Solution structure of glucagon-like peptide-1 analogues and their interactions with the endogenous receptor and albumin

Frimann, Tine Maja

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

TECHNICAL UNIVERSITY OF DENMARK

PHD THESIS

Solution structure of glucagon-like peptide-1 analogues and their interactions with the endogenous receptor and albumin

---

Author:
Tine Maja FREDERIKSEN

Department of Chemistry

March 2019

TECHNICAL UNIVERSITY OF DENMARK

PHD THESIS

Solution structure of glucagon-like peptide-1 analogues and their interactions with the endogenous receptor and albumin

Author:
Tine Maja FREDERIKSEN

Supervisors:
Günther H. PETERS
Pernille HARRIS
Jens T. BUKRINSKI

Physical and Biophysical Chemistry
Department of Chemistry

March 2019

Preface

This dissertation is submitted as partial fulfilment of the PhD degree to the Technical University of Denmark. The work described in the dissertation was carried out at the Department of Chemistry from September 2015 to February 2019 under supervision of Associate Professor Günther H. Peters, Associate Professor Pernille Harris, and Senior Advisor Jens T. Bukrinski. The project was funded by an Academic Excellence Scholarship, granted by the Department of Chemistry and an EliteForsk travel grant, granted by the Ministry of Education.

The purpose of this project was to give insights into the solution structures of acylated glucagon-like peptide-1 analogues, as well their interactions to both human serum albumin and the glucagon-like peptide-1 receptor. Subsequent to 62 simulations, countless scripts, and more than 160 asymmetric flow-field-flow fractioning measurements, a little more is known.

Chapter 1 of this dissertation gives an overall introduction to glucagon-like peptide-1 and its importance in peptide drug development. The remainder of the thesis is divided into two parts, with Part 1 covering computational studies and Part 2 covering experimental studies. In Part 1, chapter 2 gives an introduction to the binding interactions between glucagon-like peptide-1 and its receptor, chapter 3 provides an overview of the theory behind molecular dynamics simulations, and chapter 4 presents the results from the computational investigations. In addition, chapter 5 presents a paper manuscript on the calculation of entropy values. In Part 2, chapter 6 gives an introduction to the half-life extension of acylated glucagon-like peptide-1 analogues, chapter 7 gives a theoretical overview of the experimental methods. Chapter 8 was composed by Assistant Professor Katrine Qvortrup, and describes the synthetic procedure used to produce the investigated peptides, and finally, chapter 9 presents the results of the experimental investigations. Lastly, in chapter 10 is a short description of future work relevant for this thesis.

Acknowledgements

First of all I want to thank my supervisors Associate Professor Günther H. Peters, Associate Professor Pernille Harris, and Senior Advisor Jens T. Bukrinski for letting me work on this project and pursue my growing interests in computational chemistry and software development. Second, I want to thank them for pushing me out of my comfort zone behind the computer and into the lab. I cannot remember when I last learned so must new stuff within such a short period of time! Lastly, I want to express my gratitude for all the scientific discussions we have had. All three of you have been a great support and help in my journey to understand the world of physical biochemistry just a little better.

My appreciation also goes to the Department of Chemistry for funding my project with an Academic Excellence Scholarship, and to The Ministry of Education for bestowing me with an EliteForsk travel grant.

I want to acknowledge Jonas Mansoor for his work on setting up, running, and maintaining our clusters as well as for installing and renewing software and licenses.

My acknowledgement also goes to Assistant Professor Katrine Qvortrup and Master's student Maria Holm Rautenberg for their hard work on synthesising a library of GLP-1 analogues for me to study.

I also want to thank Master's student Suk Kyu Ko for providing solid and useful computational work while involved in my project, and for helping with running simulations, collecting data, and performing initial analyses.

Dierk Roessner, from Wyatt Technology Europe, should also be thanked for facilitating my external stays and for providing me with expert advise and help while visiting. On this note, a special thanks goes to PhD student Lorenzo Gentiluomo who, time after time, came to my rescue while bewildered in the lab. I also thank Albumedix for providing sample for my experiments.

A heartfelt gratitude goes to my current and previous gruop mates: Ulf Molich, Alina Kulakova, Christin Pohl, Sujata Mahapatra, Sowmya Indrakumar, Line A. Ryberg, Suk Kyu Ko, Natalia Skawińska, Pernille Sønderby, Kasper D. Tidemand, and Sindrila D. Banik, for brilliant discussions and lots of laughing. I would also like to thank my friends from the hall for making life in building 206 fun. I also want to thank Maria Blanner Bang for our talks, walks, and for putting certain things in better perspective. Lastly I give a very special thank to Line A. Ryberg and Ulf Molich for our great friendship and the adventures we have shared here at the Department of Chemistry during the last 8 years - and for plenty of cake!

My parents and brother also deserve my gratitude, for celebrating my victories along my studies and for making my pursuit for a PhD degree seem like the most obvious thing for me, even when I did not.

My most sincere gratitude goes to my fiancé, Freddie, for supporting me through the inevitable ups and downs of the PhD life, for making me laugh, and for bringing food and comfort during the last part of my work.

Tine Frederiksen

Abstract

Glucagon-like peptide-1 (GLP-1) is a hormone peptide, which is secreted to the bloodstream upon glucose uptake through nutrients. GLP-1 has one target, the GLP-1 receptor (GLP-1R), which is present in many different places in the body, e.g. the pancreas, heart, gut, and brain. Here, GLP-1 binding will activate GLP-1R and stimulate tissue specific signalling. In the pancreas, GLP-1 binding causes glucose dependent insulin secretion, which is relevant in relation to treatment of type 2 diabetes mellitus. In addition GLP-1 has the property to decrease bodyweight, and indications point to an ability to prevent cardiovascular disease, and protecting against dementia. However, GLP-1 immediatly suffers from its short half-life, making it inappropriate as a drug without sufficient modifications. Even though such modifications will in fact increase half-life, it is important that the biological effect of GLP-1R–GLP-1 binding is maintained.

Several strategies have been developed to overcome the short half-life of GLP-1. One such approach is acylation with fatty acid (FA) chains, promoting self-association and interactions to human serum albumin (HSA), which in turn will increase the systemic half-life. However, little is known about the atomic interactions of the self-associating oligomers and the effects of different acylation schemes hereon. Nor is it known how acylation of GLP-1 influences the interactions with HSA. In addition, only little knowlegde exists on GLP-1R–GLP-1 interactions on an atomic level, and none exists on interactions of GLP-1R with acylated GLP-1 analogues.

In this work, the solution structures of 11 different GLP-1 analogues, along with their interactions to HSA, were investigated. Assymetric flow-field-flow fractioning followed by light scattering was used to give insight into both oligomerisation and HSA interactions. The result showed that oligomer size is directly dependent on the length of the attached FA chain as well as the ionic strength, both with positive correlations. In addition, it was also shown that the presence of a linker decrease the oligomeric state. With regards to HSA interactions, acylated peptides containing medium to long FA chains and linkers interact partially with HSA, wherease peptides containing the same FA chains, but no linkers, interact fully with HSA.

This thesis also presents studies of binding interactions between eight different acylated GLP-1 analogues and GLP-1R by use of *in silio* modelling. To compare the results with experimental studies, molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) binding energies were calculated, and a methodology for computing entropy values was developed. The resulting MM-PBSA binding energies correlated with potency measurements. Additionaly, the computational studies could disclose that the acylation site on GLP-1 highly affects binding interactions between the FA chain and the receptor extracellular domain (ECD). If acylation is placed on GLP-1 Lys26 and Lys38, the FA chain can freely interact with the ECD, whereas, if acylation is placed on GLP-1 Lys34, the FA chain will be pointing away from the hydrophobic patch of the ECD. If there are two acylation sites, on GLP-1 Lys26 and Lys34, the two FA chains can interact with each other rather than the ECD.

Resumé

Glucagon-lignende peptid-1 (GLP-1) er et hormon peptid, der bliver udskilt til blodomløbet når glukose optages gennem næring. GLP-1 har ét mål, GLP-1 receptoren (GLP-1R), der er tilstede mange forskellige steder i kroppen, f.eks. i bugspytkirtlen, hjertet, tarmene og hjernen. Her vil GLP-1 binding aktivere GLP-1R og stimulere vævsbestemt signalprocessor. I bugspyt-kirtlen, vil GLP-1 binding forårsage glukose afhængig insulin udskillelse, hvilket er relevant i relation to behandling af type 2 diabetes mellitus. GLP-1 har yderligere egenskaber, der kan mindske kropsvægt og indikationer peger på en evne til at forebygge kardiovaskulær sygdomme og beskytte mod demens. GLP-1 lider dog umiddelbart under dets korte halveringstid, hvilket gør det uegnet som et medikament uden tilstrækkelige modifikationer. Selvom sådanne modifikationer faktisk kan øge halveringstiden, er det vigtigt at bibeholde effekten af GLP-1R–GLP-1 binding.

Flere strategier er blevet udviklet for at overkomme den korte halveringstid for GLP-1. En af disse er acylering med fedtsyre (FS) kæder, der promoverer selvassociering og interaktioner med human serum albumin (HSA), som dernæst vil forøge halveringstiden i blodet. Dog, er der ikke megen viden om de atomistiske interaktioner mellem de selvassocierende oligomere og effekten af forskellige acyleringsstrategier herpå. Det vides heller ikke hvordan acylering af GLP-1 influerer interaktionerne til HSA. Ydermere, eksisterer der kun lidt viden om GLP-1R–GLP-1 interaktioner på et atomistiske niveau, og ingen om interaktionerne mellem GLP-1R og acylerede GLP-1 afarter.

I dette studie vil opløsningsstrukturen af 11 forskellige GLP-1 afarter, og deres interaktioner med HSA, blive undersøgt. Asymmetrisk flow-felt-flow fraktionering efterfulgt at lysspredningsmålinger blev brugt til at belyse både oligomerisering og HSA interaktionerne. Resultaterne viste at, oligomerstørrelsen hænger direkte sammen med længden på den tilføjet FS kæde samt ionstyrken, begge dele korrelerer positivt. Ydermere blev det vist at tilstedeværelsen af en linker mindsker oligomerstørrelsen. Med hensyn til HSA interaktioner, interagerer acylerede peptider indeholdende medium til lange FS kæder og linkere delvist med HSA, hvorimod peptider indeholdende samme FS kæder, men ingen linker, interagerer fuldt med HSA.

Denne afhandling præsenterer også studier af bindingsinteraktioner mellem otte forskellige acylerede GLP-1 afarter og GLP-1R ved brug af *in silio* modellering. For at sammenligne resultaterne med eksperimentale studier, blev molekylær mekaniske-Poisson Boltzmann surface (overflade) areal (MM-PBSA) bindingsenergier beregnet, og en metodologi til at beregne entropiværdier blev udviklet. De resulterende MM-PBSA bindingsenergier korrelerede med effektivitetsmålinger. Ydermere kunne simuleringsstudierne give information om, at acyleringslokationen på GLP-1 påvirker bindingsinteraktioner mellem FS kæden og det ekstracellulærer domæne (ECD) i receptoren meget. Hvis acylering finder sted på GLP-1 Lys 26 og Lys38, kan FS kæden frit interagerer med ECD'et, hvorimod hvis acylering er placeret på Lys34, vil FS kæden pege væk fra det upolære område på ECD'et. Hvis der er to acyleringslokationer, på GLP-1 Lys26 og Lys34, kan de to FS kæder interagerer med hinanden fremfor at interagerer med ECD'et.

List of Abbreviations

α	Extracellular domain α-helix
A	Arts
APBS	Adaptive Poisson-Boltzmann solver
AF4	Asymmetric flow-field-flow fractioning
β1-β2	β1-to-β2 turn
βx	β-strand x
BSA C-term	C-terminal
com	Complex
cov	Covariance
CPU	Central processing unit
DLS	Dynamic light scattering
7TM	7 α-helix transmembrane
DLS	Dynamic light scattering
DPP IV	Dipeptidyl-peptidase IV
E	English
E-TM	Linker between the extracellular and transmembrane domains
ECD	Extracellular domain
ECDLx	Extracellular domain loop x
ECLx	Extracellular loop x
EDA	Essential dynamics analyses
FA	Fatty acid
GLP-1	Glucagon-like peptide-1
GLP-2	Glucagon-like peptide-2
GLP-1R	Glucagon-like peptide-1 receptor
GPCR	Glucagon protein-coupled receptor
GRPP	Glicentin-related polypeptide
GPU	Graphical processing unit
HSA	Human serum albumin
ICLx	Intracellular loop x
IP-1	Intervening peptide-1
IP-2	Intervening peptide-2
IS	Ionic strength
Lira	Liraglutide
LS	Light scattering
M	Math
MALS	Multi angle light scattering
MD	Molecular dynamics
MM-PBSA	Molecular mechanics-Poisson Boltzmann surface area
MPGF	Major proglucagon fragment
Mw	Molecular weight
N-term	N-terminal
NMA	Normal mode analyses
NPT	Constant number of particles, constant pressure, constant temperature

OXM	Oxyntomodulin
Pc	Channel pressure
PDB	Protein data bank
pep	Peptide
PES	Polyethersulfone
PME	particle-mesh-Ewald
POPC	3-palmitoyl-2-oleoyl-D-glycero-1-phosphoatidylcholine
Px	Cross flow pressure
psf	Protein structure file
QHA	Quasi harmonic analyses
rec	Receptor
RI	Refractive index
RMSD	Root mean square deviation
S/N	Signal to noise ratio
SASA	Solvent accesible surface area
SC	Short channel
SD	Standard devation
SLS	Static light scattering
T2DM	Type 2 Diabetes Mellitus
TMx	Transmembrane α-helix x
TMD	Transmembrane domain
UV	Ultra violet
var	Variance
Vd	Detector flow
vdW	van der Waals
VWD	Variable Wavelength Detector
Vx	Cross flow
W	Wide

Contents

Chapter 1

Introduction

In today's world, especially in the Western parts, lifestyle diseases are becoming an immense issue. Not only do the lifestyle of the general population become more unhealthy, it also costs society more and more due to an increase in medical needs in relation to lifestyle diseases. Such diseases include cardiovascular diseases, such as strokes and heart attacks, diabetes, dementia, overweight, and obesity. In fact, cardiovascular diseases accounted for 31 % of all deaths worldwide in 2016, and diabetes for 3 %. As a result cardiovascular diseases and diabetes land an impressive first and fourth place of deaths caused by noncommunicable diseases in the world in 2016 [1].

In combination with an increase in life-style diseases, the population gets older and older. Today, the average lifespan is higher than ever, giving rise to another type of diseases, namely neurodegenerative diseases like dementia [2]. In 2016, dementia-related diseases such as Alzheimer's and Parkinson's were on the fifth place on the list of top 10 death causes in upper-middle-income countries, and an astounding third place in high-income countries [3].

Therefore, therapeutic solutions to lower these breathtaking statistics are highly sought. Even though the 2016 statistics are record breaking, treatments for some of the mentioned noncommunicable diseases have been investigated since the beginning of the 20^{th} century. Since 1906, treatments for diabetes have been looked into in relation to insulin signalling [4], and in 1981 the incretin hormone glucagon-like peptide-1 (GLP-1) was identified [5, 6]. By 1987, it was clear, that GLP-1 was involved in glucose dependent insulin regulation [7], which makes it an obvious candidate for treatment of type 2 diabetes mellitus (T2DM). It was also discovered that GLP-1 plays a role in suppressing gastric emptying and appetite, features relevant in bodyweight regulation [8–10]. Since then, initial studies have tied GLP-1 to neuroprotective properties, which could be of great importance in the battle against dementia [11, 12]. In addition, recent studies have proven that GLP-1 is also involved in reducing the risk of developing cardiovascular diseases [13, 14].

All in all, this means that finding a durable way of controlling GLP-1 related signalling can present a solution to relieve the increasing health pressure from the increase in obesity, diabetes, cardiovascular diseases, and possibly dementia. To achieve this, an elaborate understanding of GLP-1 and its role in the body is necessary. The following will introduce GLP-1.

1.1 Introducing GLP-1

GLP-1 is a signalling hormone made of either 30 or 31 amino acid residues. The sequences of both versions are given in fig. 1.1 [15, 16].

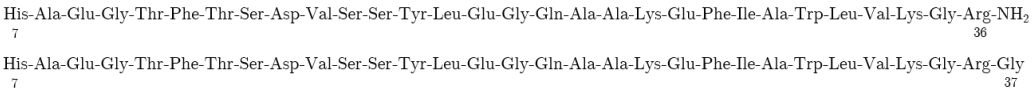


FIGURE 1.1: The two versions of the glucagon-like peptide-1. The most abundant and slightly more systemically stable version, GLP-1(7-36)NH_2 is given above. Below, is the second version, GLP-1(7-37).

The shorter version, GLP-1(7-36)-NH_2 is slightly more abundant and stable in systemic circulation than the other, longer version, GLP-1(7-37), but they share the same biological effect [17, 18]. GLP-1 adopts a helical structure in membrane-like environments. In aqueous solution, GLP-1 is unstructured, but in membrane-like environments it adopts an ordered helix structure starting from the C-terminal end of the peptide. The resulting single stranded α-helix spans from Phe12*[1]/Thr13* to Lys34*, with a dis-ordered part in the N-terminal from His7* to Thr11*/Phe12* [19–21].

GLP-1 is modified from the precursor proglucagon residues 78-107. The proglucagon gene is part of the human chromosome 2, and is transcribed into a single mRNA strand, which is then translated into a single protein (proglucagon) [22]. The proglucagon protein is present in three different tissues; the pancreatic α-cells, the intestinal L-cells, and finally, in the brain. In the pancreas, proglucagon undergoes posttranslational processing into glicentin-related polypeptide, glucagon, intervening peptide-1, and major proglucagon fragment. In the intestine and brain, proglucagon is processed into glicentin and/or oxyntomodulin, intervening peptide-2, GLP-1, and glucagon-like peptide-2 [23, 24], see fig. 1.2.

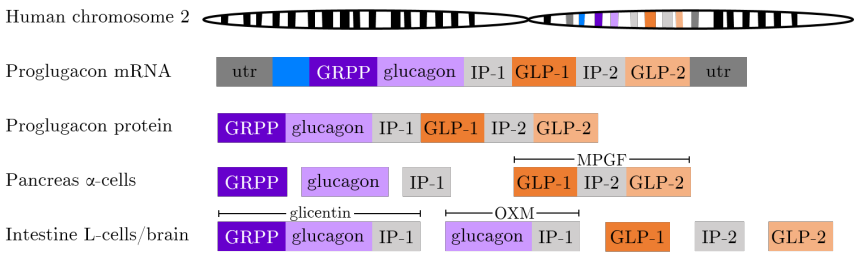


FIGURE 1.2: The human chromosome 2, which is transcribed into the proglucagon mRNA, which in turn is translated into proglucagon. Proglucaon undergoes tissue-specific posttranslational processing, resulting in glicentin-related polypeptide (GRPP), glucagon, intervening peptide-1 (IP-1), and major proglucagon fragment (MPGF) in the pancreas. In the brain and gut, proclucagon is processed into glicentin and/or oxyntomodulin (OXM), intervening peptide-2 (IP-2), GLP-1, and glucagon-like peptide-2 (GLP-2). Figure based on [25, 26].

1.2 GLP-1 function

GLP-1 has one target protein; the GLP-1 receptor (GLP-1R) [27]. GLP-1R is a class B glucagon protein-coupled receptor (GPCR) membrane protein consisting of 463 amino acid residues (29-431). It has an extracellular domain (ECD), a transmembrane domain (TMD), and an intracellular part. The TMD is made of seven α-helices

[1]GLP-1 (peptide) residues are noted with a '*', to make it distinguishable e.g. when describing binding interactions.

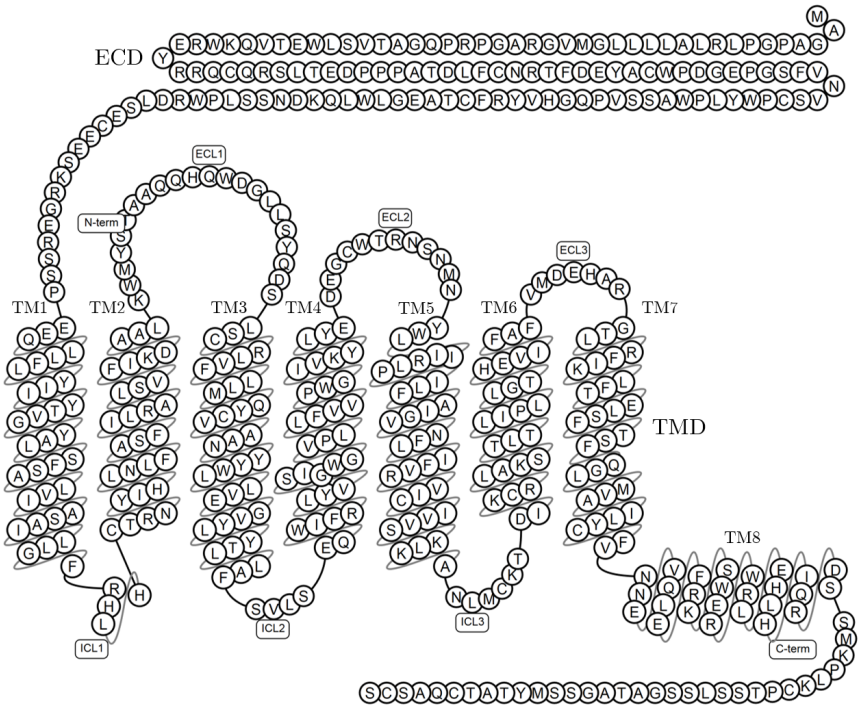


FIGURE 1.3: Snake diagram of GLP-1R. The transmembrane helices are labelled, and the two domains (ECD and TMD) are indicated. Diagram created using `www.gpcrdb.org`.

forming a hollow barrel transverse to the membrane bilayer. The ECD is the receptor N-terminal, and the intracellular part is the C-terminal [28–30]. A snake diagram of GLP-1R is given in fig. 1.3

GLP-1R has high similarity with other GPCRs and is believed to follow the same activation mechanism as reported previously for several other GPCRs [31–33]. Receptor activation was found to happen in a two-step fashion, termed the two-domain model, inducing a conformational change in the receptor which functions as the activation mechanism. First, the C-terminal part of GLP-1 binds to the inactivated receptor's ECD [34, 35]. Initial binding is believed to be induced by the partial helical structure of the GLP-1 C-terminal in the vicinity of the receptor. When the C-terminal is bound to the ECD, the N-terminal part of GLP-1 will be positioned into the top TMD part of the receptor which will induce structural changes in the receptor TM6 and ECL3 [36]. It is the structural changes in the receptor upon GLP-1 binding that activate the signalling [36]. GLP-1R activation is illustrated in fig. 1.4.

The GLP-1 C-terminal residues Phe28* and Ile29* are involved in initial receptor binding, and are crucial for activation only to ensure the interaction with the receptor ECD. On the other hand, the GLP-1 N-terminal residues His7*, Glu9*, Asp15* are directly involved in receptor activation [36, 38].

The GLP-1 receptor is present several places in the body, e.g. in the kidneys, lungs, intestine, pancreas, heart, blood vessels, and the brain [27, 39–42]. GLP-1 is secreted upon nutrition uptake in the gut and thus enters the bloodstream. In relation to diabetes treatment, it is the presence of GLP-1R in the pancreas that bears significance. Here, GLP-1 binds to GLP-1R in the pancreatic β-cells which stimulates insulin secretion [43]. Therefore, GLP-1 treatment is relevant for T2DM treatment [7]. In addition, glucose dependent GLP-1 secretion also suppresses secretion

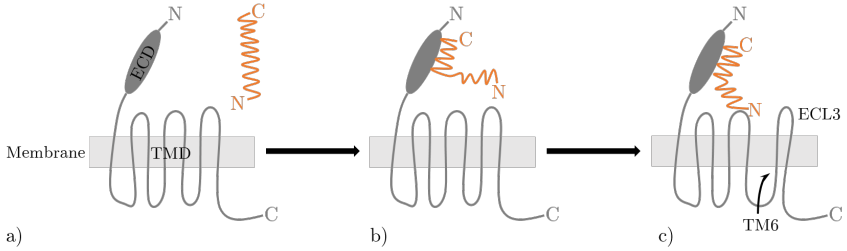


FIGURE 1.4: The two-domain model activation of GLP-1R upon binding of GLP-1. In a), the receptor (dark grey) is inactivated and GLP-1 (orange) is unbound. b) shows the first part of the two-domain model, where the C-terminal of GLP-1 binds to the receptor ECD, and c) shows the second part with the GLP-1 N-terminal bound to the activated receptor TMD part, resulting in structural changes in TM6 and ECL3. Figure based on [37].

of glucagon, which is an insulin suppressor [44]. In the gut, GLP-1R activation inhibits gastric mobility and gastric emptying, while in the brain it increases satiety. As a result, food intake is decreased, which causes reduction in body mass, and proves GLP-1 relevant in relation to treatment of obesity [45, 46]. Also in the brain, GLP-1R activation has shown neuroprotective properties by reducing amyloid-β fibrils and by decreasing neuron apoptosis, which makes GLP-1R activation in the brain a possible target for prevention of dementia [11]. GLP-1R activation in the heart causes an increase in glucose uptake, coronary flow, and heart rate, making it a possible target for treatment of cardiovascular diseases [13].

1.3 Limitations of GLP-1

Having presented GLP-1 and a selection of its very beneficial properties, there is one major problem. The half-life of GLP-1 in the bloodstream is approximate 1 to 2 min. The reason for this short half-life stems from rapid degradation via dipeptidyl peptidase 4 (DPP IV), which cleaves the peptide between Ala8* and Glu9*, leaving the inactive GLP-1(9-30/31) metabolite [47]. Additionally, GLP-1 is also cleared via the kidneys [48]. This, in itself, makes GLP-1 useless as a drug candidate without appropriate modifications.

1.4 Solutions

Alternative solutions, circumventing the shortcoming of GLP-1, i.e. its short half-life, must be found to explore GLP-1R activation as a drug treatment. One such strategy is the development of GLP-1 peptide analogues. Here, peptides resembling GLP-1 have been developed using several different approaches [49, 50].

The analogues must resemble GLP-1 enough to provide the same biological effect, but with an increased half-life appropriate for daily (or rarer) administration. Numerous strategies have been found to effectively increase the half-life of peptide drugs, e.g.: i) acylation of GLP-1 with fatty acid (FA) chains, ii) covalent binding to human serum albumin (HSA), iii) attachment of Fc fragments of immunoglobulin, and iv) backbone elogation with a Lys tag in the C-terminal [25, 51].

This study will focus on GLP-1 analogues from group i). From this category, liraglutide is worth mentioning. Liraglutide is a Novo Nordisk drug sold under the

name Victoza®. In fact, it was the first acylated GLP-1 analogue approved for T2DM treatment [25, 52] and was approved for treating T2DM in Europe in 2009 and in the US in 2010. Later, it was approved for obesity treatment in the US in 2014 and in Europe in 2015. Furthermore, liraglutide was also shown to have neuroprotective properties [53–55], as well as preventing cardiovascular diseases [56]. Liraglutide is a GLP-1(7-37) analogue that has a 16 Carbon atom FA chain attached to Lys26* via an L-γ-Glu linker. Lys34* has been mutated into Arg to ensure selective acylation at Lys26* [52]. The schematic structure of liraglutide is shown in fig. 1.7.

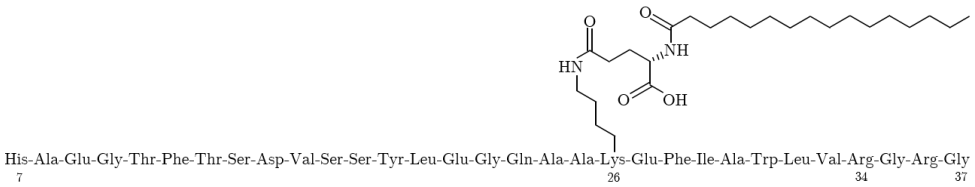


FIGURE 1.5: The sequence of liraglutide, including the attached fatty acid chain and Lys34*Arg mutation.

For liraglutide and other GLP-1 acylated analogues, the attached FA chain can interact with HSA and form oligomers induced by hydrophobic interactions, which prevents DPP IV degradation and renal clearance [52, 57, 58]. The result is a half-life extension to around 11 to 15 h [52]. In addition, liralgutide has been shown to control insulin secretion in a homeostatic fashion, meaning that administrating lirgalutide to healthy or non-diabetic patients does not lead to hypoglycemia. This, in turn, makes it possible to use for treatment of disorders in non-diabetic patients [59].

1.5 Investigated analogues

In this study, two sets of GLP-1 analogues have been investigated. They both include GLP-1(7-36)NH_2 and liralgutide. The peptides were chosen to most efficiently represent different acylation sites (Lys26*, Lys34*, and Lys38*[2]), different FA chain lengths (from 8-16 C-atoms), the presence or absence of a linker, and lastly, the presence or absence of a free acid group in the ω tail of the FA chain. The first set consists of eight peptides (including liraglutide and GLP-1(7-36)NH_2). This set incorporates different acylation sites and different FA chain lengths. Furthermore, the presence of a linker and free acid group is varied. The schematic structures for the peptides in the first set are shown in figs. 1.6 to 1.13. The second set of peptides consists of 11 peptides (including GLP-1(7-36)NH_2 and liralgutide), where the acylation site is kept constant at Lys26*. The length of the FA chain changes, as well as the presence of a linker and the presence of an acid group. Schematic structures for the second set of peptides are shown in figs. 1.14 to 1.22 (excluding GLP-1(7-36)NH_2 which are shown in figs. 1.6 and 1.7).

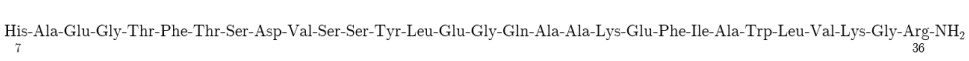


FIGURE 1.6: Schematic representation of GLP-1.

[2]Addition on an extra Lys in the C-terminal end of the GLP-1(7-37) sequence.

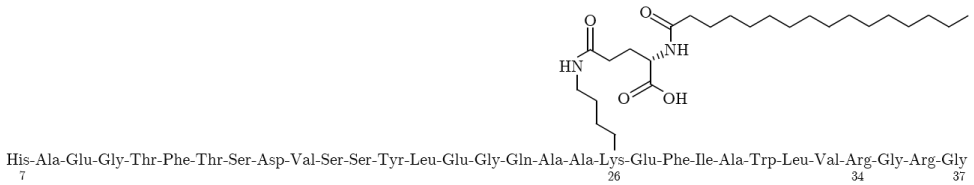


FIGURE 1.7: Schematic representation of liraglutide. Acylation scheme: 16 C-atoms FA chain and an L-γ-Glu linker, acylation on Lys26*.

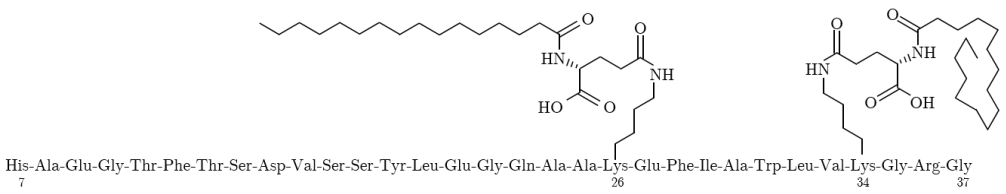


FIGURE 1.8: Schematic representation of the C16L,K^{26,34} peptide. Acylation scheme: 16 C-atoms FA chain and an L-γ-Glu linker, acylation on Lys26* and Lys34*.

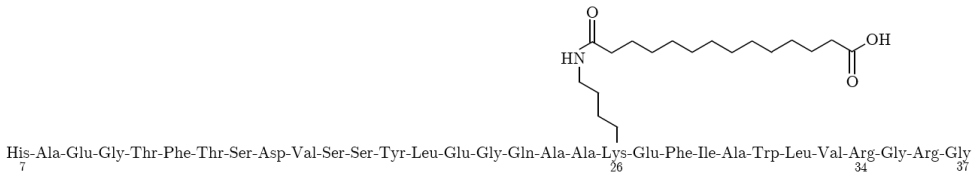


FIGURE 1.9: Schematic representation of the C14A peptide. Acylation scheme: 14 C-atoms FA chain and a free acid group in the ω chain end, acylation on Lys26*.

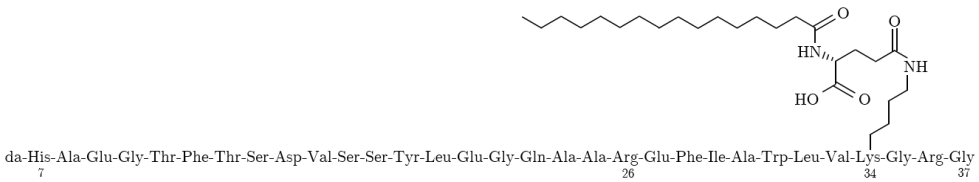


FIGURE 1.10: Schematic representation of the daC16L,K^{34} peptide. Acylation scheme: 16 C-atoms FA chain and an L-γ-Glu linker, acylation on Lys34*. The N-terminal amino group on His7* has been removed, this is denoted by 'da' (des-amino).

FIGURE 1.11: Schematic representation of the daC8,K^{34} peptide. Acylation scheme: Eight C-atoms FA chain, acylation on Lys34*. The N-terminal amino group on His7* has been removed, this is denoted by 'da' (des-amino).

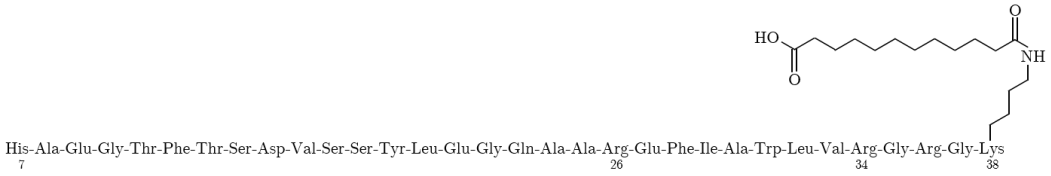


FIGURE 1.12: Schematic representation of the C12A,K^{38} peptide. Acylation scheme: 12 C-atoms FA chain and a free acid group in the ω chain end, acylation on Lys38*.

FIGURE 1.13: Schematic representation of the C14L,K^{38} peptide. Acylation scheme: 14 C-atoms FA chain and an L-γ-Glu linker, acylation on Lys38*.

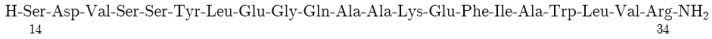


FIGURE 1.14: Schematic structure of GLP-1(14-34).

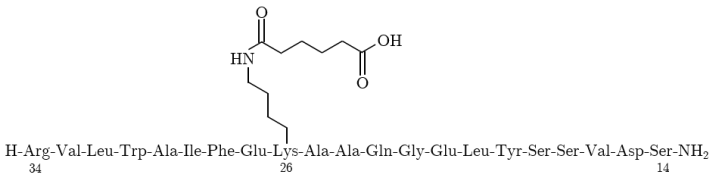


FIGURE 1.15: Schematic structure of C6Arev. The amino acid sequence of GLP-1 is reversed. Acylation scheme: Six C-atoms FA chain and a free acid group in the ω chain end, acylation on Lys26*.

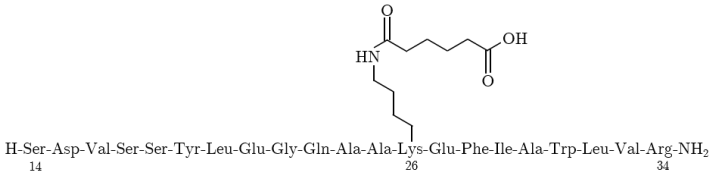


FIGURE 1.16: Schematic structure of C6A. Acylation scheme: Six C-atoms FA chain and a free acid group in the ω chain end, acylation on Lys26*.

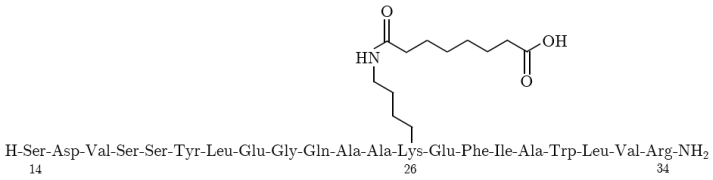


FIGURE 1.17: Schematic structure of C8A. Acylation scheme: Eight C-atoms FA chain and a free acid group in the ω chain end, acylation on Lys26*.

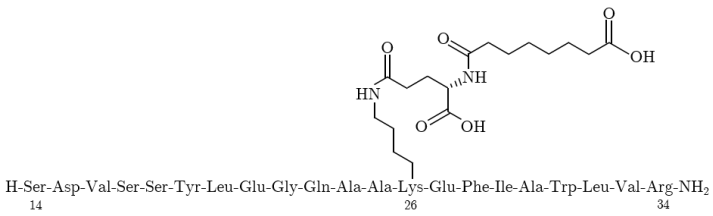


FIGURE 1.18: Schematic structure of C8LA. Acylation scheme: Eight C-atoms FA chain, an L-γ-Glu linker, and a free acid group in the ω chain end, acylation on Lys26*.

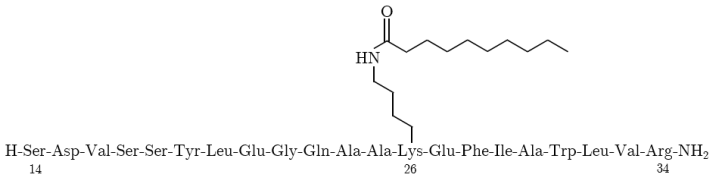


FIGURE 1.19: Schematic structure of C10. Acylation scheme: Ten C-atoms FA chain, acylation on Lys26*.

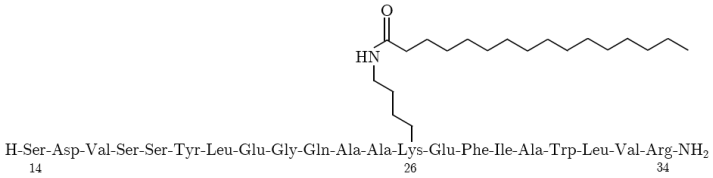


FIGURE 1.20: Schematic structure of C16. Acylation scheme: 16 C-atoms FA chain, acylation on Lys26*.

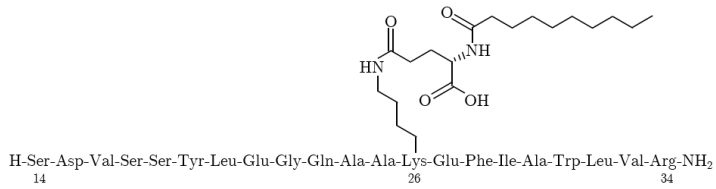


FIGURE 1.21: Schematic structure of C10L. Acylation scheme: Ten C-atoms FA chain and an L-γ-Glu linker, acylation on Lys26*.

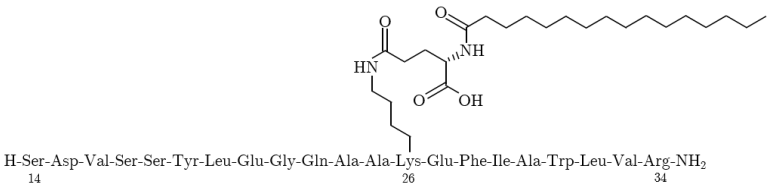


FIGURE 1.22: Schematic structure of C16L(14-34). Acylation scheme: 16 C-atoms FA chain and an L-γ-Glu linker, acylation on Lys26*.

1.6 Aim of this study

The aim of this PhD thesis was to investigate how different acylated GLP-1 analogues behave in solution, mainly in relation to oligomer formation and HSA interaction, and how they differ in binding to GLP-1R. Two approaches were used to cover these topics. The first approach covers *in silico* studies of the first set of peptides in complex with the GLP-1 receptor. The second approach involves experimental studies on the solution structures and HSA interactions of the different peptides in the second set. Part 1 of this thesis presents the *in silico* studies, and Part 2 of this thesis presents the experimental studies.

Both parts contain a more elaborate introduction to the individual topics as well as the used methods and setups.

Part 1: Computational investigations

Chapter 2

Computational introduction

This part of the thesis will cover computational studies of the GLP-1 receptor in complex with GLP-1 and varius GLP-1 analogues. Chapter 3 gives a short description to molecular dynamics (MD) simulations, chapter 4 presents a study of the binding interactions between the GLP-1 receptor and the first set of peptides. Finally, chapter 5 gives detailed descriptions of how to compute binding energies from MD simulations with a focus on ensuring converged entropy values of binding.

2.1 Binding interactions: GLP-1R–GLP-1

Peptide binding will be investigated on the basis of GLP-1 binding, since structural studies of GLP-1 binding to the receptor have been performed previously.

Many studies have been made of the biological effect of GLP-1R agonists, including the biological effect of acylated GLP-1 analogues [60, 61]. Such studies cover both potency studies as well as mutagenesis studies of GLP-1 and GLP-1R [25, 38]. However, GLP-1R is a membrane protein and has proven difficult to crystalise, and structures covering its full length have only been reported recently [62–64]. Even though such structures can give insights into GLP-1R, none of them have acylated GLP-1 analogues bound. As a result, the overall picture of GLP-1 binding was completed from different studies covering both full length GLP-1R and only the receptor ECD. In continuation hereof, MD simulations are an obvious choice for detailed investigations of GLP-1 analogue binding.

However, a good representable structure of the GLP-1R–GLP-1 complex must be obtained to perform qualified MD simulations describing GLP-1 and GLP-1 analogue binding. Since only a few structures exist of full length GLP-1R, the bigger picture covering GLP-1 binding has to be puzzled together.

Runge et al. 2008 [65] crystallised the receptor ECD in complex with Exendin-4 (PDB ID 3C59), and similarly, Underwood et al. 2010 [66] could solve the structure of the receptor ECD in complex with GLP-1 (PDB ID 3IOL). Jazayeri et al. 2017 [62] reported the full length GLP-1R structure in complex with a truncated 10 residue peptide agonist (PDB ID 5NX2), whereas Zhang et al. 2017 [63] showed oryctolagus cuniculus GLP-1R in complex with GLP-1 (PDB ID 5VAI). Last year, Liang et al. [64] presented GLP-1R in complex with Exendin-P5 (PDB ID 6B3J). Here, the full picture of GLP-1 binding is drawn based on the findings by Underwood et al. 2010, Jazayeri et al. 2017, and Zhang et al. 2017 in combination with the modelled activation studies reported by Santiago et al. [36]. The simulations performed by Santiago et al. were run on a structure mainly based on the 5VAI structure produced by Zhang et al. 2017.

Starting with the receptor ECD, the presence of a hydrophobic patch comprised of the ECD α-helix residues Leu32, Thr35, Val36, and Trp39, in combination with

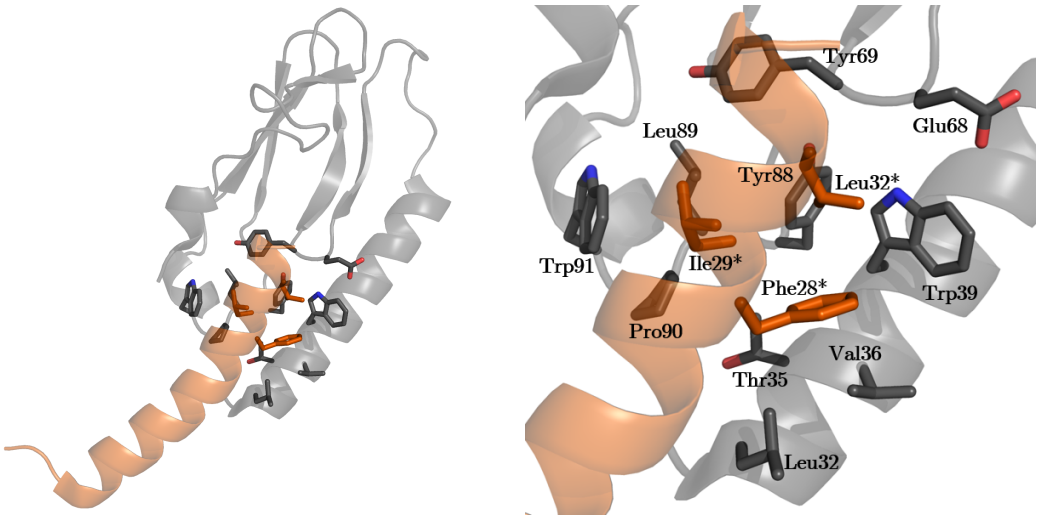


FIGURE 2.1: The hydrophobic patch of the GLP-1R ECD in complex with GLP-1. A zoom in showing the involved residues is given to the right. The receptor is shown in grey ribbon, and GLP-1 in orange ribbon. The hydrophobic residues comprising the apolar patch on the ECD are shown as grey sticks. The GLP-1 residues interacting with this patch are shown as orange sticks.

Tyr69 in the β1-to-β2 turn, and ECDL2 residues Tyr88, Leu89, Pro90, and Trp91, was reported by Runge et al. 2008. This patch was likewise reported by Underwood et al. 2010 and conserved in the stucture reported by Jazayeri et al. 2017. Figure 2.1 shows the hydrophobic patch in the ECD in complex to GLP-1, as reported in the 3IOL structure by Underwood et al. 2010. The hydrophobic residues Phe28*, Ile29*, and Leu32* in GLP-1, that interact with this hydrophobic patch are also shown in Figure 2.1.

Jazayeri et al. 2017 report interactions based on the structure between a truncated peptide agonist, Pep5, resembling GLP-1(7-17), and the TMD. In fig. 2.2, the 5NX2 structure of GLP-1R and Pep5 is shown together with the interactions between the peptide residues 1-5 (corresponding to GLP-1(7-12)) and the core part of the TMD. The remaining interactions reported by Jazayeri et al. 2017 between Pep5 (residues 6-10, corresponding to GLP-1(13-17)) and GLP-1R are shown in fig. 2.3.

The 5VAI structure comprises full length rabbit GLP-1R in complex with full length GLP-1. As a result, interactions spanning the entire GLP-1 peptide are given, contrary to the 5NX2 structure with a truncated peptide, and the 3IOL structure of the ECD only. Figure 2.4 shows the interactions based on the structure between the GLP-1 N-terminal/middle part and the core TMD.

The interactions between the GLP-1 middel/C-terminal and the top TMD/ECD are represented in fig. 2.5. In combination, this gives a set of binding interactions between GLP-1R and GLP-1 (the truncated peptide agonist in 5NX2, Pep5, is translated into GLP-1(7-17) via table 2.1) that covers the full length GLP-1 structure and both the TMD and the ECD of GLP-1R.

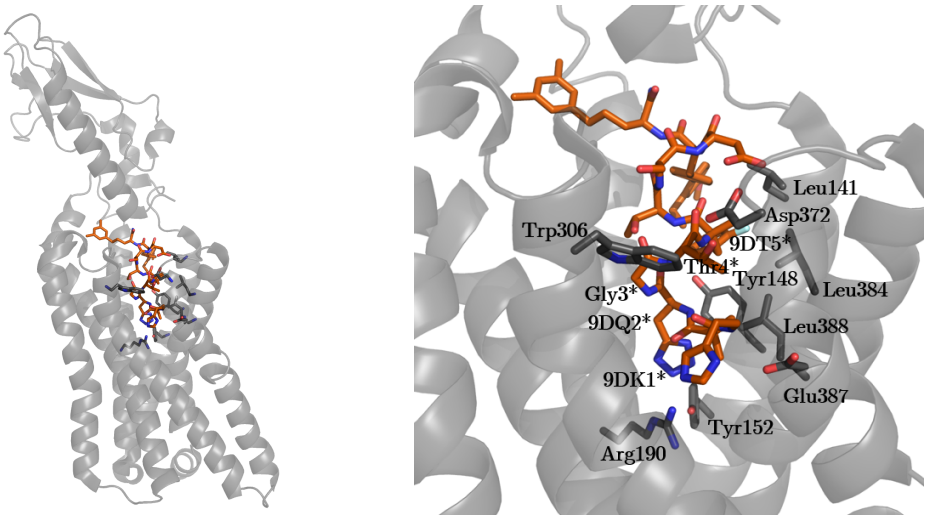


FIGURE 2.2: The interactions of Pep5 residues 1-5 (corresponding to GLP-1(7-12)) with the core part of the GLP-1R TMD. A zoom in showing the involved residues is given to the right. The receptor is shown in grey ribbon, with the interacting residues shown as grey sticks. The peptide agonist is shown as orange sticks. See table 2.2 for list of interactions.

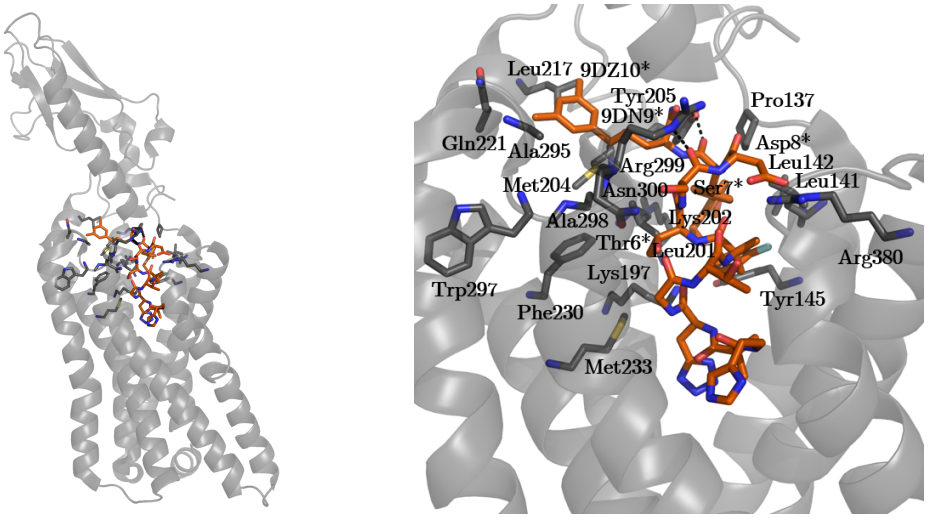


FIGURE 2.3: The interactions of Pep5 7-10 (corresponding to GLP-1(13-17)) with the top part of the GLP-1R TMD. A zoom in showing the involved residues is given to the right. The receptor is shown in grey ribbon, with the interacting residues shown as grey sticks. The peptide agonist is shown as orange sticks. Hydrogen bonds are shown as dashed lines. See table 2.2 for list of interactions.

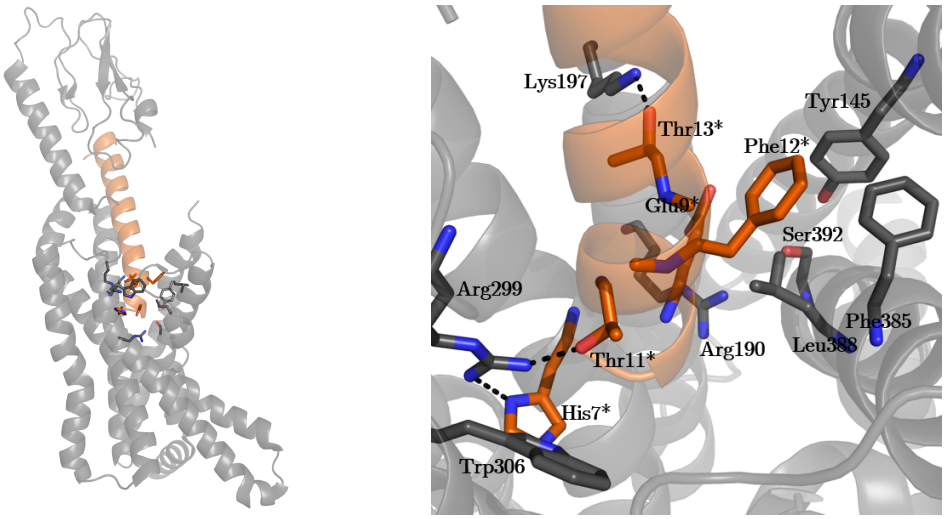


FIGURE 2.4: The interactions of GLP-1 peptide residues 7-13 with the core part of the GLP-1R TMD. A zoom in showing the involved residues is given to the right. The receptor is shown in grey ribbon, with the interacting residues shown as grey sticks. GLP-1 is shown in orange ribbon, and the peptide residues involved in interactions are shown as orange sticks. Hydrogen bonds are shown as dashed lines. See table 2.2 for list of interactions.

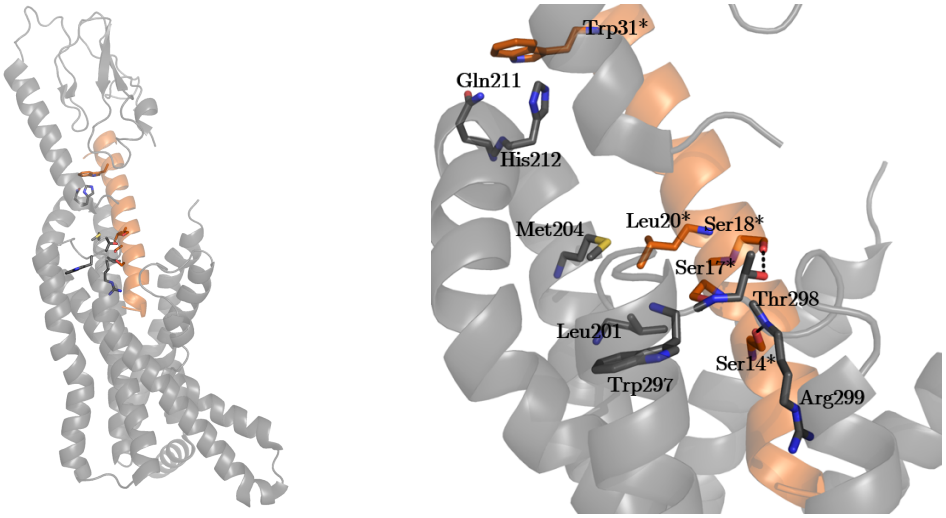


FIGURE 2.5: The interactions of GLP-1 peptide residues 14-37 with the top part of the GLP-1R TMD and ECD. A zoom in showing the involved residues is given to the right. The receptor is shown in grey ribbon, with the interacting residues shown as grey sticks. GLP-1 is shown in orange ribbon, and the peptide residues involved in interactions are shown as orange sticks. Hydrogen bonds are shown as dashed lines. See table 2.2 for list of interactions.

TABLE 2.1: Conversion scheme from Pep5 to GLP-1(7-17).

Residue #	Pep5	Residue #	GLP-1
1	9DK	7	His
1	9DK	8	Ala
2	9DQ	9	Glu
3	Gly	10	Gly
4	Thr	11	The
5	9DT	12	Phe
6	Thr	13	Thr
7	Ser	14	Ser
8	Asp	15	Asp
9	9DN	16	Val
10	9DZ	17	Ser

The interactions described in the above studies are summarised in table 2.2, including the interactions reported by Santiago et al. 2018, based on modelling studies.

TABLE 2.2: Interactions between GLP-1 and GLP-1R as reported by Jazayeri et al. 2017 (J), Santiago et al. 2018 (S), Underwood et al. 2010 (U), and Zhang et al. 2017 (Z). † indicates residue cluster formation rather than individual interaction pairs. Potency reduction from mutagenesis studies is given as reviewed by de Graaf et al. 2016 [25]. If several mutations have been reported, only the receptor residue mutation causing the highest potency reduction is reported.

GLP-1	GLP-1R	Source	Potency reduction (x-fold)
His7*	Arg299 (ECL2)	Z	85 (Arg299Ala)
His7*	Trp306 (TM5)	Z	206 (Trp306Ala)
His7*	Glu364 (TM6)	S	15 (Glu364Ala)
His7*	Glu387 (TM7)	J, S	10 (Glu387Asp)
Ala8*	Leu384 (TM7)	J	41 (Leu384Ala)
Ala8*	Leu388 (TM7)	J	208 (Leu388Ala)
Glu9*	Tyr145 (TM1)	Z	<5 (Tyr145Ala)
Glu9*	Tyr148 (TM1)	S	26 (Tyr148Ala)
Glu9*	Tyr152 (TM1)	J, S	<5 (Tyr152 His)
Glu9*	Arg190 (TM2)	J, S, Z	270 (Arg190Ala)
Glu9*	Leu388 (TM7)	S, Z	208 (Leu388Ala)
Glu9*	Ser392 (TM7)	S, Z	<5 (Ser392Ala)
Gly10*	Trp306 (TM5)	J	109 (Trp306Ala)
Thr11*	Arg299 (ECL2)	S, Z	85 (Arg299Ala)
Thr11*	Asp372 (ECL3)	J	59 (Asp372Ala)
Thr11*	Leu384 (TM7)	J	41 (Leu384Ala)
Phe12*	Leu141 (TM1)	J	<5 (Leu141Ala)
Phe12*	Leu142 (TM1)	J	Not measured
Phe12*	Tyr145 (TM1)	Z	<5 (Tyr145Ala)
Phe12*	Tyr148 (TM1)	J	26 (Tyr148Ala)
Phe12*	Leu384 (TM7)	J	41 (Leu384Ala)
Phe12*	Phe385 (TM7)	Z	<5 (Phe385Ala)

Continued on next page

Table 2.2 – *Continued from previous page*

GLP-1	GLP-1R	Source	Potency reduction (x-fold)
Phe12*	Leu388 (TM7)	J	208 (Leu388Ala)
Thr13*	Lys197 (TM2)	J, Z	630 (Lys197Ala)
Thr13*	Leu201 (TM2)	J	Not determined
Thr13*	Phe230 (TM3)	J	Not measured
Thr13*	Met233 (TM3)	J	70 (Met233Ala)
Thr13*	Trp297 (ECL2)	S	316 (Trp297Ala)
Thr13*	Thr298 (ECL2)	J	<5 (Thr298Ala)
Ser14*	Asn300 (ECL2)	J	501 (Asn300Ala)
Ser14* †	Trp297 (ECL2)	Z	316 (Trp297Ala)
Ser17* †	Thr298 (ECL2)	Z	<5 (Thr298Ala)
Ser18* †	Arg299 (ECL2)	J, Z	85 (Arg299Ala)
Asp15*	Arg380 (TM7)	J, S	1853 (Arg380Asp)
Val16*	Pro137 (E-TM)	J	Not measured
Val16*	Leu141 (TM1)	J	<5 (Leu141Ala)
Val16*	Leu142 (TM2)	J	Not measured
Val16*	Tyr145 (TM1)	J	<5 (Tyr145Ala)
Val16*	Leu201 (TM2)	J	Not determined
Val16*	Lys202 (TM2)	J	Not measured
Val16*	Tyr205 (TM2)	J	62 (Tyr205Ala)
Ser17*	Ser31 (ECD)	J	Not measured
Ser17*	Leu32 (α)	J	<5 (Leu32Ala)
Ser17*	Trp33 (α)	J	Not measured
Ser17*	Met204 (TM2)	J	334 (Met204Ala)
Ser17*	Tyr205 (ECL1)	J	62 (Tyr205Ala)
Ser17*	Leu217 (ECL1)	J	<5 (Leu217Ala)
Ser17*	Gln221 (ECL1)	J	Not measured
Ser17*	Gly295 (ECL2)	J	<5 (Gly295Ala)
Ser17*	Thr298 (ECL2)	J	<5 (Thr298Ala)
Leu20*	Leu201 (TM2)	S, Z	Not determined
Leu20*	Met204 (TM2)	S, Z	334 (Met204Ala)
Glu21*	Arg299 (ECL2)	S	85 (Arg299Ala)
Lys26*	Glu128 (ECD)	U	<5 (Glu128Ala)
Trp31*	Gln211 (ECL1)	S, Z	<5 (Gln211Asp)
Trp31*	His212 (ECL1)	S, Z	<5 (His212Ala)
Val33*	Arg121 (ECD)	U	<5 (Arg121Ala)

In addition to the binding interactions identified in structural studies, table 2.2 includes mutagenesis data as reviewed by de Graaf et al. 2016. Data is given as GLP-1 potency reduction. This gives an idea of the importance of the given receptor residue in relation to potency.

From table 2.2, it is evident that many interactions between GLP-1 and GLP-1R have been reported. However, the data are collected from multiple studies, and only a few interactions have been reported in several studies. Furthermore, many of the given interactions involve receptor residues that does not seem to be important for potency (reduction < 5-fold). This aside, a set of interactions alongside the recognition of the hydrophobic cluster formation shown in fig. 2.1 has been assigned. Both constitute important factors in GLP-1 binding and can be used to evaluate interactions between the GLP-1R and various acylated GLP-1 analogues.

2.2 GLP-1 analogues

It has been proven that acylation of GLP-1 with FA chains causes differences in effectiveness of receptor activation [52, 67, 68]. This was shown as differences in potency compared to native GLP-1. Meaning that acylation of GLP-1 can either increase or decrease binding to GLP-1R, assuming that potency is related to binding affinity. In the liraglutide development process, several acylation schemes were investigated and elaborate studies on half-life extension and potency were performed for many candidates, amongst which liraglutide was deemed the best. Here, it was shown that potency tend to increase with FA chain length up to 16 C-atoms where, for longer FA chains, potency is decreased. The presence of a linker also showed an increase in potency. Furthermore, the type of linker had an influence on the potency. It was also shown that addition of a free acid in the FA chain ω tail had a tendency to decrease potency. In addition, acylation at the N-terminal of GLP-1 decreased potency, whereas acylation in the C-terminal of the peptide did not affect potency [67].

Therefore, a set of eight peptides were chosen for simulation studies, including GLP-1(7-36)-NH_2 (denoted simply as GLP-1 in this part of the thesis) and liraglutide (shortened to lira in this part of the thesis). The chosen peptides vary with respect to acylation site, number of acylation sites, and FA chain length. In addition, the presence of an L-γ-Glu linker and a free ω acid group was also varied. A description of the chosen peptides is given in table 2.3 along with potency values (EC_{50}) as reported by Knudsen et al. 2010 [67].

TABLE 2.3: The peptides investigated with regards to binding. Cxx indicates the length of the FA chain. L indicates presence of L-γ-Glu linker. A indicates presence of a free acid group. da indicates that the backbone amine has been removed on His7*. K^{xx} gives the placement of the Lys residue on which acylation occurs. If no number is given, the acylation occurs on Lys26*.

ID	FA length	Linker	ω acid	Acyl site	Mutations	des-amino	EC_{50} (pM)
GLP-1	0	✗	✗	-	-	✗	55 ± 19
lira	16	✓	✗	26	K34R*	✗	61.0 ± 7.1
C16L,$K^{26,34}$	16	✓	✗	26, 34	-	✗	$16\,700 \pm 3700$
C14A	14	✗	✓	26	K34R*	✗	72.0 ± 0.7
daC16L,K^{34}	16	✓	✗	34	K26R*	✓	2360 ± 370
daC8,K^{34}	8	✓	✗	34	K26R*	✓	236 ± 66
C12A,K^{38}	12	✗	✓	38	K26R*, K34R*	✗	4.19 ± 0.98
C14L,K^{38}	14	✓	✗	38	K26R*, K34R*	✗	54 ± 1

Schematic structures of the peptides are given in figs. 4.2 to 4.9. in chapter 2.

2.3 Aim of the computational study

The aim of this part is to look at the differences in binding to the receptor due to the differences in acylation of GLP-1. The chosen peptides should be able to disclose information on the affect of FA length, acylation site, and the number of acylations on GLP-1R binding. In addition, the chosen peptides should also be able to show if the presence of an L-γ-Glu linker or free acid group in the FA ω tail affect binding to GLP-1R.

To do this, MD simulations of complexes containing the peptides described in table 2.3 bound to GLP-1R were performed. MD gives the time resolved positions of

the atoms and can, therefore, give insights into the dynamics of the complexes and the behaviour of the individual peptides as well as the attached FA chains.

It should be mentioned that Master's student Suk Kyu Ko was responsible for starting and running the simulations as well as collecting data and perform initial analyses. I, however, made the models and membrane systems as well as created the simulation setup (described in section 4.1). Data collection was done by my instructions: RMSD, binding interactions, and MM-PBSA energies were computed using scripts created by me. Suk Kyu Ko made the scripts for creating the heatmaps from the collected interaction data. Suk Kyu Ko has presented some of the same results in his Master's thesis.

Chapter 3

MD theory

3.1 Molecular dynamics

MD presents itself as an opportune tool for investigating protein dynamics on an atomic level. It uses classical mechanics to calculate the positions and momenta of the atoms in a system. As a result, the movement of systems as large as proteins or protein complexes can be described in the time scale of nano- to microseconds [69]. Atomic fluctuations, side chain rearrangements as well as loop movement, all happen in this time scale [70, 71].

As mentioned, MD uses classical mechanics, which is obtained by integrating Newton's equations of motion in discrete time steps. For a system comprised of N particles that interact via a potential $U(\boldsymbol{r}_i)$, $i = 1, 2, 3, \ldots, N$, the equations of motion become [72]:

$$\boldsymbol{F}_i = \frac{\mathrm{d}^2 \boldsymbol{r}_i}{\mathrm{d}t^2} m_i \tag{3.1}$$

With the forces given as

$$\boldsymbol{F}_i = -\nabla_{\boldsymbol{r}_i} U(\boldsymbol{r}_i, \ldots, \boldsymbol{r}_N) \tag{3.2}$$

Here, the force is calculted as the gradient of the potential on atom i with respect to all other atoms, where $U(\boldsymbol{r_i}, \ldots, \boldsymbol{r}_N)$ is the potential which is a function of the coordinates, m_i is the mass of the i^{th} atom, and $\boldsymbol{r}_i$ is the position of the i^{th} atom. The result is a set of second order differential equations, describing the forces in three dimensions; x, y, and z. These equations are solved numerically using an integration algorithm which will give $\boldsymbol{r}_i(t)$. Such an algorithm should successively advance the systems in discrete time steps [73]. Often, such algorithms are derived from Taylor expansions of the positions $\boldsymbol{r}_i(t)$ and velocities $\boldsymbol{v}_i(t)$.

Such integration algorithms must have a set of starting positions and initial velocities. The positions can readily be taken from three dimensional structures, like those given in the protein data bank. The velocities are most often assigned from a Maxwell-Boltzmann distribution at a given temperature [73]. The outcome of solving the integration algorithms is a trajectory describing the time evolution of the coordinates and velocities of each atom in the system. To initiate an MD simulation, it is necessary to define a solvent model, ionic strength, pH, force field, and dependent on the ensemble, temperature and pressure.

3.1.1 Force fields

In MD, an atom is not represented as a nucleus with separate surrounding electrons, but is considered as a single particle, resulting in atoms modelled as spheres with point charges [74]. The forces between such atoms are then calculated by the use of a

force field. A force field is a set of equations with associated atomic specific parameters that will describe the interactions between the atoms based on their coordinates. The atom parameters in a force field are oftentimes determined experimentally or by quantum mechanics computations [75]. Most force fields are comprised of six terms, each describing different potential energies that can be categorised as either bonded or nonbonded interactions. The bonded interactions cover the intramolecular potential energies determined from the molecular geometry, such as variations in bond length, bond angle, torsion angle and improper torsions. The nonbonded interactions cover van der Waals (vdW) and electrostatic interactions between atoms that are more than two bonds away from each other, or in different molecules [69].

3.1.2 NPT ensemble

To approximate *in vivo* conditions, MD simulations of biomolecules are often run in the NPT ensemble under constant number of particles (N), pressure (P), and temperature (T). Constant temperature can be simulated by introducing a thermostat, corresponding to the addition of two terms to the equations of motion. Those terms will simulate a fictional heat bath mass that can either transfer or receive temperature (energy) [76]. Constant pressure can be obtained by letting the volume of the simulation box expand or contract [77–79].

3.1.3 Periodic boundary conditions

MD simulations of biomolecular systems are contained in simulation boxes of confined dimensions, with often only one or two biomolecules and only around $\sim$1000 to 100 000 water molecules. In order to reproduce bulk conditions and account for the interactions with the sides of the simulation box, periodic boundary conditions are introduced. Periodic boundary conditions create an infinite number of copies of the simulated system in each direction. Each copy behaves the same, meaning that if an atom exits through one face in the central box, an identical atom will enter from the opposite face [80].

Chapter 4

Binding interactions of GLP-1 analogues to GLP-1R

4.1 Material and methods

4.1.1 GLP-1 receptor

The used receptor structure is based on the crystal structure of the human GLP-1 receptor in complex with a truncated peptide agonist. The PDB ID for the structure is 5NX2 [62]. However, the structure only contains the residues from 29 to 417, covering the ECD and the TMD, but not the intracellular C-terminal helix (residues 410 to 431). Hence, a complete receptor structure was created by making a homology model using residues 29-405 from 5NX2 (chain A) and the residues 406-431 from PDB ID 4L6R (corresponding to 404-429 in the 4L6R structure) as templates. 4L6R is another GPCR class B family member with 39 % identity to GLP-1R. The homology model was constructed using the Structure Prediction Wizard in Maestro [81], and the alignment of the GLP-1R sequence to the two template structures was created using ClustalW [82, 83], also available through Maestro [81] (see table 4.1).

TABLE 4.1: Alignment of the GLP-1R sequence to the 5NX2 and 4L6R structures. Yellow highlighted residues show the template sequences; 5NX2 residues 29-405, and 4L6R residues 406(404)-431(429). The numbers in parantheses denote original 4L6R numbering. The three dots at the end of the 4L6R sequence, indicates that it continues, but the selected template part was truncated at residue 431(429) to match the length of GLP-1R.

	29....|....|....|....|....|....|....|....|....|....|....|....|...
GLP-1R (query)	TVSLWETVQKWREYRRQCQRSLTEDPPPATDLFCNRTFDEYACWPDGEPGSFVNVSCPWYLPWASSVPQG
5NX2 chain A	TVSLWETVQKWREYRRQCQRSLTEDPPPATDLFCNRTFDEYACWPDGEPGSFVNVSCPWYLPWASSVPQG
4L6R	----------MDGE----------
	.|....|....|....|....|....|....|....|....|....|....|....|....|...
GLP-1R (query)	HVYRFCTAEGLWLQKDNSSLPWRDLSECEESKRGERSSPEEQLLFLYIIYTVGYALSFSALVIASAILLG
5NX2 chain A	HVYRFCTAEGLWLQKDNSSLPWRDLSECEESKRGERSSPEEQLLFLYIIYTVGYALSFSALVIASAILLG
4L6R	----------EIEVQKEVAKMYSS-----FQVMYTVGYSLSLGALLLALAILGG
	.|....|....|....|....|....|....|....|....|....|....|....|....|...
GLP-1R (query)	FRHLHCTRNYIHLNLFASFILRALSVFIKDAALKWMYSTAAQQHQWDGLLSYQDSLSCRLVFLLMQYCVA
5NX2 chain A	FRHLHCTRNYIHLNLFASFILRALSVFIKDAALKWMYSEAAQAHQWRGLLSYQDSLSCRLVFLFMQYCVA
4L6R	LSKLHCTRNAIHANLFASFVLKASSVLVIDGLLRTLSDGAVAG----------CRVAAVFMQYGIV
	.|....|....|....|....|....|....|....|....|....|....|....|....|...
GLP-1R (query)	ANYYWLLVEGVYLYTLLAFSVFSEQWIFRLYVSIGWGVPLLFVVPWGIVKYLYEDEGCWTRNSNMNYWLI
5NX2 chain A	ANYYWLLVEGVYLYTLLAFSVFSEQWIFRLYVSIGWGVPLLFVVPWGIVKYLYEDEACWARNSNMNYWLI
4L6R	ANYCWLLVEGLYLHNLLGLATLPERSFFSLYLGIGWGAPMLFVVPWAVVKCLFENVQCWTSNDNMGFWWI
	.|....|....|....|....|....|....|....|....|....|....|....|....|...
GLP-1R (query)	IRLPILFAIGVNFLIFVRVICIVVSKLKANLMCKTDIKCRLAKSTLTLIPLLGTHEVIFAFVMDEHARGT
5NX2 chain A	IRLPILFAIGVNFLIFVRVIAIVVSKLKANLMCKTDIKCRLAKSTLTLIALLATVEVIFAFVMDEHARGT
4L6R	LRFPVFLAILINFFIFVRIVQLLVAKLRARQMHHTDYKFRLAKSTLTLIPLLGVHEVVFAFVTDEHAQGT
	.|....|....|....|....405..|....|....|....|431.|....|....|...
GLP-1R (query)	LRFIKLFTELSFTSFQGLMVAILYCFVNNEVQLEFRKSWERWRLEHLHIQRDS
5NX2 chain A	LRFIKLFTELSFTSFQGLMVAILYCFANNEVQLEFRKSW
4L6R	LRSAKLFFDLFLSSFQGLLVAVLYCFLNKEVQSELRRRWHRWRLGKVLWEERNADLEDNWETLNDNL...

This procedure resulted in a model structure based on the available 5NX2 structure and a minor addition in the C-terminal part of the receptor. Hence, the ECD and TMD is identical to the actual 5NX2 structure, which are the parts that are investigated in the following study.

4.1.2 GLP-1R–GLP-1 complex

The placement of the peptides in the receptor was based on the 5NX2 ligand and the C-terminal part of the GLP-1 ligand in the PDB ID 3IOL structure. The former is a ten residue peptide that resembles the N-terminal part of GLP-1, called Pep5, situated in the top part of the TMD of the receptor. The latter is a crystal structure of only the GLP-1R ECD in complex with GLP-1(7-35) in a straight I shape.

TABLE 4.2: Conversion scheme from pep5 to GLP-1(7-17).

Residue #	Pep5	Residue #	GLP-1
1	9DK	7	His
1	9DK	8	Ala
2	9DQ	9	Glu
3	Gly	10	Gly
4	Thr	11	The
5	9DT	12	Phe
6	Thr	13	Thr
7	Ser	14	Ser
8	Asp	15	Asp
9	9DN	16	Val
10	9DZ	17	Ser

To begin with, the residues in Pep5 were mutated into the corresponding GLP-1(7-17) residues (see table 4.2). Hereafter, the ECD from the 3IOL structure was aligned to the ECD part in the 5NX2 structure, to ensure the right orientation of the ECD to the TMD.

After this, the structure of GLP-1, with the correct orientation with regards to both the ECD and TMD, was obtained by creating a homology model. This homology model of GLP-1 residues 7-35 was constructed (same procedure as described in section 4.1.1) using residues 7-17 from the mutated Pep5 in 5NX2 and residues 18-35 from GLP-1 in 3IOL as templates. After this, the new GLP-1(7-35) structure was elongated to GLP-1(7-36) by adding an Arg residue to the C-terminal, in Maestro. Hereafter, residues 16-21, around the bend, were minimised. Finally, residues 32-36, in the C-terminal part, were minimised. At this stage, a part of the receptor ECL1 was interlocking the backbone chain of GLP-1. Hence, the ECL1 residues (205-225) were minimised until there were no clashes with GLP-1.

The outcome was a complete model structure of GLP-1(7-36)[1] in complex with the GLP-1 receptor, shown in fig. 4.1.

[1]Denoted GLP-1 in the following.

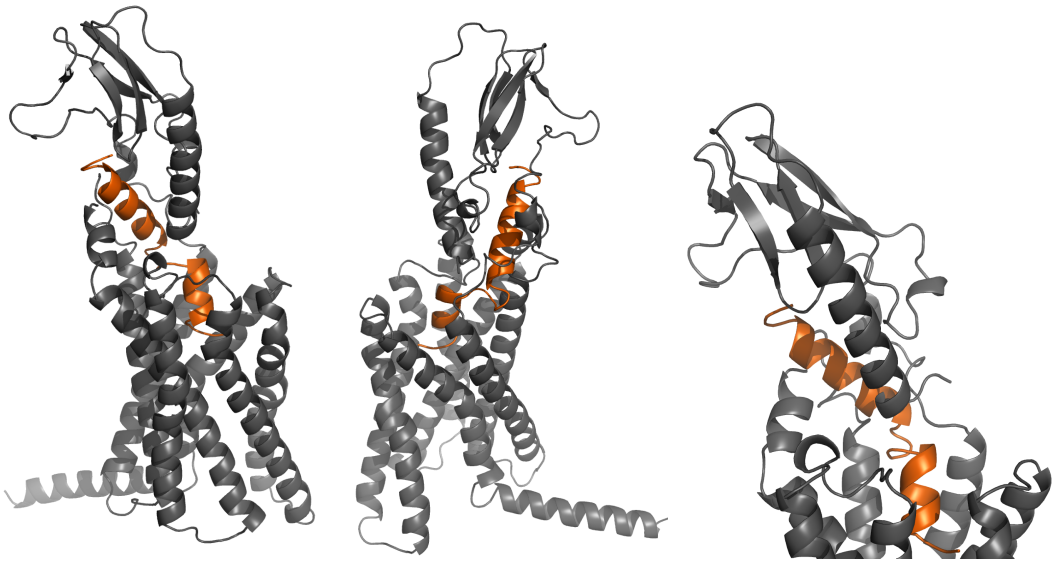


FIGURE 4.1: GLP-1 in complex with GLP-1R, as used for simulations. The receptor structure is shown in grey ribbon and the peptide in orange ribbon. The left and middle panels are the total receptor seen from either side. The right panel is a zoom in of the ECD and top TMD with GLP-1 inserted.

4.1.3 GLP-1 analogues

To create the remaining GLP-1R–peptide complexes, the GLP-1 structure (fig. 4.10 GLP-1) was modified manually in Maestro to give the seven other peptide structures (see figs. 4.2 to 4.9. for schematic structures). Liraglutide was created by aligning Lys26* in GLP-1 to the acylated Lys26* in the NMR structure of liraglutide (PDB ID 4APD) and adding the L-γ-Glu linker and C16 FA chain atoms to the GLP-1 structure. Then Lys34* in GLP-1 was mutated into Arg34*, and Gly37* was added to the C-terminal helix chain (fig. 4.10 lira). C16L,K^{26,34} was produced by altering the liraglutide structure. An extra L-γ-Glu linker with an attached 16 C-atom FA chain was added to Lys34* (fig. 4.10 C16L,K^{26,34}). C12A,K^{38} was constructed based on the GLP-1 structure where both Lys26* and Lys34* were mutated into Arg. Furthermore, the C-terminal was elongated with Arg37* and Lys38*. The FA chain atoms from liraglutide was added to Lys38*, and afterwards shortened to 12 C-atoms with an added acid group in the ω tail of the FA chain (fig. 4.10 C12A,K^{38}). C14A was constructed from liraglutide, by removing the L-γ-Glu linker and shortening the FA chain to only 14 C-atoms and adding a free acid group in the FA chain end (fig. 4.10 C14A). daC8,K^{34} was made from GLP-1, by mutating Lys26* to Arg and adding the FA chain from liraglutide to Lys34*, which was shortened to eight C-atoms. Also, the C-terminal was elongated with Arg37*, and the amino group on the His7* backbone was replaced by a H-atom (fig. 4.10 daC8,K^{34}). daC16L,K^{34} was created from daC8,K^{34} by replacing the FA chain with the L-γ-Glu linker and 16 C-atom FA chain from liraglutide (fig. 4.10 daC16L,K^{34}). C14L,K^{38} was modelled from C12A,K^{38} by exchanging the FA chain on Lys38* to the L-γ-Glu and FA chain from liraglutide, shortened to a 14 C-atom FA chain (fig. 4.10 C14L,K^{38}).

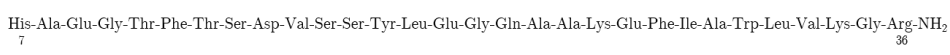


FIGURE 4.2: Schematic representation of GLP-1.

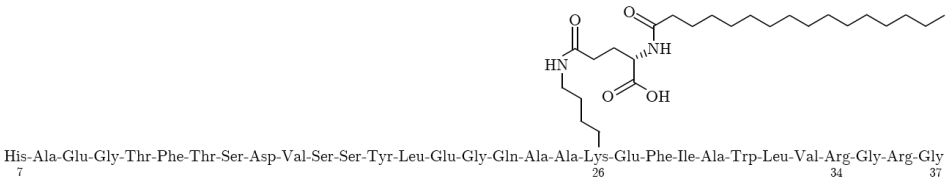


FIGURE 4.3: Schematic representation of liraglutide.

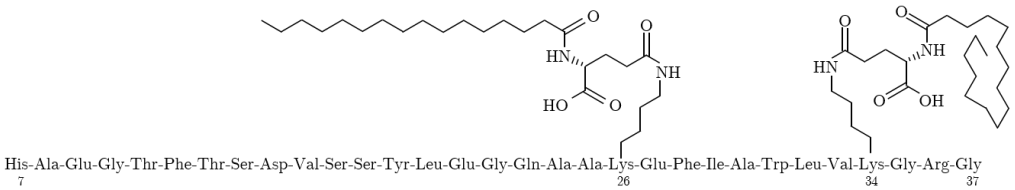


FIGURE 4.4: Schematic representation of the C16L,K^{26,34} peptide.

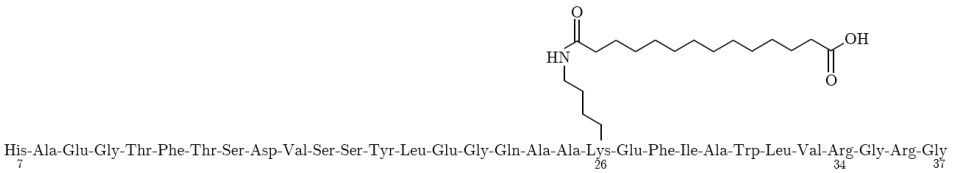


FIGURE 4.5: Schematic representation of the C14A peptide.

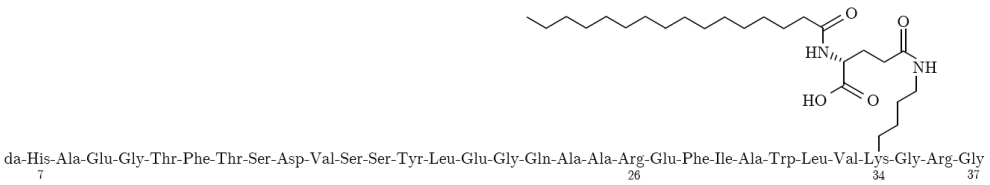


FIGURE 4.6: Schematic representation of the daC16L,K^{34} peptide. The N-terminal amino group on His7* has been removed, which is denoted by 'da' in the N-terminus.

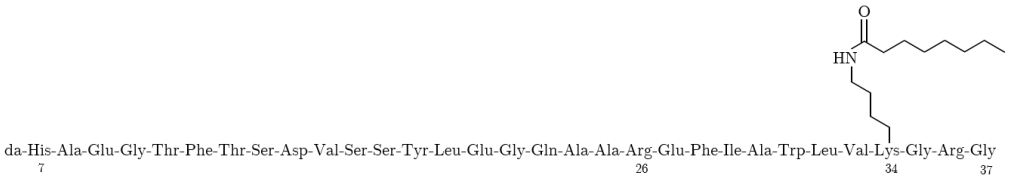


FIGURE 4.7: Schematic representation of the daC8,K^{34} peptide. The N-terminal amino group on His7* has been removed, which is denoted by 'da' in the N-terminus.

FIGURE 4.8: Schematic representation of the C12A,K^{38} peptide.

FIGURE 4.9: Schematic representation of the C14L,K^{38} peptide.

This gave eight GLP-1R–peptide systems with peptide structures docked manually into the receptor based on partial alignments to crystal structures. As a note, it should be mentioned that the complexes containing liraglutide, C16L,$K^{26,34}$, and C14A (the three peptides with acylation site at Lys26*) all had an artificial clash between the FA chain and the side chain of Tyr88 in the receptor. As a solution to this problem, the FA chain on these three peptides were moved manually out of the Tyr88 ring. Afterwards, the entire Lys26* residue was minimised.

In addition to the eight GLP-1R–peptide complexes, a system with no peptide was considered. This system is referred to as empty. The same receptor model was used, giving an empty activated receptor structure. This resulted is nine systems in total. The only difference between the systems is the docked peptide structure. Therefore, only the peptide structures are shown in fig. 4.10, since fig. 4.1 represents the peptide orientation that applies to all peptides considered in this study.

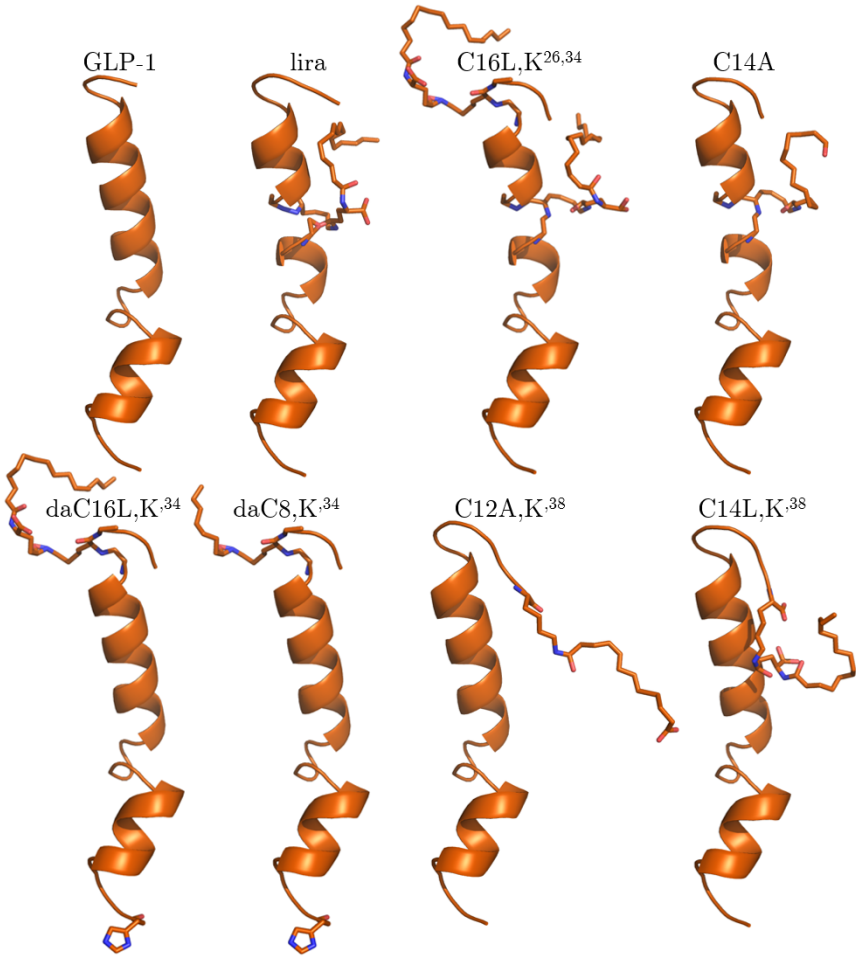


FIGURE 4.10: The eight peptides. The peptides are shown in orange ribbon with the acylated FA chains in orange sticks. For daC8,K^{34} and daC16L,K^{34} the des-amino backbone in His7* is also shown in sticks.

At this stage, all eight receptor-peptide systems are described by two individual pdb files, one containing the receptor coordinates, and one containing the peptide coordinates. The two pdb files were merged into one, and a structure file was generated using psfgen [84], a VMD [85] plugin. The 4 disulfide bonds in the receptor (Cys46-Cys71, Cys62-Cys104, Cys226-Cys296, and Cys126-Cys85) were given as patches[2]. This was also done for the empty receptor, but obviously no peptide structure was added. Thus, the nine systems were now described by one pdb file and one psf file, and ready for simulation preparation.

4.1.4 Membrane setup

Next, the complexes should be inserted into a membrane patch, but firstly, it is important to ensure that the added TIP3 water molecules contain an extra bond between the two hydrogens. This is done by adding the otherwise commented line:
`BOND OH2 H1 OH2 H2 H1 H2`
in the files `wat.top` and `top_all36_lipid.rtf`, and make the line:
`BOND OH2 H1 OH2 H2`
into a comment.

[2]See Appendix A for psfgen input file exemplified for liraglutide.

The first file is located in the directory `vmd/plugins/noarch/tcl/solvate1.6/`, and the second one in `vmd/plugins/noarch/tcl/readcharmmtop1.1/`. This ensures that all water molecules are described by three bonds, as needed when using SHAKE in AMBER[3].

Then the membrane setup can begin: The total system was inserted into a membrane patch of 80 Å × 80 Å in the x-y plane consistent of a 3-palmitoyl-2-oleoyl-D-glycero-1-phosphoatidylcholine (POPC) lipid bilayer. This was done based on the procedure described in the 'Membrane proteins tutorial' [86] using the file `memprot_align.tcl` shown in Appendix B. The flag `-top` followed by `c36`, was given to define CHARMM36 lipid naming. The positioning of the GLP-1R molecule into the membrane patch was done manually, so that the 7TM barrel is upright in the membrane with the intracellular α-helix tail running along the inside of the membrane and the top part of the barrel extruding the membrane on the extracellular side followed by the ECD. `remove_lipwat.tcl` was then used to remove lipid and water molecules that overlaps with the inserted protein complex. Since CHARMM36 naming was used to generate the membrane bilayer, the phosphate naming `P1` in the tutorial file was changed to `P` (see Appendix C).

4.1.5 Solvation and addition of ions

The next step was to solvate the entire system. This was done with use of the program Solvate [87] using the TIP3 water model [88]. The size of the solvation box was set to be 15 Å below, and 50 Å above the membrane, ensuring 10 Å of water above the top of the extracellular domain and below the intracellular α-helix. This ensures no self-interactions when using periodic boundary conditions. The size confining the above mentioned space is noted as the waterbox size in `solv_remove.tcl`, as well as the free distance between the protein and lipid molecules, which was set to 1.5 Å. An example of the `solv_remove.tcl` file is given in Appendix D.

After this, Na^+ ions were added to the water layers to counter the negative charges on the peptides and receptor. Then NaCl molecules were added to give an ionic strength of 100 mM in the water layers. This step was performed using the VMD plugin Autoionize [89].

The result was eight receptor-peptide complexes and one empty receptor, all inserted into a membrane patch with a water layer on either side of the membrane bilayer. The GLP-1R–GLP-1 system is shown as an example in fig. 4.11.

[3]See sections 4.1.6 and 4.1.7 for conversion from VMD/NAMD to AMBER and simulation details.

TABLE 4.3: The total number of atoms, water molecules, and NaCl molecules, as well as the sizes of the eight receptor-peptide complexes and the empty receptor. The systems are denoted by the peptide IDs. This notation will be used in the following.

System	No. of atoms	No. of waters	No. of NaCl	Box size (Å×Å×Å)
GLP-1	63226	13645	26	$83.5 \times 80.3 \times 121.0$
lira	63280	13640	26	$83.5 \times 80.3 \times 121.0$
C16L,K^{26,34}	63305	13628	26	$83.5 \times 80.3 \times 121.0$
C14A	63267	13643	26	$83.5 \times 80.3 \times 121.0$
daC16L,^{34}	63251	13631	26	$83.5 \times 80.3 \times 121.0$
daC8,K^{34}	63232	13638	26	$83.5 \times 80.3 \times 121.0$
C12A,K^{38}	63245	13630	26	$83.5 \times 80.3 \times 121.0$
C14L,K^{38}	63306	13643	26	$83.5 \times 80.3 \times 121.0$
empty	62886	13685	26	$83.5 \times 80.3 \times 121.0$

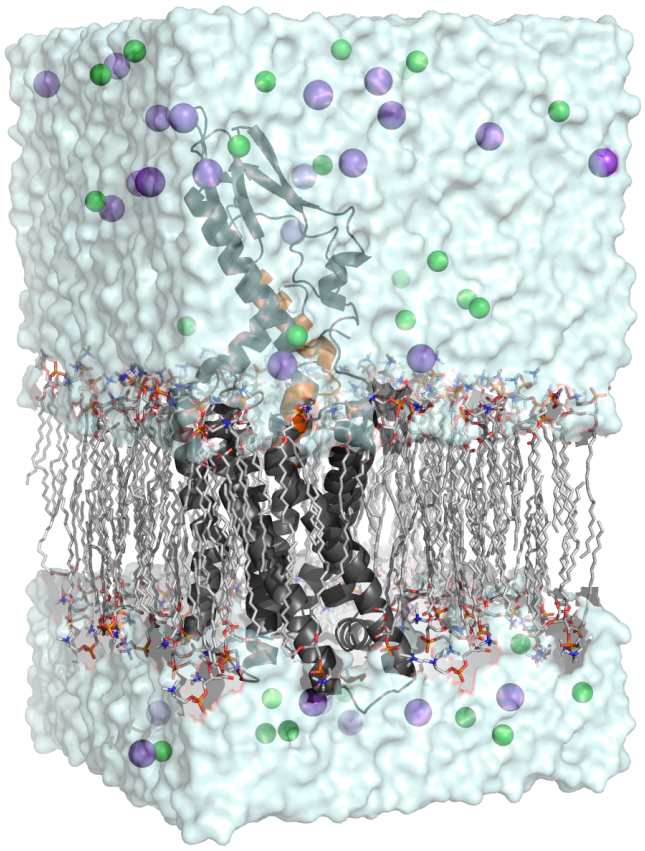


FIGURE 4.11: The GLP-1 system inserted into a membrane bilayer patch, solvated and with Na^+ and Cl^- ions. Receptor structure is shown in dark grey ribbon and GLP-1 in orange ribbon. The membrane bilayer is shown in sticks with C-atoms in light grey, phosphate in orange, oxygen in red, and nitrogen in blue. The water solvent layers are shown as transparent surfaces. Na^+ and Cl^- ions are shown as purple and green spheres, respectively.

The size and number of atoms, water molecules, and NaCl molecules in each system are given in table 4.3.

The systems are now described by a structure file and a pdb coordinate file, using CHARMM36 parameter and topology files, and are ready for minimisation, equilibration, and production run in NAMD. However, the simulations were performed in AMBER, see section 4.1.6 for an elaboration on this.

4.1.6 From CHARMM to AMBER

Since I previously worked with NAMD [90], I have made several parameter and topology adaptations for the special residues (acylated Lys residues and des-amino N-terminal His[4], using acetelyted Lys ('ALY') in the CHARMM36 force field [91]. However, our group acquired a GPU cluster, which NAMD poorly exploits. On the other hand, AMBER [92] is optimised for use on GPUs [93, 94]. Therefore, a conversion scheme from psf/pdb and CHARMM topology (`.rtf`)/parameter (`.prm`) files to coordinate (`.rst7`)/ parameter (`.parm7`) AMBER input files were needed, inherently maintaining the added changes for the special residues using the CHARMM36 force field.

Such conversion was possible by the use of the program Chamber [95] within ParmEd [96]. ParmEd is a program developed to edit parameter files, in general, and Chamber (a part of ParmEd), specifically transforms CHARMM setup files to match the input necessary for running AMBER.

When in ParmEd, Chamber was called and given the topology files `top_all36_prot_glp1a.rtf`, `top_all36_lipid.rtf` with the flag `-top` in front of both, and the parameter files `par_all36_prot_glp1a.prm`, `par_all36_lipid.prm`, with the flag `-param` in front of each of them. To cover the water molecules and ions, the string file `toppar_water_ions.str` was given with the flag `-str`. The structure and coordinate files; `filename-ionized_moved.psf` and coordinate file `filename-ionized_moved.pdb` were then given with the flags `-psf` and `-crd`, respectively:

```
$ parmed
> chamber —top top_all36_prot_glp1a.rtf —top top_all36_lipid.rtf —param
    par_all36_prot_glp1a.prm —param par_all36_lipid.prm —str
    toppar_water_ions.str —psf filename—ionized_moved.psf —crd filename—
    ionized_moved.pdb
```

After this, the residue numbering from the pdb file was added to the input:

```
> addpdb filename—ionized_moved.pdb
```

Then, the chamber parameter/topology and coordinate *output* files (equal to the AMBER *input* files) were generated:

```
> outparm filename.parm7 filename.rst7
```

The newly generated AMBER coordinate file will, per default, give the box dimensions as 0 in all directions, which should be corrected to span the actual size of the water box. This can be done using the script `vmd_box_dims.sh`[5] to measure the size of the water in the original pdb file:

```
$ vmd_box_dims.sh —i filename—ionized_moved.pdb —s water
```

[4]See Appendix E for changes to topology and parameter files, respectively.

[5]See Appendix F for script file.

The outputted size dimension should then be added to the last line of the generated `.rst7` file instead of 0 for the x, y, and z lengths.

The result will be two AMBER input files, namely `filename.parm7` and `filename.rst7`, containing the force field parameters from CHARMM36 (including the manually added ones for the special residues), which are ready to use for simulations on GPUs in AMBER.

4.1.7 Simulation setup

The systems were treated according to the AMBER tutorial on MD simulations of lipid bilayers [97]. In AMBER [92], the systems were initially minimised. Each system was simulated in quadruplets differing only in the number of initial minimisation steps. Hence, four sets of eight GLP-1R–peptide complexes and one empty GLP-1 receptor, totalling in 36 simulations were run. Each system was minimised for 5000, 7000, 10 000, or 13 000 steps to obtain different starting conditions. During the first 5000 steps, the steepest decent method was used, while the conjugate gradient method was used for the remaining steps of the minimisations. Output is written every 500 steps (see Appendix G for minimisation input file). In the following, the individual systems will be referred to as the system ID followed by the minimisation step number in parenthesis, meaning that the simulation of the GLP-1R–GLP-1 complex started from the minimisation protocol of 7000 minimisation steps, will be written as GLP-1R–GLP-1(7000).

Hereafter, the systems were subjected to a heating ramp in two steps. First, the lipids were held constant with a force constant of 10 kcal/(mol Å^2), while the system was heated from 0 K to 100 K using a Langevin thermostat [76]. This heating procedure was run for 2500 steps with a timestep of 2 fs. The SHAKE algorithm [98–100] was used to restrain hydrogen bonds, and again, the volume was held constant. Since, the CHARMM36 force field [91] was used, the nonbonded cut-off distance was set to 12 Å with a switching function for the Lennard-Jones potential from 10 Å to 12 Å. The particle-mesh-Ewald (PME) method [101, 102] was used for evaluation of long range electrostatics. The nonbonded list (pair list distance) was 14 Å and was updated whenever an atom had moved more than 1 Å since the last list update. A frame was saved to the output trajectory file every 100$^{\text{th}}$ step. See Appendix H for input file for heating ramp protocol. The second heating went from 100 K to 303 K, but this time under constant pressure at 1 bar, controlled anisotropically (since the system contains a membrane), using the Berendsen barostat [79]. Otherwise, the same input parameters used in the first heating were used here as well. The second heating was done for 50 000 steps. See Appendix I for input file for the second heating ramp.

Next, the systems were equilibrated for 25 000 steps ten times, one after the other. This was done at constant NPT (constant number of particles, pressure, and temperature) at 303 K and 1 bar, again controlled using a Langevin thermostat and Berendsen barostat respectively. The nonbonded list (pair list distance) was increased to 17 Å and was updated whenever an atom had moved more than 2.5 Å since the last list update. A frame was saved to the trajectory file every 5000$^{\text{th}}$ step. Otherwise, the same input parameters used in the second heating were used here as well. See Appendix J for equilibration input file.

Finally, the production was performed with constant NPT, 303 K and 1 bar, controlled using a Langevin thermostat and a Monte Carlo barostat respectively [77, 78]. The nonbonded list was decreased again to 14 Å and was updated whenever an atom had moved more than 1 Å since the last list update. A frame was saved to

the trajectory file every 5000$^{\text{th}}$ step. Otherwise, the same input parameters used in the second heating were used here as well. The production run was continued for 100 000 000 steps corresponding to 200 ns. See Appendix K for production run input file.

Each of the above mentioned steps were submitted to the DTU Chemistry STENO GPU cluster using five different submission scripts. They are shown in Appendix L.

4.1.8 MM-PBSA binding free energies

Binding free energies were calculated using the molecular mechanics-Poisson-Boltzmann surface area method (MM-PBSA). A more detailed description on how to compute these is given in chapter 5.

Internal energies (bond, angle, dihedral, van der Waals, and electrostatic) were calculated using the VMD plugin, NAMDEnergy, with the CHARMM36 force field and PME method. Non-polar solvation energies were calculated from the solvent accessible surface area (SASA), applying the linear relation given in eq. (4.1).

$$G_{\text{solv,non}-\text{polar}} = \gamma \text{SASA} + \beta \tag{4.1}$$

With the surface tension, $\gamma = 0.005\,42\,\text{kcal}/(\text{mol}\,\text{Å}^2)$ and $\beta = 0.92\,\text{kcal}/\text{mol}$ [103]. SASA was computed in VMD using a probe radius of 1.4 Å correspondant to the spherical size of a water molecule. The polar solvation energy was calculated using APBS (Adaptive Poisson-Boltzmann Solver) [104]. Here, dielectric constants of 80 for water (solvent) and 1 for protein (solute) were used, together with a probe radius of 1.4 Å. PQR files were generated using PDB2PQR [105, 106]. Script and input files are shown in Appendix M.

All of these energetic contributions were calculated for 400 frames, 0.5 ns apart, covering the last 100 ns of the simulations.

Entropy calculations were investigated for convergency as described in section 5.4.1. The GLP-1R–C16L,K^{26,34}(7000) simulation was chosen as a representative of all the systems, and entropy values were calculated using an increasing number of frames. The result is plotted in fig. 4.12.

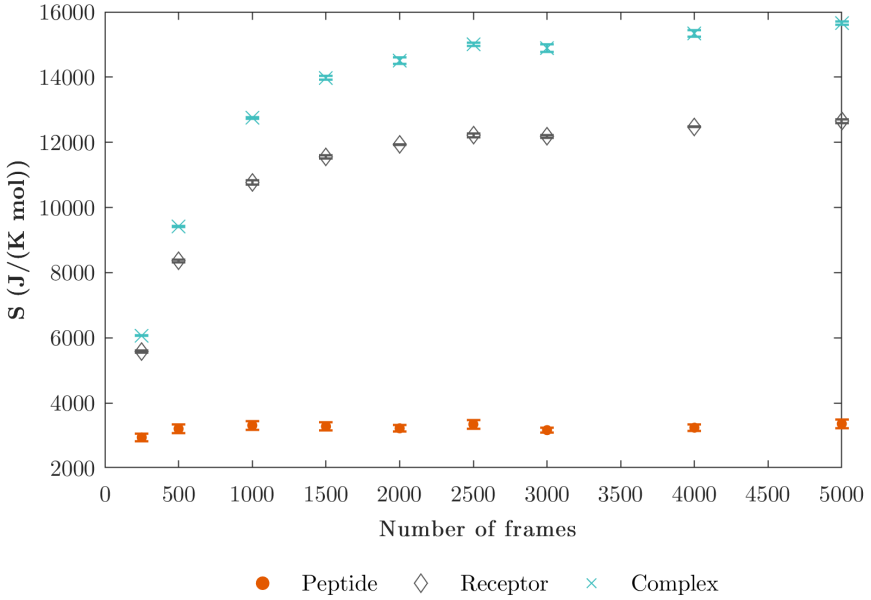


FIGURE 4.12: Test for entropy convergency: Absolute entropy as a function of frames used. Values plotted are averages calculated from two batches. Entropy values are calculated for the GLP-1R–C16L,$K^{26,34}$(7000) simulation.

From fig. 4.12, it is visible that the entropy converges when using 2500 frames in each of the two batches.

The standard deviation on the entropy differences of binding, $T\Delta S$, were calculated for 2, 3, and 4 batches with 2500 frames in each batch, resulting in SDs varying between ∼10.4 and 11.4 kcal/mol, which should be seen in the context of the absolute entropy of 15 000 J/(mol K) for the complex. Therefore, two batches will be used. Hence, all entropy values were calculated by use of two batches with 2500 frames in each batch. The frames were evenly spread out over the last 100 ns with a step size of 20 ps. When computing the entropy values, all calculations were simplified by using only the C-α atoms in the receptor and the peptide, except for the acylated residues where all C-atoms were included. The entropy values were calculated directly from the trajectory files using the three scripts shown in Appendix N, Appendix O, and Appendix P. The bash script (Appendix N) calls a VMD (Appendix O) script and a Python script (Appendix P) and a program called ProDy [107, 108].

4.2 Results and discussion

In this section, several analyses of the obtained simulations are presented, and the emerging results are discussed. Calculations are made for every 50$^{\text{th}}$ frame. This gives 400 frames, 0.5 ns apart, evenly distributed over the simulated 200 ns. This setup is mainly used for preliminary studies of the simulations, since analyses should be kept to the equilibrated part of the simulations. If the total trajectory *is* considered, it is clearly illustrated as a simulation time axis. It turns out that all systems are well equilibrated[6] after 100 ns, equalling to 200 frames taken only from the last

[6]Based on time resolved root mean square deviation and number of binding interactions throughout the simulations.

100 ns of the simulations. For an in depth justification of the chosen time interval, see section 4.2.1.

4.2.1 Structure evaluation

As the first step in the investigations of the simulated receptor systems, it is sensible to set up an evaluation scheme. On this note, the simulated systems are evaluated on three points: Overall stability, the correlation between binding free energy and experimentally measured potency for the different receptor–peptide systems, and lastly, captured interactions between the receptor and GLP-1, during the simulations, as reported previously.

Overall stability

A good estimate of the overall stability of the systems can be found by taking a look at the total root mean square deviation (RMSD) for the whole system (receptor plus peptide backbone atoms) and the total number of interactions between the receptor and the peptide. Starting with the RMSD, the criteria are convergence and a low deviation. Inspection of the RMSD plotted in fig. 4.13 shows overall convergence for all nine simulations, with a maximum deviation of 6 to 7 Å.

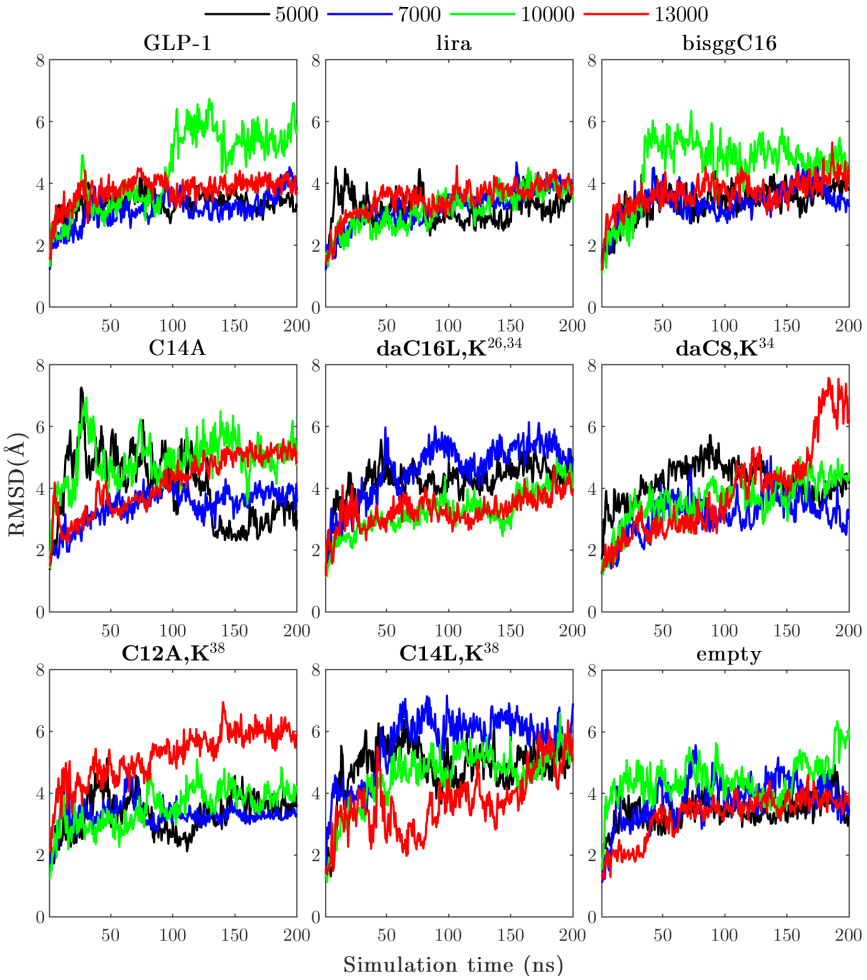


FIGURE 4.13: The total RMSD for all nine systems, calculated for the receptor plus the peptide backbone atoms (only the receptor for the empty system).

In two cases, the RMSD is increasing towards the end of the simulation (the daC8,K^{34}(13000) simulation, and the empty receptor, (10 000 steps) simulation. However, when comparing all four trajectories for each system, we see convergence. Figure 4.13 also shows that the RMSD converges after 100 ns for all systems.

After scrutinising the overall movement of the systems, it is natural to also consider the total number of interactions between the receptor and the peptides during the simulations. This will give an idea of whether or not the total interaction map also seems to stabilise. Hence, the total number of receptor–peptide interactions are plotted in fig. 4.14.

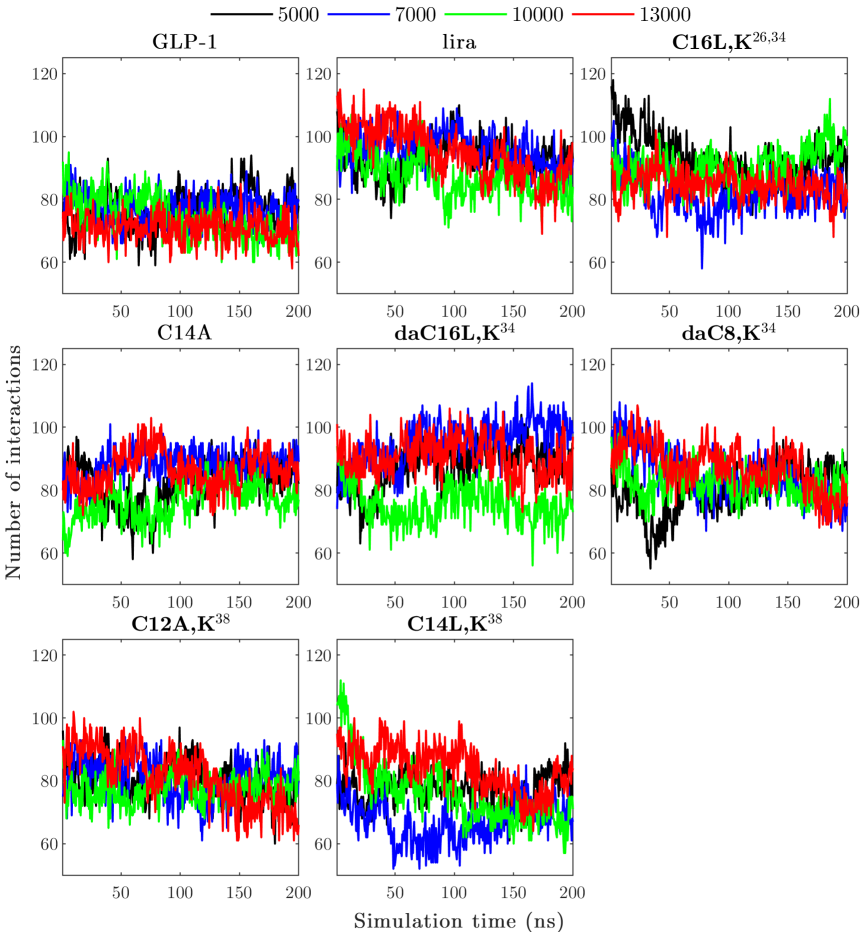


FIGURE 4.14: The total number of interactions between the receptor and the peptides are shown as a function of simulation time. An interaction is counted when a residue in the receptor is within 3 Å of a residue in the peptide. The empty receptor is not included in this study.

Here, an interaction is considered as two residues, one in the receptor and one in the peptide, being within 3 Å of each other. We see three cases of non-convergence; namely the daC8,K^{34}(13000) simulation, the C12A,K^{38}(13000) simulation, and the C14L,K^{38}(10000) simulation. However, as for the RMSD, it is only for one of the four repetitions, so overall the number of interactions during the simulations are either close to constant or converge after 100 ns. Therefore, all analyses will be calculated for the last 100 ns of the simulations henceforth.

All in all, the RMSD and total number of interactions indicate that all the systems are stable after 100 ns of simulation time and can be evaluated further.

Binding free energy and potency

Having looked at the more individual trends of the different systems, it would be prudent to consider a more overall trend amongst the simulations, prefereably one that can be related to experimental studies. Comparing calculated binding free energies with experimenatally measured potency values can do exactly this. Therefore, the binding free energies, ΔG_{bind}, are calculated using the MM-PBSA approach for the eight receptor-peptide systems. These are then compared to the potency values for

the peptides reported by Knudsen et al. 2000 [52]. In fig. 4.15, the calculated binding free energies and the potency values (in the form of pEC_{50} = -log(EC_{50}) [109]) are plotted against each other. The plotted energies are determined as the average of the MM-PBSA energies determined for the last 100 ns of each of the four simulations for each system.

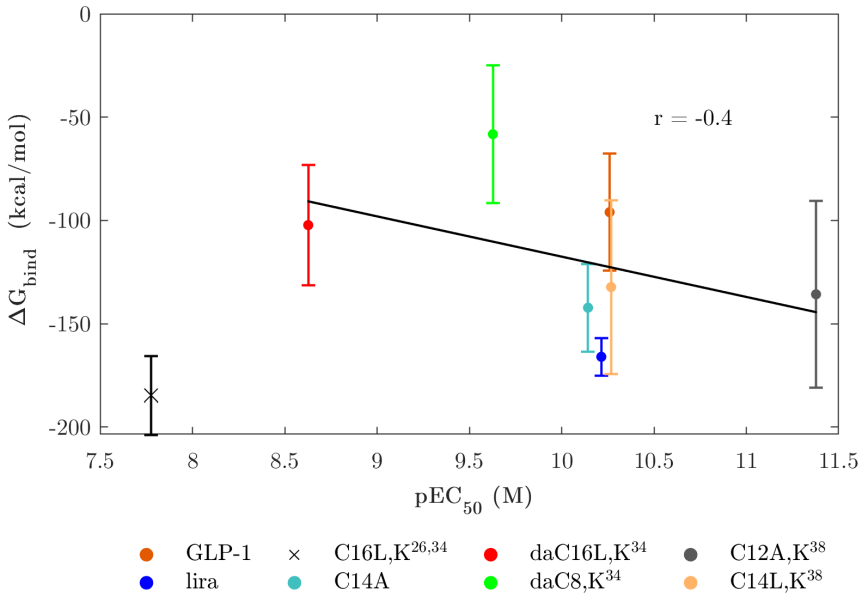


FIGURE 4.15: The calculated binding energies plotted against the potency values (shown as pEC_{50} = -log(EC_{50}). The C16L,K^{26,34} system (black ×) is not included in the calculated regression line and correlation coefficient, r, shown in the plot.

The C16L,K^{26,34} system is regarded as an outlier since it is the only peptide with two acylated FA chains, and because the measured potency is so much lower than the others: 16 700 pM compared to values between 4.19 pM and 2360 pM. Therefore, the C16L,K^{26,34} point is not considered when calculating the correlation coefficient.

Disregarding C16L,K^{26,34}, the correlation coefficient is -0.4 between the calculated energies and pEC_{50} values. This means that the potency values (EC_{50}) correlate positively with the calculated MM-PBSA energies and the measured potency values. That is, for a peptide that binds strongly, both the EC_{50} value and the binding energy is low. This adds to the validation of our simulated systems and justifies further investigations of the different GLP-1R–peptide complex simulations.

Important GLP-1R–GLP-1 interactions

To further affirm the simulations, one could consider specific interactions found in studies of GLP-1R. Since GLP-1R only has been structurally studied in complex with GLP-1 and non-acylated GLP-1 analogues, it is natural to use the simulated GLP-1 system as a control in regards to captured important interactions.

In chapter 2, a set of important interactions between GLP-1 and GLP-1R are given based on structural studies of GLP-1R bound to GLP-1 and other peptides. To ensure that the simulations are able to map the same interactions, the GLP-1R–GLP-1 system will be investigated. The interactions in question are (re)shown in table 4.4.

TABLE 4.4: Interactions between GLP-1 and GLP-1R as reported by Jazayeri et al. 2017 (J), Santiago et al. 2018 (S), Underwood et al. 2010 (U), and Zhang et al. 2017 (Z). † indicates residue cluster formation rather than individual interaction pairs. Potency reduction from mutagenesis studies is given as reviewed by de Graaf et al. 2016 [25]. For a given receptor residue, only the mutation causing the highest potency reduction is shown.

GLP-1	GLP-1R	Source	Potency reduction (x-fold)
His7*	Arg299 (ECL2)	Z	85 (Arg299Ala)
His7*	Trp306 (TM5)	Z	206 (Trp306Ala)
His7*	Glu364 (TM6)	S	15 (Glu364Ala)
His7*	Glu387 (TM7)	J, S	10 (Glu387Asp)
Ala8*	Leu384 (TM7)	J	41 (Leu384Ala)
Ala8*	Leu388 (TM7)	J	208 (Leu388Ala)
Glu9*	Tyr145 (TM1)	Z	<5 (Tyr145Ala)
Glu9*	Tyr148 (TM1)	S	26 (Tyr148Ala)
Glu9*	Tyr152 (TM1)	J, S	<5 (Tyr152 His)
Glu9*	Arg190 (TM2)	J, S, Z	270 (Arg190Ala)
Glu9*	Leu388 (TM7)	S, Z	208 (Leu388Ala)
Glu9*	Ser392 (TM7)	S, Z	<5 (Ser392Ala)
Gly10*	Trp306 (TM5)	J	109 (Trp306Ala)
Thr11*	Arg299 (ECL2)	S, Z	85 (Arg299Ala)
Thr11*	Asp372 (ECL3)	J	59 (Asp372Ala)
Thr11*	Leu384 (TM7)	J	41 (Leu384Ala)
Phe12*	Leu141 (TM1)	J	<5 (Leu141Ala)
Phe12*	Leu142 (TM1)	J	Not measured
Phe12*	Tyr145 (TM1)	Z	<5 (Tyr145Ala)
Phe12*	Tyr148 (TM1)	J	26 (Tyr148Ala)
Phe12*	Leu384 (TM7)	J	41 (Leu384Ala)
Phe12*	Phe385 (TM7)	Z	<5 (Phe385Ala)
Phe12*	Leu388 (TM7)	J	208 (Leu388Ala)
Thr13*	Lys197 (TM2)	J, Z	630 (Lys197Ala)
Thr13*	Leu201 (TM2)	J	Not determined
Thr13*	Phe230 (TM3)	J	Not measured
Thr13*	Met233 (TM3)	J	70 (Met233Ala)
Thr13*	Trp297 (ECL2)	S	316 (Trp297Ala)
Thr13*	Thr298 (ECL2)	J	<5 (Thr298Ala)
Ser14*	Asn300 (ECL2)	J	501 (Asn300Ala)
Ser14* †	Trp297 (ECL2)	Z	316 (Trp297Ala)
Ser17* †	Thr298 (ECL2)	Z	<5 (Thr298Ala)
Ser18* †	Arg299 (ECL2)	J, Z	85 (Arg299Ala)
Asp15*	Arg380 (TM7)	J, S	1853 (Arg380Asp)
Val16*	Pro137 (E-TM)	J	Not measured
Val16*	Leu141 (TM1)	J	<5 (Leu141Ala)
Val16*	Leu142 (TM2)	J	Not measured
Val16*	Tyr145 (TM1)	J	<5 (Tyr145Ala)
Val16*	Leu201 (TM2)	J	Not determined
Val16*	Lys202 (TM2)	J	Not measured
Val16*	Tyr205 (TM2)	J	62 (Tyr205Ala)

Continued on next page

Table 4.4 – *Continued from previous page*

GLP-1	GLP-1R	Source	Potency reduction (x-fold)
Ser17*	Ser31 (ECD)	J	Not measured
Ser17*	Leu32 (α)	J	<5 (Leu32Ala)
Ser17*	Trp33 (α)	J	Not measured
Ser17*	Met204 (TM2)	J	334 (Met204Ala)
Ser17*	Tyr205 (ECL1)	J	62 (Tyr205Ala)
Ser17*	Leu217 (ECL1)	J	<5 (Leu217Ala)
Ser17*	Gln221 (ECL1)	J	Not measured
Ser17*	Gly295 (ECL2)	J	<5 (Gly295Ala)
Ser17*	Thr298 (ECL2)	J	<5 (Thr298Ala)
Leu20*	Leu201 (TM2)	S, Z	Not determined
Leu20*	Met204 (TM2)	S, Z	334 (Met204Ala)
Glu21*	Arg299 (ECL2)	S	85 (Arg299Ala)
Lys26*	Glu128 (ECD)	U	<5 (Glu128Ala)
Trp31*	Gln211 (ECL1)	S, Z	<5 (Gln211Asp)
Trp31*	His212 (ECL1)	S, Z	<5 (His212Ala)
Val33*	Arg121 (ECD)	U	<5 (Arg121Ala)

The important interactions between GLP-1 and GLP-1R are mainly located in two different areas of the two participants. One area is constituted by the C-terminal/ middle part of GLP-1 and the ECD/ECD-to-TMD linker (E-TM). The other area is made up of the N-terminal part of GLP-1 and the top of the TMD structure, including the three extracellular loops.

To shed some light on the binding interactions between GLP-1R and GLP-1 in the simulations, the identified residue pair interactions (GLP-1R and GLP-1 residues within 3 Å of each other) are monitored as a function of simulation time, i.e. frames. Since there are four replicas, there will also be four sets of interacting pairs with associated frame counts. These frame counts are summed to give one total frame count per identified interaction pair. The interactions are then plotted as specific pairs in a heatmap, where the colourscale will disclose the strength of the specific interaction. The higher the frame count for a specific interaction pair, the stronger the interaction is regarded. Interactions occurring less than 25 % of the total frames are not shown, and are interpreted as negligible. A frame count corresponding to 25 % to 75 % is considered as weak to medium interactions, and 75 % or above, as strong interactions.

Starting with the TMD and GLP-1 area, the occurring interactions in this region are mapped in fig. 4.16. The heatmap (left panel) shows the interaction strength of the individual residue pairs as a total sum of the four simulation repetitions. The middle and right panels illustrate the interface of binding, with the interaction pairs highlighted in the structure according to the interaction strength.

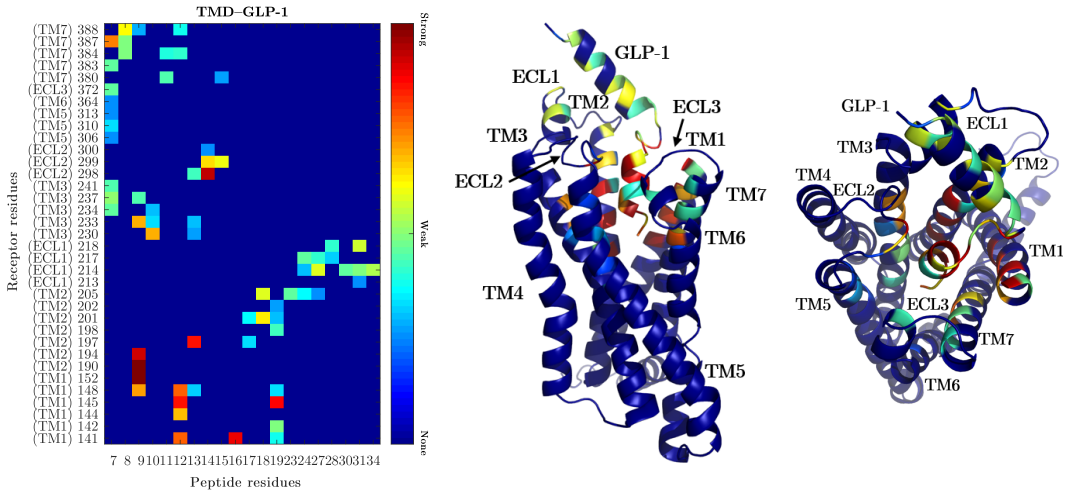


FIGURE 4.16: Interaction strength of the interacting residue pairs for the TMD and GLP-1. In the left panel is a plot with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. To illustrate the location of the binding interface between the receptor TMD and GLP-1, the structure (image taken from the last frame of the GLP-1R–GLP-1(7000) simulation) is shown from the side (middle) and top (right) with the residues coloured according to the interactions shown in the heatmap.

From the interaction plot in fig. 4.16, it is evident that His7* interacts with Glu364 in TM6 and Glu387 in TM7. This corresponds well with the fact that these particular interactions have been mapped as key interactions between GLP-1R and GLP-1 [36, 62]. From fig. 4.17, it is evident that both interactions are with the backbone amino group, rather than the His side chain. The simulations also capture interactions between His7* and other residues in TM3, TM7 and ECL3. However, those interactions are weak and might highlight the importance of the His7* side chain charge as an orientation facilitating residue, interacting with a 'cage' of charged residues in the TM3, TM7, and ECL3, to ensure that the peptide N-terminal stays inserted in the transmembrane barrel. The simulations also capture interactions between Ala8* and Leu384 and Leu388, both in TM7 as previously pointed out [62], see fig. 4.17.

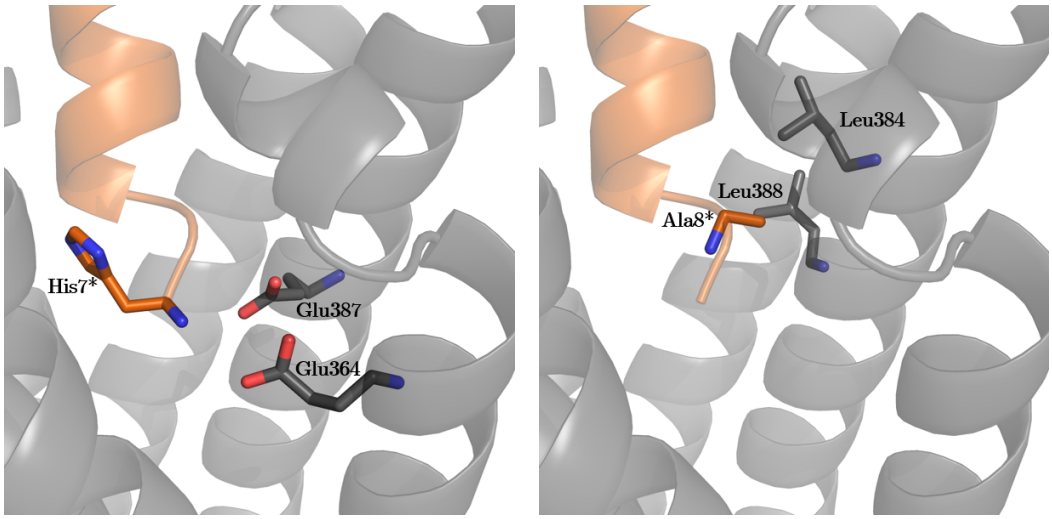


FIGURE 4.17: Interactions between His7* and Glu364/Glu387 (left) and between Ala8* and Leu384/Leu388 (right). Image taken from the last frame of the GLP-1R–GLP-1(7000) simulation.

Concerning Glu9*, the simulations capture several interactions, including the strong one to Arg190 in TM2, which is the only interaction that has been reported by three different studies [36, 62, 63]. Glu9* also interacts strongly with Tyr152 in TM7, which has been reported twice [36, 62], even though mutation studies of Tyr152 shows no significant (less than 5-fold) potency reduction [25]. Another interaction that the simulations have captured is that of Glu9* to Tyr148 in TM1 [36]. The interactions to Glu9* are shown in fig. 4.18.

The simulations only capture one interaction with Gly10*, which is to Phe230 in TM3, and not to Trp306 in TM5, as previously seen [62]. Gly10* is shown in fig. 4.18 together with both residues. Mutation of Phe230 does, however, not cause any significant potency reduction [25]. Our simulations thus could indicate that it is not an actual binding interaction that Gly10* obtains, but rather an acquired flexibility in the N-terminal end of the peptide backbone structure.

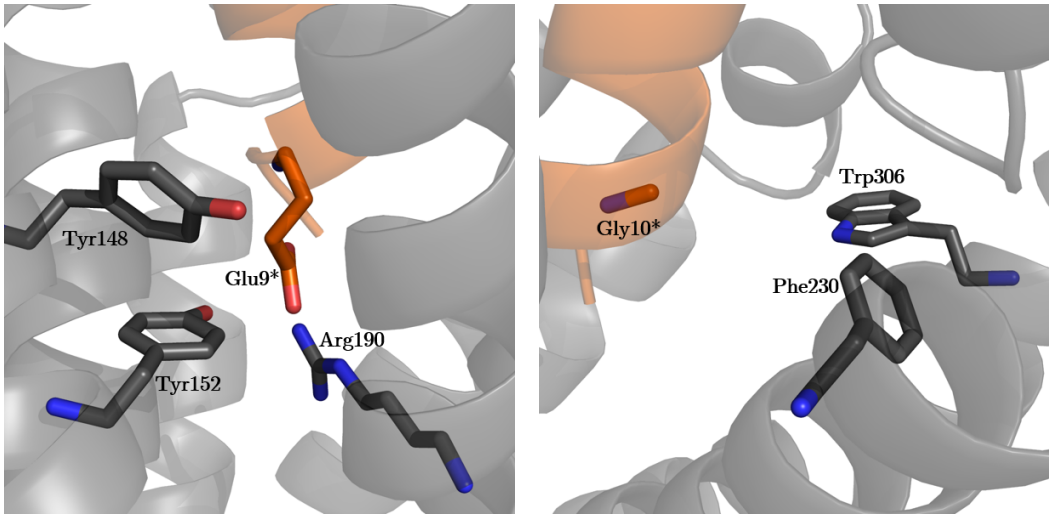


FIGURE 4.18: To the left, the interactions between Glu9* and Arg190, Tyr152, and Tyr148. To the right, the interaction between Gly10* and Phe230 rather than Trp306 (also shown). Image taken from the last frame of the GLP-1R–GLP-1(7000) simulation.

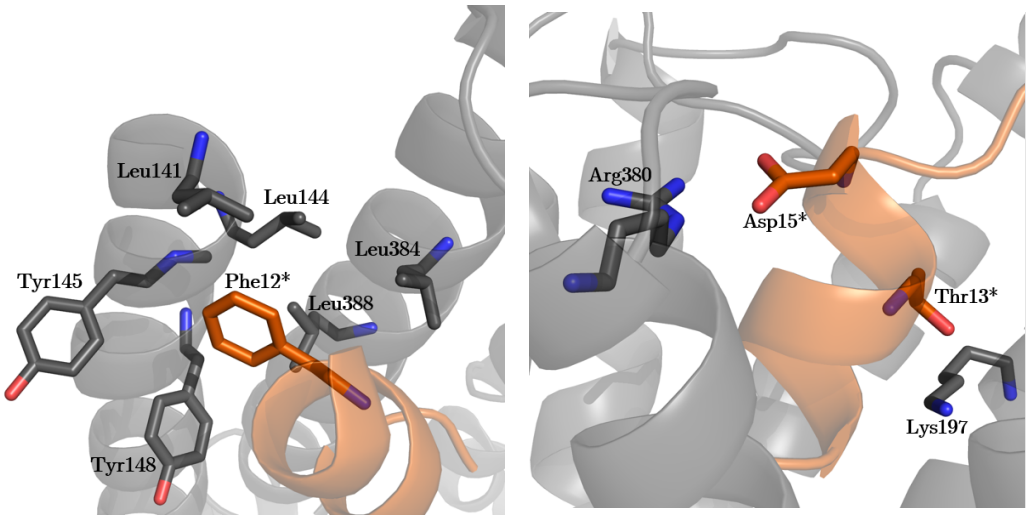


FIGURE 4.19: Interactions between Phe12* and a hydrophobic pocket in the TMD, shown to the left. The interaction between Thr13* and Lys197, as well as between Asp15* and Arg380, shown to the right. Image taken from the last frame of the GLP-1R–GLP-1(7000) simulation.

Phe12*, on the other hand, clearly interacts with a hydrophobic pocket comprised of Leu141, Tyr145, and Tyr148 that are all located in TM1, and Leu384 and Leu388 in TM7, which corresponds well with previosly results [62, 63]. Additionally, Phe12* also appears to interact with Leu144 in TM1 through the hydrophobic pocket, in our simulations. The interactions between Phe12* and the hydrophobic pocket in the TMD top are shown in fig. 4.19.

Thr13* has a strong interaction with Lys197 in TM2 as reported previously [62, 63]. Asp15* interacts weakly with Arg380 in ECL2, but the interaction seems to be important since the potency is reduced 1853-fold when Arg380 is mutated to Asp. Furthermore, this interaction has been reported previuosly [36, 62] and is illustrated in fig. 4.19.

Hence, our simulation can capture several important interactions between His7*, Glu9*, Phe12*, Thr13*, and Asp15* and the receptor transmembrane part. Additionally, it suggests that Gly10* is of importance, not in regards to interactions with the receptor, but to ensure overall flexibility of the peptide chain when binding to the activated receptor. These interactions are mainly located between the top of the receptor TMD and the N-terminal part of GLP-1.

Moving on to the ECD part, the interactions between the GLP-1 C-terminal and receptor ECD are mapped in fig. 4.20.

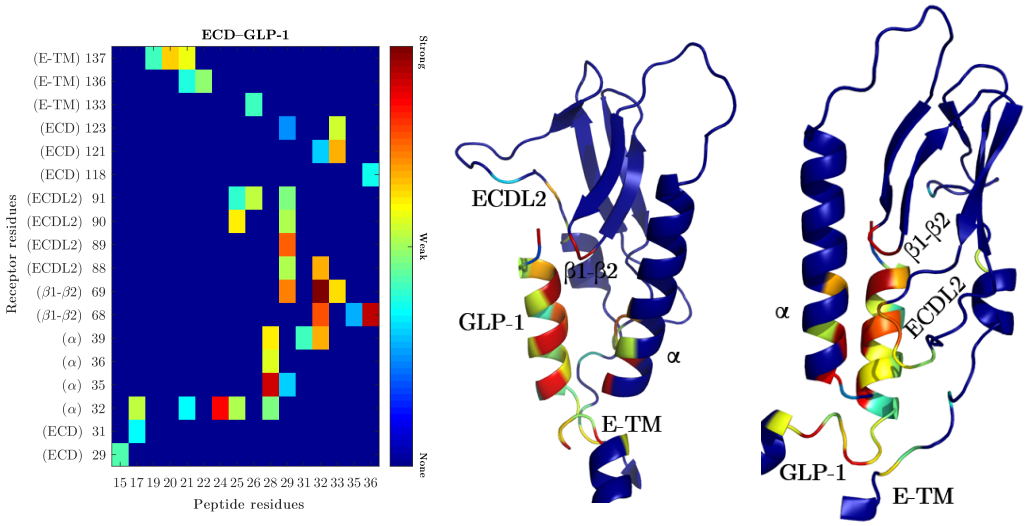


FIGURE 4.20: Interaction strength of the interacting residue pairs for the TMD and GLP-1. In the left panel is a plot with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. To illustrate the location of the binding interface between the receptor ECD and GLP-1, the structure (image taken from the last frame in the GLP-1R–GLP-1(7000) simulation) is shown from each side (middle and right) with the residues coloured according to the interactions shown in the heatmap.

The heatmap quickly shows that there are interactions present between GLP-1 and the E-TM linker, the ECD α-helix (α), the ECD loop no. 2 (ECDL2), and the β1-to-β2 turn (β1-β2), but none with the β-strands.

Starting with the middle part of GLP-1, residue Tyr19* interacts with receptor residue Pro137 in the E-TM linker, which also interacts with Leu20* and Glu21*. Glu21* also interacts with residues Leu32 (α) and Ser136 (E-TM). Gly22* only interacts with Ser136, whereas Lys26* interacts with both Trp91 and Gly133 in ECDL2 and E-TM, respectively. However, all of these interactions are identified as weak interactions, and are not mentioned previously. Nor does mutagenesis studies of those particular receptor residues indicate significant potency changes [25].

Contrary, stronger interactions are seen between Phe28* and the ECD α-helix, and between Ile29* and ECDL2. Particularly, GLP-1 residue Phe28* interacts with the hydrophobic patch comprised of ECD α-helix residues Leu32, Thr35, Val36, and Trp39, in combination with Tyr69 in the β1-to-β2 turn. Concerning GLP-1 residue Ile29*, hydrophobic interactions are present with ECDL2 residues Tyr88, Leu89, Pro90, and Trp91. This patch formation was reported several times [62, 65, 66]. The hydrophobic patch interacting with Phe28* and Ile29* is illustrated in fig. 4.21.

Leu32 in the ECD α-helix also interacts with both Ser17* and Ala24*. The former was identified before [62]. GLP-1 Leu32* show medium to strong interactions with Glu68 and Tyr69 in the β1-to-β2 turn, Trp39 in the α-helix, and Tyr88 in ECDL2. These interactions are shown in fig. 4.22.

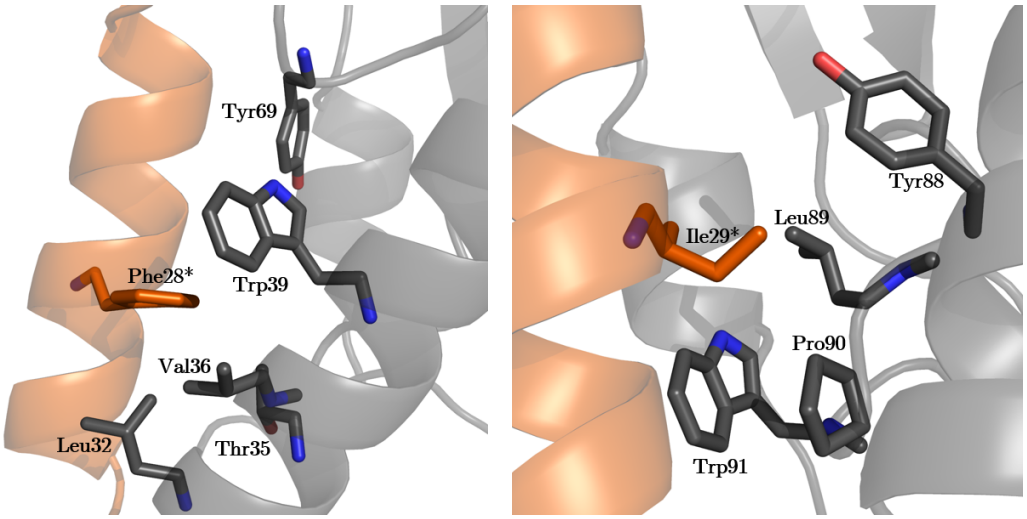


FIGURE 4.21: Interactions between Phe28* and a hydrophobic patch in the ECD, shown to the left. The interactions between Ile29* and a hydrophobic patch in the ECD, shown to the right. Image taken from the last frame in the GLP-1R–GLP-1(7000) simulation.

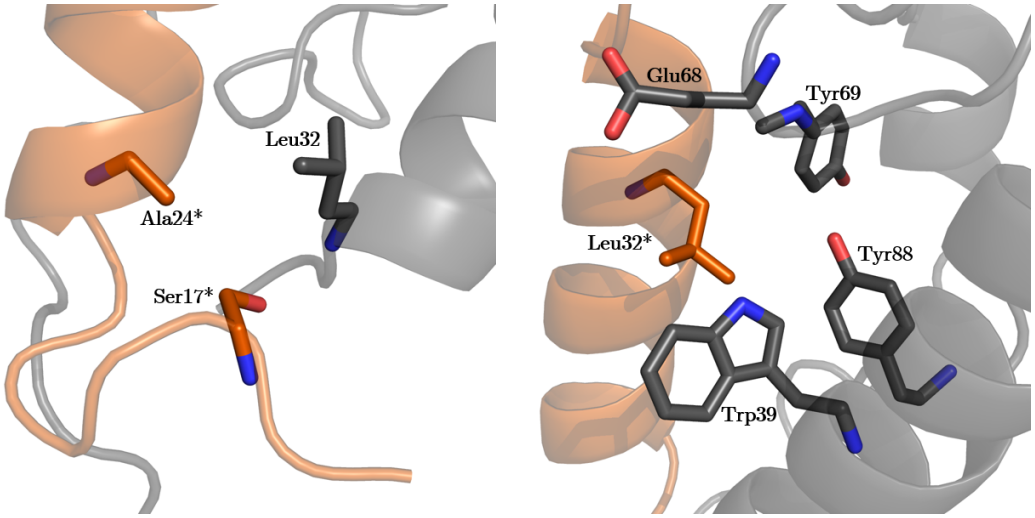


FIGURE 4.22: Interactions between Leu32 and Ser17* and Ala24* (left). Interactions between Leu32* and Glu69 and Tyr69 (right). Image taken from the last frame in the GLP-1R–GLP-1(7000) simulation.

Lastly, the bakcbone of Val33* interacts with Arg121 in the loop region prior to the E-TM linker, which has been reported previously as water mediated interaction between the backbone in Val33* and the side chain of Arg121 [66].

Considering this, our simulation system can be said to obtain and maintain the important hydrophobic patches noted in literature [62, 65, 66] between the ECD β1-to-β2 turn (Glu68 and Tyr69), α-helix (Leu32, Thr35, Val36, and Trp39), and ECD loop 2 (Tyr88, Leu89, Pro90, Trp91) and the hydrophobic residues Phe28*, Ile29*, and Leu32* in the C-terminal part of GLP-1. To sum up, the interactions identified in our GLP-1R–GLP-1 simulations are given in table 4.5. Very weak interactions are only given here, if they have been reported previously.

TABLE 4.5: Interactions identified in the GLP-1R–GLP-1 simulations. The strength of the interactions is given based on a percentage of frame counts out of the inspected total number of frames for each interaction pair. The total number of frames is based on the four replicas. Interaction pairs occurring less than 25 % of the last 100 ns are interpreted as negligible, meaning that 25 % marks the threshold for an interaction. Interactions occurring between 25 % and 75 % are considered weak to medium interactions, and frame counts of ≥75 % mark strong interactions.

GLP-1	GLP-1R	Interaction strength	Reported previously (see table 4.4)
His7*	Gln234 (TM3)	Weak	✗
His7*	Val237 (TM3)	Weak	✗
His7*	Thr241 (TM3)	Weak	✗
His7*	Asp372 (ECL3)	Weak	✗
His7*	Glu364 (TM6)	Very weak	✓
His7*	Lys383 (TM7)	Weak	✗
His7*	Glu387 (TM7)	strong	✓
Ala8*	Leu384 (TM7)	Weak	✓
Ala8*	Glu387 (TM7)	Weak	✗
Ala8*	Leu388 (TM7)	Medium	✓
Glu9*	Tyr148 (TM1)	Medium	✓
Glu9*	Tyr152 (TM1)	Strong	✓
Glu9*	Arg190 (TM2)	Strong	✓
Glu9*	Val194 (TM2)	Strong	✗
Glu9*	Met233 (TM3)	Medium	✗
Glu9*	Val237 (TM3)	Weak	✗
Gly10*	Phe230 (TM3)	Medium	✗
Thr11*	Arg380 (TM7)	Weak	✗
Phe12*	Leu141 (TM1)	Medium	✓
Phe12*	Leu144 (TM1)	Medium	✗
Phe12*	Tyr145 (TM1)	Strong	✓
Phe12*	Tyr148 (TM1)	Medium	✓
Phe12*	Leu384 (TM7)	Very weak	✓
Phe12*	Leu388 (TM7)	Very weak	✓
Thr13*	Lys197 (TM2)	Strong	✓
Thr13*	Thr298 (ECL2)	Weak	✓
Ser14*	Thr298 (ECL2)	Strong	✗
Ser14*	Arg299 (ECL2)	Medium	✗
Asp15*	Thr29 (ECD-NT)	Weak	✗
Asp15*	Arg299 (ECL2)	Medium	✗
Asp15*	Arg380 (TM7)	Very weak	✓
Val16*	Leu141 (TM1)	Strong	✓
Ser17*	Leu32 (α)	Weak	✓
Ser18*	Leu201 (TM2)	Medium	✗
Ser18*	Tyr205 (TM2)	Medium	✗
Tyr19*	Leu142 (TM1)	Weak	✗
Tyr19*	Tyr145 (TM1)	Strong	✗
Leu20*	Pro137 (E-TM)	Medium	✗
Glu21*	Pro137 (E-TM)	Weak	✗

Continued on next page

Table 4.5 – *Continued from previous page*

GLP-1	GLP-1R	Interaction strength	Reported previously
Gly22*	Ser136 (E-TM)	Weak	✗
Gln23*	Tyr205 (TM2)	Weak	✗
Ala24*	Leu32 (α)	Strong	✗
Ala24*	Leu217 (ECL1)	Weak	✗
Ala25*	Leu32 (α)	Weak	✗
Ala25*	Pro90 (ECDL2)	Medium	✗
Lys26*	Trp91 (ECDL2)	Weak	✗
Glu27*	Trp214 (ECL1)	Medium	✗
Phe28*	Leu32 (α)	Weak	✗
Phe28*	Thr35 (α)	Strong	✗
Phe28*	Val36 (α)	Weak	✗
Phe28*	Trp39 (α)	Medium	✗
Ile29*	Tyr69 (β1-β2)	Medium	✗
Ile29*	Tyr88 (ECDL2)	Weak	✗
Ile29*	Leu89 (ECDL2)	Strong	✗
Ile29*	Pro90 (ECDL2)	Weak	✗
Ile29*	Trp91 (ECDL2)	Weak	✗
Ala30*	Trp214 (ECL1)	Weak	✗
Trp31*	Trp214 (ECL1)	Weak	✗
Trp31*	Leu218 (ECL1)	Weak	✗
Leu32*	Trp39 (α)	Medium	✗
Leu32*	Glu68 (β1-β2)	Strong	✗
Leu32*	Tyr69 (β1-β2)	Strong	✗
Leu32*	Tyr88 (ECDL2)	Medium	✗
Val33*	Tyr69 (β1-β2)	Medium	✗
Val33*	Arg121 (ECD)	Medium	✓
Val33*	Leu123 (ECD)	Weak	✗
Lys34*	Trp214 (ECL1)	Weak	✗
Arg36*	Glu68 (β1-β2)	Strong	✗

In conclusion, our GLP-1R–GLP-1 simulations can capture several important interactions reported previously, in structural as well as mutagenesis studies. This adds to the justification of further analyses of our simulations.

4.2.2 Binding interface GLP-1R and GLP1 analogues

Firstly, the overall interaction pattern will be investigated. As before, this will be done by plotting the strength of the interaction pairs in heatmaps. The peptide and receptor residues are on the axes and represent interaction pairs in coloured squares. The colour illustrates the binding strength of the individual pairs. In figs. 4.23 to 4.30, the eight heatmaps for the receptor-peptide simulations are shown. They are made as a sum of the four replicas of each simulated system.

The heatmaps show the total interaction interface between the peptide and the total receptor. For the GLP-1R–GLP-1 system (fig. 4.23), there are two main interaction areas, as mentioned above.

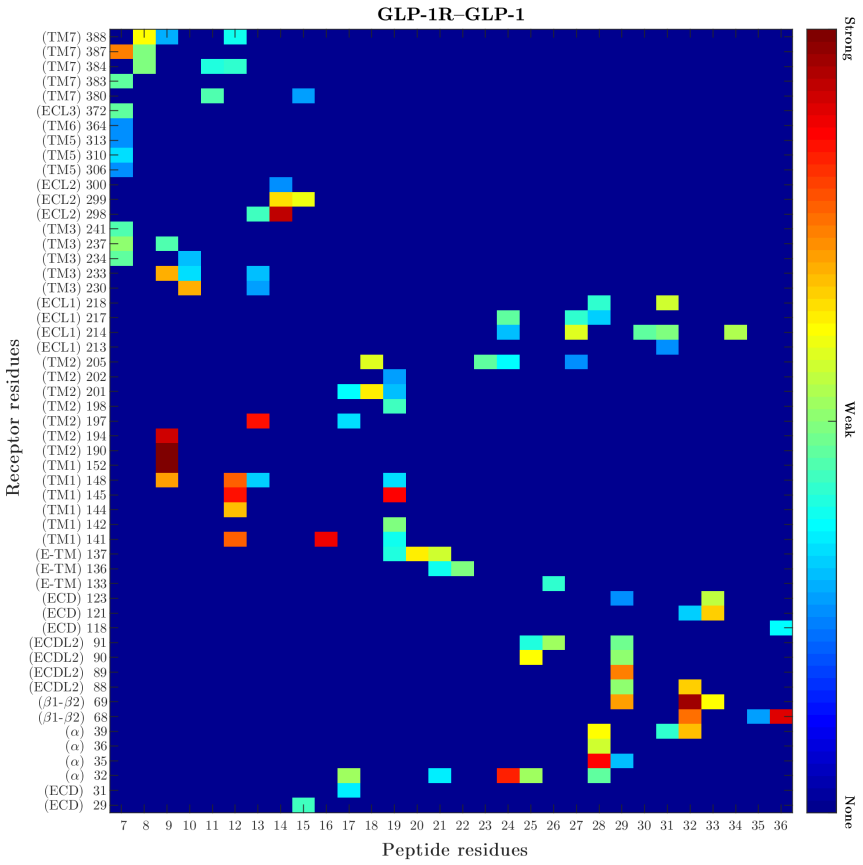


FIGURE 4.23: Interaction strength of the interacting residue pairs for GLP-1R–GLP-1 with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix Q for structure with corresponding residue colouring.

One of the areas is comprised of interactions between the N-terminal part of GLP-1 and TM1 and TM2, visible as a red cluster to the left and in the middle of the heatmap. The other involves the middle and C-terminal part of GLP-1 and the ECD, present as a red and orange cluster in the bottom right corner. Weaker interactions are also seen in two other main areas. One is between the GLP-1 N-terminal and TM3, TM5, TM6, TM7, ECL2, and ECL3, seen as a blueish to green cluster in the top left corner. The other is between the GLP-1 C-terminal and TM2 and ECL1, also visible as a blueish to green cluster in the middle, to the right.

In general, we see a similar interface pattern for the remaining seven systems, with the main difference being the strength of the interactions. This most likely stems from the fact that all the peptides have been docked manually into the activated receptor in the same position. The structural differences in the peptides are, however, enough to give rise to binding energies that correlate with experimentally measured potency values[7], so some differences are expected.

[7]It should be mentioned that the GLP-1R–C16L,$K^{26,34}$ system scores too high in regards to the binding energy, which could be due to the fact that the peptides are forced into the activated receptor.

Concerning the GLP-1R–lira system (fig. 4.24), there are two main changes to the interface binding strengths.

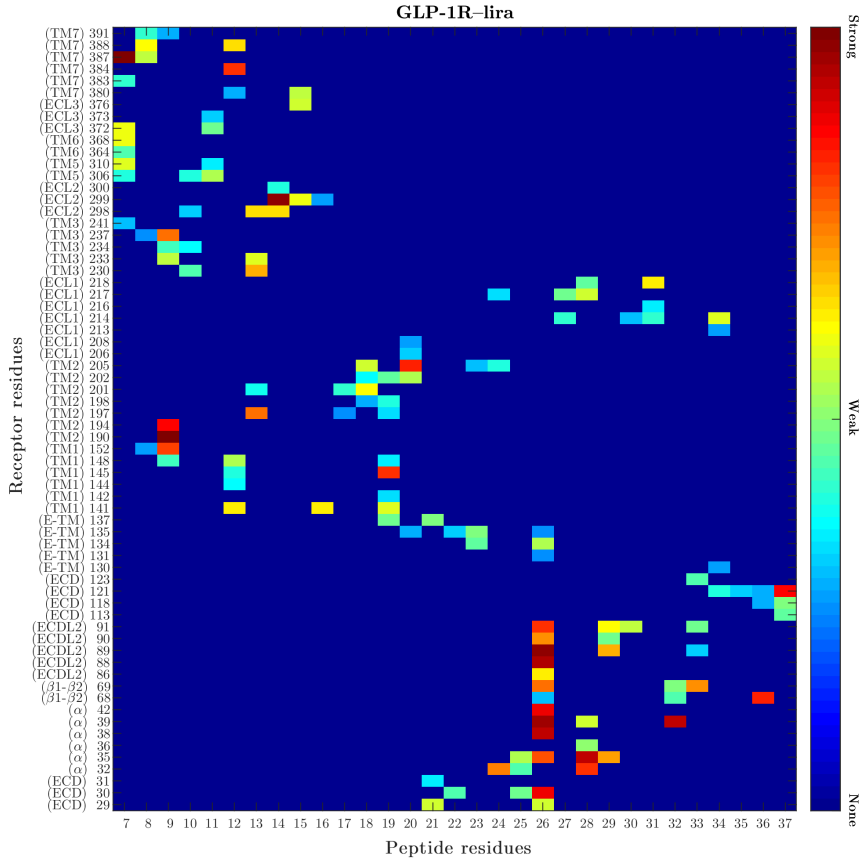


FIGURE 4.24: Interaction strength of the interacting residue pairs for GLP-1R–lira with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix R for structure with corresponding residue colouring.

The interactions between liraglutide and TM1 and TM2 are generally weaker, whereas the interactions between liraglutide and the ECD are stronger than those for GLP-1. Additionally, there are no interactions at all between the peptide N-terminal residues and the ECD. For GLP-1, residues Asp15* and Ser17* interact with the N-terminal residues of the ECD, which are absent in the liraglutide simulations.

Another prominent difference is the strong interactions that Lys26* can obtain. This is the acylation site in liraglutide. It is evident that the FA chain interacts solely with the ECD. More specifically, the α-helix, the β1-to-β2 turn, ECDL2, and some residues in the N-terminal of GLP-1R, as well as a few residues in unspecified coil/loop regions (noted simply as ECD). This leads to strong interactions between the 16 C-atom long FA chain and the ECD, and does not affect the potency, since the potency of liraglutide ((61.1 ± 7.1) pM) is in the same range as of GLP-1 ((55 ± 19) pM).

The interaction pattern of the GLP-1R–C16L,K^{26,34} system (fig. 4.25), overall, resembles that of the GLP-1R–GLP-1 system.

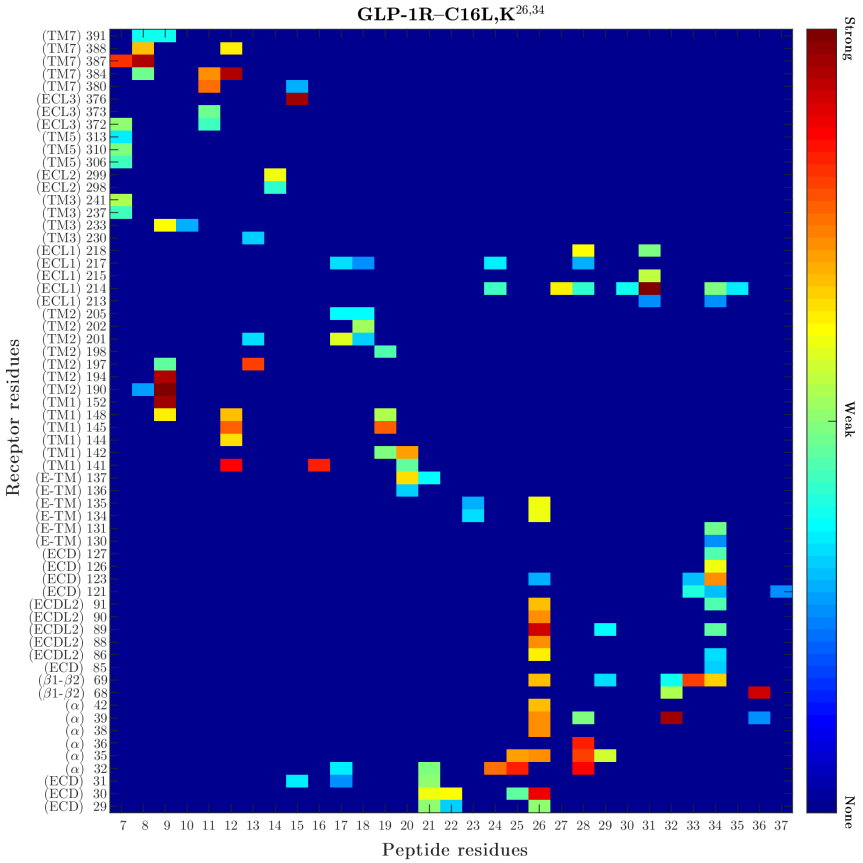


FIGURE 4.25: Interaction strength of the interacting residue pairs for GLP-1R–C16L,K^{26,34} with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix S for structure with corresponding residue colouring.

With focus on the interaction strength, the weaker interactions between the peptide C-terminal and receptor ECD, together with the stronger interactions between the peptide N-terminal and TM1 and TM2 are similar to those seen for GLP-1. On the other hand, there are several interactions between the two acylated residues, Lys26* and Lys34* and the ECD. However, compared to liraglutide, these interactions are significantly weaker. Furthermore, it is only the FA chain on Lys26* that is interacting with the α-helix, and in general, the FA chain on Lys34* only has weak to medium interactions with the receptor. C16L,K^{26,34} is the peptide that causes the highest potency reduction to $(16\,700 \pm 3700)$ pM and is the only peptide with two acylated FA chains.

Looking at the GLP-1R–C14A system (fig. 4.26), there are fewer medium and strong interactions, overall.

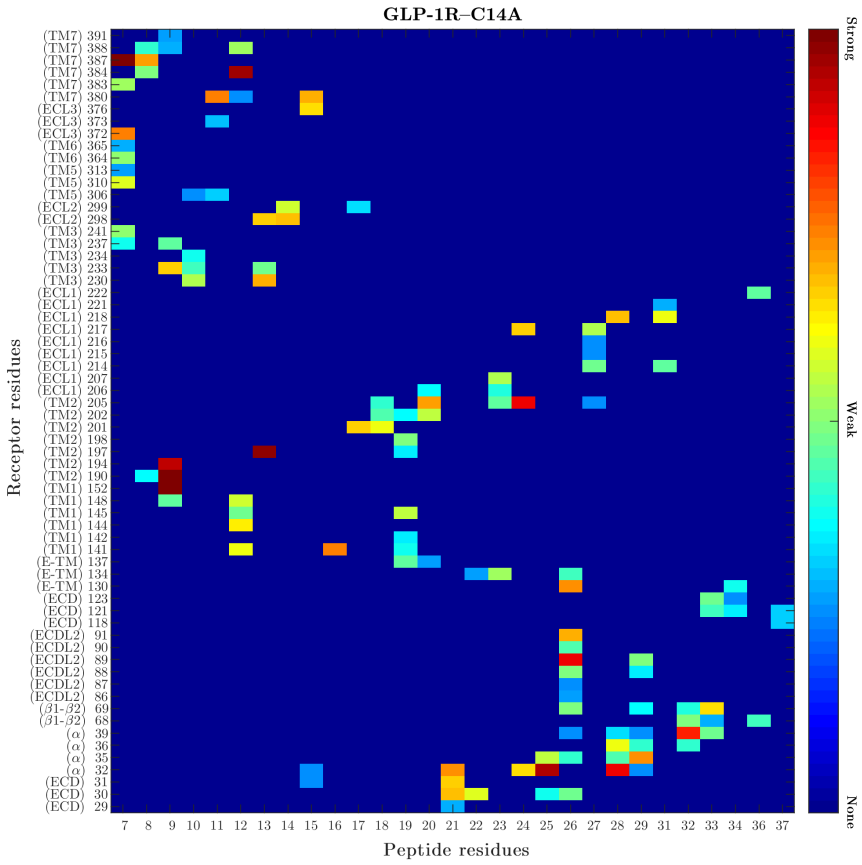


FIGURE 4.26: Interaction strength of the interacting residue pairs for GLP-1R–C14A with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix T for structure with corresponding residue colouring.

However, the appearing few strong interactions are located in both the N- and C-terminals of the peptide, interacting with TM1/TM2/TM7 for the former, and the ECD for the latter. As previosly seen, there are quite a few interactions with Lys26* (also the acylation site for the C14A peptide), but they are generally very weak to weak. The acylated FA chain for this system is shorter with only 14 C-atoms and no linker, and it has a free acid group in the end. This indicates that the free acid group can disturb some of the interactions of the apolar FA chain to the apolar patch in the ECD. This said, the potency for the C14A system is only slightly lower than that of liraglutide ((61.1 ± 7.1) pM and (72.0 ± 0.7) pM for liraglutide and C14A, respectively).

Coming to the GLP-1R–daC16L,K^{34} system (fig. 4.27), the medium and strong interactions are spread out more than seen for GLP-1.

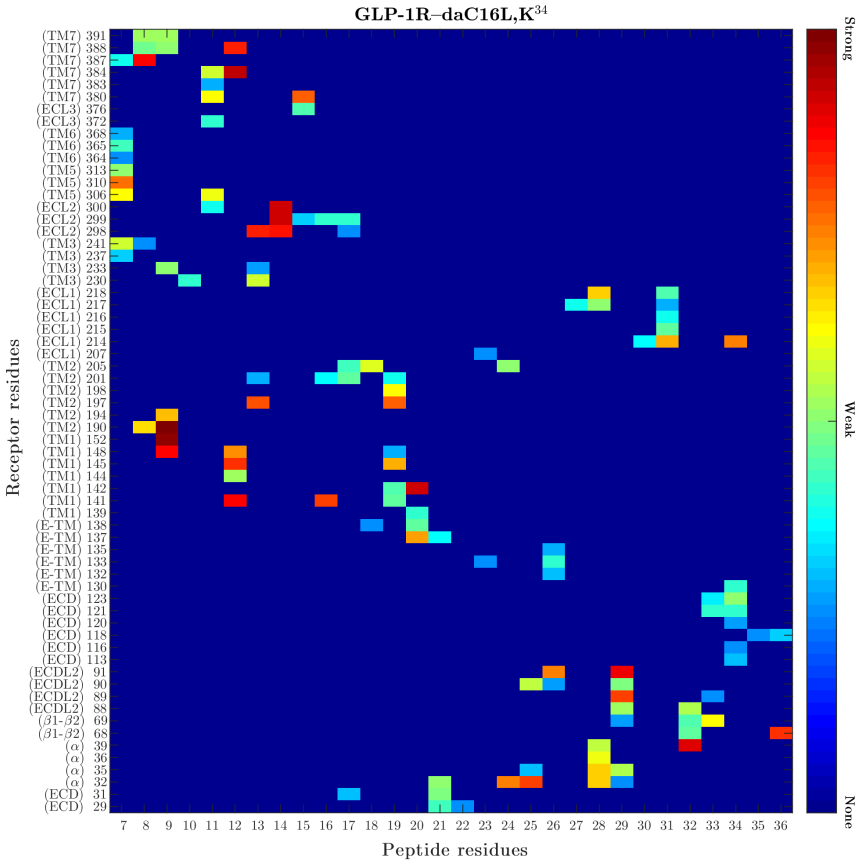


FIGURE 4.27: Interaction strength of the interacting residue pairs for GLP-1R–daC16L,K^{34} with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix U for structure with corresponding residue colouring.

Here, there are strong interactions in three areas. That is, between the peptide N-terminal and TM1/TM2, and between the peptide C-terminal and the ECD, as well as between the peptide N-terminal and ECL2/TM5/TM7 (top left corner). Even though, more strong interactions are present in this region, the medium to strong interaction between the His7* and Glu387 present in the previously described systems, is very weak here. In daC16L,K^{34}, the N-terminus residue His7* has had the free backbone amine removed, which weakens the His7* interaction to Glu387. This correlates well with the fact that Glu387 was interacting with the amine rather than the His7* side chain. This could possibly add to the reason why the potency of this peptide analogue is lower. Another thing to notice about this interaction interface is that the acylated Lys34* has fewer interactions than seen before for the FA chain and only one interaction is medium, the rest are very weak to weak. The FA chain is attached with a linker and is comprised of 16 C-atoms, as seen in e.g. liraglutide and C16L,K^{26,34}. However, its interactions with the ECD are less and weaker, indicating that the acylation on Lys34* disturbs the interaction pattern to the receptor. This correlates well with the fact the potency value is relatively high at (2360 ± 370) pM.

Continuing on to investigate the GLP-1R–daC8,K^{34} system (fig. 4.28), we see the same absence of strong interaction between His7* to Glu387, but as for daC16L,K^{34}, there are a few more strong interactions between the middle part of the peptide and the ECD α-helix, than previously seen.

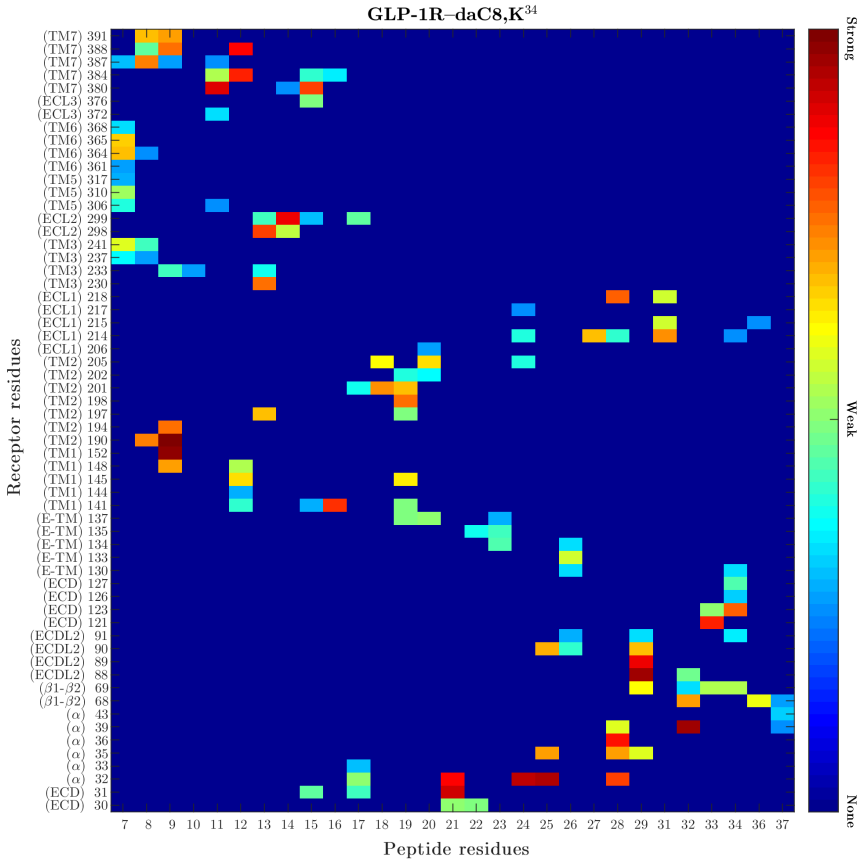


FIGURE 4.28: Interaction strength of the interacting residue pairs for GLP-1R–daC8,K^{34} with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix V for structure with corresponding residue colouring.

Here, a handful of strong interactions are present between the middle part of the peptide and the beginning of the ECD α-helix. These interactions are present in the other systems, however they appear as very weak to weak interactions. Similar to the daC16L,K^{34} peptide, the daC8,K^{34} peptide has no free amine on His7*, which again seems to show up as only a very weak interaction to Glu387, and most likely resulting in the lower potency for this peptide. The acylation site is also shared between daC16L,K^{34} and daC8,K^{34}, although, there is no linker here and the FA chain is only eight carbon atoms long. Like before, there are only a few interactions between Lys34* and the ECD, and they are mainly weak, also indicating that acylation on Lys34* prevents FA interactions with the ECD.

Next up is the GLP-1R–C12A,K^{38} system (fig. 4.29), which generally has fewer strong interactions.

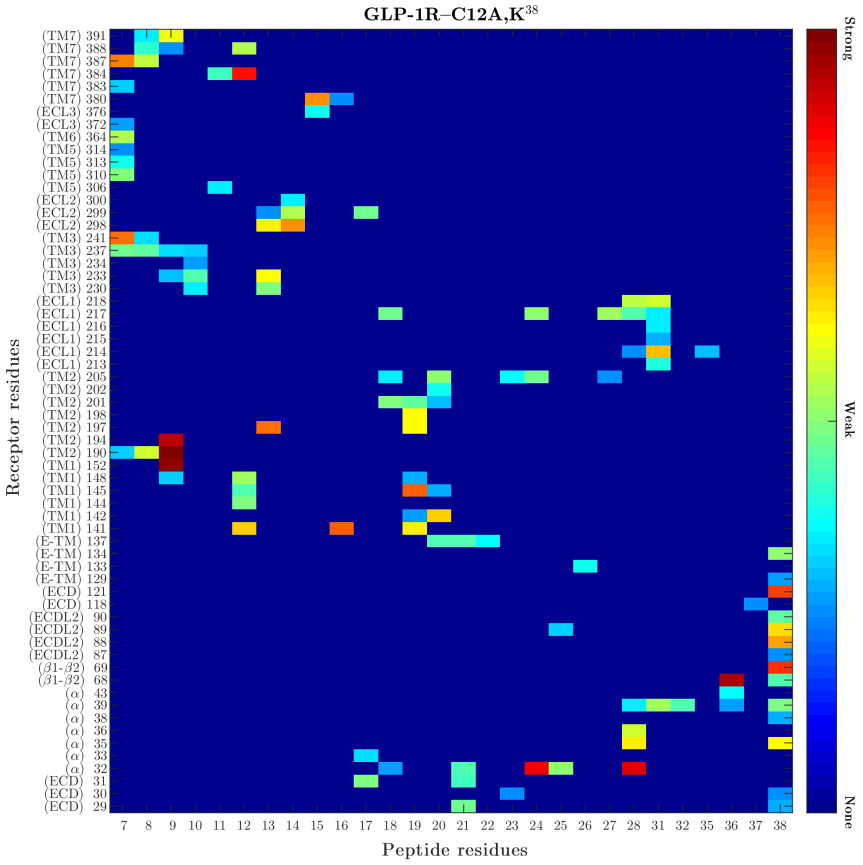


FIGURE 4.29: Interaction strength of the interacting residue pairs for GLP-1R–C12A,K^{38} with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix W for structure with corresponding residue colouring.

Each interaction area only has a couple of strong interactions, which is rather interesting, since this peptide has the highest potency of (4.19 ± 0.98) pM. For C12A,K^{38}, the acylation site is on Lys38*, and there is a free acid group at the end of the 12 C-atom FA chain. Several interactions are present between the acylated Lys38* and the ECD.

Lastly, the GLP-1R–C14L,K^{38} system (fig. 4.30), which has lost interactions between the middle part of the peptide and the N-terminal part of the ECD, is considered.

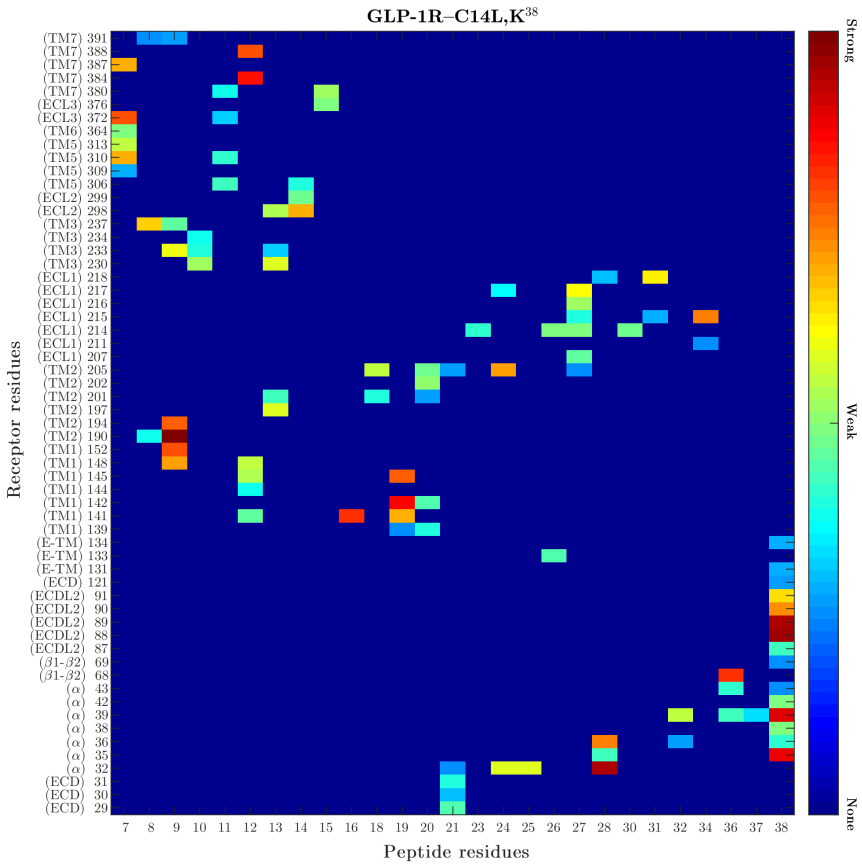


FIGURE 4.30: Interaction strength of the interacting residue pairs for GLP-1R–C14L,K^{38} with peptide residues on the x-axis and receptor residues on the y-axis. Hence, each square gives an interaction pair. The colour of the squares indicates the strength of the interactions according to the colourscale. If an interaction is seen in less than 25 % of the investigated frames, it is considered as negligible. If it occurs between 25 % and 75 % it is considered from very weak over weak to medium. Strong interactions occur 75 % or more of the investigated time. See Appendix X for structure with corresponding residue colouring.

Similar to the liraglutide system, there are no interactions at all between peptide residues Asp15* and Ser17* and the ECD. Compared to the C12acid system, which is also acylated at Lys38*, there are a few more strong interactions with the ECD. This could be a result of the purely hydrophobic FA chain end (no free acid group).

For comparison, the interactions mapped by Jazayeri et al. 2017, Santiago et al. 2018, Underwood et al. 2010, and Zhang et al. 2017 proven to be important for GLP-1 potency (>5-fold reduction) are given in Table 4.6. The interactions as they appear in the different simulation systems are marked according to their strength. Strong interactions are marked with ▲, and weak to medium interactions are marked with △. If no interaction is present, the pair is marked with ✗.

TABLE 4.6: Overview of interactions found in the different systems. The interactions shown are those identified by Jazayeri et al. 2017, Santiago et al. 2018, Underwood et al. 2010, and Zhang et al. 2017 and with receptor residue mutation causing more than a 5-fold reduction in GLP-1 potency as reviewed by de Graaf et al. 2016. Interaction pairs are given as peptide–receptor residues. If the interaction occurs in 75 %, or more, of the investigated frames (strong interaction), it is marked with ▲. If the interaction occurs in less than 25 % of the investigated frames (no interaction), it is marked with ✗. Does the interaction occur in [25 %; 75 %[of the investigated frames, it is marked with △.

Interaction	GLP-1	lira	2C16L,$K^{26,34}$	C14A	daC16L,K^{34}	daC8,K^{34}	C12A,K^{38}	C14L,K^{38}
H7*–R299	✗	✗	✗	✗	✗	✗	✗	✗
H7*–W306	△	△	△	✗	△	△	✗	✗
H7*–E364	△	△	✗	△	△	△	△	△
H7*–E387	△	▲	▲	▲	△	△	△	△
A8*–L384	△	✗	△	△	✗	△	✗	✗
A8*–L388	△	△	△	△	△	△	△	△
E9*–Y148	△	△	△	△	▲	△	△	△
E9*–R190	▲	▲	▲	▲	▲	▲	▲	▲
E9*–L388	△	✗	✗	△	△	▲	△	✗
G10*–W306	✗	△	✗	△	✗	✗	✗	✗
T11*-R299	✗	✗	✗	✗	✗	✗	✗	✗
T11*–D372	✗	△	△	✗	△	△	✗	△
T11*–L384	△	✗	△	✗	△	△	△	✗
F12*–Y148	▲	△	△	△	△	△	△	△
F12*–L384	△	▲	▲	▲	▲	▲	▲	▲
F12*–L388	△	△	△	△	▲	▲	△	▲
T13*–K197	▲	▲	▲	▲	▲	△	▲	△
T13*–M233	△	△	✗	△	△	△	△	△
T13*–W297	✗	✗	✗	✗	✗	✗	✗	✗
S14*–W297	✗	✗	✗	✗	✗	✗	✗	✗
S14*–R299	△	▲	△	△	▲	▲	△	△
S14*–N300	△	△	✗	✗	▲	✗	△	✗
D15*–R380	△	△	△	△	▲	▲	△	△
V16*–Y205	✗	✗	✗	✗	✗	✗	✗	✗
S17*–M204	✗	✗	✗	✗	✗	✗	✗	✗
S17*–Y205	✗	✗	△	✗	△	✗	✗	✗
S17*–W297	✗	✗	✗	✗	✗	✗	✗	✗
S17*–R299	✗	✗	✗	△	△	△	△	✗
S18*–W297	✗	✗	✗	✗	✗	✗	✗	✗
S18*–R299	✗	✗	✗	✗	✗	✗	✗	✗
L20*-M204	✗	✗	✗	✗	✗	✗	✗	✗
E21*-R299	✗	✗	✗	✗	✗	✗	✗	✗

Based on GLP-1 binding, table 4.6 shows that the previously reported interactions that the GLP-1R–GLP-1 simulations can not capture (✗) are not present (✗), or only weak to medium (△), in all other systems. The strong interactions in the GLP-1R–GLP-1 system are also strong in the other systems, except for Phe12*–Tyr148, which is weak to medium in all other systems, and Thr13*–K197, which is weak to medium in the GLP-1R–daC8,K^{34} system. In addition, there are only two interactions that differ from not present to strong, i.e. the Glu9*–Leu388 and Ser14*–Asn300 interactions. This indicates that the placement of the peptides, during the initial manual docking, seems to cause the same interactions to be captured for all systems, as are captured in the GLP-1 simulations. As a result, there is only a few differences in the binding scheme of the peptide–receptor residue interactions among the simulated systems. However, figs. 4.23 to 4.30, testify, that there are large differences

between the different systems when it comes to the interactions of the FA chain on the acylated residues with the receptor. Therefore, in the following, a more thorough study of the individual interactions of the FA chains will be carried out.

4.2.3 Fatty acid chain interactions

So far, the main differences in binding appears to stem from the FA chain interactions with the receptor rather than receptor-peptide residue interactions along the peptide chain. Starting with liraglutide, the binding interface surrounding the FA chain is shown in fig. 4.31 together with FA interacting ECD residues.

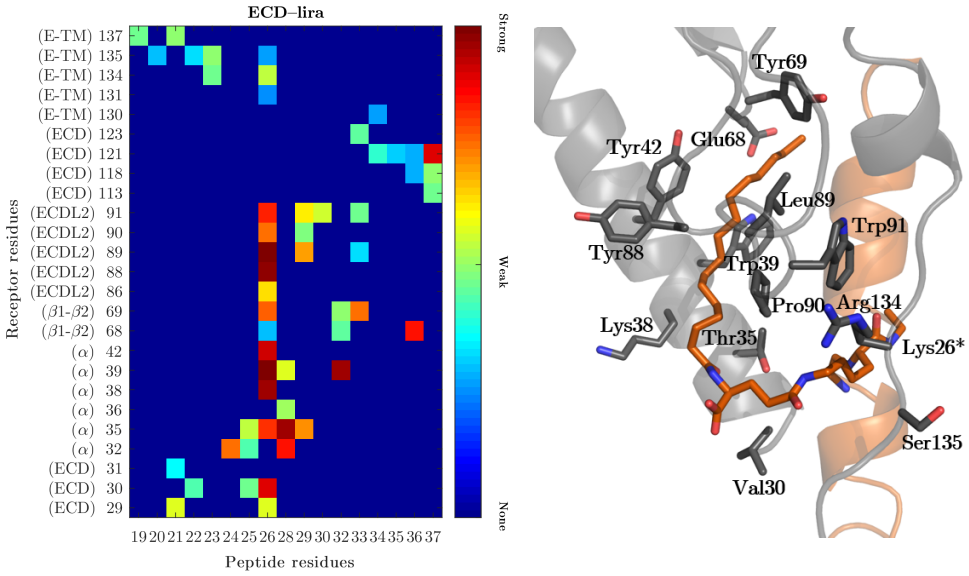


FIGURE 4.31: Heatmap showing the interactions between the receptor ECD and liraglutide, left. The ECD residues interacting with the acylated Lys26* residue in liraglutide, right. The image is taken from the last frame of the GLP-1R–lira(7000) simulation.

It is evident, that the FA chain in liraglutide forms several hydrophobic interactions with the hydrophobic patch made up of the ECD α-helix, the β1-to-β2 turn, and ECDL2. Meaning that the FA chain on liraglutide can neatly tug into the hydrophobic pocket that also interacts with Phe28*, Ile29*, and Leu32* in GLP-1. Nearly all of these interactions are strong, and only a few weak interactions are present. There are no electrostatic interactions between the polar FA chain of the acylated residue and the ECD.

Moving on to the C16L,$K^{26,34}$ peptide, there are more interactions present, but there are also two FA chains present. These interactions are shown in fig. 4.32.

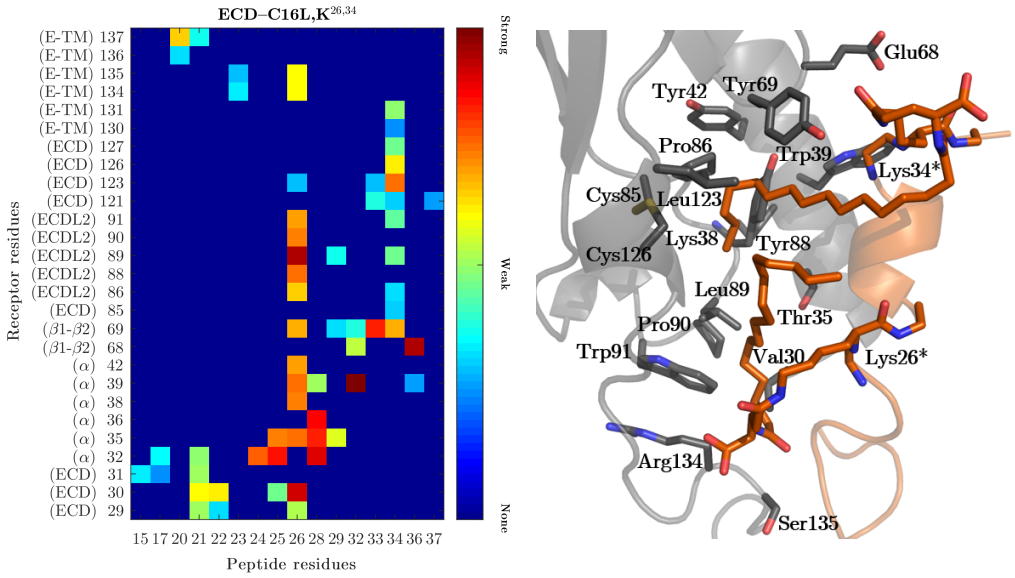


FIGURE 4.32: Heatmap showing the interactions between the receptor ECD and the C16L,K^{26,34} peptide, left. The ECD residues interacting with the acylated Lys26* and Lys34* residues in the C16L,K^{26,34} peptide, right. The image is taken from the last frame of the GLP-1R–C16L,K^{26,34}(7000) simulation.

It is only the FA chain acylated on Lys26* that can obtain strong interactions with the ECD. There are two medium interactions between the FA chain attached on Lys34*, one to Tyr69, which could be polar, and one to Leu123 which is hydrophobic. Compared to liraglutide, the interactions with the FA chain on Lys26* are weaker, and the FA chain bends back on itself and appears to interact with the FA chain on Lys34* rather than snuggling into the hydrophobic patch in the ECD. This indicates that acylation on Lys34* cannot result in as many hydrophobic interactions with the ECD as acylation on Lys26*. Furthermore, acylation on both Lys26* and Lys34* interrupts interactions with the ECD. One could speculate, if the two long FA chains interacts with each other and thus prevents initial bonding to the receptor ECD. However, since the peptide was manually docked into the activated receptor, the first step of the two-domain model was ignored. This could be the reason why this peptide has the lowest potency of around 16 000 pM. This suggests, that the simulations performed in this study represent the scenario where the peptide actually do bind (even though the potency is much lower than for any of the other peptides, it is still indicating effective binding), rather than giving a representation of the full scenario, where the peptide will not bind most of the time. This could very well be due to the bulkyness of the two FA chains interacting.

Next is the C14A peptide, whose FA interactions are shown in fig. 4.33.

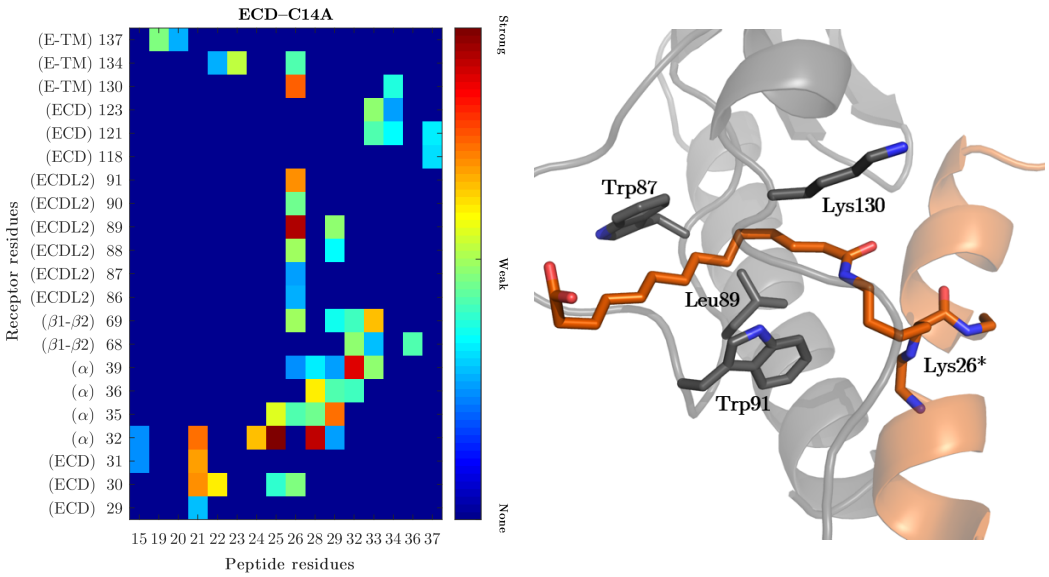


FIGURE 4.33: Heatmap showing the interactions between the receptor ECD and the C14A peptide, left. The ECD residues interacting with the acylated Lys26* residue in the C14A peptide, right. The image is taken from the last frame of the GLP-1R–C14A(7000) simulation.

Even though, the acylation site on C14A is Lys26*, like on liraglutide, there are only a few strong interactions between the FA chain and the ECD. They are all hydrophobic. It is evident that the FA chain has moved out of the hydrophobic pocket in the ECD and is 'floating' on the outside of the ECD. Suggesting that the polar acid group in the ω end of the FA chain disrupts ECD interactions. The free acid does not appear to make any interactions with any charged residues in the ECD. It rather seems to be pushed away from the hydrophobic residues in the ECD α-helix. So acylation on Lys26* does not necessarily cause hydrophobic interaction with the ECD, if the FA chain contains an acid group.

Exploring acylation on Lys34* is possible when looking at the interactions obtained between the FA chain in daC16L,K^{34} and the receptor ECD, which are shown in fig. 4.34.

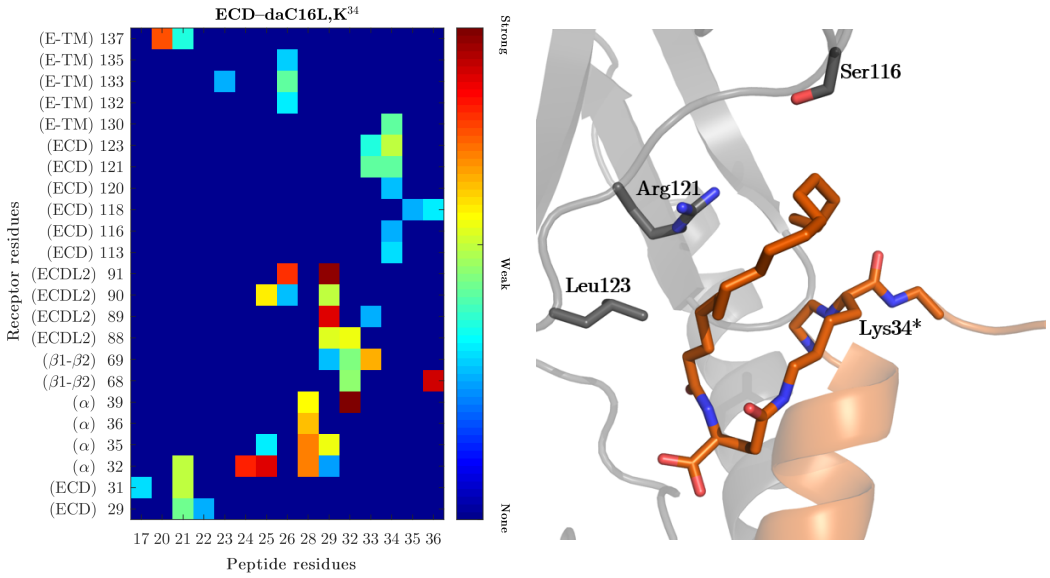


FIGURE 4.34: Heatmap showing the interactions between the receptor ECD and the daC16L,K^{34} peptide, left. The ECD residues interacting with the acylated Lys34* residue in the daC16L,K^{34} peptide, right. The image is taken from the last frame of the GLP-1R–daC16L,K^{34}(7000) simulation.

Here, it is quite clear that the FA chain cannot obtain strong interactions with the ECD. Only a handful of weaker interactions are present, and it seems as if the charged and polar residues in the region between ECDL2 and the E-TM linker are forcing the FA chain to curl up, which prevents it from sitting in the hydrophobic pocket of the ECD.

The FA interactions of the daC8,K^{34} peptide are illustrated in fig. 4.35.

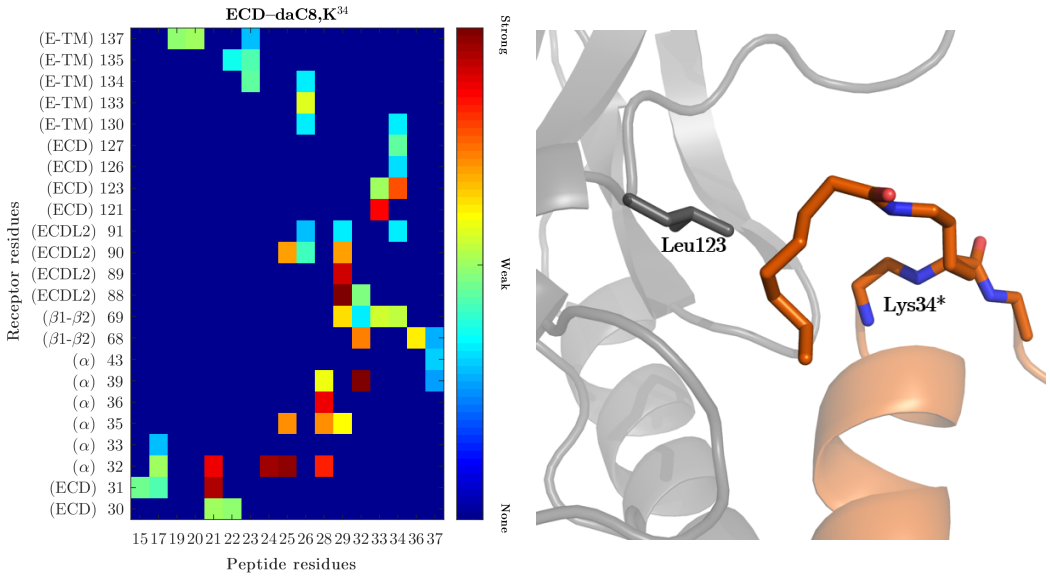


FIGURE 4.35: Heatmap showing the interactions between the receptor ECD and the daC8,K^{34} peptide (left), and residue interactions with Lys34* (right). The image is taken from the last frame of the GLP-1R–daC8,K^{34}(7000) simulation.

The daC8,K^{34} peptide also includes acylation at Lys34*, but with a short FA chain. This appears to cause even fewer interactions. Only one medium interaction to the hydrophobic Leu123 is present. Comparing to the other peptides with acylation on Lys34*, this is a reoccurring property, indicating that acylation here precludes strong hydrophobic interactions with the apolar patch on the ECD. This is possibly a result of the fact that the Lys side chain in position 34* points away from the ECD α-helix rather than towards it, as is the case for Lys26*. Furthermore, the short chain obtains fewer interactions than the longer ones, indicating that a certain length is required for the FA chain to obtain interactions with the ECD. This fits well with the lower potency for the daC8,K^{34} of (236 ± 66) pM.

Investigating acylation on Lys38*, we move on to the C12A,K^{38} peptide, whose interactions are represented in fig. 4.36.

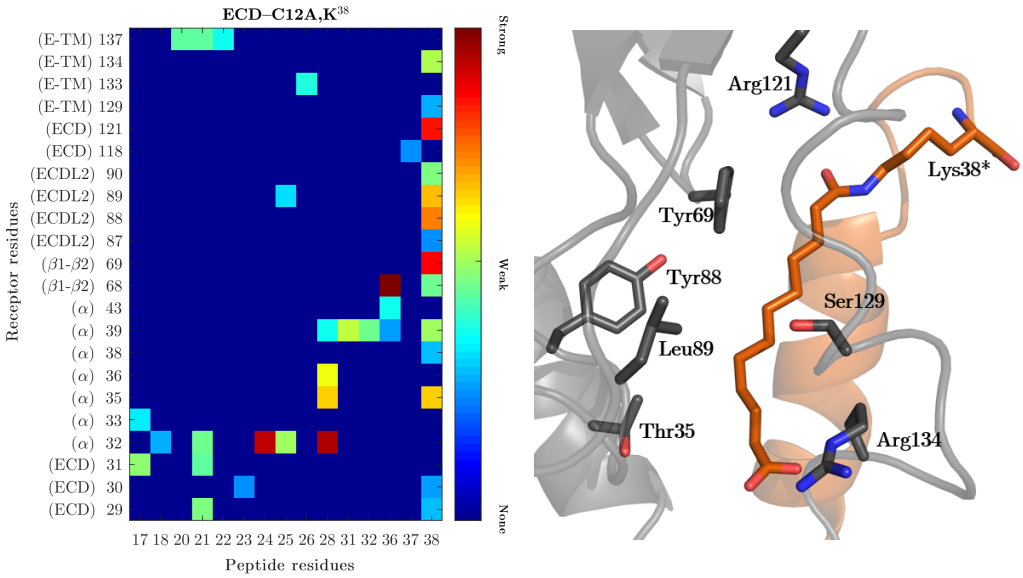


FIGURE 4.36: Heatmap showing the interactions between the receptor ECD and the C12A,K^{38} peptide, left. The ECD residues interacting with the acylated Lys38* residue in the C12A,K^{38} peptide, right. The image is taken from the last frame of the GLP-1R–C12A,K^{38}(7000) simulation.

Surprisingly, there are several interactions between the FA chain and the ECD in this system. Like with liraglutide, the FA chain is interacting with the hydrophobic patch on the ECD. Indicating that acylation on Lys38* ensures the flexibility of the peptide chain end, to give the backbone the necessary flexibility to bend and position the FA chain optimal towards the hydrophobic part of the ECD. Furthermore, the acid group on the ω end of the FA chain appears to be able to form electrostatic interactions with the polar Arg134 in the linker to the TMD. Seemingly, this interaction could retain the position of the FA chain along the peptide in the hydrophobic patch of the ECD.

The last system involves the C14L,K^{38} peptide, which is also acylated on Lys38*. The interactions observed within this system are shown in fig. 4.37.

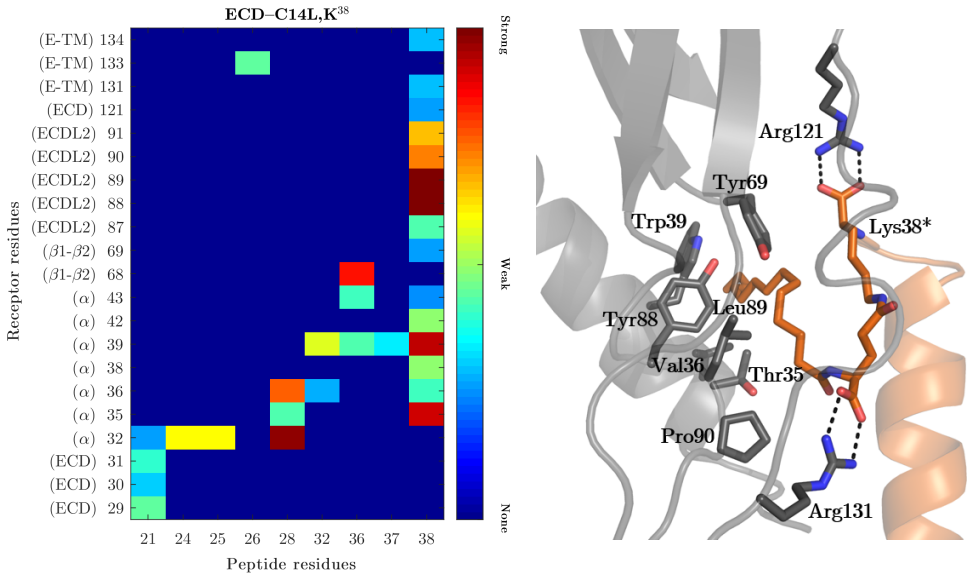


FIGURE 4.37: Heatmap showing the interactions between the receptor ECD and the C14L,K^{38} peptide, left. The ECD residues interacting with the acylated Lys26* residue in the C14L,K^{38} peptide, right. Hydrogen bonds are indicated by dashed lines. The image is taken from the last frame of the GLP-1R–C14L,K^{38}(7000) simulation.

Again, the FA chain on Lys38* can obtain several strong interactions with the hydrophobic patch of the ECD. However, there is no acid group to stabilise the FA chain end, but the linker glutamic acid forms hydrogen bonds to Arg131 and the C-terminal of the peptide chain, likewise, forms hydrogen bonds with Arg121. These interactions could retain the Lys38* side chain and linker along the peptide, facilitating interactions between the actual FA chain and the apolar part of the ECD.

A closer look at the individual interactions between the acylated FA chains and receptor ECD in the simulated systems suggests that the position of acylation is of great importance. If the FA chain is attached to Lys34*, it points away from the ECD α-helix, and the FA chain is thus prohibited from forming hydrophobic interactions with it. Is the acylation on Lys26*, the FA chain, on the other hand, points towards the ECD α-helix, and can snuggle up against the hydrophobic patch on the ECD. If acylation occurs on Lys38*, the added flexibility of an additional residue makes it possible for the peptide backbone to bend, and allow the FA chain to interact with the apolar patch. In addition to this, the results point to that acylation on both Lys26* and Lys34* interrupts interactions with the ECD and might even hinder initial binding to the ECD during the receptor activation. On an other note, the length of the FA chain also appears to be of importance, pointing to the necessity of a certain length to obtain hydrophobic interactions with the ECD. Furthermore, the presence of an ω acid group seems to influence ECD bonding, dependent on where the acylation happens. If a free acid group is present on Lys26*, interactions are prohibited, whereas if it is on Lys38*, polar interactions can stabilise the position of the FA chain along the ECD α-helix. Additionally, suggestions are made that the presence of a linker can facilitate hydrogen bonding with Arg121 and Arg131, possibly contributing to a stable placement of the FA.

4.3 Conclusion

In conclusion, the resulting simulations of the different peptide-receptor systems have been deemed stable by RMSD and number of interactions. For each of the systems, MM-PBSA energies were calculated and in turn compared to experimentally determined potency values, which they proved to correlate positively with. Furthermore, the simulated GLP-1R–GLP-1 system can obtain and maintain previously found binding interactions, indicating that analyses on the binding interactions between the respective peptides and the GLP-1 receptor are done with proper validation.

However, the simulations appear to suffer from the manual docking of the peptides into the activated receptor, bypassing the first step in the two-domain binding model. Yet, this limitation does not hamper the correlation with potencies. This is mainly visible as similar interaction patterns for all the simulated peptides, differing mainly in how the FA chains interact with the ECD, and how the peptide N-terminal His7* interacts with the TMD. Our results show:

- Des-amino His7* prevents backbone interactions to Glu387
- The major differences between the systems stems from FA interactions
- Acylation on Lys26* results in interactions between the FA chain and ECD hydrophobic patch
- Acylation on Lys34* leads to fewer FA interactions with the ECD hydrophobic patch
- Acylation on both Lys26* and Lys34* may cause FA self-interactions that could prevent initial receptor binding
- Acylation on Lys38* creates flexibility to the peptide backbone, making FA interactions with the ECD possible
- Shorter FA chains cannot interact as much with the receptor as longer FA chains

In addition to this, it appears that the presence of a linker can potentially stabilise the position of the FA chain by forming hydrogen bonds to Arg residues in the E-TM linker in the receptor. Furthermore, having a free acid group can be beneficial if the acylation occurs on Lys38*, where it can stabilise the position of the FA chain on the hydrophobic patch in the ECD. It can also be a disadvantageous, if the FA chain is attached on Lys26*, where it can disrupt interactions with the ECD.

Chapter 5

Manuscript: Entropy calculations

This chapter pose as a manuscript for a paper. It has yet to be submitted, intentionally it will be submitted to the Journal of Computational Chemistry. For the sake of consistency, the references in this chapter will be numbered in chronological order amongst the references mentioned in the other chapters. They will appear in the reference section at the end of the thesis.

5.1 Abstract

The molecular mechanics-Poisson-Boltzmann surface area method, better known as MM-PBSA, is a well evaluated method for computing binding free energy estimates of protein-peptide complexes from molecular dynamics (MD) simulations. It provides an easy screening for ranking possible drug candidates by calculation of internal and solvation energies as well as the entropic contribution. Whereas, the internal and solvation energies can be calculated from single frames using good approximations, the entropic contribution has to be calculated from several frames via less accurate approximations. However, little information is given on these type of calculations and scanty attention seems to be given to this part of the MM-PBSA calculations. We have found this to be a wide-ranging problem, since we see a large system dependent difference in how many frames needs to be included to gain converged entropy values. Yet, oftentimes, nothing is mentioned on whether or not the entropy values have stabilised using a given setup. And while it is common practise to mention the method and/or setup used for calculating the entropy values, it is frequently inadequate and normally only considers a few hundred frames, or less. Even though this will be enough for calculating internal and solvation energies, we have proven that this is *not* the case for entropy calculations. We find that several thousands of frames are needed to obtain independent entropy values, and furthermore, this is highly dependent on the size of the investigated system. Hence, our study presents a step-by-step methodology for calculating entropy values directly from the MD trajectories, as well as ensuring a setup that results in independent entropy values.

5.2 Introduction

5.2.1 MM-PBSA binding energy

Calculating stable and independent entropy values for protein-peptide interactions is essential when computing binding affinities from molecular dynamics (MD) simulations. MM-PBSA, molecular mechanics-Poisson-Boltzmann surface area, is a well-known and very capable method used for calculating binding free energies of such

simulation systems [110, 111]. Here, the binding energy, ΔG_{bind}, is calculated as the difference between the total energy of the complex (com), the total energy for the free peptide (pep), and the total energy of the free receptor (rec):

$$\Delta G_{\text{bind}} = G_{\text{MMPBSA,com}} - G_{\text{MMPBSA,rec}} - G_{\text{MMPBSA,pep}} \tag{5.1}$$

Where each $\Delta G_{\text{MMPBSA,xxx}}$ term is calculated as follows:

$$\Delta G_{\text{MMPBSA,xxx}} = E_{\text{int,xxx}} + G_{\text{solv,xxx}} - TS_{\text{xxx}} \tag{5.2}$$

xxx refers to either pep, rec, or com., and $E_{\text{int,xxx}}$ is the internal energy, $G_{\text{solv,xxx}}$ is the solvation energy, and TS_{xxx} is the entropy multiplied by the temperature. [111–116]

The first term in eq. (5.2) is the internal energy that describes the self-interactions of the atoms in the considered molecule. It contains van der Waals, electrostatic, bond, angle, and dihedral energies. All of which are easily calculated, using accurate approximations described in the force field used to compute the simulation trajectory in the first place. For instance, NAMDEnergy, a VMD [85] plugin, can be used to calculate all of the above mentioned internal energies based on atom angles and distances using e.g. the CHARMM force field parameters [91]. Thus, one value for each of the internal energies will be computed for one individual frame.

Then there is the solvation energy. This is the energy associated with transferring a molecule from vacuum into a solvent with interactions between solute and solvent. This term is split in two parts [117]; one describing the non-polar solvation energy, and one describing the polar solvation energy. The former is usually determined based on a linear relation to the solvent accessible surface area (SASA) [118]:

$$G_{\text{solv,non}-\text{polar}} = \gamma \text{SASA} + \beta \tag{5.3}$$

Where γ is the the surface tension, $0.005\,42\,\text{kcal}/(\text{mol}\,\text{Å}^2)$ and the off set, β, is 0.92 kcal/mol [103]. The SASA is readily calculated using e.g. VMD [85] from single frames. The polar solvation energy is, when using the MM-PBSA approach, computed using continuum solvent methods [119], namely the Poisson-Boltzmann equations method [120, 121]. Here, the solvent is treated as a continium phase rather than as explicit water molecules. It is given as the difference between the polar interactions of the solute in solvent, e.g. water, and the solute in a reference state, e.g. vacuum. In practise, the program, APBS (Adaptive Poisson-Boltzmann Solver) [104] can be used to compute the polar solvation energy. Also here, the energetic contribution is calculated per frame.

This leaves only the entropy. Entropy is a measure of the number of degrees of freedom in a system; the more degrees of freedom, the higher the entropy. A quantitative way of interpreting this is to imagine entropy as a measure for the number of states that a system can explore out of its phase space or total configurational space. Several approximations have been developed to compute this part, mainly based on the configuration of a molecule throughout a simulation trajectory [122]. Meaning, that the entropy should be estimated for an ensemble of configurations [123].

This means that the internal and solvation energy terms are readily determined based on atom coordinates from single MD frames, whereas the entropy value requires considerations of several frames.

MD simulations, per se, cannot cover the total phase space of a system and, therefore, cannot give the exact entropy. Due to this, it is necessary to extract a limited number of frames from a trajectory and let these represent the part of phase space sampled.

When calculating entropy contributions for large biomolecules based on MD simulations, it is crucial to make sure that the chosen frames are statistically uncorrelated and that enough frames, spanning a large enough time interval, are used to obtain converged entropy approximations. As a result, the entropic contribution to binding is often neglected or completely ignored [115]. Even if the entropy is considered, and actual investigations are conducted, it was proven tedious and lengthy, if not impossible for the absolute entropic contributions to reach convergency in the investigated time/frame setup [113, 124]. However, several studies have shown the utter importance of including the entropy when computing binding energies in biomolecules [124–126].

Thus, the focus of this paper will be a methodology for computing independent and equilibrated entropy values from time resolved atom coordinates extracted from MD simulations of protein-peptide systems.

As mentioned, the entropy value of a biomolecular system depends on the entire phase space, that is, the covered microstates of the system [123]. However, when performing simulations, we do not cover the entire phase space, only a fraction of it is explored. Hence, feasible assumptions and valid approximations to the exact entropy are introduced.

Firstly, it is assumed that the total entropy of binding can be described as a sum of its individual contributions. There are three such contributions to consider when studying binding of a macromolecular system: The rotational and translational entropy, the configurational entropy, and the solvation entropy [123]. The rotational and translational entropy is directly connected with the overall rotation and translation of the receptor and peptide. A restriction, or loss, of this entropy for the peptide is observed upon binding when the protein and peptide can no longer move independently. The configurational entropy describes the loss of internal movement within the ligand and receptor. The solvation entropy covers the entropy change from a state of ordered solvent molecules, sitting around the ligand and receptor's binding pocket, to disordered molecules in the bulk solvent. The focus of this study will be the configurational entropy since it was previously shown that both solvation, rotational and translational entropies are insignificant and can be neglected [123, 127, 128]. Therefore, we set up a tangible procedure for calculating the configurational entropy from MD output of protein-peptide systems, with the purpose of making these calculations more straightforward when computing binding affinities using the MM-PBSA method. In addition, we will present a solid investigation of the dependency of entropy on both the frame count and the system, revealing that thousands of frames is essential for robust absolute entropy calculations.

The following includes a detailed description explaining every step on the way from MD output files to final binding entropy values. Hereafter, is a thorough examination of the number of frames that have to be considered, in order to generate consistent results, and how this varies for the different systems. The investigation is done for a membrane protein-peptide system (system 1) and a non-membrane protein-peptide system (system 2). System 1 is the glucagon-like peptide-1 (GLP-1) receptor in complex with a GLP-1 analogue. The peptide analogue includes two 16 C-atoms fatty acid chains attached to the two Lys residues via an L-γ-Glu linker.

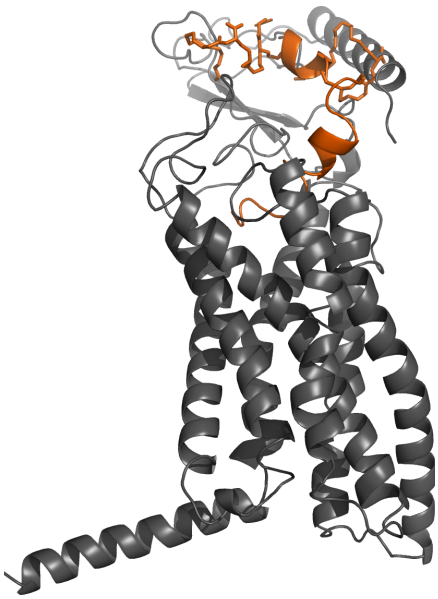


FIGURE 5.1: System 1 is a complex of the GLP-1 receptor (grey ribbon) bound to a GLP-1 analogue peptide (orange ribbon and sticks).

The receptor is constructed as a homology model, where the extracellular domain is based on the crystal structure (PDB ID 3IOL) and the transmembrane domain is based on two homologue templates (PDB IDs 4L6R and 3DQB). The structure of the GLP-1 analogue stems from the GLP-1 structure in the model developed by Coopman et al. 2011 [129]. System 1 is illustrated in fig. 5.1[1].

System 2 is the barnase receptor and barstar peptide, chain A and D in PDB ID 1BRS, respectively[2]. This system was chosen since previous entropy studies [114] have been performed on this particular system. However, the authors used a normal mode analysis approach, where we use a quasi harmonic analysis approach. Hence, system 2 will provide us with comparative knowledge. System 2 is illustrated in fig. 5.2.

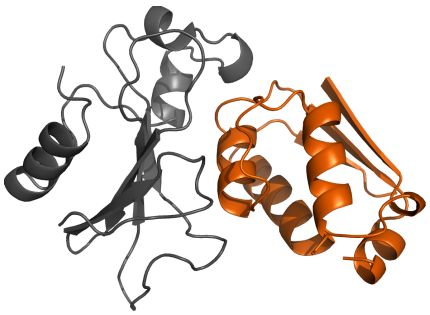


FIGURE 5.2: System 2 is a complex of the barnase receptor (grey ribbon) bound to barstar (orange ribbon).

[1]The structure setup differs from that described in sections 4.1.1 to 4.1.3 because the full length GLP-1 receptor structure (PDB ID 5NX2) was not available when the following investigations were initiated.

[2]Barstar is a 89 amino acid residue protein, but will be referred to as 'peptide' to keep consistent naming.

5.3 Methods and theory

In general, the configurational entropy can be described by the following equation:

$$S = -k_B \int p(r) \ln p(r) \, dr \tag{5.4}$$

Where k_B is the Boltzmann constant, $p(r) = \frac{e^{-\beta U(r)}}{\int e^{-\beta U(r)} dr}$ is the probability density function, and r is the $3N$-dimensional vector that represents the configuration. $U(r)$ is the potential energy of the system as a function of r, and $\beta = \frac{1}{k_B T}$, where T is the absolute temperature in K [130, 131].

Now, the trouble is to calculate the exact configurational entropy, from an MD simulation where only part of the phase space is covered. In practise, this is done by approximating the potential energy function, $U(r)$. One commonly used method for modelling $U(r)$ is based on quasi harmonic analysis [122, 130].

5.3.1 Quasi Harmonic Analysis (QHA)

As the name suggests, quasi harmonic analysis considers quasi harmonic movements. This is opposed to e.g. nomal mode analysis (NMA), which only describes the harmonic motions [122]. QHA, on the other hand, not only describes the harmonic fluctuations of the different configurations, but also partly covers the anharmonic motions, such as transformations between configurational states (different minima on the potential energy surface) [131]. There are, however, advantages and disadvantages of both methods: NMA does not account for anharmonicity, but it has been proven to provide absolute entropy values that converge faster than QHA [113, 132]. Furthermore, QHA approximates $U(r)$ by computing the determinant of the variance-covariance matrix, $\boldsymbol{\sigma}_c$, based on the Cartesian coordinates [131]. Contrary to this, NMA requires a transformation of Cartesian coordinates to a set of coordinates described by their relative position to an equilibrium configuration [133]. For our studies, we use QHA to cover partial anharmonicity and to avoid coordinate transformation. To ensure absolute entropy stability, we study the effect of frame counts and molecular system.

The Schlitter method uses QHA with $\boldsymbol{\sigma}_c$ in Cartesian coordinates, and gives the entropy, S', with good approximation to the absolute entropy [131]:

$$S' = \frac{1}{2} k_B \ln \det \left(\mathbf{1} + \left(k_B T e^2 / \hbar^2 \right) \mathbf{M} \boldsymbol{\sigma}_c \right) \tag{5.5}$$

Here, $\mathbf{1}$ is the identity matrix, $\mathbf{M}$ is the mass matrix with the atomic masses on the diagonal, and $\boldsymbol{\sigma}_c$ is the variance-covariance matrix of the Cartesian coordinates obtained directly from the MD simulation output. $\hbar$ is Planck's constant divided by 2π.

The resulting entropy value will be in units of J/K if the masses in $\mathbf{M}$ are given in kg and the variances in $\boldsymbol{\sigma}_c$ in m^2. However, atomic masses are usually given in atomic mass units, and most MD output files are based on Cartesian coordinates given in Å, not meters. Therefore, it is necessary to convert some of the input values in the Schlitter formula (eq. (5.5)). What is important here, is the fact that everything in the logarithm term has to be unit-less. Thus, the masses in $\mathbf{M}$, and the variances

(distance squared) in $\boldsymbol{\sigma_c}$ must have units that cancel out the remainder of those, which is the units of the Boltzmann constant, temperature and Planck's constant:

$$\frac{K_B T}{\hbar^2} = \frac{\mathrm{m^2\,kg\,s^{-2}\,K^{-1}\,K}}{\left(\mathrm{m^2\,kg\,s^{-1}}\right)^2} = \frac{1}{\mathrm{m^2\,kg}}$$

From this, it is clear that the masses must be given in kg (for carbon that is $m_C = 12.011\,\mathrm{amu} \times 1.660538860 \times 10^{-27}\mathrm{kg/amu} = 1.994473225 \times 10^{-26}\,\mathrm{kg}$) and the variance in $\mathrm{m^2}$ (for coordinates given in angstrom, i.e. $1\,\mathrm{\AA}^2 \times 10^{-20}\,\mathrm{m^2/\AA}^2 = 1 \times 10^{-20}\,\mathrm{m^2}$). This gives the following rewritten Schlitter entropy:

$$\begin{aligned} S' = \frac{1}{2} k_B \ln\det\Big(\mathbf{1} + \left(k_B T \mathrm{e}^2/\hbar^2\right)\mathbf{M} \\ \times 1.6605388 \times 10^{-27}\mathrm{kg/amu} \times \boldsymbol{\sigma_c} \times 10^{-20}\mathrm{m^2}\Big) \end{aligned} \tag{5.6}$$

This will give an entropy value in J/K, the unit of the Boltzmann constant.

In relation to the MM-PBSA binding energy, the unit for entropy should be given as energy per mole substance. Hence, the gas constant, R, is used instead of the Boltzmann constant (since $k_B N_A = R$) giving the entropy as $\mathrm{J/(K\,mol)}$. Lastly, a conversion to the more commonly used $\mathrm{kcal/(K\,mol)}$ can easily be done by dividing the final entropy value by 4184 kcal/J.

5.3.2 Methodology: Step-by-step guide

Simulation strategy

One of the core advantages of the MM-PBSA approach is that the binding energy can be computed based on a single MD simulation of the complex. The coordinates of the individual contributions are then extracted from this single simulation as peptide only, receptor only, and finally, the total complex [134]. Energy contributions for the two former, are then subtracted from the latter to give the binding free energy according to eqs. (5.1) and (5.2). The best way to obtain an average entropy value for the three components, is to divide the simulation into several batches. The variance-covariance matrices should then be computed for each of these batches of frames. Hereafter, one can calculate the entropy according to eqs. (5.5) and (5.6)[3].

Alignment of frames

As described in section 5.3.1, the configurational entropy depends on the variance and covariance of the coordinates for a sampled conformation. Thus, to get an actual entropy value for a conformation, e.g. the peptide, the peptide coordinates from a sufficient part of the simulation must be considered. As a result, the overall movement of the system becomes an issue, if not taken care of. This is in contrast to all the other contributions in the MM-PBSA energy calculations that are calculated as an average value over each frame. To get rid of the overall movement, and only consider the configurational changes, the individual frames of the studied simulation period has to be aligned to the first frame in each batch. This must be done for each of the individual components. That is, when calculating the entropy for the peptide,

[3] A detailed description of the simulation setup for both system 1 and 2 is given in Appendix Z.

only the peptide coordinates should be used to calculate the alignment, likewise for the receptor. When looking at the complex, it makes sense to use the receptor as reference, since this tends to move less than the ligand.

For statistical reasons, it is advisable to make several batches of frames from the simulation and then align the frames in each of these batches. This procedure will ensure the possibility to find an average entropy value representing the conformation well. Furthermore, it removes the overall rotation and translation of the system so that only the internal movement, i.e. the configurational changes, in each batch period are studied. More on this in section 5.4.1.

Selection of simplified system

In order to limit computational costs when calculating the Schlitter entropy, it is recommendable to select a simplified version of the system that still represents the overall configurational changes of the total system. This involves deciding how many atoms to include and what kind. Here, one can use the C-α atoms in the receptor and peptide chains [113]. In the case of system 1, where the peptide is acylated, it makes sense to consider the C-α atoms as well as the C atoms of the fatty acid chains. The receptor will be represented only by the C-α atoms, the same applies for barnase and barstar in system 2.

For further simplicity, only considering C atoms makes the construction of the mass matrix, **M**, much easier, since all the diagonal elements will be the same, that is, the mass of a C atom, 12.011 amu or $1.994\,473\,225 \times 10^{-26}$ kg.

Variance-covariance matrix

This section should clarify what the variance-covariance matrix covers. As the name implies, it is a measure for the variance, var, and the covariance, cov, of two or more data sets. The variance, i.e. the spread of the data, will be given on the diagonal, and the covariance, the co-dependencies of the different data sets, will be given as the off-diagonal elements, that is:

$$\mathrm{var}\,(X) = \sum_{i=1}^{N} (X_i - \bar{X})^2 / N \tag{5.7}$$

And:

$$\mathrm{cov}\,(X, Y) = \sum_{i=1}^{N} (X_i - \bar{X})\,(Y_i - \bar{Y})\,/N \tag{5.8}$$

Where X is a dataset with N points where the i^{th} point is X_i and the mean of the data points is $\bar{X}$, and where Y is a second dataset with N points where the i^{th} point is Y_i and the mean of the data points is $\bar{Y}$.

The variance-covariance matrix then becomes:

$$\boldsymbol{\sigma} = \begin{bmatrix} \sum (X_i - \bar{X})^2 / N & \sum (X_i - \bar{X})\,(Y_i - \bar{Y})\,/N \\ \sum (X_i - \bar{X})\,(Y_i - \bar{Y})\,/N & \sum (Y_i - \bar{Y})^2 / N \end{bmatrix} \tag{5.9}$$

From this it is evident that $\boldsymbol{\sigma}$ is symmetric [135].

Construction of σ: A simple example

Let's exemplify this (example data taken from StatTrek.com [135]) by considering five students' test results in math, English, and art:

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

There are three sets of data (math (M), English (E), and art (A) results) with five points in each. This will produce a 3×3 variance-covariance matrix.
The first step involves calculating the three means:

$$\bar{M} = \frac{(90+90+60+60+30)}{5} = 66$$
$$\bar{E} = \frac{(60+90+60+60+30)}{5} = 60$$
$$\bar{A} = \frac{(90+30+60+90+30)}{5} = 60$$

Now we can calculate the variances:

$$\mathrm{var}(M) = \frac{(90-66)^2+(90-66)^2+(60-66)^2+(60-66)^2+(30-66)^2}{5} = 504$$
$$\mathrm{var}(E) = \frac{(60-60)^2+(90-60)^2+(60-60)^2+(60-60)^2+(30-60)^2}{5} = 360$$
$$\mathrm{var}(A) = \frac{(90-60)^2+(30-60)^2+(60-60)^2+(90-60)^2+(30-60)^2}{5} = 720$$

Then the covariance between the different data sets can be calculated, that is between math and English, between math and art, and between art and English:

$$\begin{aligned}\mathrm{cov}(M,E) &= ((90-66)(60-60)+(90-66)(90-60)+(60-66)(60-60)\\ &\quad +(60-66)(60-60)+(30-66)(30-60))\,/5 = 360\\ \mathrm{cov}(M,A) &= ((90-66)(90-60)+(90-66)(30-60)+(60-66)(60-60)\\ &\quad +(60-66)(90-60)+(30-66)(30-60))\,/5 = 180\\ \mathrm{cov}(A,E) &= ((90-60)(60-60)+(30-60)(90-60)+(30-60)(60-60)\\ &\quad +(90-60)(60-60)+(30-60)(30-60))\,/5 = 0\end{aligned}$$

Resulting in the following covariance matrix:

$$\boldsymbol{\sigma} = \begin{array}{c} \\ M \\ E \\ A \end{array} \begin{array}{c} \begin{array}{ccc} M & E & A \end{array} \\ \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix} \end{array}$$

From this, we can see that the variance or spread is greatest among the art grades and lowest among the English grades. Furthermore, we can establish that the students that get high grades in math also get high grades in English and art. This means that if the students start getting higher math grades, they will tend to get higher grades in English and art and vice versa. We can see this from the positive covariance between these subjects. The covariance between English and art grades is, on the other hand, 0 which means that there are no detectable correlations between these. Had the covariance values been negative, it would mean that they were oppositely correlated, that is, high grades in one subject would mean low grades in the other subject [135].

Construction of σ: A high dimensional example

In the case of an MD simulation the output is atom coordinates instead of subjects, and instead of different students, it is different time steps. This will result in a data set (matrix) with the size of $n_t \times 3N$ where n_t is the number of time steps or frames during the studied simulation period, and N is the number of atoms. There will be $3N$ data points, since each atom is represented by 3 Cartesian coordinates; x, y, and z. The covariance matrix of such a data set will thus be a $3N \times 3N$ matrix. Here, the variance on the diagonal will give the spread of each x, y, and z coordinate for each of the considered atoms and the covariances on the off-diagonal will give the correlation between them.

Construction of σ: How to get it

To obtain the covariance matrix from an MD output file (e.g. a dcd file), one needs a program or piece of software that can handle compact and large data files. One such program is ProDy [107, 108].

The program is run to get covariance matrices by typing the following in a terminal window:

```
$ prody eda -s [atom selection] -v -p [prefix name] --pdb [pdb_filename.
    pdb] --aligned [dcd_filename.dcd]
```

ProDy calls the program `eda`, which is an abbreviation for essential dynamics analysis, `-s` is a flag for a selection which requires the atom selection in quotation marks in Python/PyMOL syntax [136]. `-v` specifies that the covariance matrix should be outputted, and the `-p` flag sets a prefix name for the output file. `-pdb` specifies the pdb file containing the atom names and types used to give the selection and requires a pdb filename. `--aligned` is a flag that indicates that the frames have already been aligned. Last is the name of the trajectory file to extract the variance-covariance matrix from.[4]

[4]See Appendix Y for a code snippet that calculates the variance-covariance matrix.

5.4 Results and discussion

The following contains an investigation of how to obtain independent and reliable Schlitter entropy values for peptide-receptor binding based on system 1 and system 2.

5.4.1 Calculation setup: Effect of frame number on absolute entropy

When calculating the internal and solvation energy terms, one can easily obtain a mean value and the standard deviations, since only one frame is needed for each of the calculations. This is, however, not the case for the entropy, since several frames are needed for the calculation of one variance-covariance matrix, and thus one entropy value. By dividing the simulation trajectory into batches, several entropy values can be used to obtain a mean and corresponding standard deviation (SD) for one simulation. But how many batches should be used? And how many frames should be in each batch? There are two major things to consider when answering these questions. First, it is important that enough frames, with enough spacing, are used in each batch when calculating the variance-covariance matrix. Otherwise, it will not be independent. When enough frames are used, the mean value will converge. Second, the number of batches will affect the SD value, the more batches the lower the SD value. However, the number of batches has to be balanced against the number of frames in each batch and the total number of frames available in the simulation trajectory.

To give a clear view on the mentioned effects, several setups were tested using a single simulation of system 1 and a single simulation of system 2. Firstly, we calculate the absolute entropy of system 1. This is done for the peptide, receptor, and complex, and is shown as a function of frames used to construct the variance-covariance matrices. Initially, this is done using two batches, see fig. 5.3 for the resulting absolute entropies.

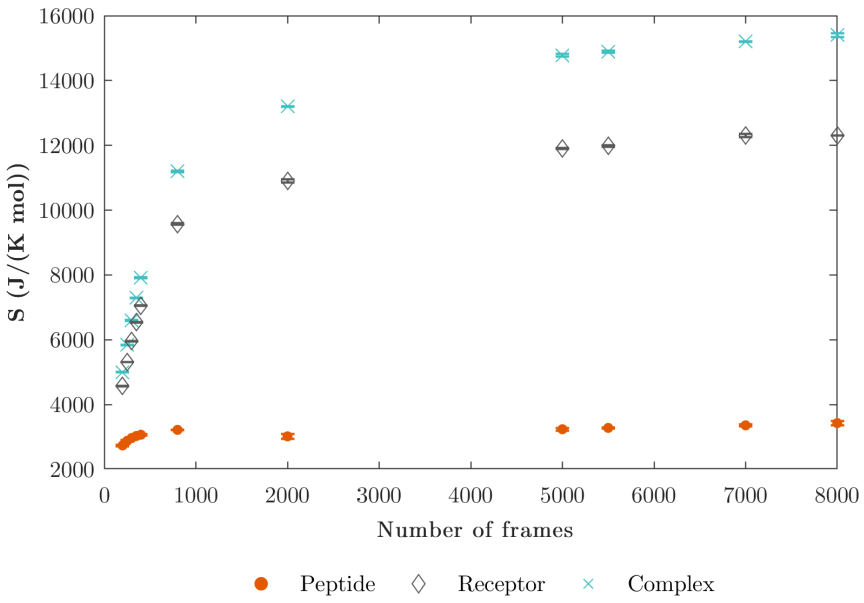


FIGURE 5.3: Absolute entropy values for the peptide, receptor, and total complex as a function of the used number of frames. Results shown for system 1 using two batches.

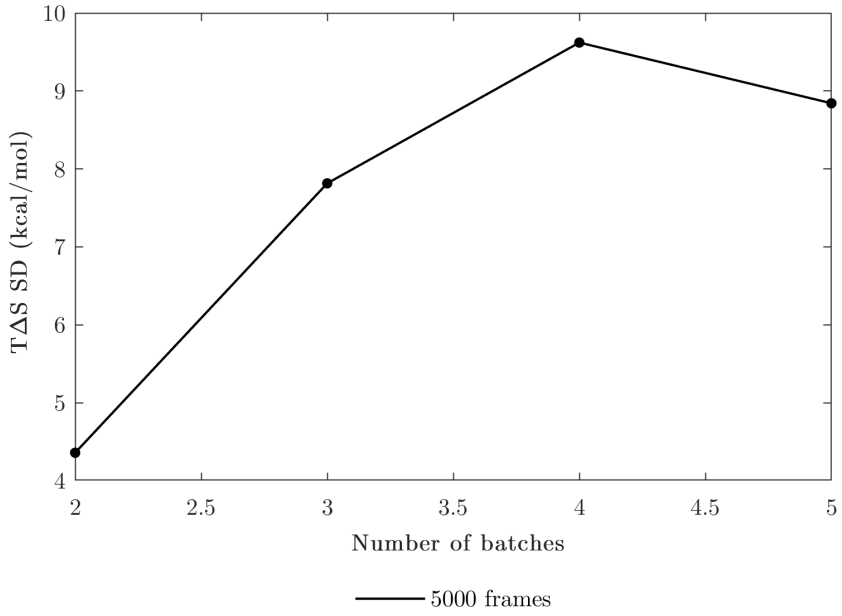


FIGURE 5.4: Standard deviation (SD) on $T\Delta S$ as a function of batches used. Each batch contains 5000 frames. Results shown for system 1.

Figure 5.3, shows how the absolute entropy for the peptide (orange dots) converge using only 800 frames per batch, whereas both receptor (grey diamond) and complex (teal $\times$) entropies do not converge until 5000 frames per batch are used. Since the simulation was run with a 2 fs step size, and the coordinates were saved every 1000^{th} steps, the time interval between each snapshot equals 2 ps. This totals to 3.2 ns for the peptide entropy to converge, and 20 ns for the receptor and complex entropy to converge. 2 ps should be just enough to ensure that the individual frames considered are not correlated [113].

Thus, when calculating absolute entropies of system 1, one has to use at least 5000 frames, covering no less than 20 ns (using two batches) to ensure converged entropy values.

The SD values on the average binding entropy ($T\Delta S = T \times (S_{\text{com}} - S_{\text{rec}} - S_{\text{pep}})/$ 4184 kcal/J) can now be investigated. The reason for choosing the standard deviation on $T\Delta S$ is because we, at this stage, already know that the entropy has stabilised.

In regards to the impact of the number of batches, the SD values on the average $T\Delta S$ are plotted in fig. 5.4. Calculations were performed for 2, 3, 4, and 5 batches, where each individual batch contained 5000 consecutive frames. This totals to 20, 30, 40, and 50 ns for 2, 3, 4, and 5 batches respectively.

It can be seen that the SD values initially increases, most likely due to the fact that a bigger part of the trajectory is covered when using more batches. It does, however, converge as the number of batches increases. Figure 5.4 shows that the standard deviation on $T\Delta S$ converges at around four batches. The preferred setup for calcualting entropy values for system 1 is, then, using four batches, each containing 5000 frames.

Having looked at system 1, we now turn to system 2, which is significantly smaller with only 197 residues as opposed to the 434 residues in system 1. We perform the same investigation as for system 1 in regards to the absolute entropy of the three components (peptide, receptor, and complex), and the result is plotted in fig. 5.5.

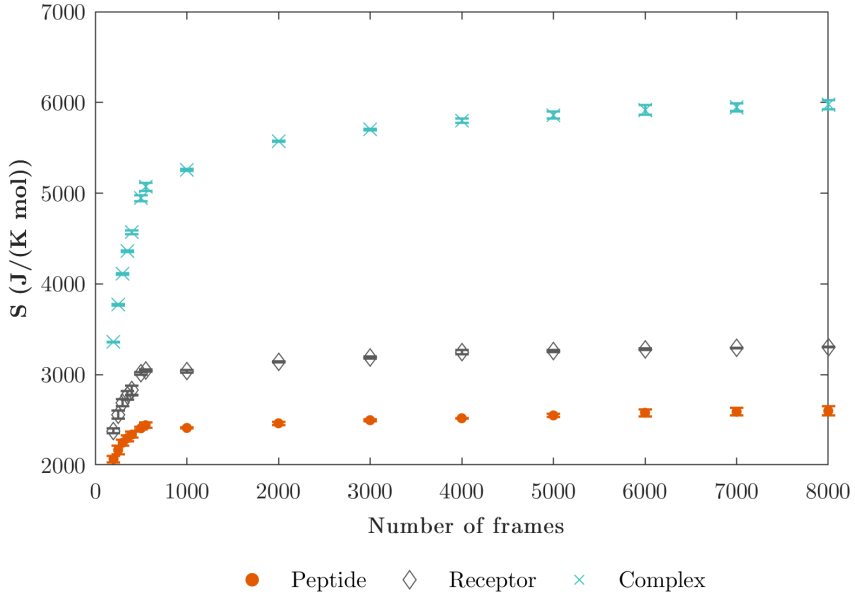


FIGURE 5.5: Absolute entropy values for the peptide, receptor and total complex as a function of the used number of frames. Results shown for system 2 using two batches.

Figure 5.5, shows how the peptide entropy (orange dots) converges first, already at around 550 frames. Contrary to system 1, the receptor entropy (grey diamonds) converges similar to the peptide, but, the complex entropy (teal $\times$) converges later, at 2000 frames. This corresponds to 2.2 ns for the peptide and receptor and 8 ns for the complex. Thus, for this smaller system, only 2000 frames has to be considered to ensure entropy convergence. That is, for system 2, the entropy converges more than twice as fast, needing less than half of the frames.

Again, we now turn to the effect of batch number on the standard deviation. SDs are calculated, like it was done for system 1, and are plotted in fig. 5.6 for system 2.

Unlike before, we do not see a convergence on the SD values within the available range of frames. This most likely stems from the fact that the entropy is calculated for a larger time interval when more batches are used. Meaning that when using more batches, the entropy change also covers a bigger part of the trajectory. However, the maximum difference between the SD values is only around 1 kcal/mol, and the maximum SD is around 2.5 kcal/mol, which is much less than what is usually seen for the internal and solvation energies. On this note, a reliable setup for entropy calculations, for systems 2, seems to be when using three batches with 2000 frames. This will cover 12 ns.

Comparing our results to those found by Liang et al. 2009 [114] who previously calculated $T\Delta S$ for system 2 using NMA. The authors report an entropy difference of binding of $T\Delta S_{\text{NMA}} = (-47.0 \pm 1.2)$ kcal/mol. With the above mentioned setup of three batches including 2000 frames each, we compute the entropy difference of binding for system 2 to be $T\Delta S_{\text{QHA}} = (-2.4 \pm 1.5)$ kcal/mol using QHA. It appears that our value does not match the one reported by Liang et el. 2009. However, there are two main reasons for this. First, the NMA entropy difference of binding calculated by Liang et el. 2009 is based on only 180 frames, covering 5 ns. Even though they use the NMA approach that converges faster than QHA, it is likely that the absolute entropy has not converged yet. Additionally, no investigations were performed by the authors to investigate the impact of the number of frames on the

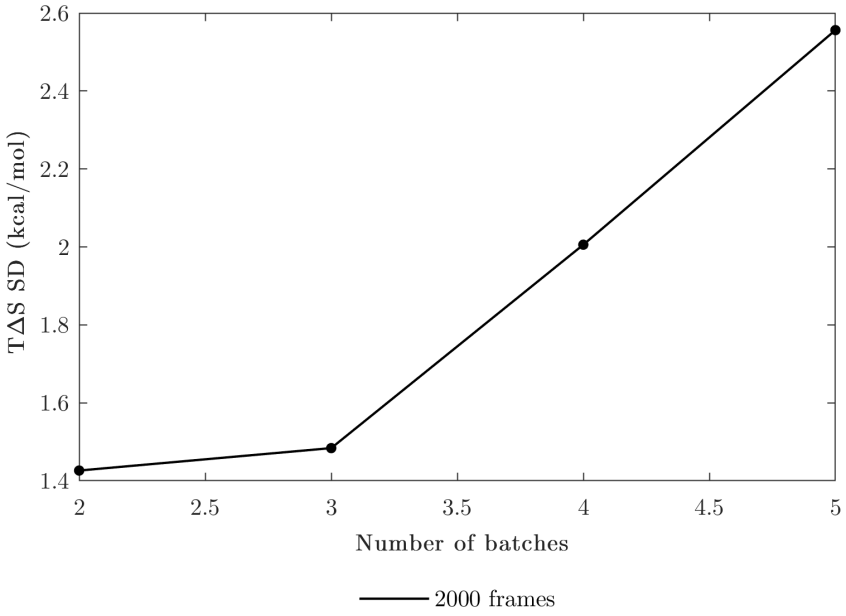


FIGURE 5.6: Standard deviation (SD) on $T\Delta S$ as a function of batches used. Each batch contains 2000 frames. Results shown for system 2.

entropy. Secondly, NMA does not account for the anharmonicity which means that QHA absolute entropies, per se, will be higher than NMA entropies. To give an idea of the actual impact of not doing such studies on the entropy, we have calculated the QHA entropy for system 2 using only 180 frames and three batches. The resulting entropy difference of binding is $T\Delta S_{\text{QHA,180 frames}} = (-84.0 \pm 1.9)$ kcal/mol. Hence, our point is that absolute entropy calculations should not be trusted unless the calculation setup has been investigated, finding the optimal setup to reach converged absolute entropies. MM-PBSA binding energies will be misleading when the entropies have not converged. This is not only in relation to the number of frames used, but also the size of the investigated system. We have proven that both have a significant impact on the calculated entropy values.

5.5 Conclusion

When performing binding free energy calculations using the MM-PBSA method, especially one part was proven tricky to accurately compute, namely the entropy. The quasi harmonic analysis method presented by Schlitter can give a fair estimate of the binding entropy using the Cartesian coordinates extracted directly from an MD simulation. However, it has been proven to be quite time consuming to obtain converged entropy values for protein-peptide systems, as a large number of frames and thus long simulations are needed. In conclusion we, therefore, recommend that a thorough test is conducted to get to a feasible setup. We have shown that one has to study the behaviour of the absolute entropy for the three complex constituents (the peptide, receptor, and complex) to reveal a solid procedure that provides converged entropy values. This is due to the great impact that the number of frames used and the size of the system, have on the resulting entropy.

Our investigation clearly shows that one must perform several calculations of the absolute entropy, for instnace starting out using two batches and increasing frame numbers. Thus, the adequate number of frames to be used can be found from the

point of convergence. Next, it is advisable to compute several entropy values - using the afore found frame count - and then increase batch numbers. From this, the standard deviation on the entropy difference of binding should be investigated for convergence, to reveal the appropriate number of batches to be used.

When considering systems of biomolecules, our study shows that several thousands of frames, summing up to tens of nanoseconds, must be included in the entropy computations to gain solid binding entropies.

Part 2: Experimental investigations

Chapter 6

Experimental introduction

This part of the thesis will cover experimental studies of varius GLP-1 analogues and their oligomerisation and interactions to HSA. Firstly, chapter 7 gives a short introduction to the relevant experimental theory. Then, a description of the peptide synthesis is given in chapter 8, followed by the experimental findings concerning oligomrisation and HSA binding of the first set of peptides in chapter 9.

6.1 GLP-1 acylation and half-life

As for binding interactions, it has been proven that different acylated GLP-1 analogues affect half-life differently. Some analogues prolong the half-life to a few hours and others up to several days [52, 68].

This extension of half-life primarily stems from two things. The first is interaction with HSA through the attached FA chain [137–139]. HSA is often targeted in relation to half-life extension of drugs since HSA is capable of binding many different molecules and has a half-life of around three weeks. When the GLP-1 analogues are bound to HSA, DPP IV cannot bind to the peptide backbone and cleave it between Ala8* and Glu9*, which results in the inactive peptide. Furthermore, the fact that the peptide has to dissociate from HSA to become available also prolongs the systemic half-life [57, 140, 141]. The other reason for half-life extension is due to oligomerisation. The attached FA chain on the GLP-1 analogues will cause the individual peptide molecules to oligomerise and form larger, stable, species [58, 142]. When oligomerised, the individual peptides have to dissociate to monomers to become biologically available [142, 143]. Furthermore, DPP IV cleavage is also prohibited when the peptides oligomerise due to steric hindrance [56]. In both cases, renal clearance is also prevented. All are properties that contribute to an *in vivo* extension of the half-life compared to native GLP-1 [52, 142]. Hence, knowledge on both oligomerisation and HSA interactions are desirable in relation to creating acylated GLP-1 analogues that will in fact increase the half-life, but without compromising the biological effect.

Acylation in itself will induce both peptide oligomerisation and HSA binding. However, studies indicate that both properties can be controlled by pH and ionic strength. For instance, liragutide could be driven from an octamer to a dodecamer by regulating pH [144], and ionic strength could determine the oligomeric state of acylated melittin, a 26 amino acid peptide from bee venom [145]. Concerning HSA interactions, ionic strength could increase binding to HSA due to screening effects.

6.2 GLP-1 analogues

In the development process of liraglutide, the half-life for a number of acylated GLP-1 analogues was measured. Here, it was evident that acylation with FA chains increased the half-life compared to native GLP-1. In fact, the half-life increased as a function of the FA chain length, although, FA chains of 12 C-atoms or longer resulted in substantially extended half-lives. As an example, FA chains of 10 C-atoms incresed the half-life to around 0.8 h and FA chains of 11 C-atoms increased the half-life to 5.1 h, whereas FA chains of 16 C-atoms extended the half-life to 16 h [67]. The presence of a linker was investigated, and addition of an L-γ-Glu linker did not seem to affect the half-life, even though it had a postive impact on potency [52]. Neither did the introduction of an ω acid in the FA chain tail affect the half-life, which on the contrary, was shown to decrease potency [52]. Both addition of linker and ω acid increases the net negative charge of the peptide, which is meant to increase HSA-peptide interaction [67].

The selection process for the investigated peptides was driven by the possibility for in-house synthesis by Assistant Professor Katrine Qvortrup. Therefore, the backbone chain length was chosen to be as short as possible without compromising the helical structure seen in GLP-1. Furthermore, only one acylation site was chosen (Lys26*) to make larger batch preparation possible. Meaning, that if only one acylation site is chosen, the two peptide parts surrounding the acylation site could be made in larger portions, and connected to different Lys residues with the desired FA chain and/or linker attached. The FA chains were chosen to cover short (six C-atoms), medium (eight C-atoms) and longer (10-16 C-atoms) chains, with and without L-γ-Glu linker and ω acids. The schematic structures of the chosen peptides are shown in figs. 6.1 to 6.11 with details given in table 6.1.

TABLE 6.1: The peptides investigated in regards to oligomerisation and HSA binding. Cxx indicates the length of the FA chain. Seq. = GLP-1 sequence. L indicates presence of L-γ-Glu linker. A indicates presence of a free acid group. rev means that the peptide chain was reversed (C to N). Cs + linker is the total number of C-atoms in the acylated residue beyond the original Lys side chain. syn = synthesised, com = commercially acquired.

ID	Seq.	FA length	Linker	ω acid	Acyl site	Mutation	Cs + linker	Origin
GLP-1(14-34)	14-34	0	✗	✗	-	K34*R	0	syn
GLP-1(7-36)	7-36	0	✗	✗	-	-	0	com
C6Arev	34-14	6	✗	✓	Lys26*	K34*R	6	syn
C6A	14-34	6	✗	✓	Lys26*	K34*R	6	syn
C8A	14-34	8	✗	✓	Lys26*	K34*R	8	syn
C8LA	14-34	8	✓	✓	Lys26*	K34*R	13	syn
C10	14-34	10	✗	✗	Lys26*	K34*R	10	syn
C16	14-34	16	✗	✗	Lys26*	K34*R	16	syn
C10L	14-34	10	✓	✗	Lys26*	K34*R	15	syn
C16L(14-34)	14-34	16	✓	✗	Lys26*	K34*R	21	syn
C16L(7-37)	7-37	16	✓	✗	Lys26*	K34*R	21	com

Please note that C16L(7-37) is liraglutide, but will be referred to as C16L(7-37) in the remaining part of the thesis.

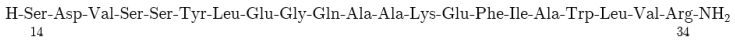


FIGURE 6.1: Schematic structure of GLP-1(14-34).

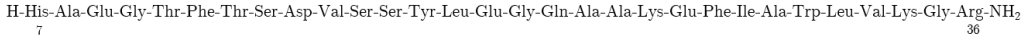


FIGURE 6.2: Schematic structure of GLP-1(7-36).

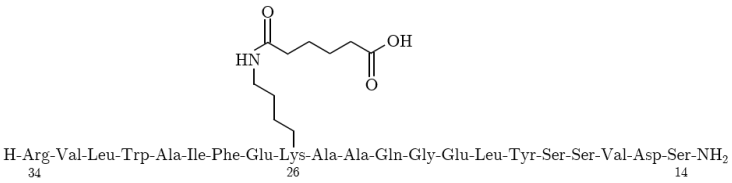


FIGURE 6.3: Schematic structure of C6Arev. The amino acid sequence of GLP-1 is reversed.

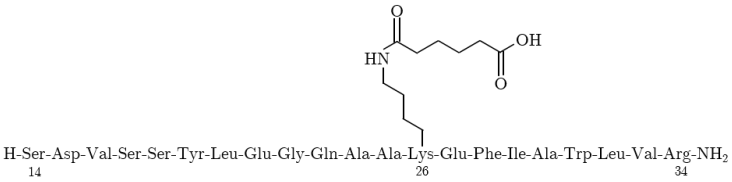


FIGURE 6.4: Schematic structure of C6A.

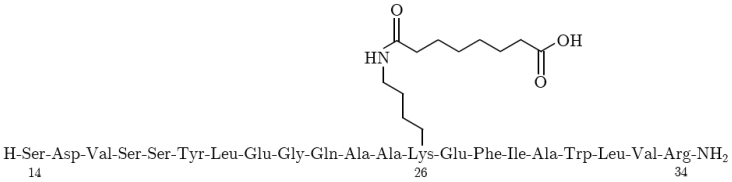


FIGURE 6.5: Schematic structure of C8A.

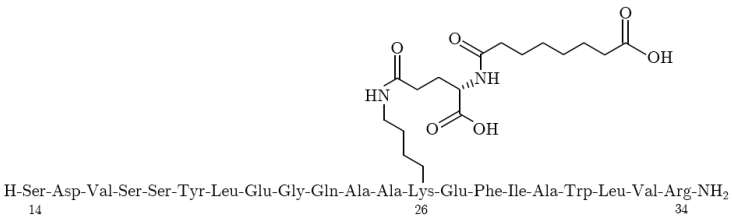


FIGURE 6.6: Schematic structure of C8LA.

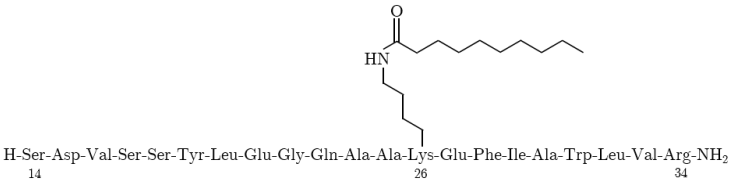


FIGURE 6.7: Schematic structure of C10.

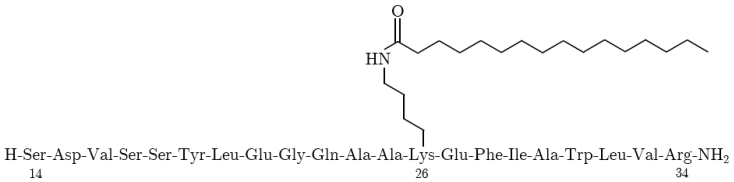


FIGURE 6.8: Schematic structure of C16.

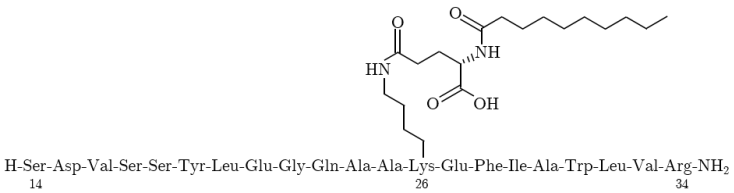


FIGURE 6.9: Schematic structure of C10L.

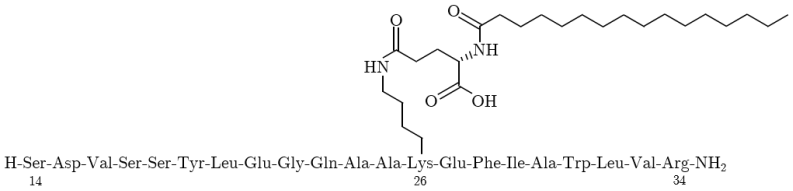


FIGURE 6.10: Schematic structure of C16L(14-34).

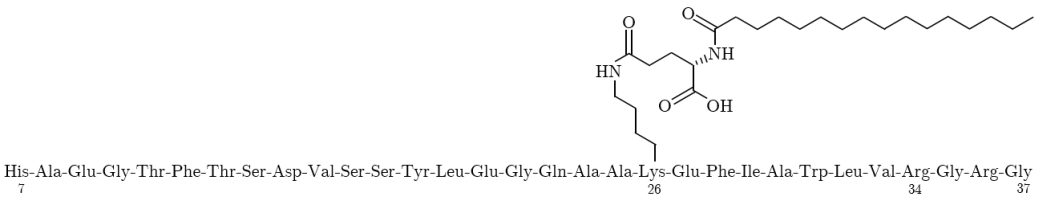


FIGURE 6.11: Schematic structure of C16L(7-37) (liraglutide).

6.3 Aim of the experimental part

The aim of the following is to look at the differences in oligomerisation and HSA interaction between the peptides presented in section 6.2. Such investigations were undertaken by use of flow-field-flow separation followed by in-line static light scattering, complimented by batch dynamic light scattering experiments, as schematically demonstrated in fig. 6.12. In combination, dynamic and static light scattering can disclose information on the molecular mass (oligomeric state) and size dimensions of the acylated peptides as well as peptide/HSA mixtures.

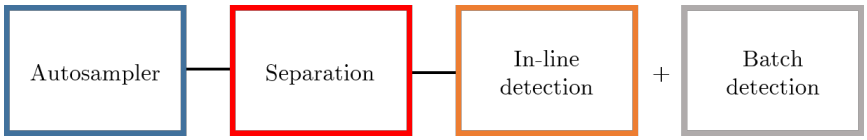


FIGURE 6.12: A schematic illustration of the experimental setup used for solution structure and HSA interaction investigations.

Chapter 7

AF4 and LS theory

7.1 Asymmetric flow-field-flow fractioning

Asymmetric flow-field-flow fractioning (AF4) is a separation method based on liquid flows. It exploits the differences in diffusion rates of molecules of different sizes (radius of hydration, R_h). Separation is obtained by applying two liquid flows perpendicular to each other above a membrane, through a channel chamber. The chamber is created by sticking two plates against each other with a spacer with a cut-out in it and a membrane in between [146]. This is demonstrated in fig. 7.1.

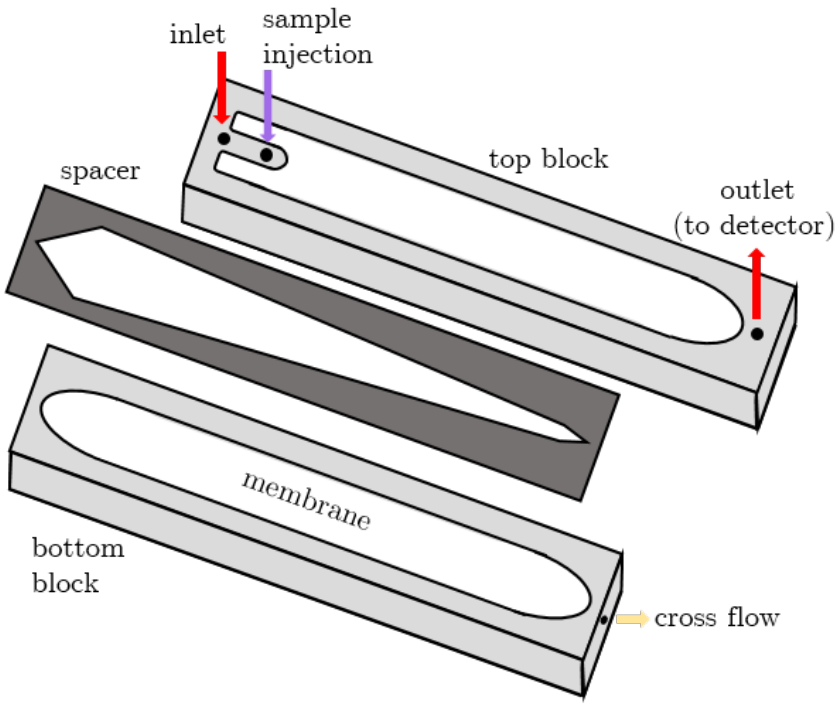


FIGURE 7.1: Schematic representation of an AF4 channel with spacer and membrane. Arrows indicate flows in and out of the channel, red is the elution flow, yellow is the cross flow, and purple the sample inlet.

The separation chamber is made up by an elongated trapezoidal shaped hole (spacer cut-out) as the side walls, and the top and bottom blocks as lid and bottom, respectively. The channel is constructed so that a carrier flow (Vd) can be run from inlet to outlet, above the membrane, and another from inlet to cross flow outlet (Vx), across the membrane (see fig. 7.2). Below the membrane is a porous frit which the cross flow permeates. This allows separation of molecules in the carrier flow above

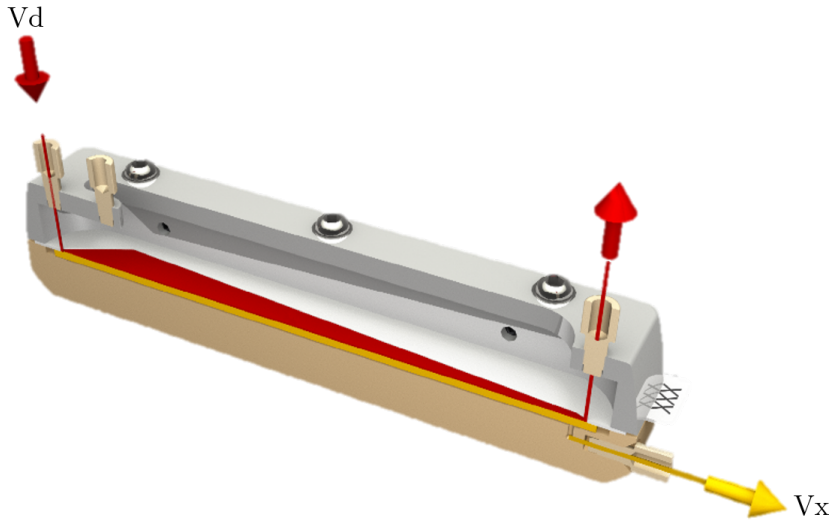


FIGURE 7.2: Carrier and cross flow through separation channel. Red arrows indicate the sample containing flow and the yellow, the cross flow. Figure adapted from Wyatt Technology Europe LSU course material [148] © by Wyatt Technology Corporation – used with permission.

the membrane, since the sample molecules are kept in the chamber by the membrane [147].

As indicated in figs. 7.1 and 7.2, the actual channel volume is shaped as an elongated trapezoidal diamond. This is done to make up for the loss of longitudinal flow velocity in the cross flow when the inlet is in one end and not through the entire top plane [149]. For demonstration purposes, the cross flow is shown as arrows perpendicular to the membrane plane. They are drawn at the maximum height off the sample volume in the channel, shown in fig. 7.3. The maximum sample height, at approximately 10 µm, is not the actual height of the spacer (which is between 350 µm to 490 µm). Meaning, that the channel volume is much larger than indicated in fig. 7.3.

In the AF4 channel, separation is obtained by injecting sample trough the sample injection inlet and then applying a focus mode where the outlet flow is reversed

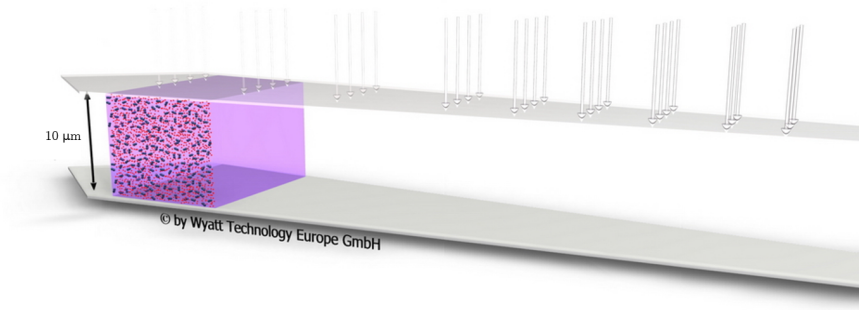


FIGURE 7.3: Representation of the assymetric channel volume to counteract the fact that the cross flow is not applied through the top plane, but through the inlet. Figure adapted from Wyatt Technology Europe LSU course material [148] © by Wyatt Technology Corporation – used with permission.

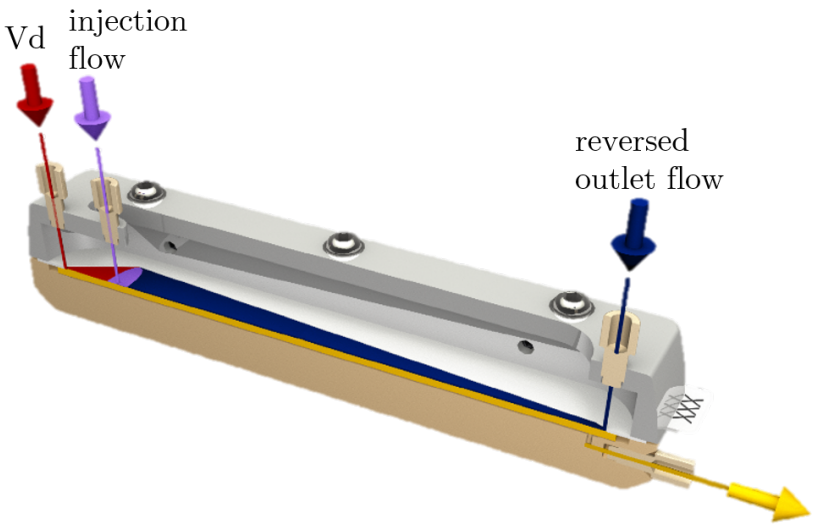


FIGURE 7.4: Injection and focus steps in the AF4 channel. The red arrow indicates the carrier flow inlet, the purple arrow is the sample injection, and the blue the reversed outlet flow. Figure adapted from Wyatt Technology Europe LSU course material [148] © by Wyatt Technology Corporation – used with permission.

to concentrate the sample in the front of the channel and create a initial separation [147]. The injection and focus step is shown in fig. 7.4. This step is demonstrated in fig. 7.5 where the inlet and the reversed outlet flows concentrate the sample at the beginning of the channel, and thus allows for a initial separation when the elution mode is applied.

To obtain full separation, elution mode is started with the carrier flow along the channel and the cross flow across the membrane. The carrier flow has a laminar flow profile. The flow profile is curved with a faster flow in the middle of the channel, and slower flow near the top and bottom of the channel [150]. Since the maximum sample height is only around 10 µm from the bottom, and the actual channel volume is at least 30 times higher than that, the curved laminar flow will be faster at the top of the maximum sample height and slower at the membrane (see fig. 7.6 for

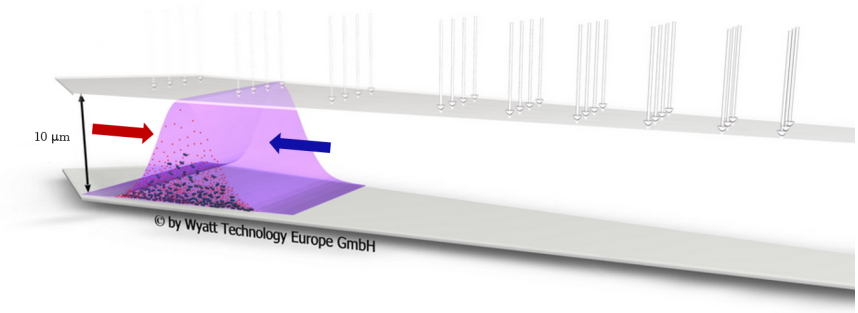


FIGURE 7.5: Sample after injection as a purple mixture, is concentrated at the beginning of the channel when the inlet flow (red arrow) and the reversed outlet flow (blue arrow) flows 'against' each other. Here, a pre-separation is obtained before elution. Figure adapted from Wyatt Technology Europe LSU course material [148] © by Wyatt Technology Corporation – used with permission.

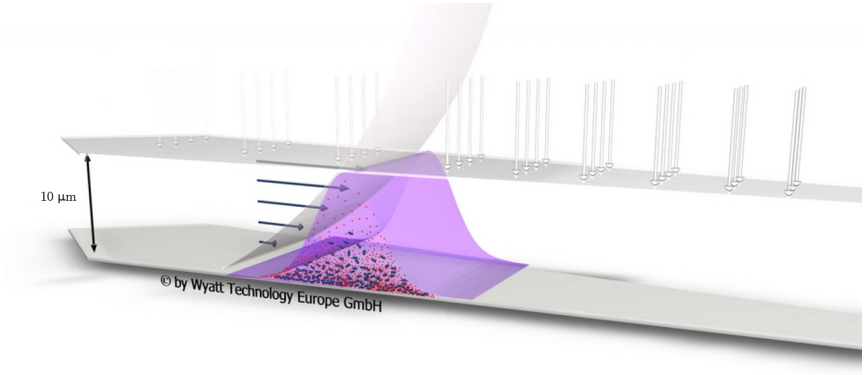


FIGURE 7.6: Elution mode in the channel. The laminar carrier flow, indicated by a curve and horizontal arrows, move the particles along the channel faster, further away from the bottom. Figure adapted from Wyatt Technology Europe LSU course material [148] © by Wyatt Technology Corporation – used with permission.

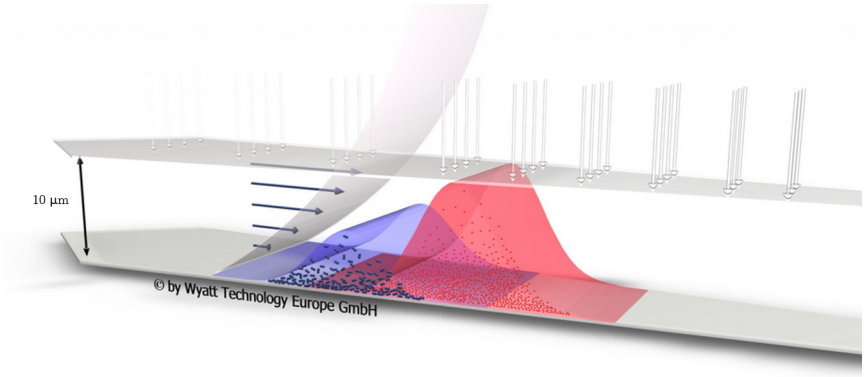


FIGURE 7.7: Initial separation of the different sized particles in the sample. The smaller particles diffuse faster away from the bottom, against the cross flow. Figure adapted from Wyatt Technology Europe LSU course material [148] © by Wyatt Technology Corporation – used with permission.

demonstration). Thus, the cross flow force the sample molecules towards the bottom of the channel, but they will counter this movement by diffusion away from the bottom. This is what makes the actual separation possible. The diffusion rate is determined by the size of the particles, and smaller particles diffuse faster than larger particles. Hence, smaller molecules will diffuse further away from the bottom and thus be carried along the channel higher and faster than larger particles, due to the laminar flow [151]. Figure 7.7 shows how the smaller particles (red) are carried away from the larger particles (blue) near the bottom wall. In fig. 7.8 there is full separation and the purple sample mixture has been separated into smaller red particles and larger blue particles. The smaller red particles will elute first, and the larger blue will elute last.

An AF4 separation run then consist of several steps: First a short elution to rinse out possible impurities before the sample is injected. This is followed by a short focus step which will be maintained while sample is injected (focus+injection) and last for 1 to 2 min afterwards, to ensure good focussing. After this, elution mode is started with the cross flow. The cross flow velocity can be changed doing this

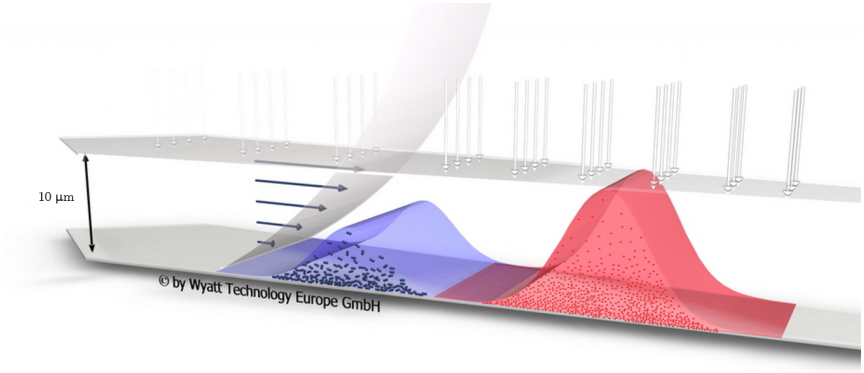


FIGURE 7.8: Full separation of the sample species. The different species have different elution times - the smaller particles elute first. Figure adapted from Wyatt Technology Europe LSU course material [148] © by Wyatt Technology Corporation – used with permission.

step, e.g. via a linear or exponential decrease [152]. Finally, there is a short elution+injection step with no cross flow. This not only ensures that all sample has left the channel, but also cleans the loop and valves from possible leftover sample or impurities.

In combination, these steps lead to sample separation within the AF4 channel which can then be investigated, in-line, by leading the carrier flow through different detectors, e.g. light scattering and refractive index detectors [153]. For the experiments carried out in this study, AF4 separation was followed by UV, multi angle light scattering, and refractive index (RI) detection.

7.2 Light scattering

Rays of light consists of magnetic (H) and electric (E) fields oscillating in waves. Figure 7.9 illustrates this.

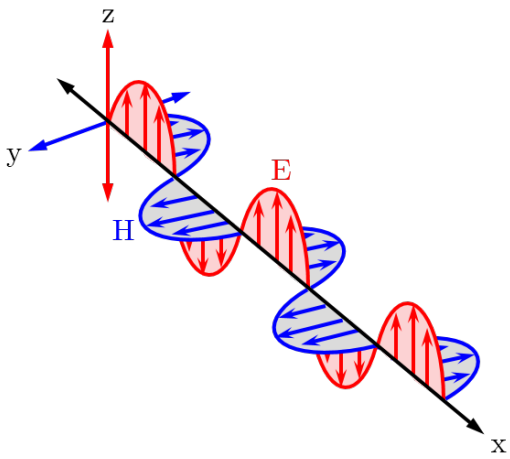


FIGURE 7.9: Light consists of a magnetic and an electric field oscillating as waves. The blue wave (H) indicates the magnetic field, and the red wave (E) indicates the electric field. Figure adapted from Wyatt Technology Europe LSU course material [154] © by Wyatt Technology Corporation – used with permission.

In relation to light scattering (LS), it is the electric field that is of importance, since it is the electric field that causes particles to scatter light. This happens when the electric field polarises the charges in the particle, so they oscillate [154]. This is shown in fig. 7.10.

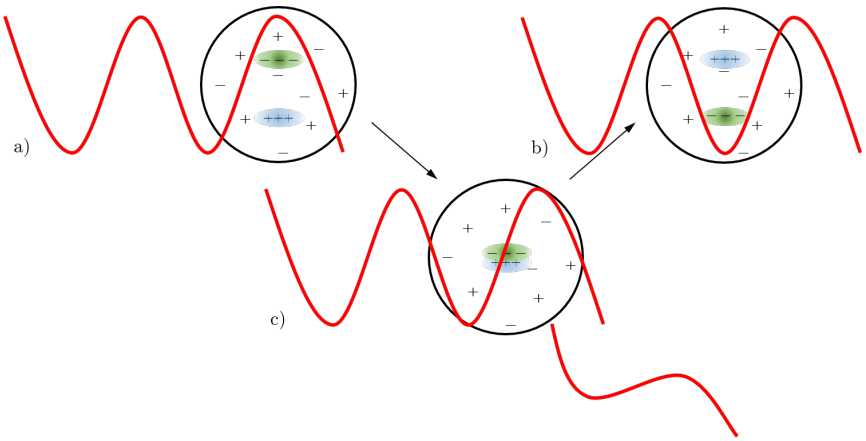


FIGURE 7.10: The principle of a particle being polarised with the electric field. a) shows how the charges are polarised with negative charges above and positive charges below. b) shows the particle with opposite polarisation due to the travelling electric field wave. c) illustrates the actual switching of the poles which reradiates (scatters) light. Figure adapted from Wyatt Technology Europe LSU course material [154] © by Wyatt Technology Corporation – used with permission.

The scattered light that is reradiated from the change in charge polarisation, shown in c) in fig. 7.10, gives rise to the principle of light scattering [154].

7.3 Static light scattering

When several scattering centres are present, the detected reradiated light will be a sum of scattered light measured as a total light intensity. Measurements of the total intensity of the scattered light, independently of time, is referred to as static light scattering (SLS). The summed light can be coherent or incoherent. If the detected light stems from independent sources, the electric fields will add independently since their phases are different. On the other hand, if the sources are identical, the resulting scattered light waves will have the same phase relationship, hence, the electric fields can interfere [154]. See fig. 7.11 for illustration.

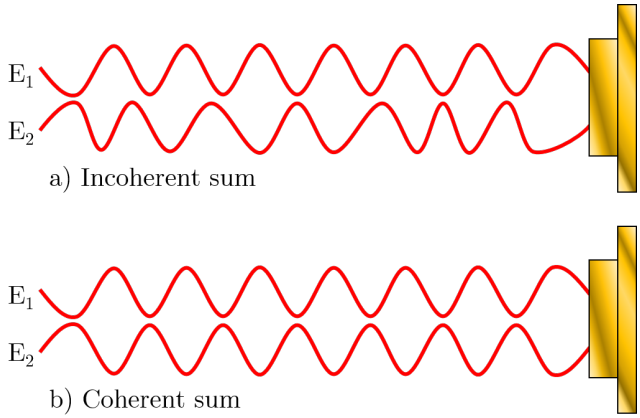


FIGURE 7.11: Coherent and incoherent addition of light. a) shows addition of two electric fields (E_1 and E_2) originating from independent scattering sources, incoherent summation. b) shows addition of two electric fields (E_1 and E_2) originating from identical scattering sources, coherent summation. Figure adapted from Wyatt Technology Europe LSU course material [154] © by Wyatt Technology Corporation – used with permission.

When two identical electrical waves interfere coherently, the relation can happen in a constructive or in a destructive fashion (see fig. 7.12) [154].

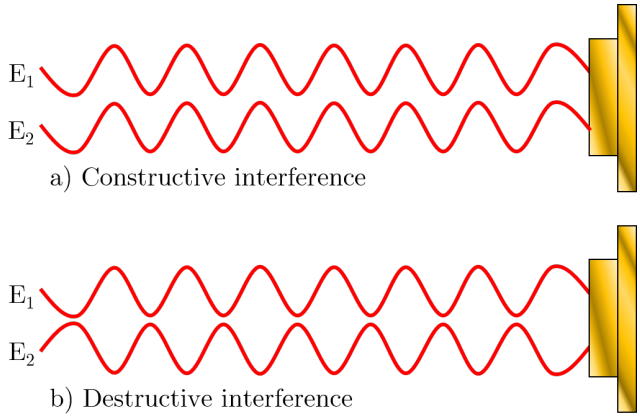


FIGURE 7.12: Coherent light can sum constructively, a) or destrcutively b). Figure adapted from Wyatt Technology Europe LSU course material [154] © by Wyatt Technology Corporation – used with permission.

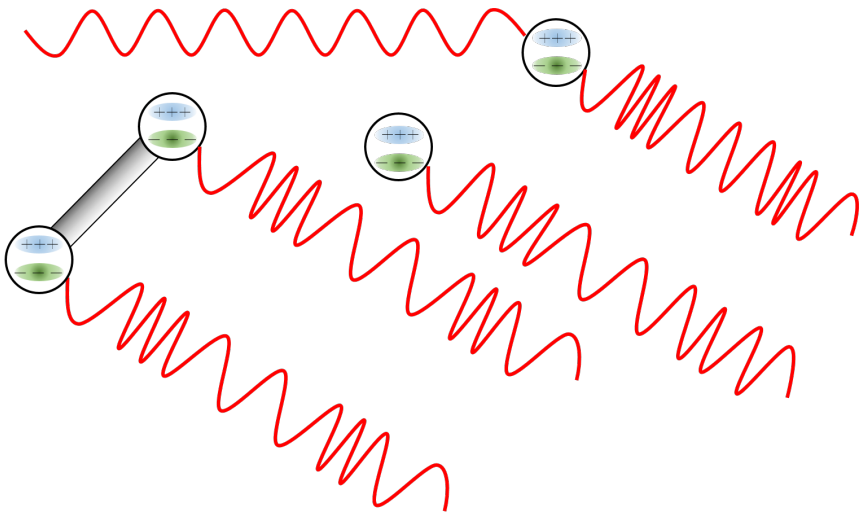


FIGURE 7.13: Coherent and incoherent addition of scattered light from a dimer (left) and monomers (right). Figure adapted from Wyatt Technology Europe LSU course material [154] © by Wyatt Technology Corporation – used with permission.

It is the very principle of coherent and incoherent light scattering that allows measurement of a scattering particle's mass. For a solution of identical scattering centres, with identical mass and scattering properties, the scattered light will add coherently if e.g. two particles are joint together and move together. Whereas, two particles not connected will move independently and randomly, based on Brownian motion. The result is incoherently added light [154]. Figure 7.13 depicts this.

The measured total scattering intensity thus depends on the difference between the incoherent and coherent addition of the scattered light. It is proportional to the specific refractive index increment (dn/dc), concentration, and finally, the mass of the particles. Hence, the molar mass of a system can be determined from static light scattering, if dn/dc and the sample concentration are known.

7.3.1 Multi angle light scattering

If the particle that scatters light is much smaller than the wavelength of the incident light source (less than 10 nm for incident light with $\lambda = 660$ nm which is used in the miniDawn TREOS detector), the scattered light will be the same at any angle in the plane perpendicular to the polarisation of the incident beam. This is called isotropic scattering and is illustrated in fig. 7.14 a). However, if the particle is larger, it will scatter the light differently dependent on the angle in the plane perpendicular to the polarisation of the incident light. This type of scatterer is called an anisotropic scatterer and is illustrated in fig. 7.14 b) [155].

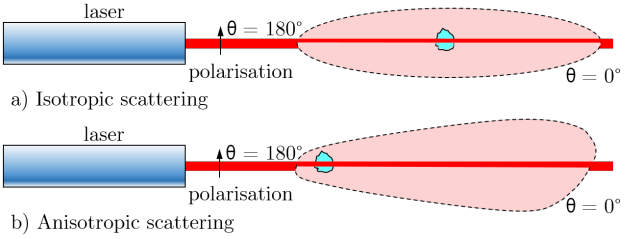


FIGURE 7.14: Isotropic scattering of a particle less than 10 nm, a). Anisotropic scattering of a particle larger than 10 nm, b). Figure adapted from Wyatt Technology Europe LSU course material [154] © by Wyatt Technology Corporation – used with permission.

Light scattering from larger particles thus happens with an angular dependency. The intensity of the scattered light will be larger for smaller angles (around 0°) than for larger angles (around 180°) as indicated in fig. 7.14 b). With multi angle light scattering (MALS), the angular dependency can be accounted for by measuring the scattered light at several angles. Hence, the size of the particle can be calculated. The size measured is the root mean square radius, or radius of gyration, R_g, which is the mass distribution around the center of mass point for the particle. Therefore, MALS gives both the molar mass and the R_g for the measured particle [156]. However, if the hydrodymanic radius, R_h is also known as well, the ratio between the two radii can give an estimate of the compactness and shape of the particle, in addition to the mass.

7.4 Dynamic light scattering

Dynamic light scattering (DLS) refers to the measurement of total scattering intensity over time. This gives an idea of the movement of particles relative to each other. Since particles in solution move via Brownian motion, the detected light intensity changes over time. This is due to the interference of the scattered light from multiple particles. Figure 7.15 illustrates how relative particle movement can cause intensity differences on the detector [157].

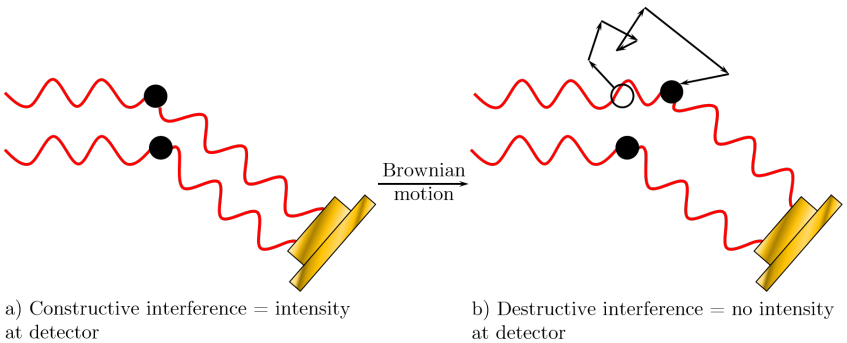
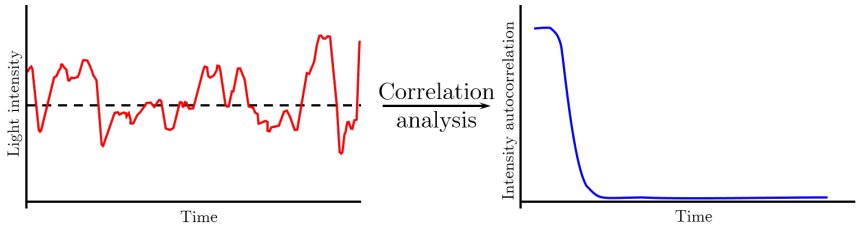


FIGURE 7.15: Two particles are positioned so that the light scattered from both of them interfere constructively, a). Due to Brownian motion, the particles have moved relative to each other and the scattered light interferes destructively, b). Figure adapted from Wyatt Technology Europe LSU course material [158] © by Wyatt Technology Corporation – used with permission.



a) Light intensity fluctuations b) Autocorrelation function

FIGURE 7.16: a) measured total intensity over time. b) autocorrelation function derived from the change in measured intensity. Figure adapted from Wyatt Technology Europe LSU course material [158] © by Wyatt Technology Corporation – used with permission.

The timescale with what these intensity fluctuations occur describes the diffusion constant of the particles. Hence, when measuring the intensity fluctuations (see fig. 7.16 a)), one can get the diffusion coefficient. This is done by composing the autocorrelation function (see fig. 7.16 b)), which describes the average overall change in intensity over time for a certain time interval [159].

The diffusion coefficient is proportional to the exponential decay rate of the autocorrelation function and is directly correlated to the hydrodynamic radius of the particle via the Stokes-Einstein relation, shown in eq. (7.1).

$$D_t = \frac{K_B T}{6\pi\eta R_h} \tag{7.1}$$

Where D_t is the diffusion coefficient, K_B is the Boltzmann constant, T is the absolute temperature in K, η is the viscosity of the solvent, and R_h is the hydrodynamic radius. Thus, R_h can be calculated from the diffusion coefficient if the solvent viscosity and temperature are known. However, it is important to note that the calculated R_h will be that of a hard sphere, with the same diffusion coefficient as that derived from the autocorrelation.

For batch measurements, the autocorrelation function can disclose whether or not the solution is mono- or multimodal: If multiple exponential time constants can be fitted to the autocorrelation function, a distribution of diffusion times is present. However, the size difference must be approximately 5 times or more[158].

A combination of MALS and DLS can give the mass, size, and shape of a particle. All components that characterise oligomers of acylated GLP-1 analogues.

Chapter 8

Peptide synthesis

Section 8.1 contains a description of the procedure used to synthesise the GLP-1 analogues GLP-1(14-34), C6Arev, C6A, C8A, C8LA, C10, C16, C10L, and C16L(14-34). Since the synthesis was executed by Assistant Professor Katrine Qvortrup and Master's student Maria Holm Rautenberg, the methodology section was created by Katrine Qvortrup. I had no involvement in preparing neither the peptides, nor the section describing the procedure. I only present what was given to me by Katrine Qvortrup to disclose the methodology used to synthesise the peptides I performed analyses on.

8.1 Methodology

8.1.1 Choosing the peptide sequence

The purpose was to study the peptides' secondary structure[1] and oligomerisation. A peptide sequence that both retains the secondary structure of GLP-1 and liraglutide, and is convenient to synthesise was chosen. The choise was based on a theoretical analysis of the secondary structures of GLP-1 and liraglutide.

By the use of data on the Protein Data Bank (PDB), the secondary structure of both GLP-1 and liraglutide were analysed. Both the original structure of GLP-1 and the analogue containing Arg34* instead of Lys34* were examined. The helix formation was found to take place between Phe12*/Thr13* and Leu32*/Arg34*, according to the PDB and PyMOL [136]. Based on this, the 14*-34* amino acid sequence was chosen for studies, to ensure as short a sequence as possible, making synthesis easier.

8.1.2 Synthetic strategy

Three different fatty acids, linked to an L-glutamyl spacer, were synthesised in two consecutive reaction steps. Initially, the carboxylic acids of the fatty acids were reacted with DCC/HOSu to activate them for the following amide coupling. The active esters were used directly (without purification) for amide coupling with L-Glu-O*t*Bu.

The peptide analogues were synthesized by the Fmoc/*t*Bu SPPS strategy, utilising a Biotage® Initiator+ AlstraTM Automated Microwave Peptide Synthesizer. As coupling reagent, N,N'-Diisopropylcarbodiimide (DIC) and additive Oxyma Pure were used. Chem-Matrix Amine resin functionalised with the Rink-Amide Linker was used, hereby obtaining the desired peptide amide analogues. To allow the desired fatty acid-functionalisation of Lysine (Lys) at position 26 in the GLP-1 analogues, the orthogonally protected Fmoc-Lys(Alloc)-OH amino acid was used in

[1]Secondary structure studies were conducted by Master's student Maria Holm Rautenberg, unrelated to the work described in this thesis.

SPPS. This group could be selectively removed and the Lys functionalised with desired fatty acids. The fatty acids containing an L-glutamyl spacer were synthesised in solution by amide coupling using the coupling reagent DCC and additive HOSu. Attachment of the fatty acids was afterwards performed using Fmoc/*t*Bu SPPS. Finally, the peptide analogues were obtained by acidic cleavage from the resin.

The crude peptide analogues were purified by preparative HPLC, and UPLC-MS and HPLC analysis were used for characterisation. For analysing the structure of the peptides, different NMR spectroscopy experiments were utilised. DQF-COSY, TOCSY and NOESY experiments were used for assigning the amino acid sequence and correlating the differences in secondary structure of each peptide analogue[2]. The peptide library is shown in figs. 8.1 and 8.2.

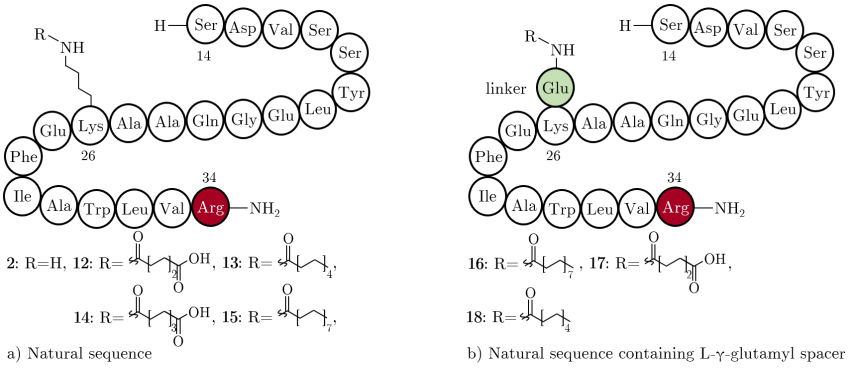


FIGURE 8.1: Library of peptides with natural GLP-1 sequence.

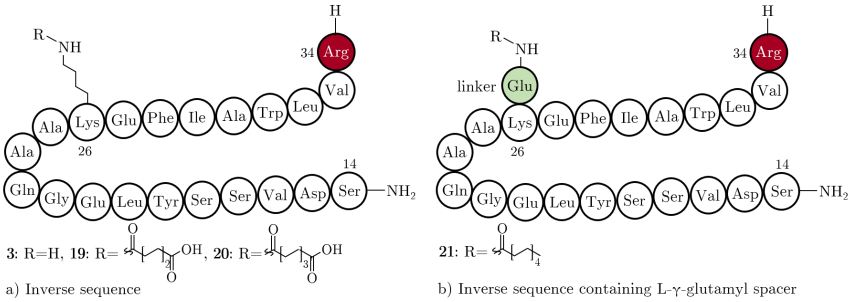


FIGURE 8.2: Library of peptides with inverse GLP-1 sequence.

UPLC-MS analysis was run on Waters AQUITY UPLC system equipped with PDA and either a SQD (for the fatty acids linked to the L-glutamyl spacer) or a SQD2 (for the peptides) electrospray MS detector. Column: Thermo accucore C18 2.6_m, 2.1_50 mm. Column temp: 50 °C. Flow rate: 0.6 mL/ min. Acid run: Solvent A1 – 0.1 % formic acid in water, Solvent B1 – 0.1 % formic acid in MeCN. Base run: Solvent A2 – 15 mM NH_4Ac in water, Solvent B2 – 15 mM NH_4Ac in MeCN/water 9:1. For determining the mass of the fatty acids linked to the L-glutamyl spacer, 5 min acid/base runs were used. For determining the mass of the peptides 5 min base runs were used. Gradient: 5 % B to 100 % B in 3 min, hold 0.1 min. Total run time – 5 min.

HPLC/ELSD analysis was run on an e2695 Waters Alliance system equipped with a 2998 PDA detector and an Agilent Technologies 1260 Infinity ELSD. Column: Symmetry C18 3.5_m, 4.6 mm_75 mm. Column temp: 20 °C. Flow rate: 1 mL/ min.

[2]These studies were performed by Master's student Maria Holm Rautenberg.

Solvent A2 – 15 mM NH_4Ac in water, Solvent B2 – 15 mM NH_4Ac in MeCN/water 9:1. Gradient: 5 % B in 0.5 min, gradient: 5 % B to 70 % B in 9.5 min, hold 2 min, gradient: 70 % B to 100 % B in 5 min, hold 3 min, run 20 min, recalibrating the column for 2 min. Total run time – 22 min.

Preparative HPLC purification of the peptides was performed on a Waters auto purification system consisting of a 2767 Sample Manager, 2545 Gradient Pump and 2998 PDA detector. Column: XBridge Peptide BEH C18 OBD Prep Column, 130 Å, 5_m, 19 mm_100 mm. Column temp: Ambient. Flow rate: 20 mL/ min. Solvent A2 – 15 mM NH_4Ac in water, Solvent B2 – 15 mM NH_4Ac in MeCN/water 9:1. Gradient: 5 % B in 3 min, gradient: 5 % B to 20 % B in 2 min, gradient: 20 % B to 50 % B in 2 min, hold 2 min, gradient: 50 % B to 70 % B in 3 min, gradient: 70 % B to 100 % B in 3 min, hold 0.5 min, run 15.5 min, recalibrating the column for 2.5 min. Total run time – 18 min.

Fmoc−SDVSSYLEGQAAK(Alloc)EFIAWLVR−NH_2
Calculated masses: m/z = 2674.32 $[C_{127}H_{181}N_{28}O_{36}]^+$; m/z = 2672.30 $[C_{127}H_{179}N_{28}O_{36}]^-$
Observed masses: m/z = 2674.7 $[M+H]^+$; m/z = 2673.0 $[M-H]^-$

Fmoc−RVLWAIFEK(Alloc)AAQGELYSSVDS−NH_2
Calculated masses: m/z = 2368.23 $[C_{108}H_{167}N_{28}O_{32}]^+$; m/z = 2366.22 $[C_{108}H_{165}N_{28}O_{32}]^-$
Observed masses: m/z = 2368.0 $[M+H]^+$; m/z = 2366.1 $[M-H]^-$

H−SDVSSYLEGQAAKEFIAWLVR−NH_2
Calculated masses: m/z = 2368.23 $[C_{108}H_{167}N_{28}O_{32}]^+$; m/z = 2366.22 $[C_{108}H_{165}N_{28}O_{32}]^-$
Observed masses: m/z = 2368.1 $[M+H]^+$; m/z = 2366.2 $[M-H]^-$

H−SDVSSYLEGQAAK[$CO(CH_2)_4COOH$]EFIAWLVR−NH_2
Calculated masses: m/z = 2496.28 $[C_{114}H_{175}N_{28}O_{35}]^+$; m/z = 2494.26 $[C_{114}H_{173}N_{28}O_{35}]^-$
Observed masses: m/z = 2496.4 $[M+H]^+$; m/z = 2493.7 $[M-H]^-$

H−SDVSSYLEGQAAK[$CO(CH_2)_8CH_3$]EFIAWLVR−NH_2
Calculated masses: m/z = 2522.37 $[C_{118}H_{185}N_{28}O_{33}]^+$; m/z = 2520.35 $[C_{118}H_{183}N_{28}O_{33}]^-$
Observed masses: m/z = 2522.4 $[M+H]^+$; m/z = 2520.5 $[M-H]^-$

H−SDVSSYLEGQAAK[$CO(CH_2)_6COOH$]EFIAWLVR−NH_2
Calculated masses: m/z = 2524.31 $[C_{116}H_{179}N_{28}O_{35}]^+$; m/z = 2522.29 $[C_{116}H_{177}N_{28}O_{35}]^-$
Observed masses: m/z = 2524.4 $[M+H]^+$; m/z = 2522.4 $[M-H]^-$

H−SDVSSYLEGQAAK[$CO(CH_2)_{14}CH_3$]EFIAWLVR−NH_2
Calculated masses: m/z = 2606.46 $[C_{124}H_{197}N_{28}O_{33}]^+$; m/z = 2604.45 $[C_{124}H_{195}N_{28}O_{33}]^-$
Observed masses: m/z = 2606.3 $[M+H]^+$; m/z = 2604.4 $[M-H]^-$

H−SDVSSYLEGQAAK[$CO(CH_2)_2C((S)-NHCO(CH_2)_{14}CH_3)COOH$]EFIAWLVR−$NH_2$
Calculated masses: m/z = 2735.50 $[C_{129}H_{204}N_{29}O_{36}]^+$; m/z = 2733.49 $[C_{129}H_{202}N_{29}O_{36}]^-$
Observed masses: m/z = 2735.5 $[M+H]^+$; m/z = 2733.7 $[M-H]^-$

H−SDVSSYLEGQAAK[$CO(CH_2)_2C((S)-NHCO(CH_2)_6COOH)COOH$]EFIAWLVR−$NH_2$
Calculated masses: m/z = 2653.35 $[C_{121}H_{186}N_{29}O_{38}]^+$; m/z = 2651.34 $[C_{121}H_{184}N_{29}O_{38}]^-$
Observed masses: m/z = 2653.2 $[M+H]^+$

$H-SDVSSYLEGQAAK[CO(CH_2)_2C((S)-NHCO(CH_2)_8CH_3)COOH]EFIAWLVR-NH_2$
Calculated masses: $m/z = 2651.41$ $[C_{123}H_{192}N_{29}O_{36}]^+$; $m/z = 2649.39$ $[C_{123}H_{190}N_{29}O_{36}]^-$
Observed masses: $m/z = 2651.4$ $[M+H]^+$; $m/z = 2650.0$ $[M-H]^-$

$H-RVLWAIFEKAAQGELYSSVDS-NH_2$
Calculated masses: $m/z = 2368.23$ $[C_{108}H_{167}N_{28}O_{32}]^+$; $m/z = 2366.22$ $[C_{108}H_{165}N_{28}O_{32}]^-$
Observed masses: $m/z = 2369.0$ $[M+H]^+$; $m/z = 2366.5$ $[M-H]^-$

$H-RVLWAIFEK[CO(CH_2)_4COOH]AAQGELYSSVDS-NH_2$
Calculated masses: $m/z = 2496.28$ $[C_{114}H_{175}N_{28}O_{35}]^+$; $m/z = 2494.26$ $[C_{114}H_{173}N_{28}O_{35}]^-$
Observed masses: $m/z = 2496.1$ $[M+H]^+$; $m/z = 2494.2$ $[M-H]^-$

$H-RVLWAIFEK[CO(CH_2)_6COOH]AAQGELYSSVDS-NH_2$
Calculated masses: $m/z = 2524.31$ $[C_{116}H_{179}N_{28}O_{35}]^+$; $m/z = 2522.29$ $[C_{116}H_{177}N_{28}O_{35}]^-$
Observed masses: $m/z = 2524.0$ $[M+H]^+$; $m/z = 2521.7$ $[M-H]^-$.

Chapter 9

Solution structures and HSA interactions

9.1 Equipment used

Data was collected at Wyatt Technology Europe's laboratory facilities in Dernbach, Germany.

9.1.1 AF4-MALS

Samples and buffers were pumped by an Agilent autosampler (Agilent, Santa Clara, California, USA) consisting of a quartinary pump in combination with a temperature controlled sampler tray and buffer degasser. Separation was obtained by use of AF4 controlled by the Eclipse AF4 (Wyatt Technology, Santa Barbara, California, USA) in combination with a short channel (SC) including a 490 µm W (wide) spacer and 1 kDa polyethersulfone (PES) membrane. The membranes used were produced at Superon. In-line detection involved UV, MALS and RI, one after the other. A VWD UV detector was used to measure UV absorbance. MALS was measured using a miniDAWN TREOS (Wyatt Technology, Santa Barbara, California, USA) with LS detectors at three different angles. RI signals were collected on an Optilab T-rEX (Wyatt Technology, Santa Barbara, California, USA).

A schematic representation of the used setup, including pump system, AF4 separation, and in-line detectors, is shown in fig. 9.1.

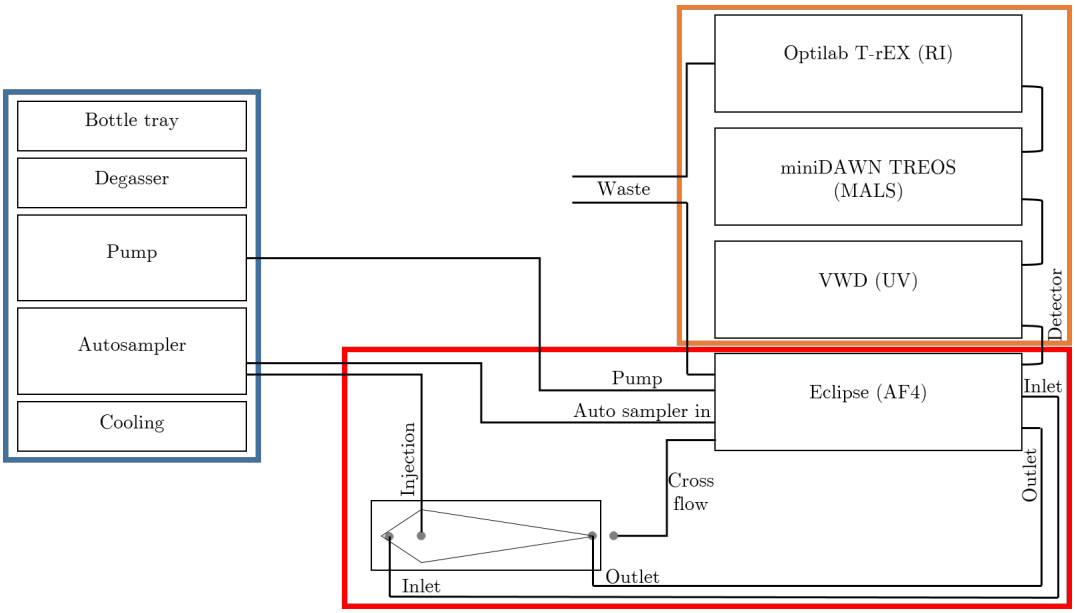


FIGURE 9.1: Schematic representation of the AF4-MALS setup. The blue box includes the pumping system. The red box includes the separation equipment, that is the Eclipse and the channel. The orange box includes the three detectors: UV, MALS, and RI.

9.1.2 DLS

Complimentary batch DLS experiments were performed on a DynaPro NanoStar (Wyatt Technology, Santa Barbara, California, USA) using single use plastic cuvettes for 5 µL sample volumes.

9.2 Buffer preparation

To test whether pH and ionic strength, in the form of varying NaCl concentrations, have an influence on possible oligomerisation and complex formation, six different buffer conditions were tested. Three with different pH values and no salt, and three with constant pH and different ionic strengths. Initially, the intention was to investigate the systems at pH 7, 8, and 9, but when solvating the peptides, it became evident that pH was too low. Therefore, pH values of 8.0, 9.6, and 10.3 were used. Salt concentrations were chosen to be 0 mM, 35 mM, 70 mM, and 140 mM. The pH value for all three salt containing buffers was 10.3, meaning, that the buffer with no salt and pH 10.3 corresponded to the 0 mM salt buffer.

pH was adjusted by adding 0.5 M NaOH. All buffers contained 8 mM phosphate and 5 % NaN_3. The azide was added solely to prevent bacterial growth in the equipment. The buffers were prepared using the following components listed in table 9.1:

TABLE 9.1: Buffer overview for preparation of 2.5 L with 8 mM phosphate and 5 % NaN_3 (0.5 g). IS = ionic strength. The ionic strengths are theoretically calculated values based on amounts of NaCl and phosphate salts.

Buffer ID	pH	[NaCl] (mM)	NaCl (g)	$NaH_2PO_4 \cdot 2\,H_2O$ (g)	Na_2HPO_4 (g)	IS (mM)
pH 8.0	8.0	0	0	0.425	2.452	21.8
pH 9.6	9.6	0	0	0.048	2.795	23.7
pH 10.3, 0 mM NaCl	10.3	0	0	3.105	0.014	8.1
35 mM NaCl	10.3	35	5.114	3.105	0.014	43.1
70 mM NaCl	10.3	70	10.227	3.105	0.014	78.1
140 mM NaCl	10.3	140	20.454	3.105	0.014	148.1

All buffers were filtered using a 0.2 µm filter.

9.3 Sample preparation

9.3.1 AF4-MALS: Peptides in varying pH

The synthesised peptides were dissolved in 8 mM phosphate and 5 % NaN_3 at pH 8.0. They were dissolved to give solutions of either 2 mg/mL (C6A, C10) or 3 mg/mL (GLP-1(7-36), GLP-1(14-34), C16L(14-34), C8AL, C16, C8A, C10L), except C6Arev, which was diluted to 0.7 mg/mL. Liraglutide[1] was used directly from formulation and diluted to 2 mg/mL. The 3 mg/mL samples were diluted to 2 mg/mL. Finally, 0.25 mL of the peptide samples were transferred to eppendorf tubes and diluted to 1 mg/mL (except for C6Arev, which was not diluted further and therefore remained at 0.7 mg/mL) to give 0.5 mL.

Due to higher peptide solubility at high pH, 2 µL of 0.5 M NaOH was added to each of the samples to increase pH, ensuring that the peptide sample was fully dissolved. As a result, the sample solutions had pH higher than 9, but not measured, since only 0.5 mL was prepared. However, when performing AF4 runs, the sample gets flushed with the carrier buffer, which had pH values of either 8.0, 9.6 or 10.3 as noted.

Hereafter, the samples were filtered with 0.22 µm filters and transferred into 0.5 mL auto sampler vials. This gave one set of samples that were used for the three runs with varying pH.

9.3.2 AF4-MALS: Peptides in varying salt concentrations

With regards to the different salt buffer runs, the initial sample solutions of either 2 mg/mL and 3 mg/mL were diluted to 1 mg/mL with 8 mM phosphate buffer at pH 10.3 with NaCl concentrations of 70 mM, 140 mM, and 280 mM. This gave three sets of sample solutions: One with 1 mg/mL peptide and 35 mM salt, one with 1 mg/mL peptide and 70 mM salt, and one with 1 mg/mL peptide and 140 mM salt. In all three sets, the sample containing C6Arev, only had a peptide concentration of 0.35 mg/mL. Like before, 2 µL of 0.5 M NaOH was added, followed by filtration with 0.22 µm filters into 0.5 mL autosampler vials.

[1] Denoted as C16L(7-37) in this part of the thesis.

9.3.3 AF4-MALS: Peptides+HSA in varying salt concentrations

Albumedix produces HSA under the salesname Recombumin® Alpha, but it contains octanoate, which will bind in the FA binding pockets on HSA. Therefore, Recombumin® Alpha was dialysed to remove excess octanoate. The buffer used contained 8 mM phosphate at pH 8.0 without octanoate.

Dialysis was conducted by injecting 0.5 mL Recombumin® Alpha in a 0.1 mL to 0.5 mL Slide-A-Lyzer® dialyses cassette with a molecular weight cut-off of 20 kDa. The cassette was then placed in 1 L 8 mM sodium phosphate buffer at pH 8.0 for 2 h. The HSA containing cassette was then submerged in a fresh batch of buffer (1 L) and left for another 2 h. The buffer was changed again and left over night. The result was around 0.3 mL HSA in 8 mM phosphate buffer at pH 8.0. The HSA concentration had fallen from 200 mg/mL to 58.67 mg/mL.

After dialyses, the resulting HSA solution was diluted to 2 mg/mL with the 8 mM phosphate buffer at pH 10.3 for AF4-MALS measurements.

Due to the limited amount of sample at this stage, (C10L had run completely out and was thus not measured in mixture with HSA) only one set of samples was prepared. This was done by mixing 250 µL of each sample at 2 mg/mL in the 8 mM phosphate buffer at pH 8.0 with 250 µL of the 8 mM phosphate buffer at pH 10.3. Due to time restrictions, AF4 runs were only performed under varying salt conditions (35 mM, 70 mM, and 140 mM NaCl) at pH 10.3. Since the samples and HSA were diluted with buffer not containing salt, the actual sample solutions did not contain the mentioned concentrations. However, in AF4 the samples get flushed with the carrier buffer, which contained salt in the respective concentrations.

9.3.4 DLS: Peptides in varying salt concentrations

Four sets of peptide samples were prepared, all with high pH and different salt concentrations of 0 mM, 35 mM, 70 mM, and 140 mM.

The 0 mM set was obtained by diluting the 2 mg/mL and 3 mg/mL samples to 1 mg/mL with 8 mM phosphate buffer with pH 8.0. A total volume of 30 µL was prepared. Due to higher peptide solubility at high pH, 2 µL of 0.5 M NaOH was added to each of the samples to increase pH and ensure fully dissolved samples. As a result, the sample solutions had pH higher than 9, but not measured, since only 30 µL was prepared.

The three remaining sets were prepared like described above but diluted with 8 mM phosphate buffer at pH 10.3 and salt concentrations of 70 mM, 140 mM, and 280 mM to give the final peptide concentrations of 1 mg/mL and salt concentrations of 35 mM, 70 mM, and 140 mM, respectively. A total volume of 30 µL was prepared. Due to higher peptide solubility at high pH, 2 µL of 0.5 M NaOH was added to each of the samples to increase pH and ensure fully dissolved samples. As a result, the sample solutions had pH higher than 10.3, but not measured, since only 30 µL was prepared.

In all four sets, the sample containing C6Arev, only had a peptide concentration of 0.35 mg/mL.

Hereafter, the samples were centrifuged for 5 min at 13 000 rpm to collect any large particles in the bottom on the eppendorf tubes. Then, 5 µL were pippetted from the top of the eppendorf, into DLS single use plastic cuvettes. To prevent evaporation, the chamber surrounding the sample chamber was filled with milliQ water, and a cap was used to close the cuvette. Thus, all samples were measured in the four different buffer solutions.

9.4 Software and sample constants

9.4.1 AF4-MALS

The Wyatt ASTRA software was used to collect and analyse the measured MALS, UV, and RI data, while Chromeleon was used to control the AF4 and pump equipment.

UV absorbance was measured at 280 nm. UV extinction coefficients (ϵ) were calculated for the peptide sequences (excluding FA chains) and HSA using ExPASy's ProtParam tool [160]. The resulting extinction coefficients were divided with the molar mass of the individual samples and used as the extinction coefficient given in Astra, see table 9.2.

TABLE 9.2: Molar mass and extinction coefficients for the peptide samples and HSA.

Peptide	Mw (Da)	ϵ ($M^{-1}\,cm^{-1}$)	Astra ϵ ($mL\,mg^{-1}\,cm^{-1}$)
GLP-1(14-34)	2369.7	6990	2.95
GLP-1(7-36)	3298.7	6990	2.11
C6Arev	2497.8	6990	2.80
C6A	2497.8	6990	2.80
C8A	2525.8	6990	2.77
C10	2524.0	6990	2.77
C8LA	2655.0	6990	2.63
C10L	2653.0	6990	2.63
C16	2608.1	6990	2.68
C16L(14-34)	2737.2	6990	2.55
C16L(7-37)	3751.2	6990	1.86
HSA	66437	34445	0.52

The refractive index increment (dn/dc) was set to 0.185 mL/g for all peptide samples as well as for HSA, and all measurements were performed at 25 °C.

9.4.2 DLS

The Wyatt DYNAMICS software was used for DLS data collection and analyses. In regards to the DLS measurements, the viscosity and refractive indices (see table 9.3) were calculated using Zetasizer v.7.1 (Malvern Instruments, Worcestershire, England). These values were calculated for the four buffers at 0 mM, 35 mM, 70 mM, and 140 mM NaCl. Values were based on the amount of added salts mentioned in table 9.1. It is important to note that the pH of the sample solutions was higher than 9.0 for the 0 mM NaCl samples, and higher than 10.3 for the 35 mM, 70 mM, and 140 mM NaCl samples. This is due to the added NaOH.

TABLE 9.3: Viscosity and refractive index values for the buffers used for DLS measurements.

Buffer (mM NaCl)	Viscosity (cP)	Refractive index
0	0.8916	1.3303
35	0.8961	1.3306
70	0.8993	1.331
140	0.9057	1.3317

9.5 AF4 method: Peptide separation

This section contains a short description of the setup procedure and some hints for problem solving, based on experiences when using the AF4 equipment in combination with low cut-off membranes.

For the peptide measurements, the AF4 separation was conducted using the method described in table 9.4, with the flow profile shown in fig. 9.2.

TABLE 9.4: Eclipse timetable for peptide sample separation.

Start time (min)	End time (min)	Duration (min)	Mode	Vx start (mL/ min)	Vx end (mL/min)
0.00	1.00	1.00	Elution	1.50	1.50
1.00	2.00	1.00	Focus		
2.00	3.00	1.00	Focus + Inject		
3.00	5.00	2.00	Focus		
5.00	35.00	30.00	Elution	1.50	1.50
35.00	40.00	5.00	Elution + Inject	0.00	0.00
40.00	41.00	1.00	Elution	1.50	1.50

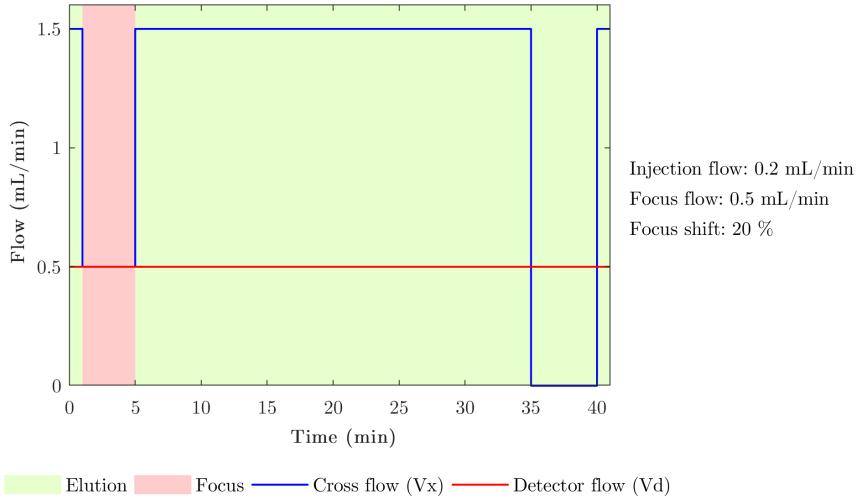


FIGURE 9.2: Flow diagram used to separate peptide samples.

All peptide experiments were conducted using the above Eclipse method, see figs. 7.2 and 7.3 for illustration of flows.

The last minute of the method is only present to prevent an error that occurs when using Chromeleon to control the Eclipse - here, Vx has to be the same in the beginning and in the end. If Voyager is used to control the Eclipse, this is not necessary.

As mentioned in section 9.1.1, the 490 µm W spacer was chosen. This was specifically done, because a thicker spacer gives a larger separation volume. The result of a larger separation volume is better resolution. However, a larger volume also means longer retention times and lower peaks. Although, when the thicker spacer is used in combination with the short channel, the retention time is durable. The low cut-off membrane should ensure that the peptide sample (Mw around 2.4 kDa to 3.7 kDa) is not lost, but the flux through the membrane will in turn be lower than for higher cut-off membranes. Furthermore, the cross flow in the method should be high enough to push the molecules against the membrane and obtain retention.

To start with, everything should be equilibrated and tweaked to ensure a correct and usable setup. This is done by running the system in elution mode with the elution buffer to be used for the measurements. When doing this, it is vital that the channel is checked for leakages. If a small leak (a few drops) is present it might be possible to fix it by tightening the screws to 9 N m. If, on the other hand, it is a more persistent leak, it will be necessary to open up the channel and change the membrane. If the channel is leaking from the beginning it might be because the membrane is bad, and then it is not worth trying to fix it. The membrane should be changed to another batch, since some batches of the 1 kDa PES membranes had some issues with the flux. This would cause leakages in the channel and problems obtaining the right cross flow pressure (Px). So if a channel is assembled with a 1 kDa PES membrane, and it is leaking or Px is very low in elution mode (less than 1 bar), it is most likely faulty, and there is no point in trying to make it work. Hence, change the membrane and try again. In general, leakages proved to be a bit of a problem when using the low cut-off membranes, therefore, it is recommendable to remember to check for leakages through the rest of the setup process and the first run. If there are no leakages until this stage, no leakages should arise during measurements.

Recommendations for the SC channel with a W 350 µm spacer and regenerated cellulose with a cut-off of 10 kDa is to use a Vd of 1 mL/ min and a Vx of 3 mL/ min. Then a Pc of 13 bar and a Px of 8 bar should be obtained. The pressures can be regulated somewhat by modifications to the setup; Pc can be regulated by modifying tubing in the setup from the Eclipse to the UV. To raise the Pc, use a thinner tubing (dark red) or longer tubing. To lower the Pc, use less tubing or thicker tubing (beige). Px can be regulated by changing Vx. The higher the Vx, the lower the Px and vice versa. Also, Vd can be lowered to get a higher Px. However, with the low cut-off membrane, it was not possible to obtain the recommended pressures using Vd of 1 mL/ min and Vx of 3 mL/ min. Therefore, Vx was lowered to around 1.5 mL/min, as shown in fig. 9.2.

After this, one must make a test run with a standard, e.g. bovine serum albumin (BSA). However, BSA is a large protein of 66.7 kDa, so it is a bit large to use as a standard for peptides. A smaller molecule that is stable, like BSA, could be used. Liraglutide has proven to be quite stable, however, it oligomerises into something between hexa- and heptamers, so it is larger than the monomeric size of liraglutide (about 22.5 kDa to 26.3 kDa, for hexa- and heptamer respectively).

Furthermore, it is important to make sure that Px does not go below 1 bar, since the CoriFlow in the Eclipse will not be able to correctly control the flowrates. If Px is too low the setup stage should be redone with a lower Vd or lower Vx.

Methodwise, it is a good idea to experiment with including a gradient in Vx that goes to zero to ensure that no sample is held back against the membrane and only elutes when the cross flow is removed.

Note: The Eclipse gives an over pressure alarm if Pc reaches 30 bar. This can sometimes happen if the tubing from the Eclipse to the UV detector is too long or if there is an issue with the membrane. I have also experienced a software malfunction where Voyager would *think* the pressure was rising and in the end set of the alarm. If Pc is rising rapidly, open the purge on the autosampler to prevent the flow from going through the channel/Eclipse. If the issue is consistent and not related to the membrane (still occurs after a membrane change), it could be the software, and a restart should be performed, otherwise, contact developers.

9.6 AF4 method: Peptide+HSA separation

For the peptide+HSA measurements, the AF4 separation was conducted using the method described in table 9.5, with the flow profile shown in fig. 9.3.

TABLE 9.5: Eclipse timetable for peptide+HSA sample separation.

Start time (min)	End time (min)	Duration (min)	Mode	Vx start (mL/ min)	Vx end (mL/min)
0.00	2.00	2.00	Elution	2.00	2.00
2.00	3.00	1.00	Focus		
3.00	4.00	1.00	Focus + Inject		
4.00	5.00	1.00	Focus		
5.00	15.00	10.00	Elution	2.00	2.00
15.00	17.00	2.00	Elution	2.00	1.10
17.00	20.00	3.00	Elution	1.10	0.60
20.00	23.00	3.00	Elution	0.60	0.33
23.00	26.00	3.00	Elution	0.33	0.18
26.00	29.00	3.00	Elution	0.18	0.10
29.00	32.00	3.00	Elution	0.10	0.05
32.00	35.00	3.00	Elution	0.05	0.03
35.00	38.00	3.00	Elution	0.03	0.01
38.00	41.00	3.00	Elution	0.01	0.01
41.00	44.00	3.00	Elution	0.01	0.00
44.00	49.00	5.00	Elution + Inject	0.00	0.00
49.00	50.00	1.00	Elution	2.00	2.00

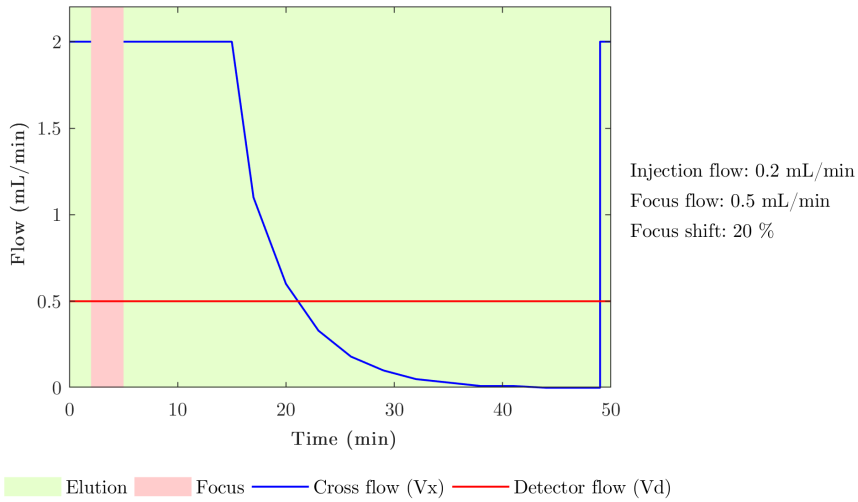


FIGURE 9.3: Flow diagram used to separate peptide+HSA mixtures.

Vx gradients were tested, and the exponential gradient shown in table 9.5 and fig. 9.3 proved to give narrower HSA peaks.

9.7 DLS method

All samples were measured three times giving 150 acquisitions of 5 s duration. Measurements were carried out at 25 °C.

Measurements were performed on the peptide samples (except C10L, which ran out) in 0 mM, 35 mM, 70 mM, and 140 mM NaCl and pH above 9.

9.8 Results and discussion

Only one run was performed on each sample, so all results presented in the following is based on only one measurement.

9.8.1 AF4-MALS: Peptides

Seven conditions were tested for all 11 peptide samples. An overview of the data quality is given in table 9.6.

TABLE 9.6: Overview of data usability in each AF4 sequence run. Each sequence includes measurements for each of the 11 samples (first column) in different buffers (top row). Red indicates low LS signal to noise ratio, green a high signal to noise ratio. The lighter colours for one condition is due to a highly noisy signal for all samples, most likely due to a batch of buffer with impurities.

	pH/[NaCl] (mM)						
Sample	8/0	9.6/0	9.6/0	10.3/0	10.3/35	10.3/70	10.3/140
GLP-1(14-34)							
GLP-1(7-36)							
C6Arev							
C6A							
C8A							
C8LA							
C10							
C16							
C10L							
C16L(14-34)							
C16L(7-37)							
Injected vol. (µL)	20	20	80	80	80	80	80

Table 9.6 shows if data can be processed for analysis. In general, data in red means that the LS signal to noise ratio is around one and data in green, in general, means that the signal to noise ratio is higher than three. The lighter colours for the sequence with pH 9.6 buffer and 20 µL injection means that the signals were very noisy with spiky peaks. This could be due to a mixture of impurities in the buffer solution for this run and the lower injection volume, therefore, this sequence run will not be considered, and is left out of the following analyses.

Overall, table 9.6 shows that the peptides without acylation (GLP-1(14-34) and GLP-1(7-36)) and those with a short FA chain of six C-atoms, do not generate usable LS signals. For those with FA chains of eight C-atoms, the signals are generally usable, but for pH 10.3, 140 mM, they are not. All measurements marked with red in table 9.6 are, therefore, not included in the analyses relying on the LS signal. For analyses relying on the UV signal, they are, however, included.

In general, there seems to be a trend with regards to the quality of the LS signal: The more C-atoms in the FA chain, the better the LS signal. This is seen as a higher signal to noise ratio (S/N) for those peptides with longer FA chains. The LS S/N is shown in figure fig. 9.4 for the three runs using salt containing buffers, and in fig. 9.5 for the three runs using pH varying buffers.

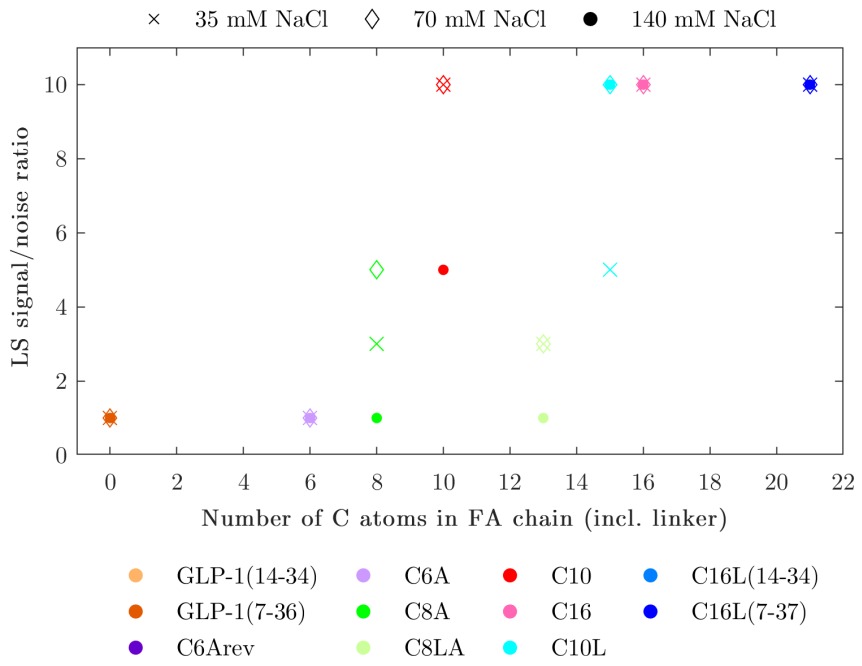


FIGURE 9.4: LS signal to noise ratio (S/N) plotted against the number of C-atoms in the FA chain, including the linker, for all peptides in the three buffers with varying salt concentrations and pH 10.3. S/N is given as numbers representing intervals: 1 = S/N: < 2, 3 = S/N: $[2-4]$, 5 = S/N: $[5-7]$, 10 = S/N: > 8.

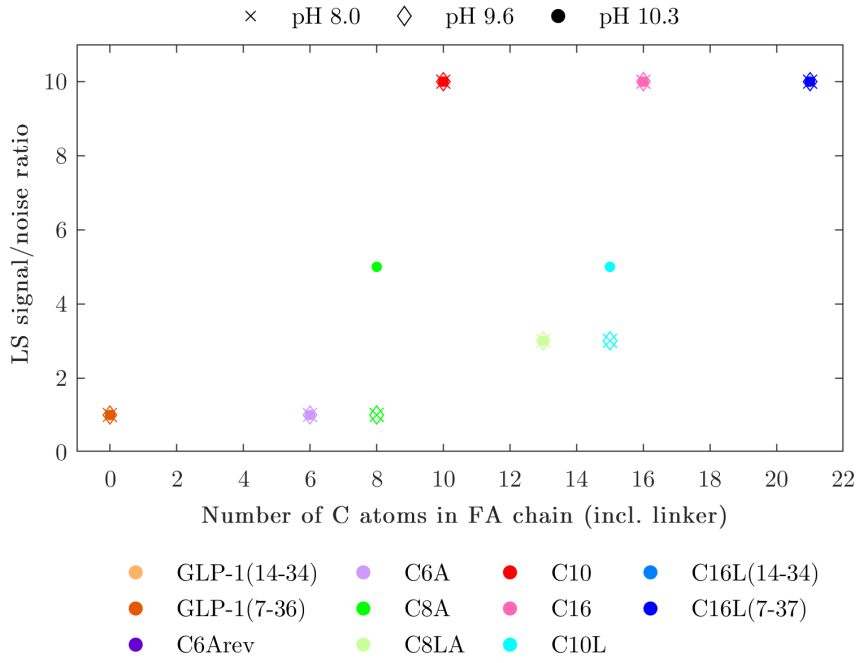


FIGURE 9.5: LS signal to noise ratio (S/N) plotted against the number of C-atoms in the FA chain, including the linker, for all peptides in the three buffers with varying pH value. S/N is given as numbers representing intervals: 1 = S/N: < 2, 3 = S/N: $[2-4]$, 5 = S/N: $[5-7]$, 10 = S/N: > 8.

In figs. 9.4 and 9.5, C16L(14-34) and C16L(7-37) (dark blues) lies on top of each

other. So does GLP-1(14-34) and GLP-(7-36) (oranges), and C6Arev and C6A (purples).

The increase in LS S/N corresponds well with the fact that the standard deviation on the total molecular weight (based on the LS signal) decreases with the size of the FA chain. LS standard deviations are plotted in figs. 9.6 and 9.7 for the salt buffers and pH buffers respectively.

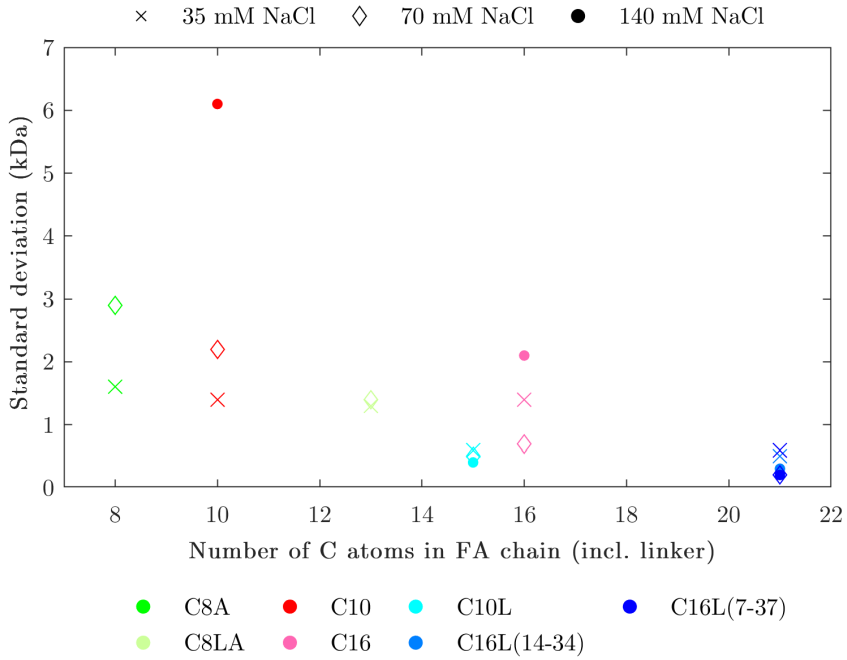


FIGURE 9.6: Standard deviation on the total molecular weight from the LS signal plotted against the number of C-atoms in the FA chain, including linker. Shown for all the peptides in the three buffers with varying salt concentration.

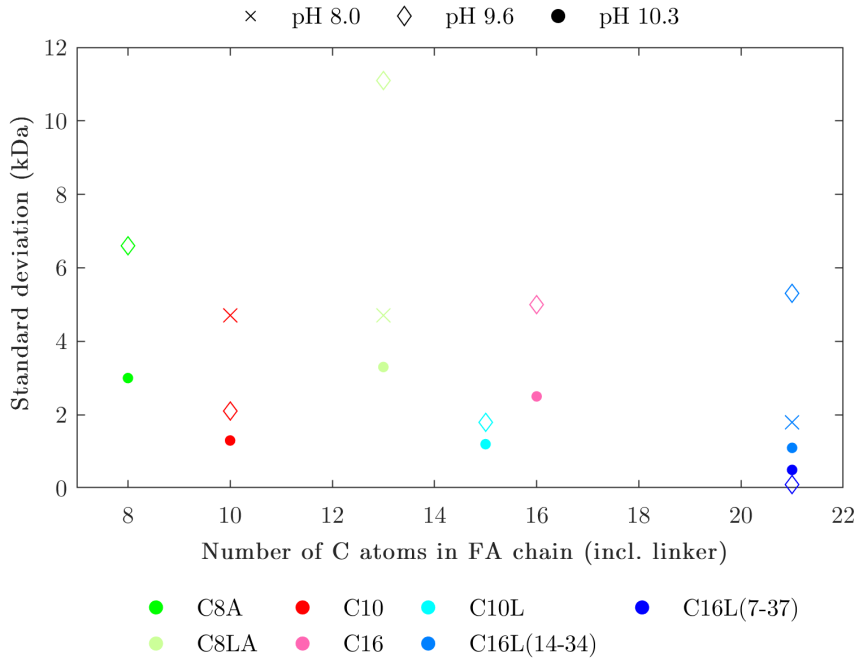


FIGURE 9.7: Standard deviation on the total molecular weight from the LS signal plotted against the number of C-atoms in the FA chain, including linker. Shown for all the peptides in the three buffers with varying pH value.

Combining these results, it is clear that the peptides with long FA chains (C10, C10L, C16, C16L(14-34), and C16L(7-37)) give rise to less noisy LS signals and lower standard deviations, indicating higher stability of the species with longer FA chains. Furthermore, the oligomeric size (total Mw divided by the molar mass of the peptide monomer) increases with the size of the FA chain. This is illustrated in figs. 9.8 and 9.9.

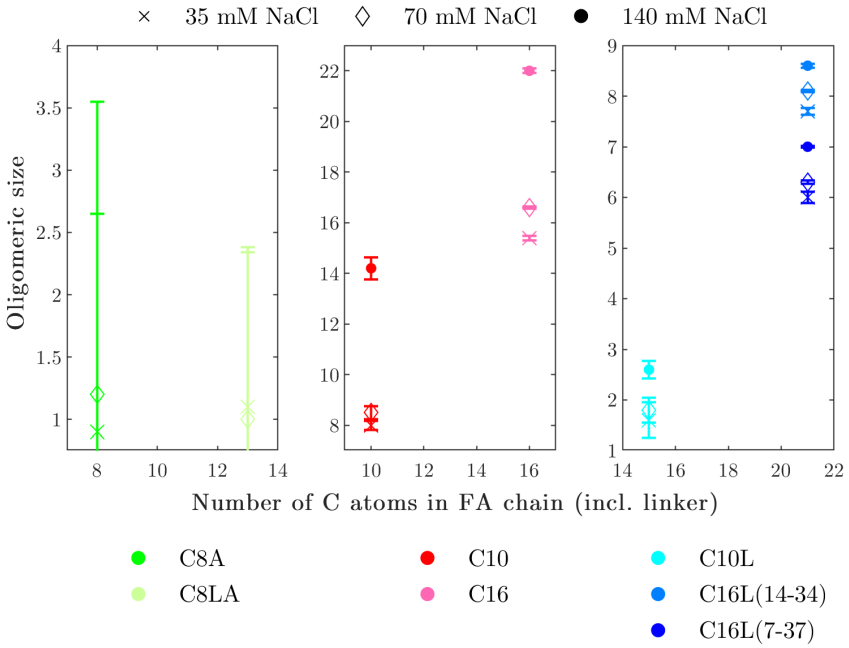


FIGURE 9.8: Oligomeric size plotted against the number of C-atoms in the FA chain, including linker, for all peptides with usable LS signal in the three buffers with varying salt concentration.

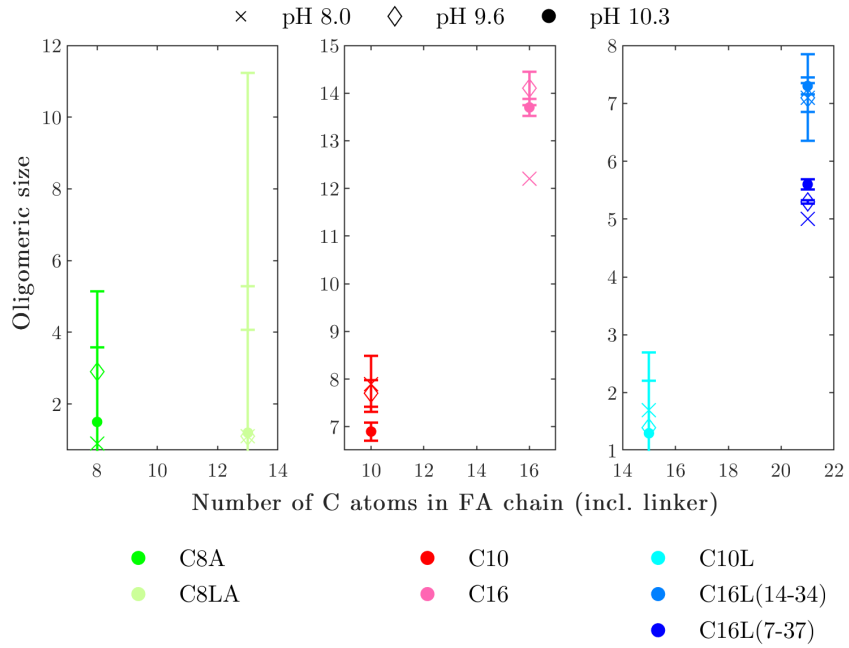


FIGURE 9.9: Oligomeric size plotted against the number of C-atoms in the FA chain for all peptides with usable LS signal in the three buffers with varying pH value.

From figs. 9.8 and 9.9, it is seen that the oligomeric size of C8A (the smallest chain) is between mono- to dimer, whereas C16L(14-34) and C16L(7-37) (the longest chains) are between hexa- and heptamers for C16L(7-37) and from octa- to nonamers for C16L(14-34). It is also worth noting that in the presence of a linker, the oligomeric size is decreased, as well as clustered closer together in the different conditions. For

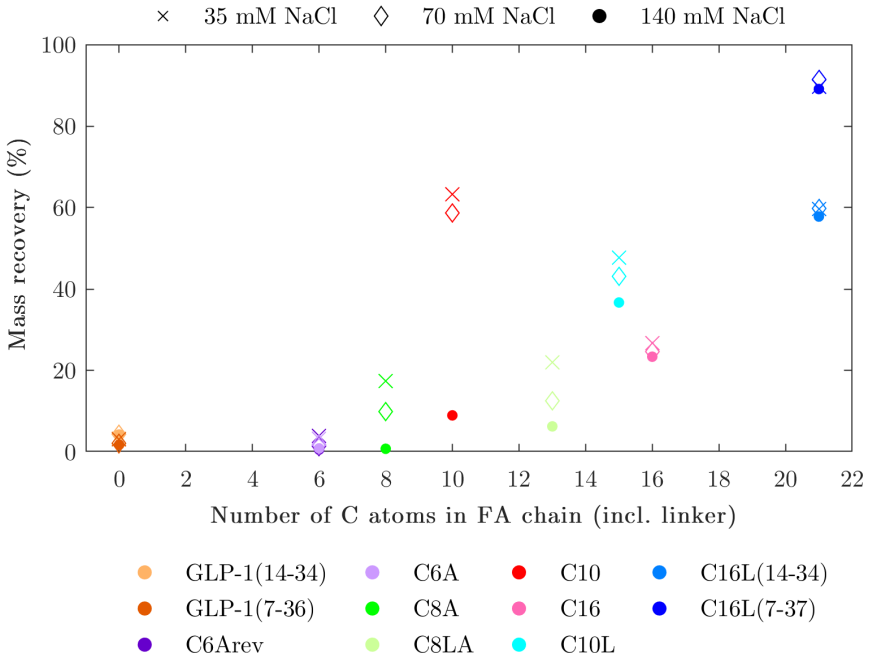


FIGURE 9.10: Mass recovery plotted against the number of C-atoms in the FA chain, including linker, for all peptides in the three buffers with varying salt concentration.

instance is C10 an octamer in the two lower salt concentrations, but a 14-mer in the high salt concentration. For C10L, the size corresponds to about a dimer in the lower salt concentrations, and a trimer in the high salt concentration. This suggests that a linker increases the stability of the oligomer, showing as a more uniformly oligomer size in the different conditions. The linker also, interestingly enough, causes the oligomers to be smaller. Additionally, fig. 9.8 also reveals that the oligomeric size of the different peptides tend to increase with increasing salt concentration. For C8A and C8LA, the errorbar is too large to conclude anything, but for the other peptides, there is an increase in size. Contrary, this trend is not present in the tested pH values (see fig. 9.9), suggesting that the pH range covered does not affect the oligomeric size of the peptides. As mentioned before, the size distribution is larger for C10 and C16 (reds) and smaller for C10L and the two C16L's (blues). This point to ionic strength induced interactions of the FA chains in the oligomers.

In addition to the oligomeric size increasing with the size of the FA chain, it should also be mentioned that the mass recovery also increases with the size of the FA chain - see figs. 9.10 and 9.11.

Other than give an idea on the stability, the mass recovery can also tell us what the likelihood of several species in the samples is. For instance, for C16L(7-37), where the mass recovery is around 90 %, it is very likely that there is only one specie present. Oppositely, for C16L(14-34), where the recovery is around 60 %, there could exist one or more other species in the one seen. Most likely, such specie(s) will be smaller and was lost through the membrane. This is based on the fact that GLP-1(7-36), GLP-1(14-34), C6A, and C6Arev does not give any LS peaks, indicating that nothing is retained by the membrane. So, for higher recoveries (C16L(7-37)), we can say that the observed specie is the only on present, but for those with lower mass recoveries (C8A, C8LA, C10, C10L, C16, and C16L(14-34)), there might be smaller

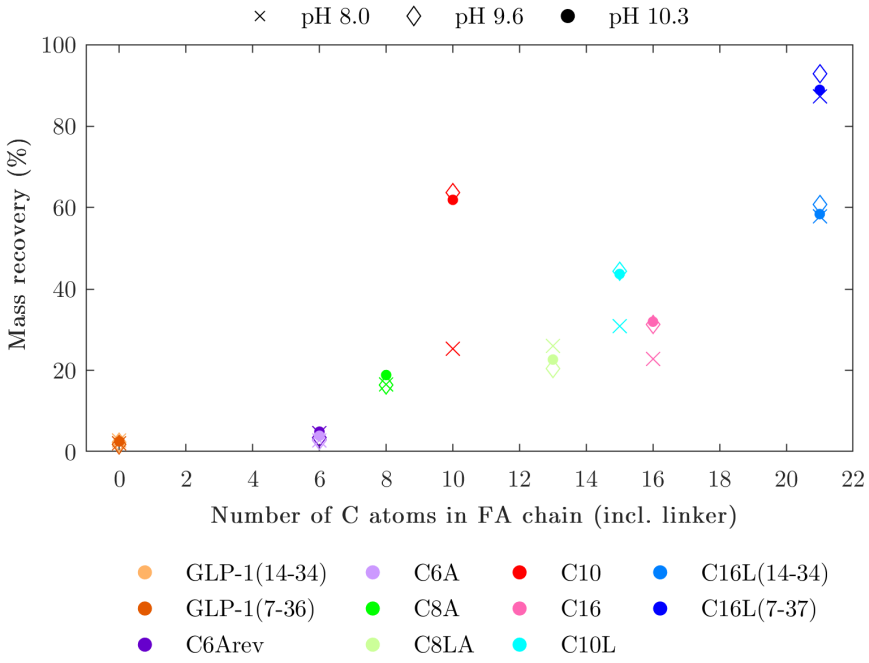


FIGURE 9.11: Mass recovery plotted against the number of C-atoms in the FA chain, including linker, for all peptides in the three buffers with varying pH value.

species presents in the samples. Of course, there is a possibility of larger species being present. The case could be that the remaining mass (the mass not eluting from the channel during the AF4 run) could be present as aggregates or other higher order oligomers, and just not eluting because the cross flow was pushing them against the membrane.

In fig. 9.10 it is seen that, increased salt concentration causes the mass recovery to decrease for GLP-1(7-36) (dark orange), C6A, C6Arev (purples), C8A, C8LA (greens), C10, C16 (reds), and C10L (blue). For GLP-1(14-34), C16L(7-37), and C16L(14-34) the mass recovery does depend on the salt concentration. One could speculate that this decrease in mass recovery with increasing salt, could be because the higher salt concentrations causes an increase in oligomeric size of the individual peptides. The reason being, that larger molecules, in theory, should be retained longer in the AF4. This prolonged retention time will increase the time of 'stress' on the peptide oligomers, and are they not stable, this increased time with stress might cause the oligomers to fall apart to monomers. The monomers could then go through the membrane and the results show as lower mass recovery. However, a plot of the mass recovery as a function of the peak point time (see figs. 9.12 and 9.13) tells us that this is not the case. Had it been the case, the retention time should have been higher for higher salt concentration.

But what we can see, is a semi-grouping of the three 'sets' of peptide structures. Peptide structures that resemble each other (greens: short FAs, reds: medium to long FA and no linker, blues: medium to long FAs with linker) seem to be retained for somewhat equal time in the AF4 channel. This indicates, that not only the length of the FA chain, but also the presence of a the L-γ-glutamyl linker is of importance. Confirming the results showing that a linker gives rise to smaller oligomers.

Furthermore, the salt buffers causes the samples to elute later than for the pH buffers with no salt. Again, this indicates that higher ionic strength stabilises the

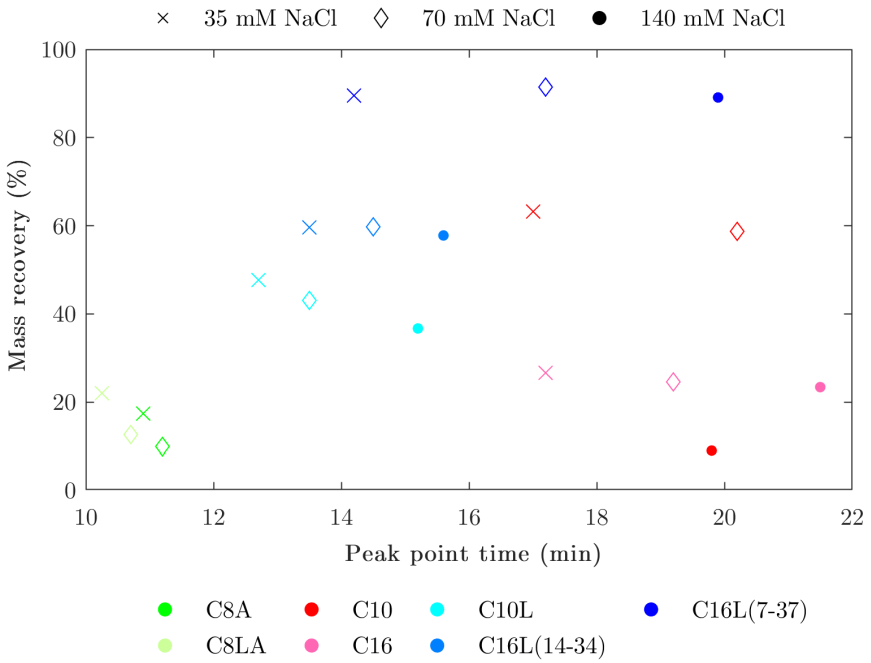


FIGURE 9.12: Mass recovery vs peak point time for all peptides in the three buffers with varying salt concentration.

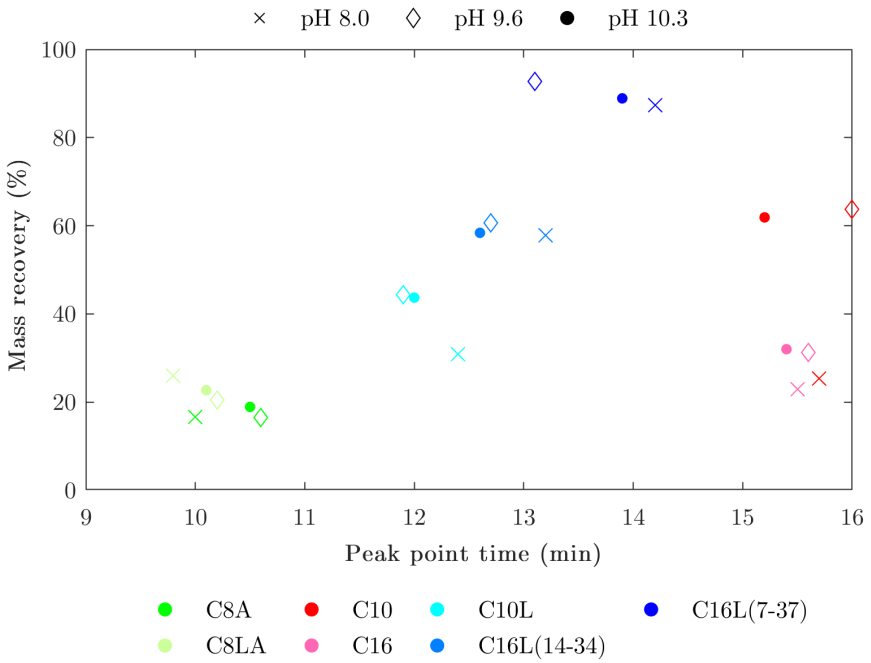


FIGURE 9.13: Mass recovery vs peak point time for all peptides in the three buffers with varying pH.

oligomers and induce interactions of the FA chains.

What is evident from figs. 9.9 and 9.13 is the fact that pH between 8.0 and 10.3 seems to have little to no impact on the structure of the oligomers. Whereas salt concentration increases the oligomeric size with at least one peptide molecule.

Returning to the quality of the LS signal in relation to the length of the FA chain. In general, the LS signal is also better for the peptides with longer FA chains, this corresponds well with previous findings. See figs. 9.14 and 9.15 for actual LS signals.

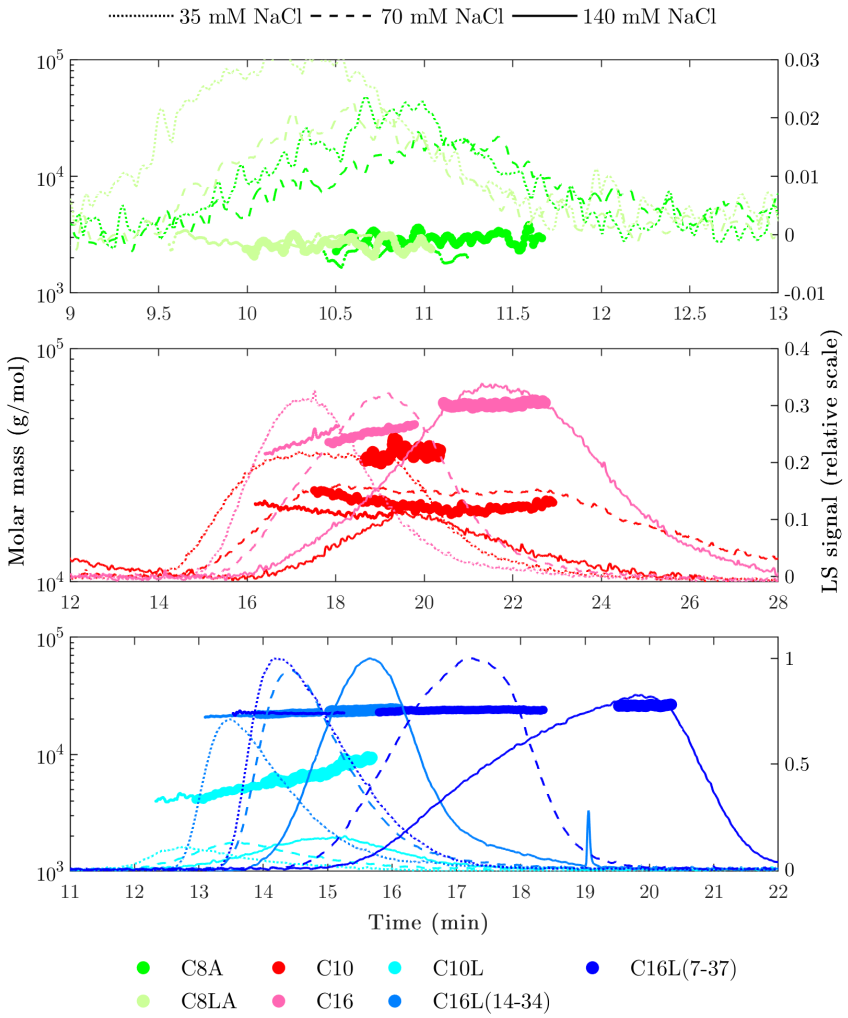


FIGURE 9.14: LS signal in the three buffers with varying salt concentration. The molecular mass, calculated from the LS signal across the peak, is given on the left y-axis and is shown as dots (dot size increases with salt concentration). On the right y-axis, the LS signal shown is scaled relatively to the UV signal, which is not shown.

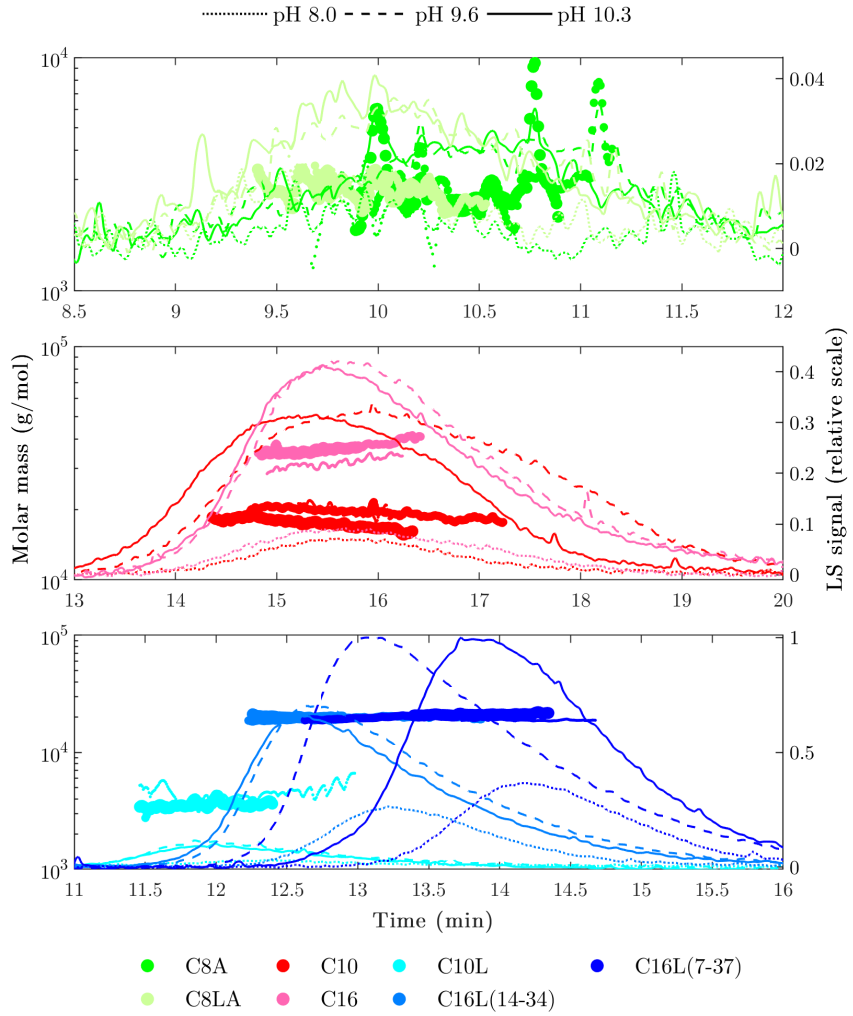


FIGURE 9.15: LS signal in the three buffers with varying pH value. The molecular mass, calculated from the LS signal across the peak, is given on the left y-axis and is shown as dots (dot size increases with salt concentration). On the right y-axis, the LS signal shown is scaled relatively to the UV signal, which is not shown.

Figures 9.14 and 9.15 clearly show that the LS signal for C8A and C8LA is rather noisy, which results in the higher standard deviations seen in figs. 9.6 and 9.7. This is a result of the lower oligomeric state and hence a lower detectable amount held back by the membrane. On this note, figs. 9.14 and 9.15 also show that the LS signal gets less tethered and noisy with the longer FA chains (reds), and even cleaner with longer FA chains including linkers (blues). Furthermore,fig. 9.14 shows an increase in retention time with increasing salt concentration for C16, C10L, C16L(14-34), and C16L(7-37). This again corresponds well with the increase in Mw, and hence oligomeric size, with increased salt concentrations, meaning that a higher ionic strength facilitates self-interactions.

Contrary to this, fig. 9.15 shows that for C10, C16, and C10L, the retention times are rather similar in each condition, with masses varying with less than a single peptide molecule for C10 and C10L, and with an increase of about two peptide molecules from pH 8.0 to higher pH values for C16 (see fig. 9.9 for oligomeric size). For C16L(14-34) and C16L(7-37), there is a noticeable shift in retention time, but less than a peptide molecule in mass (see fig. 9.9 for oligomeric size).

Overall, all runs with C16L(7-37), liraglutide, seems to be very similar under the investigated conditions. Furthermore, there is a high mass recovery and a controlled oligomerisation corresponding to one specie consisting of six to seven monomers. It could, therefore, be taken into consideration to use liraglutide as a standard for acylated peptide runs, since BSA is much larger (66.7 kDa).

It should be mentioned that, during these runs, there was no cross flow gradient. This could mean that larger species are not visible. Such species might elute when the cross flow is set to 0 in the cleaning process.

9.8.2 DLS: Peptides

DLS measurements were performed to give information on the hydrodynamic radii of the different peptide species in solution.

Experiments were run in four different conditions. All with 8 mM phosphate, 5 % NaN_3, and pH higher than 9. They had varying salt concentrations of 0 mM, 35 mM, 70 mM, and 140 mM. The sample peptide concentration was 1 mg/mL (except C6Arev, which was 0.7 mg/mL in the run with no salt and 0.35 mg/mL in the three runs with salt[2]).

A summary of the data is shown in table 9.7. In general, all samples show large aggregates. The samples were not filtered, only centrifuged, because preliminary DLS runs showed aggregates larger than the filter size, even after being filtered. Both scenarios indicate that the peptides form aggregates after a short amount of time.

TABLE 9.7: Overview of data obtained from DLS experiments. Samples were measured under four different conditions with 0, 35, 70, and 140 mM NaCl, respectively. m = average mass percentage contained in peak. If data shows several peaks of significance, px indicates the peak number. If no standard deviation is indicated, it means that only one measurement was usable. − indicates that none of the measurements were usable. × means that a second peak was not present in the given condition.

[NaCl] (mM)	0		35		70		140	
Sample	R_h (nm)	m (%)	R_h (nm)	m (%)	R_h (nm)	m (%)	R_h (nm)	m (%)
GLP-1(14-34) p1	2.1 ± 0.5	96.0	3.0 ± 0.5	97.5	4.3 ± 0.4	99.8	4.2 ± 0.2	99.9
GLP-1(14-34) p2	6.5	1.1	12 ± 3	1.0	×	×	×	×
GLP-1(7-36)	2.2 ± 0.4	99.9	3.3 ± 0.1	99.9	2.6 ± 0.1	97.1	3.3 ± 0.2	99.5
C6Arev	1.8 ± 0.9	99.9	2.0 ± 0.1	99.7	2.0 ± 0.0	92.5	−	−
C6A p1	21 ± 2	13.3	18 ± 3	38.9	21 ± 2	17.9	19 ± 3	32.0
C6A p2	101 ± 4	86.7	87 ± 7	61.1	100 ± 10	82.1	96 ± 8	68.0
C8A	1.7 ± 0.1	100	2.0	100	2.1 ± 0.2	100	2.3 ± 0.2	100
C8LA	1.6 ± 0.1	99.9	1.7 ± 0.1	99.8	1.7 ± 0.1	99.9	2.0 ± 0.0	100
C10 p1	2.0 ± 0.5	55.4	2.6 ± 0.4	77.4	6 ± 4	79.9	15.4 ± 0.2	90.6
C10 p2	10.6 ± 0.2	34.2	15 ± 3	21.9	30 ± 4	19.7	100 ± 30	9.4
C16 p1	4.4 ± 0.1	99.8	6 ± 2	99.7	4.1	89.5	4.7 ± 0.1	90
C16 p2	×	×	×	×	10 ± 1	69.5	17.0	9.1
C10L p1	2.9 ± 0.3	97.3	−	−	3.0 ± 0.2	95.9	3.3 ± 0.3	99.3
C10L p2	13 ± 6	1.3	−	−	10.6 ± 0.5	0.4	12.2 ± 0.2	0.6
C16L(14-34)	3.6 ± 0.1	99.8	3.1 ± 0.1	99.8	3.2 ± 0.1	99.9	4.8 ± 0.2	100
C16L(7-37)	2.8 ± 0.1	100	3.3 ± 0.1	99.9	3.4 ± 0.1	100	3.3 ± 0.1	100

The short non-acylated peptide, GLP-1(14-34), has one main specie of around ∼2–4 nm in all conditions, and a larger specie for 0 and 35 mM salt, but only with

[2]The lower concentration of only 0.35 mg/mL gives noisier data, but is still shown.

around 1 % mass. The other non-acylated peptide, GLP-1(7-36) only has one specie of a similar size as the small specie in GLP-1(14-34) of around 2–3 nm.

Even though C6A and C6Arev have the same FA chain attached, and the same peptide chain length (though one is reversed), there is a substantial difference between the two. The reversed amino acid chain sample only contains one specie of ∼2 nm, but C6A has two species in all conditions. Both species vary in size, but with no particular trend concerning the salt concentrations. The smaller specie is much larger than those seen for any of the non-acylated species of between 18 nm to 21 nm and the larger of around 87–101 nm.

C8A and C8LA both have only one specie of around 2 nm in all conditions. There is a small variation in size for both peptides in the different conditions, but it is not very prominent, and there is no visible pattern in the variation. Common for GLP-1(14-34), GLP-1(7-36), C6A, and C6Arev is that none of them show up in AF4-MALS runs. This suggests, that they are either so small that they are lost through the membrane, or that they interact with the membrane. It could also suggest that they are caught against the membrane during elution with constant cross flow, and not released until the channel is cleaned at the end of the runs. However, comparing to C8A and C8LA, they are similar in size, if not slightly lager, but they both show up in AF4-MALS, contrary to GLP-1(14-34), GLP-1(7-36), C6A, and C6Arev. This contradicts our previous assumption of the peptides with no, or shorter, FA chains going through the membrane. For C6A, where both species are rather large, (around 20 nm for peak 1 and around 100 nm for peak 2) the case could be, that the larger particles are pushed against the membrane throughout the AF4 run, since the cross flow is constant. As mentioned before, the 'stress' of the liquid flow during an AF4 run might be enough to break any complex formation of larger species, causing the larger particles to show up in DLS batch mode, but not in AF4-MALS.

For the two peptides with long FA chains and no linker, C10 and C16, there are significant differences. For C10, the smaller specie seems to increase in size from 2 nm to around 15 nm with increasing salt concentration. The lager specie also increases in size, actually R_h increases ten times, from 10 nm to 100 nm. The deviation of the individual acquisitions also increase with the increase in R_h. For C16, R_h of the smaller specie is more stable in the varying conditions of about 4 nm to 6 nm, and the larger specie is only doubled in size (in regards to R_h).

C10L has two species present, like C10, but the larger specie is only present with about 1 % of the total mass. The sizes of the species for C10L is close to independent of the salt concentration. The specie sizes are similar to those of C10 in no salt and 35 mM salt of ∼3 nm and 11 nm to 13 nm, for the smaller and larger specie respectively. C16L(14-34) and C16L(7-37) both only contain one specie, the former of around 3–5 nm and the later of ∼3 nm. This matches the fact that the oligomerisation is slightly higher for C16L(14-34) than for C16L(7-37).

9.8.3 AF4-MALS: Peptides+HSA

The three salt containing buffers were used for testing peptide and HSA interactions. Since GLP-1(7-36), GLP-1(14-34), C6A, and C6Arev did not give usable LS signal on their own (red fields in table 9.6) these samples are not included in the following analyses. Furthermore, the C10L sample ran out, so a mixture of C10L and HSA is not run.

An overview of the resulting data is given in table 9.8.

TABLE 9.8: Overview of usable signals from each AF4 sequence run. Each sequence includes runs for six samples (first column) in three different buffers (top row). Fields noted with '−LS' and +UV' mean no LS peak, but a UV peak for the peptide samples. '−LS' and −UV' indicate no LS peak and no UV peak for the separate peptide sample. '+LS' and +UV' mean both a UV and an LS peak for the peptide sample. All runs show mono- and dimer peaks for HSA.

	pH/[NaCl] (mM)		
Sample	10.3/35	10.3/70	10.3/140
C8A, C8LA	+UV −LS	+UV −LS	+UV −LS
C10, C16	−UV −LS	−UV −LS	−UV −LS
C16L(14-34), C16L(7-37)	+UV +LS	+UV +LS	+UV +LS

Table 9.8 shows that there is no detectable LS signal, but a UV signal, for any sample eluting before the HSA peaks for C8A and C8LA. For C10 and C16 there are neither LS nor UV peaks eluting before HSA. For C16L(14-34) and C16(7-37), there are both LS and UV peaks eluting before HSA. The actual LS and UV signals are shown in figs. 9.16 and 9.17 for clarification.

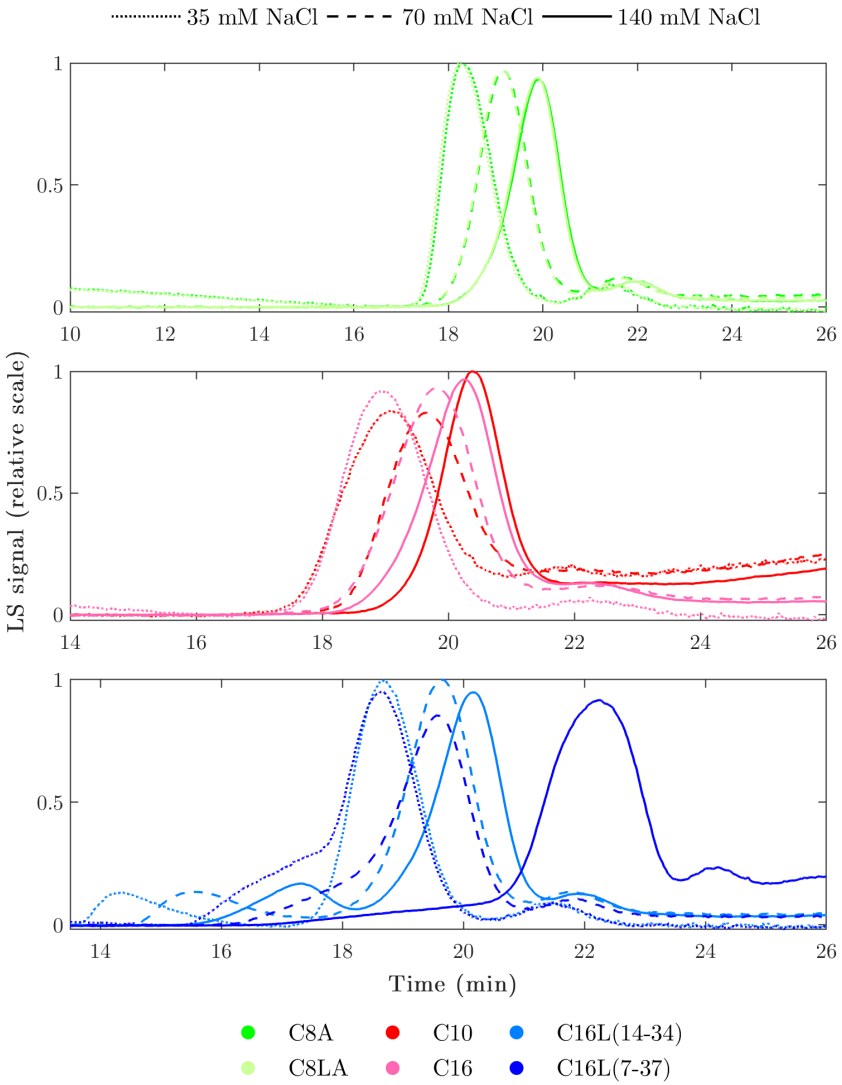


FIGURE 9.16: LS signals for the six peptide+HSA mixtures.

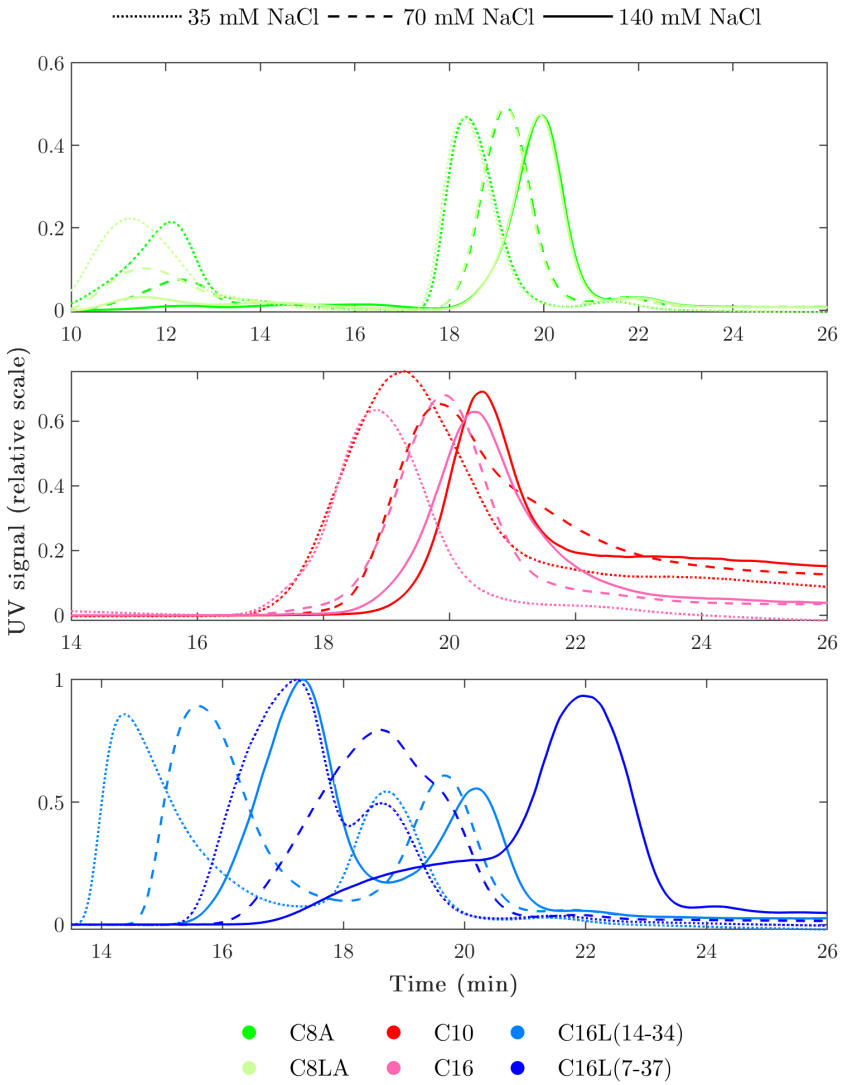


FIGURE 9.17: UV signals for the six peptide+HSA mixtures.

In fig. 9.16, LS peaks are visible for C16L(14-34) and C16L(7-37) with masses corresponding to the masses calculated from the free peptide runs. This means that there is oligomer formation in these samples when mixed with HSA.

Since the signal to noise ratio is rather small for C8A and C8LA in the peptide runs, it is likely that the 'missing' LS peaks in the mixtures are due to the large signal from HSA, which will then disguise that of the peptides as noise. For C10 and C16 there are no free peptide peaks in the LS signal for the mixture runs, but nice peaks for the peptide runs. Therefore, it is more likely that C10 and C16 actually interact with HSA. Another possibility is that the hydrodynamic radii of these species are of the same size as HSA (3.8 nm to 4.3 nm), which would mean that they would elute at the same time from the AF4 channel. Though, for C10, DLS shows that the species present are either smaller or larger than HSA. For C16, however, one specie is similar in hydrodynamic radius (4.1 nm and 4.4 nm) to HSA. For C16L(14-34) and C16L(7-37) there are peaks matching those of the free peptide, so these do not bind, or does only partly bind to HSA.

To investigate possible complex formation, the % mass recovered for the peptide peaks are compared to those of the free peptide runs. This is based on any UV

signal arising before HSA. Since the standard deviation on the LS signal is about a few kDa, it is not possible to see complex formation based on the mass of the HSA peak(s). Hence, the following is based on a comparison of the UV peptide peaks from the peptide alone, with the UV peptide peaks from the peptide+HSA mixtures.

The peptide mass recovery for the free peptide sample runs, as well as that taken from the peptide+HSA runs are plotted in fig. 9.18. Assuming that the peptide and HSA can be separated for C10 and C16, we can conclude that there is nothing eluting before HSA. This means that the % mass of the peptide peak is specified as 0 % for the peptide+HSA runs.

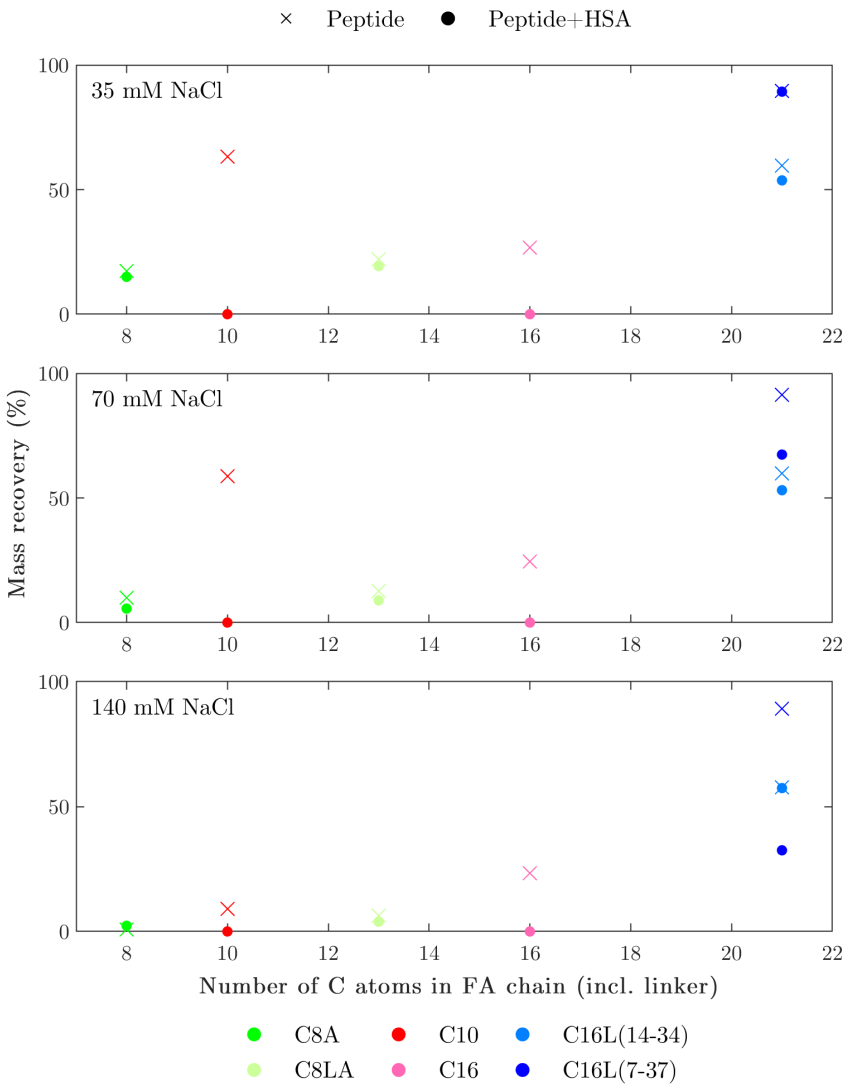


FIGURE 9.18: % mass recovered for the peptide peaks in the free peptide sample runs and the peptide+HSA runs.

From the plotted mass recoveries in fig. 9.18, it seems that the C10 and C16 samples completely bind to HSA. There is a significant decrease in % mass recovered for the peptides, from ∼60 % to 0 % for C10 in 35 mM and 70 mM NaCl and about 20 % to 0 % for C16 in all three salt concentrations. In the case of C8A and C8LA the recovered mass in the peptide peak is very low and similar for both free peptide and peptide+HSA mixtures. For C16L(14-34) and C16L(7-37), there is little to no decrease for 35 mM NaCl. This stays the same for C16L(14-34) in 70 mM and

140 mM NaCl, whereas there is a decrease for C16L(7-37) from ∼90 % to about 70 % recovered mass for the peptide peak in 70 mM NaCl, and a decrease from ∼90 % to about 30 % recovered mass for the peptide peak in 140 mM NaCl. This indicates that C16L(14-34) does not interact with HSA, whereas C16L(7-37) seems to have a low degree of interaction with HSA. The interaction between C16L(7-37) and HSA increases with increasing salt concentration. Liraglutide (C16L(7-37)) has previously been shown to bind partially to HSA [161].

The fact that C16L(7-37) does interact with HSA, and C16L(14-34) does not, is peculiar, since the attached FA chains are identical for these peptides. The only difference is the length of the amino acid chain. C16L(7-37) contains a few more charged residues (His7*, Glu9*, and Arg36*) than C16L(14-34). A possible reason for this could be that the charged residues are important for obtaining interactions to HSA. Such interactions could occur on the surface of HSA, rather than inside the FA binding pockets of HSA, which is otherwise generally believed [67, 138].

9.9 Conclusion

Our studies give several indications that the peptides form oligomers via hydrophobic interactions with the FA chains, and the longer the chain the more stable the interactions are. In combination with the better quality of the LS signal, it seems that the peptides with longer FA chains are retained better against the membrane, possibly because they do form larger and more stable oligomers. From AF4-MALS mass recoveries, it seems like C16L(7-37) is the only peptide where we for sure can say, that only one specie is present. On the other hand, DLS data, show that several peptides only have one specie present. However, it is indeed clear that there are more than one specie present for some of the peptides. Furthermore, the hydrodynamic radii of the peptides with none or short FA chains (GLP-1(7-36), GLP-1(14-34), C6Arev, and C6A) are not significantly smaller than for the peptides with longer FA chains (C8A, C8LA, C10, C10L, C16, C16L(7-37), C16L(14-34)), and thus should show up in the AF4 runs, and not be lost through the membrane. Actually, C6A only has species much larger than for any of the peptides that does show up in AF4. This indicates that the peptide is not going through the membrane, but is actually retained against it, and possibly not eluted using constant cross flow.

In regards to possible peptide interactions to HSA, we cannot say anything conclusive about C8A and C8LA due to low signals.
In conclusion, the studies show:

- The S/N for the LS signal increase with the length of the FA chain
- The SD on the LS signal decreases with the length of the FA chain
- Longer FA chains give higher oligomeric state
- Longer FA chains give higher mass recovery
- Peptides with FA chains longer than six C-atoms are more stable
- Presence of linker decreases oligomeric size
- Increased salt concentration causes increase in oligomeric size
- To vary pH from 8.0 to 10.3 has no effect on oligomeric size
- C16L(14-34) does not interact with HSA

- C16L(7-37) (liraglutide) show some degree of interactions with HSA
- C10 and C16 fully interacts with HSA

Chapter 10

Future work

10.1 Experimental

Since all experiments were measured only once, a good start would be to perform triplicates, at least duplicates, of the performed measurements. This will also ensure reproducibility. Furthermore, some additional experimenting with the AF4 methods would also be prudent. For the peptide runs, the cross flow was kept constant, and it could be interesting to see if different results will be obtained using an exponential gradient. Especially, since DLS measurements indicate larger species in some of the samples.

For better comparison of the UV and LS signals between the peptide and HSA + peptide mixtures, identical methods should be used for both type of measurements. For instance, the method used for the mixtures could be applied for the free peptides. On the same note, it would be interesting to vary the amount of added peptide to the HSA mixtures. This is to see if titration with the peptide is possible. Such experiments could reveal when HSA gets saturated with peptide.

10.2 Computational

Simulations were run on the activated receptor without peptide agonist in it. In addition, it would be interesting to compare the resulting data to simulations of the inactivated receptor with and without peptide in it. Such simulations could show if the inactivated receptor becomes active and vice versa. Doing this could hopefully help clarify the effect of the structural changes seen in the receptor upon activation. Furthermore, it would be intriguing to perform of a simulation of the C16L,$K^{26,34}$ system, with the peptide outside of the receptor and guided towards the bound state as already simulated. This could possibly shed some light on the hypothesis of the two FA chains interacting and, therefore, preventing initial binding to the receptor ECD.

10.3 Experimental output as computational input

In regards to the solution structures of the peptides, it would be very interesting to set up simulations based on the experimental output. Oligomeric states based on the mass from MALS, together with the physical sizes and the shape from R_h (DLS) and R_g (MALS), could be used to construct oligomeric structures. Such output could be used for simulation input. This will elaborate on the atomic interactions in the oligomers, and can give more detailed insights into the half-life extension due to FA induced oligomerisation.

References

[1] Geneva: World Health Organization, World Health Statistics 2018: Monitoring health for the SDGs, sustainable development goals, **2018**.

[2] Geneva: World Health Organization, Global Health Observatory (GHO) data: Life expectancy, **2018**.

[3] Geneva: World Health Organization, The top 10 causes of death, **2018**.

[4] B Moore, "On the treatment of Diabetus mellitus by acid extract of Duodenal Mucous Membrane.", *The Biochemical Journal* **1906**, *1*, 28–38.

[5] D. Shields, T. G. Warren, S. E. Roth, M. J. Brenner, "Cell-free synthesis and processing of multiple precursors to glucagon", *Nature* **1981**, *289*, 511–514.

[6] P. K. Lund, R. H. Goodman, J. F. Habener, "Pancreatic pre-proglucagons are encoded by two separate mRNAs", *Journal of Biological Chemistry* **1981**, *256*, 6515–6518.

[7] S. Mojsov, G. C. Weir, J. F. Habener, "Insulinotropin: glucagon-like peptide I (7-37) co-encoded in the glucagon gene is a potent stimulator of insulin release in the perfused rat pancreas.", *Journal of Clinical Investigation* **1987**, *79*, 616–619.

[8] N. Read, S. French, K. Cunningham, "The Role of the Gut in Regulating Food Intake in Man", *Nutrition Reviews* **1994**, *52*, 1–10.

[9] M. A. Nauck, U Niedereichholz, R Ettler, J. J. Holst, C Orskov, R Ritzel, W. H. Schmiegel, "Glucagon-like peptide 1 inhibition of gastric emptying outweighs its insulinotropic effects in healthy humans.", *The American Journal of Physiology* **1997**, *273*, E981–E988.

[10] A Flint, A Raben, A Astrup, J. J. Holst, "Glucagon-like peptide 1 promotes satiety and suppresses energy intake in humans.", *Journal of Clinical Investigation* **1998**, *101*, 515–520.

[11] K. Velmurugan, R. Bouchard, G. Mahaffey, S. Pugazhenthi, "Neuroprotective actions of glucagon-like peptide-1 in differentiated human neuroprogenitor cells", *Journal of Neurochemistry* **2012**, *123*, 919–931.

[12] L. Zhao, J. Xu, Q. Wang, Z. Qian, W. Feng, X. Yin, Y. Fang, "Protective effect of rhGLP-1 (7–36) on brain ischemia/reperfusion damage in diabetic rats", *Brain Research* **2015**, *1602*, 153–159.

[13] J. R. Ussher, D. J. Drucker, "Cardiovascular actions of incretin-based therapies", *Circulation Research* **2014**, *114*, 1788–1803.

[14] D. J. Drucker, "The Cardiovascular Biology of Glucagon-like Peptide-1", *Cell Metabolism* **2016**, *24*, 15–30.

[15] J. Holst, C. Ørskov, O. Vagn Nielsen, T. Schwartz, "Truncated glucagon-like peptide I, an insulin-releasing hormone from the distal gut", *FEBS Letters* **1987**, *211*, 169–174.

[16] C Orskov, M Bersani, A. H. Johnsen, P Højrup, J. J. Holst, "Complete Sequences of Glucagon-like Peptide-1 from Human and Pig Small Intestine", *Journal of Biological Chemistry* **1989**, *264*, 12826–12829.

[17] C. Orskov, L. Rabenhoj, A. Wettergren, H. Kofod, J. J. Holst, "Tissue and Plasma Concentrations of Amidated and Glycine-Extended Glucagon-Like Peptide I in Humans", *Diabetes* **1994**, *43*, 535–539.

[18] A. Wettergren, L. Pridal, M. Wøjdemann, J. J. Holst, "Amidated and non-amidated glucagon-like peptide-1 (GLP-1): non-pancreatic effects (cephalic phase acid secretion) and stability in plasma in humans", *Regulatory Peptides* **1998**, *77*, 83–87.

[19] K Thornton, D. G. Gorenstein, "Structure of glucagon-like peptide (7-36) amide in a dodecylphosphocholine micelle as determined by 2D NMR", *Biochemistry* **1994**, *33*, 3532–3539.

[20] X. Chang, D. Keller, S. Bjørn, J. J. Led, "Structure and folding of glucagon-like peptide-1-(7-36)-amide in aqueous trifluoroethanol studied by NMR spectroscopy", *Magnetic Resonance in Chemistry* **2001**, *39*, 477–483.

[21] X. Chang, D. Keller, S. I. O'Donoghue, J. J. Led, "NMR studies of the aggregation of glucagon-like peptide-1: Formation of a symmetric helical dimer", *FEBS Letters* **2002**, *515*, 165–170.

[22] J. W. White, G. F. Saunders, "Structure of the human ghicagon gene", *Nucleic Acids Research* **1986**, *14*, 4719–4730.

[23] S. Mojsov, G. Heinrich, I. B. Wilson, M. Ravazzola, L. Orci, J. F. Habener, "Preproglucagon gene expression in pancreas and intestine diversifies at the level of post-translational processing", *Journal of Biological Chemistry* **1986**, *261*, 11880–11889.

[24] D. J. Drucker, S. Asa, "Glucagon gene expression in vertebrate brain", *Journal of Biological Chemistry* **1988**, *263*, 13475–13478.

[25] C. d. Graaf, D. Donnelly, D. Wootten, J. Lau, P. M. Sexton, L. J. Miller, J.-M. Ahn, J. Liao, M. M. Fletcher, D. Yang, A. J. H. Brown, C. Zhou, J. Deng, M.-W. Wang, "Glucagon-Like Peptide-1 and Its Class B G Protein-Coupled Receptors: A Long March to Therapeutic Successes", *Pharmacological Reviews* **2016**, *68*, 954–1013.

[26] L. L. Baggio, D. J. Drucker, "Biology of incretins: GLP-1 and GIP", *Gastroenterology* **2007**, *132*, 2131–2157.

[27] Y. Wei, S. Mojsov, "Tissue-specific expression of the human receptor for glucagon-like peptide-I: brain, heart and pancreatic forms have the same deduced amino acid sequences", *FEBS Letters* **1995**, *358*, 219–224.

[28] B. Thorens, "Expression cloning of the pancreatic beta cell receptor for the gluco-incretin hormone glucagon-like peptide 1.", *Proceedings of the National Academy of Sciences* **1992**, *89*, 8641–8645.

[29] B. Thorens, A. Porret, L. Buhler, S. P. Deng, P. Morel, C. Widmann, "Cloning and functional expression of the human islet GLP-1 receptor. Demonstration that exendin-4 is an agonist and exendin-(9-39) an antagonist of the receptor", *Diabetes* **1993**, *42*, 1678–1682.

[30] K. E. Mayo, "International Union of Pharmacology. XXXV. The Glucagon Receptor Family", *Pharmacological Reviews* **2003**, *55*, 167–194.

[31] C. Parthier, S. Reedtz-Runge, R. Rudolph, M. T. Stubbs, "Passing the baton in class B GPCRs: peptide hormone activation via helix induction?", *Trends in Biochemical Sciences* **2009**, *34*, 303–310.

[32] D. Donnelly, "The structure and function of the glucagon-like peptide-1 receptor and its ligands", *British Journal of Pharmacology* **2012**, *166*, 27–41.

[33] K. Hollenstein, C. De Graaf, A. Bortolato, M. W. Wang, F. H. Marshall, R. C. Stevens, "Insights into the structure of class B GPCRs", *Trends in Pharmacological Sciences* **2014**, *35*, 12–22.

[34] C. Bergwitz, T. J. Gardella, M. R. Flannery, J. T. Potts, H. M. Kronenberg, S. R. Goldring, H. Jüppner, "Full Activation of Chimeric Receptors by Hybrids between Parathyroid Hormone and Calcitonin", *Journal of Biological Chemistry* **1996**, *271*, 26469–26472.

[35] S. Hoare, "Mechanisms of peptide and nonpeptide ligand binding to Class B G-protein-coupled receptors", *Drug Discovery Today* **2005**, *10*, 417–427.

[36] C. Gómez Santiago, E. Paci, D. Donnelly, "A mechanism for agonist activation of the glucagon-like peptide-1 (GLP-1) receptor through modelling & molecular dynamics", *Biochemical and Biophysical Research Communications* **2018**, *498*, 359–365.

[37] F. S. Willard, K. W. Sloop, "Physiology and emerging biochemistry of the glucagon-like peptide-1 receptor", *Experimental Diabetes Research* **2012**, *2012*, 470851.

[38] K. Adelhorst, B. B. Hedegaard, "Structure-activity studies of glucagon-like peptide-1", *Journal of Biological Chemistry* **1994**, *269*, 6275–6278.

[39] E. Alvarez, M. D. Martinez, I. Roncero, J. A. Chowen, B. Garcia-Cuartero, J. D. Gispert, C. Sanz, P. Vazquez, A. Maldonado, J. d. Caceres, M. Desco, M. A. Pozo, E. Blazquez, "The expression of GLP-1 receptor mRNA and protein allows the effect of GLP-1 on glucose metabolism in the human hypothalamus and brainstem", *Journal of Neurochemistry* **2005**, *92*, 798–806.

[40] B. P. Bullock, "Tissue distribution of messenger ribonucleic acid encoding the rat glucagon-like peptide-1 receptor", *Endocrinology* **1996**, *137*, 2968–2978.

[41] R. V. Campos, "Divergent tissue-specific and developmental expression of receptors for glucagon and glucagon-like peptide-1 in the mouse", *Endocrinology* **1994**, *134*, 2156–2164.

[42] Y. Wei, S. Mojsov, "Distribution of GLP-1 and PACAP receptors in human tissues", *Acta Physiologica Scandinavica* **1996**, *157*, 355–357.

[43] N. Mcintyre, C. Holdsworth, D. Turner, "New Interpretation of Oral Glucose Tolerance", *The Lancet* **1964**, *284*, 20–21.

[44] J. Gromada, B. Brock, O. Schmitz, P. Rorsman, "Glucagon-Like Peptide-1: Regulation of Insulin Secretion and Therapeutic Potential", *Basic and Clinical Pharmacology and Toxicology* **2004**, *95*, 252–262.

[45] A. Wettergren, B. Schjoldager, P. E. Mortensen, J. Myhre, J. Christiansen, J. J. Holst, "Truncated GLP-1 (proglucagon 78?107-amide) inhibits gastric and pancreatic functions in man", *Digestive Diseases and Sciences* **1993**, *38*, 665–673.

[46] S.-L. C. Jin, V. K. M. Han, J. G. Simmons, A. C. Towle, J. M. Lauder, P. K. Lund, "Distribution of glucagonlike peptide I (GLP-I), glucagon, and glicentin in the rat brain: An immunocytochemical study", *The Journal of Comparative Neurology* **1988**, *271*, 519–532.

[47] C. F. Deacon, M. A. Nauck, M. Toft-Nielsen, L. Pridal, B. Willms, J. J. Holst, "Both subcutaneously and intravenously administered glucagon-like peptide I are rapidly degraded from the NH2-terminus in type II diabetic patients and in healthy subjects", *Diabetes* **1995**, *44*, 1126–1131.

[48] C. Ruiz-Grande, C. Alarcón, A. Alcántara, C. Castilla, J. López Novoa, M. Villanueva-Peñacarrillo, I. Valverde, "Renal Catabolism of Truncated Glucagon-Like Peptide 1", *Hormone and Metabolic Research* **1993**, *25*, 612–616.

[49] C. F. Deacon, "Therapeutic strategies based on glucagon-like peptide 1", *Diabetes* **2004**, *53*, 2181–2189.

[50] D. J. Drucker, M. A. Nauck, "The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes", *Lancet* **2006**, *368*, 1696–1705.

[51] B. Ahrén, "GLP-1 for type 2 diabetes", *Experimental Cell Research* **2011**, *317*, 1239–1245.

[52] L. B. Knudsen, P. F. Nielsen, P. O. Huusfeldt, N. L. Johansen, K. Madsen, F. Z. Pedersen, H. Thøgersen, M. Wilken, H. Agersø, "Potent derivatives of glucagon-like peptide-1 with pharmacokinetic properties suitable for once daily administration", *Journal of Medicinal Chemistry* **2000**, *43*, 1664–1669.

[53] S Briyal, S Shah, a Gulati, "Neuroprotective and anti-apoptotic effects of liraglutide in the rat brain following focal cerebral ischemia", *Neuroscience* **2014**, *281*, 269–281.

[54] K. Talbot, H. Y. Wang, "The nature, significance, and glucagon-like peptide-1 analog treatment of brain insulin resistance in Alzheimer's disease", *Alzheimer's and Dementia* **2014**, *10*, S12–S25.

[55] P. L. McClean, C. Hölscher, "Liraglutide can reverse memory impairment, synaptic loss and reduce plaque load in aged APP/PS1 mice, a model of Alzheimer's disease", *Neuropharmacology* **2014**, *76*, 57–67.

[56] M. A. Nauck, "Liraglutide, a once-daily human GLP-1 analogue", *The British Journal of Diabetes & Vascular Disease* **2008**, *8*, S26–S33.

[57] T. Peters in *Advances in Protein Chemistry*, *Vol. 37*, Serum Albumin, **1985**, pp. 161–245.

[58] T. M. Frederiksen, P. Sønderby, L. A. Ryberg, P. Harris, J. T. Bukrinski, A. M. Scharff-Poulsen, M. N. Elf-Lind, G. H. Peters, "Oligomerization of a Glucagon-like Peptide 1 Analog: Bridging Experiment and Simulations", *Biophysical Journal* **2015**, *109*, 1202–1213.

[59] D. J. Drucker, A. Dritselis, P. Kirkpatrick, "Liraglutide", *Nature Reviews Drug Discovery* **2010**, *9*, 267–268.

[60] V. Gupta, "Glucagon-like peptide-1 analogues: An overview", *Indian Journal of Endocrinology and Metabolism* **2013**, *17*, 413.

[61] S. Madsbad, U. Kielgast, M. Asmar, C. F. Deacon, S. S. Torekov, J. J. Holst, "An overview of once-weekly glucagon-like peptide-1 receptor agonists-available efficacy and safety data and perspectives for the future", *Diabetes Obesity and Metabolism* **2011**, *13*, 394–407.

[62] A. Jazayeri, M. Rappas, A. J. H. Brown, J. Kean, J. C. Errey, N. J. Robertson, C. Fiez-Vandal, S. P. Andrews, M. Congreve, A. Bortolato, J. S. Mason, A. H. Baig, I. Teobald, A. S. Doré, M. Weir, R. M. Cooke, F. H. Marshall, "Crystal structure of the GLP-1 receptor bound to a peptide agonist", *Nature* **2017**, *546*, 254–258.

[63] Y. Zhang, B. Sun, D. Feng, H. Hu, M. Chu, Q. Qu, J. T. Tarrasch, S. Li, T. Sun Kobilka, B. K. Kobilka, G. Skiniotis, "Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein", *Nature* **2017**, *546*, 248–253.

[64] Y.-L. Liang, M. Khoshouei, A. Glukhova, S. G. B. Furness, P. Zhao, L. Clydesdale, C. Koole, T. T. Truong, D. M. Thal, S. Lei, M. Radjainia, R. Danev, W. Baumeister, M.-W. Wang, L. J. Miller, A. Christopoulos, P. M. Sexton, D. Wootten, "Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor–Gs complex", *Nature* **2018**, *555*, 121–125.

[65] S. Runge, H. Thøgersen, K. Madsen, J. Lau, R. Rudolph, "Crystal structure of the ligand-bound glucagon-like peptide-1 receptor extracellular domain", *Journal of Biological Chemistry* **2008**, *283*, 11340–11347.

[66] C. R. Underwood, P. Garibay, L. B. Knudsen, S. Hastrup, G. H. Peters, R. Rudolph, S. Reedtz-Runge, "Crystal structure of glucagon-like peptide-1 in complex with the extracellular domain of the glucagon-like peptide-1 receptor", *Journal of Biological Chemistry* **2010**, *285*, 723–730.

[67] L. B. Knudsen in *Analogue-Based Drug Discovery II*, **2010**, pp. 335–357.

[68] J. Lau, P. Bloch, L. Schäffer, I. Pettersson, J. Spetzler, J. Kofoed, K. Madsen, L. B. Knudsen, J. McGuire, D. B. Steensgaard, H. M. Strauss, D. X. Gram, S. M. Knudsen, F. S. Nielsen, P. Thygesen, S. Reedtz-Runge, T. Kruse, "Discovery of the Once-Weekly Glucagon-Like Peptide-1 (GLP-1) Analogue Semaglutide", *Journal of Medicinal Chemistry* **2015**, *58*, 7370–7380.

[69] S Hug in *Biomolecular Simulations - Methods and Protocols*, (Eds.: L Monticelli, E Salonen), Classical Molecular Dynamics in a Nutshell, Humana Press, New York, **2013**, Chapter 6.

[70] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function", *Current Opinion in Structural Biology* **2009**, *19*, 120–127.

[71] K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, D. Kern, "A hierarchy of timescales in protein dynamics is linked to enzyme catalysis", *Nature* **2007**, *450*, 913–916.

[72] G. H. Peters in *Enzyme Functionality - Design, Engineering, and Screening*, (Eds.: A. Svendsen, M. Dekker), Computer Simulations: A Tool for Investigating the Function of Complex Biological Macromolecules, New York, **2004**, Chapter 6.

[73] A. R. Leach in *Molecular Modelling: Principles and Applications*, Molecular Dynamics Simulation Methods, Pearson Education Limited, Essex, **2001**, Chapter 7.

[74] L Monticelli, D. P. Tieleman in *Biomolecular Simulations - Methods and Protocols*, (Eds.: L Monticelli, E Salonen), Force Fields for Classical Molecular Dynamics, Humana Press, New York, **2013**, Chapter 8.

[75] C Lorens, N. L. Doltsinis in *Handbook of computational chemistry*, (Ed.: J Leszczynski), Molecular Dynamics Simulation: From "Ab Initio" to "Coarse Grained", Springer Science + Business Media B.V., Dordrecht, **2012**, Chapter 8.

[76] R. J. Loncharich, B. R. Brooks, R. W. Pastor, "Langevin dynamics of peptides: The frictional dependence of isomerization rates ofN-acetylalanyl-N?-methylamide", *Biopolymers* **1992**, *32*, 523–535.

[77] K.-H. Chow, D. M. Ferguson, "Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling", *Computer Physics Communications* **1995**, *91*, 283–289.

[78] J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, B. O. Brandsdal, "Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm", *Chemical Physics Letters* **2004**, *384*, 288–294.

[79] H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola, J. R. Haak, "Molecular dynamics with coupling to an external bath", *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

[80] T. Schlick in *Molecular Modeling and Simulation: An Interdisciplinary Guide*, *Vol. 21*, **2010**, Chapter 10.

[81] Schrödinger, Schrödinger Release 2018-4: Maestro, Schrödinger, LLC, New York, NY, **2018**.

[82] J. D. Thompson, D. G. Higgins, T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic Acids Research* **1994**, *22*, 4673–4680.

[83] M. Larkin, G. Blackshields, N. Brown, R. Chenna, P. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm, R. Lopez, J. Thompson, T. Gibson, D. Higgins, "Clustal W and Clustal X version 2.0", *Bioinformatics* **2007**, *23*, 2947–2948.

[84] J. Gullingsrud, J. Saam, J. Phillips, VMD psfgen Plugin, Version 1.6.4, `https://www.ks.uiuc.edu/Research/vmd/plugins/psfgen/`, University of Illinois and Beckman Institute, Theoretical and Computational Biophysics Group, **Mar. 2017**.

[85] W. Humphrey, A. Dalke, K. Schulten, "VMD: visual molecular dynamics", *Journal of Molecular Graphics* **Feb. 1996**, *14*, 33–38.

[86] A. Aksimentiev, M. Sotomayor, D. Wells, Membrane Proteins Tutorial, `http://www.ks.uiuc.edu/Training/Tutorials/science/membrane/membrane-tutorial.pdf`, University of Illinois at Urbana-Champaign, Beckman Institute for Advanced Science, Technology, Theoretical, and Computational Biophysics Group, **Nov. 2016**.

[87] H Grubmüller, V. Groll, Solvate, `http://www.mpibpc.mpg.de/grubmueller/solvate`, Max Planck Institute for Biophysical Chemistry, **1996**.

[88] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *The Journal of Chemical Physics* **1983**, *79*, 926.

[89] I. Balabin, M. Sotomayor, L. Trabuco, Autoionize Plugin, Version 1.5, `https://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/`, University of Illinois at Urbana-Champaign, Theoretical and Computational Biophysics Group.

[90] M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kale, R. D. Skeel, K. Schulten, "NAMD: a parallel, object-oriented molecular dynamics program", *International Journal of High Performance Computing Applications* **1996**, *10*, 251–268.

[91] A. MacKerell, D Bashford, "All-atom empirical potential for molecular modeling and dynamics studies of proteins", *The Journal of Physical Chemistry. B* **1998**, *5647*, 3586–3616.

[92] D. Case, R. Betz, D. Cerutti, T. Cheatham III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. Merz, G. Monard, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. Roe, A. Roitberg, C. Sagui, C. Simmerling, W. Botello-Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, L. Xiao, P. Kollman, AMBER 2016, University of California, San Francisco, **2016**.

[93] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born", *Journal of Chemical Theory and Computation* **2012**, *8*, 1542–1555.

[94] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald", *Journal of Chemical Theory and Computation* **2013**, *9*, 3878–3888.

[95] R. Walker, M. Crowley, M. Williamson, J. Deckman, R. Betz, Chamber in ParmEd, `https://parmed.github.io/ParmEd/html/parmed.html#chamber`, University of California, San Francisco, **2010**.

[96] J. Swails, C. Hernandez, D. L. Mobley, H. Nguyen, L.-P. Wang, P. Janowski, ParmEd: a Parameter file Editor, `https://github.com/ParmEd/ParmEd`, University of California, San Francisco, **2010**.

[97] B. D. Madej, R. C. Walker, An Amber Lipid Force Field Tutorial: Lipid14 Edition, `http://ambermd.org/tutorials/advanced/tutorial16/index.html#Molecular_Dynamics`, University of California San Diego, **May 2014**.

[98] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes", *Journal of Computational Physics* **1977**, *23*, 327–341.

[99] P. Gonnet, "P-SHAKE: A quadratically convergent SHAKE in", *Journal of Computational Physics* **2007**, *220*, 740–750.

[100] S. Miyamoto, P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models", *Journal of Computational Chemistry* **1992**, *13*, 952–962.

[101] T. Darden, D. York, L. Pedersen, "Particle mesh Ewald: An N log(N) method for Ewald sums in large systems", *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.

[102] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, "A smooth particle mesh Ewald method", *Journal of Chemical Physics* **1995**, *103*, 8577–8593.

[103] D. Sitkoff, K. A. Sharp, B. Honig, "Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models", *The Journal of Physical Chemistry* **1994**, *98*, 1978–1988.

[104] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, N. A. Baker, "Improvements to the APBS biomolecular solvation software suite", *Protein Science* **2018**, *27*, 112–128.

[105] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations", *Nucleic Acids Research* **2004**, *32*, W665–W667.

[106] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, N. A. Baker, "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations", *Nucleic Acids Research* **2007**, *35*, W522–W525.

[107] A. Bakan, L. Meireles, I. Barhar, ProDy: Protein Dynamics & Sequence Analysis, **2010**.

[108] A. Bakan, L. M. Meireles, I. Bahar, "ProDy: Protein dynamics inferred from theory and experiments", *Bioinformatics* **2011**, *27*, 1575–1577.

[109] H. C. Zhang, X. L. Hu, D. Q. Yin, Z. F. Lin, "Development of molecular docking-based binding energy to predict the joint effect of BPA and its analogs", *Human and Experimental Toxicology* **2011**, *30*, 318–327.

[110] F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li, T. Hou, "Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking", *Physical Chemistry Chemical Physics* **2016**, *18*, 22129–22139.

[111] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, "Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models", *Accounts of Chemical Research* **2000**, *33*, 889–897.

[112] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, D. A. Case, "Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices", *Journal of the American Chemical Society* **1998**, *120*, 9401–9409.

[113] H. Gohlke, D. a. Case, "Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf", *Journal of Computational Chemistry* **2004**, *25*, 238–250.

[114] S. Liang, L. Li, W.-L. Hsu, M. N. Pilcher, V. Uversky, Y. Zhou, A. K. Dunker, S. O. Meroueh, "Exploring the molecular design of protein interaction sites with molecular dynamics simulations and free energy calculations.", *Biochemistry* **2009**, *48*, 399–414.

[115] H. Abroshan, H. Akbarzadeh, G. A. Parsafar, "Molecular dynamics simulation and MM-PBSA calculations of sickle cell hemoglobin in dimer form with Val, Trp, or Phe at the lateral contact", *Journal of Physical Organic Chemistry* **2010**, *23*, 866–877.

[116] S. Genheden, U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities", *Expert Opinion on Drug Discovery* **2015**, *10*, 449–461.

[117] D. Shivakumar, Y. Deng, B. Roux, "Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model", *Journal of Chemical Theory and Computation* **2009**, *5*, 919–930.

[118] R. B. Hermann, "Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area", *The Journal of Physical Chemistry* **1972**, *76*, 2754–2759.

[119] M. Orozco, F. J. Luque, "Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems", *Chemical Reviews* **2000**, *100*, 4187–4226.

[120] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome", *Proceedings of the National Academy of Sciences* **2001**, *98*, 10037–10041.

[121] N. A. Baker in *Methods in Enzymology*, *Vol. 383*, 1990, Poisson–Boltzmann Methods for Biomolecular Electrostatics, **2004**, pp. 94–118.

[122] S. Kassem, M. Ahmed, S. El-Sheikh, K. H. Barakat, "Entropy in bimolecular simulations: A comprehensive review of atomic fluctuations-based methods", *Journal of Molecular Graphics and Modelling* **2015**, *62*, 105–117.

[123] G. P. Brady, K. A. Sharp, "Entropy in protein folding and in protein—protein interactions", *Current Opinion in Structural Biology* **1997**, *7*, 215–221.

[124] R. Grünberg, M. Nilges, J. Leckner, "Flexibility and Conformational Entropy in Protein-Protein Binding", *Structure* **2006**, *14*, 683–693.

[125] N. Basdevant, H. Weinstein, M. Ceruso, "Thermodynamic Basis for Promiscuity and Selectivity in Protein-Protein Interactions: PDZ Domains, a Case Study", *Journal of the American Chemical Society* **2006**, *128*, 12766–12777.

[126] P. Kar, R. Lipowsky, V. Knecht, "Importance of Polar Solvation and Configurational Entropy for Design of Antiretroviral Drugs Targeting HIV-1 Protease", *The Journal of Physical Chemistry B* **2013**, *117*, 5793–5805.

[127] S. Swaminathan, T. Ichiye, W. Van Gunsteren, M. Karplus, "Time dependence of atomic fluctuations in proteins: analysis of local and collective motions in bovine pancreatic trypsin inhibitor", *Biochemistry* **1982**, *21*, 5230–5241.

[128] J. Brady, M. Karplus, "Configuration entropy of the alanine dipeptide in vacuum and in solution: a molecular dynamics study", *Journal of the American Chemical Society* **1985**, *107*, 6103–6105.

[129] K Coopman, R Wallis, G Robb, A. Brown, G. Wilkinson, D Timms, G. Willars, "Residues within the transmembrane domain of the glucagon-like peptide-1 receptor involved in ligand binding and receptor activation: modelling the ligand-bound receptor.", *Molecular Endocrinology* **2011**, *25*, 1804–1818.

[130] M. Karplus, J. N. Kushick, "Method for estimating the configurational entropy of macromolecules", *Macromolecules* **1981**, *14*, 325–332.

[131] J. Schlitter, "Estimation of absolute and relative entropies of macromolecules using the covariance matrix", *Chemical Physics Letters* **1993**, *215*, 617–621.

[132] H. Schäfer, A. E. Mark, W. F. Van Gunsteren, "Absolute entropies from molecular dynamics simulation trajectories", *Journal of Chemical Physics* **2000**, *113*, 7809–7817.

[133] K. Hinsen, "Analysis of domain motions by approximate normal mode calculations", *Proteins: Structure Function and Genetics* **1998**, *33*, 417–429.

[134] N. Homeyer, H. Gohlke, "Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method", *Molecular Informatics* **2012**, *31*, 114–122.

[135] StatTrek.com, Variance-Covariance Matrix, acces date 2016-09-09.

[136] Schrödinger, LLC, "The PyMOL Molecular Graphics System, Version 1.8", **2015**.

[137] M. A. Kenyon, J. A. Hamilton, "13C NMR studies of the binding of medium-chain fatty acids to human serum albumin", *Journal of Lipid Research* **1994**, *35*, 458–467.

[138] P. Kurtzhals, S. Havelund, I Jonassen, B. Kiehr, U. D. Larsen, U. Ribel, J. Markussen, "Albumin binding of insulins acylated with fatty acids: characterization of the ligand-protein interaction and correlation between binding affinity and timing of the insulin effect in vivo", *The Biochemical Journal* **1995**, *312*, 725–731.

[139] S. Myers, F. E. Yakubu-Madus, W. T. Johnson, J. E. Baker, T. S. Cusick, V. K. Williams, F. C. Tinsley, A. Kriauciunas, J. Manetta, V. J. Chen, "Acylation of Human Insulin With Palmitic Acid Extends the Time Action of Human Insulin in Diabetic Dogs", *Diabetes* **1997**, *46*, 637–642.

[140] J. Markussen, S. Havelund, P. Kurtzhals, A. S. Andersen, J. Halstrøm, E. Hasselager, U. D. Larsen, U. Ribel, L. Schæffer, K Vad, I Jonassen, "Soluble, fatty acid acylated insulins bind to albumin and show protracted action in pigs", *Diabetologia* **1996**, *39*, 281–288.

[141] P. Kurtzhals, S. Havelund, I. Jonassen, B. Kiehr, U. Ribel, J. Markussen, "Albumin Binding and Time Action of Acylated Insulins in Various Species", *Journal of Pharmaceutical Sciences* **1996**, *85*, 304–308.

[142] D. B. Steensgaard, J. K. Thomsen, H. B. Olsen, L. B. Knudsen, "The molecular basis for the delayed absorption of the once-daily human GLP-1 analoge, liraglutide", *Diabetes* **2008**, A164–A164.

[143] Y. Li, M. Shao, X. Zheng, W. Kong, J. Zhang, M. Gong, "Self-assembling peptides improve the stability of glucagon-like peptide-1 by forming a stable and sustained complex", *Molecular Pharmaceutics* **2013**, *10*, 3356–3365.

[144] Y. Wang, A. Lomakin, S. Kanai, R. Alex, G. B. Benedek, "Transformation of oligomers of lipidated peptide induced by change in pH", *Molecular Pharmaceutics* **2015**, *12*, 411–419.

[145] K. Ramalingam, S. Aimoto, J. Bello, "Conformational studies of anionic melittin analogues: Effect of peptide concentration, pH, ionic strength, and temperature-models for protein folding and halophilic proteins", *Biopolymers* **1992**, *32*, 981–992.

[146] J. C. Giddings, F. J. Yang, M. N. Myers, "Theoretical and experimental characterization of flow field-flow fractionation", *Analytical Chemistry* **1976**, *48*, 1126–1132.

[147] K. G. Wahlund, J. C. Giddings, "Properties of an asymmetrical flow field-flow fractionation channel having one permeable wall", *Analytical Chemistry* **1987**, *59*, 1332–1339.

[148] Wyatt Technology Europe GmbH, Field Flow Fractionation - Basics and Method Development, PowerPoint, Wyatt Technology Europe, Dernbach, Germany, **2016**.

[149] A. Litzén, K. G. Wahlund, "Zone Broadening and Dilution in Rectangular and Trapezoidal Asymmetrical Flow Field-Flow Fractionation Channels", *Analytical Chemistry* **1991**, *63*, 1001–1007.

[150] J. C. Giddings, F. J. Yang, M. N. Myers, "Flow field-flow fractionation: new method for separating, purifying, and characterizing the diffusivity of viruses", *Journal of Virology* **1977**, *21*, 131–138.

[151] J. Giddings, F. J. Yang, M. N. Myers, "Flow field-flow fractionation as a methodology for protein separation and characterization", *Analytical Biochemistry* **1977**, *81*, 395–407.

[152] M. Leeman, K.-g. Wahlund, B. Wittgren, "Programmed cross flow asymmetrical flow field-flow fractionation for the size separation of pullulans and hydroxypropyl cellulose", *Journal of Chromatography A* **2006**, *1134*, 236–245.

[153] M. Leeman, M. T. Islam, W. G. Haseltine, "Asymmetrical flow field-flow fractionation coupled with multi-angle light scattering and refractive index detections for characterization of ultra-high molar mass poly(acrylamide) flocculants", *Journal of Chromatography A* **2007**, *1172*, 194–203.

[154] Wyatt Technology Europe GmbH, Light Scattering University, LSU Course Manual, Wyatt Technology Europe, Dernbach, Germany, **2017**.

[155] P. J. Wyatt, "Light scattering and the absolute characterization of macromolecules", *Analytica Chimica Acta* **1993**, *272*, 1–40.

[156] B. H. Zimm, "The Scattering of Light and the Radial Distribution Function of High Polymer Solutions", *The Journal of Chemical Physics* **1948**, *16*, 1093–1099.

[157] K. S. Schmitz, G. D. J. Phillies, "An Introduction to Dynamic Light Scattering by Macromolecules", *Physics Today* **1991**, *44*, 66–66.

[158] Wyatt Technology Europe GmbH, Light Scattering University, Dyna-LSU Course Manual, Wyatt Technology Europe, Dernbach, Germany, **2016**.

[159] A. P. Minton, "Recent applications of light scattering measurement in the biological and biopharmaceutical sciences", *Analytical Biochemistry* **2016**, *501*, 4–22.

[160] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, A. Bairoch, "ExPASy: The proteomics server for in-depth protein knowledge and analysis", *Nucleic Acids Research* **2003**, *31*, 3784–3788.

[161] P. Sønderby, Solution behaviour of Human Serum Albumin and GLP-1variants, PhD thesis, DTU Chemistry, Kemitorvet 206, 2800 Kgs. Lyngby, Denmark, **2016**.

[162] S. E. Feller, Y. Zhang, R. W. Pastor, B. R. Brooks, "Constant-pressure molecular-dynamics simulation - the Langevin piston method", *Journal of Chemical Physics* **1995**, *103*, 4613–4621.

Appendices

Appendix A Generation of psf and pdb files for receptor-peptide complex - tcl script

The input file used to generate the GLP-1R–liraglutide system is given as an example.

```
mol load pdb GLP-1R_tot_5NX2_loop.pdb
set protsel1 [atomselect top "all"]
$protsel1 writepdb temp1.pdb

mol load pdb lira_no_clash.pdb
set protsel2 [atomselect top "all"]
$protsel2 writepdb temp2.pdb

package require psfgen
resetpsf
pdbalias residue HIS HSD
pdbalias atom ILE CD1 CD

topology /home/WIN/timf/Liraglutide/files/04-cgc-top-par-files/CHARMM/
    top_all36_prot_glp1a.rtf

segment SEG1 {
  pdb temp1.pdb
  first nter
  last cter
}
patch DISU SEG1:46 SEG1:71
patch DISU SEG1:62 SEG1:104
patch DISU SEG1:226 SEG1:296
patch DISU SEG1:126 SEG1:85
coordpdb temp1.pdb SEG1

segment SEG2 {
  pdb temp2.pdb
  first nter
  last cter
}
coordpdb temp2.pdb SEG2

guesscoord
writepsf GLP-1R_cryst-lira.psf
writepdb GLP-1R_cryst-lira.pdb

file delete -force -- temp1.pdb
file delete -force -- temp2.pdb

exit
```

Appendix B Insertion of complex structure into a membrane patch - tcl script

The input file used to generate a memrbane patch and insert the GLP-1R–liraglutide system is given as an example.

```
package require membrane
membrane -l POPC -x 80 -y 80 -o popc -top c36
set popc [atomselect top all]
set recmol [mol load psf GLP-1R_cryst-lira.psf pdb GLP-1R_cryst-lira.pdb]
set rec [atomselect $recmol all]
$popc moveby [vecinvert [measure center $popc weight mass]]
$popc writepdb popc_TEMP.pdb
set middle [atomselect $recmol "protein and resid 152 189 239 277 316 360
    394"]
$rec moveby [vecinvert [measure center $middle weight mass]]
display resetview
$rec move [transaxis x -45]
$rec move [transaxis y 22]
$rec move [transaxis z 55]
$rec moveby {-7 -7 0}
$rec writepdb rec_TEMP.pdb
mol delete all
package require psfgen
resetpsf
readpsf popc.psf
coordpdb popc_TEMP.pdb
readpsf GLP-1R_cryst-lira.psf
coordpdb rec_TEMP.pdb
writepsf GLP-1R_cryst-lira_popc_raw.psf
writepdb GLP-1R_cryst-lira_popc_raw.pdb
file delete rec_TEMP.pdb
file delete popc_TEMP.pdb
exit
```

Appendix C Removal of overlapping lipid and water molecules - tcl script

The input file used to remove lipid and water molecules overlapping with the protein complex, GLP-1R–liraglutide system is given as an example.

```
mol load psf GLP-1R_cryst-lira_popc_raw.psf pdb
    GLP-1R_cryst-lira_popc_raw.pdb
set POPC "resname POPC"
set all [atomselect top all]
$all set beta 0
set seltext1 "$POPC and same residue as (name P and z>0 and (x-5)^2+(y+5)
    ^2<144)"
#abs(x)<20 and abs(y)<20
set seltext2 "$POPC and same residue as (name P and z<0 and (x-5)^2+(y+5)
    ^2<144)"
set seltext3 "$POPC and same residue as (within 0.6 of protein)"
set sel1 [atomselect top $seltext1]
set sel2 [atomselect top $seltext2]
set sel3 [atomselect top $seltext3]
$sel1 set beta 1
$sel2 set beta 1
$sel3 set beta 1
set badlipid [atomselect top "name P and beta >0"]
set seglistlipid [$badlipid get segid]
set reslistlipid [$badlipid get resid]
set seltext4 "(water and not segname WCA WCB WCC WCD WF SOLV) and same
    residue as within 3 of ((same residue as (name P and beta>0)) or
    protein)"
set seltext5 "segname SOLV and same residue as within 3 of lipids"
set sel4 [atomselect top $seltext4]
set sel5 [atomselect top $seltext5]
$sel4 set beta 1
$sel5 set beta 1
set badwater [atomselect top "name OH2 and beta >0"]
set seglistwater [$badwater get segid]
set reslistwater [$badwater get resid]
mol delete all
package require psfgen
resetpsf
readpsf GLP-1R_cryst-lira_popc_raw.psf
coordpdb GLP-1R_cryst-lira_popc_raw.pdb
foreach segid $seglistlipid resid $reslistlipid {
delatom $segid $resid
}
foreach segid $seglistwater resid $reslistwater {
delatom $segid $resid
}
writepsf GLP-1R_cryst-lira_popc.psf
writepdb GLP-1R_cryst-lira_popc.pdb
exit
```

Appendix D Solvation of receptor-peptide complex systems - tcl script

The input file used to solvate the complexes above and below the membrane, the GLP-1R–liraglutide system is given as an example.

```
package require solvate
solvate GLP-1R_cryst-lira_popc.psf GLP-1R_cryst-lira_popc.pdb -o
    GLP-1R_cryst-lira_popc_water_TEMP -b 1.5 -minmax {{-38 -38 -78} {38 39
    43}}
set all [atomselect top all]
$all set beta 0
set seltext "segid WT1 to WT99 and same residue as abs(z) <25"
set sel [atomselect top $seltext]
$sel set beta 1
set badwater [atomselect top "name OH2 and beta > 0"]
set seglist [$badwater get segid]
set reslist [$badwater get resid]
mol delete all
package require psfgen
resetpsf
readpsf GLP-1R_cryst-lira_popc_water_TEMP.psf
coordpdb GLP-1R_cryst-lira_popc_water_TEMP.pdb
foreach segid $seglist resid $reslist {
delatom $segid $resid
}
writepdb GLP-1R_cryst-lira_popcw.pdb
writepsf GLP-1R_cryst-lira_popcw.psf
file delete GLP-1R_cryst-lira_popc_water_TEMP.psf
file delete GLP-1R_cryst-lira_popc_water_TEMP.pdb
exit
```

Appendix E Altered topology and paramter files for the CHARMM36 force field

Topology file

Changes were made to the original CHARMM36 topology to include the special residues in the different peptides. The used values were based on those for acetylated Lys.

```
!Modified August 2017 by TIMF for inclusion of residue 12A, 14A, C8M, D6M,
     D7M, and DAH for GLP-1 analogues.
!The residues are added alphabetically among the original residues.
!12A, 14A, C8M, D6M, and D7M are acetylated Lysine residues (addition of
    fatty acid chains).
!The added fatty acid chains are of different lenghts and added with or
    without a linker.
!12A includes acylated Lysine residue and a free acid group.
!14A includes acylated Lysine residue and a free acid group.
!C8M includes acylated Lysine residue.
!D6M includes modified Lysine residue: Lys(gamma-glu-palmitoyl/
    N-hexadecanoly-L-glutamic acid)
!(D6M was originally created by MNEL October 2013, but was modified by
    TIMF August 2017.)
!D7M includes modified Lysine residue: Lys(gamma-glu-palmitoyl/
    N-hexadecanoly-L-glutamic acid).
!DAH is a modified Histidine residue without the backbone amino group.


RESI 12A       -1.00  ! C-12 diacid *CHANGES FROM PARENT RESIDUE
GROUP
ATOM N      NH1     -0.47 !      HN   HA         O          LYSINE START
ATOM HN     H        0.31 !      |    |         //
ATOM CA     CT1      0.07 !    --N---CA--------C
ATOM HA     HB1      0.09 !          |          \
GROUP                    !          |    (The L notation has been removed
    to make recognition of C-term possible)
ATOM C      C        0.51 !          |
ATOM O      O       -0.51 !          |
GROUP                    !          |
ATOM CBL    CT2     -0.18 !     HBL1-CBL-HBL2
ATOM HBL1   HA2      0.09 !          |
ATOM HBL2   HA2      0.09 !          |
GROUP                    !          |
ATOM CGL    CT2     -0.18 !     HGL1-CGL-HGL2
aTOM HGL1   HA2      0.09 !          |
ATOM HGL2   HA2      0.09 !          |
GROUP                    !          |
ATOM CDL    CT2     -0.18 !     HDL1-CDL-HDL2
ATOM HDL1   HA2      0.09 !          |
ATOM HDL2   HA2      0.09 !          |
GROUP                    !          |
ATOM CEL    CT2     -0.02 ! *   HEL1-CEL-HEL2         (ACETYLATED LYSINE,
    ALY)
ATOM HEL1   HA2      0.09 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM HEL2   HA2      0.09 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM NZL    NH1     -0.47 ! *       NZL-HZL          (ACETYLATED LYSINE,
    ALY)
ATOM HZL    H        0.31 ! *        |                (ACETYLATED LYSINE,
    ALY)
```

```
GROUP                    !          |       LYS ENDS ACID CHAIN STARTS
ATOM CDG   C        0.51 ! *        CDG=ODG          (ACETYLATED LYSINE,
    ALY)
ATOM ODG   O       −0.51 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |
ATOM C1    CT2     −0.18 !      H1X−C1−H1Y
ATOM H1X   HA2      0.09 !          |
ATOM H1Y   HA2      0.09 !          |
GROUP                    !          |
ATOM C2    CT2     −0.18 !      H2X−C2−H2Y
ATOM H2X   HA2      0.09 !          |
ATOM H2Y   HA2      0.09 !          |
GROUP                    !          |
ATOM C3     CT2    −0.18 !      H3X−C3−H3Y
ATOM H3X    HA2     0.09 !          |
ATOM H3Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C4     CT2    −0.18 !      H4X−C4−H4Y
ATOM H4X    HA2     0.09 !          |
ATOM H4Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C5     CT2    −0.18 !      H5X−C5−H5Y
ATOM H5X    HA2     0.09 !          |
ATOM H5Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C6     CT2    −0.18 !      H6X−C6−H6Y
ATOM H6X    HA2     0.09 !          |
ATOM H6Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C7     CT2    −0.18 !      H7X−C7−H7Y
ATOM H7X    HA2     0.09 !          |
ATOM H7Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C8    CT2     −0.18 !      H8X−C8−H8Y
ATOM H8X   HA2      0.09 !          |
ATOM H8Y   HA2      0.09 !          |
GROUP                    !          |
ATOM C9    CT2     −0.18 !      H9X−C9−H9Y
ATOM H9X   HA2      0.09 !          |
ATOM H9Y   HA2      0.09 !          |
GROUP                    !          |
ATOM C10    CT2    −0.28 ! *   H10X−C10−H10Y  FROM GLU
ATOM H10X   HA2     0.09 !          |
ATOM H10Y   HA2     0.09 !          |
ATOM CD    CC       0.62 ! *        CD=OE1    FROM GLU
ATOM OE1   OC      −0.76 ! *        |
ATOM OE2   OC      −0.76 ! *        OE2(−)   FATTY ACID END

BOND N HN  N CA  CA HA
BOND CA C  C +N  !N −C
BOND CA CBL  CBL HBL1  CBL HBL2
BOND CBL CGL  CGL HGL1  CGL HGL2
BOND CGL CDL  CDL HDL1  CDL HDL2
BOND CDL CEL  CEL HEL1  CEL HEL2
BOND CEL NZL  NZL HZL
BOND NZL CDG
BOND CDG C1 C1 H1X  C1 H1Y
BOND C1 C2  C2 H2X  C2 H2Y
BOND C2 C3  C3 H3X  C3 H3Y
BOND C3 C4  C4 H4X  C4 H4Y
BOND C4 C5  C5 H5X  C5 H5Y
BOND C5 C6  C6 H6X  C6 H6Y
```

```
BOND C6 C7  C7 H7X  C7 H7Y
BOND C7 C8  C8 H8X  C8 H8Y
BOND C8 C9  C9 H9X  C9 H9Y
BOND C9 C10 C10 H10X C10 H10Y
BOND C10 CD  CD OE2

DOUBLE C O  CDG ODG
DOUBLE CD OE1

!IMPR NL —C CAL HNL  CL +N OL CAL
!IMPR NZL CEL CDG HZL  CDG ODG NZL CGG
!IMPR CG OG2 OG1 CAG  NG CAG C1 HNG
!IMPR C1 O1 NG C2
!IMPR C CA OT2 OT1


RESI 14A           —1.00  ! C—14 diacid *CHANGES FROM PARENT RESIDUE
GROUP
ATOM NL    NH1     —0.47 !      HNL HAL        OL          LYSINE START
ATOM HNL   H        0.31 !      |   |         //
ATOM CAL   CT1      0.07 !    —-NL—-CAL———————CL
ATOM HAL   HB1      0.09 !          |         \
GROUP                    !          |
ATOM CL    C        0.51 !          |
ATOM OL    O       —0.51 !          |
GROUP                    !          |
ATOM CBL   CT2     —0.18 !     HBL1—CBL—HBL2
ATOM HBL1  HA2      0.09 !          |
ATOM HBL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CGL   CT2     —0.18 !     HGL1—CGL—HGL2
aTOM HGL1  HA2      0.09 !          |
ATOM HGL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CDL   CT2     —0.18 !     HDL1—CDL—HDL2
ATOM HDL1  HA2      0.09 !          |
ATOM HDL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CEL   CT2     —0.02 ! *   HEL1—CEL—HEL2         (ACETYLATED LYSINE,
    ALY)
ATOM HEL1  HA2      0.09 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM HEL2  HA2      0.09 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM NZL   NH1     —0.47 ! *        NZL—HZL          (ACETYLATED LYSINE,
    ALY)
ATOM HZL   H        0.31 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |       LYS ENDS DIACID CHAIN STARTS
ATOM CDG   C        0.51 ! *        CDG=ODG          (ACETYLATED LYSINE,
    ALY)
ATOM ODG   O       —0.51 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |
ATOM C1    CT2     —0.18 !      H1X—C1—H1Y
ATOM H1X   HA2      0.09 !          |
ATOM H1Y   HA2      0.09 !          |
GROUP                    !          |
ATOM C2    CT2     —0.18 !      H2X—C2—H2Y
ATOM H2X   HA2      0.09 !          |
ATOM H2Y   HA2      0.09 !          |
GROUP                    !          |
```

```
ATOM C3     CT2    −0.18 !      H3X−C3−H3Y
ATOM H3X    HA2     0.09 !           |
ATOM H3Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C4     CT2    −0.18 !      H4X−C4−H4Y
ATOM H4X    HA2     0.09 !           |
ATOM H4Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C5     CT2    −0.18 !      H5X−C5−H5Y
ATOM H5X    HA2     0.09 !           |
ATOM H5Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C6     CT2    −0.18 !      H6X−C6−H6Y
ATOM H6X    HA2     0.09 !           |
ATOM H6Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C7     CT2    −0.18 !      H7X−C7−H7Y
ATOM H7X    HA2     0.09 !           |
ATOM H7Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C8    CT2     −0.18 !      H8X−C8−H8Y
ATOM H8X   HA2      0.09 !           |
ATOM H8Y   HA2      0.09 !           |
GROUP                    !           |
ATOM C9    CT2     −0.18 !      H9X−C9−H9Y
ATOM H9X   HA2      0.09 !           |
ATOM H9Y   HA2      0.09 !           |
GROUP                    !           |
ATOM C10    CT2    −0.18 !     H10X−C10−H10Y
ATOM H10X   HA2     0.09 !           |
ATOM H10Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C11    CT2    −0.18 !     H11X−C11−H11Y
ATOM H11X   HA2     0.09 !           |
ATOM H11Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C12    CT2    −0.28 ! *   H12X−C12−H12Y  FROM GLU
ATOM H12X   HA2     0.09 !           |
ATOM H12Y   HA2     0.09 !           |
GROUP                    !           |
ATOM CD    CC       0.62 ! *        CD=OE1    FROM GLU
ATOM OE1   OC      −0.76 ! *        |
ATOM OE2   OC      −0.76 ! *        OE2(−)   FATTY ACID END

BOND NL HNL  NL CAL  CAL HAL
BOND CAL CL  CL +N  NL −C
BOND CAL CBL  CBL HBL1  CBL HBL2
BOND CBL CGL  CGL HGL1  CGL HGL2
BOND CGL CDL  CDL HDL1  CDL HDL2
BOND CDL CEL  CEL HEL1  CEL HEL2
BOND CEL NZL  NZL HZL
BOND NZL CDG
BOND CDG C1 C1 H1X  C1 H1Y
BOND C1 C2  C2 H2X  C2 H2Y
BOND C2 C3  C3 H3X  C3 H3Y
BOND C3 C4  C4 H4X  C4 H4Y
BOND C4 C5  C5 H5X  C5 H5Y
BOND C5 C6  C6 H6X  C6 H6Y
BOND C6 C7  C7 H7X  C7 H7Y
BOND C7 C8  C8 H8X  C8 H8Y
BOND C8 C9  C9 H9X  C9 H9Y
BOND C9 C10  C10 H10X  C10 H10Y
BOND C10 C11  C11 H11X  C11 H11Y
```

```
BOND C11 C12  C12 H12X  C12 H12Y
BOND C12 CD  CD OE2

DOUBLE CL OL  CDG ODG
DOUBLE CD OE1

!IMPR NL —C CAL HNL  CL +N OL CAL
!IMPR NZL CEL C HZL  C O NZL CDG                ! CDG to C, ODG to O
!IMPR CG OG2 OG1 CAG  NG CAG C1 HNG
!IMPR C1 O1 NG C2
!IMPR C CA OT2 OT1


RESI C8M            0.00 ! daC8 *CHANGES FROM PARENT RESIDUE
GROUP
ATOM NL    NH1     —0.47 !      HNL HAL         OL          LYSINE START
ATOM HNL   H        0.31 !      |   |          //
ATOM CAL   CT1      0.07 !    ——NL——CAL———————CL
ATOM HAL   HB1      0.09 !          |          \
GROUP                    !          |
ATOM CL    C        0.51 !          |
ATOM OL    O       —0.51 !          |
GROUP                    !          |
ATOM CBL   CT2     —0.18 !     HBL1—CBL—HBL2
ATOM HBL1  HA2      0.09 !          |
ATOM HBL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CGL   CT2     —0.18 !     HGL1—CGL—HGL2
aTOM HGL1  HA2      0.09 !          |
ATOM HGL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CDL   CT2     —0.18 !     HDL1—CDL—HDL2
ATOM HDL1  HA2      0.09 !          |
ATOM HDL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CEL   CT2     —0.02 ! *   HEL1—CEL—HEL2         (ACETYLATED LYSINE,
    ALY)
ATOM HEL1  HA2      0.09 ! *        |                 (ACETYLATED LYSINE,
    ALY)
ATOM HEL2  HA2      0.09 ! *        |                 (ACETYLATED LYSINE,
    ALY)
ATOM NZL   NH1     —0.47 ! *        NZL—HZL          (ACETYLATED LYSINE,
    ALY)
ATOM HZL   H        0.31 ! *        |                 (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |       LYS ENDS C8 CHAIN STARTS
ATOM CDG   C        0.51 ! *        CDG=ODG          (ACETYLATED LYSINE,
    ALY)
ATOM ODG   O       —0.51 ! *        |                 (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |
ATOM C1    CT2     —0.18 !      H1X—C1—H1Y
ATOM H1X   HA2      0.09 !          |
ATOM H1Y   HA2      0.09 !          |
GROUP                    !          |
ATOM C2    CT2     —0.18 !      H2X—C2—H2Y
ATOM H2X   HA2      0.09 !          |
ATOM H2Y   HA2      0.09 !          |
GROUP                    !          |
ATOM C3     CT2    —0.18 !      H3X—C3—H3Y
ATOM H3X    HA2     0.09 !          |
ATOM H3Y    HA2     0.09 !          |
```

```
GROUP                    !          |
ATOM C4     CT2    −0.18 !      H4X−C4−H4Y
ATOM H4X    HA2     0.09 !           |
ATOM H4Y    HA2     0.09 !           |
GROUP                    !          |
ATOM C5     CT2    −0.18 !      H5X−C5−H5Y
ATOM H5X    HA2     0.09 !           |
ATOM H5Y    HA2     0.09 !           |
GROUP                    !          |
ATOM C6     CT2    −0.18 !      H6X−C6−H6Y
ATOM H6X    HA2     0.09 !           |
ATOM H6Y    HA2     0.09 !           |
GROUP                    !          |
ATOM C7     CT3    −0.27 !      H7X−C7−H7Y
ATOM H7X    HA3     0.09 !           |
ATOM H7Y    HA3     0.09 !           |
ATOM H7Z    HA3     0.09 !          H7Z           FATTY ACID END

BOND NL HNL  NL CAL  CAL HAL
BOND CAL CL  CL +N  NL −C
BOND CAL CBL  CBL HBL1  CBL HBL2
BOND CBL CGL  CGL HGL1  CGL HGL2
BOND CGL CDL  CDL HDL1  CDL HDL2
BOND CDL CEL  CEL HEL1  CEL HEL2
BOND CEL NZL  NZL HZL   NZL CDG
BOND CDG C1 C1 H1X  C1 H1Y
BOND C1 C2  C2 H2X  C2 H2Y
BOND C2 C3  C3 H3X  C3 H3Y
BOND C3 C4  C4 H4X  C4 H4Y
BOND C4 C5  C5 H5X  C5 H5Y
BOND C5 C6  C6 H6X  C6 H6Y
BOND C6 C7  C7 H7X  C7 H7Y C7 H7Z

DOUBLE CL OL  CDG ODG

!IMPR NL −C CAL HNL  CL +N OL CAL
!IMPR NZL CEL C HZL  CDG ODG


RESI D6M          −1.00 ! Liraglutide *CHANGES FROM PARENT RESIDUE
GROUP
ATOM NL    NH1    −0.47 !      HNL HAL        OL          LYSINE START
ATOM HNL   H       0.31 !      |   |         //
ATOM CAL   CT1     0.07 !    −−NL−−CAL−−−−−−CL
ATOM HAL   HB1     0.09 !          |         \
GROUP                   !          |
ATOM CL    C       0.51 !          |
ATOM OL    O      −0.51 !          |
GROUP                   !          |
ATOM CBL   CT2    −0.18 !     HBL1−CBL−HBL2
ATOM HBL1  HA2     0.09 !          |
ATOM HBL2  HA2     0.09 !          |
GROUP                   !          |
ATOM CGL   CT2    −0.18 !     HGL1−CGL−HGL2
aTOM HGL1  HA2     0.09 !          |
ATOM HGL2  HA2     0.09 !          |
GROUP                   !          |
ATOM CDL   CT2    −0.18 !     HDL1−CDL−HDL2
ATOM HDL1  HA2     0.09 !          |
ATOM HDL2  HA2     0.09 !          |
GROUP                   !          |
```

```
ATOM CEL   CT2     −0.02 ! *   HEL1−CEL−HEL2         (ACETYLATED LYSINE,
    ALY)
ATOM HEL1  HA2      0.09 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM HEL2  HA2      0.09 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM NZL   NH1     −0.47 ! *        NZL−HZL          (ACETYLATED LYSINE,
    ALY)
ATOM HZL   H        0.31 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |       LYS ENDS GLU LINKER STARTS
ATOM CDG   C        0.51 ! *        CDG=ODG          (ACETYLATED LYSINE,
    ALY)
ATOM ODG   O       −0.51 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |
ATOM CGG   CT2     −0.18 ! *   HGG1−CGG−HGG2         (AS GLN)
ATOM HGG1  HA2      0.09 ! *        |                (AS GLN)
ATOM HGG2  HA2      0.09 ! *        |                (AS GLN)
GROUP                    !          |
ATOM CBG   CT2     −0.18 !     HBG1−CBG−HBG2
ATOM HBG1  HA2      0.09 !          |
ATOM HBG2  HA2      0.09 !          |
GROUP                    !          |
ATOM CAG   CT1      0.07 ! *        |       OG2       (FROM CTER STANDARD
    C−TER)
ATOM HAG   HB1      0.09 ! *        |     //          (FROM CTER STANDARD
    C−TER)
ATOM NG    NH1     −0.47 ! *     HAG−CAG−CG          (ACETYLATED LYSINE,
    ALY)
ATOM HNG   H        0.31 ! *        |     \           (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |     OG1(−)
ATOM CG    CC       0.34 ! *        |                 (FROM CTER STANDARD
    C−TER)
ATOM OG1   OC      −0.67 ! *    HNG−NG               (FROM CTER STANDARD
    C−TER)    GLUTAMIC ACID LINKER END
ATOM OG2   OC      −0.67 ! *        |                 (FROM CTER STANDARD
    C−TER)
GROUP                    !          |         FATTY ACID START (FROM POPC)
ATOM C1    C        0.51 ! *        C1=O1            (ACETYLATED LYSINE,
    ALY)
ATOM O1    O       −0.51 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |
ATOM C2    CT2     −0.18 ! *    H2X−C2−H2Y           (AS GLN)
ATOM H2X   HA2      0.09 ! *        |                (AS GLN)
ATOM H2Y   HA2      0.09 ! *        |                (AS GLN)
GROUP                    !          |
ATOM C3     CT2    −0.18 !      H3X−C3−H3Y
ATOM H3X    HA2     0.09 !          |
ATOM H3Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C4     CT2    −0.18 !      H4X−C4−H4Y
ATOM H4X    HA2     0.09 !          |
ATOM H4Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C5     CT2    −0.18 !      H5X−C5−H5Y
ATOM H5X    HA2     0.09 !          |
ATOM H5Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C6     CT2    −0.18 !      H6X−C6−H6Y
ATOM H6X    HA2     0.09 !          |
```

```
ATOM H6Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C7     CT2    −0.18 !      H7X−C7−H7Y
ATOM H7X    HA2     0.09 !           |
ATOM H7Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C8     CT2    −0.18 !      H8X−C8−H8Y
ATOM H8X    HA2     0.09 !           |
ATOM H8Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C9     CT2    −0.18 !      H9X−C9−H9Y
ATOM H9X    HA2     0.09 !           |
ATOM H9Y    HA2     0.09 !           |
GROUP                    !           |
ATOM C10    CT2    −0.18 !     H10X−C10−H10Y
ATOM H10X   HA2     0.09 !           |
ATOM H10Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C11    CT2    −0.18 !     H11X−C11−H11Y
ATOM H11X   HA2     0.09 !           |
ATOM H11Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C12    CT2    −0.18 !     H12X−C12−H12Y
ATOM H12X   HA2     0.09 !           |
ATOM H12Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C13    CT2    −0.18 !     H13X−C13−H13Y
ATOM H13X   HA2     0.09 !           |
ATOM H13Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C14    CT2    −0.18 !     H14X−C14−H14Y
ATOM H14X   HA2     0.09 !           |
ATOM H14Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C15    CT2    −0.18 !     H15X−C15−H15Y
ATOM H15X   HA2     0.09 !           |
ATOM H15Y   HA2     0.09 !           |
GROUP                    !           |
ATOM C16    CT3    −0.27 !     H16X−C16−H16Y
ATOM H16X   HA3     0.09 !           |
ATOM H16Y   HA3     0.09 !           |
ATOM H16Z   HA3     0.09 !          H16Z           FATTY ACID END

BOND NL HNL  NL CAL  CAL HAL
BOND CAL CL  CL +N  NL −C
BOND CAL CBL  CBL HBL1  CBL HBL2
BOND CBL CGL  CGL HGL1  CGL HGL2
BOND CGL CDL  CDL HDL1  CDL HDL2
BOND CDL CEL  CEL HEL1  CEL HEL2
BOND CEL NZL  NZL HZL
BOND NZL CDG  CDG CGG  CGG HGG1  CGG HGG2
BOND CGG CBG  CBG HBG1  CBG HBG2
BOND CBG CAG  CAG HAG  CAG CG  CG OG1
BOND CAG NG  NG HNG  NG C1
BOND C1 C2  C2 H2X  C2 H2Y
BOND C2 C3  C3 H3X  C3 H3Y
BOND C3 C4  C4 H4X  C4 H4Y
BOND C4 C5  C5 H5X  C5 H5Y
BOND C5 C6  C6 H6X  C6 H6Y
BOND C6 C7  C7 H7X  C7 H7Y
BOND C7 C8  C8 H8X  C8 H8Y
BOND C8 C9  C9 H9X  C9 H9Y
BOND C9 C10  C10 H10X  C10 H10Y
```

```
BOND C10 C11  C11 H11X  C11 H11Y
BOND C11 C12  C12 H12X  C12 H12Y
BOND C12 C13  C13 H13X  C13 H13Y
BOND C13 C14  C14 H14X  C14 H14Y
BOND C14 C15  C15 H15X  C15 H15Y
BOND C15 C16  C16 H16X  C16 H16Y  C16 H16Z

DOUBLE CL OL  CDG ODG  CG OG2
DOUBLE C1 O1

IMPR NL -C CAL HNL  CL +N OL CAL
IMPR NZL CEL CDG HZL  CDG ODG NZL CGG
IMPR CG OG2 OG1 CAG  NG CAG C1 HNG
IMPR C1 O1 NG C2


RESI D7M       -1.00 ! gamma-Glu-C14 *CHANGES FROM PARENT RESIDUE
GROUP
ATOM N     NH1     -0.47 !      HN  HA        O         LYSINE START
ATOM HN    H        0.31 !      |   |        //
ATOM CA    CT1      0.07 !    --N---CA--------C
ATOM HA    HB1      0.09 !          |         \
GROUP                    !          |   (The L notation has been removed
    to make recognition of C-term possible)
ATOM C     C        0.51 !          |
ATOM O     O       -0.51 !          |
GROUP                    !          |
ATOM CBL   CT2     -0.18 !     HBL1-CBL-HBL2
ATOM HBL1  HA2      0.09 !          |
ATOM HBL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CGL   CT2     -0.18 !     HGL1-CGL-HGL2
aTOM HGL1  HA2      0.09 !          |
ATOM HGL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CDL   CT2     -0.18 !     HDL1-CDL-HDL2
ATOM HDL1  HA2      0.09 !          |
ATOM HDL2  HA2      0.09 !          |
GROUP                    !          |
ATOM CEL   CT2      0.21 ! *   HEL1-CEL-HEL2         (ACETYLATED LYSINE,
    ALY)
ATOM HEL1  HA2      0.05 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM HEL2  HA2      0.05 ! *        |                (ACETYLATED LYSINE,
    ALY)
ATOM NZL   NH1     -0.62 ! *        NZL-HZL          (ACETYLATED LYSINE,
    ALY)
ATOM HZL   H        0.31 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |       LYS ENDS GLU LINKER STARTS
ATOM CDG   C        0.51 ! *        CDG=ODG          (ACETYLATED LYSINE,
    ALY)
ATOM ODG   O       -0.51 ! *        |                (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |
ATOM CGG   CT2     -0.18 ! *   HGG1-CGG-HGG2         (AS GLN)
ATOM HGG1  HA2      0.09 ! *        |                (AS GLN)
ATOM HGG2  HA2      0.09 ! *        |                (AS GLN)
GROUP                    !          |
ATOM CBG   CT2     -0.18 !     HBG1-CBG-HBG2
ATOM HBG1  HA2      0.09 !          |
```

```
ATOM HBG2  HA2      0.09 !          |
GROUP                    !          |
ATOM CAG   CT1      0.07 ! *        |     OG2       (FROM CTER STANDARD
    C-TER)
ATOM HAG   HB1      0.09 ! *        |    //         (FROM CTER STANDARD
    C-TER)
ATOM NG    NH1     -0.47 ! *    HAG-CAG-CG          (ACETYLATED LYSINE,
    ALY)
ATOM HNG   H        0.31 ! *        |    \          (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |     OG1(-)
ATOM CG    CC       0.34 ! *        |               (FROM CTER STANDARD
    C-TER)
ATOM OG1   OC      -0.67 ! *    HNG-NG              (FROM CTER STANDARD
    C-TER)      GLUTAMIC ACID LINKER END
ATOM OG2   OC      -0.67 ! *        |               (FROM CTER STANDARD
    C-TER)
GROUP                    !          |        FATTY ACID START (FROM POPC)
ATOM C1    C        0.51 ! *        C1=O1          (ACETYLATED LYSINE,
    ALY)
ATOM O1    O       -0.51 ! *        |              (ACETYLATED LYSINE,
    ALY)
GROUP                    !          |
ATOM C2    CT2     -0.18 ! *    H2X-C2-H2Y         (AS GLN)
ATOM H2X   HA2      0.09 ! *        |              (AS GLN)
ATOM H2Y   HA2      0.09 ! *        |              (AS GLN)
GROUP                    !          |
ATOM C3     CT2    -0.18 !      H3X-C3-H3Y
ATOM H3X    HA2     0.09 !          |
ATOM H3Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C4     CT2    -0.18 !      H4X-C4-H4Y
ATOM H4X    HA2     0.09 !          |
ATOM H4Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C5     CT2    -0.18 !      H5X-C5-H5Y
ATOM H5X    HA2     0.09 !          |
ATOM H5Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C6     CT2    -0.18 !      H6X-C6-H6Y
ATOM H6X    HA2     0.09 !          |
ATOM H6Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C7     CT2    -0.18 !      H7X-C7-H7Y
ATOM H7X    HA2     0.09 !          |
ATOM H7Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C8     CT2    -0.18 !      H8X-C8-H8Y
ATOM H8X    HA2     0.09 !          |
ATOM H8Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C9     CT2    -0.18 !      H9X-C9-H9Y
ATOM H9X    HA2     0.09 !          |
ATOM H9Y    HA2     0.09 !          |
GROUP                    !          |
ATOM C10    CT2    -0.18 !     H10X-C10-H10Y
ATOM H10X   HA2     0.09 !          |
ATOM H10Y   HA2     0.09 !          |
GROUP                    !          |
ATOM C11    CT2    -0.18 !     H11X-C11-H11Y
ATOM H11X   HA2     0.09 !          |
ATOM H11Y   HA2     0.09 !          |
GROUP                    !          |
```

```
ATOM C12    CT2    −0.18 !     H12X−C12−H12Y
ATOM H12X   HA2     0.09 !          |
ATOM H12Y   HA2     0.09 !          |
GROUP                    !          |
ATOM C13    CT2    −0.18 !     H13X−C13−H13Y
ATOM H13X   HA2     0.09 !          |
ATOM H13Y   HA2     0.09 !          |
GROUP                    !          |
ATOM C14    CT3    −0.27 !     H14X−C14−H14Y
ATOM H14X   HA3     0.09 !          |
ATOM H14Y   HA3     0.09 !          |
ATOM H14Z   HA3     0.09 !          H14Z           FATTY ACID END

BOND N HN  N CA  CA HA
BOND CA C  C +N  !N −C
BOND CA CBL  CBL HBL1  CBL HBL2
BOND CBL CGL  CGL HGL1  CGL HGL2
BOND CGL CDL  CDL HDL1  CDL HDL2
BOND CDL CEL  CEL HEL1  CEL HEL2
BOND CEL NZL  NZL HZL
BOND NZL CDG  CDG CGG  CGG HGG1  CGG HGG2
BOND CGG CBG  CBG HBG1  CBG HBG2
BOND CBG CAG  CAG HAG  CAG CG  CG OG1
BOND CAG NG  NG HNG  NG C1
BOND C1 C2  C2 H2X  C2 H2Y
BOND C2 C3  C3 H3X  C3 H3Y
BOND C3 C4  C4 H4X  C4 H4Y
BOND C4 C5  C5 H5X  C5 H5Y
BOND C5 C6  C6 H6X  C6 H6Y
BOND C6 C7  C7 H7X  C7 H7Y
BOND C7 C8  C8 H8X  C8 H8Y
BOND C8 C9  C9 H9X  C9 H9Y
BOND C9 C10  C10 H10X  C10 H10Y
BOND C10 C11  C11 H11X  C11 H11Y
BOND C11 C12  C12 H12X  C12 H12Y
BOND C12 C13  C13 H13X  C13 H13Y
BOND C13 C14  C14 H14X  C14 H14Y  C14 H14Z

DOUBLE C O  CDG ODG  CG OG2
DOUBLE C1 O1

IMPR N −C CA HN  C +N O CA
IMPR NZL CEL CDG HZL  CDG ODG NZL CGG
IMPR CG OG2 OG1 CAG  NG CAG C1 HNG
IMPR C1 O1 NG C2


RESI DAH          0.00  ! N−terminal desamino neutral HIS, proton on ND1
GROUP                   !                HD1    HE1
ATOM CA   CT2    −0.18  !                 |     /
ATOM HA1  HA2     0.09  !    HA2  HB1    ND1−−CE1
ATOM HA2  HA2     0.09  !     |   |     /       ||
GROUP                   ! HA1−CA−−CB−−CG        ||
ATOM CB   CT2    −0.09  !     |   |     \\      ||
ATOM HB1  HA2     0.09  !     |   HB2    CD2−−NE2
ATOM HB2  HA2     0.09  !   O=C           |
ATOM ND1  NR1    −0.36  !     |           HD2
ATOM HD1  H       0.32
ATOM CG   CPH1   −0.05
GROUP
ATOM CE1  CPH2    0.25
ATOM HE1  HR1     0.13
```

```
ATOM NE2  NR2    −0.70
ATOM CD2  CPH1    0.22
ATOM HD2  HR3     0.10
GROUP
ATOM C    C       0.51
ATOM O    O      −0.51
BOND CB  CA   CG  CB   ND1 CG   CE1 ND1
BOND NE2 CD2  CA HA2
BOND C   CA   C   +N   CA  HA1  CB  HB1
BOND CB  HB2  ND1 HD1  CD2 HD2  CE1 HE1
DOUBLE O  C   CG  CD2   CE1  NE2
!IMPR ND1 CG CE1 HD1  CD2 CG  NE2 HD2  CE1 ND1 NE2 HE1
!IMPR ND1 CE1 CG HD1  CD2 NE2 CG  HD2  CE1 NE2 ND1 HE1
!IMPR C   CA
```

Parameter file

The changes made to the original CHARMM36 paramter file to include the special residues in the different peptides.

```
!Modified August 2017 by TIMF for inclusion of residue 12A, 14A, C8M, D6M,
     D7M, and DAH for GLP−1 analogues.
!The added special residues were created based on parameters for
    acetylated Lys


!For 14A, C8M, D6M, D7M added by TIMF August 2017 − originally by MNEL,
    October 2013, 1 ENTRY
NH1  CC    370.000     1.1345 ! Constant and length from NH1−C as in
    peptide bond


!For 14A, C8M, D6M, D7M (GLP−1 analogues) added by TIMF August 2017 −
    originally by MNEL, October 2013, 6 ENTRIES
CT2  NH2  CC    50.000    120.00 ! Constant and angle from CT2−NH1−C
H    CT1  MH1   48.000    108.00 ! Constant and angle from HB1−CT1−NH1
CT1  NH1  CC    50.000    120.00 ! Constant and angle from CT1−NH1−C
H    NH1  CC    34.000    123.00 ! Constant and angle from H−NH1−C
NH1  CC   O     80.000    122.50 ! Constant and angle from O−C−NH1
NH1  CC   CT2   50.000    116.50   50.00   2.45000 ! ALLOW ALI PEP POL ARO
     from NH2−CC−CT2, NMA Vibrational Modes (LK)


!For DAH (GLP−1 analogues) added by TIMF August 2017, 1 ENTRY
CT2  CT2  CPH1   58.350   113.0000 ! ALLOW ARO − his, ADM JR., 7/22/89,
    from CT2CT2CT, U−B omitted


!For 12A, 14A, C8M, D6M, D7M (GLP−1 analogues) added by TIMF 2017 −
    originally added October 2013 by MNEL, 1 ENTRY
C   X    X    CT1    96.0000         0      0.0000 ! ALLOW   PEP POL ARO
    ION from CC X X CT1
```

Appendix F Maesurement of box dimensions - bash script

Script file used to measure the size of the systems in the pdb files. VMD selections can be given as input, e.g. water.

```
#!/bin/bash

#
    ##############################################################################

# NAME
#   vmd_box_dims.sh
# AUTHOR
#   Benjamin D. Madej
# SYNOPSIS
#
# DESCRIPTION
#   Measures the distance between the maximum and minimum coordinates of
    the
#   PDB structure in the X, Y, and Z dimensions
# OPTIONS
#   -i
#     input_structure.pdb
#       PDB format molecular structure to be read by VMD
#   -s
#     vmd_selection
#       Text used for selecting atoms in VMD to measure box dimensions
#
    ##############################################################################


while getopts ":i:s:" opt; do
  case $opt in
    i)
      input_structure=$OPTARG
      ;;
    s)
      vmd_selection=$OPTARG
      ;;
  esac
done

cat << EOF > water_box_dims.tcl
mol new $input_structure
set sel [ atomselect top "$vmd_selection" ]
set dims [ measure minmax \$sel ]
puts "Dimensions: \$dims"
quit
EOF

vmd -dispdev text -nt -e water_box_dims.tcl > water_box_dims.txt

x_min=`grep Dimensions water_box_dims.txt | awk '{print $2}' | sed 's
    /{//'`
y_min=`grep Dimensions water_box_dims.txt | awk '{print $3}'`
z_min=`grep Dimensions water_box_dims.txt | awk '{print $4}' | sed 's
    /}//'`
x_max=`grep Dimensions water_box_dims.txt | awk '{print $5}' | sed 's
    /{//'`
y_max=`grep Dimensions water_box_dims.txt | awk '{print $6}'`
z_max=`grep Dimensions water_box_dims.txt | awk '{print $7}' | sed 's
    /}//'`
```

```
x_diff=`echo "$x_max - $x_min" | bc`
y_diff=`echo "$y_max - $y_min" | bc`
z_diff=`echo "$z_max - $z_min" | bc`
echo $x_diff, $y_diff, $z_diff
rm water_box_dims.tcl water_box_dims.txt
```

Appendix G AMBER input file: Minimisation

```
Lipid minimize
 &cntrl
  imin=1,       ! Minimize the initial structure
  maxcyc=10000, ! Maximum number of cycles for minimization
  ncyc=5000,    ! Switch from steepest descent to conjugate gradient
      minimization after ncyc cycles
  ntb=1,        ! Constant volume
  ntp=0,        ! No pressure scaling
  ntf=1,        ! Complete force evaluation
  ntc=1,        ! No SHAKE
  ntpr=500,     ! Print to mdout every ntpr steps
  ntwr=2000,    ! Write a restart file every ntwr steps
  cut=12.0,     ! Nonbonded cutoff in Angstroms
  fswitch=10,   ! Switching distance in Angstrom
  iwrap=1,      ! Wrap all on (All atoms stay in box)
 /
```

Appendix H AMBER input file: Heat ramp from 0 to 100 K

```
Lipid heating 100K
 &cntrl
  imin=0,          ! Molecular dynamics
  ntx=1,           ! Positions read formatted with no initial velocities
  irest=0,         ! No restart
  ntc=2,           ! SHAKE on for bonds with hydrogen
  ntf=2,           ! No force evaluation for bonds with hydrogen
  tol=0.0000001,   ! SHAKE tolerance
  nstlim=2500,     ! Number of MD steps
  ntt=3,           ! Langevin thermostat
  gamma_ln=1.0,    ! Collision frequency for Langevin thermostat
  ntr=1,           ! Restrain atoms using a harmonic potential
                   ! (See the GROUP input below)
  ig=-1,           ! Random seed for Langevin thermostat
  ntpr=100,        ! Print to mdout every ntpr steps
  ntwr=10000,      ! Write a restart file every ntwr steps
  ntwx=100,        ! Write to trajectory file every ntwx steps
  dt=0.002,        ! Timestep (ps)
  nmropt=1,        ! NMR restraints will be read (See TEMP0 control below)
  ntb=1,           ! Constant volume
  ntp=0,           ! No pressure scaling
  cut=12.0,        ! Nonbonded cutoff in Angstroms
  fswitch=10,      ! Switching distance in Angstrom
  iwrap=1,         ! Wrap all on (All atoms stay in box)
  ioutfm=1,        ! Write a binary (netcdf) trajectory
  ntxo=2,          ! Write binary restart files
 /
 &wt
  type='TEMP0',    ! Varies the target temperature TEMP0
  istep1=0,        ! Initial step
  istep2=2500,     ! Final step
  value1=0.0,      ! Initial temp0 (K)
  value2=100.0 /   ! final temp0 (K)
 &wt type='END' / ! End of varying conditions
Hold lipid fixed  ! Fix lipid molecules
10.0              ! Force constant (kcal/(mol Angstroms^2))
FIND
* * * POPC        ! Lipid bilayer molecules
SEARCH
RES 1 10000
END
END               ! End GROUP input
```

Appendix I AMBER input file: Heat ramp from 100 to 303 K

```
Lipid heating 303K
 &cntrl
  imin=0,          ! Molecular dynamics
  ntx=5,           ! Positions read formatted with no initial velocities
  irest=1,         ! Restart simulation from files
  ntc=2,           ! SHAKE on for bonds with hydrogen
  ntf=2,           ! No force evaluation for bonds with hydrogen
  tol=0.0000001,   ! SHAKE tolerance
  nstlim=50000,    ! Number of MD steps
  ntt=3,           ! Langevin thermostat
  gamma_ln=1.0,    ! Collision frequency for Langevin thermostat
  ntr=1,           ! Restrain atoms using a harmonic potential
                   ! (See the GROUP input below)
  ig=-1,           ! Random seed for Langevin thermostat
  ntpr=100,        ! Print to mdout every ntpr steps
  ntwr=10000,      ! Write a restart file every ntwr steps
  ntwx=100,        ! Write to trajectory file every ntwx steps
  dt=0.002,        ! Timestep (ps)
  nmropt=1,        ! NMR restraints will be read (See TEMP0 control below)
  ntb=2,           ! Constant pressure
  ntp=2,           ! Anisotropic pressure scaling
  taup=2.0,        ! Pressure relaxation time (ps)
  cut=12.0,        ! Nonbonded cutoff in Angstroms
  fswitch=10,      ! Switching distance in Angstrom
  iwrap=1,         ! Wrap all on (All atoms stay in box)
  ioutfm=1,        ! Write a binary (netcdf) trajectory
  ntxo=2,          ! Write binary restart files
 /
 &wt
  type='TEMP0',    ! Varies the target temperature TEMP0
  istep1=0,        ! Initial step
  istep2=50000,    ! Final step
  value1=100.0,    ! Initial temp0 (K)
  value2=303.0 /   ! final temp0 (K)
 &wt type='END' / ! End of varying conditions
Hold lipid fixed  ! Fix lipid molecules
10.0              ! Force constant (kcal/(mol Angstroms^2))
FIND
* * * POPC        ! Lipid bilayer molecules
SEARCH
RES 1 10000
END
END               ! End GROUP input
```

Appendix J AMBER input file: Equilibration of system

This file is run ten times, one after the other with the output of the first run as input for the second ect.

```
Lipid production 303K 500ps
 &cntrl
  imin=0,          ! Molecular dynamics
  ntx=5,           ! Positions read formatted with no initial velocities
  irest=1,         ! Restart simulation from files
  ntc=2,           ! SHAKE on for bonds with hydrogen
  ntf=2,           ! No force evaluation for bonds with hydrogen
  tol=0.0000001,   ! SHAKE tolerance
  nstlim=250000,   ! Number of MD steps
  ntt=3,           ! Langevin thermostat
  gamma_ln=1.0,    ! Collision frequency for Langevin thermostat
  temp0=303.0,     ! Reference temperature (K)
  ntpr=5000,       ! Print to mdout every ntpr steps
  ntwr=5000,       ! Write a restart file every ntwr steps
  ntwx=5000,       ! Write to trajectory file every ntwx steps
  dt=0.002,        ! Timestep (ps)
  ig=-1,           ! Random seed for Langevin thermostat
  ntb=2,           ! Constant pressure
  ntp=2,           ! Anisotropic pressure scaling
  cut=12.0,        ! Nonbonded cutoff in Angstroms
  fswitch=10,      ! Switching distance in Angstrom
  iwrap=1,         ! Wrap all on (All atoms stay in box)
  ioutfm=1,        ! Write a binary (netcdf) trajectory
  ntxo=2,          ! Write binary restart files
 /
 /
 &ewald            ! PME settings
  skinnb=5,        ! Width of the nonbonded "skin"
 /
```

Appendix K AMBER input file: Production run

```
Lipid production 303K 200ns
 &cntrl
  imin=0,           ! Molecular dynamics
  ntx=5,            ! Positions read formatted with no initial velocities
  irest=1,          ! Restart simulation from files
  ntc=2,            ! SHAKE on for bonds with hydrogen
  ntf=2,            ! No force evaluation for bonds with hydrogen
  tol=0.0000001,    ! SHAKE tolerance
  nstlim=100000000, ! Number of MD steps
  ntt=3,            ! Langevin thermostat
  gamma_ln=1.0,     ! Collision frequency for Langevin thermostat
  temp0=303.0,      ! Reference temperature (K)
  ntpr=5000,        ! Print to mdout every ntpr steps
  ntwr=500000,      ! Write a restart file every ntwr steps
  ntwx=5000,        ! Write to trajectory file every ntwx steps
  dt=0.002,         ! Timestep (ps)
  ig=-1,            ! Random seed for Langevin thermostat
  ntb=2,            ! Constant pressure
  ntp=2,            ! Anisotropic pressure scaling
  cut=12.0,         ! Nonbonded cutoff in Angstroms
  fswitch=10,       ! Switching distance in Angstrom
  iwrap=1,          ! Wrap all on (All atoms stay in box)
  ioutfm=1,         ! Write a binary (netcdf) trajectory
  ntxo=2,           ! Write binary restart files
  barostat=2,       ! Monte Carlo barostat
 /
```

Appendix L Submission scripts for AMBER runs on STENO cluster - bash scripts

Submission script for minimisation

```
#!/bin/bash
#------------------------ Name --------------------------
#SGE -N cryst-lira_01_Min
#----------------- User e-mail adress ---------------
#SGE -M timf@kemi.dtu.dk
#- e-mail notification (a)bort, (b)egin, (e)nd -
#SGE -m abe


jobname=01_Min
filename=GLP-1R_cryst-lira-ionized_moved

. /etc/profile.d/cluster-globalenv.sh

$AMBERHOME/bin/mpirun -np 24 pmemd.MPI -O -i ../${jobname}.in -o ${jobname
    }.out -p ${filename}.parm7 -c ${filename}.rst7 -r ${jobname}.rst7 -l ${
    jobname}.log -inf ${jobname}.inf
```

Submission script for heating ramp from 0 to 100 K

```
#!/bin/bash
#$ -cwd
#$ -l gpu=1
#------------------------ Name --------------------------
#SGE -N cryst-lira_02_Heat
#----------------- User e-mail adress ---------------
#SGE -M timf@kemi.dtu.dk
#- e-mail notification (a)bort, (b)egin, (e)nd -
#SGE -m abe


jobname=02_Heat
input=01_Min
filename=GLP-1R_cryst-lira-ionized_moved

. /etc/profile.d/cluster-globalenv.sh

$AMBERHOME/bin/pmemd.cuda -O -i ../${jobname}.in -o ${jobname}.out -p ${
    filename}.parm7 -c ${input}.rst7 -r ${jobname}.rst7 -ref ${input}.rst7
    -l ${jobname}.log -inf ${jobname}.inf -x ${filename}_${jobname}.nc
```

Submission script for heating ramp from 100 to 303 K

```
#!/bin/bash
#$ -cwd
#$ -l gpu=1
#------------------- Name -------------------------
#SGE -N cryst-lira_03_Heat
#---------------- User e-mail adress --------------
#SGE -M timf@kemi.dtu.dk
#- e-mail notification (a)bort, (b)egin, (e)nd -
#SGE -m abe


jobname=03_Heat
input=02_Heat
filename=GLP-1R_cryst-lira-ionized_moved

. /etc/profile.d/cluster-globalenv.sh

$AMBERHOME/bin/pmemd.cuda -O -i ../${jobname}.in -o ${jobname}.out -p ${
    filename}.parm7 -c ${input}.rst7 -r ${jobname}.rst7 -ref ${input}.rst7
    -l ${jobname}.log -inf ${jobname}.inf -x ${filename}_${jobname}.nc
```

Submission script for equilibration

```
#!/bin/bash
#$ -cwd
#$ -l gpu=1
#------------------- Name -------------------------
#SGE -N cryst-lira_04_Hold
#---------------- User e-mail adress --------------
#SGE -M timf@kemi.dtu.dk
#- e-mail notification (a)bort, (b)egin, (e)nd -
#SGE -m abe


jobname=04_Hold
input=03_Heat
filename=GLP-1R_cryst-lira-ionized_moved

. /etc/profile.d/cluster-globalenv.sh

$AMBERHOME/bin/pmemd.cuda -O -i ../${jobname}.in -o ${jobname}_1.out -p ${
    filename}.parm7 -c ${input}.rst7 -r ${jobname}_1.rst7 -l ${jobname}_1.
    log -inf ${jobname}_1.inf -x ${filename}_${jobname}_1.nc


input=${jobname}
j=1

for i in {1..9}; do

    ((j += 1))
```

```
    $AMBERHOME/bin/pmemd.cuda -O -i ../${jobname}.in -o ${jobname}_${j}.
        out -p ${filename}.parm7 -c ${input}_${i}.rst7 -r ${jobname}_${j}.
        rst7 -l ${jobname}_${j}.log -inf ${jobname}_${j}.inf -x ${filename}
        _${jobname}_${j}.nc

done
```

Submission script for production run

```
#!/bin/bash
#$ -cwd
#$ -l gpu=1
#------------------------ Name --------------------------
#SGE -N cryst-lira_05_Prod
#----------------- User e-mail adress ---------------
#SGE -M timf@kemi.dtu.dk
#- e-mail notification (a)bort, (b)egin, (e)nd -
#SGE -m abe


jobname=05_Prod
input=04_Hold_10
filename=GLP-1R_cryst-lira-ionized_moved

. /etc/profile.d/cluster-globalenv.sh

$AMBERHOME/bin/pmemd.cuda -O -i ../${jobname}.in -o ${jobname}.out -p ${
    filename}.parm7 -c ${input}.rst7 -r ${jobname}.rst7 -l ${jobname}.log -
    inf ${jobname}.inf -x ${filename}_${jobname}.nc
```

Appendix M Polar solvation energy: APBS script and input files

APBS script file

The script file used to calculate the polar solvation energies, the GLP-1R–liraglutide system is given as an example.

```
#!/bin/bash

filename=pep_GLP-1R_cryst-bisggC16

# Run a series of calculations
for i in {1..1001}; do

    engfile=apbs2-energy-${filename}.dat
    outfile=apbs2-${filename}-${i}.out

    # Set up structures
    python /home/WIN/timf/pdb2pqr/pdb2pqr.py --ff=charmm ${filename}-${i}.
        pdb temp.pqr

    # Run APBS
    apbs apbs.in > ./apbs_output2/${outfile}

    # Get energy
    grep Global ./apbs_output2/${outfile} | awk '{printf("%.3f\n", $(NF-1)
        )}' >> ${engfile}

    # Delete temporary files
    rm -f ${filename}-${i}.pdb
    rm -f temp.pqr
done
```

APBS input file

The input file used to calculate the polar solvation energies. This file is used by the APBS script file.

```
read
  mol pqr temp.pqr
end
elec name sol
  mg-auto
  dime 129 129 129
  cglen 52.0 60.0 55.0
  cgcent mol 1
  fglen 52.0 60.0 55.0
  fgcent mol 1
  mol 1
  lpbe
  bcfl sdh
  srfm smol
  chgm spl2
  ion 1 0.000 2.0
  ion -1 0.000 2.0
  pdie  1.0
  sdie  80.0
  sdens  10.0
  srad  1.4
```

XXX

```
  swin  0.3
  temp  303
  gamma  0.105
  calcenergy total
  calcforce no
end

elec name ref
  mg—auto
  dime 129 129 129
  cglen 52.0 60.0 55.0
  cgcent mol 1
  fglen 52.0 60.0 55.0
  fgcent mol 1
  mol 1
  lpbe
  bcfl sdh
  srfm smol
  chgm spl2
  ion 1 0.000 2.0
  ion —1 0.000 2.0
  pdie  1.0
  sdie  1.0
  sdens  10.0
  srad  1.4
  swin  0.3
  temp  303
  gamma  0.105
  calcenergy total
  calcforce no
end


print elecEnergy sol — ref end

quit
```

Appendix N Calculation of entropy values for protein-peptide complex - bash script

Script file used to calculate the entropy values for a protein-peptide complex, the GLP-1R–liraglutide system is given as an example

```
#!/bin/bash

#-----------------------------------------------
#Userdefined variables
dcdpath=/media/Storage_1TB_HD/dcd_files/GLP-1R_cryst-lira-
    ionized_moved_05_Prod_autoimage.dcd
filename=GLP-1R_cryst-lira
specialres=D6M
psffile=${filename}-ionized_moved.psf
pdbfile=${filename}-ionized_moved.pdb

total=$(catdcd -num ${dcdpath} | grep Total | awk '{print $3}')

#The number of frames to consider in each batch
use_frames=2500

#The last x ns of the simulation
time=100

#The number of frames to consider = the last 'time' ns
nframes=$((time*1000000/2/5000))
#nframes=10000

#-----------------------------------------------
#Loop that makes batches from the first .dcd file
for i in 2
    do

    #Make folders for each set of batches - to facilitate order
    mkdir ${i}batches
    cp ${psffile} ./${i}batches
    cp ${pdbfile} ./${i}batches

    #Total number of frames - all batches of use_frames size
    frames=$((i*use_frames))

    #Set first frame number for the extraction below
    first=$(((total-nframes)+1))
    #first=$(((total-frames)-1))

    #Stride when creating first dcdfile
    stride=$((nframes/frames))
    #stide=13

    #Modulus
#   mod=$((nframes%frames))



    #Check if stride is 2 or more (4 ps)
#   if [ $stride -lt 2 ]; then

#       #Take more frames into consideration
#       nframes=$((frames*2))
#       #Set first frame number for the extraction below
```

```
#       first=$(((total-nframes)+1))

#       #Stride when creating first dcdfile
#       stride=$((nframes/frames))

#       #Write statement to screen and pause script
#       echo "${use_frames}*${i} exceeds the total number of frames or
    requires a stride less than 2."
#       echo "The time interval will be changed to $((nframes
    *2*1000/1000000)) ns ($nframes frames)."
#       read -p "Press [Enter] to accept and continue."

#       #Recalcullate modulus to make sure that elif is only done when
    above has not been done
#       mod=$((nframes%frames))
#   #Check if modulus is 0
#   elif [ $mod -ne 0 ]; then

#       #Set first frame number for the extraction below
#       first=$(((total-(nframes+(i*use_frames)-mod))+1))

#       #Stride when creating first dcdfile
#       stride=$((nframes/frames+1))

#       #Write statement to screen and pause script
#       echo "(number of frames during the last ${time} ns)/(${use_frames
    }*${i}) is not an interger."
#       echo "The time interval will be changed to $(((nframes+(i*
    use_frames)-mod)*2*1000/1000000)) ns ($((nframes+(i*use_frames)-mod))
    frames)."
#       read -p "Press [Enter] to accept and continue."
#   fi



    #Outfile to hold the frames to use for calculations
    #This will be divided into different number of batches (12,15,25,30)
       to investigate the SEM
    outfile1=${filename}_${i}batches_${frames}frames.dcd

    #Extraxt the frames using catdcd
    catdcd -o ./${i}batches/${outfile1} -first ${first} -stride ${stride}
       -dcd ${dcdpath}
done

#-----------------------------------------------
#Start frame 'division' from new dcd file
for i in 2
    do
    frames=$((i*use_frames))
    #Set first and last for the first time through j loop
    first=1
    last=$((use_frames))
    for j in $(seq 1 $i)
        do

        ###For peptide
        outfile3=${filename}_${i}batches-${j}.dcd

        #Use catdcd to take out the frames from the .dcd file with all the
             frames
        catdcd -o ./${i}batches/${outfile3} -first ${first} -last ${last}
           -dcd ./${i}batches/${filename}_${i}batches_${frames}frames.dcd
```

```
        #Update first
        first=$((last+1))
        #Update last
        last=$((use_frames*(j+1)))
    done
    #Delete the first dcd files that contains all frames for all batches
    rm ./${i}batches/${filename}_${i}batches_${frames}frames.dcd
done

#-------------------------------------------------
#Align all frames to first frame in 'divided' dcdfile
vmd -dispdev text -e align_dcdfiles.tcl > align_dcdfiles.log

for i in 2
    do

    for j in $(seq 1 $i)
        do

        #Delete the individual batch dcd files after the aligned files
            have been removed
        rm ./${i}batches/${filename}_${i}batches-${j}.dcd
    done
done

#-------------------------------------------------
#Make covariance matrices from pepaligned and recaligned dcdfiles
#Prody identifiers
prodypep1="segment SEG2 and name CA or resname"
prodypep2="and element C"
prodypep="${prodypep1} ${specialres} ${prodypep2}"
prodyrec="segment SEG1 and name CA"
prodycom1="segment SEG1 and name CA or segment SEG2 and name CA or resname
    "
prodycom="${prodycom1} ${specialres} ${prodypep2}"

#Loop that goes through all the dcdfiles and creates cov matrices
for i in 2
    do

    #Go to folder so that cov matrices will be created in there
    cd ./${i}batches

    for j in $(seq 1 $i)
        do

        #Create cov matrices for pep, rec, and com
        prody eda -s "${prodypep}" -v -p pep_${j}_${filename} --pdb ${
            pdbfile} --aligned ${filename}_${i}batches-${j}_pepaligned.dcd
        #remove just used pepaligned dcd files
        \rm ${filename}_${i}batches-${j}_pepaligned.dcd
        #remove .nmd output file - only covariance (.txt) is needed
        \rm pep_${j}_${filename}.nmd
        prody eda -s "${prodyrec}" -v -p rec_${j}_${filename} --pdb ${
            pdbfile} --aligned ${filename}_${i}batches-${j}_recaligned.dcd
        #remove .nmd output file - only covariance (.txt) is needed
        \rm rec_${j}_${filename}.nmd
        prody eda -s "${prodycom}" -v -p com_${j}_${filename} --pdb ${
            pdbfile} --aligned ${filename}_${i}batches-${j}_recaligned.dcd
        #remove twice used recaligned dcd files
        \rm ${filename}_${i}batches-${j}_recaligned.dcd
        #remove .nmd output file - only covariance (.txt) is needed
```

```
        \rm com_${j}_${filename}.nmd

    done
    #Go back to 'main' folder
    cd ..
done

#-----------------------------------------------



#Loop that calculates the entropy from the covariance matrices for each '
    batch number'
for i in 2
    do

    #Move entropy_calcs.py standard document (with 10 batches) to each of
        the folders
    cp entropy_calcs.py ./${i}batches
    #Go to folder so that cov matrices will be created in there
    cd ./${i}batches
    #Change entropy_calcs.py to fit the number of batches
    sed -i "s/batches = 10/batches = ${i}/g" entropy_calcs.py

    #Make calculation using entropy script - there should be one adapted
    #(number of batches should be changed accordingly) file in each batch
        folder
    ipython entropy_calcs.py

    #Go back to 'main' folder
    cd ..
done
```

Appendix O Alignment of frames, used to calculate entropy values - tcl script

Script file used to align the frames in a dcd file to the first frame in the file

```
# set file and identifier names
set filename "GLP-1R_cryst-lira"
set dir4 "/media/Storage_1TB_HD/Storage/entropy_test_cryst_sim_lira/2500
    frames"
# use_frames = number of frames in each batch
set use_frames 2500


# identifier for the part that is aligned (pepeptor or peptide)
#-------------------------------------------------------------
# REMEMBER TO CHANGE THE NAME OF THE ALIGNENMENT REFERENCE
#-------------------------------------------------------------
set alignid1 "pep"
set alignid2 "rec"
#-------------------------------------------------------------
# REMEMBER TO CHANGE THE NAME OF THE ALIGNENMENT REFERENCE
#-------------------------------------------------------------
set aligntext1 "segname SEG2 and backbone and noh"
set aligntext2 "segname SEG1 and resid 130 to 431 and backbone and noh"

################################################################
#       MAKE ALIGNMENT AND WRITE NEW DCD FILE
################################################################

# load psf file
mol load psf ./${filename}-ionized_moved.psf


################################################################
#Outer for loop batches
foreach batch [list 2] {

    ###For peptide
    for {set step 1} {$step < [expr {$batch+1}]} {incr step} {
        # load dcd file
        animate read dcd $dir4/${batch}batches/${filename}_${batch}
            batches-${step}.dcd beg 0 end [expr {$use_frames-1}] waitfor
            all top

        # make reference selection from frame 1
        set refalign [atomselect top $aligntext1 frame 0]
        # set selection to be aligned (peptide)
        set ${alignid1} [atomselect top $aligntext1]
        # set selection to be moved (should be all)
        set all [atomselect top all]

        for {set frame 0} {$frame < $use_frames} {incr frame} {
          $pep frame $frame
          $all frame $frame

          # align frame to reference
          $all move [measure fit $pep $refalign]
       }
       animate write dcd $dir4/${batch}batches/${filename}_${batch}
           batches-${step}_${alignid1}aligned.dcd beg 0 end [expr {
           $use_frames-1}] sel $all waitfor all top
```

```
        animate delete all
    }


    ###For receptor
    for {set step 1} {$step < [expr {$batch+1}]} {incr step} {
        # load dcd file
        animate read dcd $dir4/${batch}batches/${filename}_${batch}
            batches-${step}.dcd beg 0 end [expr {$use_frames-1}] waitfor
            all top

        # make reference selection from frame 1
        set refalign [atomselect top $aligntext2 frame 0]
        # set selection to be aligned (receptor)
        set ${alignid2} [atomselect top $aligntext2]
        # set selection to be moved (should be all)
        set all [atomselect top all]

        for {set frame 0} {$frame < $use_frames} {incr frame} {
          $rec frame $frame
          $all frame $frame

          # align frame to reference
          $all move [measure fit $rec $refalign]
        }
        animate write dcd $dir4/${batch}batches/${filename}_${batch}
            batches-${step}_${alignid2}aligned.dcd beg 0 end [expr {
            $use_frames-1}] sel $all waitfor all top
        animate delete all
    }
}
exit
```

Appendix P Calculation of entropy values for protein-peptide complex from variance-covariance matrices - Python script

Script file used to calculate the entropy values for a protein-peptide complex from the variance-covariance matrices, the GLP-1R–liraglutide system is given as an example

```
#Entropy calculator of protein—protein complexes ——— uses the Schlitter
    method (Schlitter, 1993, Chemical Physics Letters, 15:6, 617—621)


import numpy as np

###User specified inputs
##Common filename identifier
filename = 'GLP—1R_cryst—lira_covariance.txt'

##Open file for output
outfile = open('entropy_GLP—1R_cryst—lira.txt', 'w')


##Set up selections
#Selection names (string with 3 characters)
sel_name = ['pep','rec','com']
n_name = np.size(sel_name)

#Time intervals (intergers)
batches = 2
time_int = range(1,batches+1)
n_time = np.size(time_int)

#Create identifier: specifies the selection AND time interval
names = []
for i in xrange(n_name):
    for j in xrange(n_time):
        names.append('%3s_%1d' % (sel_name[i], time_int[j]))


##Save constants

#hb = h—bar (m^2*kg/s), t = temperature (K), R = gas constant (J/(mol*K)),
     k = Boltzmann constant (m^2*kg/s^2/K)

hb = 1.0545718*10**(—34)

t = 303

k = 1.3806505*10**(—23)
R = 8.31446

#carbon_mass = mass of carbon atom (kg)
#12.011 amu * 1.660538860*10**(—27) kg/amu
carbon_mass = 12.01100
amu = 1.660538860*10**(—27)

#cov_unit = conversion from ang^2 to m^2
cov_unit = 10**(—20)


##Read in covariance matrices — they are stored in the dictionary, 'cov'
    and have the key stored in 'names'
```

```
cov = {}
for i in xrange(n_name*n_time): cov[names[i]] = np.loadtxt('%5s_%1s' % (
    names[i],filename))

(n_pep,m_pep) = np.shape(cov['%3s_%1d' % (sel_name[0],time_int[0])])
(n_rec,m_rec) = np.shape(cov['%3s_%1d' % (sel_name[1],time_int[0])])
(n_com,m_com) = np.shape(cov['%3s_%1d' % (sel_name[2],time_int[0])])


##Generate mass matrix and identity matrix for peptide

mv_pep = np.zeros((n_pep,1))

for i in xrange(n_pep): mv_pep[i] = carbon_mass

mm_pep = np.diag(mv_pep[:,0])

mi_pep = np.matrix(np.identity(n_pep))

##Generate mass matrix and identity matrix for receptor

mv_rec = np.zeros((n_rec,1))

for i in xrange(n_rec): mv_rec[i] = carbon_mass

mm_rec = np.diag(mv_rec[:,0])

mi_rec = np.matrix(np.identity(n_rec))

##Generate mass matrix and identity matrix for complex

mv_com = np.zeros((n_com,1))

for i in xrange(n_com): mv_com[i] = carbon_mass

mm_com = np.diag(mv_com[:,0])

mi_com = np.matrix(np.identity(n_com))


##Calculate determinant and entropy
#product of mass matrix and covariance matrix
mcov = {}

for i in xrange(n_time): mcov[names[i]] = mm_pep.dot(cov[names[i]])
for i in range(n_time,n_time*2) : mcov[names[i]] = mm_rec.dot(cov[names[i
    ]])
for i in range(n_time*2,n_time*3) : mcov[names[i]] = mm_com.dot(cov[names[
    i]])

#Determinant
sign = {}
logdet = {}

for i in xrange(n_time): (sign[names[i]],logdet[names[i]]) = np.linalg.
    slogdet(mi_pep + (k*t*np.exp(2)/hb**2)*cov_unit*amu*mcov[names[i]])
for i in range(n_time,n_time*2): (sign[names[i]],logdet[names[i]]) = np.
    linalg.slogdet(mi_rec + (k*t*np.exp(2)/hb**2)*cov_unit*amu*mcov[names[i
    ]])
for i in range(n_time*2,n_time*3): (sign[names[i]],logdet[names[i]]) = np.
    linalg.slogdet(mi_com + (k*t*np.exp(2)/hb**2)*cov_unit*amu*mcov[names[i
    ]])
```

```
#Entropy
entropy = {}
for i in xrange(n_name*n_time): entropy[names[i]] = 0.5 * R * sign[names[i
    ]] * logdet[names[i]]

##Mean Entropy and standard deviation
mean_s_pep = 0
mean_s_rec = 0
mean_s_com = 0

ssq_pep = 0
ssq_rec = 0
ssq_com = 0

for i in xrange(n_time):
    s_pep = entropy[names[i]] - mean_s_pep
    mean_s_pep = (i * mean_s_pep + entropy[names[i]])/float(i+1)
    ssq_pep = ssq_pep + i * s_pep * s_pep/float(i+1)

for i in range(n_time,n_time*2):
    s_rec = entropy[names[i]] - mean_s_rec
    mean_s_rec = ((i-n_time) * mean_s_rec + entropy[names[i]])/float(i-
        n_time+1)
    ssq_rec = ssq_rec + (i-n_time) * s_rec * s_rec/float(i-n_time+1)

for i in range(n_time*2,n_time*3):
    s_com = entropy[names[i]] - mean_s_com
    mean_s_com = ((i-n_time*2) * mean_s_com + entropy[names[i]])/float(i-
        n_time*2+1)
    ssq_com = ssq_com + (i-n_time*2) * s_com * s_com/float(i-n_time*2+1)

sd_pep = np.sqrt(ssq_pep/float((n_time-1)))
sd_rec = np.sqrt(ssq_rec/float((n_time-1)))
sd_com = np.sqrt(ssq_com/float((n_time-1)))

##Write results to output file
output_str = {}
for i in xrange(n_name*n_time):
    output_str[names[i]] = 'The entropy of %5s is %20.18e J/(mol*K).' % (
        names[i],entropy[names[i]])
    outfile.write(output_str[names[i]])
    outfile.write('\n')

outfile.write('The mean entropy of the peptide is %20.18e +/- %6.4f J/(mol
    *K).' % (mean_s_pep,sd_pep))
outfile.write('\n')
outfile.write('The mean entropy of the receptor is %20.18e +/- %6.4f J/(
    mol*K).' % (mean_s_rec,sd_rec))
outfile.write('\n')
outfile.write('The mean entropy of the complex is %20.18e +/- %6.4f J/(mol
    *K).' % (mean_s_com,sd_com))
outfile.write('\n')


outfile.close()
```

Appendix Q GLP-1R–GLP-1 interactions: Structure

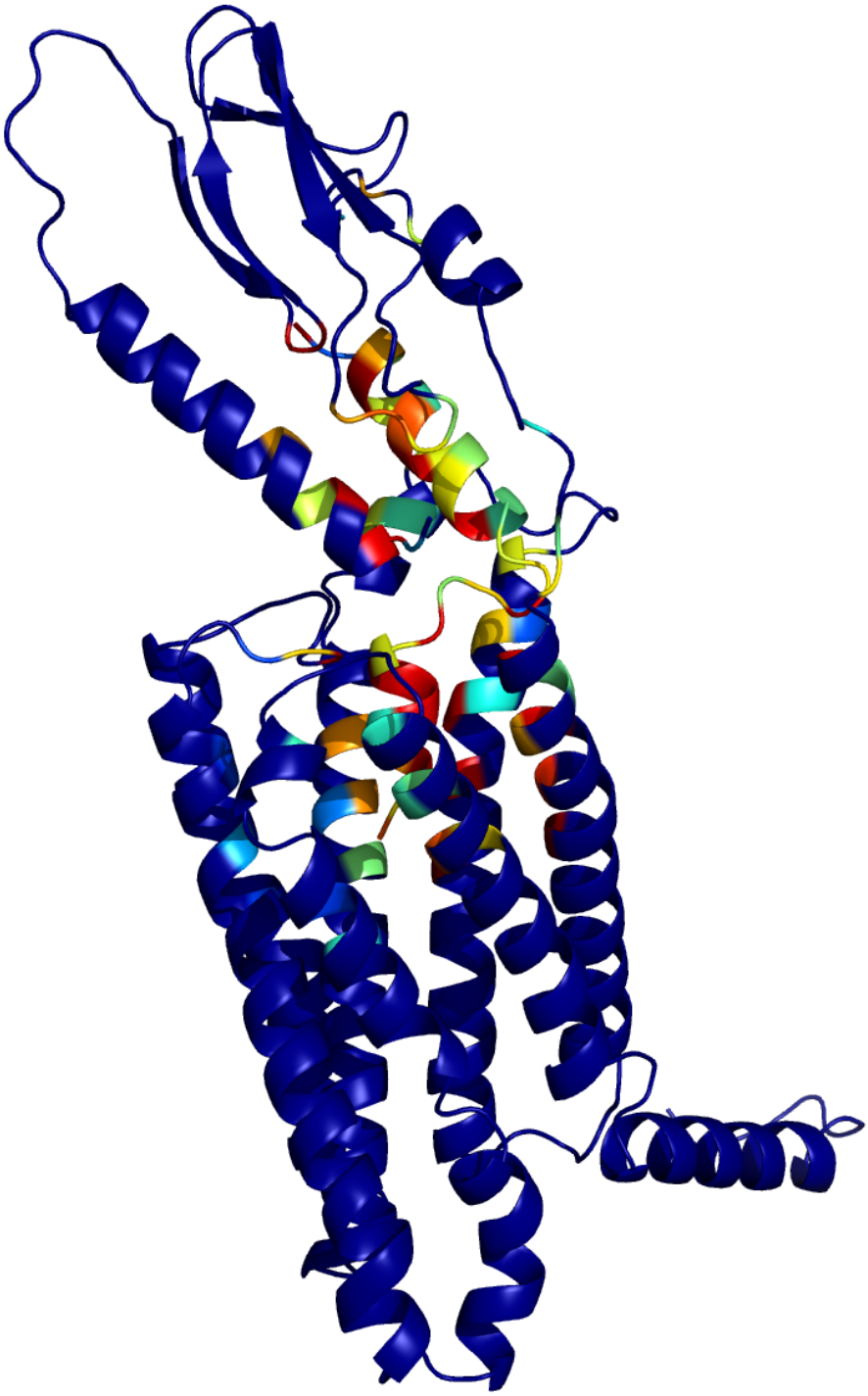


FIGURE 1: GLP-1R–GLP-1 structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–GLP-1(7000) simulation.

Appendix R GLP-1R–lira interactions: Structure

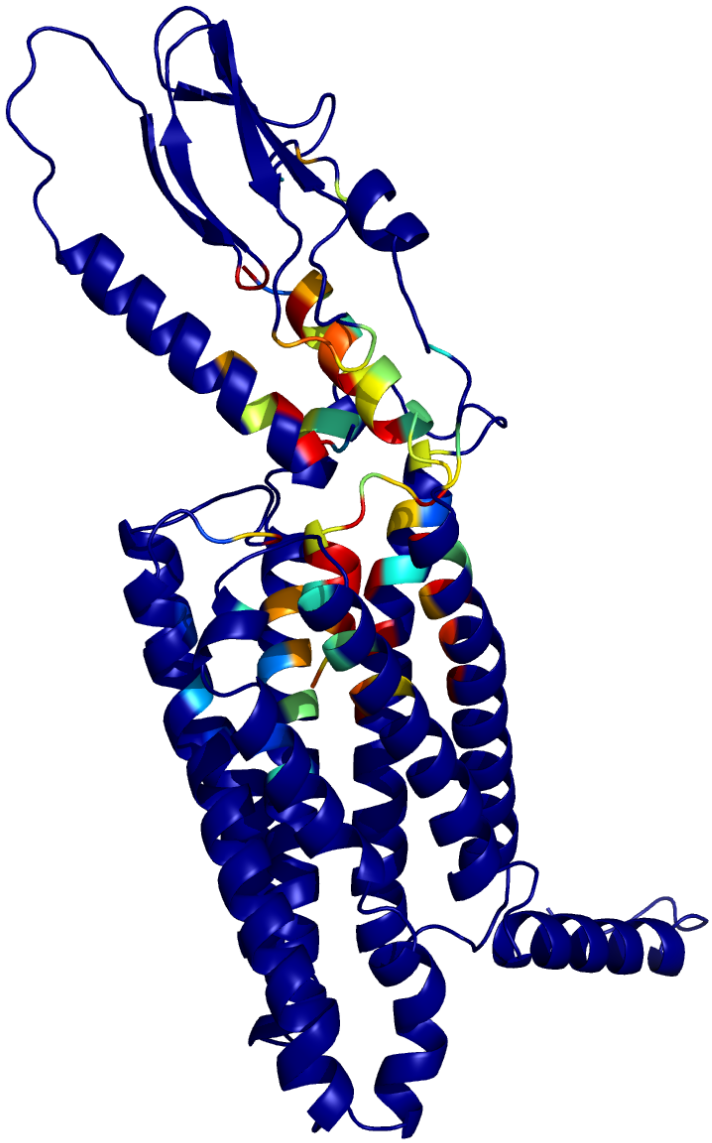


FIGURE 2: GLP-1R–lira structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–lira(7000) simulation.

Appendix S GLP-1R–C16L,K^{26,34} interactions: Structure

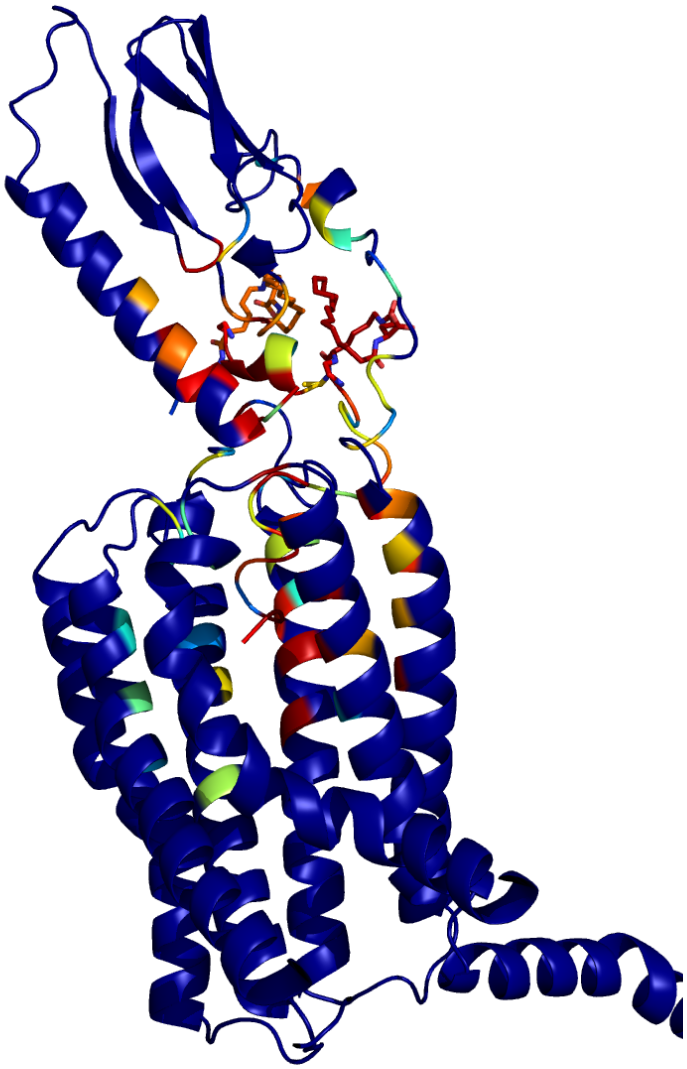


FIGURE 3: GLP-1R–C16L,K^{26,34} structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–C16L,K^{26,34}(7000) simulation.

Appendix T GLP-1R–C14A interactions: Structure

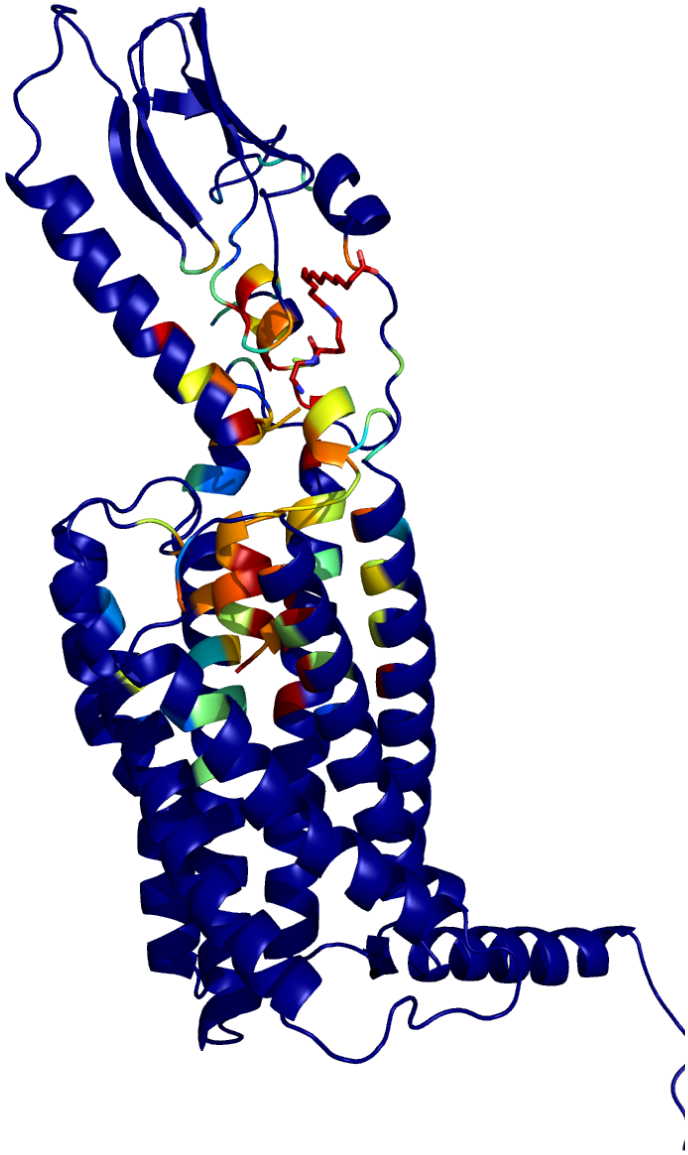


FIGURE 4: GLP-1R–C14A structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–C14A(7000) simulation.

Appendix U GLP-1R–daC16L,K^{34} interactions: Structure

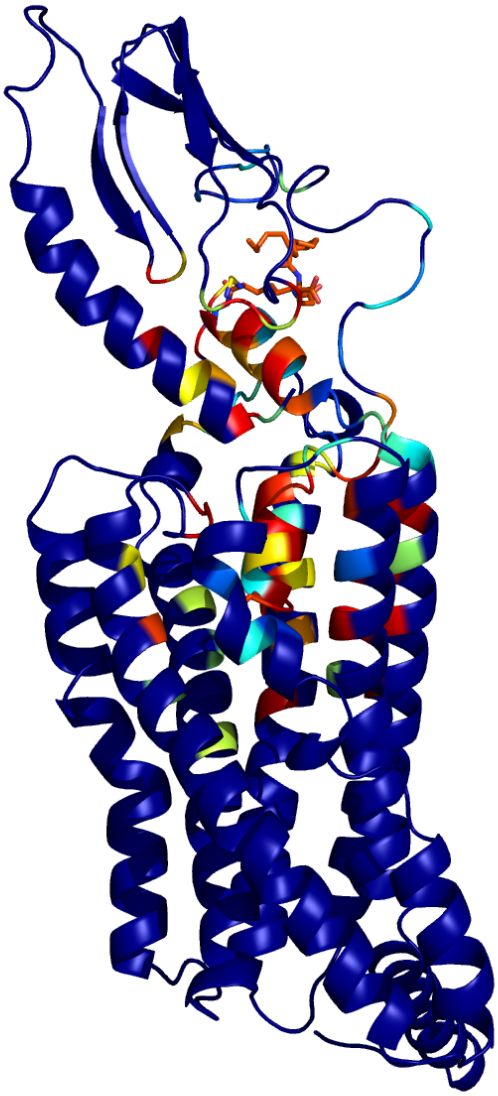


FIGURE 5: GLP-1R–daC16L,K^{34} structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–daC16L,K^{34}(7000) simulation.

Appendix V GLP-1R–daC8,K^{34} interactions: Structure

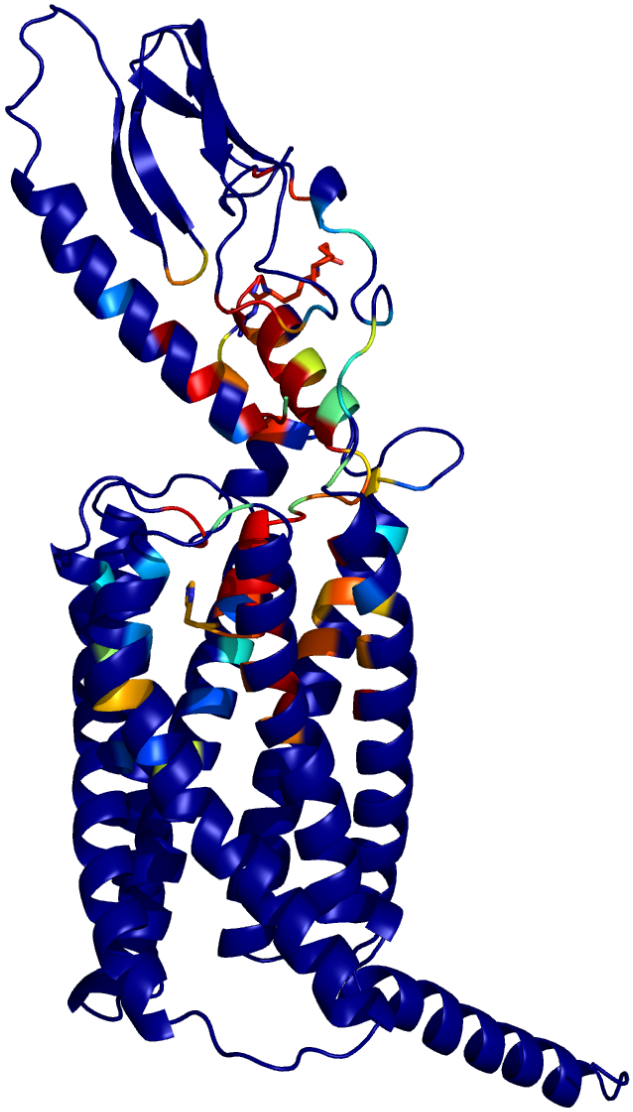


FIGURE 6: GLP-1R–daC8,K^{34} structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–daC8,K^{34}(7000) simulation.

Appendix W GLP-1R–C12A,K^{38} interactions: Structure

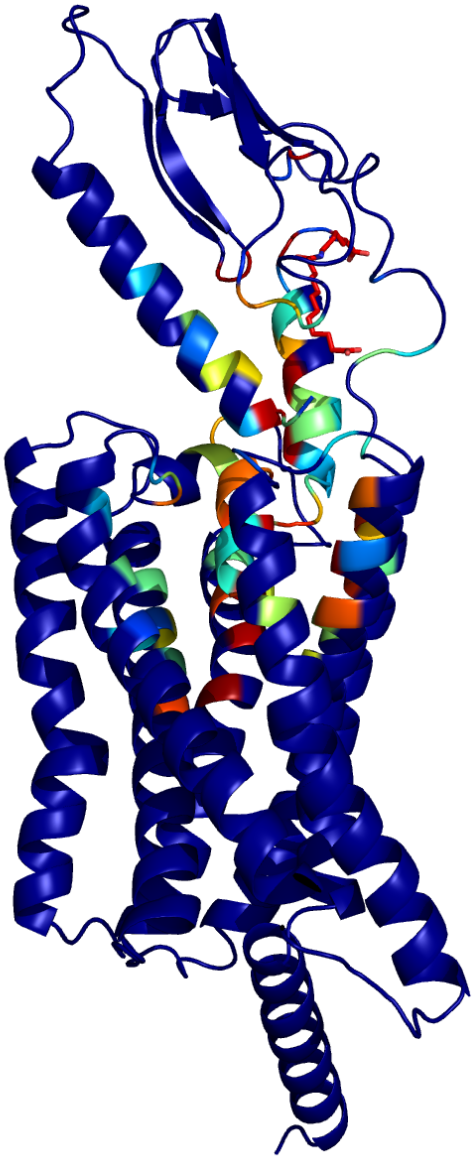


FIGURE 7: GLP-1R–C12A,K^{38} structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–C12A,K^{38}(7000) simulation.

Appendix X GLP-1R–C14L,K^{38} interactions: Structure

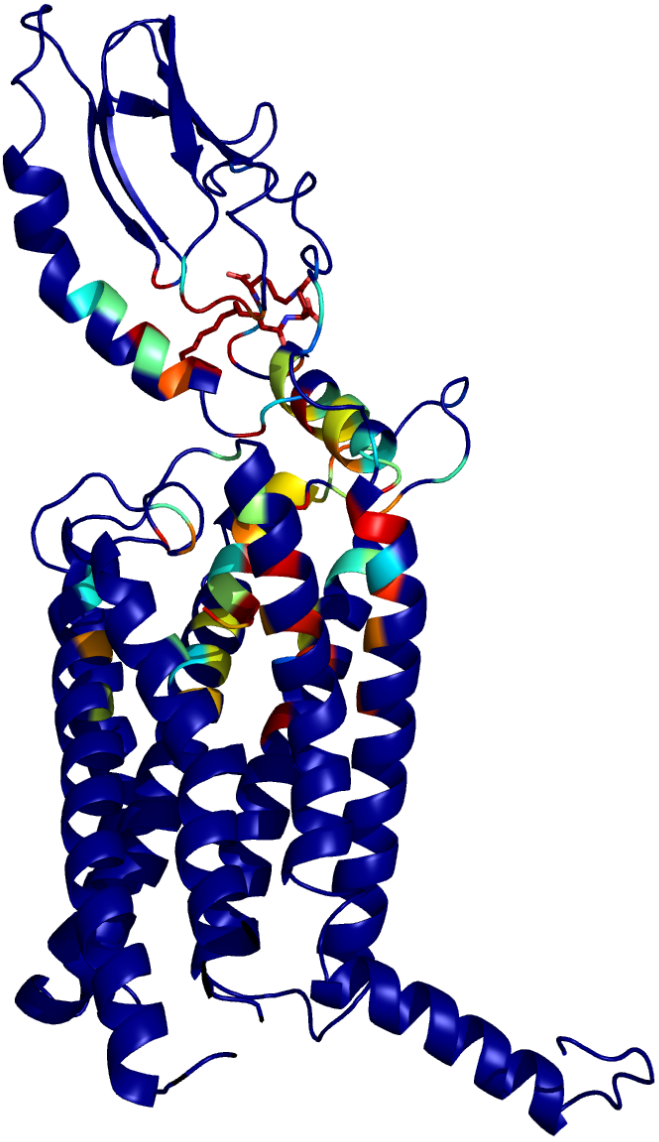


FIGURE 8: GLP-1R–C14L,K^{38} structure coloured according to the interaction strength of the interacting residue pairs. Image taken from the last frame of the GLP-1R–C14L,K^{38}(7000) simulation.

Appendix Y C code snippet: Calculation of variance-covariance matrix

Below is an example of a snippet of C code that can calculate the variance-covariance matrix. As input, it requires a file where the x-, y-, and z-coordinates of the atoms in the first frame are given as the first row, then the same for the second frame in the second row and so forth. That is, $3N$ columns and m rows where m is the number of frames.

```
//Triple for loop that populates cov
    for (int f = 0; f < m; f++){
        for (int i = 0; i < 3*n; i++){
            for (int j = i; j < 3*n; j++){
                //Only the elements on and above the diagonal
                //are calculated, since the elements below the
                //diagonal will be the same.
                cov->A[i][j] += (data->A[f][i]*data->A[f][j] +
                    ave->v[i]*ave->v[j] -
                    data->A[f][i]*ave->v[j] -
                    ave->v[i]*data->A[f][j])/(double)m;
            }
        }
    }

    for (int i = 0; i < 3*n; i++){
        for (int j = i+1; j < 3*n; j++){
            //The elements above the diagonal are copied
            //to the elements below the diagonal.
            cov->A[j][i] = cov->A[i][j];
        }
    }
```

Where `data` is a structure that contains the input coordinates in a matrix as described above, and `ave` is another structure that holds a vector with all the $3N$ coordinate averages (taken over all the studied frames). `cov` will hold the resulting covariance matrix. i and j loop over the matrix dimensions of the covariance matrix ($3N \times 3N$) and f over the number of frames, up to m. When dividing the trajectory into batches, m will be the number of frames in each batch.

Appendix Z Simulation setup for entropy tests

System 1

The homology model of the GLP-1 receptor and in complex with the GLP-1 analogue, 2C16L,K[26,34], was inserted into a POPC (3-palmitoyl-2-oleoyl-D-glycero-1-phosphoatidylcholine) bilayer membrane patch of 80 Å × 80 Å as described in the VMD [85] 'Membrane proteins tutorial' [86]. The system was then solvated using the VMD plugin [87] to ensure a water layer that surrounds the protruding protein parts by 15 Å when copying the simulation box in each direction. Water molecules are described using the TIP3 model [88]. Hereafter, the systems was neutralised with Na^+ and then adjusted to an ionic strength of 100 mM NaCl using the VMD plugin Autoionize [89]. Detailed information of the system is given in table 1.

TABLE 1: Number of atoms, water molecules, and NaCl together with the size of system 1 as used for simulation to test entropy calculation setup.

No. of atoms	No. of waters	No. of NaCl	Box size (Å×Å×Å)
61903	39216	25	83.99 × 80.96 × 108.99

Preparation and simulation was conducted as described in 'Membrane proteins tutorial' [86]. The described minimisation and equilibration setup was used initially.

Minimisation and molecular dynamics calculations were perfomed in NAMD [90] using the CHARMM36 force field [91]. Parameters and topologies were adapted to include the γ-Glu linker in the peptide as shown in Appendix E.

The intregration time step was 2 fs. The Lennard Jones potential is used to calculate vdW interactions with a cut-off of 12 Å in combination with a switching function from 10 Å to 12 Å. The pair list distance was set to 14 Å for short range interactions and updated every 20^{th} fs. Long range electrostatic interactions were calculated using the particle-mesh-Ewald method [101, 102].

The simulation production run was performed under NPT condition with constant number of atoms (N), pressure of 1 bar (P), and temperature 303 K (T). The pressure and temperature was controlled using Langevin thermo- and barostats [76, 162].

System 2

System 2 was prepared from the 1BRS pdb file using chains A and D. The system was solvated using the same procedure as for system 1. System 2 was neutralised and ionised as described for system 1. An overview of the resulting system is given table 2

TABLE 2: Number of atoms, water molecules, and NaCl together with the size of system 1 as used for simulation to test entropy calculation setup.

No. of atoms	No. of waters	No. of NaCl	Box size (Å×Å×Å)
37318	11390	21	79.98 × 72.78 × 73.96

Minimisation was carried out in NAMD [90] for 1000 steps, and system 2 was hereafter simulated using the same procedure as described for system 1.