Svoluji k zapůjčení své diplomové práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně ocitovat.

Charles University

Faculty of Science

Study programme: Biologie

Branch of study: Genetika, molekulární biologie a virologie

Bc. Michaela Černeková

Modelling Binding of Insulin Variants toward Insulin Receptor Based on Recent Structural Breakthroughs

Modelování vazby variant insulinu k insulinovému receptoru díky průlomům ve strukturní biologii

Diploma Thesis

Supervisor: RNDr. Martin Lepšík, PhD.

Consultant: RNDr. Jiří Jiráček, CSc.

Prague, 2019

Prohlášení:

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

V Praze, 26.4.2019

Acknowledgment

A very big thanks goes to my supervisor RNDr. Martin Lepšík, PhD., for giving me the opportunity to work on this project, for being always unhesitatingly helpful and determined to explain the methods, which were new to me, for his active approach to supervising and valuable advices. Further, I would like thank my consultant RNDr. Jiří Jiráček, CSc., for providing the experimental data of his research group, for comments and information about insulin receptor and its ligands, and for willingness to help me. I am also very grateful to Prof. Ing. Pavel Hobza, DrSc., FRSc, D.h.c., for allowing me to work in his research group in Institute of Organic Chemistry and Biochemistry, which was great honour and experience for me. Last but not least, I want to thank my colleagues for being helpful, namely I would like to mention Anja Muždalo, who gave me advices about molecular dynamics.

This work was supported by project: MŠMT OP VVV „Chemická biologie pro vývoj nových terapií“.

Abstract

Insulin receptor is a multi-domain signalling protein acting as a dimer. It comprises an extracellular ectodomain, a transmembrane domain and intracellular tyrosine kinase domain. Upon insulin binding, conformational changes in insulin as well as in insulin receptor occur and trigger the signaling cascade via the kinase domain.

Abnormalities in insulin and insulin receptor function cause *diabetes mellitus*, a widespread disorder which can be consequence of genetic factors as well as lifestyle and is manifested by increased level of blood glucose. A common treatment of *diabetes mellitus* is via insulin analogues with different molecular properties.

Insulin/insulin receptor interactions in the binding pocket are divided into two groups, so-called "site1" and "site2". The molecular details of the interactions in site1 are well known, while site2 residues are still not completely elucidated. It is important to shed light on the binding properties of insulin and insulin receptor, especially site2 interactions, because it could contribute to improved design of new insulin analogues. In this work, we used the very recent breakthroughs in the structural biology of insulin receptor to study the interactions by computational chemistry methods. It was thus possible to assess the noncovalent interactions and conformational changes in this system relevant to *diabetes mellitus* treatment. Especially, we observed that mutations of insulin residues, which are suggested as part of site2, can affect maintaining of the conformation necessary for binding to insulin receptor.

Keywords: insulin, insulin receptor, binding, diabetes, molecular modelling, molecular dynamics, molecular docking

Abstrakt

Inzulínový receptor je dimérny signálny proteín skladajúci sa z mnoho domén. Pozostáva z mimobunkovej časti, transmembránoveho úseku a vnútrobunkovej kinázovej domény. Počas väzby inzulínu k inzulínovému receptoru dochádza ku konformačným zmenám u oboch z nich, čo spúšťa signálnu kaskádu skrz kinázovú doménu.

Poruchy vo funkcii inzulínu a inzulínového receptoru spôsobujú *diabetes mellitus*, celosvetovo rozšírené ochorenie, ktoré môže vznikať ako následok genetických faktorov, ale aj životného štýlu a prejavuje sa zvýšenou hladinou glukózy v krvi. Bežnou formou liečby *diabetes mellitus* je podávanie inzulínových analógov s rôznymi molekulárnymi vlastnosťami.

Interakcie medzi inzulínom a inzulínovým receptorom vo väzobnej kapse sa rozdeľujú na dve skupiny nazývané „site1“ a „site2“. Zatiaľ čo molekulárne detaily interakcií v site1 sú už známe, reziduá patriace k site2 sa ešte nepodarilo celkom objasniť. Je teda dôležité ozrejmiť väzobné vlastnosti inzulínu a inzulínového receptoru, najmä interakcie v site2, keďže to by mohlo prispieť k zlepšenému návrhu nových inzulínových analógov. V tejto práci sme využili nedávne prelomy v štruktúrnej biológii inzulínového receptoru, na to aby sme študovali väzobné interakcie metódami výpočtovej chémie. Takto bolo možné pozorovať nekovalentné interakcie a konformačné zmeny v systéme, ktorý má význam v liečbe *diabetes mellitus.* Tým sa preukázalo sa najmä to, že mutácie v reziduách inzulínu, u ktorých je predpoklad, že sú súčasťou site2, možu ovplyvniť nadobúdanie konformácie dôležitej pre väzbu k inzulínovému receptoru.

Kľúčové slová: inzulín, inzulínový receptor, väzba, cukrovka, molekulové modelovanie, molekulová dynamika, molekulové dokovanie

Contents

Abbreviations

AIR - ambiguous interaction restraints (settings in docking program)

AMBER - Assisted Model Building with Energy Refinement (molecular dynamics program)

C++ - programming language

CR - cysteine-rich domain

cryo-EM - cryo-electron microscopy (structural biology method)

CUDA - Compute Unified Device Architecture (parallel computing platform)

DM - *diabetes mellitus*

FAs - fatty acids

FnIII-1/2/3 - fibronectin domain type III subtype 1/2/3

GLUT4 - glucose transporter type 4

HADDOCK - High Ambiguity Driven protein-protein Docking (docking program)

ID - insert domain

IGF-1/2 - insulin growth factor type 1/2

IR - insulin receptor

IRS-1 - insulin receptor substrate type 1

K - Kelvin (SI unit of thermodynamic temperature)

L1/2 - leucine-rich domain type 1/2

MD - molecular dynamics

NPH - neutral protamine Hagedorn

PDB - Protein Data Bank (database of three-dimensional structures)

PI - phosphatidylinositol

PKC - protein kinase type C

αCT - α-carboxy terminal

1 Introduction

Insulin receptor is tyrosine kinase of which extracellular domain (ectodomain) is able to bind insulin. The complete insulin receptor is dimer, whose monomers consist of leucine-rich domain, cysteine rich domain, second leucine rich domain, first fibronectin domain, second fibronectin domain with insert domain and third fibronectin domain in ectodomain. Further, there is transmembrane helix, juxtamembrane domain and the tyrosine kinase domain with the carboxy-terminal tail in the intracellular part. Insulin is composed of chains A and B, which are stabilized by disulphide bonds. Interactions between insulin and the insulin receptor in binding pocket are divided in two interfaces so-called "site1" and "site2". It means there are site1 residues in insulin and insulin receptor interacting with each other, and there are also site2 residues in insulin and insulin receptor interacting with each other. Site1 interactions are well known thanks to mutational analysis and X-ray crystallographic structures, but site2 interactions remain still unclear, because results of recent cryo-electron microscopy studies of insulin complexes with insulin receptor ectodomain do not fully confirmed insulin site2 residues proposed by mutagenesis.

Upon insulin binding to insulin receptor, the kinase activity is induced and signal is transferred via signal cascade resulting in metabolic as well as mitogenic effect. Abnormality in proper insulin receptor and insulin action can lead to disorder called *diabetes mellitus*, the main characteristics of which is an increased level of glucose in blood. There are two types of *diabetes mellitus*: type 1 and more common type 2. The type 2 is acquired during life and its development is due to genetic factors and, importantly, also the lifestyle. At the present time, insulin analogues are widely used in *diabetes mellitus* treatment.

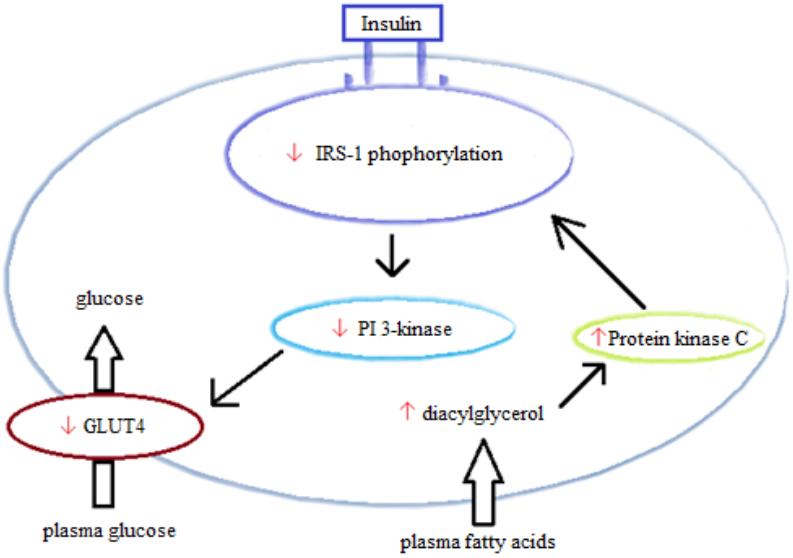
Better understanding of interactions between insulin and insulin receptor upon binding could offer new possibilities for rational design of insulin analogues. Thus, it is important to study still not completely understood site2 interactions. Using the very recent breakthroughs in insulin receptor structural biology, we have an excellent starting material for computational chemistry methods. The first steps thus were to build molecular models from the experimental structures. We further used these to simulate their dynamics and binding using molecular dynamics and docking approaches. During the simulations with insulin analogues structures we observed conformational changes and stability of noncovalent interactions, which brought an insight into binding properties. Docking of the insulin structures with the insulin receptor structures under different conditions provided several binding poses, which were then compared to the natural binding pose of insulin.

2 Literature Review

2.1 *Diabetes Mellitus*

Dysfunction in normal secretion or action of insulin causes disorder called *diabetes mellitus*. *Diabetes mellitus* (DM) is accompanied with high blood glucose level (hyperglycemia), which can lead to failure of many organs. There are two types of DM: type 1 and type 2. DM type 1 is characterized by total absence of insulin, while more common type 2 is in principle insulin resistance (Alam *et al*. 2014). At present, it is known that not only genetic factors but also the environment and lifestyle are involved in the development of DM type 2. Smoking of cigarettes, alcohol consumption, lack of physical activity, obesity or toxic substances in the environment, such as bisphenol A, could be contributing to the development of DM type 2 (Olokoba *et al*. 2012). However, the exact molecular mechanisms that cause development of DM type 2 are still unclear and subject of intensive research.

Development of the insulin resistance might be a consequence of declined glycogen synthesis in the skeletal muscle cells. It all starts with higher concentration of fatty acids (FAs) in plasma, which can be caused e.g. due to obesity. When there is elevated level of FAs, there is also elevated concentration of their metabolites, such as diacylglycerol, and they activate protein kinase C θ-type (PKC-θ). PKC-θ suppresses phosphorylation of insulin receptor substrate 1 (IRS-1) by kinase domain of insulin receptor after insulin is bound. The phosphorylated IRS-1 is necessary for phosphatidylinositol (PI) 3-kinase activity, but when IRS-1 is not phosphorylated, because of active PKC-θ, it is no able to bound PI 3-kinase, thus PI 3-kinase is then inhibited. When there is not present active PI 3-kinase, a defect of glucose transporter type 4 (GLUT4) transporters occurs because they need to be stimulated by active PI 3-kinase to be able to transfer glucose into the cells. The defect in cells not being able to accept glucose leads to the decline of glycogen synthesis [Fig.1].



[Fig.1]:

Schematic representation of one of the potential mechanisms for development of insulin resistance principle in skeletal muscles. Increased level of fatty acids in plasma activates protein kinase C, which inhibits phosphorylation of IRS-1. Unphosphorylated IRS-1 can not bind to PI 3-kinase and thus GLUT4 transporters are not stimulated to transfer plasma glucose into the cells, what leads to declined glycogen synthesis (adapted from: Petersen and Shulman 2006).

Alternatively, insulin resistance in the hepatocytes is also caused by accumulated metabolites of the FAs, because it leads to activation of protein kinase C (η-type) and it results in inhibition of insulin substrate 2 phosphorylation.

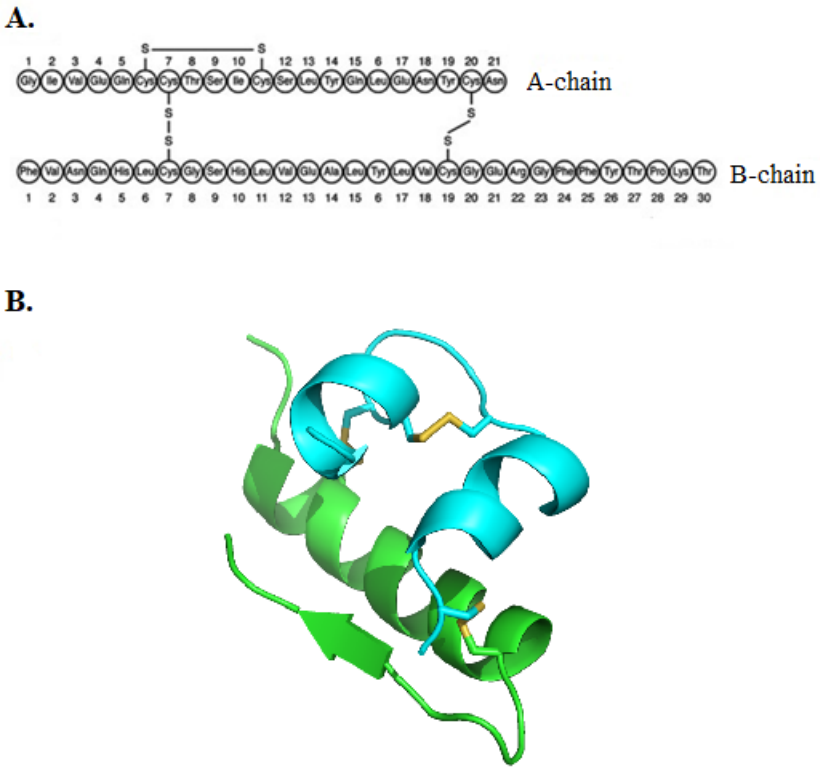
Not only obesity but also other genetic defects in adipocytes can lead to increased level of FAs, such as dysfunction of peroxisome proliferator-activated receptor γ (Petersen and Shulman 2006).

2.2 Insulin

The discovery of insulin by Frederick G. Banting and his team was enabled by prior experiments aimed at treating *diabetes mellitus*. The experiments with animals showed that the removal of their pancreas followed by the administration of the pancreatic extract back to their bodies lead to a decrease of the glucose level in their blood and urea. However, this approach could not be implemented in human medicine because of the immune reactions. The success came when Banting's team had joined James B. Collip to purify the active substance from bovine pancreas, which was used successfully in the treatment of humans (Rosenfeld 2002).

2.2.1 Structure of Human Insulin

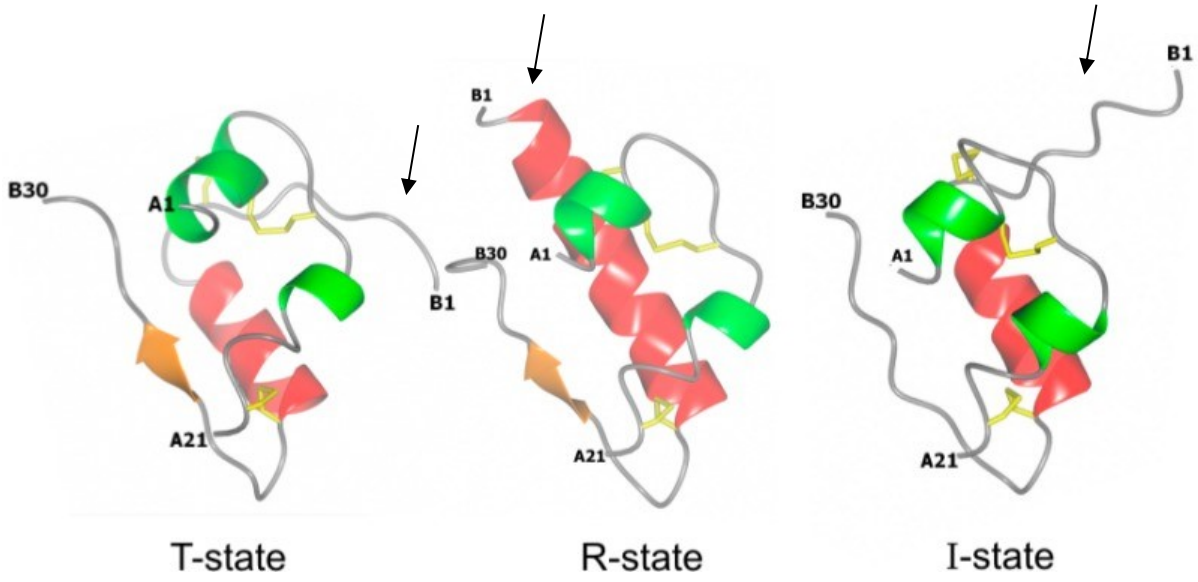
The pure bovine insulin was the first protein sequenced by Sanger's sequencing method and its structure was also determined. It was found that insulin consists of two chains, A and B. The first sequenced chain was B-chain, which was composed of 30 amino acids and the second sequenced chain was A-chain with 21 amino acids [Fig.2A] (Stretton 2002). The secondary structure of the A-chain comprises two α-helices, at the N- and C-termini. The B-chain consists of one α-helix in the central segment. The tertiary structure is formed by two interchain disulphide bonds joining both helices in the A-chain to the helix in the B-chain (Wilcox 2005). Moreover, there is one intrachain disulphide bond in the A-chain between residues Cys6 and Cys11 [Fig.2B] (Stretton 2002).



[Fig.2]:

A. The amino acid sequence of insulin A-chain (21 amino acids) and B-chain (30 amino acids) with schematic indication of disulphide bonds (adopted from: Heise and Mathieu 2017).
B. The three-dimensional structure of insulin monomer with the secondary structures shown: the A-chain is in cyan, the B-chain is in green. The disulphide bonds are shown as yellow sticks (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source Protein Data Bank structure 1BEN: Smith *et al*. 1996).

Insulin can be found in three possible states, R-, T- or I-state, depending on the conformation of the B-chain N-terminus [Fig.3]. In the R-state, the central helix of the B-chain is extended by the first eight residues of the N-terminus. There is also a variation of R-state called R^f-state, where the first three residues in B-chain N-terminus are not part of the extended helix. In the T-state, the N-terminus is not part of helix and has a different conformation. Another possibility is that the N-terminus takes on any conformation between R- and T-states and it is called I-state, for intermediate state (Kosinová *et al*. 2014) or it is also called O-state, for open state (Yao *et al*. 1999) [Fig.3]. Roles of different insulin states are not entirely elucidated, but there are suggestions that the most abundant T-state is required for a correct folding of pro-insulin (see below) and insulin, while the R-state is important for biological activity, but does not play a role in receptor binding (Kosinová *et al*. 2014).



[Fig.3]:

Structures of the three states of human insulin differing in the conformation of the N-terminus of insulin B-chain (shown by arrow) (adopted from: Kosinová *et al*. 2014).

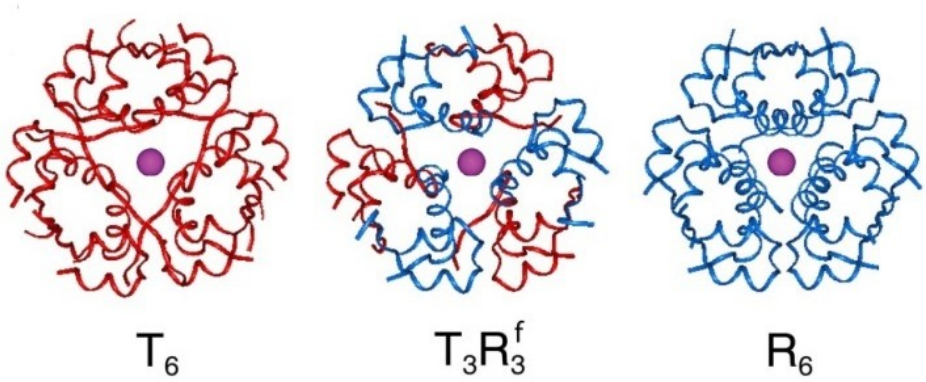
There is a structure with which insulin shares a common ancestor, the insulin-like growth factor 1 (IGF-1). There are numerous similarities and a few differences. One of the differences is that IGF-1 is composed of 70 amino acids, where residues 1-29 are homologous to B-chain of insulin and residues 42-62 are homologous to the A-chain of insulin (Rinderknecht and Humbel 1978). Structurally, IGF-1 also contains three α-helices, two in A-chain and one in B-chain, which are stabilized by three disulphide bonds, but in addition there is C-region and D-region not present in insulin (Vajdos *et al*. 2001). Regarding to functional differences between insulin and IGF1, insulin executes first and foremost a metabolic role, while IGF-1 mainly a mitogenic role, although there are slight overlaps in their function (Werner *et al*. 2008). Another related molecule to insulin is insulin growth factor 2 (IGF-2), with which insulin shares 50% homology of amino acids. It is supposed that IGF-2 is important for prenatal growth, but it was shown also its ability to bind to insulin receptor, thus there is also speculation about its metabolic role (Cianfarani 2012).

2.2.2 Maturation, Secretion and Activity of Insulin

Insulin is first synthesized by ribosomes as pre-proinsulin in endoplasmic reticulum in pancreatic β-cells, specifically in Langerhans islets. Pre-proinsulin contains signal peptide, which has to be removed in order to form proinsulin. Removal of the signal peptide occurs in the endoplasmic reticulum and proinsulin is then relocated to Golgi apparatus. The high concentration of calcium and zinc ions aids in the formation of hexamers containing proinsulin (Wilcox 2005). While the role of zinc in formation of the hexamers is well known, the role of calcium, which is present even in higher concentration then zinc, is not fully elucidated. It is suggested it plays a role in induction of hexamerization (Duboué-Dijon *et al*. 2018).

Proinsulin still differs from the insulin in that it has C-peptide, which is separated from insulin in Golgi apparatus and hexamers with final insulin peptide are formed (Wilcox 2005). Three types of hexamers can be formed according to which insulin states participate in oligomerization: R_6, T_6 or $T_3R_3/T_3R^f_3$ [Fig.4]. In R_6 hexamers, the monomers of all three dimers are in the R-state, while in T_6 hexamers all are in the T-state. In $T_3R_3/T_3R^f_3$ hexamers, each dimer is formed by one monomer in T-state and one in R-state. (Kosinová *et al*. 2014). T_6 hexamers can be assembled into T_3R_3 hexamers in presence of anions from Hofmeister series, e.g. SCN^-. Then T_3R_3 hexamers can transition into R_6 hexamers by adding phenolic compounds, which can be bound to six binding sites in the hexamer (Shneine *et al*. 2000).

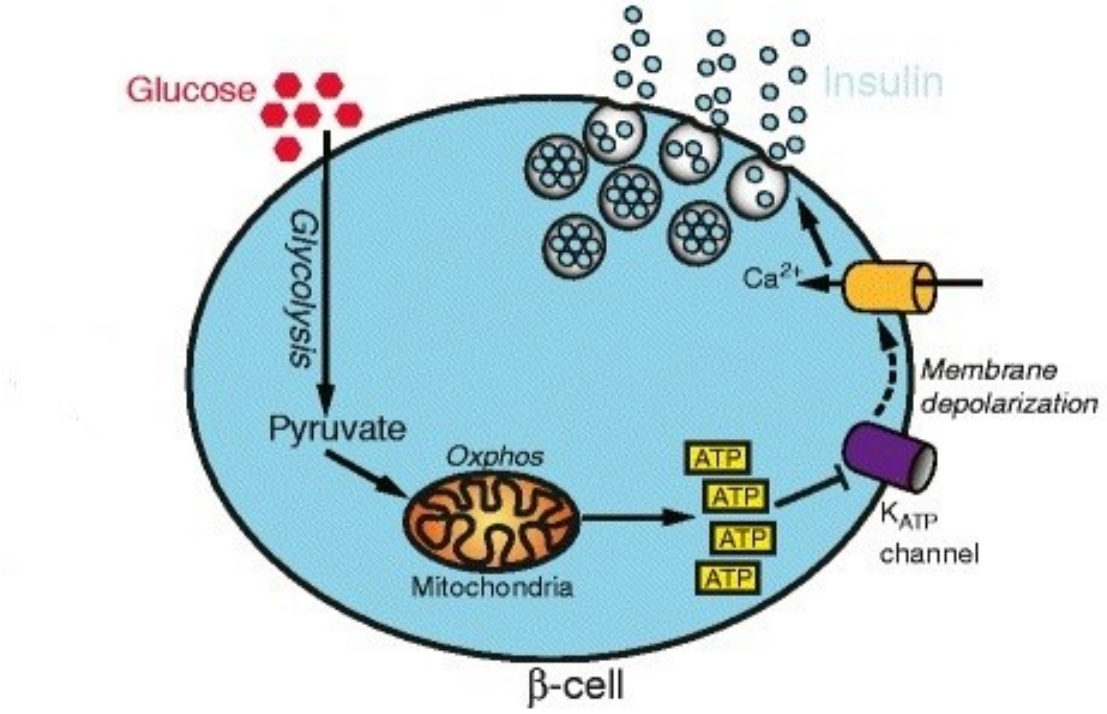
It was showed that neurotransmitters have also an effect on insulin oligomerization. Particularly, within insulin hexamer can be bound serotonin, which stabilizes T_3R_3 conformation. To other molecules, which are able to be bound by insulin hexamers belong dopamine and also arginine, which contribute to stabilized T_3R_3 conformation as well. As those three proteins are abundantly present in insulin storage granules in pancreas, it is assumed that T_3R_3 conformation is the one preferred for insulin storing. In addition, these abundant proteins in storage granules could have an impact on processing pro-insulin into mature insulin (Palivec *et al*. 2017).



[Fig.4]:

There are three types of hexamers, which can be formed depending on the insulin state. T_6 contains six insulin monomers in T-state, all in red. $T_3R_6^f$ contains three insulin monomers in T state (red) and three insulin monomers in R-state (blue). R_6 contains six insulin monomers in R-state, all in blue. The circle in the centre represents zinc ions (adopted from: Weiss *et al*. 2000).

Mature insulin is afterwards secreted in response to an impulse, most typically glucose, but it could also be amino acids or hormones. There is an important role of glucose transporters, which transfer glucose to β-cells. When glucose enters into β-cells, it starts to be subsequently metabolized by glucokinase and as a result of forming glucose-6-phosphate, ATP is generated. In consequence of raised levels of intracellular ATP, depolarization of cell membrane occurs, because ATP sensitive potassium channels close and calcium channels open. Thus the intracellular calcium level starts to be elevated and it causes insulin secretion from the β-cells [Fig.5] (Wilcox 2005).



[Fig.5]:

Pancreatic β-cell, which in case of increased glucose level produces more ATP. Elevated level of ATP molecules triggers closing of ATP sensitive potassium channels, thus opening of calcium channels, because of membrane depolarization. Ca^{2+} ions enter into β-cell and insulin is secreted (adopted from: Cantley and Ashcroft 2015).

Insulin secretion as consequence of glucose stimulus appears in two phases. First phase is represented by intensive release of insulin and lasts for 10 minutes. Second phase occurs after those 10 minutes and is characterized by more moderate release of insulin, which last between 1 and 2 hours. Insulin releasing in the second phase depends on glucose concentration in blood, while in the first phase it is not related to glucose level. It means if there is more elevated level of glucose, there is also more extensive release of insulin in the second phase (Pfeifer *et al*. 1981).

As we mentioned above, insulin activity can be regulated not only by nutrient factors such as glucose, but also by peptide hormones and amino acids. For example, gastrointestinal hormone glucagon-like peptide 1 or amino acid leucine are able to potentiate insulin secretion, while hormone somatostatin suppress insulin secretion (Wilcox 2005).

Secreted insulin can be bound only as monomer (Glidden *et al*. 2018) by insulin receptor family, including two isoforms of insulin receptors (IR-A/ IR-B) and one insulin-like growth factor-1 receptor. Activation of receptors from insulin family results in an activation of metabolic (e.g. lipid metabolism) as well as in mitogenic (proliferation, apoptosis) signaling cascades (Leibiger *et al.* 2008).

2.2.3 *Diabetes Mellitus* Treatment with Insulin Analogues

Insulin analogues are needed in the treatment of diabetic patients, because injection of natural human insulin does not fully not mimic physiological action of endogenous human insulin as it is desirable. Especially, subcutaneously injected human insulin has slow onset of actions and it is not able to reduce high glucose levels after meal as quickly and efficiently as endogenous pancreatic insulin. Realizing this, modifications of the natural human insulin were necessary. The analogues can be divided in two main groups: basal insulin analogues and rapid-acting insulin analogues (Lindholm 2002).

Long-Acting Basal Insulin Analogues

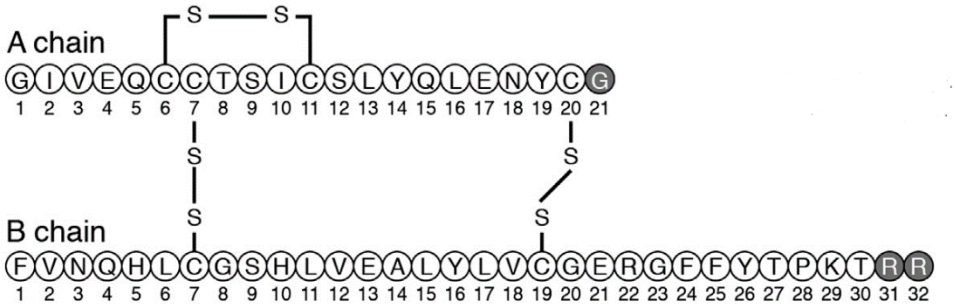
Basal insulin analogues have prolonged action and their action peak is lower than the peak of endogenous insulin. The reason of this ability is more moderate absorption into circulatory system (Mavrogiannaki and Migdalis 2012). The characteristic properties of basal insulin analogues can be acquired with several methods, for example with precipitation, formation of multi-hexameric structures after injection, PEGylation or binding with another protein (Heise and Mathieu 2017).

The first basal insulin analogue prepared by precipitation was neutral protamine “Hagedorn” (NPH). Besides the protamine, Hagedorn contains also insulin, zinc and phenol, or instead of phenol can be included phenolic derivatives (Hagedorn 1937, cited according to Hilgenfeld *et al*. 2014). Insulin is in this suspension in crystal form and after injection with protamine stays in depot. Dissolvation of accumulated insulin crystals from depot is delayed, what is the cause of the more prolonged activity of NPH analogue (Søeborg *et al*. 2012). Peak of its action occurs

after 4-6 hours and considering its prolonged activity up to 14 hours (in regard to insulin), it started to be administrated twice a day (Lepore *et al*. 2000).

With techniques of recombinant DNA came up also new possibilities how to modify insulin sequence itself. Using these techniques allowed to administrate insulin as solution, not in crystal form as before. The consequences were elevated stability and prolonged action (Heise and Mathieu 2017). The first insulin analogue containing amino acid mutations was named "NovoSol Basal". This analogue has one mutation in C-terminus of A-chain (Asn21Gly) and another mutation in C-terminus of B-chain (Thr27Arg). In addition, C-terminus of insulin B-chain is amidated in order to block carboxyl group (Markussen *et al*. 1988). NovoSol Basal showed prolonged action, but was removed from further clinical studies, because of large variability of activity between patients and decreased variability intraindividually (Vajo and Duckworth 2000).

Another basal insulin analogue "glargine" is mutated at position Asn21 in A-chain to glycine and to C-terminus of B-chain are added extra two arginines Arg31 and Arg32 [Fig.6]. The result of mutation and addition of new residues is change of isoelectric point to more neutral values, what makes this analogue less soluble in neutral pH, thus has a prolonged action (Seipke *et al*. 1992, cited according to Mane *et al*. 2012). Clinical study showed that glargine is able to lower glucose lever for 24 hours and it is enough to administrate only one injection per day (Porcellati *et al*. 2007).

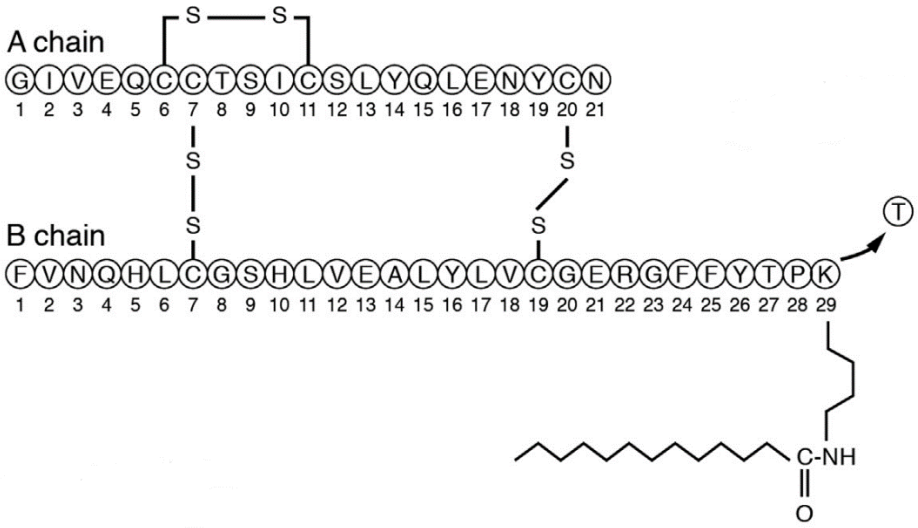


[Fig.6]:

Basal insulin analogue glargine with mutation Asn21Gly in A-chain and addition of two arginines in C-terminus of B-chain (adopted from: Tibaldi 2014).

Basal insulin analogue named "detemir" was prepared with different approach as the previous insulin analogues. Detemir does not contain any amino acid substitution, but has only

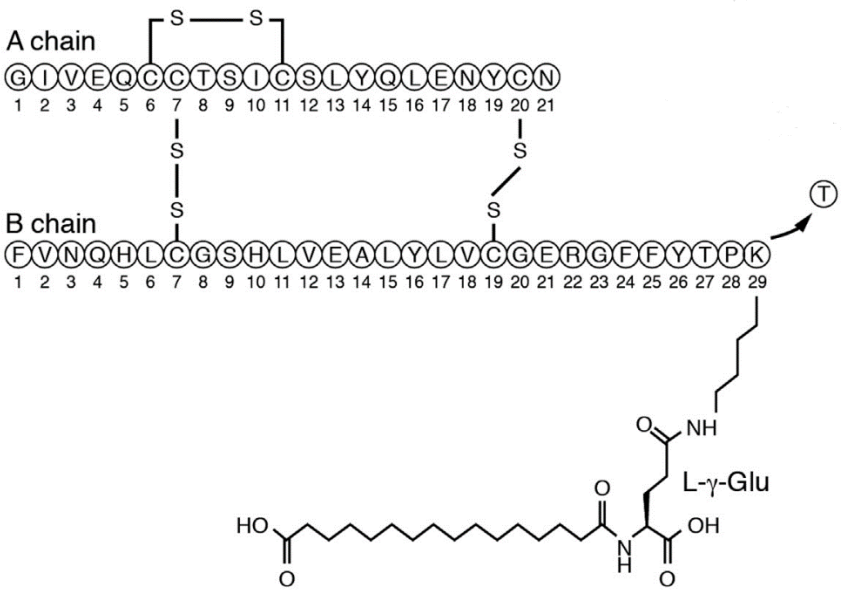
29 amino acids in B-chain (wild-type insulin has 30 amino acids) and it is because of myristoyl fatty acid is bound to lysine at position 29 in B-chain [Fig.7]. The addition of myristoyl fatty acid contributes to assembly of detemir molecules into hexamers and this is, actually, only acylated residue, which does not disrupt insulin assembly and also activity. Added myristoyl fatty acid plays also a role in reversible binding of albumin to detemir, which results in more steady activity of detemir, because albumin slows its absorption (Markussen *et al*. 1996).



[Fig.7]:

Basal insulin analogue detemir, which has removed threonine in C-terminus of B-chain and the last residue Lys29 is acylated with myristoyl fatty acid (adopted from: Tibaldi 2014).

"Degludec" is basal insulin analogue, which has also prolonged action by acetylation. The last residue in B-chain (Thr30) is removed and residue Lys29 is bound with fatty acid, however, not with the same 14-carbon fatty acid as in detemir, but with 16-carbon fatty diacid and it is done via gamma-glutamic acid spacer [Fig.8]. Degludec is injected together with zinc and phenol, what enables to form di-hexamers. After the injection, phenol is dissipating and degludec with zinc are organized into multi-hexameric chains. After zinc starts to be disappearing as well, multi-hexameric chains are disrupted and degludec maintains dimeric and then monomeric structure, which is able to penetrate into blood circulation from depot (Jonassen *et al*. 2012). Clinical studies confirmed degludec as ultra-long acting basal analogue with slow and consistent absorption. There was a high concentration of degludec in serum even after 42 hours and it was possible to detect it in serum for 120 hours after injection (Heise *et al*. 2012).



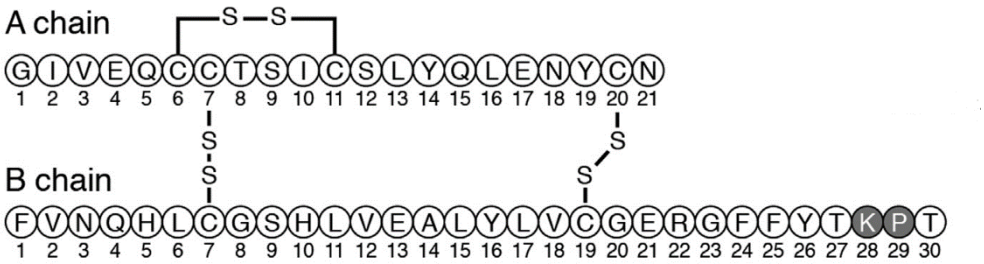
[Fig.8]:

Basal insulin analogue degludec has removed threonine in C-terminus of B-chain and the last residue Lys29 is acylated via gamma-glutamic spacer with 16-carbon fatty diacid (adopted from: Tibaldi 2014).

Rapid-Acting Insulin Analogues

There are three rapid-acting analogues, which are currently available: lispro, aspart and glulisine. They have an advantage of being absorbed quickly, thus are able to better mimic endogenous insulin secretion after meal (postprandial) than subcutaneously injected human insulin. Their presence in circulatory system is limited, which decreased risk of hypoglycemia. They are usually prescribed for postprandial application altogether with basal insulin analogues to be used for overnight application (Tibaldi 2014).

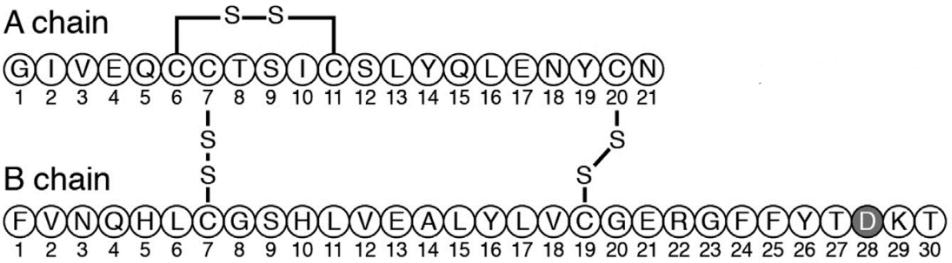
One of used rapid-acting insulin analogues is called “lispro”. Lispro has switched position of residues Pro28 and Lys29 in B-chain, what means that there is Lys28 and Pro29 [Fig.9]. The inverted arrangement of these residues results in lispro having decreased ability of self-assembly, what allows to peak two time higher in half of time than in the case of insulin (Howey *et al*. 1994). In order to improve postprandial glucose level it was suggested that lispro should be injected 15-30 minutes before meal (Rassam *et al*. 1999).



[Fig.9]:

Rapid-acting insulin analogue lispro, which has inverted amino acids, original Pro28 is here at position Pro29 and original Lys29 is there at position Lys28 (adopted from: Tibaldi 2014).

Insulin analogue "aspart" also belongs to the rapid-acting group of analogues. Aspart is mutated at position Pro28 in B-chain to aspartic acid [Fig.10]. As well as in case of lispro, this change results in decreased self-assembly into the hexamers, thus aspart is rapidly dissociated into the dimers, then monomers and is quickly absorbed. This analogue is also administrated short time before meal to ensure the postprandial glucose level decline (Lindholm and Jacobsen 2001). There is also available mixed solution of aspart and basal insulin analogue degludec at the time, which allows to inject both types of insulin analogues and it represents the first combination of different analogues. Ratio of these analogues is 70% degludec to 30% aspart and characteristic action of both analogues remains independent from each other (Havelund *et al*. 2015).

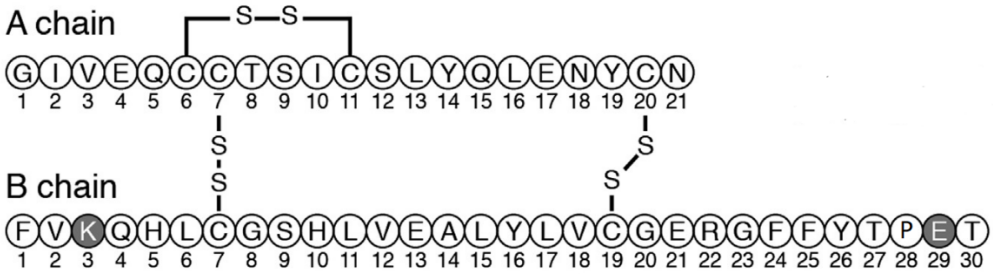


[Fig.10]:

Rapid-acting insulin analogue aspart has mutation Pro29Asp (adopted from: Tibaldi 2014).

The third approved rapid acting analogue is "glulisine", which is mutated at two positions in B-chain: Asn3 is replaced by lysine and Lys29 is replaced by glutamic acid [Fig.11]. Similar to the previous rapid-acting analogues, glulisine has also decreased ability of self-assembly into the hexamers. Therefore increased dissociation into the monomers results in rapid absorption

after injection. It can be injected in short time before meal to improve postprandial glucose level in circulation, compared to insulin, which has to be injected 30 minutes before meal to ensure similar action (Garg *et al*. 2005).



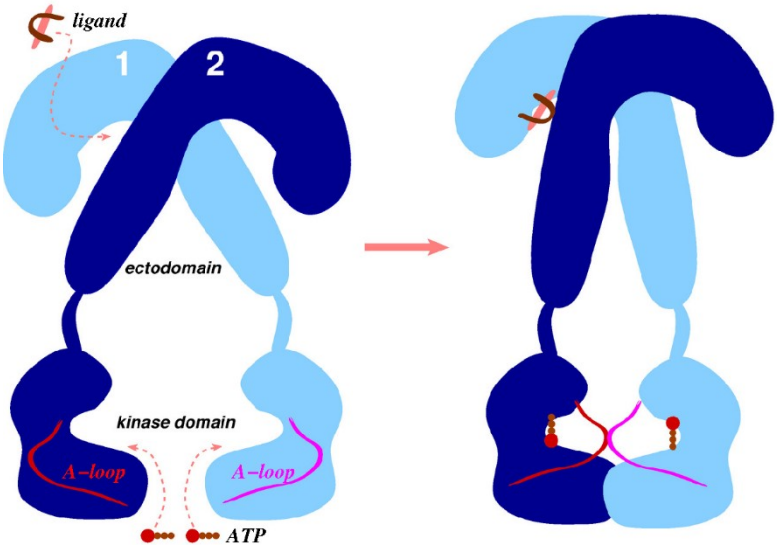
[Fig.11]:

Rapid-acting insulin analogue glulisine with mutation Asn3Lys and Lys29Glu (adopted from: Tibaldi 2014).

2.3 Insulin Receptor

2.3.1 Structure of Insulin Receptor

Insulin receptor (IR) is a signalling transmembrane protein. It binds insulin via its extracellular domain (ectodomain) and the signal is transduced to its intracellular tyrosine kinase domain [Fig.12], which activates downstream mitogenic and metabolic cascades. Besides insulin, IR can bind insulin-like growth-factor type 1 and 2 (IGF1 and IGF2), but with diverse affinity (Vashisth 2015).



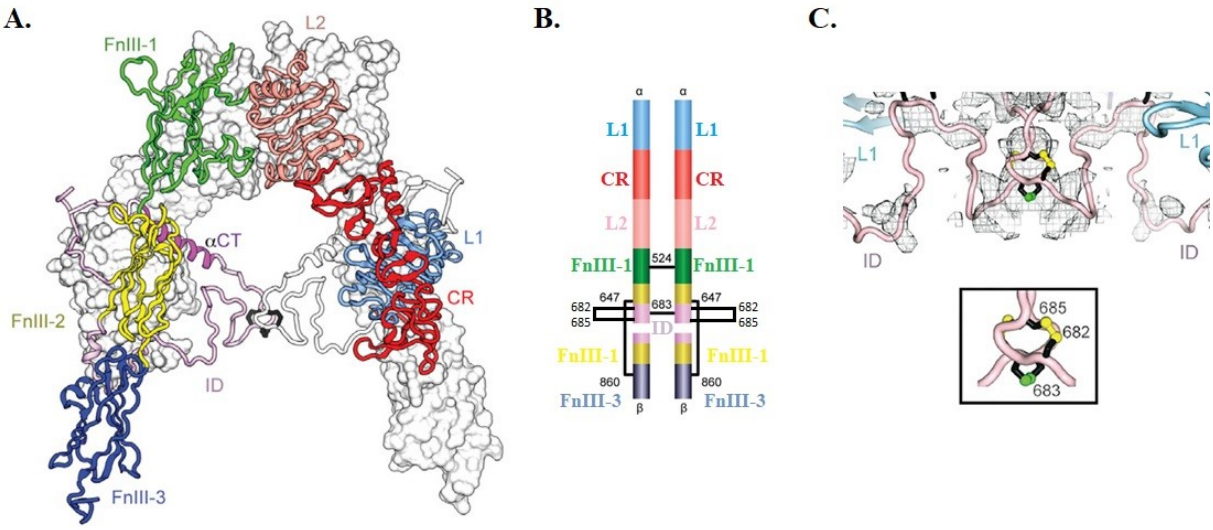
[Fig.12]:

Insulin receptor ectodomain binds ligand, which triggers signal transfer to intracellular kinase domain and activation of signal cascade (adopted from: Vahisth 2015)

IR is a homodimer with both monomers cleaved into α-chain at N-terminus and β-chain at C-terminus. Ectodomain, extracellular part of receptor, involves entire α-chain and 194 residues of β-chain of each monomer (Smith *et al*. 2010). The α-chain and β-chain of the same monomer are joined by one disulphide bond, between residues Cys647 (α-chain) and Cys860 (β-chain). However, there is another intra-monomer disulphide bond, which is intra-α-chain and it is between Cys682 and Cys685. Intermonomer disulphide bonds connecting two monomers into one dimer are formed between residues Cys524 of both monomers [Fig.13B] and another intermonomer disulphide bond is between residues Cys683 and Cys685, which form with the intramonomer disulphide bond (C682-C685) a triplet [Fig.13C] (Croll *et al*. 2016).

Both αβ-monomers have an inverted „V“ shape and each leg of the inverted „V“ consists of several domains [Fig.13A]. One leg is subdivided into a leucine-rich domain 1 (L1), followed by a cysteine-rich region (CR) and second leucine-rich domain (L2). The second leg includes three fibronectin type III domains (FnIII-1/2/3). There is an insert domain (ID) within FnIII-2 domain, in which the monomer is proteolytically cleaved into α- and β-chain and there is also structurally important α-helical motif called α-C-terminal (α-CT) helix (Croll *et al*. 2016).

Insulin receptor ectodomain ends with FnIII-3 domain at C-terminus and is followed by transmembrane helix, which continues with intracellular part of the receptor including a juxtamembrane region, a tyrosine kinase catalytic domain and C-terminal tail (C-tail) (Smith *et al*. 2010).



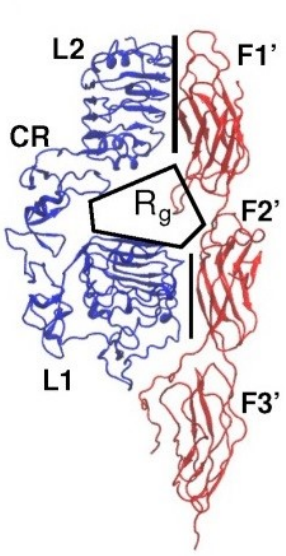
[Fig.13]:

A. Inverted „V“ structure of insulin apo-receptor ectodomain, where it can be seen secondary structure of individual domains, which are coloured differently. Above this structure is scheme of how individual domains come after in monomers and there are also included intracellular domains
B. Schematic representation of the chain organization in insulin receptor together with the location of all the disulphide bonds
C. A triplet formed from three disulphide bonds, two are intra-monomer between C682-C685 in both monomers and one is inter-monomer C683-C683 (adapted from: Croll *et al*. 2016).

There are two isoforms of insulin receptor: IR-A and IR-B, which are products of alternative splicing. Their structures differ only by extra 12 amino acids at the C-terminus of α-CT helix in the isoform B. Each isoform has its own predominant tissue distribution. IR-B is mainly located in hepatocytes, skeletal muscle cells and subcutaneous fat and IR-A is main isoform in other types of tissues, such as lymphatic tissues, brain or in embryos. It is also suggested that activation of IR-B has metabolic impact and activation of IR-A has mitogenic, which can be connected to their tissue distribution, but it is still not fully clear and more accurate evidence is missing (Jiráček a Žáková 2017).

2.3.2 Insulin Receptor Interactions with Insulin

Insulin-binding pocket on IR is located between L1-CR-L2 leg of one monomer and FnIII-1/2/3 leg of the second monomer and radius of gyration (R_g) of the binding pocket is predicted by molecular simulations to be 20Å [Fig.14] (Vashisth 2015).



[Fig.14]:

The insulin-binding pocket on IR ectodomain, which is placed between two monomers. L1-CR-L2 part of one monomer is blue and FnIII-1/2/3 part of second monomer of the same dimer is red (adopted from: Vashisth 2015).

Insulin receptor dimer interacts with insulin monomer via two cooperatively coupled interfaces called „site1"and „site2". While the former has been structurally described by X-ray crystallography (see details below) and by alanine substitutions (Smith *et al*. 2010), the structural characterization of the latter has been lagging and is the major focus of this thesis. Site1 surface of the receptor includes α-CT helix from one monomer and central β-sheet of L1 domain of other monomer [Fig.15].

[Fig.15]:
Representation of site1 and site2 surface in insulin receptor. Site1 surface is highlighted in blue and site2 surface is highlighted in red (adopted from: Vashisth 2015).

Insulin residues of A-chain and B-chain are in contact with residues of α-CT helix of receptor. Specific interactions were found based on X-ray crystallography and cryo-electron microscopy (cryo-EM) structures (see below) [Table.1].

Interactions	Insulin receptor residues	Insulin residues	Source
1.	His710 (α-CT helix)	A-chain: Val3 B-chain: Gly8, Ser9, Val12	Menting *et al*. 2013
2.	Phe714 (α-CT helix)	A-chain: Gly1, Ile2, Tyr19 B-chain: Leu11, Val12, Leu15	
3.	Phe39, Phe64, Arg65 (L1 domain)	B-chain: Val12	
4.	Phe39 (L1 domain)	B-chain: Tyr16	

[Table.1]:

Table of site1 interactions between insulin receptor and insulin residues. Particular rows of the table represent set of the residues interacting with each other, e.g. His710 (α-CT helix) interacts with Val3 in insulin A-chain and Gly8, Ser9, Val12 in insulin B-chain.

In modelling studies, other insulin residues have been suggested to be included in site1 interactions, such as Gln5, Thr8, Asn21 in A-chain and Tyr16, Phe24, Phe25, Tyr26 in B-chain (Vashisth 2015). The insulin residues belonging to site1 are components of the surface, which is involved in dimer formation (Glendorf *et al*. 2008).

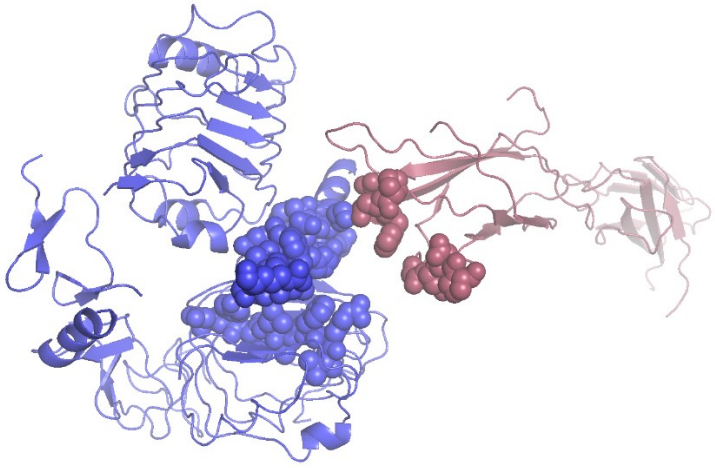
Site2 interface of insulin receptor and insulin is not known so well in detail as site1. It was suggested from receptor structures, which were obtained by X-ray crystallography that site2 residues of insulin receptor are located in the first and second fibronectin domain [Fig.15] (Croll *et al*. 2016). On the other side, the structures of insulin receptor with bound insulin obtained by cryo-electron microscopy show site2 interface formed by residues in first fibronectin domain and L2 domain of insulin receptor [Fig.16]. The insulin receptor and insulin residues, which are considered as part of site2 are shown in the table below [Table.2].

Interactions	Insulin receptor residues	Insulin residues	Source
1.	Pro495-Arg498, Arg539-Asn541 (FnIII-1 domain)	A-chain: Glu4, Gln5, Cys6, Cys7, Thr8, Ile10	Scapin *et al*. 2018
2.	Lys484, Leu552, Asp591 (FnIII-1); Ile602, Lys616, Asp620 and Pro621 (FnIII-2)	A-chain: Ser12, Leu13, Glu17 B-chain: His10, Glu13, Leu17, Val18	Vashisth 2015

[Table.2]:

Table of suggested site2 residues of insulin and insulin receptor according to different sources. The first row of the table represents site2 residues suggested from cryo-EM structure (Scapin *et al*. 2018). The second row represents different site2 residues, which were suggested from computational chemistry studies (Vashisth 2015).

Another mutagenesis study indicated that site2 insulin residues include Thr8, Ile10, Ser12, Leu13, Glu17 from A-chain and His10, Glu13, Leu17 from B-chain, what connects some insulin residues from the first row of the table [Table.2] with the residues from the second row (De Meyts 2015).



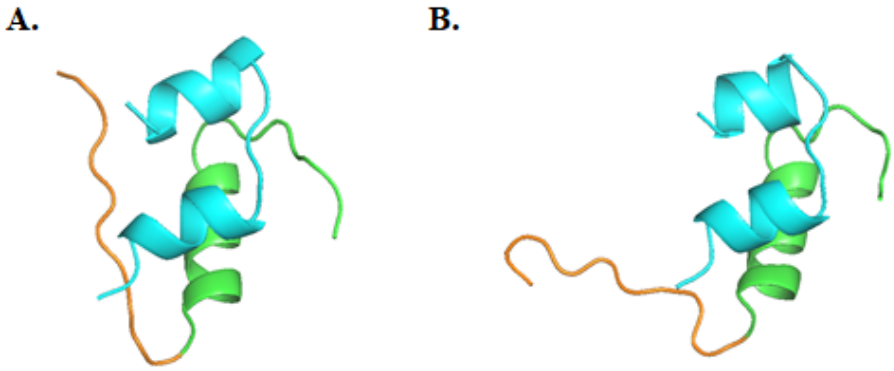
[Fig.16]:

Representation of how site1 and site2 appear in cryo-EM structure. Site1 is highlighted in blue and there are no changes compared to previous suggestions. Site2 is highlighted in red and the main difference in that FnIII-2 domain is not part of site2 (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source Protein Data Bank structure 6CE7: Scapin *et al*. 2018).

2.3.3 Binding Properties and Conformational Changes upon Ligand Binding

Insulin receptor dimer is able to bind just one insulin monomer (A/B chain) with high affinity because insulin binding leads to a conformational asymmetry of the receptor - negative cooperativity. It is connected with existence of two binding pockets, one binding pocket is formed by site1/2' and the other by site1'/2, which are placed antiparallel at both sides of insulin receptor dimer. Binding of insulin to one of the binding pockets weakens binding to the second potential binding pocket in the receptor (De Meyts 1994).

Conformational change occurring in insulin upon binding to the receptor is detachment of C-terminus of B-chain, specifically of residues from Phe25 to Thr30 [Fig.17].



[Fig.17]:

A. Insulin structure with undetached C-terminus of B-chain (orange), A-chain is blue, rest of B-chain is green (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source Protein Data Bank structure 1MSO: Smith *et al*. 2003).
B. Insulin structure with C-terminus of B-chain (orange) in detached conformation, A-chain is blue, rest of B-chain is green (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source Protein Data Bank structure 6CE7: Scapin *et al*. 2018).

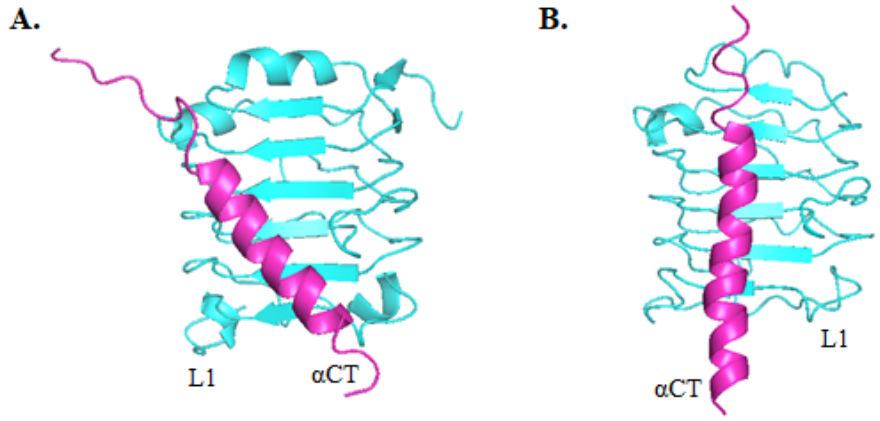
Result of the detachment is uncovering of first three residues in N-terminus of A-chain, which have hydrophobic side chains and then are able to interact with insulin receptor (Hua *et al*. 1991). Besides these residues also others belonging to hydrophobic core are unmasked, such as Tyr19, Cys20 in A-chain and Leu11, Val12, Leu15, Cys19 in B-chain (Ward *et al*. 2007). Tyr19 is then able to form hydrogen bond with Gln5 (also in A-chain), which contributes to stabilization of the hydrophobic core (Jiráček *et al*. 2010).

Upon formation of activated insulin another important feature in C-terminus of B-chain appears, namely the β-turn (between residues Gly20-Gly23) and the β-strand formed between residues Phe24-Thr27. After β-strand is detached, there is a motif of aromatic amino acids Phe24, Phe25 and Tyr26 interacting with receptor domains in site1. Thus C-terminus of insulin B-chain in detached conformation is crucial for biological activity, whereas its undetached conformation is important for folding and self-assembly (Menting *et al*. 2014).

On the other side, N-terminus of B-chain undergoes conformational changes as well. As we mentioned already, B-chain N-terminus can adopt several states (T/R/I). Insulin residues playing a role in this transformation are Asn3, His5 and Gly8 (in B-chain), because they are involved in transition from T-state to R-state. This is supported by mutagenesis studies which showed that residue 8 is also important for receptor binding. It is suggested that insulin receptor could possibly again play a role as a chaperone in transition of insulin from T- to R-state. (Kosinová *et al*. 2014).

Another conformational change occurs in α-CT segment, which is structural element important for insulin binding and it is part of site1 interface. Its conformational change is

coupled with detachment of insulin C-terminus, when there are exposed hydrophobic amino acids and they can interact with α-CT (Whittaker *et al*. 2012). If there is no insulin bound to the insulin receptor (apo-receptor), α-CT segment positioned over central β-sheet of L1 domain [Fig.18A]. α-CT in such state could not allow insulin to bind because of steric clashes and that it why it was suggested that displacement of α-CT is needed (Smith *et al*. 2010). Recent structures of insulin receptor with bound insulin proved this right and it was observed that C-tail of α-CT is displaced from its apo-conformation in order to allow detachment of C-terminus of insulin B-chain [Fig.18B] (Menting *et al*. 2014).



[Fig.18]:

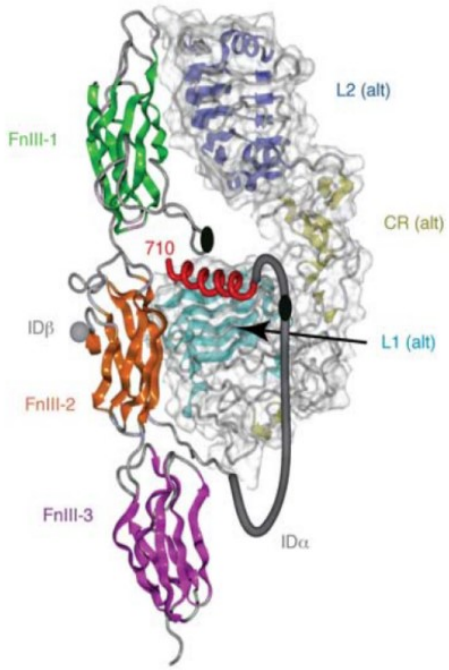
A. Conformation of αCT-segment (magenta) above L1 domain (blue) in insulin apo-receptor (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source PDB structure 4ZXB: Croll *et al*. 2016).
B. Conformation of αCT-segment (magenta) above L1 domain (blue) in insulin receptor when insulin is bound (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source PDB structure 6CE7: Scapin *et al*. 2018).

To fit insulin into binding pocket it is required to change also conformation of the monomeric leg containing three fibronectin domains. There would be large steric clash of insulin site2 surface with FnIII-1 and FnIII-2 domain without their displacement. Since moving away first and second fibronectin domain has to modify conformation of third fibronectin domain as well, which is joined with transmembrane helix, it is proposed to be the sequence of conformational changes which leads to intracellular tyrosine-kinase activation and downstream signalling (Menting *et al*. 2013). More detailed insight into how fibronectin domains could rearrange and thus how site2 surface of insulin receptor could look like upon insulin binding is contained in chapter “Structures of Insulin Receptor with Bound Insulin”.

2.3.4 Structures of Apo-form of Insulin Receptor

The first insight into insulin apo-receptor structure was provided in year 2006 using X-ray crystallography (in Protein Data Bank https://www.rcsb.org/, PDB, 2DTG structure: Mckern *et al*. 2006). This crystal structure had resolution of 3.8Å and comprised ectodomain of isoform IR-A. It revealed that insulin receptor exhibited an inverted "V" conformation. In addition, it elucidated that L1 domain is directed to the central part of receptor and two L1 domains of one dimer are distant ± 65Å from each other. Authors suggested that according to this structure, site2 region could be located in loops of FnIII-1 domain. ID domain with αCT segment was not included because of disorder. Further, it was speculated that movement of L1-CR domains of one monomer would also cause movement of FnIII-2/FnIII-3 domains, which could result in activation of intracellular kinase and signaling cascade (Mckern *et al*. 2006).

Structure 2DTG was later revised and replaced by structure with code 3LOH (Smith *et al*. 2010). The authors used the same data as for structure 2DTG, but enhanced their interpretation with new methods. Due to this approach, αCT-segment was resolved and it was shown that it consists of α-helix placed over central β-sheet of L1 domain. Also another feature of αCT segment was suggested, namely the assembly in *trans*. This means that αCT-segment of one monomer is supposed to be located over L1 domain of the second monomer and not over L1 domain of the one monomer from which it originates. In that only αCT part of ID domain was resolved, it remained still unclear how rest of the ID domain is arranged [Fig.19]. And even though conformation of the αCT in apo-receptor was shown, its conformation after insulin binding was not known (Smith *et al*. 2010).

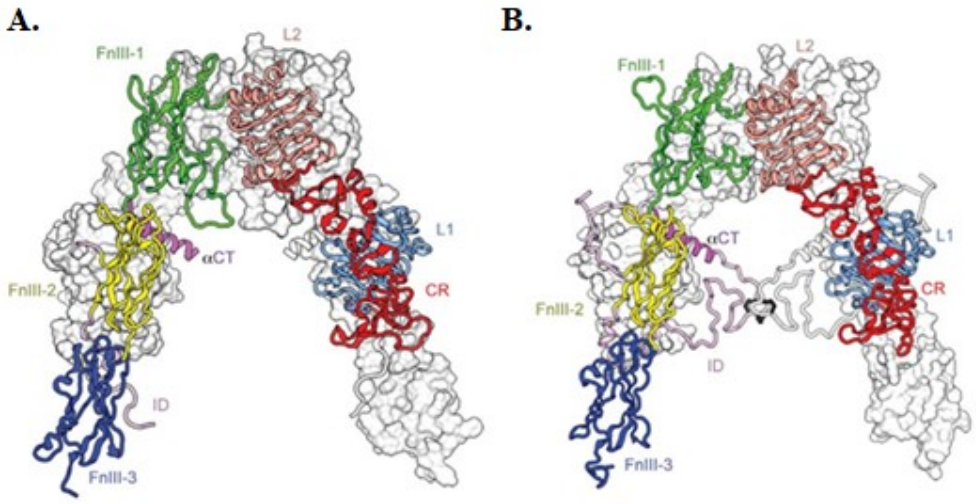


[Fig.19]:

There is shown *trans*-assembly of αCT-segment, because it is from FnIII-2 domain of one monomer, but it is placed along L1 domain of alternative monomer of the same dimer. Gray tube represents unresolved residues of ID-domain (adopted from: Smith *et al*. 2010).

In 2016 structure 3LOH was replaced by new refined structure as well generated by X-ray crystallography. This structure with code 4ZXB (Croll *et al*. 2016), which is still up-to-date, has 3.3Å resolution. 4ZXB includes revised ID domain, FnIII-1 and -3 domain and some of N-glycans were also added. Particular changes in FnIII-1 domain involve assignment of some residues to correct loops. Interesting is that one FnIII-1 loop is displaced here by more than 10Å in comparison with 2DTG and 3LOH. In FnIII-3 domain, some loops were also corrected to contain proper residues and one loop was remodeled into α-helix. Residues 638-656 in ID domain adopted a hairpin-like turn and were located near the L1 domain. Revision of the structure allows modeling of other missing parts in ID domain and authors made a more complete model named ModelS1 [Fig.20]. The structure also shows new features upon ligand binding, in particular residues Arg454, Glu453 in L2 domain and residues Asp496, Arg498, and Asp499 in FnIII-1 domain are able to create electrostatic interactions upon rotation of domains, when they could get into contact with each other. Another residues in L2 domain in apex of receptor dimer, residue Asp464, forms carboxylate-carboxylate pair with Asp464‘ of the other monomer. This feature could be responsible for a pH-dependent conformational switch. Therefore, it is suggested that a high-affinity insulin binding is associated with

deprotonation of the residue Asp464 in protonated carboxylate pair in the apo-protein (Croll *et al*. 2016).



[Fig.20]:

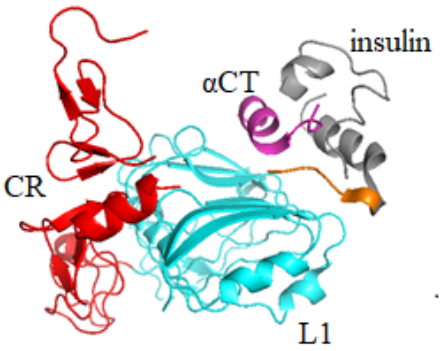
A. Obsolete structure 3LOH, which is missing part of ID-domain.
B. ModelS1 - the most complete structure of insulin apo-receptor up to date. The most significant difference compared to 3LOH, is resolving of ID-domain
(adopted from: Croll *et al*. 2016).

2.3.5 Structures of Insulin Receptor with Bound Insulin

The first structures where insulin was bound to a part of insulin receptor were determined in 2013 using X-ray crystallography. A complex with resolution of 3.9Å (PDB code: 3W11) (Menting *et al*. 2013), contains human insulin with truncated insulin receptor (domains L1-CR and αCT segment). The authors made also other three complexes varying in whether it contains insulin analogue or bovine insulin, but with worse resolution. Structure 3W11 showed for the first time how site1 binding interface looked like. Furthermore, it was the first time when the αCT segment was observed to be displaced from its apo-form. However, B-chain of insulin was not complete, because 9 residues at C-terminus were missing (Menting *et al*. 2013).

A year later, in 2014, the same structure of truncated receptor and human insulin as in the previous case was characterized by X-ray crystallography but now with improved resolution of

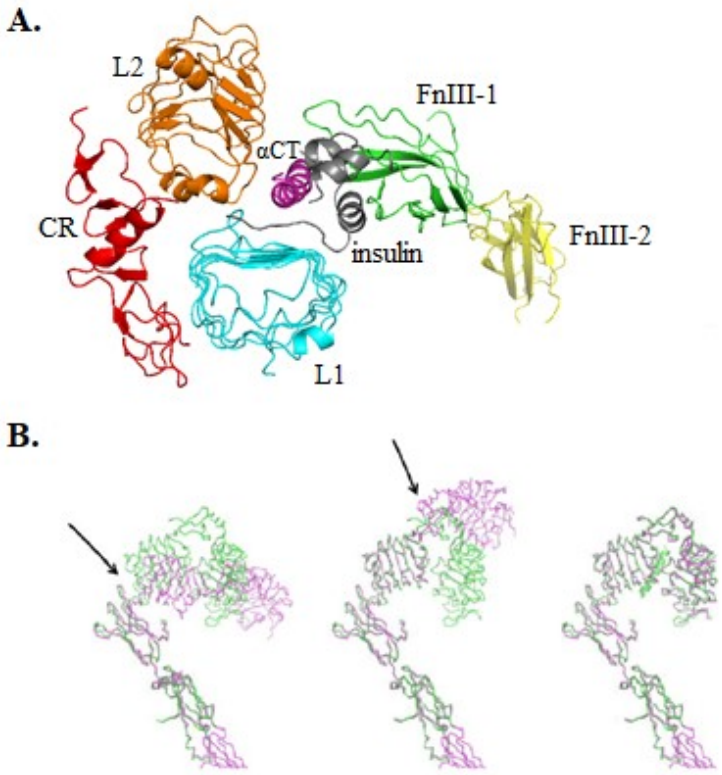
3.5Å. This structure with code 4OGA offered better insight into detachment of C-terminus of insulin B-chain [Fig.21], which was not resolved in structure 3W11. However, last three residues (Pro28, Lys29, Thr30) of B-chain C-terminus were not included (Menting *et al*. 2014).



[Fig.21]:

Structure 4OGA: L1 domain is blue, CR domain is red, αCT-segment is magenta and insulin is gray (both A-chain and B-chain) with C-terminus highlighted with orange colour
(Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source PDB structure 4OGA: Menting *et al*. 2014).

Another significant breakthrough in determining structure of insulin-bound insulin receptor was achieved by cryo-electron microscopy. Structure 6CE7 (Scapin *et al*. 2018) with resolution of 4.3Å contains besides domains L1, CR, L2 and αCT also first and second fibronectin domain, therefore it brought the first structural view into site2 interface. Site2 interface of insulin receptor in the cryo-EM structure differs from the predictions. In this structure, the site2 is not formed by the FnIII-2 domain but rather by only two loops of FnIII-1 domain (residues Pro495-498 and Arg539-Asn541) and in addition also L2 domain (residues Arg372-Tyr374). Insulin residues being part of site2 are following: Glu4-Thr8 and Ile 10, all in A-chain [Fig.22A]. Insulin receptor in this structure undertakes large conformational changes as compared to apo-receptor structure 4ZXB, because receptor in insulin-bound state has not the shape of inverted "V" anymore. The conformational changes involve rotation of whole L1-CR-L2 leg of monomer and then protrusion of CR-L1 segment [Fig.22B] (Scapin *et al*. 2018).



[Fig.22]:

A. Cryo-EM structure of insulin receptor with bound insulin. L1 domain is blue, CR domain is red, L2 domain is orange, FnIII-1 domain is green, FnIII-2 domain is yellow, αCT is magenta and insulin is gray (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source PDB structure 6CE7: Scapin *et al*. 2018).
B. Schematic representation of how insulin receptor changes its conformation upon ligand binding. It is compared to insulin apo-receptor 4ZXB (purple), the first arrow points to rotation of L1-CR-L2 leg and the second arrow points to CR-L1 protrusion
(adapted from: Scapin *et al*. 2018).

The most recent insulin receptor structure (PDB code 6HN5: Weis *et al*. 2018) in insulin-bound conformation was also obtained by using cryo-EM and has resolution 3.2Å in sphere, where insulin is bound. Region adjacent to a membrane, including FnIII-2 and FnIII-3 domain, has resolution 4.2Å, which is result of this region flexibility. Conformational changes are very similar to the previous cryo-EM insulin receptor structure as receptor is not in inverted "V" shape, but a main difference is that this newer structure is including also second and third fibronectin domain, which offer new details of signal transduction. FnIII-3 domain from one monomer contacts FnIII-3' domain from other monomer triggering movement of intracellular tyrosine kinase domains towards each other and activating trans-phosphorylation [Fig.23]. When compared to the previous cryo-EM, there is also the same aspect of site2, where only residues His10 and Glu13 in insulin B-chain interact with FnIII-1 domain of insulin receptor.

Other insulin residues predicted to be part of site2 interface as Ser12, Leu13, Glu17 in A-chain or Leu17 in B-chain are not in contact with insulin receptor, whereupon it was suggested that these residues contribute in docking of insulin to insulin receptor, which then allows insulin binding per site1 interactions (Weis *et al*. 2018).

[Fig.23]:

FnIII-3 domain from one monomer (dark blue) is brought into contact with FnIII-3'domain from another monomer (light blue) of the same insulin receptor dimer
(adopted from: Weis *et al*. 2018).

2.4 Basics of Structural Biology Methods

2.4.1 X-Ray Crystallography

X-ray crystallography is structural biology method, which is currently the most widely used approach to obtaining 3D information about macromolecular structures. There are three main steps leading to successful structure determination: crystallization, X-ray exposition and diffraction data analysis.

The key of the crystal forming rests upon high concentrations of purified protein and adjustment of proper protocol for crystal growing. The principle of crystal growth lies in protein solution of high concentration, which is activated in a way that crystals start go out of solution (Smyth and Martin 2000). However, solubility itself can cause problems, because not every protein is soluble, actually there are approximately 6000 proteins, which are insoluble. Another challenging protein category are membrane proteins requiring lipidic cubic phases or detergents, but after all the results may not be favourable (Zheng *et al*. 2015). Another factor on which depends successful analysis is crystal size and it is suggested to have a crystal long at least 100nm in the widest dimension.

Once protein is crystalized it can be moved ahead toward exposition of crystal to X-rays, which are organized into beam. The incident rays are diffracted and a diffraction pattern is generated. There are two common ways how to produce X-ray beam, the first approach includes machine called “synchrotron” enabling electron acceleration and the second approach is based on a copper anode, which is hit by electrons.

The diffraction pattern is subsequently analyzed using crystallographic software. The application of specific algorithms results in a calculated electron density map, into which a molecular model is fitted and refined. The final model with best agreement with the electron densities is deposited with the Protein Data Bank (https://www.rcsb.org/).

The major limitation of X-ray crystallography usage is difficulty to crystalize proteins, which contain flexible regions or undergo conformational changes (Smyth and Martin 2000). The other limitations are post-translational modifications, e.g. glycosylation, acetylation or methylation (Zheng *et al*. 2015).

2.4.2 Cryo-Electron Microscopy

Cryo-electron microscopy (cryo-EM) represents another method of macromolecular structure determination. For the past tens of years was the method improving until currently enabling near atomic resolution and even being able to provide structures of the flexible macromolecules, which offers an advantage over X-ray crystallography. Another significant advantage is that it is not required to have sample of that high concentration, as in X-ray crystallography, but it is sufficient to have smaller amount or even one single protein particle. There is also no need of crystallization, because it is possible to determine the structures under more native conditions, with their dynamic properties as well (Murata and Wolf 2018).

The principle of cryo-EM is observation of the samples in electron microscope. The first step is to administer purified solution of the macromolecules to a grid. Surface of the grid contains holey carbon coating and it enables to form a layer of the solution in the holes. The grid is subsequently immersed in a substance cooled by liquid nitrogen, thus are the molecules of interest frozen, what conserves their structure. The macromolecules frozen after this manner are present in several different orientations and images of that orientations are produced by electron microscope applying a beam of electrons. These images represent only 2D projection, however, by alignment of 2D-images to each other is obtained 3D projection of the structures. The 3D projection in the case of cryo-EM is not electron density map, but it is Coulomb potential density map, although, their interpretation used to be the same (Cheng 2015).

3 Aims of the Thesis

- **The overall goal of the thesis is to gain understanding into the structural and energetic determinants of insulin binding to insulin receptor. It entails three sub-goals:**
    - to model representative structures of insulin receptor as well as insulin and insulin analogues, which could be used for studying of binding interactions, especially the largely unknown site2 interface;
    - to perform molecular dynamics with wild-type insulin and insulin analogues in order to observe if the mutations in the analogues have an effect on conformational changes necessary for binding to insulin receptor;
    - to test model structures of insulin receptor via molecular docking with insulin to find out whether they are able to reproduce the original binding pose by using the previously described interactions between insulin and insulin receptor;

4 Materials and Methods

4.1 Materials

The material in this project is represented as original or modified Protein Data Bank (PDB) structures (https://www.rcsb.org/).

4.2 Methods

4.2.1 Molecular Modelling

The term molecular modelling covers a wide spectrum of approaches and possible applications, but in general it is a usage of theoretical chemistry knowledge via computer graphics and calculation to create new model structures or studying noncovalent interactions. Creating of the new models can include, for example, adding missing parts (residues, loops, domains), changing conformation or residue substitution.

To transfer a list of atoms into 3D-structure in some of the graphical programs, it is required to have the Cartesian (x,y,z) coordinates of the atoms in the particular system (Leach 2001). These coordinates provide information about the atom positions. Sometimes it is preferred to have internal coordinates instead of Cartesian coordinates, by reason of including relative positions of atoms to each other, such as bond lengths, valence angles and torsion angles (Kornilov *et al*. 1982). If the purpose is to continue working with the structure in molecular dynamics, the preference is Cartesian coordinates, but it is possible to convert coordinates if necessary.

Another important component of the system are van der Waals spheres of the atoms, which form van der Waals surface of the molecule. When an atom interferes with van der Waals sphere of another atom of the same or different molecule, it is called contact surface. A deeper interference into van der Waals surface of another atom denotes steric clash in the structure. It is important to keep in mind that bringing changes into structure, for example by rotating the

bonds of atoms or moving them in space, it also changes the energy of the conformation (Leach 2001).

There are multiple graphical programs for molecular visualization and modelling. Here, we use PyMOL (The PyMOL Molecular Graphics System, Version 2.0.7 Schrödinger, LLC., https://pymol.org/2/) and Modeller (version 9.21; https://salilab.org/modeller/). Molecular modelling in PyMOL can be done via its graphic interface if mode "Editing" is turned on. Tool "Mutagenesis" allows to substitute a residue by another one, in "Builder" interface is possible to add new atoms or amino acids, creating or deleting bonds and lot of other options. The program Modeller can be used online as webserver ModLoop (https://modbase.compbio.ucsf.edu/modloop/) by uploading a structure and then defining the residues, which can be moved in order to form loops.

4.2.2 Molecular Dynamics

Molecular dynamics (MD) is a method, which is able to bring an insight into connection between structure and function of macromolecules. The structure of macromolecules is connected with interactions, which are occurring in there and the molecular interactions have an impact on biological function of the macromolecules. Proteins, especially, may experience huge conformational changes to be able execute their activity and their dynamic features can be necessary to understand principle of the function. Molecular dynamics offers the possibility to take three-dimensional structures and use them for MD simulations in order to observe their action based on molecular interactions.

Before running an MD simulation itself, it is required to prepare structures. The preparation of the structures includes adding hydrogens, minimizing energy, adding solvent molecules and counter ions.

To carry out MD simulations, we need a force field which contains parameters for bonded (bond lengths, bond angles, bond torsions) and nonbonded (electrostatic and van der Waals) terms. Using the Newton's equations of motion, we obtain trajectories, i.e. time series of the changes of the positions of atoms under the force field.

Another of the key settings is also solvent representation, where one can choose explicit or implicit solvent. The explicit solvent represents a more natural model of solvent molecules,

e.g. to ensure hydrophobic effect, but it makes the MD simulations more time-consuming. The implicit solvent, on the other hand, is treated as a continuum, which is fast, captures the gross features of solvent but lacks the description of specific water bridging.

Currently, multiple programs can be used for producing and analyzing MD simulations: NAMD (Nelson *et al*. 1996; http://www.ks.uiuc.edu/Research/namd/), AMBER (Case *et al*. 2005; http://ambermd.org/), GROMACS (Hess *et al*. 2008; http://www.gromacs.org/) and CHARMM (Brooks *et al*. 2009; https://www.charmm.org/).

We will look more closely into program AMBER (Assisted Model Building with Energy Refinement; version 14 and 16) which was used in this project. AMBER is comprised of several codes, which can be divided into three categories: system preparation, molecular dynamics simulation and trajectory analysis. For the system preparation there are the codes under name "LEaP" providing adding hydrogen atoms, solvation of the system and creating topology files with force field parameters. As input to LeaP, PDB files can be used and there are two types of LEaP outputs, coordinate file (.rst) and parameter topology file (.top). The parameter topology file involves names and masses of atoms, there is also register of bonds, angles, and dihedrals as well as the force field parameters, while in the coordinate file are included only Cartesian (x,y,z) coordinates of all atoms in the system (Case *et al*. 2005). There are several choices of the force fields and I used the "ff14SB" protein force field (Maier *et al*. 2015). The most commonly used explicit solvent in AMBER is water model TIP3P (Price and Brooks 2004), applied also in this project as pre-equilibrated box TIP3PBOX, which was executed with command "solvateoct" building truncated octahedron of solvent around the macromolecules (AMBER 2016 Reference Manual; http://ambermd.org/doc12/). On choice of the water model depends also choice of ions, which are added to the structure in order to neutralize system to achieve the total concentration of 0.15 M. In my system with TIP3PBOX water model is used ion model for TIP3P with Na^+ and Cl^- particularly (Joung and Cheatham 2008). Another step towards accelerating MD simulations was used via hydrogen mass repartitioning (HMR), which resulted in changed mass of each hydrogen to 3 Da (AMBER 2016 Reference Manual; http://ambermd.org/doc12/).

The codes running molecular dynamics simulations in AMBER are called "sander" and "pmemd". We will focus more on the program pmemd, which is revised version of sander in that way it provides benefit of parallelization. This program delivers to each processor an information about atom coordinates, however, not all of the processors have information about all coordinates, but only about the atom coordinates of particular section, which is submitted to the specific processor. This is the time-saving difference from sander, where all processors have

information about all atom coordinates (Case *et al*. 2005). Performance of the MD in pmemd was even more boosted with version "pmemd.cuda", when it was modified in CUDA (Compute Unified Device Architecture) language in order to allow usage of graphics processing units GPU (Salomon-Ferrer *et al*. 2013). In this project, the MD simulations were performed on GPUs using pmemd.cuda. At the beginning of MD, the system was heated to a desired temperature of 300 K. The first MD simulation was 300ps long. After achieving the desired temperature, the MD run was further extended to 100 ns and finally, I did other nine runs of 100 ns MD, thus we obtained 900ns long MD (plus with the initial 100ns it is 1μs together).

After obtaining trajectory files from the MD simulation program, the next step is trajectory analysis in AMBER with codes under name "ptraj". It is necessary operation, because the output files from MD simulations are large and one would like to reduce them, for example by stripping out the solvent when it is not required for further analysis, as in our case. Another application of the program ptraj is when there are several output files from different MD runs and there is a need of compiling them into one output file. Besides this, ptraj offers including of another commands depending on user's requirements, thus it is a platform allowing wide range of assignments (Case *et al*. 2005). Version of ptraj, made in coding language C++, which is called "cpptraj" owns feature not included in classic ptraj, such us using multiple topology files at once (Salomon-Ferrer *et al*. 2013), what I used as well.

4.2.3 Molecular Docking

Molecular docking is a method, which generates 3D structures of complexes between molecules and their ligands (binding poses) and estimates the binding affinities (scores). This approach (as part of the virtual screening protocols) is very useful in drug discovery and structure-based drug design.

Docking of two macromolecules (protein-protein docking) can be performed in program HADDOCK (High Ambiguity Driven protein−protein Docking; Dominguez *et al*. 2003). The HADDOCK varies from the most of docking programs in principle, on which they are based. In the other programs is one protein usually in rigid state, what means it is stabilized in space, and the second protein is flexible and moving around the rigid protein. After that, for all poses is calculated the score (based on the Van der Waals repulsion, electrostatic interactions and also on complementarity of the protein surfaces), but this approach is more computationally

demanding and does not offer sufficient flexibility. On the other side, the HADDOCK makes advantageous use of experimental data obtained e.g. from nuclear magnetic resonance or mutagenesis. The experimental data offering information about which residues of both proteins are interacting with each other are input as so-called “ambiguous interaction restraints” (AIR). There are two types of input residues playing role as AIR: active and passive residues. The active residues resolved by mutagenesis are those, which was proved to be important for binding and are exposed to the solvent. As the passive residues are usually chosen those, which surround the active residues and have limited exposure to the solvent (Dominguez *et al*. 2003). The HADDOCK program exists also as webserver, which was used in this project. In the HADDOCK webserver (version 2.2; https://milou.science.uu.nl/services/HADDOCK2.2/haddock.php) are submitted two protein structures and for both of them has to be specified the active residues. The passive residues can be also specified or they can be automatically defaulted by the program, which picks those surrounding the active (Van Zundert *et al*. 2016).

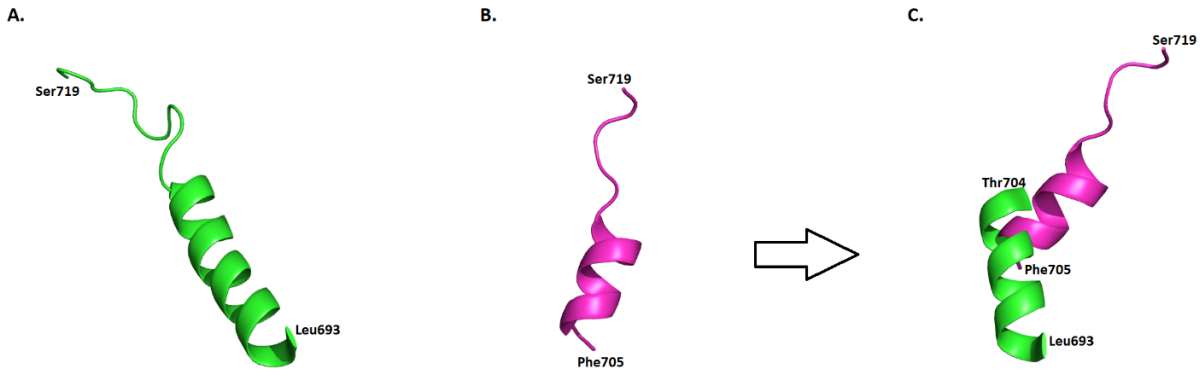
5 Results

5.1 Molecular Modelling of Insulin Receptor

5.1.1 Modelling based on Crystal Structure

There is a limited number of insulin receptor structures and they are, generally, incomplete. The most representative structure of insulin apo-receptor has PDB code 4ZXB (Croll *et al*. 2016), although there is even more complete structure derived from 4ZXB, but it is a model which is not present in Protein Data Bank. This model structure is originally called "ModelS1" and can be found in supplementary section of the paper referring to 4ZXB (Croll *et al*. 2016). Because ModelS1 contains additional features not included in 4ZXB (such as modelled ID-domain) it became my initial structure. At that time, there was no insulin receptor structure with bound insulin, where would be also present fibronectin domains, thus no closer insight into site2 was available. There was, at least, truncated insulin receptor with bound insulin (PDB code 4OGA; Menting *et al*. 2014), which shows conformational changes in αCT-segment of the receptor and in insulin B-chain upon binding.

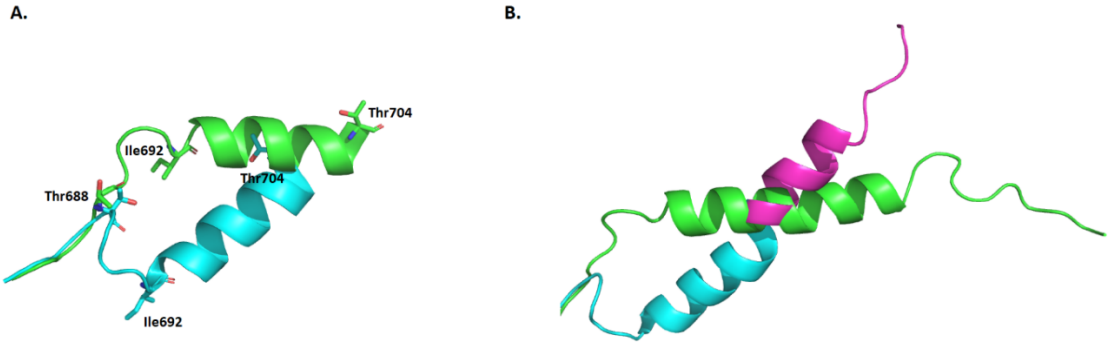
The aim here was to model sufficiently complete insulin receptor structure in state of bound insulin using existing insulin apo-receptor structure and truncated insulin receptor with bound insulin. ModelS1 includes both identical monomers of the insulin receptor dimer, but changes were made only in one half of the dimer, which involves L1-CR-L2 part of the one monomer and αCT-FnIII-1/2/3 part of the second monomer. One of the major differences visible between those two structures (ModelS1 and 4OGA) is displacement of the αCT-segment in 4OGA. The first step was to remodel αCT-segment in ModelS1, but the αCT-segment in 4OGA is truncated and consists only of the residues from Phe705-Ser719, while the αCT-segment in ModelS1 is complete (Leu693-Ser719). Thus, the approach was to substitute whole region from Phe705 to Ser719 in ModelS1 by this region from 4OGA, but as it can be seen [Fig.24] this is not enough to get a qualitative structure. The replacement by 4OGA region was made in text mode of ModelS1 PDB-file, where corresponding residues were removed and 4OGA residues were added in their place via "cat" command in Linux.



[Fig.24]:

A. αCT-segment from ModelS1, in state when insulin is not bound.
B. αCT-segment from 4OGA, in state of bound insulin. In this structure is helix truncated and starts with residue Phe705.
C. Green: Cut αCT-helix from ModelS1 (containing residues up to Thr704); magenta: truncated αCT-helix from 4OGA replaced cut part of the helix from ModelS1 and it is in its original insulin bound state (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014).

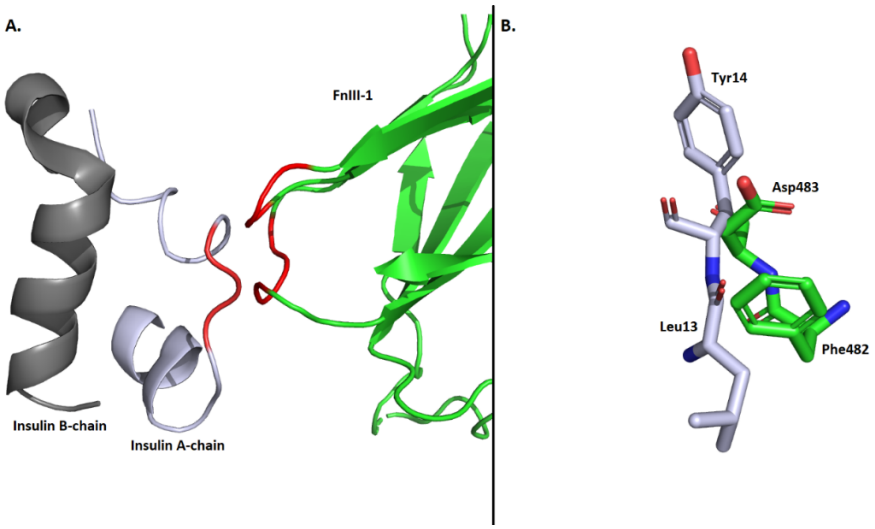
It is necessary to change location of the region downstream of the residue Phe705 (in ModelS1) in order to connect the rest of αCT-helix (Ile692-Thr704) with αCT-helix from 4OGA. It is not allowed to move with helical part from 4OGA, because it is the state occurring when insulin is bound, thus I had to adjust the rest of the helix to it. However, we did not touch the particular helical part (Ile692-Thr704), because I did not want to affect a continuity of the helix and its interactions. I moved only with residues from Thr688 to Gln691, which form loop behind the helix and serve as good tool for moving the helical part without disrupting it. This approach resulted into connected helical parts αCT-segment from ModelS1 and 4OGA [Fig.25].



[Fig.25]:

A. Alignment of the original (green) and newly modeled (blue) conformation of αCT-helical part Thr688-Thr704. Behind residues Thr688 are these structure identical, thus are overlapping. After the residues Thr688 is different conformation, residues Thr688-Ile692 are not part of helix, so it was moved with them and it caused movement of the helix as well. The new conformation (blue) is the one, which enables connection with helical part from structure 4OGA.
B. The newly modeled αCT-segment in full size, where blue part is remodeled ModelS1 and magenta part is added from 4OGA. The original conformation of αCT-segment from ModelS1 (green) is here for comparison (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source is own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014).

Another step towards modeling insulin bound conformation of the insulin receptor model is to add insulin into binding pocket, but since ModelS1 is apo-receptor, there is not enough space in the binding pocket because of steric clashes of insulin A-chain (part between helixes) and the first fibronectin domain [Fig.26].



[Fig.26]:

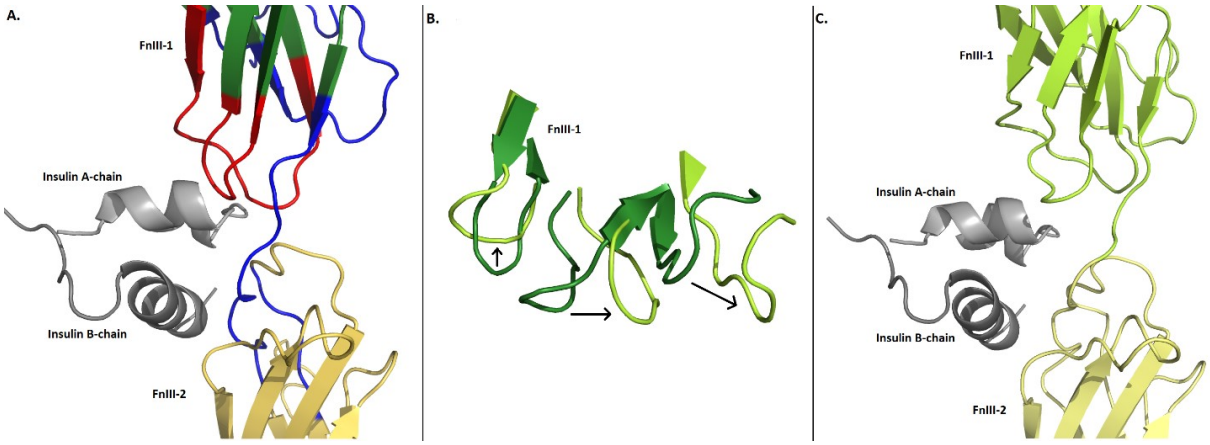
A. Red coloured of the loops of FnIII-1 domain (green) and insulin A-chain (light gray), where the critical sites of steric clashes are.
B. The specific example of insulin A-chain residues (light gray) and FnIII-1 residues (green), which are too close to each other. In case of Tyr14 (insulin) and Asp483 (FnIII-1) is there also significant overlay (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source is own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014).

The solution is to remodel loops in the first and second fibronectin domain (FnIII-1 and -2) to widen binding pocket and allow adding of insulin from structure 4OGA via "cat" command in Linux. However, not only the specific loops, which are close to insulin were remodeled, but also other loops in fibronectin domains in order to enabling a movement of the loops closer to insulin. The loops of FnIII-2 were moved also [Table.3], although, there is no direct steric clash with insulin B-chain, but it is more preferred to have residues not so close to each other because it could happen in other steps [Fig.27]. There are no steric clashes of αCT-segment with insulin B-chain anymore as it was displaced in the previous step. The result of the remodelling fibronectin domains loops is enlarged binding pocket with insulin from structure 4OGA without the steric clashes [Fig.27B/C].

Domain	Residues
FnIII-1	Arg479-Ile485
FnIII-1	Tyr507-Val531
FnIII-1	Gly550-Ala563
FnIII-1	Val588-Ala592
FnIII-2	Thr593-Pro601
FnIII-2	Lys614-Ser619

[Table.3]:

Table of the FnIII-1/2 residues, which were moved in order to enlarge binding pocket for insulin.

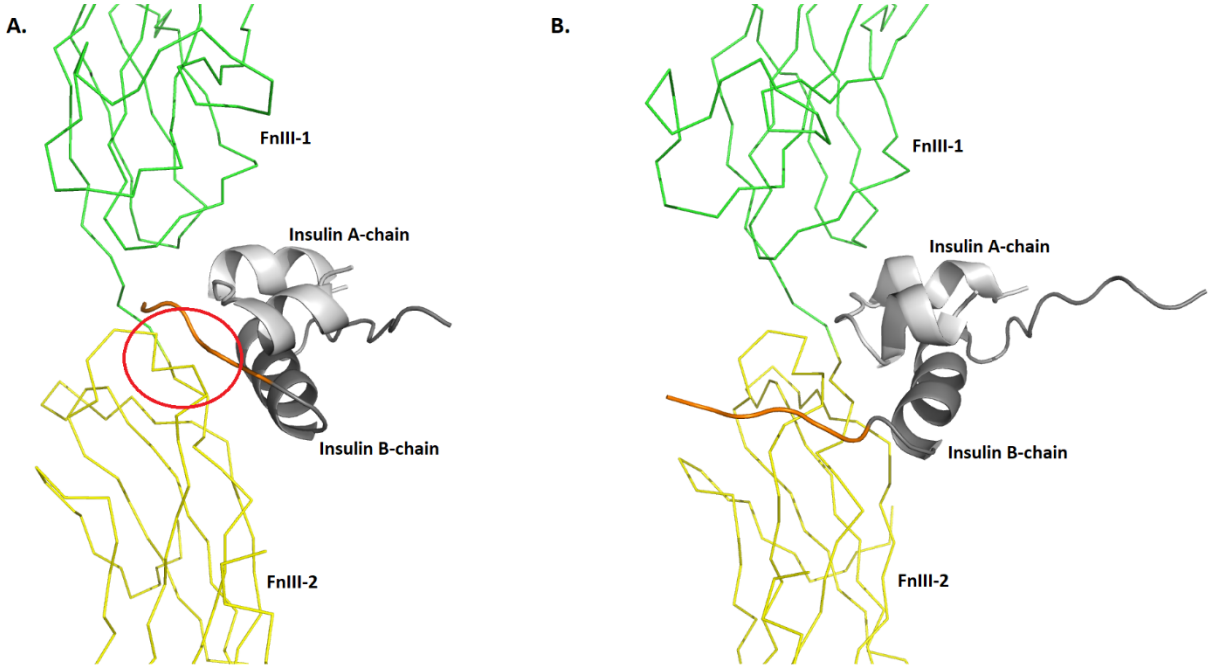


[Fig.27]:

A. Red loops (FnIII-1) are those in which were residues moved because of the steric clashes with insulin, blue loops (FnIII-1/2) were moved in order to allow moving of the red loops.
B. Alignment of the initial structure (dark green) with the newly remodeled structure (light green). The arrows indicate how the loops were moved from original position to new position.
C. View on the remodeled loops of FnIII-1 and FnIII-2 domain with insulin, when there are no steric clashes (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source is own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014).

Insulin added from the structure 4OGA needed to be also remodeled, because there were missing residues in N-terminus as well as in C-terminus of B-chain. In the C-terminus of B-chain were not included the last three residues Pro28, Lys29, Thr30 and since it is known which residues are there, it could be resolved easily only by adding these particular residues to the C-terminus of B-chain in PyMOL graphic interface. More challenging was to model N-terminus of B-chain, because there were missing first six residues (Phe1-Leu6) and as it was mentioned, N-terminus can obtain several states: R-, T- and I- (called also O) state. As the R-state was not observed to be present in active insulin (Kosinová *et al*. 2014), the choice was then T-state and there is insulin molecule in T-state in Protein Data Bank with PDB code 1MSO (Smith *et al*. 2003), in despite of it is with undetached C-terminus of B-chain it is not important, because I only needed conformation of N-terminus of insulin B-chain. Positions of the residues Phe1-Leu6 in B-chain of the insulin structure 1MSO were transferred into ModelS1 in way that I aligned 1MSO with insulin structure in remodeled ModelS1, then I moved with the insulin residues in remodeled ModelS1 to ensure connection with added residues and, finally, insert the new residues from 1MSO via “cat” command. However, there were steric clashes of the new N-terminus of B-chain with FnIII-2 domain [Fig.28A] and as I did not want to move with fibronectin domains anymore, I had to remodeled N-terminus of insulin B-chain and the final

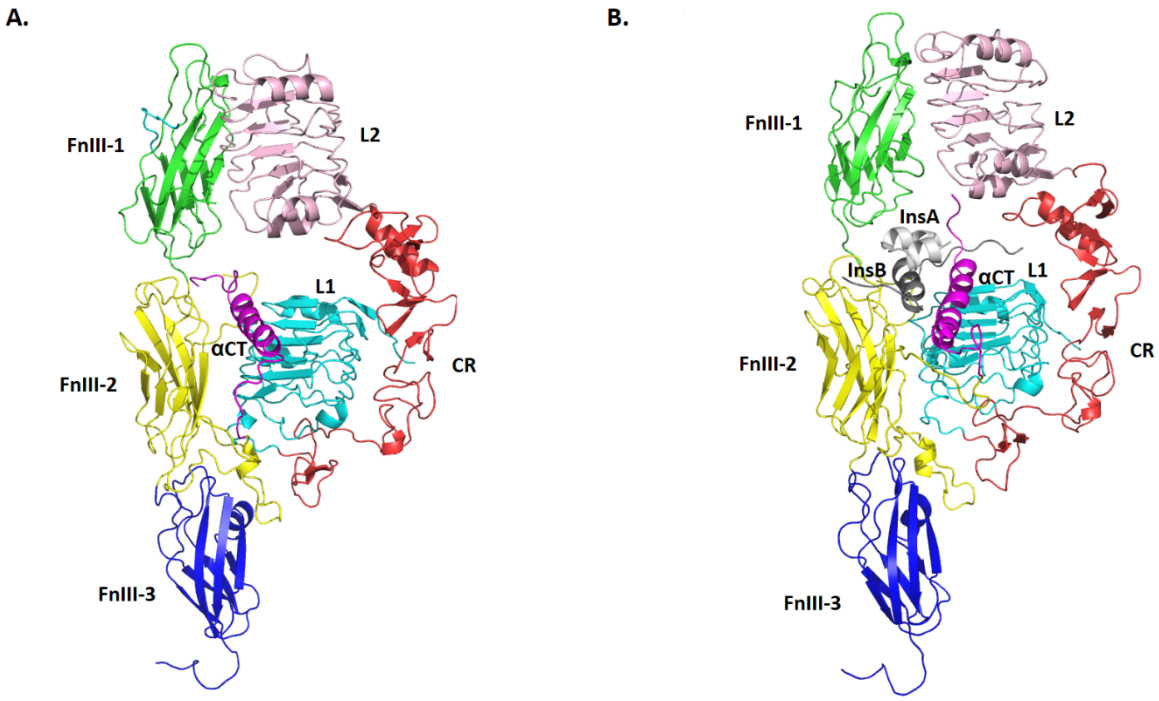
conformation of that terminus is not exact T-state, but it looks more as I-state conformation or it can be called as open (O) state [Fig28.B].



[Fig.28]:

A. There is insulin in T-state, where N-terminus of B-chain (orange) would have steric clash with the loop of FnIII-2 domain (red circle).
B. Remodeled N-terminus of insulin B-chain (orange) is now detached and has no steric clash with FnIII-2 domain (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source is own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014 and 1MSO: Smith *et al*. 2003).

In comparison to the original starting structure [Fig.29A], the final structure represents model of insulin receptor with bound insulin [Fig.29B], which has displaced αCT-segment in order to allow adding of insulin, which was also remodeled because of incompleteness, and the loops in FnIII-1 and FnIII-2 domains had to be remodeled as well.



[Fig.29]:

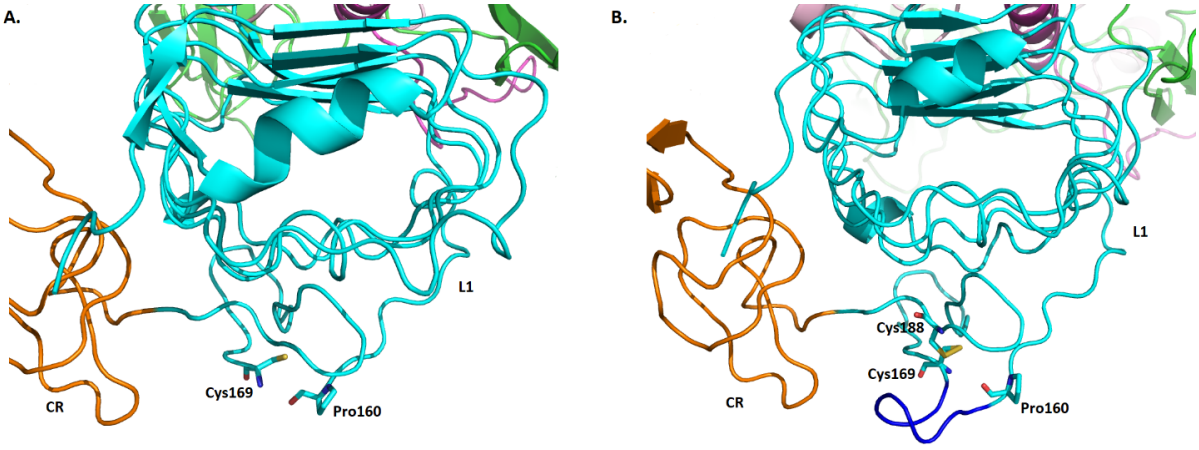
A. Side view of the original structure ModelS1, which represents insulin apo-receptor (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source ModelS1: Croll *et al*. 2016). B. Side view of the final remodeled structure, where is displaced αCT-segment, added remodeled insulin, moved loops of the FnIII-1/2 domain (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014 and 1MSO: Smith *et al*. 2003).

5.1.2 Modelling based on Cryo-EM Structure

The cryo-EM structure 6HN5 (Weis *et al*. 2018) was also incomplete and it was necessary to model several loops. For modelling was selected only the part of structure, which has higher resolution (3.2Å) and it is the part around insulin, including L1, CR, L2, FnIII-1 domain of one monomer, and αCT-segment, L2, FnIII-1 domain of another monomer. The approach for modelling was to add missing residues in PyMOL and some loops were also connected using program Modeller.

The first added loop was L1 domain loop of residues Gly161-Asn168 and these residues were taken from the structure of ModelS1 (Croll *et al*. 2016) in that way I added them into text format of PDB 6HN5 via "cat" command. I did not manually model connections between residues Pro160-Gly161 and Asn168-Cys169, but I submitted this structure in Modeller, where

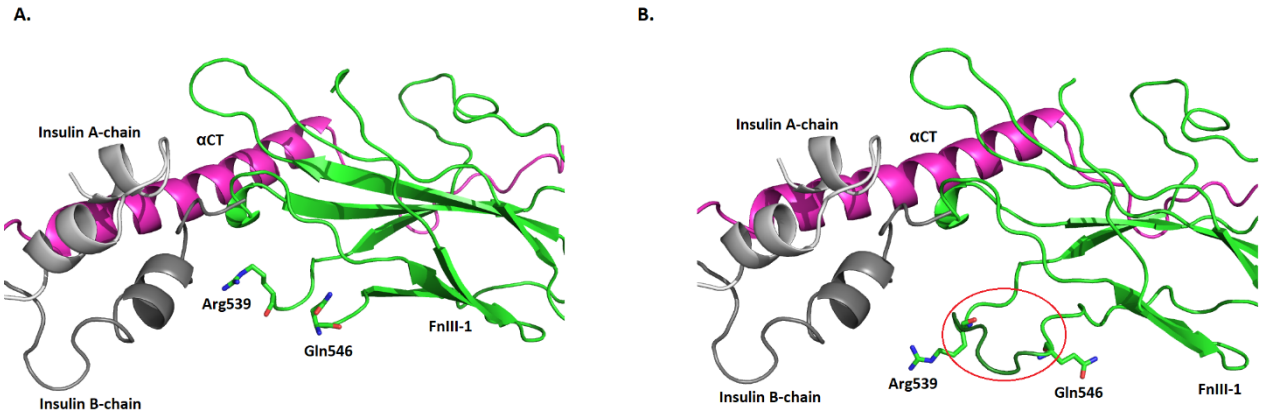
it was defined of which residues should be the loop generated, thus we defined residues Cys159-Pro170. From each end of the modeled loop was defined one residue in addition and that is the reason why it starts with Cys159 and ends with Pro170 – it ensures more adequate interconnection, if more residues from both sides are allowed to move. This loop was further remodeled in order to form disulphide bond between Cys169 and Cys188, thus it was moved with residues Pro160-Ile174 [Fig.30].



[Fig.30]:

A. The incomplete loop in L1 domain (cyan), where the gap is flanked by Pro160 on one side and Cys169 on the other side; CR domain is orange (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source PDB structure 6HN5: Weis *et al*. 2018)
B. The added residues are forming the missing loop (dark blue); CR domain is orange. Disulphide bond between Cys169-Cys188 is also represented (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on 6HN5: Weis *et al*. 2018 and ModelS1: Croll *et al*. 2016).

The same strategy was used for the loop in FnIII-1 domain, where missing residues Ser540-Ser545 were added from the recent cryo-EM structure 6CE7 (Scapin *et al*. 2018) also via "cat" command and then using program Modeller were connected the residues Arg539-Ser540 and Ser545-Gln546. The residues, which were defined as moving are from Leu538 up to Asn547. This loop of FnIII-1 domain is located close to insulin B-chain and can have important interactions with it, thus it is required to have this loop in proper conformation, which was not the one generated by Modeller. I had to remodel the loop manually in PyMOL graphic interface in way that it has similar conformation as in the structure 6CE7 [Fig.31]. Eventually, the moved residues include Pro537-Leu556.

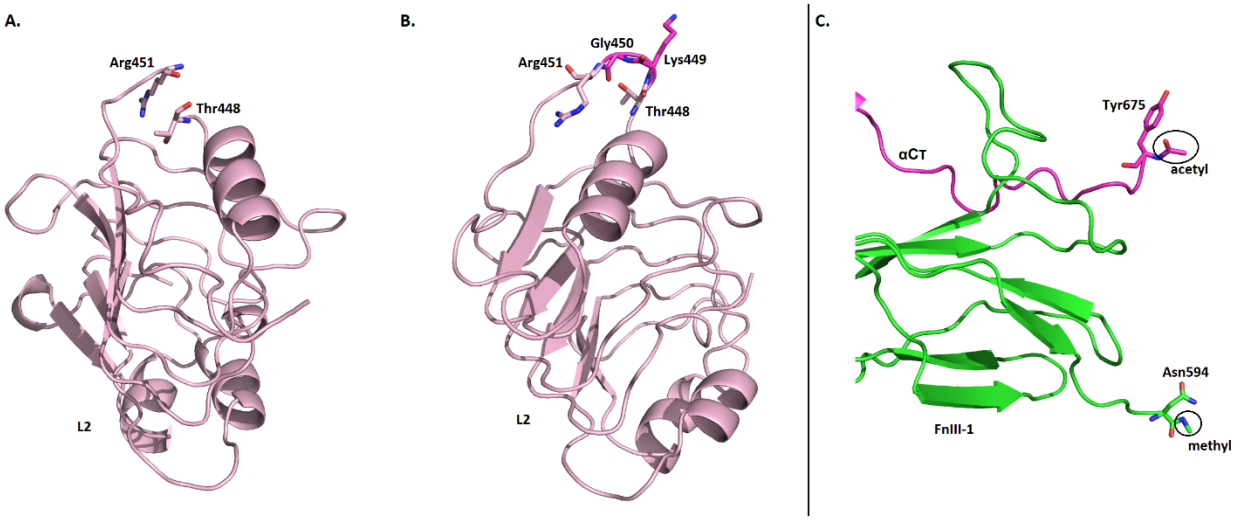


[Fig.31]:

A. The gap in loop of FnIII-1 domain, where are missing residues between Arg539 up to Gln546; αCT-segment is magenta, insulin A-chain is light gray, insulin B-chain is dark gray (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source 6HN5: Weis *et al*. 2018).
B. The modeled loop (dark green in the red circle) of FnIII-1 domain (green); αCT-segment is magenta, insulin A-chain is light gray, insulin B-chain is dark gray (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on 6HN5: Weis *et al*. 2018 and ModelS1: Croll *et al*. 2016 and 6CE7: Scapin *et al*. 2018).

Another missing residues in the structure 6HN5 were Lys449 and Gly450, which were added manually in PyMOL graphic interface, but for proper formation of the peptide bonds between the residues Thr448-Lys449 and Gly450-Arg451 was used program Modeller as well and residues defined as moving were from Gly447 up to Gln452 [Fig.32A/B].

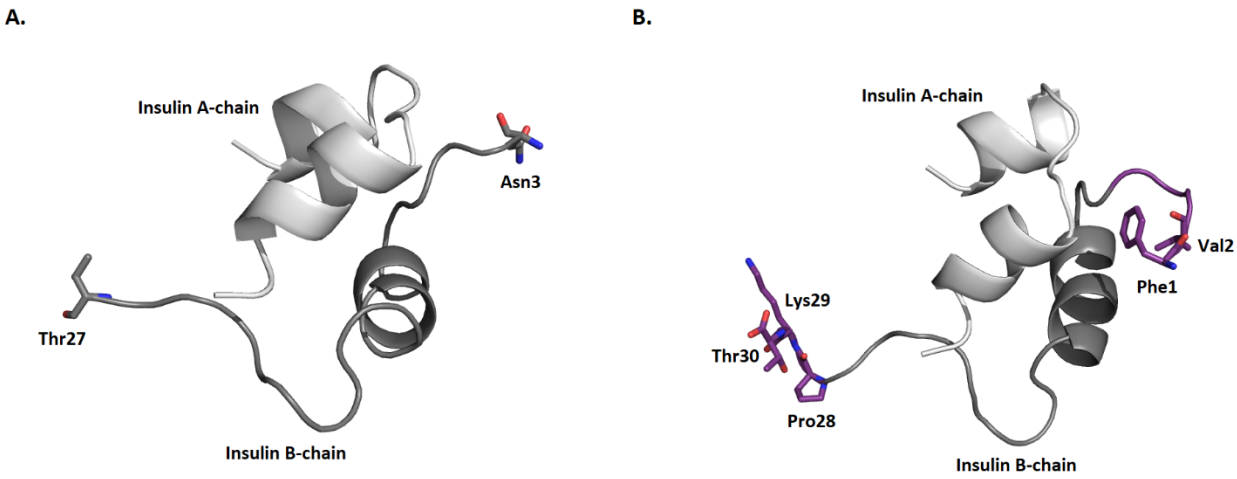
Another minor edit includes adding of residue Ser719 (αCT-segment) with deprotonated carboxyl group at the end. After residue Asn594, which is the last residues of the both chains because of cleaved domains, was added methyl group [Fig.32C]. In front of the first residue of cleaved incomplete domains was added acetyl group, particularly in front of Lys310 (cleaved beginning of the L2 domain) and Tyr675 (cleaved beginning of the αCT-segment) [Fig.32C].



[Fig.32]:

A. The incomplete loop in L2 domain (pink), because there are missing two residues between Thr448 and Arg451 (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source 6HN5: Weis *et al*. 2018).
B. The added residues Lys449 and Gly450 (magenta) complete the loop in L2 domain (pink).
C. αCT-segment (magenta) is cleaved behind residue Tyr675 and to this residue is added acetyl group. FnIII-1 domain (green) is cleaved behind residue Asn594 and to this residues is added methyl group (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on 6HN5: Weis *et al*. 2018 and ModelS1: Croll *et al*. 2016 and 6CE7: Scapin *et al*. 2018).

Insulin in this structure was missing the last three residues in C-terminus of B-chain (Pro28, Lys29, Thr30) [Fig.33] and the first two residues in N-terminus of B-chain (Phe1, Val2), which all were added in PyMOL graphic interface. After that, N-terminus of insulin B-chain was remodeled to have a conformation of T-state and the strategy was to align it with another insulin molecule in T-state (1MSO: Smith *et al*. 2003) and move residues Phe1-Leu6 (of 6HN5) into T-state [Fig.33].



[Fig.33]:

A. C-terminus of insulin B-chain (dark gray) ends with residue Thr27 and N-terminus of insulin B-chain starts with residue Asn3, thus it is incomplete insulin in the structure 6HN5; insulin A-chain is light gray (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source 6HN5: Weis *et al*. 2018).

B. The revised insulin in the structure 6HN5 now includes Pro28-Thr30 (purple) at C-terminus of B-chain (dark gray) and Phe1-Val2 (purple) at N-terminus of B-chain. In addition, N-terminal residues of the B-chain are remodeled to be in T-state conformation; the residues of N-terminus, which were moved to T-state are highlighted as purple loop downstream of residue Val2, insulin A-chain is light gray (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on 6HN5: Weis *et al*. 2018 and ModelS1: Croll *et al*. 2016 and 6CE7: Scapin *et al*. 2018 and 1MSO: Smith *et al*. 2003).

5.2 Molecular Modelling and Molecular Dynamics of Insulin and Insulin Analogues

5.2.1 Molecular Modelling of Insulin and Insulin Analogues

The initial structure was PDB structure with code 1MSO (Smith *et al*. 2003), which is insulin hexamer in T-state, but we worked only with one monomer of the hexameric structure, which was already prepared and optimized by RNDr. Martin Lepšík, PhD. from his previous research work (Žáková *et al*. 2014). This structure was mutated in PyMOL graphical interface by tool "Mutagenesis" and the mutations were based on experimental data from research group of RNDr. Jiří Jiráček, CSc. (manuscript in preparation). Each mutated variant contains only one

mutation and in case of mutations to histidine both Nδ (Hid) and Nε (Hie) versions of protonation were generated (except of mutant Leu13His). Overall, we generated 8 types of analogues with mutation in A-chain [Table.4].

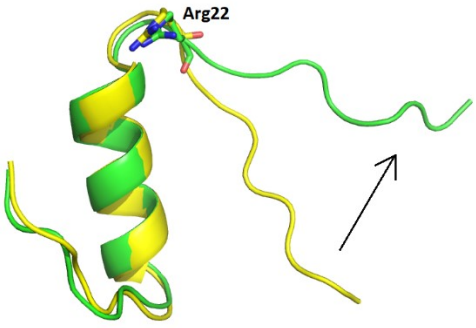
Residue in wild-type insulin (A-chain)	**Residue in mutated insulin**	
Ile10	Hie10 / Hid10	Val10
Ser12	Thr12	
Leu13	Hie13	Val13
Glu17	Hie17 / Hid17	

[Table.4]:

Table of the original amino acids (left column), which were mutated to different amino acids (right column). In case of Ile10 and Leu13 were generated two different substitution, while in case of Ser12 and Glu17 substitution with one amino acid.

The insulin structure from 1MSO (Smith *et al*. 2013) has, besides N-terminus in T-state, C-terminus of B-chain undetached. In some cases the introduction of mutation produced a steric clash or we wanted to try more residue rotamers, thus these mutated residues had to be rotated (these mutants have then "-edit" in title).

The same type of mutations was made in insulin structure with C-terminus of B-chain detached and with N-terminus of B-chain in T-state. This structure was modeled from the structure with PDB code 1MSO (Smith *et al*. 2013). Residues in insulin B-chain from Arg22 up to end of C-terminus were moved into detached conformation inspired by insulin structure from 4OGA (Menting *et al*. 2014) [Fig.34].



[Fig.34]:

Original insulin structure 1MSO (green) was modeled to have detached C-terminus of B-chain. The new model (yellow) has moved residues Arg22-Thr30. The arrow represent the direction of movement (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source 1MSO: Smith *et al*. 2003 and own).

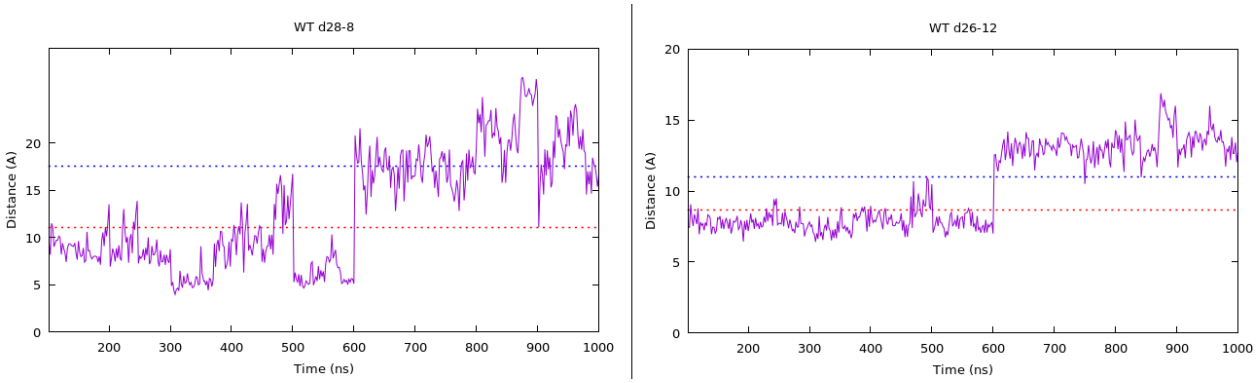
The last series of the same mutations was made with the exactly same insulin structure, which is included in our remodeled ModelS1 structure. This insulin structure has detached C-terminus of B-chain and N-terminus of B-chain in open state (O-state).

5.2.2 Molecular Dynamics of Insulin and Insulin Analogues (First Set)

The first set MD with insulin and insulin analogues contains those modeled only from the structure 1MSO (Smith *et al*. 2013), which has C-terminus of B-chain undetached and N-terminus of B-chain in T-state.

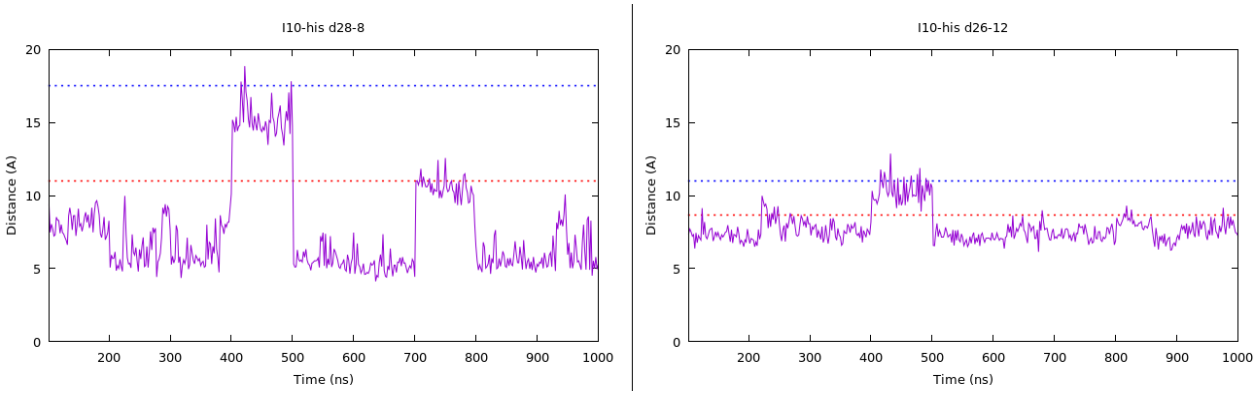
The observed feature from the last run of molecular dynamics (900ns) was if C-terminus of insulin/insulin analogue B-chain obtains detached conformation. If C-terminus of B-chain is in detached conformation, it was defined on the basis of measurement the distances between residues in C-terminus of B-chain and residues in B-chain helix. The distance was measured between residues Gly8(B-chain helix)-Pro28(C-terminus of B-chain), and between residues Val12(B-chain helix)-Tyr26(C-terminus of B-chain) (Papaioannou *et al*. 2017).

For the first pair of residues holds this: when C-terminus of B-chain is undetached, distance between residues Gly8 and Pro28 is below 11Å. When C-terminus of B-chain is detached, distance between residues Gly8 and Pro28 is greater than 17.5Å. For the second pair of residues holds this: when C-terminus of B-chain is detached, distance between residues Val12 and Tyr26 is below 8.7Å. When C-terminus of B-chain is detached, distance between distance between residues Val12 and Tyr26 is greater than 11Å (Papaioannou *et al*. 2017). The graphs showing the distance between these two pairs of residues during 900ns MD run were made for wild-type insulin as well as for insulin analogues [Fig.35-45].



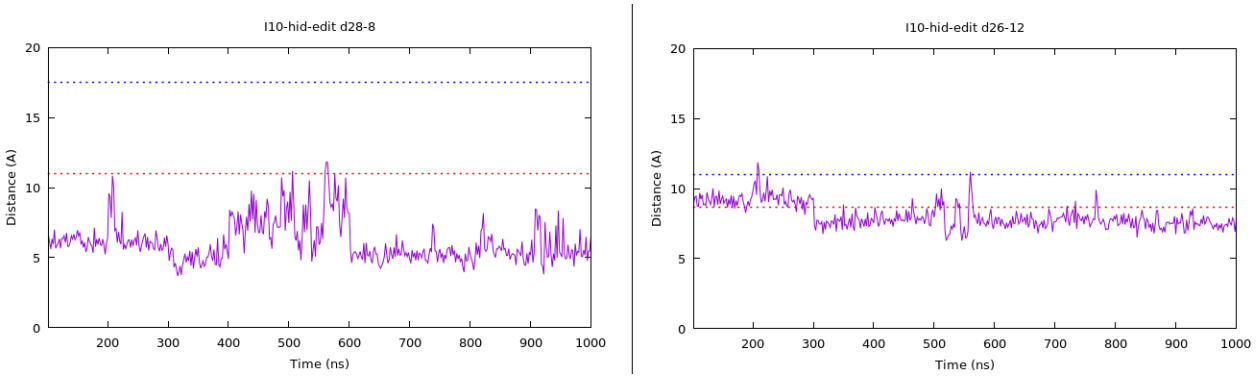
[Fig.35]:

Left: Graph of the measured distances between Gly8-Pro28 in wild-type insulin. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in wild-type insulin. Blue line is threshold 11Å, red line is threshold 8.7Å.



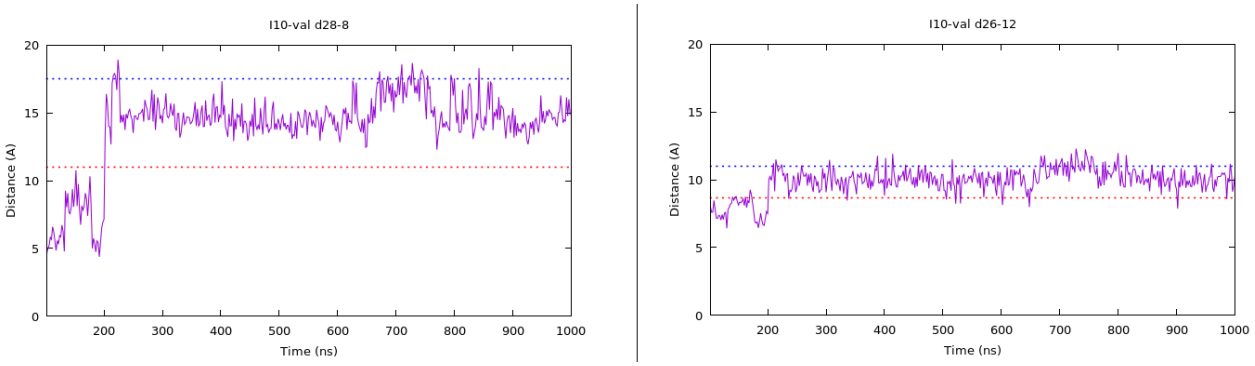
[Fig.36]:

Left: Graph of the measured distances between Gly8-Pro28 in Ile10His mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Ile10His mutant. Blue line is threshold 11Å, red line is threshold 8.7Å.



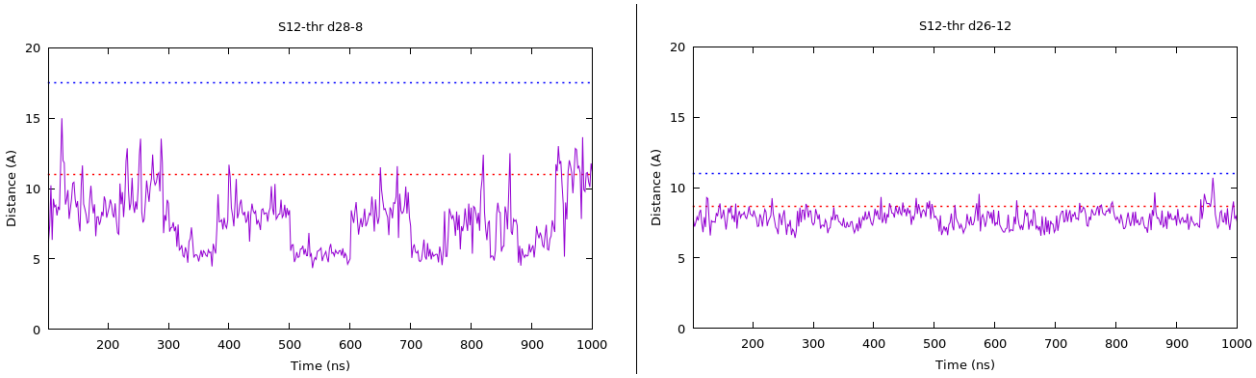
[Fig.37]:

Left: Graph of the measured distances between Gly8-Pro28 in Ile10Hid mutant (edited). Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Ile10Hid mutant (edited). Blue line is threshold 11Å, red line is threshold 8.7Å.



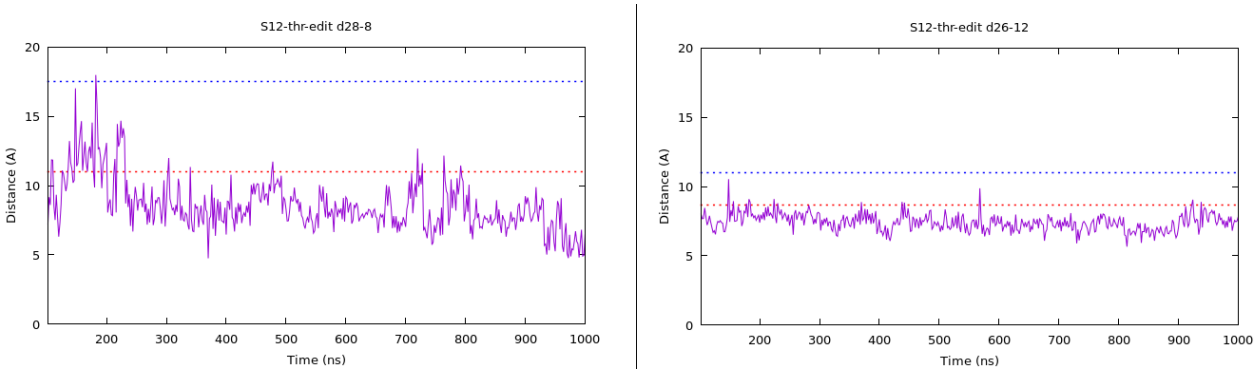
[Fig.38]:

Left: Graph of the measured distances between Gly8-Pro28 in Ile10Val mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Ile10Val mutant. Blue line is threshold 11Å, red line is threshold 8.7Å.



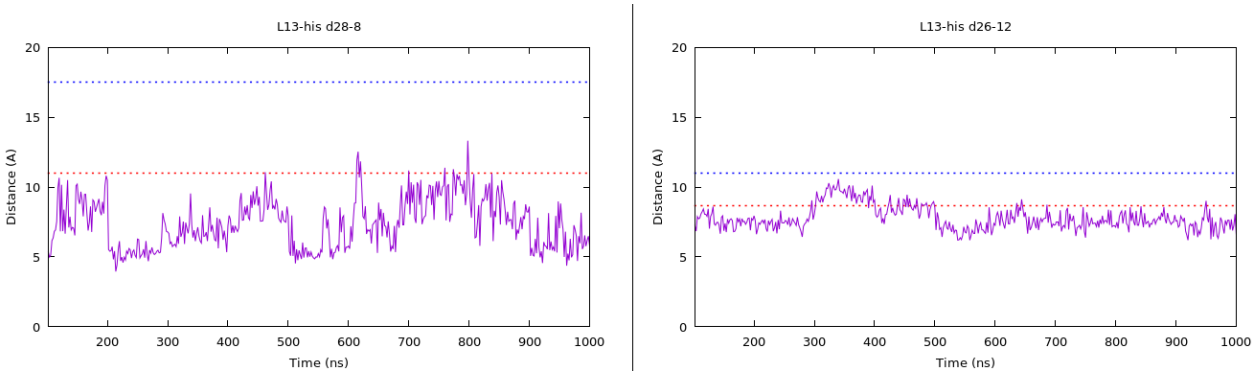
[Fig.39]:

Left: Graph of the measured distances between Gly8-Pro28 in Ser12Thr mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Ser12Thr mutant. Blue line is threshold 11Å, red line is threshold 8.7Å.



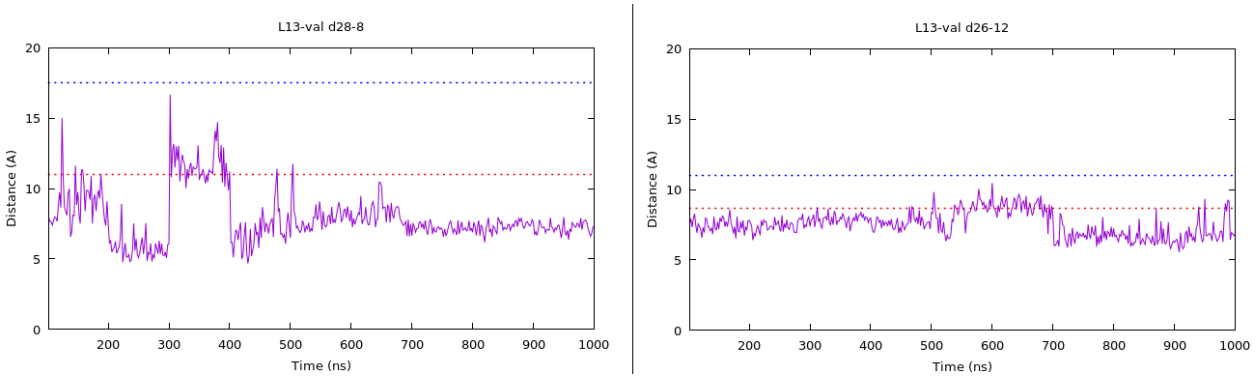
[Fig.40]:

Left: Graph of the measured distances between Gly8-Pro28 in Ser12Thr (edited) mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Ser12Thr (edited) mutant. Blue line is threshold 11Å, red line id threshold 8.7Å.



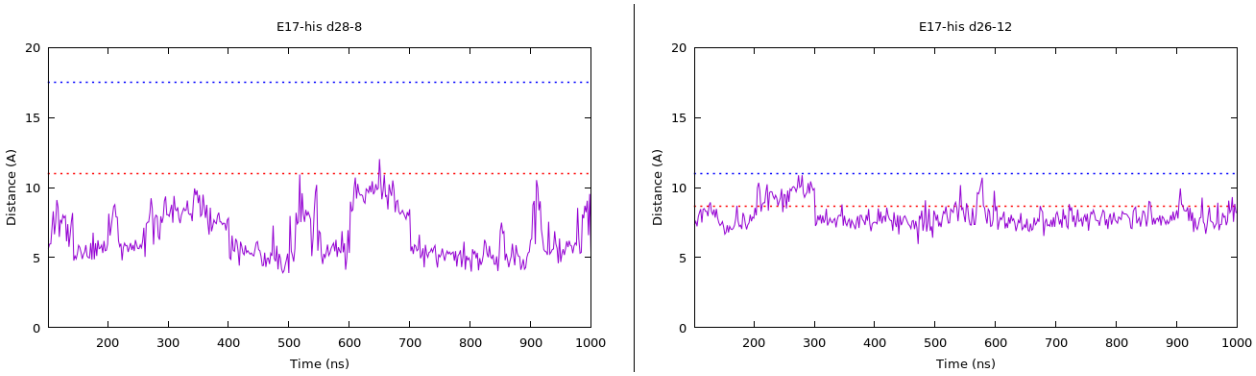
[Fig.41]:

Left: Graph of the measured distances between Gly8-Pro28 in Leu13His mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Leu13His mutant. Blue line is threshold 11Å, red line is threshold 8.7Å.



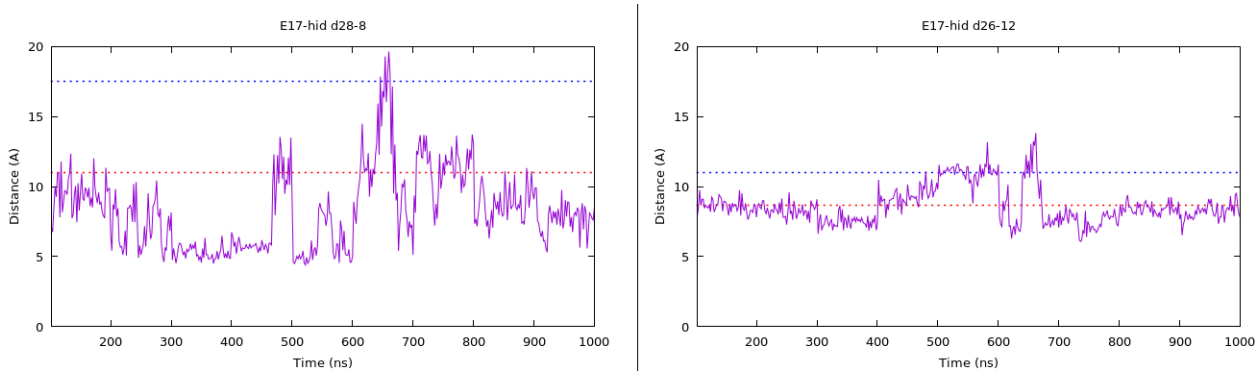
[Fig.42]:

Left: Graph of the measured distances between Gly8-Pro28 in Leu13Val mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Leu13Val mutant. Blue line is threshold 11Å, red line is threshold 8.7Å.



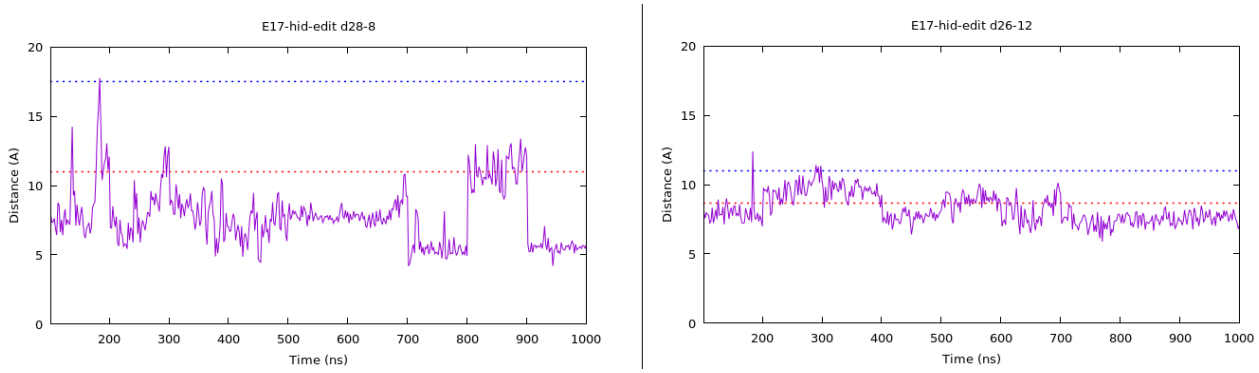
[Fig.43]:

Left: Graph of the measured distances between Gly8-Pro28 in Glu17His mutant. Blue line is threshold 17.5Å, red line id threshold 11Å below.
Right: Graph of the measured distances between Val12-Tyr26 in Glu17His mutant. Blue line is threshold 11Å, red line id threshold 8.7Å.



[Fig.44]:

Left: Graph of the measured distances between Gly8-Pro28 in Glu17Hid mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Glu17Hid mutant. Blue line is threshold 11Å, red line is threshold 8.7Å.



[Fig.45]:

Left: Graph of the measured distances between Gly8-Pro28 in Glu17Hid (edited) mutant. Blue line is threshold 17.5Å, red line is threshold 11Å.
Right: Graph of the measured distances between Val12-Tyr26 in Glu17Hid (edited) mutant. Blue line is threshold 11Å, red line is threshold 8.7Å.

Besides measuring the distances between residues of B-chain, was performed also an analysis of hydrogen bonds formation. I analyzed the mutated residues of analogues as well as the original residues of wild-type insulin (at that particular position as in mutants) during 900ns MD run. Only those hydrogen bonds, which have more than 25% occupancy during 900ns MD were processed [Table.5-7]. Mutants of residue Ile10 do not contain hydrogen bonds with occupancy higher 25% at the position 10 in A-chain.

Atom of the critical residue (A-chain)	Hydrogen bond partner	Occupancy in wild-type insulin	Occupancy in Ser12Thr	Occupancy in Ser12Thr-edit
12:O	Gln15:N-H	74%	85%	75%
12:O	Leu16:N-H	83%	67%	95%
12:OG	Gln15:N-H	32%	25%	48%
12:N-H	Gln15:OE1	34%	42%	74%
Thr12:OG1-HG1	Gln15:OE1	-	37%	78%

[Table.5]:

Hydrogen bond occupancy of residue at position 12 in B-chain of wild-type insulin as well as in Ser12Thr mutant and its edited version.

Atom of the critical residue (A-chain)	Hydrogen bond partner	Occupancy in wild-type insulin	Occupancy in Leu13His	Occupancy in Leu13Val
13:O	Glu17:N-H	97%	85%	92%
His13:ND1	His17:N-H	-	31%	-

[Table.6]:

Hydrogen bond occupancy of residue at position 13 in B-chain of wild-type insulin as well as in Leu13His and Leu13Val mutant.

Atom of the critical residue (A-chain)	Hydrogen bond partner	Occupancy in wild-type insulin	Occupancy in Glu17His	Occupancy in Glu17Hid	Occupancy in Glu17Hid-edit
17:N-H	Leu13:O	97%	71%	81%	89%
17:O	Cys20:N-H	55%	81%	79%	65%
Glu17:OE1	Arg43:NH1-HH12	79%	-	-	-
Glu17:OE1	Arg43:NH2-HH22	68%	-	-	-
Glu17:OE2	Arg43:NH1-HH12	84%	-	-	-
Glu17:OE2	Arg43:NH2-HH22	76%	-	-	-
Hid17:N-H	Tyr14:O	-	29%	25%	-

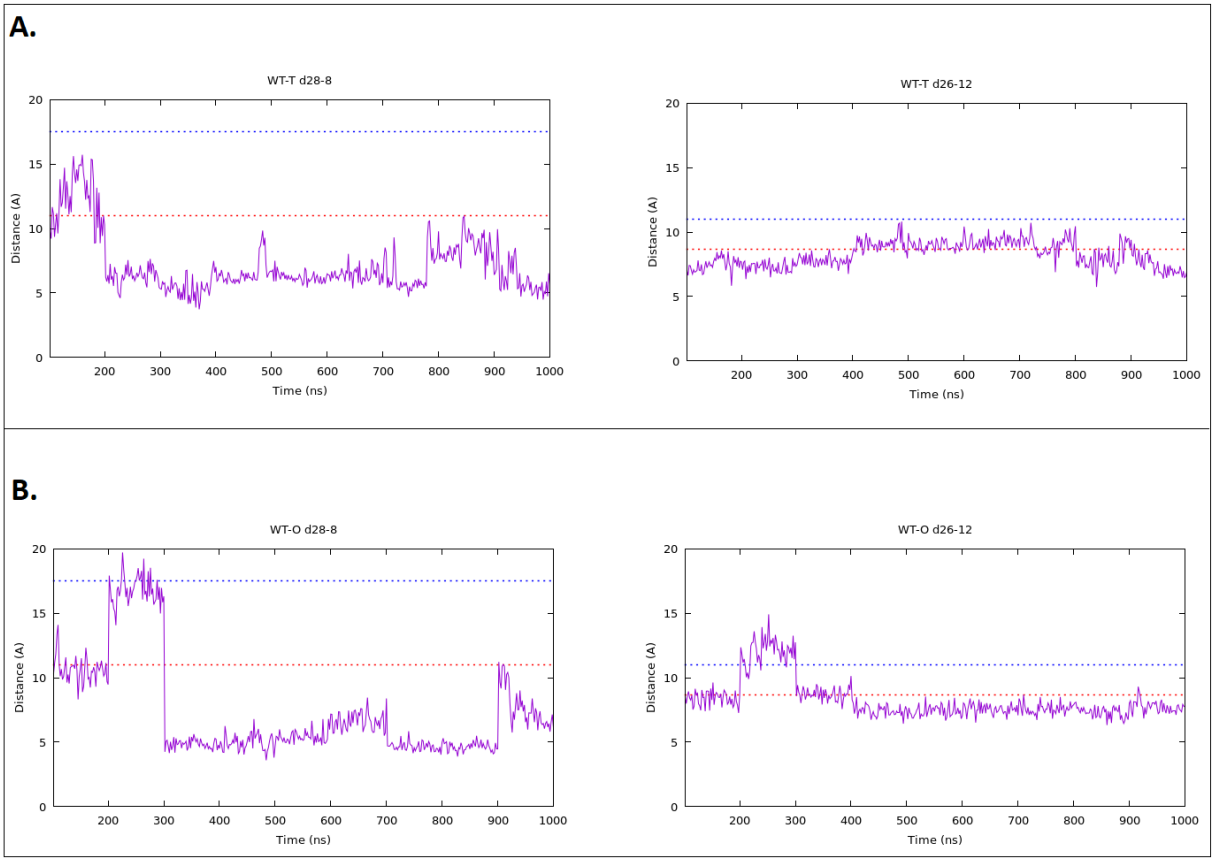
[Table.7]:

Hydrogen bond occupancy of residue at position 17 in B-chain of wild-type insulin as well as in Glu17His, Glu17Hid and its edited version.

5.2.3 Molecular Dynamics of Insulin and Insulin Analogues (Second set)

The 2.set of MD with insulin and insulin analogues structures includes those, which were modeled from the structure 1MSO (Smith *et al*. 2013) and those originally taken from remodeled structure ModelS1. All these structures have C-terminus of B-chain undetached, and N-terminus of B-chain in T-state or in O-state, thus there are two groups in this MD run: wild-type insulin and mutants with detached C-terminus and N-terminus in T-state, and wild-type insulin and mutants with detached C-terminus and N-terminus in O-state. In the 2.set of MD were not included all previous mutated residues, but besides wild-type insulin only Glu17His and Ser12Thr.

From 900ns MD run was also analyzed the distance between residues Gly8-Pro28 and Val12-Tyr26 [Fig.46-48]. The thresholds defining undetached and detached C-terminus of B-chain remain the same as in the previous MD analysis.



[Fig.46]:

A.

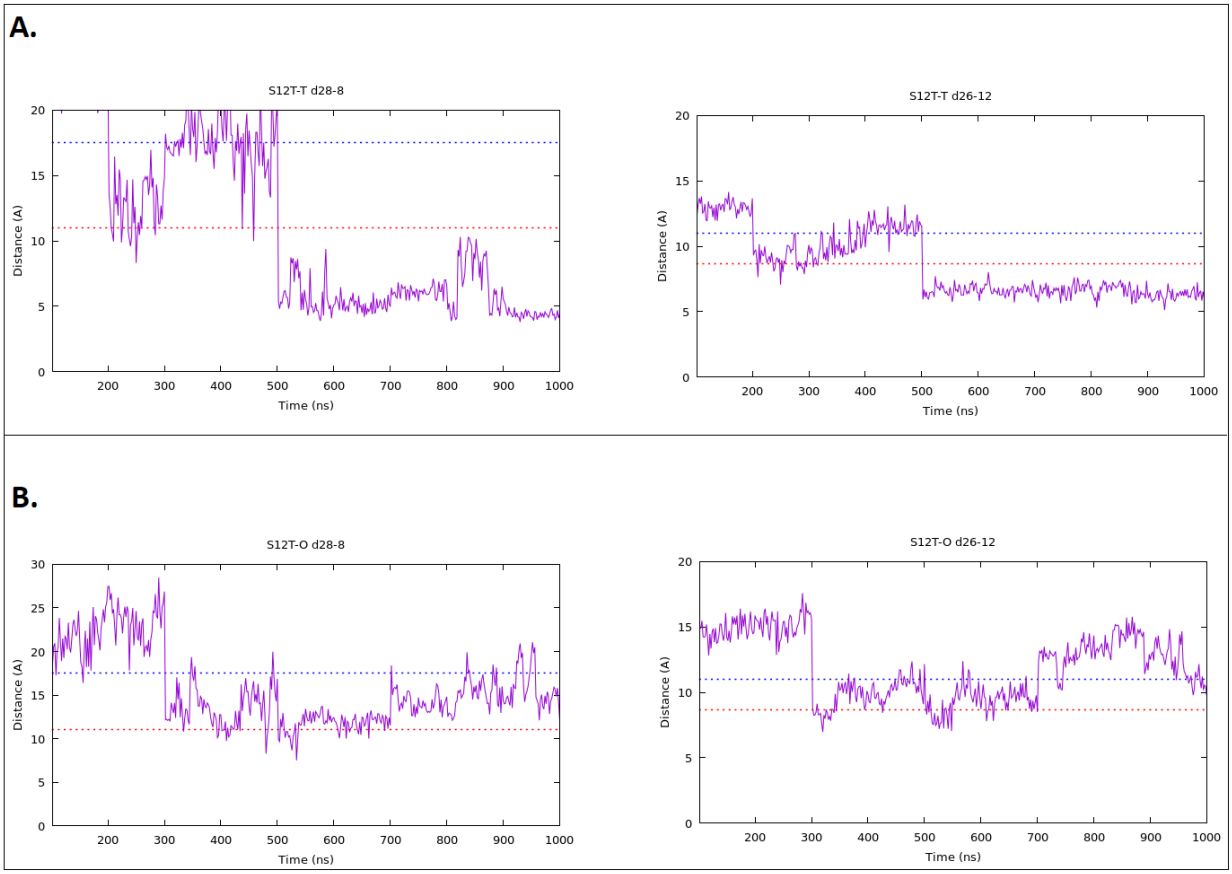
Right: Graph of measured distances between the residues Gly8-Pro28 in wild-type insulin with detached C-terminus of B-chain and N-terminus of B-chain in T-state. Blue line is threshold 17.5Å, red line is threshold 11Å.

Left: Graph of measured distances between the residues Val12-Tyr26 in wild-type insulin with detached C-terminus of B-chain and N-terminus of B-chain in T-state. Blue line is threshold 11Å, red line is threshold 8.7Å.

B.

Right: Graph of measured distances between the residues Gly8-Pro28 in wild-type insulin with detached C-terminus of B-chain and N-terminus of B-chain in O-state. Blue line is threshold 17.5Å, red line is threshold 11Å.

Left: Graph of measured distances between the residues Val12-Tyr26 in wild-type insulin with detached C-terminus of B-chain and N-terminus of B-chain in O-state. Blue line is threshold 11Å, red line is threshold 8.7Å.



[Fig.47]:

A.

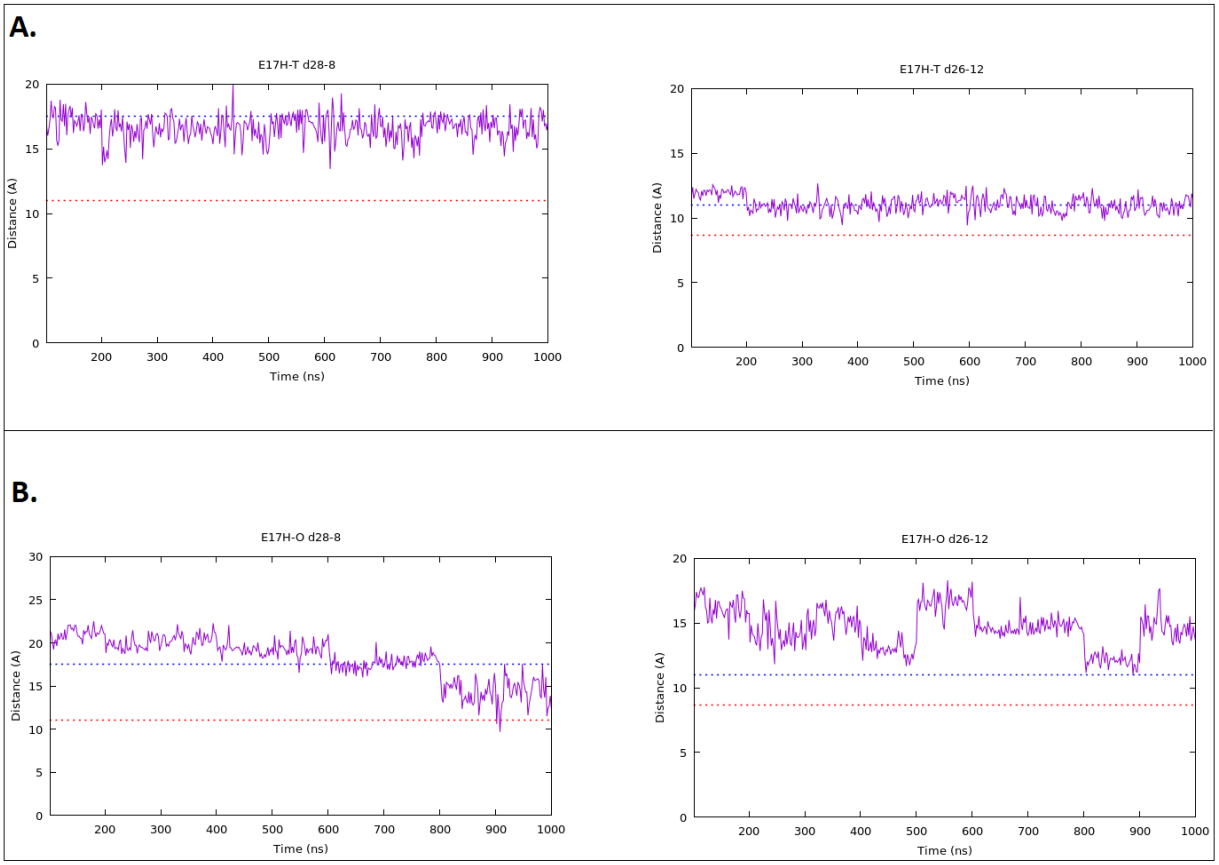
Right: Graph of measured distances between the residues Gly8-Pro28 in S12Thr mutant with detached C-terminus of B-chain and N-terminus of B-chain in T-state. Blue line is threshold 17.5Å, red line is threshold 11Å.

Left: Graph of measured distances between the residues Val12-Tyr26 in S12Thr mutant with detached C-terminus of B-chain and N-terminus of B-chain in T-state. Blue line is threshold 11Å, red line is threshold 8.7Å.

B.

Right: Graph of measured distances between the residues Gly8-Pro28 in S12Thr mutant with detached C-terminus of B-chain and N-terminus of B-chain in O-state. Blue line is threshold 17.5Å, red line is threshold 11Å.

Left: Graph of measured distances between the residues Val12-Tyr26 in S12Thr mutant with detached C-terminus of B-chain and N-terminus of B-chain in O-state. Blue line is threshold 11Å, red line is threshold 8.7Å.



[Fig.48]:

A.

Right: Graph of measured distances between the residues Gly8-Pro28 in Glu17His mutant with detached C-terminus of B-chain and N-terminus of B-chain in T-state. Blue line is threshold 17.5Å, red line is threshold 11Å.

Left: Graph of measured distances between the residues Val12-Tyr26 in Glu17His mutant with detached C-terminus of B-chain and N-terminus of B-chain in T-state. Blue line is threshold 11Å, red line is threshold 8.7Å.

B.

Right: Graph of measured distances between the residues Gly8-Pro28 in Glu17His mutant with detached C-terminus of B-chain and N-terminus of B-chain in O-state. Blue line is threshold 17.5Å, red line is threshold 11Å.

Left: Graph of measured distances between the residues Val12-Tyr26 in Glu17His mutant with detached C-terminus of B-chain and N-terminus of B-chain in O-state. Blue line is threshold 11Å, red line is threshold 8.7Å.

Hydrogen bonds analysis was done with this MD set as well to see hydrogen bonding occupancy of the critical residues at positions 12 and 17 (in B-chain) during 900ns MD run [Table.8-11].

Atom of the critical residue	Hydrogen bond partner	Occupancy in wild-type insulin (T-state)	Occupancy in Ser12Thr (T-state)
12:O	Leu16:N-H	80%	71%
12:O	Gln15:N-H	72%	77%
12:OG	Gln15:N-H	46%	45%
12:N-H	Gln15:OE1	25%	50%
12:OG1-HG1	Gln15:OE1	-	63%

[Table.8]:

Hydrogen bond occupancy of residue at position 12 in B-chain of wild-type insulin in T-state as well as in Ser12Thr in T-state.

Atom of the critical residue	Hydrogen bond partner	Occupancy in wild-type insulin (O-state)	Occupancy in Ser12Thr (O-state)
12:O	Leu16:N-H	90%	93%
12:O	Gln15:N-H	71%	71%
12:OG	Gln15:N-H	48%	54%
12:N-H	Gln15:OE1	43%	39%
12:OG	Tyr14:N-H	39%	-

[Table.9]:

Hydrogen bond occupancy of residue at position 12 in B-chain of wild-type insulin in O-state as well as in Ser12Thr in O-state.

Atom of the critical residue	Hydrogen bond partner	Occupancy in wild-type insulin (T-state)	Occupancy in Glu17His (T-state)
17:N-H	Leu13:O	93%	99%
17:O	Cys20:N-H	56%	-
17:OE1	Arg43:NH1-HH12	53%	-
17:OE2	Arg43:NH1-HH12	46%	-
17:OE2	Arg43:NH2-HH22	43%	-
17:OE1	Arg43:NH2-HH22	40%	-
17:O	Asn21:N-H	-	37%
17:O	Asn21:ND2-HD22	-	34%

[Table.10]:

Hydrogen bond occupancy of residue at position 17 in B-chain of wild-type insulin in T-state as well as in Glu17His in T-state.

Atom of the critical residue	Hydrogen bond partner	Occupancy in wild-type insulin (O-state)	Occupancy in Glu17His (O-state)
17:N-H	Leu13:O	81%	80%
17:O	Cys20:N-H	64%	40%
17:OE1	Arg43:NH1-HH12	51%	-
17:OE2	Arg43:NH1-HH12	32%	-
17:OE2	Arg43:NH2-HH22	50%	-
17:OE1	Arg43:NH2-HH22	47%	-

[Table.11]:

Hydrogen bond occupancy of residue at position 17 in B-chain of wild-type insulin in O-state as well as in Glu17His in O-state.

5.3 Docking of Insulin into Insulin Receptor Structures

5.3.1 Docking of Insulin into Remodeled Insulin Receptor ModelS1

Our remodeled structure of insulin receptor ModelS1, which underwent 4ns long MD (courtesy of RNDr. Martin Lepšík, PhD.) and is even more open, served as one input in HADDOCK webserver. There were 3 types of insulin structures, which were used as other inputs and it was insulin in T-state, R-state, and O-state. Each of the insulin structures has detached C-terminus of B-chain. Insulin in T-state was the same structure as it was used in the MD of insulin analogues, which means it is the hybrid of insulin from structure 4OGA (Menting *et al*. 2014) and from structure 1MSO (Smith *et al*. 2003). Insulin in R-state is modeled from structure with PDB code 1G7B (Smith *et al*. 2001), which has C-terminus of B-chain undetached, but it was modeled to have detached C-terminus manually in PyMOL graphic interface. Insulin in O-state was taken from our remodeled structure ModelS1.

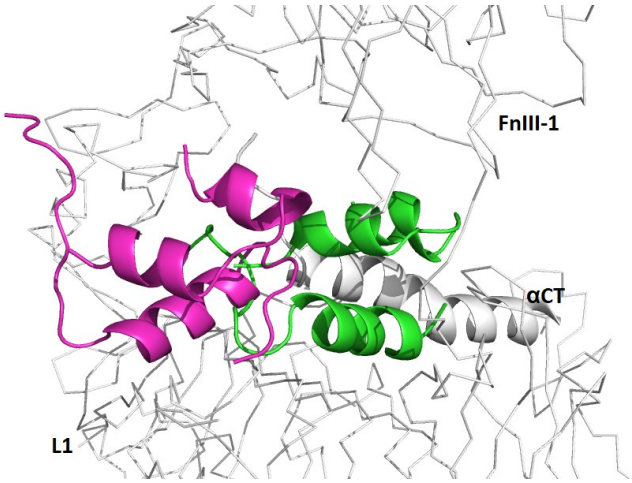
The combinations of active residues from insulin receptor and insulin structures were specified based on known site1 interactions (Menting *et al*. 2013) [Table.12]. The passive residues were set up as default.

	Insulin residues	**Insulin receptor residues**
1.combination	A-chain: G1, I2, V3, Y19 B-chain: G8, S9, L11, V12, L15, Y16	αCT: H710, F714 L1: F39, F64, R65
2.combination	A-chain: G1, I2,V3, Y19 B-chain: G8, S9, L11, V12, L15	αCT: H710, F714
3.combination	A-chain: V3 B-chain: G8, S9	αCT: H710

[Table.12]:

Table of the active residues combinations, which were used for this set of docking.

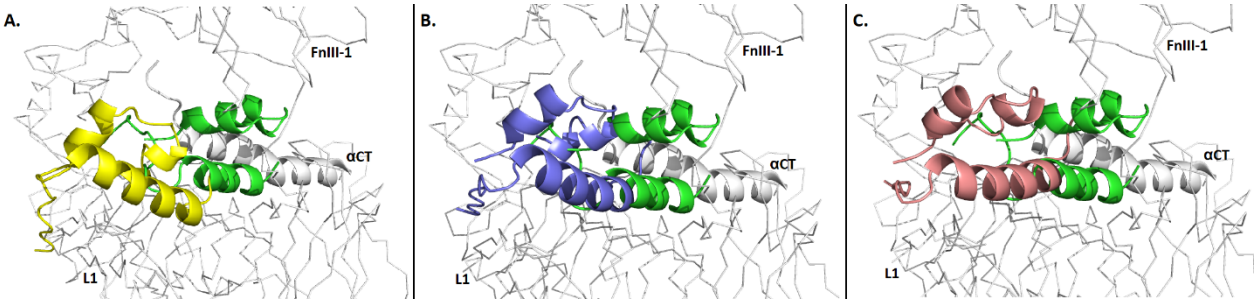
The structure from docking with insulin in T-state and remodeled insulin receptor from ModelS1, which mimics the natural binding pose the best ensued from the 2.combination of the active residues [Fig.49].



[Fig.49]:

The result of docking with insulin in T-state using 2.combination of the active residues (magenta), which mimics the natural binding pose of insulin the best (green) (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014 and 4OGA: Menting *et al*. 2014 and 1MSO: Smith *et al*. 2003).

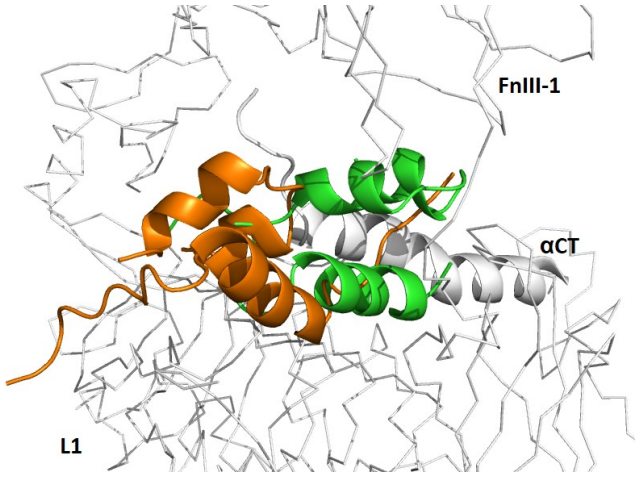
Output structures of docking insulin R-state with insulin receptor from remodeled ModelS1, which mimic the natural binding pose were found as very similar in each combination of the active residues [Fig.50].



[Fig.50]:

A. The result docking pose (yellow), which mimics the natural binding pose of insulin (green) the best with the 1.combination of the active residues and insulin in R-state.
B. The docking pose (blue) mimicking the natural binding pose of insulin (green) the best with the 2.combination of the active residues and insulin in R-state.
C. The result docking pose (pink), which with using of the 3.combination and insulin in R-state mimics the natural binding pose of insulin (green) the best
(Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014 and 1G7B: Smith *et al*. 2001).

Docking with insulin O-state and insulin receptor from remodeled ModelS1 resulted in finding the structure, which mimics the natural binding pose the best with 1.combination of the active residues [Fig.51].



[Fig.51]:

The result of docking (orange), which mimics the natural binding pose of insulin (green) the best with the 1.combination of the active residues and insulin in I/O-state (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source own based on ModelS1: Croll *et al*. 2016 and 4OGA: Menting *et al*. 2014 and 1MSO: Smith *et al*. 2003).

5.3.2 Docking of Insulin into Insulin Receptor Structure determined by Cryo-EM

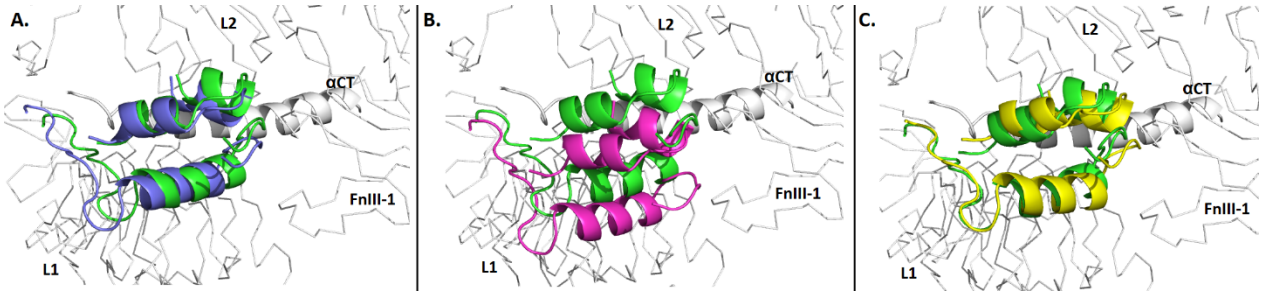
Insulin in T-state (the same structure as in the previous docking with insulin T-state) was one of the inputs in HADDOCK webserver. Another input was insulin receptor in insulin bound state from the cryo-EM structure with PDB code 6CE7 (Scapin *et al*. 2018). The first two combinations of the active residues from insulin and insulin receptor remain the same as in previous docking, but in addition, there was one different combination based on own observation of which insulin residues could interact with FnIII-1 residues of insulin receptor [Table.13].

	Insulin residues	Insulin receptor residues
1.combination	A-chain: G1, I2, V3, Y19 B-chain: G8, S9, L11, V12 ,L15, Y16	αCT: H710, F714 L1: F39, F64, R65
2.combination	A-chain: G1, I2, V3, Y19 B-chain: G8, S9, L11, V12, L15	αCT: H710, F714
3.combination	B-chain: Q4, H5, C7, S9, H10	Fn-1 domain: P495, F497, R498, R539, D542

[Table.13]:

Table of the active residues combinations, which were used for docking insulin in T-state and insulin receptor from the cryo-EM structure.

Each of the active residues combinations resulted in a structure, which is able to mimic the natural binding pose of insulin in insulin receptor. However, the 1. and 3.combination of the active residues resulted in slightly better poses than the 2.combination [Fig.52].



[Fig.52]:

A. The docking pose of the 1.combination (blue), which mimics the natural binding pose of insulin (green) the best.
B. The 2.combination resulted also in binding pose (magenta) mimicking the natural binding pose of insulin (green), but not as sufficient as with other active residues combinations.
C. The result docking pose of the 3.combination (yellow) mimicking the natural binding pose of insulin (green) the best (Figure prepared with PyMol, version 2.0.7., https://pymol.org/2/; source 6CE7: Scapin *et al*. 2018 and 4OGA: Menting *et al*. 2014 and 1MSO: Smith *et al*. 2003).

6 Discussion

- **Molecular Modelling of Insulin Receptor**

Building a reliable insulin receptor model is crucial for further analysis such as molecular dynamics. It is important to mention that at that time there was no insulin receptor structure in insulin-bound state, which includes also site2 interface (fibronectin domains). There was only truncated receptor (L1-CR and αCT-segment) with bound insulin (PDB code 4OGA: Menting *et al*. 2014), which contains site1 interface, but not site2 interface as this novel model contains. The most complete insulin apo-receptor model called ModelS1 (Croll *et al*. 2016), contains fibronectin domains, but does not contains insulin-bound features at all, however, they are now included in this novel model. The model is successfully used in molecular dynamics and scoring analysis by other colleagues.

Another remodeled structure was the newest cryo-EM structure of insulin receptor with bound insulin (PDB code 6HN5: Weis *et al*. 2018). Compared to the initial structure, in the new model were added missing loops and modeled into favourable conformation, especially the loop in FnIII-1 domain necessary for site2 interface investigation. This completed model of insulin receptor in insulin bounds state represents another structure, which is used in molecular dynamics and scoring analysis. In addition, it serves as comparison to the previous remodeled structure.

- **Molecular Modelling of Insulin and Insulin Analogues**

Insulin analogues were modeled in order to have required models in molecular dynamics. There were modeled from several available insulin structures and maintain various conformations, in which differ from the original structures 1MSO (Smith *et al*. 2003) and 4OGA (Menting *et al*. 2014), which are either incomplete or not in desired conformation. These models complement set of structures derived from insulin and are actively used also by other colleagues in their molecular dynamics analysis.

- **Molecular Dynamics of Insulin and Insulin Analogues (First Set)**

To contextualize our approach to MD of insulin and insulin analogues and measuring if C-terminus of B-chain is detached, it is necessary to show experimental data of measured affinity of the particular insulin analogues to insulin receptor. This experimental data were provided by research group of RNDr. Jiří Jiráček, CSc. (unpublished data, manuscript in preparation) [Table.14].

Insulin Analogue	Affinity of insulin analogue to insulin receptor -isoform A (in comparison to wild-type insulin - 100%)
Ile10His	16%
Ile10Val	73%
Ser12Thr	108%
Leu13His	15%
Leu13Val	91%
Glu17His	7%

[Table.14]:

Experimentally measured affinity of insulin analogues to insulin receptor. The values are present in percentage as it is comparison to affinity of wild-type insulin to insulin receptor, which is considered as 100% (unpublished data).

The values of measured insulin analogues affinities to insulin receptor show, that these analogues can be divided into two main groups: low-affinity analogues and analogues with affinity comparable to wild-type insulin. To be further specific, these residues of insulin A-chain were hypothesized to be part of so-called site2 interactions (De Meyts 2015) and this is the reason why they were chosen for mutation.

It is known that if C-terminus of insulin B-chain is undetached, can not be bound to insulin receptor because of steric clashes with αCT-segment, thus it represents inactive state of insulin (Hua *et al*. 1991). The purpose of MD analysis was to observe if during simulation C-terminus of B-chain archives detached conformation, which corresponds to active state of insulin (Hua *et al*. 1991). Papaioannou *et al*. (2017) used a similar approach for studying insulin analogues activity, when they take advantage of information that detachment of C-terminus occurs when water molecule breaks hydrophobic core of insulin, of which interactions do not keep insulin closed anymore.

There is also hypothesis that site2 residues could play a role in initial docking of insulin to insulin receptor, what allows rearrangement of insulin receptor and interactions of site1 residues with each other (Weis *et al*. 2018), thus there could be a connection between their mutation and detachment of B-chain C-terminus, necessary for binding to the receptor.

Wild-Type Insulin

We set the thresholds for detached and undetached conformation according to Papaioannou *et al*. (2017). In wild-type insulin occurs the detachment of C-terminus of B-chain in time 600ns and remains like that until the end of MD simulation. This result can be proof of the allegation about that detachment of B-chain C-terminus is possible without insulin receptor in the system and water molecule breaking the hydrophobic core of insulin is the important factor in this conformational change, as Papaioannou *et al*. (2017) suggested. It could be also proposed that wild-type insulin has favourable noncovalent interactions, which allow the water molecule to enter and break the hydrophobic core.

Mutants at Position 10 in A-chain

Insulin analogue Ile10His is considered as low-affinity analogue, because of its 16% experimentally measured affinity to insulin receptor in comparison to wild-type insulin (unpublished data of Dr. Jiráček's research group). One can assume, when compared to wild-type behavior in MD, that its low affinity could have connection with not being able to archive detached conformation of B-chain C-terminus. The measured distances of its His10:Nε protonated version did not exceed thresholds for detached conformation, in fact, they remain below threshold of undetached conformation for the most time of simulation, excepting of one slightly more significant motion towards detachment in time of 400ns, but it is still under threshold of detachment. An explanation of how a single mutation in A-chain could have contribution on conformational changes in B-chain can be in found in composition of noncovalent interactions in the analogue, because substitution of one amino acid by another amino acid with completely different chemical properties (such as isoleucine to histidine is) could have contribution on the noncovalent interactions composition in such way that water molecule is no table to enter the hydrophobic core and breaks it.

Second version of the same substitution Ile10Hid-edit, which has protonated His10:Nδ behaves in a similar way as the previous one, because C-terminus of B-chain remains undetached during the whole time of MD simulation, because there is not more significant movement toward detachment of C-terminus. This is in accord with the experimental data, which showed this analogue with low affinity. Using of both possible protonated states of histidines demonstrates no serious variability between them.

Substitution of isoleucine by valine in analogues Ile10Val resulted in experimentally measured affinity of 73% in comparison to wild-type insulin (unpublished data of Dr. Jiráček's

research group). The measured distances during MD simulation maintain intermediate values as they remain mostly between the both thresholds, what means analogue is not in inactive undetached conformation anymore, but also is not fully activated yet. One could conclude that these distances can correspond to 73% affinity.

Mutants at Position 12 in A-chain

Experimentally measured affinity of insulin analogue Ser12Thr is 108%, which make this type of analogue comparable to wild-type insulin (unpublished data of Dr. Jiráček's research group), thus one would expect the same or very similar pattern of graphs with measured distances. However, the distances did not exceed the detached threshold, but not even significantly exceed the undetached threshold, what means that analogue persevered in inactive undetached state for whole MD simulation.

The second version of this analogue Ser12Thr-edit did not obtain detached conformation during simulation as well. Values of the measured distances also remain mostly below undetached threshold.

Analysis of hydrogen bond forming at position 12 showed that hydrogen bonds partners as well as occupancy of the hydrogen bonds is in insulin analogue Ser12Thr similar to wild-type insulin. It is expected, because in wild-type insulin is at position 12 (A-chain) amino acid serine and in the analogue is threonine, which are amino acids of similar chemical properties, thus it is conservative substitution and it is not expected to have significant changes in activity. Any hydrogen bonds were not destructed by mutation, in fact one new was created. Occupancies of hydrogen bonds are also mostly comparable, only with two exceptions in analogue Ser12Thr-edit, where is elevated occupancy of the last two hydrogen bonds, what can be a result of rotation with Thr12 during modelling.

The reason why expected detachment of C-terminus of B-chain did not occur as in wild-type insulin, which has comparable affinity, can be explained by lack of sampling in this MD run. One of the options might be to extend MD simulation. As these mutated structures serve also as inputs for other different types of MD performed by other colleagues, it was showed, specifically in MD with collective variables, that detachment of B-chain C-terminus in mutant Ser12Thr occurs (unpublished data and personal communication with Anja Muždalo).

Mutants at Position 13 in A-chain

Insulin analogue Leu13His was experimentally demonstrated with low affinity 15% to insulin receptor in comparison to wild-type insulin (unpublished data of Dr. Jiráček's research group), which is in accordance with theoretical data. Values of the measured distances if B-chain C-terminus is detached are mostly below undetached threshold, which means this analogue is inactive. There are some values slightly above undetached threshold, but is negligible as it is just for very short time and the values are still not enough for detachment.

Conservative mutation of leucine to valine resulted in experimentally measured affinity 91% (unpublished data of Dr. Jiráček's research group). One would expect that detachment of B-chain C-terminus occurs as it is analogue of comparable affinity to wild-type insulin. However, the measured distances in MD simulation are mostly below undetached threshold, except for short time above, but it is still not above detached threshold, thus this analogue does not go to active conformation at all. It is the same pattern as in analogue Ser12Thr, which was also expected to obtain active detached conformation, but rather stayed undetached and inactive. This result can be interpreted the same way as Ser12Thr analogue, that there is lack of sampling and it is needed to extend MD simulation to see if the analogue overcome local minimum, which trapped the analogue in energy barrier.

Residue Leu13 in wild-type insulin forms stable hydrogen bond with residue Glu17 in A-chain and this hydrogen bond formation is comparable to those two mutants as well. There is only one new hydrogen bond in mutant Leu13His formed within the mutated residue His17 itself.

Mutants at Position 17 in A-chain

Glutamic acid at position 17 in A-chain was substituted only by histidine. In experiments was measured affinity 7% in comparison to wild-type insulin, thus this analogue has low affinity (unpublished data of Dr. Jiráček's research group). Each modeled version of the analogue stayed mostly under undetached threshold for whole MD simulation, except for relatively short time, when distances were even above detached threshold (Glu17Hid), but it is be negligible because of mostly undetached conformation in overall. Our theoretical data are in accordance with experimentally measured data of this mutant.

From hydrogen bond analysis can be seen that, by contrast to wild-type insulin, mutants lost hydrogen bond partner Arg43 (original numbering is Arg22 in B-chain), with which are formed relatively stable hydrogen bonds in wild-type insulin.

- **Molecular Dynamics of Insulin and Insulin Analogues (Second Set)**

The second set differs from the previous one in the starting structures, because not all the previous types of mutants were included, only wild-type insulin, one low-affinity analogue (Glu17His) and one analogue with affinity comparable to the wild-type insulin (Ser12Thr).

Wild-Type Insulin

Both types of starting structures maintained undetached conformation during MD and stayed in this inactive conformation until the end of simulation. Only wild-type insulin with N-terminus of B-chain in O-state returned for a short time to detached conformation, but it remained mostly undetached during simulation. The reason why detached conformation of wild-type insulin was not able to persist during simulation could be that it is not energetically favourable. Insulin in detached conformation, as starting structure, has its hydrophobic core exposed to the solvent (water molecules) and it causes destabilization of the hydrophobic core, which has then tendency to aggregate in order to reduce the exposure to solvent (personal communication with Anja Muždalo).

Mutant at Position 12 in A-chain

The insulin analogue structure with detached C-terminus of B-chain and N-terminus of B-chain in T-state gained undetached conformation definitely in time 500ns, which persevered until the end of MD simulation. The same insulin analogue, but with N-terminus of B-chain in O-state rather fluctuated in the sphere of thresholds for most time of MD simulation. The difference from wild-type insulin is that this analogue can be more stable in detached conformation and especially when N-terminus of B-chain is in O-state. It could be suggested that O-state is in this analogue ensures beneficial noncovalent interactions, which could stabilize detached conformation during MD simulation. From hydrogen bond analysis is visible that analogue Ser12Thr in O-state has the most stabilized hydrogen bond with residue Leu16 (A-chain) in comparison with the same analogue in T-state and also with comparison to wild-type insulin and lost hydrogen bond partner Tyr14:N-H. Another feature which can play a role is that the analogue in O-state is not such significant hydrogen donor for atom Gln15:OE1 as this analogue in T-state. In despite of it seems the analogue Ser12Thr is slightly more stable when detached compared to wild-type, but it is also not able to persist in active detached

conformation in overall, as well as wild-type insulin. It could be explained with the hydrophobic core exposure to the solvent, which is not energetically favourable for the analogue and it tends to aggregate (maintain less detached conformation).

Mutation at Position 17 in A-chain

Insulin analogue Glu17His is considered as low-affinity analogue and was expected its different behaving during MD simulation compared to the wild-type insulin and to the analogue Ser12Thr. The result distances showed that the analogue Glu17His in T-state had very stable conformation of B-chain C-terminus and values of these distances fluctuated near the detached threshold. The same analogue but in O-state had values of the distance mostly above detached threshold. Hydrogen bond analysis showed loss of hydrogen bond partners Arg43 (in original numbering Arg22 in B-chain) and Cys20 in the analogue in T-state, when compared to wild-type insulin, while the analogue in O-state lost hydrogen bond formation only with Arg43. One could assume that the mutual feature which differentiate the from wild-type, loss of hydrogen bond partner Arg43, could have contribution to such stable detached conformation of B-chain C-terminus, because it could provide possibilities of new hydrogen bonds formation (such as with Asn21) or lead to favourable composition of noncovalent interactions in the hydrophobic core.

It is necessary to expand sampling (e.g. extend MD simulations, create other models, try other types of MD analysis) to have even more materials for comparison, because as it is relatively novel approach to study site2 interactions via insulin site2 analogues synthesis and computational chemistry methods together, there is not enough of other comparable studies at that time.

- **Docking of Insulin into Remodeled Insulin Receptor ModelS1**

Combinations of the active residues submitted into docking program were chosen on the basis of the previous docking trials with known or suggested residues belonging to site1 interface (Menting *et al*. 2013). Docking with this insulin-receptor model did not result in insulin binding pose, which could sufficiently mimic the natural binding pose, even the best docking poses are far from the original one. I suppose the cause could be that insulin receptor has to undergo even more significant conformational changes allowing proper insulin binding into binding pocket.

- **Docking of Insulin into Insulin Receptor Structure determined by Cryo-EM**

The cryo-EM structure 6CE7 represents the first structure of insulin receptor in insulin-bound state, where also fibronectin domains and site2 interface is able to see. Docking insulin into this insulin receptor gave promising outputs, because in each of the active residues combinations was found a docking pose satisfyingly mimicking the natural binding pose of insulin. Actually, there was also results almost identical as the original structure with bound insulin. This could be evidence that the cryo-EM insulin receptor really captures the insulin-bound conformation. Remodeled ModelS1 used in the previous docking could then represent a model of intermediate conformation between apo-receptor and insulin-bound receptor. Further steps could lead to an investigation of what is between remodeled ModelS1 and the insulin-bound cryo-EM receptor structure, particularly, what triggers such conformational rearrangements of fibronectin domains. This brings us back to the site2 interactions and to the hypothesis of their role in initial docking of insulin to insulin receptor (Weis *et al*. 2018), which also contextualizes impact of their mutations on conformational changes in insulin itself.

7 CONCLUSION

- **Molecular Modelling of Insulin Receptors, Insulin and Insulin Analogues**

Two insulin receptor models were built: one from the crystal structures of the apo-receptor and insulin-bound receptor fragment and the second from a very recent cryo-EM structure. These models were further used for studying interactions, especially in site2, and for docking with insulin. Insulin and insulin analogues were prepared for molecular dynamics analysis.

- **Molecular Dynamics of Insulin and Insulin Analogues**

Two sets of molecular dynamics simulations were performed with insulin and insulin analogues differing in the starting conformation of C-terminus of B-chain: the first one with undetached and the second with detached C-terminus.

In the first set, molecular dynamics analysis showed mutations in site2 residues as able to change dynamics of B-chain C-terminus, which could affect binding to the insulin receptor, thus it could be one of the possible explanations why there was experimentally measured low-affinity by Dr. Jiráček's research group. Particularly, C-terminus of wild-type insulin maintained detached conformation, while the low-affinity analogues remained undetached, what means that water molecule was not able to enter the hydrophobic core, which could be a consequence of non-conservative mutation. The insulin analogues with comparable affinity to wild-type did not maintain detached conformation, which could be explained by lack of sampling and by analogues structures trapped in energy barrier.

In the second set, wild-type insulin was not stable in the initial detached conformation as well as insulin analogue of comparable affinity to wild-type, while low-affinity insulin analogue behaved as the most stable one. Cause of the instability could be in energetic disfavour when the starting structure is detached, because the hydrophobic core is exposed to solvent, what could trigger its aggregation.

- **Docking of Insulin into Insulin Receptor Structures**

Docking insulin into the first insulin receptor model (based on X-ray structures) has only produced poses far from mimicking the natural binding pose of insulin. I explain this by the inability of the used modelling procedure to form a sufficient size of binding pocket, thus larger conformational changes have to occur. However, docking with the cryo-EM insulin receptor structure gave us satisfying outputs, in which the docking poses mimic the natural binding pose almost identically. It indicates that the cryo-EM insulin receptor structure may be used for further modelling with insulin analogues.

8 References

Alam, U., Asghar, O., Azmi, S., & Malik, R. A. (2014). General aspects of diabetes mellitus. Handbook of clinical neurology 126: 211-222.

Brooks, B. R., Brooks III, C. L., Mackerell Jr, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., & *al*. (2009). CHARMM: the biomolecular simulation program. Journal of computational chemistry 30: 1545-1614.

Cantley, J., & Ashcroft, F. M. (2015). Q&A: insulin secretion and type 2 diabetes: why do β-cells fail?. BMC biology 13: 33.

Case, D. A., Cheatham III, T. E., Darden, T., Gohlke, H., Luo, R., Merz Jr, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. Journal of computational chemistry 26: 1668-1688.

Cheng, Y. (2015). Single-particle cryo-EM at crystallographic resolution. Cell 161: 450-457.

Cianfarani, S. (2012). Insulin-like growth factor-II: new roles for an old actor. Frontiers in endocrinology 3: 118.

Croll, T. I., Smith, B. J., Margetts, M. B., Whittaker, J., Weiss, M. A., Ward, C. W., & Lawrence, M. C. (2016). Higher-resolution structure of the human insulin receptor ectodomain: multi-modal inclusion of the insert domain. Structure 24: 469-476.

De Meyts, P. (1994). The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. Diabetologia 37: S135-S148.

De Meyts, P. (2015). Insulin/receptor binding: the last piece of the puzzle? What recent progress on the structure of the insulin/receptor complex tells us (or not) about negative cooperativity and activation. Bioessays 37: 389-397.

Dominguez, C., Boelens, R., & Bonvin, A. M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society 125:1731-1737.

Duboué-Dijon, E., Delcroix, P., Martinez-Seara, H., Hladílková, J., Coufal, P., Křížek, T., & Jungwirth, P. (2018). Binding of divalent cations to insulin: Capillary electrophoresis and molecular simulations. The Journal of Physical Chemistry B 122: 5640-5648.

Garg, S. K., Ellis, S. L., & Ulrich, H. (2005). Insulin glulisine: a new rapid-acting insulin analogue for the treatment of diabetes. Expert opinion on pharmacotherapy 6: 643-651.

Glendorf, T., Sørensen, A. R., Nishimura, E., Pettersson, I., & Kjeldsen, T. (2008). Importance of the solvent-exposed residues of the insulin B chain α-helix for receptor binding. Biochemistry 47: 4743-4751.

Glidden, M. D., Yang, Y., Smith, N. A., Phillips, N. B., Carr, K., Wickramasinghe, N. P., Ismail-Beigi, F., Lawrence, M. L., Smith, B. J., & Weiss, M. A. (2018). Solution structure of an ultra-stable single-chain insulin analog connects protein dynamics to a novel mechanism of receptor binding. Journal of Biological Chemistry 293: 69-88.

Hagedorn, H.C. (1937). Cited according to Hilgenfeld, R., Seipke, G., Berchtold, H., & Owens, D. R. (2014). The evolution of insulin glargine and its continuing contribution to diabetes care. Drugs 74: 911-927.

Havelund, S., Ribel, U., Hubálek, F., Hoeg-Jensen, T., Wahlund, P. O., & Jonassen, I. (2015). Investigation of the physico-chemical properties that enable co-formulation of basal insulin degludec with fast-acting insulin aspart. Pharmaceutical research 32: 2250-2258.

Heise, T., & Mathieu, C. (2017). Impact of the mode of protraction of basal insulin therapies on their pharmacokinetic and pharmacodynamic properties and resulting clinical outcomes. Diabetes, Obesity and Metabolism 19: 3-12.

Heise, T., Nosek, L., Bøttcher, S. G., Hastrup, H., & Haahr, H. (2012). Ultra-long-acting insulin degludec has a flat and stable glucose-lowering effect in type 2 diabetes. Diabetes, Obesity and Metabolism 14: 944-950.

Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of chemical theory and computation 4: 435-447.

Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. Advances and applications in bioinformatics and chemistry 8: 37-47.

Howey, D. C., Bowsher, R. R., Brunelle, R. L., & Woodworth, J. R. (1994). [Lys (B28), Pro (B29)]-human insulin: a rapidly absorbed analogue of human insulin. Diabetes 43: 396-402.

Hua, Q. X., Shoelson, S. E., Kochoyan, M., & Weiss, M. A. (1991). Receptor binding redefined by a structural switch in a mutant human insulin. Nature 354: 238.

Jiráček, J., & Žáková, L. (2017). Structural perspectives of insulin receptor isoform-selective insulin analogs. Frontiers in endocrinology 8: 167.

Jiráček, J., Žáková, L., Antolíková, E., Watson, C. J., Turkenburg, J. P., Dodson, G. G., & Brzozowski, A. M. (2010). Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. Proceedings of the National Academy of Sciences of the United States of America 107: 1966-1970.

Jonassen, I., Havelund, S., Hoeg-Jensen, T., Steensgaard, D. B., Wahlund, P. O., & Ribel, U. (2012). Design of the novel protraction mechanism of insulin degludec, an ultra-long-acting basal insulin. Pharmaceutical research 29: 2104-2114.

Joung, I. S., & Cheatham III, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. The journal of physical chemistry B 112: 9020-9041.

Kornilov, M. Y., Tolstukha, T. A., Kadziauskas, P., & Butkus, E. (1982). A direct method for obtaining Cartesian coordinates of atoms in molecules. Computers & Chemistry 6: 193-195.

Kosinová, L., Veverka, V., Novotná, P., Collinsová, M., Urbanová, M., Moody, N. R., Turkenburg, J.P., Jiráček, J., Brzozowski, A. M., & Žáková, L. (2014). Insight into the structural and biological relevance of the T/R transition of the N-terminus of the B-chain in human insulin. Biochemistry 53: 3392-3402.

Leach, A. R. (2001): Molecular Modelling: Principles and Applications, 2nd edition. Pearson Education Limited, Harlow.

Leibiger, I. B., Leibiger, B., & Berggren, P. O. (2008). Insulin signaling in the pancreatic β-cell. Annual Review of Nutrition 28: 233-251.

Lepore, M., Pampanelli, S., Fanelli, C., Porcellati, F., Bartocci, L., Di Vincenzo, A., Cordoni, C., Costa, E., Brunetti, P., & Bolli, G. B. (2000). Pharmacokinetics and pharmacodynamics of subcutaneous injection of long-acting human insulin analog glargine, NPH insulin, and ultralente human insulin and continuous subcutaneous infusion of insulin lispro. Diabetes 49: 2142-2148.

Lindholm, A. (2002). New insulins in the treatment of diabetes mellitus. Best Practice & Research Clinical Gastroenterology 16: 475-492.

Lindholm, A., & Jacobsen, L. V. (2001). Clinical pharmacokinetics and pharmacodynamics of insulin aspart. Clinical pharmacokinetics 40: 641-659.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of chemical theory and computation 11: 3696-3713.

Markussen, J., Diers, I., Hougaard, P., Langkjaer, L., Norris, K., Snel, L., Sørensen, A. R., Sørensen, E., & Voigt, H. O. (1988). Soluble, prolonged-acting insulin derivatives. III. Degree of protraction, crystallizability and chemical stability of insulins substituted in positions A21, B13, B23, B27 and B30. Protein Engineering, Design and Selection 2: 157-166.

Markussen, J., Havelund, S., Kurtzhals, P., Andersen, A. S., Halstrøm, J., Hasselager, E., Larsen, U. D., Ribel, U., Schäffer, L., Vad, K., & Jonassen, I. (1996). Soluble, fatty acid acylated insulins bind to albumin and show protracted action in pigs. Diabetologia 39: 281-288.

Mavrogiannaki, A. N., & Migdalis, I. N. (2012). Long-acting basal insulin analogs: latest developments and clinical usefulness. Therapeutic advances in chronic disease 3: 249-257.

McKern, N. M., Lawrence, M. C., Streltsov, V. A., Lou, M. Z., Adams, T. E., Lovrecz, G. O., Elleman, T. M., Richards, K. M., Bentley, J. D., Pilling, P. A., & *al*. (2006). Structure of the insulin receptor ectodomain reveals a folded-over conformation. Nature 443: 218-221.

Menting, J. G., Whittaker, J., Margetts, M. B., Whittaker, L. J., Kong, G. K. W., Smith, B. J., Watson, C. J., Žáková, L., Kletvíková, E., Jiráček, J., & *al*. (2013). How insulin engages its primary binding site on the insulin receptor. Nature 493: 241-245.

Menting, J. G., Yang, Y., Chan, S. J., Phillips, N. B., Smith, B. J., Whittaker, J., Wickramasinghe, N. P, Whittaker, L. J., Pandyarajan, V., Wan, Z., & *al*. (2014). Protective hinge in insulin opens to enable its receptor engagement. Proceedings of the National Academy of Sciences of the United States of America 111: E3395-E3404.

Murata, K., & Wolf, M. (2018). Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochimica et Biophysica Acta, General Subjects 1862: 324-334.

Nelson, M. T., Humphrey, W., Gursoy, A., Dalke, A., Kalé, L. V., Skeel, R. D., & Schulten, K. (1996). NAMD: a parallel, object-oriented molecular dynamics program. The International Journal of Supercomputer Applications and High Performance Computing 10: 251-268.

Olokoba, A. B., Obateru, O. A., & Olokoba, L. B. (2012). Type 2 diabetes mellitus: a review of current trends. Oman medical journal 27: 269-273.

Palivec, V., Viola, C. M., Kozak, M., Ganderton, T. R., Křížková, K., Turkenburg, J. P., Halušková, P., Žáková, L., Jiráček, J., Jungwirth, P., & Brzozowski, A. M. (2017). Computational and structural evidence for neurotransmitter-mediated modulation of the oligomeric states of human insulin in storage granules. Journal of Biological Chemistry 292: 8342-8355.

Papaioannou, A., Kuyucak, S., & Kuncic, Z. (2017). Computational study of the activity, dynamics, energetics and conformations of insulin analogues using molecular dynamics simulations: Application to hyperinsulinemia and the critical residue B26. Biochemistry and biophysics reports 11: 182-190.

Petersen, K. F., & Shulman, G. I. (2006). Etiology of insulin resistance. The American journal of medicine 119: S10-S16.

Pfeifer, M. A., Halter, J. B., & Porte Jr, D. (1981). Insulin secretion in diabetes mellitus. The American journal of medicine 70: 579-588.

Porcellati, F., Rossetti, P., Ricci, N. B., Pampanelli, S., Torlone, E., Campos, S. H., Andreoli, A. M., Bolli, G. B., & Fanelli, C. G. (2007). Pharmacokinetics and pharmacodynamics of the long-acting insulin analog glargine after 1 week of use compared with its first administration in subjects with type 1 diabetes. Diabetes care 30: 1261-1263.

Price, D. J., & Brooks III, C. L. (2004). A modified TIP3P water potential for simulation with Ewald summation. The Journal of chemical physics 121: 10096-10103.

Rassam, A. G., Zeise, T. M., Burge, M. R., & Schade, D. S. (1999). Optimal administration of lispro insulin in hyperglycemic type 1 diabetes. Diabetes Care 22: 133-136.

Rinderknecht, E., & Humbel, R. E. (1978). The amino acid sequence of human insulin-like growth factor I and its structural homology with proinsulin. Journal of Biological Chemistry 253: 2769-2776.

Rosenfeld, L. (2002). Insulin: discovery and controversy. Clinical chemistry 48: 2270-2288.

Salomon-Ferrer, R., Case, D. A., & Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. Wiley Interdisciplinary Reviews: Computational Molecular Science 3: 198-210.

Scapin, G., Dandey, V. P., Zhang, Z., Prosise, W., Hruza, A., Kelly, T., Mayhood, T., Strickland, C., Potter, C. S., & Carragher, B. (2018). Structure of the insulin receptor–insulin complex by single-particle cryo-EM analysis. Nature 556: 122-125.

Seipke, G. K, Geisen, H. P, Neubauer, C., Pittius, R., Rosskamp, R., Schwabe, D. (1992). Cited according to Mane, K., Chaluvaraju, K. C., Niranjan, M. S., Zaranappa, T. R., & Manjuthej, T. R. (2012). Review of insulin and its analogues in diabetes mellitus. Journal of basic and clinical pharmacy 3: 283-293.

Shneine, J., Voswinkel, M., Federwisch, M., & Wollmer, A. (2000). Enhancing the TR Transition of Insulin by Helix-promoting Sequence Modifications at the N-Terminal B-Chain. Biological chemistry 381: 127-133.

Smith, G. D., Ciszak, E., & Pangborn, W. (1996). A novel complex of a phenolic derivative with insulin: Structural features related to the T→ R transition. Protein science 5: 1502-1511.

Smith, B. J., Huang, K., Kong, G., Chan, S. J., Nakagawa, S., Menting, J. G., Hu, S., Whittaker, J., Steiner, D. F., Katsoyannis, P. G., & *al*. (2010). Structural resolution of a tandem hormone-binding element in the insulin receptor and its implications for design of peptide agonists. Proceedings of the National Academy of Sciences of the United States of America 107: 6771-6776.

Smith, G. D., Pangborn, W. A., & Blessing, R. H. (2001). Phase changes in $T_3R_3^f$ human insulin: temperature or pressure induced?. Acta Crystallographica Section D: Biological Crystallography 57: 1091-1100.

Smith, G. D., Pangborn, W. A., & Blessing, R. H. (2003). The structure of T6 human insulin at 1.0Å resolution. Acta Crystallographica Section D: Biological Crystallography 59: 474-482.

Smyth, M. S., & Martin, J. H. J. (2000). x Ray crystallography. Molecular Pathology 53: 8-14.

Søeborg, T., Rasmussen, C. H., Mosekilde, E., & Colding-Jørgensen, M. (2012). Bioavailability and variability of biphasic insulin mixtures. European Journal of Pharmaceutical Sciences 46: 198-208.

Stretton, A. O. (2002). The first sequence: Fred Sanger and insulin. Genetics 162: 527-532.

Tibaldi, J. M. (2014). Evolution of insulin: from human to analog. The American journal of medicine 127: S25-S38.

Vajdos, F. F., Ultsch, M., Schaffer, M. L., Deshayes, K. D., Liu, J., Skelton, N. J., & de Vos, A. M. (2001). Crystal structure of human insulin-like growth factor-1: detergent binding inhibits binding protein interactions. Biochemistry 40: 11022-11029.

Vajo, Z., & Duckworth, W. C. (2000). Genetically engineered insulin analogs: diabetes in the new millenium. Pharmacological reviews 52: 1-10.

Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, J., & Bonvin, A. M. J. J. (2016). The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. Journal of molecular biology 428: 720-725.

Vashisth, H. (2015). Theoretical and computational studies of peptides and receptors of the insulin family. Membranes 5: 48-83.

Ward, C. W., Lawrence, M. C., Streltsov, V. A., Adams, T. E., & McKern, N. M. (2007). The insulin and EGF receptor structures: new insights into ligand-induced receptor activation. Trends in biochemical sciences 32: 129-137.

Weis, F., Menting, J. G., Margetts, M. B., Chan, S. J., Xu, Y., Tennagels, N., Wohlfart, P., Thomas Langer, Müller, C. W., Dreyer M. K., & Lawrence, M. C. (2018). The signalling conformation of the insulin receptor ectodomain. Nature Communications 9: 4420.

Weiss, M., Steiner, D. F., & Philipson, L. H. (2014). Insulin biosynthesis, secretion, structure, and structure-activity relationships. Endotext [Internet]. MDText. com, Inc.

Werner, H., Weinstein, D., & Bentov, I. (2008). Similarities and differences between insulin and IGF-I: structures, receptors, and signalling pathways. Archives of physiology and biochemistry 114: 17-22.

Whittaker, J., Whittaker, L. J., Roberts, C. T., Phillips, N. B., Ismail-Beigi, F., Lawrence, M. C., & Weiss, M. A. (2012). α-Helical element at the hormone-binding surface of the insulin receptor functions as a signaling element to activate its tyrosine kinase. Proceedings of the National Academy of Sciences of the United States of America 109: 11166-11171.

Wilcox, G. (2005). Insulin and insulin resistance. Clinical biochemist reviews 26: 19-39.

Yao, Z. P., Zeng, Z. H., Li, H. M., Zhang, Y., Feng, Y. M., & Wang, D. C. (1999). Structure of an insulin dimer in an orthorhombic crystal: the structure analysis of a human insulin mutant (B9 Ser→ Glu). Acta Crystallographica Section D: Biological Crystallography 55: 1524-1532.

Zheng, H., Handing, K. B., Zimmerman, M. D., Shabalin, I. G., Almo, S. C., & Minor, W. (2015). X-ray crystallography over the past decade for novel drug discovery–where are we heading next?. Expert opinion on drug discovery 10: 975-989.

Žáková, L., Kletvíková, E., Lepšík, M., Collinsová, M., Watson, C. J., Turkenburg, J. P., Jiráček, J., & Brzozowski, A. M. (2014). Human insulin analogues modified at the B26 site reveal a hormone conformation that is undetected in the receptor complex. Acta Crystallographica Section D: Biological Crystallography 70: 2765-2774.