Infrared spectroscopic studies: from small molecules to large.

Nadejda Eremina
Abstract

Infrared (IR) spectroscopy has been around since the discovery of IR light by Friedrich Wilhelm Herschel in 1800. However, until 1940’s, IR studies involved only water and small organic molecules, because of the long measurement times and inaccessibility of the instruments. In 1940s came the first commercially available infrared spectrometers, which relied on prisms to act as dispersive elements. The most significant advances in infrared spectroscopy were made when the first Fourier-transform spectrometer was introduced. Development of Fourier-transform infrared spectroscopy (FTIR) and computers has dramatically improved the quality of infrared spectra and minimized the time required to obtain data, making it possible to investigate larger biological systems, such as proteins and nucleic acids.

This thesis has a focus on the applications of several different infrared spectroscopic techniques to a variety of biological systems as well as development of new approaches to investigate complex biological events.

The method utilizing of so-called caged compounds was applied to study the Alzheimer’s amyloid β (Aβ) peptide. Alzheimer’s disease is an incurable neurodegenerative disorder, linked to the formation of Aβ-fibrils in the brain. The molecular mechanism of the fibril formation is still unknown, however it has been noted that the peptide is pH sensitive. Addition of caged-sulfate to the Aβ samples lets one change the pH of the sample in the process of recording IR data, allowing a detailed study of fibril formation in a time-resolved manner.

Caged compounds can also be used to study enzymatic reactions, such as the production of ATP and creatine from ADP and creatine phosphate, catalyzed by creatine kinase (CK). CK in its turn has been characterized as the helper enzyme, to further develop a method that alters the nucleotide composition in a sample. With CK as a helper enzyme it became possible to study the effects of the phosphate binding on the secondary structure of sarcoplasmic reticulum Ca\(^{2+}\)ATPase and determine the structural differences between two very similar states Ca\(^{2+}\)E1ADP and Ca\(^{2+}\)E1ATP.

Drug development is held back by the need to design a special test for each potential drug to control its binding to the target protein. With the help of ATR-FTIR spectroscopy and a specially designed dialysis setup, a general method was developed to detect ligand binding events by observing IR absorbance changes in the hydration shell around the molecules.

ATR-FTIR spectroscopy was also used to determine the binding of DNA to the transcription factors (TFs) of the E2F family. The interaction between these TFs and DNA is a main part of the gene regulatory networks that control cell development, cellular processes and responses to environmental stimuli. However how they recognize their binding sites and the mechanism of binding is not yet understood. By studying the formation of the E2F-DNA complexes by IR, the changes in the secondary structure of the proteins, as well as the distortions of DNA have been observed.
List of publications


IV. **Eremina, N.**, Morgunova, E., Taipale, J., Barth, A. Interaction between Transcription Factors of E2F family and DNA Studied with Infrared Spectroscopy. *Manuscript*

Publication not included in this thesis

# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aβ</td>
<td>Amyloid beta</td>
</tr>
<tr>
<td>AD</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>ADK</td>
<td>Adenylate kinase</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine 5’-diphosphate</td>
</tr>
<tr>
<td>APP</td>
<td>Amyloid precursor protein</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5’-triphosphate</td>
</tr>
<tr>
<td>ATR</td>
<td>Attenuated total reflection</td>
</tr>
<tr>
<td>BB-CK</td>
<td>Creatine kinase homodimer in the brain</td>
</tr>
<tr>
<td>C</td>
<td>Creatine</td>
</tr>
<tr>
<td>Caged ADP</td>
<td>P$_2$-1-(2-nitro)-phenylethyladenosine 5’-diphosphate</td>
</tr>
<tr>
<td>Caged ATP</td>
<td>P$_3$-1-(2-nitro)-phenylethyladenosine 5’-triphosphate</td>
</tr>
<tr>
<td>Ca$_2$E1</td>
<td>Ca$^{2+}$-bound form of Ca$^{2+}$-ATPase</td>
</tr>
<tr>
<td>Ca$_2$E1P</td>
<td>ADP sensitive phosphoenzyme of Ca$^{2+}$-ATPase</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CK</td>
<td>Creatine kinase</td>
</tr>
<tr>
<td>CP</td>
<td>Creatine phosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Desoxyribonucleic acid</td>
</tr>
<tr>
<td>DTGS</td>
<td>Deuterated Triglycine Sulfate</td>
</tr>
<tr>
<td>E2P</td>
<td>ADP insensitive phosphoenzyme of Ca$^{2+}$-ATPase</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared spectroscopy</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared spectroscopy</td>
</tr>
<tr>
<td>MB-CK</td>
<td>Creatine kinase heterodimer in the heart</td>
</tr>
<tr>
<td>Mi$_s$-CK</td>
<td>Sarcomeric creatine kinase</td>
</tr>
<tr>
<td>Mi$_u$-CK</td>
<td>Ubiquitous creatine kinase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>MM-CK</td>
<td>Creatine kinase homodimer in the skeletal muscle</td>
</tr>
<tr>
<td>MTC</td>
<td>Mercury Cadmium Tellurium</td>
</tr>
<tr>
<td>NPE-sulfate</td>
<td>1-(2-nitrophenyl)ethyl sulfate</td>
</tr>
<tr>
<td>PEP</td>
<td>Phosphoenolpyruvate</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PK</td>
<td>Pyruvate kinase</td>
</tr>
<tr>
<td>pRB</td>
<td>Retinoblastoma protein</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SR</td>
<td>Sarcoplasmic reticulum</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
</tbody>
</table>
Outline

ABSTRACT ...................................................................................................................................... 1
LIST OF PUBLICATIONS ........................................................................................................... 2
LIST OF ABBREVIATIONS ........................................................................................................... 3
OUTLINE ...................................................................................................................................... 5

1 INTRODUCTION ......................................................................................................................... 7
  1.1 PROTEINS ............................................................................................................................... 7
  1.2 PROTEIN LIGAND BINDING .................................................................................................... 8
  1.3 WATER AND PROTEINS IN WATER ....................................................................................... 8
  1.4 DNA ........................................................................................................................................ 9
  1.5 CA²⁺-ATPASE ........................................................................................................................ 11
  1.6 CREATINE KINASE ............................................................................................................... 13
    1.6.1 Functional coupling between creatine kinase and CaATPase ........................................... 15
  1.7 PYRUVATE KINASE ............................................................................................................... 15
  1.8 TRANSCRIPTION FACTORS .................................................................................................... 17
    1.8.1 General background ........................................................................................................... 17
    1.8.2 E2F family .......................................................................................................................... 18
  1.9 AMYLOIDS ............................................................................................................................ 19
  1.10 ALZHEIMER’S DISEASE ...................................................................................................... 21
    1.10.1 Alzheimer’s disease .......................................................................................................... 21
    1.10.2 Amyloid precursor protein .............................................................................................. 21
    1.10.3 The Aβ aggregation process ........................................................................................... 23
    1.10.4 Oligomers ........................................................................................................................ 23

2 METHODS ................................................................................................................................... 25
  2.1 INFRARED SPECTROSCOPY .................................................................................................. 25
    2.1.1 Vibrational spectroscopy ................................................................................................... 25
    2.1.2 Infrared spectroscopy ........................................................................................................ 26
    2.1.3 FTIR spectrometer ............................................................................................................ 26
  2.2 ATTENUATED TOTAL REFLECTION .................................................................................... 27
  2.3 REACTION-INDUCED DIFFERENCE SPECTROSCOPY .......................................................... 28
  2.4 CAGED COMPOUND APPROACH .......................................................................................... 29
  2.5 DIALYSIS SETUP FOR THE ATR EXPERIMENTS ................................................................. 31
  2.6 FTIR SPECTROSCOPY OF PROTEINS ............................................................................... 32
  2.7 HELPER-ENZYME APPROACH ............................................................................................. 33
  2.8 CIRCULAR DICHOISM .......................................................................................................... 34

3 RESULTS AND DISCUSSION .................................................................................................... 36
  3.1 PAPER I – FORMATION OF TWO DIFFERENT TYPES OF OLIGOMERS IN THE EARLY PHASE OF PH-INDUCED AGGREGATION OF THE ALZHEIMER’S AB(12-28) PEPTIDE ............................................................................. 36
3.2 Paper II - Use of Creatine Kinase to Induce Multistep Reactions in Infrared Spectroscopic Studies

3.3 Paper III - Detection of Ligand Binding to Proteins Through Observation of Hydration Water

3.4 Paper IV: Interaction Between the Transcription Factors of the E2F Family and DNA Studied with Infrared Spectroscopy

4 Summary and Future Plans

5 Acknowledgements

6 Sammanfattning på svenska

7 References
1 Introduction

1.1 Proteins

Proteins are essential building blocks of all living organisms along with polysaccharides, nucleic acids and fatty acids. Proteins are long polymer chains of amino acids held together by peptide bonds in a specific sequence, often referred to as the primary structure of the protein.

Amino-acid residues of the protein chain can interact with each other through hydrogen bonds, forming regularly repeating patterns known as the secondary structure of the protein. Most common secondary structures include α-helices, β-sheets and turns. α-helix is a right-handed spiral, where every -NH group of the backbone is hydrogen bonded to the C=O group of an amino acid four residues away along the sequence.¹ In a β-sheet carboxyl groups in the backbone of one strand are laterally bonded to the –NH group of an adjacent strand. Depending on a relative direction of the strands β-sheets can be classified into parallel and anti-parallel. In a parallel β-sheet the N-termini of the participating strands are aligned with one another while in an antiparallel β-sheet, the sequential strands alternate directions so that the N-terminus of one strand is aligned to the C-terminus of the following strand.² Turns are defined as secondary structure motifs, where Cα atoms of the amino acids, that are separated by a few peptide bonds and are not involved in α-helixes or β-sheets, are in close enough proximity to form hydrogen bonds. Turns are classified according to the separation between the end residues and their hydrogen bond patterns. The tertiary structure of a protein describes the spatial relationships between the amino acids that are far apart in the sequence. This fold is responsible for the basic functions of a protein and is stabilized by non-covalent interactions, disulphide bonds and the hydrophobic effect. The hydrophobic effect is what drives the folding process of the protein, as the number of hydrophobic amino acids exposed to water is minimized.³⁴ And finally the highest level of protein organization is quaternary structure, which describes how several folded units associate with each other.

Proteins have a wide variety of roles in the cell, for example enzymes that catalyze most of the reactions in the cell, antibodies that bind antigens or foreign substances and target them for destruction, transporters that binds small molecules and move them from one cell location to the next, etc.⁵ An important group of enzymes that are central to many biological processes are the kinases. These enzymes transfer the terminal phosphoryl group from the high-energy donor molecules, such as e.g. ATP to specific substrates, either another nucleotide or a small molecule or to a protein. In Papers II and III several members of this group of enzymes are studied in closer detail.
1.2 Protein Ligand binding

A ligand is a substance that binds and forms a complex with a protein to serve a biological purpose. This event usually occurs by a combination of intermolecular forces, such as electrostatic interactions, hydrogen bonding and van der Waals forces. On rare occasions even covalent bonding can occur. Different type of ligands include substrates, inhibitors, activators, neurotransmitters and even nucleic acids. The strength of the binding is determined by the binding affinity, which can be described by a dissociation constant $K_d$, an equilibrium constant that measures the propensity of a ligand to dissociate from the macromolecule it is bound to. The dissociation constant is defined as follows:

$$K_d = \frac{[P] \cdot [L]}{[C]}$$

where $[P]$, $[L]$ and $[C]$ are the concentrations of protein, ligand and complex respectively. $K_d$ corresponds to the concentration of a ligand, at which half of the binding sites on the protein are occupied. This means that the smaller is the dissociation constant, the more tightly the ligand binds, or the higher is the affinity between ligand and protein.

Ligand binding often perturbs the tertiary structure of the protein. In most cases the perturbations are not very large, but small movements in protein structure do occur in all binding events. These movements usually involve flexible loops and help to maximize the interaction between protein and ligand and also to minimize the interaction with the solvent.  

1.3 Water and proteins in water

Water is a bent molecule, the distribution of its charge is asymmetric, and hence water is polar. Molecules in aqueous solutions interact with water molecules through the formation of hydrogen bonds and through electrostatic interactions. Water is both an H-bond donor and acceptor. The oxygen of water has two covalent bonds with H, and it has the possibility of forming two H-bonds with its two lone-pair electrons located on the O. Because water can form four H-bonds and the small water molecules can rotate in solution, bonds are continuously forming and breaking, producing flickering networks of hydrogen bonds in liquid water. Water in condensed phases would arrange so that two H-atoms of the neighboring water would be associated with the lone pair of electrons of the O. Statistically the water molecules have many possibilities of arrangement. Therefore water has large entropy that increases as temperature increases. Hydrophobic molecules that are put into water, disrupt the attraction patterns of water due to hydrogen bonding, leading to water molecules becoming more ordered around the solutes and thus decreasing the entropy. This is the origin of the hydrophobic effect and is the reason why
hydrocarbons do not dissolve in water. This effect also reduces the possibility of interactions between water molecules making it unfavorable for a non-polar group to be in water. The water molecules in contact with these nonpolar molecules form ‘cages’ around them, becoming more well-ordered than the water molecules free in solution. However when two such molecules come together, some of the water molecules are released, allowing them to interact freely with bulk water.\textsuperscript{7,10}

Some classes of proteins are designed to function in water and lose their activity when removed from it. Data from x-ray crystallography shows that a typical protein has about 1.5-2 water molecules per amino acid residue.\textsuperscript{11} In solution these are constantly in motion moving over and around one another as forces of attraction and repulsion continuously change with translation and rotational movement. Water increases protein plasticity and lubricates protein folding by processes such as hydrogen bonds bridging of backbone carbonyls and amides.\textsuperscript{12} In a solution with protein, water molecules will attempt to occupy all space not already occupied by protein atoms. The positional stability of the water molecules is dependent on protein properties like the geometry of protein surfaces, protein crevices and holes, polarity of the side chains and H-bonding capacity, Gibbs free energy etc.\textsuperscript{12–14}

A number of water molecules that are buried deep within a protein are often highly conserved between homologous proteins and form an integral part of the protein structure. These buried water molecules bind to the peptide groups not involved in hydrogen bonds with other peptide groups and facilitate main chain hydrogen bonding. They have much longer residence times then the average water molecule, ranging from 10 nanoseconds to 10 milliseconds. Surface ordered waters are more likely to be in surface grooves and show considerably more discrimination between polar and nonpolar side-chains compared to the deep-buried water molecules. The first hydrogen shell is relatively ordered in comparison to outer shells with well-defined time average hydrogen site.\textsuperscript{15,16}

1.4 DNA

Desoxyribonucleic acid (DNA) is another of the major macromolecules alongside proteins and polysaccharides. It encodes all of the genetic information needed for the development and functioning of a living organism.

The structure of DNA is made up of two polynucleotide chains, coiled round the same axis, forming a double helix as shown in figure 1.1. Primarily two forces stabilize the double helix. The first one is hydrogen bonding between the backbone phosphate sugars and the complementary base pairs. The second one is base-stacking interactions between the aromatic bases. The four bases found in DNA are two purines: adenine (A) and guanine (G) and two pyrimidines: cytosine (C) and thymine (T).
The B-form described by Watson and Crick is believed to predominate in cells. It is 23.7 Å wide and extends 34 Å per 10 base pairs of sequence. The double helix makes one complete turn about its axis every 10.4-10.5 base pairs in solution.\textsuperscript{17,18} The attachment of bases to the backbone sugars through glycosidic bonds is asymmetrical. This results in the formation of two different grooves on opposite sides of the base pairs, the major and minor grooves. Although the grooves are of similar depth in B-DNA, the major groove is considerably wider than the minor groove. Each groove is lined by potential hydrogen-bond donor and acceptor atoms that enable specific interactions with proteins. Many proteins that bind DNA recognize specific sequences of bases and most bind in the major groove with four hydrogen bond donors and acceptors compared to the minor groove which has only two.\textsuperscript{10}

Recently another mechanism for protein-DNA recognition was proposed, which involves the changes in the structure of the DNA helix. It was long thought that the recognition of specific DNA sequences would take place primarily in the major groove by the formation of series of hydrogen bonds between amino acids and bases, mentioned above. However in the past years, it has been shown that the DNA can assume conformations that deviate from the structure of B-form helix, to optimize the protein-DNA interface. In some case that conformational changes can be quite large, such as e.g. opening of the minor groove.\textsuperscript{19-22}

The A-form DNA has a shorter more compact helical structure. It appears during dehydration of DNA or in RNA and RNA-DNA hybrid helices. In the A-form the base-pairs are not perpendicular to the helical axis but instead they are tilted at a steep angle and are placed closer together along the helical axis. The helical pitch of A-form DNA is closer to 11 base pairs per turn in 28 Å rather than 34 Å. As a result, the A-form is about 25% shorter than the B-form. The
tilted base pairs also allow room for the 2' oxygen present in RNA chains and therefore all double helices containing at least one RNA strand are in the A-form.$^{10,23,24}$

When the self-complementary polymer d(CG)$_3$ was crystallized in high ionic strength conditions in 1979, a very unusual form of DNA called the Z-form was discovered. The Z-form helix is left-handed with only one single groove. The nucleotide bases along one strand alternate between the syn- and anti-conformation, while the backbone is arranged in a zigzag pattern, thus giving the name Z-DNA. The Z-DNA proved to be highly antigenic, as Z-DNA specific antibodies were found in several human autoimmune diseases. It has also been shown that the Z-DNA sequences tend to concentrate near the transcription start sites, and the formation of Z-DNA fragments initiates the transcription.$^{25}$

1.5 $\text{Ca}^{2+}$ATPase

P-type ATPases make up a large family of enzymes whose main role is ATP-driven transport of essential ions across biological membranes in order to maintain the cellular environment. They include among others Na$^+$, K$^+$ ATPase, H$^+$, K$^+$ATPase and Ca$^{2+}$ATPase.

![Figure 1.2 The sarcoplasmic reticulum $\text{Ca}^{2+}$ATPase: Ca$_2$E1P state. The structure includes three cytoplasmic domains, the nucleotide (N) domain, phosphorylation (P) domain, actuator (A) domain and the transmembrane (TM) domain. (PDB ID: 3BA6)
Among the enzymes mentioned above, Ca\(^{2+}\)ATPase is one of the most studied. It mediates muscle relaxation by transporting Ca\(^{2+}\) ions back into the sarcoplasmic reticulum (SR), against the concentration gradient.\(^{26}\) Ca\(^{2+}\)ATPase is made up of a single polypeptide chain of 994 amino acids and has a molecular mass of about 110 kDa. Its structure is similar to other P-type ATPases and consists of a transmembrane domain (TM), made up of 10 transmembrane helices (M1-M10), and 3 cytoplasmic domains: phosphorylation (P), nucleotide-binding (N), and actuator (A) domains, as can be seen in figure 1.2. The two cytoplasmic domains P and N are located in sequence between helices M4 and M5. Upon Ca\(^{2+}\) binding and dissociation and upon nucleotide binding to Ca\(_{2}\)E1 helices M1-M6 tend to move, while M7-M10 keep their position and seem to anchor the protein to the membrane.\(^{27}\) The Ca\(^{2+}\) binding sites are located between the transmembrane helices M4-M6 and M8.

The P-domain contains a highly conserved residue Asp351, to which phosphate is bound in the phosphoenzyme formation. The N-domain is linked to the P-domain and contains the nucleotide binding site. The A-domain is highly mobile and acts as a gate for the Ca\(^{2+}\) transport, through its connection to M1-M3 helices.

Recently the presence of a fifth domain has been proposed, a so called “core domain”, composed of the most conserved parts of the ATPase. It plays an important role in Ca\(^{2+}\)/H\(^{+}\) translocation, as it forms a communication between the phosphorylation sites and the Ca\(^{2+}\) binding sites.\(^{28}\)

\[ \text{Figure 1.3 The reaction cycle of Ca}^{2+}\text{ATPase.} \]

During the reaction cycle the Ca\(^{2+}\)ATPase undergoes several conformational changes and forms at least four phosphorylated and unphosphorylated intermediates. The model of the mechanism of Ca\(^{2+}\)ATPase, as proposed by de Meis and Vianna, is described in Figure 1.3. In the initial step of the reaction cycle, two Ca\(^{2+}\) ions from the cytoplasm bind to the state E to form the Ca\(_{2}\)E1 intermediate. The Ca\(_{2}\)E1*ATP intermediate is formed as ATP binds to the N-domain of the enzyme, resulting in ATP hydrolysis, where the \(\gamma\)-phosphate of ATP is transferred to the Asp351 residue in the P-domain. The phosphoenzyme intermediate Ca\(_{2}\)E1P is ADP sensitive, meaning that it can synthesize ATP in presence of ADP. The subsequent reaction step leads to a number of
conformational changes in the enzyme, decreasing its affinity for Ca\(^{2+}\) and releasing the ions into the SR lumen while E2P is formed. During this stage the water molecules in the phosphorylation site can exchange oxygen atoms with the phosphate, triggering release of P\(_i\), completing the cycle. During this process counter transport of 2-3 H\(^{+}\) occurs. It has been indicated that Ca\(^{2+}\) and H\(^{+}\) compete for the same binding site. This could be explained by the fact that departure of Ca\(^{2+}\) ions causes an overall negative charge in the binding site, thus destabilizing the structure, so the protons neutralize this effect.\(^{29,30}\)

### 1.6 Creatine Kinase

![3D structure of rabbit muscle creatine kinase (PDB ID: 2CRK)](image)

Creative kinase (CK) belongs to the subclass of guanidino-kinases along with glycocyamine kinase, arginine kinase etc. It transfers the phosphate group of creatine phosphate (CP) to an ADP molecule, producing ATP and unphosphorylated creatine (C), according to the following reaction: \(\text{ADP} + \text{CP} \rightarrow \text{ATP} + \text{C}\). The elevated level of CK in human blood is an important diagnostic indicator for diseases of the nervous system and the heart muscle, for malignant hypothermia, and for certain tumors.\(^{31}\)

The main function of CK is to maintain the energy balance in the cells. ATP is a direct source of energy for most energy requiring processes in biological systems. Many cells and tissues, e.g. muscle, brain, photoreceptor cells, all require large amounts of energy to be able to function properly. CK constitutes about 10% of the total soluble cytoplasmic protein and its activity is much higher than other ATP synthesizing and consuming processes.\(^{32}\) Under physiological conditions the equilibrium of creatine kinase is shifted towards ATP synthesis. During the transition from rest to muscular work a slight change in ADP concentration causes a significant
change in the concentrations of CP and C, whereas the ATP concentration remains essentially unchanged until complete exhaustion of the CP stores. The function of CK as an energy buffering mechanism means that, under metabolic conditions CK maintains the ATP/ADP ratio at a high level. While maintaining the ATP concentrations, CK prevents the rise in free ADP, which would cause inactivation of cellular ATPases and a net loss of adenine nucleotides. Along with utilizing ADP, the CK reaction also consumes protons, which are products of ATP hydrolysis, so the functional coupling of the CK with ATPases prevents local acidification of cells that are breaking down high amounts of ATP within short periods of time.

CK is also thought to function as an energy transport system, a so-called “CP-shuttle”. Here CP serves as an energy carrier connecting sites of ATP production with sites of ATP utilization via the subcellularly compartmentalized (mitochondrial and cytosolic) CK isozymes. For example the complex made up of CK, the inner mitochondrial membrane adenylate translocator and the outer membrane porin, constitutes one side of this shuttle that exports CP from mitochondria into the cytosol.

CK has three organ specific cytoplasmic isozymes with the molecular weight of ~85-kDa each: a MM-CK homodimer in the skeletal muscle, a BB-CK homodimer in the brain, and a MB-CK heterodimer in the heart. In addition to these there are also two mitochondrial isozymes: Miα-CK, the ubiquitous isozyme and Miβ-CK, the sarcomeric isozyme, that exist either as dimers or as octamers. The full-length sequence of the mitochondrial isozymes is about 35 residues longer than of the cytosolic ones. The additional residues belong to a leader peptide, which is removed proteolytically, either during or after the translocation across the mitochondrial membrane.

The CK monomer, shown in figure 1.4, consists of two domains: an α-helical N-terminal domain and a C-terminal domain, connected by a long linker. The C-terminal domain is an eight-stranded antiparallel β-sheet flanked by α-helices. The β-sheet forms a cradle with five α-helices on its convex side and one α-helix on the concave side. This helix together with the majority of the residues of the β-sheet are highly conserved among the CK species. The active site, located in the β-sheet cradle, is surrounded by a cluster of positively charged amino acids, among which there are five Arg (130, 132, 236, 292, 341) from the C-terminal domain and one Arg 96 from the N-terminal domain. These, together with two highly conserved histidines (His191 and His296), are responsible for the nucleotide binding. The binding site for creatine is located in the same area but is much smaller than the nucleotide binding site. The only direct H-bond is formed between the creatine carboxylate and the main-chain nitrogen of Val72, while the rest of the interactions occur via water molecules.
1.6.1 Functional coupling between creatine kinase and CaATPase.

It has been pointed out that the changes in myofibrillar function do not correlate with the ATP level available in the medium. This can be explained by the existence of site-specific regeneration of ATP, which creates a local pool of ATP close to the sites of ATP utilization. Several studies have shown that CK can attach itself to the SR membranes in different types of tissues near ATP consuming sites such as Ca$^{2+}$-ATPase. Local ATP regeneration is especially important for Ca$^{2+}$-uptake by sarcoplasmic reticulum (SR) when the rate of ADP production is high and luminal free calcium starts to increase. Another important observation is that Ca$^{2+}$-ATPase has enhanced affinity for the ATP rephosphorylated by the CK bound to the SR compared to ATP synthesized by the other ATP regenerating systems. At the same time the SR-bound CK proved to be a more effective competitor for ADP, released by the Ca$^{2+}$-ATPase, compared to other kinases. This apparent greater binding of ADP by SR-bound CK can be interpreted as a sign of close structural proximity of CK and Ca$^{2+}$-ATPase on the SR-membrane. The importance of this local phosphorylation of ADP by CK is not only to supply Ca$^{2+}$-ATPase with ATP, but also to keep a low level of ADP, as it had been shown that ADP inhibits Ca$^{2+}$-ATPase.$^{41-45}$

1.7 Pyruvate kinase

Pyruvate kinase (PK) is an enzyme that is involved in the final step of glycolysis, presented in figure 1.5. It catalyzes the transfer of a phosphate group from phosphoenolpyruvate (PEP) to ADP, producing one molecule of pyruvate and one molecule of ATP. This process requires manganese or potassium ion to function.

![Pyruvate kinase enzymatic reaction](image)

*Figure 1.5 Pyruvate kinase enzymatic reaction*

Pyruvate kinase is also involved in synthesizing glucose in liver. In this process PEP, instead of producing pyruvate, is converted into glucose. Similar glycolytic pathways have been found in both prokaryotes and eukaryotes, meaning that PK is present in most organisms. In mammalian tissue, four different isoforms have been discovered: M1 found in skeletal muscle, M2 found in kidney, adipose tissue and lungs, L found in liver and R found in the red blood cells.$^{46}$
M1 type rabbit muscle pyruvate kinase, presented in figure 1.6, is made up of four subunits. Each subunit folds into four domains: A, B, C and N. Domain N is a short helix-turn-helix motif, domain A is a parallel (β/α)_8 barrel, domain B is a nine-stranded β-barrel and domain C is composed of five α-helices and a five stranded β-sheet, as can be seen in Figure 1.6.

Figure 1.6 3D structure of pyruvate kinase (PDB ID: 1AQF)

The active site lies in a pocket between domains A and B, which contains the highly conserved residues Arg-72, Arg-119, Lys-269, Asp-112, Glu-271 and Asp-295. 47

So far no structure of PK with bound PEP has been reported due to slow hydrolysis of PEP. However structures with PEP analogs indicate that the side chains of Arg-72 and Lys-269 are responsible for the binding of PEP. 47,48 Cations play a crucial role in the activity of most of PKs, it has been shown that for example M1 type PKs require both monovalent cations such as K⁺ and divalent cations such as Mg²⁺ for activity. Crystal structures of rabbit muscle PK have shown that it requires two divalent cations per active site. 49 One of these coordinates directly to the protein through the carboxylate side chains of Glu-271 and Asp-295, while the second one binds to the phosphoryl groups of ATP and does not interact with the enzyme. K⁺ is located in a well-defined pocket with four ligands: Asn-74, Ser-76, Asp-112 and Thr-113. It is worth to mention that Asp-112 actively participates in binding of PEP. 47,48

PK is very important for the functioning of the human body. If there is a lack of pyruvate kinase, glycolysis slows down, stripping the cells from their main energy source, which can lead to very severe consequences. For example, red blood cells with pyruvate kinase deficiency can undergo
hemolysis, which leads to hemolytic anemia.\textsuperscript{50} PK also has a high potential to be used as a tumor marker, since one of its isoforms M2 is over-expressed by tumor cells and can therefore be quantitatively determined.

\textbf{1.8 Transcription factors}

\textbf{1.8.1 General background}

The regulation of gene transcription is central both to tissue specific gene expression and to the regulation of gene activity in response to specific stimuli. In most cases regulation occurs at the level of transcription by deciding which genes will be transcribed into primary RNA transcript. Once this has occurred, the remaining stages of gene expression, such as RNA splicing, occur automatically and result in the production of the corresponding protein. Inspection of the regulatory regions of genes that show similar patterns of transcription, revealed the presence of short DNA sequences that are common to genes with a particular pattern of regulation, but were absent from other genes, which did not show this pattern of regulation. These short DNA sequences act by binding specific regulatory proteins known as transcription factors (TFs), which regulate the transcription of the gene.\textsuperscript{51,52}

Different transcription factors have a modular structure in which specific regions of the molecules are responsible for binding DNA, while other regions produce a stimulatory or inhibitory effect on transcription. Most transcription factors have been classified according to their distinct DNA-binding domains. These include: the helix-turn-helix motif, the two cysteine - two histidine zinc finger, the multi-cysteine zinc finger, the Ets domain and the basic DNA binding domain, which usually is followed by a dimerization domain.\textsuperscript{52,53}

In addition to the DNA-binding domain, many transcription factors also contain activation or suppression domains. Just as in the case of the DNA-binding domains, there are a number of different types of activation domains. These are classified according to their composition: whether they are rich in acidic amino acids, glutamine residues or proline residues. Activation domains function by interacting with the components of the basal transcriptional complex: RNA polymerase II and various transcription factors, which assemble at the gene promoters and are essential for the transcription to occur.\textsuperscript{52} Wide variety of TFs act as inhibitors of transcription for specific genes by preventing the activating TF from binding to DNA, either by binding to its DNA binding sequence, or by forming a non-DNA binding complex with the activating TF, or by quenching.\textsuperscript{54,55}

Given the vital role of TFs in a wide variety of cellular processes, it’s not surprising that alterations in these proteins can result in disease. The most common of the human diseases related to TFs is
cancer. The growth of cells is controlled by the variety of proteins, some of which stimulate cellular growth while others inhibit it. The abnormal activation of specific genes encoding growth promoting factors, as well as inactivation of gene-encoding growth-inhibiting proteins can both lead to cancer.\textsuperscript{56}

1.8.2 \textit{E2F family}

The mammalian cell cycle is a highly regulated process that is influenced by positive and negative growth–regulatory signals during the G\textsubscript{1} stage. These signals are controlled by the transcriptional activity of E2F-family of transcription factors. The first E2F protein was discovered in 1980s as the transcriptional activator of the adenovirus E2 promoter. Further studies have shown that E2Fs also control the transcription of cellular genes important for cell division, such as genes encoding cycle regulators, the retinoblastoma protein (pRB), enzymes involved in nucleotide biosynthesis, as well as in cell death.\textsuperscript{57,58}

In mammalian cells, there are currently eight known E2F family members, divided into activator (E2F1-3) and repressor (E2F4-8) subclasses. The classic E2Fs (E2F1-6) contain one DNA-binding domain, and a dimerization domain required for the interaction with a member of the dimerization-partner family (DP1-DP4). The dimerization with DP seems to be required for the formation of functional transcription complexes, however the effects on the transcription activity are not fully understood.\textsuperscript{59-61} E2F1-6s activity is controlled through the binding of the pRB family of proteins.\textsuperscript{59,61,62} The atypical family members, E2F7 and E2F8, contain two DNA-binding domains and can form homodimers or E2F7-E2F8 heterodimers.\textsuperscript{63} The DNA-binding domain of E2Fs consists of three α-helices and a β-sheet, so-called winged-helix DNA-binding motif. Sequence comparison between the E2F family members presents a highly conserved motif RRXYD, which is responsible for the DNA base contact within the binding domain. Any changes to this sequence cause loss of DNA binding.\textsuperscript{59,63}

Many of cell replication genes contain E2F binding sites, which underlines the vital role of E2Fs in directing cell cycle progression. During the G\textsubscript{0} and G\textsubscript{1}, E2F activity is mainly mediated by E2F4 and E2F5, which are preferentially bound to p130 and inhibit the E2F-responsive genes. At the same time, the activating E2Fs are bound and inactivated by the pRB. As the cell progresses to the late G\textsubscript{1} phase, pRB and p130 are phosphorylated, releasing the activating E2Fs. p130 is targeted for ubiquitin-mediated degradation, its level falls and E2F4 and E2F5 are removed from the nucleus to the cytoplasm. As E2F1-3 get activated transcription of the S-phase genes is rapidly increased. Whether the cell actually proceeds in the cell cycle and divides, or dies, depends on the integrity of the replication process and the balance between the growth factors and E2F1.\textsuperscript{57,64}

In Paper IV of this thesis we have studied the DNA binding processes of two members of E2F family E2F1 and E2F8.
E2F1 is the original and the most extensively studied member of the E2F family. It belongs to the activator subfamily, coordinating the expression of key genes involved in cell cycle regulation and progression.\textsuperscript{65,66} It can also induce apoptosis via distinct p53-dependent and independent pathways. Transcriptional activation of p73 by E2F1 leads to the activation of p53-responsive target genes, which leads to cell death by apoptosis. Moreover, E2F1 is known to upregulate the pro-apoptotic members of the B-cell leukemia 2 family of proteins, and also downregulate of anti-apoptotic signals, by inhibiting activity of nuclear-factor-kappa-B inhibitor protein, thereby enhancing apoptosis.\textsuperscript{67,68} E2F1 also participates in DNA repair either directly at the sites of DNA damage or through modulation of DNA repair genes that are under its transcriptional control or by initiating a cascade of events that leads to apoptosis as a response to various degrees of DNA-damage.\textsuperscript{69–71} Structurally E2F1 belongs to the leucine-zipper family of proteins containing a DNA-binding leucine-zipper domain followed by a dimerization domain. It recognizes and binds to specific DNA sequences 5'-TTTSCGS-3', where S can be either G or C, by forming heterodimers with transcription factors of the DP family.\textsuperscript{61,72,73}

Given this variety of cellular functions it has been shown that E2F1 can be either an oncogene or a tumor suppressor, as there are examples in mouse models of both positive and negative effects on tumorigenesis when E2F1 is either deleted or overexpressed.\textsuperscript{58,74–76}

E2F8 controls a transcriptional network that needs to be repressed to induce liver cell polyploidization. This physiological process is essential for controlling cell size, but is unexpectedly not relevant for liver differentiation or liver regeneration.\textsuperscript{77} The function of E2F8 under normal and pathological conditions is still obscure. Its role in cancer has not been widely studied but it is known to be strongly upregulated in human hepatocellular carcinoma. In contrast to classical members of the E2F family, E2F8 contains two DNA-binding domains and regulates transcription in a DP-independent manner. Both DNA-binding domains are required for DNA-binding but the mechanism of this interaction remains unknown.\textsuperscript{64,77–79}

### 1.9 Amyloids

The term amyloid comes from early misidentification of the substance as starch. Nowadays there are two definitions of amyloids: a classical medical definition that states that an amyloid is an extracellular, protein-like deposit exhibiting β-sheet structure and a biophysical one that states that an amyloid is any polypeptide that polymerizes to a cross-β sheet motif.\textsuperscript{80}

The cross-β motif consists of several β-sheets twisting around a central axis, each sheet being composed of hydrogen bonded β-strands running perpendicular to the fiber axis.\textsuperscript{81} The β-sheets can be parallel or antiparallel, though parallel β-sheets seem to be the most common ones. An example of a cross-β motif with a parallel β-sheets is shown in figure 1.7. Further studies have
revealed two types of cross-β motifs, depending on the relative orientations of the strands in the sheets and the angle between them. Adjacent strands within the sheets are generally separated by ~ 4.7 Å, while a typical distance between the β-sheets ranges from 9-12 Å, depending on the character of the side chains. The cross-β structure is very stable as it uses fully the hydrogen bonding capacity of the backbone.6,82,83

Figure 1.7 Alzheimer’s Aβ(1-40) fibril composed of two stacked cross-β motifs. (PDB ID: 2LMN)

Among protein folds this specific pattern is unique and gives rise to a variety of functions, both good and bad for the organism. Recent studies have shown that amyloids have a variety of functions in nature. They participate in sorting, storing and releasing hormones, regulate certain pathways and mRNA translation etc., however they are mostly associated with a number of serious human diseases e.g. Alzheimer’s disease, Parkinson’s disease, Down’s syndrome, type 2 diabetes, etc.8,84–87

The mechanism of the formation of a full fibril from a peptide is still not fully understood. The suggested path for the process involves a transition from random coil to β-strands, to cross-β motifs, which then assemble from monomeric species to oligomers, to protofilaments, to shorter precursors and finally to full-length fibrils. It has been found that the intermediate species such as oligomers and protofilaments are the most toxic in the disease-associated fibrils, while the functional amyloids seem to be lacking these intermediates.82,83
1.10 Alzheimer’s disease

1.10.1 Alzheimer’s disease

Alzheimer’s disease (AD) is the most frequent, widespread neurodegenerative disorder in the elderly human population.\textsuperscript{88} The common symptoms of it are progressive memory impairment, altered behavior such as paranoia, delusions, loss of social skills, progressive decline of language function, etc.\textsuperscript{89} Since the condition was discovered in 1906,\textsuperscript{90} it has been widely studied, however what causes it and how it progresses is still not fully understood. More than 20 million people worldwide suffer from AD, 100 000 cases are reported in Sweden in the past year.\textsuperscript{91} About 95% of the patients are of 65 years old and above. Above an age of 65 years, the risk of developing the disease increases twofold for every fifth year, reaching a 50\% chance at the age of 85 years. Most cases of Alzheimer’s disease are random, with risk factors such as age, high blood pressure or head injury, however, there is also a familial form caused by various mutations. The familial form is uncommon, but it usually occurs earlier in life, with typical range between 45 to 65 years of age and is inherited from a first degree relative with a history of AD.\textsuperscript{90}

Clinical diagnosis of AD is based on patient history coupled with advanced imaging techniques such as positron emission tomography (PET) for example. Recent advances in imaging technology have led to development of highly sensitive methods that can directly detect amyloid plaques and tangles that are thought to be the main cause of the disease. Such plaques contain large amounts of amyloid-\(\beta\) peptide (A\(\beta\)), figure 1.8, which is 36-43 amino acids long and occurs mostly in its fibrillar form.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{A\(\beta\) (1-42) peptide in its native state, which is represented by an \(\alpha\)-helix imbedded in the membrane. (PDB ID: 1Z0Q)}
\end{figure}

1.10.2 Amyloid precursor protein

The peptide originates from the amyloid precursor protein (APP) by sequential proteolytic cleavages. APP is a single transmembrane protein located outside the cell, with an \(\alpha\)-helix spanning the cell membrane and a small fraction of the protein perturbing into the cell’s interior.
There are three major isoforms of APP expressed throughout the body, the most common of which is found predominantly in the synapses of neurons. One of its major roles is synaptic formation and repair. APP is translocated into the endoplasmic reticulum via its signal peptide and then posttranslationally modified through the secretory pathway. The posttranslational modification of APP includes i.e. proteolytic cleavage to generate peptide fragments. The cleavage is catalyzed by proteases from the secretase family.

Most APP molecules are cleaved by α-secretase, rather than β-secretase, near the middle of the Aβ region. This releases the large, soluble ectodomain (APPs-α) into the medium and allows the resultant 83-residue, membrane-retained, C-terminal fragment to be cleaved by γ-secretase, generating the small p3 peptide. α-secretase acts on APP molecules at the cell surface, although some processing also occurs in intracellular secretory compartments. It is thought that cleavage by α-secretase followed by γ-secretase enables the release of the APP intracellular domain into the nucleus, where it may participate in transcriptional signaling.

β-secretase has the same function as the α-secretase, it removes a large soluble ectodomain. However its cleavage site is a few residues earlier then α-secretase, leaving a 99-residue C-terminal fragment. Cleavage of this fragment in the middle of the transmembrane domain by γ-secretase generates the Aβ fragments, as shown in figure 1.9.

![Figure 1.9 A schematic of APP proteolytic cleavage (adapted from RCSB Protein Data Bank Molecule of the Month 2006)](image)
Both β-secretase and γ-secretase have more than one cleavage site resulting in multiple forms of Aβ peptide: from Aβ(1-37) to Aβ(1-43). Aβ(1-40) and Aβ(1-42) are the most occurring peptides, consisting of 28 residues of the extracellular domain of APP and 12 or 14 residues of the transmembrane domain, respectively. The Aβ(1-40) version is an amphiphilic peptide with a hydrophilic N-terminal part, a central hydrophobic segment and a hydrophobic C-terminus. The Aβ(1-42) version of the peptide has two additional hydrophobic amino acids in the C-terminus and is therefore more prone to aggregation and it has been shown that most of the peptides found in the AD plaques are the Aβ(1-42) version.  

1.10.3 The Aβ aggregation process

The aggregation process of Aβ is still not fully understood, however two kinetic models have been proposed. One is a nucleation-dependent polymerization model, that starts off with unstructured Aβ peptides that are converted into intermediate monomers containing some degree of β-sheet structure that assemble into a “nucleus” in a step called the nucleation phase. Once the nucleus is formed, it acts as a seed for exponential fibril growth. This is the elongation phase, which results in the formation of oligomers and high order aggregates. In the final steady state phase the fibrils are in equilibrium with the monomers. The second model is referred to as the template assembly model. Here the fibrils grow via the reversible addition of a soluble monomer to a pre-existing fibril, followed by a conformational change to an aggregation-competent state and hence the irreversible association onto the end of the fibril.

The fibrillation process is affected by many factors such as the initial peptide aggregation state, peptide concentration, peptide length, pH etc. It has also been proposed that metal ions such as Zn^{2+}, Fe^{3+} and Cu^{2+} have an impact on the aggregation process. 

1.10.4 Oligomers

Until 1992, formation of Aβ fibrils was considered a pathological event. However the degree of the disease did not seem to correlate with the amount of plaques found in the patients. These were also found in the cerebrospinal fluid and plasma of healthy subjects throughout their life, which indicated that Aβ fibril production is a normal metabolic event. Instead the levels of soluble Aβ in the brain matched very well with synapse loss; the higher s the concentration of the soluble Aβ, the worse the condition of the patients. Recent results have indicated, that the source of neurotoxicity are not the insoluble Aβ fibrils, but the soluble oligomers and protofilaments, which occur at the intermediate stages of fibril formation. Experiments have shown that cells in the brains of AD patients have very high amounts of oxidated proteins, lipids and DNA. It has been suggested that interactions of Aβ oligomers with Fe^{2+} or Cu^{2+} generate H_2O_2, which leads to lipid peroxidation and formation of the lipid oxidation products 4-hydroxynonenal and acrolein, which can bind to and modify proteins on cysteine, lysine and histidine residues.
Aβ oligomers can also cause mitochondrial oxidative stress and dysregulation of Ca\(^{2+}\) homeostasis, resulting in impairment of the electron transport chain, increased production of superoxide anion radicals and decreased production of ATP.\(^{105,106}\) Superoxide radical is in turn converted to \(\text{H}_2\text{O}_2\) by the activity of superoxide dismutases and can also interact with nitric oxide via nitric oxide synthase to produce peroxynitrite. Interaction of \(\text{H}_2\text{O}_2\) with Fe\(^{2+}\) or Cu\(^{+}\) generates the hydroxyl radical, which is highly reactive and can induce membrane-associated oxidative stress that contributes to the dysfunction of the endoplasmic reticulum.\(^{104}\)
2 Methods

2.1 Infrared spectroscopy

2.1.1 Vibrational spectroscopy

Maxwell’s classical theory of electromagnetic radiation considers electromagnetic radiation as electric and magnetic fields oscillating in single planes at a right angle to each other. These fields are characterized by their wavelength $\lambda$ and frequency $\nu$. Frequency is described as a number of waves that pass a given point in a unit of time and wavelength is the distance from a crest of one wave to the crest of the adjacent wave. These two values are related by following equation $\nu = \frac{c}{\lambda}$.

In vibrational spectroscopy it is more common to use another unit: the wavenumber which is defined by the number of waves in a length of one centimeter and is given by the following formula: $\tilde{\nu} = \frac{1}{\lambda} = \frac{\nu}{c}$. This unit is linear with energy of the radiation.

During the late 19th century – beginning of 20th century it was proposed that the electromagnetic radiation can be considered as a stream of particles called photons with the energy given by the Bohr equation $E = h\nu$, where $h$ is the Planck constant and $\nu$ is the equivalent of the classical frequency. These photons may be absorbed or emitted by the molecules in which case the rotational, vibrational or electric energy of the molecules will change, with the amount given by the Bohr equation. Each absorbed or emitted photon moves the atom or a molecule from one discrete quantum energy level to another. Most of the vibrational energies within the molecule fall into the infrared region of the electromagnetic spectrum. Vibrational energy of a molecule is described by its vibrational frequency.

If one considers simple case of a molecule made up of two oscillating atoms joined by a spring/bond, then the vibrational frequency of such a bond can be described by the Hooke’s law: $\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$ where $\nu$ is vibrational frequency, $k$ – the classical force constant and $\mu$ – the reduced mass of the two atoms. This means that the frequency increases if the strength of the bond increases, or if the masses of the vibrating atoms decrease. In a multi-atom system one can distinguish between different kinds of vibrations. The most common ones are the stretching vibrations, where the bonds elongate and contract, and the bending vibrations, where the angle between the two bonds changes. The stretching vibrations can also be divided into symmetric and asymmetric modes.107–109
2.1.2 Infrared spectroscopy

For a molecule to absorb an infrared photon, an electric dipole moment of the molecule must change upon vibration. Meaning that there must be two partial charges $+q$ and $-q$, separated by distance $d$ that can be perturbed by the electric field of the incoming radiation. The infrared absorption is directly proportional to the change of the dipole moment, so the larger the change in the dipole moment, the stronger absorption will be observed.

A molecule consisting of $n$ atoms has a total of $3n$ degrees of freedom. In a non-linear molecule these include three rotational degrees and three translational degrees, while the rest are the vibrational normal modes. This means that in a typical non-linear molecule, there are $3n-6$ fundamental vibrations that will be observed in the spectrum.

As mentioned above as the infrared photon is absorbed by the molecule, it induces a transition to the next energy level. Transition from the ground state to the first energy state is considered fundamental and is allowed by selection rules, while transition probabilities from the ground state to higher energy states are equal to zero. However real molecules are slightly aharmonic and these kinds of transitions can occur. They are known as overtones. Simultaneous transitions of two vibrations from the ground state to a higher energy state are known as combination bands. As an example of this the majority of peaks in the near infrared region (NIR) arise from overtones of the X-H stretching modes, while the majority of peaks in the mid infrared region (MIR) are from fundamental vibrations.$^{110-113}$

2.1.3 FTIR spectrometer

Fourier transform infrared spectroscopy (FTIR) is a method that monitors the changes in molecular vibrations, as they absorb an infrared photon. A typical FTIR spectrometer consists of the following parts: an IR-source, a laser, an interferometer and a detector.

Figure 2.1 A schematic of a Michelson interferometer.$^{114}$
A typical interferometer used for the FTIR spectrometers is a Michelson interferometer, shown in figure 2.1. In such an interferometer the light emitted by the source is split by the beam splitter into two halves, one of which is then directed onto a fixed mirror and the other continues on to a moving mirror. Reflected by the mirrors, the beams are recombined at the beam splitter and directed out towards the detector. Due to the changes of the position of the moving mirror the recombined beams create an interferogram. From the detector data is sent to the computer, which performs a Fourier transform to convert the data from an interferogram to a spectrum. The laser is a monochromatic source that is used to coordinate the movement of the mirror, ensure the alignment of interferometer and data collection with wavelength precision.\(^{110}\)

Detectors commonly used for measuring the incoming IR light are MTC detectors where MCT stands for Mercury Cadmium Tellurium and DTGS detectors, where DTGS stands for Deuterated Triglycine Sulfate. MCT is a ternary semiconductor compound, which exhibits a wavelength cutoff proportional to the alloy composition. The actual detector is composed of a thin layer (10 to 20 \(\mu\)m) of HgCdTe with metalized contact pads defining the active area. Photons with energy greater than the semiconductor band-gap energy excite electrons into the conduction band, thereby increasing the conductivity of the material.

The nitrogen-cooled MCT detector has great advantages over detectors that operate at or near room temperature. For a given scanning time, an MCT detector will produce a spectrum with a noise level 10 to 100 times lower than the noise from a DTGS detector. This low noise has two important implications. Firstly it lowers the minimum detection limits for all compounds being measured, and secondly it widens the concentration range over which valid measurements can be made.\(^{110,115}\)

### 2.2 Attenuated total reflection

Infrared spectroscopy of biological systems is often performed in a transmission mode. This means that the IR beam of the spectrometer is passing through the sample and the transmitted IR intensity is measured. This mode is however sensitive to the water vapor present in the air in the sample compartment of the spectrometer and therefore requires extensive purging with dry air to minimize water vapor contributions. Another disadvantage of transmission mode is that the sample must be diluted with an IR transparent salt, pressed into a pellet or pressed to a thin film, prior to analysis to prevent totally absorbing bands in the infrared spectrum. Attenuated total reflection (ATR) is a technique alternative to the transmission mode infrared spectroscopy.
ATR operates by measuring the changes that occur in a totally reflected beam when the beam comes into contact with the sample. An infrared beam is directed onto an optically dense crystal with a high refractive index at a certain angle. The beam then penetrates a very short distance beyond the interface and into a less-dense medium before the complete reflection occurs (figure 2.2). This is called evanescent wave and is given by the following formula

$$d = \frac{\lambda}{2\pi n_2^2 \sin^2 \theta - n_2^2},$$

where $d$ is the penetration depth of the wave, $\lambda$ is the wavelength of incoming IR light, $n_2$ is the refractive index of the sample, $n_1$ is the refractive index of the crystal and $\theta$ is the angle of the incident light. The intensity of the evanescent wave is reduced by the sample in regions of the IR spectrum where the sample absorbs. Since the evanescent wave protrudes only a few micrometers into the sample, there must be good contact between the sample and the crystal. The IR beam then exits at the opposite end of the crystal and is passed on to the detector. One of the main benefits of this technique are that the samples require virtually no preparation beforehand. The second big advantage is that the IR beam passes through a constantly purged ATR unit and never comes in contact with the air around the sample, allowing a more water vapor free measurement.

### 2.3 Reaction-induced difference spectroscopy

A typical IR spectrum contains very detailed information about the system monitored; however an average size protein has about 25000 vibrational degrees of freedom, which leads to a very crowded spectrum with many overlapping bands. In the best cases, the effects of a protein reaction can be observed directly in the IR absorption spectra, but this does not happen very often. The most common solution to this problem is obtaining an associated IR difference spectrum. This is done by subtracting a spectrum of a protein in a state B, from a spectrum of a protein in a state A as shown in figure 2.3.
The resulting difference spectrum will originate only from the molecular groups that are directly involved in the reaction, while contributions from passive groups will cancel out. The absorbance changes observed in protein reactions are usually very small, in the order of 0.1% of the maximum absorbance, therefore measuring first an absorbance of protein in state A and then in state B and then subtracting one from another, does not allow very small changes to be observed. Instead it has been common to induce a protein reaction directly in the IR cuvette: the protein is prepared in state A, and its spectrum is measured, then the reaction is triggered and the protein proceeds to state B while the absorbance spectra are being recorded using time resolved methods. Reaction-induced IR spectroscopy can be performed for example by using a dialysis setup, by letting a ligand dialyze into the protein sample and thus starting the reaction or by using a light source that breaks a photosensitive “caged” compound and triggers the release of a compound of interest into the sample.

To be able to selectively observe individual functional groups, it is possible to shift the bands of interest from their original positions by means of isotope exchange. The main principle behind isotope exchange is based on the fundamental relationship between mass and vibrational frequency meaning that increased mass of an atom will lead to a band shift to a lower wavenumber. This can be monitored either by comparing spectra of labeled and unlabeled protein samples or by initiating isotope exchange directly in the IR cuvette.

2.4 Caged compound approach

Caged compounds have been widely used in FTIR to study biological reactions since the 1980’s. Caged compounds are photosensitive compounds that can release an “effector” molecule from an
inactive precursor by flash photolysis. In Papers I and II such compounds were used to induce a rapid pH jump in the sample and to initiate an enzymatic reaction, respectively.

pH plays an important role in biological systems as it can influence the structure and properties of molecules, however it is often difficult to achieve ~1ms time resolution while studying the pH induced changes in traditional ways. This problem can be solved by photolyzing a caged sulfate in the sample and monitoring the pH-induced structural changes with rapid scan FTIR.

\[
\text{OSO}_3^- \text{CH}_3 \text{CH}_3 \text{O}^+ \text{H}^+ + \text{SO}_4^- \text{NONO}_2
\]

**Figure 2.4 Photolysis reaction of 1-(2-nitrophenyl)ethyl sulfate (caged sulfate)**

1-(2-nitrophenyl)ethyl sulfate (NPE-sulfate) belongs to the group of 2-nitrobenzyl compounds that release a proton within 100 ns after the flash. The photolysis reaction, proceeds in several steps, shown in figure 2.4. Upon the flash the primary product is formed, a nitric acid, that is in rapid equilibrium with its conjugate aci-nitro anion. The life-time of the anion varies, depending on the pH. The proton is released upon the formation of the anion, causing a pH drop to as little as ~3.5. It is then picked up again and released together with a sulfate ion upon the final breakdown of the nitric acid, lowering the pH down to 2.

This approach in infrared spectroscopy has been used previously to study the pH-induced partial unfolding of myoglobin, as well as for the studies of the secondary structure changes upon the acidification induced aggregation of the Alzheimer’s Aβ(1-28).

**Figure 2.5 Photolysis reaction of 1-(2-nitrophenyl)ethyl ADP (caged ADP)**

1-(2-nitrophenyl)ethyl ADP is a nucleotide analog in which the terminal phosphate is esterified with a blocking group, making the molecule biologically inactive. The reaction proceeds like the NPE-sulfate reaction described above. Caged ADP and ATP were originally synthesized to probe the mechanism of action of the Na\(^+\) - K\(^+\) -pump ATPase. Subsequently they have been used to study the ouabain-sensitive ATP-ADP exchange reaction in human red cell ghosts and to analyze the influence of intra- and extracellular Na\(^+\) and K\(^+\), as well as an array of different ATP/ADP dependent reactions.
2.5 Dialysis setup for the ATR experiments

As mentioned above ATR provides water-vapor free measurements. ATR spectroscopy for protein studies has mostly been applied to protein films, deposited on the ATR crystal. To study the protein-ligand interactions, the ligands must be exchanged on the top of the deposited film. This way multiple experiments can be performed on the same protein sample. However, one can easily perturb the protein films in such experiments while washing or removal of media or repeated ligand additions, lowering the quality of the experiment. Closing off the sample compartment by a dialysis membrane allows for precise variations of the sample conditions. It also allows the use of proteins in solution, minimizing the preparation time. A schematic of the experimental setup is presented in figure 2.6.

![Figure 2.6 A simple sketch of a dialysis setup where R is the reservoir, O is an O-ring, M is a dialysis membrane, IR is an infrared beam, D is a diamond plate, P is a protein and L is the ligand.](image)

In such an experiment the dialysis membrane is first soaked for a few minutes in a buffer solution. Then it is fixed to the conical end of a cylindrical ligand reservoir, made of delrin, by an O-ring. A 5 µl drop of the protein solution is placed on the total reflection surface of the ATR and another 5 µl are deposited as a hanging drop on the dialysis membrane at the bottom of the reservoir. The reservoir is then placed into the opening of a thermostating block, which is secured tightly to the ATR bench. The height of the dialysis membrane over the crystal surface is determined by the height of the bottom conical part of the reservoir. The reservoir itself is pressed down by a screw cap, which screws onto a thread on the thermostating block. The O-ring makes tight contact with the walls of the conical dip in the metal block of the ATR unit, thus, a tight closure of the sample compartment is achieved without any pressure being exerted on the total reflecting surface of the ATR. Finally the reservoir is filled by the buffer in, which the protein of interest is dissolved. The solution is constantly stirred by a small mechanical stirrer, placed on the top of the reservoir.119
2.6 FTIR spectroscopy of proteins

IR spectroscopy provides a variety of information about the molecules and their environment. Anything that changes the electron distribution of the bonds within a molecule and in its surrounding will change the vibrational frequency of the bond and thus will appear in the IR spectrum.

In the case of proteins IR spectroscopy can provide information about the secondary structure content as well as the amino-acid environment. The most important group for the analysis of a peptide/protein is the peptide group, which has nine amide modes: A, B and I-VII. However for the conformational analysis of the polypeptides only Amide I and Amide II modes are used. The Amide I (1700-1600 cm\(^{-1}\)) mode originates mainly from the stretching vibration of the C=O bonds and to some extent from the stretching vibration of the C-N bonds. This vibration is highly sensitive to the conformations of the polypeptide backbone and therefore can be used to quantify the secondary structure content of the protein. The common bands assignments for the secondary structure are presented in Table 1 below.

**Table 1**: Assignment of amide I band positions of the secondary structure motifs. \(^{129}\)

<table>
<thead>
<tr>
<th>Secondary structure</th>
<th>Band position in H(_2)O</th>
<th>Average band position</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)-helix</td>
<td></td>
<td>1654</td>
<td>1648-1657</td>
</tr>
<tr>
<td>(\beta)-sheet</td>
<td></td>
<td>1633</td>
<td>1623-1641</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1684</td>
<td>1674-1695</td>
</tr>
<tr>
<td>Random coil</td>
<td></td>
<td>1647</td>
<td>1642-1652</td>
</tr>
<tr>
<td>Turns</td>
<td></td>
<td>1672</td>
<td>1662-1686</td>
</tr>
</tbody>
</table>

In certain cases it is even possible to distinguish between parallel and antiparallel \(\beta\)-sheets. As shown in Table 1 \(\beta\)-sheets have 2 bands in the region between 1600 and 1700 cm\(^{-1}\). It has been shown that the high wavenumber band (1690 cm\(^{-1}\)) appears for planar antiparallel \(\beta\)-sheets with large number of strands, but not for small or twisted antiparallel \(\beta\)-sheets. It is also weak or not observable for the parallel \(\beta\)-sheets. \(^{130,131}\)

The Amide II (1600-1500 cm\(^{-1}\)) band arises mainly from the bending vibration of the N-H bond. Although it is sensitive to the backbone conformation, the correlation with the secondary structure content is less evident in than in the amide I region. \(^{129}\)
Water plays an important role in all biological processes. In the mid-IR range it has two main bands: a broad band around 3400 cm\(^{-1}\), due to the stretching vibrations of the O-H group and a narrow band at 1640 cm\(^{-1}\), due to the bending vibration of the O-H group. The latter band overlaps with the amide I band of the protein, which causes a severe problem in the interpretation of the spectra, thus many protein experiments are performed in D\(_2\)O instead, which shifts the water absorption away from the amide I region.\(^{132-134}\)

### 2.7 Helper-enzyme approach

A considerable extension of the existing approaches for the reaction induced infrared spectroscopy would be the possibility to induce a second reaction after the initiation of the first reaction. For these means an approach using helper enzymes has been developed. Addition of a small amount of helper enzyme, implies that the signals coming from it will be very small and will not interfere with the signal form the compound of interest, while the reaction catalyzed by the helper enzyme would still be visible.

Initially the method was used for nucleotide removal and investigation of partial reactions of Ca\(^{2+}\)ATPase with help of two enzymes adenylate kinase (ADK) and potato apyrase. ADK generates ATP from ADP (2ADP \(\leftrightarrow\) ATP + AMP) and can therefore remove ADP generated by the Ca\(^{2+}\)ATPase reaction cycle. Potato apyrase was also used to remove the produced ADP as it cleaves of the inorganic phosphate (P\(_i\)) from ADP in presence of divalent cations (ADP \(\rightarrow\) AMP + P\(_i\)). Regeneration of ATP by ADK made it possible to make repeated measurements with the same sample under prolonged time, thus studying the Ca\(^{2+}\)ATPase reaction cycle without the interference of the cage signals from the flash photolysis.

The use of potato apyrase together with Ca\(^{2+}\)ATPase allowed the dissociation of ADP from the Ca\(_2\)E1ADP state to be observed. This revealed a conformational change of the enzyme that was opposite of the conformational change upon ADP binding, but smaller in magnitude, leading to the conclusion that ADP plays an important role in stabilizing the closed conformation of the Ca\(_2\)E1 state.\(^{122}\)

The method was extended even further by using the helper enzyme to perform an isotope exchange, which enabled a study of a phosphate group which is crucial to the activity of Ca\(^{2+}\)ATPase.

The experiment was started by the photolytic release of caged ATP with an \(^{18}\)O-labeled \(\gamma\)-phosphate. As the ATP phosphorylated the Ca\(^{2+}\)ATPase, phosphoenzyme E2P was obtained with its phosphate group labeled together with an ADP molecule. As the cycle progressed more ADP was created, which was utilized by the ADK to create an unlabeled ATP molecule that could enter the Ca\(^{2+}\)ATPase cycle. The binding of an unlabeled ATP created an unlabeled E2P
phosphoenzyme. Since both states were acquired in the same sample under the same measurement it was possible to compare the two states and distinguish and assign the phosphate bands.\textsuperscript{135}

In Paper II creatine kinase is characterized as a helper enzyme and used to study the transition between the $\text{Ca}_2\text{E1ADP} \rightarrow \text{Ca}_2\text{E1ATP}$ states of the calcium pump.

\section*{2.8 Circular Dichroism}

Circular Dichroism spectroscopy (CD) is an optical spectroscopic method, used to study chiral molecules (i.e. asymmetric molecules) such as proteins and peptides, which measures the difference in absorbance of the right- and left-circularly polarized light.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2_7.png}
\caption{Linear (A) and circularly (B) polarized light}
\end{figure}

Light, like all electromagnetic radiation can be divided into two components, electric field and magnetic field that propagate orthogonally to each other. By convention the polarization of light is described by specifying only the orientation of the electric field component at a point in space over the period of oscillation. In an unpolarized light, such as regular white light from the sun or a light bulb, the electric field oscillates in all possible directions that are perpendicular to the propagation direction. An illustration of linear polarized light and circular polarized light can be seen in figure 2.7. In linear polarized light the electric field is confined to one plane along the propagation direction. Circularly polarized light occurs when the amplitude of the oscillation of the field is constant, however the field vector rotates around the axis of propagation. Circularly polarized light can be separated into right- and left-polarized. These have different refraction indexes and different extinction coefficients, which leads to different velocities of traveling through optically active media as well as different absorbance. When the light passes through an optically active sample, the magnitudes of the counter rotating electric fields will not be the same, hence the light will not be longer circular polarized but rather elliptically polarized. This difference in magnitude gives rise to a CD signal and is measured in ellipticity, which is defined as the tangent of the ratio of the minor to major axis of the ellipse.\textsuperscript{136,137}
Figure 2.8 CD spectra of different secondary structure motifs in the wavelength of 180-250 nm. The α-helix has two negative bands at 222 and 208 nm, and a positive band at ~195 nm. The spectrum of a β-sheet has in general a negative band between 210-220 nm and a positive band between 195-200 nm. Turns have a strong positive band at ~205 nm and a negative band at ~190 nm, while a random coil has a negative band at around 200 nm.  

Chiral molecules that do not possess rotational symmetry will absorb left- and right-polarized light differently. In protein the basic chiral points are the alpha carbons participating in the peptide bond, however most groups are not optically active by themselves, but rather through the interactions with the asymmetric environment. This means that CD is very sensitive to the changes in molecular structure of the protein as presented in figure 2.8. Thus it is widely used to study the conformational changes and folding and unfolding of biomolecules such as peptides, proteins and nucleic acids.  

---

Figure 2.8
3 Results and discussion

3.1 Paper I – Formation of two different types of oligomers in the early phase of pH-induced aggregation of the Alzheimer’s Aβ(12-28) peptide

In paper I we have studied the early phase of the aggregation of the Alzheimer’s peptide Aβ (12-28) with both protected and unprotected ends with help of time-resolved infrared spectroscopy and circular dichroism. The Aβ (12-28) was chosen because it consists of residues important for transition to the β-sheet enriched conformation and the amyloid formation. For this study we used two variants of Aβ (12-28); one protected, where the ends of the peptide are uncharged, providing little electrostatic effect, and one unprotected, where the ends are charged, which might influence the aggregation process. To observe the influence of pH on the peptides we recorded CD as well as FTIR spectra. In the CD measurements we observed a strong negative band at ~196 nm for pH 2 and 9, indicating the presence of random coil. However this band was not observed at pH 4 and 6, meanwhile a negative band at 217 nm and a positive band below 210 nm appeared indicating the formation of β-sheets. FTIR absorption spectra showed two bands at 1616 cm\(^{-1}\) and 1685 cm\(^{-1}\) at pH 4 and 6 indicating formation of antiparallel β-sheets. These bands were not observed at pH 2 and 9.

We proceeded by initiating aggregation of the Aβ (12-28) by a pH jump that was induced by flash photolysis of NPE-sulfate, as discussed above in Section 2.4. With the protected Aβ-peptide we observed a broad positive band at 1627 cm\(^{-1}\) appearing after the first flash, which was assigned hydrogen carbonate due to CO\(_2\) dissolved in the alkaline solution in which the peptides were prepared. We also noted two positive bands at 1687 cm\(^{-1}\) and 1601 cm\(^{-1}\), which have been previously assigned to the photolysis of NPE. By evaluating the pH dependent bands of NPE sulfate we have concluded that the pH in the sample after the first flash was ~8.2. After the second flash the pH had dropped to 4-5. In this measurement we obtained a positive band at 1622 cm\(^{-1}\) that increased in absorbance and shifted to 1617 cm\(^{-1}\), within 12 min of the measurement. Here we again observed a photolysis band at 1687 cm\(^{-1}\), however when the photolysis contributions were subtracted from the spectra, a band still remained initially at 1684 cm\(^{-1}\) and later shifting to 1686 cm\(^{-1}\). There was also an increase in a broad negative band between 1660 cm\(^{-1}\) and 1630 cm\(^{-1}\), which indicated disappearance of the random coil structures in the sample. Interestingly when an extra ice-sonication was added during the preparation of the samples, upon the first flash we observed the same behavioral pattern as upon the second flash without the ice sonication. So ice-sonication seems to reduce the pH in the samples and convert the carbonate into hydrogen carbonate prior to the flash.

We also studied the unprotected Aβ (12-28). Here we found similar results as with the extra-sonicated protected peptide i.e. a positive band at ~1620 cm\(^{-1}\) that slowly increases in amplitude,
while shifting to lower wavenumbers. However in 30% of experiments we instead found the hydrogen carbonate band at 1627 cm$^{-1}$ upon both flashes. This band remained at its position throughout the whole experiment and was not influenced by the sonication of the samples.

In conclusion we have observed formation of $\beta$-sheets in the early stages of aggregation for both protected and unprotected A$\beta$ (12-28). The 1617 cm$^{-1}$ aggregates are antiparallel in their nature, which is indicated by the weak band at 1686 cm$^{-1}$ and form rapidly when the pH drops to ~8.2. The shift from 1622 cm$^{-1}$ to 1617 cm$^{-1}$, observed after the pH drop from 8 to 4-5, points towards further growth of the $\beta$-sheets, as the wavenumber is expected to shift down as the number of strands in a sheet increases.\textsuperscript{126}

### 3.2 Paper II - Use of Creatine Kinase to Induce Multistep Reactions in Infrared Spectroscopic Studies.

Paper II characterizes creatine kinase (CK) for its use as a helper enzyme in infrared spectroscopic studies. As discussed earlier infrared spectroscopy has high time resolution and contains large amount of information about the system, so one must use reaction induced infrared spectroscopy to achieve the necessary sensitivity. One of the ways to achieve this is to use so-called caged compounds. A considerable extension of existing approaches for reaction induced infrared spectroscopy would be the possibility to induce a second reaction after the initiation of the first reaction. One of the methods for achieving this is the use of a helper enzyme to modify the compound of interest. In previous work a helper enzyme ADK has been used to convert ADP, produced by phosphoenzyme of Ca$^{2+}$ATPse, back to ATP. ADK however uses 2 ADP molecules to produce one molecule of ATP, so ATP gets depleted rather fast. Therefore in this work we have chosen to use CK as a helper enzyme since it uses just one ADP molecule to produce one ATP.

First the enzymatic reaction of the CK was characterized. Each individual component of the enzymatic reaction (ATP, ADP, creatine and creatine phosphate) was measured by infrared spectroscopy and the reaction spectrum has been calculated. This calculated spectrum was then compared to the actual reaction spectrum measured by using the caged-ADP. Comparing the two spectra we were able to assign there positive bands at 917, 992 and 1243 cm$^{-1}$ to the formation of ATP, a negative band at 944 cm$^{-1}$ to consumption of ADP and two negative bands at 979 and 1614 cm$^{-1}$ to consumption of CP. The latter band was of specific interest, because in previous works it has been assigned to C=O of the peptide backbone of the CK. However we observed a similar band in our model spectrum and since the CK concentration in our sample is too low to give such a major contribution, we have assigned the 1614 cm$^{-1}$ band to the dephosphorylation of CP.
We have also studied the kinetics of the CK enzymatic reaction in presence of Mg$^{2+}$ and Ca$^{2+}$. Previous studies have shown that ions such as Mg$^{2+}$, Ca$^{2+}$, Mn$^{2+}$ and Co$^{2+}$ can be used as activators of CK, while other metals such as Ni$^{2+}$, Cr$^{2+}$ and Cd$^{2+}$ can cause inhibitory effects. Our results show that the enzymatic activity drops drastically in presence of 6 mM Ca$^{2+}$, which indicates that Ca$^{2+}$ also causes inhibition at mM concentrations.

We proceeded by evaluating the use of CK as a helper enzyme, with Ca$^{2+}ATP$se as the protein of interest. In our multistep experiment, ADP is released from caged ADP and the released ADP binds to the ATPase $Ca_{2}E1$ to form $Ca_{2}E1ADP$. Meanwhile CK converts ADP to ATP which has a higher binding affinity to the ATPase than ADP and therefore binds to the Ca$^{2+}ATP$se more readily and phosphorylates the enzyme $Ca_{2}E1ADP$ to $Ca_{2}E1P$.

The bands observed in the Amide I region indicate conformational changes of the Ca$^{2+}ATP$se upon nucleotide binding. The positive band at 1628 cm$^{-1}$ is indicative of the binding event and has been tentatively assigned to the conformational changes in the beta-sheet of the phosphorylation domain along with the 1693 cm$^{-1}$ band. The positive shoulder at 1619 cm$^{-1}$ is considered a marker band for the binding of ADP to Ca$^{2+}ATP$se. In the spectra presented in this paper there are two new negative bands at 1607 cm$^{-1}$ and 977 cm$^{-1}$, which indicate the consumption of CP in the CK reaction. In these spectra no positive bands of ATP were observed because ATP was immediately consumed by the Ca$^{2+}ATP$se, however another positive band appeared at 1070 cm$^{-1}$ indicating the formation of free phosphate.

Absorbance changes presented in this paper are those from the conformational transition from $Ca_{2}E1ADP$ to $Ca_{2}E1P$. We have compared these to the previously published the spectrum of the phosphorylation reaction: $Ca_{2}E1ATP$ to $Ca_{2}E1P$. These reactions have the same end state but different starting states, therefore the differences between these are due to the effects of binding of the $\gamma$-phosphate of ATP.

When comparing the present spectrum of $Ca_{2}E1ADP$ to $Ca_{2}E1P$ to the spectrum of $Ca_{2}E1ATP$ to $Ca_{2}E1P$ it became clear that the positive bands at 1717, 1653, and 1550 cm$^{-1}$, found at similar positions in both spectra arise from the specific structural features of $Ca_{2}E1P$. The 1717 cm$^{-1}$ band has been attributed to the phosphorylation of Asp351, the 1653 cm$^{-1}$ band to a conformational change of an $\alpha$-helix and the 1550 cm$^{-1}$ band to a conformational change reflected by the amide II absorption. Some differences were observed as well. The negative bands at 1689 and 1627 cm$^{-1}$ in the $Ca_{2}E1ATP$ to $Ca_{2}E1P$ spectrum were replaced by positive bands at 1693, 1678 cm$^{-1}$ and 1631 cm$^{-1}$ in the $Ca_{2}E1ADP$ to $Ca_{2}E1P$, and a positive band at 1639 cm$^{-1}$ was replaced by a negative band at 1643 cm$^{-1}$. From this combination of negative and positive bands we have suggested that $\gamma$-phosphate binding induces a downshift of a $\beta$-sheet band from ~1640 to ~1630 cm$^{-1}$, which is partially reversed upon formation of $Ca_{2}E1P$. 
3.3 Paper III – Detection of ligand binding to proteins through observation of hydration water

In paper III we have presented an observation that can potentially lead to the development of a general method for detecting ligand binding to proteins by monitoring the changes in water absorption. The main advantage of this method is that it is label-free, does not need any special sample preparation and it is based on a universal event that takes place in all binding processes. In order to detect a change in water absorption we have used our dialysis accessory for the ATR-FTIR, as described in Section 2.2. The protein that was used as a model system was pyruvate kinase (PK) together with its ligands phosphoenolpyruvate (PEP) and several divalent and monovalent ions that are required for binding and activity of PK. In this paper we concentrate mainly on the low wavenumber edge of the OH stretching band of water (~3000 cm\(^{-1}\)), however we use the region below 1800 cm\(^{-1}\), studied by us previously, to confirm our statements.

The complex formation between PK and PEP is best observed by monitoring the double band at 1230-1210 cm\(^{-1}\). We have previously assigned the 1214 cm\(^{-1}\) band to bound PEP, while the 1230 cm\(^{-1}\) is due to the free PEP. As we have observed in our spectra the 1214 cm\(^{-1}\) band rises in the early stages and eventually saturates by the end of the experiment, while the 1230 cm\(^{-1}\) band rises later and continues to rise until the end of the experiment. We have calculated the difference between these two bands as a function of time and obtained a kinetic plot with two distinct phases: the initial negative drop due to complex formation and a later positive rise due to increase of free PEP concentration in the sample. As we turned towards the water absorption region we observed a similar two-phased behavior. Upon the first addition of PEP, the water absorption in the 3000 cm\(^{-1}\) region decreased as the 1214 cm\(^{-1}\) band of bound PEP increased. As the binding sites became saturated and the 1230 cm\(^{-1}\) band appears, the water absorption starts to increase. Thus we attributed the decrease of the absorption in the water region to the binding of PEP to PK and it’s increase to an increase in the free concentration of PEP. To confirm our assumptions we repeated the experiment with a lower PEP concentration. In this experiment we observed very small decrease in the water absorption upon first addition of PEP to PK. The estimated concentration of PEP that reached the sample compartment was calculated to be 10 times less than what is needed to saturate the binding sites. Lack of the signal in the ~3000 cm\(^{-1}\) region was confirmed by the lack of the conformational change bands in the 1700-1600 cm\(^{-1}\) region. During the second addition however the marker bands for the complex formation appeared and accordingly the water absorbance decreased.

We then continued to test our methods by control experiment of adding PEP to PK in absence of Mg\(^{2+}\) and K\(^{+}\), ions that are necessary for binding. This experiment did not produce a negative signal in the water absorbance region, neither did we observe the signals of complex formation in the region below 1800 cm\(^{-1}\).
Water absorption can also be used to detect binding of ligands as small as ions, as we have shown in our final experiment in Paper II. The case investigated was Mg$^{2+}$ binding to PK. It is known that Mg$^{2+}$ is necessary for PK function and the binding of PEP as discussed in the publication. During the first addition we observed first a decrease in the 3000 cm$^{-1}$ absorbance, followed by an increase. This was confirmed by the changes in the 1800 cm$^{-1}$ region, which indicated a change in the conformation of the protein upon ion binding. The second addition revealed only broad bands in the amide I and II regions, which were due to protein settling on the surface of the ATR setup. In the water region only positive signals were observed.

The observed signal appears in the region where OH and NH groups of the protein may have some absorption. Dried proteins show a band at 3070 cm$^{-1}$, which has been assigned to the amide B vibration, the lower wavenumber component of a Fermi resonance doublet involving NH stretching and amide II vibration, which is expected to be effected by the changes in the hydrogen bonding pattern. However, amide B band has considerably lower intensity than the amide I and amide II bands. The amplitude of the band observed by us is comparable to the amide I and amide II bands, so it seems unlikely that this change would be caused by the NH stretching. We have also considered the idea that the 3000 cm$^{-1}$ band comes from the OH groups in the protein. However the OH groups are not as abundant in the proteins as the NH groups, so their absorbance would be even lower. One would also expect a much narrower band for the isolated OH, while we observe a very broad decrease in absorbance between 2800 and 3100 cm$^{-1}$.

3.4 Paper IV: Interaction between the Transcription Factors of the E2F Family and DNA Studied with Infrared Spectroscopy

The interaction between the transcription factors (TFs) and the DNA is a main part of the gene regulatory networks that control cell development, cellular processes and responses to environmental stimuli. However how the TFs recognize their binding sites and the mechanism of binding is not yet well understood. In paper IV we have studied the interaction between TFs of E2F family with their DNA targets. E2Fs are especially interesting because they seem to control both cell profiliation and apoptosis, and understanding their role under abnormal cellular conditions can lead to development of new tumor-specific therapies. The family members E2F1, which can induce either cell division or apoptosis, and E2F8, which controls liver-cell polyploidization, have been chosen for this work.

To be able to interpret the complex formation spectra we have measured the spectra of DNA and TFs separately, then manually added these spectra together and compared to the measured complex spectra. The most striking change in both proteins of interest was the decrease of the 1651 cm$^{-1}$ peak and an appearance of the 1640 cm$^{-1}$ shoulder in the complex spectrum compared to the calculated spectrum. The 1651 cm$^{-1}$ is assigned to the regular α-helices in the protein, meaning that in the complex sample these disappeared, while the appearance of the 1640 cm$^{-1}$
shoulder could be indicative of the formation of long or twisted α-helices. In the amide II region we observed a similar decrease in the 1550 cm⁻¹, which is also assigned to the regular α-helices.

The second interesting difference between the measured complexes and the calculated ones was the 1608 cm⁻¹ peak. In case of E2F1-DNA complex this band showed a decrease in absorption, while in the E2F8-DNA complex it showed an increase. The 1608 cm⁻¹ peak is due to the adenine bases present in the DNA. The change in this peak indicated that adenines take active part in the complex formation between each TF and its corresponding DNA. The E2F8-DNA complex also showed an increase in the 1663 cm⁻¹ band, which is assigned to the stretching vibrations of thymine. This band tends to decrease upon the duplexation of DNA, meaning that in our samples the interaction between A-T base pairs was weakened.

Relative intensities of the antisymmetric stretching (1224 cm⁻¹) and the symmetric stretching (1086 cm⁻¹) of the PO₂⁻ of the DNA backbone were altered upon complex formation, indicating distortion in the DNA double helix upon its interaction with the proteins. This conclusion was also confirmed by the CD spectra that were measured and constructed in the same fashion as the FTIR spectra.
4 Summary and future plans

The work presented in this thesis gives insight on the versatility of IR spectroscopy in its applications for studying of biological processes. This technique is fast, cost efficient and it allows investigation of small molecules and complex biological systems alike, under physiological conditions. The main advantage and disadvantage of IR spectroscopy is the large amount of information that is present in a single IR spectrum, making it hard to interpret. However over the years several methods have been developed that help filter out the relevant information from the irrelevant.

Papers I and II show the usefulness of the caged-compound approach for investigation of small (Alzheimer’s Aβ peptide) and large (creatine kinase) systems. Alzheimer’s disease (AD) is one of the most frequent neurodegenerative disorders and it has been a topic for ongoing research for many years. Recent studies have shown that the source of the neurotoxicity in AD is not the insoluble fibrils of the Aβ-peptide but the oligomers and protofibrils that occur on the intermediate stage of the fibril formation. We have used time-resolved FTIR in combination with pH jump induced by the release of a proton from a caged sulfate to study the formation of Aβ (12-28) oligomers. Our results emphasize the complexity of the aggregation process, as the peptide seems to aggregate in a similar manner despite the difference in charge of the two versions of the peptide used. Further work on this topic could include isotope labeling of individual amino acids in the peptide and monitoring their participation in the formation of the oligomers. It is also possible to further lower the time resolution of FTIR experiments by lowering the amount of caged sulfate used in the sample so the pH changes will not be so rapid and by increasing the data recording speed of the instrument. This would give an even better insight on what happens within the peptide on the very early stages of oligomer formation.

Another application of the caged sulfate is to investigate a possibility of two separate ion paths existing in the sarcoplasmic reticulum Ca^{2+}ATPase (ongoing project). Ca^{2+}ATPase mediates muscle relaxation by transporting Ca^{2+} ions back into the sarcoplasmic reticulum, against the concentration gradient, while counter-transporting 2-3 H^{+}. The Ca^{2+} channel in the protein has been blocked by removing the calcium from the buffer solution and introducing thapsigargin, a non-competitive inhibitor of Ca^{2+}ATPase, which keeps the Ca^{2+}ATPase from adopting the conformation that binds Ca^{2+} ions. Following this we induced pH jumps in the sample, with help of caged-sulfate. This should show whether or not the protons continue having access to the ion binding site and therefore prove whether they are using the same channel as Ca^{2+} ions do or a different one. Preliminary results show protonation of aspartates between pHs 6 and 7, however there is still not enough data to conclude whether the signals come from the aspartates in the Ca^{2+}-binding site or the ones on the surface of the protein.

There is a wide variety of commercially available caged compounds with caged-sulfate being just one example. Another example used in this thesis is caged-ADP. This caged nucleotide was
used in paper II to monitor the enzymatic reaction of creatine kinase (CK), which catalyses the formation of ATP and creatine from ADP and creatine phosphate. The main aim of this experiment was however, to characterize CK as a helper enzyme. The helper enzyme approach has been developed to monitor the partial reaction in the Ca$^{2+}$-ATPase reaction cycle but is applicable to most nucleotide dependent proteins. In the initial experiment adenylate kinase (ADK) was used to regenerate ATP from ADP produced during the phosphorylation of Ca$^{2+}$-ATPase. However the ADK requires two ADP molecules for the production of one ATP, thus we proposed modifying the method by using CK as a helper enzyme, as it requires only one ADP molecule per one ATP and the production of ATP continues while there is sufficient creatine phosphate to fuel it. As another major advantage of using CK, it turned out that creatine phosphate has a distinct band that does not interfere with the bands coming from the Ca$^{2+}$-ATPase and can therefore be used to follow the CK enzymatic reaction while monitoring the Ca$^{2+}$-ATPase reaction cycle. Finally using this setup we were also able to determine the structural differences between two very similar partial reactions Ca$_2$E1ADP → Ca$_2$E1P and Ca$_2$E1ATP → Ca$_2$E1P. Comparing the two states resulted in the conclusion that the binding of ATP comes with a large conformational change of the β-sheet in the phosphorylation domain. Upon the phosphorylation of the ATPase, the β-sheet relaxes back to a structure that is intermediate between that in the ADP bound state and ATP bound state.

The second part of this thesis looks in close detail at the protein-ligand binding events. A ligand is a substance that binds and forms a complex with a protein to serve a biological purpose. Ligands come in all shapes and forms, from ions and small molecules to nucleic acids and proteins. In paper III we used ATR-FTIR spectroscopy for developing a novel method for detecting protein-ligand binding, by observing the changes in the water absorption. We have established that the water absorption decreases upon the binding of a ligand to protein, which indicates a change in the vibrations of the OH bond as the water molecules go from the bound state to bulk water. We were also able to show that this change is detectable even in the case of ions binding to the protein in question. This observation has high potential for drug development as well as for basic research because it can lead to a general method for detecting binding events that is label-free, works with both binding partners being in aqueous solution and is based on a universal process that takes place in all binding events. However further improvement is needed, for example increasing the light intensity in region of OH stretching vibrations would greatly increase the sensitivity of the method.

ATR-FTIR spectroscopy was also used in paper IV to determine the binding of DNA to the transcription factors (TFs) of the E2F family. The interaction between these TFs and the DNA is a main part of the gene regulatory networks that control cell development, cellular processes and responses to environmental stimuli. However how they recognize their binding sites and the mechanism of binding is not yet understood. By studying the formation of the E2F-DNA complexes by IR, the changes in the secondary structure of the proteins, as well as the distortions to the DNA have been observed. The observed changes are in line with the predicted changes of
the E2F-DNA interactions. However more experiments need to be made to fully characterize the binding, such as titration experiments, isotope exchange experiments and caged-DNA experiments. We have also discovered that the E2Fs have a tendency of aggregating in room temperature (unpublished work), but the reason behind this is unknown.
5 Acknowledgements

The time of my PhD studies is coming to an end and I would like to take this moment to thank all the people who have shared this time with me and helped me along the way.

I would like to start off by saying a big thank you to my supervisor Andreas, for giving me an opportunity to work in your lab and introducing me to the world of infrared spectroscopy, for constant support even through the dry periods or holding me back when I got too enthusiastic. I think that at one point during all these years, you have told me that I am trying to do way too many projects at the same time, but it worked out anyway.

I would also like to thank my co-supervisor Lena, for all the help and support in these years. Your door has always been open for me and you have always found an encouraging word to say.

I would like to say a special thank you to Astrid, for all the help and advice during my last project.

A big thank you to Stefan, for taking such good care of us - PhD students and always coming up with good suggestions.

Also I would to thank past and present members of our group: Saroj, for teaching me everything I know about IR, our long discussions about nothing and everything when it was just the two of us in the group. Maria who connected our group with the mysterious and apparently very fun other C4 corridor. Eva-Lisa, my former roomie, for relevant discussions about CaATPase and less relevant discussions about games, cats, life, oh! and introducing me to Grills eat-as-much-meat-as-you-can-fridays. Chenge and Ljubica – tea/orange/chocolate/foosball companions, not quite sure how I survived without you guys, Julia the old-new member of our group, thank you for all the long lunches and support through this last year. Maurizio, I have to admit our group has been rather quiet without you and your passion for research is very admirable (but slightly troublesome at times). Paulami, thank you for the collaboration on the Aβ paper and introducing me to the delicious indian food and motto: there is a journal for any paper. Marta, Carolina, Eva, Stefanie, Markus, Henrik, Atie and many more who joined our group during these years, thank you guys and girls for making the atmosphere in our lab so pleasant and friendly.

I would also like to thank Sebastian and Juri, for all your help with the CD interpretations and NMR experiments and your support these last couple of months. Sebastian, I’m still not sure that you’re not responsible for that famous computer crash in our lab. A very special thank you to Fatemeh, for putting up with my countless questions about CD and fluorescence instruments, and chemicals for the labs, and instructions, and more questions about instruments, because even 5 years later, I’m still not sure how to turn on those machines. To Victor and Johannes, you guys are hilarious when together and very friendly and even a bit funny when one at a time.
Thank you guys, for these past 2 years of teaching together. Axel, Weihua, Jobst, Pontus, Scarlett, Anna, Sofia, Jesper, the biophysics corridor wouldn’t be right without you guys.

The girls and guys from Jan Willems group: Mirjam, Susan, Anna, David, Thomas, thank you guys from teaching me the proper way to do pubs and several years of very fruitful collaboration in the area. And the foosball gang Johan, Diogo, Dan, Jörg, Kiavash, and Ben, I probably lost all the skills you guys have so patiently taught me. Emilie, Ann-Louise, Patricia, Karin, Patrik, Inger and Ing-Marie, those long-long hours of exam correcting would be horrible without you. Peter, Walter, Maria, Stephen, Candan, Rickard, Minttu and Narges, thank you guys for making the time of doing my PhD so much fun. Rob, I really appreciate all the advice you have given me over the time I’ve known you, and I’ve kept some of the very long sentence in the text just for you.

I would like to say a big thank you to Torbjörn for technical support, with instruments and computers and discussions during lunches. There seems to be no limit to the things you know and its always very interesting talking to you. Peter, thank you so much for dealing with all my computer problems, I can’t believe how much patience you have had with the computers in our lab. And Håkan, thank you for all your help with the technical stuff that Torbjörn doesn’t have time for. And Eddie, thank you for helping me build my very first, very own experimental setup.

Also a special thank you to Haidi, Maria, Malin, Lotta, Anita, Alex and Ann for all the help with administration, and the entertaining department get-togethers and everything else you do for us.

I would also like to say thank you to Henrik, for putting up with my whining and complaining and reading the thesis and skiing and movies. Thank you so much for all the encouragement and support. And to Pär, my organic chemistry supplier, it’s your turn soon! And Olga, Ikea soon, or just a nice quiet evening with salad, cheese, wine and a guitar? And everyone else from the russian gang and the volleyball gang, who has been with me these years.

And finally a big thank you to my family, Patrik, the master problem-solver and experiment-designer, my father, for the eternal question “How are those articles doing?” and especially my mother, for all the patience and encouragement and advice and teaching me the basics of chemistry and biochemistry and crystallography and how to make beautiful pictures in PyMol 😊
6 Sammanfattning på svenska

Infraröd (IR) spektroskopi har funnits sedan upptäckten av Friedrich Wilhelm Herschel år 1800. Den bygger på det faktum att bindningar mellan atomer inom molekyler vibrerar med frekvenser som kan beskrivas med fysikens lagar. Det enklaste sättet att beskriva vibrationsfrekvensen av en bindning mellan atomer är med hjälp av en tvåatomig oscillator:

\[ \nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \]

där \( \nu \) är vibrationsfrekvens, \( k \) - den klassiska kraftkonstanten och \( \mu \) - den reducerade massan av två atomer. Detta innebär att frekvensen ökar om styrkan av bindningen ökar, eller om massan av de vibrierande atomerna minskar. I ett fleratomssystem kan man skilja mellan olika typer av vibrationer. De vanligaste är sträckningsvibrationer, där bindningarna sträcs och drar ihop sig och böjningsvibrationer, där vinkeln mellan de två atombindningarna ändras. Sträckningsvibrationer kan också delas in i symmetriska och asymmetriska.


Det arbete som presenteras i denna avhandling ger insikter om mångsidighet av IR-spektroskopi i sina tillämpningar på en mängd olika biologiska system samt utveckling av nya metoder för undersökning av mer komplexa biologiska händelser. Denna teknik är snabb, kostnadseffektiv och möjliggör undersökning av små molekyler samt komplexa biologiska system, under fysiologiska förhållanden. Den största fördelen och nackdelen med IR-spektroskopi är den stora mängd information som ett enda IR-spektrum innerhåller, vilket gör den svår att tolka. Men under årens lopp har flera metoder utvecklats, som hjälper till med att filtrera bort relevant information från irrelevant.

Artiklarna I och II visar användbarheten av så kallad ”caged”-molekylföreningsmetoden för undersökning av små (Alzheimers Aβ-peptid) och stora (kreatinkinas) system. Alzheimers sjukdom (AD) är en av de vanligaste neurodegenerativa sjukdomarna och har varit ett ämne för pågående forskning i många år. Nya studier har visat att orsaken till neurotoxicitet i AD inte är de olösliga fibriller av Aβ-peptiden men oligomerer och protofibriller som uppkommer på mellanstadiet av fibrillbildning. Vi har använt tidsupplöst FTIR i kombination med pH-hopps inducerad genom frisättning av en proton från en ”caged”-sulfat för att studera bildandet av Aβ (12-28) oligomerer. Våra resultat understyker komplexiteten av aggregeringsprocessen, eftersom peptiden verkade aggregera på ett liknande sätt, trots skillnaden i laddning av de två
versionerna av den peptid som användes. Ytterligare arbete på detta ämne skulle kunna omfatta isotopmärkning av enskilda aminosyror i peptiden och övervaka deras deltagande i bildandet av oligomorer. Detta skulle ge en ännu bättre inblick i vad som händer inom peptiden på en mycket tidig fas av oligomerbildning.

Det finns en mängd olika kommersiellt tillgängliga ”caged” molekylföreningar, där ”caged”-sulfat bara är ett exempel. Ett annat exempel som används i denna avhandling är ”caged”-ADP. Denna nukleotid användes i artikel II för att studera en enzymatisk reaktion av kreatinkinas (CK), som katalyserar bildningen av ATP och kreatin från ADP och kreatinfosfat. Det främsta syftet med detta experiment var dock att karakterisera CK som ett hjälpzym. Hjälpzymmetoden har utvecklats för att övervaka den partiella reaktionen i Ca$^{2+}$-ATPase reaktionscykel, men kan tillämpas för att studera de flesta nukleotidberoende proteinerna. I det ursprungliga experimentet användes adenylatkinas (ADK) för att regenerera ATP från ADP som produceras under fosforyleringen av Ca$^{2+}$-ATPase. ADK kräver dock två ADP molekyler för framställning av en ATP, så vi föreslog att modifiera metoden genom att använda CK som hjälpzym, eftersom att det bara kräver en ADP-molekyl per ATP och produktionen av ATP fortsätter medan det finns tillräckligt med kreatinfosfat att underhålla den. Som en annan viktig fördel med att använda CK, visade det sig att kreatinfosfat har ett distinkt band som inte sammanfaller med de band som kommer från Ca$^{2+}$-ATPase och kan därför användas för att följa CK:s enzymatiska reaktion medan Ca$^{2+}$-ATPase reaktionscykel övervakas. Slutligen, genom att använda denna uppstånd, kunde vi också följa de strukturella skillnaderna mellan två mycket likartade partiella reaktioner, Ca$_2$E1ADP → Ca$_2$E1P och Ca$_2$E1ATP → Ca$_2$E1P. Jämförelsen mellan de två tillstånden ledde till slutsatsen att bindningen av ATP kommer med en stor konformationsförändring av β-flak i fosforyleringsdomänen. Efter fosforylering av ATPas, slappnar den β-flaken av tillbaka till en struktur som ligger mellan det i ADP-bundet tillstånd och ATP-bundet tillstånd.

ljusintensitet i området för OH-sträckningsvibrationer kraftigt öka känsligheten av metoden.

ATR-FTIR spektroskopi användes också i artikel IV för att bestämma bindning av DNA till transkriptionsfaktorer (TF) från E2F familjen. Sampelet mellan dessa TF och DNA är en viktig del av det genregleringsnät som styr cellutveckling, cellulära processer och reaktioner på stimuli från omgivningen. Men hur de känner igen sina bindningsställen och mekanismen för bindning är ännu inte klarlagd. Genom att studera bildningen av E2F-DNA komplex med IR, har vi kunnat karakterisera förändringar i den sekundära strukturen av proteinerna, liksom ändringar av DNA helix. De observerade förändringarna är i linje med de förutsagda förändringarna av E2F-DNA-interaktioner. Fler experiment behöver dock göras för att helt karakterisera bindningen, såsom titreringsexperiment, isotopbytesexperiment och ”caged”-DNA experiment.
7 References


(91) http://www.omalzheimer.se/fakta_om_alzheimer/.


56


