A Molecular Dynamics Analysis of Insulin

Submitted by:

Henry Per Andreas Wittler

Bachelor of Science, Major in Physics, 2010, Gothenburg University

Master of Science, Major in Physics with specialization in Materials and Biological Systems, 2013, Gothenburg University

A thesis submitted in total fulfilment
of the requirements for the degree of
Doctor of Philosophy

School of Molecular Sciences
College of Science, Health and Engineering
Department of Chemistry and Physics

La Trobe University
Victoria, Australia

July 2019

Contents

List of Figures

First number indicates which chapter, supplementary figure, are for. The second number are continuous across the whole supplementary.

List of Tables

List of Abbreviations

Word abbreviations. *(Abbreviations in plural ends with an s, e.g. αHs.) Throughout thesis there is a different font for software or computer-commands.*

A	Acceptor atom of HB
αH	α-helix secondary structure
BB/bb	Backbone (i.e. atom N CA C O atoms of a protein)
AC	A-chain of insulin
BC	B-chain of insulin
CHARMM	Chemistry at Harvard Macromolecular Mechanics
C	Cysteine-rich region of insulin receptor
CT	C-Terminal (of an amino-acid chain)
D	Donor atom of HB
DA	Dihedral Angle
DH	Hydrogen of donor in HB
DG/RMD	Distance-Geometry/Restrained Molecular Dynamics
DGR	Other abbreviation for DG/RMD
ES	Ensemble
F1, F2, F3	Fibronection type III domains, 1'st, 2'nd, 3'rd of insulin receptor
GMX	GROMACS (GROningen Machine for Chemical Simulations)
GRO/.gro	GROMACS file format for structural and kinetic information.
HB	Hydrogen Bond
I/i	insulin
IR	Insulin Receptor
IGF	Insulin Growth Factor (if 1 or 2 after abbr. it refers to type)
JM	Juxta-Membrane segment of IR
KPI/kpi	KP-insulin
L1, L2	Leucine-rich repeat domain, 1'st or 2'nd of insulin receptor
LB	Lower Bound (of distance or angle restraint)
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
MC/mc	Main-chain
MD	Molecular Dynamics

MS	Mean Structure (i.e. of insulin only if not indicated otherwise)
NT	N-Terminal (of an amino-acid chain)
Nr/nr	number
PDB/.pdb	Protein Data Bank (can also refer to entry with structural and experimental information, '.pdb' being the file-format containing this information)
QM	Quantum Mechanics
RMSD	Root Mean Square Deviation
RDA	Restrained DA of a DGR ensemble
RHB	Restrained HB distance of a DGR ensemble
RHH	Restrained H to H distance (inferred NOE) of a DGR ensemble
SC/sc	Side-chain
SD	Standard Deviation
TK	Tyrosine Kinase domain of IR
TM	Trans-Membrane domain of IR
VMD	Visual Molecular Dynamics
WB	Within lower and upper bound of a restraint
UB	Upper Bound (of distance or angle restraint)
μIR	micro-insulin receptor

Nomenclature for a single structure or an ensemble of them. *Only structures that are used in the intricate analysis, those noted below, are in bold.*

M1, M2:

Monomer 1 or 2 (of asymmetric dimer unit of PDB 4INS). Analysed in chapter 4.

M12:

Referring to both Monomer 1 and 2 in asymmetric unit. Analysed in chapter 4.

$\mathbf{CF}^{\mathbf{6HN5}}_{\mathbf{(A,B)}}$/***CF(A,B)6HN5***:

The residues within 10 Å of insulin (chain A,B) when bound to IR-fragment as reported in PDB 6HN5. Same for other insulin contiguous structures. Analysed in chapter 4.

$\mathbf{E}^{\mathbf{DGR}}_{\mathbf{kpi}}$/***EkpiDGR***:

Referring, to an ensemble of 20 DGR obtained structures of insulin analogue (kpi) reported

in PDB 2KJJ. Analysed in chapter 5.

$\mathbf{P_{kpi}^{MDm}}$/***PkpiMDm***:
Condition (e.g. P for physiological), indicated simulation method (e.g. MD) and replica (e.g. m) and insulin analogue specified (e.g. kpi). Analysed in chapter 6.

Nomenclature for residue selection of domain atoms or analytical property. *For atom, atom-selection and amino-acid name abbreviations and colouring, see Appendix A .*

(A# S): Locator in main text, of a particular residue, e.g. in chain A, at position #, with one letter residue-name S.

$S_P^{A\#}$: With S referring to one letter sequence-name, A to chain, # to chain number, P to property or selection of atoms. For example, G_{MC}^{A1}, are referring to glycine at chain A position 1, and the main-chain (MC) selection of atoms.

Insulin receptor domains: When distinguishing to a domain of one of the insulin receptor monomers e.g. TK (tyrosine kinase domain) and of the other monomer with an asterisk e.g. TK*. However, if explaining a domain property in general (referring to any monomer) the abbreviation without asterisk is used.

Symbols used in main text

Å	Angstrom
nm	nano-meter
nr	Number
§	Section of main text
S	Supplementary section
#	Arbitrarily numbered chain position or wildcard-atom (e.g. one of hydrogens of a methyl group)
*	Used for distinguishing domains or binding sites of one insulin receptor monomer to the other having no asterisk

Abstract

Insulin is a vital protein hormone, whose discovery and structural understanding has been of critical importance, for the treatment of diabetes mellitus. However, the molecular mechanism, how it acts as an agonist on its cognate receptor, though heavily investigated, remains incompletely answered. There is now a strong body of evidence that indicates that when insulin binds to regions of its receptor, it unlocks a cascade of movements in the receptor and in the subsequent signal pathway. Here are provided some background, for understanding some of insulins structural biology and activity of specific residues, in relation to the structural overviews, here calculated for insulin models pertaining to different environments. Especially novel is a conjectured model of the IR binding up to 4 insulin's and its physiological meaning. Furthermore, a complete intricate dynamical profile model of the solvated insulin monomer has been short of literature, which is here intricately provided by means of molecular dynamics (MD). In this thesis unprecedently long MD simulations of the insulin monomer have been sampled, providing a physiological model of its dynamics.

A large part of this work pertained to the developing of analytical overview methods, for the specific analytical geometric queries, albeit adaptable to other protein models. This overview method was partly used to obtain analytical information and binding surfaces, from already reported insulin structures, i.e. in hexamers, bound to receptor fragments, and a NMR restrained solvent model. This innovatory depiction method is used as an example to get an analytical overview of oligomer aggregates and ligand receptor binding surfaces. The example being a complementing structural analytical overview for the classical hexamer structure of Baker *et al.* [1] and the insulin high affinity bound cross-link to its IR binding region by Weis *et al.* [2]. Furthermore the NMR solvent model by Q. Hua *et al.* [3] were compared to its own restraints and an analytical overview provided. An analogous geometrical analysis was also applied to insulin in explicit solvent, from the highly dynamic and time dependent MD simulations, which were validated by means of comparing to the NMR restraints and extensive sampling. Moreover, the obtained structural analysis of this dynamic solvent MD model, is readily comparable to the other models depicted. The geometrical perspective obtained in this thesis, facilitates an understanding of solvated insulin dynamics and contiguous binding surfaces. The aim of the thesis being to aid in the development of novel insulin analogues or other molecules, for the treatment of diabetes.

Statement of Authorship

I, Henry P.A. Wittler, declare that this thesis titled, "A Molecular Dynamics Analysis of Insulin" and the work presented in it are my own. I confirm that:

- Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma.
- No other person's work has been used without due acknowledgment in the main text of the thesis.
- This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution."

Signed:

Date: 6.7.2019

Author information:

PhD student: Henry P. A. Wittler

ORCID: 0000-0002-9888-0370

Supervisor: Brian J. Smith

ORCID: 0000-0003-0498-1910

Conferences and talks:

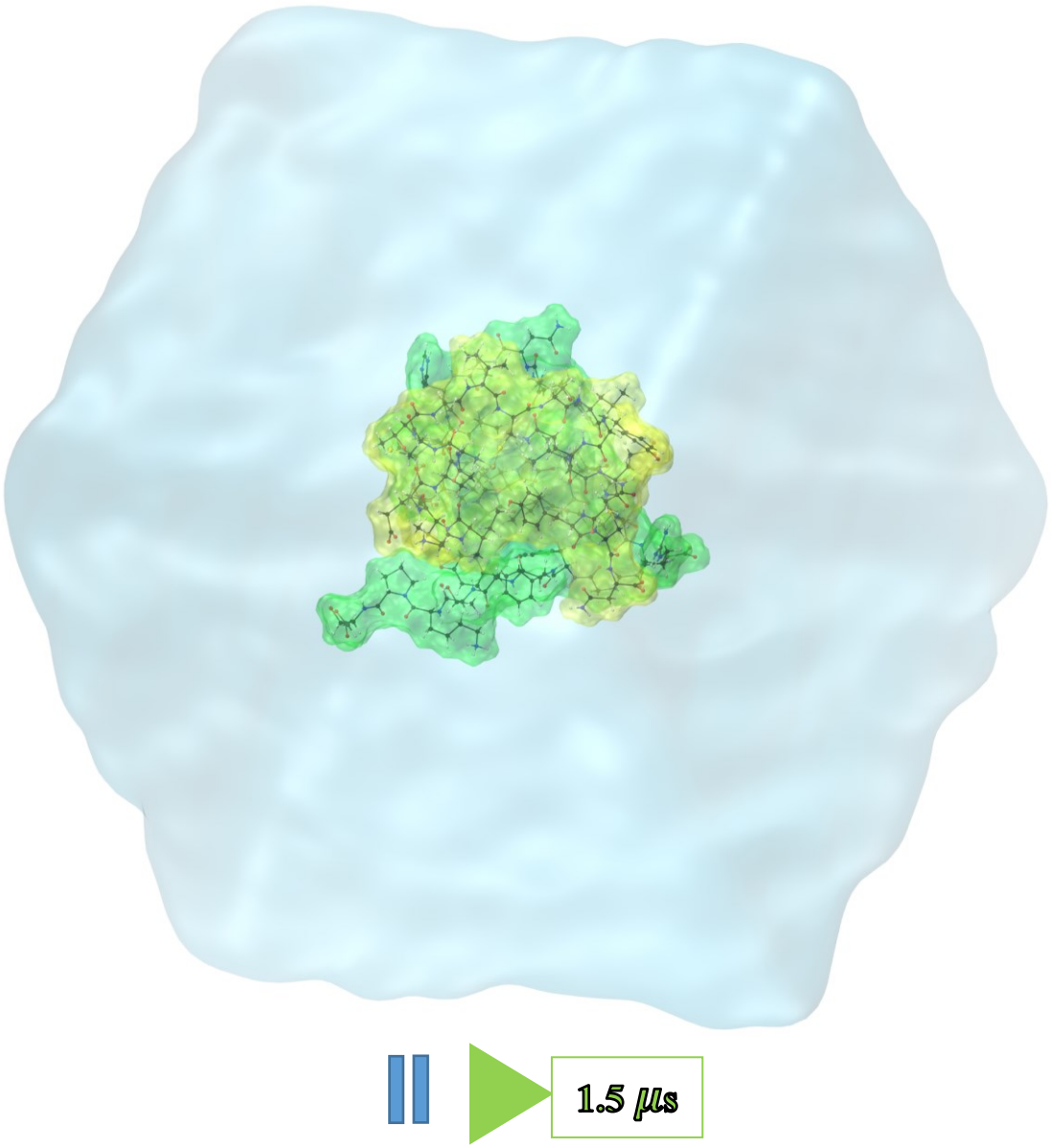
Poster presentation at conferences: MM2014 (July, Brisbane) & MM2015 (December, Sydney), Association of Molecular Modellers of Australasia (AMMA).

Presentation talk at conference: "Molecular Dynamics and Binding of Insulin and Analogues" IGF-OZ 2016 (February, Melbourne), Walter and Eliza Hall Institute of Medical Research.

Eventual pertaining journal publications will be after this published thesis, see link

https://wittler-github.github.io/A_MD_Analysis_of_Insulin/

This Thesis is dedicated to my parents Vasti and Jan-Erik Wittler and the rest of my family, friends and colleagues, who have given me love, encouragement, support and advice needed to complete this thesis. Also dedicated to the people the research may serve and help.



Chapter 1 Introduction & Prologue

> *"Indeed many findings on the structure and properties of proteins were first made with insulin as a model."*

P. De Meyts [4].

Insulin has been much investigated in fields such as physiology, medicine and biochemistry, being a central hormone in metabolism and a vital treatment for diabetes. Diabetes is a chronic disease characterised by heightened levels of blood glucose, originating from either deficient insulin production in the pancreas (i.e. type 1 diabetes), or when the cells of the body cannot effectively use insulin (i.e. type 2 diabetes) [5, 6]. Insulin was first discovered and successfully isolated around 1921, marking a major breakthrough, and was soon thereafter used to save lives of individuals with diabetes [7]. Before insulin was available as a treatment, children with diabetes had a short life expectancy and there was a dire prognosis for adulthood diabetes sufferers [4, 7, 8].

Diabetes is today a growing concern, being one of the most serious causes of sickness and mortality around the world. Today more than 400 million people live with diabetes, compared with 108 million in 1980, hence the global impact have been steadily increasing [5]. The organisation "*Diabetes Australia"* [9] has called it the biggest challenge that confronts Australia's health system. Besides the associated total cost were estimated at about 14.6 billion dollars per annum. In Australia alone, around 1.7 million people have some form of diabetes, in addition more or less than 280 develop the disease every day.

Since its initial use, a variety of insulin analogues have been developed, for optimising the safety and efficacy of treatment for specific patients [10-12]. Though already researched considerably, there is an increasing demand of understanding the insulin signalling pathway, due to this reality of epidemic diabetes. An example being to understand a putative cause of Type 2 diabetes i.e. an acquired resistance of the agonistic action of insulin [13-17]. Rational drug design would hence benefit from a more clear picture of the biophysics of insulin and its receptor binding mechanism, in addition to the following mechanism of the intracellular signal pathway [18, 19]. Moreover as conformational fluctuations promote the degradation of pharmaceutical formulations, understanding the intricate dynamics of the insulin monomer is of broad interest. Besides, the design of ultra-stable formulations

obtained with knowledge of insulin dynamics and receptor binding, has been proposed for humanitarian use in the developing world [3, 20, 21].

In addition having served as lifesaving treatment for diabetes, insulin has also been an important protein in many respects, throughout the past century [4]. An extensive amount of research and testing of insulin has been performed, since around 1921 when Banting and Best first isolated insulin containing extracts [8, 12, 22]. This field of research have been attributed with much value and several Nobel prizes, not the least for the knowledge of its molecular structure and function. Insulin was the first protein to have its primary-sequence determined in the 1950's [4, 23, 24] which was a milestone in biochemistry, revealing that a protein has a defined sequence linked by peptide bonds and consisting only of L amino acids [25]. The three-dimensional structure of insulin, in hexamer form, was first determined by X-ray crystallography in 1969 [26, 27]. Today interest has focused on understanding the structure of the receptor and its response to binding of its cognate ligand, insulin. Furthermore, to visualize and understand the intricate molecular dynamics and mechanism, by which the insulin monomer activates its receptor; how it enables signalling pathways vital for metabolism.

1.1 Statement of the puzzle and rationale of thesis

Future vision of mechanistic visualization of insulin receptor activation

Currently it is not fully visualized and understood, how insulin activates the cognate receptor at the lipid membrane surface of a cell. Particularly the structural changes that accompany the *unlocking* of the receptor by this *key* insulin, initiating a cascade of intracellular signal pathways. The structure of a protein is reputed to be important for its biological function [25]; however the structure of a molecule is dynamic and strongly influenced by its immediate environment. Indeed, a protein's function is strongly coupled to the conformational changes it makes in response to its environment. Hence, to chart and predict the complex biological physics and conformational changes in time, is also an important measure of a proteins function. Since the conformational dynamics of any protein, can indicate how its structure may fluctuate and allow for different interactions, depending on the surrounding biochemistry of the protein [28, 29].

Insulin receptor (IR) activation is evidently not just a simple collision between ligand and receptor, but rather a complex and highly choreographed process, that putatively involves

successive mechanical movements, in the transition towards the activating high-affinity cross-link conformations [30]. Empirical and theoretical studies have solved parts of the puzzle, of the IR activation mechanism [31, 32]. Notwithstanding, being able to visualize the biophysical dynamics of solvated insulin or as it binds to the solvated IR, could elucidate vital features of this puzzling process. To fully simulate an atomic-detailed "movie" of the initial ligand binding to the receptor with full subsequent conformational changes and activation, has not yet been done. However this effort may be expedient in the coming decade even, as simulation and analysis methodologies are steadily becoming more powerful. Conjointly, improvements in experimental techniques such as cryo-EM microscopy [33] and emerging X-ray free electron laser methods [34-38] are likely to provide more information on structural biology and even time resolved dynamics [39-41]. Even more the progressing advances in simulation, graphical depiction and movie-making are revolutionizing our visualization and understanding of biochemistry [42-44]. Hence a complete time resolved dynamical movie of the insulin binding mechanism and even large parts of its resulting signal pathways is here postulated to be progressively achieved. The full IR ligands binding mechanism appears as one of the more important biochemical systems to be modelled and visualized.

Computational biophysical studies of the solvated insulin monomer

The functional native state of a soluble protein is not determined by a static structure, rather it is described by a dynamic ensemble of conformations (or structures), partitioned in an energy-landscape as determined by statistical thermodynamics [45, 46]. Computational methods to simulate dynamic conformations have improved immensely, it is expected that the predictions that simulations can afford, will improve significantly with the increase in computing power over the next few decades [47, 48]. Conjointly, the method of MD [49, 50] has until now produced many applications for depicting many kinds of biologically relevant conformations [51, 52]. Even though algorithms for MD is simplistic in its approximations, it provides an effective probe on the atomic scale, i.e. for the study of biophysical structure and dynamics. Hence the technique may even be referred to as a computational recording atom-, molecular-, macromolecular-scope with regards to time and space. The technique progressively enabling simulations of macromolecules, with unprecedented size and time scales. Despite the advances in recent decades, solvated systems required for the size of the insulin receptor, has been prohibitively expensive to

simulate. Heretofore, simulating the insulin monomer in explicit solvent using long timescale MD, is a much more feasible endeavour. The opportunity to explore the solution conformations of insulin for extended timescales, offers potential to gain new insights in atomic detail. Facilitating a time-dependent statistical analysis of observables, including hydrogen bonding, torsion angles and structural distances. Such a structurally dynamical profile of insulin may provide a clearer picture of its intricate biophysics that are vital for its biology. In addition, the mean structure of a MD simulation, may tell us how the average shape allows for its multiple interactions, depending on environment, and how the monomer can readily form oligomers or bind to its receptor. Thus, in this report, the focus is to provide a clearer picture of the interactions and dynamics of the insulin monomer. Besides, the reliability of a MD simulation method depends on the choice of parameters and algorithms. Hence the MD methodology, including parameters, used in these calculations were carefully considered, to assure a more accurate view of insulin biophysics in solution.

Previous studies of insulin monomer with molecular dynamics

Since the inception of MD techniques [53], there have been several computational studies of the insulin monomer. Each study exploring different aspects of the molecule and with diverse interpretations being derived. A picosecond simulation of the crystalline form of insulin was conducted in the early 1980's [54, 55], it were an elaborate analysis, but the short timeframe could not allow extensive sampling of dynamics. Later in the 2000's, simulations over nanosecond time-frames were performed on insulin monomers and dimers, which revealed some statistics [56, 57]. The work presented in this thesis is to an extent reminiscent to these previous studies which have given some inspiration. Moreover, a complementary study of the insulin monomer has concurrently been reported [58, 59], a main investigation by them were for particular mutations, discussing their relevance in interactions within the core of the molecule. Using a similar simulation method as the latter mentioned study, in this thesis it were performed significantly longer simulations, with the aim here to give an encompassing analytical profile of insulin in solution. Some aspects of these abovementioned simulations are verified by the results in this thesis.

Verifying simulation results by statistical replicas

An important consideration when performing a MD simulation, since it is a statistical

method, is to determine whether a system has been sufficiently sampled. Moreover, how to decide if the simulation or any geometrical observable has converged. The quality of any extracted geometrical observable, from a MD simulation, depends on the statistical quality of the sampled conformational space [60]. Thus, it is common practice to perform multiple replicas with independent starting states, or otherwise validate with experiment. The statistical uncertainty of a MD simulation, depends (to an extent) on the inverse of simulation time (~$1/\sqrt{simulation\ time}$) [60]. Nonetheless, since the method is stochastic, different replicas are expected to exhibit varying or even divergent structural movements. Performing MD simulations is hence a venture, since it is not certain that one will capture the structural event that one may anticipate. Hence all derived conclusions of a simulation has to be inferred with sound biophysical reasoning, moreover weighted by available knowledge of empirical, statistical and experimental kind. Accordingly, in the work presented here, results of multiple replicas are reported, the aim being also to evaluate the accuracy and reliability of a MD simulation of a protein such as insulin.

As an aside, I have recently applied statistical replicas to also simulations in image retrieval in the field of X-ray free electron science [35].

Verifying simulation results with experiment and their derived models

There are data from various methods that can be integrated in structural biology models [61]. On the other hand there are many physical properties of biomolecules that one can estimate from MD simulations, that are comparable with those derived from various methods [25, 29]. Because insulin dissolves into a monomer in the bloodstream and finally binds to its receptor, nuclear magnetic resonance (NMR) studies of engineered monomers, have in the past attracted considerable attention. There have been many NMR investigations of monomer analogues, yielding derived insulin modelled structures [3, 20, 62-64]. In particular we have followed the work of Q. Hua *et al.* [3], who reported a "Distance Geometry/Restrained Molecular dynamics" (DG/RMD or DGR) ensemble of insulin structures. This ensemble is in the protein data bank (PDB) entry 2KJJ, which contains restraints of hydrogen distances, hydrogen bonds and dihedral angles. Their ensemble seemingly captures a structure of the insulin monomer, that gives a resemblance to a protomer in a hexamer. However, their ensemble does not capture the elaborate dynamics that can occur, for example the transient breaking and reforming of HBs. A query being can MD simulations starting from a DGR model be used to visualize the dynamics and still

satisfy the restraints.

Furthermore, analytical structural overviews of several different models of insulin in various environments, can elucidate otherwise unseen vantage-points; in this way it is provided structural overviews of also single structures, facilitating for observers to study inter-structural relationships. A relevant single structure as obtained from X-ray crystallography are monomers in a hexamer, we chose to compare it to a well known structure obtained in the later 1980's; which have been used in comparison for elucidating results in relation to insulin structure and function [3, 65-68]. Other relevant single structures are those of fragmented IR structures [2, 69-71], with the monomer bound to parts of its binding site, obtained from crystallographic and electron microscopy methods. Single structures may also be compared to analogous structural overviews of structure ensembles relating to solvent conditions as obtained by DGR and MD methods.

Obtaining an analytical structural overview of any protein system

Biological applications of MD had its beginning close around the decade 1970, since then there has been a vast rise in computational capability, and simulation is likely to play an even more important role in the future [53]. There is dawning even now a vast computation capability in many areas of science and development, not to say the least for the study of biophysical data [47]. Hence, there is a meeting demand in harvesting the computational capabilities that can be applied to biological physics and medical sciences, for obtaining reliable information. There are various simulation and/or analysis packages, for analysing single structures or ensembles of them [48, 50, 72-81], some which are still under development. However, for anyone undertaking to analyse a simulation, the specific query may be out of reach of any available package, or otherwise needing some refinement of the original code, which might not be expedient for external researchers. Hence in this field there are need of standardization and simplified approaches, so that any novice researcher may obtain reliable information from relevant vantage-points, in a way readily accessible. Analysing and seeing the whole overview of a biochemical structure intricately can be a daunting task, not the least for a simulated trajectory of thousands or millions of time-frames. A difficult task in MD society has been to see common or differing traits of say a protein in different environments, or between many replicas of a simulation for obtaining statistical profiles of any observable. This need for improvement will hence likely advance this area in the coming years. Nevertheless in this thesis, the analysis code was developed

much from the ground-up, on scientific ground laid by others, as described and indicated in main text. The analysis emphasizing on how to obtain a comparable vantage-point of geometrical properties, as pertains to either replicas of MD insulin structure ensembles and/or single structures in various environments. Hence, this analytical overview approach is depicted for several insulin systems in this thesis, which are treated separately, however facilitating a comparison between these single insulin structures or ensembles of them. The analytical overviews rely much on vector-graphics such as filetypes EMF (enhanced meta-files) and SVG (scalable vector graphics). Vector-graphics are zoomable and are an exemplary way of condensing a large amount of information into a "portable document format". The viewers Okular (UNIX) and Acrobat Reader DC supports zooming best, decently also in Microsoft Edge which work well for opening many windows. In a way the analytical over-view of insulin, is meant to be an exemplary prototype or depiction model for other biological systems, to readily compare MD replicas in between or structures obtained with other methods. These innovative depictions of structure in this thesis, may serve as inspiration for other peer researchers or even adapted in future analysis-packages. There is an era of increasing amount of data storage available, hence the approach in this thesis can serve as inspiration for condensing a large amount of biophysical data in any portable format. Even more, there is an advent of the development of overview standardized analytical "black box" software, for the readily able output and visualisation of any relevant biological physics. Hence information from simulations will become increasingly more intelligible and allow for vivid analyses for the study of biological systems [43].

1.2 Thesis outline and overview

- **Chapter 1:** Thesis introduction and prologue.

- **Chapter 2:** Here it will be provided an interpreted review regarding the structural and functional biology of insulin, from its storage form to disassembly into monomers and mechanism of binding to its cognate receptor. Specifically, reviewing relevant insulin physiology (§2.1), putative binding residues and overall mechanism, focusing on the monomers primary sequence (§2.2) tertiary structure (§2.3), as a protomer in a hexamer (§2.4), monomer binding to IR and further conjecturing a case for multiple monomers binding to IR (§2.5). *This chapter is also in one sense an enriching chapter to the rest of thesis. However as an example, the conjectured model of §2.5.2.2, is not a main*

hypothesis of this thesis, only included for awareness in this field of research, a vision of its conceivable simulation are nevertheless given in §7.2.

- **Chapter 3:** Here it will be provided the background for enabling simulation of a solvated insulin monomer and analysis of any insulin monomer structure. That is the simulation conditions and plausible ionization states at a certain pH (§3.1), moreover a brief overview of MD and the method of simulation (§3.2), followed by a description of the analysis used for the study of insulin, i.e., how to obtain an analytical structural overview of any atom-specific protein model (§3.3). Furthermore, this chapter is meant also as a reference for the subsequent results chapters with supplementary.

- **Chapter 4:** Here it will be provided intricate structural analyses, from a distinct vantage point of important structures, i.e. of a renowned high resolution insulin hexamer structure (§4.1), moreover of various lower resolution fragmented IR structures, culminating in a closer look of the high-affinity bound insulin in its signalling conformation (§4.2). The chapter ends with a concluding statement (§4.3).

- **Chapter 5**: Here it will be focused mainly on obtaining an analytical structural overview from an ensemble of highly restrained DGR solution models. The original method is restated in §5.1, after which the structures are analysed and overviewed in §5.2 and compared to its restraints from which it were derived. A structural overview and a few residue-profiles are provided in §5.2.5, elucidating some biophysical structure. The chapter ends with a concluding statement in §5.3. *This chapter is really in conjunction with chapter 6, there are much similar representations, thus for clarity, they are separated in chapters. The chapter is in a way depicting the foundation that can be built upon by sampling the conformational space with MD simulations, and how the restraints respectively relate to both ensembles of insulin structures.*

- **Chapter 6:** Here it will be investigated MD simulations with different parameters resulting in replicated distinct trajectories. Comparing each replica's respective congruence to the restraints of Q. Hua *et al.,* for rich sampling and evaluation of results, obtaining a thorough analytical dynamic overview of the most representative model. The method of obtaining the 9 replicas are stated in §6.1, after which the MD trajectory is analysed and depicted in §6.2, ending in a discussion and conclusion in §6.3.

- **Chapter 7:** Thesis summary and epilogue will be the ending concluding chapter.

Chapter 2 Construed Review of Insulin with Receptor Structure and Binding

Reviewing General Puzzle-pieces of Insulin Structure and Binding and Conjecturing a Model of Multiple Insulins Binding to One Receptor

> *"So that's intriguing, insulin disassembles to bind its receptor."*

M. Lawrence [82].

In order to interpret insulin structure and function, it is informing to get an understanding of its biology from literature, here providing a deduced perspective in the following sections.

2.1 Some insulin related physiology

The protein hormone insulin is produced in pancreatic β-cells, where in the presence of Zn^{2+} ions it self-associates into hexamers (see Figure 2.1) and are stored within vesicles [83]. In response to heightened levels of glucose in the blood, vesicle exocytosis precedes dissociation of the hexamers into dimers, then into biologically active monomers, thereby releasing it into the bloodstream as a zinc free monomer [10, 20, 84, 85]. Then the insulin monomer function as an agonist, i.e. it binds to its receptor on a cell surface and transmits a signal that activates intracellular auto-phosphorylation through its enzymatic kinase site; which further catalyses the phosphorylation of substrates and initiate various signalling pathways [25, 86, 87]. A major branch of these pathways results in glucose transporters to appear on the cells surface, which then transports glucose into the cell [25]. Hence insulin enables glucose to fuel the metabolism of cells such as those composing skeletal muscles, fat-tissues and liver.

Being a chief hormone in metabolic control and following the daily biological rhythm (e.g. cyclic blood glucose levels). In particular at times of digestion where the insulin blood content increases after a meal and within hours gradually returns to its basal level (typically between meals about 60-80 pM) [88]. Putatively insulin secretion is not continuous, but rather like oscillating wave pulses (of period ~4 minutes) and in portavenous blood in anti-synchronous phase from glucagon (an opposing hormone produced in the pancreatic α-cells causing raise of blood glucose) [89].

Most of the insulins is absorbed by the liver during the first passage, after which the oscillating pulses (and insulin concentration) of the pulses is greatly attenuated [90], yielding more or less 0.1-0.8 nM in the peripheral bloodstream. One source has stated that continuous exposing of insulin on receptors initiates their down-regulating internalisation process in the host cells, that the oscillation has meaning in reducing the receptors down-regulation and limit the need for insulin [91].

The rate of internalization or otherwise inhibition of the receptor have been inferred as proportional to insulin concentration (in the range around nM to µM) by multiple sources, indicating insulin concentration as a regulating factor [92-95]. Putatively then the subsequent insulin binding induces intracellular auto-phosphorylation and triggers internalization of the receptor complex into the intracellular endosome and lysosome system, leading to dissociation and breakdown of insulin(s) and even inactivation of receptor by phosphatase action and further recycling back to the cell membrane [96-103]. In §2.5.2.2 it is further construed and conjectured, that internalization is plausibly activated when multiple insulins have bound 1 receptor.

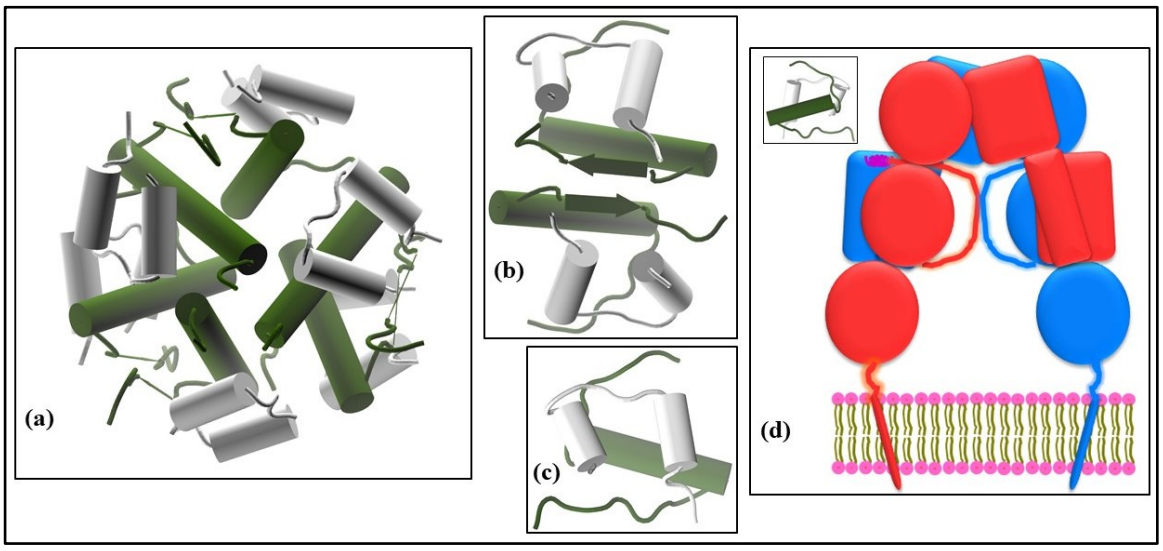


Figure 2.1*: Dissociation order illustration of insulin.* ***(a)*** *Storage hexamer of insulin.* ***(b)*** *Intermediate dimer.* ***(c)*** *Biologically active monomer.* ***(d)*** *Initial to binding of monomer (top left, same as in (c) backwards) to an insulin receptor ectodomain binding region. Receptor monomers in red and blue respectively, cell-membrane and trans-membrane domain included. Moreover an αCT helix in purple, where insulin is binding, the αCT of the second identical binding site are occluded in this schematic. Hexamer in (a) are from PDB 1TRZ (biological assembly 3), showing the R-state (more α-helix at initial NT green-chain) trimer at front and T-state (initial NT green-chain in a loop closer to insulin core) trimer at back. Dimer in (b) are an asymmetric T-state dimer from a hexamer crystal model (PDB 4INS, biological assembly 7). The monomer in a (T-state like folding) in (c) are from the NMR derived solvent model in PDB 2KJJ.*

2.2 Primary structure of human insulin and binding surfaces in an oligomer and bound to a receptor bound structure

The folded peptide hormone, a rather small globular protein, has defined surfaces of secondary and tertiary structure, which putatively are fluctuating depending on chemical environment (e.g. tumbling in solvent or stable in a crystal lattice). Nevertheless, the small surface area of insulin is suggesting that the same residues are involved in forming various binding surfaces. Extensive biochemical characterization, has before demonstrated, that there is overlap of residues involved in binding surface for monomer self-assembly as for high affinity cross-link to IR [4, 32]. Accordingly, it is here shown the primary structure of insulin in Figure 2.2, which are moreover depicting (for the folded protein) inferred residues involved in binding surfaces in two unique structures. Here it is revealed, that binding residues of insulin as a T-state monomer in a hexamer, have to some extent overlap, with those residues of insulin bound with high affinity to the IR. Albeit the two structures compared have unique atomic structures, the IR bound structure were obtained at a lower resolution and hence may be less accurate.

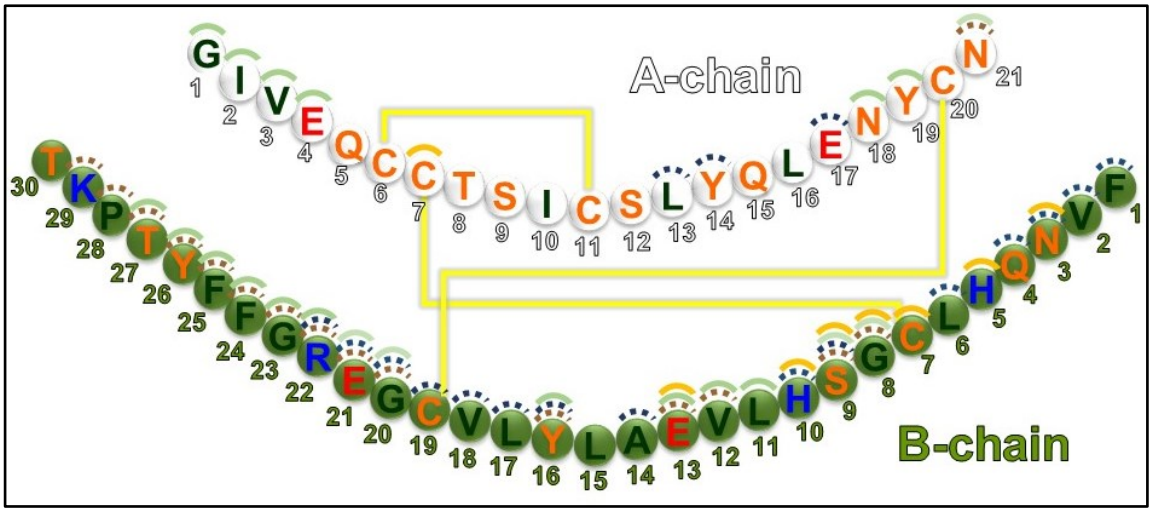


***Figure 2.2**: Primary structure of insulin and two bound structures binding surfaces. Residues shown with polarity of charge colouring (same as in Figure A1). Cysteine links are shown with yellow lines and are between intrachain C^{A6} to C^{A11} and interchain C^{A7} to C^{B7} and C^{A20} to C^{B19}. Binding surfaces (in vicinity of insulin) inferred through VMD: residue included if any non-hydrogen atom-moiety (including MC & SC atoms) were within 5.0 Å of the indicated binding region of insulin (without hydrogens). Residues in the binding surfaces of a hexamer assembled of monomers in a T-state (PDB 4INS, biological assembly 3, 1.5 Å resolution, no included hydrogens, Baker et al. [1]): "dotted arcs", brown for dimer and dark-blue for hexamer surfaces. Residues in the binding surface to IR ectodomain in a high affinity bound structure (PDB 6HN5 3.2 Å resolution, no hydrogens, Weis et al.[2]): "full arcs", light-green for site 1 and in orange for site 2. See text for further explanation.*

A note on similar residues among vertebrate species

There are similarities of residues in the primary sequence of insulin, which can be found in the very diverse species among vertebrates. Depending on classification [67, 104], the following residues may be considered highly conserved or invariant: G^{A1}, I^{A2}, V^{A3}, Y^{A19}, N^{A21}, L^{B6}, G^{B8}, L^{B11}, V^{B12}, G^{B23} and F^{B24}, in addition to the cysteines. Conservative substitutions exist also, e.g. "$F^{B25} \rightarrow$ Tyr" and "$E^{B13} \rightarrow$ Asp. The biochemical activity and function of insulin in any vertebrate species are naturally unique, even though more or less similar. Notwithstanding, studies of this homologous protein have elucidated core structure and function of human insulin. As an example, studies of bovine ($T^{A8} \rightarrow A^{A8}$, $I^{A10} \rightarrow V^{A10}$, $T^{A30} \rightarrow A^{A30}$) and porcine insulin ($T^{A30} \rightarrow A^{A30}$) have revealed biochemical activity, sequence and structure [1, 23]. Besides, porcine insulin has been reported to have full receptor binding affinity [105].

2.3 Tertiary structure of a dynamically restrained model of the monomer in solution

The inherent dynamics of the insulin monomer is evidently essential, for allowing the different binding surfaces relating to self-assembly and receptor binding [32, 63, 69]. Hence its inherent biophysics, e.g. the momentous breaking and forming of HBs, must be relevant for its mechanism. In contrast, in Figure 2.3, it is shown an insulin structure in solvent from NMR derived and restrained simulation models. This structure putatively represents an average structure of insulin in solution, which resembles the tertiary structure of a T-state protomer in a hexamer. Nevertheless, the ensemble of structures is dynamically restrained, and hence inferred to underestimate conformational fluctuations, although dynamics were to an extent inferred with other methods by Q. Hua *et al.* [3]. This solution model of the insulin monomer consists of an A-chain of 21 residues made of two α-helice's (A1-A9 & A13-20), separated by a central region of extended polypeptide. The B-chain of 30 residues, has two strands separated by a central α-helix (B8-B19). The two chains are linked by two interchain disulphide bonds between A7-B7 and A20-B19, whereas the A-chain has an intrachain disulphide bond between A6-A11. This model is congruent with previous understanding of the insulin monomer, derived from e.g. crystallography and NMR [1, 3, 4]. That is to say that the interior of the monomer is mainly hydrophobic and the surface having a composition of hydrophobic, polar, negative and positive charge; whereby one surface are largely hydrophobic, being flat and mainly aromatic and buried upon dimer

formation; another hydrophobic surface being buried when the dimers assemble to form hexamers [106].

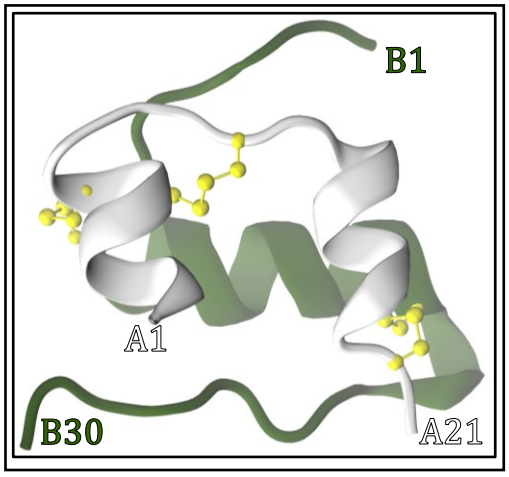


***Figure 2.3**: Tertiary structure of a restrained insulin model in solution . This depicts the NT B-chain loop in the T-state. The atoms and bonds of disulphide links coloured yellow. Shown is the 7'th model out of 20 reported for conditions 298 K, 0.1 mM salt concentration, pH 7.4. Structure obtained and published by Q. Hua et al. [3] and reported in PDB 2KJJ.*

2.4 Quaternary structure of insulin in T-state hexamer crystals

Unique structures of crystal insulin hexamers have been determined for various cosolvents and mutations of residues, for which characterization has been reviewed elsewhere [68, 106]. For the native insulin hexamer, a few different kinds have been discovered, with an equilibrium between the T and R states of its monomer components. That is to say, depending on the allosteric cosolvent; the six monomers can form two trimers T-state (T_6) [1, 26], or two trimers R-state (R_6) [107], or one trimer of each T-, R-state (T_3R_3) [108]. If an insulin monomer has the B1-B8 residues in an extended β-strand, closer to the hydrophobic core, it is called the T-state. On the other hand, when B1-B8 form a continued α-helix from B9-B19, it is referred to as the R-state. Contrasted to when the helix only extends from B4 to B19, and the B1-B3 residues being extended or "frayed", it is referred to as R^f-state [109, 110]. The intricate structure and in particular the B1-B8 residue contacts are thus not the same in the different forms of hexamers [68]. Apparently, the R-state has only been observed within hexamers, as also stated elsewhere [111]. Noteworthy is that the storage of insulin within vesicles have been speculated to take the form of T_3R_3, or even R_6 like structures, due to the R-state forms being more stable and allosteric ligand sites being preserved in many vertebrate species [112]. Remarkably, insulin putatively has averagely a T-state like conformation when it circulates in the blood as a biologically active monomer. Hence here it is chosen, to investigate the T-state hexamer surfaces more thoroughly. A detailed report of the crystal structure of 2Zn porcine insulin hexamer in T_6

form was written by Baker *et al.* [1], being often cited it has been referred to as a bible of insulin structure [18]. The original report has information of residue contacts and hydrogen bonding, of the hexamer structure, including also other atoms such as zinc and water. The hexamer consists of three asymmetrical dimer-units monomer 1 and 2, here it is inferred the protein dimer and hexamer contact surfaces, as illustrated in Figure 2.4 (and Figure 2.2). Between the monomers of a dimer are the form of the flat hydrophobic and mainly aromatic surface, the three dimers are assembled by zinc and water coordination [1, 113], with polar and non-polar residues being buried between them. The inter-dimer packing surface is looser than that of the intra-dimer. There is a strong interaction of four hydrogen bonds that stabilizes each dimer, between the strands of the two CT B-chains.

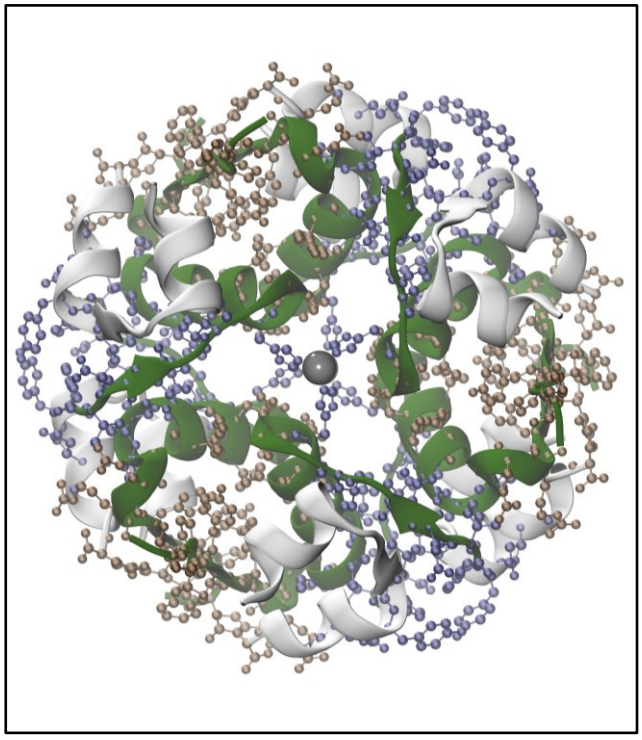


Figure 2.4*: Schematic of binding surfaces of monomers in a T_6 hexamer. Trimer of monomer 2 at front, and of monomer 1 at back. In the middle, each monomer of a trimer contributes a H^{B10} for coordination with a zinc-ion (dark-gray), respectively. Only showing the residues (both MC and SC) at approximate binding surfaces, inferred by VMD that have any non-hydrogen atom-moiety within 5 Å. However, with residues at surfaces being varying in symmetry for monomer 1 and 2 but have same residues near binding surface. Residues located at the dimer surface (brown) and at hexamer surface (dark-blue), same as in Figure 2.2 (c.f. [1, 4, 30, 68]). Structure from PDB 4INS (biological assembly 3, configuration A), the depiction here is without hydrogens, and atom-colour overlap for residues that are in both surfaces [1].*

A note on Lispro

An insulin analogue, Lispro, are used in treatment of diabetes, being the active component of pharmaceutical Humalog® (Eli Lilly and Co.) [12]. Compared to native human insulin, Lispro, has an interchange of two B-chain CT residues ($P^{B28} \Leftrightarrow K^{B29}$), hence may also be referred to as $K^{B28}P^{B29}$-insulin, or KPI (KP-insulin) for short [109]. The accompanying perturbation at the dimer interface, gives an accelerated disassembly of the zinc insulin hexamer, upon subcutaneous injection. Since the $K^{B28}P^{B29}$ interchange of KP-insulin, was

found to disrupt the otherwise hydrophobic effects of P^{B28}, with residues B20-B23, moreover weakening the hydrogen bonds between residues F^{B24} and Y^{B26}, being critical for dimer formation [109]. The interchanged residues B28-B29, is believed not to alter the receptor binding surface of the hormone, since KPI retains full biochemical potency [3]. This is particularly mentioned since the solvent model(s) of Figure 2.3 were obtained by the LisPro analogue.

2.5 Ligand binding to its cognate receptor

Insulin receptor binding has been a much studied topic and a great many pieces has been collected, some appear more blurred than others and the complete puzzle is yet to be discovered. There is a wealth of information, however incomplete, regarding the structural features that constitute the ligand receptor binding domain, and conjointly the full molecular mechanism of receptor engagement. Somewhat different views and sometimes contradicting statements has been reported through the years. Thus this section is attempting to consolidate a range of literature, regarding the insulin receptor binding puzzle, and to add a conjecture of multiple insulins activating its receptor.

2.5.1 Insulin receptor

The insulin receptor (IR) belongs to the receptor tyrosine kinase family of cell-surface receptors; whose variation, relation and physiological significance are described elsewhere [25, 114-118]. A homologous family is that of the ligand/receptor insulin-like growth factor (IGF), with hormones IGF-I and IGF-II [116, 119].

There are two different isoforms of IR (IR-A and IR-B). These differs in the sequence-length of 12 residues at the CT end of the α-subunit (αCT), with IR-A numbering inserted between 716 and 717 before the αCT end at residue 719 [120]. The number of residues of IR-A is 1346 and of IR-B is 1358 [121-124]. The IR-A form has about 1.5 stronger affinity, and a 2-fold higher dissociation rate; reasoned to be due to the extra 12 residues of IR-B, obstructing insulin binding and dissociation [125]. Congruent with the larger IGF-II peptide, binding IR-A with high-affinity and activating differing kinase activity, but not for IR-B [126, 127].

Furthermore, a structure of the unliganded IR-A ectodomain (apo receptor) in complex with four antigen-binding fragments (2xFab 83-7, 2xFab 83-14), has been solved, revealing a folded over "Λ" conformation [120, 128, 129]. The insulin receptor (domain schematic in

Figure 2.5) is a dimer of two identical αβ half-receptors, each composed of an α-subunit and a β-subunit. The extracellular α-subunit is a chain of about 720 amino-acid residues, moreover heavily glycosylated. The β -subunit (about 650 residues) starts on the extracellular side and spans the membrane to the cytoplasmic side, where it ends in the C-terminal end segment (CE). Each ectodomain part of monomer contains a leucine-rich repeat (L1), a cysteine-rich (C) region and a second leucine-rich repeat (L2) domain, followed by three fibronectin type III domains (F1, F2, F3) [130]. Where F2 contains an insert domain ($ID\alpha$, $ID\beta$) of about 120 residues, within which lies the α-β cleavage site [31, 131, 132]. Each α-subunit is linked to a β-subunit via disulphide bonding, to form an αβ receptor monomer, in addition, the two receptor monomers have binding interactions between and are linked by a few disulphide bonds [31]. The receptor monomers β-subpart are continuing CT to the F3 domain; in a single transmembrane (TM) helix; followed by an about 40-residue intracellular juxtamembrane (JM) region; followed by the TK catalytic domain; and then by an about 115-residue CT end (CE) [32, 130, 131].

The ectodomain of the IR has two identical insulin binding regions, each having two binding sites. Putatively in a binding region, site 1* are composed of L1* and the αCT helix [69, 129]. In addition, at least partly, site 2 is thought to be in the proximity of the junction between F1 and F2 domain [31, 32, 71, 131, 132], near site 1* in the apo-receptor [129]. In addition, regions of L2 has been implicated as involved in site 2 binding [4, 18, 132, 133]. The second identical insulin binding region being composed of site 1&2*. Putatively the high affinity cross-linked binding are in a 1:2 stoichiometry [18], however (as conjectured in e.g. §2.5.2.2) a 2:2 stoichiometry with equal binding affinity appears plausible. Nevertheless, the binding of ligand(s) translates as a signal across the TM domain, concomitantly causing *trans*-phosphorylation of tyrosines in the two β-subunits TK, initiating cascade signal pathways [14, 25, 31, 132, 134, 135].

There are 13 intracellular tyrosines in each endo β-subunit, where a number of them may be phosphorylated in response to ligand binding including; Y^{960} in the JM; Y^{1146}, Y^{1150}, Y^{1151} in the TK activation loop (occluding the kinase catalytic active site in apo-IR); Y^{1316}, Y^{1322} in the CE tail [136-138]. Where the auto-phosphorylation creates phospho-tyrosine recruitment sites for downstream signalling proteins such as the IR substrate (IRS) [86].

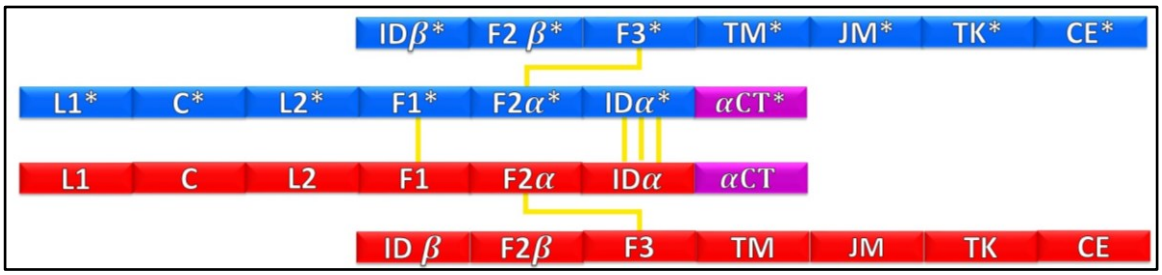


***Figure 2.5**: Schematic of the dimeric insulin receptor sequence. Left to right NT to CT direction (continuing from αCT to IDβ, wherein monomers are cleaved). One receptor monomer is coloured in red (without asterisk used also naturally in main text as a general abbreviations for both monomers), the other in blue having domain names with an asterisk ''*" (if distinguishing from other monomer in main text or other figures also), purple region for αCT segment (704-719) in both monomers. Gold-coloured links denote disulphide bonds. See text for further explanation.*

2.5.2 Characteristics of insulin receptor binding and signalling

2.5.2.1 Some previous understanding & model of one insulin binding activation

Insulin binding to the receptor homodimer, is long known to be characterized by negative cooperativity among the receptor binding regions. In addition, by curvilinear Scatchard plots, i.e. a plot of "bound/free traced ligand" as a function of bound unlabelled ligand, linear in the case of simple non-cooperative binding. In addition, by a bell-shaped dose-response curve for the tracer insulin dissociation-acceleration effect by unlabelled insulin [18, 30, 139, 140].

These features have been explained with a model, by which the α-subunits of the receptor monomers are arranged in an antiparallel symmetry, each having two ligand binding sites [30]. Insulin was proposed, to bind with low affinity to first e.g. site 1*, then to form a high affinity cross-linked binding to site 2. The formation of this cross-link, between e.g. the site 1* & 2 pair, reduces the capacity of ligand to form a cross-link of the alternate site 1 & 2* pair (i.e. negative cooperativity), though maintaining ability of insulin to bind singly to the individual components of the alternate site 1, 2* pair (i.e. a bell-shaped dose response) [31]. This model have been formulated in a mathematical "harmonic oscillator" model (Figure 2.6), that were shown to be fitting kinetic parameters, relating to insulin receptor binding [32, 125, 141]; whereby the binding of insulin to both site 1* and 2 (or alternatively site 1 and 2*), produces a conformational change in the insulin receptor, required for its activation. The insulin receptor were assumed, to change from its inactive symmetrical conformation, to an activated tilted conformation, concurrent with e.g. the site 1*, 2 pair being moved

closer, and the other site 1, 2* pair being separated [141]. This tilted (activated) conformation, being at a higher free energy state. Moreover in the absence of insulin, equilibrium being shifted strongly, to the energetically more favourable symmetrical (inactive) conformation [141]. Fitting of the insulin receptor model to experimental data, gave K_d values of 6.4 nM for site 1*, and 400 nM for site 2, and 0.19 nM for the high affinity cross-link of sites 1* and 2 (or 1 & 2*); this in good agreement with experiment [31, 141].

Several other more or less varying models of IR activation mechanism has been hypothesized and has been described elsewhere [14, 31, 32, 70, 142-144].

Previous models apparently have not discussed the possibility, that multiple insulins can bind simultaneously at different parts to the receptor and its relation to internalization.

However, very recently published as a prewrite by Gutmann *et al.* [145] of a cryo-EM IR ectodomain structure (very insulin saturated i.e. 50 μM); whereby two insulins are found bound at sites 1*&2 and 1&2*; however, two additional insulin binds at new binding regions. Their findings and discussion do appear to redefine the notion of a primary and secondary region crosslink. That their site "1", "1"" are the site 1*&2, 1&2* cross-link respectively; and the new binding cross-links, site "2" are residues in L1*, F1, in addition to site "2"" at residues in L1, F1*. The structure and discussion Gutmann *et al.* may rationalize and redefine to some extent the harmonic oscillator model and previous understanding.

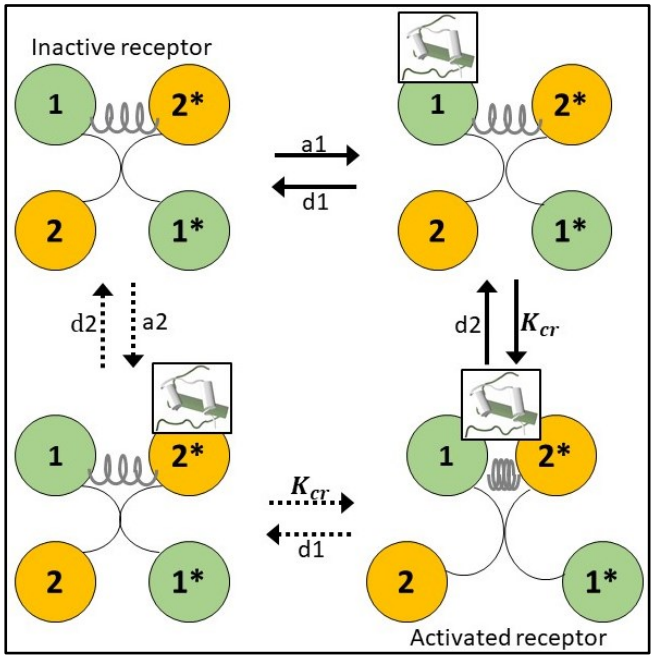


***Figure 2.6**: Simplified schematic of the IR harmonic oscillator model . Site 1 for primary and 2 for secondary binding site. Identical sites are indicated with a 1*, 2* respectively. Association constants a1 and a2 for binding sites 1, 2 respectively. Dissociation constants d1 and d2 for binding sites 1, 2 respectively. With K_{cr} being the cross-linking constant. Major pathway full arrows, supplementary pathway dashed arrows. (Schematic based on figure 1 of Knudsen et al. [125], figure 6 of Kiselyov et al. [141]. See these references for elaborate details).*

"A weight on the seesaw lever that causes the rock to fall down the mountain"

Furthermore, it is believed that the nestled receptor α-subunits functions as allosteric enzymes, which inhibits the kinase activity of the β-subunits. Since the cleavage of the α-subunits results in activated kinase in a similar fashion as insulin [124, 146], it suggests that insulin provides the binding mechanism, that have the analogy *"a weight on the seesaw lever that causes the rock to fall down the mountain".* For example indicated by the tryptic activation by cleaving peptide bond in an α-subunits (implied to be R^{576}-R^{577} in the F1 domains), correlating in an identical manner with insulin, activating the intracellular kinases, suggesting a releasing conformation change in α-subunits, that were transmitted through each TM domain [147].

2.5.2.2 A conjectured rationale for two (or four) ligand activated IR being needed for full auto-phosphorylation and internalization

So far no found literature has exclaimed the possibility that two (or 4) insulins can be bound at the same time, with equal binding affinity, and further connected it to its biological meaning and significance. Here explaining a model of two insulin activation (and if 4 insulins are required for full activation which is not certain).

Are the internalizing recycling of IR being initiated when 2 insulin ligands have bound to respective binding region cross-link, site 1*&2 and 1&2* (or 4 insulins bound if site "2", "2'" are necessary as defined by Gutmann *et al.* [145]). This idea seems plausible according to an old study by Terris *et al.* [95, 148, 149], who observed that from prebound ^{125}I-insulin, a linear relationship of extra insulin binding to the internalisation of receptors (slower without prebound ^{125}I-insulin), having a dissociation constant of 3.5 nM.

The above dissociation constant appears to bear remarkable congruence with that of the measured dissociation constant of 3.5 nM for soluble IR ectodomains, for that of two insulins binding with equal affinity, to the corresponding site 1*&2 and site 1&2* cross-links [150]. Similar values for the soluble ectodomain dissociation constant, with a linear scatchard plot, has been measured and indicated [151, 152]. Hence, I conjecture that, *in vivo*, it suggests a fast dissociation when two insulins have bound, that can be replaced by other insulins with a dissociation constant of ~3.5 nM; plausibly this should be verified by further sources until confirmed. Further speculating, that *in vivo,* the weakened binding of the two insulins might play a role, when the receptor are internalized in the cytoplasm and

recycled in the endosome with insulins broken down in the liposome [97, 153].

Furthermore, as I infer, there is a strong indication of a vital biological feature that 1 (or 2) bound insulin induces only part of full auto-phosphorylation of the internal IR domains, that 2 (or 4) or more insulin binding are required for full. This idea is strengthened by that auto-phosphorylation are strongly indicated to increase steadily; either with time (for insulin 1 uM) [136, 154]; or with increasing insulin concentration (sub nM to uM), even shown in some studies an inhibition at above 100 nM (as seen also for accelerated dissociation) [126, 135, 155-158]. In addition, even a noticed increased internalization (10^{-11} to 0.5×10^{-6}), which looks like an inverted sigmoidal curve to that of accelerated dissociation of prebound insulin [2, 126]. An increase in activating certain substrates in distinct pathways are also more or less in proportion to insulin concentration [126, 159].

A questionable order of auto-phosphorylated tyrosines upon 1 to 2 (or 2 to 4) insulin binding, considering those in both receptor β-subunits, are in the following order; Y^{1150}, Y^{1146} , Y^{1151} (in the activation loop near the TK enzymatic site) and then for Y^{1316} and Y^{1322} (in the CE domain), and less certain the order of region Y^{953} and/or Y^{960} (in the JM) [116, 136, 138, 160-162]. For these residues, it has been measured part full auto-phosphorylation at low insulin concentration (10 nM) and more or less double for higher (100 nM) [126, 127], which may be indicating a conformational change to a symmetrical IR ($\alpha_2\beta_2$) upon a second insulin binding. Further it may suggest that 2 (or 4) insulin bound are necessary to reach full intracellular auto-phosphorylation, and a regulating feature of activating signalling pathways, moreover internalising recycling.

A nestled symmetric TK dimer unit was asserted to be a plausible model for the two fully activated catalytic sites (having Y^{1150}, Y^{1146} , Y^{1151} in both subunits phosphorylated) [163], which may be indicative to that 2 (or 4) insulins are subsequently bound in order to reach such a symmetrical conformation.

That the TMs are apart in the apo-IR, and upon 1 to 2 (or 2 to 4) insulin binding acquiring a closer association (and in effect the fibronectin domain legs), indeed appears believable [163-165].

The initial apo-IR relation between the two TK domains of each receptor subunit, are however not clear. Where the putatively TM helical segments are separated, albeit conceivably inclined towards each other [166], the intracellular β-units may be inhibited by this separation. However that the TKs are initially in a dimer inhibition-state has also been suggested as plausible [138, 163]. Notwithstanding, it is believed that the TK catalytic

site activation entails releasing *cis*-inhibition, catalytic site activation-loop 3 tyrosine auto-phosphorylation, and allosteric dimer formation [163]. That before ligand binding each sub-unit is believed to have a *cis*-inhibiting conformation involving a JM critical residue Y^{972} (IR-A numbering) [158] and Y^{1150} in the kinase active site [138, 143].

The TK domains have also been postulated to respectively be unactivated and near inverted (i.e. NT/CT with respect to the IR sequence direction across membrane), and upon ligand binding released in a near "yoyo" like fashion [32]. The idea of inverted TKs is partly based upon a finding, that a CE conformational change takes place upon ligand binding, leading to a shortlived state that can bind ATP, and upon phosphorylation driving another distinct conformational change involving regions in the JM, TK and CE domains [167, 168]. Where the JM and CE domains are putatively unstructured polypeptide segments, except for possible β-turns in the JM region near Y^{953} and Y^{960} [138, 169].

The kinase catalytic active site is located in between the junction of the NT and CT TK lobes, which can bind an ATP as a substrate and facilitate phosphorylation on tyrosine residues. Putatively necessary are the auto-phosphorylation of Y^{1146}, Y^{1150} and Y^{1151} for activating the kinase towards exogenous substrates, even going from the double to triple auto-phosphorylated form may also play a role in regulation [136, 160, 170-172]. A crucial side-chain of the active site is a lysine, K^{1018}, since its mutation to alanine renders the active site inactive [173]. Putatively the TK enzymatic sites act primarily via inter-subunit (*trans*) tyrosine auto-phosphorylation, rather than of the same subunit (*cis*), appearing to even concern at least some of the 3 tyrosines of the activation loop [138, 160, 174].

Evidently, an inactivating mutation ($K^{1018} \rightarrow A$) in one or both TK domains, showed that *trans* TK auto-phosphorylation of the respective β-subunit is vital, albeit a minor (3 fold) *cis* auto-phosphorylation were found (with 1 of 2 TK active site inactivated). Further concluded, were that both TK active sites needs to be functional, in order to activate the TK catalytic sites towards exogenous substrate [175-178].

A conjecture of the broadview structural model of the entire receptor transcending from 0 to 1 to 2 (or 2 to 4) insulin ligands fully bound

Here is further conjectured a novel broadview model, of how insulin may induce large domain movements in the IR (illustrated in Figure 2.7); as was partly discussed above and more elaborated here.

To visualize the IR domain movements, one may further consider that the following

structures are at least coarsely solved: an 0 insulin bound apo-IR ectodomain [129]; an 1 insulin high affinity bound ectodomain [2]; an 2 insulin weaker affinity bound soluble ectodomain [70]; and a saturated 4 insulin bound ectodomain [145]. In addition, that the TM structure are indicatively known [166] and that the JM and CE segment may be coarsely modelled. In addition, that the TK domains have been indicated of being functional dimers in the activated state [163, 179, 180]. In addition to other literature mentioned here.

Apo-IR upon insulin binding

As I elaborate from the apo-IR by Croll *et al.* [129] (schematic ectodomain in Figure 2.7a), there are two large cavities that can accommodate an insulin binding. Considering the first binding region cavity would have insulin putatively dock at site 1* (L1*, αCT), and site 2 residues near the F1-F2 junction [120]. May be some site 2 residues are at the C*-L2*, or even in the IDα, since it looks like there is enough room for insulin to enter from two directions in respective binding regions "cavity hole".
Furthermore, the αCT-helix are closely bound to L1* and the junction in F1-F2, and further from C*-L2*. Moreover stabilizing interactions are between L1* to C* and F2-F3 (in addition to the IDs) domains respectively [164]. At least these interactions are apparently more or less disrupted by 1 (or 2) insulins binding. The favourable interactions (of e.g. L1* with F2-F3) that the binding of 1 (or 2) insulin perturbs, speculatively relaxes the entire β*-subunit, causing the release of inhibitory mechanisms, and initiation of at least part of full auto-phosphorylation. Conjectured here, is that the first insulin binding to one binding region, e.g. site 1*&2 (and maybe additional insulin to e.g. site "2"), causes either or both sub-units ($\alpha\beta$ or $\alpha\beta$*) to more or less relax, leading to the unleashing of the *trans* auto-phosphorylation kinase activity; altogether causing a change in conformation that perturbs the alternate binding site 1&2* (causing negative cooperativity).

Holo-IR with 1 insulin bound upon 2 (or 4) insulin binding

How does the 1 insulin high affinity bound IR look like in its entirety in a biological cell? Plausibly there must be (at least in part) some correspondence to the proposed IRs "signalling conformation", IRΔβ-zipInsFv, by Weis *et al.* [2], schematically corresponding to Figure 2.7b3 ectodomain (which are here provided possible other alternative for). For which insulin is bound to L1*, αCT, F1 and close to C*-L2* (and further from L2), with

the αCT closely bound to L1*, C*, L2* and F1 and also somewhat close to F1* and C-L2. The pre-tethered F3 to F3* domains have ~ 25 Å center-of-mass distance (c.f ~ 100 Å center-of-mass distance in apo-IR). In addition, one ectodomain receptor subunit L1*-C*-L2*-F1*-F2*-F3* being slightly straightened, notwithstanding leaving the other binding region perturbed (but largely intact). Interestingly, this perturbed but preserved secondary binding region (including L1, αCT*, F1*-F2* junction), plausibly indicates why a second insulin would bind at a slower rate (negative cooperativity) [2].

Since the unbound form, IRΔβ-zip, has native IR curvilinear Scatchard plots [2], it appears indicative of negative cooperativity and accelerated dissociation. Hence, I posit that 1 (or 2) more insulin can bind IRΔβ-zipInsFv and break the binding affinity of the L1 domain to the stabilizing contacts (including to F1*-F2*); relaxing the second $\alpha\beta$-subunit, whilst the other receptor subunit (L1*-C*-L2*-F1*-F2*-F3*) stays straightened. This 2 (or 4) insulin bound IRΔβ-zip, I'd surmise would have a similar schematic shape as the ectodomain in Figure 2.7c. As a deduction then, why the IRΔβ-zipInsFv has only one bound insulin, would be due to that is from incubated low insulin concentrations (~100 pM?): since moreover Weis *et al.* assumed that the 83-Fv antibodies had negligible effect on structure.

Furthermore, since the F3 domain tethered fibronectin "legs" of IRΔβ-zipInsFv, appears symmetrical, I posit are in an alike conformation, when 2 (or 4) insulins are bound with equal but lesser affinity to site 1*&2 and 1&2*. Further, I presuppose, that the "legs" already may be in a similar conformation as the, *in vivo,* fully phosphorylated IR endodomain, with the TM and intra-domains of both receptor subunits nestled and fully *trans* auto-phosphorylated.

However, I posit that the 1 (or 2) insulin bound IR, more likely have only one of its two receptors subunits relaxed, i.e. like in either of the schematics in Figure 2.7b12; whereby at least one of these schematics may represent the "/Γ"-shape (or "II") of Gutmann *et al.* [165]. Their "/Γ"-shape of ectodomain, appears to have the fibronectin legs still separated (but close), if this peradventure represents a 1 (or 2) insulin induced relaxation of 1 receptor subunit (possibly with 2 insulins simultaneously bound to the other site 1&2*). Whose relaxed subunit have at least partly *trans* auto-phosphorylated and nestled itself to the other subunit (explaining the intracellular dense region seen in their "II" conformation).

Interestingly, seen in their "T"-shape conformation [165], are the F1*-3*, F1-3 and TM* and TM appearing in close proximity. Moreover a strong "intracellular" density which were presumed to be interacting β-subunits, this density may be the fully *trans* auto-

phosphorylated activated TK dimer, that are possibly structured as indicated by Cabail *et al.* [163]. In support of the idea "T"-shape, are the very ligand saturated "T"-shaped ectodomain, with two cross-linked bound insulins, in addition to two insulins bound at top of "stalk" region (which may be stabilizing the "T"-shape) [145].

Here it is then conjectured, that the fully intra IR auto-phosphorylated signalling conformation may be for 2 (or maybe even 4) bound insulins (schematic in Figure 2.7c). Hence the schematic in Figure 2.7b3, I suppose is only the "signalling conformation" of the insulin high affinity bound region, but maybe not the ectodomain or IR as a whole.

Summary of the conjecture of apo- to holo-IR upon insulins binding

Since the accelerated dissociation of insulin follows a bell-shape, it suggests that at low concentrations (sub nM) of insulin only 1 insulin occupies the site 1*&2 cross-link (or if 2 insulins also to e.g. site "2" crosslink) will be occupied. Higher concentrations (supra nM) adds the likelihood that the perturbed alternate site 1&2* will also form a cross-link (or if additional insulin cross-linked to e.g. site "2'"), reaching the "T"-shape ectodomain, and resulting in the weaker binding affinity of the site 1*&2 and 1&2* cross-links, increasing the chance of each to dissociate and reassociate (and maybe insulins dissociating and reassociating to site "2", "2'").

Moreover, initial supra-physiological concentration of (above 100nM) is indicative of impeding the dissociation of the firstly bound insulin e.g. at site 1*&2 (maybe for insulin at e.g. site "2" also). This may be due to that with the increased concentration there is a larger likelihood that the second binding region will be occupied by two insulins binding too either site 1 and 2*, impeding this second cross-link significantly [141, 181]. However, given that one of two insulin bound at site 1 or 2* may be more amenable to dissociate, would leave the other to acquire the site 1&2* cross-link.

Speculatively then a second insulin forming a cross-link to sites 1&2* (or also to e.g. site "2'") would relax the second receptor subunit and fully activate the kinase sites, driving a full concurrent *trans* auto-phosphorylation with concurrent intertwisting of intracellular subunit domains.

Moreover since that Gutmann *et al.* [145] has found two new binding regions of insulin (however very ligand saturated) site "2" and site "2'" that respectively appears to stabilize the site *1&2 or 1&2* cross-links and maybe even impeding their dissociation, perhaps also plays some role in stabilizing the final "T"-shape of receptor. Hence Gutmann *et al.*

results and discussion, may indicate, that actually the order of insulin binding for full auto-phosphorylation are; first insulin binding site “2”; second insulin to site 1*&2 cross-link; third to the other site “2’”; fourth insulin to site 1&2* cross-link; or any other order or combination.

As the previous subsection indicated in literature, I posit that the 2 (or even 4) insulin IR binding, may signal the cell to internalize the receptor with 1 to 2 (or 3 to 4) still bound insulins, if not dissociating in the process. However, that at least a part initiation step, may actually be, that both receptor subunits intra-domains are fully auto-phosphorylated.

Closing statement

The presented simplified conjecture appears to an extent truthful, based on the literature at hand, though need further elaboration and verification, by biological and structural studies. Furthermore, the accompanying cytoplasmic signalling cascade pathways, with inherent IR substrates attaching to the activated IR intra-domains, followed by downstream linking substrates with further regulation, are also an important aspect, that may not be completely considered here [25, 86, 182]. The discussion here largely pertains to the IR-A isoform, but may hold some truth to various hybrid IR’s with various ligand binding processes, that is finetuned and specified to individual celltypes, development stage and vertebrate species [116, 118]. Since, for example, even different ligands binding to receptor IR-A, such as insulin and IGF-II, whose binding affinity and specific molecular contacts, may induce varying auto-phosphorylation behaviour and resulting biological pathways [126, 127].

The insulin binding to its cognate receptor hence is a very complex machinery, but the conjectured simplification here may be a more or less correct model, at least for the IR-A and IR-B isoforms.

Figure 2.7*: A conjectured model of the insulin receptor upon binding 1 to 2 insulins.* ***(a)*** *The apo-IR with two floating insulins in "black-frame", which are illustrated with a conjectured trans-dimer inhibition and cis-inhibition partly via inverted CT/NT TKs.* ***(b1,2,3)*** *The conjectured alternatives for binding of 1 insulin to site 1*&2 relaxing either α-β, α*-β* or both subunits and unwinding either the NT/CT TK, TK* or both, causing a part or full trans auto-phosphorylation of the tyrosines in the activation loop, CE and JM domains.* ***(c)*** *A conjectured depiction of the fully auto-phosphorylated IR. Where full trans auto-phosphorylation of the tyrosines in the activation loops, CE and JM domains in both subunits have occurred and have driven the nestling of JM, TK domains with concurrent approaching of the TM and F1-3 domains of both subunits.*

Note that the figure, doesn't explicitly depict the other optional pathway, that two additional insulins may bind to site "2" i.e. L1, F1 and another at site "2'" i.e. L1, F1*, beneath the "head" region's two insulins bound at site 1*&2, site 1&2* respectively. That e.g. 2 insulins binding to site "2" and site 1*&2 may facilitate part auto-phosphorylation and then 2 extra insulins binding to the second site "2'" and site 1&2* are required for full auto-phosphorylation. See text for further elaboration. Also note that the conjectured model, is not a main hypothesis of thesis, however included for value of idea, in this field of research.*

2.5.3 Residues putatively involved in the mechanism of insulin binding to receptor

2.5.3.1 Putative primary and secondary site insulin residues

Some suggested or putative residues involved in binding to primary site (1* or 1) binding are: G^{A1}, I^{A2}, V^{A3}, E^{A4}, Q^{A5}, Y^{A19}, N^{A21}, G^{B8}, S^{B9}, L^{B11}, V^{B12}, Y^{B16}, G^{B23}, F^{B24}, F^{B25} and Y^{B26}. Some suggested or putative ones in secondary site (2 or 2*) binding are: T^{A8}, I^{A10}, S^{A12}, L^{A13}, E^{A17}, H^{B10}, E^{B13} and L^{B17} [4, 31, 32, 142, 183]. This is stated and emphasized differently in various literature, thus specific residues are more or less certain. The insulin monomer residues involved in primary site binding, are to some extent congruent with a consensus of "classical binding residues": A1, A5, A19, A21, B12, B16, B23, B24, B25 and B26 [4, 18, 140, 184-186]. Which resulted early on from knowledge of insulin structure, and studies of biological activities and sequences from different animals [184]. The B-chain residues overlaps the ones involved in dimerization. However, there is apparently greater affinity for receptor binding, $K_d \sim 0.2$ nM (supposed main contribution by primary site), than for affinity in dimerization, $K_d \sim 7\mu$M, [141]. By an alike observation additional residues was early on proposed to be involved in receptor binding [184].

The residues of the insulin monomer involved in secondary site binding, are to an extent consistent with a so called novel binding surface [4, 125]. Which are involving mutations of L^{A13} and L^{B17} in the hexamer binding surface, proposed [30, 187] on the grounds of studies of insulin analogues with abnormal binding properties, similar to those of hagfish [188, 189] and hystricomorph insulins [190, 191]. In addition, secondary site residues: A8, A10, A12, A13, A17, B10, B13 and B17, has been indicated from mutagenesis studies [31, 192], and overlaps the hexamer surface.

2.5.3.2 Alanine mutagenesis studies

Substitutions or deletions of specific residues in native insulin, more or less changes the structure and/or receptor binding affinity of the monomer. Which is not surprising giving the dimensions of the insulin monomer.

An alanine scanning mutagenesis study were performed by Kristensen *et al.* [105], wherein they measured the receptor binding affinity of various analogues having residues mutated by alanine. They reasoned that a disruption of affinity was due either, to that the residue substituted interacted directly with receptor, or otherwise supported a conformation

required for binding to the insulin receptor. Analogues with alanine substituted at I^{A2}, V^{A3}, Y^{A19}, G^{B23} and F^{B24}, had less than 5% of native affinity. Reasoning that these residues are an essential receptor binding patch, being fully exposed when the B-chain CT strand are away from core and enabling direct interaction with receptor. Albeit, they reasoned also that structural perturbations of the monomer, could have been part of this reduction of affinity. Analogues having alanine substitutions of E^{A4}, N^{A21}, Y^{B16} and Y^{B26}, had 36-139% (omitting error) altered affinity for the receptor; they reasoned that these residues as not being main residues in functional binding. Analogues with alanine substitutions of residues S^{A9}, G^{B20} and R^{B22}, resulted in a 260-405% (omitting large error) increase in receptor affinity. With the expression yield of analogues having substituted G^{B20} or R^{B22} were very low and was thought to indicate structural consequences. Moreover, the substitution, $G^{B8} \rightarrow A^{B8}$, also had less than 5% of native affinity, thought then to perhaps play a structural role, since it is an initial residue to the α-helix of B-chain. Alanine substitutions at NT B1-B4, had about 54-134% (omitting error) native affinity, whereas at B5 it was reported to be 31% [105]. Which can be compared with the substitution, $L^{B6} \rightarrow A^{B6}$, having 1.4% receptor affinity [193], or an even further decrease, about 0.052 %, for substitution $L^{B6} \rightarrow G^{B6}$, indicating an important structural role for the leucine SC [193]. Moreover, an analogue having removed the B1-B6 residues, had about 0.041% of receptor binding potency [193], similar to the analogue having a glycine substituted at B6. The relatively smaller decreases in receptor binding affinity, are of removal of B1 to B5 [193]. Though the analogue with B1-B5 removed, has been reported to have markedly more decrease in receptor binding, than the analogue with B1-B4 removed [1, 194].

Studies by De Meyts *et al.* concluded that residues A21 and B23 to B26 are essential for negative cooperativity in hormone receptor binding, whereof substitutions or deletions at B23-B26, or deletions at A21 (substitutions tolerated), resulted in a loss of the analogue to accelerate the dissociation of prebound ^{125}I-native insulin [4, 30, 140]. In comparison, analogues with alanine substitution at B23, B24 and B25 respectively were reported as 3-10%, of B26 as about 36%, of A21 as about 66%, receptor affinity [105].

2.5.3.3 Amino-acid mutagenesis of the B-chain α-Helix

The B-chain central αH (B9-B19) is thought to function as a central recognition element, in receptor binding. Since its residues are flanked, by the putative receptor binding surfaces

of the insulin monomer. Putatively binding to the insulin receptor at the L1* β-sheet at site 1*, conjointly having residues binding to site 2. Thus, the B-chain central α-helix is appearing to be a binding motif, serving as a linker of continual binding surface [185, 195]. An amino-acid mutagenesis investigation by Glendorf *et al.* [185], showed the relevance of the B-chain solvent exposed residues, for receptor binding. They asserted that residues V^{B12}, Y^{B16} and E^{B13}, are indeed part of a binding surface of insulin. Since, Y^{B16} being situated in the proximity of V^{B12} in the dimer-forming surface, they were also early on proposed to be part of the 'classical binding surface' [184]. Furthermore, cross-linking studies has inferred that Y^{B16} maps to the L1 domain of the IR [195]. Moreover, structure activity relationships has inferred that V^{B12} interacts directly with the IR, as a "high affinity" contact [195]. Even though earlier reports, had inferred Y^{B16} to not be part of binding residues, based on not being "evolutionary conserved" [67], moreover that the substitution $Y^{B16} \rightarrow A^{B16}$, had ~69% of receptor affinity [105]. However, the later investigations [185, 195], reported alanine substitution of Y^{B16}, having lower receptor affinity ($\sim 27-34\%$). Moreover, only phenylalanine and tryptophan substitutions at B16, maintained almost native affinity [185], suggesting hydrophobic or aromatic interactions of Y^{B16} at receptor binding. The solvent exposed residue H^{B10}, has been putatively implied to interact with site 2 of the IR [31, 32]. Albeit, Glendorf *et al.* [185], stated that H^{B10} are not required for IR binding, due to that many of the other standard amino-acids, having similar or greater IR affinity. They stated though that the nearby residue E^{B13}, to be plausibly involved in receptor binding at secondary site, were only substitution to asparagine or tryptophan maintained close to native affinity.

2.5.3.4 The detachment model

The extended conformation of the B-chain CT strand facilitates the stabilizing interface of an insulin dimer, however, for a solvated insulin monomer, this strand can move more freely. A "detachment model" envisaged, that upon binding its receptor insulin undergoes a structural transition, in which the end of the B-chain disengages from the hormone, exposing an otherwise hidden binding surface [20, 69]. This exposed surface of the monomer was inferred to form a primary binding interface, with the insulin receptor, mainly through the αCT segment [32, 71, 131]. Putatively also the residues G^{A1}, I^{A2} and V^{A3} (part of hydrophobic core) becomes exposed and directly contact the IR, after this detachment

[20, 64, 66, 69]. Congruent with studies of insulin analogues, having connecting segments between NT A-chain and CT B-chain of less than 3 residues, which prevented the induced fit and were essentially without biological activity [20, 196, 197]. In contrast, the des-pentapeptide analogue (insulin without B26-B30), were reported with no decrease in receptor affinity, compared to native insulin [65].

2.5.3.5 Structures of insulin bound to site 1 and 2 in receptor fragments

Supporting the "detachment model" some researchers [69, 71, 131] reported micro-receptor (μIR) structures. With the newest structure of μIR (Figure 2.8a) as composed of insulin bounded with a fragmented complex of in part α-subunit domains L1* and αCT (or likewise L1 and αCT*). This foregoing structure reveals features of the site 1* receptor binding surface, depicting some overlap with insulin residues of the dimerization surface. A difference from the monomers in the dimer surface is that insulin has by its B-chain CT strand, its hinge-like B20-B23 β-turn rotated, moreover an outward rotation of B24-B27, locating itself between strands of the L1* β-sheet and αCT residues 714-718. Side-chains of F^{B24} and Y^{B26} are mainly towards L1* and αCT, while those of F^{B25} are directed towards αCT, congruent with previous findings of F^{B24} cross-linking to L1* and F^{B25} to αCT [198, 199]. Furthermore, the dissociation constant for the μIR were reported as, K_d~7.5 nM [69], close to the site 1* reported value of, K_d ~6.4 nM [141], this observation further indicates that this structure features a resemblance to binding site 1*.

Recently Scapin *et al.* [70], obtained structures of insulin bound to soluble IR (sIR) structures, wherein a similar view is depicted for the lower resolution (hence with a larger error range) structures of either 2 or 1 insulin bound on a soluble ectodomain receptor, denoted as "sIR+2" and "sIR+1" respectively. Notwithstanding, the interaction between insulin, αCT and L1*, have a resembling binding surface [69, 70, 142], although differing in overall structure (orientations of MC, SC etc). Hence these structures at least support the location of a site 1* binding region. However, there are suggesting contacts of A7 and range B4-B10 with F1 residues, depicting these as plausible site 2 residues. For a sIR which binds two insulins, each indicated of having a proximate dissociation constant, K_d~3.5 nM, and a fast dissociation rate [70, 150, 152]. Discordantly, since the high affinity cross-link has a higher binding affinity (K_d ~0.19 nM) for 1 insulin monomer bound per receptor dimer (1:2 stoichiometry). Hence it appears that the "sIR+2" structure doesn't depict a high affinity cross-linked state fully, however speculatively at least coarsely the double-bound

weaker affinity of the conjectured double insulin bound IR (Figure 2.7c).

Even more recently a publication by Weis *et al.* [2], revealed a cryo-EM structure of insulin bound in a high affinity cross-link, where the F3 to F3* tethered (or zippered) ectodomain receptor (IRΔβ-zipInsFv). Revealing more or less the same location of site 1*&2 as the above-mentioned structures, albeit with differences in resolution, and also of contacts and orientation of specific residues. Also remarkable is that competition binding curves of IRΔβ-zipFv, has congruent high affinity binding and characteristics to that of a holo-receptor (hIR) [2, 200]. However, the sIR can bind 2 monomers with equal affinity, and the hIR at least initially binds only 1 monomer with orders of magnitude higher affinity, additionally the sIR displays no negative cooperativity, whereas the hIR does [139, 140, 187]. The source of the sIRs lower binding affinity is not certain, it were speculated to be related to the relatively larger separation of the apo-sIRs F2 domains where the αCT and disulphide bridged $ID\alpha$s are correlated [2], however I conjecture here this may be due to that two insulins have bound in "sIR+2".

The residues vital for negative cooperativity, A21 and B23-B26 indeed each appears to have relevance in locating insulin with L1* and αCT to form the site 1* binding. As inferred by Weis *et al.*, IRΔβ-zipInsFv informs the explanation of negative cooperativity, by the destabilization of the second binding site, via disulphide coupling of αCTs, such that in theory, excess insulin could bind to this site and cause the accelerated dissociation of the first bound insulin. Nevertheless it is believable that at least the contiguous structure of insulin, in IRΔβ-zipInsFv, reveals at least an approximate depiction of the high affinity bound insulin [2]. Appearing as the most meaningful structure in this regard, we depict more clearly the contacts as in Figure 2.2 and here in Figure 2.8d, showing the residues within 5 Å of the bound insulin. The other putative site 2 residues of insulin (A8, A10, A12, A13, A17, B17) are more or less not in direct contact, as speculated also by Weis *et al.,* at least some of them may play a role in the mechanism of reaching or upholding the high affinity cross-link.

A confirming picture of the location of primary & secondary binding sites (1*&2 and 1&2*) are from an even more recent prewrite of Gutmann *et al.* [145], of a very insulin saturated IR. That appears to rationalize the insulin residues (A8, A10, A12, A13, A17, B17) as involved in IR activation, where two insulins are binding to other cross-links at regions site "2" (L1*, F1) and "2'" (L1, F1*).

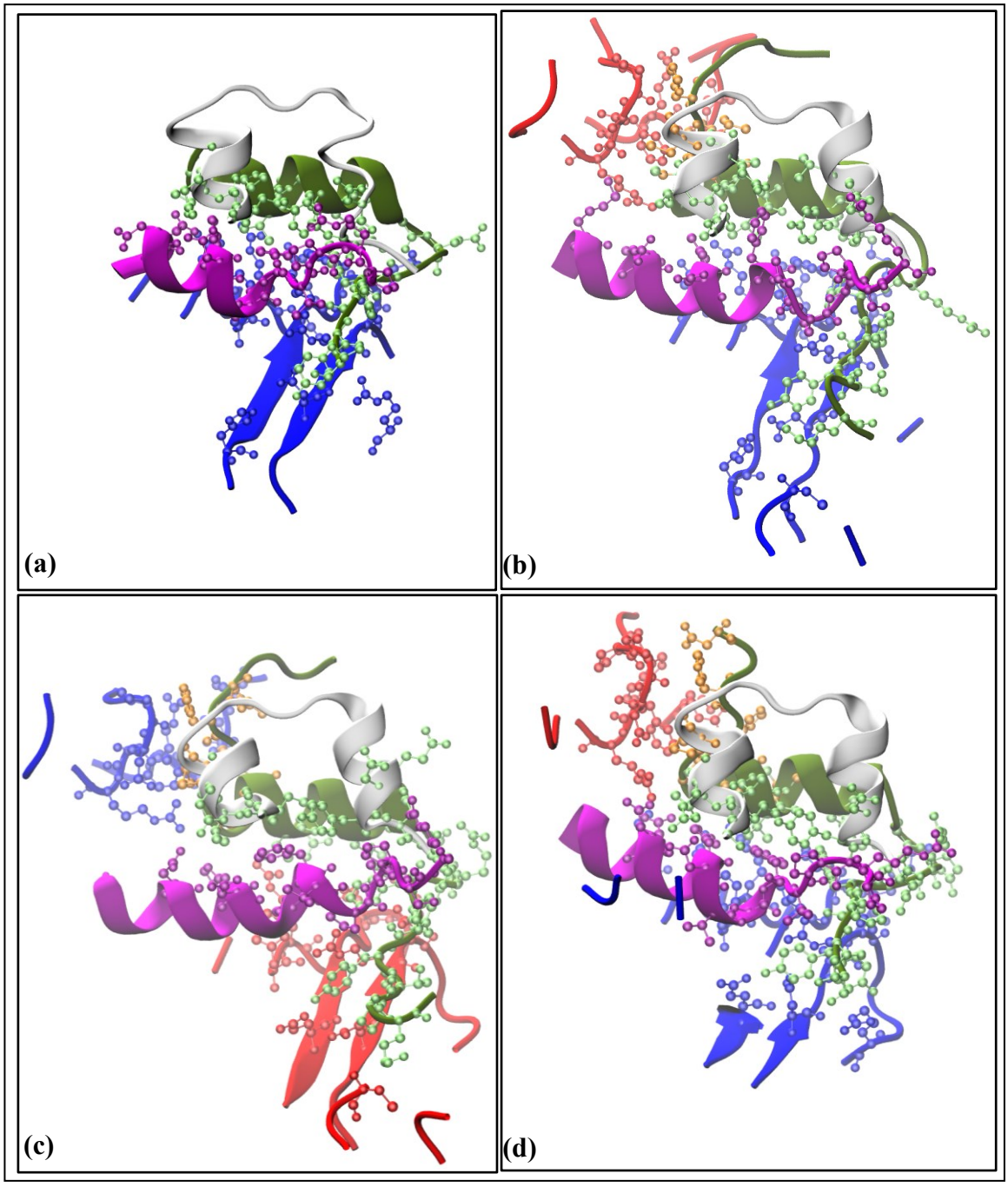


Figure 2.8**: Diverse structures of insulin bound to receptor fragments The figures are depicting more or less accurate approximations of the structural relationships of insulin with that of its site 1 and 2 residues.* ***(a) *μIR, PDB 4OGA, 3.5 Å resolution [69], (B1-B6 and B28-B30 absent).* ***(b)*** *"sIR+2", PDB 6CE9, 4.3 Å resolution [70].* ***(c)*** *"sIR+1", PDB 6CE7, 7.4 Å resolution [70].* ***(d)*** *IRΔβ-zipInsFv, PDB 6HN5, 3.2 Å resolution [2], (B1-B2 and B28-B30 absent). For each respective structure depicting the receptor αCT helix (purple same receptor monomer as red) and F1 domain (red) of one monomer, and the L1* (blue) domain of the alternate receptor monomer. However, for "sIR+1" the opposite colouring is chosen arbitrarily, due to its naming convention in its PDB. Respectively for each structure: the secondary cartoons are within 10 Å of insulin, and the ball-stick represented residues are those that have any non-hydrogen atom-moiety within 5 Å of insulin (for clarity only residues of insulin contacting to site 1 as light-green and to site 2 in orange). See text for further discussion.*

Chapter 3 Method for Simulation and Analysis

A description of how to Perform Molecular Dynamics and how to obtain an Analytical Overview for Protein Systems

"I am enough of the artist to draw freely upon my imagination. Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world."

From interview to A. Einstein [201].

"… everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

The Feynman lectures on physics [202].

To simulate a solvated nanosized molecule with classical molecular mechanics for microseconds, is indeed an endeavour that modern science allows, which were not feasible at the time when Einstein and Feynman lived. Much knowledge has been gathered e.g. for the protein insulin as seen in the previous chapter. This chapter, however, will investigate further knowledge and use the mathematical imagination of the mind, to analyse insulin structure and dynamics in a comprehensible way. Explaining some of the underlying concepts and methodology for the subsequent result chapters, for understanding the analysis of a single structure or an ensemble of them. Furthermore, describing how to perform and analyse MD simulations and their resulting trajectories.

In particular regarding concentration and construing plausible ionization states for insulin in §3.1. Following is the background and methodology behind MD simulations, which are described in §3.2. Then following is a section on how the resulting MD trajectories were postprocessed, in addition to the various analysis methods, being described in §3.3.

3.1 Inclusion of molecules and choice of amino-acid ionization states for insulin

To perform MD simulations of insulin, one need to consider what concentration of water, ions or other solvent to include. In addition, if the concentration of the insulin monomer in a NMR experiment, will be directly comparable to an insulin monomer in a "box" of a MD simulation. Furthermore, for inferring plausible ionization states, a review is provided.

3.1.1 Concentrations of solvent

The following are describing the choice, for including a number of solvent molecules, in a MD simulation.

Inclusion of water

Approximately 4995 water molecules, of the Tip3p model, are included in a "box" of a MD simulation of the insulin monomer, having a density of $\sim 985 \frac{\mathrm{g}}{\mathrm{L}}$, which gives a molarity of $54.676 \frac{\mathrm{mol}}{\mathrm{L}}$. During subsequent equilibration this will be modified, and the effect of possible different density, we concluded to be negligible, using the expected molarity of $55.5 \frac{\mathrm{mol}}{\mathrm{L}}$ [25, 28], in following inclusion of molecules calculations.

The volume that are occupied of e.g. 4995 water molecules,

$\frac{4995\ \mathrm{H_2O}}{55.5 \frac{\mathrm{mol}}{\mathrm{L}}} = 1.4945 \times 10^{-22}\ \mathrm{L} = 149.45 \times 10^{-21}\ \mathrm{mL} = 149.45\ (\mathrm{nm})^3$;

$(1\ \mathrm{L} = 1(\mathrm{dm})^3 = 10^{24}(\mathrm{nm})^3 \rightarrow 1(\mathrm{nm})^3 = 10^{-21}\ \mathrm{mL})$.

Inclusion of ions mimicking that of human blood serum

The ion concentration of electrolytes in human serum is ~ 0.14 M for sodium and ~ 0.1 M for chloride ions, other ions of lower concentration not included in MD simulation [203, 204]. Hence one can determine the number of Na^+ and Cl^- ions to include in the water box $[4995\ H_2O \times (0.1\ \frac{mol}{L}) / (55.5\ \frac{mol}{L}) = 9\ Cl^-]$. For example, at human serum pH of 7.4 [25, 203], we assume that the insulin monomer, has a net charge of $-2e$. Hence a choice of $10\ \mathrm{Na^+}$ and $8\ \mathrm{Cl^-}$ will make the system neutral, maintaining an approximate salt concentration of 0.1 M. This assignment, however, may slightly underrepresent the sodium concentration (where the $\mathrm{Na^+}, \mathrm{Cl^-}$ ratio in human plasma may be 1.21:1 to 1.54:1 [203]).

3.1.2 Concentration of a molecule in a cubic volume

Here pointing to the likelihood of insulin to occupy any space, for e.g. in receptor binding experiments. This is not related to the use of periodic boundaries and use of long range cutoffs in MD, it is merely a concentration viewpoint of occupying a nm sized space. Furthermore in a NMR experiment, the insulin monomer concentration of about 0.5-1.5 mM may be used of a solution containing separated monomers [3, 20, 62, 63]. One can verify that it is likely separated monomers they are measuring, moreover that a monomer in box calculation will be a good comparison. For a solvent having an insulin monomer concentration of 0.5 mM, it is imaginable such a solvated system divided into cubes of a volume, V_{cube} of $(5.45\ \mathrm{nm})^3$, each being large enough to encompass 1 monomer. The expected occupancy of insulin monomers, in one cubic box, is then 4.87×10^{-2} (see below calculations). This has the meaning that it is quite unlikely to find an insulin monomer in any of those cubic volumes. For example, if occupancy equals 1, then we'd expect to find an insulin monomer in every volume, V_{cube}, filling the entire solvated space. And if occupancy were instead 2, we'd expect to find 2 insulin monomers in every cube, and so on. In the monomer MD simulations, we simulate one monomer in a box of volume, V_{cube} equal $(5.45\ \mathrm{nm})^3$, hence comparable to NMR experiments of distant monomers.

Shown by calculations:

$V_{cube} = (5.45\ \mathrm{nm})^3 = 161.88 \times 10^{-27}\mathrm{m}^3 = 161.88 \times 10^{-24}\mathrm{L}\,;\ (1\ \mathrm{L} = (0.1\ \mathrm{m})^3 = 10^{-3} \times \mathrm{m}^3)$.

To further obtain the occupancy, or number of expected monomers in a volume V_{cube};

$$0.5\ \mathrm{m}\frac{\mathrm{mol}}{\mathrm{L}} \times V_{cube} = 0.5 \times 10^{-3} \times \frac{6.0221420 \times 10^{23}\ \text{monomers}}{\mathrm{L}} \times 161.88 \times 10^{-24}\ \mathrm{L}$$
$$= 4.87 \times 10^{-2}\ \text{monomers}.$$

3.1.3 Choice of ionization states of residues in insulin

This section is a review, of ionization states of insulin in various environments, as reported in the cited literature. Here we also try to infer what constant ionization state, should be for each amino-acid for a solvated insulin monomer, when simulated over microseconds. This is a lesser way of modelling reality, since ionization states are fluctuating over time, with protons diffusing in water [47, 205]. Nonetheless, having a residue-wise constant ionisation state, may still serve as an average approximation for the solvated monomer.

3.1.3.1 General about proteins amino-acid ionization states

Any protein's native state and dynamics are thermodynamically determined by temperature, pressure and type of solvent. Furthermore, by the intra- and inter-molecular electromagnetical interactions, such as hydrogen bonding, hydro-phobicity/-philicity and van der Waals forces. Moreover, a proteins structure, stability, solubility and function, depends on its net charge and the ionization state of the individual residues [206, 207]. Variations of pH (proton concentration) in the solution, will alter the charge on residues having ionisable side chains, dependant on how much these residues are exposed to solvent. At a certain pH, the ionization state equilibrium for an ionizable group, is determined by its pK_a value; for example, with pH equalling pK_a, the ionized state equals the deionized state. However, the pK_a value are dependent upon temperature and the chemical environment, of the residues ionisable group; comprised of nearby residue-moieties and solvent [25, 206]. Hence in reality the biochemistry of a protein in solution is dynamic, thus their pK_a values will also depend on the momentous conformation of structure, which may not be fully modelled by MD.

3.1.3.2 Insulin's ionizable chemical groups

For the insulin monomer, there are seven amino-acid SCs, in addition to the NT amino and CT carboxy groups, which readily ionizes between pH 1 and 14 [25, 206]. For SCs of asparagine's, glutamine's, tyrosine's and cysteine's, their ionisable groups are uncharged below their pK_a, and negatively charged above their pK_a. For SCs of histidine's, lysine's, and arginine's, their ionisable groups are positively charged below their pK_a and uncharged above their pK_a. Hence, the native human insulin monomer contains a total of 16 ionisation transferable groups; 2 NT amino $C_\alpha NH2$ (Gly^{A1}, Phe^{B1}), 2 CT $C_\alpha COOH$ (Asn^{A21}, Thr^{B30}), 4 glutamate $C_\gamma COOH$ ($Glu^{A4,A17,B13,B21}$), 2 histidine imidazole ($His^{B5,B10}$), 4 tyrosine phenolic ($Tyr^{A14,A19,B16,B26}$), 1 lysine $C_\epsilon NH2$ (Lys^{B29}), and 1 arginine guanidium (Arg^{B22}).

3.1.3.3 Reported ionization states of insulin in various environments

Diverse pK_a values of the ionisable groups in insulin, has been reported in literature [207-210], here shown in Table 3.1. In addition it are assumed that these values differ for each report, due to varying methods of determination, solvent conditions and protein aggregation etc.

Firstly shown, are continuum electrostatic calculations [208], of the bovine insulin monomer (Table 3.1: col. A, B). Which have values differing to some extent, dependent on whether they were calculated for a closed or a more open conformation ($\mathrm{Gly}^{\mathrm{A1}}$, $\mathrm{Ala}^{\mathrm{B30}}$ not in close contact). Averaging over both conformations, $\mathrm{Asn}^{\mathrm{A21}}(\mathrm{C_\alpha COOH})$ and $\mathrm{Ala}^{\mathrm{B30}}(\mathrm{C_\alpha COOH})$ has a pKa of 2.325, and the glutamates $\mathrm{Glu}^{\mathrm{A4,A17,B13,B21}}(\mathrm{C_\gamma COOH})$ a pKa of 3.7. These average values are overall congruent for the open and closed conformation, indicating that these pK_a values of glutamic and C-terminal carboxy groups are separable. Excepting the discrepancy for the closed conformation for which $\mathrm{Glu}^{\mathrm{A4}}(\mathrm{C_\gamma COOH})$ has a pKa of 2.3 and $\mathrm{Ala}^{\mathrm{B30}}(\mathrm{C_\alpha COOH})$ has a pKa of 3.7. Furthermore, the N-terminal $\mathrm{C_\alpha NH2}$, have pKa values, that suggests those being protonated at physiological pH. Authors of this report [208], performed 95 ns molecular dynamics of insulin varying from pH 7-1, for four systems (O, H, HE, HEC), whereby when pH were lowered, the ionizable sites were successively protonated in the following sequence: ($5 < pH$); $\mathrm{His}^{\mathrm{B5,B10}}$ ($2 < pH < 5$); $\mathrm{Glu}^{\mathrm{A4,A17,B13,B21}}$ ($1 < pH < 2$); C-terminal carboxy groups of $\mathrm{Asn}^{\mathrm{A21}}$, $\mathrm{Ala}^{\mathrm{B30}}$($pH < 1$).

Next are the NMR low pH titration studies of an insulin analogue ($\mathrm{Ser}^{\mathrm{B9}} \rightarrow \mathrm{Asp}$) [210] (Table 3.1: col. C), of which were reported that the pKa values of glutamic and C-terminal COOH groups are not separable and that $\mathrm{His}^{\mathrm{B5,B10}}$ had a pKa close to 7. However, throughout this investigated pH range ($pH < 7$), they report that insulin existed more or less as a dimer. Albeit, at pH 7.5, this analogue was found to be mainly monomeric, indicating that $\mathrm{His}^{\mathrm{B5,B10}}$ being deprotonated at this pH. However, another NMR study [207] has reported the histidines ($\mathrm{His}^{\mathrm{B5,B10}}$) of insulin, to have a pKa close to 7 (not in table).

Next is another NMR pH titration study of an insulin analogue ($\mathrm{Glu}^{\mathrm{B13}} \rightarrow \mathrm{Gln}$) (Table 3.1: col. D), for which insulin are stated to be in different aggregation forms when varying pH [209]. Moreover, the titratable groups were reported in sets, such that individual residues can have ambiguous values. Here the COOH groups of glutamate's and C-terminals are ambiguous, and the $\mathrm{His}^{\mathrm{B5,B10}}$ has a pKa of 6.85. Their study also identified insulin dimerization as a factor inducing perturbations to pK_a values. Which also implies that the pKa values for a monomer, are likely different from higher oligomers.

***Table 3.1**: Sources of ionization states values of insulin and analogues (column A-D). In addition are shown average pKa values of proteins in general, (column E-F). Amino acids are colour coded by type; hydrophobic green, polar orange, acidic and C-terminal C^{α}-carboxy group red, basic and N-terminal C^{α}-amino group blue.*

*Col. **A**, **B**: pKa values of monomeric bovine insulin, calculated by continuum electrostatic calculations [208]; col. A closed as in crystal structure, col. B open conformation. Col. **C**: $Ser^{B9} \rightarrow Asp$ analogue. NMR low pH titration studies at 295 K [210]. Col. **D**: $Glu^{B13} \rightarrow Gln$ analogue. NMR pH titration studies at 296 K and 0.1 mM KCl [209]. Col. **E**: Typical pK_a values of ionisable groups in proteins from ref. [25], Col. **F**: Average of 541 tabulated pK_a values of 78 folded proteins under various conditions [206].*

Residue(group)	A	B	C	D	E	F
$Asn^{A21}(C_{\alpha}COOH)$	2.2	2.2	3.17	3.4, 4.66	3.1	3.3
$Thr^{B30}(C_{\alpha}COOH)$	3.7 (Ala)	1.2 (Ala)	2.38	3.4, 4.66	3.1	3.3
$Glu^{A4}(C_{\gamma}COOH)$	2.3	3.6	2.62	3.4, 4.66	4.1	4.2
$Glu^{A17}(C_{\gamma}COOH)$	3.9	4.3	> 3.7	3.4, 4.66	4.1	4.2
$Glu^{B13}(C_{\gamma}COOH)$	3.5	4.0	2.20		4.1	4.2
$Glu^{B21}(C_{\gamma}COOH)$	4.1	3.9	3.71	3.4, 4.66	4.1	4.2
$Asp^{B9}(C_{\beta}COOH)$			2.6		4.1	3.5
His^{B5}(imidazole)	6.1	6.4	6.915	6.85	6.0	6.6
His^{B10}(imidazole)	5.3	5.6	7.04	6.85	6.0	6.6
$Phe^{B1}(C_{\alpha}NH2)$	8.7	9.5		7.85, 9.33	8.0	7.7
$Gly^{A1}(C_{\alpha}NH2)$	16.9	9.3		7.85, 9.33	8.0	7.7
$Lys^{B29}(C_{\varepsilon}NH2)$	11.2	11.9		9.83	10.8	10.5
$Tyr^{B16}(C_{\zeta}OH)$	10.8	11.2		11.23	10.9	10.3
$Tyr^{A14}(C_{\zeta}OH)$	10.7	12.3		11.23	10.9	10.3
$Tyr^{A19}(C_{\zeta}OH)$	14.6	15.6		11.23	10.9	10.3
$Tyr^{B26}(C_{\zeta}OH)$	16.0	15.0		11.23	10.9	10.3
Arg^{B22}(guanidium)	12.9	11.1		11.23	12.5	

3.1.3.4 Presumed average ionization states of the insulin monomer in solution

The two sources giving pK_a values typical of amino-acids in proteins, agree quite well (Table 3.1: col. E, F), whereby glutamates, CT carboxy and histidines are separable, similar as for the insulin monomer in Table 3.1(col. A, B). Here it is made the following informed presumption, about the ionization state for an insulin monomer without oligomerization. That is to say, by varying pH from 7-1, a separation can be made to the systems into successive protonation at pH ranges: none ($5 < pH$); histidines ($2 < pH < 5$); glutamates ($1 < pH < 2$); carboxy groups ($pH < 1$).

A more or less coarse approximation may be seen from the Henderson-Hasselbalch formula [25], which gives the ratio of protonation of a chemical group:

$$\frac{[deprotonated\ amino\ acid]}{[protonated\ amino\ acid]} = \frac{[A^-]}{[HA]} = 10^{pH-pK_{aR}} \qquad (3.1)$$

Where a difference of 1 ($pH - pK_a = \pm 1$), yields a factor difference of 10, from protonated to deprotonated amino acid, whereby the following simplifying assumption are made; if the pK_a values are separated by at least 1, between amino-acid types histidines, glutamates and CT carboxy groups, one can infer that these amino-acid types are separable. Thus, for the MD simulation to be closer to physically meaningful, the amino-acids should be separated by type, with their pKa values differing at least by 1. Which we infer from the discussion above, appear largely plausible for histidines, glutamates and CT carboxy groups. Although because of the complex nature of the dynamics, and the apparent change of pKa values upon oligomerization, this is not a strong assumption. Hence the line of reasoning here, should be regarded just as an informed presumption, of the average pKa values of an insulin monomer in solution, at different pH. Since for MD simulations, as used in this thesis, over microseconds, it is restricted to use a constant ionization state, for the ionisable residues.

3.1.3.5 Standard ionization state of pH 7.4 for insulin systems studied

Since insulin monomers bound in hexamers or separate in solution or bound to IR-fragments are compared in this thesis, each having elaborate structures. Hence, the same protonation states, representing pH 7.4, are chosen, to allow a direct comparison of these structures. This is an oversimplification, that ought to be considered for, when regarding these systems. Howbeit, for this pH the following ionization states are chosen; the carboxy-groups deprotonated, glutamates being deprotonated, histidines deprotonated, tyrosines protonated, moreover the lysine and arginine group being protonated. This choice of protonation yields a net charge of the insulin monomer of -2.

3.2 Molecular dynamics simulations

Molecular dynamics (MD) or a.k.a. Molecular Mechanics (MM) is included in the term *Computational chemistry,* which is a term that covers a range of mathematical and computational techniques in chemistry. Computational chemistry encompasses quantum mechanics of smaller molecules, to classical mechanics of larger molecules or complexed aggregates. Including also semi-classical methods, e.g. modelling active sites of enzymes with quantum-mechanical approximations and the rest of the enzyme with classical mechanics.

Molecular dynamics which is a classical molecular mechanics technique, have the goal of

explaining realistic chemical systems with a simplified atomic model. Such a simplified model may enable understanding and prediction of atom-scale properties and in effect even macro-scale properties of a chemical system [211]. Biological applications of MD had its dawning in the 1970's [212-214], since then, MD have been used for simulating more or less complex biological systems, e.g. proteins, carbohydrates, lipids, nucleic acids, or a combination thereof [29, 51, 215-217].

Simulations by MD are developed in order to give an atomic to macro molecular picture, of time averaged observables from an experiment [48]. Where average values can be obtained from experiment, but we can't see the molecules giving rise to the measured quantities, i.e. models or simulations are needed to explain what's been measured. Thus, there is mutual interdependence of experiment and simulation. Since MD simulations are calibrated and tested against experimental data [218, 219]. On the other hand MD simulations are used for giving atomistic models, enabling explanation of experimental results [220].

The surrounding information about MD are covered in great detail elsewhere [211], and is outside of the scope here. However, its concept can be simplified as following [221, 222], for MD simulations, the classical or Newtonian equation of motion are solved, for a system of N interacting atoms:

$$m_i \frac{\partial^2 \boldsymbol{r}_i}{\partial t^2} = F_i, i = 1 \dots N. \tag{3.2}$$

The negative derivatives of a potential function $V(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N)$, gives the forces:

$$\mathrm{F_i} = -\frac{\partial \mathrm{V}}{\partial \mathrm{r_i}} \tag{3.3}$$

The potential includes many terms, relating e.g. to bonded interactions such as vibration and torsion, also for non-bonded electrostatic interactions. In small time-steps the equations of motion are solved simultaneously for each atom. The system is first followed for some time, taking care that the pressure and temperature is stable at the required values. After this initial time, the system will usually reach an equilibrium state. The coordinates and velocities are written to an output file at regular intervals, which represent as a function of time, a trajectory or ensemble of the system. By averaging over an already equilibrated trajectory, many macroscopic properties can be extracted from the output file.

A particular functional and parameter form of the potential is referred to as a force-field, where numerous ones have been developed and improved over the years [223]. The reliability of a MD methodology will depend on the force-field of choice and its parametrization. Since what goes into a MD simulation program is metaphorically 1 and 0's and what comes out is 1's and 0's to analyse, i.e. GIGO (Good In Good Out). One other limitation of MD simulations is of which, how to determine if the MD simulation of e.g. a protein have been accurately sampled. Moreover, if enough sampling has been done, to observe a particular conformational change, that are of interest to the researcher. Molecular dynamics being based upon statistical mechanics, hence care have to be taken how to interpret conformational changes of e.g. a protein, that one sees in a simulation. Just because one has seen a conformational change once, does not give the statistical certainty, that it will happen every time during a certain time-period.

To assess the accuracy of a simulation, one may calculate observables and compare with experimental properties. In addition, various statistical analyses can be done to assess if a simulation or calculated observable has converged [60]. Further expected are that the statistical quality will increase with the simulation time-length, since the conformational space are sampled more exhaustively. Notwithstanding, a simplified model that has been sampled well, may be more valuable, than a detailed model with poor statistics. However, the average of several relatively shorter simulations, may give better convergence of observables, than of one long simulation [224]. To note also is that several other studies have also indeed indicated, that the calculated observables from a MD simulation may depend significantly on e.g. the starting structure and initial velocities [224-226]. Hence a proper starting structure, that reflects a reasonable biological profile, may give more trustworthy results.

3.2.1 Simulation Method for Molecular Dynamics

The MD simulations were performed using software v. 5.0.4. GROMACS [48, 49]; wherein force-field chosen was CHARMM36 (mar. 2014) with TIP3P water model, (see mdp parameters in Appendix S3.1); the method and parameters were largely suggested from tutorials, GROMACS websites and relevant articles [227-234]; the initial structure of insulin, were taken from the protein data bank [235].

3.2.1.1 Replicas

For validating the MD simulations, the conformational space was sampled by performing replicas; the three different ones of any MD system being referred to as m, n, o. Firstly, an initial structure was selected, that had the lowest RMSD to all structures in a set of structures (see §3.3.1.3). The three replicas (m, n, o) had their respective initial structure slightly altered using modeller [236, 237], resulting in varying coordinates, see Figure 3.1. In hindsight three coordinates from the original set of DGR models could have been chosen, nevertheless, they still provide an initial set of starting coordinates, that are resembling of the mean structure of the respective set of DGR structures. Each replica also having different random seeds in the initial velocity generation, which add to the stoichastic variability, in addition to the different independent starting structures causing a further variation. Hence the respective replica (m, n, o) are used to refer to a starting structure and a particular seed. Furthermore e.g. replica "m" has the same value for seed regardless if original starting structure are from PDB 2KJJ or 2HIU or if a change in temperature.

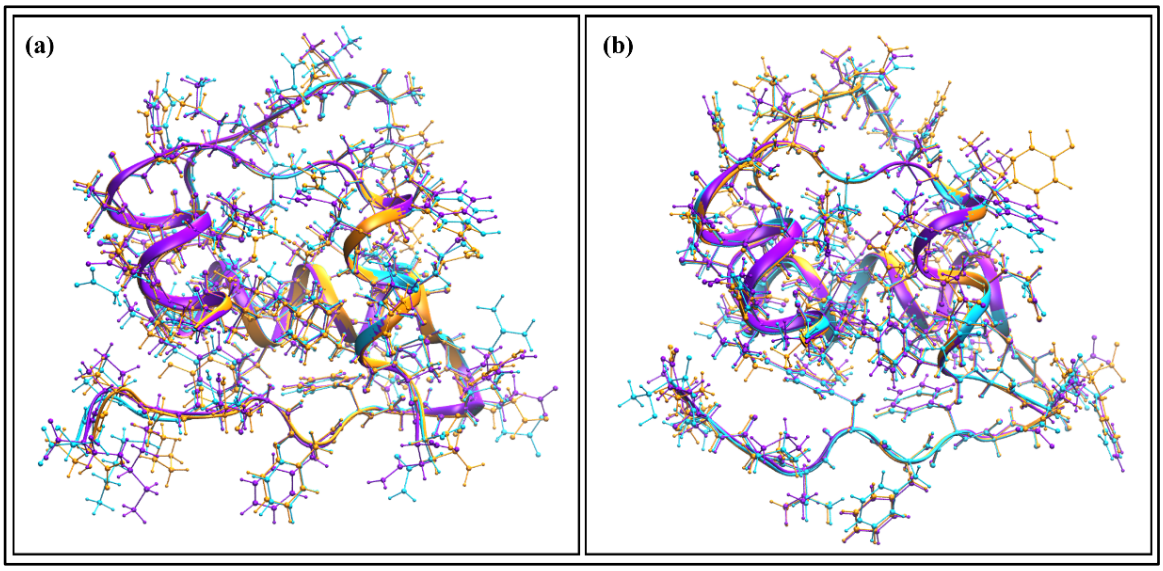


Figure 3.1**: Different initial starting coordinates for MD of a solvated insulin monomer Respectively designating replica m (purple), n (cyan), o (orange), generated via modeller from the mean structure of an ensemble of reported ensembles.* ***(a) *From PDB 2KJJ.* ***(b)*** *From PDB 2HIU.*

3.2.1.2 Simulation procedure

The following procedure is a standard example for this thesis, for simulating molecular dynamics of insulin monomers. Differences in parameters to simulations of other conditions, such as temperature and solvent, is noticed in the method sections of result

chapters.

The insulin monomer initial structure was placed in a cube of side 5.45 nm, with a minimum distance of 1.0 nm to the box edge. Then adding solvent in the box yielding 4995 water molecules. Then steepest descent minimization was performed by replacing water with a concentration of ions, 10 Na^+ and 8 Cl^-, to counterbalance an insulin monomer overall charge of $-2e$. Then, another steepest descent energy minimization was performed. Then the system was taken through three equilibrium simulations with position restraints on the protein. First under a NVT ensemble to stabilize temperature of system at 300 K, using V-rescale thermostat with coupling groups for protein and non-protein, with respective time constant of 0.1 ps. Generation of velocities was commenced only in this equilibration step, by a different random seed for each replica. Then under two NPT ensembles to stabilize the density and average of pressure to ~1 atm, with a coupling constant of 1 ps. That is equilibration with Berendsen for 400 ps, followed by equilibration with 400 ps of isotropic pressure coupling to a Parrinello-Rahman barostat, both with the same V-rescale thermostat. Then continuing with production simulation by the latter NPT ensemble without position restraint for 1499 ns. Due to the position restraint released after 0 ns and system not fully equilibrated, only the last 9 to 1499 ns (omitting 0-8 ns), were chosen to be considered in the analysis of any MD trajectory (though included in plots such as RMSD).

In equilibration and production simulation, the following parameters were used: Integration of Newton's equation of motion was performed using the Leap-Frog algorithm at 300 K with a time step of 2 fs. The cutoff scheme was Verlet with neighbour lists updated every 20 fs. Electrostatics was treated with particle-mesh Ewald (PME), using a coulomb cutoff of 1.2 nm and gridspacing of 0.1 nm, with a sixth-order interpolation. Van der Waals interactions was determined with force switching [228] between 1.0 − 1.2 nm. Covalent bonds to hydrogen atoms were constrained [228, 230] by LINCS. Periodic boundary conditions were applied in the x, y, z directions. No long range dispersion correction was applied [231, 238]. In minimization the same parameters were applied, except for using steepest descent integrator time step 1 fs and no constraints.

3.2.1.3 Nomenclature to distinguish MD ensembles

For distinguishing the resulting MD time dependent structure ensembles, with different parameters to the procedure described above, we apply a nomenclature. For example,

ensemble $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{MDm}}$, **E** for say experiment conditions, **kpi** acronym for the specific insulin analogue (KP-insulin), **MD** for molecular dynamics simulation; the notation $\boldsymbol{m}$ being a replica identifier (other two being n, o) and is included when referring to a particular replica or several replicas (e.g. as **mn**).

3.3 Analysis of any ensemble of structures

Here it is described the method of analysis for a static crystal structure and DGR or MD structure ensembles, which are presented in the results chapters. For all calculations of statistics (definitions in Appendix B), for MD time dependent structure ensembles (trajectories), it were only included times 9-1499 ns, if not otherwise indicated. Since all structures were renumbered consecutively, e.g. chain A is numbered 1 to 21 and chain B to 22-51, if considering larger structures other chains were renumbered subsequently, which are important in understanding the calculation of e.g. residue-wise distances. However, when referring to any residue in main text it is referred to its conventional chain-numbering, however consecutive numbering (if not also conventional) is shown in plots.

3.3.1 Geometry equations SASA, RGYR, RMSD, RMSF

3.3.1.1 Solvent accessible surface area

The solvent accessible surface area (SASA), is an estimate of the surface area of a molecule that is accessible to a solvent. Which were calculated for the insulin monomer, with a probe sphere of 1.4 Å. The VMD program, measure sasa [50, 239], was used to calculate this quantity.

3.3.1.2 Radius of gyration

To have a measure for the compactness of a protein structure, one can calculate the radius of gyration (RGYR):

$$r_{gyr}^2 = \left(\frac{\sum_i m_i |\mathbf{r}_i - \mathbf{r}_c|^2}{\sum_i m_i} \right) \qquad 3.4$$

, where m_i is the mass of atom "i" and $\mathbf{r}_i$ the position of atom *"i"*, $\mathbf{r}_c$ is the center of mass of the molecule. A lower RGYR value indicates a more compact structure, as such it is a

measure of a protein structure conformational compactness and stability. The VMD program, measure rgyr, was used to calculate this quantity [50, 239].

3.3.1.3 Root Mean Square Deviation

The root mean squared deviation (RMSD) is defined as [240, 241]:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} \left(\boldsymbol{r}_i(t) - \boldsymbol{r}_i(t_r)\right)^2}{N_{atoms}}} \qquad 3.5$$

, where N_{atoms} is the number of atoms whose position is considered, the vector $\boldsymbol{r_i}(t)$ is the position of atom *i* at time *t,* moreover $\boldsymbol{r}_i(t_r)$ is its position at a reference time t_r. Noting also that the variable, t, can just as well refer to a model in an ensemble of model structures. A RMSD value comparing a residue segment of a reference structure (e.g. the mean structure) to all other structures in the ensemble, were obtained for different residue sequences. For instance, seeing how much the residues B1-B5 or B25-B30 differ during a trajectory, compared to same residues in a reference structure. All RMSD calculations of this thesis includes all atoms for the residues compared.

3.3.1.4 Root Mean Square Fluctuation

The root mean square fluctuation (RMSF) is a measure of the deviation between the position of atom '*i'* and its reference position:

$$RMSF_i = \sqrt{\frac{\sum_{t=1}^{T} (\boldsymbol{r}_i(t) - \bar{\boldsymbol{r}}_\iota)^2}{T}} \qquad 3.6$$

, where *T* is the total numbers of structures in an ensemble (for $t > 8\ ns$ for MD trajectory), over which one wants to average, and $\bar{\boldsymbol{r}}_\iota$ is the ensemble averaged position of the same atom "*i"*. This equation may be used to calculate the RMSF for a particular atom over an ensemble. A difference between RMSD and RMSF, is that in the former the average is taken over the sum of atoms giving time specific values, whereas in the latter it is averaged over time giving a value for of each atom "*i"*. The average RMSF value of each residue were calculated in three quantities SC, MC (including amino or carboxy atoms for MC terminal residues) and all atoms (AA). Each atom in a residue's atom-selection is separately

calculated and the average RMSF is returned for any indicated residue, hence the notation e.g. $RMSF_{\langle SC \rangle}$ for the average RMSF of all SC atoms as a function of any residue. Further averaging over a chain or chains of residues are noted as $\langle RMSF_{\langle SC \rangle} \rangle$. The VMD program measure rmsf [50, 239] were used for this calculation.

B-factor to RMSF conversion

Some structures of insulin reported in PDB format (e.g. 4INS), includes B-factors (temperature factors or thermal parameter), for each atom of every residue. The magnitude of an atomic B-factor is formally a reasonable estimate of the mean square displacement of respective atom from its mean position, whereby the RMSF for each atom "i" can be defined as related to its B-factor value, B_i, by the equation [1, 54, 55, 242-251]:

$$RMSF_i = \sqrt{\frac{3B_i}{8\pi^2}} \qquad 3.7$$

Then the reside specific average RMSF value for any selection of atoms, SC, MC or all atoms (AA), with notation e.g. $RMSF_{\langle SC \rangle}$ as a function of any residue.

3.3.2 Script protocols for structure ensemble analysis

3.3.2.1 Postprocessing and obtaining a mean structure

The MD time dependent trajectory (or structure ensemble) obtained by method in §3.2.1, were postprocessed, whereof including concatenating the 3x500 ns, removing certain ''artefacts'' of the simulation, centring the protein in the box, and superimposing on a moiety of a structure (at a reference time). This was performed using GMX commands and tailored VMD scripts (see S3.2).

The MD simulation effectively yields a trajectory (or structure ensemble), of 1500, in 1 ns time-frames, i.e. the whole system, with protein and solvent, is outputted at time 0 ns, 1 ns, 2 ns … 1499 ns. Nevertheless, at each time-step, the protein and solvent, is in a random configuration. Hence, for enabling analysis of the protein, the protein and solvent had to be superimposed on a common moiety, e.g. a selection of atoms in the protein. Thus, translation and rotation were removed, by superimposing the whole system at a reference time of the trajectory, with regards to a moiety of the protein (C_α atoms of residues B11-

B17). This was performed with VMD program measure [50, 239] and GMX trjconv [211], that can obtain a transformation matrix [252]. This matrix when applied, will best align the coordinates of C_α atoms of residues B11-B17, for each 1 structure of the 1500 structures, with the coordinates of the same atoms at a reference time (e.g. 508 ns). The reference time, was chosen as the time, having the mean structure (MS) with lowest RMSD (Eq. 3.5) (including only all atoms of the protein), to all other structures in the 1500 structure trajectory (above equilibration time, $t > 8$ ns). Thus, for each time in this trajectory, the whole protein along with solvent, is transformed by this transformation matrix. Moreover, for each time in trajectory (or insulin and solvent structure ensemble), keeping relative coordinates of protein and solvent perfectly intact. This final ensemble trajectory was then used as input for all subsequent analysis.

Mean structure

The above obtained reference time (e.g. at 508 ns) has the protein structure with the lowest difference to all other protein structures at all other times (above 8 ns), hence this reference time contains the mean structure (MS). The MS of an ensemble of e.g. the 20 insulin structures in PDB entry 2KJJ, were simply the structure with the lowest RMSD (Eq. 3.5), to all other structures in the ensemble. Superimposing analogously via a transformation matrix the 20 structures on the MS (via C_α atoms of residues B11-B17).

3.3.2.2 Representation of structure ensembles and mean structure backbone

Flexibility of ensembles

From the MD structure ensemble (or time trajectory) obtained above, which already being superimposed on its reference time, containing the mean structure (of protein insulin), whereof was drawn 20 models on top of each other (from time 19 to 1499 ns, in 74 ns steps), using VMD program mol drawframes. For comparison to the DGR structures, reported in e.g. PDB 2KJJ, these were also superimposed on the reference time (e.g. 508 ns) of MD trajectory (via C_α atoms of residues B11-B17). Then this 20 structure ensemble was also drawn with mol drawframes.

Mean structure backbone

Separately compared are the mean protein structure of the MD and DGR ensemble, (superimposed via the C_α atoms of the B11-B17 residues). Also shown in this comparison, for the DGR and the MD ensemble, are the respective ensemble averaged positions of the C_α atoms (for MD ensemble $t > 8$ ns).

3.3.2.3 Fractional Occupancy of Protein and Solutes

The choice of solutes should influence a protein's biophysics in a MD simulation. To measure a particular presence in space, the fractional occupation during a trajectory, of protein, water and ions were calculated with VMD volmap [50, 239]. This program separates the analysis-box encompassing the solvent and protein (with their specific atom coordinates), into 3D gridcubes (of chosen sidelength 0.2 Å). Where atoms are regarded as spheres with radius being the atomic-radii of the atom-type. A number 1 is counted in any gridcube if it is encompassed by an atoms sphere; averaged over each 1×1500, hence a fractional occupancy are calculated. The whole 0-1499 ns was considered here, since an option was not available to choose the 9-1499 interval. However, the effect of the few 0-5 ns of stabilization of a MD production run is assumed to be negligible. The fractional occupancy was then visualized with VMD isosurface; whose iso-value can be tuned to show regions having equal or less than that fractional occupancy. For example, an iso-value 0.30 for a selection of molecules, will visualize the gridcube surfaces having up to 30% occupancy during a trajectory.

3.3.2.4 Distances between residues

Distances between each amino-acid residue of a protein, were obtained for every unit of a structure or trajectory ensemble. For example, the total number of residue pairs to compare in the 51 residues of the insulin monomer is 1275[$= 1 + 2.. + 50 =$51(51-1)/2]. The residue distances were calculated with regards to the geometric centre between the following atom-selections; SC to SC; CA-atom to CA-atom; SC to MC; MC to MC. These distinctions are made in order to reveal certain interactions between residues and are useful to determine if e.g. two SCs interact during a trajectory ensemble. The geometric centre of each residue's atom-selection, were obtained by adding the $\boldsymbol{r}(x, y, z)$ vector coordinates of each atom in any selection (including hydrogens), then dividing by its number of atoms. The following nomenclature are used to calculate the average distance between any atoms or geometric centre of any atom-selections:

$$\langle r_{(AS1,AS2)}^{(R1,R2)} \rangle = \frac{1}{S} \sum_{1}^{S} |\boldsymbol{r}_{AS1}^{R1} - \boldsymbol{r}_{AS2}^{R2}| \qquad 3.8$$

, where S means the total number of considered structures in an ensemble, *R1* is residue 1, *R2* residue 2, *AS1* for atom-selection 1 and *AS2* for atom-selection 2. For a MD trajectory ensemble only the 1490 insulin structures at times, 9-1499 ns, were considered in averaging. For only a single structure this equation simplifies to $r_{(AS1,AS2)}^{(R1,R2)}$. The distances were represented in matrices with the following division: First matrix; upper left ($i > j$) "SC to SC"; lower right ($i < j$) "CA to CA"; diagonal ($i = j$) distances zeroes; Second matrix: upper left ($i > j$) "SC to MC"; lower right ($i < j$) "MC to MC"; diagonal ($i = j$) distances 0's. Since the calculation of the distances were calculated residue-wise only once, for the unequal atom-selections of say "SC of R1" to "MC of R2", is that this was only calculated and depicted in matrices for the case where (R1<R2). The GRO files used in this thesis were calculated from CHARMM and VMD convention, which defines its GLY SC hydrogen as D-chiral, whereas the other amino-acids SCs are of course L-chiral, hence should be noted in these calculations.

3.3.2.5 Hydrogen bonds calculation and statistical comparison

Hydrogen bonds (HBs) being a fundamental interaction in proteins, hence it was considered for any single structure or ensembles of them. A description and specific definition used for calculating HBs are described in Appendix A.3. The calculation of HBs is a measure inherent of a structure, i.e., each structure is independent of e.g. atom-selection used in superimposing a trajectory. Hence this measure can tell something more about e.g. a proteins regional stability, than e.g. the quantity of RMSF. Here the criteria used are the distance range, "$|\boldsymbol{r}_{AD}| < 3.5$ Å", and angle range, "$\varphi < 90°$", which are considered to cover most of the HBs in proteins [253, 254]. The strongest HBs, tend to be of lesser angles, such that the three atoms of a HB lies closer to a straight line [25], reason why it are distinguished between three angles ranges in this thesis ($\varphi < 30° < 60° < 90°$). The HBs were calculated with VMD hbonds [50, 239], which were incorporated into an elaborate TCL script. This script could infer every possible HB, including multiple hydrogens of duplicate donor atoms, e.g. arginine hydrogens of its nitrogens, in addition to duplicate acceptor atoms, e.g. the two oxygen atoms of a CT COO- group. Hence all geometrically

possible HBs, with stated criteria, were included. This script also revealed, of a considered structure ensemble, the percentage for each possible HB fulfilling respective criteria of distance and angles. The HBs were represented in matrices (inverted with respect to row index *i*) with the following division: First matrix; upper left ($i > j$) "SC to SC" & "SC to MC" & "NH3+ involved as a donor hydrogen"; lower right ($i < j$) "MC to MC (including CT carboxy oxygens)"; diagonal ($i = j$) any HB within same residue. For ensembles of structures, the HBs were represented in different matrices with varying percentage of ensemble, e.g. a separate matrix including only HBs existing for more than 25% of structures.

In addition, for the insulin monomer, three structural HB classes are defined (I-III): intra-chain HBs, including stable α-helices and β-turns, that are within the AC (I) and BC (II) respectively, and inter-chain HBs between the AC and BC (III). The sum of HBs in respective class were considered (if given also the sum of all classes) in the format: "sum intra-chain HBs in AC"_"sum intra-chain HBs in BC"_sum inter-chain HBs between AC and BC"_("sum of preceding 3 sums"). A python script was written to compare the overlap between these different classes of e.g. ensembles and/or single structures. Considering that through a structure ensemble there may be e.g. 3 possible hydrogens (H1, H2, H3) of NT G^{A1}_{NH3+} within the limit of the HB criteria to e.g. CT oxygens (OT1, OT2) of T^{B30}_{COO-}, whereas in a single structure only 1 of its hydrogens (e.g. H3) and oxygen (e.g. OT1), may satisfy the criteria.

3.3.2.6 Dihedral angles

The dihedral angles (DAs) is a characteristic structural feature and hence can be used to compare insulin structures between themselves. The concept and definition I've found difficult to find explanation for in literature, however the basic concepts are explained in Appendix A.2. The actual atom definition of the DAs as used in calculations are defined in Table A1. Moreover, the ϕ, ψ and χ angles were calculated for every structure in an ensemble e.g. 0-1499 ns of a MD (insulin) structure ensemble.

3.3.3 Comparison to NMR derived restraints

One of the predominant methodologies, for studying dynamics and structure of biomolecules, are nuclear magnetic resonance (NMR). There are NMR observables e.g.

chemical shifts, ${}^{3}J$-coupling frequencies, nuclear over-hauser effect (NOE) intensities, whereof e.g. dihedral angle restraints and NOE distance upper bounds can be derived and structures can be determined [255-257]. Furthermore, it can be calculated from MD simulations geometrical observables, which are comparable with the experimentally derived ones [216, 220, 258, 259]. Particularly, NOE spectroscopy can display pairs of hydrogens that are in proximity, i.e. if two protons are less than about 5.5 Å apart. In other words, the effect provides one means of measuring the average distance and relative location of hydrogens [25], hence elucidates the three-dimensional structure of a protein in solution.

The following procedure was used for calculating NOEs from an ensemble of insulin structures. Moreover, for comparing these to respective experimental NOEs and derived hydrogen distance upper bounds [220]:

i. Locate all pairs of hydrogens that show a NOE in the experiment.
ii. Calculate the average distance between all possible pairs of hydrogens i and j, of insulin structures in ensemble. Obtaining predicted average distances, that if less than 5.5 Å are counted as NOEs, which can be compared to the corresponding experimental NOE distance bounds. The averaging of distances is described by the following equations, where a is an integer, e.g. 6 or 3 or -1, with the summation being over the whole ensemble of insulin structures, e.g. 20 for DGR and 1500 for MD ensembles:

$$R = \langle r_{ij}^{-a} \rangle^{-1/a} = \left(\frac{1}{S} \sum\nolimits_{1}^{S} \left| r_i - r_j \right|^{-a} \right)^{-1/a} \quad 3.9$$

$$R = \langle r_{ij} \rangle = \frac{1}{S} \sum\nolimits_{1}^{S} \left| r_i - r_j \right| \quad 3.10$$

, where S is total number of considered structures in an ensemble, with the latter equation being the unweighted averaging ($a = -1$). There are 381 hydrogen atoms in a native insulin monomer, meaning if the distances from each hydrogen to every other hydrogen are compared, there would then be 72390 pairs of them [$72390 = 1 + 2.. + 380 = 381(380)/2$]. When experimentally measured intensities are transformed to distance bounds, the $\langle r_{ij}^{-6} \rangle^{-1/6}$ assumption are of typical usage, with regards to structure determination [220, 260]. By increasing the variable a of Eq. 3.9, it increasingly takes a

smaller fraction of the ensembles, to fulfil, that a hydrogen distance to be averaged as less than 5.5 Å. For hydrogen pair distances having ambiguous wildcard assignment, e.g. the CH3-group of Val^{A3} and CH2-group of Glu^{A4}, it were considered only 1 out of 3 hydrogens in the CH3-group of Val^{A3}, which had the lowest averaged distance, to any of the 2 hydrogens in the CH2 group of Glu^{A4}. In addition, a minor notion is that *R* values of GLY HA1 were considered D-chiral and HA2 as L-chiral (opposite to the CHARMM convention).

3.3.3.1 Comparison to restrained hydrogen and HB distance bounds

There are some PDB entries of insulin analogues e.g. 2HIU, 2KJJ and 2JZQ, containing information about hydrogen distance restraints (RHH) and hydrogen bond restraints (RHB). For the RHH there are hydrogen identities of each of the two hydrogens in a pair, with the lower and upper bound of the distance between them. These PDB entries use the XPLOR [256, 261] format for distance restraints: "assign (resid nr and name H_i) (resid nr and name H_j) d d_{min} d_{max}", for example, "assign (resid 6 and name HN) (resid 5 and name HA) 3.400 1.600 0.00". From the PDB 2KJJ experimental NOEs, it was located the same pairs of hydrogens in an ensemble of insulin structures (by calculation of $\langle r_{ij}^{-6}\rangle^{-1/6}$). Then separating each NOE: below experimental lower bound (LB), $\langle r_{ij}^{-6}\rangle^{-1/6} < (d - d_{min})$; within bound (WB), $(d - d_{min}) < \langle r_{ij}^{-6}\rangle^{-1/6} < (d + d_{max})$; or above upper bound (UB), $(d + d_{max}) < \langle r_{ij}^{-6}\rangle^{-1/6}$. In addition, it was obtained a percentage of ensemble NOEs ending up in the respective bounds. Moreover, for every NOE higher than the UB in a structure ensemble, a violation was counted:

$$V_{ij} = \langle r_{ij}^{-6}\rangle^{-1/6} - (d + d_{max}) \qquad 3.11$$

For violations of an UB those were added to the beta value of each hydrogen in the pair. Hence specific atoms can have accumulated violations (added to beta-value) if involved in many restraints and be visualized with VMD. In addition, an average violation was obtained:

$$\langle V_{ij}\rangle = \frac{\sum V_{ij}}{(nr\ of\ violations)} \qquad 3.12$$

The same above calculations were performed for the RHBs, since the same format applies

for distance bounds between donor and acceptor atoms. However, assumed $\langle r_{ij} \rangle$-averaging for the two assigned distances in a structure ensemble, to be compared to each experimental restraint (including two restraints for D and DH to A distances).

3.3.3.2 Matrices of NOEs

How to sort the total number of calculated NOEs respectively from an insulin structure ensemble? Here it is described the procedure of obtaining 2-dimensional matrices of NOEs. For example, the 51 residues of insulin have $2601 (= 51^2)$ matrix indices, each index can be assigned a number of NOEs between residue "i" and residue "j". Thus, visualizing which residues having hydrogen pairs satisfying the NOE distance criteria (e.g. $\langle r_{ij}^{-6} \rangle^{-1/6}$ being less than 5.5 Å). In an analogous fashion, an experimental NOE matrix is built from the RHHs; adding a 1 to index (i, j) for each RHH between those residues.
The following division was used for all NOE matrices: upper left ($i > j$) "SC to SC NOEs" and "SC to MC NOEs"; lower right ($i < j$) "MC to MC NOEs"; diagonal ($i = j$) "SC and/or MC NOEs within same residue". The scoring of the number of NOEs is used to see a relative intensity of each index (i, j) of the matrix, visualizing which pairs of residues, has more comparable NOEs.

Matrix of calculated NOEs from a structure ensemble

In particular for a 1490 MD insulin structures ensemble (omitting the first 8 ns of interval 0-1499 ns) the procedure were as follows: time averaged distances, $\langle r_{ij}^{-6} \rangle^{-1/6}$, was calculated, between every hydrogen of residue i(row from bottom to up) to every hydrogen of residues j (column from left to right). For every NOE an increment of 1 was added for that index (i, j). For instance, for each of the 7 SC hydrogens of residue V^{A3}, having a time averaged distance (e.g. $\langle r_{ij}^{-6} \rangle^{-1/6}$), to each of the 9 SC hydrogens of I^{A2}, being less than 5.5 Å, the value of that 'upper left' matrix index ($i = 3,\ j = 2$) was increased by 1.
For the 20 DGR structure ensemble reported in PDB entry 2KJJ, NOE matrices were obtained the same way, just using the number of models as an input for analysis. The atom naming format for the PDB structures, was before converted to be the same as for the MD nomenclature.

Matrix of experimental NOEs

From the RHHs reported in PDB entry 2KJJ, a matrix of experimental NOEs was obtained, by simply adding a 1 to each matrix index (i, j), for every RHH existing between those residues. The 10 NOEs involving Glycine HA# (# meaning unassigned hydrogen identity) were omitted, however 6 NOEs involving HA1, HA2 were included, however the chirality of these unknown.

3.3.3.3 Comparison to restrained DAs

All possible DAs (as defined in Table A1) were calculated for ensembles of insulin structures. These were compared to respective DA restraint (RDA) from the PDB entry 2KJJ (here redepicted in Table 3.2). Where each RDA are reported with four atoms having a DA e.g. -65° of bounds $\pm$40°, hence LB -105° and UB -25°. The DA atom definition for the RDAs were assumed to be the same as in Table A1 (though SCs such as valine have two indistinguishable χ_1 angles), whereby respective calculated DA of each structure in an ensemble, were checked if within bounds (WB).

Mean DAs being awkward to define, wherefore the DAs and congruence with RDAs were visualized for the mean structure (MS). Moreover, calculating the fraction WBs of structures in ensemble.

Table 3.2*: The RDAs of a solvent model of KP-insulin . Redepicted from PDB 2KJJ, containing 47 RDAs, the bounds on all RDAs are* $\pm$40°*. Personal correspondence with peers of author of Qua et al. [3], suggests them probably being derived mainly from* ^{13}C *chemical shifts, also that the bounds may be rather large, and that the RDAs may not be totally accurate.*

DA abbr.	DA°	A10 I(χ_1)	-60	A18 N(ϕ)	-65	B6 L(ϕ)	-120	B17 L(χ_1)	180
A2 I(ϕ)	-65	A10 I(ϕ)	-120	A19 Y(χ_1)	-60	B10 H(ϕ)	-65	B17 L(ϕ)	-65
A3 V(ϕ)	-65	A12 S(χ_1)	60	A19 Y(ϕ)	-65	B11 L(χ_1)	-60	B18 V(χ_1)	-60
A4 E(ϕ)	-65	A12 S(ϕ)	-120	A20 C(ϕ)	-65	B11 L(ϕ)	-65	B18 V(ϕ)	-65
A5 Q(χ_1)	180	A13 L(ϕ)	-65	B2 V(ϕ)	-120	B12 V(χ_1)	-60	B19 C(χ_1)	-60
A5 Q(ϕ)	-65	A14 Y(χ_1)	180	B4 Q(χ_1)	180	B12 V(ϕ)	-65	B24 F(χ_1)	60
A6 C(χ_1)	-60	A14 Y(ϕ)	-65	B4 Q(ϕ)	-120	B13 E(ϕ)	-65	B25 F(ϕ)	-120
A6 C(ϕ)	-65	A15 Q(ϕ)	-65	B5 H(χ_1)	-60	B14 A(ϕ)	-65	B27 T(ϕ)	-120
A7 C(χ_1)	180	A16 L(ϕ)	-65	B5 H(ϕ)	-65	B15 L(ϕ)	-65		
A10 I(χ_2)	60	A17 E(ϕ)	-65	B6 L(χ_1)	-60	B16 Y(ϕ)	-65		

Chapter 4 An Analysis and View of Monomer Bound Complexes

Structures of Insulin Monomers Bound in Hexamers and Receptor Fragments

The insulin monomer with its relatively small surface area can assemble into oligomers and form a cross-linked binding in its IR binding region. This chapter will briefly document and calculate some geometrical properties and binding surfaces of a few structures pertaining to insulin structural biology. This in order, for complementing the structures in their original reports, and provide some basis for comparison with solvent models. In particular, the structural overview elucidates otherwise unseen view-angles of these structures. Especially interesting is a detailed analytical representation, for the lesser resolution structure, of the contiguous IR-A binding region with insulin bound at high affinity.

4.1 An analysis and view of insulin in T-state hexamer crystals

Putatively the average conformation of insulin in solution is close to that of a T-state, hence there should be some similarity of intra-monomer contacts and relative fluctuation. An overview of the hexamer crystal structure and binding surfaces were reviewed in §2.4, in particular for the T-state protomer of Baker *et al.* [1], here it is built on this model and add ed some extra unreported geometrical analysis. As modern computational methods and power can provide another vantage-point of visualizing the structure than were done in the late 1980's, hence can complement the original report. In addition, it is an evaluation system for the innovative analysis approach developed in this thesis, since it is one out of few well documented structures. Furthermore, in order, to provide a reference of comparison of structure and HBs, to the T-state resembling solution models of insulin, as investigated in this thesis.

4.1.1 Comments on methods

Many analysis approaches developed in this thesis, were readily applicable to T-state hexamer structures as reported in two PDB entries, those considered were the PDBs of 4INS

and 4E7T. First a conversion from the PDB to GRO format were done, i.e. these structures were added hydrogens and converted to charmm36(mars 2014) naming format via gmx pdb2gmx. Protonation states as of pH 7.4 were applied in conversion, allowing a more computation-able comparison to solvent models (of the next result chapters). For the crystal structures, with residues having reported atoms of occupancies about 50% of alternate configurations (A and B), for which case only the A configuration were included (as outputted by gmx pdb2gmx). The resulting GRO file format were used for calculation of the analytical properties, such as DAs, HBs and residue distances.

4.1.2 A comparison of the hydrogen bonds of a T-state dimer

Here regarding first a few T-state dimers, from two sources, and a simulation of a dimer of them, certifying their general structure and allowing a comparison. The HB network elucidates the structure, being a foundational scaffold for crystal protein structures. Here it is calculated the intra-monomer HBs of the asymmetric dimer unit, which are shown in Table 4.1 (visualized for **M12** in Figure S4.1), moreover for wider angles ($60° < \varphi < 90°$) in Table S4.2. The wider HBs may be considered weaker, since strong HBs tend to be of lesser angles. Moreover, the more encompassing angle criteria ($\varphi < 90°$) are stated to have been used by Baker *et al.* [1, 262].

The classical T-state hexamer structure as obtained by two sources

This is the most famous structure, who Baker *et al.* [1] delivered an elaborate report for, chosen for the elaborate analysis comparison, as it is better documented. However, a confirmation of this T-state crystal structure (of a T_6 hexamer) were done in a more recent report by Frankaer *et al.* [263] (PDB 4E7T). Their T-state asymmetric dimer unit of bovine insulin, has a high resemblance to the one of Baker *et al.* [1]. Apparent is that **M1**, **M2** share most HBs with **BM1**, **BM2**, as indeed their overall structure are congruent.

A shorter MD simulation of the asymmetric dimer

About 16 years later from when Baker *et al.* [1] published their work, there was published a 5 ns simulation of this dimer unit by Zoete *et al.* [57]. Here were reported the HBs of this simulated dimer unit (**MDM1** and **MDM2)**, showing these in Table 4.1, which were present for more than 50% of their simulation, with similar HB criteria ($|\boldsymbol{r}_{AD}| < 3.2$ Å $\ \& \ \varphi < 60°$).

Note that possibly a larger distance and angle range would have included a lot more potentially physically relevant HBs. Also note that Zoete *et al.* [57], uses **M1** to refer to **M2** and vice versa, so I find it uncertain as to which monomer is which, and portray it here as I believe it is meant. However, a related but varying MD methodology to the one in §3.2.1 were used in their work. At that time, a 5 ns simulation were a relatively long time-scale simulation, for a protein of this size (102 residues solvated with water). In addition, noting also, is that their simulation of **M2**, did not change much from the starting conformation until after about 2.75 ns, which indicates a lesser sampled simulation. Notwithstanding, some similarity of HBs to the one of the crystal structures are evident. However, they are in addition expected to be different, partly due to that the simulation is in a dynamic solution of explicit water, whereas in the crystal the dimer is a unit of periodic hexamer crystals.

Noted apparently strong common intra-dimer HBs in solvent as in a crystal

Comparing only the crystal structure HBs, there are evidently seen many strong HBs stabilizing this structure. However, in relation to the solution structure, there appears to be a few strong HBs that remains especially rigid noted below. First for the AC, there are some stabilizing HBs: "$\mathrm{Q}_{\mathrm{HN}}^{\mathrm{A5}} \rightarrow \mathrm{G}_{\mathrm{O}}^{\mathrm{A1}}$"; "$\mathrm{C}_{\mathrm{HN}}^{\mathrm{A6}} \rightarrow \mathrm{I}_{\mathrm{O}}^{\mathrm{A2}}$"; "$\mathrm{I}_{\mathrm{HN}}^{\mathrm{A10}} \rightarrow \mathrm{S}_{\mathrm{OG}}^{\mathrm{A9}}$"; "$\mathrm{S}_{\mathrm{HN}}^{\mathrm{A12}} \rightarrow \mathrm{Q}_{\mathrm{OE1}}^{\mathrm{A15}}$"; "$\mathrm{E}_{\mathrm{HN}}^{\mathrm{A17}} \rightarrow \mathrm{L}_{\mathrm{O}}^{\mathrm{A13}}$"; "$\mathrm{Y}_{\mathrm{HN}}^{\mathrm{A19}} \rightarrow \mathrm{L}_{\mathrm{O}}^{\mathrm{A16}}$"; "$\mathrm{C}_{\mathrm{HN}}^{\mathrm{A20}} \rightarrow \mathrm{E}_{\mathrm{O}}^{\mathrm{A17}}$". Next of the BC, the HBs: "$\mathrm{H}_{\mathrm{HN}}^{\mathrm{B10}} \rightarrow \mathrm{C}_{\mathrm{O}}^{\mathrm{B7}}$"; "$\mathrm{R}_{\mathrm{HN}}^{\mathrm{B22}} \rightarrow \mathrm{C}_{\mathrm{O}}^{\mathrm{B19}}$", are stabilizing the turn regions, with the evident BC α-helical HBs, indeed conveys the result of being most settled and rigid.

The inter-chain HBs are also interesting, since they contribute in stabilizing the two chains; apart from the disulphide links and inter-monomer/-dimer contacts. A few HBs are seen to stabilize the BC NT strand, in particular two MC HBs between residues $\mathrm{C}^{\mathrm{A11}}$ and Q^{B4}. Moreover, stabilizing the region in vicinity of the NT BC loop, for both monomers, are the HB "$\mathrm{L}_{\mathrm{HN}}^{\mathrm{B6}} \rightarrow \mathrm{C}_{\mathrm{O}}^{\mathrm{A6}}$". Moreover, the region in vicinity of the CT BC turn region are stabilized by two HBs: "$\mathrm{N}_{\mathrm{HN}}^{\mathrm{A21}} \rightarrow \mathrm{G}_{\mathrm{O}}^{\mathrm{B23}}$" and "$\mathrm{F}_{\mathrm{HN}}^{\mathrm{B25}} \rightarrow \mathrm{Y}_{\mathrm{O}}^{\mathrm{A19}}$". The only four HBs between the two monomers in crystal as in the dimer simulation, are the two duplicates of two HBs: "$\mathrm{Y}_{\mathrm{HN}}^{\mathrm{B26}} \rightarrow \mathrm{F}_{\mathrm{O}}^{\mathrm{B24}}$" and "$\mathrm{F}_{\mathrm{HN}}^{\mathrm{B24}} \rightarrow \mathrm{Y}_{\mathrm{O}}^{\mathrm{B26}}$".

***Table 4.1**: Intra-monomer HBs, for T-state crystal dimer units , i.e. monomer 1 (**M1**), monomer 2 (**M2**) of PDB entry 4INS, and an explicit solvent MD simulation of a dimer. Counting the HBs in format of number of HBs in designated chains; "AC_BC_AC&BC (sum)". Corresponding nr for X-ray with different criteria: $r_{AD} < 3.5$ Å & $\varphi < 30°$ (greyed), 60° (additional in white background), 90°(not shown); **M1** is 8_11_6(25), 19_16_8(43), 43_44_10(97); **M2** is 11_8_7(26), 19_12_9(40), 45_35_9(89). Moreover, same range for PDB entry 4E7T; **BM1** 9_10_5(24), 16_14_8(38), 40_42_10(92); **BM2** 9_7_4(20), 19_15_7(41), 44_40_9(93). For MD of asymmetric dimer in explicit solvent, HBs of more than 50% occupancy: **MDM1** is 9_9_5(23): **MDM2** is 9_10_4(23).*

Donor	Acceptor	M1	M2	BM1	BM2	MDM1	MDM2	B4 Q(HE21)	B2 V(O)				✓		
								B3 N(HD21)	B3 N(N)	✓					
								B10 H(HN)	B7 C(O)	✓	✓	✓	✓	✓	
A1 G(H3)	A4 E(OE#)	✓	✓	✓	✓			B11 L(HN)	B7 C(O)					✓	✓
A4 E(HN)	A4 E(OE#)	✓	✓	✓	✓			B11 L(HN)	B8 C(O)	✓	✓	✓	✓		
A4 Q(HN)	A1 G(O)				✓			B12 V(HN)	B8 G(O)	✓		✓		✓	✓
A5 Q(HN)	A1 G(O)	✓	✓	✓	✓	✓	✓	B13 E(HN)	B9 S(O)	✓	✓	✓	✓		✓
A5 Q(HE21)	A10 I/V(O)		✓		✓			B14 A(HN)	B10 H(O)	✓	✓	✓	✓	✓	✓
A5 Q(HE22)	A15 Q(OE1)		✓		✓			B15 L(HN)	B11 L(O)	✓	✓	✓	✓	✓	✓
A5 Q(HE21)	A19 Y(OH)	✓						B16 Y(HN)	B12 V(O)	✓	✓	✓	✓	✓	✓
A6 C(HN)	A2 I(O)	✓	✓	✓	✓	✓	✓	B17 L(HN)	B13 E(O)	✓	✓	✓	✓	✓	✓
A7 C(HN)	A3 V(O)	✓		✓		✓		B18 V(HN)	B14 A(O)	✓	✓	✓	✓	✓	✓
A7 C(HN)	A4 E(O)		✓		✓			B19 C(HN)	B15 L(O)	✓	✓	✓	✓	✓	✓
A8 T/A(HN)	A3 V(O)	✓		✓		✓		B20 G(HN)	B16 L(O)	✓	✓	✓	✓		
A8 T(HN)	A4 E(sc)						✓	B22 R(HH21)	B18 V(O)				✓		
A8 T/A(HN)	A4 E(O)		✓		✓		✓	B22 R(HN)	B19 C(O)	✓	✓	✓	✓		✓
A9 S(HN)	A4 E(O)	✓		✓				B23 G(HN)	B20 G(O)	✓	✓	✓	✓		
A9 S(HN)	A5 Q(O)		✓		✓		✓	B29 L(HZ2)	B30 (OT1)				✓		
A10 I/V(HN)	A9 S(OG)	✓	✓	✓	✓	✓	✓	B30 A(HN)	B27 T(O)	✓		✓			
A12 S(HN)	A15 Q(OE1)	✓	✓	✓	✓	✓	✓	A11 C(HN)	B4 Q(O)	✓	✓	✓	✓	✓	✓
A15 Q(HN)	A12 S(OG)	✓	✓	✓	✓			A14 A(HN)	B10 H(O)						
A15 Q(HN)	A12 S(O)	✓	✓					A21 N(HD21)	B22 R(O)				✓		
A16 L(HN)	A12 S(O)	✓	✓	✓	✓			A21 N(HN)	B23 G(O)	✓	✓	✓	✓	✓	✓
A17 E(HN)	A13 L(O)	✓	✓	✓	✓	✓	✓	B4 Q(HN)	A11 C(O)	✓	✓	✓		✓	✓
A17 E(HN)	A14 L(O)	✓	✓	✓	✓			B5 H(HE2)	A7 C(O)		✓		✓		
A18 N(HN)	A14 Y(O)						✓	B5 H(HD1)	A7 C(O)	✓		✓			
A18 N(HN)	A15 Q(O)	✓	✓	✓	✓			B6 L(HN)	A6 C(O)	✓	✓	✓	✓	✓	✓
A19 Y(HH)	A5 Q(NE2)	✓						B22 R(HE)	A21 (OT1)	✓	✓	✓			
A19 Y(HN)	A16 L(O)	✓	✓	✓	✓	✓		B22 R(HH#1)	A21 (OT1)	✓	✓	✓			
A20 C(HN)	A17 E(O)	✓	✓	✓	✓	✓		B25 F(HN)	A19 Y(O)	✓	✓	✓	✓	✓	
B3 N(HD21)	B2 V(O)	✓						B29 K(HN)	A4 (OE2)		✓		✓		

4.1.2.2 A comparison of T-state monomers calculated RMSF of B-factors from a few different sources

The average B-factor for residue-wise MC and SC atoms respectively, were depicted and discussed in the original report of Baker *et al.* [1]. Here it is also calculated (Figure S4.2), showing near identical to original values; variations assumed at least partly to be due to rounding errors (due to less decimal precision of the numbers calculated near mid-1980s). Further sources of variations are that Baker et al. included one of the OT1 or OT2 (the terminal OXT atom of PDB entry) in the SC atom B-factor averaging, whereas here it was included in that of MC. In addition, for the residues with two SC configurations, they included only the first (A-configuration) in the calculation, whereas here the average is over both configurations. Merely the **M1** K_{SC}^{B29} and A_{SC}^{B30} average B-factors appears to have some mistake in calculation or assignment in the original report. The atom-wise B-factor being presumed to be directly proportional to the mean square displacement from the atoms mean position. From which the average B-factor over MC or SC atoms gives a measure of that respective selections average mean square displacement. Conjointly, the average RMSF (Eq. 3.7), being proportional to the square root of B-factor, gives a corresponding, albeit different graph, of relative proportions in residue-wise fluctuations. Nonetheless, the average RMSF have to my knowledge not been previously calculated and depicted in any report. Only B-factor derived RMSF values of CA atoms [54, 57], by Eq. 3.7, has previously been used in comparison to shorter time scale MD models of **MDM1** and **MDM2**. Hence here the quantities of average RMSF of SC and MC atoms (or $RMSF_{\langle SC\rangle}$, $RMSF_{\langle MC\rangle}$) are respectively given in Figure 4.1, as derived from the same B-factors (see §3.3.1.4). The smaller B-factors are expected for those atoms e.g. making strong HBs and other stabilizing contacts, such as the MC in αHs. The larger B-factors are the sum of atomic vibrations and the movements of peptide moieties as a group, e.g. in less sterically hindered spaces of certain residues in the crystal. Since this is merely another view of the results of Baker et al., one should refer also to the original report if drawing unrelated conclusions from the ones in this thesis. As expected, for almost all residues, the $RMSF_{\langle SC\rangle}$ are larger than or equal for the $RMSF_{\langle MC\rangle}$; two exceptions being for the cysteine bridge at A20-B19 and of F_{SC}^{B24} being buried. The different average RMSFs at CT and NT in **M1** and **M2** is due too different and looser contacts in the crystal. The higher average RMSF of **M1** than in **M2**, for the AC and BC is seen. For the AC it was explained by that of **M1**'s fewer crystal contacts than **M2**. It appears that also the BC has less overall sterical hindrance in **M1** than

in **M2**. Another bovine T-state crystal structure [263] were considered, having a high resemblance to the porcine insulin of Baker et al.. For which, also the average RMSF was calculated from the B-factors shown in Figure S4.3. Having an overall similar profile, albeit somewhat higher fluctuations, however verifying the common elements to some extent.

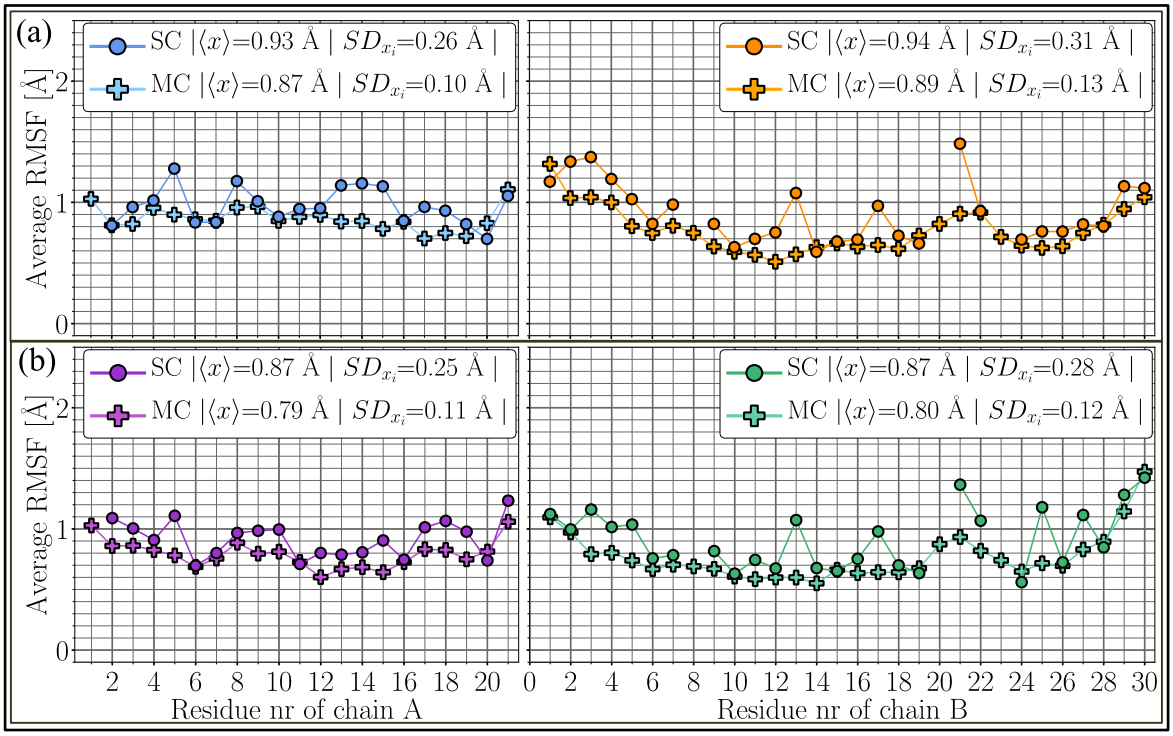


***Figure 4.1**: Average RMSF of crystal T-state monomers . For all non-hydrogen atoms of indicated selection for each residue. **(a) M1**. **(b) M2**. The statistics are separate for AC and BC where x in ⟨x⟩ refers to $RMSF_{\langle SC\rangle}$, and $RMSF_{\langle MC\rangle}$ respectively. The temperature factors (or B-factors) from PDB 4INS (biological assembly 7).*

4.1.2.3 Dihedral angles of M1 and M2

The distinct DAs of **M1** and **M2** are shown in Figure 4.2. These are recalculated DAs as relating to the Ramachandran angles (ϕ, ψ), which were reported of Baker *et al.* [1]. Both sources of DAs show near identical congruence, on average $0-1°$ off, again possibly due to here using higher decimal precision. Here also the NT, CT (ϕ, ψ) are defined as zero. In addition, by Baker *et al.* the NT ϕ-angle, were regarded as zero, however not understood logically, is the CT ψ-angle having nearer to trans orientation. In addition, here correcting an apparent typo of **M2** V_{ϕ}^{B21}. Notwithstanding, the χ-angles was not reported originally, hence the DA signature of **M12** in Figure 4.2, gives a complimenting graph and enables a reference for comparison. Albeit, considering also that only the A-conformation of alternate configurations were calculated here.

In addition, as were explained in §3.3.3.3, here the DAs was compared to 47 RDAs of a solution model (those in Table 3.2). Here it is noted that the solvent model derived RDAs has a fair correspondence in a T-state monomer. There are similar discrepancies which are noted further in the following chapters, concerning solvent models. Here there are 39 WBs (within bounds) for **M1**, those not WBs are: $\mathrm{Q}_{\chi_1}^{\mathrm{A5}}$, $\mathrm{C}_{\phi}^{\mathrm{A6}}$, $\mathrm{I}_{\chi_1,\chi_2}^{\mathrm{A10}}$, $\mathrm{Y}_{\chi_1}^{\mathrm{A14}}$, $\mathrm{Q}_{\chi_1}^{\mathrm{B4}}$, $\mathrm{L}_{\chi_1}^{\mathrm{B11}}$, $\mathrm{V}_{\chi_1}^{\mathrm{B18}}$. And for **M2** there are 37 WBs, those not WBs are: $\mathrm{I}_{\chi_1,\chi_2}^{\mathrm{A10}}$, $\mathrm{Y}_{\chi_1}^{\mathrm{A14}}$, $\mathrm{Q}_{\chi_1}^{\mathrm{B4}}$, $\mathrm{H}_{\phi,\chi_1}^{\mathrm{B5}}$, $\mathrm{L}_{\chi_1}^{\mathrm{B11}}$, $\mathrm{V}_{\chi_1}^{\mathrm{B12}}$, $\mathrm{L}_{\chi_1}^{\mathrm{B17}}$, $\mathrm{V}_{\chi_1}^{\mathrm{B18}}$. Noting that each bound of any specific DA designation, for these restraints, are fairly large ($\pm 40°$), however ϕ-angles that are not WBs are not far off, whereas the solvent exposed SCs can be understood to have some difference in angles, the core adjacent SCs of $\mathrm{L}^{\mathrm{B11}}$, $\mathrm{V}^{\mathrm{B18}}$ also have a deviating orientation. As expected, the more α-helical segments fulfil the ϕ-angle RDAs, also the disulphide bonded DAs of $\mathrm{C}_{\chi_1}^{\mathrm{A6}}$, $\mathrm{C}_{\chi_1}^{\mathrm{A7}}$ and $\mathrm{C}_{\chi_1}^{\mathrm{B19}}$. Noteworthy is for $\mathrm{L}_{\chi_1}^{\mathrm{B6}}$ in addition to $\mathrm{V}_{\phi}^{\mathrm{B2}}$, $\mathrm{Q}_{\phi}^{\mathrm{B4}}$ and $\mathrm{L}_{\phi}^{\mathrm{B6}}$ being WBs, congruent with a T-state NT BC loop. There is an angle for $\mathrm{F}_{\chi_1}^{\mathrm{B24}}$ at $\sim 60°$, which indicates this as a hingepoint, even in solution. Moreover. interesting is that $\mathrm{F}_{\phi}^{\mathrm{B25}}$, $\mathrm{T}_{\phi}^{\mathrm{B27}}$ are WBs, partly indicating that $\mathrm{F}_{\mathrm{SC}}^{\mathrm{B24}}$, $\mathrm{Y}_{\mathrm{SC}}^{\mathrm{B26}}$ are prone to be near the core, even in solvent.

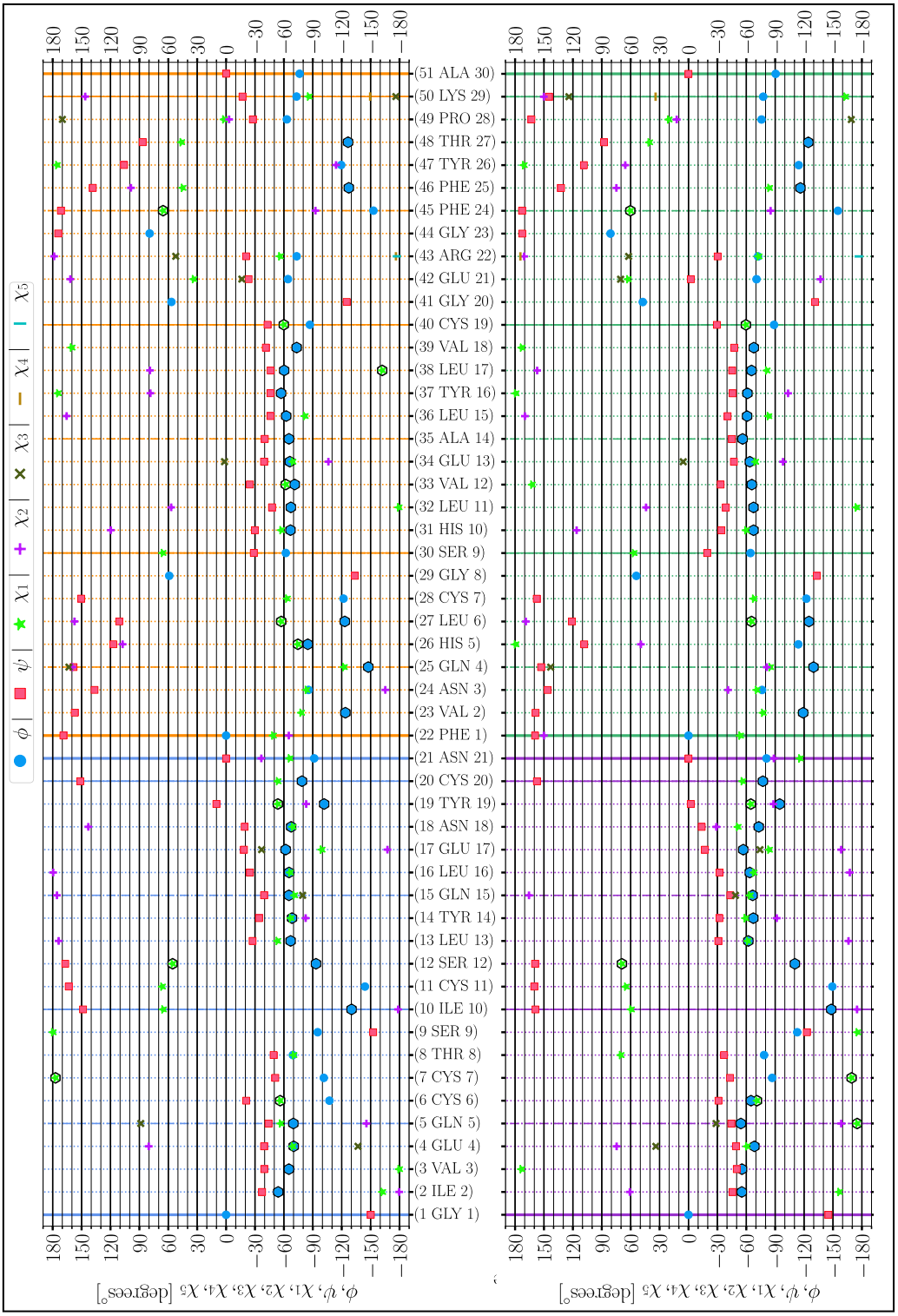


***Figure 4.2**: The dihedral angles of an asymmetric dimer .*

*(Top) **M1** (Bottom) **M2**.*

*Black hexagons (39 for **M1** 37 for **M2**) are for those DAs within bounds of the RDAs from the solution model reported in PDB 2KJJ.*

Chain residues renumbered sequentially 1-51 (nearest top graph), and with actual residue name and chain number.

Graph has zoomable vector graphics.

*Grid-lines for **M2**, **M1** respectively for AC (1-21) in purple, blue and BC (22-51) turquoise, orange.*

Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Calculated from structure in PDB 4INS (biological assembly 7).

4.1.2.4 Conformational analytical overview and a few residue-profiles of M12

The structure of **M12** are shown with chain-numbering (Figure 4.3), and with distances between residue-moieties (Figure 4.4), and with the sorted HBs between residue-moieties (Figure 4.5; same as in Table 4.1), with the residues DAs already shown above (Figure 4.2). In addition, demonstrating that this graphical zoom method is applicable to larger systems, there is a corresponding overview of symmetry and contacts of the whole hexamer (Figure S4.6, Figure S4.7). The original report of Baker *et al.* [1] has a different representation of each residue's contacts and nearby structure, and should be consulted with also. Since residue-profiles were described in the original report, they are not elaborately repeated here. However, the representation here, complements the original report, since here it is given an elaborate overview and vantage-point of its structure in just a few graphs. This conformational overview is advantageous, partly for relating to other insulin models in various environments. From the graphs it is fairly evident, which residues that have atom-contacts within 5 Å of the dimer and hexamer surface, of any atom of insulin, albeit here distances are between geometric-centres of respective atom-selections (including hydrogens). Noting some aspects of a few example residues explicitly, to make acquaintance with how to interpret the below graphs, referring to common traits in both monomers (by designating **M12)**, if not indicated otherwise individually (as **M1** and **M2)**. The same selection of residue-profiles are given for the other insulin structure models of $\mathbf{CF}_{(\mathbf{A},\mathbf{B})}^{\mathbf{6HN5}}$ (§4.2.2), $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ (§5.2.5) and $\mathbf{P}_{\mathbf{kpi}}^{\mathbf{MDm}}$ (§6.2.5); providing a direct comparison.

(A1 G): Fairly mobile in the crystal as indicated by its RMSF. Moreover, for example $\langle r_{(SC,SC)}^{A1,A4} \rangle$ is 5.6 Å (between 5.5-6.0 Å), reflecting that the saltbridge "$\mathrm{G}_{\mathrm{H3}}^{\mathrm{A1}} \rightarrow \mathrm{E}_{\mathrm{OE\#}}^{\mathrm{A4}}$", explaining proximity of respective SCs. Other intra-monomer contacts are to I^{A2}, V^{A3}, Q^{A5} and Q^{A5}; in addition to $\mathrm{K}^{\mathrm{B29}}$ in **M2** and $\mathrm{Y}^{\mathrm{A19}}$, $\mathrm{T}^{\mathrm{B30}}$ in **M1**. Note also the strong HB "$\mathrm{Q}_{\mathrm{HN}}^{\mathrm{A5}} \rightarrow \mathrm{G}_{\mathrm{O}}^{\mathrm{A1}}$".

(B4 Q): At the NT BC, close to the core, here $\langle r_{(SC,SC)}^{A11,B4} \rangle$ is between 4.0-4.5 Å, reflecting two strong HBs: "$\mathrm{C}_{\mathrm{HN}}^{\mathrm{A11}} \rightarrow \mathrm{Q}_{\mathrm{O}}^{\mathrm{B4}}$, $\mathrm{Q}_{\mathrm{HN}}^{\mathrm{B4}} \rightarrow \mathrm{C}_{\mathrm{O}}^{\mathrm{A11}}$"; however there is slightly longer distances in **M1** congruent with a distinctly larger RMSF. Intra-monomer contacts are to: V^{B2}, N^{B3}, H^{B5}, L^{B6}, S^{A9}, $\mathrm{I}^{\mathrm{A10}}$, $\mathrm{C}^{\mathrm{A11}}$, $\mathrm{S}^{\mathrm{A12}}$ and $\mathrm{L}^{\mathrm{A13}}$. In the hexamer interface this residue in **M2** has contacts to **M1** of the second dimer, likewise this residue in **M1** has contacts to **M2** of the third dimer; which are to $\mathrm{Y}^{\mathrm{B16}}$, $\mathrm{L}^{\mathrm{B17}}$, $\mathrm{V}^{\mathrm{B18}}$ and $\mathrm{G}^{\mathrm{B20}}$; moreover for **M1** it has proximity to $\mathrm{E}^{\mathrm{B21}}$ and a cross-dimer HB i.e. "$\mathrm{Q}_{\mathrm{HE22}}^{\mathrm{B4}} \rightarrow \mathrm{L}_{\mathrm{O}}^{\mathrm{B17}}$".

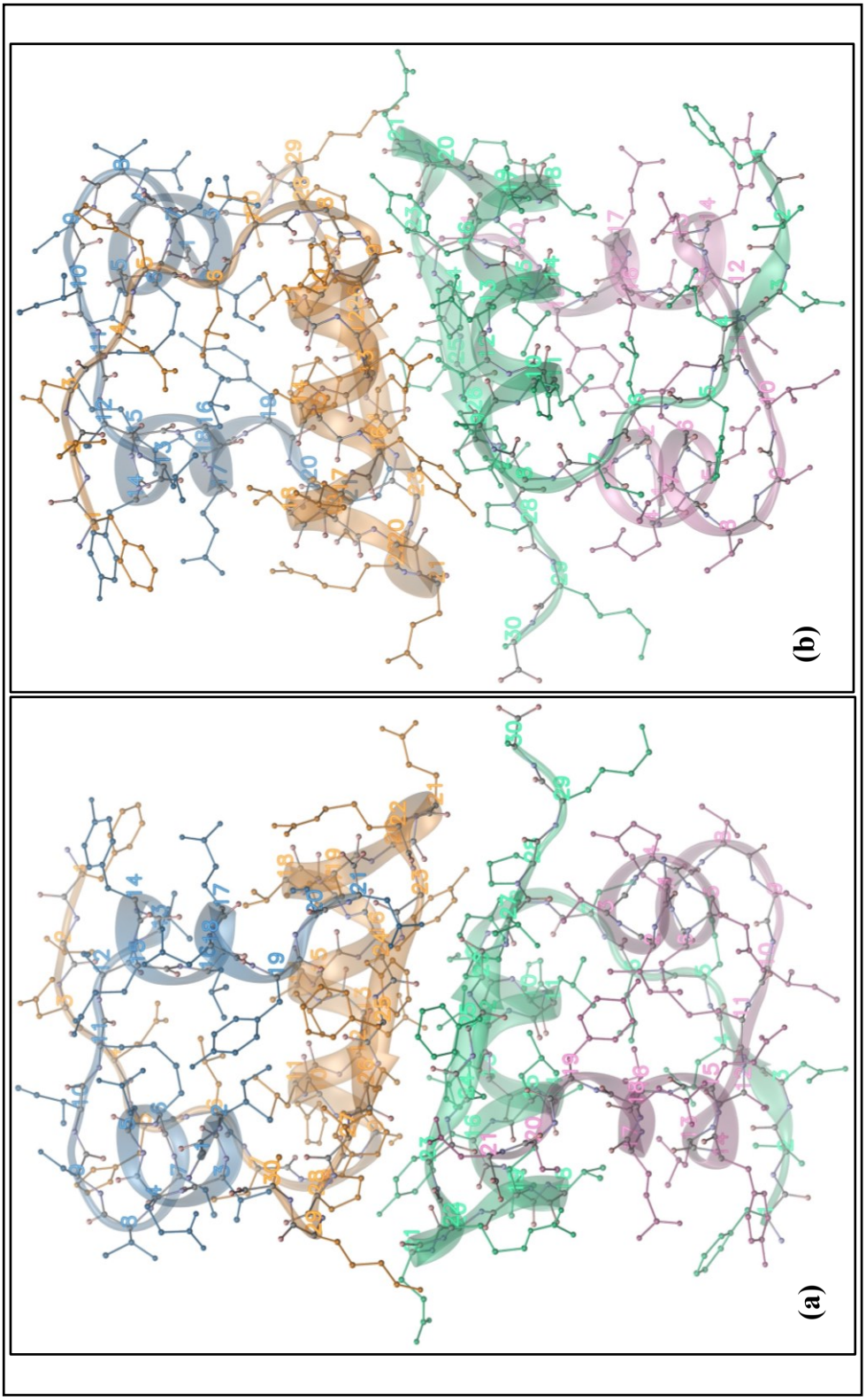


***Figure 4.3**: Structure with chain-numbering for an asymmetric dimer unit (Bottom) **M2** (AC mauve, BC turquoise). (Top) **M1** (AC blue, BC orange). **(a)** "Front", **(b)** "back" (front rotated sideways 180º). The ACs, BCs and SCs being transparent, BB "metallic pastel". The structure having only A configuration of those residues having alternate SC configurations in original PDB structure (**M1** R^{B22}, K^{B29}; **M2** Q^{B4}, V^{B12}, E^{B21}, R^{B22}, T^{B27}). Structure from PDB 4INS (biological assembly 7).*

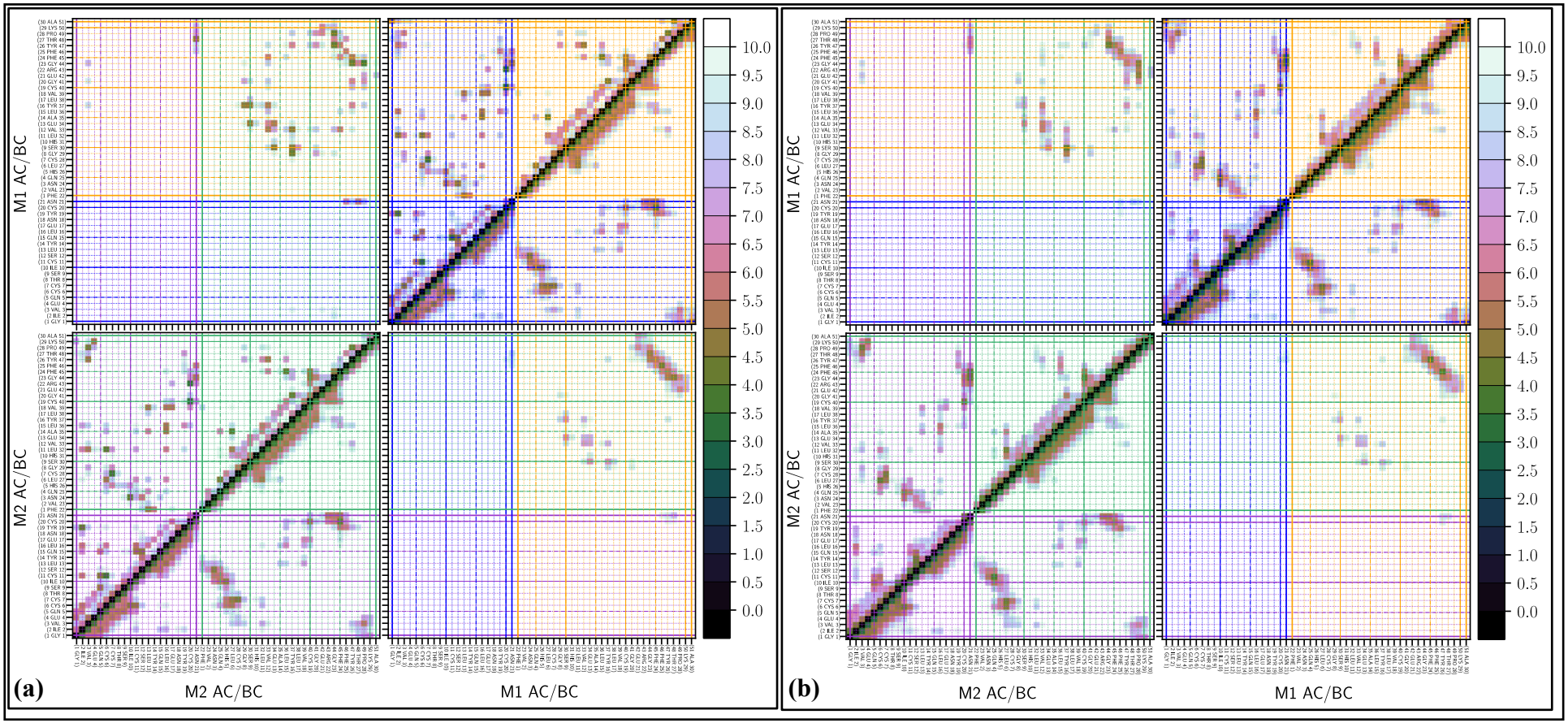


Figure 4.4*: Residue distances within 10 Å, matrix, of an asymmetric dimer. Diagonal as reference 0 Å (black), above 10 Å in white.* ***(a)*** *Upper left is SC to SC geometric centre distances. Lower right is CA to CA-atom distances.* ***(b)*** *Upper left is SC to MC geometric centre distances. Lower right is MC to MC geometric centre distances. Distances divided in 0.5 steps, c.f. most of the CA-atom distances of adjacent residues are between 3.5-4.0 Å. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. grid-lines for AC (1-21) in purple, blue and BC (22-51) turquoise, orange for* ***M2****,* ***M1*** *respectively. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Calculated from structure in PDB 4INS (biological assembly 7).*

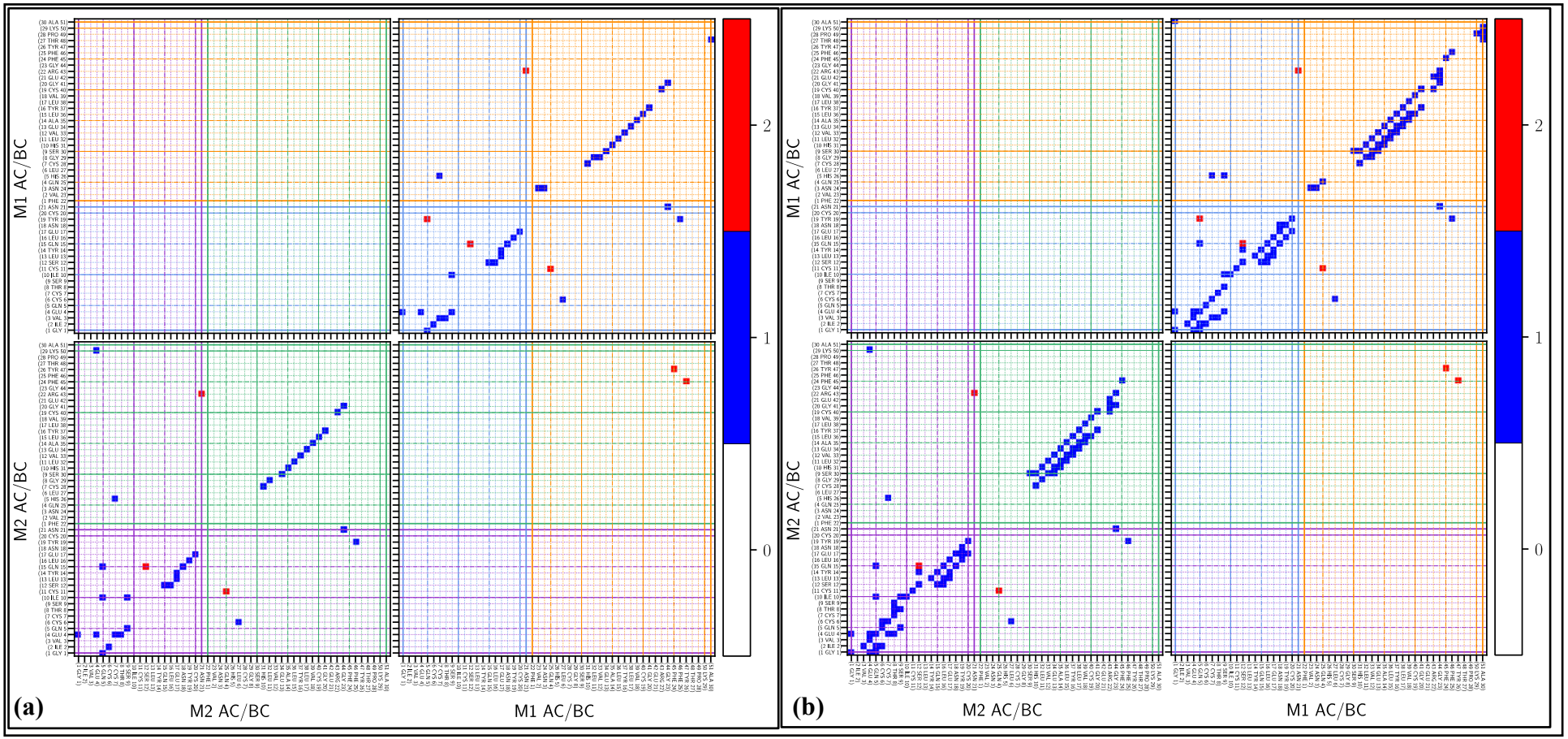


***Figure 4.5**: Hydrogen Bonds between residue-moieties of **M12**. The matrices of HBs between residues calculated with, $|\boldsymbol{r}_{AD}| < 3.5$ Å, and **(a)** $\varphi < 60°$, 87 HBs, **(b)** $\varphi < 90°$, 190 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. grid-lines for AC (1-21) in purple, blue and BC (22-51) turquoise, orange for M2, M1 respectively. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Calculated from structure in PDB 4INS (biological assembly 7).*

4.2 An analysis and view of insulin bound to receptor fragments

This section is merely an objective comparison of insulin receptor complexes that have been reported as being near representative of receptor binding, during the later years. The latest structure by Weis *et al.* [2], of insulin in its high affinity binding state, appears the one of most high precision and meaningful to intricately chart its geometry. However, for all reported structures, we present analogous information, to assist researchers in understanding each respective structure, however considering their largely different resolutions (error ranges). Assisting in clarifying not only the similarities but also the differences, quantifying the general geometry and HBs that are within 10 Å of the referred structures. There are written a complementary review about insulin binding to its receptor in §2.5 (where the same naming and nomenclature are used as here); in particular, the insulin bound complexes presented in this section are the same as introduced in §2.5.3.5.

4.2.1 A comparison of insulin contiguous residues in ectodomain fragments

Various analysis tools developed and described in §3.3, are directly applicable on the insulin contiguous residues, in addition for the bound insulin. Here, at first, the method of obtaining these insulin contiguous structures are described. Afterwards comparing the whole IR ectodomain fragments next to each other. Following, is then a comparison of the resulting insulin contiguous residue surfaces, having a closer look at some properties and how they overlap. Following that is an in depth look of the structure by Weis *et al.*, to provide an analytical view overview representation and example residue-profiles.

4.2.1.1 Method

The insulin contiguous receptor residues were excerpted from PDB entries: 3W11, 4OGA, 6CE9, 6CEB, 6CE7 and 6HN5. For each of these IR fragments the residues that had any non-hydrogen atoms within 10 Å of the insulin bound, were then included in a PDB file; with the residues order being reorganised within the file. In addition, the change, “ARG 717 to GLY”, were done, due to having missing SC atoms in PDB entries of e.g. 6CE9. Then a conversion could be made from the numbering in PDB format to a continuous numbering in GRO format (using gmx pdb2gmx which includes hydrogens). Moreover, between the different excerpted chains, the NT amino-groups were protonated, and the CT carboxy-groups deprotonated and the other residues protonated at pH 7.4 (standard

protonation state, see §3.1.3.5). The resulting GRO file format were used for calculation of the analytical properties DAs, HBs and residue distances.

Nomenclature

To distinguish insulin and its contiguous residues when bound to IR-fragments (CF), i.e. that are within 10 Å to insulin (and including insulin), a nomenclature is constructed. For example, for $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{6HN5}}$, the subfix refers to insulin chain A and B and the superfix, e.g. 6HN5, the PDB entry. The receptor monomer 1 have red colour, and monomer 2 have blue colour and are marked with an asterisk. Note that the designation of 1'st and 2'nd monomer is arbitrarily chosen, for all structures, and is merely a way of comparing these structures.

Structure before and after the residue excerption and file conversion

- The 3.9 Å resolution structure of PDB entry 3W11 (as in 3w11.pdb), referred to here as μIRa, in whole are including: insulin (chain A of 21 residues and chain B of 15 residues, B1-6 and B22-30 missing); monoclonal antibodies fab 83-7 fragment - "heavy" (chain C, 118 residues) and -"light" chain (chain D, 114 residues); parts of L1*-C* (chain E, 288 residues); αCT peptide (chain F, 11 residues). The resultant excerpted structure, $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{3W11}}$, contains in addition to insulin, residues from chain E (20 residues), chain F (11 residues).
- The 3.5 Å resolution structure of PDB entry 4OGA (as in 4oga.pdb), referred to here as μIRb, in a whole are including: insulin (chain A of 21 residues and chain B of 21 residues, B1-6 and B28-30 missing); monoclonal antibodies fab 83-7 fragment - "heavy" (chain C, 118 residues) and -"light" chain (chain D, 114 residues); parts of L1*-C* (chain E, 288 residues); αCT peptide (chain F, 15 residues). The resultant excerpted structure, $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{4OGA}}$, contains in addition to insulin, residues from chain E (35 residues) and chain F (15 residues).
- The 4.3 Å resolution structure of PDB entry 6CE9 (as in 6ce9.pdb), referred to here as "sIRa+2", as a whole are including: two respective insulin bound at different sites (chains K, N of 21 residues and chain L, O of 30 residues); Insulin receptor ectodomain–monomer 1 L1-C-L2-F1 (chain A, 562 residues) and monomer 2 L1*-C*-L2*-F1* (chain B, 562 residues); αCT (chain P 30 residues), αCT* (chain M 30 residues). The first resultant excerpted structure, $\mathbf{CF}_{(\mathbf{K,L})}^{\mathbf{6CE9}}$, contains in addition to insulin, residues from chain A (23 residues), chain B (43 residues) and chain M (19 residues). The second

resultant excerpted structure, $\mathbf{CF}_{(\mathbf{N,O})}^{\mathbf{6CE9}}$, contains in addition to insulin, residues from chain A (44 residues), chain B (23 residues) and chain P (19 residues).

- The 4.7 Å resolution structure of PDB entry 6CEB (as in 6ceb.pdb), referred to here as "sIRb+2", in whole are including: two respective insulin bound at different sites (chains K, N of 21 residues and chain L, O of 30 residues); Insulin receptor ectodomain– monomer 1 L1-C-L2-F1-F2 (chain A, 682 residues) and monomer 2 L1*-C*-L2*-F1* (chain B, 562 residues); αCT (chain M 30 residues), αCT* (chain P 30 residues). The first resultant excerpted structure, $\mathbf{CF}_{(\mathbf{K,L})}^{\mathbf{6CEB}}$, contains in addition to insulin, residues from chain A (23 residues), chain B (42 residues) and chain M (19 residues). The second resultant excerpted structure, $\mathbf{CF}_{(\mathbf{N,O})}^{\mathbf{6CEB}}$, contains in addition to insulin, residues from chain A (43 residues), chain B (23 residues) and chain P (19 residues).
- The 7.4 Å resolution structure of PDB entry 6CE7 (as in 6ce7.pdb), referred to here as "sIR+1", includes: one bound insulin (chains N of 21 residues and chain O of 30 residues); Insulin receptor ectodomain monomer 1 L1-C-L2-F1 (chain A, 528 residues) and monomer 2 L1*-C*-L2*-F1*-F2* (chain B, 659 residues); αCT* peptide (chain P, 30 residues). The resultant excerpted structure, $\mathbf{CF}_{(\mathbf{N,O})}^{\mathbf{6CE7}}$, contains in addition to insulin, residues from chain A (42 residues), chain B (21 residues), chain P (18 residues).
- The 3.2 Å resolution structure of PDB entry 6HN5, upper part (as in 6hn5.pdb), referred to here as IRΔβ-zipInsFvU, includes: one bound insulin (chains A of 21 residues and chain B of 26 residues); Insulin receptor ectodomain monomer 1 L2-F1 & αCT (chain F, 323 residues) and monomer 2 L1*-C*-L2*-F1* (chain E, 585 residues). The resultant excerpted structure, $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{6HN5}}$, contains in addition to insulin, residues from chain F (34 residues) and chain E (42 residues).
- The 4.2 Å resolution structure of PDB entry 6HN4, lower part (as in 6hn4.pdb), referred to here as IRΔβ-zipInsFvL, includes: none bound insulin; Insulin receptor ectodomain monomer 1 L1-C, F2-F3 (chain F, 498 residues) and monomer 2 F2*-F3* (chain E, 202 residues).With IRΔβ-zipInsFv referring to both upper and lower part. The Fv 83-7 epitope on domain C, were not included in PDB, moreover original authors assuming it to have effectively none disturbance on insulin binding to the holo-IR. Some unmodelled and disordered domains were ID, ID*, αCT*.

4.2.1.2 B-factors of all insulin residues contiguous to insulin

The B-factors of the IR-fragments considered here were included in respective PDB entry; however, it appears that little or no mention about their relevance in interpretation of the

data, were stated by the original authors. Nevertheless, for the insulin adjacent residues, average residue-wise B-factors, were recalculated and depicted in Figure S4.8, Figure S4.9; these are exceedingly large, on average on the order of hundreds of $Å^2$; somewhat proportional to the respective reported resolution. Notwithstanding, it has been suggested to me, that more or less, these are relatively good quality structures; not correlating with the high B-factors. Albeit, the high peaks of B-factors may be a cause for caution and concern in respective structural model.

Besides, in general crystallography, drawbacks of exceedingly large B-factors have been examined [264]; in addition to if larger than ~100 $Å^2$, will yield negligible contribution to the calculation of structure factors. Nevertheless, it has been described by Wlodamer *et al*. [265], that some other published cryo-EM structures, have pushed the available resolution to its limits; on the expense of the validity of the resulting structures and the originating density-map; moreover some with unphysical B-factors.

Nevertheless, the end-user of a structure, often not a crystallographer, should be wary to draw any conclusive statements of absolute geometry from any low-resolution structure. Though of course, random structural configurations may be expected even at optimal resolution. However in particular here to be wary in regards to the IR-fragments, whose structural aspects have widely varying certainty and may be more or less overinterpreted. This in proportion to their respectively varying resolution (and B-factors), which may indicate varying differences in flexibility and disorder. Hence, the geometrical analysis presented in this chapter concerns only the presented structures as presented in respective PDB with original reports. However, here considered the highest resolution (~3.2 Å) structure, $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{6HN5}}$, have B-factors of mean ~80 $Å^2$ (c.f. ~20 $Å^2$ for **M12**), still a large number. As generally understood [235], resolutions of around 1-1.5 Å (as for **M12**) can be considered enough to discern even atomic identities of amino-acids, however for resolution 3 Å or larger, will at most discern the contours of the protein chain. Hence the atomic identities and residue orientation of IR-fragments, has been inferred, to more or less an extent, by the procedure implemented by the original authors.

4.2.1.3 The complete ectodomain IR-fragments with bound insulin

At first here we compare the whole IR-fragments, by showing the insulin molecule in a similar orientation when bound (see Figure 4.6). This reveals an interesting comparison and may be a serendipitous view of how insulin binds its receptor. The apparent similarity of

IR domain structures in “sIRa+2”, “sIRb+2”, “sIR+1” and IRΔβ-zipInsFv, appears to may be partly due to them being modelled from PDB 4ZXB. As the first two structures (μIRa and μIRb) has features of a binding site, remarkably a resembling binding motif is found in the other structures, as is depicted here. Interestingly that in the two double-bound structures of “sIRa+2” and “sIRb+2”, there appears to be a “T”-shape conformation, with varying resolved parts of the F2 domains. One may make the conjecture, that they are resembling the coarse visualization of a “T”-shaped IR by Gutmann *et al.* [145, 165, 266, 267].

What is further interesting, is that the singly bound structures (“sIR+1” and IRΔβ-zipInsFv), shows more of a “Γ”-shape, and appears largely matching. Differing to the greatest extent is the orientation of unbound L1-C, I assume it is explained partly by “sIR+1” being untethered, whereas IRΔβ-zipInsFv is tethered by a 33-residue GCN4 leucine zipper segment at the CT ends of its β-subunits. A “/Γ”-shape structure of the activated insulin receptor, are apparently visualized by Gutmann *et al.* [165]; who however surmised it to be a subpart of the “T”-shape structure, merely representing another view-angle. As inferred by Weis *et al.* [2], IRΔβ-zipInsFv, is a representation of the “signalling conformation” of the receptor ectodomain; (however a conjectured alternative “/Γ”-shape provided in §2.5.2.2).

Interestingly, Gutmann *et al.* visualized, via single particle electron microscopy, glycosylated full-length IRs in lipid nm size discs. Where the apo-IR of “Λ”-shape were incubated 1h with saturating concentrations of insulin (1 μM) which showed 26% “/Γ”-shape (designated “II” in journal) and ~72% “T”-shape; additional insulin yielded close to 100% of the “T”-shape. Albeit, their experiments at nearer to physiological insulin concentration (~0.8 -12 nM), implied to them that 1 insulin bound is enough to induce a “T”-shape. Nonetheless, as a deduction, from my perspective, it may be that the “/Γ”-shape seen by Gutmann *et al.*, are a structure with only 1 insulin bound (or possibly with a second insulin bound to either site 1 or 2* of the alternate binding site), that represents an insulin being bound with high affinity. Other EM studies have shown either “T”- or “Y”-shaped ectodomain structures (reviewed in [18]), appear to exist with saturating insulin concentration or time of incubation [266, 267]. One may notice a somewhat less density at the top and middle of “sIRa+2”, “sIRb+2” which may partly explain a “Y”-shape appearance in EM studies.

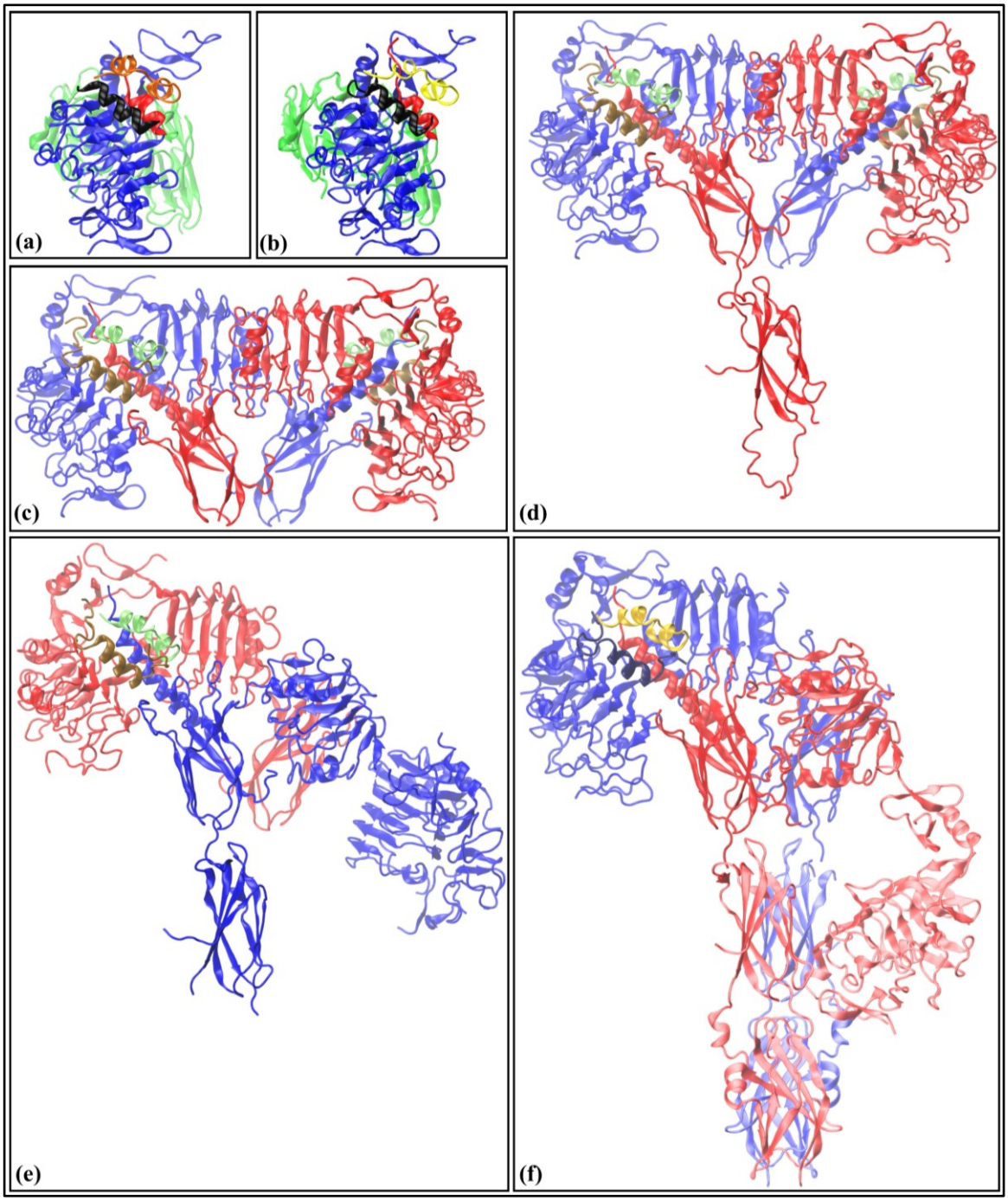


***Figure 4.6**: Comparative view of reported insulin bound IR-fragments. **(a)** µIRa, PDB 3W11: chain A(orange-red), B(black), E (blue), C and D(green). **(b)** µIRb, PDB 4OGA; chain A(yellow), B(black), E (blue), C and D(green). **(c)** "sIRa+2", PDB 6CE9; chain K & N(lime-green), L & O(brown), A and M (red), B and P(blue), chain K, L at left. **(d)** "sIRb+2", PDB 6CEB; chain K & N(lime-green), L & O(brown), A and M (red), B and P(blue), chain K, L at left. **(e)** "sIR+1", PDB 6CE7; chain N(lime-green), O(brown), A (red), B and P(blue). **(f)** IRΔβ-zipInsFvU, PDB 6HN5 darker upper part, IR Δβ -zipInsFvL, PDB 6HN4 lighter lower part; chain A(gold-yellow), B(midnight-blue), E (blue), F(red). All structures are oriented as to depict the same binding of insulin, and with fibronectin domains pointing downwards as to an imaginary cell membrane.*

4.2.1.4 Overlap of the insulin contiguous structures

For any general structure reported in a PDB entry, the intricate details to allow an understanding or comparison of other structures are not straightforwardly shown; whereof it is one reason why in this section it is included elaborate and encompassing tables and graphs relating to each structure. Just looking at the structures they may look very similar but in essence very much differ in moiety orientations; whereby it is meant also as complementary information to the original journals and PDB entries with structures. Here then, it is focused upon the insulin contiguous residues of each IR-fragment, as an overlap of each other (see Figure 4.7), with a particular facet shown (rotated from that of Figure 4.6); moreover the respective HBs are in Table 4.2. Obvious is that all the structures are fully complete atom structural solutions, despite that the resolutions vary from 3.2 to 7.4 Å; hence there are more and less large error ranges for the atomic positions. The contiguous fragments differ substantially in what residues are within 10 Å, however, there are some more or less common residues and HBs in all of them.

The highest resolution structures $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{3W11}}$, $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{4OGA}}$, $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{6HN5}}$ are not identical but has much similar structure, arising from similar resolution and are in that regard more comparable; in addition to sharing many authors in respective publications [2, 69, 71]. There is some similarity to the other more lower resolution structures, partly due to them being implied of being based partly upon IR domains (PDB 4ZXB), site 1 (PDB 3W11), in addition to the insulin molecule (PDB 1ZNI) [70]. Indeed, it is seen that only $\mathbf{CF}_{(\mathbf{K,L})}^{\mathbf{6CE9}}$, $\mathbf{CF}_{(\mathbf{N,O})}^{\mathbf{6CE9}}$, $\mathbf{CF}_{(\mathbf{K,L})}^{\mathbf{6CEB}}$ and $\mathbf{CF}_{(\mathbf{N,O})}^{\mathbf{6CEB}}$ have very near identical structure, obviously they were given the same solution, differing to some extent to $\mathbf{CF}_{(\mathbf{N,O})}^{\mathbf{6CE7}}$. The comparison made, highlights an important point, that a published structure even if reported as an atomic structure, can differ substantially in certainty to a biologically representative structure. Nonetheless, even with the more or less varying certainty of these structures, they still represent an important hallmark of insulin structural biology. Defining at least an approximate binding region where insulin cross-link to both receptor monomers.

Figure 4.7*: Overlap of insulin contiguous binding residues of various IR-fragments, with varying resolution. Superimposition on α-helical regions A1-7, A13-19, B11-17, 705-714. Backbone traced with 'Ribbon' light colour, CA-atoms as glass spheres, residue-atoms without hydrogens. Shown are the whole residues having at least one non-hydrogen atom within 10 Å from any part of insulin, respectively for each structure (with no hydrogens included):*

(a) $\boldsymbol{CF^{3W11}_{(A,B)}}$ *(dark-orange) &* $\boldsymbol{CF^{4OGA}_{(A,B)}}$ *(yellow);*

(b) $\boldsymbol{CF^{6CE9}_{(K,L)}}$ *(ochre-brown);* $\boldsymbol{CF^{6CE9}_{(N,O)}}$ *(purple);* $\boldsymbol{CF^{6CEB}_{(K,L)}}$ *(red);* $\boldsymbol{CF^{6CEB}_{(N,O)}}$ *(white);*

(c) $\boldsymbol{CF^{6CE9}_{(K,L)}}$ *(ochre-brown);* $\boldsymbol{CF^{6CE7}_{(N,O)}}$ *(lime-green).*

(d) $\boldsymbol{CF^{6HN5}_{(K,L)}}$ *(midnight-blue) &* $\boldsymbol{CF^{4OGA}_{(A,B)}}$ *(yellow).*

(e) $\boldsymbol{CF^{6HN5}_{(K,L)}}$ *(midnight-blue) &* $\boldsymbol{CF^{6CE9}_{(K,L)}}$ *(ochre-brown).*

(f) $\boldsymbol{CF^{6HN5}_{(K,L)}}$ *(midnight-blue) &* $\boldsymbol{CF^{6CE7}_{(N,O)}}$ *(lime-green).*

***Table 4.2**: Hydrogen Bonds for insulin bound to IR fragments, with varying resolution. The HBs are calculated with criteria: $r_{AD} < 3.5$ Å, & $\varphi < 30°$(shaded in gray), & $\varphi < 60°$. Stated with the total number of lower and medium angle HBs: Column **1**. $\boldsymbol{CF^{3W11}_{(A,B)}}$ with 19 &50 HBs ; **2**. $\boldsymbol{CF^{4OGA}_{(A,B)}}$ with 27 & 66 HBs; **3**. $\boldsymbol{CF^{6CE9}_{(K,L)}}$ with 24 & 59 HBs; **4**. $\boldsymbol{CF^{6CE9}_{(N,O)}}$ with 25 & 58 HBs: **5**. $\boldsymbol{CF^{6CEB}_{(K,L)}}$ with 28 & 58 HBs ; **6**. $\boldsymbol{CF^{6CEB}_{(N,O)}}$ with 26 & 59 HBs: 7. $\boldsymbol{CF^{6CE7}_{(N,O)}}$ with 21 & 63 HBs, **8**. $\boldsymbol{CF^{6HN5}_{(A,B)}}$ with 25 & 70 HBs. Here the receptor residues have numbering as in respective PDB entries, which respectively is apparently congruent with conventional numbering of the insulin receptor sequence. The HBs due to NT and CT specification of excerpted chains (numbering continuous) in brackets [].*

Donor (H)	Acceptor	1	2	3	4	5	6	7	8	A19 Y(HN)	A15 Q(O)								B23 G(HN)	B19 C(O)								✓
Insulin chain A(or K or N) & chain B(or L or O)										A19 Y(HN)	A16 L(O)	✓		✓	✓	✓	✓			B23 G(HN)	B20 G(O)		✓						
A1 G(H3)	A4 E(OE1)			✓	✓			✓		A20 C(HN)	A16 L(O)	✓		✓	✓	✓	✓			B24 F(HN)	B22 R(O)			✓	✓	✓	✓		
A4 E(HN)	A1 G(O)	✓						✓		A20 C(HN)	A17 E(O)	✓	✓			✓	✓	✓	✓	A21 N(HN)	B23 G(O)		✓						
A4 E(HN)	A4 E(OE1)		✓	✓	✓	✓	✓	✓		A20 C(HN)	A18 N(O)		✓							B5 H(HD1)	A9 S(O)						✓		
A5 Q(HN)	A1 G(O)	✓	✓							A21 N(HN)	A19 Y(O)							✓		B6 L(HN)	A6 C(O)			✓	✓	✓	✓		
A5 Q(HN)	A2 I(O)		✓	✓	✓	✓	✓	✓	✓	B4 Q(HN)	B2 V(O)						✓			[B7 C(H3)	A7 C(SG)]	✓							
A6 C(HN)	A2 I(O)	✓	✓	✓	✓	✓	✓	✓	✓	B7 C(HN)	B5 H(O)							✓		B25 F(HN)	A19 Y(O)								✓
A6 C(HN)	A3 V(O)		✓							B10 H(HN)	B7 C(O)	✓								L1, C, L2									
A7 C(HN)	A2 I(O)							✓		B11 L(HN)	B7 C(O)		✓	✓	✓	✓	✓		✓	11 M(HN)	32 H(O)		✓					✓	✓
A7 C(HN)	A3 V(O)		✓	✓	✓	✓	✓			B11 L(HN)	B8 G(O)	✓								13 I(HN)	34 E(O)		✓	✓	✓	✓	✓	✓	✓
A8 T(HN)	A3 V(O)		✓	✓	✓	✓	✓	✓	✓	B12 V(HN)	B8 G(O)			✓	✓	✓	✓	✓		14 R(HE)	12 D(OD1)	✓							
A9 S(HN)	A4 E(O)	✓						✓		B12 V(HN)	B9 S(O)	✓	✓						✓	14 R(HE)	12 D(OD1/2)	✓	✓						
A9 S(HN)	A5 Q(O)								✓	B13 E(HN)	B9 S(O)	✓	✓	✓	✓	✓	✓	✓	✓	14 R(HH21)	12 D(OD1/2)	✓	✓						✓
A10 I(HN)	A9 S(OG)		✓	✓	✓	✓	✓	✓		B14 A(HN)	B10 H(O)	✓	✓	✓	✓	✓	✓		✓	15 N(HN)	36 L(O)		✓						
A12 S(HN)	A11 C(SG)	✓								B15 L(HN)	B11 L(O)	✓	✓	✓	✓	✓	✓	✓	✓	15 N(HN)	37 L(O)	✓	✓	✓	✓	✓	✓		✓
A12 S(HN)	A15 Q(OE1)	✓	✓	✓	✓	✓	✓			B15 L(HN)	B12 V(O)								✓	16 N(HN)	14 R(O)	✓	✓	✓	✓	✓	✓	✓	
A15 Q(HN)	A12 S(OG)			✓	✓	✓	✓	✓		B16 Y(HN)	B12 V(O)	✓	✓	✓	✓	✓	✓	✓	✓	18 T(HN)	16 N(O)						✓		
A15 Q(HN)	A12 S(O)	✓								B17 L(HN)	B13 E(O)	✓	✓	✓	✓	✓	✓	✓		19 R(HH1/22)	12 D(O)			✓	✓	✓	✓	✓	✓
A16 L(HN)	A12 S(O)	✓	✓	✓	✓	✓	✓	✓	✓	B18 V(HN)	B14 A(O)	✓	✓	✓	✓	✓	✓	✓		19 R(HE)	12 D(O)								✓
A17 E(HN)	A13 L(O)	✓						✓	✓	B18 V(HN)	B15 L(O)								✓	20 L(HN)	17 L(O)								✓
A17 E(HN)	A14 Y(O)		✓	✓	✓	✓	✓			B19 C(HN)	B15 L(O)	✓	✓	✓	✓	✓	✓	✓		32 H(HE2)	12 D(OD1)						✓		
A18 N(HD22)	A14 Y(OH)							✓		B19 C(HN)	B16 Y(O)								✓	32 H(HE2)	12 D(OD2)					✓	✓	✓	
A18 N(HN)	A14 Y(O)	✓	✓					✓	✓	B20 G(HN)	B16 Y(O)		✓	✓	✓	✓	✓			33 L(HN)	60 Y(O)		✓	✓	✓	✓	✓	✓	✓
A18 N(HD21)	A15 Q(O)								✓	B21 E(HN)	B19 C(O)	✓								34 E(HN)	11 M(O)		✓					✓	✓
A18 N(HN)	A15 Q(O)			✓	✓	✓	✓			B22 R(HH21)	B19 C(O)							✓		34 E(HE21)	12 D(OD1)		✓						
A19 Y(HH)	A2 I(N)	✓								B22 R(HN)	B22 R(NE)							✓		34 E(HN)	12 D(OD1)			✓	✓	✓	✓		

Table 4.2 *continued:*

Donor (H)	Acceptor	1	2	3	4	5	6	7	8
L1, C, L2 continued									
34 E(HE21)	60 Y(OH)								✓
35 I(HN)	62 L(O)			✓	✓	✓	✓		
36 L(HN)	13 I(O)		✓	✓	✓	✓	✓	✓	
37 L(HN)	64 F(O)	✓	✓	✓	✓	✓	✓	✓	✓
38 M(HN)	65 R(O)	✓	✓	✓	✓	✓	✓		✓
41 T(HN)	39 F(O)			✓	✓	✓			✓
60 Y(HH)	34 Q(NE2)			✓	✓	✓	✓	✓	✓
65 R(HH11)	67 Y(OH)	✓	✓						
65 R(HN)	96 F(O)	✓	✓					✓	✓
65 R(HE)	97 E(OE1/2)	✓	✓	✓	✓				
65 R(HH21)	97 E(OE2)	✓	✓	✓	✓	✓	✓		
65 R(HN)	97 E(O)	✓	✓						✓
67 Y(HN)	38 M(O)	✓	✓	✓	✓	✓	✓	✓	✓
97 E(HN)	120 E(O)		✓						
99 V(HN)	97 E(O)	✓	✓						
267 K(HZ3)	276 Q(NE2)							✓	
267 K(HZ3)	276 Q(OE1)							✓	
αCT									
[705 F(H2)	707 D(OD2)]		✓						
706 E(HN)	702 R(O)			✓	✓	✓	✓		✓
706 E(HN)	703 K(O)							✓	
707 D(HN)	703 K(O)			✓	✓	✓	✓		✓
707 D(HN)	704 T(O)							✓	
707 D(HN)	707 D(OD2)		✓						
708 Y(HN)	704 T(O)			✓	✓	✓	✓		✓
708 Y(HN)	705 F(O)			✓	✓				
708 Y(HN)	707 D(OD1)		✓						
709 L(HN)	705 F(O)	✓	✓						✓
710 H(HE2)	706 E(OE2)								✓
710 H(HN)	706 E(O)	✓	✓			✓	✓	✓	✓

711 N(HN)	707 D(O)	✓		✓	✓	✓	✓	✓	✓
711 N(HN)	708 Y(O)		✓						
712 V(HN)	708 Y(O)	✓	✓	✓	✓	✓	✓	✓	✓
712 V(HN)	709 L(O)			✓	✓				✓
713 V(HN)	709 L(O)	✓	✓	✓	✓	✓	✓		✓
713 V(HN)	710 H(O)							✓	
714 F(HN)	710 H(O)		✓	✓	✓				
717 G(HN)	714 F(O)							✓	
717 R(HE)	719 S(OT1)		✓						
720 R(HN)	719 S(OG)							✓	
F1(Fibronectin-1)									
498 R(HH21)	496 D(OD2)								✓
498 R(HE)	496 D(OD1)								✓
498 R(HN)	496 D(OD2)							✓	
499 D(HN)	496 D(O)			✓	✓	✓	✓	✓	✓
541 N(HN)	539 R(O)			✓	✓	✓	✓	✓	
545 S(HG1)	541 N(O)							✓	
547 N(HD21)	538 L(O)			✓	✓	✓	✓	✓	
575 E(HN)	573 S(OG)							✓	
575 E(HN)	573 S(O)								✓
Insulin & L1, C, L2									
B24 F(HN)	15 N(ND2)		✓						✓
B24 F(HN)	15 N(OD1)		✓						✓
B26 Y(HH)	19 R(NH1)								✓
B26 Y(HH)	19 R(NE)								✓
B26 Y(HH)	34 E(OE1)							✓	
15 N(HD22)	B24 F(N)								✓
40 K(HZ3)	B16 Y(OH)		✓						
40 K(HZ2)	B21 E(OE1)								✓
65 R(HH21)	B9 S(OG)								✓
65 R(HH12)	B13 E(OE2)		✓						

Insulin & ***α***CT									
A1 G(H3)	711 N(OD1)	✓							
A3 V(HN)	711 N(OD1)	✓	✓						✓
710 H(HE2)	B8 G(O)		✓						
711 N(HD22)	A4 E(OE1)		✓						
717 R(HN)	A18 N(O)								✓
720 R(HH22)	A17 E(O)							✓	
720 R(HH21)	A18 N(ND2)				✓		✓		
720 R(HH21)	A18 N(OD1)					✓	✓		
720 R(HE)	A18 N(ND2)					✓	✓		
720 R(HH21)	A20 C(O)							✓	
720 R(HE)	A21 N(OT2)							✓	
Insulin & F1									
B7 C(HG1)	496 D(OD1)			✓	✓	✓	✓		
B10 H(HE2)	540 S(N)			✓	✓	✓	✓		
498 R(HH12)	B7 C(O)							✓	
539 R(HH12)	B10 H(ND1)								✓
539 R(HH22)	B10 H(ND1)								✓
539 R(HH12)	B13 E(OE2)								✓
L1, C, L2 & ***α***CT									
14 R(HH12)	713 V(O)	✓	✓						
14 R(HH22)	713 V(O)	✓	✓						✓
121 K(HZ2)	706 E(OE1)			✓	✓	✓	✓		✓
121 K(HZ2)	706 E(OE2)			✓	✓	✓	✓		
L1, C, L2 & F1									
121 K(HZ1)	498 R(NH1)			✓					
αCT & F1									
703 K(HZ1/2)	496 D(OD1)					✓		✓	
703 K(HZ1/2)	496 D(OD2)			✓			✓	✓	
498 R(HH11)	706 E(OE1)								✓
498 R(HH12)	707 D(OD1)								✓
498 R(HH22)	707 D(OD1)								✓

4.2.1.5 Note on the analysis of the insulin contiguous residues in micro-insulin receptors and soluble ectodomain IR-fragments

Whereas some elaborate discussion of $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{3W11}}$, $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{4OGA}}$ like structures were described from another viewpoint [32, 69, 71]. However, the analysis graphs in the supplementary and are meant as complementary, providing more detail. There are some common authors to these former structures as to the one of $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{6HN5}}$; whose intricate structure were described to a lesser extent by Weis *et al.* [2]. The structures of $\mathbf{CF}_{(\mathbf{K,L})}^{\mathbf{6ce9}}$ (and the three near identical ones) and of $\mathbf{CF}_{(\mathbf{N,O})}^{\mathbf{6ce7}}$, have lesser resolution, hence are not discussed here in scrutinizing detail. However, as they show a similar binding pose as $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{6HN5}}$, their respective analysis graphs are also included in supplementary for comparison.

4.2.2 Conformational analytical overview of $\mathbf{CF}_{(\mathbf{A,B})}^{\mathbf{6HN5}}$

This ~3.2 Å resolution structure is credibly the most meaningful (to date) in relation to an insulin high affinity bound cross-link, whereof the internal contacts of insulin and its binding contacts may hence be the most precise in this regard, which we provide an overview of here. The structure are shown with chain-numbering (Figure 4.8), and with distances between residue-moieties (Figure 4.9), and with the sorted HBs between residue-moieties (Figure 4.10; the low to medium angle HBs in Table 4.2 column 8), in addition to the DAs (Figure 4.11). Even if it was measured almost native binding for this system, I suspect a few caveats with the structure, in regards to the cryo-EM method by which it were obtained [2]. Notwithstanding a structure with resolution of about 3.2 Å, can reveal more or less the contours of the protein chain, hence the atomic structure has been inferred to some degree and may be off to some extent. Note, that only some aspects of the structure are explicitly mentioned here, insulin has many intricate contacts. Hence the graphs and tables are meant as complementary, to assist the reader to infer meaningful structure. From the graphs (calculated from structure with added hydrogens) it is fairly evident which insulin residues that have non-hydrogen atoms within 5 Å (definition here of in close vicinity) to the receptor (given in Figure 2.2). Here listing only a few residue-profiles to gain familiarity with the graphs, not necessarily all aspects of those residues noticed. Some overlapping selection of residue-profiles are given for the other insulin structure models of **M12** (§4.1.2.4), $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ (§5.2.5) and $\mathbf{P}_{\mathbf{kpi}}^{\mathbf{MDm}}$ (§6.2.5); whose analogous analytical overview provides a direct comparison.

Example residues-profiles for $CF^{6HN5}_{(A,B)}$ in order A1-21 and B1-30

(A1 G): This residue having in vicinity (non-hydrogen atoms within 5 Å): I^{A2}, V^{A3}, E^{A4}, Q^{A5}, Y^{A19}, N^{711}, F^{714} and P^{716}. With the positively charged amino group (-NH3+) having stabilizing HB like interactions to N^{711}_{O}, F^{714}_{O} and Y^{A19}_{OH} ($r_{AD} < 5$ Å, $\varphi < 60°$); moreover additional longer distance polar-charged interactions with N^{711}_{SC}, E^{A4}_{SC} and Q^{A5}_{SC}. The HB "$Q^{A5}_{HN} \rightarrow G^{A1}_{O}$" is only present for a higher angle ($\varphi < 90°$), hence might be perturbed by the binding.

(A2 I): This residue having in vicinity: G^{A1}, V^{A3}, E^{A4}, Q^{A5}, C^{A6}, C^{A7}, Y^{A19}, L^{B11}, L^{B15}, H^{710} N^{711} and F^{714}. The SC have here hydrophobic interactions to V^{A3}_{SC}, Y^{A19}_{SC}, L^{B11}_{SC}, L^{B15}_{SC} (c.f. **M12**) and H^{710}_{SC}, F^{714}_{SC}, also in vicinity to the likely less polar buried disulphide bond A6-11 [268]. The HB "$C^{A6}_{HN} \rightarrow I^{A2}_{O}$" are still present at a low angle, in addition there is a distant HB "$I^{A2}_{HN} \rightarrow N^{711}_{OD1}$; $r_{AD} = 3.94$; $\varphi = 81°$" (the latter out of criteria hence not included in graphs or tables).

(A3 V): This residue having in vicinity: G^{A1}, I^{A2}, E^{A4}, Q^{A5}, C^{A6}, C^{A7}, T^{A8}, L^{B11}, D^{707}, H^{710} and N^{711}. The SC have hydrophobic contacts with I^{A2}_{SC} and L^{B11}_{SC}, moreover is close to the more polar/charged E^{A4}_{SC}, C^{A7}_{SC}, T^{A8}_{SC}, D^{707}_{SC}, H^{710}_{SC} and N^{711}_{SC}. There is even formed a site 1 HB "$V^{A3}_{HN} \rightarrow N^{711}_{OD1}$", moreover the HB "$T^{A8}_{HN} \rightarrow V^{A3}_{O}$" is present.

(A4 E): This residue having in vicinity: G^{A1}, I^{A2}, V^{A3}, Q^{A5}, C^{A6}, T^{A8}, S^{B9} and N^{711}. The charged SC is still pointing towards G^{A1} (c.f. $\langle r^{A1,A4}_{(SC,SC)} \rangle$ is 3.96 Å between 3.5-4.0 Å), with the saltbridge being offset "$G^{A1}_{H1} \rightarrow E^{A4}_{OE1}$; $r_{AD} = 5$ Å ; $\varphi = 93°$"; showing longer range interactions with Q^{A5}_{SC} and N^{711}_{SC} (c.f. $\langle r^{A4,711}_{(SC,SC)} \rangle$ is 6.1 Å between 6.0-6.5 Å). The SC to MC HB "$T^{A8}_{HG1} \rightarrow E^{A4}_{O}$; $r_{AD} = 2.8$ Å ; $\varphi = 80°$", is present.

(A7 C – B7 C): These residues have in vicinity: I^{A2}, V^{A3}, Q^{A5}, C^{A6}, T^{A8}, S^{A9}, H^{B5}, L^{B6}, G^{B8}, S^{B9}, H^{B10}, L^{B11}, P^{495}, D^{496}, F^{497} and R^{498}. This disulphide bond is more exposed to solvent in the free monomer, however here making contacts with site 2 residues. In particular, there is a contact with D^{496}_{SC} (c.f. $r^{(A7,496)}_{(SC,SC)}$ is 5.8 Å between 5.5-6.0 Å), and to D^{498}_{SC} (c.f. $r^{(A7,498)}_{(SC,SC)}$is 7.3 Å between 7.0-7.5 Å). There are several strong saltbridges that R^{498} forms "$R^{498}_{HH21} \rightarrow D^{496}_{OD2}$ & $R^{498}_{HE} \rightarrow D^{496}_{OD1}$", in addition to E^{706}, D^{707} (only D^{707} have no moieties between to C^{A7}-C^{B7} with minimum atom distance, $r^{(A7,707)}_{(SG,OD1)}$ of 5.13 Å).

(A18 N): This residue having in vicinity: Y^{A14}, Q^{A15}, L^{A16}, E^{A17}, Y^{A19}, C^{A20}, N^{A21}, F^{B25},

V^{715}, P^{716} and R^{717}. The SC fits into a cavity of partly of moieties: Y^{A14}_{SC}, Q^{A15}_{SC}, Y^{A19}_{SC}, P^{716}_{SC} and R^{717}_{SC}; in addition to other MC moieties, such as in HB "$N^{A18}_{HD21} \rightarrow Q^{A15}_{O}$". Some MC contacts include the HBs: "$N^{A18}_{HN} \rightarrow Y^{A14}_{O}$" & "$R^{717}_{HN} \rightarrow N^{A18}_{O}$".

(A19 Y): This residue having in vicinity: G^{A1}, I^{A2}, Q^{A5}, Q^{A15}, L^{A16}, E^{A17}, N^{A18}, C^{A20}, N^{A21}, L^{B15}, G^{B23}, F^{B24}, F^{B25}, F^{714}, V^{715}, P^{716} and R^{717}. The SC maintaining similar orientation as in **M12**, here closest to I^{A2}_{SC}, Q^{A5}_{SC}, N^{A18}_{SC}, L^{B15}_{SC} and to F^{714}_{SC}, P^{716}_{SC}, stabilizing the αCT. Apparent charged and polar interactions are to G^{A1}_{NH3+} and Q^{A5}_{SC}. Some MC contacts include the HB "$F^{B25}_{HN} \rightarrow Y^{A19}_{O}$".

(A21 N): This residue having in vicinity: E^{A17}, N^{A18}, Y^{A19}, C^{A20}, R^{B22}, G^{B23}, F^{B24}, F^{B25}, Y^{B26} and R^{717}. Where the residue is in a slightly different orientation than for **M12**, however the SC is also in contact with F^{B25}_{SC}. The carboxy group "N^{A21}_{COO-}" has longer distance HBs ($r_{AD} >$ 3.5 Å) with R^{717}_{SC} and R^{B22}_{SC} (that is a shorter distance saltbridge in **M12**), appearing as interactions of saltbridge like character. This latter observation appears to explain, at least partly, why substitutions of this amino-acid are tolerated but not deletions of this residue, for retaining negative cooperativity when binding to receptor. The "ARG → ALA" substitution of B22 reported with binding affinity increase of 405% [105], may be partly explained, by a resulting stronger saltbridge of N^{A21}_{COO-} to R^{717}_{SC}.

Next focusing upon a few example B-chain residues in vicinity to site 1 or 2, whose strands have significant differences to a T-state monomer. Interesting is the orientation of B3-5, which are seen to have a large DA shift of the MCs of residues, as compared with **M12**.

(B3 N): This residue having in vicinity: I^{A10}, Q^{B4}, H^{B5}, W^{493}, P^{494} and P^{495}. The SC are closest to MCs of W^{493}, P^{494} and P^{495} in the F1 domain. Within high angle criteria there are only a few MC to MC HBs to Q^{B4} and H^{B5}. This residue being also in the hexamer surface of **M12.**

(B4 Q): This residue having in vicinity: C^{A6}, S^{A9}, I^{A10}, C^{A11}, Q^{B4}, H^{B5} and L^{B6}. The SC has merely contact with I^{A10}_{SC}. In addition in comparison to **M12**, here the HB "$Q^{A4}_{HN} \rightarrow C^{A11}_{O}$" is more separated whereas the closer HB "$C^{A11}_{HN} \rightarrow Q^{B4}_{O}$; $r_{AD} = 4.67$ Å , $\varphi = 50°$" is only within angle criteria (not in table or graph).

(B5 H): This residue having in vicinity: C^{A6}, C^{A7}, T^{A8}, S^{A9}, I^{A10}, N^{B3}, Q^{B4}, L^{B6}, C^{B7}, H^{B10}

and P^{495}. The SC stays close to B7-B10 (as in **M12**), however here it is close to N^{B3} and also close to F1 domain being in touch with P^{495}. Moreover, the MC HB "$C^{B7}_{HN} \rightarrow H^{B5}_{O}$" is present.

(B6 L): This residue having in vicinity: C^{A6}, C^{A7}, T^{A16}, Q^{B4}, H^{B5}, C^{B7}, G^{B8}, H^{B10}, C^{B11}, A^{B14} and L^{B15}. Where the SC has retained intra-monomer contacts to L^{A16}_{SC}, L^{B11}_{SC} and A^{B14}_{SC}, albeit closer to H^{B10} than in **M12**. Present is the expected intra-chain HB "$L^{B6}_{HN} \rightarrow C^{A6}_{O}$; $r_{AD} < 3.14$ Å , $\varphi = 61.72°$". Notable is the MC DAs of H^{B5}, L^{B6} and C^{B7}, forming a curve and approximately maintained as in **M12**, albeit somewhat varying; partly explained by the strong interactions and/or shape complementarity of H^{B5}_{SC} and L^{B6}_{SC} with adjacent inter-insulin residues.

(B10 H): This residue having in vicinity H^{B5}, L^{B6}, C^{B7}, G^{B8}, S^{B9}, L^{B11}, V^{B12}, E^{B13}, A^{B14}, L^{B15}, F^{497} and R^{539}. The SC being closest to L^{B6}_{SC}, E^{B13}_{SC}, A^{B14}_{SC}, F^{497}_{SC} and R^{539}_{SC}. Notable are the HBs "$R^{539}_{(HH12\ \&\ HH22)} \rightarrow H^{B10}_{ND1}$".

(B12 V): This residue having in vicinity: G^{B8}, S^{B9}, H^{B10}, L^{B11}, E^{B13}, A^{B14}, L^{B15}, Y^{B16}, F^{B24}, L^{37}, F^{39}, F^{64}, R^{65}, H^{710} and F^{714}. The SC making non-polar contacts in the primary binding interface at: L^{37}_{SC}, F^{39}_{SC}, F^{64}_{SC}, R^{65}_{SC}, H^{710}_{SC} and F^{714}_{SC}.

(B13 E): This residue having in vicinity: S^{B9},H^{B10}, L^{B11}, V^{B12}, A^{B14}, L^{B15}, Y^{B16}, L^{B17}, R^{65} and R^{539}. The SC having apparent polar and/or charged interactions with S^{B9}_{SC}, H^{B10}_{SC}, R^{65}_{SC} and R^{539}_{SC}. Indeed, there is a strong saltbridge "$R^{539}_{HH12} \rightarrow E^{B13}_{OE2}$". Noting also as an example, that the geometric centre distance (7.3 Å) of E^{B13}_{SC} & R^{539}_{SC} is far off, due to their long SCs and opposite directions. To note is that Weis *et al.* [2] states that there is a poorly disordered segment of residues 541-545, not included in structure (sequence N-D-P-K-S [129]); more or less close to of E^{B13} & R^{539}, that may have some role in binding.

(B16 Y): This residue having in vicinity: L^{B11}, V^{B12}, E^{B13}, A^{B14}, L^{B15}, Y^{B16}, L^{B17}, V^{B18}, C^{B19}, G^{B20}, F^{B24}, F^{39} and K^{40}. The SC next to L^{B17}_{SC}, G^{B20}_{SC} and F^{B24}_{SC}, moreover an apparent pi-stacking [269] to F^{39}_{SC}. In addition, the SC are towards K^{40}_{SC} which in turn forms a HB to "$K^{40}_{HZ2} \rightarrow E^{B21}_{OE1}$" (c.f. in $\mathbf{CF}^{\mathbf{4OGA}}_{(\mathbf{A},\mathbf{B})}$ where the HB "$K^{40}_{HZ3} \rightarrow Y^{B16}_{OH}$" is rather present).

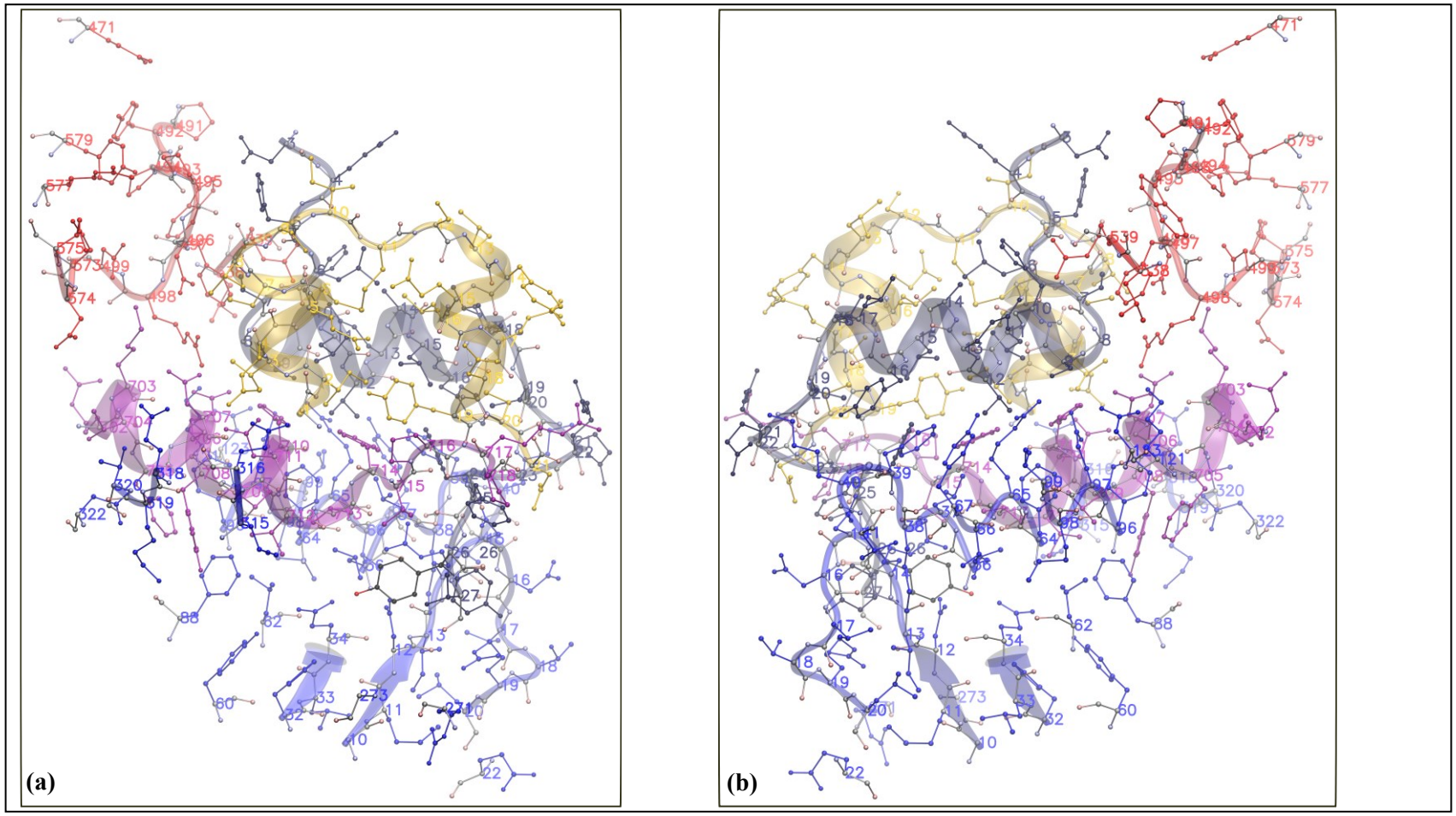


Figure 4.8**: Structure of IR fragment nearest to bound insulin, $\boldsymbol{CF}_{(A,B)}^{\boldsymbol{6HN5}}$. Shown are the whole residues having at least one non-hydrogen atom within 10 Å from any non-hydrogen atom of insulin (gold-yellow, chain A; midnight-blue, chain B), belonging to the domains of L1*, C*, L2* (blue, chain E) and the αCT (purple, chain F) and F1 (chain F). The B-conformation of Y_{SC}^{B26} shown in plain atom-colouring.* ***(a) *"Front".* ***(b)*** *"Back" (~180° sideway rotation).*

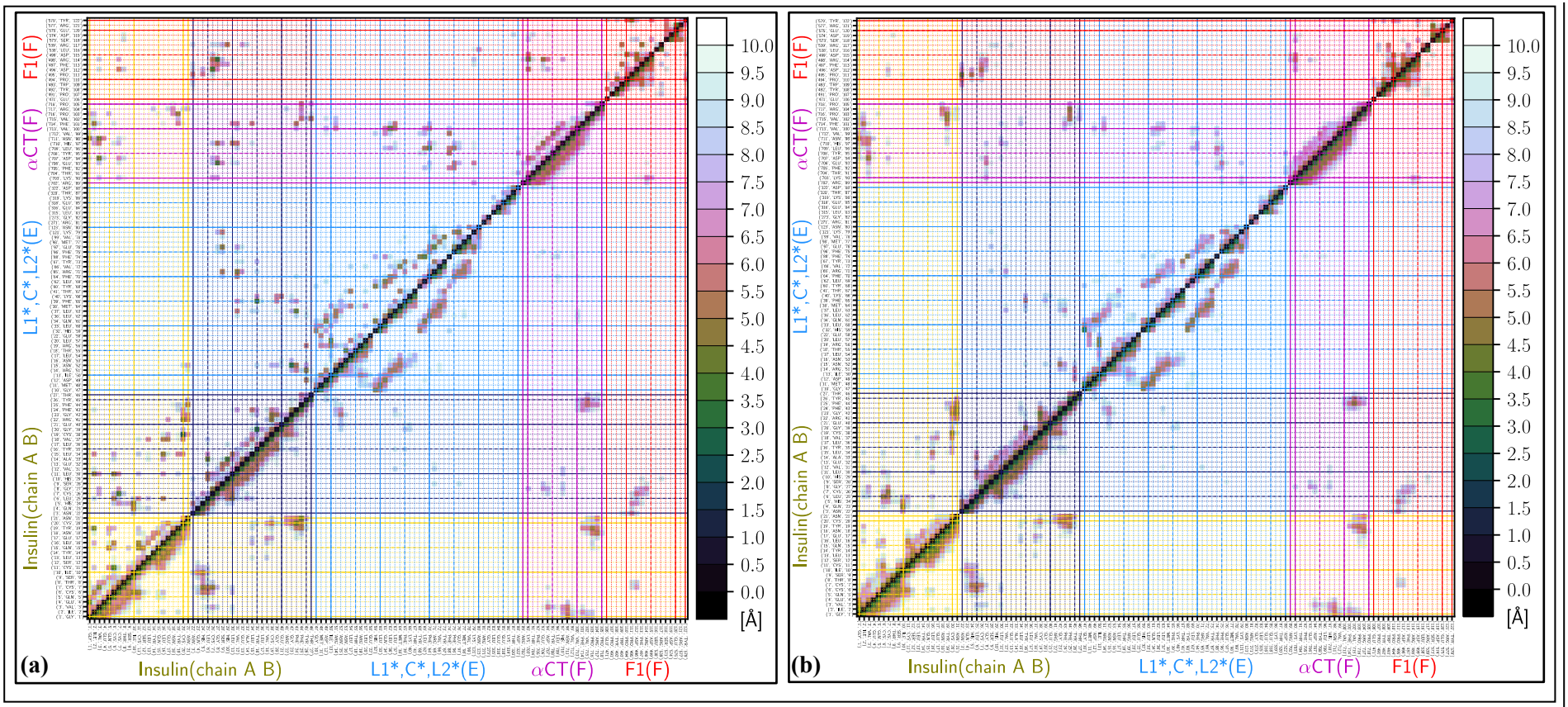


Figure 4.9: *Distances between residue-moieties of* $\boldsymbol{CF}^{\boldsymbol{6HN5}}_{\boldsymbol{(A,B)}}$. *The depictions are showing line-colouring for insulin, gold for chain A, midnight-blue for chain B. Structure with hydrogens, including the whole residues (from* PDB 6HN5*) having at least one non-hydrogen atom within 10 Å from any non-hydrogen atom of insulin, which are belonging to the domains of L1*, C*, L2* (blue, chain E) and the αCT (purple, chain F), F1(red, chain F). Distance matrices of geometric centres from following selections* ***(a)*** *SC to SC upper left, CA to CA atom lower right,* ***(b)*** *SC to MC at upper left, MC to MC lower right. Chains residues re-numbered sequentially 1-122 (nearest graph), and with actual residue name and number. Vector graphics are zoomable. Grid-lines are chain coloured, distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-122. Actual chain and residue naming as in* PDB 6HN5 *[2].*

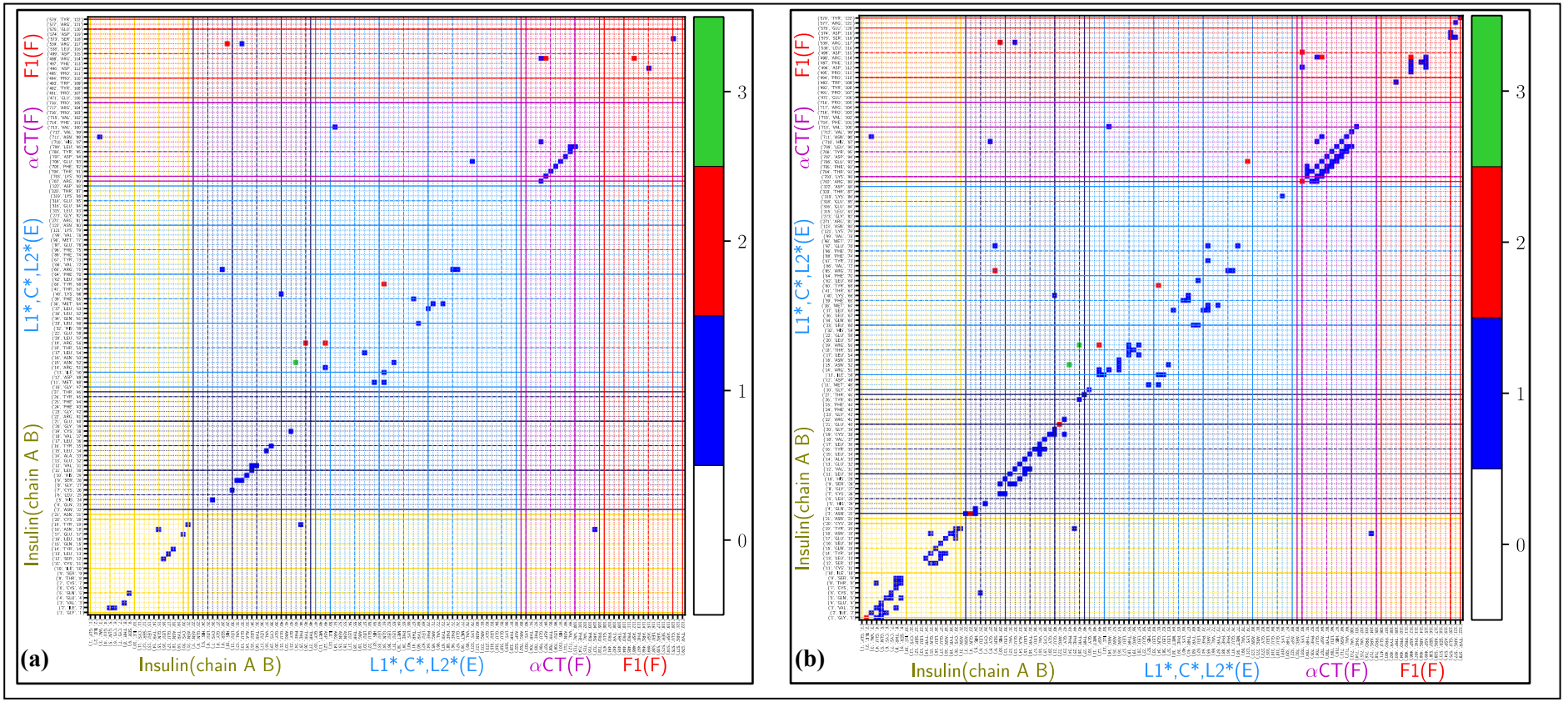


Figure 4.10: *Hydrogen Bonds between residue-moieties of* $\boldsymbol{CF}^{\boldsymbol{6HN5}}_{(\boldsymbol{A},\boldsymbol{B})}$. *The line-colouring depiction is showing for insulin, gold for chain A, midnight-blue for chain B. Structure with hydrogens, including the whole residues (from PDB 6HN5) having at least one non-hydrogen atom within 10 Å from any non-hydrogen atom of insulin, which are belonging to the domains of L1*, C*, L2* (blue, chain E) and the αCT (purple, chain F), F1(red, chain F). The matrices of HBs between residues calculated with:* $|\boldsymbol{r}_{AD}| < 3.5$ Å, *and* ***(a)*** $\varphi < 60°$, *70 HBs,* ***(b)*** $\varphi < 90°$, *188 HBs. The HBs are sorted as; SC to MC and SC to SC HBs in upper left; MC to MC HBs in lower right; diagonal any HBs in same residue. Chains residues renumbered sequentially 1-122 (nearest graph), and with actual residue name and number. Vector graphics are zoomable. Grid-lines are chain coloured, distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-122. Actual chain and residue-naming as in PDB 6HN5.*

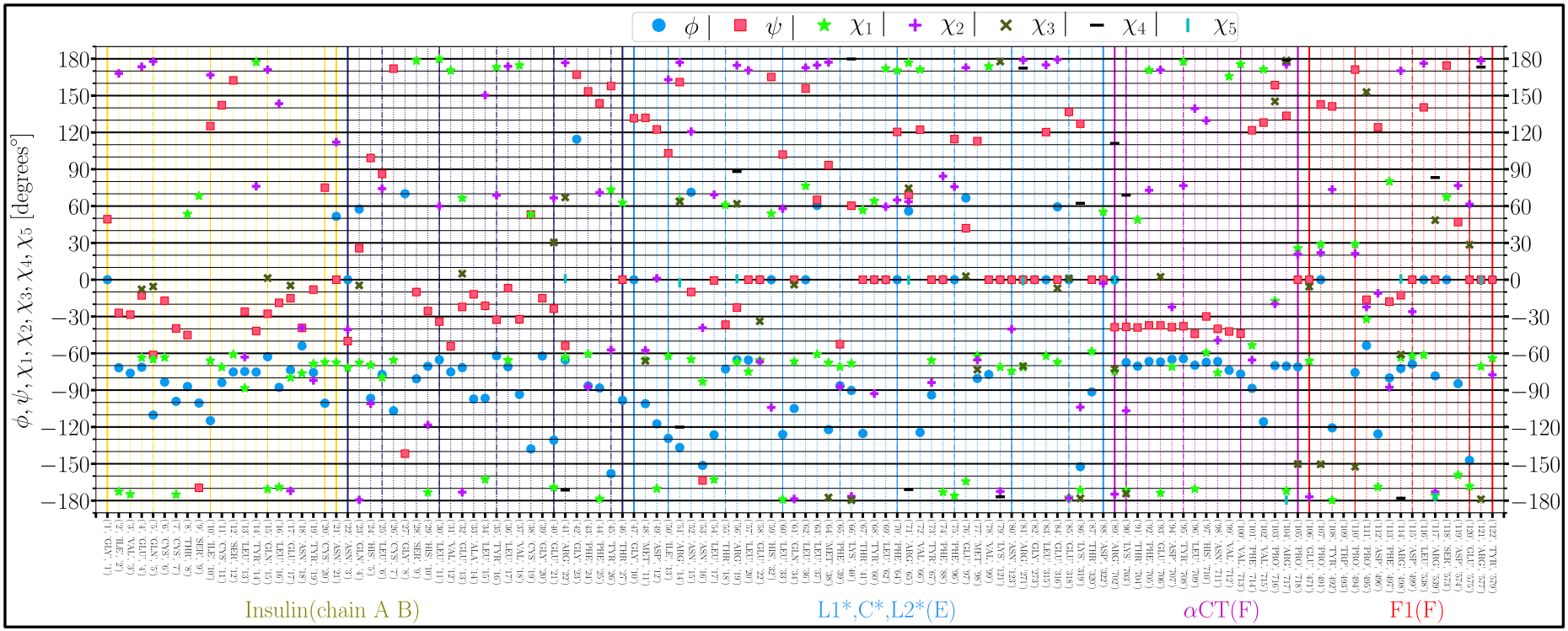


Figure 4.11: *Dihedral angles of residues in* $\boldsymbol{CF^{6HN5}_{(A,B)}}$. *The depiction is showing line-colouring insulin gold for chain A, midnight-blue for chain B. Structure with hydrogens, including the whole residues (from* PDB 6HN5*) having at least one non-hydrogen atom within 10 Å from any non-hydrogen atom of insulin, which are belonging to the domains of L1*, C*, L2* (blue, chain E) and the αCT (purple, chain F), F1(red, chain F). Chains residues renumbered sequentially 1-122 (nearest graph), and with actual residue name and number. Vector graphics are zoomable. Grid-lines are chain coloured, distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-122. Actual chain and residue-naming as in* PDB 6HN5.

4.3 Concluding statement

Science of different disciplines can complement each other, the skills to obtain the structural comparisons shown in this chapter are not common. The aspiration is that these examples, with an additional perspective, may be valuable for researchers studying the structure of biological systems such as insulin. The tools for this chapter and their assessment of PDB structures were developed as an example to unveil some hidden structural information in any protein structure (considering that resolution of some structures do not fully allow atom resolved depictions). Besides not having the elaborate insight from actually obtaining these structures, the analytical information is meant to represent additional information of the original PDB entries. Even if it seems redundant to have included intricate information for all the different systems; it serves as an example of the method used here to obtain them, on how to output detailed information for any analogue system of a PDB entry. Due to the industrious increase in computer power, worldwide, I postulate that these kinds of structural analysis and even more sophisticated types, could in the near future, be automatically outputted for any well-defined PDB structure. Which would enable any novice researcher to get an overview of any biochemical structure readily, and not having to spend a PhD-period developing the tools for it. Even more so this kind of structural overview comparison may assist researchers in obtaining better structures, before reporting them in structural databases. Interestingly, a similar analysis approach was taken for the pre-write published [145] ligand saturated "T"-shaped ectodomain, making the results to an extent comparable. The structural understanding of the IR-binding process is likely to improve in the coming years, it is plausible that the structural analysis in this work may facilitate a better understanding also of future structures. That understanding the initial engagement and molecular mechanism of insulin reaching the high affinity cross-link, is another important frontier in the study of this system. Here we postulate that it may be at least approximated in a real-time simulated "movie" and even with time-dependent intricate detail related to what's been presented in this chapter, which could provide another important vantage-point. This endeavour may be achieved even in the next decade by a technique akin to MD or a QM/MM approach, in addition visualized with encompassing analysis depictions. For example, the structural analysis of $\mathbf{CF}_{(\mathbf{A},\mathbf{B})}^{\mathbf{6HN5}}$ as depicted here, could then serve as a comparison to the "paused" scene in the movie; actually depicting the high affinity cross-linked binding, validating the common aspects.

Chapter 5 An Analysis and View of a DGR Model Ensemble

Restrained structures of insulin in solution as derived from NMR

This chapter will investigate a structural analytical overview of a solvent model of insulin. That is to say of experimental restraints and relevant geometry inherent to the DGR models, obtained from the work of Q-x. Hua *et al.* [3, 62]. Their methodology yielded an ensemble evocative of dynamic variability, but which obviously underestimates conformational fluctuations. Their ensemble, however, may still contain inherent geometrical information that are tied to insulin's function in solution; moreover indicative of being near representative of physiological conditions. The structural overview presented in this chapter will serve as a comparison for the next chapter; where it is endeavoured to sample the more far-reaching conformational space for insulin in solvent, as can be obtained with unrestrained molecular dynamics.

5.1 Restated method of the DGR model ensemble of KP-insulin

Here we rewrite the method of how the DGR model (or structure) ensemble was obtained by Q-x. Hua *et al*. [3]. This ensemble is located in PDB entry 2KJJ; having assuredly confirmed with the authors affiliates that the referred article [3] are connected to this entry; albeit it had not been updated in the protein data bank. Their distance-geometry/simulated annealing calculations, were performed using DG-II [270]; restrained molecular dynamics were calculated using X-PLOR [271, 272]. Nuclear overhauser effect related and dihedral angle restraints, were used for the restrained molecular modelling, as described [273]. The restraints (with lower and upper bounds) in total being 881 (average 17.3 per residue): which included 803 H distance restraints (RHHs); and 31 hydrogen bond restraints (RHBs); and 47 dihedral angular restraints (RDAs). The modelled structural ensemble were 40 models whereof 20 of them with the lowest energy were reported. Hydrogen bonds are stated to have been inferred from patterns of amide-proton protection in D_2O [274], for

conditions in which insulin is reported to be monomeric and stably folded [275]. A NMR analysis of amide-proton exchange in D_2O were used as probe of insulin stability and dynamics. NMR spectra were obtained in a variety of conditions. The representative NMR experimental conditions reported for these 20 structures are: temperature 298 K, pH 7.4, protein concentration 0.5 mM, ionic strength 0.1, ambient pressure, solvent 90% H_2O, 10% D_2O. Their DGR protocol enforced all the restraints simultaneously; therefore it may be an underestimation of conformational fluctuations, and also stated to not directly provide dynamic information [3]. When referring to this ensemble of 20 DGR structures, we use the following nomenclature, $\mathbf{E_{kpi}^{DGR}}$, meaning native KP-insulin (kpi), system E, simulation method DGR.

5.2 Analysis and View for $\mathbf{E_{kpi}^{DGR}}$

Here it is shown some geometrical calculations of this structure ensemble.

5.2.1 Flexibility in overall geometry for ensemble $\mathbf{E_{kpi}^{DGR}}$

The traced CA atoms of backbone of this ensemble are shown in Figure 5.1.

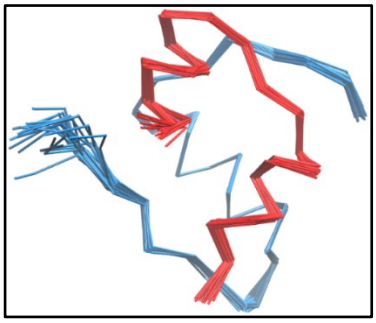


Figure 5.1*: The traced CA-atoms of the ensemble reported in PDB entry 2KJJ. That is 20 structures, reported for conditions 298 K, 0.1 mM salt concentration, pH 7.4.*

What is evident however, is that this solution structure resembles a T-state protomer similar to **M12**, likewise stated by Q-x. Hua *et al.* [3]. Which implies, that having the NT and CT BC strand closer to the core, is the most entropically and energetically favourable conformation of insulin in solution.

Some other geometrical properties of $\mathbf{E_{kpi}^{DGR}}$ is summarized in Figure 5.2 (directly related to its values of Table 6.1). Evident is that the SASA and RGYR has relatively low variation, obviously expected due to the restrained nature of this ensemble. The RMSD and Average RMSF moreover shows that there is higher difference in the BC than for the AC.

Although $RMSF_{\langle SC \rangle}$ reveals that the SC atoms has a higher fluctuation in this ensemble, in particular for the more solvent exposed residues (c.f. the difference in SC atoms of e.g. G_{SC}^{B20} and R_{SC}^{B22}); whereas $RMSF_{\langle MC \rangle}$ shows that the MC atoms has a more constrained nature. In addition, $RMSF_{\langle SC \rangle}$ and $RMSF_{\langle MC \rangle}$, do reveal slight variability at especially the terminal segments of the AC and BC. Hence indeed, this ensemble appears at least evocative of dynamic variability [3]. In addition, a mean insulin structure was obtained as explained in §3.3.1.3, this was the 7'th structure of the 20 reported; albeit they reported the 1'st structure to be most representative; nevertheless, these two structures are among the most similar.

Amide proton exchange

The complementary studies of amide-proton exchange (APE), revealed a contrasting view of the conformational dynamics of insulin. Since the amide proton (HN) of a peptide-bond unit is reactive and able to exchange with either a proton (^1H) or a deuterated one (^2H). The rate of exchange can be used to infer the amount of protection of the peptide bond, which depends on the local structural fluctuations leading to HB breakage and exposure to solvent. Their APE analysis is described in [3], hence not repeated here, however there is a few observations that appears coarsely inferable from their analysis. Due to the largely lack of protection from solvent of the NT B1-11, CT B20-30, NT A1-8 there seems to be a higher fluctuation of these regions enabling their exposure to solvent (and hence exchanges protons more readily i.e. are less protected). On the contrary, the α-helical regions A15-19 and B12-B19 appears less fluctuating and overall more protected from solvent. These observations are to an extent confirmed by the MD results of the next chapter.

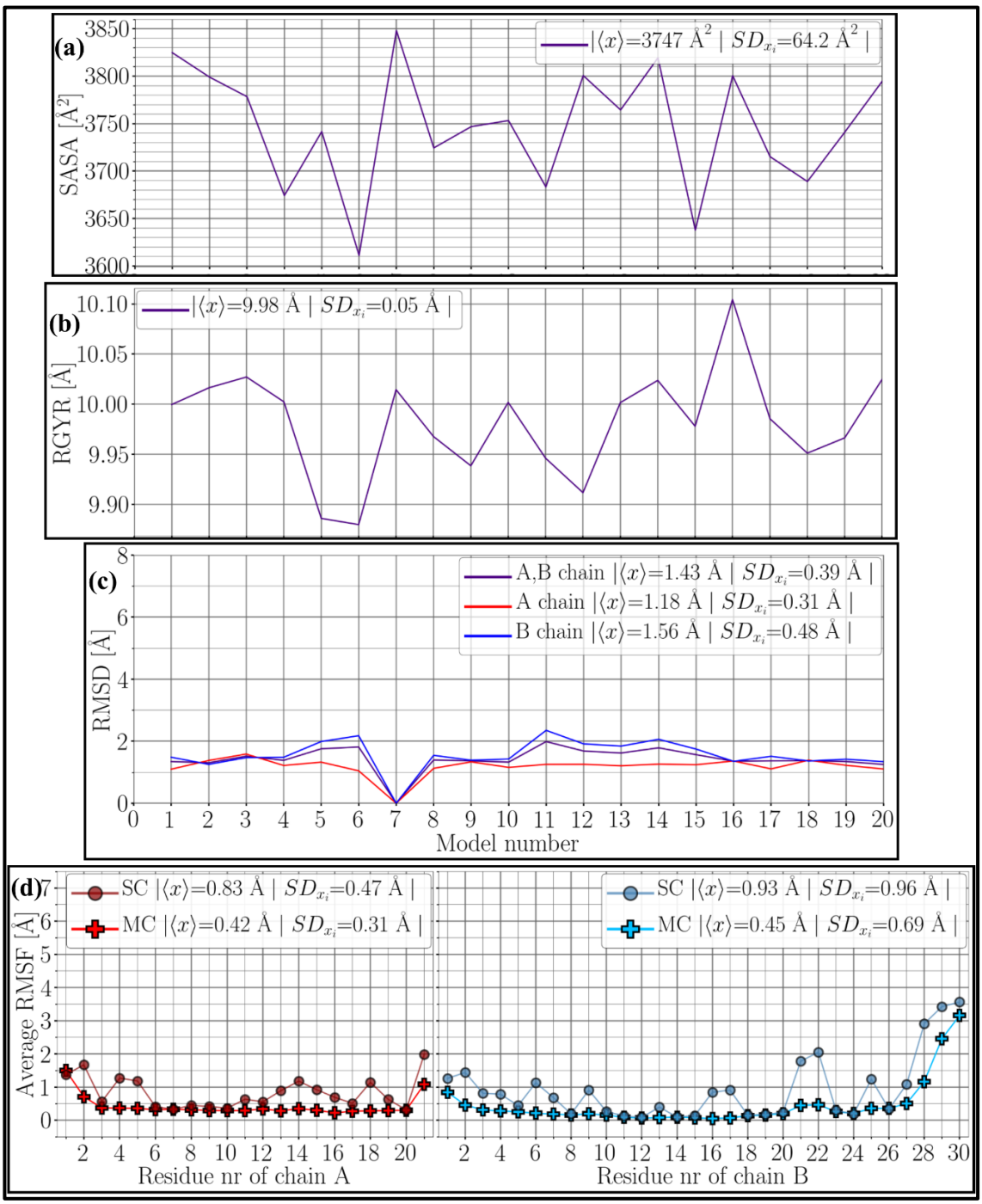


***Figure 5.2**: Flexibility properties of $\boldsymbol{E_{kpi}^{DGR}}$. **(a)** SASA. **(b)** RGYR. **(c)** RMSD. Reference as MS (7'th model), showing the deviation from this structure. **(d)** Average RMSF, i.e. residue-wise $RMSF_{\langle SC \rangle}$ and $RMSF_{\langle MC \rangle}$, respectively given for AC and BC.*

5.2.2 A statistical comparison of intra-monomer calculated HBs and RHBs

Even though $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ were inherently restrained to RHBs (in addition to RHHs and RDAs); here undertaking calculations of HBs as outlined in §3.3.2.5 and of RHBs in §3.3.3.1. Here some of the most prevalent HBs of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ are shown in Figure 5.3. The calculated HBs for all percentages of structures and angle ranges (see §5.2.5 for each HB sorted in matrices), are sorted here in Table 5.1 and Table 5.2 showing also overlap to RHBs and HBs of **M12**. Furthermore, the calculated atom distances of RHBs (separated in bounds in Table 6.4), have the atom-wise accumulated UB violations depicted in Figure 5.4 (violations shown in Table S5.3). Here it is seen that $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ match RHBs well; only three HBs with small violation of UB; with the 9 below LB are only less than 0.24 Å of 1.8 Å. We can see that as the RHBs only include distance bounds, they are not all captured by the calculated HBs of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$; albeit the "D&DH to A" bounds are generous which can otherwise restrain the angle (φ). However, we see that by increasing the angle (φ) for the calculations, the RHBs are covered to an increasing extent (c.f. Table 5.1). For a higher angle ($\varphi < 90°$) all 20 structures cover most of RHBs, except three: ("A10 I(HN) → A5 Q(O)"; "A14 Y(HN) → A12 S(O)"; "A11 C(HN) → B4 Q(O)"); which are not within 3.5 Å and this angle range. Furthermore, for the calculated HBs of **M12**; there is some overlap to RHBs (Table 5.2); almost all except 5 in AC are covered for higher angle ($\varphi < 90°$; for which **M1** have 97 and **M2** 89 HBs). In addition, there are also correspondence of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ to **M12** (c.f. Table 5.1); nonetheless its evident each structure being distinct; which can be expected due to different environments (**M12** HBs being e.g. facilitated by water in the crystal). Furthermore, Hua et al. [3] classified that the majority of HBs observed in crystal structures, as only transiently maintained in solution, including key inter-chain contacts. Where an earlier solvent study by Q. Hua et al. [276], suggested that intra-chain HBs are stabilizing upon self-assembly, in particular "B6 L(NH) → A6 C(O)" and "A11 C(NH) → B4 E(O)"; i.e. that dimerization damps the fluctuations compared to an isolated monomer. Hence, at least a part of the solution HBs, even if they break and reform transiently, should be the same to the crystal structure; indeed this is affirmed by the results here. The RHBs as derived from insulin at low pH and D_2O with 20% deuterated acetic acid, may influence the overall structure and hence also the HBs to some extent; this seems plausible when structures reported in PDB 2HIU [62] (albeit an older seemingly less refined model) obtained under that condition, differs markedly in structure to that of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$.

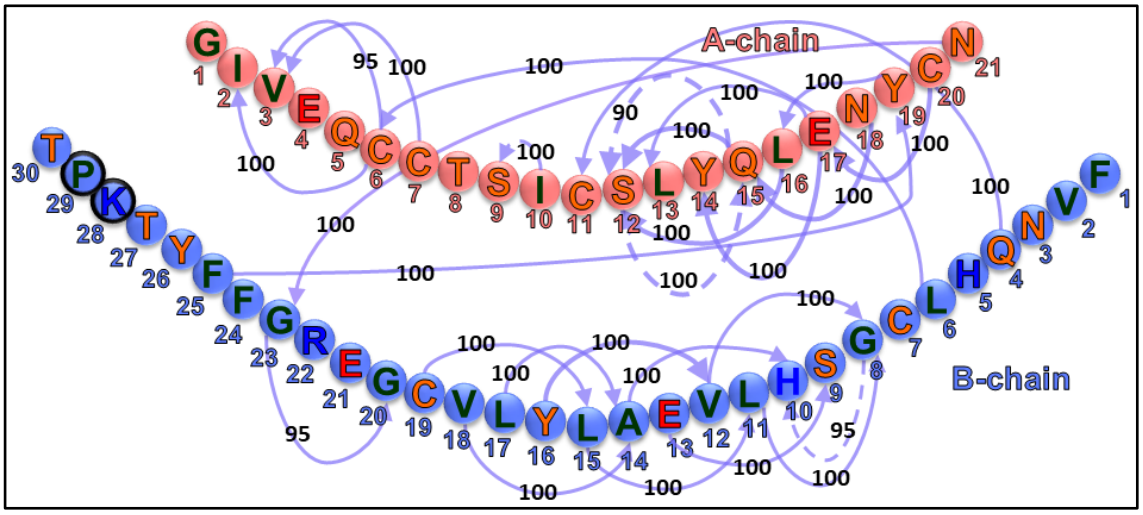


***Figure 5.3**: Calculated medium angle HBs for $\boldsymbol{E_{kpi}^{DGR}}$. The 28 HBs existing between SC and/or MC for more than 75% of the 20 models, for medium angle range ($\varphi < 60°$). Bold numbers indicate percentage of the HBs present in the 20 structures in ensemble. The CYS disulphide bonds of A6-A11, A20-B19, A7-B7 are omitted for clarity. These same HBs in model dependent graphs are depicted in Figure S5.28 with alike colour representation. Moreover the 65 HBs present for more than 5% (i.e. at least 1 out of 20 model structures) are each plotted in Figure S5.29.*

***Table 5.1**: Calculated HBs of $\boldsymbol{E_{kpi}^{DGR}}$ compared to RHBs and HBs of **M12** (Table 4.1). Where $\boldsymbol{E_{kpi}^{DGR}}$ and **M1**, **M2** are respectively calculated with criteria: "$r_{AD} < 3.5$ Å" & "$\varphi < 30°, 60°, 90°$". Note that $\boldsymbol{E_{kpi}^{DGR}}$ compares its HB occupancy at indicated percentage (the -||- symbol designates the same respective range as the indices of 2nd column), showing the overlap to RHB, **M12** (of presence above 0%). In the table the nr of HBs are sorted in "nr in AC_nr in BC_nr between AC&BC" (same format for respective nr of overlap to the other sets of HBs). For example, at medium angle "$\varphi < 60°$", there are sorted HBs 13_11_4(sum 28) present in more than or equal to 15 out of 20 structures in $\boldsymbol{E_{kpi}^{DGR}}$ (i.e. 75% or more), having overlap of "10_8_4" with RHB (i.e. 22 of 31) and "12_9_4" of **M1** (i.e. 25 of 43), and "11_8_4" with **M2** (i.e. 23 of 40). In contrast, each structure in $\boldsymbol{E_{kpi}^{DGR}}$ has other amount of total HBs (c.f. MS at model 7 for each angle range), where the average statistical number at each angle range are: $\langle x \rangle (SD_{x_i})$; 9.40(1.2), 34.8(2.1), 89.9(3.9).*

$\mathbf{E_{kpi}^{DGR}}$	Nr HB's (A_B_AB)		Nr HBs = RHB		Nr HBs = **M1**		Nr HBs = **M2**	
$\sum HB$ >= 5%	>= 5%	>= 25%	-||-	-||-	-||-	-||-	-||-	-||-
	>= 50%	>= 75%	-||-	-||-	-||-	-||-	-||-	-||-
MS	= 100 %		-||-		-||-		-||-	
$\varphi < 30°$								
-||-	7_9_8	4_6_0	4_6_4	3_4_0	4_6_4	3_4_0	5_5_4	4_3_0
24	4_5_0	2_4_0	3_3_0	2_2_0	3_4_0	2_3_0	4_3_0	2_2_0
MS	3_5_0		2_3_0		2_4_0		3_3_0	
$\varphi < 60°$								
-||-	30_22_13	18_17_4	13_9_4	12_9_4	16_11_4	15_11_4	13_10_4	13_10_4
65	16_13_4	13_11_4	11_9_4	10_8_4	13_11_4	12_9_4	13_10_4	11_8_4
MS	17_14_4		12_9_4		14_10_4		13_9_4	
$\varphi < 90°$								
-||-	74_58_18	46_49_5	13_11_4	13_11_4	40_35_5	36_35_4	37_32_5	36_32_4
150	38_42_4	33_41_4	13_11_4	13_11_4	32_32_4	29_32_4	32_30_4	29_30_4
MS	39_45_4		13_11_4		34_34_4		34_31_4	

***Table 5.2**: Calculated low angle HBs for $\boldsymbol{E_{kpi}^{DGR}}$ and RHBs The HBs of the 20 structures are shown with percentage, that of the MS with an asterisk (*). The 31 RHBs of PDB 2KJJ are indicated, whose numbers in or between respective chains can be sorted as 15_11_5 (AC_BC_AC&BC). The RHBs in common with both monomers in **M12** shaded with grey, respectively calculated with increasing angle ($\varphi < 30°,60°,90°$) 4_7_5 (shown explicitly here), 9_10_5, 10_11_5, with a few more matches in AC with individual comparison of **M1**, **M2** (c.f. Table 4.1 and Table S4.2).*

Donor (H)	Acceptor (A)	%, RHB		B12 V(HN)	B8 G(O)	*100
A5 Q(HN)	A1 G O	5		B12 V(HN)	B9 S(O)	RHB
A7 C(HN)	A3 V (O)	RHB		B13 E(HN)	B9 S(O)	*10, RHB
A8 T(HN)	A4 E(O)	RHB		B14 A(HN)	B10 H(O)	RHB
A9 S(HN)	A4 E(O)	RHB		B15 L(HN)	B11 L(O)	*100, RHB
A10 I(HN)	A5 Q(O)	RHB		B16 Y(HN)	B12 V(O)	50, RHB
A10 I(HN)	A9 S(OG)	RHB		B17 L(HN)	B13 E(O)	RHB
A12 S(HN)	A15 Q(OE1)	60, RHB		B18 V(HN)	B14 A(O)	10, RHB
A14 Y(HN)	A12 S(OG)	RHB		B19 C(HN)	B15 L(O)	*100, RHB
A14 Y(HN)	A12 S(O)	RHB		B20 G(HN)	B16 Y(O)	RHB
A15 Q(HN)	A12 S(OG)	*70		B23 G(HN)	B19 C(O)	5
A15 Q(HN)	A12 S(O)	10, RHB		B23 G(HN)	B20 G(O)	45, RHB
A16 L(HN)	A12 S(O)	RHB		A11 C(HN)	B4 Q(O)	RHB
A17 E(HN)	A13 L(O)	RHB		A21 N(HN)	B22 R(O)	10
A17 E(HN)	A14 Y(O)	RHB		A21 N(HN)	B23 G(O)	10, RHB
A18 N(HN)	A15 Q(O)	RHB		A21 N(HD2#)	B24 F(O)	5
A19 Y(HH)	A1 G (N)	5		B4 Q(HN)	A11 C(O)	20, RHB
A19 Y(HN)	A16 L(O)	*90, RHB		B6 L(HN)	A6 C(O)	10, RHB
A20 C(HN)	A17 E(O)	*100, RHB		B25 F(HN)	A19 Y(O)	20, RHB
B8 G (HN)	B10 H(NE2)	*95		B27 T(HG1)	A1 G(N)	5
B11 L(HN)	B8 G(O)	RHB				

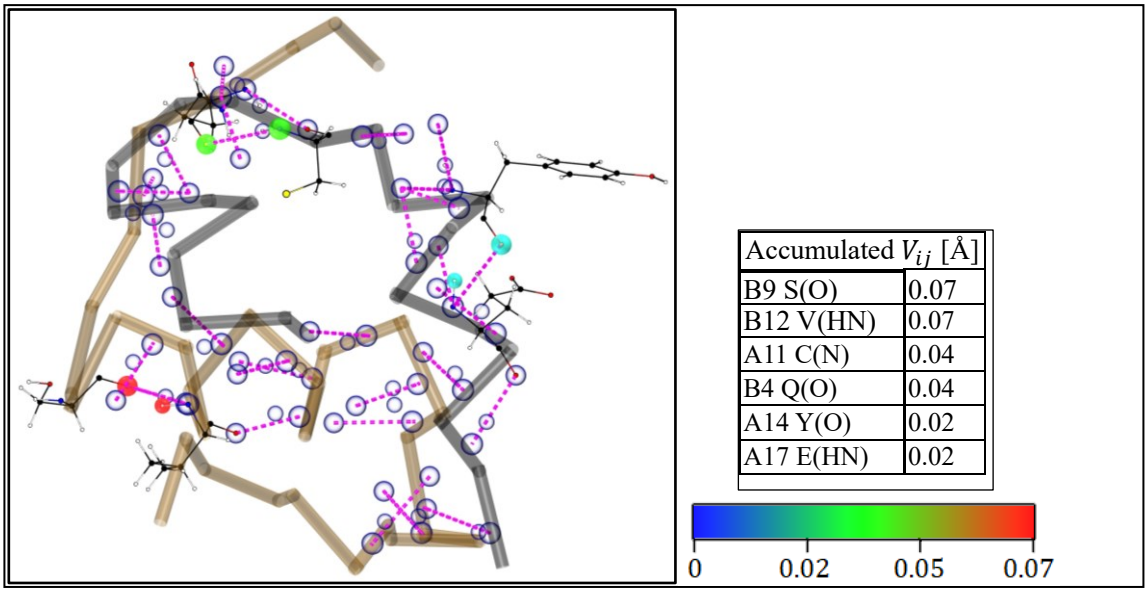


***Figure 5.4**: Visualized RHB violation for $\boldsymbol{E_{kpi}^{DGR}}$. The 31 RHB distances depicted as dotted lines (only shown between the D and A atoms). Note that each RHB have a restraint to both donor and donor hydrogen (i.e. 62 RHBs given). Showing transparent traced CA atoms with AC black and BC brown. Showing only residues that has any RHB violation, in ordinary atom-colouring, to easier see identity. The bigger coloured atoms indicated with colour bar, depicts which atoms has accumulated RHB violation. The RHB assigned atoms with no violation are blue and transparent, i.e. the very most of RHBs bounds are satisfied by $\boldsymbol{E_{kpi}^{DGR}}$. The depiction shown on model 20.*

5.2.3 Calculated NOEs of $\mathbf{E_{kpi}^{DGR}}$ and RHH bounds comparison

This section undertakes NOE calculations and experimental comparison as outlined in §3.3.3 (in particular in §3.3.3.1 and §3.3.3.2).

5.2.3.1 Number of NOEs

When calculating NOEs for larger molecules of say 129 amino-acids, an $\langle r_{ij}^{-3}\rangle^{-1/3}$ averaging may be employed [220]. For a smaller molecule as the insulin monomer, though not apparently stated by Hua *et al.* [3], it were assumed that $\langle r_{ij}^{-6}\rangle^{-1/6}$ averaging is more suitable for comparison [216, 220, 259]. With this latter measure the number of calculated NOEs for the $\mathbf{E_{kpi}^{DGR}}$ ensembles are 5458 (c.f. Table 6.5); which is a much larger number than the 793 included experimental restraints (803 with restraints having GLY HA#). Which is known from NMR experiments, that the total number of NOEs derived, may be a considerably smaller number, than would be expected from the large possible number of hydrogen pairs that may be within 5.5 Å; causes of which can be NMR spectral overlap and incomplete assignment etc [260, 277-279]. There are 381 hydrogens in the insulin monomer, obviously not all possible pairs can be within reach of 5.5 Å, since the insulin monomer is a sterically constrained protein. Nevertheless, if each hydrogen would be compared to every other, there would be 72390 pairs in comparison; which would then be 7.54% of all pairs in $\mathbf{E_{kpi}^{DGR}}$ showing a NOE (with $\langle r_{ij}^{-6}\rangle^{-1/6}$); and only 1.1% of the experimental NOEs.

5.2.3.2 Matrices overlap of calculated NOEs and RHHs

These (abovementioned) calculated NOEs from $\mathbf{E_{kpi}^{DGR}}$ and RHHs are distributed in matrices and overlapped in Figure 5.5. Here it is apparent the wider distribution of hydrogen pairs involved in NOE prediction for $\mathbf{E_{kpi}^{DGR}}$, having overlap with almost all residue pairs found in RHHs. Three experimental NOEs stand out as not counted for $\mathbf{E_{kpi}^{DGR}}$ ($\langle r_{ij}^{-6}\rangle^{-1/6} > 5.5$ Å), however still satisfying congruence (since those RHH UBs are higher than 5.5 Å), with small violation: i.e. "A11 C(HN) - B5 H(HB#)" (V_{ij} of 0.08 Å) and "B5 H(HN) - B3 N(HA&HB#)" (no violations). This congruent overlapping with $\mathbf{E_{kpi}^{DGR}}$ is not surprising, since this structure ensemble were restrained to the RHHs (along with RHB and RDA). However, the overlapping does again imply a T-state like structure indeed being a prevalent motif in solution (made clearer in §6.2.3.3).

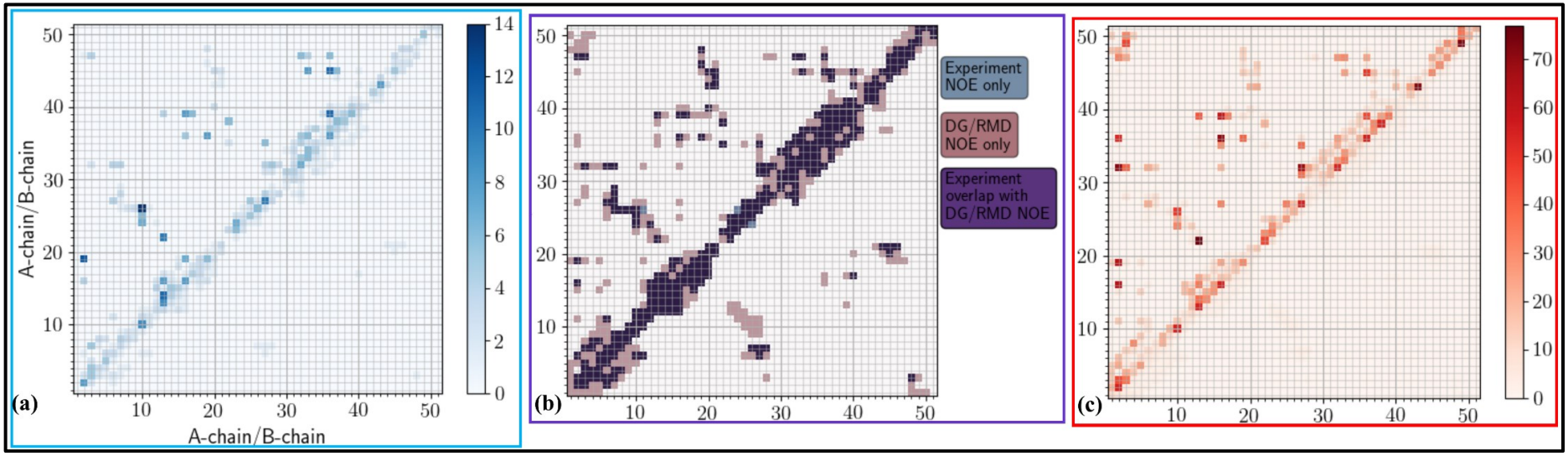


***Figure 5.5**: Matrices overlap between calculated NOEs of $\boldsymbol{E_{kpi}^{DGR}}$ and RHHs . **(a)** Some 793 experimentally derived NOEs (RHHs) from PDB entry 2KJJ. **(b)** The overlap between the RHHs and calculated NOEs of $\boldsymbol{E_{kpi}^{DGR}}$. **(c)** The 5458 calculated NOEs of $\boldsymbol{E_{kpi}^{DGR}}$. The colour-bar numbering has the counts of NOEs for any matrix-index. Matrices is obtained as explained in §3.3.3.2. The overlapped comparison of matrices is done by setting first all indices that have any NOE to 1, respectively for each matrix, then overlapping, separating colours and plotting with matplotlib [280]. Hence the overlap does not mean that all specific hydrogen pair NOEs agree between residue pairs. Note that here residues are numbered in N to C terminal direction as AC (1 to 21) continuing to BC (22 to 51).*

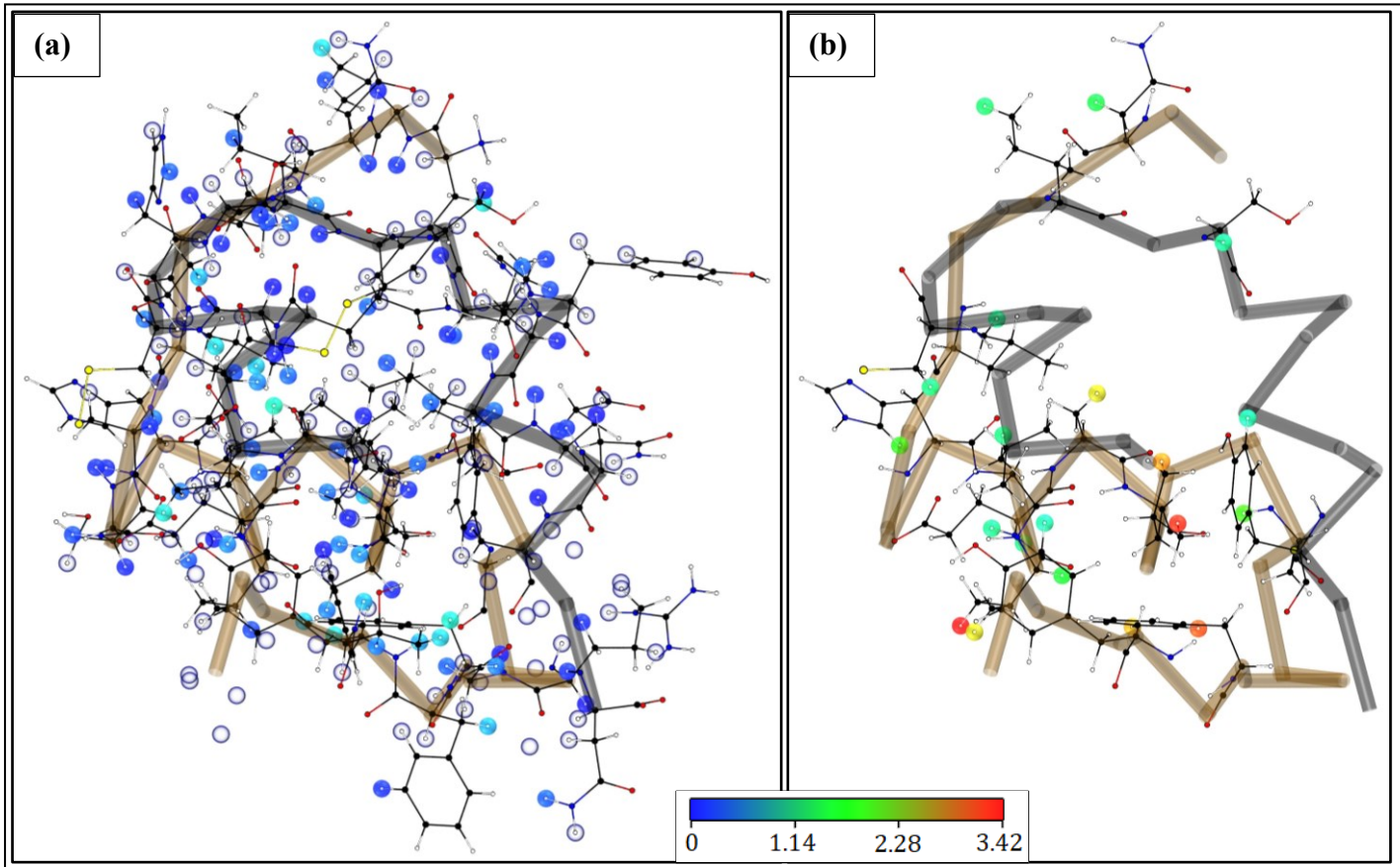
5.2.3.3 Calculated NOEs separated in the RHH bounds

As elaborated in §3.3.3.1, the calculated NOEs from the insulin structures of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$, were separated in respective RHH bound region (shown in Table 6.6). Where it is apparent that the RHHs are between residues closeby in this T-state ensemble $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$; since even the majority of the calculated NOEs (86.76%), are within the limits of the LB and UB.

Here the RHHs have an UB varying between 2.70-9.00 Å; for which there are 13.11% of the calculated NOEs that have a violation (see Table S5.3); however the very most are less than 1 Å (i.e. 12.86%).

Moreover, the average violation of the UBs is 0.0412 Å, with the accumulated violations for individual hydrogens (from different RHHs) are shown in Figure 5.6. Clear then is that the highest violations occur with hydrogens of residues in the BC, e.g.: "B12 V(HG#1)", "B15 L(HD#1)", "B24 F(HE1&HD1)"; these mentioned residues have many other violations of RHHs in between themselves, but also to their nearby residues.

Nonetheless, the RHH bounds seem to be fulfilled to a high degree. Because most violations are less than 1 Å, it may just be that the $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ are slightly distorted, due to being restrained also to the RHB and RDA bounds.



Accumulated V_{ij} [Å]	
B12 V(HG21)	3.42
B15 L(HD11)	3.3
B24 F(HD1)	3.1
B15 L(HD21)	2.8
B24 F(HE1)	2.7
B12 V(HG11)	2.6
B14 A(HB1)	2.6
B19 C(HN)	1.9
B10 H(HD2)	1.8
B12 V(HA)	1.6
B3 N(HB1)	1.5
A10 I(HD1)	1.3
B11 L(HD11)	1.3
B13 E(HA)	1.2
A7 C(HB1)	1.2
B13 E(HN)	1.2
B6 L(HD21)	1.1
A12 S(HN)	1.1
B26 Y(HE1)	1.0
A19 Y(HE1)	1.0
B11 L(HD21)	0.99
B24 F(HB1)	0.9
A3 V(HG11)	0.9
B1 F(HB1)	0.9
A7 C(HN)	0.8
B2 V(HG21)	0.8
B27 T(HG21)	0.8
⋮	⋮

***Figure 5.6**: Visualized RHH violations for ensemble $\boldsymbol{E_{kpi}^{DGR}}$. **(a)** Hydrogens of 0 Å and accumulated violation less than 1 Å (7 largest of 92 shown in table at right). **(b)** Hydrogens of accumulated violation more than 1 Å (shown in table at right). The whole residues are shown with smaller atoms in ordinary atom-colouring, if at least one hydrogen has a violation (in respective range in (a) and (b)). Bigger hydrogens with colour-bar depicts which has accumulated violation (the hydrogens of RHH having no violation, i.e. 0 Å, are in transparent blue). For example, the atom in red "B12 V(HG21)" at value 3.42 Å is apparent. Note that a violation is added to both hydrogens of a restraint. Transparent BB with AC black, BC brown (curvature at CA atoms). Shown on model 20.*

5.2.4 Calculated DAs and RDA bounds comparison

The variation in DAs of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$, is in a way reminiscent of the varying position of SC and MC atoms, in the residues among its insulin units, (c.f. Figure S5.27 and Figure 5.2d). Here the DAs of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ are compared to the respective 47 RDAs (as described in §3.3.3.3), see Figure 5.7, whereof it is seen that the DAs of respective insulin structures are mostly WBs. Note also that the bounds are within $\pm 40°$ of a specific DA in all RDAs, which may not be physically reasonable for some of the χ DAs of solvent exposed residues, e.g. "A10 I(χ_2)". For all deviations of LB or UB are merely a few degrees e.g.: "A10 I(χ_2)" of UB $100°$ has values $101.60°$ to $103.57°$; and "B11 L(χ_1)" of LB $-100°$, has values $-101.66°$ to $-104.04°$; possibly indicating these as not physically reasonable RDAs. Again, asserting the restrained nature of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$, also to the RDAs. Nevertheless many of the RDAs that are maintained here, are found also with congruence for **M12** (c.f. §4.1.2.3), implying that a similar structure is maintained in solution.

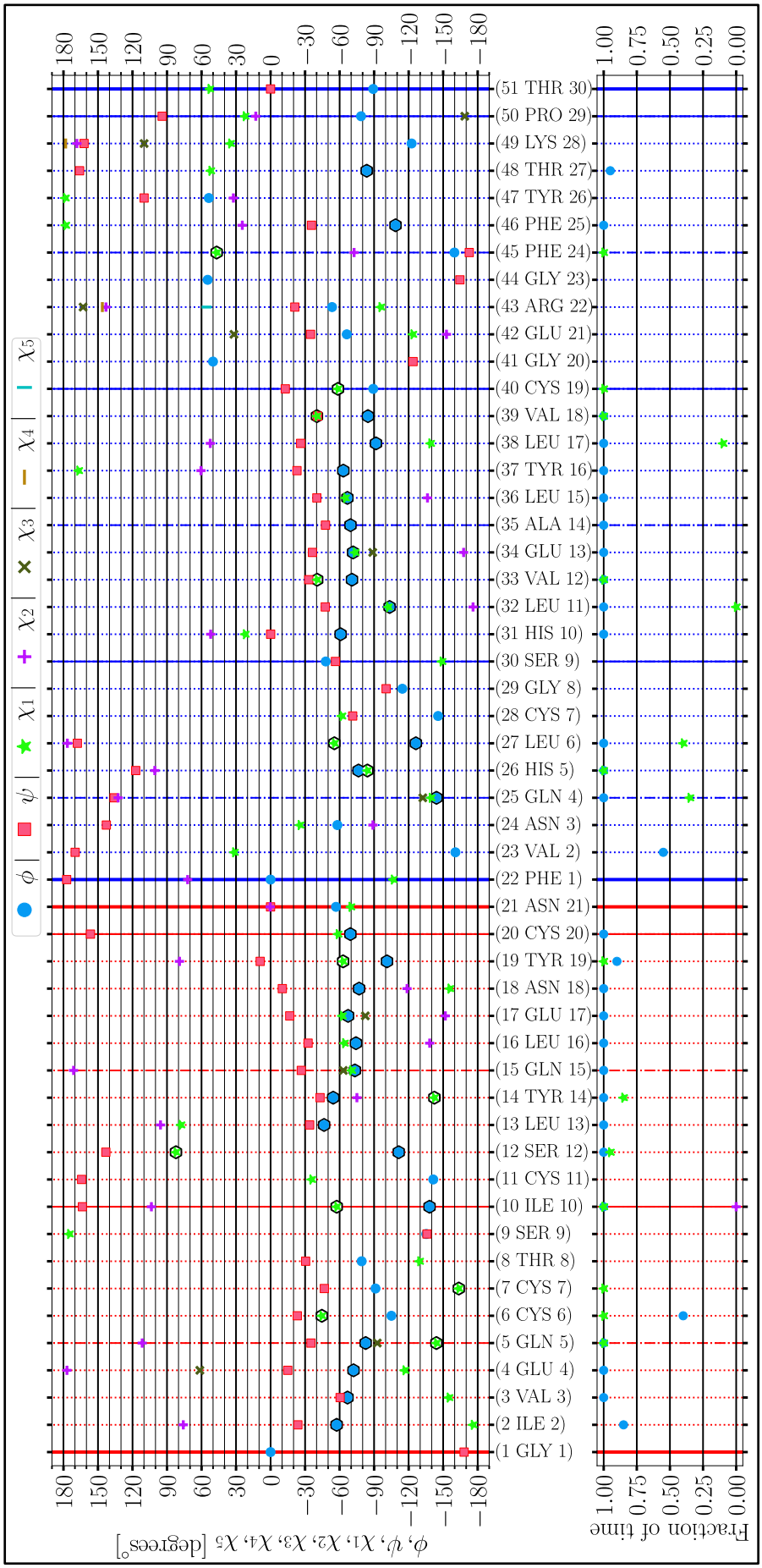


***Figure 5.7**: Dihedral angles and comparison of $\mathbf{E}_{\mathbf{kpl}}^{\mathbf{DGR}}$ to RDA bounds. Showing the DAs of MS, the ones in circled black hexagons are inside RDA bounds. Lower graph depicts the fraction, "f", of structures the respective RDA bound are fulfilled for this ensemble of 20 structures; sum of all, "$\sum f$", being 41.30 (47 if fraction would be 1 of all RDA bounds).*

5.2.5 Conformational analytical overview of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$

The structure of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ are shown with chain-numbering (Figure 5.8), and with distances between residue-moieties (Figure 5.9). The HBs presented here are the same as in §5.2.2, however here represented in HB matrices for various percentage of trajectories and three divisions of the angles: $\varphi < 30°$, in Figure S6.31ab; $\varphi < 60°$, in Figure 5.10 and Figure S6.31cd; $\varphi < 90°$, in Figure 5.11 and Figure S6.31ef. Analogous structural information are given for the other insulin structure models of **M12** (§4.1.2.4), $\mathbf{CF}_{(\mathbf{A},\mathbf{B})}^{\mathbf{6HN5}}$ (§4.2.2) and $\mathbf{P}_{\mathbf{kpi}}^{\mathbf{MDm}}$ (§6.2.5), providing a comparison.

(B4 Q): Here $\langle r_{(MC,MC)}^{A11,B4} \rangle$ is 4.27 Å (between 4.0-4.5 Å). Reflecting partly the HB "$\mathrm{Q}_{\mathrm{HN}}^{\mathrm{B4}} \rightarrow \mathrm{C}_{\mathrm{O}}^{\mathrm{A11}}$" are indeed seen for higher angles ($\varphi < 60$ & $90°$) present for at least 75% of models; however classified as without protection in APE. The other MC HB "$\mathrm{C}_{\mathrm{HN}}^{\mathrm{A11}} \rightarrow \mathrm{Q}_{\mathrm{O}}^{\mathrm{B4}}$; $\langle r_{AD} \rangle = 3.65$; $\langle \varphi \rangle = 20°$" is there however outside of distance criteria (nevertheless this HB was classified as transient due to low protection in APE).

(B25 F): Here, $\langle r_{(MC,MC)}^{A19,B25} \rangle$, being 6.2 Å (between 6.0-6.5 Å); however the HB "$\mathrm{F}_{\mathrm{HN}}^{\mathrm{B25}} \rightarrow \mathrm{Y}_{\mathrm{O}}^{\mathrm{A19}}$", is indicative as a strong HB, present for higher angle ($\varphi < 60$ & $90°$), more than 75% of models; anomalously no protection in APE, plausibly due to much vicinity to solvent and a transient breakage. In the ensemble $\mathrm{F}_{\mathrm{SC}}^{\mathrm{B25}}$ are very subtly moving about the same position, seen pointing outwards (alike for **M2**).

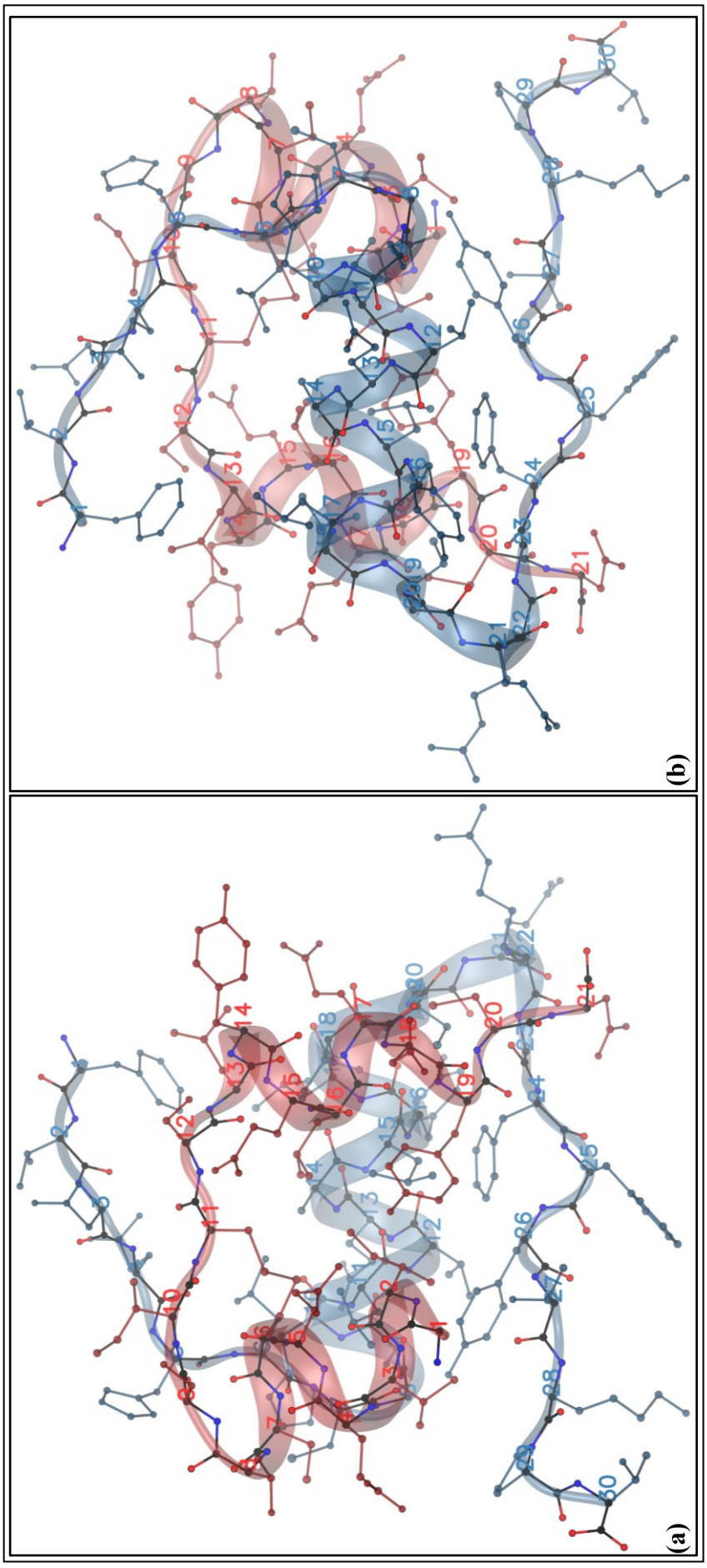


***Figure 5.8**: Structure and numbering for mean structure of ensemble* $\boldsymbol{E_{kpi}^{DGR}}$ *. **(a)** Front, **(b)** back, (front rotated sideways 180°). Hydrogens omitted for clarity. The AC and BC being transparent, and the BB are chalky and in ordinary atom colouring (SC have chain-colour in edgy-glassy look). Chain-numbering right of the CA-atoms.*

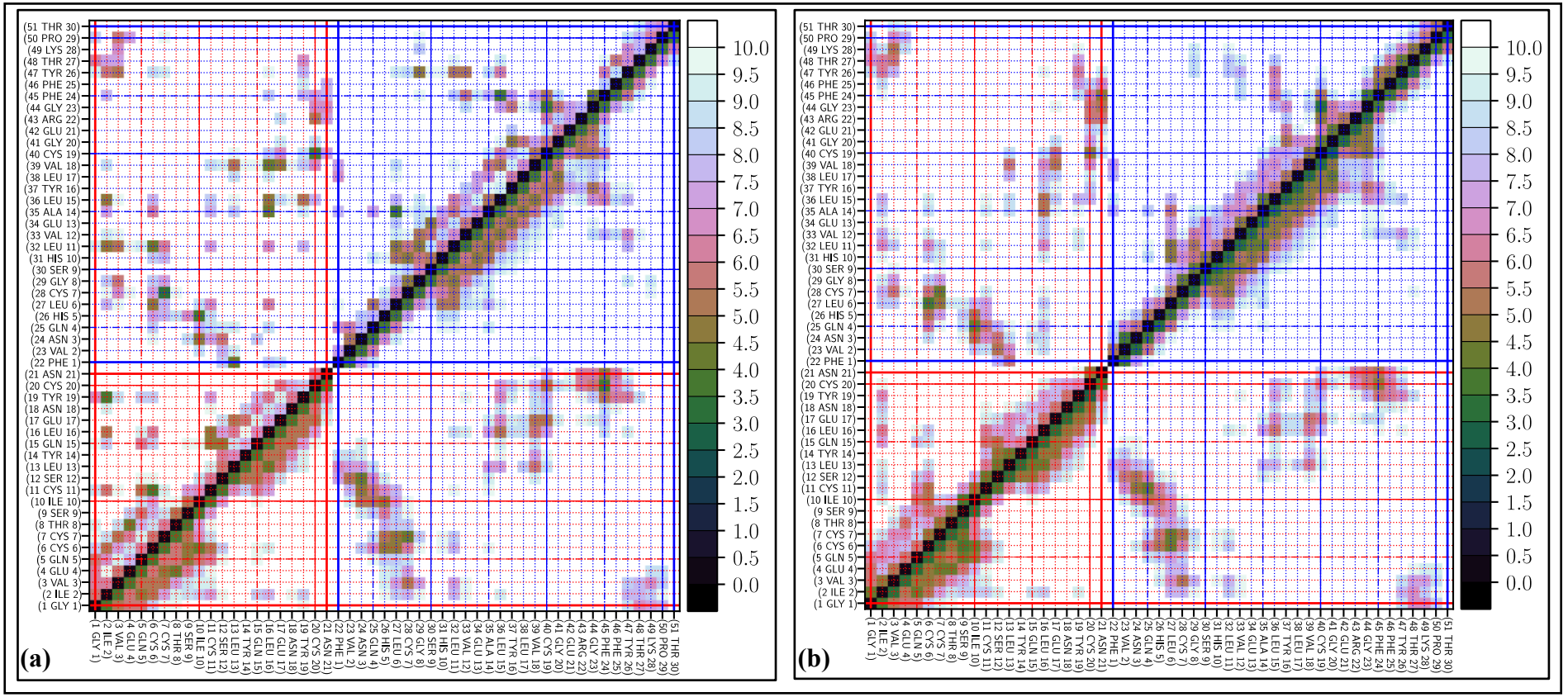


***Figure 5.9**: Average residue-moiety distances within 10 Å for ensemble* $\boldsymbol{E_{kpi}^{DGR}}$ *. **(a)** Upper left is SC to SC geometric centre distances. Lower right is CA to CA-atom distances. **(b)** Upper left is SC to MC geometric centre distances. Lower right is MC to MC geometric centre distances. Distances divided in 0.5 steps, c.f. most of the CA-atom distances of adjacent residues are between 3.5-4.0 Å. Chains residues re-numbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graphs has zoomable vector graphics. grid-lines for AC (1-21) in red and BC (22-51) blue. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Diagonal as reference 0 Å (black), above 10 Å in white.*

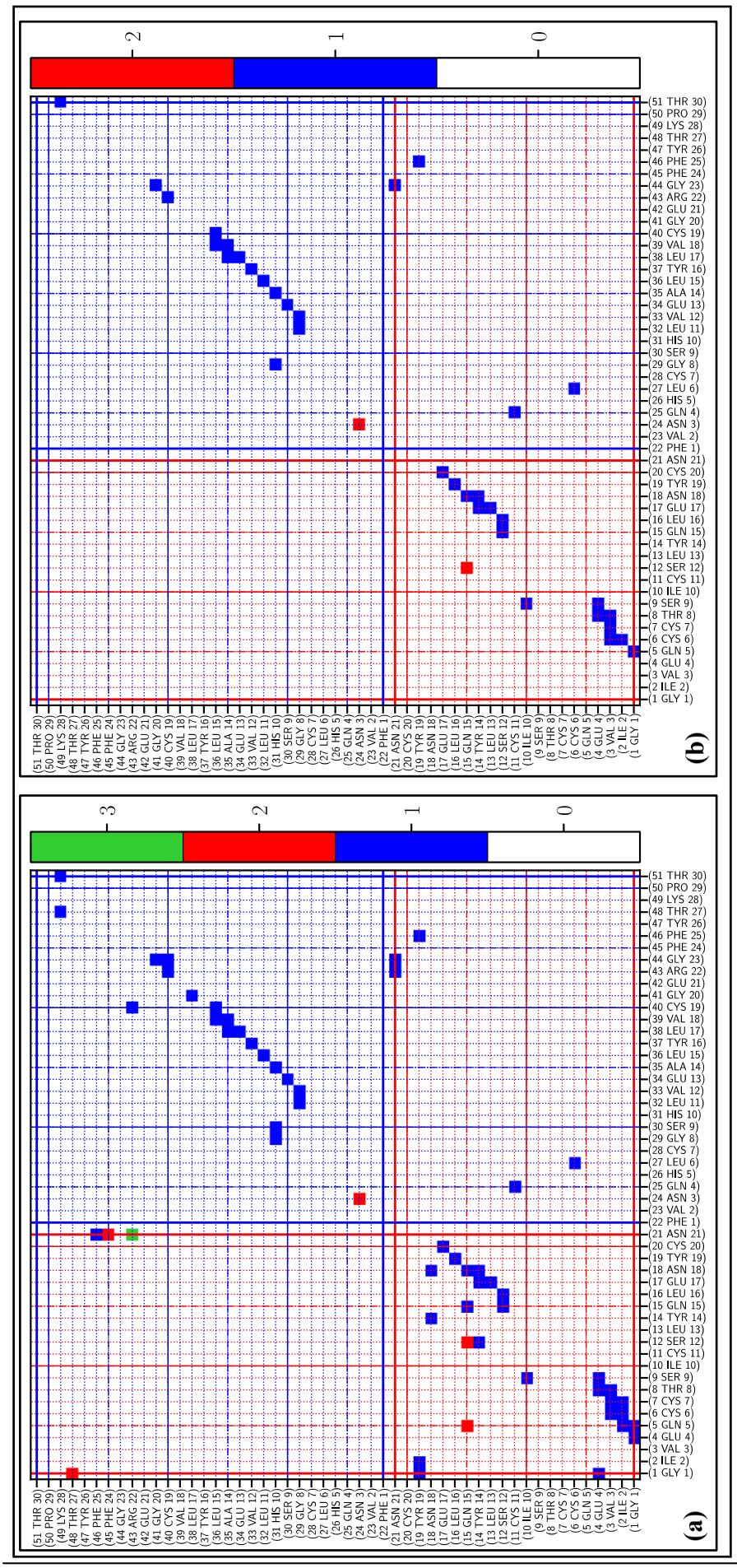


***Figure 5.10**: Hydrogen Bonds matrices between residues of ensemble $\boldsymbol{E}_{\boldsymbol{kpi}}^{\boldsymbol{DGR}}$. The HBs calculated with: "$|\mathbf{r}_{AD}| < 3.5$ Å, & $\varphi < 60°$" and presence larger **(a)** 5% of models, 65 HBs **(b)** 25% of models, 39 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. grid-lines for AC (1-21) in red, and BC (22-51) blue. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51.*

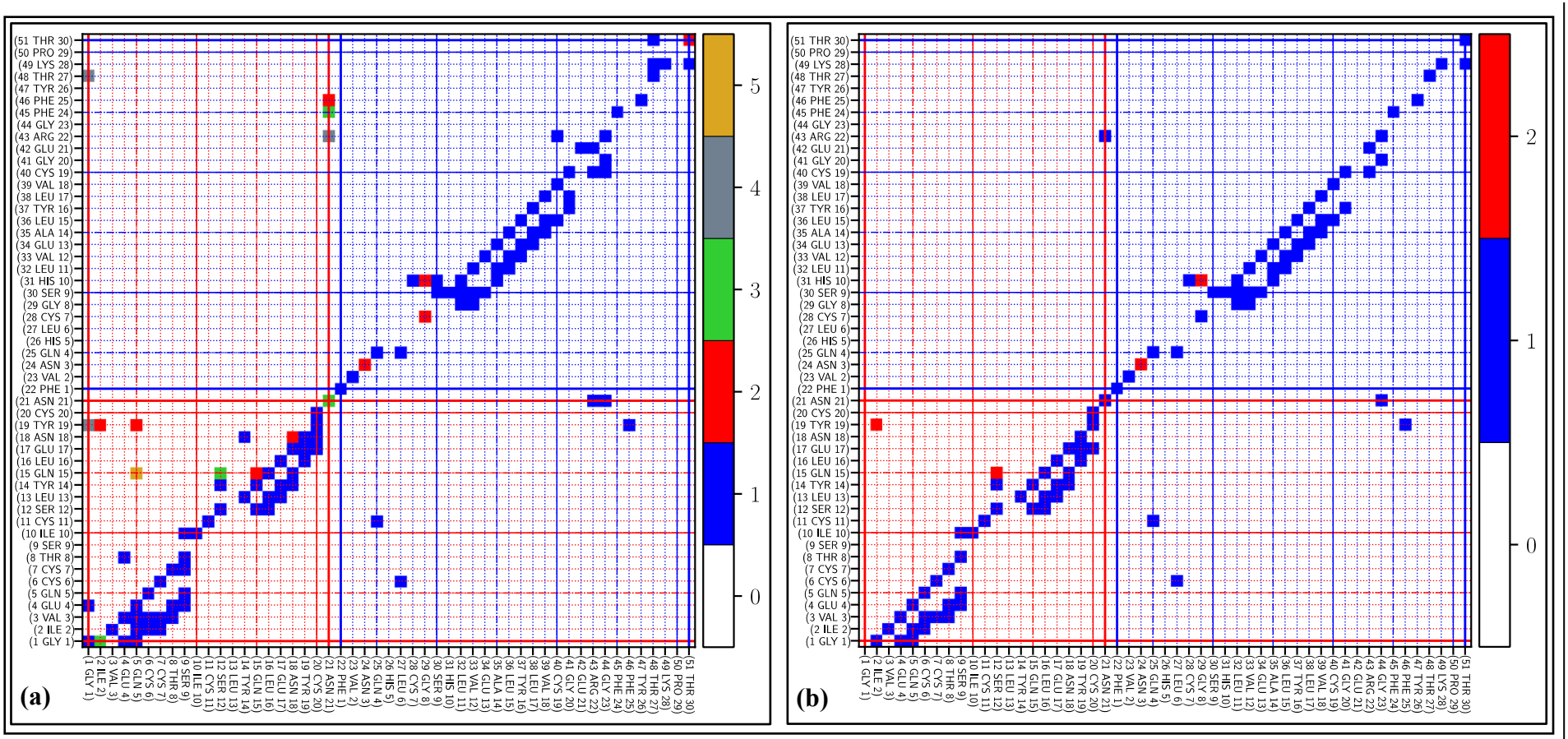


Figure 5.11: *Hydrogen Bonds matrices between residues of ensemble* $\boldsymbol{E_{kpi}^{DGR}}$. *The HBs calculated with: "*$|\boldsymbol{r}_{AD}| < 3.5$ Å, & $\varphi < 90°$*" and presence larger* ***(a)*** 5% *of models, 150 HBs* ***(b)*** *25% of models, 100 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. grid-lines for AC (1-21) in red, and BC (22-51) blue. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51.*

5.3 Concluding statement

This short chapter served as a brief analysis of a model of the solvated insulin monomer. For obtaining geometrical properties as calculated from this DGR model structure ensemble, moreover for comparison to its restraints. Not surprisingly $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ is an ensemble that largely fulfils the restraints used to derive it. As it appears, to a lesser extent, the conformational uncertainty of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ (as seen in geometry and the restraints), are evocative of realistic dynamics of the solvated monomer; hence $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$ appears to be an average or "molten" structure ensemble. Partly explained by the restraints being enforced simultaneously in the DGR protocol, and as implicated by Q. Hua *et al.,* the resulting structure ensemble underestimates conformational fluctuations. Moreover, the higher number of calculated HBs, NOEs, and DAs of the structure ensemble as compared to the respective restraints, may suggest a higher degree of conformational motion. Nevertheless, this structure ensemble cannot fully depict the realistic flexibility of the monomer; however the residues B1-5 and B25-30 appears to be indicative of relatively higher flexibility, in addition to certain residue SCs throughout the monomer. The calculated HBs (at $\varphi < 30°$) of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$, I'd deem not as representative as average HBs in real monomer dynamics; however a fair congruence was found with experimentally derived HBs, for higher angles ($\varphi < 60°$), which may imply a higher flexibility in real dynamics. The original authors performed complementary amide-proton exchange studies to infer some of the dynamics, and may be consulted with also if inferring results [3].

The restraints reported in PDB entry 2HIU has some weird numbers and may have used an older XPLOR version, hence the restraints of this PDB entry were not used in comparison. However the more recently obtained PDB entry 2KJJ, appears more reliable and are used also in restraint comparison of the MD simulations in the next chapter; wherein the DGR solution model of this chapter, are the basis for initial structures used in MD simulations of analogous conditions; to infer if this model can be improved upon and provide more dynamical information.

Chapter 6 An Analysis and View of MD trajectories

Sampling conformational space of insulin in solution by Molecular Dynamics

In this chapter we perform multiple MD simulations, using a DGR structure from the previous chapter as initial conformation. Endeavouring to explore the conformational space available to insulin in solvent. Comparing the geometrical properties in between the DGR and MD insulin structure ensembles and their respective congruence with experimental restraints. Conjointly, testing the influence of different temperature and starting conformations, having otherwise the same parameters. Each simulation goes its own path and shows distinct behaviour; however yielding similar observable averages. Out of these many MD ensembles, inferring the ones which makes the most representative models for the solvated insulin monomer. In addition, noteworthy is the one out of nine replicas occurring unanticipated sampling of another state than a T-state monomer; notwithstanding perhaps a possible conformation of insulin in solution.

6.1 Methods for MD trajectories

Method of simulation of §3.2.1 was applied for the MD simulations here. The rationale for our choice of protonation state, of each of the ionizable amino-acids for this pH of 7.4, are described at end of §3.1.3. Enabling a close to an analogous comparison with the DGR structure ensemble ($\mathbf{E_{kpi}^{DGR}}$), that have experimental conditions representative for them (§5.1); accordingly chosen to perform these MD simulations, resulting in MD ensembles denoted as $\mathbf{E_{kpi}^{MD}}$, with close to corresponding conditions at: 300 K, pH 7.4, 10 Na and 8 Cl. For increasing and stabilizing the conformational sampling, the temperature was elevated to physiological (310 K), denoted as $\mathbf{P_{kpi}^{MD}}$; in addition to the testing of another set of coordinates, denoted as $\mathbf{P_{i}^{MD}}$. Initial structures for MD protocol for ensembles $\mathbf{E_{kpi}^{MD}}$, $\mathbf{P_{kpi}^{MD}}$ are seen in Figure 3.1a and of $\mathbf{P_{i}^{MD}}$ in Figure 3.1b, where there were three replicas (m, n, o) for each respective trajectory considered.

6.2 Analysis and view for $\mathbf{P_{kpi}^{MD}}$, $\mathbf{E_{kpi}^{MD}}$, $\mathbf{P_{i}^{MD}}$

This section compares the analysis of these 9 MD trajectories; however especially that of $\mathbf{P_{kpi}^{MDm}}$ are depicted here. There is a danger of grossly misinterpreting MD simulations if not carefully and methodologically considering the results; which has also been stated by developers and users of computational chemistry simulation software [49, 211]; one reason why the scrutinizing approach is taken in this section.

6.2.1 Overall geometry and fluctuations of MD replicas

6.2.1.1 Superimposition consequence on occupancy and other analysis

This section is to give the reader a feeling of the resulting 0-1499, 1 ns time-step, trajectories (or 1500 analysis boxes used in analysis), from each replica of the MD simulations. Moreover, why the superimposed region B11-17 were used for all subsequent analysis. With each time-frame containing a box of: the insulin molecule of 786 covalently bonded atoms; 10 sodium atoms and 8 chloride ions; and an ample amount of ~4986 water molecules (see Figure 6.1). The polar and charged solvent may too some extent affect the dynamics of insulin; of course the occupancy in any volume of space (excepting the protein region), is expected to be much greater for water. By the method outlined in §3.3.2.3, the fractional occupancy was calculated of: the protein (all-atom, BB, CA-atoms respectively); and solutes, i.e. water, sodium and chloride. Note though for the protein centred in the box, for all time-frames, that the most occupied area for any of atom-selections are dependent on the atoms of superimposition. That is to say, the superimposed atoms will be the least mobile area of the atoms in the box (c.f. Figure S6.35), with rest of molecules in analysis-box moving about (with protein centred in box for all time-frames). The superimposed region chosen here to be the CA-atoms of B11-17, which is relatively stable within, also in relation to the AC αHs and turns B6-8 and B20-24. With this choice of superimposition, for the all-atom protein (Figure 6.2), it is indeed understood that the monomer samples a large space, especially seen for the BC NT/CT strands; apparent is that the most occupied areas are those closest to the monomer core. A similar profile is found for the analogous calculation of backbone and CA-atoms of protein (not shown); however respectively somewhat less sampled in space, since including less atoms.

The chloride and sodium fractional occupancy (Figure S6.32, Figure S6.33) covers almost all of space when considering the lowest isovalue. However, will be much less sampled in

distinct volumes of space, fading away to zero at less than 1% fractional occupancy. Albeit, it is interesting to note that there is a larger occupancy of Na^+ close to A14-21 and B20-24, not appearing to be due to the choice of superimposed region. The negatively charged glutamate groups Glu^{A17}, Glu^{B13}, Glu^{B21}, and carboxy group of Asn^{A21}, may at least partly explain the higher occupancy of Na^+ around these residues. Albeit only vaguely observed, is minuscule amounts larger fractional occupancy of Cl^-, found closer to positively charged residues.

The water occupancy (Figure S6.34), however was noted to shrink as more time were included. This is obviously due to the water outside the radius of the box sidelength will have less occupancy (the cubic water box rotating in space around the fixed CA-atoms of B11-17). Nevertheless, it can be seen for water the highest fractional occupancies (or most occupied regions) are above 40-50%, slowly diminishing, and at 50-78% it slowly goes to zero around B11-17 (not notably due to its charged residues). Furthermore, the same conclusion of fractional occupancies was found for the ensembles $\mathbf{P_{kpi}^{MDno}}$, $\mathbf{E_{kpi}^{MD}}$, and $\mathbf{P_{kpi}^{MD}}$, though all stochastically unique. The most conformationally different ensemble are included for $\mathbf{P_i^{MDm}}$ (Figure S6.50).

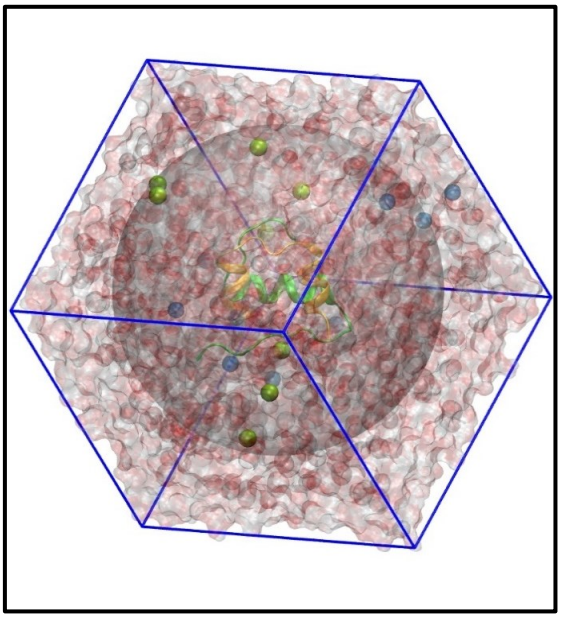


***Figure 6.1**: Simulation box of $\boldsymbol{P_{kpi}^{MDm}}$. Solutes shown in surface representation, water (H white; Oxygen red) and Na^+ (blue), Cl^- (green-yellow), for snapshot (or time-frame) 508 ns. The box has a sidelength (blue lines) of 53.87 Å, whereof a sphere with radius of half sidelength (26.935 Å), are drawn in transparent grey, centred at geometric-centre of all atoms of protein coordinates, at time-frame 508 ns. Box is rotating around the protein fixed in space about CA-atoms B11-B17. The simulation was calculated with periodic boundary conditions, but the resulting trajectory will be snapshots of 1500 boxes of atoms, rotating around the superimposed-region fixed protein. Hence the sphere represents always occupied space in the calculation of e.g. fractional occupancy as the 1500 different atom boxes are loaded in VMD.*

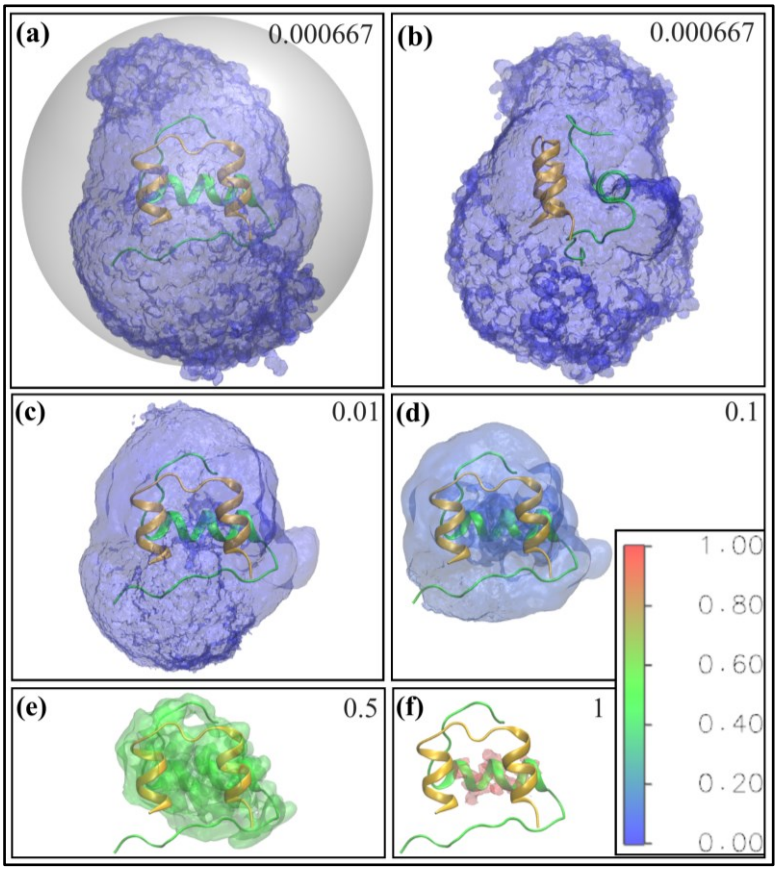


***Figure 6.2**: Protein amino-acids, all-atom, fractional occupancy for $\boldsymbol{P_{kpi}^{MDm}}$. At respective isovalue at top-right (multiply by 100 to get % of trajectory). Most occupied regions respectively when considering varying isovalues **(a)** 1/1500 (0.0667%) (In transparent grey is an imaginary sphere of radii 26.935 Å (analysis box half sidelength) centring at "protein geometric centre" of MS time-frame). **(b)** (a) from right side (rotated 90°). **(c)** 15/1500 (1%), **(d)** 150/1500 (10%), **(e)** 750/1500 (50%), **(f)** 1500/1500 (100%). The figure meaning is that a large space are occupied by protein atoms, considering all time-frames; most occupation are closer to monomer since closer to the fixed in space superimposed protein atoms.*

6.2.1.2 Flexibility of the CA-atoms

The flexible regions of the ensemble $\mathbf{P_{kpi}^{MDm}}$ are shown in Figure 6.3; mostly the NT & CT B-chain is shown to have the relatively highest fluctuation. Visible is that the deviation of the overlap of B11-17 is small, hence this region was chosen as reference (of course another region of superimposition would yield another picture of flexible regions). Indeed it does seem on average that the MSs will be close to a T-state structure as obtained by MD (by the measure in §3.3.2.1). Also notwithstanding, it is seen in Figure 6.4 (c.f. Figure S6.44, Figure S6.47), that the MSs are more or less similar to that of $\mathbf{E_{kpi}^{DGR}}$. The stochastic fluctuations of the MD ensembles are explaining these variations to some extent; in a way reflecting the congruence structurally and experimentally (explained following sections).

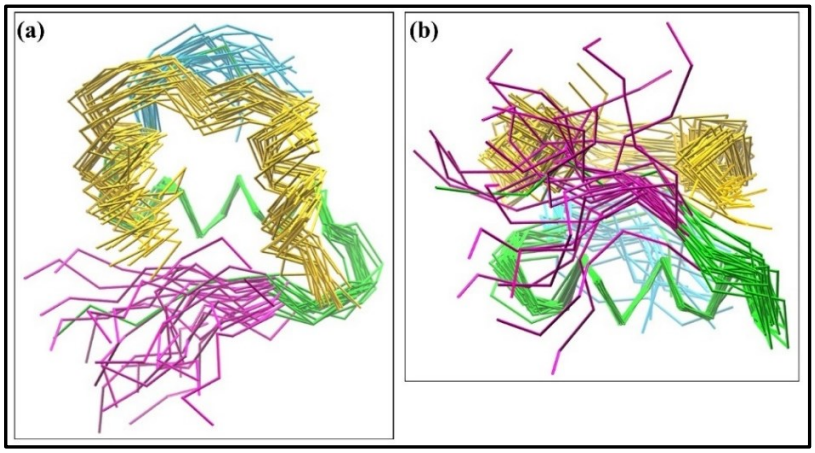


***Figure 6.3**: Flexible CA-atom regions of ensemble $\boldsymbol{P}_{\boldsymbol{kpi}}^{\boldsymbol{MDm}}$. The MS has AC yellow, BC green, in addition are 20 snapshots (i.e. every 74'th ns from 19 ns to 1499ns) with B1-5 coloured turquoise (including CT MC bond between residues B5-6), and B25-30 coloured magenta (including CT MC bond between residues B24-25). Superimposed region is B11-17. **(a)** Front. **(b)** Bottom (rotated 90° upwards).*

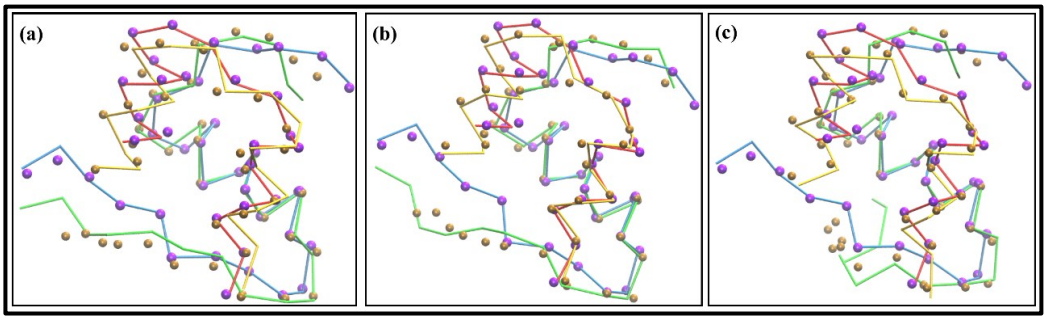


***Figure 6.4**: Traced CA-atoms of MSs of simulated ensembles $\boldsymbol{P}_{\boldsymbol{kpi}}^{\boldsymbol{MD}}$. Superimposed MS of these MD ensembles to that of $\boldsymbol{E}_{\boldsymbol{kpi}}^{\boldsymbol{DGR}}$ (7'th structure in PDB 2KJJ). Shown with respective ensemble averaged positions of C_α atoms, displayed are the C_α atoms with those averaged **r**(x,y,z) coordinates, coloured brown for $\boldsymbol{P}_{\boldsymbol{kpi}}^{\boldsymbol{MD}}$ and coloured purple for $\boldsymbol{E}_{\boldsymbol{kpi}}^{\boldsymbol{DGR}}$. That of **(a)** replica "**m**" at 508 ns. **(b)** replica "**n**" at 949 ns. **(c)** replica "**o**" at 1071 ns.*

6.2.1.3 Sampling fluctuations in structure

In Table 6.1 are the average and standard deviation for many trajectories of properties: SASA, RGYR, RMSD and RMSF (see §3.3.1.1-3.3.1.4). Following are the corresponding properties shown for the whole trajectory of $\mathbf{P_{kpi}^{MDm}}$ in Figure 6.5, Figure 6.6 and Figure 6.7 respectively. In addition one may compare with same properties of $\mathbf{E_{kpi}^{DGR}}$ in Figure 5.2, noting that it is obvious that the values, $\langle \mathrm{RMSD} \rangle$ and $\langle \mathrm{RMSF}_{\langle \mathrm{AA} \rangle} \rangle$, are many times larger than that of $\mathbf{E_{kpi}^{DGR}}$ (c.f. also $\mathrm{RMSF}_{\langle \mathrm{SC} \rangle}$ and $\mathrm{RMSF}_{\langle \mathrm{MC} \rangle}$ as a function of residues in plots). In addition, noting that the overlap between SASA and the number of water molecules within 5 Å are showing that on average around 470 water molecules are close to insulin; noting also larger peaks when the A1-10, A11-21, B1-5 and B25-B30 show large RMSDs. Furthermore, as revealed by the RMSD and average RMSF, it is even more evident that the MD trajectory is highly stochastic and fluctuating. Noteworthy is that A1-10 are overall more flexible than A11-21; in addition, there are relatively larger fluctuations of the B-chain NT & CT residues, which occasionally more or less depart from the monomer core. Furthermore, it is revealed the fluctuation of the SCs (c.f. $RMSF_{\langle SC \rangle}$), especially the solvent-exposed ones; in addition to that the MCs of some residues are highly fluctuating (c.f. $RMSF_{\langle MC \rangle}$).

Nevertheless, smaller variation is found for the average RMSF between replicas of $\mathbf{P_{kpi}^{MD}}$ (c.f. Figure S6.36); where there are less uniformity between replicas of $\mathbf{E_{kpi}^{MD}}$ (Figure S6.46) and $\mathbf{P_{i}^{MD}}$ (Figure S6.49). Even though $\mathbf{E_{kpi}^{MD}}$ are understandably more similar in nature to $\mathbf{P_{kpi}^{MD}}$, it appears that $\mathbf{E_{kpi}^{MDno}}$ have slightly higher fluctuations and $\mathbf{E_{kpi}^{MDm}}$ somewhat lower. Apparently $\mathbf{P_{i}^{MDno}}$ are more similar to $\mathbf{P_{kpi}^{MD}}$; indeed it appears that the change of initial structures, does alter the ensembles structural trajectory to a greater extent. In contrast, there is anomalous fluctuation seen for $\mathbf{P_{i}^{MDm}}$, observed also from the RMSD (Figure S6.48); from resembling a T-state undergoing a transition at ~60-120ns, wherefrom staying close to its mean structure (Figure S6.47a). Though it does occur, more or less durable, large deviations of B1-5 in the other 8 ensembles considered here; no event where $\mathrm{L}_{\mathrm{SC}}^{\mathrm{B6}}$ departs from its hydrophobic cavity for an extended period of time as in $\mathbf{P_{i}^{MDm}}$.

***Table 6.1**: Statistically calculated geometrical properties of insulin structure ensembles. For $\boldsymbol{P_{kpi}^{MD}}$ and $\boldsymbol{E_{kpi}^{MD}}$ the RMSD reference structure is the MS of $\boldsymbol{E_{kpi}^{DGR}}$; however for $\boldsymbol{P_{i}^{MD}}$ it is the MS of DGR models in PDB 2HIU (albeit approximately same values as with MS of $\boldsymbol{E_{kpi}^{DGR}}$). Calculations are including all atoms of AC and/or BC as indicated.*

	$\langle SASA \rangle (SD_{x_i})$ AC & BC	$\langle RGYR \rangle (SD_{x_i})$ AC & BC	$\langle RMSD \rangle (SD_{x_i})$ AC | BC		$\langle RMSF_{\langle AA \rangle} \rangle (SD_{x_i})$ AC | BC	
$\mathbf{E_{kpi}^{DGR}}$	3747 (64.2)	9.98 (0.05)	1.18 (0.31)	1.56 (0.48)	0.68 (0.39)	0.76 (0.85)
$\mathbf{P_{kpi}^{MDm}}$	3986 (179.1)	10.51 (0.29)	3.53 (0.93)	5.09 (1.47)	2.41 (0.62)	2.88 (2.88)
$\mathbf{P_{kpi}^{MDn}}$	4007 (190.3)	10.53 (0.36)	3.61 (0.67)	5.28 (1.60)	2.30 (0.70)	2.87 (2.89)
$\mathbf{P_{kpi}^{MDo}}$	3965 (174.1)	10.46 (0.28)	3.60 (0.77)	6.50 (1.67)	2.13 (0.57)	3.06 (3.22)
$\mathbf{E_{kpi}^{MDm}}$	3921 (128.0)	10.37 (0.20)	3.22 (0.46)	4.40 (0.99)	1.94 (0.55)	2.31 (2.13)
$\mathbf{E_{kpi}^{MDn}}$	3998 (183.0)	10.50 (0.31)	3.70 (0.89)	5.40 (1.62)	2.37 (0.70)	3.16 (3.20)
$\mathbf{E_{kpi}^{MDo}}$	4123 (161.8)	10.60 (0.29)	4.19 (0.74)	6.34 (1.65)	2.33 (0.68)	3.12 (3.04)
$\mathbf{P_{i}^{MDm}}$	4219 (206.8)	10.99 (0.43)	6.43 (1.09)	11.4 (1.67)	3.30 (0.71)	3.29 (3.24)
$\mathbf{P_{i}^{MDn}}$	3940 (139.4)	10.45 (0.19)	3.48 (0.58)	5.33 (0.74)	2.11 (0.53)	2.16 (1.83)
$\mathbf{P_{i}^{MDo}}$	3923 (167.1)	10.39 (0.28)	3.45 (0.56)	4.76 (1.34)	2.12 (0.52)	2.78 (2.28)

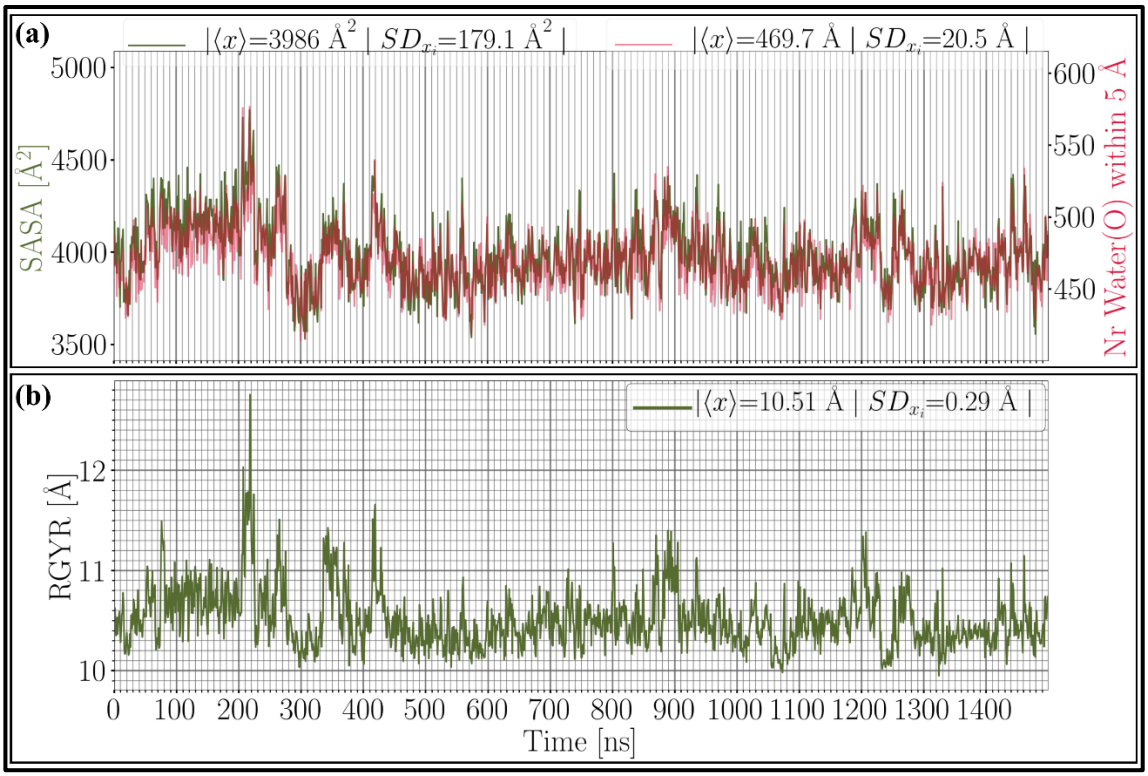


***Figure 6.5**: SASA and RGYR of the insulin structure ensemble of $\boldsymbol{P_{kpi}^{MDm}}$. Calculated for whole protein (AC & BC). **(a)** SASA, note the congruence with water near the protein at any time. **(b)** RGYR.*

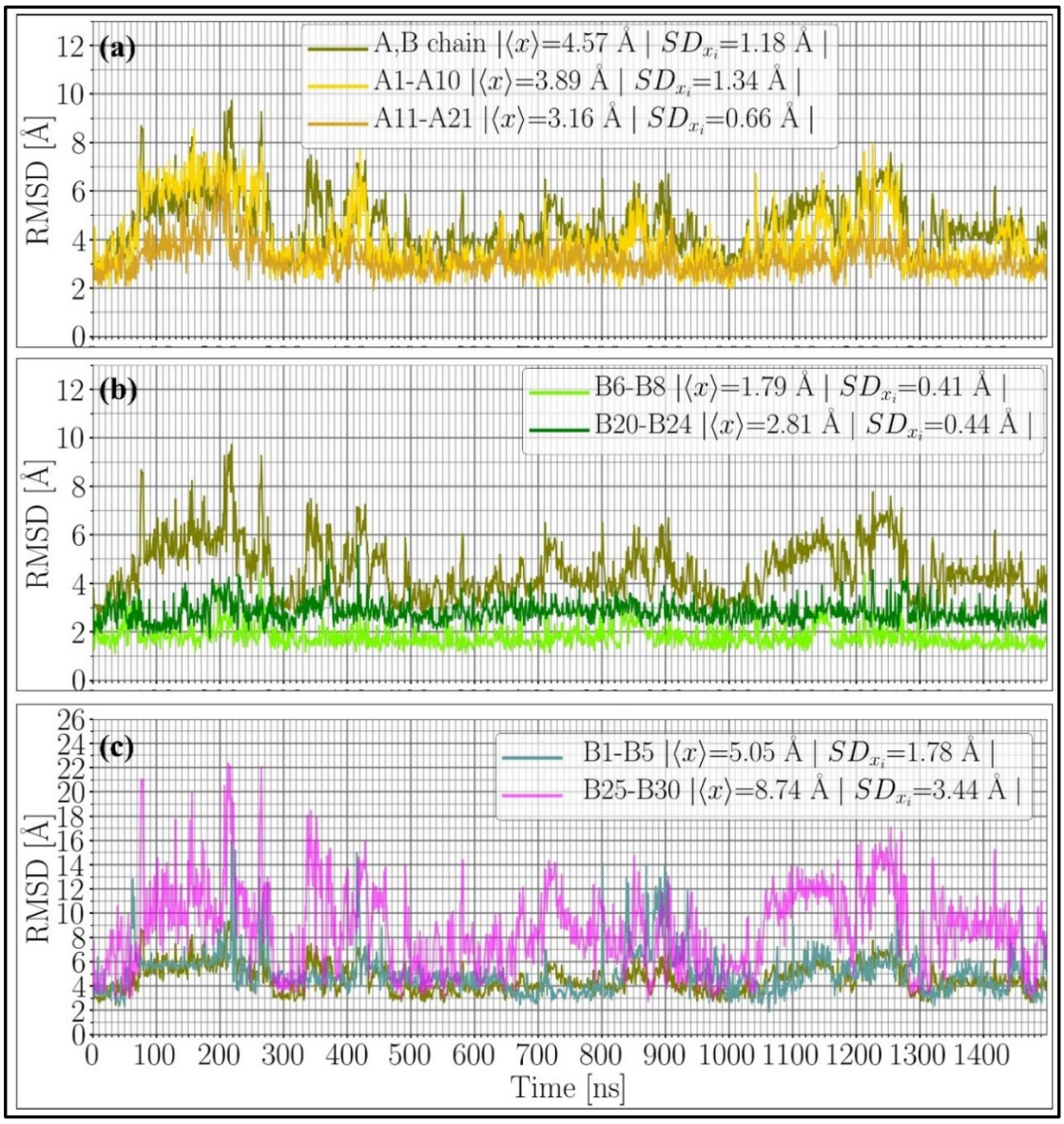


***Figure 6.6**: The RMSD of specific segments of the ensemble $\boldsymbol{P_{kpi}^{MDm}}$, for all atoms in residues. Superimposed region of trajectory are CA-atoms of B11-B17. Reference structure are MS of $\boldsymbol{E_{kpi}^{MD}}$.*

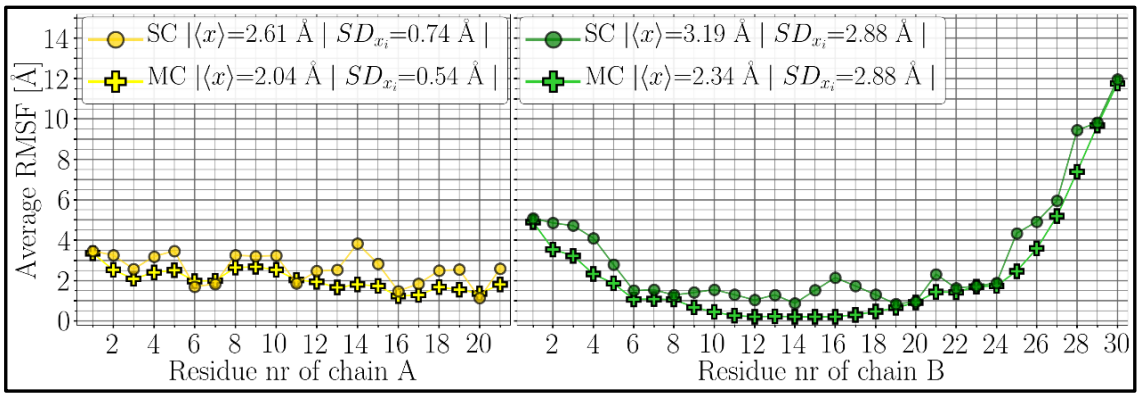


***Figure 6.7**: Average RMSF for insulin structures in ensemble $\boldsymbol{P_{kpi}^{MDm}}$ for all atoms of indicated selection for each residue (i.e. residue-wise $RMSF_{\langle SC \rangle}$, $RMSF_{\langle MC \rangle}$). Superimposed atoms of trajectory are CA-atoms of B11-B17.*

6.2.2 A statistical comparison of intra-monomer calculated hydrogen bonds

An important query is what set of HBs may be the most probable of the insulin monomer in solution. Here the calculated lower angle HBs are shown for insulin from the 9 MD trajectory ensembles in Table 6.2; in particular the more probable ones are depicted for $\mathbf{P_{kpi}^{MDm}}$ in Figure 6.8. Moreover the individual HBs are statistically counted and compared, for the three angle ranges (sorted in matrices in §6.2.5), to other sets of derived individual HBs in Table 6.3. Here obviously there is a vast amount of HBs that are sampled by MD (e.g. $\mathbf{P_{kpi}^{MDm}}$), even if the very most of them are relatively shortlived and intermittent.

The structures sampled at each nanosecond in a MD trajectory are momentous and as such the HBs are transient, with the HB trajectory percentage indicating their probability at any time (e.g. for MS of $\mathbf{P_{kpi}^{MDm}}$ at 508 ns it are 24 individual HBs present).

Plausibly the larger number of sampled HBs covers the majority within immediate conformational space (in comparison to $\mathbf{E_{kpi}^{DGR}}$ and **M12** which are constrained). Nevertheless, it is here seen that there are some correspondence to the other sets of derived HBs, sampled to a varying extent. Significantly less HBs are available in the limited conformational space of $\mathbf{E_{kpi}^{DGR}}$ (c.f. Table 5.1); also interesting is that for MD trajectories, the average number of lower angle HBs at any time is about twice (20.2 for $\varphi < 30°$); however similar values for the higher angle ranges.

By moreover counting the HB presence in all sets of MD ensembles of Table 6.2 (c.f. Table S6.4), some variation is apparent. Even though there is a great similarity among many of the HBs of all the different MD replicas, there is apparently a greater difference due to the variation in starting structures, rather than the change in temperature. In addition, as expected, the higher temperature of 10 Kelvin gives an overall higher amount of sampled HBs. Notwithstanding, $\mathbf{P_{kpi}^{MD}}$ and $\mathbf{E_{kpi}^{MD}}$, gives overall an agreeing picture of the prevalent HBs, albeit their RMSF profile looks slightly different (c.f. Figure 6.7, Figure S6.46). Similar but lesser congruence is found for $\mathbf{P_{i}^{MDno}}$, with some notably different HBs being sampled. Larger deviation is of course for $\mathbf{P_{i}^{MDm}}$ having an irregular conformational change of the NT B-chain; with a greater number of HBs due to sampling a larger part of conformational space. Furthermore, there are some larger number of HBs in agreement for the MD replica ensembles, also when just comparing individual MSs, though some variation in agreement to RHB and **M12** (c.f. Table S6.4).

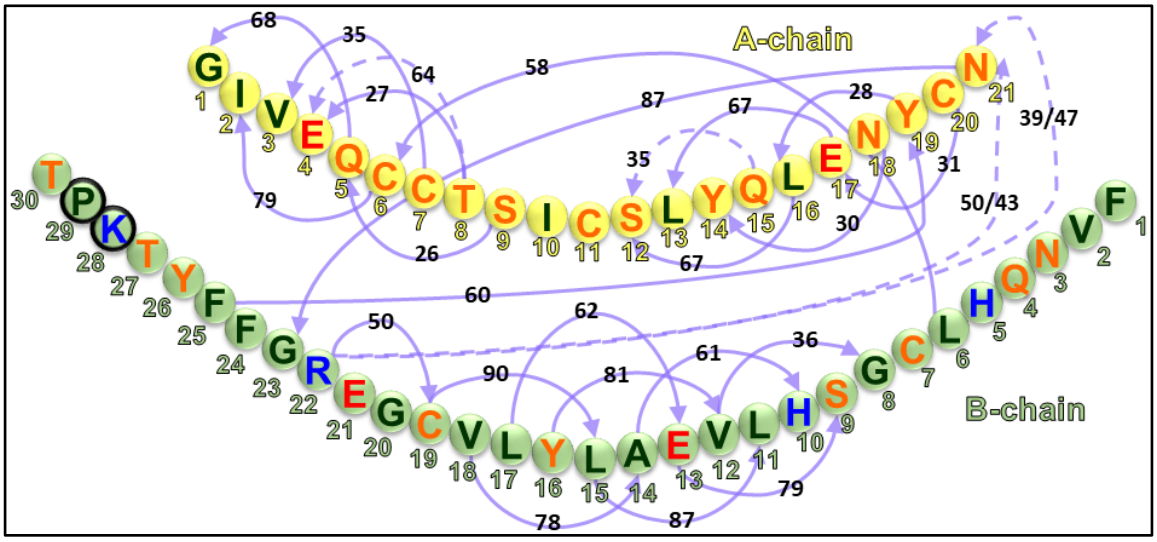


Figure 6.8: *Hydrogen bonds for ensemble* $\boldsymbol{P_{kpi}^{MDm}}$*.The 28 HBs existing between SC or MC for more than 25% of times 9-1499 ns, for a lower angle range (*$\varphi < 30°$*). Note that e.g. "B22 R(HE)→A21 N(OT#)" has presence 50/43 assigning both wildcard atoms (OT1/2), whose average are given in Table 6.2. The HBs in purple "between MC atoms" are represented by whole arrows and "SC to SC" or "MC to SC" by dotted arrows (donor start and acceptor end). These HBs are shown on an akin structure representation in Figure S6.38. These same HBs in time-dependent graphs are included in Figure S6.39, in addition to the other HBs present for more than 5%.*

***Table 6.2**: Intra-monomer HBs for $\boldsymbol{P_{kpi}^{MD}}$, $\boldsymbol{E_{kpi}^{MD}}$, $\boldsymbol{P_{i}^{MD}}$ above 5% presence for low angle ($\varphi < 30°$). Note for wildcards (#) the average of all possible interactions is given, as such the average probability is given here; note however that the HBs with a "#" respectively combine to a larger presence during a trajectory. With the huge amount of HBs below 5% omitted for clarity, even if present in any other MD replica. The HBs in common to RHB and **M12** are shaded with grey, respectively calculated with increasing angle: ($\varphi < 30°$,60°,90°) 4_7_5, 9_10_5, 10_11_5; (shown here for $\varphi < 30°$, same ones shown in grey in Table 5.2). These same (atom-specific) HBs in time-dependent graphs are included in Figure S6.39 for $\boldsymbol{P_{kpi}^{MD}}$.*

Donor (H)	Acceptor	$\mathbf{P_{kpi}^{MDm}}$	$\mathbf{P_{kpi}^{MDn}}$	$\mathbf{P_{kpi}^{MDo}}$	$\mathbf{E_{kpi}^{MDm}}$	$\mathbf{E_{kpi}^{MDn}}$	$\mathbf{E_{kpi}^{MDo}}$	$\mathbf{P_{i}^{MDm}}$	$\mathbf{P_{i}^{MDn}}$	$\mathbf{P_{i}^{MDo}}$
A1 G(H#)	A4 E (OE#)	12	*11	*11	11	*12	*12	13	11	*11
A4 E(HN)	A4 E(OE#)	11	10	*9	9	12	11	13	*10	10
A5 Q(HN)	A1 G(O)	*68	*61	*71	71	*74	*71	64	*70	66
A6 C(HN)	A2 I(O)	*79	*67	*84	*81	*82	*76	*77	*81	*73
A6 C(HN)	A3 I(O)									5
A7 C(HN)	A3 V(O)	35	37	*33	31	*39	*65	*70	*29	56
A8 T(HN)	A3 V(O)									8
A8 T(HG1)	A4 E(O)	*64	*63	*61	*63	*66	*56	*64	*69	56
A8 T(HN)	A4 E(O)	*27	*31	26	28	*35	*30	37	21	27
A8 T(HN)	A5 Q(O)	5	6	8	8	7	7	9	9	
A9 S(HG1)	A5 Q(O)	*23	18	31	*31	19	13	15	*33	6
A9 S(HN)	A5 Q(O)	26	23	25	15	29	*40	43	16	16
A10 I(HN)	A7 C(O)							7		
A11 C(HN)	A6 C(O)							8		
A12 S(HG1)	A15 Q(OE1)		*10		*9		6		8	
A12 S(HN)	A15 Q(OE1)	*16	20	19	*23	12	13		22	
A15 Q(HN)	A12 S(OG)	35	30	*33	*33	*36	28	42	31	35
A15 Q(HE22)	A15 Q(O)				6	6				
A15 Q(HE22)	A19 Y(OH)		7		5	6				
A16 L(HN)	A12 S(O)	*67	*61	*61	*65	*70	*52	*62	*63	*60
A17 E(HN)	A13 L(O)	*67	*75	72	*79	*72	*67	*84	83	*76
A18 N(HN)	A14 Y(O)	30	40	*45	*39	*38	*46	*73	*67	*51
A18 N(HN)	A15 Q(O)	8					5			
A18 N(HD2#)	A15 Q(OE1)	5				5		9	*6	*11
A19 Y(HN)	A15 Q(O)	8	11	10	11	12	10	*40	30	*14
A19 Y(HN)	A16 L(O)	28	22	*24	24	21	*22	5	6	19
A20 C(HN)	A16 L(O)		7	8	6	6	7	34	21	10
A20 C(HN)	A17 E(O)	31	23	*22	26	23	*21		*12	24
A21 N(HN)	A18 N(O)							19	*28	
A21 N(HD2#)	A21 N(OT#)			6						
B1 F (H#)	B4 Q (OE1)				5				5	
B2 V(HN)	B10 H(ND1)						14			
B4 Q(HE2#)	B2 V(O)	10	*13	9	*10	9	6	8	5	
B4 Q(HE2#)	B5 H(O)	5			8				7	
B5 H(HE2)	B26 Y(OH)							*12		
B6 L(HN)	B10 H(ND1)						5			
B7 C (HN)	B5 H(ND1)							6		
B10 H(HN)	B7 C(O)	10	6	7	9	12			5	
B11 L(HN)	B7 C(O)	23	34	26	24	29	*40		27	*42
B11 L(HN)	B8 G(O)	7			6				6	
B12 V(HN)	B8 G(O)	*36	*48	39	*34	*39	60	*66	*34	*67
B13 E(HN)	B9 S(O)	*79	*83	*83	*82	*76	82	72	*84	*79
B14 A(HN)	B10 H(O)	*61	68	*66	*66	66	71	59	*61	*66

B15 L(HN)	B11 L(O)	*87	*85	*89	*86	*87	*86	61	*77	*85
B16 Y(HN)	B12 V(O)	*81	*80	*80	*80	*82	*79	*76	*82	*82
B17 L(HN)	B13 E(O)	*62	*58	61	*62	*66	61	34	44	*62
B18 V(HN)	B14 A(O)	*78	81	*76	*82	*83	*84	*69	*85	*82
B19 C(HN)	B15 L(O)	*90	*93	*92	92	*94	*93	*73	*81	*89
B20 G(HN)	B16 L(O)	*22	22	*28	20	24	22			20
B20 G(HN)	B17 L(O)				*			17	17	5
B22 R(HN)	B19 C(O)	*50	*53	51	51	55	50			*42
B22 R(NE)	B19 C(O)									6
B22 R(HE)	B20 G(O)							51	11	
B22 R(HH21)	B20 G(O)							13		
B23 G(HN)	B19 C(O)				*		6			
B26 Y(HN)	B16 Y(OH)							59	69	11
B26 Y(HN)	B24 F(O)			*7						
B29 L(HN)	B27 T(OG1)									5
B30 T(HG1)	B27 T(O)							9		
B30 T(HG1)	B30 T(OT#)			7	6	*	5			
A3 V(HN)	B26 Y(OH)			8						
A11 C(HN)	B3 N(O)									38
A11 C(HN)	B4 Q(O)	7	*39	8	24	10	8		10	20
A19 Y(HH)	B25 F(O)		12	10	21	16	*33			*14
A19 Y(HH)	B26 F(O)									6
A21 N(HD22)	B22 (O)			5						
A21 N(HN)	B23 G(O)	*87	*89	87	*90	*91	*90			*67
A21 N(HD21)	B25 F(O)		9							
B3 N(HN)	A11 C(O)									17
B4 Q(HN)	A11 C(O)		9		8					
B5 H(HD1)	A7 C(O)	9			11					
B5 H(HN)	A9 S(O)									*39
B5 H(HD1)	A9 S(O)	8			10					
B6 L(HN)	A6 C(O)	*58	62	54	*79	*41	10		*71	61
B8 G(HN)	A7 C(SG)						9			
B22 R(HH#2)	A17 E(OE#)							15	*26	5
B22(HH11)	A20 C(O)							14		
B22 R(HH11)	A21 N(OD1)							27		
B22 R(HE)	A21 N(OT#)	*47	*45	*46	*47	*48	47	5		*33
B22 R(HH21)	A21 N(OT#)	*43	*42	*44	*44	*44	*44	5		31
B25 F(HN)	A19 Y(O)	*60	53	23	*49	*50	26			37
B26 Y(HH)	A4 E(OE#)							12		
B26 Y(HH)	A19 Y(OH)								5	
B29 L(HN)	A4 E(OE#)							6		
B30 T(HN)	A4 E(OE#)					8				
B30 T(OG1)	A4 E(OE#)					7				

Table 6.3 continued.

***Table 6.3**: Calculated and compared intra-monomer HBs of $\boldsymbol{P_{kpi}^{MDm}}$ to those in other systems. The HBs of $\boldsymbol{P_{kpi}^{MDm}}$, $\boldsymbol{E_{kpi}^{DGR}}$ (in Table 5.2) and **M12** (Table 4.1) are respectively calculated with criteria: "$r_{AD} < 3.5$ Å" & "$\varphi < 30°, 60°, 90°$". Note that $\boldsymbol{P_{kpi}^{MDm}}$ compares its HB occupancy at indicated percentage (the -||- symbol designates the same respective range as the indices of 2nd column), showing the overlap to all other HBs, in respective angle range, of sets RHB, $\boldsymbol{E_{kpi}^{DGR}}$, **M12**. That is to say, $\boldsymbol{P_{kpi}^{MDm}}$ compares its HB presence at each indicated percentage to all other sets whose HBs are of presence above 0%. The total nr of HBs are as sorted in "nr in AC_nr in BC_nr between AC&BC", same format for respective nr of overlap to the other sets of HBs. For example, at low angle "$\varphi < 30°$", there are sorted HBs "12_9_7(sum 28)" present in $\boldsymbol{P_{kpi}^{MDm}}$ (i.e. in 25% or more of interval 9-1499 ns), having overlap of "6_7_3" with RHB (i.e. 16 of 31 HBs), "4_6_3" with $\boldsymbol{E_{kpi}^{DGR}}$ (i.e. 13 of 150 HBs), "4_9_3" with **M1** (i.e. 16 of 25) and "7_8_4" with **M2** (i.e. 19 of 26). In contrast, each structure in $\boldsymbol{P_{kpi}^{MDm}}$ (i.e. in each ns of interval 9-1499 ns), has a lower total amount of HBs (c.f. MS at time-frame 508 ns for each angle range), where the average statistical number, at each angle range are: $\langle x \rangle (SD_{x_i})$; 20.2(2.9), 35.9(3.4), 88.8(5.3).*

$\boldsymbol{P_{kpi}^{MDm}}$	Nr HB's (A_B_AB)		Nr HBs = RHB		Nr HBs = $\boldsymbol{E_{kpi}^{DGR}}$		Nr HBs = to **M1**		Nr HBs = to **M2**	
$\sum HB$ > 0%	>= 5%	>= 25%	-||-	-||-	-||-	-||-	-||-	-||-	-||-	-||-
	>= 50%	>= 75%	-||-	-||-	-||-	-||-	-||-	-||-	-||-	-||-
MS	= 100 %		-||-		-||-		-||-		-||-	
$\varphi < 30°$										
-||-	26_15_10	12_9_7	8_9_4	6_7_3	5_6_3	4_6_3	7_9_5	4_9_3	10_8_5	7_8_4
358	5_8_4	1_5_1	2_7_3	0_5_1	1_5_3	0_5_1	2_8_3	1_5_1	3_8_4	1_5_1
MS	9_10_5		4_8_3		2_6_3		3_9_3		5_8_3	
$\varphi < 60°$										
-||-	41_30_19	19_13_7	12_11_5	8_9_3	20_15_4	12_10_3	15_13_7	10_12_4	17_12_6	12_11_4
514	11_11_7	7_9_4	6_9_3	5_8_2	9_9_3	7_8_2	7_11_4	6_9_2	8_11_4	6_9_2
MS	13_14_6		6_8_3		9_9_3		7_11_3		8_10_3	
$\varphi < 90°$										
-||-	83_96_24	50_45_9	12_11_5	10_10_3	43_47_4	35_35_3	36_39_8	31_34_5	42_35_6	35_32_4
746	39_33_7	27_26_4	9_10_3	7_9_2	32_31_3	25_25_2	28_32_4	22_26_2	31_31_4	23_26_2
MS	42_40_8		8_8_3		25_28_3		22_27_3		27_25_3	

6.2.2.1 RHB bounds comparison

As in above section inferred, a vast number of HBs were sampled by MD, moreover part-coverage of the RHBs were found (Table 6.3). Here the distances were separated in the bounds of the respective RHBs (calculated as in §3.3.3.1) and shown in Table 6.4. As an example the atoms and their accumulated violations being depicted in Figure 6.9 for replica $\mathbf{P_{kpi}^{MDm}}$ (individual violations shown in Table S6.5); from whence it is obvious the lesser congruence of RHBs (as compared to $\mathbf{E_{kpi}^{DGR}}$ c.f. Figure 5.4). Overall the best fit is found for $\mathbf{P_{kpi}^{MDm,n}}$ and $\mathbf{E_{kpi}^{MDm}}$.

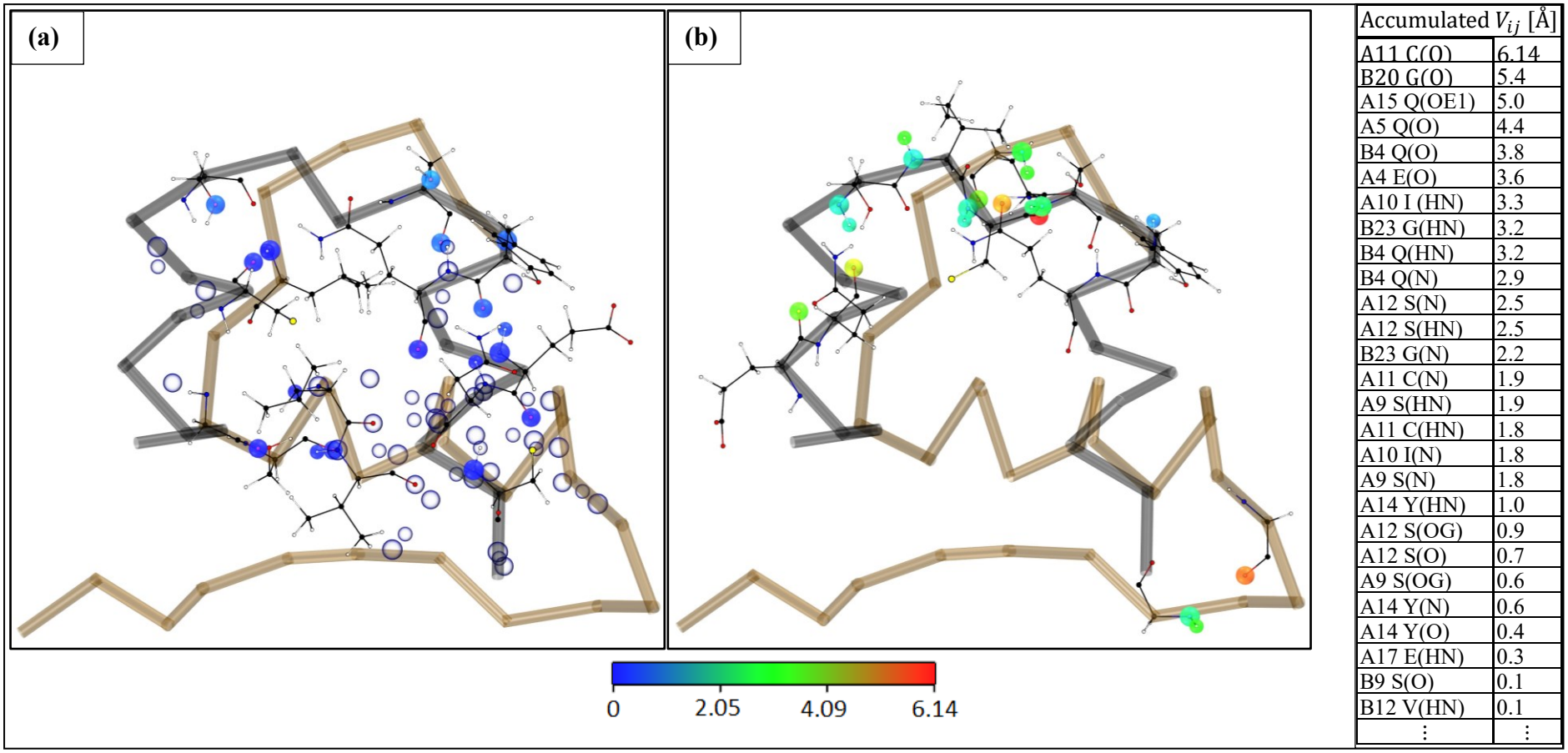
Nevertheless, seen for all MD replicas, are null or minor violations of most of the RHBs in the BC α-helix, also for adjacent RHBs e.g. "B6 L(N&HN) → A6 C(O)" (a larger violation for $\mathbf{E_{kpi}^{MDo}}$ and $\mathbf{P_{i}^{MDm}}$). Hence the MD simulations strengthens a notion, that the BC α-helical region are indeed relatively stable, i.e. in a solution as it is in a crystal (**M12**).

Mostly minor violations are found at the α-helix of the CT A-chain, largest for "A12 N(HN)→A15 N(OE1)". Some other violations are in segment A4-A12, in addition between "B4 Q(N&HN) → A11 C(O)" and "A11 C(NH) → B4 Q(O)" (minor in $\mathbf{E_{kpi}^{DGR}}$). The latter observation may partly be understood from the more fluctuating N-terminal AC and BC, c.f. Figure 6.7 (assumed constrained in $\mathbf{E_{kpi}^{DGR}}$ of Figure 5.2d).

However, some intra-chain RHBs are sampled well, which have minor or null violation for $\mathbf{P_{kpi}^{MD}}$, $\mathbf{E_{kpi}^{MD}}$; i.e. "A21 N(HN)->B23 G(O)" (larger for $\mathbf{P_{i}^{MD}}$) and "B25 N(HN) → A19 Y(O)" (larger for $\mathbf{P_{kpi}^{MDo}}$, $\mathbf{P_{i}^{MD}}$). In addition, a large violation of "B23 G(HN) → B20 G(O)" (c.f. Figure 6.9), which is explained by the peptide bond of B22-B23 flipping away, breaking this HB over most of the time. Moreover, for the MD replicas a large amount of HBs are accessible (c.f. Table 6.3, Table 6.2), transiently breaking and reforming. Hence indicating a range of possible HB distributions, depending on which parts of the energy conformational landscape being sampled by MD; influenced especially by varying starting structure.

***Table 6.4**: Calculated D & DH to A distances separated in RHB bounds , given in percentage and number. The 31 RHBs (62 with D & DH to A distance bounds) are from PDB 2KJJ. Format given at the top, where the lines are for separation in bounds. For example, of $\boldsymbol{E_{kpi}^{DGR}}$ there are 3 violations (0-1 Å) between atoms i.e. 4.84% of the 62 restraints, that have an average violation of 0.0020 Å.*

Ense-mble	NOE bounds % (nr)			$\langle V_{ij} \rangle$ [Å]
	$\langle r_{ij} \rangle < LB$	$LB < \langle r_{ij} \rangle < UB$	$UB < \langle r_{ij} \rangle$	
	Amount of V_{ij} % (nr)			
	0-1 Å	1-2 Å	2-3 Å	> 3 Å
$\mathbf{E_{kpi}^{DGR}}$	14.52% (9)	80.65% (50)	4.84% (3)	0.0020
	4.84% (3)	0% (0)	0% (0)	0% (0)
$\mathbf{P_{kpi}^{MDm}}$	0.00% (0)	58.06% (36)	41.94% (26)	0.5077
	22.58% (14)	8.06% (5)	8.06% (5)	3.23% (2)
$\mathbf{P_{kpi}^{MDn}}$	0.00% (0)	54.84% (34)	45.16% (28)	0.4189
	27.42% (17)	12.90% (8)	3.23% (2)	1.61% (1)
$\mathbf{P_{kpi}^{MDo}}$	0.00% (0)	53.23% (33)	46.77% (29)	0.5716
	24.19% (15)	11.29% (7)	6.45% (4)	4.84% (3)
$\mathbf{E_{kpi}^{MDm}}$	0.00% (0)	56.45% (35)	43.55% (27)	0.4112
	27.42% (17)	9.68% (6)	4.84% (3)	1.61% (1)
$\mathbf{E_{kpi}^{MDn}}$	0.00% (0)	53.23% (33)	46.77% (29)	0.5969
	27.42% (17)	4.84% (3)	9.68% (6)	4.84% (3)
$\mathbf{E_{kpi}^{MDo}}$	0.00% (0)	53.23% (33)	46.77% (29)	0.7907
	24.19% (15)	4.84% (3)	9.68% (6)	8.06% (5)
$\mathbf{P_{i}^{MDm}}$	0.00% (0)	40.32% (25)	59.68% (37)	2.0449
	30.65% (19)	6.45% (4)	3.23% (2)	19.35% (12)
$\mathbf{P_{i}^{MDn}}$	0.00% (0)	48.39% (30)	51.61% (32)	0.8503
	25.81% (16)	8.06% (5)	11.29% (7)	6.45% (4)
$\mathbf{P_{i}^{MDo}}$	0.00% (0)	53.23% (33)	46.77% (29)	0.5800
	24.19% (15)	12.90% (8)	4.84% (3)	4.84% (3)



Accumulated V_{ij} [Å]	
A11 C(O)	6.14
B20 G(O)	5.4
A15 Q(OE1)	5.0
A5 Q(O)	4.4
B4 Q(O)	3.8
A4 E(O)	3.6
A10 I (HN)	3.3
B23 G(HN)	3.2
B4 Q(HN)	3.2
B4 Q(N)	2.9
A12 S(N)	2.5
A12 S(HN)	2.5
B23 G(N)	2.2
A11 C(N)	1.9
A9 S(HN)	1.9
A11 C(HN)	1.8
A10 I(N)	1.8
A9 S(N)	1.8
A14 Y(HN)	1.0
A12 S(OG)	0.9
A12 S(O)	0.7
A9 S(OG)	0.6
A14 Y(N)	0.6
A14 Y(O)	0.4
A17 E(HN)	0.3
B9 S(O)	0.1
B12 V(HN)	0.1
⋮	⋮

Figure 6.9: *Visualized RHB violations for* $\boldsymbol{P_{kpi}^{MDm}}$. ***(a)*** *Atoms (D, DH and A) of accumulated* V_{ij} *of 0-1 Å (8 largest of 18 shown in table at right).* ***(b)*** *Atoms (D, DH and A) of accumulated* V_{ij} *above 1 Å (all shown in table at right). The bigger atoms with colour-bar depicts which has accumulated* V_{ij} *(the atoms of RHB having no violation, i.e. 0 Å, are in transparent blue). The whole residues are shown with smaller atoms in ordinary atom-colouring, if at least one atom has a violation (in respective range of 0-1 Å or above 1 Å). Note that an acceptor atom can have large accumulated violations due to two restraints assigned to it, which is the case for "A11 C(O)" of bond "B4 Q(N&HN)->A11 C(O)" having the largest value of 6.14 Å. Lines between restrained atoms not shown. Showing transparent traced CA atoms with AC black and BC brown (with curvature at CA-atoms). Depictions shown on MS (i.e. at time-frame 508 ns).*

6.2.3 Calculated NOEs of MD replicas and RHH bounds comparison

This section undertakes NOE calculations and experimental comparison as outlined in §3.3.3 (in particular in §3.3.3.1 and §3.3.3.2).

6.2.3.1 Number of NOEs

The number of calculated NOEs for the MD replicas and $\mathbf{E_{kpi}^{DGR}}$ are shown in Table 6.5, with different R averaging included for comparison. Obviously, by increasing the variable "a" of Eq. 3.9, it increasingly takes a smaller fraction of the insulin structure ensembles, to fulfil, that a hydrogen pair distance to be averaged as less than 5.5 Å. The MD replicas increases more in NOEs compared to $\mathbf{E_{kpi}^{DGR}}$, apparently due to the sampling of the more extensive conformational space. With the $\langle r_{ij}^{-6}\rangle^{-1/6}$ averaging it were calculated from e.g. $\mathbf{P_{kpi}^{MDm}}$ 6455 NOEs, i.e. 8.88% of all 72390 hydrogen-pairs of insulin; considering also the tethered dynamics of the protein. Hence the MD replicas, as for $\mathbf{E_{kpi}^{DGR}}$, predict a much larger number of NOEs than those derived from experiment (as were noted in §5.2.3.1).

Table 6.5*: Number of calculated NOEs from MD replicas, for different R averaging (Equation 3.9, 3.10). The experimental NOEs included from PDB 2KJJ being 793 (omitting glycine with unassigned HA# atoms out of the 803 reported).*

	$\mathbf{E_{kpi}^{DGR}}$	$\mathbf{P_{kpi}^{MDm}}$	$\mathbf{P_{kpi}^{MDn}}$	$\mathbf{P_{kpi}^{MDo}}$	$\mathbf{E_{kpi}^{MDm}}$	$\mathbf{E_{kpi}^{MDn}}$	$\mathbf{E_{kpi}^{MDo}}$	$\mathbf{P_{i}^{MDm}}$	$\mathbf{P_{i}^{MDn}}$	$\mathbf{P_{i}^{MDo}}$
$\langle r_{ij}^{-6}\rangle^{-1/6}$	5458	6455	6546	6471	6356	6579	6522	5843	6120	6759
$\langle r_{ij}^{-3}\rangle^{-1/3}$	5125	4418	4312	4387	4551	4400	4276	4033	4575	4526
$\langle r_{ij}^{-1}\rangle^{-1/1}$	4892	3642	3569	3620	3693	3545	3453	3469	3764	3583
$\langle r_{ij}\rangle$	4701	3104	3117	3136	3263	3095	3084	3004	3240	3071

6.2.3.2 Matrices overlap of calculated NOEs and RHHs

Here comparing the overlapping intra-monomer NOEs of $\mathbf{P_{kpi}^{MDm}}$ with those of $\mathbf{E_{kpi}^{DGR}}$ and its restraints (RHH); shown in Figure 6.10. From the overlap matrix with RHH, it can be seen, that almost all residue-pairs have some overlap, excepting a few ones not counted as a calculated NOE ($\langle r_{ij}^{-6}\rangle^{-1/6}$ > 5.5 Å), however having a fair congruence: "A11 C(HN) - B5 H(HB#)" (V_{ij} is 0.61 Å of UB 6.50) and "B5 H(HN) - B3 N(HA)" ($\langle r_{ij}^{-6}\rangle^{-1/6}$ is 5.67 albeit WBs) and "A3 N(HN) - B27 T(HN)" (V_{ij} is 0.17 Å of UB 5.50 Å). Overall evident is again

the plausibly over-restrained structure ensemble of $\mathbf{E_{kpi}^{DGR}}$, which has another distribution of calculated NOEs than what is sampled from $\mathbf{P_{kpi}^{MDm}}$; however the overlap is striking. Comparatively it may be seen the overlap with NOEs from $\mathbf{P_{i}^{MDm}}$ (in Figure S6.51), since this structure ensemble were shapeshifting from a T-state alike conformation after 60-120 ns, and are indeed showing less congruence with the RHHs. Furthermore, that the other replicas have an insulin MS more alike a T-state conformation, are apparently related to a better congruence with the RHH.

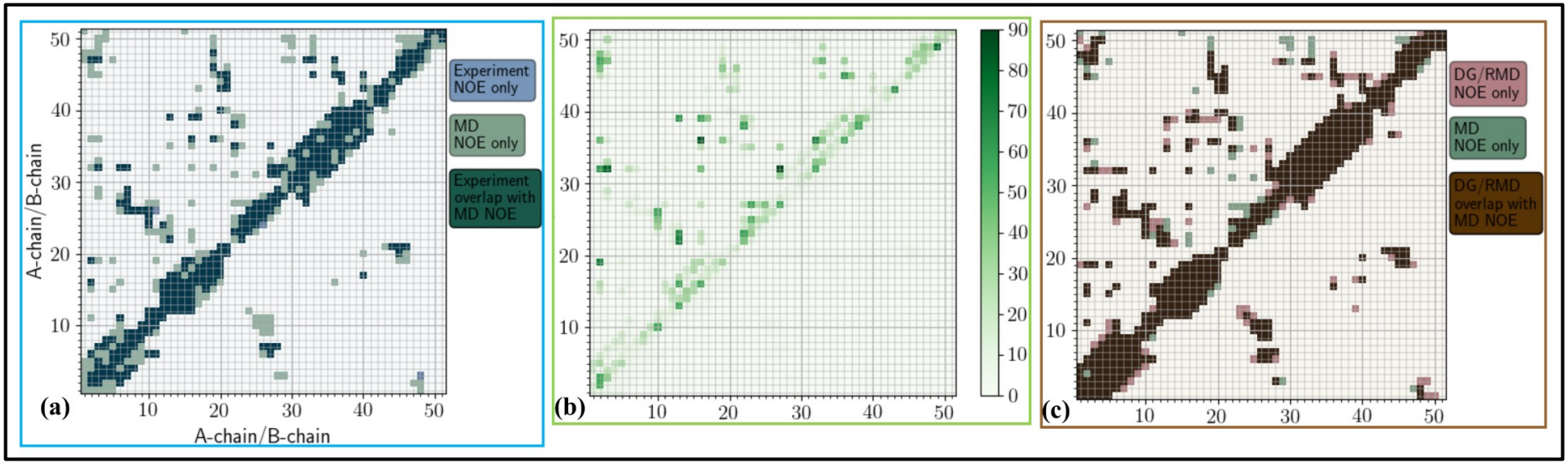


***Figure 6.10**: Overlap of experimental RHHs, calculated NOEs of $\boldsymbol{P_{kpi}^{MDm}}$ and $\boldsymbol{E_{kpi}^{DGR}}$. **(a)** The overlap of RHHs and calculated NOEs of $\boldsymbol{P_{kpi}^{MDm}}$. **(b)** Calculated NOEs of $\boldsymbol{P_{kpi}^{MDm}}$, the colour-bar numbering shows the counts of NOEs for any matrix-index. **(c)** Overlap between calculated NOEs of $\boldsymbol{P_{kpi}^{MDm}}$ and $\boldsymbol{E_{kpi}^{DGR}}$. Experimental RHHs are from PDB entry 2KJJ. Matrices is obtained as explained in §3.3.3.2. The overlapped comparison of matrices is done by setting first all indices that have any NOE to 1, respectively for each matrix, then overlapping, separating colours and plotting with matplotlib* [280]. *Hence the overlap does not mean that all specific hydrogen pair NOEs agree between residue pairs. Note that here residues are numbered in NT to CT direction as AC (1 to 21) continuing to BC (22 to 51).*

6.2.3.3 Calculated NOEs separated in RHH bounds

As elaborated in §3.3.3.1, here it is separated into bounds the RHH congruent NOEs; calculated from the MD obtained insulin structure ensembles (see Table 6.6). Apparent is that most of the calculated NOEs are within the LB and UB (if below LB only slightly less); however a considerable fraction is above the UB, i.e. violations (Eq. 3.11). To illustrate for one replica ($\mathbf{P_{kpi}^{MDm}}$), a depiction are shown in Figure 6.11, displaying the accumulated violations of any hydrogen (c.f. Table S6.5). Interestingly, smaller violations are spread out at hydrogens that are closer to the core and around the α-helices, moreover lesser at e.g. the C-terminal BC (B20-30). The largest violations mostly involve SC atoms, e.g. for "B10 H(HD2)" arising mainly from assignments to "B9 S(HA)" and "B11 L(HD11&HD21)", "B12 V(HG11&HG21)". Some other larger violations are between "B13 E(HA&HG1)" and "B12 V(HG21)", moreover between "A12 S(HN)" and "B3 N(HB1)".

Though there are common violations, the lesser number and accumulation of violations for $\mathbf{E_{kpi}^{DGR}}$ is understandable, being restrained to RHH (c.f. §5.2.3).

Apparently as with the congruence with RHBs, here $\mathbf{E_{kpi}^{MDo}}$ and $\mathbf{P_{i}^{MDn}}$ show the second most violations of the UBs (having higher variability in amount but also common RHH violations); which in all likelihood implies that these are not the most representative, out of the nine considered MD replicas. However, for 6 of the replicas it shows a similar though varying distribution among the bounds, with the average violation ($\langle V_{ij} \rangle$), being of similar magnitude.

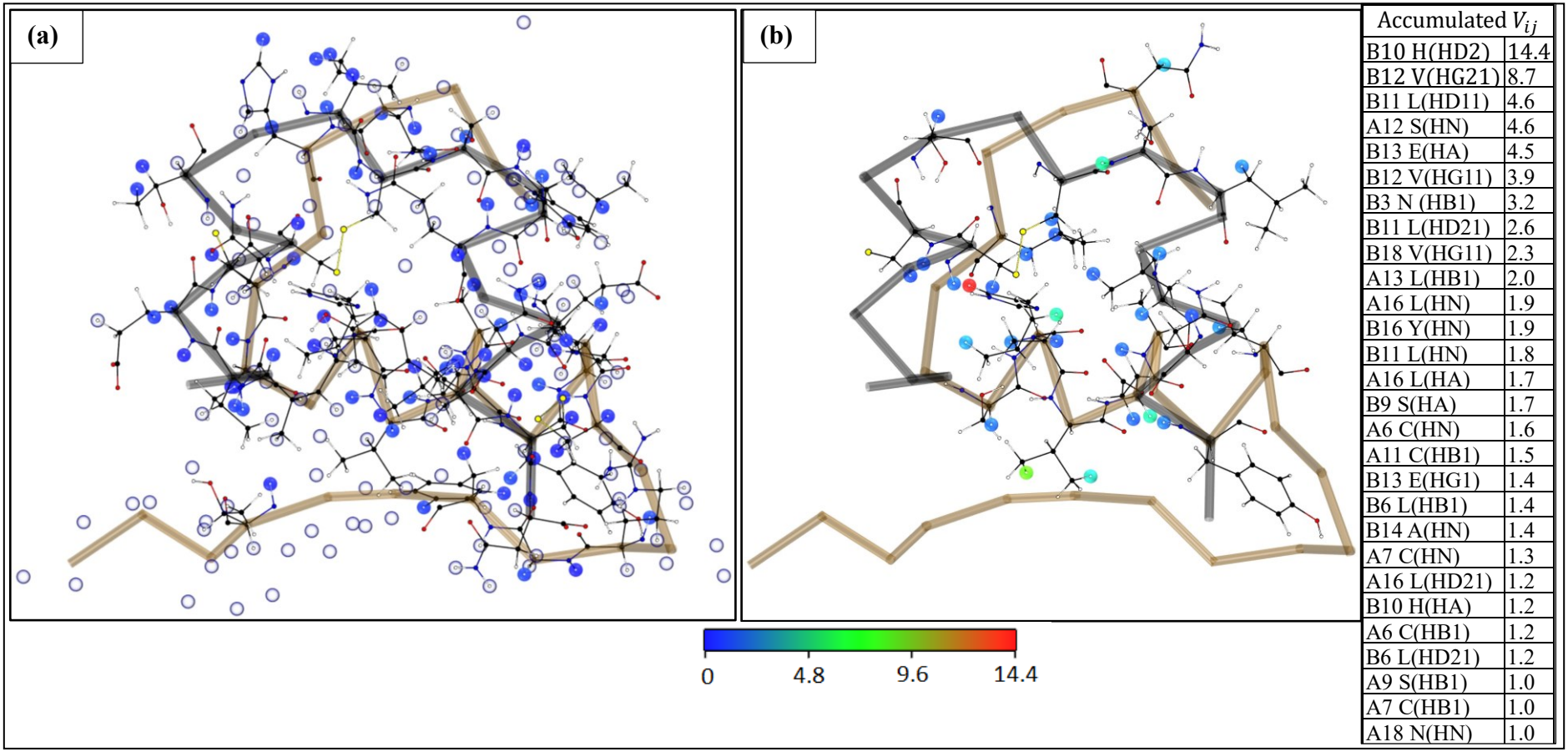
The worst agreement of the bounds is obviously for $\mathbf{P_{i}^{MDm}}$, for which only about 70.37% are WBs and the violations are many and large (c.f. Figure S6.52); being understood by the sampling of additional conformational space. An apparently similar and analogous comparison of MD structure ensembles of non-native nature by Zagrovic *et al.* [220], also gave relatively similar congruence with bounds determined for respective native structures. That is in particular of a shapeshifting protein "Villin" (36 residues; 474 hydrogen restraints [281]; $\langle r_{ij}^{-6} \rangle^{-1/6}$ averaging; 76.6% WB) and another high temperature denatured state of "Lysozyme" (129 residues; 1632 hydrogens [282]; $\langle r_{ij}^{-3} \rangle^{-1/3}$ averaging; 73.1% WB); presumedly a better agreement if compared to a MD ensemble of native structures. Hence, Zagrovics study illustrated that it may be relatively easy to match NOE bounds even with non-native MD structure ensembles [220, 283].

Nevertheless, here the congruence of the averagely T-state structures (e.g. $\mathbf{P_{kpi}^{MDm}}$), is still

remarkable, suggesting that the MD simulation methodology captures structures that fits the restraints quantitively well; however with the above noted uncertainty.

***Table 6.6**: Calculated NOEs of structure ensembles divided in RHH bounds, given in percentage and number. The restraints are from PDB 2KJJ of 793 experimental NOEs.*

Ense-mble	NOE bounds % (nr) $\langle r_{ij}^{-6}\rangle^{-1/6} < LB$	$LB < \langle r_{ij}^{-6}\rangle^{-1/6} < UB$	$UB < \langle r_{ij}^{-6}\rangle^{-1/6}$	$\langle V_{ij}\rangle$ [Å]
NOE nr	Amount of V_{ij} % (nr) 0-1 Å	1-2 Å	2-3 Å	> 3 Å
$\mathbf{E_{kpi}^{DGR}}$	0.13% (1)	86.76% (688)	13.11% (104)	0.0412
5458	12.86% (102)	0.25% (2)	0% (0)	0% (0)
$\mathbf{P_{kpi}^{MDm}}$	0.00% (0)	87.39% (693)	12.61% (100)	0.0663
6455	10.97% (87)	0.88% (7)	0.50% (4)	0.25% (2)
$\mathbf{P_{kpi}^{MDn}}$	0.00% (0)	87.01% (690)	12.99% (103)	0.0705
6546	11.10% (88)	1.26% (10)	0.38% (3)	0.25% (2)
$\mathbf{P_{kpi}^{MDo}}$	0.13% (1)	86.38% (685)	13.49% (107)	0.0716
6471	11.48% (91)	1.13% (9)	0.63% (5)	0.25% (2)
$\mathbf{E_{kpi}^{MDm}}$	0.00% (0)	84.87% (673)	15.13% (120)	0.0724
6356	13.11% (104)	1.39% (11)	0.38% (3)	0.25% (2)
$\mathbf{E_{kpi}^{MDn}}$	0.00% (0)	85.25% (676)	14.75% (117)	0.0783
6579	12.61% (100)	1.39% (11)	0.5% (4)	0.25% (2)
$\mathbf{E_{kpi}^{MDo}}$	0.13% (1)	83.10% (659)	16.77% (133)	0.1119
6522	13.24% (105)	2.40% (19)	0.63% (5)	0.5% (4)
$\mathbf{P_{i}^{MDm}}$	0.38% (3)	70.37% (558)	29.26% (232)	0.4029
5843	15.89% (126)	5.55% (44)	3.53% (28)	4.29% (34)
$\mathbf{P_{i}^{MDn}}$	0.38% (3)	80.45% (638)	19.17% (152)	0.1468
6120	14.63% (116)	2.40% (19)	1.26% (10)	0.88% (7)
$\mathbf{P_{i}^{MDo}}$	0.38% (3)	84.24% (668)	15.38% (122)	0.0820
6759	13.62% (108)	0.88% (7)	0.63% (5)	0.25% (2)



Accumulated V_{ij}	
B10 H(HD2)	14.4
B12 V(HG21)	8.7
B11 L(HD11)	4.6
A12 S(HN)	4.6
B13 E(HA)	4.5
B12 V(HG11)	3.9
B3 N (HB1)	3.2
B11 L(HD21)	2.6
B18 V(HG11)	2.3
A13 L(HB1)	2.0
A16 L(HN)	1.9
B16 Y(HN)	1.9
B11 L(HN)	1.8
A16 L(HA)	1.7
B9 S(HA)	1.7
A6 C(HN)	1.6
A11 C(HB1)	1.5
B13 E(HG1)	1.4
B6 L(HB1)	1.4
B14 A(HN)	1.4
A7 C(HN)	1.3
A16 L(HD21)	1.2
B10 H(HA)	1.2
A6 C(HB1)	1.2
B6 L(HD21)	1.2
A9 S(HB1)	1.0
A7 C(HB1)	1.0
A18 N(HN)	1.0

***Figure 6.11**: Visualized RHH UB violations of $\boldsymbol{P_{kpi}^{MDm}}$. **(a)** Hydrogens of accumulated V_{ij} of 0 or less than 1.0 Å. **(b)** Hydrogens of accumulated V_{ij} equal or above 1.0 Å (shown in table at right). The bigger hydrogens with colour-bar depicts which has accumulated V_{ij} (the hydrogens of RHH having no violation, i.e. 0 Å, are in transparent blue). The whole residues are shown with smaller atoms in ordinary atom-colouring, if at least one hydrogen has a violation (in respective range in (a) and (b)). The atom in red "B10 H(HD2)" shows the largest accumulated V_{ij} of 14.4 Å. Analogous MD simulations of 6 replicas with B5, B10 protonated show more or less same (also for "B10 H(HD2)" but varying violations and higher at "B5 H (HD2&HE1)". Note that a violation (V_{ij}) is added to both hydrogens of a restraint. Transparent bb with AC black, BC brown (curvature at CA atoms). Depictions shown on MS (time-frame at 508 ns).*

6.2.4 RDA bounds comparison

The insulin structure ensembles from respective MD replica has unrestrained motion (apart from the physically geometric force-field); observed especially for the DAs in a way that reflects the more or less fervid motion of respective residue's SC and MC atoms (c.f. Figure 6.6 and Figure 6.7). The time-dependent DAs is a feature for depicting any structural change in a residue at any time of a MD trajectory; hence these are included for $\mathbf{P_{kpi}^{MDm}}$ in Figure S6.37.

Furthermore, here it is compared how well the DAs compare to the respective 47 RDAs (as described in §3.3.3.3), see Figure 6.12. The better congruence for $\mathbf{E_{kpi}^{DGR}}$ are obviously due to being restrained to the bounds of RDAs (c.f. §5.2.4), whereas for the MD replicas having distinct trajectories, are also seen to a fair extent satisfy the RDAs (see Table 6.7). Interestingly, the DAs outside of RDA bounds are predominantly of residues at the more flexible regions and those exposed to solvent. Here $\mathbf{E_{kpi}^{MDo}}$ and $\mathbf{P_{i}^{MDm}}$ shows the least fraction of DAs that are WBs of the RDAs. An example of an interesting DA to note is the "B24 F(χ_1) ~60°", being satisfied (for all except $\mathbf{P_{i}^{MDm,n}}$); apparently pivotal for making the SC fit in the hydrophobic core.

Table 6.7*: Summed fraction of DAs within RDA bounds for DGR and MD insulin structure ensembles . The RDAs are those from PDB 2KJJ. The amount, "f", is obtained by adding the fraction, "f", of ensemble structures satisfying each respective RDA (c.f. Figure 5.7, Figure 6.12).*

	$\mathbf{E_{kpi}^{DGR}}$	$\mathbf{P_{kpi}^{MDm}}$	$\mathbf{P_{kpi}^{MDn}}$	$\mathbf{P_{kpi}^{MDo}}$	$\mathbf{E_{kpi}^{MDm}}$	$\mathbf{E_{kpi}^{MDn}}$	$\mathbf{E_{kpi}^{MDo}}$	$\mathbf{P_{i}^{MDm}}$	$\mathbf{P_{i}^{MDn}}$	$\mathbf{P_{i}^{MDo}}$
$\sum f$	41.30	36.71	35.40	35.56	36.98	35.56	32.84	32.07	35.74	34.89

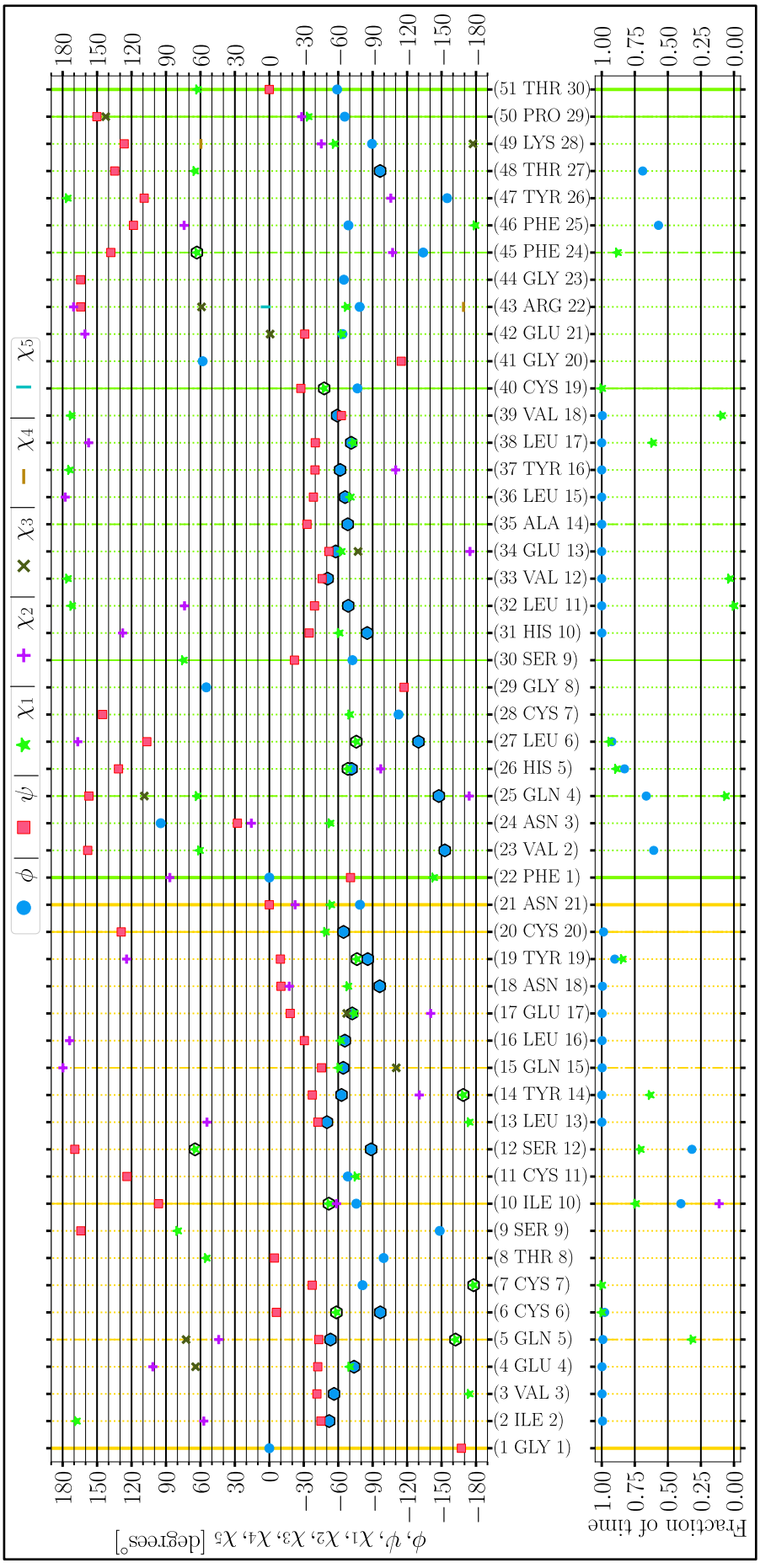


***Figure 6.12**: Dihedral angles and comparison of $\mathbf{P}_{\mathbf{kpi}}^{\mathbf{MDm}}$ to RDAs. Shown are the insulin mean structure DAs, the ones in circled black hexagons are WBs of RDA. Lower graph depicts the fraction of time, "f", the respective restraint is fulfilled during the trajectory, sum of all, "$\sum f$", being 36.71 (47 if "f" could be 1 of all bounded DAs).*

6.2.5 Conformational analytical overview and residue-profiles of $\mathbf{P_{kpi}^{MDm}}$

Since the MD replicas are highly flexible and dynamic, having independent trajectories that are varyingly congruent (in a way reflected by their respective RMSFs and HBs). However, here is provided a structural overview of 1 out 9 replicas showing the most optimal congruence with other replicas and experimental restraints. Only a selection of the 51 residues are included below; however the reader can infer some structure themselves, if wanting to figure out the specific structure for any particular residue. Much aspects of $\mathbf{P_{kpi}^{MDm}}$ are already presented in the previous sections; moreover, here the insulin structure are shown with chain-numbering (Figure 6.13), and with average and time-dependent distances between residue-moieties (Figure 6.14; Figure S6.41). The atom-specific HBs presented here are the same as in §6.2.2; however here sorted in HB matrices for various percentages of trajectory (9-1499 ns) and three divisions of angle-ranges: $\varphi < 30°$, in Figure S6.42 (counting atom-specific HBs whereas averages of wildcards are in Table 6.2); $\varphi < 60°$, in Figure 6.15; $\varphi < 90°$, in Figure 6.16. Again the time-dependent and residue-specific DAs and strongest lower-angle HBs are in Figure S6.37, Figure S6.39 respectively. Analogous depictions are given for the other insulin structure models of **M12** (§4.1.2.4), $\mathbf{CF_{(A,B)}^{6HN5}}$ (§4.2.2) and $\mathbf{E_{kpi}^{DGR}}$ (§5.2.5), providing a direct comparison.

(A1 G): Here this residue are stabilized by the low-angle HB "$Q_{HN}^{A5} \rightarrow G_{O}^{A1}$; 68 % presence; $\boldsymbol{\varphi} < 30°$", moreover by a strong saltbridge "$G_{H\#}^{A1} \rightarrow E_{OE\#}^{A4}$; estimated ~70% accumulated presence; $\boldsymbol{\varphi} < 30°$", (interestingly these two interactions are seen in **M12** albeit perturbed in $\mathbf{CF_{(A,B)}^{6HN5}}$ and in contrast much less present in $\mathbf{E_{kpi}^{DGR}}$). One may note how relatively fluctuating distances are to E^{A4}, Q^{A5}, N^{A21}, T^{B30} etc (Figure S6.41). In particular, the distance to T^{B30} are highly fluctuating with an average distance being off the matrix depiction (c.f. $\langle r_{(CA,CA)}^{A1,B30} \rangle$ is 13.2 Å i.e. beyond 10 Å in Figure 6.14).

(A2 I): Here this residue have a low-angle HB "$C_{HN}^{A6} \rightarrow I_{O}^{A2}$; 79 % presence; $\varphi < 30°$" (same HB in **M12** of Table 4.1). Whose SC is also seen to be flexible as noted in the RMSF (Figure 6.7), being in the core there are some similar contacts as in **M12** whose RMSF (Figure 4.1) is more constrained. In comparison, it is seen in the time-dependent distances to Y^{A19}, L^{B11}, L^{B15}, F^{B25}, T^{B27} (Figure S6.41) and DAs (Figure S6.37), that there are quite some flexibility in the core.

(A21 N): The CT carboxy group appears important in negative cooperativity. Interestingly there is a saltbridge "$R^{B22}_{HE\&HH21} \rightarrow N^{A21}_{OT\#}$", seen for most of the other replicas (seen also in **M12,** and close but not within criteria in $\mathbf{CF^{6HN5}_{(A,B)}}$, and minorly in $\mathbf{E^{DGR}_{kpi}}$); appearing to stabilize the orientation of the SC and MC (in an akin conformation as found in binding surfaces of **M12** and $\mathbf{CF^{6HN5}_{(A,B)}}$, but more varying for structures in $\mathbf{E^{DGR}_{kpi}}$). Interestingly the MC HB "$N^{A21}_{HN} \rightarrow G^{B23}_{O}$" is most of time present (even in **M12** and $\mathbf{E^{DGR}_{kpi}}$), plausibly making nearby region more stable.

(B4 Q): At the NT BC, a quite mobile residue, staying still somewhat close to core, here $\langle r^{A11,B4}_{(MC,MC)} \rangle$ is 6.9 Å (between 6.5-7.0 Å); reflecting that the HBs "$C^{A11}_{HN} \rightarrow Q^{B4}_{O}$; $Q^{B4}_{HN} \rightarrow C^{A11}_{O}$" are indeed seen for higher angles ($\varphi < 60$ & 90°) though only present for less than 10 %.

(B6 L): The SC indeed functions as a hinge-point, overall showing a good fit into the hydrophobic pocket, with minor fluctuations in its position (Figure 6.7). Even an intermittent breakage of the low-angle HB "$L^{B6}_{HN} \rightarrow C^{A6}_{O}$; 58% presence; $\varphi < 30$°" (c.f. Figure S6.39), as in the other replicas deviating mostly for $\mathbf{E^{MDo}_{kpi}}$, $\mathbf{P^{MDm}_{i}}$ with 10%, 0.5% presence respectively (Table 6.2). This observation, may point to an amenability of forming an R-state (in allosteric hexamer forms), where the HB "$L^{B6}_{HN} \rightarrow C^{A6}_{O}$" are not formed, and may be partly explaining the improbable transition of B1-8 in $\mathbf{P^{MDm}_{i}}$.

(B22 R): Interestingly the SC does not have a saltbridge with E^{B21}_{SC} (c.f. $\langle r^{B21,B22}_{(SC,SC)} \rangle$ is 8.8 Å), due to these residue being in a loop with side-chains oriented away. Where the SC are rather towards the polar/charged residue-moieties of A17-21, in particular there is a saltbridged HB "$R^{B22}_{HH21\&HE} \rightarrow E^{A21}_{OT\#}$" (see stability in Figure S6.39). Appearing like longer range charged interactions with E^{B17}_{SC}, albeit there are zero HBs present above 5% (Figure 6.16a), compare with some 9 unique HBs very minorly present (0.13-2.75%) between SC to SC and/or MC "$R^{B22}_{HH\#\#} \rightarrow E^{A17}_{OE\#\&O}$; $\varphi < 90$" (Figure S6.43e).

(B25 F): The RMSF (Figure 6.7) and time-dependent DAs (Figure S6.37) of this residue reveals a highly mobile residue. There is the anticipated MC HB "$F^{B25}_{HN} \rightarrow Y^{A19}_{O}$; 60% presence; $\varphi < 30$", albeit transient (Figure S6.39), it is a contributing hinge HB of the BC CT strand (c.f. $\langle r^{A19,B25}_{(MC,MC)} \rangle$ is 5.9 Å). During the trajectory, F^{B25}_{SC} are very mobile, however it is seen in the MS (Figure 6.13) to be pointing outwards, alike for $\mathbf{E^{DGR}_{kpi}}$ (Figure 5.8) and **M12** (Figure 4.3).

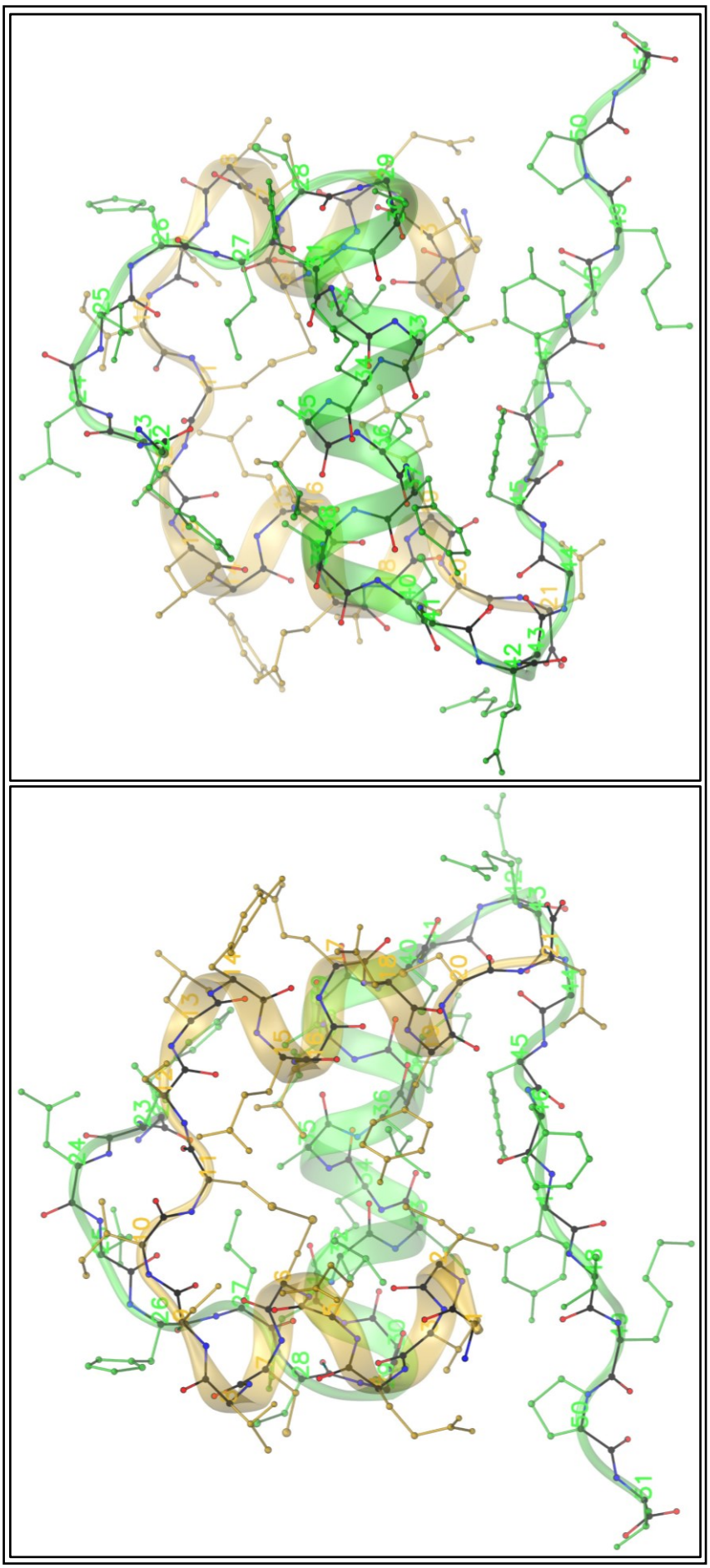


***Figure 6.13**: Structure and numbering for mean structure of ensemble $\mathbf{P}_{kpi}^{MDm}$, i.e. the MS at time-frame 508 ns. **(a)** Front, **(b)** back, (front rotated sideways 180°). The hydrogens are omitted for clarity, moreover the AC and BC being transparent, and the BB are chalky and in ordinary atom colouring, and SCs have chain-colour as "edgy-glassy". The chain-wise residue-numbering is at right of each of the CA-atoms.*

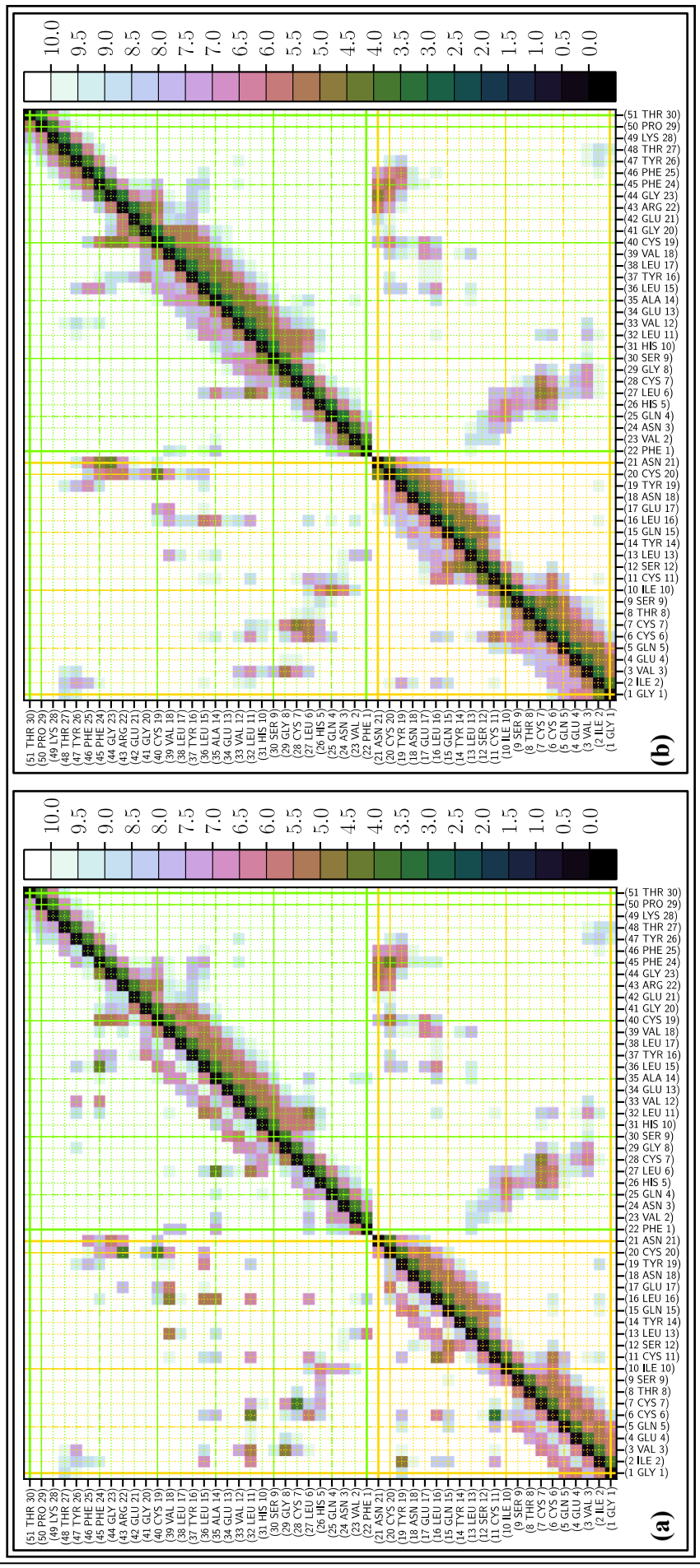


***Figure 6.14**: Average residue-moiety distances within 10 Å of* $\mathbf{P}_{kpi}^{\mathbf{MDm}}$. ***(a)** Upper left is SC to SC geometric centre distances. Lower right is CA to CA-atom distances. **(b)** Upper left is SC to MC geometric-centre distances. Lower right is MC to MC geometric-centre distances. Distances divided in 0.5 steps, c.f. the CA-atom distances of adjacent residues are between 3.5-4.0 Å. Chains residues re-numbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. grid-lines for AC (1-21) in gold-yellow and BC (22-51) lawn-green. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Diagonal as reference 0 Å (black), above 10 Å in white.*

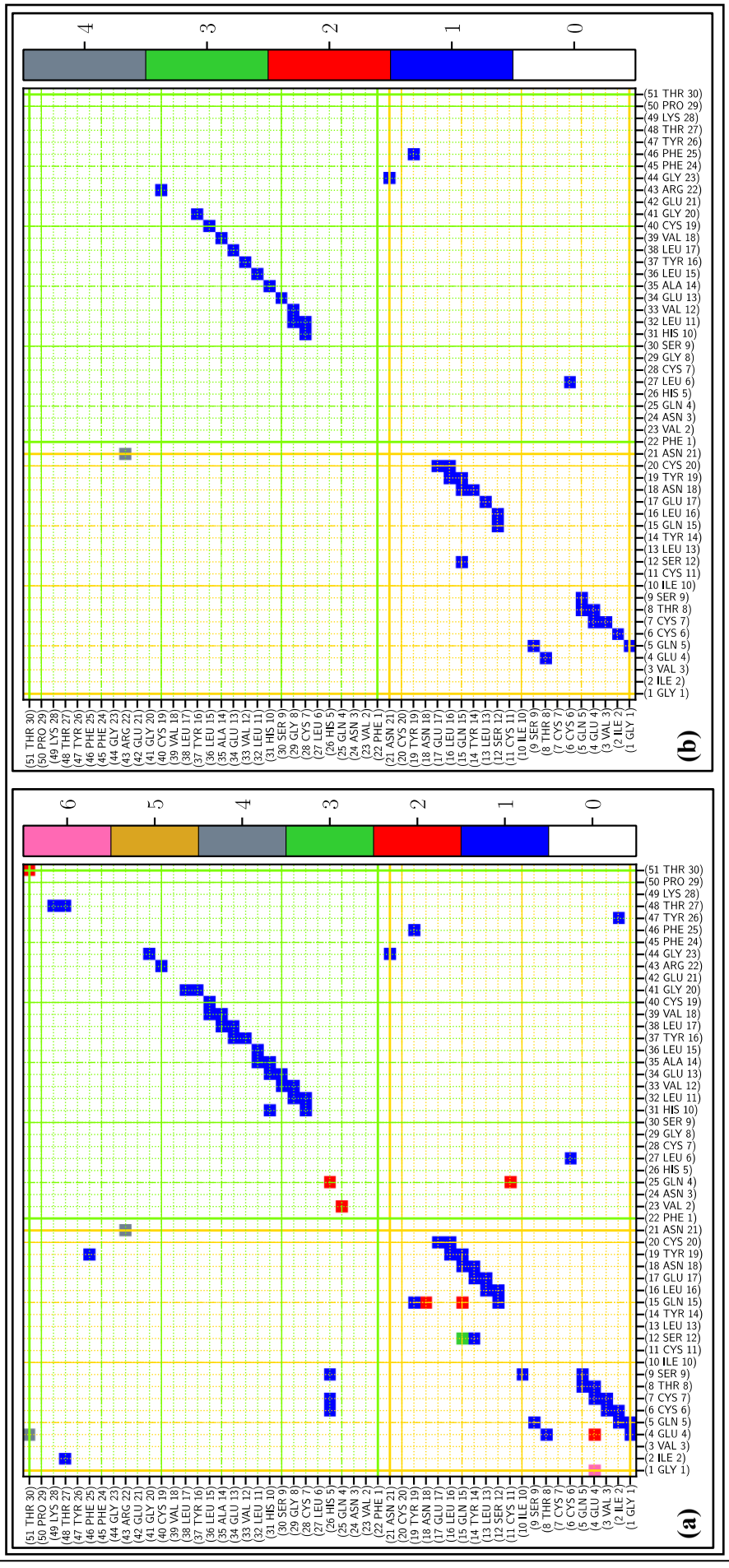


***Figure 6.15**: Sorted HBs in matrices between insulin residues of $\mathbf{P}_{\mathbf{kpl}}^{\mathbf{MDm}}$, calculated with "$|\mathbf{r}_{AD}| < 3.5$ Å & $\varphi < 60°$". **(a)** The 90 HBs present at least 5% of time (9-1499 ns). **(b)** The 39 HBs present at least 25% of time. The atom-specific HBs are sorted as "SC to MC", "SC to SC" HBs in upper left; "MC to MC" HBs in lower right; diagonal any HBs within a residue. Chains residues re-numbered sequentially 1-51 (nearest graph), and with actual residue name and number. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. grid-lines for AC (1-21) in gold-yellow, and BC (22-51) green. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51.*

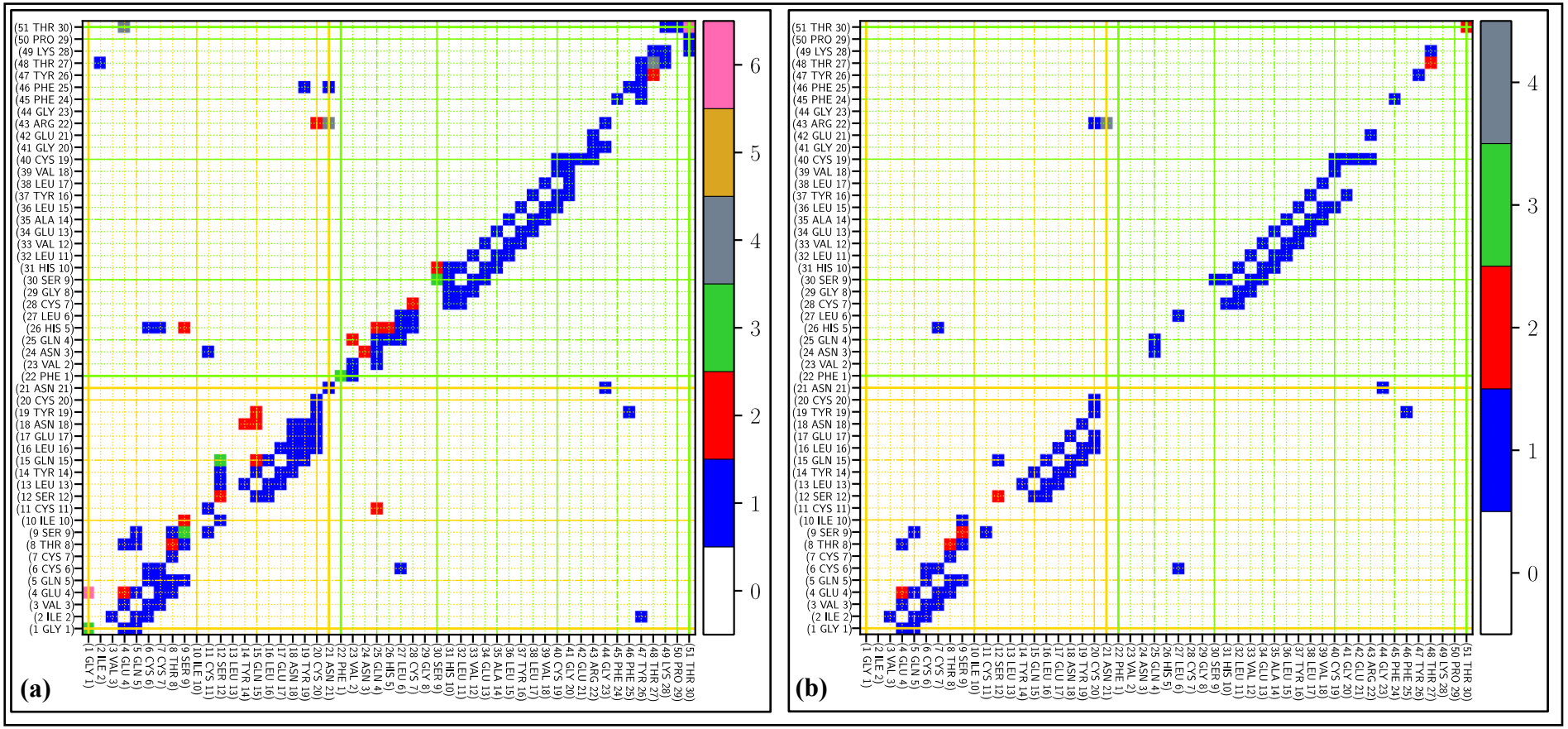


Figure 6.16: *Sorted HBs in matrices between insulin residues of* $\boldsymbol{P_{kpi}^{MDm}}$ *, calculated with "*$|\boldsymbol{r}_{AD}| < 3.5$ Å & $\varphi < 90°$*". **(a)** The 203 HBs present at least* 5*% of time (9-1499 ns). **(b)** The 104 HBs present at least* 25*% of time. The atom-specific HBs are sorted as "SC to MC" and "SC to SC" HBs in upper left; "MC to MC" HBs in lower right; diagonal any HBs within a residue. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics, grid-lines for AC (1-21) in gold-yellow, and BC (22-51) green. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51.

6.3 Discussion & conclusion

This chapter, building a solvated insulin model on top of that as obtained by Q. Hua *et al.* [3], whose experimental restraints and geometry were compared. The MD replicas indeed showed a much higher fluctuation and flexibility than $\mathbf{E_{kpi}^{DGR}}$, hence there was much more dynamical information obtained by MD. There is some uncertainty if these MD replicas, actually provide the most representative picture of the dynamics of insulin in solution (at conditions considered). The restrained DGR structure ensemble understandably matched its own restraints better; however the MD replicas being highly flexible in nature, showed a fair degree of correspondence to these restraints. Some uncertainties arise, if the RHBs (due to being derived for another solvent condition) is actually a good representation of HBs for a physiological solution condition. Moreover, the large amount of HB configurations sampled by MD (e.g. $\mathbf{P_{kpi}^{MDm}}$) and as calculated from $\mathbf{E_{kpi}^{DGR}}$ and **M12**, indicates a plasticity of this protein under various conditions.

The RHHs indeed showed a fair degree of congruence, albeit especially the restraints including H_{HD2}^{B10} had many violations. However speculatively, there may be some ionization state variation or dynamics that are not captured by this MD method (if not a mistake in the RHH assignment). Considering especially that the SC of histidine may have some sort of proton exchanging dynamics near physiological pH. However, 6 MD replicas with H^{B5} and H^{B10} protonated did not change the H_{HD2}^{B10} violation anomality, significantly. Though H^{B10} is involved in Zn^{2+}-coordination in storage-hexamer and apparently in the IR high affinity cross-link in contact with charged E_{SC}^{B13} and R_{SC}^{539}, there may be some effect due to the charged solvent (seemingly deuterioacetic acid) used in obtaining the RHHs.

Moreover that some RDAs, e.g. those of χ-angles, are freely rotational in space, hence does not seem reasonable to have bounds on, albeit may represent probable configurations.

The stochastic nature of the 9 different MD replicas, even if merely a few parameters between, revealed varying structural congruence with restraints. Starting from another set of three replicas, of markedly different initial starting structures, indeed had a larger effect, than merely changing temperature by 10 Kelvin. Hence this MD technique, apparently will sample another part of conformational space, if started in a markedly different energy-landscape. Indeed, a very different conformation was sampled by $\mathbf{P_i^{MDm}}$, which would have been missed if this simulation was run for less than 60 ns. Deviating MD replicas, showing abnormal conformations were also seen when I simulated other pH conditions, often 1 of 3

replicas (not shown in this thesis). This latter observation may have meaning for other less known proteins than insulin, to sample and study multiple replicas, in order to ascertain that a particular structural profile is reasonable. In addition, a MS from a simulation isn't necessarily "the" representative physical structure: for example, the calculated HBs of the MS does only depict these in a transient structure and not their probability during a full trajectory. In addition, the MD replicas that have a large sampled conformational space, do display structural variances at any time, though for most times constrained to a smaller space. Notwithstanding, the MSs of at least six out of nine replicas, have indeed on average a resembling structure to that of $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$. Hence it is to an extent verified, that the most probable average motif of insulin in solution is of T-state character. The most representative replica of the nine considered here, appears to be $\mathbf{P}_{\mathbf{kpi}}^{\mathbf{MDm}}$, albeit at least five others of near similar profile; this is in regards to having most experimental and structural congruence with the restraints, also with other experimentally derived HBs.

Moreover, ascertained by the HBs of MD replicas (also for **M12)** and RHB is the stability of the HBs in the BC αH. With known importance of the solvent exposed residues of the BC α-helix (c.f. §2.5.3), i.e. in regard to being emphasized as a central recognition element in binding to the IR; in addition involving residues that binds to both site 1* & 2 [185, 195]. Hence it is conceivable that the stability of this central BC α-helix, are vital for the insulin monomers ability to dock with the receptor. Moreover, the MD replicas points to a much higher possible flexibility of insulin dynamics, and also an ability of structural movements and transitions. The latter observation may have more meaning than so far known, in relation to the process of reaching the high affinity bound state.

The importance of verifying MD simulation results, before publishing them as accurate findings, has at times been carelessly overlooked in the field of computational chemistry. The number of replicas included in the analysis may seem redundant; however, the intention of the chapter is also to show differences in results due to small variations in parameters. Here it was observed that no MD ensemble are exactly the same, an important point was noted: that the intricate results can differ substantially if markedly different starting structures. Nevertheless, the combined analysis still gave reasonable statistical profiles of the most probable movements and behaviour. A full structural overview was only provided for one of the replicas, hence this were chosen as a model of solvated insulin under physiological conditions.

Chapter 7 Summary & Epilogue

Henry: "I think that one day any matter can be simulated realistically … "

Brian: "Absolutely!"

Similar wording from phone conversation, in year 2013, by novice PhD student prospect in Sweden and future supervisor in Melbourne.

To visualize and depict the complex biochemical molecules of life by the use of computational physics, is indeed an area that shows great promise in the decades to come; that so far only a glimpse has been realised of the capacity available to discover [284]. As of today, the simulation and experiment capability, allows to obtain models of rather stochastic nature. Future potential of visualization may provide a more determinant picture. The molecular dynamics model of the solvated insulin monomer is of stochastic character; however replicated many times for the particular conditions; here also dependent upon another solution model, inferred from NMR data of its dynamics and structure. The main aim of this thesis has been to provide a structural overview, of various models of the insulin monomer in different environments, to serve as a chart for understanding of its structure and dynamics.

7.1 Summary of thesis

This thesis has focused on models of the insulin monomer, the *key* in IR activation and a *building-block* in oligomeric storage forms. It was developed a conjecture of the insulin activation pathway in a novel framework. Moreover, models of insulin simulated by MD in explicit solvent were developed, analysed and compared to other experimental observables and their derived models. A similar analytical representation of the various models was developed, in order to chart structure and readily able make comparisons. Hence the analysis method of chapter 3 and the results chapters 4 to 6 (with supplementary) are intimately mathematically congruent, wherefore only example residue-profiles are provided. That is for allowing the reader to infer structure and dynamics themselves, one reason why the extensive figures of vector graphics are included in the supplementary.

Chapter summaries

- **Chapter 1:** Thesis introduction and prologue were the beginning chapter.

- **Chapter 2**: Here was provided a brief review of the insulin biology and where the solution monomer fits. Further reviewing plausible residues involved in its binding mechanism. Here introducing the structure of a T-state hexamer, since it is close in structure to a solvated monomer. In addition, reviewing IR activation, even though it remains somewhat enigmatic a lot of puzzle-pieces has been collected, the information was taken from a subset of literature, whereof some information which are rather cryptic. Notwithstanding, a conjectured model for the initial apo-IR binding zero, one, two (or 4) insulins were pieced together from various knowledge of its structural and functional biology. Nevertheless, this picture will likely be further verified how accurate it is, i.e. in the coming years as the full insulin IR binding process is unravelled and visualized.

- **Chapter 3**: Here was written a methodology for performing an MD simulation and analysis of insulin and proteins in general. Some biophysics and an MD simulation procedure were described for simulating a solvated protein. Methodology for relevant analysis of MD simulations (or any structure ensemble) were developed and explained. The methods applied are not exclusive to insulin, and can be applied to any protein of interest, as such it can serve as a guideline for other MD practitioners.

- **Chapter 4:** Here it was calculated some geometrical properties of published atomic structures, with highly varying atomic resolution. The analytical overview representation of e.g. HBs and distance matrices revealed structure in an innovative way for the **M12** asymmetric units from a highly resolved T-state hexamer crystal. The structure of **M12** is an interesting comparison to other insulin structures, in this chapter to the lower resolution approximative structure of the high affinity bound insulin ($\mathbf{CF}_{(\mathbf{A},\mathbf{B})}^{\mathbf{6HN5}}$), which apparently rationalizes at least some structural biology expectations. The highly varying resolution of the so far published atomic-structural IR-fragments, also of $\mathbf{CF}_{(\mathbf{A},\mathbf{B})}^{\mathbf{6HN5}}$, partly suggests that this picture can be improved upon. That one of the next important milestones is to confirm the observation by providing an alike highly detailed analytical overview of insulin reaching the high affinity cross-link to the apo-receptor, by means of e.g. simulation and analysis.

- **Chapter 5:** Here it was calculated and depicted an analytical overview of a DGR solution model that were obtained by Q. Hua *et al.,* comparing this insulin structure ensemble to its own restraints. Giving the conclusion that it contains rational geometry, but scarcely reflects any realistic biophysical fluctuations, not surprisingly appearing as a restrained average. Hence the analogous depiction of this insulin structure ensemble to that obtained by MD in chapter 6, enabled a straightforward overview and provided a basis for comparison.

- **Chapter 6:** Here it was obtained multiple MD trajectories starting from varying coordinates and temperature. Culminating in an analytical overview of its structure that are readily comparable to the other insulin structural models (as presented in the other result chapters). Obviously, there are a wide range of conformational plasticity of the systems compared. However also a great deal of common structural elements, e.g. the HBs of BC α-helices, and various HBs stabilizing the monomer. Hence this study has revealed plausible HBs, relative degree of fluctuation etc, important for insulin's structure, moreover its function. Hydrogen bonds are inarguably one of the most fundamental geometrical property of biochemical structures. Nevertheless, as congruent the dynamic model of insulin may seem, with the other presented models with experimental data; this MD model may need to be verified or improved upon by further research or validation. The intricate details provided for HBs and distances etc may serve as a comparison chart for direct measurements of observables, hence can serve as a reference. Notwithstanding, it must be emphasized that it is merely another solution model presented in this work. This model however has in effect been validated to an appreciable extent in this thesis, by scrutinising its degree of validation from experimental statistics and empirical observations. Some general observations of this MD model hence have a lot of truth to it, but one have to remember, it is a computer simulation that are of stochastic nature, whose real QM biophysical nature in solution may even hold hidden information of its dynamics. Nevertheless, this is a thorough classical mechanics modelling study of the solvated insulin monomer. A model of a *key* with its intricate dynamical structure, that upon binding fits the *lock* of the IR's binding regions, opening a cascade of vital biochemical pathways.

- **Chapter 7:** Thesis summary and epilogue.

Significance of work

The significance and innovation of this work is the filling in of missing puzzle-pieces and improving the depiction and visualization of various aspects of insulin structural biology; a puzzle that is yet not completely solved. The work is another step towards allowing a better visual and comprehensive understanding of insulins (and proteins in general) structural and dynamical biology. Since, rational drug design benefits, from having as clear of a picture and understanding as possible, of the physiological and molecular biology of insulin, which still in parts are unascertained. This report aimed to facilitate and improve current and future understanding; which it has done by pointing to apparently unseen aspects of molecular biology and gathered structural data from various different models of insulin; culminating in an experimentally rational dynamical model of the solvated insulin monomer.

7.2 Improvements in future analysis and visualization

Matter in a biological system is a fuzz of moving mass and energy. Though many biochemical conformations can be approximated by MD or a.k.a. classical molecular mechanics (MM); there are also quantum mechanical (QM) effects involved, e.g. diffusion of protons through solvent and breakage and forming of covalent bonds [285-289]. The computational biochemistry field, hence anticipates the rise of improved hybrid QM and MD methods [290]. Even in the last two decades an increasing use of QM/MM have been for the studying energetic properties of proteins and enzymes reactions; in particular relevant here are for kinase catalysed phosphorylation reactions [291-297].

In the past 25 years computer capacity for computational chemistry has increased immensely. A maybe not very realistic prediction based on Moore's law [47, 290] is that the computer power might be scaled by a million fold by 2040; however dependent on if world development are either destructive or benevolent and if there are innovations in software and hardware; hence this estimate may be off (lower or higher) by many orders of magnitude. Concomitantly improvements in the field of computational chemistry, along with algorithms used in QM/MM methods; possibly in conjunction with microscopy techniques; here postulated to allow simulations of intricate details of biochemistry of much larger size- and time-scales. There will also likely be an increase of "black box" calculations, making it easier for practitioners of computational chemistry to output reliable results; as of today it may be a varyingly cumbersome process to get an analytical overview of

simulated biological systems. Hence it is not difficult to imagine that eventually a realistic simulation and analysis of the insulin (and akin molecules) to cognate receptor binding processes will be possible; visualizing completely an intricate mechanism on how insulin activates its cognate receptor.

Improving simulation of insulin analogues and hexamer storage forms

The attempt in this thesis has been to, with relatively simple means, capture structural transitions, that may be important for the solvated insulin's biological function.

The solvated insulin model of this thesis, were obtained by classical MD simulation of multiple 1499 ns replicas, having various starting coordinates and temperatures. Since the main model coordinates were well chosen from a refined DGR model; which evidently sampled a substantial part of conformational space; appearing in large parts reasonable experimentally. However, the choice of a more deviant initial structure, apparently gave quite different dynamical behaviour and sampling. There are other advanced methods such as accelerated MD and Replica Exchange MD [298], that could possibly sample conformational space in a more exhaustive manner; overcoming some high energy barriers between different conformational states, that may otherwise not be sampled. Indeed choice of: pH, starting coordinates, temperature, parameters and force-field, may more or less influence the observables from a MD simulation [299]; to a varying extent ascertained by other insulin MD simulations performed (not included).

Furthermore, having constant ionization states in a MD simulation may not fully model nature, since it should dynamically depend on momentous chemical environment. There are other methods such as constant pH MD, that possibly can sample the effect of protonation state variability on conformational space, in a more realistic way [205, 300]. Even choice of solvent model used for MD simulations may play a role [301-305]; there are however progress in adaptive QM/MM solvent treatment [306-309], e.g. the ions Mg^{2+}, Zn^{2+} has been better modelled with a QM/MM approach [310]. Nevertheless, we have taken care in choosing parameters and choice of force-field for the MD simulations. Notwithstanding, considering that a solvated insulin system is relatively small (~6000 atoms), an interesting comparison would be to treat it by QM/MM; to see if some biophysics (e.g. histidine ionization state variation) can be better understood [311], and if results are comparable to this thesis.

An interesting validation of a force-field adapted for crystallography, would be to perform an at least μs long simulation of the periodic hexamer structures, with the pertaining experimental conditions; confirming the atomic fluctuations and average structure of the asymmetric periodic unit **M12** by Baker *et al.* [1]. Furthermore, with present-day QM/MM it might be possible to perform accurate longer time-scale simulation, of hexamer packing within pharmaceutical storage forms; even to resemble storage within entire vesicles inside the pancreas. Which might provide further insights, e.g. how insulin can be better stored for use in diabetes treatment.

Simulating insulin and other ligand receptor binding

Albeit present binding sites appears largely well-defined: there may be useful procedures to dock insulin to the already available IR-fragments, in order to obtain more precise definitions of insulin binding [312, 313]. However, the simulating of insulin binding (and unbinding upon dissociation) to its cognate receptor, is likely to eventually be fully simulated by QM/MM. Which would then be a milestone in understanding the biophysical dynamics of this binding process and its activated signalling pathways. Further it would also serve as a model or extension for accurate simulations of other ligand receptor systems; particularly for the related types of IGF hormones and cognate receptors. Would this be achieved reliably then possibly single amino-acid mutations of either insulin or IR may be studied, in order to better understand various types of disease [124]. Analytical overviews of the IR to be compared with other IGF ligands and receptors, would possibly provide important referencing. Moreover, it could provide a test-system for various other related alternate ways of activating the IR. For example one could visualize how a 24-residue peptide having the TM sequence can activate the TK domains by putatively disturbing the TM domains [314].

In principle it may be possible even now, to start with the unbound monomer with the apo-IR ectodomain structure (known to a substantial degree [129]); with the CT F3 and F3* possibly tethered to a point in space or to a lipid membrane (with appropriate assumptions about dynamic spring-forces); which can then be solvated, equilibrated and simulated of reaching the high affinity cross-link via MD (if not so far unidentified vital QM effects are necessary in the ectodomain).

Some pioneering simulation attempts of parts in the ligand receptor puzzle has been performed by Vashisth *et al.* [315-320].

Vision of simulating and analysing the hypothetical mechanism of 1 to 2 (or more) insulins binding to IR with consequental tyrosine kinase activation

An attempt of an atomic model architecture depiction by Kidmose *et al.* [321] of the full human insulin receptor, were combined by an ectodomain structure [129], TM [166] (inserted into model lipid bilayer membrane) and with separated TK domains [322].
Here however the model are as presented in §2.5 (in particular §2.5.2.2, Figure 2.7) which are represented as an atomic 3D model in Figure 7.1, which is a summarizing depiction of an imaginative ligands receptor binding simulation.
The apo-IR structure of Figure 7.1a, have its ectodomain modelled from "model S1" [129], whereas TM domains are from PDB 2MFR [166], JM and CE segments loosely modelled and the unactivated TK domains are from PDB 1IRK [180], with the unbound insulin monomers being the dynamical model of this thesis (i.e. the structure in Figure 6.13).
The two different alternatives in Figure 7.1b1&2, of a 1 insulin bound IR ectodomain; half activated ectodomains are more or less devised (altered) from the combined "model S1" , PDB 6HN4 & 6HN5; with half-activated TK domains modelled by one subunit of PDB 4XLV and one from PDB 1IRK. The insulin bound "signalling conformation" ectodomain of Figure 7.1b3 are merely PDB 6HN4 & 6HN5, with the fully activated TK domain modelled by PDB 4XLV. A structure of the here fully activated conjectured 2 insulin bound T-shape ectodomain receptor Figure 7.1c, was here modelled by insulins, αCTs and F1-L1-C-L2 domains from PDB 6CE9 and F3-F2 domains from PDB 6HN4, where the fully activated TK domains are modelled by PDB 4XLV.
The above structural conformations inspires the idea that at least eventually, one can construct an entire structural "box" containing a resolved full-atom representation of the human insulin apo-receptor, solvated with water and other relevant molecules and ions, in an energy-minimized and equilibrated state, to be given in a PDB entry. The need for constructing a full-scale structurally reliable model, moreover with the developing of a suitable MD or QM/MD method is postulated here to be paramount, if these advances are made reliably, one may then in principle add a physiological probability of insulin monomers into this box simulating the mechanism of binding. Then initially before binding, the solvated insulin in this box, could verify the common aspects of that of the structural dynamics as presented in this thesis. Following analysing and viewing the full atomic-mechanism of insulins binding conforming and twisting with the receptor monomers, reaching the high affinity bound conformation and beyond.

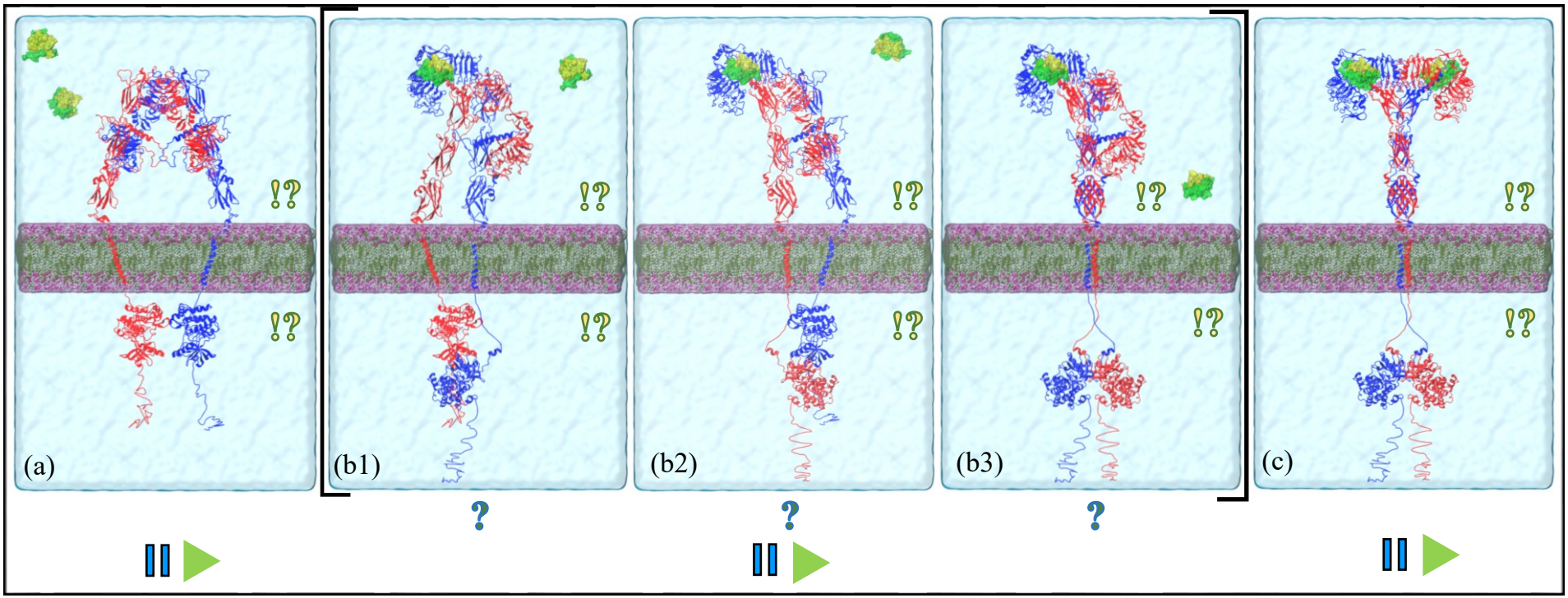


***Figure 7.1**: The figure depicts an imaginary view of the insulin receptor activation mechanism as would be played in a realistic MD/QM simulated movie. One may even imagine that this is an excerpted box of a much larger system entailing more insulins and other biomolecules of the insulin signalling pathway. Hypothetically a simulation would involve all atoms mostly modelled by MD and QM for the phosphorylation of tyrosine's in the activation loop and in other sites that promotes the recruitment of the consequent IR substrates being links in the signalling pathway. The exclamation and question marks designate that this presented model is based on more or less certain elements of what the apo-, holo-IR ecto and cytoplasmic domains looks like. Here only explicitly showing the secondary structure as a ribbon model, insulins in all-atom surface, implicitly shown are all-atoms, moreover possible extra unbound/bound insulins, moreover molecules such as e.g. ATP/ADP, ions and possibly other supporting or scaffolding structures. Directly comparable to schematic of Figure 2.7 in §2.5.2.2. Picture made in VMD, see text for further explanation.*

Acknowledgements and note on contribution

This work was supported by a La Trobe University Postgraduate Research Scholarship, in addition this work was supported by a La Trobe University Full Fee Research Scholarship. This work was performed whilst a member in the group specialising in molecular modelling lead by Brian J. Smith, who with members M. Encisco, M. Thomas, N. Meftahi N. Smith, R. Jin, S. He, M. Walker, A. Lucke, and others at department, have been assisting in various matters. Foremost Brian has provided supervisor support and advice for the whole duration of this challenging project. The earlier stage data-collection and code-development performed at La Trobe University, along with facilities and support from various staff, whereas the later stage computing-analysis and writing was off-campus and in Sweden, funded by my parents.

Various staff at my department and especially at the international and graduate research departments of La Trobe University enabled me to complete this difficult project overseas. Useful insights and structures are come from Brian's group collaborating with e.g. groups of M. Lawrence, M. Menting, C. Ward at the Walter and Eliza Hall Institute, in addition to M. Weiss, N. Wickramasinghe and others from Case Western Reserve University, Cleveland, USA, in addition to many others. The MD simulations were calculated on Trifid supercomputer of VPAC (Victorian Partnership for advanced computing) and the Avoca BlueGene supercomputers of the Victorian Life Sciences Initiative (VLSCI), now a days called Melbourne BioInformatics, whose societies has provided a lot of assistance. The software used for MD simulation were GROMACS [48, 49], I've been in personal contact with the developers, mainly D.v.d Spoel, who I visited during a short term, who along with other developers at forums, J. Lemkul, M. Abraham, E. Lindahl and many others, have provided help on how to use the software.

People providing generous discussion or comments, who with benevolent remarks improved the thesis: thanks to Brian J. Smith, Janni Boding Christensen, Tamanna Saiyed, Pierre De Meyts; and thanks to examination reviewers Neha Gandhi and Harish Vashisth.

Thanks also to the people contributing to the good work described in the literature of the bibliography, on which this thesis is built upon.

Chapter 2: The review was written by myself, where inspiration and references are as noted in the text of the chapter. Some validating dialogue from Pierre De Meyts on the matters of the chapter.

Chapter 3: Has been written by myself, with initial and ongoing inspiration about how to perform MD simulation and analysis by supervisor, members of group and others at the MD society (in particular at GMX and VLSCI). Also the methodology of the MD simulation, was largely inspired from Justin Lemkuls tutorial [232], which I refined as are indicated in the cited literature. Inspiration for performing automatic multi replica simulations on VPAC and VLSCI computers, were given by Michael Kuiper. Furthermore, with the much used software Visual Molecular Dynamics (VMD) software [50, 239], I wrote many elaborate analysis scripts, based on its functions. With the VMD code being based on the Tcl/Tk [323, 324] language. Moreover, in the coding language Python scientific programming [325-328], I wrote the scripts used for statistical calculations. In addition the plotting scripts I have written in matplotlib [280].

Chapter 4, 5, 6: The analysis of results and writing is performed by me, with main feedback and discussion from Brian J. Smith. The hexamer and IR-fragments structures with original reports are as cited, albeit the structural analysis overview provided by me. There were helpful discussion provided on the DGR model and its restraints [3], by members of the authoring group, mainly Nalinda Wickramasinghe and Michael Weiss.

Appendix A Atoms in Amino-acids Naming and Bonding

A.1 Amino-acid structure and naming nomenclature

Here it is defined the structure and atom-naming of some amino-acids (see Figure A1), whereby the residue and atom-names are used as abbreviations when referring to them in the main text, figures or tables. The atom-naming follows the convention that were implemented in GMX v.5.04. CHARMM36 (mars 2014 version), that were used for the MD simulations.

Consecutive chains e.g. for the insulin monomer of two chains, the A-chain (AC) and B-chain (BC), are sometimes named with the residue-numbers restarting at the beginning of a subsequent chain, i.e. the chain numbering may respectively be A1-A21, B1-B30 (same if for contiguous fragmented chains). In addition for consecutive chains, e.g. for insulin the chain-numbering continues from the CT AC to the NT of BC, it is designated as B22-B51, if proper in the text it may be written as B1(22)-B30(51) to clarify in explaining analysis (same if for contiguous fragmented chains).

The MC atoms are defined (if not noted otherwise) as anything not being a SC atom, i.e. the following atoms; HN, N, C, O, CA. The physiological protonation state NT and CT atoms (H1, H2, H3, OT1, OT2) are designated as MC atoms when calculating atom or residue distances. The SC atoms are defined as any other atoms of residue connected to a CA atom, not being part of the MC atoms.

When referring to an amino acid within insulin, it is distinguished with its position in a polypeptide, and if adequate referring to an atom or an entire selection of atoms or a property. Where the following nomenclature are used, e.g. for insulin: Leucine at position 6 (or 27) in BC, as L^{B6}, also if not clear from context, when referring to a DA (e.g. χ_1), or an explicit atom e.g. HN, or an entire atom-selection e.g. SC or MC, then respectively $L^{B6}_{\chi_1}$, L^{B6}_{HN} or L^{B6}_{SC} may be used. Albeit interchangeably the nomenclature "B6 L" and in particular "B6 L(HN)" may be used especially when referring to an atom, e.g. "HN", in HBs or in hydrogen pair distances. In addition, as an example, nomenclature "$L^{B6}_{HN} \rightarrow C^{A6}_{O}$", may be used when referring to a specific HB.

The colour scheme of atoms in thesis (if not indicated otherwise) are as in general chemistry convention, i.e. the following: hydrogen as white, carbon as black, oxygen as red, nitrogen as blue, sulphur as yellow. In figures depicting structures the MC is atom-coloured, however the SC atoms may have an alike respective colour as implied in Figure A1.

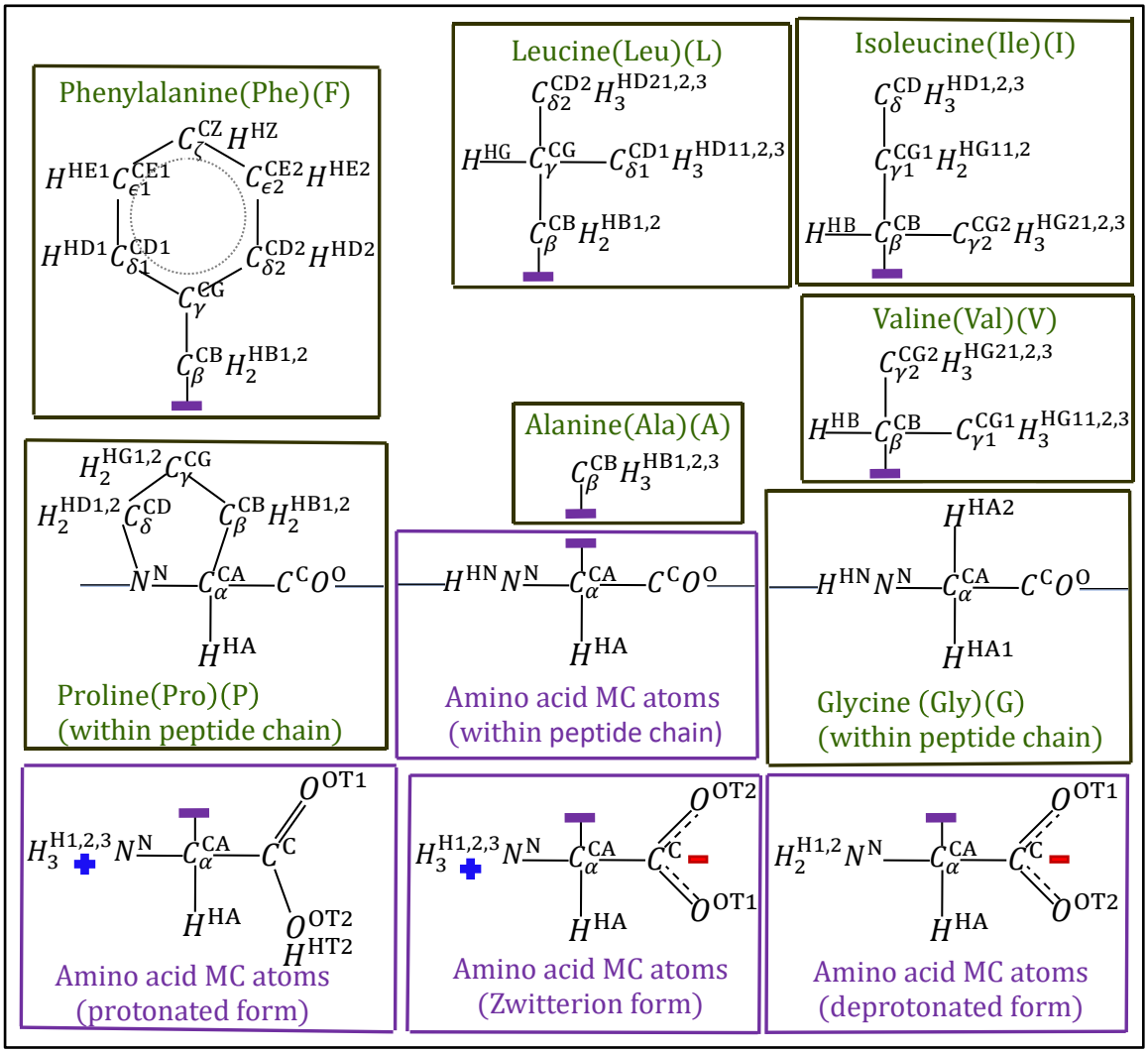


Figure A1*: Structure and atom-naming nomenclature for amino-acids . Here shown is the backbone atoms and the less-polar or hydrophobic (a.k.a. non-polar) residues. Tetrahedral sp3 hybridized carbons shown in Fischer projection (see Figure A2). The atom notation is shown with an example, for the MC central alpha atom which is named, C_{α}^{CA}, where main letter C means carbon-atom, subscript α means alpha position, and the superscript CA is the atom-name used in simulations. Thus, we may refer to notation C_{α}, in a general sense in the text, and possibly its atom-name CA when referring to results derived from simulations, or interchangeably.*

The logic for the greek alphabet used in naming are in order (lowercase:uppercase): $(\alpha{:}A), (\beta{:}B), (\gamma{:}\Gamma), (\delta{:}\Delta), (\epsilon{:}E), (\zeta{:}Z), (\eta{:}H)$, where in atom-naming G is used for γ, and D is used for δ, instead of the uppercase greek symbols.

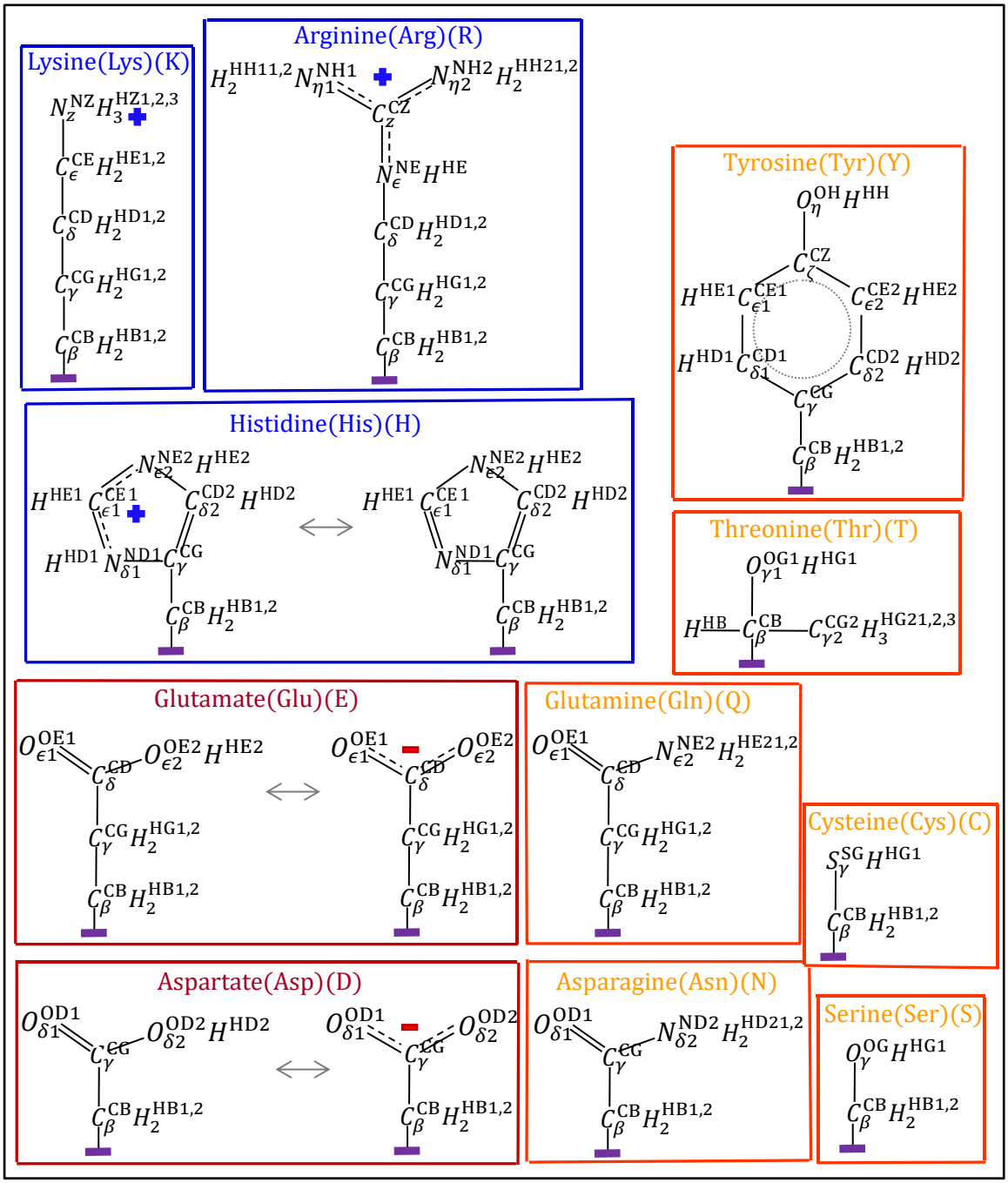


Figure A1 continued: *Atom-naming format for amino-acids. Here shown are acidic, basic and polar amino acids. Indicating different protonation states with arrows between the two forms. For the deprotonated case of Histidine either* $N_{\delta 1}^{ND1}$, $N_{\epsilon 2}^{NE2}$ *can be protonated, depending on the adjacent chemical environment.*

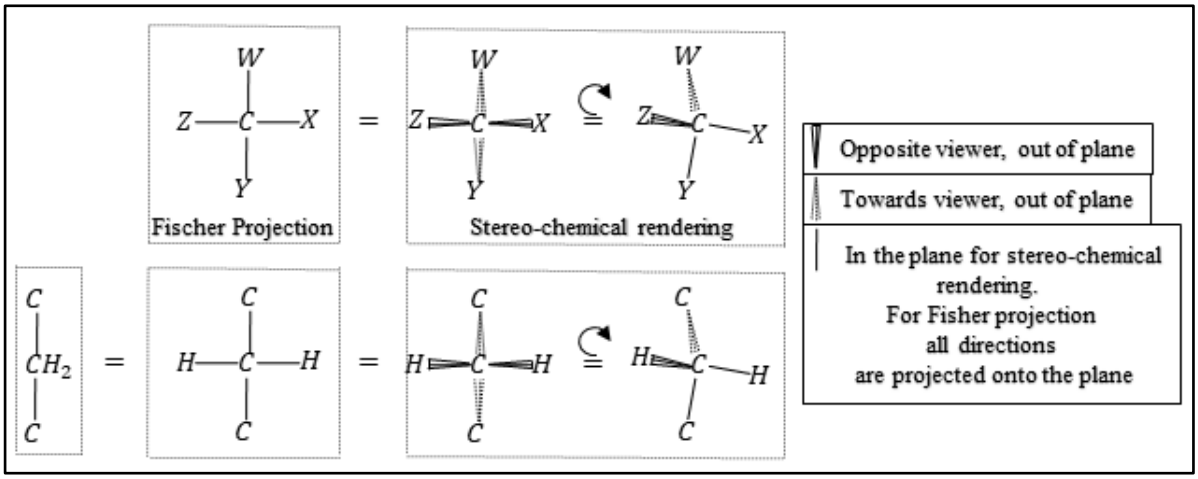


***Figure A2**:A Fischer projection, equalling a stereo-chemical rendering which after rotation has different orientation in space of each bond. That is, for a tetrahedral geometry as for a sp^3 hybridized carbon with four single bonds to other atoms. The angle between two of any of the bonds are* $109.5°$. *Above generic example. Below with example atoms, note the molecular formula equivalence of the* $-CH_2-$. *That is to say that hydrogens can be implicitly included, still maintaining the information of stereo chemistry. In a Fischer projection, the bonds to the central carbon are represented by horizontal and vertical lines, from the substituent atoms to the carbon atom at the centre of the cross. By convention, the horizontal bonds are assumed to project out of the page toward the viewer, whereas the vertical bonds are assumed to project behind the page away from the viewer [25].*

A.2 Dihedral Angle Definition

A protein being constituted of polypeptide chains and inherently flexible related to the allowed degree of torsional rotation around covalent bonds. The peptide bond connecting different residues in a protein is rigid, because of its in part double-bond character [25] (Figure A3), thus restraining the conformational space of a protein.

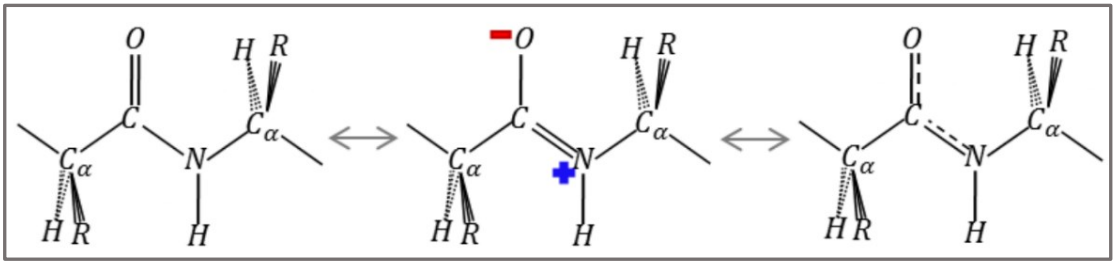


***Figure A3**: Resonance of the peptide-bonds in a protein . The peptide bond is kinetically stable. With a rate of hydrolysis being extremely slow, closing in to 1000 years, in an aqueous solution, with the absence of a catalyst [25].*

However, the single covalently bonded atoms of each residue, are freely rotatable, though restricted by steric-hindrance from other chemical groups. To calculate these rotations, it is

consensually defined a dihedral angle between 4 atoms, as defined in Figure A4. In addition, the nomenclature used for describing main-chain angles are described in Figure A5.

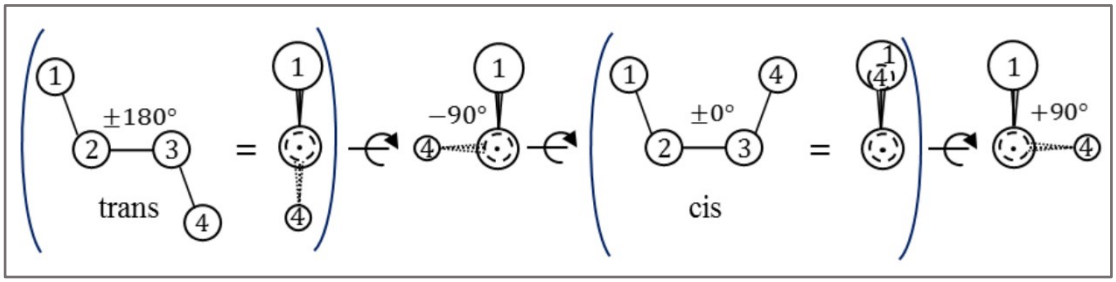


Figure A4*: Definition of a dihedral angle of bonded atoms . From left to right: clock-wise* $+90°$ *rotations of atom 4 relative to atom 1, where the central bond of atoms 2-3 is the rotating axis.*

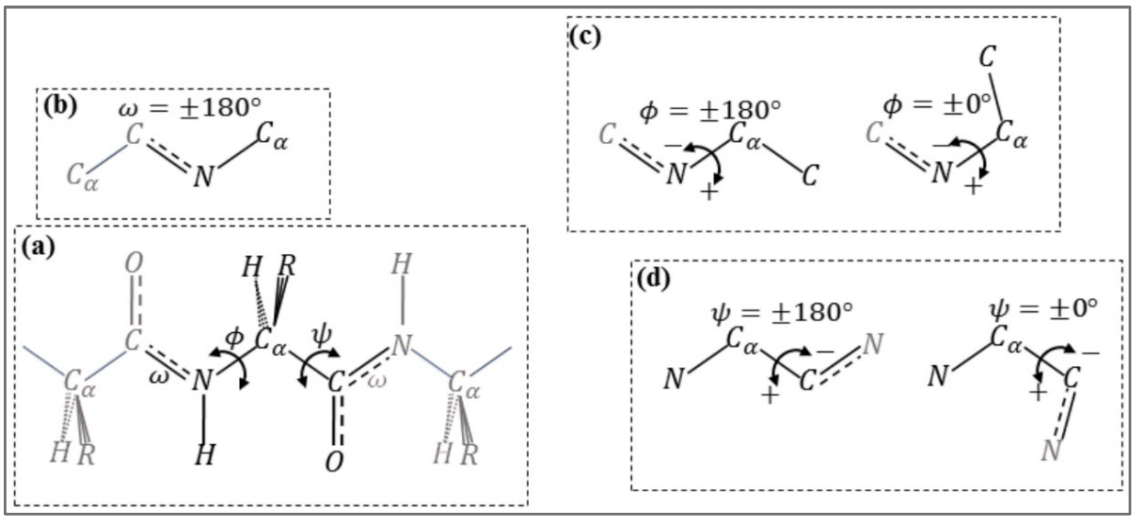


Figure A5*: Definition of the peptide main-chain dihedral angles .* ***(a)*** *Excerpt of a peptide chain, indicating the* ω, ϕ, ψ *dihedral angles. Atoms of the residue in black for which the angles belong, adjacent residues in grey.* ***(b)*** *The amide-bond dihedral angle,* ω*, here showing the most common trans (*$\pm 180°$*) configuration. The N-terminal residue of a peptide chain do not have this* ω *angle.* ***(c, d)*** ϕ, ψ *angles, which respectively can range between cis (*$\pm 0°$*) and trans (*$\pm 180°$*) conformation. Restrained by steric hindrance from chemical groups such as the side-chain (R) or main-chain. The* ϕ *angle is defined to be zero at the N-terminus of an amino-acid chain, while the* ψ *angle can vary. Likewise, the* ψ *angle is defined to be zero at the C-terminus of an amino-acid chain, while the* ϕ *angle can vary.*

In particular, it may be defined dihedral angles for the singly covalent bonds between atoms of side-chains, referred to as χ angles. Where the number of χ angles for each standard amino-acid is varying between 1 and 5 ($\chi_1 \dots \chi_5$). The definition for each χ angle are involving 4 specific atoms. The ω angle tells if the peptide bond is cis or trans, as for insulin monomer analogues in this thesis, all peptide-bonds are trans. In fact, almost all peptide bonds in proteins are trans, due to the steric hindrance that would otherwise occur between the adjacent residues side-chains, if the peptide-bond between were cis. The most common

cis ω angle peptide-bond, being for X-Pro (X arbitrary amino-acid), where for the cis and trans form similar steric hindrance arises [25]. The χ angles have free rotation, restricted by moieties steric hindrance and bonding to intra- or inter-residue chemical groups.

***Table A1**: Definition of main-chain and side-chain dihedral angles. The convention for atom-naming and colouring is the one of Figure A1. The dihedral angles belong to a specific residue whose atom-names is in black, if atom-names are grey they belong to adjacent residues as in Figure A5. For the main-chain dihedral angles the order of atoms is written from NT to CT direction.*

Angle	Rotation Axis	Atoms of Angle	$\pm$0° angle
ω	C-N	CA-C-N-CA	CA cis to CA
ϕ	N-CA	C-N-CA-C	C cis to C
ψ	CA-C	N-CA-C-N	N cis to N
VAL (χ_1)	CA-CB	N-CA-CB-CG1	N cis to CG1
ILE (χ_1)	CA-CB	N-CA-CB-CG1	N cis to CG1
ILE (χ_2)	CB-CG1	CA-CB-CG1-CD	CA cis to CD
LEU (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
LEU (χ_2)	CB-CG	CA-CB-CG-CD1	CA cis to CD1
PRO (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
PRO (χ_2)	CB-CG	CA-CB-CG-CD	CA cis to CD
PRO (χ_3)	CG-CD	CB-CG-CD-N	CB cis to N
PHE (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
PHE (χ_2)	CB-CG	CA-CB-CG-CD1	CA cis to CD1
MET (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
MET (χ_2)	CB-CG	CA-CB-CG-SD	CA cis to SD
MET (χ_3)	CG-SD	CB-CG-SD-CE	CB cis to CE
CYS (χ_1)	CA-CB	N-CA-CB-SG	N cis to SG
SER (χ_1)	CA-CB	N-CA-CB-OG	N cis to OG
THR (χ_1)	CA-CB	N-CA-CB-OG1	N cis to OG1
TYR (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
TYR (χ_2)	CB-CG	CA-CB-CG-CD1	CA cis to CD1
ASN (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
ASN (χ_2)	CB-CG	CA-CB-CG-OD1	CA cis to OD1
GLN (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
GLN (χ_2)	CB-CG	CA-CB-CG-CD	CA cis to CD
GLN (χ_3)	CG-CD	CB-CG-CD-OE1	CB cis to OE1
ASP (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
ASP (χ_2)	CB-CG	CA-CB-CG-OD1	CA cis to OD1
GLU (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
GLU (χ_2)	CB-CG	CA-CB-CG-CD	CA cis to CD
GLU (χ_3)	CG-CD	CB-CG-CD-OE1	CB cis to OE1
HIS (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
HIS (χ_2)	CB-CG	CA-CB-CG-ND1	CA cis to ND1
LYS (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
LYS (χ_2)	CB-CG	CA-CB-CG-CD	CA cis to CD
LYS (χ_3)	CG-CD	CB-CG-CD-CE	CB cis to CE
LYS (χ_4)	CD-CE	CG-CD-CE-NZ	CG cis to NZ
ARG (χ_1)	CA-CB	N-CA-CB-CG	N cis to CG
ARG (χ_2)	CB-CG	CA-CB-CG-CD	CA cis to CD
ARG (χ_3)	CG-CD	CB-CG-CD-NE	CB cis to NE
ARG (χ_4)	CD-NE	CG-CD-NE-CZ	CG cis to CZ
ARG (χ_5)	NE-CZ	CD-NE-CZ-NH1	CD cis to NH1

A.3 Hydrogen Bond Definition

Hydrogen bonds are ubiquitous in biochemistry, for example, they are responsible for specific base-pair formation in the DNA double helix. Moreover a chief factor in determining the secondary, tertiary and quaternary structure of proteins [329]. Furthermore hydrogen bonding interactions are responsible for many of the properties of water, that makes it such a special solvent [25]. Hydrogen bonds (HBs) are hence a fundamental electrostatic interaction, here defined as in Figure A6. The H atom in a HB, is partially shared by two electronegative atoms, such as nitrogen or oxygen. A HB donor group consists of two atoms, the H atom with a covalent bond to a more electronegative atom. The electronegative atom having a partial negative charge (δ^-), since the electron density is taken from the H atom, giving the H a partial positive charge (δ^+). Thus, the H will be electrostatically attracted, to another electronegative acceptor atom having negative charge. The HBs are much weaker than covalent bonds, having energies ranging from 4 to 20 kJ mol^{-1}. Whereas, for example, a typical nitrogen-hydrogen covalent bond ($N-H$) of length 0.9 Å, has bond energy of 391 kJ mol^{-1} [330]. Moreover, HBs are somewhat longer than covalent bonds. Where bond lengths; from the hydrogen to acceptor atom, are ranging from 1.5 Å to 2.6 Å; or conjointly bond lengths, from D to A atoms, ranging from about 2.4 Å to 3.5 Å (or less if angle φ not zero). The criteria's of distance, "$|\boldsymbol{r}_{AD}| < 3.5$ Å", and angle "$\varphi < 90°$", are considered to cover most of the classical HBs in proteins [253, 254, 331]. The strongest HBs, tend to be of lesser angles, such that the three atoms of a HB lie along a straight line [25], hence distinguishing between three angle ranges ($\varphi < 30,60,90°$).

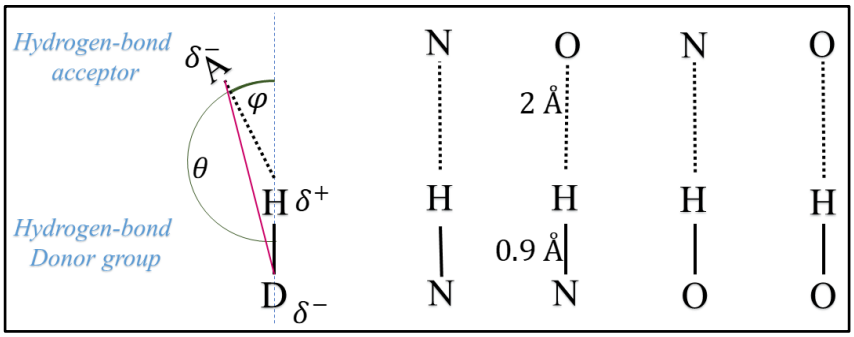


***Figure A6**: Hydrogen bond definition. The donor group with the electronegative atom D (of coordinate $\boldsymbol{r}_D$), covalently bonded to a hydrogen H (of coordinate $\boldsymbol{r}_H$), forming a HB with an electronegative acceptor atom A (of coordinate $\boldsymbol{r}_A$). The atoms partial charges (δ^+, δ^-) are indicated. Where the electronegative atoms nitrogen and oxygen are ubiquitous in e.g. proteins, with example distances of the HB O--H-N shown [25]. The following geometric criterion are for the formation of a HB. That is, the vector length distance from A to D, to be less than a cut-off distance (e.g. $|\boldsymbol{r}_{AD}| < 3.5$ Å). In addition, that the angle, $\varphi = 180° - \theta$, to be less than a cut-off angle (e.g. $\varphi < 30°$, 60° or 90°) where θ is the angle formed between $\boldsymbol{r}_{AH}$ and $\boldsymbol{r}_{DH}$.*

Appendix B Math definitions used in analysis

Some relevant math used in this thesis, for proper definition and for understanding the description of analysis methods.

B.1 Vector algebra

Defining some vector algebra [252, 332]. For a three-dimensional Cartesian coordinate vector $\boldsymbol{r}(x, y, z)$**,** the length of it is defined as:

$$|\boldsymbol{r}| = \sqrt{x^2 + y^2 + z^2} = r \tag{7.1}$$

Designating the vector coordinates of atoms in space, for instance those of a hydrogen bond, $\boldsymbol{r}_D$ of donor D, $\boldsymbol{r}_H$ of donor hydrogen H, and $\boldsymbol{r}_A$ for acceptor atom A. Assigning a coordinate difference between e.g. the following two coordinates:

$$\begin{aligned} \boldsymbol{r}_A - \boldsymbol{r}_H &= (x_A - x_H, y_A - y_H, z_A - z_H) = (x_{AH}, y_{AH}, z_{AH}) \\ &= \boldsymbol{r}_{AH} \end{aligned} \tag{7.2}$$

To get the angle, θ, between two vectors $\boldsymbol{r}_{AH}$ and $\boldsymbol{r}_{DH}$, there is a relation to their dot-product,

$$\boldsymbol{r}_{AH} \cdot \boldsymbol{r}_{DH} = x_{AH}x_{DH} + y_{AH}y_{DH} + z_{AH}z_{DH} = |\boldsymbol{r}_{AH}|\,|\boldsymbol{r}_{DH}|\cos(\theta)$$

, and solving for the angle

$$\theta = \arccos\left(\frac{\boldsymbol{r}_{AH} \cdot \boldsymbol{r}_{DH}}{|\boldsymbol{r}_{AH}|\,|\boldsymbol{r}_{DH}|}\right) \tag{7.3}$$

B.2 Mean, variance and standard deviation

Some statistics used in this thesis [60, 332, 333]. For a number, n, of data-values, $x_1, x_2, \cdots, x_n$, we define an arithmetic average of

$$\langle x \rangle = \left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \tag{7.4}$$

The mean error or standard deviation (SD), of an individual measurement or sampling x_i:

$$SD_{x_i} = \sqrt{\left(\frac{\sum_{i=1}^{n}(x_i - \langle x \rangle)^2}{n-1}\right)} \tag{7.5}$$

, where the variance is defined as $(SD_{x_i})^2$. The standard deviation (mean error) of the average $\langle x \rangle$,

$$SD_{\langle x \rangle} = \frac{SD_{x_i}}{\sqrt{n}} \tag{7.6}$$

, the latter equation not included in any graphs, though can be directly inferred with a calculator. Note that all statistics were calculated for any observable in a MD ensemble, even if showing full trajectory (0-1499 ns); only times after the first 8 ns (9-1499 ns) were considered in statistics. The script calculating these statistical quantities for any time-dependent variable, I wrote in Python [325, 327, 328].

Supplementary Chap. 3

Researchers publishing new biomolecular structures, using any simulation methodology, are encouraged, to describe their computational procedures more fully [334]. One reason is so that others can be able to reproduce their results. Similar reason is so that others may receive the benefit of understanding the results in more depth, not having to figure out or guess vital details on their own. Nonetheless, the reliability of a MD simulation method, depends on the choice of parameters and force-field. For these reasons, it is also recommended by developers of the GROMACS (a.k.a. GMX) software, to include sufficient information about the MD simulation in publications, using their software [48, 334]. Accordingly, here is provided this information, so that the work presented can be scrutinized and further improved upon. Moreover, to provide reproducibility of our results, and guidance for anyone continuing similar work.

Here GMX commands and parameters are shown in S3.1. In addition, it is provided in S3.2 the GMX commands for post-processing a MD trajectory, along with the protocol for mean protein structure calculation. The elaborate scripts I've written for analysis and plotting for this thesis is not included, since that would take up many, many pages and would require very intrinsic explanation. However, the underlying basics of the analysis are already described in §3.3.

The trajectory or structure data may appear also at 1 or 2 databanks (links https://welcome.gpcrmd.org/, https://www.rcsb.org/); where the eventual links, accompanying publications, explanations (and if corrections), data and code may be provided in a github webpage/repository, and if anything missing can be added on demand: https://wittler-github.github.io/A_MD_Analysis_of_Insulin/.

S3.1 Simulation commands and parameters

Commands for performing a MD simulation of a protein in GMX v. 5.0.4. CHARMM36 (v. mars14) on a UNIX and PBS system:

(1) Make Gromacs topology

```
gmx pdb2gmx -f ins.pdb -o gmx.gro -ignh -merge all -water tip3p -ff charmm36
```

(2) Define box

```
gmx editconf -f gmx.gro -o newbox.gro -c -d 1.0 -bt cubic -box 5.45 5.45 5.45
```

(3) Adding water solvent

gmx solvate -cp newbox.gro -o solv.gro -p topol.top -cs spc216.gro

(4) Replace water molecules (SOL) with NA and CL

gmx grompp -f em.mdp -c solv.gro -p topol.top -po em_ions_out.mdp -o ions.tpr

echo SOL | gmx genion -s ions.tpr -o solv_ions.gro -p topol.top -pname NA -pq 1 -np 10 -nname CL -nq -1 -nn 8

(5) Energy Minimization

gmx grompp -f em.mdp -c solv_ions.gro -p topol.top -po emout.mdp -o em.tpr

qsub em.com (gmx mdrun -deffnm em -ntomp 16 -ntmpi 1)

(6) Equilibration NVT 100 ps

gmx grompp -f nvt.mdp -c em.gro -p topol.top -po nvtout.mdp -o nvt.tpr

qsub nvt.com (gmx mdrun -deffnm nvt -ntomp 16 -ntmpi 1)

(7) Equilibration NPT (Berendsen) 400 ps

gmx grompp -f npt_B.mdp -c nvt.gro -t nvt.cpt -p topol.top -po npt_Bout.mdp -o npt_B.tpr

qsub npt_B.com (gmx mdrun -deffnm npt_B -ntomp 16 -ntmpi 1)

(8) Equilibration NPT (Parrinello-Rahman) 400 ps

gmx grompp -f npt_PR.mdp -c npt_B.gro -t npt_B.cpt -p topol.top -po npt_PRout.mdp -o npt_PR.tpr

qsub npt_PR.com (gmx mdrun -deffnm npt_PR -ntomp 16 -ntmpi 1)

(9) 1'st Production MD Run 500ns

gmx grompp -f 1pmd.mdp -c npt_PR.gro -t npt_PR.cpt -p topol.top -po 1pmdout.mdp -o 1pmd.tpr

qsub 1pmd.com (gmx mdrun -deffnm 1pmd -ntomp 16 -ntmpi 1)

(10) 2'nd Production MD Run 500ns

gmx grompp -f 2pmd.mdp -c 1pmd.gro -t 1pmd.cpt -p topol.top -po 2pmdout.mdp -o 2pmd.tpr

qsub 2pmd.com (gmx mdrun -deffnm 2pmd -ntomp 16 -ntmpi 1)

(11) 3'rd Production MD Run 500ns

gmx grompp -f 3pmd.mdp -c 2pmd.gro -t 2pmd.cpt -p topol.top -po 3pmdout.mdp -o 3pmd.tpr

qsub 3pmd.com (gmx mdrun -deffnm 3pmd -ntomp 16 -ntmpi 1)

Table S3.1*: Standard Parameters used in MD simulation, here e.g.* $\boldsymbol{P_{kpi}^{MDm}}$ *. Col.* ***I****; energy minimization (em.mdp). Col.* ***II****; NVT Equilibration (nvt.mdp). Col.* ***III****; NPT Equilibration, Berendsen (npt_B.mdp). Col.* ***IV****; NPT equilibration Parrinello-Rahman (npt_PR.mdp). Col.* ***V****; NPT, for trajectory 0-1499ns (1pmd.mdp, 2pmd.mdp, 3pmd.mdp).*

Parameters	**I**	**II**		**III**		**IV**		**V**	
	EM 50 ps	NVT 100 ps		NPT B 400 ps		NPT PR 400		NPT 0-1499 ns	
define		-DPOSRES		-DPOSRES		-DPOSRES			
emtol	100								
emstep	0.01								
integrator	steep	md		md		md		md	
dt	0.001	0.002		0.002		0.002		0.002	
nsteps	50000	50000		200000		200000		2.5e+8	
nstcomm	100	100		100		100		100	
continuation		no		yes		yes		yes	
constraints	none	h-bonds		h-bonds		h-bonds		h-bonds	
constraint_algorithm	lincs	lincs		lincs		lincs		lincs	
cutoff-scheme	Verlet	Verlet		Verlet		Verlet		Verlet	
ns_type	grid	grid		grid		grid		grid	
pbc	xyz	xyz		xyz		xyz		xyz	
rlist	1.2	1.2		1.2		1.2		1.2	
nstlist	10	10		10		10		10	
rcoulomb	1.2	1.2		1.2		1.2		1.2	
rcoulomb-switch	0	0		0		0		0	
rvdw	1.2	1.2		1.2		1.2		1.2	
rvdw_switch	1.0	1.0		1.0		1.0		1.0	
coulombtype	PME	PME		PME		PME		PME	
coulomb-modifier	Potential-shift-Verlet	Potential-shift-Verlet		Potential-shift-Verlet		Potential-shift-Verlet		Potential-shift-Verlet	
pme_order	6	6		6		6		6	
fourierspacing	0.1	0.1		0.1		0.1		0.1	
vdw-type	Cut-off	Cut-off		Cut-off		Cut-off		Cut-off	
vdw-modifier	force-switch	force-switch		force-switch		force-switch		force-switch	
DispCorr	no	no		no		no		no	
tcoupl		V-rescale		V-rescale		V-rescale		V-rescale	
tc-grps		Protein	Non-Protein	Protein	Non-Protein	Protein	Non-Protein	Protein	Non-Protein
tau_t		0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
ref_t		310	310	310	310	310	310	310	310
pcoupl		no		Berendsen		Parrinello-Rahman		Parrinello-Rahman	
pcoupltype				isotropic		isotropic		isotropic	
tau_p				1		1		1	
ref_p				1		1		1	
compressibility				4.5e-5		4.5e-5		4.5e-5	
refcoord_scaling				com		com			
gen_vel		yes		no		no		no	
gen_temp / gen_seed		310 / 173529							

S3.2 Postprocessing and obtaining a mean structure

Here it is provided additional GMX commands (Linux) and VMD script protocols. To complement what is written in §3.3.2.1.

(1) Use gmx make_ndx to output atom-indices for residues B11-B17(B32-B38) and CA atoms

```
echo -e "a CA & r 32-38\n q \n" | gmx make_ndx -f ../CPGMX/npt_PR.gro -o resfix.ndx &> LogCenterProteinBox.log
```

(2) Concatenate the three trajectories each 500ns (500000ps), with starting points of each indicated

```
echo -e "0 \n 499000 \n 999000" | gmx trjcat
-f ../CPGMX/1pmd.trr ../CPGMX/2pmd.trr ../CPGMX/3pmd.trr -o 123pmd.trr -settime &>> LogCenterProteinBox.log
```

(3) Protein centred in the box and simulation artefacts removed

```
echo 'Protein' 'System' | gmx trjconv -s ../CPGMX/1pmd.tpr -f 123pmd.trr -o 123pmdc.trr -pbc mol -center &>> LogCenterProteinBox.log
```

(4) Run VMD script MS.tcl, obtaining time of mean protein structure:

(i) A reference time is chosen, at the first iteration step. To be a random time above $8\ ns$ *(equilibration time) of the molecular dynamics trajectory. All times of the ensemble trajectory, were consecutively superimposed on this reference time. Using the transformation matrix which minimizes the RMSD of B11-B17* C_α *atoms of each time to the reference time.*

(ii) Then we obtained from this transformed trajectory, the time having the mean protein structure. By calculating the frame with lowest RMSD (eq. 3.5)) to all other times (above equilibration time, $t > 8\ ns$*). This frame was then used as reference time input in step (i). And another iteration of step (ii) yielded our representable mean-protein time (e.g. 101 ns) of the trajectory, including both protein and solvent. The iterations are not as necessary for our choice of a stable region (B11-B17 of an insulin monomer). But for a choice of a more fluctuating moiety, it could require several iterations of step (i) and (ii), to obtain a time containing the mean protein structure.*

(5) Output from trajectory the whole ‘System’(protein and solvent) of time having the mean protein structure (e.g. 101 ns, or 101000 ps) obtained in previous step, in structure-file MS.gro

```
echo 'System' | gmx trjconv -s ../CPGMX/1pmd.tpr -f 123pmdc.trr -dump 101000 -o MS/MS.gro &>> LogCenterProteinBox.log
```

(6) Using MS.gro to superimpose the time ensemble trajectory on. Using the B11-B17 CA-atom atom-index (from step 1) for least-square measurement. Outputting the whole ‘System'.

```
echo 'CA_&_r_32-38' 'System' | gmx trjconv -s MS/MS.gro -f 123pmdc.trr -o MS/123pmdcfMS.trr -n resfix.ndx -fit rot+trans &>> LogCenterProteinBox.log
```

Supplementary Chap. 4

S4.1 T-state Monomers packed in Hexamers

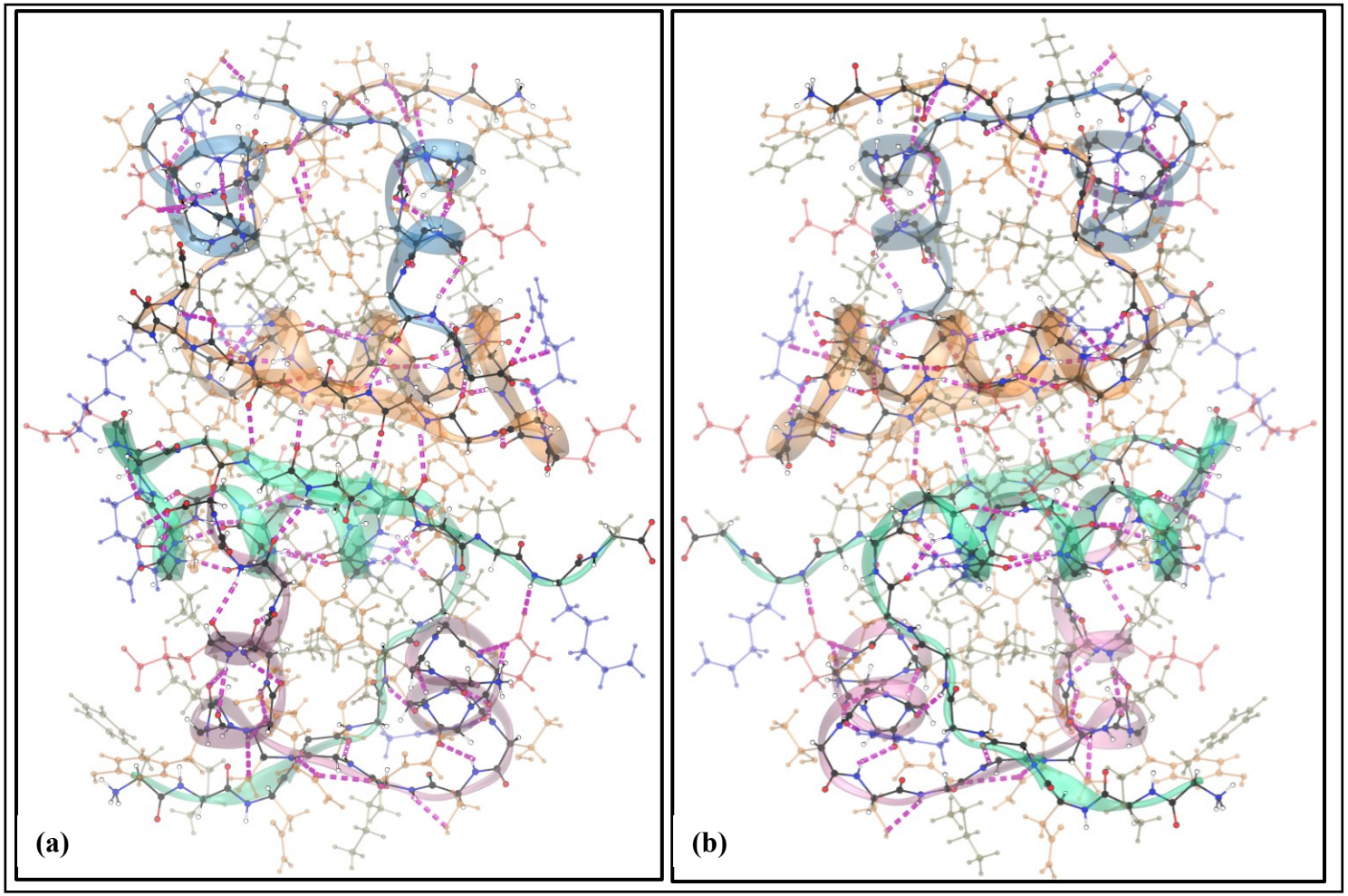


Figure S4.1*: Structure and HBs for an asymmetric dimer unit . (Bottom)* ***M2*** *(AC mauve, BC turquoise). (Top)* ***M1*** *(AC blue, BC orange).* ***(a)*** *"Front",* ***(b)*** *"back" ("front" rotated sideways 180°). The ACs, BCs and SCs being transparent, BB chalky (residue colouring as in Figure A1), HBs in dotted purple, calculated 87 HBs with criteria* $|\boldsymbol{r}_{AD}| < 3.5$ Å & $\varphi < 60°$*. The structure having only A configuration of those residues having alternate SC configurations in original PDB structure (****M1*** R^{B22}, K^{B29}*; M2* Q^{B4}, V^{B12}, E^{B21}, R^{B22}, T^{B27}*). Structure from PDB 4INS (biological assembly 7).*

***Table S4.2**: Intra-monomer higher angle HBs, for crystal dimer units, of PDB entry 4INS (**M12**), moreover for PDB entry 4E7T (**BM12**). Counting the HBs in format; AC_BC_AC&BC(sum). Corresponding nr with different criteria: $r_{AD} < 3.5$ Å & higher angle region ($60° < \varphi < 90°$); **M1** 24_28_2(54); **M2** 26_23_0(49); **BM1** 24_28_2(54); **BM2** 25_25_2(52).*

Donor	Acceptor	M1	M2	BM1	BM2		B6 L(HN)	B6 L(O)			✓	
							B9 S(HN)	B9 S(OG)	✓	✓	✓	✓
A1 G(H2)	A1 G(O)	✓	✓				B10 H(HN)	B9 S(OG)				✓
A3 V(HN)	A2 I(N)	✓	✓	✓	✓		B10 H(HN)	B9 S(N)	✓	✓	✓	✓
A4 E(HN)	A1 G(O)	✓	✓	✓			B11 N(HN)	B10 H(N)	✓	✓	✓	✓
A4 E(HN)	A3 V(N)	✓	✓	✓	✓		B12 H(HN)	B9 S(O)	✓	✓	✓	✓
A4 E(HN)	A4 E(OE2)				✓		B12 H(HN)	B11 L(N)	✓	✓	✓	✓
A5 Q(HN)	A2 I(O)	✓	✓	✓	✓		B13 E(HN)	B10 H(O)	✓	✓	✓	✓
A5 Q(HN)	A4 E(N)	✓	✓	✓	✓		B13 E(HN)	B12 V(N)	✓	✓	✓	✓
A6 C(HN)	A5 Q(N)	✓	✓	✓	✓		B14 A(HN)	B11 L(O)	✓	✓	✓	✓
A6 C(HN)	A6 C(SG)		✓		✓		B14 A(HN)	B13 E(N)	✓	✓	✓	✓
A7 C(HN)	A6 C(N)	✓	✓	✓	✓		B15 L(HN)	B12 V(O)	✓	✓	✓	✓
A8 T/A(HN)	A7 C(N)	✓	✓	✓	✓		B15 L(HN)	B14 A(N)	✓	✓	✓	✓
A8 T/A (HN)	A8 T/A (OG1)		✓				B16 Y(HN)	B13 E(O)	✓	✓	✓	✓
A9 S(HN)	A8/A T(OG1)		✓				B16 Y(HN)	B15 L(N)	✓	✓	✓	✓
A9 S(HN)	A8/A T(N)	✓	✓		✓		B17 L(HN)	B14 A(O)	✓	✓	✓	✓
A10 I/V(HN)	A10 I/V(O)	✓	✓		✓		B17 C(HN)	B16 Y(N)	✓	✓	✓	✓
A11 C(HN)	A11 C(O)	✓	✓	✓	✓		B18 V(HN)	B15 L(O)	✓	✓	✓	✓
A12 S(HN)	A12 S(OG)	✓	✓	✓	✓		B18 V(HN)	B17 L(N)	✓	✓	✓	✓
A14 Y(HN)	A12 S(OG)	✓	✓	✓	✓		B19 C(HN)	B18 V(N)	✓	✓	✓	✓
A14 Y(HN)	A13 L(N)	✓	✓	✓	✓		B20 G(HN)	B19 C(N)	✓	✓	✓	✓
A15 Q(HE21)	A5 Q(OE1)	✓		✓			B22 R(HN)	B20 G(O)		✓		
A15 Q(HE22)	A5 Q(OE1)			✓			B22 R(HN)	B21 E(N)	✓	✓	✓	✓
A15 Q(HN)	A12 S(O)			✓	✓		B23 G(HN)	B21 E(O)	✓			
A15 Q(HN)	A14 Y(N)	✓	✓	✓	✓		B23 G(HN)	B22 R(N)	✓	✓	✓	✓
A16 L(HN)	A13 L(O)	✓	✓	✓	✓		B24 F(HN)	B24 F(O)	✓	✓	✓	✓
A16 L(HN)	A15 Q(N)	✓	✓	✓	✓		B25 F(HN)	B25 F(O)	✓			
A17 E(HN)	A16 L(N)	✓	✓	✓	✓		B29 K(HN)	B28 P(N)	✓		✓	
A18 N(HN)	A17 E(N)	✓	✓	✓	✓		B29 K(HZ1)	B30 A(OT1)				✓
A18 N(HD21)	A18 N(N)	✓		✓			B30 A(HN)	B28 P(O)	✓		✓	
A19 Y(HN)	A17 E(O)		✓	✓	✓		B30 A(HN)	B29 L(N)	✓		✓	
A19 Y(HN)	A18 N(N)	✓	✓	✓	✓		A1 G(H1)	B30 A(OT1)	✓			
A20 C(HN)	A20 C(SG)				✓		A21 N(HD22)	B24 F(N)				✓
A20 C(HN)	A19 Y(N)	✓	✓	✓	✓		A21 N(HD22)	B24 F(O)			✓	
B2 V(HN)	B2 V(O)			✓			B4 N(HN)	A11 C(O)				✓
B4 Q(HN)	B4 Q(O)	✓		✓	✓		B5 H(HD1)	A9 S(O)	✓		✓	

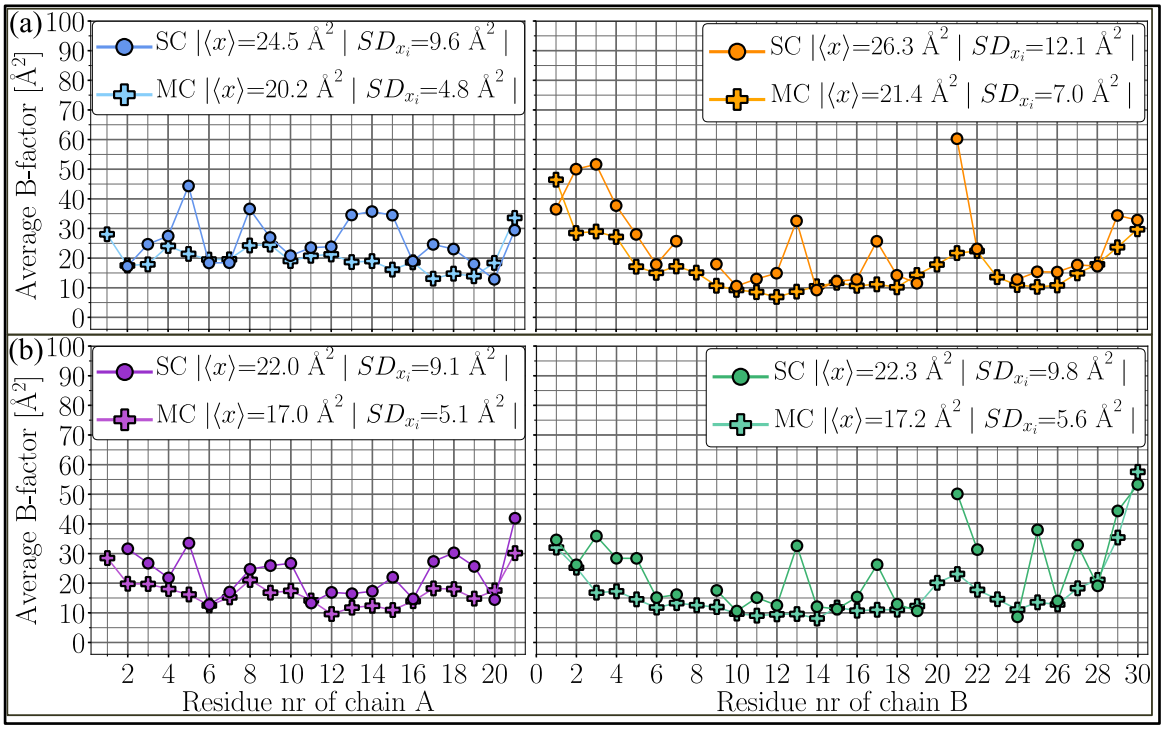


***Figure S4.2**: Average B-factor of a crystal T-state dimer unit for all atoms of either SC or MC . **(a) M1**. **(b) M2**. The statistics are separate for AC and BC where x in ⟨x⟩ refers to residue-wise mean B-factor of atoms referred, SC or MC, respectively. The temperature factors (or B-factors) from PDB 4INS (biological assembly 7).*

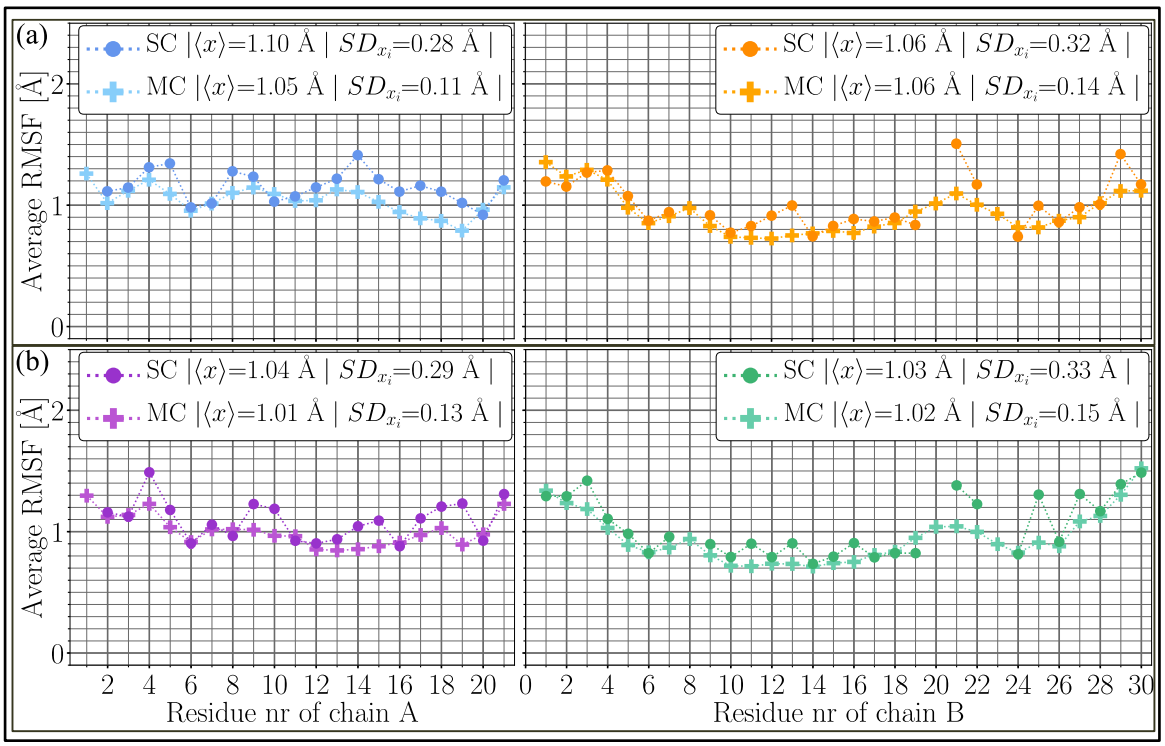


***Figure S4.3**: Average RMSF of crystal T-state monomersfor all non-hydrogen atoms of indicated selection for each residue. **(a) M1**. **(b) M2**. The statistics are separate for AC and BC where x in ⟨x⟩ refers to $RMSF_{\langle SC \rangle}$, and $RMSF_{\langle MC \rangle}$ respectively. The temperature factors (or B-factors) from PDB 4E7T.*

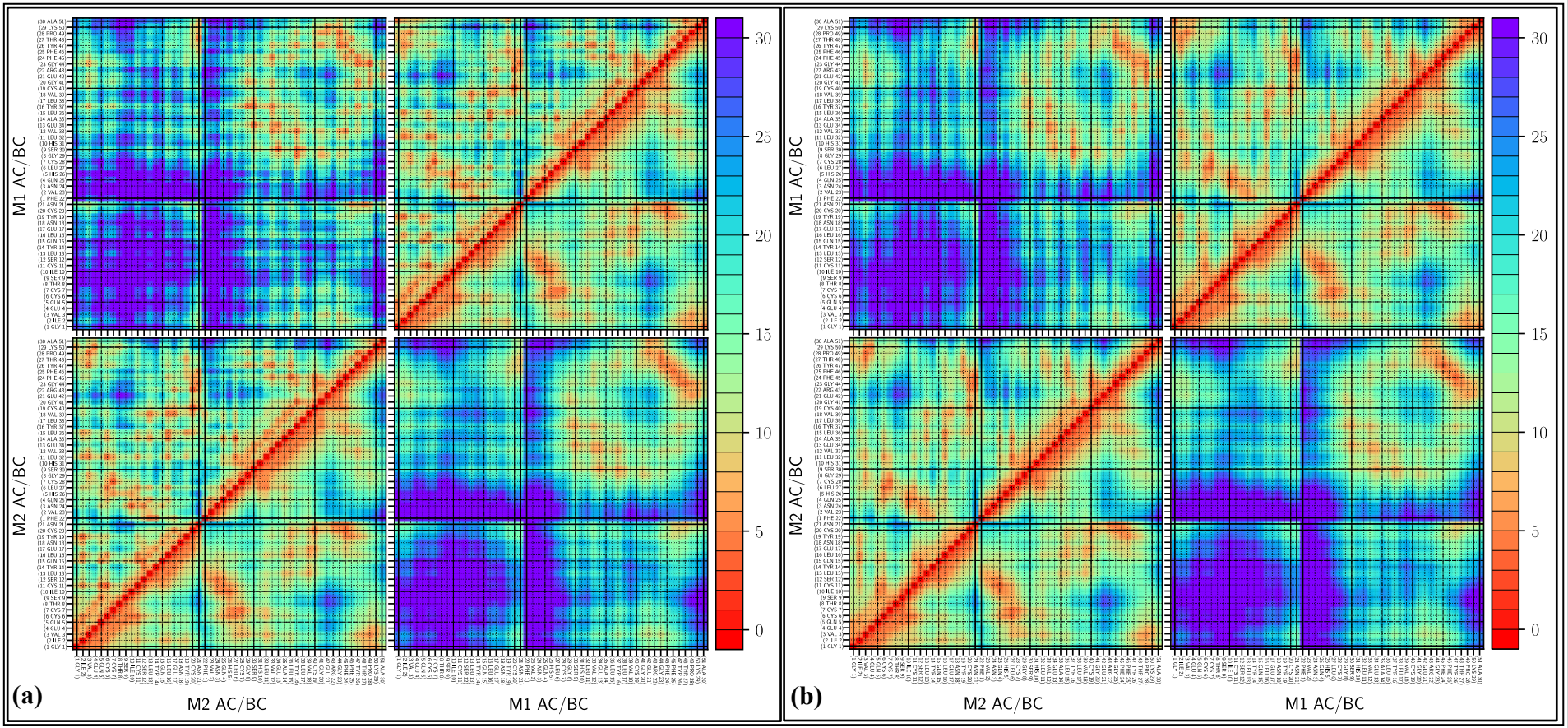


***Figure S4.4**: Residue-moiety distances within 30 Å, of an asymmetric dimer . Distances divided in matrix as: **(a)** Upper left is SC to SC geometric centre distances. Lower right is CA to CA-atom distances. **(b)** Upper left is SC to MC geometric centre distances. Lower right is MC to MC geometric centre distances. Diagonal as reference 0 Å (red), above 30 Å in purple. Distances divided in 0.5 steps, c.f. most of the CA-atom distances of adjacent residues are between 3.5-4.0 Å. The numbering is as follows M2 AC (1-21), BC (22-51); M1 AC (1-21), BC (22-51). Largest SC distance between any residues of **M2**, **M1** are 40.83 Å of **M1** S_{SC}^{A9} and **M2** S_{SC}^{A9}. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines, in black, are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Calculated from structure in PDB 4INS (biological assembly 7).*

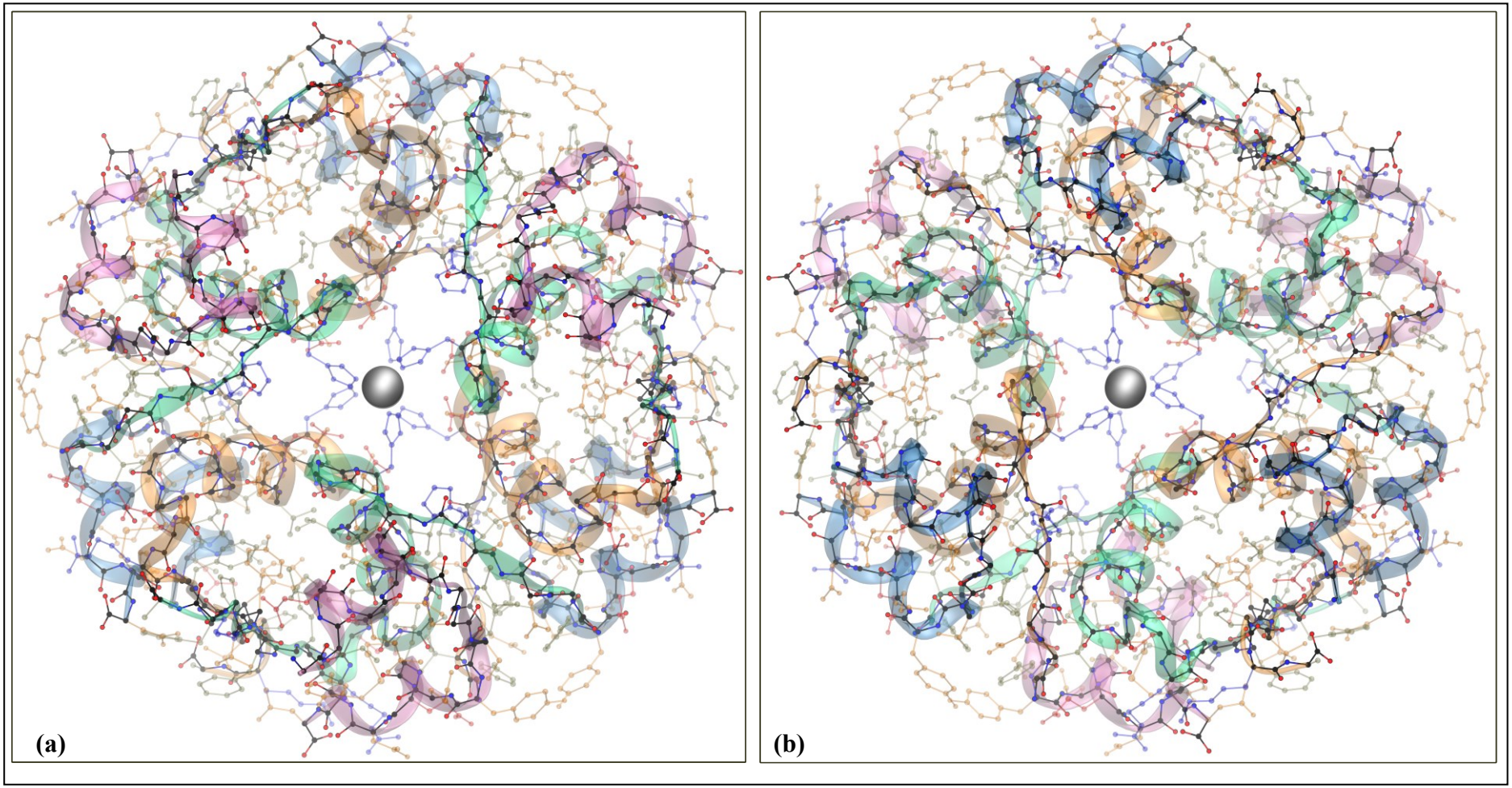


Figure S4.5*: Structure for a hexamer of asymmetric dimer units .* ***(a)*** *At front are trimer of* ***M2****.* ***(b)*** *At front are trimer of* ***M1.*** *"(a)" rotated sideways 180°. The ACs, BCs and SCs being transparent, BB (colouring as in Figure A1). The plausible HBs stabilizing the intra-hexamer structure (omitted), calculated 278 HBs with criteria* $|\boldsymbol{r}_{AD}| < 3.5$ Å & $\varphi < 60°$*. Residue* H^{B10} *plays an important role, in stabilising the dimers in a hexamer, through zinc coordination. Structure from* PDB 4INS *(biological assembly 3).*

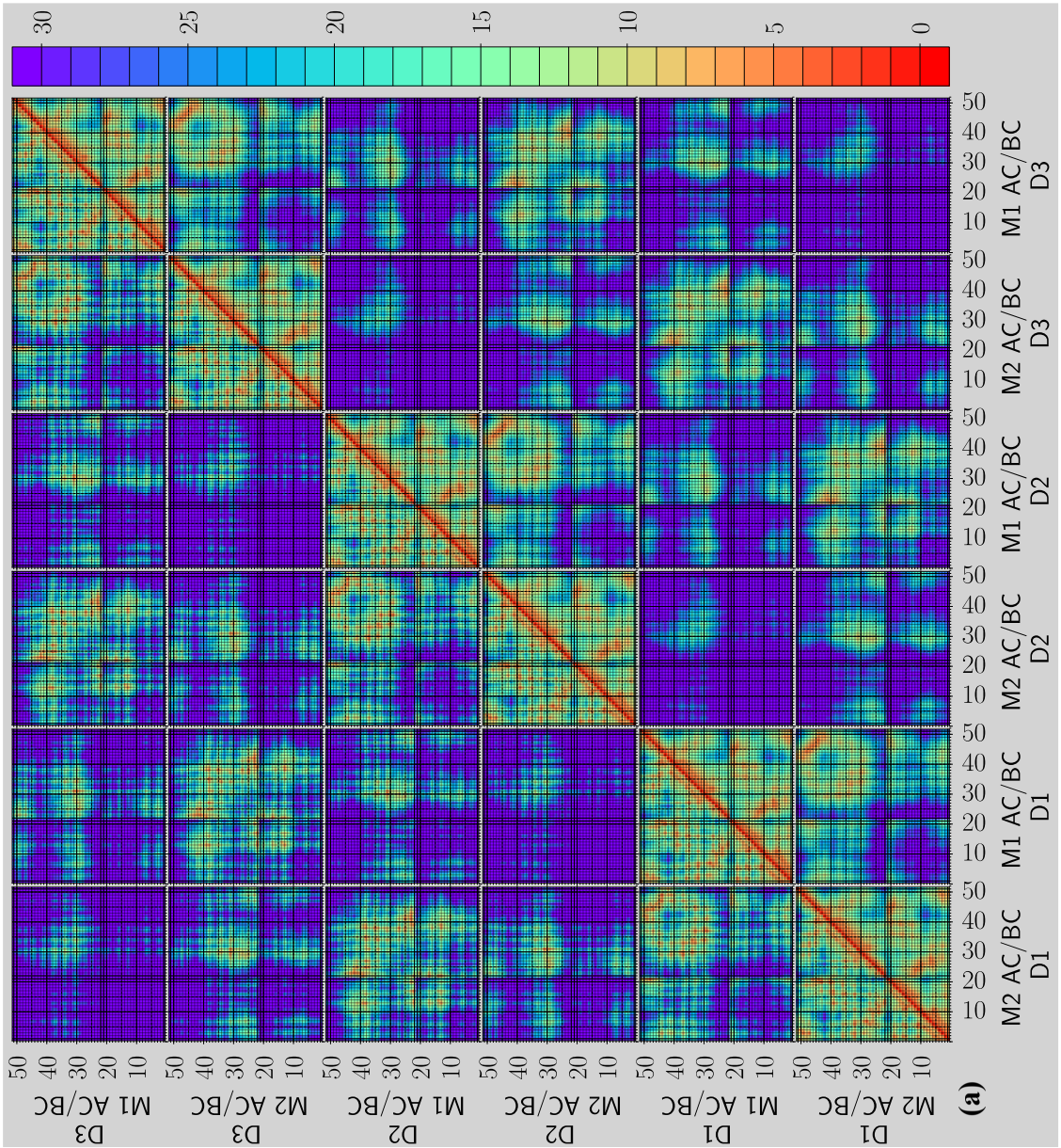


***Figure S4.6**: Residue-moiety distances within 30 Å of three asymmetric dimers (**D1**, **D2**, **D3**) of **M1** and **M2** an asymmetric unit. Diagonal as reference 0 Å (red), above 30 Å in purple. Matrix divided;**(a)** Upper left being SC to SC and lower right is CA to CA-atom distances. **(b)** Upper left being SC to MC and lower right is MC to MC geometric centre distances. Distances divided in 1.0 steps, c.f. most of the CA-atom distances of adjacent residues are between 3.0-4.0 Å. The 36 matrices shown are divided in order of (i, j) direction, that is (**D1 M1 M2** ... **D3 M1 M2**, **D1 M1 M2** ... **D3 M1 M2**). So if referring to one matrix e.g. (**D2 M1**, **D1 M2**) and any residues between any of these monomers indicated as e.g. (F^{B1}, F^{B1}) and if one residues touching several as (F^{B1}, L^{A13} & Y^{A14} & E^{A17} & V^{B18}). The furthest away SC distance of hexamer is 49.76 Å of (**D1 M1** $L^{B29(50)}$, **D3 M2** $A^{B30(51)}$). The graph has vector graphics and is zoomable. Grid-lines, in black, are distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Calculated from structure in PDB 4INS biological assembly 3. Further explanation of contacts here, continued next page:*

The symmetry of the three dimer units of the hexamer is readily seen in this matrix. Where it is further verifiable that the intra-dimer contacts are the same for each dimer. Moreover, revealed here is the distances between each of the six monomers, which also readily reveals the symmetry of contacts between the monomers.

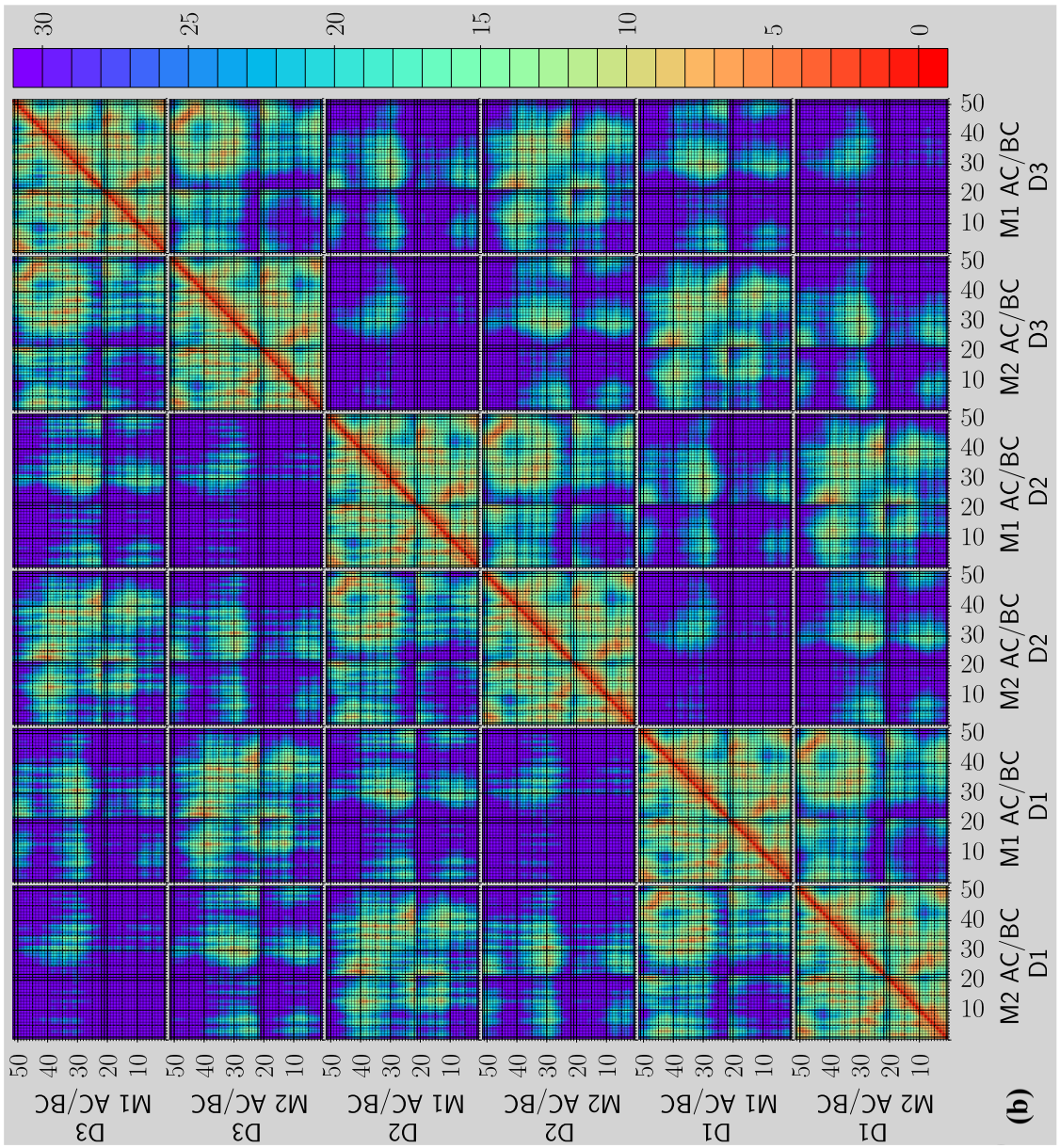


***Figure S4.6 continued**: The most residue contacts are for the **M1** of one dimer to that of **M2** of other dimers, three of those matrices shown for SC (**D2 M1**, **D1 M2**), (**D3 M2**, **D1 M1**), (**D3 M1**, **D2 M2**), whose matrices in itself appears essentially symmetric. Of these three some adjacent residues are e.g. diagonally (L^{A13}, L^{A13}), (Y^{A14}, Y^{A14}), (F^{B1}, F^{B1}), (E^{B13}, E^{B13}), (L^{B17}, L^{B17}). Moreover e.g. some more or less symmetrical off-diagonal adjacent contacts are (F^{B1}, L^{A13} & Y^{A14} & E^{A17} & V^{B18}), (V^{B2}, C^{B19} & G^{B20} & E^{B21} & R^{B22}), (L^{B17}, L^{A13} & Q^{B4} & L^{B6} & A^{B14}). There are 6 matrices essentially symmetric that corresponds to each of the three residue distance contacts of the same monomer between dimers, i.e. (**D2 M2**, **D1 M2**), (**D2 M1**, **D1 M1**), (**D3 M2**, **D1 M2**) etc. The closest of these are near the zinc-coordinated region of each trimer (H^{B10}, G^{B8} & S^{B9} & H^{B10}), (S^{B9}, E^{B13}), (E^{B13}, E^{B13}), the H^{B10} residues of each trimer being close due to coordination by zink.*

*The remainder three matrices are the distances between the furthest away **M1** and **M2** of separate dimers, e.g. (**D3 M1**, **D1 M2**). The closest residues (albeit not in direct contact) being around (S^{B9}, S^{B9}), (E^{B13}, E^{B13}), (L^{B17}, L^{B17}) and their respective off-diagonal closest residues.*

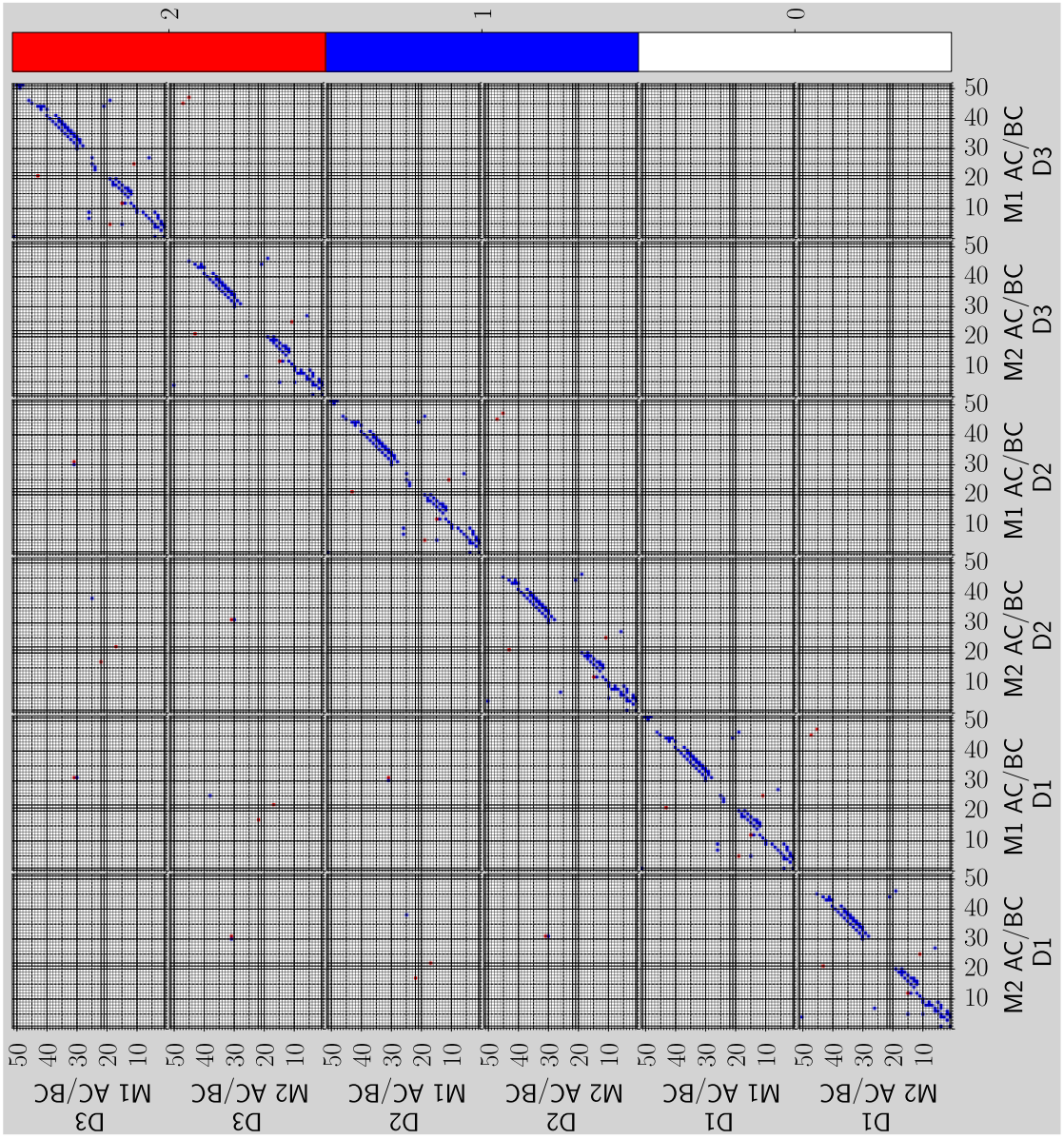


***Figure S4.7**: Hydrogen Bonds between residue-moieties of asymmetric hexamer. The matrices of HBs between residues calculated with: $|\mathbf{r}_{AD}| < 3.5$ Å, and $\varphi < 90°$, 604 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal are any HBs. Each of the 36 matrices has their chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. grid-lines for AC (1-21) in purple, blue and BC (22-51) turquoise, orange for **M2**, **M1** respectively. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Calculated from structure in PDB 4INS biological assembly 3.*

Further explanation of non intra-dimer HBs here:

*Three HBs (with $\varphi < 90°$) within monomers of trimers, i.e. between respectively (**D2 M2**, **D1 M2**), (**D3 M2**, **D2 M2**), (**D3 M1**, **D1 M1**), (S^{B9}_{HG1}, H^{B10}_{ND1}), (H^{B10}_{NE2}, H^{B10}_{HE2}), (H^{B10} H^{B10}_{HE2}), and in another order for (**D2 M1**, **D1 M1**), (**D3 M1**, **D2 M1**), (**D3 M2**, **D1 M2**), (H^{B10}_{ND1}, S^{B9}_{HG1}), (H^{B10}_{NE2}, H^{B10}_{HE2}), (H^{B10}_{HE2}, H^{B10}_{NE2}). The H^{B10} are zinc coordinated, with a zinc ion in between, however its interesting that it is in hydrogen bonding distance also. There are 5 HBs respectively for (**D2 M1**, **D1 M2**), (**D3 M2**, **D1 M1**), (**D3 M1**, **D2 M2**), (F^{B1}_{H1} & F^{B1}_{H2}, E^{A17}_{OE2}), (E^{A17}_{OE2}, F^{B1}_{H1} & F^{B1}_{H2}), (Q^{B4}_{HE22}, L^{B17}_{O}), (order "(L^{B17}_{O}, Q^{B4}_{HE22})" for "(**D3 M2**, **D1 M1**)").*

S4.2 Properties of IR-fragments contiguous to insulin

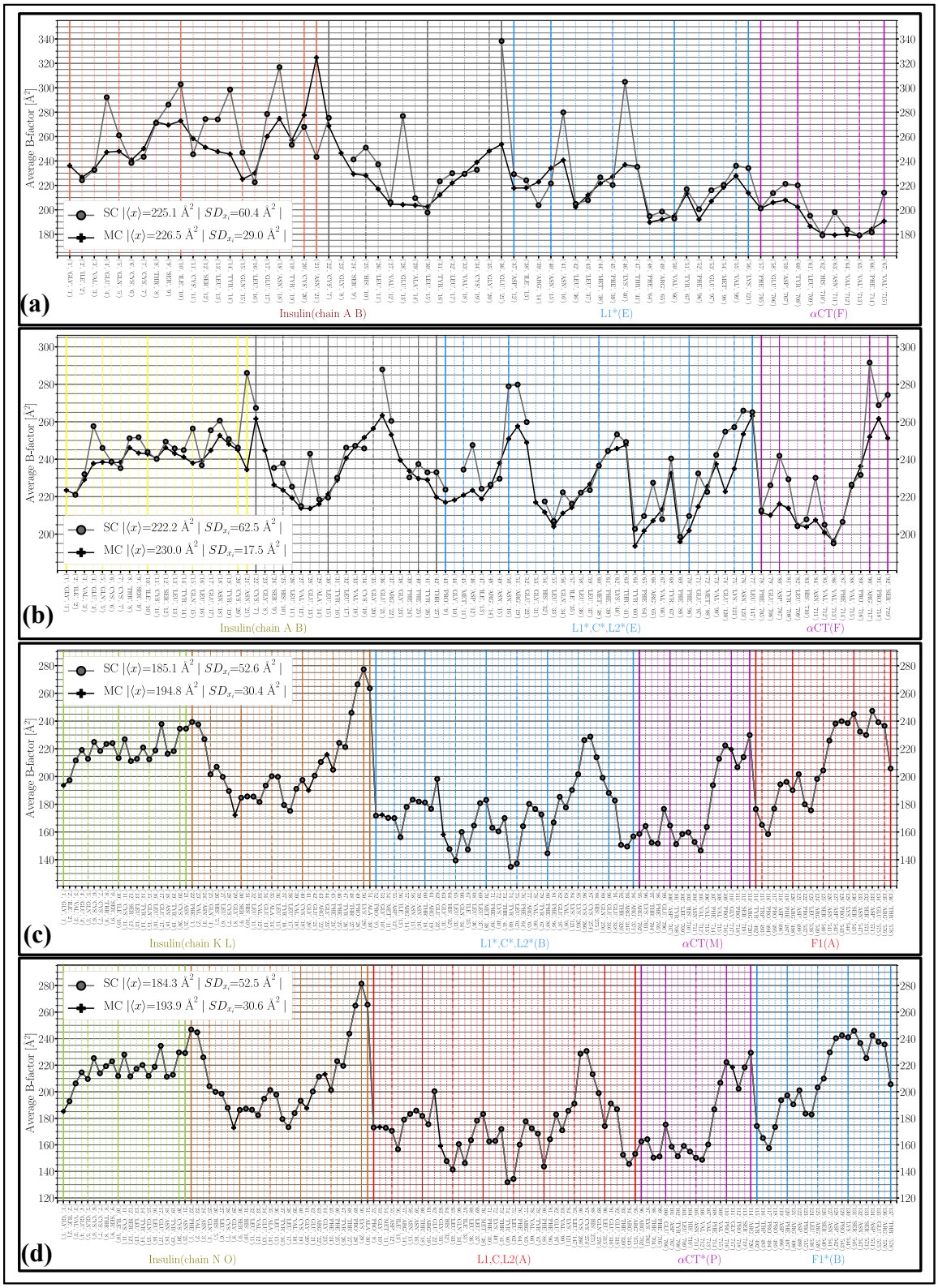


***Figure S4.8**: Average B-factors for insulin contiguous residues in IR-fragments **(a)** $CF^{3W11}_{(A,B)}$. **(b)** $CF^{4OGA}_{(A,B)}$. **(c)** $CF^{6CE9}_{(K,L)}$. **(d)** $CF^{6CE9}_{(N,O)}$. The graphs have zoomable vector graphics.*

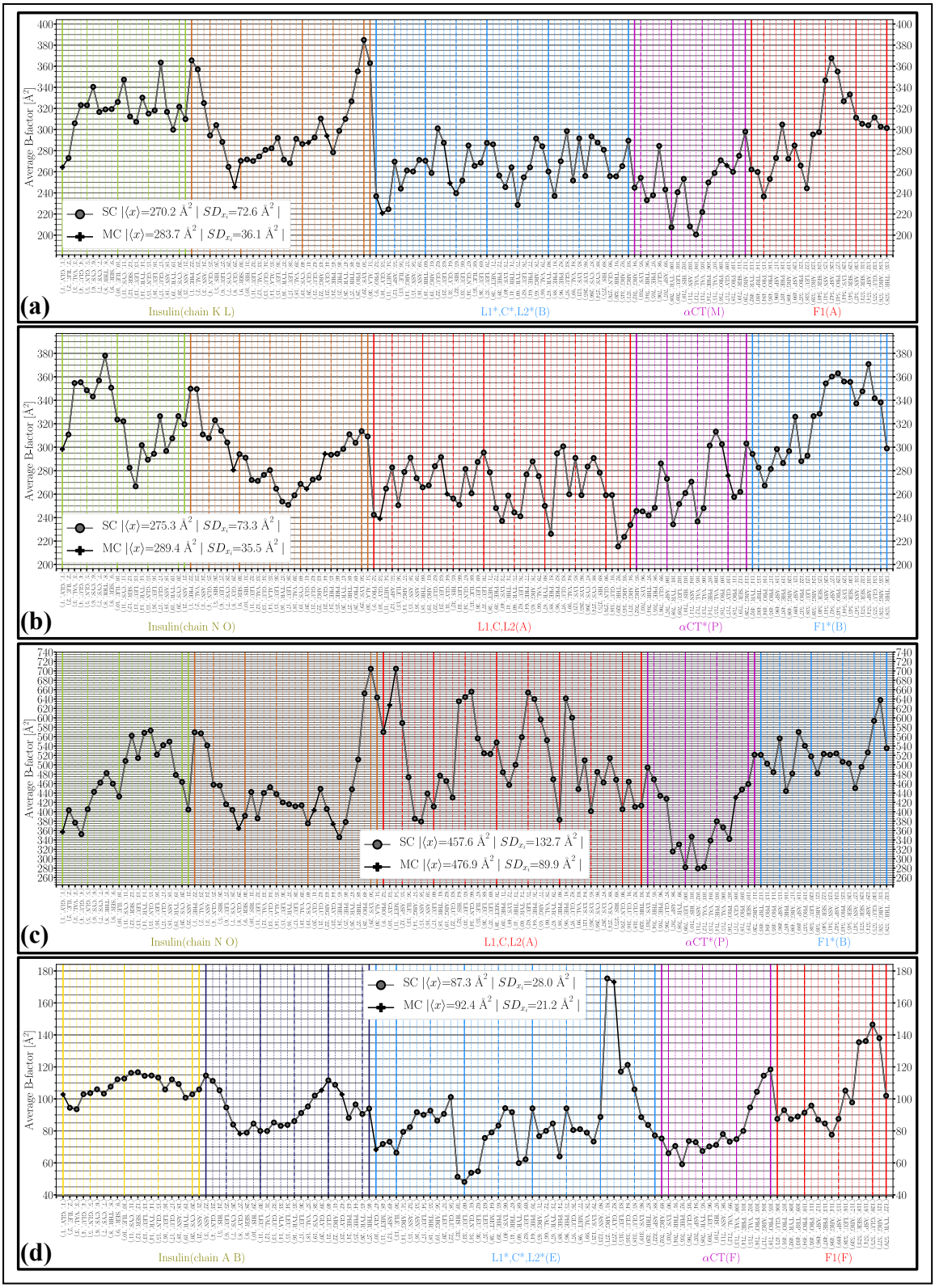


***Figure S4.9**: Average B-factors for insulin contiguous residues in IR-fragments*

***(a)** $CF^{6CEB}_{(K,L)}$. **(b)** $CF^{6CEB}_{(N,O)}$. **(c)** $CF^{6CE7}_{(N,O)}$. **(d)** $CF^{6HN5}_{(A,B)}$. The graphs have zoomable vector graphics.*

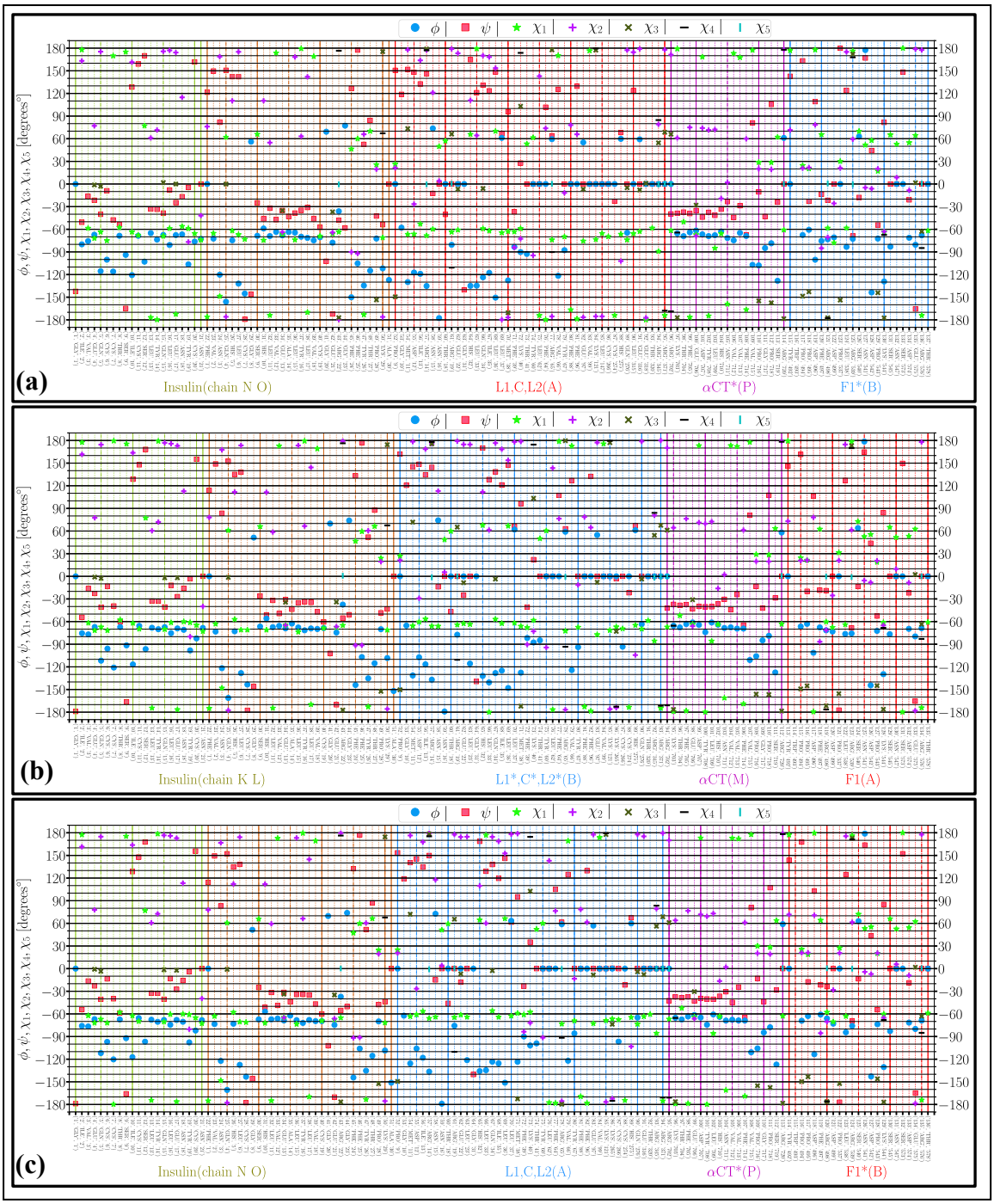


***Figure S4.10**: Dihedral angles for insulin contiguous residues in IR-fragments . For structures **(a)** $\boldsymbol{CF^{6CE9}_{(N,O)}}$, **(b)** $\boldsymbol{CF^{6CEB}_{(K,L)}}$, **(c)** $\boldsymbol{CF^{6CEB}_{(N,O)}}$. The graphs are zoomable vector graphics.*

Figure S4.11*: Structure of IR fragment nearest to bound insulin,* $\boldsymbol{CF^{3W11}_{(A,B)}}$*. Shown are the whole residues having at least one non-hydrogen atom within 10 Å from any non-hydrogen atom of insulin (dark-orange, chain A; gray-black, chain B), which are belonging to the domains of L1*, (blue, chain E) and the αCT (purple, chain F).* ***(a)*** *'Front'.* ***(b)*** *'Back' ~180° sideway rotation.*

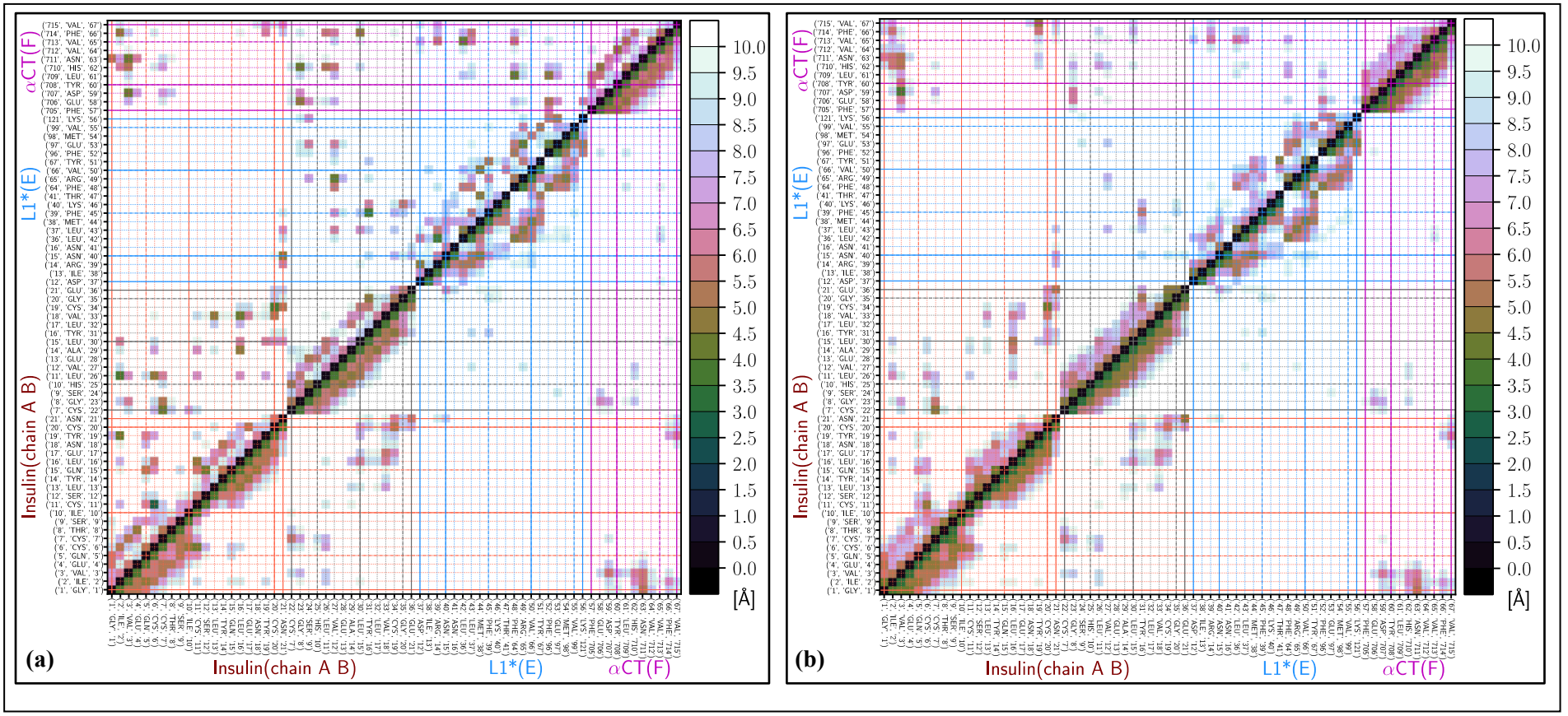


Figure S4.12: *Distances between residue-moieties of* $\boldsymbol{CF}^{\boldsymbol{3W11}}_{\boldsymbol{(A,B)}}$ *. The line-colouring depiction is showing insulin, orange for chain A, dark-grey for chain B. Structure with hydrogens, including the whole residues (from PDB 3W11) having at least one non-hydrogen atom within 10 Å from any non-hydrogen atom of insulin, which are belonging to the domains of L1* (blue, chain E) and αCT (purple, chain F). Distance-matrices of geometric centres from following selections* ***(a)*** *SC to SC upper left, CA to CA atom lower right,* ***(b)*** *SC to MC at upper left, MC to MC lower right. Chains residues renumbered sequentially 1-67 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-67. Actual chain and residue-naming as in PDB 3W11 [71].*

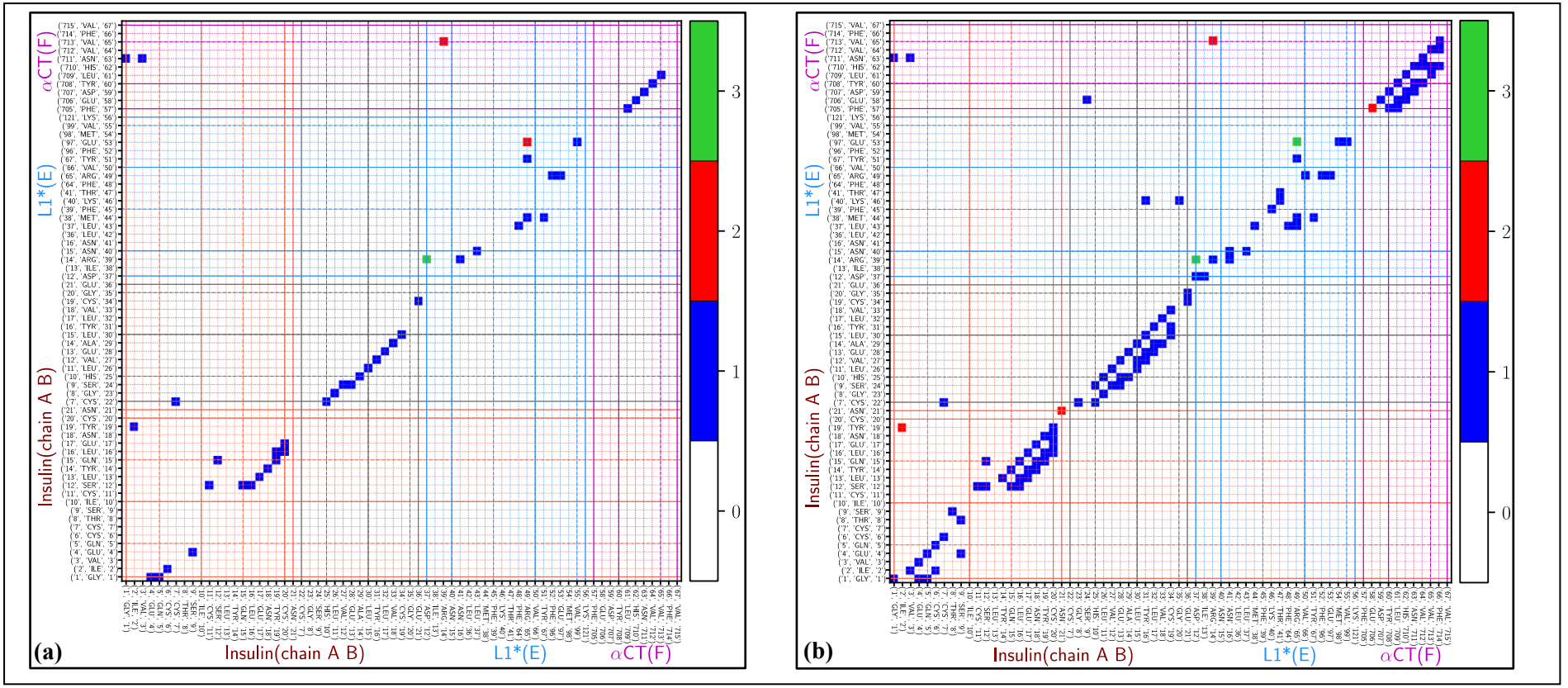


***Figure S4.13**: Hydrogen bonds between residue-moieties of $\boldsymbol{CF}^{\boldsymbol{3W11}}_{(\boldsymbol{A},\boldsymbol{B})}$. The line-colouring depiction is showing insulin in orange for chain A, and dark-grey for chain B. Structure with hydrogens, including the whole residues (from PDB 3W11) having at least one non-hydrogen atom within 10 Å from any non-hydrogen atom of insulin, which are belonging to the domains of L1* (blue, chain E) and the αCT* (purple, chain F). The matrices of HBs between residue-moieties calculated with $|\boldsymbol{r}_{AD}| < 3.5$ Å and **(a)** $\varphi < 60°$, 50 HBs, **(b)** $\varphi < 90°$, 123 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs. Chains residues renumbered sequentially 1-67 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-67. Actual chain and residue-naming as in PDB 3W11 [71].*

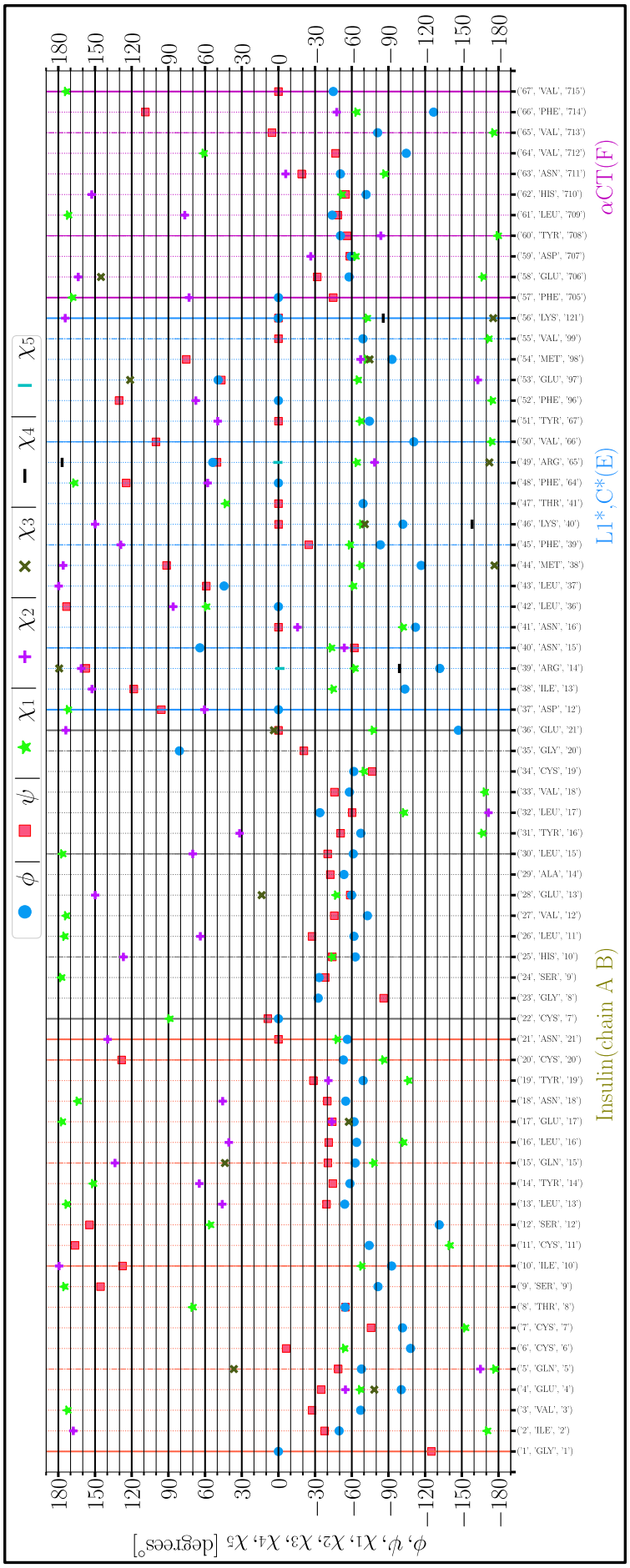


Figure S4.14: *Dihedral angles of residues in* $\mathbf{CF}^{\mathbf{3W11}}_{(\boldsymbol{A,B})}$. *The line-colouring depiction is showing insulin in orange for chain A, and grey for chain B. Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1* , C* (blue, chain E) and the αCT* (purple, chain F).*

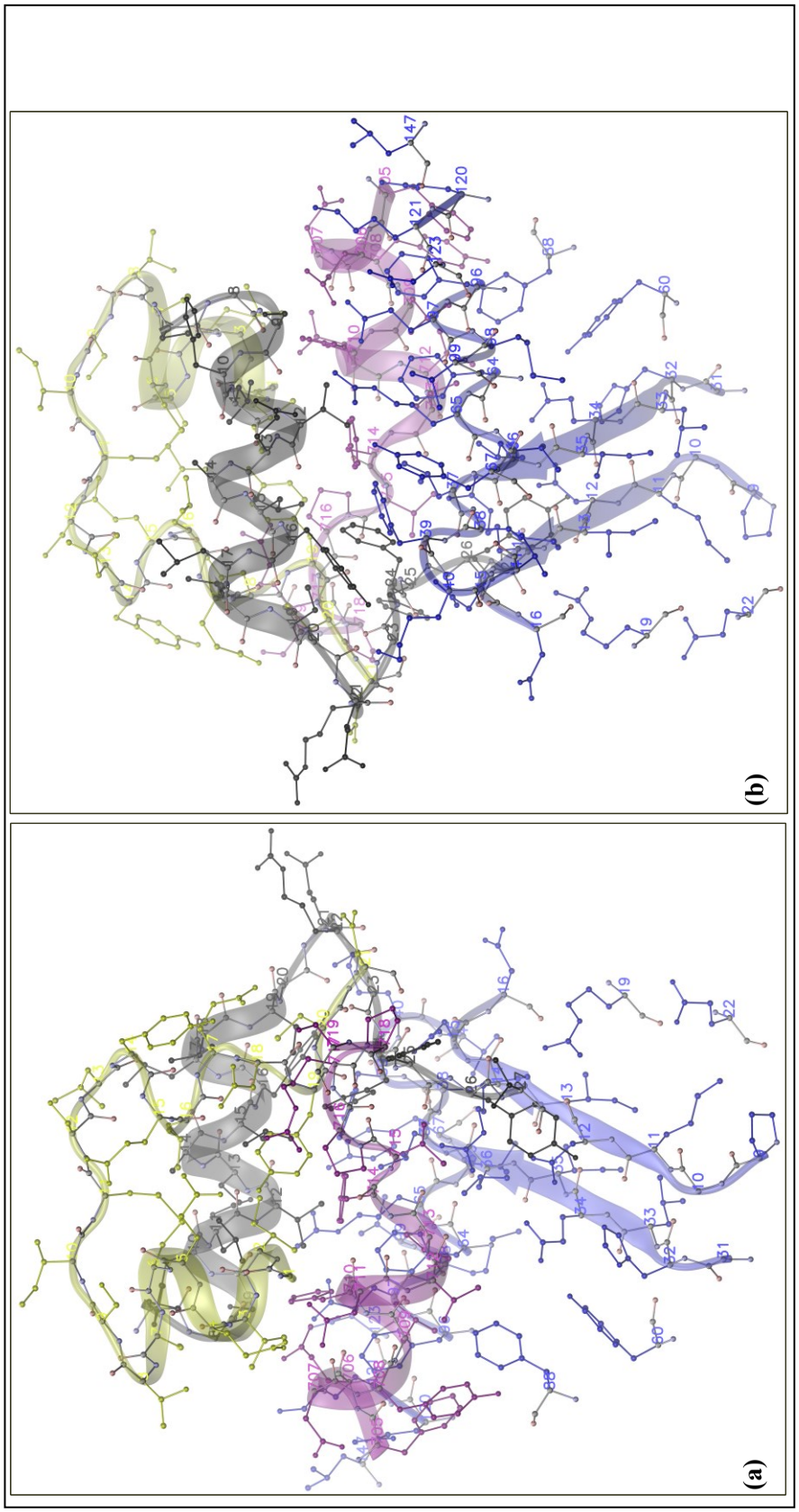


***Figure S4.15**: Structure of IR fragment nearest to bound insulin from* $\mathbf{CF}^{\mathbf{40GA}}_{(A,B)}$ *. Shown are the whole residues having at least one atom within 10 Å from any atom of insulin (yellow, chain A; gray, chain B); which are belonging to the domains of L1*, (blue, chain E) and the αCT (purple, chain F). **(a)** 'Front'. **(b)** 'Back' ~180° sideway rotation.*

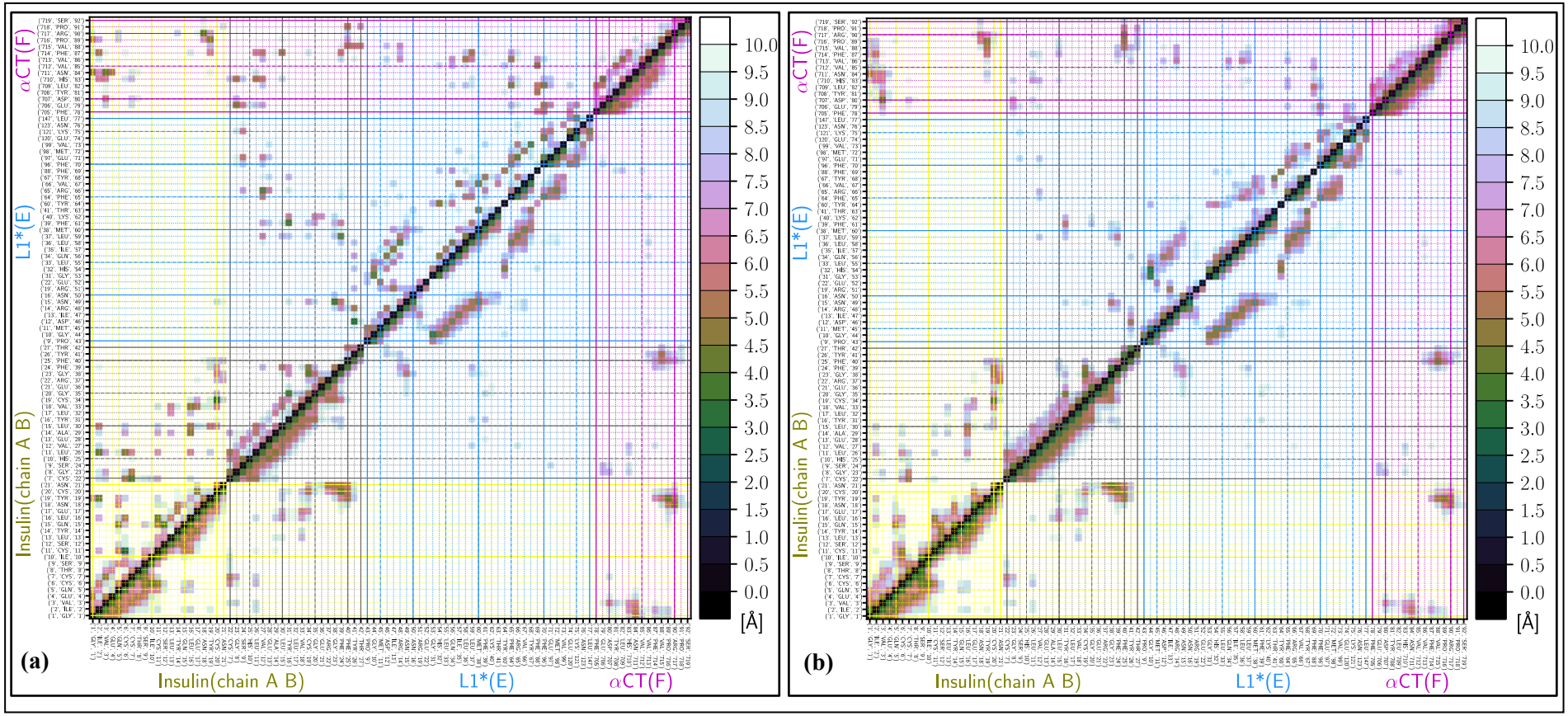


***Figure S4.16**: Distances between residue-moieties of $\mathbf{CF}^{4OGA}_{(A,B)}$. The line-colouring depiction is showing insulin, yellow for chain A, grey for chain B. Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1* (blue, chain E) and the αCT* (purple, chain F). Distance matrices of geometric centres from following selections **(a)** SC to SC upper left, CA to CA atom lower right, **(b)** SC to MC at upper left, MC to MC lower right. Chains residues renumbered sequentially 1-92 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-92. Actual chain and residue-naming as in PDB 4OGA [69].*

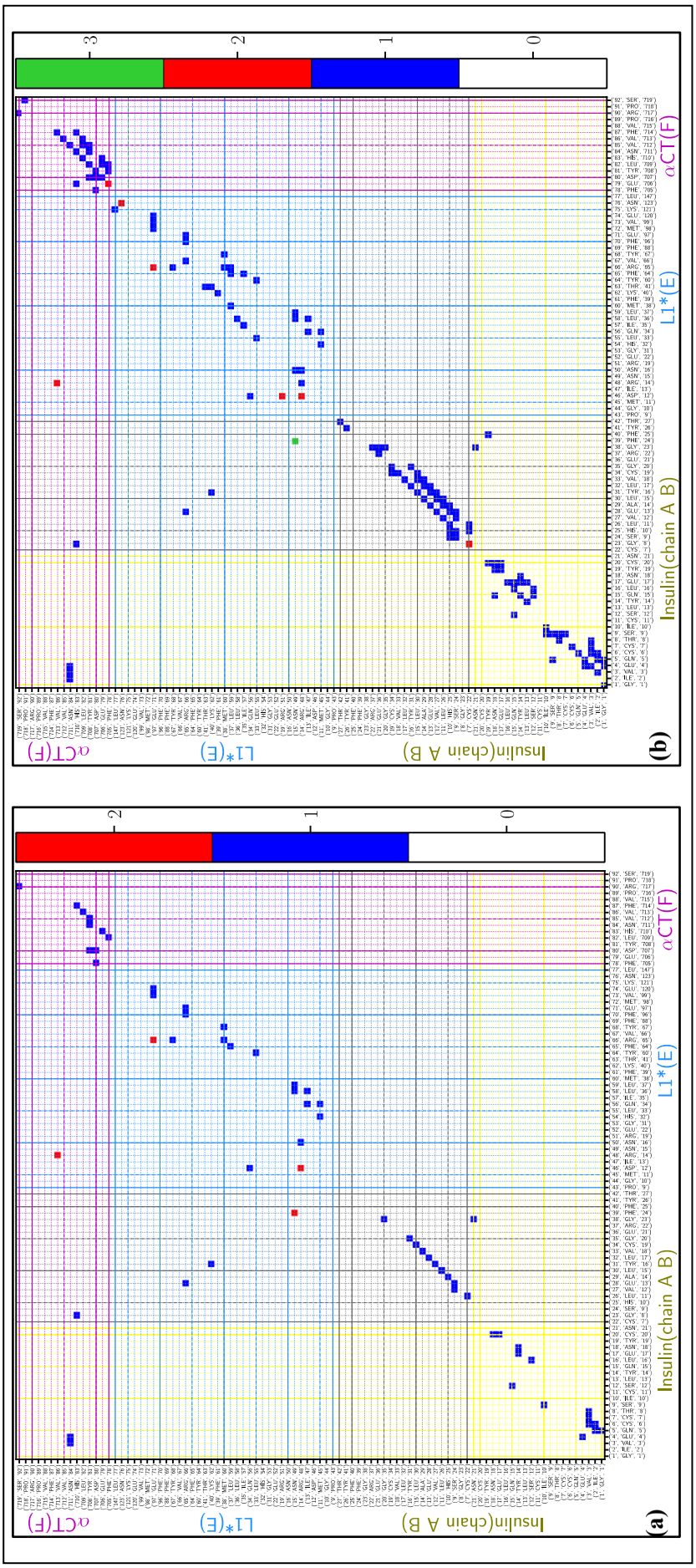


Figure S4.17**: Hydrogen Bonds between residue-moieties of* $\boldsymbol{CF^{4OGA}_{(A,B)}}$ *. The line-colouring depiction is showing insulin, yellow for chain A, grey for chain B. Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1* (blue, chain E) and the αCT* (purple, chain F). The matrices of HBs between residues calculated with:* $|\boldsymbol{r}_{AD}| < 3.5$ *Å, and* ***(a) $\varphi < 60°$*, 66 HBs,* ***(b)*** $\varphi < 90°$*, 155 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs. Chains residues renumbered sequentially 1-92 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-92. Actual chain and residue-naming as in PDB 4OGA [69].*

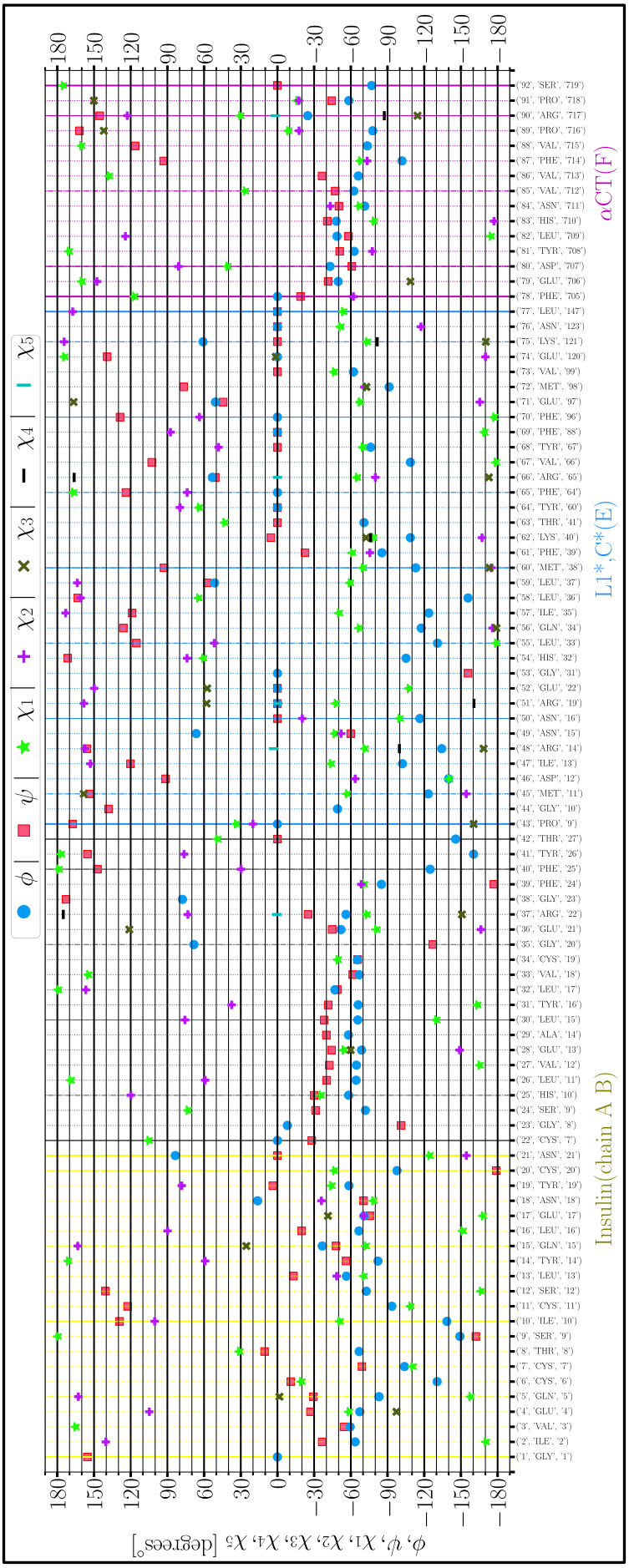


Figure S4.18: *Dihedral angles of residues in* $\mathbf{CF}^{\mathbf{40GA}}_{(\mathbf{A,B})}$ *. Shown are the whole residues having at least one atom within 10 Å from any part of insulin (yellow, chain A; gray, chain B), which are belonging to the domains of L1*,C* (blue, chain E) and the αCT (purple, chain F).*

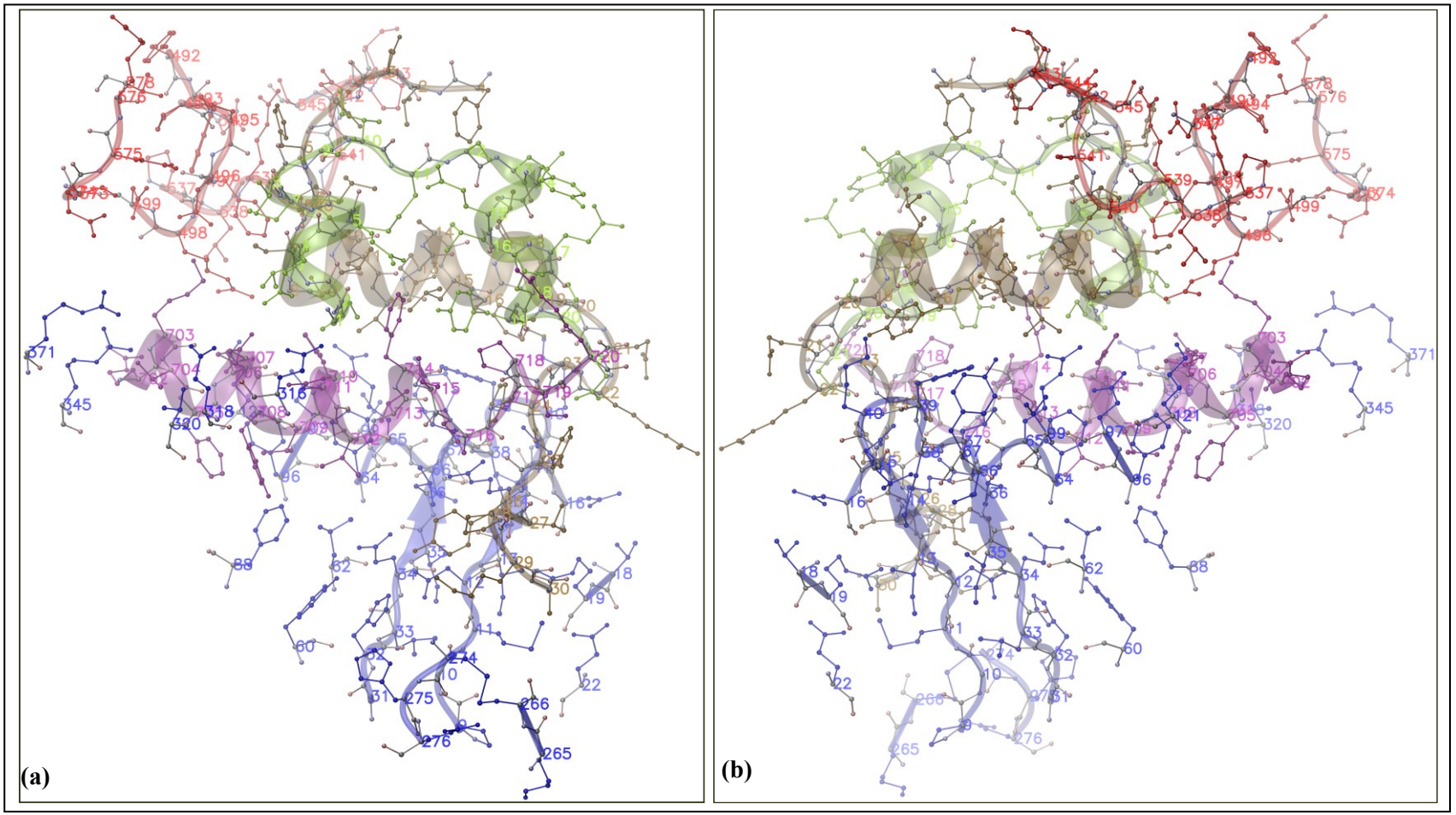


Figure S4.19**: Structure of IR fragment nearest to bound insulin from* $\boldsymbol{CF_{(K,L)}^{6CE9}}$ *. Shown are the whole residues having at least one atom within 10 Å from any part of insulin (lime-green, chain A; ochre-brown, chain B), which are belonging to the domains of L1*, C*, L2* (blue, chain B) and the αCT (purple, chain M), F1* (chain A).* ***(a) *'Front'.* ***(b)*** *'Back' ~180° sideway rotation.*

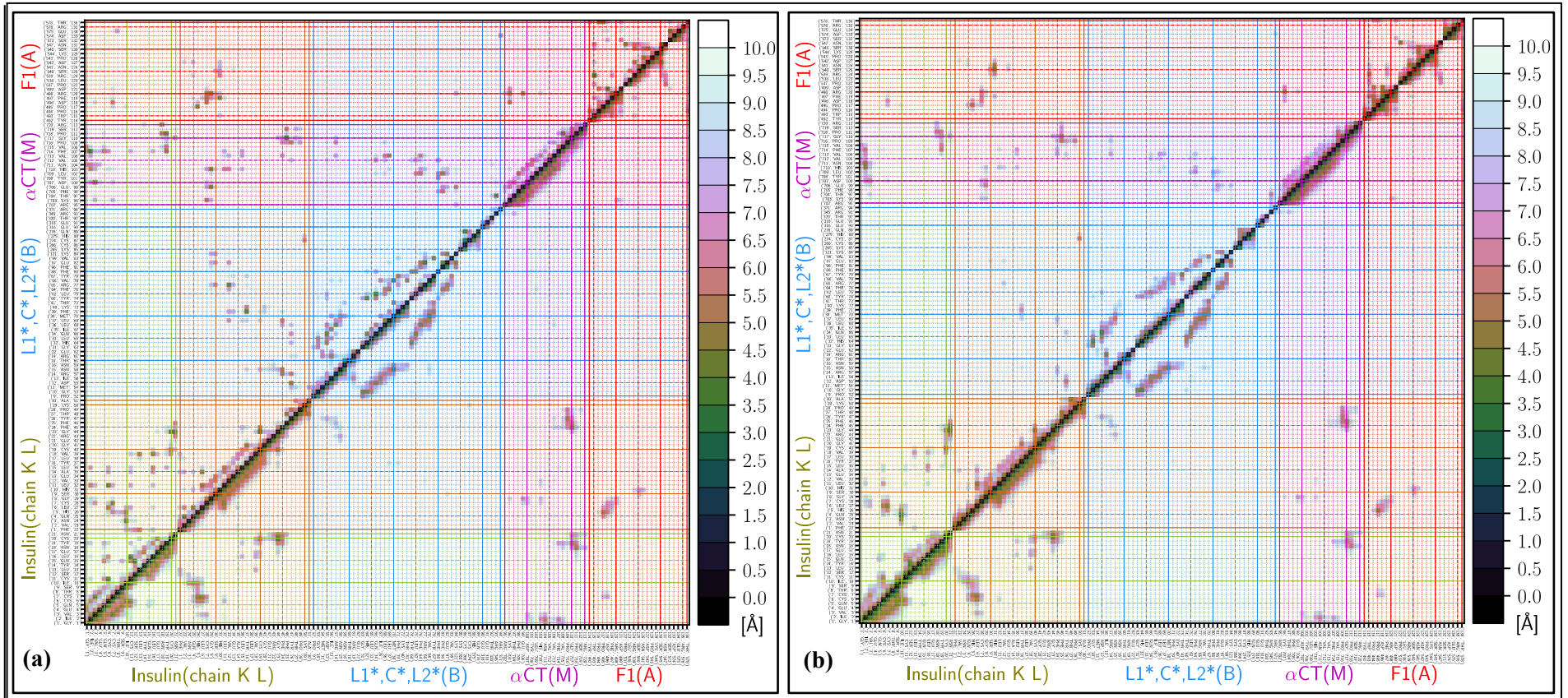


Figure S4.20: *Distances between residues of* $\boldsymbol{CF}^{\boldsymbol{6CE9}}_{\boldsymbol{(K,L)}}$ *. The depiction is showing insulin chain K(lime-green), L(ochre-brown). Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1*, C*, L2* (blue, chain B) and the αCT (purple, chain M), F1(red, chain A). Residue-distance matrix, of geometric centres of following selections* ***(a)*** *SC to SC upper left, CA to CA atom lower right,* ***(b)*** *SC to MC at upper left, MC to MC lower right. Chains residues renumbered sequentially 1-136 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-136. Actual chain and residue-naming as in PDB 6CE9 [70].*

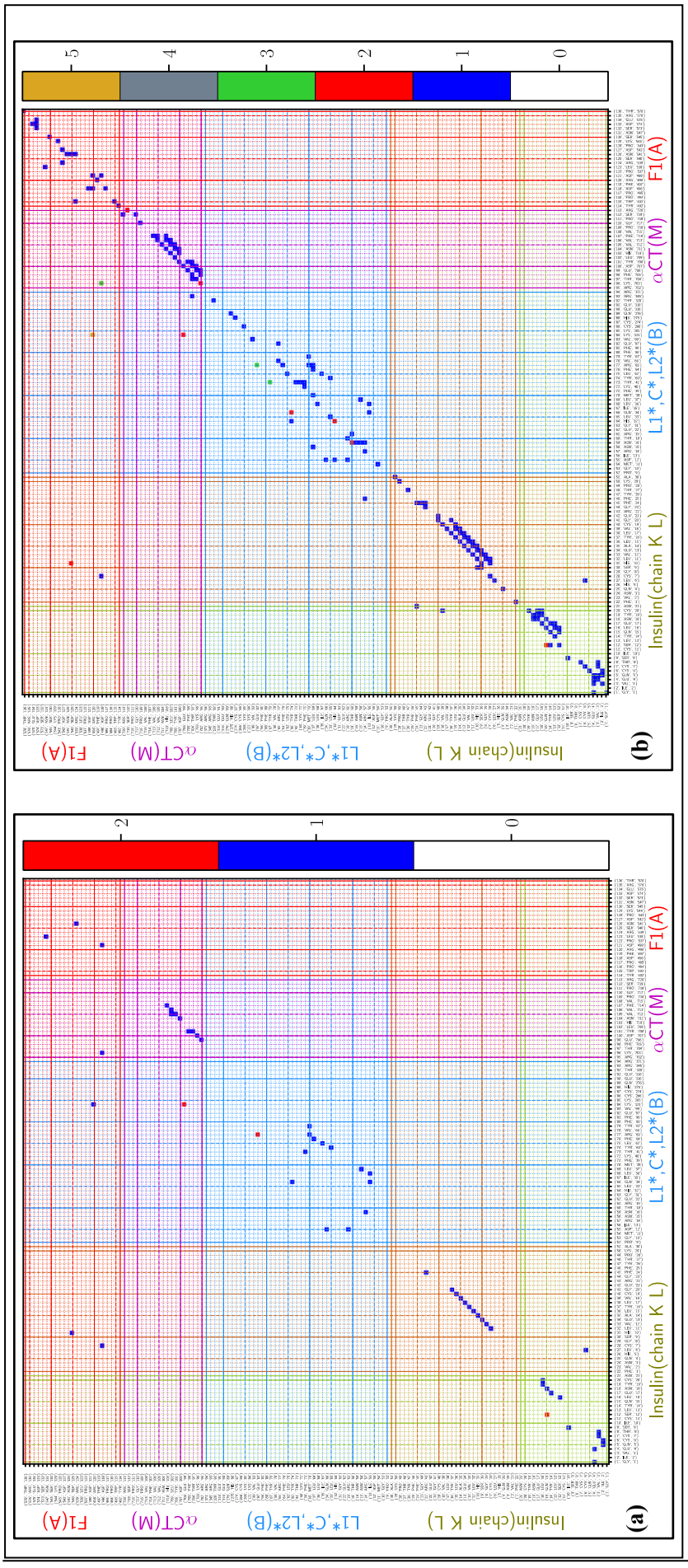


Figure S4.21**: Hydrogen Bonds between residues of* $\mathbf{CF}^{\mathbf{6CE9}}_{(K,L)}$ *. The depiction is showing insulin chain K(lime-green), B(brown). Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1*, C*, L2* (blue, chain B) and the αCT (purple, chain M), F1(red, chain A). The matrices of HBs between residues calculated with:* $|\mathbf{r}_{AD}| < 3.5$ *Å, and* ***(a) $\varphi < 60°$*, 59 HBs,* ***(b)*** $\varphi < 90°$*, 206 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs. Chains residues renumbered sequentially 1-136 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-136. Actual chain and residue-naming as in PDB 6CE9 [70].*

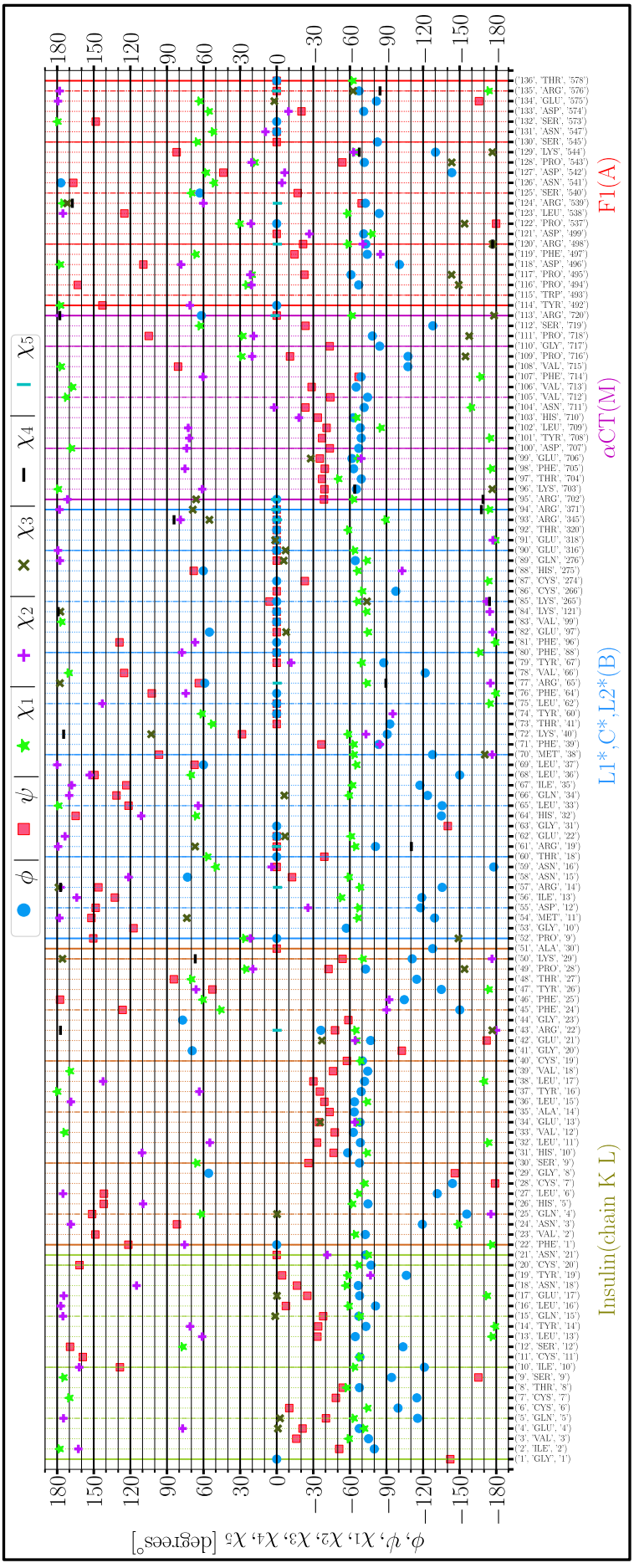


***Figure S4.22**: Dihedral angles of residues in* $\boldsymbol{CF^{6CE9}_{(K,L)}}$ *. The depiction is showing insulin chain K(lime-green), B(brown). Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1*, C*, L2* (blue, chain B) and the αCT(purple, chain M), F1(red, chain A).*

Figure S4.23: *Structure of IR fragment nearest to bound insulin from* $\boldsymbol{CF^{6CE7}_{(N,O)}}$. *Shown are the whole residues having at least one atom within 10 Å from any part of insulin (lime-green, chain A; ochre-brown, chain B), which are belonging to the domains of L1, C, L2 (red, chain A) and the αCT* (purple, chain P), F1 (blue, chain B).* ***(a)*** *'Front'.* ***(b)*** *'Back' ~180° sideway rotation.*

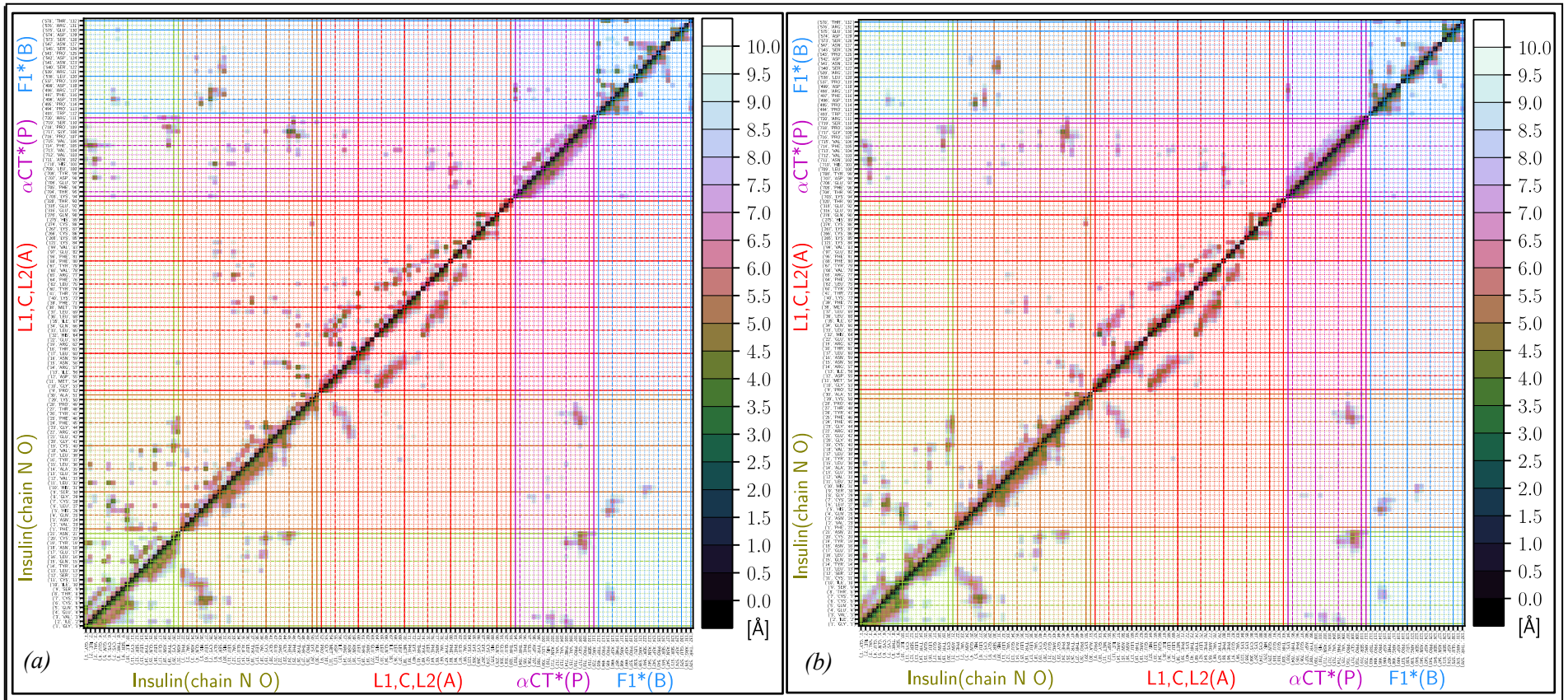


Figure S4.24: *Distances between residues of* $\boldsymbol{CF^{6CE7}_{(N,O)}}$ *. The depiction is showing insulin chain N(lime-green), O(ochre-brown). Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1, C, L2 (red, chain A) and the αCT* (purple, chain P), F1(blue, chain B). Residue-distance matrix, of geometric centres of following selections* ***(a)*** *SC to SC upper left, CA to CA atom lower right ,* ***(b)*** *SC to MC at upper left, MC to MC lower right. Chains residues renumbered sequentially 1-132 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-132. Actual chain and residue-naming as in PDB 6CE7 [70].*

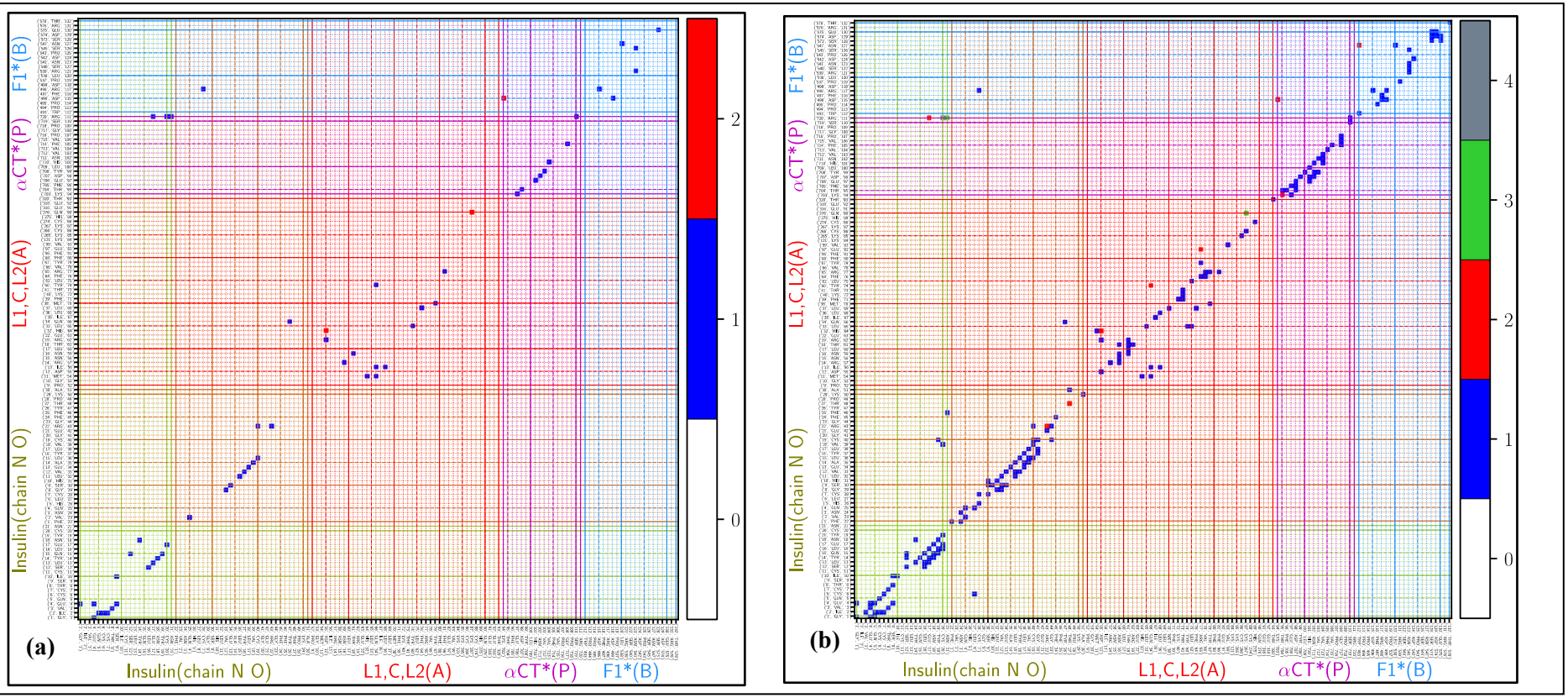


Figure S4.25: *Hydrogen Bonds between residues of* $\boldsymbol{CF}^{6CE7}_{(N,O)}$. *The depiction is showing insulin chain N(lime-green), O(brown). Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1, C, L2 (red, chain A) and the αCT* (purple, chain P), F1*(blue, chain B). The matrices of HBs between residues calculated with:* $|\boldsymbol{r}_{AD}| < 3.5$ Å, *and* ***(a)*** $\varphi < 60°$, *63 HBs,* ***(b)*** $\varphi < 90°$, *208 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs. Chains residues renumbered sequentially 1-132 (nearest graph), and with actual residue name and number. Zoomable vector graphics, grid-lines are chain coloured distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-132. Actual chain and residue-naming as in PDB 6CE7 [70].*

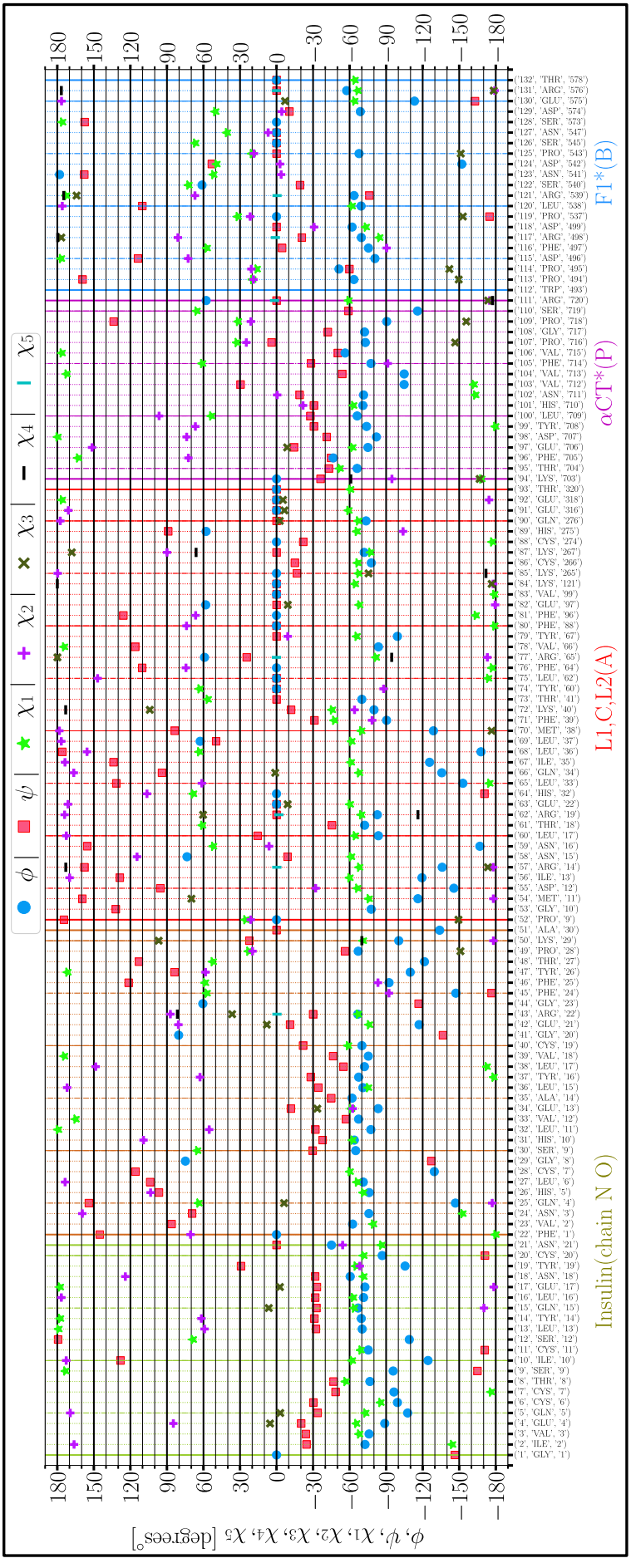


***Figure S4.26**: Dihedral angles of residues in* $\mathbf{CF}^{\mathbf{6CE7}}_{(\mathbf{N,O})}$ *. The depiction is showing insulin chain N(lime-green), O(brown). Moreover, the whole residues having at least one atom within 10 Å from any part of insulin, which are belonging to the domains of L1, C, L2 (red, chain A) and the αCT* (purple, chain P), F1*(blue, chain B).*

Supplementary Chap. 5

S5.1 Structure ensemble $\mathbf{E}_{\mathbf{kpi}}^{\mathbf{DGR}}$

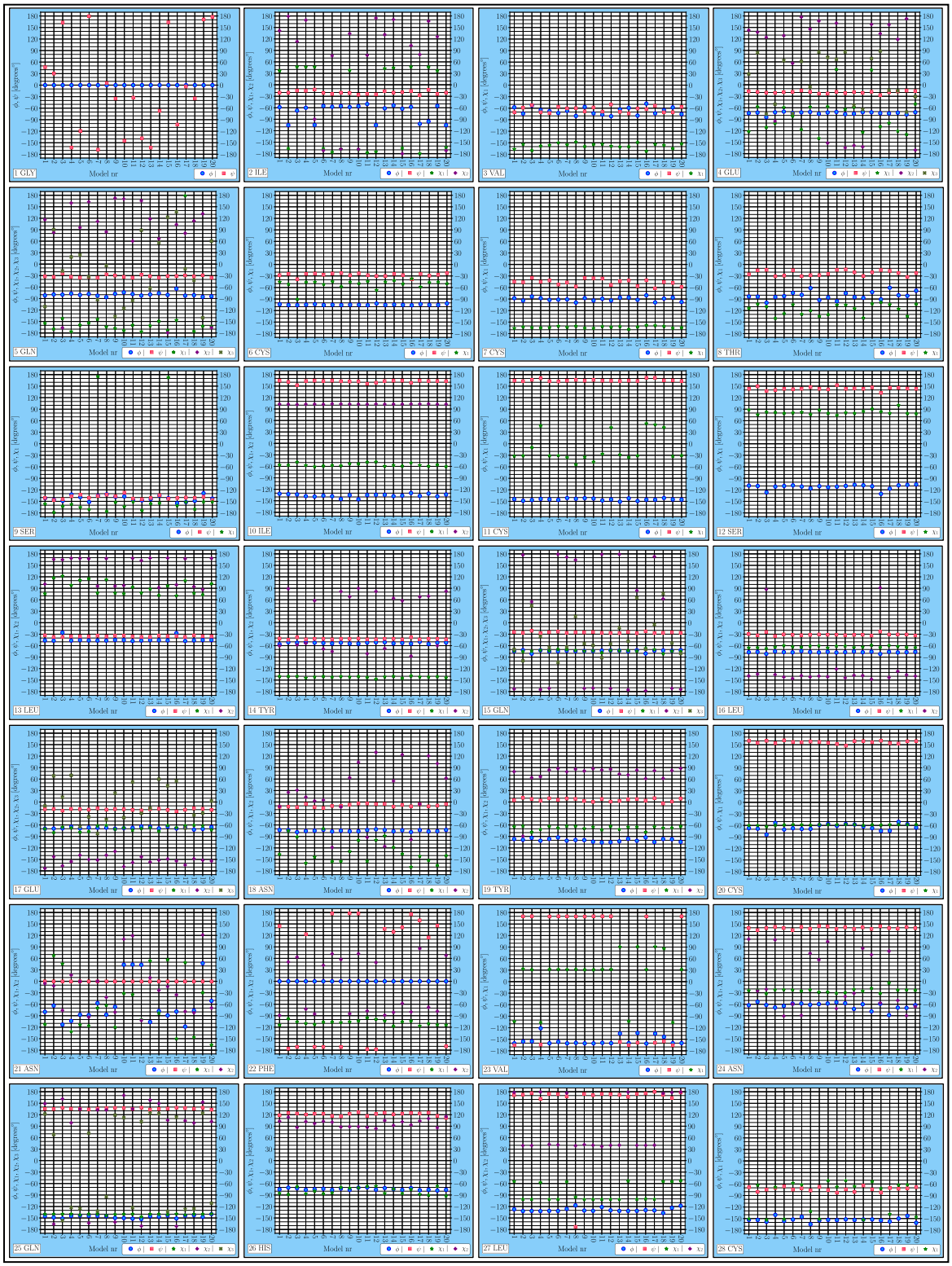


***Figure S5.27, 1st page**: The variation of DAs of $\boldsymbol{E_{kpi}^{DGR}}$, i.e. in the 20 models, for each of the 51 residues.*

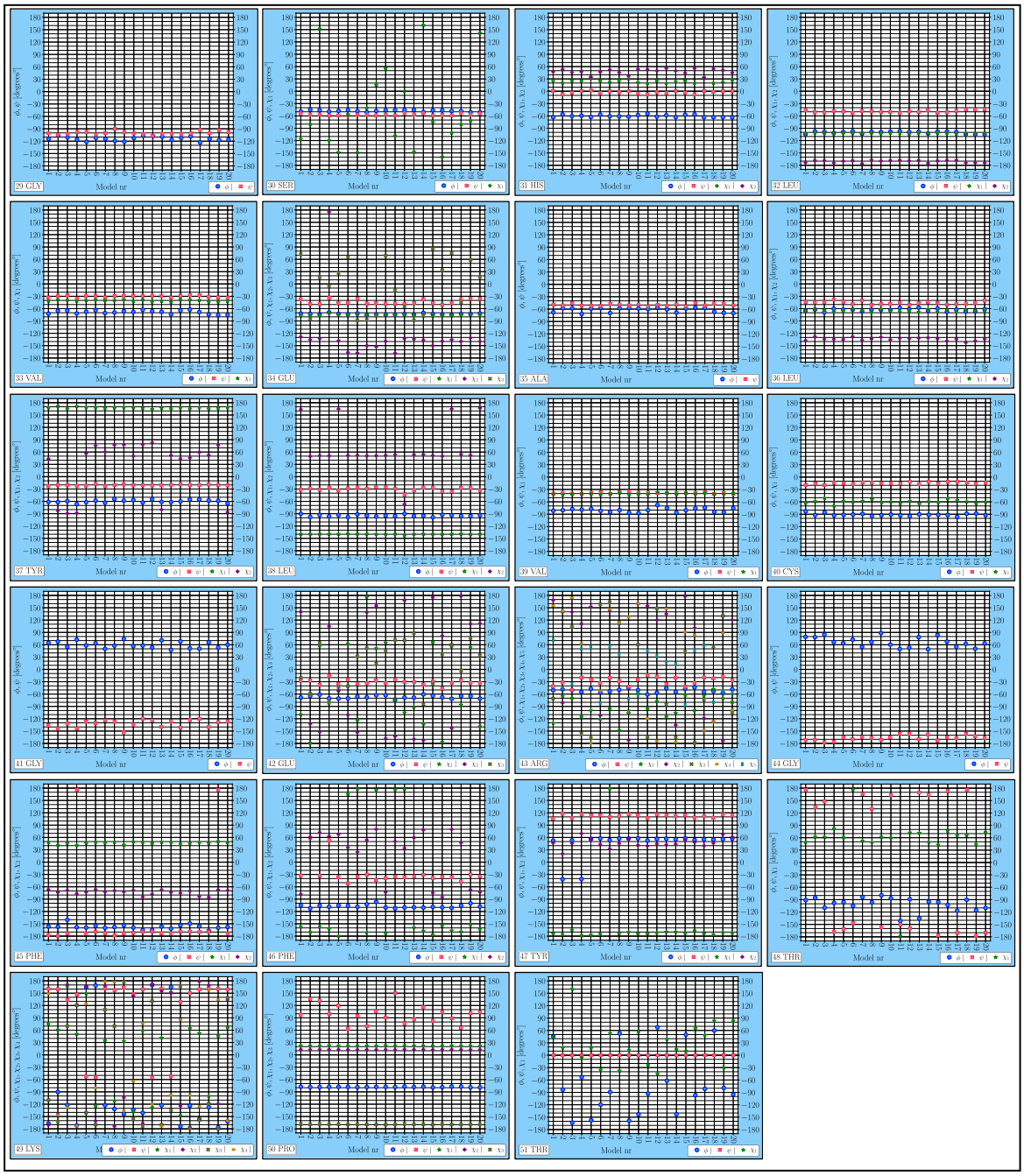


Figure S5.27, ***2nd page***.

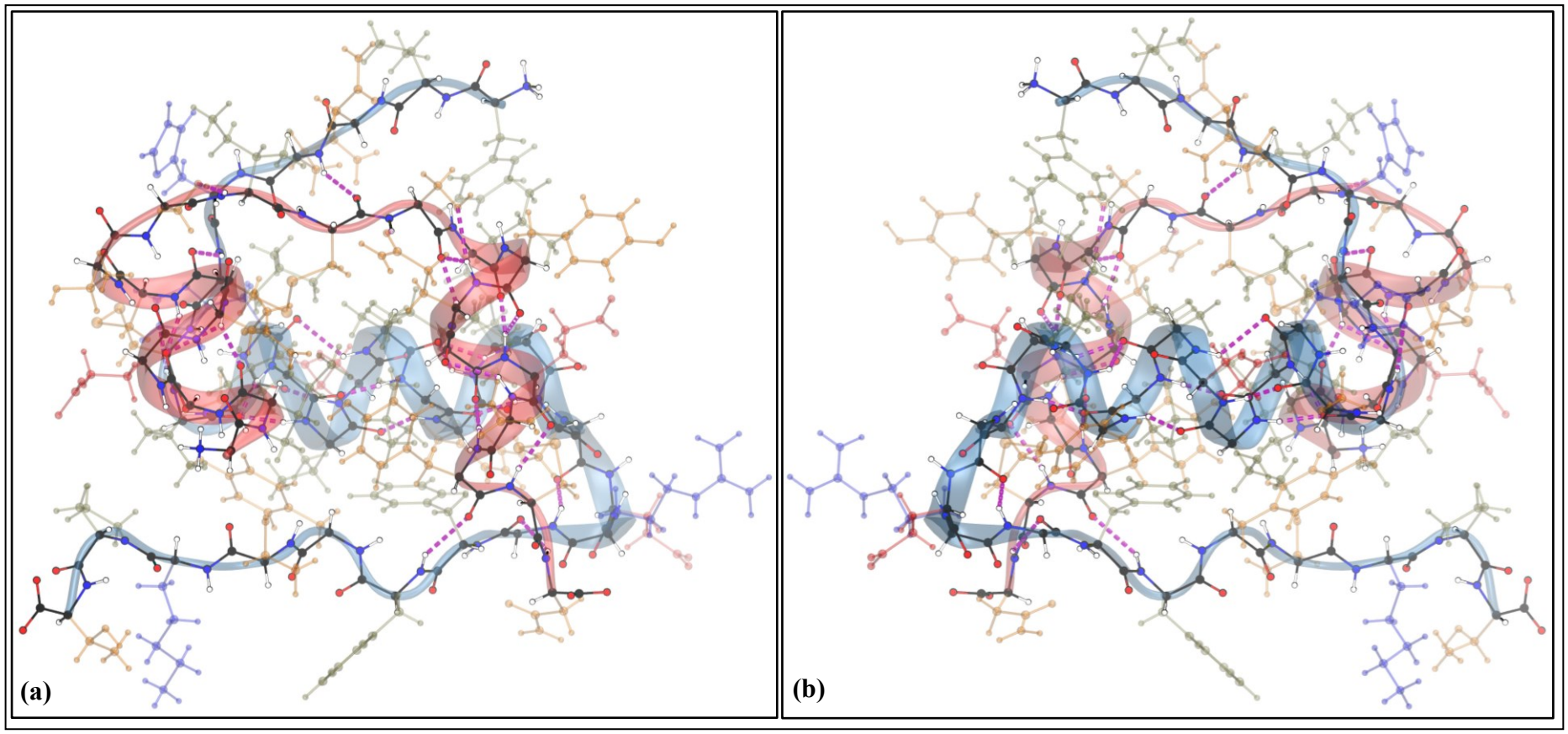


Figure S5.28: *Structure and medium-angle HBs for* $\boldsymbol{E_{kpi}^{DGR}}$ *.* ***(a)*** *Front,* ***(b)*** *back, (front rotated sideways 180°). The structure is the MS of structure ensemble, with the AC, BC and SCs shown as transparent, with the MC being chalky and in ordinary atom colouring (SC colouring and HBs is alike to that of Figure 5.3). The 28 HBs are in dotted purple, existing within the criteria "*$|\boldsymbol{r}_{AD}| < 3.5$ Å *&* $\varphi < 60°$*", and present for at least 75% (i.e. in 15 of the 20 models).*

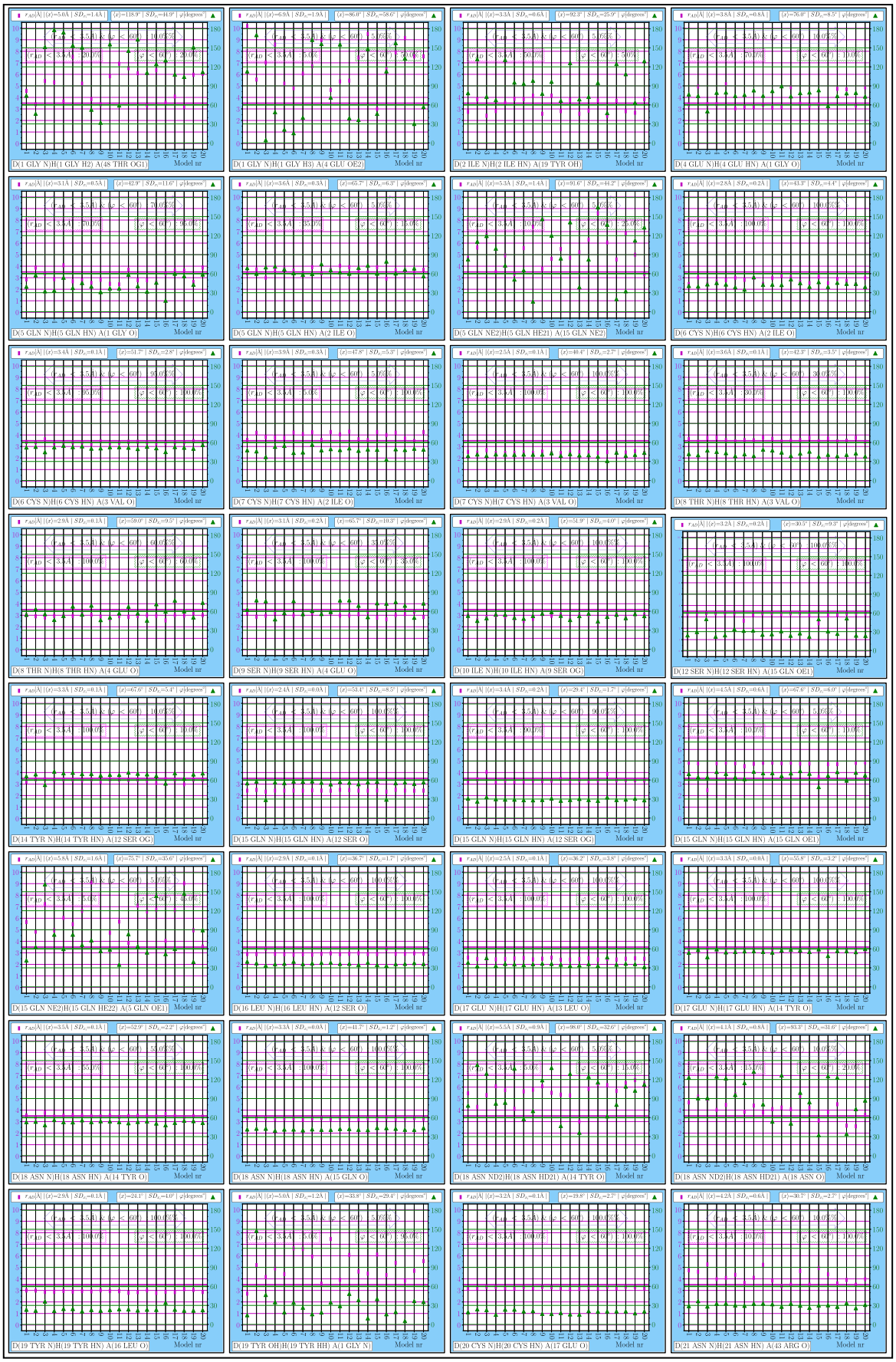


***Figure S5.29, 1st page**: Some 65 HBs of $\boldsymbol{E_{kpi}^{DGR}}$, present for at least 1 in 20 structures (where medium angle is; $\varphi < 60°$). The duplicate "%%" in fulfilling both criterias is meant to be only one "%" sign.*

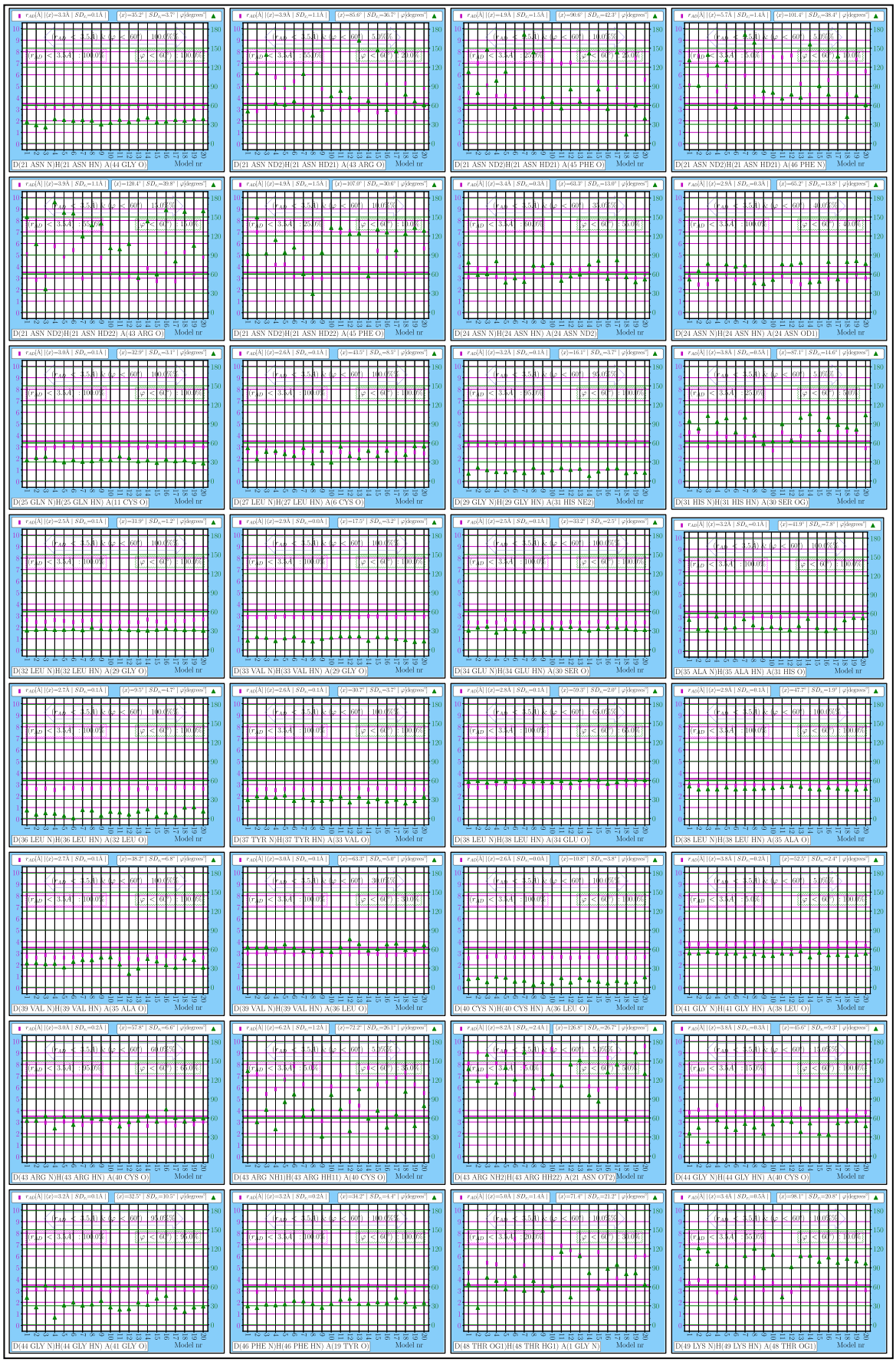


Figure S5.29, ***2nd page****: Omitting the 65'th* HB *($T_{HN}^{B30} \rightarrow K_{O}^{B29}$) present for 25% of the 20 models.*

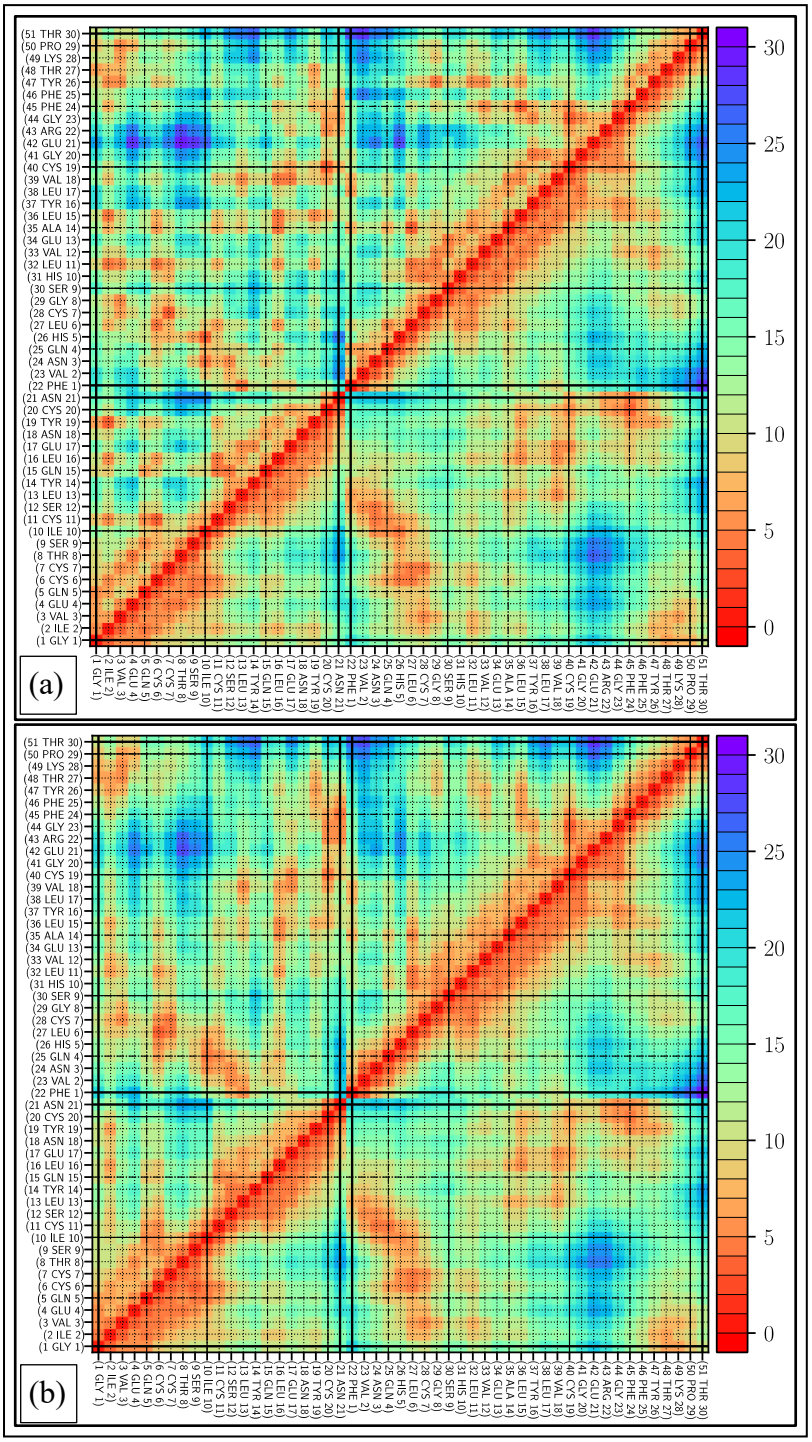


***Figure S5.30**: Residue-moiety average distances within 30 Å, matrix, of* $\boldsymbol{E_{kpi}^{DGR}}$.

Diagonal as reference 0 Å (red), above 30.0 Å in purple.

***(a)** Upper left is SC to SC geometric centre distances; lower left is CA to CA-atom distances.*

***(b)** Upper left is SC to MC geometric centre distances; lower left is MC to MC-atom distances.*

Largest distance is 29.73 Å between SCs of A8 and B21(42).

Distances divided in 1.0 steps, c.f. most of the CA-atom distances of adjacent residues have between 3-4 Å.

Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number AC (1-21), BC (1-30). Grid-lines in black, are chain coloured, distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Zoomable vector graphics.

.

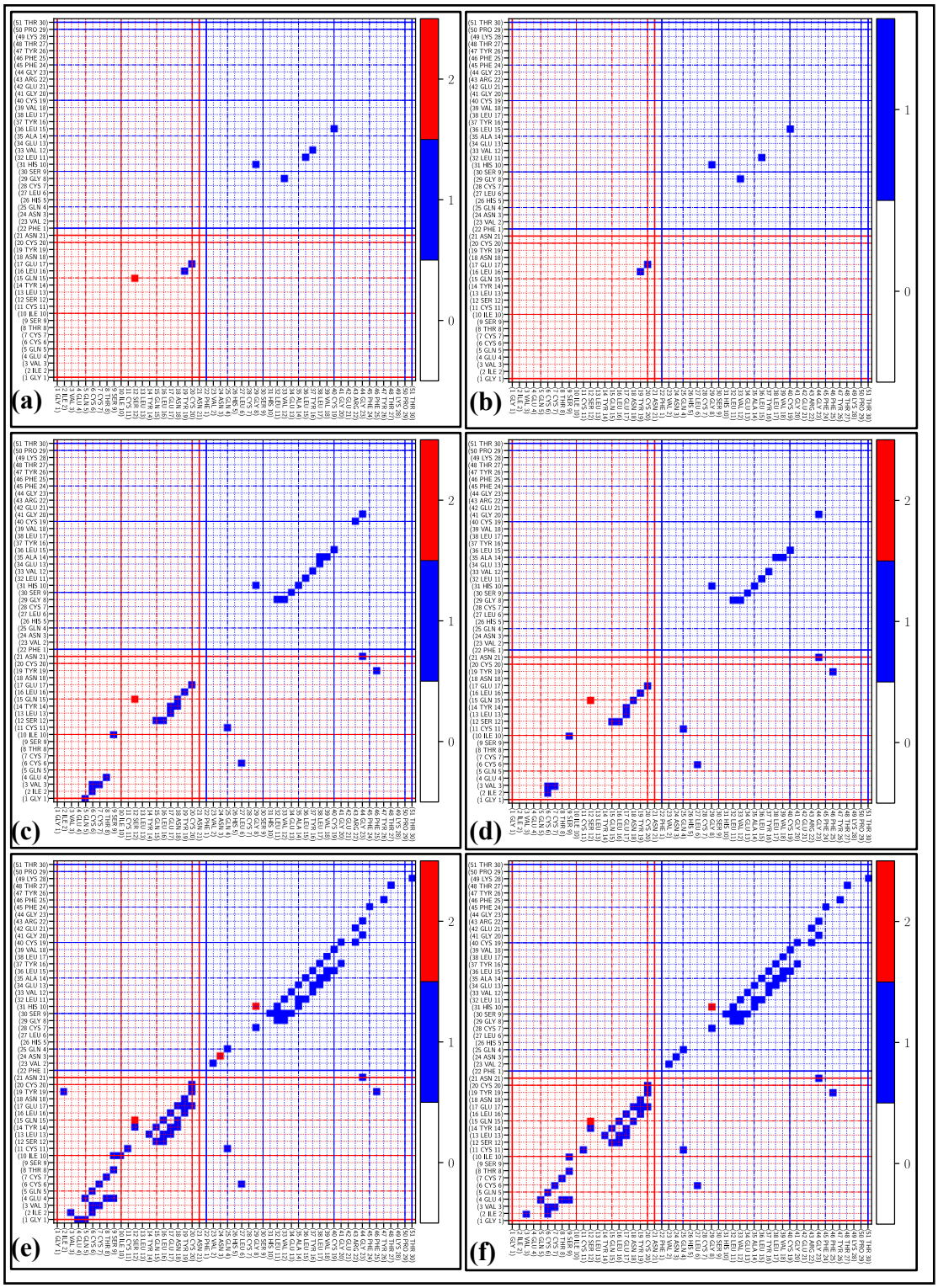


***Figure S6.31**: Sorted HBs between residues in the structure ensemble $\boldsymbol{E_{kpi}^{DGR}}$. For a presence larger than a certain percentage of structures and angle region. The HBs were calculated with: $|\boldsymbol{r}_{AD}| <$ 3.5 Å, and $\varphi < 30°$, **(a)** 50% of structures, 9 HBs, **(b)** 75% , 6 HBs; $\varphi < 60°$, **(c)** 50% , 33 HBs, **(d)** 75%, 28 HBs; and $\varphi < 90°$, **(e)** 50%, 84 HBs, **(f)** 75%, 78 HBs. The HBs are sorted as SC to MC and SC to SC HBs in upper left, and MC to MC HBs in lower right, diagonal any HBs in same residue. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graphs has zoomable vector graphics. grid-lines for AC (1-21) in red, and BC (22-51) blue. Grid-lines distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51.*

***Table S5.3**: Upper bound RHH violations by respective calculated distances of $\boldsymbol{E_{kpi}^{DGR}}$.*

Here shown is the 73 largest of 104 UB violations for restrained hydrogen distances (RHHs) from PDB 2KJJ (where only 793 RHHs were included since omitting GLY HA# out of the 803).

RHH V_{ij} [Å]					
B24 H(HD1)	B15 L(HD21)	1.23	A5 Q(HN)	A3 V(HG11)	0.36
B26 Y(HE1)	B12 V(HG21)	1.05	B19 C(HN)	B18 V(HG11)	0.36
B12 V(HA)	B11 L(HD21)	1.00	B24 F(HE1)	B12 V(HG21)	0.36
B24 F(HE1)	B12 V(HG11)	0.98	B1 F(HZ)	B17 L(HB1)	0.34
B13 E(HA)	B12 V(HG21)	0.97	B24 F(HZ)	B15 L(HD11)	0.34
B24 F(HD1)	B24 F(HB1)	0.93	B11 L(HA)	B14 A(HB1)	0.31
B24 F(HE1)	B15 L(HD21)	0.90	A21 N(HD21)	B23 G(HA1)	0.31
B13 E(HN)	B14 A(HB1)	0.89	B13 E(HG1)	B12 V(HG11)	0.30
A12 S(HN)	B3 N(HB1)	0.85	A14 Y(HN)	A13 L(HB1)	0.29
B1 F(HB1)	B2 V(HG21)	0.76	B4 Q(HB1)	B6 L(HB1)	0.29
A6 C(HB1)	B6 L(HD21)	0.75	B5 H(HE1)	A10 I(HG11)	0.29
B19 C(HN)	B15 L(HD11)	0.73	B14 A(HN)	B12 V(HG11)	0.29
B10 H(HD2)	B14 A(HB1)	0.71	B13 E(HN)	B12 V(HG11)	0.26
B19 C(HN)	B15 L(HD21)	0.70	B18 V(HB)	A16 L(HD21)	0.26
A7 C(HB1)	A8 T(HG21)	0.69	B17 L(HN)	B15 L(HD11)	0.24
B26 Y(HB1)	B27 T(HG21)	0.69	A10 I(HA)	A10 I(HD1)	0.24
B14 A(HB1)	B11 L(HD11)	0.66	B26 Y(HN)	B15 L(HD11)	0.23
B3 N(HB1)	A10 I(HD1)	0.63	B9 S(HN)	B10 H(HB1)	0.22
B10 H(HD2)	B12 V(HG21)	0.59	A12 S(HN)	A15 Q(HG1)	0.21
B27 T(HN)	B25 F(HB1)	0.58	B11 L(HN)	B12 V(HB)	0.19
A19 Y(HE1)	B15 L(HB1)	0.56	B11 L(HN)	B13 E(HA)	0.18
B26 Y(HN)	B24 F(HD1)	0.56	A7 C(HN)	A3 V(HG11)	0.17
A7 C(HN)	A3 V(HG21)	0.51	B18 V(HA)	A13 L(HD21)	0.17
A7 C(HB1)	A4 E(HB1)	0.51	A7 C(HN)	A9 S(HB1)	0.14
B12 V(HA)	B11 L(HD11)	0.49	B1 F(HB1)	B2 V(HG11)	0.13
A19 Y(HE1)	B27 T(HB)	0.48	B25 F(HE1)	B27 T(HG21)	0.12
B17 L(HN)	B16 Y(HB1)	0.47	B27 T(HN)	B26 Y(HB1)	0.12
B24 F(HE1)	B15 L(HD11)	0.47	A17 E(HN)	A17 E(HG1)	0.11
B4 Q(HN)	A10 I(HD1)	0.46	B8 G(HN)	B7 C(HA)	0.10
B24 F(HA)	B15 L(HD11)	0.44	B10 H(HD2)	B11 L(HD11)	0.10
B24 F(HD1)	B15 L(HD11)	0.41	B16 Y(HN)	B12 V(HA)	0.10
B24 F(HZ)	B12 V(HG21)	0.38	B26 Y(HN)	B25 F(HB1)	0.10
B15 L(HN)	B12 V(HG11)	0.38	B3 N(HA)	B2 V(HG11)	0.09
A7 C(HA)	A3 V(HG11)	0.37	A16 L(HN)	A18 N(HN)	0.09
B7 C(HN)	B6 L(HD21)	0.37	A21 N(HN)	B22 R(HB1)	0.09
B10 H(HD2)	B12 V(HG11)	0.37	B3 N(HA)	B4 Q(HG1)	0.08
B18 V(HN)	B15 L(HD11)	0.37	⋮	⋮	⋮

Supplementary Chap. 6

For chapter 6, each MD simulation of an insulin system, few replicas are emphasized in the main text. For seeing some difference to other replicas, they are provided in the supplementary, with indicated system identity and explanation for figures. These results from other replicas are not included directly, since it would obscure the main text. However, differences are explained in the main text of respective result chapter, with referral to these figures. The replicas are included, since they provide information about differing structural behaviours in the ensemble. Moreover, statistical variance of the MD simulation method used.

S6.1 Overview Trajectory Ensembles $\mathbf{P_{kpi}^{MD}}$, $\mathbf{E_{kpi}^{MD}}$, $\mathbf{P_{i}^{MD}}$

***Table S6.4**: Lower angle intra-monomer HBs of MD ensembles compared to observed HBs .Where 'RHB' being the restrained HB's of Table 5.2, and **M1**, **M2** the x-ray structures of Table 4.1, and all HBs compared for the lower angles ($\varphi < 30°$). Note MD replicas compares its HB presence at indicated percentage to RHB and **M1**, **M2** whose HBs are of presence above 0%, except when comparing between replicas where the same percentage applies.*

System	Nr of HBs AC_BC_AC&BC		Nr HBs = RHB (15_11_5)		Nr HBs = **M1** (8_11_6)		Nr HBs = **M2** (11_8_7)	
Nr HBs	>= 5%	>= 25%	-||-	-||-	-||-	-||-	-||-	-||-
> 0%	>= 50%	>= 75%	-||-	-||-	-||-	-||-	-||-	-||-
MS	= 100 %		-||-		-||-		-||-	
$\mathbf{P_{kpi}^{MD}}$ m	26_15_10	12_9_7	8_9_4	6_7_3	7_9_5	4_9_3	10_8_5	7_8_4
358	5_8_4	1_5_1	2_7_3	0_5_1	2_8_3	1_5_1	3_8_4	1_5_1
MS	9_10_5		4_8_3		3_9_3		5_8_3	
n	27_14_11	9_10_8	7_8_5	4_7_4	6_9_5	2_9_4	9_8_6	5_8_5
380	5_8_3	0_5_1	2_7_3	0_5_1	2_8_3	0_5_1	3_8_3	0_5_1
MS	8_8_4		3_5_2		3_7_2		5_6_3	
o	27_16_11	11_11_6	7_8_4	4_8_2	6_9_4	2_9_2	9_8_5	5_8_3
340	5_8_2	1_5_1	2_7_2	0_5_1	2_8_2	1_5_1	3_8_2	1_5_1
MS	12_8_2		4_7_0		5_6_0		7_6_1	
m=n=o	24_12_8	9_9_6	7_8_4	4_7_2	6_9_4	2_9_2	9_8_5	5_8_3
223	5_8_2	0_5_1	2_7_2	0_5_1	2_8_2	0_5_1	3_8_2	0_5_1
MS	4_4_2		1_4_0		2_4_0		3_4_1	
$\mathbf{E_{kpi}^{MD}}$ m	28_19_12	11_9_7	7_9_5	5_7_3	6_9_6	3_9_3	9_8_6	6_8_4
276	5_8_2	2_5_2	2_7_2	1_5_2	2_8_2	1_5_2	3_8_2	1_5_2
MS	10_10_5		3_6_3		2_7_3		4_6_4	
n	28_13_13	10_10_7	7_8_4	4_7_3	6_9_4	2_9_3	9_8_5	5_8_4
329	5_8_1	1_5_1	2_7_1	0_5_1	2_8_1	1_5_1	3_8_1	1_5_1
MS	10_8_5		4_6_3		2_7_3		6_6_4	
o	27_16_10	10_10_7	8_8_4	4_7_2	7_9_4	2_9_2	10_8_5	5_8_3
333	6_8_1	1_5_1	3_7_1	0_5_1	2_8_1	1_5_1	3_7_1	1_5_1
MS	12_5_3		6_4_1		5_4_1		7_4_1	
m=n=o	25_12_9	10_9_6	7_8_4	4_7_2	6_9_4	2_9_2	9_8_5	5_8_3
179	5_7_1	1_5_1	2_7_1	0_5_1	2_7_1	1_5_1	3_7_1	1_5_1
MS	5_3_2		2_3_1		1_3_1		2_3_1	
$\mathbf{P_{i}^{MD}}$ m	28_16_12	12_10_1	5_7_0	4_7_0	4_8_0	2_8_0	7_7_1	5_7_0
448	7_9_0	2_1_0	3_6_0	1_1_0	2_7_0	1_1_0	3_6_0	1_1_0
MS	8_6_1		3_3_0		1_4_0		2_3_0	
n	28_19_7	11_10_4	7_8_2	3_7_1	6_8_2	2_8_1	9_7_2	4_7_1
350	6_7_1	2_5_0	2_6_1	1_5_0	2_6_1	1_5_0	3_6_1	1_5_0
MS	11_7_2		3_6_1		3_7_1		4_6_1	
o	27_15_15	9_10_9	6_8_4	4_7_3	6_9_4	2_9_3	8_8_5	5_8_4
364	7_8_2	1_5_0	3_7_2	1_5_0	2_8_2	0_5_0	3_7_2	0_5_0
MS	7_10_5		2_7_1		2_9_1		3_8_2	
n=o	24_11_4	8_9_1	6_7_2	3_7_1	5_8_2	2_8_1	8_7_2	4_7_1
261	6_6_1	1_5_0	2_6_1	1_5_0	2_6_1	0_5_0	3_6_1	0_5_0
MS	4_7_0		1_6_0		1_7_0		2_6_0	

S6.2 Trajectory Ensemble P_{kpi}^{MDm}

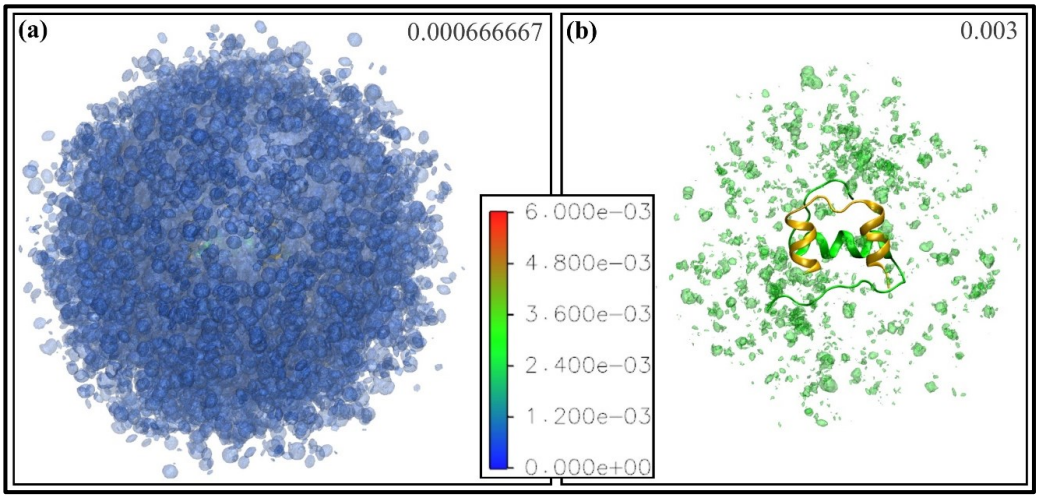


***Figure S6.32**: Chloride fractional occupancy for P_{kpi}^{MDm}. **(a)** That of 1/1500 (0.06667%) (in transparent grey are a sphere of radius half sidelength of the analysis square box), **(b)** 4.5/1500 (0.3%). Isovalues that is one hundredth of the percentage shown. Meaning that any chloride ion is found very rarely in any of the 0.2 Å sidelength cubes, that the box is divided in for calculating fractional occupancy.*

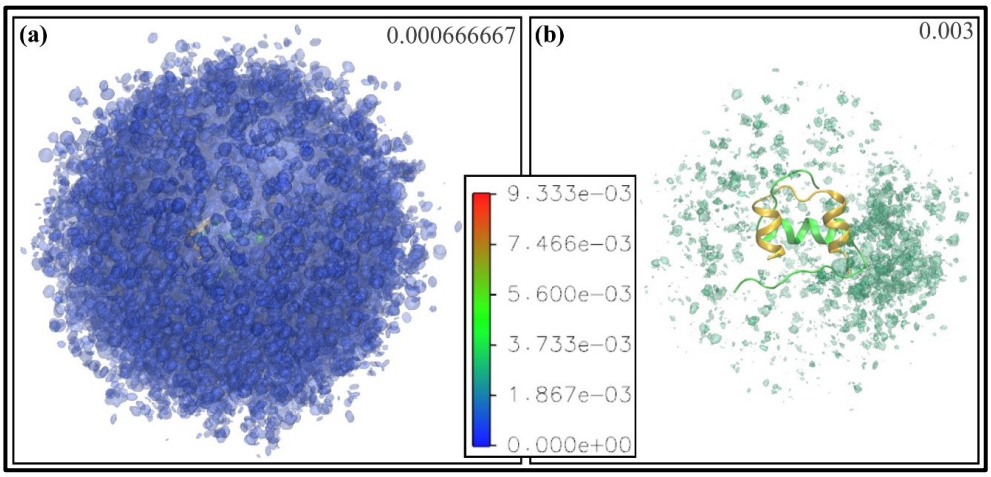


***Figure S6.33**: Sodium fractional occupancy for P_{kpi}^{MDm} **(a)** That of 1/1500 (0.06667%) (in transparent grey are a sphere of radius half sidelength of the analysis square box), **(b)** 4.5/1500 (0.3%). Isovalues that is one hundredth of the percentage shown. Isovalues that is one hundredth of the percentage shown. Meaning that any sodium ion is found very rarely in any of the 0.2 Å sidelength cubes, that the box is divided in for calculating fractional occupancy, however a slight preference at the charged residues around the B-chain turn.*

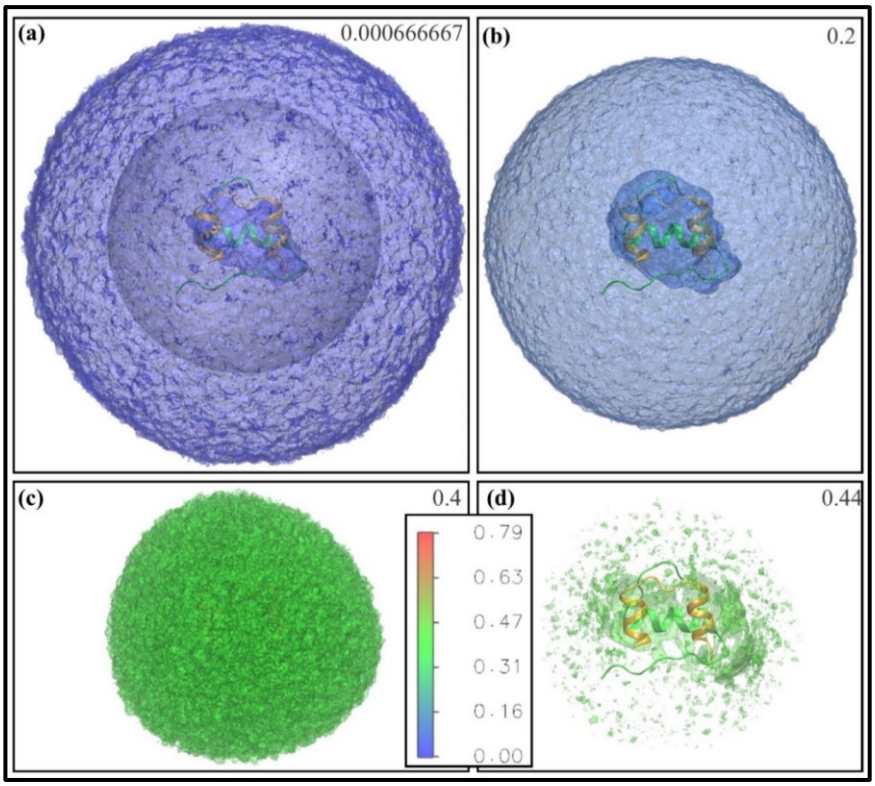


***Figure S6.34**: Water fractional occupancy for $\boldsymbol{P_{kpi}^{MDm}}$ **(a)** That occupied at least by 1/1500 (0.06667%) (in transparent grey are a sphere of radius half sidelength of the square analysis-box), **(b)** 300/1500 (20%), **(c)** 600/1500 (40%), **(d)** 660/1500 (44%). Isovalues shown are thus one hundredth of the percentage. Meaning that in the analysis-box water-atoms is found often and mostly within the analysis-box sphere, encompassing any of the 0.2 Å sidelength cubes, that space are divided in for calculating fractional occupancy. Showing there is a slight preference of water to the superimposed region (here CA-atoms in B11-B17).*

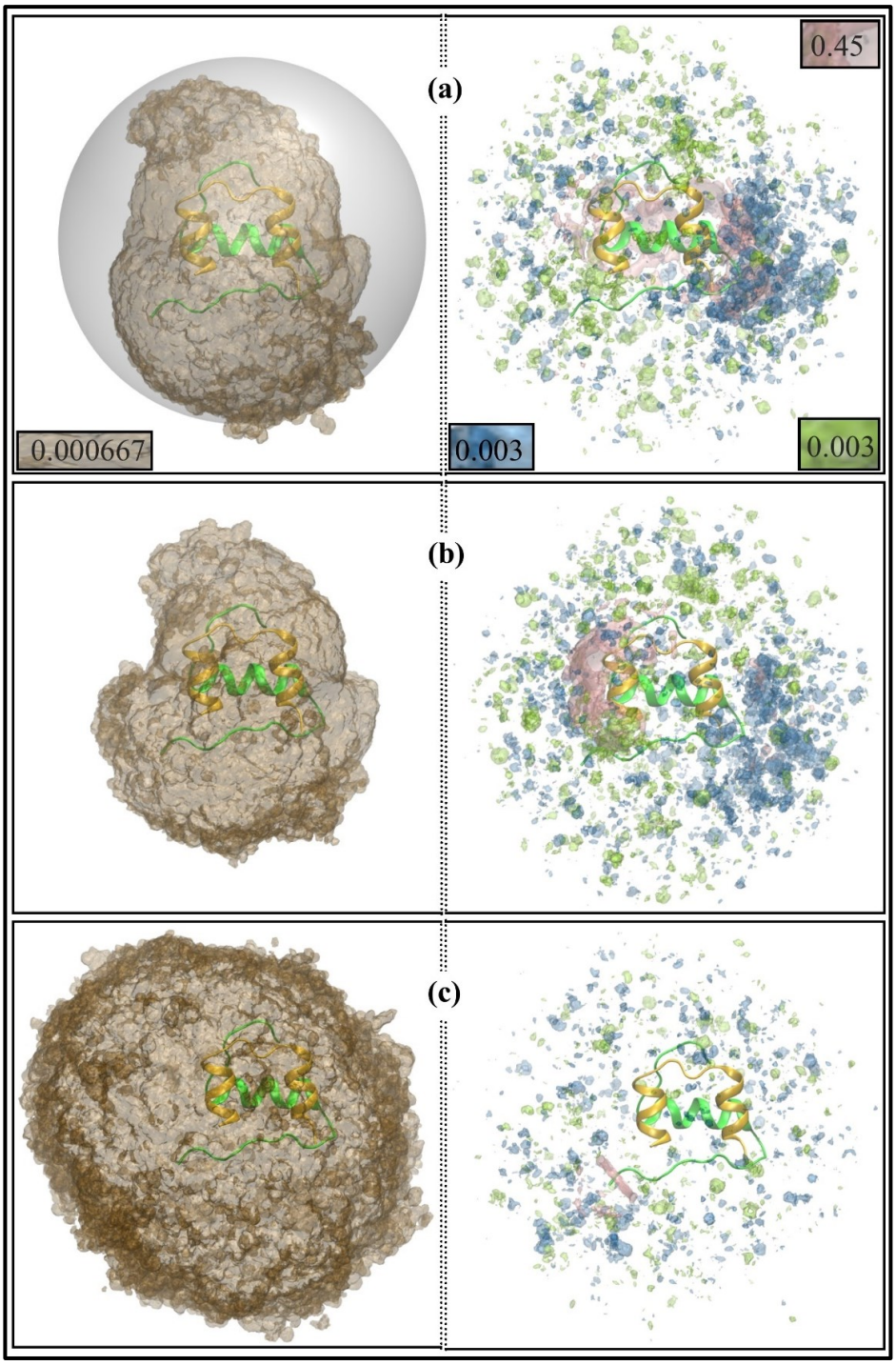


***Figure S6.35**: Superimposition effect on occupancies for ensemble $\boldsymbol{P_{kpi}^{MDm}}$. Superimposed CA-atoms of residues **(a)** B11-17 (imaginary sphere shown with radius 26.935 Å (analysis-box half sidelength) centrering at geometric centre of protein at frame 508), **(b)** A2-8, **(c)** B50-51. The motion of the protein and solutes box is relative to superimposed region in space, hence the different space covered by the calculated occupancies. Indicated isovalues for protein (brown), water (pink), sodium (blue), chloride (yellow-green). Respectively, was chosen the respective residues in (a-c) for superimposition instead of B11-17 in S3.2. However, the protein is for (a-c) and the whole trajectories, centred in the solutes filled box of sidelength 53.78 Å, as in step (3) of S3.2.*

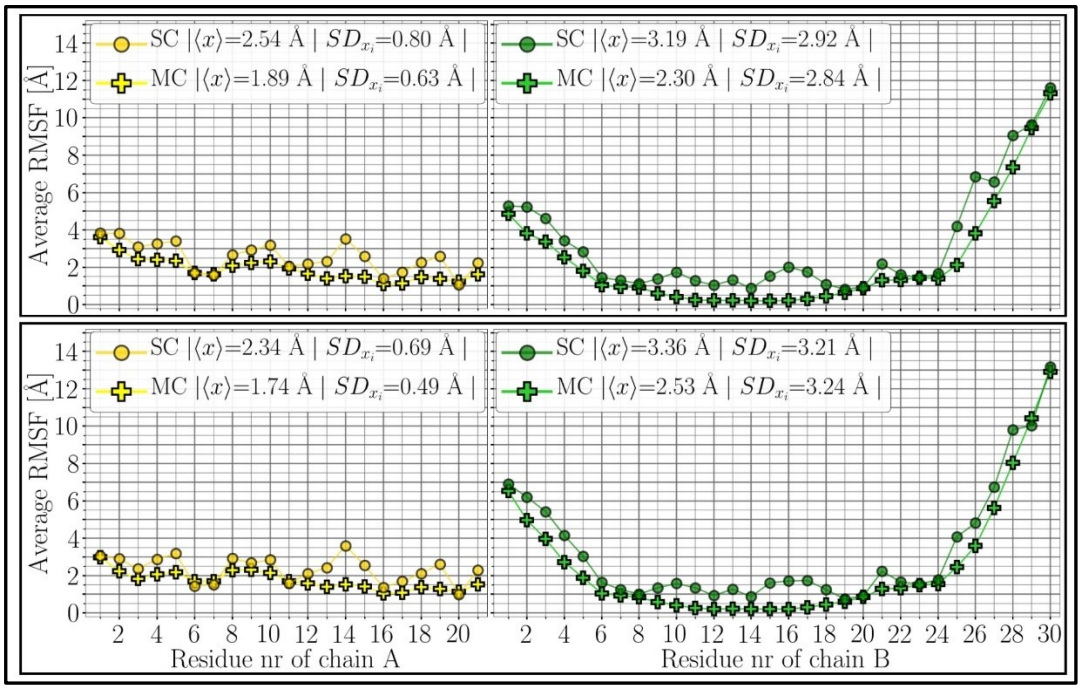


***Figure S6.36**: Average RMSF of insulin in trajectory ensemble $\boldsymbol{P_{kpi}^{MDno}}$, for all atoms of indicated selection (SC or MC) for each residue, i.e. $RMSF_{\langle SC \rangle}$ and $RMSF_{\langle MC \rangle}$ of replica "**n**" (top) and "**o**"(bottom).*

***Figure S6.37, 1st page**: The time-dependent DAs of $\mathbf{P}_{\mathbf{kpi}}^{\mathbf{MDm}}$, i.e. of MD trajectory, 0-1499 ns (plotted every 5th data point at 0,5,10,15…1495 ns) for each of the 51 residues.*

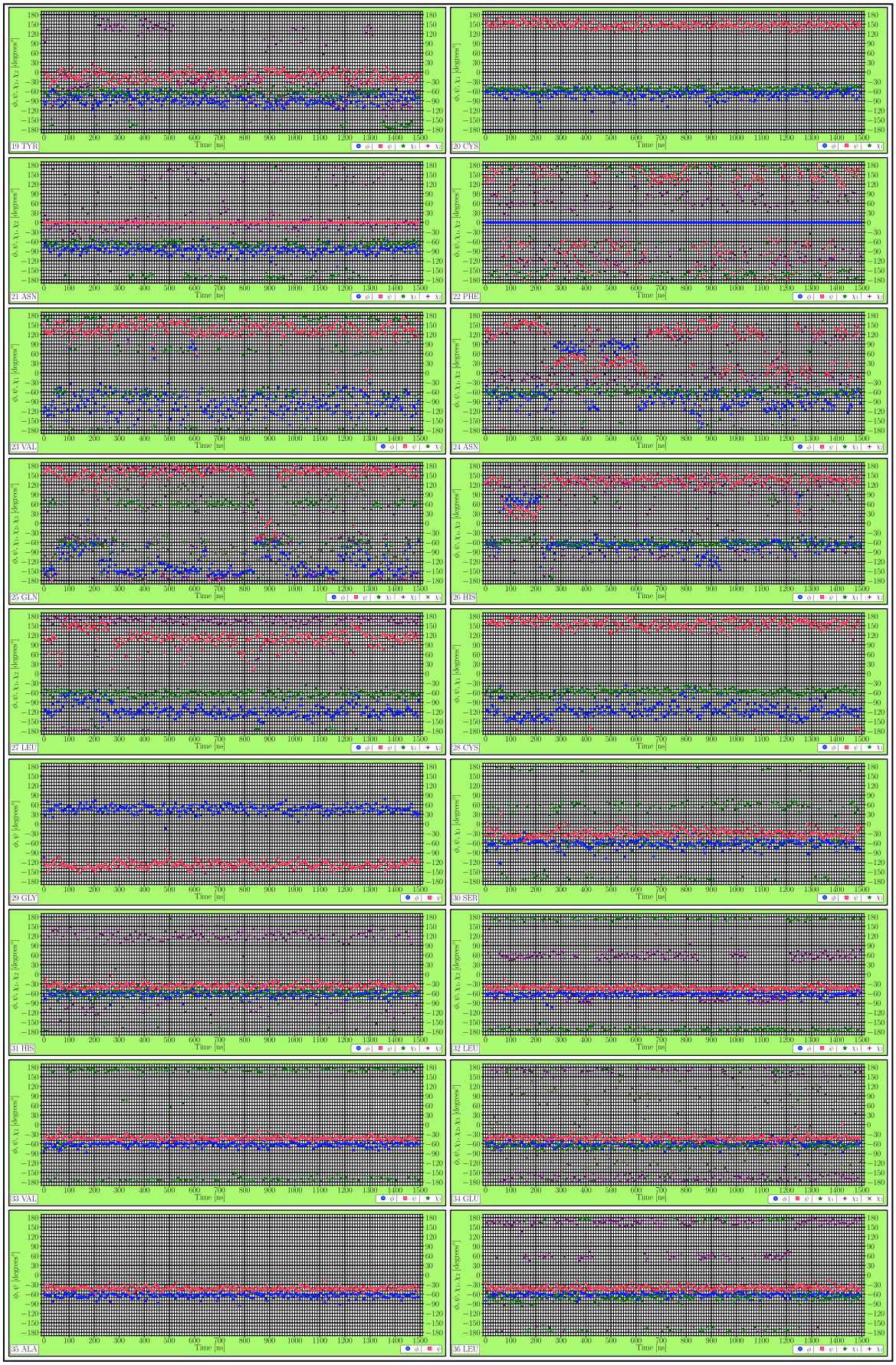


Figure S6.37, 2nd page.

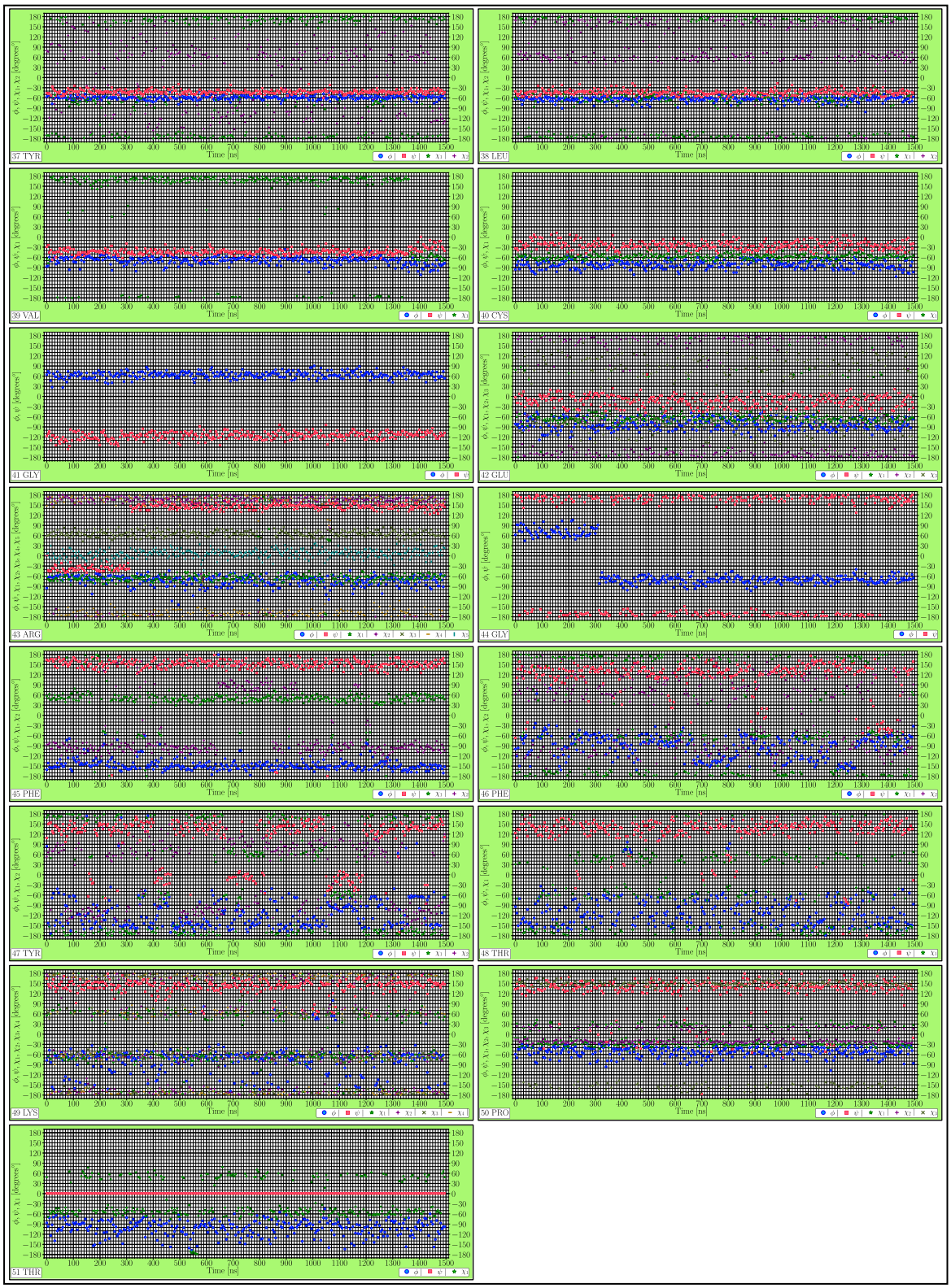


Figure S6.37, 3rd page.

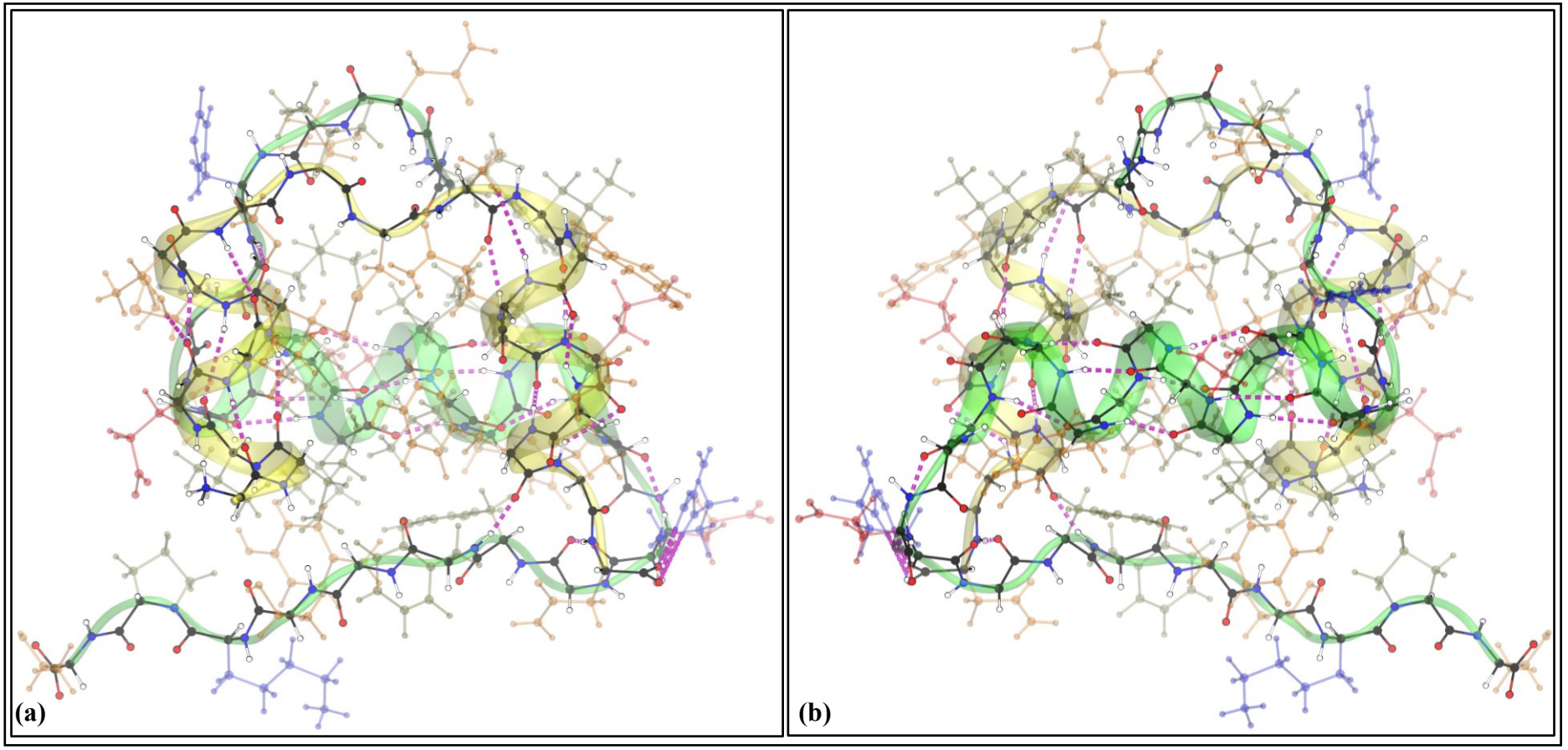


Figure S6.38*: Structure and low-angle HBs for insulin in ensemble* $\boldsymbol{P_{kpi}^{MDm}}$ *. **(a)** Front, **(b)** back (front rotated* $180°$ *sideways). The structure is the MS, with the AC and BC and SCs being transparent and BB chalky (in ordinary atom-colouring). Depicted are the 28 HBs with criteria "* $r_{AD} < 3.5$ Å *&* $\varphi < 30°$*"; present for more than 25% of 9-1499 ns; each HB (donor hydrogen to acceptor atom) are as purple dotted lines (regardless if between MCs or SCs).*

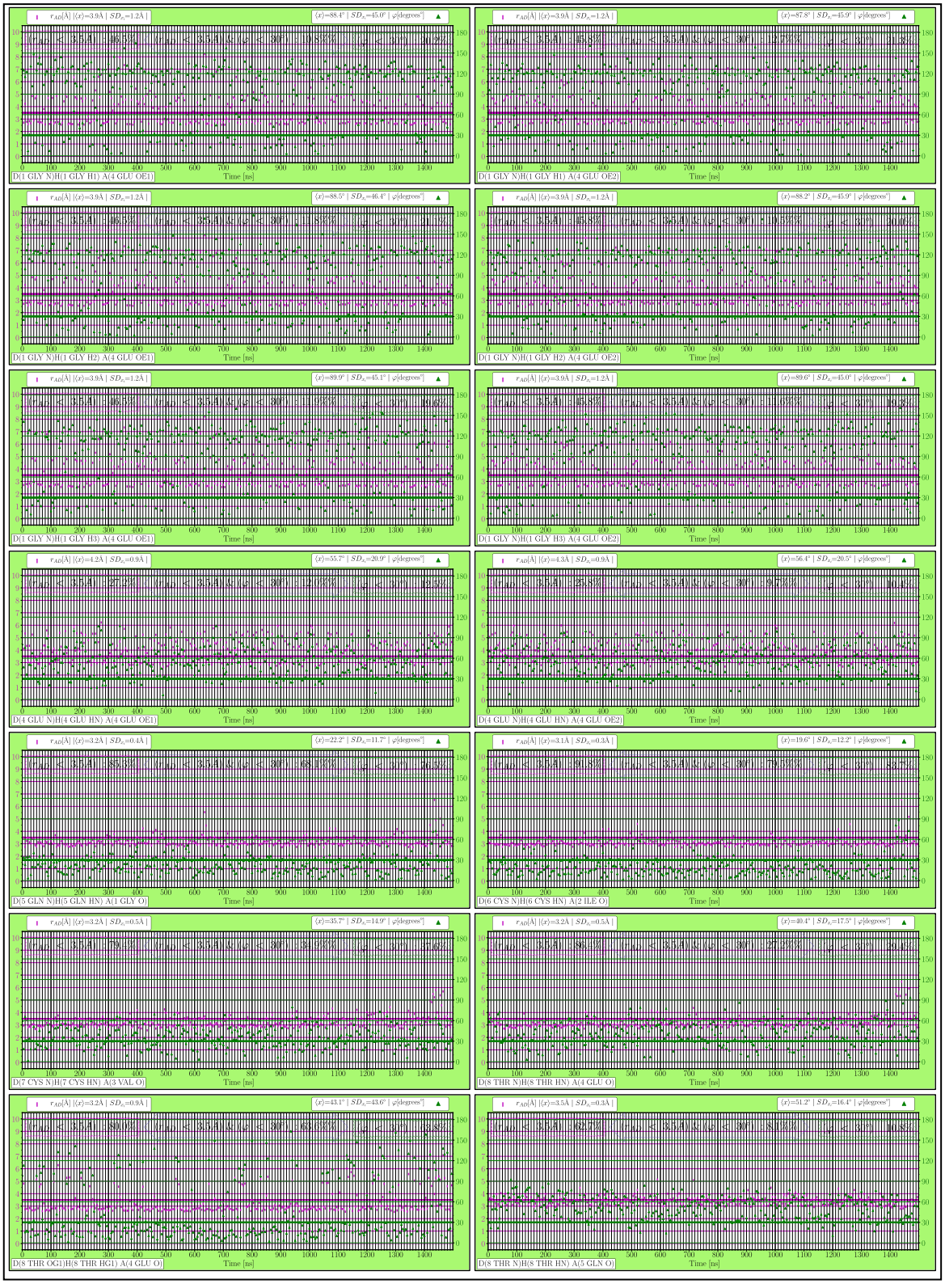


***Figure S6.39, 1st page**: Lower angle time-dependent HBs of $\boldsymbol{P_{kpi}^{MDm}}$, i.e. in MD trajectory, 0-1499 ns (plotted every 5th data point at 0,5,10,15…1495 ns), for any individual HB present for more than 5% of times 9-1499 ns. Statistics are calculated for the range 9-1499 ns. The duplicate "%%" in fulfilling both criterias is meant to be only one "%" sign.*

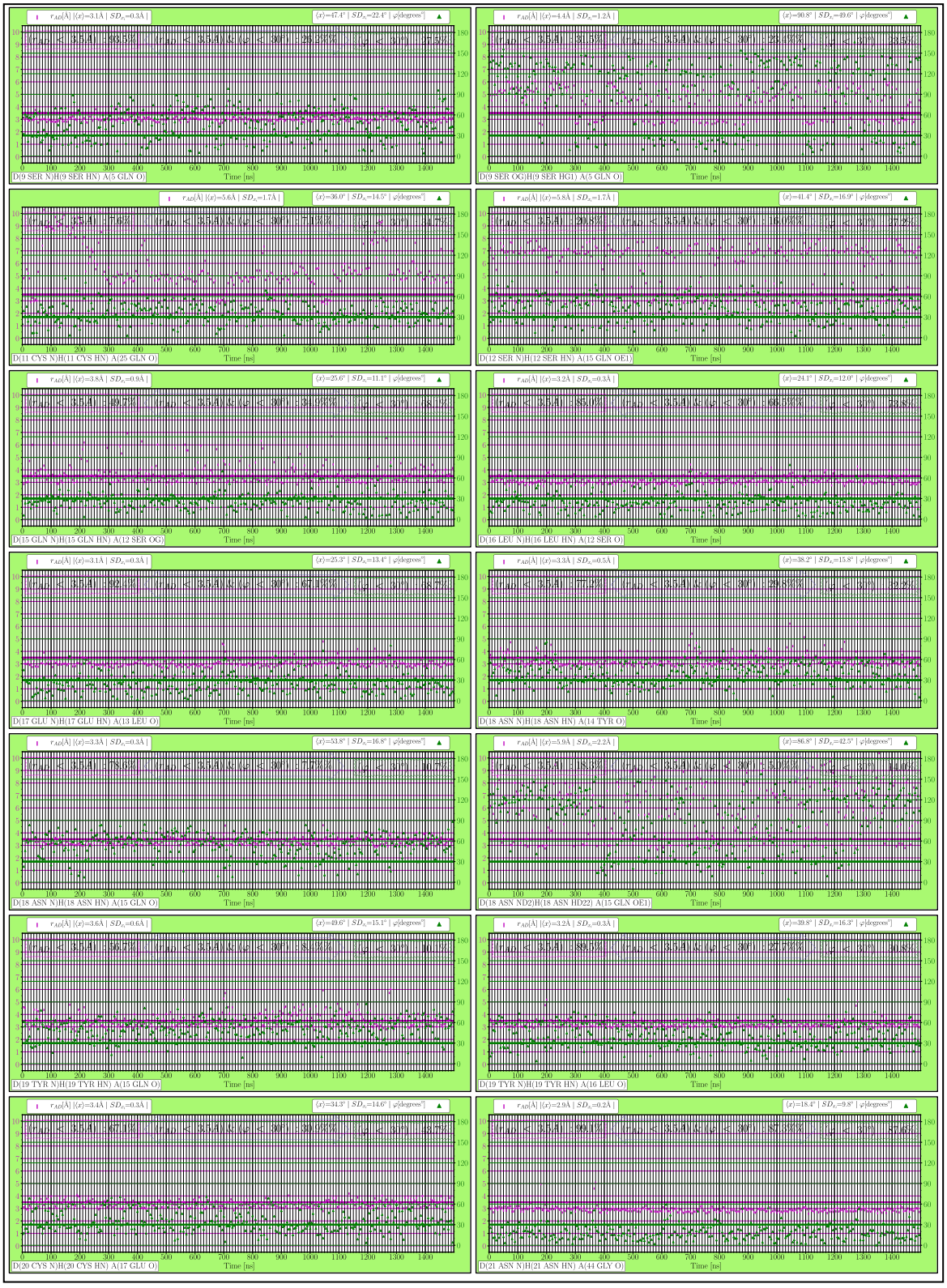


Figure S6.39, 2nd page.

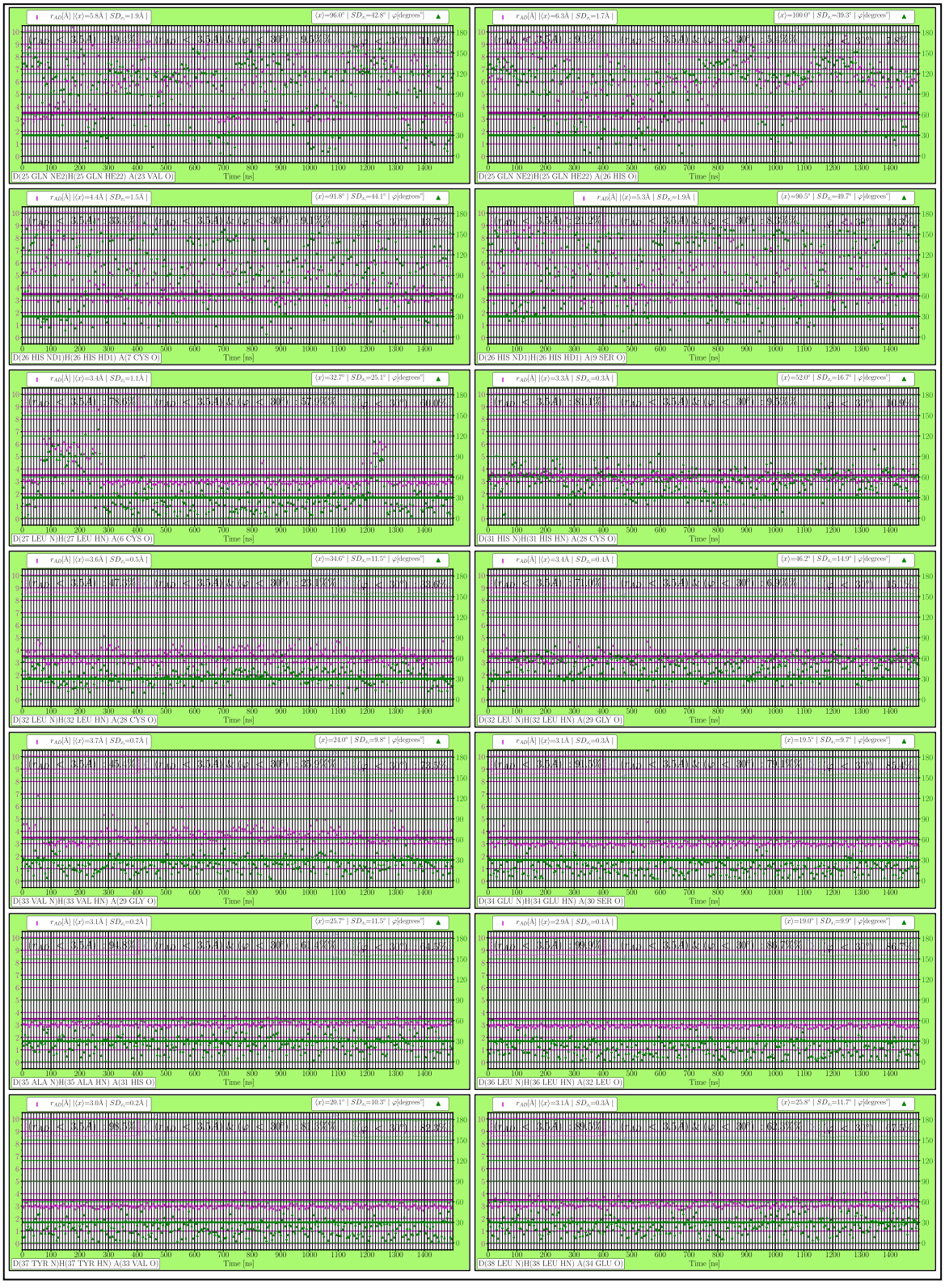


Figure S6.39, 3rd page.

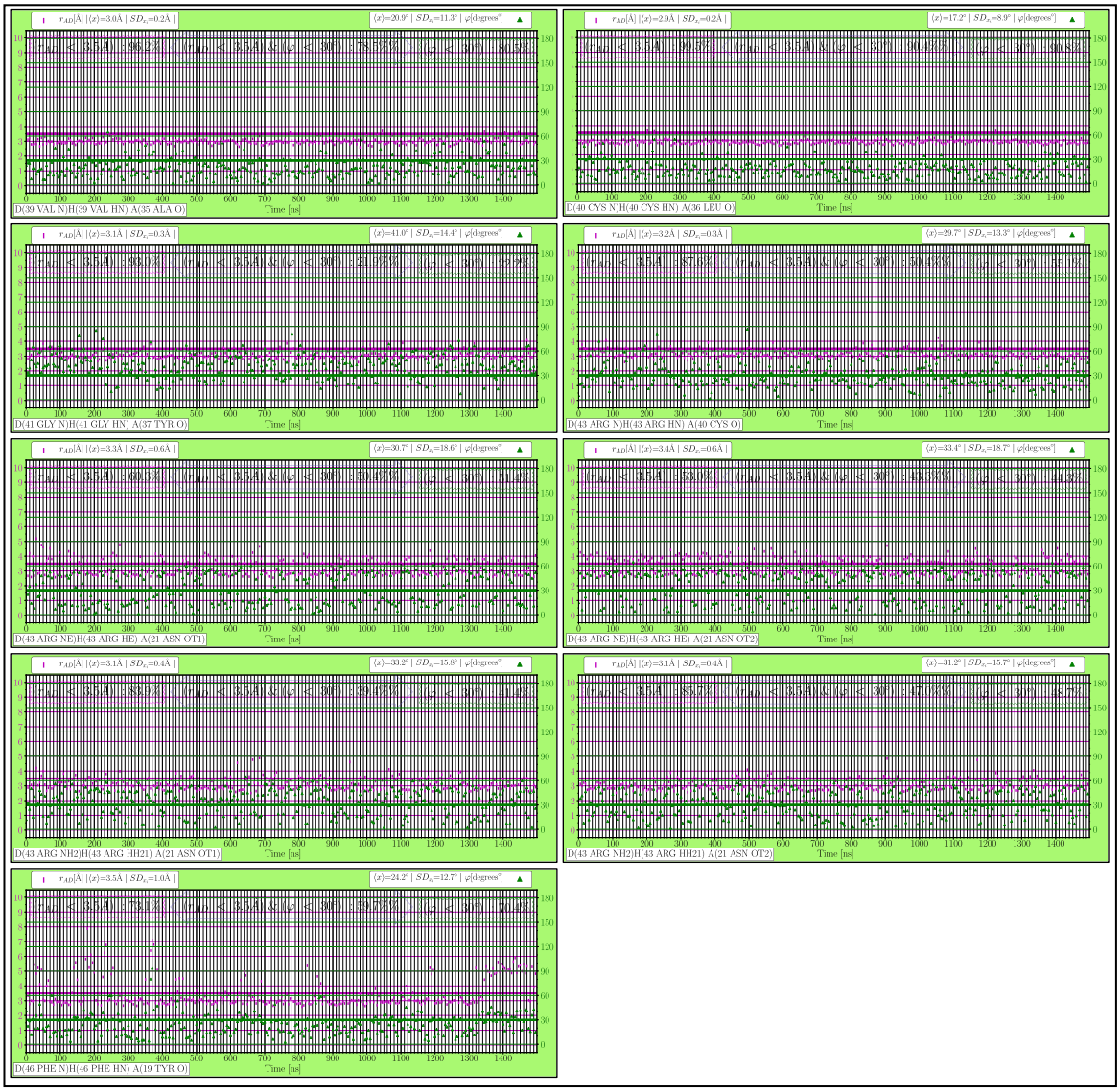


Figure S6.39, 4th page.

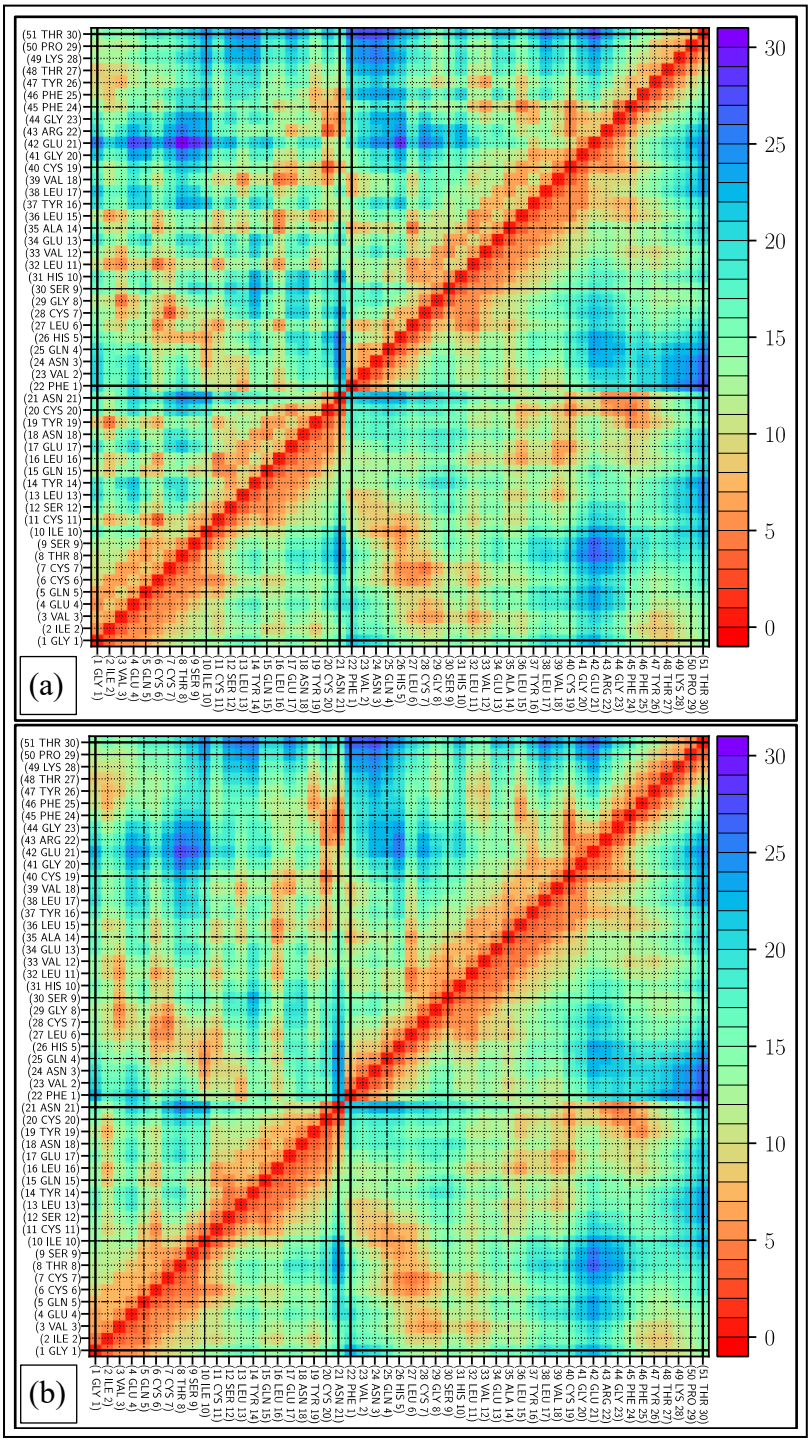


***Figure S6.40**: Residue-moiety average distances within 30 Å, matrix, of* $\boldsymbol{P_{kpi}^{MDm}}$.

Diagonal as reference 0 Å (red), above 30.0 Å in purple.

***(a)** Upper left is SC to SC geometric centre distances; lower left is CA to CA-atom distances.*

***(b)** Upper left is SC to MC geometric centre distances; lower left is MC to MC-atom distances.*

Largest distance is 30.269865 between SCs of A8 and B21(42).

Distances divided in 1.0 steps, c.f. most of the CA-atom distances of adjacent residues have between 3-4 Å.

Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number AC (1-21), BC (1-30). Grid-lines in black, are chain coloured, distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51. Zoomable vector graphics.

***Figure S6.41, 1st page**: Selection of time-dependent residue-moiety distances of $\boldsymbol{P_{kpi}^{MDm}}$, i.e. of MD trajectory, 0-1499 ns (plotted every 5th data point at 0,5,10,15…1495 ns), only a selection of the in total 1275 residue-pairs that are compared* $(= (51 + (51 - 1)/2))$*, sorted in increasing residue-number (1 to 51).* Statistics are calculated for the range 9-1499 ns.

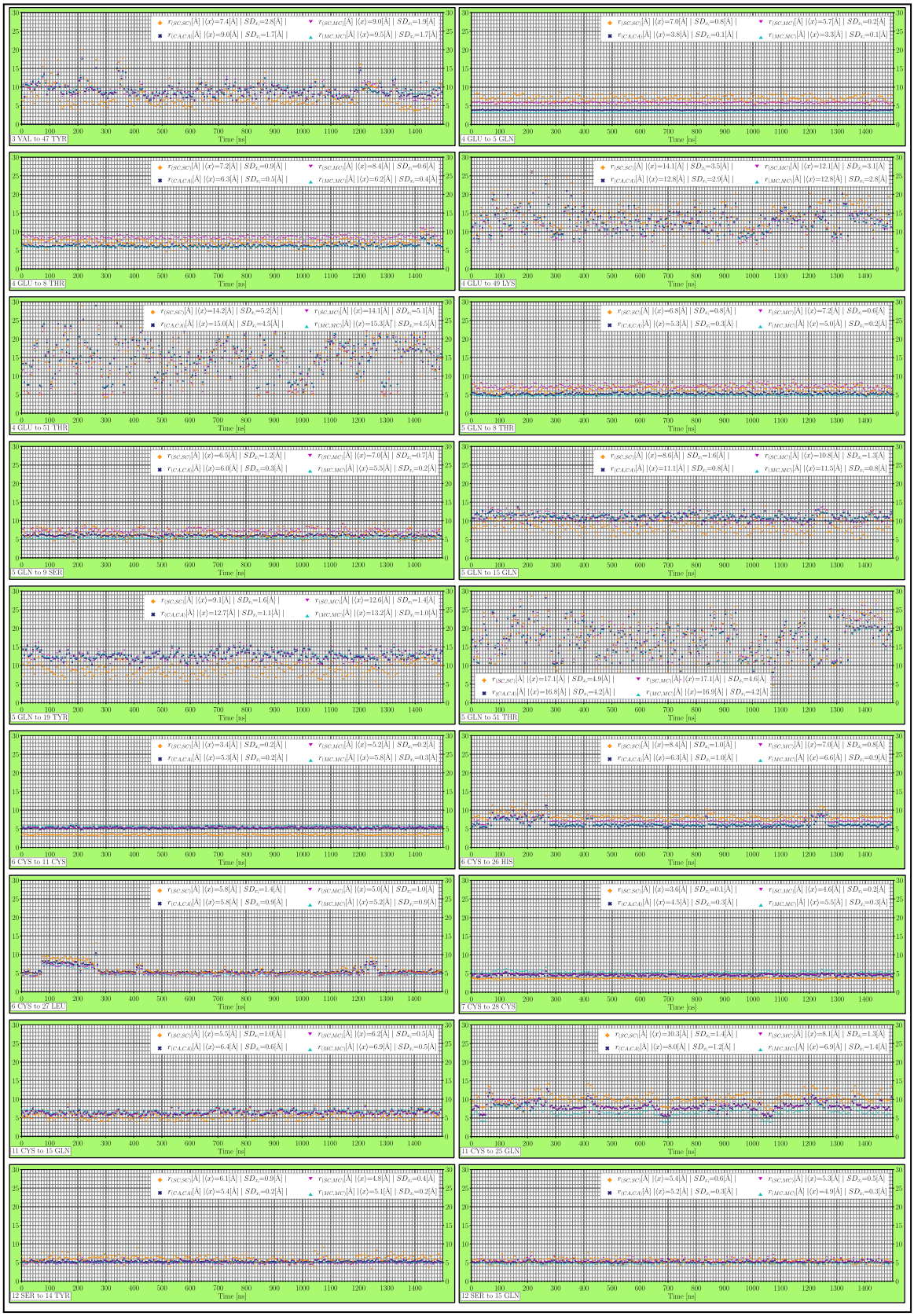


Figure S6.41, ***2nd page.***

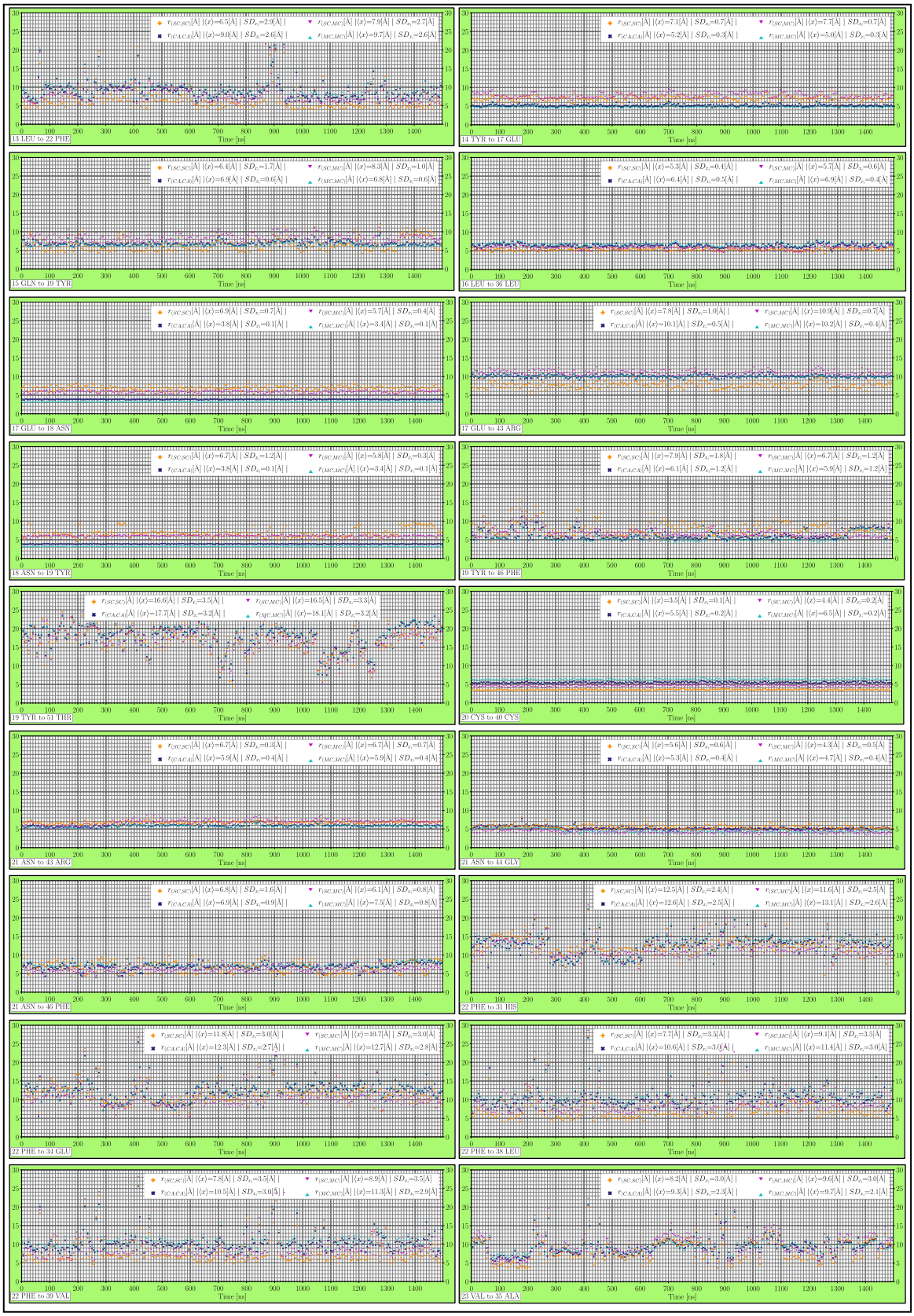


Figure S6.41, ***3rd page.***

Figure S6.41, 4th page.

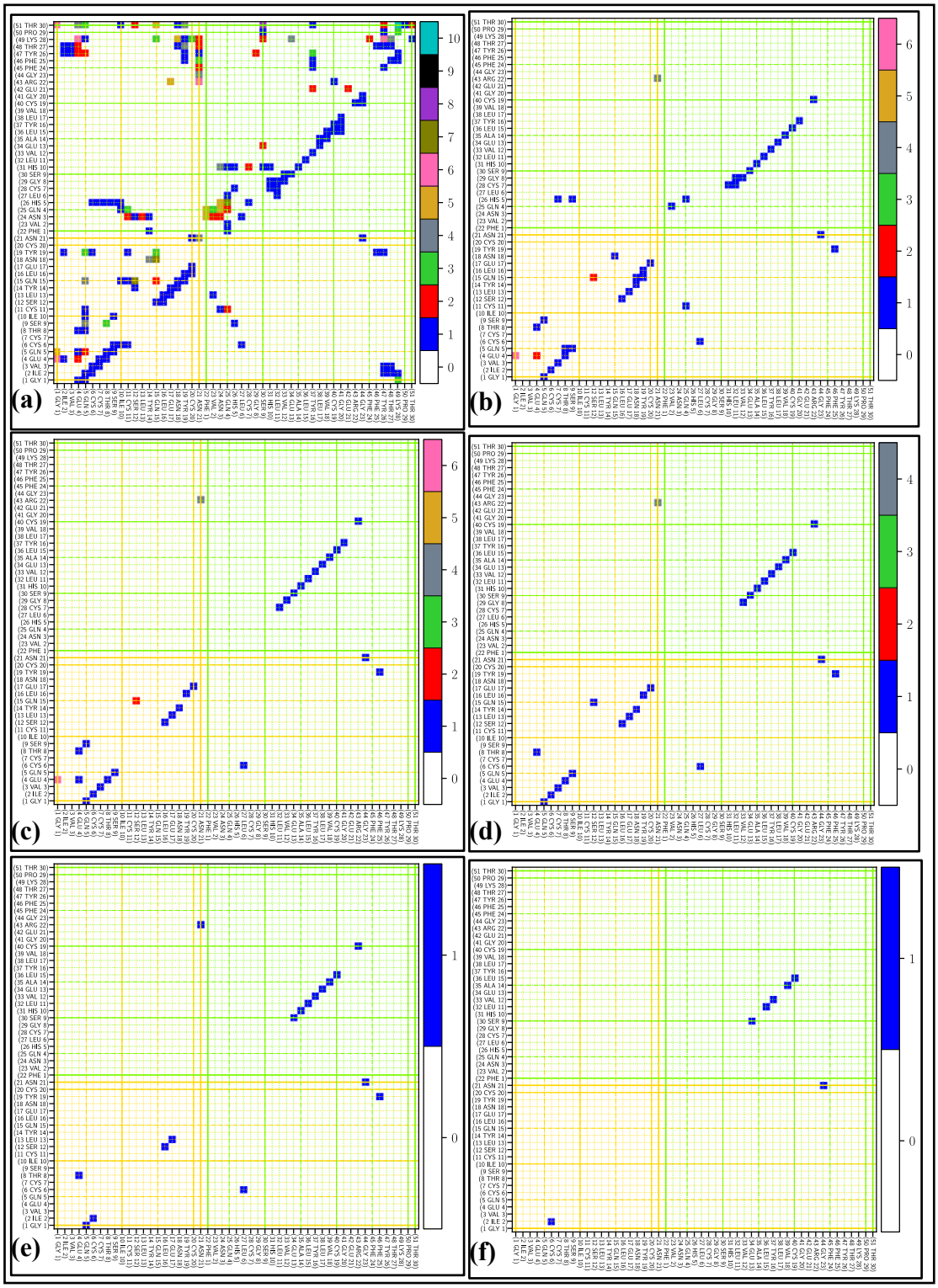


Figure S6.42: *Sorted HBs between insulin residues* $\boldsymbol{P_{kpi}^{MDm}}$ *.Calculated with "* $|\boldsymbol{r}_{AD}| < 3.5$ Å, *&* $\varphi < 30°$*", presence larger than:* ***(a)*** *0% of time (358 HBs);* ***(b)*** *5% of time (51 HBs each graphed in Figure S6.39),* ***(c)*** *10% of time (39 HBs);* ***(d)*** *25% of time (28 HBs each depicted in Figure S6.38); and* ***(e)*** *than 50% of time (17 HBs);* ***(f)*** *75% of time (7 HBs). The HBs are sorted as "SC to MC", "SC to SC" and "NH3+" HBs in upper left; "MC to MC" HBs in lower right; any HBs within a residue in diagonal. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graphs has zoomable vector graphics. Grid-lines for AC (1-21) in gold-yellow, and BC (22-51) green, distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51 (number closest to graph in both i & j direction). The same HBs are also counted in Table 6.3.*

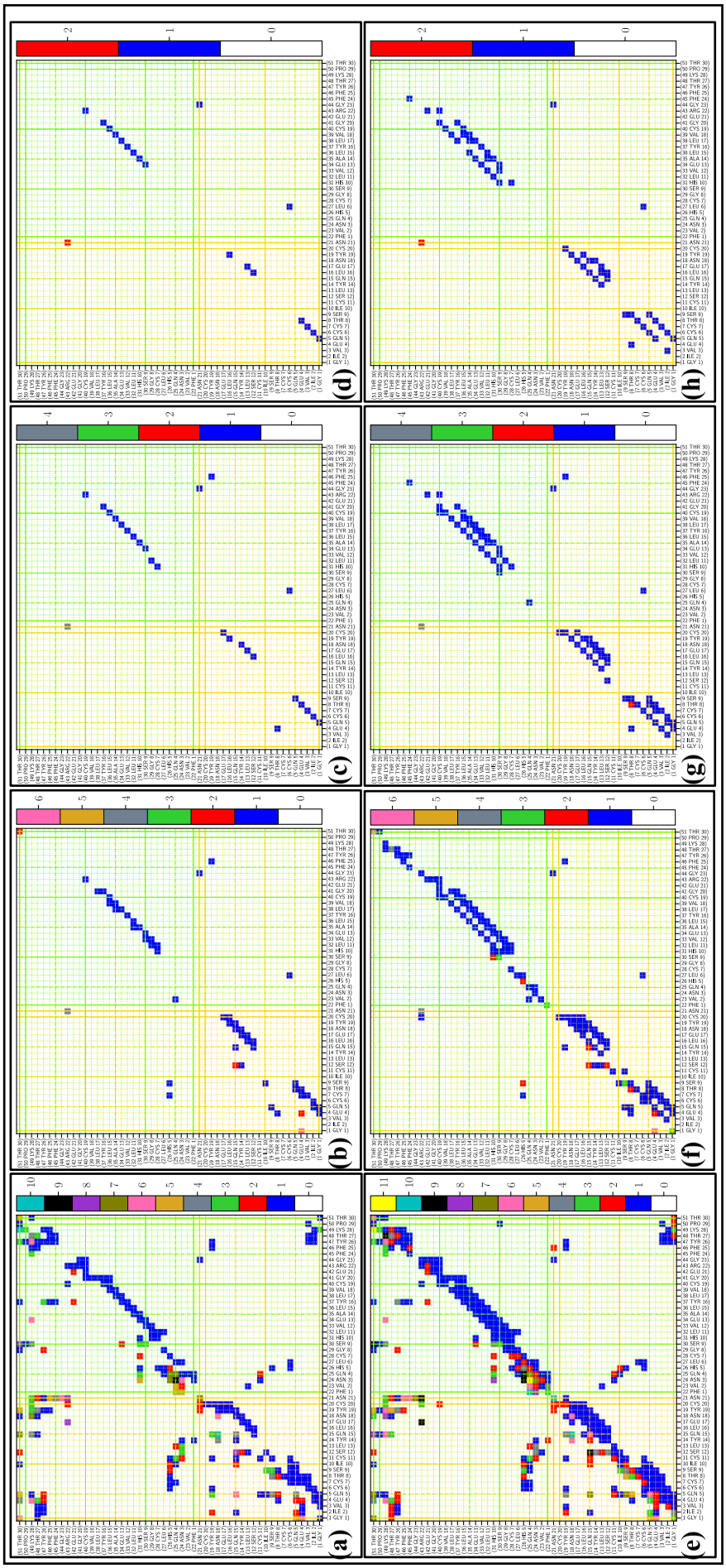


***Figure S6.43**: Hydrogen Bonds between residues of ensemble $\mathbf{P}_{kpi}^{\mathbf{MDm}}$. The HBs calculated with: $|\mathbf{r}_{AD}| < 3.5$ Å, and (Top) $\varphi < 60°$, presence larger than **(a)** 0% of time, 514 HBs, **(b)** 10% of time, 65 HBs, **(c)** 50% of time, 29 HBs, **(d)** 75% of time, 20 HBs. And (Bottom) $\varphi < 90°$, presence larger than **(e)** 0% of time, 746 HBs, **(f)** 10% of time, 160 HBs **(g)** 50% of time, 79 HBs **(h)** 75% of time, 57 HBs. The HBs are sorted as "SC to MC", "SC to SC" and "NH3+" HBs in upper left; "MC to MC" HBs in lower right; any HBs within a residue at diagonal. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Chains residues renumbered sequentially 1-51 (nearest graph), and with actual residue name and number. Graph has zoomable vector graphics. Grid-lines for AC (1-21) in gold-yellow, and BC (22-51) green, distinguished at NTs, CTs and at every 10,5 and 1 steps of 1-51.*

***Table S6.5**: Upper bound violations by respective calculated distances of $\boldsymbol{P_{kpi}^{MDm}}$. Here shown is the 26 UB violations of restrained HB distances (RHBs) and 78 largest of 100 UB violations of hydrogen distances (RHHs) from PDB 2KJJ (where only 793 RHHs were included since omitting GLY HA# out of the 803).*

RHB V_{ij} [Å]			RHH V_{ij} [Å]								
B4 Q(HN)	A11 C(O)	3.24	B10 H(HD2)	B12 V(HG11)	3.93	B7 C(HN)	B6 L(HB1)	0.57	A7 C(HN)	A4 Q(HA)	0.20
B23 G(HN)	B20 G(O)	3.24	B10 H(HD2)	B11 L(HD11)	3.54	B15 L(HN)	B18 V(HG11)	0.55	A6 C(HN)	A5 Q(HA)	0.19
B4 Q(N)	A11 C(O)	2.90	B10 H(HD2)	B12 V(HG21)	2.78	B16 Y(HN)	B12 V(HA)	0.55	A19 Y(HN)	A17 E(HA)	0.19
A10 I(HN)	A5 Q(O)	2.70	B10 H(HD2)	B11 L(HD21)	2.64	B14 A(HN)	B12 V(HG21)	0.53	B16 Y(HN)	B15 L(HA)	0.19
A12 S(N)	A15 Q(OE1)	2.51	B13 E(HA)	B12 V(HG21)	2.58	A2 I(HN)	A19 Y(HE1)	0.50	B19 Y(HN)	B18 V(HA)	0.19
A12 S(HN)	A15 Q(OE1)	2.46	A12 S(HN)	B3 N(HB1)	2.48	A7 C(HB1)	A4 E(HB1)	0.50	B12 V(HN)	B11 L(HA)	0.18
B23 G(N)	B20 G(O)	2.19	A12 S(HN)	A11 L(HB1)	1.55	A17 E(HN)	A17 E(HG1)	0.49	B13 E(HN)	B12 V(HA)	0.18
A11 C(N)	B4 Q(O)	1.94	B10 H(HD2)	B9 S(HA)	1.50	A16 L(HN)	A18 N(HN)	0.49	B17 L(HN)	B16 Y(HA)	0.18
A9 S(HN)	A4 Q(O)	1.88	B11 L(HN)	B13 E(HA)	1.46	A6 C(HN)	A7 C(HB1)	0.48	B3 N(HB1)	A10 I(HD1)	0.18
A11 C(HN)	B4 Q(O)	1.83	B13 E(HG1)	B12 V(HG21)	1.41	A14 Y(HN)	A13 L(HG)	0.48	A3 V(HN)	B27 T(HN)	0.17
A9 S(N)	A4 Q(O)	1.75	A16 L(HA)	A13 L(HB1)	1.35	A12 S(HN)	A15 Q(HN)	0.47	A5 Q(HN)	A4 E(HA)	0.17
A10 I(N)	A5 Q(O)	1.73	A6 C(HB1)	B6 L(HD21)	1.16	A18 N(HN)	A16 L(HA)	0.40	A16 L(HN)	A15 Q(HA)	0.17
A14 Y(HN)	A12 S(O)	0.59	A7 C(HN)	A9 S(HB1)	1.04	B8 G(HN)	B11 L(HD11)	0.39	B11 L(HN)	B10 H(HA)	0.17
A10 I(HN)	A9 S(OG)	0.58	A12 S(HA)	A16 L(HD21)	0.96	B10 H(HA)	B13 E(HA)	0.37	B16 Y(HN)	B17 L(HN)	0.17
A14 Y(N)	A12 S(OG)	0.50	B16 Y(HN)	B18 V(HG11)	0.93	A5 Q(HN)	A8 T(HG21)	0.34	A3 V(HN)	A2 I(HA)	0.16
A14 Y(HN)	A12 S(OG)	0.45	B4 Q(HB1)	B6 L(HB1)	0.83	B13 E(HN)	B14 A(HB1)	0.33	A5 Q(HN)	A3 V(HG11)	0.16
A17 E(HN)	A14 Y(O)	0.31	B16 Y(HB1)	B12 V(HG21)	0.81	B24 F(HD1)	B24 F(HB1)	0.33	B5 H(HE1)	A10 I(HG11)	0.16
B12 V(HN)	B9 S(O)	0.14	B14 A(HA)	B18 V(HG11)	0.79	B17 L(HN)	B15 L(HD21)	0.30	B19 C(HA)	B18 V(HG21)	0.16
A17 E(N)	A14 Y(O)	0.12	B5 H(HN)	B4 Q(HG1)	0.71	B12 V(HN)	B11 L(HB1)	0.28	A2 I(HN)	A3 V(HN)	0.15
B6 L(N)	A6 C(O)	0.12	B22 R(HN)	B23 G(HA2)	0.69	A13 L(HA)	A16 L(HD21)	0.27	A20 G(HN)	A18 N(HA)	0.15
A20 C(N)	A17 E(O)	0.08	B14 A(HN)	B10 H(HA)	0.68	A14 Y(HN)	A13 L(HD11)	0.27	B4 Q(HN)	A10 I(HB)	0.15
A14 Y(N)	A12 S(O)	0.07	A6 C(HN)	B11 L(HD11)	0.66	A17 E(HG1)	B18 V(HG21)	0.27	B11 L(HN)	B9 S(HA)	0.15
A10 I(N)	A9 S(OG)	0.06	A16 L(HN)	A13 L(HB1)	0.65	A6 C(HN)	A2 I(HA)	0.26	B15 L(HG)	B18 V(HG21)	0.15
A18 N(HN)	A15 Q(O)	0.04	B24 F(HZ)	B12 V(HG21)	0.61	B3 N(HB1)	A10 I(HG11)	0.25	A4 E(HA)	A8 T(HB)	0.14
B11 L(HN)	B8 G(O)	0.03	A11 C(HN)	B5 H(HB1)	0.61	B4 Q(HN)	B3 N(HB1)	0.25	A14 Y(HN)	A13 L(HA)	0.14
B6 L(HN)	A6 C(O)	0.01	A16 L(HN)	A13 L(HA)	0.57	B5 H(HE1)	A10 I(HD1)	0.22	B13 E(HN)	B14 A(HN)	0.12

S6.3 Trajectory Ensemble, E_{kpi}^{MD}

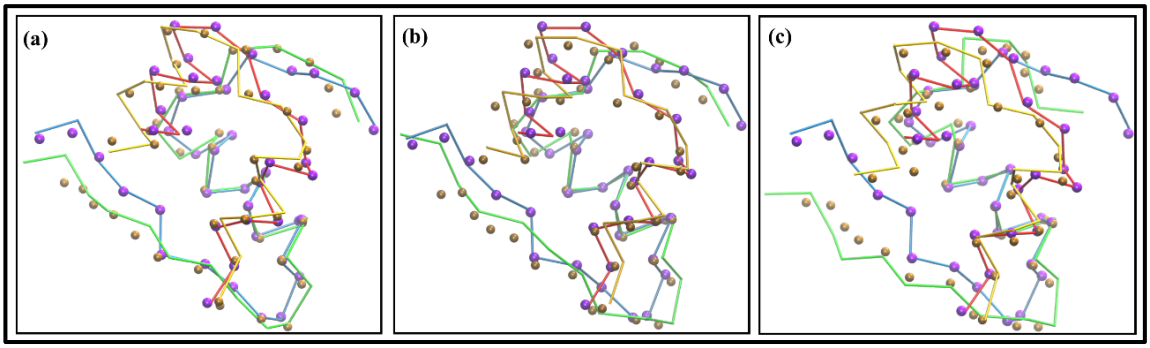


Figure S6.44*: Traced CA-atoms of MSs of* E_{kpi}^{MD} *That of replica* ***(a)*** *"**m**" at 101 ns.* ***(b)*** *replica "**n**" at 126 ns,* ***(c)*** *"**o**" at 1214 ns.*

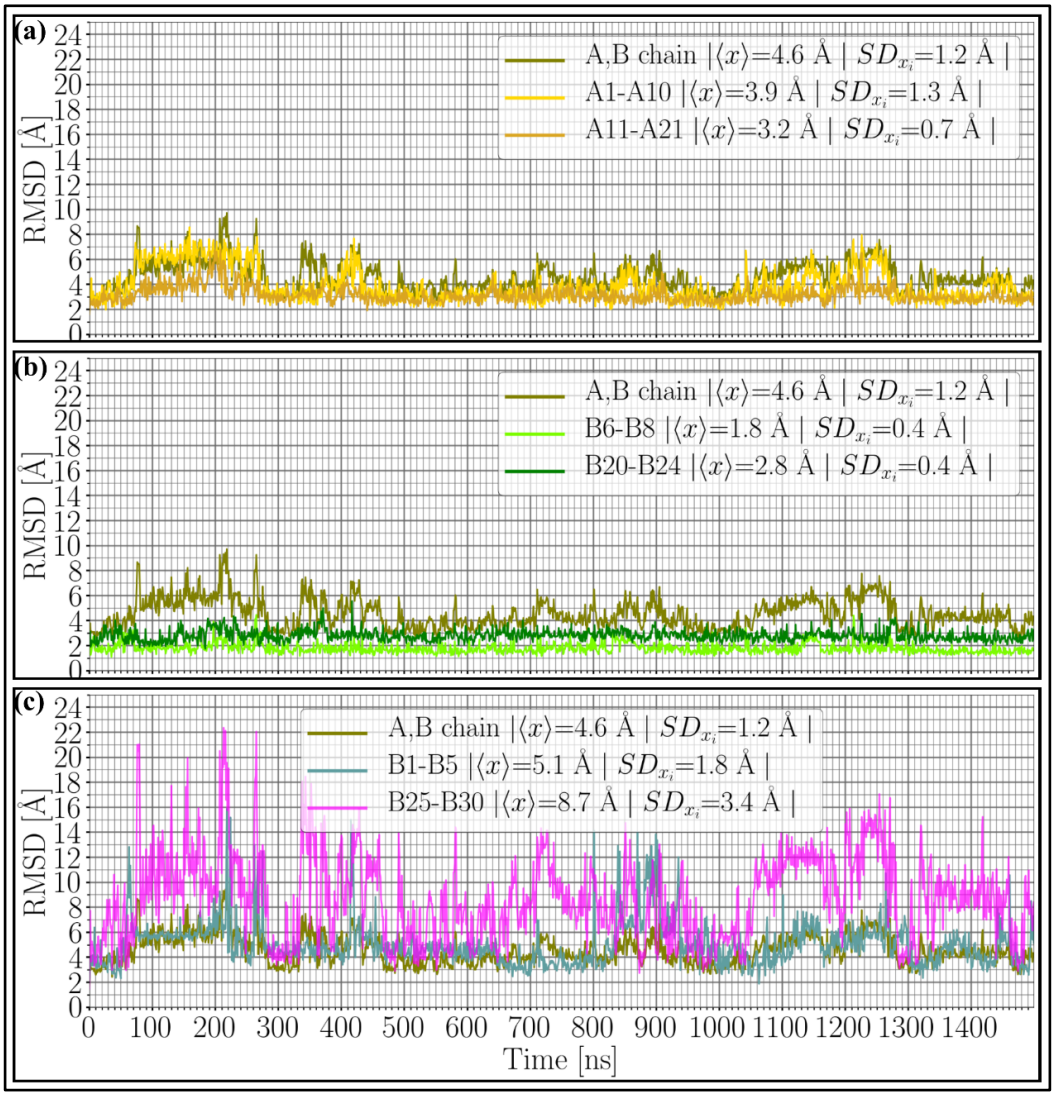


***Figure S6.45**: RMSD of specific regions of the ensemble* $\boldsymbol{E_{kpi}^{MDm}}$ *. Calculations of RMSD includes all atoms of compared residues. Reference structure is MS of* $\boldsymbol{E_{kpi}^{DGR}}$ *(7'th structure).*

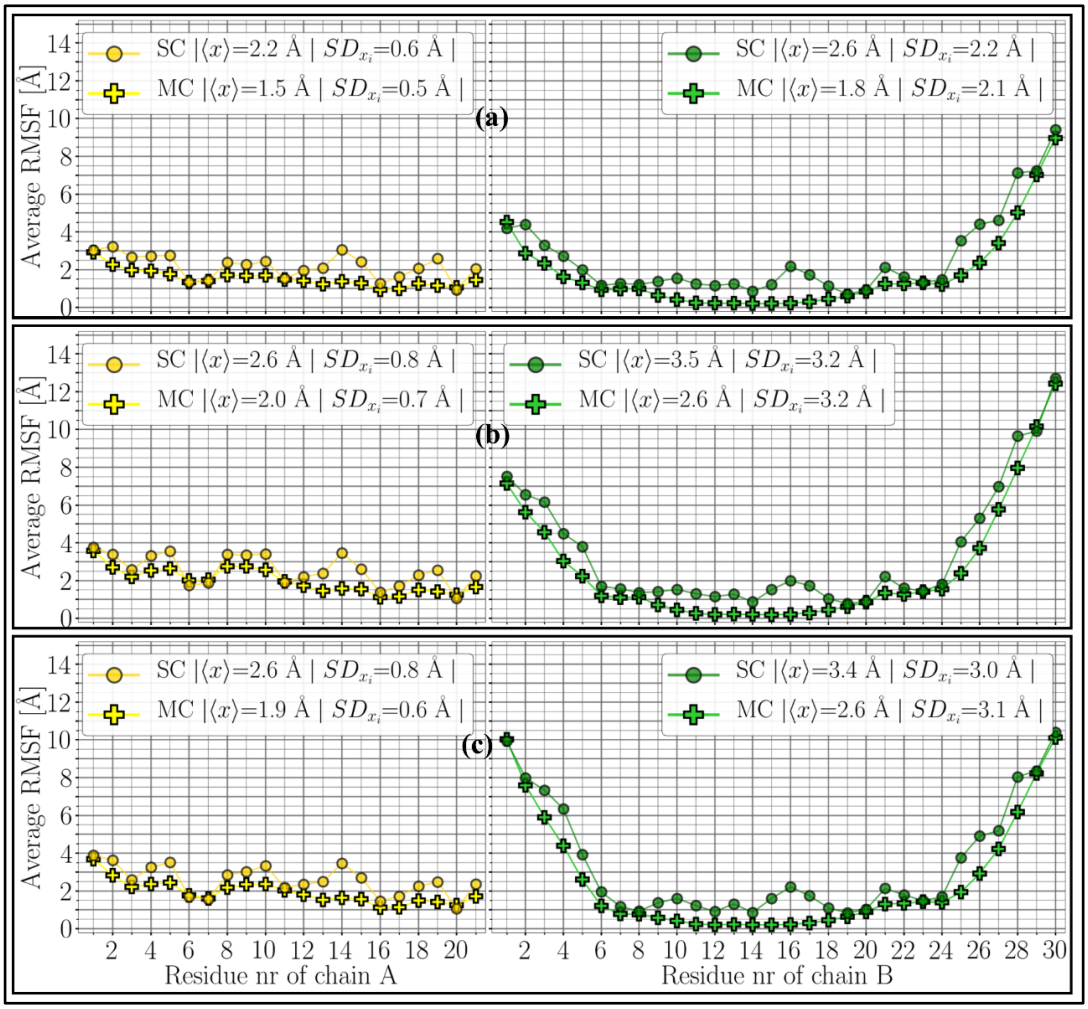


***Figure S6.46**: Average RMSF for ensemble $\boldsymbol{E_{kpi}^{MD}}$, of all atoms of indicated selection (SC or MC), for each residue (i.e. $RMSF_{\langle SC \rangle}$ and $RMSF_{\langle MC \rangle}$), of replica **(a)** "**m**", **(b)** "**n**", **(c)** "**o**".*

S6.4 Trajectory Ensemble, P_i^{MD}

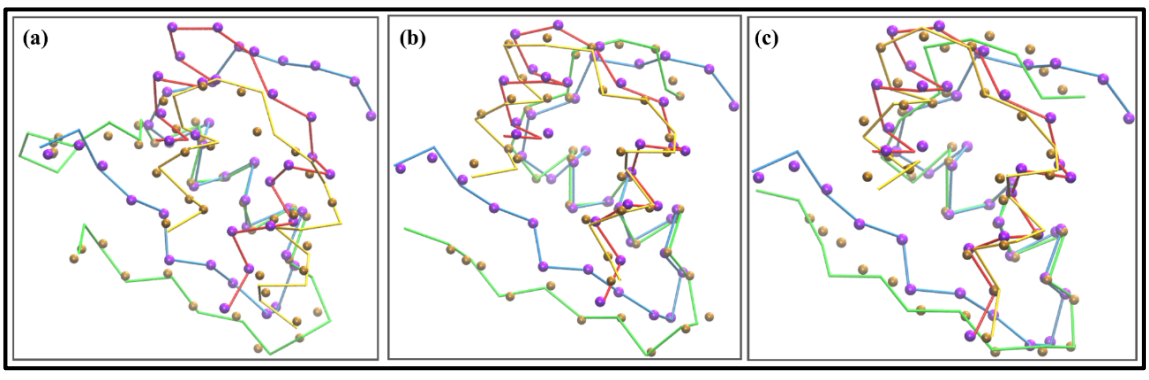


***Figure S6.47**: Traced CA-atoms of MSs of $\boldsymbol{P_i^{MD}}$. That of respective replica **(a)** "**m**" at 1051 ns, **(b)** "**n**" at 1045 ns, **(c)** "**o**" at 800 ns. Compared to traced atoms of MS of $\boldsymbol{E_{kpi}^{DGR}}$ (7'th structure).*

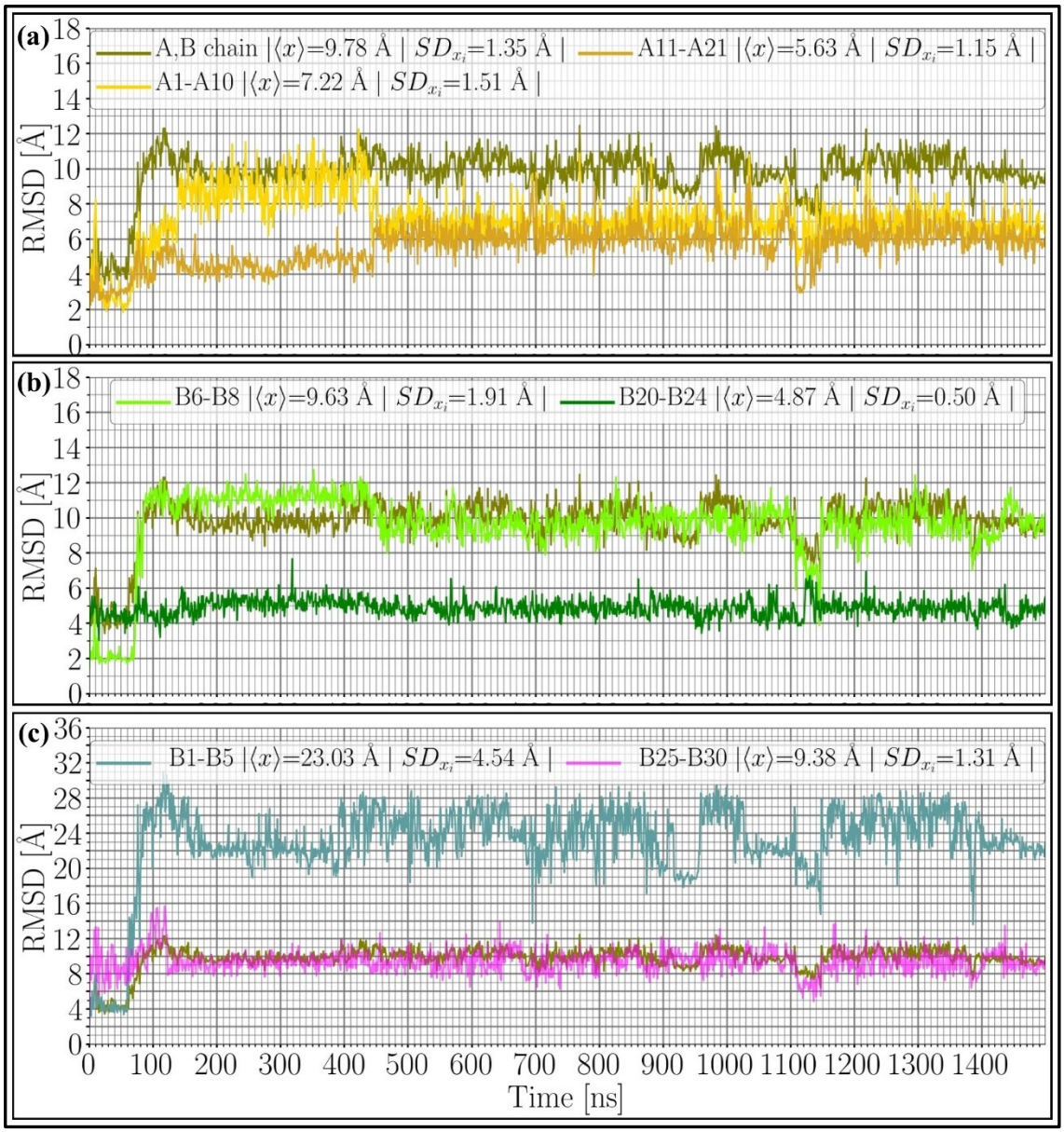


***Figure S6.48**: RMSD of specific regions of the ensemble $\boldsymbol{P_i^{MDm}}$. The RMSD includes all atoms of compared residues. The reference structure is mean structure of the 10 structures of PDB 2HIU.*

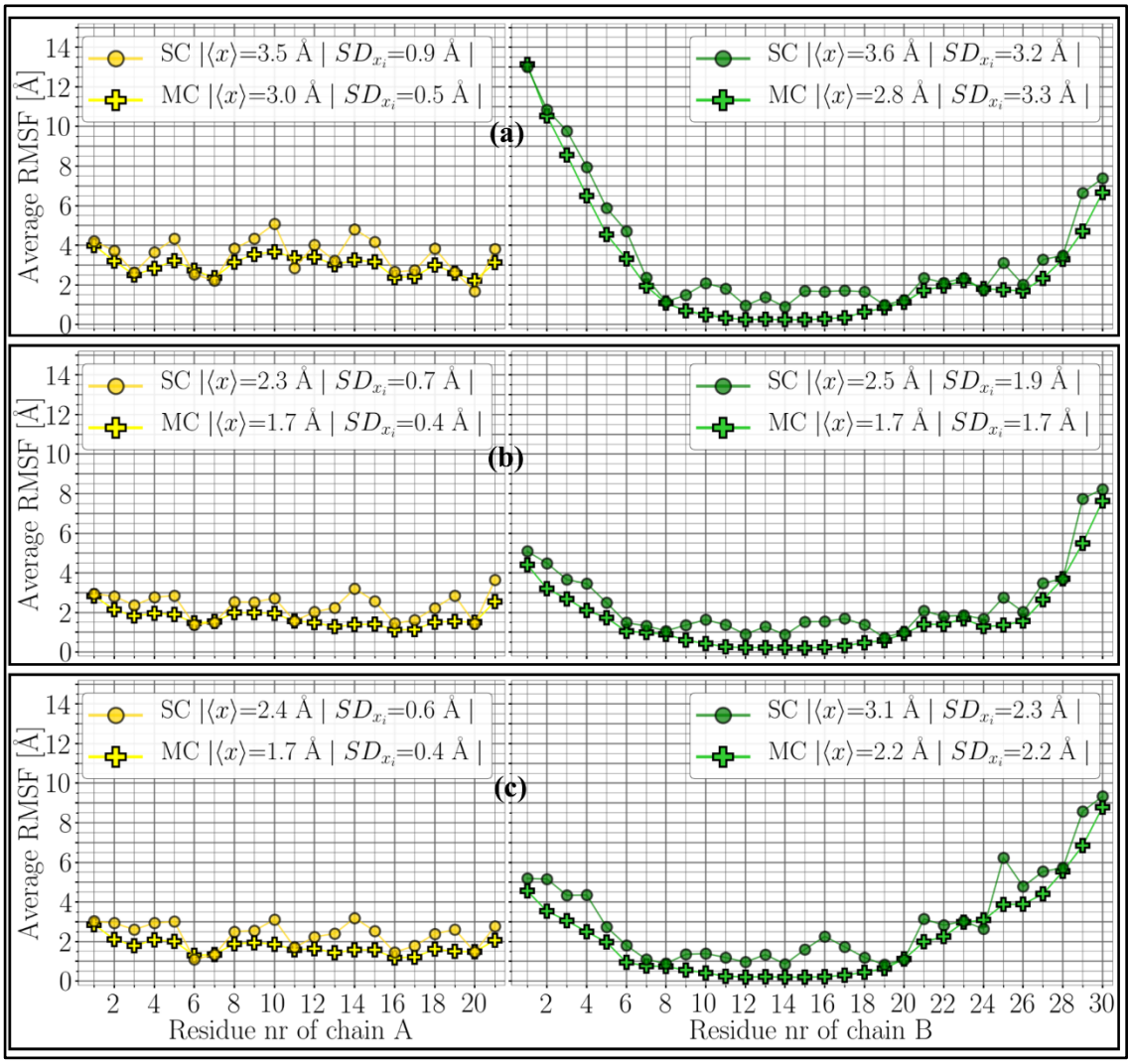


***Figure S6.49**: Average RMSF of ensemble $\boldsymbol{P_i^{MD}}$ for all atoms of indicated selection (SC or MC) for each residue (i.e. $RMSF_{\langle SC \rangle}$ and $RMSF_{\langle MC \rangle}$) of replica **(a)** "**m**", **(b)** "**n**", **(c)** "**o**".*

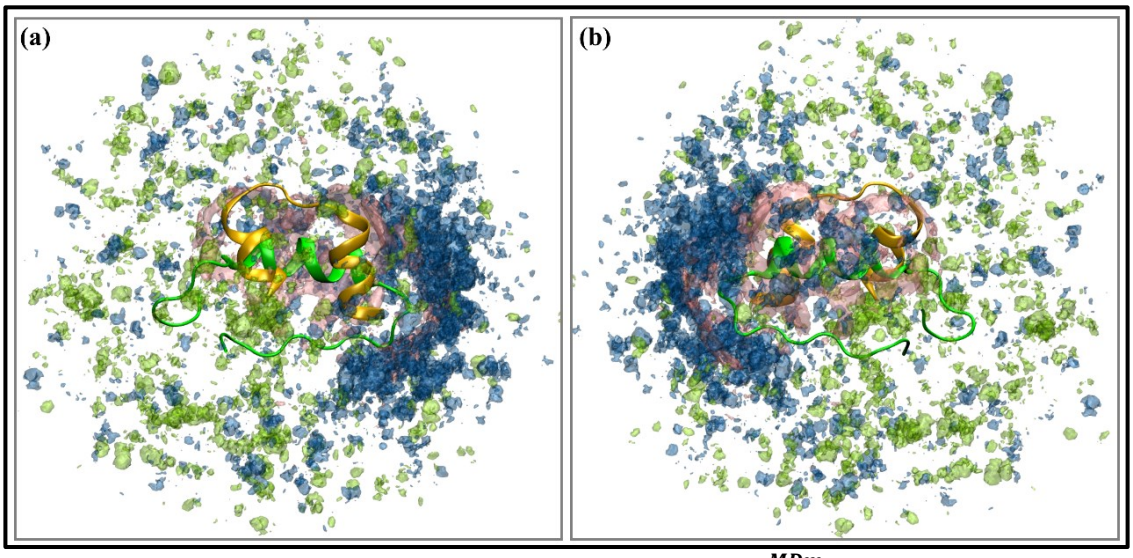


***Figure S6.50**: Fractional occupancy for solutes of ensemble $\boldsymbol{P_i^{MDm}}$. Obtained same way as in §6.2.1.1 and Figure S6.35(a).*

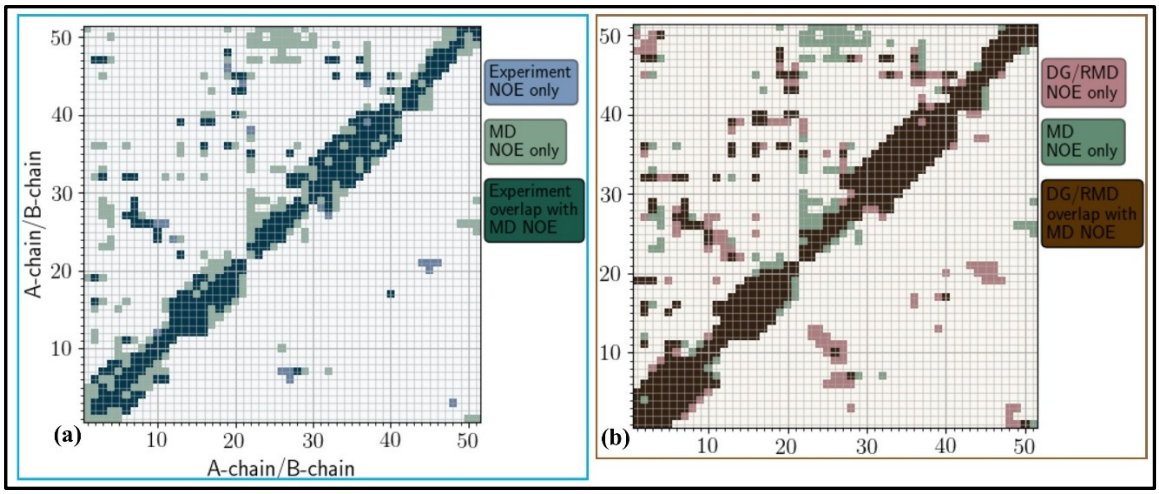


***Figure S6.51**: Comparison of experimental and calculated NOEs of $\boldsymbol{P_i^{MDm}}$, with $\langle r_{ij}^{-6}\rangle^{-1/6}$ averaging. Overlap with **(a)** experimental NOEs (RHH) from PDB entry 2KJJ **(b)** with calculated NOEs of $\boldsymbol{E_{kpi}^{DGR}}$.*

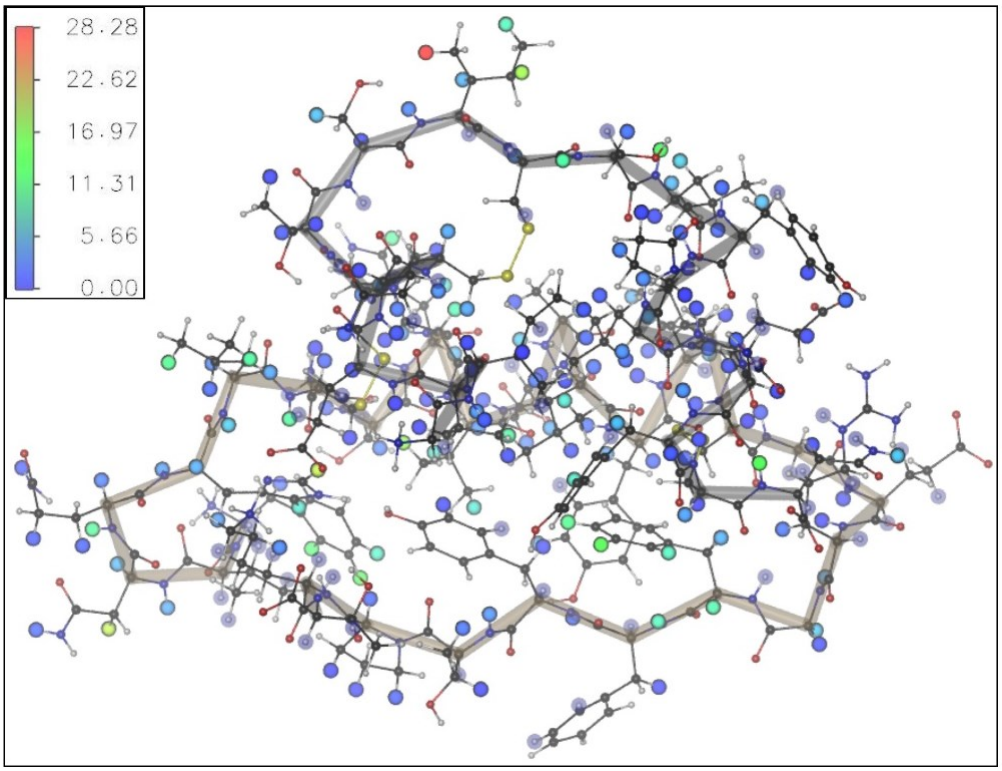


***Figure S6.52**: Visualized RHH UB violations for $\boldsymbol{P_i^{MDm}}$. Transparent bb with AC black, BC brown (curvature at CA atoms). Smaller atoms with no RHH assignment in ordinary atom-colouring. The bigger atoms and the colour-bar depicts which atoms has accumulated V_{ij} (the RHH hydrogens with no violation being in transparent blue). The atom in red (I_{HG21}^{A10}) shows the largest accumulated violation at 28.28 Å. Depiction shown on MS at 1051 ns.*

BIBLIOGRAPHY

[1] E. N. Baker *et al.*, "The Structure of 2Zn Pig Insulin Crystals at 1.5 Å Resolution," *Philosophical Transactions of the Royal Society of London. B, Biological Sciences,* vol. 319, no. 1195, pp. 369-456, 1988.

[2] F. Weis *et al.*, "The signalling conformation of the insulin receptor ectodomain," *Nature Communications,* vol. 9, no. 1, p. 4420, 2018.

[3] Q.-x. Hua, W. Jai, and M. A. Weiss, "Conformational Dynamics of Insulin," *Frontiers in Endocrinology,* Original Research vol. 2, 2011.

[4] P. De Meyts, "Insulin and its receptor: structure, function and evolution," *BioEssays,* vol. 26, no. 12, pp. 1351-1362, 2004.

[5] "Global report on diabetes," *World Health Organisation,* 2016.

[6] R. F. Dods, "Chap. 6, Classification system for diabetes mellitus," in *Understanding Diabetes A Biochemical Perspective* L. Ebook, Ed. 1 ed.: John Wiley & Sons, 2013, pp. 183-198

[7] C. C. Quianzon and I. Cheikh, "History of insulin," *Journal of Community Hospital Internal Medicine Perspectives,* vol. 2, no. 2, p. 18701, 2012.

[8] "The Discovery of Insulin," *Nobelprize.org. Nobel Media AB*.

[9] Organisation - Diabetes Australia [Online].

[10] B. Guerci and J. P. Sauvanet, "Subcutaneous insulin: pharmacokinetic variability and glycemic variability," *Diabetes & Metabolism,* vol. 31, no. 4, pp. 4S7-4S24, 2005.

[11] S. L. Noble, E. Johnston, and B. Walton, "Insulin lispro: a fast-acting insulin analog," *Am Fam Physician,* vol. 57, no. 2, pp. 279-86, 289-92, 1998.

[12] I. Hartman, "Insulin Analogs: Impact on Treatment Success, Satisfaction, Quality of Life, and Adherence," *Clinical Medicine & Research,* vol. 6, no. 2, pp. 54-67, 2008.

[13] D. Goodsell, "Insulin Receptor," *Molecule of the Month, February* 2015.

[14] S. A. Tatulian, "Structural Dynamics of Insulin Receptor and Transmembrane Signaling," *Biochemistry,* vol. 54, no. 36, pp. 5523-5532, 2015.

[15] M. White, "IRS proteins and the common path to diabetes," *American Journal of Physiology,* vol. 46, no. 3, pp. E413-E422, 2002.

[16] S. Fröjdö, H. Vidal, and L. Pirola, "Alterations of insulin signaling in type 2 diabetes: A review of the current evidence from humans," *BBA - Molecular Basis of Disease,* vol. 1792, no. 2, pp. 83-92, 2009.

[17] J. A. Archer, P. Gorden, and J. Roth, "Defect in insulin binding to receptors in obese man. Amelioration with calorie restriction," *The Journal of Clinical Investigation,* vol. 55, no. 1, pp. 166-174, 1975.

[18] P. De Meyts and J. Whittaker, "Structural biology of insulin and IGF1 receptors: implications for drug design," *Nature Reviews Drug Discovery,* vol. 1, no. 10, pp. 769-783, 2002.

[19] S. A. Ross, E. A. Gulve, and M. Wang, "Chemistry and Biochemistry of Type 2 Diabetes," *Chemical Reviews,* vol. 104, no. 3, pp. 1255-1282, 2004.

[20] Q.-x. Hua *et al.*, "Design of an Active Ultrastable Single-chain Insulin Analog," *The Journal of Biological Chemistry,* vol. 283, no. 21, pp. 14703-14716, 2008.

[21] M. D. Glidden *et al.*, "Solution structure of an ultra-stable single-chain insulin analog connects protein dynamics to a novel mechanism of receptor binding," *Journal of Biological Chemistry,* vol. 293, no. 1, pp. 69-88, 2018.

[22] F. G. Banting, C. H. Best, J. B. Collip, W. R. Campbell, and A. A. Fletcher, "Pancreatic Extracts in the Treatment of Diabetes Mellitus," *Canadian Medical Association Journal,* vol. 12, no. 3, pp. 141-146, 1922.
[23] A. O. W. Stretton, "The first sequence. Fred Sanger and insulin," *Genetics,* vol. 162, no. 2, pp. 527-532, 2002.
[24] "The Nobel Prize in Chemistry 1958," *Nobelprize.org. Nobel Media AB*.
[25] J. M. B. Lubert Stryer, John L. Tymoczko. , *Biochemistry 7th edition*. W. H. Freeman and Company, 2011.
[26] M. J. Adams *et al.*, "Structure of Rhombohedral 2 Zinc Insulin Crystals," *Nature,* vol. 224, no. 5218, pp. 491-495, 1969.
[27] "The Nobel Prize in Chemistry 1964," *Nobelprize.org. Nobel Media AB*.
[28] R. B. Phillips, J. Kondev, and J. Theriot, "Chap. 5 Mechanical and Chemical Equilibrium in the Living Cell. ," in *Physical Biology of the Cell*: Garland Science, 2009, pp. 167-214.
[29] X. Z. Ke-li Han, Ming-jun Yang, "Chap. 8, Simulating Protein Folding in Different Environmental Conditions," in *Protein Conformational Dynamics* J. D. Lambris, Ed. no. 805): Springer International Publishing, 2014, pp. 171-198.
[30] P. De Meyts, "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,* journal article vol. 37, no. 2, pp. S135-S148, 1994.
[31] P. De Meyts, "Insulin/receptor binding: The last piece of the puzzle?," *BioEssays,* vol. 37, no. 4, pp. 389-397, 2015.
[32] C. W. Ward, J. G. Menting, and M. C. Lawrence, "The insulin receptor changes conformation in unforeseen ways on ligand binding: Sharpening the picture of insulin receptor activation," *BioEssays,* vol. 35, no. 11, pp. 945-954, 2013.
[33] E. Callaway, "The revolution will not be crystallized: a new method sweeps through structural biology," *Nature,* vol. 525, no. 7568, p. 172, 2015.
[34] H. N. Chapman *et al.*, "Femtosecond X-ray protein nanocrystallography," *Nature,* vol. 470, p. 73, 2011.
[35] H. P. A. Wittler, G. A. v. Riessen, and M. W. M. Jones, "The influence of noise on image quality in phase-diverse coherent diffraction imaging," *Journal of Optics,* vol. 18, no. 2, p. 024001, 2016.
[36] R. Neutze, G. Brändén, and G. F. X. Schertler, "Membrane protein structural biology using X-ray free electron lasers," *Current Opinion in Structural Biology,* vol. 33, no. Supplement C, pp. 115-125, 2015.
[37] R. Neutze, G. Huldt, J. Hajdu, and D. van der Spoel, "Potential impact of an X-ray free electron laser on structural biology," *Radiation Physics and Chemistry,* vol. 71, no. 3, pp. 905-916, 2004.
[38] J. C. H. Spence, "XFELs for structure and dynamics in biology," *IUCrJ,* vol. 4, no. Pt 4, pp. 322-339, 2017.
[39] "Future perfect," *Nature Chemical Biology,* Editorial vol. 11, p. 889, 2015.
[40] J. C. H. Spence, "X-ray lasers for structure and dynamics in biology," *IUCrJ,* vol. 5, no. Pt 3, pp. 236-237, 2018.
[41] R. Neutze, "Snapshots of a protein quake," *Science,* vol. 350, no. 6259, pp. 381-381, 2015.
[42] S. I. O'Donoghue *et al.*, "Visualizing biological data—now and in the future," *Nature Methods,* vol. 7, p. S2, 2010.
[43] H. Miao *et al.*, "Multiscale Molecular Visualization," *Journal of Molecular Biology,* vol. 431, no. 6, pp. 1049-1070, 2019.

[44] S. Gui, D. Khan, Q. Wang, D.-M. Yan, and B.-Z. Lu, "Frontiers in biomolecular mesh generation and molecular visualization systems," *Visual Computing for Industry, Biomedicine, and Art,* vol. 1, no. 1, p. 7, 2018.
[45] G. Wei, W. Xi, R. Nussinov, and B. Ma, "Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell," *Chemical reviews,* vol. 116, no. 11, p. 6516, 2016.
[46] K. Gaalswyk, M. I. Muniyat, and J. L. Maccallum, "The emerging role of physical modeling in the future of structure determination," *Current Opinion in Structural Biology,* vol. 49, pp. 145-153, 2018.
[47] D. W. Borhani and D. E. Shaw, "The future of molecular dynamics simulations in drug discovery," *Journal of Computer-Aided Molecular Design,* journal article vol. 26, no. 1, pp. 15-26, 2012.
[48] M. J. Abraham *et al.*, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX,* vol. 1–2, pp. 19-25, 2015.
[49] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry,* vol. 26, no. 16, pp. 1701-1718, 2005.
[50] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics,* vol. 14, no. 1, pp. 33-38, 1996.
[51] G. Zhao *et al.*, "Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics," *Nature,* vol. 497, p. 643, 2013.
[52] G. Gonzalez-Gutierrez, Y. Wang, G. D. Cymes, E. Tajkhorshid, and C. Grosman, "Chasing the open-state structure of pentameric ligand-gated ion channels," *The Journal of General Physiology,* vol. 149, no. 12, p. 1119, 2017.
[53] M. Karplus, "Molecular dynamics of biological macromolecules: A brief history and perspective," *Biopolymers,* vol. 68, no. 3, pp. 350-358, 2003.
[54] P. Krüger, W. Straßburger, A. Wollmer, W. F. van Gunsteren, and G. G. Dodson, "The simulated dynamics of the insulin monomer and their relationship to the molecule's structure," *European Biophysics Journal,* journal article vol. 14, no. 8, pp. 449-459, 1987.
[55] S. Yun-yu, Y. Ru-huai, and W. F. van Gunsteren, "Molecular dynamics simulation of despentapeptide insulin in a crystalline environment," *Journal of Molecular Biology,* vol. 200, no. 3, pp. 571-577, 1988.
[56] M. Falconi, M. T. Cambria, A. Cambria, and A. Desideri, "Structure and Stability of the Insulin Dimer Investigated by Molecular Dynamics Simulation," *Journal of Biomolecular Structure and Dynamics,* vol. 18, no. 5, pp. 761-772, 2001.
[57] V. Zoete, M. Meuwly, and M. Karplus, "A Comparison of the Dynamic Behavior of Monomeric and Dimeric Insulin Shows Structural Rearrangements in the Active Monomer," *Journal of Molecular Biology,* vol. 342, no. 3, pp. 913-929, 2004.
[58] A. Papaioannou, S. Kuyucak, and Z. Kuncic, "Molecular Dynamics Simulations of Insulin: Elucidating the Conformational Changes that Enable Its Binding," *PLoS ONE,* vol. 10, no. 12, p. 0144058, 2015.
[59] A. Papaioannou, S. Kuyucak, and Z. Kuncic, "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,* vol. 11, pp. 182-190, 2017.
[60] A. Grossfield and D. M. Zuckerman, "Quantifying uncertainty and sampling

quality in biomolecular simulations," *Annual reports in computational chemistry,* vol. 5, pp. 23-48, 2009.

[61] A. B. Ward, A. Sali, and I. A. Wilson, "Integrative Structural Biology," *Science,* vol. 339, no. 6122, pp. 913-915, 2013.

[62] Q.-X. Hua, S. N. Gozani, R. E. Chance, J. A. Hoffmann, B. H. Frank, and M. A. Weiss, "Structure of a protein in a kinetic trap," *Nat Struct Mol Biol,* vol. 2, no. 2, pp. 129-138, 1995.

[63] Q.-x. Hua, S. Nakagawa, S.-Q. Hu, W. Jia, S. Wang, and M. A. Weiss, "Toward the Active Conformation of Insulin: stereospecific modulation of a structural switch in the B-chain," *Journal of Biological Chemistry,* vol. 281, no. 34, pp. 24900-24909, 2006.

[64] Q. X. Hua, S. E. Shoelson, M. Kochoyan, and M. A. Weiss, "Receptor binding redefined by a structural switch in a mutant human insulin," *Nature,* vol. 354, no. 6350, pp. 238-241, 1991.

[65] R. G. Mirmira, S. H. Nakagawa, and H. S. Tager, "Importance of the character and configuration of residues B24, B25, and B26 in insulin-receptor interactions," *Journal of Biological Chemistry,* vol. 266, no. 3, pp. 1428-1436, 1991.

[66] S. H. Nakagawa and H. S. Tager, "Importance of aliphatic side-chain structure at positions 2 and 3 of the insulin A chain in insulin-receptor interactions," *Biochemistry,* vol. 31, no. 12, pp. 3204-3214, 1992.

[67] J. M. Conlon, "Evolution of the insulin molecule: insights into structure-activity and phylogenetic relationships," *Peptides,* vol. 22, no. 7, pp. 1183-1193, 2001.

[68] G. G. Dodson and J. L. Whittingham, "Insulin: Sequence, Structure and Function - A Story of Surprises," in *Insulin & Related Proteins - Structure to Function and Pharmacology*, M. L. Dieken, M. Federwisch, and P. De Meyts, Eds. Dordrecht: Springer Netherlands, 2002, pp. 29-39.

[69] J. G. Menting *et al.*, "Protective hinge in insulin opens to enable its receptor engagement," *Proceedings of the National Academy of Sciences,* vol. 111, no. 33, pp. E3395-E3404, 2014.

[70] G. Scapin *et al.*, "Structure of the insulin receptor–insulin complex by single-particle cryo-EM analysis," *Nature,* vol. 556, p. 122, 2018.

[71] J. G. Menting *et al.*, "How insulin engages its primary binding site on the insulin receptor," *Nature,* vol. 493, no. 7431, pp. 241-245, 2013.

[72] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations," *Journal of Computational Chemistry,* vol. 32, no. 10, pp. 2319-2327, 2011.

[73] A. Caflisch, F. Rao, G. Settanni, M. Cecchini, and M. Seeber, "Wordom: a program for efficient analysis of molecular dynamics simulations," *Bioinformatics,* vol. 23, no. 19, pp. 2625-2627, 2007.

[74] E. Lindahl, B. Hess, and D. van der Spoel, "GROMACS 3.0: a package for molecular simulation and trajectory analysis," *Molecular modeling annual,* vol. 7, no. 8, pp. 306-317, 2001.

[75] D. R. Roe and T. E. Cheatham, "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data," *Journal of Chemical Theory and Computation,* vol. 9, no. 7, pp. 3084-3095, 2013.

[76] T. J. Callahan, E. Swanson, and T. P. Lybrand, "MD display: An interactive graphics program for visualization of molecular dynamics trajectories," *Journal of Molecular Graphics,* vol. 14, no. 1, pp. 39-41, 1996.

[77] L. Laaksonen, "A graphics program for the analysis and display of molecular dynamics trajectories," *Journal of Molecular Graphics,* vol. 10, no. 1, pp. 33-34,

1992.
[78] D. E. S. R. Desmond Molecular Dynamics System, New York, NY, 2019. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2019.
[79] B. R. Brooks *et al.*, "CHARMM: the biomolecular simulation program," *Journal of computational chemistry,* vol. 30, no. 10, pp. 1545-1614, 2009.
[80] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *Journal of Computational Physics,* vol. 117, no. 1, pp. 1-19, 1995.
[81] W. R. P. Scott *et al.*, "The GROMOS biomolecular simulation program package.(Groningen Molecular Simulation)," *Journal of Physical Chemistry A,* vol. 103, no. 19, p. 3596, 1999.
[82] "Lawrence (2014): How insulin binds its receptor - an event relevant to multiple disease states," *YouTube channel WalterandElizaHall* 2014.
[83] Y. V. Li, "Zinc and insulin in pancreatic beta-cells," *Endocrine,* vol. 45, no. 2, pp. 178-189, 2014.
[84] S. O. Emdin, G. G. Dodson, J. M. Cutfield, and S. M. Cutfield, "Role of zinc in insulin biosynthesis," *Diabetologia,* vol. 19, no. 3, pp. 174-182, 1980.
[85] P. E. MacDonald and P. Rorsman, "Oscillations, Intercellular Coupling, and Insulin Secretion in Pancreatic β Cells (Primer)," *PLoS Biology,* vol. 4, no. 2, p. e49, 2006.
[86] R. S. Alan and C. R. Kahn, "Insulin signalling and the regulation of glucose and lipid metabolism," vol. 414, p. 799, 2001.
[87] C. M. Taniguchi, B. Emanuelli, and C. R. Kahn, "Critical nodes in signalling pathways: insights into insulin action," *Nature Reviews Molecular Cell Biology,* vol. 7, no. 2, p. 85, 2006.
[88] H. Iwase, M. Kobayashi, M. Nakajima, and T. Takatori, "The ratio of insulin to C-peptide can be used to make a forensic diagnosis of exogenous insulin overdosage," *Forensic Science International,* vol. 115, no. 1, pp. 123-127, 2001.
[89] E. Grapengiesser, B. Hellman, A. Salehi, and S. S. Qader, "Glucose Induces Glucagon Release Pulses Antisynchronous with Insulin and Sensitive to Purinoceptor Inhibition," *Endocrinology,* vol. 147, no. 7, pp. 3472-3477, 2006.
[90] J. J. Meier, J. D. Veldhuis, and P. C. Butler, "Pulsatile Insulin Secretion Dictates Systemic Insulin Delivery by Regulating Hepatic Insulin Extraction In Humans," *Diabetes,* vol. 54, no. 6, pp. 1649-1656, 2005.
[91] B. Hellman, E. Gylfe, E. Grapengiesser, H. Dansk, and A. Salehi, "Insulin oscillations--clinically important rhythm. Antidiabetics should increase the pulsative component of the insulin release," *Läkartidningen,* vol. 104, no. 32-33, pp. 2236-2239, 2007.
[92] M. Krupp and M. D. Lane, "On the mechanism of ligand-induced down-regulation of insulin receptor level in the liver cell," *Journal of Biological Chemistry,* vol. 256, no. 4, pp. 1689-1694, 1981.
[93] R. S. Bar, P. Gorden, J. Roth, C. R. Kahn, and P. De Meyts, "Fluctuations in the affinity and concentration of insulin receptors on circulating monocytes of obese patients: effects of starvation, refeeding, and dieting," *The Journal of Clinical Investigation,* vol. 58, no. 5, pp. 1123-1135, 1976.
[94] J. L. Carpentier, M. Fehlmann, E. Van Obberghen, P. Gorden, and L. Orci, "Insulin receptor internalization and recycling: mechanism and significance," *Biochimie,* vol. 67, no. 10, pp. 1143-1145, 1985.
[95] S. Terris and D. F. Steiner, "Binding and degradation of 125I-insulin by rat hepatocytes," *Journal of Biological Chemistry,* vol. 250, no. 21, pp. 8389-98, 1975.

[96] L. Zaliauskiene, S. Kang, C. G. Brouillette, J. Lebowitz, R. B. Arani, and J. F. Collawn, "Down-regulation of cell surface receptors is modulated by polar residues within the transmembrane domain," *Molecular biology of the cell,* vol. 11, no. 8, p. 2643, 2000.
[97] C. W. Mike Lawrence, " Insulin receptor and insulin action," *Diapedia* vol. 51040851452 rev. no. 25, 2014.
[98] K. Siddle, "Signalling by insulin and IGF receptors: supporting acts and new players," (in English), vol. 47, no. 1, p. R1, 2011.
[99] F. C. Kosmakos and J. Roth, "Insulin-induced loss of the insulin receptor in IM-9 lymphocytes. A biological process mediated through the insulin receptor," *Journal of Biological Chemistry,* vol. 255, no. 20, pp. 9860-9, 1980.
[100] J. R. Gavin, J. Roth, D. M. Neville, P. De Meyts, and D. N. Buell, "Insulin-Dependent Regulation of Insulin Receptor Concentrations: A Direct Demonstration in Cell Culture," *Proceedings of the National Academy of Sciences,* vol. 71, no. 1, pp. 84-88, 1974.
[101] J. L. Carpentier, E. Van Obberghen, P. Gorden, and L. Orci, "Surface redistribution of 125I-insulin in cultured human lymphocytes," *The Journal of Cell Biology,* vol. 91, no. 1, pp. 17-25, 1981.
[102] J.-L. Carpentier, H. Gazzano, E. Van Obberghen, M. Fehlmann, P. Freychet, and L. Orci, "Internalization and recycling of 125I-photoreactive insulin-receptor complexes in hepatocytes in primary culture," *Molecular and Cellular Endocrinology,* vol. 47, no. 3, pp. 243-255, 1986.
[103] M. D. Hollenberg, "Mechanisms of receptor-mediated transmembrane signalling," *Experientia,* vol. 42, no. 7, pp. 718-727, 1986.
[104] C. Chakraborty, S. S. Roy, M. J. Hsu, and G. Agoramoorthy, "Can computational biology improve the phylogenetic analysis of insulin?," *Computer Methods and Programs in Biomedicine,* vol. 108, no. 2, pp. 860-872, 2012.
[105] C. Kristensen *et al.*, "Alanine Scanning Mutagenesis of Insulin," *Journal of Biological Chemistry,* vol. 272, no. 20, pp. 12978-12983, 1997.
[106] E. J. Dodson *et al.*, "Insulin Assembly: Its Modification by Protein Engineering and Ligand Binding," *Philosophical Transactions: Physical Sciences and Engineering,* vol. 345, no. 1674, pp. 153-164, 1993.
[107] U. Derewenda *et al.*, "Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer," *Nature,* vol. 338, no. 6216, pp. 594-596, 1989.
[108] G. Bentley, E. Dodson, G. U. Y. Dodson, D. Hodgkin, and D. A. N. Mercola, "Structure of insulin in 4-zinc insulin," *Nature,* vol. 261, no. 5556, pp. 166-168, 1976.
[109] E. Ciszak, J. M. Beals, B. H. Frank, J. C. Baker, N. D. Carter, and G. D. Smith, "Role of C-terminal B-chain residues in insulin assembly: the structure of hexameric LysB28ProB29-human insulin," *Structure,* vol. 3, no. 6, pp. 615-622, 1995.
[110] E. Ciszak and G. D. Smith, "Crystallographic Evidence for Dual Coordination Around Zinc in the T3R3 Human Insulin Hexamer," *Biochemistry,* vol. 33, no. 6, pp. 1512-1517, 1994.
[111] M. A. Weiss, "The Structure and Function of Insulin: Decoding the TR Transition," *Vitamins and Hormones,* vol. 80, pp. 33-49, 2009.
[112] M. F. Dunn, "Zinc–Ligand Interactions Modulate Assembly and Stability of the Insulin Hexamer – A Review," *Biometals,* vol. 18, no. 4, pp. 295-303, 2005.
[113] S. Mukherjee, S. Mondal, A. A. Deshmukh, B. Gopal, and B. Bagchi, "What Gives an Insulin Hexamer Its Unique Shape and Stability? Role of Ten Confined

Water Molecules," *The Journal of Physical Chemistry B,* vol. 122, no. 5, pp. 1631-1637, 2018.
[114] M. A. Lemmon and J. Schlessinger, "Cell Signaling by Receptor Tyrosine Kinases," *Cell,* vol. 141, no. 7, pp. 1117-1134, 2010.
[115] S. R. Hubbard and W. T. Miller, "Closing in on a mechanism for activation," *eLife* vol. 3, p. e04909, 2014.
[116] R. Sarfstein and H. Werner, "Chap. 7 The INSR/IGF1R Receptor Family " in *Receptor Tyrosine Kinases: Family and Subfamilies*, D. L. Wheeler and Y. Yarden, Eds. 1st ed. ed.: Springer International Publishing, 2015, pp. 297-320.
[117] D. L. Wheeler, *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Disease* (Receptor Tyrosine Kinases). New York, NY: Springer New York, 2015.
[118] A. Belfiore, F. Frasca, G. Pandini, L. Sciacca, and R. Vigneri, "Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease," *Endocrine reviews,* vol. 30, no. 6, p. 586, 2009.
[119] A. A. Butler and D. Leroith, "Minireview: tissue-specific versus generalized gene targeting of the igf1 and igf1r genes and their roles in insulin-like growth factor physiology," *Endocrinology,* vol. 142, no. 5, p. 1685, 2001.
[120] C. W. Ward, M. C. Lawrence, V. A. Streltsov, T. E. Adams, and N. M. McKern, "The insulin and EGF receptor structures: new insights into ligand-induced receptor activation," *Trends in Biochemical Sciences,* vol. 32, no. 3, pp. 129-137, 2007.
[121] K. Louise *et al.*, "Agonism and antagonism at the insulin receptor," *PLoS ONE,* vol. 7, no. 12, p. e51972, 2012.
[122] A. Ullrich *et al.*, "Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes," *Nature,* vol. 313, no. 6005, pp. 756-761, 1985.
[123] Y. Ebina *et al.*, "Expression of a Functional Human Insulin Receptor from a Cloned cDNA in Chinese Hamster Ovary Cells," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 82, no. 23, pp. 8014-8018, 1985.
[124] C. R. Kahn and M. F. White, "The insulin receptor and the molecular mechanism of insulin action," *The Journal of clinical investigation,* vol. 82, no. 4, p. 1151, 1988.
[125] L. Knudsen, P. De Meyts, and Vladislav V. Kiselyov, "Insight into the molecular basis for the kinetic differences between the two insulin receptor isoforms," *Biochemical Journal,* vol. 440, no. 3, pp. 397-403, 2011.
[126] E. Harinda and E. Briony, "Ligand binding affinity at the insulin receptor isoform A (IR-A) and subsequent IR-A tyrosine phosphorylation kinetics are important determinants of mitogenic biological outcomes," *Frontiers in Endocrinology,* vol. 6, pp. 1-11, 2015.
[127] F. Frasca *et al.*, "Insulin Receptor Isoform A, a Newly Recognized, High-Affinity Insulin-Like Growth Factor II Receptor in Fetal and Cancer Cells," *Molecular and Cellular Biology,* vol. 19, no. 5, pp. 3278-3288, 1999.
[128] N. M. McKern *et al.*, "Structure of the insulin receptor ectodomain reveals a folded-over conformation," *Nature,* vol. 443, no. 7108, pp. 218-221, 2006.
[129] T. I. Croll *et al.*, "Higher-Resolution Structure of the Human Insulin Receptor Ectodomain: Multi-Modal Inclusion of the Insert Domain," *Structure (London, England : 1993),* vol. 24, no. 3, pp. 469-476, 2016.
[130] S. R. Hubbard, "Structural biology: Insulin meets its receptor," *Nature,* vol. 493, no. 7431, pp. 171-172, 2013.

[131] B. J. Smith *et al.*, "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,* vol. 107, no. 15, pp. 6771-6776, 2010.
[132] L. Whittaker, C. Hao, W. Fu, and J. Whittaker, "High-Affinity Insulin Binding: Insulin Interacts with Two Receptor Ligand Binding Sites," *Biochemistry,* vol. 47, no. 48, pp. 12900-12909, 2008.
[133] C. Hao, L. Whittaker, and J. Whittaker, "Characterization of a second ligand binding site of the insulin receptor," *Biochemical and Biophysical Research Communications,* vol. 347, no. 1, pp. 334-339, 2006.
[134] R. Roth and D. Cassell, "Insulin receptor: evidence that it is a protein kinase," *Science,* vol. 219, no. 4582, pp. 299-301, 1983.
[135] O. M. Rosen, R. Herrera, Y. Olowe, L. M. Petruzzelli, and M. H. Cobb, "Phosphorylation activates the insulin receptor tyrosine protein kinase," *Proceedings of the National Academy of Sciences,* vol. 80, no. 11, pp. 3237-3240, 1983.
[136] J. R. Flores-Riveros, E. Sibley, T. Kastelic, and M. D. Lane, "Substrate phosphorylation catalyzed by the insulin receptor tyrosine kinase. Kinetic correlation to autophosphorylation of specific sites in the beta subunit," *Journal of Biological Chemistry,* vol. 264, no. 36, pp. 21557-21572, 1989.
[137] H. E. Tornqvist and J. Avruch, "Relationship of site-specific beta subunit tyrosine autophosphorylation to insulin activation of the insulin receptor (tyrosine) protein kinase activity," *Journal of Biological Chemistry,* vol. 263, no. 10, pp. 4593-4601, 1988.
[138] S. R. Hubbard, "The Insulin Receptor: Both a Prototypical and Atypical Receptor Tyrosine Kinase," *Cold Spring Harbor Perspectives in Biology,* vol. 5, no. 3, 2013.
[139] P. De Meyts, J. Roth, D. M. Neville, J. R. Gavin, and M. A. Lesniak, "Insulin interactions with its receptors: Experimental evidence for negative cooperativity," *Biochemical and Biophysical Research Communications,* vol. 55, no. 1, pp. 154-161, 1973.
[140] P. D. Meyts, E. Van Obberghen, J. Roth, A. Wollmer, and D. Brandenburg, "Mapping of the residues responsible for the negative cooperativity of the receptor-binding region of insulin," *Nature,* vol. 273, no. 5663, pp. 504-509, 1978.
[141] V. V. Kiselyov, S. Versteyhe, L. Gauguin, and P. De Meyts, "Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation," *Molecular Systems Biology,* vol. 5, pp. 243-243, 2009.
[142] L. Ye *et al.*, "Structure and dynamics of the insulin receptor: implications for receptor activation and drug discovery," *Drug Discovery Today,* vol. 22, no. 7, pp. 1092-1102, 2017.
[143] N.-O. Yunn, J. Kim, Y. Kim, I. Leibiger, P.-O. Berggren, and S. H. Ryu, "Mechanistic understanding of insulin receptor modulation: Implications for the development of anti-diabetic drugs," *Pharmacology & Therapeutics,* vol. 185, pp. 86-98, 2018.
[144] P. De Meyts, "Insulin and IGF-I Receptor Structure and Binding Mechanism," in *Mechanisms of Insulin Action*, A. R. S. J. E. Pessi, Ed. New York, NY: Springer New York, 2007, pp. 1-32.
[145] T. Gutmann *et al.*, "Cryo-EM structure of the complete and ligand-saturated insulin receptor ectodomain," *bioRxiv,* p. 679233, 2019.
[146] S. Tamura, Y. Fujita-Yamaguchi, and J. Larner, "Insulin-like effect of trypsin on

the phosphorylation of rat adipocyte insulin receptor," *Journal of Biological Chemistry,* vol. 258, no. 24, pp. 14749-14752, 1983.

[147] S. E. Shoelson, M. F. White, and C. R. Kahn, "Tryptic activation of the insulin receptor. Proteolytic truncation of the alpha-subunit releases the beta-subunit from inhibitory control," *The Journal of biological chemistry,* vol. 263, no. 10, p. 4852, 1988.

[148] S. Terris and D. F. Steiner, "Retention and degradation of 125I-insulin by perfused livers from diabetic rats," *The Journal of Clinical Investigation,* vol. 57, no. 4, pp. 885-896, 1976.

[149] S. Terris, C. Hofmann, and D. F. Steiner, "Mode of uptake and degradation of 125I-labelled insulin by isolated hepatocytes and H4 hepatoma cells," *Canadian Journal of Biochemistry,* vol. 57, no. 6, pp. 459-468, 1979.

[150] J. Markussen, J. Halstrøm, F. C. Wiberg, and L. Schäffer, "Immobilized insulin for high capacity affinity chromatography of insulin receptors," *Journal of Biological Chemistry,* vol. 266, no. 28, pp. 18814-18818, 1991.

[151] J. Whittaker, P. Garcia, G. Q. Yu, and D. C. Mynarcik, "Transmembrane domain interactions are necessary for negative cooperativity of the insulin receptor," *Molecular Endocrinology,* vol. 8, no. 11, pp. 1521-1527, 1994.

[152] J. Brandt, A. S. Andersen, and C. Kristensen, "Dimeric Fragment of the Insulin Receptor α-Subunit Binds Insulin with Full Holoreceptor Affinity," *Journal of Biological Chemistry,* vol. 276, no. 15, pp. 12378-12384, 2001.

[153] J.-L. Carpentier, "Insulin receptor internalization: molecular mechanisms and physiopathological implications," *Diabetologia,* journal article vol. 37, no. 2, pp. S117-S124, 1994.

[154] R. A. Kohanski and M. D. Lane, "Kinetic evidence for activating and non-activating components of autophosphorylation of the insulin receptor protein kinase," *Biochemical and Biophysical Research Communications,* vol. 134, no. 3, pp. 1312-1318, 1986.

[155] M. Kasuga, Y. Fujita-Yamaguchi, D. L. Blithe, and C. R. Kahn, "Tyrosine-specific protein kinase activity is associated with the purified insulin receptor," *Proceedings of the National Academy of Sciences,* vol. 80, no. 8, pp. 2137-2141, 1983.

[156] H.-U. Häring, M. Kasuga, and C. R. Kahn, "Insulin receptor phosphorylation in intact adipocytes and in a cell-free system," *Biochemical and Biophysical Research Communications,* vol. 108, no. 4, pp. 1538-1548, 1982.

[157] M. F. White, H. U. Haring, M. Kasuga, and C. R. Kahn, "Kinetic properties and sites of autophosphorylation of the partially purified insulin receptor from hepatoma cells," *The Journal of biological chemistry,* vol. 259, no. 1, p. 255, 1984.

[158] S. Li, N. D. Covino, E. G. Stein, J. H. Till, and S. R. Hubbard, "Structural and Biochemical Evidence for an Autoinhibitory Role for Tyrosine 984 in the Juxtamembrane Region of the Insulin Receptor," *Journal of Biological Chemistry,* vol. 278, no. 28, pp. 26007-26014, 2003.

[159] P. Ahorukomeye *et al.*, "Fish-hunting cone snail venoms are a rich source of minimized ligands of the vertebrate insulin receptor," *eLife,* vol. 8, p. e41574, 2019.

[160] L. Wei, S. R. Hubbard, W. A. Hendrickson, and L. Ellis, "Expression, Characterization, and Crystallization of the Catalytic Core of the Human Insulin Receptor Protein-tyrosine Kinase Domain," *Journal of Biological Chemistry,* vol. 270, no. 14, pp. 8122-8130, 1995.

[161] J. M. Tavaré, R. M. O'Brien, K. Siddle, and R. M. Denton, "Analysis of insulin-receptor phosphorylation sites in intact cells by two-dimensional phosphopeptide mapping," *The Biochemical journal,* vol. 253, no. 3, p. 783, 1988.
[162] J. M. Tavaré and R. M. Denton, "Studies on the autophosphorylation of the insulin receptor from human placenta. Analysis of the sites phosphorylated by two-dimensional peptide mapping," *Biochemical Journal,* vol. 252, no. 2, pp. 607-615, 1988.
[163] M. Z. Cabail, S. Li, E. Lemmon, M. E. Bowen, S. R. Hubbard, and W. T. Miller, "The insulin and IGF1 receptor kinase domains are functional dimers in the activated state," *Nature Communications,* vol. 6, p. 6406, 2015.
[164] J. M. Kavran *et al.*, "How IGF-1 activates its receptor," *eLife,* vol. 3, 2014.
[165] T. Gutmann, K. H. Kim, M. Grzybek, T. Walz, and Ü. Coskun, "Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor," *The Journal of Cell Biology,* 2018.
[166] Q. Li, Y. L. Wong, and C. Kang, "Solution structure of the transmembrane domain of the insulin receptor in detergent micelles," *Biochimica et Biophysica Acta (BBA) - Biomembranes,* vol. 1838, no. 5, pp. 1313-1321, 2014.
[167] V. Baron, P. Kaliman, N. Gautier, and E. Van Obberghen, "The insulin receptor activation process involves localized conformational changes," *Journal of Biological Chemistry,* vol. 267, no. 32, pp. 23290-4, 1992.
[168] V. Baron *et al.*, "Insulin binding to its receptor induces a conformational change in the receptor C-terminus," *Biochemistry,* vol. 29, no. 19, pp. 4634-4641, 1990.
[169] J. M. Backer *et al.*, "The insulin receptor juxtamembrane region contains two independent tyrosine/beta-turn internalization signals," *The Journal of Cell Biology,* vol. 118, no. 4, pp. 831-839, 1992.
[170] B. Zhang, J. M. Tavaré, L. Ellis, and R. A. Roth, "The regulatory role of known tyrosine autophosphorylation sites of the insulin receptor kinase domain. An assessment by replacement with neutral and negatively charged amino acids," *Journal of Biological Chemistry,* vol. 266, no. 2, pp. 990-996, 1991.
[171] M. Dickens and J. M. Tavaré, "Analysis of the order of autophosphorylation of human insulin receptor tyrosines 1158, 1162 and 1163," *Biochemical and Biophysical Research Communications,* vol. 186, no. 1, pp. 244-250, 1992.
[172] M. F. White, S. E. Shoelson, H. Keutmann, and C. R. Kahn, "A cascade of tyrosine autophosphorylation in the beta-subunit activates the phosphotransferase of the insulin receptor," *The Journal of biological chemistry,* vol. 263, no. 6, p. 2969, 1988.
[173] C. K. Chou *et al.*, "Human insulin receptors mutated at the ATP-binding site lack protein tyrosine kinase activity and fail to mediate postreceptor effects of insulin," *Journal of Biological Chemistry,* vol. 262, no. 4, pp. 1842-7, 1987.
[174] J. Lee, T. O'Hare, P. F. Pilch, and S. E. Shoelson, "Insulin receptor autophosphorylation occurs asymmetrically," *Journal of Biological Chemistry,* vol. 268, no. 6, pp. 4092-8, 1993.
[175] A. L. Frattali, J. L. Treadway, and J. E. Pessin, "Transmembrane signaling by the human insulin receptor kinase. Relationship between intramolecular beta subunit trans- and cis-autophosphorylation and substrate kinase activation," *Journal of Biological Chemistry,* vol. 267, no. 27, pp. 19521-8, 1992.
[176] J. L. Treadway *et al.*, "Transdominant inhibition of tyrosine kinase activity in mutant insulin/insulin-like growth factor I hybrid receptors," *Proceedings of the National Academy of Sciences,* vol. 88, no. 1, pp. 214-218, 1991.
[177] T. O'Hare and P. F. Pilch, "Separation and characterization of three insulin

receptor species that differ in subunit composition," *Biochemistry,* vol. 27, no. 15, pp. 5693-5700, 1988.
[178] M. Böni-Schnetzler, J. B. Rubin, and P. F. Pilch, "Structural requirements for the transmembrane activation of the insulin receptor kinase," *Journal of Biological Chemistry,* vol. 261, no. 32, pp. 15281-7, 1986.
[179] S. R. Hubbard, "Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog," *The EMBO Journal,* vol. 16, no. 18, pp. 5572-5581, 1997.
[180] R. H. Stevan, W. Lei, and A. H. Wayne, "Crystal structure of the tyrosine kinase domain of the human insulin receptor," *Nature,* vol. 372, no. 6508, p. 746, 1994.
[181] A. E. Whitten *et al.*, "Solution Structure of Ectodomains of the Insulin Receptor Family: The Ectodomain of the Type 1 Insulin-Like Growth Factor Receptor Displays Asymmetry of Ligand Binding Accompanied by Limited Conformational Change," *Journal of Molecular Biology,* vol. 394, no. 5, pp. 878-892, 2009.
[182] P. D. Meyts, "The Insulin Receptor and Its Signal Transduction Network " *National Center for Biotechnology Information* 2016.
[183] C. W. Ward and M. C. Lawrence, "Landmarks in Insulin Research," *Frontiers in Endocrinology,* vol. 2, p. 76, 2011.
[184] R. A. Pullen *et al.*, "Receptor-binding region of insulin," *Nature,* vol. 259, no. 5542, pp. 369-373, 1976.
[185] T. Glendorf, A. R. Sørensen, E. Nishimura, I. Pettersson, and T. Kjeldsen, "Importance of the Solvent-Exposed Residues of the Insulin B Chain α-Helix for Receptor Binding," *Biochemistry,* vol. 47, no. 16, pp. 4743-4751, 2008.
[186] J. P. Mayer, F. Zhang, and R. D. DiMarchi, "Insulin structure and function," *Peptide Science,* vol. 88, no. 5, pp. 687-713, 2007.
[187] L. Schäffer, "A model for insulin binding to the insulin receptor," *European Journal of Biochemistry,* vol. 221, no. 3, pp. 1127-1132, 1994.
[188] S. Gammeltoft and J. Gliemann, "Degradation, receptor binding affinity, and potency of insulin from the Atlantic hagfish (Myxine glutinosa) determined in isolated rat fat cells," *Journal of Biological Chemistry,* vol. 252, no. 2, pp. 602-608, 1977.
[189] S. O. Emdin, O. Sonne, and J. Gliemann, "Hagfish Insulin: The Discrepancy Between Binding Affinity and Biologic Activity," *Diabetes,* vol. 29, no. 4, pp. 301-303, 1980.
[190] R. Horuk, T. L. Blundell, N. R. Lazarus, R. W. J. Neville, D. Stone, and A. Wollmer, "A monomeric insulin from the porcupine (Hystrix cristata), an Old World hystricomorph," *Nature,* vol. 286, no. 5775, pp. 822-824, 1980.
[191] M. Bajaj *et al.*, "Coypu insulin. Primary structure, conformation and biological properties of a hystricomorph rodent insulin," *Biochemical Journal,* vol. 238, no. 2, pp. 345-351, 1986.
[192] M. Jensen, "Analysis of structure-function relationships of the insulin molecule by alanine scanning mutagenesis," M. Sc. Thesis, The Receptor Systems Biology Laboratory Hagedorn Research Institute, Copenhagen University, 2000.
[193] S. H. Nakagawa and H. S. Tager, "Implications of invariant residue LeuB6 in insulin-receptor interactions," *Journal of Biological Chemistry,* vol. 266, no. 18, pp. 11502-11509, 1991.
[194] G. Schwartz and P. G. Katsoyannis, "Synthesis of des(tetrapeptide B1-4) and des(pentapeptide B1-5) human insulin. Two biologically active analogs," *Biochemistry,* vol. 17, no. 21, pp. 4550-4556, 1978.
[195] K. Huang *et al.*, "How Insulin Binds: the B-Chain α-Helix Contacts the L1 β-

Helix of the Insulin Receptor," *Journal of Molecular Biology,* vol. 341, no. 2, pp. 529-550, 2004.

[196] U. Derewenda, Z. Derewenda, E. J. Dodson, G. G. Dodson, X. Bing, and J. Markussen, "X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule: A completely inactive analogue," *Journal of Molecular Biology,* vol. 220, no. 2, pp. 425-433, 1991.

[197] S. H. Nakagawa and H. S. Tager, "Perturbation of insulin-receptor interactions by intramolecular hormone cross-linking. Analysis of relative movement among residues A1, B1, and B29," *Journal of Biological Chemistry,* vol. 264, no. 1, pp. 272-279, 1989.

[198] T. Kurose *et al.*, "Cross-linking of a B25 azidophenylalanine insulin derivative to the carboxyl-terminal region of the alpha-subunit of the insulin receptor. Identification of a new insulin-binding domain in the insulin receptor," *Journal of Biological Chemistry,* vol. 269, no. 46, pp. 29190-29197, 1994.

[199] B. Xu *et al.*, "Diabetes-Associated Mutations in Insulin: Consecutive Residues in the B Chain Contact Distinct Domains of the Insulin Receptor," *Biochemistry,* vol. 43, no. 26, pp. 8356-8372, 2004.

[200] P. A. Hoyne *et al.*, "High affinity insulin binding by soluble insulin receptor extracellular domain fused to a leucine zipper," *FEBS Letters,* vol. 479, no. 1, pp. 15-18, 2000.

[201] G. S. Viereck, "What Life Means to Einstein. An interview by George Sylvester Viereck," *The Saturday Evening Post,* pp. 17, 110, 113-114, 117, 1929.

[202] R. P. Feynman, R. B. Leighton, and M. L. Sands, "Chap. 3, The relation of physics to other sciences," in *The Feynman lectures on physics* vol. 1: Reading, Mass : Addison-Wesley Pub. Co., 1963, pp. 3-10.

[203] L. Chen, "The myth of 0.9% saline: neither normal nor physiological," *Critical Care Nursing Quarterly,* vol. 38, no. 4, pp. 385-389, 2015.

[204] H. Li, S.-r. Sun, J. Q. Yap, J.-h. Chen, and Q. Qian, "0.9% saline is neither normal nor physiological," *Journal of Zhejiang University-Science B,* vol. 17, no. 3, pp. 181-187, 2016.

[205] S. Donnini, F. Tegeler, G. Groenhof, and H. Grubmüller, "Constant pH Molecular Dynamics in Explicit Solvent with λ-Dynamics," *Journal of Chemical Theory and Computation,* vol. 7, no. 6, pp. 1962-1978, 2011.

[206] C. N. Pace, G. R. Grimsley, and J. M. Scholtz, "Protein Ionizable Groups: pK Values and Their Contribution to Protein Stability and Solubility," *Journal of Biological Chemistry,* vol. 284, no. 20, pp. 13285-13289, 2009.

[207] C. Bryant *et al.*, "Acid stabilization of insulin," *Biochemistry,* vol. 32, no. 32, pp. 8075-8082, 1993.

[208] J. Haas *et al.*, "Primary Steps of pH-Dependent Insulin Aggregation Kinetics are Governed by Conformational Flexibility," *ChemBioChem,* vol. 10, no. 11, pp. 1816-1822, 2009.

[209] N. C. Kaarsholm, S. Havelund, and P. Hougaard, "Ionization behavior of native and mutant insulins: pK perturbation of B13-Glu in aggregated species," *Archives of Biochemistry and Biophysics,* vol. 283, no. 2, pp. 496-502, 1990.

[210] M. D. Soerensen and J. J. Led, "Structural Details of Asp(B9) Human Insulin at Low pH from Two-Dimensional NMR Titration Studies," *Biochemistry,* vol. 33, no. 46, pp. 13727-13733, 1994.

[211] D. v. d. S. M.J. Abraham, E. Lindahl, B. Hess, and the GROMACS development team, "Chap 1. Introduction," in *GROMACS User Manual version 5.0.4* www.gromacs.org, 2014, pp. 1-6.

[212] M. Levitt and A. Warshel, "Computer simulation of protein folding," *Nature,* vol. 253, p. 694, 1975.
[213] A. Warshel, "Bicycle-pedal model for the first step in the vision process," *Nature,* vol. 260, p. 679, 1976.
[214] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature,* vol. 267, p. 585, 1977.
[215] A. Hazel, C. Chipot, and J. C. Gumbart, "Thermodynamics of Deca-alanine Folding in Water," *Journal of Chemical Theory and Computation,* vol. 10, no. 7, pp. 2836-2844, 2014.
[216] R. Chen and A. Mark, "The effect of membrane curvature on the conformation of antimicrobial peptides: implications for binding and the mechanism of action," (in English), *European Biophysics Journal,* vol. 40, no. 4, pp. 545-553, 2011.
[217] C. Kandt, W. L. Ash, and D. Peter Tieleman, "Setting up and running molecular dynamics simulations of membrane proteins," *Methods,* vol. 41, no. 4, pp. 475-488, 2007.
[218] S. A. Showalter and R. Brüschweiler, "Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field," *Journal of Chemical Theory and Computation,* vol. 3, no. 3, pp. 961-975, 2007.
[219] O. F. Lange, D. van der Spoel, and B. L. de Groot, "Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data," *Biophysical Journal,* vol. 99, no. 2, pp. 647-655, 2010.
[220] B. Zagrovic and W. F. van Gunsteren, "Comparing atomistic simulation data with the NMR experiment: How much can NOEs actually tell us?," *Proteins: Structure, Function, and Bioinformatics,* vol. 63, no. 1, pp. 210-218, 2006.
[221] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, "Molecular dynamics simulations: advances and applications," *Advances and Applications in Bioinformatics and Chemistry : AABC,* vol. 8, pp. 37-47, 2015.
[222] C. Mura and C. E. McAnany, "An introduction to biomolecular simulations and docking," *Molecular Simulation,* vol. 40, no. 10-11, pp. 732-764, 2014.
[223] D. Vlachakis, E. Bencurova, N. Papangelopoulos, and S. Kossida, "Chap. Seven, Current State-of-the-Art Molecular Dynamics Methods and Applications," in *Advances in Protein Chemistry and Structural Biology*, vol. 94, R. Donev, Ed.: Academic Press, 2014, pp. 269-313.
[224] S. Genheden, C. Diehl, M. Akke, and U. Ryde, "Starting-Condition Dependence of Order Parameters Derived from Molecular Dynamics Simulations," *Journal of Chemical Theory and Computation,* vol. 6, no. 7, pp. 2176-2190, 2010.
[225] H. D. A., B. R. Andrew, Z. Weixing, S. T. E., and G. W. H., "Dynamics of the Hck-SH3 domain: Comparison of experiment with multiple molecular dynamics simulations," *Protein Science,* vol. 9, no. 1, pp. 95-103, 2000.
[226] A. N. Koller, H. Schwalbe, and H. Gohlke, "Starting Structure Dependence of NMR Order Parameters Derived from MD Simulations: Implications for Judging Force-Field Quality," *Biophysical Journal,* vol. 95, no. 1, pp. L04-L06, 2008.
[227] P. r. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, "Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models," *Journal of Chemical Theory and Computation,* vol. 6, no. 2, pp. 459-466, 2010.
[228] A. D. Mackerell, M. Feig, and C. L. Brooks, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum

mechanics in reproducing protein conformational distributions in molecular dynamics simulations," *Journal of Computational Chemistry,* vol. 25, no. 11, pp. 1400-1415, 2004.

[229] A. D. MacKerell *et al.*, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," *The Journal of Physical Chemistry B,* vol. 102, no. 18, pp. 3586-3616, 1998.

[230] R. B. Best *et al.*, "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles," *Journal of Chemical Theory and Computation,* vol. 8, no. 9, pp. 3257-3273, 2012.

[231] "Note on CHARMM Gromacs," *http://www.gromacs.org*, October 2014.

[232] J. Lemkul, "Gromacs Tutorials," *http://www.mdtutorials.com/gmx/* 2015.

[233] N. Todorova, F. S. Legge, H. Treutlein, and I. Yarovsky, "Systematic Comparison of Empirical Forcefields for Molecular Dynamic Simulation of Insulin," *The Journal of Physical Chemistry B,* vol. 112, no. 35, pp. 11137-11146, 2008.

[234] V. Zoete, M. Meuwly, and M. Karplus, "Investigation of glucose binding sites on insulin," *Proteins: Structure, Function, and Bioinformatics,* vol. 55, no. 3, pp. 568-581, 2004.

[235] H. M. Berman *et al.*, "The Protein Data Bank," *Nucleic Acids Research,* vol. 28, no. 1, pp. 235-242, 2000.

[236] A. Šali and T. L. Blundell, "Comparative Protein Modelling by Satisfaction of Spatial Restraints," *Journal of Molecular Biology,* vol. 234, no. 3, pp. 779-815, 1993.

[237] A. Fiser *et al.*, "Tools for comparative protein structure modeling and analysis," *Nucleic Acids Research,* vol. 31, no. 13, pp. 3375-3380, 2003.

[238] J. Huang and A. D. MacKerell, "CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data," *Journal of Computational Chemistry,* vol. 34, no. 25, pp. 2135-2145, 2013.

[239] J. Hsin, A. Arkhipov, Y. Yin, J. E. Stone, and K. Schulten, "Using VMD: An Introductory Tutorial," in *Current Protocols in Bioinformatics*: John Wiley & Sons, Inc., 2002.

[240] I. Kufareva and R. Abagyan, "Methods of protein structure comparison," *Methods in molecular biology (Clifton, N.J.),* vol. 857, pp. 231-257, 2012.

[241] C. Oliviero and P. Sándor, "A normalized root-mean-square distance for comparing protein three-dimensional structures," *Protein Science,* vol. 10, no. 7, pp. 1470-1473, 2001.

[242] A. Kuzmanic and B. Zagrovic, "Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors," *Biophysical Journal,* vol. 98, no. 5, pp. 861-871, 2010.

[243] G. N. Phillips, Jr., "Comparison of the dynamics of myoglobin in different crystal forms," *Biophysical journal,* vol. 57, no. 2, pp. 381-383, 1990.

[244] B. Halle, "Flexibility and packing in proteins," *Proceedings of the National Academy of Sciences,* vol. 99, no. 3, pp. 1274-1279, 2002.

[245] L. Meinhold and J. C. Smith, "Fluctuations and Correlations in Crystalline Protein Dynamics: A Simulation Analysis of Staphylococcal Nuclease," *Biophysical Journal,* vol. 88, no. 4, pp. 2554-2563, 2005.

[246] W. C. Lu, C. Z. Wang, E. W. Yu, and K. M. Ho, "Dynamics of the trimeric AcrB transporter protein inferred from a B-factor analysis of the crystal structure," *Proteins: Structure, Function, and Bioinformatics,* vol. 62, no. 1, pp. 152-158, 2006.

[247] N. Glykos, "On the application of molecular-dynamics simulations to validate thermal parameters and to optimize TLS-group selection for macromolecular refinement," *Acta Crystallographica Section D,* vol. 63, no. 6, pp. 705-713, 2007.
[248] L.-W. Yang, E. Eyal, C. Chennubhotla, J. Jee, A. M. Gronenborn, and I. Bahar, "Insights into Equilibrium Dynamics of Proteins from Comparison of NMR and X-Ray Data with Computational Predictions," *Structure,* vol. 15, no. 6, pp. 741-749, 2007.
[249] C.-H. Lu *et al.*, "On the relationship between the protein structure and protein dynamics," *Proteins,* vol. 72, no. 2, p. 625, 2008.
[250] D.-W. Li and R. Brüschweiler, "All-Atom Contact Model for Understanding Protein Dynamics from Crystallographic B-Factors," *Biophysical Journal,* vol. 96, no. 8, pp. 3074-3081, 2009.
[251] Z. Hu and J. Jiang, "Assessment of biomolecular force fields for molecular dynamics simulations in a protein crystal," *Journal of Computational Chemistry,* vol. 31, no. 2, p. 371, 2010.
[252] D. C. Lay, *Linear Algebra and Its Applications*. Pearson Education, 2006, p. 576.
[253] "Hydrogen Bonds in Proteins: Role and Strength," *Encyclopedia of Life Sciences (ELS),* pp. 1-7, 2010.
[254] I. K. McDonald and J. M. Thornton, "Satisfying Hydrogen Bonding Potential in Proteins," *Journal of Molecular Biology,* vol. 238, no. 5, pp. 777-793, 1994.
[255] S. Neal, M. Berjanskii, H. Zhang, and D. S. Wishart, "Accurate prediction of protein torsion angles using chemical shifts and sequence homology," *Magnetic Resonance in Chemistry,* vol. 44, no. S1, pp. S158-S167, 2006.
[256] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. Marius Clore, "The Xplor-NIH NMR molecular structure determination package," *Journal of Magnetic Resonance,* vol. 160, no. 1, pp. 65-73, 2003.
[257] G. A. Bermejo and C. D. Schwieters, "Protein Structure Elucidation from NMR Data with the Program Xplor-NIH," in *Protein NMR: Methods and Protocols*, R. Ghose, Ed. New York, NY: Springer New York, 2018, pp. 311-340.
[258] T. A. Soares, X. Daura, C. Oostenbrink, L. J. Smith, and W. F. Gunsteren, "Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme," *Journal of Biomolecular NMR,* vol. 30, no. 4, pp. 407-422.
[259] X. Daura, I. Antes, W. F. van Gunsteren, W. Thiel, and A. E. Mark, "The effect of motional averaging on the calculation of NMR-derived structural properties," *Proteins: Structure, Function, and Bioinformatics,* vol. 36, no. 4, pp. 542-555, 1999.
[260] G. K. C. Roberts, Ed. *NMR of macromolecules: a practical approach*. Oxford: IRL Press, 1993.
[261] C. D. Schwieters, J. J. Kuszewski, and G. Marius Clore, "Using Xplor–NIH for NMR molecular structure determination," *Progress in Nuclear Magnetic Resonance Spectroscopy,* vol. 48, no. 1, pp. 47-62, 2006.
[262] E. K. Hohne, G., "New interpretation of helical structures in polypeptides," *Stuida Biophysica,* vol. 87, pp. 23-28, 1982.
[263] C. G. Frankaer, M. V. Knudsen, K. Noren, E. Nazarenko, K. Stahl, and P. Harris, "The structures of T6, T3R3 and R6 bovine insulin: combining X-ray diffraction and absorption spectroscopy," *Acta Crystallographica Section D,* vol. 68, no. 10, pp. 1259-1271, 2012.
[264] O. Carugo, "How large B-factors can be in protein crystal structures," *BMC bioinformatics,* vol. 19, no. 1, pp. 61-61, 2018.

[265] A. Wlodawer, M. Li, and Z. Dauter, "High-Resolution Cryo-EM Maps and Models: A Crystallographer's Perspective," *Structure,* vol. 25, no. 10, pp. 1589-1597.e1, 2017.
[266] C. N. Woldin, F. S. Hing, J. Lee, P. F. Pilch, and G. G. Shipley, "Structural Studies of the Detergent-solubilized and Vesicle-reconstituted Insulin Receptor," *Journal of Biological Chemistry,* vol. 274, no. 49, pp. 34981-34992, 1999.
[267] K. Christiansen, J. Tranum-Jensen, J. Carlsen, and J. Vinten, "A Model for the Quaternary Structure of Human Placental Insulin Receptor Deduced from Electron Microscopy," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 88, no. 1, pp. 249-252, 1991.
[268] A. J. Saunders, G. B. Young, and G. J. Pielak, "Polarity of disulfide bonds," *Protein Science,* vol. 2, no. 7, pp. 1183-1184, 1993.
[269] G. B. McGaughey, M. Gagné, and A. K. Rappé, "π-Stacking Interactions: alive and well in proteins," *Journal of Biological Chemistry,* vol. 273, no. 25, pp. 15458-15463, 1998.
[270] T. F. Havel and M. E. Snow, "A new method for building protein conformations from sequence alignments with homologues of known structure," *Journal of Molecular Biology,* vol. 217, no. 1, pp. 1-7, 1991.
[271] A. T. Brünger, *X-PLOR: Version 3.1 : a System for X-ray Crystallography and NMR*. Yale University Press, 1992.
[272] T. F. Havel, "An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance," *Progress in Biophysics and Molecular Biology,* vol. 56, no. 1, pp. 43-78, 1991.
[273] Q.-X. Hua *et al.*, "Mini-proinsulin and mini-IGF-I: homologous protein sequences encoding non-homologous structures1," *Journal of Molecular Biology* vol. 277, no. 1, pp. 103-118, 1998.
[274] Q. X. Hua and M. A. Weiss, "Toward the solution structure of human insulin: sequential 2D proton NMR assignment of a des-pentapeptide analog and comparison with crystal structure," *Biochemistry,* vol. 29, no. 46, pp. 10545-10555, 1990.
[275] M. A. Weiss *et al.*, "Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: assignment of aromatic resonances with application to protein folding, structure, and dynamics," *Biochemistry,* vol. 28, no. 25, pp. 9855-9873, 1989.
[276] Q. X. Hua, W. Jia, B. H. Frank, and M. A. Weiss, "Comparison of the Dynamics of an Engineered Insulin Monomer and Dimer by Acid-Quenched Amide Proton Exchange," *Journal of Molecular Biology,* vol. 230, no. 2, pp. 387-394, 1993.
[277] J. F. Doreleijers, M. L. Raves, T. Rullmann, and R. Kaptein, "Completeness of NOEs in protein structures: A statistical analysis of NMR data," *Journal of Biomolecular NMR,* journal article vol. 14, no. 2, pp. 123-132.
[278] O. Jardetzky, G. C. K. Roberts, Ed. *NMR in molecular biology*. New York : Academic Press, 1981.
[279] W. K., *NMR of proteins and nucleic acids*. New York: John Wiley & Sons, 1986.
[280] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering,* vol. 9, no. 3, pp. 90-95, 2007.
[281] C. J. McKnight, P. T. Matsudaira, and P. S. Kim, "NMR structure of the 35-residue villin headpiece subdomain," *Nature Structural Biology,* vol. 4, pp. 180-184, 1997.
[282] H. Schwalbe *et al.*, "A refined solution structure of hen lysozyme determined using residual dipolar coupling data," *Protein Science,* vol. 10, no. 4, pp. 677-688,

2001.
[283] R. Bürgi, J. Pitera, and W. F. van Gunsteren, "Assessing the effect of conformational averaging on the measured values of observables," *Journal of Biomolecular NMR,* vol. 19, no. 4, pp. 305-320, 2001.
[284] R. Nussinov, C.-J. Tsai, A. Shehu, and H. Jang, "Computational Structural Biology: Successes, Future Directions, and Challenges," *Molecules (Basel, Switzerland),* vol. 24, no. 3, 2019.
[285] G. R. Fleming, G. D. Scholes, and Y.-C. Cheng, "Quantum effects in biology," *Procedia Chemistry,* vol. 3, no. 1, pp. 38-57, 2011.
[286] S. F. Huelga and M. B. Plenio, "Quantum dynamics of bio-molecular systems in noisy environments," *Procedia Chemistry,* vol. 3, no. 1, pp. 248-257, 2011.
[287] F. Gatti, *Molecular Quantum Dynamics From Theory to Applications*. Berlin/Heidelberg : Springer, 2014.
[288] D. Abbott, P. C. W. Davies, A. K. Pati, and S. R. Penrose, *Quantum Aspects of Life*. Imperial College Press, 2008.
[289] T. Lazaridis and G. Hummer, "Classical Molecular Dynamics with Mobile Protons," *Journal of Chemical Information and Modeling,* vol. 57, no. 11, pp. 2833-2845, 2017.
[290] S. Grimme and P. R. Schreiner, "Computational Chemistry: The Fate of Current Methods and Future Challenges," *Angewandte Chemie International Edition,* vol. 57, no. 16, pp. 4170-4176, 2018.
[291] U. Ryde, "Chapter Six - QM/MM Calculations on Proteins," in *Methods in Enzymology*, vol. 577, G. A. Voth, Ed.: Academic Press, 2016, pp. 119-158.
[292] J. McClory, G.-X. Hu, J.-W. Zou, D. J. Timson, and M. Huang, "Phosphorylation Mechanism of N-Acetyl-l-glutamate Kinase, a QM/MM Study," *The Journal of Physical Chemistry B,* vol. 123, no. 13, pp. 2844-2852, 2019.
[293] E. Bellomo, A. Abro, C. Hogstrand, W. Maret, and C. Domene, "Role of Zinc and Magnesium Ions in the Modulation of Phosphoryl Transfer in Protein Tyrosine Phosphatase 1B," *Journal of the American Chemical Society,* vol. 140, no. 12, p. 4446, 2018.
[294] J. McClory, D. J. Timson, W. Singh, J. Zhang, and M. Huang, "Reaction Mechanism of Isopentenyl Phosphate Kinase: A QM/MM Study," *The Journal of Physical Chemistry B,* vol. 121, no. 49, pp. 11062-11071, 2017.
[295] A. Pérez-Gallegos, M. Garcia-Viloca, À. González-Lafont, and J. M. Lluch, "SP20 Phosphorylation Reaction Catalyzed by Protein Kinase A: QM/MM Calculations Based on Recently Determined Crystallographic Structures," *ACS Catalysis,* vol. 5, no. 8, pp. 4897-4912, 2015.
[296] P. Ojeda-May, Y. Li, V. Ovchinnikov, and K. Nam, "Role of Protein Dynamics in Allosteric Control of the Catalytic Phosphoryl Transfer of Insulin Receptor Kinase," *Journal of the American Chemical Society,* vol. 137, no. 39, p. 12454, 2015.
[297] L. Yu, L. Xu, M. Xu, B. Wan, L. Yu, and Q. Huang, "Role of Mg2+ ions in protein kinase phosphorylation: insights from molecular dynamics simulations of ATP-kinase complexes," *Molecular Simulation,* vol. 37, no. 14, pp. 1143-1150, 2011.
[298] D. v. d. S. M.J. Abraham, E. Lindahl, B. Hess, and the GROMACS development team, "Chap. 3 Algorithms," in *GROMACS User Manual version 5.0.4* www.gromacs.org, 2014, pp. 11-66.
[299] D. B. Kony, P. H. Hünenberger, and W. F. van Gunsteren, "Molecular dynamics simulations of the native and partially folded states of ubiquitin: Influence of methanol cosolvent, pH, and temperature on the protein structure and dynamics,"

Protein Science, vol. 16, no. 6, pp. 1101-1118, 2007.
[300] J. A. Wallace and J. K. Shen, "Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange," *Journal of Chemical Theory and Computation,* vol. 7, no. 8, pp. 2617-2629, 2011.
[301] H. Dominguez, "Molecular dynamics simulations to study the solvent influence on protein structure," *Chemical Physics Letters,* vol. 651, pp. 92-96, 2016.
[302] H. Zhang, C. Yin, Y. Jiang, and D. van der Spoel, "Force Field Benchmark of Amino Acids: I. Hydration and Diffusion in Different Water Models," *Journal of Chemical Information and Modeling,* vol. 58, no. 5, pp. 1037-1052, 2018.
[303] D. van Der Spoel, P. J. van Maaren, and H. J. C. Berendsen, "A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field," *The Journal of Chemical Physics,* vol. 108, no. 24, pp. 10220-10230, 1998.
[304] D. van der Spoel and P. J. van Maaren, "The Origin of Layer Structure Artifacts in Simulations of Liquid Water," *Journal of Chemical Theory and Computation,* vol. 2, no. 1, pp. 1-11, 2006.
[305] J. S. Hub, C. Caleman, and D. Van Der Spoel, "Organic molecules on the surface of water droplets an energetic perspective," *Phys. Chem. Chem. Phys.,* vol. 14, no. 27, pp. 9537-9545, 2012.
[306] N. Bernstein *et al.*, "QM/MM simulation of liquid water with an adaptive quantum region," *Phys. Chem. Chem. Phys.,* vol. 14, no. 2, pp. 646-656, 2011.
[307] D. Xenides, B. R. Randolf, and B. M. Rode, "Hydrogen bonding in liquid water: An ab initio QM/MM MD simulation study," *Journal of Molecular Liquids,* vol. 123, no. 2, pp. 61-67, 2006.
[308] T. Urbic, "Ions increase strength of hydrogen bond in water," *Chemical Physics Letters,* vol. 610-611, pp. 159-162, 2014.
[309] S. Pezeshki and H. Lin, "Molecular dynamics simulations of ion solvation by flexible-boundary QM/MM: On-the-fly partial charge transfer between QM and MM subsystems," *Journal of Computational Chemistry,* vol. 35, no. 24, pp. 1778-1788, 2014.
[310] S. Riahi, B. Roux, and C. N. Rowley, "QM/MM molecular dynamics simulations of the hydration of Mg(II) and Zn(II) ions.(Report)," *Canadian Journal of Chemistry,* vol. 91, no. 7, p. 552, 2013.
[311] L. E. Ratcliff, S. Mohr, G. Huhs, T. Deutsch, M. Masella, and L. Genovese, "Challenges in Large Scale Quantum Mechanical Calculations," vol. 7, ed, 2016.
[312] Y. Zhang and M. F. Sanner, "AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes," *Bioinformatics,* 2019.
[313] M. Kurcinski, M. Jamroz, M. Blaszczyk, A. Kolinski, and S. Kmiecik, "CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site," *Nucleic Acids Research,* vol. 43, no. W1, pp. W419-W424, 2015.
[314] J. Lee, M. Miyazaki, G. R. Romeo, and S. E. Shoelson, "Insulin Receptor Activation with Transmembrane Domain Ligands," *Journal of Biological Chemistry,* vol. 289, no. 28, pp. 19769-19777, 2014.
[315] H. Vashisth and C. F. Abrams, "Docking of insulin to a structurally equilibrated insulin receptor ectodomain," *Proteins: Structure, Function, and Bioinformatics,* vol. 78, no. 6, pp. 1531-1543, 2010.
[316] H. Vashisth and C. Abrams, "Docking of Insulin to its Receptor," *Biophysical Journal,* vol. 96, no. 3, Supplement 1, p. 673a, 2009.
[317] H. Vashisth, L. Maragliano, and Cameron f. Abrams, ""DFG-Flip" in the Insulin

Receptor Kinase Is Facilitated by a Helical Intermediate State of the Activation Loop," *Biophysical Journal,* vol. 102, no. 8, pp. 1979-1987, 2012.
[318] H. Vashisth, "Theoretical and Computational Studies of Peptides and Receptors of the Insulin Family," *Membranes,* vol. 5, no. 1, p. 48, 2015.
[319] H. Vashisth and H. Mohammadiarani, "All-Atom Structural Models of the Transmembrane Domains of Insulin Receptor and Type-1 Insulin-Like Growth Factor Receptor," *Biophysical Journal,* vol. 110, no. 3, Supplement 1, p. 58a, 2016.
[320] H. Vashisth, "Flexibility in the Insulin Receptor Ectodomain Enables Docking of Insulin in Crystallographic Conformation Observed in a Hormone-Bound Microreceptor," *Membranes,* vol. 4, no. 4, pp. 730-746, 2014.
[321] Rune T. Kidmose and Gregers R. Andersen, "Interacting with the Human Insulin Receptor," *Structure,* vol. 24, no. 3, pp. 351-352, 2016.
[322] T. Anastassiadis *et al.*, "A Highly Selective Dual Insulin Receptor (IR)/Insulin-like Growth Factor 1 Receptor (IGF-1R) Inhibitor Derived from an Extracellular Signal-regulated Kinase (ERK) Inhibitor," *Journal of Biological Chemistry,* vol. 288, no. 39, pp. 28068-28077, 2013.
[323] B. Wheeler, *Tcl/Tk 8.5 Programming Cookbook*. Packt Publishing, 2011, p. 236.
[324] A. P. Nadkarni, *The Tcl Programming Language: A Comprehensive Guide*. CreateSpace Independent Publishing Platform, 2017, p. 668.
[325] F. Perez and B. E. Granger, "IPython: A System for Interactive Scientific Computing," *Computing in Science & Engineering,* vol. 9, no. 3, pp. 21-29, 2007.
[326] T. E. Oliphant, "Python for Scientific Computing," *Computing in Science & Engineering,* vol. 9, no. 3, pp. 10-20, 2007.
[327] K. J. Millman and M. Aivazis, "Python for Scientists and Engineers," *Computing in Science & Engineering,* vol. 13, no. 2, pp. 9-12, 2011.
[328] S. v. d. Walt, S. C. Colbert, and G. Varoquaux, "The NumPy Array: A Structure for Efficient Numerical Computation," *Computing in Science & Engineering,* vol. 13, no. 2, pp. 22-30, 2011.
[329] E. N. Baker and R. E. Hubbard, "Hydrogen bonding in globular proteins," *Progress in Biophysics and Molecular Biology,* vol. 44, no. 2, pp. 97-179, 1984.
[330] S. S. Zumdahl, "Chap. 13, Bonding General Concepts," in *Chemical Principles*Sixth ed.: Brooks/Cole, Cengage Learning, 2009, pp. 592-659.
[331] J. A. Ippolito, R. S. Alexander, and D. W. Christianson, "Hydrogen bond stereochemistry in protein structure and function," *Journal of Molecular Biology,* vol. 215, no. 3, pp. 457-471, 1990.
[332] J. Ö. Carl Nordling, *Physics Handbook for Science and Engineering*, 8 ed. Sweden: Studentlitteratur, Lund, 2006.
[333] K. F. Riley, *Mathematical Methods for Physics and Engineering*. Cambridge University Press, 2006.
[334] A. Elofsson, B. Hess, E. Lindahl, A. Onufriev, D. van der Spoel, and A. Wallqvist, "Ten simple rules on how to create open access and reproducible molecular simulations of biological systems," *PLOS Computational Biology,* vol. 15, no. 1, p. e1006649, 2019.