominus} = 1$ mol/l. The activity coefficient γ_i represents non-ideal behavior of the a substance.

However, the activity coefficient of a given type of ions is experimentally unavailable, and thus we introduce the term mean activity coefficient $\gamma_{\pm}$

$$\gamma_{\pm,X_{\nu_+}Y_{\nu_-}} = \sqrt[\nu_+ + \nu_-]{\gamma_+^{\nu_+}\gamma_-^{\nu_-}} \tag{2.9}$$

where X represents a cation and Y an anion.

Another very important quantity is pH

$$\mathrm{pH} = -\log_{10} a_{(\mathrm{H_3O^+})} \tag{2.10}$$

which is defined as the negative logarithm of the activity of $\mathrm{H_3O^+}$ ions and pK_a

$$\mathrm{p}K_a = -\log_{10} K_a \tag{2.11}$$

which is a p-function of the dissociation constant K_a and it is characteristic for each acid (or base) in a particular molecule.

2.1.1 Ideal behavior: Henderson-Hasselbalch equation

Ideal behavior of substances is characterized by the fact that mutual interactions between molecules do not matter (the activity coefficient γ_i is equal to 1) and the overall acid-base behavior depends only on the concentrations of individual substances. Thus, for ideal behavior, the dissociation constant becomes the ratio of equilibrium concentrations of the dissociated components to the undissociated form of an acid

$$K_a(\mathrm{acid}) = \frac{[\mathrm{A^-}][\mathrm{H_3O^+}]}{[\mathrm{HA}]} \tag{2.12}$$

and of a base

$$K_a(\mathrm{base}) = \frac{[\mathrm{B}][\mathrm{H_3O^+}]}{[\mathrm{BH^+}]} \tag{2.13}$$

where brackets [ ] indicate the relative concentrations $c_{(i,rel)}$ of the given forms of individual substances.

The ratio of concentration of the charged form of a given substance to its total concentration (the sum of charged and uncharged forms) is called the degree of ionization α. For an acid

$$\alpha(\text{acid}) = \frac{[\text{A}^-]}{[\text{A}^-]+[\text{HA}]} \tag{2.14}$$

and for a base

$$\alpha(\text{base}) = \frac{[\text{B}^+]}{[\text{B}^+]+[\text{B}]} \tag{2.15}$$

From Equations 2.12 and 2.15, we can express the equation for pH for an acid

$$\text{pH}(\text{acid}) = \text{p}K_a + \log_{10}\frac{[\text{A}^-]}{[\text{HA}]} \tag{2.16}$$

and for a base.

$$\text{pH}(\text{base}) = \text{p}K_a + \log_{10}\frac{[\text{B}]}{[\text{BH}^+]} \tag{2.17}$$

We then can modify Equations 2.16 and 2.17 to express the degree of ionization for the acid α(acid)

$$\alpha(\text{acid}) = \frac{1}{1+10^{(\text{p}K_a-\text{pH})}} \tag{2.18}$$

and an analogous expression for the base.

$$\alpha(\text{base}) = \frac{1}{1+10^{-(\text{p}K_a-\text{pH})}} \tag{2.19}$$

These four equations (2.16-2.19) are forms of the Henderson-Hasselbalch equation[2].

Numerical value of pK_a is equal to pH when the ratio between the dissociated and non-dissociated form of the acid (or base) is 1:1.

In Figure 2.1 we can see the plot of the degree of ionization α depending on pH for an acid and base. For the acid, at pH = pK_a−1 the degree of ionization equals approximately 0.1 and at pH = pK_a−2 α(acid)$\doteq$ 0.01. This trend is centrally symmetric with respect to the pK_a value, so if pH = pK_a+1 then α(acid)$\doteq$0.9 and if pH = pK_a+2 then α(acid)$\doteq$0.99. All these approximate values for acid and base are shown in Table 2.1.

pH	α(acid)	α(base)
pKa - 2	0.01	0.99
pKa - 1	0.1	0.9
pKa	0.5	0.5
pKa + 1	0.9	0.1
pKa + 2	0.99	0.01

Table 2.1: Approximate values of degree of ionization for various pH

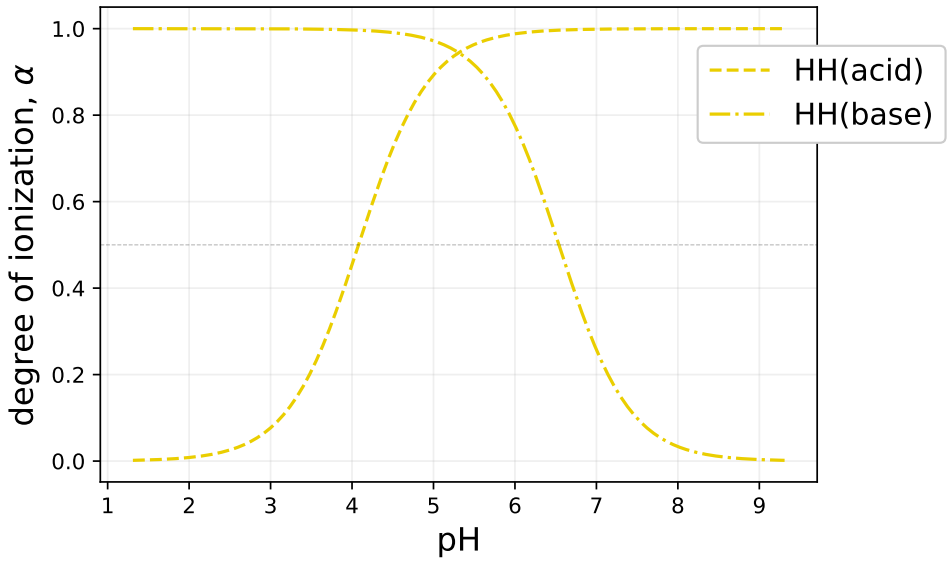


Figure 2.1: Degree of ionization as a function of pH for an acid and a base plotted using Henderson-Hasselbalch equation for an ideal molecule. $\mathrm{p}K_a(\text{acid}) = 4.08$, $\mathrm{p}K_a(\text{base}) = 6.54$

2.1.2 Added salt: the Debye–Hückel correction

So far we have not considered any salt in the solution at all. However, when we add salt to an acid (or a base) solution, the acid-base behavior changes significantly. Such a change can be described using the p-function of the mean activity coefficient $\gamma_\pm$

$$\mathrm{p}\gamma_\pm \equiv -\log_{10}\gamma_\pm = \frac{A\sqrt{I}}{1 + Ba\sqrt{I}} \tag{2.20}$$

which depends on the ionic strength I of the solution

$$I = \frac{1}{2}\sum_i c_i {z_i}^2 \tag{2.21}$$

where A, B and a are are coefficients resulting from Debye-Hückel theory[3]. Their values are in Table 2.2.

For a monovalent salt such as NaCl, the concentration c_{salt} is equal to the ionic strength I of the solution. Equation 2.20 is part of Debye–Hückel theory[3] and it is called the Debye–Hückel expression for the mean activity coefficient and for which we use the abbreviation DHc.

Table 2.2: Values of DHc coefficients [22]

Constant	A [$\mathrm{mol}^{-1/2}$ $\mathrm{dm}^{3/2}$]	B [nm^{-1} $\mathrm{mol}^{-1/2}$ $\mathrm{dm}^{3/2}$]	a [nm]
Value	0.5085	0.3281	0.9

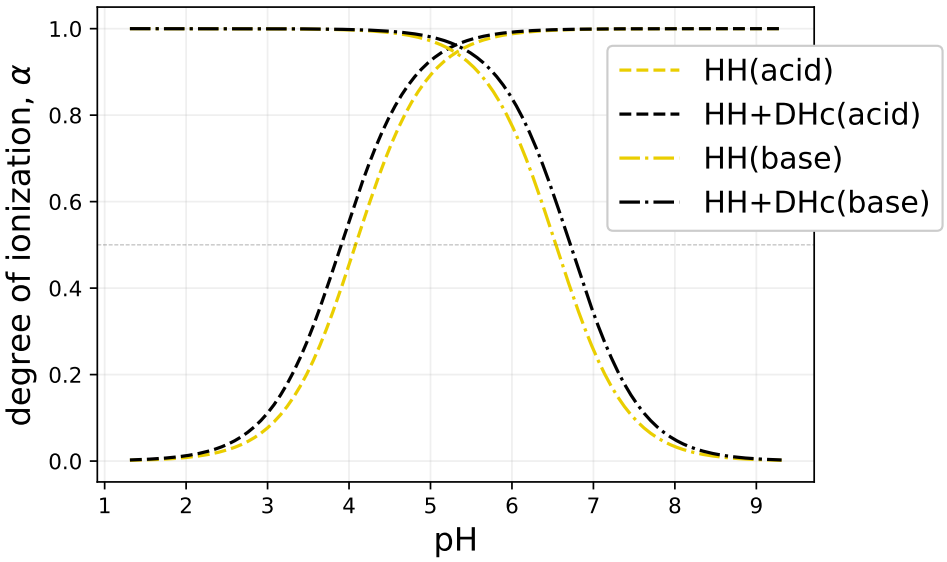


Figure 2.2: Degree of ionization as a function of pH for an acid and a base plotted using Henderson-Hasselbalch equation for an ideal molecule and Henderson-Hasselbalch equation corrected using DHc, $\mathrm{p}K_a(\mathrm{eff}) = \mathrm{p}K_a(\mathrm{HH}) - \mathrm{p}\gamma$ for an acid, $\mathrm{p}K_a(\mathrm{eff}) = \mathrm{p}K_a(\mathrm{HH}) + \mathrm{p}\gamma$ for a base

By implementing the DHc into the Henderson-Hasselbalch equation, we obtain a modified expression for the degree of ionization of the acid $\alpha(\mathrm{acid})$

$$\alpha(\mathrm{acid}) = \frac{1}{1 + 10^{(\mathrm{p}K_a - \mathrm{pH} - \mathrm{p}\gamma)}} \tag{2.22}$$

and for the base $\alpha(\mathrm{base})$.

$$\alpha(\mathrm{base}) = \frac{1}{1 + 10^{-(\mathrm{p}K_a - \mathrm{pH} + \mathrm{p}\gamma)}} \tag{2.23}$$

The change of the acid-base behavior is shown graphically in Figure 2.2.

2.2 More complex molecules: Charge regulation

The acid-base behavior of a real molecule is not only affected by the added salt but also by interactions between individual ionizable groups. The interaction is influenced, among other things, by the distance between these ionizable groups and if this distance is defined, for example, by bonds between individual ionizable groups, then their resulting acid-base behavior strongly deviates from ideal behavior. We can see an example of this deviation in Figure 2.3 on the HVD molecule (peptide Ac-NH-histidine-valine-aspartic acid-$\mathrm{CONH_2}$). There are two ionizable groups of opposite polarity, quite spacialy close to each other and for that reason $\mathrm{p}K_a$ of histidine is significantly shifted compared to $\mathrm{p}K_a$ of an isolated molecule so its degree of ionization is increased[23].

This shift in the $\mathrm{p}K_a$ value can be described by quantity called $\Delta\mathrm{p}K_a$

$$\Delta \mathrm{p}K_a = \mathrm{p}K_a(\mathrm{eff}) - \mathrm{p}K_a(\mathrm{ref}) \tag{2.24}$$

which is the difference between the effective value of pK_a, for example pK_a value of histidine in the HVD molecule, and some reference value, for example pK_a value of histidine under ideal conditions.

We then can also monitor the difference in ΔpK_a obtained by various methods, for example simulation and experiment. This variable is called $\Delta\Delta$pK_a.

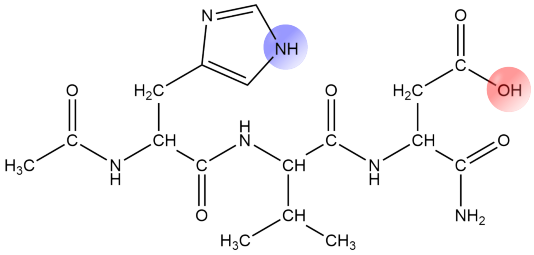


Figure 2.3: HVD molecule

$$\Delta\Delta \mathrm{p}K_a = \Delta \mathrm{p}K_a(\mathrm{simulation}) - \Delta \mathrm{p}K_a(\mathrm{experiment}) \tag{2.25}$$

2.2.1 Hill equation

However, for more complex molecules, not only the pK_a value but also the slope of the titration curve (Figure 2.1) may change. This means that the degree of ionization is affected in different ways, depending on the pH value. The slope of the titration curve is characterized by the Hill coefficient n in the Hill equation, for an acid

$$\alpha(\mathrm{acid}) = \frac{1}{1 + 10^{n(\mathrm{p}K_a - \mathrm{pH})}} \tag{2.26}$$

and for a base.

$$\alpha(\mathrm{base}) = \frac{1}{1 + 10^{-n(\mathrm{p}K_a - \mathrm{pH})}} \tag{2.27}$$

2.3 Acid-base behavior of peptides

A peptide is a molecule that consists of individual amino acids linked by a peptide bond. An oligopeptide consists of $\approx$ 2-10 amino acids, a polypeptide contains $\approx$ 11-100 amino acids, and if the chain is longer than 100 amino acids, we call it a polypeptide (or protein if it is produced by living organisms).

There are 21 proteinogenic α-amino acids, 20 of which contain a side chain. The term α-amino acid means that the amino group -NH_2 is attached to the same carbon (alpha carbon C_α) as the attached -COOH group. Several of these 20 side chains of individual amino acids are ionizable (they contain a basic or acidic group and undergo chemical reactions - Equation 2.1 or 2.2). All amino-acids used in this work are listed in Table 2.3.

When we connect individual amino acids into a peptide chain, we must in principle have a free N-terminus at one end of the chain and a C-terminus at the other end (if we are not considering cyclic peptides). These ends are also ionizable. In our work, we explicitly write these ionizable ends in the sequences of studied molecules using abbreviations "n"

Table 2.3: All amino acids, n-terminus and c-terminus used in this work. The one-letter abbreviation of acidic amino acids is highlighted in red and basic amino acids are highlighted in blue (this colored marking is also used for sequences of studied peptides in this work).

Amino acid	Abbreviation		Side chain	Group
Alanine	Ala	A	neutral	methyl group
Arginine	Arg	**R**	basic	
Asparagine	Asn	N	neutral	
Aspartic acid	Asp	**D**	acidic	carboxyl group
Glutamic acid	Glu	**E**	acidic	carboxyl group
Glutamine	Gln	Q	neutral	
Glycine	Gly	G	—	—
Histidine	His	**H**	basic	imidazole
Isoleucine	Ile	I	neutral	
Leucine	Leu	L	neutral	
Lysine	Lys	**K**	basic	
Methionine	Met	M	neutral	
Phenylalanine	Phe	F	neutral	benzyl group
Serine	Ser	S	neutral	hydroxyl group
Tyrosine	Tyr	**Y**	acidic	4-OH-benzyl group
Valine	Val	V	neutral	isopropyl group
N-terminus	—	**n**	—	—
C-terminus	—	**c**	—	—

and "c". It should be noted that the abbreviations "n" and "c" in peptide sequences are not standard abbreviations. However, we write them down because they also undergo acid-base reactions (Equation 2.1 and 2.2) and thus contribute to the charge regulation.

In general, we can say that peptides and proteins are substances where charge regulation plays an important role, and the acid-base behavior of specific amino acids in the chain is strongly influenced by its immediate surroundings.

2.4 Radius of gyration R_g

One of the quantities describing the effective size of a peptide is a the radius of gyration R_g, that for a chain composed of N monomer units of the same mass m_i can be calculated as

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\vec{r_i} - \vec{r_{cm}}|^2} \tag{2.28}$$

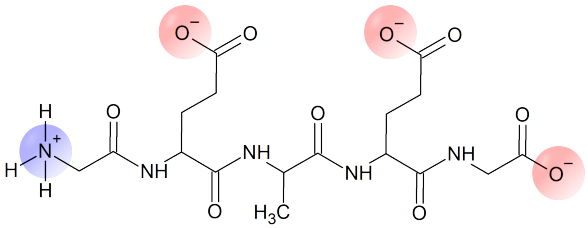
where r_i is the position vector of a given monomer unit and r_{cm} is the position vector of the chain center of mass.

3. Systems of our interest

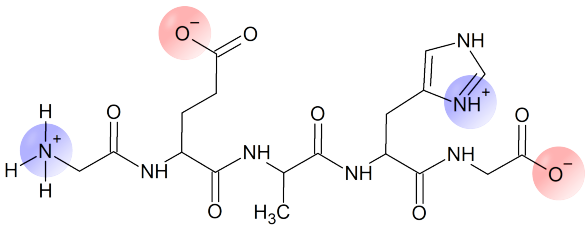
3.1 Short peptides

In our first project we focused specifically on 4 oligopeptides - **n**G**E**A**E**G**c**, **n**G**E**A**H**G**c**, **n**G**H**A**E**G**c**, **n**G**H**A**H**G**c** (Figure 3.1). There are two ionizable side chains in each sequence, and we work with all their possible combinations: both acidic, both basic, acidic at the n-terminus with basic at the c-terminus and basic at the n-terminus with an acid at the c-terminus.

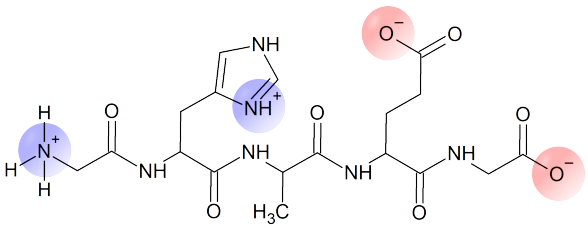
For these olipeptides, we focused on values of the degree of ionization for individual pH values, on effective pKa values resulting from titration curves and on the overall changes in titration curves with respect to ideal behavior (shift of titration curve and the change of its slope).



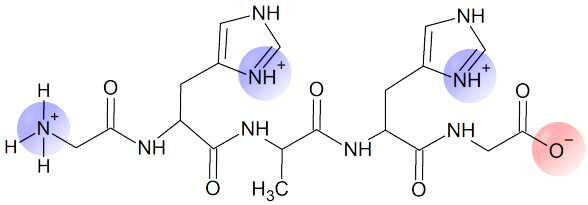
(a) **n**G**E**A**E**G**c** sequence



(b) **n**G**E**A**H**G**c** sequence



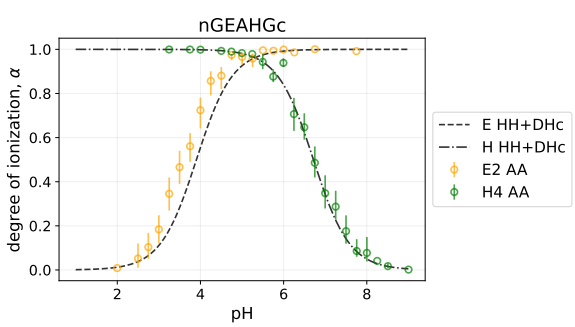
(c) **n**G**H**A**E**G**c** sequence



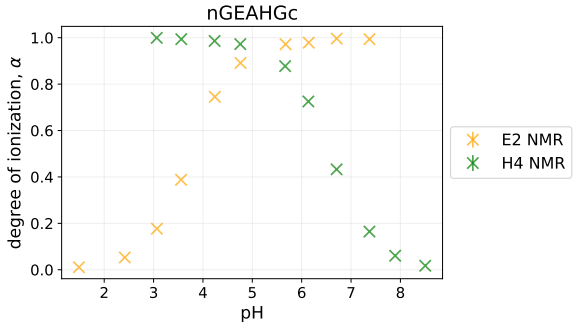
(d) **n**G**H**A**H**G**c** sequence

Figure 3.1: Peptide sequences investigated in the first part of our work

In Dobrev et al.[8], they present data for all-atom (AA) simulation and NMR experiment for all of the four sequences. In Figure 3.2a, we can see example data of the AA simulation and Figure 3.2b shows example data for the NMR experiment. Data for other sequences can be found in the Appendix (Figure 7.1-7.8).



(a) All-atom (AA) data for **n**G**E**A**H**G**c**



(b) NMR data for **n**G**E**A**H**G**c**

Figure 3.2: Degree of ionization as a function of pH for nGEAHGc peptide from Dobrev et al.[8]. E2 represents glutamic acid at position 2 in the sequence and H4 histidine at position 4. The black dashed line without dots represents the ideal behavior for glutamic acid described by the Henderson-Hasselbalch equation corrected by the Debye-Huckel correction (E HH+DHc) and the black dashed line with dots represents this ideal behavior for histidine (H HH+DHc).

3.2 The Aβ-(1–42) peptide at a charged surface

In our second project, we studied the acid-base behavior of the Aβ-(1–42) peptide in the presence of a charged surface, which is supposed to represent nanoparticles with a charged surface. We chose a charged surface (plate) instead of a charged spherical particle due to the large radius of a nanoparticle relative to the size of the peptide. Blanco et al.[24] claim that R_g of the Aβ-(1–42) peptide is around 2 nm and Cabaleiro-Lago et al.[19] studied the adsorption of amyloid peptide on charged nanoparticles with a size of 120 nm. The ratio of particle sizes can be seen in Figure 3.3. Given these data, we decided to simulate our system as a peptide in the presence of a charged surface.

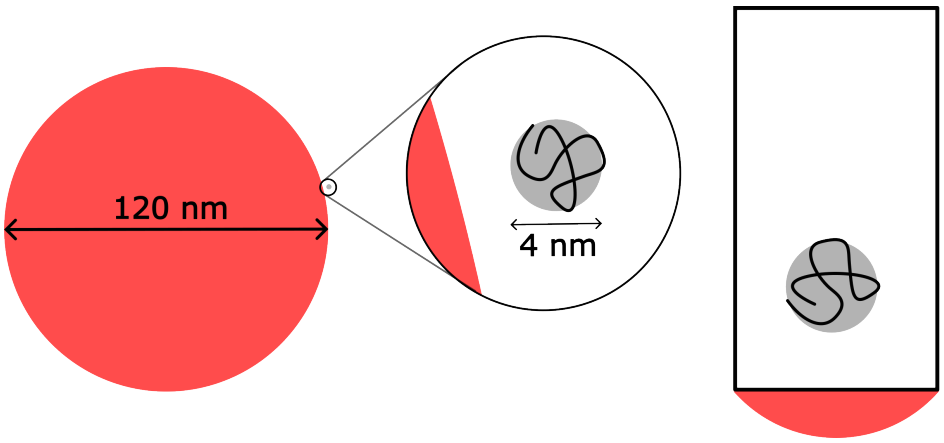


Figure 3.3: Scheme of a nanoparticle and a Aβ-(1–42) peptide. Red bead represents a nanoparticle, gray bead the effective size of the peptide (sphere with a radius of R_g) and the black tangled line represents a Aβ-(1–42) peptide

From a number of amyloid peptide alloforms, we chose the Aβ-(1-42) peptide, which is the major form of the amyloid peptides found in the plaques[16].

The sequence of the Aβ-(1–42) peptide is:
nDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIAc

where acidic ionizable groups are highlighted in red and basic ionizable groups in blue. Overall, this molecule contains 7 acidic and 6 basic side chains plus an acidic c-terminus and a basic n-terminus so there are up to 8 negative and 7 positive charges per molecule. The scheme of the Aβ-(1-42) peptide can be seen in Figure 3.4.

It should be noted that the Aβ-(1–42) peptide already contains elements of secondary structure - in an apolar microenvironment there are two helical regions connected by a flexible kink[16], in water it adopts a collapsed coil structure[25] and in the form of aggregates (fibrils) they take a β-sheet conformation[17].

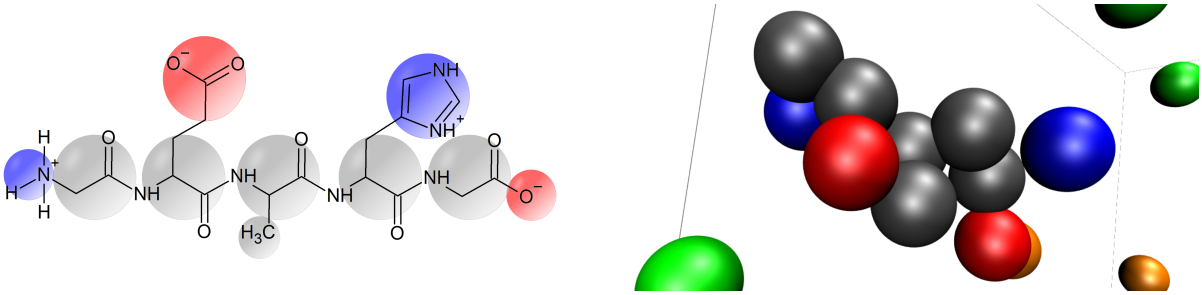
Nevertheless, we have not (yet) included any of these secondary structure elements in our coarse-grained model. But in the future we would like to include them in our model by using additional parameters of the Martini force field mimicking the secondary structure[26].

Figure 3.4: Scheme of Aβ-(1–42) peptide.

4. Our approach - molecular simulation

4.1 Simulation model of peptides

In our simulation we use the coarse-grained model, specifically the two-bead model[12]. Molecules represented by the coarse-grained model are generally only an approximation of the real molecules. Individual parts containing multiple atoms are replaced by individual beads. In our case, each amino acid is represented by two beads, therefore the two-bead model. One bead is for the backbone of the chain and the second one is for the side chain. If an amino acid has no side chain, it is represented by only one bead. In our work this is true for glycine. The two-bead model can be seen in Figure 4.1.



(a) Our CG model of nGEAHGc peptide (b) Simulation snapshot of nGEAHGc peptide

Figure 4.1: Our coarse-grained model of a short peptide

This two-bead model is characterized by several parameters as well as interactions between individual beads.

The most intuitive parameter is the size of the beads and the distances between them, bond lengths. The bead diameter was 0.35 nm for all beads. Some of the bond lengths were obtained using the all-atom (AA) simulation by Lunkad et al.[27], bonds: C_α-C_α, C_α-Glu, C_α-His. The rest of the bond lengths used in the Short peptides were determined using the Avogadro software[28], bonds: C_α-n-terminus, C_α-c-terminus, C_α-Ala. Other bond lengths in the Aβ-(1–42) peptide, not determined by AA simulation from Lunkad et al.[27] or Avogadro software[28], had a default bond length of 0.4 nm. These bond lengths are the distances between centers of individual beads and they are shown in Table 4.1.

A bonding interaction between individual beads is represented by a harmonic potential

$$U_h(r) = \frac{k_h}{2}(r - r_0)^2 \tag{4.1}$$

where k_h is the stiffness constant, r_0 is the equilibrium distance between two beads connected by a chemical bond (bond lengths in Table 4.1) and r is the current distance between these two beads.

Table 4.1: Bond lengths for the two-bead model

Bond	Bond length [nm]	Data source
C_α-C_α	0.382	Lunkad et al.[27]
C_α-Glu	0.329	Lunkad et al.[27]
C_α-His	0.452	Lunkad et al.[27]
C_α-Ala	0.153	Avogadro software[28]
C_α-n-terminus	0.146	Avogadro software[28]
C_α-c-terminus	0.246	Avogadro software[28]
default bond	0.400	—

A non-bonding interaction between all particles in a simulation is represented by the short-range Lennard-Jones interaction

$$U_{LJ}(r) = 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right) \tag{4.2}$$

where ϵ represents the depth of the potential well, which is a measure of interaction strength of two particular particles. σ represents a distance at which the potential between particles is zero, it is a distance at which two particles can approach each other, and r is the current distance between two particles. The parameter σ is the size of a bead so it is 0.35 nm for all particles in the simulation and ϵ was set to the value of $k_B T$ for all particles.

A non-bonding interaction between all charges in a simulation is represented by the long-range Coulomb interaction

$$U_{coul}(r) = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_i q_j}{r_{ij}} \tag{4.3}$$

where ϵ_0 represents vacuum permittivity, ϵ_r relative permittivity of a solvent (in our case water), q_i and q_j represent to particular charges and r_{ij} is the distance between them.

For molecular simulations we have other technical parameters that we can change. One of them is the cutoff for various potentials. We introduce this cutoff due to computational complexity of the simulation. We calculate all interactions below this cutoff value in real space and all interactions above this value (beyond the cutoff limit) are calculated in reciprocal space. This approach is called the Ewald summation[29].

4.2 Simulation methods

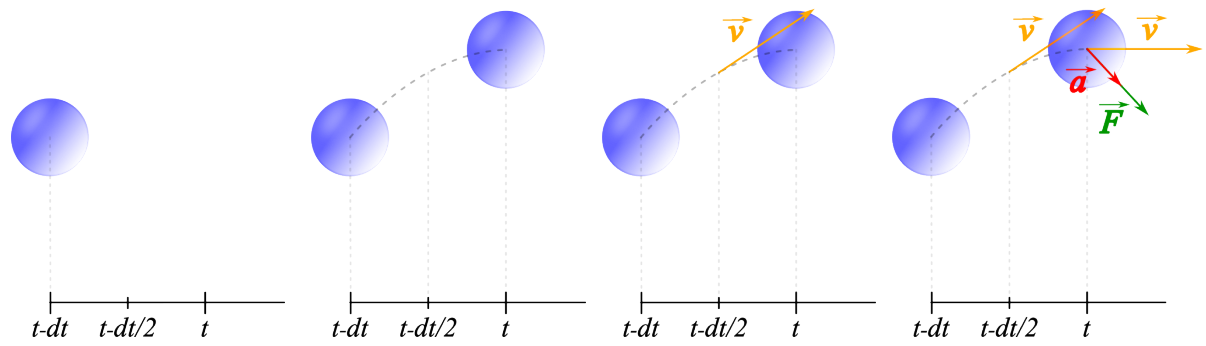
Simulation methods can be divided into several categories. Among the most important are Molecular dynamics and Monte Carlo methods. Molecular dynamics is a method of statistical sampling of the phase space where we use integration of motion equations. A particle in a phase space is defined by its position and momentum. Unlike molecular dynamics, the Monte Carlo method is used to statistically sample the configuration space

by randomly generating configurations of a given system. Particles in this configuration space are defined only by their position.

4.2.1 Motion in a simulation: The velocity Verlet method

One of the most widely used methods to track temporal evolution of particle positions is the velocity Verlet method.

The whole process of tracking temporal evolution of particle positions begins with generation of particles in completely random positions (at $t = t - dt$, Figure 4.2a). Then, using an integrator other than the velocity Verlet we generate a position of the particle at time t (Figure 4.2b). Further, using Equation 4.5, we calculate the velocity in half of the time step dt as the difference of individual positions divided by the time step dt (Figure 4.2c). This step is just an approximation, but for a sufficiently small value of the time step dt, there is almost no deviation. In the last step, we calculate the force at a given time t (according to Equation 4.6). By dividing the force by the mass of the particle we obtain its acceleration a and finally we calculate the velocity at time t (according to Equation 4.7).



(a) Random position and velocity (b) First integrator (not velocity Verlet) (c) Velocity at $t - dt/2$ (d) Velocity and force at t

Figure 4.2: Process before the velocity Verlet

After the process in Figure 4.2 we are at the starting point for the velocity Verlet cycle. It starts by calculating the velocity at half of the time step dt, at $t + dt$, $v(t + dt/2)$ using

$$v(t + dt/2) = v(t) + \frac{F(x(t), v(t - dt/2), t)}{m} dt/2 \tag{4.4}$$

shown in Figure 4.3a. Then we calculate the new position of the particle at time $t + dt$, $x(t + dt)$ using

$$x(t + dt) = x(t) + v(t + dt/2)dt \tag{4.5}$$

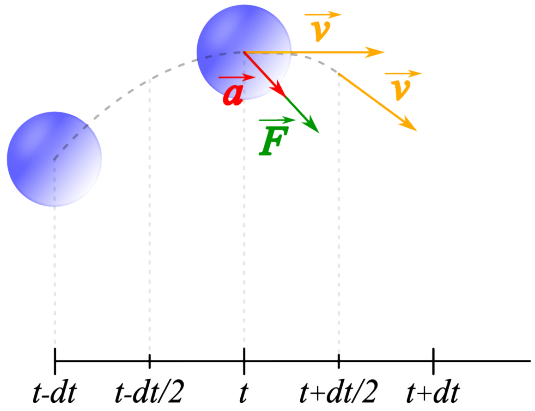
shown in Figure 4.3b, the total force acting on the given particle F

$$F = F(x(t + dt), v(t + dt/2), t + dt) \tag{4.6}$$

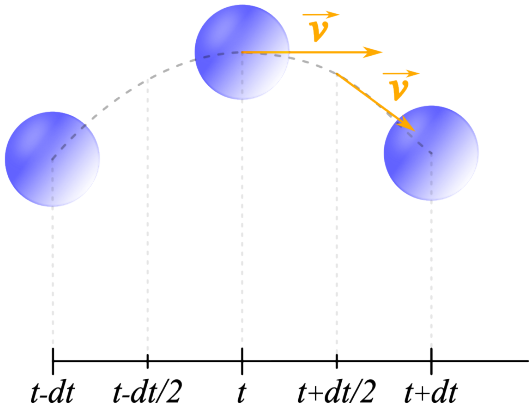
shown in Figure 4.3c and the new velocity at time $t + dt$, $v(t + dt)$

$$v(t + dt) = v(t + dt/2) + \frac{F(x(t + dt), t + dt)}{m} dt/2 \tag{4.7}$$

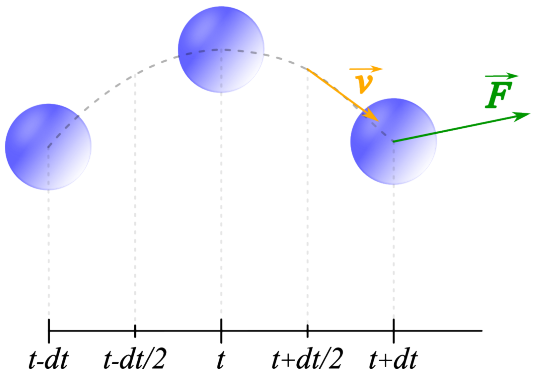
shown in Figure 4.3d. At the end of this cycle we are back in the starting point we were in in Figure 4.2d and the cycle can start again. A schematic of the pre-velocity Verlet process along with two velocity Verlet cycles is shown in Figure 4.4 (pink area represents pre-velocity Verlet process, green area is for the first cycle of the velocity Verlet and the orange one is for its second cycle).



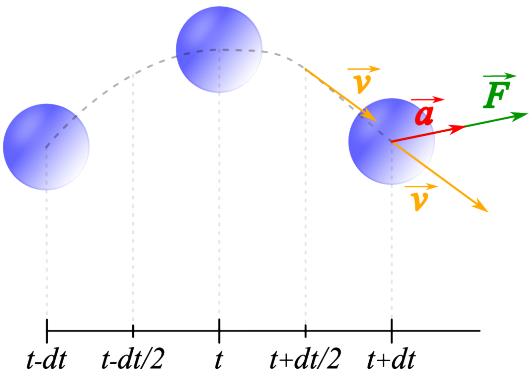
(a) Step 1



(b) Step 2



(c) Step 3



(d) Step 4

Figure 4.3: The velocity Verlet cycle

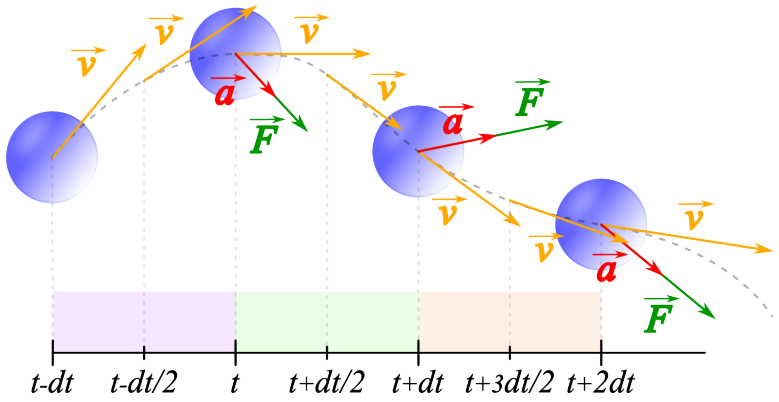


Figure 4.4: The velocity Verlet method

4.2.2 Constant temperature: The Langevin thermostat

The Langevin thermostat is defined using the Langevin equation which is a differential equation describing the Brownian motion of particles in a liquid and several other physical systems[30].

At first, we can define velocity $\vec{v}(t)$ using angular momentum $\vec{p}$ as

$$\frac{d\vec{r}}{dt} = \vec{v}(t) = \frac{\vec{p}}{m} \tag{4.8}$$

and force $\vec{F}$ using acceleration $\vec{a}$ or using the change in angular momentum $\vec{p}$ over time t as

$$\vec{F} = m\vec{a} = m\frac{d\vec{v}}{dt} \overset{cont.m}{=} \frac{d(m\vec{v})}{dt} = \frac{d\vec{p}}{dt} \tag{4.9}$$

where m is a mass of a particle.

Then the total force $\vec{F}$ acting on the particle

$$\vec{F} = \frac{d\vec{p_i}}{dt} = f_i(\vec{x_i}, \vec{v_i}, t) - \gamma m\vec{v_i}(t) + \sqrt{2\gamma m k_B T}\frac{d\vec{w}}{dt} \tag{4.10}$$

is calculated as the sum of the deterministic force calculated from the interactions $f_i(\vec{x_i}, \vec{v_i}, t)$, a negative term containing the friction coefficient γ and a term representing the thermal force containing the temperature T. The symbol k_B represents the Boltzmann constant and $\vec{w}$ is a (3N-dimensional) Wiener process[30].

To maintain a constant temperature T we have to add random forces mimicking collisions with solvent particles at a given temperature (the thermal force term in Equation 4.10), but at the same time we have to add friction forces (the negative term with γ in the same equation) that will partially compensate the thermal forces. Nevertheless, Langevin dynamics is already an advanced method and a more detailed description is beyond the scope of this paper.

4.2.3 Reactions in a simulation

4.2.3.1 The constant-pH method (cpH)

The constant-pH method (cpH) is a Monte Carlo method designed to study acid-base behavior at a given pH.

In this method, during the sampling of the system, a randomly selected ionizable group is changed according to the acid-base reaction, for an acid

$$\mathrm{HA} \rightleftharpoons \mathrm{A}^- + \mathrm{X}^+ \tag{4.11}$$

or for a base.

$$\mathrm{BH}^+ \rightleftharpoons \mathrm{B} + \mathrm{X}^+ \tag{4.12}$$

These equations basically represent Equations 2.1 and 2.2 with a few differences. First of all, in our simulation we do not have an explicit solvent (individual water particles) and thus we do not have a given solvent in the equations. And secondly, there is not directly a hydrogen ion H^+, but a general counterion X^+. Generally, this is because in a real system a hydrogen ion can react with a basic buffer to form a BH^+ ion and the electroneutrality of the system must be preserved in our simulation. We can choose an arbitrary counterion that will be formed during the acid-base reactions (Equations 4.11 and 4.12). For our simulation we chose the sodium ion Na^+.

When we perform the change of a randomly selected ionizable group (Equations 4.11 or 4.12), we calculate the probability of acceptance of this reaction P_{cpH}.

$$P_{\mathrm{cpH}} = \min(1, e^{-\beta\Delta U \pm \ln(10)(\mathrm{pH}-\mathrm{p}K_a)}) \tag{4.13}$$

4.2.3.2 The reaction ensemble MC method (RxMC)

This method ensures a constant chemical potential of the salt particles throughout the simulation box and thus the desired salt concentration in a bulk. The chemical potential of salt particles is directly related to the equilibrium cosntant K_{NaCl}

$$K_{\mathrm{NaCl}} = e^{\frac{\mu_{\mathrm{Na}^+}+\mu_{\mathrm{Cl}^-}}{k_B T}} \tag{4.14}$$

We can think of this approach as if our simulation box were connected to an infinitely large reservoir filled with salt particles with a constant chemical potential of these particles, and ion pairs were exchanged between this reservoir and our simulation box.

$$\varnothing \rightleftharpoons \mathrm{Na}^+ + \mathrm{Cl}^- \tag{4.15}$$

The Equation 4.15 is characterized by its equilibrium constant K_{NaCl}

$$K_{\mathrm{NaCl}} = (a_{\mathrm{Na}^+})(a_{\mathrm{Cl}^-}) = \left(\frac{c_{\mathrm{Na}^+}}{c^\ominus}\gamma_{\mathrm{Na}^+}\right)\left(\frac{c_{\mathrm{Cl}^-}}{c^\ominus}\gamma_{\mathrm{Cl}^-}\right) = \left(\frac{c_{\mathrm{salt}}}{c^\ominus}\gamma_\pm\right)^2 \tag{4.16}$$

where c_{salt} is the desired salt concentration in bulk, $c^\ominus$ is the standard concentration (1 mol/l) and $\gamma_\pm$ is the dimensionless mean activity coefficient. The input parameter of

this method is the chemical potential of the salt ion pair, which we try to estimate using the desired salt concentration and the mean activity coefficient $\gamma_{\pm}$. The coefficient can be calculated using Equation 2.20.

4.3 Data analysis

The direct output of the simulation is the time evolution of a given quantity. The quantity can be, for example, the total energy of the system and the kinetic energy. However, it can also be a charge of a side chain of a particular amino acid in a specific molecule. Then, we can simply average the values from this time series and thus obtain the sample mean $\overline{X}$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{n\to\infty}{=} \langle X \rangle \tag{4.17}$$

where n is the number of values we average (number of samples).

The value of the sample mean $\overline{X}$ approaches the true mean $\langle X \rangle$ when the number of our samples approaches infinity.

However, how do we determine that we have a sufficient number of samples and that our value of the sample mean well represents the true mean of the given quantity? For statistical data processing, there are several procedures for approaching the true mean of a given quantity and for determining its accuracy, standard deviation σ_X^2.

$$\sigma_X^2 = \langle (\overline{X} - \langle X \rangle)^2 \rangle \tag{4.18}$$

Since we start from a completely random configuration, our entire simulation can be divided into two parts: the equilibration and the production part. The equilibration is discarded and not included in further statistical data processing.

However, we also need to visually check that the production part graphically looks like random noise and that there are no long-term fluctuations in the order of 10 % of the entire production part. If there are long-term fluctuations, we need to run a longer simulation.

Furthermore, we must also take into account the fact that our data are correlated, they depend on each other. When we let the system evolve over time, we observe for some time that the new configuration depends on the configuration of the previous one. In the plot of time evolution of a particular quantity we observe a certain increasing or decreasing trend. If we had uncorrelated data (they would not depend on each other), we could calculate the standard deviation $\sigma_{X_i}^2$

$$\sigma_{X_i}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} \tag{4.19}$$

Nevertheless, the simulation data are mostly correlated and therefore the calculation needs to be adjusted accordingly. We can modify Equation 4.19 to

$$\sigma_{X_i}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}(1+2\tau) = \frac{\sum_{i=1}^{n}\Delta {X_i}^2}{n}(1+2\tau) \tag{4.20}$$

where τ is the autocorrelation time. It is the time scale on which the system loses memory of its previous state.

Therefore, for the calculation of standard deviation of the simulation (correlated) data, we need to know the estimate of the autocorrelation time τ. For this purpose we can use the block method.

At first, we calculate the sample mean $\overline{X}$ of all data from the production part using Equation 4.17 and calculate the standard deviation σ_X^2 using Equation 4.19 (using the procedure as if our data were uncorrelated). Then we divide the data into several blocks, b, (we chose b = 16) where each block contains k samples and the total number of samples is n (then $n = bk$). We calculate the arithmetic mean of each block $\overline{X}_j$ (Equation 4.17) and the standard deviation σ_B^2 of these means $\overline{X}_j$ from $\overline{X}$.

$$\sigma_B^2 = \frac{\sum_{i=1}^{b}(\overline{X}_j - \overline{X})^2}{b} \tag{4.21}$$

The error of the given quantity ϵ is calculated as

$$\epsilon = \sqrt{\frac{\sigma_B^2}{b}} \tag{4.22}$$

Finally, we calculate the autocorrelation time τ_{corr}

$$\tau_{\text{corr}} = \Delta t \frac{k}{2}\frac{\sigma_B^2}{\sigma_{X_i}^2} \tag{4.23}$$

and the number of effective samples n_{eff}.

$$n_{\text{eff}} = \frac{n}{2\tau_{\text{corr}}} \tag{4.24}$$

We chose an arbitrary criterion to have at least 5 effective samples per block, $n_{\text{eff}} \geq 5b$. When this condition is met, the blocks are mutually uncorrelated and the result of our simulation can be written as $result = \overline{X} \pm \epsilon$.

5. Simulation protocol

The simulation itself was performed using the ESPResSo software (Extensible Simulation Package for Research on Soft Matter)[31]. For Short peptides we used ESPResSo version 4.2.0 and for the Aβ-(1–42) peptide version 4.2.1. The peptide model was built using the software pyMBE: the Python-based Molecule Builder for ESPResSo[14].

5.1 Short peptides

This simulation was performed in a cubic simulation box. The contents of the box consisted of one peptide molecule, explicit salt particles (NaCl) and an implicit solvent that was set there (the solvent was not represented by particles, but only by the solvent permittivity value). The peptide was represented by the two-bead model. During this simulation, we used the velocity Verlet method for calculating particle motion, the Langevin thermostat to maintain a constant temperature, and the constant-pH method.

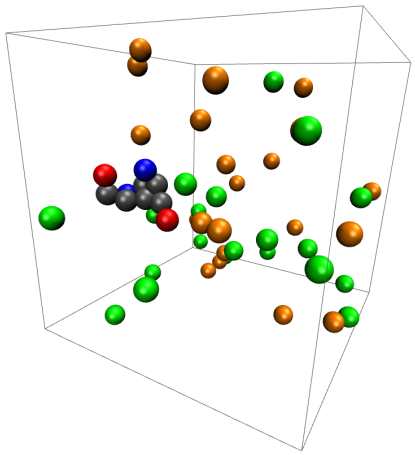


Figure 5.1: Simulation box with pentapeptide nGEAHGc and salt particles. Dark gray beads represent non-ionizable groups, blue ones are for basic ionizable groups, red ones are for acidic ionizable groups, orange ones are for sodium ions Na^+ and green ones are for chloride ions Cl^-.

All parameters mentioned in Section 4.1, peptide and salt concentrations used in our simulation and approaches from Dobrev et al.[8] are summarized in Table 5.1.

Table 5.1: Parameters of our simulations, compared to the simulations and NMR experiment from Dobrev et al.[8]

Parameter	Our work CG model	Dobrev et al.[8] AA model	Dobrev et al.[8] NMR
k_h [kJ mol^{-1} nm^{-2}]	992	1000	—
ϵ [$k_B T$]	1	—	—
σ [nm]	0.35	—	—
T [K]	298.15	300	300
$c_{peptide}$ [mol l^{-1}]	0.00769	0.00769	0.01
c_{salt} [mol l^{-1}]	0.15	0.15	0.15

5.2 The Aβ-(1–42) peptide at a charged surface

Unlike the previous section 5.1, where we performed the simulation in a cubic simulation box, this time we used a different shape of the box. We inserted a square base of our simulation box to the xy plane at $z = 0$, where we placed a grid of charges representing the surface of the nanoparticle. These charges had a fixed position and a constant charge. We chose the length of the simulation box in the z-axis L_z to be 4 times longer than the length in the x- (or y-) axis ($4L_x = 4L_y = L_z$). We used this setting in order to minimize the surface of the base (xy plane) and thus the number of charges on the surface due to lower computational demands (in general, fewer charges in the simulation box mean lower computational complexity). However, in order to be able to simulate adsorption on the one hand and the peptide in the bulk on the other hand, we had to extend the system in the z-axis as already mentioned.

Even before we put anything in the simulation box, we have to somehow ensure that the mirror images in the z-axis do not interact with each other; to really ensure "space nearby the nanoparticle" and "space far from the nanoparticle" (in other words bulk). For this purpose we chose the Electrostatic Layer Correction (ELC) ESPResSo functionality. However, ELC is already an advanced method and a detailed description of its principle is beyond the scope of this paper[32][33]. What ELC does in practice is it inserts an extra layer in the xy plane that cancels the electrostatic contribution of the periodic images in the z-direction. Thus it converts the 3D method to a 2D method (we only consider periodic images in the x and y direction). This layer is above the L_z (so the actual dimensions of the simulation box even with ELC gap are [L_x, L_y, (L_z+ELC gap)]).

When we have a simulation box prepared in this way, we can insert one peptide molecule, explicit salt particles (NaCl) and an implicit solvent into the simulation box (the solvent is not represented by particles, but only by the solvent permittivity value). The peptide was represented by the two-bead model. During this simulation, we used the velocity Verlet method for calculating particle motion, the Langevin thermostat to maintain a constant temperature, the constant-pH method and the reaction ensemble MC method to ensure a constant chemical potential of the salt in all parts of the simulation box. All the parameters are in Table 5.2.

Table 5.2: Parameters of our simulations, compared to the simulations from Blanco et al.[24]. Blanco et al.[24] used a hard sphere potential instead of the LJ potential, so it is not possible to compare values of ϵ and σ.

Parameter	Our work reference	Our work with charged surface	Blanco et al.[24] reference
k_h [kJ mol^{-1} nm^{-2}]	992	992	240
ϵ [k_BT]	1	1	—
σ [nm]	0.35	0.35	—
T [K]	298.15	298.15	298.15
c_{peptide} [mol l^{-1}]	0.0008	0.0008	0.0002
c_{salt} [mol l^{-1}]	0.13	0.13	0.1

Figure 5.2 shows the final form of the simulation box together with all its contents.

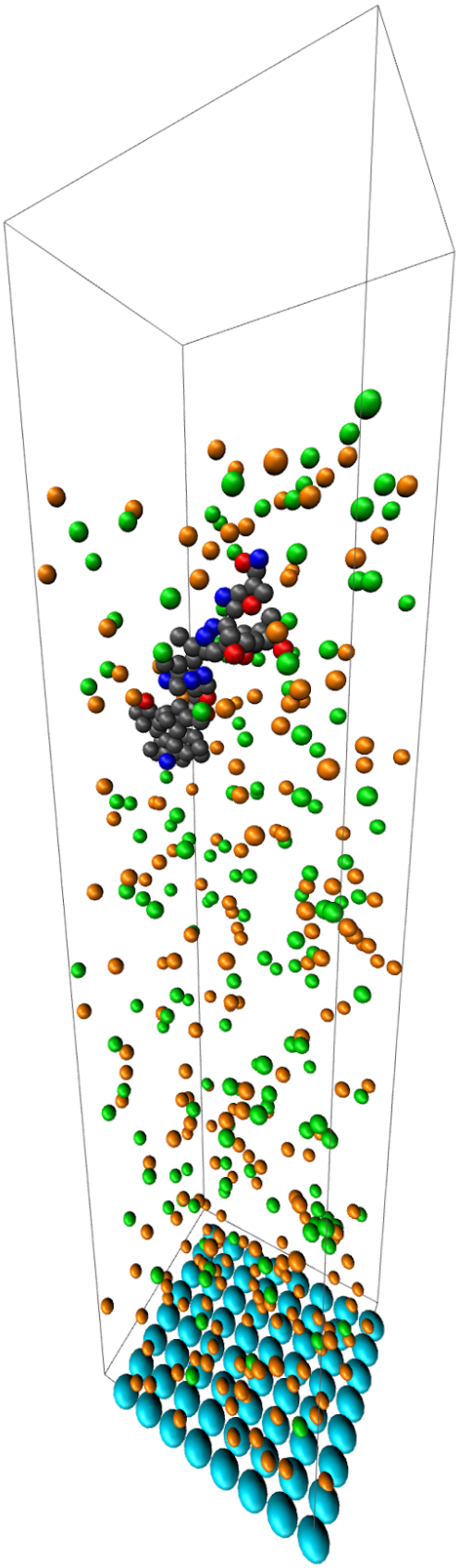


Figure 5.2: Simulation box with Aβ-(1–42) peptide and salt particles in the presence of a charged surface. In the lower part, the particles of the negatively charged surface are shown in cyan color. The sodium ions are shown in orange and chloride ions are represented in green. An increased concentration of sodium ions can be seen near this surface compared to the rest of the simulation box. In the middle part, we can see the Aβ-(1–42) peptide, its non-ionizable groups are depicted in dark gray, acidic ionizable groups are red, and basic ionizable groups are depicted in blue. In the very top part of the simulation box, we can see zero concentration of salt particles. This is because there is an ELC gap in which there are no particles at all.

6. Results and discussion

6.1 Short peptides

6.1.1 Degree of ionization

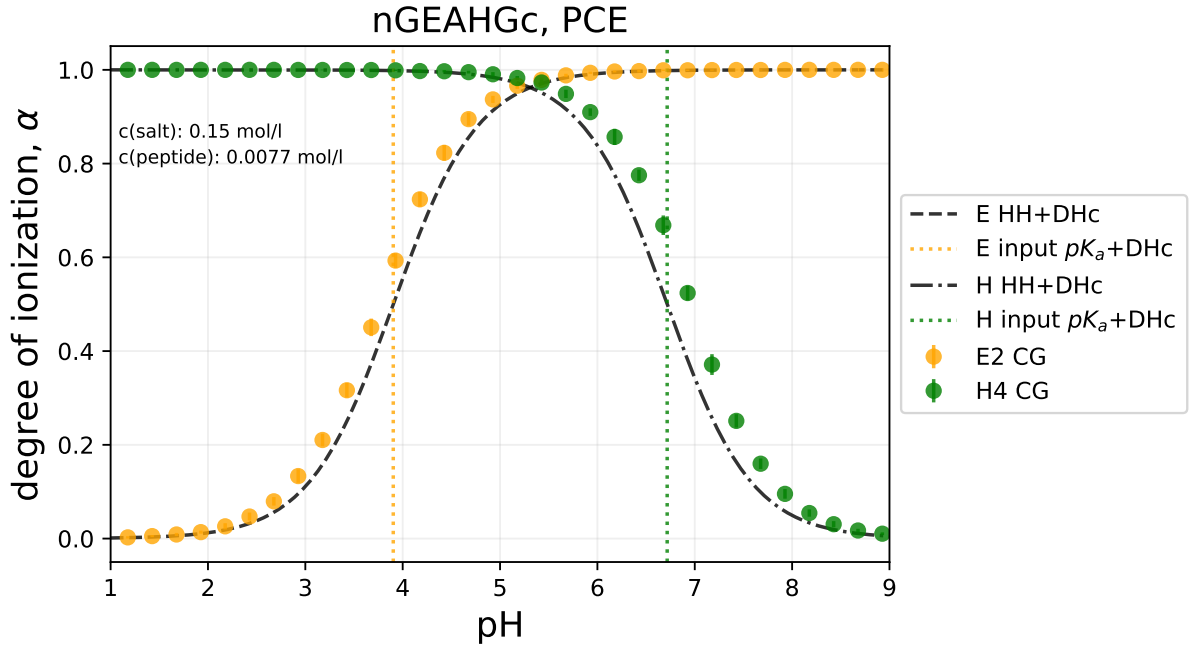


Figure 6.1: Degree of ionization as a function of pH for nGEAHGc sequence with permanently charged ends from CG simulation. E2 represents glutamic acid at position 2 in the sequence and H4 histidine at position 4. Each orange or green circle in Figure 6.1 represents one simulation performed using the constant-pH ensemble method. The black dashed line without dots represents the ideal behavior for glutamic acid described by the Henderson-Hasselbalch equation corrected by the Debye-Huckel correction (E HH+DHc) and the black dashed line with dots represents this ideal behavior for histidine (H HH+DHc). The vertical orange dashed line represents the DHc-corrected simulation pKa input for glutamic acid (E input pK_a+DHc), and the vertical green dashed line represents the DHc-corrected simulation pKa input for histidine (H input pK_a+DHc). All simulation results in this figure were simulated with permanently charged ends (PCE) of the peptide.

In Figure 6.1, there is an example of the coarse-grained (CG) model results where we can see the pH dependence of the degree of ionization of two side chains in nGEAHGc peptide, namely the glutamic acid at position 2 (E2) and histidine at position 4 (H4). We simulated short peptides with permanently charged ends because in Dobrev et al.[8] they also used permanently charged ends in the AA simulation and we wanted to make a comparison as representative as possible. For the salt and the peptide concentration

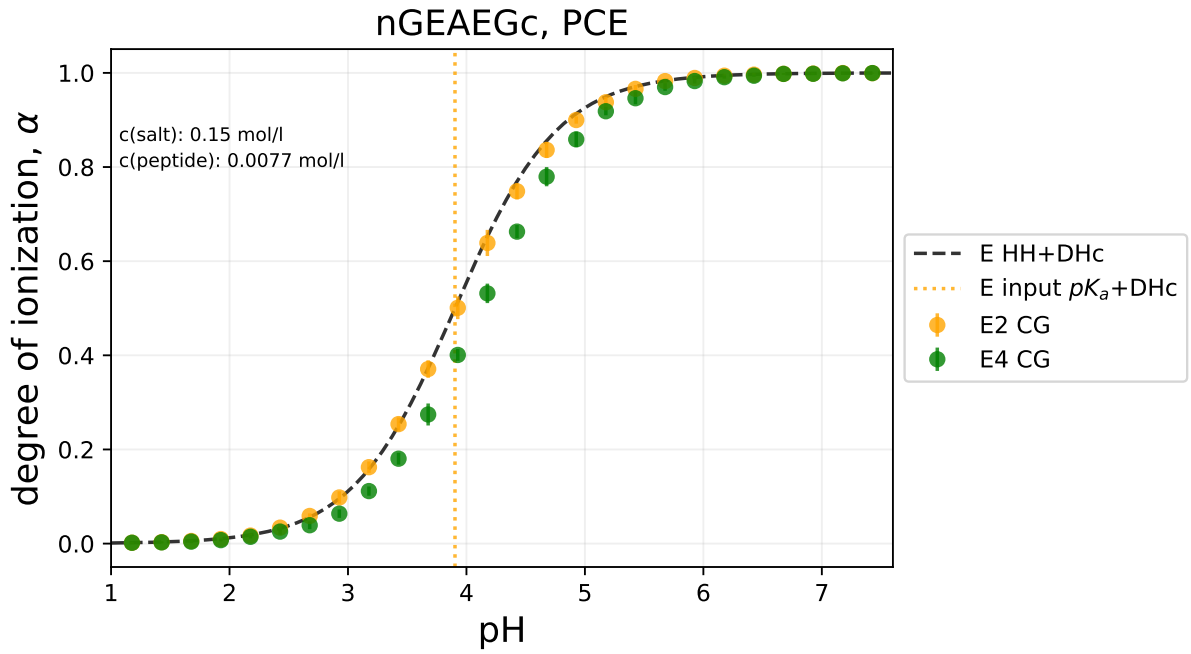


Figure 6.2: Degree of ionization as a function of pH for nGEAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figure 6.1.

we chose the same values as Dobrev et al.[8] used for their AA simulation (also because of a representative comparison). However, we also wanted to study the influence of the ends ionizability so we also performed a simulation with non-permanently charged ends (NPCE). For this aim we chose an arbitrary pK_a set (Nozaki et al.[34]) and took the pK_a values of n-terminus and c-terminus from this pK_a set.

We can notice that our results for E2 and H4 systematically deviate from the ideal curve corrected using Debye–Hückel correction (HH+DHc) in Figure 6.1. Both in the case of E2 and H4 it is an increase in the degree of ionization compared to ideal behavior (HH+DHc). A rough first estimate of the cause of this increase in ionization can be attributed to the fact that in the **n**G**E**A**H**G**c** sequence there are ionizable groups distributed on the alternation. The positively charged side chain (H4) is located between the two negatively charged ionizable groups (E2 and the c-terminus) and the negatively charged side chain (E2) is located between the two positively ionizable groups (the n-terminus and H4). In contrast, in Figure 6.2 we can observe a decrease in the degree of ionization for individual side chains in the **n**G**E**A**E**G**c** sequence where the E4 side chain is located between the two negatively charged ionizable ends (E2 and c-terminus).

However, it should be noted that these changes in the degree of ionization depend not only on the order of ionizable groups in the sequence, but also on other parameters, namely on the geometry of the molecule and the choice of the pK_a value as an input parameter of the simulation.

Results for all other sequences are similar and can be found in Appendix (Figures 7.9-7.16).

6.1.2 Data fitting and obtaining pK_a(eff) values

From all CG data we can obtain effective pK_a values, pK_a(eff). For this purpose we simply fit our data with either the Henderson-Hasselbalch equation (Equation 2.18 for an acid and 2.19 for a base) or the Hill equation (Equation 2.26 for an acid and 2.27 for a base). Then we get the effective pK_a value pK_a(eff) as a fitting parameter. The advantage of the Hill equation is that we can find out the slope of the titration curve of our fitted data (n parameter in Equation 2.26 and 2.27).

In Figure 6.3 we can see the same results as in Figure 6.1 but here they are fitted by the Henderson-Hasselbalch equation (Equation 2.18).

$$\text{res}_\alpha = \alpha(\text{simulation}) - \alpha(\text{fit}) \tag{6.1}$$

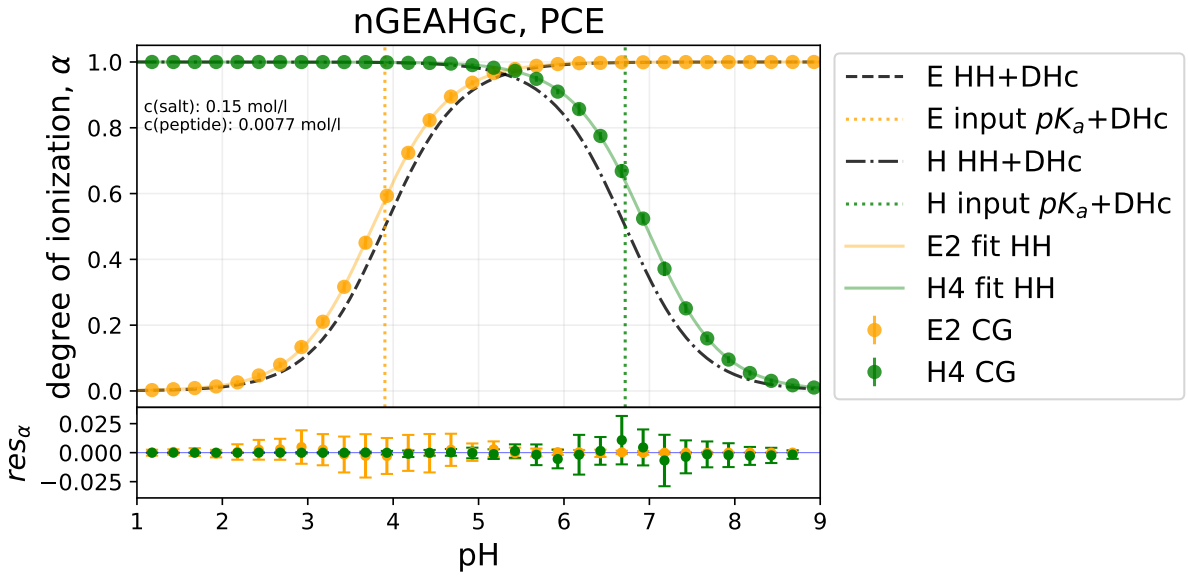


Figure 6.3: Fit using the Henderson-Hasselbalch equation of degree of ionization as a function of pH for nGEAHGc sequence from CG simulation, PCE. The fit of E2 is represented by the solid orange line and the solid green line represents the fit for H4. We can also see the residuals at the bottom of the plot, that is the difference between the value of the degree of ionization obtained by simulation α(simulation) and the value of the degree of ionization from the given fit α(fit) (Equation 6.1). An explanation of the other labels can be found in Figure 6.1.

For the nGEAHGc sequence, both fits, the Henderson-Hasselbalch fit in Figure 6.3 and the Hill fit in Figure 6.4, are almost the same and our CG data are equally distributed around the fit in the pH range 1-9.

However, for sequences with two same side chains (nGEAEGc and nGHAHGc) distribution of our CG data around the fit is pretty different. For the nGEAEGc sequence we can observe a significant wave in residuals, positive res$_\alpha$ value for pH $\doteq$ pK_a−0.5 and negative res$_\alpha$ value for pH $\doteq$ pK_a+0.5 (Figure 6.5). This observation is even more significant for the NPCE (Figure 6.6). It means that at a pH value which is a bit lower

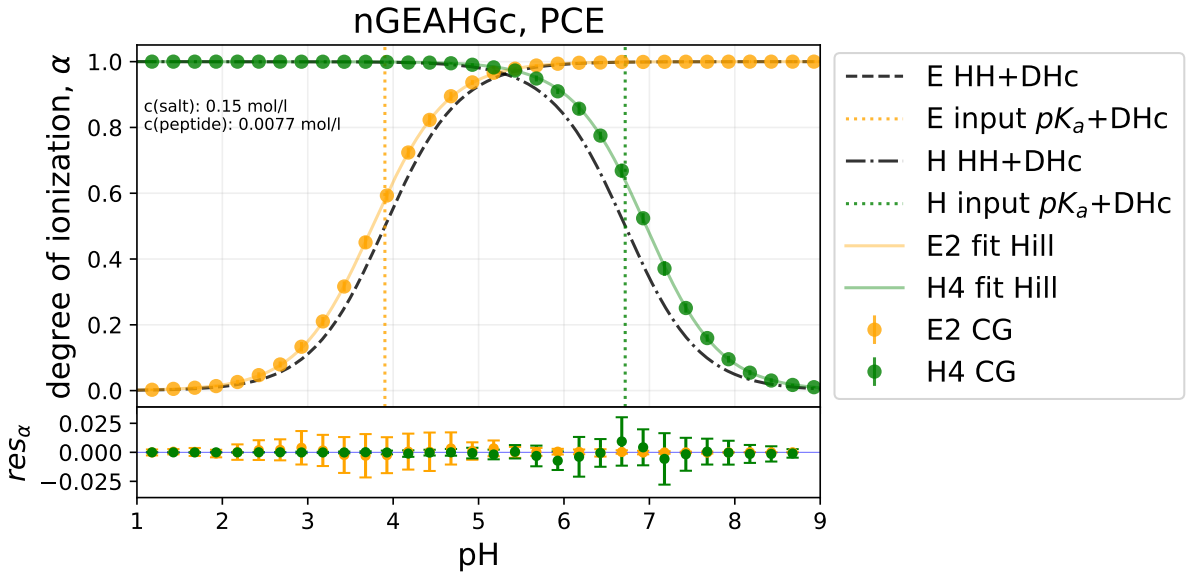


Figure 6.4: Fit using the Hill equation of degree of ionization as a function of pH for nGEAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

than the pK_a value the degree of ionization of a particular amino acid in the oligopeptide is higher than in the case of an ideal isolated amino acid. At a bit higher pH value the situation is opposite, the degree of ionization is lower. For the nGHAHGc sequence we can also observe this wave - lower degree of ionization for pH $\doteq$ pK_a−0.5 and higher for pH $\doteq$ pK_a+0.5 (Figures 6.7 and 6.8, also bigger wave for NPCE).

In the Figure 6.6, the wave occurs because at pH < 4 only the n-terminus is charged, 1 positive charge in the whole molecule, and thus the ionization of another acidic group is supported compared to ideal behavior. Then, at a pH value a bit higher than the pK_a value the degree of ionization is lower compared to the ideal behavior because the number of negative charges in the molecule increases and these charges repel each other. There is an analogous explanation for the nGHAHGc sequence in Figure 6.8.

Results for all other sequences for the data fitting can be found in the Appendix (Figures 7.17-7.32).

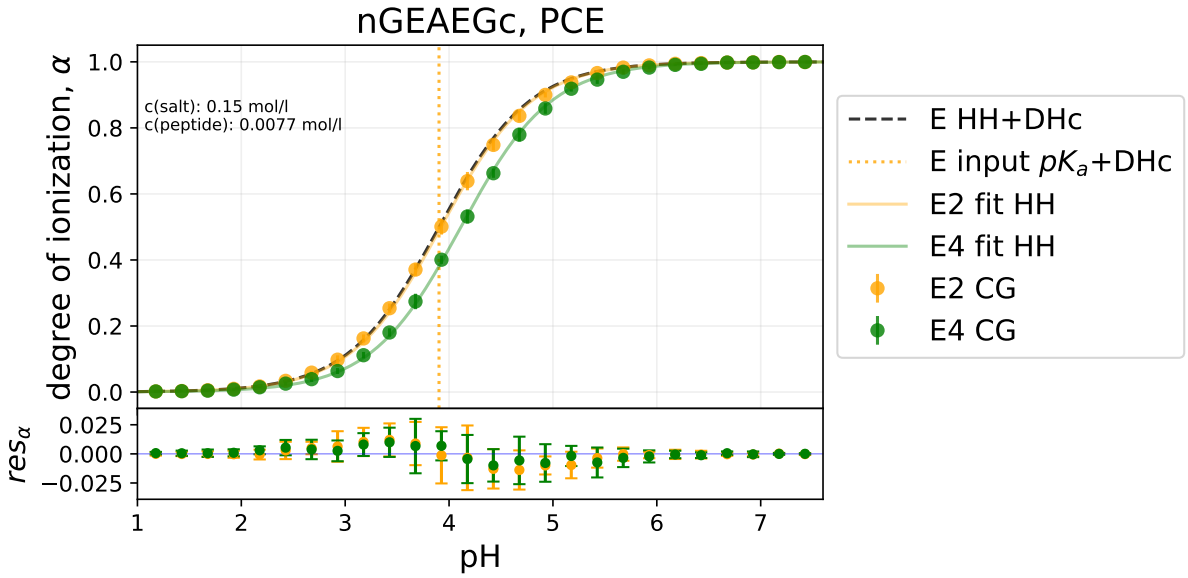


Figure 6.5: Fit using the Henderson-Hasselbalch equation of degree of ionization as a function of pH for nGEAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

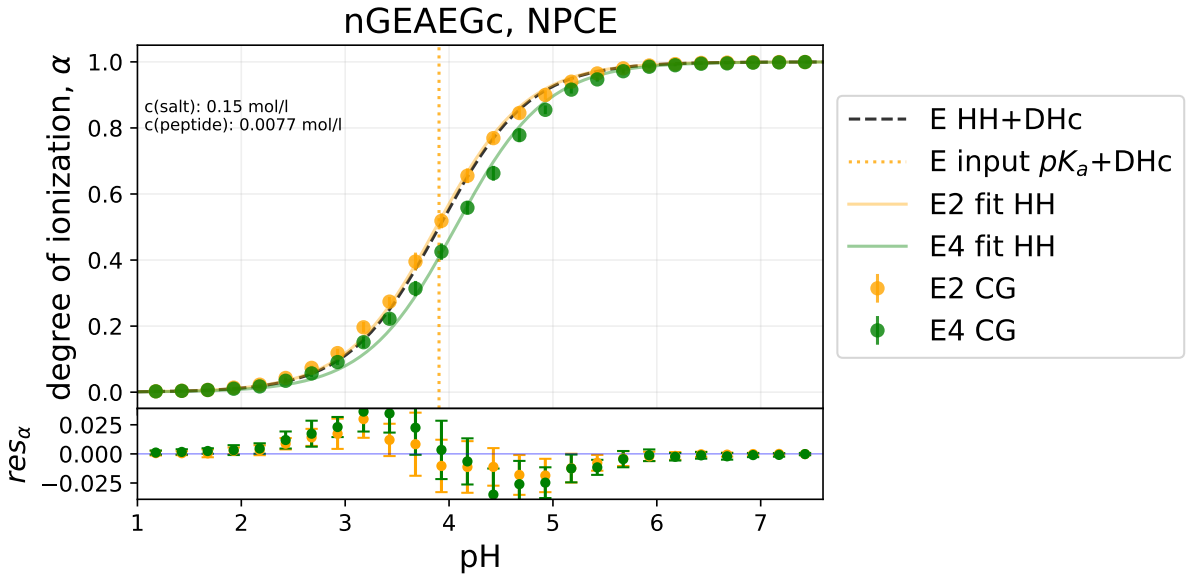


Figure 6.6: Fit using the Henderson-Hasselbalch equation of degree of ionization as a function of pH for nGEAEGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

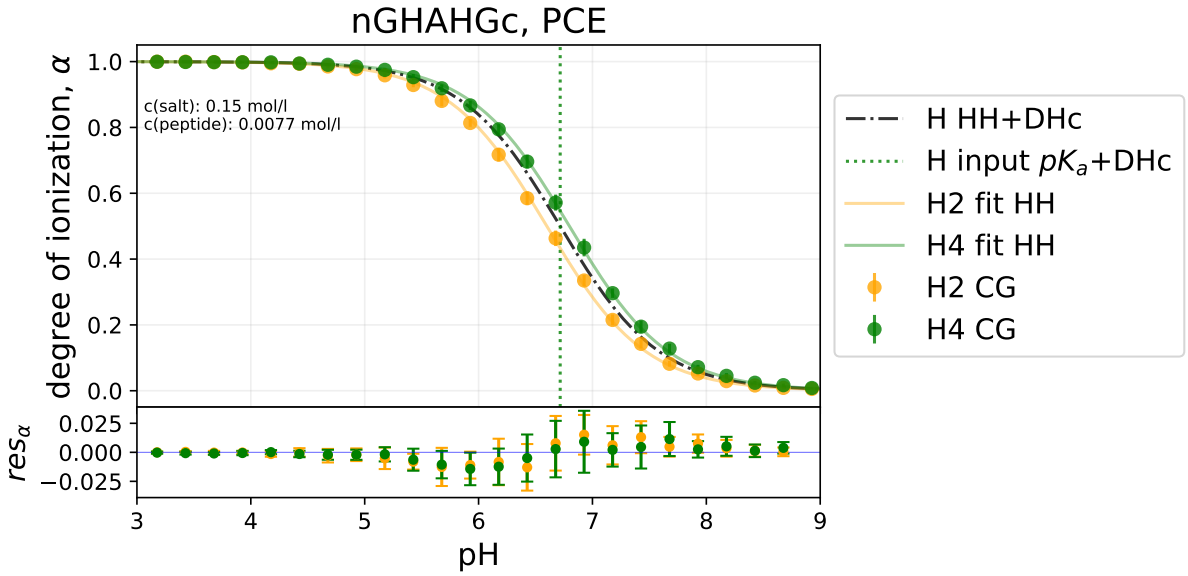


Figure 6.7: Fit using the Henderson-Hasselbalch equation of degree of ionization as a function of pH for nGHAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

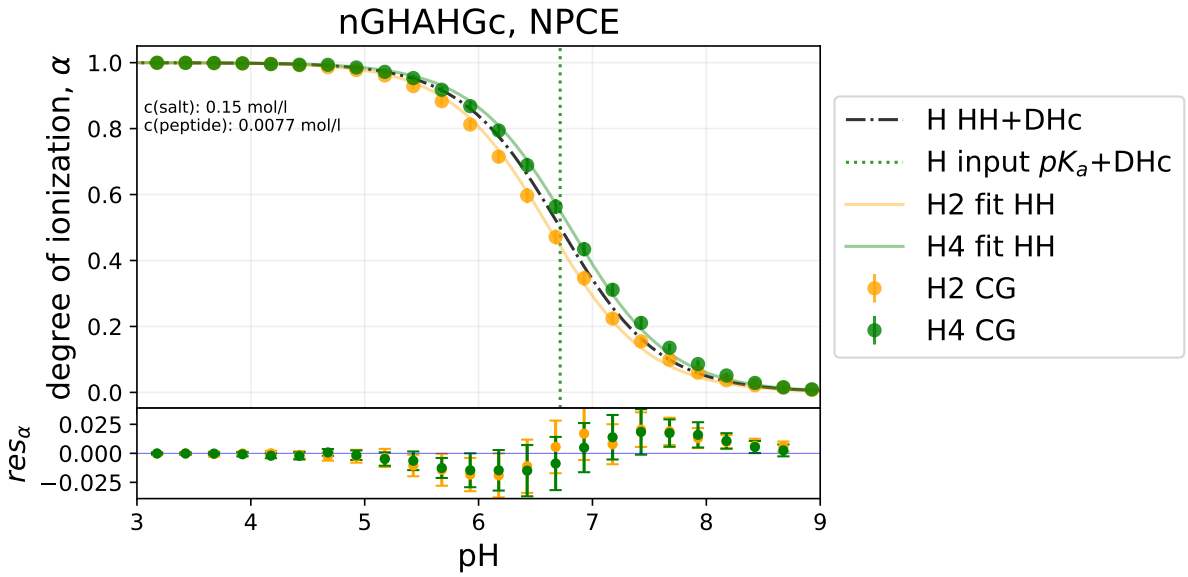


Figure 6.8: Fit using the Henderson-Hasselbalch equation of degree of ionization as a function of pH for nGHAHGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

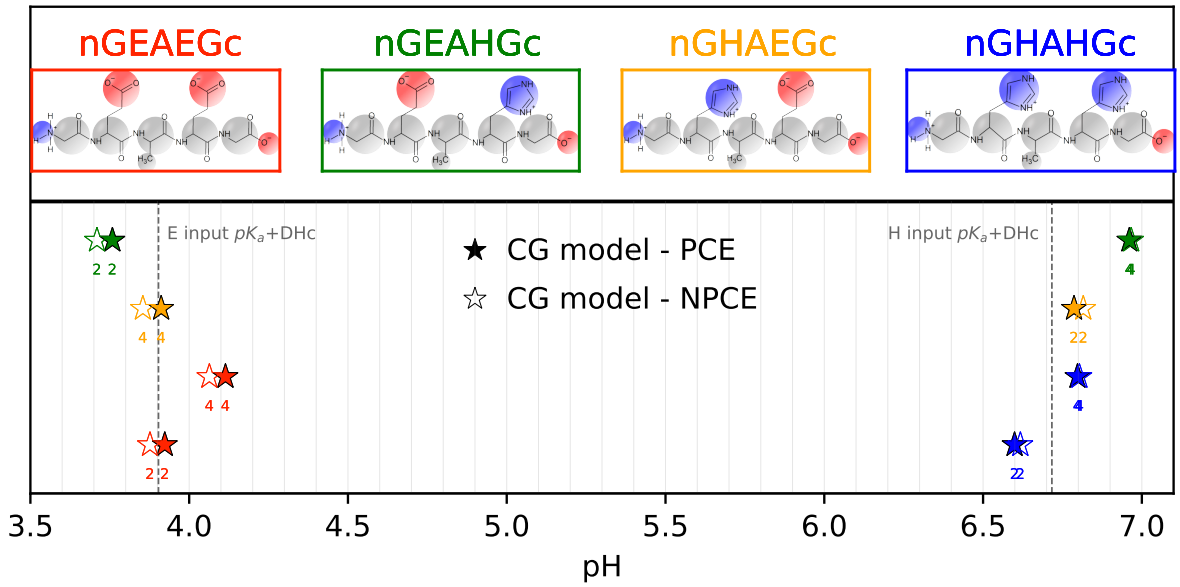


Figure 6.9: pK_a(eff) values obtained using our CG results and input pK_a values from Dobrev et al.[8]. Values of pK_a(eff) are represented by stars in the plot, filled stars are for permanently charged ends (PCE) and an empty stars are for non-permanently charged ends (NPCE). Every sequence is represented by its own color which is the color of the frame around the schematic of that molecule in Figure 6.9. The small number below the star represents the position of an amino acid in the molecule, 2 is closer to the n-terminus and 4 is closer to the c-terminus. Dashed vertical lines represent input pK_a values corrected by the Debye–Hückel correction.

6.1.3 pK_a(eff) and ΔpK_a values for various pK_a sets

We also wanted to study the effect of various pK_a values as a simulation input, so we chose several pK_a sets and performed the simulation at these values. At first, we ran a simulation on pK_a values for glutamic acid and histidine from Dobrev et al.[8] and pK_a values for n-terminus and c-terminus were taken from an arbitrarily chosen pK_a set - Nozaki et al.[34]. And then we performed a simulation for two other arbitrarly chosen pK_a sets, Nozaki et al.[34] and Thurlkill et al.[35]. All the pK_a values are in Table 6.1. When we ran the simulation for permanently charged ends (PCE) we set up the pK_a of n-terminus for value of 20 and pK_a of c-terminus for value of -5.

Table 6.1: pK_a values for various pK_a sets

pK_a set	n-terminus (n)	glutamic acid (E)	histidine (H)	c-terminus (c)
Dobrev et al.[8]	7.5	4.08	6.54	3.8
Nozaki et al.[34]	7.5	4.4	6.3	3.8
Thurlkill et al.[35]	8.00	4.25	6.54	3.67

We can see the results for the pK_a(eff) obtained using our CG results and input pK_a values from Dobrev et al.[8] in Figure 6.9. Results for the other pK_a sets are in Figure

7.33 in Appendix. Exact values of pK_a(eff) can be found in Tables 6.2 and 6.3.

Table 6.2: pK_a(eff) values for various pK_a sets for permanently charged ends (PCE)

PCE	Side chain	CG model Dobrev pK_a set	CG model Nozaki pK_a set	CG model Thurlkill pK_a set	AA model[8]
nGEAEGc	E2	3.92	4.23	4.09	3.81±0.05
	E4	4.11	4.44	4.29	4.09+0.05−0.04
nGEAHGc	E2	3.75	4.07	3.92	3.60±0.05
	H4	6.96	6.72	6.95	6.77±0.05
nGHAEGc	H2	6.79	6.55	6.79	6.24±0.04
	E4	3.91	4.23	4.08	3.92±0.04
nGHAHGc	H2	6.60	6.35	6.60	6.07±0.05
	H4	6.80	6.56	6.78	6.58±0.05

Table 6.3: pK_a(eff) values for various pK_a sets for permanently charged ends (NPCE)

NPCE	Side chain	CG model Dobrev pK_a set	CG model Nozaki pK_a set	CG model Thurlkill pK_a set	NMR[8]
nGEAEGc	E2	3.88	4.21	4.06	4.06±0.01
	E4	4.06	4.40	4.24	4.04±0.01
nGEAHGc	E2	3.71	4.05	3.91	3.76±0.02
	H4	6.97	6.74	6.96	6.59±0.02
nGHAEGc	H2	6.82	6.59	6.80	6.14±0.01
	E4	3.85	4.20	4.05	3.81±0.01
nGHAHGc	H2	6.62	6.38	6.62	6.23±0.01
	H4	6.80	6.57	6.80	6.66±0.03

Green cells in Tables 6.2 and 6.3 represent results of our CG model which agree with the result of Dobrev et al.[8] on the same line.

Another quantity which we can monitor is delta ΔpK_a (Equation 2.24) which in this case we can calculate as

$$\Delta\mathrm{p}K_a = \mathrm{p}K_a(\mathrm{eff}) - \mathrm{p}K_a(\mathrm{HH+DHc}) \tag{6.2}$$

As we can notice, ΔpK_a for PCE almost does not depend on the input pK_a values or the chosen pK_a set (see Table 6.4). For NPCE, the values differ slightly for individual pK_a sets although they are still very close to each other (see Table 6.5).

Green cells in Table 6.5 represent results of our CG model which agree with the result of Dobrev et al.[8] on the same line.

Table 6.4: $\Delta \mathrm{p}K_a$ values for various $\mathrm{p}K_a$ sets for permanently charged ends (PCE)

PCE	Side chain	CG model			AA model[8]
		Dobrev $\mathrm{p}K_a$ set	Nozaki $\mathrm{p}K_a$ set	Thurlkill $\mathrm{p}K_a$ set	
nGEAEGc	E2	0.01	0.01	0.02	−0.27
	E4	0.21	0.21	0.22	0.01
nGEAHGc	E2	−0.15	−0.15	−0.16	−0.48
	H4	0.24	0.24	0.24	0.23
nGHAEGc	H2	0.07	0.08	0.07	−0.30
	E4	0.01	0.01	0.00	−0.16
nGHAHGc	H2	−0.12	−0.13	−0.12	−0.47
	H4	0.08	0.08	0.07	0.04

Table 6.5: $\Delta \mathrm{p}K_a$ values for various $\mathrm{p}K_a$ sets for permanently charged ends (NPCE)

NPCE	Side chain	CG model			NMR[8]
		Dobrev $\mathrm{p}K_a$ set	Nozaki $\mathrm{p}K_a$ set	Thurlkill $\mathrm{p}K_a$ set	
nGEAEGc	E2	−0.03	−0.01	−0.01	−0.01
	E4	0.16	0.18	0.16	−0.03
nGEAHGc	E2	−0.19	−0.17	−0.16	−0.31
	H4	0.25	0.26	0.25	0.04
nGHAEGc	H2	0.10	0.11	0.09	−0.41
	E4	−0.05	−0.03	−0.02	−0.26
nGHAHGc	H2	−0.10	−0.10	−0.10	−0.32
	H4	0.09	0.09	0.08	0.11

Then, we can calculate difference between $\Delta \mathrm{p}K_a$ from the simulation and from an experiment using Equation 2.25. Results of the $\Delta\Delta \mathrm{p}K_a$ are in Table 6.6.

Orange cells in Table 6.6 represent results of our CG model which are closer to the value of 0 than the result of Dobrev et al. on the same line (which means that this particular line in Table 6.6 our CG model is more accurate than the AA model from Dobrev et al.).

Table 6.6: $\Delta\Delta \mathrm{p}K_a$ values for various $\mathrm{p}K_a$ sets for permanently charged ends (NPCE)

NPCE	Side chain	CG model			Dobrev et al.
		Dobrev $\mathrm{p}K_a$ set	Nozaki $\mathrm{p}K_a$ set	Thurlkill $\mathrm{p}K_a$ set	
nGEAEGc	E2	−0.02	0	0	−0.25
	E4	0.19	0.21	0.19	0.05
nGEAHGc	E2	0.12	0.14	0.15	−0.16
	H4	0.21	0.22	0.21	0.18
nGHAEGc	H2	0.51	0.52	0.50	0.10
	E4	0.21	0.23	0.24	0.11
nGHAHGc	H2	0.22	0.22	0.22	−0.16
	H4	−0.02	−0.02	−0.03	−0.08

6.1.4 Hill coefficient n

The very last monitored variable is the Hill coefficient n (see in Equation 2.26 and 2.27). In the case of ideal behavior, in the Henderson-Hasselbalch equation, n equals 1. In case of $n > 1$ the slope of the titration curve is greater then for an ideal curve and the middle part of the curve is more vertical. For $n < 1$ the slope of the titration curve is less steep than for an ideal curve and the middle part of the curve is more horizontal. Values of the n coefficient can be obtained by simple fitting of the data from CG simulation using the Hill equation (Equation 2.26 and 2.27). The Hill coefficients for the CG model, the AA model and the NMR experiment can be found in Table 6.7 and 6.5.

Table 6.7: Hill coefficients for various $\mathrm{p}K_a$ sets for permanently charged ends (PCE)

PCE	Side chain	CG model			AA model[8]
		Dobrev $\mathrm{p}K_a$ set	Nozaki $\mathrm{p}K_a$ set	Thurlkill $\mathrm{p}K_a$ set	
nGEAEGc	E2	0.95	0.96	0.96	1.08+0.1−0.09
	E4	0.96	0.96	0.96	1.04+0.09−0.08
nGEAHGc	E2	1.00	0.99	1.00	1.05+0.1−0.08
	H4	1.01	1.00	1.00	0.97+0.08−0.07
nGHAEGc	H2	1.01	1.00	1.01	0.92+0.06−0.05
	E4	0.99	0.99	1.00	1.12+0.08−0.07
nGHAHGc	H2	0.95	0.95	0.95	0.91+0.07−0.06
	H4	0.96	0.94	0.94	1.02+0.09−0.07

Green cells in Tables 6.7 and 6.8 represent results of our CG model which agree with the result of Dobrev et al.[8] on the same line.

In Figure 6.10, we can see that the all-atom model has significantly larger error bars than our coarse-grained model and also the AA model deviates from the value of 1 in the opposite direction than the experimental value in 5 out of 8 cases.

Table 6.8: Hill coefficients for various pK_a sets for permanently charged ends (NPCE)

NPCE	Side chain	CG model			NMR[8]
		Dobrev pK_a set	Nozaki pK_a set	Thurlkill pK_a set	
nGEAEGc	E2	0.92	0.92	0.93	0.90±0.02
	E4	0.88	0.89	0.90	0.79±0.02
nGEAHGc	E2	0.96	0.97	0.96	0.94±0.03
	H4	0.98	0.98	1.00	0.91±0.03
nGHAEGc	H2	0.96	0.96	1.00	0.96±0.01
	E4	0.93	0.94	0.92	0.90±0.02
nGHAHGc	H2	0.92	0.93	0.94	1.06±0.03
	H4	0.93	0.92	0.95	0.85±0.04

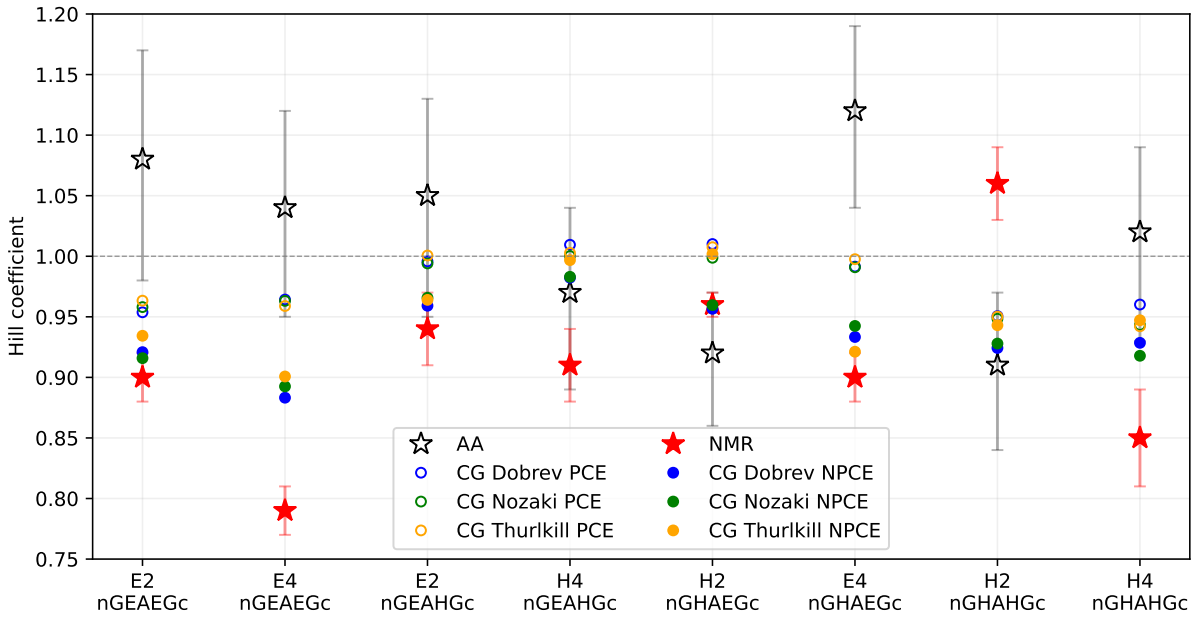


Figure 6.10: Hill coefficients for all side chains of all studied sequences. AA symbolizes the results of the all-atom simulation and together with NMR represent the results from Dobrev et al.[8]. Our coarse-grained results are shown by empty circles for permanently charged ends and full circles for non-permanently charged ends.

6.2 The Aβ-(1–42) peptide at a charged surface

6.2.1 Simulation of salt particles at a charged surface

First, we had to test whether the RxMC method itself reliably maintains a constant chemical potential in the simulation box and the desired salt concentration in the bulk. For this purpose, we ran a simulation of salt particles at a charged surface without the Aβ-(1-42) peptide. As input parameters, we chose the test desired salt concentration, 0.01M, and the activity coefficient calculated using the DHc (Equation 2.20). In this simulation of salt particles at a charged surface, we achieved a concentration of 0.01M in the bulk. The resulting density profile for salt particles is shown in Figure 6.11. Based on this simulation of salt particles in the presence of a charged surface without Aβ-(1–42) peptide, we conclude that the RxMC method reliably maintains a constant chemical potential in the simulation box and the desired salt concentration in bulk.

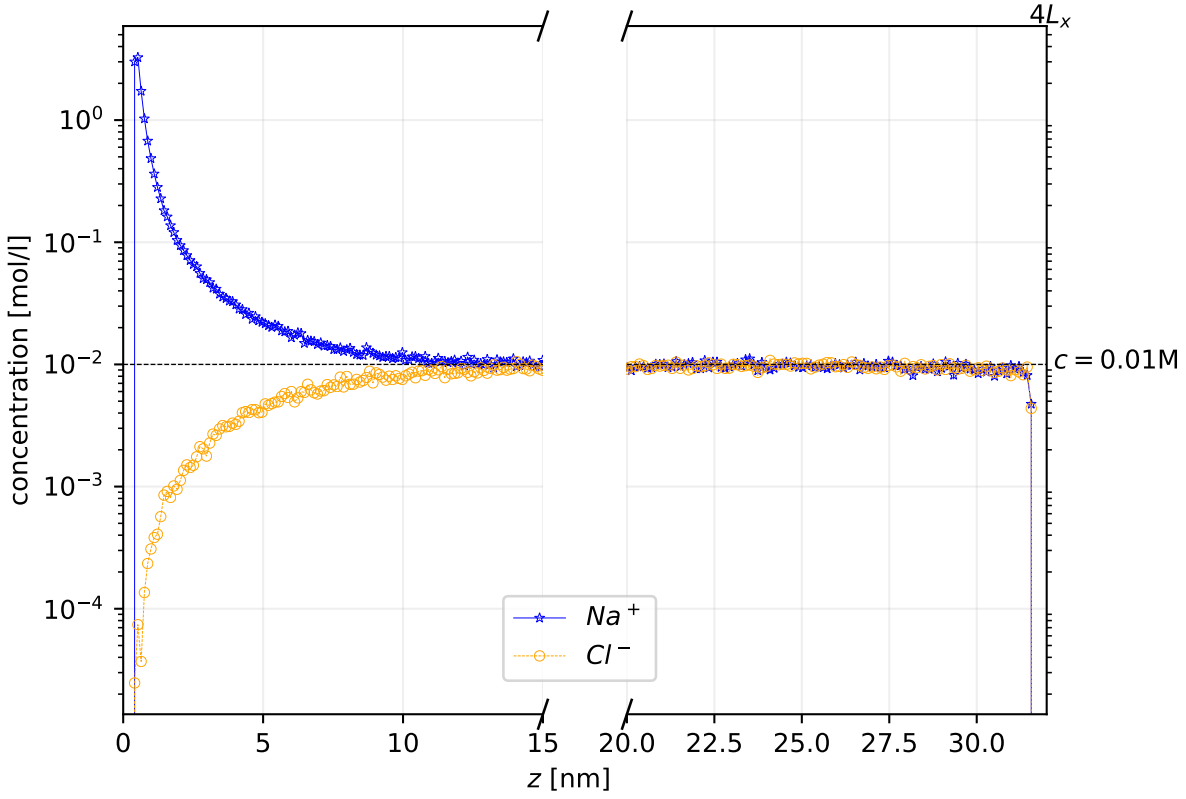


Figure 6.11: Density profile of Na^+ and Cl^- particles in a simulation of salt particles at a charged surface. Size of the simulation box: $4L_x = 4L_y = L_z = 32$ nm.

6.2.2 Verification of the suitability of selected parameters

Before running the simulation to obtain results for the acid-base behavior of the Aβ-(1–42) peptide in the presence of a charged surface, we had to use test simulations to determine the correct technical parameters of the simulation that will not affect the physical and chemical properties of our study system.

6.2.2.1 Box size and density profiles of salt particles

The first parameter was the box size determined on the density profile of the salt particles. A test simulation designed to determine the correct box size contained the Aβ-(1–42) peptide in the presence of a charged surface and salt particles (NaCl). The RxMC method was used in this simulation to ensure a constant chemical potential of the salt throughout the simulation box. The input parameters were a salt concentration of 0.15 M and the mean activity coefficient for this concentration calculated using DHc (Equation 2.20).

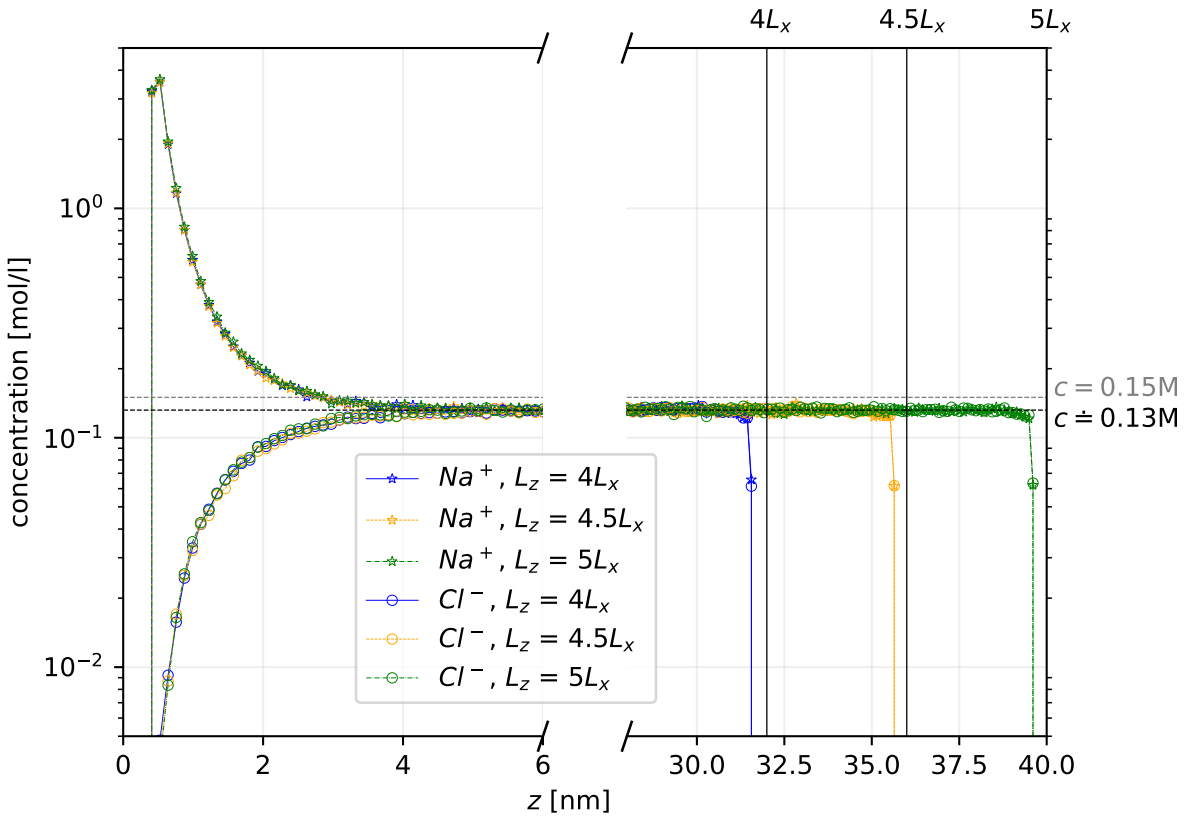


Figure 6.12: Density profile of Na^+ and Cl^- particles in simulation boxes with various simulation box edge lengths in z-direction, L_z. Size of the simulation box: $L_x = 8$ nm.

It can be seen from Figure 6.12 that for these lengths the density profile near the surface is very similar. Therefore, out of these tested box sizes we chose the smallest box size due to the shortest computation time.

Although we chose the required salt concentration to be 0.15M, the final concentration in the bulk was a bit lower, around 0.132M. It was due to the fact that the DHc applies only to a certain range of salt concentrations, approximately for concentrations lower than 0.01M. To achieve a concentration of 0.15M it is necessary to use more precise relations suitable for other ranges of salt concentrations, for example, the McInnes approximation[36]

$$-\log_{10}\gamma_{\pm} = \frac{A|z_{+}z_{-}|\sqrt{I}}{1+1.5\sqrt{I}} \tag{6.3}$$

where A represents coefficient and its value can be found in Table 2.2, or the Davies equation[37].

$$-\log_{10}\gamma_{\pm} = 0.5|z_{+}z_{-}|\left(\frac{\sqrt{I}}{1+\sqrt{I}} - 0.2I\right) \tag{6.4}$$

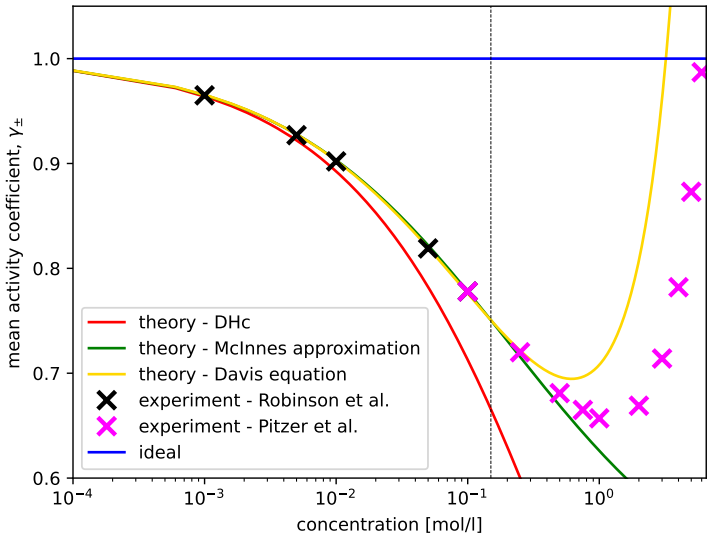


Figure 6.13: Mean activity coefficient depending on NaCl concentration in an aqueous solution. Black crosses represent experimental data from Robinson et al.[38] and pink crosses are for experimental data from Pitzer et al.[39].

6.2.2.2 Thickness of the ELC gap

Another parameter was the thickness of the ELC gap. Here it was necessary to verify that the density profile does not deviate in any way from the required salt concentration in the places farthest from the surface, around 30 nm on the x-axis in Figure 6.14. Here we can see that even a very thin ELC gap layer, 0.8 nm, effectively shields the charged surface effect in periodic boundary conditions. To be safe, we chose a slightly wider ELC gap of 4 nm.

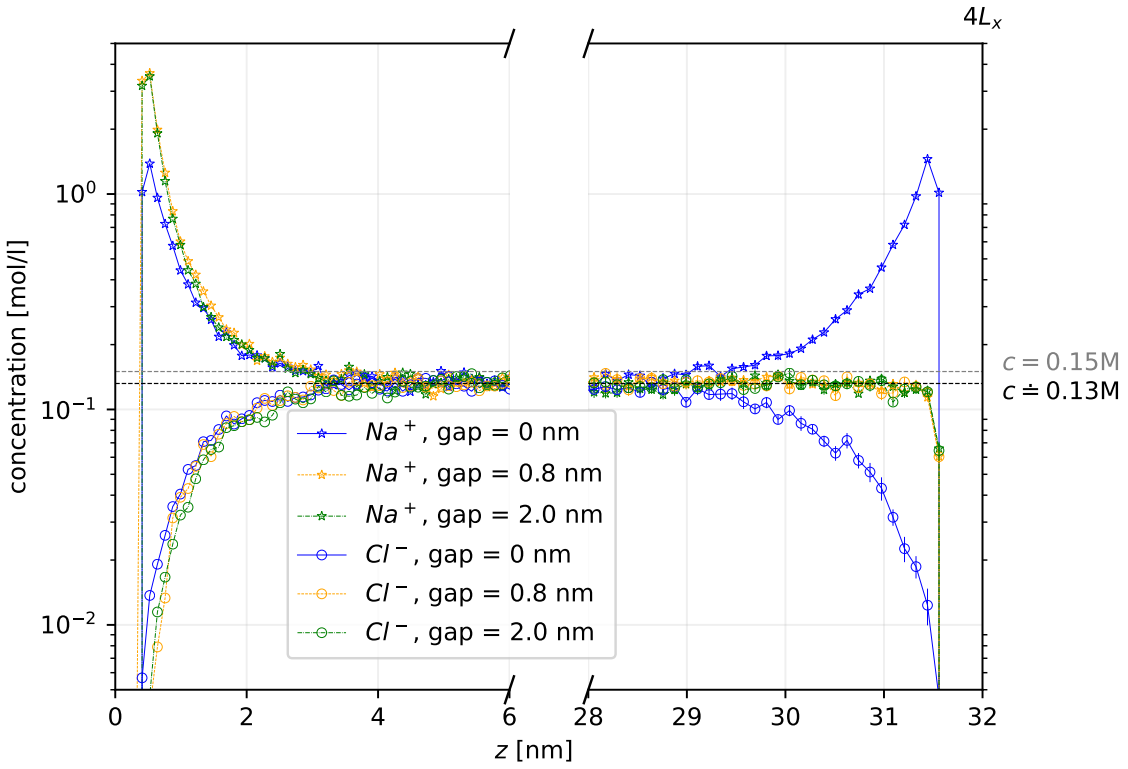


Figure 6.14: Density profile of Na^+ and Cl^- particles in simulation boxes with various ELC gap thicknesses. Size of the simulation box: $4L_x = 4L_y = L_z = 32$ nm.

6.2.3 Net charge on the peptide

After determining all the necessary parameters, we ran a simulation of the Aβ-(1–42) peptide in the presence of a charged surface and salt particles to determine the acid-base behavior of the Aβ-(1–42) peptide in the presence of a charged surface. We first focused on the net charge on the peptide. In Figures 6.16 and 6.17, we compared the data of the net charge as a function of pH for the peptide in the presence of a charged surface to two reference simulations.

One reference simulation was performed by us. In this simulation we set up exactly the same peptide and salt concentration as we set for the simulation with the charged surface and the peptide was also represented by CG, 2-bead model. However in this simulation there was no charged surface. In Figure 6.15, there is a snapshot of our reference simulation. We found the second reference simulation in Blanco et al.[24] where they studied the Aβ-(1–42) peptide using the coarse-grained simulation, however they used the 1-bead model. This model only has a backbone, there are no side chains, and it differs from the 2-bead model mainly in the geometry of the molecule.

An overview of all the parameters of the simulations can be seen in Table 6.9.

As we can notice in Figures 6.16 and 6.17, at low value of pH, $1 <$ pH < 4, the peptide has a positive charge and its net charge is strongly influenced by the negatively charged

Table 6.9: Parameters of all the simulations of Aβ-(1–42) peptide

	Our simulation of the peptide without surface	Our simulation of the peptide with charged surface	Simulation of the peptide from literature[24]
$L_x L_y L_z$ [nm^3]	8×8×32	8×8×32	22×22×22
model	CG, 2-bead	CG, 2-bead	CG, 1-bead
c_{peptide} [mol l^{-1}]	0.0008	0.0008	0.0016
c_{salt} [mol l^{-1}]	0.13	0.13	0.1
presence of the surface	no	yes	no

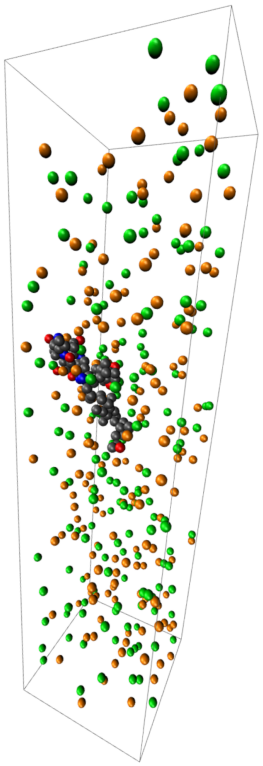


Figure 6.15: Snapshot of our reference simulation of the Aβ-(1–42) peptide. The sodium ions are shown in orange and chloride ions are represented in green. In the middle part, we can see the Aβ-(1–42) peptide, its non-ionizable groups are depicted in dark gray, acidic ionizable groups are red, and basic ionizable groups are depicted in blue.

surface. The net charge of the peptide in the presence of a charged surface is higher than the net charge of the Aβ-(1–42) peptide in both reference systems.

In Figure 6.17, we can also see a slight shift of the isoelectric point pI towards lower pH values compared to the ideal curve.

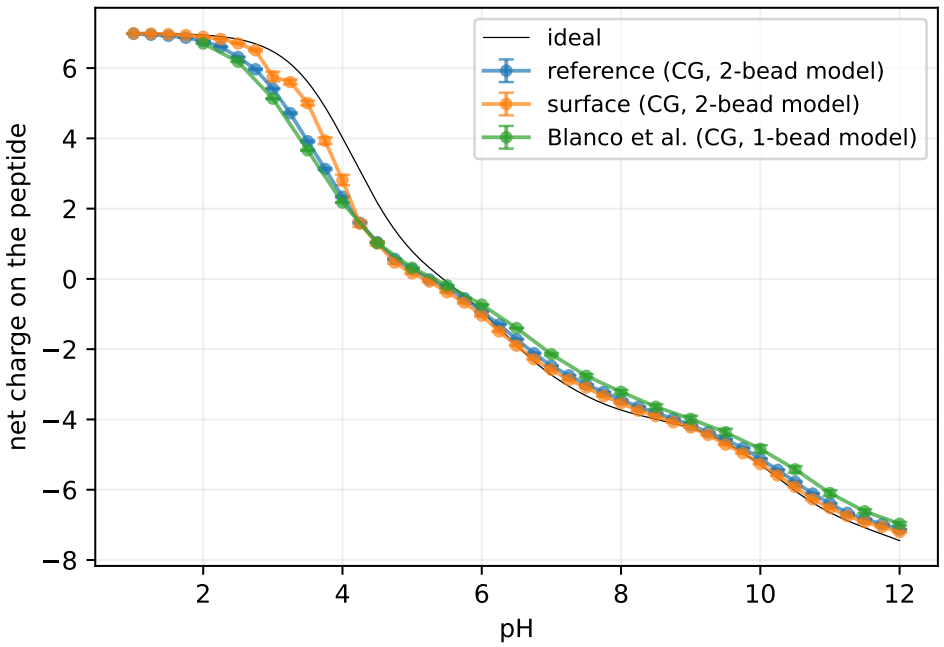


Figure 6.16: Net charge on the Aβ-(1–42) peptide as a funtion of pH. Our reference simulation is shown in orange, our simulation of the Aβ-(1–42) peptide in presence of a charged surface in blue and reference simulation from Blanco et al.[24] is shown in green.

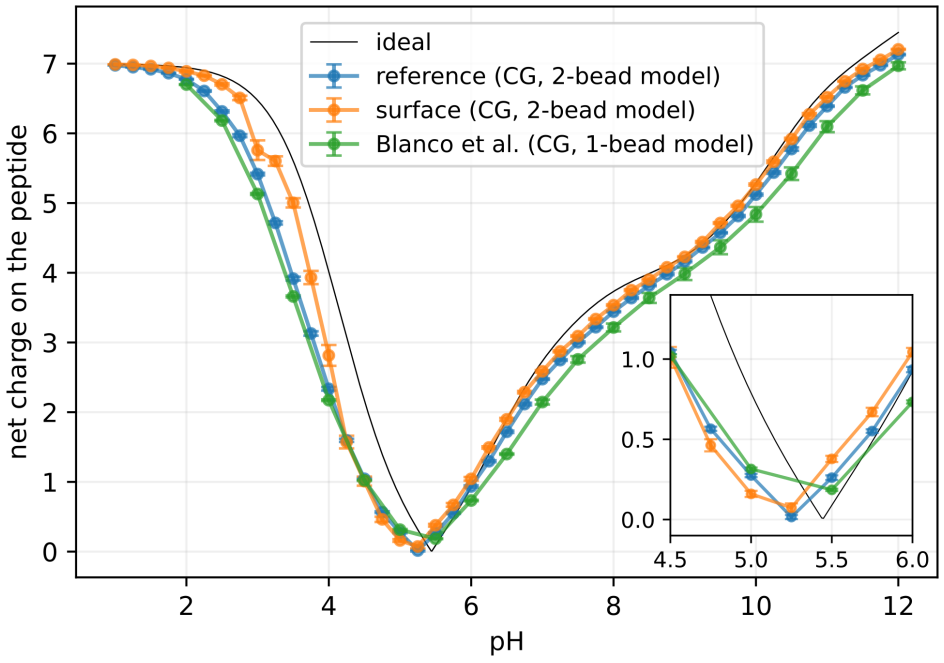


Figure 6.17: Net charge on the Aβ-(1–42) peptide as a funtion of pH in absolute value. An explanation of the individual labels can be found in Figure 6.16.

6.2.4 Radius of gyration R_g

In Figure 6.18, we can see the dependence of the radius of gyration R_g on pH for the 3 simulation systems described in Section 6.2.3.

All the curves in Figure 6.18 have a minimum around the isoelectric point pI. This is quite an expected fact, since the isoelectric point pI is characterized by a zero net charge of the peptide. Therefore at pH = pI, both positive and negative charges are present on the peptide, and the peptide assumes a collapsed conformation. In contrast, at lower pH, $1 <$ pH < 4, the peptide has an overall positive charge, it assumes a stretched conformation, and the more the mass is spread out in space, the greater the radius of gyration. The higher pH region, $6 <$ pH < 12, has an analogous explanation as region $1 <$ pH < 4 but there is an overall negative charge on the peptide.

As we can notice in Figure 6.18, similarly to the net charge on the peptide, R_g of the Aβ-(1–42) peptide in our reference system differs from R_g of the Aβ-(1–42) peptide in the presence of a charged surface mainly at low pH values, $1 <$ pH < 4.

Interestingly, the radius of gyration for the 1-bead model is significantly higher than that of the 2-bead model. We have not yet found an explanation for this observation.

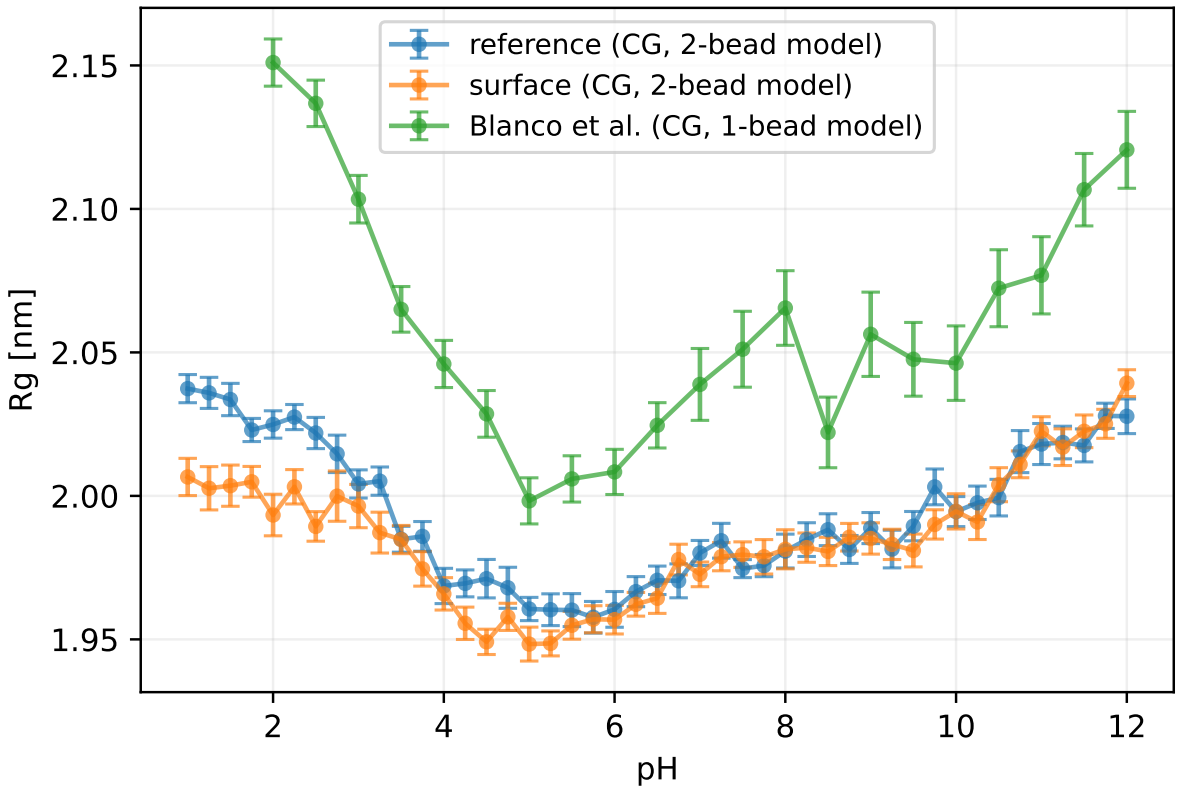


Figure 6.18: Radius of gyration of the Aβ-(1–42) peptide. An explanation of the individual labels can be found in Figure 6.16.

6.2.5 Probability density of the Aβ-(1–42) peptide

Figure 6.19 shows that for low pH values, $1 < \text{pH} < 4$, the probability density is significantly higher near the surface, up to 5 nm from the surface, than in other parts of the simulation box. This is an indication that most likely at low pH, $1 < \text{pH} < 4$, adsorption of the Aβ-(1–42) peptide to the negatively charged surface occurs. However, our simulation of the Aβ-(1–42) peptide in the presence of a charged surface was not sufficiently sampled and thus these results can only be considered preliminary.

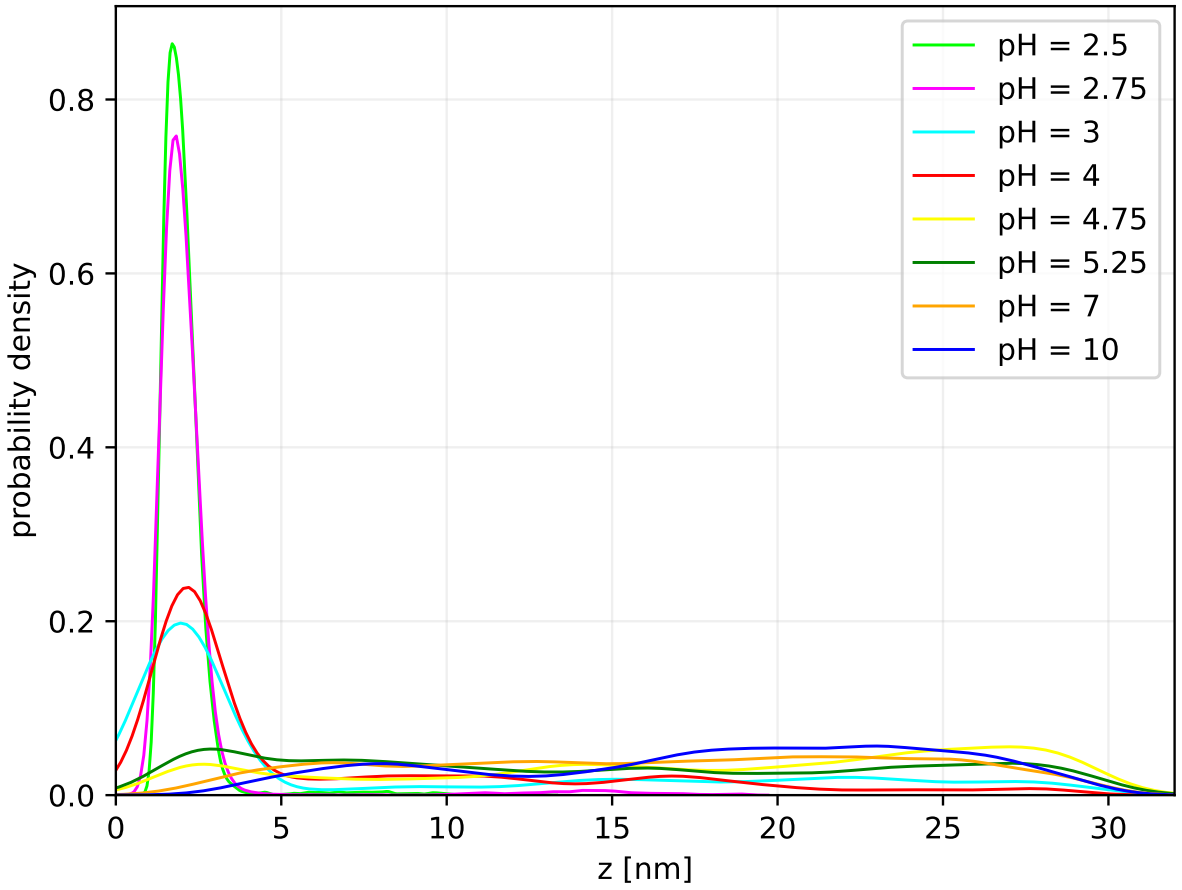


Figure 6.19: Probability density function of observing the Aβ-(1–42) peptide at a particular distance z from the charged surface at various pH values

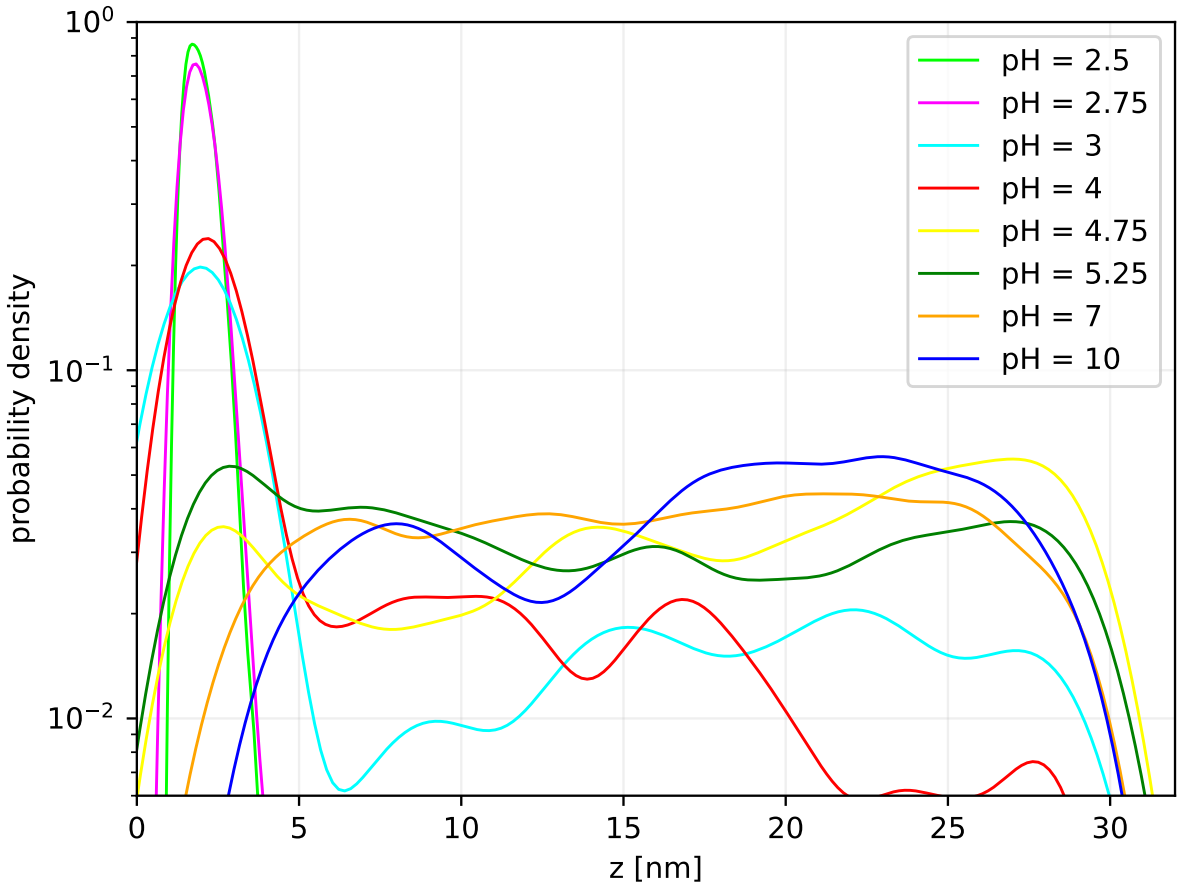


Figure 6.20: Probability density function of observing the Aβ-(1–42) peptide at a particular distance z from the charged surface at various pH values, y-axis in log scale

7. Conclusion

7.1 Short peptides

In this project, we have used our coarse-grained model to investigate the degree of ionization of the side chains of four oligopeptides - nGEAEGc, nGEAHGc, nGHAEGc and nGHAHGc. We then compared our results with other methods, with a computationally more demanding all-atom model and with NMR experiment. The goal was to find out how accurate our model is in determining the acid-base behavior of short peptides.

After a thorough examination of our data we can say that our results of a coarse-grained model systematically deviate from the ideal behavior described by the Henderson Hasselbalch equation corrected by the Debye–Hückel correction due to the presence of salt in the solution. This deviation was characterized in most cases of an acid (Glu) as an increase in the degree of ionization for pH $\doteq$ $\mathrm{pK_a}$ $-$ 0.5 and as a decrease in the degree of ionization for pH $\doteq$ $\mathrm{pK_a}$ + 0.5. For a base (His) it was an increase in the degree of ionization for pH $\doteq$ $\mathrm{pK_a}$ + 0.5 and a decrease at pH $\doteq$ $\mathrm{pK_a}$ $-$ 0.5. This systematic deviation was even more significant for the oligopeptides with two identical side chains and for the setting of non-permanently charged ends in the peptide. This happens because of the non-zero charge on the peptide at the extremes of pH, pH $<$ pKa(acid) and pH $>$ pKa(base). With a non-zero charge on the peptide, ionization of the ionizable group with opposite polarity is always promoted, as a compensation for the given non-zero charge on the peptide. However, at a pH closer to neutral, pH = pK_a(acid) + 0.5 or pH = pK_a(base) $-$ 0.5, more charges are already present on the peptide, repulsion between individual charges occurs and ionization of non-ionized groups is suppressed.

When we then compared our results with the reference data from Dobrev et al.[8], we found that our results differ significantly in $\mathrm{pK_a}$(eff) values compared to the AA model or the NMR experiment (despite the fact that we used the same input $\mathrm{pK_a}$ values as in the AA simulation). Nor the $\Delta\mathrm{pK_a}$ values were in good agreement with the reference data from Dobrev at al.[8]. Nevertheless, the final $\Delta\Delta\mathrm{pK_a}$ values are in most cases very small. On the other hand, the Hill coefficients n of our coarse-grained model are more consistent with both the AA model and the NMR experiment than the pK_a(eff) values.

In conclusion, our coarse-grained model describes acid-base behavior qualitatively rather than quantitatively. However, if we chose some more suitable $\mathrm{pK_a}$ values for these short peptides, then our results could be even better and perhaps even quantitative.

7.2 The Aβ-(1–42) peptide at a charged surface

In our second project, we have tried to determine the conditions under which the Aβ-(1-42) peptide adsorbs onto a charged surface. Despite some inaccuracies that we want to improve in the future, we can draw some conclusions from our preliminary results.

First, for pH lower than 4, we have observed for the Aβ-(1–42) peptide an increase in the net charge on the peptide in the presence of a charged surface compared to the reference state without a charged surface.

Furthermore, at this pH, $1 < \text{pH} < 4$, we have observed a slight increase in the radius of gyration R_g for the Aβ-(1–42) peptide in the presence of a charged surface compared to the peptide in the reference system without a charged surface. This conclusion agrees with the result for the net charge on the peptide. If we have several negative charges in a molecule, then these charges repel each other, the molecule stretches and thus R_g increases. However, it was surprising that the radius of gyration values for our 2-bead model were significantly lower than R_g values for the 1-bead model from Blanco et al.[24] for the pH range 1-9. We do not yet have an explanation for this observation.

Last but not least, we focused on the probability of occurrence of the Aβ-(1–42) peptide at a charged surface. From this observation, we found that at pH, $1 < \text{pH} < 4$, there is a significantly higher probability of the Aβ-(1–42) peptide occurrence at the charged surface compared to $\text{pH} > 4$. For $\text{pH} > 5$, the probability of occurrence was already evenly distributed throughout the system and there was no preferential site of occurrence.

From these data, we can say that at $\text{pH} < 4$ there is strong adsorption of the amyloid peptide on the negatively charged surface.

However, it should be noted that the system with the amyloid peptide in the presence of a charged surface was undersampled because we did not use any preferential sampling. Our data is therefore of lower quality, and in the future it will be necessary to focus on a way to better sample our system.

Appendix

Reference data for short peptides

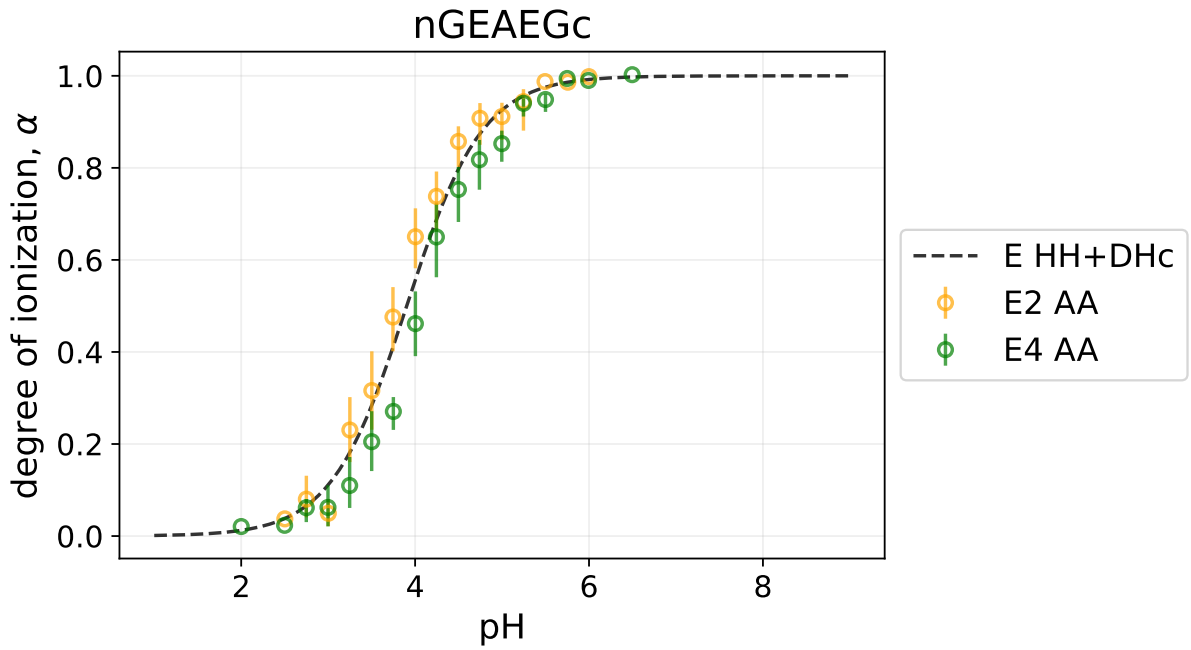


Figure 7.1: All-atom data for nGEAEGc sequence from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

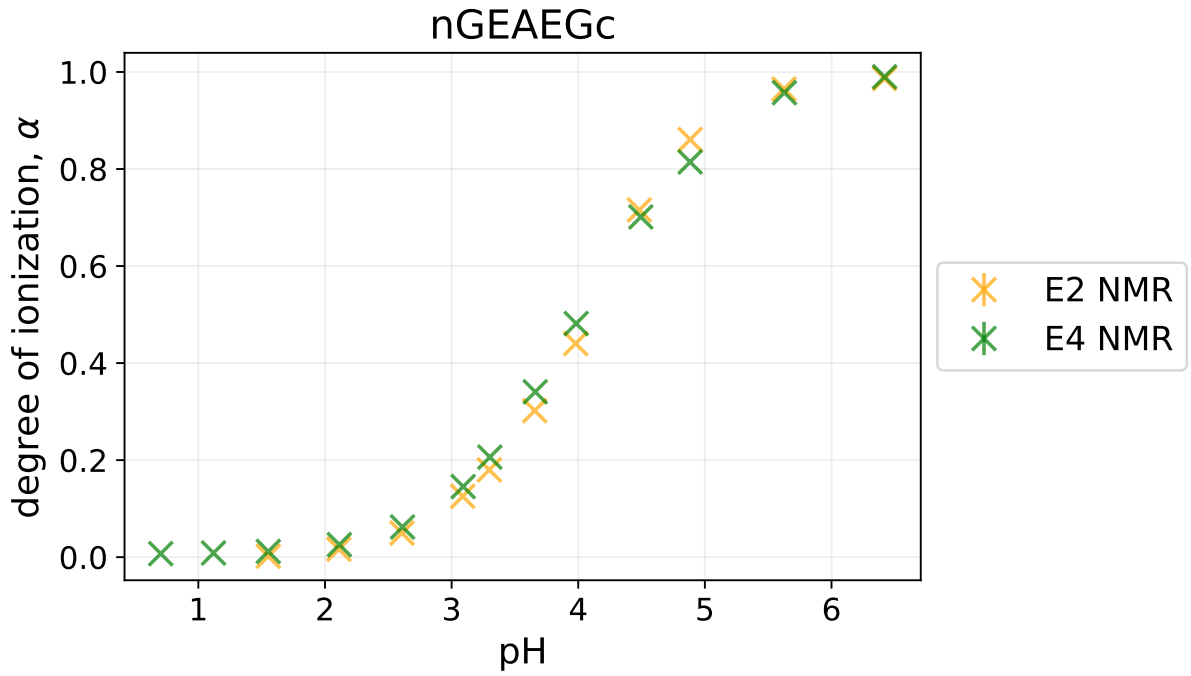


Figure 7.2: NMR data for nGEAEGc sequence from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

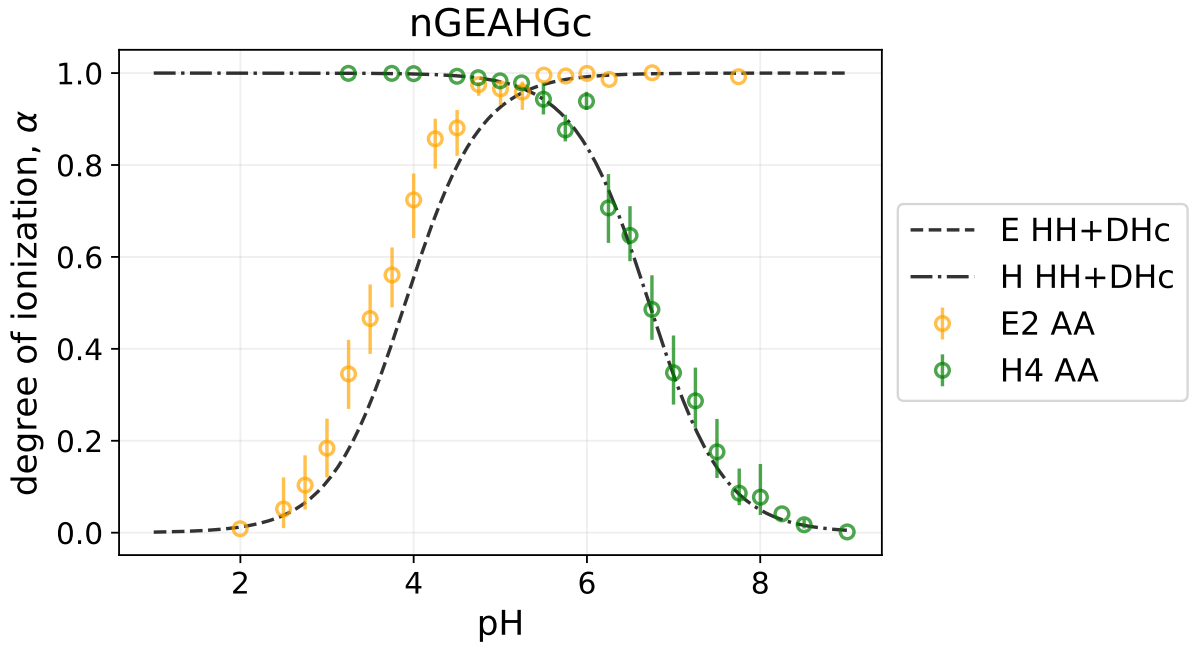


Figure 7.3: All-atom data for nGEAHGc sequence from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

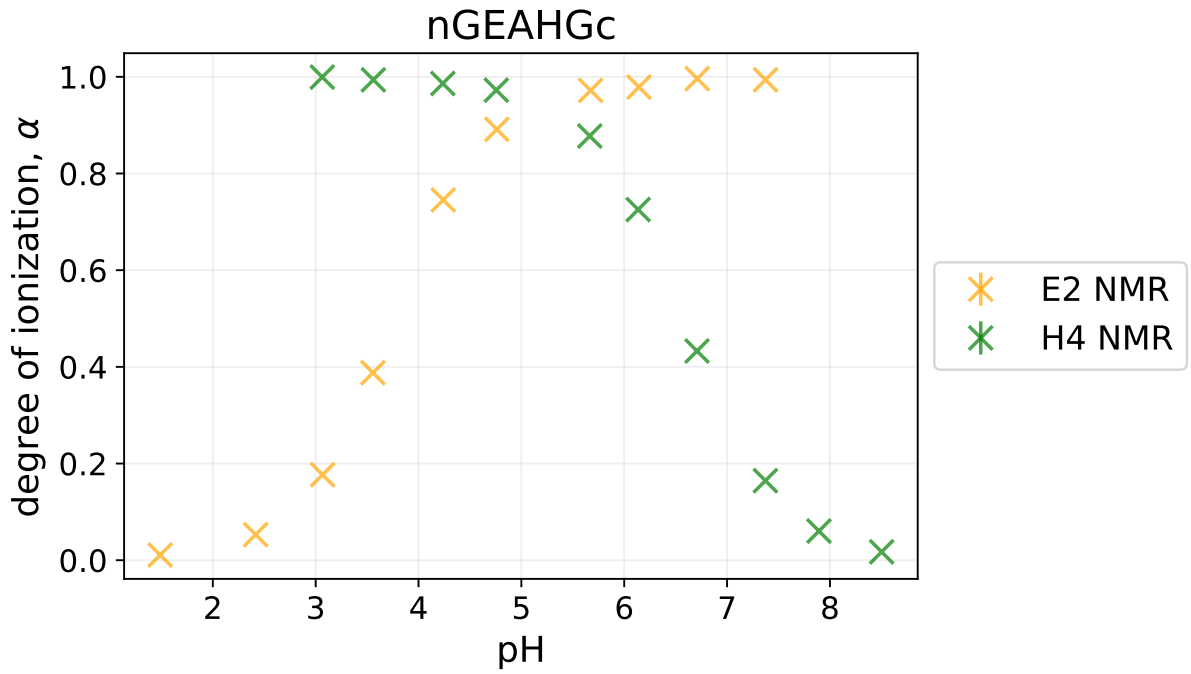


Figure 7.4: NMR data for nGEAHGc sequence from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

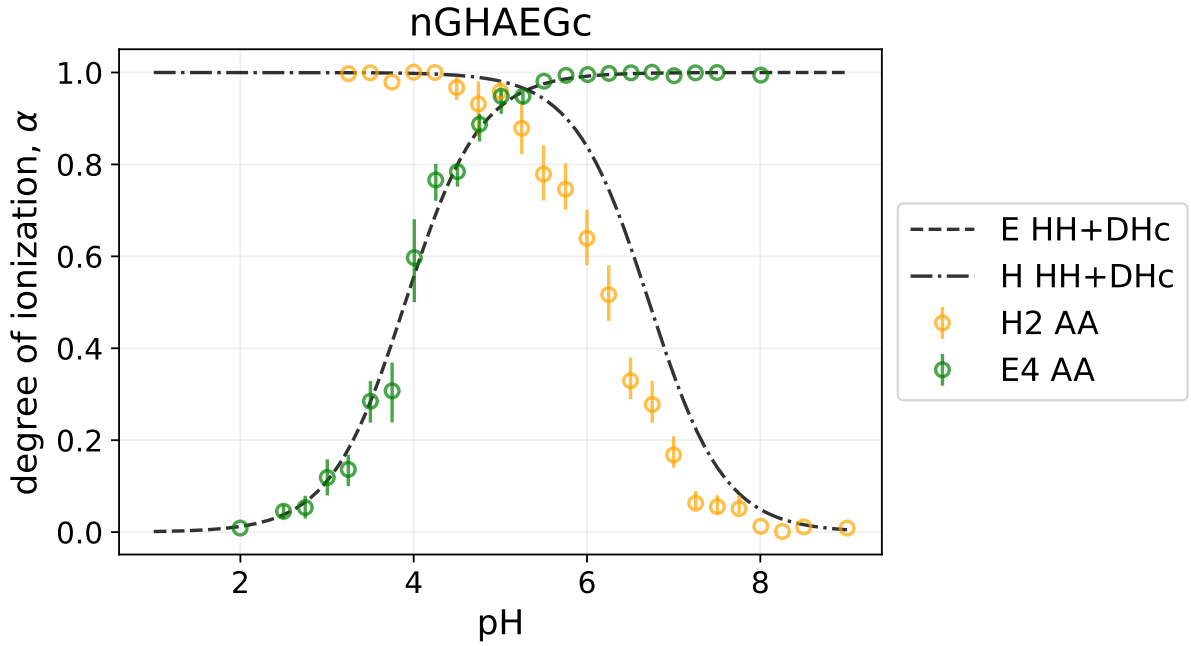


Figure 7.5: All-atom data for nGHAEGc sequence from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

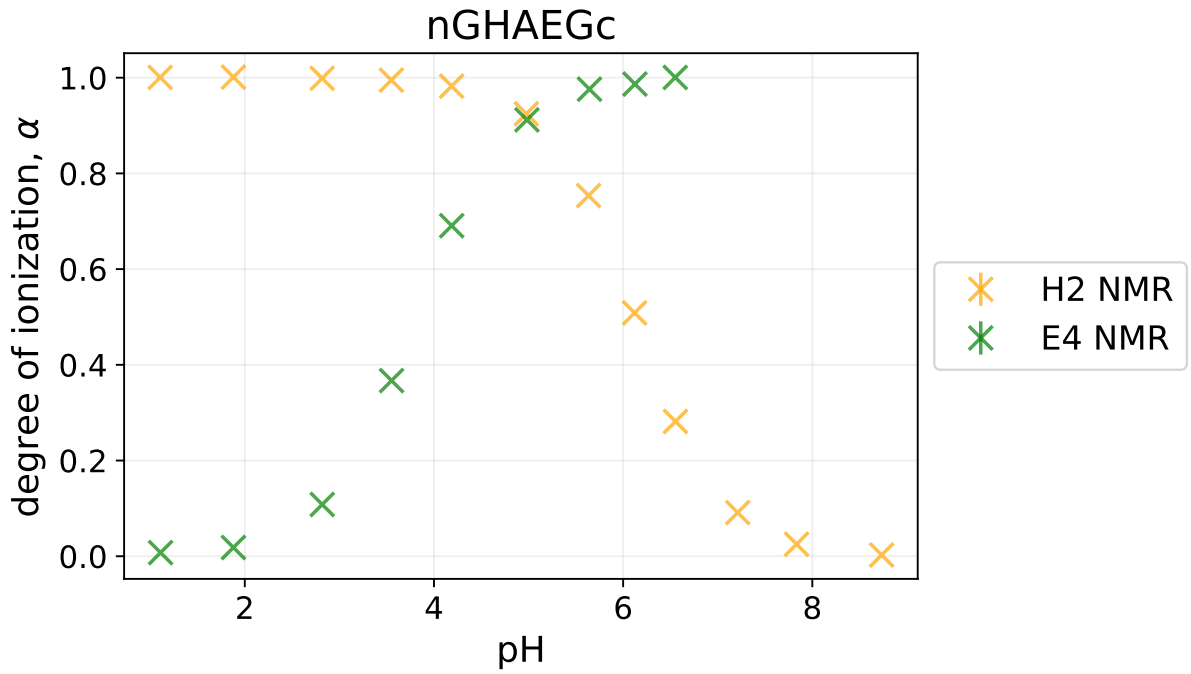


Figure 7.6: NMR data for nGHAEGc sequence from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

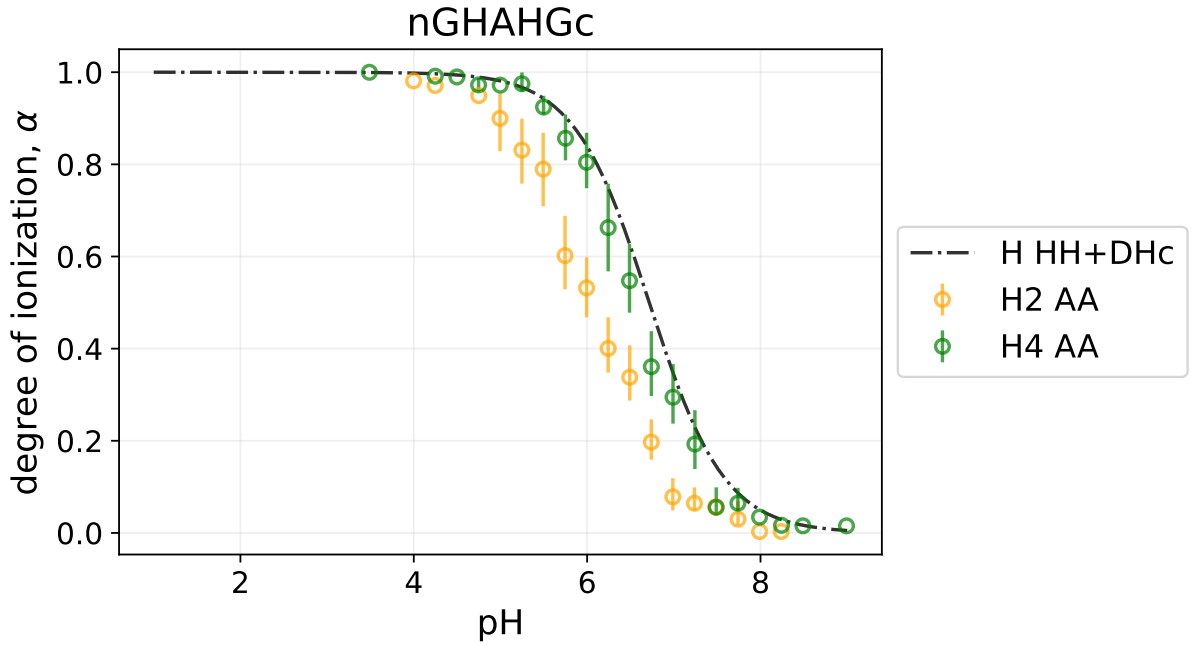


Figure 7.7: All-atom for nGHAHGc sequence data from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

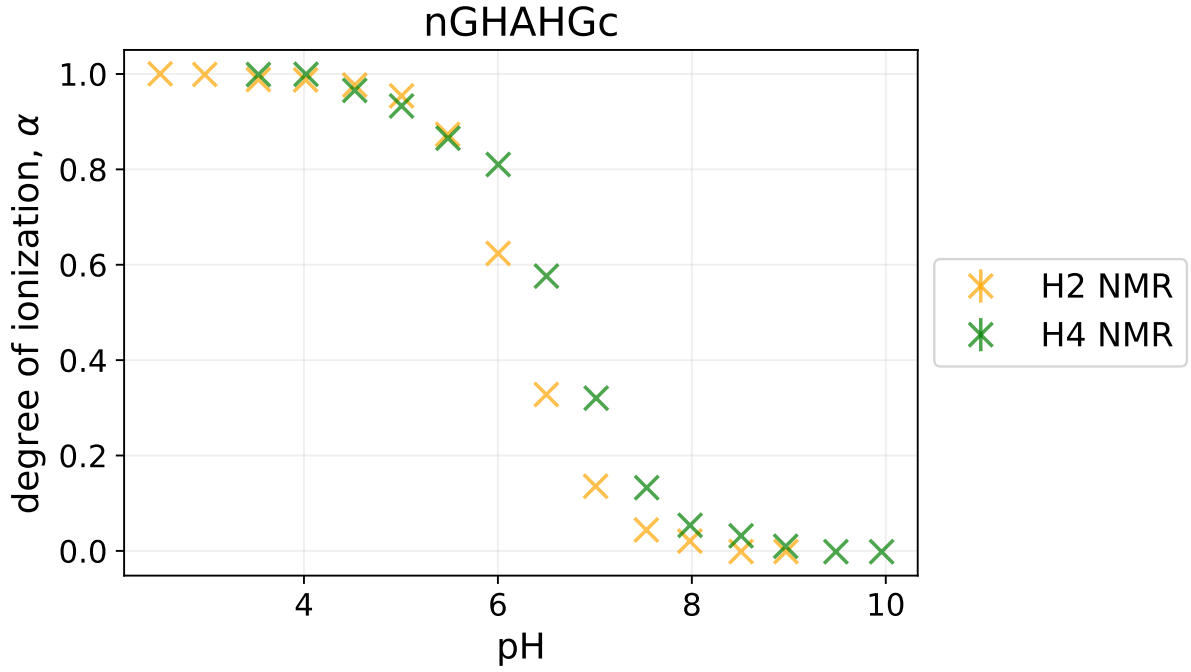


Figure 7.8: NMR data for nGHAHGc sequence from Dobrev et al.[8]. An explanation of the individual labels can be found in Figure 6.1.

Degree of ionization of short peptides

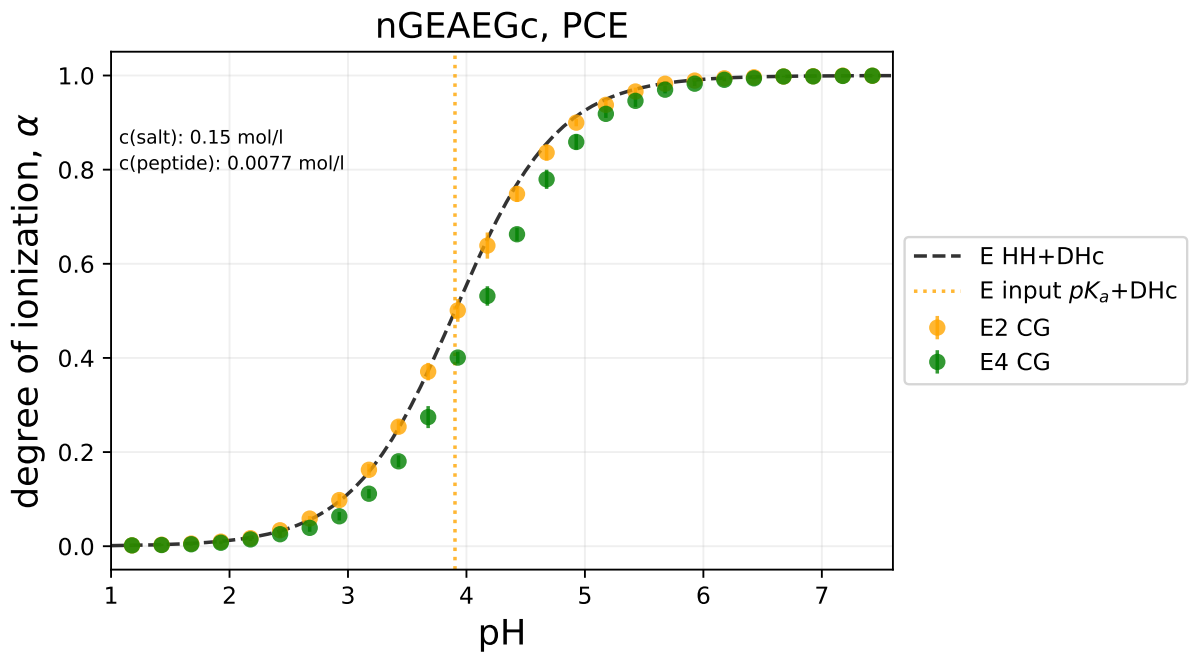


Figure 7.9: Degree of ionization for nGEAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figure 6.1.

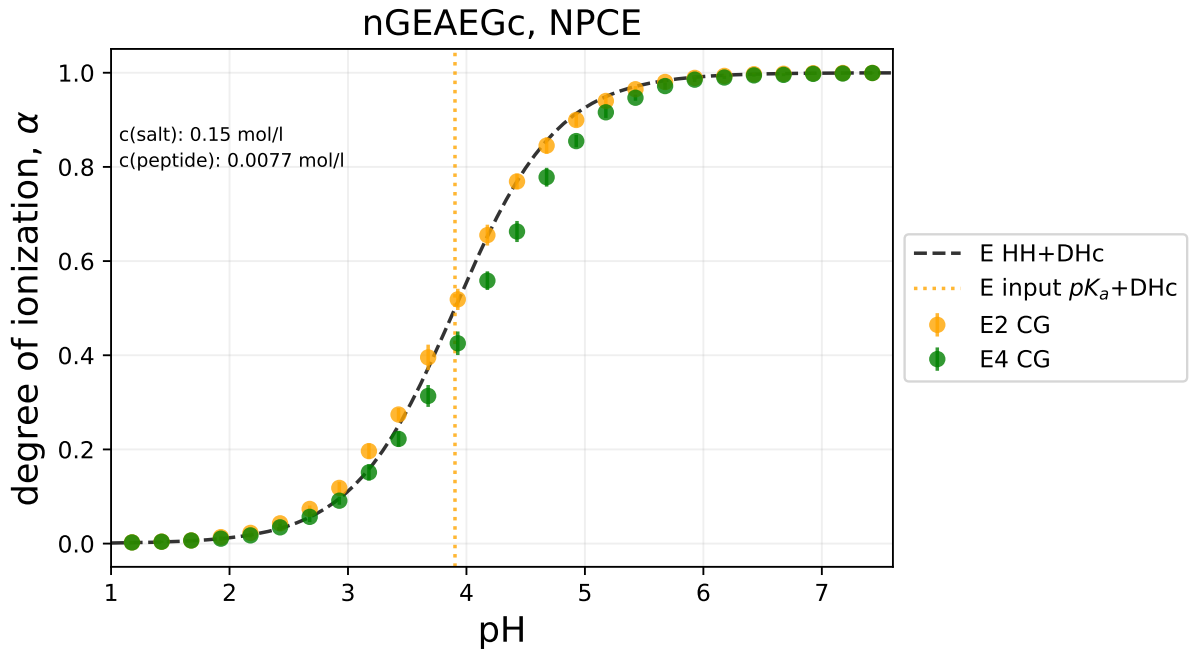


Figure 7.10: Degree of ionization for nGEAEGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figure 6.1.

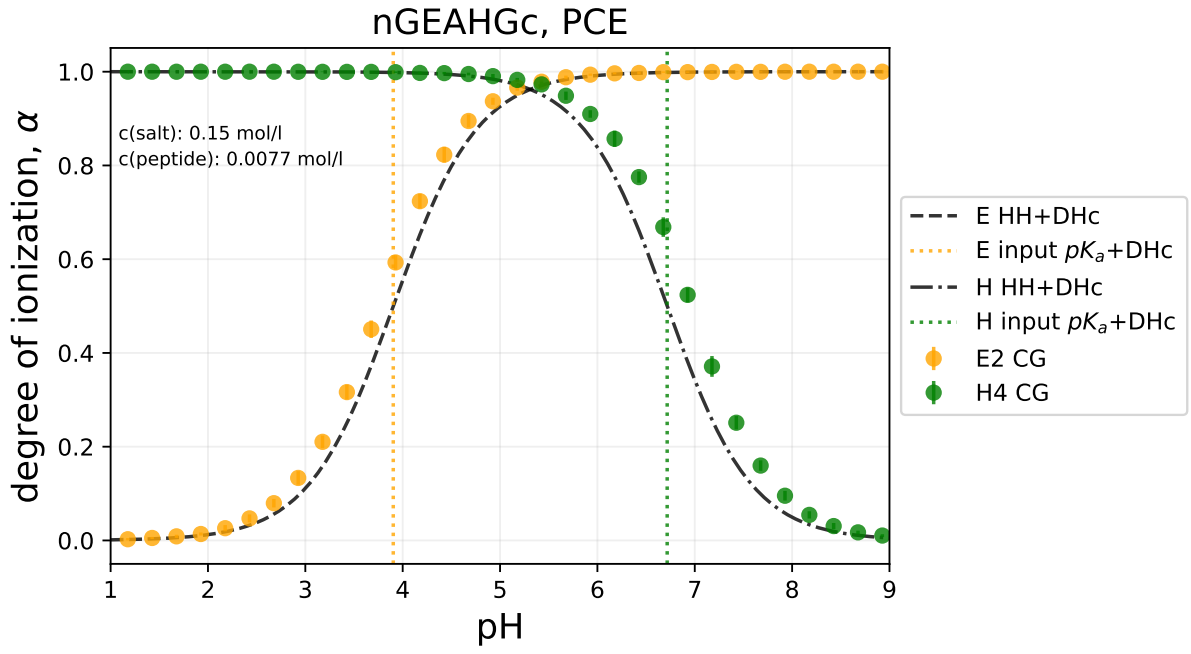


Figure 7.11: Degree of ionization for nGEAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figure 6.1.

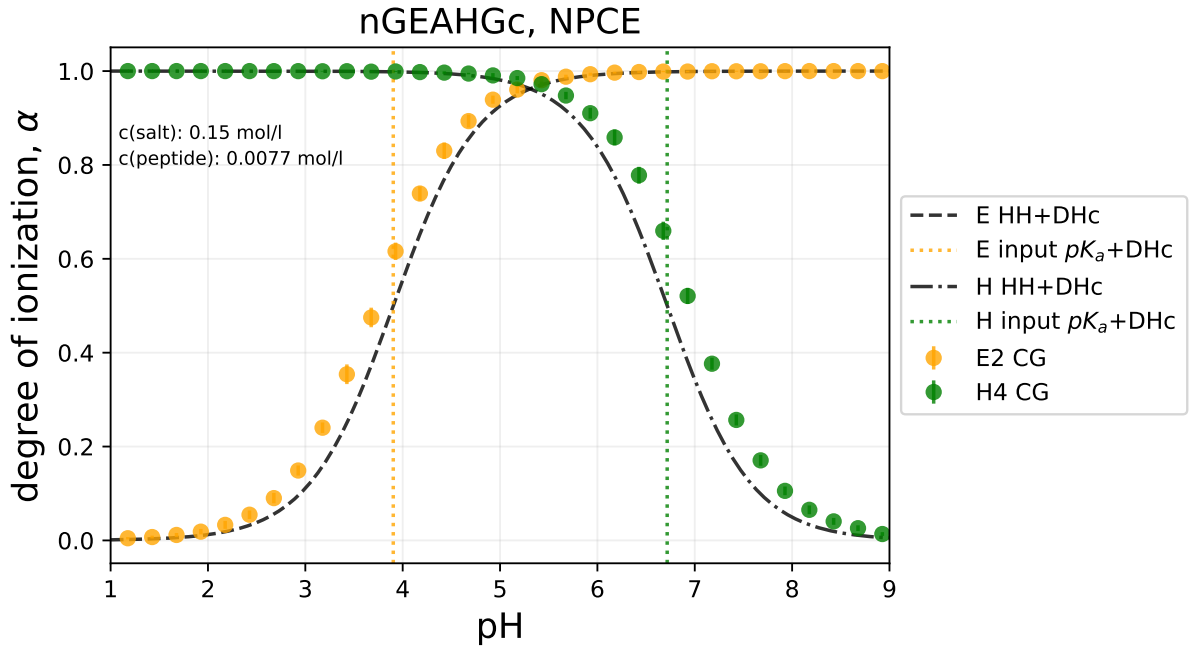


Figure 7.12: Degree of ionization for nGEAHGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figure 6.1.

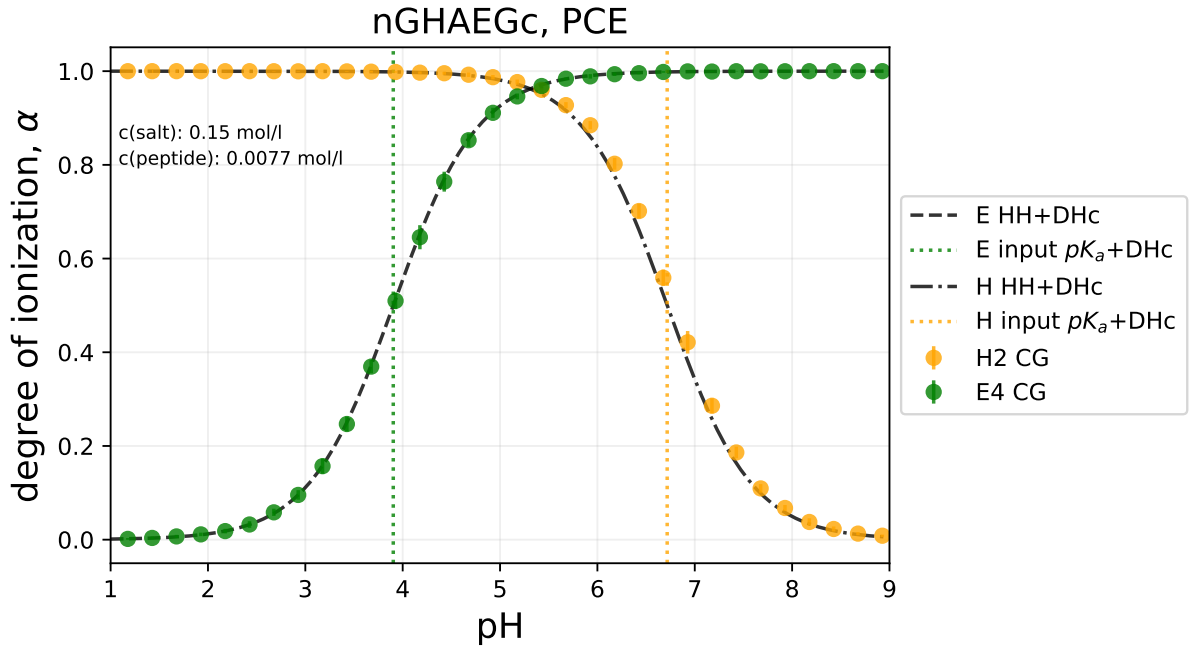


Figure 7.13: Degree of ionization for nGHAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figure 6.1.

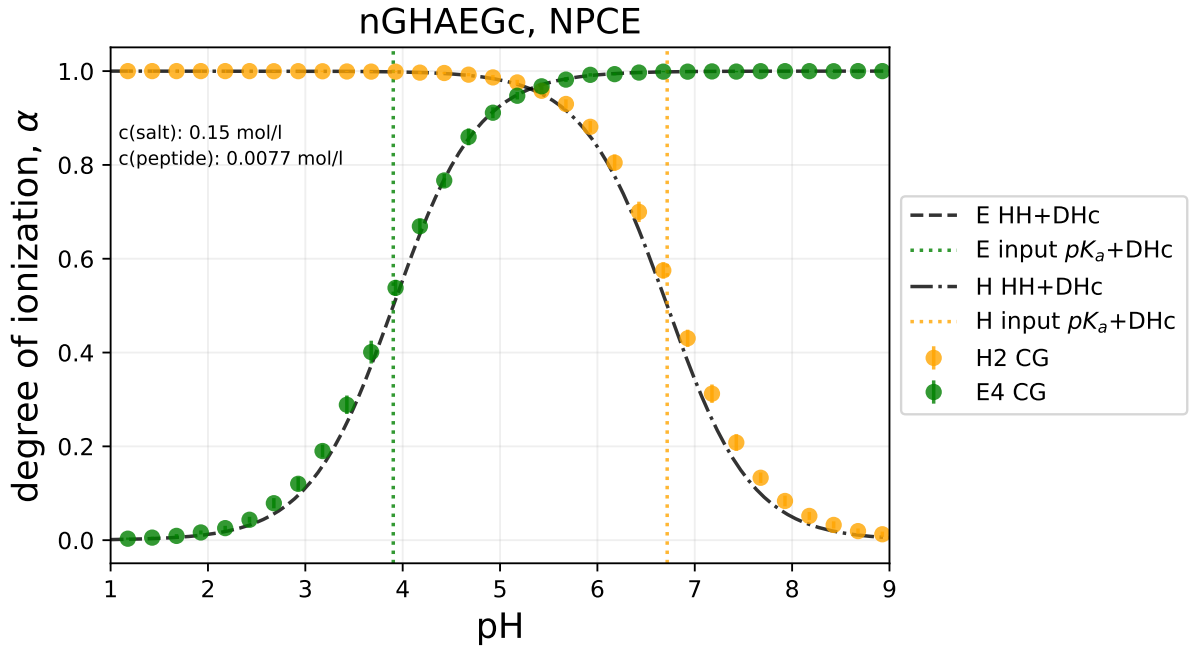


Figure 7.14: Degree of ionization for nGHAEGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figure 6.1.

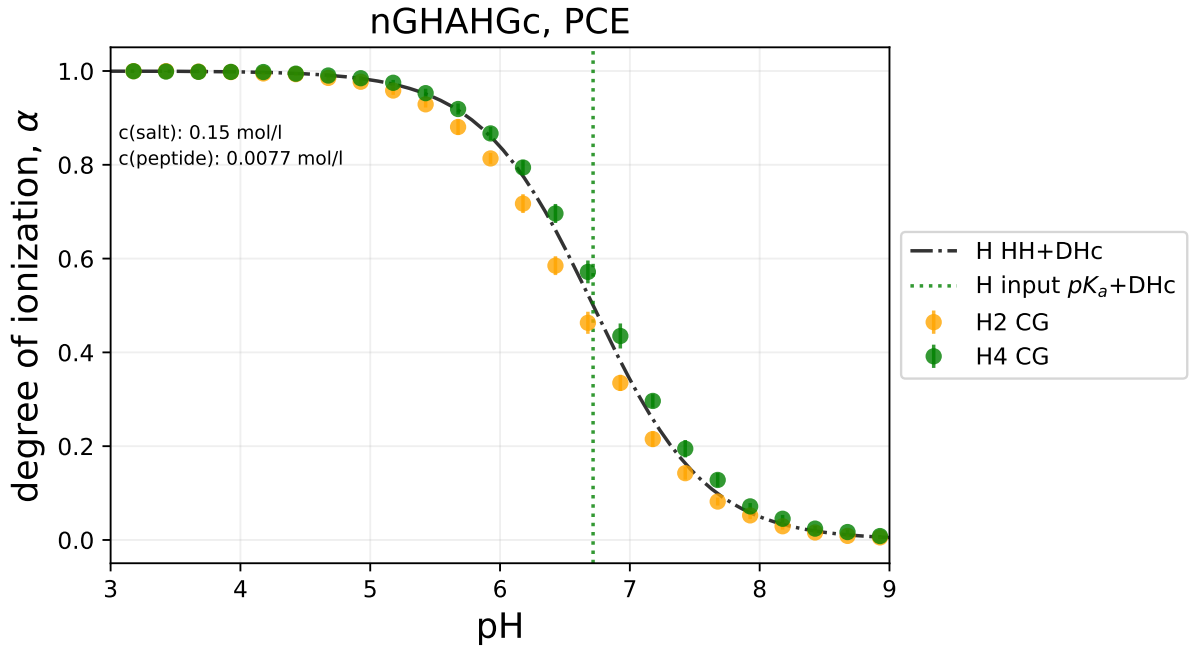


Figure 7.15: Degree of ionization for nGHAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figure 6.1.

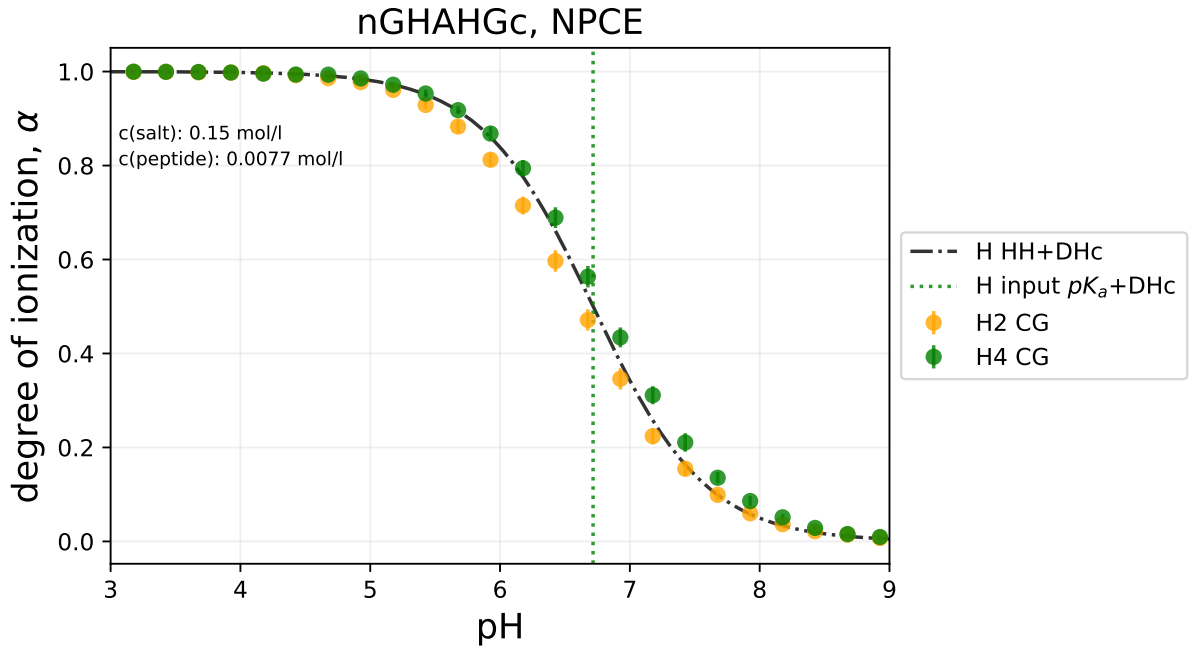


Figure 7.16: Degree of ionization for nGHAHGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figure 6.1.

Fits of degree of ionization of short peptides

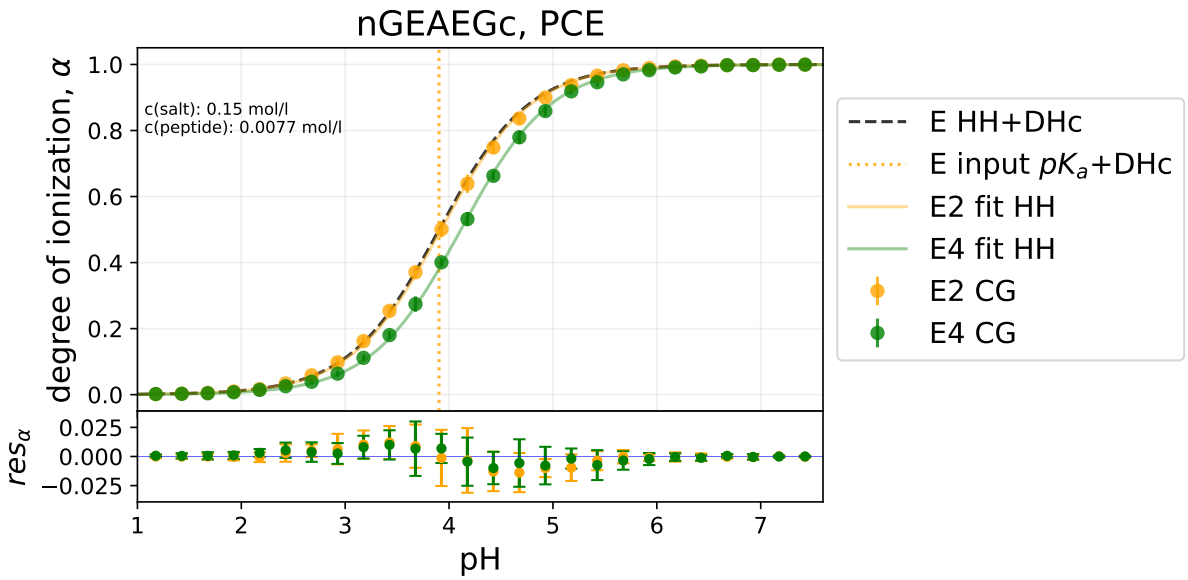


Figure 7.17: Fit using Henderson-Hasselbalch equation of degree of ionization for nGEAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

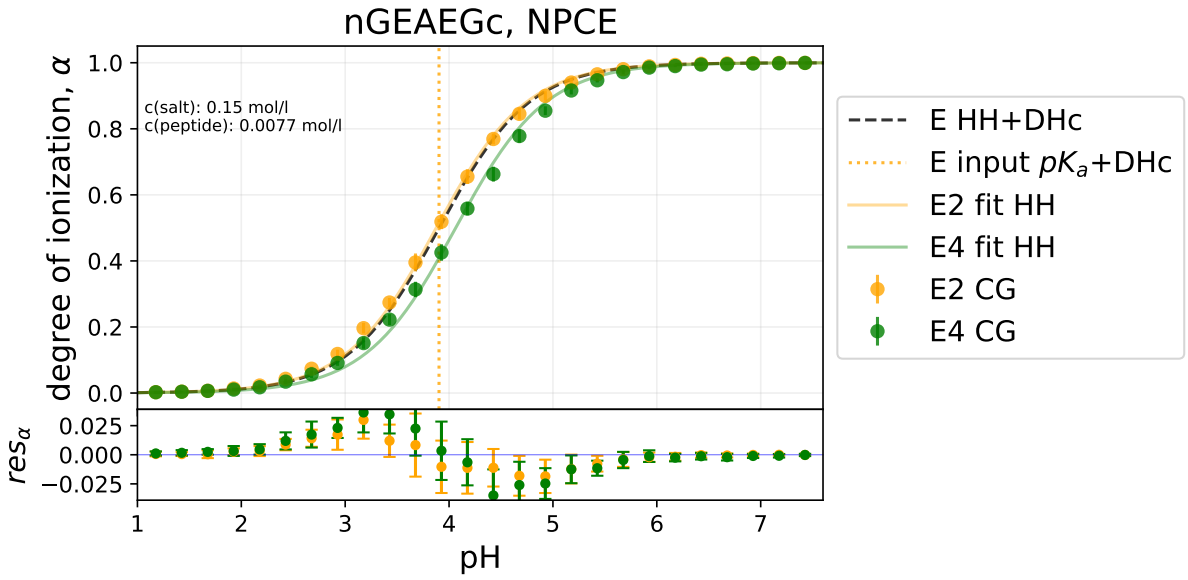


Figure 7.18: Fit using Henderson-Hasselbalch equation of degree of ionization for nGEAEGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

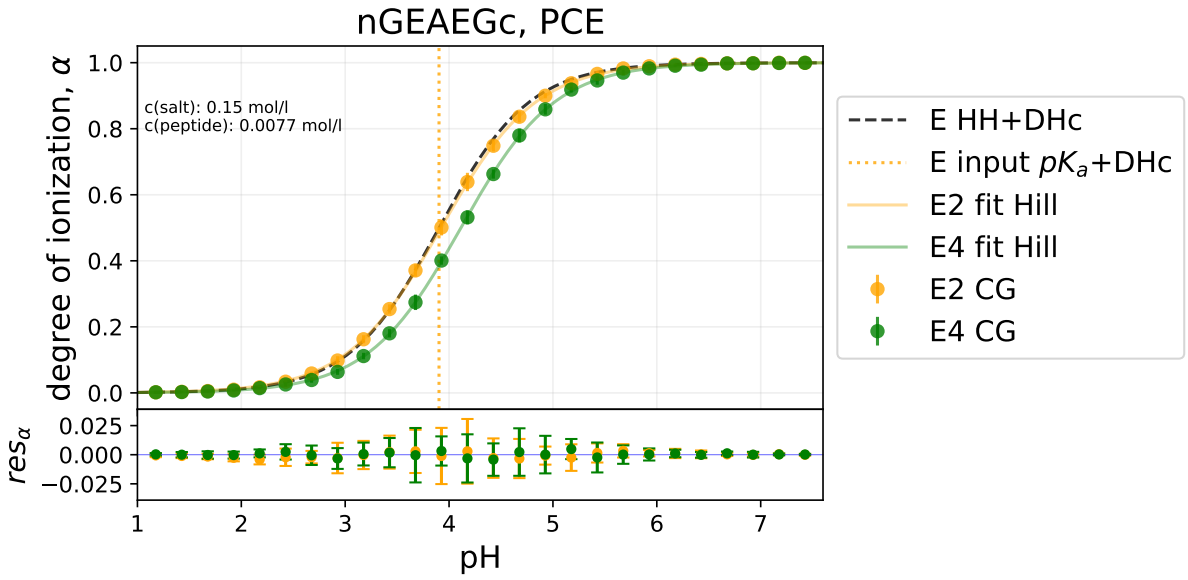


Figure 7.19: Fit using Hill equation of degree of ionization for nGEAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

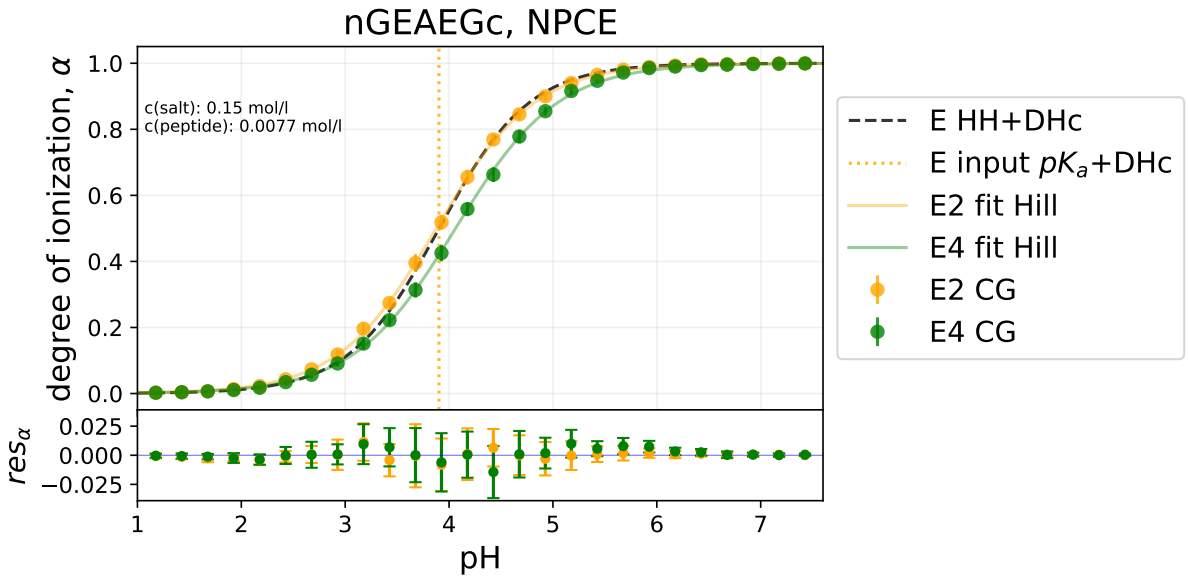


Figure 7.20: Fit using Hill equation of degree of ionization for nGEAEGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

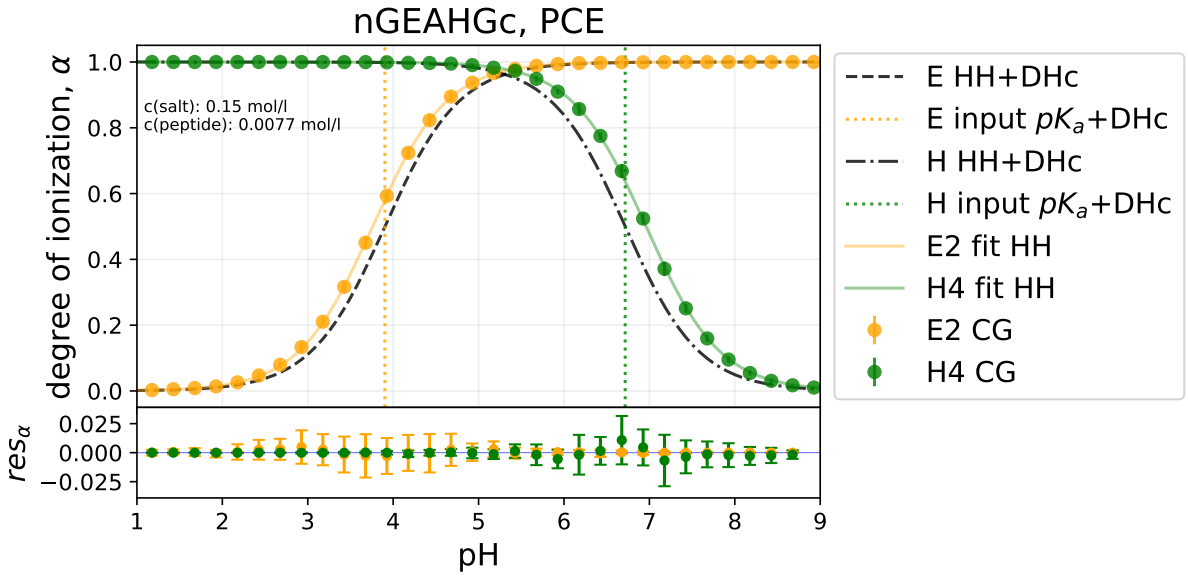


Figure 7.21: Fit using Henderson-Hasselbalch equation of degree of ionization for nGEAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

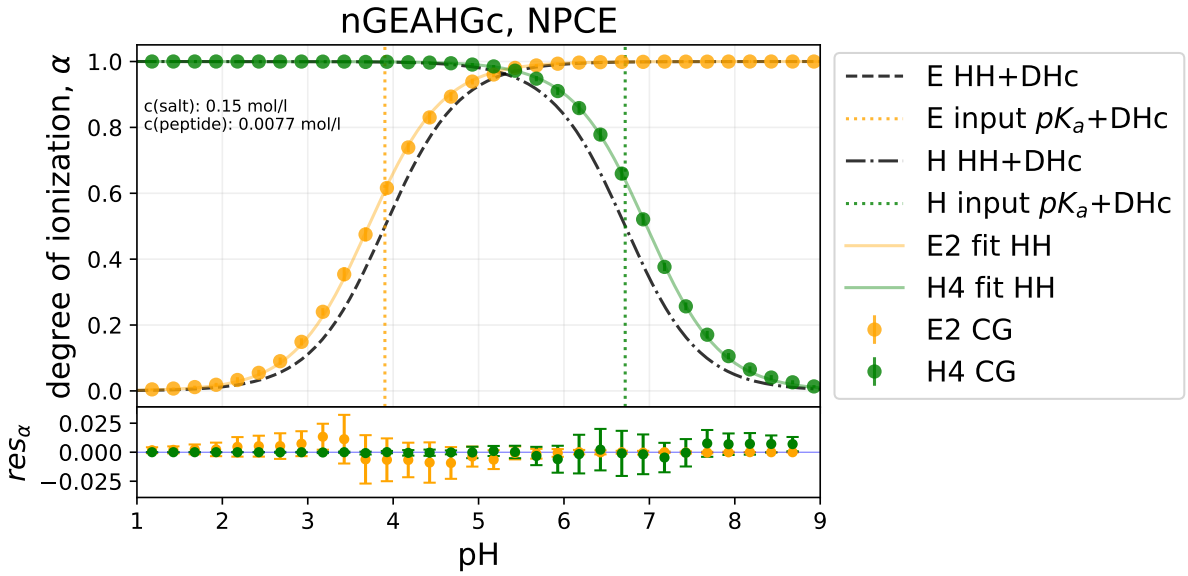


Figure 7.22: Fit using Henderson-Hasselbalch equation of degree of ionization for nGEAHGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

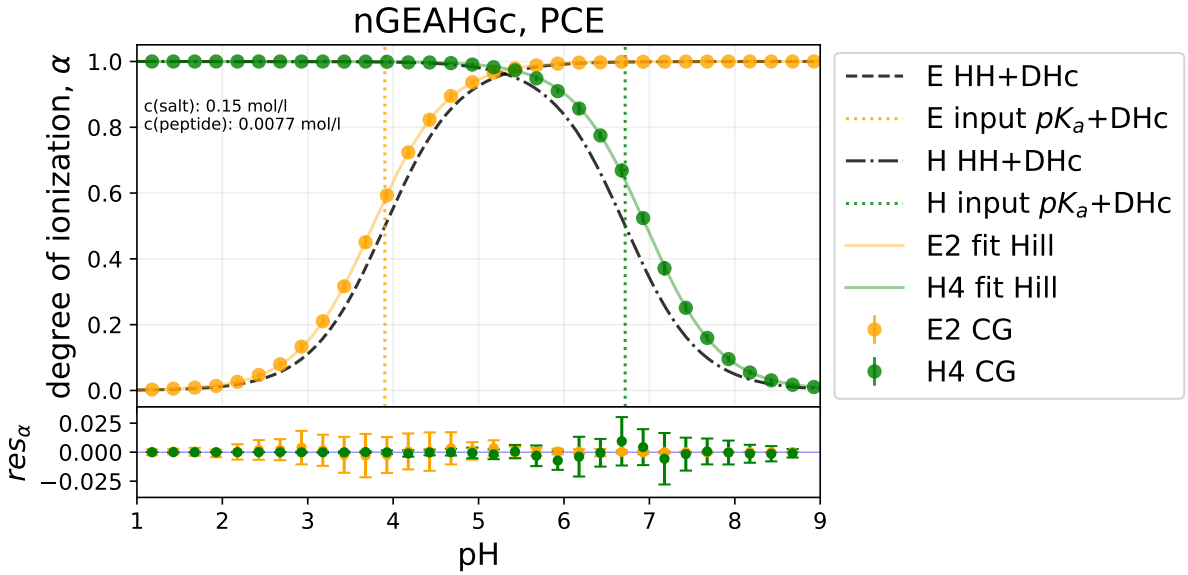


Figure 7.23: Fit using Hill equation of degree of ionization for nGEAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

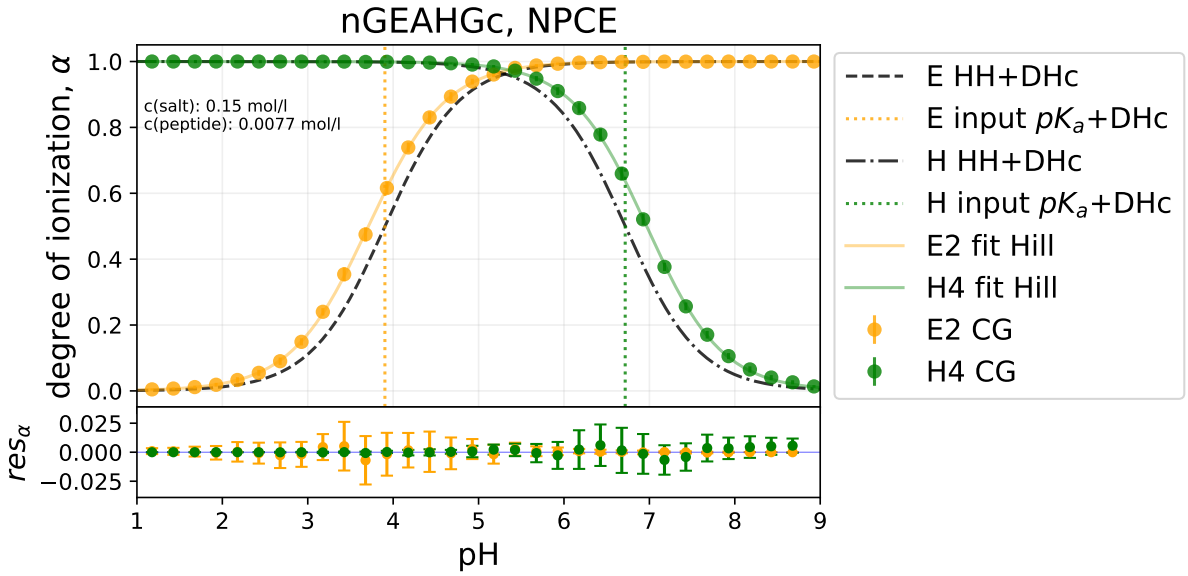


Figure 7.24: Fit using Hill equation of degree of ionization for nGEAHGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

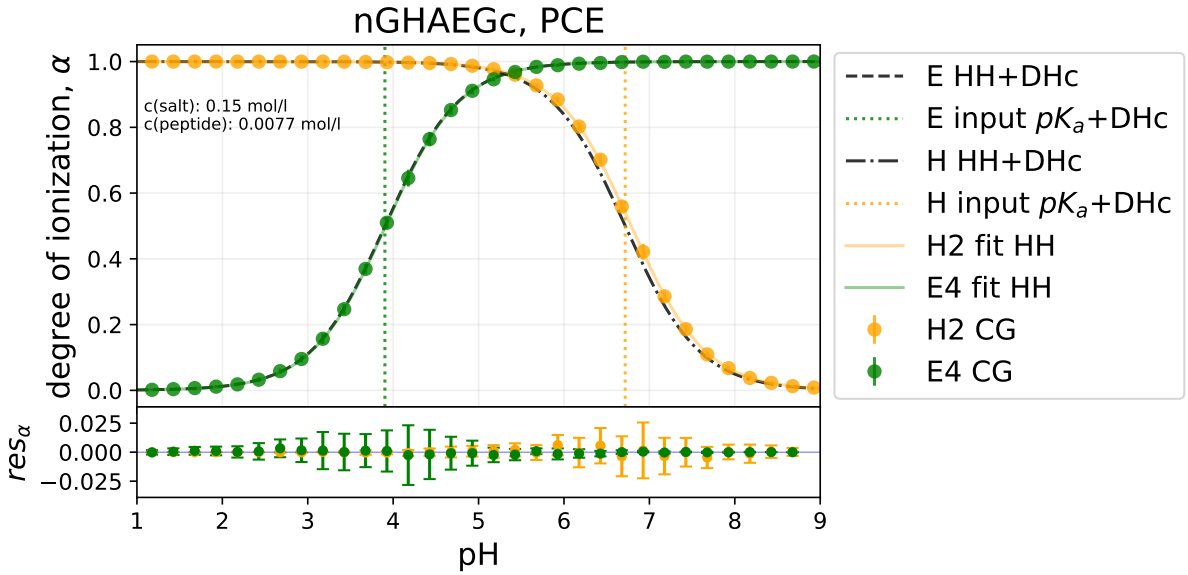


Figure 7.25: Fit using Henderson-Hasselbalch equation of degree of ionization for nGHAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

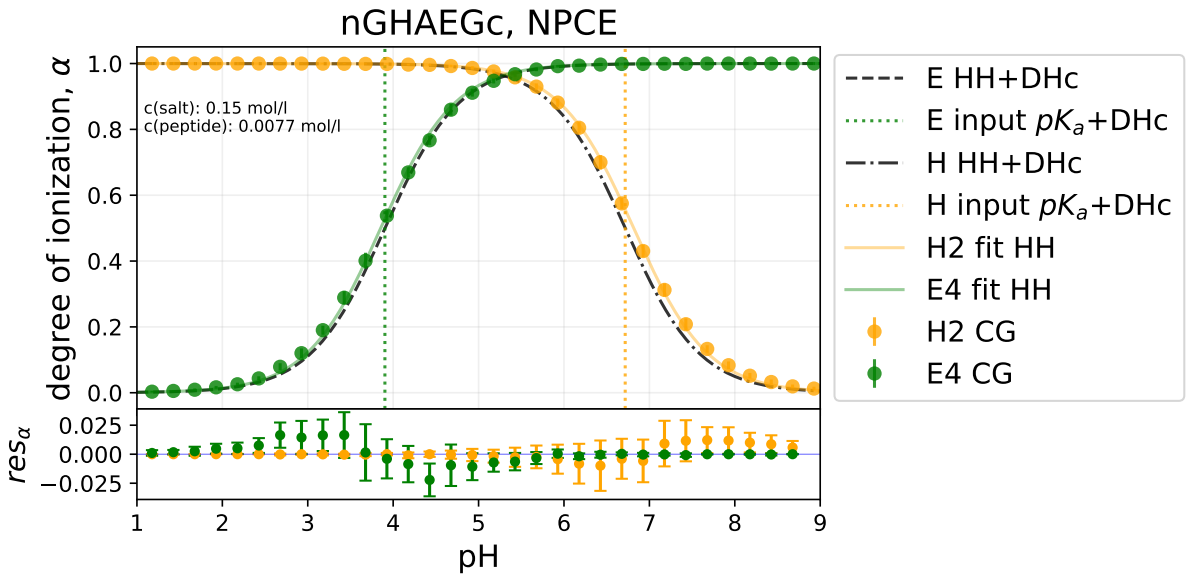


Figure 7.26: Fit using Henderson-Hasselbalch equation of degree of ionization for nGHAEGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

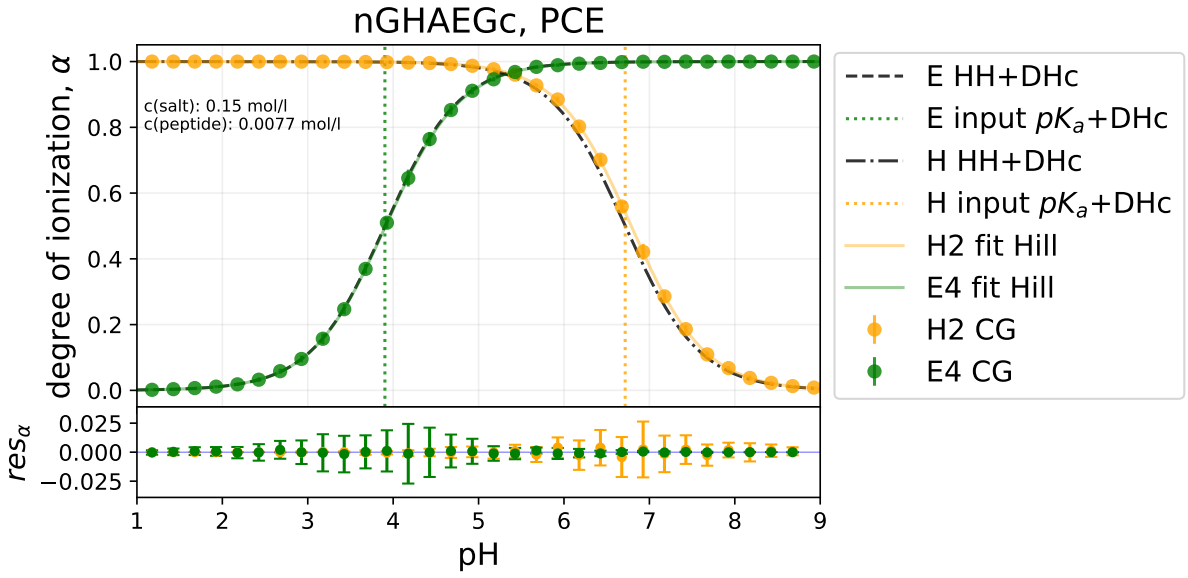


Figure 7.27: Fit using Hill equation of degree of ionization for nGHAEGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

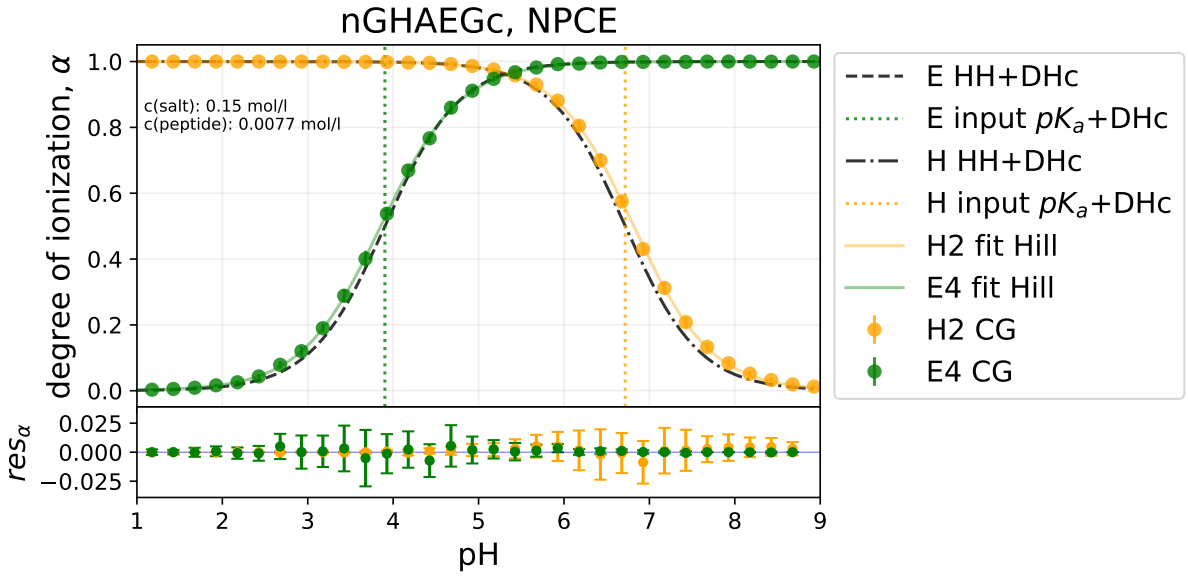


Figure 7.28: Fit using Hill equation of degree of ionization for nGHAEGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

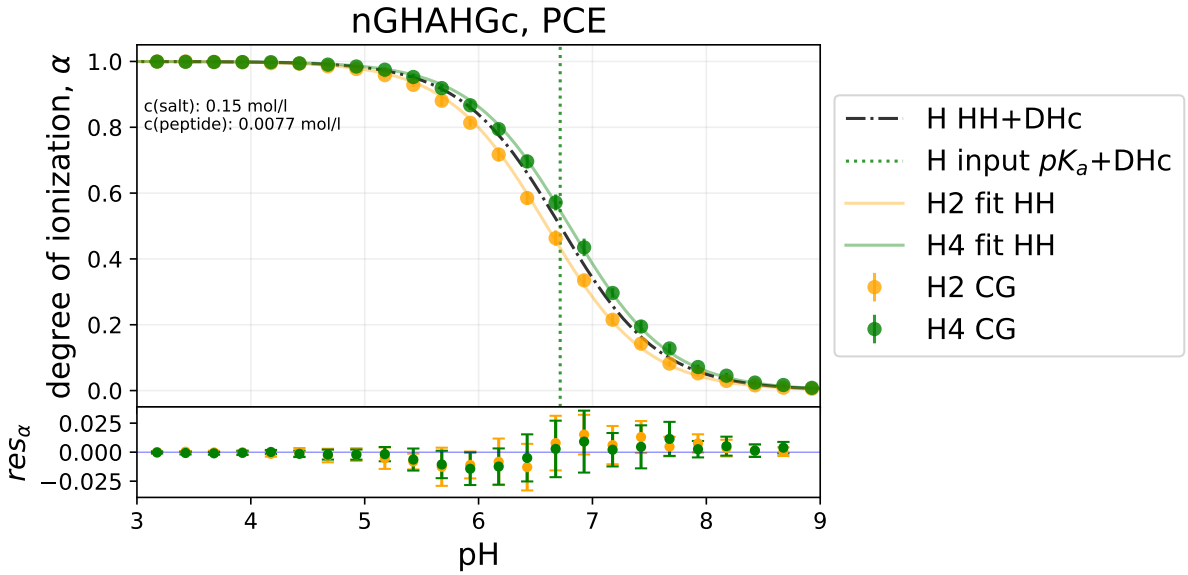


Figure 7.29: Fit using Henderson-Hasselbalch equation of degree of ionization for nGHAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

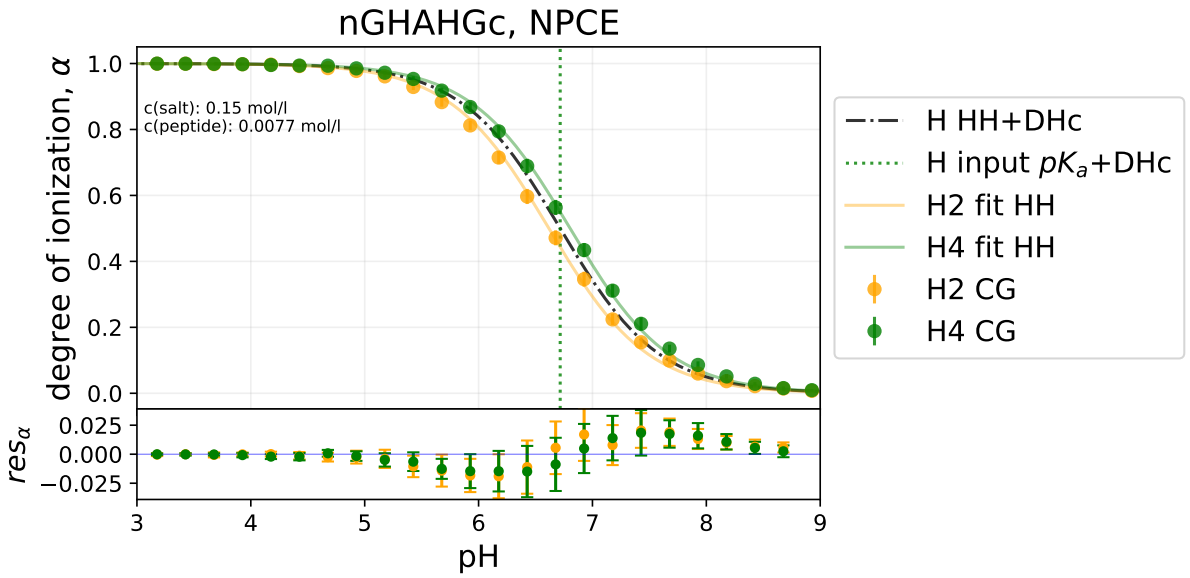


Figure 7.30: Fit using Henderson-Hasselbalch equation of degree of ionization for nGHAHGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

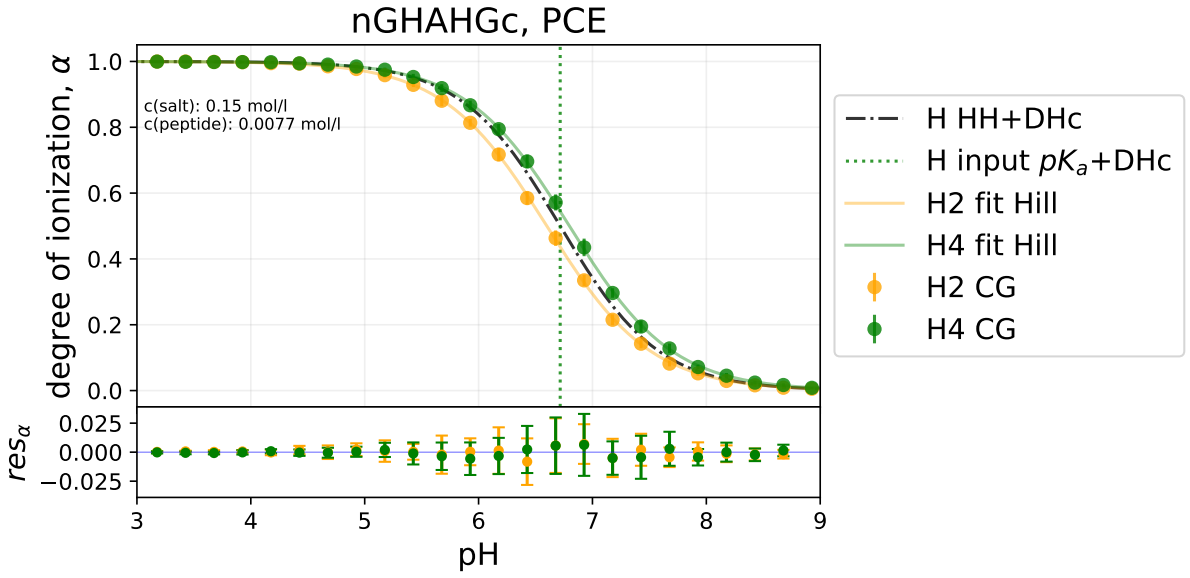


Figure 7.31: Fit using Hill equation of degree of ionization for nGHAHGc sequence from CG simulation, PCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

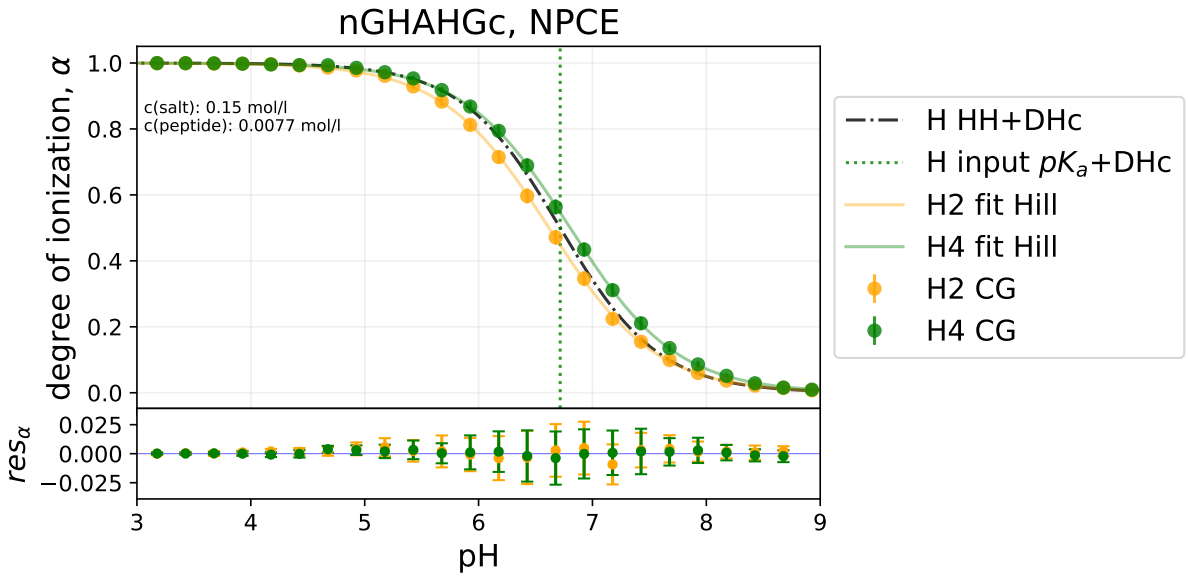


Figure 7.32: Fit using Hill equation of degree of ionization for nGHAHGc sequence from CG simulation, NPCE. An explanation of the individual labels can be found in Figures 6.1 and 6.3.

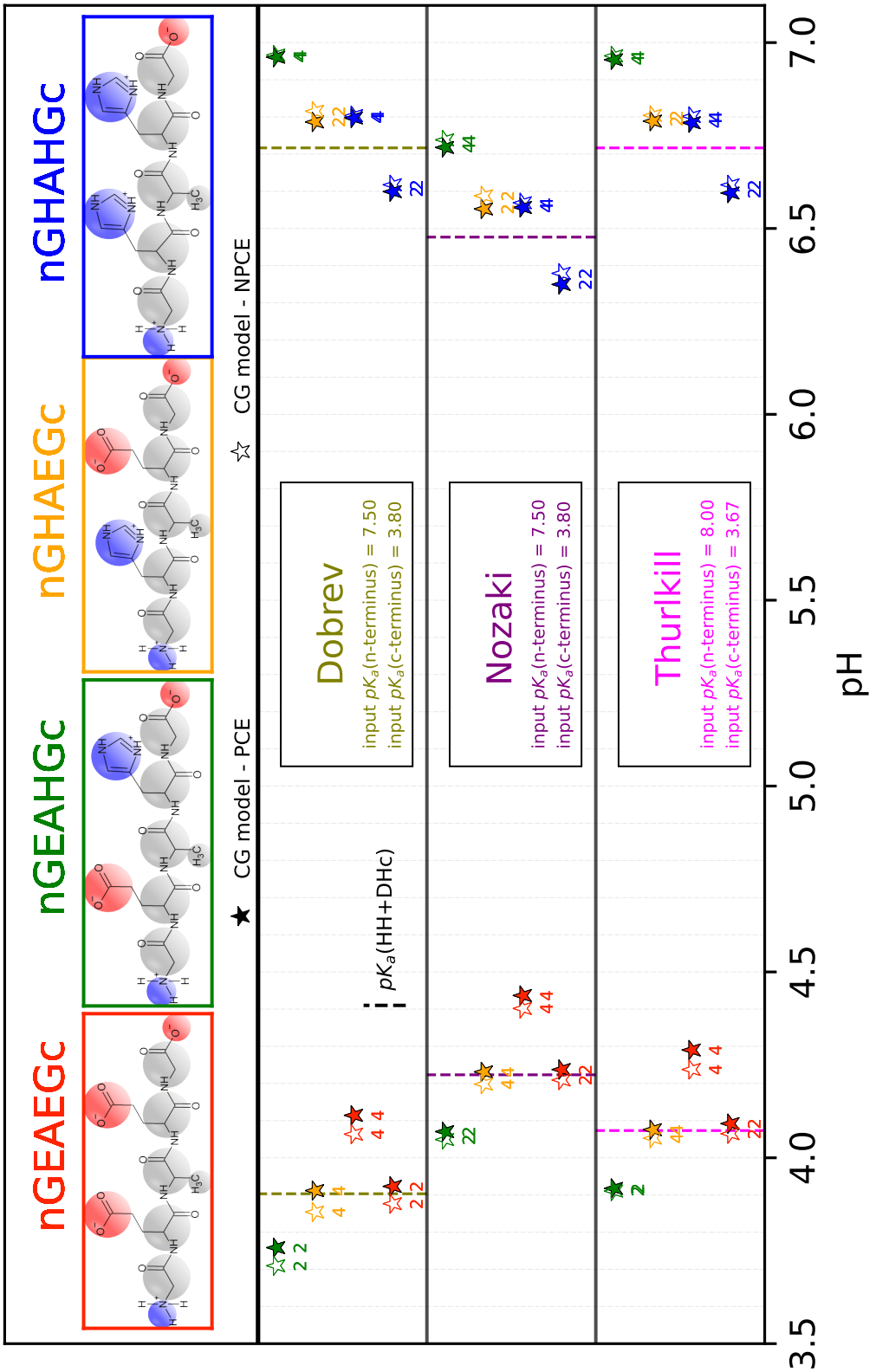


Figure 7.33: Various pKa sets. An explanation of the individual labels can be found in Figures 6.9.

Bibliography

[1] Johannes N. Brönsted. "Einige Bemerkungen über den Begriff der Säuren und Basen". In: *Recueil des Travaux Chimiques des Pays-Bas* 42.8 (1923), pp. 718–728. DOI: https://doi.org/10.1002/recl.19230420815. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/recl.19230420815.

[2] Lawrence J. Henderson. "CONCERNING THE RELATIONSHIP BETWEEN THE STRENGTH OF ACIDS AND THEIR CAPACITY TO PRESERVE NEUTRALITY". In: *American Journal of Physiology-Legacy Content* 21.2 (1908), pp. 173–179. DOI: 10.1152/ajplegacy.1908.21.2.173.

[3] Peter Debye and Erich Hückel. "Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen". In: *Physikalische Zeitschrift* 24.185 (1923), p. 305.

[4] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (June 1953), pp. 1087–1092. ISSN: 0021-9606. DOI: 10.1063/1.1699114. eprint: https://pubs.aip.org/aip/jcp/article-pdf/21/6/1087/18802390/1087_1_online.pdf. URL: https://doi.org/10.1063/1.1699114.

[5] Wilfred K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970), pp. 97–109. ISSN: 00063444. (Visited on 05/15/2024).

[6] Berni J. Alder and Thomas E. Wainwright. "Studies in Molecular Dynamics. I. General Method". In: *The Journal of Chemical Physics* 31.2 (Aug. 1959), pp. 459–466. ISSN: 0021-9606. DOI: 10.1063/1.1730376. eprint: https://pubs.aip.org/aip/jcp/article-pdf/31/2/459/18817177/459_1_online.pdf.

[7] Loup Verlet. "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Phys. Rev.* 159 (1 July 1967), pp. 98–103. DOI: 10.1103/PhysRev.159.98.

[8] Plamen Dobrev et al. "Probing the Accuracy of Explicit Solvent Constant pH Molecular Dynamics Simulations for Peptides". In: *Journal of Chemical Theory and Computation* 16.4 (2020). PMID: 32192342, pp. 2561–2569. DOI: 10.1021/acs.jctc.9b01232. eprint: https://doi.org/10.1021/acs.jctc.9b01232.

[9] Kamil Tamiola et al. "pepKalc: scalable and comprehensive calculation of electrostatic interactions in random coil polypeptides". In: *Bioinformatics* 34.12 (Jan. 2018), pp. 2053–2060. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty033. eprint: https://academic.oup.com/bioinformatics/article-pdf/34/12/2053/48936007/bioinformatics_34_12_2053.pdf.

[10] Jonas Landsgesell et al. "Simulations of ionization equilibria in weak polyelectrolyte solutions and gels". In: *Soft Matter* 15 (6 2019), pp. 1155–1185. DOI: 10.1039/C8SM02085J.

[11] Lucie Nová, Filip Uhlík, and Peter Košovan. "Local pH and effective pKA of weak polyelectrolytes – insights from computer simulations". In: *Phys. Chem. Chem. Phys.* 19 (22 2017), pp. 14376–14387. DOI: 10.1039/C7CP00265C.

[12] Raju Lunkad et al. "Quantitative prediction of charge regulation in oligopeptides". In: *Mol. Syst. Des. Eng.* 6 (2 2021), pp. 122–131. DOI: 10.1039/D0ME00147C.

[13] Roman Staňo et al. "Multivalent counterions accumulate in star-like polyelectrolytes and collapse the polymer in spite of increasing its ionization". In: *Soft Matter* 16 (4 2020), pp. 1047–1055. DOI: 10.1039/C9SM02318F.

[14] David Beyer et al. *pyMBE: the Python-based Molecule Builder for ESPResSo*. 2024. arXiv: 2401.14954 [cond-mat.soft].

[15] Sebastian P. Pineda et al. "Charge Regulation Triggers Condensation of Short Oligopeptides to Polyelectrolytes". In: *JACS Au* (2024). DOI: 10.1021/jacsau.3c00668. eprint: https://doi.org/10.1021/jacsau.3c00668.

[16] Orlando Crescenzi et al. "Solution structure of the Alzheimer amyloid β-peptide (1–42) in an apolar microenvironment". In: *European Journal of Biochemistry* 269.22 (2002), pp. 5642–5648. DOI: https://doi.org/10.1046/j.1432-1033.2002.03271.x. eprint: https://febs.onlinelibrary.wiley.com/doi/pdf/10.1046/j.1432-1033.2002.03271.x.

[17] Marina D Kirkitadze, Margaret M Condron, and David B Teplow. "Identification and characterization of key kinetic intermediates in amyloid β-protein fibrillogenesis11Edited by F. Cohen". In: *Journal of Molecular Biology* 312.5 (2001), pp. 1103–1119. ISSN: 0022-2836. DOI: https://doi.org/10.1006/jmbi.2001.4970.

[18] Shigeki Kobayashi et al. "Dependence pH and proposed mechanism for aggregation of Alzheimer's disease-related amyloid-β(1–42) protein". In: *Journal of Molecular Structure* 1094 (2015), pp. 109–117. ISSN: 0022-2860. DOI: https://doi.org/10.1016/j.molstruc.2015.03.023.

[19] Celia Cabaleiro-Lago et al. "Dual Effect of Amino Modified Polystyrene Nanoparticles on Amyloid β Protein Fibrillation". In: *ACS Chemical Neuroscience* 1.4 (2010). PMID: 22778827, pp. 279–287. DOI: 10.1021/cn900027u. eprint: https://doi.org/10.1021/cn900027u.

[20] Celia Cabaleiro-Lago et al. "Inhibition of Amyloid β Protein Fibrillation by Polymeric Nanoparticles". In: *Journal of the American Chemical Society* 130.46 (2008). PMID: 18954050, pp. 15437–15443. DOI: 10.1021/ja8041806. eprint: https://doi.org/10.1021/ja8041806.

[21] Sara Linse et al. "Nucleation of protein fibrillation by nanoparticles". In: *Proceedings of the National Academy of Sciences* 104.21 (2007), pp. 8691–8696. DOI: 10.1073/pnas.0701250104. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.0701250104.

[22] Robert M. Garrels and Charles L. Christ. *Solutions, Minerals and Equilibria*. San Francisco: Freeman, Cooper and Company, 1965.

[23] Brett H. Pogostin et al. “pKa Determination of a Histidine Residue in a Short Peptide Using Raman Spectroscopy”. In: *Molecules* 24.3 (2019). ISSN: 1420-3049. DOI: 10.3390/molecules24030405.

[24] Pablo M. Blanco et al. “Influence of macromolecular crowding on the charge regulation of intrinsically disordered proteins”. In: *Soft Matter* 17 (3 2021), pp. 655–669. DOI: 10.1039/D0SM01475C.

[25] Shengsheng Zhang et al. “The Alzheimer’s Peptide Aβ Adopts a Collapsed Coil Structure in Water”. In: *Journal of Structural Biology* 130.2 (2000), pp. 130–141. ISSN: 1047-8477. DOI: https://doi.org/10.1006/jsbi.2000.4288.

[26] Luca Monticelli and Emppu Salonen. *Biomolecular Simulations: Methods and Protocols*. Vol. 924. Jan. 2013, pp. 533–565. ISBN: 978-1-62703-016-8. DOI: 10.1007/978-1-62703-017-5.

[27] Raju Lunkad et al. “Role of pK_A in Charge Regulation and Conformation of Various Peptide Sequences”. In: *Polymers* 13.2 (2021). ISSN: 2073-4360.

[28] Avogadro Chemistry. *Avogadro*. Version 1.93.0. URL: http://avogadro.cc/.

[29] Paul P. Ewald. “Die Berechnung optischer und elektrostatischer Gitterpotentiale”. In: *Annalen der Physik* 369.3 (1921), pp. 253–287. DOI: https://doi.org/10.1002/andp.19213690304. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19213690304.

[30] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, June 2017.

[31] Florian Weik et al. “ESPRESSO 4.0 – an extensible software package for simulating soft matter systems”. In: *The European Physical Journal Special Topics* 227.14 (2019), pp. 1789–1816. DOI: 10.1140/epjst/e2019-800186-9.

[32] Axel Arnold, Jason de Joannis, and Christian Holm. “Electrostatics in periodic slab geometries. I”. In: *The Journal of Chemical Physics* 117.6 (Aug. 2002), pp. 2496–2502. ISSN: 0021-9606. DOI: 10.1063/1.1491955. eprint: https://pubs.aip.org/aip/jcp/article-pdf/117/6/2496/19040427/2496_1_online.pdf.

[33] Jason de Joannis, Axel Arnold, and Christian Holm. “Electrostatics in periodic slab geometries. II”. In: *The Journal of Chemical Physics* 117.6 (Aug. 2002), pp. 2503–2512. ISSN: 0021-9606. DOI: 10.1063/1.1491954. eprint: https://pubs.aip.org/aip/jcp/article-pdf/117/6/2503/19039315/2503_1_online.pdf.

[34] Yasuhiko Nozaki and Charles Tanford. “[84] Examination of titration behavior”. In: *Enzyme Structure*. Vol. 11. Methods in Enzymology. Academic Press, 1967, pp. 715–734. DOI: https://doi.org/10.1016/S0076-6879(67)11088-4.

[35] Richard L. Thurlkill et al. “pK values of the ionizable groups of proteins”. In: *Protein Science* 15.5 (2006), pp. 1214–1218. DOI: https://doi.org/10.1110/ps.051840806. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1110/ps.051840806.

[36] Jiří Dvořák and Jiří Koryta. *Elektrochemie*. Praha: Academia, 1983.

[37] Cecil W. Davies. “397. The extent of dissociation of salts in water. Part VIII. An equation for the mean ionic activity coefficient of an electrolyte in water, and a revision of the dissociation constants of some sulphates”. In: *J. Chem. Soc.* (0 1938), pp. 2093–2098. DOI: 10.1039/JR9380002093.

[38] Robert A. Robinson and Robert H. Stokes. *Electrolyte solutions*. London: Butterworths, 1970.

[39] Kenneth S. Pitzer, J. Christopher Peiper, and R. H. Busey. “Thermodynamic Properties of Aqueous Sodium Chloride Solutions”. In: *Journal of Physical and Chemical Reference Data* 13.1 (Jan. 1984), pp. 1–102. ISSN: 0047-2689. DOI: 10.1063/1.555709. eprint: https://pubs.aip.org/aip/jpr/article-pdf/13/1/1/9766132/1_1_online.pdf.