



DISSERTATION

Titel der Dissertation

„Chemistry and NMR studies of Beta-Catenin-
TCF4 interaction – a potential antitumor target“

Verfasser

Morkos A. Henen, M.Sc.

angestrebter akademischer Grad

Doktor der Naturwissenschaften

Wien, 2012

Studienkennzahl lt.
Studienblatt:

A 094 490

Dissertationsgebiet lt.
Studienblatt:

Molekulare Biologie

Betreuerin / Betreuer:

Professor Dr. Robert Konrat

To the Soul of the Peaceful Warrior; My Father
To My Mother
To Matta, Mariam & George
To My Angel; Joy

Acknowledgements

1. I would like to express my deep thanks to **Prof. Robert Konrat** for supervising this thesis. With his support and caring about every single item, I have learnt a lot. His continuous encouragement and optimism make the lab a friendly place to be in every day.
2. I also acknowledge **Prof. Georg Kontaxis** for his help especially in explaining the secrets of NMR techniques and related software. He always has such creative ideas that one can never imagine.
3. I'm also very appreciative of the help of **Dr. Nicolas Coudeville**. His amazing guidance, especially on how to process and how to interpret the spectra, was very valuable throughout the whole research project.
4. My gratitude also goes to **Karin Ledolter** or (lab mother) as we all used to call her. She offered me the first guide in the protein expression and the use of lab infrastructure. Talking and discussing with her, even outside the work, made things much easier for me
5. Saying thanks to the best friend and lab colleague; **Leonhard Geist**, won't be enough. His competence particularly in cloning and python scripting was a big addition for me. Apart from the lab, he was the one that I can confide in and complain to about work or private problems.
6. My dear friends in Konrat group were always such a great aid for me; **Gerald Platzer, Thomas Schwarz, Gonul Kizilsavas** and **Sandra Haiderer**. We used to have fruitful and enjoyable talks on daily basis over different obstacles we face in the lab.
7. I'm also admiring the help of **Kristina Djinovic group members** in the crystallization work and their advices.
8. I appreciate so much the help of **Dr. Hanspeter Kählig**, the organic chemistry institute, in doing the ^{19}F -NMR experiments.

Eventually, I would like to thank **OeAD** - Österreichischer Austauschdienst for their financial support.

Table of contents

1. Introduction

1.1. Chemistry of protein	8
1.2. Protein domain	11
1.3. IDPs	11
1.4. Meta structure	13
1.5. Protein structure elucidation	15
1.6. Protein ligand interaction	19
1.7. Fragment based drug design	21
1.8. β -Catenin & TCF4	23

2. Theoretical

2.1. NMR screening in drug discovery	31
2.2. NMR experiments adopted during this investigation	32
2.3. Dynamic combinatorial library	48
2.4. ITC	50
2.5. Protein expression	52

3. Experimental

3.1. Material	59
3.2. Methods, Results and Discussions	70

4. References	133
----------------------	------------

Abstract

Aberrant Wnt/Beta-Catenin signaling is observed in a wide spectrum of human malignancies. Activation of this pathway is involved in colon and ovarian cancer. Therefore, inhibitors of the TCF4/Beta-Catenin interaction are of great potential as antitumor agents. We used the recently described protein meta-structure approach to identify novel small molecule compounds as binders for Beta-Catenin. The identified ligands have been tested for binding with the target protein using 1D ^1H -STD, 1D AFP NOESY, 2D NOESY and ^{19}F -NMR. The identified binders were demonstrated to compete with the authentic Beta-Catenin binding partner TCF4.

In addition, a dynamic combinatorial library has been designed to synthesize new ligands for Beta-Catenin. Moreover, 2D & 3D-NMR experiments were used to assign the backbone signals of full length TCF4 and two different truncations at different pH values. The data obtained from assignment at pH=4.5 and pH=2 were compared with the predicted random coil shifts. Secondary structure propensities were also compared which allowed us to estimate changes in secondary structure propensities upon pH change. Finally, further structural insight into the TCF4/Beta-Catenin complex was provided by a combined X-ray crystallography and NMR study employing selective lysine reductive-methylation reaction. To highlight the importance of our work, the resulting prospects for the development of new powerful anticancer drugs with a high safety profile will also be discussed.

INTRODUCTION

Introduction

1.1. Chemistry of Proteins

Proteins are polypeptides of varying length consist of amino acids linked together *via* peptide bond. The binding between two amino acids through the addition-elimination interaction between the carboxylic group of one amino acid and the amino group of the other generates the peptide bond. Peptide bond is characterized of being planar due to amide resonance with only 2 degrees of freedom (phi and psi torsion angles).

Proteins consist of one or more polypeptide chains. The number and sequence of amino acids differs from one protein to the other.

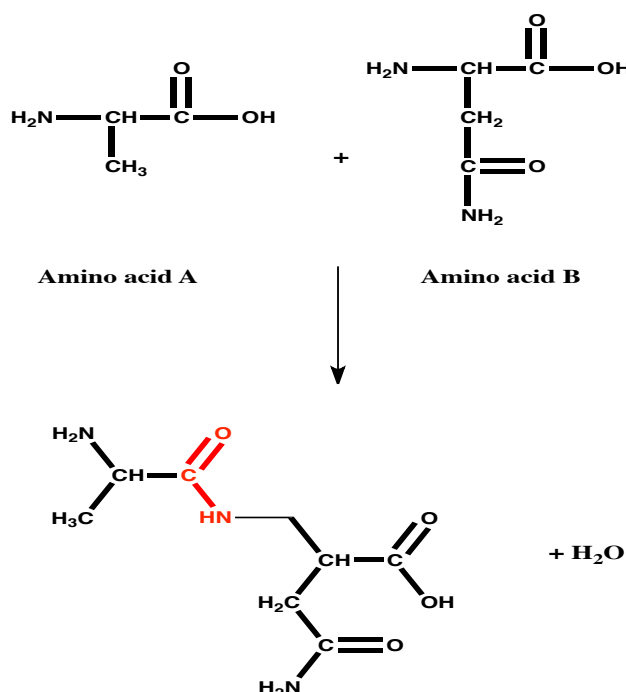


FIG. 1 Peptide bond formation.

Proteins are complex macromolecules and to account for this enormous complexity, Linderstrom-Lang¹ described his classical schemes in which he proposed a four levels system for the protein structure: primary, secondary, tertiary and quaternary structure.

a- Primary structure gives just the linear amino acid sequence. In a study by Sobolevsky *et. al.*², they demonstrated the presence of conserved octamers in most of the sequenced genomes. In the species where these conserved octamers aren't found, analogs of them with one or two residues different are detected. Based on this, it could be predicted that the conserved sequence and the modules they represent,

Introduction

should serve as a basis for molecular reconstruction of the last common ancestor of modern species.

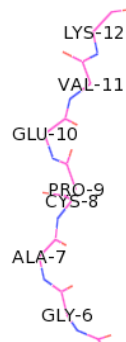


FIG. 2 Protein primary structure.

b- The secondary structure is the next ‘level up’ in protein structure that arises from the primary structure. Hydrogen bonds play the main role in the type of secondary structure that a protein possesses. The main two secondary elements of proteins are; α -helix and β -sheet. In α -helix³, each N-H group forms a hydrogen bond to the backbone C=O of an amino acid that is 4 residues earlier. This H-bonding helps in directing many polypeptides to take the cylindrical shape that characterizes this secondary structure.

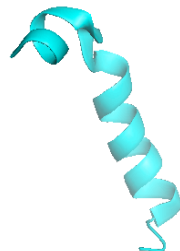


FIG. 3 Protein α -helix secondary structure.

In β -sheet, the chain of amino acids in the polypeptide folds back and forth upon itself instead of forming a helix. This type of folding exposes hydrogen atoms, which can lead to extensive mesh of hydrogen bonding.

Introduction

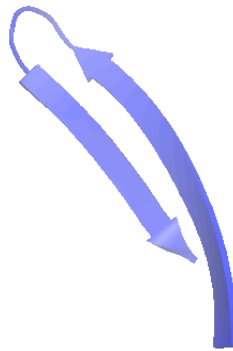


FIG. 4 Protein β -sheet secondary structure showing 2 antiparallel beta-strands.

c- The tertiary structure is the native (3D) conformation and the overall folding pattern of the protein. Once the polypeptide has a definite secondary structure, the side chain of each amino acid is positioned in a certain way. Hence, if additional chemical bonds (e.g. hydrogen bonds, covalent bonds or other atomic interactions) are taking place, the polypeptide chain will accommodate to this and gives a unique 3D structure.

The tertiary folding of the polypeptide forms exposed regions or grooves in the molecule, which may be involved in the binding sites for ligand or other protein.

In 1976, Levitt and Chothia⁴ studied the structure of different 31 proteins and showed that the arrangement of secondary structure units is statistically significant and these secondary units when adjacent in the sequence, they tend to form few number of specifically defined folding units.

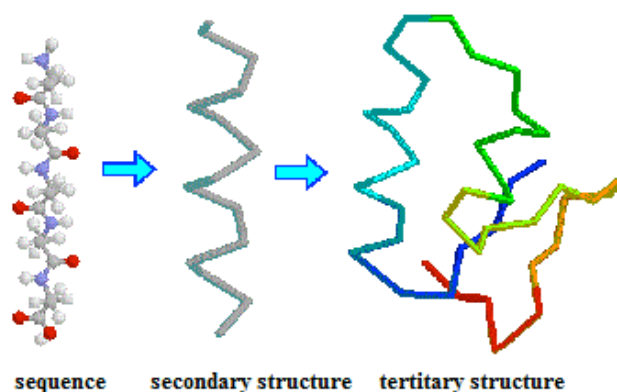


FIG. 5 Protein from primary to secondary to tertiary structure. (Figure is adapted from web site: <http://www.cs.nott.ac.uk/~nxk/TEACHING/G6DHLL/COURSEWORK/2003-2004/courseworks.htm>).

Introduction

d- If the protein consists of more than one polypeptide, the organization of protein subunits to form an oligomeric arrangement is called the quaternary structure (e.g., hemoglobin structure⁵). The polypeptide subunits bind together either by non-covalent bonding (mainly hydrophobic interaction) or by covalent linkages.

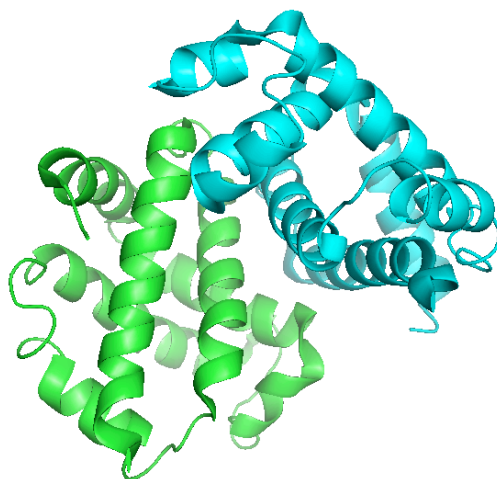


FIG. 6 Quaternary structure of hemoglobin protein. (PDB entry 2DHB).

1.2. Protein Domain

Several polypeptide chains fold into two or more parts; each displays α -helix or β -sheet secondary structure element. These segments are called domains which represent the regions of the polypeptide which can function, and exist independently of the rest of the protein⁶. One protein can have several domains and one domain can appear in a variety of different proteins like armadillo repeat domain in catenin family.

1.3. Intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) account for a large scale of eukaryotic proteins, as it's believed that 50% of these proteins contain disordered areas of more than 30 amino acids⁷. In addition, it was found by Fong *et al*⁸ that on average three quarters of the protein complexes have as much as one third of their residues in the unstructured state. IDPs play usually a major role in the regulatory processes of translation, transduction and cell signaling and they are found predominantly in the

Introduction

cell nucleus. While naturally folded proteins have global energy minima, the IDPs fail to adopt one single conformation or defined 3D structure⁹. This unique nature of the IDPs is mainly related to the low content of hydrophobic residue chains in the sequence, which leads to the absence of a folded core in the IDPs. On the other hand, they are rich with high content of polar and charged residues.

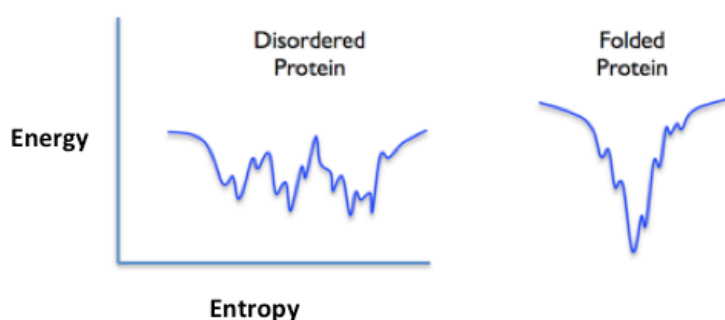


FIG. 7 Difference between IDPs and folded proteins in energy minima. While folded ones have global energy minima, IDPs sample different conformations without definite 3D structure.

Due to this sequence difference, the IDPs have intrinsic lower tendency to form aggregates than denatured globular proteins although some IDPs are associated with amyloid formation diseases. As most of the globular proteins tend to aggregate when there's change in the solution condition or disruption of the sequence, the IDPs have naturally evolved to avoid amyloid formation¹⁰.

Although lacking a definite 3D structure, the IDPs even in the highly unfolded state possess secondary structure propensity¹¹. It was shown by Marsh *et al*⁷ that in the IDP family of synuclein, α -synuclein protein, which is prone to amyloid formation and highly related to Parkinson's disease and amyloid plaques of Alzheimer disease, has 34% difference in the sequence in the amyloid formation region with the γ -synuclein that forms only fibrils at much higher concentration and slower rate. This difference in tendency to form amyloid between the two proteins was rationalized based on the difference in the secondary structure propensity between them. α -Synuclein has β -structure propensity in the amyloid-forming region, while γ -synuclein has a tendency to form α -helical structure, which could possibly stabilize the intermediate state and prevents oligomerization in comparison to the β -structure intermediate in α -synuclein. Many IDPs move to more ordered or folded state upon binding to their targets and usually this process to be enthalpic driven⁹. Due to the high capture radius of these

Introduction

proteins for specific binding site, they weakly bind to the targeted protein at large distance and then they start to fold as they approach the binding site (fly-casting mechanism) which provides a kinetic advantage¹².

1.4. Protein Meta-Structure Concept

The knowledge about the protein chemistry and structure is crucial for different fields like health care and biotechnology. The main reason for developing meta-structure concept is to have the ability to predict protein structure and possible ligands based only on sequence (without knowledge of the secondary or tertiary structure of the protein).

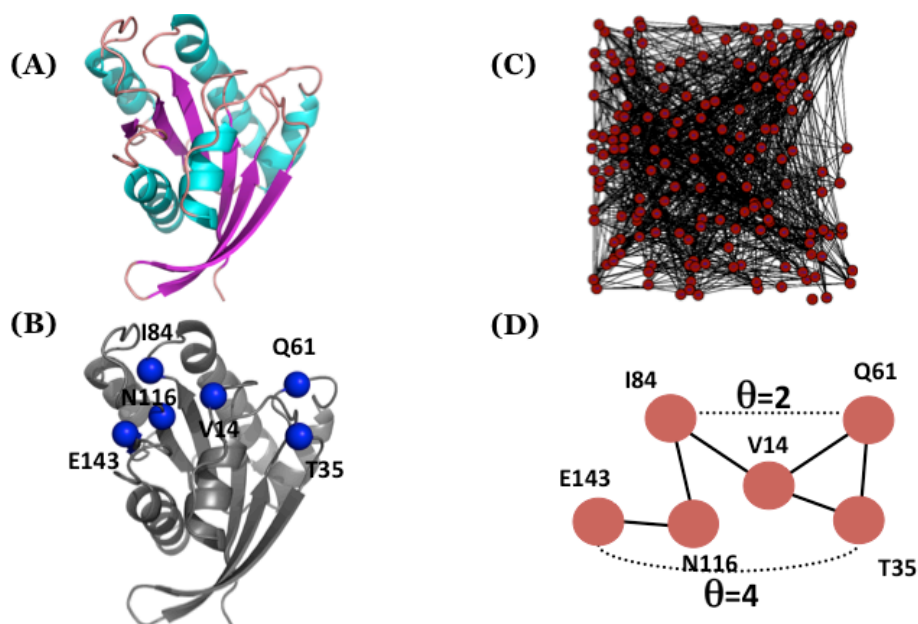


FIG.8 The protein meta-structure concept. The 3D structure is converted into a topological map by calculating the network of amino acids interactions. (a) 3D structure of a protein (b) Location of the network including residues Q61–T35–V14–I84– N116–E143. (c) Topological map of calculated from the 3D structure (d) Extraction of the parameter (shortest path length θ) between residues. (Figure is adapted from Konrat¹³).

Meta-structure concept was developed by Prof. Robert Konrat¹³ in 2009. In this technique, 3D structure of different protein were taken from PDB, and each 3D is converted into a topological map in which each node represent an amino acid and each edge indicates a neighborhood between the two residues (neighborhood means

Introduction

that the distances between the $C\alpha$ - $C\alpha$ are less than 8 \AA). From this topological map, the shortest path length connecting two residues in the network quantifies a topological parameter called θ . This parameter depends on the primary sequence distance between the two amino acids and their nature. Parameter θ is statistically evaluated using the PDB¹⁴, and stored as a pairwise distribution functions.

Parameter θ is used then alone to calculate the two Meta-structure parameters; compactness and secondary structure. Compactness (C_i) is calculated and plotted against the residue position. Large compactness values indicate stable part of the protein or deeply buried domain in the structure. While small values indicate flexible loop or unstructured region. On the other hand, Secondary structure parameter (S_i) is also calculated and plotted against the residue position. Typically, residues located in α -helices secondary structures give positive S_i values while residues located in β -sheets give negative S_i values.

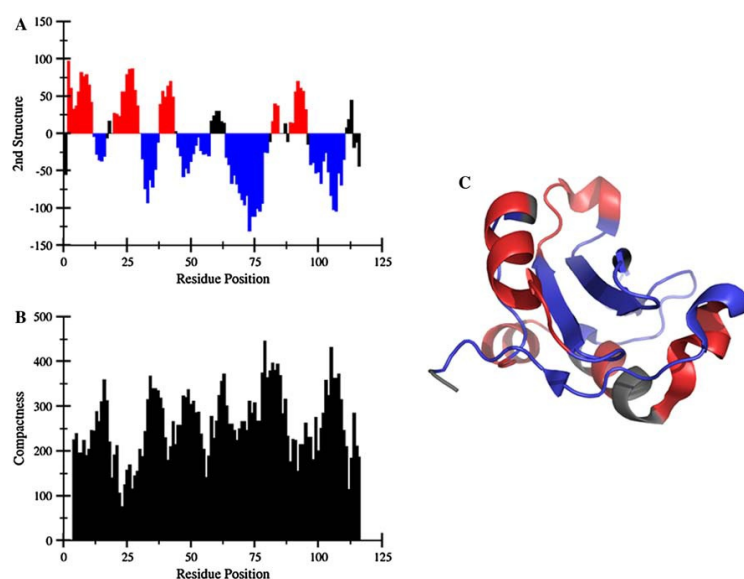


FIG. 9 The two protein meta-structure parameters. Residue secondary structure and compactness plot of the PI3-kinase p85 *N*-terminal SH2 domain. (PDB: 2IUG). Comparison of the calculated local secondary structural features (a) and compactness (b) as a function of residue position, and 3D protein structure (c). In the 3D structure of the SH2 domain (c) residues are color coded according to the meta-structure calculated local secondary structure results (α -helix: red; β -sheet: blue; loosely define structure: black). (Figure is adapted from Konrat¹³).

1.5. Protein Structure elucidation

Many techniques are adopted to study different aspects of proteins structure, but two techniques allow a resolution at the level of distinguishing individual atoms: *X-ray crystallography* and *Nuclear Magnetic Resonance* technique (NMR).

Early in the 20th century, X-ray crystallography has been applied to determine the structure of organic and inorganic crystals. The start was with the explication of the structure of salt crystal, which for example allowed Linus Pauling to study atomic distances from which he developed his theory of the chemical bond. Based on the data acquired from salt crystals, Pauling introduced the hypothesis of the alpha helical and beta strand secondary structures¹⁵. Both of which have been confirmed by X-ray crystallographic analysis for the first time exploiting crystals of myoglobin in the early 60ies by Kendrew and co-workers¹⁶.

Loose protein structure parts represent a challenge for the X-ray technique. The introduction of nuclear magnetic resonance techniques; NMR, can be used to overcome this problem. In contrast to protein crystals, which are crucial for X-ray diffraction, liquid-state NMR can be used to detect structure and dynamic of protein solution. Hence, these flexible structures could be resolved successfully. On the other hand, NMR has the problem of size limitation as large proteins are being considered a challenge for NMR due to fast relaxation rates.

Moreover, with the technique of cryo-electron microscopy (a technique in which sample is spread as a thin film and frozen), complexes with resolution as small as 5-10Å can be detected. One of the main advantages of this technique is that the sample isn't stained or fixed in anyway which means that the protein is visualized in its native form in contrast to X-ray technique in which the sample is fixed in the form of crystal.

1.5.1. *X-ray crystallography*

X-ray crystallography visualizes the pattern of diffraction of X-rays that are shot through the crystal. The recorded diffraction pattern is a result of the interaction between X-rays and the electrons in the different atoms. The crystal works as an

Introduction

amplifier because of the crystal lattice, which is packed with thousands of molecules and gives a unique diffraction pattern that is used to extract information about the structure.

Growth of protein crystal sometimes is a tricky process; there's no definite rule for the conditions at which the protein can crystallize but the whole process depends on trial and error, then some optimization. Polyethylene glycol is a widely used precipitant in the crystallization because of its efficiency and weak denaturation properties.

Noteworthy, membrane proteins always tend to precipitate out of the solution due to unfavored interaction and this's considered as a challenge for the crystallography technique. To overcome this and keep them soluble in aqueous solution, addition of detergents to membrane proteins is required, but on the other hand, addition of these detergents could interfere with the normal arrangement of the protein complex resulting in unexpected diffraction pattern.

The diffraction pattern can be converted into a 3D structure by creating an electron density map. The mathematical formula to do this is called *Fourier transformation* of the structural factor F_{hkl} . The phase problem in X-ray technique is considered as an obstacle because the location of the electrons can't be extracted from the diffraction pattern.

The three approaches that are used for phase solving include the following:

a- Multi-wavelength anomalous dispersion (MAD)

Elements absorb X-ray as well as emitting them, but this absorption drops sharply at λ just below their characteristic emission wavelength. This sudden change in absorption as function of λ is called absorption edge. The absorption edge for light atoms as C, H or O isn't near the wavelength used in normal X-ray wavelengths, hence, they all scatter normally. While in case of heavy atoms they scatter anomalously in this wavelength range used. As a result, phase delay is observed. For example, a heavy-atom derivative of methionine called Selenomethionine can be

Introduction

incorporated into the protein structure *via* the use of selective media during protein expression. Anomalous diffraction is then recorded at a synchrotron facility.

b- Multiple Isomorphous replacement (MIR)

In this techniques the crystals are soaked in a heavy atom but with one condition that addition of this heavy atom doesn't change more than 1% of the unit cell. Afterwards, Patterson difference map between native and heavy atom containing crystal reveals the location of the heavy atom and hence we can get the amplitude and phase. It's noteworthy that these procedures should be done with 2 different isomorphous derivatives otherwise we could have two possibilities for the phase.

c- Molecular replacement (MR)

Herein, the phase from structural factors of a known similar structure is used as initial estimation of the phase of our new protein. Similarity between the reference protein and the new one should be not less than 25%.

Protein initial models have to be refined. This is often accomplished by comparing the experimental data with the calculated data. The difference between the experimental structure and the calculated one is given as *R-factor*. R-factor is a measure of the difference between the calculated and observed structural factors of the built model and it's used to validate the model.

1.5.2. Nuclear magnetic resonance (NMR)

NMR provides a highly sensitive way to detect protein structure. It differs from X-ray in that it measures mainly *distances* between atomic nuclei (not electron density). With NMR, a strong, high frequency magnetic field affects atomic nuclei (spins) that have magnetic spin (^1H , ^{13}C , or ^{15}N). As a result of applying this strong magnetic field, some spins start to precess aligned with the magnetic field (α) and others aligned against the field (β). The difference between the populations of these two states is what creates Z-magnetization (Boltzmann equilibrium magnetization). Radiofrequency is then applied to take this magnetization to the transverse plane where the magnetization starts to resonate at certain frequency, which creates an

Introduction

oscillatory current that is detected as time domain and then it's fourier transformed into a frequency domain. The distance and type of neighboring nuclei determines the resonance frequency of the stimulated atomic nuclei. This dependence on next neighbors is parameterized by chemical shift and is considered as a parameter for the chemical environment of each spin. Different sets of NMR experiments have been designed and used starting from 1D to 6D experiments, which use the network connection between different nuclei to transfer magnetization *via* bond or *via* space. NMR can also be used as a powerful technique for detection and quantification of protein-ligand interaction; this will be discussed in details through different parts in this thesis.

The following advantages make NMR spectroscopy a method of choice for biological macromolecules:

- Protons are the most abundant spins in the protein structure.
- It's possible to label carbon and nitrogen atoms of the protein to do multidimensional NMR.
- NMR gives an idea about the dynamic of the protein, which is not the case on using x-ray crystallography.

The limiting criterion in protein NMR is the size of the protein. Currently, the size limit for proteins size eligible for solution NMR is 40 KDa.

In addition, although liquid-state NMR doesn't require having the protein in ordered crystal lattice form, it's demanding in the sense that it needs high concentrations of soluble protein (hence, NMR-derived structures are called *solution structures*)

Introduction

1.6. Protein-ligand interaction

Protein-ligand interaction is a key process required to perform diverse of biological functions, hence, interfering with this interaction is a therapeutic concept. Ligand is a small molecule that binds a biomolecule as protein to perform a specific biological function. The binding occurs by intermolecular forces, for example, hydrogen bonds, ionic bonds and van der Waals forces, this association is usually reversible (dissociation). Association between ligand and the biomolecule leads sometimes to conformational changes and this leads to modification of the function of the biomolecule. The interaction between a protein and a ligand is usually specific, so the protein can discriminate between many different molecules and bind only one particular molecule or one of a number of very closely related (i.e. chemically and structurally similar) molecules. Protein–ligand interactions range from weak and transient to strong and persistent, depending on the strength of the non-covalent bonds.

To describe the affinity between a protein and ligand, the term dissociation constant (K_D) is used¹⁷. K_D is the molar concentration of the ligand at which half of a specific site on the protein is occupied. The lower the K_D the higher the affinity between the protein and the ligand. Dissociation constant can be determined by several methods, one of the most important techniques is isothermal titration calorimetry¹⁸.

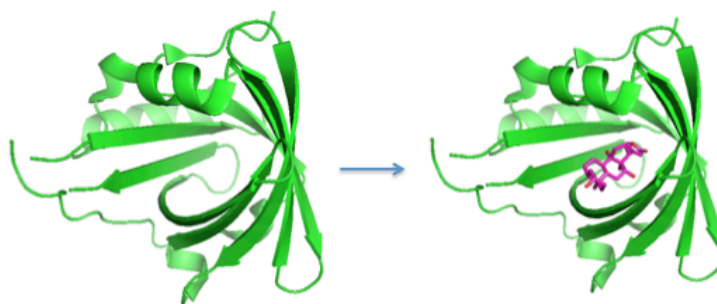


FIG. 10 Ligand binding site, which is often formed in a cleft or a pocket on the protein surface. Protein is shown in green while ligand is shown in magenta.

Finding ligands for the targeted biomolecule can be achieved *via* different methods.

Examples of known methods are:

Introduction

1-Protein-structure-similarity clustering (PSSC): it's used to investigate potential ligands for specific protein¹⁹. In this technique, proteins that share the similar structure rather than sequence identity are the ones of interest, and hence, it serves as a guiding mechanism to select natural products to target structurally similar proteins. The main disadvantage of this technique is that the 3D structure of the binding motif should be conserved for comparison, but the lack of 3D information about a lot of proteins that have pharmacological interest stands as a limitation for using this method.

2- Recently, meta-structure approach can be used to design a library of different fragments based on the similarity between the targeted protein and other proteins in the database¹³. The meta-structure parameters are used as a powerful comparison tool between proteins based on sequence. Only by knowing the sequence of the protein, it's possible to detect the meta-structure parameters, which are then aligned with other proteins in the database to predict structure and possible ligands by doing protein-Meta-Structure similarity clustering (PMSSC). As it depends only on the sequence, so it can be used even for proteins that have unknown 3D structure and of biological relevance. Figure 11 shows the difference between the use of PSSC and PMSSC for finding possible protein ligands.

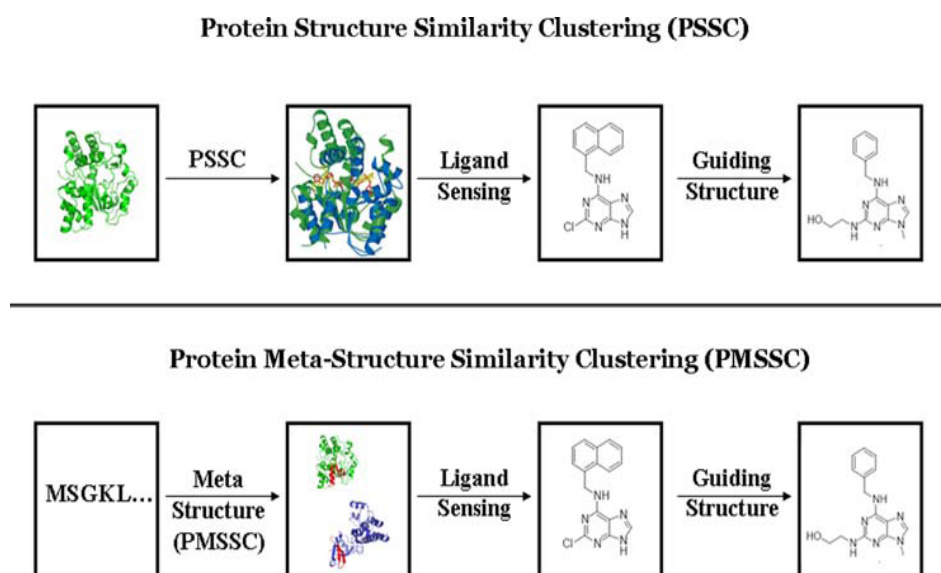


FIG. 11 Meta-structure alignment as a powerful tool to detect possible ligands from drug data bank in comparison to PSSC. (Figure is adapted from Konrat¹³).

Introduction

1.7. Fragment Based Drug Design.

The indispensable starting point of FBDD is the identification of small molecule that are weak binders in the size range of 100– 300 Da and characterized by being drug-like²⁰. In comparison, High-throughput screening (HTS) in which diversity of compounds are distributed well-plates where the targeted protein is added and the interaction is then measured using different methods like measuring the reflectivity²¹, the HTS screen can contain up to 10^5 compounds and this covers only small fraction of possible small molecules. On top of that, HTS sometimes fails to provide compounds of biological relevance in drug industry because the compounds tested are of non-drug nature and this makes them not-applicable in the pharmaceutical industry²⁰.

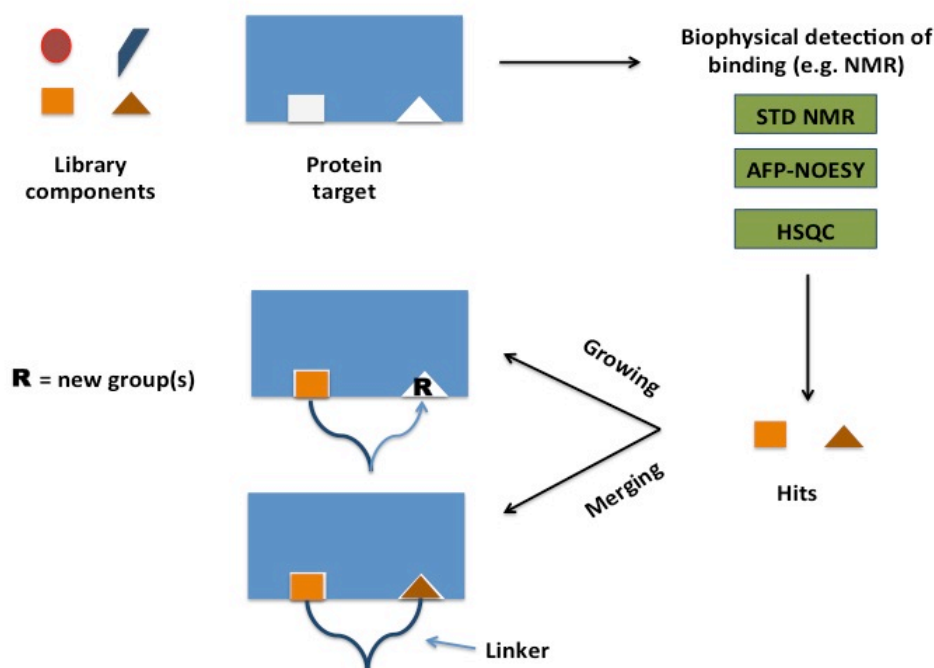


FIG.12 The main principles for fragment based drug design, which include library design, a method to test the binding of the fragments, and then the development into larger compounds using growing or linking.

The fragment library usually based on selection of a list of available chemical compounds with exclusion of toxic or reactive compounds²². More focused libraries also can be designed taking in account substructure of know pharmacologically active compounds²³. The challenging point in FBDD is that the fragments tested are usually

Introduction

of weak binding to the tested biomolecule, hence a sensitive method are needed to test the binding. These fragments are then developed into larger compounds either *via merging/linking* fragments together or *via growing* the fragments to enhance the binding. In *linking*, two or more fragments that are known to be binding to different sites on the target are linked together with a linker to have the product of the K_D of the linked fragments. On the other hand, *growing* means to augment the binding *via* addition of functional groups.

There are wide variety of techniques that can be used in order to measure binding between the targeted protein and the selected fragments like NMR and surface plasmon resonance²⁴. Indeed, detection of the binding between these fragments and the proteins is challenging due to the weak binding (K_D usually in the mM range). Therefore, NMR provides a unique technique due to its high sensitivity and reliability in this area. The first published fragment based lead discovery by NMR was described by Abbott group²⁵ and the binding was detected by the perturbation of the HSQC spectrum of labeled protein upon addition of the ligands. Different NMR techniques can be used to elucidate the binding events; some of them have been used in this investigation and will be discussed in details.

Introduction

1.8. β -Catenin & TCF4

1.8.1. Structure and function of Catenins

The word catenin originated from the Latin word catena which has the meaning of a chain²⁶. With the exception of structurally unrelated α -Catenin²⁷, all human catenin family members contain a central Armadillo repeat domain that contains between 9 to 12 repeats (each repeat is 3 α -helices formed of about 40 amino acids)²⁸. This Armadillo repeat region folds to produce a super-helix of helices that bears a positively charged groove which is indispensable for binding²⁹. The catenin proteins family is classified into three subfamilies depending on their function³⁰.

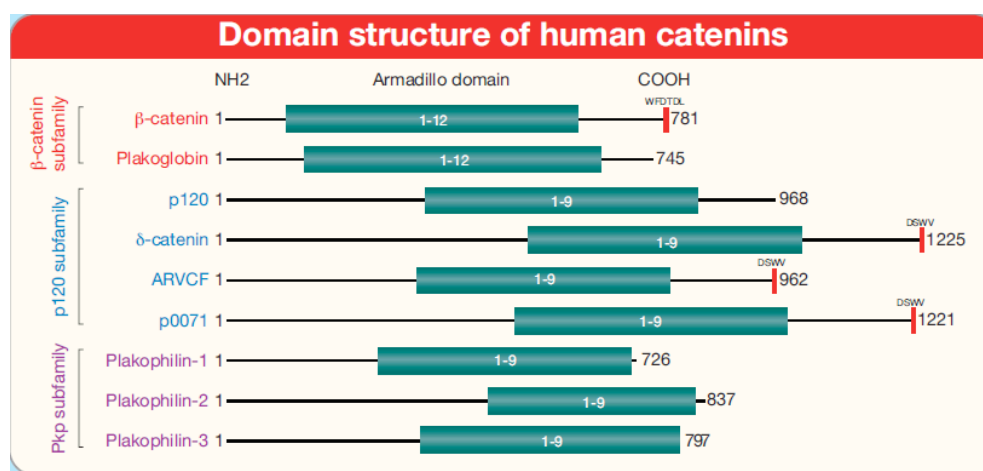


FIG. 13 Classification of catenin into three subfamilies. (Figure is adapted from McCrea and Gu³¹).

β -Catenin is the most relevant member in catenin family³¹. It consists of *N*-terminal domain (NTD), armadillo repeat domain and *C*-terminal domain (CTD). The armadillo repeat region consists of 12 repeats and it's the main region involved in protein binding while the *N*-terminal domain is essential for the process of proteosomal degradation of the protein, and *C*-terminal domain contributes to gene transactivation.

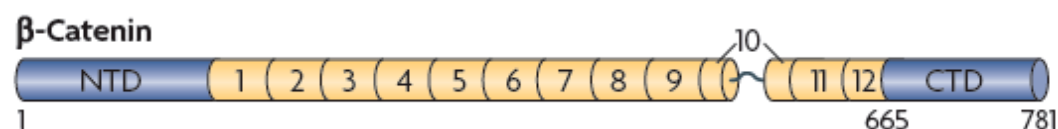


FIG. 14 Structure of full-length β -Catenin. (Figure is adapted from Mosimann *et al.*³²).

β -Catenin is involved mainly in two functions:

Introduction

- a- Binding to cadherin at cell-cell junctions (where β -Catenin binds with E-cadherin and α -Catenin to link them to the actin cytoskeleton)³³.
- b- It's the main component in canonical Wnt signaling pathway

1.8.2. Canonical Wnt signaling

The Wnt family of proteins consists of 19 known human members³⁴, and of size 40 KDa³⁵. These secreted Wnt are modified by glycosylation and can bind to Frizzled family receptors on the plasma membrane. Two domains on Wnts bind to Frizzled receptors and one of them contains palmitoleic acid lipid which projects inside the pocket of the receptor³⁶.

Interactions of Wnts with their receptors are associated with three signaling pathways; canonical Wnt/ β -Catenin pathway, non-canonical Wnt/ Ca^+ pathway and planar cell polarity pathway. In this research we are concerned mainly about the *canonical pathway* in which β -Catenin is playing the crucial role.

In the absence of Wnt ligands, β -Catenin in the cell cytoplasm is degraded by phosphorylation of the serine and threonine amino acids of the *N*-terminal by the destruction protein complex³⁷. This complex consists of five different proteins like the tumor suppressor protein APC; an active serine-threonine kinase. On the other hand, in presence of Wnt ligands, they bind to frizzled receptor on the cell membrane and this results in recruiting the destructive phosphorylation complex. Subsequently, β -Catenin escapes degradation and is translocated into the nucleus where it interacts with the DNA-binding T-cell factor/ lymphoid enhancer factor (TCF4/LEF) family proteins, replacing TCF4-bound corepressors leading to the conversion of TCF into a transcriptional activator of Wnt target genes³⁸ such as cyclin D1, c-Myc, survivin and COX-2. This Canonical pathway is involved in different cellular behavior such as embryogenesis, growth, cell division and apoptosis³².

Introduction

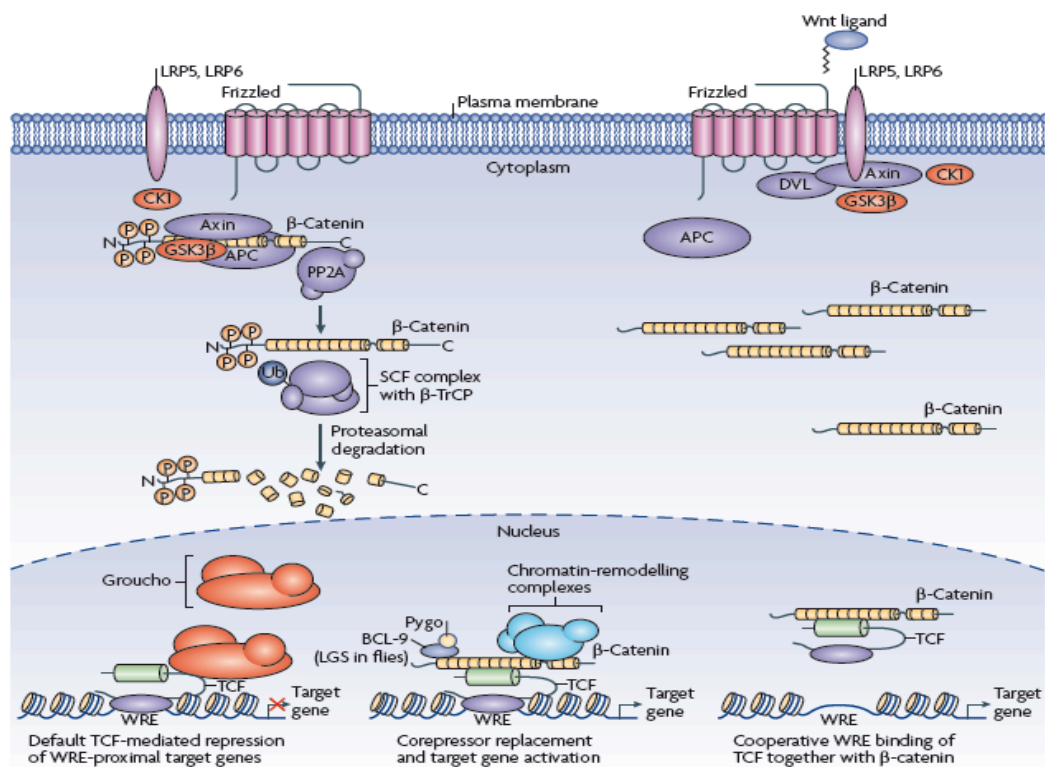


FIG. 15 Canonical Wnt signaling pathway. On the left, in absence of Wnt ligands. On the right, in presence of the Wnt ligands. (Figure is adapted from Mosimann *et al.*³²).

Noteworthy, upon mutation of β -Catenin (*N*-terminal serines or threonines amino acids)³⁹ or any of the phosphorylation destruction complex members, β -Catenin escapes from degradation and accumulates in the cell. These mutations impair the regulation of β -Catenin levels and affect mainly the tissues that depend on Wnt for repair. This overexpression of β -Catenin lead to continuous formation of complex with TCF4, which is associated with number of human tumors including colorectal, lung, breast, cervical, skin and hepatic tumors⁴⁰. Thus, β -Catenin is considered as a potential oncogene. From this perspective, blocking of the Wnt signaling pathway as possible antitumor therapeutic intervention could be achieved *via* targeting the complex between TCF4 and β -Catenin.

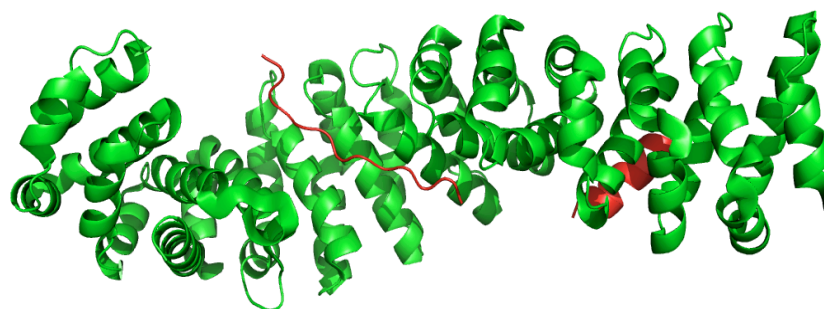


FIG.16 Structure of the complex between armadillo repeat region of β -Catenin (green) and TCF4 (red). TCF4 residues 13-25 runs as extended region in the twist formed by the armadillo repeat region. While residues 40-50 of TCF4 runs as α -helix in the shallow groove formed by repeats 3,4, and 5 of β -Catenin. (Protein Databank Entry: 1JPW).

1.8.3. β -Catenin/TCF4 complex as antitumor target

1.8.3.1. Disruption of β -Catenin/TCF4 as a new antitumor target

The Wnt/ β -Catenin pathway is linked to cancer *via* the activation of TCF4/ β -Catenin activated genes. Abnormal activation of the Wnt/ β -Catenin pathway is a frequent early event in intestinal epithelial cells during the development of colon cancer⁴¹, which promotes the expression of Wnt/ β -Catenin responsive genes, including cyclin D1, myc, and PPAR- δ , which play crucial roles in colorectal tumors⁴².

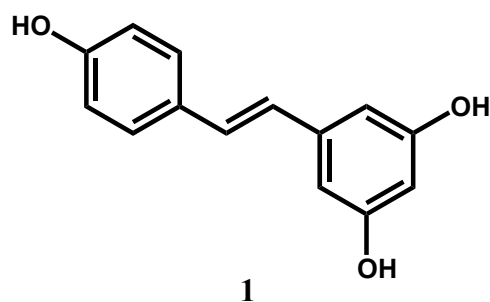
Based on that, disruption of β -Catenin/TCF4 complex has been considered a prominent target to design new antitumor drugs⁴³. It was shown that disruption of such complex results in down regulation of the oncogene c-myc, which was shown to be extensively expressed in colorectal cancer⁴⁴.

Recently, several Wnt inhibitors were identified in high throughput screening that target the upstream signaling of β -Catenin in order to promote its degradation⁴⁵.

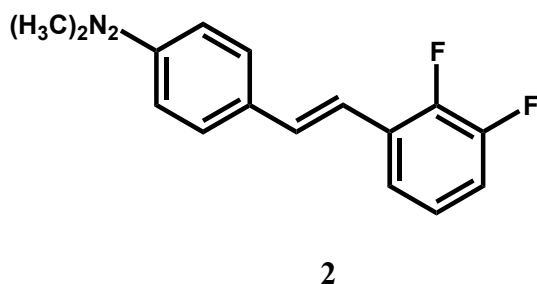
1.8.3.2. Literature survey on antitumor compounds related to β -Catenin

Natural products found in food are potentially ideal therapeutic agents for cancer. Resveratrol (**1**) is a phytoalexin produced in plants particularly in red grapes and blueberries⁴⁶. Resveratrol has been suggested to induce β -Catenin degradation and inhibits its nuclear localization in colorectal cancer cells⁴⁷.

Introduction

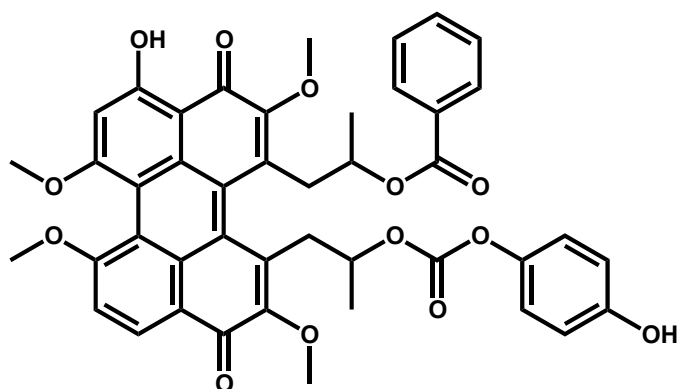


Zhang *et al*⁴⁸ reported the synthesis and evaluation of novel resveratrol analogues as anticancer that work against Wnt signaling. It was proved that (*E*)-4-(2,6-difluorostyryl)-*N,N*-dimethylaniline (**2**) is highly active as antitumor agent targeting Wnt pathway.

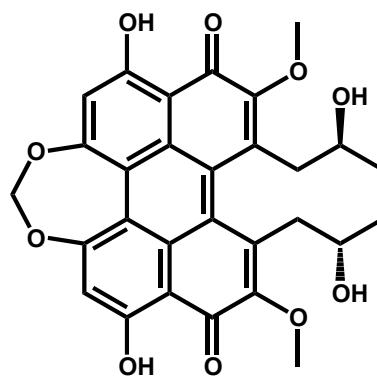


In addition, Lepourcelet *et al*⁴³ screened about 7000 natural low molecular weight compounds at 10 μ M concentrations for their ability to disrupt the β -Catenin/TCF4 complex. Six compounds isolated from microbial origin showed reproducible dose-dependent inhibition of β -Catenin/TCF4 interaction. These compounds showed diverse structural properties although they share the polyhydroxylated scaffolds. These compounds are; PKF115-584 (**3**), CGP049090 (**4**) and PKF222-815 (**5**) which are isolated from fungal organisms, whereas PKF118 (**6**), PKF118-310 (**7**) and ZTM000990 (**8**) and they are isolated from Actinomycete strains.

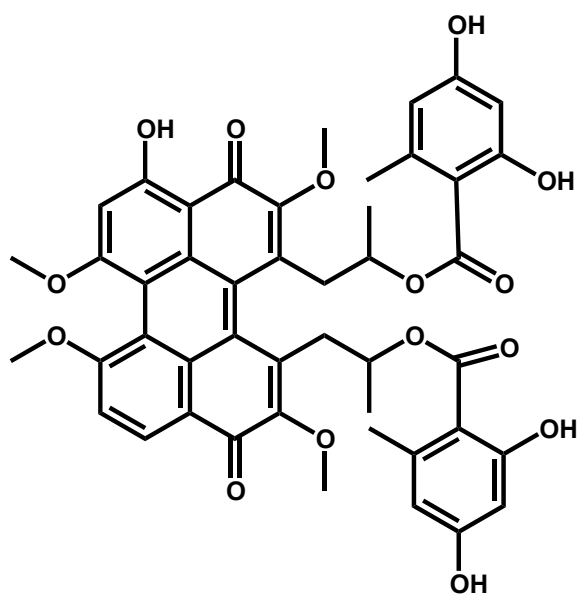
Introduction



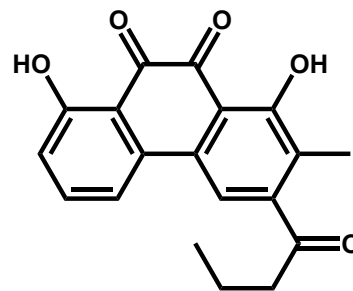
3



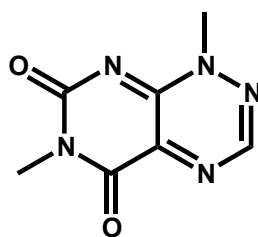
4



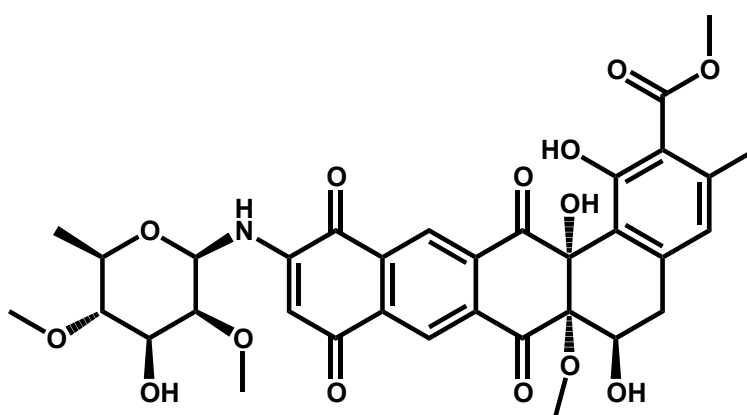
5



6



7

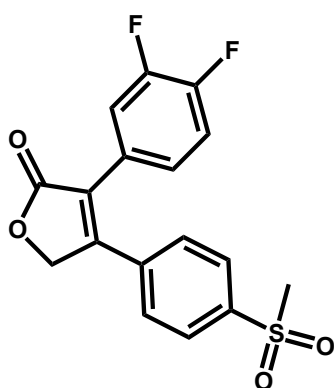


8

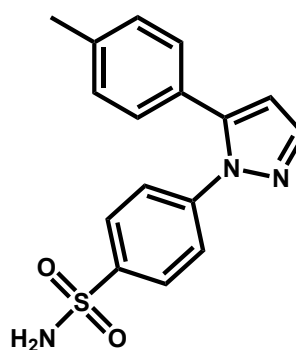
Introduction

1.8.3.3. Drugs that target the Wnt target genes

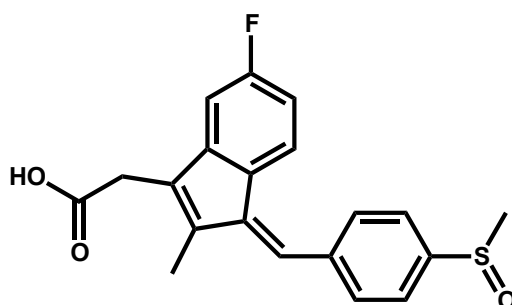
Amongst the Wnt target genes are PPAR- γ and COX-2, both of them have been proved to play a role in the development of colorectal tumors. Both of these genes are inhibited by non-steroidal antiinflammatory drugs (NSAIDs), which have been shown in turn to inhibit colon cancer. COX-2 produces eicosanoids (such as prostaglandins) from arachidonic acid and these eicosanoids are themselves PPAR- γ ligands that stimulate transcription. PPAR- γ is upregulated in early carcinogenesis by the hyper-activity of TCF4/ β -Catenin complex³⁴. Treatment of mutant mice with the selective COX-2 inhibitor; MF tricyclic (**9**) was shown to reduce polyp number⁴⁹ and the treatment of patients with familial adenomatous polyposis with COX-2 inhibitor; celecoxib (**10**) showed significant decrease in the number of colorectal polyps⁵⁰. Also levels of nuclear β -Catenin in patients with adenomatous polyposis were reduced significantly when treated with the NSAID sulindac (**11**) for 6 months. In spite of this, the treatment with the selective COX-2 inhibitors is still controversial due to the possibilities of developing blood clots and strokes⁵¹.



9



10



11

THEORETICAL

2.1. NMR screening in drug discovery:

In fragment based drug design (FBDD), different fragments are tested for binding to the target biomolecule. As mentioned before, the problem with this method is the weak affinity of these fragments, which limits other conventional methods from detecting this. Therefore, NMR serves here as a sensitive technique, that can be used to demonstrate binding of those low affinities fragments.

Two main methods are used for protein-ligand binding investigation using NMR; either observing target resonance or ligand resonance.

2.1.1. *Ligand resonance observing method*

Different sets of experiments have been designed in order to see the protein-ligand interaction like saturation transfer difference NMR, ^{19}F -NMR, water-LOGSY and NOE-ROE experiments. Usually these methods provide the advantage of having short time measurement experiments in addition to the absence of limitation of the macromolecule size. The main limitation of this method is the lack of ability to detect tight binding ligands as the slow dissociation rate inhibits the transfer of the magnetization on the small bound fraction to the bulk unbound one which is the one that is recorded in this kind of experiments⁵².

In order to overcome the drawback of high affinity limitation, different approaches have been adopted such as reporter screening⁵³ and spin labeling⁵⁴.

2.1.2. *Target resonance observing method*

Herein, the changes of the macromolecule chemical shift (e.g. protein) are observed upon the binding to the ligand. One of the main advantages of this method is the possibility to detect the binding site on the protein if the assignment of the target is available. On the other hand, the biggest obstacle in this process is the size limit of the macromolecules which is usually limited to 40-50 KDa⁵⁵, as at larger size, the spectrum usually suffer from faster relaxation properties which lead in turn to line broadening. Moreover, this method is limited by the need to prepare expensive labeled sampled (e.g. ^{15}N and/or ^{13}C labeled).

2.2. NMR experiments adopted during this investigation:

2.2.1. Protein-Ligand Binding elucidation NMR experiments:

2.2.1.1. 1D ^1H -Saturation Transfer Difference NMR: (STD)⁵⁶

Saturation transfer difference approach can provide a detailed picture on the interaction between ligands and protein targets, and also to detect possible epitopes. The STD experiment can also be applied as screening technique in the dynamic combinatorial libraries to identify high-affinity ligands. The method requires only small amount of the unlabeled target protein and a 50-100-fold excess of ligand(s).

In this technique, RF irradiation of the protein spins at a resonance where no ligand signals are present is done; this is achievable because of the unique chemical environment of the protein and because of the huge line width of the protein signal in comparison to sharp peaks of the small ligands, so the protein spins have significant signal intensity even in segments outside the spectral window of low-molecular-weight ligands as in the negative ppm region (less than zero ppm) or above 10 ppm. This selective protein irradiation leads to saturation of the entire protein via spin diffusion. If the ligand binds the protein, saturation will spread onto the ligand. So the ligand spins that are closer to the protein receive more intermolecular NOE. As a result, intensity of the ligand signal is attenuated. Subtraction of resulting spectrum from the reference 1D ^1H -spectrum (where the RF irradiation is done so far from the area of interest) yields the positive 1D ^1H -STD spectrum, which contain only signals from the binding ligand (s). If the ligand is not binding, the resulting subtraction spectrum will be just a zero ^1H -STD spectrum. The experiment is repeated in the absence of the protein to be sure that a zero ^1H -STD spectrum is produced (negative control experiment).

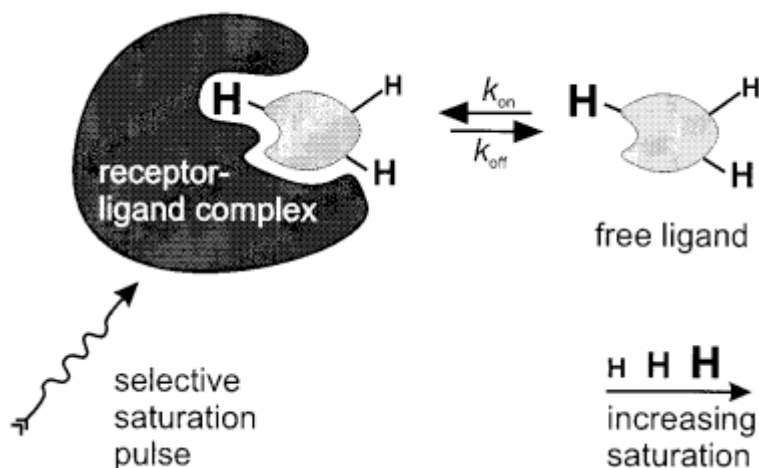


FIG. 17 Schematic representation of the saturation transfer difference NMR experiment. The degree of saturation of the ligand protons is represented *via* the size of the proton in the figure. The larger the proton size in the figure means the closer it's to the protein surface and the more amount of magnetization it gets. (Figure is adapted from Mayer and Meyer⁵⁶).

2.2.1.2. ¹⁹F-NMR screening

¹⁹F atom has unique chemical properties that attracted great attention in the pharmaceutical industry. It can affect the physico-chemical properties of the molecule, which is translated into better penetration into the membrane for example. In addition, C-F bond is 7 Kcal stronger than the C-H bond and that makes it more resistant to enzymatic attack. ¹⁹F-NMR ligand screening offers unique advantages; a) high sensitivity and the natural abundance of the ¹⁹F nucleus, b) no interference from protonated buffers or reagents, c) absence of overlap because of the wide dispersion of the peaks, d) the ¹⁹F transverse relaxation rate R_2 in bound state is highly sensitive to binding due to high exchange contribution⁵⁷. This R_2 sensitivity to binding is directly reflected on the peak width. Hence, when a fluorinated ligand binds to a protein, a broadening of the ¹⁹F-peaks involved in the binding site is observed. In addition, competitive-based fluorine screening has been described as a tool for detection of binding constant. In this technique, the fluorinated ligand acts as a (spy) molecule that could be displaced by non-fluorinated ligand, which gives the advantage of avoiding the overlap between the spy, and the non-fluorinated ligand.

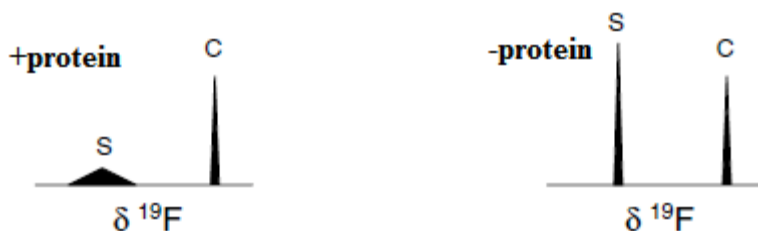


FIG. 18 ^{19}F -NMR of fluorinated ligand (s) in presence and in absence of the binding protein showing broadening of the peak upon binding in comparison to the control sample (c). (The figure is adapted from Dalvit⁵⁷).

2.2.1.3. 2D-homonuclear ^1H - ^1H NOESY

NOESY is very illuminating NMR experiment in terms of structure and it gives a clear idea about the neighborhood between protons can be extracted. The main building block of NOE experiment consists of 90° pulse to label the protons with their chemical shifts, followed by evolution time. Afterwards, another 90° pulse is applied which takes the magnetization to the z-axis, followed by a mixing time delay (τ_m) during which the perturbed spins return to equilibrium *via* reshuffling of the spins population and hence transfer the magnetization to other spins in the system *via* cross relaxation (σ) in distance range 5 \AA . Finally, another 90° pulse is applied to bring the magnetization back to the transverse plane, and followed by detection⁵⁸.

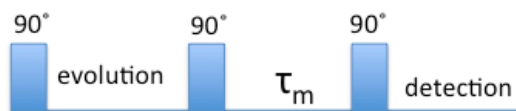


FIG. 19 The building block of NOE experiment pulse sequence.

In the observed spectrum, we see mainly two types of peaks; the diagonal peaks which appear at offset (Ω_1, Ω_1) of spin 1, and cross peaks, which appear at (Ω_1, Ω_2) for spin 1 and 2 respectively. The cross peaks are the ones that carry the information about possible interaction between spin 1 and 2. The intensity of these peaks at short mixing time is ($\sigma \tau_m$). In case of fast tumbling molecules, the cross relaxation rate is positive so we see negative peak, while for large slowly tumbling molecules, σ is negative and the cross peaks appear with positive sign (same as the diagonal peaks). The change of sign of the NOE cross relaxation rate arises from the fact that $\sigma = W_2 -$

Theoretical

W_0 where W_2 is the probability to have double quantum transition, while W_0 is the probability to have zero quantum transition. For small molecules, W_2 is larger than W_0 which makes the sign of the cross relaxation rate positive, while for large molecules, W_2 starts to decline while W_0 continues to linearly increase resulting in negative sign for σ .

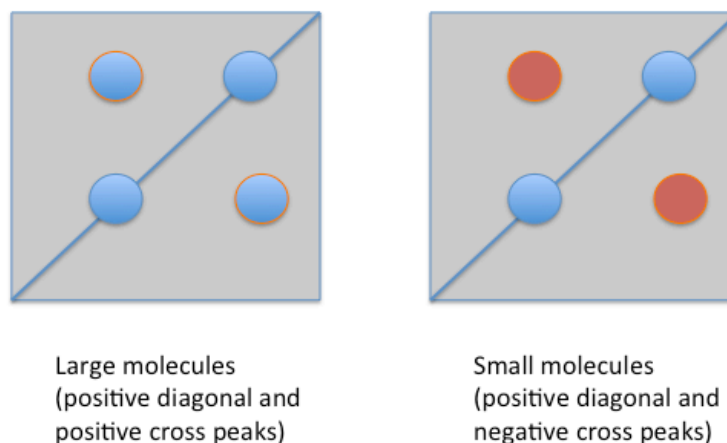


FIG. 20 2D homonuclear ^1H - ^1H NOE showing diagonal and cross peaks signs.

The cross relaxation rate is proportional to the reciprocal of distance between the two spin (r^{-6}), hence usually it's observed if the distance between the 2 spins is in the range of 5 Å. In addition, the cross peak intensity is dependent on the mixing time, where the magnetization starts to build up, followed by a maximum and then the magnetization starts to decay. Therefore, one has to be careful with choosing the right mixing time depending on what kind of NOE wanted to be observed; in case of spin diffusion, longer mixing time should be chosen, as the intensity is now dependent on quadratic term ($\sigma^2 \tau_m^2/2$).

Measuring 2D ^1H - ^1H NOE at different mixing time for a small ligand in presence of protein can provide an efficient tool to observe change of the intensity of the cross peaks and to map the ligand protons that are imbedded in the binding site. From the NOE build up curves, it's possible to see spin diffusion effect on some protons of the ligand which can be translated into binding between these protons and the protein.

Theoretical

2.2.1.4. Cross relaxation during Adiabatic Fast Passage⁵⁹.

The nuclear overhauser effect (NOESY) is an information-rich NMR technique for structure determination studies and it refers to the longitudinal cross relaxation. In addition, it's also proved that Rotating Frame overhauser spectroscopy (ROESY) can provide unique information about dynamics of proteins⁶⁰. ROESY refers to the transverse cross relaxation.

In this method, the NMR experiment measures homonuclear (¹H-¹H) cross relaxation rates (NOEs and ROEs) during adiabatic past passage (AFP) which is convenient method for the examination of protein ligand binding.

Unlike conventional AFP methods, the RF field is not small but of comparable strength to the sweep frequency ($\gamma B_1 \approx \Delta\omega$) and so leading to significant contribution of transverse relaxation to the effective spin lock relaxation rate.

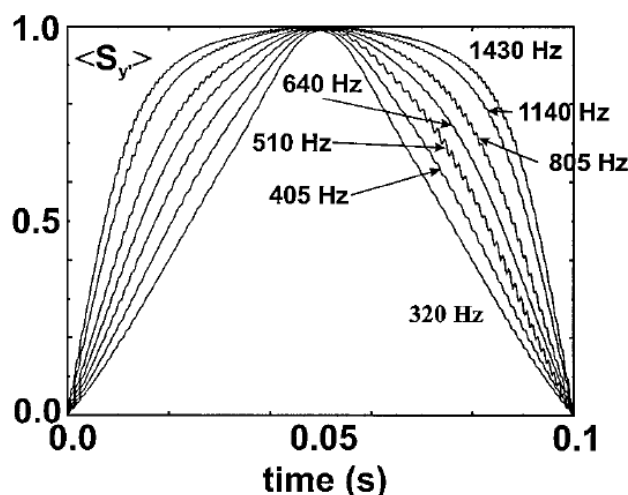


FIG. 21 The effect of the strength of the applied AFP on the amount of the transverse magnetization. The figure shows that increasing γB_1 increases the amount the net amount of transverse magnetization during the sweep. In the middle of the sweep, the transverse magnetization reaches the on-resonance condition. (Figure is adapted from Konrat and Tollinger⁶¹).

The pulse scheme for this experiment is a conventional NOESY in which the AFP that has parabolic phase shift in order to create a time dependent offset, replaces the original longitudinal mixing period. Therefore, the effective relaxation rate for two spins i and j is

$$\sigma_{\text{eff}}^{ij} = \sigma_{\text{NOE}} \cos \theta_i \cos \theta_j + \sigma_{\text{ROE}} \sin \theta_i \sin \theta_j$$

Where θ is the angle between the offset and the effective fields, σ_{NOE} is the NOE cross

Theoretical

relaxation and σ_{ROE} is the ROE cross relaxation.

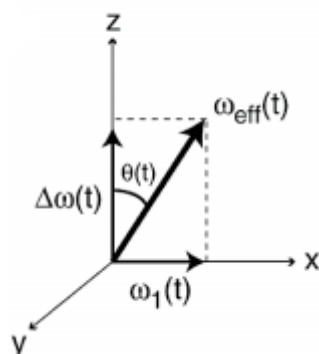


FIG. 22 Frame of the Adiabatic spin lock shows the following: offset $\Delta\omega(t)$, r.f. field $\omega_1(t)$, effective field $\omega_{\text{eff}}(t)$ and the angle $\theta(t)$ between offset and the effective field. (Figure is adapted from Auer *et al.*⁵⁹).

For small molecules like ligands, both NOE and ROE enhancements are the same (+50%) and the spin lock relaxation rate is independent on the strength of the applied AFP strength during mixing time. For large molecules, NOE and ROE enhancements aren't the same as they have different value and different sign (NOE becomes negative -100%, while ROE remains positive and has higher value +200%). As a result, for large molecules, the enhancement will be dependent on the AFP strength. As the strength of the AFP increases, the relaxation shifts from NOE to ROE regime. During this moving from the NOE to ROE, there will be a zero crossing. For rigid macromolecules with no internal mobility, the zero crossing happens at angle $\theta = 35.5^\circ$.

Theoretical

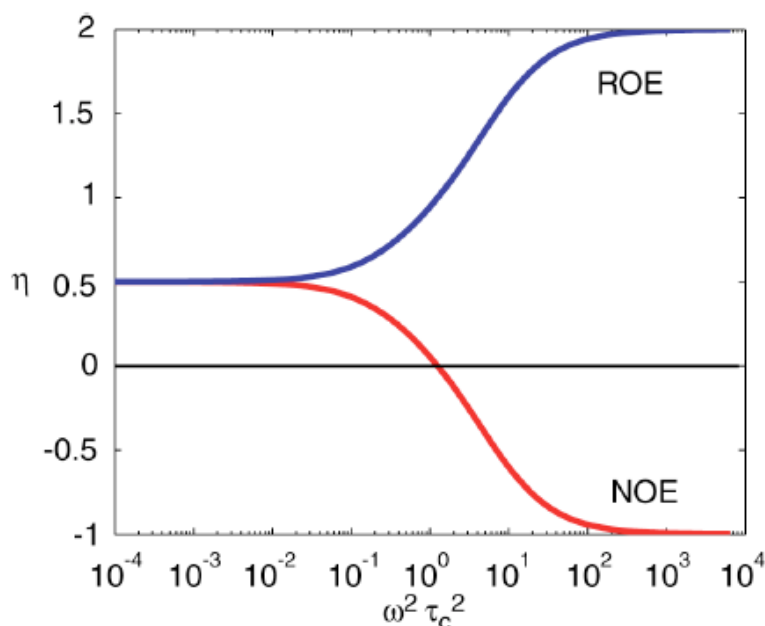


FIG. 23 NOE and ROE enhancement (η is the enhancement; $\omega^2\tau_c^2$ is a parameter of size based on correlation time and the larmor frequency). (Figure is adapted from Auer *et al.*⁵⁹).

It's noteworthy that this zero crossing angle can be less than 35.5° because of high internal mobility (order parameter $S^2 > 0.5$), or it can be larger in case of spin diffusion.

2.2.2. Protein Reductive Methylation

Protein reductive methylation was first introduced in 1968 by Means and Feeny⁶². It's used to enhance the crystallization. Some proteins are resistant for crystallization because of the flexible nature of the surface residues. Hence, lysine residues, which are characterized by polar, flexible and long side chain, can be localized in order crystal lattice formation. Therefore, reductive methylation of the lysine side chain was found to have the ability to increase the hydrophobic surface of the protein in aqueous media, which favors the protein-protein hydrophobic interaction, and hence better packing of the protein in the crystal lattice. This in turn is reflected into better crystallization conditions⁶³.

Synthesis of selectively lysine-methylated protein can be recruited by NMR to have a sensitive method for detection of protein conformational changes induced by small molecules binding and protein-protein interaction. Methyl groups provide a wonderful probe for NMR due to; a) high sensitivity as it has 3-fold symmetry, b) high side chain mobility which means higher relaxation time, c) using selective labeling methyl

Theoretical

groups allow the labeling of unlabeled proteins from all kind of sources. Methylation doesn't significantly perturb protein structure, it retains the positive charge⁶⁴ and doesn't induce protein-protein interactions⁶⁵. On top of that, reductive methylation of large proteins provides favorable relaxation properties because of the reduced order parameter (S^2) of the lysine side chain, which can be in turn recruited for proteins like β -Catenin that are problematic for conventional NMR measurements. So, ^{13}C methyl groups added to lysine side chain can be detected with higher sensitivity than those on aliphatic amino acids in the ^{13}C labeled proteins.

In this technique, two ^{13}C -methyl groups are introduced on each lysine amino acid in a non-labeled protein sample⁶⁶.

To achieve this selective methylation of the lysine amino acids, ^{13}C -labeled formaldehyde is added to the protein. The formed imine is subsequently reduced in presence of borane ammonium complex to give secondary amine that carries one ^{13}C -labelled methyl group. Afterwards, excess of ^{13}C -labelled Formaldehyde and reducing agent is added to the protein resulting in formation of tertiary amine carrying two ^{13}C -labelled methyl groups for each lysine residue.

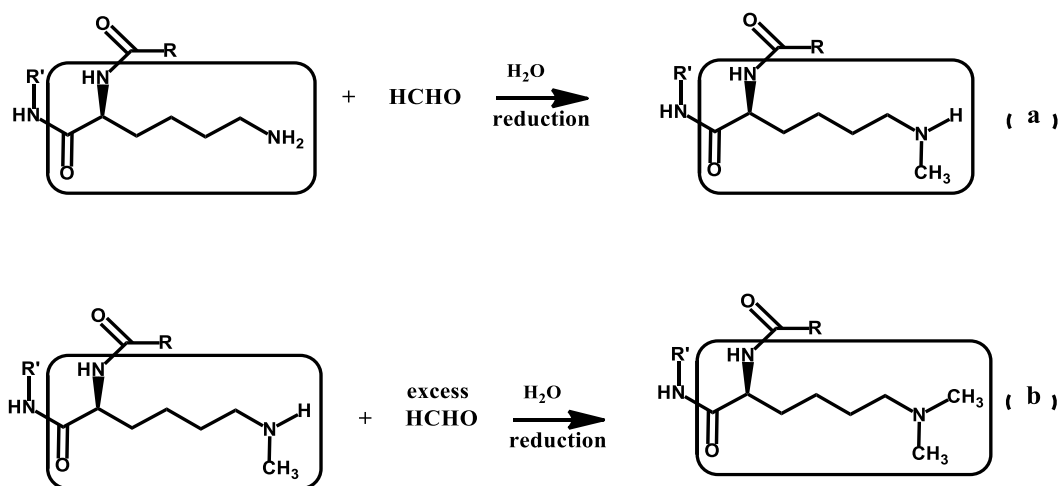


FIG. 24 Reductive methylation reaction showing the two steps addition of two methyl groups to an ϵ -amino group of a lysine residue or the N -terminal amino group through the formation of Schiff's base intermediate. (a) Initial step of the reaction to form monomethylated lysine. (b) In the presence of sufficient formaldehyde, the reaction proceeds to give the di-methylated lysine derivatives.

In this investigation, synthesis of ^{13}C -methylated lysines of β -Catenin was adopted by using the previously described method⁶⁴ to yield a tertiary amine (di-methylated derivative).

Theoretical

^1H - ^{13}C 2D-HSQC NMR spectra were recorded for the newly synthesized methylated β -Catenin alone, in the presence of its binding partner; TCF4 and with the detected binding ligand; sodium 2-phenoxybenzoate. From this experiment it was possible to confirm the similarity in the binding site between TCF4 and sodium 2-phenoxybenzoate on the β -Catenin.

2.2.3. Protein assignment NMR experiments:

“The figures of magnetization transfer for this section are taken from the site <http://www.protein-nmr.org.uk/spectra.html> developed by Biochemistry Department, University of Oxford South Parks Road Oxford OX1 3QU, UK.”

In 2D and 3D NMR assignment experiments, the main building block is INEPT (insensitive nuclei enhanced by polarization transfer). In INEPT, the magnetization is transferred from the highly sensitive nuclei (usually ^1H) with the highest gyromagnetic ratio (hence, highest Boltzman equilibrium magnetization) to the less sensitive one (like ^{14}N or ^{13}C). Then using reverse INEPT, the magnetization can be transferred back to the sensitive nuclei for detection.

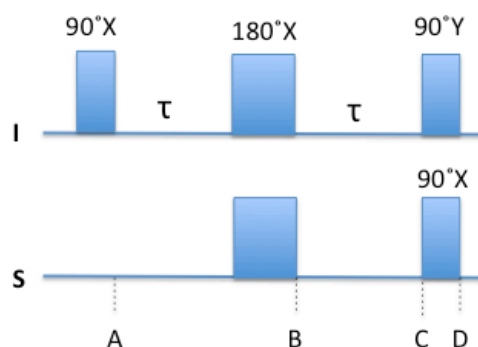


FIG. 25 Pulse sequence of INEPT experiment done on two spins (I and S) where I is the most sensitive nuclei and S is the less sensitive one.

In figure 25, the pulse sequence of INEPT is shown. Spin I is the spin with the highest sensitivity (e.g. ^1H) while spin S is the less sensitive one (e.g. ^{13}C or ^{15}N). At point A, after a 90° pulse on spin I, the operator will be I_y . During the delay τ , evolution of scalar coupling between spins I and S, so the operator will be $I_y \cos(\pi\mathbf{J}_{\text{IS}}\tau) + 2I_xS_z \sin(\pi\mathbf{J}_{\text{IS}}\tau)$. At point B, after applying 180° pulses on both spins the

Theoretical

operator will be $-I_y \cos(\pi J_{IS}\tau) - 2I_x S_z \sin(\pi J_{IS}\tau)$. At point C, the operator will be $-I_y \cos(\pi J_{IS}2\tau) - 2I_x S_z \sin(\pi J_{IS}2\tau)$. So if $2\tau = 1/2J_{IS}$, the only remaining term will be a pure antiphase magnetization $-2I_x S_z \sin(\pi J_{IS}2\tau)$. At point D, after applying 90° pulses on both spins, the operator will be $-2I_z S_y$, which means that we started with transverse magnetization on spin I and we ended with transverse magnetization on spin S (coherence transfer).

2.2.3.1. 2D 1H - ^{15}N HSQC (heteronuclear single quantum coherence spectrum):

This is the most standard and initial NMR experiment and creates H-N correlations (via 1 bond scalar coupling). It's considered as the fingerprint for the proteins and requires ^{15}N labeled protein.

The main regions that are observable in the spectrum include the following:

- a- Backbone amide groups
- b- Tryptophan side-chain $N\epsilon$ - $H\epsilon$ (appears downfield)
- c- Asparagine/Glutamine side-chain $N\delta$ - $H\delta 2$ / $N\epsilon$ - $H\epsilon 2$ groups respectively.
- d- The Arginine $N\epsilon$ - $H\epsilon$ (appears as folded peak with opposite sign to the other peaks because the $N\epsilon$ chemical shift of arginine is outside the spectral width recorded)

Mostly in this experiment, it starts with the equilibrium magnetization on 1H rather than on the heteronucleus because proton has higher Boltzmann equilibrium magnetization, which enhances the sensitivity as mentioned before. In addition, proton magnetization returns to its equilibrium value faster than the one of other heteronuclei so the time needed for waiting between scans is less and hence reduces the experiment time.

The pulse sequence for HSQC experiment starts mainly with INEPT to transfer magnetization from the 1H to ^{15}N followed by evolution time (t_1) which allows to label the spins of ^{15}N spins with their chemical shifts, followed by back INEPT to transfer the magnetization back to 1H for detection (t_2).

Theoretical

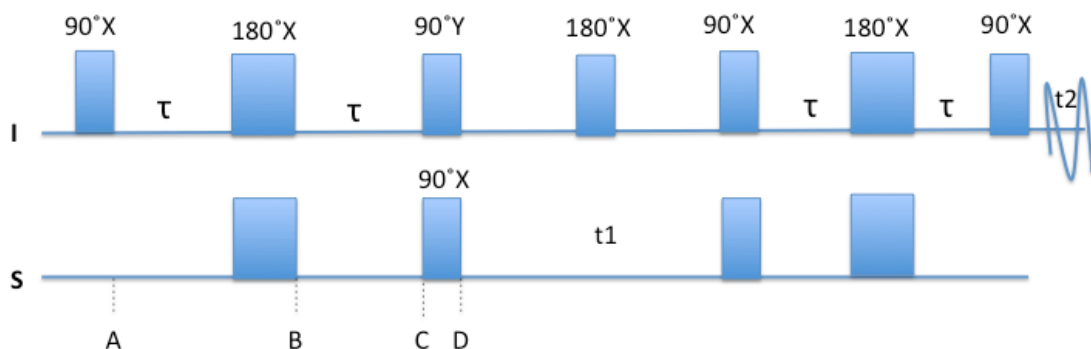


FIG. 26 Pulse sequence of HSQC experiment done on two spins (I and S) where I is the most sensitive nuclei and S is the less sensitive one.

It's also noteworthy to mention that there's a short version of this experiment called sofast in which only 30° or 40° soft pulses instead of 90° pulse are applied to selectively excite backbone amide protons spins, hence, this reduces the time needed for the bulk equilibrium magnetization to build up again and this in turn means more time saving between scans.

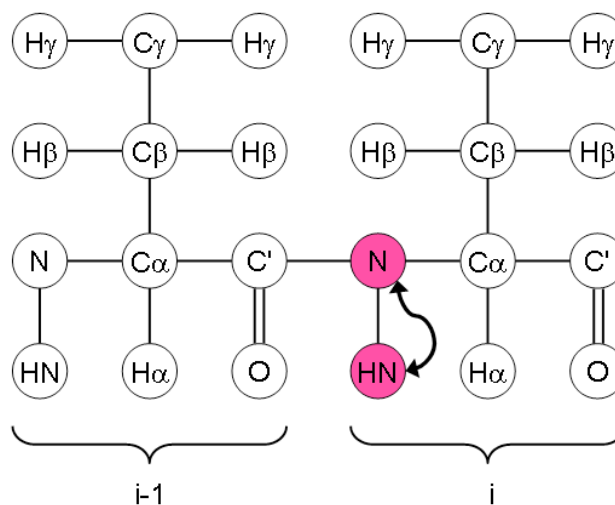


FIG. 27 Magnetization transfer in 2D ^1H - ^{15}N HSQC. (Figure is adapted from site <http://www.protein-nmr.org.uk/spectra.html>).

2.2.3.2. 3D ^1H - ^{15}N - ^{13}C HNCA⁶⁷:

Here the magnetization is passed from ^1H to ^{15}N on the backbone and then *via* the N-C α scalar coupling to the $^{13}\text{C}\alpha$ and then back again to ^{15}N and ^1H hydrogen for detection. The chemical shift is evolved for $^1\text{H}^{\text{N}}$, $^{15}\text{N}^{\text{H}}$ and $^{13}\text{C}\alpha$, resulting in a 3-dimensional spectrum. 2 Peaks of both C α s (C α_i and C α_{i-1}) are visible

Theoretical

in the spectrum for each NH because the amide nitrogen is coupled with both $C\alpha$ s. The peaks of $C\alpha_i$ appear more intense in comparison to the ones of $C\alpha_{i-1}$ because of the higher coupling to the directly attached $C\alpha$.

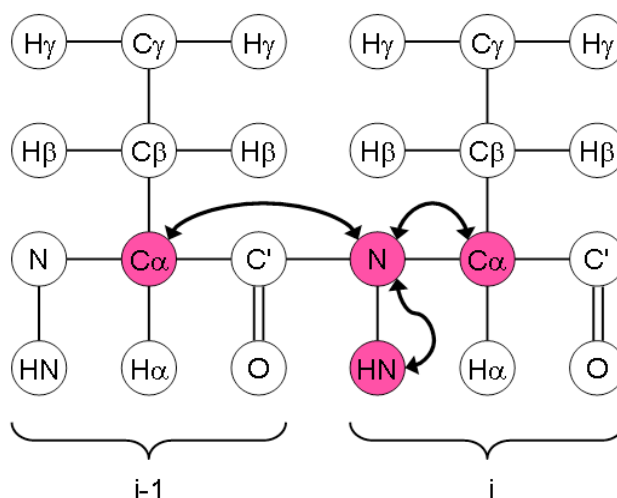


FIG. 28 Magnetization transfer in 3D ^1H - ^{15}N - ^{13}C HNCA. (Figure is adapted from site <http://www.protein-nmr.org.uk/spectra.html>).

2.2.3.3. 3D ^1H - ^{15}N - ^{13}C HN(co)CA⁶⁸:

This experiment is homologous to the HNCA with one modification that the magnetization is passed from ^1H to ^{15}N and then to ^{13}CO . From here it is transferred to $^{13}\text{C}\alpha$. The chemical shift is evolved only for the $C\alpha$ but not for the CO. This results in a spectrum, which is similar to the HNCA, but only one peak is observed for each NH corresponding to the $C\alpha$ of the preceding residue ($C\alpha_{i-1}$).

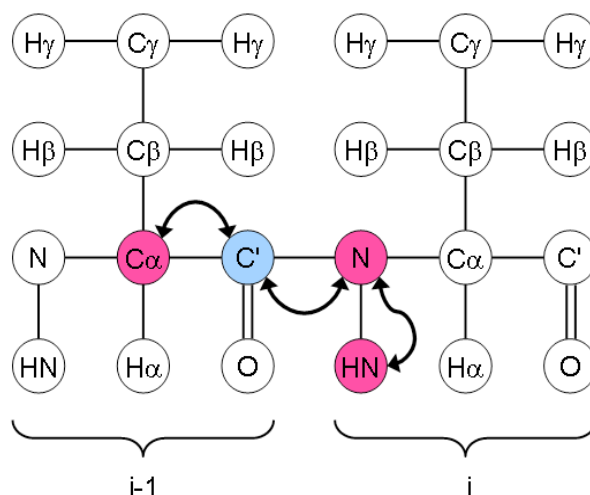


FIG. 29 Magnetization transfer in 3D HN(CO)CA. (Figure is adapted from site <http://www.protein-nmr.org.uk/spectra.html>).

Theoretical

2.2.3.4. 3D ^1H - ^{15}N - ^{13}C HNCACB⁶⁹:

Magnetization is transferred from $^1\text{H}_\alpha$ and $^1\text{H}_\beta$ to $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$, respectively, and then from $^{13}\text{C}_\beta$ to $^{13}\text{C}_\alpha$ then to $^{15}\text{N}^{\text{H}}$ and to $^1\text{H}^{\text{N}}$ for detection. So for each NH group there are two recorded C_α and C_β peaks belonging to the i and $i-1$ residues. The signal intensity for the i peaks are more than those of the $i-1$ because of the stronger scalar coupling. In This experiment, signal intensity is distributed between among four resonances in the ^{13}C dimension and this makes it much less sensitive experiment in comparison to HNC O or HNCA.

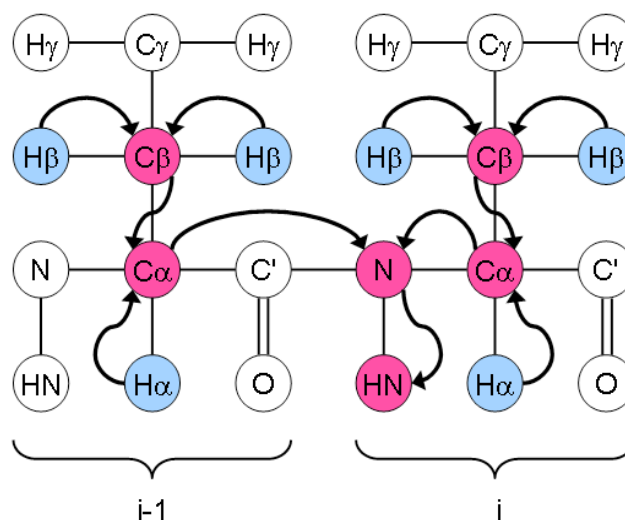


FIG. 30 Magnetization transfer in 3D HNCACB. (Figure is adapted from site <http://www.protein-nmr.org.uk/spectra.html>).

2.2.3.5. 3D ^1H - ^{15}N - ^{13}C HNC O ⁶⁷:

Magnetization is passed from ^1H to ^{15}N and then selectively to ^{13}CO via the $^{15}\text{N}^{\text{H}} - ^{13}\text{CO}$ J-coupling. Magnetization is then passed back via ^{15}N to ^1H for detection. Herein, we observe only one peak for each NH corresponding to the CO_{i-1} , but nothing is observed for the CO_i because the scalar coupling between ^{15}N and $^{13}\text{CO}_i$ is approaching zero.

HNC O experiment is much more sensitive in comparison to HNCA. The reduced sensitivity of HNCA experiment is due to:

- a- Distribution of magnetization in HNCA on two coupling partners (C_α_i and C_α_{i-1}).
- b- Smaller ^1J coupling in HNCA

Theoretical

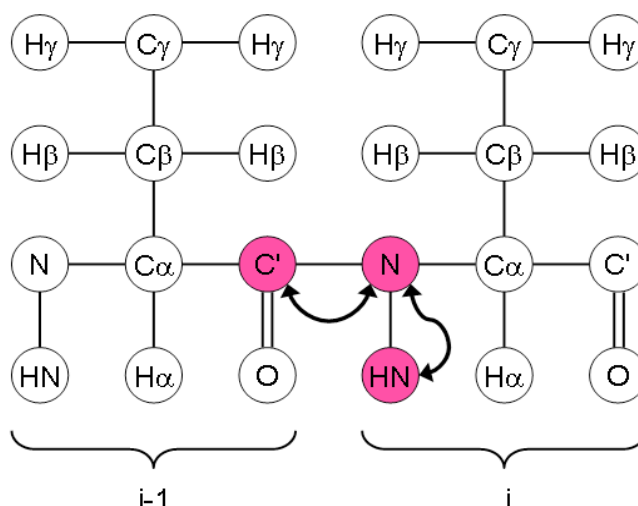


FIG. 31 Magnetization transfer in 3D HNCO. (Figure is adapted from site <http://www.protein-nmr.org.uk/spectra.html>).

2.2.3.6. 3D ^1H - ^{15}N - ^{13}C HN(ca)CO⁷⁰:

The Magnetization is transferred from ^1H to ^{15}N and then *via* the N- $\text{C}\alpha$ scalar coupling to the $^{13}\text{C}\alpha$. From there it is transferred to the ^{13}CO *via* the $^{13}\text{C}\alpha$ - ^{13}CO scalar coupling. For detection, the magnetization is moved back in the opposite direction till finally it reaches ^1H . The chemical shift is only evolved on ^1H , ^{15}N and ^{13}CO . Because the amide nitrogen is coupled both to the $\text{C}\alpha$ of its own residue and that of the preceding residue, both these transfers occur and transfer to both ^{13}CO nuclei occurs. So for each NH group, two carbonyl groups are observed in the spectrum. But because the coupling between N_i and $\text{C}\alpha_i$ is stronger than that between N_i and $\text{C}\alpha_{i-1}$, the H_i - N_i - CO_i peak generally ends up being more intense than the H_i - N_i - CO_{i-1} peak.

Theoretical

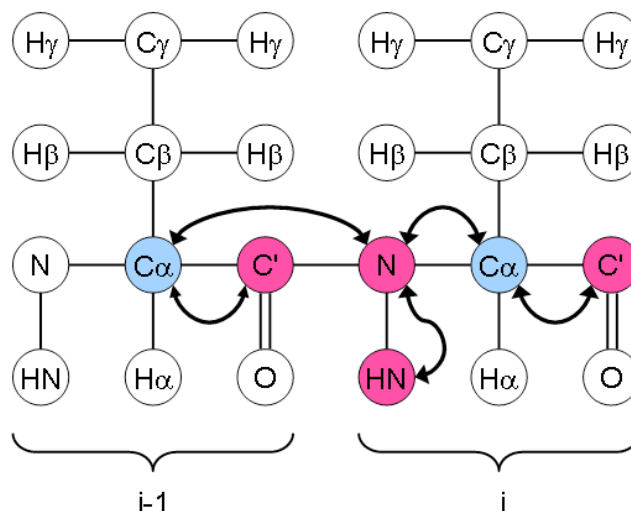


FIG. 32 Magnetization transfer in 3D HN(ca)CO. (Figure is adapted from site <http://www.protein-nmr.org.uk/spectra.html>).

For protein assignment in this work, it starts with observing the $C\alpha$ and $C\beta$ chemical shifts of some amino acids in the sequence, which have characteristic values. For example, alanine has a unique upfield value for the $C\beta$ around 20 ppm, while serine and threonine are the opposite, as they have downfield characteristic $C\beta$ chemical shift around 60 and even more downfield than their $C\alpha$, on the other hand, glycine residue has no $C\beta$ at all. Detecting these amino acids is a good start to connect a series of amino acids in the sequence using the different measured spectra. Figure 33 shows 4 strips from HNCACB spectrum and how these strips can be connected to predict a chain of residues.

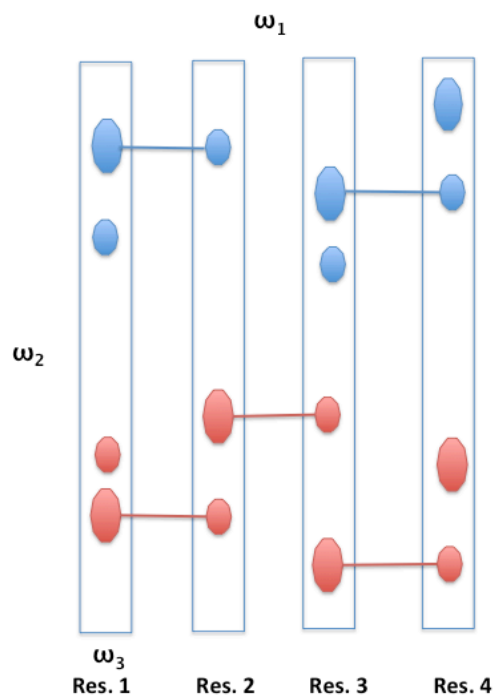


FIG. 33 Assignment of residues chain using HNCACB spectrum. 4 strips at different ^{15}N chemical shifts are shown. The blue peaks are the $\text{C}\beta$ peaks while the red ones are the $\text{C}\alpha$ peaks. ω_3 is ^1H , ω_2 is ^{13}C , while ω_1 is ^{15}N .

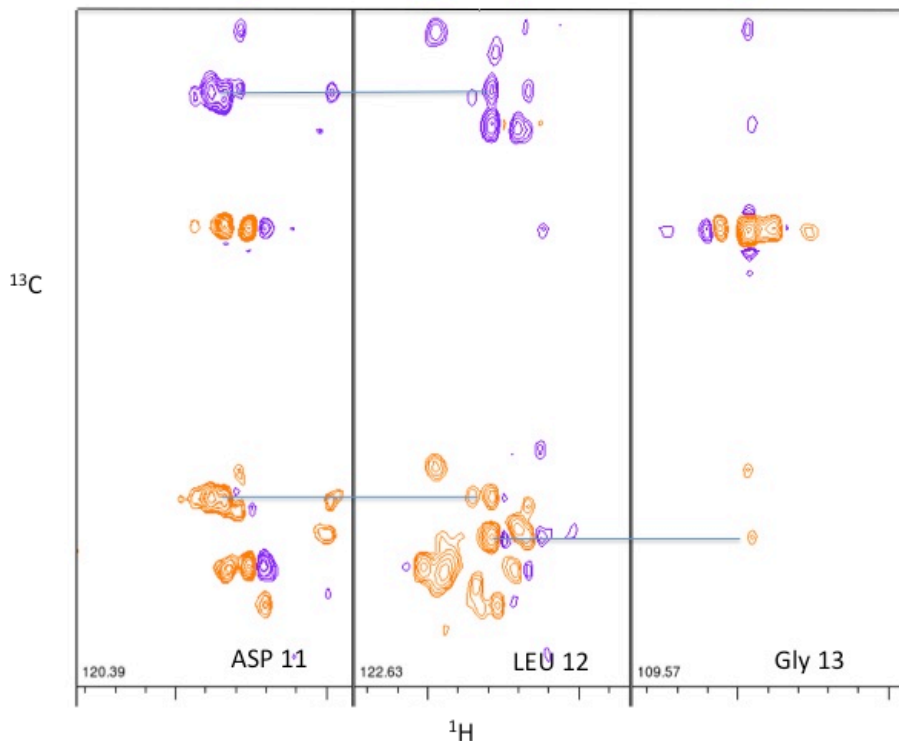


FIG. 34 Sample of assignment of TCF4 protein at $\text{pH}=4.5$ showing 3 different strips of ^1H - ^{13}C spectra at 3 different ^{15}N frequencies in the HNCACB spectrum and how to connect the three residues together. $\text{C}\alpha$ peaks are in orange, while $\text{C}\beta$ peaks are in violet.

Theoretical

2.3. Dynamic Combinatorial Chemistry

Dynamic combinatorial chemistry is an outstanding new approach that combines the library generation and screening processes⁷¹. One of the key issues in designing dynamic libraries is the choice of the reversible chemical reaction that can interconvert the library components. This reversible interconversion between the library components makes the library mainly governed by thermodynamics rather than kinetics⁷². Unfortunately, the reversibility of the reaction between the building blocks of the library puts a limitation on the type of reactions that can be used for such a technique.

In addition, the second crucial point in designing a dynamic combinatorial library is to have the availability to do in-situ measuring of the binding between the library components and the target protein. The reactions that can be used for dynamic library shouldn't interfere with the method used to test the binding.

A number of such reactions have been tested, which include transacylation⁷³, imine exchange⁷⁴, isomerization⁷⁵, thiol-disulfide exchange⁷⁶ and ligand exchange in coordination complexes⁷⁷.

Herein, we showed that imine formation is a fast and efficient reversible reaction that can be used to generate systems of potential β -Catenin inhibitors under neutral conditions. These systems were also subjected to direct binding analysis by saturation transfer difference (STD) 1D NMR spectroscopy, a rapid and efficient technique for protein–ligand-binding studies⁷⁸, in which the optimal imine interactions with the biological target can be readily deduced.

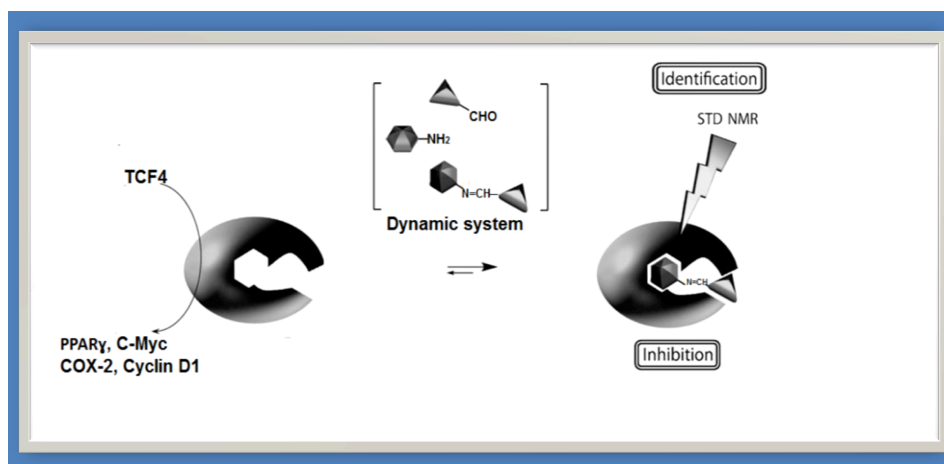


FIG. 35 Concept of direct STD NMR identification of β -Catenin inhibitors from a virtual dynamic imine system. (The figure is adapted from Caraballo *et al.*⁷⁹).

Theoretical

The addition of 2 amines to 2 aldehydes through nucleophilic addition was done in the NMR tube in absence and in presence of the targeted protein β -Catenin resulting in the synthesis of four new imines. ^1H STD NMR was performed for each separate component of the library and for the library components together by tracking the proton signal of formyl and imine groups.

This combination of dynamic system formation with STD NMR spectroscopy, resolving complex systems by the identification of the best ligands, can be easily expanded to accommodate a wider variety of compounds, resulting in efficient and rapid mapping of protein inhibitory potentials.

2.4. Isothermal titration calorimetry (ITC)

ITC is a non-destructive technique that directly measures the change in enthalpy (ΔH), stoichiometry (n) and the dissociation constant (K_D) for bimolecular binding interaction. ITC can also be used to give complete thermodynamic and kinetic profile of a certain reaction in a one or two-hour experiment⁸⁰. In addition, the samples can be recovered to be used for other measurements since the technique isn't invasive and doesn't require temperature elevation. However, the experiment must be well-designed and certain information like reactants concentrations need to be understood⁸¹. It's also very crucial to consider the sample preparation, the experiment design and the instrument used⁸².

A typical ITC instrument consists of two identical cells made of an inert material that is highly efficient thermal conducting; one is a reference containing water, buffer or solvent to provide a reference of comparable heat capacity. The other cell, a sample cell which typically contains 0.5-1.5 ml of the protein sample in which ligand is injected at a fixed temperature in the range 5-60°C. Ligand which is in the syringe and titrated into the sample cell should have the same buffer system and salt conditions as the protein. The main drawback about using ITC is that it's not up for large scale analysis so it can't be used for example to investigate binding of libraries.

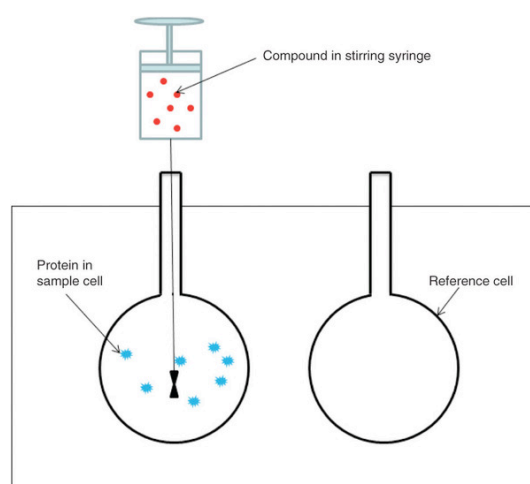


FIG. 36 Schematic presentation of ITC instrument. (Figure adapted from Zhou *et al.*⁸³).

Upon making injection of ligand, heat is either released or consumed (the process is endothermic or exothermic). These small heat effects are detected as the thermal power ($\mu\text{cal/s}$) needed to compensate for the ligand induced heat effect and to put the

Theoretical

cell back into thermal equilibrium. Initial injection of ligand into the cell gives large heats since almost all injected ligand binds to the targeted protein. However, as the titration continues, binding sites become more and more occupied and the amount of the heat per injection declines. Eventually, all the binding sites are filled (saturation) and the heat change corresponds only to dilution heats. Sample of raw data obtained from ITC is shown in figure 37.

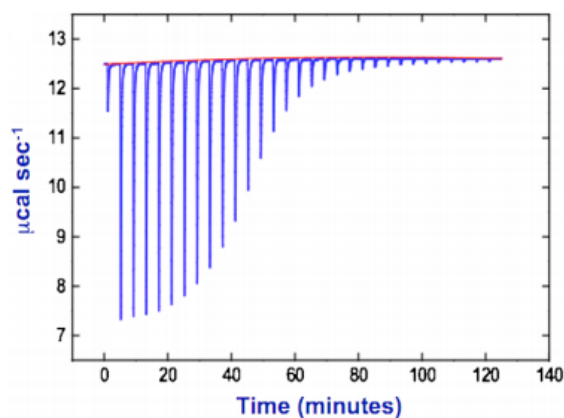


FIG. 37 Typical ITC raw data for a 1:1 interaction. Each peak corresponds to one injection of ligand into protein solution. The peaks indicative of the power used to compensate for the temperature difference to maintain thermal equilibrium.

Integrating these peaks then treats the raw data and doing correction to a per mole basis that allows having the binding isotherm. Figure 38 shows the data processing result, which gives a sigmoidal curve when the heat is plotted as a function of ligand/protein molar ratio.

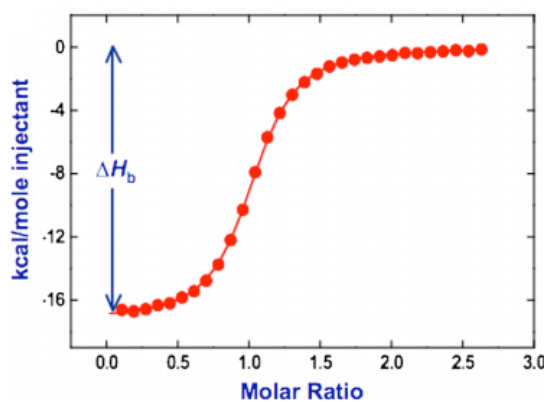


FIG. 38 ITC processed data gives sigmoidal curve from which binding isotherm is observed.

2.5. Protein Expression

The production of purified recombinant proteins is of great interest for the protein chemists⁸⁴. In addition, Many criteria control the choice of the expression approach like cell growth features, expected expression levels, post-translational modifications (e.g. phosphorylation), and biological activity of the targeted protein⁸⁵. Different hosts are being used currently by biologists to synthesize and express proteins, for example, *Escherichia coli*, yeast, protozoa, insects and mammalian cells. In this research, *E.coli* had been used as a host to express both proteins of interest; human armadillo repeat domain of β -Catenin and its binding partner TCF4.

2.5.1. *Escherichia coli* as a host for proteins expression

Although different expression hosts are available for protein production, the most commonly used for industrial purposes still remains the gram-negative bacteria; *E. coli*⁸⁶. A disadvantage for therapeutic use of produced recombinant proteins in *E. coli* is the accumulation of endotoxins (lipopolysaccharide), which are pyrogenic in humans and other mammals. Proteins that are intended to be used for pharmaceutical industry must be purified in a second step to become endotoxin-free⁸⁷. It's noteworthy that the *E.coli* lactose utilization (lac operon) has been used as the regulatory element to produce the targeted polypeptides. For lac operon, IPTG, which is a lactose analogue, is used as inducer and it allows the production of considerable amount of the targeted protein. Moreover, pET vectors are also commonly used in the *E.coli* process of protein production. The main disadvantage of using pET vectors is the production of large amount of mRNA, which can in turn cause ribosome destruction and cell death⁸⁸.

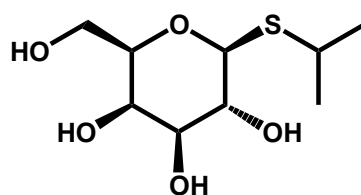
Alternative hosts are now considered to be used as more accessible and affordable. For example, eukaryotic systems including mammalian, yeast and insect cell expression are becoming easier to use and less expensive⁸⁹. Cell-free protein synthesis also has great potential for overcoming some of the problems of soluble protein expression⁹⁰.

2.5.2. *Expression and induction*

To start the expression, the plasmid containing the fusion target protein sequence is transformed into the respective *E.coli* cell line. When starter culture

Theoretical

reaches a suitable optical density (OD_{600}), it's diluted 1:100 in the expression culture depending on the protein we need. If unlabeled protein is targeted, the expression culture is L.B., which contains the appropriate antibiotics. On the other hand, if ^{15}N , and/or ^{13}C labeled protein is targeted, the expression culture is done in minimal medium (e.g. M9) enriched with ^{15}N labeled ammonium chloride and/or ^{13}C labeled glucose. Afterwards, the expression culture is incubated with shaking until OD_{600} reaches 0.5. In order to start the induction of protein synthesis, isopropyl- β -D-thiogalactopyranosid (**12**) IPTG, is added to the expression medium⁹¹.



12

The expression culture is then incubated depending on the nature of the protein and then the cells are harvested and resuspended in the appropriate buffer, which depends on the nature of the expressed protein. In this investigation, lysis and PBS buffers have been used to suspend the pellets for β -Catenin and TCF4 respectively.

2.5.3. Dithiothreitol, a protective reagent for SH Group

Proteins that have cysteine amino acid in their composition tend to form inter- and intramolecular disulfide linkages. In order to inhibit the formation of this disulfide linkage, a reducing agent; dithiothreitol DTT is added to the buffers used for pellets resuspension and for purification⁹².

DTT is a strong reducing agent, with a redox potential of -0.33 V at pH 7, while the pKa of SH groups is around ~ 8.3 . The reduction of a disulfide bond continues by two consecutive thiol-disulfide exchange reactions and ends with the formation disulfide-bonded six membered ring.

Theoretical

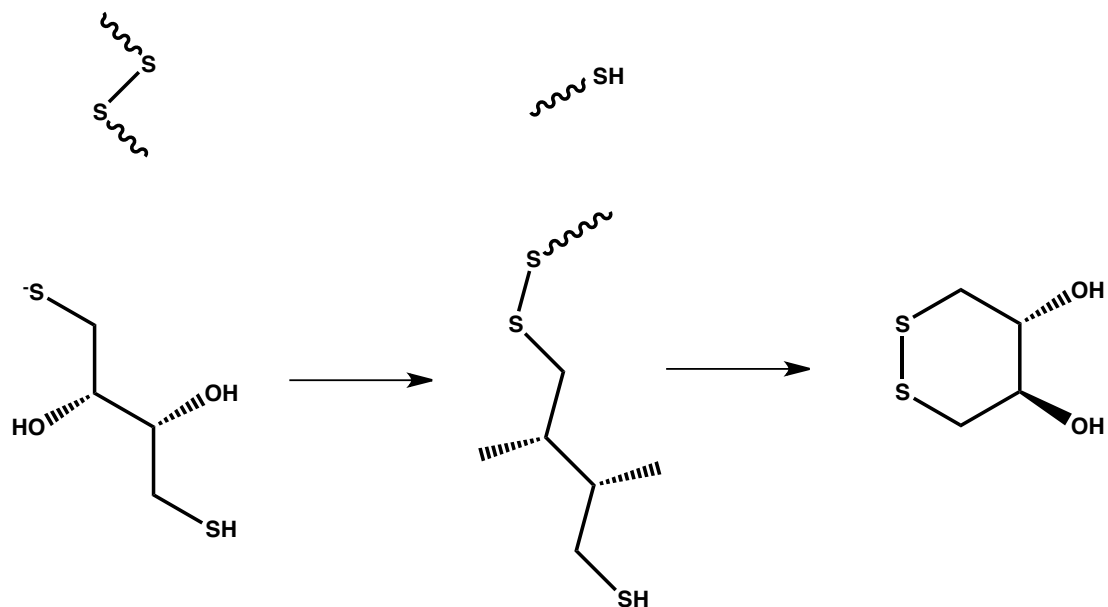


FIG. 39 Two steps reduction of protein disulfide linkage using DTT.

2.5.4. Chromatographic purification of protein

In this research, proteins purification was performed on fast performance liquid chromatography (FPLC) (Äktaexplorer from amersham pharmacia biotech, Sweden). The purification techniques used are affinity chromatography and size exclusion chromatography.

2.5.4.1. Affinity chromatography

2.5.4.1.1. Glutathione-S-transferase tag (GST)

GST is a 26 KDa tag that is fused to the protein in order to enhance the solubility, to allow easy production, selectivity and protection against intracellular protease cleavage. The targeted protein is expressed in a plasmid vector that contains a GST gene (e.g. *pGEX-6P*). Later on, the fused protein can be easily purified on a GSTrap column, which has immobilized glutathione, then the bound protein is eluted using 10mM of reduced glutathione containing buffer⁹³. Moreover, the GST can be fused to the *N*- or *C*-terminal of the protein of interest. In this investigation, armadillo repeat region of β -Catenin and all its mutants were expressed as GSTagged proteins and were purified using GSTrap FF 5 ml column (Amersham Pharmacia) which has a glutathione ligand linked *via* 10 carbon linker to highly cross linked 4% agarose.

Theoretical

2.5.4.1.2. Polyhistidine tag (*His-tag*)

One of the most widely used tags in affinity chromatography purification is the polyhistidine tag (which consists usually of 6 histidine residues). Histidine has imidazole ring that has electron donor group forming coordinate bond with the immobilized metal ion matrix in the hitrap column. The charged column has a Ni⁺²-NTA matrix (NTA is nitrilotriacetic acid which is a chelating agent in highly cross linked agarose matrix and binds Ni⁺² by four coordinate sites), which binds to the histag, then the bound protein is eluted using high imidazole containing buffer which replaces the target protein from the column⁹⁴. In this research, TCF4 and its mutants were expressed as his-tagged proteins and purified using hitrap-column 5ml (Amersham Pharmacia).

Theoretical

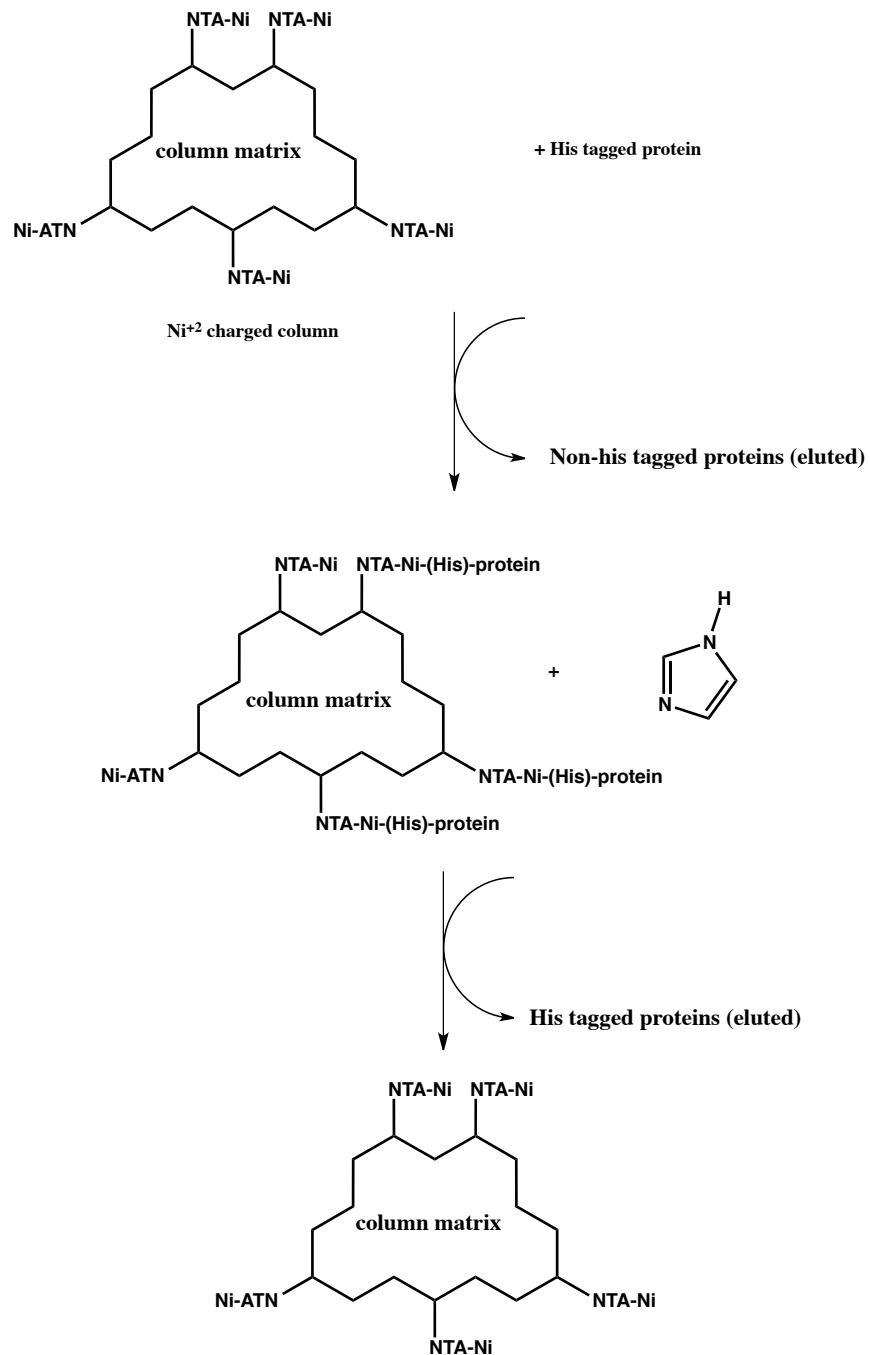


FIG. 40 Schematic representation of the purification protocol of His-tagged protein using Ni^{+2} -NTA matrix column.

Theoretical

2.5.4.2. Size exclusion chromatography

Size exclusion columns consist of dextran covalently bound to highly cross-linked agarose. The separation range of these columns is different depending on the pore size. Column 16/60 Supradex 75 prep grade column gives excellent resolving power of proteins in the molecular weight range of 3-70 KDa (mean particle size 34 μm). This column has been used as a final purification step for β -Catenin and TCF4.

2.5.5. Protein concentration & purity measurements

For measuring the protein samples in NMR, it's crucial to know the exact concentration of our sample and to prove its purity.

Protein concentration can be measured accurately using the Lambert-Beer law ($A = \epsilon cl$) where A is the absorbance, ϵ is the molar extinction coefficient of the protein $\text{M}^{-1}\text{cm}^{-1}$, c is the molar concentration M , and l is the path length cm . If the protein contains tryptophan(s) residues, the extinction coefficient of 1 tryptophan is ($5690 \text{ M}^{-1}\text{cm}^{-1}$ at $\lambda=280\text{nm}$); the ϵ of the whole protein is calculated by multiplying 5690 by the total number of tryptophan residues. Furthermore, the absorbance is measured at wavelength 280 nm ⁹⁵.

In order to elucidate the purity and correct size of the expressed protein, the sodium dodecyl sulfate (SDS) gel electrophoresis technique is used⁹⁶. The separation of the protein is based on its size and charge. The gel consists mainly of acrylamide, bisacrylamide, SDS, and a buffer. Upon applying electric field across the gel, the negatively charged proteins migrate towards the positive (+) electrode. According to the protein molecular weight, it will diffuse differently through the gel. If it's a small protein, it will move more easily through the pores in the gel and diffuse faster, while if it's a large size one, it will experience more difficulty to diffuse and so, it will move slower. Hence, Proteins can be separated according to size (molecular weight). Moreover, a molecular weight marker is used in order to act as a reference for the protein sample size (Dalton Mark VII-L Standard Mixture, Sigma Aldrich).

EXPERIMENTAL

(materials, methods & discussion)

Experimental

3.1. Materials:

3.1.1. Microorganisms:

Organism	Strain
<i>Escherichia (E.) coli</i>	BL (21 DE3) Rosette (DE3) pLys S DH5 α
All strains derived from Novagen	

3.1.2. Media for Bacterial Growth:

1-L.B. Media For β -Catenin expression	Components/ Amounts
Luria Broth Media	20 g Luria Broth
Autoclaved for 15 min. at 121°C	Ampicillin in concentration 100mg/ml
Store at 4 °C	Add distilled water to volume 1000ml

Experimental

2-M9 Media For TCF4 expression	Components/ Amounts
Autoclaved for 15 min. at 121 °C	6 g Na ₂ HPO ₄ x 2 H ₂ O
Store at 4 °C	3 g KH ₂ PO ₄
	0.5 g NaCl
	1 g NH ₄ Cl
	Ampicillin in concentration 100mg/ml
	Chloramphenicol in concentration 25mg/ml
	Add distilled water to volume 1000ml
	Prior to use add:
	20 ml of 20% glucose
	2 ml of 1 M MgSO ₄
	0.3 ml CaCl ₂
	10 ml trace elements solution

3-M9 Media For ¹⁵N-labelled protein expression	Components/ Amounts
Autoclaved for 15 min. at 121°C	s. M9 medium
Store at 4 °C	Use ¹⁵ NH ₄ Cl instead of NH ₄ Cl

Experimental

4-M9 Media For ¹⁵N, ¹³C-labelled protein expression	Components/ Amounts
Autoclaved for 15 min. at 121°C	s. M9 medium
Store at 4 °C	use ¹⁵ NH ₄ Cl instead of NH ₄ Cl
	add 3 g ¹³ C-Glucose after autoclaving

3.1.3. Buffers and solutions:

1-Ampicillin	Components/ Amounts
Ampicillin	
100 mg/ml (≈0.27 M)	1 g ampicillin
Sterilize by filtration (0.2 μm).	H ₂ O to a final volume of 10 ml
Store at -20 °C.	

2-CaCl₂	Components/ Amounts
1 M	14.7 g CaCl ₂ × 2H ₂ O
Autoclave at 121°C for 15 min.	H ₂ O to a final volume of 100 ml

3-Chloramphenicol	Components/ Amounts
Chloramphenicol	
25 mg/ml	250 mg chloramphenicol
Store at -20 °C.	EtOH to a final volume of 10 ml

Experimental

4-Coomassie staining solution(for protein SDS page)	Components/ Amounts
0.5% (w/v) Coomassie	2.5 g Coomassie Brilliant Blue R-250
50% (v/v) MeOH	250 ml MeOH (98%)
10% (v/v) HOAc	50 ml HOAc (99%)
Store light protected.	H ₂ O to a final volume of 500 ml

5-Destaining solution (for protein SDS page)	Components/ Amounts
30% (v/v) EtOH	300 ml EtOH (98%)
10% (v/v) HOAc	100 ml HOAc. (99%)
	H ₂ O to a final volume of 1 l

6-DNA sample buffer (5x) (for agarose gel electrophoresis)	Components/ Amounts
20 mM EDTA	100 µl EDTA 1 M (pH 8.0)
30 % (w/v) Glycerine	5 mg bromphenol blue
0.5 % (w/v) SDS	1.7 ml glycerine
0.1 % (w/v) Bromphenol blue	125 µl SDS 20 %
Store at 4°C.	H ₂ O to a final volume of 5 ml

Experimental

7-dNTP Mix (For synthesis of TCF4 clones)	Components/ Amounts
5 mM Datp	5 μ l dATP 100 mM
5 mM dCTP	5 μ l dCTP 100 mM
5 mM dGTP	5 μ l dGTP 100 mM
5 mM dTTP	5 μ l dTTP 100 mM
Aliquote and store at -20 °C.	H ₂ O to a final volume of 100 μ l

8-Dithiothreitol (DTT)	Components/ Amounts
1 M	
Sterilize by filtration (0.2 μ m).	1.54 g dithiothreitol
Store at -20 °C.	H ₂ O to a final volume of 10 ml

9-EDTA buffer	Components/ Amounts
0.05 M	9.306 g EDTA
	H ₂ O to a final volume of 500 ml add 5 M NaOH to reach pH=7

10-Glucose	Components/ Amounts
1 M	
Sterilize by filtration (0.2 μ m).	19.8 g Glc \times H ₂ O
Store at 4 °C.	H ₂ O to a final volume of 100 ml

Experimental

11-HEPES buffer (for β-Catenin reductive methylation)	Components/ Amounts
10 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid	2.38 g HEPES
150mM NaCl	8.76 g NaCl
1mM DTT	1 ml DTT
Autoclave at 121 °C for 15 min.	800 ml H ₂ O
	Adjust pH to 7.4 by adding conc HCl and bring volume to 1 l.

12-Imidazole buffers (for TCF4 chelating HiTrap column)	Components/ Amounts
20 mM trizma base	2.42 g Trizma base
500 mM NaCl	29.2 g NaCl
2.5% w/v glycerol	25 g glycerol
a-10mM imidazole (low-imidazole buffer)	a-0.68 g imidazole
b-500 mM imidazole (Hi-imidazole buffer)	b-34.04 g imidazole
	900 ml H ₂ O
	Adjust pH to 7.4 by conc HCl
	Bring volume to 1 l.

Experimental

13-Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Components/ Amounts
1 M Sterilize by filtration (0.2 μ m). Store at 4 °C.	2.4 g IPTG (238.25 g/mol) H ₂ O to a final volume of 10 ml

14-Kanamycin	Components/ Amounts
25 mg/ml Sterilize by filtration (0.2 μ m). Store at -20 °C.	250 mg Kanamycin monosulfate H ₂ O to a final volume of 10 ml

15-Laemmli buffer (Electrophoresis running buffer for protein SDS page)	Components/ Amounts
25 mM Tris 190 mM Glycine 0.1 % SDS.	250 mg Kanamycin monosulfate H ₂ O to a final volume of 10 ml

Experimental

16-Lysis buffer (β-Catenin)	Components/ Amounts
PBS buffer (pH 7.4)	0.26 g KH_2PO_4
1 mM Na_3VO_4	2.17 g $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$
20 mM DTT	8.71 g NaCl
	0.09 g Na_3VO_4
	20 ml DTT
	800 ml H_2O , adjust pH to 7.4 and bring volume to 1 l.

17-Nickel Chloride solution (For charging of Chelating Hitrap Column)	Components/ Amounts
0.1 M NiCl_2	1.29 g anhydrous NiCl_2
	Add H_2O to bring volume to 100 ml

18-PBS buffer (β-Catenin & TCF4)	Components/ Amounts
140mM NaCl	8.17 g NaCl
2.7 mM KCl	0.2 g KCl
10mM Na_2HPO_4	1.77 g Na_2HPO_4
1.8mM KH_2PO_4	0.24 g KH_2PO_4
2mM DTT	2 ml DTT
	800 ml H_2O , adjust pH to 7.4 and bring volume to 1 l.

Experimental

19-Protein size marker (For protein SDS page)	Components/ Amounts
1.4 µg/µl total protein Store at -80°C.	2.1 mg SDS7 Dalton mark (1 vial) Protein sample buffer 1 × to a final volume of 1.5 ml

20-2x Protein sample buffer (For protein SDS page)	Components/ Amounts
120 mM Tris × HCl, pH 6.8	1.2 ml Tris × HCl 1 M (pH 6.8)
6 % (w/v) SDS	3 ml SDS 20 %
20 % (v/v) Glycerine	2 ml glycerine
0.01 % (w/v) Bromphenol blue	1 mg bromphenol blue
10 % (v/v) β-Mercaptoethanol	H ₂ O to a final volume of 9 ml
Store at -20 °C. Thaw only once.	Add 100 µl of β-mercaptoethanol to 900 µl of buffer, then store at 4°C.

21-Stop Mix (5x) for DNA restriction	Components/ Amounts
20 mM EDTA	0.1ml 1 M EDTA (pH=8.0)
30 % Glycerin	1.7 ml Glycerin
0.5 % SDS	0.125 ml 20 % SDS
0.1 % Bromphenolblue	2 ml 0.5% Bromphenolblue
	H ₂ O to a final volume of 5 ml.

Experimental

22-TBE buffer (Running buffer for agarose gel)	Components/ Amounts
10×	
pHcalc 8.3	108 g tris base
0.89 M Tris	55 g H ₃ BO ₃
0.89 M Boric acid	7.44 g EDTA Na ₂ ×2H ₂ O
0.02 M EDTA	H ₂ O to a final volume of 1 l
Autoclave at 121 °C for 15 min.	

23-TEV-protease buffer	Components/ Amounts
50 mM trizma base	6.057 g tris base
0.5 mM EDTA	0.186 g EDTA
1mM DTT	1 ml DTT
Autoclave at 121 °C for 15 min.	800 ml H ₂ O
	Adjust pH to 8 by adding conc HCl and bring volume to 1 l.

Experimental

24-Tris.HCl buffer (for β-Catenin size exclusion chromatography)	Components/ Amounts
100 mM trizma base	12.1 g trizma base
150mM NaCl	8.76 g NaCl
1mM DTT	1 ml DTT
Autoclave at 121 °C for 15 min.	800 ml H ₂ O
	Adjust pH to 7.4 by adding conc HCl and bring volume to 1 l.

25-Tris.Glutathione buffer (elution buffer for β-Catenin GSTrap column)	Components/ Amounts
50 mM trizma base	6.05 g trizma base
20mM reduced Glutathione	6.14 g reduced glutathione
Autoclave at 121 °C for 15 min.	800 ml H ₂ O
	Adjust pH to 7.4 by adding conc HCl and bring volume to 1 l.

3.2. Methods, Results and Discussions:

3.2.1. Expression of GST-tagged Armadillo Repeat region of β -Catenin protein:

1-Transformation:

To 100 μ l of BL21 (DE3) *E.coli* cell line, 10 ng of β -Catenin *Pgex6p* vector (Ampicillin resistant) were added in eppendorf tube. The mixture was left in ice bath for 10 minutes, then in water bath at 42 °C for 1.5 minutes, in ice bath again for 2 minutes. To this mixture, 300 μ l of L.B. medium was added in the eppendorf (1 bacteria cell line volume: 3 L.B. medium), then shaken for 45 minutes in shaker at 37 °C. An inoculum of 50 μ l of the plasmid/bacteria mixture was transferred to Luria–Broth (LB)-Agar plate that contains ampicillin, spreaded then incubated in the incubator at 37 °C overnight.

2- Pre-culture & overnight culture:

In a mixture of 50 ml sterilized L.B. & 50 μ l ampicillin, one colony of the grown bacteria was added and then put in the shaker overnight at 30 °C.

3- Expression Culture:

To 1 l of sterilized L.B., 1 ml ampicillin was added. Then 10 ml of the overnight culture (to have 1:100 concentration) were supplemented. Shaking took place at 37 °C and then Optical Density (OD_{600}) was measured every 45 minutes until (OD_{600}) =0.5 was reached. Finally, 1 ml of 4 M IPTG was added to the 1 l (to have final concentration of 0.4 mM IPTG). The mixture was shaken overnight at 20 °C.

4- Harvesting:

The OD_{600} was measured to be sure that it increased overnight. Centrifugation (4000 rpm, 4 °C, 20 minutes) was used to harvest the cells. The

Experimental

pellets were resuspended in 25 ml lysis buffer and the aliquot was frozen in liquid nitrogen for 5 minutes and stored at -80°C until purification.

5- Sonication & DNA precipitation:

The sample is thawed and a few amount of lysozyme is added, then cells were lysed by ultrasonication. The sample was centrifuged (18000 rpm, 4 °C, 20 minutes) to get rid of the broken cells. To the clear supernatant, 10% w/v streptomycin solution was added to precipitate the DNA in a ratio 1:10. The sample was again centrifuged (18000 rpm, 4 °C, 20 minutes) to remove the precipitated DNA.

6-Chromatographic purification:

The chromatographic purification strategy has been applied in a series to run on a currently existing FPLC chromatographic system ÄKTAE Explorer (Amersham Biosciences, Uppsala, Sweden). Chromatographic profiles continuously monitor absorbance (260 and 280 nm).

a- GSTrap FF 5 ml column:

The buffer solutions were degased using a water jet pump and filtrated through a ZapCapS bottle top filter (0.2 µm). The column was equilibrated with 5-column volume of the binding buffer (PBS buffer). The sample supernatant obtained from the last step was loaded on the column in a flow rate 1 ml/minute and the loading was repeated five times to ensure binding of the GST-tag protein because of the low binding kinetics.

Afterwards, the column was washed with PBS buffer until the absorbance return to zero and the base line is stable. The sample was then eluted using the elution buffer (Tris/Glutathione buffer). Based on the chromatogram, the protein fractions were concentrated to a volume of 3ml to be used for the next size exclusion chromatography.

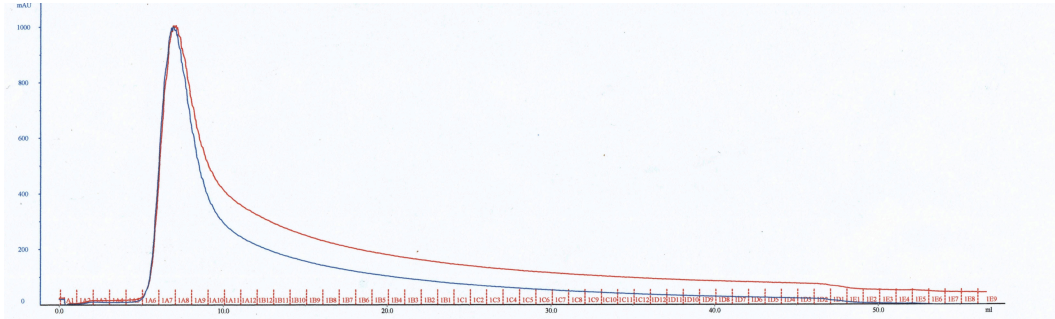


FIG. 41 Chromatogram of the GST-tagged armadillo repeat β -catenin after loading on GSTrap FF column and subsequent elution with buffer containing glutathione. The fractions containing the protein are tracked by the absorbance at 260 nm (in red) and 280 nm (in blue).

b- Gel filtration Chromatography (size exclusion Hiload column 16/60):

The column is equilibrated with 1.5 column volume of Tris-HCl buffer until the base line is stable. The sample Obtained from the GSTrap FF column was injected through a 5 ml loop and then the sample fractions were eluted using the Tris-HCl running buffer. The protein comes out from the column based on its size (85 KDa). Based on the obtained chromatogram, the protein fractions were collected.

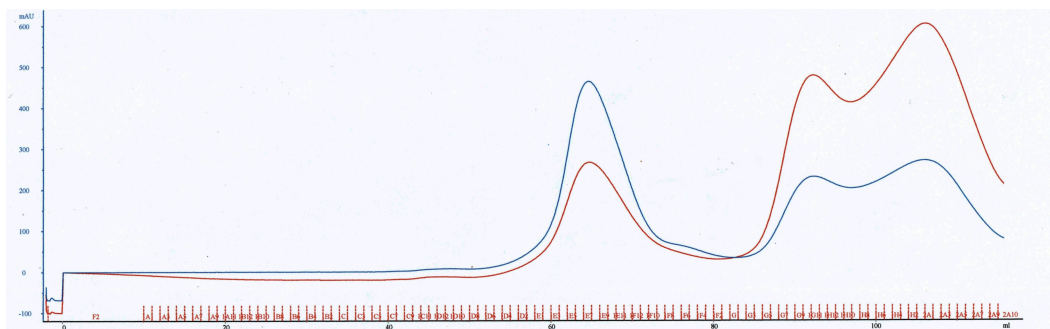


FIG. 42 Chromatogram of the GST-tagged armadillo repeat β -Catenin after loading on Hiload 16/60 and subsequent elution. The fractions containing the protein are tracked by the absorbance at 260 nm (in red) and 280 nm (in blue). The first peak is the protein's while the second is DNA.

7- SDS polyacrylamide gel Electrophoresis (SDS-page of proteins)

The gel was put into a running cell filled with Laemmli buffer. The different sample fractions mixed with 2x protein buffer (10 μ l protein + 10 μ l 2x buffer) were loaded onto the gel. Use protein size marker for referencing. The gel was run at 200 V for 40 min. Afterwards, the gel was stained with Coomassie staining solution for at least 30 min. Later, the gel got destained using the destaining solution until the gel is colourless again, except the protein bands. A picture of the gel was taken with a Gel Doc System (Sony).

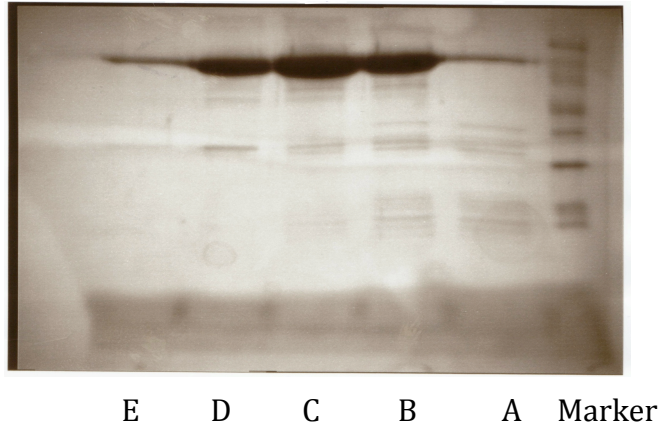


FIG. 43 Purified human GST-tagged armadillo repeat β -Catenin protein. Coomassie stained 15% SDS-polyacrylamide gels show pure fractions of recombinant protein after GSTrap and gel filtration columns. Numbering of lanes (A-E) corresponds to different collected fractions based on the chromatogram of the gel filtration step.

8- Concentration and quantification of the protein sample:

The protein sample is transferred to a concentration device (centricon), which has a cut off value of 10 KDa. The centricon was then centrifuged (4000 rpm, 4 °C) for the suitable time until the required concentration is reached. The excess buffer passes through the membrane while the protein sample is retained above the membrane.

To quantify the concentration of the final volume of the protein, UV absorbance at 280 nm was measured. The extinction coefficient for β -Catenin is $75000 \text{ M}^{-1}\text{cm}^{-1}$. The protein was obtained in a maximum concentration of 80 μM . At higher concentrations, it started to aggregate and precipitate.

The expressed armadillo repeat domain of human β -Catenin is the core region (the residues 134-671). It consists of 12 repeats; each is 3 α -helices. The sample was sent for sequencing in LGC GENOMIC Company, Berlin, Germany.

Experimental

Human β -Catenin armadillo repeats Sequence

10 20 30 40 50 60
HAVVNLINYQ DDAELATRAI PELTKLINDE DQVVVNKAAV MVHQLSKKEA SRHAIMRSPQ

70 80 90 100 110 120
MVSAIVRTMQ NTNDVETARC TAGTLHNLSH HREGLLAIFK SGGIPALVKM LGSPVDSVLF

130 140 150 160 170 180
YAITTLHNLL LHQEGAKMAV RLAGGLQKMV ALLNKTNVKF LAITTDCLQI LAYGNQESKL

190 200 210 220 230 240
IILASGGPQA LVNIMRTYTY EKLLWTTSRV LKVLVCSSEN KPAIVEAGGM QALGLHLTDP

250 260 270 280 290 300
SQRLVQNCLW TLRNLSDAAT KQEGMEGLLG TLVQLLGSDD INVVTCAAGI LSNLTCNNYK

310 320 330 340 350 360
NKMMVCQVGG IEALVRTVLR AGDREDITEP AICALRHLS RHQEAEMAQN AVRLHYGLPV

370 380 390 400 410 420
VVKLLHPPSH WPLIKATVGL IRNLALCPAN HAPLREQGAI PRLVQLLVRA HQDTQRRTSM

430 440 450 460 470 480
GGTQQQFVEG VMEEIVEGC TGALHILARD VHNRI VIRGL NTIPLFVQIL YSPIENIQRV

490 500 510 520 530
AAGVLCELAQ DKEAAEAIEA EGATAPLTEL LHSRNEG VAT YAAAVLFRMS EDKPDYK

Number of amino acids: 538

Molecular weight: 58773.2

Theoretical pI: 8.26

Experimental

Amino acid composition:

Residue	Number of residues	%Composition
Ala	57	10.6%
Arg	30	5.6%
Asn	27	5.0%
Asp	18	3.3%
Cys	11	2.0%
Gln	29	5.4%
Glu	31	5.8%
Gly	34	6.3%
His	19	3.5%
Ile	31	5.8%
Leu	71	13.2%
Lys	22	4.1%
Met	16	3.0%
Phe	6	1.1%
Pro	19	3.5%
Ser	24	4.5%
Thr	34	6.3%
Trp	3	0.6%
Tyr	10	1.9%
Val	46	8.6%

Total number of negatively charged residues (Asp + Glu): 49

Total number of positively charged residues (Arg + Lys): 52

Experimental

Atomic composition:

Carbon	C	2580
Hydrogen	H	4247
Nitrogen	N	747
Oxygen	O	761
Sulfur	S	27

Formula: $C_{2580}H_{4247}N_{747}O_{761}S_{27}$

Total number of atoms: 8362

Extinction coefficients:

Extinction coefficient is in units of $M^{-1} \text{ cm}^{-1}$, at 280 nm measured in water.

Ext. coefficient 32025

Abs 0.1% (=1 g/l) 0.545, assuming all pairs of Cys residues form cystines

Ext. coefficient 31400

Abs 0.1% (=1 g/l) 0.534, assuming all Cys residues are reduced

GSTag

10 20 30 40 50 60
MSPILGYWKI KGLVQPTRL L LEYLEEKYEE HLYERDEGDK WRNKKFELGL EEPNLPYYID

Experimental

70 80 90 100 110 120
GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA YSKDFETLKV

130 140 150 160 170 180
DFLSKLP EML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL DAFPKLVCFK

190 200 210 220
KRIEAIPQID KYLKSSKYIA WPLQGQATF GGGDHPPKSD LVPRGSP

Number of amino acids: 227

Molecular weight: 26389.6

Theoretical pI: 5.91

Number of amino acids: 227

Molecular weight: 26389.6

Theoretical pI: 5.91

Experimental

Amino acid composition:

Residue	Number of residues	%Composition
Ala	10	4.4%
Arg	10	4.4%
Asn	4	1.8%
Asp	18	7.9%
Cys	4	1.8%
Gln	5	2.2%
Glu	17	7.5%
Gly	15	6.6%
His	6	2.6%
Ile	13	5.7%
Leu	28	12.3%
Lys	21	9.3%
Met	9	4.0%
Phe	8	3.5%
Pro	15	6.6%
Ser	10	4.4%
Thr	6	2.6%
Trp	4	1.8%
Tyr	14	6.2%
Val	10	4.4%

Total number of negatively charged residues (Asp + Glu): 35

Total number of positively charged residues (Arg + Lys): 31

Experimental

Atomic composition:

Carbon	C	1204
Hydrogen	H	1861
Nitrogen	N	303
Oxygen	O	337
Sulfur	S	13

Formula: $C_{1204}H_{1861}N_{303}O_{337}S_{13}$

Total number of atoms: 3718

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.

Ext. coefficient 43110

Abs 0.1% (=1 g/l) 1.634, assuming all pairs of Cys residues form cystines

Ext. coefficient 42860

Abs 0.1% (=1 g/l) 1.624, assuming all Cys residues are reduced

Total weight of the synthesized human GST-tagged armadillo repeat β -Catenin =
58773.2 + 26389.6 = 85 K.D.

Experimental

3.2.2. Expression of His-tagged Transcription factor (TCF4):

1-Transformation:

To 100 μ l of Rosetta BL5 *E.coli* cell line (chloramphenicol resistant), 10 ng of TCF4 *pet-20b* vector (Ampicillin resistant) were added in eppendorf tube. The mixture was left in ice bath for 10 minutes, then in water bath at 42°C for 1.5 minutes, in ice bath again for 2 minutes. 300 μ l of L.B. medium were then added (1 bacteria cell line volume: 3 L.B. medium), then the eppendorf tube was shaken for 45 minutes at 37°C. 50 μ l of the plasmid/bacteria mixture was applied on Luria-Broth (LB)-Agar plate which contains chloramphenicol and ampicillin. After spreading the inoculum, the plate was left for 5 minutes, inverted and kept in incubator at 37°C overnight.

2- Pre-culture:

To 3 ml of sterilized L.B., 3 μ l chloramphenicol and 3 μ l ampicillin were added. Afterwards, one colony of the grown bacteria was pipette into the prepared L.B. medium and then put in the shaker for 4 hours at 37°C.

3- Overnight culture:

To 50 ml of sterilized L.B., 50 μ l chloramphenicol and 50 μ l ampicillin were added. 100 μ l of the previously prepared pre-culture was mixed with this mixture. The prepared culture was shaken overnight at 32°C.

4- Expression Culture:

To 1 l of sterilized M9 (¹⁵N or ¹³C & ¹⁵N labeled), 1 ml of chloramphenicol and 1 ml ampicillin were added. In addition, 10 ml of the overnight culture (to have 1:100 ratio) were added. Shaking took place at 37°C and then Optical Density (OD₆₀₀) was measured every 45 minutes until (OD₆₀₀)

Experimental

=0.5 was reached. Finally, 1 ml of 4 M IPTG was added to the 1 l (to have final concentration of 0.4 mM IPTG). The mixture was shaken for 4 hours at 37°C.

5- Harvesting:

The cells were harvested by centrifugation (4000 rpm, 4 °C, 20 minutes). The pellets were resuspended in 25 ml PBS buffer and the aliquot was frozen in liquid nitrogen for 5 minutes and stored at -80°C until purification.

6- Sonication:

After thawing the sample, one tablet of protease cocktail inhibitor was added, and then cells were lysed by ultra sonication. The sample was centrifuged (18000 rpm, 4 °C, 20 minutes) to get rid of the broken cells.

7-Chromatographic purification:

a- Hitrap chelating column:

The buffer solutions were degased using a water jet pump and filtrated through a ZapCapS bottle top filter (0.2 µm).

The column was recharged with 3 ml of 0.1 M NiCl₂ and washed with 5-column volume of H₂O.

The column was equilibrated with 5 column volume of the binding buffer (low imidazole buffer), then with 5 column volume of the eluting buffer (hi-imidazole buffer), and eventually, with 5 column volume of the binding buffer again.

The TCF4 sample supernatant obtained from the last step was loaded on the Ni⁺² charged column in a flow rate 1 ml/minute to allow binding of the protein histidine tag to the Ni⁺². The column was washed with low-imidazole buffer till a stable base line was obtained. The sample fractions were eluted during a gradient elution over 75 minutes (100% target concentration of the hi-imidazole buffer). Finally, based on the chromatogram, the protein fractions were concentrated to a volume of 3 ml to be used for the next size exclusion chromatography.

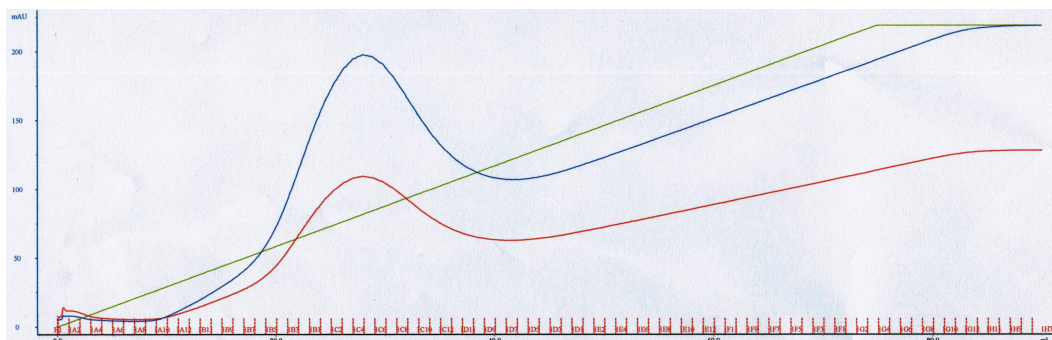


FIG. 44 Chromatogram of the His-tagged TCF4 after loading on Hitrap chelating column and subsequent elution with buffer gradient of low and hi-imidazole. The fractions containing the protein are tracked by the absorbance at 260 nm (in red) and 280 nm (in blue). The green line shows the gradient increase in the concentration of imidazole.

b- Gel filtration Chromatography (size exclusion Hiload column 16/60):

The column was equilibrated with 1.5 column volume of PBS buffer until the base line is stable. The sample Obtained from the previous chromatographic step was concentrated to 3 ml and injected through a 5 ml loop. Afterwards, the sample fractions were eluted using the PBS running buffer. The protein comes out from the column according to its size (7 KDa). Based on the obtained chromatogram, the protein fractions were collected.

Experimental

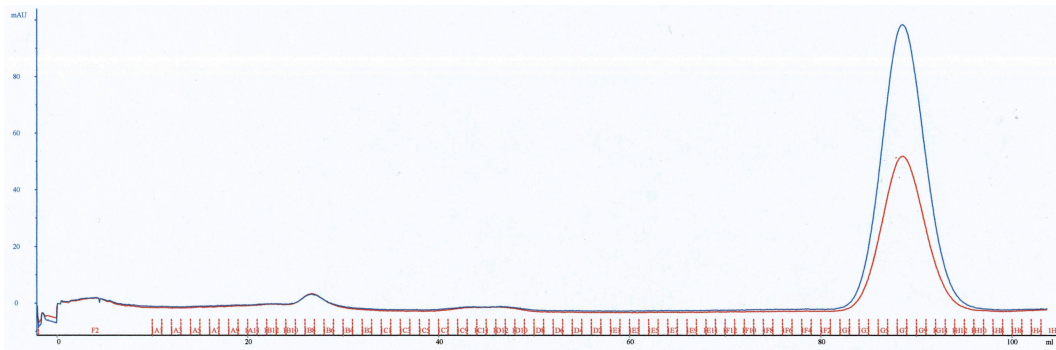


FIG. 45 Chromatogram of the His-tagged TCF4 after loading on Hiload 16/60 and subsequent elution with PBS buffer. The fractions containing the protein are tracked by the absorbance at 260 nm (in red) and 280 nm (in blue). The large peak is the protein's.

8- SDS polyacrylamide gel Electrophoresis (SDS-page of proteins)

The gel was run in the same way described before and different fractions based on the chromatogram have been tested.

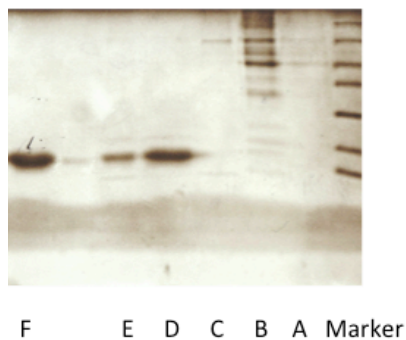


FIG. 46 Purified His-tagged TCF4 protein. Coomassie stained 15% SDS-polyacrylamide gels show pure fractions of recombinant protein after Chelating hitrap and gel filtration columns. Numbering of lanes (A-F) corresponds to the fractions collected based on the chromatogram of the gel filtration step.

Experimental

9- Concentration and quantification of the protein sample:

The protein sample is transferred to a concentration device (centricon) with a cut off value smaller than the molecular weight of the protein. The centricon was then centrifuged (4000 rpm, 4 °C) until the required concentration is reached.

To quantify the concentration of the obtained protein, UV absorbance at 280nm was measured. The extinction coefficient for His-tagged TCF4 is 5500 M⁻¹cm⁻¹. The protein was obtained in maximum concentration 1 mM. The sample was sent for sequencing in LGC GENOMIC Company, Berlin, Germany.

His-tagged Tcf4 sequence

MPQLNGGGGD¹⁰ DLGANDELIS²⁰ FKDEGEQEEK³⁰ SSENSSAERD⁴⁰ LADVKSSLVN⁵⁰ ESETNKNWLE⁶⁰

HHHHH

Number of amino acids: 65

Molecular weight: 7200.5

Theoretical pI: 4.52

Amino acid numbers and percent composition:

Residue	Number of residues	%Composition
Ala	3	4.6%
Arg	1	1.5%
Asn	6	9.2%
Asp	6	9.2%
Cys	0	0%
Gln	2	3.1%

Experimental

Glu	10	15.4%
Gly	6	9.2%
His	5	7.7%
Ile	1	1.5%
Leu	6	9.2%
Lys	4	6.2%
Met	1	1.5%
Phe	1	1.5%
Pro	1	1.5%
Ser	8	12.3%
Thr	1	1.5%
Trp	1	1.5%
Tyr	0	0%
Val	2	3.1%

Total number of negatively charged residues (Asp + Glu): 16

Total number of positively charged residues (Arg + Lys): 5

Atomic composition:

Carbon	C	299
Hydrogen	H	459
Nitrogen	N	91
Oxygen	O	115
Sulfur	S	1

Formula: C₂₉₉H₄₅₉N₉₁O₁₁₅S₁

Experimental

Total number of atoms: 965

Extinction coefficients:

Ext. coefficient 5500

Abs 0.1% (=1 g/l) 0.764

Total weight of the synthesized human His-tagged TCF4 = 7 K.D.

3.2.3. Expression of his-tagged truncated TCF4 mutants

Two truncated versions out of TCF4 have been expressed; fragment which has the residues from 10-30 and a fragment, another that has the residues from 32-52.

I-Cloning of DNA fragments

*In the PCR tube mix, the following components have been added

5 µl of 10x PFU buffer+MgSO₄

1 µl of 10 µM sense primer

1 µl of 10 µM anti-sense primer

2 µl of 50 ng/ µl template TCF4-plasmid

5 µl of dNTP

0.5 µl PFU-DNA polymerase

35.5 µl water (to have a final volume of 50 µl)

*PCR machine program:

1- 95 °C 30 seconds

2- 95 °C 30 seconds

3- 55 °C 1 minute

4- 68 °C 30 seconds

5- procedures from 2-4 were repeated 30 times

Experimental

6- 72 °C 10 minutes

7- hold at 4°C

*Digestion of the DNA with restriction enzymes (Not1, Nco1):

The PCR products were purified using PCR purification kit (Qiagen), and then to the eluted purified DNA, the following were added

3 µl of buffer O⁺

1 µl Not 1 restriction enzyme (10 u/ µl)

1 µl Nco1 restriction enzyme (10 u/ µl)

Digestion was done at 37 °C overnight

*Ligation:

50 ng of vector pET-gB1 (kanamycin resistant)

6-fold molar excess of insert DNA

2 µl of 10x T4 DNA ligase buffer

0.5 µl T4 DNA ligase

H₂O to final volume of 20 µl

The mixture is kept at 24°C for 1 h.

*Transformation of the ligated DNA into DH-5α

The 20 µl of the ligation mixture was transformed into 100 µl aliquot of *E.coli* DH-5α strain. After overnight growth, one colony was transferred into 10 ml of sterilized L.B. + 10 µl of kanamycin and allowed to grow overnight at 37°C. The L.B. was then centrifuged and the pellets were treated with miniprep kit (Qiagen) to obtain the purified DNA plasmid.

Experimental

II- Expression of the targeted ¹⁵N-labeled TCF4 mutants

The DNA plasmids of the 2 TCF4 mutants were transferred into 100 µl of Rosetta BL5 *E.coli* cell line (chloramphenicol resistant) and the same procedures used to express the wild type TCF4 were followed.

His-tagged mutant-Tcf4 sequence (N-terminal part)

10 20 30 40 50 60
MKHHHHHHHPM KQYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK
70 80 90 100
TFTVTEGSGS GSENLVFQGA **MGCEDLGAND ELISFKDEGE QEE**

Where the bold letters represent the sequence of the truncated TCF4 peptide, while the rest is the histag and the Gb1 fusion part.

Number of amino acids: 103

Molecular weight: 11426.3

Theoretical pI: 4.76

Extinction coefficient: 11460

His-tagged mutant-Tcf4 sequence (C-terminal part)

10 20 30 40 50 60
MKHHHHHHHPM KQYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK
70 80 90 100
TFTVTEGSGS GSENLVFQGA **MGSENSSAE RDLADVKSSL VNE**

Where the bold letters represent the sequence of the truncated TCF4 peptide, while the rest is the histag and the Gb1 fusion part.

Experimental

Number of amino acids: 103

Molecular weight: 11309.3

Theoretical pI: 5.39

Extinction coefficient: 11460

III- Cleavage of the Gb1- fusion using Tev-protease

In order to perform the NMR measurements on the expressed TCF4 mutants, it was crucial to get rid of the Gb1 fusion using the Tev-protease enzyme. Tev protease minimal recognition site is the sequence (ENLYFQG), where the cleavage occurs between Q and G residues.

The expressed truncated TCF4 peptides were dialyzed against Tev-buffer overnight. 1 mg of Tev protease was added to each 50 mg of the protein and the mixture was incubated overnight at 4°C. The mixture was then transferred into a centricon with cut-off 3 KDa, therefore, the cleaved peptide that has a size of 2.5 KDa would pass to the flow through, while the Gb1 fusion protein was retained on the membrane. The flow through was then dialyzed against water and froze dried. Eventually, for NMR measurements, the dried sample was dissolved in 450 µl of PBS buffer and 50 µl of D₂O.

3.2.4. NMR experiments to measure the binding between the meta-structure predicted ligands and the human armadillo repeat region of β -Catenin:

Two ligands have been taken at the beginning of the project as lead compounds to be investigated. One compound was predicted by meta-structure to be binding with the armadillo repeat core region of human β -Catenin; fluorescein sodium (**13**). The predicted ligand was hypothesized based on similarity of the met-structure parameters between catenin and human serum albumin.

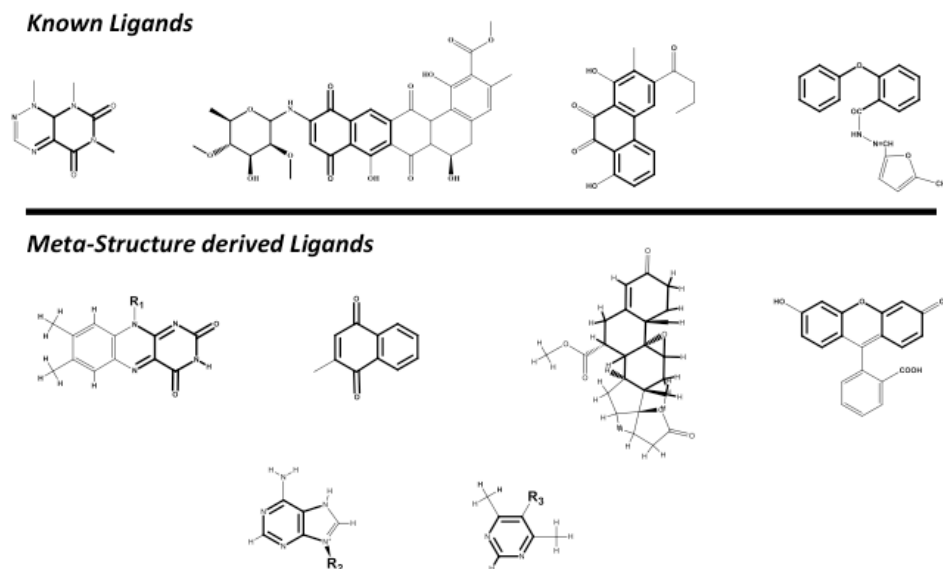
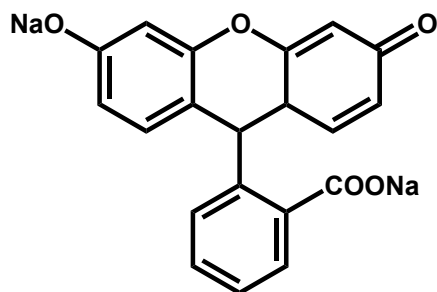


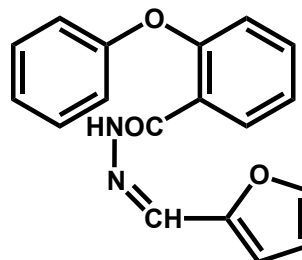
FIG. 47 Similarity between beta-catenin known ligands and ligands derived from meta-structure calculations. The similar structures are highlighted in black. (Figure is adapted from Henen *et al.*⁹⁷)

In addition, *N*-(furan-2-ylmethylene)-2-phenoxybenzohydrazide (**14**), which was reported as a potential ligand for catenin based on NMR and ITC studies⁸³, was also considered.

Experimental

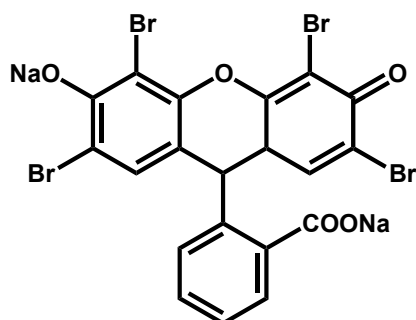


13

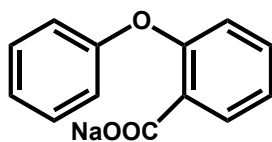


14

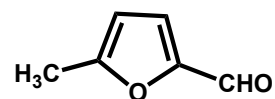
In order to investigate the binding of different ligands with β -Catenin, fluorescein sodium (**13**) and a fluorescein derivative; Eosin Y sodium (**15**) have been tested for binding using 1D ^1H -saturation transfer difference NMR (STD) experiment. Moreover, we considered testing the binding of compound (**14**) as two separate parts; sodium 2-phenoxybenzoate (**16**) and 5-methylfurfural (**17**) using the same NMR experiment.



15



16



17

3.2.4.1. ^1H -Saturation Transfer Difference NMR (^1H -STD):

All spectra were acquired at 25 °C on a Varian INOVA 500 MHz spectrometer. The ligands were purchased from Sigma/Aldrich Company. Data were collected from a ligand and the expressed β -Catenin mixture in a final concentration of 1 mM ligand and 30 μM protein. The ligands and the protein mixtures were all in Tris-buffer at pH=7.4 with 10% D_2O to lock the field. For each ligand 1D ^1H -NMR experiment was recorded prior to The STD experiment. Moreover, negative control STD experiments were performed on 1mM solutions of the ligands in absence of the protein. Selective pre-saturation of the protein protons was achieved by a train of 50–100 Gauss shaped pulses of 50 ms with a strength of $(c/2p)B_1 = 26$ Hz, the protein resonances were removed from the spectra using $T1\rho$ filter. The on resonance irradiation of the protein was done at -1.0 ppm. Off-resonance irradiation was applied at 100 ppm, where no protein signals could be detected. The usual relaxation delay was set to 2 s. To eliminate protein resonances from the spectrum, a spin-lock filter ($T1\rho$ -filter) was used. The spectra were processed by multiplication with an exponential line broadening function of 0.3–1 Hz prior to Fourier transformation. The data were processed and analyzed with NMRPipe software⁹⁸. In addition, STD was measured for 0.5 mM of compound **16** and with saturation time 1s to calculate STD-AF as a parameter of binding. The STD-AF was calculated from the following equation⁹⁹:

$$\text{STD - AF} = \varepsilon(I_0 - I_{\text{sat}})/I_0$$

Where ε is the ratio between ligand and protein concentration, $I_0 - I_{\text{sat}}$ is the normalized signal intensity of the ligands' protons after protein saturation, and I_0 the signal intensity of the ligand in the reference 1D experiment.

Experimental

Table 1: Recording parameters of the performed 1D ¹H-STD experiments

Spectral width [Hz]	6000.2
Number of points	2996
Number of scans	128

1D-¹H NMR spectra for 1 mM of compounds **13**, **15**, **16** and **17** were initially recorded and they showed the following spectral data (table 2).

Table 2: 1D ¹H-NMR data of compounds 13, 15, 16 and 17.

Compound Number and concentration (w/v)	¹H-NMR (Tris-buffer/10% D₂O)
13 0.037 g/ 100 ml	6.57 (m, 4H, H ₁ & H ₂ , & H ₇ & H ₈), 6.72 (m, 2H, H ₄ & H ₅), 7.26 (d, 1H, H _{3'}), 7.71 (dt, 1H, H _{5'}), 7.79 (dt, 1H, H _{4'}), 7.99 (d, 1H, H _{6'}).
15 0.068 g/ 100 ml	7.25 (d, 1H, H _{3'}), 7.48 (s, 2H, H ₁ & H ₈), 7.65 (m, 2H, H _{4'} & H _{5'}), 7.78 (d, 1H, H _{6'}).
16 0.0236 g/ 100 ml	6.75 (d, 3H, H _{2'} , H _{4'} , H _{6'}), 6.86 (d, 2H, H _{3'} , H _{5'}), 6.99 (t, 1H, H ₆), 7.08 (t, 1H, H ₄), 7.21 (m, 1H, H ₅), 7.45 (d, 1H, H ₃)
17 0.011 g/ 100 ml	1.8 (s, 3H, CH ₃), 2.95 (s, 2H, CH ₂), 6.35 (s, 1H, H ₃), 7.45 (s, 1H, H ₂)

Results & Discussion

1D ^1H -STD NMR was recorded later for each ligand and protein mixture. Compounds **13**, **15**, and **16** showed positive STDs as the resonances of these compounds are observed while compound **17** showed zero STD. Since magnetization transfer in STD experiment moves from the protein to the small molecule ligand and is detected as signal enhancement of the ligand after its dissociation from the protein, this observation strongly support that these compounds which showed positive STD are binding to the β -Catenin. The negative control experiments all showed zero STD as expected. (Results are shown in figure 48). Moreover, sodium 2-phenoxybenzoate (**16**) showed STD-AF value of 0.63.

Experimental

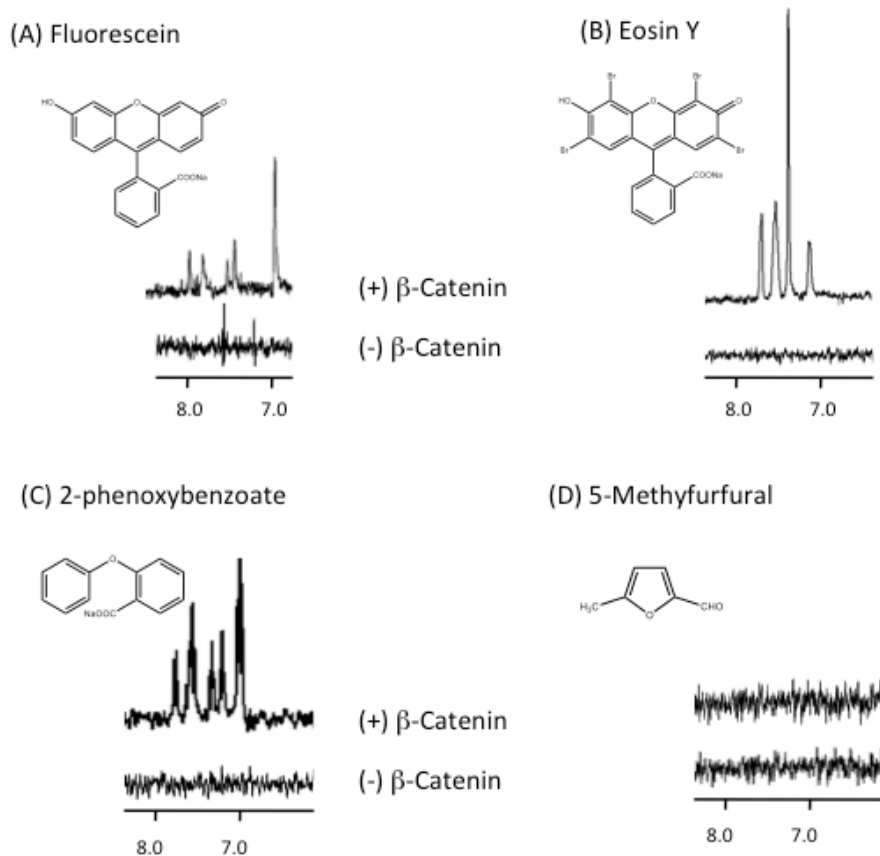
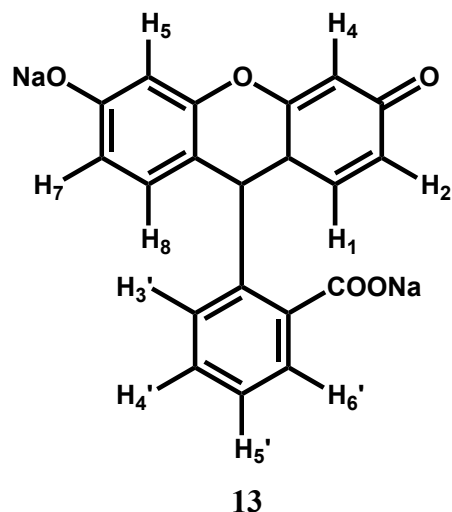


FIG. 48 1D ^1H STD spectra of 30 μM solution of β -Catenin with 1 mM aqueous solution of different ligands. The corresponding negative control in absence of the protein is shown below each spectrum. A) Fluorescein sodium. B) Eosin Y. C) Sodium 2-phenoxybenzoate. D) 5-Methylfurfural. (Ligands A and B figures are adopted from Henen *et al.*⁹⁷).

3.2.4.2. *Cross-Relaxation during Adiabatic Fast Passage:*

In this experiment, AFP-NOESY measures the homonuclear (^1H - ^1H) cross relaxation rates (NOEs and ROEs) during adiabatic fast passage (AFP). This involved adiabatic fast passage radio frequency (RF) pulses with a parabolic phase modulation leading to a linear frequency sweep through a considerably large spectral window. The experiment was done to a sample of fluorescein sodium (**13**) without protein and to a mixture of 1 mM of fluorescein sodium (**13**) and 30 μM of β -Catenin in tris-buffer/10% D_2O at different AFP spin lock power. Initially, to invert protons number H_1 , H_2 , H_7 and H_8 , a selective 180° IBURP1 pulse centered at 6.5 ppm with inversion bandwidth 0.34 ppm was applied. On the other hand, to invert protons number H_4 , H_5 , and H_6 , a selective 180° IBURP pulse centered at 7.61 ppm with inversion bandwidth 0.34 ppm was applied (used IBURP pulses had a length of 9.58 ms and power of 24 db). Afterwards, different strength AFP spin lock fields were applied 5 kHz downfield of the carrier frequency for 400 ms and swept over 10 kHz in the upfield direction with 10% apodization “at the beginning (sinusoidal) and at the end (cosinusoidal)”. Eventually, a read-out 90° pulse was applied.

The NOE between different protons in compound **13** was measured as a function of AFP spin lock power ($\sin^2\theta$). The different strength of the applied spin lock field gave rise to the following $\sin^2\theta$: 0.0607, 0.0760, 0.0948, 0.1180, 0.146, 0.1808, 0.2221, 0.2710, 0.3277, 0.3919, 0.4623, 0.5365, and 0.6110.



Results & Discussion

In the first set of experiments, peaks of protons H₁, H₂, H₇ and H₈ were inverted and the NOE effect on peaks of protons H₄ and H₅ was observed. The data shown in figure 49A clearly demonstrated that the behavior of compound **13** is now a behavior of large molecule and these protons are binding to the protein present in the solution.

In the second set of experiments, peaks of protons H_{4'}, H_{5'} and H_{6'} were inverted, and the NOE effect on peaks of protons H₁, H₂, H₇ and H₈ was observed. The data shown in figure 49B demonstrated the behavior of large molecule as the NOE and ROE differ in their sign (NOE enhancement is -100% and ROE enhancement is +200%) leading to dependence of the cross relaxation rate is on the strength of the applied AFP spin lock field. It can be concluded from this that these protons bind to the armadillo repeat region of human β -Catenin.

It can be noticed from the results here, the change of the peak intensity as a function of the applied AFP with no sign change and consequently absence of zero crossing. The absence of zero crossing is due to spin diffusion which

Experimental

indicates that the whole molecule of compound **13** is deeply embedded in the binding pocket of the protein and being surrounded by hydrophobic patch.

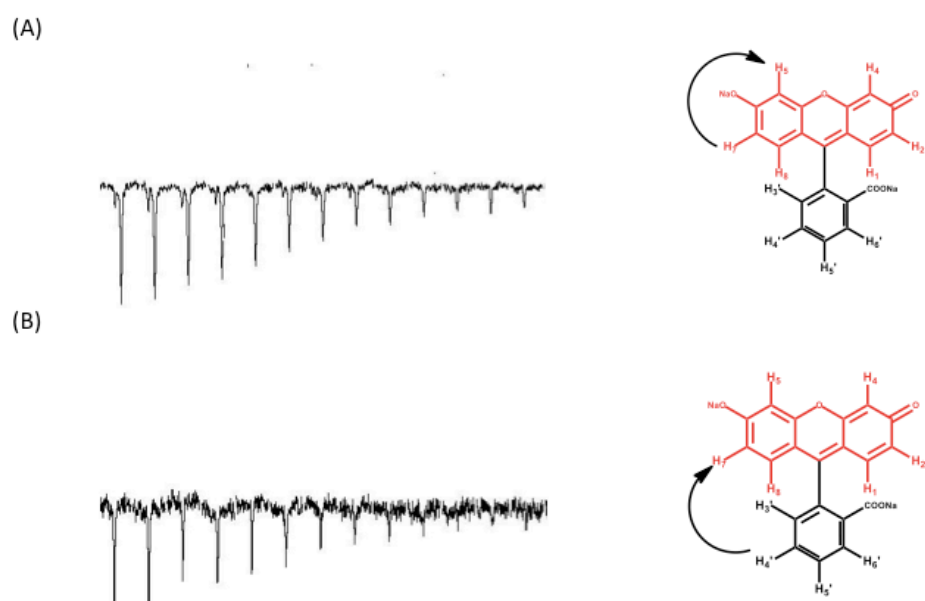


FIG. 49 Selective AFP-NOESY traces of the protein complex formed between Fluorescein sodium and human β -Catenin. (A) The NOE between H₁, H₂, H₇, H₈ and H₄, H₅ of Fluorescein was measured as a function of AFP spin lock power. (B) The NOE between H_{4'}, H_{5'}, H_{6'} and H₁, H₂, H₇, H₈ of Fluorescein was measured as a function of AFP spin lock power. (Figure is adapted from Henen *et al.*⁹⁷).

3.2.5. Isothermal calorimetry (ITC):

The binding between fluorescein sodium (**13**) with catenin was measured using ITC to detect the thermodynamics of binding. Different concentrations of the protein and the ligand have been tried to achieve saturation and considerable data. The initial concentrations were 20 μM β -Catenin : 4 mM of compound **13**. The measurements were done at 25 °C using μ -ITC calorimeter (*Microcal*). Samples of the protein and ligand were all dissolved in the same batch of tris-buffer to avoid biased results. The ligand was injected from a syringe into the cell containing the protein through 20 successive injections. To get rid of the heat dilution effect, a blank experiment was done in which the ligand was injected into a cell containing the buffer alone and the result was subtracted from the original results. Thermodynamic parameters were estimated by non-linear least square method using procedures incorporated in the “origin” software package (*Microcal*).

Results & Discussion

The isotherm showed that the fluorescein sodium is binding to the protein in binding ratio (stoichiometry N) 16:1. The dissociation constant (K_D) of the binding between this ligand and the protein is 79.3 μM . The enthalpy of the reaction (ΔH) is -3694 cal/mol indicating an exothermic reaction, and the entropy of the reaction is 6.25 cal/mol/deg. The raw data and the fit data are shown in figure 50(A &B).

Experimental

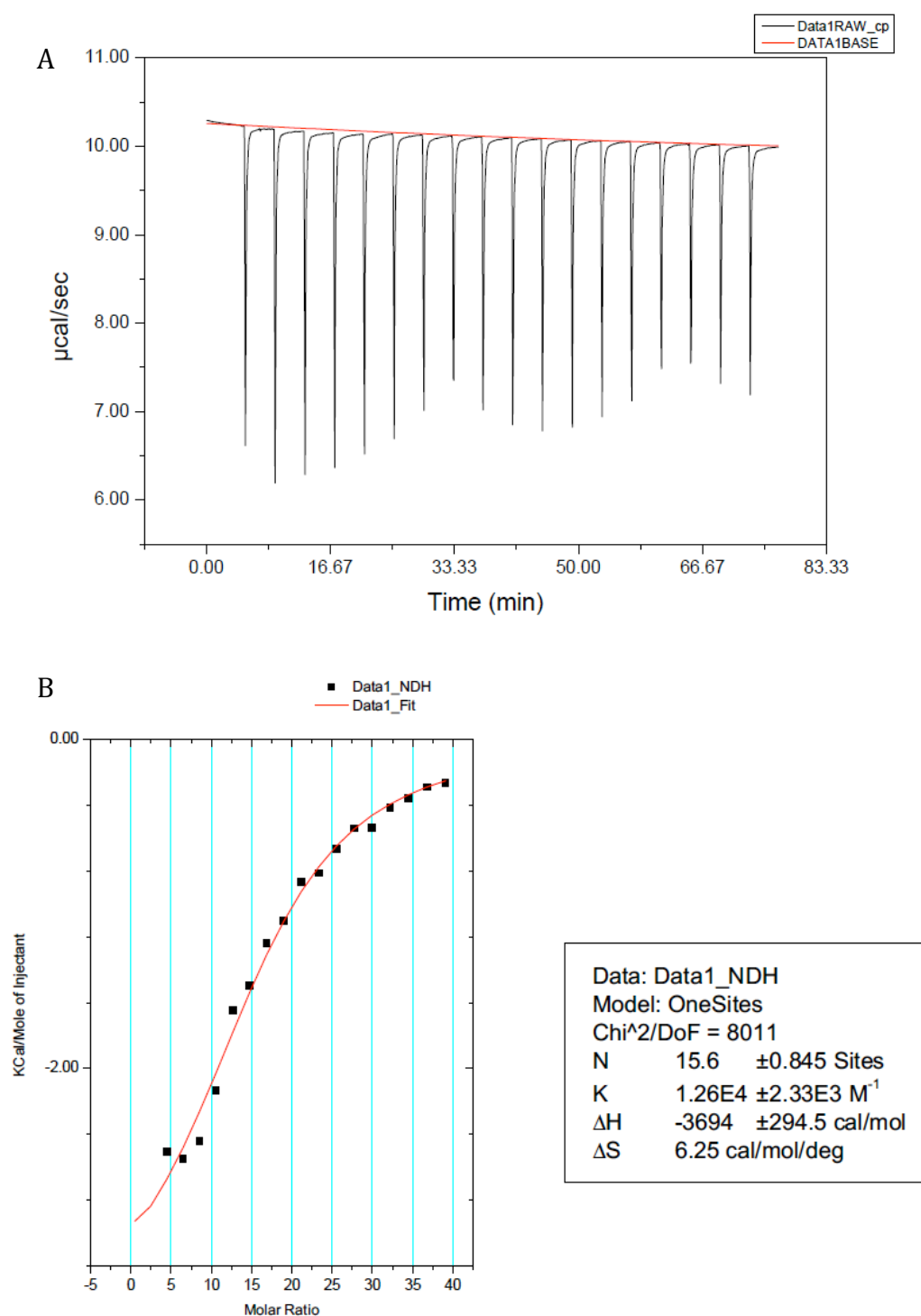
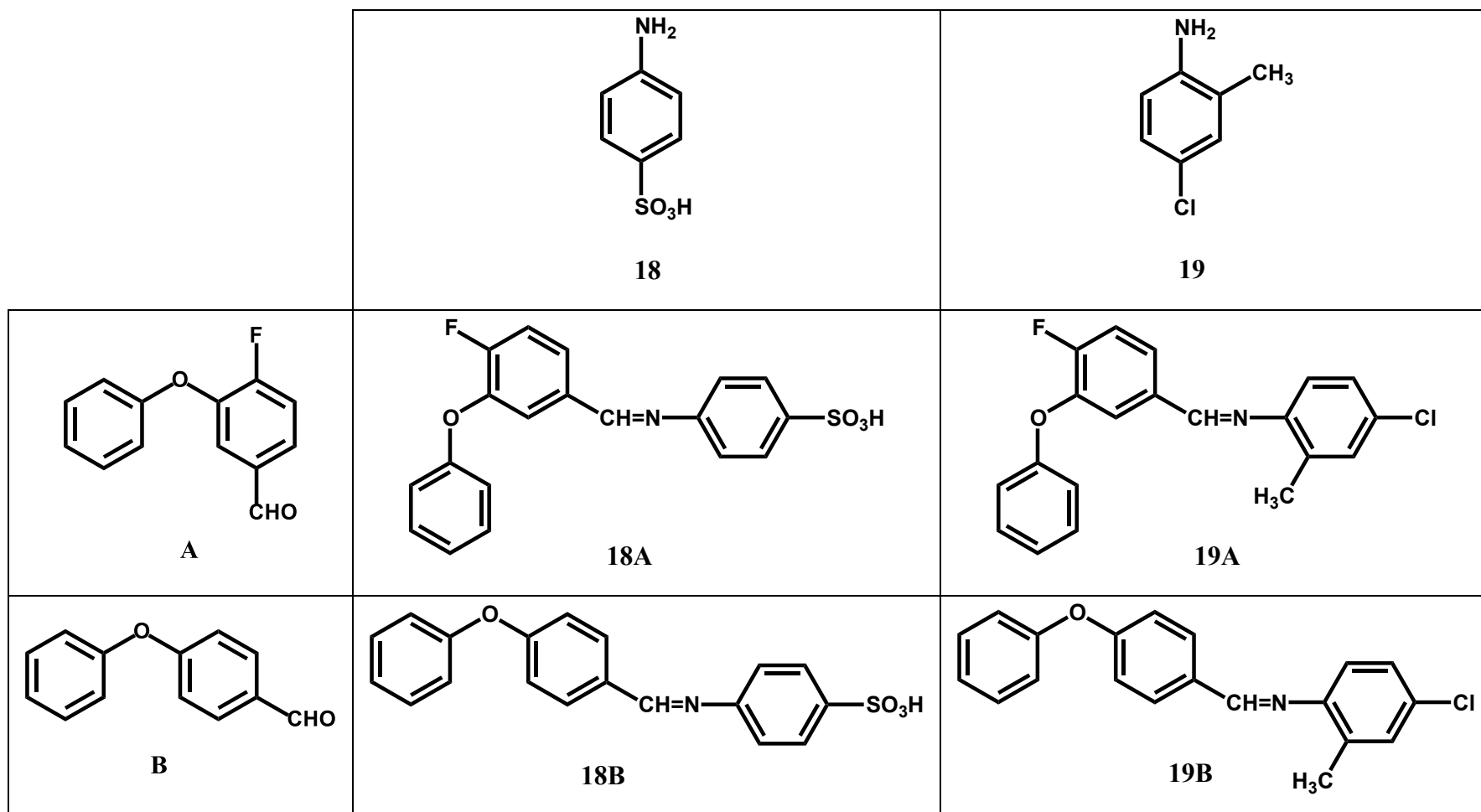


FIG. 50 The raw data of the calorimetric measurement between 20 μM β -Catenin and 4 mM fluorescein sodium (A), and the processed data showing the stoichiometry, binding constant, enthalpy and entropy (B).

Experimental

3.2.6. Dynamic Combinatorial Chemistry:

Table 3. Virtual dynamic imine system formation



Experimental

To probe imine formation reaction for its performance in dynamic chemistry, a prototype system was designed and tested for binding to the human armadillo repeat region of β -Catenin. 1D ^1H -STD NMR experiments were subsequently adopted for binding studies of the library with β -Catenin.

Thus the nucleophilic addition reaction between two aldehydes; 4-fluoro-3-phenoxybenzaldehyde (**A**) & 4-phenoxybenzaldehyde (**B**), and two amines; 4-aminobenzenesulfonic acid (**18**) & 4-chloro-2-methylaniline (**19**), resulted in spontaneous synthesis of 4 novel different imines in equilibrium with the initial building blocks. Each single component in the library was tested separately for binding with the targeted protein. Then each single aldehyde with the two amines was tested. Eventually the whole (library) was tested. The components of the library have been chosen based on their similarity with the two known lead compounds; **13** & **14**.

The reactions as well as the measurements were done in the NMR tube at final concentration 1 mM of each separate ligand and 30 μM of protein. All the ligands were dissolved in deuterated DMSO in stock of 40 mM and the added to protein solution to have a sample with final volume 500 μl [ligand(s) concentrations 1mM and protein concentration 30 μM at pH=7.4]. All spectra were acquired at 25 $^\circ\text{C}$ on a Varian INOVA 800 MHz spectrometer.

Results & Discussion

a- Testing of each single compound (18, 19, A and B).

A complex of 1 mM 4-chloro-2-methylaniline (**19**) and 30 μM β -Catenin showed zero STD indicating that this amine doesn't bind to the targeted protein (fig. 51)

Experimental

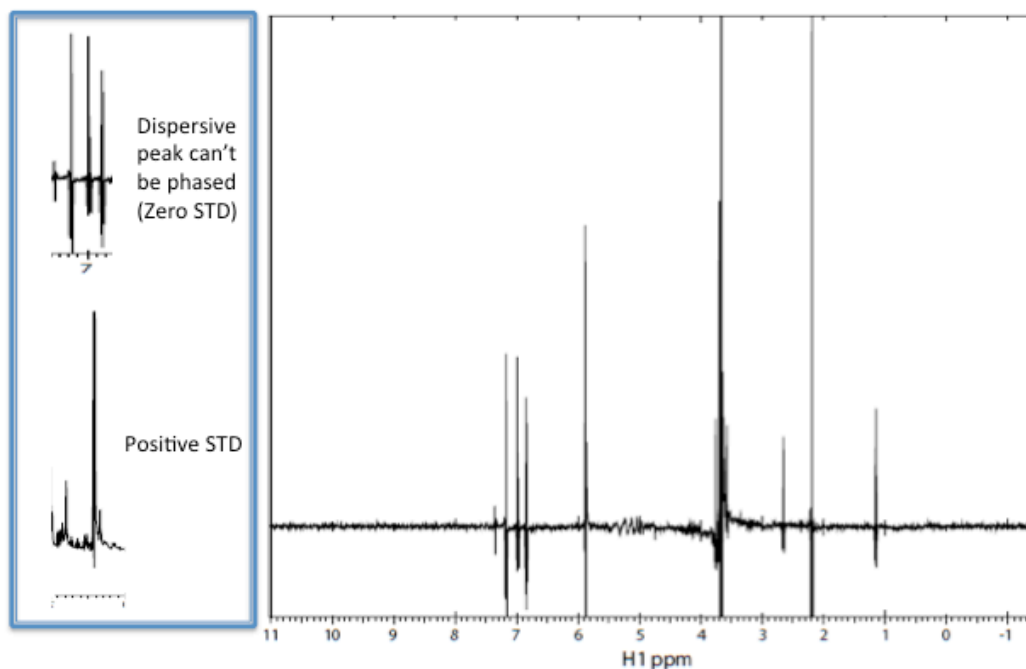


FIG. 51 1D ^1H -STD NMR of the protein complex formed between compound **19** and human β -Catenin. Zero STD was obtained (the peaks observed are dispersive peaks can't be phased and result from subtraction artifact). In the blue box, illustration of the difference between positive STD and artifact zero-STD.

Moreover, a complex of 1mM 4-aminobenzenesulfonic acid (**18**) and 30 μM β -Catenin showed zero STD indicating that this amine didn't bind to the targeted protein.

In addition, a complex of 1mM of 4-fluoro-3-phenoxybenzaldehyde (**A**) and 30 μM targeted protein showed positive STD indicating that this aldehyde binds to β -Catenin.

Eventually, a complex of 1mM of 4-phenoxybenzaldehyde (**B**) and 30 μM targeted protein showed positive STD indicating that this aldehyde binds to β -Catenin.

Experimental

b- Testing of each aldehyde (A or B) with the 2 amines (18 and 19).

Based on the data that showed that both aldehydes (A and B) are binding to the targeted protein, while both amines showed no binding to the targeted protein, it was noteworthy to test the binding of each aldehyde in the presence of both amines. This leads to the formation of two novel imine compounds in each case. 1D ¹H-STD was done to measure the binding and to see the equilibrium between the free aldehyde (CHO) peak and the newly formed imine peak (N=CH).

- *4-Fluoro-3-phenoxybenzaldehyde (A) with amines 18 & 19.*

To 25 μ l 20 mM compound **A**, 12.5 μ l of 40 mM compound **18** and 12.5 μ l of 40 mM compound **19** were added to have a final volume of 50 μ l in the NMR tube. The mixture was kept at room temperature for 20 minutes. To this prepared mixture, 450 μ l of 33 μ M the expressed human β -Catenin in tris-buffer/10%D₂O was added to have a final concentration 30 μ M protein and 1 mM of each ligand. 1D ¹H-NMR was measured for this mixture. It was observed the reduction in the aldehyde peak intensity at 9.82 ppm and the appearance of new imines peak at 9.20 ppm. This proved the synthesis of two new imines **18A** and **19A** in equilibrium with the starting blocks (fig. 54). In addition, STD-AF value was calculated for 0.5 mM of compound **19A** to compare it with the value obtained for compound **16**. Compound **19A** showed STD-AF value of **3.64** (i.e. STD-AF for compound **16** was **0.63**)⁹⁷. This can be justified by better binding of the newly formed imine **19A** in comparison to the small fragment **16**.

Experimental

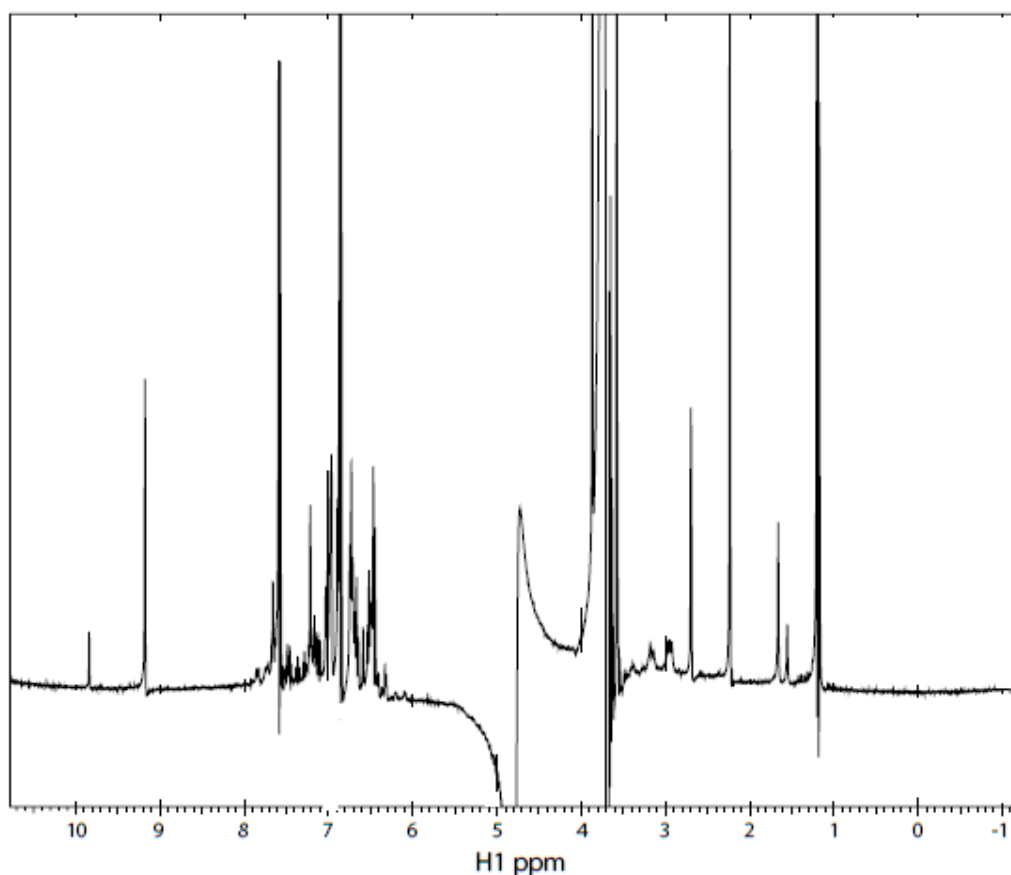


FIG. 52 1D ¹H-NMR of the reaction between aldehyde A and the 2 amines 18& 19 in presence of β -Catenin. The newly synthesized compounds 18A and 19A showed a new imine peak at 9.20 ppm.

The formed protein complex was subjected to 1D ¹H-STD NMR to see the binding between the newly formed imines and the targeted protein. It was found that positive STDs were recorded for the imine.

Experimental

- *4-phenoxybenzaldehyde (B) with amines 18 & 19.*

To 25 μ l 20mM compound **B** in deuterated DMSO, 12.5 μ l of 40 mM compound **18** in deuterated DMSO and 12.5 μ l of 40 mM compound **19** in deuterated DMSO were added to have a final volume of 50 μ l in the NMR tube. The mixture was allowed to stand at room temperature for 20 minutes. To this prepared mixture, 450 μ l of 33 μ M the synthesized human β -Catenin in tris-buffer/10%D₂O was added to have a final concentration 30 μ M protein and 1 mM of each ligand. 1D ¹H-NMR was recorded for this mixture. Watching the peaks at 9.82 & 9.29 that corresponding to the aldehyde and the imine functional groups respectively monitored the equilibrium between aldehyde and the formed imine (fig. 53).

The formed protein-ligand complex was subjected to 1D ¹H-STD NMR to see the binding between the newly formed imines and the target protein. It was found that positive STDs were recorded for the imines.

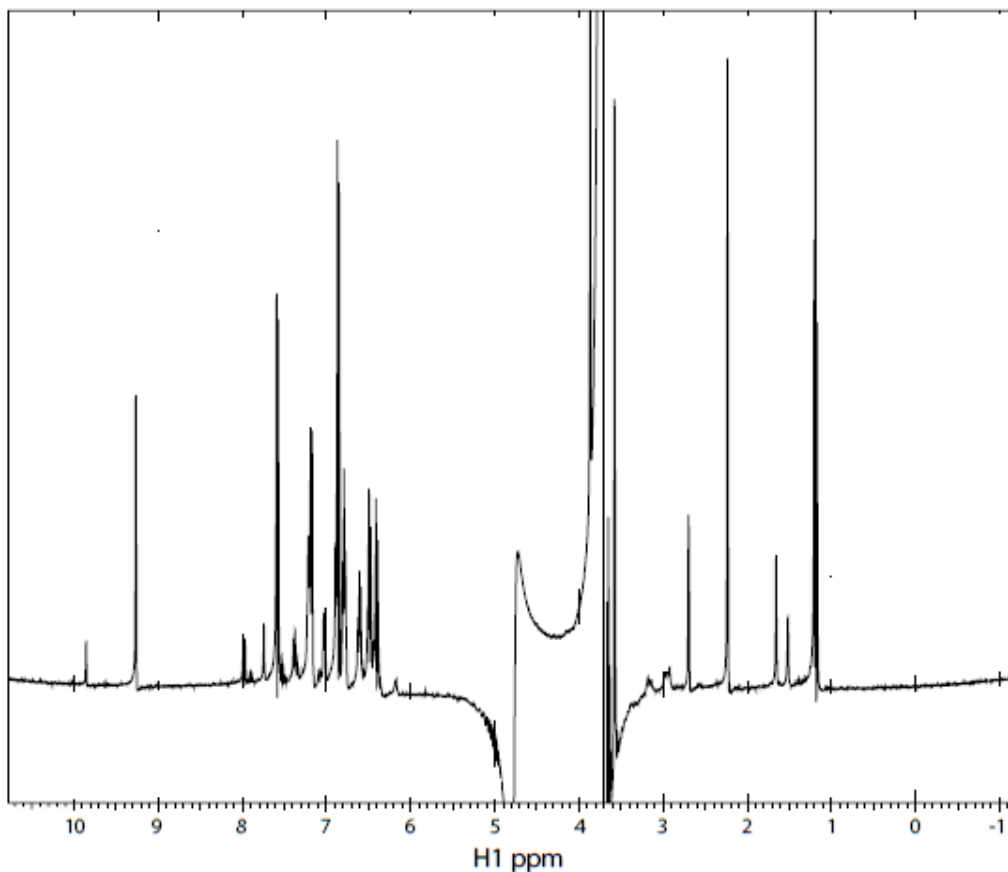


FIG. 53 1D ^1H -NMR of the reaction between aldehyde **B** and the 2 amines **18** & **19** in presence of β -Catenin. The newly synthesized compounds **18B** and **19B** showed a new imine peak at 9.29 ppm.

*c- Testing of the “whole library” (aldehydes **A** and **B**) with the 2 amines (**18** and **19**).*

It was noteworthy to test how the whole library would behave when we have the 2 aldehydes (**A** and **B**) with the 2 amines (**18** and **19**) mixed with the target protein.

Experimental

A mixture of deuterated DMSO solutions of 12.5 μl 40 mM compound **A**, 12.5 μl of 40 mM compound **B**, 12.5 μl of 40 mM compound **18** and 12.5 μl of 40 mM compound **19** was prepared to have a final volume of 50 μl in the NMR tube. The mixture was allowed to stand at room temperature for 20 minutes. To this prepared mixture, 450 μl of 33 μM the human β -Catenin in tris-buffer/10%D₂O was added to have a final concentration 30 μM protein and 1 mM of each ligand. 1D ¹H-NMR was recorded for this mixture. It was observed the reduction in both aldehyde peaks intensity at 9.82 and 9.84 ppm and the appearance of two new imines peaks corresponding to compounds **18A**, **18B**, **19A** and **19B** at 9.20 and 9.29 ppm (fig. 54). The newly formed imines showed positive STD indicating that the 4 imines are binding to catenin (fig. 55)

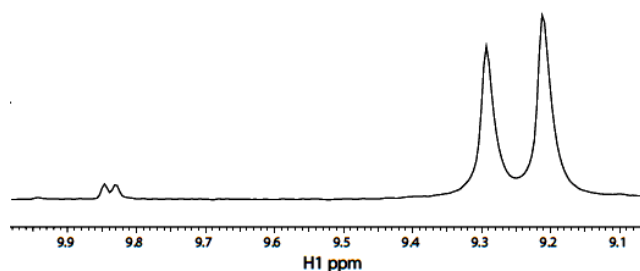


FIG. 54 zoomed 1D ¹H-NMR of the reaction between aldehydes A and B and the 2 amines 18 & 19 in presence of β -Catenin. The newly synthesized compounds 18A,18B, 19A and 19B showed new imine peaks at 9.20 and 9.28 ppm. A reduction in aldehyde peaks at 9.82 and 9.84 was observed.

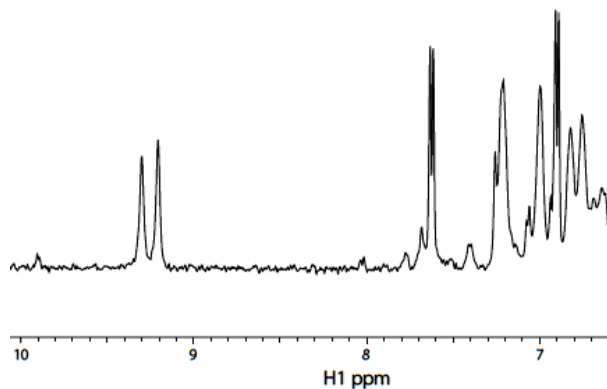


FIG. 55 ^1H - STD NMR of the 4 newly formed imines in presence of β -Catenin showing positive STD.

d- ^{19}F -NMR for compound 19A

Compound **19A** was tested for binding with armadillo repeat region of β -Catenin using ^{19}F -NMR. Firstly, compound **19A** (mixture of compound **19** and compound **A**) in 10% DMSO and Tris-buffer was measured using ^{19}F -NMR. It showed 3 different peaks at -129.06, -125.9 and -121.9 (corresponding to the *E*- and *Z*- conformers and the free aldehyde). The experiment was measured again after addition of 30 μM of catenin. The peak at -129.06 got broadened while the peak at -125.9 got less affected. Broadening of the ^{19}F peaks is indication of binding between compound **19A** and catenin.

Experimental

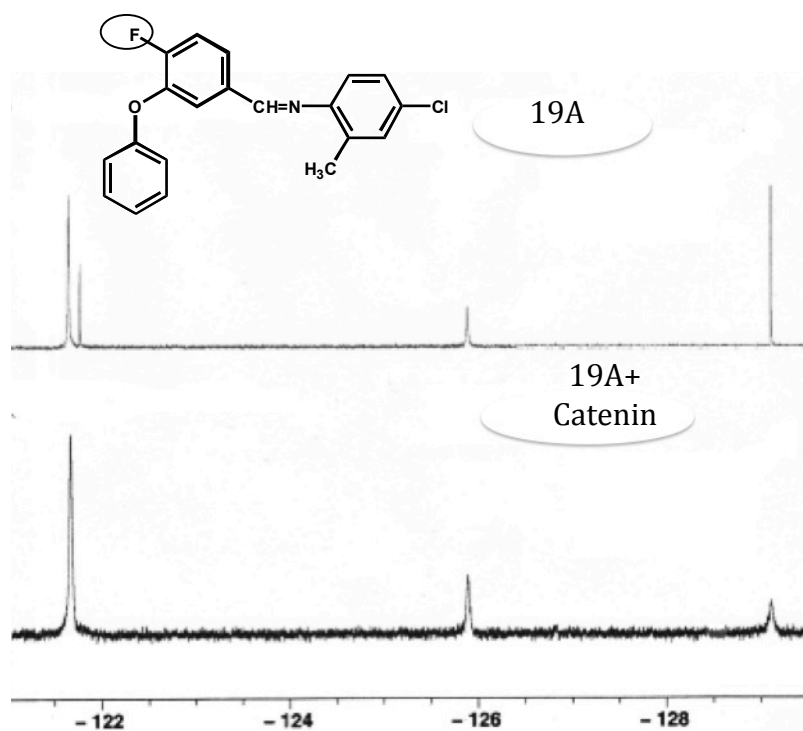


FIG. 56 ^{19}F -NMR for compound **19A** alone and compound **19A** with the targeted protein.

The experiment was repeated for a mixture of **19A**, catenin and $60\ \mu\text{M}$ TCF4 and for a mixture of **19A**, catenin and 1mM sodium 2-phenoxybenzoate (**16**). It was found that in presence of TCF4, the peaks are still broadened indicating that TCF4 doesn't substitute **19A** from the binding site. On the other hand, in presence of sodium 2-phenoxybenzoate (**16**), the peak at -129.06 got sharper indicating a competition between **19A** and **16** on the binding site of β -Catenin.

Experimental

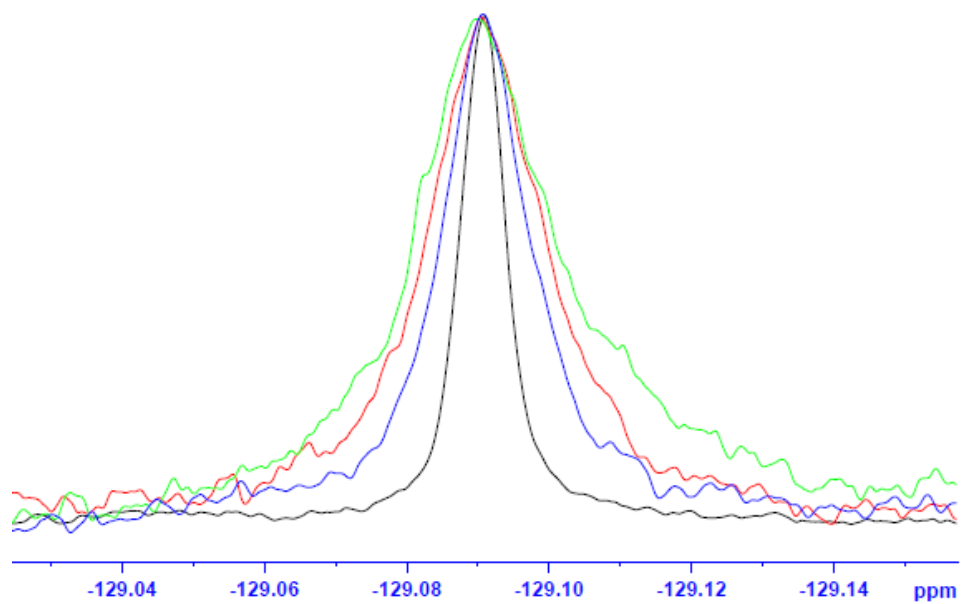


FIG. 57 ^{19}F -NMR peak at -129.06 ppm. Black: 19A alone in tris-buffer. Red: 19A + catenin. Green: 19A +catenin+TCF4. Blue: 19A+catenin+ sodium 2-phenoxybenzoate.

Experimental

Structure Summary:

From the previous data, it was shown that both the free aldehydes (**A** and **B**) bind to the human armadillo repeat region of β -Catenin, while the free amines (18 and 19) didn't show any binding to the target protein. From these data, and from what was previously shown in 1D AFP-NOESY of observed NOEs between the protons of the xanthen ring in fluorescein, it can be concluded that OXYDIBENZENE moiety is a potential pharmacophore for β -Catenin ligands and when it's present in the compound, the whole compound is embedded in the core region of β -Catenin.

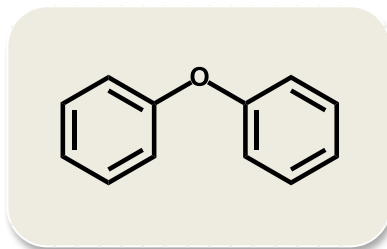


FIG. 58 Oxydibenzene moiety was shown to be a potential pharmacophore in β -Catenin binders.

3.2.7. Synthesis of selectively ^{13}C -labelled reductively methylated β -Catenin

Reductive methylation was performed similar to what was described by Means and Feeney. The protein was dialyzed into 10 mM HEPES buffer at pH 7.4. 0.25 ml of 3.2 M borane ammonium complex ($\text{NH}_3\text{-BH}_3$) and 0.5 ml of 3.2 M ^{13}C -labelled formaldehyde were added to 0.5 ml of 40 μM protein and the reaction was incubated under stirring at 4°C. The addition of borane ammonium complex and ^{13}C -formaldehyde was repeated and the reaction was incubated for more 2 h. After addition of 0.12 ml of 3.2 M borane ammonium complex, the reaction mixture was incubated at 4°C with stirring overnight. Adding glycine to 200 mM concentration ended the reaction. Dialysis against tri-buffer at pH 7.4 was performed to get rid of excess reagents and unwanted products. The sample was concentrated to 40 μM . Three sets of 2D HSQC NMR experiments were done on the ^{13}C -methylated β -Catenin. All NMR experiments were done on Varian Inova 600 MHz spectrometer.

Experimental

Table 4: Recording parameters of the performed 2D ^1H - ^{13}C HSQC experiments on the methylated lysine β -Catenin.

Spectral width [Hz] $F1 \times F2$	5000.0 x 7022
Number of points $F2$ (^1H)	1024
Number of increments $F1$ (^{13}C)	115
Number of scans	24
Decoupler offset Hz (dof)	-10870

Results & Discussion

a- 2D ^{13}C - ^1H HSQC for the free methylated lysine β -Catenin:

For 40 μM of the free ^{13}C -methylated β -Catenin in tris-buffer/10% D_2O , ^{13}C - ^1H HSQC was measured (fig. 59A). Lysine side chains that are commonly found on the protein exterior experience unobstructed rotation. Hence, all appeared as one intense peak at ^1H , ^{13}C ; (2.83, 40.5) ppm.

The other methylated lysines appeared at ^1H , ^{13}C ; (2.45, 41.9) ppm, (2.75, 52.5) ppm, (2.86, 42.5) ppm, (2.95, 46) ppm and (3.00, 46.5) ppm.

b- 2D ^{13}C - ^1H HSQC for the methylated lysine β -Catenin with 120 μM TCF4 :

To 40 μM of the free ^{13}C -methylated β -Catenin in tris-buffer/10% D_2O , 120 μM of TCF4 was added. ^{13}C - ^1H HSQC was measured (fig. 59B). Lysine side chains that are commonly found on the protein exterior appeared as one intense peak at ^1H , ^{13}C ; (2.91, 40.5) ppm. The other methylated lysines appeared at ^1H , ^{13}C ; (2.77, 45.00) ppm, (2.75, 52.5) ppm, (2.86, 42.5) ppm, (2.95, 46) ppm and (3.00, 46.5) ppm. It was clear that one peak which appeared at ^1H , ^{13}C ; (2.45, 41.9) ppm was shifted to ^1H , ^{13}C ; (2.77, 45.00) ppm when the binding partner

Experimental

TCF4 was added. In addition, peaks at (2.95, 46) ppm and (3.00, 46.5) ppm had less intensity upon binding with TCF4.

c-2D ^{13}C - ^1H HSQC for the methylated lysine β -Catenin with 1 mM sodium 2-phenoxybenzoate (**16**):

From previously done experiments, we proved that fluorescein analogs bind to the core region of the synthesized β -Catenin. It was of great interest to know if these compounds bind to the same binding site(s) as TCF4. In order to do this, 40 μM of the free ^{13}C -methylated β -Catenin in tris-buffer/10% D_2O was added to 1 mM of sodium 2-phenoxybenzoate (**16**). ^{13}C - ^1H HSQC was measured.

Lysine side chains that are commonly found on the protein exterior appeared as one intense peak at ^1H , ^{13}C ; (2.91, 40.5) ppm. The other methylated lysines appeared at ^1H , ^{13}C ; (2.81, 45.00) ppm, (2.75, 52.5) ppm, (2.75, 58.5) ppm, (2.86, 42.5) ppm, (2.95, 46) ppm and (3.00, 46.5) ppm.

It was proved that binding to TCF4 or sodium 2-phenoxybenzoate (**29**) shifted the lysine peak at ^1H , ^{13}C ; (2.45, 41.9) ppm to the same new position. Moreover, binding with compound 2, led to appearance of new peak at ^1H , ^{13}C ; (2.75, 58.5) ppm which could be belonging to one of the exterior lysines was included in binding (fig. 59C)

It can be concluded, that both TCF4 and sodium 2-phenoxybenzoate have similar binding sites on the armadillo repeat region of β -Catenin.

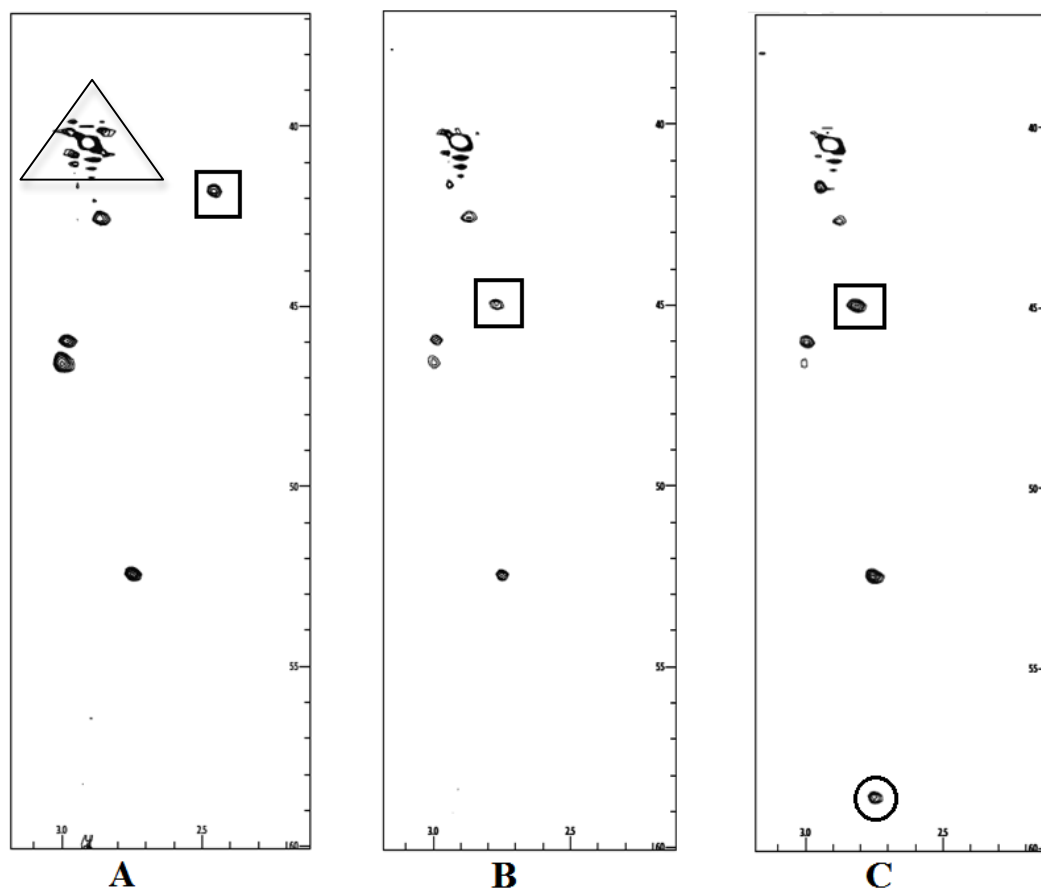


FIG. 59 ^{13}C - ^1H HSQC spectra of (A) Free form of methylated catenin, (B) methylated catenin with TCF4, and (C) methylated catenin with sodium 2-phenoxybenzoate. The peaks of lysines that are on the surface of the protein appeared on top of each other and highlighted with the triangle. The lysine peak that is highlighted with square shifted upon binding to TCF4 and to sodium 2-phenoxybenzoate as well. The peak highlighted with the circle appeared only upon binding with phenoxybenzoate.

3.2.8. Backbone assignment of TCF4 and its truncated versions using 2D and 3D NMR experiments

Data collection for residue assignment of TCF4 were done on a Varian Unity 500 & 600 MHz Spectrometers at 25°C. Data were collected from a ^{15}N - and ^{13}C -labeled TCF4 which was prepared at a final concentration of 1 mM (pH 7.4, 4.5 and 2). Final recording parameters for NMR experiments are depicted in table 5.

Table 5. Recording parameters of the performed assignment NMR experiments.

2D ^1H - ^{15}N -HSQC	
Spectral width [Hz] $F1 \times F2$	5000.0 x 7022
Number of points $F2$ (^1H)	1024
Number of increments $F1$ (^{14}N)	115
Number of scan	24
3D HNCA	
Spectral width [Hz] $F1 \times F2 \times F3$	4500.0 x 1500 x 69999.7
Number of points $F3$ (^1H)	720
Number of increments $F1$ (^{13}C)	80
Number of increments $F2$ (^{14}N)	28
Number of scans	24

Experimental

3D HNCOCA	
Spectral width [Hz] F1 x F2 x F3	4500.0 x 1500.01 x 6999.7
Number of points F3 (¹ H)	558
Number of increments F1(¹³ C)	50
Number of increments F2(¹⁴ N)	36
Number of scans	24
3D HNCACB	
Spectral width [Hz] F1 x F2 x F3	8000.0 x 1800.01 x 6999.7
Number of points F3 (¹ H)	896
Number of increments F1(¹³ C)	90
Number of increments F2(¹⁴ N)	28
Number of scans	24
3D HNCO	
Spectral width [Hz] F1 x F2 x F3	2000.0 x 1500.01 x 6999.7
Number of points F3 (¹ H)	560
Number of increments F1(¹³ C)	50
Number of increments F2(¹⁴ N)	36
Number of scans	20

Experimental

3D HNCACO	
Spectral width [Hz] F1 x F2 x F3	2000.0 x 1500.01 x 6999.7
Number of points F3 (^1H)	562
Number of increments F1(^{13}C)	50
Number of increments F2(^{14}N)	28
Number of scans	32

For assignment of TCF4, the spectra measured at pH=7.4 were so ambiguous, so we moved down with the pH value to 4.5 which gave the best spectral data. In addition, the two truncated versions of TCF4; *N*-terminal and *C*-terminal peptides were also assigned at the same pH and the values obtained were compared with those of the full length. Moreover, a comparison between the chemical shift values of the assigned residues and the predicted ones of random coil was done to extract the possible secondary structure of TCF4 in the unbound form.

Table 6 shows the assigned chemical shift values for the full length TCF4 and the peptides.

Experimental

Table 6. Backbone assignment of ^{15}N - ^{13}C labeled wild type-TCF4 and its mutants

<i>Full Res #</i>	<i>res</i>	<i>H</i>	<i>N</i>	<i>Cα</i>	<i>Cβ</i>	<i>Pept Res #</i>	<i>H</i>	<i>N</i>	<i>Cα</i>	<i>Cβ</i>
4	Leu	8.56	124.6 7	54.94	41.91					
5	Asn	8.63	119.7 1	52.96	38.44					
6	Gly	8.44	109.1 2	44.96	-					
9	Gly	8.46	109.6 6	44.95	-					
10	Asp	8.34	120.3	53.9	40.65					
11	Asp	-	-	-	-	11	8.36	121.4 8	54.16	40.97
12	Leu	8.28	122.9 7	55.08	41.58	12	8.26	123.5 6	55.48	41.78
13	Gly	8.48	109.4 7	45.04	-	13	8.48	109.5	45.09	-
14	Ala	8.13	123.5 6	52.26	18.92	14	8.09	123.5 2	52.36	18.2
15	Asn	8.19	118.9 2	53.85	40.52	15				
16	Asp	8.29	120.9 2	54.18	41.56	16	8.27	120.8 7	54.86	40.95
17	Glu	8.32	120.7	56.31	29.5	17	8.33	121.0 1	56.47	28.91
18	Leu	8.19	122.7 6	54.92	41.66	18	8.2	122.7 6	55.25	41.09
19	Ile	8.04	122.0 8	60.6	38.21	19	8.02	122.0 4	61.07	38.91
20	Ser	-	-	-	-	20	8.42	117.8 3	-	-
21	Phe	8.29	122.7	55.23	41.52	21	8.26	123.0 1	57.78	41.85
22	Lys	8.63	-	52.96	38.4	22	8.16	123.5 7	56.15	-

Experimental

23	Asp	8.35	122.3	53.97	40.65	23	8.37	122.6 3	54.19	41.21
24	Glu	8.73	122.1 1	56.78	29.3	24	8.48	122.1 4	56.97	30.01
25	Gly	8.57	109.7 9	44.91	-	25				
26	Glu	8.26	120.6 5	56.18	29.6	26	8.24	120.6 2	56.35	29.06
27	Gln	8.48	121.3 1	55.51	29.04	27	8.47	121.5 4	55.31	29.63
28	Glu	8.45	122.4 9	56.37	29.49	28	8.5	123.3	56.72	30.33
29	Glu	8.43	123.0 9	55.96	32.53					
30	Lys									
31	Ser									
32	Ser									
33	Glu	8.52	122.7 7	56.14	29.54					
34	Asn	8.43	119.3 8	52.92	38.5					
35	Ser	8.34	116.5 4	58.31	63.14					
36	Ser	8.42	117.9 9	58.08	63.18					
37	Ala	8.27	125.7 5	52.45	18.8	37	8.25	125.7	52.41	-
38	Glu	8.29	119.5 9	56.45	29.33	38	8.26	119.5 5	56.51	-
39	Arg	8.19	121.4 2	55.98	30.51	39	8.15	121.2 8	55.94	-
40	Asp	8.39	121.4 1	54.12	40.4	40	8.35	121.2 8	54.15	-
41	Leu	8.15	122.8 3	55.08	-	41	8.12	122.9 1	55	-
42	Ala	8.2	123.4 9	52.49	18.76	42	8.19	123.5 1	52.49	-
43	Asp	8.18	119.2 9	53.89	40.45	43	8.14	119.0 7	53.92	-

Experimental

44	Val	7.99	120.2 4	62.09	31.92	44	7.94	120.3 7	62.31	-
45	Lys	8.39	124.0 7	56.28	32.2	45	8.35	124.2 4	56.11	-
46	Ser	8.32	116.1 7	58.38	62.98	46				
47	Ser					47				
48	Leu	8.18	123.6 9	55.19	41.73	48	8.16	123.5 4	54.9	-
49	Val	8	120.4 4	62.05	32.17	49	7.97	120.1 8	61.76	-
50	Asn					50	8.48	122.8 4	52.79	-
51	Glu	7.84	127.0 8	58.74	-	51	7.91	126.4 6	57.64	30.84
52	Ser									
53	Glu									
54	Thr	8.06	113.4 8	62.36	69.01					
55	Asn	8.05	113.5 2	-	-					
56	Lys									
57	Asn									
58	Trp	8	119.0 9	57.04	29.22					

Experimental

umber. It can be clearly seen from this, that there's no significant difference between both which proves that TCF4 in the unbound form act as a random coil and lacks any definite secondary structure. The predicted chemical shift values for TCF4 were extracted using the online project (NCIDP) developed by Molecular Dynamics and NMR Group / University of Groningen, The Netherlands¹⁰⁰.

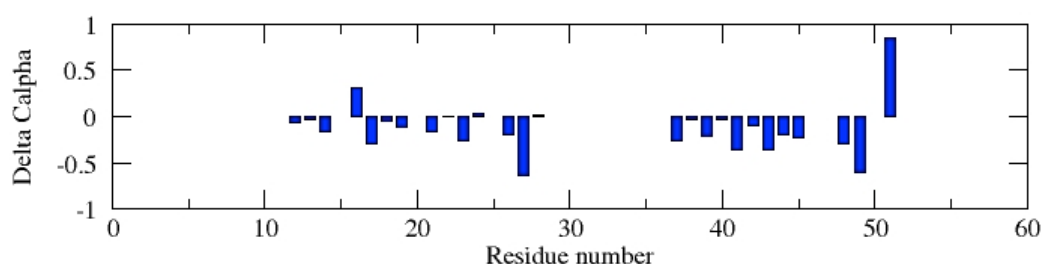


FIG. 61 C_α chemical shift difference between the assigned TCF4 peptides at pH=4.5 and the predicted random coil chemical shift values. The figures shows the absence of secondary structure elements from the full length free TCF4.

3.2.9. Secondary structure propensity for TCF4 (SSP)

From the pdb structure “entry 1JDH”, it’s possible to observe that TCF4 has structural variability. It forms 2 alpha helices regions upon binding to β -Catenin. These areas are residues 21-31 and 38-49. This structure modulation on TCF4 can be rationalized based on the environmental effect, which is the binding partner in this case. The residues on catenin that are close to these regions on TCF4, which form alpha helices, are mainly lysine residues. When these residues are protonated, they are acting as conjugated acid (H-bond donor or simply H-donor like any acid), and this in turn results in stabilizing the formation of alpha-helix structures. Hence, at low pH, which is mimicking of the presence of protonated basic residues, partial protonation to the negatively charged side chains on TCF4 stabilizes the formation of alpha helix structures. Figure number 62

Experimental

shows the structure of the complex of catenin-TCF4 illustrating the basic residues on catenin that approach the alpha helices on TCF4.

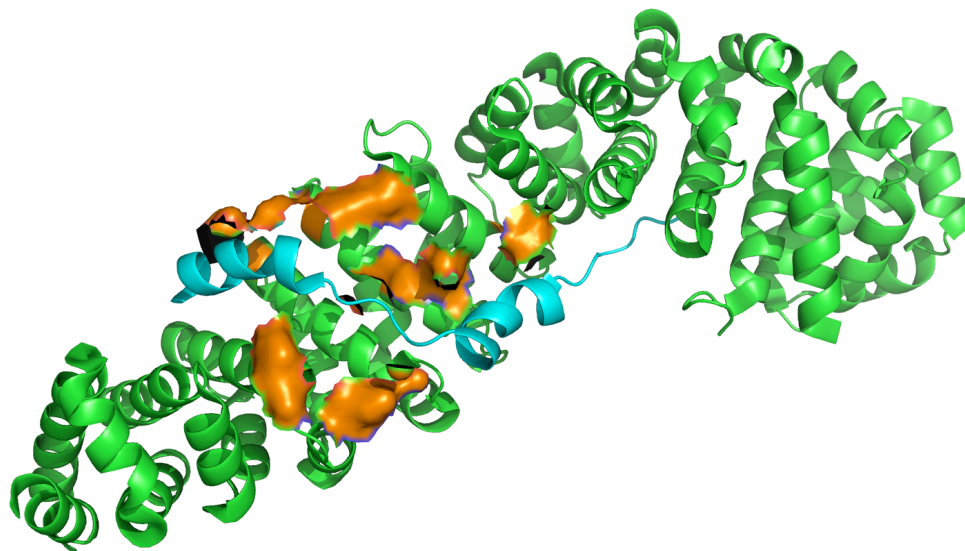


FIG. 62 Crystal structure of β -Catenin-TCF4 complex (pdb entry: 1JDH). The catenin is highlighted in green, TCF4 in blue and the basic residues on catenin that are in contact with TCF4 are in orange.

In order to investigate this hypothesis and to predict the possible secondary structure elements in TCF4, the assignment of the protein was also done at pH=2 (to mimic the partial-protonation event upon binding). The Chemical shifts of N, $^1\text{H}^{\text{N}}$, $\text{C}\alpha$ and $\text{C}\beta$ were extracted and compared regarding SSP with the ones obtained at pH=4.5 using the pearl script described by Marsh *et al*⁷. It was shown that at pH=2, two regions on TCF4 (residues 23-27, and residues 36-42) showed alpha helix secondary structure propensity, while the rest of residues are in beta or random coil propensity. These findings match with the crystal structure in which two regions of TCF4 form alpha helices upon (partial protonation=binding).

Experimental

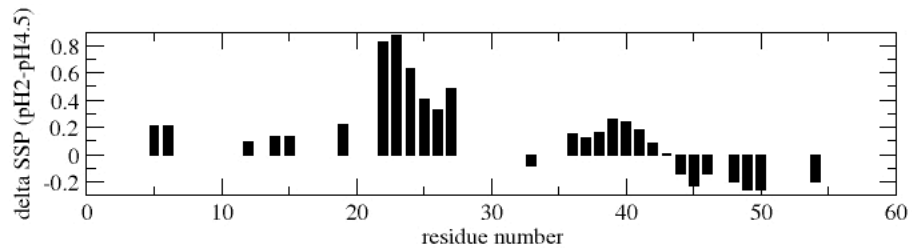


FIG. 63 Plot of the difference in secondary structure propensity of TCF4 between pH 2 and pH 4.5. It's shown that at pH=2, regions of residues 23-27 and 36-42 have alpha helix propensity in contrast to the same regions at pH=4.5.

3.2.10. Crystallization & structure elucidation of armadillo repeat region of human β -Catenin

The following robotics have been employed for the crystallization work: a) an Alchemist II liquid handling robot (Rigaku) for pipetting of non-commercial 24-well Linbro format (hanging drop plates), b) a Phoenix liquid handling system (Art Robbins Instruments) for preparation 96-well crystallization trials using commercially available conditions, c) a Minstrel DT imaging system (Rigaku) equipped with the Atlantis software for automatic imaging. Images were screened using the CrystalTrak software (Rigaku). For commercial screens and optimizations in 96-well format pipetted by the Phoenix liquid handling system, 3-well IntelliPlates (Art Robbins instruments) were used. 24-well optimizations, which were set up by hand, were performed in 24-well Linbro format plates, pregreased (Crystalgen). Crystals were stored in a crystallization room at either 22 °C.

The protein sample was expressed and purified in tris-buffer and concentrated to 6.8mg/ml concentration. Different commercial JCSG conditions have been tried (QIAGEN). The protein was crystallized upon using screen in the following conditions: 0.2 M potassium thiocyanate, and 20 % w/v polyethyleneglycol 3350 at pH=7.4. Optimization windows were tried to get the optimal crystals. The crystals obtained from condition of monoethylene polyethyleneglycol 20% w/v 2000 + potassium thiocyanate 0.1 M were sent for data collection. The crystals were kept in a solution of 40% PEG 3350 and 0.1 M potassium thiocyanate for cryogenic protection.

Data collection:

Diffraction data were collected either in the laboratory using a Bruker Microstar (Bruker AXS Inc.) rotating anode at 1.54 Å wavelength or using ESRF

Experimental

beamline (Grenoble, France) at wavelength 0.939 Å.

The phase problem was solved using Zebra fish β -Catenin as a model for molecular replacement (PDB entry: 2Z6G)¹⁰¹ by software PHASER¹⁰². Model refinement was done using phenix.refine from the PHENIX suite¹⁰³. The built model validation was done using a web server Molprobity¹⁰⁴. Details of data collection and refinement statistics are illustrated in table 7.

Table 7: Data collection and refinement statistics for human armadillo repeat region of β -Catenin.

Source	ID14-4 (ESRF)
Wavelength (Å)	0.939
Resolution (Å)	45-2.4 (245-2.4)
Space group	P ₂₁₂₁₂₁
Unit cell (Å)	a= 51.326 b= 73.62 c=133.63
Molecules/ a.u.	1
Unique reflections	20594 (1356)
Completeness (%)	99.1 (90.2)
I/ σ (I)	10.6 (1.8)
Software used for integration	XDS
Software used for scaling	XSCALE
Software used for phasing	PHASER

Experimental

$R_{\text{crys}}/R_{\text{free}}$	0.183/0.266
R.M.S.D bonds (Å)	0.008
R.M.S.D angles (Å)	1.13
Average B factor protein (Å ²)	52.06
Average B factor solvent (Å ²)	45.6
Software used for refinement	Refmac/Phenix

Results & Discussion

The structure of human armadillo repeats region of β -Catenin was solved at a resolution 2.4 Å. After solving the phase problem using molecular replacement and building the refined model, the structure showed conserved 12 repeats of alpha helices forming a super helix, each repeat is 3 alpha helices. The region of residues 559-567 showed different conformations although this isn't the binding site of TCF4.

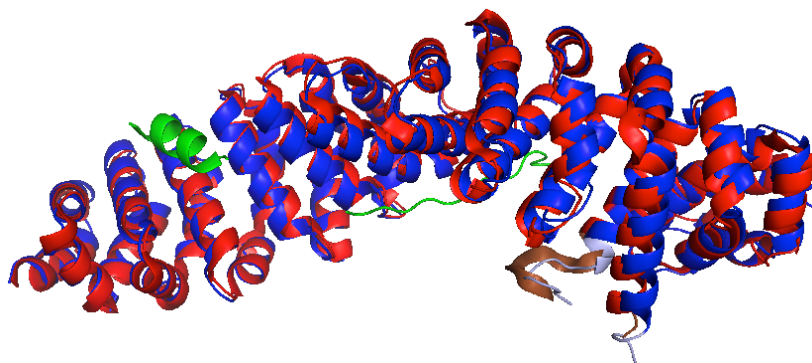


FIG. 64 Structure of the solved *free* human armadillo repeat domain of β -Catenin at resolution 2.4 Å is shown in blue, while the structure of the human one *bound* to TCF4 is shown in red, TCF4 is shown in green (pdb entry: 1JPW). The area of residues between 559-567 showed different conformation (highlighted in light blue for the free form and in brown for the bound form).

Experimental

Query	133	HAVVNLINIQDDAELATRAIPELTKLLNDEDEQVVVNKAAVMVHQLSK	179
		HAVVNLINIQDDAELATRAIPELTKLLNDEDEQVVVNKAAVMVHQLSK	
Sbjct	134	HAVVNLINIQDDAELATRAIPELTKLLNDEDEQVVVNKAAVMVHQLSK	180
Query	180	KEASRHAIMRSPQMVSIAIVRTMONTNDVETARCTSGTLHNLSSHREGLLAIFKSGGIPAL	239
		KEASRHAIMRSPQMVSIAIVRTMONTNDVETARCT+GTLHNLSSHREGLLAIFKSGGIPAL	
Sbjct	181	KEASRHAIMRSPQMVSIAIVRTMONTNDVETARCTAGTLHNLSSHREGLLAIFKSGGIPAL	240
Query	240	VNMLGSPVDSVLFHAITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAITTDC	299
		V+MLGSPVDSVLF+AITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAITTDC	
Sbjct	241	VKMLGSPVDSVLFYAITTLHNLLLHQEGAKMAVRLAGGLQKMVALLNKTNVKFLAITTDC	300
Query	300	LQILAYGNQESKLIILASGGPQALVNIMRTYTYEKLLWTTSRVLKVLVSVCSNKPATIVEA	359
		LQILAYGNQESKLIILASGGPQALVNIMRTYTYEKLLWTTSRVLKVLVSVCSNKPATIVEA	
Sbjct	301	LQILAYGNQESKLIILASGGPQALVNIMRTYTYEKLLWTTSRVLKVLVSVCSNKPATIVEA	360
Query	360	GGMQALGLHLTDPSQRLVQNCLWTLRNLSDAATKQEGMEGLLGLTLVQLLGSDDINVTCA	419
		GGMQALGLHLTDPSQRLVQNCLWTLRNLSDAATKQEGMEGLLGLTLVQLLGSDDINVTCA	
Sbjct	361	GGMQALGLHLTDPSQRLVQNCLWTLRNLSDAATKQEGMEGLLGLTLVQLLGSDDINVTCA	420
Query	420	AGILSNLTCNNYKKNMMVCQVGGIEALVRTVLRAGDREDITEPAICALRHILSRHQDAEM	479
		AGILSNLTCNNYKKNMMVCQVGGIEALVRTVLRAGDREDITEPAICALRHILSRHQ+AEM	
Sbjct	421	AGILSNLTCNNYKKNMMVCQVGGIEALVRTVLRAGDREDITEPAICALRHILSRHQEAEM	480
Query	480	AQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAI PRLVQLL	539
		AQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAI PRLVQLL	
Sbjct	481	AQNAVRLHYGLPVVVKLLHPPSHWPLIKATVGLIRNLALCPANHAPLREQGAI PRLVQLL	540
Query	540	VRAHQDTQRRRTSMGGTQQQFVEGVRMEEIVEACTGALHILARDIHNRIVIRGLNTIPLFV	599
		VRAHQDTQRRRTSMGGTQQQFVEGVRMEEIVE+CTGALHILARD+HNRIVIRGLNTIPLFV	
Sbjct	541	VRAHQDTQRRRTSMGGTQQQFVEGVRMEEIVEGCTGALHILARDVHNRIVIRGLNTIPLFV	600
Query	600	QLLYSPIENIQRVAAGVLCELAQDKEAAEAIEAEGATAPLTELHLSRNEGVATYAAAVLF	659
		QLLYSPIENIQRVAAGVLCELAQDKEAAEAIEAEGATAPLTELHLSRNEGVATYAAAVLF	
Sbjct	601	QLLYSPIENIQRVAAGVLCELAQDKEAAEAIEAEGATAPLTELHLSRNEGVATYAAAVLF	660
Query	660	RMSEDKPDYK	670
		RMSEDKPDYK	
Sbjct	661	RMSEDKPDYK	671

FIG. 65 alignment of Zebra fish catenin (query) and the human catenin (sbjct). The different residues are highlighted with red.

REFERENCES

References

1. Richards, F. M., Linderstrom-Lang and the Carlsberg Laboratory - the View of a Postdoctoral Fellow in 1954. *Protein Sci* **1992**, *1* (12), 1721-1730.
2. Sobolevsky, Y.; Trifonov, E. N., Protein modules conserved since LUCA. *J Mol Evol* **2006**, *63* (5), 622-634.
3. Pauling, L.; Corey, R. B.; Branson, H. R., The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America* **1951**, *37* (4), 205-11.
4. Levitt, M.; Chothia, C., Structural Patterns in Globular Proteins. *Nature* **1976**, *261* (5561), 552-558.
5. Perutz, M. F.; Muirhead, H.; Cox, J. M.; Goaman, L. C., Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature* **1968**, *219* (5150), 131-9.
6. Phillips, D. C., The three-dimensional structure of an enzyme molecule. *Scientific American* **1966**, *215* (5), 78-90.
7. Marsh, J. A.; Singh, V. K.; Jia, Z. C.; Forman-Kay, J. D., Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation. *Protein Sci* **2006**, *15* (12), 2795-2804.
8. Fong, J. H.; Shoemaker, B. A.; Garbuzynskiy, S. O.; Lobanov, M. Y.; Galzitskaya, O. V.; Panchenko, A. R., Intrinsic Disorder in Protein Interactions: Insights From a Comprehensive Structural Analysis. *Plos Comput Biol* **2009**, *5* (3).
9. Dyson, H. J.; Wright, P. E., Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Bio* **2005**, *6* (3), 197-208.
10. Naranjo, Y.; Pons, M.; Konrat, R., Meta-structure correlation in protein space unveils different selection rules for folded and intrinsically disordered proteins. *Mol Biosyst* **2012**, *8* (1), 411-416.
11. Dyson, H. J.; Wright, P. E., Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv Protein Chem* **2002**, *62*, 311-340.
12. Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G., Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97* (16), 8868-+.
13. Konrat, R., The protein meta-structure: a novel concept for chemical and molecular biology. *Cellular and molecular life sciences : CMLS* **2009**, *66* (22), 3625-39.
14. Kontaxis, G.; Delaglio, F.; Bax, A., Molecular fragment replacement approach to protein structure determination by chemical shift and dipolar homology database mining. *Methods in enzymology* **2005**, *394*, 42-78.
15. Pauling, L.; Corey, R. B., Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America* **1951**, *37* (5), 235-40.
16. Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C., Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **1960**, *185* (4711), 422-7.

References

17. Jiao, M.; Li, H. T.; Chen, J.; Minton, A. P.; Liang, Y., Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria. *Biophysical journal* **2010**, *99* (3), 914-23.
18. Velazquez-Campoy, A.; Freire, E., Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nature protocols* **2006**, *1* (1), 186-91.
19. Balamurugan, R.; Dekker, F. J.; Waldmann, H., Design of compound libraries based on natural product scaffolds and protein structure similarity clustering (PSSC). *Mol Biosyst* **2005**, *1* (1), 36-45.
20. Bienstock, R. J.; American Chemical Society. Division of Chemical Information.; American Chemical Society. Division of Computers in Chemistry., *Library design, search methods, and applications of fragment-based drug design*. American Chemical Society: Washington, DC, 2011; p p.
21. Hann, M. M.; Oprea, T. I., Pursuing the leadlikeness concept in pharmaceutical research. *Current opinion in chemical biology* **2004**, *8* (3), 255-263.
22. Siegal, G.; Ab, E.; Schultz, J., Integration of fragment screening and library design. *Drug Discovery Today* **2007**, *12* (23-24), 1032-1039.
23. Gianti, E.; Sartori, L., Identification and Selection of "Privileged Fragments" Suitable for Primary Screening. *J Chem Inf Model* **2008**, *48* (11), 2129-2139.
24. Proll, F.; Fechner, P.; Proll, G., Direct optical detection in fragment-based screening. *Analytical and bioanalytical chemistry* **2009**, *393* (6-7), 1557-62.
25. Hajduk, P. J., SAR by NMR: putting the pieces together. *Molecular interventions* **2006**, *6* (5), 266-72.
26. Ozawa, M.; Baribault, H.; Kemler, R., The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. *The EMBO journal* **1989**, *8* (6), 1711-7.
27. Benjamin, J. M.; Nelson, W. J., Bench to bedside and back again: molecular mechanisms of alpha-catenin function and roles in tumorigenesis. *Seminars in cancer biology* **2008**, *18* (1), 53-64.
28. Choi, H. J.; Weis, W. I., Structure of the armadillo repeat domain of plakophilin 1. *Journal of molecular biology* **2005**, *346* (1), 367-76.
29. Shapiro L., W. W. I., Structure and biochemistry of cadherins and catenins. *Cold Spring Harb. Perspect. Biol.* **2009**, *1* (a003053).
30. Phillips, B. T.; Kimble, J., A new look at TCF and beta-catenin through the lens of a divergent *C. elegans* Wnt pathway. *Developmental cell* **2009**, *17* (1), 27-34.
31. McCrea, P. D.; Gu, D., The catenin family at a glance. *Journal of cell science* **2010**, *123* (Pt 5), 637-42.
32. Mosimann, C.; Hausmann, G.; Basler, K., Beta-catenin hits chromatin: regulation of Wnt target gene activation. *Nature reviews. Molecular cell biology* **2009**, *10* (4), 276-86.
33. Mareel, M.; Boterberg, T.; Noe, V.; VanHoorde, L.; Vermeulen, S.; Bruyneel, E.; Bracke, M., E-cadherin/catenin/cytoskeleton complex: A regulator of cancer invasion. *J Cell Physiol* **1997**, *173* (2), 271-274.
34. Camilli, T. C.; Weeraratna, A. T., Striking the target in Wnt-y conditions: intervening in Wnt signaling during cancer progression. *Biochemical pharmacology* **2010**, *80* (5), 702-11.

References

35. Tanaka, K.; Kitagawa, Y.; Kadowaki, T., Drosophila segment polarity gene product porcupine stimulates the posttranslational N-glycosylation of wingless in the endoplasmic reticulum. *Journal of Biological Chemistry* **2002**, *277* (15), 12816-12823.
36. Franch-Marro, X.; Wendler, F.; Griffith, J.; Maurice, M. M.; Vincent, J. P., In vivo role of lipid adducts on Wingless. *Journal of cell science* **2008**, *121* (10), 1587-1592.
37. Huang, H.; He, X., Wnt/beta-catenin signaling: new (and old) players and new insights. *Current opinion in cell biology* **2008**, *20* (2), 119-25.
38. Hikasa, H.; Ezan, J.; Itoh, K.; Li, X.; Klymkowsky, M. W.; Sokol, S. Y., Regulation of TCF3 by Wnt-dependent phosphorylation during vertebrate axis specification. *Developmental cell* **2010**, *19* (4), 521-32.
39. Liu, C.; Li, Y.; Semenov, M.; Han, C.; Baeg, G. H.; Tan, Y.; Zhang, Z.; Lin, X.; He, X., Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **2002**, *108* (6), 837-47.
40. Endo, K.; Ueda, T.; Ueyama, J.; Ohta, T.; Terada, T., Immunoreactive E-cadherin, alpha-catenin, beta-catenin, and gamma-catenin proteins in hepatocellular carcinoma: relationships with tumor grade, clinicopathologic parameters, and patients' survival. *Human pathology* **2000**, *31* (5), 558-65.
41. Akiyama, T., Wnt/beta-catenin signaling. *Cytokine & growth factor reviews* **2000**, *11* (4), 273-82.
42. Tetsu, O.; McCormick, F., Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **1999**, *398* (6726), 422-6.
43. Lepourcelet, M.; Chen, Y. N.; France, D. S.; Wang, H.; Crews, P.; Petersen, F.; Bruseo, C.; Wood, A. W.; Shivdasani, R. A., Small-molecule antagonists of the oncogenic Tcf/beta-catenin protein complex. *Cancer cell* **2004**, *5* (1), 91-102.
44. Sandgren, E. P.; Quafe, C. J.; Pinkert, C. A.; Palmiter, R. D.; Brinster, R. L., Oncogene-induced liver neoplasia in transgenic mice. *Oncogene* **1989**, *4* (6), 715-24.
45. Chen, B.; Dodge, M. E.; Tang, W.; Lu, J.; Ma, Z.; Fan, C. W.; Wei, S.; Hao, W.; Kilgore, J.; Williams, N. S.; Roth, M. G.; Amatruda, J. F.; Chen, C.; Lum, L., Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nature chemical biology* **2009**, *5* (2), 100-7.
46. Athar, M.; Back, J. H.; Tang, X.; Kim, K. H.; Kopelovich, L.; Bickers, D. R.; Kim, A. L., Resveratrol: a review of preclinical studies for human cancer prevention. *Toxicology and applied pharmacology* **2007**, *224* (3), 274-83.
47. Hope, C.; Planutis, K.; Planutiene, M.; Moyer, M. P.; Johal, K. S.; Woo, J.; Santoso, C.; Hanson, J. A.; Holcombe, R. F., Low concentrations of resveratrol inhibit Wnt signal throughput in colon-derived cells: implications for colon cancer prevention. *Molecular nutrition & food research* **2008**, *52 Suppl 1*, S52-61.
48. Zhang, W.; Sviripa, V.; Kril, L. M.; Chen, X.; Yu, T.; Shi, J.; Rychahou, P.; Evers, B. M.; Watt, D. S.; Liu, C., Fluorinated N,N-dialkylaminostilbenes for Wnt pathway inhibition and colon cancer repression. *Journal of medicinal chemistry* **2011**, *54* (5), 1288-97.
49. Oshima, M.; Dinchuk, J. E.; Kargman, S. L.; Oshima, H.; Hancock, B.; Kwong, E.; Trzaskos, J. M.; Evans, J. F.; Taketo, M. M., Suppression of intestinal polyposis in Apc delta716 knockout mice by inhibition of cyclooxygenase 2 (COX-2). *Cell* **1996**, *87* (5), 803-9.

References

50. Steinbach, G.; Lynch, P. M.; Phillips, R. K.; Wallace, M. H.; Hawk, E.; Gordon, G. B.; Wakabayashi, N.; Saunders, B.; Shen, Y.; Fujimura, T.; Su, L. K.; Levin, B., The effect of celecoxib, a cyclooxygenase-2 inhibitor, in familial adenomatous polyposis. *The New England journal of medicine* **2000**, *342* (26), 1946-52.
51. Roth, S. H.; Anderson, S., The NSAID Dilemma: Managing Osteoarthritis in High-Risk Patients. *Physician Sportsmed* **2011**, *39* (3), 62-74.
52. Fernandez, C. J., W, New approaches for NMR screening in drug discovery. *Drug Discovery Today* **2004**, *1* (3), 277-283.
53. Siriwardena, A. H.; Tian, F.; Noble, S.; Prestegard, J. H., A straightforward NMR-spectroscopy-based method for rapid library screening. *Angew Chem Int Ed Engl* **2002**, *41* (18), 3454-7.
54. Jahnke, W., Spin labels as a tool to identify and characterize protein-ligand interactions by NMR spectroscopy. *Chembiochem : a European journal of chemical biology* **2002**, *3* (2-3), 167-73.
55. Kay, L. E., NMR studies of protein structure and dynamics. *J Magn Reson* **2005**, *173* (2), 193-207.
56. Mayer, M.; Meyer, B., Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society* **2001**, *123* (25), 6108-17.
57. Dalvit, C., Ligand- and substrate-based F-19 NMR screening: Principles and applications to drug discovery. *Progress in nuclear magnetic resonance spectroscopy* **2007**, *51* (4), 243-1.
58. Kumar, A.; Wagner, G.; Ernst, R. R.; Wuthrich, K., Buildup Rates of the Nuclear Overhauser Effect Measured by Two-Dimensional Proton Magnetic-Resonance Spectroscopy - Implications for Studies of Protein Conformation. *Journal of the American Chemical Society* **1981**, *103* (13), 3654-3658.
59. Auer, R.; Kloiber, K.; Vavrinska, A.; Geist, L.; Coudeville, N.; Konrat, R., Pharmacophore mapping via cross-relaxation during adiabatic fast passage. *Journal of the American Chemical Society* **2010**, *132* (5), 1480-1.
60. Desvaux, H.; Berthault, P.; Birlirakis, N.; Goldman, M., Off-Resonance Roesy for the Study of Dynamic Processes. *J Magn Reson Ser A* **1994**, *108* (2), 219-229.
61. Konrat, R.; Tollinger, M., Heteronuclear relaxation in time-dependent spin systems: (15)N-T1 (rho) dispersion during adiabatic fast passage. *Journal of biomolecular NMR* **1999**, *13* (3), 213-21.
62. Means, G. E.; Feeney, R. E., Reductive Alkylation of Amino Groups in Proteins. *Biochemistry* **1968**, *7* (6), 2192-&.
63. Sledz, P.; Zheng, H. P.; Murzyn, K.; Chruszcz, M.; Zimmerman, M. D.; Chordia, M. D.; Joachimiak, A.; Minor, W., New surface contacts formed upon reductive lysine methylation: Improving the probability of protein crystallization. *Protein Sci* **2010**, *19* (7), 1395-1404.
64. Rayment, I., Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Method Enzymol* **1997**, *276*, 171-179.
65. Kurinov, I. V.; Mao, C.; Irvin, J. D.; Uckun, F. M., X-ray crystallographic analysis of pokeweed antiviral protein-II after reductive methylation of lysine residues. *Biochem Bioph Res Co* **2000**, *275* (2), 549-552.
66. Abraham, S. J.; Hoheisel, S.; Gaponenko, V., Detection of protein-ligand interactions by NMR using reductive methylation of lysine residues. *Journal of biomolecular NMR* **2008**, *42* (2), 143-148.

References

67. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A., Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *J Magn Reson* **2011**, *213* (2), 423-41.
68. Bax, A.; Ikura, M., An efficient 3D NMR technique for correlating the proton and ¹⁵N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly ¹⁵N/¹³C enriched proteins. *Journal of biomolecular NMR* **1991**, *1* (1), 99-104.
69. Grzesiek, S.; Bax, A., An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *Journal of Magnetic Resonance* **1992**, *99* (1), 201-207.
70. Clubb, R. T.; Thanabal, V.; Wagner, G., A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue H-1(N), N-15, and C-13(') Chemical-Shifts in N-15-C-13-Labeled Proteins. *Journal of Magnetic Resonance* **1992**, *97* (1), 213-217.
71. Eliseev, A. V., Emerging approaches to target-assisted screening of combinatorial mixtures. *Current opinion in drug discovery & development* **1998**, *1* (1), 106-15.
72. Otto, S.; Furlan, R. L. E.; Sanders, J. K. M., Dynamic combinatorial chemistry. *Drug Discovery Today* **2002**, *7* (2), 117-125.
73. Swann, P. G.; Casanova, R. A.; Desai, A.; Frauenhoff, M. M.; Urbancic, M.; Slomczynska, U.; Hopfinger, A. J.; LeBreton, G. C.; Venton, D. L., Nonspecific protease-catalyzed hydrolysis synthesis of a mixture of peptides: Product diversity and ligand amplification by a molecular trap. *Biopolymers* **1996**, *40* (6), 617-625.
74. Huc, I.; Lehn, J. M., Virtual combinatorial libraries: Dynamic generation of molecular and supramolecular diversity by self-assembly (vol 94, pg 2106, 1997). *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94* (15), 8272-8272.
75. Eliseev, A. V.; Nelen, M. I., Use of molecular recognition to drive chemical evolution .1. Controlling the composition of an equilibrating mixture of simple arginine receptors. *Journal of the American Chemical Society* **1997**, *119* (5), 1147-1148.
76. Hioki, H.; Still, W. C., Chemical evolution: A model system that selects and amplifies a receptor for the tripeptide (D)Pro(L)Val(D)Val. *Journal of Organic Chemistry* **1998**, *63* (4), 904-905.
77. Klekota, B.; Miller, B. L., Selection of DNA-binding compounds via multistage molecular evolution. *Tetrahedron* **1999**, *55* (39), 11687-11697.
78. Meyer, B.; Peters, T., NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Edit* **2003**, *42* (8), 864-890.
79. Caraballo, R.; Dong, H.; Ribeiro, J. P.; Jimenez-Barbero, J.; Ramstrom, O., Direct STD NMR Identification of beta-Galactosidase Inhibitors from a Virtual Dynamic Hemithioacetal System. *Angew Chem Int Edit* **2010**, *49* (3), 589-593.
80. Buurma, N. J.; Haq, I., Advances in the analysis of isothermal titration calorimetry data for ligand-DNA interactions. *Methods* **2007**, *42* (2), 162-172.
81. Velazquez-Campoy, A.; Freire, E., ITC in the post-genomic era... ? Priceless. *Biophys Chem* **2005**, *115* (2-3), 115-124.

References

82. O'Neill, M. A. A.; Gaisford, S., Application and use of isothermal calorimetry in pharmaceutical development. *Int J Pharmaceut* **2011**, *417* (1-2), 83-93.
83. Trosset, J. Y.; Dalvit, C.; Knapp, S.; Fasolini, M.; Veronesi, M.; Mantegani, S.; Gianellini, L. M.; Catana, C.; Sundstrom, M.; Stouten, P. F. W.; Moll, J. K., Inhibition of protein-protein interactions: The discovery of druglike beta-catenin inhibitors by combining virtual and biophysical screening. *Proteins* **2006**, *64* (1), 60-67.
84. Schmidt, F. R., Recombinant expression systems in the pharmaceutical industry. *Appl Microbiol Biot* **2004**, *65* (4), 363-372.
85. Goeddel, D. V., Systems for Heterologous Gene-Expression. *Methods in enzymology* **1990**, *185*, 3-7.
86. Goulding, C. W.; Perry, L. J., Protein production in Escherichia coli for structural studies by X-ray crystallography. *J Struct Biol* **2003**, *142* (1), 133-143.
87. Petsch, D.; Anspach, F. B., Endotoxin removal from protein solutions. *J Biotechnol* **2000**, *76* (2-3), 97-119.
88. Baneyx, F., Recombinant protein expression in Escherichia coli. *Curr Opin Biotech* **1999**, *10* (5), 411-421.
89. Greene, J. J., Host cell compatibility in protein expression. *Methods Mol Biol* **2004**, *267*, 3-14.
90. Sitaraman, K.; Esposito, D.; Klarmann, G.; Le Grice, S. F.; Hartley, J. L.; Chatterjee, D. K., A novel cell-free protein synthesis system. *J Biotechnol* **2004**, *110* (3), 257-63.
91. Hansen, L. H.; Knudsen, S.; Sorensen, S. J., The effect of the lacY gene on the induction of IPTG inducible promoters, studied in Escherichia coli and Pseudomonas fluorescens. *Current microbiology* **1998**, *36* (6), 341-7.
92. Ruegg, U. T.; Rudinger, J., Reductive cleavage of cystine disulfides with tributylphosphine. *Methods in enzymology* **1977**, *47*, 111-6.
93. Terpe, K., Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **2003**, *60* (5), 523-33.
94. Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G., Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **1975**, *258* (5536), 598-9.
95. Gill, S. C. v. H., P. H., *Anal. Biochem.* **1989**, *182* (2), 319-326.
96. Laemmli, U. K., Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. *Nature* **1970**, *227* (5259), 680-&.
97. Henen, M. A.; Coudevylle, N.; Geist, L.; Konrat, R., Toward Rational Fragment-Based Lead Design without 3D Structures. *Journal of medicinal chemistry* **2012**, *55* (17), 7909-19.
98. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of biomolecular NMR* **1995**, *6* (3), 277-293.
99. Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M., Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. *Chem-Eur J* **2010**, *16* (26), 7803-7812.
100. Tamiola, K.; Acar, B.; Mulder, F. A. A., Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **2010**, *132* (51), 18000-18003.

References

101. Xing, Y.; Takemaru, K. I.; Liu, J.; Berndt, J. D.; Zheng, J. J.; Moon, R. T.; Xu, W. Q., Crystal structure of a full-length beta-catenin. *Structure* **2008**, *16* (3), 478-487.
102. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of applied crystallography* **2007**, *40*, 658-674.
103. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **2010**, *66* (Pt 2), 213-21.
104. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research* **2007**, *35* (Web Server issue), W375-83.

Zusammenfassung

Abweichungen im Wnt/Beta-Catenin Signalweg kommen in vielen verschiedenen menschlichen Krebserkrankungen vor. Unter anderem wird dieser Signalweg in Darm- und Eierstockkrebs aktiviert. Aus diesem Grund haben Inhibitoren der TCF4/Beta-Catenin Interaktion großes Potential als Antitumormittel. Mit Hilfe des vor kurzem beschriebenen „Meta-Structure“ Ansatzes haben wir eine Reihe von niedermolekularen Molekülen identifiziert, die an Beta-Catenin binden und als Ausgagspunkt für zukünftige Entwicklungen dienen könnten. Um diese Interaktionen zu testen, wurden verschiedene Methoden angewandt: 1D ^1H -STD, 1D AFP NOESY, 2D NOESY und ^{19}F -NMR. Auch die Fähigkeit der neu identifizierten Liganden mit dem bekannten Bindungspartner TCF4 von Beta-Catenin zu konkurrieren wurde getestet.

Zur weiteren Untersuchung der Interaktionen und um weitere neue Liganden synthetisieren und identifizieren zu können, wurde ein sogenannter „Dynamic Combinatorial Library“ Ansatz (Dynamische, kombinierbare Chemikalienbibliothek) adaptiert und mithilfe . Weiters wurden zwei- und dreidimensionale NMR-Experimente durchgeführt, mit welchen die ^1N - ^{15}H Signale des TCF4 Proteinrückgrats, und zweier separate Teile hiervon, bei pHs von 4,5 und 2,0 zugeordnet werden konnten. Ein Vergleich der so bestimmten Frequenzen mit den errechneten Werten eines theoretischen, gänzlich unstrukturierten Proteins gleicher Aminosäuresequenz („Secondary Structure Propensity“) erlaubte es, Aussagen über die Wahrscheinlichkeit von Sekundärstrukturen bei verschiedenen pH-Werten zu machen. Zusätzlich zu der Untersuchung neuer Liganden wurde auch die Struktur des TCF4/Beta-Catenin Komplexes näher untersucht. Hierfür wurden sowohl röntgenkristallographische Experimente verwendet, als auch die selektive Reduzierung von Lysin-Seitenkette mit $^{13}\text{HCHO}$, welche ein Messen des daraus resultierenden ^{13}C - ^1H Spektrums erlaubt. Um die Relevanz meiner Untersuchungen zu unterstreichen, werden die sich ergebenden Möglichkeiten für die Entwicklung eines neuen Antitumormittels mit möglichst hohem Sicherheitsprofil erörtert.

Curriculum Vitae

Morkos A. Henen

Department of Structural and Computational Biology
Max F. Perutz Laboratories, Prof. Robert Konrat group, Vienna Biocenter
Campus Vienna Biocenter 5, 1030 Wien, AUSTRIA

morkos.henen@univie.ac.at

33 years old

CAREER HISTORY

Oct 2009 to date: PhD student at the Max F. Perutz Laboratories

“Chemistry and NMR studies of the Beta-Catenin-TCF4 interaction – a potential antitumor target”

Supervised by Prof. Robert Konrat,
Department of Structural & Computational Biology, Max. F.
Perutz Laboratories, Vienna, Austria

2000/2009: Associate lecturer of pharmaceutical organic chemistry-
Faculty of Pharmacy-Mansoura University, Egypt.

EDUCATION

2008: Master degree in pharmaceutical organic chemistry
Faculty of pharmacy- Mansoura University Egypt

2000: Bachelor in pharmaceutical sciences (Very good, with honor
degree)
Faculty of pharmacy- Mansoura University, Egypt.

TEACHING EXPERIENCE

2000/2009: - Basic practical courses of organic chemistry for pharmacy
students (pre, first and second year students) – 20 h/week

Scholarships

Oct2009-Jun2012: - Grant from the Austrian agency for international mobility
and cooperation in education, science and research.
Vienna/Austria

RESEARCH EXPERIENCE & INTERESTS

- Experience in using NMR techniques to measure protein-ligand interaction. I also have experience in linking organic chemistry approaches with structural biology like protein reductive methylation and dynamic combinatorial libraries as tools for fragment based drug design.

- In addition, I have experience in using X-ray for protein structure elucidation.
- In my research work, I have adopted the meta-structure approach to find new ligands for β -Catenin; a promising antitumor target. Different ligands have been tested for binding with the targeted protein.
- Moreover, I have worked on designing a small dynamic combinatorial library to synthesize and test new ligands for catenin adopting the fragment based drug design approach. For testing the binding, different NMR techniques have been addressed like measuring cross relaxation during adiabatic fast passage, saturation transfer difference NMR and ^{19}F -NMR.
- I was also interested in using 2D and 3D NMR techniques to assign the binding partner of beta-catenin; TCF4. Therefore, the protein was assigned and its secondary structure propensity was also determined.
- Eventually, I used ^{13}C -selective reductive methylation of lysine residues to map the binding site on β -Catenin using 2D ^1H - ^{13}C HSQC.

PUBLICATIONS

2. Towards rational fragment-based lead discovery without 3D structures

Henen M. A., Coudeville N., Geist L., Konrat R., *J. Med. Chem.* (2012) 55 : 7909-7919

1. [1,2,4]Triazolo[4,3-a]quinoxaline: synthesis, antiviral, and antimicrobial activities

Henen M.A., El Bialy S, Goda F.E., Nasr M.A., Eisa H.M., *Med. Chem. Res.* (2012) 21 :2368-2378

CONFERENCES

1- *XXVth International Conference on Magnetic Resonance in Biological System, Lyon, 19-24th August 2012*

2- *Magnetic Moments in Central Europe, Slovakia, March 2011*