TECHNISCHE UNIVERSITÄT MÜNCHEN

Department für Chemie

Lehrstuhl für Biomolekulare NMR-Spektroskopie

Structural and functional studies of the transcription factor FOXP2 and its regulation

Gesa Richter

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Mathias Feige
Prüfende der Dissertation:	1. Assoc. Prof. Dr. Tobias Madl
	2. Prof. Dr. Michael Sattler

Die Dissertation wurde am 18.12.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 12.05.2020 angenommen.

Dedicated to Ben

Declaration I hereby declare that parts of this thesis are already submitted to scientific international journals or already published:

Richter G, Koyani C, Bourgeois B, Ulz P, Heitzer E, von Lewinski D, Malli E, Madl T. *β-catenin is regulating FOXP2 activity by its disordered region.* FEBS, *in revision*

Abbreviations and Definitions

µM	micromolar
APC	adenomatous poliposis coli
BME	β-mercaptoethanol
cAMP	cyclic adenosine monophosphate
CBP	CREB (adenosine 3',5'-monophosphate response element-binding protein)-binding protein
CD	circular Dichroism
CMV	cytomegalovirus
CSP	chemical shift perturbations
CTBP1	C-terminal binding protein 1
dd H_2O	distilled deionized water
D_2O	deuterium oxide
DEG	differentially expressed gene
E.coli	*Escherichia coli*
EDTA	ethylendiamintetraacetate
EMSA	electromobility shift assay
FAM	fluorescein amidite
FC	fold change
FOXP1	Forkhead-box protein P1
FOXP2	Forkhead-box protein P2
$FOXP2^{\Delta helix}$	FOXP2 protein construct lacking residue 264-272
$FOXP2^{IDR}$	FOXP2 protein construct from residue 247 to 341
$FOXP2^{FH}$	FOXP2 forkhead domain from residue 504 to 594
$FOXP2^{FH-IDR}$	FOXP2 construct forkhead domain linked to IDR
$FOXP2^{R553H}$	FOXP2 forkhead domain substitution of arginine 553 to histidine
FOXP3	Forkhead-box protein P3
FOXP4	Forkhead-box protein P4
FPLC	fast protein liquid chromatography
Fw	forward
GSK3α	glycogen synthase kinase 3α
GSK3β	glycogen synthase kinase 3β
HD	Huntington's disease

HEK-293T	human embryonic kidney 293T cells
HSQC	heteronuclear single quantum coherence
ICAT	β-catenin interacting protein 1
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IPTG	isopropyl-β-D-thiogalactopyranosid
ITC	isothermal Calorimetry
kDa	kilo dalton
LB	luria Bertani
LEF1	Lymphoid enhancer-binding factor 1
M	molar
min	minute
mL	milliliter
mM	millimolar
MRI	magentic Resonance Imaging
mRNA	messenger ribonucleic acid
myr	million years ago
NaCl	sodium chloride
NaF	sodium fluoride
NFATC2	Nuclear factor of activated T-cells, cytoplasmic 2
NH_4Cl	ammonium chloride
nM	nanomolar
nm	nanometer
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser enhancement spectroscopy
OD_{600}	optical density at 600 nm
PBS	phosphate buffered saline
PIAS	Protein inhibitor of activated STAT
PKA	protein kinase A
RNA	ribonucleic acid
rpm	rotations per minute
RT	room temperature

RT-PCR	reverse-transcriptase polymerase-chain-reaction
Rv	reverse
S. cerevisiae	*Saccharomyces cerevisiae*
s	second
SAXS	small-angle x-ray scattering
SEC	size exclusion chromatography
TCF7L2	Transcription factor 7-like 2
TEMED	tetramethylethylenediamine
TEV	tobacco etch virus
TFE	trifluorethanol
Wnt	Wingless related integration site
WT	wildtype

Abstract (English):

The transcription factor Forkhead-box-protein P2 (FOXP2) is a highly conserved key regulator of embryonal development. Mutations such as R553H in the Forkhead DNA-binding domain of FOXP2 results a speech disorder. Other mutations result in various cancers, indicating that FOXP2 is an important player in signaling pathways. In this thesis I investigated FOXP2 function by determining structural features and characteristics. In order to understand the regulation of this transcription factor I identified new interaction partners, which provide information about the complex network of FOXP2 regulation. Beside biophysical methods such as NMR, ITC and SAXS I used cell-based assays to further investigate the interaction between FOXP2 and a crucial player in Wnt-signaling, the co-activator β-catenin. In my studies I discovered two interaction sites within FOXP2, one is an intrinsically disordered region (IDR) with interesting evolutionary features, as it contains the two residues which differ between human and chimpanzee. Next, I found that FOXP2 is forming a back fold leading to an intramolecular interaction, which might have impact on interactions with other proteins or might influence the DNA-binding affinities and thus the transcriptional activity of FOXP2. In order to investigate the effect of β-catenin and the back-fold within FOXP2 on the transcriptional activity of FOXP2, I performed RNA-Sequencing. Those data revealed not only various signaling pathways but also that both β-catenin and the intramolecular interaction affect the transcriptional activity of FOXP2 and thus might play an important role in human cells during embryonal development or diseases such as cancer. Beside this I discovered a novel phosphorylation site in the IDR region of FOXP2, which binds to the interaction partners I found earlier within this project. Using NMR I could show, that this phosphorylation is not affecting the interaction between FOXP2 and the novel binding partners. This phosphorylation might play a role in the regulation of FOXP2 itself or affecting the binding to other interaction partners.

Summarizing I discovered novel structural features of FOXP2, which play a crucial role in its, so far, poorly understood regulation mechanism. Additionally, I found various novel interaction partners, which give clues about the interaction network of FOXP2 and thus help to understand, how FOXP2-linked diseases are developing. I discovered a novel phosphorylation site in a region, which is important for protein-protein interactions and with evolutionary importance. This novel posttranslational modification might be crucial for protein-protein-interactions or the regulation of FOXP2 function.

Zusammenfassung:

Der Transkriptionsfaktor Forkhead-Box Protein P2 (FOXP2) ist ein stark konservierter Regulator in embryonaler Entwicklung. Mutationen, wie R553H in der Forkhead-DNA bindenden Domäne von FOXP2 führen zu einer Sprachstörung. Andere Mutationen führen zu verschiedenen Krebstypen, FOXP2 scheint daher eine wichtige Rolle in Signalwegen zu spielen. In dieser Arbeit habe ich die Funktion von FOXP2 durch das Bestimmen von strukturellen Eigenschaften untersucht. Um die Regulation dieses Transkriptionsfaktors zu verstehen, habe ich neue Interaktionspartner entdeckt, welche Indizien auf das komplexe Netzwerk der Regulation von FOXP2 geben. Neben biophysikalischen Methoden wie NMR, ITC und SAXS, habe ich zell-basierte Methoden verwendet, um die Interaktion zwischen FOXP2 und einem wichtigen Protein des Wnt-Signalwegs, dem Co-Aktivator β-catenin, zu untersuchen. In dieser Studie konnte ich zwei Bindestellen innerhalb FOXP2 lokalisieren, eine davon ist eine unstrukturierte Region mit interessanten evolutionären Eigenschaften, da sie die beiden Aminosäuren enthält, welche sich zwischen Mensch und Schimpanse unterscheiden. Daneben entdeckte ich, dass FOXP2 mit sich selbst interagieren kann. Diese Eigenschaft könnte einen Einfluss auf Interaktionen mit anderen Proteininteraktionen oder die DNA-Bindeaffinität haben und so die Transkriptionsaktivität von FOXP2 haben. Um den Effekt von β-catenin und der Interaktion innerhalb FOXP2s auf die Transkriptionsaktivität von FOXP2 zu untersuchen, nutze ich RNA-Seq. Diese Daten zeigten nicht nur einige neue Signalwege, aber auch, dass β-catenin und die intramolekulare Interaktion die Transkriptionsaktivität von FOXP2 beeinflussen und somit eine wichtige Rolle in Zellen während der Embryonalentwicklung oder Krankheiten wie Krebs spielen. Daneben entdeckte ich eine neue Phosphorylierung in FOXP2 IDR. Mittels NMR konnte ich zeigen, dass diese Phosphorylierung nicht die Bindung zu den anderen entdeckten Proteinen beeinflusst. Diese Phosphorylierung könnte eine Rolle bei der Regulation von FOXP2 selber spielen oder die Bindung zu anderen Interaktionspartnern beeinflussen. Zusammengefasst habe ich neue strukturelle Eigenschaften von FOXP2 entdeckt, welche eine wichtige Rolle in der, noch wenig verstandenen, Regulationsmechanismen spielen könnten. Zusätzlich entdeckte ich einige neue Interaktionspartner, welche Informationen auf das Interaktionsnetzwerk von FOXP2 zulassen und somit helfen, die Entwicklung FOXP2-abhängiger Krankheiten zu verstehen. Ich entdeckte eine neue Phosphorylierung, in einer wichtigen Region für

Protein-Protein Interaktionen und evolutionären Eigenschaften. Diese neue Post-translationelle Modifikation könnte wichtig für andere Protein-Protein-Interaktionen oder für die Regulation von FOXP2 sein.

List of contents

Chapter 1

1. Introduction

1.1 Transcription and transcription factors

The human genome consists of 3.2 billion DNA base pairs which are coding for more than 20.000 proteins. Major events such as differentiation, metabolism, communication and cell division are the results of the regulation of genetic information in order to enable cells to perform target-driven functions and tasks. Thus, each cell type containing the identical genomic information, require different proteins for their function. For the proper regulation of the expression of those, proteins called transcription factors play the most important role on the molecular level. In general, protein expression starts with the transcription of a certain DNA sequences into RNA[1]. The process of transcription can be divided in three steps: initiation, elongation and termination. During the initiation the enzyme RNA polymerase II binds direct or via transcription factors to specific promoter regions on the DNA and unwinds it to the single-stranded state[2]. At a transcription start site a part of the DNA sequence, called template strand, is transcribed to mRNA. Transcription initiation is regulated by transcription factors, acting either as activators or repressors, which are sometimes associated with coactivators or corepressors[2]. During the elongation the RNA is synthesized according to the DNA sequence leading to increasing mRNA molecules[2]. In the last step of transcription, the termination, a polyadenylation occurs in order to label the end of the mRNA[2,3]. Not much is known about the interplay between DNA organizing proteins and transcription factors, which enable the recruitment of the RNA polymerase to the promoter site of a gene for RNA synthesis. However, studies in prokaryotes have shown, that if promoter regions are occupied by other proteins bound to DNA, the transcription is disturbed due to the prevention of transcription factor binding or DNA separation[4-6]. In this thesis, the term transcription factor is used to refer to transcriptional activators and repressors that affect the transcription of target genes via specific promotors regions. Those proteins are defined to contain at least one DNA-binding domain which is able to bind specific DNA sequences, thereby affecting the transcriptional activity of RNA-polymerases and thus the transcription and translation of proteins[7]. In the human genome approximately 2600 proteins contain one or more DNA-binding domain assuming that those function as transcription factors[8]. Thus more than 10% of all genes in the human genome code for transcription factors, which

makes it the largest protein family in humans. In order to regulate protein expression transcription factors stabilize or block the binding of the RNA-polymerase to DNA[2], recruit coactivators or corepressors to the protein-DNA complex[9] and catalyze acetylation or deacetylation of histone proteins[10]. Such signaling pathways are the driving force behind many biochemical and molecular networks and thus a key players in development and diseases. The understanding of these complex regulatory elements just emerged in the last decades. Despite various biochemical and molecular biological studies, various key mechanisms still remain elusive.

1.2 FOX transcription factors

Forkhead-Box (FOX) proteins comprise a well conserved class of transcription factor proteins in eukaryotes. They are highly diverse in terms of function, ranging from proliferation, cell growth, differentiation and longevity to embryonal development and homeostasis. Many genes encoding for FOX proteins have been identified and were grouped and classified into 19 subclasses (FOXA-FOXS) based on sequence conservation[11]. FOXO-proteins regulate for example metabolism, cell proliferation, stress tolerance, cancer and possibly lifespan[12-14]. FOXG proteins play a role in brain and telencephalon development, mutations are linked to microcephaly and brain malformations[15-17]. FOXM gene is known as a human proto-oncogene, the protein encoded by this gene is involved in cell cycle progression and cell proliferation[18]. Abnormal upregulation of FOXM1 is involved in oncogenesis of basal cell carcinoma[19] and many other cancers such as liver, breast, lung, colon and brain. FOXP2 protein is crucial for language development in human and animals and misregulation are linked to cancer[20-25].

Until now, 43 FOX proteins with specific functions have been discovered, however, all of them harbor a specific motif which is unique for this class of transcription factors: the forkhead box domain (later on called FH domain), a sequence of 80 to 100 amino acids forming a folded domain that directly binds DNA in a sequence-specific manner.

1.3 Function of FOXP2

The family of the FOXP proteins belong to an evolutionary important family of transcription factors. The importance of the FH domain has already been described for FOXP2 as a point mutation results in an inherited speech disorder (see chapter 1.4). As transcription factor, FOXP2 regulates hundreds of genes in various embryonic, postnatal and adult tissues, where it plays a role during embryonal development[26-28]. FOXP2 seems to be mainly active in embryonal development, similar to other FOX-proteins. It has been shown, that FOXP2 regulates genesis of some progenitors and neurons in the mammalian cortex, which is known to be a key center for human speech[29]. Besides, it has been found to be expressed in various brain areas such as cortical plate, basal ganglia, thalamus, inferior olives and cerebellum and was shown to be important for the development of brain areas involved in motor control[27]. Next to its function in brain development, it has also been shown, that FOXP2 plays a role in skull shaping and bone remodeling, as well as regulating the strength and length of hind limbs and maintenance of joint cartilage[30]. These recent results lead to the conclusion, that FOXP2 might have played a role in the evolution of the human bipedal locomotion[30]. One major functional aspect of FOXP2 activity derives from its heterodimerization with its paralogs FOXP1 and FOXP4. A study has shown that different combinations of FOXP1/2/4 dimerization severely affect gene expression[31]. This property may have oncogenic consequences.[24] FOXP1 and FOXP2 are widely co-expressed in specific environments[32] and interplay during the embryonal development of lung and esophagus[33].

1.4 Misfunction of FOXP2

The elucidation of its role in human speech development began with the study of a multigenerational British family known as the KE family. Half of the members suffered from developmental verbal dyspraxia, an inherited speech disorder. Affected members had difficulties in sequencing mouth movements resulting in an inappropriate ability of articulation. Beside oral language deficits, also written language was impaired. Members with this speech disorder showed difficulties in understanding complex language tasks such as grammar, including both expressive and receptive skills[34]. This disorder is caused by a missense mutation localized in the FH domain of FOXP2, which

is the only mutation that differ between unaffected and affected members. Hereby, an arginine at residue position 553 is replaced by a histidine (R553H) (Fig.1.1).

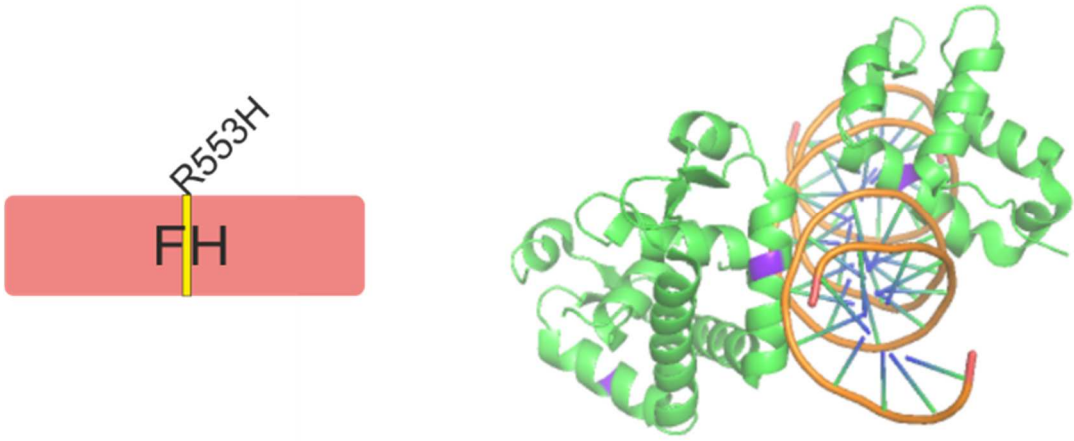

Fig.1.1: left: schematic representation of location of the point mutation R553H causing the speech disorder. Right: X-ray 3D structure of a monomer and dimer FOXP2 FH (green) bound to DNA molecules (orange) (PDB code: 2A07). Point mutation is indicated as purple.

This arginine residue seems to be crucial for the DNA-binding properties of this domain as it is located close to the DNA binding region (Fig.1.1) and, additionally, responsible for the nuclear localization of FOXP2[35]. As a consequence this missense mutation leads to a non-functional FOXP2[36]. Additionally, this point mutation causes severe changes in the brain organization, especially those brain areas involved in speech. Via Magnetic Resonance Imaging (MRI) scientist found different brain regions active in patients with the R553H mutation compared to healthy individuals[37]. Apart from its popular role in speech disorders, it has been shown, that FOXP2 is misregulated in many cancers. In hepatocellular carcinoma tissue of patients, FOXP2 was significantly reduced compared to non-tumorous tissue and linked to poor survival[38]. Also in breast cancer tissue, expression of FOXP2 was lower than in healthy breast tissue and was associated to decreased survival. It was shown, that FOXP2 overexpression inhibits breast cancer cell migration[25], whereas FOXP2 silencing mediates breast cancer metastasis[39]. Thus FOXP2 is a novel suppressor of breast cancer metastasis[25]. Beside those examples showing the suppressive behavior of FOXP2 on cancer progression, a few other cancer types has been reported, where overexpression of FOXP2 is linked to increased cancer development[38,40-42]. Also B-cell lymphoma patients showed poorer survival, if FOXP2 expression was high[43], highlighting the complexity of FOXP2 function in cell development. Also several other FOX family members are directly involved in cancer initiation, maintenance and progression in adults[18,40,44].

1.5 Evolutionary biology of FOXP2

Via its misfunction as result of certain mutations, *FOXP2* has been the first gene found to be related to human speech. Human speech is an individual attribute to humans, its complex pattern is unique among other animals. Interestingly, the protein coding sequence of FOXP2 is among the 5% most-conserved proteins in vertebrates. This strong conservation and the strong expression in various, especially fetal, brain regions lead to the idea of a strong role in development.

Between our closest ancestor, the chimpanzee and the humans, FOXP2 differ only within two amino acids (substitutions N303T and S325N) (Fig. 1.2). FOXP2 has the same sequence in chimpanzee, gorilla and rhesus macaque. Between the mouse and human FOXP2 there are only 3 amino acids which differ (E80D, N303T and S325N) (Fig. 1.2)[45]. Thereby the amino acid substitution at position 325 specific for human creates a potential phosphorylation site and shows a different secondary structure propensity than the sequence of chimpanzee[45], which might lead to a different regulation of FOXP2 by posttranslational modifications or interaction with co-factors[45]. Interestingly, the one substitution which differs between mouse and chimpanzee (E80D) must have occurred during the roughly 130 million years ago (myr) of evolution that separated the ancestor of humans and chimpanzee from the mouse. By contrast, the two amino acids, which differ between chimpanzee and human, must have occurred in the relatively short evolution time of 4.6-6.2 myr, when the human lineage diverged from the chimpanzees. The ratio of these changes is significant compared to all other lineages harboring the FOXP2 protein. However, the hypothesis, that these changes might be a proof of recent positive selection in human evolution, was disproven in a recent study[46].

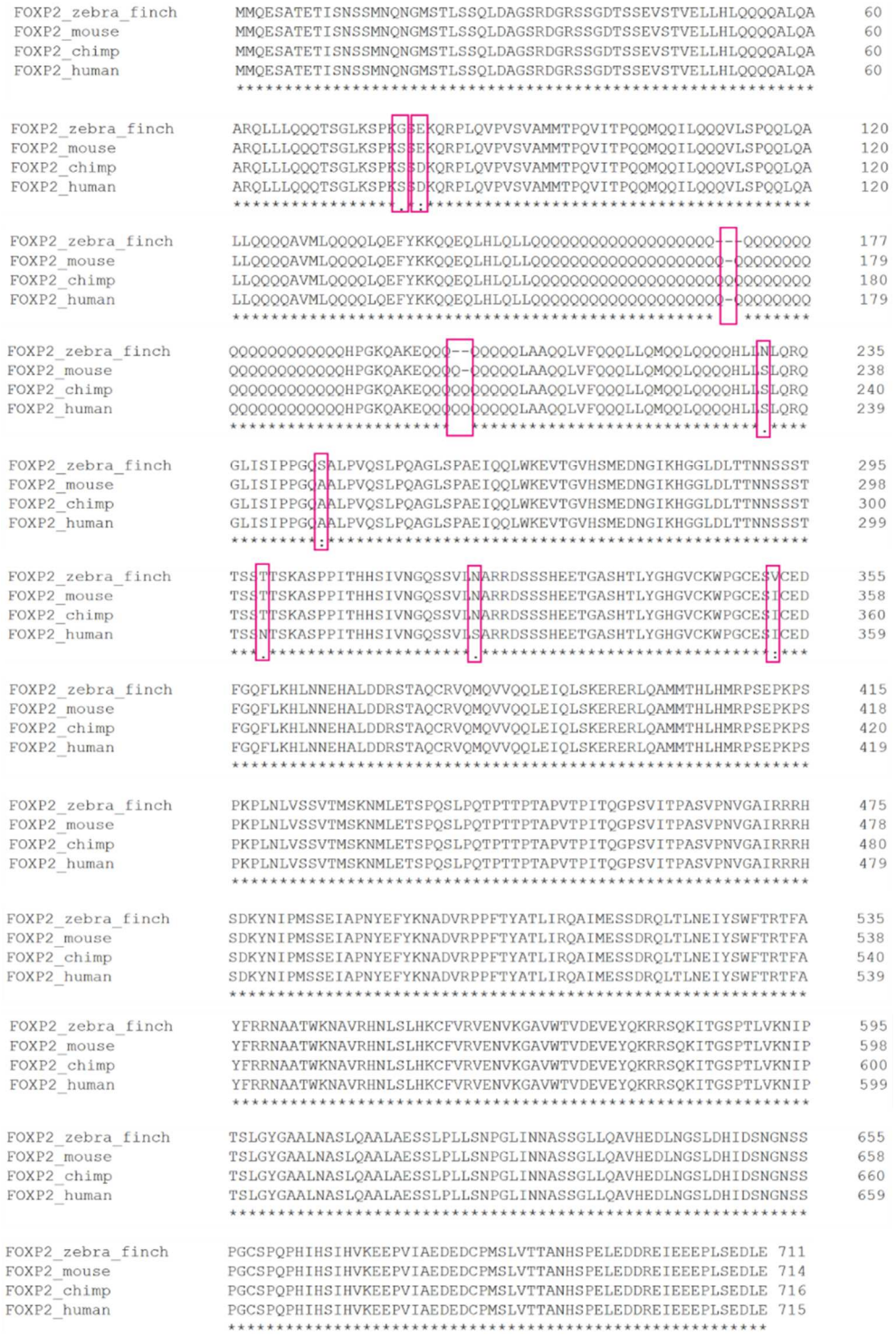

Fig.1.2: protein sequence alignment of FOXP2 in human, chimp, mouse and zebra finch (https://www.ebi.ac.uk/Tools/msa/clustalo). Differences between species are indicated with pink boxes.

1.6 Animal studies on FOXP2

As FOXP2 is involved in speech development of humans, the question raised whether FOXP2 is also important for vocalization in other species. The conservation of the protein among different species makes it an important evolutionary factor. The human ortholog of FOXP2 differs from zebra finch, mouse and chimpanzee orthologs only in seven, three or two amino acids, respectively[45] (Fig. 1.2). This fact, as well as the fact that only humans developed a complex speech pattern, led to studies with various animal models. In order to find an appropriate animal model, the learning of vocalization must be similar to humans. E.g. zebra finches imitate their parents or tutors to learn the species-specific song pattern, which makes them an interesting model for human speech development, as humans learn languages by imitating their educators. FOXP2 expression levels in the brain of zebra finches are fluctuating depending on the period of singing and the age. Zebra finches with FOXP2 knockout[47] or overexpression[48] in the striatal song nucleus Area X showed decreased spine density and impaired song learning and -production. This lacking imitation ability indicates that behavior-linked regulation of FOXP2 is more critical for vocalization than absolute FOXP2 levels[47-49]. Mice are an optimal model to study the function of proteins in mammals. To study FOXP2 function in mice, Shu *et al.* disrupted one or two copies of the FOXP2 gene. Those knockout mice showed significantly reduced number of ultrasonic vocalizations compared to wildtype mice, suffered from severe motor impairments and died within 3 weeks after birth[50]. Mice carrying the mutation found in the affected members of the KE family developed severe motor abnormalities, severe impairment of ultrasonic vocalization [51], cerebellar abnormalities and deficits in motor-skill learning[52]. Additionally, researchers were interested in the learning abilities of mice with the human FOXP2. Schreiweis *et al.* designed a humanized mouse model, thus harboring the human FOXP2, which showed not only differences in striatal neuroplasticity but also accelerated learning leading to the conclusion, that the human FOXP2 evolution led to differential tuning of corticostriatal systems involved in learning processes and thus contributed to adapting the human brain for speech acquisition[53].

1.7 Structural organization of FOXP2

The protein FOXP2 belongs to the FOXP-family, which consists of four members, FOXP1, FOXP2, FOXP3 and FOXP4. All members harbor a zinc-finger, leucine zipper and the characteristic Forkhead-domain (Fig.1.3). FOXP1, 2 and 4 share many similarities compared to FOXP3, which lacks the poly-Glutamine tract (further on called poly-Q).

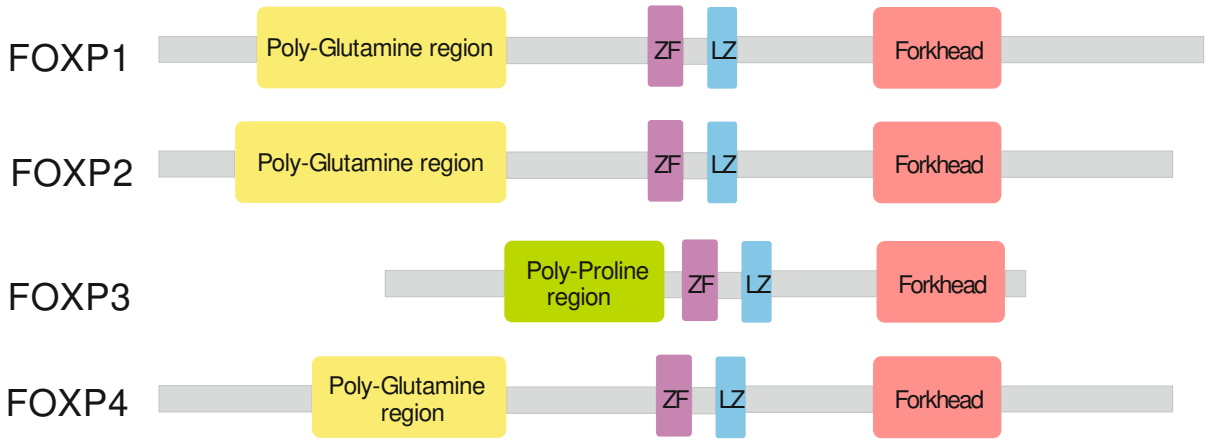

Fig.1.3: schematic representation of the structural organization of FOXP proteins. Domains are colored in different colors, IDRs are colored in grey.

The Forkhead-domain, however, remains conserved in all members, keeping one feature for the FOXP-family. FOXP2 is expressed in various organs including lung, gut, muscle and liver where it is crucial for the proper embryonal development.

Looking at the domain organization, FOXP2 consists of 715 amino acids forming four structured domains which are linked with intrinsically disordered regions (IDRs) (Fig.1.4).

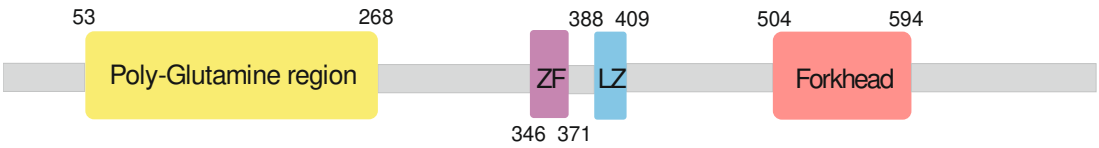

Fig.1.4: schematic representation of the structural organization of FOXP2. Domains are colored in different colors, IDR are colored in grey.

N-terminally it contains a poly-Q, which is the longest of its kind in any human protein in healthy individuals[54]. It stretches from residue 53 to 268 and contains a region with 40 glutamines in a row. Poly-Q proteins are known to be involved in nuclear localization and transcriptional regulation. The most known proteins, which contain poly-Q tracts are associated to the development of neurodegenerative disease such as Alzheimer's[55], Parkinson's[56] or Huntington's disease (HD)[57]. FOXP2 has been shown to be linked to HD. A study discovered that knockdown of FOXP2 in healthy mice mimicked HD associated behavioral deficits, whereas an overexpression in a mouse model with HD associated deficits improved their behavioral phenotype[58].

The second domain forms a zinc-finger from residue 346-371. Zinc fingers are usually small and the most common protein motif in the proteome of mammals[59-61]. Almost half of all human transcription factors contain this domain. It functions as promoter of protein-DNA, protein-RNA or protein-protein interactions[60,62]. The tertiary structure of zinc fingers is normally stabilized by a zinc ion, but not interfering in the interaction with binding partners[63]. As the affinity from a zinc finger to DNA is normally in the higher µM-range[64], proteins contain usually a few zinc fingers in a row[65-67]. However, FOXP2 contains only one zinc finger domain. Except the primary sequence, not much is known about the zinc finger of FOXP2. There is no evidence that this zinc finger is involved in DNA or protein binding. Using yeast-two-hybrid system Li *et al.* showed, that removing of the zinc finger do not have significant impact on the transcriptional activity of FOXP2[68]. However, for the subfamily member FOXP1, it has been shown, that the zinc finger is involved in homo- and heterodimerization[69].

The third domain is a leucine zipper. They are able to form coiled coil motifs with hydrophobic parts and thereby promote homo- and heterotypical protein-protein interactions[70,71]. Homodimers bind mostly to palindromic DNA sequences whereas the heterodimers are able to bind any combination of DNA sequences[72-74]. About the function of the leucine zipper in FOXP2, not much is known. In the previous mentioned study Li *et al.* showed that the leucine zipper is essential for FOXP2 activity. They also hypothesized, that FOXP2 must dimerize for successful DNA binding[68]. The FOXP proteins are conserved and it has been shown, that co-expressed FOXP1, FOXP2 and FOXP4 interact with themselves and each other. Studies have indicated, that the leucine zipper is necessary for those interactions and thus important for the function of FOXP2[68,75,76].

The fourth domain is the forkhead domain (FH), which acts as the main DNA-binding domain of FOXP2 and is highly conserved in the FOX family. It is the most studied domain of FOXP2 due to its importance in its function. In 2006 Stroud *et al.* published the X-ray structure of the FH bound to DNA[77], thereby six FH domains bind as two dimers and two monomers to two DNA molecules (Fig.1.5). The two dimers form a so called swapping dimer, a known feature in protein complexes.

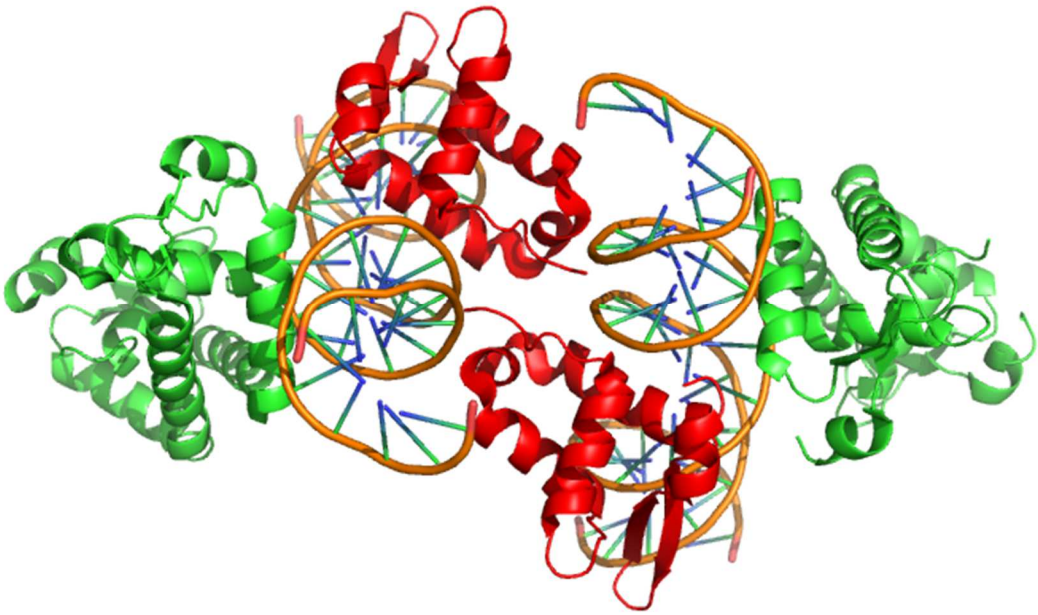

Fig.1.5: X-ray 3D structure of FOXP2 FH bound to DNA molecules (PDB code: 2A07). Dimers are colored in green, monomer in red, DNA in orange.

It has been hypothesized that through this feature FOXP2 is able to bring two remote DNA segments close together[77]. Beside this, studies with *S. cerevisiae* have shown that domain swapping can also regulate the time point of DNA-replication[78], thus this feature might regulate FOXP2 activity. The dimerization of FOXP proteins is known to be unique among FOX proteins, as other FOX transcription factors are known to be monomeric upon DNA binding. Compared to other subclasses, the forkhead domain of FOXP2 is located at the C-terminus instead of the N-terminus.

Another structural feature of FOXP2 is the presence of two nuclear localization signals (NLS) located in the FH domain[35]. If one of both NLS is disrupted, the nuclear localization is slightly affected but remaining. Only disruption of both NLS lead to the abrogation of FOXP2 nucleo-cytoplasmic shuttling[35]. The R553H mutation, which is

leading to the speech disorder, disrupts the nuclear localization leading to increased levels of FOXP2 in the cytoplasm[35], the molecular mechanism behind this phenomenon remains unclear. In combination with wildtype FOXP2 the mutant version can still be imported in the nucleus via heterodimerization with the wildtype protein leading to increased levels in the nucleus[35].

1.8 Regulation of FOXP2

As it is important for various biological processes, FOXP2 must be tightly regulated. So far not much is known about the regulatory processes of FOXP2 and its transcriptional activity. A few interaction partners are known such as Protein inhibitor of activated STAT (PIAS)[75], C-terminal binding protein 1 (CTBP1)[79] and Nuclear factor of activated T-cells, cytoplasmic 2 (NFATC2)[80].
Additionally, FOXP2 undergoes post-translational modifications such as SUMOylation (Small Ubiquitin like modifier)[81]. A previous study has shown, that this is a crucial regulatory mechanism of FOXP2 activity[75]. Posttranslational modifications (PTMs) are important regulators of transcription factors. Thereby functional groups are added to certain amino acids to diversify and extend protein function beyond what is dictated by gene transcripts[82]. They reversibly or irreversibly alter the structure and properties of proteins through biochemical reactions, leading to diverse functions[83]. A variety of PTMs allow eukaryotic cells to dynamically regulate signaling and physiological processes. As analytical methods have improved, the biological influences of many types of PTMs have been identified and are characterized in many systems. Besides alternative splicing, they provide the proteome with an enormous capacity for biological diversity and regulate a plethora of processes including cell growth and differentiation, programmed cell death, intracellular transport and cell–cell communication between the intracellular and extracellular environment[84]. In particular PTMs are affecting protein-protein interactions, protein-DNA interactions, gene expression and signal transduction[84]. Common types of eukaryotic PTMs that belong to the first class of protein modifications are phosphorylation, acetylation, alkylation and glycosylation. These PTMs typically occur in ‘regulatory’ protein regions that are intrinsically disordered, but also loop regions of folded domains. Intrinsically disordered regions (IDRs) are therefore a common target for enzymatic modifications as fast cellular signaling responses usually require modifying enzymes to rapidly access individual

protein PTM sites, which is easier, when modifiable amino acids are solvent exposed such as IDRs[85-87]. Phosphorylation of serine and threonine residues constitutes the most abundant PTM in eukaryotes[88,89].
Disruptions of established PTMs or misregulation of protein kinases are often linked to cancers, cardiovascular, brain diseases, diabetes and several metabolic disorders. Thus, there have been many studies investigating the mechanism how PTMs regulate different cellular signaling processes, in both preventive and curative sense.

For members of the FOXP-family various posttranslational modifications have been found, but the effect of those is only known for a few phosphorylations. A phosphorylation of a tyrosine in the forkhead domain of FOXP3 is linked to inhibited carcinogenesis and transcription, another phosphorylation of S418 is linked to inhibited cell growth and induced transcription (https://www.phosphosite.org).
Phosphorylations, the attachment of a phosphoryl group, occur on serine, threonine and tyrosine side chains through a phosphoester bond formation, on histidine, lysine and arginine through phosphoamidate bonds, and on aspartic acid and glutamic acid through mixed anhydride linkages. This modification forms the most studied PTM in eukaryotes, as misregulations of kinases often result in diseases, especially in cancer[90]. Phosphorylations are commonly mediated by enzymes called protein-kinases, which are also regulated by phosphorylation, thereby forming a dependent regulation network. The removal of a phosphate group from a residue is mediated by enzymes called phosphatases, making phosphorylations a dynamic regulatory tool[91].
For FOXP2 few PTMs are already known, including six phosphorylated residues (Fig.1.6).Only for one FOXP2 phosphorylation the function and effects are known. The phosphorylation of the FH domain at position S557, which lies nearby to the mutation causing the speech disorder (R553H), was shown to decrease the DNA binding affinity of the FH and thus might be involved in the regulation of FOXP2 transcriptional activity[92].

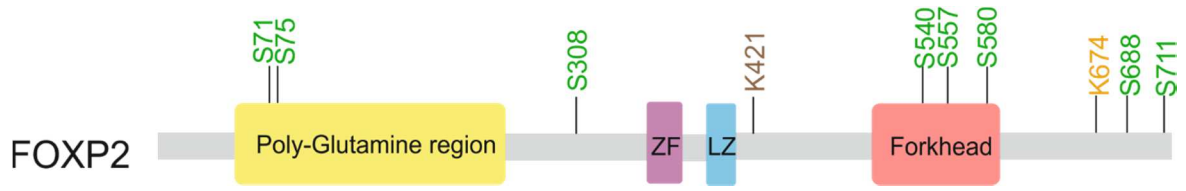

Fig.1.6. known post translational modifications of FOXP2. Phosphorylations (green), Ubiquitination (brown), Sumoylation (orange) (derived from https://www.phosphosite.org/).

To better understand the regulatory mechanisms which control FOXP2 activity we aimed to find clues about pathways, which are regulated by FOXP2 to get ideas about regulatory elements and processes in cells. Recently, other FOXP proteins have been found to be linked to the Wnt signaling pathway, which raises the question, whether and how FOXP proteins are regulated by the Wnt pathway and thus gives new insight into human development and mechanism of various diseases like cancer linked to both protein families.

Using Mass-Spectrometry, Walker *et al.* showed that FOXP1 enhances Wnt signaling, an important signal transduction pathway in early embryo development of metazoans, by forming a co-complex with β-catenin, Transcription factor 7-like 2 (TCF7L2) and CREB (adenosine 3',5'-monophosphate response element-binding protein)-binding protein (CBP)[93]. Also FOXP3 misregulation is linked to the Wnt pathway activation in lung cancer by promoting tumor growth and metastasis and forms a complex with β-catenin, as shown by Co-Immunoprecipitation (Co-IP)[94]. Wnt signaling operates in both vertebrates and invertebrates. The Wnt signaling pathways are central regulatory elements in a remarkably diverse range of functions during the embryonic development and adult homeostasis controlling cell fate specification[95], cell proliferation[96,97] and cell migration[98]. Disruptions in this highly-conserved signaling pathway result in various diseases including cancer and neurodegenerative diseases[99-101].

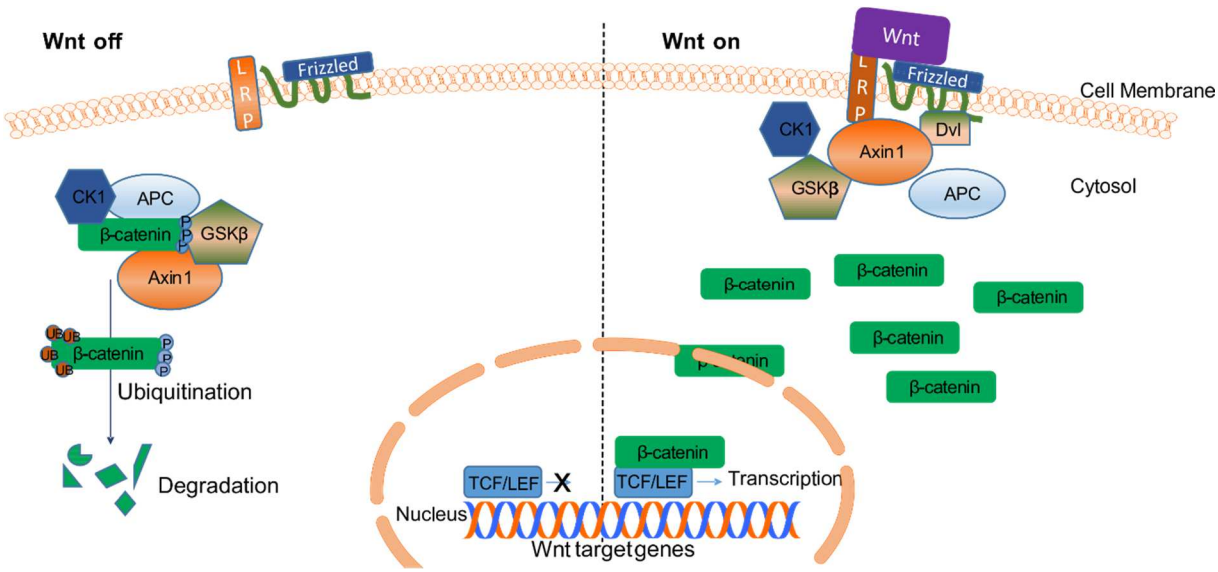

Fig.1.7: schematic representation of the Wnt-signaling pathway.

β-catenin is a crucial player in this pathway acting as transcriptional co-activator for many transcription factors. In the inactive state of the Wnt signaling pathway, a protein

complex called the β-catenin destruction complex degrades β-catenin continuously in the cytoplasm, preventing high β-catenin levels and thus its translocation to the nucleus. In the active state the destruction complex is falling apart leading to higher cytosolic levels of β-catenin, following its nuclear translocation and acting as activator for various transcription factors including the TCF/LEF[102-104] pathway, but also FOX proteins (Fig.1.7).

Fig.1.8: schematic representation of the structural organization of β-catenin.

Until now, it remains elusive whether or not β-catenin interacts, apart from FOXP1 and FOXP3, also with FOXP2. Therefore, in this thesis, I investigated a possible link between FOXP2 and the Wnt pathway, as both are important in embryonal development. For this aim I focused on the interaction with β-catenin as transcriptional co-activator (Fig.1.8), which is regulating, when active, various transcription factors such as TCF/LEF[102-104], Hypoxia Induced Factor 1 α under hypoxic conditions[105] or FOXO proteins under oxidative stress[106].
These studies arise the question, whether and how FOXP proteins are regulated by the Wnt pathway. Until now, it remains elusive whether or not β-catenin directly interact with FOXP proteins and what are the molecular mechanisms allowing regulation of FOXP proteins activity and functions.

1.9 Aim of this thesis

In this thesis I describe my PhD project, which was the main project during my doctoral studies. For this and all other side projects I used structural biology methods, such as solution NMR spectroscopy, Small angle x-ray scattering or isothermal titration calorimetry in order to answer specific research questions. All projects intended to decipher unprecedented mechanisms in signal transduction, and metabolism research.

FOXP2 is a transcription factor with a huge impact on embryonal development, cancer and speech development in humans. It is regulating more than 1000 genes. Due to its biological role in animals it must be tightly regulated to ensure the proper transcription

of genes in certain stages of development. So far, its regulation and network is still elusive.

By combining Nuclear magnetic resonance (NMR) spectroscopy and Small-angle X-ray (SAXS) methods I studied the structural functions of FOXP2 *in vitro* to obtain more information about its function and properties. Next, I was searching for novel binding partners interacting with FOXP2 using biophysical methods such as NMR, ITC and SAXS, which could regulate FOXP2 activity. To confirm regulatory elements I was using cell-based assays such as Co-IP and RNA-Seq to validate the impact of those binding-partners on the activity of FOXP2 and thus get more information about its function. Beside this I was seeking for novel post translational modifications, which could as well play a role in the regulation of FOXP2 using NMR. With all these studies I aimed to elucidate the, so far, unknown regulation of FOXP2 in human embryonal development and its role in diseases such as cancer, autism and schizophrenia.

Chapter 2

2 Material and Methods

2.1 Materials

2.1.1 Buffers and solutions

Kanamycin (50 mg/ml)

5 g Kanamycin (Applichem Panreac) in 100 ml of MilliQ H_2O. Filter sterilized.

LB growth medium for *E.coli*

2 % (w/v) lysogeny broth (Roth) (+ 0.1 % (v/v) Kanamycin)

10 x salt solution

1 M KH_2PO_4 (VWR 99.8%)

0.5 M K_2HPO_4 (VWR 99.8%)

0.6 M Na_2HPO_4 (Applichem Panreac anhydrous >99%)

0.14 M K_2SO_4 (VWR 99-101%)

pH 7.2 adjusted with HCl (12.2 M) and NaOH (5 M)

Trace element solution

41 mM $CaCl_2$ x 2 H_2O (Applichem Panreac 97%)

22 mM $FeSO_4$ x 7 H_2O (VWR 84%)

6 mM $MnCl_2$ x 4 H_2O (Applichem Panreac)

3 mM $CoCl_2$ x 6 H_2O (Applichem Panreac)

1 mM $ZnSO_4$ x 7 H_2O (VWR 99.9%)

0.1 mM $CuCl_2$ x 2 H_2O (VWR)

0.2 mM $(NH_4)_6Mo_7O_{24}$ x 4 H_2O (VWR 98.5%)

17 mM EDTA (Alfa Aesar 99+%)

Minimal medium growth medium for isotope labeled or non-labeled proteins

900 ml MilliQ H_2O

100 ml 10 x salt solution

5 ml Trace element solution

5 ml 1 M $MgCl_2$ (Alfa Aesar 99%)

1g $^{15}NH_4Cl$ (Sigma Aldrich 98%)/2 g $^{14}NH_4Cl$ (Alfa Aesar 99.5%)

2g $^{13}C_6H_{12}O_6$ (Cambridge Isotope Laboratories 99%)/ 6 g $C_6H_{12}O_6$ (Roth 99%)

0.1% (w/w) Kanamycin (1000x)

Lysis buffer (lysis buffer for structured proteins)

50 mM Tris (VWR ultra pure)

150 mM NaCl (VWR 99.6%)

20 % glycerol (VWR 99%)

20 mM imidazole (Millipore 99.0%)

2 mM BME (Roth 98%)

pH 7.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Urea lysis buffer (lysis buffer for disordered proteins)

50 mM Tris (VWR ultra pure)

150 mM NaCl (VWR 99.6%)

6 M Urea (Roth >99.5%)

20 M imidazole (Millipore 99.0%)

2 mM BME (Roth 98%)

pH 7.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Washing Buffer

50 mM Tris (VWR ultra pure)

150 mM NaCl (VWR 99.6%)

20 mM imidazole (Millipore 99.0%)

2 mM BME (Roth 98%)

pH 7.5 adjusted with HCl (12.2 M) and NaOH (5 M)

High-salt Washing Buffer

50 mM Tris (VWR ultra pure)

1 M NaCl (VWR 99.6%)

20 mM imidazole (Millipore 99.0%)

2 mM BME (Roth 98%)

pH 7.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Elution Buffer

50 mM Tris (VWR ultra pure)

150 mM NaCl (VWR 99.6%)

500 mM imidazole (Millipore 99.0%)

2 mM BME (Roth 98%)

pH 7.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Phosphate buffer for NMR, ITC and SAXS measurements

50 mM Na_2HPO_4/NaH_2PO_4 (VWR ultra pure)

150 mM NaCl (VWR 99.6%)

2 mM BME (Roth 98%)

pH 6.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Tris buffer for NMR, ITC and SAXS measurements

50 mM Tris (VWR ultra pure)

150 mM NaCl (VWR 99.6%)

2 mM TCEP (Roth 98%)

pH 7.5 adjusted with HCl (12.2 M) and NaOH (5 M)

HEPES buffer for ITC and EMSA experiments

10mM HEPES

100mM KCl

2mM BME

pH 6.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Tris buffer for CD experiments

20mM Tris

100mM NaF

0-50% of TFE

pH 7.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Dye 10X buffer for EMSA experiments:

10mM Tris

1mM EDTA

50% v/v glycerol

0.001% w/v bromphenol blue

HEPES 10X binding buffer for EMSA experiments:

100mM HEPES

10mM EDTA

1M KCl

1mM BME

50% v/v glycerol

pH 6.5 adjusted with HCl (12.2 M) and NaOH (5 M)

Tris Acetate-EDTA 10X buffer for running gel-electrophoresis

400mM Tris

200mM Acetic acid glacial

10mM EDTA

pH 6.5 adjusted with HCl (12.2 M) and NaOH (5 M)

2.1.2 Instruments

FPLC

Äkta pure (GE Healthcare Life science)

NMR spectrometer

700 MHz Avance III NMR spectrometer (Bruker Biospin) equipped with a TCI cryoprobe and 600 MHz Avance Neo NMR spectrometer (Bruker Biospin) equipped with a TXI 600S3 probehead. Software for processing: NMRpipe (IBBR) and Topspin (Version 4.0.3), Software for analysing: CCPNMR Assignment software (version 3.0.b1).

SAXS

SAXSpace (Anton Paar Gmbh), equipped with 1D Mythen Detector, slit collimation, cooled Autosampler (ASX), Analysis with the Software SAXSAnalysis (version 3.00.044).

ITC

Microcal VP-ITC (Malvern), Analysis with Software Origin (MicroCal, version 7.0).

RNA-Seq

NextSeq 500 (Illumina), Analysis with Software DESeq2[107]

q-PCR

7900HT Fast-Real-Time PCR System (Applied Biosystems, Singapore)

EMSA

Typhoon 9400, Variable Mode Imager (GE Healtcare Life Science)

2.2 Methods:

2.2.1. Molecular biological methods:

2.2.1.1 Plasmid preparation

Plasmids were obtained from Genscript (Piscataway, New Jersey, USA) in desired vector. For bacterial expression petM11 vector containing a kanamycin resistance gene for positive selection, a z-tag containing a Poly-histidine-tag with TEV-cleavage site were used. For mammalian expression the full-length FOXP2 sequence in a pCMV-3Tag-1a vector with a SV40 and CMV promoter and kanamycin resistance was kindly provided from Prof. Dr. Wolfgang Enard from the Ludwig Maximilians-University in Munich. Plasmids were amplified using competent *E.coli* cells (Top10 from XX) cultured in LB-Medium. DNA was then isolated and purified using the Wizard® *Plus* SV Minipreps DNA Purification System (Promega).

2.2.1.2 Cloning and mutagenesis

Mutations were generated by site-directed mutagenesis using Q5 High-Fidelity Polymerase (New England Biolabs).

For the 50µl PCR the reagents were mixed as follows:

10 µl Q5 High Fidelity DNA Polymerase Buffer (5x)

1 µl dNTPs (10 mM)

2.5 µl Forward Primer (10 µM)

2.5 µl Reverse Primer (10 µM)

0.25 µl Q5 High Fidelity DNA Polymerase (2 U/µl)

1 µl DNA Template (20 ng/µl)

32.75 µl Nuclease-free water

Cycling conditions:

Step 1: Initial Denaturation (98°C) 2 minutes

Step 2: Denaturation (98°C) 30 seconds

Step 3: Annealing (primer-dependent) 30 seconds

Step 4: Extension (72°C) 4.5 minutes

Go to Step 2 and repeat 25 times

Step 5: Final Extension (72°C)

Step 6: 2 minutes Hold (10°C)

Step 7: ∞

Following primers were used:

Construct	Primer sequence 5' to 3'	Optimized for	vector
FOXP2^{FH}	Fw1: CGTCCGCCGTTCACCTAC Rv1: GGATCCTTATTCCAGGTCTTCTG Fw2: CCCGACGCTGtaaAAAAACATTCC Rv2: GAGCCGGTAATTTTCTGAC	*E.coli* cells	pETM11
FOXP2^{R553H}	FW: AAACGCTGTCCACCATAATCTGAG RV: TTCCACGTAGCTGCATTAC	*E.coli* cells	pETM11
FOXP2$^{503\ STOP}$	FW: CGCTGTGTGTAGCCGCCGTTCA RV: TTTTTGTAGAATTCGTAGTTC	*E.coli* cells	pETM11
FOXP2^{S330E}	FW: CCGTCGCGATGAAAGTAGTCATGAG RV: GCGCTCAGTACGGAGAGGAT	*E.coli* cells	pETM11
FOXP2$^{\Delta polyQ}$	FV: GGTCAGGCCGCACTGCCG RV: ATGCAGCAGTTCGACCGTAGAAACTT	*E.coli* cells	pETM11
FOXP2$^{247-715}$	FwW: GGCCAGGCAGCACTTCCT Rv: TTAGGTTTCACAAGTCTCGAGTCATT	Mammalian cells	pCMV 3tag
FOXP2$^{345-715}$	FW: TCTCTATGGCCATGGAGTTT RV: AAGCCGAATTCCACCACA	Mammalian cells	pCMV 3tag
FOXP2$^{504-715}$	FW: TCAGACCTCCATTTACTTATGC RV: AGCCGAATTCCACCACAC	Mammalian cells	pCMV 3Tag
FOXP2 $^{\Delta 264-272}$	FW: GTGACTGGAGTTCACAGTATG RV: AGGACTTAAGCCAGCTTG	Mammalian cells	pCMV 3Tag

Tab.1: primers used for side-directed mutagenesis

PCR efficiency was confirmed by running a 1% Agarose DNA gel with a part of the PCR product. 1 µl of DpnI restriction enzyme was added into the PCR tube and mixed by pipetting. After incubation at 37°C for one hour, 2 µl of the PCR reaction were mixed with 15 µl of Nuclease-free water and heated for 20 minutes at 80°C. 2 µl of 10X T4 Ligase buffer (NEB) and 1 µl of T4 polynucleotide Kinase (PNK-NEB) were added to the chilled tube and incubated for 30 minutes at 37°C. After a ligation step of 2 hours at room temperature using 1 µl of T4 Ligase (NEB), the plasmids were transformed into competent high copy *E.coli* cells (Top10) and platted on agarose plates containing Kanamycin. Only *E.coli* cells containing the plasmid are able to grow due to antibiotic resistance in plasmid of interest. Colonies were picked from agarose plate after incubation overnight at 37°C and DNA was isolated, purified and sent for sequencing to GATC biotech (Eurofins genomics).

2.2.1.3 Protein expression

100 ng of plasmid DNA was added to competent *E.coli* DE3 cells. After an incubation step of 10 min on ice, a heat shock of 42°C for 45 s was applied in an Eppendorf thermoblock to introduce the plasmid of interest in the competent cells. Cells were incubated on ice for another 5 min to recover from the heat shock. For antibiotic resistance development, cells were incubated for one hour in 1 ml of LB media without antibiotics at 37°C shaking in an Eppendorf thermoblock. After incubation, 400µl of the cell solution were plated on LB-agar plates containing kanamycin for selection and incubated overnight at 37°C. Colonies were picked next day and grown in 10 ml of LB media + kanamycin for selection at 37°C in an incubator overnight. The overnight culture of *E.coli* DE3 with the genetic information of the protein of interest was used to inoculate a main culture (1 L LB media or minimal medium (^{13}C or ^{12}C glucose and/or ^{15}N or ^{14}N ammonium chloride supplemented) + Kanamycin for selection). When grown to an optical density at 600 nm (OD_{600}) of approximately 0.8-1 at 37°C, the protein expression was induced with 1 mM IPTG at 20°C overnight. After expression of proteins, cells were harvested by centrifugation (15 min, 6000 rpm, 4°C).

2.2.1.4 Protein purification

Harvested *E.coli* cells were resuspended in 20ml lysis buffer after IPTG-inducted protein expression to isolate the recombinant expressed proteins. The pelleted cells were then flash-frozen using liquid nitrogen and thawn in a water bath. Cells were sonicated for 24 minutes (1s on/1s off; 70 % intensity) in ice-cold water bath and centrifuged for 30-45 minutes at 4 °C (12 000 rpm). The supernatant contains all the soluble components including the recombinantly expressed protein. In the first step proteins were purified using Ni-NTA (Ni-NTA Agarose as stationary phase, Thermo Fisher) gravity columns. The purification is based on binding of His-tagged proteins to Ni^{2+} residues in the column. The column was first equilibrated with 25 mL of washing buffer and the supernatant of lysed *E.coli* cells was applied on the column. The flow through was collected and applied a second time to obtain a higher yield. Non-specifically bound proteins were removed by washing the column with approximately 40 ml of washing buffer and non-specific bound DNA was removed by washing the column with 50 mL of high-salt washing buffer. The elution of non-specific proteins was monitored using NanoDrop™ (Peqlab). The protein of interest was eluted with 10-15 ml of elution buffer. To avoid a non-native conformation of the protein, the Z-tag was removed using TEV protease. The concentration of the eluted protein was determined using NanoDropTM (Peqlab) and 2 (w/w) % of TEV protease was added to the protein solution. The sample was incubated for at least 8 hours at 4 °C for cleavage. HiPrep 26/10 Desalting (50 ml, GE Healthcare) on an Äkta pure FPLC system (GE Healthcare) was used for buffer exchange. Column was equilibrated with 60 ml of washing buffer at a flow rate of 5 ml/min. The protein elution was applied on the column and eluted fractions were collected depending on the UV absorbance at 280 nm measured by Äkta system detector. In order to avoid Z-tag or TEV protease contamination, the protein was applied on a HisTrap (5 mL, GE Healthcare) on an Äkta pure FPLC system and the flow through containing the cleaved protein was collected.

I observed that the protein constructs $FOXP2^{FH}$ and $FOXP2^{R553H}$ still bind with cleaved Z-tag to the HisTrap column. The FOXP2 FH was then purified from Z-Tag and TEV protease using a HiTrap Heparin HP (5ml) column equilibrated with washing buffer and eluted using an increasing gradient of high-salt washing buffer (100 % high-salt buffer in 15 ml).

The FOXP2^{R553H} construct did not bind to the HiTrap Heparin HP (5ml) column, thus it was loaded on a HisTrap and eluted without Z-tag at a washing step with high-salt washing buffer before the His-tag was eluted with elution buffer.

For all experiments, proteins were purified with Size Exclusion Chromatography (SEC). High molecular weight constructs were purified using Superdex 200 10/300 Increase column, low molecular weight constructs were purified using Superdex 75 300/10 column in final experimental buffer.

To obtain higher protein concentrations, Amicon Ultra Centrifugal Filters (Millipore) with a cut off of 30 kDa, 10 kDa or 3 kDa were used. Centrifugation steps of 5 – 10 min at 4 °C at 3500 rpm were followed by mixing of the protein solution in the filter to avoid precipitation of the protein due to the concentration gradient. Concentrations of purified proteins were determined using NanoDropTM (Peqlab) and the corresponding extinction coefficient.

2.2.1.5 DNA preparation:

For DNA interaction studies we used a target sequence for FOXP2 published by a previous publication[108]. We obtained primers with following sequence from IDT-DNA: fw: 5'GCG CTC TTG TTT ACA GCT 3', rv: 5'AGC TGT AAA CAA GAG CGC 3'. The obtained DNA was dissolved in desired buffer at a concentration of 1 to 2mM concentration. Both primers were mixed at same concentrations and boiled at 95°C for 10 min and then cooled on ice for 30min to obtain annealed, double-stranded DNA. Concentration was then estimated using NanoDropTM (Peqlab) and the extinction coefficient. This step was confirmed using 1D ^{1}H NMR spectroscopy, as only double-strand DNA show up as signal at 13-15ppm.

2.2.2 Structural Biological methods

2.2.2.1 NMR experiments

Interaction experiments:

The most common NMR experiment for protein-interaction studies is the ^{1}H-^{15}N Heteronuclear single quantum coherence (HSQC) spectrum. The obtained signals of this experiment represent the H-N correlations of mainly the backbone amide groups, but for tryptophan, asparagine and glutamine side chains are also visible. The ^{1}H-^{15}N HSQC is regarded as the fingerprint of a protein and looks different depending on the chemical environment of the protein. Upon binding of unlabeled interaction partners (proteins, peptides, DNA, RNA, small molecules) the chemical environment changes, thus the signals differ to the reference spectrum of the labeled protein in absence of binding partner.

For those studies ^{13}C and/or ^{15}N isotopically labelled recombinant proteins were produced in *E.coli*. Samples for NMR measurements contained 100 µM ^{15}N labelled $FOXP2^{IDR}$, $FOXP2^{FH}$ or $FOXP2^{R553H}$ constructs in phosphate buffer with 10% D_2O added for the lock signal. ^{1}H-^{15}N HSQC NMR spectra were recorded at 298 K on a 600 MHz Bruker Avance Neo NMR spectrometer equipped with a TXI 600S3 probehead or on a 700 MHz Bruker Avance III NMR spectrometer equipped with a TCI cryoprobe. NMR spectra with full-length β-catenin were prepared in Tris buffer, as full-length β-catenin is only stable at a pH of more than 7. All spectra were recorded with a recycle delay of 1.0 s, spectral widths of 15.9/30 ppm, centered at 4.7/118.0 ppm in $^{1}H/^{15}N$, with 1,024 and 256 points, respectively, and 16 scans per increment.

Assignment experiments:

Assignment experiments are used to identify the different signals obtained from NMR experiments such as ^{1}H-^{15}N HSQC and link them to the residues of the studied protein. This can be achieved by running triple resonance experiments with NMR and a ^{15}N ^{13}C labeled protein sample. ^{1}H-^{15}N pairs are recognized as belonging to neighboring residues when corresponding frequencies match. In this way chains of sequentially connected residues can be built. This chain might be interrupted by missing NMR

signals or invisible residues such as prolines. Afterwards, those chains can then be mapped to the protein sequence. The backbone assignment is based on two main spectra: HNCACB and HN(CO)CACB. The HNCACB links each NH group with the Cα and Cβ chemical shift of its own residue (i) and the preceding residue (i+1), whereas the HN(CO)CACB only links NH groups with the Cα Cβ chemical shifts of the residue before (i-1). These two experiments are suitable for small to medium proteins, for larger proteins the signal-to-noise may not be great and assignment using a HNCA, HN(CO)CA, HNCO and HN(CA)CO might be the better experiments. The HNCANNH experiment is specific for intrinsically disordered proteins (IDPs), it links the NH group of its own (i) with the Cα of the residue before (i-1).

Three dimensional assignment experiments of $FOXP2^{IDR}$ were recorded on a 600MHz Bruker Avance and $FOXP2^{FH}$ were recorded on a 700 MHz Bruker Avance III NMR spectrometer, both equipped with a TCI cryoprobe at 298 K. HNCACB spectra (hncacbgp3d) were recorded with spectral widths 13.6543/24.00/62.00 ppm, centered at 4.7/118/39 ppm in $^{1}H/^{15}N/^{13}C$, with 1024, 64, 200 points, respectively. HN(CO)CACB spectra (hncocacbgp3d) were recorded with spectral widths 16.02/24.00/62.00 ppm, centered at 4.7/118/39 ppm in $^{1}H/^{15}N/^{13}C$, with 1024, 64, 200 points, respectively. HN(CA)NNH spectra (hncannhgp3d, hncannhgp3d.2) were recorded with spectral widths 13.66/24.00/24.00 ppm or 13.66/24.00/4.00 ppm, centered at 4.7/118/118 ppm or 4.7/117/4.7 ppm in $^{1}H/^{15}N/^{13}C$, with 1024, 64, 200 points or 1024, 64, 100 points, respectively.

Chemical shift perturbations:

Assuming a two-site exchange, the binding of two proteins results in different resonance frequencies ω_{free} and ω_{bound} and their difference $\Delta\omega=\omega_{bound} - \omega_{free}$. How the different species appear in a spectrum depends on the dissociation constant (K_d) and the exchange rate (k_{ex}).

In NMR spectroscopy there are three different exchange regimes based on k_{ex} and the difference in resonance frequency ($\Delta\omega$).

• Fast exchange where $k_{ex} >> | \Delta\omega |$

In fast exchange interaction a single peak appears at a population weighted average chemical shifts.

• Intermediate exchange where $k_{ex} \sim |\Delta\omega|$

In intermediate exchange signals are severely broadened. The position and intensity of the peaks are highly uncertain, making the interpretation more complicated.

• Slow exchange $k_{ex} << |\Delta\omega|$

For interactions in slow exchange each state and the corresponding frequencies can be individually observed.

Secondary structure propensity:

Assignment of $^{13}C\alpha$ and $^{13}C\beta$ chemical shift were performed on FOXP2 IDR human and chimpanzee in order to predict the propensity of the corresponding residues to form alpha helix or beta-stranded secondary structure elements. The chemical shifts of the protein backbone are sensitive to the local backbone geometry and can therefore provide information on the propensity of secondary structural elements. This allows to derive secondary structure elements and dihedral angles from chemical shifts. The $^{13}C\alpha$ and $C\beta$ chemical shift difference between FOXP2 IDR and random coil residues is dependent on the protein secondary structure and can be calculated using following formula:

$$\Delta\delta = \delta_{\text{observed}} - \delta_{\text{random coil}}$$

Random coil chemical shifts were predicted using the ncIDP library, which is optimized for IDPs, harboring many proline residues[109]. Positive chemical shift differences values are observed if the corresponding amino-acids tend to form α-helical structures and negative values it they are likely to be in a β-stranded secondary structure. Mapping the transient secondary structure elements within IDR is essential, as characteristic for protein-protein interaction or PTMs sites.

Relaxation experiments:

$^{1}H^{15}N$ HetNOE experiments were measured for FOXP2 IDR at a concentration of 300 µM on a 600 MHz Bruker Avance Neo NMR spectrometer equipped with a TXI 600S3

probehead at 298 K. Experiments were recorded with spectral widths 16.0176/19.0033 ppm, d1 3 sec, centered at 4.7/118.5 ppm, with 2048, 512 points.

In-cell phosphorylation experiments:

Samples for NMR phosphorylation assays contained 50 µM ^{15}N labelled FOXP2 IDR constructs in NMR buffer with 300 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 5 mM ATP, PhosphoSTOP™ (1x, Roche) mixed with 500µl of HEK-lysate (15 mg/ml, estimated using Pierce BCA protein assay kit, ThermoFisher) or 500µl of soluble mouse brain cell extract (wildtype, C57Bl/6) and 10% D_2O added for the lock signal. 20 ^{1}H ^{15}N HSQC NMR spectra were recorded in a row at 298 K on a 600 MHz Bruker Avance Neo NMR spectrometer equipped with a TXI 600S3 probehead. All spectra were recorded with a recycle delay of 1.0 s, spectral widths of 15.9/30 ppm, centered at 4.7/118.0 ppm in $^{1}H/^{15}N$, with 1,024 and 256 points, respectively, and 16 scans per increment.

2.2.2.2 CD experiments

CD spectra were recorded on a Jasco J-715 Spectropolarimeter spectrometer at 25°C between 190 and 260 nm using 15 µM of either human FOXP2 or chimpanzee FOXP2 in Tris buffer with increasing concentrations of TFE: 0%, 5%, 10%, 15%, 30%, 40% or 50%.

2.2.2.3 ITC experiments

For ITC experiments, FOXP2 FH, FOXP2 IDR and β-catenin were prepared in Tris buffer and FOXP2 FH, FOXP2 FH-IDR and DNA in HEPES buffer. Binding affinities of FOXP2 FH to FOXP2 IDR, β-catenin or DNA were determined at 10 °C with 24 injections of each 4µl. The concentration in the cell was 10 µM for all measurements, the concentrations in the syringe was 100-200 µM, estimated right before the measurement using NanoDropTM (Peqlab). The ITC data were analysed with the programme MicroCal Origin software version 7.0 and binding information such as stoichiometry (N), K_D (binding affinity) and ΔH (binding enthalpy) calculated.

2.2.2.4 SAXS experiments

SAXS data were recorded with an in-house SAXS instrument (SAXSspace, Anton Paar, Graz, Austria) equipped with a Kratky camera, a sealed X-ray tube source and a Mythen2 R 1 K Detector (Dectris). Thereby Axin-1/GSK3ß complex and the buffer for background subtraction where loaded via an ASX autosampler and measured in a flow cell. The scattering patterns were measured with a 180-min exposure time (180 frames, each 1 min). Radiation damage was excluded on the basis of a comparison of individual frames of the 180-min exposures, wherein no changes were detected. Obtained SAXS data were processed using the SAXSanalysis package (Anton Paar, version 3.0) and analyzed using the ATSAS package (version 2.8.2, Hamburg, Germany). The data were desmeared using GIFT (PCG-Software). The forward scattering (I(0)), the radius of gyration (Rg), the maximum dimension (Dmax), and the interatomic distance distribution function (P(r)) were computed with GNOM. To calculate surface models based on the P(r) functions DAMMIF was employed, which uses GNOM files as input. For each structure, 50 simulated annealing runs were and the resulting models were superimposed, averaged and filtered using DAMAVER. Matching models were then clustered by DAMCLUST.

2.2.3 Fluorescence-based methods

2.2.3.1 EMSA experiments

EMSA experiments were performed according to the protocol of Hellman and Fried[110]. For studies of FH bound to FAM-labeled DNA a 10% polyacrylamide gel containing 400mM Tris, 25mM EDTA at pH 7.8, 40% acrylamide-bisacrylamide dd. H_2O, 0.05g Ammonium persulfate and 12 µL of TEMED were used. FOXP2 FH was prepared in HEPES buffer. Fluorescine amidite (FAM)-labeled DNA was ordered as primer from IDT DNA GmbH with following target sequence: fw: 5'CGCG CTC TTG TTT ACA GCT 3', rv: 5'AGC TGT AAA CAA GAG CGCG 3', whereby only the FW primer contained the FAM label at the 5’ end, to obtain one label per annealed DNA molecule. Both primers (labelled and unlabled) were then dissolved in HEPES buffer to an equal concentration, then mixed in equal parts, boiled at 95°C for 10 min and then cooled on ice for 30 min. Concentrations was estimated measuring the FAM label with the

NanoDropTM (Peqlab) at wavelength 488nm and then calculated by the extinction coefficient of the label. Samples were prepared by mixing DNA and FH in different ratios with 10X binding buffer. Samples contained 200nM of DNA in each well. FH was added in increasing concentrations. Gel was preran with dye-labeled loading buffer to determine the migrations pattern for each gel. Samples were then loaded in rinsed wells and ran at 50V for 45-60 min. Gels were then removed from gel-chamber and DNA was detected using a FAM filter.

2.2.4 Cell-based methods:

2.2.4.1 Cell culture

Experiments were kindly performed by Dr. Chintan Koyani from the Medical University in Graz.

2.2.4.2 Co-Immunoprecipitation

Experiments were kindly performed by Dr. Chintan Koyani from the Medical University in Graz.

2.2.4.3 RNA-Isolation

RNA was isolated for RNA-Seq analysis and q-PCR analysis. Therefore six conditions of each five biological replicates were used:

1. Condition: mock transfected cells as control

2. Condition: FOXP2 wildtype overexpression

3. Condition: CHIR treated cells

4. Condition: $FOXP2^{\Delta helix}$ overexpression

5. Condition: FOXP2 wildtype overexpression and CHIR treatment

6. Condition: $FOXP2^{\Delta helix}$ overexpression and CHIR treatment

U2OS cells were used, as they do not contain endogenous FOXP2 and only cytoplasmic β-catenin. CHIR treatment (99021, Tocris) was used to translocate β-catenin in the nucleus via GSK3 β inhibition.

2.2.4.4 RNA-Seq Preparation and Analysis

First, RNA integrity and quality was determined using a bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA). A cDNA library was prepared using the TruSeq Stranded mRNA Sample Prep Kit (Illumina Inc., San Diego, USA) according to manufactures recommendation. Briefly, 1µg of total RNA was used for first-strand synthesis performed on a random hexamer and SuperScript II (Life Technologies, Carlsbad, CA, USA). Second-strand synthesis was performed using dUTP and the Illumina specific Second Strand Marking Master Mix. After end repair and A-tailing indexed adaptors were ligated to the cDNA fragments. Those fragments were then enriched using PCR for 15 cycles and purified using AMPure XP Beads (Bechman Coulter Inc., Brea, CA, USA). The final libraries were quality checked and quantified by q-PCR. On average 61.7 million reads were obtained (range 48-87) per sample. The obtained data were analyzed with the DeSeq-bioconductor software in order to obtain information about differentially expressed genes. I obtained technical support from Christine Beichler and bioinformatical support for the DESeq analysis from Peter Ulz and Prof. Ellen Heitzer from the Medical University Graz.

2.2.4.5 Real-time RT-PCR

To quantify mRNA from chosen genes in the total RNA samples, the Kit Luna Universal One-Step RT-qPCR Kit (NEB Laboratory) was used according to the protocol. Therefor 25ng of total RNA was used as input. For analysis the ΔΔCT-method was performed, samples were normalized to the house-keeping gene TBP (Tata-box binding protein) and calculated the log2 FC compared to the Mock-transfected cells (cells treated only with transfection chemicals).

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
BMP4	ATGATTCCTGGTAACCGAATGC	CCCCGTCTCAGGTATCAAACT
CCND1	GCTGCGAAGTGGAAACCATC	CCTCCTTCTGCACACATTTGAA
CD44	CTGCCGCTTTGCAGGTGTA	CATTGTGGGCAAGGTGCTATT
HPRT1	CCTGGCGTCGTGATTAGTGAT	AGACGTTCAGTCCTGTCCATAA
JUN	TCCAAGTGCCGAAAAAGGAAG	CGAGTTCTGAGCTTTCAAGGT
LEF1	TGCCAAATATGAATAACGACCCA	GAGAAAAGTGCTCGTCACTGT
TIAM1	CCTGTGTCTTACACTGACTCTTC	CATCCCCGTAAAGCCTGCTC

Tab.2: primer sequenes used for q-PCR

Chapter 3

3 Results

3.1 Intrinsically disordered region of FOXP2 contains secondary structures propensity

The human isoform and the chimpanzee isoform of FOXP2 differ only in two residues. Residue 303 is an asparagine in humans and a threonine in chimpanzee, residue 325 is a serine in humans and an asparagine in chimpanzees (Fig. 1.2). At the beginning of this thesis I was interested in the impact of these two amino acid substitutions on the structure of FOXP2, as they seem to have a huge impact on the function of FOXP2. For those studies we started with the IDR from residue 247 to residue 341 between the poly-Q region and the zinc finger, as this region contains both amino acids which differ between human and chimpanzee. Even though the region is intrinsically disordered, it can adopt transient secondary structures which might indicate favored sites for protein-protein interactions. After expression and purification of $FOXP2^{IDR}$ first CD-experiments were performed with both proteins to detect possible secondary structures in a fast and simple way. By adding increasing amounts of Trifluorethanol (TFE) to the sample the formation of possible secondary structures can be followed. TFE is a chemical, which is known to induce and stabilize secondary structures[111]. The CD experiments in absence of TFE show a curve typical for disordered proteins for both, human and chimpanzee $FOXP2^{IDR}$. The CD curves of α-helical proteins are typical and harbor two maxima of CD signals at 208 nm and 222 nm, which can be observed with the human $FOXP2^{IDR}$ at 30% TFE. Then, with increasing TFE concentration the α-helical propensity of $FOXP2^{IDR}$ becomes stronger so that the signal intensity at 208nm and 222nm increases progressively (Fig. 3.1, left). This α-helical propensity of the 248-340 disordered region of human FOXP2 is a first clue that this region could be involved in protein-protein interaction.

The same experiment for mouse $FOXP2^{IDR}$ was recorded in order to show potential differences to form α-helical structures between human and mouse FOXP2 due to the two amino acids substitutions (Fig. 3.1, right).

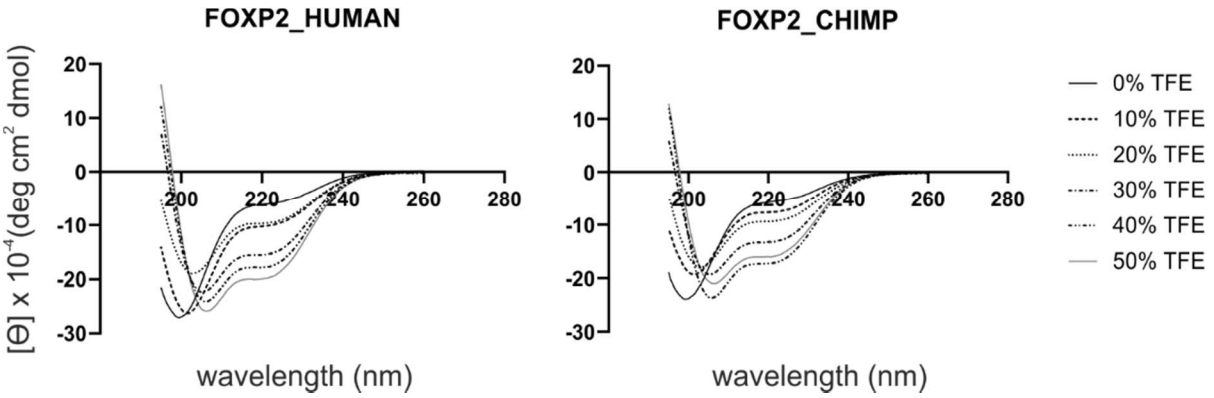

Fig.3.1: CD-spectra of human FOXP2^{IDR} (left graph) and chimpanzee FOXP2^{IDR} (right graph). Curves show the structure of FOXP2^{IDR} with different concentrations of TFE as secondary structure inducing chemical.

As the effect of TFE is discussed to produce false positive data, NMR was used to confirm this secondary structure propensity in absence of TFE. NMR is especially useful, as it can not only tell, if there is an α-helix or β-sheet propensity, it can also provide information about, which residues are involved in the formation of those secondary structures elements. To get these information it is necessary to assign the protein of interest. NMR-derived chemical shift values from the $^{13}C_\alpha$ and C_β can be used to provide information about secondary structure elements propensity (i.e. 2.2.2.1). Therefore data about chemical shifts of $^{13}C_\alpha$ and $^{13}C_\beta$ nuclei were used in order to determine deviations from random coil chemical shifts. Hereby, positive chemical shift differences indicate a tendency to form α-helical structures and negative chemical shift differences indicate a tendency to form β-stranded structures. After assigning the $^{13}C_\alpha$ and C_β chemical shifts of the IDR of human and chimpanzee FOXP2, the chemical shift differences (observed – random coil) were calculated (Fig. 3.2). Indeed, in the region of residue 264 to 272 those calculations revealed positive values meaning, that in region there is a propensity of forming an α-helix. Negative values would indicate a β-sheet formation propensity. As a secondary structure element is not formed by only one amino acid, positive or negative calculated values must stretch over at least 4 amino acids in a row.

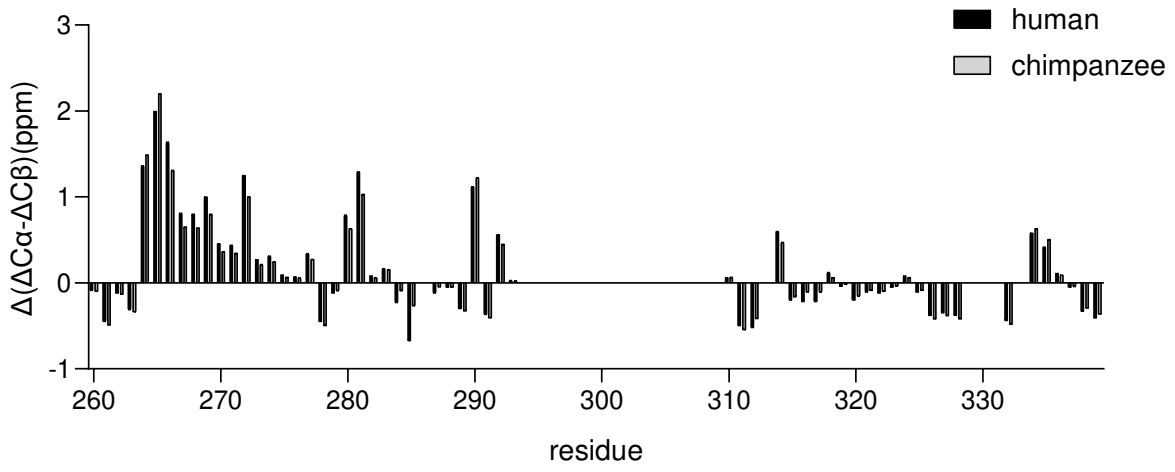

Fig.3.2: Cα-Cβ chemical shift of FOXP2^{IDR} human and chimpanzee.

In addition, relaxation experiments are a useful tool for the determination of flexibility of protein regions. Secondary structure motifs are more rigid, therefore HetNOE values are positive, whereas highly flexible disordered regions have largely negative values. A hetNOE experiment with ^{15}N labeled IDR of human FOXP2 was recorded to confirm the existence of a secondary structure propensity. Hereby, after calculation positive values from residue 264 to 272 were observed proving the existence of a less flexible region (Fig. 3.3). For the rest of the sequence negative values were observed, which proof that the rest is fully disordered.

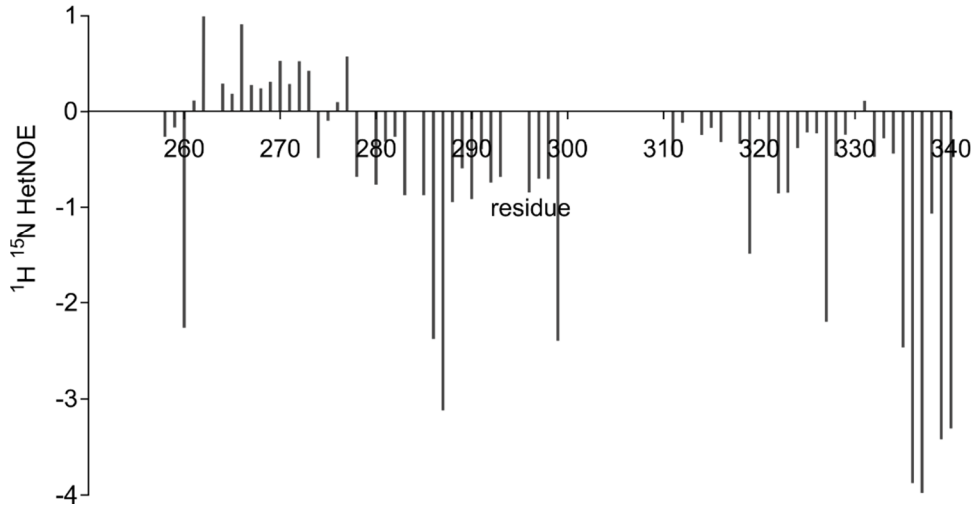

Fig.3.3: HetNOE experiment indicates a less flexible region between residue 263 and 273 of human FOXP2 in agreement with the presence of a secondary structure element in this region.

These data indicate that the IDR between the Poly-Q region and the zinc finger of FOXP2 indeed shows an α-helical propensity. This secondary structure could be necessary for protein-protein interactions and thus an important element in FOXP2 function.

3.2 Region with α-helical propensity is a protein-protein interaction site

To confirm, that this region is indeed important for binding to other interaction partners we used different protein constructs to check possible interactions. FOXP2 plays a crucial role in embryonal development and therefore must be tightly regulated, thus possible protein partners involved in signaling pathways were tested. As already known FOXP1[93] and FOXP3[94] are linked to the Wnt signaling pathway. Lymphoid enhancer-binding factor 1 (LEF1) is a transcription factor, which is expressed in multiple tissues during embryonal development and a key player of the Wnt-/β-catenin signaling pathway and its regulation. Thereby it activates the transcription of Wnt targets in presence of β-catenin[112,113]. It contains a β-catenin binding domain at the N-terminus and a DNA-binding domain called high mobility group box (HMG box) from residue 299 to 367[102]. In order to investigate if LEF1 interacts with FOXP2 and thereby competing with β-catenin binding, we used NMR. First, a construct of LEF1 spanning from residue 1 to 299 (LEF^{1-299}) was used to determine possible binding events between the IDR of FOXP2 and the β-catenin-binding site of LEF1. A ^{1}H ^{15}N HSQC reference spectrum with ^{15}N labeled $FOXP2^{IDR}$ was recorded, then increasing amounts of unlabeled $LEF1^{1-299}$ was added and further ^{1}H ^{15}N HSQC spectra recorded (Fig. 3.4). No differences in ^{1}H-^{15}N cross-peak chemical shifts or intensity were detected, thus no binding event occurred between both proteins.

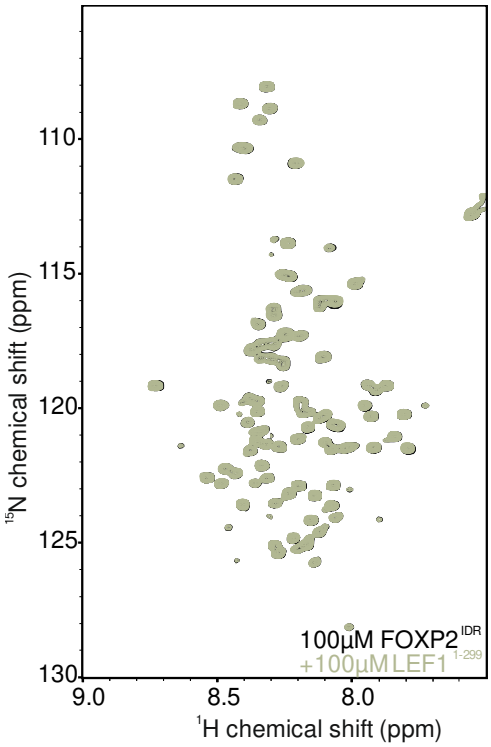

Fig. 3.4: ^{1}H ^{15}N HSQC with ^{15}N labeled $FOXP2^{IDR}$, reference (black, 100µM), titrated with unlabeled LEF^{1-299} (olive, 100µM).

Next, the other part of LEF1 from residue 288-399 ($LEF^{288-399}$), which contains the HMG-box domain, was tested. Thereby a reference spectrum of ^{15}N labeled $FOXP2^{IDR}$ was recorded and titrated with increasing amounts of unlabeled $LEF^{288-399}$. Our assignment was then used in order to localize the involved binding sites on $FOXP2^{IDR}$. Indeed, chemical shift perturbations (CSPs) of the ^{1}H-^{15}N $FOXP2^{IDR}$ cross-peaks were found upon addition of unlabeled $LEF^{288-399}$ indicating the direct binding of $FOXP2^{IDR}$ to $LEF^{288-399}$ around the residues of FOXP2 which were before found to be involved in the α-helix formation (Fig. 3.5).

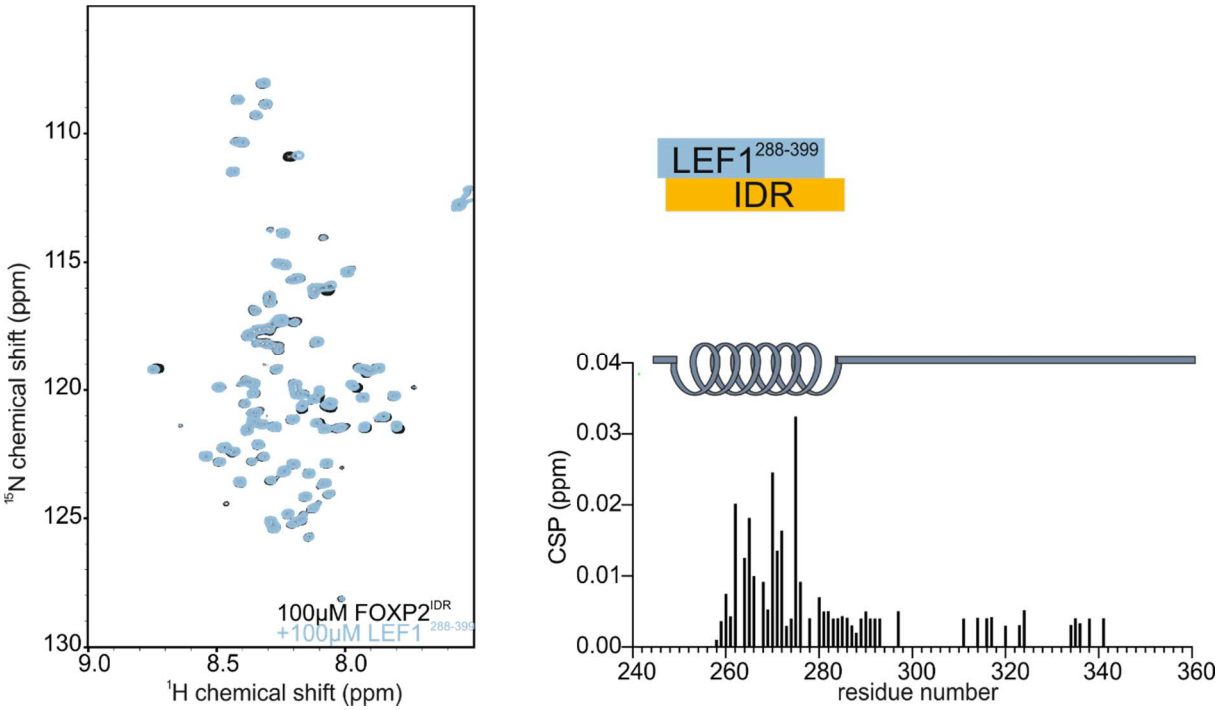

Fig.3.5: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{IDR} (black, 100µM), titration with unlabeled LEF$^{288-399}$ in light blue (200µM). Right: CSP plot of affected residues of FOXP2^{IDR}.

β-catenin interaction protein 1 (ICAT) is known to be a negative regulator of the Wnt signaling pathway by preventing the interaction between β-catenin and LEF1[114,115]. Our data have shown a direct interaction between LEF1 and FOXP2, thus we were wondering, if there might be competition between ICAT, LEF1 and FOXP2^{IDR}. For this another reference spectrum of ^{15}N labeled FOXP2^{IDR} was recorded and increasing concentrations of unlabeled ICAT were added. Indeed, CSPs were detected for the same FOXP2 residues than those affected by LEF$^{288-399}$ addition suggesting a competition between LEF1 and ICAT for FOXP2 binding (Fig. 3.6).

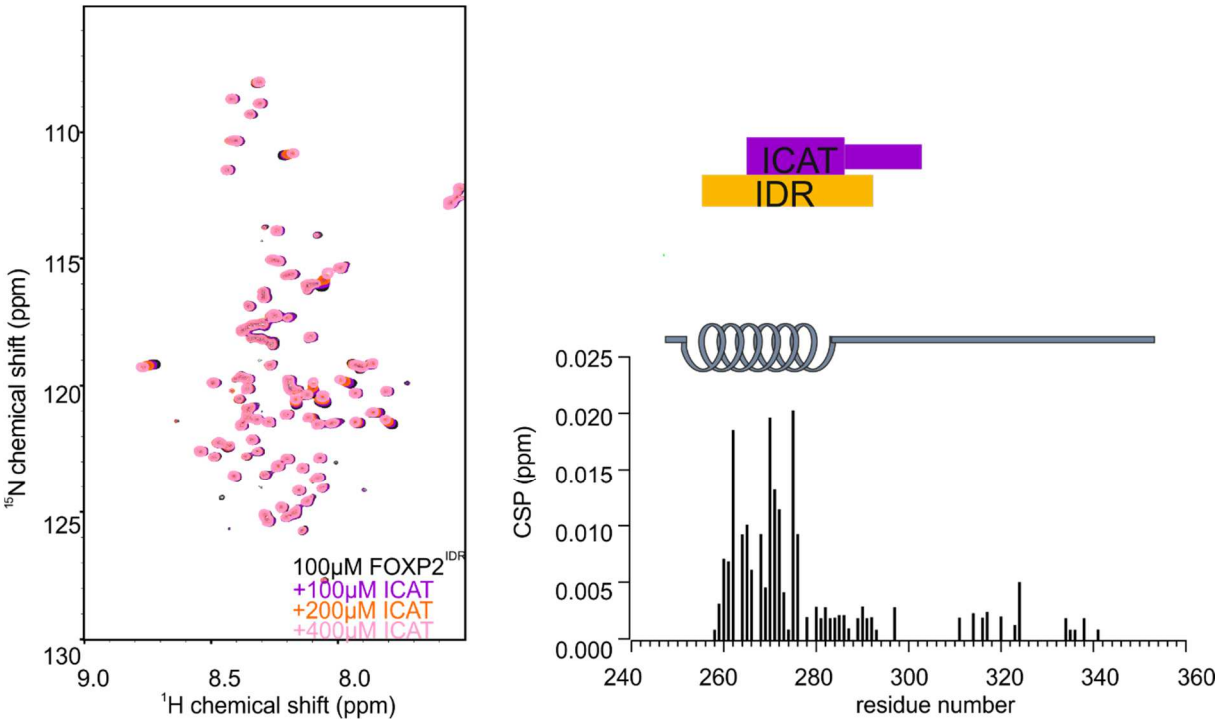

Fig.3.6: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{IDR}, reference in black, titration with unlabeled ICAT in magenta (100µM), orange (200µM) and rose (400µM). Right: CSP plot of affected residues of FOXP2^{IDR}.

3.3 FOXP2 interacts with β-catenin

The Wnt signaling pathway was studied extensively in correlation with embryonal and cancer development. It has already been show, that few FOX proteins are related to the Wnt signaling pathway and, some of them, interacting with a crucial player of the Wnt signaling pathway, the transcriptional co-activator β-catenin[93,94,106,116,117]. As FOXP2 functions as transcription factor, is could be regulated by β-catenin, as its active form occur in the nucleus upon Wnt-activation. As FOXP1[93] and FOXP3[94] were shown to interaction with β-catenin via Co-IP, thus enhancing the Wnt signaling pathway, we hypothesized that FOXP2 is a novel binding partner of β-catenin.

In order to determine if β-catenin and the IDR of FOXP2 are interacting, NMR was used as fast and sensitive method. A ^{1}H ^{15}N HSQC reference spectrum with ^{15}N labeled FOXP2^{IDR} was recorded. Then increasing amounts of unlabeled β-catenin^{FL} (β-catenin full-length) was added. The FOXP2 ^{IDR} reference spectrum shows less signal as compared to the previous ones, as the used buffer had a pH of 7.5. β-catenin has an pI-value of 5.8, thus it is insoluble at a pH of 6.5 or less. As seen in figure 3.7

chemical shift perturbations of ^{1}H-^{15}N FOXP2^{IDR} cross-peaks were observed upon β-catenin addition.

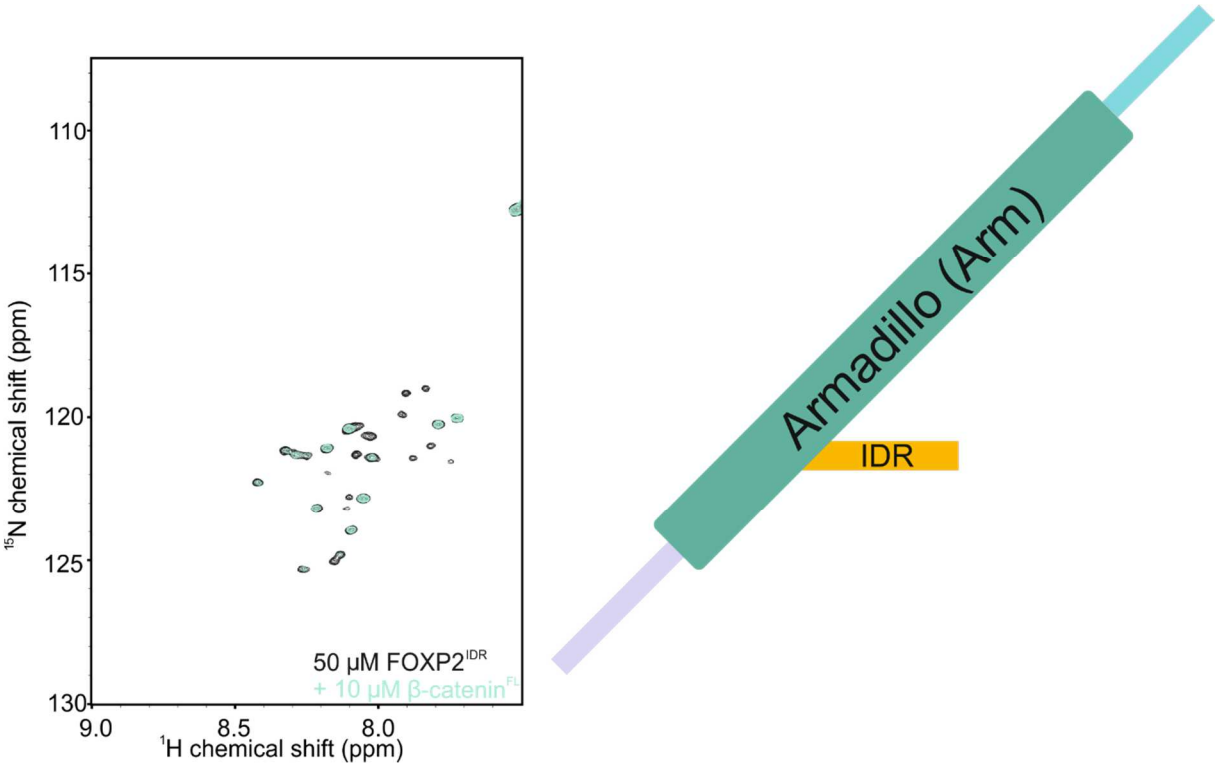

Fig.3.7: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{IDR} (black, 50µM), titration with unlabeled β-catenin^{FL} (light-blue, 10µM).

As β-catenin is a co-activator of transcription factors the question arose, if the IDR is the only binding site for β-catenin. From the molecular side it would make sense, if β-catenin would also bind close to the DNA-binding domain in order to affect DNA-binding and thus the transcription of FOXP2 target genes. To investigate this hypothesis ^{15}N labeled FOXP2^{FH} was used and a ^{1}H ^{15}N HSQC was recorded as reference. The recorded spectrum displayed characteristic signals for a folded protein. Then increasing amounts of unlabeled β-catenin was added. Indeed, CSPs were observed indicating the interaction between the FH domain of FOXP2 and β-catenin (Fig. 3.8). Thus we found two binding site of β-catenin to FOXP2.

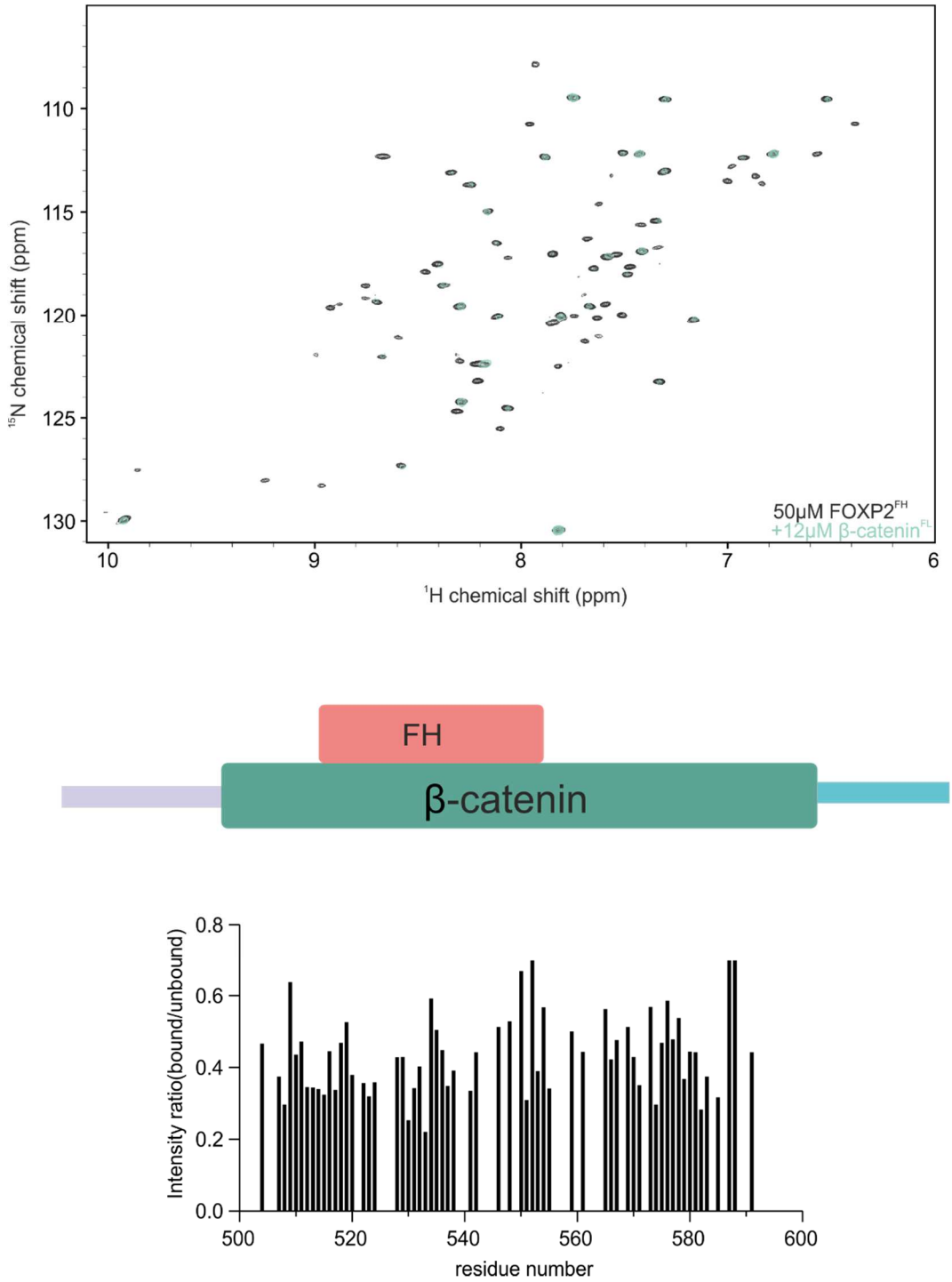

Fig.3.8: Above: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH} (black, 50µM), titration with β-catenin^{FL} (light-blue, 10µM). Below: Intensity plot indicating the binding-affected residues.

In order to validate this interaction in cells Co-IP experiments were used in HEK-293 T cells. Both, FOXP2 and β-catenin, are expressed endogenously in this cell type in sufficient amounts, thus overexpression was not necessary. Indeed, we could pull-down β-catenin with FOXP2 antibodies and *vice versa*. This indicated the interaction

could also take place in human cells and at endogenous level (Fig. 3.9, cell culture and western blot carried out by Dr. Chintan Koyani from Medical Unversity Graz).

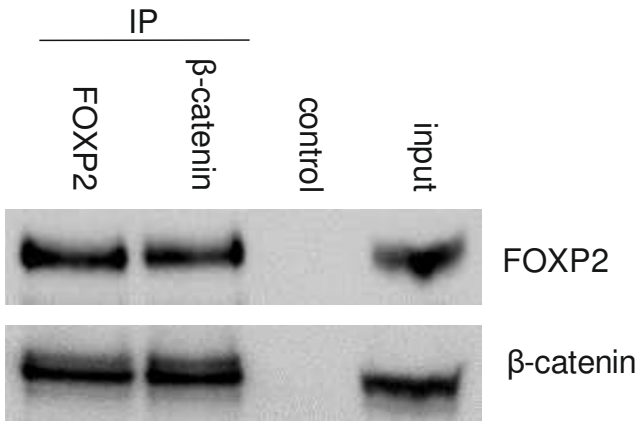

Fig.3.9: Co-IP (carried out by Dr. Chintan Koyani, Medical University Graz) of endogenous FOXP2 and β-catenin from HEK-293-T cells.

In order to specify the binding site of β-catenin to FOXP2 we used different β-catenin constructs. β-catenin contains of a flexible N-terminus, followed by a domain called Armadillo region, and a flexible C-terminus. It is known, that the Armadillo domain is binding to other IDRs of interaction partners[118-121], thus we tried to express and purify this folded domain of β-catenin. This construct has been expressed by *E.coli* but could not be concentrated to sufficient concentrations. Thus we continued with another construct containing the N-terminal part of the domain from residue 141 to 305 (β-catenin$^{141-305}$). This part is known to bind to Axin-1[122] and LEF-1[123]. After recording a reference ^{1}H ^{15}N HSQC spectrum with ^{15}N labeled FOXP2^{IDR} increasing concentration of unlabeled β-catenin$^{141-305}$ were added. Indeed, CSPs of ^{1}H ^{15}N FOXP2^{IDR} cross-peaks were observed, indicating that the N-terminal region of the armadillo domain of β-catenin is binding to the IDR of FOXP2 (Fig. 3.10). Using the ^{1}H ^{15}N FOXP2 backbone assignment the binding site of β-catenin$^{141-305}$ to the FOXP2^{IDR} could be determined at residues 264-272 and thus corresponding to the region with α-helical propensity.

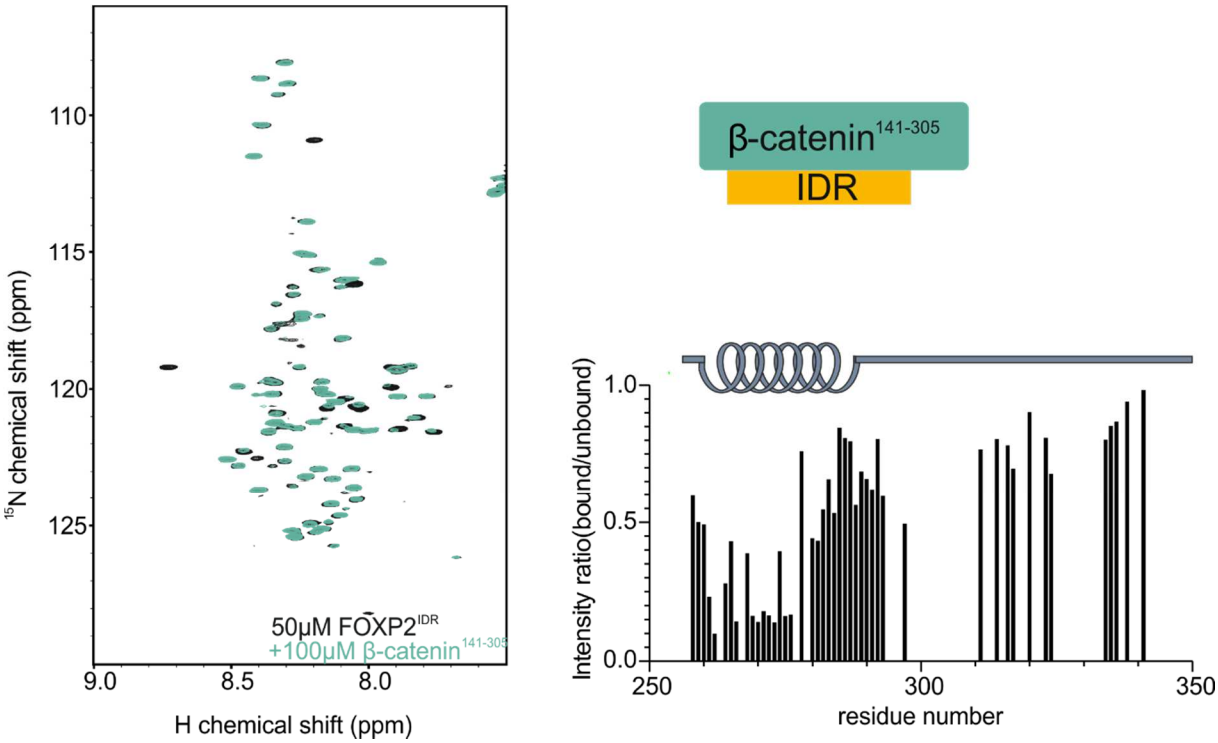

Fig.3.10: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{IDR} (black, 50µM), titration with unlabeled β-catenin$^{141-305}$ (green, 100µM). Right: Intensity plot of residues of FOXP2^{IDR} indicating the binding region.

Also with ^{15}N labeled β-catenin$^{141-305}$ and addition of unlabeled FOXP2^{IDR} CSPs of ^{1}H ^{15}N β-catenin$^{141-305}$ cross-peaks as well as signals broadening could be detected, proofing the interaction between both proteins (Fig. 3.11)

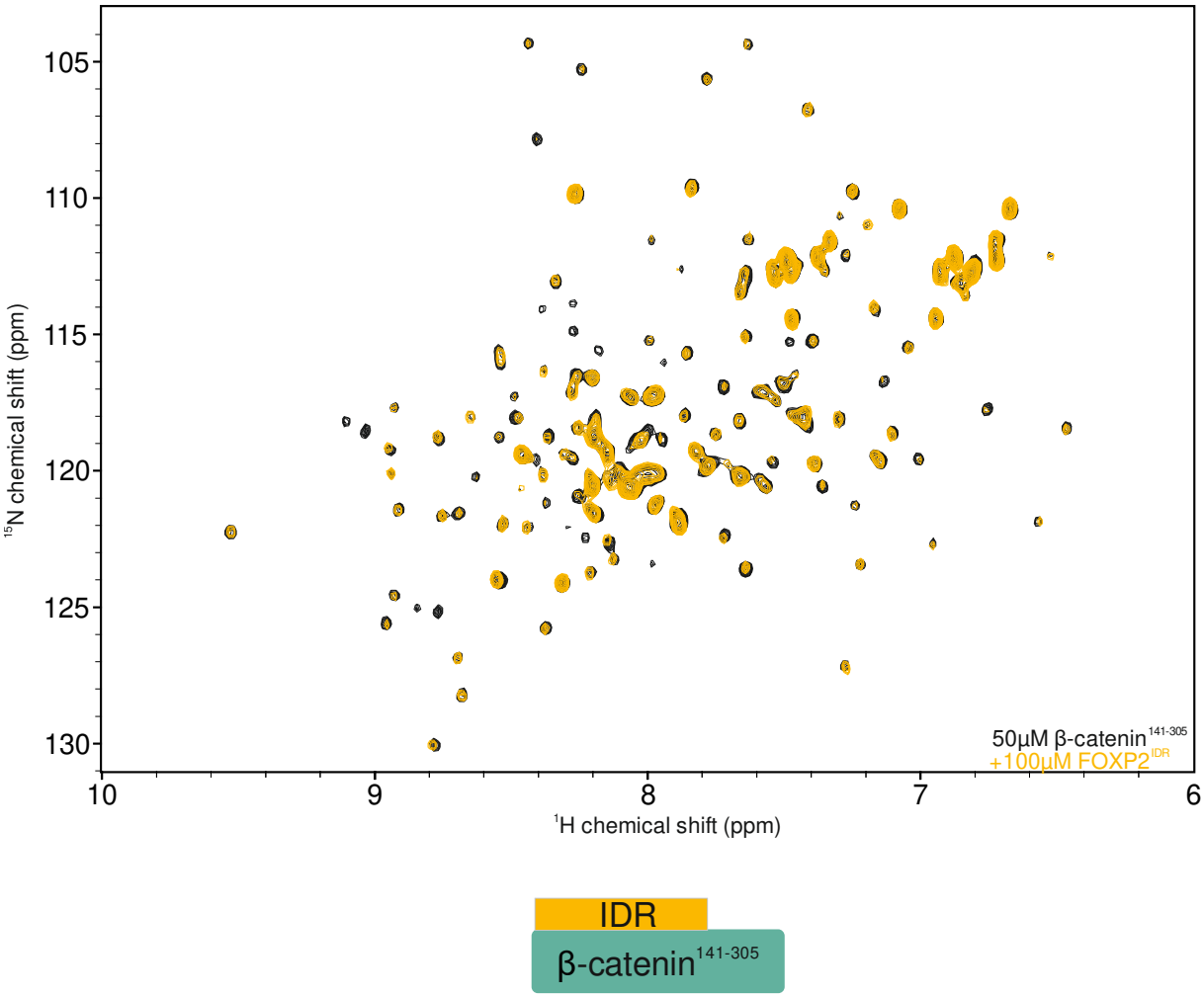

Fig.3.11: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled β-catenin$^{141-305}$ (black, 50µM), titration with unlabeled FOXP2^{IDR} (yellow, 100µM).

Next, we also wanted to specify the binding of β-catenin to the FH domain of FOXP2. First, the part of β-catenin, which also binds to the IDR, was used. For this ^{15}N labeled FOXP2^{FH} was used and increasing amounts of unlabeled β-catenin$^{141-305}$ added. When comparing the reference spectrum with the titration spectra, no differences were observed (Fig. 3.12).

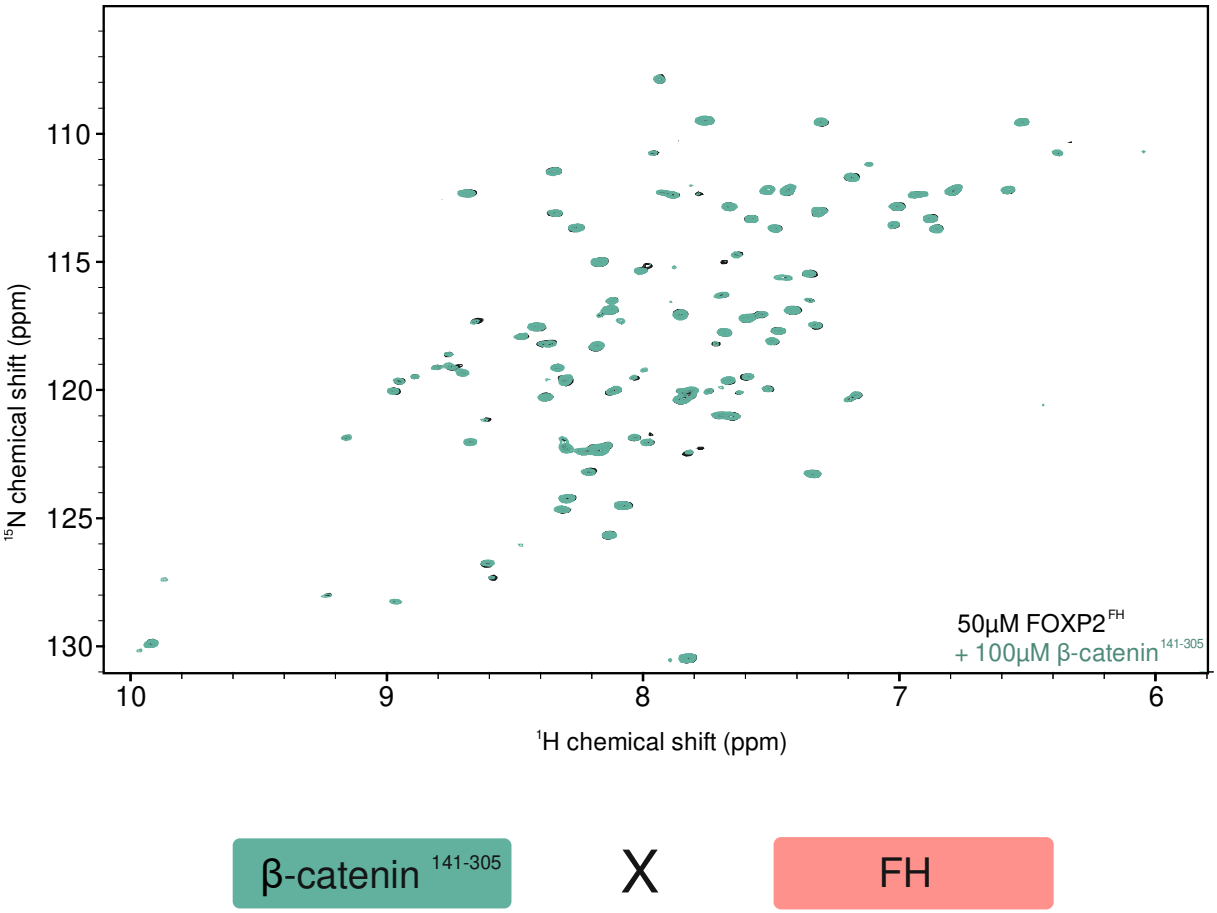

Fig.3.12: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH} (black, 50µM), titration with unlabeled β-catenin$^{141-305}$ (green, 100µM).

Thus, the data show that the part of β-catenin, which binds to the IDR of FOXP2 is not binding to the FH domain. Then, the flexible N-terminus of β-catenin, ranging from residue 1 to 140 (β-catenin$^{1-140}$), was used. Indeed, by comparing the reference spectrum of the FH to spectra with addition of unlabeled β-catenin$^{1-140}$ CSPs of ^{1}H ^{15}N FOXP2^{FH} cross-peaks as well as signals broadening could be detected proving the interaction between the N-terminal IDR of β-catenin and the FH domain of FOXP2 (Fig. 3.13).

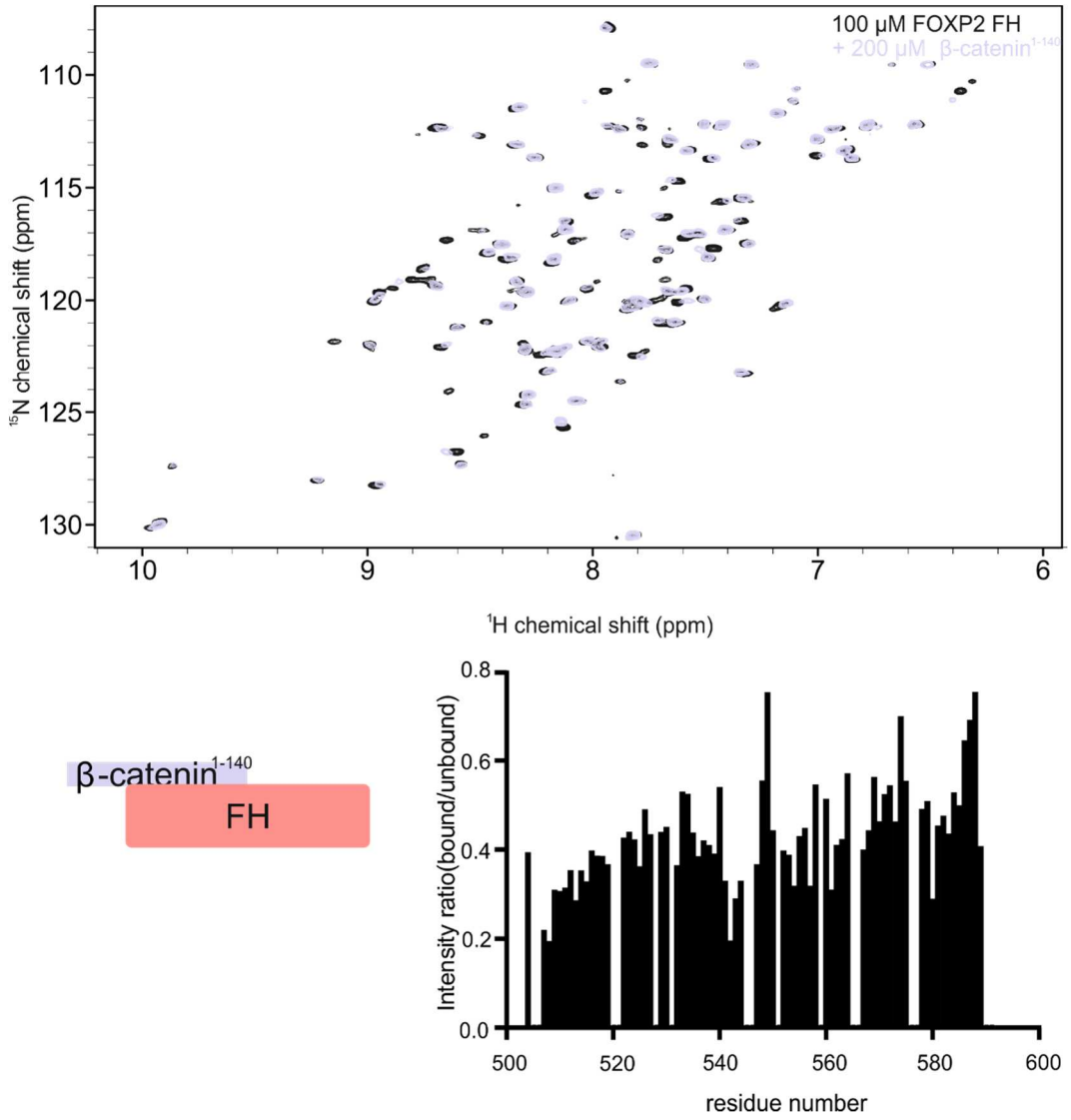

Fig.3.13: Above: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH} (black, 100µM), titration with unlabeled β-catenin$^{1-140}$ (light purple, 200µM). Below: Intensity plot indicating the binding-affected residues.

To test the C-terminal IDR of β-catenin (residue 666-781, β-catenin$^{666-781}$), titrations with ^{15}N FOXP2^{FH} and unlabeled β-catenin$^{666-781}$ were also performed. Indeed, also here CSPs of ^{1}H ^{15}N FOXP2^{FH} cross-peaks as well as signal broadening indicated the interaction between the C-terminal IDR of β-catenin and the FH domain of FOXP2. In order to locate the binding site on the FOXP2 FH domain the obtained ^{1}H ^{15}N NMR signals were assigned in order to link each signal to the corresponding residue. Then, an intensity ratio plot of the FOXP2^{FH} ^{1}H ^{15}N cross-peaks in free and β-catenin bound forms were calculated (Fig. 3.14).

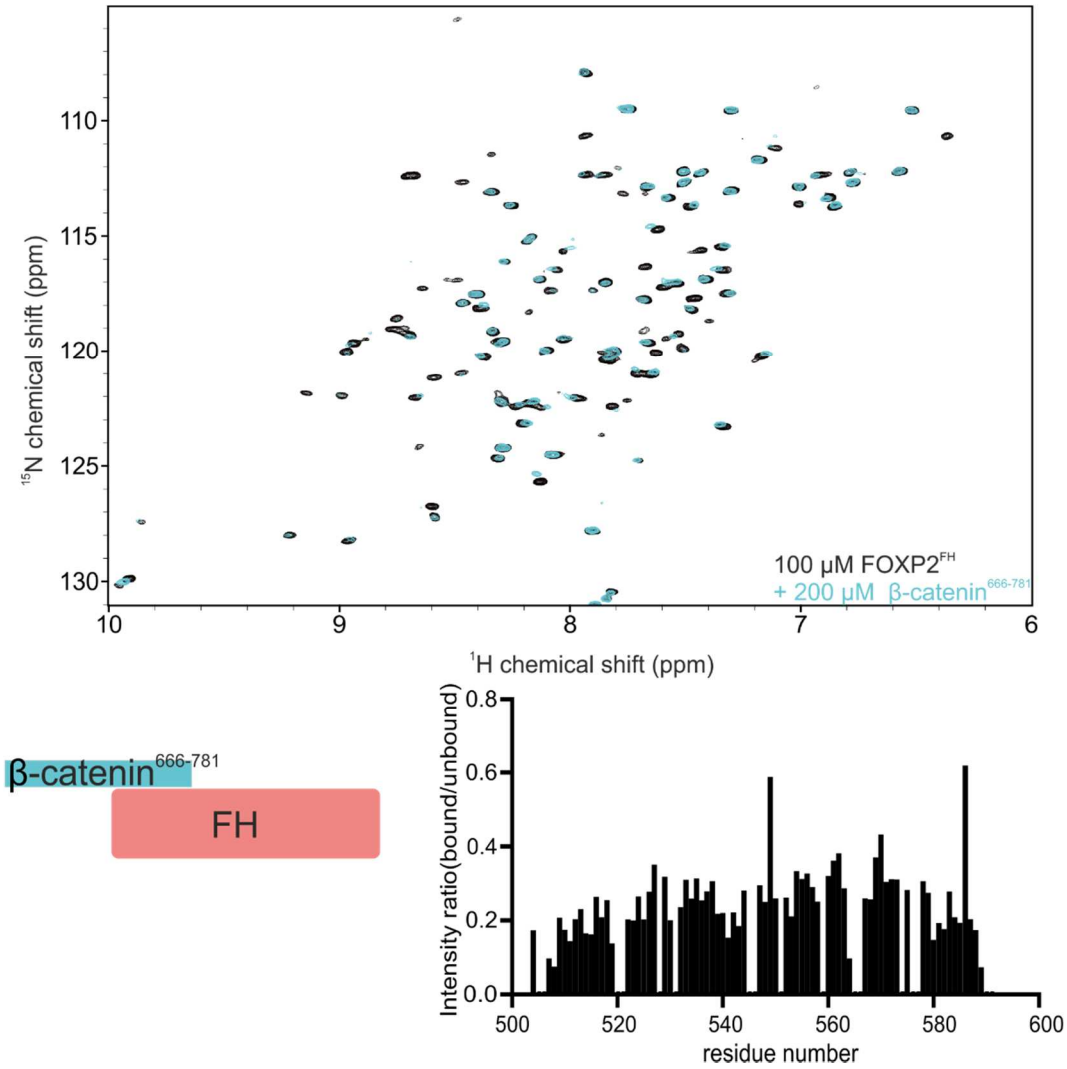

Fig.3.14: Above: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH} (black, 100µM), titration with unlabeled β-catenin$^{666-781}$ (light blue, 200µM). Below: Intensity plot indicating the affected residues.

Analysis of these intensity ratio plots for both N- and C-terminal part of β-catenin showed, that both parts interact with the same residues on the FH (Fig. 3.15).

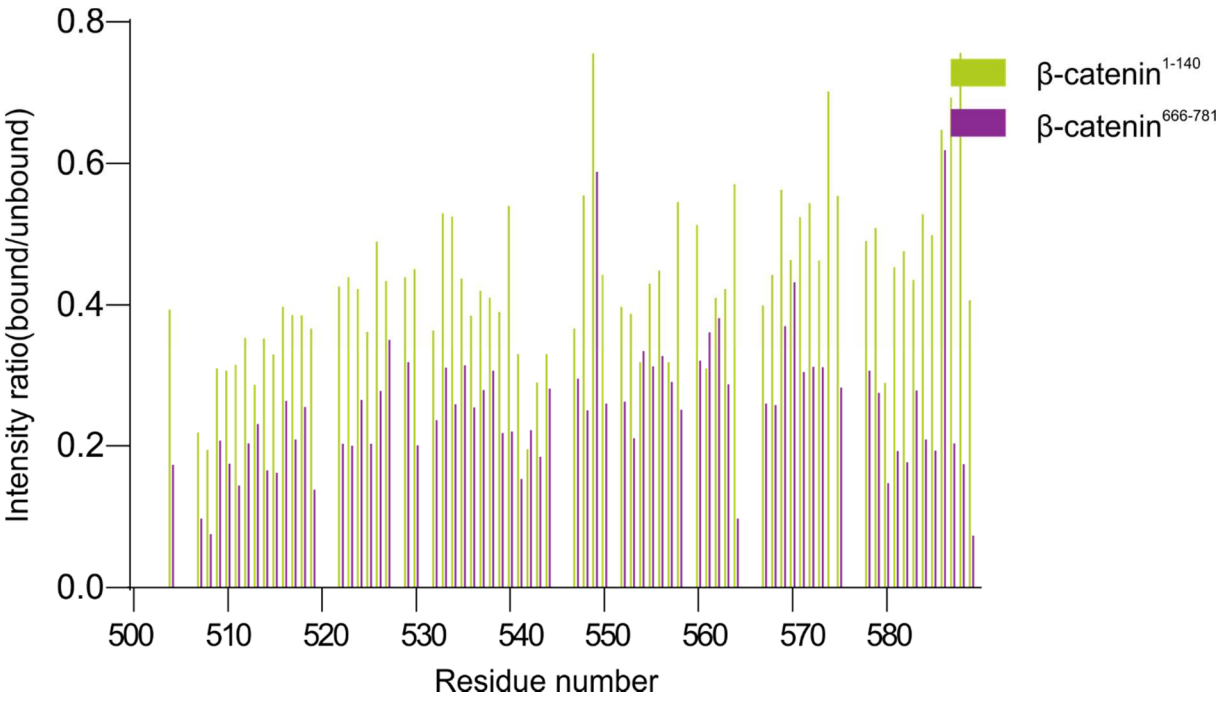

Fig.3.15: overlay of intensities of FH residues bound to β-catenin$^{1-140}$ and β-catenin$^{666-781}$.

In order to investigate the binding regions of β-catenin on FOXP2 in cells, several FOXP2 constructs were designed, decreasing size progressively regarding its functional domains and regions (Fig. 3.16) and their interaction with endogenous β-catenin were tested using further Co-IP experiments.

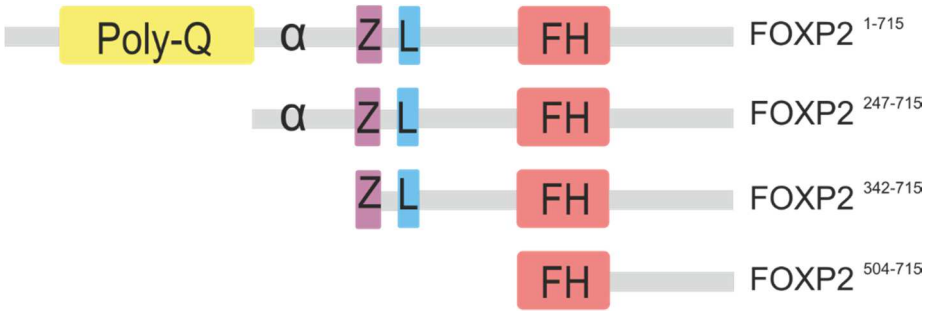

Fig.3.16: FOXP2 constructs designed for Co-IP.

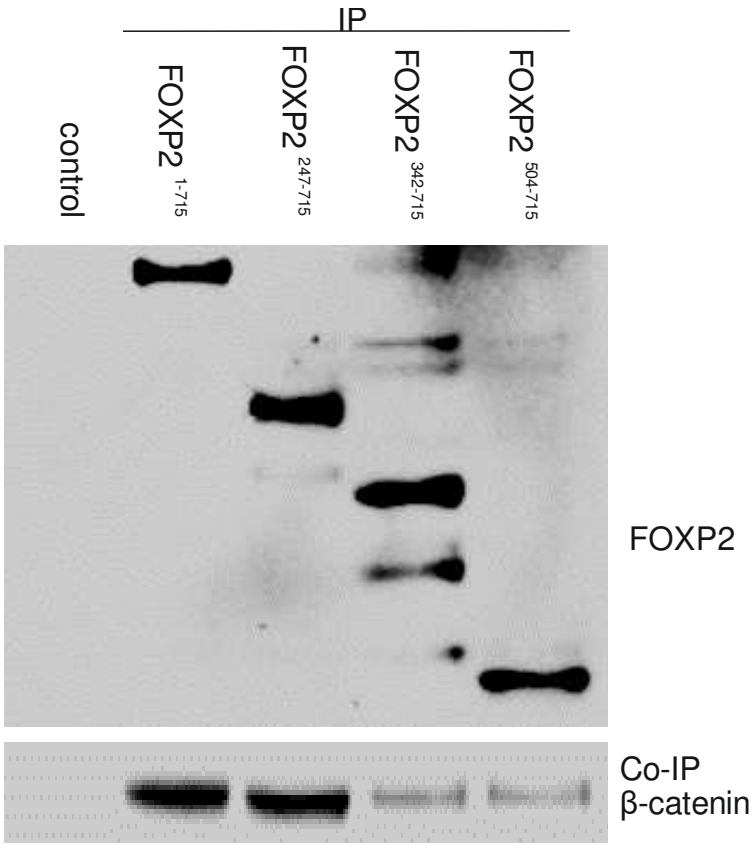

Fig.3.17: Co-IP using U2OS cells with different overexpressed FOXP2 constructs pulling down endogenous β-catenin (carried out by Dr. Chintan Koyani, Medical University Graz).

For this set-up U2OS cells where used, as this cell line contains only endogenous β-catenin and not FOXP2, thus allowing to examine β-catenin binding only to the exogenously expressed FOXP2 constructs. The different FOXP2 constructs were overexpressed in comparable amounts and then Co-IP experiments were performed using either FOXP2 or β-catenin as bait. As expected due to the previous Co-IP results, full-length FOXP2 showed a clear pull-down of β-catenin (Fig. 3.17, cell culture and western blot carried out by Dr. Chintan Koyani from Medical Unversity Graz). Deletion of the N-terminal part of FOXP2 from residue 1 to 246, lacking the poly-Q region showed similar amounts of co-immunoprecipitated β-catenin compared to the full-length construct assuming that this region is not involved in β-catenin binding. In contrast further deletion of the IDR from residue 247 to 341 showed significantly decreased amounts of pulled-down β-catenin suggesting that this FOXP2 region is involved in β-catenin binding. Further deletion of the region 342 to 503 containing the zinc finger and leucine zipper do not alter the amount of pulled-down β-catenin, suggesting that this C-terminal region containing the FH domain is sufficient for β-catenin binding. Taken together, these results confirm our previous NMR data that not

only the FH domain, but also the IDR of FOXP2 around residue 247-341, are both involved in the interaction with β-catenin.

3.4 FOXP2^{IDR} is interacting with FOXP2^{FH}

The two binding sites of β-catenin to FOXP2 are not located in physical proximity, which raised the question, if there might be a mechanism to bring both binding sites close to each other in order to facilitate the binding to β-catenin and other binding partners. In order to check if there might be an intramolecular interaction between the two β-catenin binding sites, ^{15}N labeled FOXP2^{IDR} was used to record a ^{1}H ^{15}N HSQC, then an increasing amount of unlabeled FOXP2^{FH} was added. Indeed, CSPs of ^{1}H ^{15}N FOXP2^{IDR} cross-peaks indicated the direct interaction between the IDR and FH of FOXP2 (Fig. 3.18).

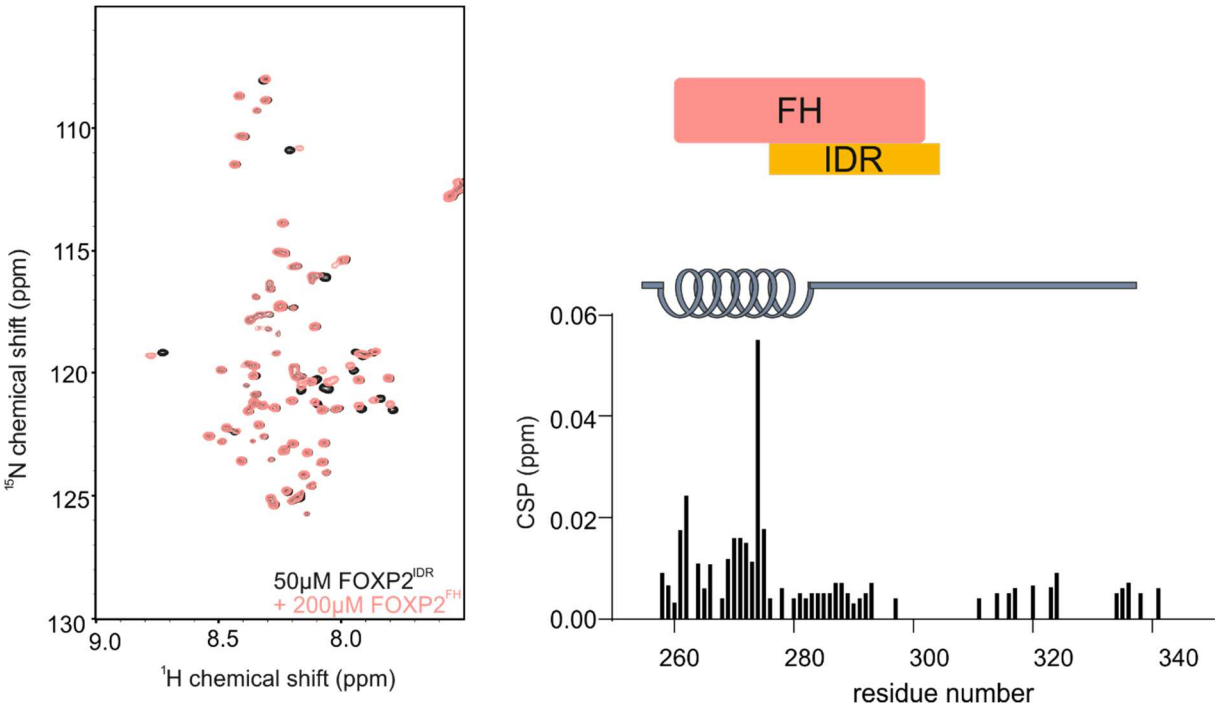

Fig.3.18: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{IDR} (black, 50µM), titration with unlabeled FOXP2^{FH} (coral, 200µM). Right: CSP plot indicating the affected residues around region with α-helical propensity.

Using our NMR assignment the binding site of the FH domain on FOXP2^{IDR} could be determined to the region forming the α-helix. To identify the binding site on the FH we recorded another ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH} and added increasing amounts of unlabeled IDR. Also in this direction CSPs of ^{1}H ^{15}N FOXP2^{FH} cross-peaks and signal broadening were observed indicating the direct interaction between both

protein regions (Fig. 3.19). According to the assignment of the FH the binding site could be determined to be partially overlapping with the binding site of β-catenin.

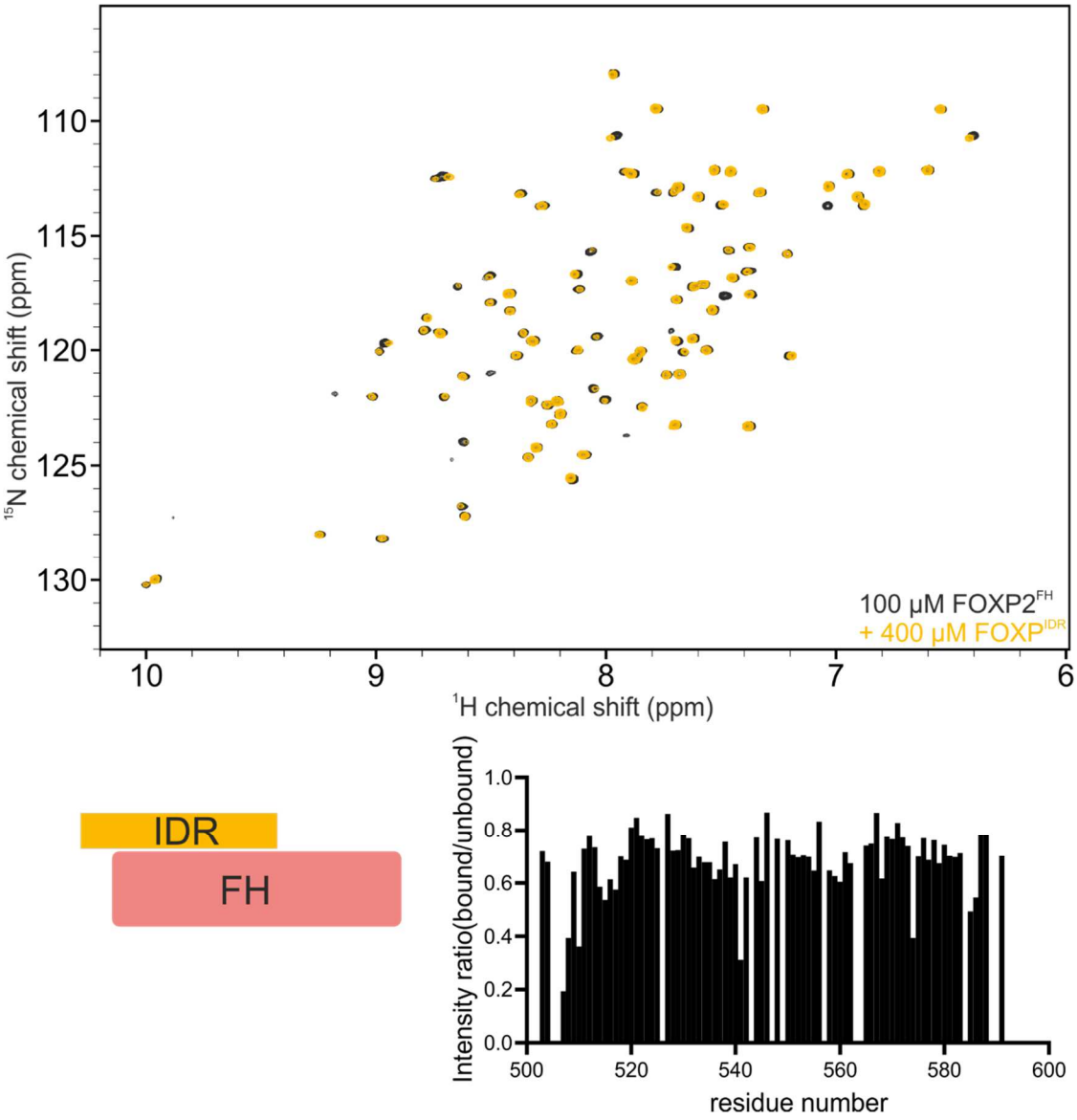

Fig.3.19: Above: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH} (black, 100µM), titration with unlabeled FOXP2^{IDR} (yellow, 400µM). Below: Right: intensity plot indicating the affected residues.

To test if β-catenin and the FOXP2 FH domain are still interacting, when the FOXP2 IDR with α-helical propensity is deleted, a construct of FOXP2^{IDR} lacking residue 264-272 was designed, further called FOXP2$^{IDR\Delta helix}$. First, ^{15}N FOXP2$^{IDR\Delta helix}$ was used to record a ^{1}H ^{15}N HSQC reference spectrum, then increasing amounts of unlabeled β-catenin$^{141-305}$ (Fig. 3.20) or unlabeled FOXP2^{FH} (Fig. 3.21) were added. In both cases, no CSPs and signal broadening were detected, proving, that the region with α-helical propensity is indeed crucial for the interaction with both partners.

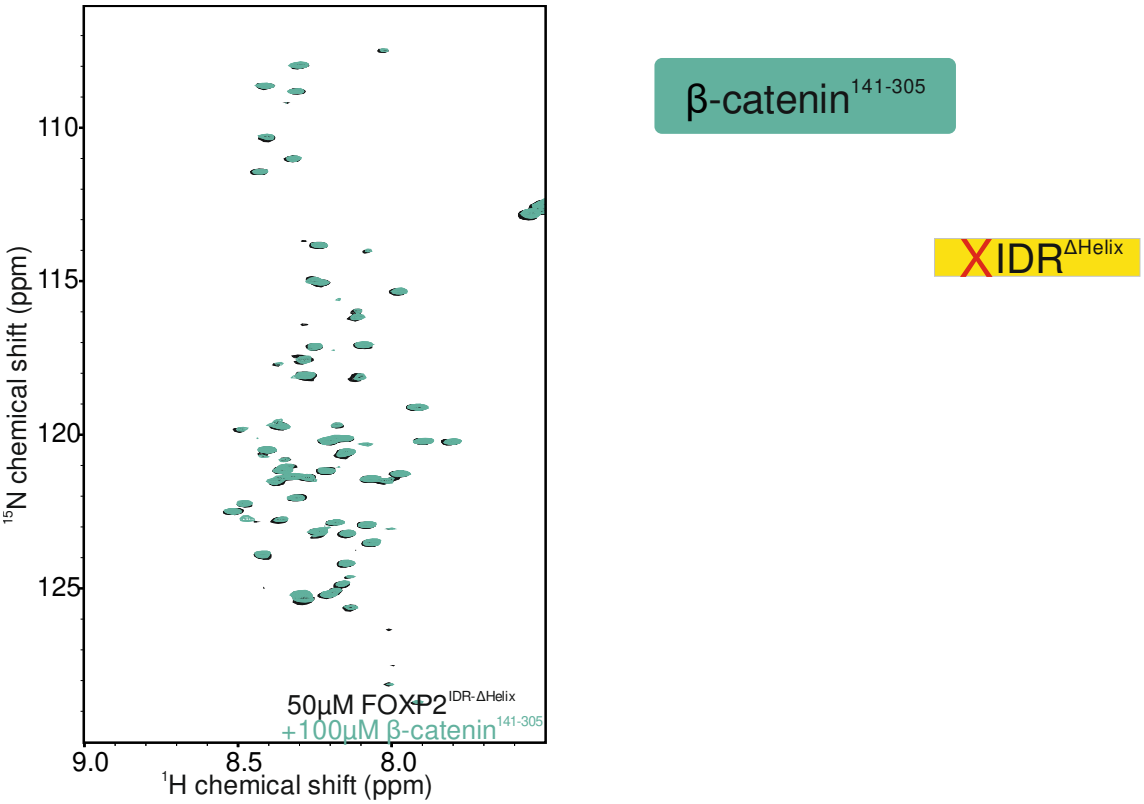

Fig.3.20: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2$^{IDR\Delta helix}$ (black, 50μM), titration with β-catenin$^{141-305}$ (green, 100 μM).

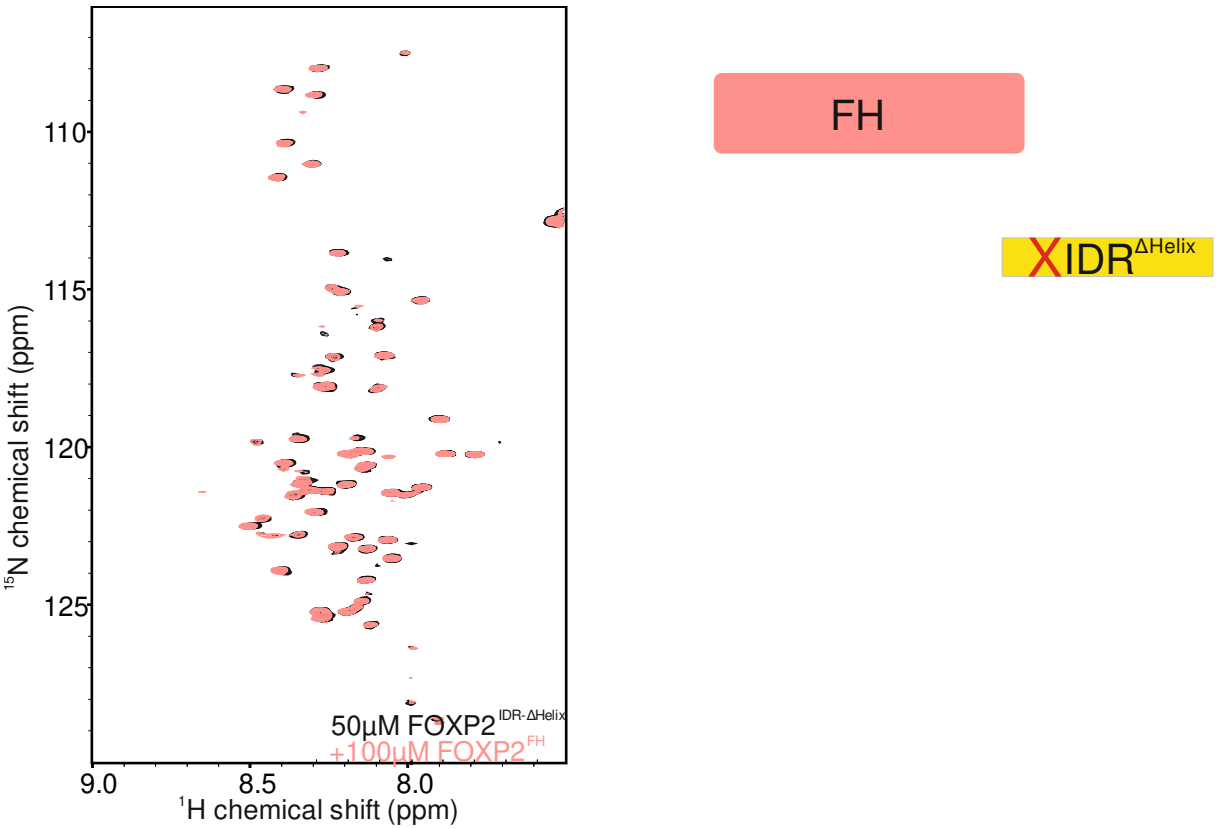

Fig.3.21: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2$^{IDR\Delta helix}$ (black, 50μM), titration with FOXP2^{FH} (coral, 100 μM).

To test, if the intramolecular interaction is affecting the binding to β-catenin, a construct containing the FH linked to the region with α-helical propensity (residue 264-272) and separated with a GS-linker was designed, further on called FOXP2$^{FH\text{-}IDR}$. This construct should facilitate the intramolecular interaction given their artificial proximity. Unfortunately, the expression of full-length FOXP2 was not successful. A reference ^{1}H ^{15}N HSQC of ^{15}N labeled FOXP2$^{FH\text{-}IDR}$ was recorded and β-catenin^{FL} titrated in increasing amounts. CSPs of ^{1}H ^{15}N FOXP2$^{FH\text{-}IDR}$ cross-peaks and signal broadening were observed proofing the interaction between FOXP2$^{FH\text{-}IDR}$ and β-catenin^{FL} (Fig. 3.22), thus the intramolecular interaction between the IDR and the FH domain of FOXP2 is not competitive with FOXP2 binding to β-catenin^{FL}.

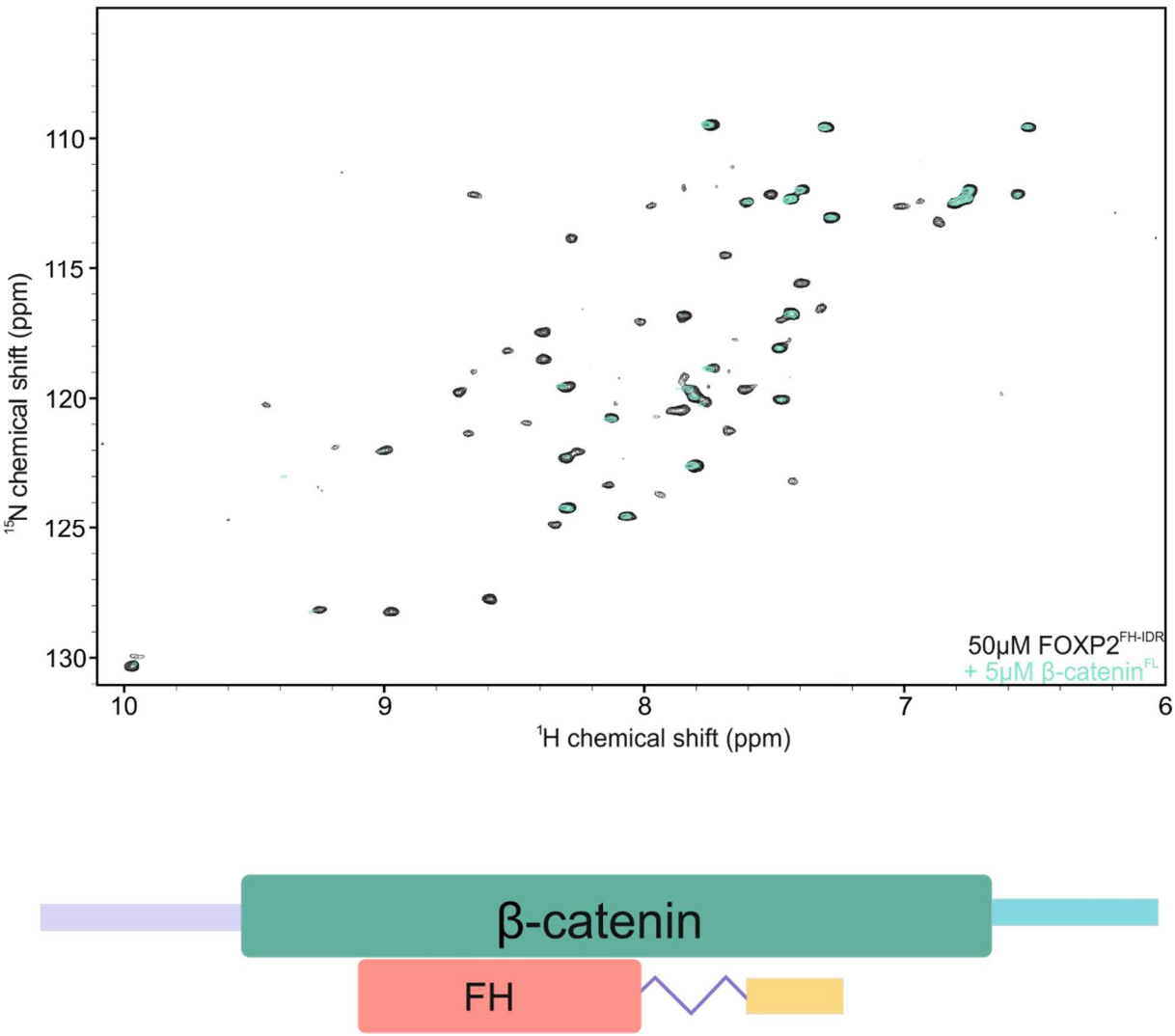

Fig.3.22: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2$^{FH\text{-}IDR}$ (black, 50µM), titration with unlabeled β-catenin^{FL} light-green, 5µM).

As β-catenin^{FL} is often aggregating upon interaction, the shorter, flexible C-terminus of β-catenin was used in order to investigate the binding between this part and the FH domain of FOXP2 in presence of the region with α-helical propensity. After recording a reference ^{1}H ^{15}N HSQC of FOXP2 $^{FH-IDR}$ increasing amounts of β-catenin$^{666-781}$ where added (Fig. 3.23). Signal broadening of ^{1}H ^{15}N FOXP2^{FH-IDR} cross-peaks were observed proofing the direct interaction between the FH-IDR construct and the C-terminus of β-catenin.

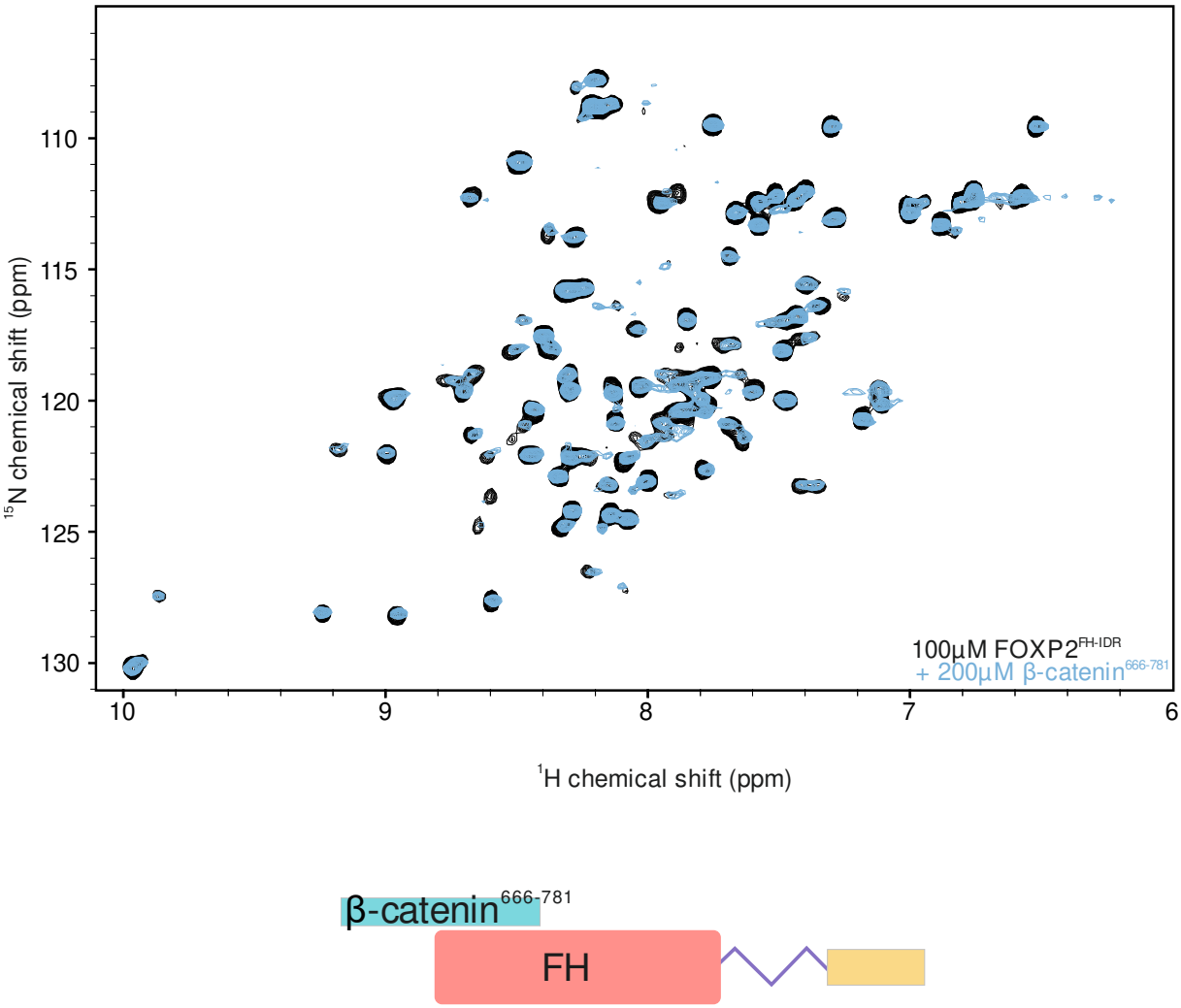

Fig.3.23: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH-IDR} (black, 50µM) titration with unlabeled β-catenin$^{666-781}$ (light-blue, 200µM).

Both, β-catenin and the IDR are binding to the same residues of the FH, thus a competitive network is possible. Nevertheless, these results indicate, that the interaction between FOXP2 and β-catenin is still occurring when the FOXP2 FH domain is bound to the IDR region. Therefore, precise binding constant in both states should be obtained in order to get more detailed about possible competitive or synergic binding modes.

To this end ITC was used. ITC is a powerful biophysical method to determine binding affinities (K_D), binding enthalpy (ΔH) and stoichiometry (N) on the basis of the thermodynamic behavior of biomolecule interactions. While ITC is very sensitive to strong and semi-strong interactions, it loses its sensitivity when it comes to weaker interactions. Thus, interactions with a K_D of higher µM range can be detected by NMR, but sometimes not by ITC.

First the interaction between FOXP2^{IDR} versus β-catenin$^{141-305}$ was tested. Therefore, 100µM concentrated β-catenin$^{141-305}$ was step-wise titrated in 10µM concentrated FOXP2^{IDR} at a temperature of 10°C leading to temperature changes upon the endo- or exothermic interaction (Fig. 3.24).

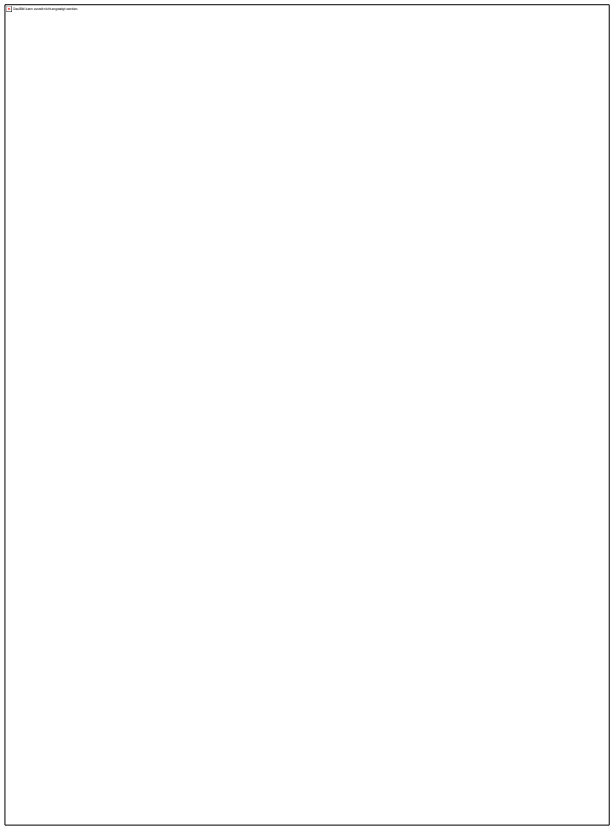

Fig.3.24: Isothermal calorimetry of the binding between β-catenin$^{141-305}$ (100µM) and FOXP2^{IDR} (10µM) at 10°C.

From this experiment a K_D of 8µM and a N= 0.98 (stoichiometry) could be calculated, but these values should not be taken with caution, as the titration did not show plateaus on both ends of the curve meaning that the K_D and N could still vary. Next, a titration of 100µM FOXP2^{IDR} and 10µM β-catenin^{FL} at 10°C was not giving any signal (Fig. 3.25).

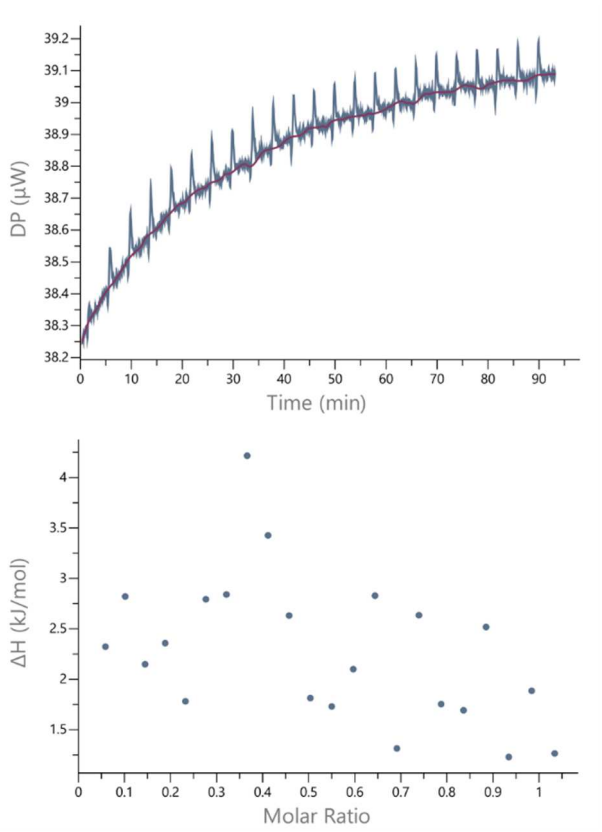

Fig.3.25: Isothermal calorimetry of the binding between FOXP2^{IDR} (100µM) and β-catenin^{FL} (10µM) at 10°C .

One explanation could be, that β-catenin^{FL} is sensitive to stirring and aggregates fast. Next, the interaction between the FH and the N- (Fig. 3.26) and C-terminus (Fig. 3.27) of β-catenin was investigated. In the titrations of FH (200µM) vs the N-terminus of β-catenin (20µM) at 10 °C no signal has been observed. After changing buffers, temperatures and concentrations no signal could have been detected. In this case the binding seem to be too weak to be detected with ITC. However, the titration between FH (200µM) and the C-terminus of β-catenin (20µM) at 10°C resulted in a binding signal with a K_D of 4µM and N of 1. These results indicate that the C-terminus of β-catenin binds stronger to the FH than the N-terminus of β-catenin.

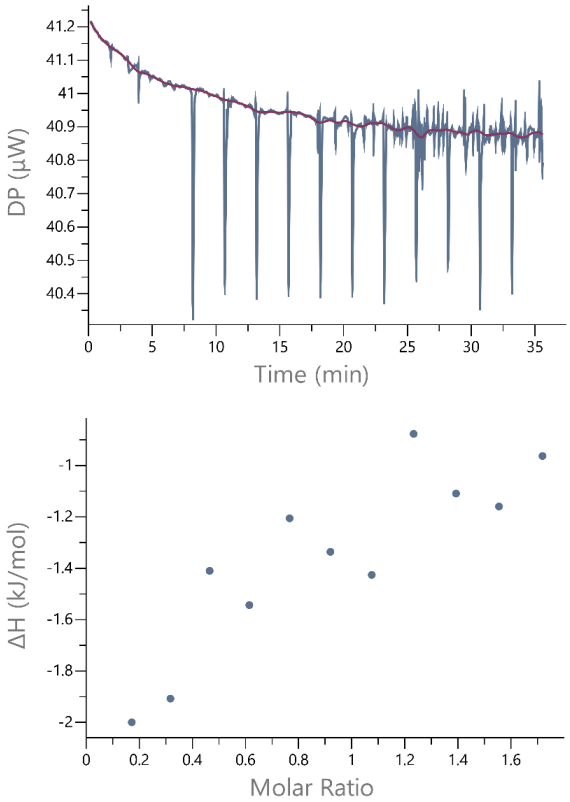

Fig.3.26: Isothermal calorimetry of the binding between β-catenin$^{1\text{-}140}$ (200µM) and FOXP2^{FH} (20µm) at 10°C.

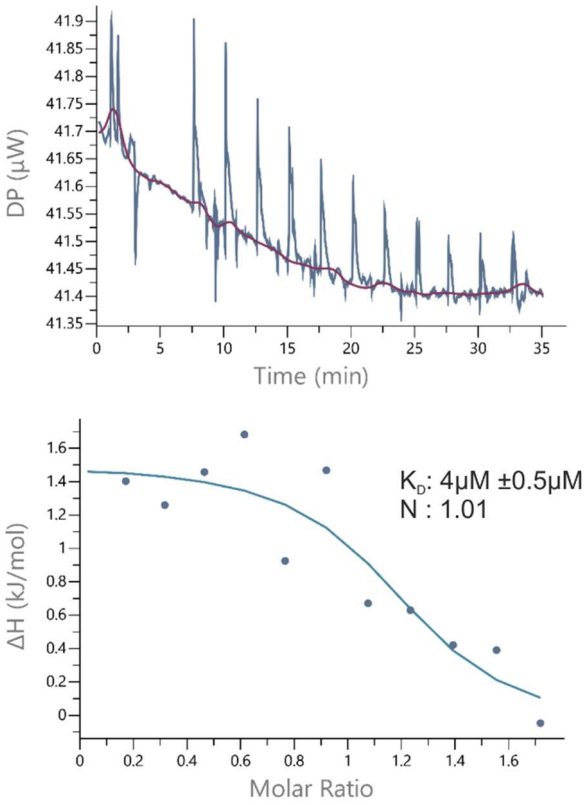

Fig.3.27: Isothermal calorimetry of the binding between β-catenin$^{666\text{-}781}$ (200µM) and FOXP2^{FH} (20µM) at 10 °C.

Then the FOXP2^{FH-IDR} was used to determine if there is a change of affinity to β-catenin$^{1-140}$ or β-catenin$^{666-781}$ compared to the binding between FOXP2^{FH} and those β-catenin constructs, which could reveal information about the molecular network between both proteins. By titrating β-catenin$^{1-140}$ (200µM) to FOXP2^{FH-IDR} (20µM) at 10°C no signal was detected (Fig. 3.28), thus, compared to the measurement of β-catenin$^{1-140}$ versus FOXP2^{FH} (Fig. 3.26), the presence of the IDR, and thus the possible interaction between FH and IDR does not change the interaction to the N-terminus of β-catenin.

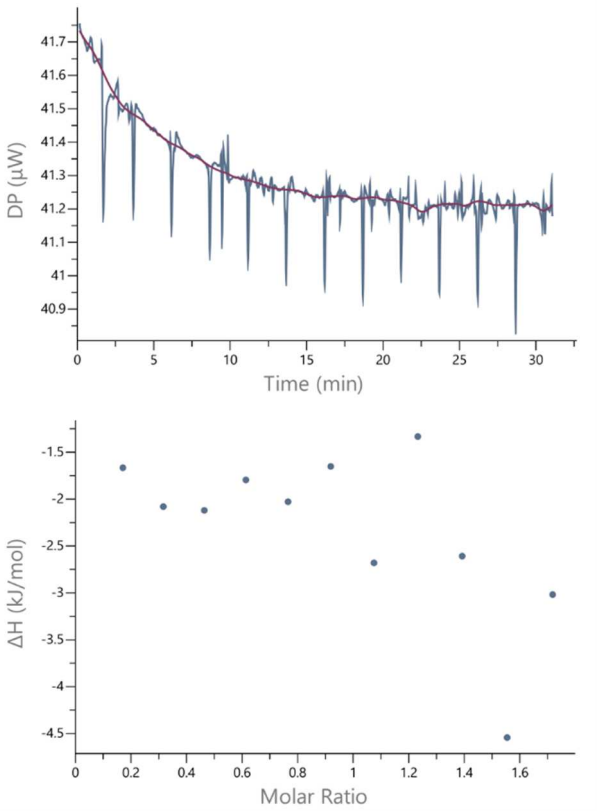

Fig.3.28: Isothermal calorimetry of the binding between β-catenin$^{1-140}$ (200µM) and FOXP2^{FH-IDR} (20µM) at 10°C.

Then, the interaction between FOXP2 FH-IDR and β-catenin$^{666-781}$ was investigated. By titrating β-catenin$^{666-781}$ (200µM) to FOXP2^{FH-IDR} (20µM) at 10 °C no signal was detected (Fig. 3.29) thus compared to the measurement of β-catenin$^{666-781}$ versus FOXP2^{FH} (Fig. 3.27), the presence of the IDR, and thus the possible interaction between FH and IDR does change the interaction to the C-terminus of β-catenin, as the interaction between FH and β-catenin$^{666-781}$ does not take place anymore.

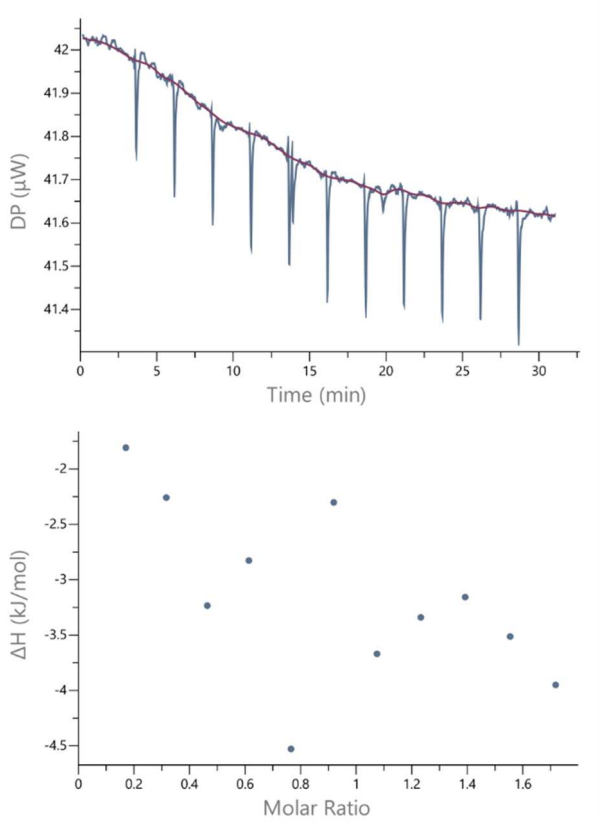

Fig.3.29: Isothermal calorimetry of the binding between β-catenin$^{666-781}$(200µM) and FOXP2^{FH-IDR} (20µM) at 10°C.

3.5. FOXP2 binds to DNA as dimer and is influenced by the intramolecular interaction

To investigate the DNA binding abilities of the FH and the influence of β-catenin and the intramolecular backfold NMR, SAXS, ITC and EMSA was used. The DNA target sequence was chosen from a FOXP2 target as already published[108]. First, the binding site of the DNA was localized on the FH domain using NMR to check if the interaction overlaps with the interaction site of β-catenin and the IDR. Thereby a ^{1}H ^{15}N HSQC of ^{15}N labeled FOXP2^{FH} was recorded as reference and then added increasing amounts of DNA to the sample (Fig. 3.30).

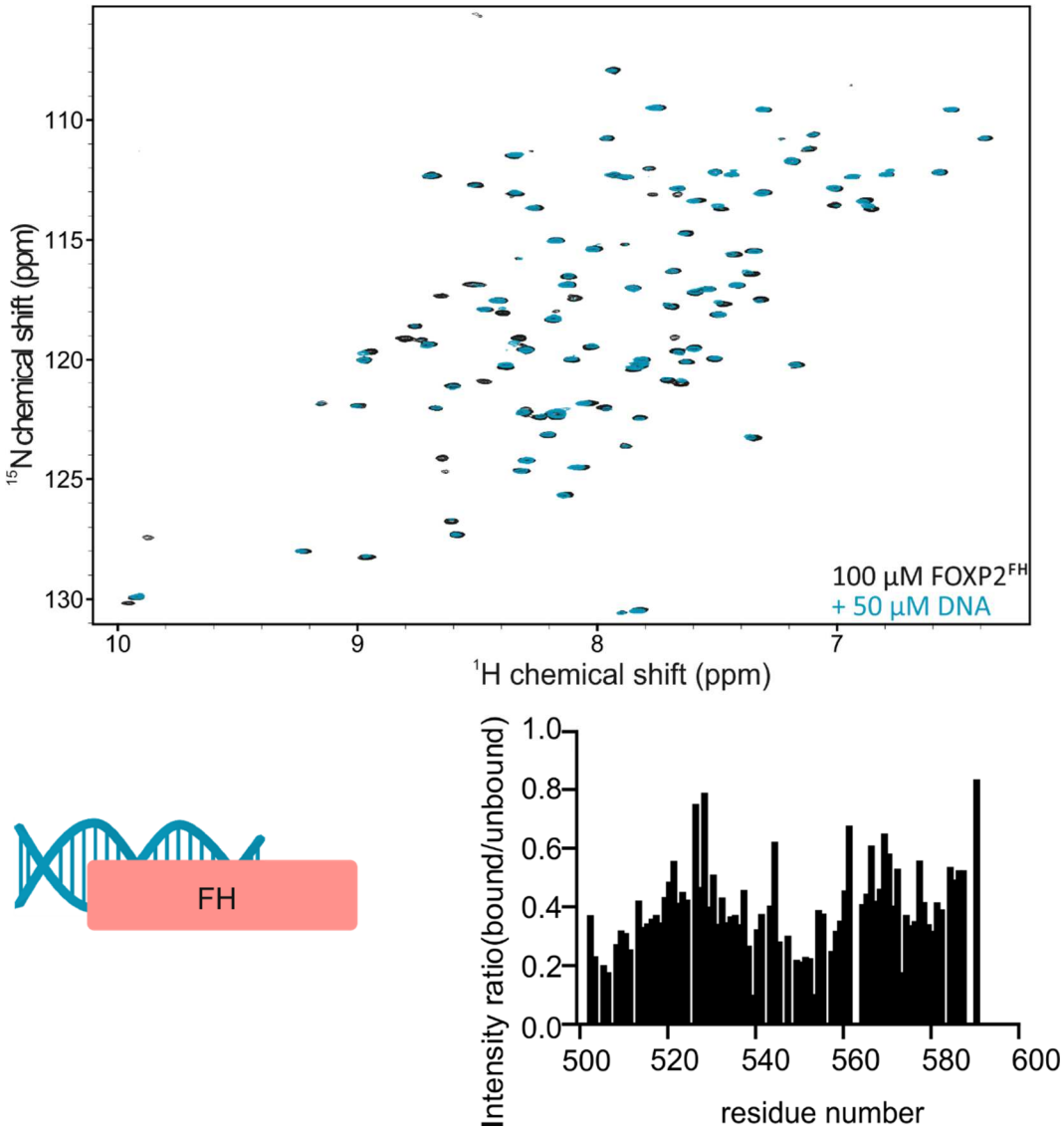

Fig.3.30: Above: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH} (black, 100µM), titration with DNA (blue, 50µM). Below: CSP plots indicating the affected residues.

Indeed, various ^{1}H ^{15}N cross peaks show CSPs and line broadening upon DNA addition. Using the FH assignment the binding site could be localized to similar residues which are affected by the N- and C-terminus of β-catenin. Not only those, but also much more residues were affected. These affected FH residues might not only come from the binding to the DNA, but also from the expected dimerization of the FH domain[77].

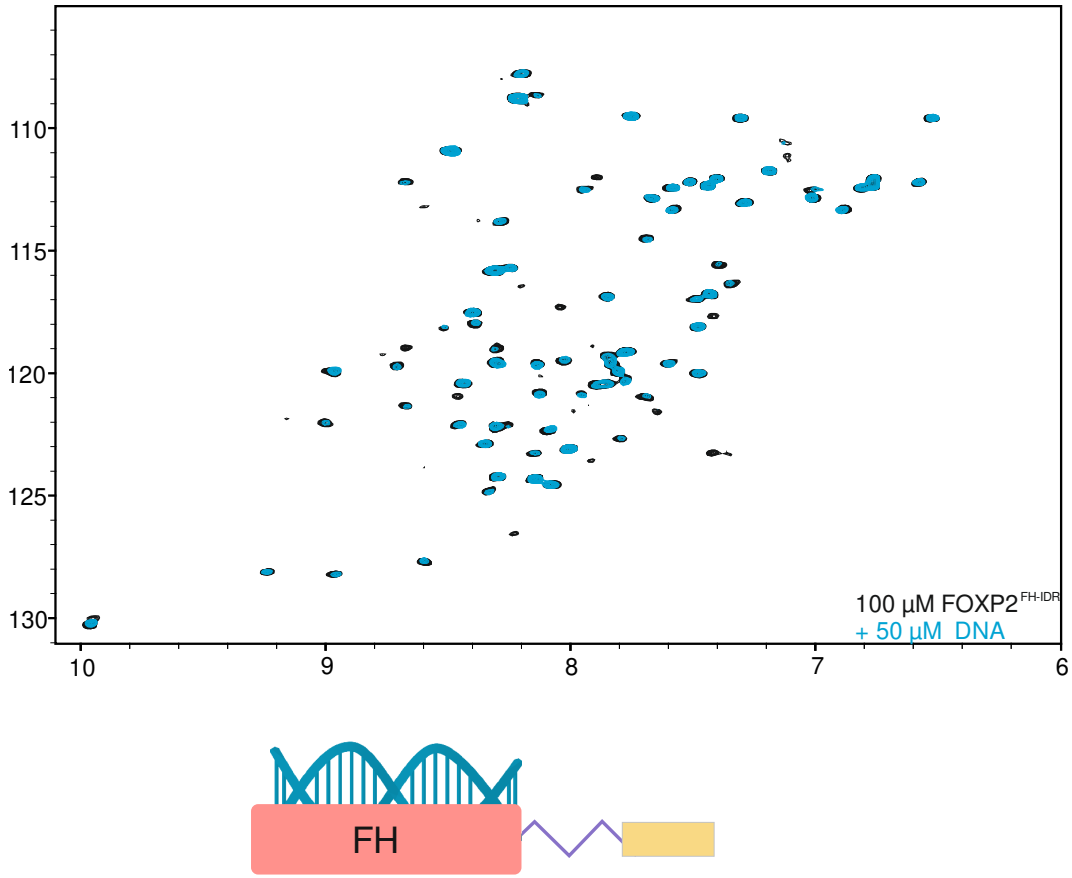

Fig.3.31: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{FH-IDR} (black, 100µM), titration with DNA (blue, 50µM).

Then the FOXP2^{FH-IDR} was investigated using NMR in order to detect differences in DNA binding compared to the FOXP2 FH domain without present IDR. By comparing the reference ^{1}H ^{15}N HSQC of the FH-IDR construct alone or titrated with DNA, various CSPs of ^{1}H ^{15}N cross-peaks of the FH-IDR could be detected proofing the interaction between the protein with the DNA (Fig. 3.31). Thus the presence of the IDR is not preventing the binding of DNA to the FH. Conclusions about affinity differences between both constructs to DNA cannot be made here as both constructs are different and thus the spectra differ.

To investigate the known dimerization with this DNA target, SAXS was used. Beside samples containing i) only FOXP2^{FH} and ii) only DNA, iii) equal amounts of FOXP2^{FH} and DNA and iv) two equivalent amounts of FOXP2^{FH} and one equivalent DNA were recorded in order to characterize the expected dimerization. The buffer-subtracted curves already indicated different intensities in the low-q region (Fig. 3.32), assuming that there is a difference of size between the FH alone and the FH + DNA.

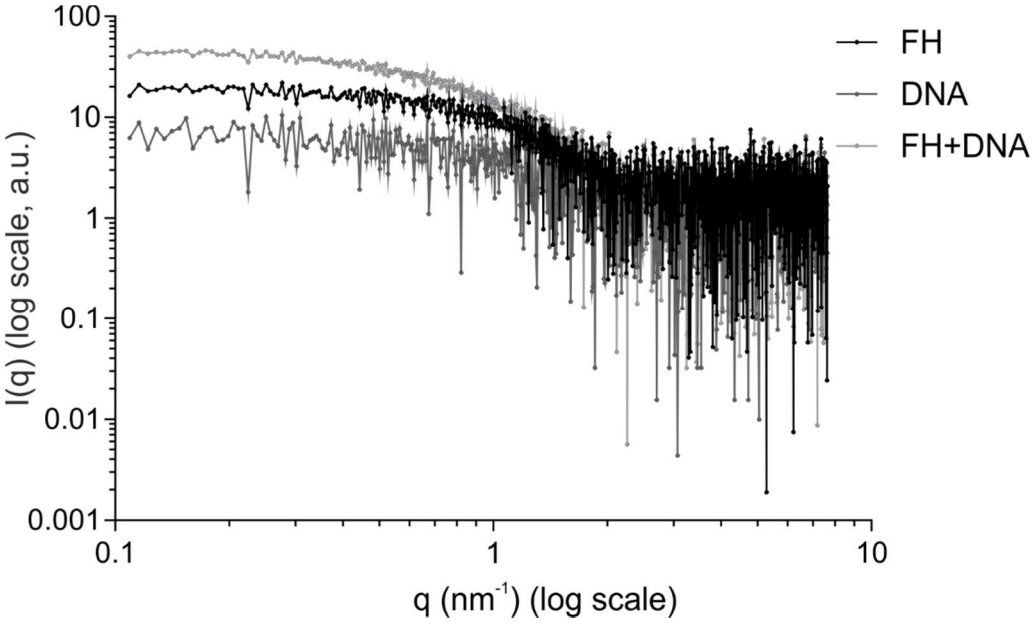

Fig.3.32: Buffer-subtracted raw-curve related to SAXS measurements of FH (225µM, black), DNA (225µM, grey), 1:1 ratio of FH:DNA (225µM:225µM, light grey).

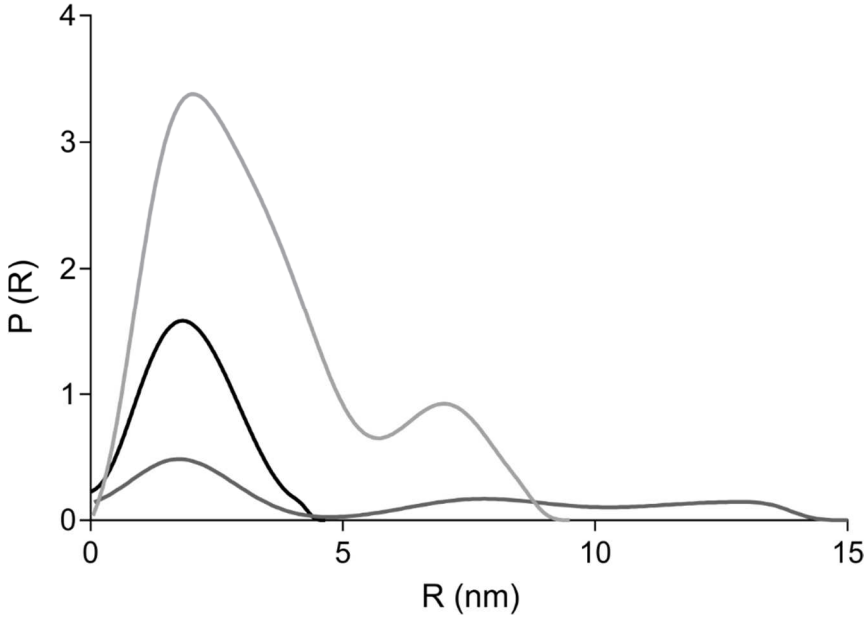

Fig.3.33: p(r) curves related to SAXS measurements of FH (black), DNA (grey), 1:1 ratio of FH:DNA (light-grey).

The p(r)-curve then confirmed that the mix between FOXP2 FH and DNA indeed leads to an increased size of the FH domain upon DNA addition (Fig. 3.33). The porod-volume then gave insights in the molecular weight of the different samples, which resulted in 10kDa for the FH alone, 31kDa for the 2:1 mix between FH:DNA, 29kDa for the 1:1 mix and 13kDa for the DNA. These data indicate, that the FH is indeed forming a dimer upon DNA binding or that two DNA molecules bind to one FH molecule.

In order to investigate the binding affinities, especially in correlation with β-catenin and the IDR-induced backfold, ITC was used. A previous study already investigated the DNA-binding affinities of the FH of FOXP2 via ITC[124]. Thereby a K_D of 67nM (±4.3nM) as binding affinity and a stoichiometry of N= 0.96 was determined between FH and DNA[124]. As the same DNA-sequence and same buffer was used, similar results were expected. By titrating the FOXP2^{FH} (100µM) in less-concentrated DNA (10µM) at 10°C a strong binding was detected (Fig. 3.34).

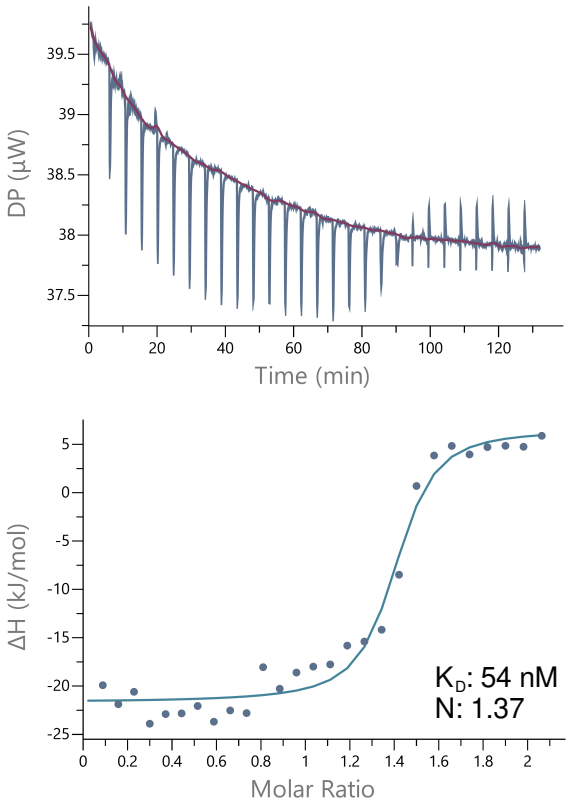

Fig.3.34: Isothermal calorimetry of the binding between FOXP2^{FH} (100µM) and DNA (10µM) at 10°C.

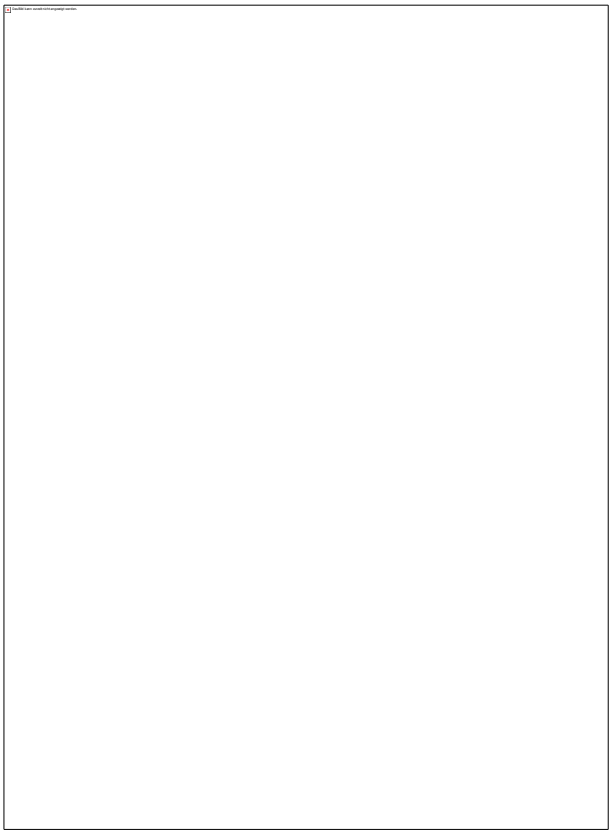

Fig.3.35: Isothermal calorimetry of the binding between FOXP2^{FH-IDR} (100µM) and DNA (10µM) at 10°C.

Our obtained K_D of 54nM (± 2.3nM) is similar to the recent published data. The N-value gives information about the stoichiometry and thus the number of molecules bound to the interaction partner. The obtained N-value was 1.49, which does not directly indicate dimerization of the FH upon DNA binding and thus not proofing our SAXS data ultimately.

In order to test if the IDR has an influence on the DNA binding and thus represent a regulatory element, the construct FOXP2^{FH-IDR} (100µM) was used and titrated with the DNA (10µM) at 10°C (Fig. 3.35). Interestingly, the presence of the IDR result to an increased affinity of the FH domain of FOXP2 to the DNA, as we obtained a K_D of 9nM (±1nM), which suggests that the intramolecular interaction between the FH domain and the IDR enhances FOXP2 ability to interact with DNA. The N-value did not differ significantly from the previous measurement (N = 1.44).

Next, the effect of β-catenin on the DNA-binding to the FOXP2^{FH} domain was tested, as its role as co-activator for FOXP2 activity is unknown. First, ITC was used in order to compare DNA binding of the FOXP2^{FH} in presence of β-catenin^{FL} (Fig. 3.36).

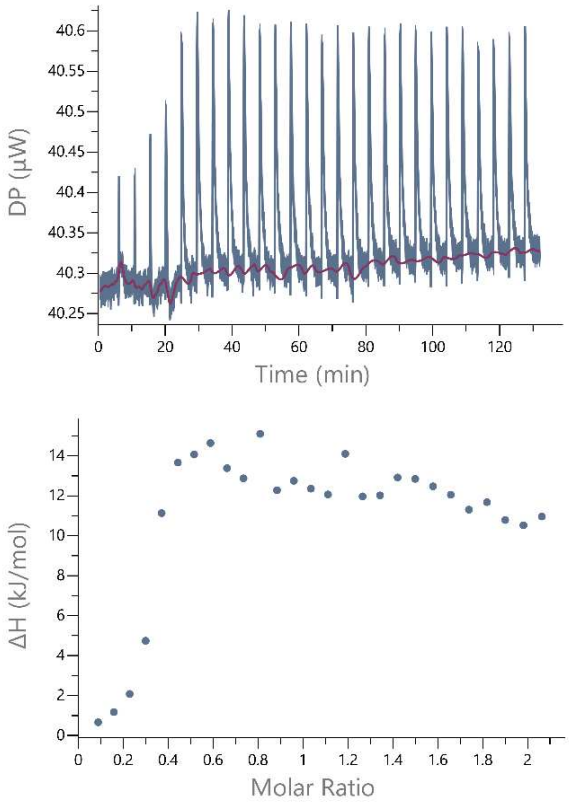

Fig.3.36: Isothermal calorimetry of the binding between DNA (100µM) and FOXP2^{FH}(10µM) + β-catenin^{FL}(20mM) at 10°C.

Therefore DNA (100µM) was titrated to FOXP2^{FH} (10µM) at same concentrations as above in presence of 2-fold excess of β-catenin^{FL} (20µM) in order to ensure saturation of the binding between FOXP2^{FH} and β-catenin. No signal was detected for this titration, β-catenin^{FL} resulted in aggregation of the sample in the cell. Thus, the effect of the smaller constructs of β-catenin were tested on DNA binding, as they are more stable and therefore the results more reliable than with β-catenin^{FL}. By titrating the DNA (100µM) in FOXP2^{FH} (10µM) plus a 5-fold excess of β-catenin$^{1-140}$ (50µM) at 10°C, no significant signal could be detected. In order to complete the data set DNA (100µM) was titrated in FOXP2^{FH} (10µM) plus a 5-dolf excess of β-catenin$^{666-781}$ (50µM). Here no signal was detected neither. This indicates, that the interaction of N– and C-terminus of β-catenin to the FOXP2^{FH} abrogates binding of the FOXP2^{FH} to the DNA suggesting that β-catenin is a transcriptional repressor of FOXP2.

In order to understand the role of the IDR backfold in this process, DNA (100µM) was titrated into FOXP2^{FH-IDR} (10µM) plus a 5-fold excess of β-catenin$^{141-305}$ (50µM) (Fig. 3.37). The obtained K_D of 19nM (±2.7nM) (N=1.67) was different compared to the ITC measurement of DNA into FOXP2^{FH-IDR} alone (Fig. 3.35). Here, the presence of β-

catenin slightly decrease the DNA binding affinity of the FOXP2$^{FH\text{-}IDR}$ construct as compared to the strong effect on the FOXP2^{FH} alone.

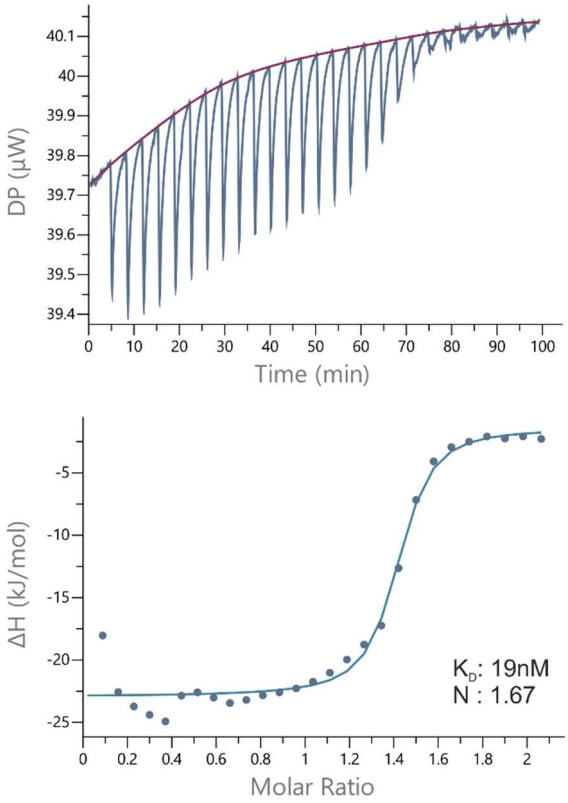

Fig.3.37: Isothermal calorimetry of the binding between DNA (100µM) and FOXP2$^{FH\text{-}IDR}$ (10µM) + β-catenin$^{141\text{-}305}$ (30µM) at 10°C.

Another method to detect DNA binding affinities is EMSA. Recently a study published EMSA data with FOXP2^{FH} and DNA[125]. Therefor FAM-labeled DNA was used to be detectable with a fluorescence filter. FOXP2^{FH} was added in increasing amounts from 0µM to 200µM to the DNA and then separated via a bisacrylamid gel (Fig. 3.38, left graphic). This gel is not disrupting molecular complexes, thus changes in size can be detected. The results show, that upon FH addition, the amounts of free DNA decrease and a clear band at higher molecular weight appeared corresponding to a FH-DNA complex. At a FOXP2^{FH} concentration of 200µM most of DNA stuck in the well of the gel indicating a complex of very high molecular weight.

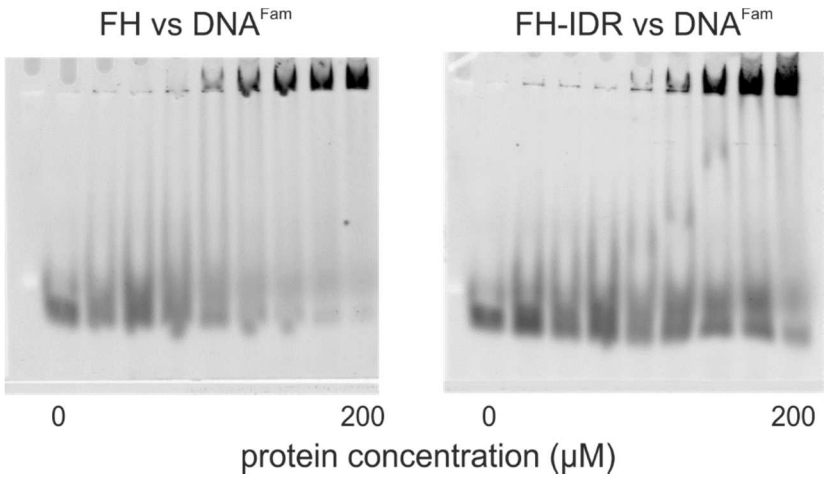

Fig.3.38: EMSA gel of fluorophore-labeled DNA and increasing amounts (0, 1.5, 3.1, 6.1, 12,3, 25, 50, 100, 200µM) of FOXP2^{FH} (left) and FOXP2^{FH-IDR} (right).

The same was repeated with the FOXP2^{FH-IDR} construct (Fig. 3.38, right graphic). Interestingly, the binding to the DNA seems to be less strong than with the FH alone, which does not support the previous ITC-data.

3.6 FOXP2 mediates transcription of Wnt target genes

Until now the complex regulation mechanisms of the transcription factor FOXP2 and its broad biological functions remain elusive. To determine possible regulatory elements which control FOXP2 activity and discover the biological role of β-catenin and the intramolecular interaction in FOXP2 function, we performed RNA-Sequencing (RNA-Seq) experiments.

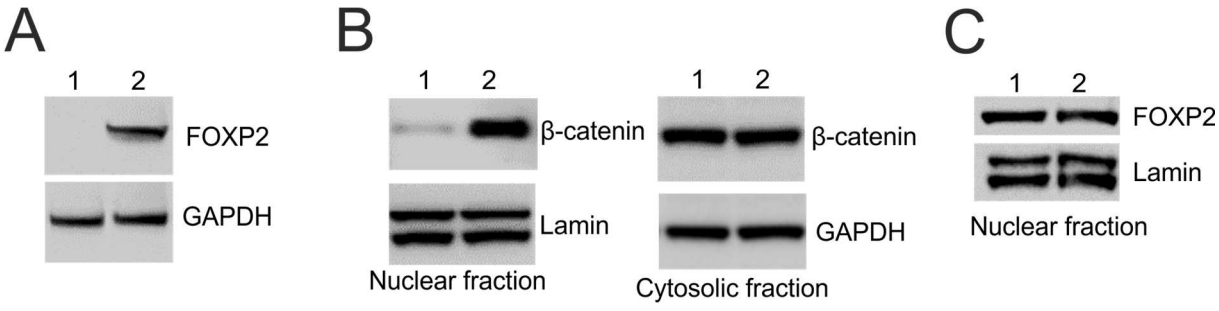

Fig.3.39: Western blot (carried out by Dr. Chintan Koyani, Medical University Graz) showing A) Lane 1 = untransfected U2OS cells, lane 2 = overexpression of FOXP2^{WT}, B) Lane 1 = without CHIR treatment, lane 2 = with CHIR treatment in both, cytosolic and nuclear fraction, C) Lane 1 = overexpression of FOXP2^{WT}, lane 2 = overexpression of FOXP2$^{\Delta\text{-helix}}$.

For the set-up U2OS cells were used, as they contain no endogenous FOXP2 and only cytoplasmic and thus inactive β-catenin (Fig. 3.39, cell culture and western blot carried out by Dr. Chintan Koyani from Medical University Graz). Six conditions were used with each 5 replicates (Tab. 3).

Condition	**Expression**
Control	Mock-transfected cells
FOXP2	FOXP2 WT overexpression
CHIR	CHIR treatment
$FOXP2^{\Delta\text{-helix}}$	FOXP2 $FOXP2^{\Delta\text{-helix}}$ overexpression
FOXP2 + CHIR	FOXP2 WT overexpression + CHIR treatment
$FOXP2^{\Delta\text{-helix}}$ + CHIR	FOXP2 $FOXP2^{\Delta\text{-helix}}$ overexpression + CHIR treatment

Tab.3: overview of conditions used for RNA-Seq in U2OS cells.

The overexpression of FOXP2 (Fig. 3.39A) and the equal expression of FOXP2 wildtype (WT) and a construct lacking the region with α-helical propensity ($FOXP2^{\Delta\text{-helix}}$) was confirmed by western blot (Fig. 3.39C).

The chemical CHIR was used to translocate the cytoplasmic β-catenin in the nucleus of the U2OS cells in order to affect FOXP2 activity. The successful translocation was proven with western blot (Fig. 3.39B). Functional enrichment analysis was performed for all differentially expressed genes (DEGs) (FC >2/<-2, p-value <0.05). These data resulted in significant enrichment of upregulated and downregulated DEGs. By comparing the control with the cells overexpressing FOXP2, 3054 genes were found to be significantly upregulated and 4555 genes to be downregulated (Fig. 3.40).

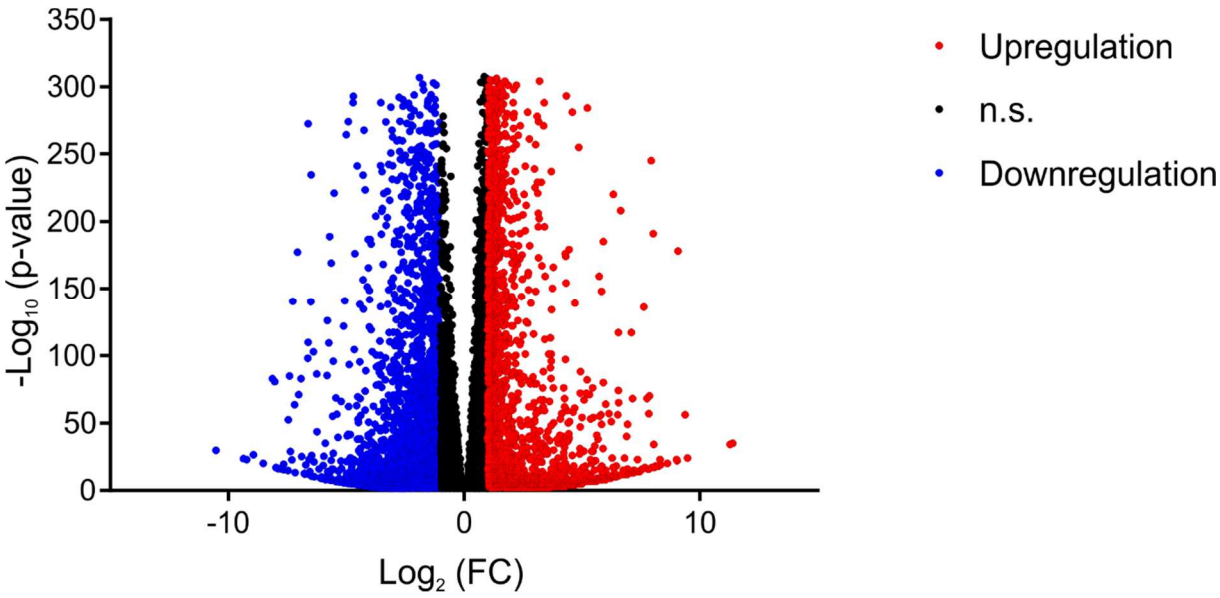

Fig.3.40: Volcano plot displaying the up-(red) and downregulated (blue) DEGs of FOXP2 overexpression compared to control.

In order to make functional interpretation for the gene expression changes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was performed based on Fisher's extract test using DAVID 3.8[126]. The GO database is currently the most widely-used gene annotation system for gene functions and products and provide a better understanding of the links between genes and diseases[127]. The KEGG database, on the other side, combines genetic information with functional information and can thus be used to understand relationships between genes and enriched pathways[128]. 3095 upregulated DEGs were found in 29 significant KEGG pathways. 'PI3K-Akt signaling pathway' was the pathway with most genes involved (51 genes) (Tab.4).

KEGG pathways	Upregulated	Downregulated
	Ribosome biogenesis in eukaryotes	Axon guidance
	TNF signaling pathway	ECM-receptor interaction
	NF-kappa β signaling pathway	Basal cell carcinoma
	Cytokine-cytokine receptor interaction	Cell adhesion molecules (CAMs)
	RNA transport	Arrythmogenic right centricular cardiomyopathy
	Legionellosis	Pathways in cancer
	Apoptosis	Hypertrophic cardiomyopathy
	NOD-like receptor signaling pathway	Calcium signaling pathway
	Hematopoietic cell lineage	Hippo signaling pathway
	MAPK signaling pathway	Dilated cardiomyopathy
Biological functions		
	rRNA processing	Homophilic cell adhesion
	Inflammatory response	Cell adhesion
	Ion transport	Extracellular matrix organization
	Response to liposaccharides	Nervous system development
	Sodium ion transmembrane transport	Synapse assembly
	Maturation of SSU-rRNA	Calcium-dependent cell-cell adhesion
	Neurotransmitter transport	Heart development
	Cellular response to liposaccharide	Negative regulation of gene expression
	Ion transmembrane transport	Glycosaminoglycan catabolic process
	Ribosomal large subunit biogenesis	Axon guidance

Tab.4: List of most significant up- and downregulated KEGG pathways and biological functions upon FOXP2 overexpression compared to control.

Additionally, 136 significant biological functions were found. The biological function with most genes was ‘signal transduction’ (152 genes). 4555 downregulated genes were associated with 31 significant pathways (KEGG) and 222 biological functions. The pathway with most genes involved was ‘pathways in cancer’ (76 genes) and the biological function with most genes was ‘cell adhesion’ (130 genes). Interestingly, in the gene set of the downregulated genes Wnt-related GO-terms appeared in five biological functions as ‘Wnt signaling pathway’ (p-value: 0.0013), ‘negative regulation of Wnt signaling pathway’ (p-value: 0.0083), ‘negative regulation of canonical Wnt signaling pathway’ (p-value: 0.021), ‘Wnt signaling pathway, calcium modulating

pathway' (p-value: 0.033) and 'positive regulation of Wnt signaling pathway, planar cell polarity pathway' (p-value: 0.058). Additionally, it was found in the Top 15 of the KEGG pathway analysis leading to a strong effect of FOXP2 overexpression on Wnt-pathways. To get a more specific idea of the effect of FOXP2 on the Wnt-pathway the expression of 25 known Wnt targets in the RNA-Seq data set was visualized in a heatmap (Fig. 3.38). FOXP2 overexpression significantly changes the expression of 17 observed Wnt targets giving a first clue about the regulation mechanisms of FOXP2. Visualization in a heatmap shows the effect of FOXP2 overexpression on those genes (Fig. 3.41).

To validate the gene expression changes from the RNA-Seq data real-time q-PCR was performed on a list of few Wnt targets. As expected, the expression levels of the selected genes were similar to the RNA-Seq data confirming the previous results (Fig. 3.42).

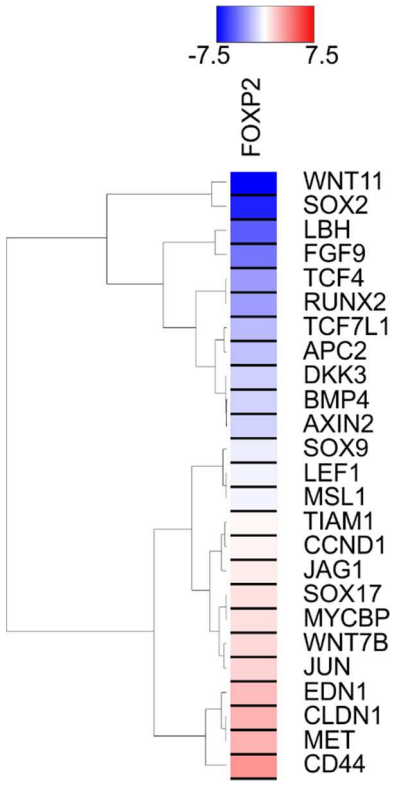

Fig.3.41: Heatmap of Wnt genes changed by FOXP2 overexpression compared to control, colors indicate normalized log_2 FC values, negative (blue), positive (red), neutral (white).

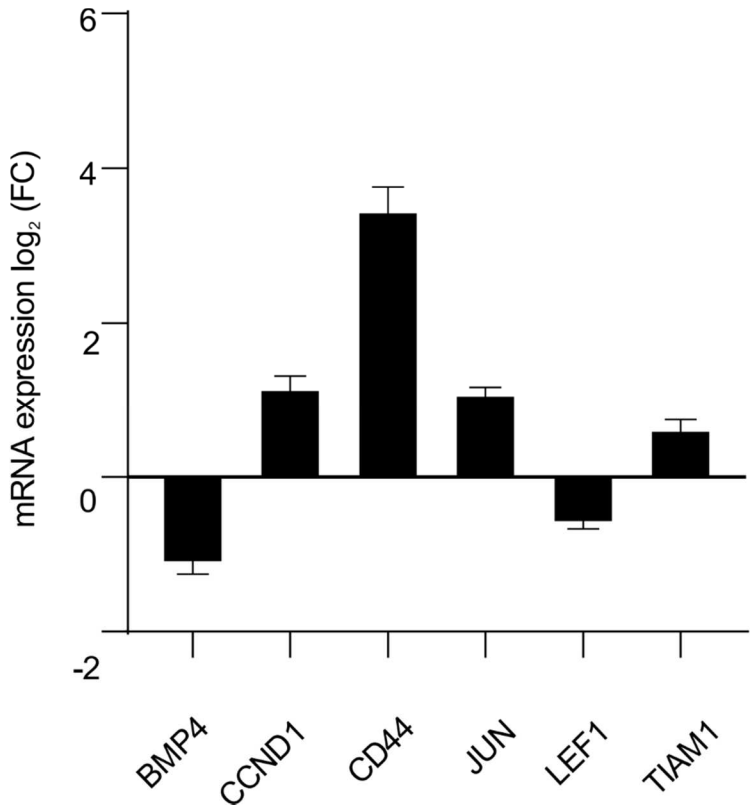

Fig.3.42: q-PCR quantification of Wnt genes changed by FOXP2 overexpression compared to control.

In order to assess the role of the intramolecular interaction between the IDR with α-helical propensity and the FH of FOXP2 and its effect on the regulation of its function we performed RNA-Seq analysis on cells expressing a FOXP2 construct lacking the α-helix element in this IDR (further called FOXP2$^{\Delta\text{-helix}}$). By comparing the genes changed between FOXP2 fulllength compared to control and FOXP2$^{\Delta\text{-helix}}$ compared to control various differences were found between both data sets (Fig. 3.43).

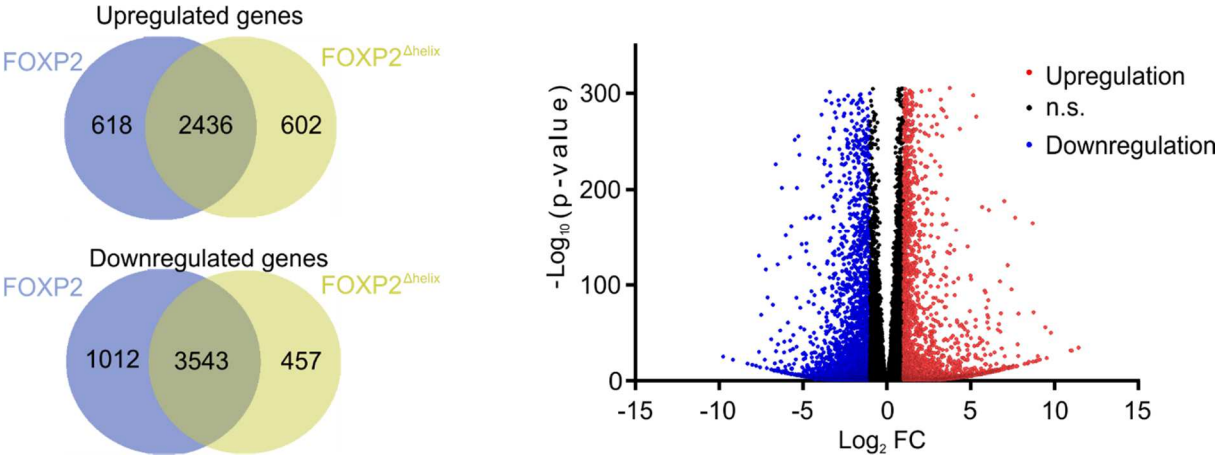

Fig.3.43: Left: Venn diagram showing the differences between FOXP2 and FOXP2$^{\Delta\text{-helix,}}$ both compared to the control. Right: Volcano plot showing significantly up-(red) and downregulated (blue) genes compared to control.

Next, KEGG-pathway and GO term biological function analysis was performed on this data set (Tab.5). Interestingly, the biological function with most upregulated genes was 'signal transduction' (166 genes involved, 8,6e-6). Compared to the wildtype FOXP2, FOXP2$^{\Delta\text{-helix}}$ showed only once the Wnt signaling pathway in the significant different KEGG pathways and biological functions (Tab.5).

KEGG pathways	**Upregulated**	**Downregulated**
	Ribosome biogenesis in eukaryotes	Pathways in cancer
	MAPK signaling pathway	ECM-receptor interaction
	TNF signaling pathway	Axon guidance
	NF-kappa B signaling pathway	Basal cell carcinoma
	Cytokine-cytokine receptor interaction	Focal adhesion
	Mineral absorption	ABC transporter
	Transcriptional misregulations in cancer	Arrythmogenic right ventricular cardiomyopathy (ARVC)
	Hematopoietic cell lineage	PI3K-Akt signaling pathway
	Bile secretion	Dilated cardiomyopathy
	Legionellosis	Hypertrophic cardiomyopathy (HCM)
Biological functions		
	rRNA processing	Homophilic cell adhesion
	Angiogenesis	Cell adhesion
	Maturation of LCU-rRNA	Nervous system development
	Signal transduction	Extracellular matrix organization
	Mitochondrial translational elongation	Synapse assembly
	Response to cAMP	Skeletal system development
	Response to cytokine	Chemical synaptic transmission
	Mitochondrial translational termination	Calcium-dependent cell-cell adhesion
	Apoptotic mitochondrial changes	Axon guidance
	Mitochondrion organization	Negative regulation of endopeptidase activity

Tab.5: List of most significant up- and downregulated KEGG pathways and biological functions upon FOXP2$^{\Delta\text{-helix}}$ overexpression compared to control.

66% of 3656 upregulated genes in total were similar between both constructs. 618 genes were unique for FOXP2 full-length and 602 genes were unique for the FOXP2 Δα-helix construct.

70% a total of 5012 genes, which were downregulated, were similar in both conditions. 1012 genes were unique for FOXP2 full-length and 457 genes were unique for FOXP2$^{\Delta\text{-helix}}$. The heatmap in Fig. 3.44 shows the effect of FOXP2$^{\Delta\text{-helix}}$ on a selection of Wnt-target genes. Not much differences could be detected between both data sets.

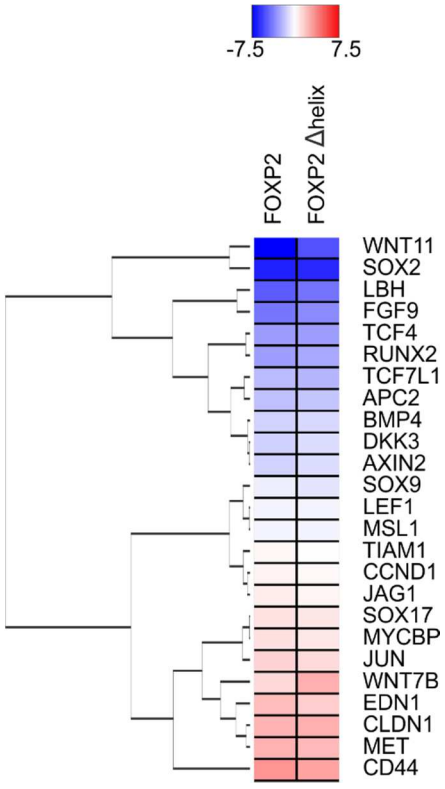

Fig.3.44: Heatmap of Wnt genes changed by FOXP2 and FOXP2$^{\Delta\text{-helix}}$ overexpression compared to control, colors indicate normalized $\log_2$ FC values, negative (blue), positive (red), neutral (white).

Q-PCR showed, that there are very slight differences between a few selected Wnt targets. (Fig. 3.45)

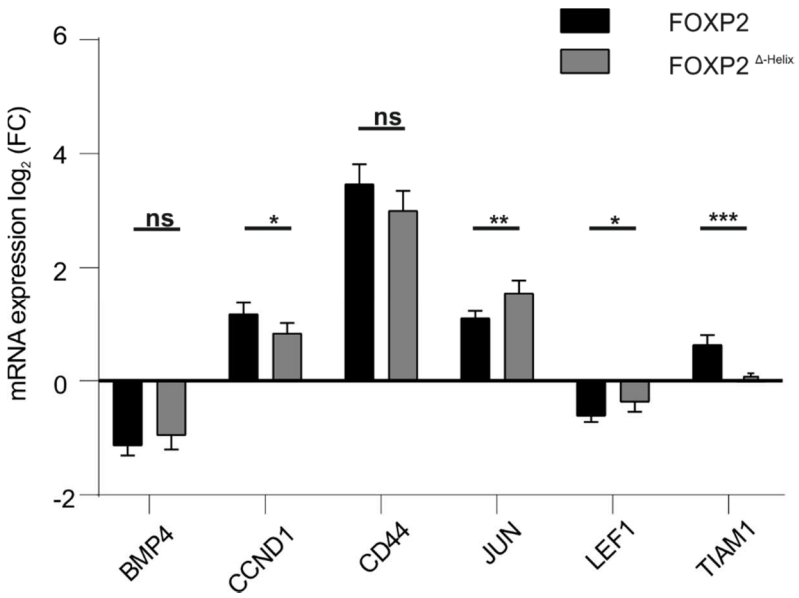

Fig.3.45: q-PCR quantification of Wnt genes changed by FOXP2 or FOXP2$^{\Delta\text{-helix}}$ overexpression compared to control.

As these comparisons do not show accurately the specific differences between FOXP2 and FOXP2$^{\Delta\text{-helix}}$ expressing cells, as they only show the differences of each condition compared to the control cells, the genes of the FOXP2 condition were directly compared to the FOXP2$^{\Delta\text{-helix}}$ condition to detect all differences between both (Fig. 3.46).

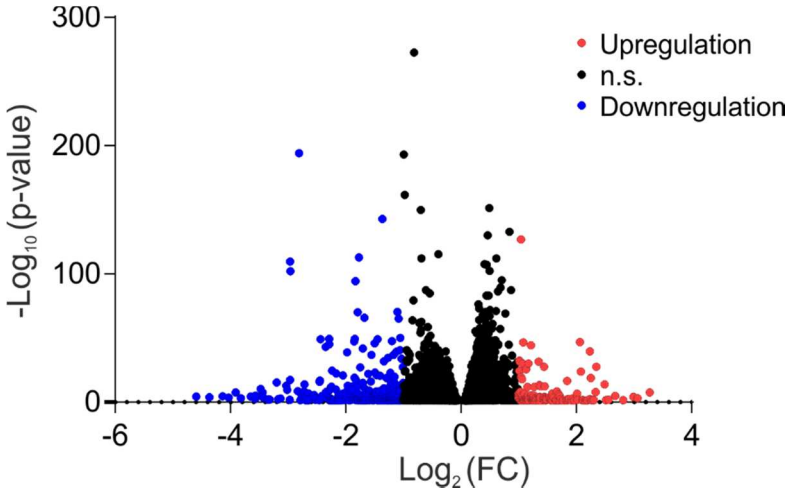

Fig.3.46: Volcano plot showing significantly up-(red) and downregulated (blue) genes of FOXP2$^{\Delta\text{-helix}}$ compared to FOXP2.

Here much less genes were found to be differentially expressed compared to the comparisons before. Thus, in comparison to FOXP2 156 genes were found to be significantly up- and 362 genes downregulated in the FOXP2$^{\Delta\text{-helix}}$ condition. For the

upregulated fraction 9 biological pathways and none KEGG pathways were found to be involved. The biological function with most genes was ‘immune response’. For the downregulated fraction 24 biological functions and 11 KEGG pathways were found to be involved. Interestingly, the biological function with most genes was ‘signal transduction’ and the KEGG pathway with most genes the ‘MAPK signaling pathway’. Thus more genes are downregulated with the FOXP2 construct lacking the α-helix, thus the backfold indeed seem to have a regulatory effect on FOXP2 activity either due to the intramolecular interaction or the interaction with FOXP2 binding partners.

3.7 FOXP2 is regulated by the Wnt signaling pathway

To investigate the effect of β-catenin on FOXP2 transcriptional activity, a part of the cells was treated with CHIR and compared to cells only expressing FOXP2 and cells only treated with CHIR. This chemical is inhibiting both kinases GSK3α and GSK3β, which are responsible for the cytoplasmic phosphorylation of β-catenin and thus play a crucial role in cellular β-catenin degradation. Using CHIR we were able to enrich β-catenin in the nucleus due to its accumulation in the cytoplasm (Fig. 3.39B).
In order to find clues about β-catenin associated regulation of FOXP2 function the gene sets of FOXP2, CHIR and FOXP2 + CHIR, each compared to the control were compared (Fig. 3.47).

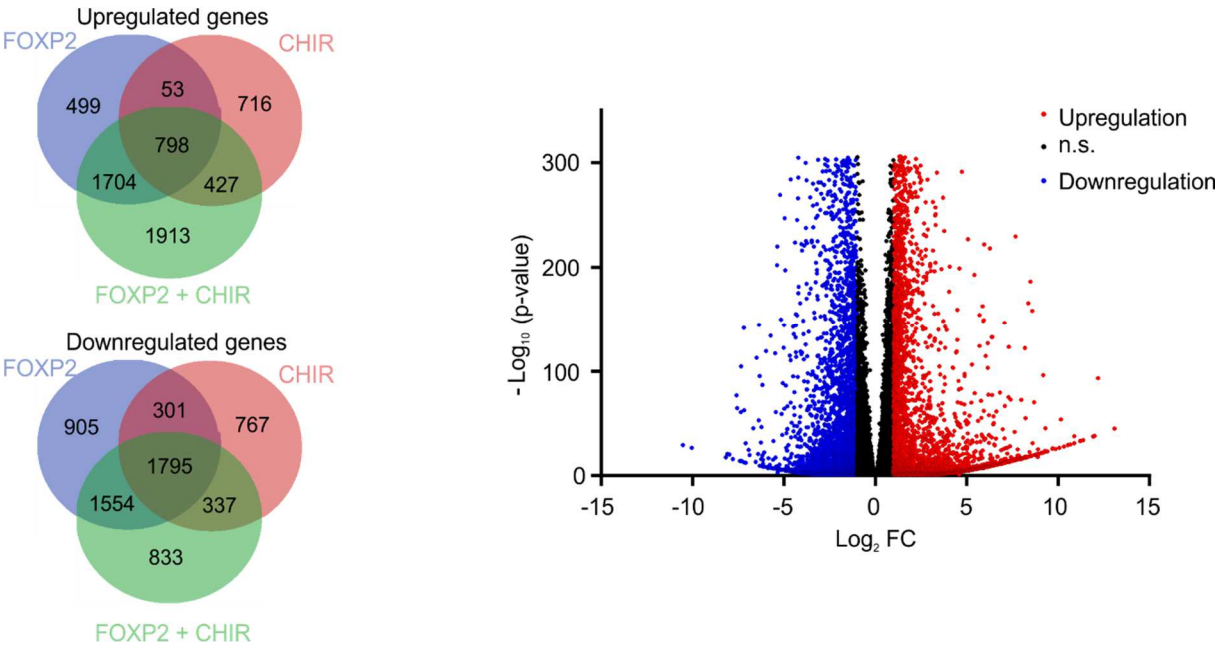

Fig 3.47: Left: Venn diagram showing the differences between FOXP2, CHIR and FOXP2 + CHIR, all compared to the control. Right: Volcano plot showing significantly up-(red) and downregulated (blue) genes of each condition compared to control.

Thereby in the upregulated fraction 798 genes were found to be similar in all data sets. 499 genes were unique for FOXP2, 716 unique for CHIR. More importantly 1913 genes were unique for FOXP2 + CHIR, thus could highlight β-catenin dependent regulation of FOXP2 function and vice-versa. 1704 genes were similar in both FOXP2 and FOXP2 + CHIR, leading to the suggestion, that those genes are FOXP2 dependent but β-catenin independent. The smallest fraction of genes is similar between FOXP2 and CHIR. In the downregulated part the fraction with most genes was found to be similar in all conditions (1795). The second largest fraction is similar between FOXP2 and FOXP2 + CHIR (1554). 905 genes were unique for FOXP2, 767 unique for CHIR and 833 unique for FOXP2 + CHIR. In Tab. 6 the most significant KEGG pathways and biological functions are displayed, derived from genes, which only show up in the FOXP2+CHIR condition. Interestingly, Wnt related pathways did not showed up significantly.

KEGG pathways	Upregulated	Downregulated
	Ribosome biogenesis in eukaryotes	ECM-receptor interaction
	Cytokine-cytokine receptor interaction	Arrhythmogenic right ventricular cardiomyopathy (ARVC)
	TNF signaling pathway	Focal adhesion
	NF-kappa B signaling pathway	Hypertrophic cardiomyopathy (HCM)
	Mineral absorption	Pathways in cancer
	MAPK signaling pathway	Dilated cardiomyopathy
	Epithelial cell signaling in *Heliobacter pylori* infection	PI3K-Akt signaling pathway
	Legionellosis	Axon guidance
	Rheumatoid arthritis	Rap1 signaling pathway
	Phagosome	Small cell lung cancer
Biological functions		
	rRNA processing	Homophilic cell adhesion
	Maturation of SSU-rRNA	Cell adhesion
	Inflammatory response	Extracellular matrix organization
	Sodium ion transmembrane transport	Synapse assembly
	Ribosome biogenesis	Nervous system development
	(cellular) Response to lipopolysaccherides	Calcium-dependent cell-cell adhesion
	Response to cAMP	Skeletal system development
	Ion transport	Response to drug
	Ribosomal large subunit biogenesis	Heart development
	Mitochondrial translational termination	Positive regulation of synapse assembly

Tab.6: List of most significant up- and downregulated KEGG pathways and biological functions upon FOXP2 overexpression + CHIR treatment compared to control.

In the heatmap the effect of the CHIR treatment on the conditions is visualized (Fig. 3.48). The condition with FOXP2 + CHIR is supposed to display the effect of the interaction between FOXP2 and β-catenin. The confirmation by q-PCR supported the previous RNA-Seq data (Fig. 3.49).

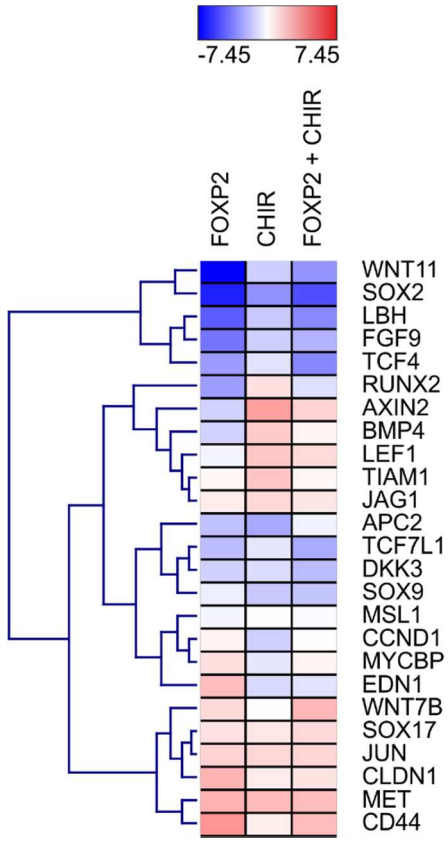

Fig.3.48: Heatmap of Wnt genes changed by FOXP2 overexpression, CHIR treatment and FOXP2 overexpression + CHIR treatment compared to control, colors indicate normalized log_2 FC values, negative (blue), positive (red), neutral (white).

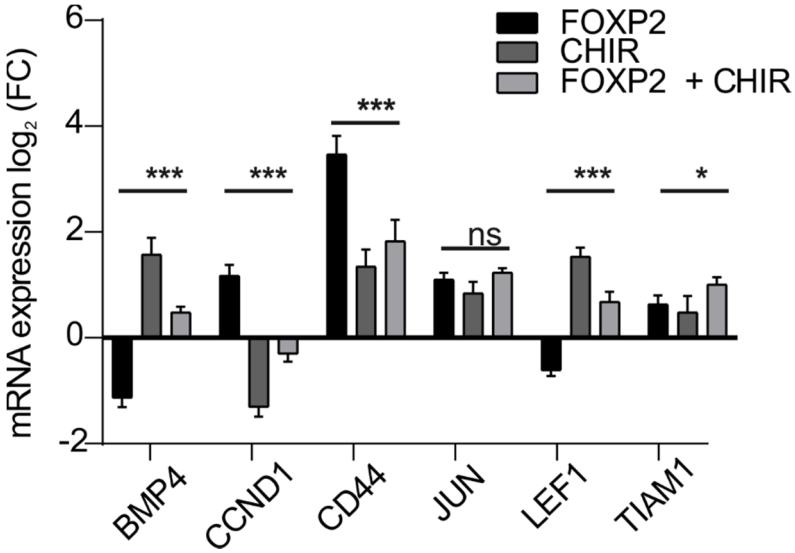

Fig.3.49: q-PCR quantification of Wnt genes changed by FOXP2 overexpression, CHIR treatment or FOXP2 overexpression and CHIR treatment compared to control.

To investigate the effect of β-catenin on FOXP2$^{\Delta\text{-helix}}$, a part of the used cells were treated with CHIR and compared to cells only expressing FOXP2$^{\Delta\text{-helix}}$ and cells only treated with CHIR.

The gene sets being unique for FOXP2 plus CHIR and FOXP2$^{\Delta\text{-helix}}$ plus CHIR were compared in order to find differences between the transcriptional activities of both FOXP2 constructs in presence of β-catenin. Thereby 963 upregulated genes were found to be similar in both data sets (Fig. 3.50).

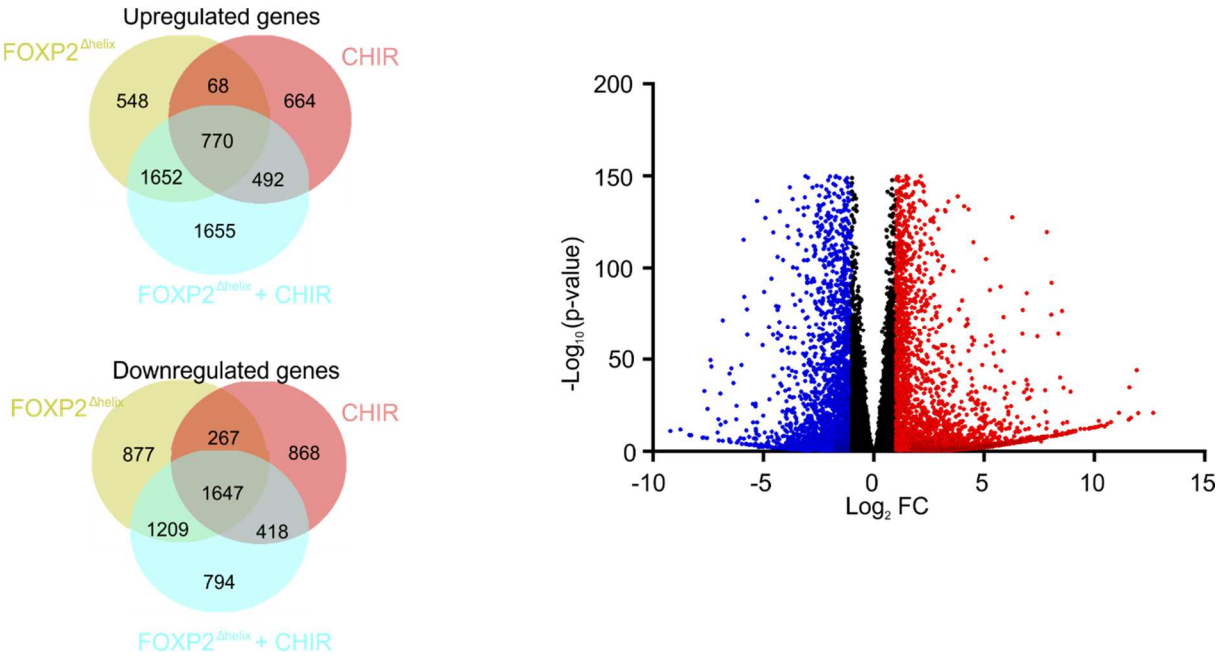

Fig 3.50: Left: Venn diagram showing the differences between FOXP2 $^{\Delta\text{-helix}}$, CHIR and FOXP2 $^{\Delta\text{-helix}}$ + CHIR, all compared to the control. Right: Volcano plot showing significantly up- (red) and downregulated (blue) genes of each condition compared to control.

KEGG-pathway and GO term biological function analysis indicated ´signal transduction´ (196 genes, 5, 4e-3) as biological function and ´Cytokine-cytokine receptor interaction´ (59 genes, 3, 4e-5) as KEGG-pathway with most significant DEGs. In the data set of the most downregulated genes cell adhesion (113 genes, 5, 5e-13) appeared as biological function and PI3K-Akt signaling pathway (70 genes, 2e-6) as KEGG pathway with most significant DEGs. The heatmap visualization displays the effect of CHIR on the regulation of Wnt targets in presence or absence of FOXP2 and FOXP2$^{\Delta\text{-helix}}$ (Fig. 3.51). Q-PCR confirmed the RNA-Seq results (Fig. 3.52).

KEGG pathways	Upregulated	Downregulated
	Ribosome biogenesis in eukaryotes	Focal adhesion
	Cytokine-cytokine receptor interaction	Arrhythmogenic right ventricular cardiomyopathy (ARVC)
	NF-kappa B signaling pathway	ECM-receptor interaction
	Rheumatoid arthritis	PI3K-Akt signaling pathway
	MAPK signaling pathway	Axon guidance
	Legionellosis	Hypertrophic cardiomyopathy (HCM)
	Nicotine addiction	Dilated cardiomyopathy
	Bile secretion	Rap1 signaling pathway
	TNF signaling pathway	Pathways in cancere
	Carbohydrate digestion and absorption	Proteoglycans in cancer
Biological functions		
	rRNA processing	Homophilic cell adhesion
	Inflammatory response	Cell adhesion
	Ion transport	Extracellular matrix organization
	Response to lipopolysaccheride	Nervous system development
	Sodium ion transmembrane transport	Synapse assembly
	Maturation of SSU-rRNA	Calcium-dependent cell-cell adhesion
	Neurotransmitter transport	Negative regulation of viral genome replication
	Cellular response to lipopolysaccharide	Glycosaminoglycan catabolic process
	Ion transmembrane transport	Chemical synaptic transmission
	Ribosomal large subunit biogenesis	Axon guidance

Tab.7: List of most significant up- and downregulated KEGG pathways and biological functions upon FOXP2$^{\Delta\text{-helix}}$ overexpression + CHIR treatment compared to control.

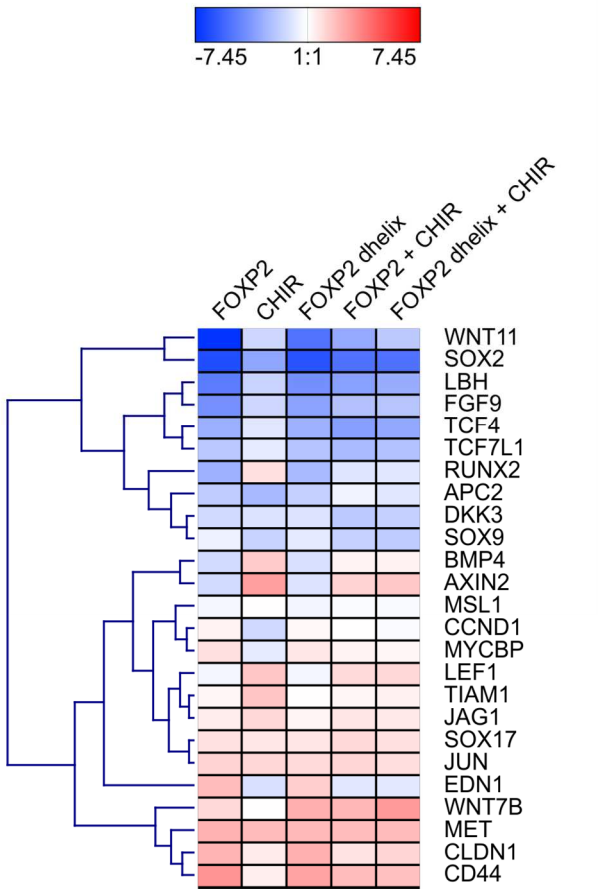

Fig 3.51: Heatmap of Wnt genes changed in all conditions compared to control, colors indicate normalized log_2 FC values, negative (blue), positive (red), neutral (white).

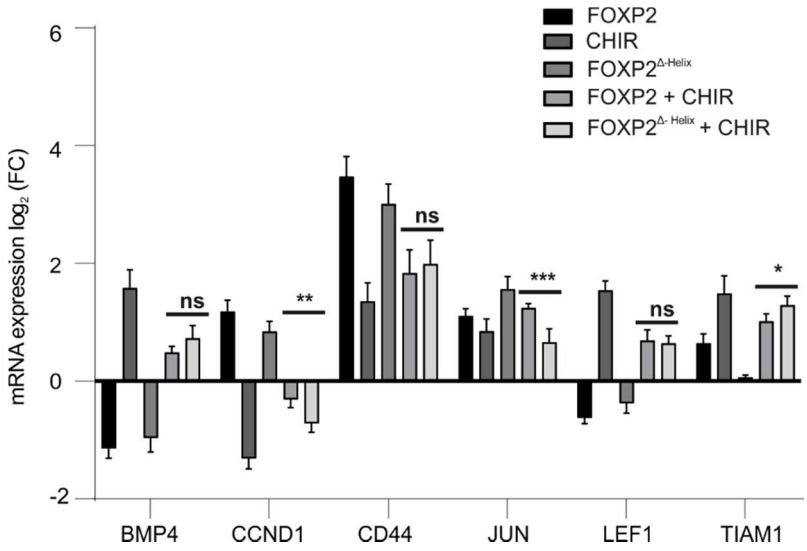

Fig 3.52: q-PCR quantification of Wnt genes changed by all investigated conditions compared to control.

3.8 FOXP2^{IDR} contains a phosphosite

In addition to the already known phosphorylations, we were interested to find futher in the IDR of FOXP2, as PTMs are often occurring on IDRs.
In order to detect PTMs such as phosphorylations, a few methods have been developed. Common ones are western blot, Enzyme-linked Immunosorbent assay (ELISA) or intracellular Flow Cytometry. They all have in common that the detection is sometimes not reliable, as the difference between a serine with or without phosphate group is difficult to detect.
Mass spectrometry (MS) has been one of the methods of choice to detect protein post-translational modifications as its high sensitivity and resolution. In principle, the method relies on the preliminary digestion of the protein sample and analysis of the resulting peptides. However, if different phosphorylation sites are in close proximity and on the same peptide after digestion, MS cannot identify the exact location of these sites. MS/MS might give the correct solution, but the instability of phosphate groups is often a problem. The quantitative analysis is also unreliable due to phosphate instability, and differential ionization efficiencies of the peptides. Therefore, MS usually is used with other biochemical techniques to give more accurate results of phosphorylated residues.

NMR can be used to study PTMS from an analytical perspective, but also from a mechanistic, functional and structural point of view. Most of PTMs is forming by reversible, covalent additions of small, chemical entities, such as phosphate groups, acyl chains, alkyl chains, or various sugars, to the side-chains of individual protein residues[89]. Covalent PTMs introduce local alterations in the chemical environments of individual protein residues that can then be detected as characteristic chemical shift changes in a NMR spectrum. Enormous advantages of NMR for PTM detection is the exact localization of the PTM and the ability to follow those in real-time.

In order to investigate, if, beside the known phosphorylation at S308, there is another phosphosite in the FOXP2^{IDR} construct, a NMR-based technique using cell-lysate and isotope labeled recombinantly expressed protein was performed in order to detect the effect of the kinases in the lysate on the isotope labeled protein[129]. According to this technique, first isotope labeled protein is recorded for a reference spectrum, then lysate

or recombinant kinases are added to the labeled protein and further spectra are recorded allowing the detection and tracking of phosphorylations and other PTMs.

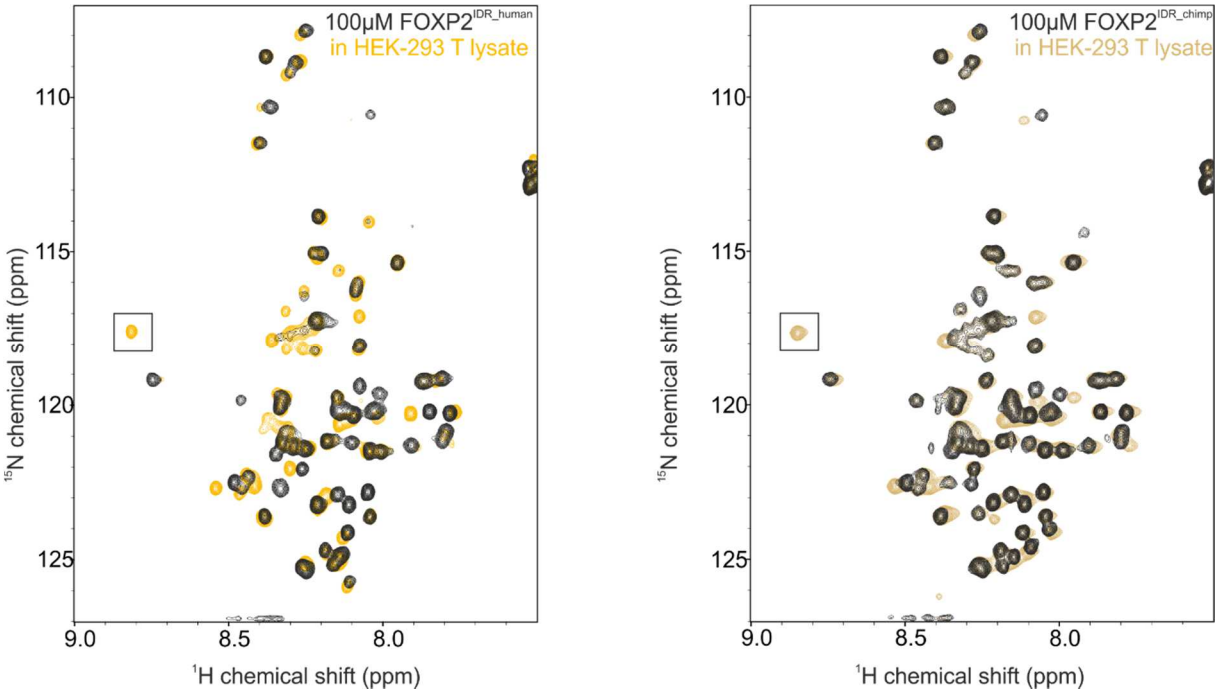

Fig 3.53: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{IDR} (black, 100µM, left: human, right: chimpanzee), HEK-293 T cell lysate (yellow, after 3h).

We used ^{15}N labeled human FOXP2^{IDR} construct and recorded a reference ^{1}H,^{15}N HSQC. Then human cell lysate from HEK 293 T cells was added and 20 ^{1}H,^{15}N HSQCs recorded successively. Those cells express FOXP2 endogenously, therefore kinases specific for FOXP2 phosphorylation are present. In order to avoid the removal of the phosphate group from a possible phosphorylated residue, phosphatase inhibitor within the lysate was used. As seen in figure 3.53, the ^{1}H, ^{15}N HSQC spectrum with lysate (blue) compared to the reference ^{1}H, ^{15}N HSQC spectrum (black) is different, as a signal appears on higher ^{1}H chemical shift (around 8ppm) after around 3h. This signal is typical for a phosphorylation event, as the phosphate group on the residue is more electron negative then without and therefore a direct proof for the phosphorylation of the FOXP2^{IDR} construct. In order to assign the phosphorylated residue, TEV-protease to cleave the Hexa-Histidine-tag from the ^{15}N, ^{13C} labeled FOXP2^{IDR} was not used, then the NMR experiments with the lysate performed. Next, the tagged protein was repurified from the lysate using Ni-NTA beads and then cleaved the Hexa-Histidine-tag from the protein and assigned the pure phosphorylated FOXP2^{IDR} sample. Afterwards 3D NMR experiments were recorded on that sample in order to assign the

construct with the phosphorylated residue. After completing the assignment, we determined Ser330 as residue being phosphorylated in HEK-293 T cell lysate.

To investigate, if this phosphorylation also occur in the chimpanzee protein, the experiments with the chimpanzee FOXP2^{IDR} construct was repeated. Indeed, the same phosphorylation on position 330 appeared in HEK-cell lysate (Fig. 3.52 right).

As FOXP2 is crucial for proper brain development, brain cells were an interesting environment to test, if this phosphorylation also occur in other cell types. As human brain is not accessible, brain tissue from mice were used. After lysing those, the phosphorylation assays were repeated with the human and chimpanzee FOXP2^{IDR} construct (Fig. 3.54). Indeed, the same phosphorylation occurred at the same frequency in the spectra for both species in mice brain cell lysate (Fig. 3.55).

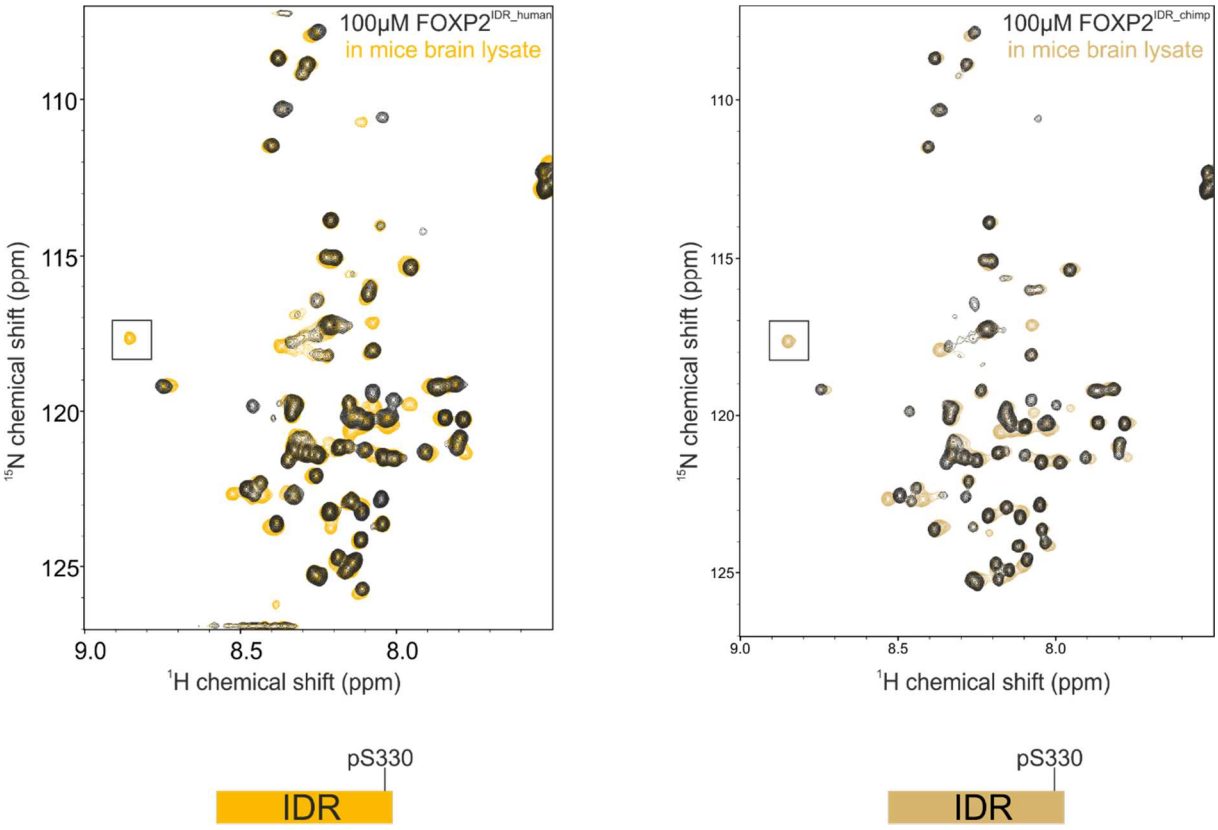

Fig 3.54: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{IDR} (black, 100µM, left: human, right: chimpanzee), mice brain cell lysate (yellow, after 4h).

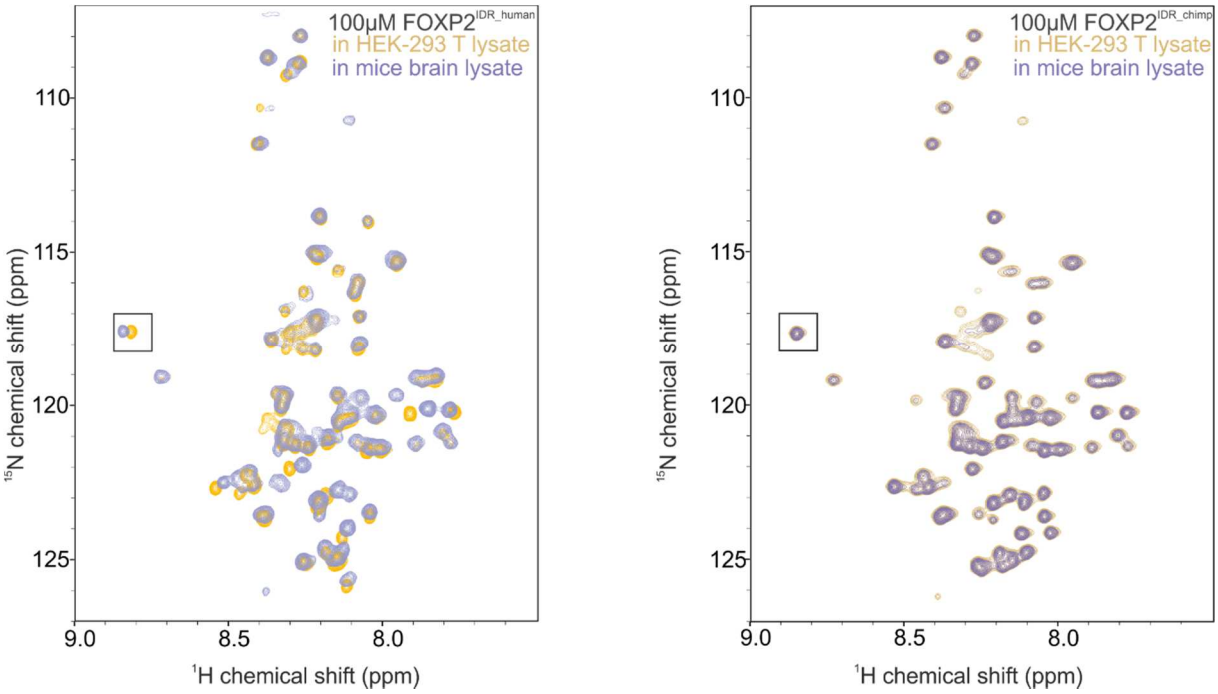

Fig 3.55: ^{1}H ^{15}N HSQC with ^{15}N labeled $FOXP2^{IDR}$ (each 100µM, left: human, right: chimpanzee), in HEK 293 T cell lysate (yellow, after 3h) overlaid with spectra recorded with mice brain cell lysate (purple, after 4h).

Now, we wanted to know, which kind of kinase this was and if it was the same in both cell types. In order to get an idea, we used a kinase prediction tool (kinasephos.mbc.nctu.edu.tw). By providing the sequence around Ser330, the software predicted protein kinase A (PKA), protein kinase C (PKC), Casein kinase 1 (CK1) as possible kinases for this motif (R-R-X-S). In order to check, if one of those predicted kinase indeed is responsible for the phosphorylation of FOXP2 S330, a recombinantly expressed protein kinase A was purchased (Promega, V5161). Protein kinase A, also known as cAMP-dependent kinase, is one of the most studied kinases and belongs to the serine/threonine kinases. It is mainly active in the regulation of the catabolism, but also important in cell proliferation by acting on transcription factors[130] and occurs in cytoplasm and nucleus. In order to investigate the effect of the kinase on FOXP2 a reference ^{1}H, ^{15}N HSQC spectrum with ^{15}N labeled $FOXP2^{IDR}$ human/chimpanzee was recorded. Then, 400 units of PKA were added to the sample and after incubation for one hour another ^{1}H, ^{15}N HSQC spectrum was recorded (Fig. 3.56). Indeed, a peak at higher chemical shift (8.85 ppm at ^{1}H axis) appeared compared to the reference spectrum, indicating the phosphorylation of an amino acid in both species.

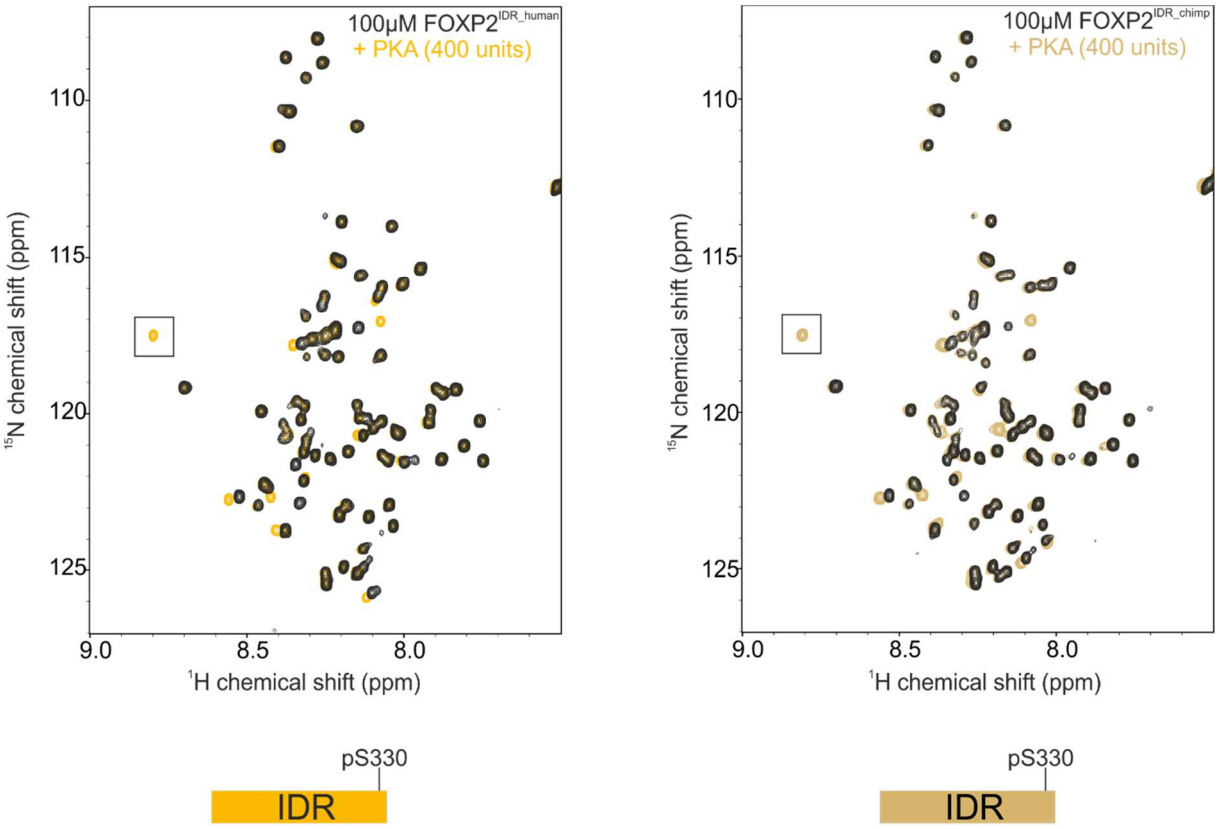

Fig 3.56: ^{1}H ^{15}N HSQC with ^{15}N labeled $FOXP2^{IDR}$ (black, 100µM, left: human, right: chimpanzee), PKA titrations (yellow, after 40min).

By comparing the spectra with the spectra of the cell lysate-experiments (Fig. 3.57), we concluded, that the phosphorylation is also taking place at Ser330. These data indicate, that it's PKA, which is responsible for the phosphorylation of S330 of FOXP2. PKA indeed is expressed in HEK-293 T cells and also known to be present in brain cells (see expression levels at www.proteinatlas.org).

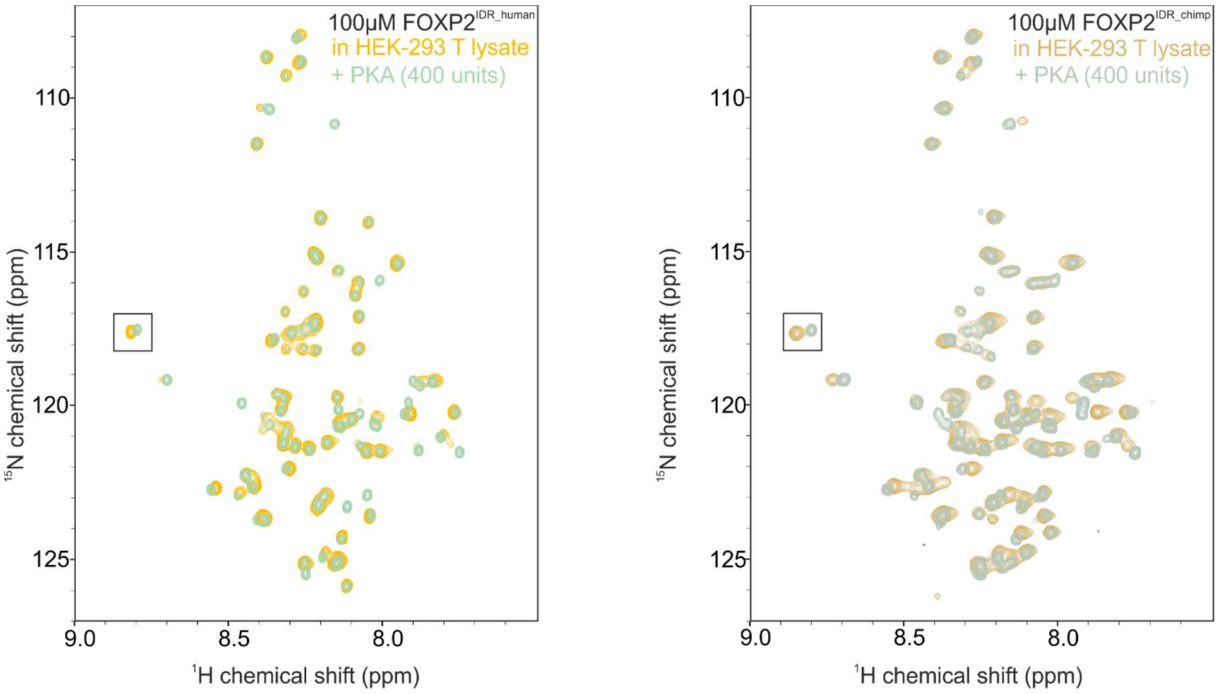

Fig 3.57: ^{1}H ^{15}N HSQC with ^{15}N labeled $FOXP2^{IDR}$ (each 100µM, left: human, right: chimpanzee), in HEK 293 T cell lysate (yellow, after 3h) overlaid with spectra recorded with PKA (green, after 40min).

To investigate the effect of the phosphorylation on binding partners, a phospho-mimicking mutant of $FOXP2^{IDR}$ was designed by replacing the Ser330 to a Glutamic acid ($FOXP2^{S330E}$). Hereby a similar structure of the phosphorylated serine is created by mutating the serine to a glutamic acid, as this amino acid harbors a negative charge[131].

As phosphorylations can have a regulatory effect on transcription factors, we tested the effect of the phospho-mimicking mutant on the newly discovered interaction partners. First, a ^{1}H, ^{15}N HSQC spectrum of ^{15}N labeled $FOXP2^{S330E}$ was recorded as reference. Then increasing amounts of unlabeled β-$catenin^{141-305}$ was added (Fig. 3.58).

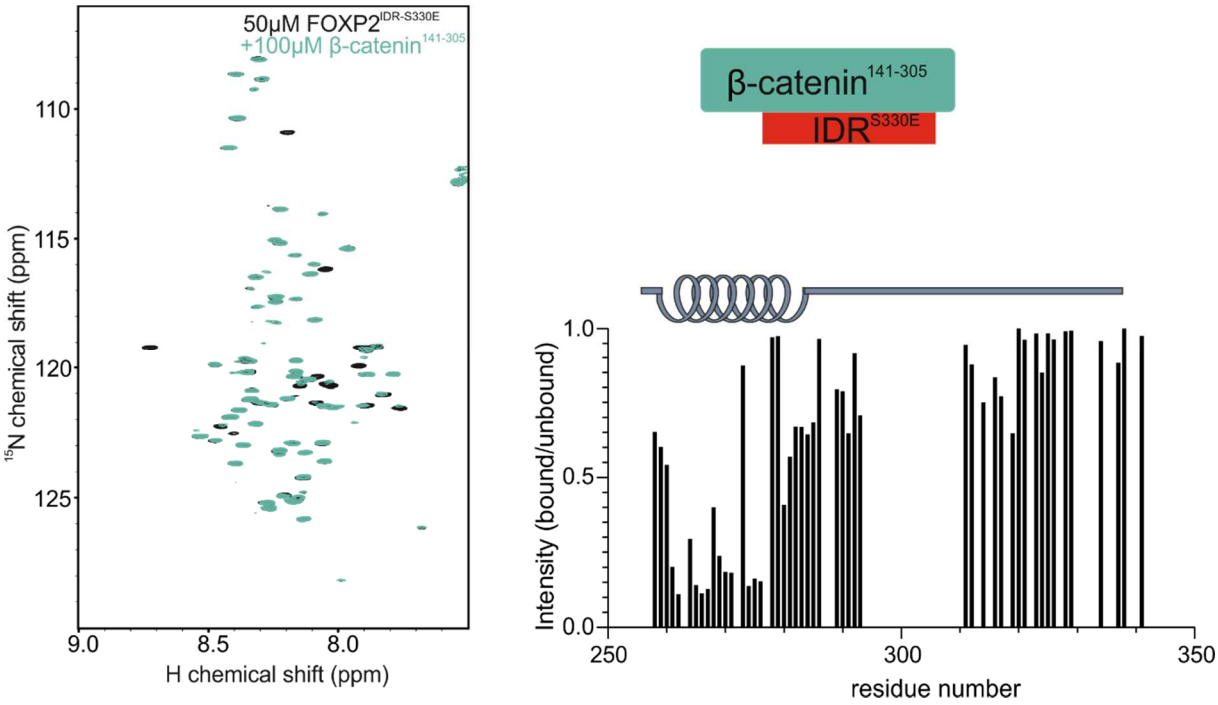

Fig 3.58: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{S330E} (black, 100µM), titrations with β-catenin$^{141-305}$ (green, µM). Right: Intensity plot indicating the affected residues around region with α-helical propensity.

According to those results, the phospho-mimicking mutant is not affecting the binding to β-catenin as similar CSPs were observed than with FOXP2^{IDR}. Then a possible effect of the phospho-mimicking mutant on other newly discovered FOXP2 interaction partners of this project was investigated (i.e. LEF1, ICAT). After recording a reference spectrum with ^{15}N labeled FOXP2^{S330E} increasing amounts of LEF$^{288-399}$ (Fig. 3.59) or ICAT (Fig. 3.60), were added.

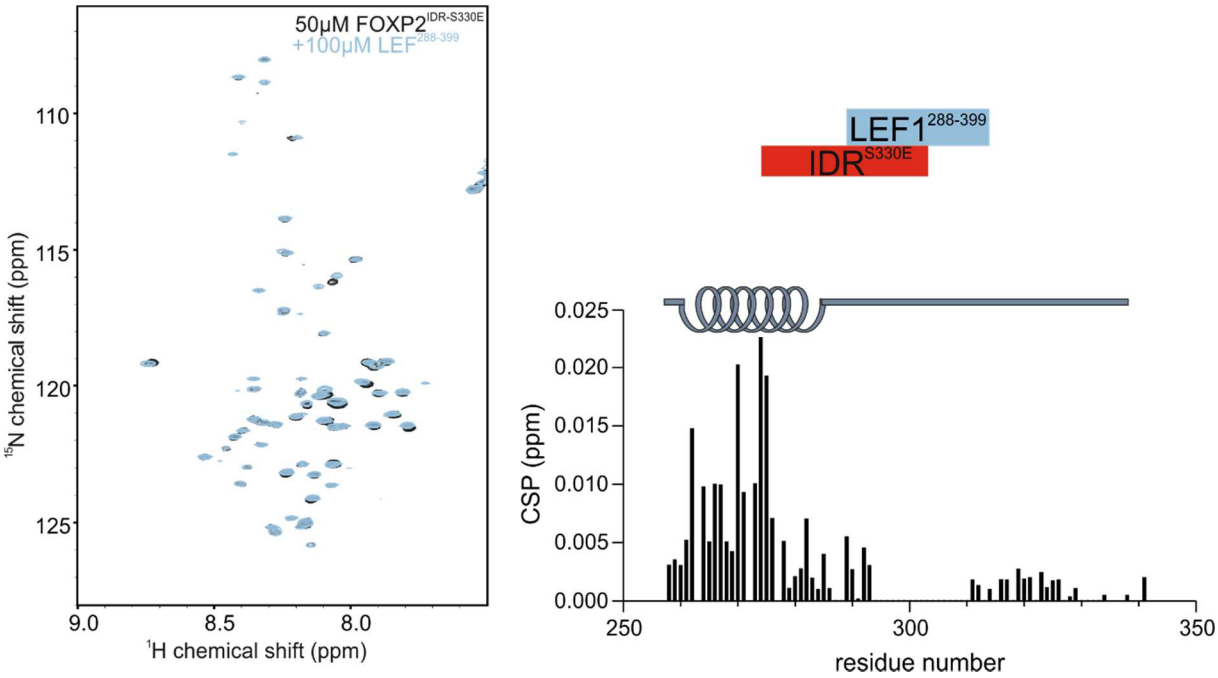

Fig 3.59: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{S330E} (black, 100µM), titrations with LEF$^{288-399}$ (blue, µM). Right: CSP plot indicating the affected residues around region with α-helical propensity.

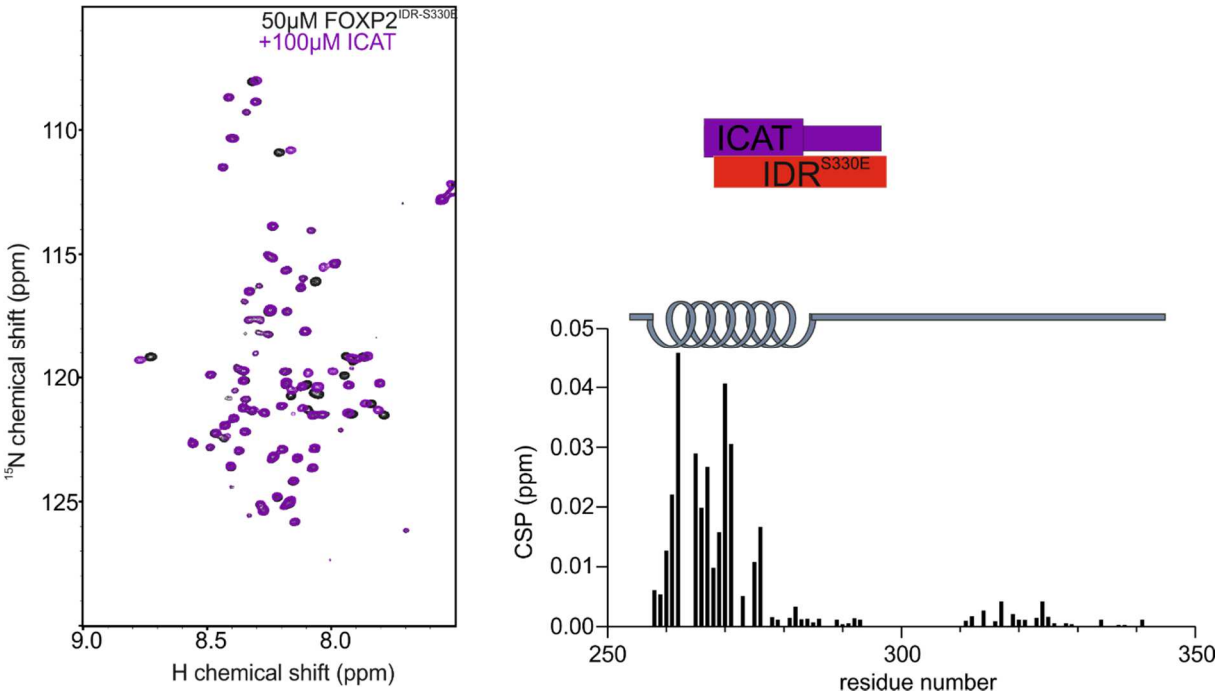

Fig 3.60: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled FOXP2^{S330E} (black, 100µM), titrations with ICAT (purple, µM). Right: CSP plot indicating the affected residues around region with α-helical propensity.

ICAT and LEF$^{288-399}$ were still binding to the IDR despite the negative charge at position 330. Next, it was tested, if the FH of FOXP2 is still binding to FOXP2^{S330E} (Fig. 3.61).

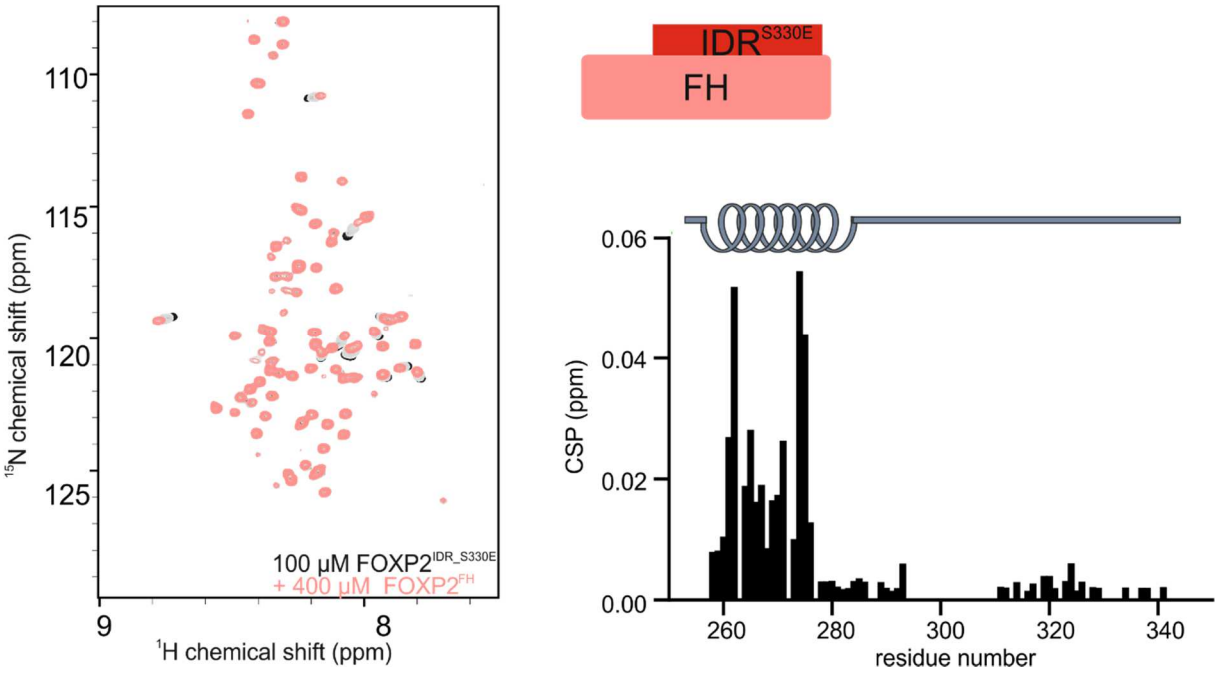

Fig 3.61: Left: ^{1}H ^{15}N HSQC with ^{15}N labeled $FOXP2^{S330E}$ (black, 100µM), titrations with $FOXP2^{FH}$ (coral, 400µM). Right: CSP plot indicating the affected residues around region with α-helical propensity.

Indeed, also here the interaction still took place. This indicates, that for the interaction partners I found in my thesis binding to the IDR of FOXP2, the phosphorylation does not seem to have an effect. These results could be due to the fact that the phosphorylation site is not located close to the region with α-helical propensity (residue 264-272), thus the impact of the phosphorylation on binding partners binding to the region with α-helical propensity might be less likely.

3.9 R553H mutant binds to DNA

In order to test which effect the mutant responsible for the speech disorder has on protein-protein and protein-DNA interactions, another FH construct was designed including the point mutation ($FOXP2^{R553H}$). The first differences between wildtype and mutant already appeared during the expression and purification. While the wildtype FH was completely soluble after elution of the NiNTA step, the mutant was mainly insoluble. To solve this, lysis buffer with high concentrations of Urea was used, in order to denaturate the protein and making it soluble. To confirm the proper refolding of the FH after removal of Urea, a ^{1}H, ^{15}N HSQC was recorded and compared to the wildtype spectrum (Fig. 3.62). Indeed, the forkhead of the mutant had similar signals than the wildtype FH indicating that the $FOXP2^{R553H}$ refolded during purification.

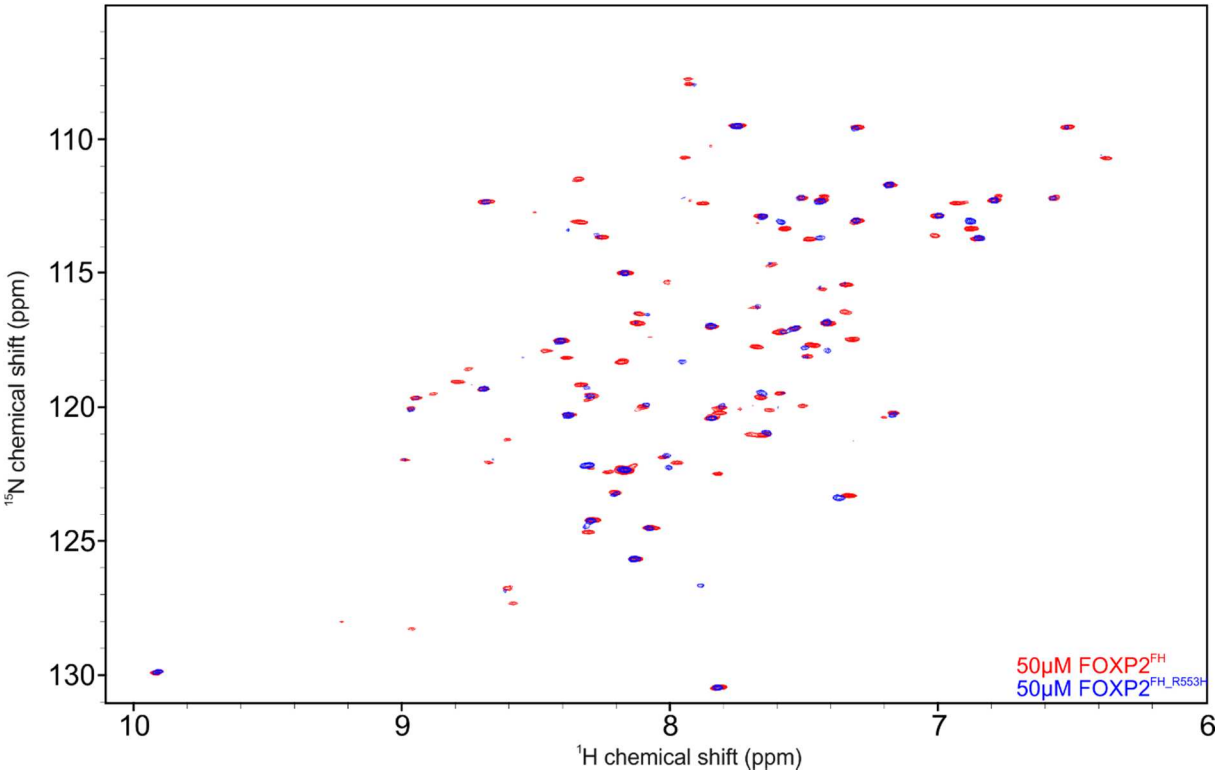

Fig.3.62: ^{1}H, ^{15}N HSQC of the FOXP2^{R553H} compared to the wildtype FH proofing the proper refolding of the FH-mutant.

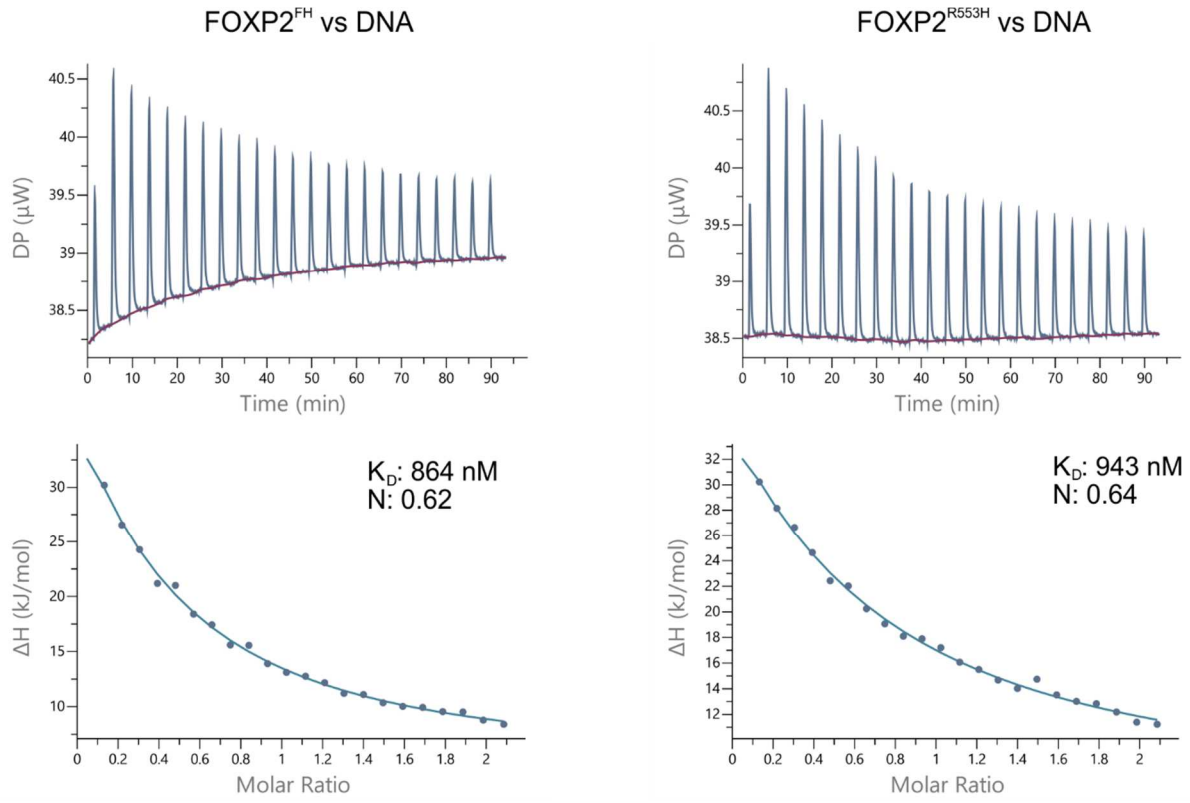

Fig. 3.63: Isothermal calorimetry comparison of the binding between FOXP2^{FH} and DNA (left) and FOXP2^{R553H} and DNA (right).

Also the behavior of wildtype and mutant differed in phosphate buffer, when DNA was added. While the wildtype remained soluble, the mutant precipitated. Therefore HEPES buffer was used, as the mutant did not show precipitation in this buffer composition. Additionally, while the wildtype FH was binding to a heparin column due to its global positive charge, the mutant did not bind. This can be explained, as the pI of the wildtype is 8.5 at a pH of 7, the pI of the mutant is 7.6, therefore the mutant protein is less positively charged than the wildtype protein. According to these information, one could conclude, that the DNA binding affinity of the mutant is changed due to its point mutation. A study showed that the mutant where not showing any DNA binding in a luciferase-assay[75].

As *in-cell* interactions must be rather strong to be detected, *in vitro* interactions can also be detected in µM range. Thus, we tested if the mutant is binding to DNA using ITC. Interestingly, by titrating DNA into FOXP2^{R553H} the ITC-derived dissociation constants corresponding to DNA binding to either WT or mutated FOXP2^{FH} are similar (864nM vs 943nM) compared to the FH wildtype (Fig. 3.63). Thus, in our experiments, no difference between WT and mutant could be detected.

Chapter 4

4 Discussion

4.1 FOXP2 contains two protein-protein interaction sites

FOXP2 belongs to a broad family of transcription factors, which are known to be involved in embryonal development. As it is responsible for various biological functions, a tight regulation of its transcriptional activity must take place. This can occur by protein partners interacting and thus influencing the activity of FOXP2. So far, only few interaction partners have been identified, their role in regulation of FOXP2 remains elusive. In studies using experiments such as Co-IP CTBP1[79], NFAT, PIAS1[75], TBR1[132], FOXP1 and FOXP4[31] have been identified to interact with FOXP2. Additionally, variations of homo- and heterodimerization of FOXP1, FOXP2 and FOXP4 have been shown to influence significantly the expression of known FOXP2 targets involved in embryonal neuronal development[31]. Another mechanism of regulation are post-translational modifications. FOXP2 harbour a few phosphorylation sites, an ubiquitination site and a sumoylation site. Only for one phosphorylation the effect is known so far[92].

In this thesis I first investigated an IDR within FOXP2, which is located between the polyQ-region and the zinc finger (residue 247-331) and might play a crucial role in the evolutionary development of humans, as both amino acids, which differ between human and chimpanzee, are located in this IDR. Indeed, a region with α-helical propensity was found, which thus might function as protein-protein interaction site. IDRs are crucial players in regulation of proteins, as they facilitate binding to other partners due to less flexibility and solvent exposure and can influence the regulation[133,134]. This α-helical propensity was observed in both species (human and chimpanzee), indicating that the two different amino acids do not affect the formation of the α -helix and thus behave similarly.

To determine if this alpha-helix indeed facilitate protein protein interactions NMR was used as sensitive methods for protein-protein interactions detection and indeed we determined various proteins as novel interactions partners to the IDR of FOXP2 with α-helical propensity. The transcription factor LEF1 and the regulator ICAT, both involved in the Wnt-pathway, are directly interacting with the IDR of FOXP2, indicating that those interactions might be competitive under certain circumstances. These interactions could be a competitive mechanism in order to regulate transcription of

targets, such as the Wnt targets. Thereby FOXP2^{IDR} could interact with the DNA-binding domain of LEF1 and thereby changing the transcriptional activity of LEF1. Then the FH of FOXP2 could thereby bind to DNA and regulate the transcription of other genes. All recorded interactions were affecting the residues around the previously determined α-helix, indicating the important function of this secondary structure element in the interaction network of FOXP2. As already known, FOXP1 and FOXP3 are involved in Wnt signaling pathway[93,94], thus we tested a possible link between FOXP2 and this pathway, which is important in embryonal development and adult homeostasis. Thereby we discovered β-catenin as novel interaction partner of FOXP2. Using NMR two binding sites were determined, the FH domain and the IDR between PolyQ-region and zinc finger. Using different constructs of β-catenin the first part of the armadillo region of β-catenin was determined to interact with the IDR of FOXP2 and the disordered N- and C-termini of β-catenin are interacting with the FH of FOXP2. These results indicate, that one molecule of β-catenin could bind to one molecule of FOXP2. Interestingly, the N- and C-terminus of β-catenin bind to the same region at the FH but interact with different affinities. Then, the interaction was proven with endogenous FOXP2 and β-catenin via Co-IP in human cells, whereby it was shown that FOXP2 and β-catenin interact in HEK-293-T cells confirming a – at least - indirect interaction between both proteins. Using different constructs we could show with Co-IP, that β-catenin has at least two binding sites on FOXP2 confirming our NMR data. With ITC the binding between FH and the C-terminus of β-catenin was determined as slightly stronger than the binding between FH and N-terminus, which indicates that the interaction between FH and C-terminus is the favorable interaction and only upon interaction of the C-terminus with another protein partner might facilitate the binding between FH and N-terminus. This competitive element thus might be a regulatory function on the transcriptional activity of FOXP2. Another possibility for this competitive network might be other interaction partners of FOXP2, such as other FOX proteins. In absence of those, β-catenin might affect the transcriptional activity of FOXP2 by binding weakly to the FH domain. In presence of other interaction partners, β-catenin might bind to those because of possible higher affinities leading to a disruption of affecting FOXP2 and its function. A further possibility would be, that other proteins have a stronger affinity to the FH of FOXP2 than β-catenin and thus compete with it for binding to FOXP2. In line with this hypothesis, FOXP2 could compete with TCF for

interaction with β-catenin. The transcriptional activity of TCF might be decreased, if FOXP2 is overexpressed in the cells and interacting with β-catenin.
Considering the fact, that β-catenin is binding to two binding sites, which are located far from each other, a possible link between both binding sites was investigated. Indeed, NMR experiments have shown, that the IDR is interacting with the FH domain creating a backfold within the protein. If this interaction is happening as intramolecular or intermolecular could not be determined, but given that fact, that the concentration of FOXP2 in the nucleus must be in low µM range to make an intermolecular interaction possible, the intramolecular interaction was taken as the most likely state. This interaction provides variations of possible regulatory elements. In order to test the simplest idea, that the IDR backfold on the FH domain affects the DNA binding affinity of FOXP2, NMR and ITC experiments were performed. Indeed, a difference in binding affinity was observed, if an artificial construct containing the FH linked to the IDR via a GS-linker was titrated with DNA compared to the FH alone. The data have shown that the DNA binding is stronger in presence of the IDR, indicating that this IDR acts as a regulatory element in FOXP2 function. As the binding between FH and the IDR of FOXP2 is weak, this interaction might only take place, when there are no other proteins binding to the FH or the IDR itself, except their affinity is lower than the one between FH and IDR. Thus other proteins might interrupt this intramolecular interaction by either binding to the FH or to the IDR of FOXP2 and thus regulate the function of FOXP2. Another function of this backfold might be the translocation of cofactors affecting the transcriptional activity of FOXP2. This hypothesis is supported by our findings, that the α-helix within the IDR of FOXP2 is an interaction site for various proteins and thus might play an important role in recruiting other proteins to the FH and thereby affecting the transcriptional activity.

4.2 FOXP2 is involved in the regulation of the Wnt-signaling pathway

In order to investigate the effect of β-catenin and the intramolecular backfold on the transcriptional activity of FOXP2, RNA Seq was performed with six different conditions. By overexpressing FOXP2 in U2OS cells more genes were found to be downregulated than upregulated, which fits with other studies claiming FOXP2 as transcriptional repressor[24,135]. Thus the data in this thesis confirm the effect of FOXP2 on Wnt-signaling. Beside this pathway various pathways involved in brain development,

morphology and learning were significantly changed upon FOXP2 overexpression, such as cell adhesion molecules, Hippo signaling and Wnt signaling. Based on those data the Wnt-signaling pathway was found to be significantly downregulated upon FOXP2 overexpression in U2OS cells which has been shown before[136,137]. Various other FOX proteins are already known to regulate Wnt pathway. Beside FOXP1[93] and FOXP3[94], also FOXK1 and FOXK2[138] have been found to be linked to Wnt-signaling regulation. Thus FOX transcription factors play a crucial role in this important signaling way and thus understanding the molecular mechanisms behind such networks might help to understand the development of diseases such as cancer and Alzheimers' disease.

In order to investigate the effect of β-catenin on the transcriptional activity of FOXP2 cells overexpressing FOXP2 were treated with CHIR, a chemical inducing the nuclear localization of β-catenin. With this treatment, FOXP2 and β-catenin were both localized in the nucleus of the cell and thus could interact with each other. Indeed, we observed that in the condition of FOXP2 overexpression plus CHIR treatment the number of upregulated genes increased significantly compared to the condition of FOXP2 overexpression without CHIR treatment. This results could proof that, while Wnt signaling is inactive, FOXP2 acts as a repressor and upon Wnt activation, β-catenin concentrations increase in the nucleus and inhibit FOXP2 activity leading to loss of its function as transcriptional repressor. Additionally, by comparing the targets changed upon FOXP2 overexpression alone and the condition of FOXP2 overexpression plus β-catenin induction, many differences occurred. These data indicate that the transcriptional activity of FOXP2 is affected by the presence of β-catenin and thus other targets are regulated by FOXP2. Nevertheless, it cannot be proven, that these changes occurred because of the direct interaction of both proteins, but the effect can also come from differences in cofactors and affinities. An option would be that β-catenin binds to proteins, which were affecting the activity of FOXP2 in absence of β-catenin, thus the effect of those on FOXP2 is disbanded and the transcriptional activity changed.

Further results in this thesis show the direct interaction of the IDR to the FH domain of FOXP2, forming probably a backfold-like conformation. In order to find proofs for a regulatory effect of the IDR on FOXP2 activity RNA-Seq was performed using a FOXP2 variant lacking this α-helix. Indeed, various differences were found by comparing this data set to the data set of the $FOXP2^{FL}$ overexpression. For a few genes it seems that even they are downregulated by $FOXP2^{FL}$, the variant lacking the α-helix leads to an

further increased transcriptional repression. However, by comparing the by the variant and FOXP2^{FL} significantly regulated genes we found that 30% of the downregulated and 33% of upregulated genes differ between both conditions. Interestingly, we found that the Wnt signaling pathway is suppressed less by the FOXP2$^{\Delta\text{-helix}}$ construct than by FOXP2^{FL}. Thus, FOXP2^{FL} activity seems to be regulated by the α-helix as the variant lacking this secondary structure element regulates various genes which are not regulated by the full-length transcription factor. These data suggest that the α-helix is indeed a regulatory element in FOXP2 itself as FOXP2^{FL} is suppressing certain pathways such as the Wnt-signaling pathway more efficient than FOXP2 lacking the α-helix. Thus this region might indeed function as interaction side for various co-factors such as β-catenin to the FH and thus regulating the activity of FOXP2. Additionally, the newly discovered backfold mechanism might play a role in the regulation of FOXP2 by affecting the DNA binding affinities on FOXP2 target genes. In this thesis we discovered while using ITC that an artificial construct, containing the FH and the α-helix, is binding with higher affinities to a DNA target than the FH alone. This must not be true for other DNA targets but indicates that the IDR indeed is regulating the DNA binding affinities of FOXP2. For other targets the interaction to DNA might be disturbed by the IDR or co-factors which are recruited to the FH via the intramolecular interaction. To obtain more insight in this mechanisms, other studies must take place in order to understand how the backfold is changing the transcriptional activity of FOXP2.

4.3 FOXP2 is phosphorylated by protein kinase A

Beside novel interaction partners, which regulate FOXP2 activity, also PTMs such as phosphorylation were investigated. As IDRs in transcription factors are solvent-accessible, they are often target of PTMs and thus regulated. In this thesis FOXP2^{IDR} was investigated to find possible novel PTMs. This region was chosen, as it is intrinsically disordered and additionally, it contains the two amino acids which differ between human and chimpanzee, thus another regulation mechanism might be an interesting and important feature between both species. In this region, already one phosphorylation is known in humans at residue Ser308 (https://www.phosphosite.org), but the effect of this phosphorylation is still unknown. Using HEK-293T-cell lysate and NMR, a new phosphorylation site was discovered at Ser330, in both human and chimpanzee FOXP2. Interestingly, this phosphorylation is located close to the residue,

which are different between human and chimpanzee (N303T, S325N)[45]. In our NMR experiments this phosphorylation did not change the binding to interaction partners, which were discovered earlier in this thesis. It neither changed the intramolecular interaction to the FH, thus a direct effect on the DNA binding ability of FOXP2 was not observed. Thus, our *in vitro* data do not show any effect of this phosphorylation event in FOXP2 regulation at molecular level. However, *in cell* assays were not performed, thus it remains open, whether this phosphorylation is affecting the transcriptional activity of FOXP2. Thus RNA-Seq experiments with cells either expressing FOXP2 as phospho-mimicking mutant or a mutant containing an alanine instead of the phosphorylated residue to prevent phosphorylation by endogenous kinases could be carried out, in order to investigate the effect of the phosphorylation on the transcriptional activity of FOXP2.

Further on, PKA was determined as possible responsible kinase. PKA is a kinase, which is dependent on cellular levels of cAMP and phosphorylates proteins exposing the motif arginine-arginine-X-serine. By these phosphorylations proteins can be activated or deactivated. As PKA has various proteins as targets, its function varies with the cell type its occurring into. A few examples are stimulating glycogenolysis, glycolysis and epithelial sodium channel and affecting renin secretion. Disfunctions of this kinase are linked to various diseases such as cardiovascular diseases and cancer[139]. As FOXP2 is mainly expressed in the brain, the phosphorylation of PKA might play a role in the reward system by transfer/translate the dopamine signal into cells in the nucleus accumbens[140]. Another possibility would be its role in memory formation, as PKA knockdown in Drosophila melanogaster showed decreased learning ability and memory retention[141]. Thus, PKA might play a role on FOXP2 while learning. As the motif around Ser330 is highly conserved in mammals, birds, reptiles and amphibia, this phosphorylation occurs in various other species, where it could be responsible for memory and learning. The cellular effect of the phosphorylation of FOXP2 by this kinase could not be determined in this thesis, thus further work must be done. It would be possible, that this phosphorylation increases/decreases the binding affinity to certain partners, which then regulate the transcriptional activity of FOXP2 and thus its function.

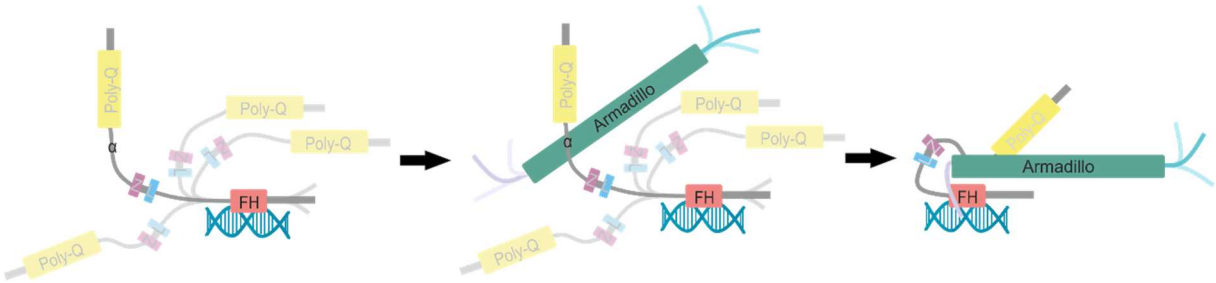

Fig.4: possible model of the regulation of FOXP2 by β-catenin.

Concluding, in this thesis I found the Wnt signaling pathway to be regulated by FOXP2 for the first time. Thus three of the four members of the FOXP family are linked to this pathway which acts in embryonal development and cancer. I could map two interaction sites of β-catenin to FOXP2, these results reveal an evolutionarily conserved function of β-catenin that is independent of TCF signaling. Both binding sites seem to be important for the function and regulation of FOXP2 activity as I observed several changes in regulated pathways if the first binding site was lacking. Thus we come to the hypothesis that the α-helix in the IDR acts as a recruiter for co-factors (Fig. 4). Those play crucial roles in signal transduction and regulation[9]. This regulatory element seems also to be important in the regulation by β-catenin, as we found various genes changed in expression when overexpressing cells with FOXP2 and treated with CHIR compared to cells overexpressing $FOXP2^{\Delta\text{-helix}}$ with CHIR treatment, leading to other biological and molecular functions.

Concluding we found regulatory elements within FOXP2 which must be closer determined. We found β-catenin regulating FOXP2 activity, thus further studies on this pathway should be carried on.

5 References

1. Proudfoot, N.J., Furger, A. & Dye, M.J. Integrating mRNA processing with transcription. *Cell* **108**, 501-12 (2002).
2. Watson, J.D. *Molecular biology of the gene*, xxxiv, 872 pages (Pearson, Boston, 2014).
3. Lykke-Andersen, S. & Jensen, T.H. Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie* **89**, 1177-82 (2007).
4. Soares, D. et al. Archaeal histone stability, DNA binding, and transcription inhibition above 90 degrees C. *Extremophiles* **2**, 75-81 (1998).
5. Xie, Y. & Reeve, J.N. Transcription by Methanothermobacter thermautotrophicus RNA polymerase in vitro releases archaeal transcription factor B but not TATA-box binding protein from the template DNA. *J Bacteriol* **186**, 6306-10 (2004).
6. Wilkinson, S.P., Ouhammouch, M. & Geiduschek, E.P. Transcriptional activation in the context of repression mediated by archaeal histones. *Proc Natl Acad Sci U S A* **107**, 6777-81 (2010).
7. Mitchell, P.J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-8 (1989).
8. Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. & Teichmann, S.A. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14**, 283-91 (2004).
9. Xu, L., Glass, C.K. & Rosenfeld, M.G. Coactivator and corepressor complexes in nuclear receptor function. *Curr Opin Genet Dev* **9**, 140-7 (1999).
10. Narlikar, G.J., Fan, H.Y. & Kingston, R.E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475-87 (2002).
11. Kaestner, K.H., Knochel, W. & Martinez, D.E. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev* **14**, 142-6 (2000).
12. Bourgeois, B. & Madl, T. Regulation of cellular senescence via the FOXO4-p53 axis. *FEBS Lett* **592**, 2083-2097 (2018).
13. Burgering, B.M. & Medema, R.H. Decisions on life and death: FOXO Forkhead transcription factors are in command when PKB/Akt is off duty. *J Leukoc Biol* **73**, 689-701 (2003).
14. Prasad, S.B. et al. Down Regulation of FOXO1 Promotes Cell Proliferation in Cervical Cancer. *J Cancer* **5**, 655-62 (2014).
15. De Filippis, R. et al. Expanding the phenotype associated with FOXG1 mutations and in vivo FoxG1 chromatin-binding dynamics. *Clin Genet* **82**, 395-403 (2012).
16. Florian, C., Bahi-Buisson, N. & Bienvenu, T. FOXG1-Related Disorders: From Clinical Description to Molecular Genetics. *Mol Syndromol* **2**, 153-163 (2012).
17. Kortum, F. et al. The core FOXG1 syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J Med Genet* **48**, 396-406 (2011).
18. Myatt, S.S. & Lam, E.W. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer* **7**, 847-59 (2007).
19. Teh, M.T. et al. FOXM1 is a downstream target of Gli1 in basal cell carcinomas. *Cancer Res* **62**, 4773-80 (2002).
20. Kalinichenko, V.V. et al. Foxm1b transcription factor is essential for development of hepatocellular carcinomas and is negatively regulated by the p19ARF tumor suppressor. *Genes Dev* **18**, 830-50 (2004).
21. Kim, I.M. et al. The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. *Cancer Res* **66**, 2153-61 (2006).
22. Liu, M. et al. FoxM1B is overexpressed in human glioblastomas and critically regulates the tumorigenicity of glioma cells. *Cancer Res* **66**, 3593-602 (2006).
23. Wonsey, D.R. & Follettie, M.T. Loss of the forkhead transcription factor FoxM1 causes centrosome amplification and mitotic catastrophe. *Cancer Res* **65**, 5181-9 (2005).
24. Herrero, M.J. & Gitton, Y. The untold stories of the speech gene, the FOXP2 cancer gene. *Genes Cancer* **9**, 11-38 (2018).

25. Chen, M.T. et al. Downregulation of FOXP2 promotes breast cancer migration and invasion through TGFbeta/SMAD signaling pathway. *Oncol Lett* **15**, 8582-8588 (2018).
26. Cesario, J.M., Almaidhan, A.A. & Jeong, J. Expression of forkhead box transcription factor genes Foxp1 and Foxp2 during jaw development. *Gene Expr Patterns* **20**, 111-9 (2016).
27. Lai, C.S., Gerrelli, D., Monaco, A.P., Fisher, S.E. & Copp, A.J. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* **126**, 2455-62 (2003).
28. Lu, M.M., Li, S., Yang, H. & Morrisey, E.E. Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Mech Dev* **119 Suppl 1**, S197-202 (2002).
29. Tsui, D., Vessey, J.P., Tomita, H., Kaplan, D.R. & Miller, F.D. FoxP2 regulates neurogenesis during embryonic cortical development. *J Neurosci* **33**, 244-58 (2013).
30. Xu, S. et al. Foxp2 regulates anatomical features that may be relevant for vocal behaviors and bipedal locomotion. *Proc Natl Acad Sci U S A* **115**, 8799-8804 (2018).
31. Sin, C., Li, H. & Crawford, D.A. Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. *J Mol Neurosci* **55**, 437-48 (2015).
32. Bacon, C. & Rappold, G.A. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Hum Genet* **131**, 1687-98 (2012).
33. Shu, W. et al. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* **134**, 1991-2000 (2007).
34. Watkins, K.E., Dronkers, N.F. & Vargha-Khadem, F. Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain* **125**, 452-64 (2002).
35. Mizutani, A. et al. Intracellular distribution of a speech/language disorder associated FOXP2 mutant. *Biochem Biophys Res Commun* **353**, 869-74 (2007).
36. Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. & Monaco, A.P. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519-23 (2001).
37. Liegeois, F. et al. Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat Neurosci* **6**, 1230-7 (2003).
38. Yan, X. et al. Downregulation of FOXP2 promoter human hepatocellular carcinoma cell invasion. *Tumour Biol* **36**, 9611-9 (2015).
39. Cuiffo, B.G. et al. MSC-regulated microRNAs converge on the transcription factor FOXP2 and promote breast cancer metastasis. *Cell Stem Cell* **15**, 762-74 (2014).
40. Campbell, A.J. et al. Aberrant expression of the neuronal transcription factor FOXP2 in neoplastic plasma cells. *Br J Haematol* **149**, 221-30 (2010).
41. Khan, F.H. et al. Acquired genetic alterations in tumor cells dictate the development of high-risk neuroblastoma and clinical outcomes. *BMC Cancer* **15**, 514 (2015).
42. Jia, W.Z. et al. MicroRNA-190 regulates FOXP2 genes in human gastric cancer. *Onco Targets Ther* **9**, 3643-51 (2016).
43. Wong, K.K. et al. FOXP2-positive diffuse large B-cell lymphomas exhibit a poor response to R-CHOP therapy and distinct biological signatures. *Oncotarget* **7**, 52940-52956 (2016).
44. Katoh, M., Igarashi, M., Fukuda, H., Nakagama, H. & Katoh, M. Cancer genetics and genomics of human FOX family genes. *Cancer Lett* **328**, 198-206 (2013).
45. Enard, W. et al. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869-72 (2002).
46. Atkinson, E.G. et al. No Evidence for Recent Selection at FOXP2 among Diverse Human Populations. *Cell* **174**, 1424-1435 e15 (2018).
47. Haesler, S. et al. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol* **5**, e321 (2007).
48. Heston, J.B. & White, S.A. Behavior-linked FoxP2 regulation enables zebra finch vocal learning. *J Neurosci* **35**, 2885-94 (2015).

49. Murugan, M., Harward, S., Scharff, C. & Mooney, R. Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron* **80**, 1464-76 (2013).
50. Shu, W. et al. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proc Natl Acad Sci U S A* **102**, 9643-8 (2005).
51. Fujita, E. et al. Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. *Proc Natl Acad Sci U S A* **105**, 3117-22 (2008).
52. Groszer, M. et al. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr Biol* **18**, 354-62 (2008).
53. Schreiweis, C. et al. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc Natl Acad Sci U S A* **111**, 14253-8 (2014).
54. Butland, S.L. et al. CAG-encoded polyglutamine length polymorphism in the human genome. *BMC Genomics* **8**, 126 (2007).
55. Reid, S.J. et al. TBP, a polyglutamine tract containing protein, accumulates in Alzheimer's disease. *Brain Res Mol Brain Res* **125**, 120-8 (2004).
56. Tomas-Zapico, C. et al. alpha-Synuclein accumulates in huntingtin inclusions but forms independent filaments and its deficiency attenuates early phenotype in a mouse model of Huntington's disease. *Hum Mol Genet* **21**, 495-510 (2012).
57. Orr, H.T. & Zoghbi, H.Y. Trinucleotide repeat disorders. *Annu Rev Neurosci* **30**, 575-621 (2007).
58. Hachigian, L.J. et al. Control of Huntington's Disease-Associated Phenotypes by the Striatum-Enriched Transcription Factor Foxp2. *Cell Rep* **21**, 2688-2695 (2017).
59. Elrod-Erickson, M., Benson, T.E. & Pabo, C.O. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451-64 (1998).
60. Hanas, J.S., Hazuda, D.J., Bogenhagen, D.F., Wu, F.Y. & Wu, C.W. Xenopus transcription factor A requires zinc for binding to the 5 S RNA gene. *J Biol Chem* **258**, 14120-5 (1983).
61. McCarty, A.S., Kleiger, G., Eisenberg, D. & Smale, S.T. Selective dimerization of a C2H2 zinc finger subfamily. *Mol Cell* **11**, 459-70 (2003).
62. Klug, A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* **43**, 1-21 (2010).
63. Pavletich, N.P. & Pabo, C.O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-17 (1991).
64. Chou, C.C., Wei, S.Y., Lou, Y.C. & Chen, C. In-depth study of DNA binding of Cys2His2 finger domains in testis zinc-finger protein. *PLoS One* **12**, e0175051 (2017).
65. Wolfe, S.A., Nekludova, L. & Pabo, C.O. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**, 183-212 (2000).
66. Persikov, A.V. et al. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res* **43**, 1965-84 (2015).
67. Nunez, N. et al. The multi-zinc finger protein ZNF217 contacts DNA through a two-finger domain. *J Biol Chem* **286**, 38190-201 (2011).
68. Li, S., Weidenfeld, J. & Morrisey, E.E. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Mol Cell Biol* **24**, 809-22 (2004).
69. Wang, B., Lin, D., Li, C. & Tucker, P. Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. *J Biol Chem* **278**, 24259-68 (2003).
70. Yu, Y.B. Coiled-coils: stability, specificity, and drug delivery potential. *Adv Drug Deliv Rev* **54**, 1113-29 (2002).
71. Tropsha, A., Bowen, J.P., Brown, F.K. & Kizer, J.S. Do interhelical side chain-backbone hydrogen bonds participate in formation of leucine zipper coiled coils? *Proc Natl Acad Sci U S A* **88**, 9488-92 (1991).

72. Krylov, D., Olive, M. & Vinson, C. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. *EMBO J* **14**, 5329-37 (1995).
73. Glover, J.N. & Harrison, S.C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373**, 257-61 (1995).
74. Landschulz, W.H., Johnson, P.F. & McKnight, S.L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-64 (1988).
75. Estruch, S.B., Graham, S.A., Deriziotis, P. & Fisher, S.E. The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. *Sci Rep* **6**, 20911 (2016).
76. Estruch, S.B. et al. Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. *Hum Mol Genet* **27**, 1212-1227 (2018).
77. Stroud, J.C. et al. Structure of the forkhead domain of FOXP2 bound to DNA. *Structure* **14**, 159-66 (2006).
78. Ostrow, A.Z. et al. Conserved forkhead dimerization motif controls DNA replication timing and spatial organization of chromosomes in S. cerevisiae. *Proc Natl Acad Sci U S A* **114**, E2411-E2419 (2017).
79. Estruch, S.B., Graham, S.A., Chinnappa, S.M., Deriziotis, P. & Fisher, S.E. Functional characterization of rare FOXP2 variants in neurodevelopmental disorder. *J Neurodev Disord* **8**, 44 (2016).
80. Wu, Y. et al. FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell* **126**, 375-87 (2006).
81. Meredith, L.J. et al. The Key Regulator for Language and Speech Development, FOXP2, is a Novel Substrate for SUMOylation. *J Cell Biochem* **117**, 426-38 (2016).
82. Wang, Y.C., Peterson, S.E. & Loring, J.F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res* **24**, 143-60 (2014).
83. Walsh, C.T., Garneau-Tsodikova, S. & Gatto, G.J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-72 (2005).
84. Deribe, Y.L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat Struct Mol Biol* **17**, 666-72 (2010).
85. Radivojac, P. et al. Intrinsic disorder and functional proteomics. *Biophys J* **92**, 1439-56 (2007).
86. Iakoucheva, L.M. et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**, 1037-49 (2004).
87. Xie, H. et al. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res* **6**, 1917-32 (2007).
88. Cohen, P. The origins of protein phosphorylation. *Nat Cell Biol* **4**, E127-30 (2002).
89. Khoury, G.A., Baliban, R.C. & Floudas, C.A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* **1**(2011).
90. Day, E.K., Sosale, N.G. & Lazzara, M.J. Cell signaling regulation by protein phosphorylation: a multivariate, heterogeneous, and context-dependent process. *Curr Opin Biotechnol* **40**, 185-192 (2016).
91. Wurzenberger, C. & Gerlich, D.W. Phosphatases: providing safe passage through mitotic exit. *Nat Rev Mol Cell Biol* **12**, 469-82 (2011).
92. Blane, A., Dirr, H.W. & Fanucchi, S. A Phosphomimetic Study Implicates Ser557 in Regulation of FOXP2 DNA Binding. *Protein J* **37**, 311-323 (2018).
93. Walker, M.P. et al. FOXP1 potentiates Wnt/beta-catenin signaling in diffuse large B cell lymphoma. *Sci Signal* **8**, ra12 (2015).
94. Yang, S. et al. FOXP3 promotes tumor growth and metastasis by activating Wnt/beta-catenin signaling pathway and EMT in non-small cell lung cancer. *Mol Cancer* **16**, 124 (2017).
95. Nusse, R. Wnt signaling and stem cell control. *Cell Res* **18**, 523-7 (2008).

96. Masckauchan, T.N., Shawber, C.J., Funahashi, Y., Li, C.M. & Kitajewski, J. Wnt/beta-catenin signaling induces proliferation, survival and interleukin-8 in human endothelial cells. *Angiogenesis* **8**, 43-51 (2005).
97. Kaldis, P. & Pagano, M. Wnt signaling in mitosis. *Dev Cell* **17**, 749-50 (2009).
98. Micalizzi, D.S., Farabaugh, S.M. & Ford, H.L. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *J Mammary Gland Biol Neoplasia* **15**, 117-34 (2010).
99. Johnson, M.L. & Rajamannan, N. Diseases of Wnt signaling. *Rev Endocr Metab Disord* **7**, 41-9 (2006).
100. Nusse, R. & Clevers, H. Wnt/beta-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985-999 (2017).
101. Logan, C.Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**, 781-810 (2004).
102. Behrens, J. et al. Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature* **382**, 638-42 (1996).
103. Huber, O. et al. Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. *Mech Dev* **59**, 3-10 (1996).
104. Brunner, E., Peter, O., Schweizer, L. & Basler, K. pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in Drosophila. *Nature* **385**, 829-33 (1997).
105. Kaidi, A., Williams, A.C. & Paraskeva, C. Interaction between beta-catenin and HIF-1 promotes cellular adaptation to hypoxia. *Nat Cell Biol* **9**, 210-7 (2007).
106. Essers, M.A. et al. Functional interaction between beta-catenin and FOXO in oxidative stress signaling. *Science* **308**, 1181-4 (2005).
107. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
108. Nelson, C.S. et al. Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets. *Nucleic Acids Res* **41**, 5991-6004 (2013).
109. Tamiola, K., Acar, B. & Mulder, F.A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* **132**, 18000-3 (2010).
110. Hellman, L.M. & Fried, M.G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* **2**, 1849-61 (2007).
111. Sonnichsen, F.D., Van Eyk, J.E., Hodges, R.S. & Sykes, B.D. Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry* **31**, 8790-8 (1992).
112. Eastman, Q. & Grosschedl, R. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr Opin Cell Biol* **11**, 233-40 (1999).
113. Boras-Granic, K., Chang, H., Grosschedl, R. & Hamel, P.A. Lef1 is required for the transition of Wnt signaling from mesenchymal to epithelial cells in the mouse embryonic mammary gland. *Dev Biol* **295**, 219-31 (2006).
114. Tago, K. et al. Inhibition of Wnt signaling by ICAT, a novel beta-catenin-interacting protein. *Genes Dev* **14**, 1741-9 (2000).
115. Daniels, D.L. & Weis, W.I. ICAT inhibits beta-catenin binding to Tcf/Lef-family transcription factors and the general coactivator p300 using independent structural modules. *Mol Cell* **10**, 573-84 (2002).
116. Zhang, N. et al. FoxM1 promotes beta-catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis. *Cancer Cell* **20**, 427-42 (2011).
117. Dai, Y. et al. Loss of FOXN3 in colon cancer activates beta-catenin/TCF signaling and promotes the growth and migration of cancer cells. *Oncotarget* **8**, 9783-9793 (2017).
118. Huber, A.H., Nelson, W.J. & Weis, W.I. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90**, 871-82 (1997).

119. Graham, T.A., Weaver, C., Mao, F., Kimelman, D. & Xu, W. Crystal structure of a beta-catenin/Tcf complex. *Cell* **103**, 885-96 (2000).
120. Eklof Spink, K., Fridman, S.G. & Weis, W.I. Molecular mechanisms of beta-catenin recognition by adenomatous polyposis coli revealed by the structure of an APC-beta-catenin complex. *EMBO J* **20**, 6203-12 (2001).
121. Poy, F., Lepourcelet, M., Shivdasani, R.A. & Eck, M.J. Structure of a human Tcf4-beta-catenin complex. *Nat Struct Biol* **8**, 1053-7 (2001).
122. Xing, Y., Clements, W.K., Kimelman, D. & Xu, W. Crystal structure of a beta-catenin/axin complex suggests a mechanism for the beta-catenin destruction complex. *Genes Dev* **17**, 2753-64 (2003).
123. Sun, J. & Weis, W.I. Biochemical and structural characterization of beta-catenin interactions with nonphosphorylated and CK2-phosphorylated Lef-1. *J Mol Biol* **405**, 519-30 (2011).
124. Morris, G. & Fanucchi, S. A Key Evolutionary Mutation Enhances DNA Binding of the FOXP2 Forkhead Domain. *Biochemistry* **55**, 1959-67 (2016).
125. Haussermann, K., Young, G., Kukura, P. & Dietz, H. Dissecting FOXP2 Oligomerization and DNA Binding. *Angew Chem Int Ed Engl* **58**, 7662-7667 (2019).
126. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
127. Jelier, R. et al. Literature-aided interpretation of gene expression data with the weighted global test. *Brief Bioinform* **12**, 518-29 (2011).
128. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**, D109-14 (2012).
129. Theillet, F.X. et al. Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts. *Nat Protoc* **8**, 1416-32 (2013).
130. Caretta, A. & Mucignat-Caretta, C. Protein kinase a in cancer. *Cancers (Basel)* **3**, 913-26 (2011).
131. Chen, Z. & Cole, P.A. Synthetic approaches to protein phosphorylation. *Curr Opin Chem Biol* **28**, 115-22 (2015).
132. Deriziotis, P. et al. De novo TBR1 mutations in sporadic autism disrupt protein functions. *Nat Commun* **5**, 4954 (2014).
133. Cortese, M.S., Uversky, V.N. & Dunker, A.K. Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol* **98**, 85-106 (2008).
134. Dunker, A.K., Brown, C.J. & Obradovic, Z. Identification and functions of usefully disordered proteins. *Adv Protein Chem* **62**, 25-49 (2002).
135. Oswald, F. et al. The FOXP2-Driven Network in Developmental Disorders and Neurodegeneration. *Front Cell Neurosci* **11**, 212 (2017).
136. Devanna, P., Middelbeek, J. & Vernes, S.C. FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Front Cell Neurosci* **8**, 305 (2014).
137. Vernes, S.C. et al. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet* **7**, e1002145 (2011).
138. Wang, W. et al. FOXKs promote Wnt/beta-catenin signaling by translocating DVL into the nucleus. *Dev Cell* **32**, 707-18 (2015).
139. Turnham, R.E. & Scott, J.D. Protein kinase A catalytic subunit isoform PRKACA; History, function and physiology. *Gene* **577**, 101-8 (2016).
140. Wand, G., Levine, M., Zweifel, L., Schwindinger, W. & Abel, T. The cAMP-protein kinase A signal transduction pathway modulates ethanol consumption and sedative effects of ethanol. *J Neurosci* **21**, 5297-303 (2001).
141. Horiuchi, J., Yamazaki, D., Naganos, S., Aigaki, T. & Saitoe, M. Protein kinase A inhibits a consolidated form of memory in Drosophila. *Proc Natl Acad Sci U S A* **105**, 20976-81 (2008).

6 Appendix

6.1 Protein sequences

6.1.1 FOXP2^{FL}

MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSEVSTVELLHLQQQQALQAARQLLL
QQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQLQALLQQQQAVMLQ
QQQLQEFYKKQQEQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
HPGKQAKEQQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQQHLLSLQRQGLISIPPGQAALPVQSL
PQAGLSPAEIQQLWKEVTGVHSMEDNGIKHGGLDLTTNNSSSTTSSNTSKASPPITHHSIVNGQSSVL
SARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLKHLNNEHALDDRSTAQCRVQMQVVQ
QLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNLVSSVTMSKNMLETSPQSLPQTPTTPTAPVTPI
TQGPSVITPASVPNVGAIRRRHSDKYNIPMSSEIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTL
NEIYSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKRRSQKITGSPTL
VKNIPTSLGYGAALNASLQAALAESSLPLLSNPGLINNASSGLLQAVHEDLNGSLDHIDSNGNSSPGCS
PQPHIHSIHVKEEPVIAEDEDCPMSLVTTANHSPELEDDREIEEEPLSEDLE

6.1.2 FOXP2$^{247-715}$

GQAALPVQSLPQAGLSPAEIQQLWKEVTGVHSMEDNGIKHGGLDLTTNNSSSTTSSNTSKASPPITHH
SIVNGQSSVLSARRDSSSHEETGASHTLYGHGVCKWPGCESICEDFGQFLKHLNNEHALDDRSTAQ
CRVQMQVVQQLEIQLSKERERLQAMMTHLHMRPSEPKPSPKPLNLVSSVTMSKNMLETSPQSLPQT
PTTPTAPVTPITQGPSVITPASVPNVGAIRRRHSDKYNIPMSSEIAPNYEFYKNADVRPPFTYATLIRQAI
MESSDRQLTLNEIYSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKR
RSQKITGSPTLVKNIPTSLGYGAALNASLQAALAESSLPLLSNPGLINNASSGLLQAVHEDLNGSLDHID
SNGNSSPGCSPQPHIHSIHVKEEPVIAEDEDCPMSLVTTANHSPELEDDREIEEEPLSEDLE

6.1.3 FOXP2$^{342-714}$

LYGHGVCKWPGCESICEDFGQFLKHLNNEHALDDRSTAQCRVQMQVVQQLEIQLSKERERLQAMMT
HLHMRPSEPKPSPKPLNLVSSVTMSKNMLETSPQSLPQTPTTPTAPVTPITQGPSVITPASVPNVGAIR
RRHSDKYNIPMSSEIAPNYEFYKNADVRPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNA
ATWKNAVRHNLSLHKCFVRVENVKGAVWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGAALNASL
QAALAESSLPLLSNPGLINNASSGLLQAVHEDLNGSLDHIDSNGNSSPGCSPQPHIHSIHVKEEPVIAE
DEDCPMSLVTTANHSPELEDDREIEEEPLSEDLE

6.1.4 FOXP2$^{504-715}$

RPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVRVENVKGA
VWTVDEVEYQKRRSQKITGSPTLVKNIPTSLGYGAALNASLQAALAESSLPLLSNPGLINNASSGLLQA
VHEDLNGSLDHIDSNGNSSPGCSPQPHIHSIHVKEEPVIAEDEDCPMSLVTTANHSPELEDDREIEEEP
LSEDLE

6.1.5 FOXP2^{IDR}

GQAALPVQSLPQAGLSPAEIQQLWKEVTGVHSMEDNGIKHGGLDLTTNNSSSTTSSNTSKASPPITHH
SIVNGQSSVLSARRDSSSHEETGASHT

6.1.6 FOXP2$^{IDR_\Delta 264-272}$

GQAALPVQSLPQAGLSPTGVHSMEDNGIKHGGLDLTTNNSSSTTSSNTSKASPPITHHSIVNGQSSVL
SARRDSSSHEETGASHT

6.1.7 FOXP2^{S330E}

GQAALPVQSLPQAGLSPAEIQQLWKEVTGVHSMEDNGIKHGGLDLTTNNSSSTTSSNTSKASPPITHH
SIVNGQSSVLSARRDESSHEETGASHT

6.1.8 FOXP2^{FH}

RPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVRVENVKGA
VWTVDEVEYQKRRSQKITGSPTL

6.1.9 FOXP2$^{FH\text{-}IDR}$

RPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAATWKNAVRHNLSLHKCFVRVENVKGA
VWTVDEVEYQKRRSQGGSGGSGGSGGSQAGLSPAEIQQLWKEVTGVHS

6.1.10 FOXP2^{R553H}

RPPFTYATLIRQAIMESSDRQLTLNEIYSWFTRTFAYFRRNAATWKNAVHHNLSLHKCFVRVENVKGA
VWTVDEVEYQKRRSQKITGSPTL

6.1.11 β-catenin^{FL}

MATQADLMELDMAMEPDRKAAVSHWQQQSYLDSGIHSGATTTAPSLSGKGNPEEEDVDTSQVLYE
WEQGFSQSFTQEQVADIDGQYAMTRAQRVRAAMFPETLDEGMQIPSTQFDAAHPTNVQRLAEPSQ
MLKHAVVNLINYQDDAELATRAIPELTKLLNDEDQVVVNKAAVMVHQLSKKEASRHAIMRSPQMVSAI
VRTMQNTNDVETARCTAGTLHNLSHHREGLLAIFKSGGIPALVKMLGSPVDSVLFYAITTLHNLLLHQE
GAKMAVRLAGGLQKMVALLNKTNVKFLAITTDCLQILAYGNQESKLIILASGGPQALVNIMRTYTYEKLL
WTTSRVLKVLSVCSSNKPAIVEAGGMQALGLHLTDPSQRLVQNCLWTLRNLSDAATKQEGMEGLLGT
LVQLLGSDDINVVTCAAGILSNLTCNNYKNKMMVCQVGGIEALVRTVLRAGDREDITEPAICALRHLTS
RHQEAEMAQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAIPRLVQLLV
RAHQDTQRRTSMGGTQQQFVEGVRMEEIVEGCTGALHILARDVHNRIVIRGLNTIPLFVQLLYSPIENI
QRVAAGVLCELAQDKEAAEAIEAEGATAPLTELLHSRNEGVATYAAAVLFRMSEDKPQDYKKRLSVEL
TSSLFRTEPMAWNETADLGLDIGAQGEPLGYRQDDPSYRSFHSGGYGQDALGMDPMMEHEMGGH
HPGADYPVDGLPDLGHAQDLMDGLPPGDSNQLAWFDTDL

6.1.12 β-catenin$^{1\text{-}140}$

MATQADLMELDMAMEPDRKAAVSHWQQQSYLDSGIHSGATTTAPSLSGKGNPEEEDVDTSQVLYE
WEQGFSQSFTQEQVADIDGQYAMTRAQRVRAAMFPETLDEGMQIPSTQFDAAHPTNVQRLAEPSQ
MLKHAVVNLI

6.1.13 β-catenin$^{141\text{-}305}$

NYQDDAELATRAIPELTKLLNDEDQVVVNKAAVMVHQLSKKEASRHAIMRSPQMVSAIVRTMQNTND
VETARCTAGTLHNLSHHREGLLAIFKSGGIPALVKMLGSPVDSVLFYAITTLHNLLLHQEGAKMAVRLA
GGLQKMVALLNKTNVKFLAITTDCLQILA

6.1.14 β-catenin$^{666\text{-}781}$

KPQDYKKRLSVELTSSLFRTEPMAWNETADLGLDIGAQGEPLGYRQDDPSYRSFHSGGYGQDALGM
DPMMEHEMGGHHPGADYPVDGLPDLGHAQDLMDGLPPGDSNQLAWFDTDL

6.1.15 LEF$^{1\text{-}299}$

MPQLSGGGGGGGGDPELCATDEMIPFKDEGDPQKEKIFAEISHPEEEGDLADIKSSLVNESEIIPASN
GHEVARQAQTSQEPYHDKAREHPDDGKHPDGGLYNKGPSYSSYSGYIMMPNMNNDPYMSNGSLSP
PIPRTSNKVPVVQPSHAVHPLTPLITYSDEHFSPGSHPSHIPSDVNSKQGMSRHPPAPDIPTFYPLSPG
GVGQITPPLGWQGQPVYPITGGFRQPYPSSLSVDTSMSRFSHHMIPGPPGPHTTGIPHPAIVTPQVK
QEHPHTDSDLMHVKPQHEQRKEQEPKRPHI

6.1.16 LEF$^{288-399}$

QRKEQEPKRPHIKKPLNAFMLYMKEMRANVVAECTLKESAAINQILGRRWHALSREEQAKYYELARK
ERQLHMQLYPGWSARDNYGKKKKRKREKLQESASGTGPRMTAAYI

6.1.17 ICAT

MNREGAPGKSPEEMYIQQKVRVLLMLRKMGSNLTASEEEFLRTYAGVVNSQLSQLPPHSIDQGAEDV
VMAFSRSETEDRRQ

6.2 Top 10 up/downregulated genes of RNA Seq data of each condition compared to control, log2 fold change in brackets

	Upregulation	Downregulation
FOXP2	FOXP2 (13)	PRND (-10.5)
	MMP1 (11.3)	KANK4 (-9.3)
	LPAR6 (11.2)	HTR2A (-9.2)
	CTD-2532D12.4 (9.4)	SOSTDC1 (-8.9)
	MYCT1 (9.3)	AC074289.1 (-8.5)
	IL24 (9)	MYOCD (-8.1)
	ENAM (9)	CXCL14 (-8.1)
	AF121898.3 (9)	FOXS1 (-8)
	SPRR2D (8.6)	RP11-798L4.1 (-7.9)
	MAL (8.6)	ABCB5 (-7.9)

	Upregulation	Downregulation
CHIR	GZMB (12)	MYOCD (-9.2)
	MUCL1 (11.4)	FOXS1 (-8.8)
	CXCL6 (10.7)	NWD1 (-8.5)
	MTHFD2P1 (10.7)	GALNt15 (-8.3)
	C1orf168 (10.1)	CD180 (-8.1)
	LINC00161 (9.9)	YPEL4 (-7.9)
	RP11-415C15.2 (9.8)	SYT8 (-7.7)
	RP11-95P13.3 (9.8)	HTR2A (-7.2)
	CST1 (9.6)	TNFSF15 (-7.2)
	FILIP1 (9.5)	CTD-2334D19.1 (-6.8)

	Upregulation	Downregulation
FOXP2$^{\Delta helix}$	FOXP2 (12.6)	PRND (-9.7)
	MAL (11.4)	PLP1 (-9.7)
	LPAR6 (11)	KANK4 (-9.1)
	MMP1 (10.9)	SERPINB2 (12)
	MUC5AC (9.7)	CCDC67 (-8)
	AGXT (9.5)	CXCL14 (-7.8)
	MYCT1 (9.4)	RP11-798L4.1 (-7.7)
	CTD-2532D12.4 (9.1)	GABRR1 (-7.7)
	LINC00659 (9)	AC074289.1 (-7.6)
	SRL (8.9)	MYOCD (-7.6)

	Upregulation	Downregulation
FOXP2 + CHIR	FOXP2 (14)	PRND (-10.5)
	LPAR6 (13)	SOSTDC1 (-10)
	MYCT1 (12.1)	IGFBP5 (-9.1)
	SERPINB2 (12)	PLP1 (-8.2)
	RP11-753N8.1 (11.9)	HTR2A (-8)
	SPRR2D (11.3)	CXCL14 (-8)
	CTD-2532D12.4 (11.2)	SLC2A12 (-8)
	MAL (11.2)	AC074289.1 (-7.9)
	MMP1 (10.8)	R3HDML (-7.8)
	GZMB (10.8)	GALNT15 (-7.6)

	Upregulation	Downregulation
FOXP2$^{\Delta helix}$ + CHIR	FOXP2 (13.8)	PRND (-9.2)
	MAL (12.7)	SOSTDC1 (-8.7)
	LPAR6 (12.6)	IGFBP5 (-8.7)
	MYCT1 (11.9)	ABCB5 (-8.3)
	RP11-753N8.1 (11.9)	APOL4 (-8.2)
	MUC5AC (11.6)	GABRR1 (-7.7)
	AGXT (11.5)	RP1.193H18.2 (-7.7)
	GZMB (11.5)	CXCL14 (-7.7)
	SPRR2D (10.7)	FABP7 (-7.5)
	MUCL1 (10.5)	KIAA1210 (-7.5)

7 Acknowledgements

Hereby I want to thank the following people, who either contributed scientifically to this work, made the last years memorable or supported me in any other way:

• Prof. Dr. **Tobias** Madl who gave me the opportunity to do my PhD in his lab and provided the interesting and challenging projects. I would like to thank him for giving me the opportunity to undertake this project, with continued support, lively discussions and advice throughout my whole doctoral period. Thanks to him I was able to participate in different conferences and present my results.

• All people at the TUM during the time, the Madl group was taking most spectrometer time ☺ thanks to **Michael, Gerd, Ralf** and all other members of the Sattler group, you helped me throughout my master thesis with smaller and bigger problems. Thanks to **Gülden, Steffi and Hannah**, it was such as pleasure to work with you, beside all discussion during lunch.

• Many thanks to the **Madl group** for the awesome working atmosphere, for the exciting discussions about our projects and our after work activities. It was such a great time in Graz, you are awesome!

• **Jennifer** Herrmann und **Juliane** Mayr as internship student for their help and great work.

• to **Sarah**: You were the best office neighbor one could imagine. We went both through tough times, always supporting each other. You are amazing and will find your way, doesn't matter where and how ☺. **Julia**: It was such a great time with you, not only in the lab, but also during all the time we had out of lab. **Therese and Anna**: thank you so much, not only for help in the lab, but also for the lovely discussions. To all those girls: I really hope, we never loose sight of each other!

• to my family: **Jon, Laurids und Papa**, without you this would not have been possible. Thanks for the support during this time.

• to **Ben**: we met at my very first meeting in the Madl AG, and you haven´t been only my supervisor in my master thesis but an awesome partner in lab & life. Thanks for the all-day and all-night support, the immense patience with me and your love ♥