UNIVERSITÀ DEGLI STUDI DI PADOVA

CHARLES UNIVERSITY IN PRAGUE

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE DEL FARMACO

CORSO DI LAUREA MAGISTRALE

TESI DI LAUREA

STRUCTURAL CHARACTERIZATION OF PROTEIN-DNA RESPONSE ELEMENT TERNARY COMPLEX

Relatore
Prof.ssa Alice Sosic
Università degli Studi di Padova
Correlatore
RNDr. Petr Novák, Ph.D.
Charles University in Prague

Laureanda
Federica Serlini
Matricola 2018992

Anno accademico
2022-2023

"And once the storm is over,
you won't remember how you made it through,
how you managed to survive.
You won't even be sure,
whether the storm is really over.
But one thing is certain.
When you come out of the storm,
you won't be the same person who walked in.
That's what this storm's all about."

– Kafka on the Shore, Haruki Murakami

Abstract

English version

Transcription factors are key elements of several biological processes such as control of cell cycle progression, differentiation of cells during development, immune response, or maintenance of intracellular metabolic balance. Their ability to regulate the transcription of specific sets of genes makes the cells responsive to diverse stimuli. They are dynamic elements participating in the last step of a signalling cascade with the ability to bind the DNA cooperating with other transcription factors and with numerous cofactors. They are elements of interest since dysregulation of the transcription leads to many diseases including cancer. For this reason, the complicated network of transcription factors and cofactors and their interactions is an active research field.

In this work, two transcription factors TEAD1 and FOXO4 were analysed to retrieve some information about their structural change in conformation when they are interacting with the DNA response element and when they are in a mixture together either with or without the DNA. Based on previous evidence of a possible interaction between TEAD1 and FOXO4, the experiment was set on the use of isotopic cross-linkers conjugated with LC-MS analysis.

The two proteins were studied in different conditions in the presence of the isotopic-

ally labelled DSA cross-linking agents. The positions of the cross-linked amino acids combined with the known cross-linking agent length that imposes a distance constraint between functional groups within a protein or a complex gave information on the 3D structure of TEAD1, FOXO4 and their complex. Moreover, the isotopically labelled DSA cross-linker allowed the quantification of conformational changes of proteins under different conditions and the protein interaction dynamics.

It was identified how the proteins change conformation when they are in the only presence of DNA. In the case of TEAD1, 7 cross-links were depicted and among them, 3 highlighted a change in the structure of the protein in the presence of the DNA. On the contrary, for FOXO4 3 cross-links were identified which are slightly changing when the transcription factor interacts with the DNA. Furthermore, when TEAD1 and FOXO4 are mixed in the presence of the DNA, it was noticed that they interact and change conformation to adjust themselves in the formation of a ternary complex TEAD1-FOXO4-DNA and a significant presence of a cross-link confirmed an interaction of TEAD1 and FOXO4 to form the complex with the DNA.

Italian version

I fattori di trascrizione sono elementi fondamentali alla base di diversi processi biologici, tra cui la regolazione del ciclo cellulare e del differenziamento cellulare durante lo sviluppo, il controllo della risposta immunitaria e il mantenimento dell'omeostasi cellulare. La loro capacità di regolare la trascrizione di specifici geni permette alla cellula di rispondere a diversi stimoli esterni. Sono elementi dinamici che intervengono alla fine della cascata di trasduzione quando, attivati da diverse reazioni chimiche, legano il DNA e ne permettono la trascrizione. L'interazione con il DNA avviene per la maggior parte dei casi in collaborazione con altri fattori di trascrizione o numerosi cofattori. Alterazioni di questo meccanismo portano all'insorgenza di molteplici patologie tra cui il cancro. Per questo motivo, l'intricata rete in cui partecipano i fattori di trascrizione con i cofattori e le interazioni che creano è un campo della ricerca attivo.

In questa tesi, due fattori di trascrizione, in particolare TEAD1 e FOXO4 sono stati analizzati per ottenere informazioni riguardo cambi conformazionali che avvengono nell'int erazione con il DNA e nell'interazione proteina-proteina in assenza o in presenza del DNA. Basandosi su precedenti evidenze di una possibile interazione tra TEAD1 e FOXO4, è stato deciso di svolgere l'esperimento coniugando una reazione di cross-linking con l'analisi LC-MS.

Le due proteine sono state analizzate in condizioni differenti utilizzando DSA cross-linker marcato con isotopi nella sua versione leggera e pesante. Conoscendo la posizione degli aminoacidi coinvolti nel legame con il cross-linking agent e sapendo che il cross-linking agent ha una lunghezza definita che determina la distanza tra i gruppi funzionali che lega, é stato possibile ottenere informazioni su come TEAD e FOX04 riarrangiano

la loro conformazione Oltretutto, DSA cross-linker ha permesso, essendo marcato con isotopi, la quantificazione dei cambiamenti conformazionali avvenuti nelle proteine nelle condizioni studiate e le dinamiche di interazione delle proteine.

Sono stati caratterizzati i cambi conformazionali delle proteine nella sola presenza del DNA. Nel caso di TEAD1, 7 cross-links sono stati individuati e tra questi, tre in particolare indicano che TEAD cambia conformazione nell'interazione con il DNA. Al contrario, per FOXO4 i tre cross-links individuati rimangono quantitativamente invariati anche quando il fattore di trascrizione lega il DNA Inoltre, quando TEAD1 e FOXO4 sono in soluzione insieme, nell'aggiunta del DNA è stato notata un'interazione tra le due proteine per formare un complesso ternario TEAD1-FOXO4-DNA. La presenza significativa del cross-link ha confermato quest'interazione.

Contents

List of Acronyms

AcN Acetonitrile.

AKT/PKB Serine–threonine kinase AKT/protein kinase B.

DBD DNA binding domain.

DMSO Dimethyl sulfoxide.

DSA Disuccinimidyl adipate.

DSBU Ureido-4,4´-dibutyric acid bis.

DTT Dithiothreitol.

FOX Forkhead box.

HEPES 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

IPTG Isopropyl-β-D-thiogalactopyranoside.

LDS-PAGE Polyacrylamide gel electrophoresis in the presence of litium dodecyl sulfate.

MES 2-(N-morpholino)ethanesulfonic acid.

PPIs Protein-protein interactions.

Sd Scallop gene.

TAZ Transcriptional coactivator with PDZ-binding motif.

TEAD Trascriptional enhancer associated domain.

TEFs Transcriptional enhancer factors.

TF Transcription factor.

YAP Yes/src associated protein.

Chapter 1

Introduction

1.1 Protein at the base of life

All biological processes are carried out by what are called macromolecules. This macro group includes polysaccharides, lipids, nucleic acids, and proteins. Each of these biomolecules fulfills its own unique function and occupies different spaces within a cell or a system. Since dynamics is at the base of all the biological mechanisms, these macromolecules move and interact with one another to better accomplish their duties. Among these biomolecules, proteins play a fundamental role in all the cellular functions such as structural organisation, metabolism, transport, information processing, and communication [1]. The reason why they have many roles is that they rarely act alone. They are prompt to create networks and physiochemical connections with other biomolecules or other proteins. Because of that, the study of protein interactions has become increasingly important [2–4]. Especially, it has been found that protein-protein interactions are essential to cellular life; these multi-protein complexes or molecular machines are involved in core processes such as transcription, translation, cell-cell adhesion signalling cascades, cell cycle control, and protein synthesis and degradation.

1.1.1 History of discoveries

The concept of protein was first introduced by Mulder and Berzelius, two scientists whose assumption was to have found a primordial molecule characterizing the organic mater-

ial. A few years later, pepsin was discovered after noticing digestive activity in animal stomach and a variety of fermentations were noticed in different organisms [5]. Also, the description of trypsin isolation and further studies on trypsin reported the presence of an inhibitor able to combine with the enzyme and to create an inactive compound but in a reversible way [6]. The same ability was identified between an antigen and an antibody. These discoveries brought up the concept of protein-protein interactions. At this point the definition of enzymes was 'the chemical bases of life' but still something different from a protein. Even though physical interactions seemed to have a fundamental role, it first urged researchers to better understand the nature of proteins. With the advance in techniques and technologies, the use of ultracentrifugation, purification, and the possibility to study proteins from their sequence composition to their structures and interactions, it was confirmed that enzymes are of the same origins of proteins and furthermore, that proteins rarely act alone but they mostly associate with other molecules or other proteins to better exert their functions. It was also reported the vastness of protein functions and the fact that they create a thick network of interactions [7, 8]. Stimuli coming from the outside trigger a cascade of signalling events which involves the intricate networks of protein-protein interaction, protein-ligand interactions, and protein-DNA-RNA interaction. A cascade of signals allows cells to respond to a stimulus coming from the outside by promoting gene expression. In this process, proteins are mainly involved and cover a wide range of roles such as receptors, signalling proteins, and controllers of gene expression. Proteins which are both capable of binding a specific DNA sequence and either activate or repress the transcription of a DNA template are called transcription factors [9].

1.2 Transcription factors

Originally, the term 'transcription factor' referred to a protein involved in transcription or, at least, capable of altering the expression of the genes [10, 11]. Nowadays, the term is strictly related to proteins both capable of recognizing and binding specific sequences of DNA within regulatory regions and recruiting or interacting with proteins that participate in transcription regulation [10]. These properties make the transcription factors key elements for a cell to respond to diverse stimuli by regulating the transcription of specific sets of genes [12, 13]. They are fundamental to many biological processes such as control of cell cycle progression, differentiation of cells during development, immune response, or maintenance of intracellular metabolic balance. Eventually, they form the core of the RNA polymerase II complex, an event that must occur when every gene expression is about to start. Among all human proteins, approximately 1600 are now labelled as transcription factors. Even though the number seems high, the process in which they are involved is so complex that they would never cover the requests if they worked on their own. In fact, they temporarily assemble and interact with a variety of other proteins and co-factors [11]. In order to make the gene expression regulation happen in the proper way and at the right moment, transcription factors bind the DNA cooperating with other transcription factors and furthermore, with numerous cofactors [10, 11]. As all the processes that need fine-tuned and multi-level regulations, and that involve the coordination of a wide range of dynamic elements, some errors and alterations can lead to an unbalance. This condition, if prolonged, may lead to various diseases. It is no surprise that studies in cancer and human disorders have highlighted an overrepresentation of transcription regulators and nucleic acid binders as main effectors due to mutations affecting their func-

tionalities [11, 12, 14]. For that reason, the complicated network of transcription factors, cofactors, and their interactions is an active research field. As aforementioned, transcription factors regulate the gene expression binding DNA motifs or cis-regulatory elements of target genes in specific regions and, they can interact with gene regulatory regions. The regulatory elements in the DNA include promoters and sequences that are defined as either enhancers if they promote the transcription of the gene or silencers if they repress the transcription. The DNA binding domain is the specific domain that allows transcription factors to create a sequence-specific contact with the DNA. There is a variety of DNA-binding structural motifs which can be used to recognise target sequences on the DNA. Among transcription factors, there are similarities in amino acid composition and structures of DNA binding domains. These analogies are used as parameters to make a classification into families and subfamilies. The major families that were identified in eukaryotes are C2H2-zinc finger, basic leucine zipper, nuclear hormone receptor, and basic helix-loop-helix. The advance in experimental methods led to the discovery of more DNA binding domains. Currently, roughly 100 DNA families are known but there are proteins identified as transcription factors with a still unknown DNA binding domain (DBD). The most present DBDs are C2H2 zinc fingers and homeodomain followed, in quantity, by helix-loop-helix, leucine zipper and Forkhead DBD types 2. Depending on the DBD TFs are more likely to bind specific motifs on the DNA. The binding to the DNA is played by noncovalent interactions established between the amino acids composing the DBD of the protein and the bases of the DNA response motif or the DNA's sugar-phosphate backbone. Other noncovalent interactions may independently contribute to the binding like in homeodomain. This domain consists of three α -helices; the third one at the C ter-

minal is inserted in the major groove of the DNA and three amino acid residues interact with TAAT base-core binding motif on the DNA through hydrogen bonds and Van der Waals contacts. The whole complex is toughened up by several water-mediated contacts together with ionic interaction between amino acid side chains and the sugar-phosphate backbone. It must be considered that the interaction between DNA and TFs goes beyond the mere sequence and structural conformation of proteins: cofactors, epigenetic modifications, and cooperative binding with other factors participate in the process. Even though transcription factors can bind the DNA as monomers, they are more prompt to collaborate with each other to form dimers or multimers. Protein-protein interactions and indirect co-operations are strategies acquired to assemble the complex. Concerning protein-protein interactions, TFs can form functional homo or hetero dimers or multimers. Other types of cooperation can also occur through a process known as indirect co-operativity or collaborative competition, in which an agglomerate of TFs collectively competes with the same histone octamer for access to the underlying DNA. Effector domains of TFs are mainly involved in these interactions. The ones involved in the interactions with components of the preinitiation complex (which is composed of general TFs and RNA polymerase II) are called transactivation domains. These domains make either direct Protein-protein interactions (PPIs)s with other transcription factors or components of the mediator complex. One of their characteristics is that they often appear as unstructured regions that acquire a structure only upon binding with a partner which plays as a template for their shaping. Important types of effector domains are the ones that mediate the cooperation among different transcription factors. The cooperative binding can also be considered as a strategy for a combinatorial regulation which enables the cell to integrate signals com-

ing from different pathways. Then, there are effector domains (other specific interfaces) which interact with histone-modifying enzymes to model the chromatin to facilitate or repress the access of the transcription machinery to transcription start sites.

1.3 Transcriptional enhancer associated domain

Trascriptional enhancer associated domain (TEAD), alias Transcriptional enhancer factors (TEFs), are family members of transcription factors involved in the regulation of organ development. In humans, four TEAD proteins have been identified sharing a highly evolutionary conserved TEA DBD and Yes/src associated protein (YAP) binding domain separated by a proline-rich region. The founding member of the TEA family was TEF-1 later named TEAD1; it was identified as a small nuclear protein able to upregulate transcription by binding to the GT-IIC motif (5'-ACATTCCAC-3') of the SV40 enhancer in HeLa cells [15]. The earliest studies on the Hippo pathway and TEAD were performed in Drosophila melanogaster. It was discovered that Scallop gene (Sd) is the single TEAD family gene coding for a TEAD family member. Sd can bind to DNA, and it controls neural progenitor proliferation in the central nervous system and, combined with nerfin-1 (INSM1 in humans), controls medulla neuron fate [16]. Latterly, other proteins comprising the TEA domain were noticed in a range of eucaryotic species such as AbaA in Aspergillus nidulans and Tec1 in Saccharomyces cerevisiae, and TEF1/TEAD1 in vertebrates [17]. There are four members of TEAD family in vertebrate species and they seem to be highly conserved: mammalian TEADs share approximately 98% homology with Sd in the DNA-binding domain and 50% homology in the co-factor-binding domain. The high degree of conservation throughout the TEAD proteins can indicate the import-

ance of their role in different organisms. Furthermore, the four mammalian TEADs share roughly 87.9% homology in the DBD and 72% in the co-factor-binding domain [18]. It is well-known that these proteins are dependent on co-factors: YAP and Transcriptional coactivator with PDZ-binding motif (TAZ) are required to exert the TEAD transcriptional activity. YAP and TAZ participate in the Hippo signalling pathway and are regulators of cell proliferation, stemness, and differentiation. An additional TEAD interacting cofactor family, the vestigial-like proteins (VGLL1-4), interacts with TEADs and it is known to regulate development, proliferation, and differentiation in different tissues, such as skeletal and cardiac muscle [19]. TEAD1 and TEAD2 can interact with the p160 protein family which consists of Src1, Src2, and Src3. They are known to bind to nuclear receptors and activate transcription in a hormone-dependent manner [20]. TEAD proteins bound to co-factors exert their transcriptional function binding the promoter and enhancer regions [21, 22]. Not only the activation of transcription but also the recruitment of epigenetic modifiers and mediators complex happens in the presence of TEAD- YAP/TAZ complex. Furthermore, TEAD proteins form complexes with other transcription factors to coordinate the transcription of target genes.

1.3.1 TEAD structure

As aforementioned, in mammals the TEAD family comprises four members TEAD1 (TEF-1, NTEF-1), TEAD2 (ETF, ETEF-1, TEF-4), TEAD3 (DTEF-1, TEF5, ETFR-1), and TEAD4 (RTEF1, TEF-3, ETFR-2, FR-19) [23]. They allocate the same DNA binding domain, indeed called TEA domain, which is a conserved domain made of 77 amino acids. In addition to TEA domain, TEAD proteins have the so-called YAP binding

domain placed at the C-terminal. At the N-terminal, preceding the TEA DNA binding domain there is a region that along with the proline-rich region composes the variable parts of the protein. Even though high-resolution structure for the full-length protein is not available yet, structures of the individual domains have been explored and separately solved.

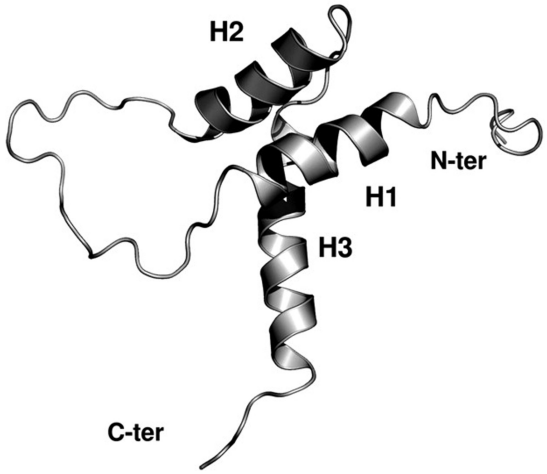


Figure 1.1: *NMR structure of TEA domain [24].*

1.3.2 TEA domain

The TEA domain is composed of a helix turn helix homeodomain fold. The folded globular structure is made of three α-helices, H1, H2, and H3 of which H1 and H2 are nearly antiparallel and they both fold on the opposite side of the N-terminal end of H3. The three α- helices are connected by two loops: the long L1 and the shorter L2. The NMR study revealed that H1-H2 contact forms a hydrophobic patch and that this surface is more likely involved in protein-protein interactions and that H3 helix is mainly dedicated to the interaction with the DNA. Also, the L2 loop which precedes the H3 helix is part of the recognition region [24]. An expansion of this knowledge was possible by using X-ray diffraction to resolve the structure of TEAD-DBD. Differently from what was

reported before, in the TEAD-DBD structure the L1 loop was found to mediate sequence-independent interactions with the minor groove of the DNA. What is more, TEAD1-DBD lacking the L1 loop forms a domain-swapped dimer rather than a monomer present in the seen structure of TEAD1-DNA. A further study on the interface of TEA domain-DNA complex was performed; the X-ray diffraction of the whole human TEAD4 TEA domain-DNA (residues 36-139) revealed that the N-terminal of H3 helix is embedded into the major groove of the DNA duplex and it interacts through specific non-covalent interactions, mainly hydrogen and salt bridges with bases of the DNA. The interaction also involves the L1 region which however contacts the minor groove. It seemed that the main role of the L1 loop is to stabilise the complex predominantly by hydrophobic packing [25, 26].

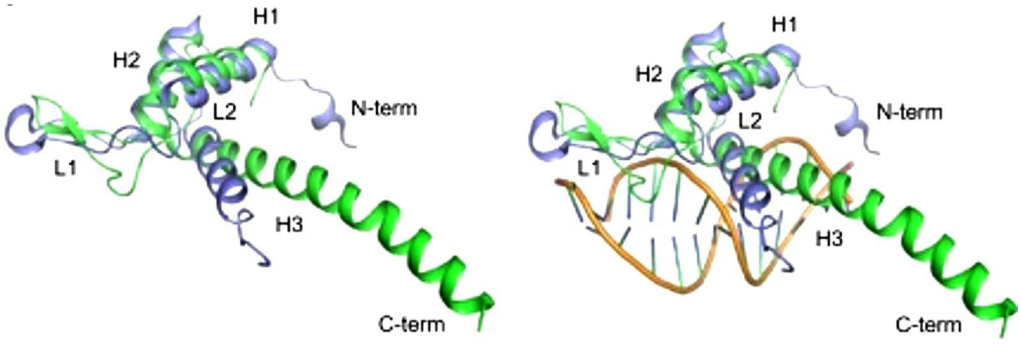


Figure 1.2: *Structural superposition of TEAD-DBD in apo state (on the left) and in complex with DNA (on the right). Comparison of apo state TEAD1-DBD solved by NMR (in figure, blue) and crystallographic structure of DNA complexed TEAD4-DBD (in figure, green). Helix H3 and loop L1 were identified as the DNA recognition regions. Helix H3 is prolonged and 30° rotated in the bound form. Adapted from [26].*

1.3.3 TEAD cofactor- TEA cofactor binding domain

Since TEAD proteins are not able to activate the transcription themselves, they interact with different co-factors mainly with the coactivators YAP and its paralog TAZ. This is the reason why the YAP binding domain is relevant and well-described. This domain is also referred to as transactivation domain: after forming a complex with cofactors through

this region TEADs can perform their role. The first high-resolution study of the structure of this domain was performed on the human TEAD2, and it was noticed that it utilises an immunoglobulin-like fold with a core of two β-sheets packing against each other to form a β-sandwich and two helix-turn-helix motifs capping the openings on each end of the β-sandwich [27]. The TEAD YBD is composed of β-sandwich fold with four α-helix and is post-translationally modified by palmitic acid. The palmitic acid is attached to a conserved cysteine residue inside a central hydrophobic pocket of the protein. It seems that the palmitoylation serves for the overall stability of the protein. Crystal structures of the TEAD YAP domain interacting with the cofactor YAP peptide (amino-acids 50-100) highlighted the interfaces at which the two elements interact. The YBD-TBD complex is composed of three interfaces. Interface 1 consists of an anti-parallel β-sheet between the YBD and YAP residues 52–56. From YAP residues 61–73, several hydrophobic interactions contributed by an α-helix stabilise interface 2. In interface 3 both hydrogen bonds and hydrophobic interactions coming from a short loop, referred to as the ω loop, are created with YAP residues [27, 28]. The three interfaces are conserved between TEAD1-4 and stabilise the binding to YAP cofactor. Further studies of the structure identified that the ω loop serves as the primary energetic driving force at interface 3 for the binding with YAP. Moreover, it was found that the interaction with YAP is formed through a short, unfolded segment of YAP which falls in an ordered conformation upon binding TEAD-YBD surface. The TEAD-YAP complex was also crystallised with other known cofactors such as the Vgll1 peptide, Vgll4 peptide, and TAZ peptide. It was found that Vgll1 binds to the interface 1 and 2 of TEAD and Vgll4 uses a different conformation to create a bridge between two YBD molecules [29, 30]. Interestingly, TAZ

was found to bind to TEAD in two different conformations. In one of the conformations, the binding is like the YAP manner to attach TEAD YBD domain, with non-covalent interactions between interfaces 2 and 3. In the other one, TAZ molecule binds two TEAD molecules to form a heterotetramer. However, in both cases, the same binding interfaces as in the TEAD-YAP complex are used [31]. The two different ways to interact with TEAD by YAP and TAZ might impact biological function differences between TAZ and YAP transcription profiles.

1.4 Role and function

A variety of cues triggers the interaction of TEADs and their co-activators and induces TEAD transcription activity. Stimuli including change in cell morphology, cytoskeletal tension, cell-cell adhesion, metabolic stress, stiffness, and attachment to the extracellular matrix and other vastity of factors play a role in the YAP/TAZ and TEAD co-localisation, thus contributing to TEAD activity. Other signalling pathways, such as TGFβ, Wnt/β-catenin and EGFR can crosstalk with Hippo signalling and influence TEAD activity [23]. TEADs can be subjected to post-translational modifications such as phosphorylation and palmitoylation; phosphorylation of TEAD1 impairs the DNA binding activity and palmitoylation gives more stability to the proteins [32]. This complex network of signals orchestrates the tissue growth to the correct size and shape. TEAD proteins are fundamental during the early development stages and cellular senescence. They exert the regulation of progenitor proliferation, stem cell identity, and lineage specification. Also, by coupling external signals to cell growth and specification they control tissue size and function. During a cell's life, TEADs play different essential roles in maintain-

ing the homeostasis of various tissues and in contributing to tissue responses to injury and mechanical stimuli. For example, TEADs are required for maintaining function in some populations of cells such as progenitor and differentiated cells [33]. In response to mechanical stimuli, they regulate tissue growth or morphology [34]. Signals may be Hippo-dependent, whereas others may be Hippo-independent. Being guardians of the pro-growth gene transcription and regulators of the balance between progenitor cell proliferation and differentiation, it is not unexpected that they can be mainly implicated in diseases. Hypoactivation of TEADs can result in organ dysgenesis and can cause proliferative diseases such as cancer. TEAD mutations may affect interactions with coactivators, such as YAP and TAZ cofactors which are participants of the Hippo pathway that controls organ size by regulating cell proliferation, apoptosis, and stem cell self-renewal.

1.4.1 TEAD subclass function

Even though TEAD1-4 share a high homology, these proteins are differentially expressed in a tissue- and development-dependent manner and the pattern of their expression varies. In fact, each of the TEAD protein isoforms seems to have its own roles during the organism development and cell homeostasis. TEAD1 is expressed at organ and tissue levels such as the lungs, heart, kidney, liver, brain and skeletal muscles. An important role is covered during embryonic development where TEAD1 regulates cardiac muscle growth and myocardium differentiation. This was verified in mice embryos lacking TEAD1 that were showing lethal heart defects. In adult heart tissue, TEAD1 is needed to maintain its physiological function. TEAD1 is also involved in skeletal and smooth muscle formation and function and in development of the neural system through regulation of Foxa2 gene.

It was initially identified as a binder to the GT-IIC motif and SpH enhancers of the SV40 in HeLa cells whose sequence is 5'-ACATTCCAC-3' [15]. Since similar TEAD binding motifs were discovered in many muscle-specific cytidine-adenosine-thymidine genes, a consensus sequence 5'-CATTCCT-3', called M-CAT (muscle CAT) was established [35–37]. For defining the affinity of TEAD1 binding to the DNA duplex, protein-binding chips with randomised DNA duplexes were designed. The results lead to a sequence corresponding to ANATVCZN, in which V can be A, T, or G; Z can be A, T, or C; and N can be any base. Lastly, using ChIP-Seq, HT-SELEX which are high throughput sequencing techniques the consensus binding motif of TEAD proteins was shortened to 5'-ATTCC-3' [38]. TEAD2 is the most abundant TEAD protein during the first seven days of development in mice. Later, during the development TEAD2 regulates neuronal development controlling Pax3 expression. It has been seen that TEAD2 and its coactivator YAP are co-expressed with Pax3 in the dorsal neural tube. It has a role in neuron proliferation. An inactivation of TEAD2 is compensated by TEAD1 during the first stage of embryonic development. TEAD3 function remains something which is not clear yet. It is expressed in the extraembryonic tissue which forms the placenta later in the development. TEAD4 was initially thought to be responsible for the regulation of the trophectoderm genes needed for the trophectoderm cell differentiation. Verily, when TEAD4 was inactivated in mice a disruption in the trophectoderm (precursor of placenta) formation was seen with failure in an embryo's implantation. A post implantation inactivation in mice did not affect the normal development. Subsequent studies found out that TEAD4 is present not only in the nucleus but also in mitochondria and that its main function is keeping homeostasis and preventing an overaccumulation of reactive oxygen species.

1.5 Role in cancer

Cancer is a complex disease which still does not have a defined cause, but it is the result of a dysregulation of the homeostasis and mutations. Tumorigenesis is the loss of balance and gain of several malignant aspects such as dedifferentiation, fast proliferation, metastasis, evasion of apoptosis, immunosurveillance, and dysregulation in the metabolism. Mutations in oncogenes and tumour suppressor genes might affect their expression levels or activities. Having said that, TEAD proteins can be deemed possible cancer effectors once dysregulated. Also, they were found to upregulate expression of numerous other genes connected with cell proliferation like CYR61 and CTGF both involved in migration and adhesion, genes known as oncogenes or tumour markers (MYC, Mesothelin) and genes encoding glucose transporters GLUT1 and GLUT3. It has been seen that multiple tumours overexpress TEAD genes including prostate cancers [39], colorectal cancers [40], breast cancers [41] or gastric cancers [42]. As mentioned before, TEAD proteins cannot interact with the DNA until they have interacted interacting with other proteins- their coactivators. The well-known coactivators are YAP and its paralog TAZ. The two proteins are effectors of the Hippo signalling pathway. Stimuli activate the pathway at the level of the receptors and trigger a cascade of signals which leads to the formation of a nuclear complex among YAP or TAZ and TEAD proteins and the activation of the translation. Hippo signalling pathway plays a major role in organ size control, cell proliferation and tumorigenesis [43, 44]. Mechanical signals, cellular stress, extracellular stimuli, or cell-cell contact are transmitted at the level of cytoplasm through kinases to large tumour suppressor kinase 1/2 (LATS1/2) which, once activated, phosphorylates YAP or TAZ. The phosphorylation inactivates the proteins which are not available for

TEAD binding and vice versa. The Wnt/β-catenin pathway [45] and alternative Wnt pathway [46], the TGFβ pathway [47] or the LKB1-AMPK signalling [48] are found to regulate TEAD activity by promoting their interaction with YAP/TAZ coactivators. Moreover, in a Hippo independent way other coactivators interact with TEAD proteins. The VGLL4 protein which is part of the Vestigial-like protein (VGLL) family binds to TEAD on a similar interface of YAP/TAZ competing with them for the binding site. It was found that an increased concentration of VGLL4 inhibits the effect of the deregulated Hippo pathway causing increased nuclear translocation of YAP/TAZ by blocking its binding site on TEAD [30]. Post translational modifications were described which seem to regulate the activity of TEAD proteins. TEAD proteins can be phosphorylated on serine and threonine residues of the third helix thus disrupting the DNA binding of the protein to the DNA. For the physiological function and the proper folding and stability of TEAD proteins, the addition of palmitic acid on the cysteine residue in the hydrophobic core of YAP binding domain was found to be crucial. Eventually, cellular stress induces a cytoplasmatic translocation of TEAD proteins: TEAD forms a complex with p38 and is translocated to the cytoplasm. In this way, YAP/TAZ are unable to activate their target genes.

1.6 Forkhead box

The Forkhead box (FOX) family of transcription factors is comprised of more than 100 members which were divided into 19 subclasses designated by a letter from A to S according to the sequence homology within the Forkhead box and other functional domains [49]. The origin of FOX genes was tracked in unicellular organisms, and during evolution, it has expanded through a series of duplication events. As with all transcription factors, FOX TFs either activate or repress gene transcription or exert their function by interacting with other transcription factors. This family of TFs is involved in essential processes including development, differentiation, proliferation, apoptosis, stress resistance, and metabolism [50]. The first identified member of the forkhead class of TFs was found in Drosophila melanogaster: it was seen to be essential for the proper formation of terminal structures of the embryo. The mutant model showed a transformation of the foregut resembling a head-like structure [51, 52]. The class O of forkhead TFs (FOXO) has been identified in numerous species such as *C. elegans*, *D. rerio*, *D. melanogaster*, *G. gallus*, mouse, and human. Four mammalian FOXO proteins were identified: FOXO1 (FKHR), FOXO3a (FKHRL1), FOXO4(AFX), and FOXO6. The FOXO transcription factors are homologous to the Caenorhabditis elegans transcription factor DAF-16 (abnormal DAuer Formation-16), which regulates lifespan downstream of the C. elegans insulin receptor DAF-2. In mammalian cells, similarly, FOXO1, FOXO3, and FOXO4 members are regulated by insulin-like growth factor (IGF) signalling where, in response to insulin and growth factor stimulation, they are phosphorylated by the Serine–threonine kinase AKT/protein kinase B (AKT/PKB). The phosphorylation on Thr32, Ser253, and Ser315 residues enables the formation of a complex with 14-3-3 proteins, inhibiting the

transcription activation and promoting their transition from the nucleus to the cytosol [53]. FOXO TFs are ubiquitously expressed with a variation in the expression level depending on the type of organs or tissues. The repression of FOXO activity is directed by post-translational modifications such as the just mentioned AKT/PKB mediated phosphorylation, non-AKT/PKB-mediated phosphorylation, acetylation and ubiquitylation. There are several sites near or within the FOXO DNA-binding domain for posttranslational modifications which enable or attenuate the affinity of FOXO direct interaction with the DNA or protein-protein interactions. Many of these FOXO post-translational modifications regulate fox activity towards the increased cellular ROS species which triggers cellular oxidative-stress resistance. FOXO proteins play pleiotropic roles in developmental and adult-age cellular/biological processes. These transcription factors have been related to different functions such as stem cell maintenance and differentiation (MSTN), proliferation (Cyclin D), regulation of glucose and lipid metabolism (G6PC), apoptosis (TRAIL), oxidative-stress resistance (Catalase), DNA damage repair, cell cycle arrest (Cyclin D), and maintenance of the homeostasis. FOXOs are implicated in a broad range of important cellular functions (related to the decision of the fate of the cells), signalling pathways (phosphatidylinositol 3-kinase/protein kinase B (PI3K-AKT), and the transformation of growth factor beta (TGF-β), WNT/β-catenin, insulin, Sonic-Hedgehog and Jagged-Notch) [54]. Moreover, FOXO proteins act downstream of numerous oncogenic pathways including NF-κB-IKKβ (nuclear factor-κB and NF-κB kinase-β inhibitor, respectively), ERK, PI3K–AKT. For this they have been defined as tumour suppressors with anti-proliferative and pro-apoptotic properties[55, 56]. It is no surprise, as it was for TEAD transcription factors, that dysregulations of the functioning of these molecules

bring cancer progression, tumorigenesis and also ageing. For this reason, the key Insulin signalling and oxidative-stress signalling have been well described to discover the involvement of FOXO family in the pathways. These two signalling pathways crosstalk when they end at a common FOXO point. the increase of ROS triggers the JNk signalling cascade. Thus, JNK JNK-mediated phosphorylation of residues located in the transactivation domain of FOXO TFs may contribute to FoxO interactions with the basal transcription machinery. JNK mediates the interaction between insulin and oxidative-stress signalling. AKT/PKB phosphorylation of FOXO proteins induces foxo binding to 14-3-3 proteins which leads to the repression of FOX TF activity and a translocation of these proteins from the nucleus to the cytoplasm. On the contrary JNK phosphorylation induces FOX activation and transcription regulation. Thus, insulin and oxidative stress balance FOXO function through multiple mechanisms and cross-talking; the outcome of the opposing actions is highly context dependent [57]. Moreover, it has been reported that FOXO protein activation comes from the acquisition of opposite roles among themselves. For example, FOX04 activates cellular senescence which serves as defence to prevent cell-hyper-proliferation. On the contrary FOXO1 and FOXO3 participate in resisting senescence. As it happens in cardiac microvascular endothelial cells and vascular SMCs, the activation of PI3K/ATK pathway leads to the phosphorylation of FOXO1 and FOXO3 transcription factors inhibiting their activity on antioxidant and promoting the increased cellular ROS levels and senescence. The increased level of ROS involved JNK-induced activation of FOXO4 and the subsequent senescence process. FOXO4 shows the opposite effects of FOXO1 and FOXO3 in cellular senescence. These opposite effects depend on the activation of individual FOXO proteins by the prevailing conditions of cellular stress

[58, 59]. Therefore, the preservation of the cell homeostasis is a matter of the balance in activation among distinct FOXOs.

1.6.1 FOXO structure

The Forkhead transcription factors share a highly conserved 110-amino acid DNA binding domain known as forkhead box or winged-helix domain. The amino acids in this domain fold into three α-helices (H1, H2 and H3), three β-strands (S1, S2 and S3) and two wing-like loops (W1 and W2) [60]. The topological arrangement of the domain is H1-S1-H2-H3-S2-W1-S3-W2. The strand S1 found in between the two helices H1-H2, interacts with S2-S3 strands to form a three-stranded antiparallel β sheet. The helix H3 plays an important role as a main DNA- recognition element: it binds to the major groove roughly perpendicular to the DNA axis through bases-specific interactions [60, 61]. Residues involved in this interaction are conserved among all the forkhead proteins and make either a direct or water-mediated amino acid-base contact [62]. The highly conserved residues N–X–X–R–H–X–X–S/T of helix H3 interact with the major groove of the DNA. The included in the H2-H3 turn and both wings W1 and W2 are the less conserved regions. It has been thought that these less conserved regions of the forkhead domain are necessary for the DNA-binding specificity and the stability of FOX-DNA complexes. Some differences in the domain distribution have been identified among the forkhead proteins such as the presence of an additional short 310-type helix between α-helices H2 and H3 [63–65], an additional α-helix at the C-terminus of forkhead domain. Two consensus sequences have been identified for FOXO proteins: 5 ⁽-GTAAA (T/C) AA-3⁽, known as the Daf-16 family member-binding element [66]; and 5⁽ (C/A) (A/C) AAA(C/T) AA-3⁽, present in

the IGFBP-1 promoter region and known as the insulin-responsive sequence [67, 68]. Both sequences include the core sequence 5\'-(A/C) AA(C/T) A-3\' recognised by all FOX proteins. The crystal structure of FOXO3-DNA complex provided the first insight of FOXO-DNA interaction. The H3 helix docks perpendicularly to the major groove of the DNA. The well-conserved residues of helix H3 create contact with the Daf-16 family member-binding element consensus sequence on the DNA. The loop regions defined as wings W1 and W2 were found to participate in DNA binding. The W1 interacts with the phosphate backbone of DNA stabilizing the protein-DNA complex. Also, the W2 wing is involved in stabilisation of the FOXO-DNA complex. It seems to contact the major groove phosphate bone of the DNA and to allow the binding with the target DNA. All FOXO proteins contain a five amino acid insertion (sequence KGDSN) between the H2-H3 helices. The role of this sequence is still unveiled even though in FOXO4-DNA structure it seems to adopt a different conformation and interacts with the phosphate groups of DNA backbone [69]. This structure could help to stabilise the FOXO-DNA complex. Another region involved in the stabilisation of the complex is the N-terminal region of the forkhead domain. Ordered water molecules play an important role at the FOXO/DNA interface and they are an integral part of the sequence-specific DNA contacts [64, 69]. It is well-known that FOXO activity is regulated through posttranslational modifications. Most of the sites involved in the modifications were found within the forkhead DBD. It was investigated where the protein kinase B phosphorylates FOXO protein. It phosphorylates three sites of FOXO: one of them is in the W2 close to the conserved residues which are involved in the DNA binding. The other two sites accepting the phosphorylation are the 14-3-3 protein binding motifs. One is in the wing 2 and

the other one is located at the N-terminus of the FOXO molecule. The 14-3-3 proteins inhibit the DNA-binding affinity of FOXO factors [70]. Another kinase which participates in the phosphorylation of the forkhead domain of FOXO proteins is the oxidative stress-regulated mammalian Ste-20-like kinase-1 (MST1) [71]. MST1 phosphorylates four highly conserved serine residues thus causing the disruption of the FOXO–14-3-3 complex and the translocation of FOX proteins to the nucleus. Two of these residues are located within the helix H3, and the other two lie at the C-terminus of wing W1. When phosphorylated, these sites decrease the DNA-binding affinity of the FOX forkhead domain [72]. At the C-terminal there are sites which are mainly acetylated. The acetylation impacts the FOXO DNA-binding affinity, therefore decreasing it [64, 73]. FOXO4 is a member of the forkhead family transcription factors O subclass. It is also known as acute leukaemia fusion gene from chromosome X (AFX) which was initially demonstrated to fuse with the mixed-lineage leukaemia zinc finger transcription factor owing to a t(X;11) chromosomal translocation in acute leukaemia [74]. Initially, gene knockout studies in mice did not reveal any abnormalities in the absence of FOXO4 [75] It was shown later that, as all FOXO isoforms, it plays an important role in several cell processes. In mammals, it was found to be expressed in almost every tissue, including the brain, lung, kidney, skin, muscles, and prostate. It was noticed that is mostly abundant in skeletal muscle tissues. FOXO4 has a role in regulating the antioxidant state, apoptosis, cell cycle, and cellular homeostasis. It also serves as a tumour suppressor by induction of cell cycle arrest (from thesis). It mediates cell responses to treatment with antitumour agents. The most important function of FOXO4 is the control of the oxidative stress of the cells. FOXO4 promotes detoxification through transcriptional activation of anti-

oxidative enzymes, such as superoxide dismutase 2 and catalase and on the other way round reactive oxidative species regulate FOXO4 activity. On one hand, the oxidative stress activates post-translational modification of the protein in particular JNK-mediated phosphorylation, SIRT1-mediated deacetylation, and MDM2-mediated monoubiquitination induce its nuclear translocation [76–78]. On the other hand, NLK-inhibited FOXO4 monoubiquitination causes downregulation of FOXO4 activity and it inactivates the cellular antioxidative defence systems thus raising the rate of cell death [79].

1.7 FOXO4

The involvement of FOXO4 in tumorigenesis as a tumour suppressor strictly relates to its anti-tumoral activities. It seems to regulate anti-proliferative (e.g., p27kip1, cyclin D1, GADD45) and pro-apoptotic functions (e.g., Bim, Bcl-6, Bax) [80]. FOXO4 plays a fundamental role in the regulation of angiogenesis, erythropoiesis, and glucose metabolism which are mechanisms essential for tumour expansion. At the presence of hypoxic conditions, FOXO4 activates the downregulation of hypoxia-inducible factor 1 α protein (HIF1α) and the suppression of vascular endothelial growth factor (VEGF), erythropoietin, and glucose transporter 1 expression. These factors inhibit angiogenesis [81]. Since FOX04 is implicated in pivotal cell functions and anti-tumoral activities and, its dysregulation implies the rising of a variety of pathological diseases it is considered a target for cancer and other disease treatments.

1.7.1 FOXO4 structure

FOXO4 gene encodes a protein of 505 amino acids whose structure is conserved throughout a variety of species. It is characterised by four domains including a highly conserved forkhead winged-helix DNA-binding domain (FHD), a nuclear export sequence (NES), a nuclear localisation sequence (NLS), and a C-terminal transactivation domain (TAD). The FHD domain is involved in the recognition of the consensus core recognition motif 5'-GTAAACAA-3' on the DNA. The NLS domain interacts with the nuclear import machinery for the translocation of FOXO4 into the nucleus. The NES domain interacts with exportins and mediates the export of the TF. Phosphorylation of NLS element can result in FOXO4 inactivation with following export of FOX04 from the nucleus [82]. The earliest studies on the structure of FOXO4-DBD were conducted by using NMR and it was confirmed the expected forkhead winged-helix fold which comprises three -helices (H1–H3), a fourth short 310-helix H4 located between H2 and H3, and a short twisted three-stranded antiparallel -sheet comprising of strands S2 and S3 flanked at the N terminal and C terminal (composed by wings W1 and W2) disordered and highly flexible regions. These two regions are involved in the stabilisation of the protein DNA complex (the N-terminal unstructured part preceding the first helix and the C-terminal W2 wing loop). FOXO4-DBD uses its helix H3 to dock into the major groove through base-specific contacts, while the N-terminal loop preceding the first helix H1, the turn region between helices H4-H3, and the flexible wing W1 located between strands S2-S3 make additional contacts with the phosphate groups of DNA. As it was for the rest of the forkhead DBD the highest variability of these regions plays a role in the modulation of the DNA-binding specificity. The conserved motif of helix H3 (Asn148-X-X-Arg-His-X-X-Ser/Thr) inter-

acts with the bases of the DNA via both direct and water-mediated hydrogen bonds as well as van der Waals contacts as it was seen for all the FOX factors. In contrast with the other FOXO-DBD-DNA structures, the loop between helices H2-H3 differs in conformation. It participates in the DNA binding. It has been noticed that the highly ordered water molecules which line at the binding interface create a sequence-specific manner interaction with the DNA [69].

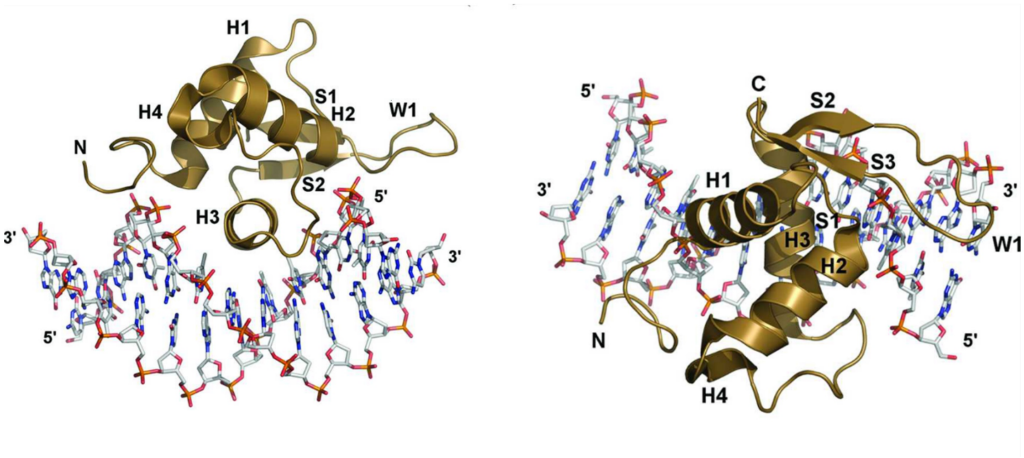


Figure 1.3: *Structure of the FOXO4-DBD·DNA complex.* ***a****. The overall view of FOXO4-DBD bound to DNA. The complex structure is formed by insertion of the H3 helix to the major groove and is further stabilised by sequence-independent contacts of the N-terminus and wing W1 with the phosphate groups of DNA.* ***b****. 90° rotation of complex in a., toward its horizontal axis. Adapted and modified from [69].*

1.8 Cancer involvement of TEAD and FOX

The important role that TEAD and FOX proteins play as tumour suppressors brings the attention to their possible implication in tumour progression related to some disruptions in carrying their functions or in pathways they are mainly involved as effectors. It has been seen that multiple tumour types overexpress TEAD genes, and this overexpression was correlated to the expression of pro-growth factors, such as CTGF, receptor tyrosine

kinase AXL, Cyr61, survivin and Myc. TEAD expression levels are also used as a prognostic marker in different type of cancers including prostate cancer, gastric cancer, and colorectal cancer. Furthermore, TEADs interact with YAP and TAZ which are well-defined oncogenes, and they are seen to be overexpressed in many cancers [43, 83]. A high presence of YAP/TAZ coactivator proteins leads to an excessive number of co-activators that are not phosphorylated by the Hippo kinases and that can activate TEADs transcriptional activity entering the nucleus. Also, it is well-known that TEAD is required for YAP oncogenesis and that YAP oncogenesis is TEAD-dependent. For this reason, TEAD proteins have emerged as promising therapeutic targets. For what concerned FOX proteins, they participate in several oncogenic pathways such as, PI3K–AKT, ERK and NF-κB-IKKβ (nuclear factor-κB and NF-κB kinase-β inhibitor, respectively). FOXOs are well-described as tumour suppressors for their anti-proliferative and pro-apoptotic properties. They are guardians of p27KIP1, CDKN1A/p21, FasL, Trail and Bim genes which participate in cell death and cell cycle arrest [56]. FOXO are part of the PI3K/AKT molecular cascade which is seen to be altered in many human cancers; it is involved in several key cellular processes such as differentiation, survival, proliferation/growth, metabolism, and migration. FOXOs acts as tumour suppressors due to their functions in promoting apoptosis and senescence, and inhibiting cell cycle progression, angiogenesis, metastasis and impacting cancer metabolism. Under certain conditions FOXOs can promote tumour development maintaining cancer stem cells active, reactivation of PI3K-AKT and drug resistance [84]. Having said that, TEAD and FOX become matter of interest as therapeutical targets. But since some structural studies are missing regarding these two protein families and some DNA interaction characterisations it is important to

improve methodologies to acquire a satisfying knowledge.

1.9 TEAD and FOXO4 transcription factors possibly in complex with DNA

In the laboratory this thesis work is accomplished, many experiments were run to develop refined structural mass spectrometry methods for the characterisation of transcription factor complexes with their cognate response motifs and to apply these methods to structurally characterise the interaction between the DNA binding domains of FOXO4 and TEAD1 proteins and their DNA response motifs. Interestingly, searching for the regulatory regions of human genes for TF binding sites, it has been noticed that TEAD and FOXO4 binding motifs collocate alongside each other. To unveil the hypothesis, several dsDNA mimicking the TEAD/FOXO4 interacting genomic regions were prepared and it was evaluated the ability of the two TFs to bind the DNA binding domains at the same time through a gel shift assay and native MS. The test was run on five DNA constructs which differed in orientation of the two binding sites and in the number of base pairs separating the two binding sites. In four of them the three-element complex (TEAD1-DBD, FOXO4-DBD and the dsDNA). The highest amount of the tertiary complex was seen in samples containing dsDNA construct where the two binding motifs overlapped in one base pair. The ternary complex was also found in both samples where the motifs were separated by one base pair, the presence of the ternary complex was not confirmed for the dsDNA construct where the two motifs were separated by two base pairs. The increased complex formation was identified in samples where two motifs are closer to each other, thus suggesting the two TFs interact with each other. This proof of evidence paid the base

for this thesis research study.

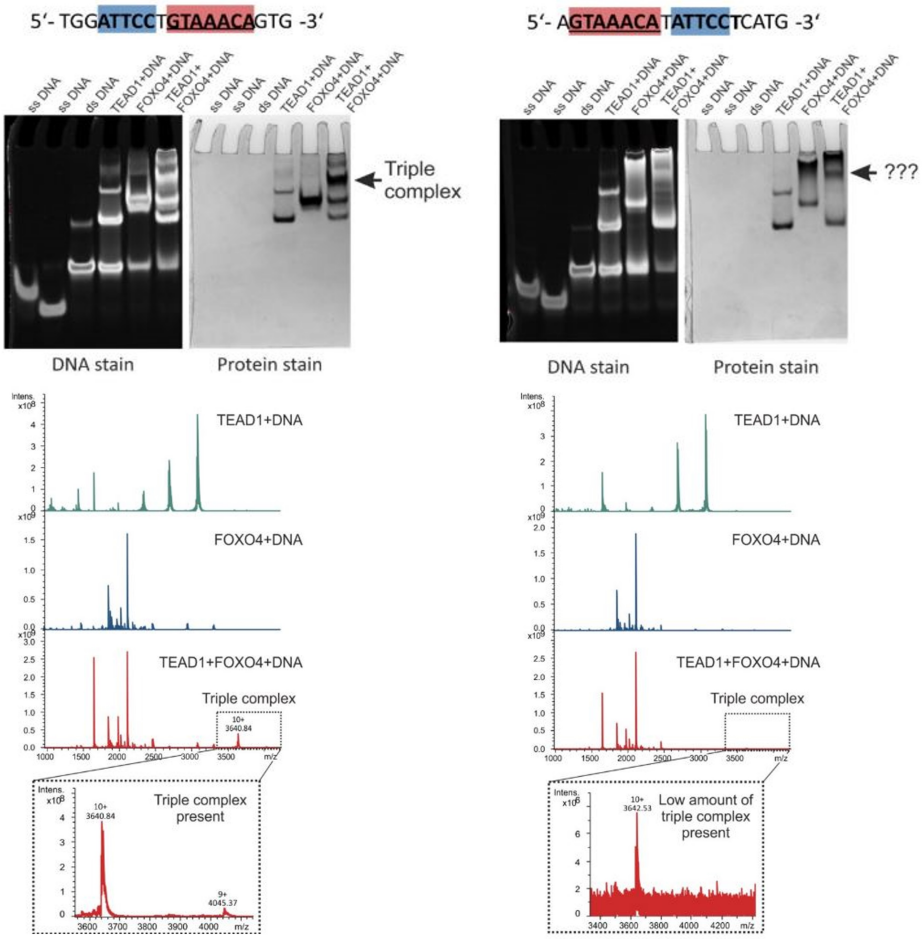


Figure 1.4: *Confirmation of presence of a complex of TEAD1-DBD, FOXO4-DBD and one dsDNA construct. In the left column, the complex could be observed directly from the gel shift assay and the native MS confirmed its identity, in the right column it was barely visible due to low amount of the complex being formed. The sensitive native MS analysis was necessary to observe its presence. The position of the TEAD1 binding motif is highlighted in blue and that of FOXO4 in red in the oligonucleotide sequence in the top part of the picture. Their mutual position is probably the reason for the different amounts of the complex formed.*

Chapter 2

Aim of the thesis

The structural dynamics of protein is still a challenge for researchers. Proteins are crucial to most of life's functionalities, and any deregulation in which they are involved can lead to diseases. The aim of this thesis is to contribute to the characterisation of change in conformation of two proteins, in particular TEAD1 and FOXO4 transcription factors in the absence and the presence of the oligonucleotide and their interaction in forming a ternary complex with the DNA by the combination of mass-spectrometry technique and chemical cross-linking. The specific goals were:

- Set different experimental conditions to identify any change in conformation of the proteins or any protein-protein interaction.
- Chemical cross-linking of the proteins and MS-analysis.
- Quantitative analysis of the results
- Characterisation of the protein structure using models executed in Pymol.

Chapter 3

Material and Methods

3.1 Proteins and DNA oligonucleotides

TEAD1 sequence:

GSSADKPIDNDAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIA

RYIKLRTGKTRTRKQVSSHIQVLARRKSRD

FOXO4 sequence:

GSGAVTGPRKGGSRRNAWGNQSYAELISQAIESAPEKRLTLAQIYEWMVRTVPY

FKDKGDSNSSAGWKNSIRHNLSLHSKFIKVHNEATGKSSWWMLNPE

Oligonucleotide sequence:

NO2 F (5'-TGGATTCCTGTAAACAGTG-3') (Sigma Aldrich, US)

NO2 R (5'-CACTGTTTACAGGAATCCA-3') (Sigma Aldrich, US)

3.2 Buffers and solutions

- Protein isolation and equilibration buffer

HEPESbuffer, 20mM HEPES, 0.5M NaCl, 20mM Imidazole, DNAse (300 U), RNAse (300U) (pH 7.4).

- Size exclusion chromatography, dissolution of proteins, and preparation of solutions

HEPES buffer, 20mM HEPES, and 50mM NaCl solution (pH 7.5).

- Polyacrylamide gel electrophoresis in the presence of litium dodecyl sulfate (LDS-PAGE)

MES Running Buffer 20x concentrated pH 7.25.

LDS sample buffer 4× concentrated containing 100 mM Dithiothreitol (DTT).

Staining solution 45% (v/v) methanol, 10% (v/v) acetic acid, 0.25% (w/v) Coomassie brilliant blue R-250.

Destaining solution 55% (v/v) H_2O, 35% (v/v) ethanol, 10% (v/v) acetic acid.

- Digestion of samples

Dilution solution, 50 mM ethylmorpholine buffer (pH 8.5) with 10% AcN. Trypsin dissolution, 50 mM ethylmorpholine buffer (pH 8.5) with 10% AcN.

3.3 Protein expression

E. Coli BL21-CodonPlus (DE3)-RIPL (Agilent Technologies, USA) was used for recombinant production of the proteins. The batch was transformed using the pET-28 (+) plasmid (Generey, China), containing the gene of the full-length TEAD1-DBD and FOXO4-DBD transcription factors which was cloned immediately adjacent to an *N*-terminal His-tag of the pET-28a (+) plasmid to express the respective *N*-terminal His-6 tagged recombinant protein. The bacteria were grown in Terrific Broth medium at 37 °C until OD 600 = 0.9. The expression of the protein was induced by the addition of 0.5mM Isopropyl-β-D-thiogalactopyranoside (IPTG) in the broth with consequent incubation at 18 °C for more than 16 hours. The bacteria were harvested by centrifugation (8000 × g, 15 min, 4°C) and stored at -80 °C for further use.

3.4 Protein purification

3.4.1 Protein isolation and affinity chromatography

Bacterial pellets overexpressing FOXO4-DBD and TEAD1-DBD were re-suspended in 20mM HEPES, 0.5M NaCl, 20mM Imidazole, pH 7.4 buffer containing protease inhibitor cocktail, DNAse (300 U), RNAse (300U), and the disintegration of the bacterial pellet was carried out by ultrasonication in ice bath. The soluble fraction was separated by centrifugation (70 000 × g, 60 min, 4 °C), filtered using a 0.22 µm syringe filter (Millipore, Merck, USA), and the liquid part, the supernatant, was loaded on Ni^{2+} -charged Bio-Scale Mini Nuvia IMAC (5ml; Bio-Rad Laboratories, USA) already equilibrated in HEPES buffer. The resin was washed with the HEPES buffer, and the protein was eluted by a linear gradient of 20-500 mM imidazole. The fractions containing FOXO4-DBD and TEAD1-DBD were dialysed using a solution 20mM HEPES, 50mM NaCl, pH 7.4, and the *N*-terminal His 6 -tag was cleaved during dialysis using TEV protease 1 (1:100, w/w), overnight at 4 °C. ESI-FTICR MS analysis was carried out to verify the cleavage efficacy. The His 6-tagged TEV protease was separated from FOXO4-DBD and TEAD1-DBD by affinity chromatography on Bio-Scale Mini Nuvia IMAC (1 ml; Bio-Rad Laboratories, USA), and the flow-through fraction was loaded on EconoFit High S column (5 ml; Bio-Rad Laboratories, USA) to perform cation-exchange chromatography. The proteins were further eluted by a linear gradient of NaCl ranging 50- 500 mM NaCl. The collected fractions of the proteins were frozen in liquid nitrogen and stored at -80 °C until further use.

3.4.2 Size exclusion chromatography

SEC chromatography was used to assess the purity of the proteins and change the buffer. It was performed on SEC70 column (Bio-Rad Laboratories, USA). The column had been previously flashed with 100 ml of EtOH, 20% v/v. The constant flow was set at 1000 µl/min. The buffer used was HEPES buffer (20mM HEPES, and 50mM NaCl solution at pH 7.5). The eluate was monitored at a wavelength of 280 nm for the whole chromatography. The column was connected to a fraction collector, where a volume of 800 µl per fraction was collected. Fractions corresponding to the highest absorbance of protein concentration were collected and measured on a spectrophotometer Denovix DS-11.

3.4.3 Determination of protein concentration

The concentration of both TEAD1 and FOXO4 proteins was measured with spectrophotometer (DeNovix DS-11 FX+). Both proteins were dissolved in 20mM HEPES, and 50mM NaCl solution at pH 7.5 (HEPES buffer). In the case of TEAD1 1mM Dithiothreitol DTT was added to the solution. The wavelength used was 280 nm and the molar extinction coefficient for TEAD1 was 15220 l/mol*cm and for FOXO4 was 32000 l/mol*cm.

3.5 Gel electrophoresis

3.5.1 Polyacrylamide gel electrophoresis in the presence of litium dodecyl sulfate (LDS-PAGE)

TEAD1 and FOXO4 and their corresponding cross-linked forms were analysed using LDS-PAGE to verify the integrity of the proteins. The apparatus was filled with 20× concentrated running buffer (2-(N-morpholino)ethanesulfonic acid (MES) buffer, Bio-Rad Laboratories, USA). All samples were prepared in the same way. 10 µl of the sample was mixed with 2 µl of 4× concentrated LDS sample buffer (Invitrogen) containing 100 mM DTT as the reducing agent in a 3:1 (v/v) ratio. To fully denature proteins, samples were mixed, sonicated for 3 min, and heated at 95 °C for 5 min then, loaded onto a precasted NuPage 4–12% Bis–Tris gel. In order to identify the proteins' relative molecular weight, 10 µl of protein marker Novex Sharp Pre-stained Protein Standard (Bio-Rad Laboratories, USA) was loaded on the gel as well. The voltage for this technique was set up to 200 V max at 130 mA and the electrophoresis had been running for approximately 45 min. After separation, the gels were stained by Coomassie Brilliant Blue R250 and destained with ethanol, water, and acetic acid in the ratio 55:35:10 [85]. The same procedure was used to analyse the reactions of the experiment of this thesis.

3.6 Chemical cross-linking experiment characterisation and set up

3.6.1 DSA experiment

The Disuccinimidyl adipate (DSA) experiment was set up to have different conditions which were numbered from 0 to 5. The 0 condition is the negative control. Conditions 1-

5 are used to study the change in conformation of TEAD1 and FOX04 in the presence of the oligonucleotide and the possible interactions between the proteins with and without the presence of the oligonucleotide. Experiment conditions:

- 0: TEAD1 + DSAC12/ DSAC13 FOXO4 + DSAC12/ DSAC13
- 1: FOXO4 + DSAC12 versus FOXO4+ NO2 + DSAC13
- 2: TEAD1 + DSAC12 versus TEAD1 + NO2 + DSAC13
- 3: TEAD1 + FOXO4 + DSAC12 versus TEAD1 + FOXO4 + NO2 + DSAC13
- 4: TEAD1 + NO2 + DSAC12 versus TEAD1 + NO2 + FOX04 + DSAC13
- 5: FOXO4 + NO2 + DSA C12 versus FOXO4 + NO2 + TEAD1 + DSAC13

DSAC12 is non-labelled light form of the cross-linker. DSAC13 is the labelled heavy form of the cross-linker.

The volume and the concentrations of each element in the final solutions were calculated to have an equal molar concentration of FOXO4 and TEAD1 in every solution. The amount of the oligonucleotide was quantified to have a ratio of 1:1:1 (FOXO4: TEAD1: OLIGO) in the reactions. The same was done for the volume of the cross-linker in solution. The cross-linker must have a ratio 1:20 over the proteins. The experiment included 18 reactions for each protein to obtain 5 different conditions (1-5). Firstly, 4 solutions (A-D) plus one control solution per each condition (E-CTRL) were created mixing either FOXO4 or TEAD1 or both the proteins together depending on the desired condition. At the proteins were also added 500nM NO2 oligonucleotide, 15 g/l cross-linkers solution, and the HEPES buffer (pH 7.5) up to 16,54 µl final volume per each reaction. FOXO4, TEAD1, and NO2 have the same molar concentration in solution which is 46 µM.

3.6.2 DSBU experiment

The volume and the concentrations of each element in the final solutions were calculated to have an equal molar concentration of FOXO4 and TEAD1 in every solution. The amount of the oligonucleotide was quantified to have a ratio of 1:1:1 (FOXO4: TEAD1: OLIGO) in the reactions. The same was done for the volume of the cross-linker in solution. The cross-linker must have a ratio 1:20 over the proteins. The experiment included 6 reactions where TEAD1 and FOXO4 are mixed together to have two different conditions C and D. At the proteins were also added 500nM NO2 oligonucleotide, 20 g/l cross-linkers solution, and the HEPES buffer (pH 7.5) up to 22 µl final volume per each reaction. FOXO4, TEAD1, and NO2 have the same molar concentration in solution which is 30 µM.

3.7 Sample preparation

3.7.1 Oligonucleotide preparation

The NO2 oligonucleotide (Sigma Aldrich, US) was used. It contains the binding motif of both TEAD1-DBD and FOX04-DBD. The forward sequence 5'-TGGATTCCTGTAAACAGT G-3' was annealed to the reverse sequence 5'-CACTGTTTACAGGAATCCA-3'. The desired duplex samples were obtained by dissolving corresponding forward/reverse strands in water, mixing them in an equimolar ratio, and heating the solution at 95°C for 5 min. Finally, each mixture was allowed to slowly cool down at room temperature to generate a 19-bp dsDNA construct.

3.7.2 DSA reaction preparation

A previously described protocol was followed to perform cross-linking reactions [86]. The cross-linking reagent DSA was used and added after mixing all the other elements in the solution. Each A and D sample from 1 to 5 conditions was added with a non-labelled cross-linker (DSAC12), whereas each B and C sample from 1 to 5 conditions was added with the ^{13}C-labelled version (DSAC13). Please refer to the table 3.1.

Table 3.1: DSA reaction assessment

SOLUTIONS	TEAD1	FOXO4	NO2	BUFFER	DSAC12	DSAC13	DSAC12/C13
0A TEAD1	3.45	X	X	12.49	X	X	0.60
0B FOXO4	X	6.82	X	9.12	X	X	0.6
1A	3.45	X	X	12.49	0.6	X	X
1B	3.45	X	1	11.49	X	0.6	X
1C	3.45	X	X	12.49	X	0.6	X
1D	3.45	X	1	11.49	0.6	X	X
1E	3.45	X	1	11.49	X	X	0.6
2A	X	6.82	X	9.12	0.6	X	X
2B	X	6.82	1	8.12	X	0.6	X
2C	X	6.82	X	9.12	X	0.6	X
2D	X	6.82	1	8.12	0.6	X	X
2E	X	6.82	1	8.12	X	X	0.6
3A	3.45	6.82	X	5.07	1.2	X	X
3B	3.45	6.82	1	4.07	X	1.2	X
3C	3.45	6.82	X	5.07	X	1.2	X
3D	3.45	6.82	1	4.07	1.2	X	X
3E	3.45	6.82	1	4.07	X	X	1.2
4A	3.45	X	1	10.89	1.2	X	X
4B	3.45	6.82	1	4.07	X	1.2	X
4C	3.45	X	1	10.89	X	1.2	X
4D	3.45	6.82	1	4.07	1.2	X	X
4E	3.45	6.82	1	4.07	X	X	1.2
5A	X	6.82	1	7.52	1.2	X	X
5B	3.45	6.82	1	4.07	X	1.2	X
5C	X	6.82	1	7.52	X	1.2	X
5D	3.45	6.82	1	4.07	1.2	X	X
5E	3.45	6.82	1	4.07	X	X	1.2

Volumes are in µl. X stays for 0 µl, no volume added.

Both cross-linkers were purchased from Creative Molecules (Canada) and were dissolved in Dimethyl sulfoxide (DMSO) to 14.6mM concentration. The amount of cross-linker stock added to the substrate was calculated to provide a 20:1 molar excess over the pro-

tein. After 30 min incubation at room temperature corresponding non-labelled and ^{13}C-labelled samples were mixed in 1:1 ratio to enable quantitative determinations. So, each A sample was mixed with each B sample referred to the same condition, and each C sample was mixed with each D sample referred to the same condition. Negative control samples treated with a 1:1 mixture of both labelled/unlabelled cross-linkers, were prepared in parallel to enable proper comparison. To demonstrate the cross-link formation and quantification reproducibility, the experiment was performed in triplicates for each of the 5 conditions. The samples were subsequently characterised according to an established bottom-up approach in which digestion with a protease (i.e., trypsin) was carried out before LC-MS analysis.

3.7.3 DSBU reaction preparation

After mixing TEAD1 and FOXO4 in buffer, half of the volume of the solution was added with NO2, and the other half was added with buffer to have the same volume of the solutions. Triplicates of C and D conditions were added with 1 µl of cross-linker. The amount of cross-linker stock added to the substrate was calculated to provide a 20:1 molar excess over the protein. After 30 min incubation at room temperature C and D solutions were mixed together. The samples were subsequently characterised according to an established bottom-up approach in which digestion with a protease (i.e., trypsin) was carried out before LC-MS analysis.

3.8 Sample preparation for mass spectrometry

3.8.1 Digestion of samples

The cross-linking reaction mixtures were diluted 2x with 50 mM ethylmorpholine buffer (pH 8.5) with 10% AcN. Every reaction was added with 2 µl of benzoate solution containing 99 µl of HEPES and 1 µl of benzoate. The mixtures were left 30min at room temperature. Trypsin was dissolved in 100 mM ethylmorpholine buffer (pH 8.5) with 10% AcN to give a final concentration 20:1 (w/w) trypsin : protein. The digestion was carried out at 37 °C for 2h. A second round of trypsin was performed at 50°C overnight.

3.9 Mass spectrometry measurements

All measurements were performed using a 15T-SolariX XR FT-ICR mass spectrometer (Bruker Daltonics, Billerica, USA).

3.9.1 Intact protein analysis by ESI-Mass Spectrometry

10 µg of TEAD1/FOXO4 was desalted on protein micro trap column (C4 phase, Michrom Bioresources, USA) according manufacturer instruction, and eluted in 500 µl of 80% acetonitrile/ 0.1% formic acid. The protein was analysed using a solariX XR FT-ICR mass spectrometer (Bruker Daltonics, Germany) equipped with 15T superconducting magnet. The instrument was internally calibrated using Agilent tuning mix (Agilent Technologies, USA). Mass spectra were acquired in the positive mode over the m/z range 250–2500 with 1 M data points transient and 0.2 s ion accumulation, 8 scans were accumulated per spectrum. Data acquisition was performed using solariXControl and

interpreted by DataAnalysis 5.2.

3.9.2 Bottom-up mass spectrometry

The LC-MS analysis was performed as described in [87]. The peptide mixture was reconstituted with 30 µl of 0.1% formic acid, and 5 µl was injected using an autosampler Agilent 1290 series UPLC system on a desalting precolumn Luna Omega Polar C18 (5 µm, 100 Å, 0.3 × 30 mm). After 5 min of desalting at a flow rate of 20 µL/min, the peptides were separated by an analytical column Luna Omega Polar C18 (3 µm, 100 Å, 0.3 × 150 mm) heated to 50 °C at a flow rate of 10 µL/min with an acetonitrile gradient from 5% to 35% in 35 min. The chromatographic system was directly coupled with an electrospray ionization source of a solariX XR FT-ICR mass spectrometer (Bruker Daltonics, Germany) equipped with a 15 T superconducting magnet. The eluted peptides were analysed in a positive mode with one million transient data points over the range of 250–2500 m/z. The final spectrum was created by averaging the four following spectra with accumulation ions in the collision cell for 0.2 s. The mass spectrometer was operated in data-independent mode. The analyte at m/z 922.0098 (ESI-TOF tuning mix, Agilent Technologies, USA) was used as a lock mass. Fragmentation data was acquired with 15 eV collision voltage and 0.2 s MS/MS ion accumulation, and 2 scans were accumulated per spectra.

3.10 Data analysis

3.10.1 DSA data processing

LinX algorithm identifies cross-linked peptides by matching experimental data to a theoretical library generated based on the protein sequence, protease specificity, cross-linker reactivity and composition, and protein chemical modification. The parameters for the search were defined as follows: enzyme–trypsin (specificity – cleavage after lysine and arginine, not cleaved after modification, 3 missed cleavages); variable modification – oxidation on methionine, dead-end cross-link (hanging DSA ^{12}C6x or hanging DSA ^{13}C6x) and a fixed modification – TEAD1 Cysteine Carbamidation. Crosslinks- lys-lys DSA ^{12}C6x and lys- lys DSA ^{13}C6x, *N*-terminus-lys DSA ^{12}C6x and *N*-terminus-lys DSA ^{13}C6x. mass error 1 ppm. All cross-links identified by the LinX algorithm were manually confirmed by examination of the raw ms and ms/ms data using Data Analysis 4.4.

3.10.2 DSBU data processing

MeroX compares the acquired MS/MS spectrum to theoretical fragment ions of the respective cross-link. The MeroX software scans MS/MS spectra for two matching doublet signals indicative of an interpeptide cross-link. The search was performed with the following setup: protease –trypsin/LysC with three missed cleavage; variable modification – oxidation of methionine, stable modification – carbamidation of cysteines; mass shift for DSBU cross-linker and reporter ions – 196.085, 85.053 and 111.032; mass error – 10 ppm for precursor and 15 ppm for fragments. All cross-links identified in triplicate were manually checked in raw spectra.

3.11 Visualisation of cross-links

Models of TEAD1- DBD in complex with M-CAT DNA determined by X-ray diffraction (5GZB, PDB structure), FOXO4-DBD bound to DNA determined by X-ray diffraction (3L2C, PDB structure), and TEAD1-FOXO4-DNA complex modeled by a colleague, Dr. Jiří Černý were used. The homology model of the combined FOXO4 and TEAD1 complex with DNA 18-mer (TGGATTCCTGTAAACAGTG) was based on the available crystal structures 5gzb [26] and 3l2c [69]. The necessary nucleotide substitutions were modeled using the MMB program [88]. The backbone conformations of the resulting model were set using MMB to the nearest parent conformation using dinucleotide conformation class (NtC) alphabet as assigned at the https://dnatco.datmos.org web service [89, 90]. The geometry of the initial model was further equilibrated during a 200 ns molecular dynamics simulation in GROMACS 2021.4 [91] using the ff14SB [92] force field for the proteins and tumuc1 force field [93] for the DNA. The model was placed in a rectangular box with the 10 Å shortest distance from the walls. The box was filled with TIP3P model water and Na^+ and Cl^- ions were added to reach charge charge-neutral system with 100 mM salt concentration. The system was simulated with non-covalent cutoffs of 10 Å at 300K and 1 bar with the V-rescale modified Berendsen thermostat and Parrinello-Rahman barostat.

Chapter 4

Results and discussion

Researchers have been focusing on protein characterisation for a long time to retrieve some more information on the role of proteins including transcription factors. The information about the structural dynamics of transcription factors and interactions they are involved in, like protein-protein interactions and DNA binding protein interaction, is crucial for a deep understanding of their action and the mechanism they adopt to recognise the DNA. The proteins are part of a complex and labile system, therefore, contrary to the DNA binding domain, other regions of the proteins controlling the modulation of transcription activity, or mediating ligand interactions are complicated to delineate [94]. A combination of mass-spectrometry-based techniques and chemical cross-linking is efficient in identifying the spacial position and change in conformation of different proteins. In a chemical cross-linking reaction, the selected reagent known as cross-linker, with its two reactive groups parted by a spacer with a known length is added to a solution containing the protein or a complex. Thus, the cross-link connects two functional groups of amino acid side chains imposing a maximum distance determined by the spacer length. The cross-linked protein or complex undergoes an enzymatic cleavage to peptides which are resolved by reverse phase liquid chromatography and analysed by mass spectrometry. This thesis aimed to describe TEAD1 and FOXO4 change in conformation and their possible protein-protein interaction using LC-MS analysis and chemical cross-linking reaction.

4.1 Protein expression, purification and characterisation

TEAD1 and FOXO4 were recombinantly produced in *E.Coli* BL21 bacteria. After the expression induced by IPTG to obtain TEAD1-DBD and FOXO4-DBD proteins and the elution of the proteins on a resin, the proteins were purified by affinity chromatography and stored at -80 °C. The proteins were subjected to a SEC chromatography with HEPES buffer to exchange the buffer. The proteins were collected in the elution time from 10 to 20 according to the graph in figure 4.1.

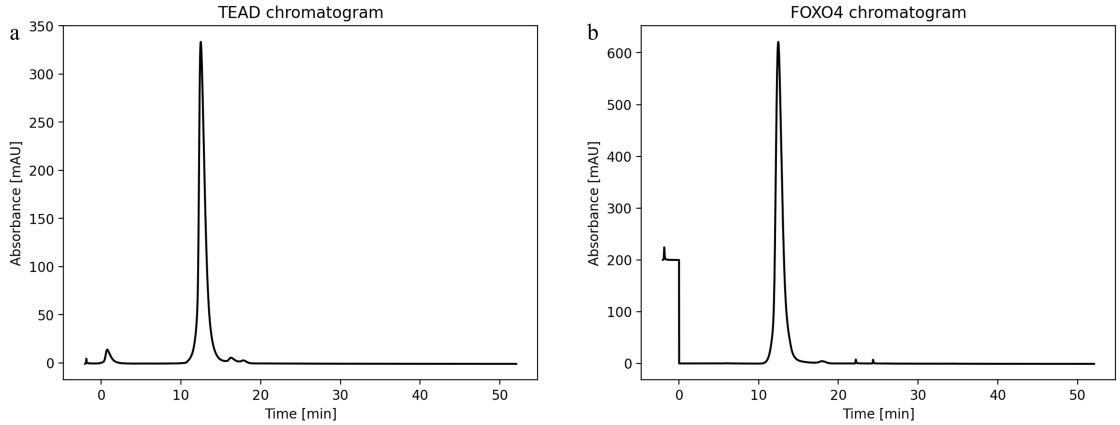


Figure 4.1: *Record of size exclusion chromatography of TEAD1 and FOXO4. The black line represents TEAD1-DBD and FOXO4-DBD elution absorbance measured by spectrophotometer at a wavelength of 280 nm.* ***a****. TEAD1-DBD elution. The peak of absorbance represents the protein.* ***b****. FOXO4-DBD elution. The peak of absorbance represents the protein.*

A LDS-PAGE was performed to verify the integrity of the proteins. The performed LDS-PAGE in figure 4.2 revealed the presence of intact proteins. In lane 1 and 3, two clear bands at molecular weight of approximately 15 kDa are detectable.

Finally, the protein samples were desalted and submitted to MS analysis to characterise the integral proteins. TEAD1 theoretical $[M+H]^+$ mass is 9668.087 Da. Deconvoluted mass spectrum determined $[M+H]^+$ 9669.091 with 1.00 ppm error and that confirms the TEAD1-DBD to be successfully expressed and prepared. Refer to figure 4.3. Whereas

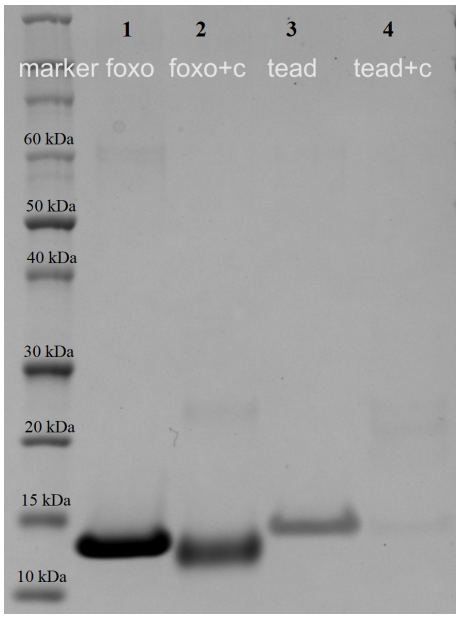


Figure 4.2: *LDS-PAGE gel of TEAD1 and FOXO4.* ***Lane 1*** *FOXO4 control.* ***Lane 2****, FOXO4 after cross-linking reaction with DSA.* ***Lane 3****, TEAD1 control.* ***Lane 4****, TEAD1 after cross-linking reaction with DSA. The marker is Novex Sharp Pre-stained Protein Standard.*

FOXO4 theoretical $[M+H]^+$ mass is 14818.5635 Da. Deconvoluted mass spectrum determined $[M+H]^+$ 14818.559 with 0.00 ppm error and that confirms the FOXO4-DBD to be successfully expressed and prepared. Refer to figure 4.4

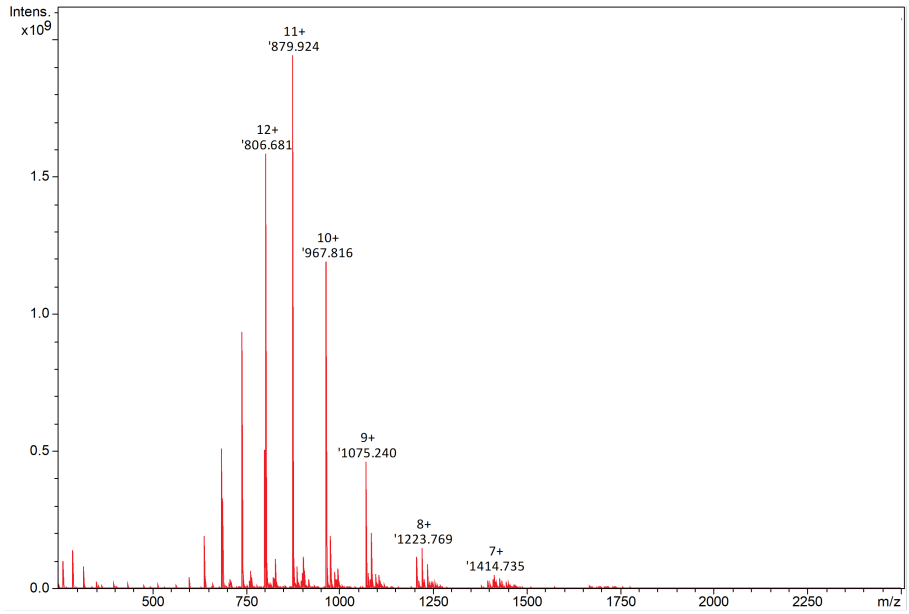


Figure 4.3: *Deconvoluted mass spectrum of TEAD1-DBD protein obtained using FT-ICR MS instrument.*

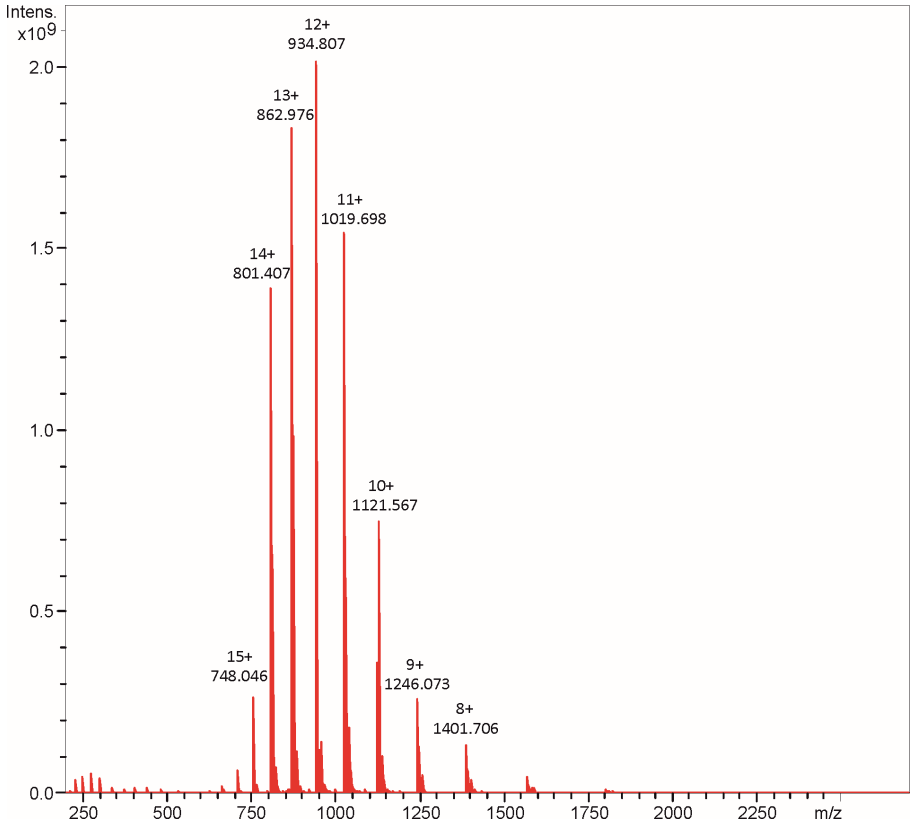


Figure 4.4: *Deconvoluted mass spectrum of FOXO4-DBD protein obtained using FT-ICR MS instrument.*

4.2 Experiments setting

4.2.1 DSA

The characterisation of TEAD1 and FOXO4 was conducted through an experiment of five different conditions to depict any change in conformation once the proteins are in solution with the oligonucleotide in a protein-DNA complex, and any contact and change in conformation when they are in a ternary complex TEAD1-FOXO4-DNA or a protein-protein complex. The samples were subjected to protein-protein cross-linking. Two different cross-linkers were used to better understand the dynamics. For the quantification of the conformational changes, Disuccinimidyl adipate (DSA) was used. DSA, an amine-reactive cross-linker that reacts with solvent-accessible lysine-NH_2 groups, was used in

its non-labelled light (^{12}C6) and labelled heavy (^{13}C6) forms. The cross-linker is a 9 Å long spacer that creates a bridge between the cross-linked residues, imposing a rigid interresidue distance. One form of the cross-linker was added to the proteins at defined conditions, while the second form was added to the proteins under different conditions concerning the first ones. The mixtures containing the light and heavy forms of DSA were then mixed in a ratio of 1:1 to enable a subsequent quantification of the cross-links. Part of the reactions were run on a LDS-PAGE gel to verify the cross-link reaction. According to the figure 4.5a, in lane 1, the band at 10 kDa corresponds to TEAD1 cross-linked with DSA C12/C13. The band close to 15 kDa represents the fraction of TEAD1 non-cross-linked. In lanes 2-3, the band at 10 kDa corresponds to TEAD1 cross-linked with DSA. The band close to 15 kDa represents the fraction of TEAD1 non cross-linked, whereas the band at 20 kDa can be related to artifacts of TEAD1 dimers cross-linked and not cross-linked. In lanes 4-5, the band at 10 kDa corresponds to TEAD1 and FOXO4 cross-linked with DSA. The band close to 15 kDa represents the fraction of TEAD1 and FOXO4 non cross-linked, whereas the band at 20 kDa can be related to artifacts of TEAD1 dimers and TEAD1-FOXO4-DNA complex cross-linked and not cross-linked. The same can be said for lanes 6-7 where bands of non-cross-linked TEAD1 and FOXO4 are not clearly distinguishable probably because, in this case, TEAD1 and FOX4 were in solution together from the beginning and they tend to associate. According to the figure 4.5b, in lane 1, the band at approximately 10 kDa corresponds to FOXO4 cross-linked with DSA C12/C13. The band at 20 kDa can be an artifact of FOXO4 dimers. In lanes 2-3, the band at 10 kDa corresponds to FOXO4 cross-linked with DSA and the band at 10 kDa can be artifacts of FOXO4 dimers. In lines 4-5, the bands (which are not that distinguishable)

at approximately 10 kDa corresponds to FOX04 and TEAD1 cross-linked, but also to the proteins cross-linked within the complex. The band at 20 kDa can be artifacts of TEAD1 and FOXO4 dimers, and complex of TEAD1-FOXO4-DNA.

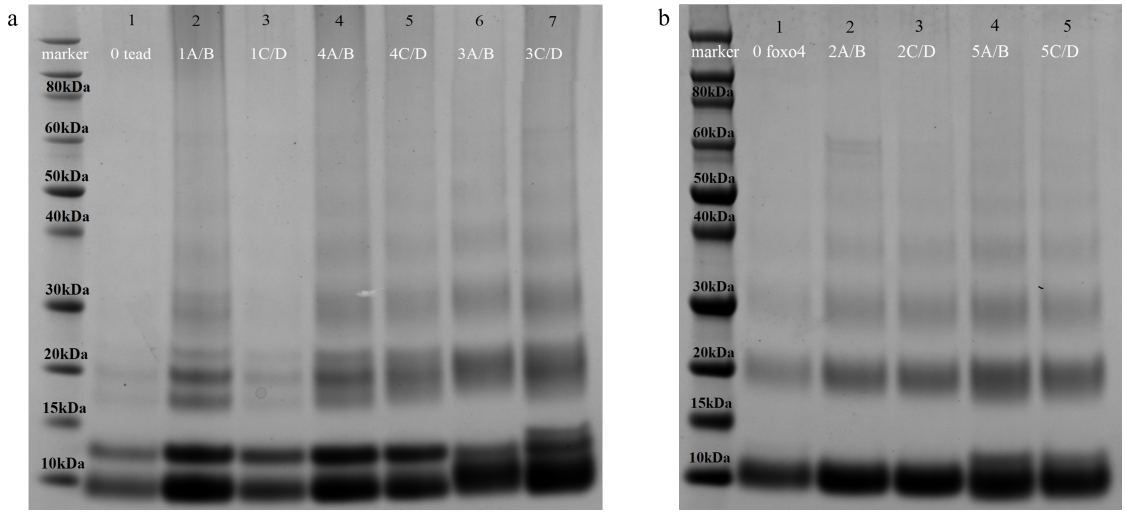


Figure 4.5: *LDS-PAGE gels of TEAD1 and FOXO4 cross-link reactions. The marker is Novex Sharp Pre-stained Protein Standard.*
***a**. LDS-PAGE of TEAD1.*
Lane 1, TEAD1 cross-linked with C12/C13.
Lane 2, TEAD1 cross-linked with DSA C12 in the absence and with DSA C13 in the presence of DNA.
Lane 3, TEAD1 cross-linked with DSA C13 in the absence and with DSA C12 in the presence of DNA.
Lane 4, TEAD1 cross-linked in solution with DNA with DSA C12 in the absence of FOXO4 and with DSA C13 in the presence of FOXO4.
Lane 5, TEAD1 cross-linked in solution with DNA with DSA C13 in the absence of FOXO4 and with DSA C12 in the presence of FOXO4.
Lane 6, TEAD1 and FOXO4 cross-linked with DSA C12 in the absence of DNA and with DSA C13 in the presence of DNA.
Lane 7, TEAD1 and FOXO4 cross-linked with DSA C13 in the absence of DNA and with DSA C12 in the presence of DNA.
***b**. LDS-PAGE of FOXO4.*
Lane 1, FOXO4 cross-linked with C12/C13.
Lane 2, FOXO4 cross-linked with DSA C12 in the absence OF DNA and with DSA C13 in the presence of DNA.
Lane 3, FOXO4 cross-linked with DSA C13 in the absence and with DSA C12 in the presence of DNA.
Lane 4, FOXO4 cross-linked in solution with DNA with DSA C12 in the absence of TEAD1 and with DSA C13 in the presence of TEAD1.
Lane 5, FOXO4 cross-linked in solution with DNA with DSA C13 in the absence of TEAD1 and with DSA C12 in the presence of TEAD1.

4.2.2 DSBU

To validate the previous experiment, a second cross-linker was used. DSBU (BuUrBu), is a urea-based cross-linker that reacts rapidly and irreversibly with amino functions in proteins, mainly neighbouring lysine groups. However, NHS esters are also reactive toward other nucleophiles, such as serine, threonine, and tyrosine residues, to a lower extent. The cross-linker has a 12.5 Å length that defines the distance between the cross-linked lysine functions. The distance constraints obtained from the analysis of the chemical cross-linkers gave some valuable information about the proteins' behavior throughout all the experiment conditions.

4.3 Data analysis and quantification

The LC-MS data obtained after samples digestion and analyses by LC-MS for the DSA experiment were searched using LinX software able to rapidly assign, evaluate, and validate mass-spectrometric data sets. Refer to figure S1. The software visualises and validates every detected cross-link in its heavy and light form. At a specific scatter, cross-linked peptides with the assigned charged molecular masses were manifested in spectra as doublet peaks. Since the program cannot determine the most abundant population of cross-links in the mixture, the quantification was done taking into account the cross-link ion intensities. Concerning DSBU MeroX2.0 software was used to analyze the LC-MS data and identify the cross-links. The candidates were manually selected for their intensities.

Quantification is an essential part of data analysis. The intensities of the monoisotopic peak given by the heavy and light form of the DSA cross-linker for each detected cross-

link were used in a ratio. The most abundant isoform of the cross-link represents how the proteins change in their conformation depending on the set condition. The experiments were performed in triplicates. For the quantification, the intensities of each cross-link per triplicate were considered. An average was calculated. Only the cross-links that were found in all the triplicates were taken into account.

4.4 Visualisation of results

Models of TEAD1- DBD in complex with M-CAT DNA determined by X-ray diffraction (5GZB, PDB structure), FOXO4-DBD bound to DNA determined by X-ray diffraction (3L2C, PDB structure) and TEAD1-FOXO4-DNA complex modeled by a colleague, Dr. Jiří Černý, were exerted in PyMOL to visualise the cross-links. The distance between the α-carbon atoms of two cross-linked lysines can reach a maximum of 25-30 Å for DSA (spacer length 9 Å) and 30–35 Å for DSBU (spacer length 12.5 Å) if the length and flexibility of lysine are considered. Most of the identified cross-links fell within this distance. Even though sometimes the cross-links did not fall within the distances, they were retained since it is not known if the model used fits perfectly the structure of the proteins and their conformational changes in solution.

4.5 Characterisation of TEAD1 and FOXO4 with DSA experiment

4.5.1 TEAD1 change in conformation when the oligonucleotide is added to the solution

The experiment was set in a way to unveil TEAD1 change in conformation when the oligonucleotide is added to the solution. The light form of the DSA cross-linker was

used to identify all the possible cross-links when the protein is free in solution, whereas the heavy form of the cross-linker was used to describe the cross-links that form when the oligonucleotide is added to the solution. Seven cross-links were detected for this experiment. Please refer to Table4.1 where all the depicted cross-links are summarised with the quantifying ratio, as well as the standard deviation. The low values of standard deviation suggest that differences between measurements were minimal.

Table 4.1: Summary of identified cross-links for experiment TEAD1+DSAC12 vs TEAD1+NO2+DSAC13

Peptide	**Modified AA**	**DSA C12:C13**
62-66x82-84	K.64-K.82	65:35 ± 0.08
57-61x62-66	K.59-K.64	48:52 ± 0.04
69-80x82-84	K.69-K.82	27:73 ± 0.04
62-66x69-80	K.64-K.69	52:48 ± 0.07
57-61x69-80	K.59-K.69	73:27 ± 0.03
39-50x57-61	K.46-K.59	7:93 ± 0.01
57-61x82-84	K.59-K.82	87:13 ± 0.02

Data obtained after LC-MS analysis of digested TEAD1 protein. The identified peptides comprising the linked lysines and their quantifying ratios with the standard deviation are reported.

For visualisation of the data in PyMOL, 5GZB PDB structure was selected as aforementioned. All the identified cross-links on TEAD1 structure fell within 30 Å distance as imposed by the cross-linker length. Three cross-links out of the seven identified seemed to be most characterised when TEAD1 is free in solution: k.64-k.82 whose quantifying ratio is 65:35, K.59-K.69 whose quantifying ratio is 73:27, and K.59-K.82 whose quantifying ratio is 87:13, represented in blue in figure4.6. Then, when the oligonucleotide is added to the solution different cross-links were depicted: K.69-K.82 whose quantifying ratio is 27:73, and K.46-K.59 whose quantifying ratio is 7:93, represented in light blue in figure4.7. Only two out of seven cross-links remain neutral to the changed condition: K.59-K.64 whose quantifying ratio is 48:52 and K.64-K.69 whose quantifying ratio is 52:48, represented in yellow in figure4.6, 4.7. The cross-links K.64-K.82, K.59-K.69,

and K.59-K.82 which involve the lysines within H2 (k.59), L2 (k.64), and H3 (k.82), are less represented when the DNA is in solution. Since the cross-links comprising lysines within H2 and L2 are not affected by the presence of the oligonucleotide in solution, a plausible explanation might be that helix H3 bends towards the groove creating a distance between this helix itself and the Helix H2 and the loop L2 which cannot be covered by the cross-linker. On the contrary, K.46-K.59 seems to be well characterised when the DNA is present, confirming that loop L1 might lose its flexibility upon DNA binding to get close to helix H2. In this way, it better interacts with the minor groove. The same is true for the cross-link K.69-K.82 which involves two lysines of helix H3 suggesting that H3 is actually bending when it lies on the DNA. This behaviour was seen in previous studies where H3 and L1 were described to be involved in DNA binding. H3 is inserted into the major groove and rotates 30 degrees relative to helices H1 and H2 to better fit the groove. L1, instead, interacts with the minor grove [25, 26].

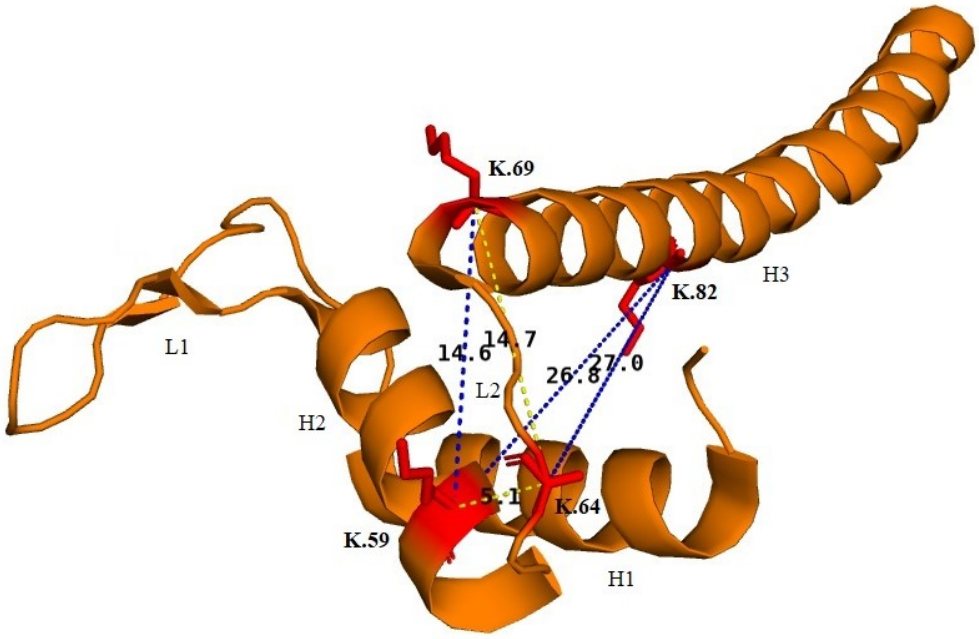


Figure 4.6: *Visualisation of the identified cross-links in X-ray diffraction model of TEAD1- DBD in complex with M-CAT DNA (5GZB). Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (blue = ratio 60-100, yellow=40-60)*

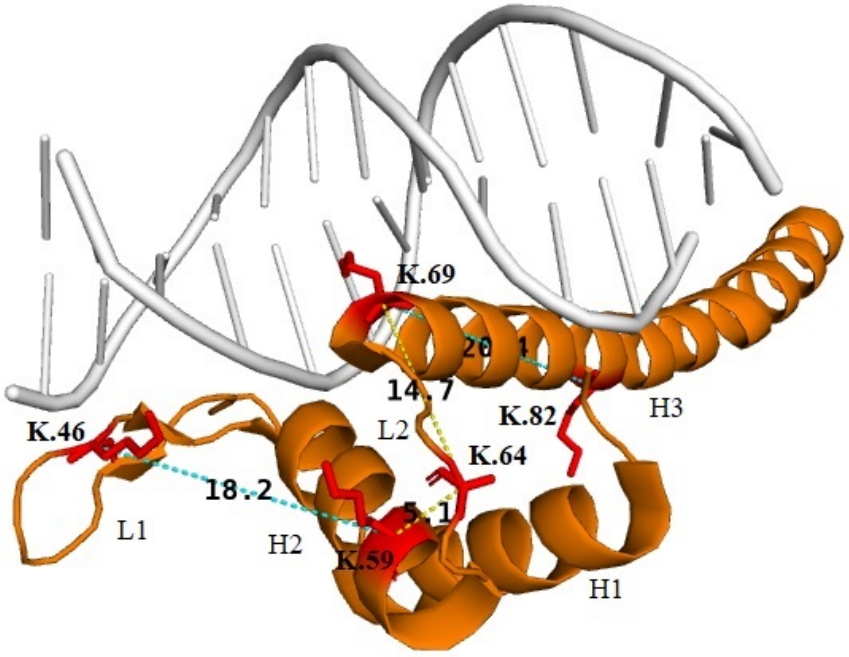


Figure 4.7: *Visualisation of the identified cross-links in X-ray diffraction model of TEAD1- DBD in complex with M-CAT DNA (5GZB). Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (blue = ratio 60-100, yellow=40-60)*

4.5.2 FOXO4 change in conformation when the oligonucleotide is added to the solution

The experiment was set in a way to unveil FOXO4 change in conformation when the oligonucleotide is added to the solution. The light form of the DSA cross-linker was used to identify all the possible cross-links when the protein is free in solution, whereas the heavy form of the cross-linker was used to describe the cross-links that form when the oligonucleotide is added to the solution. Three cross-links were detected for this experiment. Please refer to Table4.2 where all the depicted cross-links are summarised with the quantifying ratio as well as the standard deviation. The low values of standard deviation suggest that differences between measurements were minimal.

Table 4.2: Summary of identified cross-links for experiment FOXO4+DSAC12 vs TEAD1+NO2+DSAC13

Peptide	Modified AA	DSA C12:C13
1-9x10-14	N.term-K.10	47:53 ± 0.03
10-14x 16-38	K.10 -K. 37	42:58 ± 0.04
1-9x16-38	A.N-terminus -K.37	31:69 ± 0.02

Data obtained after LC-MS analysis of digested FOXO4 protein. The identified peptides comprising the linked and their quantifying ratios with the standard deviation are reported.

For visualisation of the data in PyMOL, 3L2C PDB structure was selected as aforementioned. Two cross-links out of the three identified seemed to be neutral to the different conditions. They are present both when FOXO4 is free in solution and when the DNA is added to the mixture: N.term-K.10 whose quantifying ratio is 47:53, and K.10 -K. 37 whose quantifying ratio is 42:58. Only one cross-link appeared when the DNA is added to the solution: A.N-terminus -K.37 whose quantifying ratio is 31:69. The PDB model used was manually modeled to add the missing residues at the N-terminus. The refinement was done by ModRefinement https://zhanggroup.org/ModRefiner. All the identified cross-links on FOXO4 structure fell within 30 Å distance as imposed by the cross-linker

length apart from the A.N-terminus-K.37 cross-link, but as explained the model does not represent the real conformation of the protein. This can be supported by what was found in literature [69]. Having said that, it seems that the protein is in a well-packed conformation and just a few lysines are exposed to the chemical cross-linking. It is noticeable the change in conformation of FOXO4 when attaches the DNA that enables the formation of A.N-terminus -K.37 cross-link. As said before, the model did not represent the whole structure of the protein and it was manually adjusted. Therefore, it is more likely that the spatial arrangement of residues in the N-term in the figure differs from the real one. From the results obtained, it can be made an assumption on the behaviour of FOXO4 in the presence of the DNA. It seems that the A.N-terminus bends toward the backbone of the DNA and the same appears to happen also for the loop H1 which is in between helices H1-H2 which comprises K.37. This would explain the appearance of the cross-link which can now form due to the reduced distance between the two protein's regions. On one hand, those unstructured regions that are flexible might allow the susceptible groups to get close enough to each other to allow cross-link formation. On the other, the change in conformation can be seen as a decreased flexibility of the protein structure upon binding that leads to the formation of the constraint by the cross-linker. This can be related to what was seen in previous studies where cross-links were found to form between some residues of the N-terminus and some residues of the wing W1. Thus, it is supported that the N-terminus and wing W1 tend to get close creating additional contacts with the phosphate groups of DNA. Moreover, significant conformational changes upon DNA binding were observed not only in the unstructured C- and N-terminal regions but also in loop H1-H2. Contrary to the experiment of this thesis, a plethora of cross-links

formed within the helix H3 and the loop between helices H2 and H3 was observed. These regions are involved in the DNA binding; more specifically, H3 interacts with the major groove of the DNA and the loop between the helices with the minor grove of the DNA.

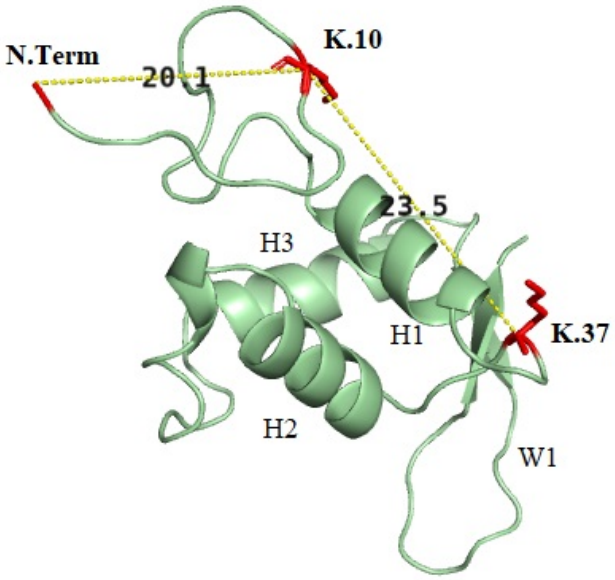


Figure 4.8: *Visualisation of the identified cross-links in X-ray diffraction model of FOXO4-DBD bound to DNA (3L2C). Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (blue = ratio 60-100, yellow=40-60)*

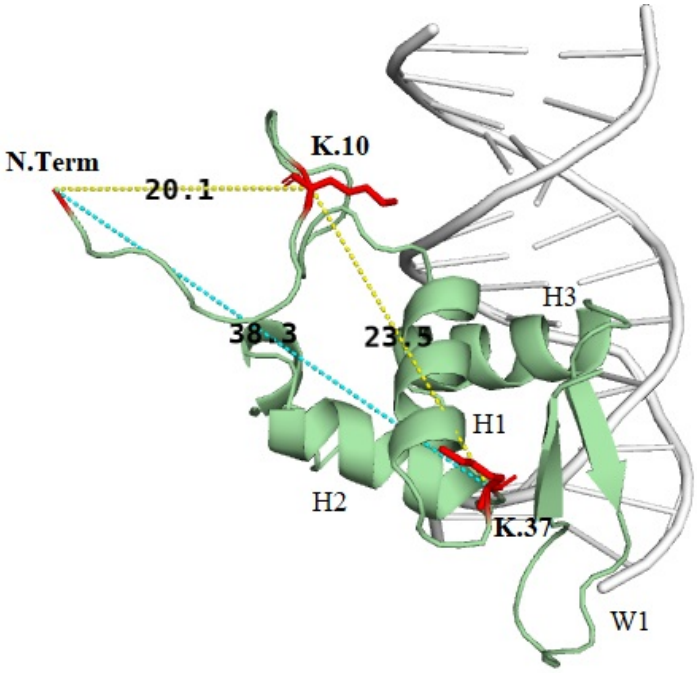


Figure 4.9: *Visualisation of the identified cross-links in X-ray diffraction model of FOXO4-DBD bound to DNA (3L2C). Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (light blue = ratio 60-100, yellow=40-60)*

4.5.3 TEAD1 change in conformation in the interaction with DNA when FOXO4 is in solution

The experiment was set in a way to disclose TEAD1 change in conformation when FOXO4 is added to the TEAD1-DNA mixture. The light form of the DSA cross-linker was used to identify all the possible cross-links when the protein is in solution with the DNA, whereas the heavy form of the cross-linker was used to describe the cross-links that form when the other transcription factor FOXO4 is added to the solution. Ten cross-links were observed. Please refer to Table4.3 where all the detected cross-links are summarised with the quantifying ratio as well as the standard deviation. The low values of standard deviation suggest that differences between measurements were minimal.

Table 4.3: Summary of identified cross-links for experiment TEAD1+NO2+DSAC12 vs TEAD1+NO2+ FOXO4+DSAC13

Protein	**Peptide**	**Modified AA**	**DSA C12:C13**
TEAD1-TEAD1	62-66x82-84	K.64-K.82	60:40 ± 0.05
TEAD1-TEAD1	57-61x82-84	K.59-K.82	63:37 ± 0.08
TEAD1-TEAD1	57-61x62-66	K.59-K.64	49:51 ± 0.01
TEAD1-TEAD1	69-80x82-84	K.69-K.82	59:41 ± 0.00
TEAD1-TEAD1	62-66x69-80	K.64-K.69	53:47 ± 0.07
TEAD1-TEAD1	57-61x69-80	K.59-K.69	60:40 ± 0.08
TEAD1-TEAD1	39-50x57-61	K.46-K.59	61:39 ± 0.02
FOXO4-TEAD1	1-9x62-66	A.N-terminus-K.64	42:58 ± 0.01
FOXO4-TEAD1	10-14x37-46	K.10-K.38	66:34 ± 0.01
FOXO4-TEAD1	16-38x69-80	K.37-K.69	27:73 ± 0.02

Data obtained after LC-MS analysis of digested TEAD1 and FOXO4 proteins. The identified peptides comprising the linked lysines and their quantifying ratios with the standard deviation are reported.

For the visualisation of the data in PyMOL, a proper model of the ternary complex was created comprising TEAD1-FOXO4-DNA. All the identified cross-links fell within 30 Å distance as imposed by the cross-linker length. Four cross-links out of ten observed are most describing TEAD1 conformation when it is in solution with the DNA: K.64-K.82 whose quantifying ratio is 60:40, K.59-K.82 whose quantifying ratio is 63:37, K.59-K.69 whose quantifying ratio is 60:40, K.46-K.59 whose quantifying ratio is 61:39 represented in blue in figure 4.10 . Unlike what was seen in the experiment dedicated to the behaviour of TEAD1 in the absence and presence of the DNA, the cross-links that there seemed to exist mostly when TEAD1 was free in solution K.64-K.82 K.59-K.82 K.59-K.69 4.6 4.7, here are seen to characterise TEAD1 conformation when attached to the DNA 4.11. These discrepancies can be explained also according to what was found in literature. An NMR study performed on TEAD11-DBD/M-CAT [24] and a crystal structure study of the TEAD14-DBD/M-CAT complex [26] differ in the description of how the protein interacts with the DNA. In the first study, loop L2 and partially helix H2 were seen to be involved in the DNA binding along with helix H3 and loop L1 whereas this was not observed to happen on the crystal structure. This may indicate the protein has two slightly different binding mechanisms. This can be comparable to what was detected in this thesis work. It might be that loop L2 which is an unstructured region, and helix H2 participate in the binding of the DNA bending towards the rotated helix H3. Thus, enabling cross-links between those two regions. Concerning K.46-K.59, it confirmed the possible rearrangement of the loop L1 to better attach the DNA. Posing this as the starting point for this experiment, it was noticed that once FOXO4 is added to the solution it leads to a slight change in conformation in TEAD1. The quantifying ratios of the cross-links

less characterising TEAD1 conformation when FOXO4 interacts with the complex do not significantly report a huge change. TEAD1 is quite stable in the DNA binding interaction. Three out of ten detected cross-links K.59-K.64 whose quantifying ratio is 49:51, K.69-K.82 whose quantifying ratio is 59:41, and K.64-K.69 whose quantifying ratio is 53:47, represented in yellow in figure 4.104.11, are neutral to the changed condition. These cross-links conform what was seen in the TEAD1 experiment where the same cross-links especially K.59-K.64 and K.64-K.69 seems to be neutral to both the presence and absence of the DNA in solution and, furthermore to the presence of FOXO4 in solution. The unchanged amount of cross-link k.69-K.82 within helix H3 in both conditions confirmed that helix H3 probably is only involved in DNA binding and is not affected by FOXO4 interaction.

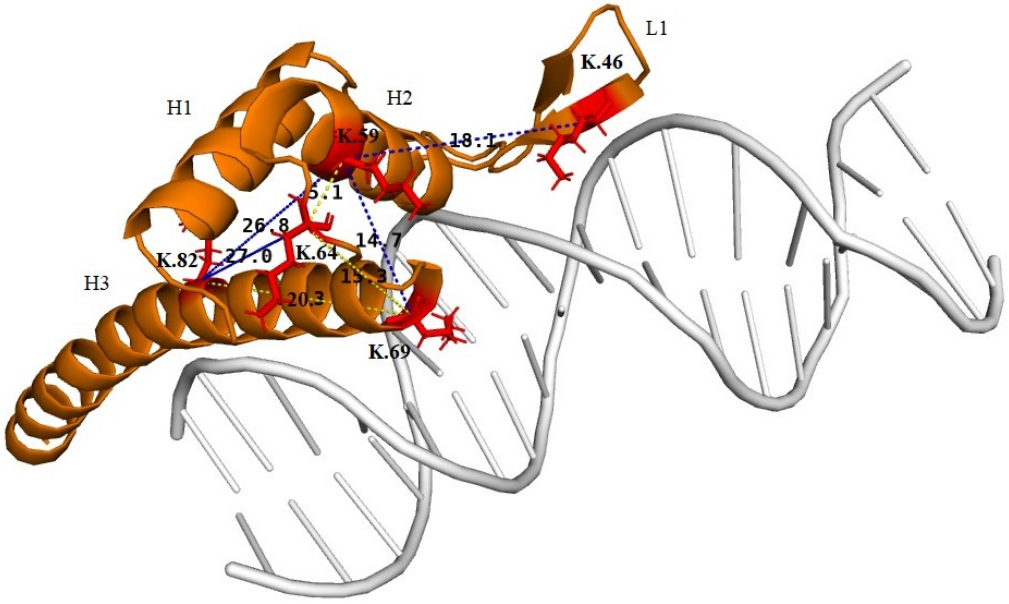


Figure 4.10: *Visualisation of the identified cross-links in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (blue = ratio 60-100, yellow=40-60)*

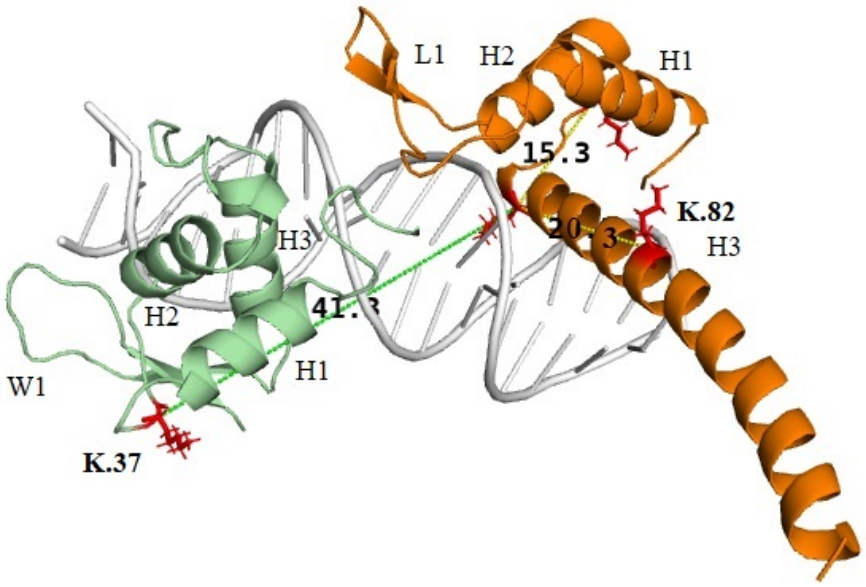


Figure 4.11: *Visualisation of the identified cross-links in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (light blue = ratio 60-100, yellow/light green=40-60)*

4.5.4 FOXO4 change in conformation in the interaction with DNA when TEAD1 is in solution

The experiment was set in a way to disclose FOXO4 change in conformation when TEAD1 is added to the FOXO4-DNA mixture. The light form of the DSA cross-linker was used to identify all the possible cross-links when the protein is in solution with the DNA, whereas the heavy form of the cross-linker was used to describe the cross-links that form when the other transcription factor TEAD1 is added to the solution. Six cross-links were observed. Please refer to Table3.3 where all the detected cross-links are summarised with the quantifying ratio as well as the standard deviation. The low values of standard deviation suggest that differences between measurements were minimal.

For the visualisation of the data in PyMOL, a proper model of the ternary complex comprising TEAD1-FOXO4-DNA was created. All the identified cross-links fell within 30

Table 4.4: Summary of identified cross-links for experiment FOXO4+NO2+DSAC12 vs FOXO4+NO2+ TEAD1+DSAC13

Protein	**Peptide**	**Modified AA**	**DSA C12:C13**
FOXO4-FOXO4	1-9x10-14	A.N-terminus-K.10	61:39 ± 0.03
FOXO4-FOXO4	1-9x16-38	A.N-terminus-K.37	56:44 ± 0.00
FOXO4-FOXO4	10-14x16-38	K.10-K.37	55:45 ± 0.08
FOXO4-TEAD1	1-9x82-84	A.N-terminus-K.82	36:64 ± 0.03
FOXO4-TEAD1	1-9x62-66	A.N-terminus-K.64	35:65 ± 0.02
FOXO4-TEAD1	16-38x69-80	K.37-K.69	12:88 ± 0.05

Data obtained after LC-MS analysis of digested FOXO4 and TEAD1 proteins. The identified peptides comprising the linked lysines and their quantifying ratios with the standard deviation are reported.

Å distances as imposed by the cross-linker length. Two cross-links A.N-terminus-K.37 whose quantifying ratio is 56:44 and K.10-K.37 whose quantifying ratio is 55:45 are neutral to the changed condition. Only A.N-terminus-K.37 is represented in yellow in figure 4.12 since FOXO4 in this model is missing some residues at the N-terminus and it was not possible to model it as did before. These cross-links were detected in both the presence and absence of TEAD1 in solution. A.N-terminus-K.37 and K.10-K.37 cross-links were depicted also in FOXO4 experiment where the change in conformation of FOXO4 was described in the presence of the DNA. In fact, it might be that the N-term and the loop between H1 and H2 adjust to the DNA to create a contact with the backbone of the latter. It was reproved by the experiment where the conformation of FOXO4 was described in the presence of the DNA before TEAD1 is added to the solution. Moreover, these experiments might attest that these regions were not subjected to any change in conformation when TEAD1 comes to interact within the complex. A.N-terminus-K.10, whose quantifying ratio is 61:39, is most represented in the absence of TEAD1 in solution and its characterisation is less once TEAD1 is within the complex 4.13.

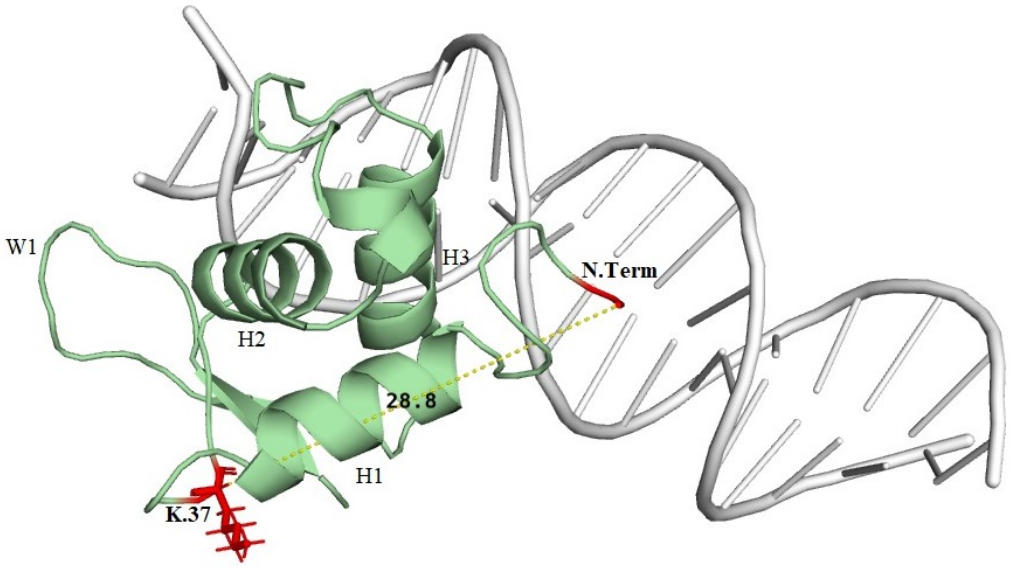


Figure 4.12: *Visualisation of the identified cross-links in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (light blue = ratio 60-100, yellow/light green=40-60)*

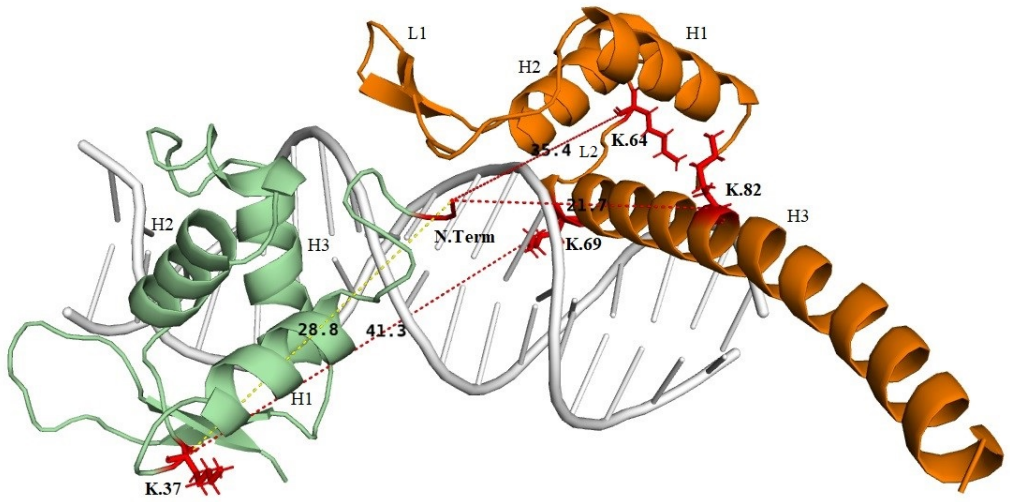


Figure 4.13: *Visualisation of the identified cross-links in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (red = ratio 60-100, yellow=40-60)*

4.5.5 TEAD1 and FOXO4 change in conformation and interaction when free in solution and in the presence of the DNA

The experiment was set in a way to disclose how TEAD1 and FOXO4 interact in solution and their conformations in the absence of the oligonucleotide and how they interact and change in conformation when the oligonucleotide is added to the mixture. The light form of the DSA cross-linker was used to identify all the possible cross-links when the proteins are free in solution, whereas the heavy form of the cross-linker was used to describe the cross-links that form when the oligonucleotide is added to the solution and the elements are in a ternary complex TEAD1-FOXO4-DNA. Sixteen cross-links were described. Please refer to Table4.5 where all the detected cross-links are summarised with the quantifying ratio as well as the standard deviation. The low values of standard deviation suggest that differences between measurements were minimal.

Table 4.5: Summary of identified cross-links for experiment TEAD1+FOXO4+DSAC12 vs TEAD1+FOXO4+NO2+DSAC13

Protein	Peptide	Modified AA	DSA C12:C13
TEAD1-TEAD1	62-66x82-84	K.64-K.82	88:12 ± 0.00
TEAD1-TEAD1	57-61x62-66	K.59-K.64	83:17 ± 0.08
TEAD1-TEAD1	69-80x82-84	K.69-K.82	66:34 ± 0.02
TEAD1-TEAD1	62-66x69-80	K.64-K.69	76:24 ± 0.02
TEAD1-TEAD1	39-50x57-61	K.46-K.59	45:55 ± 0.00
FOXO4-FOXO4	1-9x10-14	A.N-terminus-K.10	47:53 ± 0.07
FOXO4-FOXO4	1-9x16-38	A.N-terminus-K.37	30:70 ± 0.08
FOXO4-FOXO4	10-14x16-38	K.10-K.37	18:82 ± 0.00
FOXO4-TEAD1	1-9x82-84	A.N-terminus-K.82	5:95 ± 0.02
FOXO4-TEAD1	1-9x62-66	A.N-terminus-K.64	83:17 ± 0.06
FOXO4-TEAD1	1-9x57-61	A.N-terminus-K.59	61:39 ± 0.07
FOXO4-TEAD1	10-14x37-46	K.10-K.38	69:31 ± 0.01
FOXO4-TEAD1	1-9x69-80	A.N-terminus-K.69	80:20 ± 0.05
FOXO4-TEAD1	16-38x82-84	K.37-K.82	83:17 ± 0.04
FOXO4-TEAD1	51-58x67-81	K.56-K.69	55:45 ± 0.01
FOXO4-TEAD1	16-38x57-61	K.37-K.59	85:15 ± 0.07

Data obtained after LC-MS analysis of TEAD1 protein. The identified peptides comprise the linked lysines and their quantifying ratios with the standard deviation.

For the visualisation of the data in PyMOL, a proper model of the ternary complex com-

prising TEAD1-FOXO4-DNA was created. Concerning the characterisation of TEAD1 five cross-links were identified. Overall the sixteen identified cross-links in table4.5 quantitative analysis is in the Supplementary material S1. All the detected cross-links fell within 30 Å distance imposed by the cross-linker length. Four of them are mostly described when TEAD1 is free in solution with FOXO4: K.64-K.82 whose quantifying ratio is 88:12, K.59-K.64 whose quantifying ratio is 83:17, K.69-K.82 whose quantifying ratio is 66:34 and K.64-K.69 whose quantifying ratio is 76:24 4.14. One of the cross-links seems to be neutral to the different conditions: K.46-K.59 whose quantifying ratio is 45:55. These data can be compared to the ones of the experiment where TEAD1 was characterised as free in solution and in solution with the DNA. As it was seen there, also here K.64-K.82, K.59-K.64, K.64-K.69, and K.69-K.82 cross-links mostly form when TEAD1 is free in solution and they are lowly represented once it anchors the DNA. On one hand, these results confirmed that helix H3 comprising k.82 and k.69 might bend and arrange itself in the major groove creating a distance in between helix H2-loop L2 and helix H3 itself which cannot be longer covered by the cross-linker. On the other hand, the bending might allow the cross-linker within the helix to form. The cross-link K.46-K.59 which involves loop L1 and helix H2 respectively, previously was seen to form once the protein attaches to the DNA. Here it was noticed that it appeared both in the presence and absence of the DNA. It might be that the presence of FOXO4 in solution made the conformation slightly change. Overall, it can be affirmed that the contacts with FOXO4 did not affect the conformation of TEAD1 so, they might be frequent transient interactions. Regarding FOXO4 three cross-links were detected within the protein. All the identified cross-links fell within 30 imposed by the cross-linker length. Two of them

mostly formed when FOXO4 is in solution within the ternary complex with TEAD1 and the DNA: A.N-terminus-K.37 whose quantifying ratio is 30:70, and K.10-K.37 whose quantifying ratio is 18:82. Only A.N-terminus-K.10 whose quantifying ratio is 47:53 is neutral to the changed condition. These results match with the ones obtained in the experiment of characterisation of FOXO4 both in the presence and absence of the DNA. It can be reaffirmed that when FOXO4 is anchored to the DNA the N-term might bend toward the DNA backbone getting close to loop comprising K.37 enabling the cross-link formation. In the model proposed, the cross-links involving K.10 and N-term cannot be represented since some residues at the N.term are missing. By means of this experiment, it was possible to depict protein interactions when they are free in solution or within the ternary complex. Not all the observed cross-links fell within 30 Å distance imposed by the cross-linker length when they were fit in the ternary complex model like K.37-K.82 whose quantifying ratio is 83:17 and K.56-K.69 whose quantifying ratio is 55:45. The manner is mostly described when the proteins are free in solution, so even though the cross-link visualised in the model did not fit within the distance it must be considered that the proteins are free to move in solution, so the model does not represent the real dynamic situation. The same can be said for K.37-K.59 whose quantifying ratio is 85:15. For the latter, which seems to describe interactions in both conditions, the structure of the ternary complex with visualised cross-links confirmed that it cannot fit because the lysines are too distant and the presence of the DNA creates a steric impairment to the cross-links to form. But, it can be retained when the proteins are free in solution since the proteins are free to move. For all the cross-links that involve the N-term region of FOX04, it must be considered that the model is missing some residues at the N-term, so

the region should be longer than it is in the model. Through all the cross-links describing interactions between the two proteins, only one described an interaction happening when the proteins are in the ternary complex. It was noticed the presence of A.N-terminus-K.82 whose quantifying ratio is 5:95. This result is significant since the cross-link could fit perfectly in the model and it produced a high intensity during the LC-MS analysis. It is a relevant result that could implement what was seen in a previous study, where it was hypothesised a possible interaction between the two proteins.

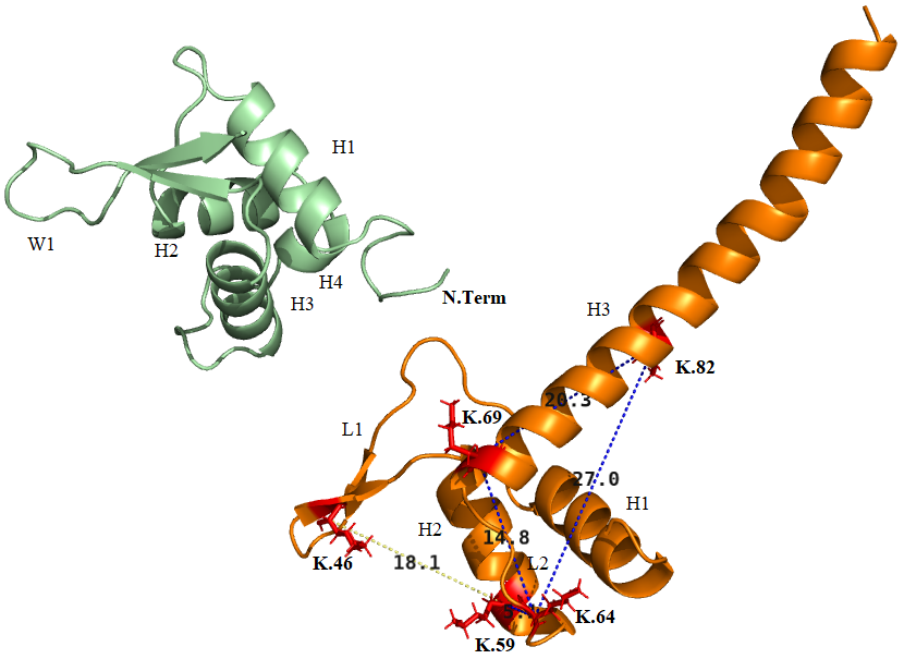


Figure 4.14: *Visualisation of the identified cross-links in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (blue = ratio 60-100, yellow/light green=40-60)*

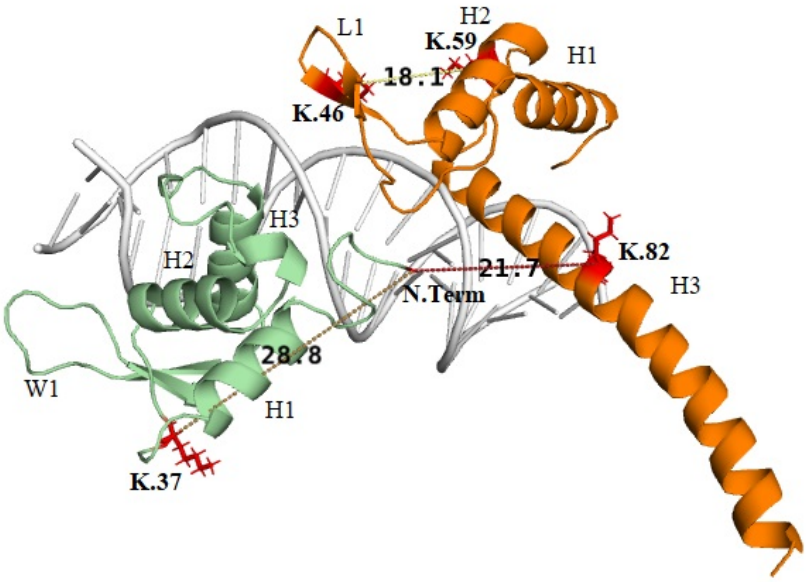


Figure 4.15: *Visualisation of the identified cross-links in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (red= ratio 60-100)*

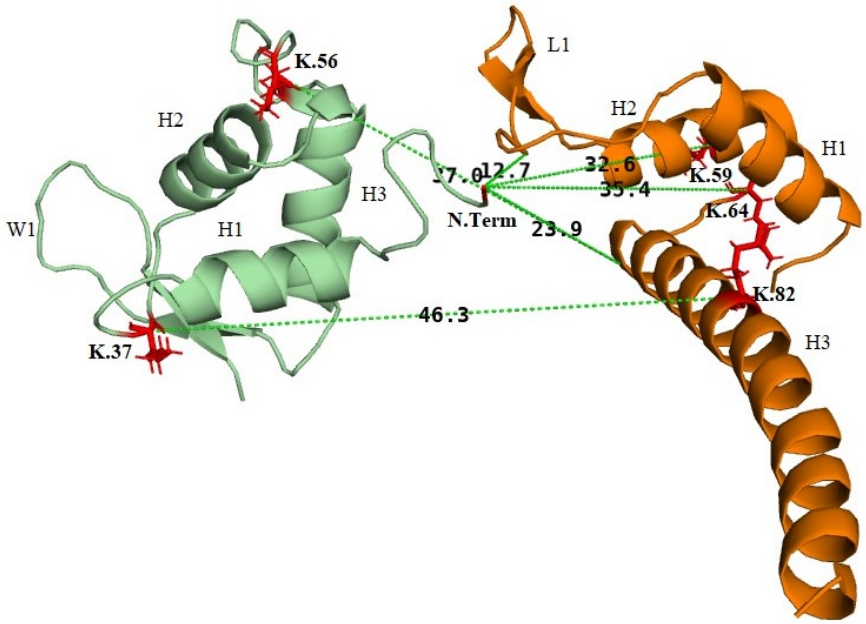


Figure 4.16: *Visualisation of the identified cross-links in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons. The colour of the lines corresponds to quantifying ratio (light green= ratio 60-100, yellow=ratio 40-60)*

4.6 Characterisation of TEAD1 and FOXO4 with DSBU experiment

The experiment was set to verify, by means of another cross-linker, if the same cross-links depicted in DSA experiment could be detected in this experiment in order to further confirm TEAD1 and FOXO4 behaviour in the different conditions examined. Two different conditions were analysed, D1 where TEAD1 and FOXO4 are free in solutions without the DNA, and C1 where the proteins are in solution with the DNA in a ternary complex TEAD1-FOXO4-DNA.

4.6.1 TEAD1 and FOXO4 change in conformation and interaction without and with DNA in solution

Ten cross-links were distinguished. Please refer to the table where all the identified cross-links are reported. For visualisation of the data in PyMOL, the complex model of the ternary complex comprising TEAD1-FOXO4-DNA was used. All the identified cross-links fell within 35 Å distance as imposed by the cross-linker length. Regarding TEAD1, four cross-links were identified when the protein is free in solution with FOXO4: S.72-K.82, K.59-K.69, K.46-K.59 and K.59-K.82. The same cross-links were seen in the DSA experiment resembling the same condition. Five cross-links were described when the protein is in solution with FOXO4 and the DNA: K.59-k.64, K.69-K.82, K.59-K.69, K.38-K.59 and K.46-K.59. This result can be aligned with the one obtained for the DSA experiment where TEAD1 and FOXO4 were studied in the ternary complex. The presence of the same cross-links was detected. It might be concluded that passing from the free-state in solution to the bound-state within the ternary complex, TEAD1 changes in conformation. More in detail, helix H3 fits the major groove of the DNA, and loop L1 moves towards the

DNA backbone, while helix H2 and loop L2 slightly move, but mainly preserve their conformation. Also, it might be assumed that TEAD1 acquires a stable conformation when anchored to the DNA without being affected by the presence of FOXO4. Concerning FOXO4, two cross-links were depicted: N-term-K.10 and N-term-K.37. These cross-links were also seen in the DSA experiments. It seems that FOXO4 does not change that much in conformation either in attaching to the DNA, within the ternary complex, or free in solution. Some interactions between the two proteins were described, but since DSBU is not a quantifying cross-link, it can just be supposed they are transient interactions. It is relevant to this thesis work, the fact that an interaction between FOXO4 and TEAD1 was observed. In the DSA experiments, where TEAD1 and FOXO4 were analysed in the ternary complex, a cross-link between N.terminus of FOXO4 and K.82 of TEAD1 was detected, thus leading to an assumption of contact between the two proteins. The cross-link found in this experiment involves a different lysine of TEAD1 K.59 which is part of helix H2 and the same N.terminus of FOXO4. It might be concluded that the two proteins interact to form a complex with the DNA and that the interaction involves mainly N.terminus of FOXO4 interchangeably associating with different residues of TEAD1.

Table 4.6: Summary of identified cross-links for condition D1

Protein	**Peptide**	**Modified AA**
TEAD1-TEAD1	69-80X82-86	S.72-K.82
TEAD1-TEAD1	57-61X69-80	K.59-K.69
TEAD1-TEAD1	39-50X57-61	K.46-K.59
TEAD1-TEAD1	57-61X82-86	K.59-K.82
FOXO4-FOXO4	0-9X10-14	A.N terminus-K.10
FOXO4-FOXO4	0-9X16-38	A.N terminus-K.37
FOXO4-TEAD1	0-9X57-61	A.N terminus-K.59
FOXO4-TEAD1	69-80X37-50	S.79-K.38
FOXO4-TEAD1	16-38X69-80	K.37-K.69

Table 4.7: Summary of identified cross-links for condition C1

Protein	Peptide	Modified AA
TEAD1-TEAD1	57-61X62-66	K.59-K.64
TEAD1-TEAD1	69-80X82-86	K.69-K.82
TEAD1-TEAD1	57-61X69-80	K.59-K.69
TEAD1-TEAD1	38-46X57-61	K.38-K.59
TEAD1-TEAD1	39-50X57-61	K.46-K.59
FOXO4-FOXO4	0-9X10-14	A.N terminus-K.10
FOXO4-FOXO4	0-9X16-38	A.N terminus-K.38
FOXO4-TEAD1	0-9X57-61	A.N terminus-K.59

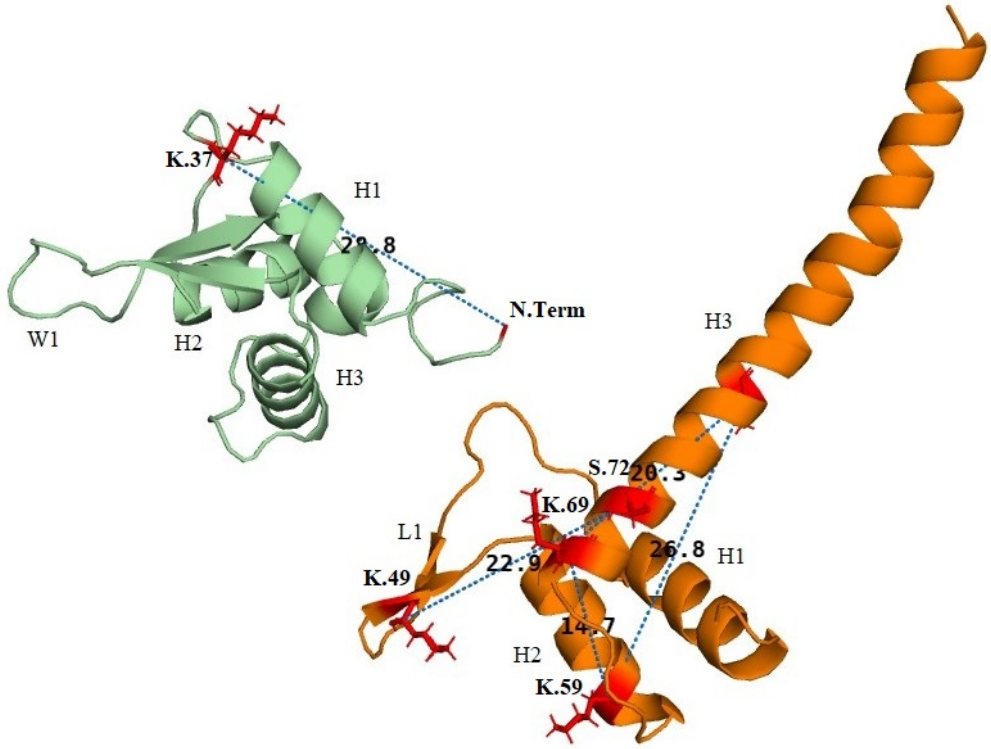


Figure 4.17: *Visualisation of the identified cross-links D1 condition in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons.*

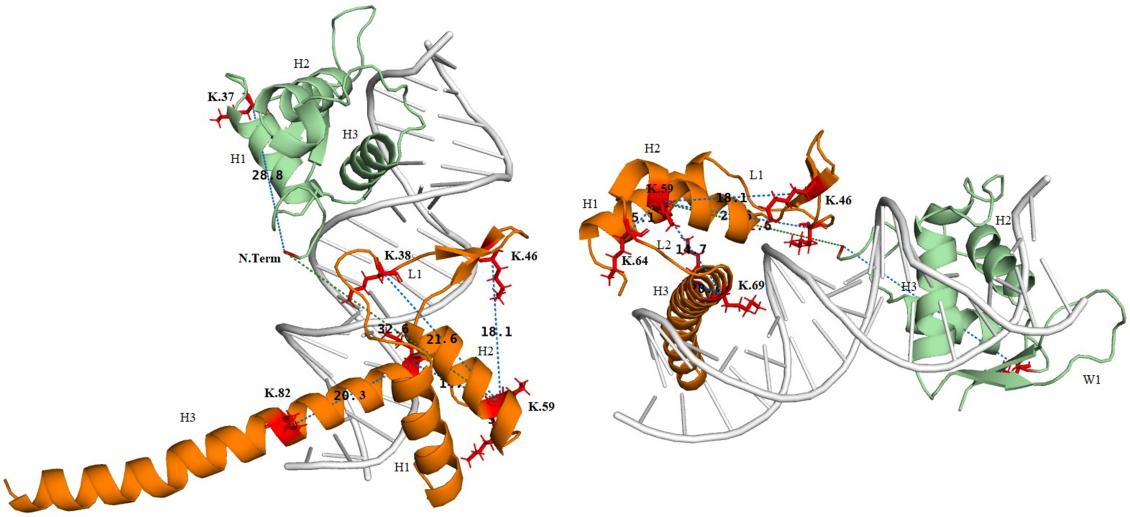


Figure 4.18: *Visualisation of the identified cross-links C1 condition in the model created by Dr. Jiří Černý. Side chains of cross-linked lysines are visualised and coloured. Lines indicate distances between alpha carbons.*

Chapter 5

Conclusion

Proteins are crucial to the functioning of all living systems. The network in which they are involved is vast, dynamic, and labile. They establish protein-protein interactions that lead to conformational changes and exertion of different functions. Proteins both capable of recognizing and binding specific sequences of DNA within regulatory regions and recruiting or interacting with proteins that participate in transcription regulation are called 'transcription factors'. It is known that any dysregulation in which they are engaged might result in diseases, especially to cancer when related to the high expression of some genes. The focus of this study was to unveil the structural change in conformation of TEAD1 and FOXO4 transcription factors when they are in different conditions either interacting with the DNA in the presence or in the absence of the other protein, or when they are mixed together in solution with or without the DNA. The work was conducted through a mass spectrometry analysis in combination with chemical cross-linking reaction which allows to work in solution. In the different conditions, the proteins were cross-linked with isotopically labelled DSA cross-linking agents. One form of the cross-linker was added to the proteins at defined conditions, while the second form was added to the proteins under different conditions concerning the first ones. In this way, it was possible to quantify the conformational changes characterising the proteins. The positions of the cross-linked amino acids combined with the known cross-linking agent length, which imposes a distance constraint between functional groups within a protein or a complex, gave some

information on the 3D structure of TEAD1, FOXO4, and their complex. In-silico models were executed in PyMOL to visualise the cross-links detected and quantified after LC-MS analysis. Along with what was already found in literature, TEAD1 in the interaction with the DNA acquires a peculiar conformation where helix H3 bends towards the major groove of the DNA creating a distance between helix H2 and loop L2. At the same time, loop L1 gets closer to helix H2 to contact the minor groove of the DNA. But it was also noticed that H2 and loop L2 are partially moving towards helix H3. It can be that the protein acquires more than one conformation especially when in solution there is another protein, as it was for the condition set. FOXO4 in the interaction with the DNA seems to bend the N-terminal towards the backbone of the DNA, the same is for the loop between H1 and H2. It acquires a more rigid conformation. The main regions involved in FOXO4 conformational change are those considered flexible. Based on previous evidence of a possible interaction between TEAD1 and FOXO4, a condition was set to identify any possible interaction once the two proteins are mixed together with the DNA.

Finally, it can be concluded that the regions of TEAD1 involved in conformational changes upon DNA binding are helix H3, loop L1, and partially helix H2 and loop L2. FOXO4 seems to preserve a defined conformation where just the N-terminus and the loop between H1 and H2 move to adjust to the DNA. Moreover, it was confirmed an interaction of TEAD1 and FOXO4 to form the complex with the DNA. The findings of this research can open the door to a thorough understanding of the behavior of the two transcription factors TEAD1 and FOXO4, which are known to play a role in the development of cancer. Additionally, it would be interesting to investigate the synergistic attachment to the DNA that results in a complex, as the role it exerts is still unknown.

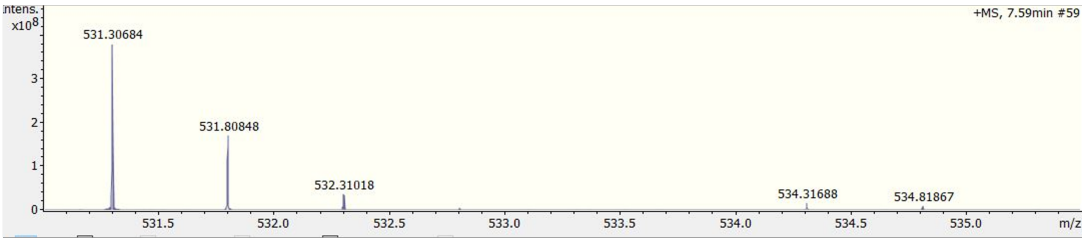
Acknowledgements

In this place, I would like to thank my supervisor Petr Novák for welcoming me to his laboratory and proposing this interesting research project and my tutor Zdeněk Kukačka for teaching me and guiding me through this experience. Thanks to Alice Sosic who supported me during my whole stay abroad and during my graduation preparation. Thanks to my sister, Paola, who is the jewel of my family, for being there every time. I would also like to thank my friends in Padua who were a great company during this master's. Thanks to Martine whose support and vicinity were extremely important to me, to Soroush who is a great talking companion, and to Mahmoud whose support was fundamental during the writing of this thesis and who bolstered me during my experience abroad. Thanks to my special friends Francesca in Brescia and Francesca in Bologna because even at a distance they were always keen on helping me. Thanks to some friends in Prague, Anna who turned my experience abroad into an amazing journey, and Claudio who suffered with me the pain of walks that were, definitely, extremely long. And, finally, thanks to my family who made all this possible.

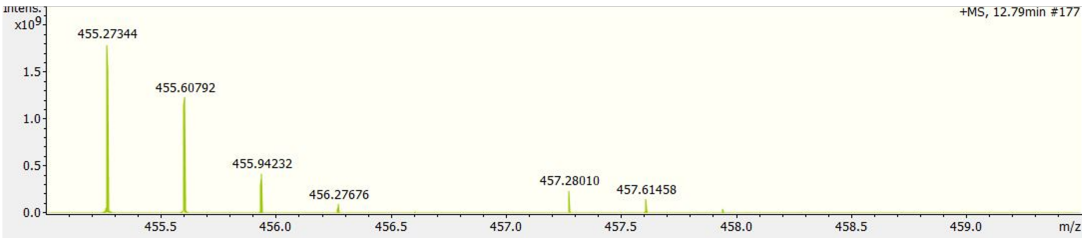
Supplementary material

Data analysis and quantification

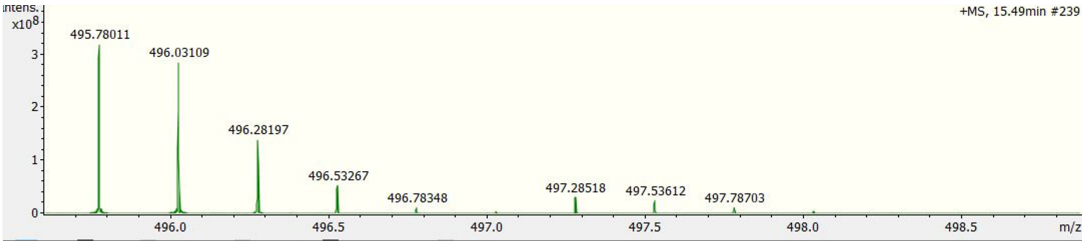
a



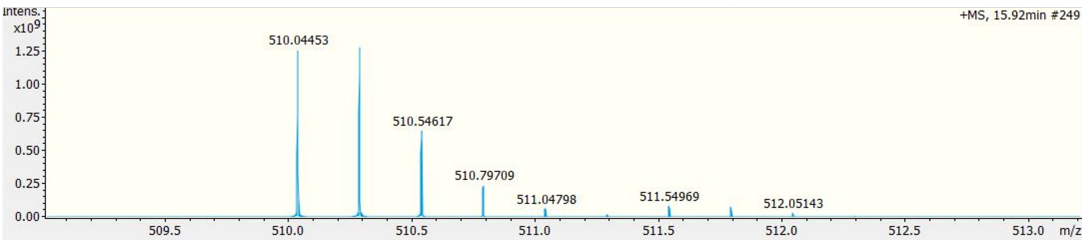
b



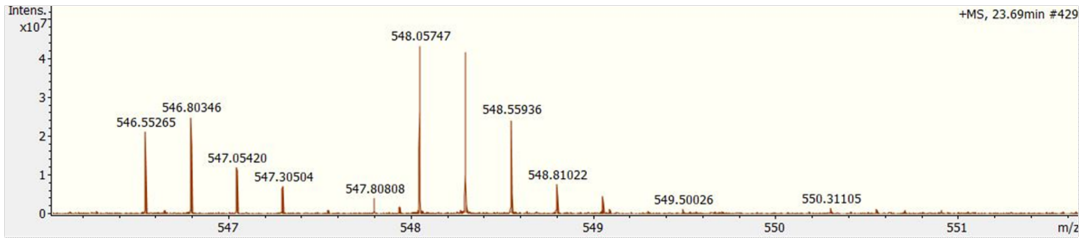
c



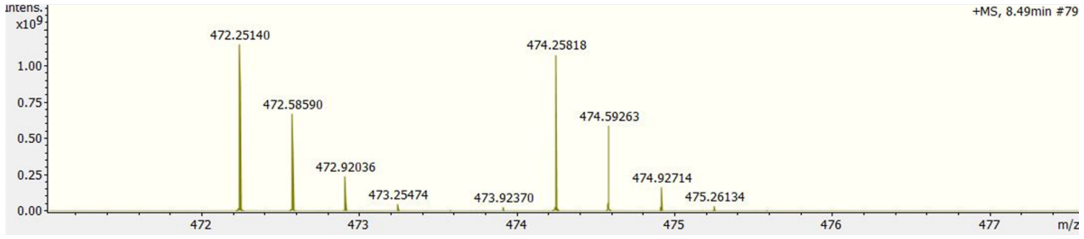
d



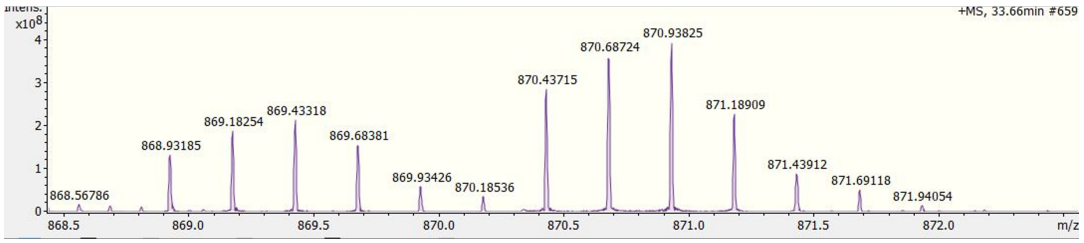
e


Intens.
x10^7
+MS, 23.69min #429
548.05747
546.80346
546.55265
548.55936
547.05420
547.30504
547.80808
548.81022
549.50026
550.31105
547
548
549
550
551
m/z

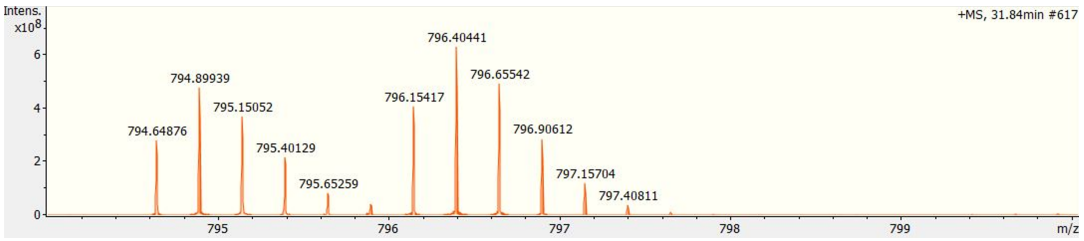
f


Intens.
x10^9
+MS, 8.49min #79
472.25140
474.25818
472.58590
474.59263
472.92036
474.92714
473.25474
475.26134
473.92370
472
473
474
475
476
477
m/z

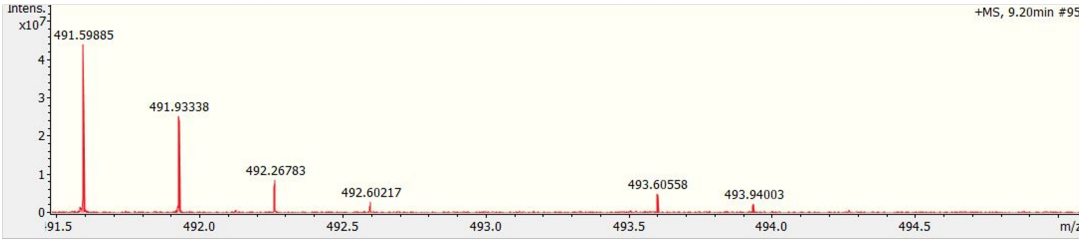
g


Intens.
x10^8
+MS, 33.66min #659
870.93825
870.68724
870.43715
871.18909
869.43318
869.18254
869.68381
868.93185
871.43912
871.69118
869.93426
870.18536
868.56786
871.94054
868.5
869.0
869.5
870.0
870.5
871.0
871.5
872.0
m/z

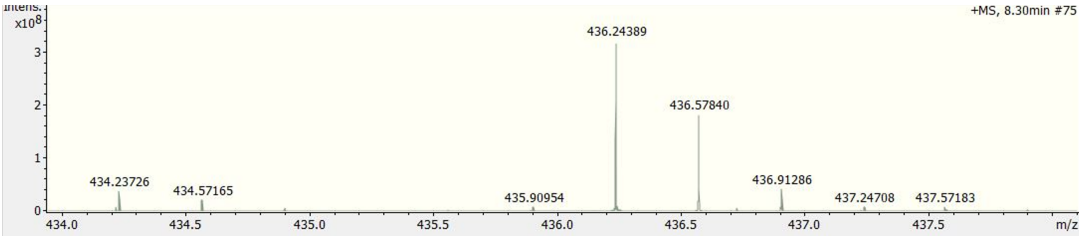
h


Intens.
x10^8
+MS, 31.84min #617
796.40441
796.65542
794.89939
796.15417
795.15052
794.64876
796.90612
795.40129
797.15704
795.65259
797.40811
795
796
797
798
799
m/z

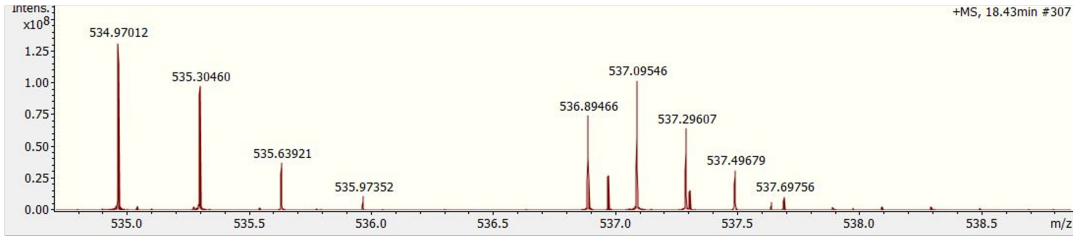
i


Intens.
x10^7
+MS, 9.20min #95
491.59885
491.93338
492.26783
492.60217
493.60558
493.94003
491.5
492.0
492.5
493.0
493.5
494.0
494.5

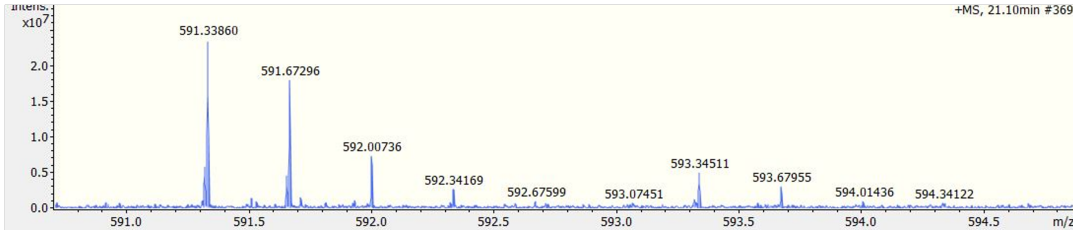
l


Intens.
x10^8
+MS, 8.30min #75
436.24389
436.57840
434.23726
436.91286
434.57165
435.90954
437.24708
437.57183
434.0
434.5
435.0
435.5
436.0
436.5
437.0
437.5
m/z

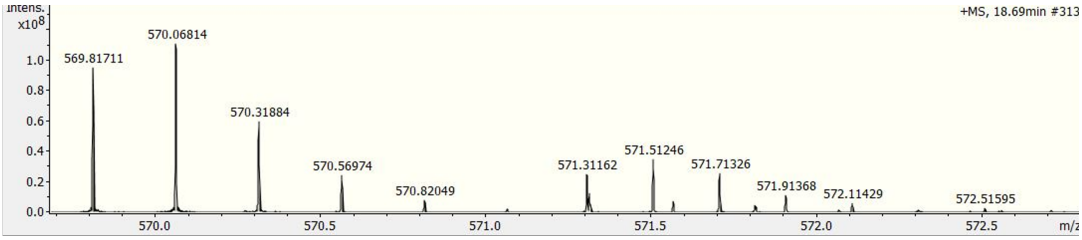
m


Intens.
x10^8
+MS, 18.43min #307
534.97012
535.30460
535.63921
535.97352
536.89466
537.09546
537.29607
537.49679
537.69756
m/z

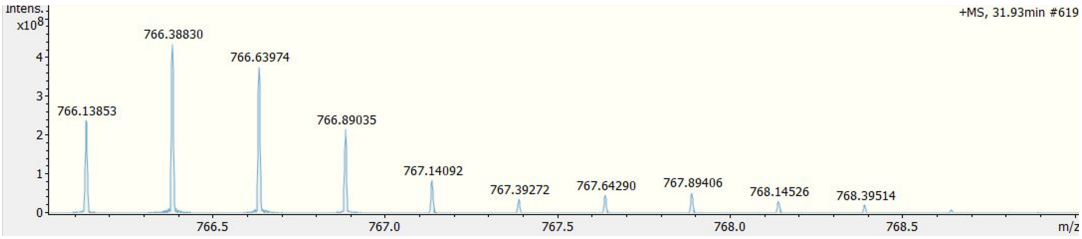
n


Intens.
x10^7
+MS, 21.10min #369
591.33860
591.67296
592.00736
592.34169
592.67599
593.07451
593.34511
593.67955
594.01436
594.34122
m/z

o


Intens.
x10^8
+MS, 18.69min #313
569.81711
570.06814
570.31884
570.56974
570.82049
571.31162
571.51246
571.71326
571.91368
572.11429
572.51595
m/z

p


Intens.
x10^8
+MS, 31.93min #619
766.13853
766.38830
766.63974
766.89035
767.14092
767.39272
767.64290
767.89406
768.14526
768.39514
m/z

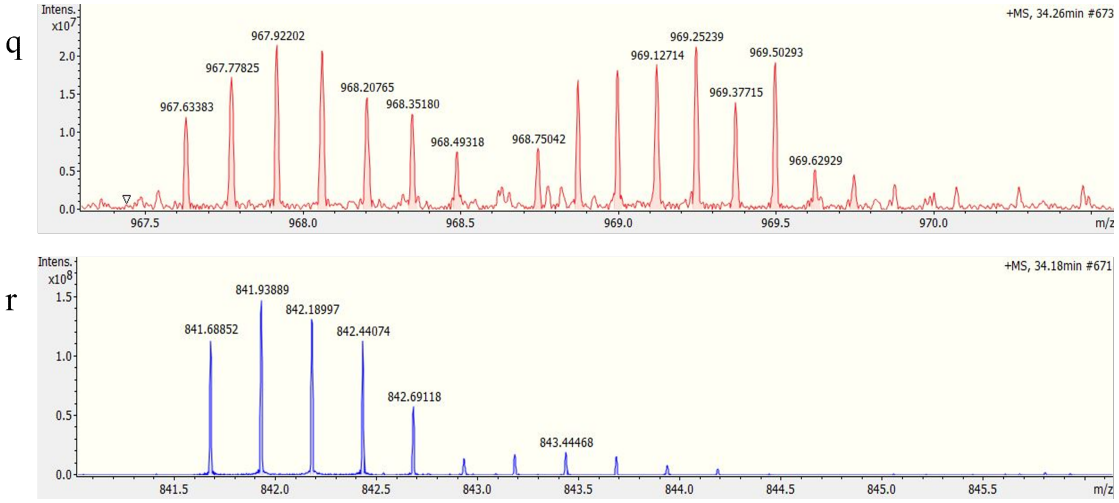


Figure S1: *Mass spectra of cross-linked residues in experiment 3 (refer to 3) by a 1:1 mixture of DSAC12/DSAC13. Mass spectrum of the identified cross-links used for the quantification of conformational changes in the proteins.* ***a****. Peptide 62-66x82-84, cross-link K.64-K.82 with charged mass C12 531.306 m/z and charged mass C13 534.316 m/z.* ***b****. Peptide 57-61x62-66, cross-link K.59-K.64 with charged mass C12 455.273 m/z and charged mass C13 457.280 m/z.* ***c****. Peptide 69-80x82-84, cross-link K.69-K.82 with charged mass C12 495.780 m/z and charged mass C13 497.285 m/z.* ***d****. Peptide 62-66x69-80, cross-link K.64-K.69 with charged mass C12 510.044 m/z and charged mass C13 511.549 m/z.* ***e****. Peptide 39-50x57-61, cross-link K.46-K.59 charged mass C12 46.552 m/z and charged mass C13 548.057 m/z.* ***f****. Peptide 1-9x10-14, cross-link A.N-terminus-K.10 with charged mass C12 472.251 m/z and charged mass C13 474.258 m/z.* ***g****. Peptide 1-9x16-38, cross-link A.N-terminus-K.37 with charged mass C12 869.183 m/z and charged mass C13 870.688 m/z.* ***h****. Peptide 10-14x16-38, cross-link K.10-K.37 with charged mass C12 794.899 m/z and charged mass C13 796.404 m/z.* ***i****. Peptide 1-9x82-84, cross-link A.N-terminus-K.82 with charged mass C12 491.598 m/z and charged mass C13 493.605 m/z.* ***l****. Peptide 1-9x62-66, cross-link A.N-terminus-K.64 with charged mass C12 434.236 m/z and charged mass C13 436.243 m/z.* ***m****. Peptide 1-9x57-61, cross-link A.N-terminus-K.59 with charged mass C12 534.970 m/z and charged mass C13 536.894 m/z.* ***n****. Peptide 10-14x37-46, cross-link K.10-K.38 with charged mass C12 591.338 m/z and charged mass C13 593.345 m/z.* ***o****. Peptide 1-9x69-80, cross-link A.N-terminus-K.69 with charged mass C12 569.817 m/z and charged mass C13 571.322 m/z.* ***p****. Peptide 16-38x82-84, cross-link K.37-K.82 with charged mass C12 766.388 m/z and charged mass C13 767.894 m/z.* ***q****. Peptide 51-58x67-81, cross-link K.56-K.69 with charged mass C12 967.633 m/z and charged mass C13 968.750 m/z.* ***r****. Peptide 16-38x57-61, cross-link K.37-K.59 with charged mass C12 841.939 m/z and charged mass C13 843.444 m/z.*

Bibliography

[1] Pascal Braun and Anne-Claude Gingras. 'History of protein–protein interactions: From egg-white to complex networks'. In: *Proteomics* 12.10 (2012), pp. 1478–1498.

[2] Daniel J Mandell and Tanja Kortemme. 'Computer-aided design of functional protein interactions'. In: *Nature chemical biology* 5.11 (2009), pp. 797–807.

[3] AS Ivanov et al. 'Direct molecular fishing in molecular partners investigation in protein–protein and protein–peptide interactions'. In: *Russian Journal of Bioorganic Chemistry* 42 (2016), pp. 14–21.

[4] Joël Janin, Ranjit P Bahadur and Pinak Chakrabarti. 'Protein–protein interaction and quaternary structure'. In: *Quarterly reviews of biophysics* 41.2 (2008), pp. 133–180.

[5] Harold Hartley. 'Origin of the word 'protein''. In: *Nature* 168.4267 (1951), pp. 244–244.

[6] SG Hedin. 'Trypsin and antitrypsin'. In: *Biochemical Journal* 1.10 (1906), p. 474.

[7] JH Kastle. 'On the vital activity of the enzymes'. In: *Science* 13.333 (1901), pp. 765–771.

[8] THE SVEDBERG. 'Mass and size of protein molecules'. In: *Nature* 123.3110 (1929), pp. 871–871.

[9] Michael J Lee and Michael B Yaffe. 'Protein regulation in signal transduction'. In: *Cold Spring Harbor perspectives in biology* 8.6 (2016), a005918.

[10] Debra L Fulton et al. 'TFCat: the curated catalog of mouse and human transcription factors'. In: *Genome biology* 10 (2009), pp. 1–14.

[11] Samuel A Lambert et al. 'The human transcription factors'. In: *Cell* 172.4 (2018), pp. 650–665.

[12] Juan M Vaquerizas et al. 'A census of human transcription factors: function, expression and evolution'. In: *Nature Reviews Genetics* 10.4 (2009), pp. 252–263.

[13] Luigi Lania, Barbara Majello and Giuliana Napolitano. 'Transcriptional control by cell-cycle regulators: A review'. In: *Journal of cellular physiology* 179.2 (1999), pp. 134–141.

[14] Domenico Accili and Karen C Arden. 'FoxOs at the crossroads of cellular metabolism, differentiation, and transformation'. In: *Cell* 117.4 (2004), pp. 421–426.

[15] JH Xiao et al. 'One cell-specific and three ubiquitous nuclear proteins bind in vitro to overlapping motifs in the domain B1 of the SV40 enhancer.' In: *The EMBO journal* 6.10 (1987), pp. 3005–3013.

[16] Joseph HA Vissers et al. 'The Scalloped and Nerfin-1 transcription factors cooperate to maintain neuronal cell fate'. In: *Cell reports* 25.6 (2018), pp. 1561–1576.

[17] André Landin-Malt et al. 'An evolutionary, structural and functional overview of the mammalian TEAD1 and TEAD2 transcription factors'. In: *Gene* 591.1 (2016), pp. 292–303.

[18] Jeffrey K Holden and Christian N Cunningham. 'Targeting the hippo pathway and cancer through the TEAD family of transcription factors'. In: *Cancers* 10.3 (2018), p. 81.

[19] Ajaybabu V Pobbati and Wanjin Hong. 'Emerging roles of TEAD transcription factors and its coactivators in cancers'. In: *Cancer biology & therapy* 14.5 (2013), pp. 390–398.

[20] Borja Belandia and Malcolm G Parker. 'Functional interaction between the p160 coactivator proteins and the transcriptional enhancer factor family of transcription factors'. In: *Journal of Biological Chemistry* 275.40 (2000), pp. 30801–30805.

[21] Chi Zhu et al. 'A non-canonical role of YAP/TEAD is required for activation of estrogen-regulated enhancers in breast cancer'. In: *Molecular cell* 75.4 (2019), pp. 791–806.

[22] Christoffer Tamm, Nathalie Böwer and Cecilia Annerén. 'Regulation of mouse embryonic stem cell self-renewal by a Yes–YAP–TEAD2 signaling pathway downstream of LIF'. In: *Journal of cell science* 124.7 (2011), pp. 1136–1144.

[23] Hyunbin D Huh et al. 'Regulation of TEAD transcription factors in cancer biology'. In: *Cells* 8.6 (2019), p. 600.
[24] Asokan Anbanandam et al. 'Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain'. In: *Proceedings of the National Academy of Sciences* 103.46 (2006), pp. 17225–17230.
[25] Dong-Sun Lee et al. 'A potential structural switch for regulating DNA-binding by TEAD transcription factors'. In: *Journal of molecular biology* 428.12 (2016), pp. 2557–2568.
[26] Z Shi et al. 'DNA-binding mechanism of the Hippo pathway transcription factor TEAD4'. In: *Oncogene* 36.30 (2017), pp. 4362–4369.
[27] Wei Tian et al. 'Structural and functional analysis of the YAP-binding domain of human TEAD2'. In: *Proceedings of the National Academy of Sciences* 107.16 (2010), pp. 7293–7298.
[28] Ze Li et al. 'Structural insights into the YAP and TEAD complex'. In: *Genes & development* 24.3 (2010), pp. 235–240.
[29] Ajaybabu V Pobbati et al. 'Structural and functional similarity between the Vgll1-TEAD and the YAP-TEAD complexes'. In: *Structure* 20.7 (2012), pp. 1135–1140.
[30] Shi Jiao et al. 'A peptide mimicking VGLL4 function acts as a YAP antagonist therapy against gastric cancer'. In: *Cancer cell* 25.2 (2014), pp. 166–180.
[31] Hung Yi Kristal Kaan et al. 'Crystal structure of TAZ-TEAD complex reveals a distinct interaction mode from that of YAP-TEAD complex'. In: *Scientific reports* 7.1 (2017), p. 2035.
[32] Cameron L Noland et al. 'Palmitoylation of TEAD transcription factors is required for their stability and function in Hippo pathway signaling'. In: *Structure* 24.1 (2016), pp. 179–186.
[33] Annemiek Beverdam et al. 'Yap controls stem/progenitor cell proliferation in the mouse postnatal epidermis'. In: *Journal of Investigative Dermatology* 133.6 (2013), pp. 1497–1505.
[34] Craig A Goodman et al. 'Yes-Associated Protein is up-regulated by mechanical overload and is sufficient to induce skeletal muscle hypertrophy'. In: *FEBS letters* 589.13 (2015), pp. 1491–1497.
[35] Janet H Mar and Charles P Ordahl. 'M-CAT binding factor, a novel trans-acting factor governing muscle-specific transcription'. In: *Molecular and cellular biology* 10.8 (1990), pp. 4271–4283.
[36] Sarah B Larkin, Iain KG Farrance and Charles P Ordahl. 'Flanking sequences modulate the cell specificity of M-CAT elements'. In: *Molecular and cellular biology* 16.7 (1996), pp. 3742–3755.
[37] IK Farrance, Janet H Mar and CP Ordahl. 'M-CAT binding factor is related to the SV40 enhancer binding factor, TEF-1.' In: *Journal of Biological Chemistry* 267.24 (1992), pp. 17234–17240.
[38] Aziz Khan et al. 'JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework'. In: *Nucleic acids research* 46.D1 (2018), pp. D260–D266.
[39] JF Knight et al. 'TEAD1 and c-Cbl are novel prostate basal cell markers that correlate with poor clinical outcome in prostate cancer'. In: *British journal of cancer* 99.11 (2008), pp. 1849–1858.
[40] Yun Liu et al. 'Increased TEAD4 expression and nuclear localization in colorectal cancer promote epithelial–mesenchymal transition and metastasis in a YAP-independent manner'. In: *Oncogene* 35.21 (2016), pp. 2789–2800.
[41] L Jia et al. 'KLF5 promotes breast cancer proliferation, migration and invasion in part by upregulating the transcription of TNFAIP2'. In: *Oncogene* 35.16 (2016), pp. 2040–2051.
[42] Y Zhou et al. 'TEAD1/4 exerts oncogenic role and is negatively regulated by miR-4269 in gastric tumorigenesis'. In: *Oncogene* 36.47 (2017), pp. 6518–6530.
[43] Bin Zhao et al. 'TEAD mediates YAP-dependent gene induction and growth control'. In: *Genes & development* 22.14 (2008), pp. 1962–1971.
[44] Bin Zhao et al. 'Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control'. In: *Genes & development* 21.21 (2007), pp. 2747–2761.
[45] Luca Azzolin et al. 'Role of TAZ as mediator of Wnt signaling'. In: *Cell* 151.7 (2012), pp. 1443–1456.
[46] Hyun Woo Park et al. 'Alternative Wnt signaling activates YAP/TAZ'. In: *Cell* 162.4 (2015), pp. 780–794.
[47] Samantha E Hiemer, Aleksander D Szymaniak and Xaralabos Varelas. 'The transcriptional regulators TAZ and YAP direct transforming growth factor β-induced tumorigenic phenotypes in breast cancer cells'. In: *Journal of Biological Chemistry* 289.19 (2014), pp. 13461–13474.

[48] Morvarid Mohseni et al. ‘A genetic screen identifies an LKB1–MARK signalling axis controlling the Hippo–YAP pathway’. In: *Nature cell biology* 16.1 (2014), pp. 108–117.
[49] Fabian Zanella, Wolfgang Link and Amancio Carnero. ‘Understanding FOXO, new views on old transcription factors’. In: *Current cancer drug targets* 10.2 (2010), pp. 135–146.
[50] Maria L Golson and Klaus H Kaestner. ‘Fox transcription factors: from development to disease’. In: *Development* 143.24 (2016), pp. 4558–4570.
[51] Detlef Weigel et al. ‘The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo’. In: *Cell* 57.4 (1989), pp. 645–658.
[52] Astrid Eijkelenboom and Boudewijn MT Burgering. ‘FOXOs: signalling integrators for homeostasis maintenance’. In: *Nature reviews Molecular cell biology* 14.2 (2013), pp. 83–97.
[53] Anne Brunet et al. ‘Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a)’. In: *Molecular and cellular biology* 21.3 (2001), pp. 952–965.
[54] Bérénice A Benayoun, Sandrine Caburet and Reiner A Veitia. ‘Forkhead transcription factors: key players in health and disease’. In: *Trends in Genetics* 27.6 (2011), pp. 224–232.
[55] Anne Brunet et al. ‘Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor’. In: *cell* 96.6 (1999), pp. 857–868.
[56] Z Fu and DJ Tindall. ‘FOXOs, cancer and regulation of apoptosis’. In: *Oncogene* 27.16 (2008), pp. 2312–2319.
[57] Meng C Wang, Dirk Bohmann and Heinrich Jasper. ‘JNK signaling confers tolerance to oxidative stress and extends lifespan in Drosophila’. In: *Developmental cell* 5.5 (2003), pp. 811–816.
[58] Xuemei Zhu, Xichen Zheng and Yi Wu. ‘Cleaved high molecular weight kininogen stimulates JNK/-FOXO4/MnSOD pathway for induction of endothelial progenitor cell senescence’. In: *Biochemical and biophysical research communications* 450.4 (2014), pp. 1261–1265.
[59] Benjamin Bourgeois and Tobias Madl. ‘Regulation of cellular senescence via the FOXO 4-p53 axis’. In: *FEBS letters* 592.12 (2018), pp. 2083–2097.
[60] Kirk L Clark et al. ‘Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5’. In: *Nature* 364.6436 (1993), pp. 412–420.
[61] Ketan S Gajiwala and Stephen K Burley. ‘Winged helix proteins’. In: *Current opinion in structural biology* 10.1 (2000), pp. 110–116.
[62] Kuang-Lei Tsai et al. ‘Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins’. In: *Journal of Biological Chemistry* 281.25 (2006), pp. 17400–17409.
[63] Katarina Psenakova et al. ‘Forkhead domains of FOXO transcription factors differ in both overall conformation and dynamics’. In: *Cells* 8.9 (2019), p. 966.
[64] Michael M Brent, Ruchi Anand and Ronen Marmorstein. ‘Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification’. In: *Structure* 16.9 (2008), pp. 1407–1416.
[65] Johan Weigelt et al. ‘Solution structure of the DNA binding domain of the human forkhead transcription factor AFX (FOXO4)’. In: *Biochemistry* 40.20 (2001), pp. 5861–5869.
[66] Tatsuo Furuyama et al. ‘Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues’. In: *Biochemical Journal* 349.2 (2000), pp. 629–634.
[67] T Obsil and V Obsilova. ‘Structure/function relationships underlying regulation of FOXO transcription factors’. In: *Oncogene* 27.16 (2008), pp. 2263–2275.
[68] DR Calnan and A Brunet. ‘The foxo code’. In: *Oncogene* 27.16 (2008), pp. 2276–2288.
[69] Evzen Boura et al. ‘Structure of the human FOXO4-DBD–DNA complex at 1.9 Å resolution reveals new details of FOXO binding to the DNA’. In: *Acta Crystallographica Section D: Biological Crystallography* 66.12 (2010), pp. 1351–1357.
[70] Jan Silhan et al. ‘14-3-3 protein masks the DNA binding interface of forkhead transcription factor FOXO4’. In: *Journal of Biological Chemistry* 284.29 (2009), pp. 19349–19360.
[71] Maria K Lehtinen et al. ‘A conserved MST-FOXO signaling pathway mediates oxidative-stress responses and extends life span’. In: *Cell* 125.5 (2006), pp. 987–1001.

[72] Kuang-Lei Tsai et al. 'Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification'. In: *Nucleic acids research* 35.20 (2007), pp. 6984–6994.
[73] Hitomi Matsuzaki et al. 'Acetylation of Foxo1 alters its DNA-binding ability and sensitivity to phosphorylation'. In: *Proceedings of the National Academy of Sciences* 102.32 (2005), pp. 11278–11283.
[74] Arndt Borkhardt et al. 'Cloning and characterization of AFX, the gene that fuses to MLL in acute leukemias with at (X; 11)(q13; q23)'. In: *Oncogene* 14.2 (1997), pp. 195–202.
[75] Taisuke Hosaka et al. 'Disruption of forkhead transcription factor (FOXO) family members in mice reveals their functional diversification'. In: *Proceedings of the National Academy of Sciences* 101.9 (2004), pp. 2975–2980.
[76] Armando van der Horst et al. 'FOXO4 is acetylated upon peroxide stress and deacetylated by the longevity protein hSir2SIRT1'. In: *Journal of Biological Chemistry* 279.28 (2004), pp. 28873–28879.
[77] Arjan B Brenkman et al. 'Mdm2 induces mono-ubiquitination of FOXO4'. In: *PloS one* 3.7 (2008), e2819.
[78] Nancy D De Ruiter, Boudewijn MT Burgering and Johannes L Bos. 'Regulation of the Forkhead transcription factor AFX by Ral-dependent phosphorylation of threonines 447 and 451'. In: *Molecular and cellular biology* 21.23 (2001), pp. 8225–8235.
[79] Anna A Szypowska et al. 'Oxidative stress-dependent regulation of Forkhead box O4 activity by nemo-like kinase'. In: *Antioxidants & redox signaling* 14.4 (2011), pp. 563–578.
[80] Shuai Jiang et al. 'Novel role of forkhead box O 4 transcription factor in cancer: bringing out the good or the bad'. In: *Seminars in Cancer Biology*. Vol. 50. Elsevier. 2018, pp. 1–12.
[81] Tracy Tzu-Ling Tang and Laurence A Lasky. 'The forkhead transcription factor FOXO4 induces the down-regulation of hypoxia-inducible factor 1α by a von Hippel-Lindau protein-independent mechanism'. In: *Journal of Biological Chemistry* 278.32 (2003), pp. 30125–30135.
[82] Ian G Macara. 'Nuclear transport: randy couples'. In: *Current biology* 9.12 (1999), R436–R439.
[83] Siew Wee Chan et al. 'A role for TAZ in migration, invasion, and tumorigenesis of breast cancer cells'. In: *Cancer research* 68.8 (2008), pp. 2592–2598.
[84] Mohd Farhan et al. 'Role of FOXO transcription factors in cancer metabolism and angiogenesis'. In: *Cells* 9.7 (2020), p. 1586.
[85] Curtis M Wilson. '[18] Staining of proteins on gels: Comparisons of dyes and procedures'. In: *Methods in enzymology*. Vol. 91. Elsevier, 1983, pp. 236–247.
[86] Zdenek Kukacka et al. 'Mapping protein structural changes by quantitative cross-linking'. In: *Methods* 89 (2015), pp. 112–120.
[87] Lukáš Fojtík et al. 'Fast fluoroalkylation of proteins uncovers the structure and dynamics of biological macromolecules'. In: *Journal of the American Chemical Society* 143.49 (2021), pp. 20670–20679.
[88] Samuel Coulbourn Flores and Russ B Altman. 'Turning limited experimental information into 3D models of RNA'. In: *Rna* 16.9 (2010), pp. 1769–1778.
[89] Jiřı Černỳ et al. 'A unified dinucleotide alphabet describing both RNA and DNA structures'. In: *Nucleic acids research* 48.11 (2020), pp. 6367–6381.
[90] Jiřı Černỳ et al. 'Structural alphabets for conformational analysis of nucleic acids available at dnatco. datmos. org'. In: *Acta Crystallographica Section D: Structural Biology* 76.9 (2020), pp. 805–813.
[91] Mark James Abraham et al. 'GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers'. In: *SoftwareX* 1 (2015), pp. 19–25.
[92] James A Maier et al. 'ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB'. In: *Journal of chemical theory and computation* 11.8 (2015), pp. 3696–3713.
[93] Korbinian Liebl and Martin Zacharias. 'Tumuc1: a new accurate DNA force field consistent with high-level quantum chemistry'. In: *Journal of Chemical Theory and Computation* 17.11 (2021), pp. 7096–7105.
[94] Yoshiaki Minezaki et al. 'Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation'. In: *Journal of molecular biology* 359.4 (2006), pp. 1137–1149.